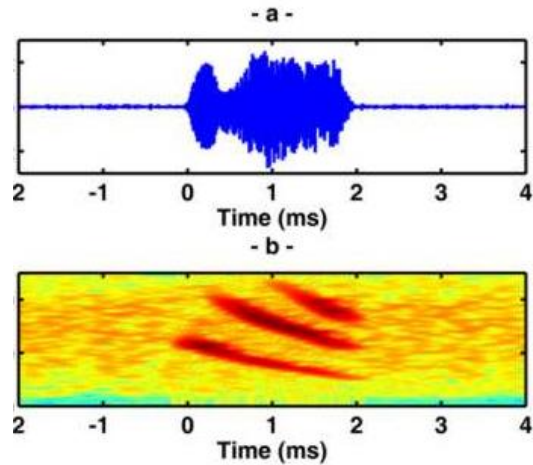


11. týden – HLAS A ŘEČ II.

Akustické a jazykové analýzy

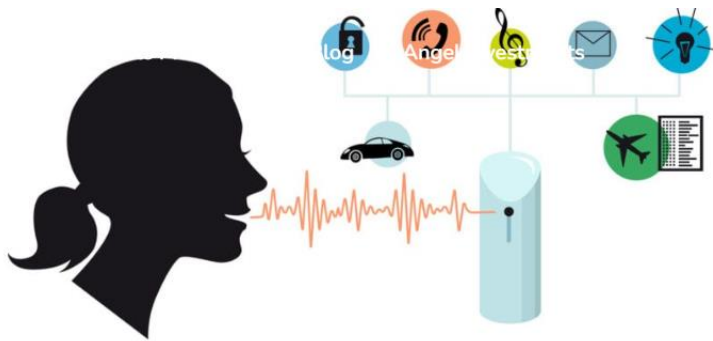


- **Lab 11a: Sonifikace animálních modelů**

- Hlasová komunikace animálních modelů
 - Netopýři
 - Laboratorní myši
- Fázový vokodér
 - Časové a frekvenční modifikace

- **Lab 11b: Strojová analýza řeči a jazyka**

- Lingvistické analýza
 - Statistická analýza textu v MATLABu
 - Vizualizace výsledků



Od akustických analýz k jazykovým

Analýza	Souvislost s řečí a jazykem	Charakteristiky
Akusticko-fonetická	Studuje fyzikální vlastnosti hlasu a řeči	Formanty, spektrum, amplituda, trvání, spektrální analýza
Lingvistická (textová)	Zkoumá jazyk a jeho strukturu pomocí NLP	Frekvence slov, syntaktické vzory, distribuce slov, sémantika

Zatímco sonifikace pracuje s fyzickými vlastnostmi zvuku, lingvistická analýza se zaměřuje na obsah a strukturu jazyka.

Sonifikace

- **Sonifikace je proces převodu dat do zvukové podoby.**
 - Umožňuje analýzu akustických signálů sluchem.
 - Laboratorní potkani a netopýři komunikují v ultrazvukovém spektru.
 - Cílem je převést tyto signály do slyšitelného spektra.
- **Sonifikace animálních modelů umožní:**
 - Lepší porozumění vzorcům komunikace.
 - Identifikaci změn v chování animálních modelů.

Sonifikace

• Netopýři



- Ultrazvuková echolokace (**20 – 100 kHz**)
 - Manévrování v lese (velmi vysoké frekvence 50 – 80 kHz)
 - Sociální vokalizace (nižší frekvence 20 – 50 kHz)

• Laboratorní myši



- Ultrazvuková vokalizace USV (**30 – 100 kHz**)
 - Sociální interakce – krátké pulzy v přítomnosti jiných jedinců.
 - Páření – samci vydávají složité sekvence.
 - Stres a bolest – změny v délce a intenzitě zvuků.
 - Odloučení mláďat – volání mláďat na matku.

Fázový vokodér

- **Oblíbený a výkonný audio nástroj**
 - **Zvuková studia**
 - Úprava délky záznamu při zachování frekvenčních vlastností
 - Frekvenční posuny (půltóny, oktávy)
 - **Řečové technologie**
 - TTS synzéza
 - Rozpoznávače
- **Fázový vokodér jsou techniky, které:**
 - segmentují signál,
 - transformují jej do frekvenční oblasti,
 - provedou modifikace ve spektru,
 - signál rekonstruují v časové oblasti.

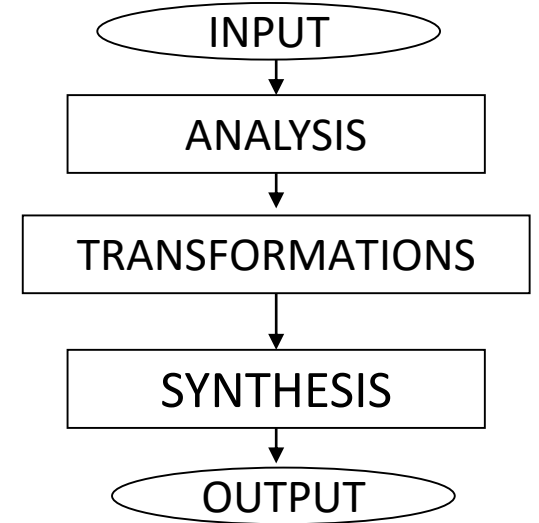
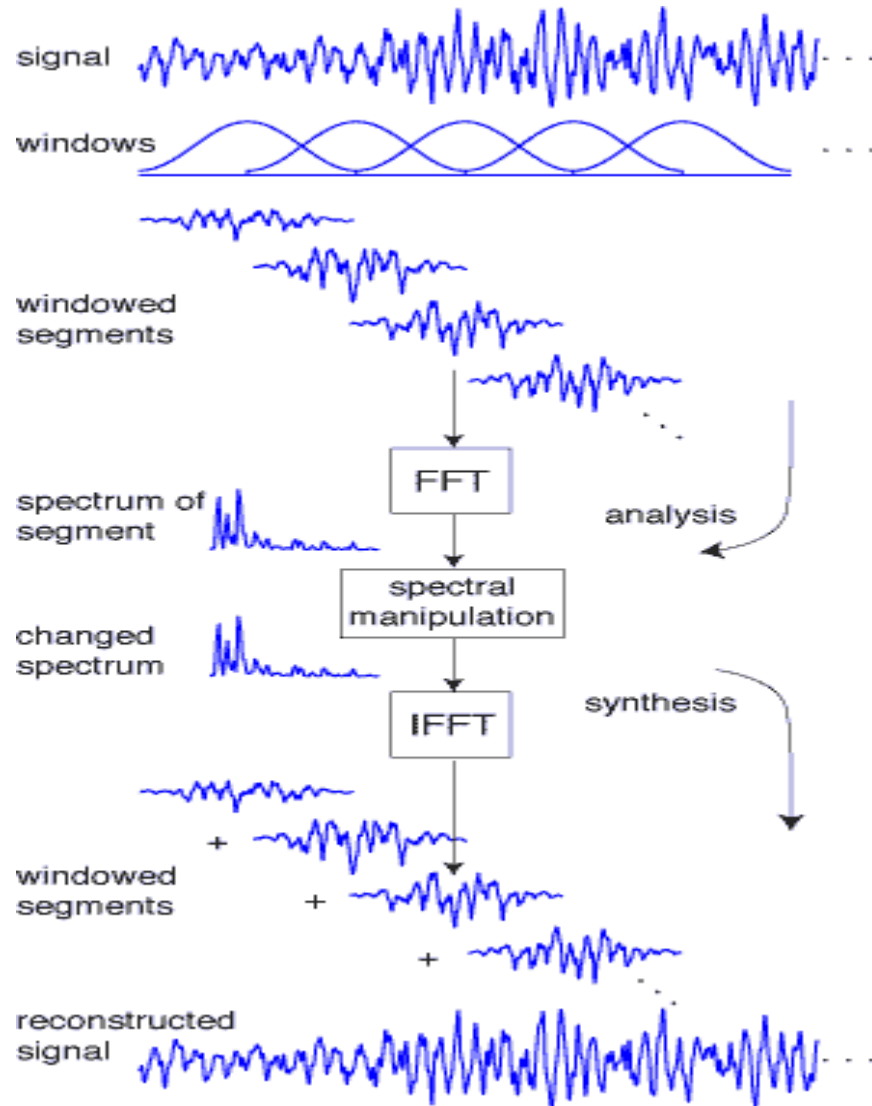
Fázový vokodér

Krátkodobá Fourierova transformace

$$X_m = \frac{1}{n} \sum_{k=0}^{n-1} x_k e^{-j \frac{2\pi mk}{n}}$$

Re-syntéza

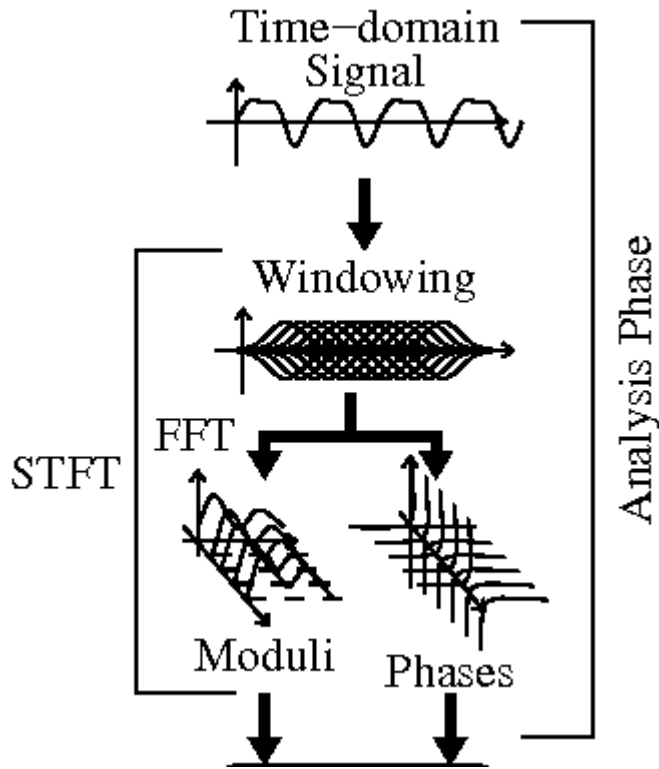
$$x_k = \frac{1}{n} \sum_{m=0}^{n-1} X_m e^{j \frac{2\pi mk}{n}}$$



Fázový vokodér – analýza

- **Délka segmentu (M):** počet analyzovaných vzorků signálu.
 - delší segment = lepší frekvenční a nižší časové rozlišení.
 - kratší segment = lepší časové a menší frekvenční rozlišení.
 - ovlivňuje dobu výpočtu (doporučená délka pro řeč je 30 .. 40 ms)
- **Typ okna:** jakékoliv nepravoúhlé (např., Hamming, Hanning, Blackman).
- **Počet vzorků FFT:** optimalizováno na 2^n (např., 1024).
 - Obvykle je velikost FFT nejbližší mocninou dvou, která je dvojnásobkem délky segmentu.
 - Např., pro délku segmentu 512, bude počet vzorků FFT roven 1024.
- **Překrytí (overlap nebo hop size):** zlomek M, např., M/8.

