Digital Signal Processing

Jiří Málek

Part I

Application of Digital Signal Processing

Utilization of DSP:

- *Biomedical applications* (diagnostics, patient monitoring, prevention)
- *Communication* (encoding, decoding, encrypting, filtration)
- *Control systems* (servomechanisms, autopilots)
- *Signal analysis* (signal modeling, classification, compression)
- *Image processing* (image modifications, computer vision)
- *Multimedia* (movies, digital television, video-conferences)
- *Musical and sound applications* (recording, reproduction, special effects)
- *Speech applications* (denoising, compression, recognition, synthesis)

DSP Applications II

Too many areas to cover in a single lecture. We will focus on:

- Processing of musical signals
- Processing of biomedical signals (ECG)
- Speech enhancement - denoising
    - Spectral thresholding
    - Spectral subtraction

Part II

Processing of musical signals

Processing of musical signals I

- Audio signals generated by musical instruments have rather simple mathematical model

$$x(t) = a(t) \sum_{m=1}^{\infty} c_m \cos(2\pi m f_0 t + \phi_m) \quad (1)$$

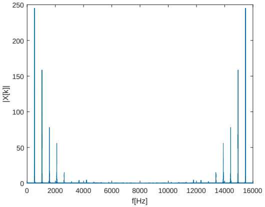
where f_0 is a pitch, c_m is an amplitude (loudness) and ϕ_m is a phase of the mth harmonic component. $a(t)$ is the signal "envelope" (time dynamics); low-frequency signal modulating the sound amplitude.

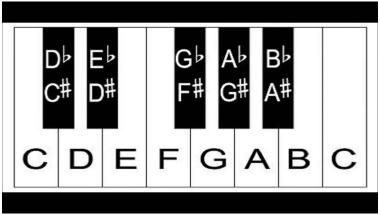
- DETAILS: Basic musical theory
- **Example:** The same tone played on various instruments

Violoncello, *Classical guitar*
Flute, *French horn*

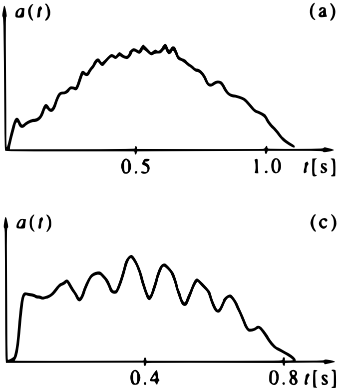
Basic musical theory

- European music has twelve halftones in an octave (c-c'), which can be well imagined through piano keys.
- Each tone has a pitch given by its "fundamental frequency" and higher harmonics which determine a color/timbre of the tone (this differs tones played on various instruments).
- If the tone has fundamental frequency f_0 (e.g., a' has frequency 440 Hz), then a tone higher by one halftone has frequency $f_0 \cdot 2^{1/12}$ (that is a#' has frequency 466 Hz).
- Tones which differ by an octave have double frequency (geometric series with ratio $2^{1/12}$)



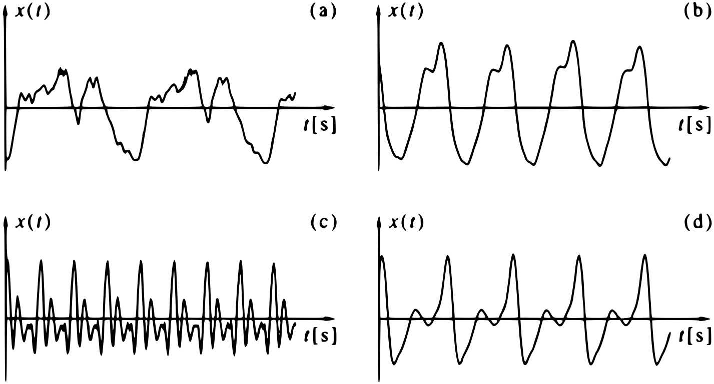


Left: DFT magnitude spectrum of tone c" on flute (basic frequency approximately 523Hz)
Right: Piano keys (SOURCE: `piano-keyboard-guide.com`)



Tone Dynamics: (a) Violoncello, (b) Classical guitar, (c) Flute, (d) French horn

SOURCE: BOAZ PORAT, A Course in Digital Signal Processing



Time course: (a) Violoncello (220 Hz), (b) Classical guitar (440 Hz), (c) Flute (880 Hz), (d) French horn (440 Hz)

SOURCE: BOAZ PORAT, A Course in Digital Signal Processing

Additive synthesis:

- The very basic synthesis uses the mathematical model in (1)
- To create a tone of a specific instrument, the knowledge of its time dynamics and harmonic structure is needed
- Advantage: Tone modification can be easily achieved by a change of several model parameters
- Disadvantage: Imperfect realism, the model (1) does not take all sound phenomena into consideration

Examples:

Generated melody - French horn, *Recorded melody - French horn*, *Generated melody - Trumpet*

Sample-based synthesis

- Basic elements are recorded samples of true musical instruments
- Acords are created by summation of partial samples
- Advantage: The generated tones are very realistic
- Disadvantage: Realistic tone modification requires "complicated" DSP techniques
- Disadvantage: Memory requirements

Examples:

Violoncello, *French horn*,
Violoncello + *French horn*

Sound effects using DSP:

- DSP is used to generate *sound effects* - artificially created or modified sounds, which are used in the movies, electronic music, computer games and life performances
- To create the sound effects, one can use linear (adaptive) filtering or various nonlinear (irreversible) mothods
- More about creation the creation of sound effects can be learned in subject **Digital Audio Engineering** (DAI, in Czech), which is tutored by *Prof. Ing. Zbyněk Koldovský, PhD.*

Example: Digital sound effects applied to the sound of electric guitar

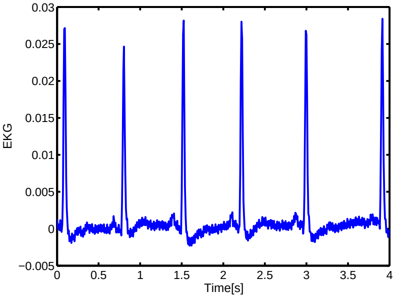
Original sound	*Vibrato*
Delay	*Chorus*
Tremolo	*Wah-wah*

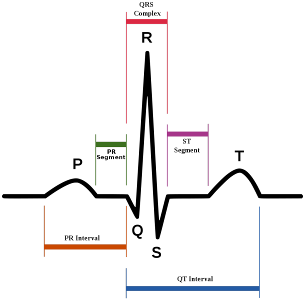
Part III

Processing of biomedical signals

Processing of ECG

- A lot of different biomedical signals exist, their analysis is differs for each specific signal
- **Electrocardiogram** is a electrical signal reflecting the heart activity
- ECG of a healthy heart has a characteristic shape - QRS complex
- The time-domain analysis of the (shape of) QRS complex allows the physicians to state diagnosis
- The sampling frequency of ECG is usually $Fs = 500Hz$ and more
- ECG can contain many unwanted signals called *artifacts*
- *Narrow-band noise:* isoline drift (slow pacient movements, breathing), voltage
- *Wide-band noise:* Myopotencials (muscle movements), sudden isoline changes (insuficient electrode contact)
- Artifact removal by fitration is performed only in the case, when the signal cannot be remeasured correctly
- **Fidelity criteria** do not allow linear filtering of the voltage noise in the strictest case
- Often, the filtering can be performed by (adaptive) FIR filters with linear phase. If the IIR filter is applied, its nonlinear phase needs to be compensated.





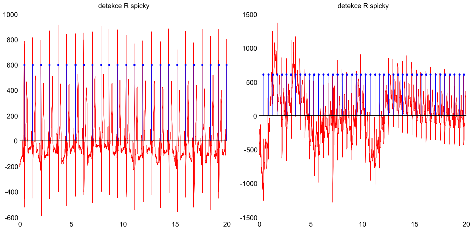
ECG and its stages

SOURCE (right image): en.wikipedia.org

Example: Detection of the R pitch within the ECG signal

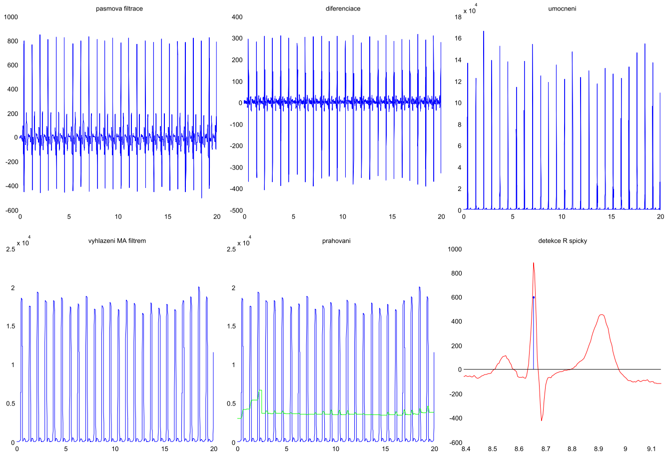
1. Bandpass filtering, pass-band 2-30Hz
2. Differentiation: $y[n] = x[n] - x[n-1]$,
   $y[n] = 0.5(x[n] - x[n-2])$
3. Squaring - emphasizes large and attenuates small values
4. Moving average filtering/smoothing
5. Thresholding
6. Detection of R pitches

