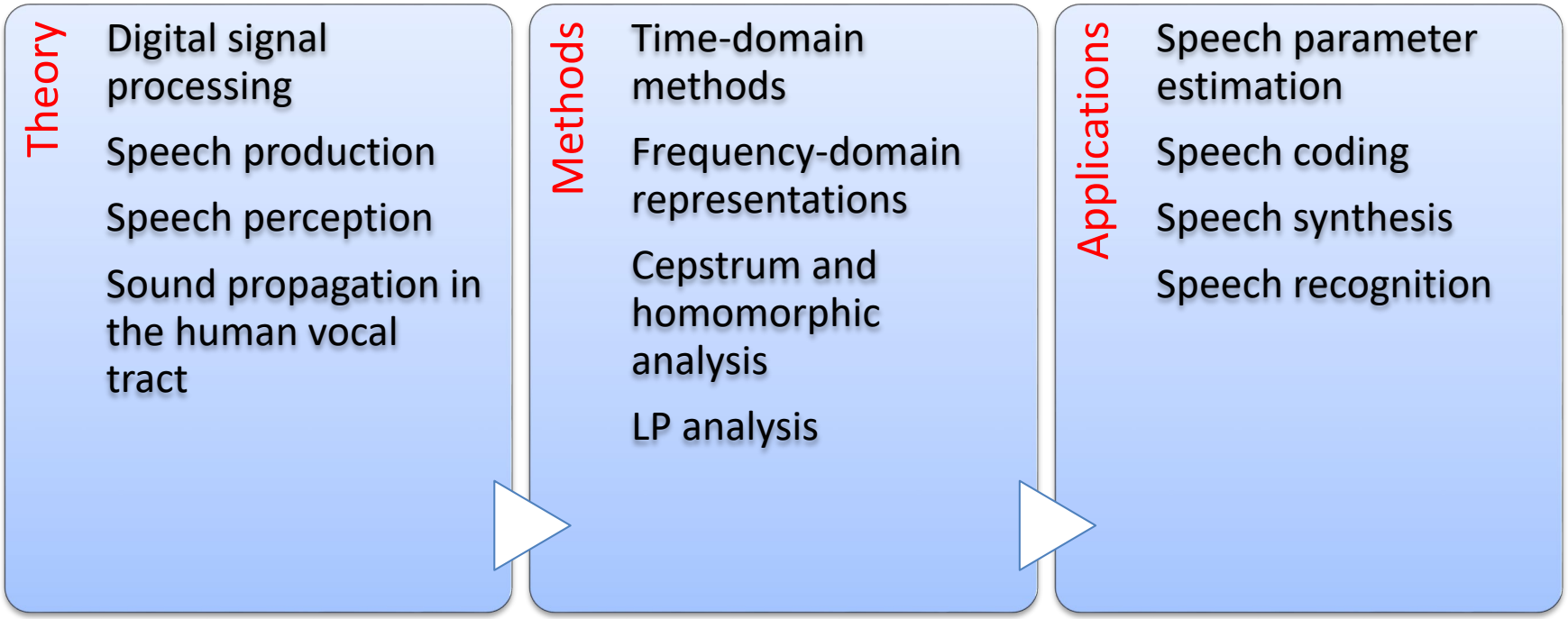
语音信号处理基础
Fundamentals of Speech Signal Processing

课程复习

Course Structure



Chapter 1
Introduction to Speech Signal Processing

- Some basic concepts
  - Speech signal and speech signal processing
  - The speech chain
    - Speech production model
    - Speech perception model
  - Applications of speech signal processing
    - Speech coding/synthesis/recognition and understanding
  - History of speech signal processing

Chapter 2
Review of Fundamentals of Digital Signal Processing

- DSP and discrete signal
- LTI system
- Transform representations
  - z-transform: convergence region
  - DTFT
  - DFT: sampling in frequency ⇔ time domain aliasing
- Digital filters
  - FIR/IIR filter design using Matlab tools
- Sampling
  - The sampling theorem
  - Decimation and interpolation

Chapter 3
Acoustic Theory of Speech Production

- Speech production mechanism
  - Human vocal apparatus
  - Process of speech production
  - Vocal cords and two types of excitation: voiced/unvoiced
  - Source-System model of speech production
- The speech signal
  - Waveforms and spectrograms
    - Silence-background / unvoiced / voiced
    - Wideband / narrowband spectrogram; Formants
  - Phonemes of English
    - Vowel: tongue position; formants
    - Consonants: distinctive features; place/manner of articulation
  - Initials / Finals / Tones in Mandarin Chinese

Chapter 4
Hearing, Auditory Models, and Speech Perception

- The auditory system
  - The composition and the black box model
  - Paired physical attribute and psychophysical observation
- Human ear
  - Function of outer/middle/inner ear
  - Basilar membrane => a bank of filters
- Perception of sound
  - Physical: Intensity / Intensity level / SPL; frequency
  - Psychophysical: Loudness level / Loudness; pitch
- Masking
  - Pure tone masking
  - Noise masking and critical bandwidth
  - Temporal masking
- Auditory model: Perceptual Linear Prediction
- Intelligibility and quality measurement for SP systems

