

Zvuk a multimédia I.

Adam J Sporka

sporkaa@fel.cvut.cz

Petr Skřípal

petr.skripal@seznam.cz

Multimédia a Animace, zima 2010/2011

Přednášky o zvuku v MMA

- Základy (21.10.2010)
 - Co patří pod pojem „zvuk“
 - Fyzikální principy
 - Vnímání zvuku
- Práce se zvukovým signálem (4.11.2010)
 - Zpracování v časové oblasti (hlasitost, střih, ...)
 - Zpracování ve frekvenční oblasti (filtry, pitch shifting, ...)
 - Zvukové efekty
 - Syntéza zvuku
- Práce s hudbou (11.11.2010)
 - Rozhraní pro tvorbu hudby
 - Procedurální audio

+ workshop: Ozvučení klipů z 3. cvičení

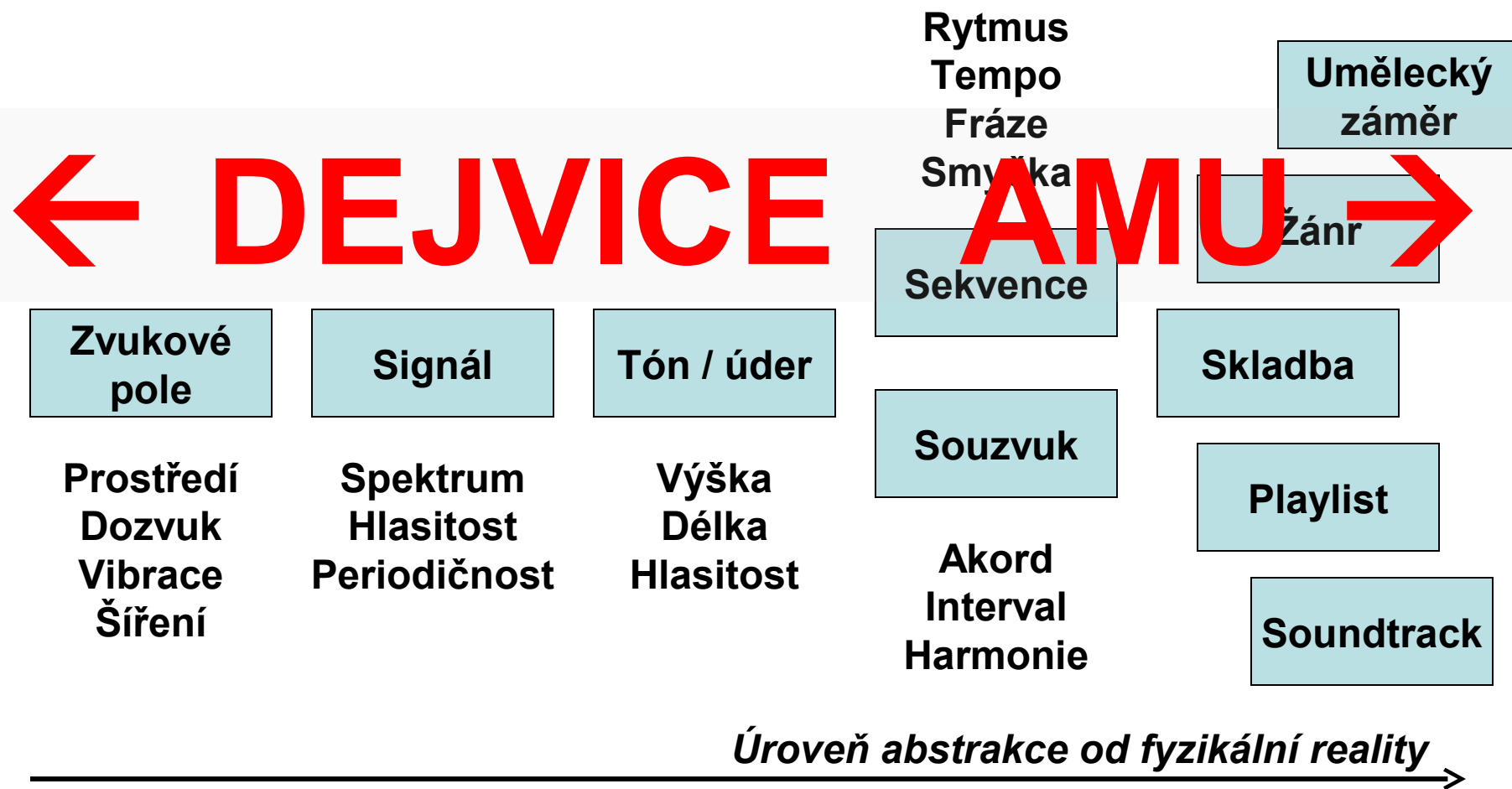


Zvuk: Nejednoznačný pojem

- „Fyzikální“ pohled
 - Zvukové pole ... vibrace v celém prostoru, a jak se šíří
 - Signál ... průběh vibrací v jednom bodě
 - Obsah signálu ... signatura toho, co vibrace způsobilo
- „Lidský“ pohled
 - Tón, ruch, úder ...
 - „Komorní A na piánu.“
 - Vzorek, sample, smyčka ...
 - „Kousek písničky, kterou remixujeme.“
 - Akustické vlastnosti nějakého zařízení
 - „Sluchátka XY mají hezkej zvuk.“

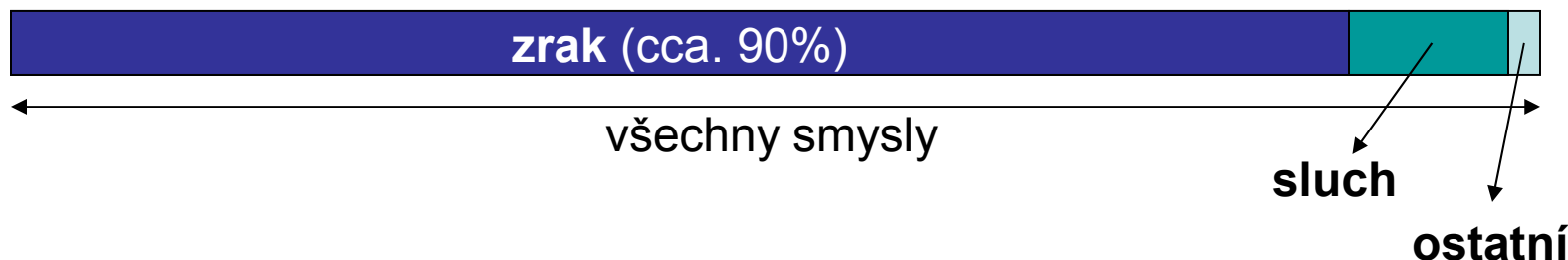


Úrovně popisu zvuku



Co se zvukem v multimédiích?

- Z hlediska kognitivních věd:
 - Zvuk nese pouze doplňkovou informaci.



- Z hlediska multimédií:
 - Zvuk vytváří atmosféru („dej tam nějakou smutnou muziku, a ať tam někdo naříká“)
 - Častý předpoklad: „Cool music implies cool stuff.“
 - Často vede ke cliché

Co se zvukem v multimédiích?

- Zvuk musí věrohodně doplnit vizuální stránku díla.
 - <http://www.youtube.com/watch?v=EbnZQuZtvGE>
<matrix-audio-swap.mp4>
- Absolutní autentičnost zvukového záznamu není často důležitá.
 - „Jeskyně velká 500m nebo 480m.“
- **Uvěřitelnost** je důležitější než realističnost
 - Hyperrealismus
 - “Očekávatelné parametry.”



Jak nepoužít zvuk

<http://www.hrubasy.cz>



Zvuk může být i vstup

- Rozpoznávání řeči

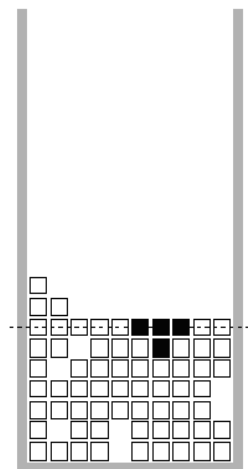
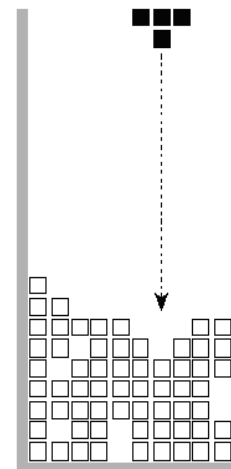
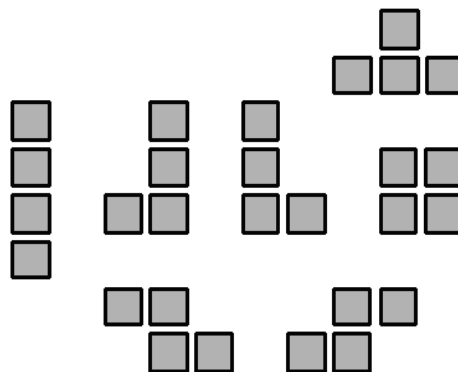
- Převod nahrávky mluvčího na text

<http://www.youtube.com/watch?v=kX8oYoYy2Gc>

Neverbální hlasový vstup (NVVI)

- Použití neřečových zvuků pro vstup informace

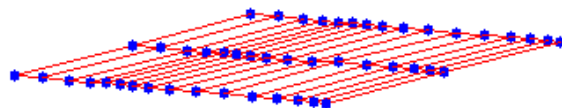
http://www.youtube.com/watch?v=Yv_M1rcM...



