

Seminární práce č. 2

Analýza hlasových záznamů

Adam Legner

F22130

Počítačové zpracování signálů (PZS)

Univerzita J. E. Purkyně

Zadání

Úkolem bylo analyzovat hlasové záznamy z databáze VOICED (Voice ICar fEDerico II). Databáze obsahuje nahrávky samohlásky "a" od zdravých i nemocných pacientů, vzorkované na 8000 Hz. Cílem bylo rozlišit zdravé a patologické hlasy pomocí spektrální analýzy, HNR a MFCC.

Postup

Načtení dat

Data jsem načetl pomocí knihovny WFDB, která umí číst formát PhysioNet. Každý záznam má svůj .hea soubor (hlavička) a .txt soubor (samotný signál). Ke každému záznamu je navíc info soubor s diagnózou, věkem a pohlavím pacienta. Celkem 208 záznamů – 57 zdravých a 151 patologických.

Zobrazení signálů

Vykreslil jsem časové průběhy zdravého a patologického hlasu. Na ose x je čas v sekundách, na ose y amplituda signálu. Zdravý hlas má pravidelnější průběh – periody se opakují rovnoměrně. Patologický hlas vykazuje nepravidelnosti v amplitudě a periodicitě, což odpovídá poruchám kmitání hlasivek.

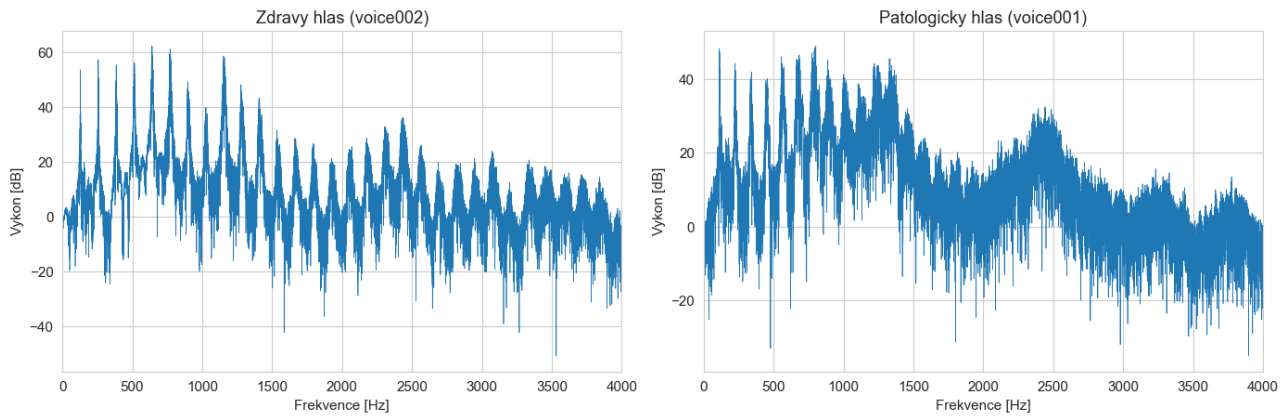


Spektrální analýza (FFT)

Signál jsem převedl do frekvenční oblasti pomocí rychlé Fourierovy transformace (FFT). FFT rozloží signál na jednotlivé frekvenční složky – výsledkem je spektrum, kde na ose x jsou frekvence a na ose y jejich výkon v dB. Zdravý hlas má výrazné harmonické špičky (násobky základní frekvence), zatímco u patologického hlasu jsou špičky méně zřetelné a je v něm více šumu.

Pro každý záznam jsem dále spočítal spektrální centroid – vážený průměr frekvencí, kde vahou je výkon na dané frekvenci. Udává, kde leží "těžiště" spektra. U zdravých hlasů vyšel kolem 824 Hz, u patologických 815 Hz. Rozdíl je malý, sám o sobě pro klasifikaci nestačí.

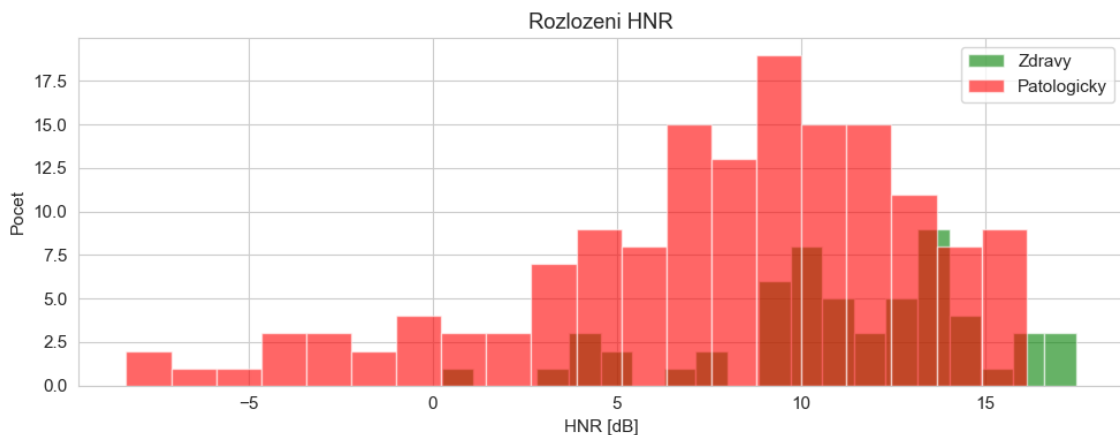
Vykonova spektra (FFT)



