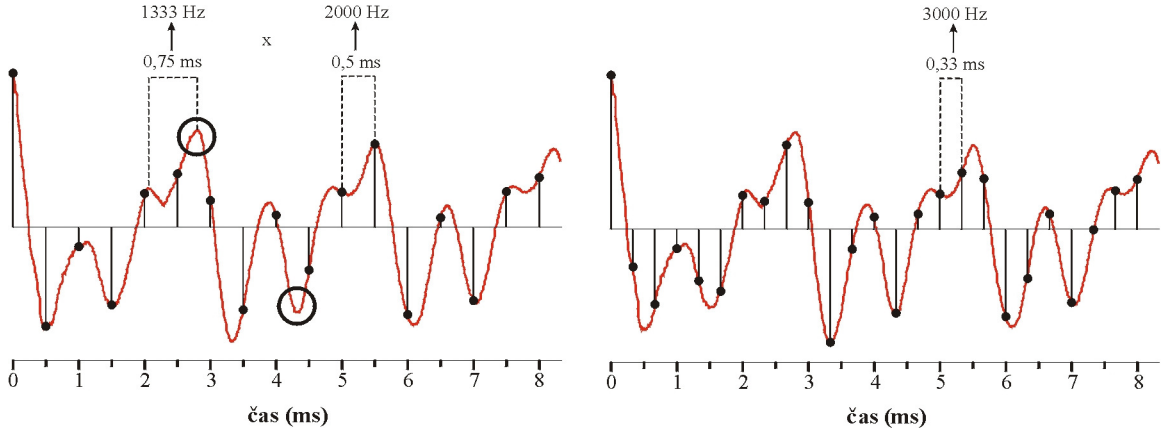
7. Analýza řečového signálu

1. převod analogového signálu na digitální

a) <u>vzorkování</u> - vzorkovací frekvence (f_s) udává, kolikrát za 1 sekundu provádíme měření

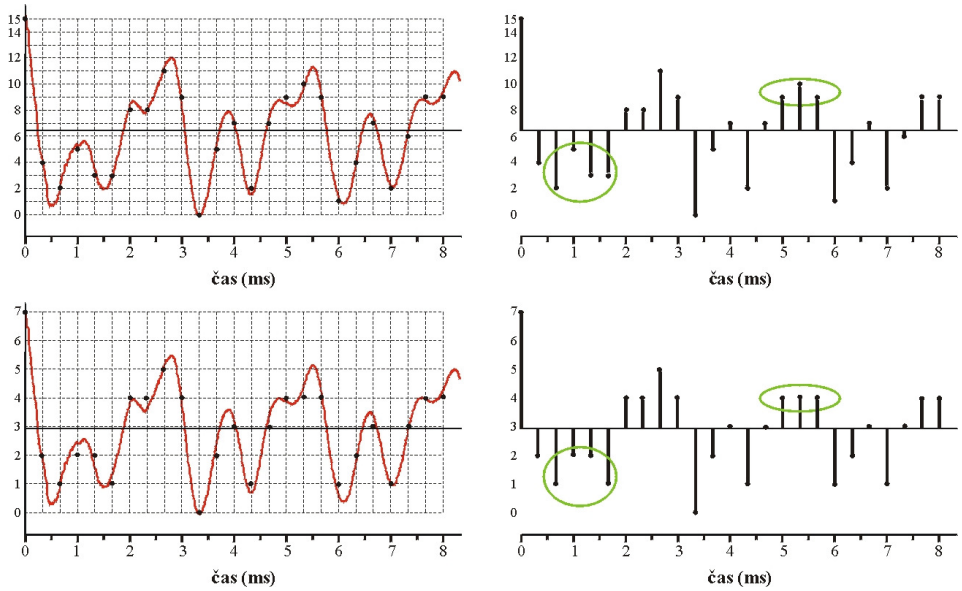
- CD je vzorkováno na 44,1 kHz, řeč mezi 16 a 32 kHz

