- **Testing of the Product / Technology:** This system is being evaluated very rigorously every day by a large setof user groups regularly with new set of test sentences. Rigorous validation modules allow the technology to be tested and upgraded regularly.
- **IPR / Open-source:** IPR: IIIT Hyderabad and MCIT. (Also to be available as open system)
- **Potential beneficiaries:**

  a) General Indian population who needs to read English documents.

  b) Web sites, search engines and portal that would need translation services from English to Indian languages

  c) Government and non-government organizations that require automated translation services from English to Indian languages.
- **User-agency tie-up:** NA
- **Name and address of the Resource Person:**

Prof. Rajeev Sangal
Language Technologies Research Centre
International Institute of Information Technology
Gachibowli
Hyderabad
India 500 019
Phone: (91) (40) 2300 1412, 2300 1967 x 144
Fax: (91) (40) 2300 1413
Email: ltrc@iiit.net

6.3 Text to Speech System

6.3.1

- **Name of the Technology**: Bengali Text to Speech Synthesis System
- **Nature of Technology**: Human Machine Interface System (It converts the given written Bengali text (in ISSCI) into unintonated phonetically clear speech)
- **Level**: Technology
- **Technical Description:** Speech generation is the process, which allows the transformation of a string of phonetic and prosodic symbols into a synthetic speech signal. The quality of the result is a function of the quality of the string, as well as of the quality of the generation process itself

  In the past few decades, various Researchers have worked in the area of Speech Synthesis and Recognition and have developed different algorithms and methodologies for different speech technology development. In area of Speech Synthesis there are a number of different methodologies like Formant, Articulatory, Sinusoidal and Concatenation Synthesis.

  In the last decade there has been a significant trend for development of speech synthesizers using Concatenative based Synthesis techniques. There are a number of different methodologies for Concatenative Synthesis like TDPSOLA, PSOLA, MBROLA and **Epoch Synchronous Non-Over Lapping Add (ESNOLA)**.

  In TDPSOLA method based di-phone concatenative technique has inherent problem in introducing intonation and prosody. In manipulating pitch for introduction of intonation the phonetic quality is often seriously compromised as because it is only pitch synchronous. Only pitch synchrony does not guaranty the preservation of phonetic quality.

  MBROLA has the same problems. Along with those, to accommodate intonation a large multiplicity of the diphones is required. This is a major problem in building all necessary elements of the signal dictionary.

ESNOLA technique provides the complete control on implementation of intonation and prosody. It allows judicial selection of signal segment so that smaller fundamental parts of the phonemes may be used as units reducing both the number and the size of the signal elements in the dictionary. Further the methodology of concatenation provides adequate processing for proper matching between different segments during concatenation. The use of special type of basic signal segment makes the size of signal dictionary very small so there is a possibility of its implementation in low-cost, general-purpose electronic devices.

Recently CDAC (Kolkata) has produced user friendly complete TTS for Bangla using the ESNOLA technique with very high intelligibility and naturalness of phonetic quality. It was possible because the ESNOLA supports introduction of jitter, shimmer and complexity perturbations. Recent development in ESNOLA technique has also shown the capability of dealing with the complexity mismatch and pitch mismatch across the concatenation boundary. A tentative demonstration model of intonated Bangla speech is also ready at CDAC (Kolkata). The final version awaits development of intonation and prosodic rules, which are not available in any Indian languages. The methodology of developing rule bases has also been completed. Absence of adequate databases has held up further development.

Since indigenous technology of adequate refinement for development of TTS in Indian Languages is already available we do not see any advantage in wasting time over experimenting with the technologies available for European Languages, particularly when we know their deficiencies vis-à-vis ESNOLA technique and the fact that the phonetic structure of Indian Languages differ significantly from that of European Languages. We intend to develop TTS for all the major Indian Languages either at CDAC (Kolkata) or to fully support these developments any where through technology transfer. The estimated time for the complete development is around one year provided the rule for phonology; intonation and prosody are carried out simultaneously.

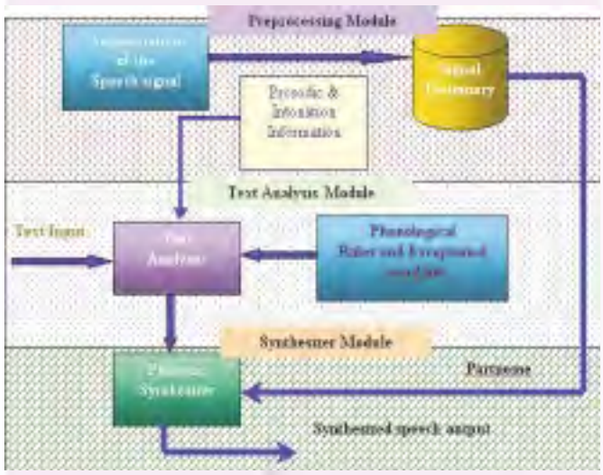


Figure 1. Basic Block Diagram of TTS System using ESNOLA Technique

The above block diagram (Fig.-1) describes the basic part of the ESNOLA technique for the development of text-to speech synthesis system.

