

Univerzita Jana Evangelisty Purkyně, Ústí nad Labem

Počítačové zpracování signálu KI/PZS

Seminární práce

Klasifikace zvukových záznamů

ZS 2024/25

Martin Kučera

Osobní číslo: F22125

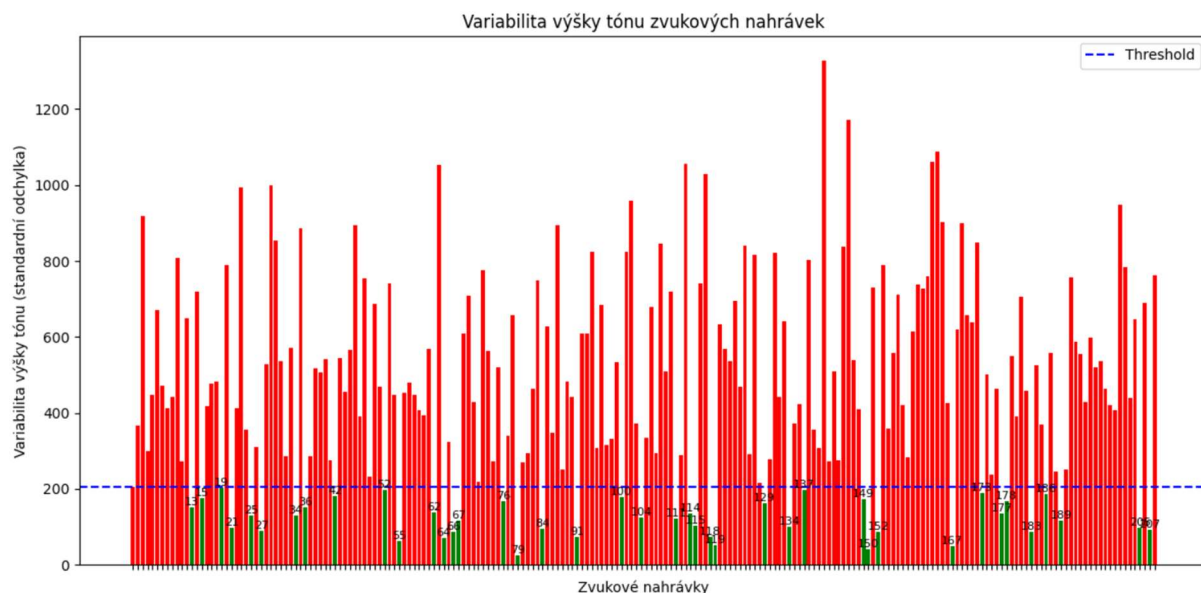
Zadání

Ve zdrojové databázi najdete celkem 208 hlasových záznamů písmene a. Pomocí Vámi vybrané techniky v časové nebo frekvenční oblasti klasifikujte zvukové záznamy na dobré a patologické. V případě patologických poté klasifikujte jednotlivé poruchy. Jejich výčet najdete buď v hlavičkových souborech nebo v propisu databáze. Pro klasifikaci do jednotlivých skupin použijte veškeré techniky, které jste si v rámci kurzu osvojili včetně Fourierovy a kepstrální analýzy. Úspěšnost Vašeho postupu porovnejte s anotacemi, resp. rozřazením do skupin, které provedli experti, kteří data pořizovali.

Postup

Před samotným načtením audio nahrávek jsem si poslechl vadný a zdravý audio záznam. Klasickým projevem patogenního záznamu pro mě byl rozdílný pitch, tedy změna tóniny hlasu.