Fázový vokodér – analýza



```

% Spectral analysis
% (Short-Time Fourier Transform)

```

```

X = []; k=1;
for start = 0:window_shift:N-window_length,
    frame = sig(start+1:start+window_length)...
        .*hamming(window_length);
    X(:,k)=fft(frame);
    k=k+1;
end

```

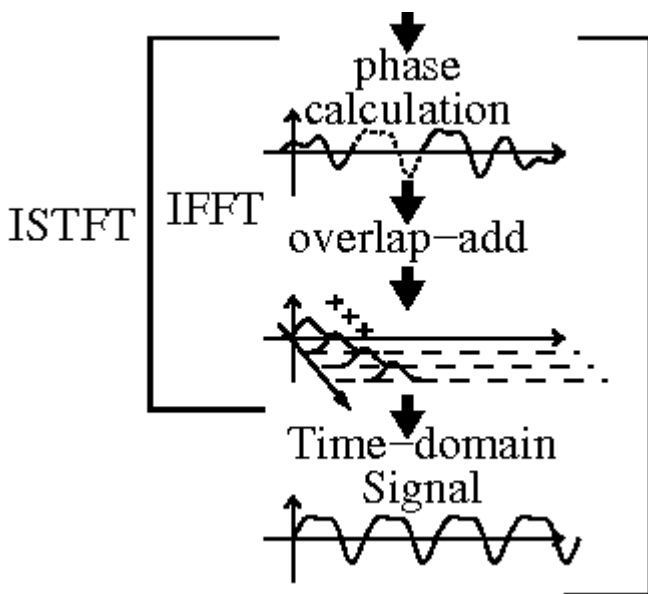
Fázový vokodér – syntéza

- **Inverzní Fourierova transformace:**

- Modifikované komplexní spektrum se transformuje zpět do časové oblasti.

- **Metoda Overlap-Add (překrývání a přidávání):**

- Jednotlivé nově transformované segmenty signálu se v časové oblasti sčítají s překrytím, čímž se rekonstruuje spojitý audio signál.
- Překrývání segmentů zajišťuje plynulé přechody a potlačuje artefakty.

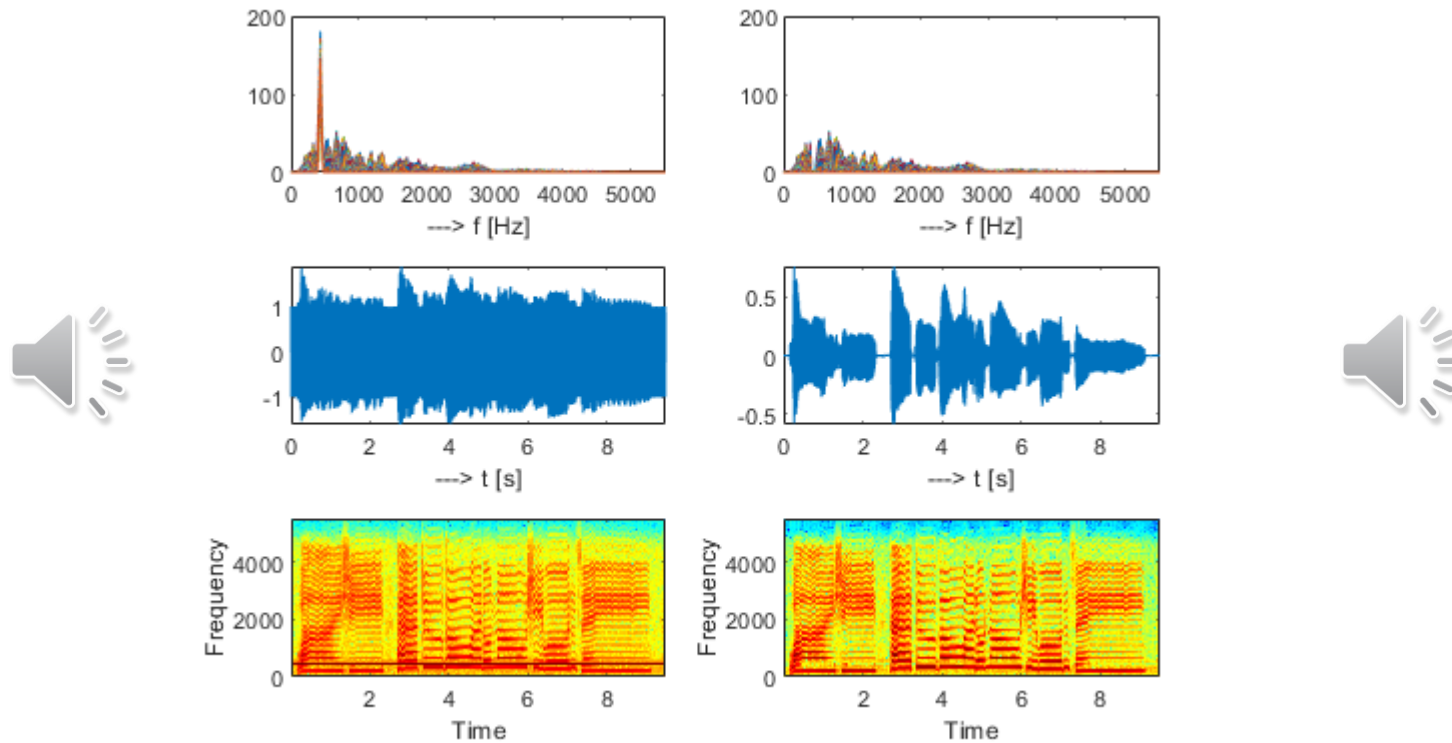


```

% Synthesis (Inverse FFT and overlap-add method)
k=1; N_new=size(Y,2)*window_shift+window_length;
signal_y=zeros(N_new,1);
for start = 0:window_shift:(size(Y,2)-1)*window_shift,
    segment=real(ifft(Y(:,k),window_length))...
        .*hamming(window_length);
    signal_y((start+1):(start+window_length))...
        =signal_y((start+1):(start+window_length))+segment;
    k=k+1;
end
  
```

Fázový vokodér – manipulace ve spektru

- Transformace: filtrace ve spektru
 - Potlačení 430 Hz



Fázový vokodér – manipulace ve spektru

• Transformace: změna délky signálu v čase

- $R = N2 / N1$ (činitel prodloužení)
- $R > 1 \rightarrow$ prodloužení (délka signálu se zvětší)
- $R < 1 \rightarrow$ zkrácení (délka signálu se zmenší)

○ Interpolace amplitudy

○ Modifikace fáze

- Nová časová mřížka
- Výpočet rozdílů fází
- Mapování rozdílů fází do nové mřížky
- Rekonstrukce fáze

○ Rekonstrukce spektra

- Kombinace amplitudy a fáze

Xmag

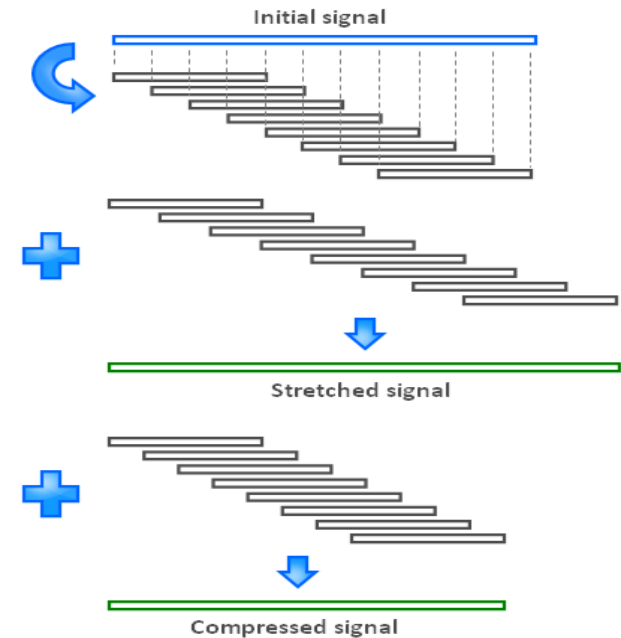
```
new_grid = floor(0:1/R:size(X, 2)-2) + 1;
```

```
D = diff(angle(X'))';
```

```
D_new = D(:, new_grid);
```

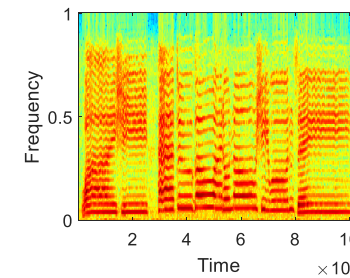
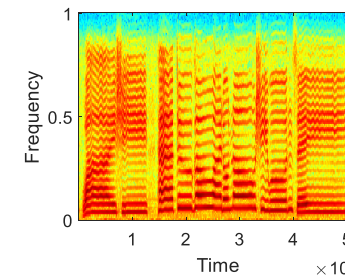
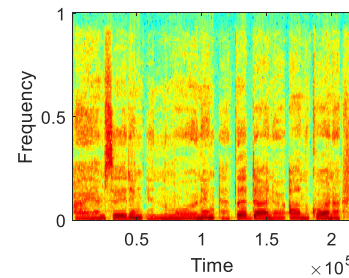
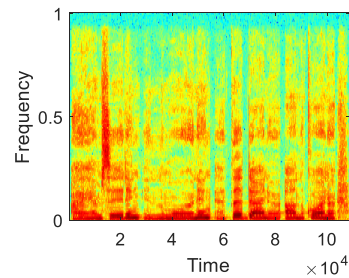
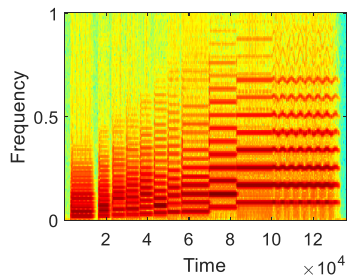
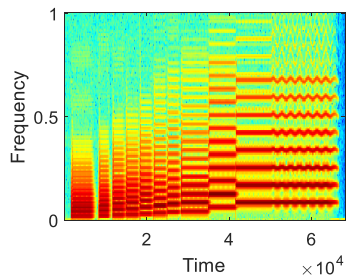
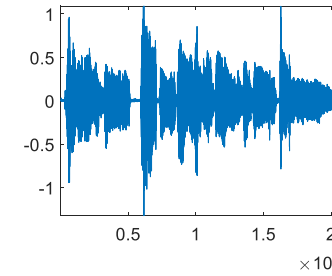
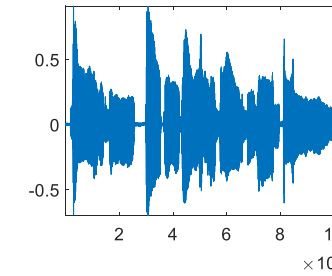
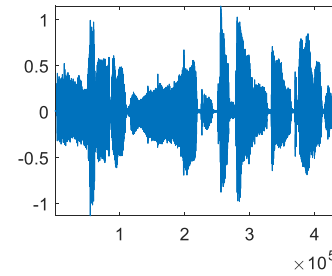
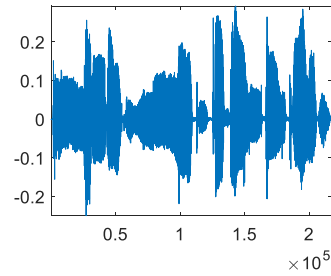
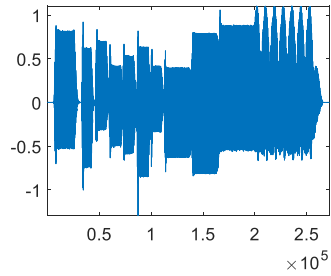
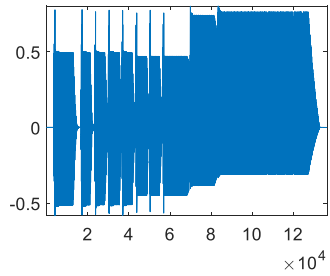
```
phaseX = cumsum(D_new)';
```

```
Y = (Xmag .* exp(j * phaseX))';
```



