

Základná zložka multimediálneho signalu – zvuk

- zvuk = vibrácia šírená prostredím (zvyčajne vzduchom)
 - pozdĺžne vlnenie v stlačiteľnom médiu
 - striedanie sa stlačenia a roztiahnutia
- základné vlastnosti: výška, farba, intenzita (hlasitosť)
 - výška (frekvencia) zvuku = **základná (dominantná) frekvencia** frekvencia pri rozložení pomocou pomocou FT, STFT **resp. subjektívne dominantná frekvencia**. Človek vníma rozsah 20Hz-20kHz, mimo je infrazvuk a ultrazvuk.
 - farba = obsah **násobkov** k základnej frekvencii - **aliquótne tóny**. Celočíselné násobky = **harmonické tóny**, neceločíselné = **čiastkové tóny**. Pravouhlý, pílovitý, sínusový signál s tou istou periódou majú sú istú základnú frekvenciu, ale iné vyššie harmonické zložky (inú farbu)
 - hlasitosť zvuku (viď pokračovanie)
 - Rýchlosť šírenia zvuku (v plynach závisí od tlaku p_c (pozitívne) a hustoty plynu ρ (negatívne))

Základné vlastnosti (pokačovanie)

○ Hlasitosť zvuku

- Akustický tlak p (efektívna hodnota zmien tlaku v prostredí (typicky atmosferického tlaku))
- Hladina akustického tlaku (SPL Sound Pressure Level) je definovaná ako
 - $L_p = 20 \log \left(\frac{p}{p_0} \right) [dB]$, kde
 - **prah počuteľnosti** $p_0 = 20 \mu Pa$. ($L_p = 0 dB$)
- **Prah bolesti** $\Rightarrow p_1 = 20 - 63 Pa$ ($L_p = 120-130 dB$), môže poškodiť ucho
 - atmosferický tlak pri hladine mora je cca 101325 Pa, t.j. 5000x väčší ako 20Pa, t.j. zmeny sú menšie než 0.2 promile – t.j. **zanedbateľné**)
- Hladina intenzity zvuku (miera **akustického výkonu**)
 - $L = 10 \log \left(\frac{I}{I_0} \right) [dB]$
 - kde $I_0 = 10^{-12} W m^{-2}$ (**akustický výkon** na jednotku plochy=**Intenzita zvuku** odpovedajúca prahu počuteľnosti)

Psychoakustika

- Ako človek vníma zvuky
 - závislé od fyziológie ucha
- Vnímanie hlasitosti
 - Ľudské ucho je najcitlivšie pri 3-6kHz
 - izofóny (ISO 226:2003)-krivky pre rovnaké vnímanie hlasitosti v závislosti od frekvencie a L_p
- Maskovanie zvukov
 - frekvenčné maskovanie
 - časové maskovanie (dopredné (hlasný-> tichý) do 50ms, spätné (tichý->hlasný) do 10ms)
- Vnímanie frekvencie
 - temperované ladenie = tón o oktávu vyššie má dvojnásobnú frekvenciu
 - oktáva sa delí na 12 poltónov
 - tón komorné A má 440Hz
 - poltóny sú od seba vzdialené 100 centov
 - cvičené ucho rozozná rozdiel 5 centov = 1/20 poltónu
 - pri súčasnom počúvaní, vďaka „záznejom“ je rozlíšenie oveľa väčšie
- vnímanie fázy
 - aj na jej základe je schopný človek lokalizovať smer

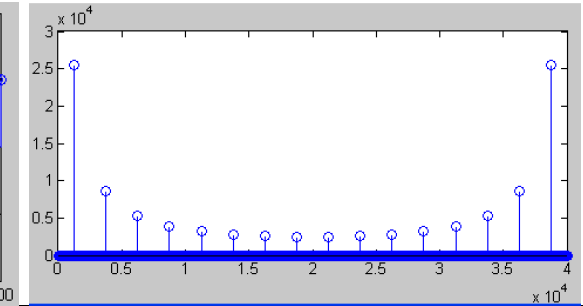
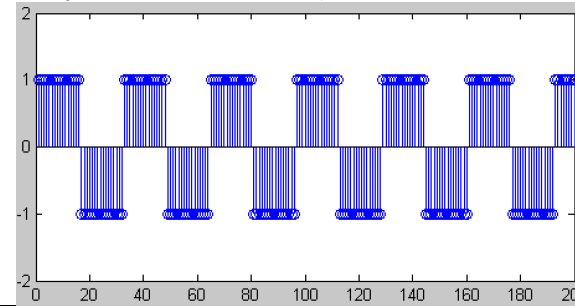
Zvuky

- čistý tón = sinusoida
 - zafarbený tón (obsahuje vyššie harmonické)
- šumy
 - biely šum (autokorelácia je jednotkový impulz) – má konštantnú spektrálnu hustotu energie
 - ružový šum – logaritmický pokles energie
 - sivý šum – má psychoakusticky konštantne vnímanú hlasitosť šumu v celom spektre
- zvuky hudobných nástrojov
 - pozor v spektre môžu byť význačné zložky aj pod základnou frekvenciou tónu
 - spôsobené obálkou (nábeh, útlm, držanie, doznenie) tónu
 - často neobsahujú len vyššie harmonické, ale aj veľa šumu
- ľudský hlas
 - okrem základného tónu obsahuje veľa alikvótnych tónov
 - charakteristické frekvenčné zložky = formanty