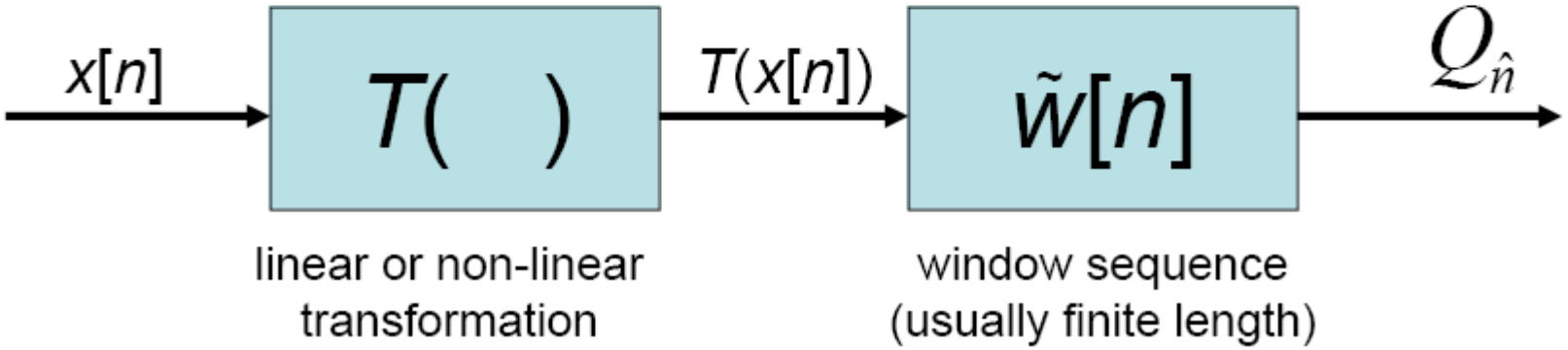
Chapter 5
Sound Propagation in the Human Vocal Tract

- Basic physic model
  - Wave equations: A(x,t), c, Rau
  - Uniform lossless tube => Formant positions
  - Effects of losses => Formant positions/bandwidths
    - Vibration of tube walls / Friction and thermal conduction / Lip radiation
  - VT transfer functions for vowels
  - Voiced excitation
- Concatenate tube models
  - Lossless tube conjunction
    - Signal-flow representation and reflection coefficient
    - Boundary conditions
  - Lossless two tube model with lip/source configuration
  - Digital filter representation => all-pole model
  - General synthesis model $H(z) = G(z)V(z)R(z)$

Chapter 6
Time Domain Methods in Speech Processing

- Speech analysis model
- Short-time analysis of speech

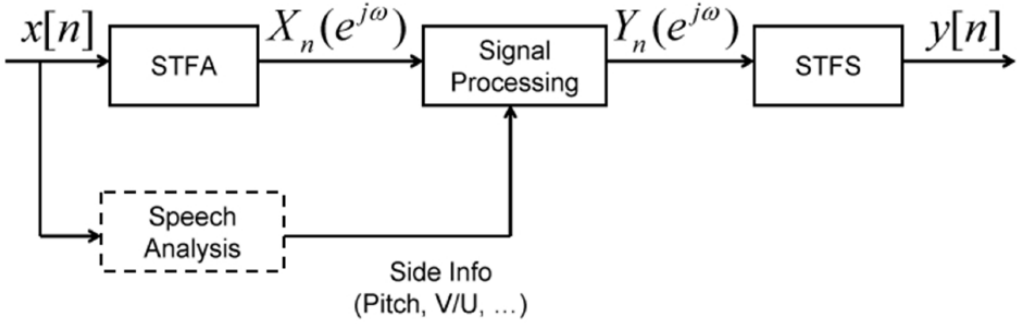
$$Q_{\hat{n}} = \left(\sum_{m=-\infty}^{\infty} T(x[m])\, \tilde{w}[n-m] \right) \Bigg|_{n=\hat{n}}$$



Chapter 6
Time Domain Methods in Speech Processing

- Short-time energy
  - Discriminate voiced/unvoiced sounds from silence
  - Effects of windows
    - Lowpass filtering
    - Frequency response of RW and HW
  - Recursive short-time energy for AGC
- Short-time magnitude
- Short-time average ZC rate
  - Discriminate voiced and unvoiced speech
- Short-time autocorrelation
  - F0 detection
  - Modified autocorrelation
- Short-time AMDF

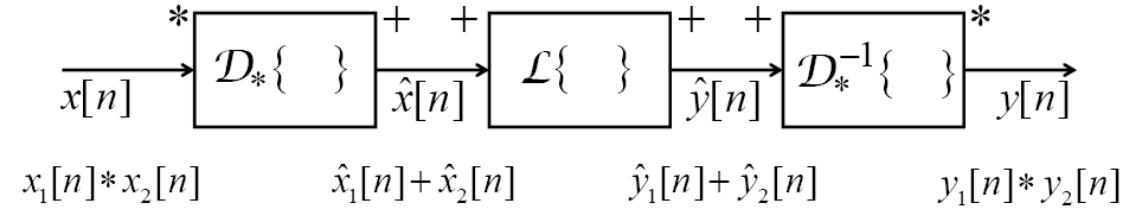
Chapter 7
Frequency-Domain Representations



