Matrubhasha – An Integrated Speech Framework for Indian Languages

Authors: Sireesha Sarma, R.K.V.S. Raman, S.Sridevi, Rekha Thomas
CDAC (Formerly NCST) Electronics City Bangalore 560100
{sireesha, raman, shridevi, rekha}@ncb.ernet.in

Abstract

In this paper, we introduce Matrubhasha, an integrated Text to Speech Solution in Indian languages, to computer-enable rural masses. In the context of digital divide and the need to bridge the digital divide, we try to present the perception of Matrubhasha and it's features that can bridge the digital divide.

Introduction

The word "Divide" is a term that the people of India are familiar with. As a developing economy, the country has had to constantly deal with disparities between people at different levels. Whether it is the divide between the rich and the poor, or the chasm between the literate and the uneducated, India has had to grapple with issues that have divided its people into the haves and the have- nots.

More recently, the divide factor has taken a more ominous, more debilitating form. Despite the Internet's democratizing potential, it has been recognized from the outset of the digital revolution that there is a very real danger that there will be two different societies in India, the "information rich" and the "information poor." In the present era of human computer interaction, the educationally under privileged and the rural communities of India are being deprived of technologies that pervade the growing interconnected web of computers and communications. In the present era of human computer interaction, the educationally under privileged and the rural communities of India are being deprived of technologies that pervade the growing interconnected web of computers and communications. As of now, the benefits of IT are only being enjoyed by a select few Indians and not by the populace at large. The main problem behind this is the availability of information on technology and its usage in English, a language known to only about 2% of Indians. The result is digital divide that keeps the Indian masses away from reaping the benefits of advancements in the fields of Electronics, Telecom and IT.

The Digital Divide-the yawning gap between citizens who have access to and know computers and those who will never interface with the new age devices-is expected to further alienate the country's underprivileged people, particularly those residing in the country's rural hinterland. The advances we make are totally inaccessible by a large number of countrymen. The Digital Divide is a scourge that needs to be eradicated if India wants to move forward, integrate with the global village and participate in the world-wide economy.

In this information age, storage and retrieval of information in a convenient manner has gained importance. Because of the near-universal adoption of World Wide Web as a repository of information for unconstrained and wide dissemination, information is now broadly available on the Internet and is accessible from remote sites. However, the interaction

between the computer and the user is largely through keyboard and screen-oriented systems. In the current Indian context, this restricts the usage to a miniscule fraction of the population, who are both computer-literate and conversant with written English. For Indians who speak no (or little) English, the barriers to the Information Age are almost insuperable.

How do we bridge the digital divide? How do we overcome the barrier between the haves and have-nots of information access?

The main problem is that the interaction between the computer and the user is largely through keyboard and screen-oriented systems. In the current Indian context, this restricts the usage to a miniscule fraction of the population, who are both computer-literate and conversant with written English.

There are two perspectives to this problem; one is that of people literate but not computer literate or those who are not conversant with English, and second those people who are completely illiterate.

Developing Operating Systems, software and content in Indian Languages might help one fraction of the people who are literate, but even then the digital divide remains between the literates and the educationally under-privileged.

In order to enable a wider proportion of population to benefit from Information technology, there is a dire need for an interface other than keyboard and screen-interface that is widely in use at present. Speech, being a natural means of communication among human beings, can also provide a consummate platform for man-machine interaction also. It is also desirable that human-machine interface permits one's native language of communication. In the context of a multi-lingual country like India, this can be of immense value to our country where literacy rate is considerably low. It is out of debate that speech technology promises to be the next generation user interface and in the context of digital divide being addressed speech interface in ones native language is the best solution.

Conception of Matrubhasha

India is a country with more than 15 official languages and each language is spoken in a variety of dialects. Further, each dialect also may differ in accent across various geographical dimensions. The same language is spoken in different forms across different places. It would be good if the TTS system spoke the dialect in the accent that is understood by common man.