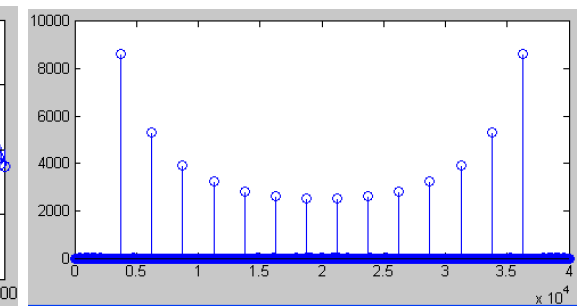
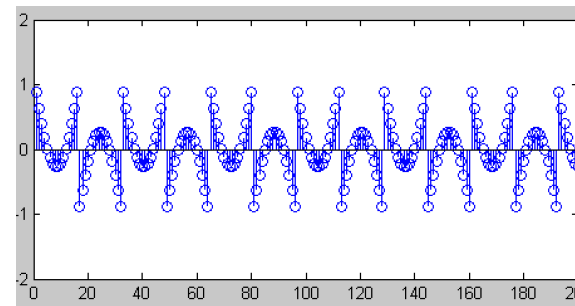
Základná frekvencia a jej výpadok – experiment1

```
clear all; close all;
Fs=8000; N=5; F1=250;
mysig=[];
myperiod=[ones(1,Fs/F1/2) -ones(1,Fs/F1/2)];
for idx=1:(F1*N)
    mysig=[mysig myperiod];
end
stem(mysig);
ylim([-2,2]); xlim([0,200]);
figure;
mySp=fft(mysig);
stem(abs(mySp));
display("hram orig. signal, stlac klavesu");
sound(mysig, Fs); pause;
figure;
sigLen=Fs*N;
z1=F1*N;
mySp(z1+1)=0;
mySp(sigLen-z1+1)=0;
stem(abs(mySp));
figure;
mysigA=real(ifft(mySp));
stem(mysigA);
ylim([-2,2]); xlim([0,200]);
display("hram filt. signal");
sound(mysigA, Fs);
```

Originálny signal v čase a jeho spektrum
(pravouhlý signal s frekvenciou 250Hz, t.j. perióda má 32 vzoriek
 $\text{pri}f_{vz} = 8000\text{Hz}$)

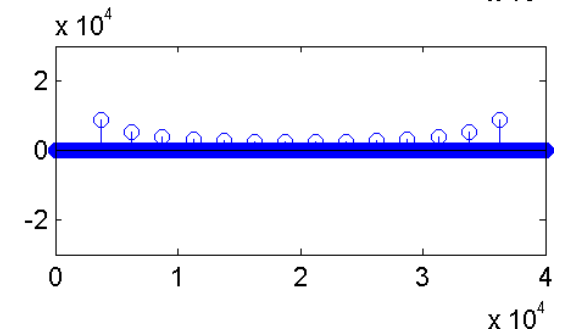
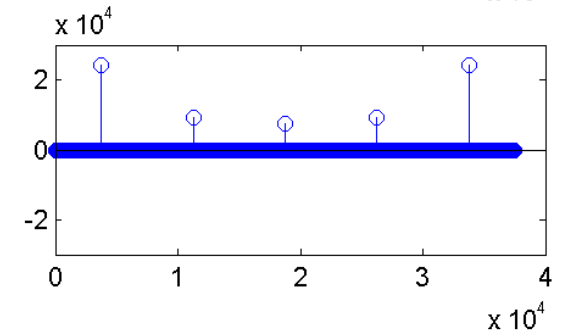
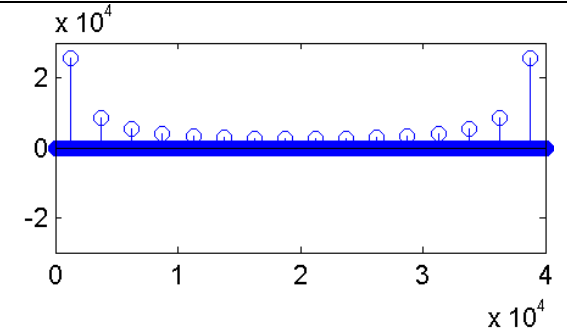
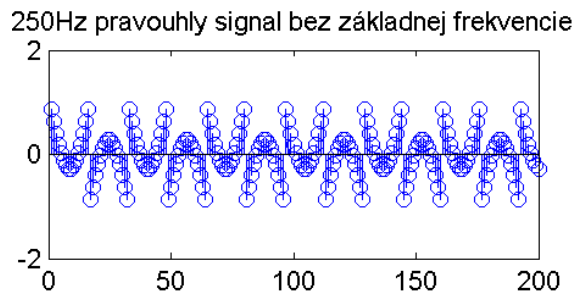
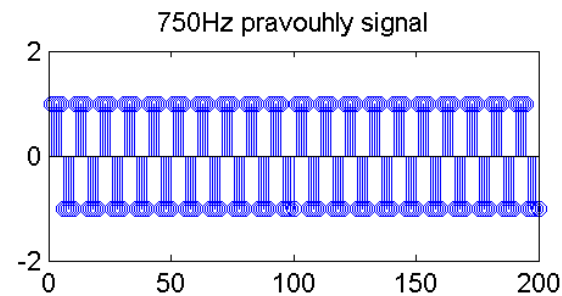
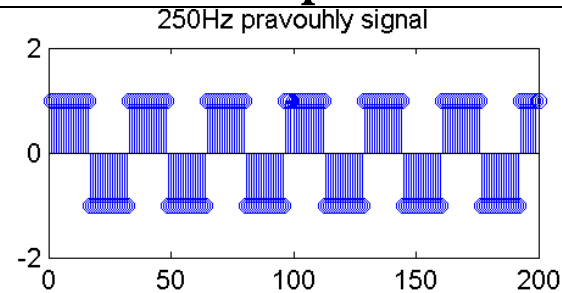


