

Multimédia

Gregor Rozinaj, Renata Rybárová, Ivan Minárik, Juraj Kačur

Autori: Gregor Rozinaj, Renata Rybárová, Ivan Minárik, Juraj Kačur
Názov diela: Multimédiá
Vydalo: České vysoké učení technické v Praze
Spracoval(a): Fakulta elektrotechnická
Kontaktná adresa: Technická 2, Praha 6, Česká republika
Tel.: +420 2 2435 2084
Tlač: (iba elektronická)
Počet strán: 131
Vydanie: 1.

ISBN 978-80-01-05312-6

Oponent: Jozef Čepko

Innovative Methodology for Promising VET Areas
<http://improvet.cvut.cz>



**Program
celoživotného
vzdelávania**

Tento projekt bol financovaný s podporou Európskej Komisie.
Táto publikácia (dokument) reprezentuje výlučne názor autora a Komisia nezodpovedá za akékoľvek použitie informácií obsiahnutých v tejto publikácii (dokumente).

VYSVETLIVKY



Definícia



Zaujímavosť



Poznámka



Príklad



Zhrnutie



Výhody



Nevýhody

ANOTÁCIA

Tento výučbový modul obsahuje základné informácie o multimédiách. V prvej časti sú predstavené základné znalosti o analógových a digitálnych signáloch, digitalizácii signálov, časovej a frekvenčnej reprezentácii signálov. Ďalšia časť obsahuje informácie o spracovaní signálov, analógových a digitálnych technológiách, filtroch a popis komunikačného kanála. Na záver sa modul venuje kompresným technikám audio a video signálov a samotnej aplikácii teórie spracovania signálov v multimédiách: syntéze reči, rozpoznání reči a obrazu.

CIELE

Hlavným cieľom tohto výučbového kurzu je oboznámiť študentov so základmi spracovania signálov, najmä s tými technikami, ktoré sa využívajú pri spracovaní multimédií. Študentovi sú predstavené základné princípy Fourierovej transformácie, digitálnych filtrov, lineárnych systémov, kompresných techník a praktické použitie všetkých vedomostí v modernej vede.

LITERATÚRA

- [1] DÚHA, J., GALAJDA, P., KOTULIAK, I., LEVICKÝ, D., MARCHEVSKÝ, S., MIKÓCZY, E., PODHRADSKÝ, P. a kol. *Multimedia ICT technologies, network platforms and multimedia services*, Vydavateľstvo STU Bratislava, 2005, ISBN 80-227-2310-X.
- [2] GAMEC, J. *Spracovanie multimediálnych signálov*, publikované v rámci projektu ESF NGN, kód projektu v ITMS: 13120110126, Bratislava, 2007.
- [3] TALAFOVÁ, R. – ROZINAJ, G. – CEPKO, J. – VRABEC, J. *Multimedia SMS Reading in Mobile Phone* In: INTERNATIONAL JOURNAL of MATHEMATICS AND COMPUTERS IN SIMULATION, Issue 1, Volume 1, ISSN 1998-0159, 2007.
- [4] KOTULIAKOVÁ, J. – ROZINAJ, G. *Číslicové spracovanie signálov I*, FABER Bratislava, 1996.
- [5] KOTULIAKOVÁ, J. – ROZINAJ, G. – POLEC, J. – PODHRADSKÝ, P. a kolektív. *Číslicové spracovanie signálov II*, FABER Bratislava, 1997
- [6] HARDESTY, L. *Explained: The Discrete Fourier Transform*, 2009, <http://web.mit.edu/newsoffice/2009/explained-fourier.html> [online]
- [7] MINÁRIK, I. *Coding of audio signals at low speed*, Diploma Thesis, 2011
- [8] MARMOL, F.G., et al. *ANALYSIS: State of The Art on Identity, Security and Trust*, Deliverable D3.1, HBB-Next FP7-ICT-2011-7, 2012

- [9] DEVENTER, O. et al. *ANALYSIS: Multi-User, Multimodal & Context Aware Value Added Services*, Deliverable D5.1, HBB-Next FP7-ICT-2011-7,2012
- [10] LEVICKÝ, D., RIDZOŇ, R. *Multimédiá a multimediálne technológie*, publikované v rámci projektu ESF NGN, kód projektu v ITMS: 13120110126, Vydalo Vydavateľstvo STU v Bratislave v spolupráci s AGROGENOFOND Nitra, 2007, ISBN 978-80-227-2604-7
- [11] PODHRADSKÝ, P. *Fourierov rad a Fourierova transformácia*, publikované ako študijný materiál pre predmet inžinierskeho štúdia "Analogové a digitálne signály a systavy I", študijný program Telekomunikácie, FEI STU Bratislava, 2003

Obsah

1	Signály	8
1.1	Čo je signál?	8
1.2	Dôležité signály	12
1.3	Digitalizácia analógového signálu.....	20
1.4	Modulácia	24
2	Časová a frekvenčná reprezentácia	30
2.1	Fourierova transformácia	30
2.2	Diskrétna Fourierova transformácia	34
2.3	Spektrum	36
2.4	Ortogonálne transformácie	43
3	Analógové a digitálne technológie.....	48
3.1	Multiplexovanie.....	48
3.2	Lineárny diskretný časovo – invariantný systém	52
3.3	Filtre	59
4	Komunikačný kanál.....	69
4.1	Komunikačný kanál.....	69
5	Metódy kompresie.....	72
5.1	Kompresia audio signálov	75
5.2	Kompresia reči	81
5.3	Kompresia statických obrazov	87
5.4	Kompresia videa.....	94
6	Spracovanie multimédií.....	104
6.1	Syntéza reči	104
6.2	Rozpoznanie obrazu	107
6.3	Animácia tváre	110
6.4	Rozpoznávanie reči	113
6.5	Multimodálne rozhranie	124
7	Príklady v Matlabe.....	127
7.1	Príklady v Matlabe	127

Pod pojmom *Multimédia* rozumieme integráciu textu, obrazu a zvuku za účelom sprostredkovania informácie. Každý z uvedených troch modalít má viacero reprezentácií. Pod textom rozumieme okrem bežného textu aj hypertext, tabuľky, webové stránky a pod. Obrazová modalita má pestrú škálu skupín ako: statický obraz, dynamický obraz, grafika, animácia, video atď. Špecifické oblasti zvuku sú hudba, reč, tóny, zvukové signalizácie apod.

Multimédia vnímame našimi zmyslami. Okrem zraku a sluchu, ktorými vnímame doteraz spomenuté modalitty, vnímame svet aj ďalšími tromi zmyslami: čuch, chuť a hmat a preto vo všeobecnosti zahŕňame do pojmu multimédia signály, ktoré pôsobia na všetkých päť zmyslov. Multimediálnu komunikáciu často ohraničujeme na komunikáciu, kde potrebujeme aspoň dva zmysly aj keď sa v multimediálnych aplikáciách často venujeme aj jednozmyslovej komunikácii.

Multimediálna aplikácia je spravidla podmienená možnosťou interakcie, t.j. človek je schopný riadiť chod multimediálnej aplikácie. Z tohto pohľadu napr. klasickú TV nepovažujeme za multimediálny signál, lebo program nevieme riadiť.

Médium je prostriedok, ktorým sa informácia vyjadruje, vníma, ukladá alebo prenáša. Multimédium je médium založené na viacerých modalitách.

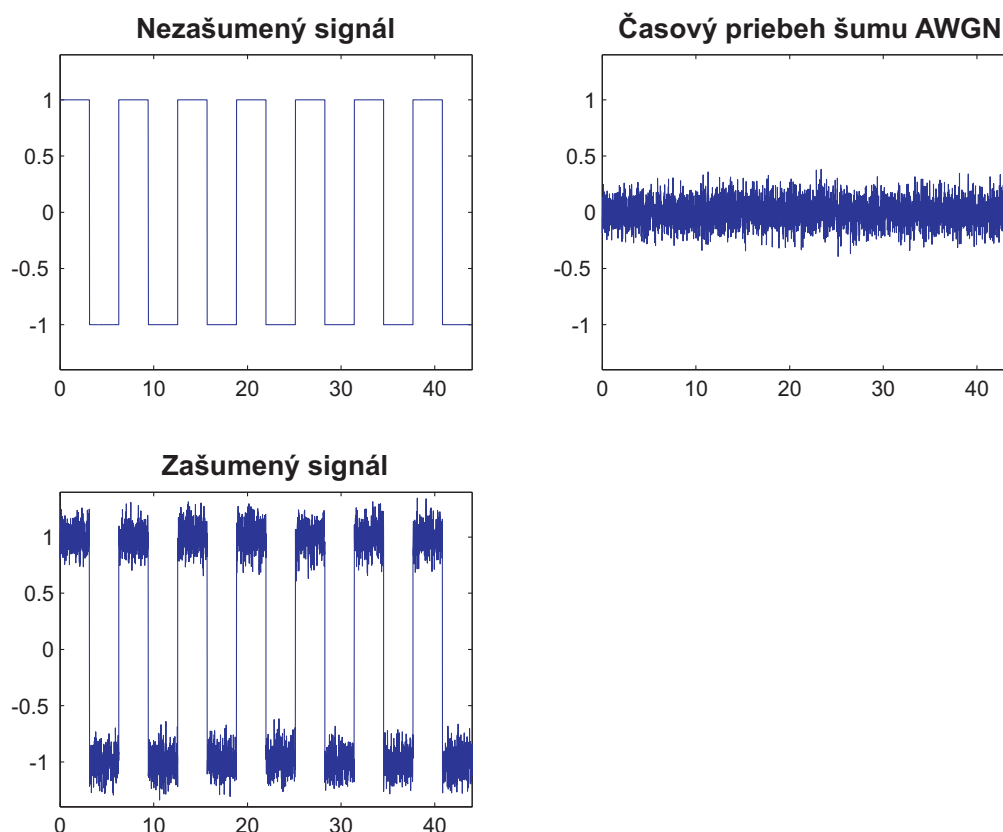
1 Signály

1.1 Čo je signál?



Význam slova signálu v tomto kurze označuje funkciu, ktorá komplexne a úplne popisuje správanie sa určitého javu.

Signál vo fyzickom svete je v zásade akákoľvek veličina, ktorá vykazuje zmenu v čase (napríklad hlas) alebo v priestore (obraz), ktorá môže poskytnúť informáciu o stave fyzického systému, alebo prenášať správu medzi pozorovateľmi. Skutočný signál je vždy zmiešaný aj so šumom.



Obr. 1.1: Skutočný signál – signál+šum

V oblasti elektrotechniky a spracovania signálu rozlišujeme dva druhy signálov – analógový a digitálny signál.

Analógový signál



Analógový signál je akýkoľvek spojitý signál, pričom môže nadobúdať teoreticky ľubovoľné hodnoty.



Analógový signál môže nadobúdať ľubovoľné hodnoty v čase a amplitúde. Reprezentuje správanie sa charakteristickej vlastnosti alebo javu v čase.

Typickým analógovým signálom je elektrický signál, alebo teplota meniac sa v čase. Analógový signál môže byť deterministický (vieme ho presne definovať, napr. matematickým vzorcom) alebo stochastický (jeho priebeh je náhodný a nevieme ho predpovedať).

Digitálny signál



Digitálny signál je reprezentovaný sekvenciou diskrétnych (zvyčajne preddefinovaných) hodnôt.



Digitálny signál môže mať iba konečné množstvo vzoriek v danom časovom úseku. Jednoduchý príklad ako dostať diskretný signál je vzorkovanie spojitého (analógového) signálu. Príkladom digitálneho signálu je napríklad teplota vzduchu meraná každých 5 minút alebo nuly a jednotky používané v počítačoch.

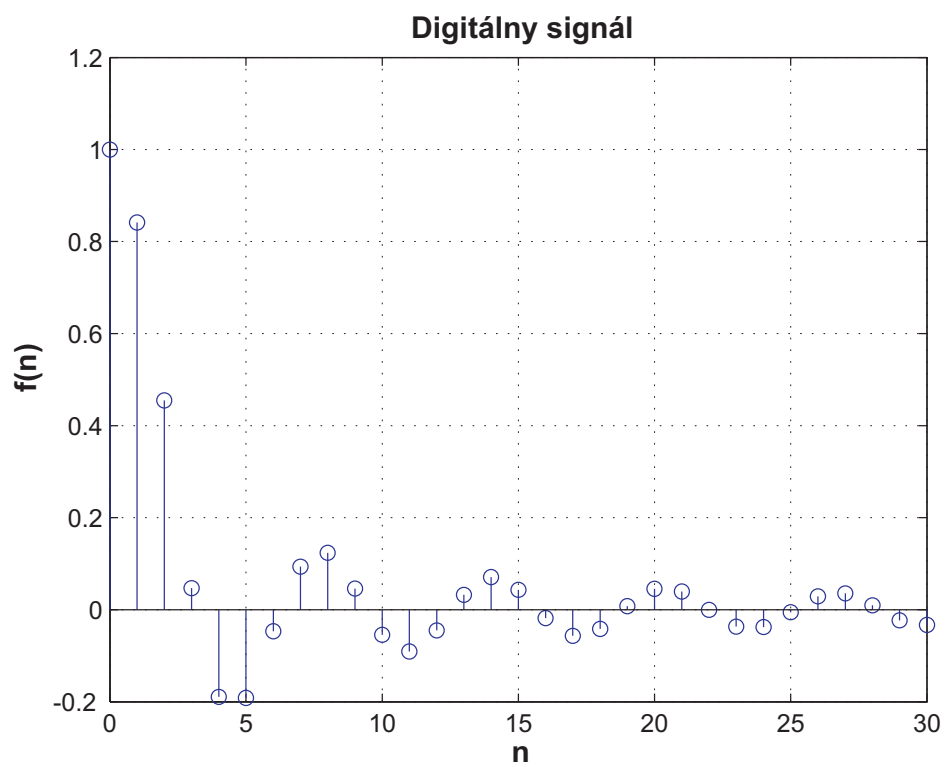
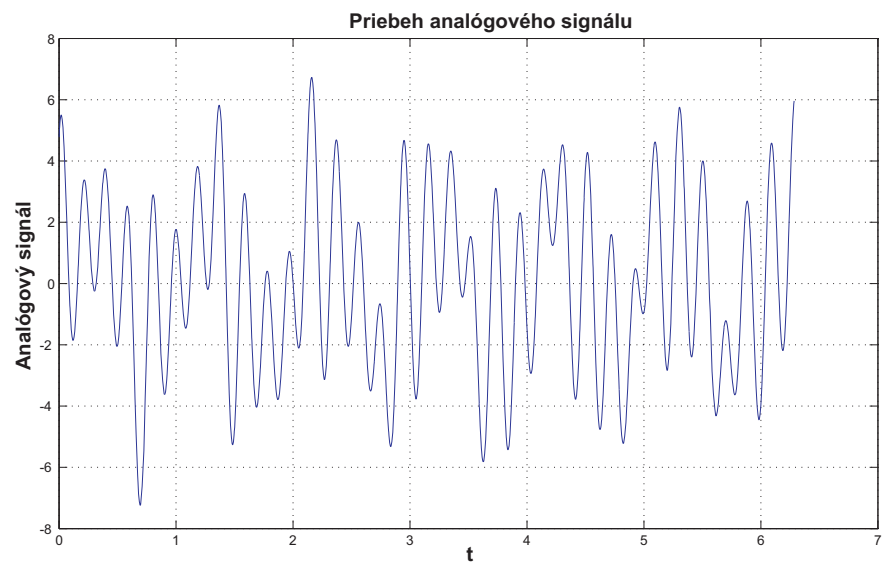
Všetky procesy v prírode sú analógové (predstavte si napríklad graf zobrazujúci teplotu meniacu sa v čase alebo rýchlosť auta v čase).

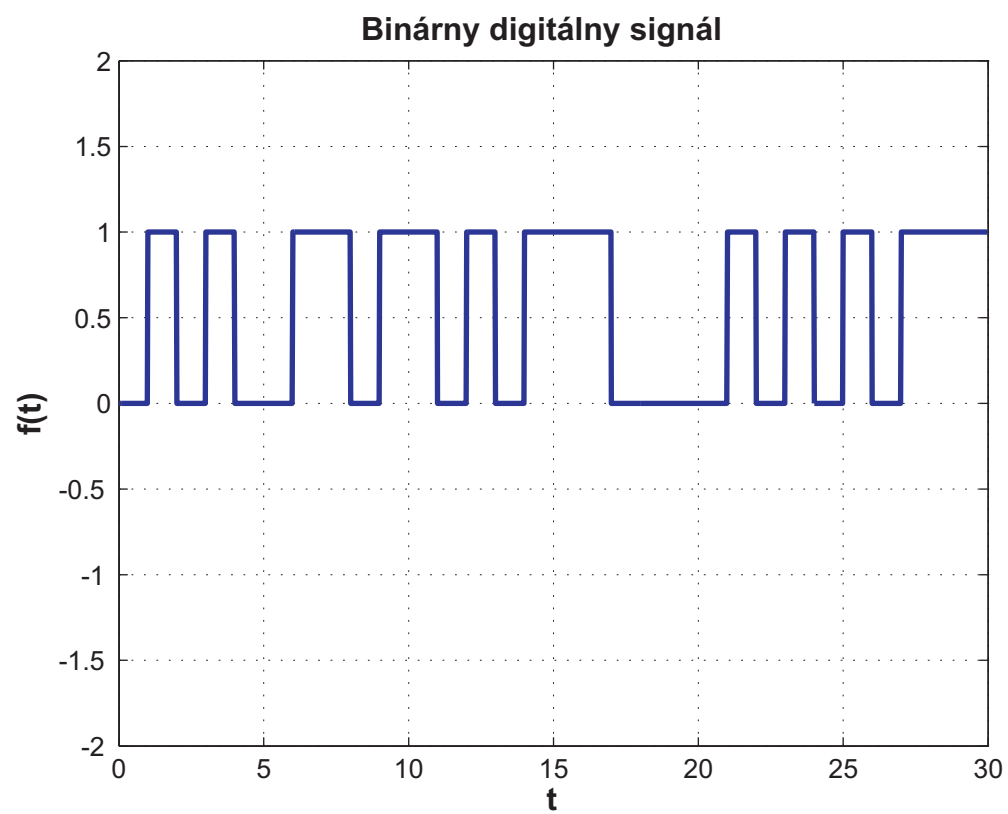
+

Hlavná výhoda spracovania analógového signálu je, že sa pri tom nestráca žiadna informácia. Avšak digitálne signály sú spracovateľné oveľa jednoduchšie (napríklad hudobné CD je digitálne – a môže byť veľmi jednoducho konvertovateľné na MP3). Digitálny signál je aj menej náchylný na degradáciu šumom.

–

Analógový signál je zložitejší na spracovanie a namáhavejšie sa s ním pracuje (predstavte si vinylovú platňu, kvalita je síce lepšia ako CD, ale skúste si pesničku z platne dať do telefónu). Výhodou analógového signálu je však to, že sa nestráca informácia ako pri digitálnom signáli cez proces vzorkovania a kvantizácie (predstavte si tabuľku hodnôt s teplotami za jeden deň – je to iba 24 číselných hodnôt).





Obr. 1.2: Príklad analógového a digitálneho signálu

1.2 Dôležité signály

V predchádzajúcej časti bol vysvetlený význam slova signál. Neboli však spomenuté prípady signálov s viac ako jednou dimenziou. V tejto časti budú predstavené najdôležitejšie signály v oblasti digitálneho spracovania signálu a multimédií.

Jednorozmerné signály



Signál, ktorý je funkciou jednej nezávislej premennej sa nazýva jednorozmerný signál. Zvyčajne je premennou čas t (napríklad $f(t)=5t$), v prípade diskretných signálov je zvyčajne miesto času diskretná premenná n , reprezentujúca číslo vzorky (napríklad $f(n) = n+1$).

V nasledovných definíciách a vzorcoch x reprezentuje množinu reálnych čísiel $\{R\}$ a n reprezentuje prirodzené čísla $\{N\}$.

Diracova delta funkcia δ je všeobecná funkcia reálnych čísiel, kde každá hodnota okrem nuly je rovná nule. Delta funkcia sa niekedy predstavuje ako nekonečne vysoká a nekonečne tenká čiara vychádzajúca z nuly. Ak by sme zráтали plochu, ktorú tento impulz zaberá, dostaneme číslo 1. V oblasti spracovania signálu je táto funkcia často známa aj pod menom jednotkový impulz.

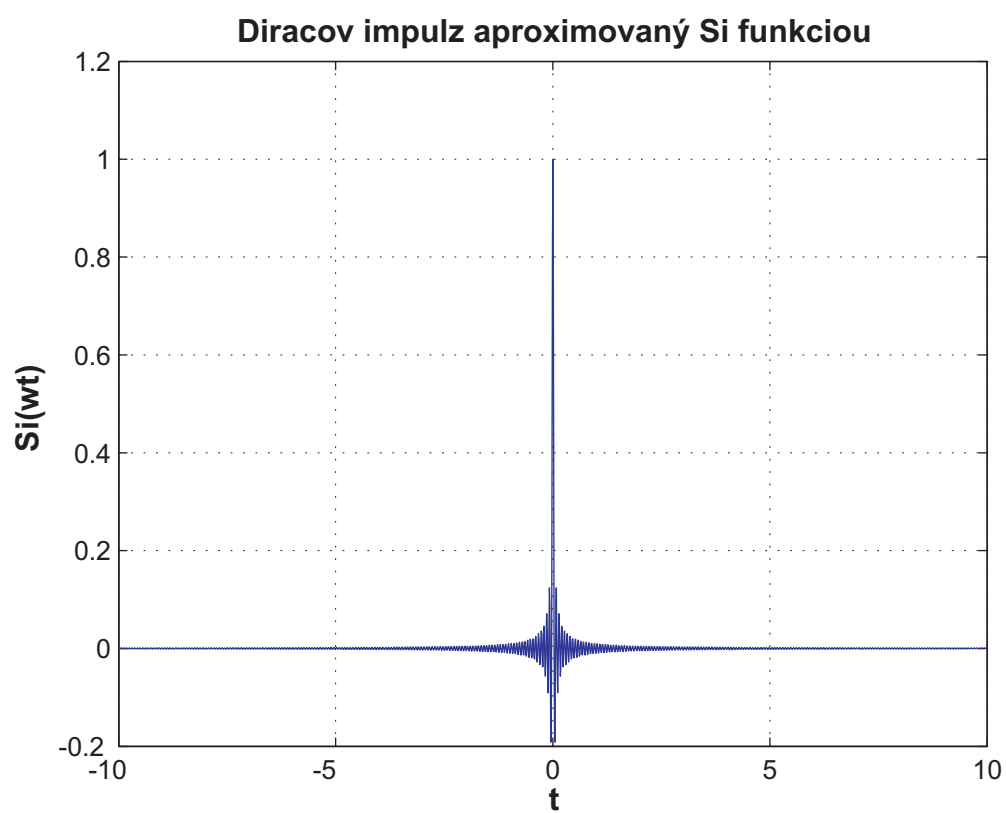
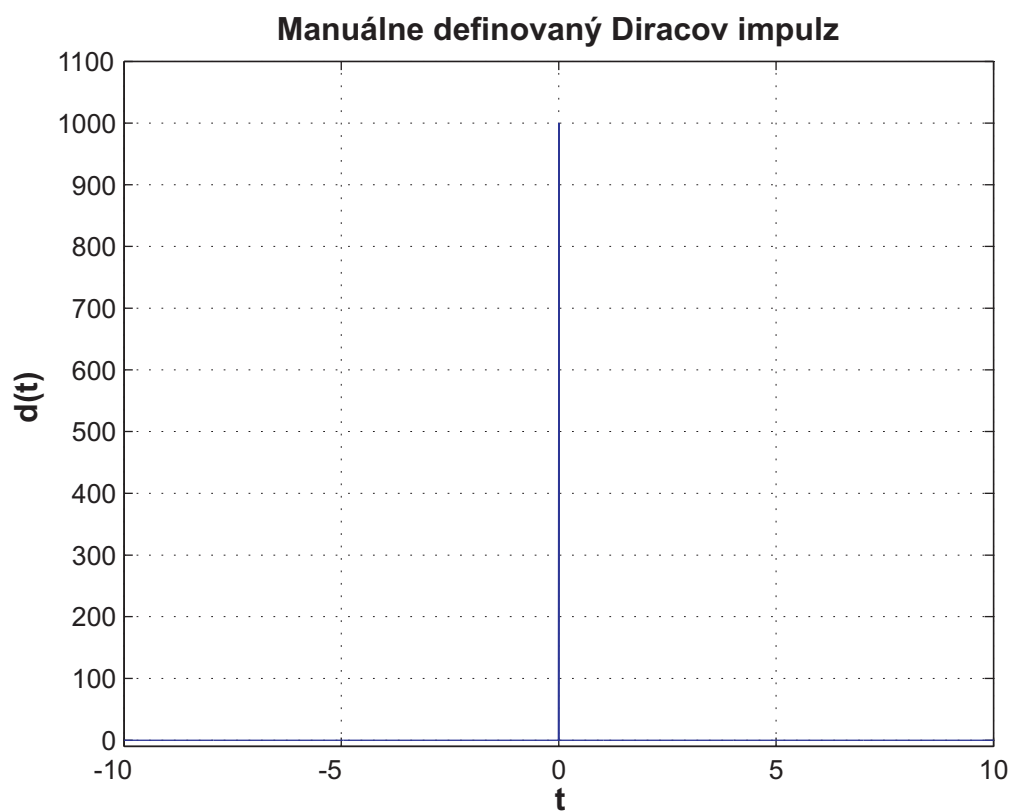
Matematická definícia:

$$\delta(x) = \begin{cases} +\infty & x = 0 \\ 0 & x \neq 0 \end{cases}$$

niekedy postačí aj obmedzená definícia

$$\int_{-\infty}^{\infty} \delta(x) dx = 1.$$

Na obrázku nižšie (Obr. 1.3) je zobrazená ideálna a aproximovaná Diracova delta funkcia. Aproximovaná Diracova funkcia je zobrazená pre lepšie vysvetlenie, ako môžeme v reálnom svete získať delta funkciu.

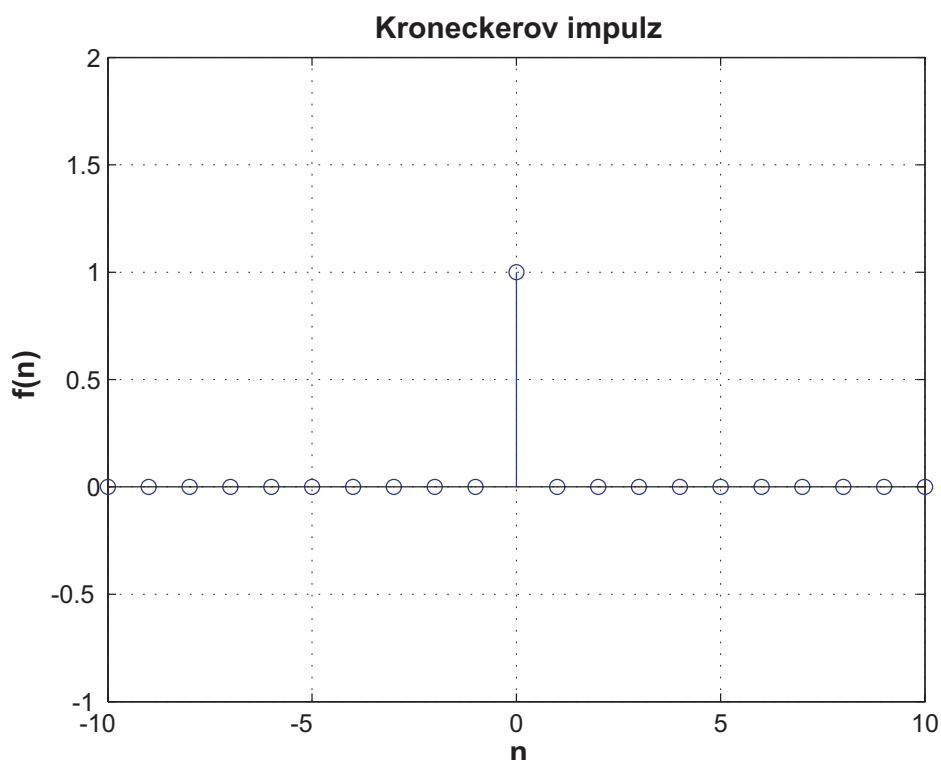


Obr. 1.3: Diracova delta funkcia – ideálna a približená funkciou $sinc()$.

V diskretnom svete je ekvivalentom Diracovej delta funkcie Kroneckerova delta funkcia. Vo svete digitálneho spracovania signálu je často známa pod jednoduchším menom impulz, alebo jednotkový impulz. Výstup, alebo odpoveď, ktorú dostaneme od digitálneho elementu (funkcie, súčiastky), nazývame odpoveď na jednotkový impulz.

Matematická definícia:

$$\delta(n) = \begin{cases} 1 & n = 0 \\ 0 & n \neq 0 \end{cases}$$



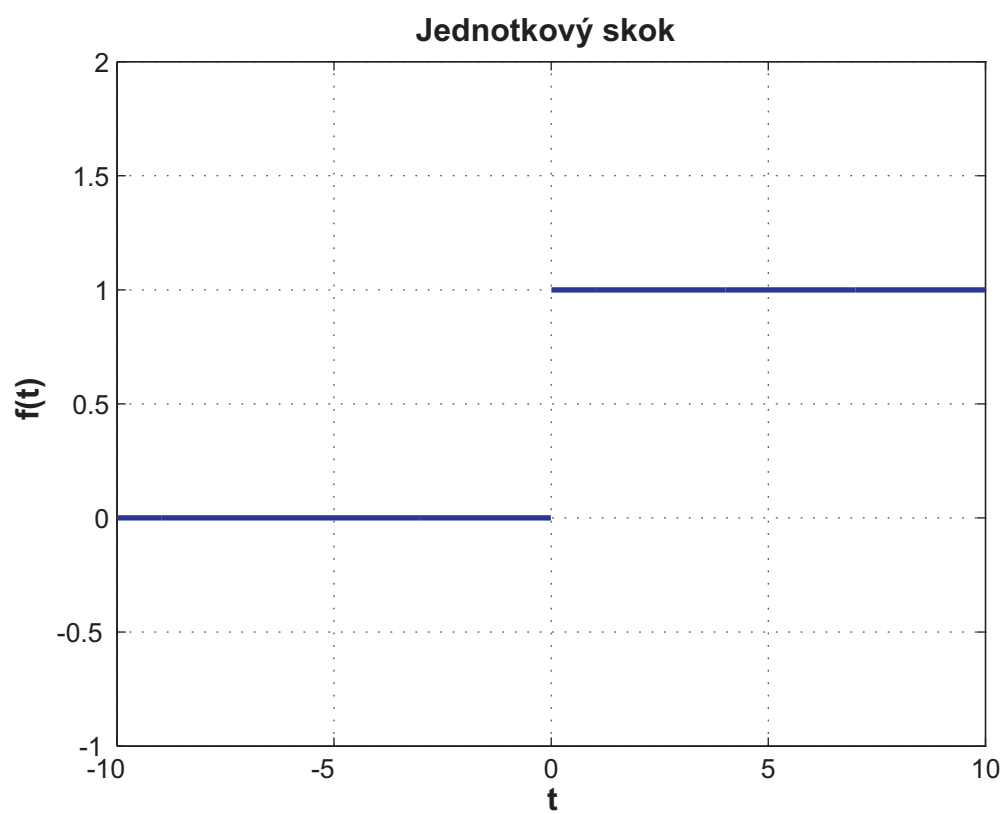
Obr. 1.4: Kroneckerova delta funkcia

Funkcia jednotkového skoku, zvyčajne označená ako u , je nespojitou funkciou, ktorej hodnota je nulová pre všetky záporné hodnoty premennej a jedna pre všetky pozitívne hodnoty. Funkcia reprezentuje signál, ktorý sa zapne v špecifikovanom čase a zostane donekonečna zapnutý. Funkcia jednotkového skoku je integrálom Diracovej delta funkcie.

$$u(x) = \int_{-\infty}^x \delta(s) ds$$

Matematická definícia:

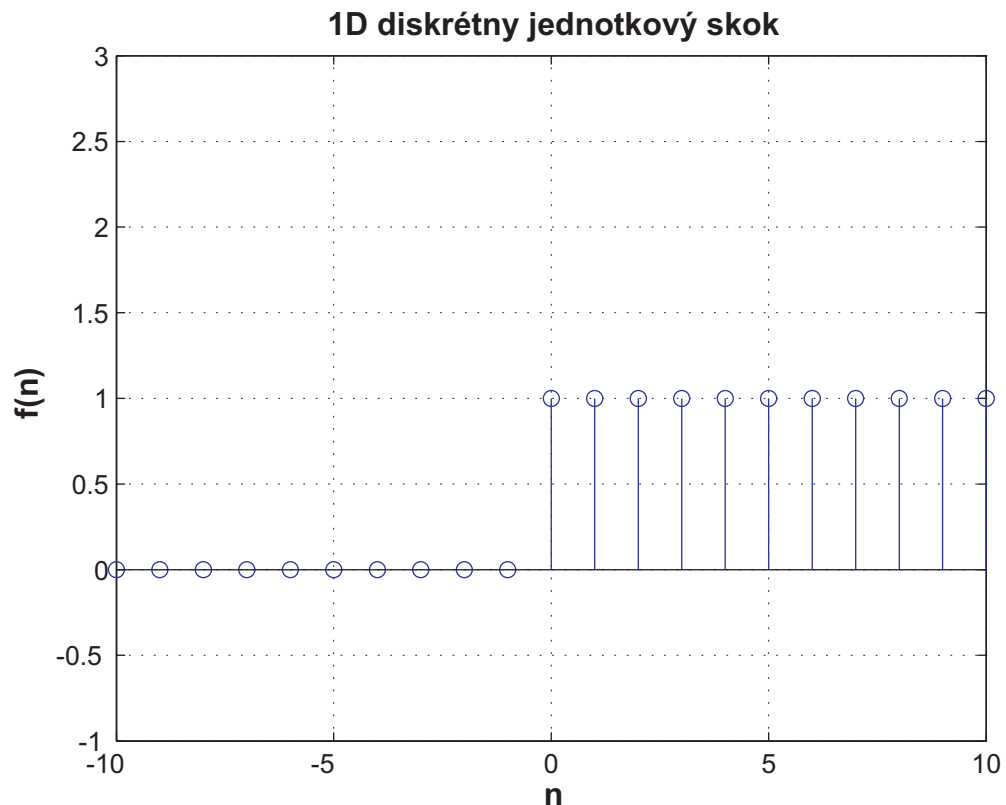
$$u(x) = \begin{cases} 0 & x < 0 \\ 1 & x \geq 0 \end{cases}$$



Obr. 1.5: Jednotkový skok (spojitý)

Diskrétna forma jednotkovej funkcie:

$$u(n) = \begin{cases} 0 & n < 0 \\ 1 & n \geq 0 \end{cases}$$



Obr. 1.6: Jednotkový skok (diskrétny)

Ďalšou množinou signálov sú periodické signály. Periodické funkcie (ktoré popisujú opakujúce sa signály) sú funkcie, ktoré opakujú svoje hodnoty v pravidelných intervaloch. Patria sem všetky trigonometrické funkcie (sínus, kosínus, tangens, kotangens – všetky majú periódu 2π). Ak periódu označíme P , potom matematická definícia periodickej funkcie je:

$$f(x) = f(x + P)$$

Dvojrozmerné signály

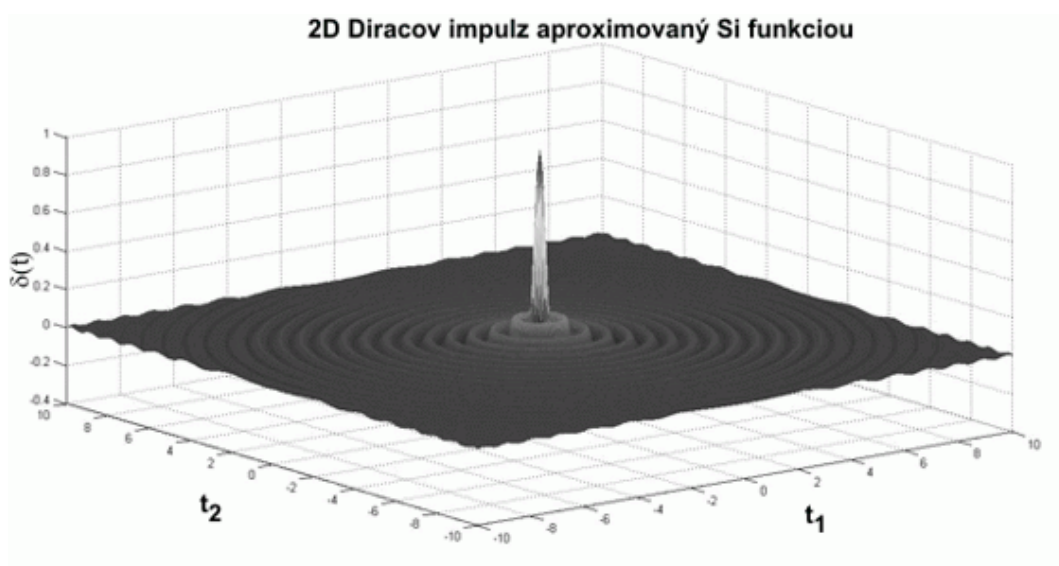
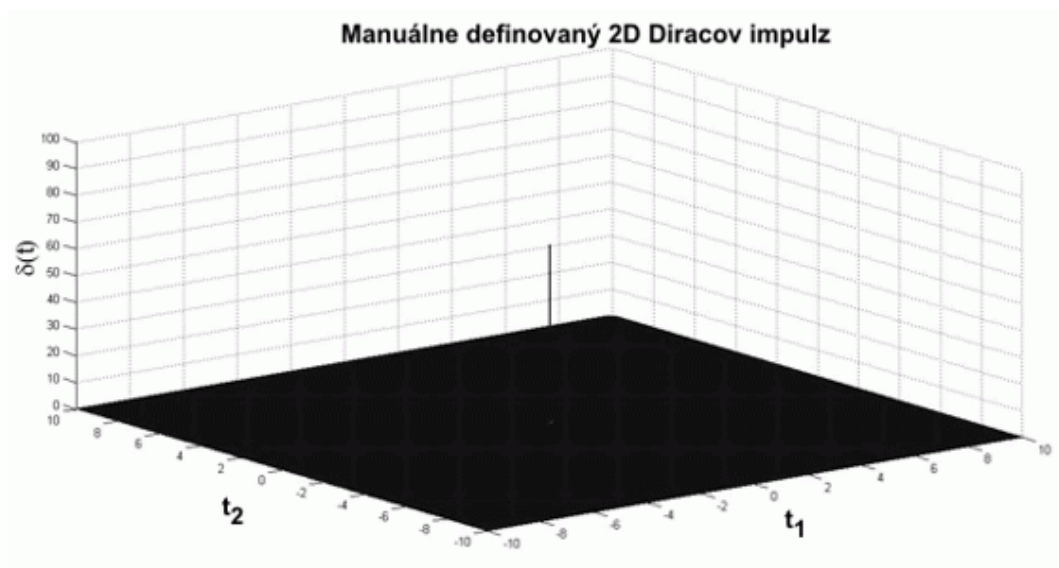


Signál, ktorý je funkciou dvoch nezávislých premenných, sa volá dvojrozmerný signál. Typickým príkladom dvojrozmerného signálu je obrázok. Obrázok sa skladá zo svetelnej a luminiscenčnej zložky. Aj 2D obraz môže byť spojitý, napríklad klasická fotografia alebo obraz, alebo diskretný, napríklad digitálna fotografia.

Všetky dôležité signály uvedené pre jednorozmerné signály je možné definovať aj pre dva rozmery. Uvádzame iba matematické definície.

Diracova delta funkcia

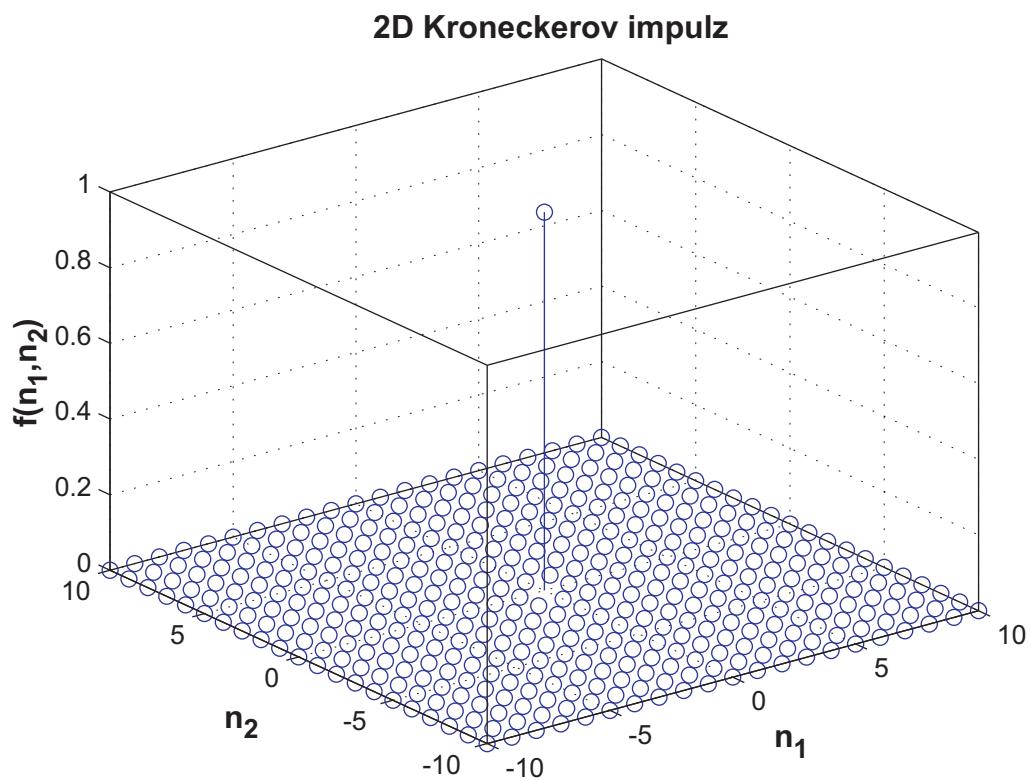
$$\delta(x_1, x_2) = \begin{cases} +\infty & x_1, x_2 = 0 \\ 0 & x_1, x_2 \neq 0 \end{cases}$$



Obr. 1.7: 2D Diracova delta funkcia – ideálna a aproximovaná funkciou *sinc()*.

Kroneckerova delta funkcia

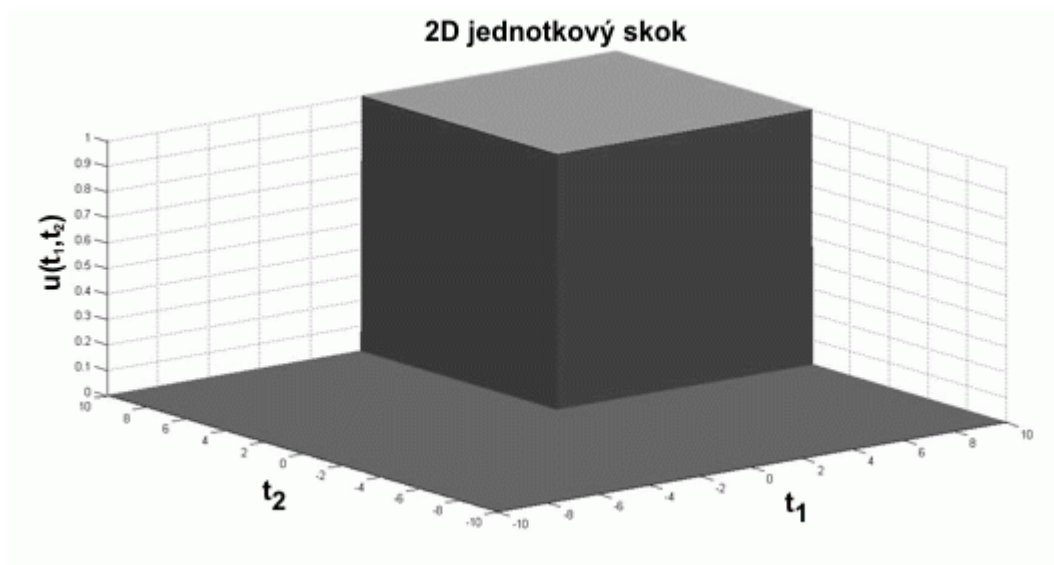
$$\delta(n_1, n_2) = \begin{cases} 1 & n_1, n_2 = 0 \\ 0 & n_1, n_2 \neq 0 \end{cases}$$



Obr. 1.8: 2D Kroneckerova delta

2D jednotkový skok (spojitý)

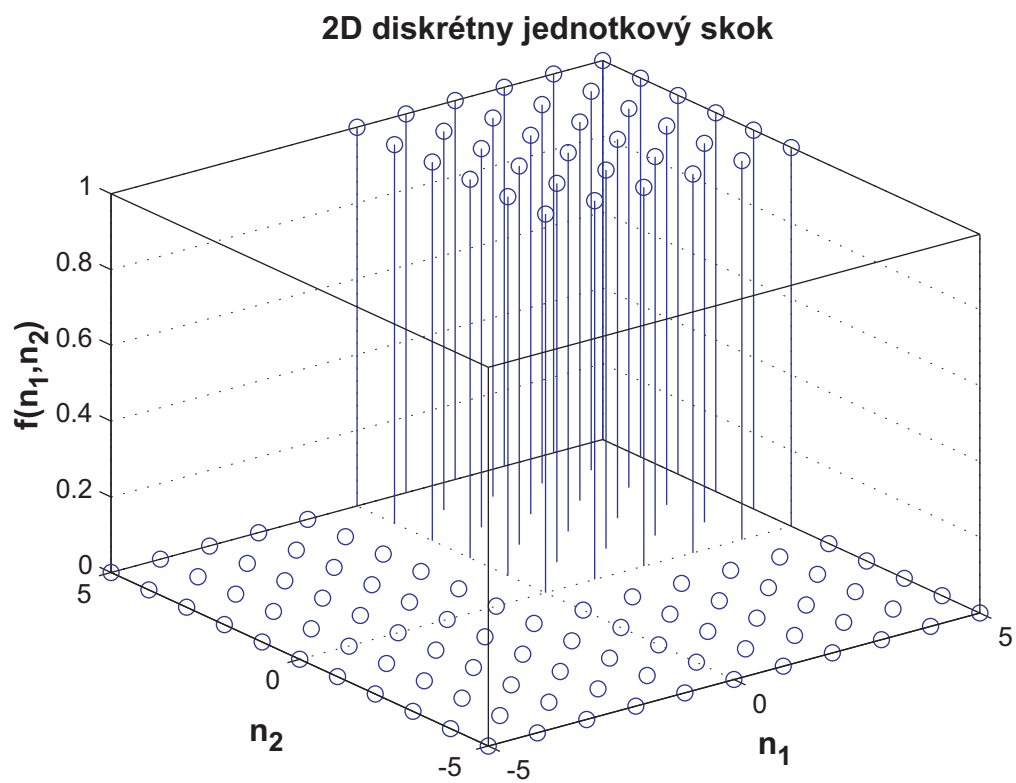
$$u(x_1, x_2) = \begin{cases} 0 & x_1, x_2 < 0 \\ 1 & x_1, x_2 \geq 0 \end{cases}$$



Obr. 1.9: 2D Jednotkový skok (spojitý)

Jednotkový skok (diskrétny)

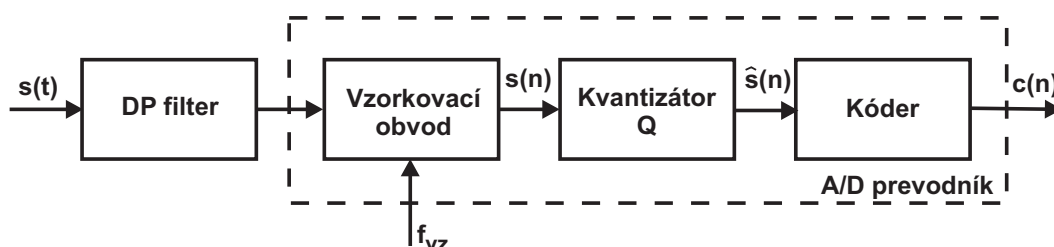
$$u(n_1, n_2) = \begin{cases} 0 & n_1, n_2 < 0 \\ 1 & n_1, n_2 \geq 0 \end{cases}$$



Obr. 1.10: 2D Jednotkový skok (diskrétny)

1.3 Digitalizácia analógového signálu

Digitalizácia analógového signálu je proces, v ktorom je analógový signál (reprezentujúci napríklad zvuk, video, obraz,...) transformovaný do digitálnej formy. Signál je najskôr navzorkovaný, kvantovaný a kódovaný. Výsledkom je sekvencia (postupnosť) binárnych čísel, ktoré sa môžu ďalej spracovávať.