- **vzorkovací teorém** - f_s musí být vyšší než dvojnásobek nejvyšší frekvenční složky signálu
  - na jednu periodu nejvyšší harmonické musejí připadnout alespoň dva vzorky
  - ilustrace - horní vlna má nejvyšší frekvenční složku 1333 Hz a vzorkovací frekvenci 1000 Hz
    - některé vrcholy nebo propady nejsou při vzorkování zohledněné ➔ **aliasování**



b) <u>kvantizace</u> - velikost intervalů amplitudy - čím jsou intervaly menší, tím je měření přesnější

- ilustrace ukazuje rozdíl mezi rozlišením na 16 a 8 možných hodnot amplitudy

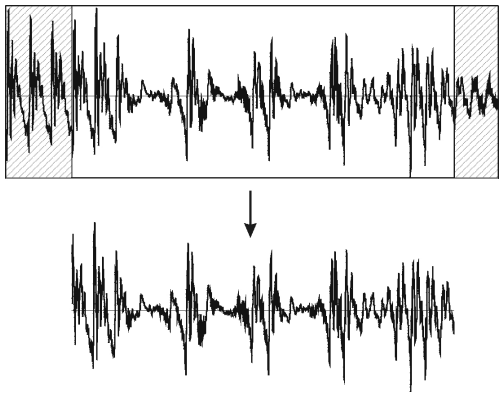


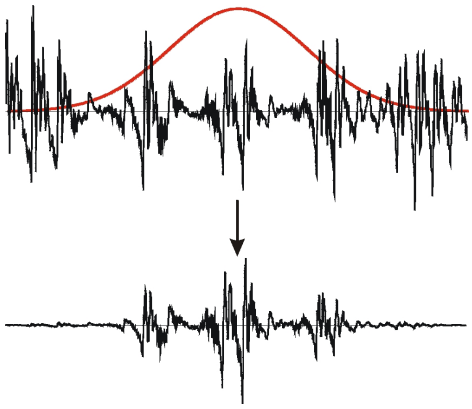
- každý vzorek je kvantizován pomocí binárních jednotek **bitů**
- čísla 0 a 1 lze uložit do jednoho bitu, čísla 0-3 potřebují dva bity (4 hodnoty), čísla 0-7 tři bity
- **bitové rozlišení** udává, na kolik intervalů celý rozsah amplitudy vzorkujeme
- 8bitové rozlišení znamená 256 možných intervalů, 16bitové rozlišení znamená 65536 intervalů

2. okna

- když se chceme zaměřit na určitou část signálu, používáme okna, která „překrýváme“ přes vlnu
- **obdélníkové okno** - základní typ okna, jednoduše vyřízneme část signálu
  - pro počítač to znamená, že signál náhle začíná a náhle končí ➔ silné vysokofrekvenční složky

- proto se používá různých algoritmů pro stanovení jiných tvarů oken (vyhlazení okrajových částí)
- na okno „přiložíme“ křivku, která stoupá od 0 k 1 a zase klesá zpět na 0, a vlnu jí vynásobíme
- **Gaussovo okno** - je nejvýhodnější, protože jediné dokáže blokovat „postranní laloky“
  - dříve se kvůli úspornosti používala jiná okna (např. Hammingovo, Hanningovo, Welchovo)



