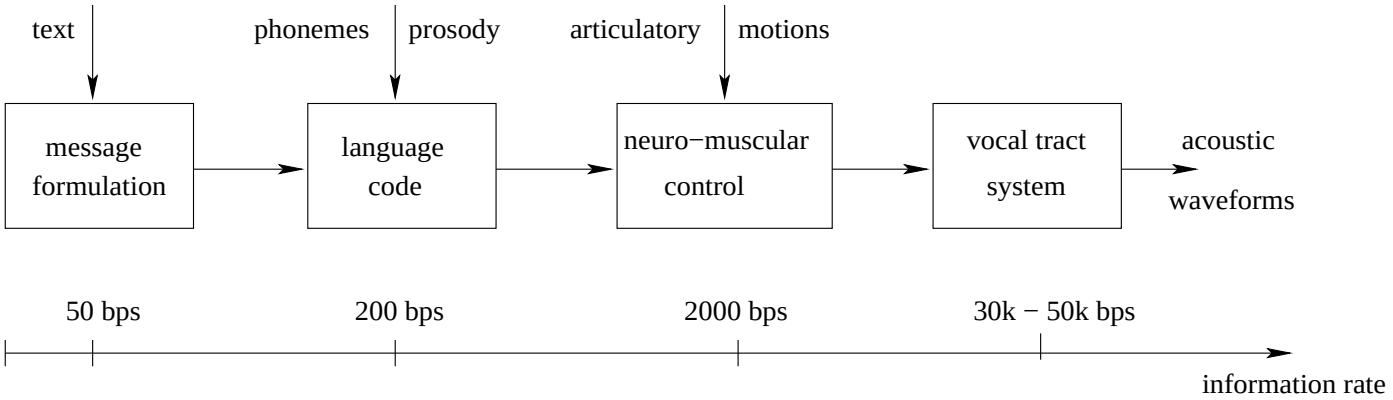
Speech Coding

Introduction to the Problem

- information rates of speech production process



- source channel coding theorem
- compact disc: 128/96 kbps
- AM radio: 64 kbps
- telephone line: 16 kbps (via a band of $200 - 3400$ Hz)

Speech Coding Classifications

- waveform codecs
  direct coding of the sequence of speech signals; low complexity, high-quality but high data rate (bandwidth requirement).
- source codecs (a.k.a. vocoders)
  coding of the vocal tract parameters, pitch and the voice/unvoiced information; high complexity, poor but intelligible quality; low data rate.
- hybrid codecs
  source codec + excitation parameter coding; even higher complexity but achieves better speech quality without significantly increasing the data rate.

