Master Thesis
Computer Science
Thesis No: MCS-2010-24
2012

Speech Translation into Pakistan Sign Language

Ahmed Abdul Haseeb
Asim Illyas

School of Computing
Blekinge Institute of Technology
SE – 371 79 Karlskrona
Sweden

This thesis is submitted to the School of Computing at Blekinge Institute of Technology in partial fulfillment of the requirements for the degree of Master of Science in Computer Science. The thesis is equivalent to 20 weeks of full time studies.

Contact Information:

Authors:
Ahmed Abdul Haseeb
E-mail: abdulhaseeb.ahmad@gmail.com
Asim Ilyas
E-mail: asimilyas85@hotmail.com

University advisor:

Prof. Sara Eriksén
School of Computing,

Blekinge Institute of Technology	Internet	: www.bth.se/com
SE – 371 79 Karlskrona	Phone	: +46 457 38 50 00
Sweden	Fax	: + 46 457 271 25

ABSTRACT

Context: Communication is a primary human need and language is the medium for this. Most people have the ability to listen and speak and they use different languages like Swedish, Urdu and English etc. to communicate. Hearing impaired people use signs to communicate. Pakistan Sign Language (PSL) is the preferred language of the deaf in Pakistan. Currently, human PSL interpreters are required to facilitate communication between the deaf and hearing; they are not always available, which means that communication among the deaf and other people may be impaired or nonexistent. In this situation, a system with voice recognition as an input and PSL as an output will be highly helpful.
Objectives: As part of this thesis, we explore challenges faced by deaf people in everyday life while interacting with unimpaired. We investigate state of art work done in this area. This study explores speech recognition and Machine translation techniques to devise a generic and automated system that converts English speech to PSL. A prototype of the proposed solution is developed and validated.
Methods: Three step investigation is done as part of thesis work. First, to understand problem itself, interviews were conducted with the domain experts. Secondly, from literature review, it is investigated whether any similar or related work has already been done, state of the art technologies like Machine translation, speech recognition engines and Natural language processing etc. have been analyzed. Thirdly, prototype is developed whose validation data is obtained from domain experts and is validated by ourselves as well as from domain experts.
Results: It is found that there is a big communication gap between deaf and unimpaired in Pakistan. This is mainly due to the lack of an automated system that can convert Audio speech to PSL and vice versa. After investigating state of the art work including solutions in other countries specific to their languages, it is found that no system exists that is generic and automated. We found that there is already work started for PSL to English Speech conversion but not the other way around. As part of this thesis, we discovered that a generic and automated system can be devised using speech recognition and Machine translation techniques.
Conclusion: Deaf people in Pakistan lack a lot of opportunities mainly due to communication gap between deaf and unimpaired. We establish that there should be a generic and automated system that can convert English speech to PSL and vice versa. As part of this, we worked for such a system that can convert English speech to PSL. Moreover, Speech recognition, Machine translation and Natural language processing techniques can be core ingredients for such a generic and automated system. Using user centric approach, the prototype of the system is validated iteratively from domain experts.

Keywords: Sign language, Pakistan sign language, Speech to sign language converter, speech and Pakistan sign language, Sign language interpreting machine, user centric, prototype validation

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF ACRONYMS

ABSA	Anjuman Behbood-e-Samat-e-Atfal (Deaf Institute Name)
AI	Artificial Intelligence
AUX	Auxiliary
API	Application Programming Interface
BSL	British Sign Language
DET	Determiner
GUI	Graphical User Interface
IL	Intermediate Language
MT	Machine Translation
NISE	National Institute of Special Education
NP	Noun Phrase
PAD	Pakistan Association of Deaf
PC	Personal Computer
PP	Prepositional Phrases
PSL	Pakistan Sign Language
PST	Part of Speech Tagging
PSLIM	Pakistan Sign Language Interpreting Machine
PTB	Penn Tree Bank
S	Sentence
SAPIs	Speech Application Programming Interface
SDA	Sir Syed Deaf Association (Deaf Institute Name)
SL	Sign Language
TESSA	The Text and Sign Support Assistant
TTS	Text-To-Speech
VP	Verb Phrase

1 INTRODUCTION

1.1 Research Domain

The research area we have chosen within Computer Science is an overlapping area between Natural language processing and Machine translation. The main goal is to facilitate the deaf people while interacting with unimpaired. In this regard, we have worked on speech translation into Pakistan Sign Language. Natural language processing can enhance the way of using computers while Machine translation can break the language barrier. Communication is a basic human need. People without hearing impairment have the ability to listen and speak but deaf people cannot hear and if they are by birth deaf then they cannot even speak. So deaf people use signs to communicate, these signs are known as sign language. Each country has its own sign language. In Pakistan, there is a sign language known as Pakistan Sign Language (PSL). It's a dynamic set of facial expressions and hand positions. Currently, human PSL interpreters are necessary to support and enhance communication with deaf people. PSL interpreters are not always available. Resultantly, communication between the deaf and other people may be impaired or nonexistent. So a translator or some interpreting mechanism is required for communication between the deaf and hearing.

Many countries have computerized interpreting mechanisms which are available to communicate with deaf persons but unfortunately no such mechanism is available in Pakistan. Hence a human translator is needed in order to communicate with deaf people. Some attempts have been made in Pakistan like GesTalk. This system requires a Data Glove and a static gesture made by the user. The system pronounces an alphabet [3].

Some attempts are made in other countries for sign languages and these systems take speech as an input and then convert it into appropriate signs. For example, in Great Britain the system known as TESSA (Text and Sign Support Assistant) is designed to translate speech into British sign language but this system is specific to post office [1].

1.2 Sign Language

Deaf people use signs to communicate. When a deaf person wants to say something, he/she can make some signs using his/her hands. Each particular sign means a distinct letter, word or expression. One can compare a sign with a word in our (hearing people) case e.g. the sign in the picture means what ("kiya" in Urdu and "vad" in Swedish).

Fig 1.1 Sign for "What" in Pakistani Sign Language

Fig 1.2 Sign for "What" in Swedish Sign Language [41]

Combination of signs will make a sentence just like words make sentences which are understood by other people. So a language of signs is there i.e. Sign Language (SL) which follows all rules of linguistics e.g. SL follows a grammar just as Urdu follows

Urdu Grammar or English follows English Grammar and has a vocabulary of distinct signs just as Urdu has a vocabulary of distinct words. One important thing is that every country has its own SL e.g. PSL in Pakistan, BSL in Great Britain, just like every country has its own language e.g. Urdu in Pakistan.

Everybody does not know Sign Language. Therefore an interpreting mechanism is required for communication between the hearing and the deaf, i.e. either an interpreter or a mechanism which is able to take speech as input and after analyzing it sentence by sentence, display appropriate signs using SL signs database, dynamically i.e. **Speech to Sign.**

1.3 History & Development of Pakistan Sign Language

The history and development of Pakistan Sign Language (PSL) can be traced in the work of the following four organizations [4]:

- Sir Syed Deaf Association (SDA), Rawalpindi.
- Anjuman Behbood-e-Samat-e-Atfal (ABSA), Karachi
- National Institute of Special Education (NISE), Islamabad
- Pakistan Association of Deaf (PAD), Karachi

1.3.1 Sir Syed Deaf Association (SDA), Rawalpindi

By the efforts of Syed Iftikhar Ahmad the result was the publication of PSL dictionary. This dictionary contained 750 signs related to different topics.

The main dilemma of this attempt was that it included the signs used in Rawalpindi region only. But the benefit of this effort was that it was an independent admirable process in the area. One of the biggest accomplishments of this project was to bring in single handed Urdu alphabetic finger spelling. This has been used by deaf population without any debate [6].

1.3.2 Anjuman Behbood-e-Samat-e-Atfal (ABSA), Karachi

ABSA (Anjuman-e-Behbood-e-Samat-e-Atfal) Research Group was established in 1986 with deaf and hearing adults as members. The aim of this group was to document and standardize the Pakistan Sign Language.

Work on this project started in November 1986 with the support of the Norwegian Church Aid Organization and the Norwegian Association of the Deaf [4]. During 1987 to 1998, ABSA Sign Language Research Group published many booklets. Some of them are mentioned below [4]:

- A Dictionary of Pakistan Sign Language, 1987
- Relationships in Sign Language, 1989
- Time and Seasons in Sign Language, 1989
- The Anatomy and Body Actions in Sign Language, 1990
- Numeration in Sign Language, 1992
- A Dictionary of Pakistan Sign Language (2nd Edition) 1995.
- Shareer Bander, Storybook, 1998

1.3.3 NISE Research Project

National Institute of Special Education (NISE) was established in 1976 and the main responsibility was development of manpower for special education centers being run under supervision of Voluntary Welfare Organizations, Federal Government and Provincial Governments, throughout the country. Soon after its establishment, it was realized that there was huge need for development of proper teaching methods in special education and especially for the education of hearing impaired children [4].

Special attention was given to training of teachers in oral/auditory method of teaching in the earlier stage of the special education development in Pakistan. Help was also taken from foreign consultants at that time.

NISE became conscious about the hearing impaired children in the institutes because they were using all types of signs for the means of communication. That's why the danger of adoption in expressing the meaning in different signs was sensed not only in different regions but also in the same center or school. It was realized that they needed to develop a basic core sign language for Pakistan having regional variations, if possible [4].

1.3.4 PAD Sign Language Research Group, Karachi

Pakistan Association of the Deaf (PAD), Karachi was established as an NGO through the efforts of ten dynamic deaf youths as a club for deaf men and women who were through with their Matriculation. It is affiliated with the World Federation of the Deaf (WFD) in Pakistan. PAD hosted a national seminar to discuss the following issues [4]:

a) Teachers of the deaf are not competent to pass on education in Sign Language. Therefore, children did not benefit from the education being taught by hearing teachers who had limited or no knowledge of the sign language.

b) Deaf children were taught through speech without signs so knowledge was limited and therefore failure rates were high in the schools.

c) In most of the cases, sign language was not given enough importance by the schools dealing with deaf children. Therefore, deaf relied more on rote learning.

PAD's sign language research group has been working over the past several years analyzing the existing sign language in the country. The research group realized that there was simply nothing to facilitate learning for deaf children at Nursery level in sign and visuals. So they gathered information from different teaching methodologies adopted at different schools for the deaf in the country. It was felt that there was a need to develop activity books that combine PSL signs and provided exercises alongside to foster learning and trial all in one. After thorough deliberation and consultation, PAD's sign language research group published following 4 books in sign language [4]:

- Workbook of Alphabet signs in Urdu
- Workbook of Alphabet signs in English
- 500-word dictionary with new words and modified signs
- Traffic signs for Deaf drivers

Extension of this work has now been focused on devising sign words for environment related vocabulary and Urdu grammar signs.

1.4 Problem Definition / Aims and Objectives

1.4.1 Purpose

The main purpose of this research is to explore the challenges currently facing the deaf while interacting with unimpaired people and then providing a solution that can facilitate this communication. One part of this goal is to explore the challenges and possible solutions for computer supported speech translation into Pakistan Sign Language.

There are solutions for PSL to speech conversion [3] (research has just started and no complete product is yet available) but there is no solution or even initial level of research going on at the moment that targets speech to PSL. Therefore deaf people have to rely on human interpreters who can convert speech of hearing people to PSL.

This research of English speech to PSL is the first initiative in this problem area. English is taught as secondary language all over Pakistan in all educational institutes and the medium of communication is English in almost all institutes. Thus the choice of developing a prototype that translates from English spoken language to PSL was more appropriate in current scenario to avoid having to address issues of local dialects of Urdu and minority languages in use in parts of Pakistan. We believe that this research work will act as an example for even other groups working on PSL to speech systems as we are targeting a much more automated and generic solution.

1.4.2 Aims and Objectives

This study explores the challenges faced by deaf people while interacting with unimpaired. In order to understand and realize these challenges, sign languages, Pakistan Sign Language and history and development of PSL is also explored. The main goal of this thesis was to develop a prototype of an automated and generic system that converts English speech to PSL that can be used to facilitate the communication among deaf and unimpaired. This work explores and analyzes existing systems targeting this problem area. This research utilizes state of art speech recognition, Machine Translation and Natural language processing techniques to reach towards a solution that is generic and automatic. To make the system, we have to have a logic engine which work as a cookbook/DB for identification of parts of speech that makes its more generic. Our cookbook itself has data for identifying the input. Maturity and generic factor depends upon the data been passed to the cookbook. The devised solution is developed as a prototype using a user centric approach. The main users we have focused on this case are instructors and teachers of deaf people.

1.4.3 Research Questions

The main research question of this thesis is:

"How can English audio speech be converted into PSL in a generic and an automatic way?"

We have divided our main research question into the following sub questions in order to reach the desired solution:

RQ 1: What are the challenges faced by deaf people when interacting with unimpaired, and how can technology support and enhance this interaction?

RQ 2: How can we adopt Machine translation techniques to convert speech into Pakistan sign language (PSL)?

1.4.4 Outcomes

After investigating the problem under consideration, examining possible solutions, analyzing their suitability for our goal and exploring in depth Speech recognition, Machine translation and Natural Language processing techniques, we proposed a system that is based on state of the art speech recognition, MT and NLPs techniques. Thus, we developed a prototype which takes English speech as an input and converts it into Pakistan Sign Language.

2 RESEARCH METHODOLOGY

A research methodology gives a strategy or an idea about how to attain research goals. The choice of appropriate implementation of a methodology is of importance to attain results and to maintain quality in the research [61]. The following section describes our research process.

2.1 Research Process

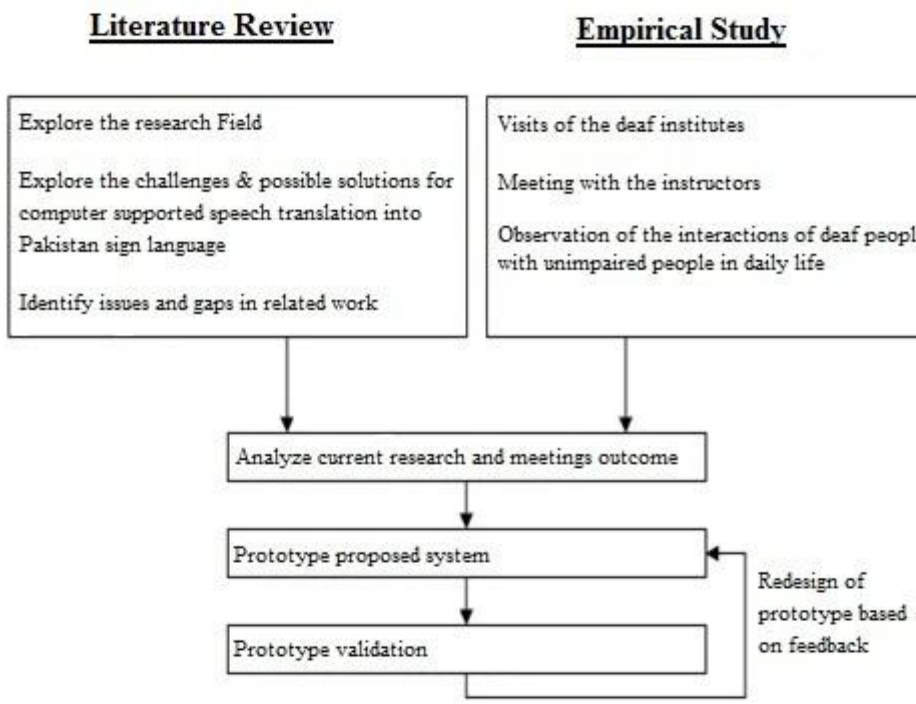


Fig 2.1 Framework of Methodology

Our research process consists of literature review, empirical study, analysis of the findings from literature and empirical study, prototyping the proposed solution, prototype validation and improvement iteratively. Note that literature review and empirical study went on almost side by side.

Going further into this process, our research has undergone a number of phases. We have done a literature review related to our particular domain. As part of this literature review, we read existing work already done related to our problem domain, analyzed if this can be used directly or indirectly towards the solution of our problem, extracted relevant problems, issues and results from existing work done. Specific as well as general study was carried out to find out gaps in our research area. Then literature about state of the art techniques like Machine Translation and Natural language processing were read and analyzed in detail to use towards the solution of our problem. Moreover, the empirical study was carried out at institutes for deaf persons in Pakistan. We visited institutes to collect information about current use of Pakistan sign language. Interviews and meetings with instructors have been carried out in order to gain better knowledge about Pakistan sign language, how it is taught and used today in Pakistan. Interviews with these domain experts and observing directly deaf people in everyday life situations helped us better understand the challenges faced by deaf people. We also obtained a set of PSL signs from these instructors that we used in our prototype.

As an outcome of literature review, collection of information from deaf institutes, interviews with instructors at deaf institutes and observing deaf people in everyday life situations, we devised a solution that converts English speech to PSL, developed a prototype system, validated it ourselves and with the help of instructors of deaf institutes for validation purpose. We also have discussed limitations of the prototype and possible future work.

A high level description of our research methodology is given in figure 2.1. The results of literature review are given in detail in Chapter 3. In this section, we shall briefly describe our empirical study. Detailed information about the interviews is given in chapter 4. We have described the prototype of our solution in Chapter 5, including detailed prototype validation.

To get relevant information about the specific domain, the approach of using interviews is the most important method [14]. We did this as described in the following section.

2.1.1 Interviews and observations

We have used interviews with the domain experts to get understanding of how the deaf people are taught in the institutes and how deaf people interact with unimpaired people. From the interviews we learned the following things:

- Deaf people in Pakistan use PSL (Pakistan Sign Language) for communication.
- English is taught as secondary language all over Pakistan in all educational institutes, thus the choice of developing a prototype that translates from English spoken language to PSL was not a controversial one but rather a natural way of avoiding having to address issues of local dialects of Urdu and minority languages in use in parts of Pakistan.
- Deaf people are taught by instructors who not only need to know the domain expertise of the area that they are teaching like Math, History, Science etc. but they also need to know PSL very well in order to teach the deaf. As such, it becomes very difficult for deaf institutes to get instructors who know both.
- Whenever deaf people need to communicate with unimpaired people in any situation, they either need to hire a translator or request the unimpaired people to write everything for them. Translators are very difficult to get all the time and they are very expensive as well. Moreover, using writing by unimpaired becomes very slow process and not all unimpaired people want to do this. We observed this phenomena ourselves as instructors of the institutes gave us ten deaf persons with whom we worked to understand their feelings and challenges in everyday life. In this way, we experienced going with deaf people in shopping malls, banks, post offices etc. and with their permission, we observed their interaction. We have concluded that sometimes their interaction with unimpaired people becomes very slow and embarrassing.

Based on above findings, we concluded that there is definitely a need for an automated system that can facilitate communication between deaf and unimpaired people.

From our interviews and observations, we came to know that ideally there should be a two way communication system that can convert Urdu or English audio to PSL and vice versa. Moreover, this must be generic and not domain specific for banks, post offices and hospitals but it should be generic enough and should not depended on any manual intervention etc. so that it can be used anywhere.

The literature review, interviews and observations were not sequential but after doing some initial literature reviewing, these processes were run in parallel.

From the literature review, we got the following information to incorporate:

- The system will be generic; it includes all vocabulary of technical and non-technical jargons to assist the instructors.
- There is no such system that provides two way communications that can convert Urdu or English audio to PSL and vice versa.
- There is no such system that provides even one way communication that can convert Urdu or English audio to PSL.
- There are some systems for machine translation between speech and sign language that exist in the world but have problems like domain specificity and need for manual human intervention etc.
- Details of these systems are given in sections 3.3 and 3.4. From their design, technologies used, their goals etc. we came to the conclusion that we cannot use these systems or even extend them to achieve our goal.

