Dialogue systems

Digital Signal Processing

Luděk Bártek

Laboratory of Searching and dialogue, Fakulty of Informatics, Masaryk University, Brno

spring 2023

- The sound characteristics do not change on short time intervals – short term analysis.
- The time interval is called micro segment – length 10 — 40 ms.
- Short term analysis methods:
  - Time domain methods – the values of samples are processed directly.
  - Frequency domain methods – the sample values are transformed into the frequency characteristics that are processed.
- Corti's apparatus modelling – simulates the resonance of particular Corti's apparatus threadbares using differential equations.

- The short-time analysis methods assume that the signal period doesn't change in the window surroundings.
- The error caused by this assumption is compensated by using the "window".
    - Window – a sequence of weights of the samples in the micro segment.
- The weights should correspond to the influencing the sample by its surrounding.
- Commonly used window functions types::
    - the rectangular window function
    - the Hamming's window function.

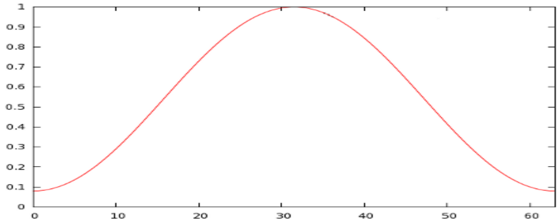
Digital Signal Processing

Hamming's Window Function

- Assumes that the samples closer to the centre of the micro segment are less influenced by surroundings than the samples close to the micro segment boundaries.
- The weight function:

$$w(n) = \begin{cases} n = 0 \ldots N-1 & 0,54 - 0,46 cos(\frac{2\pi n}{(N-1)}) \\ n < 0 \vee n \geq N & 0 \end{cases}$$

- The weight function graph on the micro segment:



- It assumes that:
  1. either the micro segment samples are not influenced by micro segment surrounding for our purpose
  2. or all samples are influenced in the same way.
- All the micro segment samples has assigned the same weight:

$$w(n) = \begin{cases} 0 \leq n < N & 1 \\ n < 0 \vee n \geq N & 0 \end{cases}$$

Time-domain Digital Signal Analysis

- Based on sample values not on spectral characteristics.
- Methods:
  - short-time energy
  - short-time intensity
  - zero crossing rate
  - first order difference
  - autocorrelation function
  - ...

Time-domain Analysis

Short-time energy

- Uses average energy function in the segment:

$$E(n) = \sum_{k=-\infty}^{\infty} (s(k)\omega(n-k))^2$$

  - $s(k)$ – time k sample
  - $\omega(n-k)$ – time k weight according the weight function
- Calculates the window average energy.
- The square root increases the sound signal dynamics.
- Usage:
  - automatic separation of speech (signal) and silence
  - characteristic for simple word classifiers
  - separation of voiced and unvoiced speech parts.

Time-domain Analysis

Short Term Intensity

- Function of signal intensity at given segment.

$$I(n) = \sum_{k=-\infty}^{\infty} |s(k)|\omega(n-k)$$

  - $|s(k)|$ – absolute value of sample in time k.
  - $\omega(n-k)$ – value of weight window corresponding the time k
- Usage – same as the short term energy.
- Doesn't increase the speech signal dynamics as much as the short term energy.

Time-domain Analysis

Short-time Zero Crossing Rate

- Counts the digital signal signum changer.

$$Z(n) = \sum_{k=-\infty}^{\infty} |sgn[s(k)] - sgn[s(k-1)]|\omega(n-k)$$

- Variant – local extreme count.
- Both methods may be negative affected by the background noise.
- Usage:
  - silence detection
  - signal start and end detection even in noisy signal
  - formants approximation
  - simple word classifiers characteristic

- Returns similarity of sequences of the micro segment (the bigger value the more similar sequences shifted by *m* samples).

$$R(m,n) = \sum_{k=-\infty}^{\infty} (s(k)\omega(n-k))(s(k+m)\omega(n-k+m))$$

- When the signal period is P then the R(m,n) maximum is when m=0, P, 2P, ...
- Assumes the segment length *2P* at least.
- Usage:
  - The F_0 period.
  - Base for LPA coefficients calculation.

