

Univerzita Jana Evangelisty Purkyně v Ústí nad Labem

Přírodovědecká fakulta

Katedra informatiky

## **II. seminární práce – Počítačové zpracování signálu (KI/PZS)**

### **Klasifikace zvukových záznamů**

Meloshyn Serhii st99554 f22638

## 1. Zadání

Cílem práce je načíst hlasové záznamy písmene „a“ z databáze VOICED (PhysioNet) a pomocí metod časové a frekvenční analýzy provést klasifikaci záznamů na zdravé a patologické. U patologických záznamů je dále cílem rozlišit konkrétní typ poruchy dle anotací v hlavičkových souborech a doprovodných souborech databáze. V rámci řešení jsou využity techniky probírané v kurzu, zejména FFT a kepstrální analýza.

## 2. Data a použité nástroje

Datový zdroj: VOICED Database (PhysioNet), 208 záznamů samohlásky „a“.

Formát dat: standard PhysioNet (hlavičkový soubor \*.hea, signální data, doplňující soubor \*-info.txt).

Použité knihovny v Pythonu:

- wfdb – pouze pro načítání záznamů (bez předpřipraveného zpracování signálu)
- numpy, pandas – práce s daty
- scipy (signal) – detekce píků ve spektru
- matplotlib – vizualizace průběhů a spekter

## 3. Postup řešení

### 3.1 Načtení signálu a diagnózy

Pro každý záznam se načte hlavička a signálová data. Diagnóza je převzata z doplňujícího souboru \*-info.txt (řádek obsahující klíčové slovo „Diagnosis“). Následně je záznam zařazen do jedné ze skupin (healthy, hyperkinetic dysphonia, hypokinetic dysphonia, reflux laryngitis).

### 3.2 Předzpracování

Signál je převeden do časové osy pomocí vzorkovací frekvence fs. Před analýzou je možné aplikovat základní normalizaci amplitudy a odstranění stejnosměrné složky (DC offset). Pro frekvenční analýzu je vhodné použít okénkování (např. Hann) pro omezení spektrálního prosakování.

### 3.3 Časová analýza

V časové oblasti je vizualizován průběh signálu (amplituda v čase). Zdravé vs. patologické záznamy se mohou lišit stabilitou amplitudy, mírou šumu a pravidelností periodického kmitání hlasivek.

### 3.4 Frekvenční analýza (FFT)

Pro každý signál je vypočteno amplitudové spektrum pomocí rychlé Fourierovy transformace (FFT). Zpracovány jsou pouze kladné frekvence. Dominantní píky odpovídají základní frekvenci a jejím harmonickým složkám.

### 3.5 Odhad základní frekvence (F0) a píků

Pomocí detekce lokálních maxim ve spektru (scipy.signal.find\_peaks) se určí dominantní píky. Základní frekvence F0 je určena jako první významný pík. Dále se hodnotí pravidelnost harmonických (průměrná vzdálenost mezi píky) a variabilita (směrodatná odchylka vzdáleností).

### 3.6 Kepstrální analýza

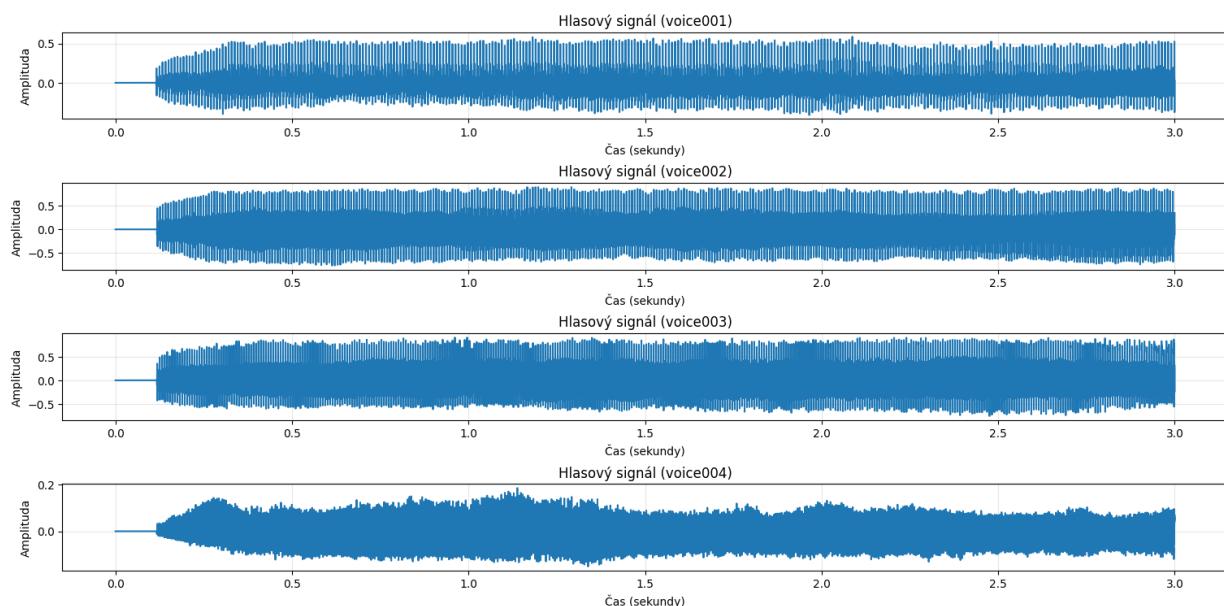
Kepstrum je vypočteno jako inverzní FFT z logaritmu amplitudového spektra. V kepstru se hledá významný vrchol odpovídající periodicitě (quefrency), který souvisí s periodou hlasového signálu. Jako ukazatel pravidelnosti se používá CPP (cepstral peak prominence) nebo výška hlavního kepstrálního vrcholu v daném intervalu.

