

# Počítačové zpracování řeči

Luděk Bártek

Fakulta Informatiky  
Masarykova Univerzita Brno

podzim 2013

# Obsah

- 1 Základy fonetiky
- 2 Digitalizace akustického signálu

# Fonetika

- Přírodní věda na pomezí lingvistiky, anatomie, fyziologie a fyziky (akustiky)
- Zkoumá zvukovou stánku jazyka z různých aspektů jazyka:.
  - fyziologickou činnost mluvidel
  - akustickou podstatu zvuků.
- Dělení fonetiky:
  - artikulační - tvorba fónů ve zvukovém ústrojí
  - akustická - přenos zvuků prostředím, jejich frekvence, ...
  - percepční - jak jsou zvuky přijímány,

# Základy fonetiky

- Foném - elementární zvukový segment, který je schopný diferencovat vyšší znakové jednotky jazykového systému (morfémy).
- Fonémy:
  - samohlásky:
    - základní frekvence
    - formanty
  - samohlásky:
    - znělé - na vzniku se spolupodílí hlasivky
    - neznělé.
- Koartikulace - vzniká změnou parametrů řečového ústrojí při přechodu z jedné hlásky na druhou
  - další řečové jednotky:
    - alofón
    - difón
    - trifón
- Fonetický přepis - Jednoznačný a přesný zápis mluvené řeči. Psaná a mluvená forma téže promluvy se mohou lišit.

# Fonetický přepis

- Přesný a jednoznačný zápis mluvené řeči.
- IPA - International Phonetic Alphabet
  - součást standardu UNICODE.
- Národní fonetické abecedy.
- TTS - většinou využívají 7bitový ASCII přepis znaků z IPA (SAMPA - Speech Assessment Methods Phonetic Alphabet) např. DITe
- pravidla pro přepis:
  - změna znělosti na hranici znělá souhláska - neznělá souhláska
  - měkčení souhlásek, pokud následuje *i*, *ě*
  - ...
- Občas může být přepis regionálně závislý:
  - *sh*:
    - Čechy - sch
    - Morava - zh

THE INTERNATIONAL PHONETIC ALPHABET (2005)

CONSONANTS (PULMONIC)	VOICELESS					VOICED					VOICELESS ALVEOLAR		GLIDAL
	Bilabial	Labiodental	Dental	Alveolar	Postalveolar	Bilabial	Labiodental	Dental	Alveolar	Postalveolar	Palatal	Velar	
Nasal	m	ɱ		n	ɳ	ɱ	ɱ	ɳ	ŋ	ɴ			
Plosive	p	ɸ	t̪	t	ʈ	c	ɟ	k	q	g	ç	ʁ	ʕ
Fricative	f	ɸ	v	θ	ð	s	z	ʃ	ʒ	ʂ	ʂ	ʁ	ʕ
Approximant				ɹ		ɻ	ɻ	ɻ	ɻ	ɻ	ɻ	ɻ	ɻ
Liquids				l		ɭ		ɭ					
Lateral fricative				ɬ		ɮ		ɮ					
Lateral approximant				l		ɭ		ɭ					
Lateral fricative				ɬ		ɮ		ɮ					

When symbols appear in pairs, the one to the right represents a normally rounded consonant, except for approximant & the first seven dorsal articulation points & it is responsible. Light grey letters are unclassified consonants of the IPA.

CONSONANTS (NON-PULMONIC)				CONSONANTS (D-ARTICULATED)			
Voiced (fricative or plosive)	Vibrant (trill)	Uvular	Ejectives	M	W	U	G
Voiced fricative	ʙ	ʀ	ʕ	ʙ	ʀ	ʕ	ʕ
Labiodental fricative	ɸ	ɸ	ɸ	ɸ	ɸ	ɸ	ɸ
Dental fricative	θ	ð	θ	θ	ð	θ	ð
Alveolar fricative	s	z	s	s	z	s	z
Postalveolar fricative	ʃ	ʒ	ʃ	ʃ	ʒ	ʃ	ʒ
Palatal fricative	ç	ʁ	ç	ç	ʁ	ç	ʁ
Velar fricative	x	ʁ	x	x	ʁ	x	ʁ
Uvular fricative	ʁ	ʕ	ʁ	ʁ	ʕ	ʁ	ʕ
Glottal fricative	h	ʕ	h	h	ʕ	h	ʕ