1. FYZIKÁLNÍ ASPEKTY

Fyzikální definice zvuku

- **Zvuk** je každé podélné (v pevných látkách případně také příčné) **mechanické vlnění** v látkovém prostředí, které je schopno vyvolat v lidském uchu sluchový vjem



- Prostředí: plyny, kapaliny, pevné látky
 - Ve vakuu se zvuk šířit nemůže

Definice zvuku

- Variace tlaku v prostředí (akustický tlak)
 - Přenášeny zvukovými vlnami
 - Rychlost ve vzduchu asi 330 m/s (1188 km/h)
 - Charakter šíření závisí na frekvenci zvuku
- Oscilace v intervalu slyšitelných frekvencí (20 Hz – 20 kHz)



< 20 Hz
infrazvuk
> 20 kHz
ultrazvuk



Co lidé neslyší...

- ~ 220 kHz



- ~ 100 kHz



- ~ 4 Hz



- ~ 7 Hz



Co známe z fyziky...

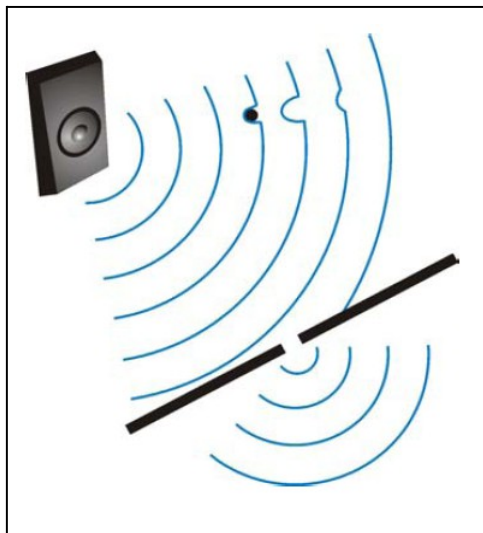
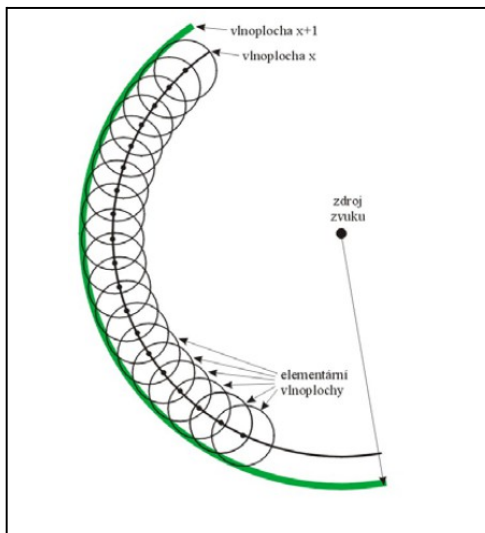
Huygensův princip

→ vlnoplocha

- obálka elementárních vlnění
- každý bod vlnoplochy je zdrojem elementárního vlnění

→ ohýb a lom zvuku

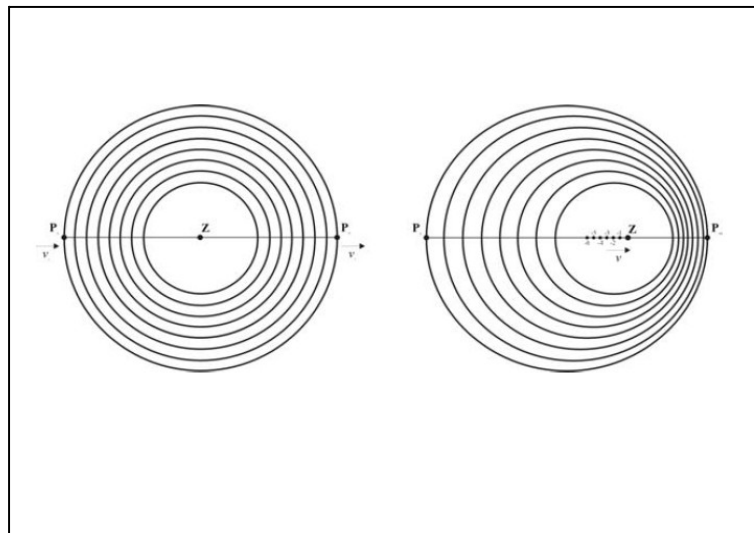
- lépe se ohýbají dlouhé vlny (= nízké frekvence)



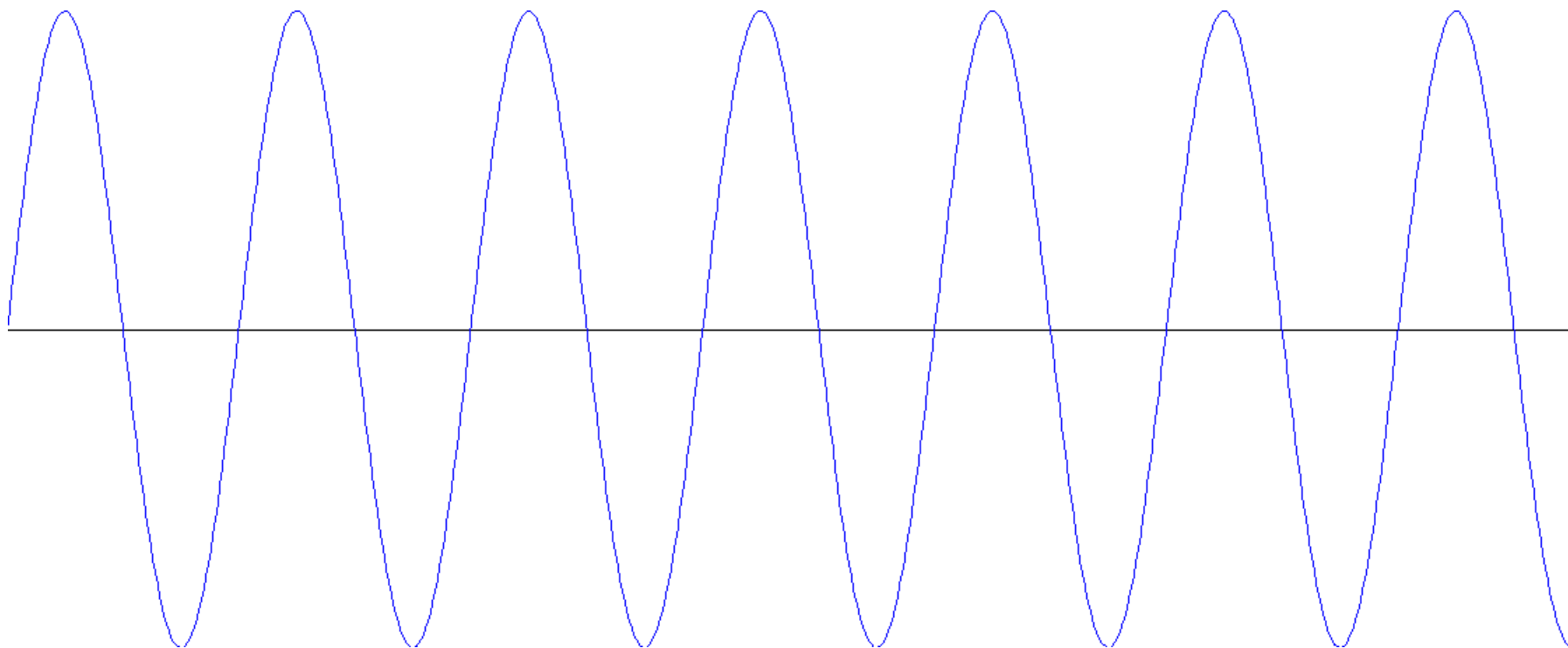
Dopplerův efekt

- vzájemný pohyb zdroje zvuku a pozorovatele
- pozorovatel vnímá zvuk jiné frekvence, než je skutečná frekvence

kmitání zdroje zvuku



Jednoduchý tón



$$A(t) = \sin(\omega t)$$

<D>



Zvukový signál

Sinusový / harmonický / periodický pohyb:

- **T - perioda T [s]**
- Popis spojitého zvukového signálu v čase

- **f - frekvence [Hz]**

- **A - amplituda [dB]**

$$f(t) = A \sin(2\pi f t + \varphi)$$

- **φ - fáze [°]**

- **λ vlnová délka [m]**

- Reprodukce:

- Generování zvukového pole, které nese tento (stejný) signál

$$\lambda = \frac{c_{\text{zvuk}}}{f}$$

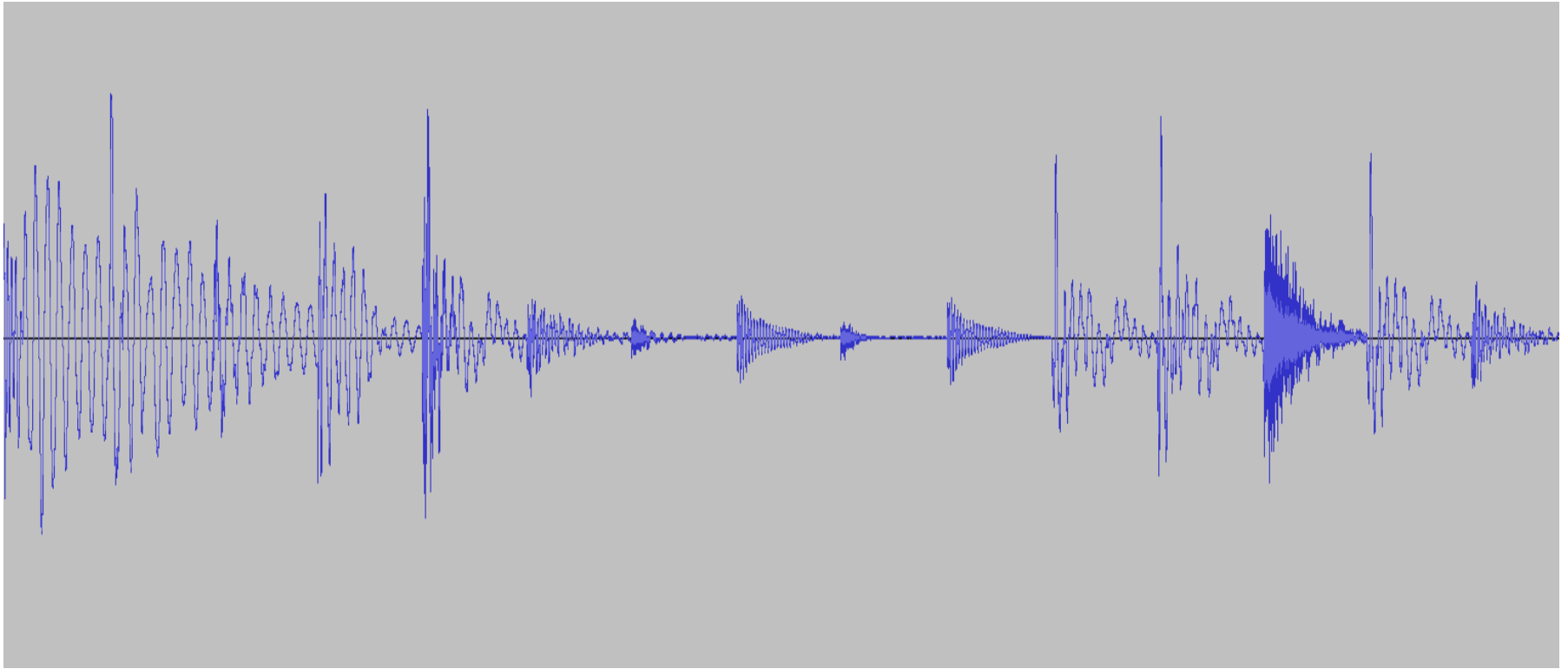


Zvukový signál

- Zvukové pole × zvukový signál
- Zvukový signál: Popis oscilací v čase
- Záznam zvuku – reprezentace signálu:
 - Mechanické vychýlení
 - Membrána reproduktoru
 - Oscilace struny nebo ramene ladičky
 - Zářezy vinylových desek
 - Variace napětí
 - Úroveň či směr magnetizace



Složený zvuk



$A(t) = \text{funkce_popisujici_udery_do_bicich}(t)$



Zvuk vs Obraz

Zvuk

- Jednorozměrný signál
 - $I(t) : \mathbb{R} \rightarrow \mathbb{R}$ (*analog*)
 - $I[t] : \mathbb{R} \rightarrow \mathbb{R}$ (*discrete*)
- Vnímán ve frekvenční oblasti
- Není perzistentní
- Sekvenční přístup k informaci

Obraz

- Dvourozměrný “signál”
 - $I(x,y) : \mathbb{R}^2 \rightarrow \mathbb{R}$
 - $I[x,y] : \mathbb{R}^2 \rightarrow \mathbb{R}$
- Vnímán v prostorové oblasti
- Perzistentní
- Náhodný přístup k informaci



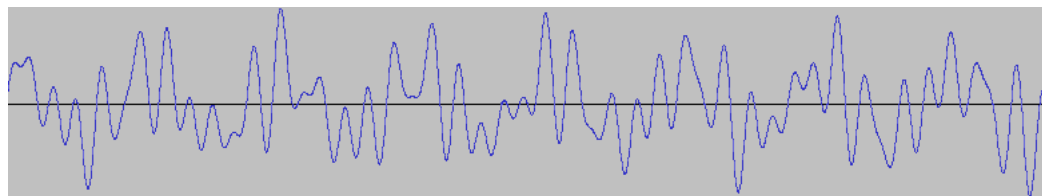
Kmitočtové spektrum signálu

- **Většina zvuků nejsou jednoduché tóny.**
 - Dají se popsat jako „součet“ jednoduchých tónů
 - Jinak řečeno: zvuk je superpozicí elementárních vln na jednotlivých frekvencích.
- **Spektrum:** Množství energie signálu v různých frekvenčních rozsazích.

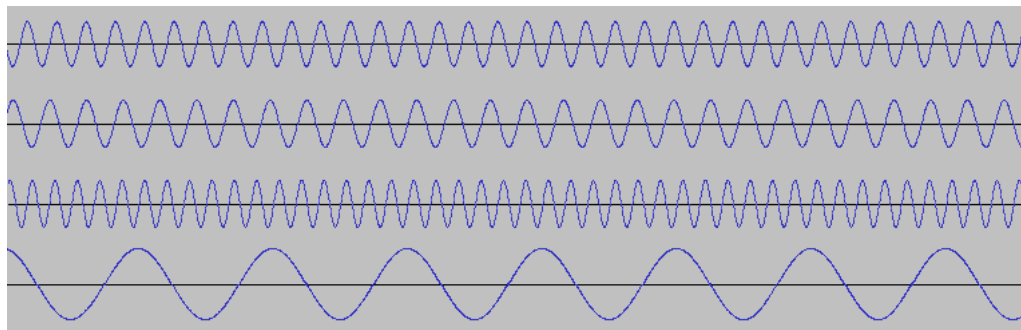


Superpozice vln

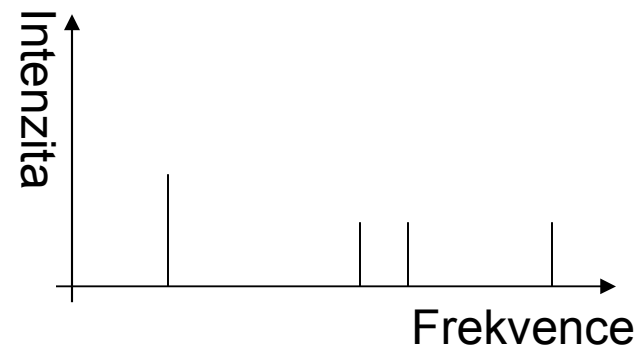
- Spektrum: Množství energie signálu v různých frekvenčních rozsazích.



=



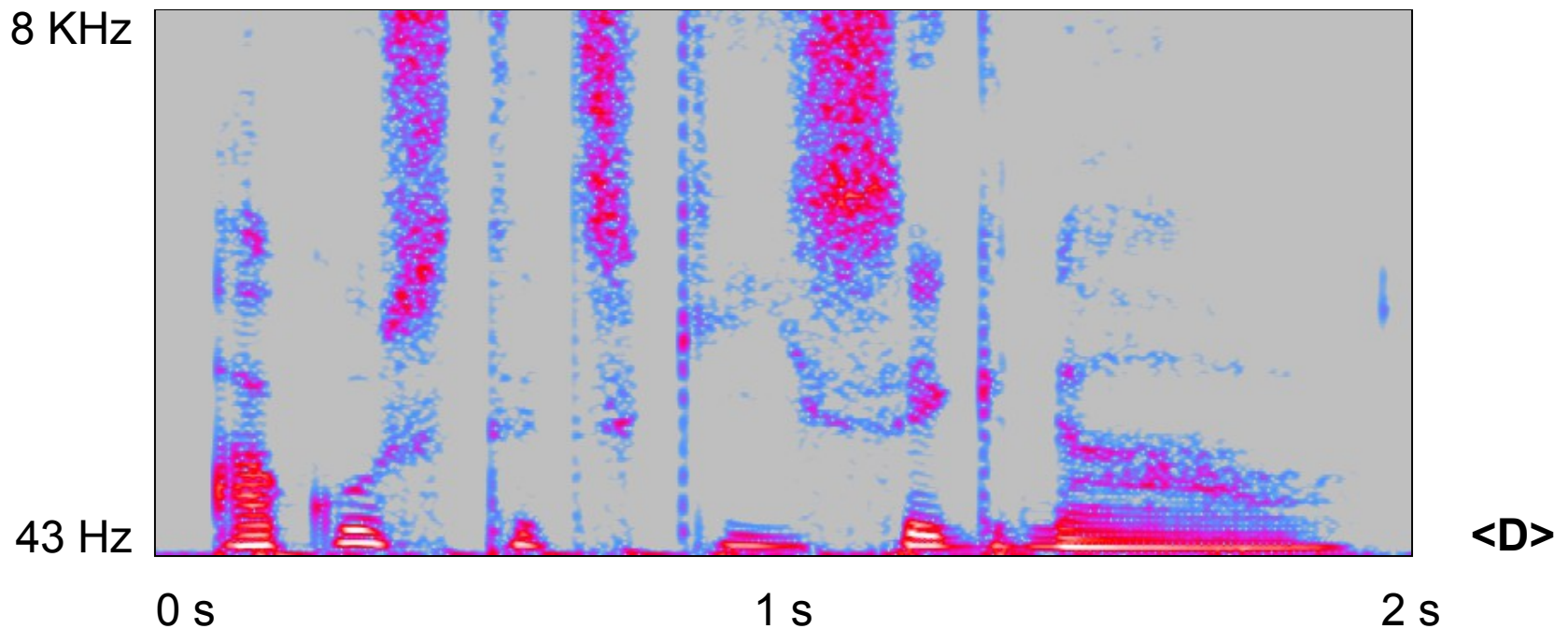
Spektrum



<D>

Spektrogram

- Zobrazení průběhu frekvenčního spektra signálu v čase



Typy spekter

- **Harmonické**

- *Samohlásky, trumpeta...*
- Pravidelné kvaziperiodické signály
- U tohoto druhu jasně vnímáme výšku tónu

