CMP4205 Audio and Speech Signal Processing

Period per Week			Contact Hour per Semester	Weighted Total Mark	Weighted Exam Mark	Weighted Continuous Assessment Mark	Credit Units
LH	PH	TH	CH	WTM	WEM	WCM	CU
30	30	00	45	100	60	40	3

Rationale

Speech processing has been one of the main application areas of digital signal processing for several decades now, and as new technologies like voice over IP, automated call centers, voice browsing and biometrics find commercial markets, speech seems set to drive a range of new digital signal processing techniques for some time to come. This course provides not only the technical details of ubiquitous techniques like linear predictive coding, Mel frequency cepstral coefficients, Gaussian mixture models and hidden Markov models, but the rationale behind their application to speech and an understanding of speech as a signal.

Objectives

- To provide students with the knowledge of basic characteristics of speech signal in relation to production and hearing of speech by humans.
- To describe basic algorithms of speech analysis common to many applications.
- To give an overview of applications (recognition, synthesis, coding) and to inform about practical aspects of speech algorithms implementation.
- To give the student practical experience with the implementation of several components of speech processing systems.

Course Content

1. ***Introduction to Digital processing of speech signals***
   - Recording; sampling, quantization
   - Speech spectra; continuous
   - Fourier transform; what do we get when we sample
   - Random signals, power spectral density
   - Modification of speech ; linear filters
   - Frequency response of a filter
2. ***Pre-processing of speech:***
   - Dc removal, preemphasis, frames, basic parameters.
   - Spectrogram.
   - Speech production; articulatory organs - vocal cords and vocal tract vs. excitation and filter.
   - Characteristics in time and frequency,
   - Influence of excitation and filter
   - What can be seen on long- and short-term spectrograms.
   - How to separate excitation and filter; cepstrum, MFCC.
3. ***Linear-predictive model:***

- Separation of vocal tract characteristics from excitation - applications in coding and recognition
- Prediction of a sample from past samples - linear prediction (LP)
- Error of LP; Obtaining the error using a single filter
- Determination of vocal tract characteristics using LP analysis
- Spectrum estimated by LP
- Features derived from LP - LAR and LSF
- LPC-cepstrum

4. Determination of fundamental frequency (F0)

- Terminology
- Characteristics of F0 for males, females and children
- Use in speech processing systems
- Methods based on autocorrelation function
- NCCF. Long-term predictor and cepstral analysis for F0 determination
- Reliability and problems of F0 detectors

5. *Coding I:*

- Aims of coding
- Bit-rate, objective and subjective measurements of quality
- Classification of coders according to bit-rate
- Waveform coders.
- Vocoders - LPC.
- Vector quantization in speech coding

6. Coding II.

- CELP, Coding in GSM networks: GSM, GSM-EFR, GSM-HR, Voice over IP
- Introduction to speech recognition - the task, classification of recognizers: isolated words - connected words - continuous speech, speaker dependent - speaker independent.
- Basic function blocks.
- Voice activity detection (VAD) for isolated words.

7. Recognition using DTW.

- Recognition based on distance of speech frames - various definitions of distance. \
- Timing: linear modification, dynamic programming (Dynamic Time Warping DTW).

8. *Hidden Markov models (HMM I.)*:

- Introduction, motivations and relation to DTW
- Structure of the model
- Gaussian distributions
- State sequences
- Probability of a sequence of states, Baum-Welch and Viterbi probabilities
- Training of models: Baum-Welch, recognition: Viterbi
- Token passing
- Connected words
- Continuous speech with large vocabulary: recognition of small units - phonemes
- Phonetics: vowels and consonants, characteristics, classification of phonemes

- International phoneme alphabets: IPA, SAMPA, TIMIT
- Co-articulation
- Applications in recognition: context-dependent triphones
- Large vocabulary, Language modeling, lattice rescoring, forced alignment

9. Features for recognition

- Suppression of pitch, de-correlation, link with spectral envelope
- LPCC, MFCC, de-correlation: PCA, LDA, HLDA, channel robustness: normalization. - delta, delta-delta
- TRAPs a FeatureNet, neural nets
- Tools for speech processing

10. *Speech synthesis*

- Structure of the synthesizer
- Conversion of written text to speech: text-to-speech
- Text normalization
- Prosody (melody, accents, timing) in synthesis
- Units for synthesis - manual and automatic selection, corpus-based synthesis
- Generation of signal in time and frequency domains: PSOLA and HNM. Applications, SW for synthesis: EPOS, MBROLA, Festival

11. *Further topics in speech processing*

- Speaker identification/verification (principles, false acceptation, false rejection, cost function, optimal operation point, EER)
- Phoneme recognition
- LVCSR
- Recognizer merging
- Very Low Bit Rate coding
- audio-video recognition
- Speech databases

Learning Outcomes

On completing this course the student should be able to:

- Describe the key aspects of typical speech signals
- Express the speech signal in terms of its time domain and frequency domain representations and the different ways in which it can be modelled;
- Derive expressions for simple features used in speech classification applications;
- Explain the operation of example algorithms covered in lectures, and discuss the effects of varying parameter values within these;
- Synthesize block diagrams for speech applications, explain the purpose of the various blocks, and describe in detail algorithms that could be used to implement them;
- Implement selected components of speech processing systems, including speech recognition and speaker recognition, using a modeling software package
- Deduce the behavior of previously unseen speech processing systems and hypothesize about their merits.

Recommended and Reference Books

[1] Deller, J. R., Proakis, J. G. and Hansen, J. H. L, 1993, *Discrete-Time Processing of Speech Signals*, Macmillan, Toronto

[2] Quatieri, T. F, 2002, *Discrete-Time Speech Signal Processing*, Prentice-Hall, New Jersey

[3] O'Shaughnessy, D, 1987, *Speech Communication: Human and Machine*, Addison-Wesley, Reading, MA

[4] Rabiner, L. R., and Juang, B.-H, 1993, *Fundamentals of Speech Recognition*, Prentice-Hall, New Jersey.

[5] Rabiner, L. R., and Schafer, R. W, 1978, *Digital processing of speech signals*, Prentice-Hall, New Jersey

[6] Huang, A. Acero, H. Hon, and R. Reddy, 2001, *Spoken Language Processing: A Guide to Theory, Algorithm and System Development*, Prentice- Hall