VOEELS	SUPRASEGMENTALS	TONE
Front: near-front, central, near-back, back	Primary stress: ˈ Secondary stress: ˈ	Low: ˩
Close: i, y, ɨ, ɥ, ɨ̞, ɥ̞, ɨ̞̞, ɥ̞̞	Secondary stress: ˈ	High: ˥
Near-close: ɪ, ʏ, ɨ̞̞̞, ɥ̞̞̞	Long: ː	High falling: ˥˩
Close-mid: e, ø, ɘ, ɚ, ɞ, ɤ, ɟ, ɟ̞	Short: ˑ	High rising: ˩˥
Mid: ə, ɜ̞, ɝ̞, ɞ̞	Intermittent: ˑ	Low: ˩
Open-mid: ɛ̞, ɜ̞̞, ɝ̞̞, ɞ̞̞	Intermittent: ˑ	High falling: ˥˩
Near-open: ɛ̝, ɜ̝, ɝ̝, ɞ̝	Intermittent: ˑ	Low falling: ˩˥˩
Open: ɛ, ɜ, ɝ, ɞ	Major intermitted break: ˑ	Uphang: ˩˥
Voiced at right & left of glides are normally unrounded.	Global rise: ˩˥˩	Downstep: ˩˩
	Global fall: ˥˩˩	Dipping: ˩˩˩

DIACRITICS			
PLACE	MANNER	VOICE	OTHERS
Labial	Labial	Labial	Labial
Dental	Dental	Dental	Dental
Alveolar	Alveolar	Alveolar	Alveolar
Postalveolar	Postalveolar	Postalveolar	Postalveolar
Palatal	Palatal	Palatal	Palatal
Velar	Velar	Velar	Velar
Uvular	Uvular	Uvular	Uvular
Glottal	Glottal	Glottal	Glottal

Podrobnosti viz stránky International Phonetic Association  
(<http://www.langsci.ucl.ac.uk/ipa/>)

# Česká fonetická abeceda

- Krátké samohlásky:
  - a a pata, εε led, ii lid, o/ɔo rod, uu ruka
- Dlouhé samohlásky:
  - a: a:, á pátá; ε: e:, é léto; i: i:, í lípa; o:/ɔ: o:, ó tón; u: u:, ú úkol
- Dvojhlásky:
  - au au, au auto: εu eu, eu euforie' ou/ɔu ou ou houba
- Souhlásky:
  - m m matka' ŋ ŋ, tramvaj' n n nos ŋ ŋ kůň ŋ ŋ banka
  - p p pes; b b babička; t t táta; d d dům; c ě ěapka; k k kost; g g gram
  - ts c co' dʒ ʒ, dz' leckdo' fʃč čáp' dʒ, dž džem
  - f f fuj' v v voda' s s sen' z z zub' ʃ š šíp' ʒž žena' x x, ch chléb' xʃabych byl' fɦ hra
  - r r rak' r̩ ř řeka' r̩ ř, ř rybář
  - j j já; l l les
  - r̩ r̩ krk' ! ! vlk' m̩ m̩ rožmberk

## Ukázka textu v abecedě SAMPA spolu s výsledkem.

- Fonetický přepis věty "Čeština je krásná řeč"
  - tSeSTina je kra:sna: r/etS
  - Syntetizovaná věta "Čeština je krásná řeč."  
(data/cestina.wav)



# Fonetika - samohlásky

- Krátké samohlásky - a, e, i, o, u
- Dlouhá samohlásky - á, é, í, (ó), ú
- Dvojhlásky - (eu), (au), ou
- Samohlásky:
  - základní hlasivkový tón - 100 - 400 Hz
  - formanty - rezonancí v dutinách hlasového traktu zesílené části akustického spektra

# Formanty

- Určující pro rozpoznávání samohlásek
- Formant  $F_1$  vzniká rezonancí v dutině ústní
- Formant  $F_2$  vzniká rezonancí v dutině hrdelní
- Hlavní formanty - spektrální poloha a intenzita může být dána:
  - muž
  - žena
  - dítě
  - individuálně
- Vyšší formanty  $F_3$  -
  - výskyt bývá individuální