- **Neharmonické**

- *Zvony, gongy, některé bubny*
- Slabé či nejasné vnímání výšky tónu

- **Šum**

- *Souhlásky, některé perkusivní nástroje*
- Komplexní superpozice mnoha harmonických a neharmonických zvuků
- Vnímání “vysoký” vs „nízký“, výška tónu však není zachycena

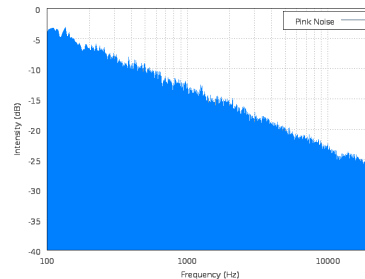


Šum

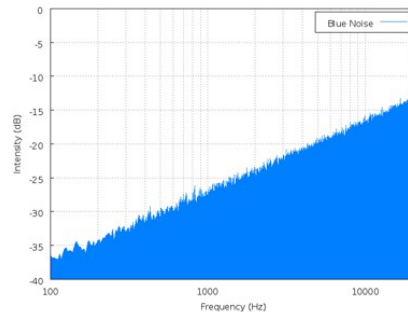
- **Druhy:**

- Bílý
- Růžový
- Hnědý
- Modrý
- Purpurový
- Šedý
- Červený
- Oranžový
- Zelený
- Černý (ticho)

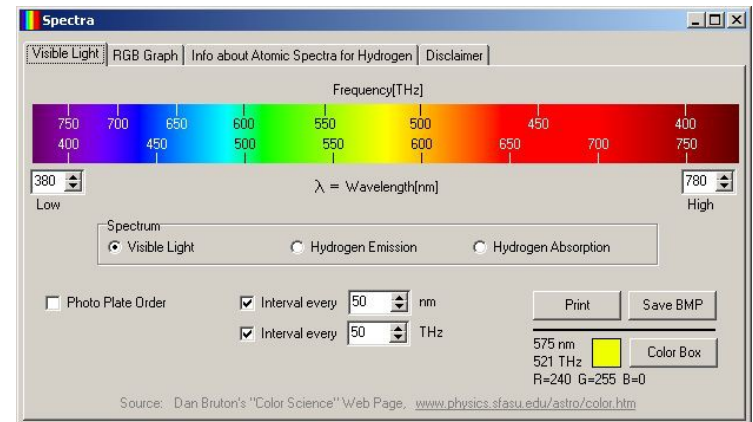
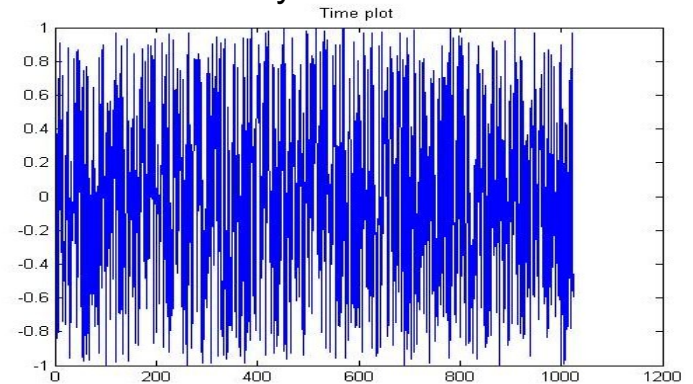
Růžový šum



Modrý šum



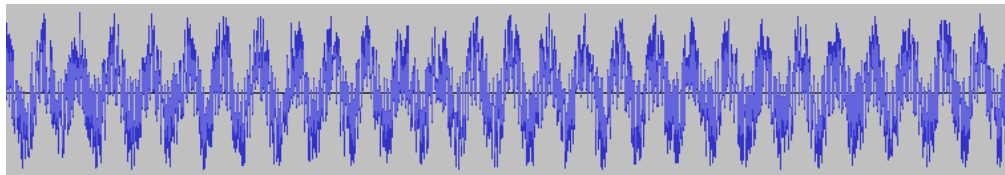
Bílý šum



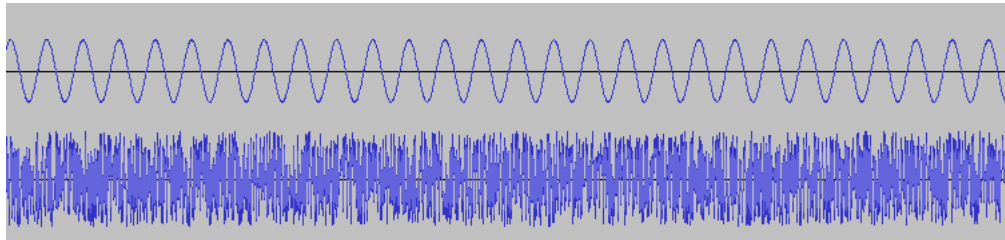
Zdroj: <http://encyklopedie.seznam.cz/heslo/445253-barvy-sumu>



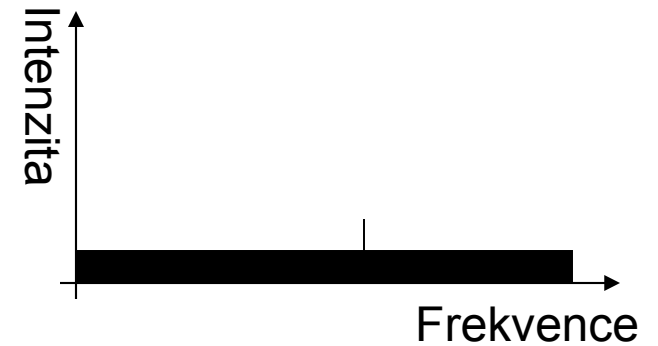
Signál + šum



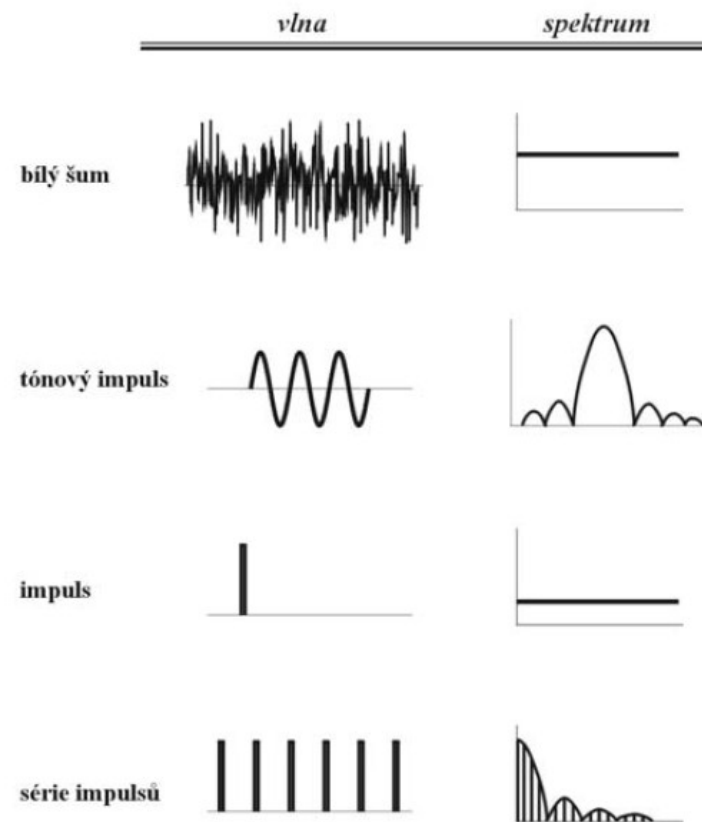
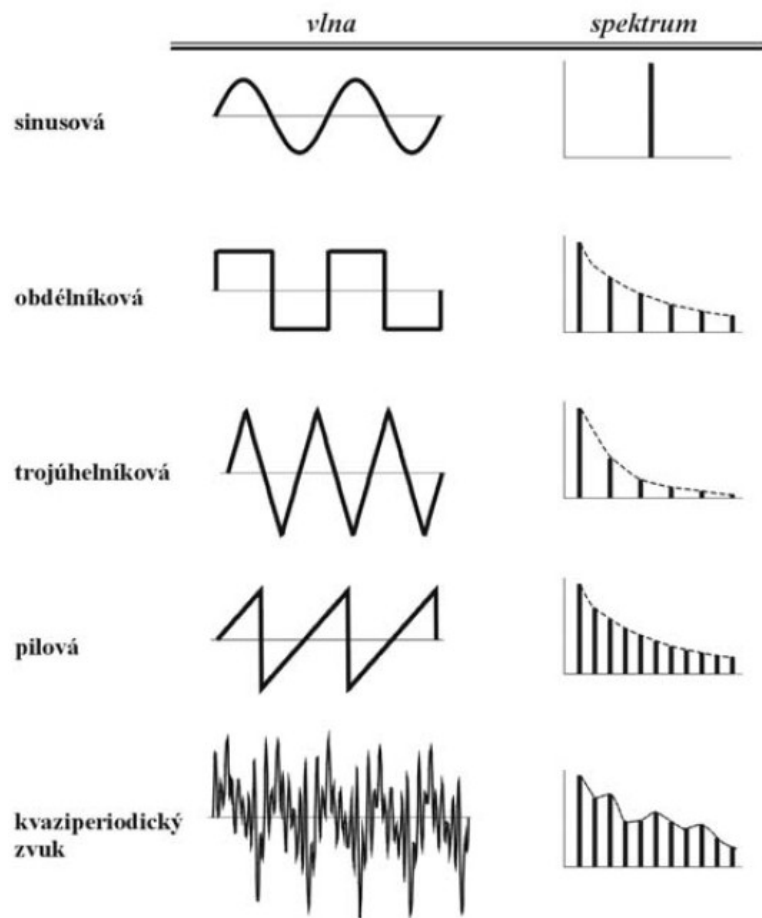
=