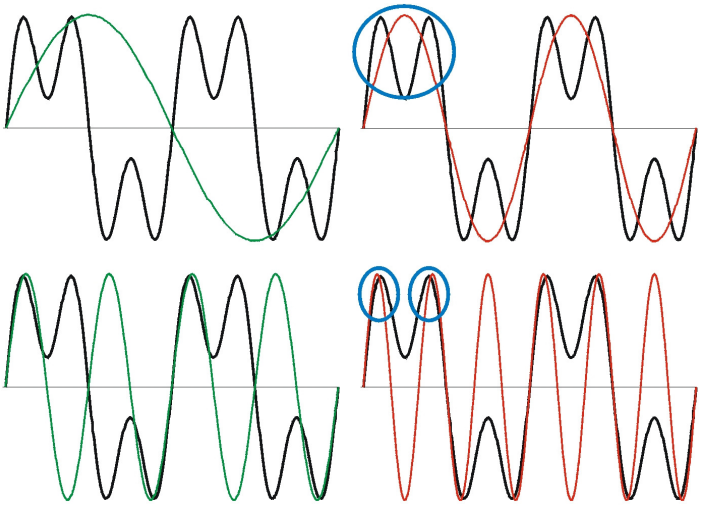


3. Fourierova transformace

- tři hlavní problémy aplikace Fourierova teorému při analýze řeči
  1) řečové zvuky jsou kvaziperiodické → proto používáme okno (jeho délka se stává periodou)
  2) digitální analýza užívá diskrétní signál → transformace **DFT** (Discrete Fourier Transform)
  3) složité výpočty → algoritmus DFT nazvaný **FFT** (Fast Fourier Transform)

- **podstata FFT** - počítá se **korelace** sinusových a kosinusových složek se složenou vlnou
  - vysoká korelace znamená, že složená vlna a daná sinusovka stoupají a klesají spolu

- ilustrace - složená vlna o složkách 100 a 300 Hz
  - u sinusovek 50 a 200 Hz rozdílné průběhy (nízká korelace), u 100 a 300 Hz vysoká korelace

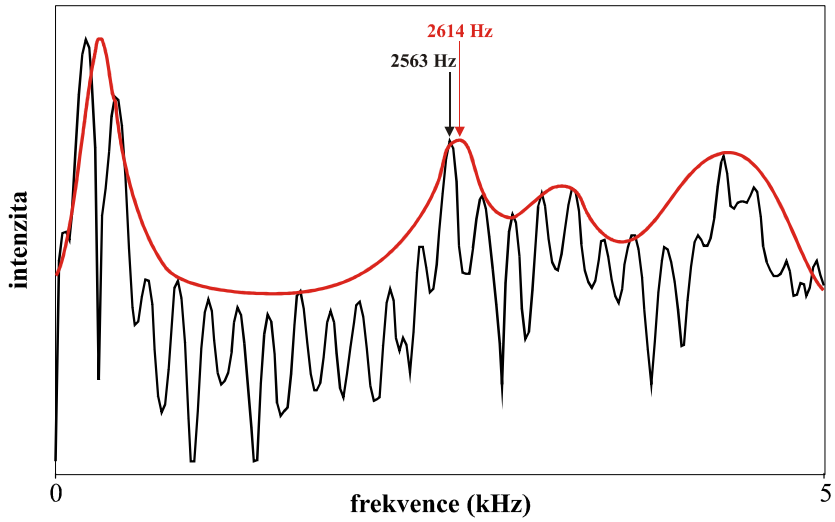


4. další spektrální reprezentace (získání vyhlazeného spektra)

- **lineární predikce (LPC)** - daný vzorek chápán jako lineární kombinace vzorků předcházejících