It consists of there part 1. Preprocessing module 2. Text analysis module 3. Synthesizer module.

- **Preprocessing module:** In this module the required speech segment database is created from the pre-recorded natural speech signal. In our system we called the segment as pratneme. The advantage of using partnemes as the basic unit is the simplicity of introducing intonation and prosodic rules into the synthesized speech signals. Though prosody and intonation have not been implemented in the present developed system due to the lack of intonation and prosodic rule but the implementation methodology development is tested.
- **Text analysis module:** The Text analysis module is the front-end language processor of the Text-to-Speech System, which accepts input text and generates corresponding phoneme string and stress markers. On many occasions the Text Analyzer consists of a natural language processing module (NLP), capable of producing a phonetic transcription of the text read, together with the desired intonation and rhythm (often termed as prosody).

- **Synthesizer module:** It is the task of the Synthesizer module to combine splices of pre-recorded speech and generate the synthesized voice output. A sequence of segments is first deduced from the phonemic input of the synthesizer. If required, the prosodic events may be assigned to individual segments based on the information extracted by the NLP.

  The Synthesizer Module functions in the following way:

  The Phoneme string input from the Text Analyzer is assigned tokens based on the indexing of the segmented partneme voice signals.

  Modification of pitch, amplitude and duration of the vowels to implement the prosodic and intonational data may be done.

  The selected segments are concatenated to get the raw output signal.

  Spectral smoothing is performed on the concatenation points to remove mismatch and other spectral disturbances to generate the final voice output.

  Specification: unlimited, flat, phonetically clear Bengali concatenative synthesizer using ESNOLA technique.

  O/S: Windows and NT.

  Front-end:

  It also available in dll from which can be integrated with other application.

- **Expandability**: Technology can be used for development of text to speech synthesis system in other Indian language.

  Intonation and prosodic incorporation can be done for naturalness.

- **Portability**: this system can be easily integrated with other application for information disbursing in local languages i.e. telephone address enquiry system (197 pariseba)

- **Availability of documentation**: documentation is available

- **Name and address of the Resource Person:**
  A.B.Saha (amiya.saha@erdcical.org)
  Shyamal Das Mandal
  (shyamal.dasmandal@erdcical.org)
  C-DAC
  Kolkata 700 091
  Web : http://www.cdacindia.com

6.3.2

- **Name of the Technology** : ***Malayalam Text to Speech (TTS) System (SUBHASHINI™)***
- **Nature of Technology** : Human Machine Interface System
- **Level** : Product
- **Technical Description:** The Malayalam Text to Speech system SUBHASHINI™ is a Windows based software, which converts Malayalam Text files into fairly intelligible speech output. The software is integrated with a text editor having ISCII, ISFOC and UNICODE support. The editor supports INSCRIPT Key board layout.

  The TTS is based on Speech synthesis by diaphonic concatenation and consists of the following three modules together with the user interface module.

  • Diaphone Database

  • Text Processing module

  • Speech Synthesiser

Block Diagram of the system is given in figure 2

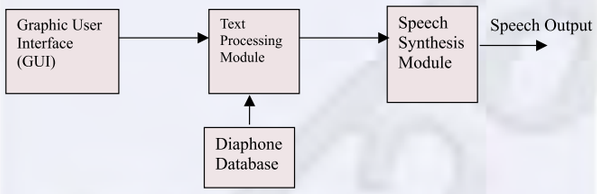


Figure 2.1 Block diagram of TTS

The Diaphone database consists of 2500 diaphones segmented from recorded words. All the commonly used allophones are also considered.

The text-processing module organizes the input sentences into manageable lists of words. It also identifies the punctuation symbols, abbreviation, acronyms and digits in the input data and tags the input data. These are then processed and converted to phonetic language – a language that the speech engine is able to recognise.

ISFOC Data (Input from file or User entered data)

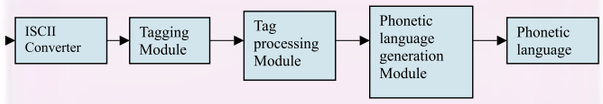


Figure 2.2 Block diagram for Text processing module

The concatenation of diaphones corresponding to the text is done in the Synthesis module and we get speech output. We are using the MBROLA speech engine for speech synthesis.

Developed on VB platform, Subhashini runs on Windows operating system.

