Advanced Topics in Speech Processing (IT60116)

K Sreenivasa Rao

Dept of CSE

IIT Kharagpur

Objectives of the course

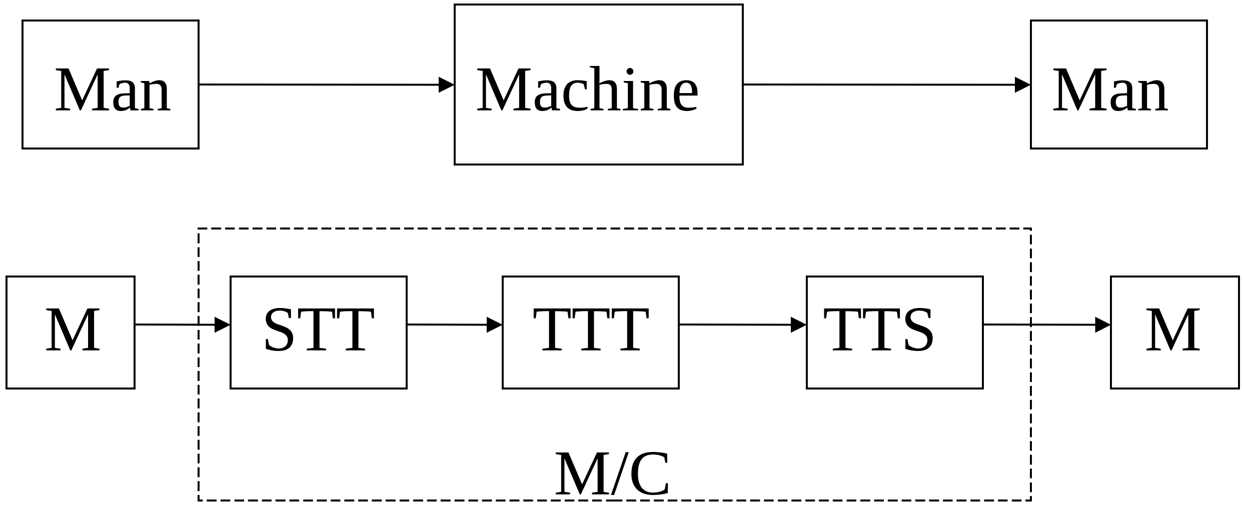
- Illustrating usefulness of signal processing tools for the analysis and processing of speech.
- Highlighting efficient utilization of communication resources by exploiting the speech production and perception properties.
- Explaining various issues involved in the processing of speech for human-computer interaction.
- Describing existing speech processing systems like speech recognition, speaker recognition and text-to-speech synthesis.
- Giving exposure to the research issues involved in the speech processing area.

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 ?



Need for human-machine interface

- Automatic dictation
- Voice response systems
- Voice based person authentication system
- Forensic investigation application
- Language identification
- Information (data) retrieval (voice-based)
- Speech-to-speech conversion applications
- Speech enhancement
- Speech coding

Speech tasks required for M/C interface

Speech Recognition : Speech-to-Text

Speech Synthesis (TTS) : Text-to-Speech

Speaker Recognition : Speech – Speaker identity

Language Identification : Speech – Language identity

Speech Enhancement : Noisy speech – Clean speech

Speech coding : Speech – Encoder – Channel – Decoder – Speech

Features for various speech tasks

- Features to characterize sound units
- Features to characterize a speaker
- Features to characterize the language
- Features to represent the articulator movements
- Features to characterize speech, nonspeech, noise and reverberation
- Features for coding and reproduction of speech

Scope of the Course

- Introduction (1)
- Acoustic Phonetics (3)
- Speech signal processing methods (4)
- Speech signal analysis approaches (8)
- Modeling techniques for developing speech systems (12)
- Speech systems (12)

Scope of the Course (Cont..)

- Introduction

- Acoustic phonetics
  - Classification of sound units
  - Production aspects
  - Time and frequency domain realizations
  - Excitation and vocal tract system characteristics

- Processing of speech signals
  - Spectral domain representation
  - Source-system representation
  - DFT and its properties
  - Pole-zero realizations

Scope of the Course (Cont..)

- Speech analysis methods
  - Filterbank analysis
  - Time-domain features of speech signal
  - Linear Prediction (LP) analysis of speech
  - Cepstrum analysis
  - Sinusoidal analysis of speech
  - Harmonic plus Noise model (HNM) analysis of speech
  - Group-delay analysis of speech

Scope of the Course (Cont..)

- Models used for developing speech systems
  - Introduction to statistical pattern recognition
    - Classification strategies
    - Probability density estimation techniques
  - Vector quantization (VQ)
  - Gaussian Mixture Model (GMM)
  - Hidden Markov Model (HMM)
  - Neural Networks (NN)
  - Support Vector Machines (SVM)

Scope of the Course (Cont..)

- Speech systems
  - Speech coding
  - Speech recognition
  - Speaker recognition
  - Speech synthesis
  - Language recognition
  - Speech enhancement

Assignments

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
12. Development of prototype speech recognition system

Assignments (Cont..)

13. Voiced/unvoiced (speech/nonspeech) detection (Energy, ZCR, ACF, AMDF)
14. Vowel recognition using Filter-bank approach
15. Vowel recognition using ZCR (realization of filter-bank using ZCR)
16. Vowel recognition using:
    - vector quantization
    - GMM
    - HMM
    - Multivariate Gaussian distribution
    - K-Nearest Neighbour
    - ANN
    - SVM

Assignments (Cont..)

17. Prototype speech recognizer
18. Speaker recognition
19. Language Identification
20. Text-to-Speech synthesis
21. Speech Enhancement (Noise subtraction)

Text books

1. L. R. Rabiner and R. W. Schafer, *“Digital processing of speech signals”,* Pearson Education, LPE, New Delhi, 2005.
2. L. R. Rabiner and B. H. Juang, *“Fundamentals of speech recognition”,* Pearson Education, LPE, New Delhi, 2003.
3. D O’Shaughnessy, *“Speech communication: Human and Machine”,* Second Edition, IEEE Press, NY, USA, 1999.
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”,* Pearson 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”,* Printice-Hall, Inc., 2001

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)

Course Evaluation Details

- Mid-Sem: 30%
- End-Sem: 50%
- Assignment-I: 10%
- Assignment-II: 10%
- Attendance: 10%
- Best two of (Assig1, Assig2, Attend)