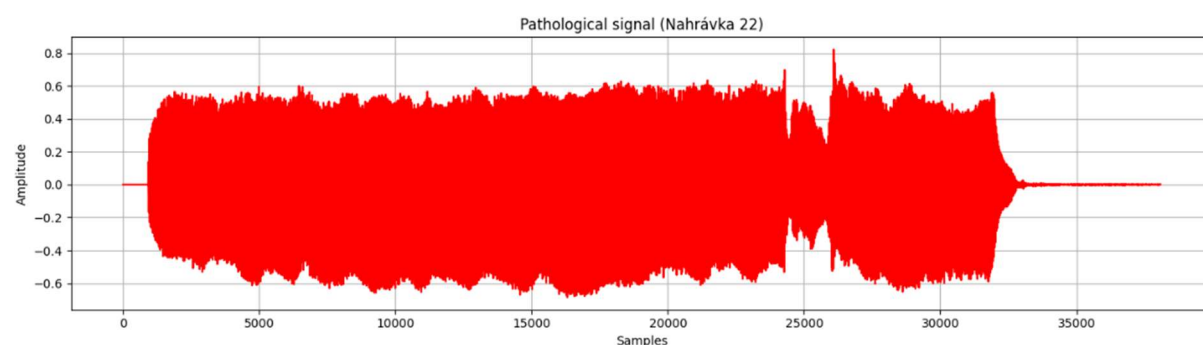
Výpočet variability tóniny hlasu a klasifikace



Obrázek 1 - variabilita tonu

Předběžnou klasifikaci nahrávek jsem klasifikoval 41 zdravých a 167 patogenních záznamů, pouze pomocí nastavení prahu variability tóniny hlasu.

Samotný tón ale nemusí znamenat nutně patogenní záznam, jelikož tón ovlivňuje i pohlaví, věk a celkový projev jedince.

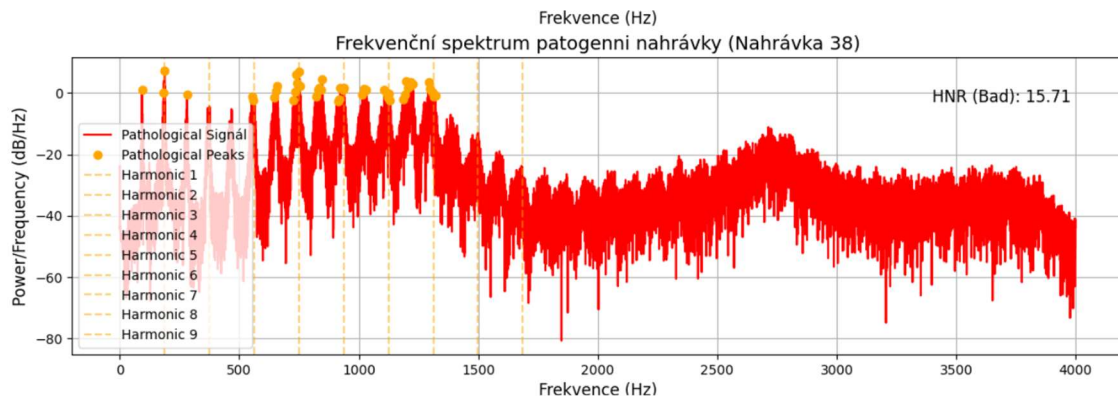


Obrázek 2 - propad u patogenní nahrávky

Fourierova transformace

Klasifikace čistě podle tónu hlasu je velice nepřesná a proto jsem se rozhodl provést výpočet Fourierovy transformace. FFT (Fast Fourier Transform) převede zvukový signál z časové domény do frekvenční domény.

Vypočítá se hustota výkonového spektra (PSD), která udává sílu jednotlivých frekvencí v signálu. Vráťi kladné frekvence a odpovídající PSD, protože negativní složky nejsou relevantní pro analýzu. Následně se identifikují dominantní frekvence (peaks) pomocí funkce `find_peaks` v PSD. Podle dominantních frekvencí se určí prvních 10 harmonických složek a vypočte HNR, což je poměr mezi harmonickou a šumovou energií.