2.1.2 Prototype Solution

For the solution to our problem, we went through state of the art techniques like Machine learning, natural language processing and speech recognition. After learning about them in detail (our understanding is given in section 3.3, 3.4), we came to conclusion that these are techniques that can be utilized to achieve our goal. In this way, we applied innovative idea of using natural language processing techniques and speech recognition.

After concluding this, we explored all the state of the art technologies related to speech recognition, Machine learning and natural language processing. From there we could see some options and then after more investigation, we decided to use the best possible options (details are given in section 3.5).

We had previous knowledge of database systems and their use, programming languages etc. For implementation of our desired prototype, we did investigation for possible programming languages, frameworks and tools etc. to decide the best for our case; details are given in section 5.5. Deeper knowledge of GUI was required in this case although we had previous experience of GUI development, for this sophisticated prototype, we put efforts into gaining expertise in GUI development considering Human Computer Interaction.

We were also in continued contact with deaf institutes instructors post interviews. With their feedback, we iteratively improved our prototype. For the prototype development itself, we used agile methodology, we used pair programming, we re-factored the code in the form of iterations using sprint methodology and we did extensive unit testing, functional testing and end to end testing. Each iteration also known as a sprint, we were defining user stories, dividing each user story into tasks and completing them in the same sprint. At the end of each sprint, we took the final working prototype version to the instructors, demonstrated and got feedback.

Instructors provided feedback regarding the prototype and finally they accepted it. Detail of prototype validation is given in section 5.4.

Based on the amount of effort it needed and the thesis time, we scoped our thesis so that we provided one way communication from audio to PSL. Moreover, we chose English as the language for audio input.

Below mentioned section gives an overview of the period of time during which we carried out literature review, empirical studies, prototype design and its development. We also highlighted the points during the process when instructors were consulted, and the process of field studies with deaf people were carried out.

2.1.3 Thesis Plan

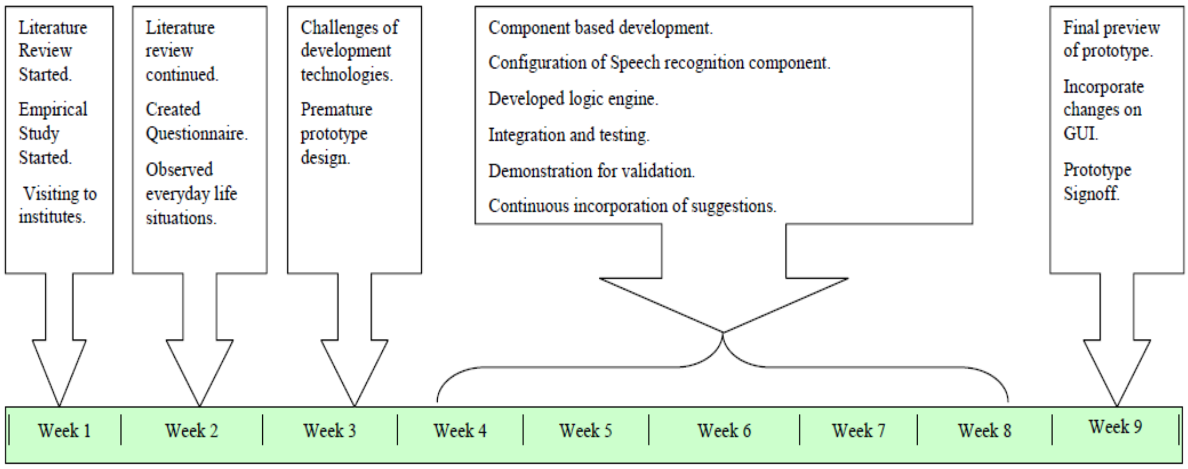


Fig 2.2 Thesis Timeline showing the prototype development process which lasted 9 weeks during the entire 20 week thesis project

Week 1

- We started with the literature review; we went through existing work which was already performed in our problem domain, scrutinized if this can be helpful directly or indirectly leading to the solution of our problem, extracted relevant problems, issues and results from existing work. Specific as well as general study was carried out to find out gaps in our research area.
- To make empirical study possible we went to the deaf institutes for the appointment and gave them the introduction of what we are up to and explained the scope of our research domain.

Week 2

- The literature review continued; we were going through the state of the art techniques like Machine Translation and Natural language processing.
- We then devised a generic questionnaire for the concerned personals and after completion of initial literature review, we decided to conduct interviews. In this week we carried out all the interviews and details of these interviews are mentioned in section 4.7.1.
- We were also present in the lecture rooms and we keenly observed the communication between the instructors and students. To analyze the communication with unimpaired people we accompanied students in everyday life situations.
- During our interview and observation conceived during daily life situations we jotted down the main points which were helpful for developing the solution and validation.

Week 3

- With the help of literature review and the empirical study we were now able to address the actual challenges we were going to face in the development of the prototype. So we then started off with the initial requirement; based upon, what we conceived so far for the development of prototype.
- To implement our prototype, we had various options of development technologies. We compared existing tools for the development of prototype keeping in mind pros and cons of each and also considering hands on experience on those tools.
- A premature flow chart of devised prototype was designed.

Week 4, 5, 6, 7, 8

- Component based development of the prototype was started.
- We initially noshed and trained Speech recognition component.
- Based upon output of Speech recognition component we analyzed, designed and developed logic engine.
- Incorporated input data set videos and pictures for data base development which were taken from instructors.
- Integration of Speech recognition component with logic engine.
- Testing phase was stared as below for two components:
    - Did extensive unit testing and functional testing of logic engine.
    - After integration of Speech Recognition component with logic engine, end to end testing was performed.
    - Validated the logic engine outputs.
- Integrated Video rendering component in our GUI.
- Overall testing and validation phase was started as below:
    - After integration of graphic output component with logic engine, end to end testing was performed.
    - Validated the final outputs by functional testing. This was done in two phases (logs enabled and automated tests).
- We demonstrated the working of the prototype to instructors and got feedback in each week.
- Based upon the feedback from instructors we incorporated suggested changes.
- Due to 70% automated testing, it was convenient for us to re-factor and improve the prototype based on the feedback from the instructors.
- Aforementioned points are discussed in details in chapter 5.

Week 9

- Asked the instructors to comment on the user experience, based upon GUI we developed.
- Instructors gave us the suggestions which were based upon look and feel of the prototype, and we incorporated those changes in our prototype.
- After incorporation of above changes we demonstrated our prototype to the instructors for validation and they finally approved it.

3 LITERATURE REVIEW

3.1 Purpose and Method

We have done literature review in order to better understand PSL, our problem area, existing work done directly or indirectly related to the problem, analyzing the suitability of existing solutions to the problem, analyzing their possible extensions or using some of their ideas to reach the solution. As part of this study, we explored state of art Machine translation and Natural language processing techniques, speech to text conversion techniques, studied and analyzed existing language text converter systems that used MT and NLP techniques, explored existing frameworks and technologies and analyzed them to find out the most suitable for devising the solution and then development of a prototype.

We have searched the information from books, digital libraries and search engines; such as ACM, IEEE, Springer, Inspec, Google Scholar etc.

Following section describes the concept of Machine translation and explains how it can be utilized for automatic speech translation. MT is given in detail as this is a critical concept in our problem domain.

3.1.1 Machine Translation (MT) Introduction

Different people have described MT in different ways. All of them refer to same concept in one way or another. For example, MT is a subfield of natural language processing, which explores the utilization of computer software to translate speech or text from one language to another [48]. Another researcher has defined MT as conventional and standard name for computerized systems that are accountable for production of translations, with or without human assistance. Previously, machine translation was also called ‘automatic translation’ and ‘mechanical translation’. These terms are hardly used now in English [46]. Therefore, in this thesis report, we have only used the terminology of “Machine Translation”.

MT is not a new concept, and it has been one of the oldest dreams of humanity. These dreams have become a reality in the twentieth century, when computer programs have the capability of translating extensive range of texts from one natural language to another. From the earliest days of MT, it has been claiming high potential. But reality is not reflective of these claims. So there is no translating machine that can take text in one language and generate a perfect translation to any other language without human involvement or support [46].

Most of the time, MT carries out substitution of words from one language to another and do more complex translations for better handling of different linguistics, idiom translation and phrase recognition etc [48].

3.1.2 Why Does Machine Translation Matter?

MT is a significant topic and its importance is possibly increased socially, commercially, politically and philosophically as the 20th century ends and 21st century begins [20].

The sociopolitical meaning of translation forms social or political importance of MT in those societies where more than one language is generally spoken. If single language is adopted, which is a practicable alternative rather than to widespread the use of translation, this alternative is not so attractive because the dominance factor is involved for the speakers of other languages and those languages become second class and finally disappear [20]. Such type of loss matters to everyone as the loss of language frequently leads to the disappearance of an individual culture [21].

As far as ordinary human interaction is concerned, translation is necessary for communication (one of the core ingredients to fully assemble a society). If you are allowed to express yourself in your own language in a different society, it seems to be important that one should be dependent on the availability of the translation [20]. In this regard, demand for translation in the modern world seems to be more important. Scarcity of human translators is a strong motivating factor to spend time for automatic translation (automation - as much as possible) [20]. Considering that modern societies can't force a common language on their members, we can conclude that machine translation has become a social and political necessity for modern societies [20].

3.1.3 The Recent Growth of Machine Translation

The main MT users are the multinational companies, as they require technical documentation and operational manuals in a number of languages. These machine translating systems are dependent on mainframe or client-server setup and raw output is generated. To get better quality of the translation, post editing of raw output is done. Another good solution is to do pre-editing of the input text i.e. it is mostly controlled with a standard language. This technique minimizes the costly editing processes. So both types of MT are in use and millions of page translation is done using these techniques every year [21].

These days, commercial as well as home users of PC products use independent professional translators. There are many sellers of mainframe systems like Systran, Fujitsu and Metal [46] as they uphold most of the famous features of their mainframe products. So some products are aimed for the professional translators and some are purely developed for personal home users [21]. In this regard many vendors introduced products by making three way product divisions: one is categorized for corporations (enterprise), second one is for independent translators or for translation agencies (professional) and third one is for home users who want infrequent translations (personal) [20].

Nowadays, many vendors provide network based translation services. The pioneer of online translation of electronic mail was CompuServe and then Systran made it possible to do online text translation and for web page translation in Babelfish service of AltaVista. Many other vendors also have followed Systran for translation such as Google, FreeTranslation, Reverso, ProMT, etc. The demand flow of internet information leads this area to stand alone PC software products that are now able to translate electronic mail messages and Web pages [20].

A few years back, Microsoft initiated a service for automatic translation known as "Windows Live Translator". This service permits to translate limited text up to 500 words or a web page from English to Dutch, Spanish, Italian, Korean, Chinese, Russian and German. Microsoft also used Systran to produce the translations and also give an option to translate computer related texts using a machine translation system. Moreover, the translation technology of Microsoft has been used to translate technical materials including MSDN Library [51].

After providing information about MT in general and recent developments on the market for MT solutions, the next section describes it in a broader context of speech translation.

3.1.4 Speech Translation

Translation is the understanding of the meaning of the text and successful production of that text into another equivalent text that shows the same message in another language. The text that is input for translation is called source text and the one that is translated is called target language or target text [49].

The objective of automatic speech translation was always a vision for MT and recent development in speech technology makes that practical. In the beginning, research was on small scale but when the industry realized its importance and need, major work was started [20].

The speech translation process includes three steps: the first one is to convert the speech into appropriate text; second step is to convert source language text to target language text and third is conversion of target language text to corresponding speech.

Second step is the most difficult part as the process involves in-depth knowledge of syntax, grammar, semantics and idioms etc. In order to understand how this step can be solved by using MT techniques, we have given detailed level information regarding machine translation in this regard.

3.1.4.1 Representing Linguistic Knowledge

The syntax is divided into two different types of analysis of sentences. The first type is categorized in constituent or phrase structure analysis. Division of these sentences is categorized into parts as verbal, nominal and so on. The second type is to categorize grammatical relations such as Subject, Object, Head and other parts of speech in the sentence [47].

Constituent Structure

Composition of sentence is made up of words and categorized into parts of speech together with Verbs, Nouns, Adverbs, Adjective, Prepositions, Auxiliary and Determiners (abbreviated as V, N, ADV, A, P, AUX, DET) [47]. For the formation of grammatical or well-formed sentences, some sets of rules are used and these rules are known as grammar of language. The following example shows difference of grammatical and non-grammatical sentence [47]:

(1) a. Put some paper in the printer.
b. Printer some put the in paper.

Rule '1a' indicates that the sentence is grammatical but '1b' is not.

English grammar has some rules by which sentences are formed or created. In this regard a sentence consists of noun phrase such as the 'the user' pursued by a representation or an auxiliary verb 'should', this is also followed by a verb phrase such as 'clean the printer' as mentioned below [47]:

(2) The user should clean the printer.

To elaborate it further, some explanation of noun phrase is required. So a noun phrase can consist of a determiner or an article such as 'the', 'or', 'a' and 'a noun'; such as printer that is mentioned in the below mentioned example ‘3a’. In ‘3b’ the determiner can be omitted according to the context of sentence [47].

(3) a. the printer
  b. printers

The abbreviations of specific terms are mentioned, to deeply understand and visualize meaning of the labeled bracketing of the strings of words in the following example.

Sentence is abbreviated as ‘S’, noun phrase as ‘NP’, verb phrase ‘VP’, auxiliary ‘AUX’ and determiner ‘BET’ [47]:

(4) a. User should clean the printer.
  b. $[_S\ [_{NP}\ [_N \text{ users }]][_{AUX} \text{ should }][_{VP}\ [_V \text{ clean }][_{NP}\ [_{DET} \text{ the }][_N \text{ printer }]]]]$

The following example shows that auxiliary verb is optional and the verb phrase can consist of just a verb (such as stopped) [47]:

(5) a. The printer should stop.
  b. The printer stopped.

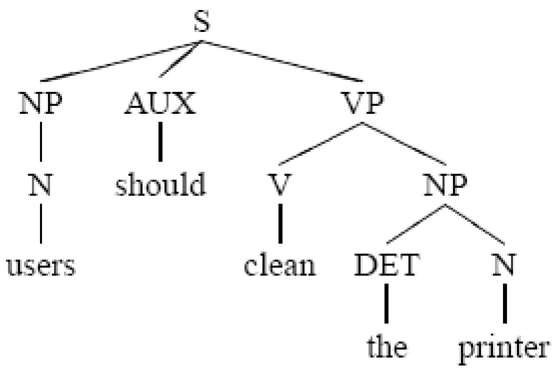


Figure 2.1 Grammars and Constituent Structure [47]

The following example shows that NP and VP can contain prepositional phrases (PPs), these prepositions consist of (on, in, with, etc.) and NPs [47]:

(6) a. The printer stops on occasions.
  b. Put the cover on the printer.
  c. Clean the printer with a cloth.

All the examples mentioned above are parts of the sentence that cannot be used to form independent sentences [47].

Grammatical Relations

In the above part we tried to give information about how sentences are made and categorized into parts of speech but in this section other information representation like sentential compliment, subject and object are discussed.

In different languages representation of the subjects and objects varies and this can be further explained with the help of an example. In English, usually ‘Subjects’ are noun phrase that come before the Verb and ‘Objects’ usually occur directly after the verb. In Japanese the flow of the world is Subject, Object, Verb and in Irish it is like Verb, Subject, Object. From examples we judged that the element structure of the language is very much different. That can be explained in terms of similarity of grammatical relations [47].

Meaning

The information represented above is useful in two ways for MT. Firstly, it's easy to use the natural language processing when simple and enhanced explanation of what is involved in translation is included. Secondly, it is used to understand the source text in context with meaning of the sentences. This improves the translation of MT system but at present this seems to be an impossible mission and to some extent it is [47]. Three kinds of knowledge are involved in understanding meaning of the language.

1. **Semantic knowledge**.
   It's the knowledge about what language (individual words and sentences) means, independent of the situation they appear in.
2. **Pragmatic knowledge**.
   It's the knowledge of the expression meaning in terms of conditions and specifically on the occasions of use.
3. **Real world or common sense knowledge**.
   It's the knowledge of context of the language or sentence.

Speech translation is certainly the most creative field of computer based translation and it's an attraction for funding and publicity. Now a days there is a widespread practical use of MT in many multinational companies, but the researchers are very well familiar that there is much to be done to improve quality [21].

3.2 Challenges Involved in MT

Human Translators actually have expert competence in at least five different kinds of knowledge [47], as follows:

- Knowledge of the source language.
- Knowledge of the target language. This permits them to create texts that are acceptable in the target language.
- Knowledge of various communications between source language and target language. This is the knowledge of how individual words can be translated.
- Knowledge of the subject matter, including ordinary general knowledge and 'common sense'. The knowledge of the source language allows them to understand what the text to be translated means.
- Knowledge of the social conventions, culture, customs and expectations of speakers of the source and target languages.

As part of our literature review, three MT systems for language text conversions are analyzed that are given in the following section.

3.3 Working of Three MT Systems

Machine Translation is a system that maps one language into another language by electronic means. In reality, natural language requires special handling when it is formulated by the grammar of input language to generate target language in consistence with the grammar of target language.

In this study, working of the following three MT systems are analyzed in detail to get better understanding of how commercial translators would work and help out to analyze the challenges regarding these techniques. These are not for sign language but their study helped us to understand the use of MT techniques for language text conversions:

- Techland internet Translator
- KIT-FAST
- KANT

3.3.1 Techland internet Translator

Techland internet Translator system has the typical architecture of the Machine translation system based on transfer with the following consequent stages of processing [42]:

1. Text segmentation based on grammar (Finite States Automata):

In this very first stage of internet translator, text is segmented into tokens and these tokens consist of words, numbers, punctuation marks etc. Within this stage, finite state automation is used to compile the tokens that are described by a regular expression. Each token's information (type, font size and color) is stored to preserve the layout of the text [42].

2. Morphological analysis:

At this stage all the known words are stored by using a monolingual dictionary (Mdic) and unknown words are simply treated as nouns [42].

3. Part of Speech Tagging (PST):

In the early stages of the internet translation, the use of PST was not included in the system but that was manually included as problems with quality of parser occur [42]. The purpose of PST is to mark the words in text (based on its definition and context) to appropriate part of speech [52].

4. Transfer: implemented in the form of set of rules:

The aim of transfer is to transform a parsed tree into a tree of rough target language structure. This functionality initiates from root of the tree and continued to depth of the tree in a depth-first algorithm style. During this process some usual parser errors can also be corrected [42].

5. Target syntactic structure synthesis:

This is last part of the internet translator. The tree produced by transfer is modified so that the target language syntax is to be accurate [42].

3.3.2 KIT-FAST

The Kit-Fast project was initiated for translation of German text into English sentences using Machine Translation system. Basic component of this machine translator consists of syntactic analyzer, which constructs structure of the input text. The analyzed structure consists of chains of phrases and of which the smallest is basic symbols of that language and the biggest chains are the sentences. Over all working is based on the construction of trees as symbols are represented by leaf nodes and sentences are represented by root of the trees [43].

