

Seminární práce II – Počítačové zpracování signálu (KI/PZS)

Autor: Arsalan Safi
Rok: 2025/2026

1. Zadání

Zadáním seminární práce bylo zpracování hlasových záznamů z databáze PhysioNet – VOICED Database. Cílem bylo pomocí technik v časové a frekvenční oblasti klasifikovat záznamy hlásky „a“ na normální a patologické. V případě patologických záznamů bylo požadováno rozlišení jednotlivých poruch uvedených v hlavičkových souborech. Pro zpracování byly použity metody zahrnující Fourierovu a kepstrální analýzu. Úspěšnost klasifikace byla porovnána s anotacemi expertů.

2. Použitá data a načtení záznamů

Databáze VOICED obsahuje záznamy hlasu od 208 subjektů (58 normálních a 150 patologických). Každý subjekt má typické soubory PhysioNet (hlavička .hea, datový soubor .dat). Knihovna WFDB byla použita pouze pro načtení dat (rdrecord/rdheader). Pro samotné zpracování signálu nebyly použity žádné hotové zpracovatelské funkce z WFDB.

3. Postup řešení

Nejprve byly z hlavičkových souborů automaticky parsovány anotace (normální vs. patologický a typ patologie). Poté byl načten samotný audio signál a předzpracován odečtením DC složky a normalizací amplitudy.

Příznaky (features) byly extrahovány po krátkých rámcích (např. 25 ms s posunem 10 ms). Použity byly tři skupiny příznaků:

- Časová oblast: RMS, ZCR, šíkmost a špičatost (agregace průměr + směrodatná odchylka)
- Frekvenční oblast (FFT): spektrální centroid, šířka pásma, roll-off, spektrální flatness
- Kepstrální oblast: MFCC (mel filterbank → log energie → DCT; agregace mean+std koeficientů)

4. Klasifikace

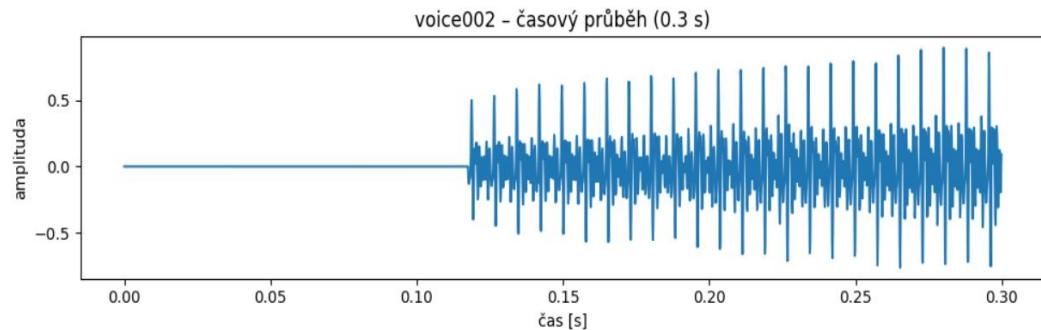
Pro klasifikaci byly použity dva klasifikátory: k-Nearest Neighbors (kNN) a Support Vector Machine (SVM, RBF). Vstupní příznaky byly standardizovány (StandardScaler). Byly řešeny dvě úlohy:

- a) binární klasifikace: normální vs. patologický hlas,
- b) multitřídní klasifikace: normální + jednotlivé typy patologií.

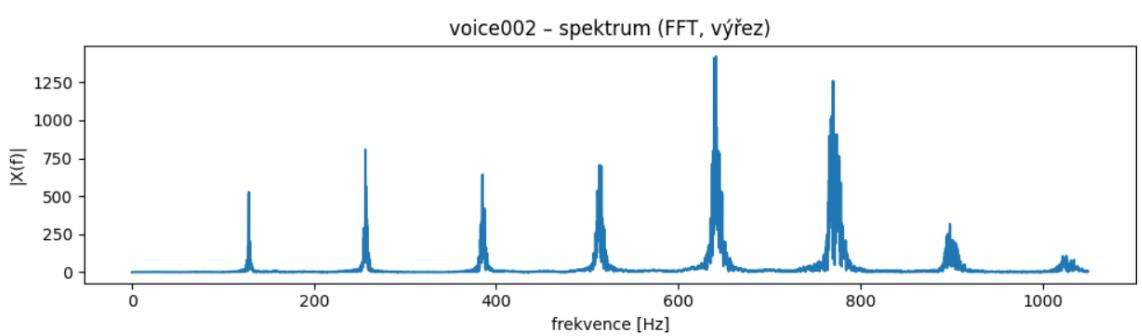
5. Výsledky

Úspěšnost klasifikace byla vyhodnocena na oddělené testovací množině (stratifikační split). Pro binární úlohu byly reportovány metriky accuracy, balanced accuracy a F1. Pro multitřídní úlohu byly reportovány accuracy a macro F1 a dále konfuzní matice a klasifikační report po třídách.

