Abstract

The focus of speech synthesis research has recently shifted from read speech to conversational and expressive styles of speech. The expressive speech synthesis has its significance in various real-life applications such as aids for visually challenged persons, talking books for children, multimedia and telecommunication applications. In this work, we are focusing on one such application of expressive style in Text-to-Speech (TTS) systems, i.e., storytelling style speech synthesis. In this speech synthesis, the goal is to generate a storytelling style speech from a TTS system which will convey story style similar to a human storyteller. Speech is a natural way of communication for human beings and they expect the synthesized speech to be very natural and expressive. There exists many problems in developing TTS system, which influences the quality of synthesis. One of the biggest problems is modeling the prosody. To synthesize a natural and story-style speech, the story TTS system must use better prosody models. These models should capture the prosodic patterns of the story-style speech narrated by a human storyteller and incorporate this information in the synthesized speech.

The main focus of the current research is to develop a high quality story TTS system in Hindi. In this regard, we have developed two types of story TTS systems: the first system consists of neutral TTS with appropriate story-specific information generation and incorporation modules and the second system consists of story speech corpus-based TTS. For improving the quality of storytelling style speech from the developed TTS systems, suitable prosody models are designed for predicting the desired story-specific prosody. The prosodic information considered in this work includes pause, duration, intonation and intensity patterns. For modeling, the pause patterns word-terminal syllable features are examined with and without discourse information. The duration, intonation and intensity information associated with storytelling style speech is predicted by exploiting the story-genre information in addition to the syllable level features. For developing the above prosody models, classification and regression trees (CART), feed-forward neural networks (FFNN), support vector machines (SVM) and extreme learning machines (ELM) are explored. The proposed prosody models are evaluated using objective measures. Moreover, these prosody models are integrated to story TTS systems, and their impact on the quality of synthesized speech is evaluated by conducting the listening tests.

The major contributions of this thesis can be summarized as follows:

- Development of story TTS using neutral TTS system in Hindi with appropriate story-specific information generation and incorporation modules. It includes design, development and integration of story-specific prosody rule-set generation and incorporation to neutral TTS.
- Development of story TTS using story speech corpus.
- Modeling of story-specific pause patterns is proposed with and without discourse information.
- Modeling of story-specific prosody (i.e., duration, intonation and intensity) are proposed based on story genre information.

Keywords: *Story TTS system, storytelling style speech synthesis, story-specific prosody rules, prosody modeling, story-semantics, discourse, story genre, classification and regression trees, feed-forward neural networks, support vector machines and extreme learning machines.*