Obr. 1.11: Digitalizácia analógového signálu (PCM metóda)

Medzi základné metódy kódovania multimediálnych signálov v časovej oblasti používané v multimediálnych telekomunikáciách možno zaradiť:

- pulzne kódovú moduláciu (PCM),
- diferenčnú PCM (DPCM),
- adaptívnu DPCM (ADPCM).

Signál sa po obmedzení DP filtrom (antialiasing filter) podrobí vzorkovaniu vo vzorkovacom obvode, čím sa získa postupnosť jeho vzoriek.



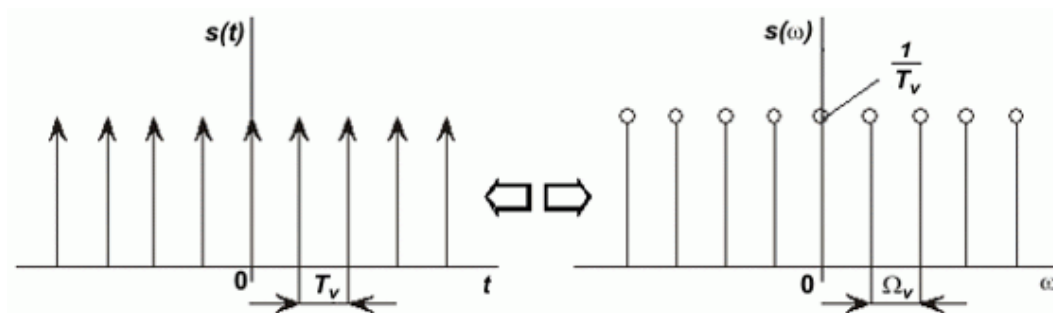
Vzorkovanie je prevod spojitého signálu na diskretný signál. Vzorkuje sa v definovaných časových bodoch, takže vzorky zodpovedajú hodnotám v čase a/alebo priestore. Vzorkovacia rýchlosť je určená zo vzorkovacej teóremy (známej ako Shannon-Kotelnikova teória), t.j. vzorkovacia frekvencia musí byť minimálne dva krát väčšia ako je hodnota najvyššej frekvencie signálu.

Veľkosť vzorkovacej frekvencie F_{vz} je daná šírkou pásma vzorkovaného signálu, napr. pre telefónny rečový signál s frekvenčným pásmom 300-3400 Hz (do 4 kHz) sa volí $F_{vz} = 8$ kHz.

Vzorkovacia teória môže byť matematicky vyjadrená ako $T_{vz} \leq \frac{1}{2F_m}$.

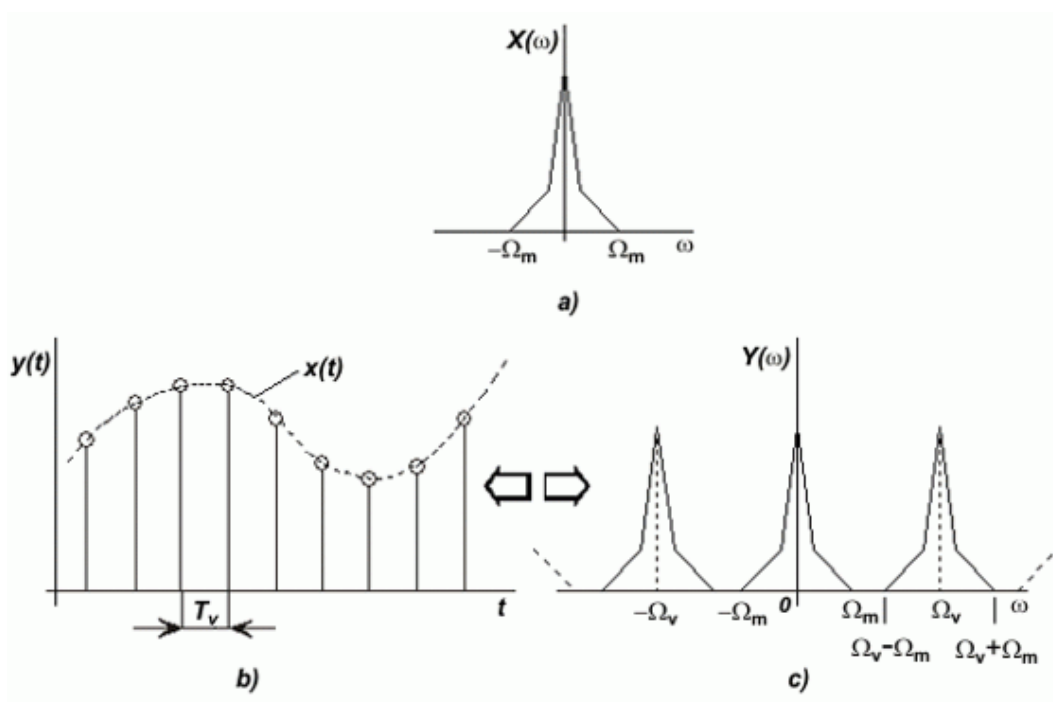
Minimálna vzorkovacia frekvencia $F_{vz \min} = 2F_m$ sa volá Nyquistova frekvencia.

Vzorkovaný signál, označme ho ako $y(t)$, je možné vyjadriť ako súčin priebehu $x(t)$ a vzorkovacej funkcie $s(t)$, ktorá je reprezentovaná nekonečnou sériou jednotkových (Diracových) impulzov. Spektrum vzorkovacej funkcie, ktorej impulzy sú od seba vzdialené o čas $t = T_{vz}$, je tiež nespojitý, frekvenčne neobmedzený a vzdialenosť jednotlivých zložiek na frekvenčnej osi je Ω_{vz} .



Obr. 1.12: Vzorkovacia funkcia v časovej a frekvenčnej oblasti

Spektrum $Y(\omega)$ vzorkovaného signálu $y(t)$ je výsledkom konvolúcie spektier signálov $X(\omega)$ a $S(\omega)$.



Obr. 1.13: a) Spektrum nevzorkovaného signálu $X(\omega)$ b) Časový priebeh ideálne vzorkovaného signálu $y(t)$ c) Spektrum ideálne vzorkovaného signálu $Y(\omega)$

Ak pre “periódu“ opakovania spektier platí $\Omega_v < 2\Omega_m$ dochádza k prekryvaniu zložiek spektra. Tento jav sa nazýva aliasing. K prekryvaniu spektier dochádza aj v dôsledku nie konečnej šírky spektra signálu $x(t)$. To spôsobí, že rekonštruovaný signál zo vzoriek je iný ako pôvodný signál.

Ďalším krokom digitalizácie je kvantovanie. Vzorky signálu sú nahradené zodpovedajúcimi kvantizačnými úrovňami (pevne určenými hodnotami ako napríklad prirodzené čísla), čím sa získa postupnosť kvantovaných vzoriek signálu. Kvantizačné úrovne sa získajú delením amplitúdy na menšie intervaly. Dĺžka intervalu sa volá krok kvantovania. V prípade, že všetky kroky kvantovania majú rovnakú dĺžku, hovoríme o lineárnom kvantovaní, v opačnom prípade ide o nelineárne kvantovanie.

- Najväčšou nevýhodou tohto procesu je kvantizačná chyba alebo šum. Je to rozdiel medzi analógovým vstupom do analógovo-digitálneho prevodníka (**ADC** *analog-digital convertor*) a digitálnym výstupom. Chyba je nelineárna a závislá od konkrétneho signálu. Chyba kvantovania spôsobuje aj problém pri spätnej konverzii digitálneho signálu na analógový. Preto signál nie je nikdy konvertovaný do identickej podoby, je iba aproximovaný z kvantizačných úrovní.

Ďalším krokom v digitalizácii je proces kódovania.



V tomto procese sa priradí k danej kvantovanej vzorke binárny kód, čím vzniká postupnosť kódových slov.

PCM metóda je medzinárodný štandard pre kódovanie a prenos multimediálnych signálov. Základný princíp tejto metódy je znázornený na obrázku nižšie.

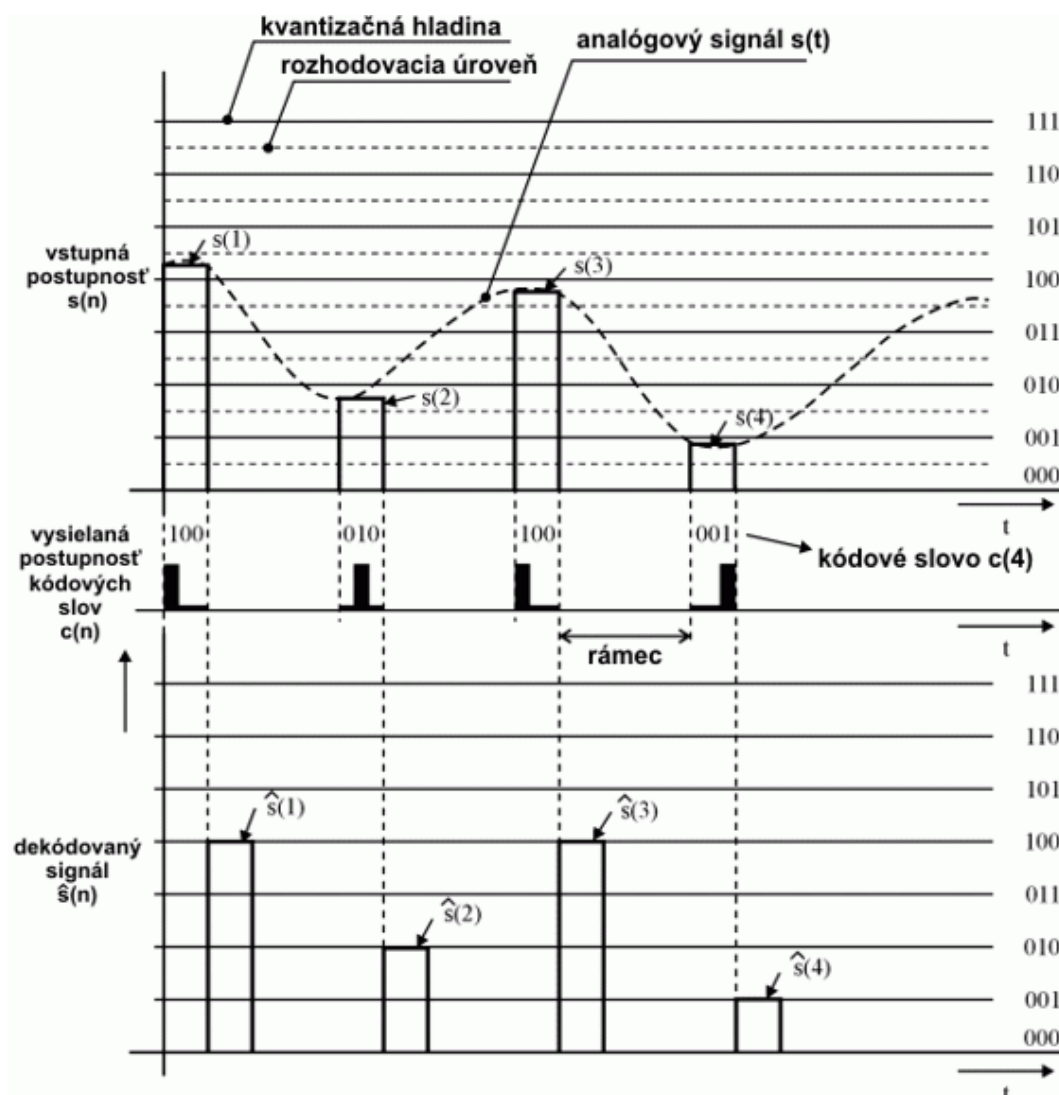


Prvé systémy na báze PCM používali 7 bitovú dĺžku kódového slova N , t.j. počet kvantizačných úrovní bol 128. Ak predpokladáme, že vzorkovacia frekvencia $F_{vz} = 8 \text{ kHz}$ a $N = 8$, potom potrebná prenosová rýchlosť na prenos reči v telefónnom pásme je $8 \cdot 10^3 \cdot 8 = 64 \text{ kbit/s}$.

- + Metóda PCM v porovnaní s analógovými metódami prenosu signálov má výhodu najmä v odolnosti prenášaného digitálneho signálu voči rušeniu.

- Na druhej strane, nevýhodou je požiadavka na širšie frekvenčné pásmo pri prenose.

Časové priebehy signálu kódovaného PCM sú znázornené na obrázku (Obr. 1.14).



Obr. 1.14: Časové priebehy v PCM metóde kódovania multimediálnych signálov

Lineárna PCM používa rovnaké kvantizačné kroky v celom pracovnom rozsahu kvantovania. Z toho vyplýva, že rozsah (dynamika) spracovateľného vstupného signálu závisí od počtu a veľkosti kvantizačných krokov. Počet kvantizačných krokov (stupňov) pri tom istom rozsahu vstupného signálu ovplyvňuje aj veľkosť kvantizačnej chyby. Pri väčšom počte kvantizačných krokov sa síce znižuje veľkosť kvantizačnej chyby, ale zvyšujú sa nároky na prenosovú rýchlosť. Tento prístup má teda obmedzenie, ktoré možno odstrániť nelineárnym rozložením kvantizačných krokov, čo je podstatou nelineárnej PCM.

Nelineárna PCM používa nelineárne rozloženie kvantizačných krokov tak, že ich veľkosť sa smerom k vyšším amplitúdam vstupného signálu zväčšuje. Modifikovaný prístup je kompresia dynamiky vstupného signálu na vysielačnej strane a expanzia dynamiky na prijímačnej strane. Kompresor na vysielačnej strane malé hodnoty vzoriek zosilní a veľké zoslabí. Expandér na prijímačnej strane naopak vráti vzorkám pôvodný rozsah a správny pomer ich hodnôt.

Dekódovanie je rovnaký proces ako kódovanie ale v opačnom poradí. Výstup dekodéra je postupnosť kvantovaných vzoriek.

1.4 Modulácia

Analógový signál, tak ako bol definovaný v prvej časti, je premenlivý signál spojený v čase aj amplitúde. Ak by sme mali napríklad graficky znázorniť striedavý prúd, vyzeral by ako vlna, s napätím meniacim sa zo záporných na kladné hodnoty. Sú tri faktory, ktoré je potrebné posúdiť: frekvencia, amplitúda a fáza.



Frekvencia je rýchlosť akou prúd osciluje nad a pod nulovú úroveň, resp. strieda svoju polaritu.

Keď prúd vystúpi nad nulovú hodnotu, potom ide späť pod nulovú hodnotu a následne sa vráti na nulu, hovoríme, že spravil jeden „cyklus“. Frekvencia sa udáva v počte cyklov za jednu sekundu a jej jednotkou je Hertz (Hz). Takže ak má analógový signál 500 cyklov za sekundu, hovoríme, že frekvencia je 500 Hertzov (500 Hz).



Za amplitúdu je považovaná hodnota najväčšej výšky a hĺbky grafickej vlny. Inak povedané, je to maximálna absolútna hodnota signálu.

Postupne, ako sa analógový signál šíri priestorom, amplitúda vlny sa znižuje. Táto vlastnosť sa nazýva útlm. Analógové vlny nie sú až tak ovplyvnené problémami s útlmom, avšak niekedy musia byť aj ony zosilnené. Amplitúda analógových vln sa meria vo wattoch, ampéroch alebo voltach. Jednotkou merania však často býva decibel. Decibel (dB) umožňuje jednoduché porovnanie dvoch rôznych hodnôt pomocou podielu (v tomto prípade ide o podiel amplitúd signálu). Decibel je bezrozmerná logaritmická jednotka.



Fáza popisuje rozdiel medzi začiatočným stavom cyklu jedného signálu v porovnaní so začiatočným stavom cyklu druhého signálu, t.j. popisuje ich posun. Jeden signál je vždy referenčným, druhý signál je fázovo posunutý.

Fázovo posunutý signál vzniká malým oneskorením signálu, ktoré spôsobí, že vrcholy a hĺbky nie sú synchronne v porovnaní s referenčným signálom. Úroveň nesynchronizovanosti je meraná v stupňoch. Ak je fázový posun signálu 180 stupňov, znamená to, že posunutý signál začína tam, kde sa referenčný signál po zostupe z vrcholu dotýka nuly.

Je dôležité dobre poznať amplitúdu, frekvenciu ako aj fázu, keďže signál dokáže prenášať údaje práve cez zmenu týchto vlastností v čase.

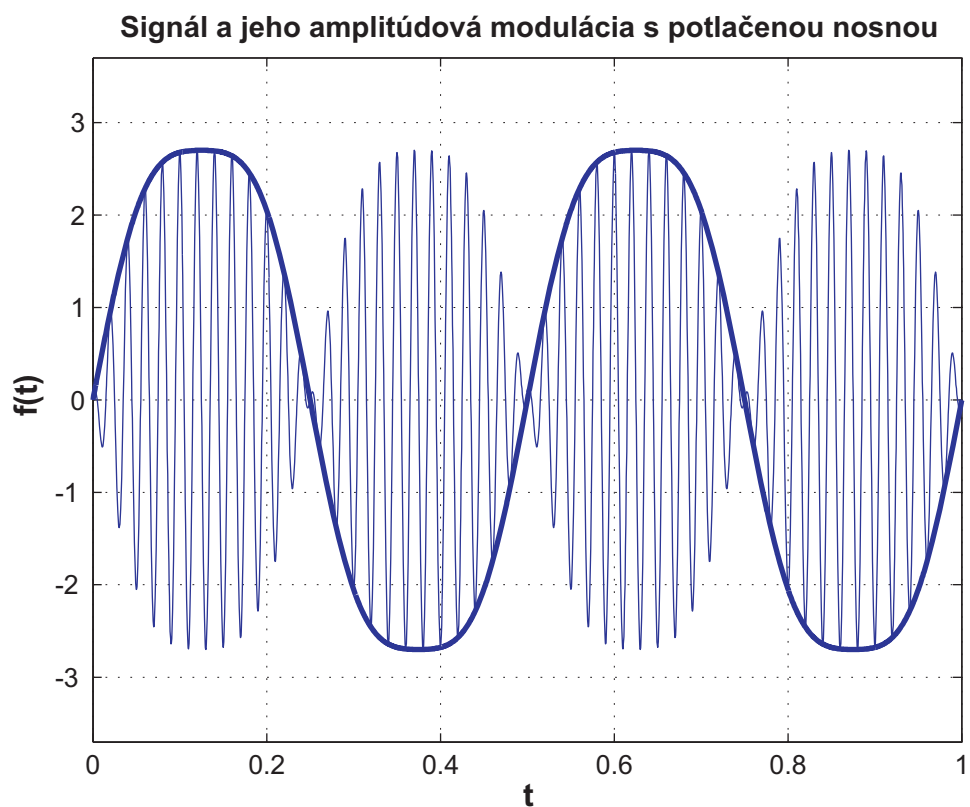


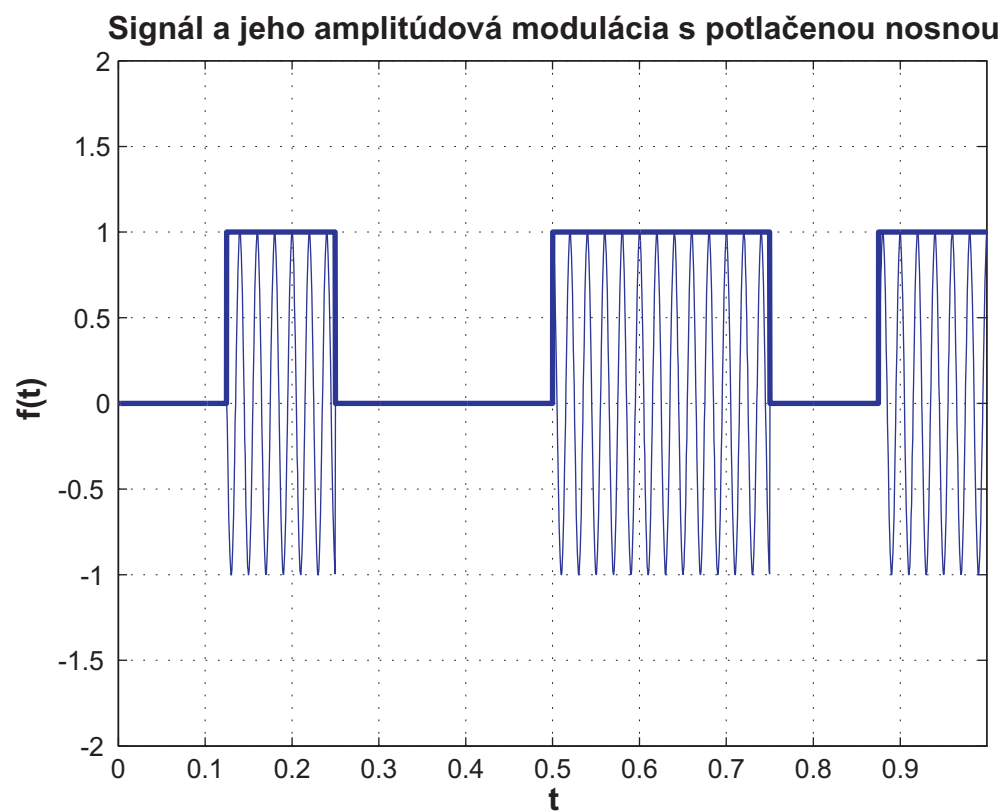
Modulácia je proces vkladania správy, napríklad digitálnych bitov, alebo analógového audio signálu, do iného signálu, ktorý môže byť fyzicky vysielaný.

Jedným z dôvodov modulácie je umožniť prenos viacerých signálov v jednom kanáli s obmedzenou šírkou pásma. Každý signál je modulovaný tak, aby sa zmestil do prenosového subpásma. Výhodou teda je, že po zoskupení týchto

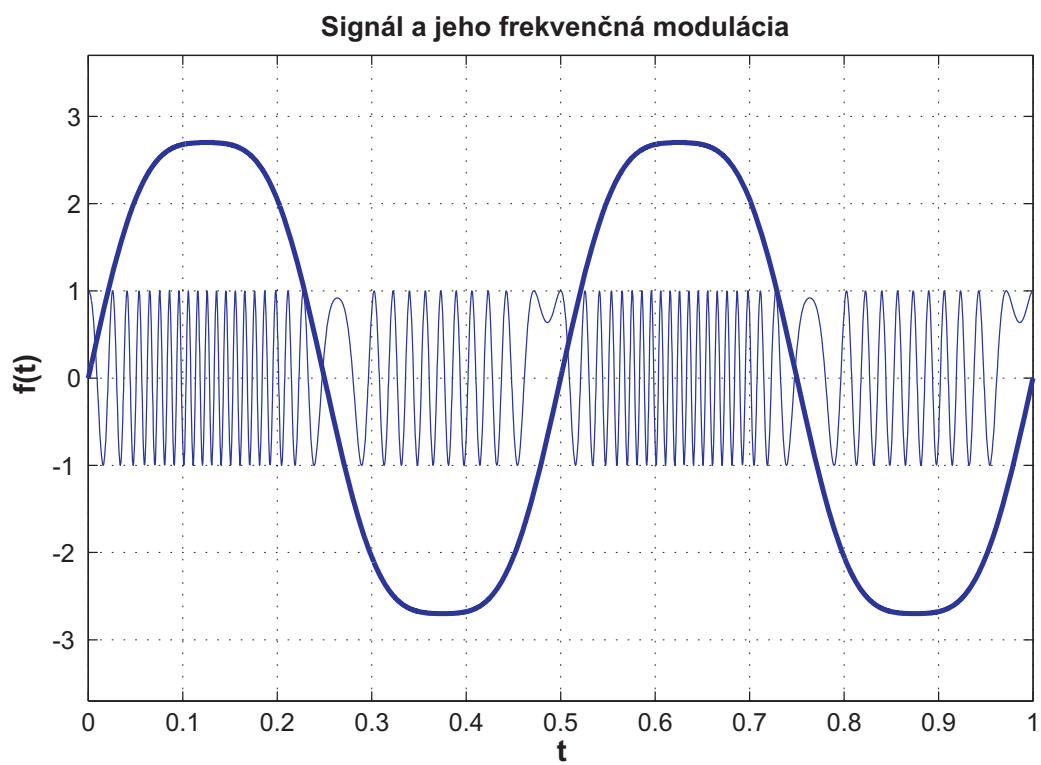
signálov (čo je možné, keďže každý má individuálne subpásmo) môžeme použiť iba jedno prenosové médium (ktorého pásma je zložené zo subpásiem).

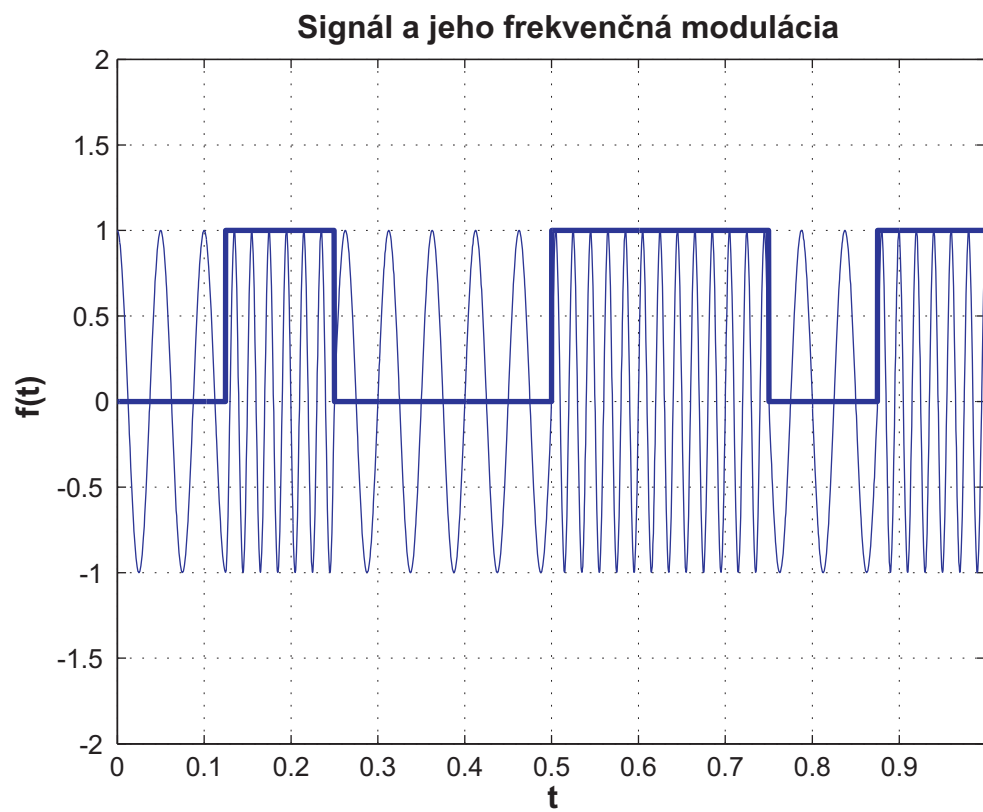
Pri analógovej modulácii sa modulácia aplikuje kontinuálne podľa spojitého analógového signálu. Existuje veľa druhov analógových modulácií, najjednoduchšie sú *amplitúdová modulácia (AM)*, *fázová modulácia (PM)* a *frekvenčná modulácia (FM)*.



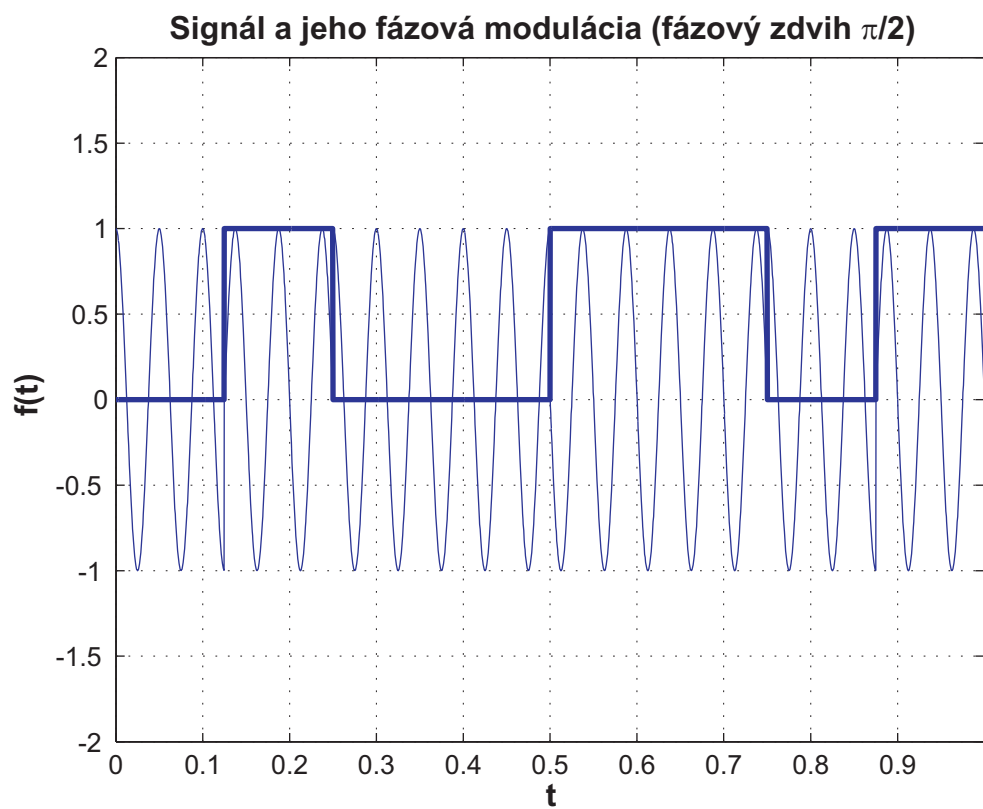


Obr. 1.15: Amplitúdová modulácia analógového a digitálneho signálu





Obr. 1.16: Frekvenčná modulácia analógového a digitálneho signálu



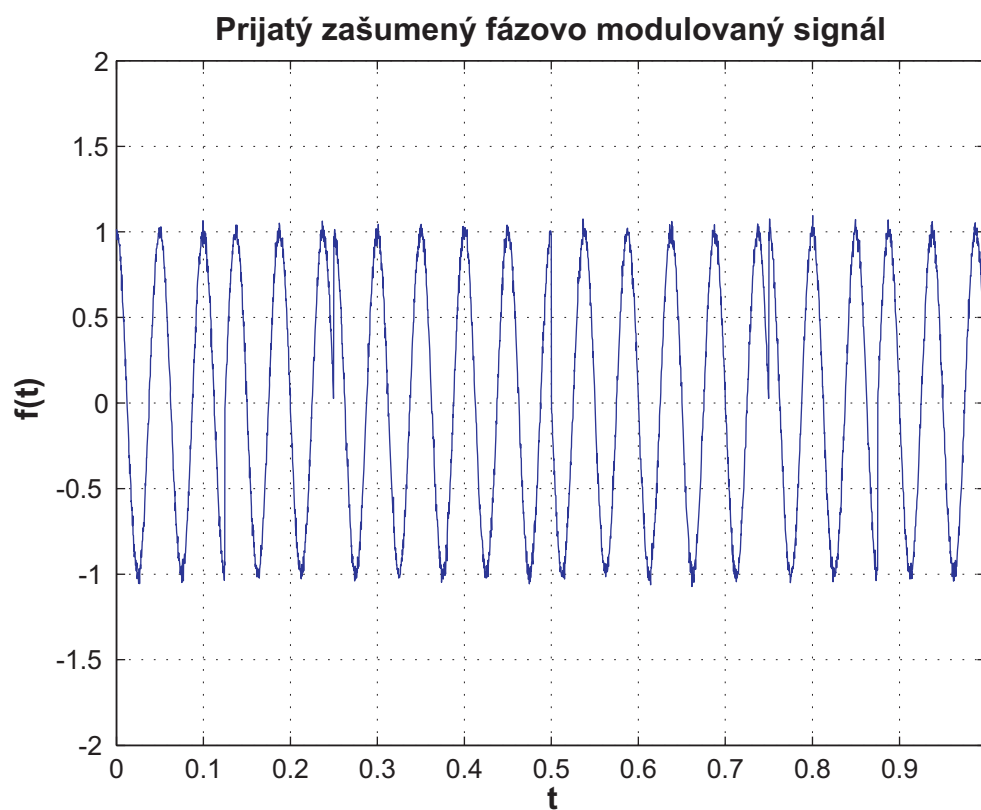
Obr. 1.17: Fázová modulácia digitálneho signálu

AM pracuje zmenou sily prenášaného signálu na základe posielanej informácie – amplitúda je priamo úmerná sile informačného signálu. Nosná vlna je modulovaná (modifikovaná) vstupným signálom za účelom prenosu informácie. Nosná vlna má zvyčajne oveľa vyššiu frekvenciu ako vstupný informačný signál. Amplitúda nosnej vlny je modulovaná vstupným signálom (t.j. prenášanou informáciou) pred samotným prenosom. Vstupná vlna modifikuje amplitúdu nosnej vlny a tým pádom definuje “obálku” celého priebehu.

FM prenáša informáciu cez nosnú vlnu modifikáciou okamžitej frekvencie. Pri analógovej modulácii je zmena medzi momentálnou a štandardnou frekvenciou nosnej vlny priamo úmerná okamžitej hodnote amplitúdy vstupného signálu.

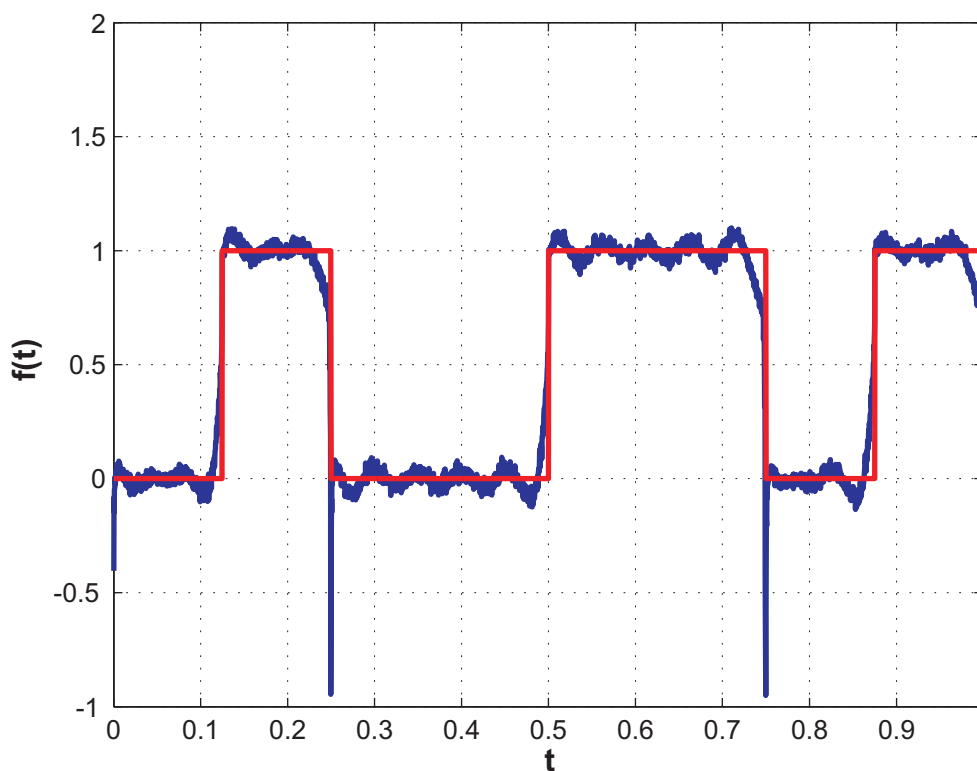
PM je forma modulácie, kde je informácia reprezentovaná ako zmena okamžitej fázy nosnej vlny. Zjednodušene povedané, modifikácia fázy podľa dátového signálu je fázová modulácia. PM nie je široko používanou technológiou pri rádiových prenosoch z dôvodu potreby komplexného prijímacieho aparátu. Pri PM je tiež potrebné ošetriť nejednoznačnosti ako napríklad pri zmenách fázy o $+180^\circ$ alebo -180° . PM sa však úspešne používa v digitálnych hudobných syntetizátoroch.

Demodulácia je proces získania informácie z modulovaného nosného signálu. Demodulácia samozrejme závisí od toho, aký parameter signálu bol modifikovaný, a teda aká modulácia bola použitá (amplitúdová, frekvenčná, alebo fázová). Napríklad pre signál s amplitúdovou moduláciou je možné použiť synchronný detektor. Avšak pre iné druhy, napríklad FM a PM modulácie je potrebné použiť príslušný FM, resp. PM demodulátor.



Obr. 1.18: Prijatý namodulovaný signál

Demodulácia prijatého zašumeného signálu a porovnanie s originálom



Obr. 1.19: Demodulovaný signál porovnaný s pôvodným signálom

2 Časová a frekvenčná reprezentácia

2.1 Fourierova transformácia

Fourierova transformácia, pomenovaná podľa Josepha Fouriera, je dôležitá v matematike, technických vedách a fyzike.



Jednoducho povedané, Fourierova transformácia predstavuje matematické vyjadrenie funkcie času v závislosti na frekvencii. Táto funkcia je známa ako frekvenčné spektrum.

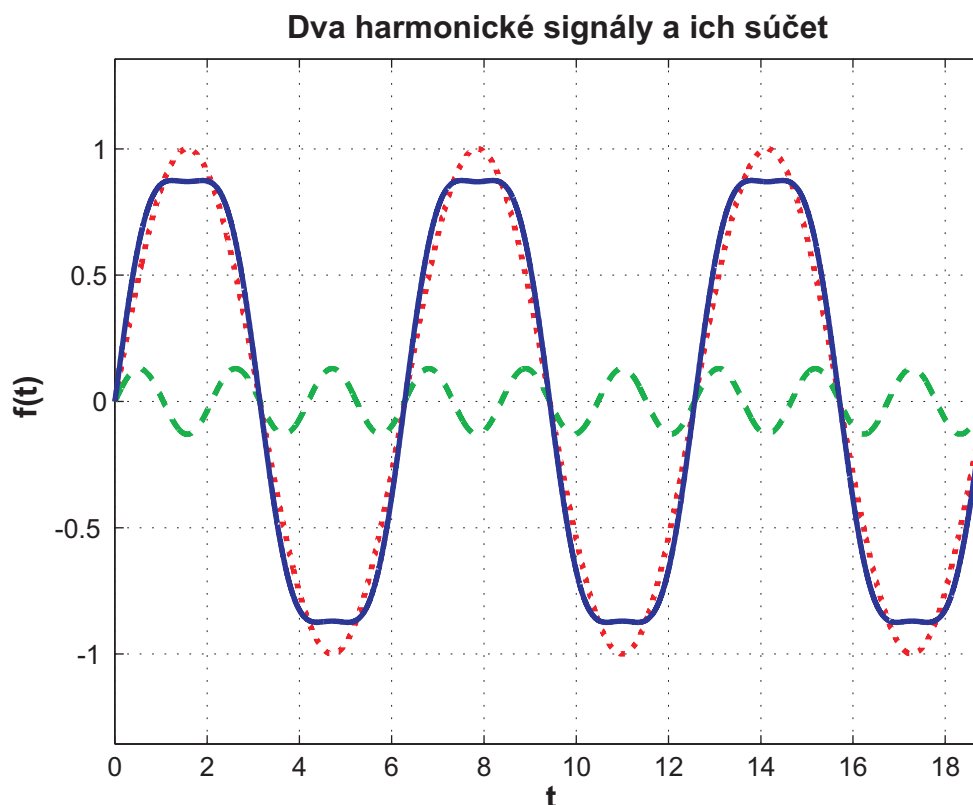


Je dôležité poznamenať, že Fourierova transformácia sa používa len pre neperiodické analógové signály. Pre periodické analógové signály sa používa Fourierov rad.

Povedzme, že máme funkciu $f(t)$, ktorá mapuje nejaké časové hodnoty t na nejaké hodnoty $f(t)$.

Teraz skúsime aproximovať f ako sumu jednoduchých harmonických kmitov, t.j. sínusových vln určitej frekvencie ω . Samozrejme, sú tam niektoré frekvencie, ktoré sa dobre hodia na aproximáciu f a niektoré ju menej dobre aproximujú. Potrebujeme teda nejakú vhodnú funkciu $f(\omega)$, ktorá nám povie, do akej miery sú dané oscilácie s frekvenciou ω reprezentované v aproximácii f .

Zoberme si napríklad funkciu (zobrazená čiernou čiarou) znázornenú na Obr. 2.1:



Obr. 2.1: Dve harmonické frekvencie tvoriace signál

ktorá je definovaná ako $f(t) = \sin(t) + 0.13\sin(3t)$. Kmitanie (bodkovaná čiara) s $\omega=1$ má najväčší vplyv na výsledok, takže môžeme povedať, že sa $F(1)=1$. Druhá vlna ($\omega = 3$, prerušovaná čiara) má malý vplyv a jej amplitúda je oveľa menšia. Teda hovoríme, že sa $F(3)=0.13$. Ďalšie frekvencie nemusia byť vôbec uvedené v aproximácii, teda píšeme, že pre tieto sa $F(\omega)=0$.

Teraz, keď poznáme $F(\omega)$ nie len pre niektoré ale pre všetky možné frekvencie ω , môžeme perfektne aproximovať našu funkciu $f(t)$. A to je to, čo robí spojitá Fourierova transformácia.

Fourierova transformácia teda vezme nejakú funkciu času $f(t)$ a vráti k nej príslušnú funkciu $F(\omega)=\mathbf{FT}(f)$, t.j. jej *Fourierovu transformáciu*. FT popisuje, do akej miery je niektorá z daných frekvencií reprezentovaná vo funkcii f . Ide iba o inú reprezentáciu $f(t)$ s rovnakou informáciou, ale so zápisom v inej doméne (napríklad frekvenčnej). Často krát však môže byť problém vyriešený jednoduchšie v inom zastúpení (spočíva v nájdení vhodného súradnicového systému).

Danú Fourierovu transformáciu môžeme integrovať cez všetky frekvencie, zostaviť váhované sínusové vlny a dostať opäť naše f . Tento proces voláme *spätná Fourierova transformácia IFT*.

Najdôležitejšie je, že Fourierova transformácia má veľa výhodných matematických vlastností (napr. konvolúcia je v frekvenčnej doméne iba

násobením). Často krát je oveľa jednoduchšie pracovať s Fourierovou transformáciou ako s funkciou samotnou. Takže funkciu transformujeme, vykonáme jednoduchú matematickú operáciu a následne transformujeme naspäť.



Povedzme, že chceme redukovať šum v digitálnom obraze. Namiesto manipulovania s funkciou obrazu: Pixel→Svetlosť, radšej transformujeme celok a pracujeme s $F(\text{obraz})$: Frekvencia→Amplitúda. Tie časti vysokých frekvencií, ktoré zapríčiňujú šum, môžu byť jednoducho odstránené – $F(\text{image})(\omega)=0, \omega>\dots\text{Hz}$.



Fourierova transformácia (zvyčajne známa ako dopredná transformácia) je definovaná ako:

$$F(\omega) = \int_{-\infty}^{\infty} f(t)e^{-j\omega t} dt$$

a inverzná Fourierova transformácia je definovaná:

$$f(t) = \frac{1}{2\pi} \int_{-\infty}^{\infty} F(\omega)e^{j\omega t} d\omega$$

Fourierova funkcia $F(\omega)$ je frekvenčná reprezentácia signálu $f(t)$, tiež nazývaná spektrálna funkcia. Spektrálna funkcia závisí od reálnej premennej ω , takže môže byť definovaná ako:

$$F(\omega) = A(\omega) + jB(\omega) = \text{Re}\{F(\omega)\} + j\text{Im}\{F(\omega)\} = |F(\omega)|e^{j\varphi(\omega)}$$

kde $|F(\omega)|$ je absolútna hodnota a $\varphi(\omega)$ je fázové spektrum.

V každej rovnici je j definované ako $j = \sqrt{-1}$. Komplexná mocnina je srdcom transformácie. Komplexná mocnina je v podstate komplexné číslo, kde obe zložky, reálna aj imaginárna, sú sínusoidy. Exaktný vzťah sa nazýva Eulerova rovnica $e^{j\varphi} = \cos\varphi + j\sin\varphi$, ktorá vedie k zaujímavej (a krásnej) rovnosti $e^{j\pi} + 1 = 0$. Je jednoduchšie pracovať s komplexnými mocninami ako s trigonometrickými funkciami, poskytujú aj oveľa jednoduchší spôsob zápisu sínusoid (je jednoduchšie napísať $e^{j\varphi}$ ako napísať $\cos\varphi + j\sin\varphi$).



Komplexné mocniny (alebo sínusy a kosínusy) sú periodické funkcie a súbor komplexných mocnín je kompletný a ortogonálny. Preto Fourierova transformácia môže reprezentovať akúkoľvek čiastočnú spojitú funkciu a chyba meraná metódou najmenších štvorcov (least-square error) medzi funkciou a jej fourierovou reprezentáciou je minimálna.

Existuje aj iný úplný a ortogonálny súbor periodických funkcií; napríklad Walshova funkcia (štvorcové vlny), ktorá je užitočná pre digitálne elektronické zariadenia (viac o tom v kapitole Ortogonálne transformácie).



Prečo sa vždy stretávame s komplexnými mocninami pri riešení fyzikálnych problémov? Prečo sú monochromatické vlny sínusového charakteru a nie periodické sledy štvorcových alebo trojuholníkových vln? Dôvod je, že derivácie komplexných mocnín sú jednoducho prepočítané komplexné mocniny. Inými slovami, komplexné mocniny sú funkcie diferenciálnych operátorov (operátor rozdielu). Väčšina fyzikálnych systémov sa riadi podľa pravidiel diferenciálnych rovníc. Teda analógové elektronické filtre budú konvertovať sínusovú vlnu na inú sínusovú vlnu s rovnakou frekvenciou (ale nie nevyhnutne s rovnakou amplitúdou a fázou), kým filtrovaná štvorcová vlna už viac nebude štvorcovou vlnou. Táto vlastnosť komplexných mocnín robí Fourierovu transformáciu užitočnou v rozmedzí od rádiového šírenia až po kvantovú mechaniku.



Fourierova transformácia je definovaná tiež pre dvojrozmerné signály a matematicky je definovaná ako:

$$F(\omega_1, \omega_2) = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} f(x_1, x_2) e^{-j\omega_1 x_1} e^{-j\omega_2 x_2} dx_1 dx_2$$

a inverzná transformácia je definovaná:

$$f(x_1, x_2) = \frac{1}{2\pi} \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} F(\omega_1, \omega_2) e^{j\omega_1 x_1} e^{j\omega_2 x_2} d\omega_1 d\omega_2$$

2.2 Diskrétna Fourierova transformácia

Diskrétna Fourierova transformácia **DFT** sa používa na získanie spektra diskrétnych signálov. Aby sme popísali signál vo frekvenčnej doméne, potrebujeme konečný počet sínusoid.

Pre znázornenie čo robí DFT použijeme nasledujúci príklad. MP3 prehrávač posielá reproduktoru audio informácie v podobe kolísavého napätia elektrického signálu. Výsledkom je pohyb vzduchových častíc a produkcia zvuku. Kolísanie audio signálu v čase môže byť graficky znázornené: os x predstavuje čas a os y napätie elektrického signálu. Toto vyzerá ako nepravidelný vlnový hlavolam, ktorý je v skutočnosti sumou niekoľkých pravidelných hlavolamov, ktoré predstavujú rôzne frekvencie zvuku. Frekvencia znamená len rýchlosť s akou sa molekuly vzduchu pohybujú tam a naspäť, alebo kolísanie napätia.



