10.5 Vaachak : Text to Speech System

Text to Speech Synthesis in Vaachak 3.0 using Unit Selection process

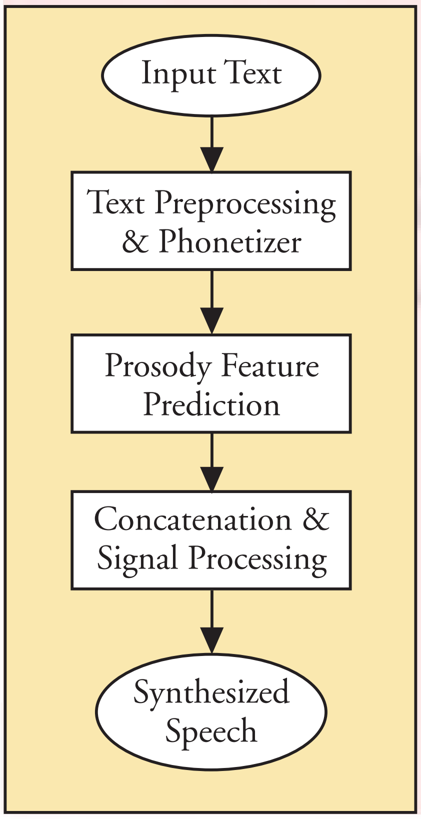
Introduction

Text to Speech (TTS) systems are built to convert given text to a speech waveform having both intelligibility and naturalness. The challenge for speech scientists and researchers the world over has been to build speech systems that come as close as possible to natural, human speech. At the same time, this objective needs to be balanced with the need to achieve this on off the shelf, commercially available, inexpensive computing power and at performance levels which allow productive utilisation in useful applications such as voice portals, screen readers, e books, etc.

Conventional TTS Technology

Conventional Concatenative TTS systems (including previous versions of Vaachak) normally have three modules, viz:

- The text processing module
- Prosody prediction module
- Concatenation signal processing module.


Input Text
Text Preprocessing & Phonetizer
Prosody Feature Prediction
Concatenation & Signal Processing
Synthesized Speech

