Proceedings of the Technical Sessions, 21 (2005) 1-8
Institute of Physics – Sri Lanka

Text To Speech Conversion - Tamil Language

N. Philips, D.U.J. Sonnadara and M.K. Jayananda
Department of Physics, University of Colombo, Colombo 00300, Sri Lanka

ABSTRACT

A text-to-speech conversion engine was developed for Tamil language that could successfully convert any arbitrary Tamil text into spoken utterance. The back-end of the engine was a sound database created with 343 *phonemes* which were chosen as the basic units of Tamil language. The text-to-speech engine splits the given text into the basic units of Tamil, extracts the correct *phonemes* from the database and concatenates the basic unit waveforms in the correct sequence to produce the required sound waveform. The front-end of the engine is a GUI interface that allows users to input Tamil text in letters or even numbers. The text can be stored to be used at a later time. The Tamil font "*TSC_Avarangal*" was selected as the font to develop the user interface. The sound editing software "*Sound Forge*" was used when preparing the phonemes with correct pitch. This work shows that for languages such as Tamil or Sinhala text-to-speech engines can be successfully developed with the available resources.

Keywords: Text to speech, Tamil language, Speech recognition

1. INTRODUCTION

From the invention of computers, the art of making them talk has always entranced computer scientists. After all, voice is one of the best alternatives for eye-strain exercise that one must go through when working with lengthy documents. Also voice serves better when it comes to users who are not familiar with computers or visually handicapped personnel. Research is being done throughout the world to improve the human-computer interface and one of the promising areas is the text-to-speech (TTS) conversion. The term text-to-speech refers to the conversion of input text into spoken utterance. The input text may consist of number of words, sentences, paragraphs, numbers and even abbreviations. The TTS conversion process should identify the text without any ambiguity and generate the corresponding sound output with acceptable clarity. This means that the quality of the output of the TTS engine should be made as close to the natural speech as possible.

In general, TTS conversion can be carried out in three ways. They are Formant based, Parameter based and Concatenation based [1].

In this work, the concatenation based technique has been chosen to develop a TTS engine. Since preliminary studies have been already carried out by several researches for Sinhala language [2, 3], this work was focused on Tamil. The method of concatenation includes two phases - namely "*offline phase*" and "*online phase*". The offline phase includes the basic unit selection, identification of language rules (phonetic rules and prosodic rules) and creation of the sound database. The online phase includes splitting of input text into basic units and converting them into speech after applying the Tamil language rules. The system takes an arbitrary text file and processes the contents letter by letter and passes them through the stages of text analysis & parsing (i.e.

identification of basic units), application of language rules and finally concatenation and synthesis to produce the speech output.

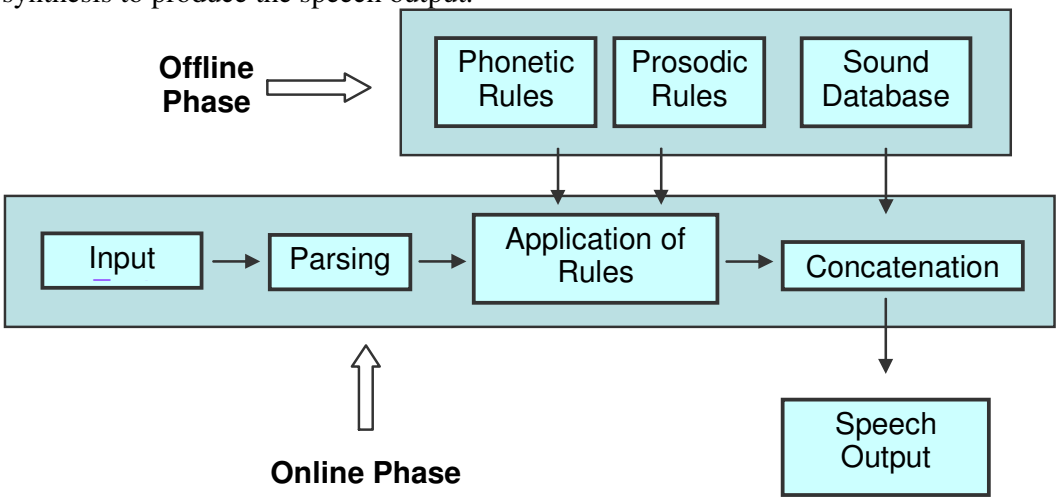

Figure 1: Block Diagram of the TTS engine

2. LANGUAGE RULES

2.1 Tamil Language

Tamil language is rich, old and it has its own heritage. The script for Tamil language was conceived since time immemorial and it has reached the present form after undergoing many changes at various periods of time.