Odfiltrovaná základná frekvencia 250Hz



Základná frekvencia a jej výpadok – experiment2

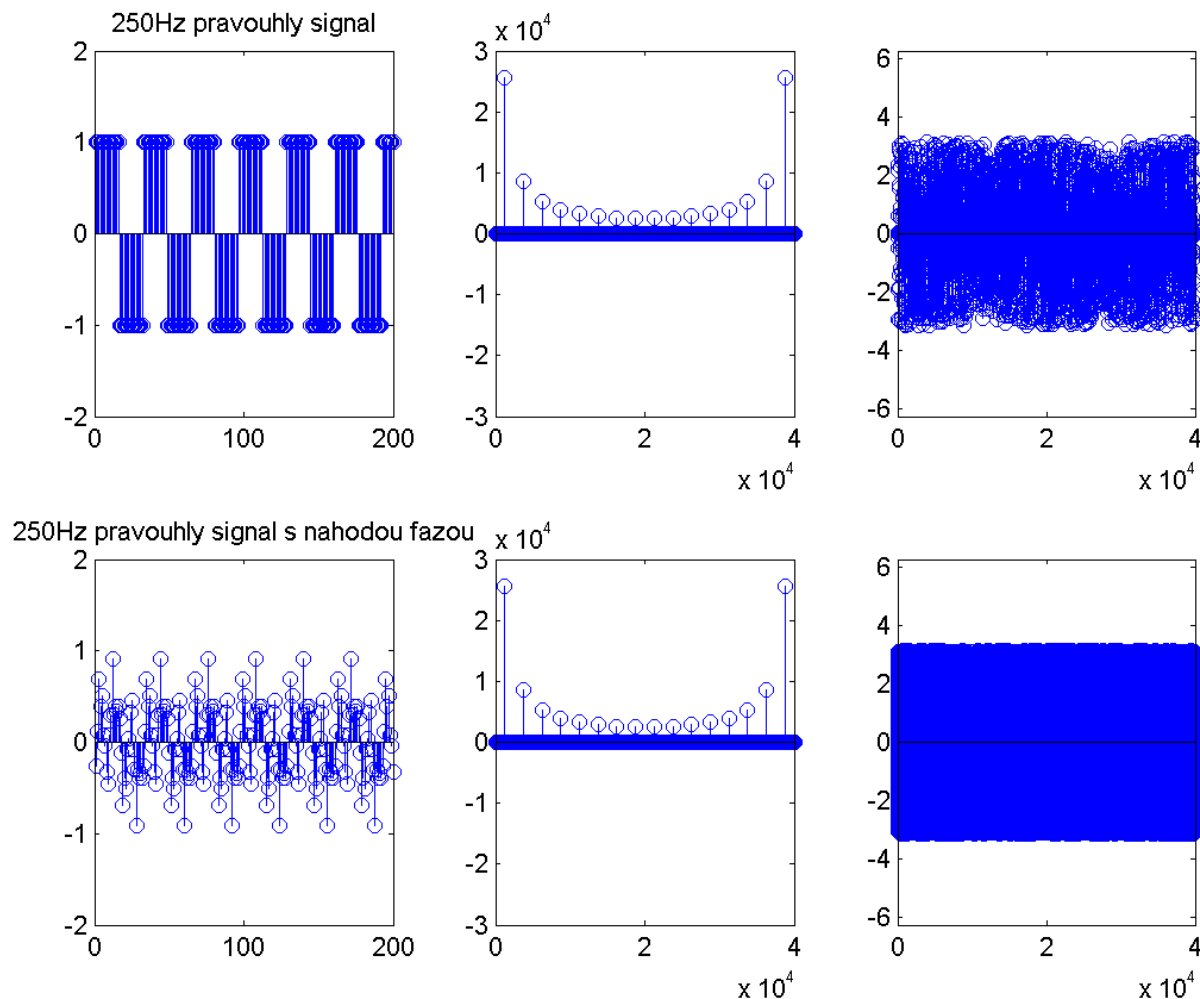
```
clear all; close all;
Fs=8000; N=5; F1=250; F2=750;
sigLen=Fs*N;
subplot(3,2,1);
spY=30000;
mysig=[];
myperiod=[ones(1,Fs/F1/2) -ones(1,Fs/F1/2)];
for idx=1:(F1*N); mysig=[mysig myperiod]; end
stem(mysig);
ylim([-2,2]); xlim([0,200]);
title(sprintf('%dHz pravouhly signal',F1));
subplot(3,2,2);
mySp=fft(mysig);
stem(abs(mySp));
xlim([0,sigLen]); ylim([-spY,spY]);
subplot(3,2,3);
mysig2=[];
myperiod2=[ones(1,round(Fs/F2/2)) -ones(1,round(Fs/F2/2))];
for idx=1:(F2*N); mysig2=[mysig2 myperiod2]; end
stem(mysig2);
ylim([-2,2]); xlim([0,200]);
title(sprintf('%dHz pravouhly signal',F2));
subplot(3,2,4);
mySp2=fft(mysig2);
stem(abs(mySp2));
xlim([0,sigLen]); ylim([-spY,spY]);
subplot(3,2,5);
sigLen=Fs*N; z1=F1*N; mySp(z1+1)=0;
mySp(sigLen-z1+1)=0;
stem(abs(mySp));
xlim([0,sigLen]); ylim([-spY,spY]);
subplot(3,2,6);
mysigA=real(ifft(mySp));
stem(mysigA);
ylim([-2,2]);
xlim([0,200]);
title(sprintf('%dHz pravouhly signal bez základnej frekvencie',F1));
print(gcf,'missingBase.png','-dpng');
sound(mysig, Fs); display("hram 250Hz signal, stlac klavesu"); pause;
sound(mysig2, Fs); display("hram 750Hz signal, stlac klavesu"); pause;
sound(mysigA, Fs); display("hram filt. 250Hz signal, stlac klavesu"); pause;
```