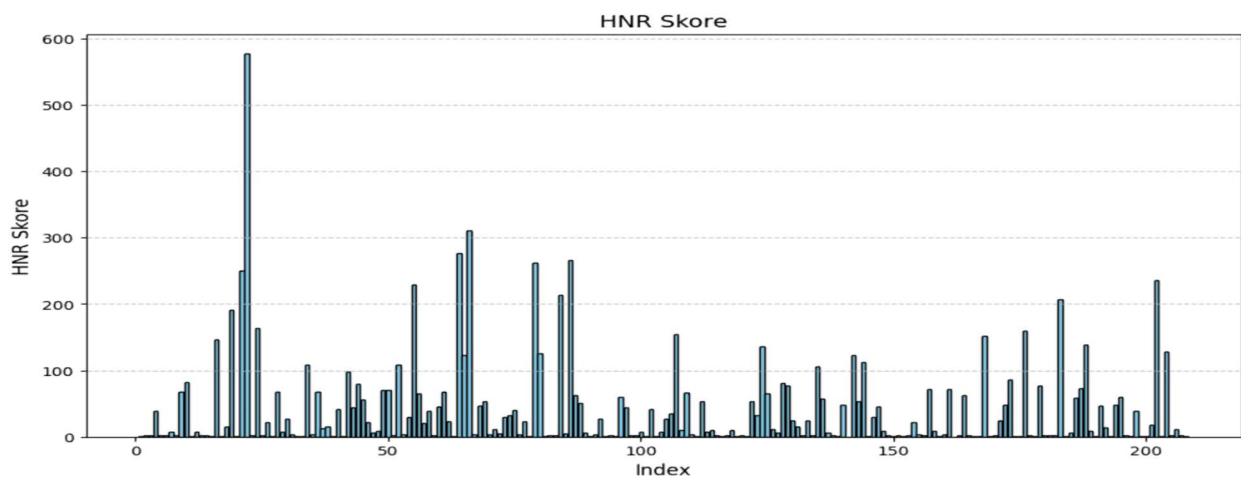


Obrázek 3 - frekvenční spektrum

Co tato analýza odhaluje?

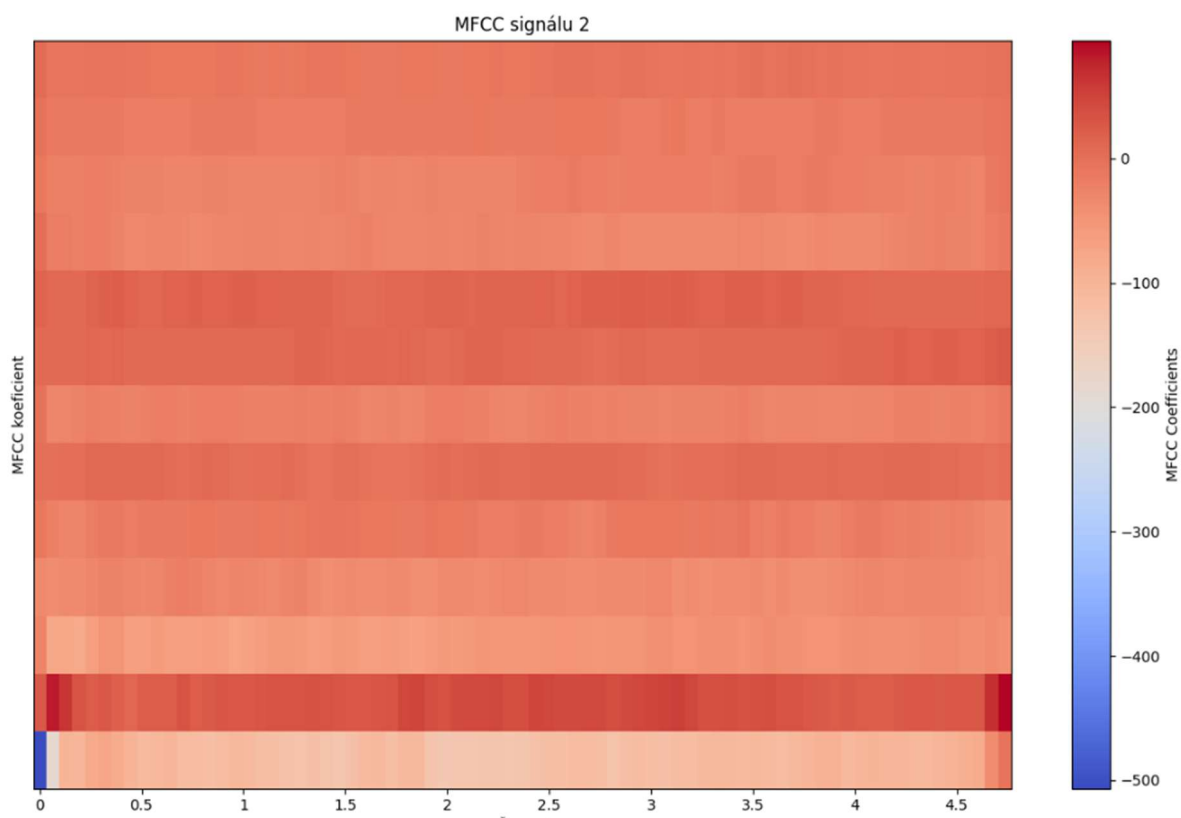
- Hlavní frekvence hlasu – umožňuje určit, zda se v signálu objevují přirozené harmonické složky.
- Přítomnost a rozložení harmonických složek – zdravý hlas má jasnější harmonickou strukturu, zatímco patologický hlas má často rozpadlé nebo oslabené harmonické frekvence.
- HNR (harmonic-to-noise ratio) – vysoká hodnota značí čistý hlas, nízká hodnota naznačuje zvýšený šum nebo nepravidelnosti v signálu.