More about biomedical signal processing can be learned in subject **Modern methods of signal processing** (MMZ, in Czech), which is lectured by *Prof. Ing. Zbyněk Koldovský, PhD.*



SOURCE: doc. Ing. ROMAN ČMEJLA CSc., Biomedical Signals, lectures

Example: Detection of the R pitch within the ECG signal



SOURCE: doc. Ing. ROMAN ČMEJLA CSc., Biomedical Signals, lectures

Part IV

Speech Enhancement

Speech Enhancement

Broad set of algorithms, which aims to increase speech quality and to remove various distortions within the speech signal.

- **Quality:** clarity, subjective intelligibility, compatibility with a subsequent signal processing technique
- **Distortion:** Aditive noise, reverberation, nonlinear distortions (clipping), influence of codecs/recording device/sensors
- **Evaluation:** Objective (computable criteria, recognition accuracy), subjective (listening tests)

Speech denoising:

- Identification and removal (suppression) of non-desired components within speech.
- Which component is non-desired (i.e., noise) depends on the situation.
- According to available information about noise signals, a denoising algorithm is chosen.

Assumption:

- *Stationary, always active:* Spectral subtraction
- *Directional, unknown direction:* Blind Source Separation
- *Directional, known direction:* Target cancellation filters, beamforming
- *Extensive sample database:* Machine learning principles

Part V

Spectral thresholding, Spectral subtraction

Denoising by spectral thresholding I

- Simple method for removal of additive background noise, provided that the SNR is high enough for the most frequency components of speech
- **Assumption:** Noise $v[n]$ has lower instantaneous power than speech $s[n]$, i.e., frequency bins with the lowest power corresponds to noise
- Based on STFT and additive model of the noisy signal ($x[n] = s[n] + v[n]$)
- **Disadvantage:** The assumption hold to a limited extend for a real-world signal, the method removes also the weak components of speech
- One of methods called jointly Spectral Masking

Algorithm:

1. Computation of the Short-time Fourier Transform (STFT): $X[m,k] = \mathrm{STFT}(x[n])$ (m-frame index, k-frequency bin index)
2. Thresholding/masking: $\hat{S}[m,k] = W[m,k] \cdot X[m,k]$, where $W[m,k]$ is a spectral mask (here a binary thresholding function)

$$W[m,k] = \begin{cases} 0, & |X[m,k]| \leq T \\ 1, & |X[m,k]| > T \end{cases} \quad (2)$$

   and T is a suitable threshold.
3. Reconstruction using an inverse STFT: $\hat{s}[n] = \mathrm{ISTFT}(\hat{S}[m,k])$.

Perfect signal reconstruction using ISTFT I

- Processing of signal $s[n]$ via STFT:
  1. Signal segmentation into frames of length M weighted by a window $w[n]$, i.e.,
  $$s_m[n] = s[n] \cdot w[n - mR], \tag{3}$$
  where R is a shift of the time frames.
  2. Computation of DFT for each segment $S[m,k] = DFT(s_m[n])$.
- It is possible to reconstruct $s[n]$ perfectly from its STFT image $S[m,k]$, provided that the *shape* and the *shift* of $w[n]$ is suitable.
- This is possible to inversibility of DFT, if $s[n]$ can be reconstructed from weighted frames $s_m[n]$.
- **Condition:** (for the shape of $w[n]$)

$$\begin{aligned} s[n] = \sum_{m=-\infty}^{\infty} s_m[n] = \sum_{m=-\infty}^{\infty} s[n]w[n-mR] \\ = s[n] \sum_{m=-\infty}^{\infty} w[n-mR]. \end{aligned} \tag{4}$$

Perfect signal reconstruction using ISTFT II

- Using equation (4) we derive the condition

$$\sum_{m=-\infty}^{\infty} w[n-mR] = \text{const.}, \tag{5}$$

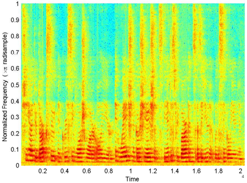
i.e., the sum of overlapping shifted windows must be equal to a constant.