1. Posledný (dolný signál) je jednoznačne na počutie hlbší ako stredný signál !
Prečo? Ucho „nepočúva“ len harmonické frekvencie ale vníma aj periodické deje!
T.j. ucho počuje aj „obálku“/periódu/“záznej“ výsledného signálu.
2. Vidíme, že vieme vytvoriť dej periodický s periódou T bez harmonickej frekvencie periodickej s periódou T . Sú to dve rozličné veci. Ucho to počuje.
3. Napr. ak máme frekvencie s periódou $2T$ a $3T$, výsledná perióda je $6T$, t.j. frekvencia výsledného „deja“ je nižšia ako frekvencie jednotlivých zložiek.

Aký má vplyv rozfázovanie zložiek signálu? Počúť to?

```
clear all; close all;
Fs=8000; F1=250; N=5;
sigLen=Fs*N;
spY=30000;
subplot(2,3,1);
mysig=[];
myperiod=[ones(1,Fs/F1/2) -
ones(1,Fs/F1/2)];
for idx=1:(F1*N); mysig=[mysig
myperiod]; end
stem(mysig);
ylim([-2,2]); xlim([0,200]);
title(sprintf('%dHz pravouhly
signal',F1));
subplot(2,3,2);
mySp=fft(mysig); myAbs=abs(mySp);
stem (myAbs); xlim([0,sigLen]); ylim([-
spY,spY]);
subplot(2,3,3);
stem (angle(mySp)); xlim([0,sigLen]);
ylim([-2*pi,2*pi]);
subplot(2,3,5);
stem (myAbs); xlim([0,sigLen]); ylim([-
spY,spY]);
subplot(2,3,6)
randphase = 2*pi.*rand(1,sigLen)-pi;
stem(randphase); xlim([0,sigLen]);
ylim([-2*pi,2*pi]);
subplot(2,3,4);
mySp2=myAbs.*exp(i*randphase);
mysigA=real(ifft(mySp2));
stem(mysigA); ylim([-2,2]);
xlim([0,200]);
title(sprintf('%dHz pravouhly signal s
nahodou fazou',F1));
print(gcf,'randomPhase.png','-dpng');
sound(mysig, Fs); display("hram orig.
250Hz signal, stlac klavesu"); pause;
sound(mysigA, Fs); display("hram 250Hz
signal s nahodnou fazou");
```

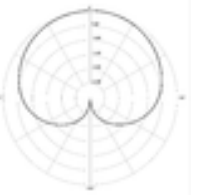
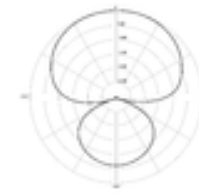
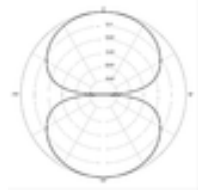
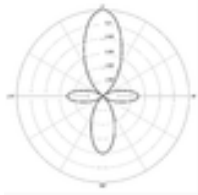


Pravouhlému signálu (horný riadok) sme úplne náhodne „prerobili“ fázy všetkých spektrálnych koeficientov. Počúť iba veľmi slabú zmenu, hoci v čase je zmena úplne drastická (hodi vidieť, že signál si zachoval periódu).

Záznam zvuku

- Mikrofón

- parametre mikrofónov: typ, frekvenčný rozsah, smerová charakteristika, citlivosť, šum, impedancia
 - Citlivosť: meria sa pri 1kHz signále so SPL=94dB (t.j. $p_1=1\text{Pa}$) – koľko voltov bude na výstupe mikrofónu (udáva sa buď v mV, alebo v dBV/Pa, napr. -54.5dBV/Pa=1.85mV)
 - Delenie Podľa konštrukcie: kapacitné (spravidla kvalitnejšie), dynamické
 - Podľa impedancie: vysokoimpedančné, nízkoimpedančné
 - Podľa smerovosti: všesmerové, smerové (smerovosť je závislá od frekvencie):

Všesmerová	subsrdcová	srdcová	supersrdcová	hypersrdcová	Osmičková	Smerová
						

- Skupina mikrofónov (stereo ...)