Samotné HNR nám řekne asi nejvíce a také jde nejlépe zakomponovat do finálního výpočtu, proto jsem si vypočetl HNR skóre všech záznamů.



Kepstrální analýza (MFCC)

Tato analýza se používá k získání klíčových rysů zvuku a dobře reprezentuje jak lidské ucho vnímá zvuk. Např. dysfonie, která je vedena v anotacích se dobře rozpozná díky nestabilní MFCC struktuře, celkově lze touto metodou odhalit hlasové abnormality.



Obrázek 4 - kepstrální analýza

Pro každý záznam bylo vypočteno MFCC Skóre.

SPOJENÍ HNR, MFCC, ZCR A SPECTRAL SKORE

Pro finální verdikt jsem se rozhodl spojit předchozí dvě skóre (HNR a MFCC) a k nim přidat pro větší přesnost ZCR a Spectral skóre.

Výhody Zero-Crossing Rate (ZCR):

- Udává frekvenci změny polarity signálu (kolikrát signál protne osu X za sekundu).
- Vysoké ZCR může znamenat, že signál obsahuje více šumu, což bývá typické pro patologické hlasy.

Výhody Spektrálního centroidu:

- Udává těžiště frekvenčního spektra – tedy, kde je soustředěna většina energie signálu.
- Patologické hlasy často mají nižší spektrální centroid kvůli zvýšené hrubosti a poruchám ve vyšších frekvencích.

Každý zvukový záznam je popsán jako vektor složený z různých rysů:

- HNR (jedna hodnota)
- MFCC (průměr všech 13 koeficientů)
- ZCR (jedna hodnota)
- Spektrální centroid (jedna hodnota)

Tyto hodnoty jsou sloučeny do jednoho celkového vektoru charakterizujícího hlasový záznam.

Závěr

Pomocí pravidlového klasifikátoru jsem roztřídil záznamy, přičemž jsem použil manuálně nastavené prahové hodnoty pro jednotlivé rysy hlasu.