$$x(n) = a_1[x(n-1)] + a_2[x(n-2)] + a_3[x(n-3)] + \ldots - e(n)$$

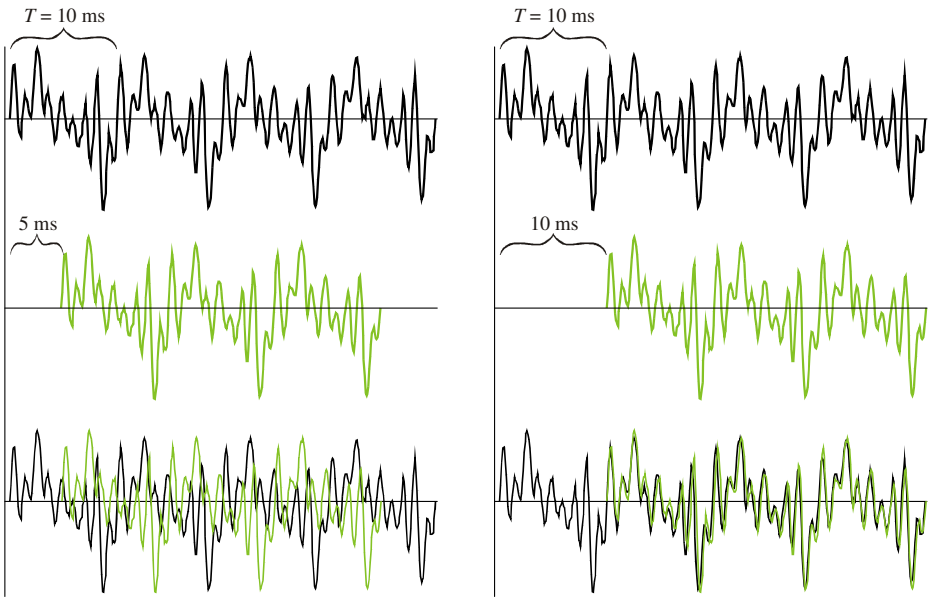
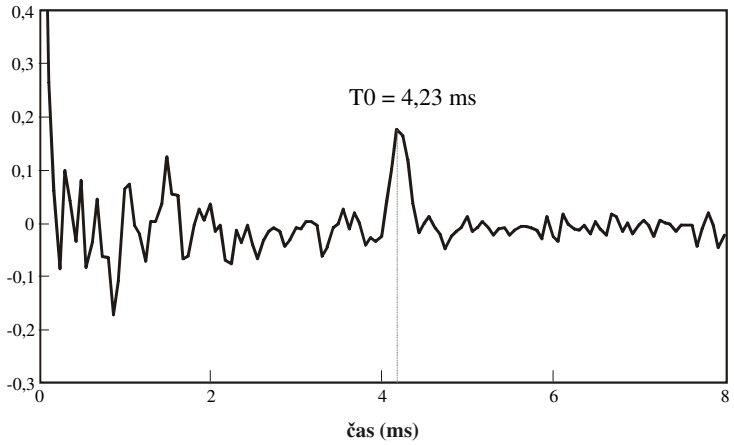
  - koeficienty LPC (a_n) reprezentují rezonance vokálního traktu (matematický model)
  - použití správného počtu předcházejících vzorků (= řád LP) aproximuje tvar spektra
  - prokládání spektra probíhá pomocí polynomických funkcí (5 formantů = 10.-11. řád LPC)



- **kepstrum** - na spektrum (výsledek FT) znovu aplikujeme FT
  - odděluje od sebe dvě hlavní složky spektra, hlasivkový signál a filtr vokálního traktu
  - odfiltrování „zdrojové“ části a následné provedení inverzní FT také ukáže vyhlazené spektrum

5. extrakce základní frekvence

- **autokorelace** - porovnáváme vlnu s kopií sama sebe s určitým fázovým posunem
  - cílem je najít takový fázový posun, aby korelace byla co nejvyšší ➔ to odpovídá jedné periodě
  - na základě fázového posunu vlny vyhledá intervaly mezi hlasivkovými pulzy
  - autokorelace může být nepřesná (poloviční nebo dvojnásobná f0, součet periody f0 + F1)



- **kepstrum** - spektrum ze zvukové vlny oddělí jednotlivé složky, které vznikly jejich superpozicí
  - na spektrum znovu aplikujeme Fourierovu transformaci a oddělíme od sebe jeho součásti
    ➔ zdrojový (hlasivkový) signál a filtr vokálního traktu
  - jakákoli periodicita bude ukázána jako oddělená složka
  - jde vlastně o spektrum spektra, přesněji o Fourierovu transformaci logaritmu spektra
  - z frekvenční domény spektra se opět vracíme do časové domény, jednotkou jsou milisekundy
  - ilustrace ukazuje schématický výsledek kepstrální analýzy vokálu /e/
    - nejvýraznější hrot odpovídá základní periodě (T0)