### 3.7 Klasifikace

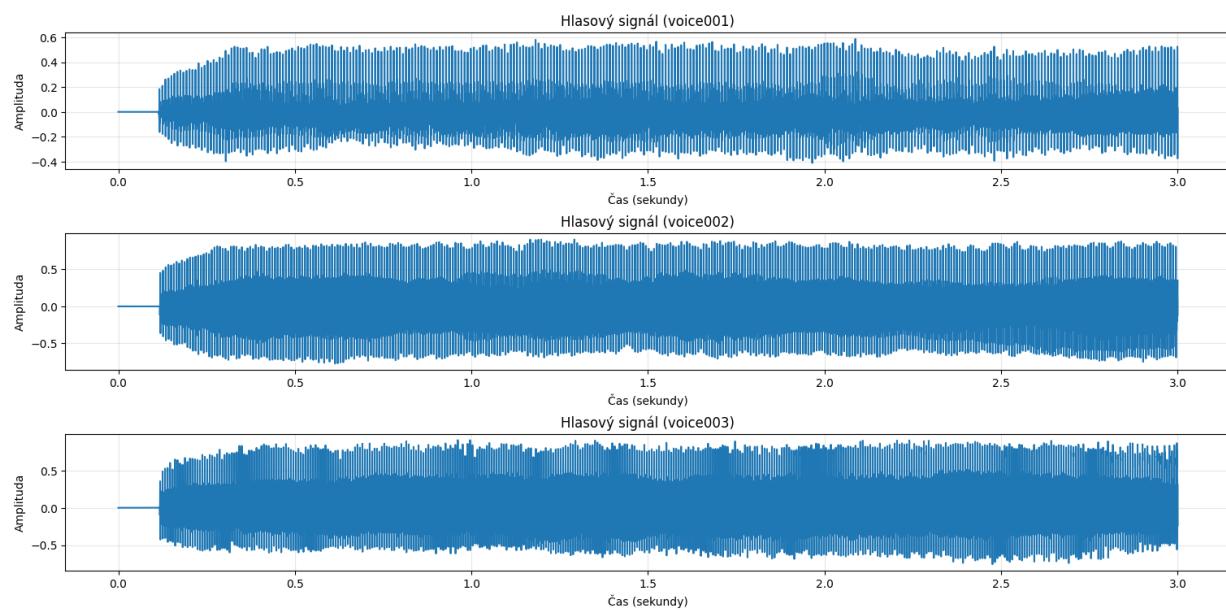
Klasifikace je provedena ve dvou krocích: (1) zdravý vs. patologický, (2) při patologii určení typu poruchy. Pro rozhodování se používají extrahované příznaky z FFT a kepstra (např. F0, pravidelnost harmonických, CPP). V základní verzi lze použít pravidlový klasifikátor (prahy nad/ pod zvolenými hodnotami) nebo jednoduchý model (např. k-NN) nad příznakovým vektorem.