The Machine translation system also employs two textual representations. First one represents structural information of a text and the second one represents the text content [43].

3.3.3 KANT System

KANT (Knowledge-based, Accurate Natural-language Translation) is the first system that combines semi-automated achievements of knowledge using set of software tools and deep analysis on source and generating text [53]. Knowledge based machine translation produces fast and high quality translation to several languages along with improvement in the source text clarity and MT analysis phase [44].

KANT system has separately designed architecture as the text generator and text parser are independent components. Resultantly, any source language supported by the system can be translated into any target language supported by the system. This separately designed architecture technique allows the knowledge sources of different languages to be combined more efficiently in new applications in order to support various source and target language combinations [44].

3.4 Sign Language Conversion Systems

In the field of information and communication technology, research and development for deaf people has been carried out in many ways and one concept is speech to text conversion and from that text to sign language. In this context, research has been done in the field of telecommunications to do communication bridging between deaf and hearing users in the region of South Africa [54].To achieve this goal, advancement has been made using state of the art technologies of multimedia i.e. text, voice and video [54].

Services provided by this system are via email, SMS (Short Message Service), video telephony, chat and fax. The architecture of this service system is dependent on two things. First, if two users have same communication device or service (most of the time these devices are non-audio), speech to sign language service is not utilized.

Actual service is provided when one of them has different device or service (i.e. when both speech and video is involved). That is, the most important factor involved is if one of them uses different language (one is hearing and one is deaf), state of the art technology is used for communication between the two and this technology was named as ‘Relay Telephony’ [54]. This relay telephony converts speech to text and video and vice versa by internet. These are semi-automated as most of the services were taking up by human operator to convert a speaking user’s speech to signs/text or deaf user’s signs/text to speech [54].

Text based examples of relay for internet is Typetalk by British telecom, (www.typetalk.org), National Relay Service (NRS) and TalkingText (www.talingtext.net). Video based project examples of relay for internet is Sorenson (www.sorensonvrs.com), UmptiDumpti and SignTel (http://signtelinc.com). These all are commercial services [54].

The drawback of this solution is that this technology is not fully operational as in many under developed countries; these relay services are not yet provided due to technological and financial reasons. One of the big reasons for not choosing this architecture is that it is semi-automated, which means human assistance is required to convert speech to text or sign or vice versa.

Another system in this context is TESSA, which converts English speech into British Sign Language (BSL). It is an experimental system designed to assist communication (completion of operation) between deaf and Post office clerk's speech to sign language by displaying signs in the screen by specially developed avatar (virtual human/graphical model of the human).

The developers of this project claim that they have used a set of pre-defined phrases as this system is meant to be used in post office and much of the communication is predictable. The sub-system used for speech recognition was 'Entropic speech recognizer' [1]. Working of this sub-system is dependent on the speaker's voice adaptation, and this process takes approximately an hour to create individual's model. These models are stored on the user basis and can be used for later use. The system architecture consists of a language model recognizer (called probabilistic language model recognizer) pursued by a language processor. Overall working of this language processor is to map the output of the recognizer to correct phrase [1].

In summary, working of TESSA is coordination of above components with each other. The key working functionality makes possible communication between speech recognition sub-system, avatar module and overall controlling of transaction process. The system used TCL (Tool Command Language) for sign assembly and C++ for avatar module. The communication between the components was performed using remote procedure call system (TCP/IP socket connections) [1].

The limitation of this system is that it operates in a restricted domain of conversation. Moreover, a major factor of noise occurs when language is spoken by the clerk. It is also Speaker dependent.

Considering all the information given in this section, we concluded that none of above mentioned systems fully matches our target goal, they can't be extended nor modified; rather we need to go for a different approach. In order to reach to the solution, we used the ideas of MT language text conversion systems.

4 INTERVIEW PROCESS - COMPREHENSION OF REAL WORLD

After completion of initial literature review, we decided to conduct interviews; as interviewer can track in depth information around a specific topic. In this way, we have carried literature review, interviews and observations of deaf people side by side. In our domain of interest, face-to-face meeting was vital to collect information within different institutes to explore the challenges and opportunities of Speech Translation into Pakistan Sign Language.

In this chapter, methods are explained that were used for collection of data. Several interviews were conducted with four different special children instructors from four different special children institutes. These instructors have been working in special children institutes for many years. The reason to choose these institutes was that they are well established and highly reputed in their field. Before performing the interviews, planning was done to select those suitable personnel or institutes which were experienced and had expertise in their field of Pakistan sign language in that specific region. So we presented our premeditated potential theme to number of instructors to collect more detailed information about needs and requirements for MT solutions for translation between English spoken language and PSL.

4.1 Conducting Interview

Several interviews were performed for gathering appropriate information of the current challenges related to the teaching and understanding of sign language. Main theme of the interviews was assessment of converting speech into Pakistan sign language. The existing techniques and technologies have not been fully adopted by the organizations for support of deaf people.

The special person institutes and some organizations are involved directly or indirectly towards development of systems, which are handy for the communication of both deaf and other persons. As there is no big budget for development of such applications and there are lots of difficulties in taking up these kinds of projects, our purpose to conduct the interviews was to explore the challenges and possible solutions for computer supported speech translation into Pakistan sign language.

4.2 Planning Approach

Appropriate preparation is important to accomplish the requirements of designed prospective to perform successful interviews. As a result, we have given emphasis on the important factors like time schedule, proper planning, management and guidance support.

Prior to accomplishing the interviews, we did a literature review and had casual meetings with the deaf person’s instructors. Adding together this whole preparation helped us design the questions which were used in the interviews with instructors.

4.3 Detailed Questions

We have planned several questions to carry out interviews to target the interviewee staff of special person's institutions. These questions were intended to achieve good quality feedback and recommendations to our research area from different personnel of institute.

4.4 Questionnaire Design

Appropriate interview questions were planned before carrying out interview from special person's instructors of institute. Concerning articles, literature review, researcher's ideas and views, several questions were planned to be put to the selected professionals during the interviews. Majority of questions were related to our research field as different suggestions from their work and communication; both with deaf and with each other, around issues concerning teaching sign language were closely related to research area.

Designed questions were simple and according to our research area. Questions were asked directly (face-to-face) in their own work environment and the details of those questions are mentioned in **Appendix A**.

4.5 Purpose of Interview

Interview is a method or process that is used to gather information from the subject of research approach [18]. Semi-structured or open ended interviews were conducted with different special person's instructors for collection of information. Several instructors were interviewed to get information about practices, techniques and methods they used to teach sign language. Mainly it was about the gathering of data on their work and communication.

The aim of intended interview questions was to attain the feedback and recommendation from the instructors to get deeper understanding of what the gaps and challenges are facing them and other concerned personnel while communicating with deaf people.

4.6 Interview Instrument

In our recommended study, we conducted interviews with different personnel of four different institutes. These selected personnel were specially trained for the education of deaf persons. Different instruments and tools were used to record these interviews. Face to face interviews were conducted and the interviewees granted permission to record their interviews.

Technologies used to record interviews were laptops and video cameras. Mainly the interviews and discussion were recorded through video cameras.

4.7 Interview Realization

Wide ranging questions formulate foundation of any interview to attain the purpose. A number of instructors of special person's institute were interviewed depending on their teaching specialization. Before carrying out the interview, we talked to the heads of the institutes and informed them about our research work so that they could provide us access to relevant personnel. In total, four different instructors from four different institutions were interviewed.

Interviews were carried out during a period of around one month. These interviews were conducted separately. Detailed discussion about our interview questionnaire was done with all interviewees and they provided feedback to the best of their understanding and abilities but the problem was to get the appointment from their busy schedule. Details of the interviewees are mentioned below.

4.7.1 Details of Interviewees

The interviewees gave us their explicit permission to use their real names in our research report.

Interviewee 1

We carried out first interview with Dr. Ahmed Hassan and had some interaction with special children at Hassan Academy for Deaf and Mute. He is principal/head of academy at Rawalpindi and has more than 15 years of experience and expertise in his field. He is well known for his excellence in deaf education. His main areas of expertise are Inclusive Education, Speech Therapy, Behavior Therapy, Scientific Approach, Independent Living Skills and Formal & Informal Education.

During the interview, our main focus was on the kind of problems that will be faced by the students if interpreters were replaced by automatic system like PSLIM. Different other factors related to sign language were also discussed which were related to our research domain.

Interviewee 2

We carried out the second interview with Zahoor Hussain at Sir Syed Deaf Association. He is responsible for teaching of computer science and he is the key player for introducing the first digital sign language book named "Pakistan first computer Sign language". During our conversation he told us that much larger inventory of vocabulary is needed to run such a higher level, wide-ranging computer literacy program for hearing impaired. But on what they have so far developed, it was decided to publish Volume-1 of the book. He also informed us that his team is still working on the task for developing advanced vocabulary.

The main theme of this interview was to get potential reflections from theoretical work. We also discussed the set of laws, rules, policy and guidelines in defining the new sign vocabulary especially in case of computer sign language.

Interviewee 3

The third interview was carried out with Dr. Jamal Nasir. He is also principle/head at Government school for Deaf and Mute, Swaan Camp, Rawalpindi. This interview was mainly based on the interview questionnaire basically on the interview questioners and general discussion about field domain.

Interviewee 4

The fourth interview was carried out with Mrs. Tayyaba. She is program coordinator for special education of mute and deaf children at National Institute of Special Education, Islamabad. She provided knowledge in development of the sign gesture data base for our prototype. As she had vast experience in the field of teaching, she gave demo of various sign gestures which we have recorded by video camera. These sign gestures are used in the database of our prototype.

5 PSLIM ARCHITECTURE AND DESIGN

As part of our solution to this thesis problem, we also have proposed and developed our own prototype. It consists of three major components - Speech Recognition, Logic Engine and Graphics. For Speech Recognition System we have used a freeware tool Microsoft Speech SDK v5.0 and the rest of two components are fully contributed by us as part of this thesis. System is used as standalone desktop application. This system is devised as components based so that one can replace any component to better versions expected to be available in the future. Moreover, one can update the PSL vocabulary as it evolves in the future. The overall architecture is shown in figure 5.1. In next section, we have described each component, its input, output, responsibilities and functioning.

5.1 User input and Speech Recognition

This component is used to take user input in the form of Speech or direct Text. This also provides software configuration settings interface. If speech is used as an input then it is converted in to Text using speech recognition system. This is further divided into subsystems as show in figure 5.1 whose high level description is given below.

5.1.1 Speech

This component is used to take speech input. This component uses Speech API-SAPI.

Software Configuration Settings

Our prototype system can be configured in various ways. Such kind of administration is provided through Software Configuration settings component. Following are the list of configuration options. We have provided very brief overview of each option here. More details are provided in **Appendix D**.

Exit

It is used to exit from the application.

5.1.2 Speech properties

This button can be used to change speech properties like:

1. **Language:** We used Microsoft English ASR Version 5 Engine for our prototype.
2. **Recognition Profile:** It stores information about how to recognize your voice. Changes can be made in a profile to recognize a different voice or a different noise environment.
3. **Setting:** This option is used for the recognition profile settings. Changes can be done to pronunciation sensitivity, accuracy vs. response time and background adaptation.
   - **Pronunciation Sensitivity:** This setting is used to control the margin of error in your pronunciation, whether the system should reject your command if it is not certain what you said (high sensitivity) or always act upon your command even if it is unsure about what you said (low sensitivity).
   - **Accuracy vs. Response Time:** There are two options to choose. One is higher accuracy with slower response time and second one is lower accuracy with faster response time. These options are available because speech recognition uses high

amount of PC's resources, especially when speech to text conversion is in process.

- **Train Profile:** This component is used to improve speech recognition accuracy as it trains the system according to the pronunciation style of the user.

4. **Engine Training:** A training wizard is used to improve recognition quality. There are lists of training sessions that can be used to train the system for better performance. List of instructions are shown and one has to follow those to complete the session.
5. **Microphone Configuration:** Microphone configuration wizard can be used to test and tune microphone for use with speech recognition. The wizard provides guidance for proper microphone placement, volume adjustment etc.

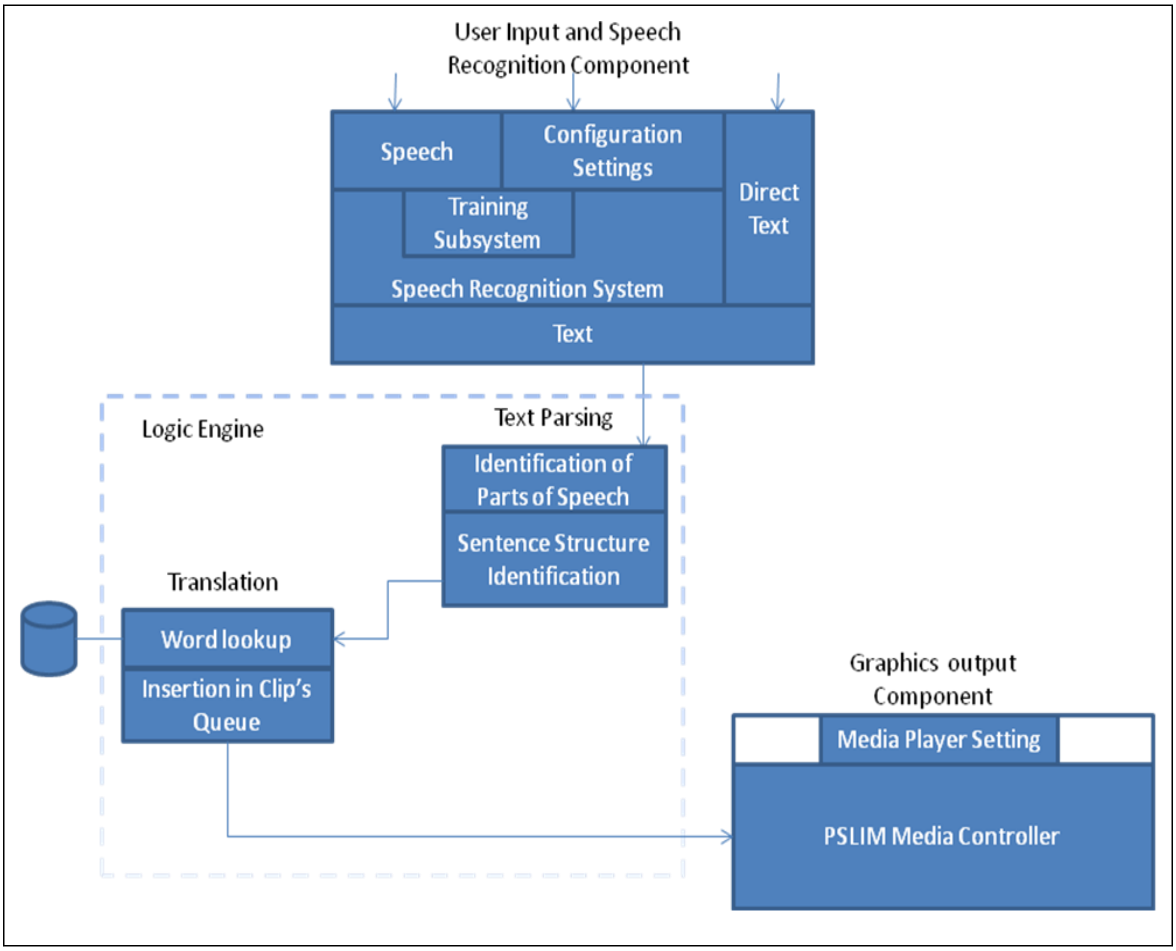


Figure 5.1: System Architecture Diagram

6. **Path Configuration:** By using these configuration properties, one can configure paths of various add-ins of the application such as path of the Clip Database, Word Net database etc.
7. **Display Switch mode:** It is provided to switch between large window and small window of the output panel.
8. **Help:** Provides various help menus of the system.
9. **Speech input:** This option has two characteristics:

- **Enable Speech:** This option is unchecked by default. In this case, system will not accept speech input. But if this optioned is enabled, it starts taking speech input from microphone.
- **Recognition Report:** Two controls are provided for speech related reports;
    i. It notifies the user that the system has started or ended accepting audio input.
    ii. It notifies the user whether speech has been recognized or not.

10. **Text input:** If the speech option is disabled, the system automatically accepts text input from its specified panel.
11. **Current Recognition:** Spoken speech is shown as a text in this panel as it is recognized by the speech recognition system.
12. **Text input list:** Given that the speech is disabled, this panel shows the list of input text strings given to the system.
13. **Translation Input:** In the case when speech is enabled and input is coming from the speech, this item will show list of text strings which are processed by the translation module. Note that this shows list of text strings as compared to the Current Recognition that displays only the current text.
14. **Current Translation:** This contains the most recent translated sentence.
15. **Clips View:** This Panel displays the video clips of the signs.

5.1.3 Speech Recognition System and Training Subsystem

This component is used to convert speech to Text. As part of this, it uses Speech API-SAPI. Speech recognition uses set of audio/sound models while processing input speech signals and maps those signals for recognition of spoken speech. Prototype contains default models for this purpose. Further to this, it contains a speech adaptation/speech training subsystem to train the speech recognizer to the voice of individual user hence customizing the system for better accuracy. This process takes about twenty minutes and the individual user models are stored in the system for later use. Speech training greatly increases accuracy of the system.

When speech training is completed, the system can be used for normal speech input that it converts to text and notifies the user about the activities/progress going on and put that text in the queue for further processing.

Text

An interface is also provided to the user to directly input the Text. In this case, Speech Recognition Component is not used. Text coming directly from user or coming via Speech recognition system is then given to the Logic Engine.

5.2 Logic Engine

Logic Engine is core part of the Prototype. It comprises of the Text Parsing and Translation Component. Input for this component is Text and output is the ordered list of clips.

5.2.1 Text Parsing

This component is divided into two parts. It includes “Identification of Parts of Speech” and “Sentence Structure Identification”. These are explained as follows:

Identification of Parts of Speech

The text is tokenized i.e. the syntax is divided into two different types for analysis of sentences. The first type is categorized in constituent or phrase structure analysis. In this case, division of these sentences is categorized into parts as verbal, nominal and so on. The second type is to categorize grammatical relations such as Subject, Object, Head and other various parts of speech of the sentence. After this categorization, each token is then identified as a part of speech according to English grammar rules. This identification is performed with the help of Sharp NLP. With the help of this identification process we have developed our own database for the identification of the rules with the desired output of our system.

Sentence Structure Identification

After identification of the parts of speech, the nature of the sentence is identified i.e. whether it is declarative, imperative, interrogative or some kind of conditional sentence.

5.2.2 Translation

In a very abstract level, this component maps the words and expressions to corresponding video sign clips. This is divided into following components:

Word lookup

It does look up in the database to find equivalent sign (video clip path) for the word under consideration. In case it is not available, it does look up for the word meanings and then tries to find the sign mapping against meaning of the word e.g. gorgeous into beautiful. Tags associated with words helps choose the best meaning for that word in particular context. It also looks for subject specific terms such as RAM, CPU etc.

Moreover, the database contains gender categorization and source to target grammar mappings as well. While doing this word lookup and getting the final signs, these gender categorization and grammar mappings are well considered and the Meta information is present in the database.

Insertion in the Clip's Queue

Paths of sign clip files are then inserted in the queue for further processing.

5.3 Graphics Output Component

The lucid way to represent text in the form of sign language by the application would be to accumulate the video recordings of the sign language interpreter. But our major focus was to generate a proper sign language phrases in response to output from the speech recognizer.

