Coding of Speech

Wolfgang Leister (NR)
Sverre Holm (Ifi)
Anders Lund (Ifi)

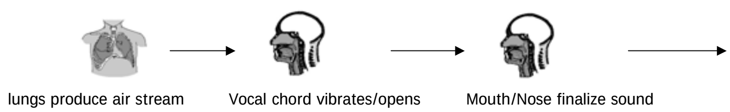
INF5080 - 2005
Ifi, UiO

Overview

- Characteristics of speech
- speech coding
- Speech Synthesis Model
- Waveform encoding
- Vocoders
- Hybrid coders

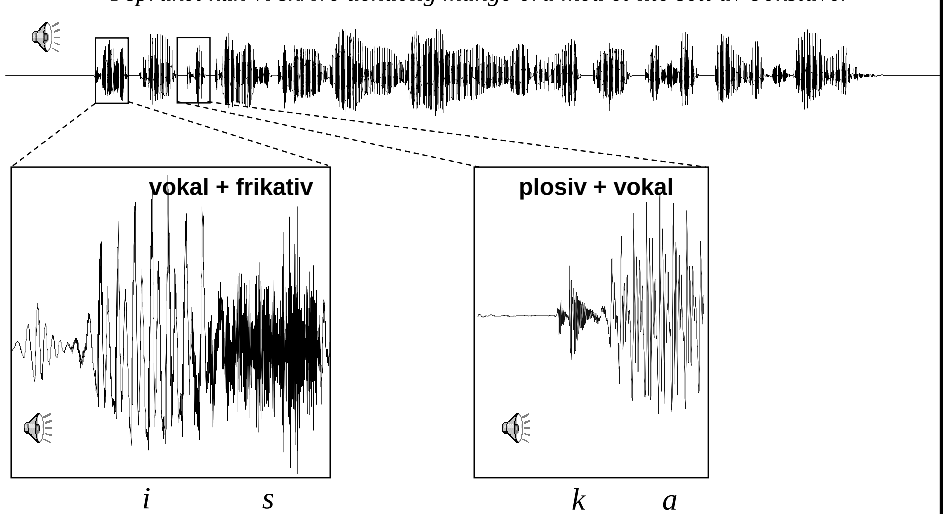
Characteristics of Speech signals

- Nonuniform probability distribution of speech amplitude
- Nonzero autocorrelation between successive speech samples
- Nonflat nature of the speech spectra
- Existence of voiced and unvoiced segments in speech
- Quasiperiodicity of voiced speech signals
- Speech is bandlimited