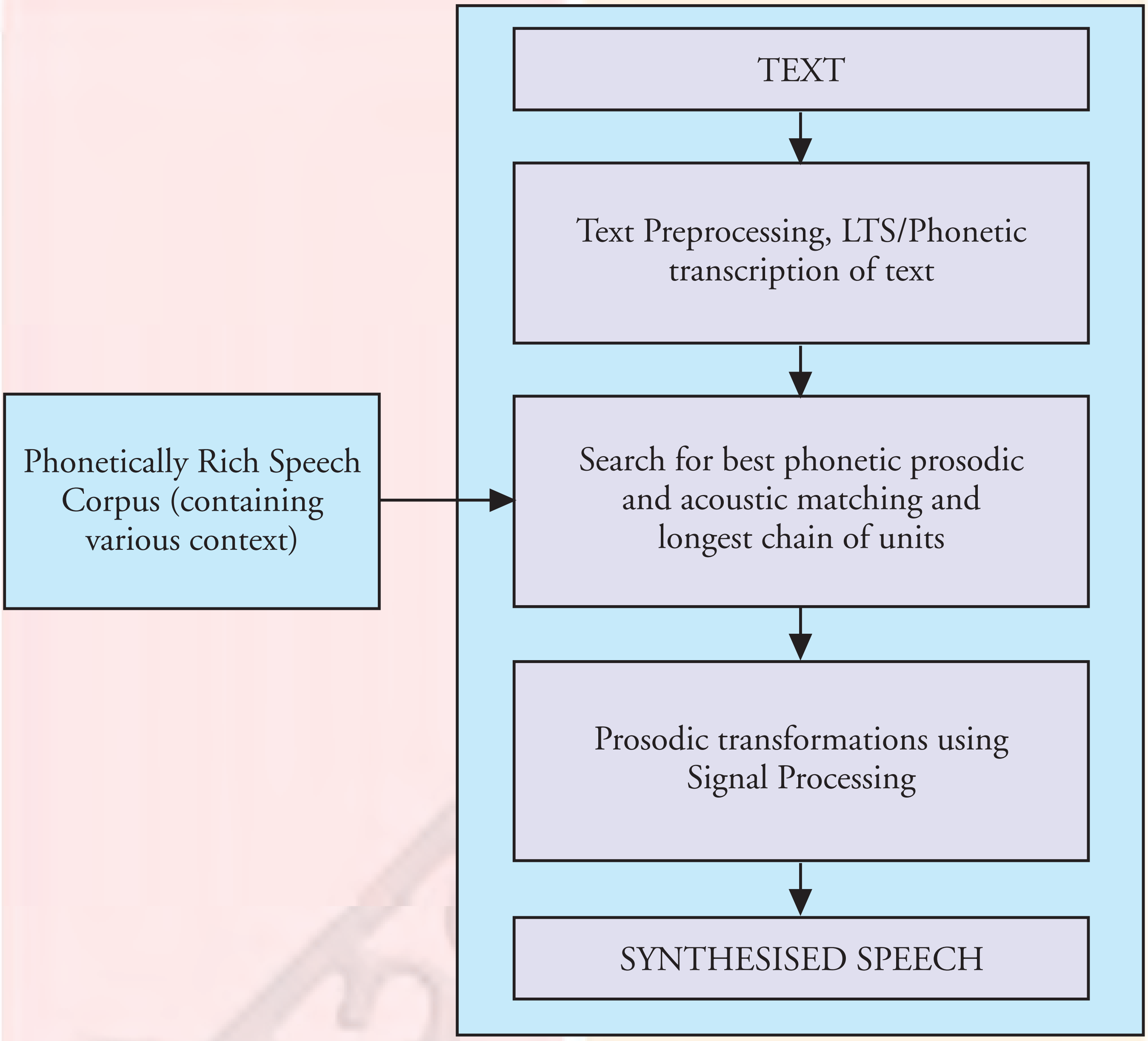
The text taken as input is first converted into a sequence of phonetic transcriptions (phonemes, diphones, demi-syllables or syllables) with high-level prosodic descriptions, such as stress, focus, breaks, etc. Then, an 'appropriate' set of prosodic contours, such as fundamental frequency, duration and amplitude, is calculated by the prosody module. At last, a pitch and duration modification algorithm, such as PSOLA, is applied to prestored units to guarantee that the prosodic features of synthetic speech meet the predicted target values. These systems have the advantages of flexibility in controlling of prosody.

The Unit Selection Approach

Modern techniques of speech synthesis (employed in Vaachak 3.0) widely employ techniques of storing multiple realizations of each unit with different prosody in a continuous speech corpus. These techniques employ various present schemes to select a proper unit from multiple instances with varied spectral features to achieve better smoothness between concatenated units.

Vaachak 3.0 uses a unit selection method that takes variations of both spectral and prosodic features into account to reduce the extent of signal processing that is required to correct the prosodic characteristics of selected instances. It makes the best decision on unit selection by minimizing the concatenated cost of a whole utterance. Since the largest available and suitable units are selected for concatenating, distortion caused by mismatches at concatenated points is minimized. Informal listening texts have indicated that the results produced by this manner produce speech of significantly higher quality, with greater listener comprehension and satisfaction. Typical Unit selection schemes are applied on speech databases which are between 1 – 2 hours in length.

This approach is based on the ultimate assumption that a very large speech corpus would be available that contains enough prosodic and spectral varieties for all synthetic units. This assumption is valid under a constraint that the whole corpus retains the same speaking style, which is referred as the "relax reading style", the same speech rate and the same timbre. Since no pitch or duration modification will be applied to the selected units before concatenation, a two-module TTS structure is adopted in our approach. It bypasses the prosody module that generates numerical prosodic features in most of the conventional TTS systems.