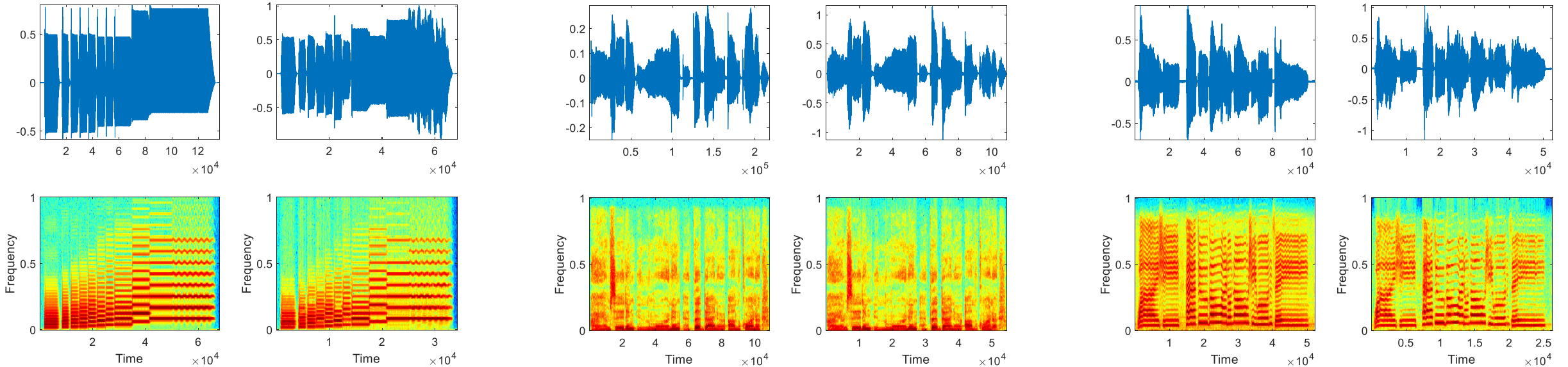
Fázový vokodér – manipulace ve spektru

- Prodloužení v čase: $R = 2$



Fázový vokodér – manipulace ve spektru

- Zkrácení v čase (komprese): $R = 0.5$



Fázový vokodér – manipulace ve spektru

• Transformace: změna výšky tónu (frekvenční transpozice)

○ Postup:

1. Prodloužení signálu v čase

- Činitel prodloužení: $R = N2 / N1$
 - $R > 1$ pro prodloužení
 - $R < 1$ pro zkrácení (kompresa)

2. Převzorkování prodlouženého signálu

- Převzorkováním s poměrem $N1 / N2$ se vrátíme k původní délce signálu, avšak s posunutou výškou tónu

○ Příklad 1:

- Posun o oktávu nahoru: $R = 2 / 1$, pak převzorkování $1 / 2$ (interpolace / decimace).
- Posun o oktávu dolů: $R = 1 / 2$, pak převzorkování $2 / 1$.

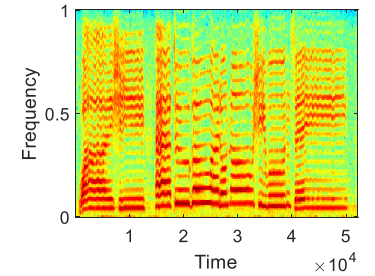
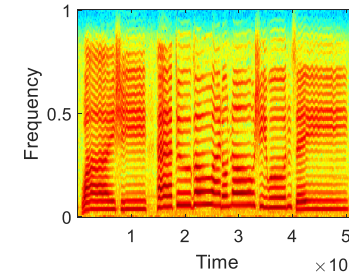
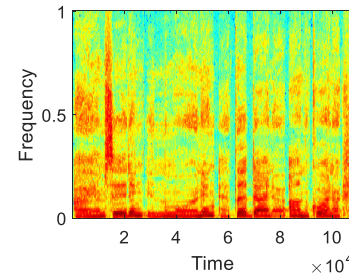
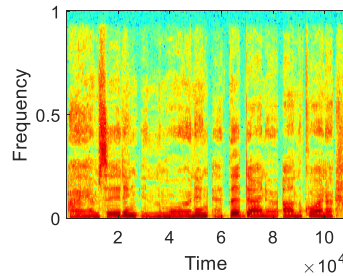
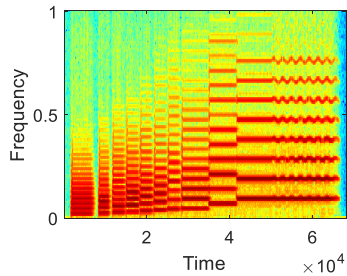
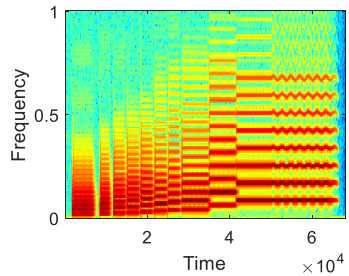
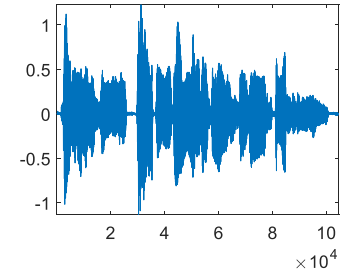
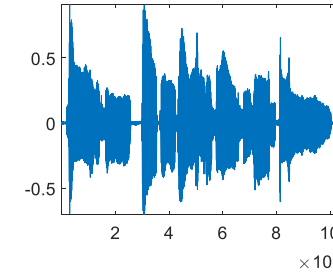
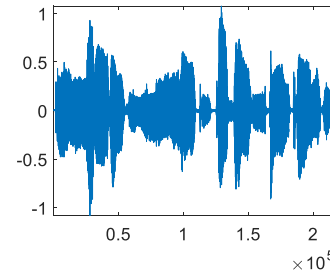
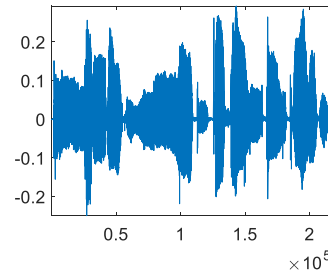
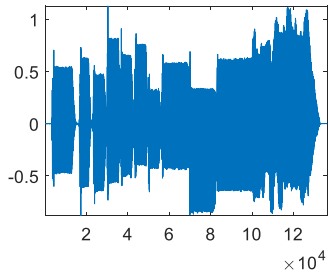
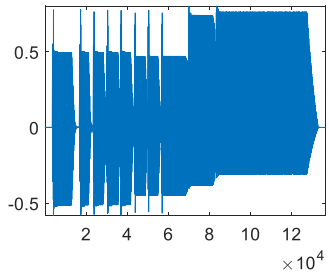
○ Příklad 2:

- Posun o půltóny ST nahoru: $ST = 12$; % Počet půltónů (12 = posun o oktávu)
 $N1 = \text{round}(\text{window_shift}); N2 = \text{round}(2^{(ST/12)} * N1);$
 $R = N2/N1; \text{resample}(y, N1, N2)$
- Posun o půltóny dolů: $ST = -12$; % Počet půltónů