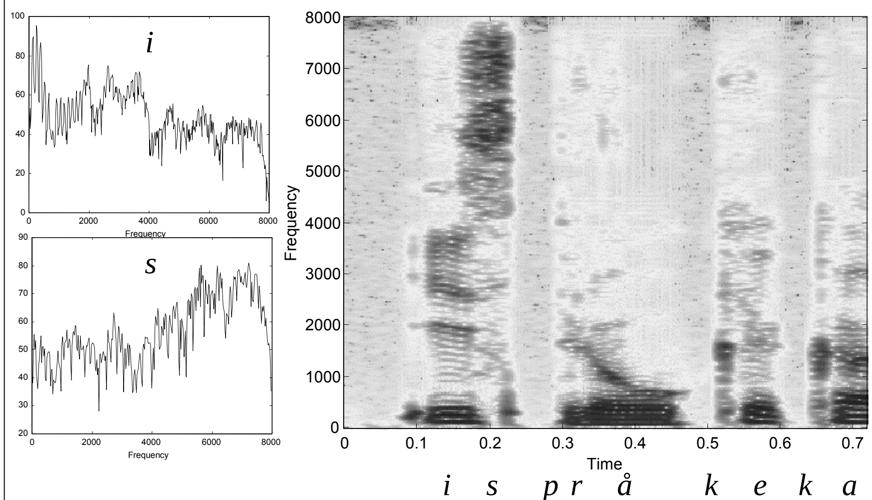
Speech

I språket kan vi skrive uendelig mange ord med et lite sett av bokstaver



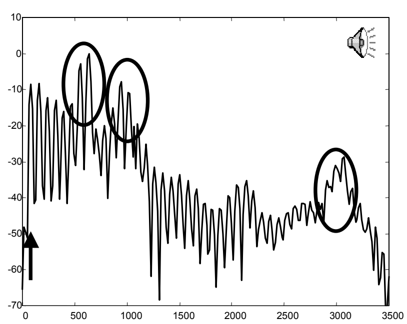
Spectrogram

i språket kan



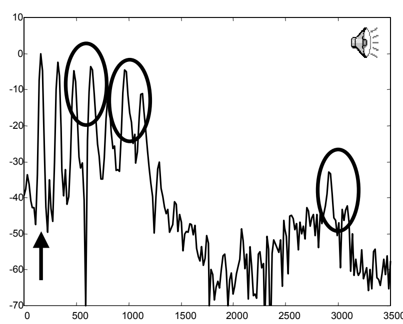
Different voice characteristics – same vocal

female *a*



f_0 = 160 Hz

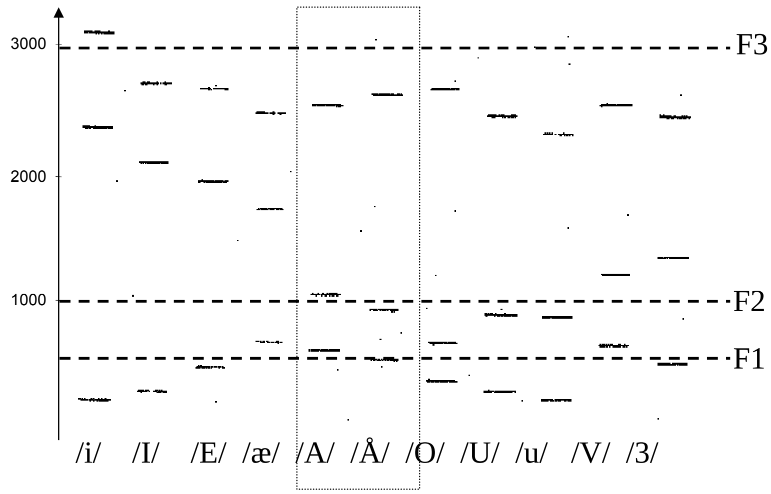
male *a*



f_0 = 80 Hz

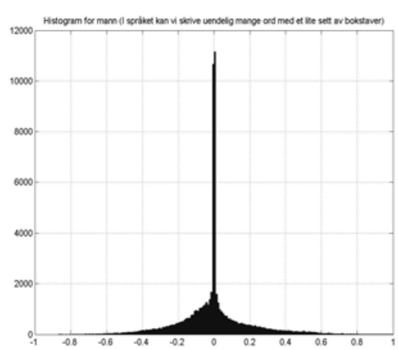
Formants: F1 = 600, F2 = 1000, F3 = 3000