Spektrum



Typy vln



Transformace do spektra

- Fourierova transformace

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

- Zpětná Fourierova transformace

$$x(t) = \int_{-\infty}^{\infty} X(\omega) e^{j\omega t} d\omega$$



2. DIGITALIZACE ZVUKU

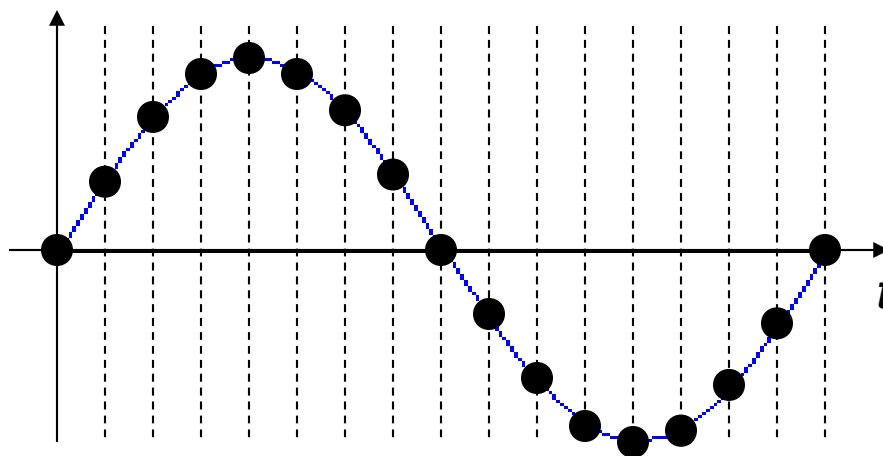
Počítačové zpracování signálu

- Počítač:
 - Omezená velikost paměti
 - Omezené rozlišení paměťových buňek (bit/buňka)
- Důsledek:
 - Signál zpracováván jako sekvence diskrétních hodnot



Vzorkování zvuku (A/D přev.)

- Periodické měření úrovně signálu tak, aby bylo možné rekonstruovat původní signál.
- Uložení sekvence těchto měření (vzorků) do paměti

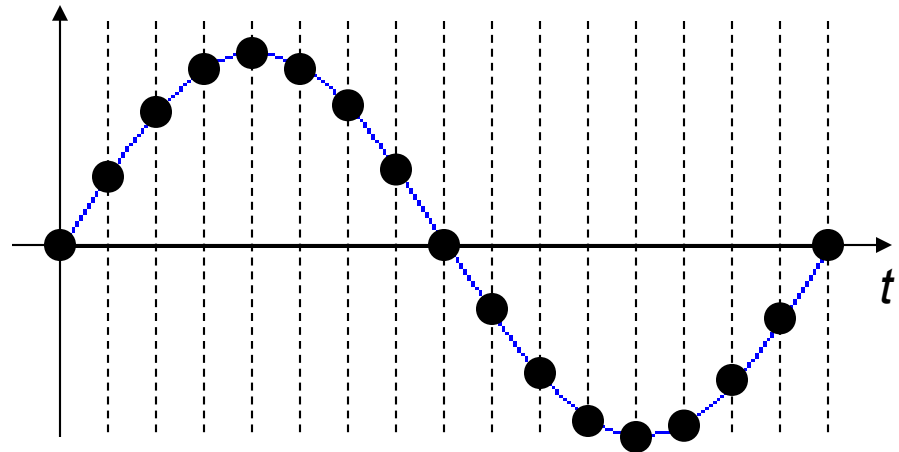


0.000	– 0.383
0.383	– 0.707
0.707	– 0.924
0.924	– 1.000
1.000	– 0.924
0.924	– 0.707
0.707	– 0.383
0.383	0.000
0.000	

D/A převodník

- Rekonstrukce analogového signálu
- Sekvence čísel řídí generátor signálu

0.000	– 0.383
0.383	– 0.707
0.707	– 0.924
0.924	– 1.000
1.000	– 0.924
0.924	– 0.707
0.707	– 0.383
0.383	0.000
0.000	

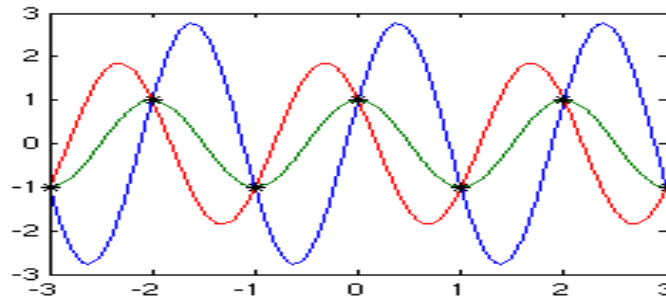


Vzorkovací frekvence

- = jak často pořizujeme vzorky
- **Nyquistova frekvence (Shannonův teorém)**
 - Maximální frekvence signálu, kterou je ještě možné zaznamenat při použití dané vzorkovací frekvence

$$f_N = f_s / 2$$

- Při vzorkovací frekvenci blízké f_N nebude již však věrně zaznamenáno množství energie na této frekvenci



Vzorkovací frekvence

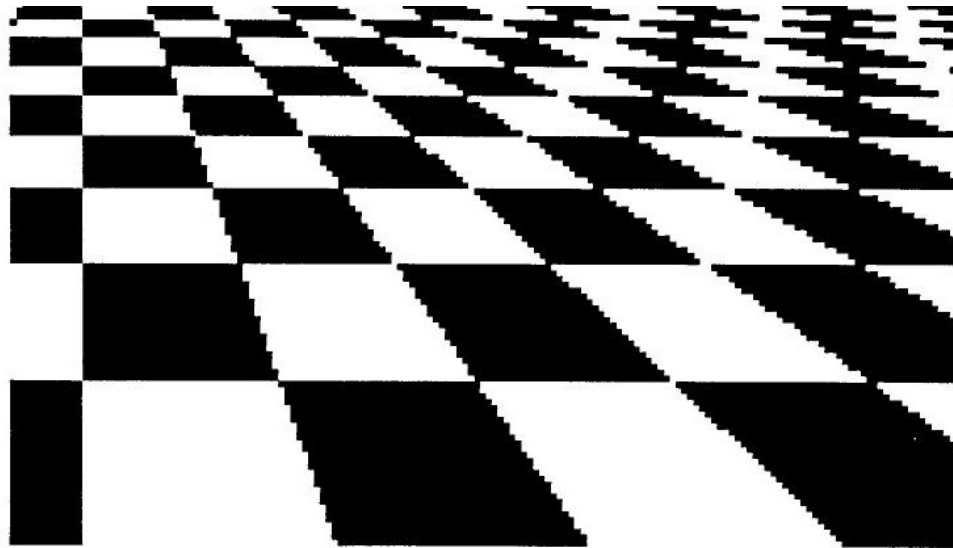


Figure 3.2: *Texture mapping with point sampling, with more extreme compression. Note moire near horizon.*

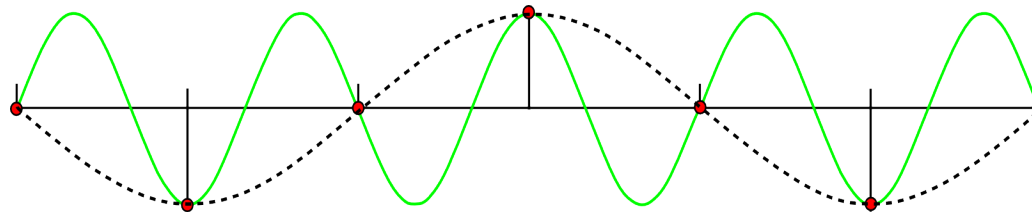
http://www.cs.berkeley.edu/~sequin/CS184/LECT_09/L28.html