## 4. Výsledky a grafické výstupy

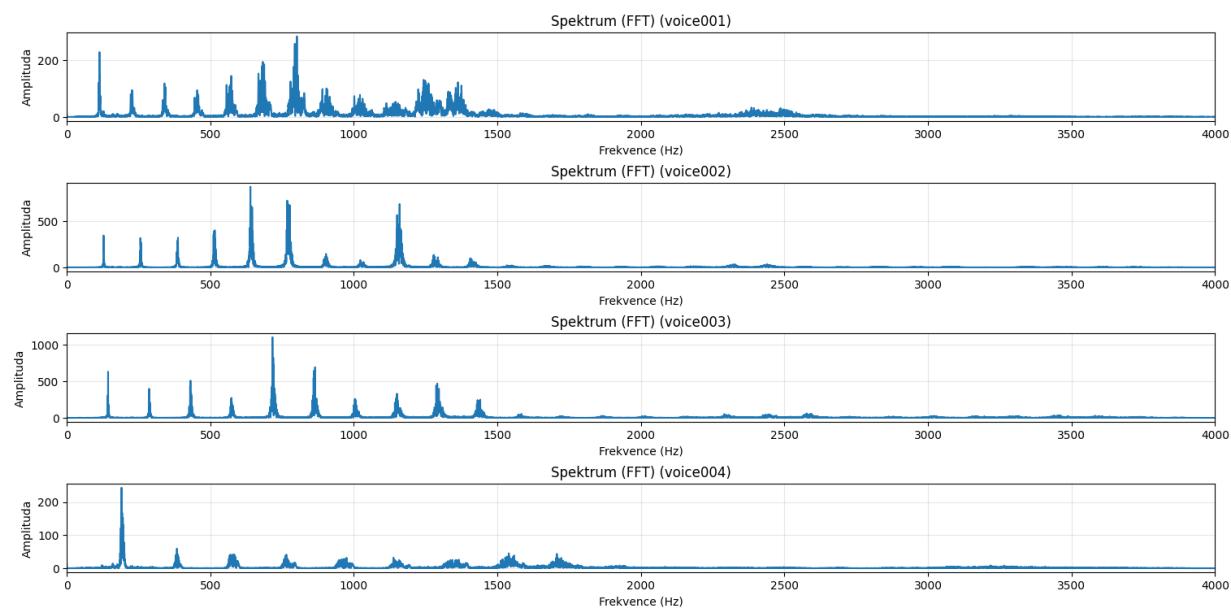
Níže jsou uvedeny grafy vytvořené během zpracování (časový průběh, spektrum a kepstrum) pro vybrané záznamy voice001–voice004.



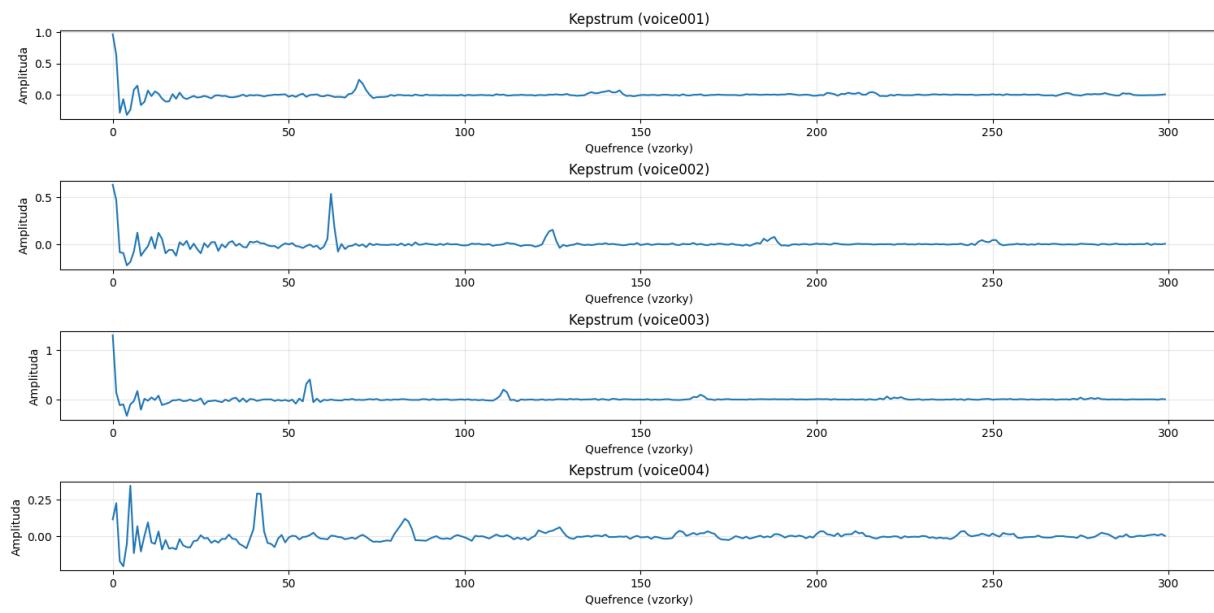
Obrázek 1: Časové průběhy vybraných hlasových záznamů (voice001–voice004).



Obrázek 2: Detailní časové průběhy (voice001–voice003).



Obrázek 3: Amplitudové spektrum (FFT) pro vybrané záznamy (voice001–voice004).



*Obrázek 4: Kepstrum pro vybrané záznamy (voice001–voice004).*