More emphasis is given to the programmers' perspective of the TTS than the linguists' perspective. Keeping aside the technical aspects, who else than a linguist can be more aware of the issues related to pronunciation and other rules involved in converting written text to spoken form? Unfortunately, the linguists can be least involved in TTS because of lack of technical detail. Hence TTS is deprived of the contribution from the linguists. This gap needs to be bridged, between the linguists and the technology for building a text to speech system that produces intelligible speech subject to the linguistic detail of the language.

The above facts were uncovered during a survey conducted on Text to Speech in Indian Languages by the Matrubhasha Team. Matrubhasha is a project carried on at CDAC (Formerly NCST) Bangalore, as a part of the digital divide bridging activities. It was undertaken with the intention of making end user applications speak and listen to the masses in any Indian language that they are comfortable to communicate in.

Indian Languages – A Brief Scenario

A small study on Indian languages revealed that Indian scripts are basically phonetic in Nature. And there is one-to-one correspondence between the written and spoken forms of most of the Indian languages. Another interesting factor is that almost all the Indian languages have a common phonetic base.

The phonetic nature, common phonetic base and one-one correspondence between written and spoken forms can be well exploited to build a phonetic based language base for each Indian language so as to build a TTS system

- that uses Rule based text to phoneme conversion with emphasis on linguistic features of the language
- that has Listen and Adjust feature to model the speech output
- that uses XML format to develop language bases so as to provide platform independent language bases.

All the above ideas ultimately lead to the conceptualisation of Matrubhasha – A Speech Solution to bridge the digital divide by making rural masses digi-savvy.

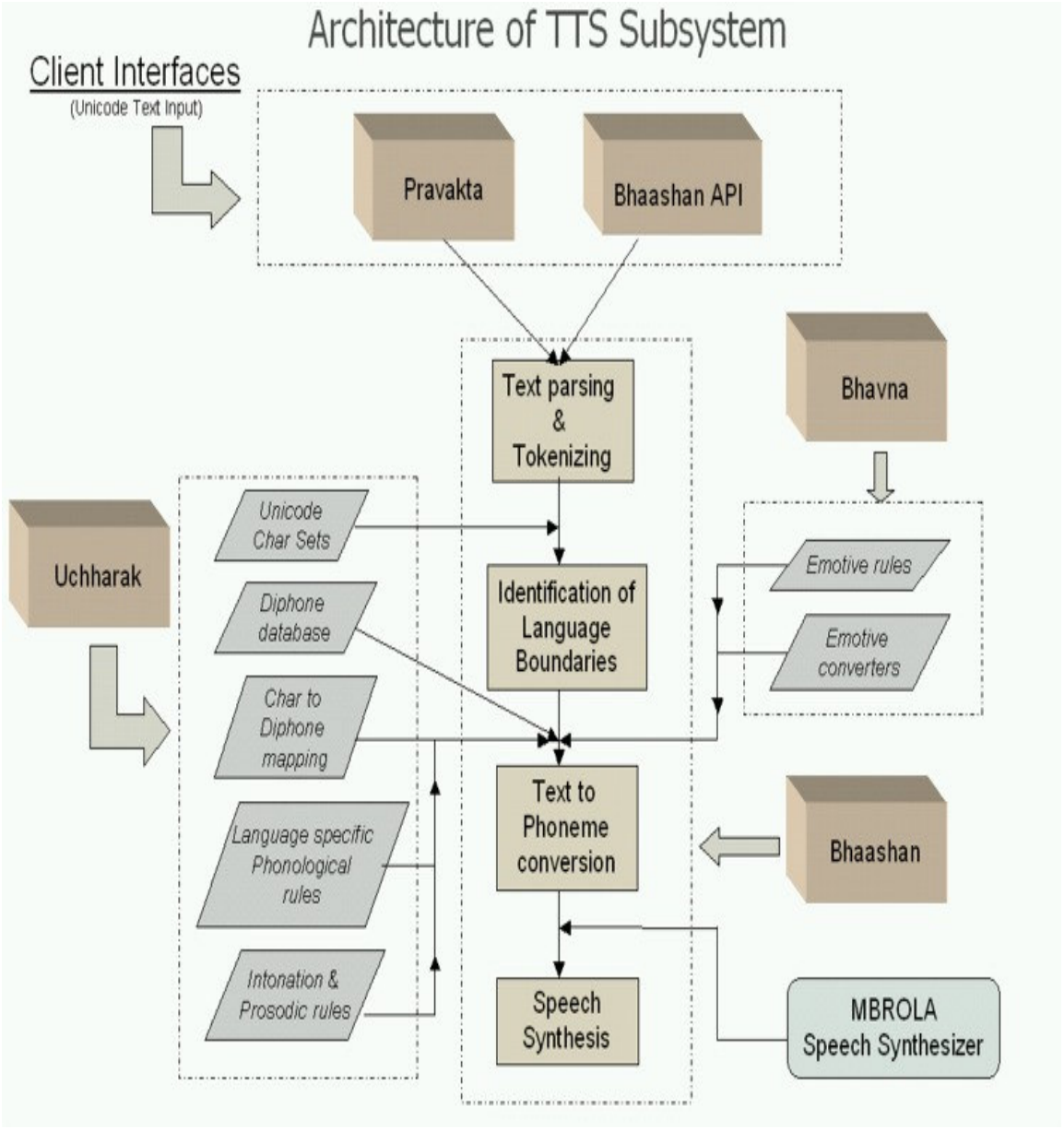
The main idea behind Matrubhasha is that,