$$2^{(s/12)} = \frac{\text{hop}_s}{\text{hop}_a}$$

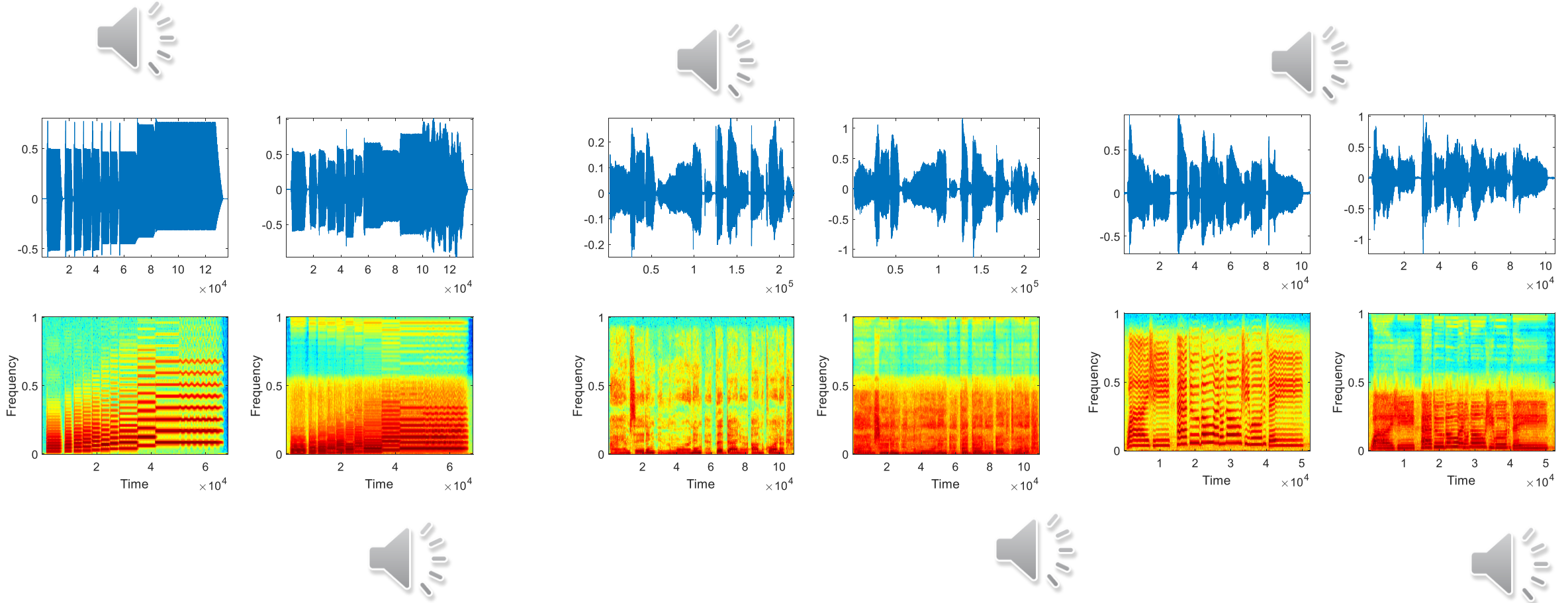
Fázový vokodér – manipulace ve spektru

- Zvýšení frekvence - $f \uparrow$



Fázový vokodér – manipulace ve spektru

- Snížení frekvence - $f \downarrow$

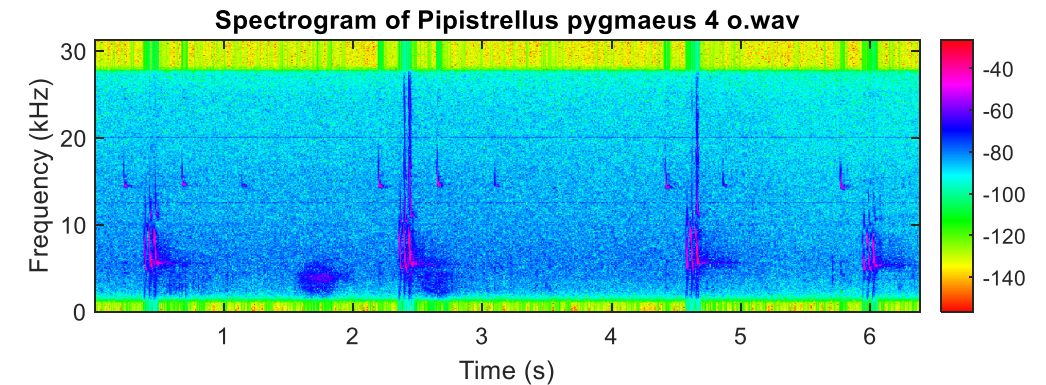
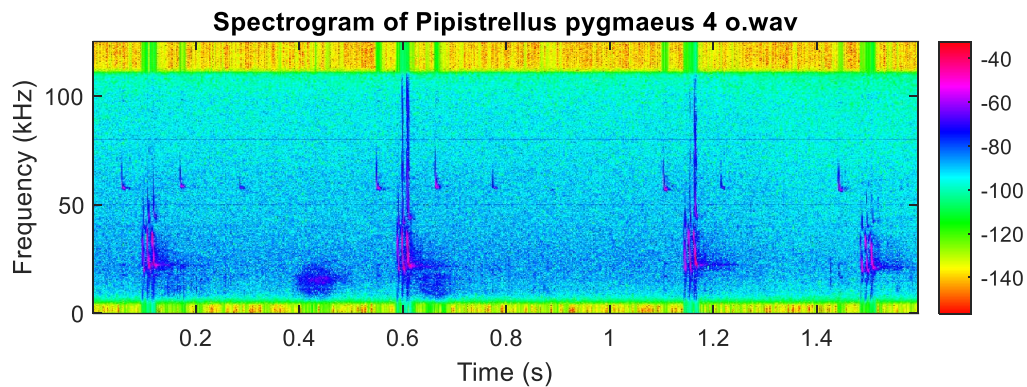
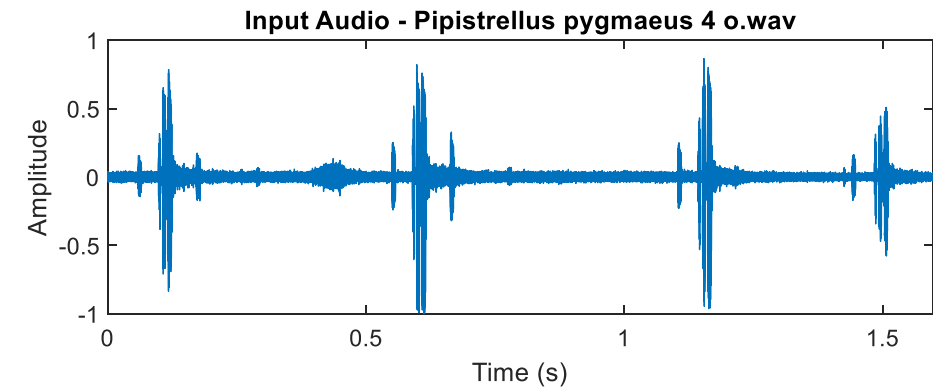
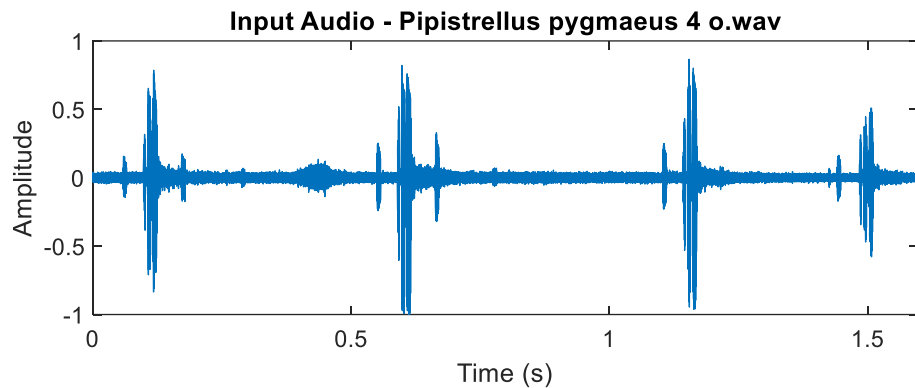


Sonifikace laboratorních zvířat

Sonifikace změnou rychlosti přehrávání

$F_s = 250 \text{ kHz}$

$F_s/4$



Vztah mezi F_s a frekvenčním obsahem

- Každé zdvojnásobení zpomalení ($F_{s_new} = F_s/2$) posune výšku tónu o 1 oktávu dolů ($ST = 12$ půltónů).

Faktor zpomalení (sF)	Posun výšky tónu (půltóny, ST)	Posun v oktávách
1× (normální rychlost)	0 půltónů	0 oktáv
2× pomaleji	-12 půltónů	-1 oktáva
4× pomaleji	-24 půltónů	-2 oktávy
8× pomaleji	-36 půltónů	-3 oktávy
16× pomaleji	-48 půltónů	-4 oktávy
32× pomaleji	-60 půltónů	-5 oktáv

$$ST = 12 \times \log_2(sF)$$

$$sF = 2^{(ST / 12)}$$

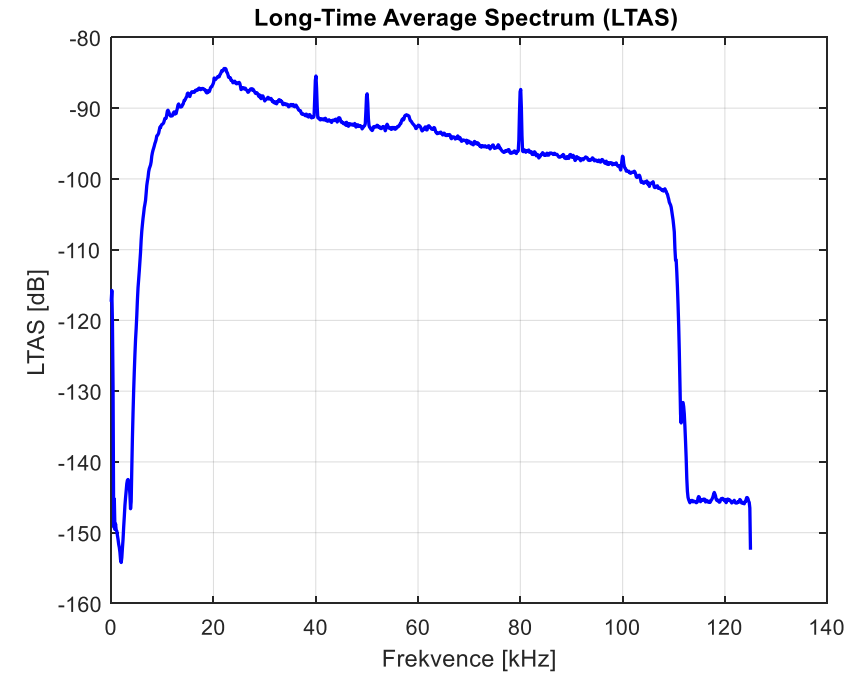
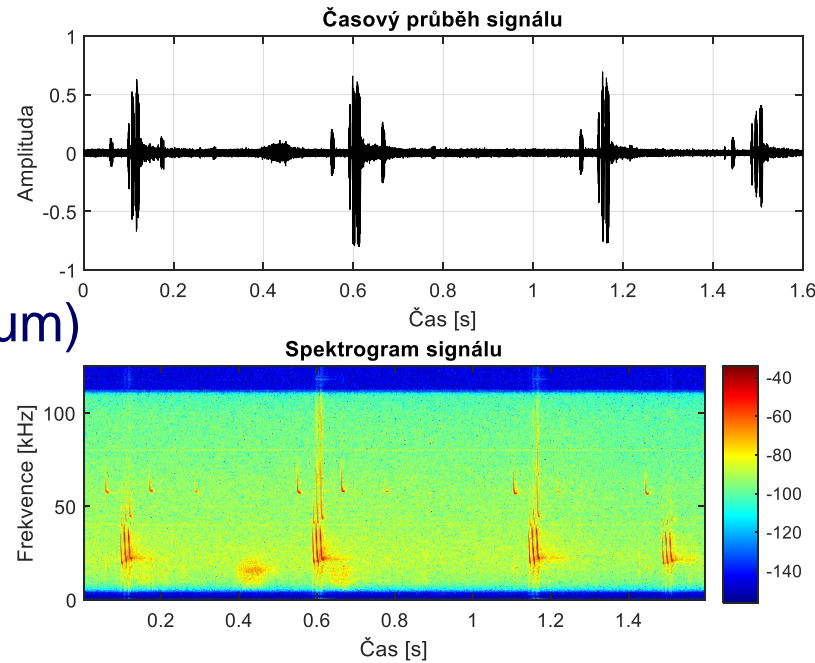
ST semitone
sF slow down factor
TS time stretching

- Netopýři ~ - 2 oktávy
- Myši ~ - 3 oktávy

Výběr parametrů sonifikace

- Výpočet:

- LTAS
(Long-Time Average Spectrum)
- F_{max} (~ -20 dB)
- $sF = F_{max} / F_{max_audio}$
- $ST = 12 \times \log_2(sF)$



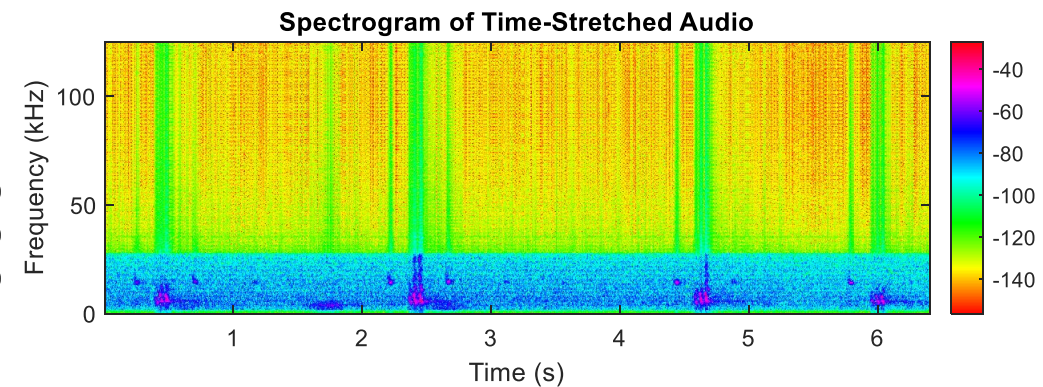
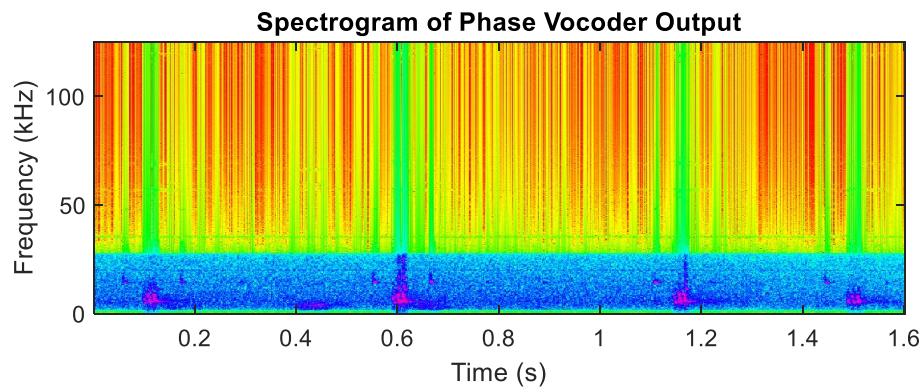
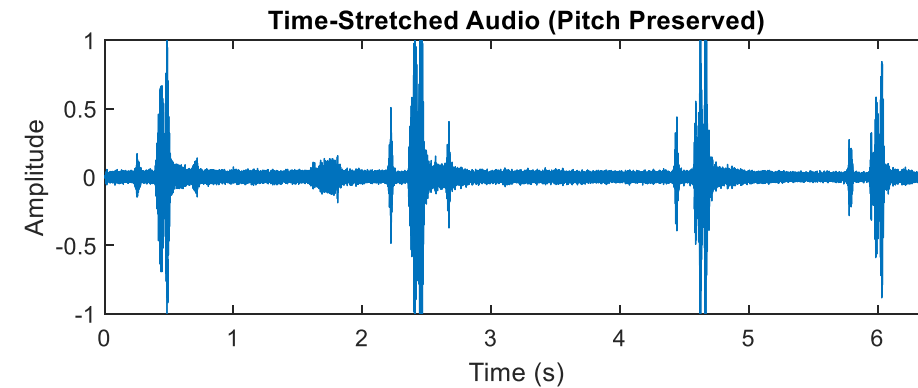
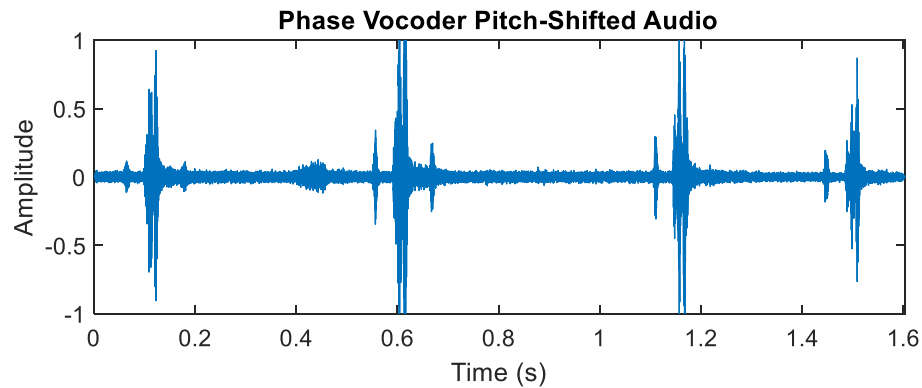
Animal	Fs (kHz)	Fmax (kHz)	Fmax_audio (kHz)	sF	ST (půltóny)
Příklad	250	80	6	13.33	41.6