- Pri stereofónnom zázname sa spravidla používa jeden mikrofón, najčastejšie s dvomi snímačmi otočenými naľavo napravo (systém XY)

- Mikrofónové pole

- pri lepšie zachytenie hovoriaceho a odfiltrovanie rušivých vplyvov
- experimentálne až 4096 mikrofónov

Reprodukcia

○ Slúchadlá

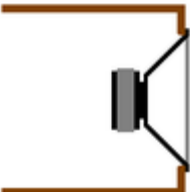
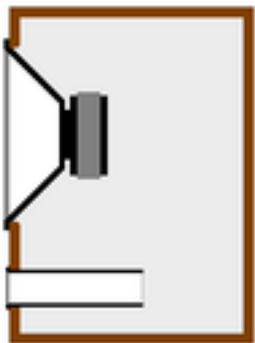
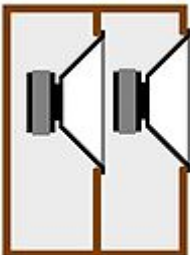
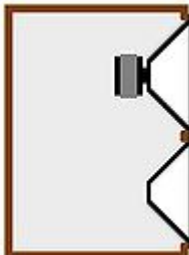
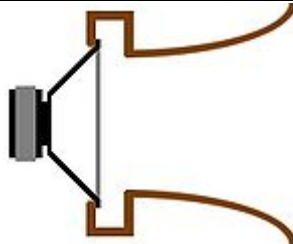
- zatvorená konštrukcia (väčšia únava ucha -nie je cirkulácia vzduchu vo zvukovode)
- otvorená konštrukcia (vyššia priepustnosť hluku zvonku)

○ Repoduktory – elektroakustický menič

- Parametre: typ, veľkosť, výkon, impedancia, citlivosť ...
- Výkon: iba cca 1% elektrického výkonu sa premení na akustický!
 - RMS (root mean square) Max efektívny výkon meraný pri 1kHz sínusovom signále a 10% skreslení
 - DIN výkon: Max efektívny výkon meraný pri pri 1kHz sínusovom signále a 1% skreslení
 - MPO (hudobný špičkový výkon) – spravidla dvojnásobný ako RMS
 - P.M.P.O. – krátkodobý maximálny výkon (nereálny výkon)
- Skreslenie (závislé od aktuálneho výkonu)
 - THD= Total Harmonic distortion
 - Meria čistotu harmonického signálu a sleduje výskyt jeho vyšších harmonických

$$\circ THD_R = \frac{\sqrt{V_2^2 + V_3^2 + V_4^2 + \dots}}{\sqrt{V_1^2 + V_2^2 + V_3^2 + \dots}} [\%]$$

- Citlivosť: je v dB a charakterizuje, aký akustický tlak reproduktor vytvori 1m od neho ak má príkon 1 Watt pri sínusovom signále (napr. $2.83V_{RMS}$ na 8Ω reproduktore)
 - Domáce reproduktory: citlivosť 86-95dB, **efektivita 0.5-4%**
 - Na koncertoch: citlivosť 103-110dB, **efektivita 10-20%**
- reproduktor (podľa rozsahu prenášaného pásma kmitočtov): širokopásmový (20Hz-20kHz), subwoofer , nízkopásmový (<1.5kHz), strednopásmový (300Hz-6kHz), vysokopásmový (>2kHz, smerové), kombinovaný, špeciálny (siréna/alarm)
- Reproduktorová sústava = reproduktory+ozvučnica+pasívne/aktívne el. obvody
 - ozvučnica

Otvorená	Uzatvorená	Basreflexová	Izobarická	Pasívny žiarič	Zvukovod	Prenosová cesta
						