- there should be one single engine and different language bases where by just using different language rule files, the same engine could speak different languages, and thus speak multiple languages from the same document.
- there should be tools with a linguist friendly user interface which help the linguists to create language bases for any language and any dialect and create rules in the language base so as to bring out correct pronunciation by imposing those rules during conversion of written text to spoken phonemic form.
- the main intention of Matrubhasha is not delivering one speech enabled application but to provide an application programming interface (API) that speech enables any application.
- the API can also be extended to create plug-ins to general purpose utilities like office products and internet browsers.

Matrubhasha uses Mbrola[1] as the underlying speech synthesizer for TTS. Mbrola is a speech synthesizer based on concatenation of diphones, which is developed by the TCTS lab of 'Faculte Polytechnique de mons' (Belgium) and provided free for non-commercial applications. Mbrola takes a list of phonemes as input, together with prosodic information (duration of phonemes and a piecewise linear description of pitch), and produces speech samples on 16 bits (linear), at the sampling frequency of the diphone database used. Matrubhasha aims at synthesizing speech from a given Unicode input, by converting it into a list of phonemes, together with prosodic, information in the format required by Mbrola speech synthesizer.

Matrubhasha – Architecture

The Architecture of Matrubhasha TTS framework is as follows:


Architecture of TTS Subsystem
Client Interfaces
(Unicode Text Input)
Pravakta
Bhaashan API
Text parsing & Tokenizing
Bhavna
Uchharak
Unicode Char Sets
Diphone database
Char to Diphone mapping
Language specific Phonological rules
Intonation & Prosodic rules
Identification of Language Boundaries
Emotive rules
Emotive converters
Text to Phoneme conversion
Bhaashan
Speech Synthesis
MBROLA Speech Synthesizer

Subsystems:

The sub systems of TTS framework are as follows:

Bhaashan Subsystem:

The key subsystem of Matrubhasha is Bhaashan, API for TTS, which enables the user to incorporate speech capabilities into the application software developed. All the other subsystems of Matrubhasha implement or call the functions provided in Bhaashan API for various tasks to be accomplished. Bhaashan API is available in various programming languages like C/C++, Java and COM.

Uchharak Subsystem:

Uchharak subsystem of Matrubhasha is the subsystem that is related to the rule bases of the system. Uchharak provides a robust interface to create and manage rule bases. These rule bases decide the prosody and pronunciation rules pertaining to each language. All the rule bases that are created using Uchharak are stored as XML files, that consist of details such as phonemes, letter to sound rules pertaining to each language. Each XML file consists of the following details for each language:

- The character set of that language
- The Unicode value of each character
- The glyph for each character
- The phoneme pertaining to each character as in Mbrola database
- The phonetisation rules pertaining to each character, when it is encountered alone and for various combinations of that character/character type with other characters/character types.