Harmonický šum (HNR)

HNR (Harmonic-to-Noise Ratio) měří, kolik energie v signálu tvoří harmonická (periodická) složka a kolik šum. Počítám ho přes autokorelaci – autokorelace signálu ukazuje, jak moc se signál podobá sám sobě po posunutí o určitý počet vzorků. Periodický signál má v autokorelaci výrazný peak na vzdálenosti odpovídající jedné periodě.

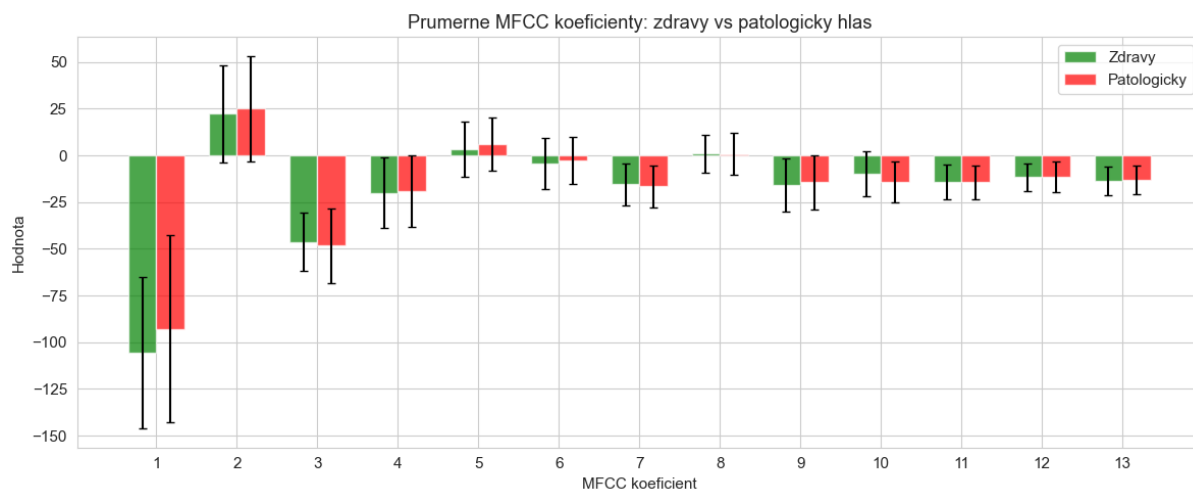
Hledám maximum autokorelace v rozsahu lagů odpovídajících F_0 50–500 Hz. Z výšky tohoto peaku odvodím HNR v dB. Čím vyšší peak, tím je signál periodičtější a méně zašuměný. Zdravé hlasy vyšly v průměru 11.09 dB, patologické 7.87 dB – rozdíl je výraznější než u spektrálního centroidu.



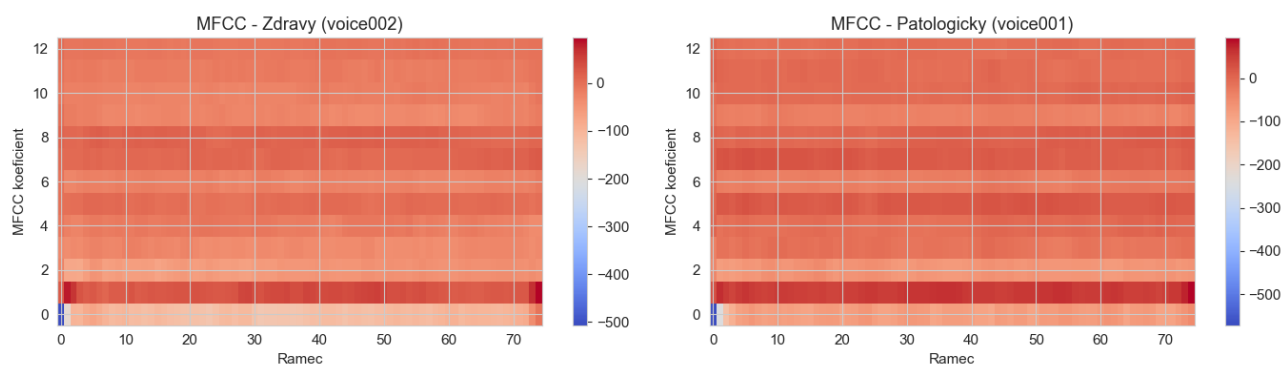
Kepstrální analýza (MFCC)