Frequency Domain Signal Analysis

- Transforms digital signal from time domain into the frequency domain.
- It uses the Fourier transformation most often.
- Commonly used types of frequency domain analysis:
  - short term Fourier transformation
  - short term discrete Fourier transformation
  - Fast Fourier transformation
  - cepstral analysis
  - linear predictive analysis
  - . . .

- Based on Fourier transformation:

$$S(\omega, t) = \sum_{k=-\infty}^{\infty} s(k)w(t-k)e^{-i\omega k}$$

  - Regular Fourier transformation can be obtained by fixating time t
  - $|S(\omega, t)|$ – acoustic spectra component amplitude corresponding to frequency ω on time t.
  - $w(n)$ – window weight function.
- It expects the periodic function on input – sound is periodic on short intervals only.
- When using it we assume the micro segment is repeating periodically.

- Used to express spectral properties of sequence of samples with period of N sample, final sequences of length N samples.
- Coefficients $X(k)$ calculations using DFT:

$$X(k) = \sum_{n=0}^{N-1} x(n) e^{-i\frac{2\pi}{N}kn} = \sum_{n=0}^{N-1} x(n) W_N^{-kn}$$

  - $|X(k)|$ – k-the spectra coefficient intensity, frequency depends on the micro segment length N and on the sampling frequency.
  - $x(n)$ – the micro segment n-th sample.
  - $W_n = e^{i\frac{2\pi}{N}} = cos(2\pi/N) + i \cdot sin(2\pi/N)$.
- Calculation of n-th sample based on coefficient $X(k)$ – IDFT:

$$x(n) = \frac{1}{N} \sum_{n=0}^{N-1} X(k) e^{i\frac{2\pi}{N}kn} = \frac{1}{N} \sum_{n=0}^{N-1} X(k) W_N^{kn},$$

- Spectral coeficient calculation using DFT – n^2 complex number operations.
- Using FFT – $N \cdot log_2 \frac{N}{2}$ multiplication operations.
- FFT requires the length of analysed segment to be power of 2.
  - uses the divide et impera method to optimize the DFT calculation
  - it separates the calculation of even and the odd members of the sum
  - previous can be understood as transformation of two vectors $(x_0, x_2, \ldots, x_{N-2})$ a $(x_1, x_3, \ldots, x_{N-1})$, they differs only in member $(e^{-i\frac{2\pi}{N}})^k$ and the transformation itself is the same.

Frequency Domain Signal Analysis

Cepstral Analysis

- Based on the vocal tract activity model.
- Speech oscillations can be modelled as the response of a linear system to stimulation consisting of a sequence of pulses for voiced speech and noise for unvoiced speech.
- Cepstra – $X(k) = IFFT(log|FFT(x(k))|)$
- Cepstral analysis allows us to separate the pulses ($f_0, \ldots$) and the vocal tract parameters.
- Usage:
  - Speech phonetic structure evaluation – voice, f_0 period, formants, ...
  - speech recognition
  - speaker verification and identification
  - ...

Frequency Domain Signal Analysis

Linear Predictive Analysis

- One of most effective acoustic signal processing methods – ensures very accurate estimations of parameters at a relatively low computational load.
- Based on the assumption that the $s(k)$ can be described as the linear combination of N previous samples and the pulse functions $u(k)$:

$$s(k) = -\sum_{i=1}^{N} a_i s(k-i) + Gu(k)$$

  where the G is the amplification coefficient and the N is the model order.
- Usage:
  - vocal tract model spectral characteristics determination
  - the information about sonority and the f_0 can be determined from the prediction error
  - a_i coefficient carry the spectral properties information – can be used as the speech recognition feature