

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

Klasifikace hlasových záznamů z databáze VOICED

Autor: Arsalan Safi

1. Zadání

Zadáním bylo zpracování hlasových záznamů hlásky „a“ z databáze PhysioNet – VOICED Database. Cílem bylo klasifikovat záznamy na normální a patologické a u patologických dále rozlišit typ poruchy dle anotací v hlavičkách. Úspěšnost byla vyhodnocena porovnáním s referenčními anotacemi (expertní štítky).

2. Data a načtení

Databáze VOICED obsahuje 208 subjektů (57 normálních, 151 patologických). Každý záznam je uložen ve formátu PhysioNet (hlavička .hea, datový soubor .dat). Knihovna WFDB byla použita pouze pro načtení dat (rdrecord/rdheader); pro samotné zpracování signálu byly použity vlastní implementace.

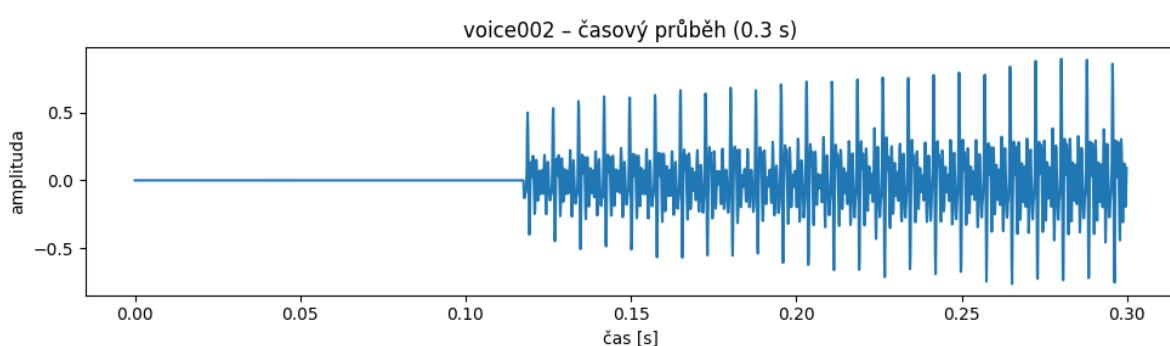
3. Postup řešení

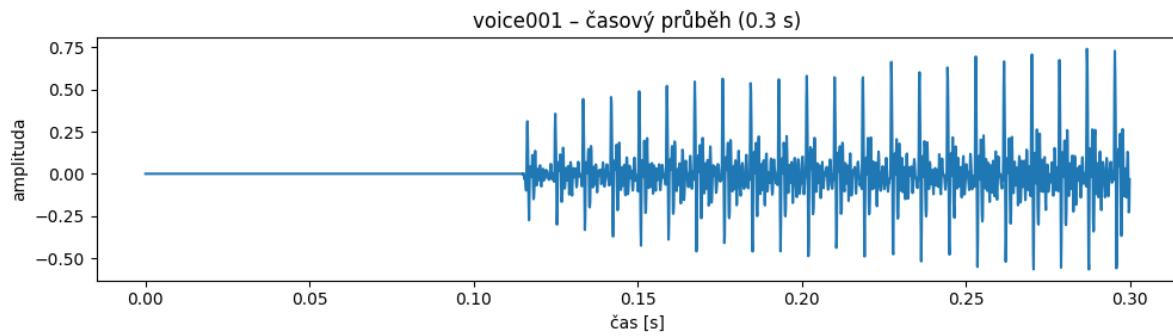
Ze souborů hlaviček byly automaticky parsovány štítky: (i) binární normal/pathological a (ii) multitřídní diagnóza. Signál byl předzpracován odečtením DC složky a normalizací amplitudy. Příznaky byly počítány po krátkých rámcích (cca 25 ms s posunem 10 ms) a agregovány (mean + std) na úroveň záznamu.