This component gets required sign clips from the queue and renders the video clips on the specified place on the GUI. Sentence boundaries are also implemented to identify the end of sentence through a special item on the queue and notify audience about it. This component is internally divided into components like Media Player Settings, clips queue reader, video clips rendering and playing, identifying end of clips and playing next clip.

Media player settings

This component is provided to configure various graphical output settings. For example, video format identification, video-display resolution and other control necessary for playing a video file.

PSLIM Media Controller

This takes clips as input using Clips queue reader, renders the clip on the specified place on the GUI and identifies end of clip.

5.4 Prototype Validation

We have done functional and non-functional validation of our prototype. For functional validation, we have verified accuracy of English speech to PSL. For non-functional requirements, we have verified response time of the system and resource utilization including CPU and memory.

5.4.1 Functional validation

We chose a set of input and their desired output. This is done in two phases. In the first phase, we have run the system with logs enabled. An ordered set of English speech input was given to the system and system output for each input was logged into a file in terms of the PSL sign clip output. These files were then verified through automated scripts (containing statically the given input and the desired output) to check how many outputs were correct. Once we had achieved considerable accuracy from the prototype, we showed this behavior to the instructors we had interviewed in the beginning. Those instructors then used the prototype and validated the English to PSL accuracy and ease of use. During design and development of the system, we kept on using log output and automated scripts way to validate the system. Then at certain points of time, we showed the system to domain experts and got it validated.

For automated tests, we ran a set of size 200. These 200 inputs and corresponding desired outputs were taken from domain experts and this data was our bench mark to validate if our prototype was working fine or not. For the validation to expert domains, we used a set of size 50. More specifically, this data set contains one, two or three possible outputs corresponding to one input. First being fully accurate, second is less and third the least. If the output obtained is none of these then output is considered incorrect. In this way, incorrect result leads to 0, matching with 1st gives value of 1, 2nd to 0.80 and 3rd to 0.70.

In this way, initially, the accuracy result was 61%.We analyzed incorrect results and found that one reason for inaccuracy is speech recognition component, specifically due to lack of enough training of individual speech. After training, we got 78 % accuracy.

Remaining 5% inaccuracy is due to speech recognition engine quality and 17 % is due to Machine translation limitations, e.g. same words in different contexts have different meanings [22].

As we have mentioned that we interviewed the instructors of the deaf institutes from different areas of Pakistan, we also demonstrated our prototype system iteratively to all of these instructors and incorporated their feedback over the period of time. Hence our system can be used across Pakistan.

5.4.2 Non-functional validation

We verified response time of the system and resource utilization including CPU and memory. We used windows XP with Intel(R) Core(TM) i5 CPU 2.40 GHz, 2.0 GB of RAM. None of other applications were running on the system except our prototype.

Following are the average values taken on 200 readings (1 reading consist of input and corresponding output):

Average Response time: 571 ms
Average CPU utilization: 8.49%
Average Memory utilization: 363MB

5.5 Technologies for our solution

This section provides an overview of our chosen technologies for the solution and other available options.

In order to implement our prototype, we had various options of development technologies. Besides our main goals of the master thesis, we have also considered cost effectiveness, time limitations of our master thesis and its scope in order to choose following technologies:

- **.NET with C#:** is used for the development of our project as .NET framework is a Microsoft freeware product including Microsoft Speech Application Programming Interface (API).
- **Sharp NLP:** is used in our project for parsing the text and finding the meaning of it. It is needed after Speech recognizer API converts human spoken audio into readable text. As part of its working, Sharp NLP uses word net.
- **WordNet:** the purpose for choosing WordNet is to have a combination of dictionary and thesaurus which is more intuitively usable in order to support automatic text analysis as done by Sharp NLP.
- **Active Media Library**: is used in our project as it gives facilities to play video files in an application. It can run all those formats which the Windows Media Player on the host machine can give. In short, it plays the video of PSL.
- **Microsoft Office Access:** is used in our project as relational database management system. The analyzed text as provided by Sharp NLP is stored in this database.

Now for each of above chosen technologies, we had other options that are given as follow:

Chosen Technology	**Other options available**
.NET with C#	Java
Sharp NLP	**Java:** Stanford's JavaNLP, OpenNLP and Gate **Python:** Natural Language Toolkit (NLTK)
WordNet	None
Microsoft Active Media Library	Java Media Framework API (JMF)
Microsoft Access	Oracle

5.6 Limitations

Below are the limitations of the prototype. We have categorized the limitations as following.

PSL Limitation:

- Due to limited vocabulary of PSL itself, our prototype at the moment works under its limitation.

Speech recognition and Machine translation techniques Limitation:

- The system is automatic and generic but it brings inherent limitation of speech recognizer and Machine Translation, i.e. inaccuracy. This system has an accuracy of 61% when used without training speech recognition engine. This means for the best accurate results, this system needs to be deployed at places like deaf institutes, banks, post offices and shops etc. In case of deaf institutes, speech recognition engine training will be done for instructors, customer representatives in case of banks and post offices and sales persons in case of shops. Moreover, deaf person can train the system for all friends who are unimpaired. In everyday life, like going on a road or in a park, this will give less accurate results. But as discussed in the future work, we expect that this application can be provided as iphone application that means even many unimpaired people can have this application with them all the time, get it train with their voice and then deaf people can interact with unimpaired even in ad hoc situation like going on a road or in a park etc. If the system is trained then its accuracy reaches up to 78%. Remaining 5% inaccuracy is due to speech recognition engine quality and 17 % is due to Machine translation limitations, e.g. same words in different contexts have different meanings [22].

Due to less than 95% of accuracy [5], this system cannot be used as mission critical application like hospital emergency department, fire emergency instruction and a military order.

6 DISCUSSION

The main objective of this study was to first explore the difficulties currently facing the deaf while interacting with unimpaired people and then providing a solution that can facilitate this communication.

With above main objective, we explored the current solutions targeting this area of problem domain, analyzed if they suit our main goal, or could be extended. Moreover, considering our main problem domain, we analyzed their drawbacks so that we could devise a solution that fully solves our main problem. On the way to achieve our goal, we explored the challenges and possible solutions for computer based speech translation into Pakistan sign language, devised working prototype of a system whose input is the audio English and output is the Pakistan sign language.

In order to fulfill our objective, we divided it into sub-objectives formulated as research questions. Following are the research questions of our study that we have attempted to answer in this thesis report.

> **RQ 1:** What are the challenges faced by deaf people when interacting with unimpaired, and how can technology support and enhance this interaction?
>
> **RQ 2:** How can we adopt Machine translation techniques to convert speech into Pakistan sign language (PSL)?

The literature review helped us gaining an overview of the research topic and identifying the gap between current research and practice. To get the relevant information about the specific domain, interviews were conducted. Visits of institutes in Pakistan were performed to collect data and information about current use of PSL. Interviews and meetings with instructors were conducted in order to gain better knowledge about PSL, how is it taught and used today in Pakistan and to validate the proposed prototype. Moreover, observations were done with deaf while they interact with unimpaired in everyday life situations to help better understand their problems faced by them.

Answer to research question 1: Deaf people need to be taught by instructors that not only need to know the domain expertise of the area that they are teaching (like Math, History, Science etc.) but they also need to know PSL very well in order to teach to deaf. Due to this, it becomes very difficult for deaf institutes to get instructors having knowledge of both. Hence, deaf people are sometimes taught by instructors that have no knowledge of PSL. Consequently, deaf people cannot get same quality of education as unimpaired people.

Furthermore, whenever deaf people need to communicate with unimpaired people in any situation, they either need to hire a translator or request the unimpaired people to write everything for them. Translators are very difficult to get all the times and they are very expensive as well. Moreover, using writing by unimpaired becomes very slow process and all unimpaired people do not want to do this. In this regard, deaf people find it difficult to do what they want to do at shopping malls, banks and post offices etc. Moreover, deaf people in Pakistan mostly end up having friends that are also deaf.

Through our field studies and interviews we learned that Pakistan sign language is important for communication with deaf people in Pakistan, but that very few unimpaired people know this language, whereas spoken English is a common language shared by most unimpaired people in Pakistan. It became clear that machine translation between Pakistan sign language and spoken English could be extremely helpful in supporting and enhancing interaction between deaf people and unimpaired people in Pakistan. A solution supporting two-way communication was not feasible within the framework of our master thesis project. Thus we decided to focus on developing a prototype for speech translation from spoken English to Pakistan Sign Language. The process of iteratively developing a prototype of such a one-way solution can also be understood as a way of further developing our understanding of the problem area.

Answer to research question 2: After study and analysis of current solutions targeting this area of problem domain, we concluded that none of the existing solutions fully attains our goal. These systems are either domain specific or human intervention is required, i.e. not automated. We also analyzed them so that if we can extend them towards the solution of our main problem area but we concluded that they cannot be extended. Details are given in section 3.4. So we chose a totally different solution for our goal that better solves our problem. Our solution involves speech recognition, machine translation and natural language processing techniques supplemented by GUI assistance. Software Application Programming Interface (SAPI) provides a high-level interface between an application and speech engine. It implements all the low-level details needed to manage and control the real-time operation of various speech engines. There are two basic types of SAPI engines i.e. text-to-speech (TTS) system and speech recognizers (SR) system. TTS system synthesizes text strings and files into spoken audio using synthetic voices and SR converts human spoken audio into readable text. For more accurate conversion of speech it involves training profile as well. Given a source as text, Machine translation techniques can be utilized in order to intelligently parse the text. Output of the parsed text can then be used to obtain corresponding signs of a sign language. In this way, Machine translation involves sentence analysis including constituent or phrase structure analysis and categorization of grammatical relations. This also includes gender categorization and grammar mappings. So in this way Machine translation techniques are adopted by our solution in order to convert speech to sign language.

After getting answers to RQ 1 and RQ 2, we also have answered RQ 2 by devising a prototype solution in which speech is first converted into a text. Text is then parsed and translated into PSL. Detailed design and architecture of this prototype system is given in chapter 5 and its validation is given in section 5.4.

6.1 Validity Threats

Creswell [13] identifies various validity threats for empirical methods and data collection procedures. We discuss internal and external validity threats here with respect to our work.

Internal validity threat is defined as the threat of the credibility of the selection of data used for validation and credibility of the participants.

External Validity Threat is defined as a threat to the validity of generalizing a finding to a bigger class.

In our case, we have tried to overcome the internal threat in the following ways.

We have chosen four domain experts of PSL. These are instructors of deaf institutes from different areas in Pakistan. So the information provided about the use of PSL, the challenges as faced by institutes and deaf people in everyday life represent all regions of Pakistan and the data given by these domain experts to validate our prototype as given in prototype validation section is credible. Due to lack of PSL instructors in Pakistan and their tight schedule, we could get hold of only four experts. It would be better to take interviews and prototype validation and feedback from more than four domain experts.

In order to overcome external threat, we have chosen big data set for prototype validation. This constitutes of 200 data set members. Moreover, this data set is carefully chosen to belong to different aspects of daily life usage so that we can generalize the prototype validation. From the prototype ease of use point of view, we got feedback from four instructors and deaf students. These deaf students age range lies in 13 to 35. So we have an external threat in the prototype that the ease of use may not work well for the deaf people who are not in this age group.

7 CONCLUSIONS AND FUTURE WORK

7.1 CONCLUSIONS

This research has investigated a computer based solution to facilitate communication among deaf people and unimpaired. Investigation was performed using literature review and visits to institutes to gain a deeper knowledge about sign language and specifically how is it used in Pakistan context. Secondly, challenges faced by deaf people to interact with unimpaired are analyzed by interviews with domain experts (instructors of deaf institutes) and by directly observing deaf in everyday life situations.

We conclude that deaf people rely on sign language for communication with unimpaired people. Deaf people in Pakistan use PSL for communication, English is taught as secondary language all over Pakistan in all educational institutes, deaf people are taught by instructors that not only need to know the domain expertise of the area that they are teaching like Math, History and Science etc. but they also need to know PSL very well in order to teach the deaf. It becomes very difficult for deaf institutes to get instructors that know both.

Whenever deaf people need to communicate with unimpaired people in any situation, they either need to hire a translator or request the unimpaired people to write everything for them. Translators are very difficult to get all the time and they are very expensive as well. Moreover, using writing by unimpaired becomes very slow process and not all unimpaired people want to do this. We observed this phenomena ourselves as instructors of the institutes provided us the opportunity to work with deaf people to understand their feelings and challenges in everyday life. In this way, we used to go with deaf people in shopping malls, banks, post offices etc. and with their permission, we observed their interaction. We have concluded that sometimes their interaction with normal people becomes very slow and embarrassing.

Based on above findings, we concluded that there is definitely a need for an automated system that can facilitate communication between deaf and unimpaired people. These factors lead to the subsequent objective of this research.

The main objective of this thesis is to identify a generic and an automated system without any human intervention that converts English speech into PSL as a solution to bridge the communication gap between deaf and unimpaired.

It is identified that existing work done related to this problem area doesn't fulfill our objective. Current solutions are either very specific to a domain, e.g. post office or need human intervention i.e. not automatic. It is identified that none of the existing systems can be extended towards our desired solution.

We explored state of the art techniques like Machine translation, Speech recognition and NLP. We have utilized these in our proposed solution. Prototype of the proposed solution is developed whose functional and non-functional validation is performed. Since none of existing work exactly matches to our problem statement, therefore, we have not compared the validation of our prototype to any existing system. We have validated prototype with respect to our problem domain. Moreover, this is validated iteratively from the domain experts, i.e. experts of PSL and the English to PSL human translators. We found this user centric approach very useful to help better understand the problem at the ground level, keeping our work user focused and then realization of user satisfaction level throughout the process.

This work has opened a new world of opportunities where deaf can communicate with others who do not have PSL knowledge. Having this system, if it is further developed from a prototype to a functioning system; deaf institutes will have wider scope of choosing best instructors for a given domain that may not have PSL expertise. Deaf people will have more opportunities to interact with other members of the society at every level as communication is the basic pillar for this.

The automatic speech to sign language is an attractive prospect; the impending applications are exhilarating and worthwhile. In the field of Human Computer Interface (HCI) we hope that our thesis will be an important addition to the ongoing research.

7.2 FUTURE WORK

With this thesis, we have devised a solution and developed its prototype that can be used to convert English Speech to PSL and is automatic and generic. At each step, special consideration is given for future improvement as explained below;

a) PSLIM is component based and each component is loosely coupled.
b) 90% code coverage is done with automated tests.
c) Each component has got well defined interfaces for input and output while interacting with other components.
d) Speech recognition component can be replaced with better quality one in the future.
e) Logic engine can be improved by using future improved Machine translation techniques. There is a lot of research going on in the area of Machine translation as the techniques will improved especially for higher accuracy in the future, logic engine can be updated.
f) PSLIM can be updated to be deployed on iphones and in this way this can be used in everyday life like walking in a park etc. Then this can be used by the unimpaired people as well who can train this system on their iphone with their voice and can help deaf people interact with them conveniently and with high accuracy.
g) This system can be used by Human Computer Interaction and Machine translation research groups to use as a platform where they can demonstrate their optimizations in the core algorithms in more end user point of view considering English speech to PSL conversion as one use case.
h) As the PSL will evolve, this can be updated in PSLIM.
i) Government of Pakistan can further deploy this system at important places like shopping malls and hospital information desks etc.
j) The combination of methods, using field studies and user-centric iterative prototyping with future users of the system, provided many new insights into the needs of users in this area; however we could only apply it to a limited extent and would like to explore further how user-centric design and development can inform better design of services.

REFERENCES

[1] S. Cox, M. Lincoln, J. Tryggvason, M. Nakisa, M. Wells, M. Tutt, and S. Abbott, "TESSA, a system to aid communication with deaf people," *Proceedings of the fifth international ACM conference on Assistive technologies*, 2002, p. 212.

[2] R. Munoz and A. Montoyo, "Advances on natural language processing," Data & Knowledge Engineering, vol. 61, 2007, pp. 403–405.

[3] Y. Khambaty, R. Quintana, M. Shadaram, S. Nehal, M.A. Virk, W. Ahmed, and G. Ahmedani, "Cost effective portable system for sign language gesture recognition," IEEE International Conference on System of Systems Engineering, 2008. SoSE'08, 2008, pp. 1–6.

[4] Government of Pakistan, National Institute of Special Education, Directorate General of Special Education, Ministry of Special Education and Social Welfare (1991). Pakistan Sign Language Book I and II.

[5] Yngve, V. H. (1957). A framework for syntactic translation. Mechanical Translation, 4(3):59–65. Available on http://www.mt-archive.info/MT-1957-Yngve.pdf

[6] Sir Syed Deaf Association, Rawalpindi (1989). Pakistan Sign Language.

[7] Microsoft, .NET Framework Conceptual Overview. 2010, [cited 2010-01-30]; Available from: http://msdn.microsoft.com/en-us/library/zw4w595w(v=VS.100).aspx.

[8] Microsoft, Speech API Overview (SAPI 5.4). 2010, [cited 2010-01-29]; Available from: http://msdn.microsoft.com/en-us/library/ee125077(v=VS.85).aspx.

[9] Microsoft, .NET Framework Class Library. 2010, [cited 2010-01-29]; Available from: http://msdn.microsoft.com/enus/library/system.speech.recognition.audiolevelupdatedeventargs.aspx.

[10] Codeplex, SharpNLP - open source natural language processing tools. 2010, [cited 2010-02-03]; Available from: http://sharpnlp.codeplex.com/.

[11] Wikipedia, Woednet Knowledge Structure. [cited 2010-01-26]. Available from: http://en.wikipedia.org/wiki/WordNet#Knowledge_structure.

[12] Codeplex, WordNet::SQLConverter. 2010, [cited 2010-02-07]; Available from: http://wnsqlconverter.codeplex.com/.

[13] J.W. Creswell, Research design: Qualitative, quantitative, and mixed methods approaches, Sage Pubns, 2008.

[14] D. Davy and C. Valecillos, "Summary of a literature review of qualitative research in technical communication from 2003 to 2007," 2009.

[15] S.S. Fels, "Glove-TalkII: Mapping Hand Gestures to Speech Using Neural Networks An Approach to Building Adaptive Interfaces," Citeseer, 1994.

[16] S.I. Amari, "Learning patterns and pattern sequences by self-organizing nets of threshold elements," IEEE Transactions on Computers, vol. 100, 1972, pp. 1197–1206.

[17] Wikipedia, ALPAC. [cited 2010-04-29]. Available from: http://en.wikipedia.org/wiki/ALPAC.

[18] S.E. Hove and B. Anda, "Experiences from conducting semi-structured interviews in empirical software engineering research," 11th IEEE International Symposium Software Metrics, 2005, 2005, p. 10.

[19] B. Kitchenham and S. Charters, "Guidelines for performing systematic literature reviews in software engineering," Version, vol. 2, 2007, pp. 2007–01.

[20] W.J. Hutchins and H.L. Somers, An introduction to machine translation, Academic Press New York, 1992.

[21] C. Goutte, N. Cancedda, and M. Dymetman, Learning machine translation, The MIT Press, 2009.

[22] M. W Madsen, "The Limits of Machine Translation" [cited 2010-04-29]. Available from: https://docs.google.com/fileview?id=0B7-4xydn3MXJZjFkZTllZjItN2Q5Ny00YmUxLWEzODItNTYyMjhlNTY5NWIz&hl=en

[23] R. Yang, S. Sarkar, and B. Loeding, "Enhanced level building algorithm for the movement epenthesis problem in sign language recognition," IEEE Conference on Computer Vision and Pattern Recognition, 2007. CVPR'07, 2007, pp. 1–8.