Sonifikace laboratorních zvířat

Sonifikace s použitím fázového vokodéru

ST = - 24

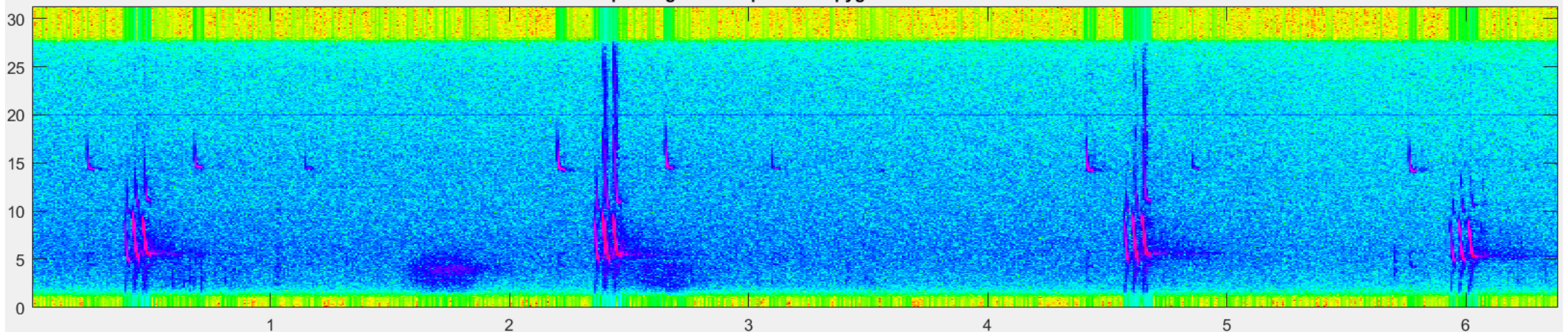
TS = 4



Sonifikace laboratorních zvířat

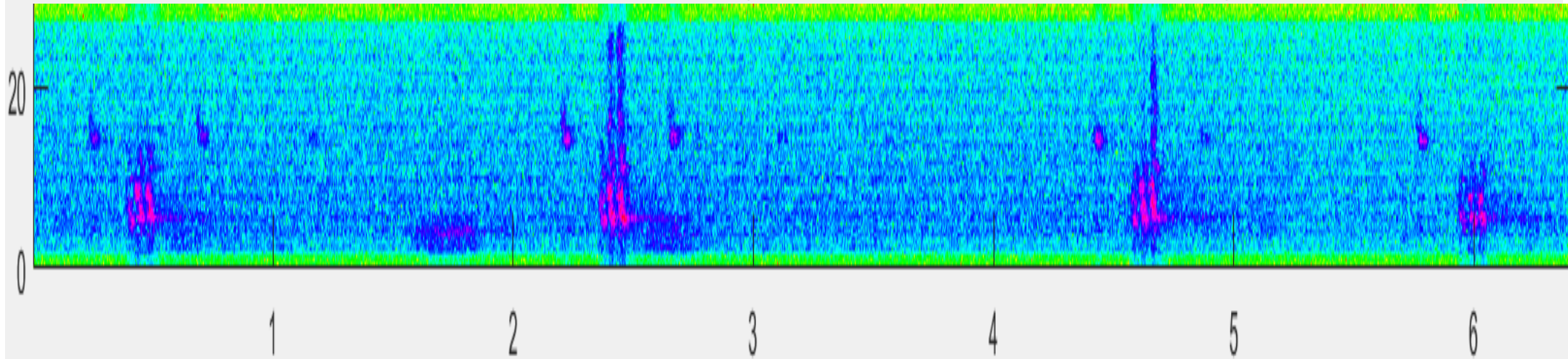
Porovnání pomalého přehrávání a stejné změny pomocí fázové vokodéry (snížení frekvencí a prodloužení)

$F_s/4$



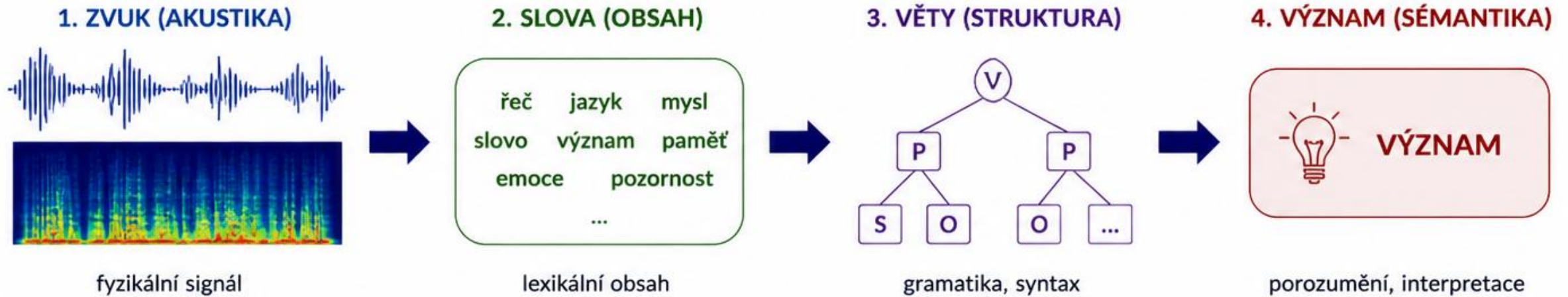
SF = 24

TS = 4



Strojová analýza řeči a jazyka

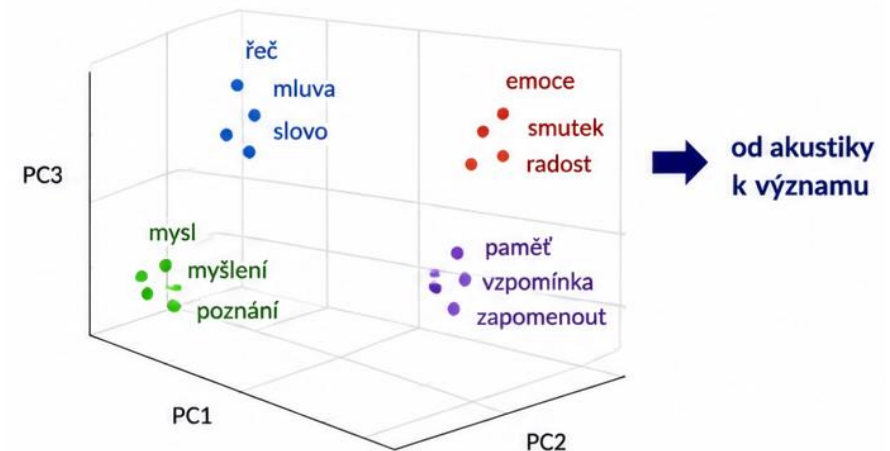
ŘEČ = komplexní signál (fyzický + jazykový)



Řeč odráží kognitivní procesy



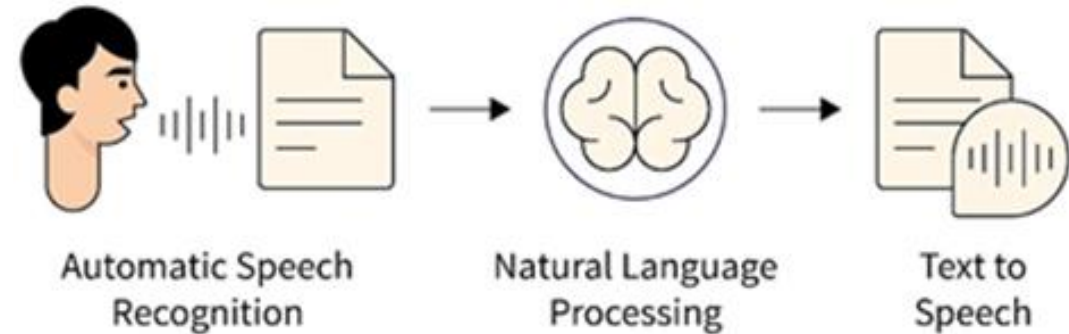
Analýza významu pomocí embeddingů



Strojová analýza řeči a jazyka

1 Převod řeči na text (Speech-to-Text)

- Automatické nástroje: Google Speech2Text, Speechnotes, Notta, Descript
- Využití akustických modelů (HMM-GMM, DNN, CNN-TDNN)
- Problémy: akcenty, šum, homofony



2 Lemmatizace – normalizace slov, převedení na infinitiv

- Použití MorFlex a dalších slovníků

3 Korpusy pro analýzu jazyka

- Český národní korpus – rozsáhlá databáze autentických textů
- Pomáhá trénovat a testovat NLP modely



4 Prozodická analýza

- Studuje tempo, intonaci, rytmus řeči

5 Lingvistická analýza

- nástroje pro statistickou analýzu v MATLABu – viz další text

Jak funguje převod řeči na text?

Hlavní kroky převodu řeči na text:

1 Zvukový signál

- Záznam mluvené řeči → převedení na digitální formu
- Časté reprezentace: MFCC, FFT analýza

2 Akustický model

- Převod zvuku na fonémy (HMM, DNN, CNN-TDNN)
 - Nejnovější modely: Whisper, Wav2Vec

3 Jazykový model

- Zohledňuje pravděpodobné posloupnosti slov
- Pomáhá eliminovat chyby přepisu pomocí korpusů

4 Textový výstup

- Finální přepis, syntaktické analýzy

Problémy převodu řeči na text:

- *Různé akcenty a dialekty*
- *Pozadí šumu ovlivňuje přesnost*
- *Homofony (stejná výslovnost, různý význam)*
- *Spontánní řeč (pauzy, opakování, slang)*

Úvod do lingvistické analýzy textu

Co je NLP a proč ho potřebujeme?

- Natural Language Processing (NLP) umožňuje automatickou analýzu psaného i mluveného textu.

- **Hlavní metody NLP:**

Metoda	Co zkoumá?	Příklad aplikace
Statistická analýza	Četnost slov, délka slov	Wordcloud, histogram
Syntaktická analýza	Gramatická struktura textu	POS tagging, syntaktické stromy
Sémantická analýza	Význam slov v kontextu	Překladače, analýza sentimentu
Pragmatická analýza	Význam na základě situace	Chatboti (GPT, Alexa, Siri)

Cíl analýzy textu:

- Předzpracování textu pro další analýzy (NLP, entropie, sémantika).
- Vyčištění textu od stopwords a interpunkce.