Vzorkovací frekvence

- **Aliasing**

- Vzorkování vyšších frekvencí, než je Nyquistova frekvence

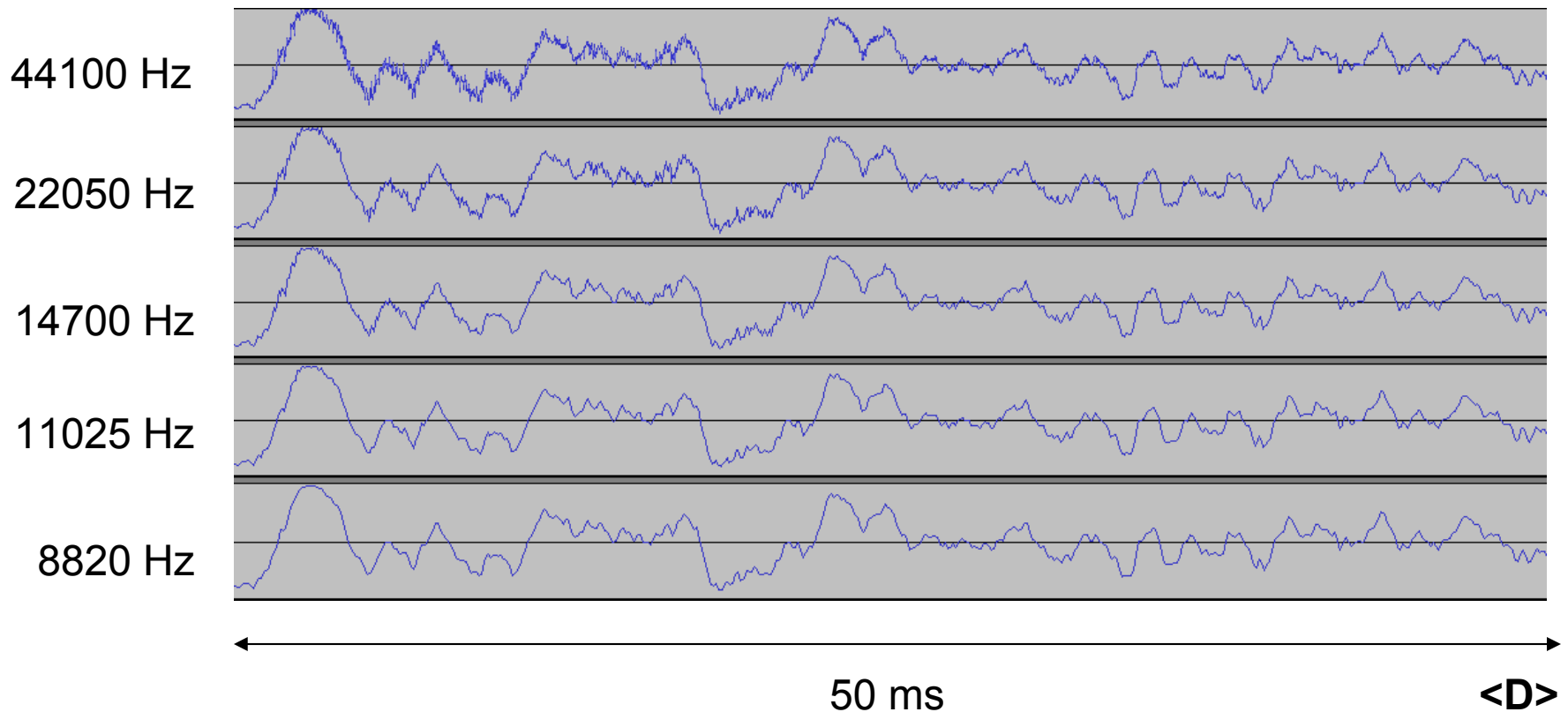


- Jak předejít?

- Filtr vyšších frekvencí než jejich Nyquistova frekvence

Jakou frekvenci vzorkování použít?

- Čím vyšší frekvence, tím věrnější výšky



Jakou frekvenci vzorkování použít?

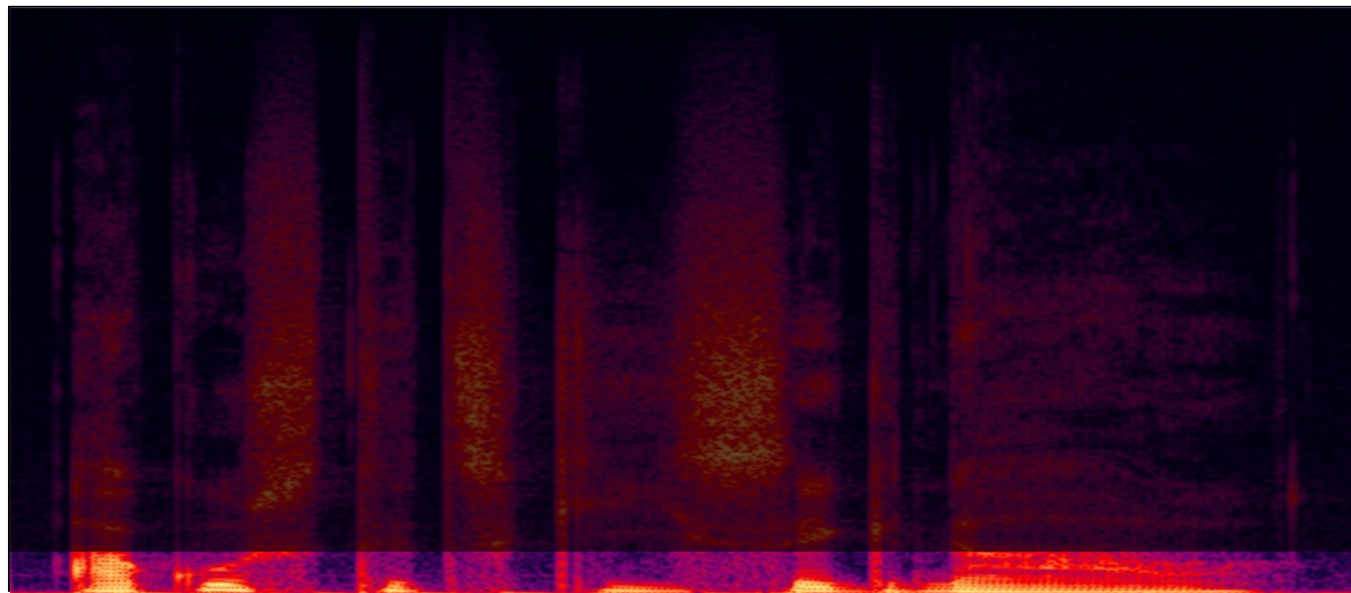
- cca 8 kHz Telefonní síť
- 44,1 kHz CD
- 48 kHz miniDV, DVD, DAT
- 96, 192 kHz High-Definition DVD audio