- STFA and STFS
- STFT—2 interpretations
  - DTFT interpretation
    - Signal recovery from STFT
    - Effect of window length
  - Linear filter interpretation
    - Modulation-lowpass filter / bandpass filter-demodulation
- Sampling rates of STFT
  - Time/frequency sampling rate for exact recovery => total sampling rate
- Overlap addition method / Filter Bank Summation
  - Condition for exact reconstruction
  - Comparisons
- Applications
  - Vocoder; speed-up/slow-down

Chapter 8
The Cepstrum and Homomorphic Speech Processing

- Homomorphic system for convolution



- Characteristic system
  - DTFT
    - issue of phase unwarping
    - Complex and real cepstrum
  - z-transform
    - Cepstrum alanysis of rational z-transforms
    - Cepstrums of minimum/maximum-phase signals; pulse train
- Homomorphic Analysis of Speech Signal
  - Complex cepstrum for speech model
  - Short-time cesptrums from speech
    - Polynomial roots
    - Recursive calculation for minimu/maximum-phase signal
    - DFT: aliasing; lifter
- Application
  - Distance measure; MFCC; vocoder

Chapter 9
Linear Predictive Analysis of Speech Signals

- Basic principal
  - Speech production model

$$s(n) = \sum_{k=1}^{p} a_k s(n-k) + Gu(n)$$

$$H(z) = \frac{S(z)}{GU(z)} = \frac{1}{1 - \sum_{k=1}^{p} a_k z^{-k}}$$

  - linear prediction model

$$\tilde{s}(\hat{n}) = \sum_{k=1}^{p} \alpha_k s(\hat{n}-k)$$

$$P(z) = \frac{\tilde{S}(z)}{S(z)} = \sum_{k=1}^{p} \alpha_k z^{-k}$$

$$e(\hat{n}) = s(\hat{n}) - \tilde{s}(\hat{n}) = s(\hat{n}) - \sum_{k=1}^{p} \alpha_k s(\hat{n}-k)$$

$$A(z) = \frac{E(z)}{S(z)} = 1 - \sum_{k=1}^{p} \alpha_k z^{-k}$$

- Determine α_k by minimizing prediction error
  - Autocorrelation method
    - Equations of Toeplitz matrix
  - Covariance method
    - Equations of symmetric matrix

Chapter 9
Linear Predictive Analysis of Speech Signals

- Frequency domain interpretation
  - LPC spectrum: short-time spectrum estimation with removal of excitation fine structure
  - Relationship with short-time autocorrelation
- Solutions of LPC equations
  - Autocorrelation method: Levision-Durbin algorithm
- Prediction error signal and its spectrum
- Properties of the LPC Polynomial
  - Minimum-phase property
  - Formant estimation from LPC roots
- Relationship with lossless tube model
- Alternative representations
  - Spectral sentsitivity
  - PARCOR / roots / IR / LP cepstrum / log area ratio / LSP

Chapter 10
Algorithms for Estimating Speech Parameters

- Speech/Non-speech detection
  - Rule-based method using log energy and zero crossing rate
- Voiced/Unvoiced/Background classification
  - Bayesian approach
- Pitch detection
  - Autocorrelation: center clipping / doubling error
  - STFT: log harmonic product spectrum
  - Cepstral pitch detector
  - LPC-based pitch detector
- Formant estimation
  - Cepstral-based formant estimation
  - LPC-based formant estimation

Chapter 11
Speech Coding

- Waveform coding
  - Sampling speech signal
  - Instance quantization
    - Uniform: SNR
    - Non-uniform: mu-law / A-law companding; SNR
  - Adaptive quantization
    - Step size adaptation / gain adaptation; AGC
  - Differential quantization
    - Delta Modulation: 1-order prediction and 1-bit quantization
    - Differential PCM: more than 1-order prediction => ADPCM

Chapter 11
Speech Coding

- Model-based (analysis/synthesis) coding
  - Vector quantization
  - Open-loop speech coder
  - Close-loop (analysis by synthesis) speech coder
    - Generate excitation by error minimization with perceptual weighting
    - Use a set of basis functions
      - Multipulse LPC
      - Code-excited LPC
      - Self-excited LPC
- Speech coding standards
  - bit rate / quality / complexity / delay /bandwidth
  - Quality measures

Chapter 12
Automatic Speech Recognition

- Plug-in MAP decision rule for ASR
- Three elements
  - Acoustic model: HMM
  - Language model: N-gram
  - Decoding and search: Viterbi search
- Other ASR related problems
  - VAD
  - Acoustic feature extraction: MFCC, PLP
  - Noise robustness
  - Confidence measure
  - System combination
  - Word accuracy calculation

Chapter 13
Speech Synthesis

- Composition of a speech synthesis system
  - Front-end and back-end
- Back-end
  - Unit selection and waveform concatenation
  - Statistical parametric speech synthesis
  - Comparisons
- HMM-based parametric speech synthesis
  - Model training and parameter generation