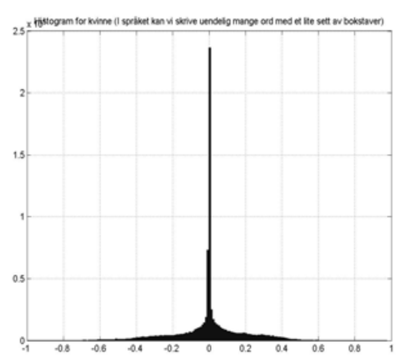
Formants



Nearest match

Histogram





- nearly exponential distribution

Speech coding

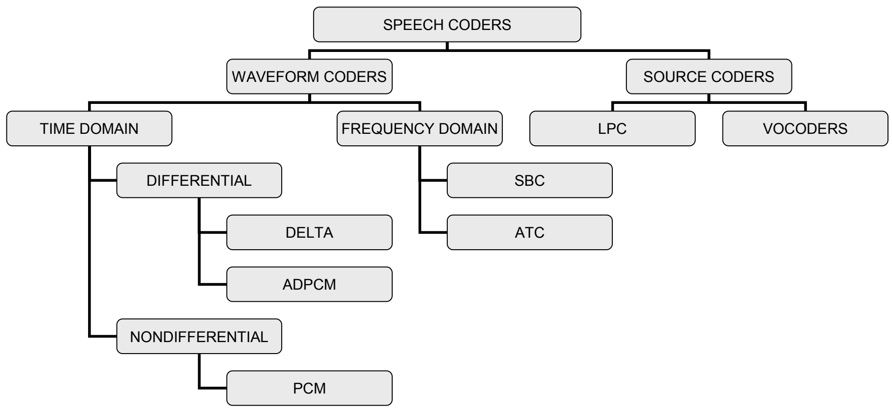
$\rightarrow$ Sampling $\rightarrow$ Quantization $\rightarrow$ Coding $\rightarrow$

Sampling	Quantization	Coding
•Sampling rate	•uniform •non-uniform •adaptive •vector	•Waveform coder •Model based / source coder •hybrid

Quantization Techniques

- Quantization is the process of mapping a continuous range of amplitudes of a signal into a finite set of discrete amplitudes
  - Uniform quantization
  - Nonuniform quantization
  - Adaptive quantization
  - Vector quantization

Types of speech coders



Source: R. Z. Zaputowycz

ADPCM

- Adaptive Differential Pulse Code Modulation
- Exploits redundancies present in speech signal
- Bit rate: 32 kbps,
  - half the std 64 kbps PCM rate, same voice quality
- Differential PCM:
  - the output is the difference between the current amplitude value and the previous one.
- ADPCM:
  - use linear predictor instead
  - Difference between actual sample and predicted value is encoded for transmission (prediction error)
  - Prediction is based on the knowledge of the autocorrelation properties of speech.

Frequency Domain Coding

- The speech signal is divided into a set of frequency components which are quantized and encoded separately.
- Sub-band coding (SBC)
  - Divides the speech signal into many smaller sub-bands and encodes each sub-band separately according to some perceptual criterion.
- Adaptive (Block) transform coding (ATC)
  - Codes the short-time transform of a windowed sequence of samples and encodes them with number of bits proportional to its perceptual significance

Sub-band Coding / SBC

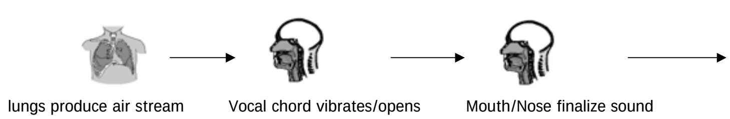
- Speech is divided into four or eight sub-bands by a bank of filters.
- Each sub-band is sampled at a bandpass Nyquist rate and encoded with different accuracy in accordance to a perceptual criteria.
- Bit rates: 9.6 kbps to 32 kbps.
- In this range, speech quality is roughly equivalent to ADPCM

Adaptive Transform Coding (ATC)

- Bit rates: 9.6 kbps to 20 kbps
- Block transformations of windowed input segments of the speech waveform
- Each segment is represented by a set of transform coefficients, which are separately quantized and transmitted.
- One of the most frequently used transforms is the discrete cosine transform (DCT)

Model Based Speech Coding

- Uses characteristics of speech organs:
  - sound generation:
    voiced sounds by chords
    unvoiced sounds from turbulences in tube
  - Model of tube is similar to a linear prediction model, or recursive filter of order 10.
- Linear prediction is basic model for speech coders and speech analysis



- Examples:
  - coder LPC-10 (Dept of Defence): 2.4 kbit/s
  - GSM 6.10 GSM fullrate: 13 kbit/s

Vocoders

- Analyze the voice signal at the transmitter, and transmit parameters derived from the analysis.
- Synthesize at receiver using those parameters
- Much more complex than waveform coders and achieve very high economy in transmission rate
- Less robust, and their performance tend to be talker dependent