21 KHz

<D>

10 Hz

<D>



0 s

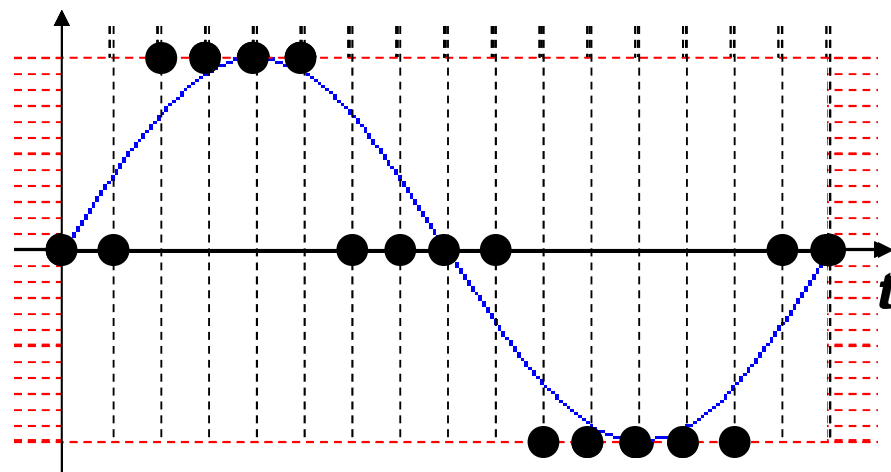
1 s

2 s



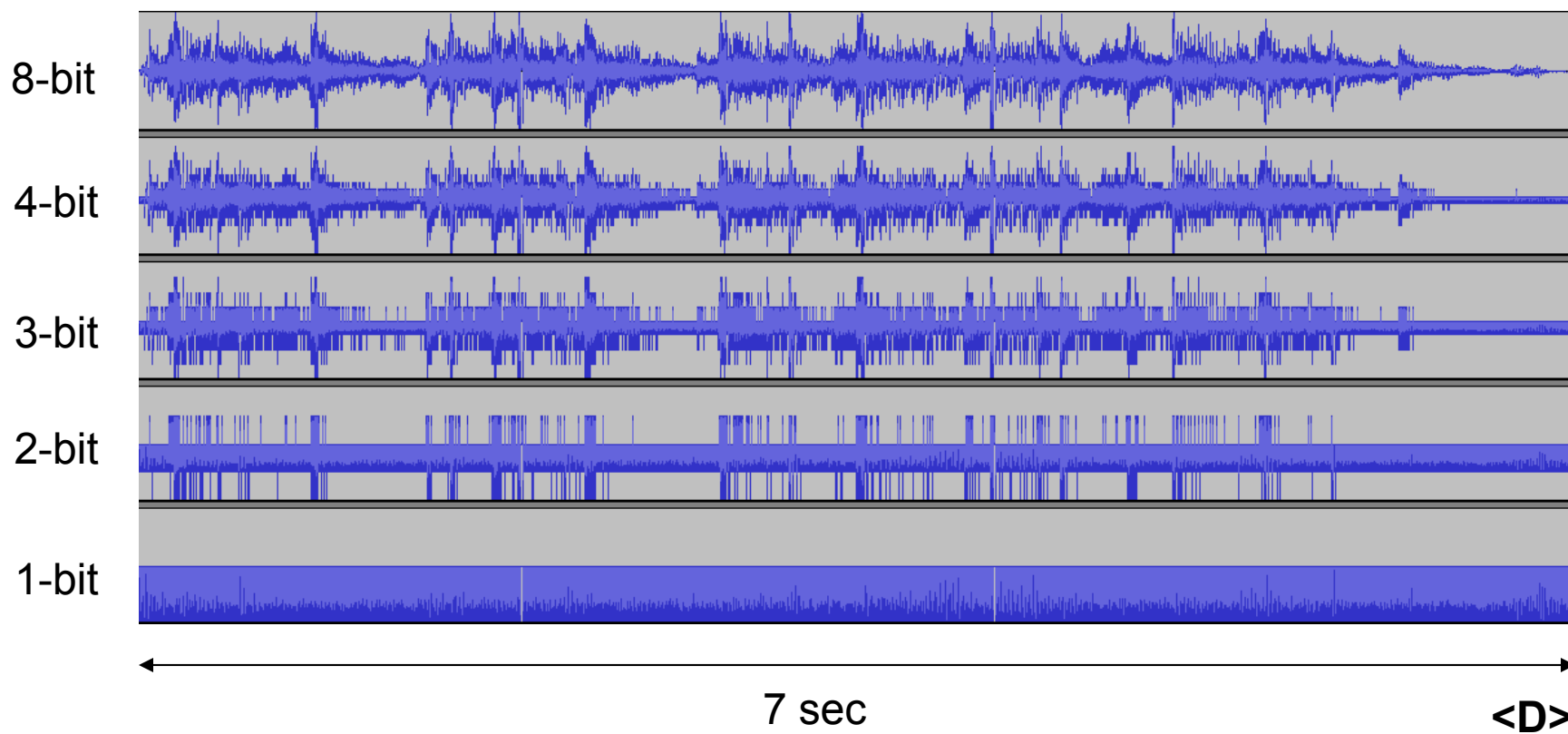
Hloubka rozlišení

- Naměřené hodnoty ukládány v diskrétním číselném oboru
 - Hloubka rozlišení udávána v bitech
 - Dnes typicky 8, 12, 16, 24, 32 bitů
- chyba měření každého vzorku
- zkreslení signálu



Hloubka rozlišení

- Srovnání:



Hloubka rozlišení

- Typická rozlišení:
 - 1-bit PC Speaker ☺
 - 4-bit Atari 800 XL
 - 8-bit Sound Blaster (Creative Labs)
 - 12-bit LP běžných digitálních kamer
 - 16-bit CD Audio
 - 24-bit DVD
- Potřeba většího rozlišení:
 - Možnost následného zpracování
 - Obdoba detailu fotografie



Dynamický rozsah

- Poměr mezi nejnižší a nejvyšší použitou amplitudou
- Škála úrovní hlasitosti signálu, při kterých je signál věrně reprezentován.
- Dynamický rozsah lidského ucha: 120 dB
- Jednotka: dB

$$X_{dB} = 20 \log_{10} \left(\frac{X}{X_0} \right)$$

- Př.: 8 bitů = 256 různých úrovní amplitudy signálu.
 $X = 1, X_0 = 256 \rightarrow X_{dB} = \underline{48dB}$
- Př.: 24 bitů $\rightarrow$ 144 dB

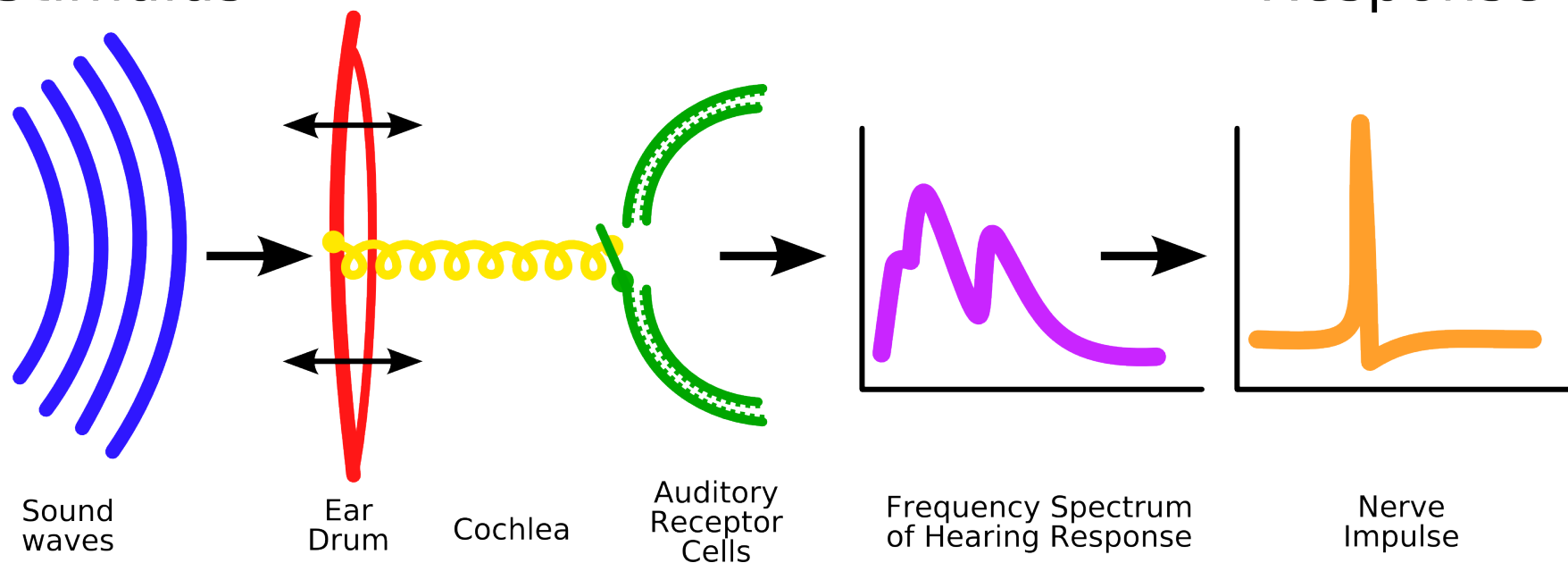


3. PSYCHOAKUSTIKA

Sluchový systém

Stimulus

Response



By: Tim Gollisch, Andreas M. V. Herz, and [Public Library of Science](#)
Diagram used under Creative Commons Attribution 2.5 license
DOI: 10.1371/journal.pbio.0030026.g001.



Výška tónu

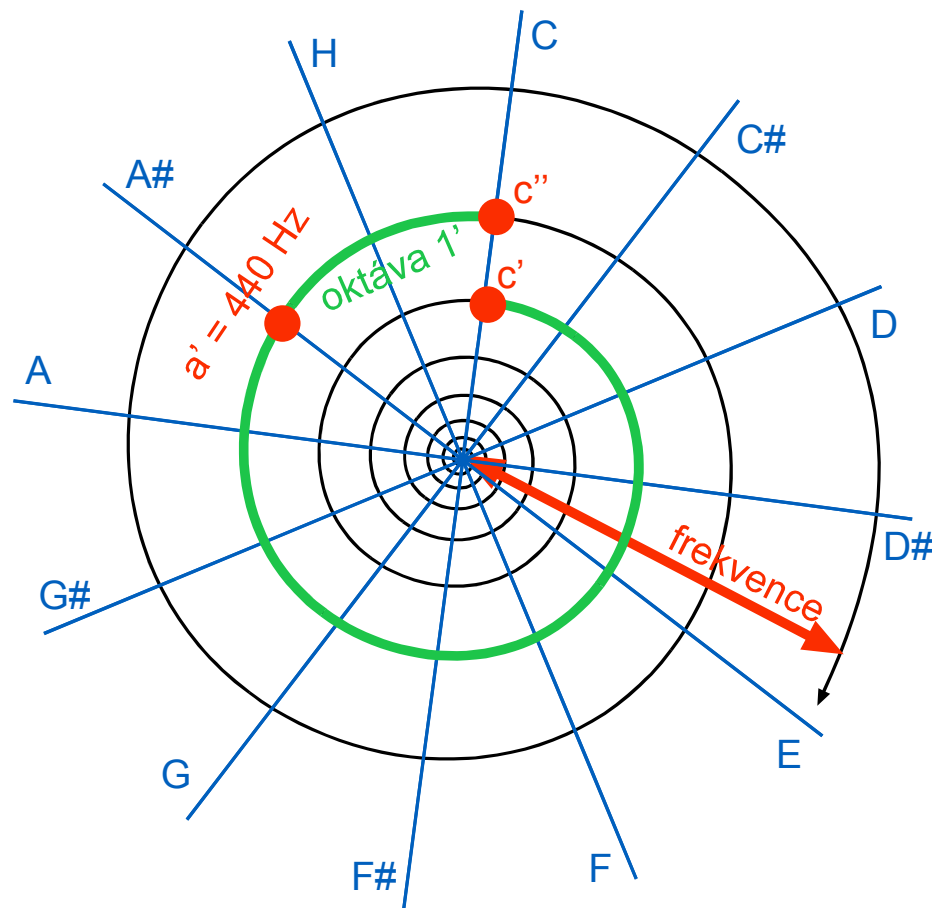
- Fundamentální frekvence zvuku [Hz]
 - Nejnižší frekvenční složka zvuku
- Výška tónu
 - Dle ANSI: Atribut, dle kterého je možné subjektivně porovnávat a řadit různé tóny od nízkých po vysoké
 - To, jak je vnímána základní (fundamentální) frekvence zvuku.
 - Může se lišit od skutečné frekvence zvuku.
- Logaritmické vnímání výšky
 - Stejný poměr frekvencí vnímán jako stejná vzdálenost (interval) mezi tóny
 - Př:

$f_1 = 440 \text{ Hz}$	$f_2 = 880 \text{ Hz}$	$f_3 = 1760 \text{ Hz}$
$f_2 - f_1 = 440 \text{ Hz}$	$f_3 - f_2 = 880 \text{ Hz}$	
$f_2 / f_1 = 2$	$f_3 / f_2 = 2$	



Vztah výšky a názvu tónu

- Oktáva může znamenat
 - Interval mezi tóny
 - Dohodnuté označení intervalu absolutních výšek tónů
- Název tónu:
 - Třída výšky
 - C, C#, D, D#, E, ...
 - Označení oktávy
 - Jednočárkovaná (od c' do c'')
 - Dvojčárkovaná ...
- Moderní konvence ve Střední Evropě
 - $a' = 440$ Hz
 - Též “komorní a”



Percepce amplitudy

- Škály hlasitosti
 - sestup hlasitosti po 1, 3 a 5 dB <D> <D> <D>
- Různá **hlasitost**, ale stejná intenzita <D>
- **Výška** tónu (pitch) <D> <D> <D>
 - přijde až od určité délky tónu
 - závislá na fundamentální frekvenci F_0

Zdroj: <http://www.ece.uvic.ca/~aupward/p/demos.htm>

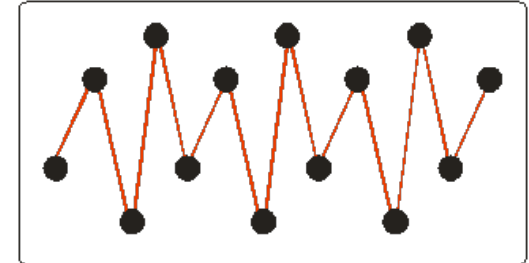


Percepce času

- **Seskupování a segregace**

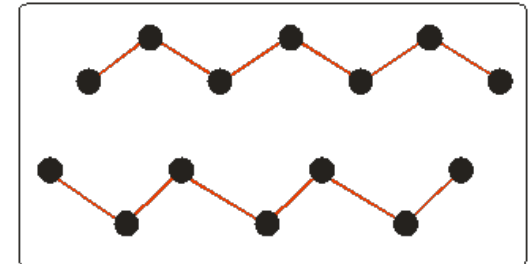
- melodie je rozpoznána jen, pokud jsou změny v čase dostatečně pomalé

<D>



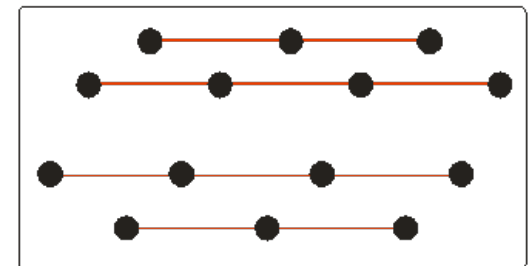
- o něco zrychlená sekvence
→ dvě skupiny tónů

<D>



- další zrychlení
→ melodie je již neslyšitelná

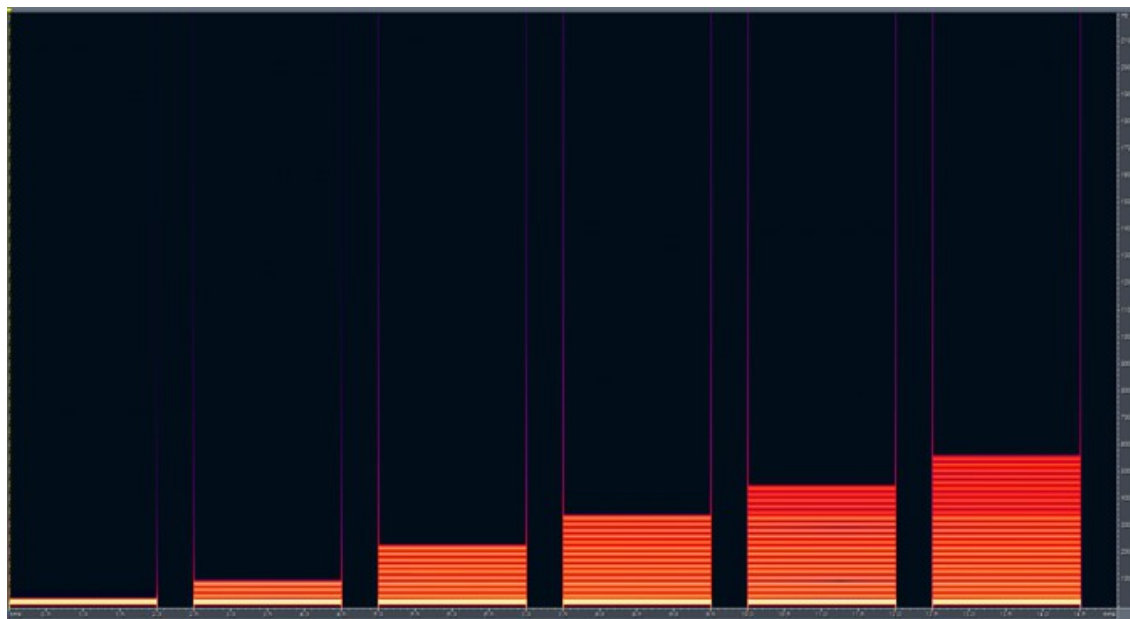
<D>



Percepce frekvencí

- **Kvalita zvuku**
 - Závisí na vyšších harmonických frekvencích

<D>



Hranice percepce

- **Maskování (ve frekvenční oblasti)**

- neslyšitelnost určitých tónů, ačkoliv je jejich intenzita nad prahem slyšitelnosti.
- dva čisté tóny vzájemně blízkých frekvencí: slabší je zcela přehlušen silnějším. <D>
- čím silnější je tón, tím širší oblast frekvencí kolem sebe maskuje.
- čisté tóny snadněji maskují tóny vyšších frekvencí<D> než nižších.

<D> <D>



Zvukové iluze

- Binaurální rázy
 - Dva tóny o různé výšce → slyšitelné rázy
 - Podobný vjem i když levé ucho přijímá signál o jiné výšce než pravé ucho
 - Vyzkoušejte doma se sluchátky

<binaural modes>



Zvukové iluze

- *Deutsch Scale Illusion*

- Dvě škály proti sobě
- Tóny škál proložené do obou uší
- A: Co je v jednotlivých kanálech
- C: Častý vjem

left right 240

A. SOUND PATTERN

B.

C. PERCEPTION

The pattern that produces Deutsch's scale illusion (A), and a way that it is often perceived (C). The notation in (B) shows how the pattern is composed of ascending and descending scales.

– http://www.philomel.com/musical_illusions/example_scale_illusion.php



Zvukové iluze

- *Shepard's Ascending Tones (Shepard's Paradox)*
 - Nekonečně dlouhý vzestup výšky tónu
- Padající zvony
 - Vjem snižující se výšky tónu
 - Skutečnost: Výška tónu se zvyšuje



Zvukové iluze

- McGurk Effect
 - (McGurk, MacDonald, 1976)
 - Kombinace zvukové a vizuální modality
 - Zvuk: „ba ba“
 - Obraz: „ga ga“
 - Vjem: „da da“
 - <http://www.youtube.com/watch?v=eQoYKuNcCpU>
- Melodie rozložená do více oktáv



To, co není slyšet, ...

... nemusí se ukládat

- Je možné vynechat část zvukové informace

✂ → ztrátové zvukové formáty (nepoužívat pro pracovní formáty 😊)

... nevadí, pokud se uloží

- Při kompresi můžeme generovat takový šum, který zůstane pod prahem slyšitelnosti



Datový tok - bitrate

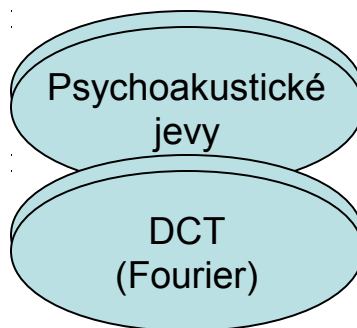
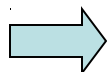
- Množství přenesených informací (dat), např:
 - CD DA - (44.1kHz / 16 / stereo [2 kanály])
 - $44100 \text{ samples/sec} * 16 \text{ bits/sample} * 2$
 - 1.4 Mbit/s



- Komprese zvuku

Bezeztrátové

- CA – DA
- WAV (RAW)
- FLAC



Ztrátové

- ATRAC (minidisc)
- MP3, AAC (.mp4)
- Vorbis OOG