„DFT robí matematicky to, čo fyzikálne robí ľudské ucho: rozkladá signál na jednotlivé zložky frekvencie. Na rozdiel od analógového signálu z gramofónu, digitálny signál z MP3 prehrávača je len sled čísel, ktoré reprezentujú veľmi krátke vzorky skutočného zvuku: napríklad CD-kvalita digitálne zvukového prehrávača zbiera 44 100 vzoriek za sekundu. Ak vyberieme niekoľko po sebe idúcich hodnôt z digitálneho signálu: 8, 128 alebo 1 000 – DFT ich reprezentuje ako vážený súčet rovnakého počtu frekvencií. („Váženú“ len znamená, že niektoré frekvencie sa významnejšie podieľajú na celku ako iné.)“



Diskrétna Fourierova Transformácia (DFT) je matematicky vyjadrená ako:

$$X(k) = \sum_{n=0}^{N-1} x(n) e^{-n2\pi jk/N}$$

a inverzná diskretná Fourierova transformácia (**IDFT**) je definovaná ako:

$$x(n) = \frac{1}{N} \sum_{k=0}^{N-1} X(k) e^{n2\pi jk/N}$$

kde N je počet vzoriek diskretného signálu a $n = 0, 1, 2, \dots, N-1$, $e^{2\pi jk/N}$ je veľmi často nahradené ako Ω ($\Omega = e^{2\pi jk/N}$).

DFT sledu N -bodových časových vstupov je N -bodové frekvenčné spektrum s Fourierovou frekvenciou k s rozsahom od $-(N/2 - 1)$, cez nultú frekvenciu nazývanou jednosmerná zložka, až po najvyššie Fourierove frekvencie $N/2$. Každá spektrálna zložka predstavuje celé číslo sínusových periód prezentovaných v časovom slede. Amplitúdy a fázy sú reprezentované amplitúdami A_k a fázami $\varnothing_k$ týchto sínusoid. Pre zhrnutie, každá spektrálna zložka môže byť popísaná ako $X(k) = A_k e^{j\varnothing_k}$.

Diskrétna Fourierova transformácia je definovaná tiež pre dvojrozmerné signály a môže byť reprezentovaná ako séria rozšírení funkcie obrazu (cez 2D priestor).



Definície doprednej a inverznej **2D FT** sú nasledovné:

$$X(\omega_1, \omega_2) = \sum_{n_1=-\infty}^{\infty} \sum_{n_2=-\infty}^{\infty} x(n_1, n_2) e^{-j\omega_1 n_1} e^{-j\omega_2 n_2}$$

$$x(n_1, n_2) = 1/(2\pi)^2 \int_{-\pi}^{\pi} \int_{-\pi}^{\pi} X(\omega_1, \omega_2) e^{j\omega_1 n_1} e^{j\omega_2 n_2} d\omega_1 d\omega_2$$

Spektrum $X(\omega_1, \omega_2)$ je komplexné keď aj postupnosť $x(n_1, n_2)$ je reálna. Spektrum môže byť definované tiež ako suma reálnych a imaginárnych častí alebo ako súčin magnitúdy a fázy.

$$X(\omega_1, \omega_2) = |X(\omega_1, \omega_2)| e^{j\theta_x(\omega_1, \omega_2)} = X_R(\omega_1, \omega_2) + jX_I(\omega_1, \omega_2)$$

2.3 Spektrum



Frekvenčné spektrum časového signálu je reprezentácia tohto signálu vo frekvenčnej doméne. Frekvenčné spektrum môže byť generované cez Fourierovu transformáciu (alebo diskretnú Fourierovu transformáciu) signálu a výsledné hodnoty sú zvyčajne prezentované ako amplitúda a fáza, obe vykreslené pomerne k frekvenciám.

Signál môže byť reprezentovaný ako amplitúda, ktorá sa mení s časom, má odpovedajúce frekvenčné spektrum. Toto zahŕňa známe pojmy ako viditeľné svetlo (farba), hudobné noty, rádio/TV programy a tiež pravidelná rotácia Zeme. Keď sú tieto fyzikálne javy reprezentované vo forme frekvenčného spektra, určitý fyzikálny popis ich interných procesov sa stáva oveľa jednoduchším. Často krát frekvenčné spektrum jasne ukazuje harmonické zložky, znázornené ako viditeľné špičky alebo čiary, ktoré poskytujú jasný náhľad do mechanizmu, ktorý vytvára celý signál.

Spektrálna analýza je technický proces dekompozície celého signálu na jednoduchšie časti. Ako je popísané vyššie, mnoho fyzikálnych procesov je lepšie popísaných ako suma viacerých individuálnych frekvenčných komponentov. Akýkoľvek proces, ktorý kvantifikuje rôzne množstvo (t.j. amplitúda, výkon, intenzita alebo fáza) v pomere k frekvencii, sa môže nazývať spektrálna analýza.

Spektrálna analýza môže byť vykonávaná na celom signáli (zvyčajne periodikom signáli). Prípadne môže byť signál (väčšinou neperiodický alebo kvázi periodický) rozdelený na kratšie časti, nazývané rámce, a spektrálna analýza môže byť aplikovaná na tieto individuálne časti.

Fourierova transformácia funkcie vytvára frekvenčné spektrum, ktoré obsahuje všetky informácie o originálnom signáli, ale v inej forme. To znamená, že originálna funkcia môže byť úplne zrekonštruovaná (syntetizovaná) inverznou Fourierovou transformáciou.

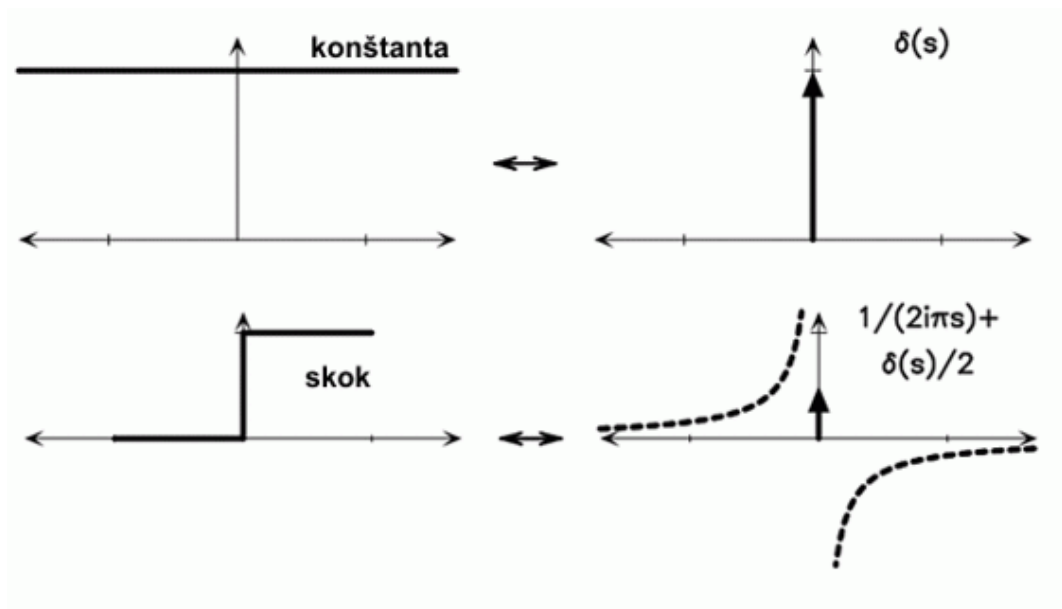


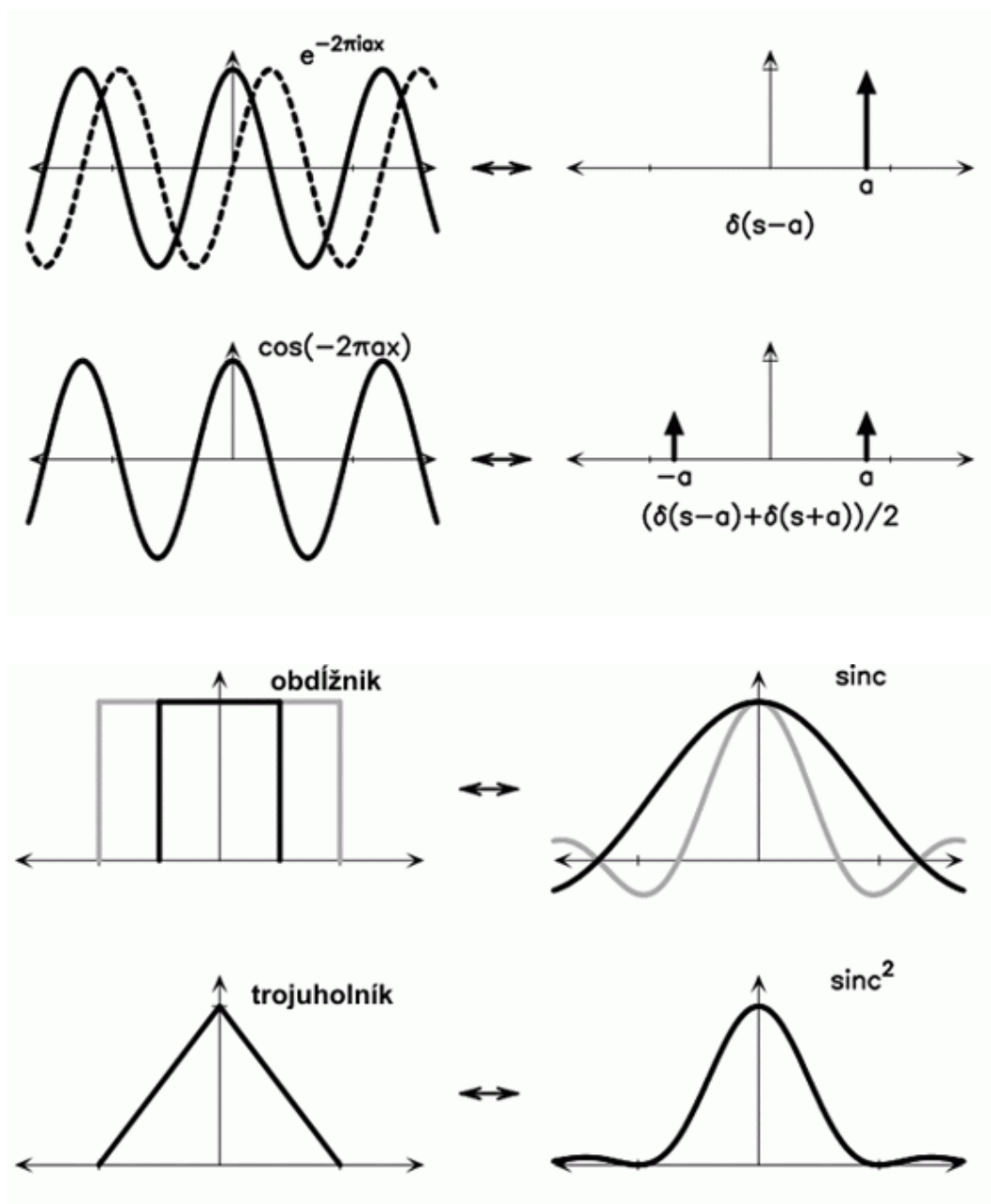
Pre dokonalú rekonštrukciu spektrálna analýza musí zachovať amplitúdu aj fázú každého frekvenčného komponentu. Tieto dve informačné časti môžu byť reprezentované ako dvojrozmerný vektor, ako komplexné číslo alebo ako magnitúda (amplitúda) a fáza v polárnej súradnicovej sústave. Známa technika v oblasti spracovania signálu je brať do úvahy mocninu amplitúdy alebo výkonu; v tomto prípade výsledný graf sa odvoláva na spektrum výkonu.

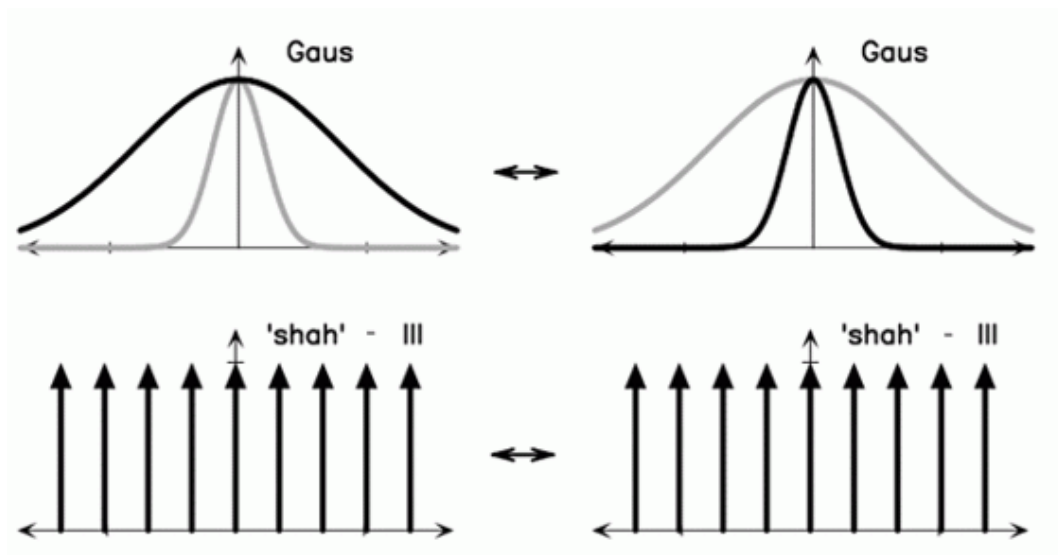
Nasledujúca tabuľka zhŕňa typy signálov a ich spektrum.

Signál	Spektrum
spojitý periodický	diskrétne neperiodické
spojitý neperiodický	spojité neperiodické
diskrétny periodický	diskrétne periodické
diskrétny neperiodický	spojité periodické

Nasledujúci obrázok (Obr. 2.2) ukazuje základné páry Fourierovej transformácie (na obrázku sú zobrazené iba amplitúdy). Tieto môžu byť kombinované použitím nižšie uvedenej teórie o Fourierovej transformácii na generovanie Fourierových transformácií rôznych funkcií.

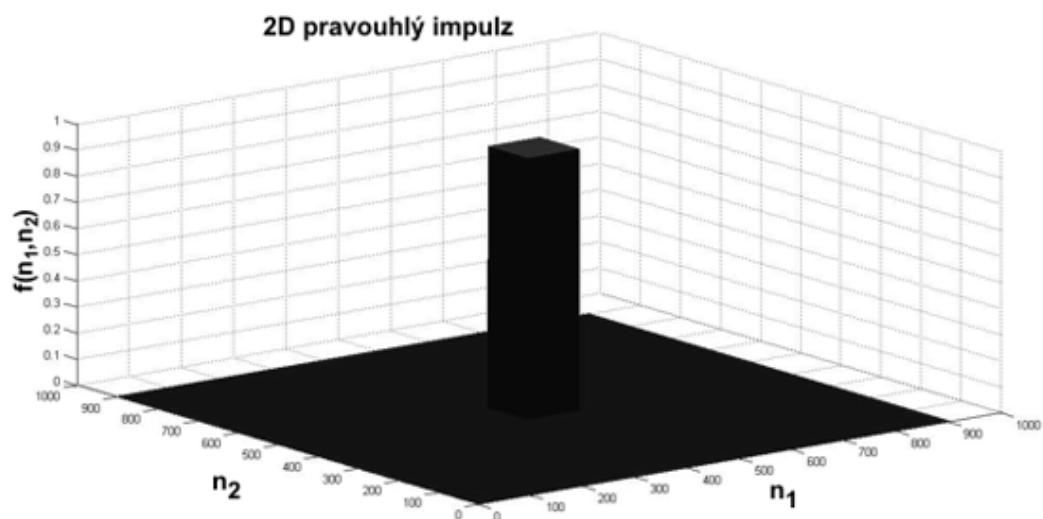


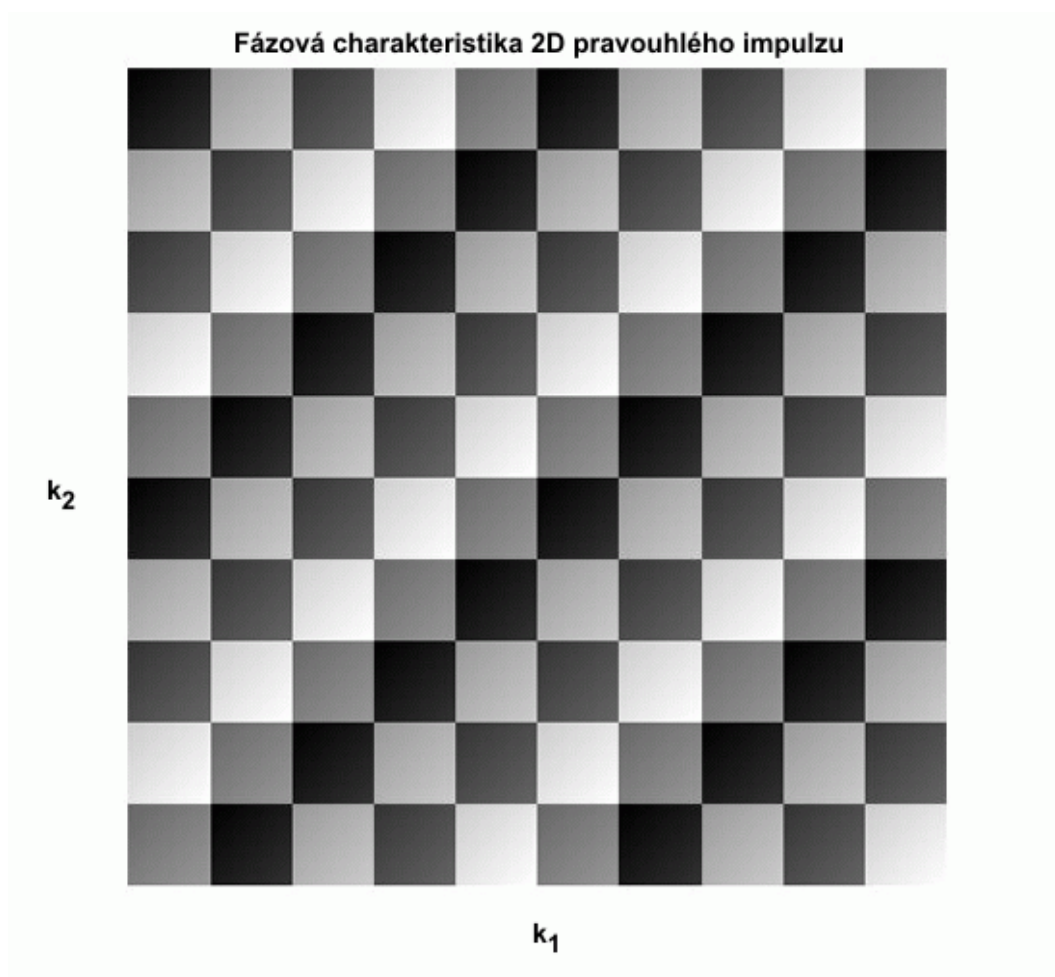
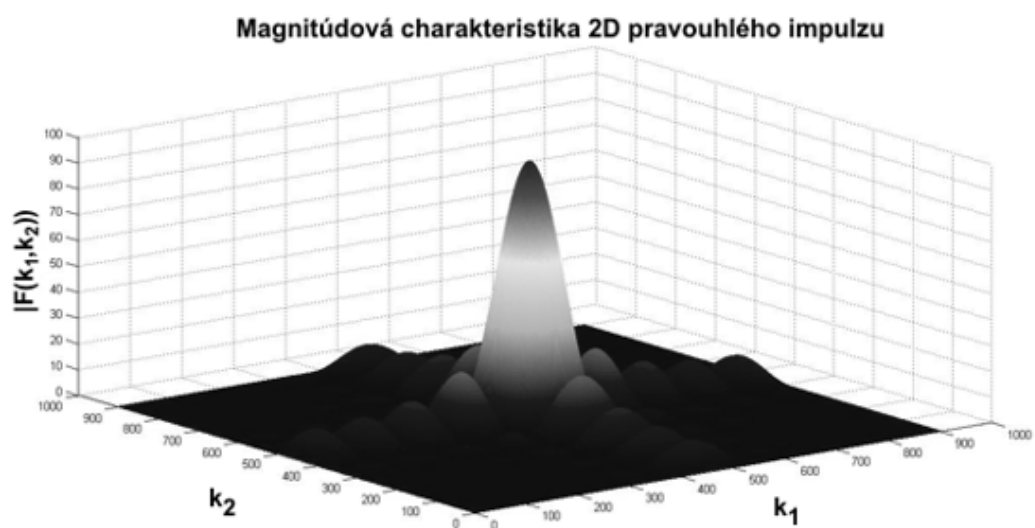




Obr. 2.2: Základné páry Fourierových transformácií

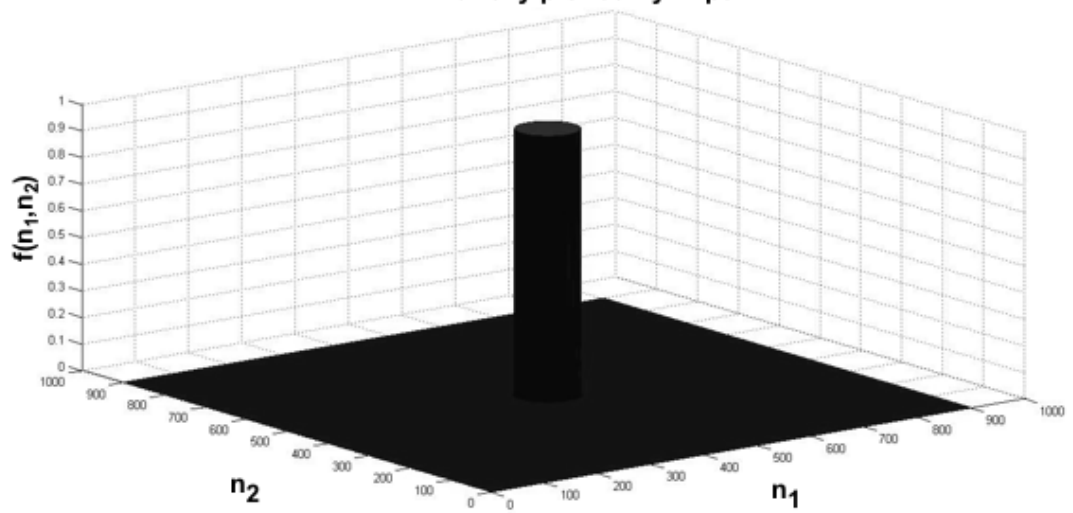
Obrázok nižšie (Obr. 2.3 a Obr. 2.4) zobrazuje dvojrozmerné signály a ich spektrá (magnitúdovú a fázovú charakteristiku).



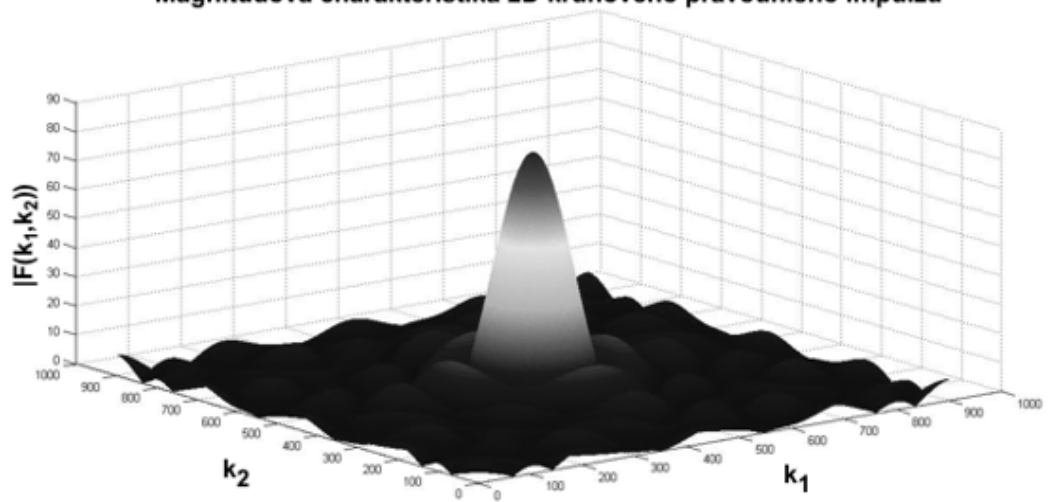


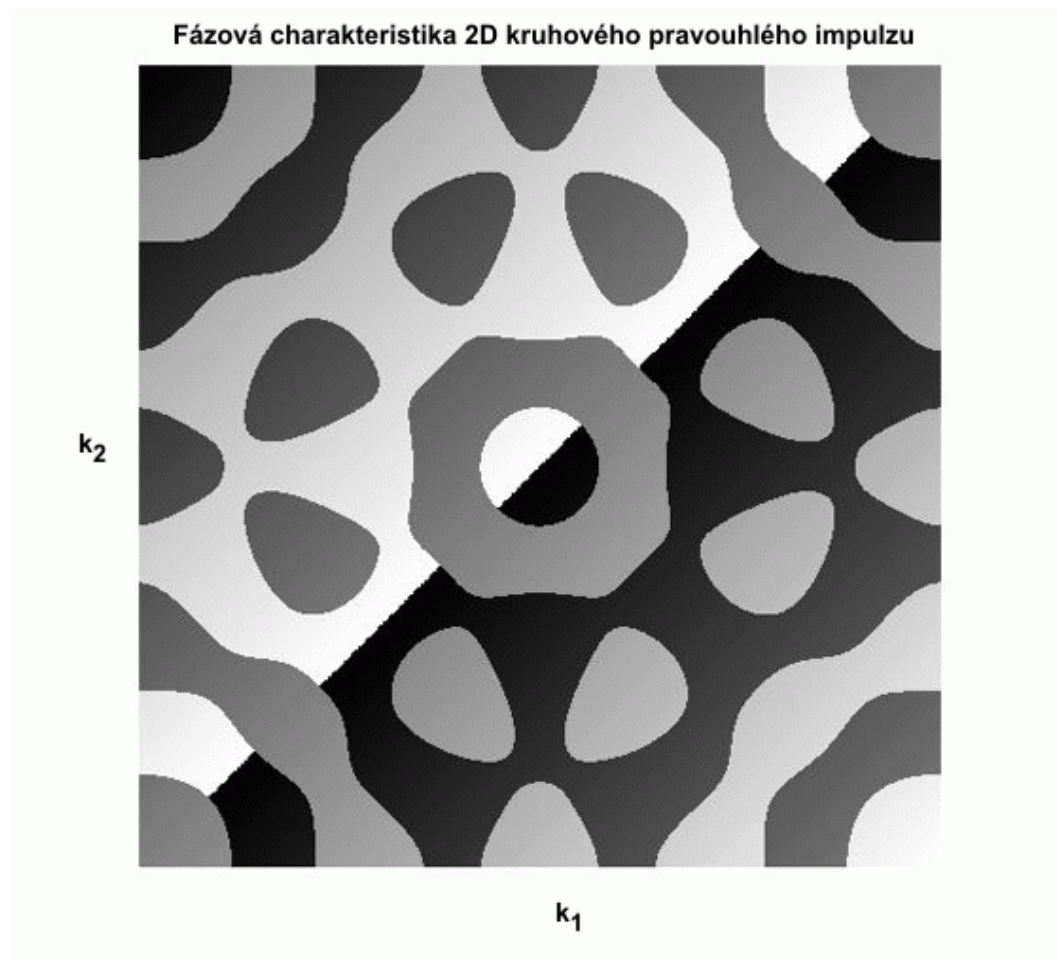
Obr. 2.3: 2D štvorcová funkcia, magnitúdová a fázová frekvenčná charakteristika

2D kruhový pravouhlý impulz



Magnitudová charakteristika 2D kruhového pravouhlého impulzu





Obr. 2.4: 2D kruhová funkcia, magnitúdová a fázová frekvenčná charakteristika

2.4 Ortogonálne transformácie



Ortogonalne transformácie umožňujú reprezentovať časový priebeh signálu vo forme jeho zovšeobecneného spektra. Transformácia sa používa, pretože v spektrálnej oblasti je potom možné vykonávať potrebné matematické operácie, ktoré umožňujú napr. rýchlejšie vyhodnotenie signálov, dokážu vylúčiť redundanciu, atď.

Diskrétné ortogonálne transformácie nachádzajú široké využitie v mnohých oblastiach pri kompresii údajov, rozpoznávaní obrazov, analýze a syntéze rečových signálov, atď. Naším hlavným cieľom v tomto kurze sú jednorozmerné ortogonálne funkcie a transformácie.

Najjednoduchším spôsobom matematického vyjadrenia jednorozmerných signálov je lineárna kombinácia (súčet násobkov) niektorých základných (bázových) funkcií. Také vyjadrenie jednorozmerných signálov je veľmi výhodné najmä v lineárnych sústavách, lebo umožňuje riešiť mnohé úlohy na základe princípu superpozície. Vyžaduje sa, aby základné funkcie $u(k, t)$ sa ľahko vyčísľovali a mali jednoduchý tvar a týmto požiadavkám najviac vyhovujú ortogonálne funkcie.



Matematická definícia ortogonálnych funkcií $u(0, t), u(1, t), \dots, u(N - 1, t)$ na intervale $\langle t_1, t_2 \rangle$ je nasledovná:

$$\int_{t_1}^{t_2} u(k, t) u(m, t) dt = 0, \quad k \neq m$$

$$\int_{t_1}^{t_2} u^2(k, t) dt = U_k, \quad k \neq m$$

V prípade, že $U_k = 1$, funkcie nazývame ortonormálne (t.j. majú veľkosť jedna).

Príklad matematického vyjadrenia jednorozmerného signálu $x(t)$, pomocou lineárnej kombinácie niektorých základných (bázových) funkcií:

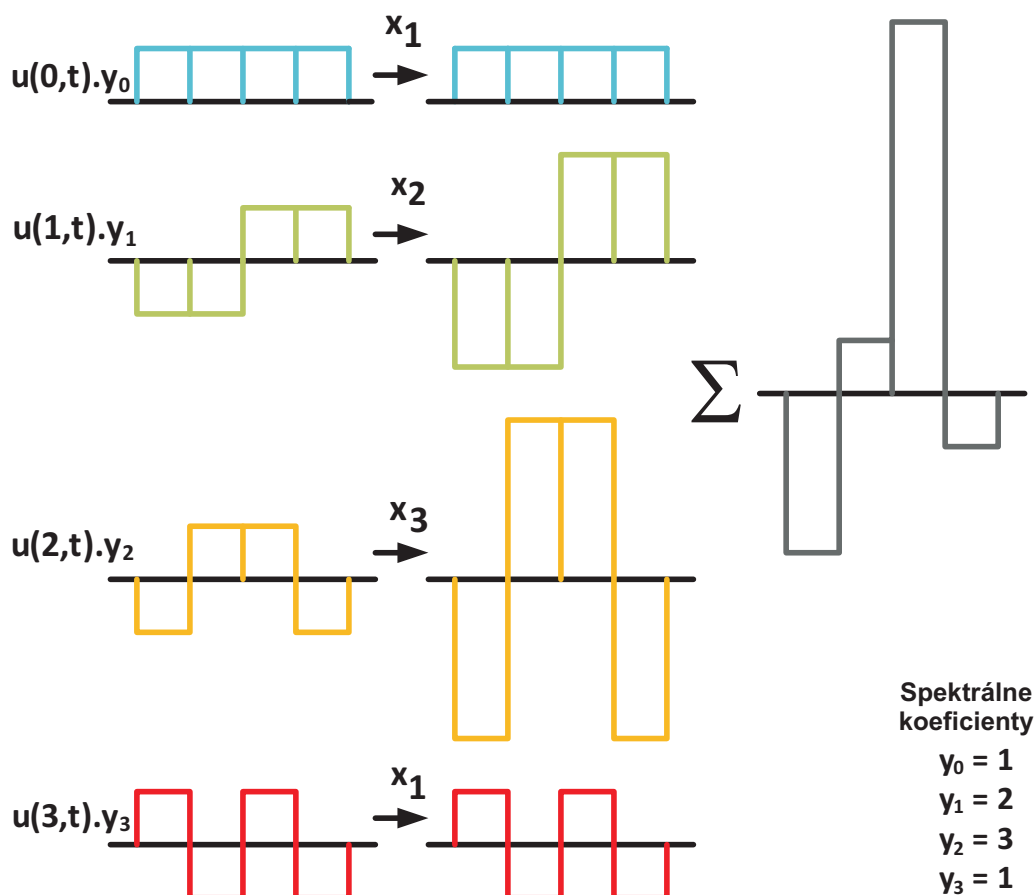
$$x(t) \cong \sum_{k=0}^{N-1} y_k u(k, t)$$

kde y_k sú spektrálne koeficienty definované ako:

$$y_k = \frac{1}{U_k} \int_{t_1}^{t_2} x(t) u(k, t) dt$$

Preštudujte si nasledujúci príklad na obrázku Obr. 2.5.

System prvých 4 Walshových funkcií



Obr. 2.5: Príklad matematického vyjadrenia jednorozmerných signálov, pomocou lineárnej kombinácie niektorých bázových funkcií

Medzi najviac používané ortogonálne funkcie v oblasti spracovania signálov patria funkcie menom Walsh, Haar a Rademacher.



Rozklad (aproximácia) diskrétného signálu $x(nT)$, kde T je diskretizačná perióda s počtom vzoriek M podľa niektorej diskkrétnej ortogonálnej bázy, je daný vzťahom:

$$x(nT) = \sum_{k=0}^{M-1} y_k u(k, nT) \quad n = 0, 1, \dots, M-1$$

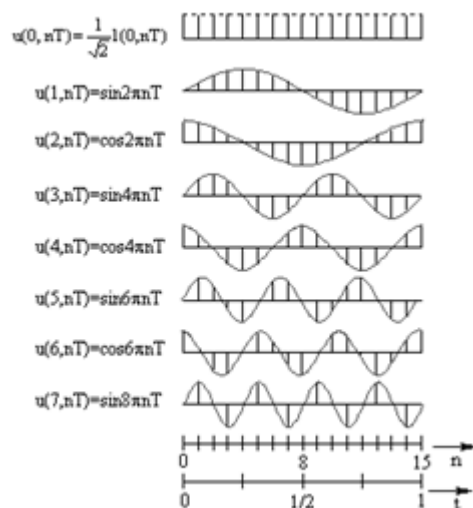
pričom optimálne koeficienty:

$$y_k = \sum_{n=0}^{M-1} x(nT) u(k, nT)$$

Pri harmonických signáloch je parametrom funkcie frekvencia. V prípade neharmonických signálov sa používa pojem sekvencia, ktorá je určená počtom

priesečníkov nulovej úrovne za sekundu. V súvislosti s diskretnými funkciami sa sekvencia určí z počtu zmien znamienka vzoriek za sekundu.

Diskrétné ortogonálne bázové funkcie môžeme získať diskretizáciou funkcií spojitej ortogonálnej bázy harmonických funkcií.



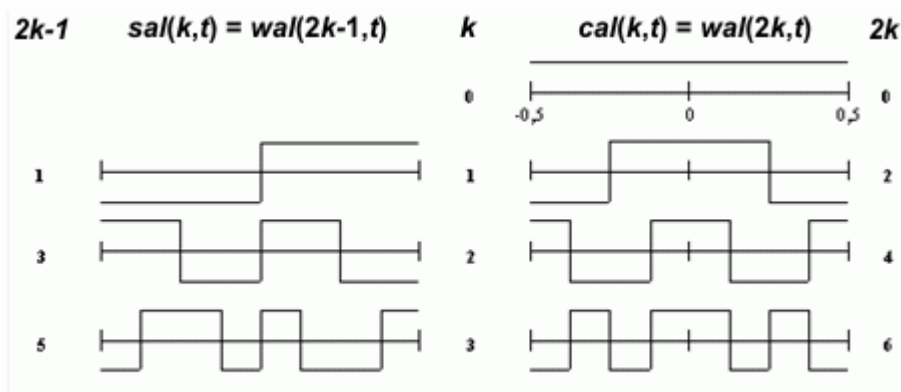
System Walshových funkcií



Walshove funkcie tvoria usporiadanú množinu pravouhlých impulzov, ktoré majú len dve možné hodnoty amplitúdy (+1 alebo -1) a predstavujú úplný systém ortogonálnych funkcií. Sú závislé od dvoch argumentov, a to času (t) a od poradového čísla (k). Značíme ich $wal(k, t)$ a môžeme rozlišovať párne (kosínusové-Walshove) $cal(k, t)$ a nepárne (sínusové-Walshove) $sal(k, t)$ funkcie, ktoré sú dané:

$$cal(k, t) = wal(2k, t), \quad k = 0, 1, 2, \dots$$

$$sal(k, t) = wal(2k-1, t), \quad k = 0, 1, 2, \dots$$



Obr. 2.6: Príklad párných a nepárných bázových Walshových funkcií

Podľa usporiadania Walshových funkcií tieto delíme na tri skupiny:

1. usporiadanie Walshove (sekvenčné) $walw(k,t)$
2. usporiadanie Paleyho (dyadické) $walp(k,t)$
3. usporiadanie Hadamardove (prirodzené) $walh(k,t)$

Jednotlivé usporiadania majú svoje realizačné opodstatnenia. Bázy obsahujú tie isté funkcie, ktoré sú však odlišne zoradené. Ako príklad je možné uviesť Hadamardovo (prirodzené) usporiadanie.



Výhodou tohto usporiadania je jednoduchá tvorba bazových funkcií väčších rozmerov.

Na obrázku nižšie je prvých osem spojitých aj diskretných Walshových funkcií s prirodzeným usporiadaním, pričom maticový zápis diskretných Walshových funkcií dáva v tomto prípade Hadamardovú maticu $U_h(3)$ s rozmerom 8×8 .



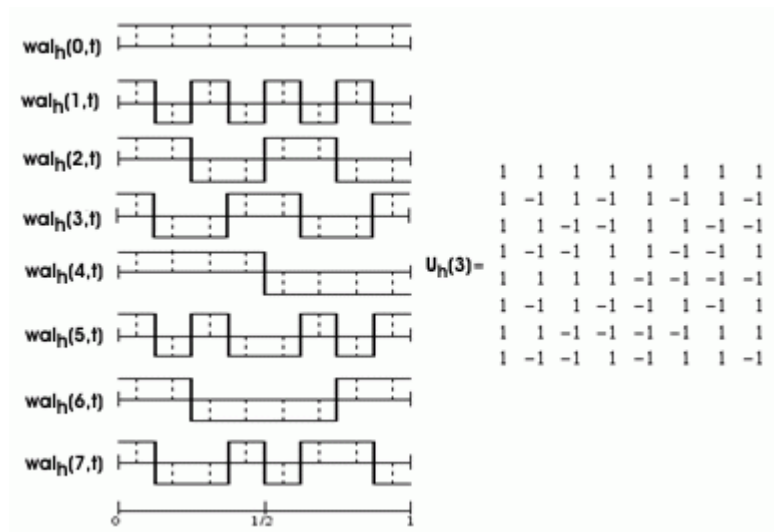
Vo všeobecnosti Hadamardovú maticu $U_h(r)$ s rozmerom $M \times M$, kde $M=2r$, môžeme vypočítať pomocou kroneckerovského (priameho) súčinu matic z $U_h(r-1)$.

$$U_h(r) = \begin{bmatrix} 1 & 1 \\ 1 & -1 \end{bmatrix} \otimes U_h(r-1) = \begin{bmatrix} U_h(r-1) & U_h(r-1) \\ U_h(r-1) & -U_h(r-1) \end{bmatrix}$$

kde $U_h(0)=1$



Kroneckerovský súčin, označený ako $\otimes$, je operácia s dvomi maticami ľubovoľných rozmerov, ktorej výsledok je matica s maticovými prvkami.



Obr. 2.7: Prirodzené usporiadanie spojitých, diskretných a matice Walshových funkcií

Existujú aj ďalšie transformácie obsahujúce harmonické bazové funkcie. Okrem diskretnej Fourierovej transformácie sem patrí *diskrétna kosínusová transformácia* (*discrete cosine transformation*, **DCT**), *diskrétna sínusová transformácia* (*discrete sinus transformation*, **DST**), *diskrétna Hartleyho transformácia* (*discrete Hartley transformation*, **DHYT**).

3 Analógové a digitálne technológie

3.1 Multiplexovanie

Multiplex spája niekoľko analógových alebo digitálnych signálov do jedného telekomunikačného spoju. Multiplexer (alebo mux) sa využíva najmä na zvýšenie objemu dát, ktoré sa môžu preniesť v danom pásme s určitou šírkou a v určitom čase.

Naopak, demultiplexer (alebo demux) je zariadenie s jedným vstupom, ktoré vezme vstupný signál a rozdelí ho do viacerých výstupných signálov. Multiplexer sa zvyčajne používa s doplnkovým demultiplexerom na strane prijímača.

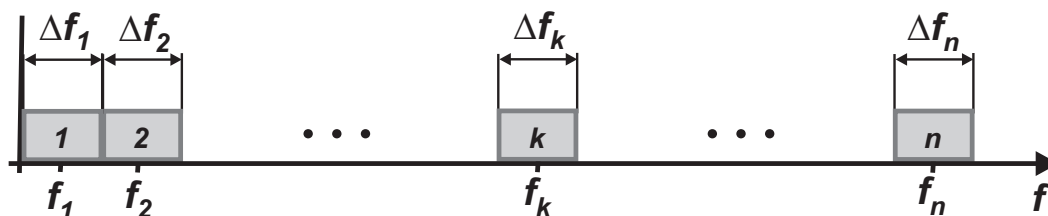
Frekvenčný multiplex

Frekvenčný multiplex (Frequency division multiplexing, FDM) je najmä analógová technológia. Frekvenčný multiplex spája niekoľko signálov do jedného spoločného zdroja tak, že sa vysielajú v odlišných frekvenčných pásmach.

- + Dôvodom využívania systémov s frekvenčným rozdeľovaním (frekvenčným multiplexom) je možnosť súčasného vysielania viac signálov z rôznych kanálov, ako aj fakt, že môžu byť vysielané signály z rôznych priestorovo odľahlých miest.
- Nedostatkom je vzájomné ovplyvňovanie kanálov, ktoré vzniká prekrývaním spektier signálov, neideálnosťou filtrov a objavením sa neželaných frekvenčných zložiek vplyvom presluchoch v elektrických obvodoch.



Pri frekvenčnom multiplexe sú jednotlivým signálom na frekvenčnej osi pridelené navzájom sa neprekrývajúce frekvenčné pásma $\Delta f_1, \Delta f_2, \dots, \Delta f_n$. Obyčajne $\Delta f_1 = \Delta f_2 = \dots = \Delta f_k = \Delta f_n$. Spektrá signálov príslušných kanálov sa musia zmestiť do vyhradených frekvenčných pásiem Δf_n .



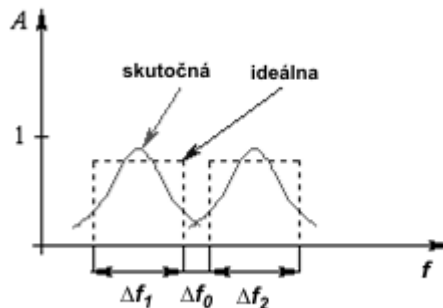
Obr. 3.1: Rozdelenie frekvenčného spektra príslušným kanálom

Frekvenčné pásmo prenosovej cesty F určujú hraničné frekvencie kanálov, čiže minimálna frekvencia intervalu Δf_1 (Δf_{1min}) a maximálna frekvencia intervalu Δf_n (Δf_{nmax}).

$$F = \Delta f_{\max} - \Delta f_{\min}$$

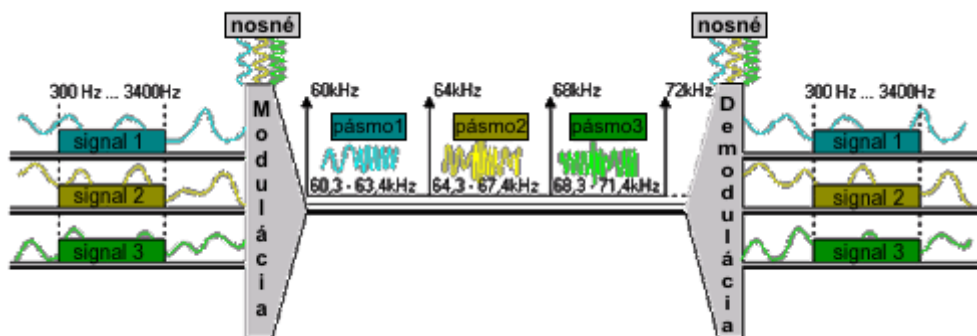
Dôležitou súčasťou frekvenčného multiplexu na strane vysielača aj prijímača sú pásmové filtre (PP), ktoré na vysielačnej strane ohraničujú frekvenčné pásma príslušných kanálov. Na prijímačnej strane zase filtre rozdeľujú signály, ktoré po prechode demodulátorom môžu byť prijaté prijímacím prevodníkom. Pri výbere frekvenčného pásma filtrov sa predovšetkým berie do úvahy požadovaná rýchlosť prenosu systému.

V reálnych filtroch útlm za hranicami frekvenčného pásma sa nemení skokom, ale spojito, preto sa medzi frekvenčnými pásmami jednotlivých filtrov vynecháva ochranné frekvenčné pásmo Δf_0 . Pomer $\Delta f_0 / \Delta f = k_f$ závisí od amplitúdovo-frekvenčných charakteristík použitých pásmových filtrov (Obr. 3.2) (viac o filtroch v kapitole Filtre).



Obr. 3.2: Amplitúdovo frekvenčné charakteristiky (ideálna a skutočná) pásmových filtrov

Prostredníctvom modulácie rôznymi signálmi (nosný signál, nosná, angl. carrier) sa posúvajú jednotlivé (napr. hovorové) pásma do subpásiem a až tak sa prenášajú. Po demodulácii na prijímačnej strane sú k dispozícii jednotlivé hovorové pásma vo svojej pôvodnej forme. Týmto postupom sa dajú prenášať viaceré hovorové pásma súčasne prostredníctvom jedného vedenia.



Obr. 3.3: Princíp frekvenčného multiplexu

Frekvenčný multiplex sa používal najmä v analógových telekomunikačných systémoch. V súčasnosti ho využívajú hlavne TV spoločnosti (každý kanál je na

mierne odlišnej frekvencii – prepínaním kanálu na TV preladíme na iné frekvenčné pásmo) a v optických komunikáciách.

Časový multiplex

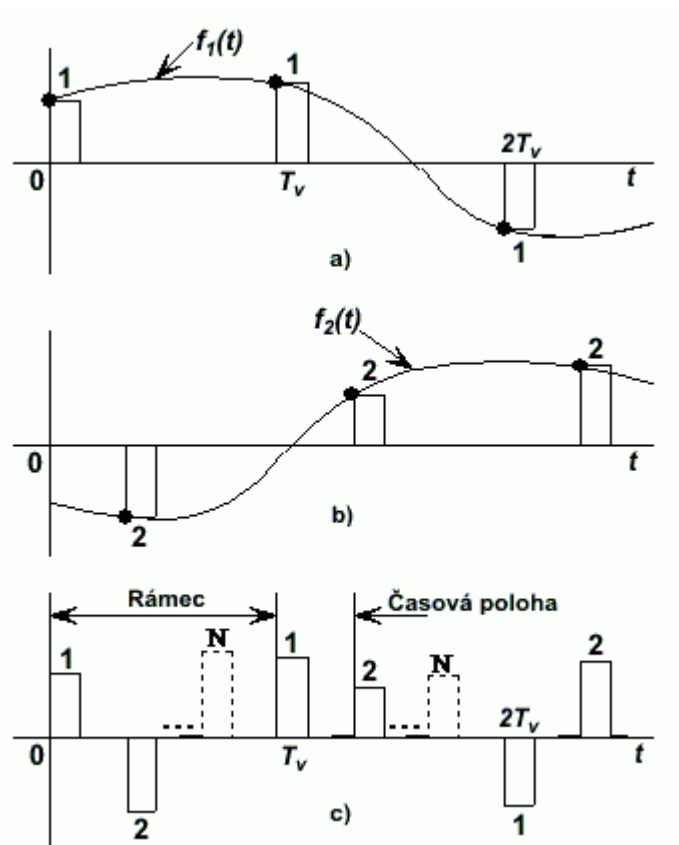


Časový multiplex (Time division multiplex TDM) sa využíva v sieťových aj telefónnych systémoch a robí presne to, čo hovorí samotný názov: periodicky vyberie vzorky z niekoľkých signálov s nižšou rýchlosťou, spojené ich prenesie cez jeden prenosový kanál s vyššou rýchlosťou a na konci obnoví pôvodný signál. Vzorky vyberá periodicky na základe času.