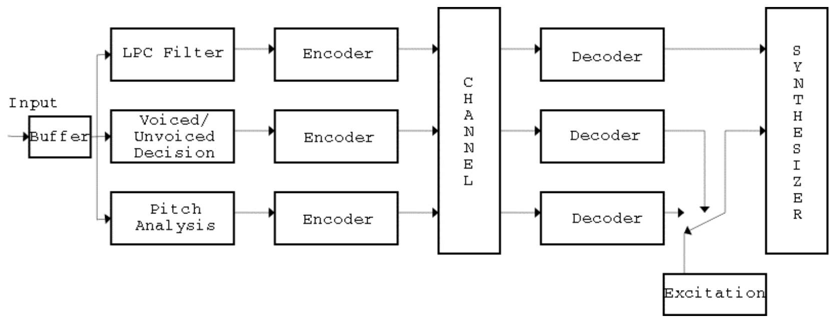
- Vocoding systems
  - linear predictive coder (LPC) - often used
  - Channel vocoder
  - Formant vocoder
  - Voice excited vocoder

Linear Predictive Coders (LPC)

- Extract significant features of speech from the time waveform
- Good quality voice at 4.8 kbps (lower rates with poorer quality possible)
- Prediction principles similar to those in ADPCM.

- Instead of transmitting quantized values of the error signal representing the difference between the predicted and actual waveform, the LPC system transmits only selected characteristics of the error signal

Block diagram of a LPC coding system



LPC Continued

- At the receiver, the received information about the error signal is used to determine the appropriate excitation for the synthesis filter.
- The synthesis filter is designed at the receiver using the received predictor coefficients
- In practice many LPC coders transmit the filter coefficients which already represent the error signal and can be directly synthesized by the receiver.

Speech coders used in various 1G and 2G wireless systems

Standard	Service Type	Speech Coder Type Used	Bit Rate (kbps)
GSM	Cellular	RPE-LTP	9.6, 13
CD-900	Cellular	SBC	16
USDC (IS-136)	Cellular	VSELP	8
IS-95	Cellular	CELP	1.2, 2.4, 4.8, 9.6,13.4, 14.4
IS-95 PCS	PCS	CELP	13.4, 14.4
PDC	Cellular	VSELP	4.5, 6.7, 11.2
CT2	Cordless	ADPCM	32
DECT	Cordless	ADPCM	32
PHS	Cordless	ADPCM	32
DCS-1800	PCS	RPE-LTP	13
PACS	PCS	ADPCM	32

PCM coding of speech and audio

	Frequency area [Hz]	Sampling rate [kHz]	Bit pr sample	Bit rate [kbit/s]
Telephone, speech	300-3400	8	8	64
Broadband speech	50-7000	16	8	128
Medium, audio	10-11000	24	16	384
High quality audio	10-22000	48	16	768

Telephone speech

- 8 kHz sampling, 12-bit quantization
- A-law (Europe) and μ-law (North-America) compression to 8 bit => 8 * 8 = 64 kbit/s
- Instant compression with nearly logarithmic characteristics
- Result is constant SNR, independent of level

NICAM

- Near-Instantaneous Companded Audio Multiplex
- Adaptive PCM
- TV-audio in Europe (BBC)
- 32 kHz sampling, 14-bit quantizing
- Compression to 10 bit based on level in blocks of 1 ms.
- Rate: 2 x 32 x 10 + side info = 728 kbit/s

Sources & Literature

- Theodore S. Rappaport, *Wireless communications – Principles and practice Second edition*, Prentice Hall 2002
- http://www-mobile.ecs.soton.ac.uk/jason/speech_codecs

Repetition
Speech Coding

- Characteristics of speech
- Waveform Coders
- Source Coders / Model driven encoding
  - model speech generation process
  - predict value
  - store parameters and prediction error

End of part