[24] L.G. Zhang, X. Chen, C. Wang, Y. Chen, and W. Gao, "Recognition of sign language subwords based on boosted hidden Markov models," Proceedings of the 7th international conference on Multimodal interfaces, 2005, p. 287.

[25] E. Efthimiou, G. Sapountzaki, K. Karpouzis, and S.E. Fotinea, "Developing an e-Learning platform for the Greek Sign Language," Computers Helping People with Special Needs, pp. 626–626.

[26] Z. Krňoul, J. Kanis, M. Železnỳ, and L. M\üller, "Czech text-to-sign speech synthesizer," Machine Learning for Multimodal Interaction, 2008, pp. 180–191.

[27] M. Lincoln and S. Cox, "A comparison of language processing techniques for a constrained speech translation system," 2003 IEEE International Conference on

Acoustics, Speech, and Signal Processing, 2003. Proceedings.(ICASSP'03), 2003.

[28] D. Burmeister, “Requirements of deaf user of information visualization an interdisciplinary approach,” Proceedings Seventh International Conference on Information Visualization IV 2003, pp. 433–438.

[29] Y. Aoki, S. Tanahasi, A. Burger, H. Hattori, and H. Wakoh, “Development of a CG-animation system for sign languagecommunication between different languages,” Information, Communications and Signal Processing, 1997. ICICS., Proceedings of 1997 International Conference on, 1997.

[30] M. Shirali-Shahreza and S. Shirali-Shahreza, “Encouraging Persons with Hearing Problem to Learn Sign Language by Internet Websites,” Eighth IEEE International Conference on Advanced Learning Technologies, 2008. ICALT'08, 2008, pp. 1036–1037.

[31] R. Yang, S. Sarkar, and B. Loeding, “Enhanced level building algorithm for the movement epenthesis problem in sign language recognition,” IEEE Conference on Computer Vision and Pattern Recognition, 2007. CVPR'07, 2007, pp. 1–8.

[32] M.P. Paulraj, S. Yaacob, H. Desa, C.R. Hema, W.M. Ridzuan, and W.A. Majid, “Extraction of head and hand gesture features for recognition of sign language,” Electronic Design, 2008. ICED 2008. International Conference on, 2008, pp. 1–6.

[33] P. Petrantonakis, V. Kosmidou, M. Nikolaraizi, S. Koutsogiorgou, and L.J. Hadjileontiadis, "“SEE and SEE”: An Educational Tool for Kids with Hard of Hearing,” Eighth IEEE International Conference on Advanced Learning Technologies, 2008. ICALT'08, 2008, pp. 1032–1033.

[34] A. Braffort, “A gesture recognition architecture for sign language,” Proceedings of the second annual ACM conference on Assistive technologies, 1996, p. 109.

[35] S. Min, S. Oh, G. Kim, T. Yoon, C. Lim, Y. Lee, and K. Jung, “Simple Glove-Based Korean Finger Spelling Recognition System,” Computational Science and Its Applications–ICCSA 2007, pp. 1063–1073.

[36] L. Van Zijl, “South African sign language machine translation project,” Proceedings of the 8th international ACM SIGACCESS conference on Computers and accessibility, 2006, p. 234.

[37] S. Furui, “Automatic speech recognition and its application to information extraction,” Proceedings of the 37th annual meeting of the Association for Computational Linguistics on Computational Linguistics, 1999, pp. 11–20.

[38] D. Hakkani-T\ür, G. Riccardi, and G. Tur, “An active approach to spoken language processing,” ACM Transactions on Speech and Language Processing (TSLP), vol. 3, 2006, pp. 1–31.

[39] R.H. Liang and M. Ouhyoung, “A real-time continuous gesture recognition system for sign language,” Third IEEE International Conference on Automatic Face and Gesture Recognition, 1998. Proceedings, 1998, pp. 558–567.

[40] C.H. Wu, Y.H. Chiu, and C.S. Guo, “Text generation from Taiwanese sign language using a PST-based language model for augmentative communication,” IEEE Transactions on Neural Systems and Rehabilitation Engineering, vol. 12, 2004, pp. 441–454.

[41] Database Home, Digital Version of Swedish Sign Language Dictionaries, [cited 2010-05-06]; Available from: http://130.237.171.78/fmi/iwp/cgi?-db=Digital%20version%20av%20Svenskt%20teckenspråkslexikon&-loadframes.

[42] M. Kłopotek, S.T. Wierzchoń, and K. Trojanowski, Intelligent information processing and Web mining, Springer, 2003.

[43] W. Weisweber and C. Hauenschild, A Model of Multi-Level Transfer for Machine Translation and Its Partial Realization, Technische Universitaet Berlin, 1990.

[44] E. Nyberg and T. Mitamura, “The KANT system: Fast, accurate, high-quality translation in practical domains,” Proceedings of COLING, 1992, p. 15thAPPENDIX A.

[45] A. Trujillo, Translation engines: techniques for machine translation, Springer Verlag, 1999.

[46] W.J. Hutchins and H.L. Somers, An introduction to machine translation, Academic Press New York, 1992.

[47] D. Arnold, L. Balkan, R.L. Humphreys, S. Meijer, and L. Sadler, Machine translation, NCC Blackwell, 1994.

[48] Wikipedia, Machine Translation. [cited 2010-01-28]. Available from: http://en.wikipedia.org/wiki/Machine_translation.

[49] Wikipedia, Translation. [cited 2010-01-29]. Available from: http://en.wikipedia.org/wiki/Translation_process.

[50] Wikipedia, Cognition. [cited 2010-01-30]. Available from: http://en.wikipedia.org/wiki/Cognitive.

[51] Google Operating System, Microsoft Launches Translation Service. 2007, [cited 2010-02-12]; Available from: http://googlesystem.blogspot.com/2007/09/microsoft-launches-translation-service.html.

[52] Wikipedia, Parts-of-speech tagging. [cited 2010-01-29]. Available from: http://en.wikipedia.org/wiki/Part-of-speech_tagging.

[53] E. Nyberg, T. Mitamura, and J. Carbonell, “The KANT Machine Translation System: From R&D to Initial Deployment,” *LISA Workshop on Integrating Advanced Translation Technology, Washington DC*, 1997.

[54] M. Glaser and W.D. Tucker, “Telecommunications bridging between Deaf and hearing users in South Africa,” Proc. CVHI 2004.

[55] Wikipedia, .NET Framework. [cited 2010-01-25]. Available from: http://en.wikipedia.org/wiki/.NET_Framework.

[56] Freshmeat, Processing Corpora with Python and the Natural Language Toolkit. 2010, [cited 2010-05-28]; Available from: http://freshmeat.net/articles/processing-corpora-with-python-and-the-natural-language-toolkit.

[57] AI Computer Vision, SharpNLP vs NLTK C# Review. 2010, [cited 2010-05-28]; Available from: http://samibadawi.blogspot.com/2010/03/sharpnlp-vs-nltk-called-from-c-review.html.

[58] Wikipedia, WordNet. [cited 2010-05-29]. Available from: http://en.wikipedia.org/wiki/WordNet.

[59] VisDic, General Information. 2010, [cited 2010-05-29]; Available from: http://nlp.fi.muni.cz/projekty/visdic/.

[60] Microsoft, .NET Framework Class Library. 2010, [cited 2010-02-14]; Available from: http://msdn.microsoft.com/en-us/library/ms554855 (v=VS.100).aspx.

[61] C. Wohlin, Experimentation in software engineering: an introduction. Boston: Kluwer, 2000.

APPENDIX A

INTERVIEW QUESTIONS

- What difficulties are you facing while interacting with Deaf person?
- Are you feeling that PSL is lacking something?
- Indicate the content areas in which you are responsible for providing instruction? (E.g. English, Urdu, Mathematics, Social studies, Science. Etc
- Currently what are your modes of communication?
- What are your instructional strategies?
- What challenges you face while tutoring the students?
- Describe the types of communication you use with students? (Like oral communication, or other signing systems?)
  What type and amount of support you receive from institutes who thoroughly worked on PSL?
- Is there any computer based system currently used by you to teach deaf students?
- How much familiar are you with Computer and Software use?
- What difficulties can be faced by instructors when using Computers and Software?
- What you recommend for the improvement of PSLIM?
- What flaws you think are in this system?
- Is it user friendly?
- How much time it took you to get familiarizes with PSLIM?
- Are you satisfied with 78% accuracy of PSLIM?
- What is your opinion regarding our future work plans?

APPENDIX B

PROTOTYPE VALIADATION INPUT DATA

We have given a sub set of sentences that we have used for the prototype validation. This is in context that all of these are spoken by an unimpaired person that will be converted to PSL so that deaf person can understand. We have divided the set of system input into categories in terms of situations that a deaf person will be interacting in. e.g. While moving around, a deaf person being asked about the current time or the answer by the hearing person to the question asked by the deaf person about the current time.

Time query:

What time is it?

1. Excuse me. Can you tell me the time, please?
2. Yes, of course. It's seven o'clock.

1. Thank you.
2. No problem.

What time is it?

1. What time is it?
2. It's half past three.

Shopping for a Jacket

1. Can I help you?
2. Yes, I'm looking for a jacket.

1. What size are you?
2. I'm an extra large.

1. How about this one?
2. Yes, that's nice. Can I try it on?

1. Certainly, there are changing rooms over there.
2. Thank you.

1. How does it fit?
2. It's too large. Do you have a large size?

1. Yes, here you are.
2. Thank you. I'll have it, please.

1. OK, how would you like to pay?
2. Do you accept credit card?

1. Yes, we do. Visa, Master Card and American Express.
2. OK, here's my Visa.

1. Thank you. Have a nice day!
2. Thank you, goodbye.

Checking In

1. Good morning. Can I have your ticket, please?
2. Here you are.

1. Thank you. Would you like smoking or non-smoking?
2. Non-smoking, please.

1. Would you like a window or an aisle seat?
2. An aisle seat, please.

1. Do you have any baggage?
2. Yes, I have a suitcase and a carry-on bag.

1. Here's your boarding pass. Have a nice flight.
2. Thank you.

Passport Control

1. Good morning. Can I see your passport?
2. Yes, here it is!

1. Thank you very much. Are you a tourist or on a business trip?
2. I'm a tourist.

1. That's fine. Have a pleasant time.
2. Thank you.

Ordering a Meal

1. Hi. How are you doing this afternoon?
2. Fine, thank you. Can I see the menu, please?

1. Certainly, here you are.
2. Thank you. What's today's special?

1. Grilled tuna with rice.
2. That sounds good. I'll have that.

1. Would you like something to drink?
2. Yes, I'd like a coke.

1. Here you are. Enjoy your meal!
2. Thank you.

1. Can I get you anything else?
2. No thanks. Can I have the bill?

1. That'll be Rs. 675.
2. Here you are. Keep the change!

1. Thank you! Have a good day!
2. Bye.

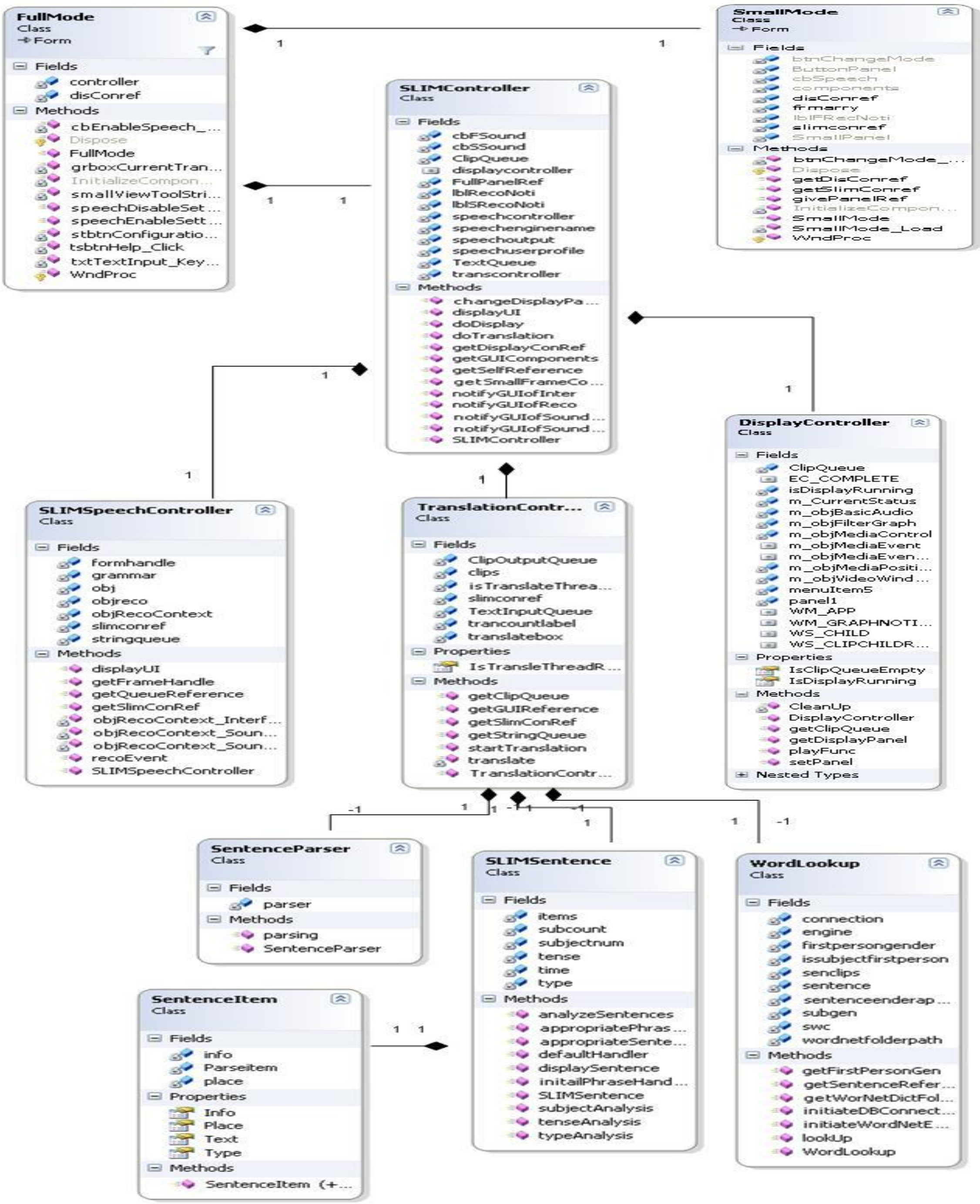
APPENDIX C

DESIGN & TESTING OF PSLIM

This section provides an overview of the entire prototype design and testing.

Class Diagram:

Class diagram helps to model the static view of an application. Mentioned below is the class diagram of PSLIM:



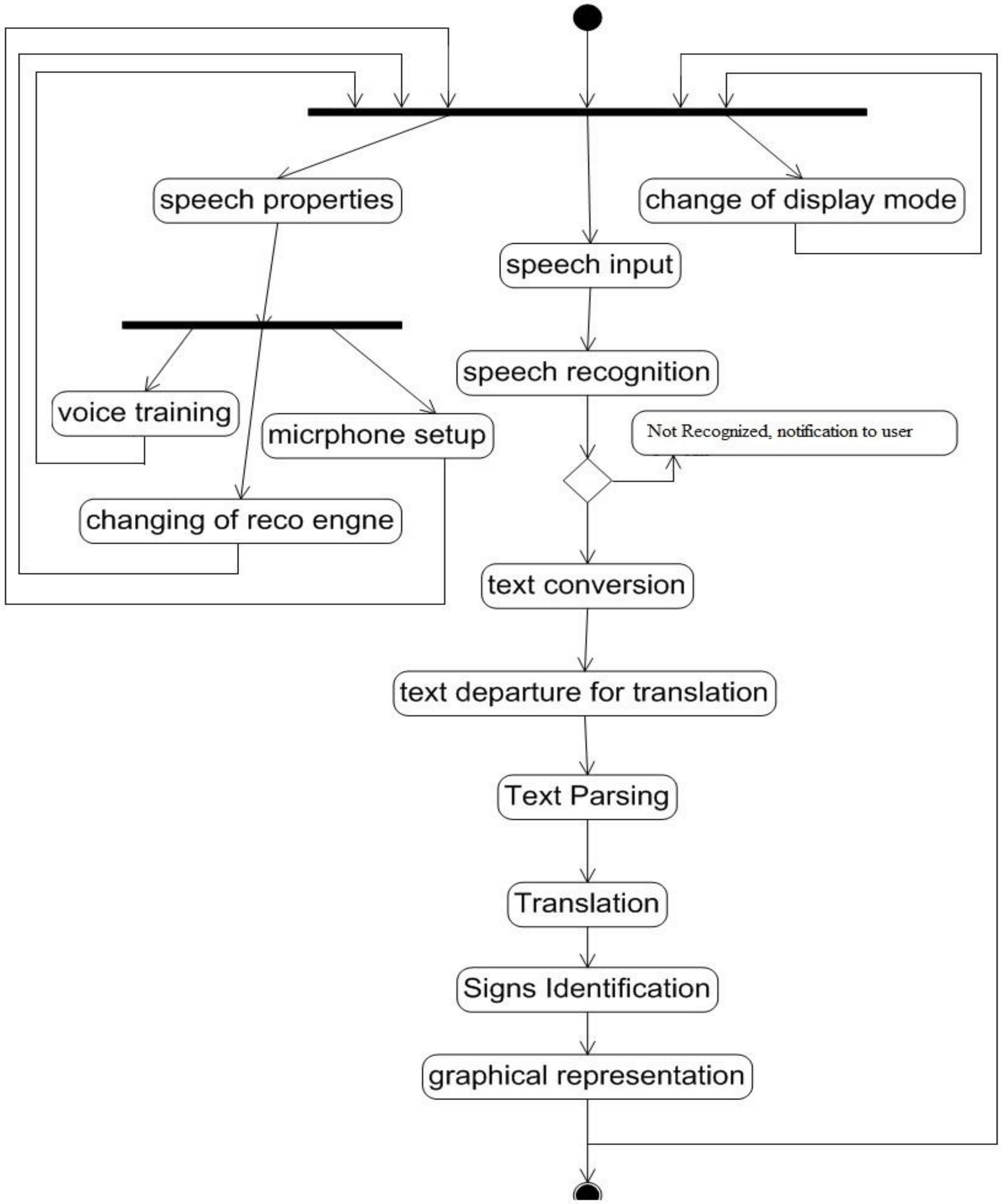
Class Diagram

Following table presents the modules/classes used in the development of PSLIM.

Modules/Classes			
No	Name	**Description**	**Related Classes**
1	Form1	Control the GUI for Big Window.	Windows Forms
2	SmallMode	Controls the GUI for small Window.	Windows Forms
3	PSLIMController Class	Contains all functionality Modules and Queues, and handles communication between all modules.	
4	PSLIMSpeechController	Converts Speech input in to text for further processing.	
5	TranslationController	Translate text into sign language.	
6	DisplayController	Controls the graphical representation.	
7	PSLIMItem	Data structure for individual word with its Part of speech and other information.	
8	SentenceParser	Parse a string of English sentence into Penn Treebank parse tree.	
9	PSLIMSentence	Analyze and reorder the parsed sentence for translation.	
10	Wordlookup	Replace each word with its corresponding sign clip.	