TEXT
Text Preprocessing, LTS/Phonetic transcription of text
Phonetically Rich Speech Corpus (containing various context)
Search for best phonetic prosodic and acoustic matching and longest chain of units
Prosodic transformations using Signal Processing
SYNTHESISED SPEECH

Conclusion

At current technology levels, the debate on which approach is necessarily the better approach for constructing a Text to Speech System still remains unresolved. To identify the 'better' TTS sytem, it is also important to qualify the primary requirements which a user may have – since each of the previous and current approaches have their own strengths and weaknesses. While a particular approach may be geared towards producing extremely high quality speech, a different approach may be highly computing power efficient. The time when (today's) high end computing capability is freely available may not be too far away, but at the moment it is a constraint that can not be taken too lightly (especially in a country such as ours). Till such time, it seems as if the debate will continue.

Vaachak 3.0 will be available for commercial use very shortly. Visit the Prologix website (www.prologixsoft.com) to experience a free, online demonstration of Vaachak 2.0 and speech samples from the upcoming Vaachak 3.0.

A brief comparative of previous approaches with each of their strengths and weaknesses is given below:

	Formant Synthesis	Diphone Concatenation	Unit Selection
Hardware Resource Requirements	Low	Medium	High
Quality of Speech	Low	Medium	High
Ease of creating multiple voices	High	Medium	Medium
Ease of modifying speech parameters (pitch, speed)	High	Medium	High
Ideal for :	Low resource platforms, handhelds, thin clients, etc. Environments where high speech quality is not a pre requisite – rarely used in commercial	Medium resource platforms (desktops/telephony), platforms such as next generation mobile phones, PDAs, digital slates, etc. Environments where information dispersal is critical – e.g. weather portals, financial updates, etc.	High resource environments such as servers and high end desktops Environments where high quality of speech is paramount and TTS readout is required for specific domains – such as directory information, weather, deployments business, news, etc.

Applications

Text to Speech Systems have found use in a wide variety of applications. Potential applications for TTS technology include:

1. Screen Readers for Visually impaired: Used to help people with visual disabilities to access text in Hindi. This application can help the visually handicapped to browse websites, read and reply to email and work as efficiently with computers as people with normal sight. Prologix is currently working closely with the National Association of Blind to develop and deploy a Hindi Screen Reader using Vaachak.
2. Information Delivery on Kiosks: TTS software can also be used to make information browsing on a kiosk convenient for people who do not have the ability to read or write. With the option of hearing the information out in a language of their choice, this enables important information to be easily delivered to each segment of our society.
3. Unified Messaging/ SMS on fixed phones: Using a TTS, it is now possible for people to access their emails over a telephone (where the text would be read out automatically by the software). In addition, TTS solutions can also be deployed to deliver SMS messages to fixed phones that do not have a display by converting them into a voice message.
4. Information delivery over telephone: Along with emails, TTS software can also be used to deliver other important information such as financial updates, railway reservation status, phone directory assistance, examination results, etc. through Interactive Voice Response Systems. Users could choose to call into these systems using a telephone and have this information converted into speech and played out to them in real time.
5. Talking Slates, Books/ Literacy aids: TTS software can be used to great effect in innovative applications such as Talking Slates or Books that can be used to help children learn to read and write.
6. Telematics: Along with other computing equipment and applications, Text to Speech systems also find a prominent application in the world of Telematics. On-board warning systems which warn the driver of possible failures in the automobile could be implemented using Text to Speech technology. Similarly guidance systems that help drivers reach a destination by guiding them on the best route to take could also be applications implemented using TTS Technology.

(For more information, contact:
Prologix Software Solutions Pvt. Ltd.
Tel. : (+91-522) 2721387, 2721382/ 3/ 4
E-mail : contact@prologixsoft.com
Website : www.prologixsoft.com)