The phonetisation rules are used to assign the duration and pitch pairs to be assigned to each phoneme of that particular character depending upon the position of that particular character, the preceding character succeeding character etc. These rules are sometimes generic and some times they are language specific and at times they can be specific to that particular sequence of characters also. There are rules that apply in any case which are applied according to priorities assigned to them.

Uchharak also provides a speech-testing interface where in the user can test the speech output for given rules and duration, pitch pairs assigned to each character/ character combination and also provides a robust way to manage the rules, edit the duration, and pitch pairs assigned to each character.

Bhavna Subsystem:

Bhavna subsystem of Matrubhasha is the subsystem related to emotional speech synthesis. Bhavna provides a robust interface to create/Manage emotion base. An emotion base is an XML file which consists of the values of certain factors that decide the prosody of synthesized speech like, pitch contour, phrase contour, stress, average base frequency and jitter etc. The values of the above mentioned factors are decided by the system by analysing the context of the text. Bhavna makes use of the Uchharak rule base files to get the language pertinent rules and depending on those phonetisation rules, adds the emotion factors so as to synthesize emotional speech. The design of Bhavna is essentially based on Emofilt[2] – an emotional speech synthesizer developed as a part of Mbrola.

Bhavna also provides a speech-testing interface where in the user can test the speech output for given emotion base.

Anuvaachak Subsystem:

Anuvaachak is an experimental system developed to enable the linguists to develop Speech Synthesis Systems with various accents. Anuvaachak is loosely based on the work done by Mike Hamilton on Cross Language Synthesis and XLANG[3]. Anuvaachak provides a much richer development environment for linguists and language experimenters than the other 2 mentioned above. With Anuvaachak it is possible to generate Speech Synthesis Systems for Hindi or Tamil or any Indian Language with Arabic or Hebrew or any other accent for which an MBROLA diphone database is available. Anuvaachak is mainly targeted to encourage the development of Speech Accent Fonts.

Pravakta Subsystem:

Pravakta is a plug-in component. This is available in the form of COM component, which can be used in any of the applications where COM is supported. This component provides the Text To Speech functionality for the application in which it is used. Pravakta makes use of the Uchharak rule base file for speech synthesis and Bhavna emotion base file if emotional speech output is required. Pravakta is plugged in to various applications like MS Office, OpenOffice and Internet browsers so that speech can be synthesized for the content input.

The most significant feature of Matrubhasha is that it provides complete language modelling support. Any linguist or to that point any user having considerable knowledge of a language can model the language as well as its diglossia without any awareness of the slightest technical detail. Further, Matrubhasha also provides ‘Rule-Helper’ and ‘Speech-Tester’ interfaces with Uchharak, which help the user to create rules for a particular character depending upon its occurrence in a word, or the user can test the prosody for a character or a combination of characters and adjust the prosody by hearing to how it is pronounced.

Matrubhasha provides Emotion-Modelling Support, Plug-in Support for various end user applications like office suites and Internet browsers, Prosody modelling support, Dialect Modelling Support, and above all, its API model makes Matrubhasha programmer friendly

also by allowing any programmer to incorporate speech capabilities into the application developed by him/her.

Conclusion:

In the Indian scenario of 18 official languages with different dialects and accents, API incorporating speech capabilities into applications that helps one to model the different dialects is very helpful to bridge the digital divide. The rule-base approach that is chosen instead of a lexicon-based approach is surely a compromise on naturalness, but in case of Matrubhasha it is a trade-off between the robustness of the language modelling support and the human-sounding speech. Matrubhasha, we hope one good agent to bridge the digital divide.

References:

1. T. DUTOIT, V. PAGEL, N. PIERRET, F. BATAILLE, O. VAN DER VREKEN, "The MBROLA Project: Towards a Set of High-Quality Speech Synthesizers Free of Use for Non-Commercial Purposes" Proc. ICSLP'96, Philadelphia, vol. 3, pp. 1393-1396 .

2. F. Burkhardt, Simulation emotionaler Sprechweise mit Sprachsyntheseverfahren, Dissertation an der TU-Berlin, Shaker Verlag 2001

3. Mike Hamilton , Cross Language Synthesis with MBROLA