Advanced Topics in Speech Processing
CS60116

K Sreenivasa Rao
Professor
Department of Computer Science and Engineering
IIT Kharagpur

https://cse.iitkgp.ac.in/~ksrao/cou-atsp-2025

Introduction

- Objective
- What is speech?
- What is speech processing?
- Applications of speech processing
- How speech processing applications are developed?
- Scope of the course
- Assignments – Hands-on
- Text books & References

Objective

- Gross-level understanding of how machines process speech
- Contribution of multiple disciplines in speech processing
  - Computer Science
  - Electronics & Communications
  - Electrical Engineering
  - Mechanical & Industrial Engg
  - Healthcare
  - Financial & Banking
  - Audio & Music
  - Linguistics & Phonetics

What is Speech?

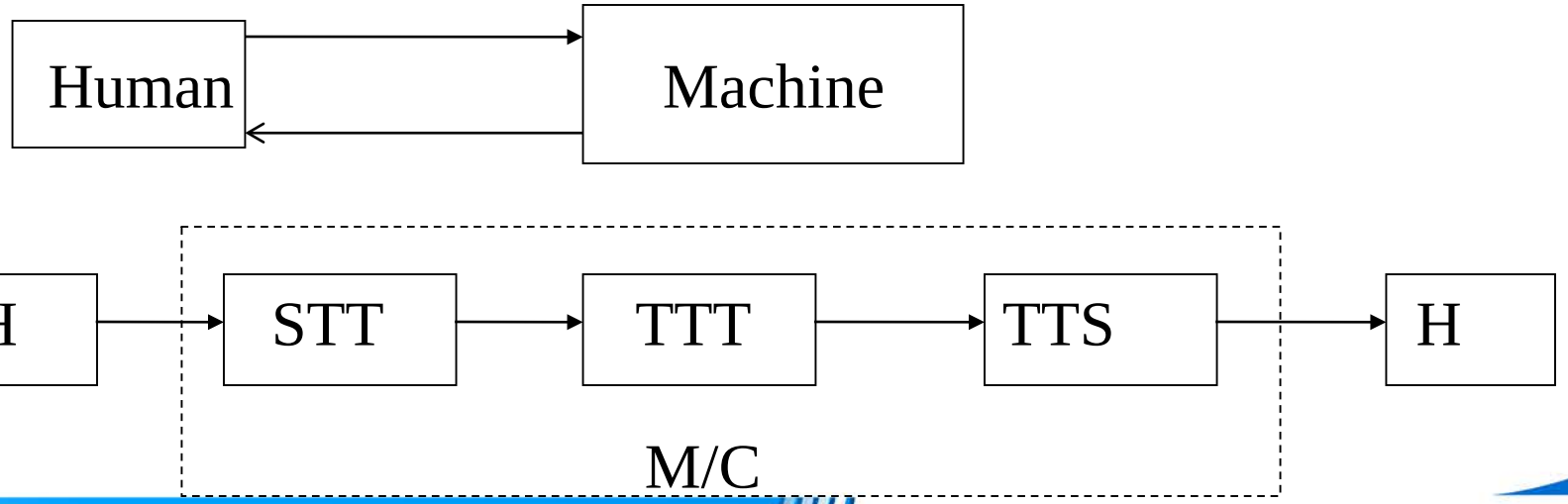
- Produced from human’s vocal apparatus (Mouth)
- Natural mode of communication among humans
- Best communication aid for humans

Need for speech processing

Speech : Natural mode of communication among human beings
Message, Speaker information and Language information

Language constraints : Legal sound units
Legal sequence of legal sound units

Why speech processing ?



Speech Processing : Applications

- Speech Recognition
  - ✓ Speaking interface with machines
  - ✓ Automatic dictation system
  - ✓ Healthcare
- Speech Synthesis
  - ✓ Speaking interface with machines
  - ✓ Voice response system
  - ✓ Screen readers, reading story books, etc..
  - ✓ Aid for visually challenged people
- Speaker Recognition/Speaker Verification
  - ✓ Voice based person authentication system
  - ✓ Forensic investigation application

Speech Processing : Applications

- Language identification
- Information (data) retrieval (voice-based)
- Speech enhancement
- Pathological Speech Analysis : Detection & Classification of Voice Disorders
- Speech coding
- Paralinguistic analysis
  - ✓ Emotion recognition
  - ✓ Speaker style modeling (speaking rate, pronunciation, etc..)