Activity Diagram

Mentioned below activity diagram illustrates the workflow of PSLIM.

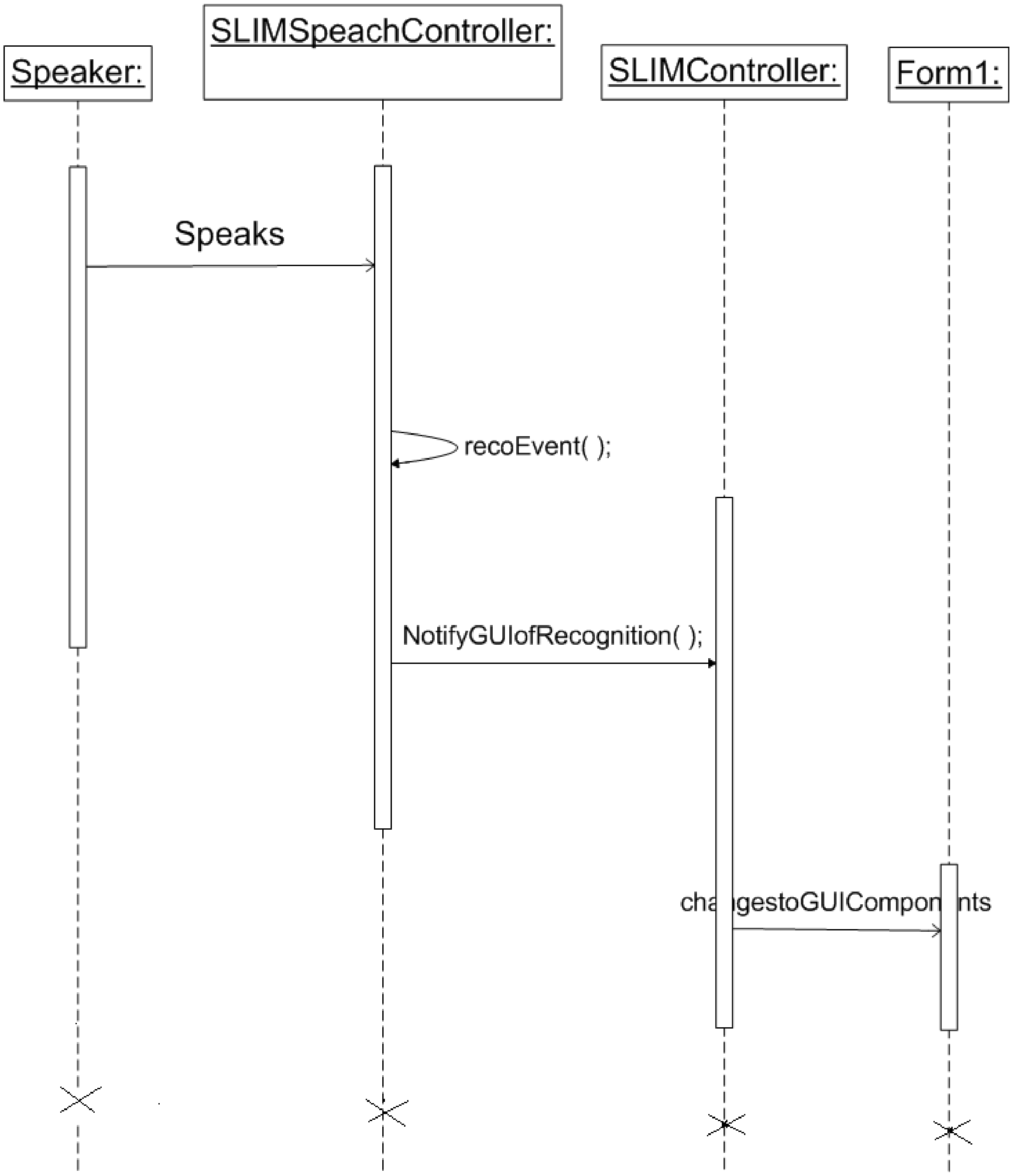


Activity diagram

Sequence Diagrams

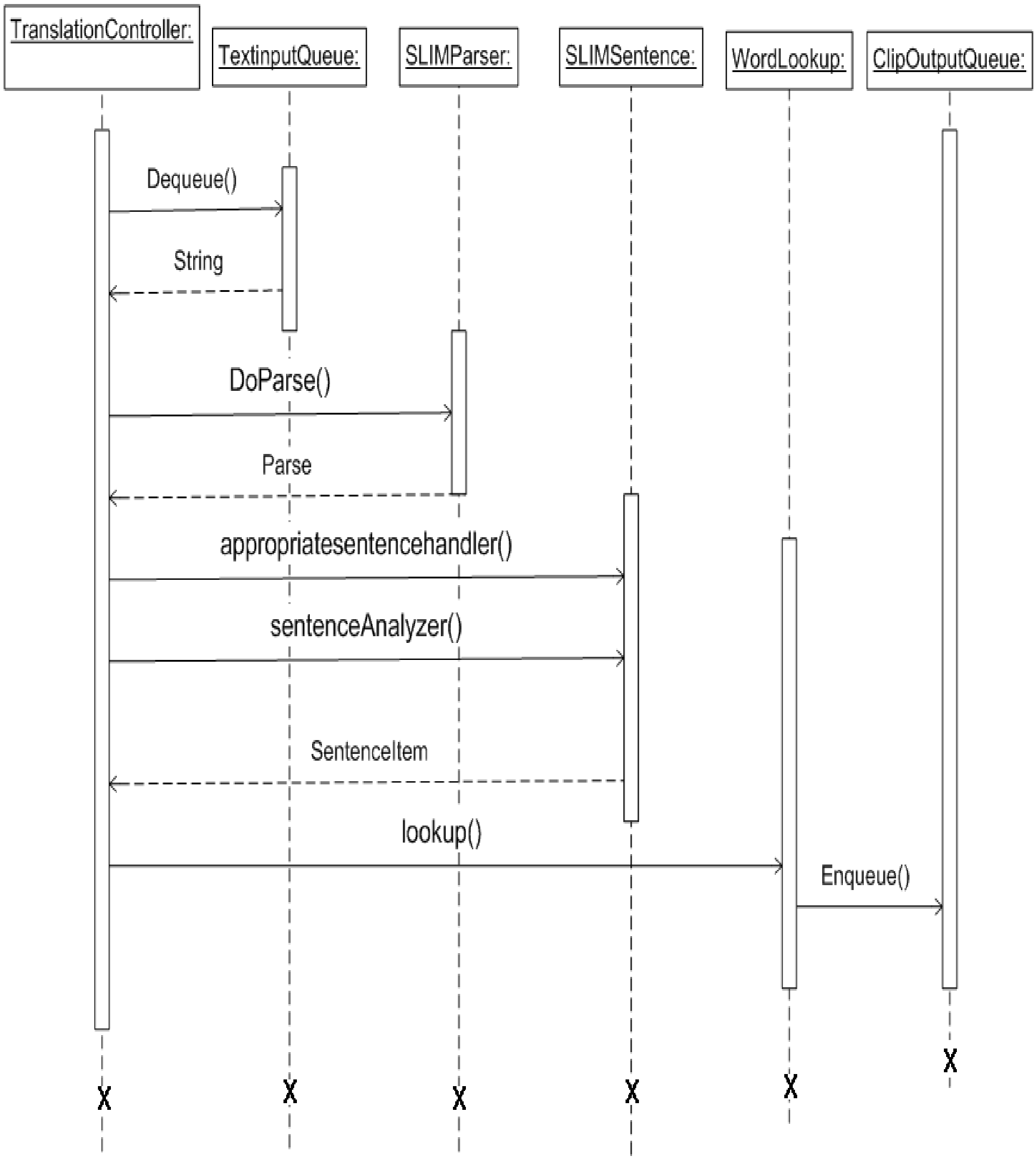
It shows time based dynamics of the interaction of PSLIM.

Speech Recognition



Sequence diagram of speech recognition

Translation



Sequence diagram of speech Translation

Testing Strategies

The basic strategies that were used for testing was following

- Black Box testing
- White Box testing
- Unit Testing
- Regression Testing
- System Testing

Each of the testing strategy is discussed as following:

a. Black Box Testing

In Black Box testing we have only tested the functionality without any regard to the code written. If the functionality which was expected from a component is provided then black box testing is completed.

Test Case 1. Speech to Text				
Description:	Using this test case Speech to text conversion can be checked			
No.	**Valid Input**	**Invalid Input**	**Output for valid input**	**Output for invalid input**
1	Clearly spoken sentence	Unclearly spoken sentence	Correct conversion into text	Incorrect conversion into text

Test Case 2. Translation				
Description:	Using this test case Translation from English to PSLIM can be checked			
No.	**Valid Input**	**Invalid Input**	**Output for valid input**	**Output for invalid input**
1	Correct English Sentence	Incorrect English Sentence	Sign Language Sentence delivering the same message as English sentence	Sign Language Sentence delivering unpredictable message

b. White Box Testing

In White Box testing we have tested internal code of every component and it was checked that the code written is efficient in utilizing various resources of the system like memory or the utilizing of input and output. Open source tools were used to perform memory leaks detection, and I/O optimization analysis. As part of white box testing strategy, we have used unit testing as well.

c. Unit Testing

In unit testing we have checked that all the individual components were working properly. Before integration of the entire components unit testing is essential because it gives a confidence that all the components individually are working fine and ready to be integrated with the other ones.

d. Regression Testing

In Regression testing the software was tested against the boundary conditions. Various input fields were tested against abnormal values and it was tested that the software does not behave abnormally at any time.

e. System Testing

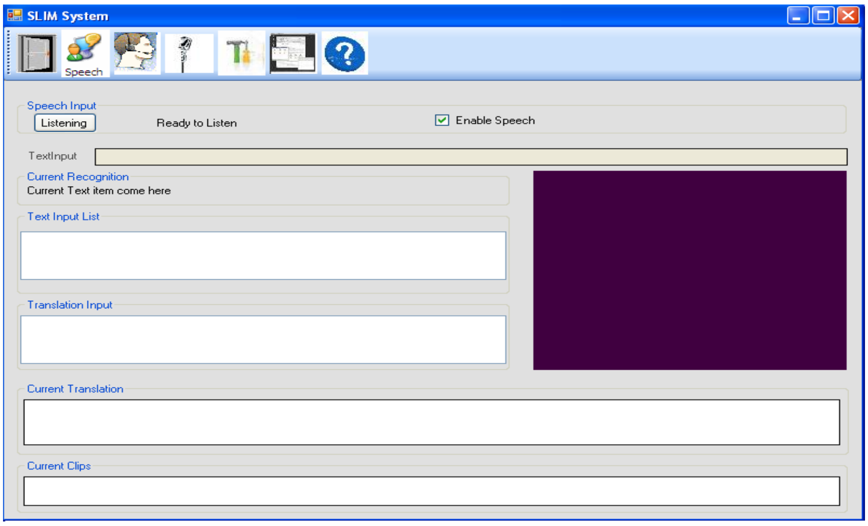
When all the units were working properly then we performed system testing where we have checked all the integrated components as a whole and looked for possible discrepancies, which could have arisen after the integration. A full end-to-end system is validated as given in section 5.4.

APPENDIX D

USERS MANUAL

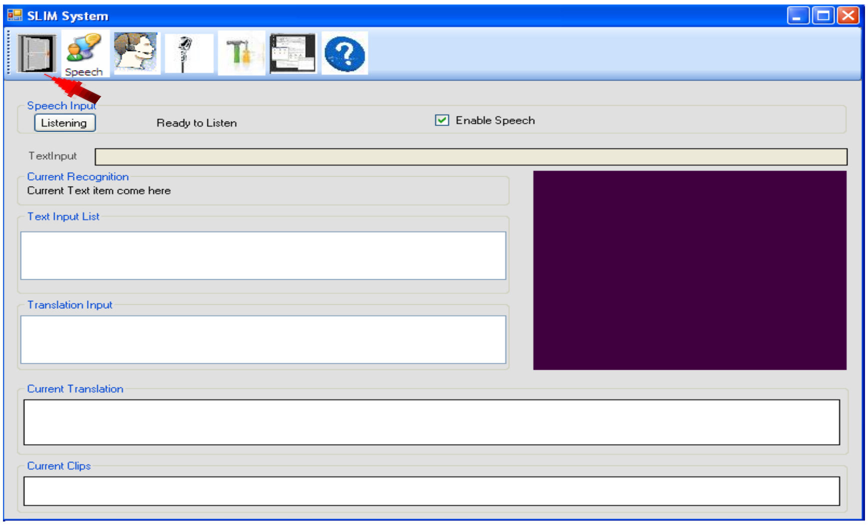
Main Window

The main GUI of the software looks like this.



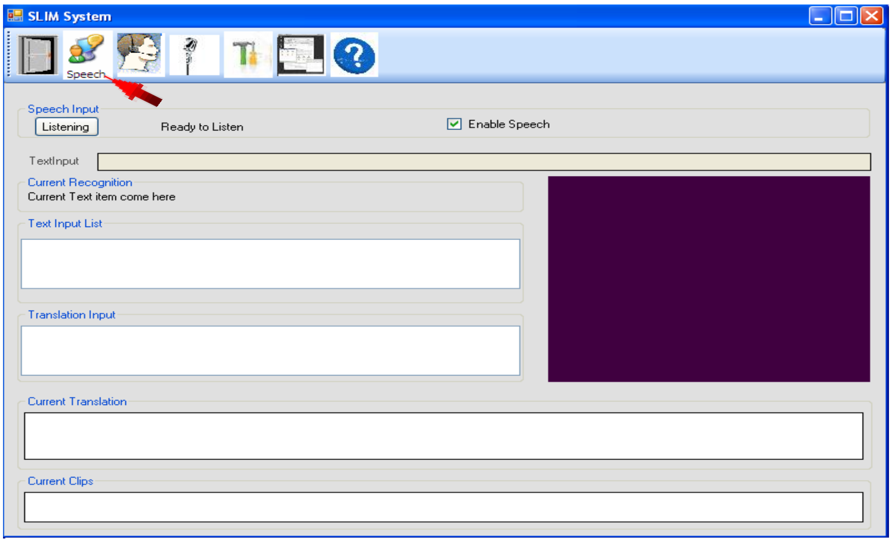
1.1 Exit

To close the application, click on the button as mentioned below in the figure.



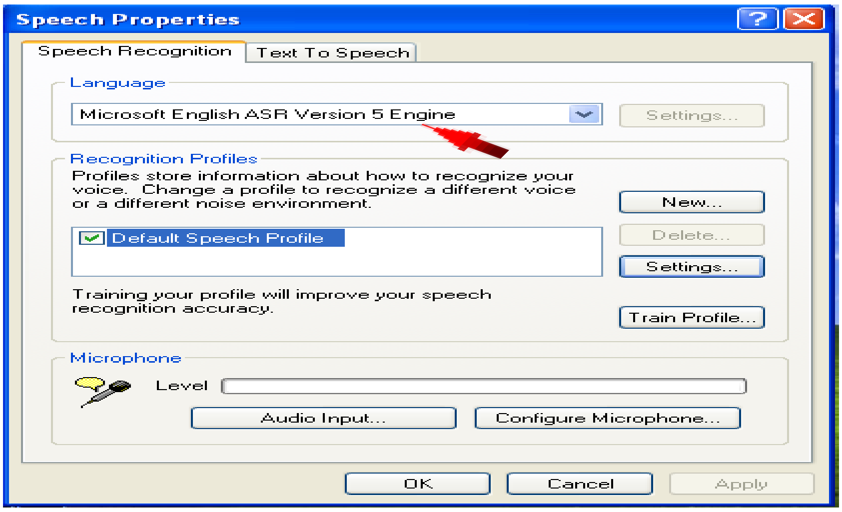
1.2 Speech Properties

In order to change speech properties click the Speech Properties Icon in the menu bar as shown below.



1.2.1 Language

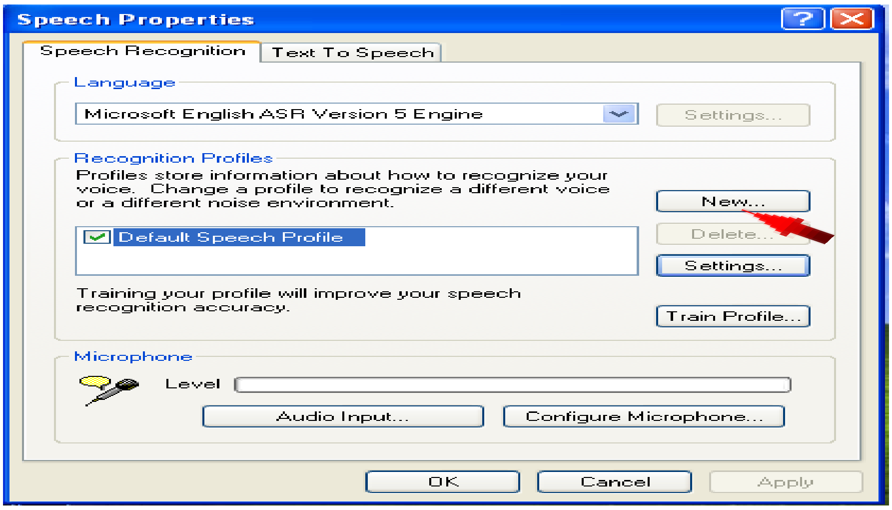
Click Speech Recognition tab and select the language (Microsoft English ASR Version 5Engine).



1.2.2 Recognition Profile

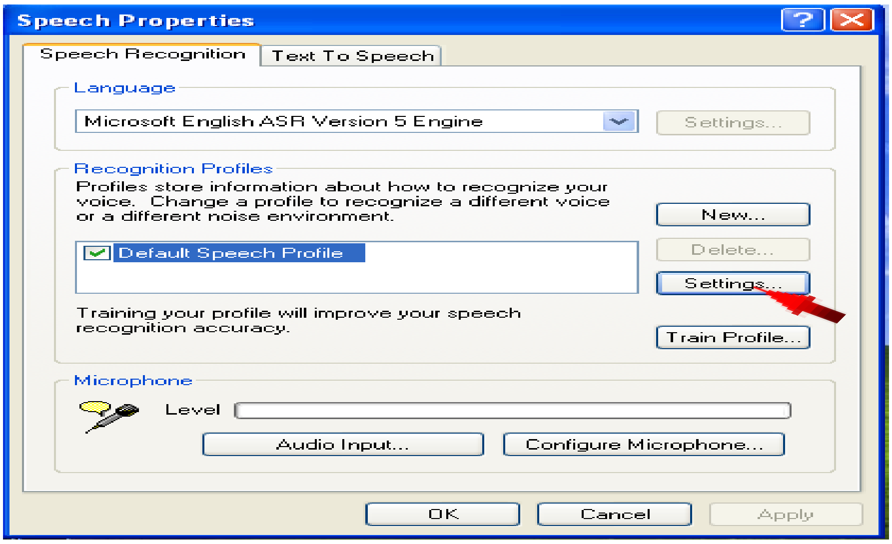
1.2.2.1 New

Click the new button in order to change the profile name.



