

Study and Implementation of a Tamil Text-to-Speech Engine

S.Subramanian and K.M.Ganesh.

Students, III yr B.E.CSE,
Department of Computer Science and Engineering,
PSG College of Technology,
Coimbatore - 04, Tamil Nadu, India.

Under the Guidance Of,
Dr. S.N.Sivanandam, Head Of Department,
Department of Computer Science and Engineering,
PSG College of Technology,
Coimbatore -04, Tamil Nadu, India

**கேட்டார்ப் பிணிக்ருந் தகையவாய்க் கேளாரும்
வேட்ப மொழிவதாஞ் சொல்.**

ABSTRACT

Today, There is a wide spread talk about improvement of the human interface to the computer, as no longer people want to sit and type out required data or do they want to read data from the monitor, since there is a painstaking effort to be taken, which involves strain to the eyes. In this aspect, we all know that Speech Synthesis is becoming one of the most important steps towards improving the human interface to the computer, the current focus on research should also be on Tamil, considered as the Language of Gods, which has very less or null ambiguity in speech rules. We present various methodologies of implementation, prospects and challenges of implementation of a Tamil TTS Engine with regard to speech synthesizer and its high level applications. The Implementation of this TTS is done using the concatenation method (used in many popular speech engines available today.) There are discussions on the lexical analyzer, phoneme identifier, voice mapping, speech synthesizer and voice modulation which are the integral parts of a Text To Speech Engine. Also various applications of the Tamil TTS engine have been discussed with an example implementation. Finally it is concluded that the application of Tamil TTS engine as a computer interface will make the lives of many people easy and more productive. The Official Website of the Kural Osai TTS Engine is www.kuralosai.com

Key Words:

Voice Synthesis, Tamil TTS, Speech Synthesis, Local Language Interface development [LLID], Interface Simplification and Design.

Introduction.

The art of making PC's talk has always entranced the human community. After all, Voice is one of the best alternatives for hours of eyestrain involved in going through any

document. Also Voice is a better interface when it comes to un-educated people rather than Graphic User Interface in English. So research is being done through out the world for improving the Human Interface to the computer and one of the best options found out till date is the ability of a computer to speak to humans. Here comes the role of the TTS engines. TTS engines are those small piece of software which will speak out the text inputted to it, as if reading from a news paper. One of the best ones is the Lernout & Hauspie Engine, released around the year 1999. After this there has been may development found around the world in the development of TTS Engines in various languages like English, French, German etc and even in Hindi. This has not been tried out till date, (according to our knowledge) in Tamil. So here is the first step towards making computers speak to Tamilians around the world.

This field of TTS in Tamil has remained untouched for such a long time due to the various facts like,

1. The Complexity of the Tamil Language.
2. The problems, which are posed by the Tamil Grammar.
3. Less Knowledge of Pure Tamil.
4. Huge gap and difference between the Tamil spoken, which is full of slang and the pure Tamil, which is written.

The Various Parts of the TTS Engine are as follows. (Fig 1)

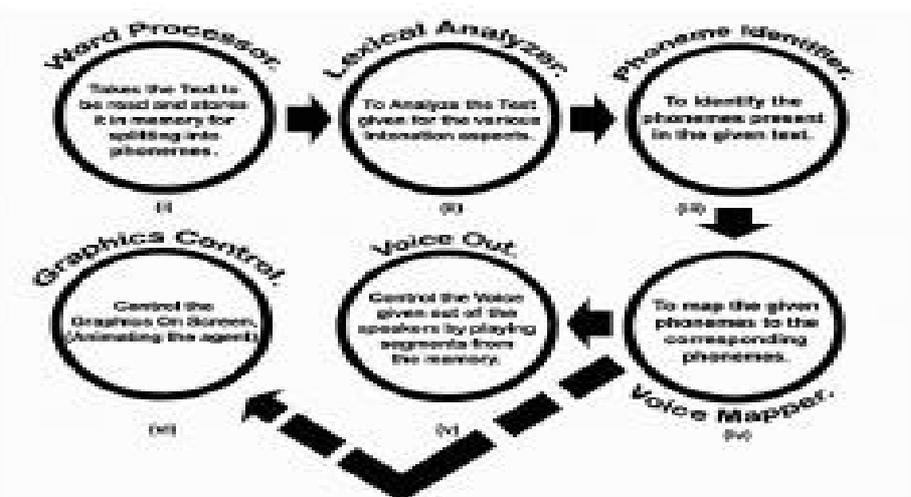


Fig 1.

- | | |
|---|----------------------------------|
| 1. Word Processor | 2. Phoneme Identifier. |
| 3. Text to Voice Database Mapping Tool. | 4. Voice Out & Graphics Control. |

Description and Broad Design of the Kural Osai TTS Engine: -

1) Word Processor:

This is similar to a normal Word Processor, which can include Tamil Fonts. (e.g. One similar to Murasu Anjal) and has the options to enter text phonetically or using the Tamil keyboard. This will get the text entry from the user who wants the data to be spoken out. This

also has shortcut keys and buttons to denote the speed of speaking out, volume and stress. The forthcoming parts will then process these files, which are usually stored using the Mylai font.

This has the following symbols.

1. /s [0 - 100]. (Default value = 50), indicates speed.
2. /v [0 - 100]. (Default value = 50), indicates volume.
3. /t, indicating stress.

Note:

After each word the values of speed, volume etc will be set to default.