- Reconstructed signal is then similar to the original, up to a scaling.
- The conventional symmetric windows must be slightly modified to fulfill (5), .e.g, last sample must be set to zero.
- MATLAB: Windows for perfect reconstruction can be obtained using parameter `'periodic'`, e.g., `w = hamming(512,'periodic')`
- For different windows is the allowed shift R in (5) different:

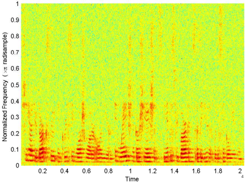
Window	Shift $R \in \mathcal{Z}$
Rectangle	M/p
Bartlett	M/p
Hann	$(M/2)/p$
Hamming	$(M/2)/p$
Blackmann	1

($p \in \mathcal{Z}$ - a free parameter)

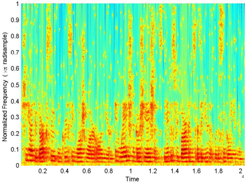
Example: Removal of Gaussian Noise



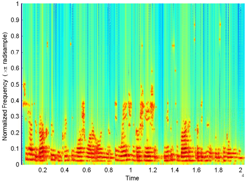
(a)



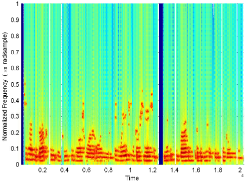
(b)



(c)



(d)



(e)

(a) Clean Speech, (b) Mixture with gaussian noise (SNR=10dB), (c) Enhanced ($T = 0.008 \cdot max(|X[m,k]|)$) (d) Enhanced ($T = 0.02 \cdot max(|X[m,k]|)$) (e) Enhanced ($T = 0.035 \cdot max(|X[m,k]|)$)

Denoising by spectral subtraction I

- A method for removal of stationary additive noise
- **Assumption:** Noise $v[n]$ has stationary spectrum $V[k]$ and is active separately during some time instant (i.e., target speech $s[n]$ is not active, which needs to be detected - Voice activity detection)
- Based on STFT and additive model of the noisy speech ($x[n] = s[n] + v[n]$)
- **Disadvantage:** Changes in noise spectrum need to be detected or the spectrum must be periodically updated (i.e., $V[m,k]$ is time-variant)

Algorithm:

0. Estimation of magnitude spectrum $V[k]$ (smoothing of $|\text{STFT}(v[n]))$ (estimation in interval, where $x[n] = v[n]$, i.e., $s[n] = 0$).
1. Computation of STFT: $X[m,k] = \text{STFT}(x[n])$
2. Computation of magnitude spectrum $|X[m,k]|$ and phase $\phi[m,k]$:

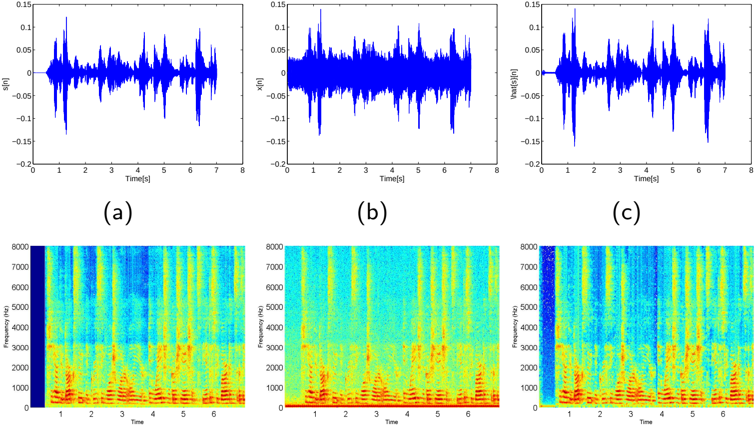
$$\phi[m,k] = \tan^{-1} \frac{\text{Im}(X[m,k])}{\text{Re}(X[m,k])} \tag{6}$$

3. Estimation of speech magnitude spectrum $|\hat{S}[m,k]|$:

$$|\hat{S}[m,k]| = \begin{cases} |X[m,k]| - |V[k]|, & |X[m,k]| - |V[k]| > 0 \\ 0, & else \end{cases} \tag{7}$$

4. Reconstruction $\hat{s}[n] = \text{ISTFT}(|\hat{S}[m,k]| \cdot e^{j\phi[m,k]})$

Example: Fan noise subtraction



(a) Clean speech,
(b) Noisy signal with a fan noise (SNR=5dB),
(c) After subtraction

Thank you for attention!