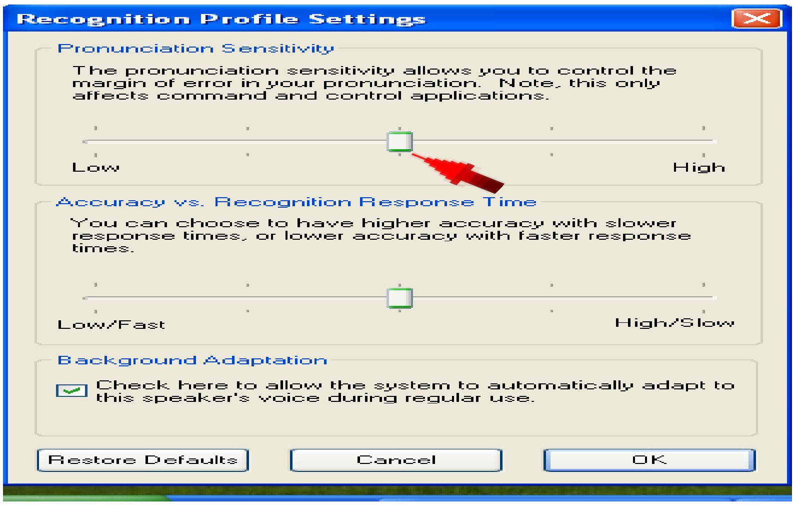
1.2.3 Setting

Settings button is for the recognition profile settings. If changes needed in pronunciation sensitivity, accuracy vs. response time and background adaptation one can click setting button as shown below.



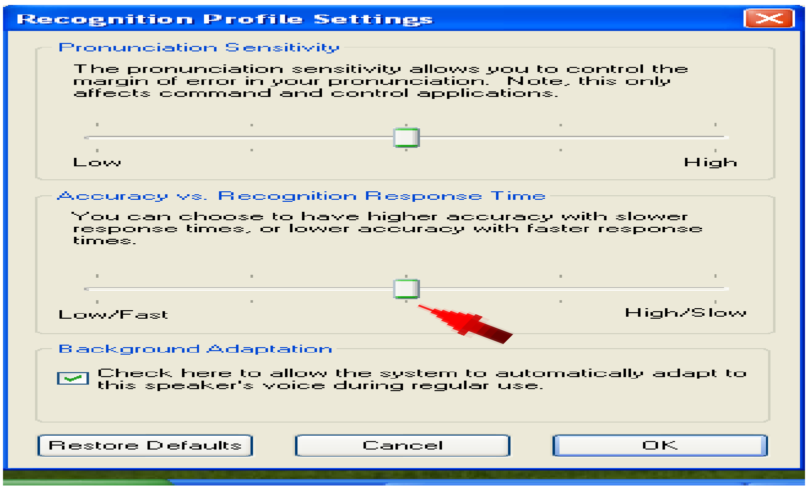
1.2.3.1 Pronunciation Sensitivity

The pronunciation setting allows to control the margin of error in user's pronunciation, whether the system should reject user's command if it is not certain what is said (high sensitivity) or always act upon the command even if it is unsure about what is said (low sensitivity).



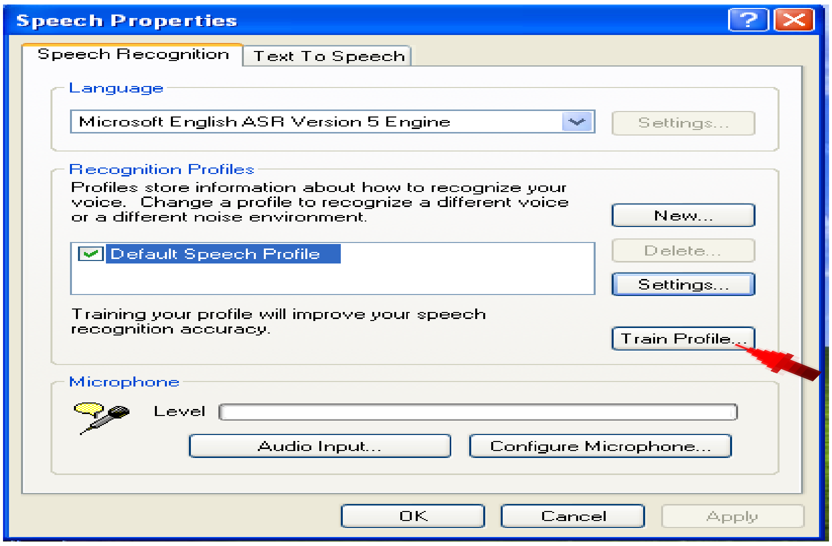
1.2.3.2 Accuracy vs. Recognition

You can choose higher accuracy with slower response times or lower accuracy with faster response times.



1.2.4 Train Profile

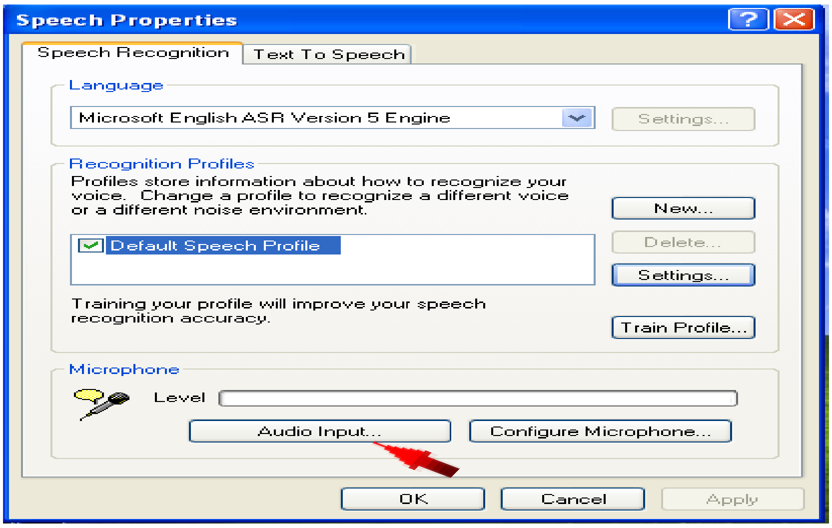
In order to improve your speech recognition accuracy, click the Train Profile button. Training wizard can also be opened from main GUI of application.



1.2.5 Microphone

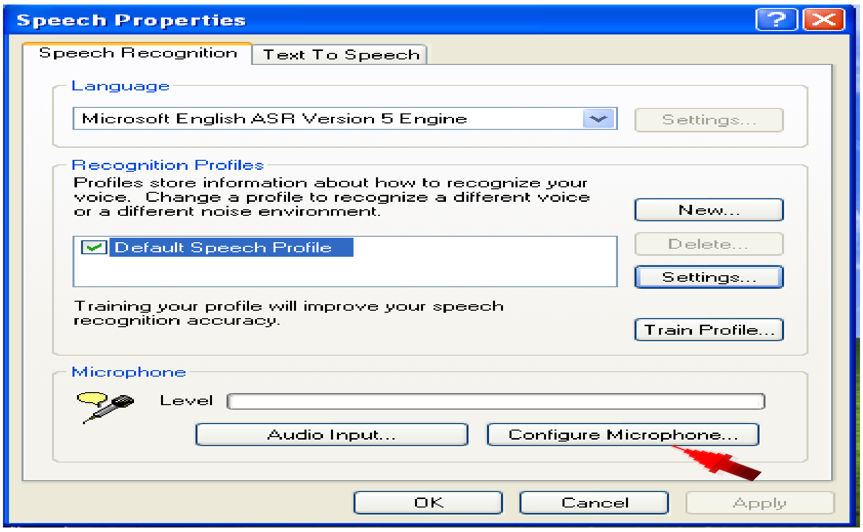
1.2.5.1 Audio Input

Click the audio input button to change the audio input device.



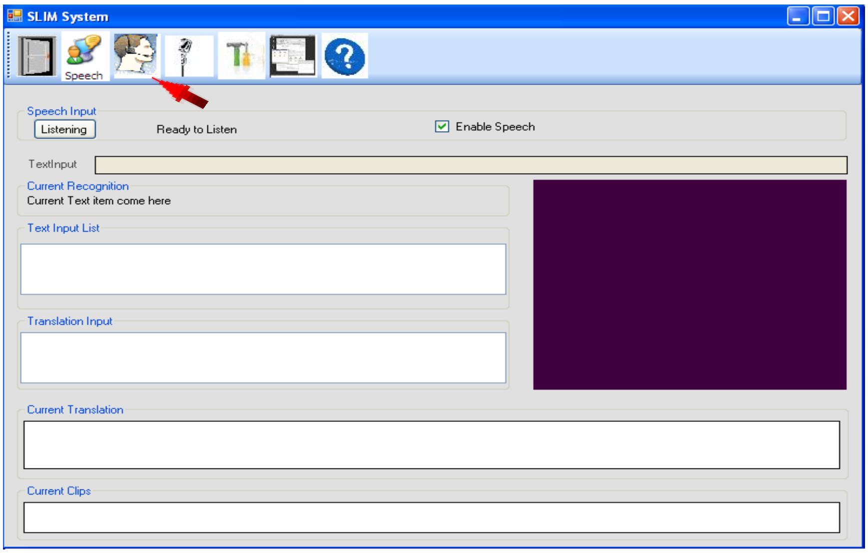
1.2.5.2 Configure Microphone

Click the configure microphone button to configure the microphone setting according to user's need. Microphone configuration can also be performed from GUI application.

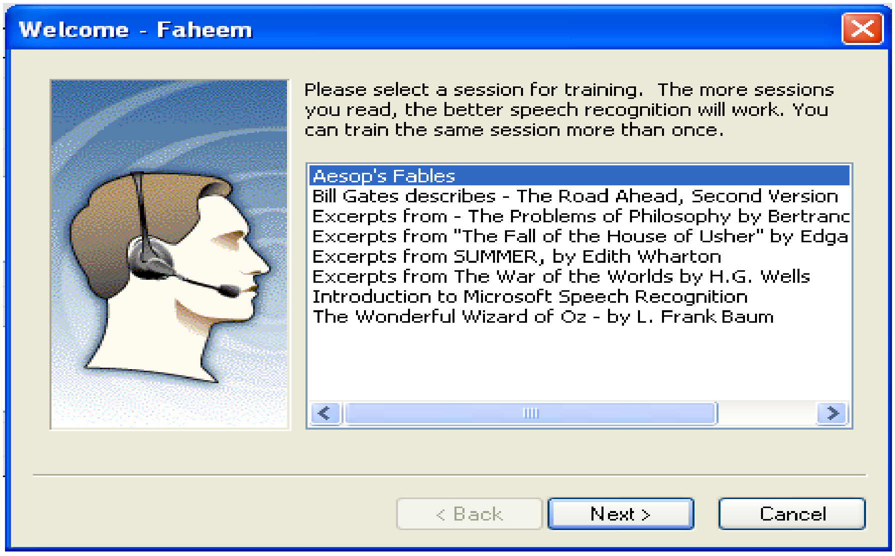


1.3 Engine Training

By clicking on this button, Engine Training wizard will appear. User can train it by following the instructions, so it will improve the recognition quality.

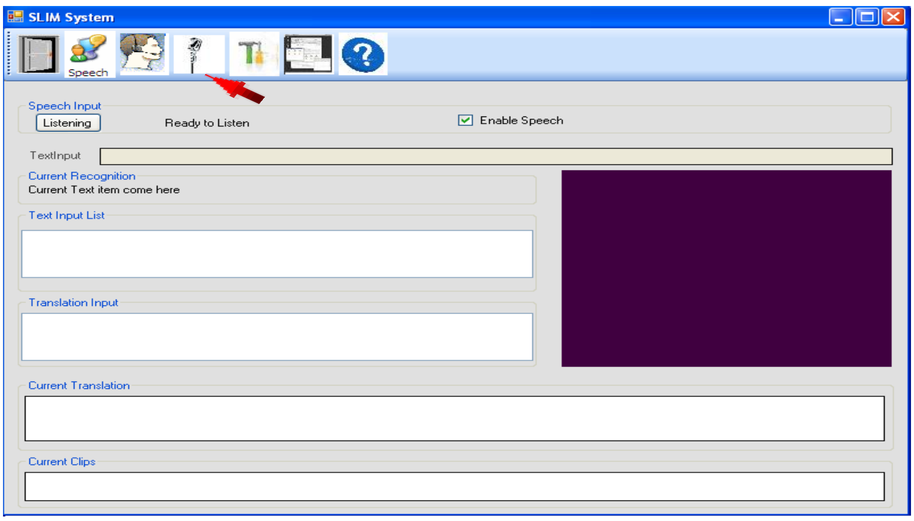


Following Window will appear by clicking Engine Training Wizard. Follow the instructions to train the profile.



1.4 Configure Microphone

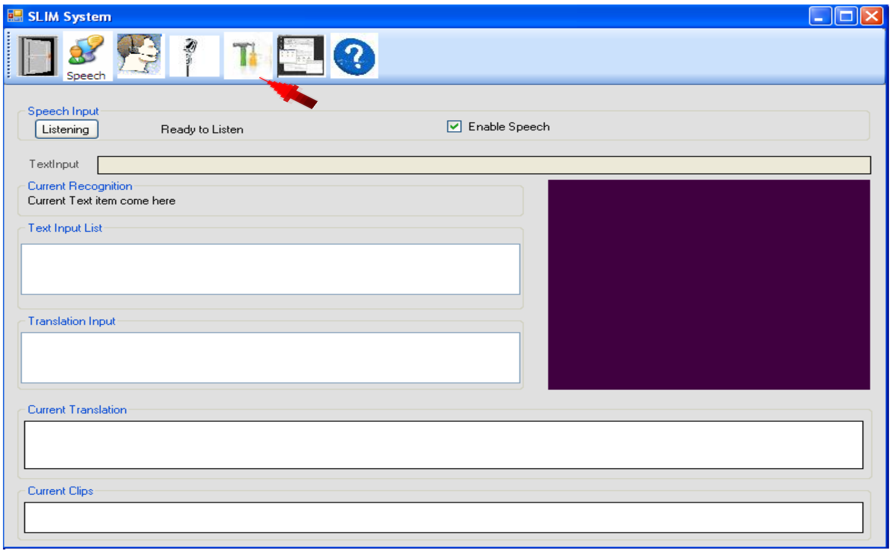
When "configure microphone" button is clicked, a new window opens. By using this, we can configure microphone.



Following Window will appear after clicking Configure Microphone button. Follow the instruction on the wizard to configure the microphone.

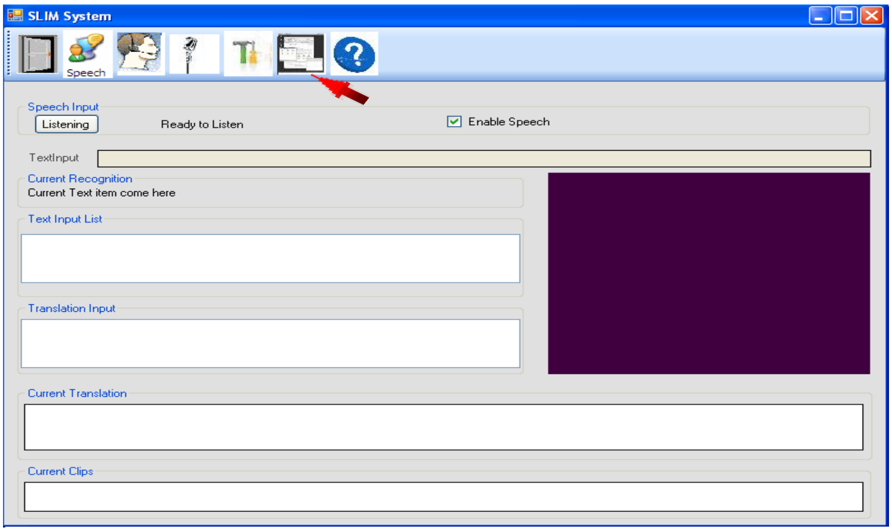
1.5 Configuration

When "Configuration" button is clicked a new window appears. By using this, we can configure paths of various add-ins of the application such as path of the Clip Database, Word Net database etc.

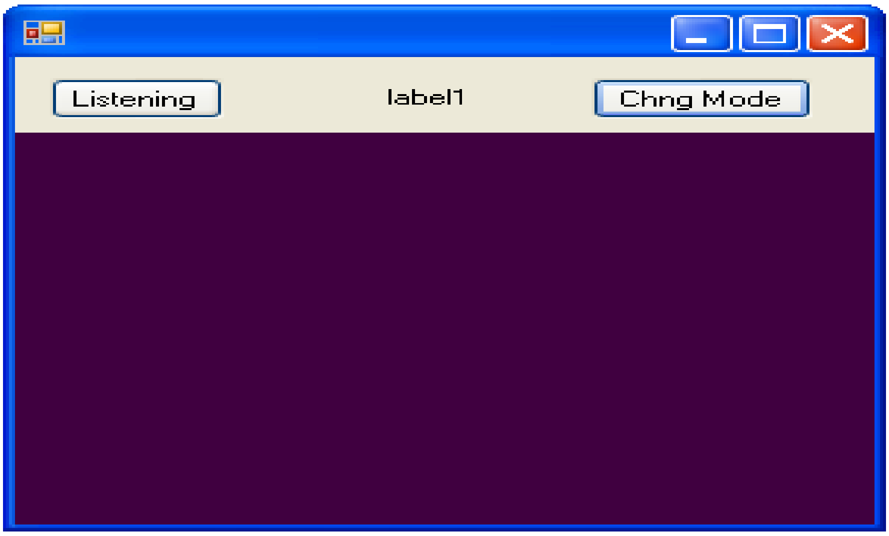


1.6 Switch Mode

When below mentioned button is clicked, main window disappears and small window of the application appears.

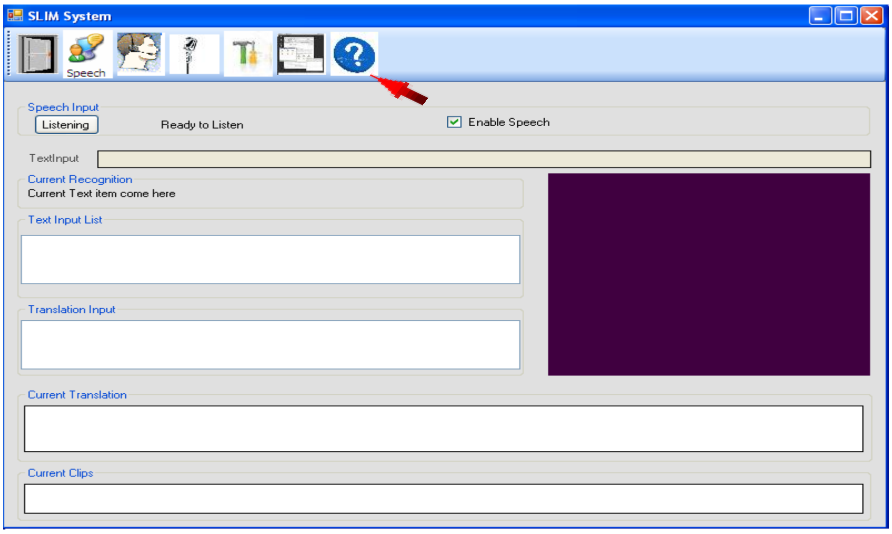


Following window appears when switch mode button is clicked.



1.7 Help

When Help button is clicked, help Manu appears.