Použité příznaky:

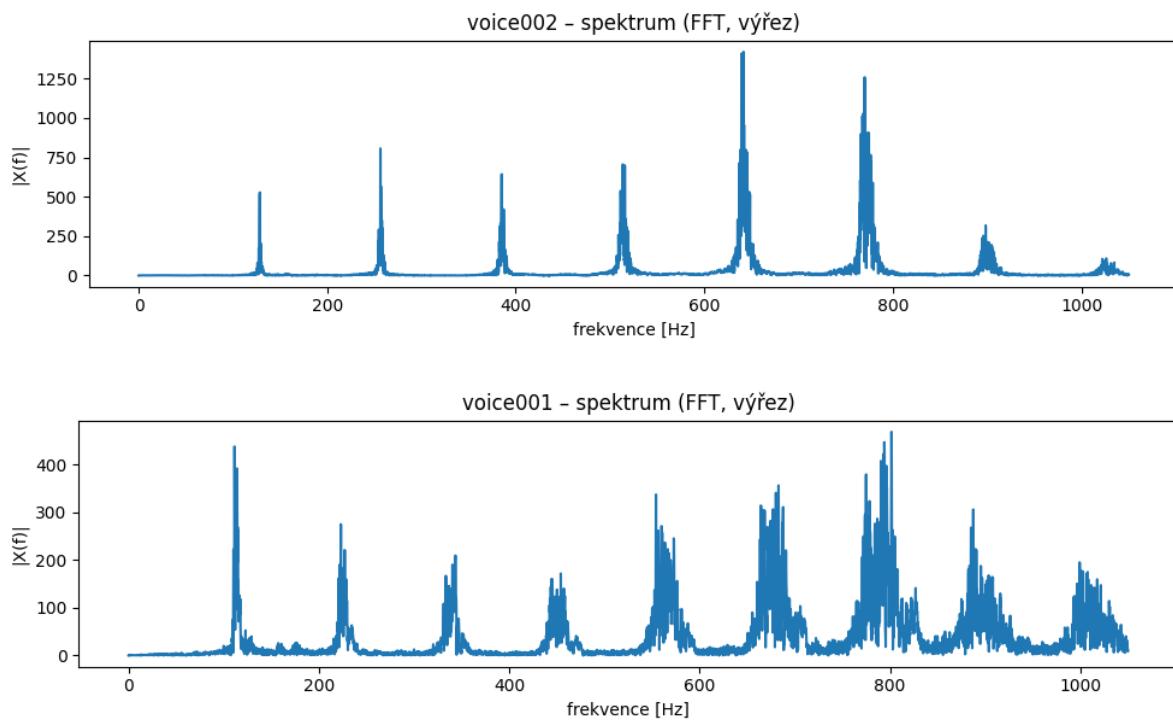
- Časová oblast: RMS, ZCR, šikmost, špičatost
- Frekvenční oblast (FFT): spektrální centroid, šířka pásma, roll-off, spektrální flatness
- Kepstrální oblast: MFCC (mel filterbank → log energie → DCT; aggregace mean + std koeficientů)