The current Tamil writing system consists of *12-vowels*, *18-consonants*, *216-consonant vowel segments* (i.e. 18×12 = 216) and one letter called "*aaydham*". There are a total of 247 characters in Tamil.

- **Vowels:**

Table 1: 12 Vowels and their pronunciations

Vowels	Pronunciation
அ	a
ஆ	aa / A
இ	i
ஈ	ii / I
உ	u
ஊ	uu / U
எ	e
ஏ	ee / E
ஐ	ai
ஒ	o
ஓ	oo / O
ஔ	au

- **Consonants:**

Table 2: 18 Consonants and their pronunciations

Consonants	Pronunciation
க்	k
ங்	ng
ச் ஞ்	ch nj
ட்	t
ண்	N
த்	th
ந் ப்	nth p
ம்	m
ய்	y
ர்	r
ல் வ்	l v
ழ்	zh
ள்	L
ற்	tR
ன்	n

These consonants can be subdivided into 3 groups. They are,

i. வல்லினம் (reads as "vallinam"):- க்(k), ச் (ch), ட் (t), த் (th), ப் (p), ற் (tR)
ii. மெல்லினம் (reads as "mellinam"):- ங் (ng), ஞ் (nj), ண் (N), ந் (nth), ம் (m), ன் (n)
iii. இடையினம் (reads as "idayinam"):- ய் (y), ர் (r), ல் (l), வ் (v), ழ் (zh), ள் (L)

- **Aaydham:** There is one special type of characters in Tamil named "aaydham" and it is written as "ஃ" (pronounced as "h"). But this character is very rarely used in spoken and written Tamil. There are only a very few Tamil words containing this character.

- **Consonant-vowel segments:** These consonant-vowel segments are produced when any one of 18-consonants are followed by any one of 12-vowels. That is altogether there are 18×12 = 216 consonant-vowel segments available in Tamil.

2.2 Tamil phonetic rules

Most of the Tamil letters have unique sound with their script. But some of the letters have a single script, but sounds differently depending on the place they occur [4]. So the letters which have different sounds and the place of their occurrences have to be defined.

Tamil has a single script for வல்லினம் (vallinam) consonants and their consonant vowel segments. i.e. for க் (k), ச் (ch), ட் (t), த் (th), ப் (p), ற் (tR) and their consonant-vowel segments. But when these "வல்லினம்" (vallinam) consonants are followed by one of 12-vowels (i.e. for "வல்லினம்" (vallinam) consonant-vowel segments), they represent different sounds depending on the place they occur. According to these different sounds, there are three basic sounds for each syllable; namely hard, soft and normal. But for some letters, two of these sounds are the same.

- **Normal sound:** A normal sound appears when it is preceded by its own consonant and when it is occurring in the beginning of the word.

Table 5: Normal Sounds in Tamil language

Tamil letter	Pronunciation
க	ka
ச	cha
ட	ta
த	tha
ப	pa
ற	tRa

This is the same, when the other vowels follows the "வல்லினம்" (vallinam) consonants

- **Hard Sounds:** Hard sound appears when these letters are preceded by "மெல்லினம்" (mellinam) consonants.

Table 3: Hard Sounds in Tamil language

Tamil script	Pronunciation	Preceded by
க	ga	ங்
ச	cha	ஞ்
ட	da	ண்
த	dha	ந்
ப	bha	ம் / ன் / ண்
ற	dRa	ன்

This is the same, when the other vowels follow the "வல்லினம்" (vallinam) consonants.

- **Soft Sound:** Soft sound appears when these letters appear in the middle of a word, but it should not be preceded by its own consonant

Table 4: Soft Sounds in Tamil language

Tamil letter	Pronunciation
க	ha
ச	sa
ட	da
த	dha
ப	bha
ற	Ra

This is the same, when the other vowels follows the "வல்லினம்" (vallinam) consonants. .

The other consonants, i.e. "மெல்லினம்" (mellinam) and "இடையினம்" (idayinam) and their consonant-vowel segments as well as the vowels have unique sounds with the script.