Vstupné zariadenie, multiplexer, vyberá vzorky z rôznych signálov, spojí ich do jedného a vysiela ich jedným kanálom. Týmto spôsobom je viacero vzoriek z rôznych zdrojov spojených a preložených do vysokorýchlostného kanála. Toto je možné realizovať vďaka tomu, že zdroje vysielať dáta relatívne malou rýchlosťou (ako 300 baud) a výsledný kanál združuje rýchlosti zo všetkých zdrojov (1200 baud). Na konci prenosovej cesty iný multiplexer rozdelí prenášané dáta a pošle ich do cieľa, znova tou istou menšou rýchlosťou akou vstúpili do TDM systému.

Túto technológiu využívajú najmä telefónne spoločnosti, ktoré musia prepojiť obrovské množstvo telefónnych hovorov cez obmedzený počet káblov. V prípade, že hovor je rozdelený na vzorky a obnovený rýchlejšie ako to dokáže ľudské ucho zistiť, nikto to nezistí. Práve preto telefónne prenosy využívajú TDM s vysokorýchlostnými nosičmi, a takto prenesú niekoľko hovorov naraz.

— Jedna nevýhoda multiplexov využívajúcich časový multiplex je, že pridávajú časové polohy aj vtedy, keď zdroj nič nevysiela. To vedie k neefektívnosti.



Obr. 3.4: Časový multiplex. a) informácia 1, b) informácia 2, c) multiplexované postupnosti impulzov

Synchronizácia

Synchronizácia vzorkovacieho procesu je nevyhnutná na to, aby sa udržali správne pozície postupností impulzov. Tá sa zabezpečuje postupnosťou hodinových impulzov, ktorá je referenčná pre všetky vzorkovacie obvody. V prijímači musí existovať podobná postupnosť hodinových impulzov, ktorá zabezpečí správne preberanie jednotlivých impulzov príslušných informácií. Inak by prijímač/demultiplexer nevedel určiť, do akého cieľového kanála patria jednotlivé zdrojové signály a ako naspäť signál zrekonštruovať. Postupnosti hodinových impulzov vo vysielacom a prijímači musia mať definované fázové pomery a vzájomne sa synchronizujú. Časová synchronizácia je preto nevyhnutne dôležitá pre správne fungovanie TDM technológií.



Synchronizácia hodinových impulzov sama o sebe ešte negarantuje správne preberanie informácií. Nevyhnutná je k tomu ešte aj rámcová synchronizácia. Pre synchronizáciu hodinových impulzov aj rámcovú synchronizáciu sa zvyčajne vymedzuje jedna časová poloha, v ktorej sa umiestnia špeciálne impulzy (napr. väčšie ako je predpokladaná najväčšia hodnota v nasledujúcej časovej polohe). Začiatok každého rámca je tak jednoznačne detekovateľný jednoduchými prahovacími obvodmi.

3.2 Lineárny diskretný časovo – invariantný systém

Lineárny, konečný, časovo – invariantný systém (LKI) sa priamo využíva v seizmológií, spracovaní signálov, obvodoch, teórii kontroly a iných technických oblastiach. Analýza spojitého LKI a diskretného LKI je veľmi podobná. Diskrétna oblasť je menej technicky náročná, preto sa zameriame na tieto systémy.



LDKI (*Lineárny, diskretný, konečný, časovo – invariantný systém*) má na vstupe a na výstupe jeden signál, pre ktoré platia nasledujúce vlastnosti:

- Systém je lineárny. To znamená, že ak vstup $x_1(n)$ a $x_2(n)$ generuje výstup $y_1(n)$ a $y_2(n)$ a ak a_1 a a_2 sú konštanty, potom vstupný signál $a_1x_1(n) + a_2x_2(n)$ generuje výstupný signál $a_1y_1(n) + a_2y_2(n)$.
- Systém je nemenný v čase. To znamená, že ak vstupný signál $x(n)$ generuje výstupný signál $y(n)$, potom pre každé reálne číslo s platí, že posunutý vstupný signál v čase $\hat{x}(t) = x(t - s)$ generuje posunutý výstupný signál v čase $\hat{y}(t) = y(t - s)$.

Diferenčná rovnica

Prvý dôležitý poznatok o správaní sa lineárneho, diskretného, konečného, časovo invariantného systému je, že odpoveď systému na ľubovoľný vstup je presne daný odpoveďou systému na jeden špecifický vstup v čase 0, a tým je Kroneckerov impulz (definovaný v kapitole Dôležité signály). Túto odozvu systému nazývame impulzovou odpoveďou. Teraz si zadefinujeme vzťah, ktorý definuje výstup generovaný vstupným signálom $x(n)$.



$$y(n) = \sum_{k=0}^N a_k x(n-k) - \sum_{k=1}^N b_k y(n-k)$$

Rovnica opisuje rekurzívny LDKI (so spätnou väzbou). Vzorky výstupného signálu sú dané ako lineárna kombinácia váhovaných vzoriek vstupného signálu. Váhové koeficienty sú označené ako a_k a b_k . Systémy opísané touto rovnicou nazývame systémy s nekonečnou impulzovou odpoveďou (*infinite impulse response*, **IIR**).

Špeciálnym prípadom je diferenčná rovnica pre nerekurzívny systém. V tomto prípade vstupný signál závisí iba od vzoriek vstupného signálu, nie je závislý od predchádzajúcich vzoriek výstupného signálu. Systém je s konečnou impulzovou odpoveďou (*finite impulse response*, **FIR**) a je definovaný ako:



$$y(n) = \sum_{k=0}^N a_k x(n-k)$$

Konvolúcia



Konvolúcia je ďalšia možnosť, ako opísať LDKI systém. Ak poznáme impulzovú odpoveď systému označenú ako $h(n)$ a vstupný signál je daný ako $x(n)$, potom výstupný signál môžeme vyjadriť ako:

$$y(n) = \sum_{k=0}^{D_y-1} x(k)h(n-k) = x(n) * h(n)$$

Kde operátor $*$ je konvolučný súčin. Dĺžka výstupného signálu je definovaná vzťahom $D_y = D_x + D_h - 1$ kde D_x je dĺžka vstupného signálu a D_h je dĺžka impulzovej odpovede.

Princíp činnosti je založený na princípe superpozície lineárnych systémov. Výstupný signál sústavy je daný súčtom vážených a posunutých impulzových charakteristík. Pre správne porozumenie si preštudujte nasledujúci príklad.



Majme FIR systém s impulzovou odpoveďou $h(n) = \{1, 2, 3\}$. Na jej vstup privedieme signál $x(n) = \{x(0), x(1), x(2), x(3)\}$.

Výstupný signál určíme pomocou konvolúcie.

Dĺžka impulzovej odpovede $h(n)$ je $D_h = 3$ a dĺžka vstupného signálu $x(n)$ je $D_x = 4$, potom dĺžka výstupu je $D_y = 6$. Konvolúciu veľmi jednoducho vypočítame pomocou tabuľky.

Konvolúcia

n	0	1	2	3	4	5	6
x(0)	x(0)	2 x(0)	3 x(0)	0	0	0	0
x(1)		x(1)	2 x(1)	3 x(1)	0	0	0
x(2)			x(2)	2 x(2)	3 x(2)	0	0
x(3)				x(3)	2 x(3)	3x(3)	0
y(n)	y(0)	y(1)	y(2)	y(3)	y(4)	y(5)	0

V jednotlivých riadkoch tabuľky sa nachádzajú príslušné vzorky vstupného signálu vážené impulzovou charakteristikou. Posúvanie riadkov smerom doprava zodpovedá oneskoreniu príslušnej vzorky vstupu. V poslednom riadku sú vzorky výstupného signálu a dostaneme ich superpozíciou hodnôt v jednotlivých stĺpcoch pre $n = 0, 1, 2, \dots$ Napríklad:

$$y(1) = 2 \cdot x(0) + x(1)$$

$$y(2) = 3 \cdot x(0) + 2 \cdot x(1) + x(2), \text{ atď.}$$

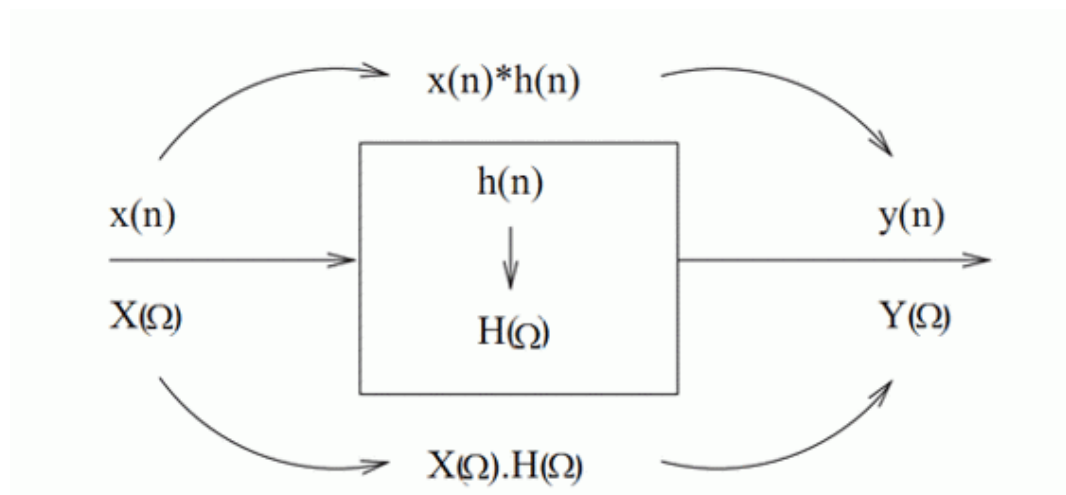
Konvolučný súčin dvoch postupností dostaneme tak, že jednu z nich usporiadame v opačnom poradí a potom ju podsúvame pod druhú sprava a v každom kroku určíme súčet vzniknutých súčinov.

Prenosová funkcia

Prenosová funkcia $H(\omega)$ reprezentuje vzťah medzi vstupným a výstupným signálom LDKI systému pri nulových počiatočných podmienkach vo frekvenčnej oblasti. Prenosová funkcia je odvodená z diferenčnej rovnice alebo impulzovej charakteristiky. V oboch prípadoch sa použije DFT na transformáciu do frekvenčnej oblasti. Ako bolo definované vyššie, platí $y(n) = h(n) * x(n)$ a po DFT dostaneme vzťah:



$$Y(\omega) = H(\omega) \cdot X(\omega) \Rightarrow H(\omega) = \frac{Y(\omega)}{X(\omega)}$$



Obr. 3.5: Súvis medzi vstupným, výstupným signálom a ich obrazmi

Je zrejmé, že $h(n)$ a $H(\omega)$ charakterizujú ten istý systém v rôznych doménach. Ak aplikujeme DFT na diferenčnú rovnicu, prenosová funkcia bude opísaná pomocou váhových koeficientov a_k a b_k , ktoré sú identické s koeficientmi v diferenčnej rovnici:

$$H(\omega) = \frac{a_0 + a_1 e^{-j\omega} + a_2 e^{-j2\omega} + \dots + a_N e^{-jN\omega}}{1 + b_1 e^{-j\omega} + b_2 e^{-j2\omega} + \dots + b_M e^{-jM\omega}}.$$

Vzťah zapísaný ako racionálna lomená funkcia je výhodnejší, pretože delením čitateľa menovateľom dostaneme priamo vzorky impulzovej odpovede. Pre IIR systémy je počet vzoriek nekonečný. V prípade FIR systému nie je prenosová funkcia v tvare zlomku, pretože menovateľ je rovný 1.

$$H(\Omega) = a_0 + a_1 e^{-j\Omega} + a_2 e^{-j2\Omega} + \dots + a_N e^{-jN\Omega}$$

Prenosová funkcia je veľmi dôležitá a pomocou $H(\Omega)$ sú určené frekvenčné charakteristiky. Tie sa využívajú najmä v teórii filtrov.

Frekvenčné charakteristiky

Frekvenčné charakteristiky určujú dynamiku systému.



Je to pomer amplitúdy (alebo magnitúdy) a fázy výstupu ako funkcie frekvencie k vstupu.

Jednoducho povedané, ak je na vstup systému privedená funkcia sínus s danou frekvenciou, odozva LDKI bude s tou istou frekvenciou, amplitúda a fáza budú mať hodnoty pomerné k vstupu.

Frekvenčná charakteristika prenosovej funkcie $H(\Omega)$ je definovaná:



$$H(\Omega) = |H(\Omega)| \cdot e^{j\varphi(\Omega)} = \text{Re}\{H(\Omega)\} + j\text{Im}\{H(\Omega)\}$$

$$\text{kde } |H(\Omega)| = \sqrt{\text{Re}\{H(\Omega)\}^2 + \text{Im}\{H(\Omega)\}^2} \quad \text{a} \quad \varphi(\Omega) = \arctan \frac{\text{Im}\{H(\Omega)\}}{\text{Re}\{H(\Omega)\}}.$$

Absolútna hodnota prenosovej funkcie $|H(\Omega)|$ sa nazýva magnitúdová frekvenčná charakteristika a $\varphi(\Omega)$ sa nazýva fázová frekvenčná charakteristika. $\varphi(\Omega)$ nie je spojitá funkcia, ale vykazuje 180 stupňové skoky. Ak odstránime tieto skoky, dostaneme fázovú frekvenčnú charakteristiku $\Theta(\Omega)$, ktorá bude spojitá. Odstránenie skokov umožní zmena znamienka pri magnitúdovej frekvenčnej charakteristike vždy pri každom skoku fázovej frekvenčnej charakteristiky $\varphi(\Omega)$. Vzťah medzi amplitúdovou a magnitúdovou frekvenčnou charakteristikou je nasledovný:

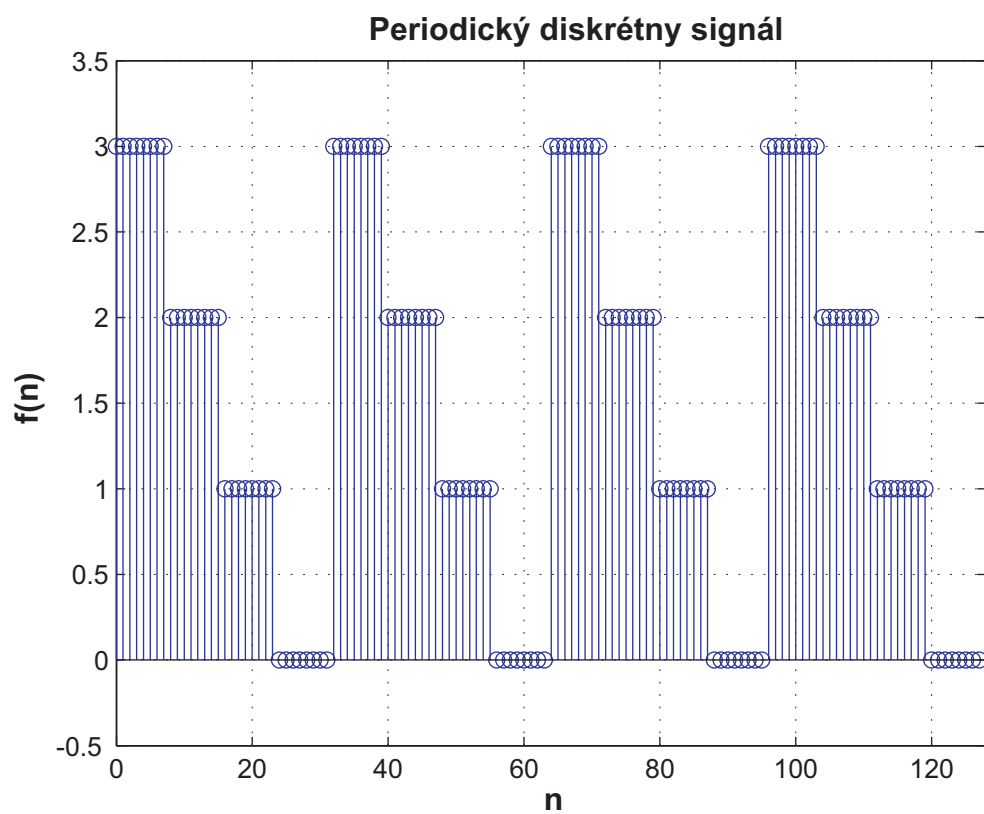
$$A(\Omega) = \pm M(\Omega).$$

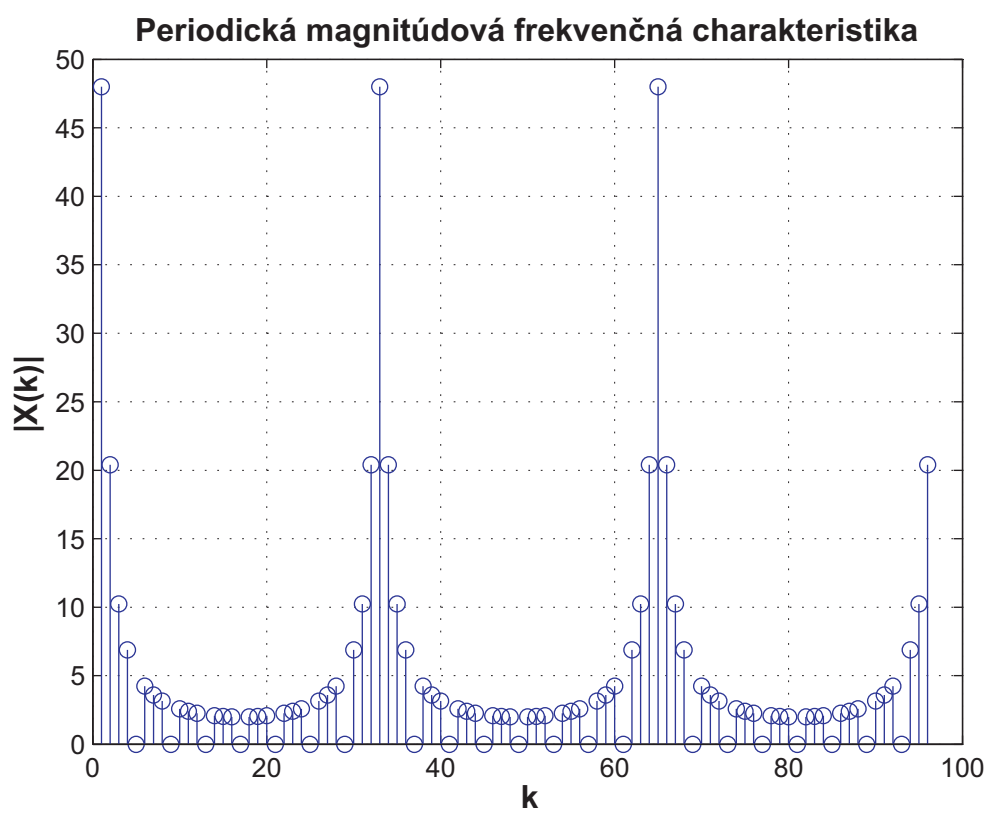
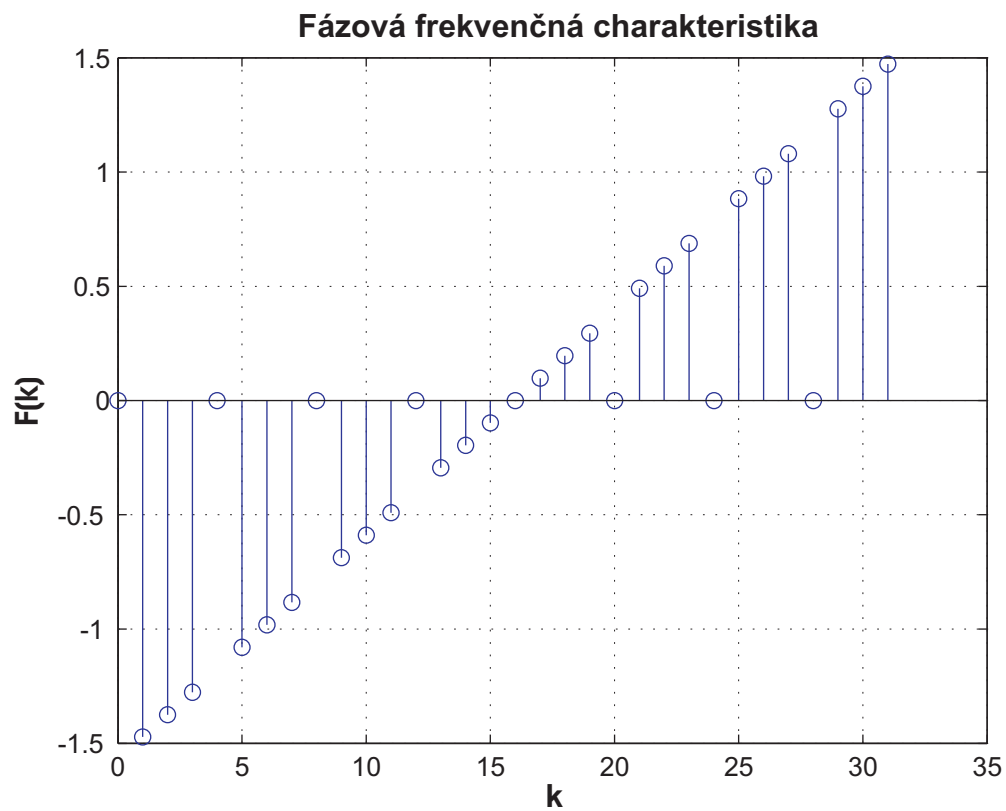
Zohľadnením informácií vyššie spomenutých môžeme definovať pre frekvenčné charakteristiky:

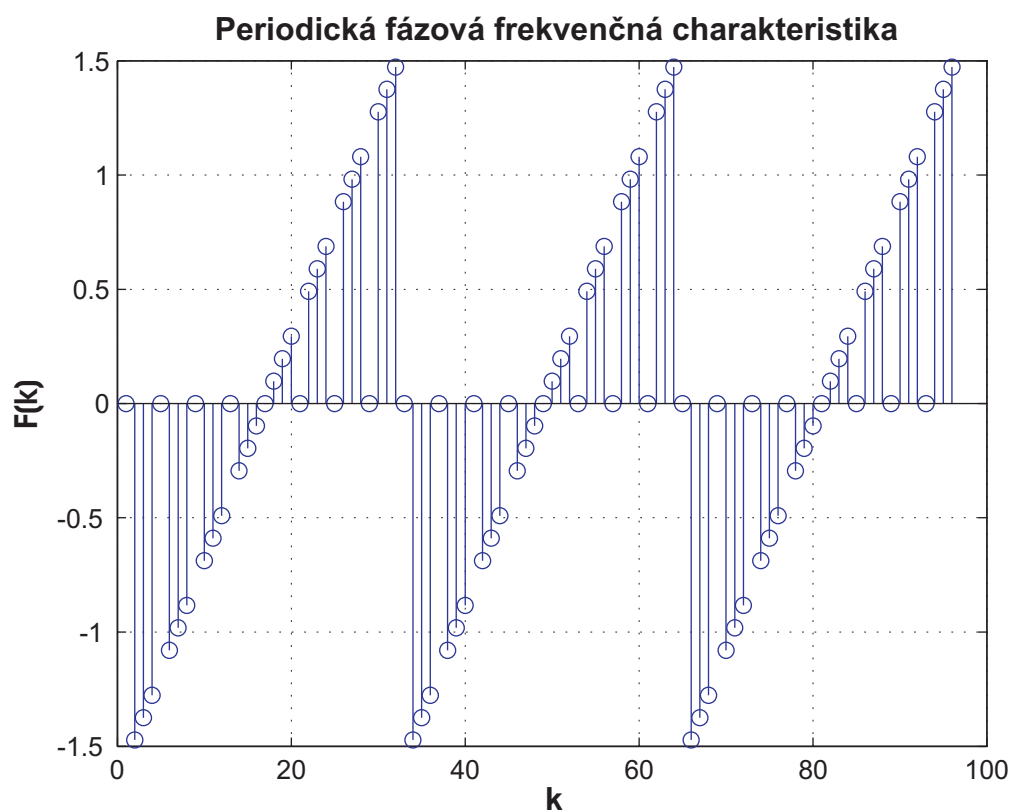


$$H(\Omega) = A(\Omega) \cdot e^{j\varphi(\Omega)}$$

$$H(\Omega) = M(\Omega) \cdot e^{j\varphi(\Omega)}$$







Obr. 3.6: Príklad prenosovej funkcie LDKI systému, magnitúdová a fázová frekvenčná charakteristika (jedna a viac periód)

3.3 Filtre



V spracovaní signálov, filter je zariadenie alebo proces, ktorý odstráni zo signálu nechcenú časť alebo vlastnosť.

Najčastejšie to znamená odstránenie niektorých frekvencií kvôli potlačeniu rušenia signálov a zníženiu šumu v pozadí. Avšak filtre nepracujú iba vo frekvenčnej doméne, obzvlášť v oblasti spracovania obrazu je viac veličín, ktoré sa môžu filtrovať.

Nedostatkom filtrácie je strata informácie spojennej s odfiltrovanou časťou. Kombinácia signálu s Fourierovým priestorom je alternatívny prístup na odstránenie iba určitých frekvencií zo signálu.



Je niekoľko rôznych klasifikácií filtrov, ktoré sa v mnohých ohľadoch prekrývajú. Neexistuje teda jednoznačné delenie, hierarchia. Filtre môžeme podľa rôznych kritérií rozdeliť na:

- analógové alebo digitálne
- diskkrétne v čase alebo diskkrétne v hodnote
- lineárne alebo nelineárne
- s nekonečnou impulzovou odpoveďou (IIR) alebo s konečnou impulzovou odpoveďou (FIR) patriace medzi diskkrétne a digitálne filtre.

Dôležité pojmy používané na klasifikáciu lineárnych filtrov:

- Dolnopriepustný filter – nízke frekvencie sa prepúšťajú, vysoké frekvencie sa tlmia.
- Hornopriepustný filter – vysoké frekvencie sa prepúšťajú, nízke frekvencie sa tlmia.
- Pásmový priepust – iba frekvencie z frekvenčného pásma sa prepúšťajú.
- Pásmový záver – iba frekvencie z frekvenčného pásma sa tlmia.
- Úzkopásmový filter – neprepúšťa iba jednu špecifickú frekvenciu, je to extrémny prípad pásmového záveru.

Filtre môžu byť implementované rôznymi technológiami. Tá istá prenosová funkcia sa môže realizovať viacerými spôsobmi. To znamená, že matematické vlastnosti a parametre sú rovnaké, ale fyzická realizácia je odlišná. Komponenty v jednotlivých technológiách sú však zvyčajne rovnaké a v príslušných filtroch spĺňajú tú istú úlohu.

- Elektronické filtre boli pôvodne vyrobené iba z pasívnych prvkov ako odpor, cievka a kondenzátor. Aktívne technológie zjednodušili výrobu a otvorili nové možnosti v špecifikácii filtrov.

- Digitálne filtre pracujú so signálmi definovanými v digitálnej podobe. Podstata digitálnych filtrov je v tom, že majú priamo v mikrokóde implementovaný matematický algoritmus zodpovedajúci požadovanej prenosovej funkcii.

Lineárne analógové filtre



Lineárny obvod opísaný spojitými funkciami je pravdepodobne najčastejšie myslený ako filter v oblasti spracovania signálov. Pre zjednodušenie sa zaviedol pojem filter. Úlohou týchto filtrov je určité frekvencie tlmiť a iné prepúšťať.

Taký filter musí byť nutne aj lineárny filter. Nelinearita by mohla spôsobiť to, že sa vo výstupnom signáli budú vyskytovať aj frekvencie, ktoré sa vo vstupnom signáli nenachádzali.

Najznámejšie skupiny lineárnych, časovo spojitých filtrov rozdelené podľa spôsobu návrhu sú:

- Chebyshev filter, má najlepšiu aproximáciu vzhľadom na ideálnu charakteristiku ľubovoľného filtra pre dané usporiadanie a zvlnenie.
- Butterworth filter, má maximálne vyrovnanú frekvenčnú charakteristiku.
- Bessel filter, má maximálne vyrovnané fázové oneskorenie.

Rozdiel medzi týmito skupinami filtrov je v tom, že každý typ používa inú polynomickeú funkciu na aproximáciu ideálnej frekvenčnej charakteristiky. Výsledkom sú rôzne prenosové funkcie. Takže podľa toho, akú frekvenčnú charakteristiku potrebujeme (zvlnenú, vyrovnanú) si vyberieme typ filtra.



Filtre sa používajú predovšetkým v oblasti telekomunikácii – sú nevyhnutnou súčasťou viacerých technologických pokrokov a boli zdrojom značných ziskov pre telekomunikačné spoločnosti. Nemalo by teda byť žiadnym prekvapením, že vývoj prvých filtrov bol úzko spojený s vývojom prenosových trás.

Digitálne filtre



V elektronike, výpočtovej technike a matematike sa za digitálny filter považuje systém, ktorý pracuje iba s navzorkovaným signálom (diskrétnym v čase). Jeho podstatou je zredukovať alebo zvýrazniť niektoré vlastnosti signálov.

Toto je hlavný rozdiel v porovnaní s analógovými filtermi, ktoré reprezentuje elektrický obvod pracujúci so spojitým signálom v čase. Analógový signál môže byť filtrovaný digitálnym filtrom v prípade, ak sa najskôr navzorkuje a je reprezentovaný ako postupnosť vzoriek. Potom môže prejsť digitálnym filtrom a späť sa zrekonštruovať na prefiltrovaný analógový signál. Pri analógových filtroch je vstupný signál priamo privedený do elektrického obvodu.



Digitálny filter je charakterizovaný svojou prenosovou funkciou alebo diferenčnou rovnicou. Matematickou analýzou prenosovej funkcie sa dá vyjadriť odpoveď na ľubovoľný vstup.

Filtre s konečnou impulzovou odpoveďou



Filter s konečnou impulzovou odpoveďou, nazývaný FIR – je filter, ktorý má konečnú impulzovú odpoveď, pretože sa po určitom čase ustáli na nule.

FIR filtre sú typické filtre bez spätnej väzby, takže výstupný signál závisí iba od vstupného signálu. Ak máme N vzoriek vstupného signálu, potom $N-1$ je stupeň filtra.



Diferenčná rovnica opisujúca systém v časovej oblasti:

$$y(n) = \sum_{k=0}^N a_k x(n-k)$$

Definícia prenosovej funkcie:

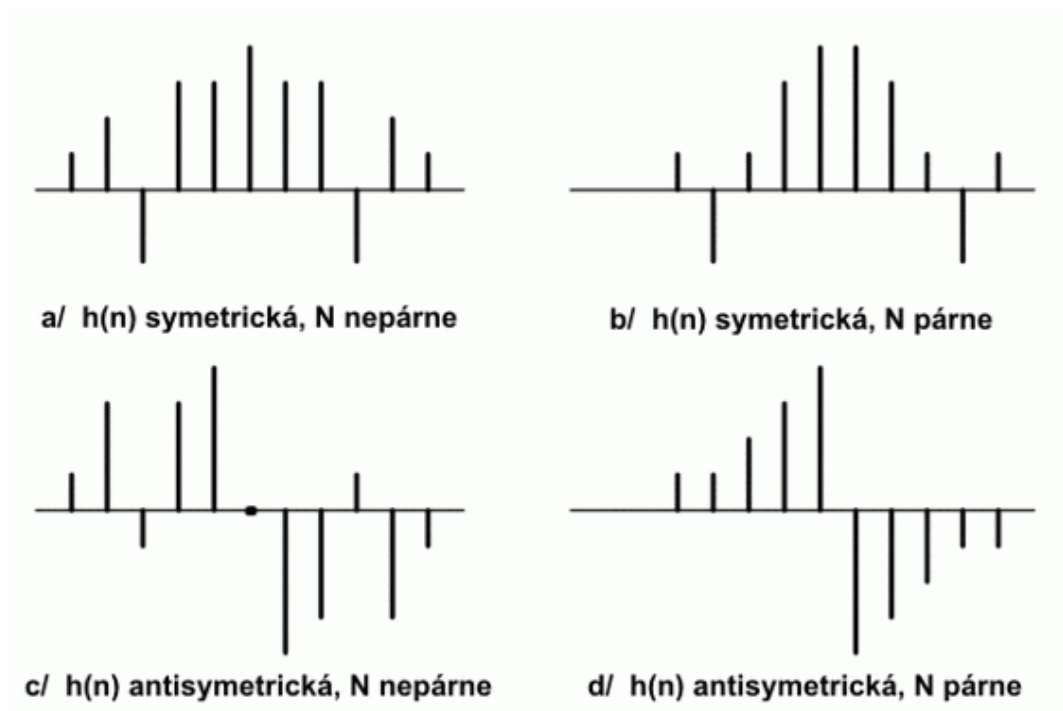
$$H(\Omega) = a_0 + a_1 e^{-j\Omega} + a_2 e^{-j2\Omega} + \dots + a_N e^{-jN\Omega}$$



Hlavné výhody FIR filtrov sú:

- Nedochádza k akumulácii chyby pri sčítavaní vzoriek, pretože nemajú spätnú väzbu.
- Jednoduchá implementácia.
- Stabilita. Sú bez spätnej väzby, takže všetky póly* sú umiestnené v začiatku (zjednodušene povedané „akoby póly neboli“). Systém je stabilný práve vtedy, ak absolútna hodnota každého pólu je menšia ako jeden.
- Je jednoduché zabezpečiť lineárnu fázovú charakteristiku. Tá je zabezpečená vtedy, ak je impulzová odpoveď systému symetrická alebo antisymetrická.

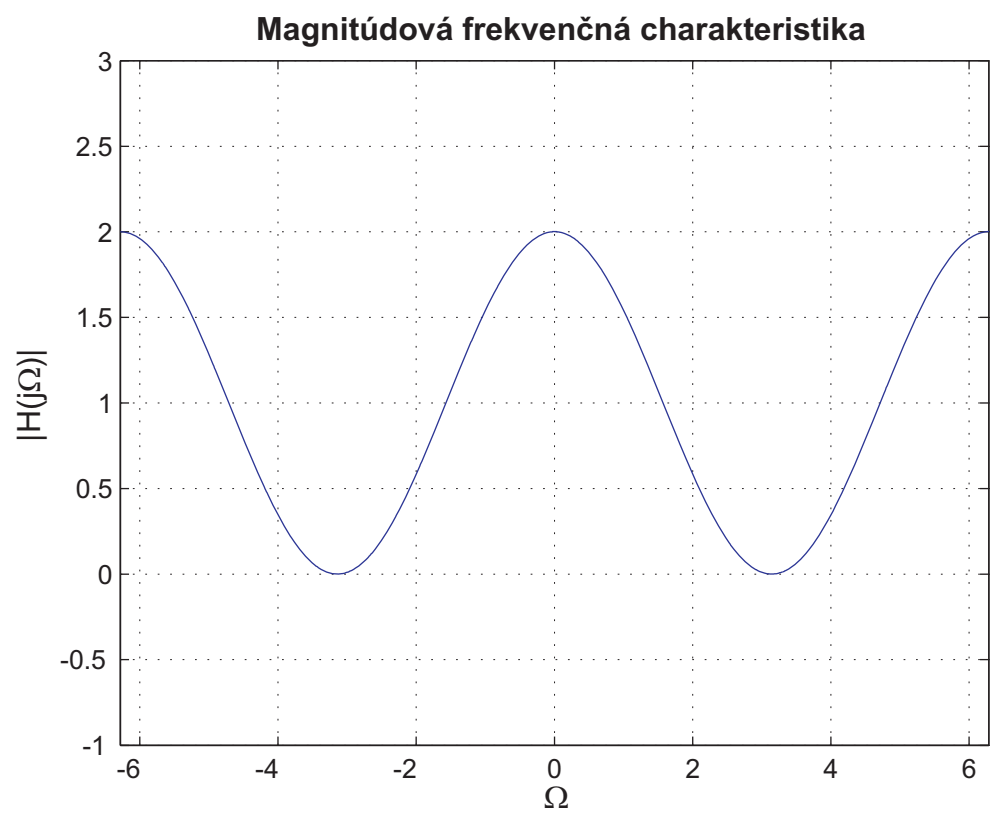
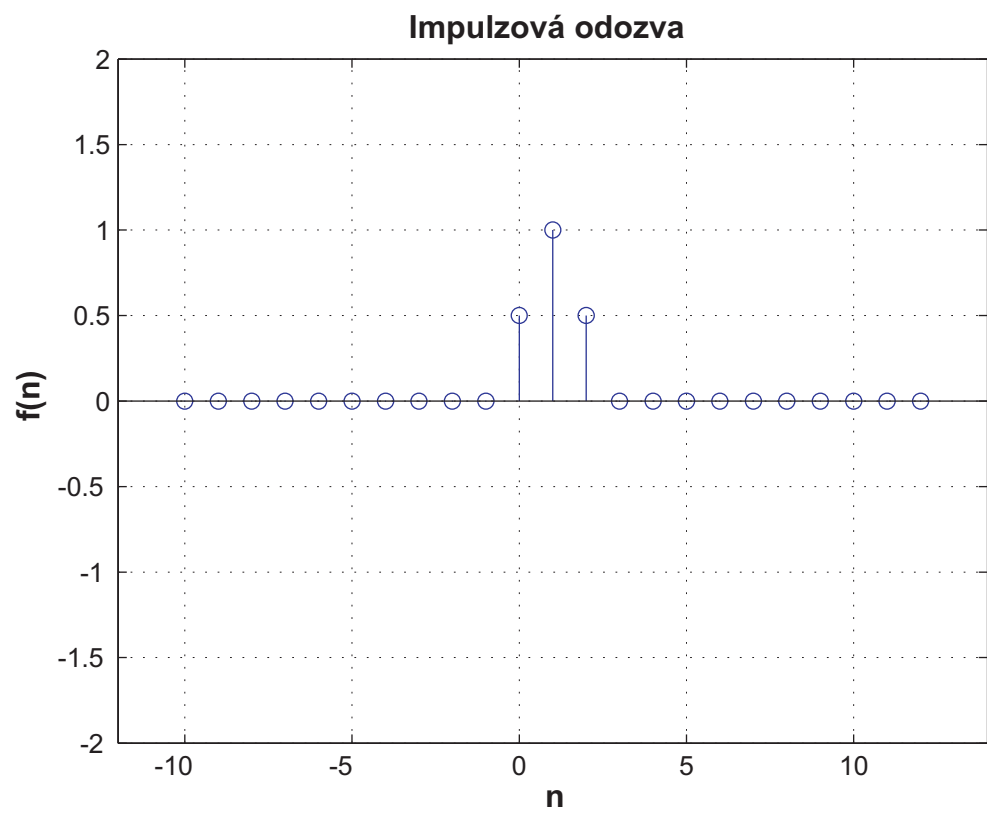
*Poznámka: póly sú korene menovateľa prenosovej funkcie.

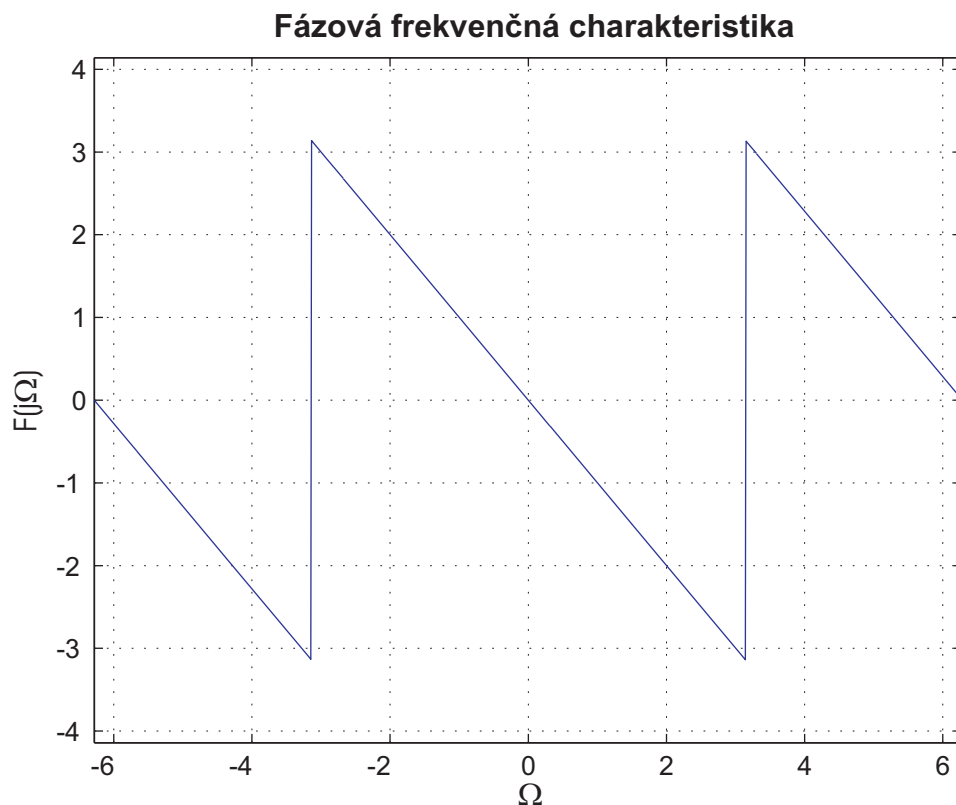


Obr. 3.7: Impulzová odpoveď (na zabezpečenie lineárnej impulzovej charakteristiky)

Ak sa povie navrhnuť filter, znamená to vybrať také koeficienty, aby mal filter požadované charakteristiky. Tie sú zvyčajne uvedené v špecifikácii filtra, ktorá najčastejšie predstavuje frekvenčnú charakteristiku filtra. Existuje niekoľko metód na hľadanie koeficientov, medzi najviac používané patria:

- Intuitívna metóda
- Metóda frekvenčného vzorkovania
- Metóda s oknovými funkciami
- Metóda vážením najmenších štvorcov





Obr. 3.8: Impulzová odpoveď, magnitúdová a fázová frekvenčná charakteristika FIR filtra

Filtre s nekonečnou impulzovou odpoveďou



Filter s nekonečnou impulzovou odpoveďou, nazývaný IIR – je filter, ktorého impulzová odpoveď nikdy nenadobudne nulovú hodnotu.

Vzorka výstupného signálu je daná sumou N vzoriek vstupného signálu, váhovanými koeficientmi a_k a výstupného signálu, váhovanými koeficientmi b_k . Je to zrejmé aj z diferenčnej rovnice opisujúcej IIR systém.



Diferenčná rovnica opisujúca IIR filter:

$$y(n) = \sum_{k=0}^N a_k x(n-k) - \sum_{k=1}^N b_k y(n-k)$$

a definícia prenosovej funkcie:

$$H(\Omega) = \frac{a_0 + a_1 e^{-j\Omega} + a_2 e^{-j2\Omega} + \dots + a_N e^{-jN\Omega}}{1 + b_0 + b_1 e^{-j\Omega} + b_2 e^{-j2\Omega} + \dots + b_M e^{-jM\Omega}}$$

Prenosová funkcia je daná podielom dvoch polynómov, preto stabilita sústavy nie je zaručená. Ako bolo spomenuté, systém je stabilný vtedy, ak absolútna hodnota každého pólu je menšia ako jeden. V prípade nestability existuje niekoľko

spôsobov, ako sústavu stabilizovať, napríklad pomocou fázovacích článkov alebo PLSI algoritmom.

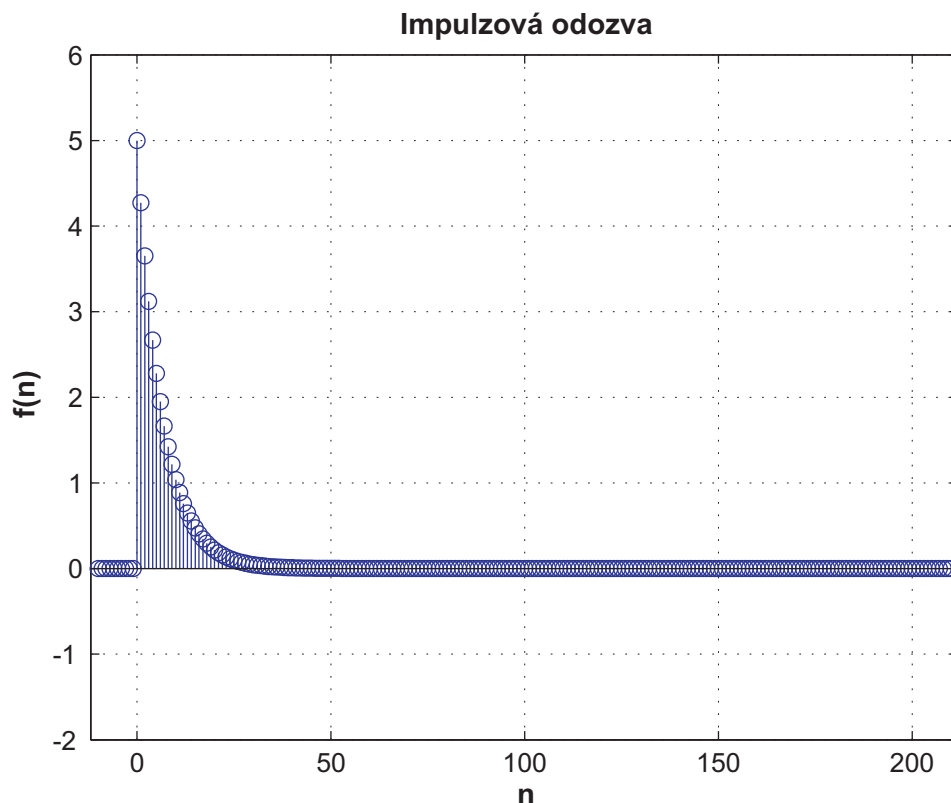
Pri návrhu filtra musíme zvoliť koeficienty tak, aby mal systém predpísanú frekvenčnú charakteristiku. V prípade IIR filtra to znamená určiť stupeň čitateľa, menovateľa a koeficienty a_k a b_k . Metódy návrhu IIR filtrov sa delia na dve skupiny.