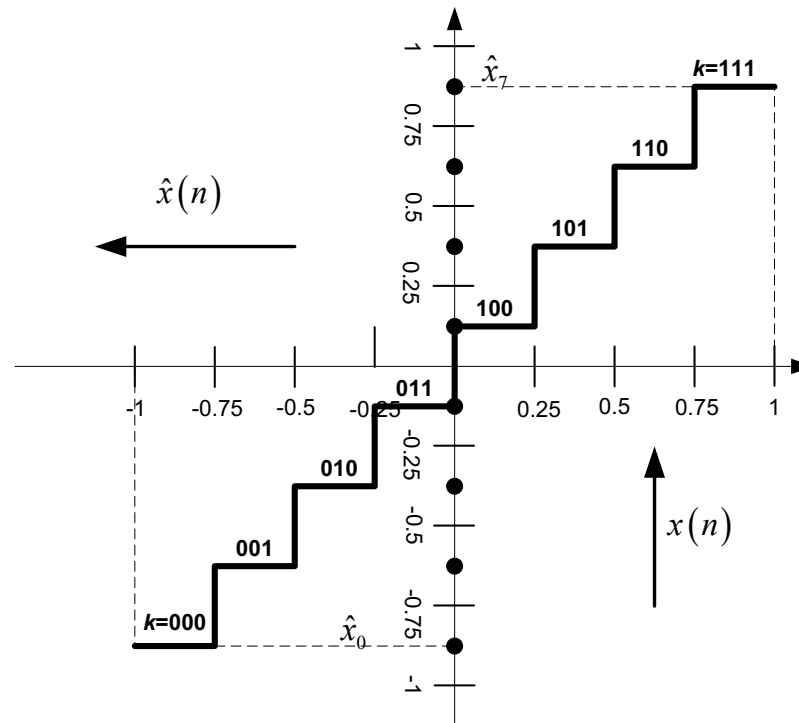
Analógové spracovanie audia

- obmedzenie rozsahu (orezanie)
- odfiltrovanie 50hz
- vyrovnanie hlasitosti (na základe merania priemernej hlasitosti), americké normy majú 300 ms merací interval, britské 10ms, ...

A/D, D/A konverzia audia

- problematika zopakovaná na prednáške 2
- používané vzorkovacie frekvencie
 - 8kHz – telefónia
 - 44100 - Audio CD
 - 48kHz – DVD audio
 - 88.2kHz (používa sa v štúdiách, keď cieľom je audio CD)
 - 96kHz – Blue ray disky, HD DVD
 - 192kHz – Profi level pre 48kHz a 96kHz ciele
 - 2822400Hz = 64 násobok 44,1kHz (t.j. oversampling s faktorom 64) – používané v DSD (direct stream digital audiu), 1 bitové vzorky používajúce sigma-delta moduláciu (viď nižšie), nie PCM, pozri prípadne aj [\[zvuk\]](#)

Príklad: lineárny skalárny kvantizátor s $K=8$, $R=3$, $x(n) \in (-1,1)$. Guličkami sú označené rekonštrukčné hodnoty $\hat{x}(k)$



Dá sa odvodiť, že priemerný výkon chybového signálu kvantizačnej chyby $e(n)$ je: $P_e = \frac{\Delta^2}{12}$

Pridaním ďalšieho bitu znížime Δ na polovicu, t.j. P_e klesne na $\frac{1}{4}$. To znamená zmenu o $10 \log_{10} \frac{1}{4} \approx -6dB$

T.j. : Pridaním 1 bitu rozlíšenia sa zvýši odstup signál/šum o cca 6dB.

Uvážením $\Delta = (B - A) / 2^R$ dostávame $P_e = P_x 2^{-2R}$ resp. $SNR = P_x / P_e = 2^{2R}$ resp.

$SNR_{dB} = 10 \log_{10} 2^{2R} \approx 6RdB$ Napr. pri Audio CD vzorkách 16bitov môžeme dosiahnuť $SNR=96dB$

A/D, D/A konverzia audia (pokračovanie)

- Kvantizácia (audio depth)
 - $SNR \approx 6.02 R$ [dB], kde R je počet bitov audio vzoriek
 - 8bit - s využitím A-law, u-law – telefónia
 - 16bit – audio CD
 - dosahuje odstup signál/šum $(S/N) = 6\text{dB/bit} \cdot 16 = 96\text{dB}$
 - 24bitov – Blu-ray Disky, DVD Audio, štúdiová kvalita (mastering)
 - v praxi limitované schopnosťami hardvéru na cca 125dB, čo odpovedá 21 bitom
 - pri použití oversamplingu a tvarovania šumu sa dá dosiahnuť aj viac ...
- Dynamický rozsah zvuku
 - pomer medzi „najslabším“ a „najsilnejším“ signálom
 - udáva sa v dB
 - odpovedá počtu použitých bitov Q
- Dithering [\[zvuk\]](#)
 - na „rozptýlenie“ kvantizačnej chyby sa pridáva ešte šum predtým ako sa signál kvantuje
 - (napr. sme dlho tesne „pod“ rozhodovacím prahom ... ako vylepšiť výslednú hodnotu?)
 - znižuje skreslenie ale zväčšuje šum pozadia

Matlab – načítanie/analýza/úprava

- Demo súbory: auda_say1-10_8khz.wav, auda_say1-10_44khz.wav

<code>[y, Fs] = audioread('auda_say1-10_44khz.wav')</code> <code>sound(y, Fs);</code>	Načítanie a prehranie
<code>[m d] = audioinfo('auda_say1-10_44khz.wav')</code>	Zistenie informácií o súbore
<code>sound(y, Fs*2);</code>	Prehratie dvojásobnou rýchlosťou
<code>sound(y, Fs/2);</code>	Prehratie polovičnou rýchlosťou
<code>audiowrite(y, Fs, 'auda_say_out.wav')</code>	Uloženie zmeneného audia ...

- audiosúbory sa môžu sčítavať (mixovať $y+z+xx$...),
- filtrovať (konvolvovať), transformovať (filtrovať), frekvenčne analyzovať (spektrogramy)
- kopírovať - `[y;y;y]`
- zobrazovať - `plot(y)`
- otočiť v čase - `flipud(y)`
- pridať echo signál $b_echo(n) = b(n) + b(n-N)$;
- úprava dynamiky (tlmenie, zosilnenie), orezanie, `lin2mu`, `mu2lin`
- krájanie, strihanie, kopírovanie, nadpájanie (pozor na body nespojitosti)
- hrať sa zo stereom, odčítať navzájom kanály

...

