Course Syllabus

- Digital Signal Processing
- Human Speech Production
- Auditory Models and Speech Perception
- Time-Domain and Frequency-Domain Methods
- The Cepstrum
- Linear Predictive Analysis
- Estimation of Speech Parameters
- Speech Recognition

History of Speech Processing

- as old as the study of human languages, as new as the latest chip
- (Bell) methods for efficient and effective ways to communicate speech signals over the telephone
- closely related to DSP since 1960
- depends on the development of IC technology, DSP algorithm, and computer architecture (SoC)
- related to image processing, video processing, radar and sonar, medical diagnosis, consumer electronics

Speech Chain

- Speech is basically used for the transmission of messages.
- speech chain: message formulation, speech generation, speech transmission, speech recognition, and message understanding
- The brain controls the time-varying vocal tract shapes for producing the intended sound sequences
- information rates
  - text: $\sim$ 50 bps
  - phone sequence and prosody: $\sim$ 200 bps
  - acoustic waveform: $\sim 64k$ bps

Speech Perception Model

- First, speech waveform is converted to a spectral representation. The basilar membrane in the inner ear acts as a non-uniform spectral analyzer.
- The spectral features are transduced (by auditory nerves) into a set of sound features.
- Sequence of sound features are converted to phonemes, words, and sentences by a language translation process.
- The conversion to the meaning or understanding.

Speech Signal

- By speech signal we often refer to the acoustic waveform.
- There are additive noise and channel distortion.
- Digital speech processing begins in the acoustic waveform domain.

Speech Stack

- fundamental science and technology
  - DSP theory, acoustics, linguitics, physiology, psychology, etc.
- signal representations
  - samples, (short-time) spectrum, cepstrum, linear prediction, etc.
- processing algorithms
  - speech detection, pitch detection, formant analysis, etc.
- applications
  - coding, synthesis, recognition, conversion, etc.

Digital Signal Processing

- digital signals and systems
- convolution
- linear constant-coefficient difference equation
- frequency-domain representation
- sampling
- ideal A/D and D/A converters

Human Speech Production

- human vocal tract
- phonemes (sounds) of American English
- linguistic units: syllables, words, sentences
- phonetic transcription
- co-articulation (within-word and cross-word)
- sound classes (vowels, consonants, diphthongs, semi-vowels, etc.)
- articulation features (places and manners)

Auditory Models and Speech Perception

- speech-chain from production to perception
- anatomy of the ear
- basilar membrane
- critical bands
- sound perception
  - sound intensity and loudness
  - fundamental frequency and pitch
  - spectral and temporal masking effects
- auditory models
- speech perception experiments
- measurement of quality and intelligibility

Time-Domain and Frequency-Domain Methods

Short-Time Analysis

- energy
- zero-crossing
- auto-correlation function

Frequency-Domain Methods

- discrete-time Fourier transform
- discrete Fourier transform
- short-time Fourier transform
- spectrogram (spectrographic displays)

Cepstrum

- the inverse DTFT of the logarithm of the magnitude of the DTFT, of a signal

$$c[n] = \frac{1}{2\pi}\int_{-\pi}^{\pi} \log|X(e^{j\omega})|e^{j\omega n}d\omega, \quad (1)$$

where

$$X(e^{j\omega}) = \sum_{n=-\infty}^{\infty} x[n]e^{-j\omega n} \quad (2)$$

- short-time cepstrum
- hormomorphic systems and filtering
- cesptral analysis of all-pole models

Linear Predictive Analysis

- simplified model of speech production
- mathematical formulation
- the auto-correlation method
- the covariance method
- frequency-domain interpretation
- solution of the LPC equation
- Levinson-Durbin algorithm
- alternative representations

Speech Parameter Estimation

- speech background/silence discrimination
- voiced/unvoiced/silence detection
- pitch period estimation
- formant estimation

Automatic Speech Recognition

- challenges
- overall recognition process
- ASR formulation
- recognition models
- training algorithm
- search algorithm
- performance evaluation
- research directions

A Robot (from Gold and Morgan)

- Robot: *Tell me a task, and I will do it for $5 an hour.*
- Alfred: *Sounds great. Can you paint?*
- Robot: *My painting is of the highest quality.*
- Alfred: *See that paint brush and bucket of paint? Take them out and paint the porch.*
- Robot: *Your request will be fulfilled.* (An hour later) *The task is completed. Please deposit $5.*
- Alfred: *Good deal! Come back again!*
- Robot: *(While leaving) Oh, by the way, it wasn't a Porsche. It was a Honda.*