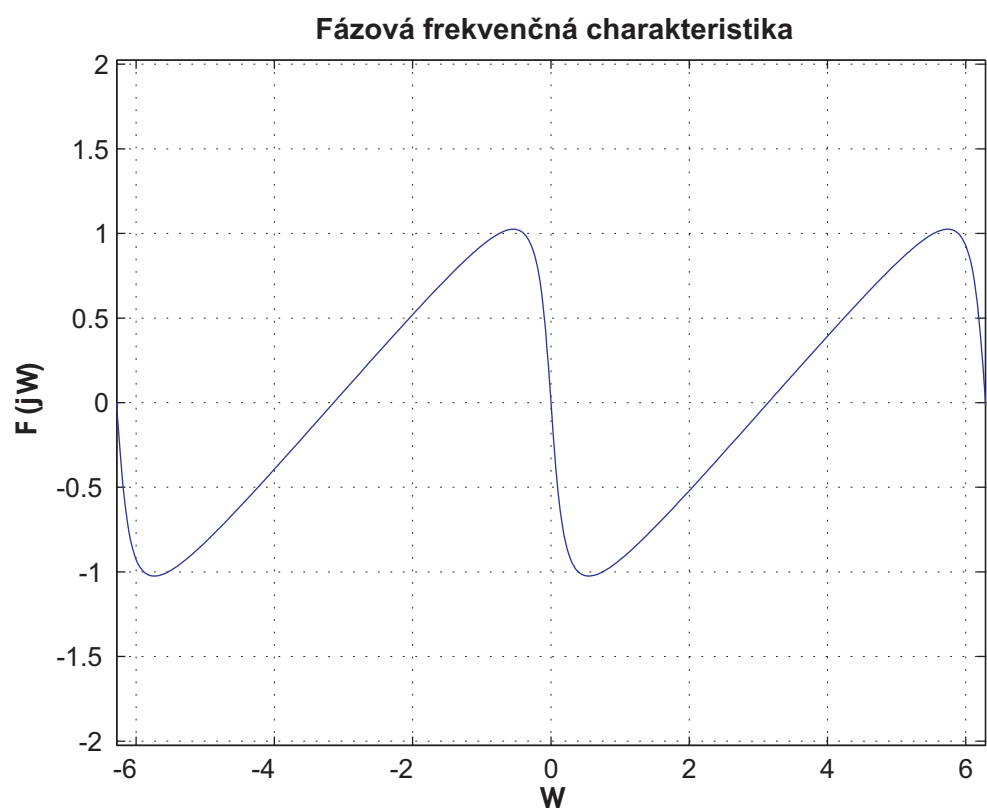
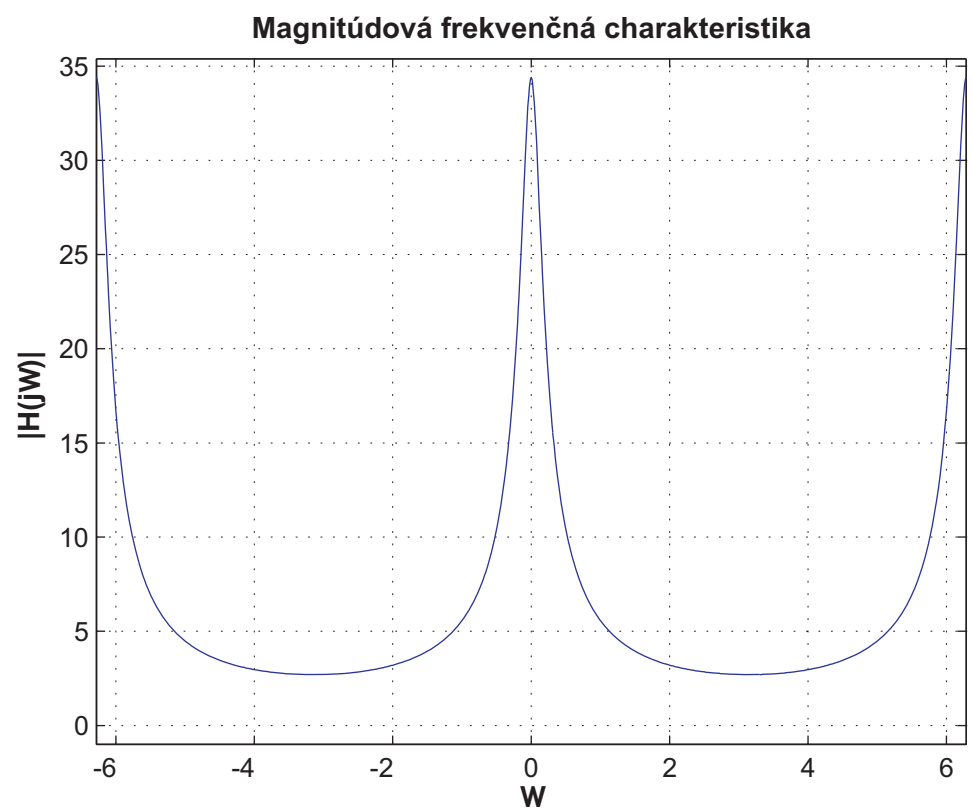
Do prvej skupiny patria priame metódy návrhu:

- Intuitívna metóda
- Metóda frekvenčného vzorkovania
- Pronyho metóda

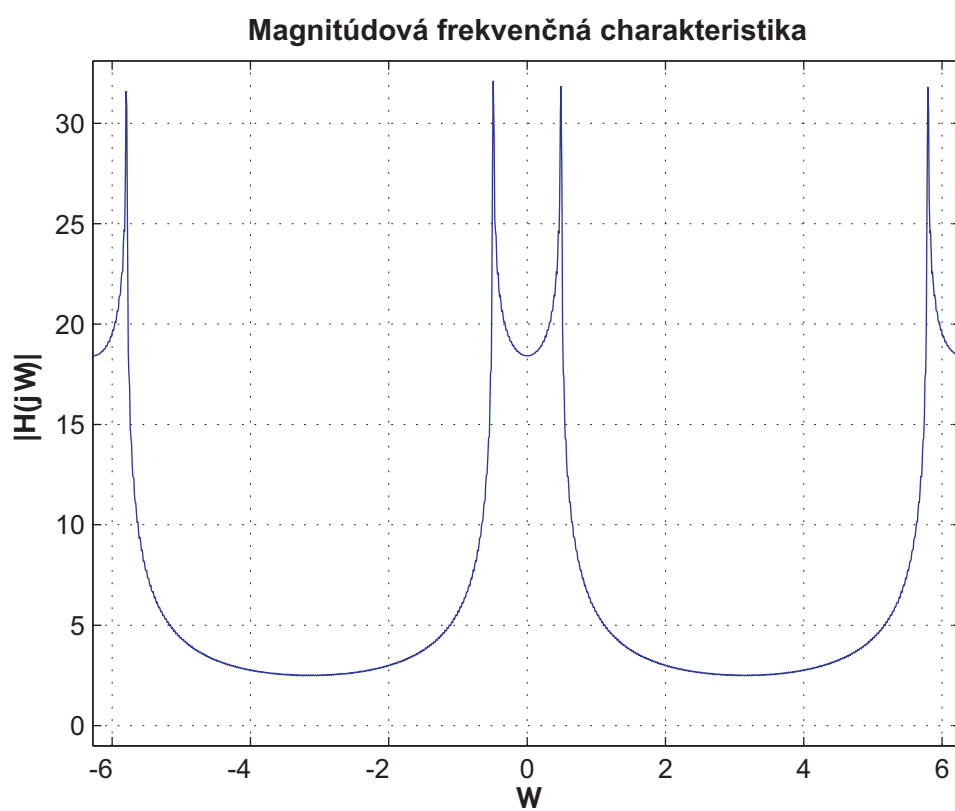
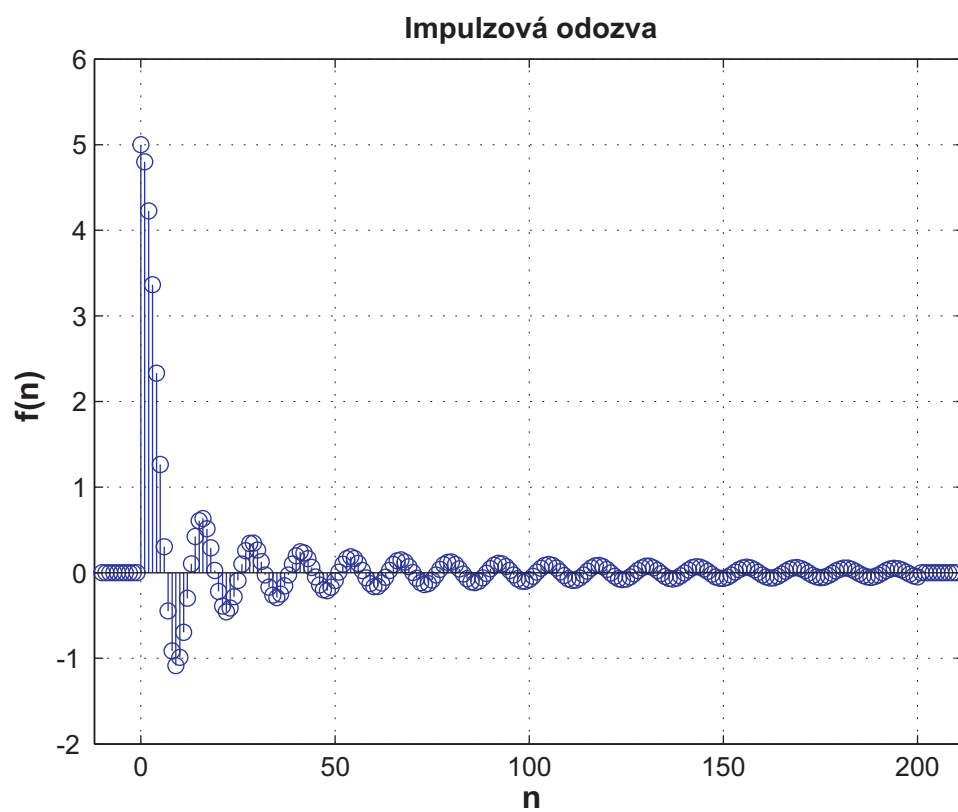
Druhá skupina návrhu sa volá nepriame metódy a vychádza z prepracovaných metód návrhu analógových filtrov. Pri návrhu sú formulované požiadavky na IIR filter. Potom sa navrhne analógový filter (Chebyshev filter, Butterworth filter, Eliptický filter), ktorý je pretransformovaný do digitálnej oblasti aplikovaním diskretizačných techník. Medzi najpoužívanéjšie patria:

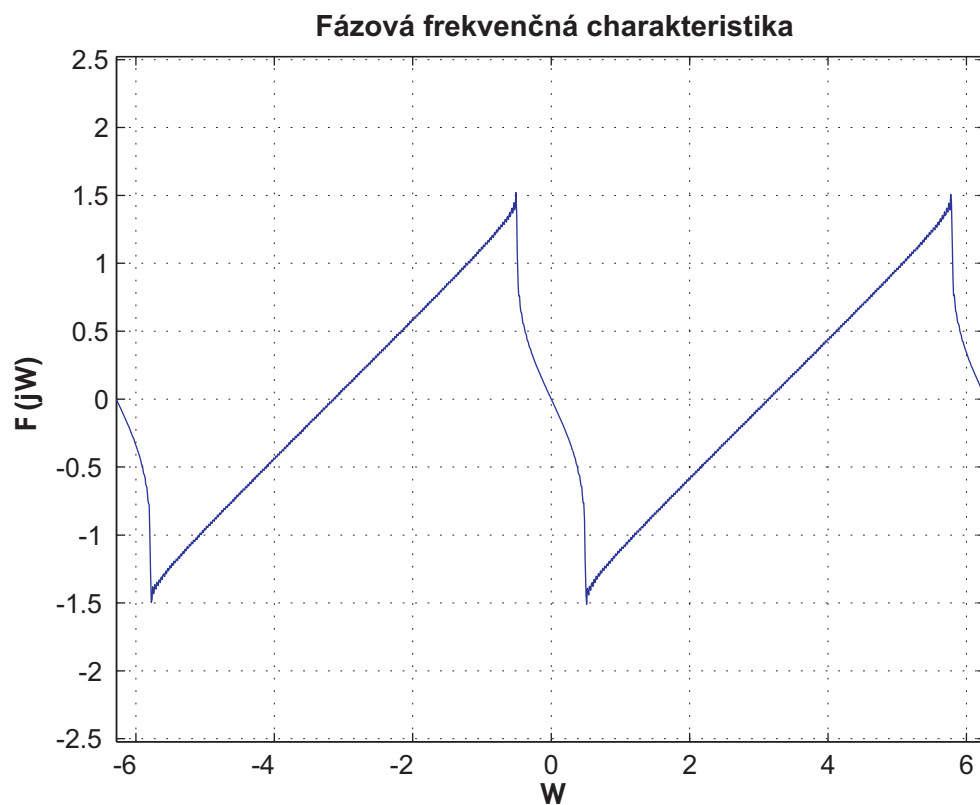
- Metóda bilineárnej transformácie
- Metóda invariantnosti impulzovej charakteristiky





Obr. 3.9: Impulzová odpoveď, magnitúdová a fázová frekvenčná charakteristika IIR filtra (exponenciálna funkcia)





Obr. 3.10: Impulzová odpoveď, magnitúdová a fázová frekvenčná charakteristika IIR filtra (sinc funkcia)

-
- +
 Hlavnou výhodou IIR filtrov je, že napriek rekurzivite majú menej členov a teda spotrebujú menej operačnej pamäte ako ekvivalentné FIR filtre. Preto sú výhodné v oblasti číslicového spracovania signálov.

 - Nevýhodou IIR filtrov je, že môžu byť nestabilné. Implementácia IIR filtrov je komplikovanejšia ako implementácia FIR filtrov.

4 Komunikačný kanál

4.1 Komunikačný kanál



Typický komunikačný systém obsahuje vysielací systém (vysielač), prijímací systém (prijímač) a prenosovú cestu, čo v súhrne tvorí telekomunikačný kanál. Telekomunikačný kanál potom predstavuje súbor technických prostriedkov umožňujúcich iba jednosmerný prenos signálu medzi dvoma miestami bez ohľadu na druh použitých prostriedkov.

Telekomunikačný okruh tvorí pár vzájomne priradených protismerných kanálov umožňujúci protismernú komunikáciu. Protismerná komunikácia môže byť buď simplexná (prenos signálu okruhom striedavo v jednom či druhom smere) alebo duplexná (prenos signálu okruhom súčasne v oboch smeroch).

Vzhľadom na signál prenášaný kanálom rozlišujeme dva druhy komunikačných systémov: analógový a digitálny. V nasledujúcej časti budú opísané oba systémy a ich základné funkčné bloky.

Analógový komunikačný kanál



Analógová komunikácia je spôsob prenášania dát, ktorý využíva spojitý signál na prenos údajov ako reč, obraz, video, atď.

Analógový prenos je jednoduchší ako digitálny a umožňuje prenášať informácie od jedného bodu k jednému alebo viacerým bodom. Na prenos dát analógovými signálmi sa využíva proces modulácie. Preto je nutné, aby vysielač aj prijímač mali kompatibilné zariadenia – vhodný modulátor na strane vysielača a demodulátor na strane prijímača.



Zariadenie, ktoré sa používa na vysielanie aj prijímanie signálu, obsahuje v sebe modulátor aj demodulátor a nazýva sa MODEM (*MODulator+DEModulator*).



Analógové obvody nezahŕňajú v sebe kvantizáciu informácie ako digitálne obvody, z čoho plynie ich veľká nevýhoda náhodných obmien chýb a poškodenie (degradácie) signálu, ktoré nevieme opraviť. Výsledkom je veľké zašumenie zvuku alebo videa. Platí, čím väčšia vzdialenosť, tým väčšie zašumenie.

Digitálny komunikačný kanál

Vysielač

Aj keď je to číslicový komunikačný systém, je schopný spracovať aj analógový signál (napríklad hlas), ktorý je *analógovo-číslicovým prevodníkom* (*analog-*

digital convertor, **ADC**) konvertovaný do číslicového tvaru. V tomto analógovo-číslícovom prevodníku prebieha proces digitalizácie signálu. Detailný postup je vysvetlený v kapitole Digitalizácia analógového signálu.

Ďalším blokom, ktorý už pracuje s digitálnym signálom je zdrojový kóder, ktorého všeobecnou vlastnosťou je, že efektívne konvertuje každý diskretný symbol do vhodnej číslicovej reprezentácie (zvyčajne binárnej). Je nutné poznamenať, že v praxi obvykle býva proces prvotného priradenia číslicovej reprezentácie súčasťou ADC prevodníka. Digitálna správa vo väčšine prípadov obsahuje redundanciu (nadbytočnosť).



Principiálnou vlastnosťou zdrojového kódera je, že odstraňuje redundanciu. Efektívnejší kóder účinnejšie odstraňuje redundanciu, čo v konečnom dôsledku vedie k menšiemu počtu bitov (binárnych číslic, „0“ a „1“) potrebných pre prenos informácie. Ako príklad odstránenia nadbytočnosti možno uviesť: „a tak ďalej“, v písanom texte sa často používa „atď.“.

V niektorých systémoch, ktoré nedisponujú funkciou kanálového kódovania, je priamo výstup zdrojového kódera transformovaný do tvaru, ktorý je vhodný pre prenos kanálom. Ako bolo uvedené pri zdrojovom kóderi, jeho úlohou je redukovať nadbytočnosť v digitálnom signáli. Naopak je to pri kanálovom kódovaní, ktorého úlohou je pridať nadbytočnosť, ktorá slúži na detekciu a opravu chýb. Bez redundantnej informácie by nebolo možné zistiť a odstrániť chyby spôsobené šumom a interferenciou. Najjednoduchšie formy redundancie sú opakovanie a parita.

Posledným zariadením na strane vysielateľa je modulátor.



V prípade digitálnej modulácie, analógový nosný signál je modulovaný diskretným signálom.

Druh použitej modulácie závisí od druhu kanála a prenášaných dát. Niektoré modulátory sú určené pre zašumené kanály (ako rádio alebo WIFI), niektoré sú určené pre kanály s minimálnym rušením (ako optické káble).

Prijímač



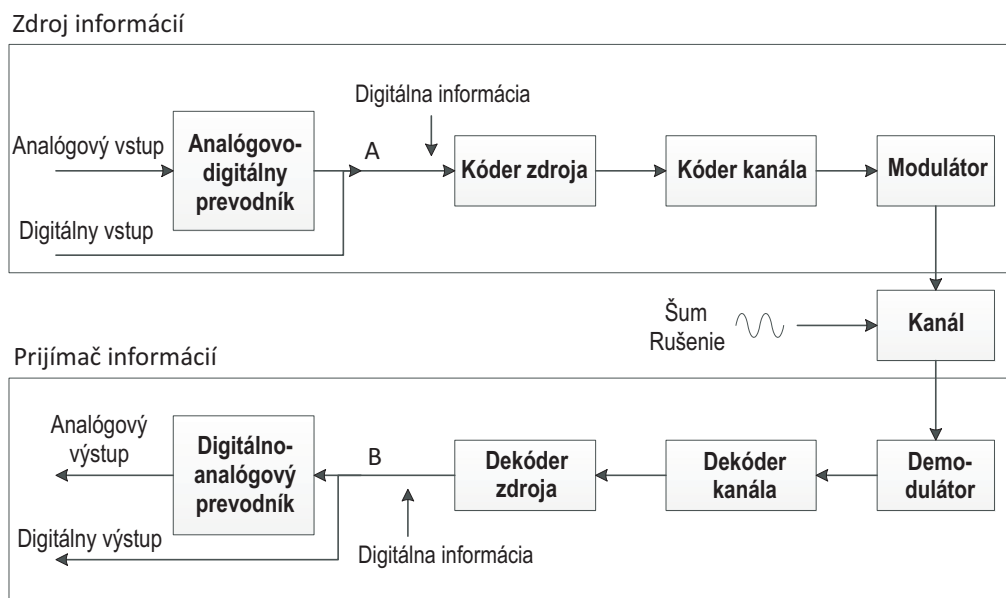
V prijímači sa realizujú inverzné operácie k operáciám vysielateľa. Veľmi dôležitou úlohou v návrhu a optimalizácii digitálneho komunikačného systému je minimalizovať chyby pri prenose.

V prijatom signáli sa zvyčajne vyskytujú chyby zapríčinené šumom kanála a tieto chyby musia byť opravené na strane prijímateľa. Toto je úlohou demodulátora, ktorý mení numerické hodnoty na hodnoty signálu (napr. číslo 1 zodpovedá hodnote signálu 5, kde 5 môže byť ampér, volt, atď.). Podľa toho, ako sú prenášané parametre (amplitúda, fáza a frekvencia) v nosnom signáli, sa vyberie spôsob demodulácie. Ak je signál modulovaný lineárnou moduláciou ako AM, použije sa synchronný detektor. V prípade uhlovej modulácie je potrebné použiť FM demodulátor alebo PM demodulátor. Tieto demodulátory sú realizované rôznymi elektrickými obvodmi.

Poslaním kanálového dekodéra je na výstupe rekonštruovať v čo najpresnejšom tvare výstup zdrojového kódéra vysielajú. V praxi sa vždy vyskytnú chyby v prenášanom signáli. Redundantná informácia pridaná kanálovým kódérom vysielajú mu umožňuje korigovať tieto chyby vnesené kanálom a demodulátorom (čo nie je možné pri analógovom prenose).

Zdrojový dekodér realizuje presne inverznú funkciu ako vykonával zdrojový kódér vysielajú, to znamená, že pridáva predtým odstránenú redundanciu. V prípade, ak výstup má zostať digitálny, potom výstup zo zdrojového dekodéra je výstupom celého prijímača.

Ak je potrebný analógový výstup (napr. reč) údaje zo zdrojového dekodéra postupujú do číslicovo-analógového prevodníka (*digital-analog convertor*, **DAC**). Číslicovo-analógový prevodník v súlade s teóriou vzorkovania rekonštruje pôvodnú analógovú správu.



Obr.4.1: Bloková schéma digitálneho komunikačného systému

5 Metódy kompresie



Cieľom kompresie je odstránenie nepodstatných alebo nadbytočných informácií z pôvodných dát tak, aby sme ich mohli ukladať alebo prenášať efektívnejšie: ušetríme tak napríklad prenosový kanál alebo ukladací priestor. Kompresiou kódujeme užitočnú informáciu menším počtom bitov, než bol použitý v pôvodnej reprezentácii.

Množstvo dát, ktoré v súčasnosti vytvoria multimediálne zariadenia, je obrovské. Ak by sme chceli uložiť 6-megapixelový obrázok (3000x2000 pixelov s 8 bitmi/farba) na US kľúč v nekomprimovanom tvare, potrebovali by sme vyše 17 megabajtov priestoru (presne 18 miliónov bajtov). Rovnaký obrázok skomprimovaný pomocou kodéra JPEG by zabral okolo 3 megabajtov ukladacieho priestoru (asi 3 200 000 bajtov) a nepostrehli by ste žiaden rozdiel. Rovnako to platí nielen pri obrázkoch, ale aj pri audio a video signáloch, a to ako pri ukladaní, tak aj pri prenose digitálnymi sieťami.

Kompresné algoritmy môžeme rozdeliť na:

- Bezstratové kódovanie
- Stratové kódovanie



Bezstratové kódovanie znižuje počet bitov hľadaním a odstraňovaním štatistickej nadbytočnosti, teda informácií, ktoré sa opakujú alebo sa môžu získať z iných informácií v signále. Bezstratová kompresia (označovaná tiež ako kompakcia) umožňuje rekonštrukciu pôvodného signálu bez akýchkoľvek zmien. Je preto vhodná na archivačné účely.



Na druhej strane stratová kompresia sa snaží znížiť množstvo dát odstraňovaním informácií, ktoré človek kvôli nedokonalostiam svojich zmyslov nemôže zachytiť. Pri použití tejto metódy kompresie nie je rekonštruovaný signál nikdy rovnaký ako pôvodný signál. Metódy sú však vhodné na aplikácie, kde sme obmedzení ukladacím priestorom alebo šírkou prenosového kanála.

Vo všeobecnosti sú stratové algoritmy oveľa efektívnejšie ako bezstratové algoritmy, no zrekonštruovaný signál sa vždy líši od toho pôvodného. Kvalita komprimovaného signálu je teda kompromisom medzi množstvom informácií odstránených z pôvodného signálu a požadovanou veľkosťou súboru komprimovaného signálu.

Miera zmien v signáli, ktoré sú ešte akceptovateľné bez pozorovateľného rozdielu oproti pôvodnému signálu, sa nazýva práve pozorovateľné skreslenie (*just-noticeable distortion*) (pri zvuku a reči) alebo práve pozorovateľný rozdiel (*just-noticeable difference*) (pri obrázkoch a videu), **jnd**.

Kompresný pomer, vyjadrujúci efektívnosť kompresnej metódy, je daný vzťahom:

$$C_r = \frac{N_n}{N_k},$$

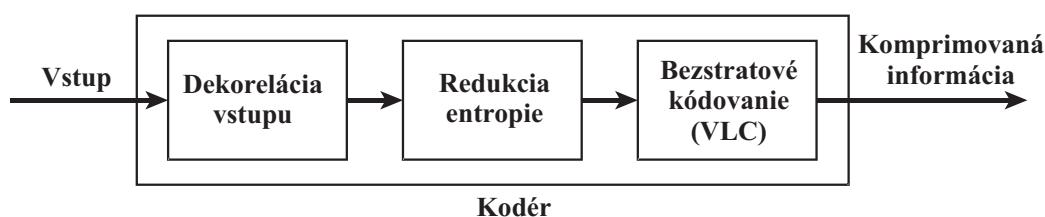
kde N_n je počet bitov nekomprimovaného signálu a N_k je počet bitov komprimovaného signálu.

Softvér, ktorý vykonáva kódovanie pôvodného signálu do jeho komprimovanej podoby a dekódovanie komprimovaného signálu do pôvodnej podoby, sa nazýva kodek (skratka z **kodér-dekodér**).

Model kompresie

Proces kompresie sa skladá z troch základných fáz:

1. Dekorelácia vstupných dát
2. Redukcia entropie
3. Bezstratové kódovanie



Obr. 5.1: Všeobecná kompresná schéma

Dekorelácia vstupných dát odstraňuje zo signálu duplicitné informácie. Môže sa vykonať rôznymi spôsobmi, resp. v rôznych oblastiach reprezentácie signálu:

- Časová (alebo priestorová) oblasť
- Parametrická oblasť
- Rozklad signálu

V časovej (alebo priestorovej) oblasti sú metódy založené na lineárnej predikcii, ktorá predpokladá, že susedné vzorky v čase (alebo priestore) sú korelované (podobajú sa). Prístup pomocou rozkladu signálu rozkladá signál na subpásma, v ktorých je možné samostatne sledovať energiu. Do parametrickej oblasti prechádzame rôznymi transformáciami tak, aby sme dokázali získať charakteristické parametre na kódovanie a prenos vstupných dát.

Proces kvantizácie redukuje entropiu. Hoci kvantizácia vždy prináša stratu informácií, dostatočne presná kvantizácia túto stratu znižuje na minimum.

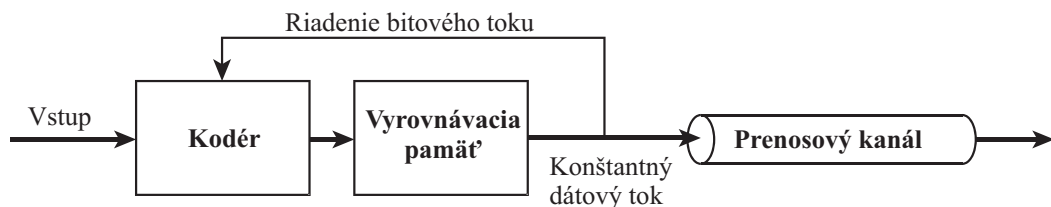
Posledným krokom v procese kódovania je bezstratové kódovanie, nazývané aj entropické kódovanie. Počas neho sa štatisticky častým kombináciám bitov

(symbolom) priradí kratšie kódové slovo a štatisticky menej častým kombináciám bitov dlhšie kódové slovo. Takéto kódovanie sa tiež nazýva kódovanie s variabilnou dĺžkou (variable length coding, VLC), pretože kódové slová majú rôznu dĺžku. Najefektívnejším algoritmom na vykonanie bezstratového VLC entropického kódovania je Huffmanovo kódovanie.

Na prenos komprimovanej informácie prostredníctvom prenosového kanála existujú dva režimy:

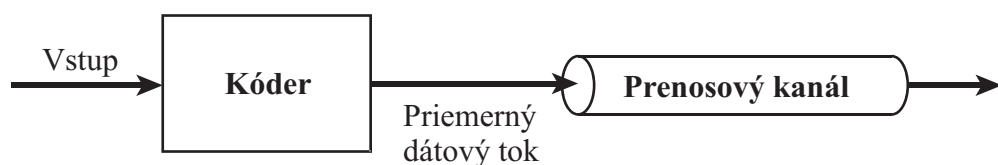
- Konštantný dátový tok
- Premennivý dátový tok

Pri použití konštantného dátového toku sa výstup kódéra ukladá do vyrovnávacej pamäte, z ktorej sa zabezpečuje konštantný dátový tok do prenosového kanála. Dátový tok ovplyvňuje kóder tak, aby produkoval rovnaký dátový tok. Takýmto spôsobom môže kóder ovplyvňovať krok kvantovania a tým aj kvalitu rekonštruovaného signálu.



Obr. 5.2: Kóder s konštantným dátovým tokom

Pri použití premenlivého dátového toku vyrovnávacia pamäť nie je potrebná a kóder tak môže použiť krok kvantovania podľa požiadaviek kódovaného vstupného signálu. Kvalita rekonštruovaného signálu sa potom nemení.



Obr. 5.3: Kóder s premenlivým dátovým tokom

Hoci sa môže zdať premenlivý dátový tok najlepšou voľbou, platí to iba v prípade, že kapacita prenosového kanála je dostatočná aj pre maximálny dátový tok, ktorý môže kóder s premenlivým dátovým tokom produkovať. Ak však máme k dispozícii obmedzenú kapacitu prenosového kanála, na zachovanie nepretržitého toku dát bude vhodnejší kóder s konštantným dátovým tokom.

5.1 Kompresia audio signálov

Na kompresiu všeobecných audio signálov máme na výber veľké množstvo kódovacích štandardov a kompresných prístupov. Mnohé z nich sa zameriavajú na špecifické druhy zvuku (napr. reč) alebo vlastnosti (výpočtová náročnosť, oneskorenie, atď.).



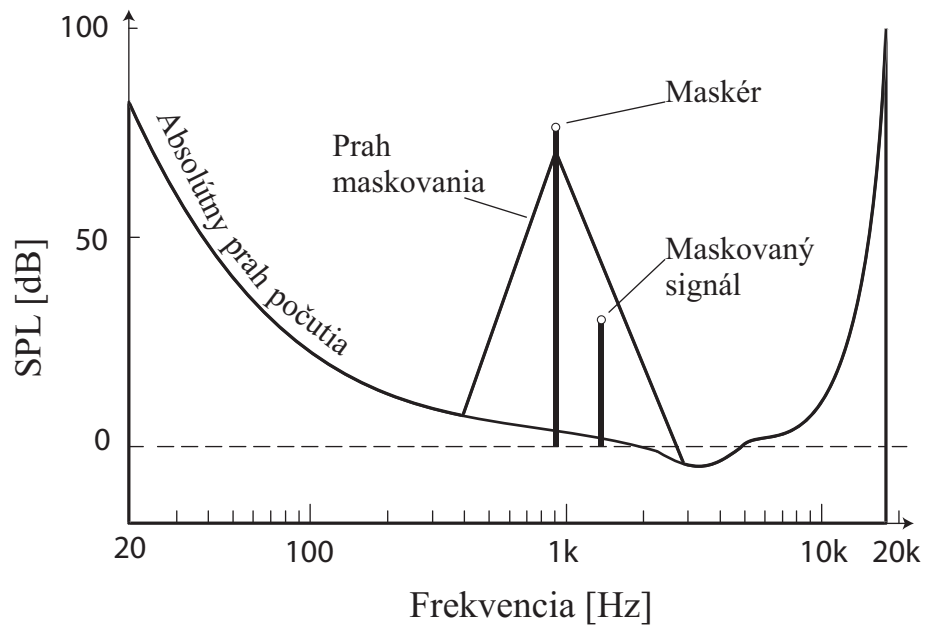
Vzorkovacia frekvencia popisuje, koľko vzoriek bolo zo signálu získaných každú sekundu. Všeobecne platí, že čím vyššiu vzorkovaciu frekvenciu použijeme, tým presnejšiu, kvalitnejšiu a vernejšiu nahrávku získame. Medzi najpoužívanéjšie vzorkovacie frekvencie patria: 8 kHz, 16 kHz, 22,5 kHz, 32 kHz, 44,1 kHz alebo 48 kHz pre každý audio kanál.

Sluchové maskovanie

Sluchové maskovanie je jav, ktorý sa prejavuje v dôsledku nedokonalosti ľudského sluchového aparátu. Uši nie sú schopné počuť zvuky všetkých intenzít, čo je vyjadrené tzv. absolútnym prahom počutia. Okrem toho jeden hlasný zvuk často prekryje blízko sa nachádzajúci tichší zvuk. To sa môže stať v časovej a frekvenčnej oblasti, preto maskovanie delíme na:

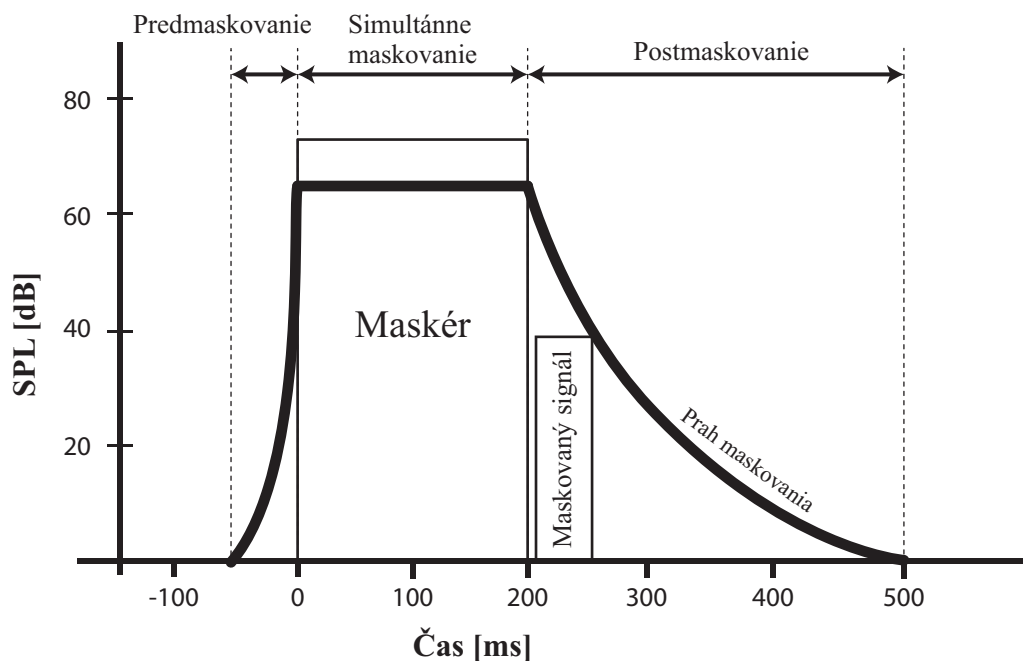
- časové (nesimultánne) maskovanie
- frekvenčné (simultánne) maskovanie

Hlasný zvuk sa označuje aj ako maskér. Ak sa dva zvuky vyskytnú v tom istom čase, môže nastať simultánne maskovanie. Maskér vytvorí prah maskovania, pod ktorým nebude počuť žiadne iné zvuky. Ak signál v blízkosti maskéra nemá dostatočnú intenzitu a padne pod prah, bude maskovaný. Nasledujúci obrázok ukazuje, ako maskér dokáže prekryť tichý signál vo frekvenčnej oblasti. Kombinácia prahov maskovania viacerých maskérov a absolútneho prahu počutia vytvára globálny prah maskovania, ktorý sa môže v čase meniť. Proces určovania maskovacích prahov sa označuje ako psychoakustická analýza.



Obr. 5.4: Maskovanie vo frekvenčnej oblasti. Ak je intenzita maskéra dostatočne vyššia než intenzita signálu a signály sú dostatočne blízko seba, signál nebude počuteľný a počuť bude iba maskéra.

Počas časového, nesimultánneho maskovania môže maskér prekryvať signál, ktorý sa vyskytne tesne pred (predmaskovanie) alebo po (postmaskovanie) zaznení maskéra. Opäť však platí, že intenzita maskéra musí byť oveľa vyššia než intenzita signálu.



Obr. 5.5: Maskovanie v časovej oblasti

Frekvenčné maskovanie je v súčasnosti už preskúmané dostatočne presne a má, ako neskôr ukážeme, široké využitie v mnohých audio kodekoch. Na druhej strane

časové maskovanie zatiaľ nebolo prebádané tak dôsledne, čo je dôsledkom jeho pomerne krátkeho trvania. Postmaskovanie trvá priemerne najviac 300 ms od skončenia samotného maskéra a predmaskovanie trvá dokonca len menej ako 50 ms. Tieto časy sú príliš krátke na to, aby ich bolo možné dôsledne analyzovať, pretože kodeky pracujú zvyčajne s rámcami dĺžky aspoň 20 ms, takže na predmaskovanie sú použiteľné iba 2 až 3 rámce.

MPEG

V súčasnosti je väčšina audio kodekov založená na práci Expertnej skupiny pre pohyblivý obraz *Motion Picture Experts Group* (**MPEG**), ktorá je súčasťou Medzinárodnej štandardizačnej organizácie (*International Standards Organization*, **ISO**). Počas jej existencie skupina uviedla niekoľko audio formátov, ktoré sa celosvetovo používajú.

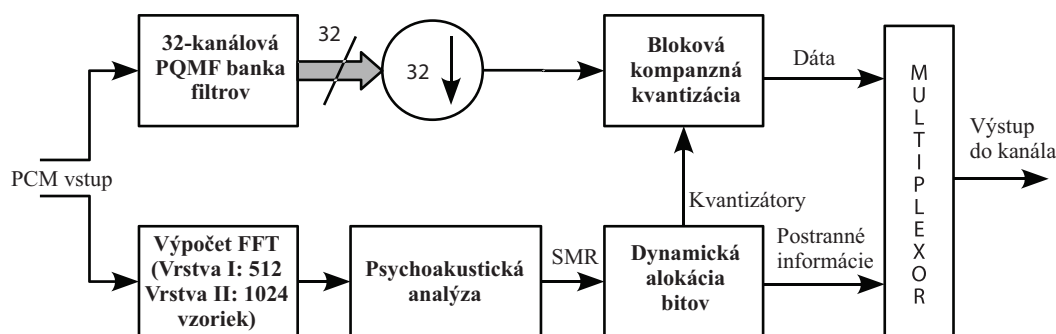
Ako bude ďalej zrejmé, kodeky z rodiny MPEG sú založené na stratovom kódovaní, čo znamená, že modifikujú pôvodný audio signál a rekonštruovaný signál nie je nikdy zhodný s pôvodným.

MPEG-1



Štandard MPEG-1 predstavuje flexibilnú kódovaciu techniku, ktorá využíva viacero metód, napr. subpásmové kódovanie, analýzu bankou filtrov, transformačné kódovanie, entropické kódovanie a psychoakustickú analýzu. Pracuje so vzorkovacími frekvenciami 32, 44,1 alebo 48 kHz so 16 bitmi/vzorka a výstupný dátový tok sa pohybuje od 32 do 192 kbit/s na jeden kanál. Štandard ponúka 4 režimy kódovania kanálov: mono, stereo, duálne mono a spojené stereo (iba vrstva III).

Architektúra štandardu obsahuje 3 vrstvy, ktoré sa líšia výpočtovou náročnosťou, oneskorením a kvalitou výstupu. Vrstvy I (mp1) a II (mp2) sú si podobné a líšia sa iba v niekoľkých detailoch. Obe používajú rýchlu Fourierovu transformáciu (*fast Fourier transform*, **FFT**), avšak vrstva I využíva okno s veľkosťou 512 vzoriek, zatiaľ čo vrstva II používa 1024-vzorkové okno. Maximálne podporované rozlíšenie v subpásmovej kvantizácii je v prípade vrstvy I 15 bitov/vzorka a v prípade vrstvy II 16 bitov/vzorka. Hoci sa tieto rozdiely zdajú byť minimálne, ukázalo sa, že vrstva II poskytuje rovnakú či dokonca vyššiu kvalitu výstupu pri bitovom toku 128 kbit/s než vrstva I s bitovým tokom 192kbit/s na kanál.



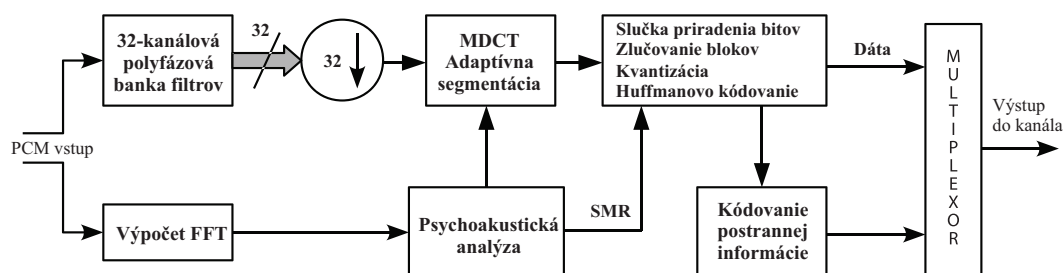
Obr. 5.6: Zovšeobecnená schéma kodéra MPEG-1 vrstva I a II

Proces kompresie v oboch vrstvách I a II pracuje so vstupným PCM signálom, ktorý rozkladá na 32 subpásiem. Počas rozkladu sa vykoná FFT, ktorej výstup prejde psychoakustickou analýzou a určením jnd. V závislosti od prahu maskovania sa pre každé subpásmo stanoví najvhodnejší krok kvantovania tak, aby boli dodržané požadovaný dátový tok a úroveň maskovania. Výstup kodéru sa a záver zakóduje Huffmanovým entropickým kódovaním.

- + Hoci MPEG-1 vrstva II poskytuje prijateľné výsledky, prevládajúcim formátom je MPEG-1 vrstva III, všeobecne známa svojou skratkou mp3. Vychádza z vrstiev I a II, pridáva však viaceré nové techniky, ktoré vedú k nižšiemu dátovému toku (okolo 64 kbit/s na kanál) pri zachovaní kvality svojich predchodcov.

Algoritmus pracuje s oknom 1152 vzoriek, ktoré rozdelí na dve tzv. granuly po 576 vzoriek. Každá z týchto granúl prejde hybridnou bankou filtrov (sústavou pásmových priepustí slúžiacich na rozdelenie vstupu na subpásma: každé subpásmo sa potom môže spracúvať samostatne), aby sa zvýšilo frekvenčné rozlíšenie. Každé subpásmo sa transformuje do frekvenčnej oblasti použitím modifikovanej diskkrétnej kosínusovej transformácie (*Modified discrete cosine transform*, **MDCT**). Potom sa iteratívne vykoná priradenie bitov a kvantizácia: počas každej iterácie sa procesom analýzy pomocou syntézy odhadne úroveň kvantizačného šumu.

- + Modifikovaná diskrétna kosínusová transformácia vychádza z diskkrétnej Fourierovej transformácie, je však špeciálne navrhnutá pre signály, ktorých bloky vzoriek sa prekrývajú. Vstupný signál rozkladá (transformuje) na sústavu kosínusových funkcií. V porovnaní s Fourierovou transformáciou, ktorej výstupom je sústava komplexných čísel, výstupom MDCT je sústava reálnych čísel charakterizujúcich kosínusové funkcie. Okrem toho výstupom DFT je rovnaký počet koeficientov ako je počet vzoriek vstupného signálu, zatiaľ čo pri MDCT je vďaka jej práci s prekryvom na výstupe polovičný počet koeficientov.



Obr. 5.7: Zovšeobecnená schéma kodéra MPEG-1 vrstva III

K pôvodnej vrstve III vznikli dve rozšírenia, MP3pro a mp3 surround. MP3pro pridáva techniku nazývanú replikácia spektrálneho pásma (*Spectral Band Replication* (**SBR**)), ktorá sa používa pri nižších dátových tokoch na odstránenie pôvodných vysokých frekvencií. Tie je možné zrekonštruovať pomocou prídavných informácií z komprimovaného signálu. mp3 surround umožňuje zakódovanie 5.1 kanálového zvuku (5 plnopásmových a jeden nízko-frekvenčný – basový – kanál) do dvoch kanálov mp3. Z nich je možné zrekonštruovať 5.1-kanálový zvuk pomocou postranných informácií. Ak dekodér rozšírenie MP3pro alebo mp3 surround nepodporuje, postranné informácie ignoruje a súbor prehrá ako štandardný súbor mp3.

MPEG-2

Štandard MPEG-2 je formálnym pokračovateľom MPEG-1. Zahŕňa 2 režimy, jeden spätne kompatibilný s MPEG-1 (Backward Comptible, MPEG-2 BC) a druhý spätne nekompatibilný (Non-Backward Comptible, MPEG-2 NBC), ktorý prináša nové metódy a techniky kódovania.

Jedinými zmenami MPEG-2 BC oproti MPEG-1 je podpora pre nižšie vzorkovacie frekvencie (LSF) a viackanálové kódovanie podobné rozšíreniu mp3 surround. Formát MPEG-2 NBC sa označuje aj ako pokročilé kódovanie zvuku (Advanced Audio Coding, AAC) a je zostavený ako súprava nástrojov na efektívne kódovanie. Čím viac nástrojov sa použije, tým lepšia kompresia sa dosiahne, pričom kvalita výstupu zostane zachovaná. Cenou je však vyššia výpočtová náročnosť a oneskorenie. Na rozdiel od MPEG-1 formát MPEG-2 NBC nepoužíva na analýzu signálu hybridnú banku filtrov, ale iba MDCT v kombinácii s novými oknovými funkciami. Formát MPEG-2 sa stal súčasťou rodiny štandardov MPEG-4.

MPEG-4 AAC

Formát MPEG-4 AAC sa pokúša pokoriť vládu formátu mp3. Prináša podporu pre vzorkovacie frekvencie od 8 po 96 kHz, 1 až 48 audio kanálov plus 15 basových a 15 dátových kanálov s rozlíšením 8, 16, 24 alebo 32 bitov/vzorka. Formát AAC s nízkou zložitou (Low Complexity (LC) AAC) predstavuje pôvodný kodek MPEG-2 AAC a je vhodný na kódovanie reči pri dátovom toku 8-12 kbit/s. Formát AAC s vysokou efektivitou (High Efficiency (HE) AAC) prináša podporu technológie SBR (verzia 1) a kanálového režimu parametrické stereo (verzia 2), ktorý je založený na profile spojené stereo štandardu MPEG-1 vrstva III.

Ogg Vorbis



Zvukový kodek Vorbis patrí medzi najúspešnejšie kodeky s otvoreným zdrojovým kódom. Od roku 2000, kedy bol štandardizovaný, sa stal priamym súperom formátu mp3 skupiny MPEG. Podporuje vzorkovacie frekvencie od 8 do 192 kHz, maximálne 255 kanálov a bitový tok jeho výstupu je variabilný.

Proces kódovania sa odlišuje od štandardov MPEG: signál sa najskôr transformuje pomocou MDCT. Vo frekvenčnej oblasti sa určí hrubý odhad spektrálnej obálky (krivky, ktorá spája jednotlivé amplitúdové koeficienty frekvenčného spektra) pomocou lomenej funkcie, ktorý sa nazýva *floor*. Rozdiel medzi floorom a skutočným spektrom sa potom zakóduje viacprechodovou vektorovou kvantizáciou.

Ogg Vorbis má vyššie nároky na pamäť než mp3, pretože jeho záhlavie obsahuje kódovú tabuľku entropického kódovania (mp3 má tabuľku pevne danú) a nastavenia dekodéra. Napriek tomu je formát veľmi vhodný na kompresiu všeobecných audio signálov a poskytuje rovnakú alebo vyššiu kvalitu zvuku pri rovnakom bitovom toku ako kodek mp3.

Windows Media Audio

Windows Media Audio (WMA) je proprietárny kodek vytvorený spoločnosťou Microsoft ako odpoveď na licenčné obmedzenia formátu mp3. Existuje viacero verzií kodeku: WMA 9 je priamym súperom mp3 s podporou pre vzorkovacie frekvencie do 48 kHz so 16 bitmi/vzorka a výstupným dátovým tokom od 64 do 192 kbit/s, podporujúc CBR aj VBR.

Verzia WMA 10 Professional rozširuje možnosti kodeku v súboji s MPEG-4 AAC pridaním vzorkovacej frekvencie 96 kHz s 24 bitmi/vzorka pre 7.1 kanálov. Ak zariadenie nie je schopné prehrať 7.1 nahrávku, signál sa automaticky degraduje (vzorkovacia frekvencia, počet bitov na vzorku, zníženie počtu kanálov) na úroveň vhodnú pre zariadenie.



WMA 10 poskytuje tiež režim pre kompresiu reči nazývaný WMA 10 Voice, ktorý poskytuje bitový tok od 4 až do 20 kbit/s. Jeho zaujímavosťou je schopnosť dynamicky prepínať medzi hlasovou a štandardnou verziou kodeku, ak je spracúvaný signál príliš zložitý. Okrem toho WMA 10 poskytuje bezstratový režim, ktorý údajne dokáže zredukovať veľkosť pôvodného PCM signálu na polovicu až tretinu.



Kodek WMA 10 Professional poskytuje vyššiu subjektívnu kvalitu pri 64 kbit/s v porovnaní s MPEG-4 AAC v2 pri 70% porovnaní.

5.2 Kompresia reči

Hoci je reč vo svojej podstate audio signál, má niektoré špecifické vlastnosti, ktoré umožňujú použitie radikálnejších kompresných techník než pri všeobecnom zvuku. V prvom rade sa rečový signál považuje za médium na vyjadrenie informácie. Informácia nemusí byť po rekonštrukcii vyrozprávaná úplne rovnako ako v origináli, aby bola zrozumiteľná. Z toho vyplýva možnosť odstránenia niektorých vlastností zvuku. Napríklad štandardný telefónny hovor je vzorkovaný s frekvenciou 8 kHz (v porovnaní s 44,1 kHz vzorkovacou frekvenciou pri bežnom audio), čo znamená zachytenie šírky pásma iba 4 kHz. Napriek tomu toto pásmo obsahuje väčšinu energie reči a teda informácie.