Ex) /s[75]/v[75]அ/தன்னை /s[75]/v[75]ஓ/தரு /s[75]/v[75]தெ/தய்வம்.

Will play the words at a speed of 75%, volume 75% and the stress will be correspondingly given to the letters just after /t.

2) Lexical Stress Analyzer.

This part analyzes the given text to check for the grammar of Tamil so as to add the correct stress to various parts of the sentence, other than those, which are explicitly mentioned so that this will help in removing the monotonous effect, which will be found when the sentence is read out normally. This part gives the required modulation to be present in the sentence.

3) Phoneme Identifier.

Phonemes are basic sound units of a language. There are various phonemes in a language, which form the basics of talking that language. In the case of English, there are about 39 phonemes but there are a lot of ambiguities in English. In Tamil, there is very less ambiguity due to the fact that there are 247 letters (not including the VADA MOZHI characters) and pronunciation of them are very distinct. The various letters form a large list of about 300 phonemes in Tamil, which form the basis for the speaking of Tamil.

When a text sentence is inputted into this part, this will divide the various words into their respective phonemes by parsing the various characters and their neighboring characters. For example,

The word அன்னை is parsed as, A னு N னு NAI, where each one have their respective sounds. Similarly any word can be processed. In case of some Tamil words like Ramba, which when written in pure Tamil, Starts with an அ, are also considered since the next letter is also parsed in this phase. The recording of phonemes is based on the various Tamil Pronunciation and Phoneme Splitting Rules quoted by NANNOOL, with some small changes due to the fact that Tamil has developed a lot till then.

4) Text To Voice Mapping Tool :

This part consists of two sub segments.

- (i) Data Base Mapper.
- (ii) Sound Modulator.
- (iii) Sound Concatenator.
- (iv) Prosody Manager.

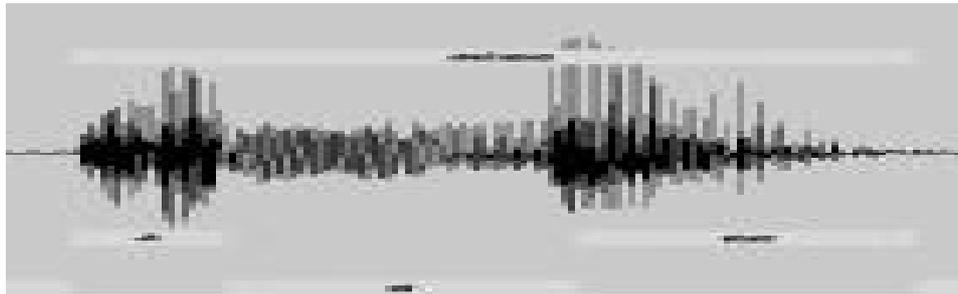
The Data Base Mapping part of the TTS is the One, which maps the various phonemes to their corresponding sounds present in the sound database (phoneme database). In the case of most commonly used words, there is also an existing database of phonetics of these words for faster data base mapping so that the online data flowing in can also be managed.

The Sound Modulator is the one which decides the time duration, pitch, volume and stress according to the instructions provided in the text using the various commands like /s, /v etc. This varies the pitch volume etc, by applying various Signal Processing effects like FFT on the sound from the database. Thus after this phase, the output will be the sound file of the corresponding phonemes.

Then the Sound Concatenator will join these sound files, according to the speed required for each of the letters and the overall speed. For example, the KURIL letters will need a lesser time to be spelt out and the NEDIL letters require more. Also this determines the combination of the consecutive letters which gives rise to the coordinated word sounds.

The Prosody manager is the part, which adds the required effects to the sound file to add emotions to the given voice. This part also requires some parsing of the whole sentence, which leads to variations in the effects.

Figure-2 gives a schematic sound file, which has been formed from the various component phonemes.



5) Voice Out and Graphics Control :

This is the last part of the TTS, which plays the sound data produced by the previous section. Also, this part controls the On screen graphics control, which can be implemented as a plug-in part, by giving it the various emotion controls, which is received from the Prosody Manager, so that the Onscreen display can adjust to the corresponding emotions.

Uses and Future Expansion Possibilities of this TTS:

There are a lot of uses for such a system like this, which will enable the simplification of the user interface, and do a very large part in bring the PC market to the common man. This will enable the development of very easily usable versions of the computers, which will be profusely helpful for the common man, since this will bring him all the required online data etc, on to his desktop, in his own language. These TTS are practically used in systems

like Chat bots, Telephone Interfaces to Online Help Systems, Interactive Sales Computers that can be employed in various fields like online reservation of tickets, Online News Readers (For Flash News On Desktop) etc.

The various future expansion possibilities for this TTS are,

- 1) Inclusion of more Voices and Graphical Faces in place of existing Robot's voice and face.
- 2) Improvement of the smoothness in the case of inclusion of prosody and the naturalization of the voices and interface.

Conclusion :

Thus, after these discussions and the design review, we can come to an conclusion that the interface to computers can be further simplified by employing the Tamil TTS and this will surely help in getting the computers still more closer to the common man.

References:

1. Nannool.
2. Discrete Time Signal Processing, Oppenheim.A.V. and Schaffer.
3. Introduction to TTS engines - The Microsoft Website.

Acknowledgements :

Special Thanks to:

1. Dr. Radhakrishnan, Principal, PSG College Of Technology.
2. Dr. S.N.Sivanandam, H.O.D., Dept. of CSE, PSG College Of Technology.
3. Our Sponsors.