4. Grafické výstupy práce se signálem

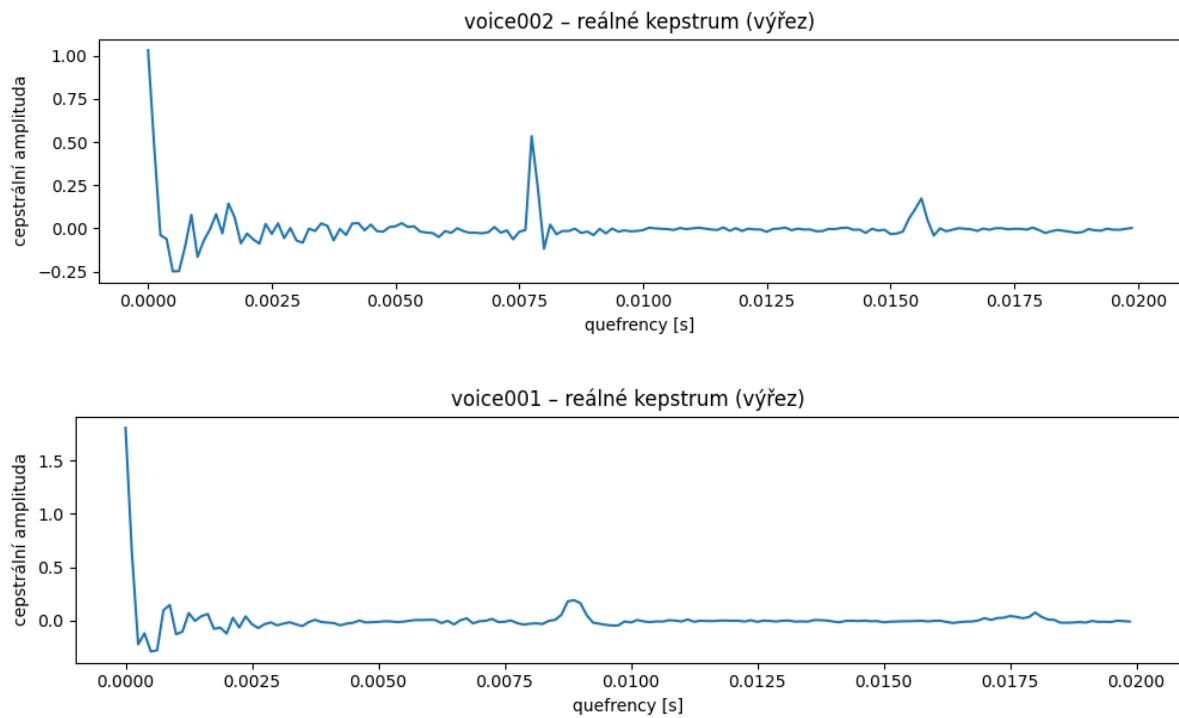




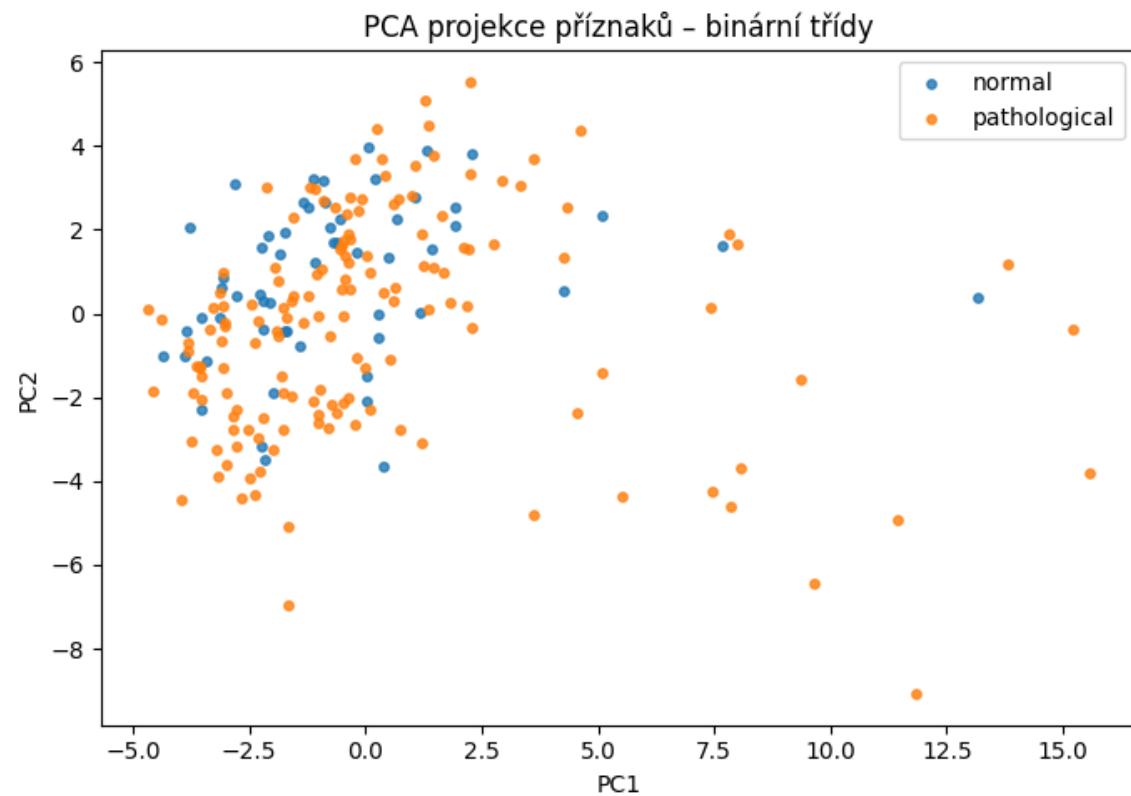
Obrázek 1: Časový průběh (0.3 s) – normal (voice002) a pathological (voice001).