- **Portability/Expandability/Scalability:** The system can be expanded to handle intonation/prosody. Integration with website to readout web content is also possible.
- **Readiness of Transfer of Technology (ToT):** Ready for ToT.
- **Availability of documentation:** Available.
- **Testing of the Product/Technology:** Third party testing to be done.
- **Potential beneficiaries:** Speech and sight disabled, Telecommunication, Railways etc.
- **User-agency tie-up:** Nil
- **Name and address of the Resource Person:**
  R. Ravindra Kumar
  Additional Director & Coordinator
  RCILTS-Malayalam, RCILTS- Malayalam
  C-DAC, Thiruvananthapuram
  Email: ravi@erdcitvm.org.

6.3.3

- **Name of the Technology:** ***Text-to-speech For Oriya Language***
- **Nature of the Technology :** Human Machine Interface System
- **Level:** (Product / Technology / Sub-system): Technology & Product
- **Technical Description of the Technology / Product including Basic block diagram, Algorithm used, O/S used, Front-end / user interface, and Specification of the Technology / Product:** *ANNEXURE-I*
- **Representative Snapshot / screenshot of the Technology / Product:** ANNEXURE-II
- **Scalability / Portability / Expandability:** ALL
- **Readiness of Transfer of Technology (ToT):** YES
- **Availability of documentation:** YES
- **Testing of the Product / Technology:** SUBMITTED TO MCIT
- **IPR / Open-source:** IPR - SW1179/2003
- **Potential beneficiaries:** COMMON MAN, BLIND & ILLITERATES
- **User-agency tie-up:** MODULAR INFOTECH, PUNE
- **Name and address of the Resource Person:**
  Dr(Mrs) Sanghamitra Mohanty
  RC-ILTS-ORIYA,
  Department of CSA,
  Utkal University,Bhubaneswar - 751004

ANNEXURE – I

TTS system provides an interface through which a user enters certain text/document and it is the software that reads it as natural as a human. The basic approach followed here is, first to analyse the document (language, font etc.), and then extract words from the text, try to parse individual words into vowels and consonants respectively. Then corresponding to these vowels and consonants existing (previously stored in the database) ".wav" files are concatenated and played.

Technologies Behind :-

- Creating the wave file database: - For creation of such a database we studied a lot of recorded words and sentences and try to break them into vowels and consonants by minute hearing. Then we analyse those cut pieces and store the appropriate and generalised form in the database.
- Extraction of exact words of a given sentence:- The same words in different sentences have different stress due to its position in the sentence. Appropriate hidden vowels are detected form the words extracted. For example considering a word "ସମୟ" (*SAMAYA*) in *Oriya* is parsed as follows :- ସ୍ + ଅ + ମ + ୟ୍ + ଅ The format of vowel and consonant break point is shown in the figure 1.
- Choosing of appropriate '.wav' file from the database.

  Considering the above example, may be the vowels we get after parsing are the same as 'ଅ' (for *Oriya),* but it is not exactly the same 'a.wav' we concat in every case. Thus, we analyse vowels broadly in three categories as ma:tra: in

  - Beginning
  - Middle
  - End

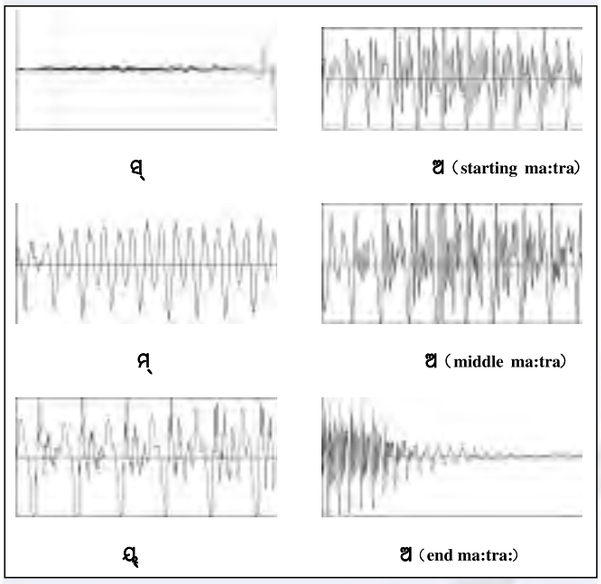
  And this is observed that the duration of ma:tra:s say 'ଅ' here, varies from each other, i.e. 'ଅ' in middle is not the same as that in the end. Again accordingly we need to get the appropriate ".wav" files from the database. As observed in the example the durations of 'ଅ'(*Oriya*) are as follows:

  Starting ma:tra: - 0.065 sec

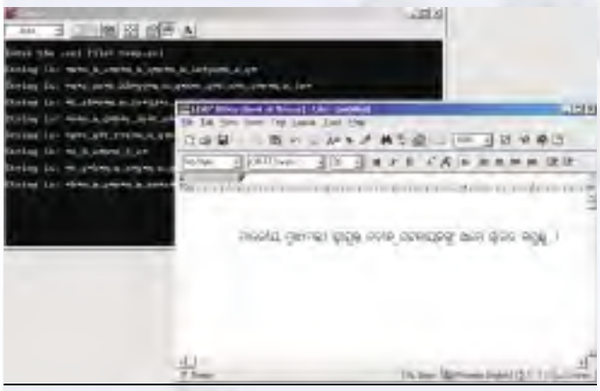
  Middle ma:tra: - 0.105 sec

End ma:tra: - 0.116 sec

- It is observed that concatenation of the wave files is not that natural as expected. This is due to the certain transitions between the characters in the actual pronunciation. Thus we are developing a robust algorithm for the generation of naturalness in the TTS output.