Typická analýza - spektrogram

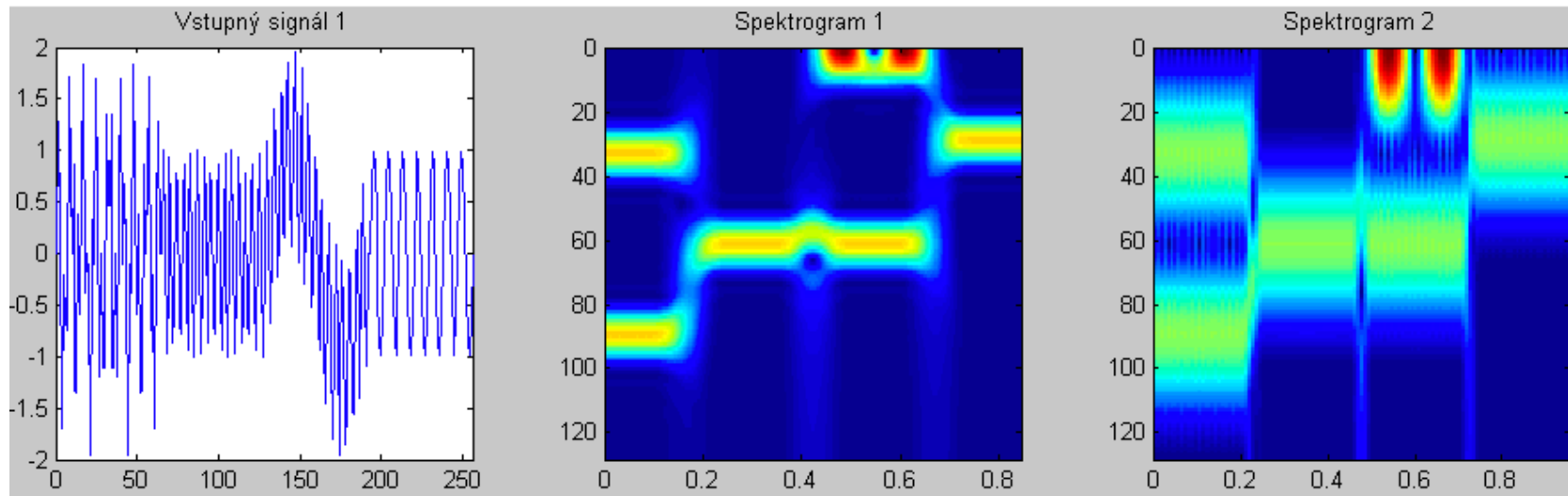
Nový druh fourierovej transformácie – krátkodobá fourierova transformácia (STFT - short time fourier transform)

- zisťujeme frekvenčný obsah iba v nejakom čase
- signál v čase pre násobíme oknovou funkciou a zistíme frekvenčný obsah tohto produktu
- výsledkom je 2D „plachta“
- $F(\omega, \tau) = \int_{-\infty}^{\infty} f(t)g(t - \tau)e^{-j\omega t} dt = \langle f(t)g(t - \tau), e^{j\omega t} \rangle$
 - je dôležité mať hladkú „oknovú funkciu“ – prečo?

Lebo násobenie v čase = konvolúcia v spektre (pravouhlé okno má si funkciu v spektre → rozvlní spektrum, vzniknú „falošné“ hodnoty)

- spektrogram = $|F(\omega, \tau)|$, t.j. zobrazujeme veľkosť

Príklad:



```
twopi=2*pi
t = 0:twopi/(64-1):twopi;
my_combi2=zeros(1,256);
my_combi2(1:64)=2*sin(7*t).*sin(15*t);
my_combi2(65:128)=sin(15*t);
my_combi2(129:192)=sin(1*t)+sin(15*t);;
my_combi2(193:256)=sin(7*t);
input1=my_combi2;
```

```
% plot input1
subplot(1,3,1);
plot(input1);
set(gca,'Xlim',[0 256])
title('Vstupný signál 1');
```



```
% plot SPECTROGRAM for input1
subplot(1,3,2)
[b,f,t]=specgram(input1,256,256,hanning(40),39);
imagesc(t,f,abs(b)), title('Spektrogram 1');

% plot SPECTROGRAM for input1
subplot(1,3,3)
[b,f,t]=specgram(input1,256,256,hanning(12),10);
imagesc(t,f,abs(b)), title('Spektrogram 2');
```

Typická úprava signálu – filtrácia

Filtrácia

- Definícia: Filtrácia = odstránenie nejakej neželanej resp. zachovanie iba nejakej želanej časti signálu
- Zvyčajne sa jedná o odstránenie/zachovanie nejakých frekvenčných zložiek, resp. rozsahu frekvencií
- Filtráciu môžeme vykonať
 - v čase: pomocou konvolúcie, resp. prechodom signálu systémom LDKI/LSKI
 - ak chceme dokonale zadržať nejakú frekvenciu, musí mať jej prenosová funkcia NULU pri príslušnom $z = e^{j\Omega}$
 - v spektra: pomocou násobenia, resp. inej úpravy spektra
 - ak chceme dokonale zadržať nejakú frekvenciu je nutné vynulovať (prenásobiť nulou) príslušné spektrálne koeficienty

Ako pomocou LDKI systému môžeme robiť filtráciu?