How to Develop Speech Applications?

- Speech Recognition : Features to characterize sound units
- Speech Synthesis : Parameters of Vocal-tract and Excitation
- Speaker Recognition : Features to characterize speaker
- Language Identification : Features to characterize language
- Information (data) Retrieval (voice-based) : Pattern discovery
- Speech Coding : Features for efficient coding and reproduction
- Speech Enhancement : Features to characterize speech, non-speech, noise, reverberation, etc...

Scope of the Course

- Introduction & Acoustic Phonetics
- Speech Signal Processing & Analysis Methods
- Speech Modelling Techniques
- Speech Systems

Scope of the Course (in detail)

Module	Topics
1	Introduction
2	Acoustic Phonetics
3	Representation and processing of speech
4	Speech analysis using time and frequency domain techniques
5	Linear prediction analysis of speech
6	Vector quantization and spectral distortion measures
7	Statistical pattern recognition
8	Probability density estimation & Gaussian mixture models
9	Hidden Markov models
10	MLP, SVM & DL
11	Speech coding & Speech recognition
12	Speech synthesis, speaker/language recognition and speech enhancement

Assignments : Hands-on

1. Familiarity with speech recording, playback and editing software
2. Effect of sampling and quantization
3. Recording and analysis of speech sounds
4. Time domain analysis of speech
5. Spectral analysis of speech using STFT
6. Spectral analysis using different windows
7. Sinusoidal analysis/synthesis of speech
8. Linear predication analysis/synthesis of speech
9. Cepstral analysis of speech
10. Estimation of pitch and formants from speech
11. Synthesis of vowels

Assignments : Hands-on

12. Voiced/unvoiced (speech/non-speech) detection (Energy, ZCR, ACF, AMDF)
13. Vowel recognition using Filter-bank approach
14. Vowel recognition using ZCR (realization of filter-bank using ZCR)
15. Vowel recognition using: VQ, GMM, HMM, KNN, ANN, SVM, DNN, CNN
16. Prototype speech recognizer
17. Speaker recognition
18. Language Identification
19. Text-to-Speech synthesis
20. Speech Enhancement (Noise subtraction)

Text Books

1. L. R. Rabiner and B. H. Juang, *"Fundamentals of speech recognition",* Pearson Education, LPE, New Delhi, 2003.
2. L. R. Rabiner and R. W. Schafer, *"Digital processing of speech signals",* Pearson Education, LPE, New Delhi, 2005.
3. D O'Shaughnessy, *"Speech communication: Human and Machine",* Second Edition, IEEE Press, NY, USA, 2012.
4. J. R. Deller, Jr., J.H.L. Hansen and J.G. Praokis, *"Discrete-time procesing of speech signals"* IEEE Press, NY, USA, 1999.
5. T. F. Quateri, *"Discrete-time speech signal processing: Principles and practice",* Pear Education, LPE, New Delhi, 2004.
6. B. Gold and N. Morgan, *Speech and Audio Signal Processing*, Wiley Student Edition, Singapore, 2004.
7. J. Benesty, M. M. Sondhi and Y. Huang, *"Springer Handbook on Speech Processing* Springer publishers, 2008.
8. X. Huang, A. Acero and H. W. Hon, *"Spoken Language Processing",* PHI, 2001.

Course Evaluation Details

- Mid-Sem: 30%
- End-Sem: 50%
- Class Test - I: 10%
- Class Test - II: 10%

References

1.IEEE Trans. Audio, Speech and Language Processing

2.Speech Communication (Elsivier)

3.Computer Speech and Language (Elsivier)

4.Journal of acoustical society of America (JASA)

5.IEEE Int. Conf. Acoust., Speech, Signal Processing (ICASSP)

6.Int. Conf Speech Processing (INTERSPEECH)