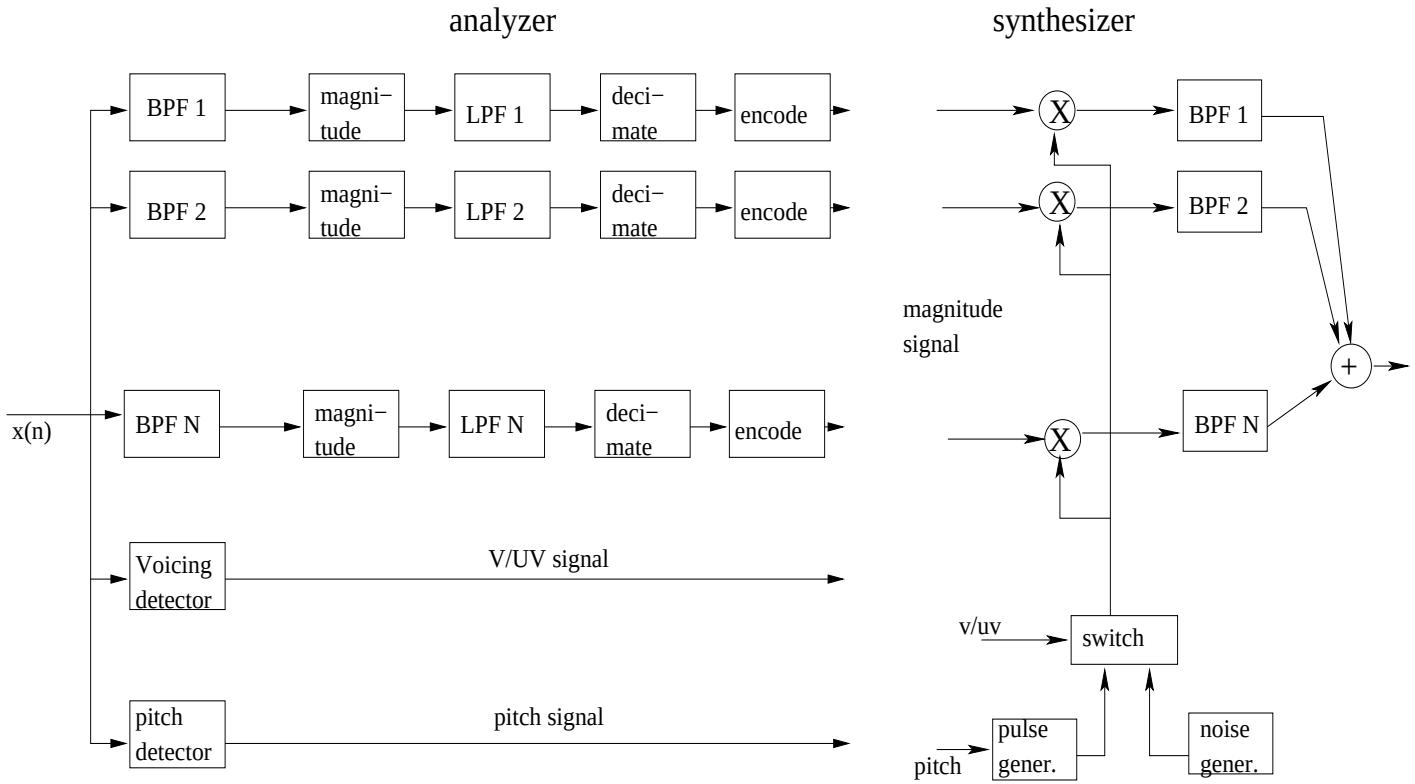
Waveform Codecs

- direct transmission: 8000 Hz $*$ 16 bits/sample $=$ 128 kbps
- PCM (pulse code modulation): the data rate is reduced to 96 kbps; to have constant signal to quantization noise ratio, log coding is often used.
- DPCM (differential PCM): code the difference between adjacent samples, and reduce data rate to 64 kbps
- ADPCM (adaptive DPCM): adaptively modifying the predictor and quantizer.
- SBC (sub-band coding): can allocate more bits to those bands which are perceptually more important.
- ATC (adaptive transform coding): transform-based and transmit the coefficients in the transform. 16 kbps.

Overview of Vocoders

- goal: to transmit speech at low bit-rate
- transmitting parameters
  - vocal tract configuration
  - pitch information
  - voiced/non-voiced decision
- design parameters
  - rate R (2400 bps)
  - number of filters N ($15 - 25$ channels)
  - filter bandwidth specification (stepwise increasing)
  - update rate (50 Hz)

System Diagram



$$R = N * r + v + p$$

Analyzer

- Bandpass filtering: ideally, each band contains a harmonic
- Magnitude: apply rectification or Hilbert transform
- Lowpass filter: to filter out the pitch ripple
- Decimation: reduce the data rate by downsampling
- Encode: use the allocated bits for quantization

Further Bit Rate Reduction

- exploiting correlation in the time and frequency domain
- vector quantizations
- reduce the number of parameters to be transmitted
- segmentation based coding: transmit larger linguistic units

Frame Fill

- Represent the omitted frame with its non-omitted neighbors by choosing the closer one or by a linear combination
- Transmit the representations of omitted frames and the non-omitted frames
- The synthesizer at the receiver uses such information to fill in the omitted frames

Vector Quantizations

- In the sample space, choose prototype points (a.k.a. patterns) and index them. An index/pattern pair is called a codeword and the collection of all codewords is called the codebook.
- Given a new data point, need to find the closest prototype point and use its index to represent the data.
- The main issues here are how the patterns are selected/updated, the metric, and the algorithms to find the closest codeword.

Segmentation Methods for BW Reduction

- the recognition/synthesis approach
  - the analyzer recognizes the speech and the sequence of linguistic units is transmitted, which is synthesized in the receiver.
  - does reduce the bit rate to approximately $50 - 100$ bps
  - quite a bit of information is lost, such as the speaker's identity and the prosody.
  - requires high-accuracy recognizers.
- strongly correlated sequence of frames are merged into segments and the vector quantization is applied on these segments.

Hybrid Codecs

- goal: the low-rate of vocoders and the high quality of waveform codecs.
- the rough representation of excitation signal is refinedly coded and transmitted.
- analysis-by-synthesis mothod: synthesizing speech to decide which candidate is the best.
- CELP (Code-Excited Linear Prediction): the excitation sequence is matched against the codewords in a codebook and the codeword index is transmitted, along with the LP parameters.

Miscellaneous Aspects of Speech Coding

- Encryptions
- Robustness (channel coding)
- Wideband vs. Narrowband
- Applications
  - Mobile Phones
  - Internet Phones
  - POTS
  - Military Scenarios