Obrázek 2: FFT spektrum (výřez) – normal (voice002) a pathological (voice001).



Obrázek 3: Reálné kepstrum (výřez) – normal (voice002) a pathological (voice001).



Obrázek 4: PCA projekce příznaků do 2D – binární třídy (normal vs. pathological).

5. Klasifikace a vyhodnocení úspěšnosti

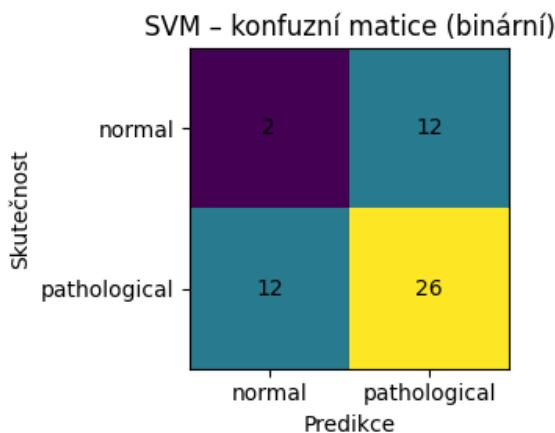
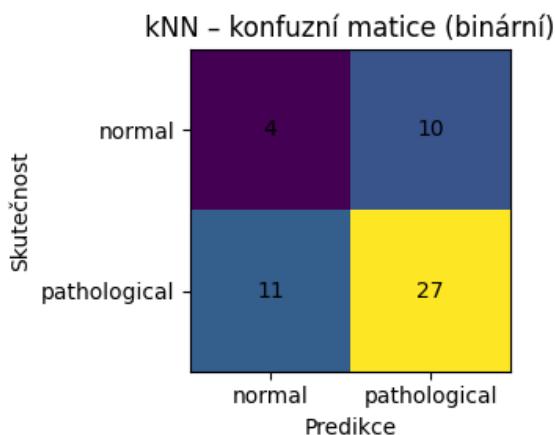
Klasifikace byla provedena pomocí kNN a SVM (RBF). Data byla rozdělena stratifikovaným splitem na trénovací a testovací část; příznaky byly standardizovány (StandardScaler). Úspěšnost je reportována na testovací množině.

5.1 Binární klasifikace (normal vs. pathological)

Souhrn metrik (accuracy, balanced accuracy, F1):

model	acc	bacc	f1
kNN (k=5)	0.596154	0.498120	0.720000
SVM (RBF)	0.538462	0.413534	0.684211

Z tabulky je patrné, že accuracy je přibližně 0.54–0.60, ale balanced accuracy je nižší (≈ 0.41 –0.50), což odpovídá nevyváženosti tříd (151 patologických vs. 57 normálních). F1 je relativně vyšší díky převaze pozitivní (pathological) třídy.