V druhom rade je rečový signál jednoduchý v porovnaní napríklad s nahrávkou rockovej kapely – zvyčajne je v ňom zachytený jediný rečník bez ďalších hudobných nástrojov. Okrem toho na dosiahnutie čo možno najčistejšej reči môžeme použiť algoritmy na potlačenie šumu a ruchov v pozadí.

Metódy používané na kompresiu reči môžeme rozdeliť do nasledujúcich skupín:

- Kódovanie tvaru vlny
 - v časovej oblasti
 - vo frekvenčnej oblasti
- Vokóдеры
 - lineárne predikčné kódovanie
 - formantové kódovanie

Na účely hodnotenia kvality rôznych algoritmov na spracovanie reči existuje miera zrozumiteľnosti. Popisuje, ako je reč pochopiteľná a zrozumiteľná. Do úvahy sa pritom berú rôzne vlastnosti, napríklad hlasnosť reči, nelineárne skreslenie, úroveň šumu na pozadí, ozveny a dozvuky a mnohé iné. Na porovnanie zrozumiteľnosti existujú dve hlavné stupnice: index prenosu reči (Speech Transition Index, STI) a všeobecná škála zrozumiteľnosti (Common Intelligibility Scale, CIS), ktoré majú rozsah od 0 (najhoršie) po 1 (najlepšie), resp. 0% až 100%. Vo všeobecnosti sa reč považuje za zrozumiteľnú, ak algoritmus dosiahne na stupnici skóre aspoň 0,5 (alebo 50%).

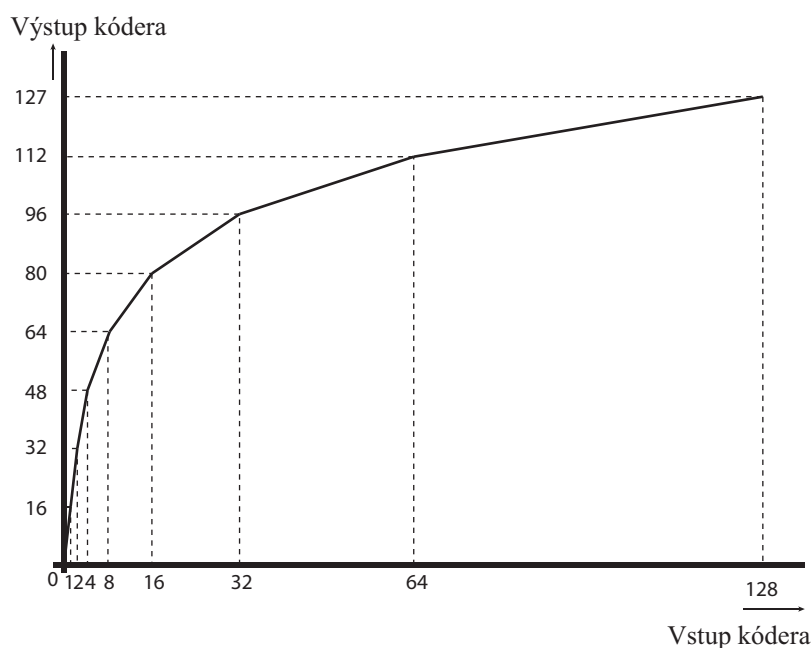
Časová oblasť

Kódovanie tvaru vlny v časovej oblasti je reprezentované technikami PCM. Zatiaľ čo lineárna PCM používa rovnaké rozostupy medzi kvantizačnými úrovňami, nelineárna PCM využíva nelineárnu kvantizačnú stupnicu alebo jej modifikáciu v podobe kompresie dynamiky (kompanziu) vstupného signálu na strane vysielača a jej expanziu na strane prijímača.

Odporúčanie G.711 definuje dve kompanzné charakteristiky μ -zákon (μ -law) (USA a Japonsko) a A-zákon (A-zákon) (Európa). Na príklad charakteristika A-zákona je daná vzťahom:

$$F(x) = \text{sgn}(x) \begin{cases} \frac{A|x|}{1 + \ln(A)}, & |x| < \frac{1}{A} \\ \frac{1 + \ln(A|x|)}{1 + \ln(A)}, & \frac{1}{A} \leq |x| \leq 1 \end{cases}$$

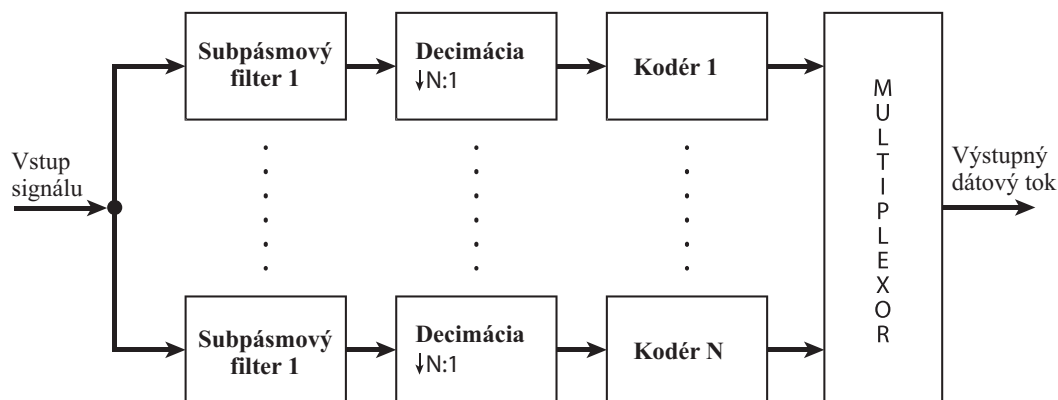
kde $\text{sgn}(x) = \pm 1$ pre kladnú alebo zápornú hodnotu x a A je parameter kompresie. Zvyčajne $A=87,7$.



Obr. 5.8: Príklad kompanznej krivky A-zákona. Vyššie frekvencie (na horizontálnej osi reprezentované vyššími číslami) sú kódované menším počtom hodnôt než nižšie frekvencie.

Frekvenčná oblasť

Vo frekvenčnej oblasti sa využívajú metódy subpásmového a adaptívneho transformačného kódovania. Pri subpásmovom kódovaní (Subband Coding, SBC) sa rečový signál pomocou sady pásmových priepustí (banky filtrov, Obr. 5.9) rozdelí na niekoľko frekvenčných pásem a na zníženie počtu vzoriek sa zdecimuje. Každé pásmo sa potom kóduje samostatne, najčastejšie použitím metódy ADPCM, ktorá umožňuje flexibilnú kvantizáciu a priradovanie bitov.

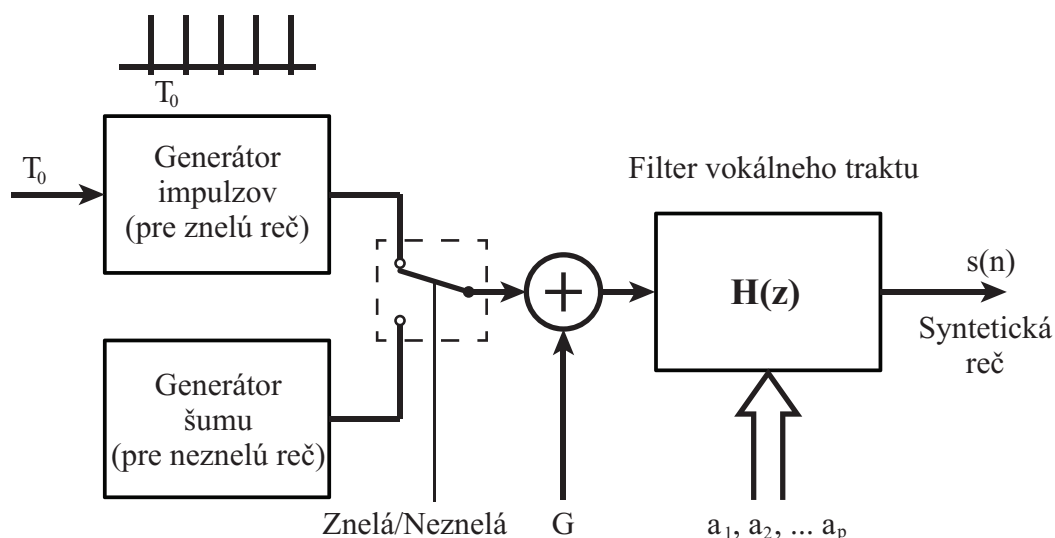


Obr. 5.9: Príklad banky filtrov

Alternatívne sa namiesto metódy ADPCM môžu použiť techniky založené na adaptívnom transformačnom kódovaní (*Adaptive Transform Coding*, **ATC**). Tu sa signál transformuje do frekvenčnej oblasti pomocou FFT alebo inej transformácie a rozdelí sa na subpásmo. Potom sa jednotlivým vzorkám v subpásmach dynamicky pridelia bity podľa ich potreby.

Lineárne predikčné kódovanie

Prirodzená ľudská reč sa môže chápať ako odozva vokálneho traktu rečníka na budiaci signál, v našom prípade na vzduch vydychovaný z pľúc. Výstupný signál sa modeluje zmenami vlastností vokálneho traktu (hlasiviek, ústnej dutiny, zubov, atď.). Ak sa na tento proces pozrieme z pohľadu analýzy signálov, môžeme výstupný signál reprezentovať pomocou budiaceho signálu a filtra predstavujúceho hlasový trakt s parametrami meniacimi sa v čase. Tieto koeficienty sa prepočítavajú približne každých 10 až 30 ms. Hoci na popis koeficientov funkcie vokálneho traktu existuje mnoho metód, najpoužívanejšou je metóda založená na lineárnej predikcii, z čoho vychádza jej názov lineárne predikčné kódovanie (*Linear Prediction Coding*, **LPC**).



Obr. 5.10: Všeobecná schéma LPC dekodéra

LPC koeficienty minimalizujú kvadratickú odchýlku medzi originálnymi a predikovanými vzorkami reči. Ako je vidieť, model LPC generátora reči pozostáva z dvoch častí vychádzajúcich z vyššie uvedeného textu:

- Budenie vokálneho traktu
- Filter vokálneho traktu

Budenie vokálneho traktu reprezentuje generátor impulzov a generátor šumu, medzi ktorými je možné prepínať v závislosti od znelosti reči. Budiaci signál sa ďalej zosilňuje zosilňovačom (G) na požadovanú úroveň.



Hlasivky sa chvejú, vibrujú s určitou základnou frekvenciou f_0 , z ktorej vieme odvodiť základnú hlasivkovú periódu T_0 . Čím vyššia je základná frekvencia, tým vyššia je výška reči.

V závislosti od toho, či sa hlasivky použijú alebo nie, rozdeľujeme zvuky reči na:

- **znelé** – zvuky sú vyslovené pomocou chvenia hlasiviek, cez ktoré prechádza vzduch z pľúc, napr. „a“, „v“, „z“. Všetky samohlásky sú znelé.
- **neznelé** – zvuky sú vyslovené iba za pomoci toku vzduchu podobnému šumu a hlasivky sa nezapájajú, napríklad „s“, „c“, „f“.

Filter vokálneho traktu je daný lineárnym predikčným filtrom s konečnou odpoveďou (FIR filter), ktorého prenosovú funkciu charakterizuje vzťah:

$$H(z) = \frac{G}{1 + \sum_{i=1}^p a_i \cdot z^{-i}} = \frac{S(z)}{E(z)},$$

kde a_i sú koeficienty filtra a p je rád filtra. Ak $S(z)$ predstavuje výstupnú vzorku a $E(z)$ je budenie, nasledujúcu vzorku $s(n)$ získame ako lineárnu kombináciu predošlých vzoriek s budením $G \cdot e(n)$:

$$s(n) = G \cdot e(n) - \sum_{i=1}^p a_i \cdot z^{-1} = G \cdot e(n) - a_1 \cdot s(n-1) - \dots - a_p \cdot s(n-p)$$

Na použitie LPC generátora reči potrebujeme pre každý segment určiť nasledujúce parametre:

- znelosť segmentu
- základnú periódu T_0
- parametre filtra (zosilnenie G a koeficienty a_i)

+

Bitový tok rečového signálu zakódovaného pomocou LPC sa pohybuje v rozmedzí od 1,2 do 2,4 kbit/s a jeho zrozumiteľnosť je okolo 80-85%.

—

Rekonštruovaný rečový signál však znie strojovo, čo je zapríčinené dvoma hlavnými dôvodmi:

1. Je ťažké segmentovať reč presne na znelé a neznelé segmenty, pretože v prirodzenej reči sa znelosť prelína a kombinuje.
2. Základná perióda (ktorá charakterizuje hlas rečníka) sa v prirodzenej reči mení častejšie, než je dĺžka rámcov a zmeny nie sú periodické.

Na potlačenie nedokonalostí metódy LPC existujú postupy, ktoré kódujú rezíduum (rozdiel) medzi originálnym signálom a signálom generovaným LPC.

Metóda lineárnej predikcie budenej rezíduom (*Residually Excited Linear Prediction*, **RELP**) prenáša priamo rozdiel medzi originálnym a rekonštruovaným signálom. Na strane prijímača sa LPC koeficienty použijú na rekonštrukciu signálu a na presnejšiu rekonštrukciu sa rezíduum pripočíta.

Nástupcom algoritmu RELP je lineárna predikcia budená kódom (*Code Excited Linear Prediction*, **CELP**). Algoritmus je založený na princípe analýzy syntézou a v uzavretej slučke perceptuálne optimalizuje syntetizovaný signál. Potom sa prehľadá kódová kniha a vyberie sa najvhodnejšia budiaca funkcia. Spolu s LPC koeficientmi sa prenáša iba pozícia v kódovej knihe. Budiaca funkcia sa môže zakódovať pomocou vektorovej kvantizácie.

—

Metóda CELP dosahuje bitový tok od 4 do 8 kbit/s. Jej nevýhodami sú relatívna výpočtová náročnosť a oneskorenie na úrovni 35 ms.

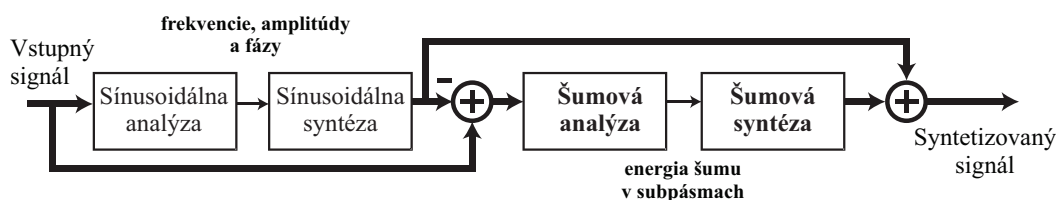


Modifikácia CELP-u s nízkym oneskorením (Low Delay CELP, LD-CELP) znižuje oneskorenie na 2 ms pri bitovom toku 16 kbit/s a stala sa súčasťou štandardu G.728 organizácie ITU-T. Ďalší kodek založený na metóde CELP je

Speex, ktorý vytvorila organizácia Xiph.Org, autor kodeku Ogg Vorbis, a má otvorený zdrojový kód.

Sínusoidálne kódovanie

Sínusoidálne kódovanie vychádza z predpokladu, že každý audio signál je kombináciou deterministickej a stochastickej zložky. Deterministickú zložku tak môžeme reprezentovať harmonickými funkciami (sínus, kosínus) a stochastickú zložku pomocou šumu alebo inou parametrizáciou. Princiálna schéma takého kodéra je uvedená nižšie. Sínusoidy sú spojenia v čase sa meniacich frekvencií, o ktorých predpokladáme, že tvoria jeden súvislý tón.



Obr. 5.11: Všeobecná schéma sínusoidálneho kodéra

Tento model však nedokáže dostatočne spracovať rýchlo sa meniace zvuky, preto bola do metódy pridaná tretia zložka, tzv. transients, ktoré modelujú rýchle zmeny v signáli. To vedie k modelu sínusoidy+transienty+šum (STN).

Iným rozšírením základného SN modelu je model Harmonické + samostatné línie + šum (HILN). V tomto prístupe sa sínusoidy rozdelia do dvoch skupín, harmonických a individuálnu. V harmonickej časti sú sínusoidy reprezentované ako harmonické násobky základných frekvencií a ukladajú sa iba násobiče. Potom sa kódujú individuálne sínusoidy a rezíduum sa spracúva ako šum.

+ Od sínusoidálneho kódovania sa očakáva dobré spracovanie jednoduchých signálov, pozostávajúcich zvyčajne z harmonických zvukov, napríklad reč. Techniku využíval kodek SVOPC, prvý kodek spoločnosti SKYPE, ktorý dosahoval dobrú kvalitu pri 20 kbit/s a bol odolný voči stratám paketov.

— Výpočtová náročnosť kodeku SVOPC však viedla k vytvoreniu nového kodeku SILK, založeného na LPC.

Na základe kodeku SILK s pridaním vlastností kodeku **CELT** (*Constrained Energy Lapped Transform*) bol v septembri 2012 štandardizovaný nový kodek s názvom Opus. Kodek je schopný využiť dobrý výkon kodeku SILK pri nízkych frekvenciách s nízkym oneskorením kodeku CELT pri vyšších frekvenciách, medzi ktorými dokáže prepínať na požiadanie. Kodek je vysoko vhodný na kódovanie reči aj všeobecného audia, a to aj pre online aplikácie, ako sú VoIP a živé prenosy.

5.3 Kompresia statických obrazov



Cieľom kompresie obrazov je redukcia nepotrebných informácií v obraze na zníženie šírky potrebného prenosového pásma alebo ukladacieho priestoru. Podobne ako pri audio signáloch existujú pri obraze stratové aj bezstratové algoritmy, ktoré sa používajú v závislosti od cieľového použitia.



Pri analýze obrazov automaticky predpokladáme používanie 2-rozmerných verzií transformácií.



Na to, aby sme mohli v počítači ukladať obrázky, existujú rôzne farebné priestory. Uvádzame tie najpoužívanejšie.

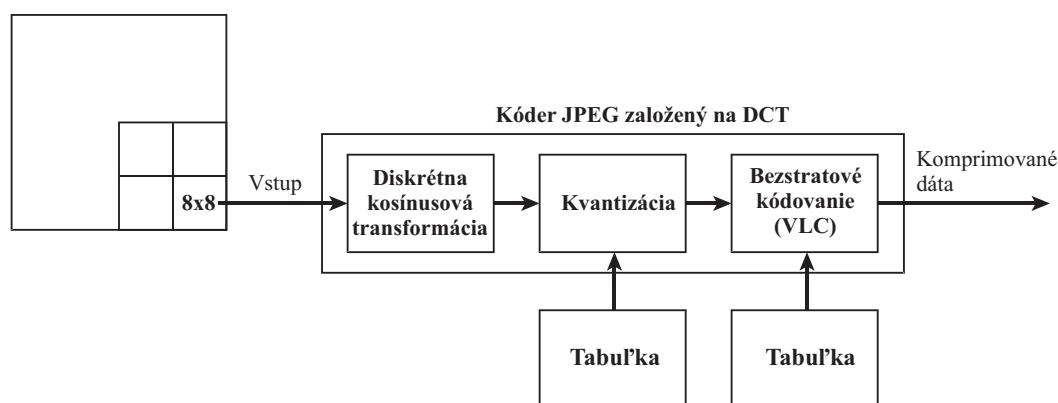
Vybrané farebné priestory

Skratka	Význam	Vysvetlenie
RGB	Red, Green, Blue (červená, zelená, modrá)	Každý pixel je daný kombináciou 3 farieb svetla. Kombinácia najvyšších úrovní všetkých 3 farieb dáva bielu. Používa sa pri obrazoch vytváraných svetlom.
RGBA	Red, Green, Blue, Alpha (červená, zelená, modrá, alfa)	Rovnaký význam ako pri RGB. Pridaný alfa kanál popisuje priehľadnosť.
YCbCr alebo YUV	Jas (Y), Modrá farebnosť (U), Červená farebnosť (V)	Jas sa mení od čiernej po bielu. Modrá a červená farebnosť sa vypočítajú z daného zdroja RGB. Hoci sa YCbCr označuje aj ako YUV, existujú medzi nimi rozdiely pri výpočte zložiek farebnosti.
CMYK	Cyan, Magenta, Yellow, Black (azúrová, purpurová, žltá, čierna)	Každý bod obrazu je daný kombináciou štyroch farieb. Kombinácia najvyšších úrovní všetkých 4 farieb dáva čiernu. Používa sa v tlači.

JPEG

Štandard JPEG je jedným z najpoužívanějších stratových obrazových formátov na ukladanie fotografií. Je pomenovaný podľa spojenej expertnej skupiny pre obraz (Joint Picture Experts Group), ktorá štandard uviedla v roku 1986. Dosahuje kompresný pomer 10:1 s minimálnou postrehnuteľnou stratou kvality.

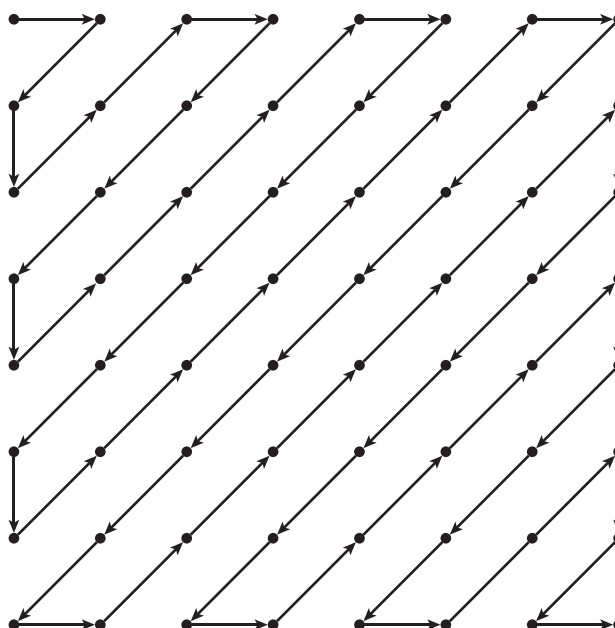
Algoritmus JPEG je založený na 2-rozmernej diskretnej kosínusovej transformácii (DCT). Vstupný obraz sa skonvertuje do farebného priestoru $YCbCr$, ktorý poskytuje lepšie vlastnosti než povedzme RGB. Obraz sa potom rozdelí na neprekrývajúce sa bloky veľkosti 8x8 bodov, ktoré sa transformujú pomocou DCT. Získané koeficienty sa kvantizujú a málo významné koeficienty sa odstránia (tu dochádza k stratovej kompresii). Koeficienty sa potom zoradia do 1-rozmernej sekvencie a bezstratovo zakódujú. Obrazové komponenty (Y , C_B a C_R) sa kódujú postupne.



Obr. 5.12: Všeobecná schéma kódéra JPEG. Vstupom je blok 8x8 bodov jasovej alebo farebnostnej zložky.

Kľúčom v kompresii JPEG algoritmom je kvantizácia. Kvantizácia je nelineárna, keďže ľudské oko je citlivejšie na zmeny v nízkych frekvenciách. Na to, aby bola kvalita, resp. kompresný pomer, meniteľná, bol zavedený faktor kvality q_f , ktorý sa mení v rozsahu 1 až 100 a upravuje kvantizačnú maticu.

Na zoradenie koeficientov slúži tzv. cik-cak čítanie, ktoré začína v ľavom hornom rohu. Ak sa každý blok 8x8 zakóduje spolu, kódovanie sa označuje ako základné (baseline) JPEG. Iným prístupom je kódovanie najskôr ľavých horných rohov všetkých obrazových blokov a potom nasledujúcich koeficientov zo všetkých blokov. Takýto postup sa označuje ako progresívny JPEG a jeho výhodou je postupná rekonštrukcia obrazu už počas jeho preberania. JPEG navyše ponúka hierarchický režim, kde sa obraz kóduje pyramídovým spôsobom vo vrstvách. Každý pixel vyššej vrstvy vznikne použitím určitej operácie na blok 2x2 nachádzajúci sa priamo pod ním. Každú vrstvu dokáže dekódér dekódovať samostatne, čo umožňuje použiť obraz pri rôznych rozlíšeniach.



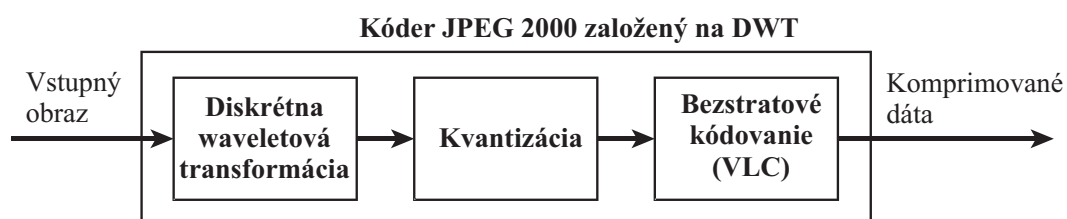
Obr. 5.13: Zorad'ovanie koeficientov do 1-rozmerného vektora pomocou cik-cak čítania



Štandard JPEG podporuje tiež bezstratové kódovanie založené na predikčnom kódovaní a bezstratovom VLC kódovaní, čím sa vynecháva DCT a úpravy spektra. Typický kompresný pomer je približne 2:1.

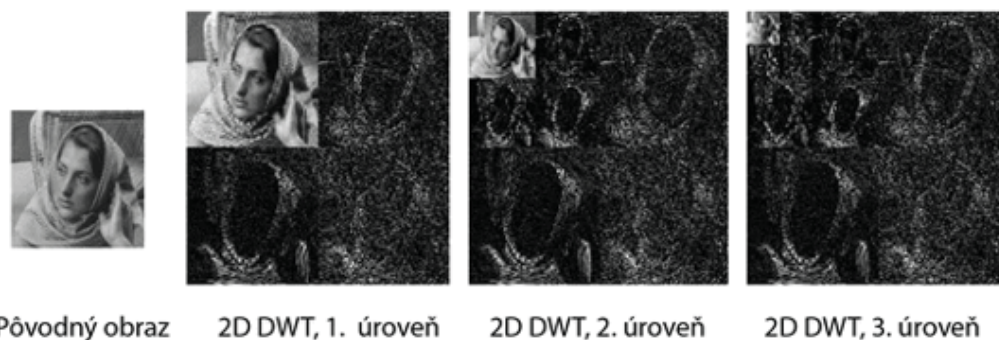
JPEG 2000

Formát JPEG 2000 (Obr. 5.14) sa snaží nahradiť nedokonalý, no stále výrazne obľúbenejší formát JPEG. Rozširuje možnosti svojho predchodcu, zlepšuje pomer kvalita/kompresia, umožňuje škálovateľnú stratovú a bezstratovú kompresiu. Medzi ďalšie vylepšenia patrí kódovanie oblasti záujmu, pri ktorom sa dôležité časti obrazu kódujú presnejšie než zvyšok.



Obr. 5.14: Všeobecná schéma kódera JPEG 2000

Transformačná funkcia sa zmenila z DCT na 1D diskretnú waveletovú transformáciu (DWT). Pôvodný obraz sa transformuje pomocou waveletu, kvantizuje a entropicky zakóduje. Hlavný rozdiel medzi DCT a waveletovou transformáciou je v tom, že DWT rozdeľuje bloky obrazu na podbloky, ktoré sa rozkladajú na podbloky, atď.



Obr. 5.15: Príklad rozkladu obrazu pomocou 2D DWT. Všimnite si, ako sú detaily rozdelené do blokov. Každý blok je tvorený DWT transformáciou v horizontálnom, vertikálnom a diagonálnom smere. Nižšia úroveň sa rozkladá do blokov vyššej úrovne.



Wavelet je časť funkcie v tvare vlnky, ktorá, na rozdiel od povedzme funkcie sínus idúcej od nekonečna do nekonečna, má stanovený začiatok, amplitúdu a koniec. Môže mať rôzne tvary, ktoré si vyberáme v závislosti od charakteru analyzovaného signálu.

Waveletová transformácia v podstate hľadá podobnosť waveletu s časťou analyzovaného signálu. Keďže wavelet má svoj začiatok a koniec, môžeme ho rozťahovať v ľubovoľnej mierke. Ak použijeme presne stanovené kroky mierky, získame viacúrovňové waveletové spektrum.

DWT transformácia vytvára 2 súbory vzoriek, vzniknuté filtráciou dolno-, resp. hornopriepustným filtrom. Na úspešnú rekonštrukciu signálu však potrebujeme poznať iba vzorky z hornopriepustného filtra každej úrovne rozlíšenia. Tie predstavujú detaily, ktoré musíme pridať k odhadu signálu v nižšej úrovni, aby sme mohli zrekonštruovať vyššiu úroveň.

GIF

Formát **GIF** (*Graphics Interchange Format*) je na Internete stále populárnym formátom obrázkových súborov. Bol uvedený v roku 1987 a predstavuje bitmapový formát s podporou 8-bitovej farebnej palety, priehľadnosti a má dobrý kompresný pomer. Kvôli obmedzenej palete farieb (255 farieb) má obmedzené použitie pre obrazy s vysokou vernosťou, ako sú fotografie. Je však vhodný na obrázky s obmedzeným počtom farieb, s ostrými hranami a minimálnymi farebnými prechodmi, ako sú napríklad logá. Druhá verzia formátu (1989) prináša podporu priehľadnosti.

GIF na kompresiu obrazových údajov používa algoritmus Lempel-Ziv-Welch (LZW), ktorý farbám vo farebnej palete priradzuje bajtové sekvencie zo slovníka.



Napriek tomu, že v súčasnosti existujú pokročilejšie algoritmy, napríklad PNG, formát GIF si stále udržuje svoju popularitu vďaka svojej podpore pre animácie umiestnením viacerých obrázkov jeden nad druhý. Táto vlastnosť bola využitá na vytváranie plnofarebných (24-bitových) obrázkov a animácií umiestnením troch

8-bitových rámcov na seba, kde každý rámec obsahuje časť 24-bitovej farebnej palety.

PNG

Portable Network Graphics (prenosná sieťová grafika) je bitmapový obrazový formát určený na nahradenie formátu GIF, ktorý je zaťažený licenčnými aj technickými obmedzeniami. Bol predstavený v roku 1996 a medzinárodným štandardom ISO/IEC sa stal v roku 2004. **PNG** podporuje farebné priestory RGB a RGBA s 8 bitmi/farba (24-bitové RGB a 32-bitové RGBA).

Formát PNG je veľmi flexibilný vďaka svojej štruktúre podobnej kontajneru. Obraz sa vytvorí po kúskoch, ktoré umožňujú rozložiť obrazovú informáciu, podporujú vrstvenie a streamovanie dát.

Bezstratová kompresia pracuje v dvoch krokoch:

- Predkompresia (filtrovanie)
- Kompresia

Počas predkompresie sa množstvo obrazových dát zredukuje metódou podobnou DPCM: hodnota pixelu sa ukladá ako rozdiel medzi ním a pixelom vľavo, nad ním, vľavo nad ním alebo ich kombinácie. Na každý riadok pixelov je možné použiť iný filter. Potom sa hodnoty komprimujú algoritmom DEFLATE, ktorý nahrádza rovnaké reťazce referenciami, a Huffmanovo kódovanie používa nie na celý obraz, ale na jednotlivé bloky dát.



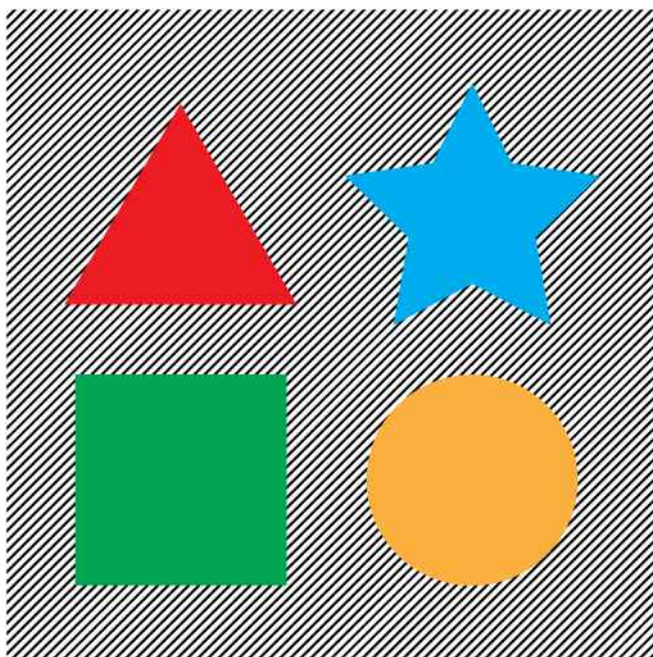
Prvý návrh na názov formátu bol PING – „PING is not GIF“ (PING nie je GIF).

—

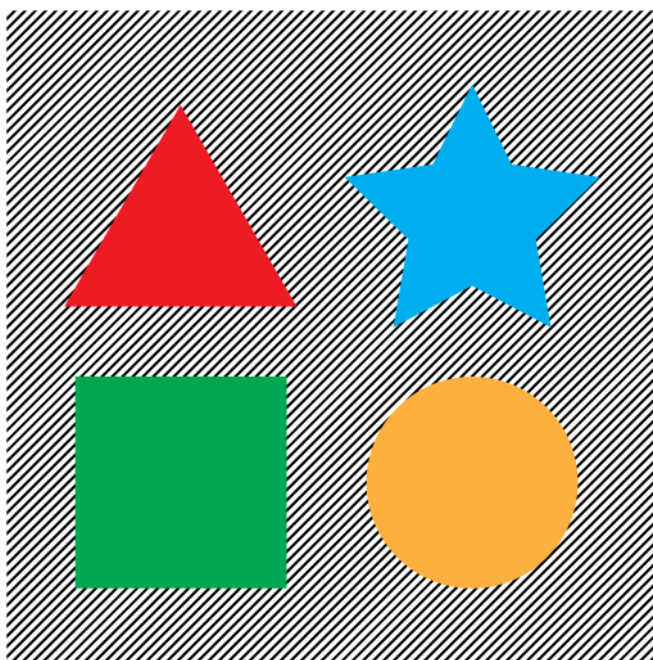
V porovnaní s formátom JPEG vytvára formát PNG z fotografií s jemnými prechodmi väčšie súbory.

+

Napriek tomu má formát JPEG veľké problémy pri spracovaní ostrých prechodov a hrán, ako je text, línie alebo grafika, a na veľkých plochách rovnakej farby, kde vytvára artefakty. PNG je schopný lepšej kompresie, po ktorej nezostávajú žiadne artefakty, takže je ideálny na použitie na webe.



Obr. 5.16: Porovnanie obrázkov zakódovaných vo formáte JPEG (vyššie) a PNG (nižšie). Pri JPEG verzii sú na hranách viditeľné nedokonalosti na hranách. Veľkosť súboru JPEG je 61 kB, súbor PNG má iba 11 kB.



WebP

Najmladší z obrazových formátov, WebP, uviedli laboratória spoločnosti Google v roku 2010. Prezентujú ho ako nový otvorený štandard, ktorý chce súťažiť so stále populárnym JPEG-om. WebP v sebe spája najlepšie vlastnosti JPEG (dobré spracovanie plnofarebnej grafiky), JPEG 2000 (stratová aj bezstratová

kompresia), PNG (priehľadnosť v stratovom aj bezstratovom režime) a GIF (podpora animácií).

Stratový algoritmus je založený na postupoch vo video formáte VP8. Kompresia je založená na predikcii obrazových blokov z troch blokov nad a jedného naľavo od analyzovaného bloku s použitím jedného zo štyroch režimov: horizontálny, vertikálny, DC (jedna farba) a TrueMotion. Zle predikované a nepredikované bloky sa potom komprimujú v blokoch 4x4 pixelov pomocou CDT alebo Walsh-Hadamartovej transformácie. Výstup sa entropicky zakóduje.

Okrem bežných techník ako je slovníkové a Huffmanovo kódovanie používa bezstratový algoritmus pokročilé metódy ako rozličné entropické kódy pre rôzne farebné kanály alebo dočasná pamäť nedávno použitých farieb.



V porovnaní s obrazovými formátmi JPEG a PNG sa WebP javí prekonať svojich konkurentov vo svojich zameraniach najmenej o 20%. WebP je v súčasnosti podporovaný na Linux a Windows pomocou pluginov a podporujú ho aj prehliadače Firefox, Chrome a Opera.

5.4 Kompresia videa

Video, alebo pohyblivý (dynamický) obraz, je v dnešnej dobe najpoužívanjšou zložkou multimedialneho obsahu. Množstvo video obsahu rapídne stúpa a v porovnaní s inými typmi médií zaberá výrazne najviac priestoru. S rastúcimi možnosťami osobných zariadení vo vytváraní a prehrávaní videa sa množstvo prenášaných dát zväčšuje každý deň. Preto je potrebné znížiť veľkosť videa na zníženie nákladov na prenos a ukladanie.

Video sekvencia pozostáva zo samostatných snímok, alebo obrazov, ktoré je v podstate možné kódovať rovnako ako statické obrazy. Avšak pri sekvencii rýchlo sa meniacich obrazov sa predpokladá výskyt toho istého objektu s meniacou sa pozíciou v čase, čo znamená, že dve nasledujúce snímky sú si podobné, iba s minimálnymi rozdielmi. Preto ak zakódujeme iba tieto rozdiely, môžeme dosiahnuť ešte vyššiu kompresiu.

Metódy kódovania videa môžeme rozdeliť na dve hlavné kategórie:

- Kódovanie v časovej oblasti
- Kódovanie na báze modelovania

Metódy kódovania v časovej oblasti využívajú transformačné kódovanie s medzisnímkovou predikciou a sú využívané v mnohých formátoch. **Metódy kódovania na báze modelovania** sa používajú pri aplikáciách s obmedzeným prenosovým pásmom, ako je videotelefónia, s dátovým tokom do 64 kbit/s. Metódy využívajú typický vzťah scén pri videotelefonovaní s nemeniacim sa obsahom, minimálnym pohybom a zníženou frekvenciou snímkovania.

Medzisnímková predikcia a kompenzácia pohybu

Na zníženie časovej nadbytočnosti medzi dvomi snímkami sa používa medzisnímková predikcia s kompenzáciou pohybu, ktorá pracuje v dvoch krokoch:

- Odhad pohybu
- Kompenzácia pohybu

Počas odhadu pohybu sa vytvára vektor pohybu, ktorý popisuje relatívny pohyb obrazových blokov z predošlej do aktuálnej snímky. Ak je pozícia predvídateľná, potom stačí preniesť iba vektor pohybu. Na odhad pohybu sa používajú dva algoritmy:

- Iteračné algoritmy (pel recursive algorithms)
- Algoritmy s blokovým prehľadávaním (block matching algorithms)

Iteračné algoritmy sa iteratívnym spôsobom snažia minimalizovať chybu predikcie. Keďže veľmi závisia od miestnych štatistických vzdialeností, nemôžu sa použiť na odhad väčších vzdialeností a sú tak vhodné pre video s malým

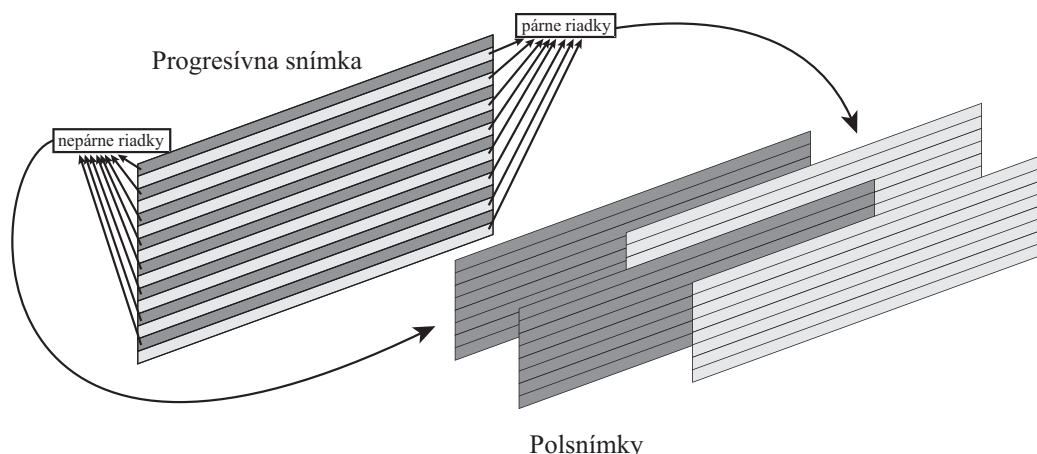
pohybom, napr. videotelefóniu. Algoritmy s blokovým prehľadávaním predpokladajú, že všetky časti obrazového bloku sa pohybujú rovnakým smerom. Aktuálna snímka sa rozdelí na bloky, pre ktoré sa v predošlej snímke nájdu najpodobnejšie bloky.

Algoritmy na kompenzáciu pohybu využívajú vektory pohybu na presúvanie každého bloku z predošlej snímky na novú pozíciu v aktuálnej snímke. Tým vznikne predikčná snímka, ktorá sa zakóduje a prenáša.

Techniky kompresie videa

Prekladané video

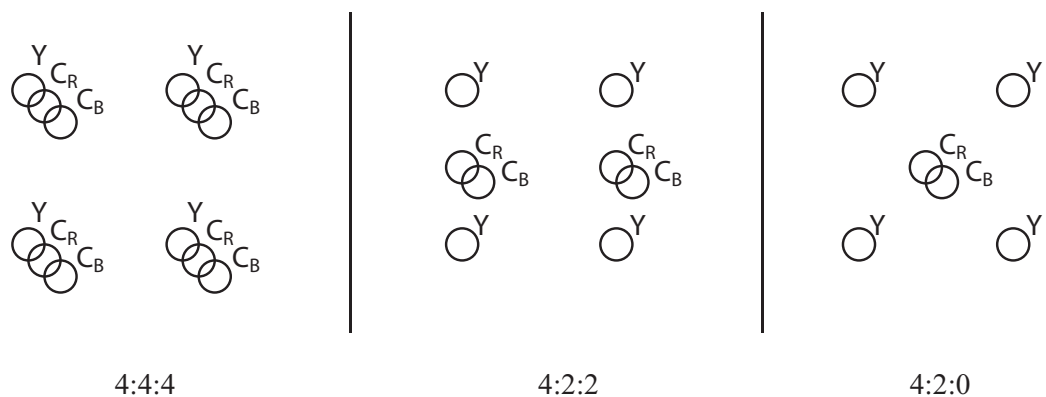
Štandardné (progressívne alebo neprekladané) video pozostáva z 25 (Európa) alebo 29,97 (USA) snímok za sekundu. Prekladané video sa skladá z dvojnásobného počtu polsnímok za sekundu (50 v Európe alebo 59,94 v USA). Každá polsnímka obsahuje buď párne alebo nepárne riadky celých video snímok, ktoré sa dekodujú a prehrávajú v prekladanom poradí.



Obr. 5.17: Rozdiel medzi progresívnym a prekladaným videom. Celá snímka sa rozdelí na párne a nepárne riadky, čím sa vytvoria polsnímky, ktoré sa striedavo uložia do sekvencie. Video má dvojnásobnú snímkovú rýchlosť, avšak pre polsnímky.

Farebné podvzorkovanie

Ďalšou technikou kódovania je farebné podvzorkovanie. Každý pixel video snímky je kódovaný vo farebnom priestore YUV pozostávajúcom z troch subpixelov: jeden jasový (luminančný) a dva farebnostné (chrominančné) pixely. Štvorec štyroch takýchto pixelov sa označuje 4:4:4. Keďže ľudské oko je citlivejšie na jasovú zložku než na zmeny farby, čo nám umožňuje znížiť počet chrominančných pixelov na polovicu, či dokonca na štvrtinu. V prvom prípade sú štyri luminančné pixely pokryté dvoma a dvoma chrominančnými pixelmi (4:2:2) a v druhom prípade iba dvoma, resp. jedným a jedným chrominančným pixelom (4:2:0 alebo 4:1:1).



Obr. 5.18: Príklady farebného podvzorkovania. Ľavý obrázok ukazuje, že každý luminančný subpixel má svoju dvojicu chrominančných subpixelov. Uprostred dva luminančné subpixely zdieľajú pár chrominančných subpixelov. Napravo sú pre štyri luminančné subpixely dostupné iba dva chrominančné.

Podobne ako pri statických obrazoch sa kompresné techniky pri videu opierajú o hybridné metódy kombinujúce kódovanie v časovej oblasti s transformačným kódovaním (ako napríklad diskretná kosínusová (DCT) alebo diskretná waveletová transformácia (DWT)). Predikčná snímka vytvorená v procese kompenzácie pohybu sa odčíta od aktuálnej snímky a vytvorí sa chybová snímka. Tá sa po blokoch zakóduje pomocou DCT, transformačné koeficienty sa kvantizujú, zoradia cik-cak čítaním a zakódujú pomocou kódu s premenlivou dĺžkou (VLC). Potom sa posiela len VLC sekvencia.

Na strane prijímača sa rekonštrukcia vykoná inverzným postupom, teda inverzným VLC, inverznou kvantizáciou a IDCT.

Popísaný postup sa s malými odlišnosťami používa v kódovacích štandardoch MPEG aj H26x.

MPEG

Expertnú skupinu pre pohyblivý obraz (Motion Picture Experts Group, MPEG) zriadili organizácie ISO a IEC, aby vytvorila štandardy pre kompresiu videa a zvuku. Ako výsledok boli navrhnuté nasledujúce štandardy zameriavajúce sa na rôzne aplikácie:

- MPEG-1
- MPEG-2
- MPEG-4

Štandard MPEG-1 sa zameriava na interaktívne systémy na báze médií CD-ROM. MPEG-2 rozširuje schopnosti MPEG-1 pre digitálnu televíziu a TV s vysokým rozlíšením (HDTV). MPEG-4 sa zameriava na multimedialne aplikácie s veľmi nízkymi bitovými tokmi.

MPEG-1

Štandard bol vyvinutý na kódovanie video signálov s dostatočnou kvalitou pri bitovom toku 1,4 Mbit/s. Podporuje rýchle pretáčanie a posun vpred a pozastavenie obrazu. Hoci typická veľkosť video snímky je 352x288 pixelov (CIF), kodek podporuje veľkosť snímky do 720x576 pixelov pri 30 snímkach/sekunda a bitovom toku 1,86 Mbit/s.

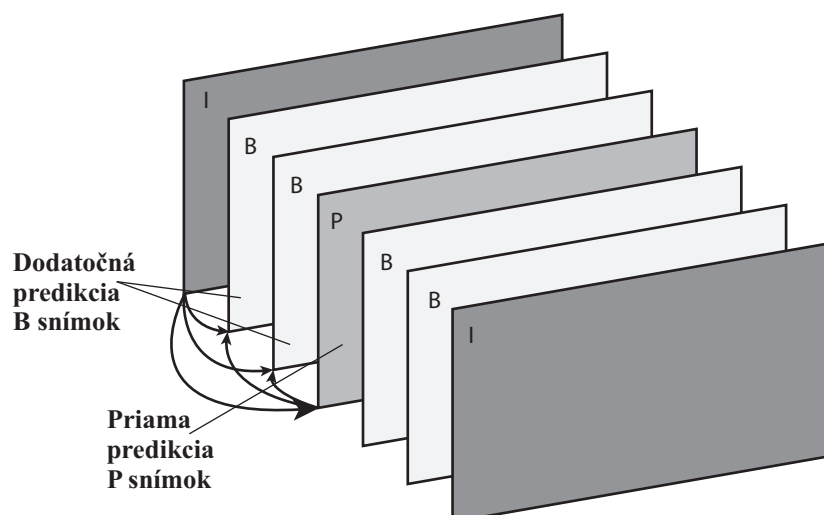
Medzislímkové kódovanie v MPEG-1 je založené na medzislímkovej predikcii a DCT kódovaní. Sú definované 3 typy makroblokov (sád štyroch lúminančných a dvoch chrominančných blokov), z ktorých vyplývajú tri druhy snímkov:

- **I snímky** – Intra snímky – kódovanie snímkov
- **P snímky** – Predikované snímky – medzislímkové kódovanie s priamou predikciou
- **B snímky** – Obojsmerné snímky – dvojsmerná predikcia/interpolácia

Snímky I sa v blokoch 8x8 kódujú DCT, kvantizujú a po cik-cak zoradení entropicky zakódujú. Nevykonáva sa žiadny odhad pohybu, takže snímka sa správa ako fotografia, je nezávislá od ostatných snímkov a slúži ako bod zastavenia pri rýchlom pretáčaní a posune vpred.