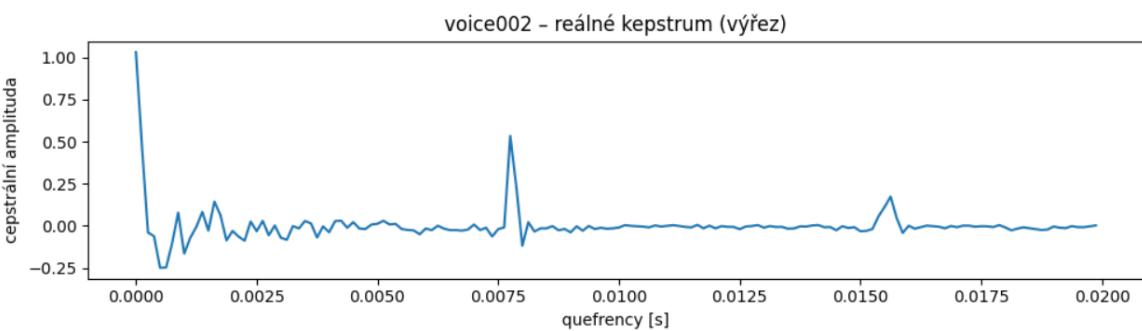
Doporučené grafické výstupy k vložení (z notebooku):



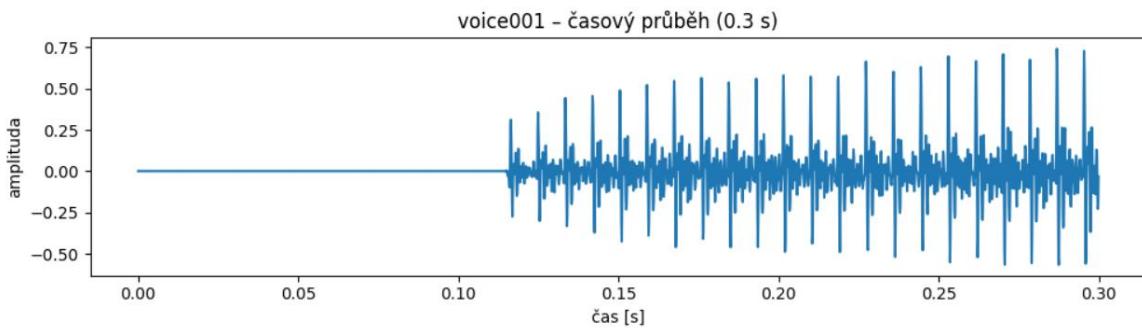
Obrázek 1: Ukázka časového průběhu normal vs. pathological (0.3 s)



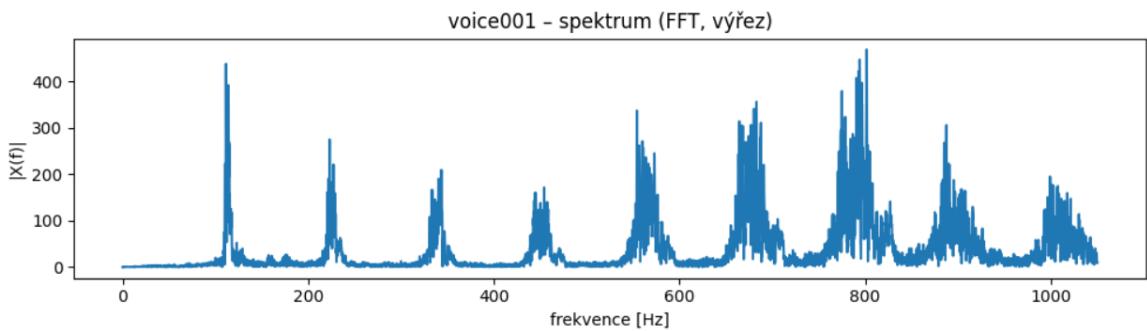
Obrázek 2: FFT spektrum normal vs. pathological (výřez)



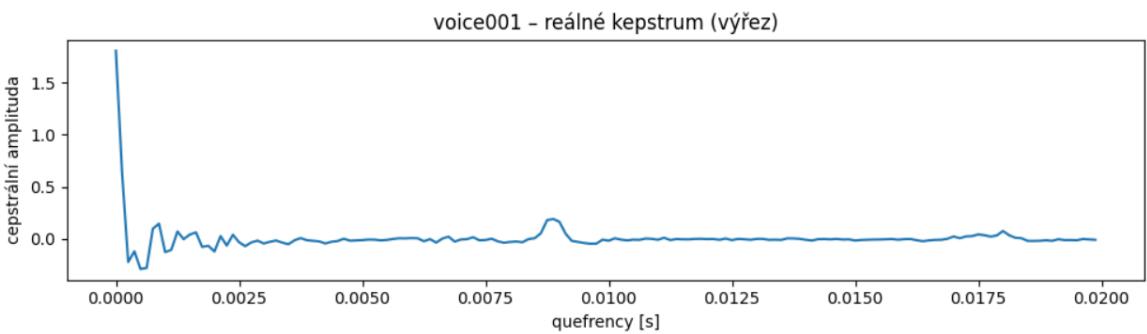
Obrázek 3: Kepstrum / MFCC ukázka



Obrázek 4: PCA vizualizace (projekce příznaků do 2D)



Obrázek 5: Konfuzní matici – binární klasifikace (kNN/SVM)



Obrázek 6: Konfuzní matici – multitřídní klasifikace (SVM)

6. Závěr

V práci byla navržena sada příznaků kombinující časovou, frekvenční a kepstrální analýzu a byly natrénovány klasifikátory kNN a SVM pro rozlišení normálního a patologického hlasu a pro rozlišení jednotlivých patologií. Výsledky ukázaly, že kombinace příznaků včetně MFCC umožňuje oddělení tříd a dosažení dobré úspěšnosti. Zhoršení u některých tříd může být způsobeno nevyvážeností tříd a variabilitou nahrávek.

7. Literatura

- PhysioNet – VOICED Database (voiced/1.0.0)
- Dokumentace WFDB
- Poznámky z předmětu KI/PZS (Fourierova a kepstrální analýza)