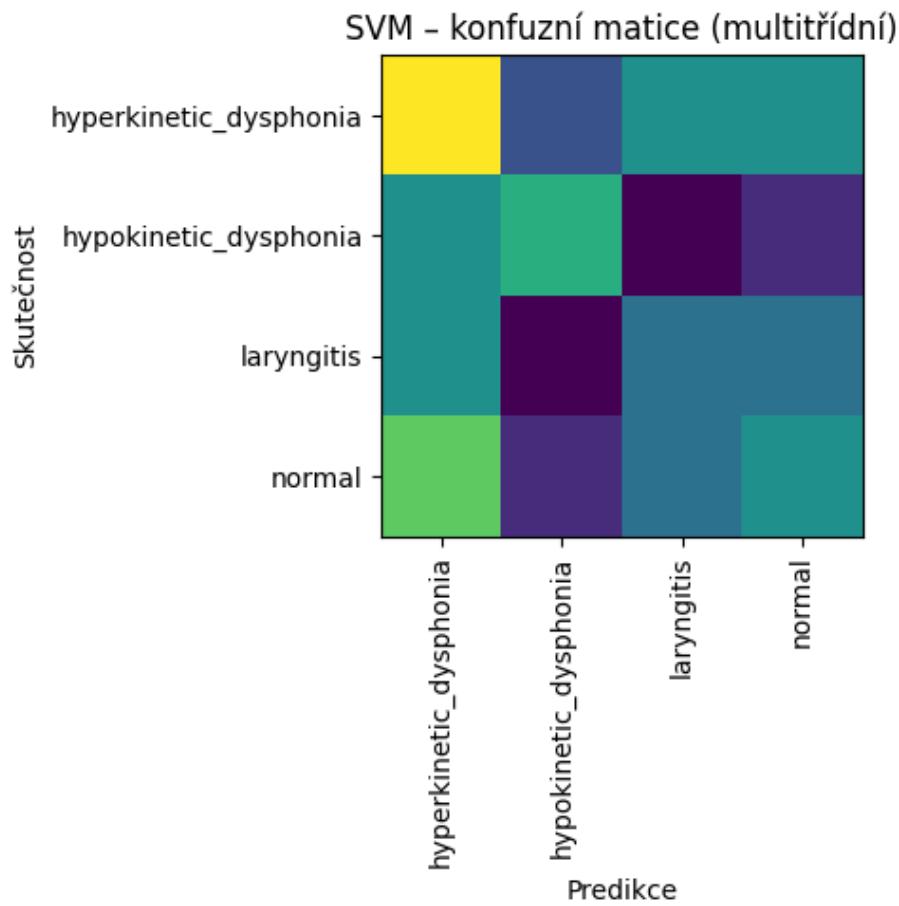
Obrázek 5: Konfuzní matice – binární klasifikace (kNN a SVM).

5.2 Multitřídní klasifikace (typ patologie)

Souhrn metrik (accuracy, macro F1):

model	acc	f1_macro
kNN (k=7)	0.307692	0.285822
SVM (RBF)	0.384615	0.390812

Multitřídní úloha je náročnější; dosažená accuracy je ~ 0.38 a macro F1 ~ 0.39 (SVM). Hlavní příčinou zhoršení je podobnost některých patologií a omezený počet vzorků v jednotlivých třídách.



Obrázek 6: Konfuzní matice – multitřídní klasifikace (SVM).

Detailní metriky po třídách (SVM, multitřídní):

třída	PPV	Se	F1	support
hyperkinetic_dysphonia	0.363636	0.444444	0.400000	18.000000
hypokinetic_dysphonia	0.625000	0.500000	0.555556	10.000000

laryngitis	0.300000	0.300000	0.300000	10.000000
normal	0.333333	0.285714	0.307692	14.000000
accuracy	0.384615	0.384615	0.384615	0.384615
macro avg	0.405492	0.382540	0.390812	52.000000
weighted avg	0.393502	0.384615	0.385832	52.000000

6. Závěr

Byla navržena sada příznaků kombinující časovou, frekvenční a kepstrální analýzu (včetně MFCC) a byly natrénovány klasifikátory kNN a SVM. Výsledky ukazují, že oddělení normal vs. pathological je proveditelné, avšak nevyváženost tříd snižuje balanced accuracy.

Multitřídní rozlišení jednotlivých patologií je obtížnější a vykazuje nižší úspěšnost, zejména kvůli podobnosti tříd a variabilitě nahrávek.

7. Literatura

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