• MATLAB kód:

- 1 Načte textový soubor
- 2 Převeďte na malá písmena a odstraní interpunkci
- 3 Rozdělí text na jednotlivá slova a odstraní stopwords

• Výstupy MATLAB analýzy:

- Wordcloud nejčastějších slov
- Histogram délky slov
- Výpočet entropie textu

Statistická analýza textu v MATLABu

% Načtení textového souboru

```
filename = 'text_new.txt';  
fid = fopen(filename, 'r');  
textData = fscanf(fid, '%c');  
fclose(fid);
```

% Úprava textu (malá písmena, odstranění interpunkce)

```
textData = lower(textData);  
textData = regexprep(textData, '[^\w\s]', '');
```

% Rozdělení textu na slova

```
words = strsplit(textData);  
words = words(~cellfun('isempty', words));
```


Entropie textu – jak moc je text předvídatelný?

Shannonova entropie měří informační obsah textu

$$H(\text{text}) = - \sum_{\text{word} \in \text{text}} \frac{\text{freq}(\text{word})}{\text{len}(\text{text})} \log_2 \left(\frac{\text{freq}(\text{word})}{\text{len}(\text{text})} \right)$$

- **Nízká entropie** (předvídatelný text)
 - Dnes je pondělí. Zítra bude úterý. Pak přijde středa.
- **Vysoká entropie** (složitý nebo chaotický text)
 - G7bxk vfwosd ljqp m ktuzh cbxnr

Entropie textu 1: 6.5410

Entropie textu 2: 7.8304

Podmíněná entropie – predikce slov

- Odhaduje pravděpodobnost výskytu slova na základě předchozích slov
- Bigramy – modely předpovídají další slovo podle předchozího

Nízká entropie

- Mám rád jablka a hrušky

Vysoká entropie

- Rychlý kluk sní světlo obloha

Podmíněná entropie
pro kontext 1: 0.8991

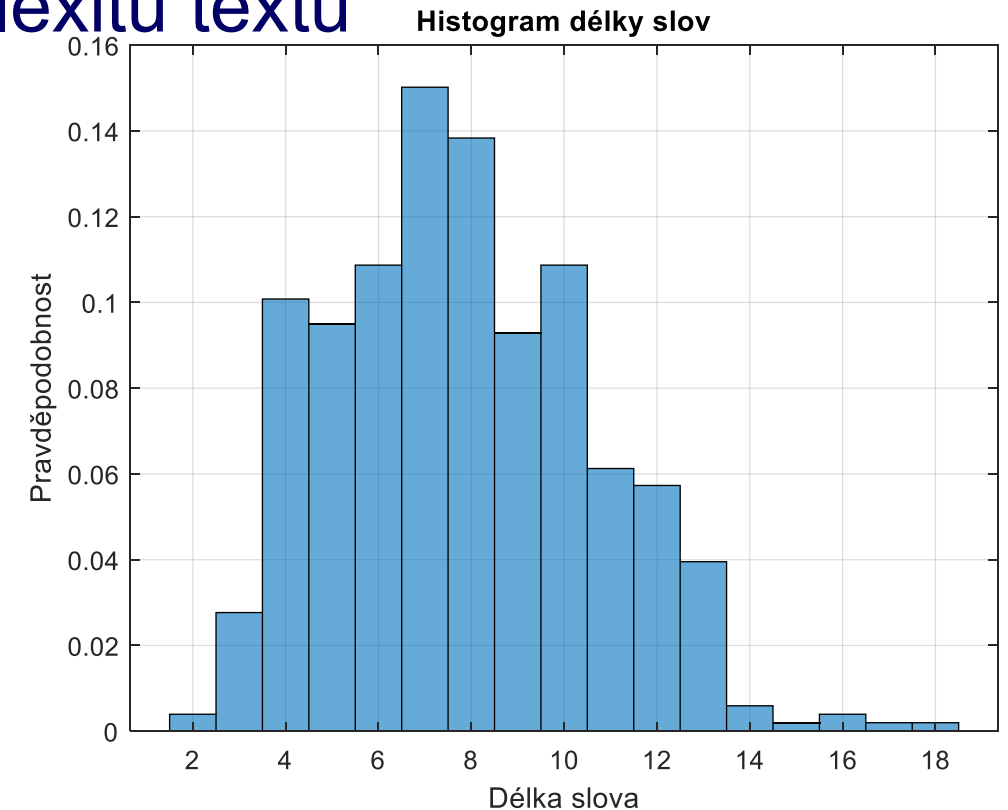
Podmíněná entropie
pro kontext 2: 1.2713

$$H_{\text{cond}}(\text{text}) = - \sum_{(c,w) \in \text{text}} \frac{\text{freq}(c, w)}{\text{len}(\text{text})} \log_2 \left(\frac{\text{freq}(c, w)}{\text{freq}(c)} \right)$$

of times word w occurs after context c (points to $\text{freq}(c, w)$)
 context (preceding n-gram) (points to c)
 word (points to w)
 means # of n-grams here (not just words) (points to $\text{len}(\text{text})$)
 total # of times context c occurs, with or without word w (points to $\text{freq}(c)$)

Vizualizace výsledků

- Používá se v NLP a textové analýze k rychlému zjištění klíčových informací
- Pomáhá lépe pochopit strukturu a komplexitu textu
 - **Histogram délky slov:**
Jak dlouhá slova jsou nejčastější?
 - **Wordcloud:**
Jaká slova dominují v textu?
 - **Entropie:**
Jak entropie souvisí s předvídatelností textu?