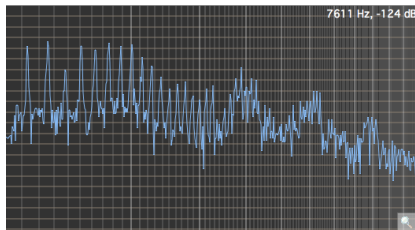
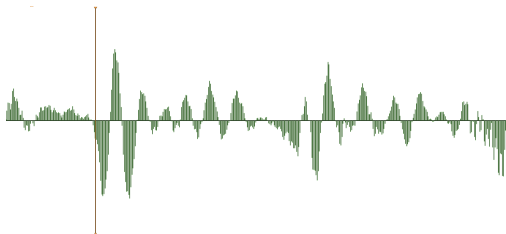
# Formanty F1 a F2 pro české samohlásky

hláska	F1 [Hz]	F2 [Hz]
a	750 - 1100	1100 - 1500
e	500 - 700	1500 - 2000
i	300 - 500	2000 - 3000
o	500 - 700	900 - 1200
u	300 - 500	600 - 1000

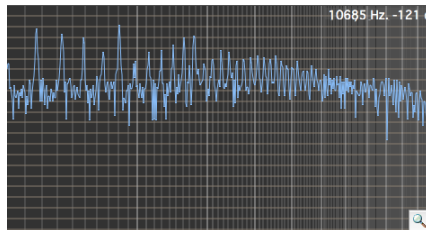
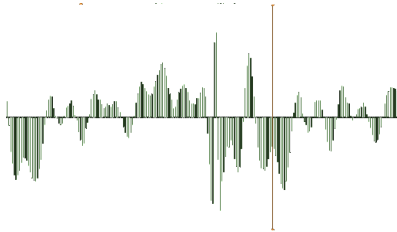
# Četnost výskytu samohlásek

- e - 10 %
- a, o, i - 6 — 7 %
- í - 4 %
- další jen s nepatrnou frekvencí:
  - á, u, é, ou, ú
  - ó, au, eu

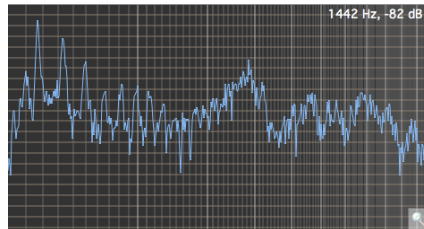
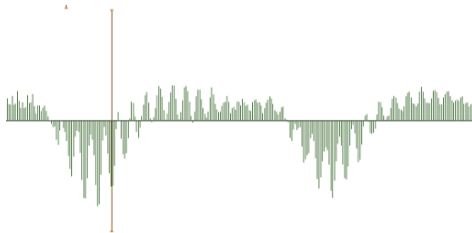
# České samohlásky - a



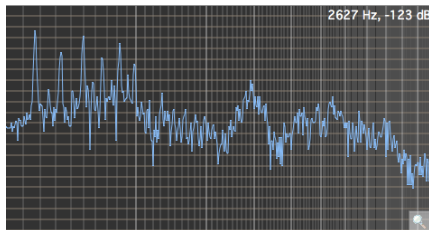
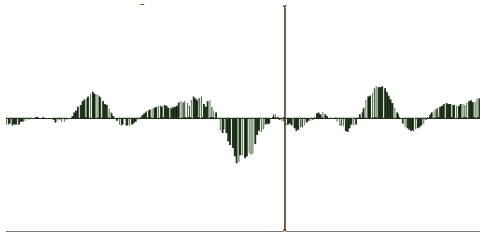
# České samohlásky - e



# České samohlásky - i

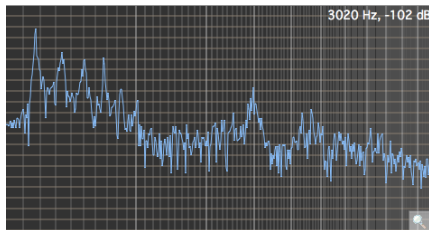
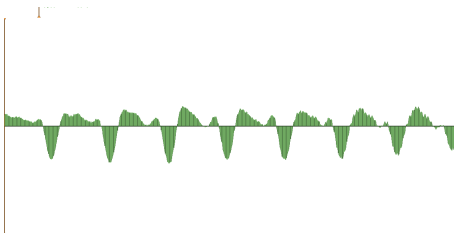


# České samohlásky - o

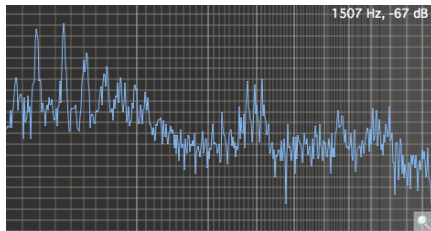
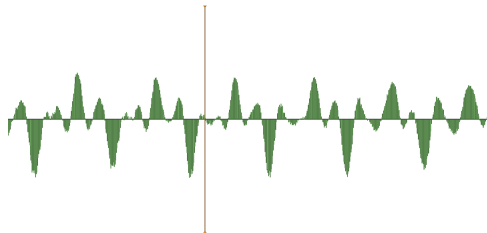