Snímky P sa kódujú s použitím medzislímkovej predikcie a porovnaním s predošlou snímkou. Tým sa vytvorí chybová snímka (reprezentovaná vektorom pohybu), ktorá sa rozdelí na makrobloky veľkosti 16x16 pixelov. Tie sa zakódujú pomocou DCT, kvantizujú a entropicky zakódujú podobne ako pri I snímkach. Napriek tomu tieto snímky neobsahujú celú obrazovú informáciu, pretože závisia od predošlých snímkov slúžia len ako referenčné snímky pri predikcii, nie pre rýchle prehrávanie.

Snímky B sú získané iba pomocou doprednej a spätnej predikcie z I alebo P snímkov. B snímky zvyčajne slúžia iba na výplň alebo pridanie detailov v rýchlych scénach medzi I a P snímkami, pretože obsahujú rovnaké informácie ako P snímky. Z B snímkov sa nevykonáva predikcia.



Obr. 5.19: Radenie I, P a B snímok a závislosti medzi nimi

Rámce je možné kombinovať flexibilne podľa požiadaviek aplikácií. Sekvencia IIIIIIIII ponúka výborný prístup k snímkam a rýchlemu prehrávaniu a posunu vpred, avšak nízka kompresia pri I snímkach zvyšuje nároky na prenosové prostriedky. Zvyčajne sa preto používa sekvencia IBBPBBPBBPBB(I) (nazývaná Group of pictures, GOP), pri ktorej sa I snímka používa iba približne dvakrát za sekundu.

MPEG-2

MPEG-2 je rozšírením predošlého štandardu (so spätnou kompatibilitou) s možnosťou prekladaného videa, vylepšeným maximálnym rozlíšením obrazu, TV kvalitou videa pri bitových tokoch 4 až 8 Mbit/s a HDTV kvalitou videa pri 20 Mbit/s.

Vlastnosti dekodérov sú dané prostredníctvom profilov a úrovní. Každá úroveň definuje sadu parametrov určujúcich cieľovú aplikáciu videa, zatiaľ čo profily určujú zložitosť použitých algoritmov. Nasledujúce tabuľky podrobne popisujú úrovne a profily.

Prehľad úrovní v MPEG-2

Úroveň	HIGH (vysoká)	HIGH 1440 (vysoká 1440)	MAIN (hlavná)	LOW (nízka)
Parametre	1920x1152 pix. 60 sn/s 80 Mbit/s	1440x1152 pix. 60 sn/s 60 Mbit/s	720x576 pix. 30 sn/s 15 Mbit/s	352x288 pix. 30 sn/s 4 Mbit/s

Profil	Algoritmy
High (vysoký)	Všetky funkcie profilu Spatial Scalable, navyše režimy kódovania s 3-vrstvovým priestorovým a SNR škálovaním Farebný model YUV 4:2:2 pre náročné úlohy
Spatial Scalable (priestorové škálovanie)	Všetky funkcie profilu SNR Scalable, navyše režim kódovania s 2-vrstvovým priestorovým škálovaním Farebný model YUV 4:2:0
SNR Scalable (SNR škálovanie)	Všetky funkcie profilu Main, navyše režim kódovania s 2-vrstvovým SNR (signal-to-noise ratio) škálovaním Farebný model YUV 4:2:0
Main (hlavný)	Bez škálovania, kódovanie prekladaného videa Náhodný prístup k snímkam, režim predikcie s B snímkami Farebný model YUV 4:2:0
Simple (jednoduchý)	Podporuje všetky funkcie profilu Main okrem režimu predikcie s B snímkami Farebný model YUV 4:2:0

Škálovanie videa

Škálovanie umožňuje dekodérom prehrávanie videa s nízkym dátovým tokom, ak nie sú schopné prehrať video s vysokým dátovým tokom. Dekodér prijme video s nízkou kvalitou a doplnujúce informácie, ktoré mu umožnia zvýšiť kvalitu videa. Pomocou SNR škálovania sa DCT koeficienty škálujú nahrubo, čím vznikne video s nízkym dátovým tokom. Potom sa rozdiel medzi hrubou kvantizáciou a skutočnými hodnotami znova kvantizuje pomocou jemnejšej kvantizácie a táto informácia sa prenáša samostatne. To umožňuje zvýšenie kvality videa na požiadanie. Pri priestorovom škálovaní sa video kóduje najskôr pri nižšom rozlíšení a vyššie rozlíšenie sa dosiahne doplnkovými dátami. Ak dekodovacie zariadenie nepodporuje vyššie rozlíšenie, vynechá doplnkové dáta a dekóduje iba video s nízkym rozlíšením. Časové škálovanie funguje podobne. Vytvorí sa video s nízkym počtom snímok a pomocou doplnkových dát je možná rekonštrukcia videa s vyšším počtom snímok. Priestorové a časové škálovanie sa dá kombinovať, čo prináša variabilitu pri kódovaní videa, napríklad na podporu HDTV aj TV v štandardnom rozlíšení.



Organizácie ISO a ITU-T vytvárali štandard MPEG-2 spoločne, pričom ITU-T ho nazvalo H.262.

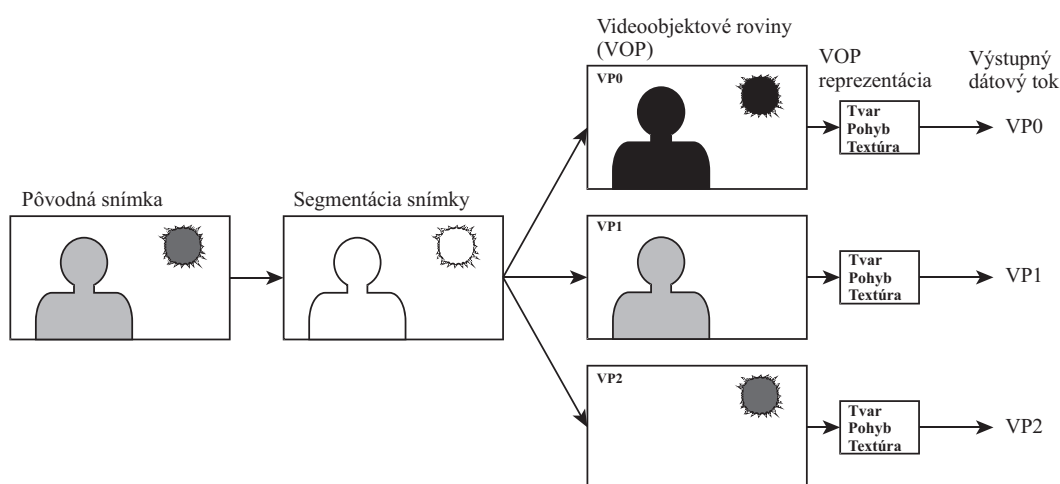
MPEG-4

Štandard MPEG-4 bol vyvinutý na podporu veľmi nízkych bitových rýchlostí do 64 kbit/s. Jeho cieľom je podpora videa na Internete, v mobilných zariadeniach a sieťach a podpora interaktivity s objektmi na scéne. To vyžadovalo vylepšenie kompresných metód, ktoré využívajú kódovanie video objektov v prirodzenom (štandardnom) a syntetickom (kódovanie modelov objektov) videu.



Štandard MPEG-4 existuje v dvoch verziách. Prvá z nich sa označuje Časť 2 a používa ju množstvo kodekov vrátane DivX, XviD, Nero Digital a ďalších. Druhá verzia sa označuje Časť 10 alebo tiež MPEG-4/H.264 AVC (Pokročilé kódovanie videa), a používa sa v kodekoch x264, Quicktime alebo pre médiá HD video, ako napríklad Blu-ray Disc.

Kódovanie prirodzeného videa sa deje prostredníctvom detekcie a kódovania videoobjektových rovín (video object planes, VOP). Každá VOP obsahuje informácie o tvare a textúre objektu na scéne. Sekvencia VOP-ov predstavujúcich ten istý objekt sa nazýva videoobjekt (VO). Každý videoobjekt je možné kódovať s iným dátovým tokom, čo umožňuje flexibilné priradovanie bitového toku a ďalšie úkony s objektmi (škálovanie, rotácia zmena jasu a farby).



Obr. 5.20: Ukážka použitia video objektov

Videoobjekt je určený svojím tvarom daným binárnou maskou alebo maskou v odtieňoch sivej. Kódovanie pohybu je založené na podobných princípoch, aké sú použité v predošlých štandardoch MPEG, ale aplikujú sa na videoobjektové roviny, čím sa vytvárajú IVOP, PVOP a BVOP snímky. Priestorová nadbytočnosť sa odstráni pomocou DCT a časová nadbytočnosť použitím kompenzácie pohybu. Textúry videoobjektov sa kóduje prostredníctvom Shape Adaptive (SA) DCT, modifikácie DCT. Alternatívne sa môže použiť kódovanie pomocou SA DWT (diskrétnaj waveletovej transformácie).

Syntetické video umožňuje vytváranie umelých objektov a ich vkladanie medzi existujúce videoobjekty na scéne. Hlavným dôvodom je umožnenie animácie tváre pre multimediálne aplikácie.

Štandardy H.261 a H.263

Štandard H.261 bol publikovaný v roku 1990 s použitím vo videotelefónii a video konferenciách s nízkymi dátovými tokmi od 64 do 1920 kbit/s a nízkym oneskorením. Zjednocuje rôzne televízne štandardy s rôznym počtom riadkov a frekvenciou polsnímok (PAL a SECAM s 625 riadkami pri 50 Hz, NTSC s 525 riadkami pri 60 Hz). V základe kodek používa rozlíšenie CIF (352x288) a QCIF

(176x144), z ktorých jedno rozlíšenie sa používa pre video konferencie s viacerými účastníkmi a druhé pre videotelefóniu, v ktorej sa zvyčajne prenáša iba hlava a ramená jedinej osoby.

Rozlíšenia CIF a QCIF sa skladajú zo skupín blokov (GOB): CIF z GOB 1-12 a QCIF z GOB 1, 3 a 5. Každá GOB sa skladá z 33 makroblokov pozostávajúcich zo 6 blokov: 4 jasových (luminančných) a dvoch farebných (chrominančných – C_R , C_B), každý zložený z 8x8 pixelov.

Kodek H.261 využíva iba I snímky (označované ako kľúčové snímky) a P snímky, ktoré sa získajú pomocou predikcie pohybu z I snímok alebo predošlých P snímok. Štandard nevyužíva B snímky.

Kódovací algoritmus využíva hybridné blokové kódovanie s medzisnímkovou predikciou a kompenzáciou pohybu a transformačné kódovanie na báze DCT, ktoré je podobné kódovaniu MPEG-1. Po odstránení priestorovej a časovej nadbytočnosti sa každý blok transformuje DCT, kvantizuje, pomocou cik-cak čítania zoradí do vektora a zakóduje Huffmanovým (bezstratovým) kódovaním. Navyše sa používa iteračný filter, ktorý vyhladzuje rozdiely medzi blokmi predikovaného obrazu, čím sa vylepšuje medzi snímková predikcia.

Iteračný filter pracuje so sekvenciou snímok a odstraňuje blokové artefakty vzniknuté DCT transformáciou každého bloku. Jeho úlohou je vyhladiť ostré hrany medzi blokmi snímky. Vyhladzovanie prebieha opakovane v slučke, až kým sa dosiahne žiadaná hranica. Hoci spracovanie slučky môže zaberať viac času než samotné dekódovanie, no po procese odhadu pohybu tak môže byť potrebné zakódovať menší vektor pohybu.

Štandard H.263 prináša v porovnaní s H.261 efektívnejšie kódovanie. Vďaka použitiu niektorých techník z MPEG-1 dosiahol kodek zníženie bitovej rýchlosti o 50%, pričom zachováva rovnakú subjektívnu kvalitu. V porovnaní s H.261 štandard H.263 prináša širšiu podporu video formátov (SQCIF, 4CIF, 16CIF), vylepšený odhad vektoru pohybu, modifikované VLC kódovanie a zavádza PB snímky.

Odhad pohybu v H.263 pracuje s polpixelovou (half-pel) predikciou. Zatiaľ čo vektory pohybu v H.261 pracujú s celými číslami, vektory v H.263 sú reprezentované s presnosťou 0,5. Okrem toho sa vektor pohybu makrobloku odhaduje porovnaním s vektormi pohybu okolitých makroblokov (výpočet mediánu) a prenáša sa iba rozdiel medzi odhadovaným a skutočným vektorom pohybu (nazýva sa mediánová predikcia).

Režim PB snímok pracuje podobným spôsobom ako v kodeku MPEG-1. P snímka sa získa z I snímky alebo P snímky. B snímka sa získa obojsmernou predikciou z okolitých snímok. Rozdiel medzi MPEG-1 a H.263 je v tom, že B snímka v H.263 je uložená priamo v P snímke, čím sa vytvára PB snímka. Je to výhodné najmä pre video s nízkym dátovým tokom.



Rozšírenie štandardu H.263, štandard H.263+, prináša odolnosť voči chybám prenosu, dynamické rozlíšenie scény a škálovanie snímok.

MPEG-4 AVC/H.264

Ďalšia (a novšia) verzia štandardu MPEG-4, označovaná ako pokročilé kódovanie videa (Advanced Video Coding, AVC), je v súčasnosti najpoužívanejším štandardom. Spravovaný je spoločne organizáciami ISO a ITU-T a je vhodný najmä na kompresiu videa vo vysokom rozlíšení.

Štandard prináša množstvo vylepšení predošlých štandardov oboch organizácií, ako je vyššie rozlíšenie farebnej informácie, škálovateľné kódovanie videa a viacpohľadové kódovanie videa, ktoré umožňuje kódovanie viacerých uhlov videa, a teda podporuje stereoskopické (3D) video.

Variabilná veľkosť blokov umožňuje precíznu segmentáciu pohyblivých regiónov s veľkosťou od 16x16 do 4x4 pixelov. Z jedného makrobloku môže byť odvodených viacero vektorov pohybu smerujúcich k rôznym referenčným obrazom. Algoritmus kompenzácie pohybu pracuje so štvrtpixelovou presnosťou (v porovnaní s polpixelovou presnosťou u H.263), čo umožňuje vyššiu presnosť vektorov pohybu. DCT transformácia bola vylepšená a upravená tak, aby poskytovala jednoznačné dekódovanie. Navyše sa na hladké oblasti môže použiť sekundárna Hadamardova transformácia, ktorá ešte viac zlepšuje kompresný pomer.

Okrem toho bolo zavedené bezstratové kódovanie makroblokov, ktoré umožňuje dokonalú reprezentáciu vybraných oblastí obrazu. Pracuje v dvoch režimoch: PCM makroblok alebo vylepšený bezstratový makroblok (efektívnejší). Entropické kódovanie používa nové algoritmy na kódovanie prvkov syntaxe a kvantizovaných hodnôt transformačných koeficientov: Kontextovo adaptívne binárne aritmetické kódovanie a Kontextovo adaptívne kódovanie s premenlivou dĺžkou, ktoré kódujú efektívnejšie než v predošlých štandardoch.

Množstvo ďalších vylepšení zachováva rovnakú subjektívnu kvalitu videa ako predošlé štandardy, no pri polovičnom, či dokonca menšom bitovom toku, čo je zjavné najmä pri vysokom dátovom toku a videu s vysokým rozlíšením.

Podobne ako pri MPEG-2 štandard MPEG-4 AVC/H.264 podporuje kódovacie profily na použitie pre rôzne cieľové aplikácie a úrovne definujúce požadovaný výkon dekodéra.

WebM

WebM je audio a video kodek s otvoreným zdrojovým kódom od spoločnosti Google určený na použitie s HTML5 videom. Je to multimediálny kontajner založený na kontajneri Matroska, ktorý obsahuje zvuk kódovaný štandardom Ogg Vorbis a video kódované kodekom VP8.

Kodek VP8 vyvinula spoločnosť On2 Technologies a po tom, ako ju v roku 2010 odkúpila spoločnosť Google, bol uvoľnený pod licenciou s otvoreným zdrojovým kódom. Hoci VP8 využíva mnohé techniky zavedené štandardmi MPEG a H.26x, prináša ďalšie vylepšenia, ktoré umožňujú zachovanie vysokej subjektívnej kvality pri znížení výpočtovej náročnosti. Niektoré z nich teraz uvedieme.

Kodek využíva tzv. konštruovanú referenčnú snímku, ktorá slúži ako referenčná snímka pre kompenzáciu pohybu niekoľkých snímok. Vzhľad konštruovanej referenčnej snímky nie je určený, takže je ponechaný na voľbu návrhárov. Proces iteračného filtrovania, ktorý odstraňuje blokové artefakty po znížení priestorovej nadbytočnosti (DCT transformácii) môže pri každom bloku použiť rôzny počet blokov v sekvencii. Entropické kódovanie využíva najmä binárne aritmetické kódovanie, ktoré sa prispôsobuje každej snímke zvlášť.



Okrem už spomínaných kódovacích štandardov existuje množstvo ďalších video formátov, napríklad Windows Media Video (WMV) od spoločnosti Microsoft alebo Ogg Theora, založený na staršom kodeku VP3 od On2 Technologies, a mnohé ďalšie štandardy, ktoré nepokrývajú organizácie ISO ani ITU-T.

6 Spracovanie multimédií

6.1 Syntéza reči



Pod syntézou reči rozumieme vytváranie reči, ktorá je čo najpodobnejšia ľudskej, za pomoci stroja, ktorý je označovaný aj ako syntetizátor.

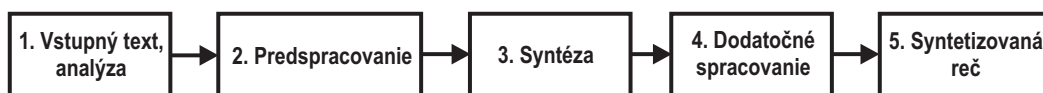
Existuje viacero typov syntetizátorov, avšak cieľom každého z nich je v zásade to isté: reprodukovať zadaný text v čo najzrozumiteľnejšej a najprirodzenejšej forme.

Existujú štyri základné prístupy:

- syntéza pomocou segmentov (napríklad difónová syntéza, kde sú malé, dopredu nahraté vzorky ľudskej reči spájané tak, aby tvorili zmysluplné slová, viď nižšie),
- formantová syntéza (celá reč je generovaná počítačom a nie ľudskými vzorkami, je jednoduchá na implementáciu, avšak znie umelo),
- artikulačná syntéza (prístup založený na modeli ľudského vokálneho traktu a reálnych artikulačných procesoch, nie je veľmi rozšírená),
- HMM syntéza (matematický model generujúci reč na základe kritérií maximálnej pravdepodobnosti).

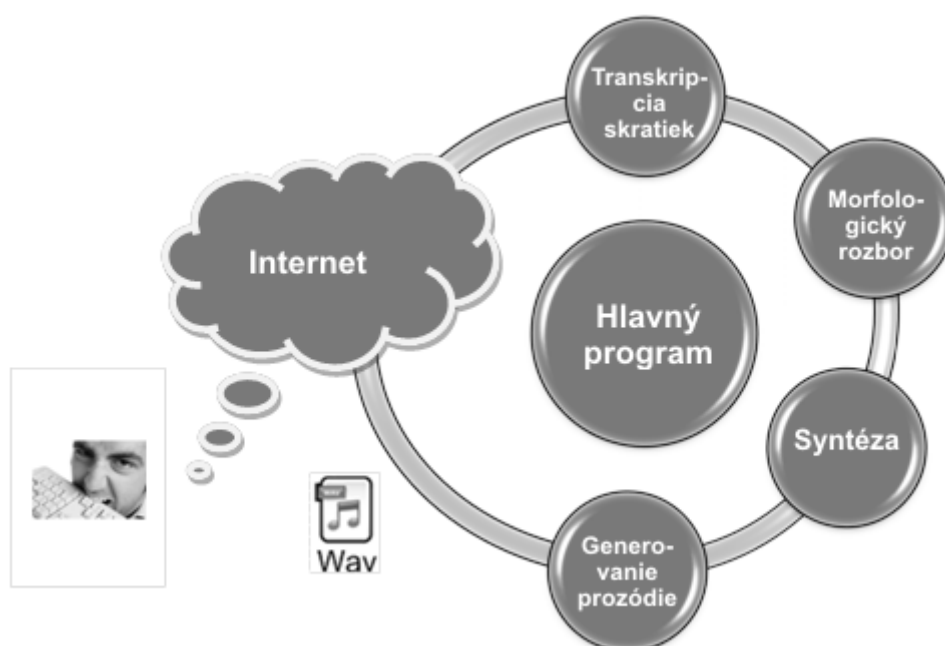
Na obrázku (Obr. 6.1) je blokový diagram všeobecného syntetizátora. Samozrejme, tento diagram je značne zjednodušený pre naše potreby a niektoré elementy (ako napríklad spätné väzby nachádzajúce sa v niektorých učiacich sa syntetizátoroch) sú vynechané. Avšak v podstate každý syntetizátor pozostáva z nasledovných častí:

- vstup textu
- predspracovanie – slová alebo vety sú „preložené“ do špeciálnej formy, resp. jazyka, ktorému rozumie počítač: čísla a všetky skratky sú prepísané do slov v správnej forme, všetky písmená sú napísané v špeciálnej abecede – SAMPA abeceda. SAMPA (z ang. Speech Assessment Methods Phonetic Alphabet) je fonetický preklad používajúci iba tlačiteľné ASCII znaky,
- samotná syntéza,
- post procesing – finálnu syntézu je ešte potrebné spraviť viac prirodzenou, predovšetkým modifikáciou prozódie (tempa reči),
- výstupný blok poskytujúci syntetizovanú reč.



Obr. 6.1: Bloková schéma všeobecného syntetizátora

Pre dosiahnutie čo najprirodzenejšej syntézy reči musia syntetizátory plniť komplexné úlohy ako predspracovanie a následné spracovanie (postprocessing). Pre čo najdokonalejšie podanie reči by tieto systémy mali byť adaptívne a schopné učiť sa. Taký systém by pozostával zo štyroch základných modulov: fonetický preklad slov, určenie slovného druhu (týka sa predovšetkým germánskych a slovanských jazykov používajúcich skloňovanie), fonetického prekladu skratiek a modulu modifikácie prozódie.



Obr. 6.2: Modulárna architektúra syntetizátora

V nasledujúcom príklade sa zameriame na difónovú syntézu.

Príklad difónovej syntézy reči

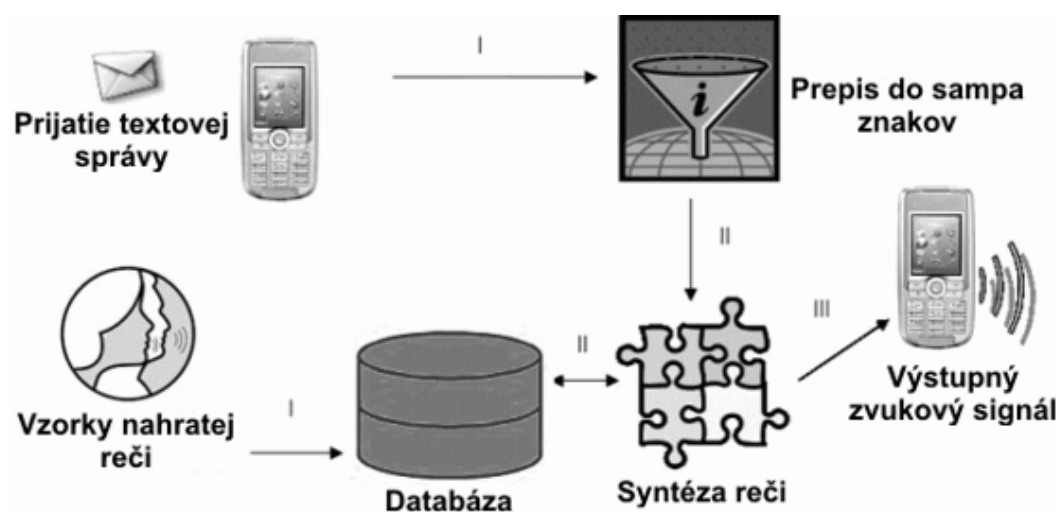
Nasledujúci príklad opisuje využitie difónového syntetizátora. Hlavnou výhodou tohto riešenia je prirodzene znejúci hlas a malá databáza. Slovenský jazyk má iba 1550 difón, a práve to robí nasledujúce riešenie veľmi praktickým (najmä v porovnaní s inými druhmi syntéz, kde je veľkosť databázy oveľa väčšia, ako napr. pri korpusovej syntéze).

Difóna je hneď po fonéme jedna zo základných jednotiek v reči. Fonéma je najmenšia komplexná jednotka zvukového systému jazyka. Jednu difónu tvoria dve po sebe nasledujúce fonémy. Hranice difóny sú v polovici týchto foném, takže jej dĺžka v časovej oblasti nie je dvojnásobná oproti fonémam, ale približne

rovnaká. Výhodou difón oproti fonémam je, že lepšie sledujú prechodovú zmenu zvuku medzi jednotlivými hláskami a majú hranice v ich strede, kde sú časové priebehy najstabilnejšie.

Vo všeobecnosti je počet difón v reči rovný počtu druhej mocniny foném. Tým dostaneme všetky kombinácie. V skutočnosti je ale počet difón menší, lebo nie všetky kombinácie dokážeme povedať, resp. v jazyku sa nenachádzajú. Reálny počet difón zistíme iba dôkladným študovaním jazyka. Difónová databáza obsahuje reálne nahrávky reči, ktoré sú rozdelené na menšie časti – difóny. Je viacero možností, ako vytvoriť a nahráť databázu, napríklad výber takých slov, aby v nich boli obsiahnuté všetky difóny zo slovníka. Slová v slovníku nemusia mať žiaden význam alebo súvis, dôraz je kladený na čo najmenší počet nahrávok.

Návrh rečového syntetizátora je na obrázku (Obr. 6.3). Princíp syntézy je tam znázornený veľmi zjednodušene.



Obr. 6.3: Návrh rečového syntetizátora

Vstupný text sa musí zosyntetizovať na reč. Najskôr sa musí text prepísať do SAMPA abecedy. Preto sa v prvom kroku syntézy všetky znaky prepíšu do SAMPY. V druhom kroku je prepis v SAMPE prepísaný podľa pravidiel výslovnosti pre konkrétny jazyk (v našom prípade slovenský jazyk). Potom sa pre zodpovedajúce difóny vyberú nahrávky z databázy a pospájajú sa. Výstupom je zosyntetizovaný text.



Uvádzame niekoľko príkladov využitia syntézy v praxi: osobní rečovní asistenti, asistenti v mobilných telefónoch pre nevidiacich, mestskí sprievodcovia, dopravný grafikon a navigačný systém, internetové multimodálne služby, aplikácie pre dokumentáciu nehôd, služby manažovania času alebo inventár. V posledných rokoch sa stali veľmi populárne syntéza v slovníkoch priamo na internete (online), alebo čítačky kníh s implementovanou syntézou textu, najmä pre anglický jazyk.

6.2 Rozpoznanie obrazu

Typický problém s obrazom je detekcia, či sa v obrázku nachádza špecifický objekt, alebo či má špecifickú vlastnosť. Tento problém je v súčasnosti vyriešený iba pre špecifické objekty (napríklad detekcia tváre), ale nie pre všeobecné objekty (napríklad zoznam všetkých vecí, čo sa nachádzajú na danom obrázku). Z tejto príčiny rozoznávame viacero oblastí rozpoznávania a detekcie, ako napríklad:



- Čiarové kódy – čiarový kód pozostáva z čiernych čiar rôznej hrúbky, ktoré reprezentujú číslce. Šírky sú samozrejme relatívne, aby mohli byť kódy skenované z rôznej diaľky a rôznej veľkosti.
- 2D kódy – tiež známe ako QR kódy, alebo maticové kódy sú v zásade rozšírením čiarových kódov o ďalšiu dimenziu. 2D kódy dokážu uložiť oveľa väčšie množstvo informácií ako jednoduché čiarové kódy, vrátane URL adries, textu alebo čísiel. Môžu mať rôznu veľkosť aj hustotu a majú implementované RSkódy, ktoré majú schopnosť detekcie a opravy chýb.
- Optické rozpoznanie znakov (Optical Character Recognition OCR) – rozpoznávanie znakov (písmen a čísiel) z obrázkov tlačeného, alebo písaného textu. Tento proces je v zásade špecifickým prípadom rozpoznávania vzorov.
- Identifikácia odtlačkov prstov – druh algoritmu rozpoznávajúceho vzory, ktorý porovnáva základné charakteristiky odtlačkov (a pattern based algorithm compares basic fingerprint patterns (oblúky a slučky) na skúmanom obrázku s uloženým originálom.
- Špecifické detekcie – obraz je prehľadaný za účelom identifikácie špecifických podmienok, napríklad detekcia pohybu v bezpečnostných kamerách.
- Rozpoznávanie tváre – obrázok je prehľadaný za účelom identifikácie špecifických čŕt ľudskej tváre
- Rozpoznávanie objektov – jeden, alebo viacero dopredu špecifikovaných, alebo naučených druhov objektov je automaticky rozpoznávaných. Príkladom môžu byť program ako napríklad Nokia Lens, alebo Google Goggles automaticky zobrazujúce popis rôznych objektov na obrazovke v reálnom čase.

Každý typ rozpoznávania používa vlastný špecifický algoritmus, ktorý môže byť jednoduchý, ale aj zložitý – napríklad použitie štatistických metód alebo neurónových sietí. Ako príklad algoritmu rozpoznávania uvádzame jednoduchý algoritmus detekcie tváre s bodmi pre čelo a brady.

Detekcia bodov tváre

Detekcia bodov tváre je založená na farebnosti ľudskej kože a morfolologickej charakteristike ľudskej hlavy. Výstup algoritmu pre detekciu kože je dôležitý na oddelenie významných bodov skúmanej tváre. Prvým bodom pre správnu detekciu je odstránenie šumu pomocou dolno-priepustného filtra. Po takomto filtri nám nebudú vznikať nechcené (a nesprávne) detekcie. Ďalším dôležitým krokom je zistenie hrán. Na to sa používa Sobelov filter. Po týchto nutných úpravách je obrázok predspracovaný na detekciu. Nasleduje samotná detekcia jednotlivých častí tváre (ako sú oči, nos a ústa, brada,...) založená na známych charakteristikách týchto bodov.

Detekcia brady prebieha v oblasti pod dolnou perou, postupujeme smerom nadol. Horizontálne prehľadávanie končí asi v tretine obrázku, pretože predpokladáme, že budú splnené podmienky na vstupnú fotografiu (t.j. že brada sa nebude nachádzať ďalej). Hľadáme oblasť, ktorá je charakteristická relatívne pokojným priebehom s následným prudkým stúpaním, ktoré označuje koniec brady.

Detekcia vychádza z porovnávania horizontálnej vzdialenosti dvoch bodov. Oba body ležia na hrane – to znamená na hranici profilu. Y-ová súradnica druhého bodu je však o 10 bodov (10 pixelov – je to hodnota určená konkrétne v algoritme a vychádza z požiadaviek na vstupnú fotku) väčšia. Ak horizontálna vzdialenosť takýchto dvoch bodov (teda rozdiel x – ových súradníc) je väčšia ako je povolená vzdialenosť, tak bod je označený ako bod brady.

Oblasť vyhľadávania bodu na čele sme definovali v prvej tretine obrázku v horizontálnom smere aj vertikálnom smere. Detekcia prebieha na obrázku so zvýraznenými hranami (zo Sobelovho filtra) a na obrázku, na ktorom je výsledok detekcie kože (Skin detect). Hľadaný bod sa nachádza približne na mieste, kde končí pokožka a začínajú vlasy.

Pri detekcii bodu definujúceho čelo vychádzame najskôr zo Sobelovho filtra. Nájdeme bod, ktorý leží na hrane. Potom začneme prehľadávať obrázok „Skin detect“ v horizontálnom smere, pričom postupujeme od detekovaného bodu smerom vpravo. Farebné zložky bodov pokožky na „Skin detect“ obrázku majú hodnotou 0 (čierna farba), farebné zložky ostatných častí tváre majú hodnotu 255 (biela farba). Keď nájdeme prvý bod s hodnotou 255 (v našej oblasti prehľadávania to znamená, že sme našli časť hlavy, kde už začínajú vlasy), zistíme rozdiel x – ových súradníc. Ak je menší ako stanovená hranica, tak bod je náš hľadaný bod – čelo.



Obr. 6.4: Detekcia brady



Obr. 6.5: Detekcia čela

Na základe výsledkov detekcie bodov tváre sa môže deformovať všeobecný 3D model hlavy a môže byť personalizovaný.

6.3 Animácia tváre

Existuje veľa metód na animáciu a modelovanie tváre. Pred tým, než sa vyberie správny spôsob, je potrebné definovať požiadavky na konkrétnu animáciu. Príklad takýchto požiadaviek na modelovanie a animáciu je nasledovný:

- Musí byť animácia nepretržitá, kontinuálna?
- Výkon musí byť optimalizovaný pre PC alebo mobilný telefón?
- Musí sa animácia synchronizovať s rečou?
- Kvalita animácie musí byť čo najlepšia, výsledná animácia by mala byť čo najviac podobná skutočnosti.



Sú dva základné prístupy pre animáciu tváre: dvojrozmerná alebo trojrozmerná a animácia v reálnom čase alebo spätne vypočítaná animácia.

Animácia v reálnom čase dovoľuje používateľovi interaktívne zasahovať do animácie a redukuje čas potrebný pre prípravu animácie. Nevýhodou tejto metódy je kvalita obrazu, keďže čas a tým pádom aj výkon na prípravu obrazu je obmedzený. Kalkulácia potrebná pre vyrátanie každého obrázku by nemala byť dlhšia ako približne 0,05 s, pretože je potrebné dosiahnuť aspoň 20 obrázkov za sekundu. (20 FPS – obrazy (frames) za sekundu). Kvalita trojrozmernej animácie je samozrejme subjektívne lepšia ako kvalita dvojrozmernej animácie, pretože výsledná animácia je viac prirodzená.

V oblasti animácie tváre sú najznámejšie metódy ako interpolácia, parametrizácia, simulácia svalov atď.

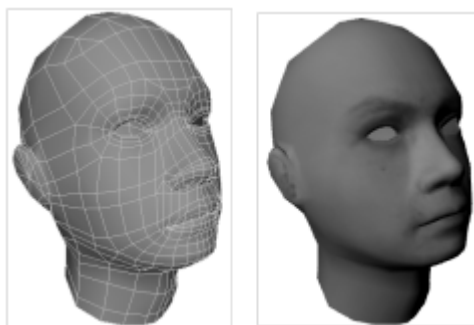


Je veľa rozličných techník na modelovanie ľudskej tváre v priestore, napríklad polygonálne modelovanie, modelovanie pomocou parametrických plôch, modelovanie menších častí.

Hovoriaca hlava v mobilnom telefóne

Táto časť predstavuje projekt aplikácie syntézy reči spolu s animáciou tváre na mobilnom telefóne. Cieľom tohto projektu bolo vytvorenie multimediálneho komunikačného systému neprenášajúceho video a hlas, ale iba text a pomocné informácie pre animáciu tváre. Výsledkom bola mobilná Java aplikácia fungujúca na mobilných telefónoch, ktorá dokázala čítať *krátke textové správy (SMS)*. Po prijatí krátkej textovej správy SMS sa na obrazovke zjavil animovaný obraz hovoriacej hlavy vytvorený z fotografie odosielateľa (fotografia bola dopredu uložená v databáze telefónu). Táto hlava následne čítala prijatú správu.

V našom príklade sa model aj vizémy potrebné pre animáciu nachádzali priamo vo formáte objekt (OBJ), čo znamená, že hlava nemala iba čiary a kontúry, ale aj textúru.



Obr. 6.6: Neutrálly model a textúra tváre

Vizéma je špecificky deformovaný model tváre. Deformácia samozrejme nie je ľubovoľná, ale tvár je deformovaná tak, aby to vyzeralo, že tvár vyslovuje príslušnú fonému. Model vizémy má tie isté čiary a body a schému ako nedeformovaná tvár. Mení sa iba pozícia uzlov modelu. Na základe tejto zmeny je model schopný pomerne jednoducho a s malým nárokom na výkon robiť interpolácie uzlov ako aj zmenu orientácie povrchov definovaných príslušnými uzlami.

Samotná animácia je realizovaná spomenutými interpoláciami vizém. Neutrálly model je prečítaný zo súboru (spolu s ostatnými modelmi a vizémami). Interpolácia sa robí medzi modelmi. Animácia je robená medzi jednotlivými modelmi (model reprezentuje rôzne sa tváriaceho človeka, resp. rôznu mimiku tváre). Animácia je vytváraná v reálnom čase. Model je deformovaný na základe výsledkov detekcie bodov tváre (opísané v predošlej kapitole).



Obr. 6.7: Vizémy

Animácia sa prehráva spolu so syntetizovanou rečou. Rečový syntetizátor použitý v tomto projekte je identický s difónovým syntetizátorom, opísaným v kapitole Syntéza reči. Veľmi dôležité je zosynchronizovať animáciu a syntetizovanú reč.



Obr. 6.8: Ukážka aplikácie na mobilnom telefóne: personalizovaný model (prvý) a všeobecný model (druhý a tretí)

6.4 Rozpoznávanie reči

Rečový signál vytvárajú ľudské orgány reči a pôvodne je reprezentovaný vlnením vzduchu. Okrem iného v sebe obsahuje lexikálnu zložku, ktorá je pre rozpoznávanie reči kľúčová. Lexikálna informácia je zakódovaná v akustickom signáli ako sekvencia akusticky odlišných zvukov. Každý jazyk obsahuje vlastnú sústavu základných zvukov, ktoré sú priradené k fonémam. Napríklad v slovenčine ich je 51. Nanešťastie, bežne vyslovené fonémy sa navzájom ovplyvňujú v časovej aj vo frekvenčnej oblasti, takže vytvárajú akusticky odlišné zvuky. Okrem toho sa každá fonéma líši s každým jej vyhovorením a u každého rečníka (obsahuje charakteristické informácie o rečníkovi). V každom reálnom prostredí ešte navyše existuje aditívny a konvolutórny šum, ktorý situáciu ďalej sťažuje. A aby toho nebolo málo, každý jazyk obsahuje obrovskú slovnú zásobu, zvyčajne niekoľko stoviek tisíc slov, z ktorých každé môže nadobúdať viacero tvarov (pádov, časov, atď.). Je jasné, že úloha je dosť zložitá a výpočtovo náročná.



Bolo vyvinutých mnoho systémov, takže existujú ich základné klasifikácie: systémy s malou, strednou a veľkou slovnou zásobou, systémy závislé a nezávislé od rečníka, systémy na báze foném, slov alebo fráz, systémy pre samostatné slová (diktačné) alebo aj spojitú reč a podobne. V súčasnosti už boli prezentované systémy pracujúce s viac ako 100 tisíc slovami v reálnom čase.

Po mnoho desaťročí sa veľké úsilie vynakladalo na stavbu a využitie systémov ASR (Automated Speech Recognition) v oblastiach ako sú systémy na prístup k informáciám, dialógové systémy, atď., avšak až s rastúcimi technologickými možnosťami sa začínajú vytvárať náročnejšie systémy: diktačné systémy, či dokonca automatický prepis prirodzenej reči. Tieto pokročilé systémy musia byť schopné prevádzky v reálnom čase, byť nezávislé od rečníka, dosahovať vysokú presnosť a podporovať slovníky pozostávajúce zo stoviek tisíc slov. V súčasnosti existujú systémy na rozpoznanie spojenej alebo diktovanej reči v reálnom čase, pričom veľkosť slovníka sa pohybuje v rádoch stoviek tisíc slov.

Metódy extrakcie vlastností reči

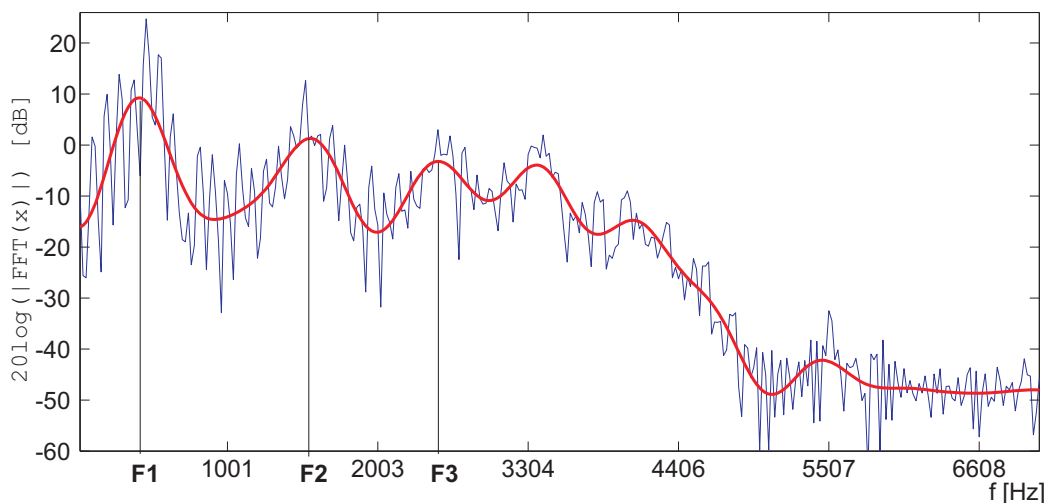
Ako už bolo povedané v úvode, reč je komplexný signál tvorený hlasovými orgánmi človeka. Takýto signál v sebe obsahuje rôzne typy informácií ako: čo sa hovorí, kto hovorí, akú má náladu, odkiaľ pochádza (dialekt), aké má sociálne pozadie, je chorý, má rečovú vadu a mnoho ďalších. Tieto všetky informácie sú nepriamo „zakódované“ pomocou mozgu a hlasových orgánov do výstupného rečového signálu.



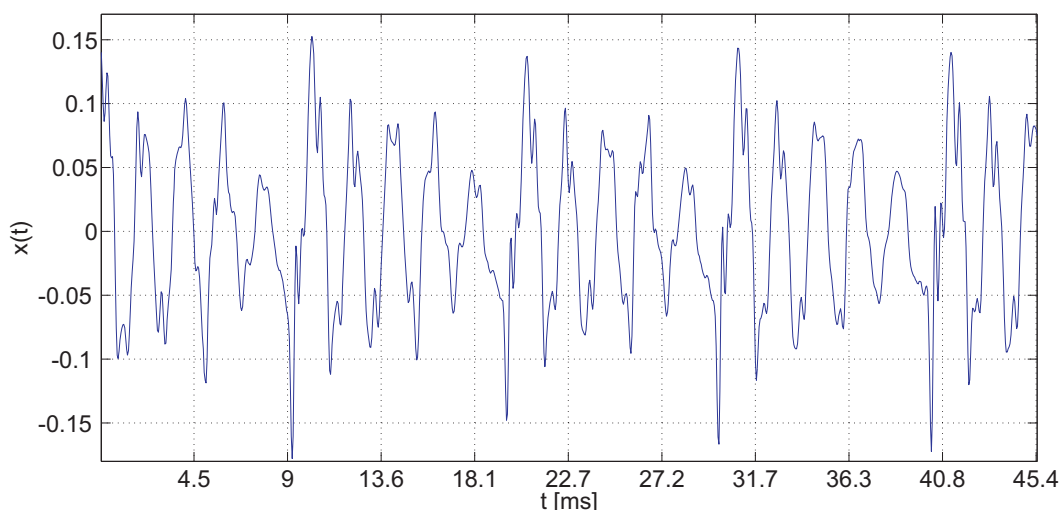
Podľa šírky efektívneho spektra reči a počtu potrebných kvantizačných úrovní sa dá odhadnúť informačný tok rečového signálu, ktorý sa pohybuje okolo 100 kbit/s. Keď sa ale bude skúmať informačná hodnota lexikálnej zložky obsiahnutej v rečovom signáli je podstatne nižšia cca. 10 bit/s.

Je teda zrejmé, že z reči je potrebné extrahovať len veľmi malé množstvo informácie, ktoré bude odrážať práve lexikálnu zložku. Táto úloha je však veľmi komplikovaná, lebo proces „kódovania“ všetkých zložiek do výsledného rečového signálu je zložitý a nie úplne reverzibilný. Preto v prvom kroku návrhu ASR systému sa rozhoduje, ktorú techniku výberu „lexikálnych“ vlastností použiť. Je množstvo metód, ktoré buď simulujú tvorbu rečového signálu (lineárny model tvorby reči) alebo častejšie sa snažia simulovať ľudský sluchový systém, matematicky ho popísať, zjednodušiť pre praktické spracovanie a prípadne ho prispôbiť pre správne a jednoduché používanie s vybranými typmi klasifikačných metód. To má vplyv na efektívne potlačenie aditívnych a konvolučných šumov, ktoré sú prítomné v bežných podmienkach. Predpokladá sa totiž, že evolúcia selektívnym výberom „nastavila“ sluchové orgány spolu s príslušnými časťami mozgu tak, aby tie boli čo najviac citlivé práve na reč a lexikálnu informáciu, čím potláčajú ostatné rušivé vplyvy.

Žiadané vlastnosti metód extrahujúcich lexikálnu informáciu sú okrem redukcie bitovej náročnosti aj vysoká citlivosť na zmeny zvuku, ktoré sú u ľudí vnímané ako rôzne a naopak mala by byť „hluchá“ na tie, ktoré ostávajú nepovšimnuté pre náš sluchový systém. Ukázalo sa, že významnými indikátormi rozdielu sú tzv. formantové frekvencie. Na nasledujúcich obrázkoch sú nakreslené frekvenčné spektrum signálu a jeho časový priebeh. Vo frekvenčnom spektre sú vyznačené polohy formantových frekvencií (z čoho je jasný aj ich význam) ako aj obálka spektra, ktorá súvisí s postavením hlasových orgánov pri vyslovení hlásky „e“. Na obrázku časového priebehu hlásky „e“ je dobre pozorovať hlasivkovú frekvenciu.



Polohy formantových frekvencií a obálka spektra pre hlásku „e“



Časový priebeh samohlásky „e“.

Aby bolo zrejmé ako formanty súvisia s vyslovovaním rôznych samohlások, je v tabuľke naznačené aké sú priemerné polohy prvých troch formantových frekvencií pre mužov a ženy, z nich je zřejmý kladný posun ich polohy pre ženy.

Polohy formantových frekvencií pre samohlásky oddelene pre mužov a ženy

Hláška	Muži			Ženy		
	F1 [Hz]	F2 [Hz]	F3 [Hz]	F1 [Hz]	F2 [Hz]	F3 [Hz]
a	730	1100	2450	850	1200	2800
e	530	1850	2500	600	2350	3000
i	400	2000	2550	430	2500	3100
o	570	850	2400	590	900	2700
u	440	1000	2250	470	1150	2700

Na základe množstva experimentov bolo zistené, že nasledujúce fenomény sú posudzované ako výrazne vnímateľné:

- Rozdiely v počte formantových frekvencií
- Výrazné rozdiely v polohe formantových frekvencií
- Výrazné rozdiely v šírke pásiem formantových frekvencií
- Rozdiel energií signálu (aj frekvenčných zložiek) je vnímaný skôr pomerovo a nie absolútne. Človek teda vníma intenzitu zvuku nelineárne, t.j. hlasitejšie časti potláča a slabšie zosilňuje. Matematicky sa táto vlastnosť dá aproximovať logaritmom.

