

Chapter 2

The Feature Extraction Algorithm for the Production of Emotions in Text-to- Speech (TTS) System for an Indian Regional Language

Jagadish S Kallimani

M S Ramaiah Institute of Technology, India

V K Ananthashayana

M S Ramaiah Institute of Technology, India

Debjani Goswami

IBM Technologies, India

ABSTRACT

Text-to-speech synthesis is a complex combination of language processing, signal processing and computer science. Ubiquitous computing (ubicom) is a post-desktop model of human-computer interaction in which information processing has been thoroughly integrated into everyday objects and activities. Speech synthesis is the generation of synthesized speech from text. This chapter deals with the development of a Text to Speech (TTS) Synthesis system for an Indian regional language by considering Bengali as the language. This chapter highlights various methods which may be used for speech synthesis and also it provides an overview on the problems and difficulties in Bengali text to speech conversion. Variations in the prosody (speech parameters – volume, pitch, intonation, amplitude) of the speech yields the emotional aspects (anger, happy, normal), which are applied to our developed TTS system.

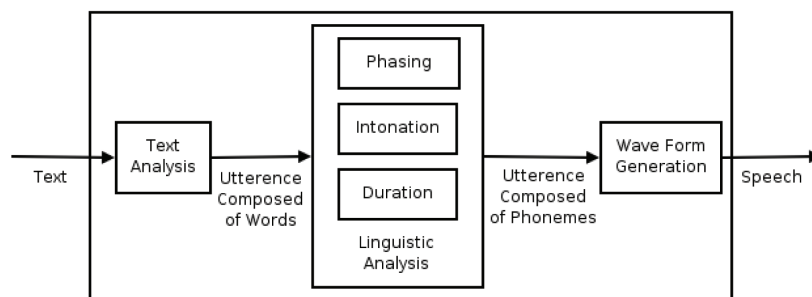
INTRODUCTION

Speech Synthesis is the process of converting input data into spoken language. The goal of a Text to

Speech synthesis system is to convert any computer readable text into a human sounding synthetic speech. A text-to-speech system is composed of two parts: a front-end and a back-end. The front-end has two major tasks. First, it converts raw text contain-

DOI: 10.4018/978-1-60566-996-0.ch003

Figure 1. Block diagram of text-to-speech synthesis



ing symbols like numbers and abbreviations into the equivalent written-out words. This process is called text normalization, pre-processing, or tokenization. The front-end then assigns phonetic transcriptions to each word, and divides and marks the text into prosodic units, like phrases, clauses, and sentences. The process of assigning phonetic transcriptions to words is called text-to-phoneme or grapheme-to-phoneme conversion. Phonetic transcriptions and prosody information together make up the symbolic linguistic representation into an audible sound. The back-end often referred to as the synthesizer—then converts the symbolic linguistic representation into an audible sound. Diagrammatically a typical TTS system can be represented as in Figure 1.

There are several approaches which may be used for the development of TTS. They are concatenate Synthesis, Formant synthesis and Pre-recorded Synthesis systems. In Concatenate synthesis, the text is phonetically represented by the combination of its syllables. These syllables are concatenated at run time and they produce phonetic representation of text. Here, the vocabulary is unlimited and voice quality is good, but cannot produce multiple featured voices and also needs large storage space. Formant synthesis is based on manipulating formants. Formants are the distinguishable frequency components of human speech. Here the voice is generated by simulation of the behavior of human vocal cord, vocabulary is unlimited, storage space is

low and can produce multiple featured voices but the voice is robotic, which is not appreciable by the users. In pre-recorded, a database of pre recorded words is maintained. The voice quality obtained here is good but the vocabulary is limited and demands large storage. The Concatenative synthesis may follow several methodologies (Epoch Synchronous Non-Overlapping Add (ESNOLA) Approach Concatenative Text to Speech Synthesis - A Technical Report, 2005). They are Time Domain Pitch Synchronous OverLap Add (TDPSOLA), Pitch Synchronous Overlap Add (PSOLA), Multi-Band Re-synthesis OverLap Add (MBROLA), and Epoch Synchronous Non Overlapping Add (ESNOLA).

The remaining chapter is arranged as follows: Section 2 presents a brief background on available TTS systems for various languages. The method follows, complexities involved in Bengali TTS system, steps and algorithm adopted are available in Section 3. The procedure for recording the speech and interface developed are also discussed here. Prosodic analysis is discussed in section 4. Conclusion and future works are suggested in Section 5.

BACKGROUND

In the recent years, several works has been carried out in the field of Text to Speech synthesis. The history of speech synthesis dates back to 1939,

11 more pages are available in the full version of this document, which may be purchased using the "Add to Cart" button on the publisher's webpage:

www.igi-global.com/chapter/feature-extraction-algorithm-production-emotions/41094

Related Content

Optimal Advertisement Spending in a Duopoly with Incomplete Information

Luis E. Castro and Nazrul I. Shaikh (2018). *International Journal of Business Analytics* (pp. 1-21).

www.irma-international.org/article/optimal-advertisement-spending-in-a-duopoly-with-incomplete-information/205640

Should Investors Trust Equity Analysts?

Xiaomin Guo (2015). *International Journal of Business Analytics* (pp. 45-61).

www.irma-international.org/article/should-investors-trust-equity-analysts/126245

Digitalization in Accounting Through Changing Technology and Accounting Engineering as an Adaptation Proposal

Azize Esmeray and Murat Esmeray (2020). *Handbook of Research on Strategic Fit and Design in Business Ecosystems* (pp. 354-376).

www.irma-international.org/chapter/digitalization-in-accounting-through-changing-technology-and-accounting-engineering-as-an-adaptation-proposal/235581

Using Artificial Neural Networks to Forecast Market Response

Leonard J. Parsons and Ashutosh Dixit (2004). *Neural Networks in Business Forecasting* (pp. 23-46).

www.irma-international.org/chapter/using-artificial-neural-networks-forecast/27243

The Roles of Knowledge Management and Organizational Innovation in Global Business

Kijpokin Kasemsap (2016). *Business Intelligence: Concepts, Methodologies, Tools, and Applications* (pp. 1759-1783).

www.irma-international.org/chapter/the-roles-of-knowledge-management-and-organizational-innovation-in-global-business/142700