Rysy hlasu a nastavené prahové hodnoty:

- Nízké HNR → patologický hlas (více šumu). **hnr_threshold = 23**
- Nízké MFCC průměry → patologický hlas (menší amplitudy, nepravdělnosti). **mfcc_threshold = -65**
- Vysoké ZCR → patologický hlas (více změn polarity, signál je méně stabilní). **zcr_threshold = 0.3**
- Vysoký spektrální centroid → patologický hlas (více energie ve vyšších frekvencích, možné narušení). **spectral_centroid_threshold = 1000**

Konečným výsledkem je 69,71% úspěšnost klasifikace (viz výpis)

```
Extrahovaná diagnóza pro soubor voice208.he: hyperkinetic dysphonia  
Počet správně klasifikovaných záznamů: 145 z 208  
Přesnost klasifikace: 69.71%
```

Výpis jednotlivých shod a neshod:

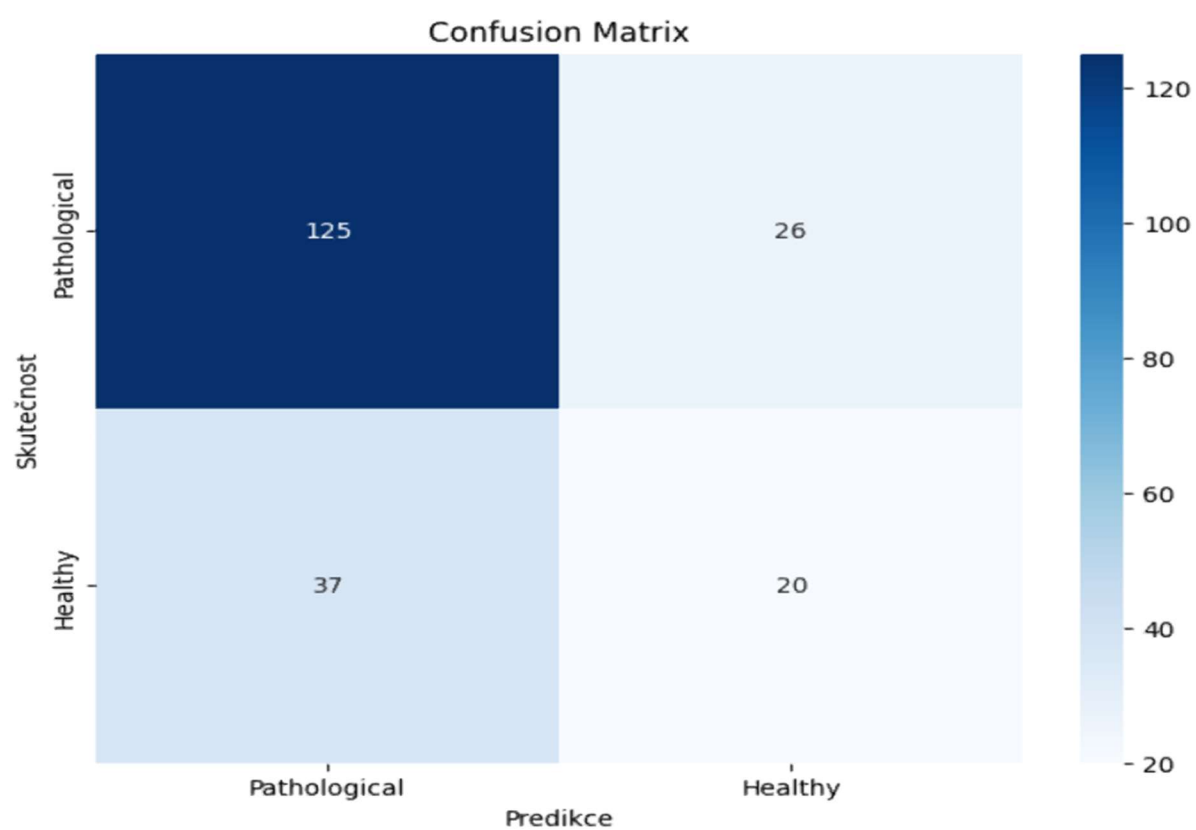
Příklad 1: Klasifikace: Pathological, Anotace: Pathological, Shoda

Příklad 2: Klasifikace: Healthy, Anotace: Healthy, Shoda

Obrázek 5 - výpis výsledků



Obrázek 6 - výsledek (přesnost)



Obrázek 7 - výsledek (confusion matrix)