Naopak, vo všeobecnosti existujú nasledovné spektrálne rozdiely, ktoré neprispievajú k výrazne odlišnému vnemu zvukov:

- Sklon spektra v tvare: $X(\omega)=S(\omega)*\omega^\alpha$, kde α je faktor sklonu
- Orezanie nízko-frekvenčných zložiek ležiacich pred prvou formantovou frekvenciou F1
- Orezanie vysoko-frekvenčných zložiek ležiacich nad treťou formantovou frekvenciou F3
- Potlačenie úzkeho frekvenčného pásma a to aj v pásme potrebnom na dosiahnutie základnej zrozumiteľnosti

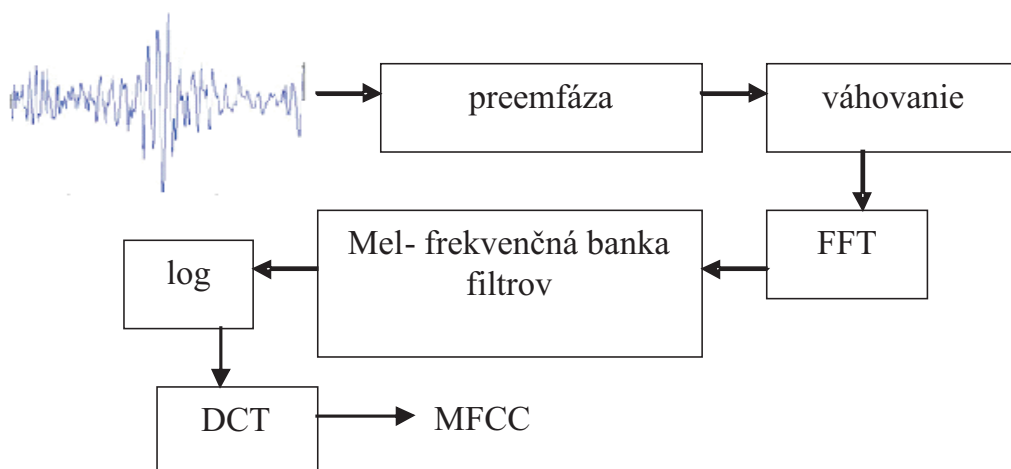
Navyše, vlastnosti extrahovaných príznakov by nemali byť citlivé na aditívny (šum pozadia) a konvolutórny (vplyv nahrávacieho zariadenia a miestnosti - ozveny) šum. Tieto podmienky je ale ťažko súčasne splniť. Preto je vhodné, aby aspoň rečové príznaky boli prezentované tak, že tieto rušenia budú v priestore vlastností ľahko lokalizovateľné a potlačiteľné. Navyše, dobré príznaky musia byť ľahko extrahovateľné súčasnými matematickými postupmi a technickým vybavením v reálnom čase.



Súčasne najpreferovanejšie akustické príznaky sú MFCC a PLP, ktoré sú navrhnuté na zachytávanie pozícií a širok formantov, ktoré sú akusticky vnímateľné. Samozrejme existuje celý rad ďalších metód, ktoré môžu dosahovať ešte lepšie vlastnosti, ale tie sú zväčša špecializované na užšiu oblasť praktického použitia, napr. sú určené pre konkrétne prostredie.

Výpočet MFCC koeficientov

Skratka MFCC znamená Melove frekvenčné keprálne koeficienty. Tieto koeficienty teda opisujú modifikované keprálne koeficienty, ktoré sú počítané nie vo frekvenčnej ale Melovej škále, ktorá lepšie odráža psychoakustické vlastnosti ľudského vnemu. Postup výpočtu MFCC koeficientov je graficky znázornený na obrázku.



Postup výpočtu MFCC

Výpočet PLP koeficientov

Skratka PLP znamená perceptuálne lineárne predikčné koeficienty. Tieto koeficienty tiež opisujú modifikované kepstrálne koeficienty, ale ich postup extrakcie je komplexnejší ako v prípade MFCC, lebo zahŕňa viac psychoakustických javov. Ich výpočet je nasledovný: segmentácia signálu a Hammingove váhovanie, výpočet spektra (fft), transformácia na Barkovú frekvenčnú os, vyhľadanie (DP filtrácia) Barkovho spektra, ktoré simuluje funkciu kritické pásma, vzorkovanie spektra, ktoré zodpovedá šírke kritického pásma, teda 1 bark, váhovanie kopírujúce krivku rovnakej hlasitosti, transformácia energie do hlasitosti, výpočet modelu produkcie reči (LPC), prepočet modelu do kepstra. I napriek zložitejšiemu výpočtu často MFCC a PLP poskytujú podobné výsledky.

Pomocné koeficienty – dynamika a energia

Reč je úzko spätá s plynutím času, preto je veľmi dôležité v základných akustických parametroch obsiahnuť aj ich vývoj v čase. Na tento účel sa používajú dynamické koeficienty, ktoré aproximujú prvú a druhú deriváciu príznakov v čase. Ďalším parametrom je energia signálu. Samotná energia však prirodzene nenesie dostatok akusticky diskriminačnej informácie, preto sa používa len zriedka, čo ale neplatí o jej dynamických vlastnostiach v čase, ktoré v hrubý črtách opisujú plynutie reči.



Na záver by sa malo poznamenať, že úloha extrakcie príznakov pre svoju komplexnosť nie je do teraz celkom vyriešená a je stále vynakladané množstvo úsilia v tejto oblasti. Existuje veľa príznakov a metód ich výpočtu, ale zatiaľ žiadna úplne nespĺňa všetky požadované vlastnosti.

Techniky slúžiace na rozpoznávanie reči

Po extrahovaní vhodných príznakov reči je každé prehovorenie opísané ako časová postupnosť vektorov príznakov. V ďalšom kroku je preto potrebné porovnať neznámu postupnosť príznakov, ktorá môže zodpovedať buď slovu, fráze alebo aj celej vete so vzorom, o ktorom vieme akému vyhovoreniu patril. Tieto vyhorenia však museli existovať vo fáze tréningu. Podľa toho či rozpoznávanie je založené na rozpoznávaní menších akustických jednotiek ako fonémy, či slabiky alebo väčších ako sú slová či celé frázy, rozlišujeme systémy založené na rozpoznávaní a modelovaní jednotiek menších ako slova, napr. hlásky, celých slov, alebo fráz. Pri celých frázach alebo slovách je možné postihnúť tzv. koartikulačný jav, kedy výslovnosť jednotlivých hlások je ovplyvnená výslovnosťou konkrétneho slova. To však naráža na problém, že všetky slová alebo frázy musia byť v tréningovej databáze obsiahnuté viac krát, čo je pri veľkých slovníkoch nerealizovateľné. Lepšou voľbou je použiť jednotky menšie ako slovo, ktorých je menší počet, tie ale nemusia postihnúť fenomén koartikulácie. Ďalej sa systémy delia podľa toho či rozpoznávajú izolované slová alebo spojitú reč, čo je zložitejší problém, keďže nie je jasné, kde jednotlivé slová začínajú a končia. Špecifický problém porovnania vzorov medzi referenčnou (známou) a neznámou vzorkou reči spočíva v tom, že aj rovnaké vyhovorenia

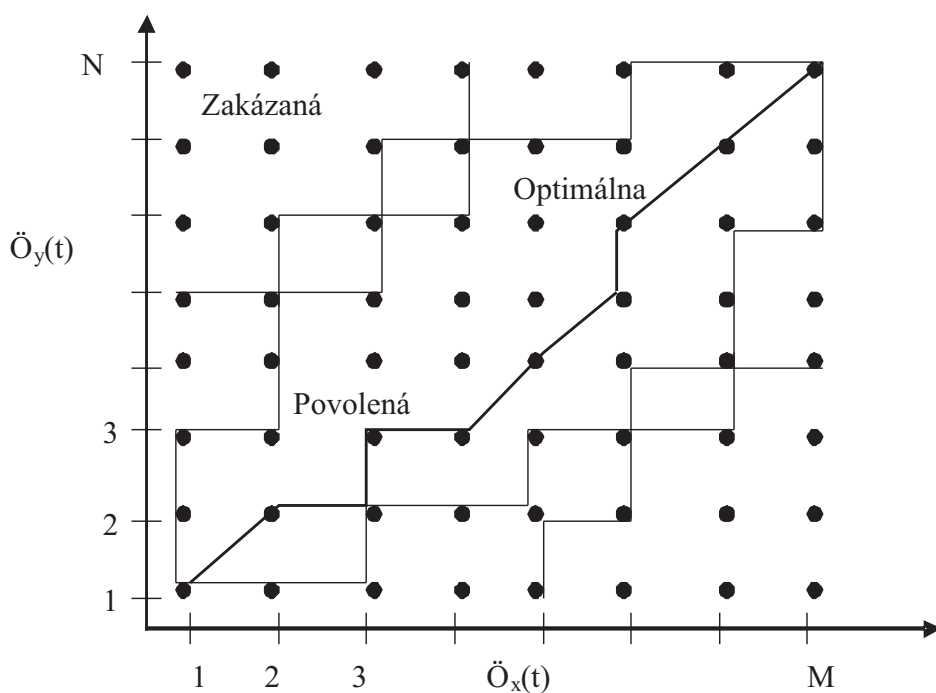
môžu a skoro vždy aj majú rôzne dĺžky (počty vektorov príznakov). Situácia sa komplikuje tým, že dochádza k nerovnomernému predlžovaniu alebo skracovaniu jednotlivých častí (hlások) v rámci slov. Preto nie je vhodné použiť jednoduché princípy lineárneho skrátenia (decimácia) lebo predĺženia (lin. interpolácia) postupnosti vektorov. Existuje viacero metód, ktoré sa pokúšajú riešiť tento problém, ale najúspešnejšie z nich sú DTW a HMM. Preto v ďalšom texte budú obe v jednoduchosti opísané.

DTW

DTW (dynamic time warping) je metóda, ktorá slúži na akustické porovnávanie dvoch vzorov (príznakov reči) referenčného a vstupného slova, pričom sú zohľadnené špecifické vlastnosti reči. Ide hlavne o nelineárnu časovú mierku pri porovnávaní príznakov prislúchajúcich ku vzorovému a testovanému slovu. Táto vlastnosť je nutná na elimináciu vplyvu rôznych dĺžok výpovedí rovnakého slova. Ďalej v sebe zahŕňa riešenie častého prípadu variácií rôznych dĺžok jednotlivých hlások vo vnútri slova, čo sa nedá riešiť jednoduchým lineárnym prečíslovaním (natiahnutím, resp. stlčením) časovej osi referenčného alebo testovaného vyhovorenia.

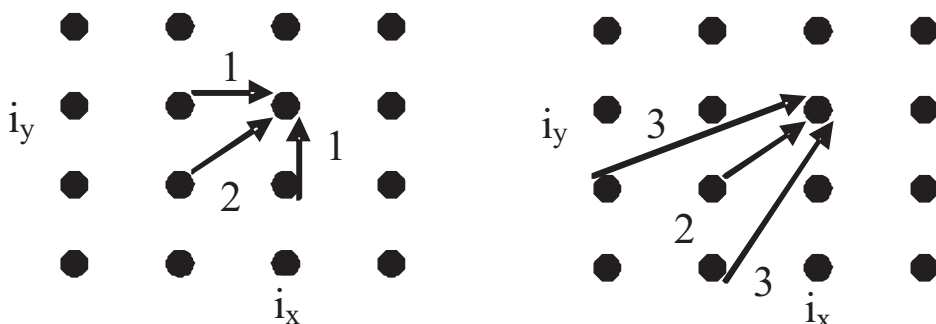
— V základnom použití tejto metódy sa predpokladá poznanie presných hraníc oboch porovnávaných vyhovorení. To si v praxi vyžaduje dodatočné použitie algoritmu detekcie reči.

Jadro metódy vychádza z princípu dynamického programovania riešiaceho problém nájdenia minimálnej cesty medzi dvoma bodmi na stanovený počet krokov. Teda hľadá sa nelineárne mapovanie (ohýbanie času) indexov (vektorov) neznámeho slova na referenčné, aby bola dosiahnutá najmenšia akustická vzdialenosť medzi nimi. Toto mapovanie musí ale spĺňať isté prirodzené požiadavky ako sú: začiatkové a konečné vektory slov sa musia mapovať na seba, funkcia času pre indexy referenčných ako aj testovacích vektorov musí byť rastúca, a musí existovať nejaký rozumný prah maximálnej časovej vzdialenosti medzi mapovaným indexom referenčným a testovacím. Pri samotnom výpočte sa používajú dve matice, jedna tzv. lokálna, kde sú vypočítané akustické vzdialenosti jednotlivých vektorov príznakov reči medzi sebou a druhá globálna, v ktorej sa kumulujú tieto chyby pri hľadaní možných minimálnych ciest s prihliadnutím na rôzne lokálne obmedzenia na pohyb vpred. Ukážka globálnej matice vzdialeností je na obrázku.



Globálna matica vzdialenosti pri DTW s optimálnou cestou, povolenou a zakázanou oblasťou.

Na ďalšom obrázku sú zase pre úplnosť zobrazené niektoré z používaných lokálnych ciest aj s váhovaním.



Dve možné lokálne cesty pre pohyb v globálnej matici aj s váhovaním smerov.



Metóda mala významné postavenie v začiatkoch budovania jednoduchých ASR systémov, ale vzhľadom pre potreby kontinuálneho rozoznávania reči, kde je treba slová radiť jednoduchým spôsobom za sebou, pričom sa ešte vyžaduje aj nezávislosť od hovoriaceho, stratila svoj význam a bola pre väčšie aplikácie úspešne nahradená štatistickou metódou modelovania reči pomocou HMM.

Skryté Markovové Modely (HMM)

Metóda skrytých Markovových modelov je založená na štatistickom modelovaní rečového signálu, respektíve jeho častí, ktoré sa návrhár systému rozhodne použiť, t.j. fonémy, slabiky, slová, frázy atď. Jej veľkou výhodou je jej nezávislosť od

hovoriaceho, lebo model reči je možné trénovať pre rôznych hovoriacich súčasne ako aj ľahké spájanie jednotlivých modelov za sebou, čo vedie až ku spojenému rozpoznávaniu reči. Metóda vychádza z konceptu Markovových reťazcov prvého radu, ktoré matematicky jednoduchým spôsobom opisujú časový sled udalostí pomocou stavov a ich pravdepodobností. Vďaka tomu je výpočtovo veľmi efektívna aj napriek tomu, že úplne neodráža prirodzený časový vývoj bežnej reči.



Markovov reťazec je daný stavmi S_1 až S_N , prechodovou maticou $P_{N \times N}$ pravdepodobností prechodu medzi stavmi a vektorom počiatočných pravdepodobností π , nasledovne:

$$P = \begin{bmatrix} p_{11} & p_{12} & \dots & p_{1N} \\ p_{21} & p_{22} & \dots & p_{2N} \\ \dots & \dots & \dots & \dots \\ p_{N1} & p_{N2} & \dots & p_{NN} \end{bmatrix}, \pi = [P(S_0 = 1), P(S_0 = 2), \dots, P(S_0 = N)]$$



Potom pravdepodobnosť výskytu nejakej postupnosti stavov na nejakom modeli je daná nasledovne:

$$P(S_1, S_2, \dots, S_{N-1}, S_N) = \pi(s_1) p_{12} \dots p_{(N-1)N}$$

Takýto model však neobsahuje žiadne spojenie stavu s vektorom príznakov. Ak sa predpokladá, že stav bude korešpondovať s nejakou stacionárnou časťou reči napr. nejaká samohláska, potom je zrejmé, že minimálne pre viacero hovoriacich bude existovať aj viacero realizácií vektora príznakov. Preto stav v sebe musí ešte obsahovať aj dodatočnú pravdepodobnosť s akou generuje takéto vektory v danom stave, t.j. $P(X/S_i)$. Kombináciu týchto pravdepodobností potom vznikne výsledný model nazývaný skrytý Markovov model. Existujú tri základné možnosti ako modelovať pravdepodobnosť pozorovania v stave. Podľa toho rozlišujeme tri typy HMM:

- Diskrétny HMM
- Spojitý HMM
- Polo-spojité HMM

Diskrétny HMM modely predpokladajú, že vstupná postupnosť vektorov príznakov je konečná čo sa realizuje pomocou vektorovej kvantizácie (VQ). Potom každý vstupný vektor je nahradený niektorým z konečnej množiny vektorov z kódovej knihy. Ak má VQ L vektorov, potom v každom stave je potrebné poznať pravdepodobnosť týchto vektorov, teda len L čísel.



Tento postup je pomerne jednoduchý a rýchly. Jeho ďalšou výhodou je istá robustnosť voči malým skresleniam spôsobených šumom.

- Pri procese VQ sa však dopúšťame trvalej kvantizačnej chyby. Navyše, pre malý počet trénovacích vzoriek a malý rozsah kódovej knihy sa dosahujú pomerne veľké chyby.

Spojité HMM predstavujú v súčasnosti overený štandard, ktorý poskytuje dobré výsledky pre široké spektrum použitia.

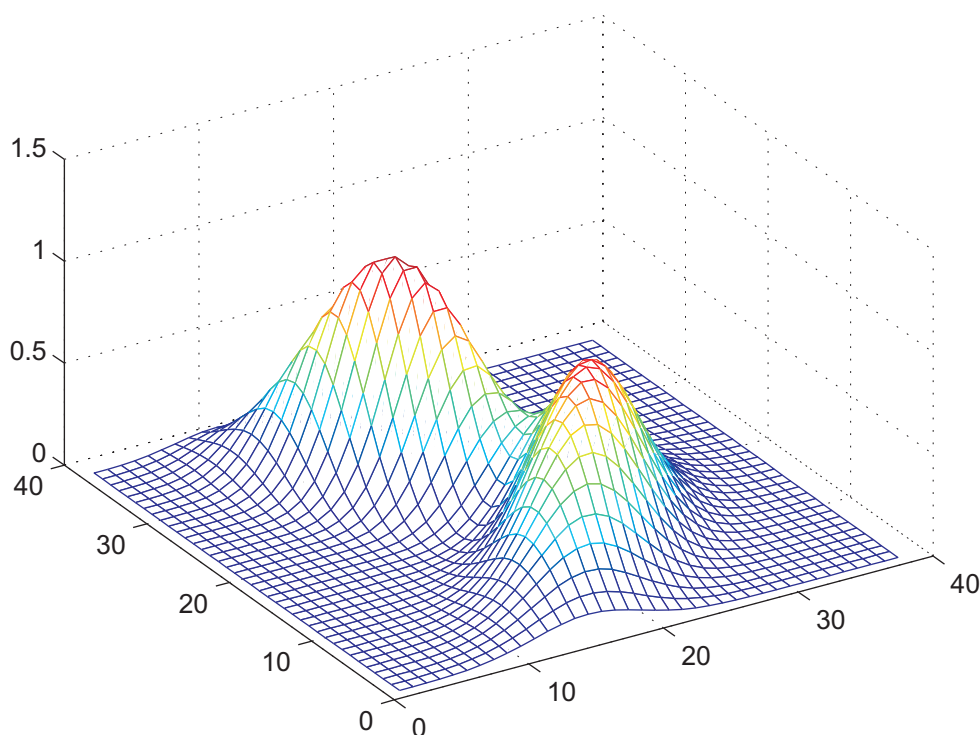


Pravdepodobnosť pozorovanie vektora príznakov v danom stave $P(X/S_i)$ je daná zmesou Gaussových rozdelení nasledovne:

$$f(x/S_i) = \sum_{i=1}^P c_i \frac{1}{\sqrt{(2\pi)^d |U_i|}} e^{-\frac{1}{2}(x-\mu_i)^T U_i^{-1} (x-\mu_i)}$$

Z uvedeného vyplýva, že takýto model je daný vektormi stredných hodnôt (μ), kovariančnými maticami (Σ) a váhami (c_i).

Zmes dvoch Gaussových 2 rozmerných rozdelení je zobrazená na obrázku.

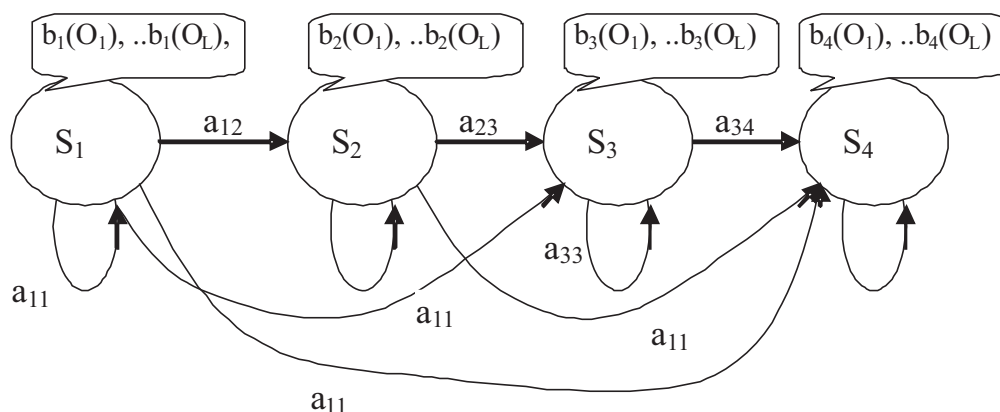


Opis 2D priestoru pomocou dvoch Gaussových zmesí.

- + Takýto model je schopný dobre popísať priestor parametrov aj pri malom počte vzoriek.
- Jeho nevýhodou je ale vyššia výpočtová aj pamäťová náročnosť.

Polo-spojité HMM sa snažia kombinovať výhody oboch. Postup je taký, že vytvorí jednotný opis viacrozmerného priestoru pomocou množiny Gaussových funkcií (môže byť ich aj niekoľko stoviek) na rozdiel od VQ, kde sú oblasti dané len centroidmi. Tieto zmesi sú následne využívané všetkými stavmi spoločne, t.j. nebude pre každý stav vytvorená vlastná množina Gaussových funkcií, čím sa zredukuje množstvo dát a uľahčí sa aj samotný výpočet. Potom v jednotlivých stavoch sa už len pamätá pravdepodobnosť pozorovania vektorov na danej zmesi pre daný stav.

Príklad ľavo pravého 4 stavového diskretného HMM modelu je znázornený na obrázku.



Štvorstavový ľavo pravý diskretný HMM model.



Potom pravdepodobnosť pozorovania reťazca vektorov príznakov o dĺžke T na modeli λ s N stavmi sa počíta rekurzívne pomocou pomocnej premennej α :

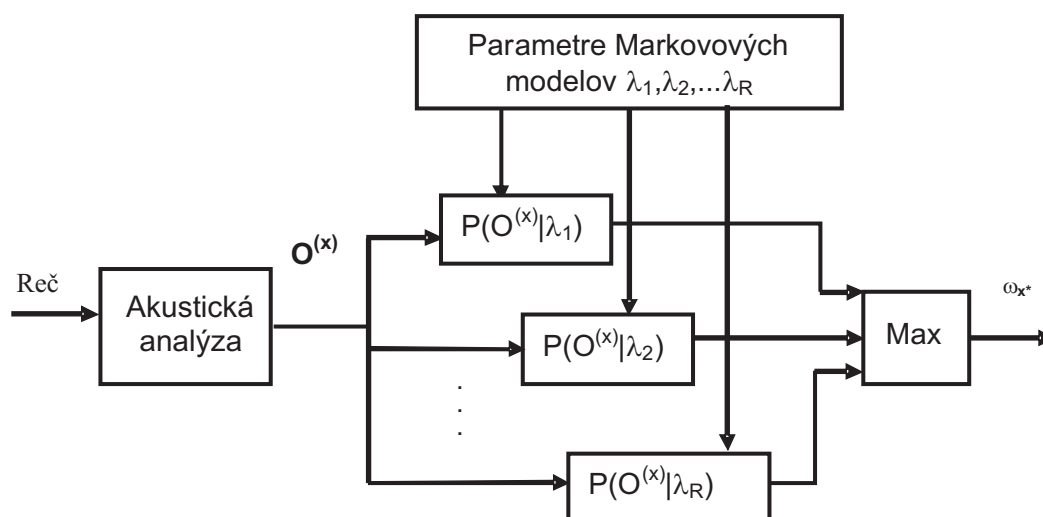
$$\alpha_t(j) = \left[\sum_{i=1}^N \alpha_{t-1}(i) * p_{ij} \right] * P(O_t / S_j),$$

kde $j = 1, \dots, N, t = 2, \dots, T$ a $\alpha_1(j) = \pi(j) * pP(O_1 / S_j)$

Výsledná pravdepodobnosť je daná:

$$P(O / \lambda) = \sum_{i=1}^N \alpha_T(i)$$

Potom rozpoznávanie sa vykoná zvolením HMM modelu (λ) s najvyššou pravdepodobnosťou tak, ako je schematicky znázornené na obrázku.



Princíp rozpoznávania pomocou HMM.



Veľkou výhodou HMM modelov je možnosť zreťazenia viacerých modelov do reťazca, teda vyjadrenie akejkoľvek dĺžky môže byť vybudované použitím súboru základných modelov.



V praktickej situácii sa kvôli rýchlosti a reťazaniu modelov počíta iba cesta s maximálnou pravdepodobnosťou, napr. $P(x_1, S_{t1}, x_2, S_{t2}, \dots, x_T, S_{tT})$, a späťne je určená postupnosť skrytých stavov, resp. modelov použitím Viterbiho algoritmu.

Zložitá otázka je trénovanie takýchto modelov, t.j. nájsť neznáme parametre, aby pre množinu pozorovaní dali maximálnu pravdepodobnosť. Neexistuje analytické riešenie tejto úlohy a preto sa robí iteračne, pričom je zaručené nájdenie len lokálneho maxima. Existuje viacero spôsobov výpočtu, ale najznámejší je tzv. Baum-Welche algoritmus. Okrem trénovacích stratégií je tiež dôležitá štruktúra HMM modelu. Zvyčajne je použitý model zľava doprava pre rečové jednotky. Je všeobecne akceptované, že každý akustický stav je modelovaný tromi stavmi HMM za účelom zachytenia jeho začiatku, stredu a končiacej časti. Na základe teórie klasifikácie sú modely HMM náchylné na nadmerné preučenie, preto sa na detekciu a prevenciu tohto fenoménu musí použiť validačný súbor.



V súčasnosti najmodernejšie postupy využívajú diskriminatívne trénovania (MMI, MCE), alebo modelovanie pozorovaní v stavoch pomocou tzv. Support Vector Machines a Neurónových sietí (Deep belief networks).

6.5 Multimodálne rozhranie

Multimodálne rozhrania sú v súčasnosti veľmi populárne. Každý o nich hovorí – o tom, že sú veľmi prirodzené a obľúbené medzi používateľmi. Multimodálne rozhrania ponúkajú riešenie mnohých problémov s používateľským rozhraním a taktiež umožňujú aplikovať nové služby.

Multimodálne rozhranie predstavuje kombináciu viacerých modalít, čiže spôsobov interakcie s počítačovými systémami. Multimodálne rozhranie slúži na identifikáciu hovoriaceho rozpoznaním tváre a hlasu, teda využíva rozpoznanie a autentifikáciu používateľa pomocou foriem ako hlas, tvár, atď. Okrem toho sa multimodálne rozhranie používa aj na ovládanie *set top boxu* (**STB**) pomocou hlasu alebo gest.

Jedným z príkladov, kde je integrované multimodálne rozhranie, je projekt HBB-Next. Úlohou projektu je napomôcť zlučovaniu vysielania a internetu vyvinutím technológie s ohľadom na koncových používateľov. Projekt má obohatiť klasickú TV o sociálny aspekt – implementovať sociálne siete, možnosť prístupu k filmom z viacerých zariadení (TV, PC, mobil) a odporúčania k filmom. Taktiež má napomôcť mixovaniu vysielaných filmov, doplnkových videí na internete a filmov vytvorených používateľom.

HBB-Next je založený na modulárnej architektúre. Moduly v **HBB** (*Hybrid Broadcast Broadband* – hybridná širokopásmová televízia) sú navrhnuté tak, aby navzájom spolupracovali. Napríklad, ak používateľ vstúpi do miestnosti, systém ho rozpozná a automaticky sa nastaví podľa jeho požiadaviek. Ak potom používateľ otvorí aplikáciu AppStore, systém mu dovolí otvoriť, kúpiť a inštalovať zvolenú aplikáciu bez ďalšej autentifikácie. Každá aktivita alebo operácia používateľa môže vyžadovať niekoľko-násobnú autentifikáciu určitým bezpečnostným stupňom. Tu napríklad vidieť spoluprácu modulu rozpoznania tváre používateľa a modulu, ktorý zabezpečuje autentifikáciu.

Identifikácia viacerých hovoriacich vychádza z nahraného audio signálu, ktorý môže obsahovať vyhovorenia viacerých osôb. Táto úloha sa môže rozdeliť do niekoľkých kategórií podľa viacerých detailov. Jednou kategóriou sú prípady, keď hovoriaci je známy a chceme zistiť, či sa jeho hlas nachádza v nahranom signáli – to je možné vtedy, ak bol súčasťou tréningovej fázy detektora. Potom sa úloha mení na identifikáciu jedného konkrétneho hovoriaceho. V prípade, že žiaden hovoriaci nie je známy, musia sa použiť techniky klastrovania a segmentácie. Cieľom väčšiny aplikácií tohto druhu je to, aby nepretržite fungovali: „počúvali“ prichádzajúci zvuk zakódovaný v PCM vzorkách, detegovali prítomnosť hlasu (*voice activity detection*, **VAD**), dokázali rozpoznať ticho a šum v pozadí a rozpoznali prekrývanie rečníkov. Ak aplikácie zaznamenávajú dlhší hlasový záznam, musia identifikovať hovoriaceho s určitou presnosťou. Pri identifikácii jedného hovoriaceho je pritom cieľom identifikovať ho bez ohľadu na významové hľadisko vety.

Každý systém na identifikáciu hovoriaceho pozostáva z dvoch hlavných častí. Prvá je rozpoznanie rečových parametrov z nahraného audio signálu, druhá je klasifikačná metóda, ktorá na základe už vybraných parametrov s určitou pravdepodobnosťou určí hovoriaceho. Tieto systémy sú väčšinou navrhované na

konkrétny účel a zatiaľ nie sú univerzálne - navrhovateľ aplikácie musí na základe typu úlohy vybrať správnu kombináciu metód a ich modifikácií.

V prípade rozpoznania rečových príkazov sa implementuje systém na rozpoznanie izolovaných slov. Najúspešnejšie a najpoužívanejšie sú tie, ktoré pracujú na princípe HMM, štatistického modelovania reči, obzvlášť tie, ktoré používajú ako základnú jednotku pre modelovanie kontextovo závislé fonémy. Ak sa používa pevne určená skupina príkazov, potom sa môžu na modelovanie použiť celé slová, a tak dosiahnuť potenciálne lepšia presnosť (lepšie zachytený koartikulačný efekt).

Vo všeobecnosti sú dve kategórie postupov sledovania gest používateľa: vzhľad a postup založený na 3D modeli. Postup založený na 3D modeli porovnáva vstupné parametre končatiny v 2D projekcii 3D modelu končatiny. Postup založený na vzhľade používa obrazové vlastnosti na modelovanie vizuálneho vzhľadu končatiny a porovnáva s vybranými obrazovými vlastnosťami z video vstupu. Pri druhom postupe výsledok závisí od schopností výpočtového zariadenia. Ak sa používa RGB kamera, metódy sa zameriavajú na sledovanie farby kože alebo tvaru gestikulujúcej časti tela. Avšak postup vysoko závisí od svetelných podmienok rovnako ako od stability popredia a pozadia sledovaného subjektu. Taktiež sa vo vyhodnocovanej oblasti nemôžu objaviť iné objekty farby kože alebo objekty v tvare končatiny, pretože môžu zmiast' algoritmus. Infračervená hĺbková kamera používa svoj vlastný zdroj IR svetla, a je preto oveľa odolnejšia voči svetelným podmienkam scény. Navyše je kamera schopná poskytnúť hĺbkovú mapu, pseudo 3D obraz scény, čo môže byť veľmi užitočné pri sledovaní gestikulujúcich častí tela, napr. ruky.

V súčasnosti je niekoľko metód, ako sa dá implementovať ovládanie a kontrola očami. Najdôležitejším bodom je správna detekcia očnej zreničky. V HBB-Next projekte bola použitá najjednoduchšia a najprirodzenejšia metóda, ktorá využíva RGB kameru zariadenia Kinect. Princíp je nasledujúci: osoba sediaci pred monitorom bude bez pohybu pozeráť na vyznačené body na obrazovke. Aplikácia odmeria vzdialenosť hlavy od monitora. Pre potreby ďalších výpočtov je potrebné zadať aj veľkosť monitora a následne sa na základe Pytagorovej vety dajú vypočítať parametre pohybu zreničiek – ich uhol a pohyb. Aplikácia určí maximálny rozsah pohybu zreničky tak, že nasmeruje používateľa, aby sa pozrel na okraje monitora, a vypočítajú sa aj uhly odklonenej zreničky od základnej pozície. Pri známej vzdialenosti hlavy od monitora vieme teda prepočítať rozptyl a uhly pri zmene polohy zreničky tak, aby bolo stále zabezpečené ovládanie.

Techniky pre rozpoznanie tváre boli podrobne popísané v kapitole Rozpoznanie obrazu. V reálnom systéme je definovaný zoznam požiadaviek podľa toho, čo všetko systém musí/mal by/môže:

- Systém musí identifikovať používateľa v miestnosti na základe jeho tváre, ak patrí do skupiny lokálnych používateľov.
- Systém musí rozpoznať používateľa a porovnať jeho profil s uloženými profilmi, bez prístupu na internet.

- Systém môže identifikovať používateľa v miestnosti na základe jeho tváre, pričom používateľ je neznámy (nepatrí medzi lokálnych používateľov).
- Systém môže rozpoznať používateľa aj v tmavom prostredí.

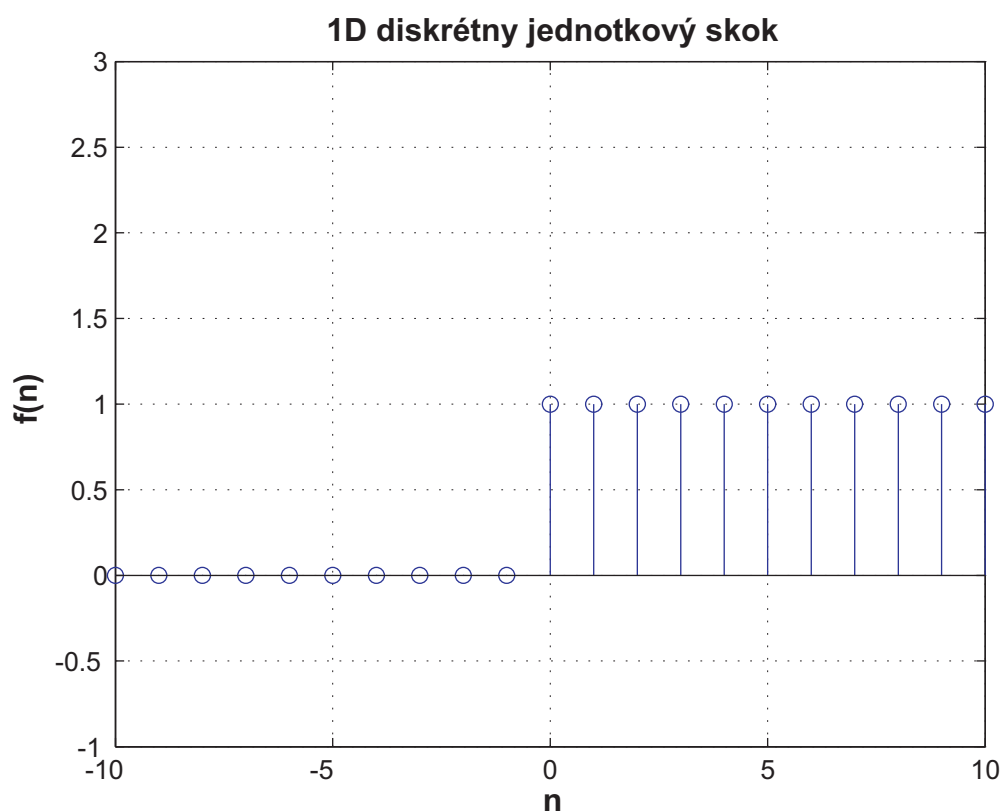
7 Příklady v Matlabe

7.1 Příklady v Matlabe



Příklad 1

Napište kód v Matlabe, který vykreslí jednorozměrný, diskrétní jednotkový skok.



Obr. 7.1: Výsledek: Diskrétní jednotkový skok

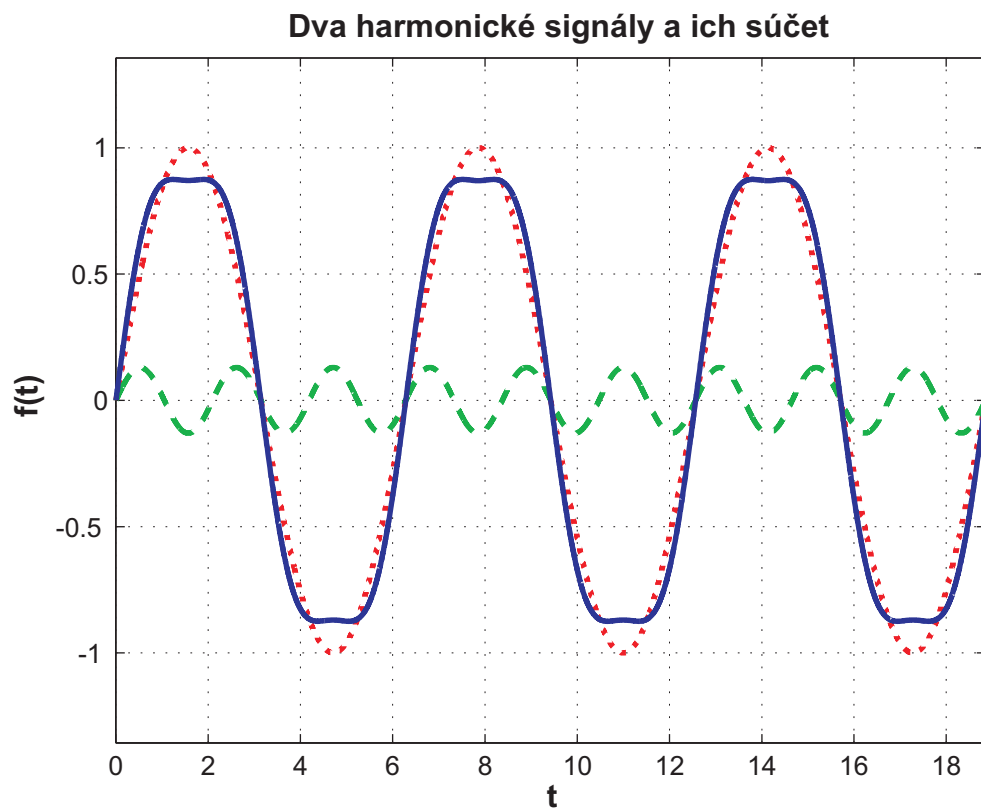
```
t=-10:20;      %definícia časovej periódy signálu
step=heaviside(t); %heaviside(x) je
                funckia v Matlabe,
                ktorej hodnoty sú 0
                pre x < 0, 1 pre x > 0,
                a 0.5 pre x = 0.

step(t==0)=1;
figure;        %príkaz na vykreslenie
                obrázku
stem(t,step);
grid on;
xlabel('n');
ylabel('f(n)');
title('1D diskretný jednotkový skok');
axis([-10 10 -1 3]);
```



Príklad 2

Napište kód v Matlabe, ktorý generuje sínusový signál.



Obr. 7.2: Výsledok: Dva harmonické signály a ich súčet

```

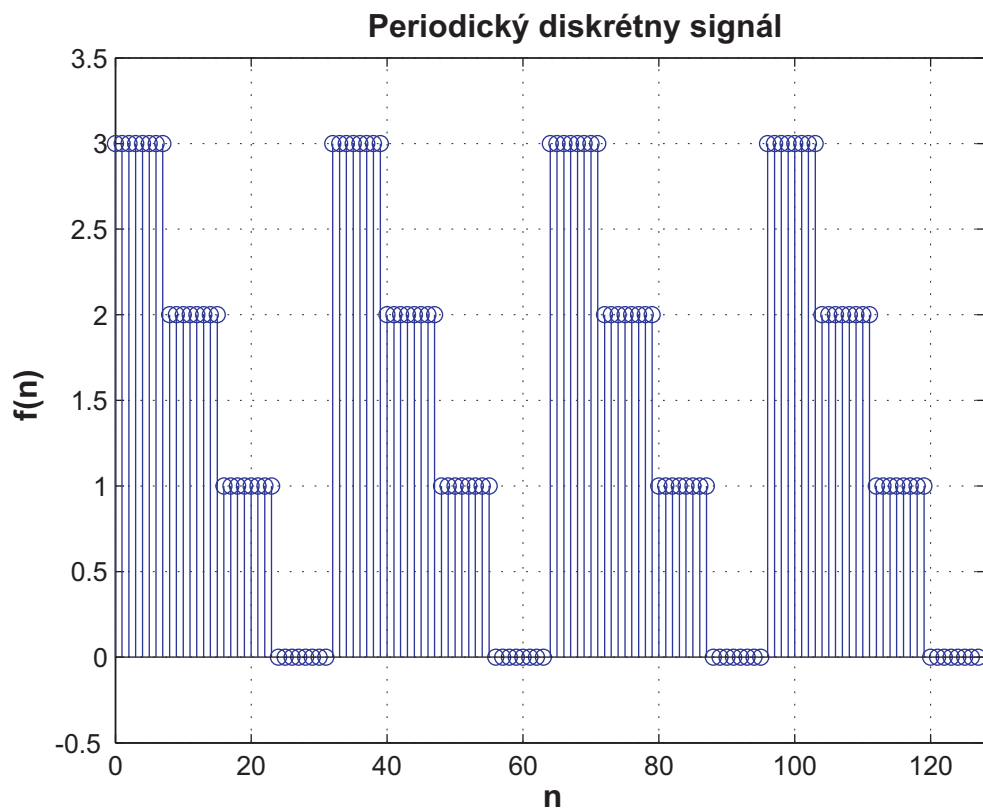
range=6*pi;           %časový rozsah signálu
t=0:0.001:range;      %časové body
A=[1 0.13];          %vektor amplitúd
w=[1 3];              %vektor frekvencií [Hz]
phi=[0 0];            %vektor fáz
sig1=A(1)*sin(w(1)*t+phi(1)); % definovanie
                        jednotlivých
                        signálov
sig2=A(2)*sin(w(2)*t+phi(2));
signal=sig1+sig2;
figure;
plot(t,sig1,':r','LineWidth',2);
hold on;
plot(t,sig2,'--g','LineWidth',2);
hold on;
plot(t,signal,'LineWidth',2);
grid on;
axis([0 rozsah -1.2*sum(A) 1.2*sum(A)]);
xlabel('t [s]');
ylabel('f(t)');
title(Harmonické signály a ich súčet);

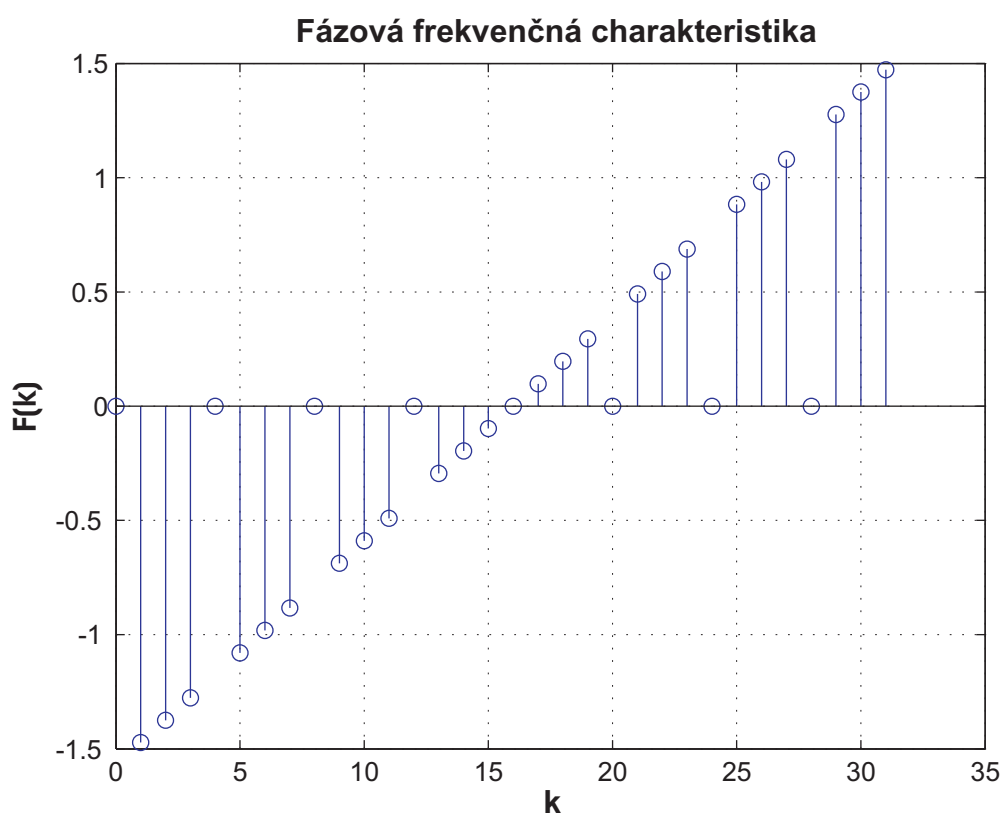
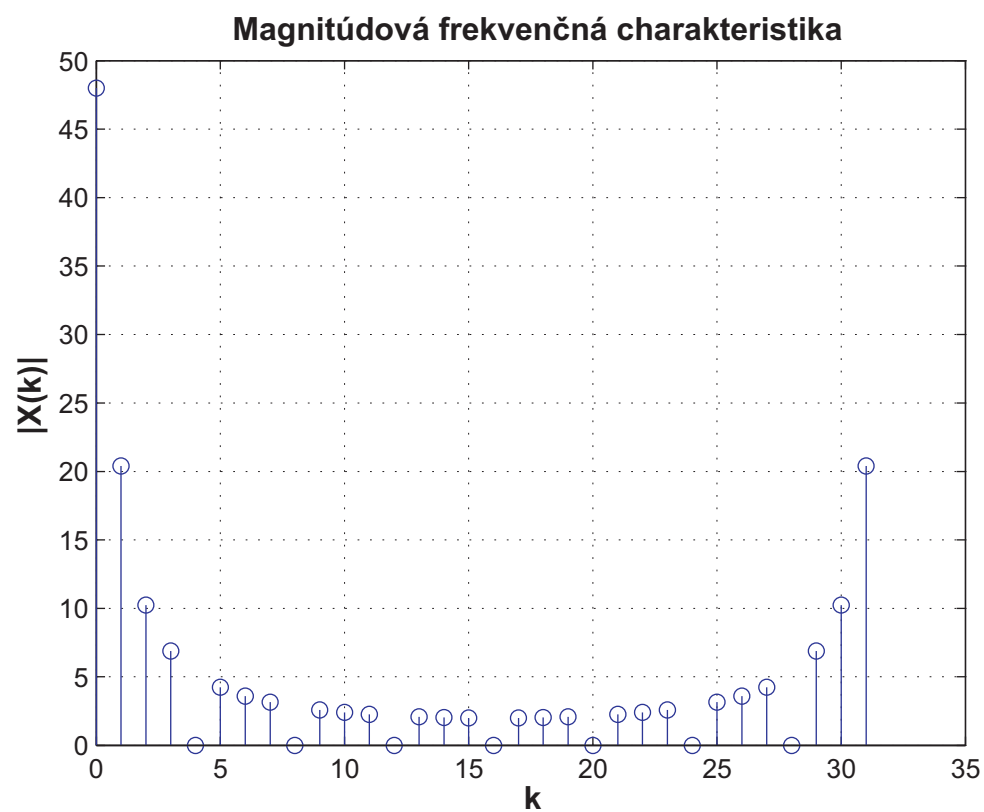
```



Príklad 3

Napište kód v Matlabe pre určenie diskretnéj Fourierovej transformácie a frekvenčných charakteristík.





Obr. 7.3: Príklad prenosovej funkcie LDKI systému, magnitúdová a fázová frekvenčná charakteristika

RIEŠENIE

```
count=32;
Ts=4/count;      %vzorkovacia frekvencia
per=4;           %počet periód
syms k;          %symbolické premenné
syms n;
signal=[3.*ones(1,count/4) 2.*ones(1, count /4)
        ones(1, count /4) zeros(1, count /4)];
fn=[];
for n=1:per
    fn=[fn signal];
end
n=0:count*per-1;

figure;
stem(n,fn);
title('Diskrétny signál');
axis([0 length(fn) min(abs(fn))-0.5
      max(abs(fn))+0.5]);
grid on;

figure;
Xk=fft(signal);      %diskrétna Fourierova
                    transformácia
os=0:length(Xk)-1;
stem(os,abs(Xk));    %magnitúdová frekvenčná
                    charakteristika
title('Magnitúdová frekvenčná charakteristika');
grid on;

figure;
stem(os,angle(Xk));  %fázová frekvenčná
                    charakteristika
title('Fázová frekvenčná charakteristika');
grid on;
```