Zadanie: vytvorme takú LDKI sústavu, ktorá pri $f_{vz} = 44100\text{Hz}$ (ako pri CD) neprepustí harmonický signal 440Hz . Ako takéto niečo vytvoriť? (príklad na ďalšej strane).

Ako pomocou LDKI sústav odfiltrovať zvolenú frekvenciu?

Uvažujme sústavu s $H(z) = 1 + z^{-1} \Rightarrow h(n) = (1, 1)$



$$x(n) * h(n) = y(n)$$

ak $x(n)$ obsahuje frekvenciu $f_{vz}/2$ táto sa prechodom cez LDKI sústavu s daným $h(n)$ stratí.

Napr.: - máme signál $x(t)$
- navzorujeme -1-
ho pri $f_{vz} = 2\text{Hz}$: $x(n) = (1, -1, 1, -1, 1, -1, 1)$

potom $y(n) = x(n) * h(n)$

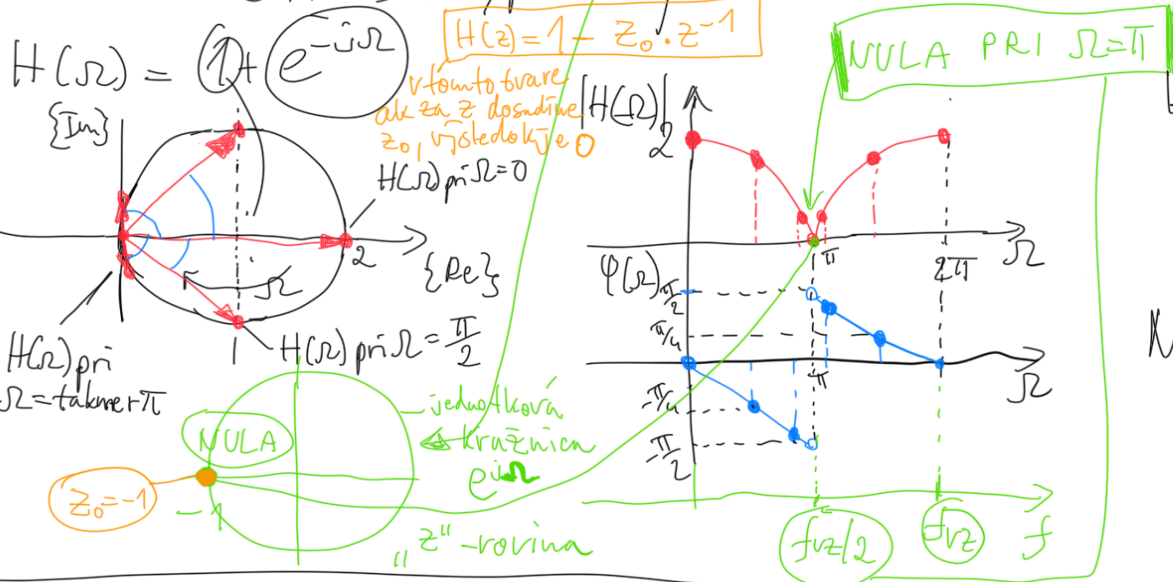
$$x(n): +1, -1, +1, -1, +1, -1, +1$$

$$h(n): 1, 1$$

$$x(n) * h(n) = (1, 0, 0, 0, 0, 0, 1)$$

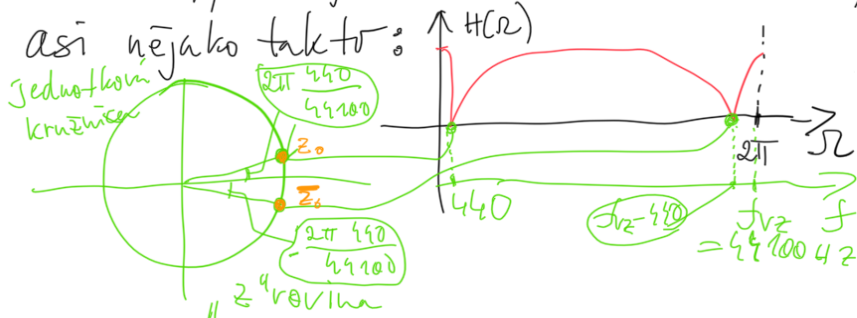
vidíme, že 1 Hz signál sa pri $f_{vz} = 2$ sústavou nepreniesol!

Na základe ADSS 1 vypočítame frekvenčné charakteristiky:



Na základe uvedeného, ako musí vyzerať frekvenčná charakteristika sústav, keď $f_{vz} = 44100$ a nechceme aby prepustila 440 Hz?

asi nejako takto:



$$z_0 = \frac{2\pi \cdot 440}{44100}$$

$$H(z) = (1 - z_0 z^{-1})(1 - z_0^* z^{-1})$$

$$= 1 - 2 \operatorname{Re}\{z_0\} z^{-1} + z_0 z_0^* z^{-2}$$

$$\Rightarrow h(n) = (1, -2 \operatorname{Re}\{z_0\}, 1)$$