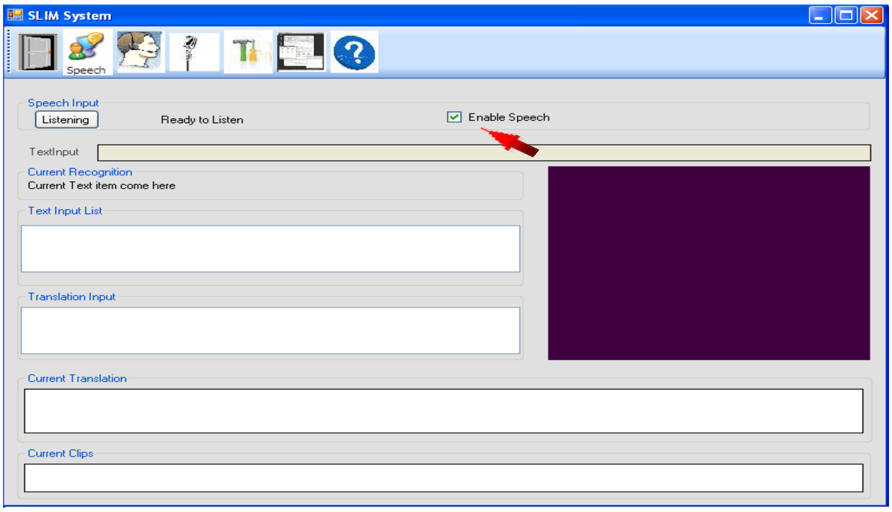


1.8 Speech Input

This group shows events related speech.

1.8.1 Enable Speech

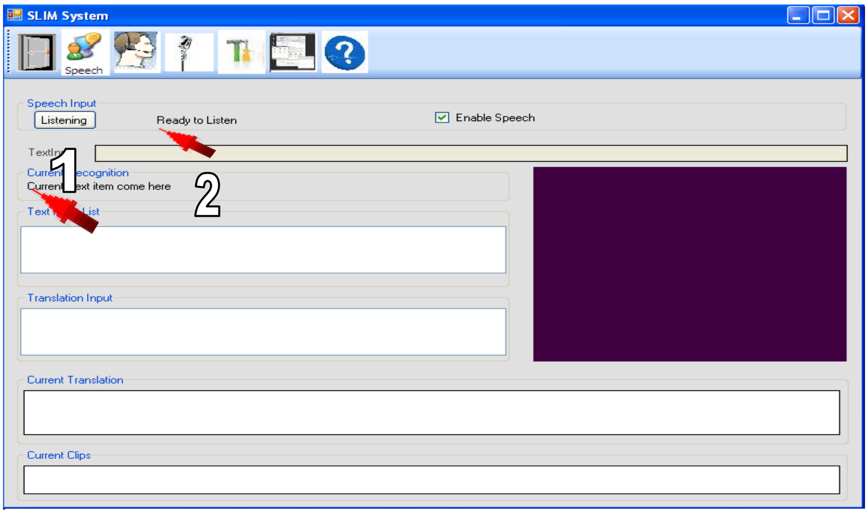
This buttons is by default unchecked. In that case system will not accept speech input. Enable this option in order to give speech input.



1.8.2 Recognition report

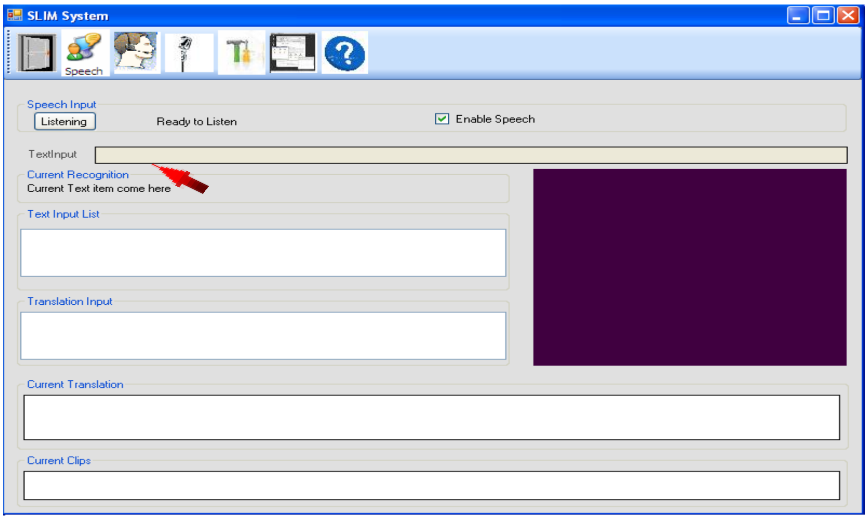
These two controls are for providing speech related reports.

1. It notifies the user whether speech has been recognized or not.
2. It notifies the user that the system has started or ended accepting audio input.



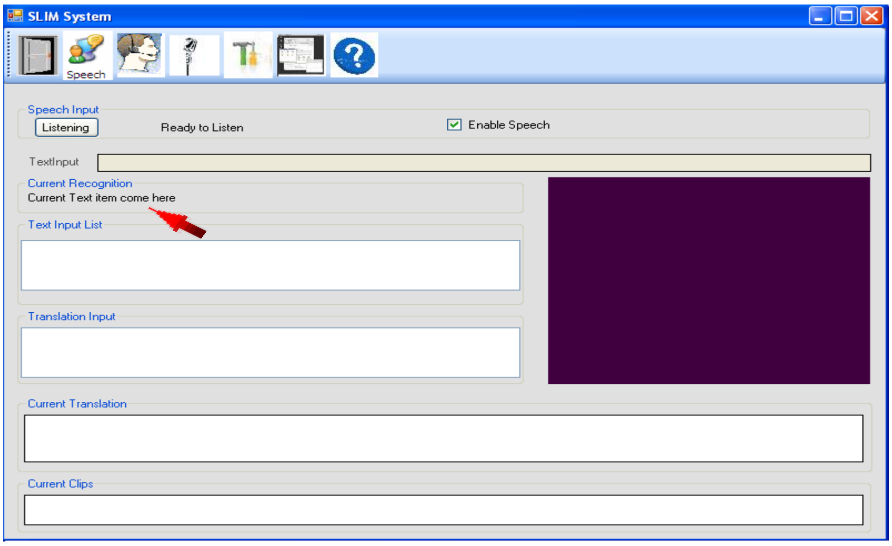
1.9 Text Input

If speech is disabled than system will accept text input. Text can be written in the mentioned area to be translated to PSL.



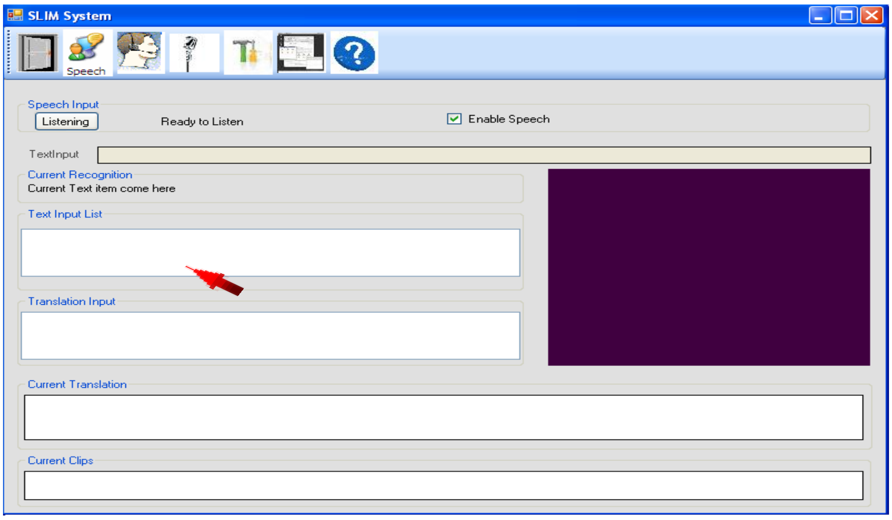
1.10 Current Recognition

If speech is enabled then this group will show the text recognized by the speech recognition engine.



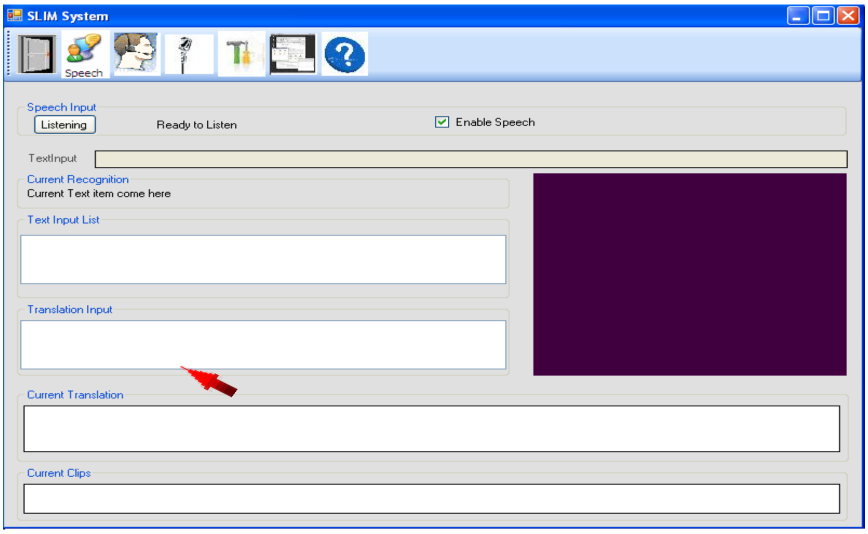
1.11 Text Input List

This group show the list of input text strings given to the system.



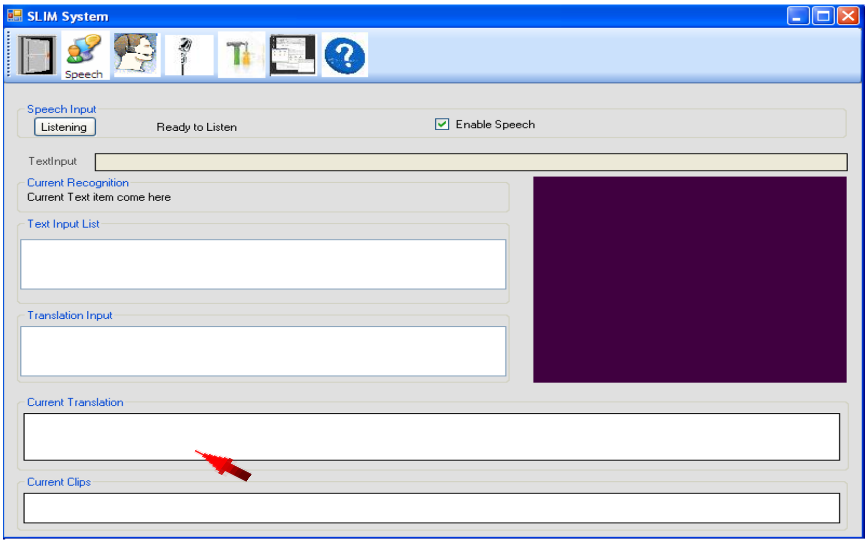
1.12 Translation input

This group will show list of text strings which are processed by the Translation module.



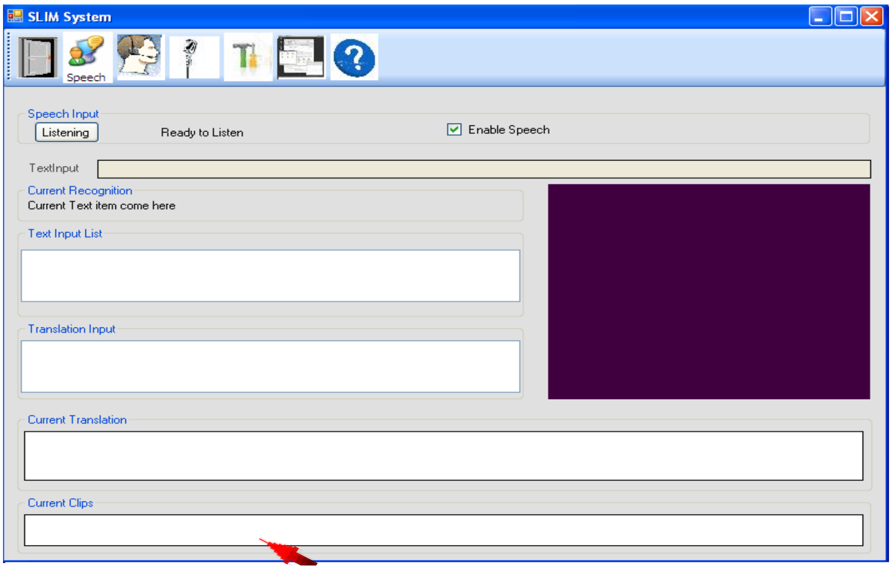
1.13 Current Translation

This group will show the last item translated by the Translation module.



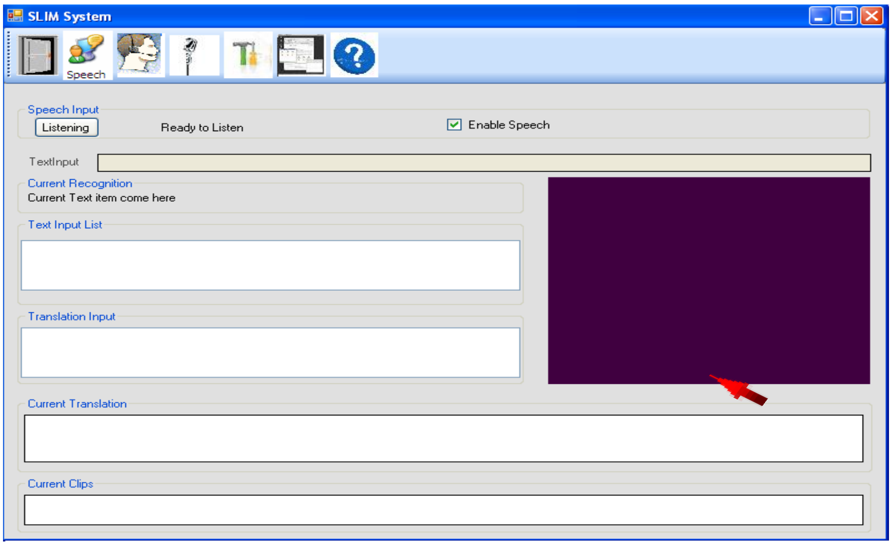
1.14 Current Clips

This group shows the list of clips of the current sentence displayed by the Display module.



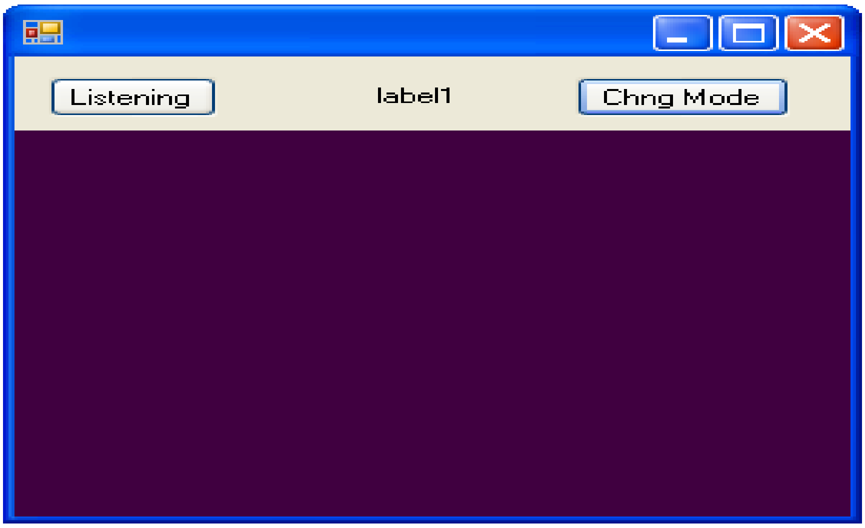
1.15 Clips View

This Panel displays the video clips of the signs.



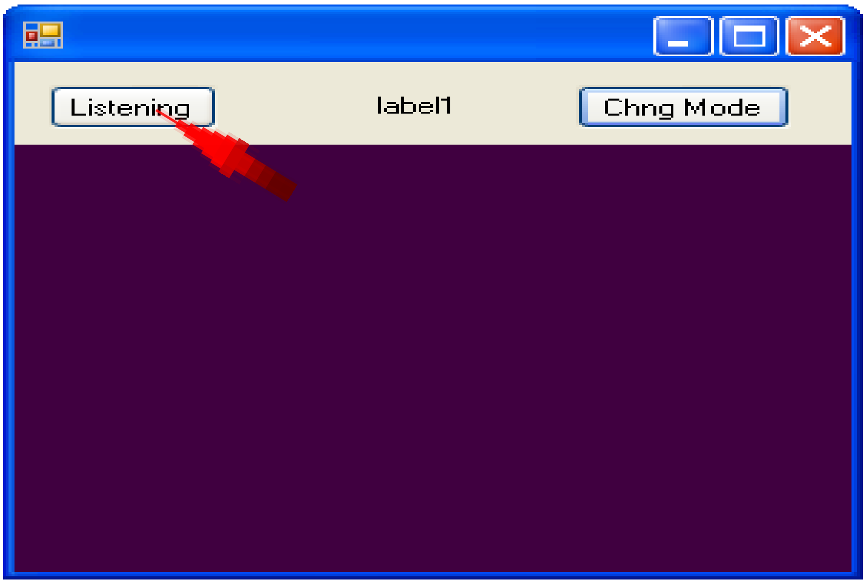
7.3 Small Window

The Small Window of the system looks like this.



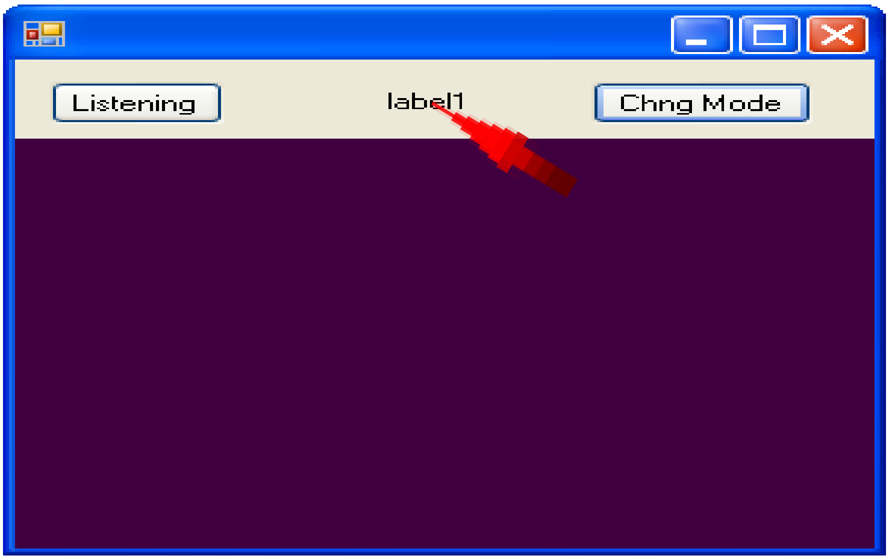
2.1 Listening

It notifies the user that the system has started or ended accepting audio input.



2.2 Label

It notifies the user whether speech has been recognized or not.



2.3 Change Mode

When we click on this button Small Window disappears and the Main Window appears.

2.4 Small Panel

This Panel displays the video clips of the signs.