



CSXX0247: Text-to-Speech Systems for Indian Languages

L-T-P-Cr: 2-0-2-3

Pre-requisites: Formally, none, however it will be helpful to have some experience with computational linguistics, signal processing, or speech science.

Objectives/Overview:

Any system that depends on natural human-machine communication must have speech synthesis. In order to learn more about speech generation and its acoustic characteristics, phonetic researchers also employ speech synthesis systems. An introduction to text-to-speech (TTS) synthesis systems and techniques is provided in this advanced course. A variety of speech synthesis techniques are discussed, including formant synthesis, concatenative synthesis, statistical (HMM and DNN based) synthesis, state-of-the-art unit selection, and more modern end-to-end techniques. Additionally discussed are the natural language processing and linguistic text analysis components commonly found in TTS systems.

Course Outcomes (COs):

At the end of the course, a student will be able to learn:

Sr. no.	CO
CO1	History, methods and TTS system components.
CO2	Various approaches to speech synthesis.
CO3	Various foundation models of TTS for Indian languages.
CO4	HMMs for Acoustic Modelling and Linguistic text analysis.
CO5	Parametrical, Neural Network (NN), and Deep Neural Network (DNN) based synthesis.
CO6	Applications of TTS systems in various Indian languages.

CO-PO mapping:

Course Outcomes	Program Outcomes											
	PO-1 (Engineering knowledge)	PO-2 (Problem analysis)	PO-3 (Design/development of solutions)	PO-4 (Conduct investigations of complex problems)	PO-5 (Modern tool usage)	PO-6 (The engineer and society)	PO-7 (Environment and sustainability)	PO-8 (Ethics)	PO-9 (Individual and team work)	PO-10 (Communication)	PO-11 (Project management and finance)	PO-12 (Life-long learning)
CO1	✓	✓	✓	✓	-	-	-	-	-	-	-	-
CO2	✓	✓	✓	✓	✓	-	-	-	-	-	-	-
CO3	✓	✓	✓	✓	✓	-	-	-	-	-	-	-
CO4	✓	✓	✓	✓	✓	-	-	-	-	-	-	-
CO5	✓	✓	✓	✓	✓	-	-	-	-	-	-	-
CO6	✓	✓	✓	✓	✓	✓						

Unit I:

Lectures: 4

hours

Introduction: history, methods; TTS system components; Acoustic Phonetics.

Unit II:

Lectures: 5

hours

Formant synthesis, Diphone synthesis, Unit Selection synthesis.

Unit III:

Lectures: 6

hours

Foundation models for Indian languages: Duration modeling, Intonation modeling, Punctuation Model (Lexicon) and language modelling (N-Gram Language models).

Unit IV:

Lectures: 6

hours

HMMs for Acoustic Modelling - Monophone, Triphone Speech Synthesis: unit selection, statistical parametric synthesis (HTS); Linguistic text analysis.

Unit V:

Lectures: 4

hours

Parametrical and Neural Network (NN) based synthesis, NN & DNN for Acoustic Modelling - Hybrid modelling- Hybrid-NN, TTS system evaluation.

Unit VI:

Lectures: 3

hours

Applications of TTS systems in Indian languages.

Text book:

1. Taylor, Paul: Text-to-Speech Synthesis. Cambridge University Press, 2009.