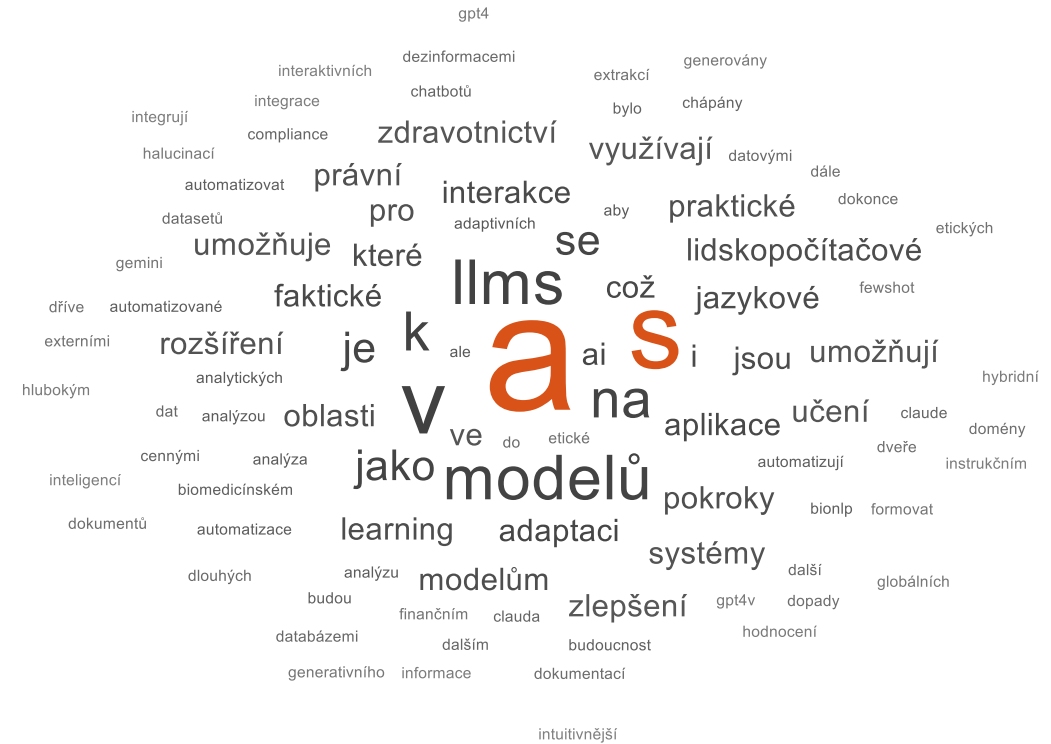
Vizualizace výsledků

- Porovnání českého lemmatizovaného (vlevo) a neupraveného (vpravo) textu

Wordcloud textu B



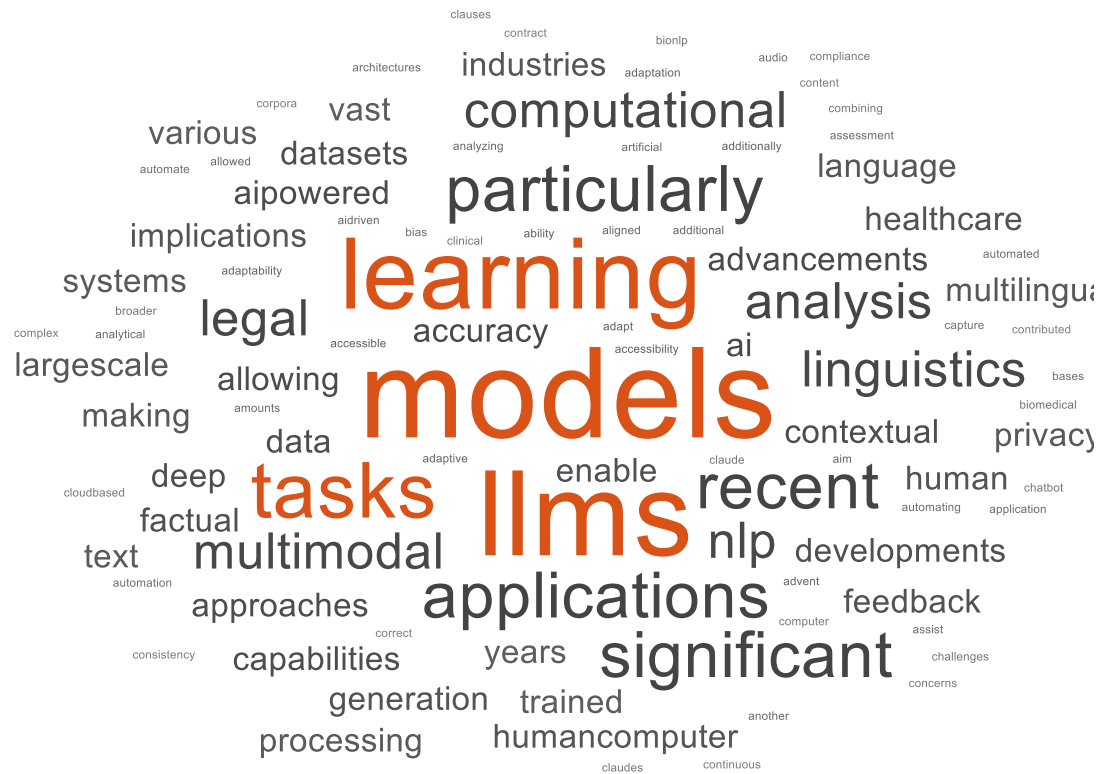
Wordcloud textu B



Vizualizace výsledků

- Porovnání anglického (vlevo) a českého (vpravo) textu

Wordcloud textu B bez stopwords

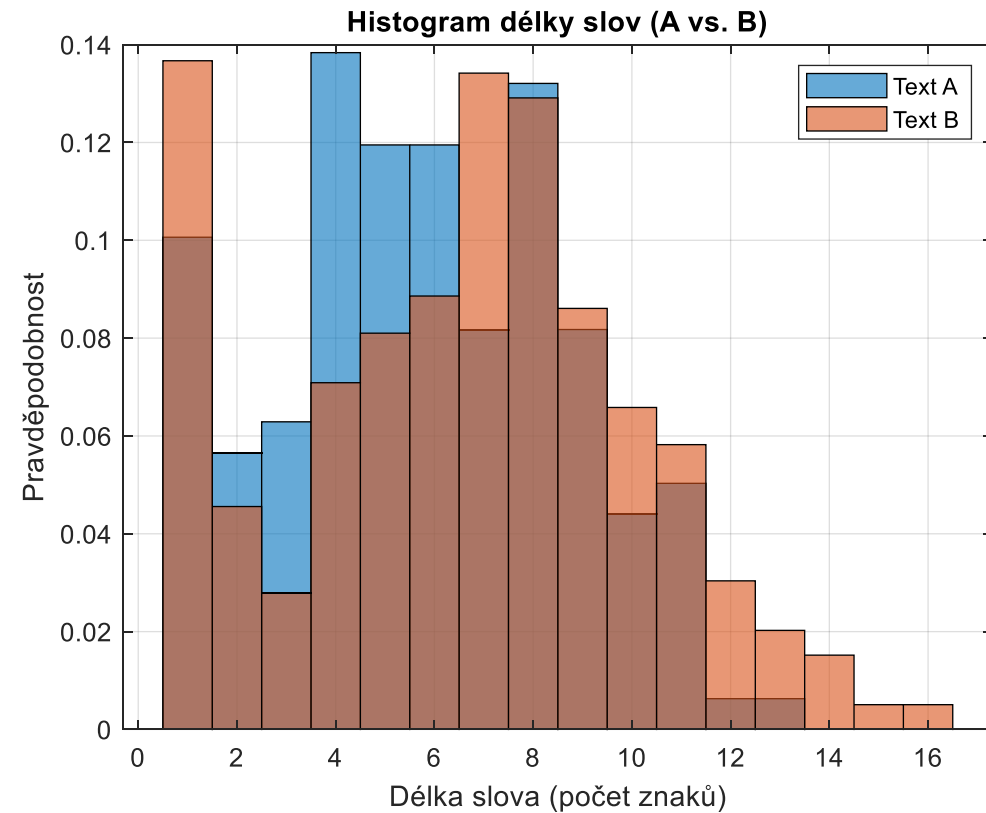
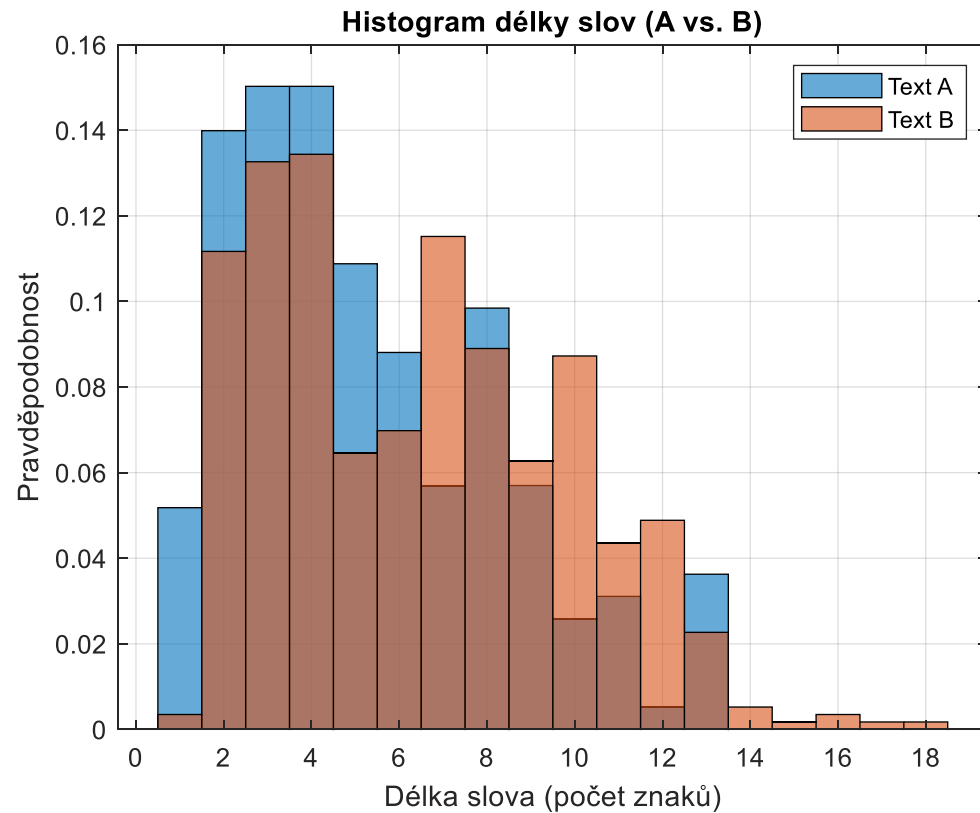


Wordcloud textu B bez stopwords



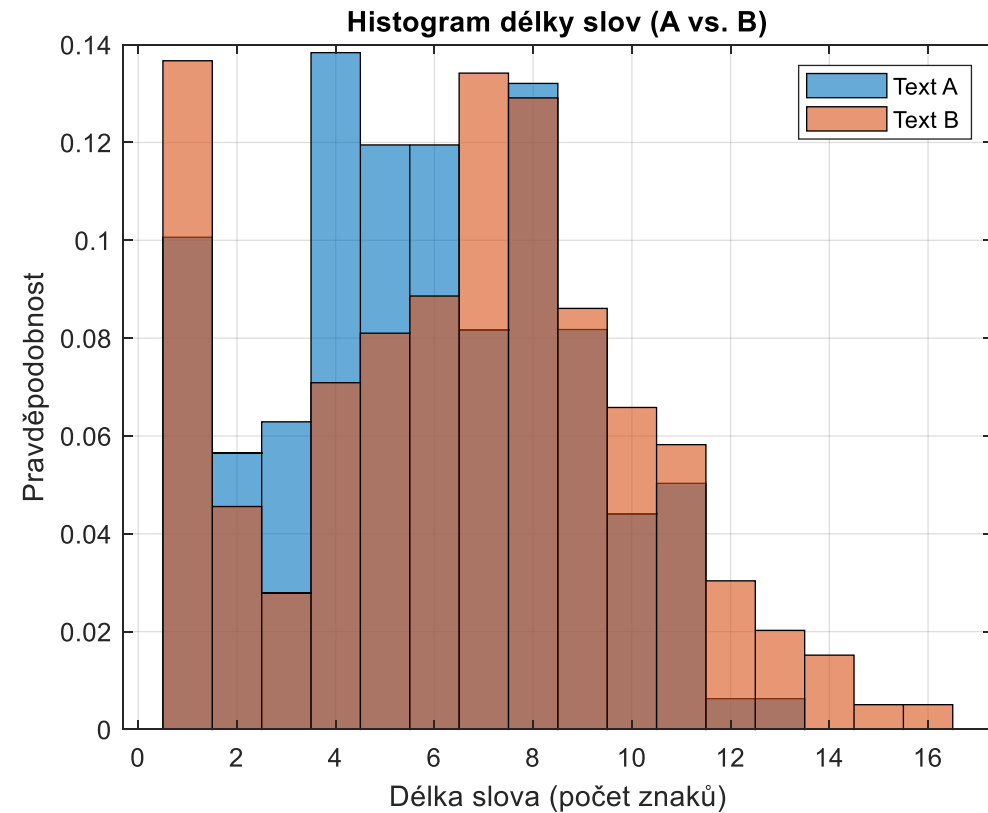
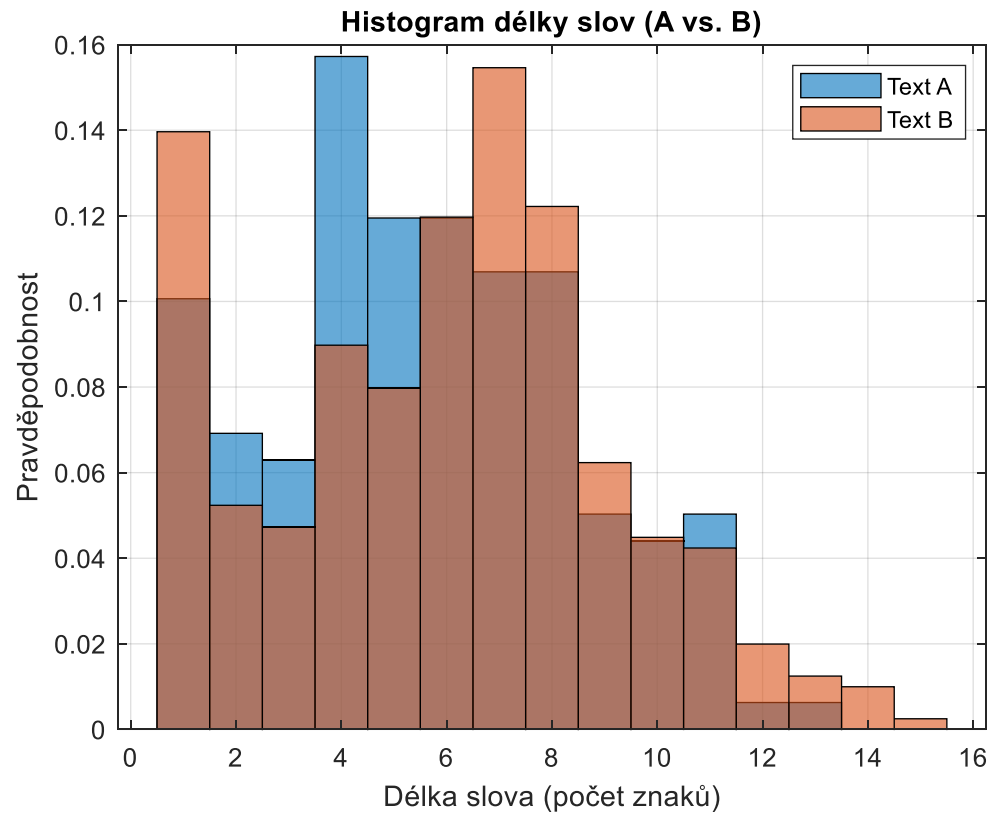
Vizualizace výsledků