(Figure – 1)



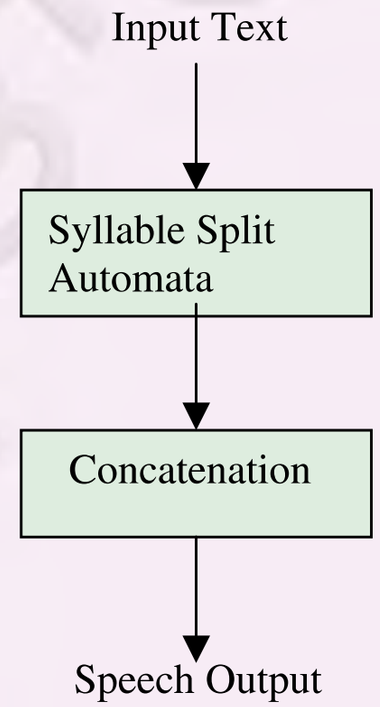
ANNEXURE – II

(Screen shot of ORITTS peaks the .aci file provided by the user and utters)

6.3.4

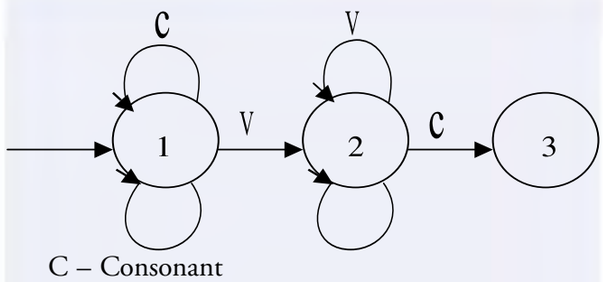
- **Name of the Technology :** ***Ethiroli (Tamil Text to Speech System)***
- **Nature of Technology :** Human Machine Interface System–Language Technology Product
- **Level :** Product
- **A Technical Description :** Ethiroli is a Text to Speech package for Tamil. When Tamil text is given as input, it splits the text into syllables using "CCVC" model. Once the syllables are identified in sequence from the given text, the sound files corresponding to the syllables are concatenated and played in sequence to give the speech output. Ethiroli uses linguistic rules to remove homographic ambiguities.

Block Diagram



The following are the important modules handled in Ethiroli:

- **Syllable Split Automata**: Tamil input text is split into syllables using "cvcc" model. The "cvcc" model is explained by the following diagram



C – Consonant

V - Vowel

Concatenation

Concatenation is a technique for producing sound from the text. It uses set of sounds for all the basic sound elements (syllables) that can occur in the language. The given text is split into syllables, the corresponding sounds are concatenated and played. The advantage of this method is that it gives smooth output without much processing time. It needs the entire list of syllables in the Tamil language.

- **User Interface:**

Input

* Input can be given from a file (TAM encoded) OR

* The following Tamil keyboard drivers are provided for typing

Tamilnet 99 - Standard Tamil keyboard

Typewriter (Thattachu) - Standard Tamil Typewriter keyboard

English 1 (Aangkilam 1) - Transliterated format - Similar to TAB font

English 2 (Aangkilam 2) - Transliterated format- Similar to TAM font

* Save the typed text by "Save" button.

* Clear the input by "Clear" Button.

Output

* "Play" button to get the audio output for the input text.

* To stop the audio use "Stop" button.

- **Specifications of the Technology:** Ethiroli was developed using Visual Basic 6.0 and Access 98 in Windows environment.
- **Minimum Requirement:** Any keyboard driver to type Tamil Font encoding scheme in TAM
- **Representative Snapshots:**

A. Expandability: Yes

B. Portability: Yes

- **Readiness of Transfer of Technology:** Yes
- **Availability of Documentation :** User and Technical manuals are available.
- **Testing of the Product :** The product has been tested with files from corpus.
- **Open Source:** Yes
- **Potential Beneficiaries** : Visually Handicapped Persons, Public and students.
- **User Agency Tie up**: No
- **Name and address of the Resource Person**:
Dr. T.V.Geetha
Dr.Ranjani Parthasarathy,
RCILTS Tamil,
Anna University Chennai 600 025