Figure 2 shows an example for the above sound variation. Consider the letter "க" (ka). As mentioned above it can be pronounced in three different ways, depending on the place it occurs. The wave forms of these three different cases are shown here.

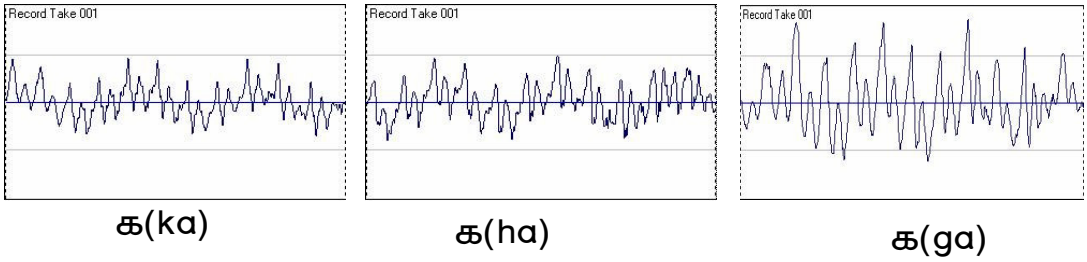

Figure 2: The wave form of "க" (ka) while occurring in different places

2.3 Tamil Prosodic rules

The prosody is a physical phonetic effect which is being employed for expression. This is used to produce natural sounding speech. Prosody consists of systematic perception and recovery of the speaker's intentions based on pauses, pitch, duration and loudness. Pitch is the most expressive part of the speech signal. People try to express their emotions through pitch variations. Therefore a detailed analysis of pitch and duration has to be made in order to get natural sounding speech. Since this is a very tedious task with many iteration, in this work only the duration of each unit was considered. Insertion of proper amount of silence when required is also an important task. Variations that should be introduced in places such as full-stop, coma, colon and semicolon were also taken into account.

3. IMPLEMENTATION

The implementation of the TTS engine, namely, the offline phase, the online phase and the development of the user interface is discussed in this section.

3.1 Offline Phase

The offline process of the system includes selection of basic units and the creation of the sound database.

- **Selection of basic unit:**

The collection of basic units must contain all the sounds/words in Tamil. This unit of speech can be syllables, phoneme, diphone, words, phrases or even sentences. Of these, the longest unit is the sentence and the shortest unit is the phoneme/diphone. If we

choose *sentence* as our basic unit, then the number of concatenations will be low. Therefore the quality will be high. But the number of units will be infinite and it will be impossible to design such a system to work with any given script.

If we choose *phoneme* as our basic unit, then the number of concatenations will be high. Therefore the quality will be low. However, the size of the database is also an important factor to be considered.

If the *syllable* is chosen as the basic unit, then to implement sound variations of each syllable when they occur in different places may not be possible; since a syllable consists of Tamil scripts only.

Diphone means the sound which is appearing in the middle of two adjacent syllables. If it is chosen as the basic unit, the concatenation distortion will be low. But the number of units that is to be store in the database is high; since it is necessary to store the sounds of every adjacent syllable.

After considering all these facts, *phoneme* has been chosen as the basic unit for this work. This is the basic sound unit of a language. In case of English, there are 26 letters, 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 with very distinct pronunciations among them. The various letters form a large list of about 343 phonemes in Tamil, which forms the basis for Tamil speech.

- **Creation of sound database**

Since phonemes were chosen as the basic unit, all phonemes had to be recorded and stored in the database. The quality of the sound database is a critical factor as far as the quality of the output of the TTS is concerned.

When recording speech, it is required to have at least 8 kHz sampling rate and 8-bit sampling. This produces poor quality speech, but it can be understood. The quality can be improved by increasing the number of bits in sampling to 12 bits or 16 bits. Increasing in digitizing levels will increase the signal to noise ratio. Increasing the sampling rate above 8 kHz, i.e. to 10 kHz, 16 kHz, or 20 kHz improves the frequency content. The higher the sampling frequency, the higher the quality. It is reasonable to have 16 kHz sampling rate for high quality speech recording and play back.