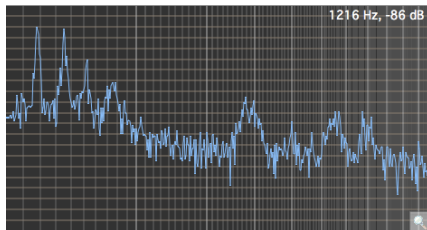
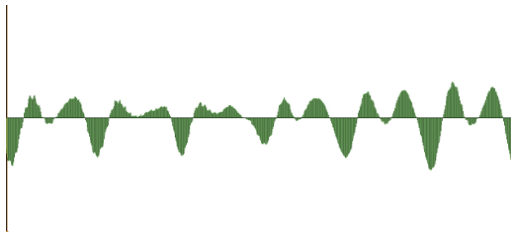
# České samohlásky - u



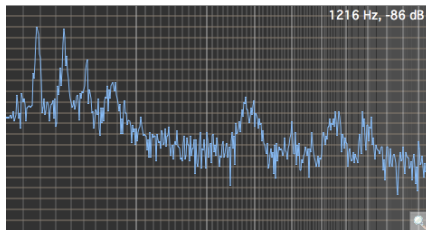
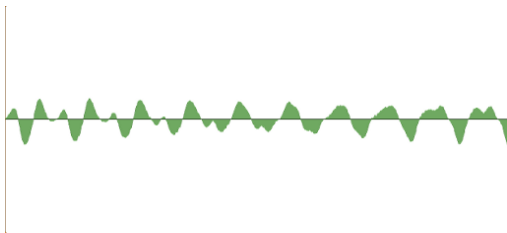
# České dvojhlásky - au



# České dvojhlásky - ou



# České dvojhlásky - eu



# Souhlásky (konsonanty)

- Zvukově dynamické děje.
- Pojem formantu ztrácí význam
  - tónový charakter mají pouze části některých souhlásek.
- Klasifikace:
  - znělé (sonorní)
  - neznělé (šumové)
  - fonetické dále podle místa a způsobu artikulace na:
    - retné - m, b, p, w, v, f
    - zubní - n, d, t, dh, th
    - dásňové - c, z, s, dz
    - patrové - ň, ě, ř, ž, š
    - závěrové (okluzívy, ražené, explozívy) - b, d, ě, g, p, ř
    - úžinové - v, z, ž, f, th, s, ř, ...

# Znělé a neznělé souhlásky

- Znělé souhlásky
  - charakteristické přítomností základního tónu
  - na vytváření se aktivně podílejí hlasivky.
- Neznělé souhlásky
  - hlasivky jsou pasivní (otevřené)
- Párové
  - neliší se artikulací, pouze znělostí
  - např. b-p, d-t. z-s, ...

# Porovnání párových souhlásek (waveform)

Figure : Souhláska s

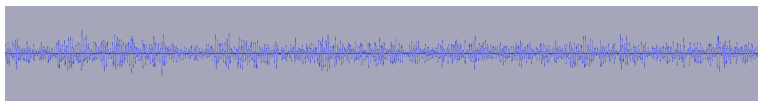
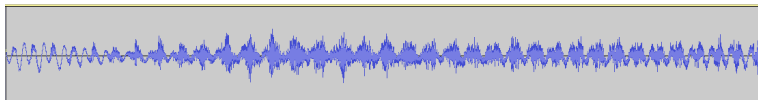


Figure : Souhláska z



# Porovnání párových souhlásek (spektrum)

Figure : Spektrum souhlásky s

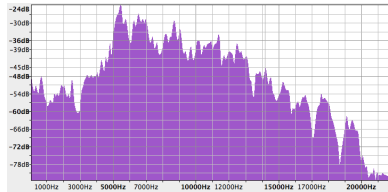
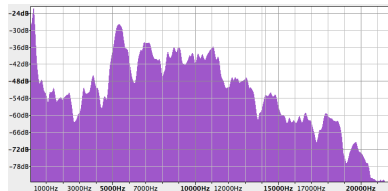