- Porovnání anglického (vlevo) a českého (vpravo) textu



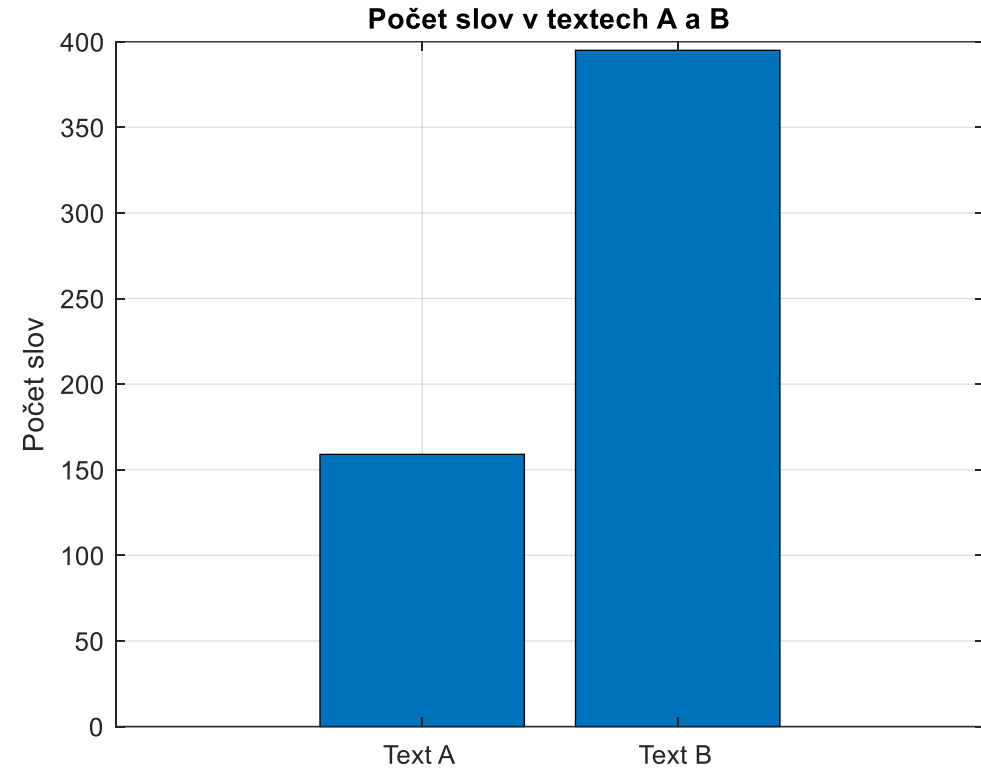
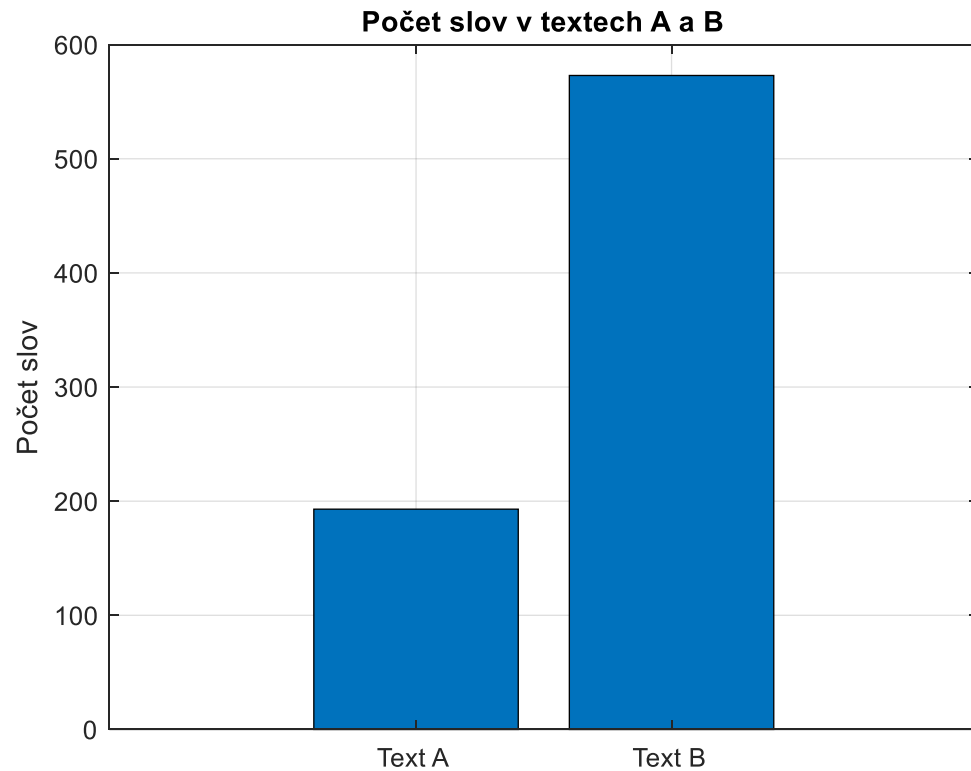
Vizualizace výsledků

- Porovnání českého lemmatizovaného (vlevo) a neupraveného (vpravo) textu



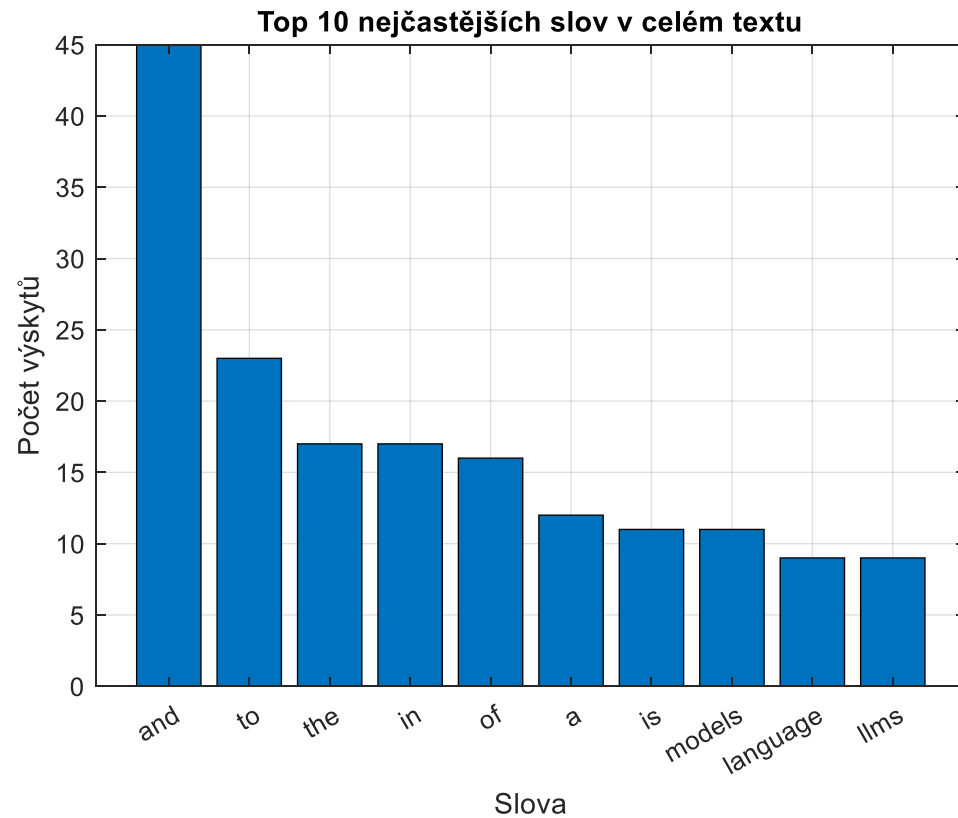
Vizualizace výsledků

- Porovnání anglického (vlevo) a českého (vpravo) textu



Vizualizace výsledků

- Porovnání anglického (vlevo) a českého (vpravo) textu



Vizualizace výsledků

- Porovnání českého lemmatizovaného (vlevo) a neupraveného (vpravo) textu



Výsledky

eng

Počet odstavců v textu: 11 s textem + 1 bez textu

Entropie textu A: 6.5410

Entropie textu B: 7.8304

Podmíněná entropie pro kontext v textu A: 0.8991

Podmíněná entropie pro kontext v textu B: 1.2713

Počet slov v textu A: 193

Počet slov v textu B: 573

Průměrná délka slov v textu A: 5.29

Průměrná délka slov v textu B: 6.44

cz

Počet odstavců v textu: 12

Entropie textu A: 6.7041

Entropie textu B: 7.8631

Podmíněná entropie pro kontext v textu A: 0.5382

Podmíněná entropie pro kontext v textu B: 0.7392

cz_lemm

Počet odstavců v textu: 11

Entropie textu A: 6.6664

Entropie textu B: 7.8644

Podmíněná entropie pro kontext v textu A: 0.5762

Podmíněná entropie pro kontext v textu B: 0.7500

Sémantická reprezentace textu

Jak poznáme, že jsou si texty významově podobné?

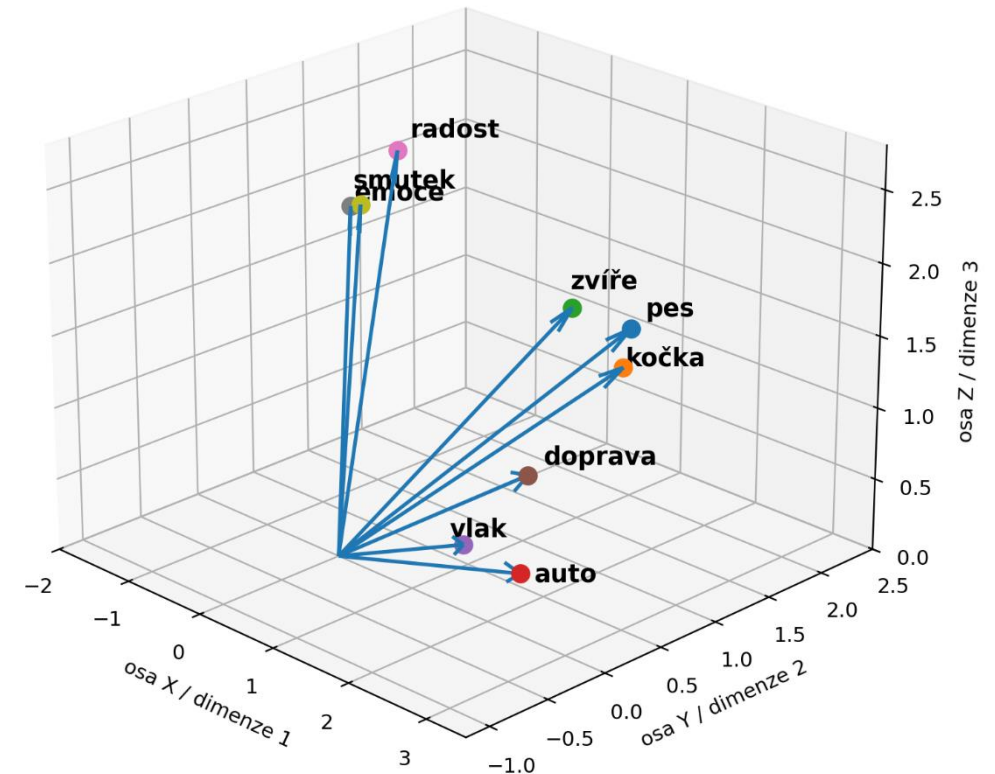
text → vektor (embedding)

podobné texty → blízké vektory

Zjednodušený embeddingový prostor

text → numerická reprezentace →
→ redukce dimenze → clustering →
→ vizualizace

„Pes běží“ \approx „Zvíře utíká“
„Pes běží“ \neq „Dnes prší“



Cíl analýzy textu:

- Sémantická analýza textu
- Vizualizace výsledků

• MATLAB kód:

- 1 Načte textový soubor a model s embeddingy
- 2 Převéde slova/výrazy na vektory
- 3 Vykreslí PCA zdrojových výrazů.

• Výstupy MATLAB analýzy:

- PCA graf s kmeans clustery
- Interpretace výsledků

Vizualizace výsledků

- PCA graf zdrojových výrazů s clustery zdrojových textů

ID	skupina	text
T01	narativ	Ráno jsem vstal, udělal si snídani a odešel do práce. Dopoled
T02	narativ	Včera jsem šla na procházku do lesa. Cestou jsem posloucha
T03	narativ	O víkendu jsme navštívili přátele. Nejprve jsme si dali oběd a j
T04	narativ	Student si přečetl článek a poté odpověděl na otázky. Většíně
T05	narativ	Večer jsem si pustil film a připravil si večeři. Po filmu jsem ješt
T06	odborne	Fotosyntéza je proces, při kterém rostliny využívají světelnou r
T07	odborne	Elektrický proud vzniká pohybem nabitých částic. Jeho veliko
T08	odborne	Buňka obsahuje organely, které zajišťují základní životní funkce
T09	odborne	Zvuk je mechanické vlnění, které se šíří prostředím a je vnímá
T10	odborne	Programovací jazyk umožňuje zapisovat algoritmy a řešit složi
T11	narusena_koherece	Ráno bylo otevřené a stůl čekal na práci, která přišla později r
T12	narusena_koherece	Procházka šla kolem myšlenky, která byla hlasitější než cesta.
T13	narusena_koherece	Článek odpověděl na otázky dříve, než student pochopil dveř
T14	narusena_koherece	Večeře sledovala film, zatímco kniha připravovala večer.
T15	narusena_koherece	Les poslouchal hudbu a kroky se dívaly na ticho mezi stromy.
T16	repetitivni	Ráno jsem vstal a šel do práce. Ráno jsem vstal a šel do prác
T17	repetitivni	Dnes je hezký den. Dnes je hezký den. Dnes je hezký den.
T18	repetitivni	Čtu knihu a čtu knihu a čtu knihu. Pořád čtu knihu a čtu knihu
T19	repetitivni	Jdu ven a jdu ven a jdu ven. Pak jdu ven znovu.
T20	repetitivni	Píšu text a píšu text a píšu text. Pořád píšu stejný text.