MFCC (Mel-Frequency Cepstral Coefficients) zachycují tvar spektrální obálky signálu. Postup výpočtu: signál se rozdělí na krátké rámce, pro každý rámeček se spočítá výkonové spektrum, to se převede na mel-škálu (nelineární škála odpovídající lidskému vnímání výšky), aplikuje se logaritmus a nakonec diskrétní kosinová transformace. Výsledkem je sada koeficientů, které kompaktně popisují spektrální vlastnosti hlasu.

Vypočítal jsem 13 MFCC koeficientů pro každý záznam pomocí librosa. Největší rozdíly mezi zdravými a patologickými hlasy se projevily u prvních koeficientů (MFCC 1–4), které zachycují hrubý tvar spektra.



Kepstrogramy – vizualizace MFCC v čase pro zdravý a patologický hlas:



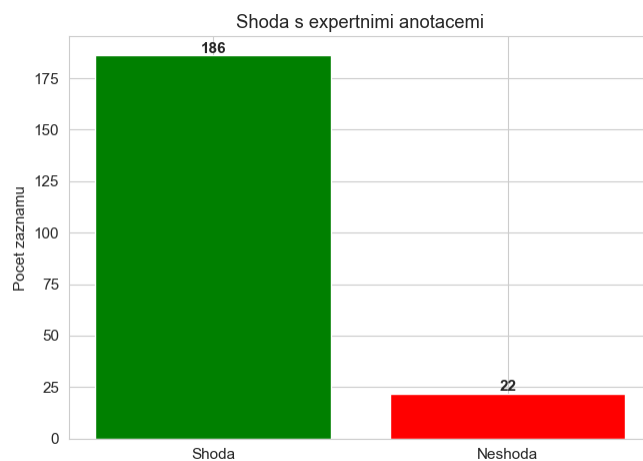
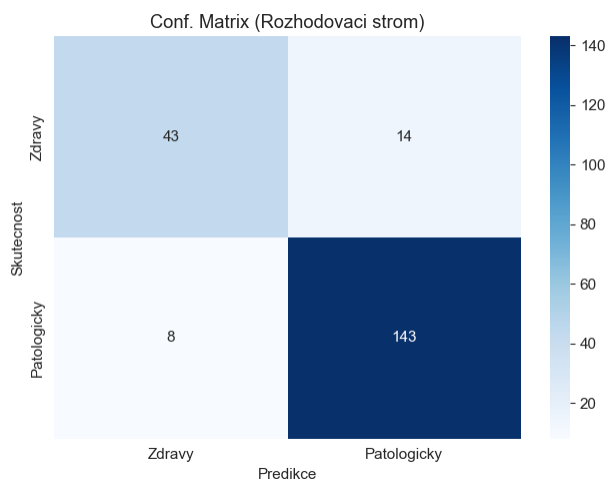
Kombinovaná klasifikace

Jako příznaky jsem použil spektrální centroid, HNR a 13 MFCC koeficientů (celkem 15 příznaků na záznam). Klasifikoval jsem rozhodovacím stromem (DecisionTreeClassifier, max_depth=5). Strom postupně dělí data podle jednotlivých příznaků – v každém uzlu vybere příznak a práh, který nejlépe oddělí zdravé od patologických.

Kvalitu jsem ověřoval stratifikovanou 5-fold křížovou validací – data se 5× rozdělí na trénovací a testovací část, model se pokaždé natrénuje na 4/5 dat a vyhodnotí na zbylé 1/5. Výsledek CV: cca 70 %. Na celých trénovacích datech přesnost kolem 89 %.

Porovnání s expertními anotacemi

Predikce modelu jsem porovnal s diagnózami od lékařů. Výsledky jsou vidět v matici záměn a grafu shody.



Závěr

Analyzoval jsem 208 hlasových záznamů z databáze VOICED. Použil jsem spektrální analýzu (FFT, spektrální centroid), HNR počítaný přes autokorelaci a 13 MFCC koeficientů. Rozhodovací strom s těmito příznaky dosáhl na křížové validaci přesnost kolem 70 %, na trénovacích datech 89 %. Samotný jeden příznak příliš nestačí, kombinace více příznaků pomáhá.