Figure : Spektrum souhlásky z



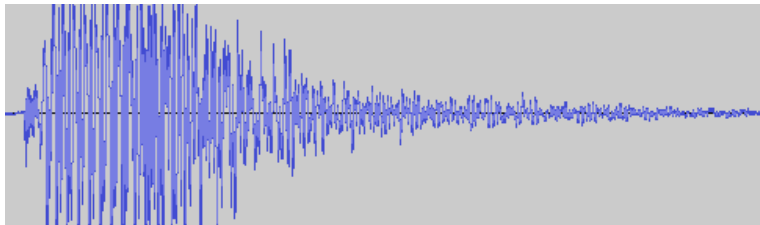


# Podle způsobu tvorby

- Okluzívy
  - závěrové souhlásky
  - vytvořena překážka výdechovému proudu vzduchu:
    - jazyk
    - zuby
    - rty
  - (p, b), (t, d), (ť, ď). (k, g), m, n, ň
- Frikativy
  - úžinové
  - zúžení výdechové cesty při artikulaci
  - (s, z), (š, ž), (f, v), (ch, h), l, j, r, ř
- Semiokluzívy
  - polouzávěrové
  - vytváří se jak překážkou, tak zúžením výdechové cesty
  - c, č

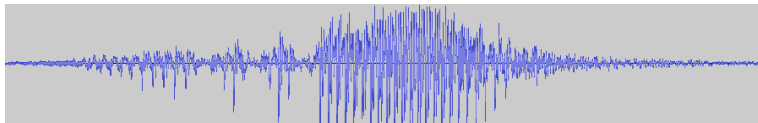
# Ukázka okluzívy

Figure : Souhláska p



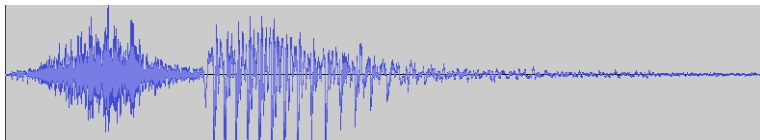
# Ukázka frikativy

Figure : Souhláska r



# Ukázka semiokluzívy

Figure : Souhláska c



# Koartikulace

- Modifikace fonému v řečovém kontextu.
- Nutnost přenastavit řečový trakt na další foném.
- Způsobuje problémy při:
  - syntéze řeči
  - rozpoznávání řeči.

# Ukázka vlivu koartikulace - původní fonémy

Figure : Souhláska p

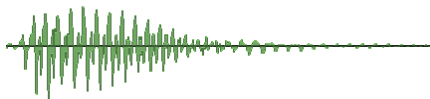
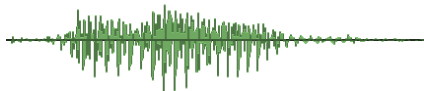


Figure : Samohláska a



## Ukázka vlivu koartikulace (2.)

Figure : Spojené hlásky p a a

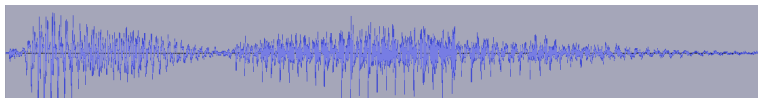
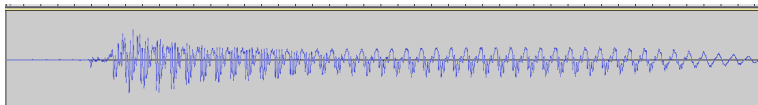


Figure : Slabika pa



# Digitalizace zvuku

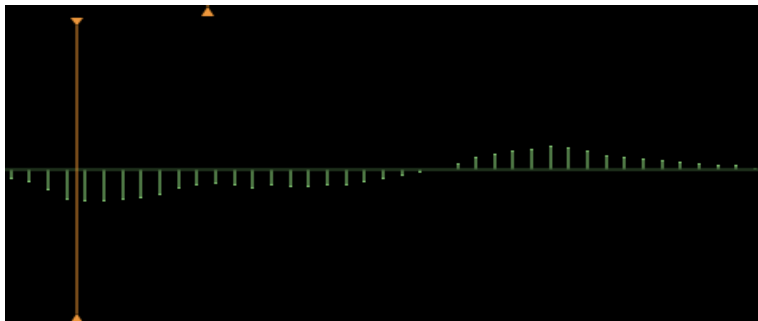
- Cíl - převod spojitého signálu na posloupnost digitálních hodnot vhodných pro uchování v počítači.
- Postup digitalizace:
  - 1 Vzorkování - převod reálných vstupních hodnot na posloupnost diskrétních reálných čísel.
  - 2 Kvantizace - převod posloupnosti reálných čísel na posloupnost celých čísel.
  - 3 Kódování - způsob uložení a kódování posloupnosti celočíselných hodnot získaných v kroku 2.