The implementation of the sound database was carried out in 4 steps.

i. **Recording sounds:** The recording of the sounds was carried out by using a native Tamil speaker. It was done through a microphone which has the frequency response of 30 – 16000 Hz in a noise free room. Recording was carried out with a sample size of 16 bits and a sampling frequency of 8 kHz. Even though the sampling frequency is small, it was good enough to produce a understandable voice output.

ii. **Extracting phonemes:** The recorded sound file was a continuous file consisting of all the units (vowels, consonants and consonant-vowel segments). The portion corresponding to each and every phoneme was extracted from this file and saved as

a separate *.wav file. This was done using the sound editing software "*Sound Forge*".

iii. **Pitch marking:** This means finding the actual start and end point of each phoneme. This is essential as the waveforms are concatenated at the pitch marks. Therefore the start and the end of each phoneme was detected by careful listening and the part in between those two was selected and saved as a separate *.wav file. Thereafter the actual duration of each phoneme was decided by "trial and error".

iv. **Editing sound files:** The sound files which were pitch marked, were not in the acceptable quality for most cases. Therefore they had to be edited to enhance the quality such that they formed a good quality output when combined. During this step, the noise reduction facilities available with the "*Sound Forge*" were also utilized.

3.2 Online Phase

The online phase includes work that is carried out in the background while the program is running. This involves three important steps, namely text analysis & parsing, application of language rules and concatenation of language units to produce TTS conversion.

Text analysis & Parsing: This phase analyzes the input text and finds single glyph letters, double glyph letters and triple glyph letters. Then it splits the input text into sequence of basic units of speech.

Application of language rule: After splitting the input text into its basic units, Tamil language rules are applied to extract the correct phoneme from the database. Language specific phonetic rules were incorporated into this phase for synthesizing a natural sounding speech.

Concatenation: During the final stage, the concatenation of the waveforms of the basic units in the correct sequence is carried out to produce the required sound waveform. The wave forms of the basic units are extracted from the database and concatenated.

3.3 User Interface

A simple graphical user interface (GUI) was developed using Visual Basic 6.0 to test the TTS engine. The interface consists of a text box to enter the text in Tamil and a button to play the speech sound. The input text can be any Tamil word or a number or a sentence. The interface also provides the user an option to save the entered text or even to open the pre-saved text to produce the TTS conversion. The Tamil font which was used in the development of the user interface was "*TSC_Avarangal*". The software which was used to enter Tamil text is called "*eKalappai 1.0*" [5].

4. CONCLUSIONS

This work shows that a TTS conversion engine that could convert any arbitrary Tamil text to spoken utterance can be developed successfully for Tamil language. The back-

end of the engine, created with 343 *phonemes* which were chosen as the basic units of Tamil language, serves as the heart of the system. The sound editing software "*Sound Forge*" was used when preparing the phonemes with correct pitch. The quality of the sound output can be further increased by using a sample speed better than 8 kHz to record the phonemes. The developed text-to-speech engine was successful in splitting the given text into the basic units, extracting the correct *phonemes*, matching with the database and concatenating the basic unit waveforms in the correct sequence to produce the required sound waveform. The front-end of the engine developed with Visual Basic is a GUI interface that allows users to input Tamil text in letters or even numbers. The interface allows users to store text to be retrieved at a later time and play back. The Tamil font "*TSC_Avarangal*" was selected as the font to develop the user interface.

In this work, emphasis was given for fine tuning the duration of the basic sound unit for proper concatenation. However, further improvements must be carried out by adjusting the pitch to make the output close to the natural speech.

The work clearly shows that for languages such as Tamil or Sinhala, text-to-speech engines can be successfully developed with the available resources.

REFERNCES

1. G.L.Jayavardhana Rama, A.G. Ramakrishnan, M.Vijay Venkatesh, R. Murali Shankar, Thirukkural, *A Text-to-Speech Synthesis System, Dept. Electrical Eng., Indian Institute of Science*, Bangalore, India (2001) www.infitt.org/ti2001/papers/jayavar.pdf
2. Asiri Nanayakkara and M.K. Jayanada, *Text to Speech conversion in Sinhala Language*, 17th National IT Conf., Sri Lanka (1998)
3. S. G. K. Savinda Gunawardana, *Text to Speech Synthesizer for Sinhala Language*, 22nd National IT Conf., Sri Lanka (2003)
4. Tamil Design Guide (2002) http://www.tdil.mit.gov.in/TamilDesignGuideOct02.pdf
5. *eKalappai 1.0* (2002) http://www.tamil.net/newtamil/ekalappai_1.html