# Vzorkování

- Transformace spojitého časové závislého signálu  $s(t)$  na časově diskrétní posloupnost  $s_n(T) = 0, 1, 2, \dots$ 
  - $T$  - perioda vzorkování.
  - Pokud nemá dojít ke ztrátě informace, musí být vzorkovací frekvence aspoň dvojnásobkem nejvyšší frekvence, která je signálu obsažena.
- Po čase  $T$  je sejmuta a dána na výstup (ke kvantizaci) hodnota ze vstupního snímače.
  - většinou okamžitá úroveň napětí nebo proudu na vstupu.
- Oblasti použití
  - digitální zpracování zvuku
    - audio CD
    - mp3 - navíc použita ztrátová komprese
    - miniDisc - navíc použita ztrátová komprese ATRAC
    - DAT
    - ...
  - digitální zpracování signálu obecně (digitalizace dat z různých analogových měřících zařízení, digitální zpracování obrazu, ...)

# Ukázka digitalizovaného signálu



# Shannonův vzorkovací teorém

- Analogový signál  $s(t)$  lze rekonstruovat z hodnot vzorků  $s_n(T)$  následovně:

$$s(t) = \sum_{n=-\infty}^{\infty} s_n(T) \frac{\sin(\pi(\frac{t}{T} - n))}{\pi(\frac{t}{T} - n)}$$

právě tehdy když je vzorkovací frekvence alespoň dvojnásobkem nejvyšší frekvence obsažené ve vstupním signálu.

- Důsledky:
  - Vzorkovací frekvence by měla být alespoň dvojnásobkem nejvyšší frekvence vstupního signálu.
  - Je-li menší dochází ke zkreslení složek vyšších frekvencí.
  - Spor příznivců a odpůrců audio CD - je 44kHz dostačující vzorkovací frekvence pro hudbu?

# Kvantizace

- Převod reálných navzorkovaných hodnot na celočíselné hodnoty.
- Počet celočíselných hodnot = počet úrovní kvantování
  - 256
  - 65 536
  - 16 777 216
- Kvantizační krok - reálný interval přiřazený kvantizované jednotce.
  - Na vstupu je signál s amplitudou 128 mA (-128 - 127 mA).
  - 8bitová kvantizace - 256 kvantizačních úrovní
  - kvantizační krok =  $\frac{256[mA]}{256} = 1[mA]$ .
- Běžně používané kvantizace – 8, 16, 24 bitů.
- Realizováno pomocí A/D převodníků
  - součást zvukových karet
  - mobilních telefonů
  - ...

# Běžně používané parametry digitalizace zvuku

- Vzorkovací frekvence:
  - 8 kHz - telefonní kvalita
  - 16 kHz - běžná řeč
  - 22 kHz - rozhlasová kvalita
  - 44 kHz - audio CD
  - 48 kHz - DVD
- Kvantizace:
  - 8 bitů
  - 16 bitů
  - 24 bitů
- Počet audio kanálů
  - 1
  - 2
  - 4
  - 6 (5.1, 5 směrových kanálů + basy)

# Způsoby kódování signálu

- PCM - přímé ukládání hodnot získaných kvantizací.
  - Výhody – jednoduché na zpracování, nedochází k další ztrátě informací.
  - Nevýhody:
    - často malé rozdíly mezi hodnotami sousedních vzorků – značná redundance dat,
    - konstantní hodnota kvantizačního kroku (závisí na parametrech AD převodníku) – v případě malé amplitudy vstupního signálu – ztráta informace (signál nepřekročí kvantizační krok), v případě velké amplitudy – hodnota překročí rozsah – zkreslení signálu. Oba případy brání kvalitní rekonstrukci původního signálu.

# Kódování průběhu vlny

## Řešení nevýhod PCM

- Diferenční PCM
  - Uchovávání rozdílů sousedních vzorků místo uchovávání jejich hodnot.
  - Hodnota rozdílu bývá podstatně menší než hodnota vzorku - lze uchovat pomocí méně bitů.
- Adaptivní PCM
  - Kvantizační krok se určuje na základě amplitudy vstupního signálu.