

VYSOKÉ UČENÍ TECHNICKÉ V BRNĚ  
FAKULTA INFORMAČNÍCH TECHNOLOGIÍ

Řešení ISS projektu 2024/25 ("Podle hlasu  
poznáte je ...")

Hana Liškařová

# Obsah

<b>1</b>	<b>Úvod</b>	<b>2</b>
<b>2</b>	<b>Zvolené řešení</b>	<b>2</b>
2.1	Mel-frekvenční keprální koeficienty . . . . .	2
2.1.1	Kroky výpočtu MFCC . . . . .	2
2.2	Měření podobnosti signálů . . . . .	2
<b>3</b>	<b>Implementace</b>	<b>3</b>
3.1	Načtení signálů . . . . .	3
3.2	Zobrazení grafů . . . . .	3
3.2.1	Testovací signály . . . . .	3
3.2.2	Referenční signály . . . . .	4
<b>4</b>	<b>Extrakce MFCC příznaků</b>	<b>5</b>
4.1	Výpočet vzdáleností a korelací . . . . .	5
4.2	Klasifikace signálů . . . . .	5
4.3	Výstup . . . . .	6
<b>5</b>	<b>Výsledky</b>	<b>7</b>
<b>6</b>	<b>Závěr</b>	<b>8</b>

# 1 Úvod

Cílem projektu je přiřadit k sobě referenční nahrávky - nahrávky zvuku motoru čtyř známých aut na konstantních otáčkách a testovací nahrávky, které pocházejí od stejných aut, ale motor běží na jiných otáčkách. S tím že je možné, že jedna nahrávka nepatří k žádnému ze čtyř známých aut.

## 2 Zvolené řešení

Dle doporučeného postupu řešení jsem si nejprve poslechla jednotlivé nahrávky a pokusila se přiřadit je k sobě čistě podle sluchu a vzhledu signálu nahrávek. Tímto postupem jsem však měla malou úspěšnost.

Zvolené řešení vychází z analýzy a zpracování zvuku pomocí **Mel-frekvenčních keprstrálních koeficientů**, dále jen MFCC. Tato metoda umožňuje převést časový průběh zvuku na sadu příznaků, která lépe reprezentuje zvuk a umožňuje přesnější klasifikaci.

### 2.1 Mel-frekvenční keprstrální koeficienty

*Mel-frekvenční keprstrální koeficienty (MFCC)* jsou jedním z nejdůležitějších příznaků používaných v analýze zvuku a řeči. MFCC převádí časový signál na sadu příznaků, které lépe odpovídají tomu, jak lidské ucho vnímá frekvence. MFCC bere v úvahu **Melovu stupnici**, která lépe modeluje lidské vnímání frekvencí.

#### 2.1.1 Kroky výpočtu MFCC

- Rozdělení signálu na rámce – Zvuk je rozdělen na malé časové úseky (20-40 ms).
- Aplikace **Hammingova okna** – Slouží k odstranění ostrých přechodů na okrajích rámců.
- **Fourierova transformace** (DFT) – Převádí časový signál na frekvenční spektrum.
- Aplikace "**Melovy filtrační banky**" (anglicky *Mel-filterbank*) – Frekvenční spektrum je filtrováno pomocí trojúhelníkových filtrů na Melově stupnici.
- Logaritmická amplituda – Z amplitudy se počítá logaritmus.
- **Diskrétní kosinová transformace** (DCT) – Redukuje dimenzi spektra a extrahuje koeficienty MFCC.

Použitím MFCC získáváme příznaky, které jsou méně citlivé na hluk a jiné rušivé vlivy, což je důležité pro robustní klasifikaci.

Použití MFCC jsem zvolila kvůli možnosti lepší reprezentace zvuku (kompaktní) a redukce dimenze do menšího počtu koeficientů pro rychlejší výpočet. Implementací MFCC jsem také získala příznaky, které jsou méně citlivé na hluk a jiné rušivé vlivy.

### 2.2 Měření podobnosti signálů

Kromě MFCC jsem implementovala dvě metody pro měření podobnosti signálů.

#### 1. Eukleidovská vzdálenost

- Měří rozdíl mezi vektory příznaků testovacích a referenčních signálů.

#### 2. Pearsonův korelační koeficient

- Měří shodu (korelaci) mezi dvěma signály.
- Čím vyšší je korelační index, tím jsou si signály podobnější.

## 3 Implementace

Implementace se skládá z několika fází, které jsou popsány níže.

### 3.1 Načtení signálů

Pro tuto část implementace jsem použila již připravený kód ze zadání projektu.

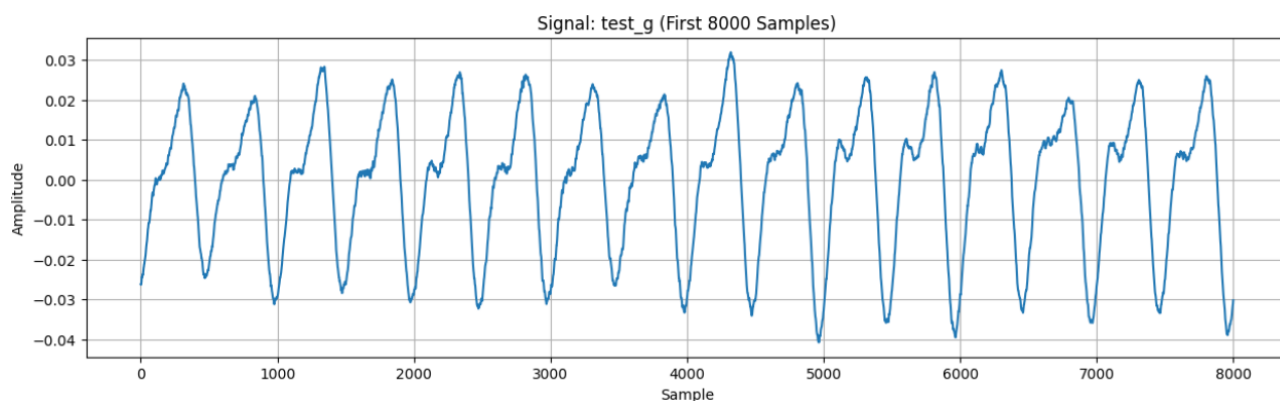
Program nejprve načte referenční a testovací signály z datových souborů (WAV souborů).

### 3.2 Zobrazení grafů

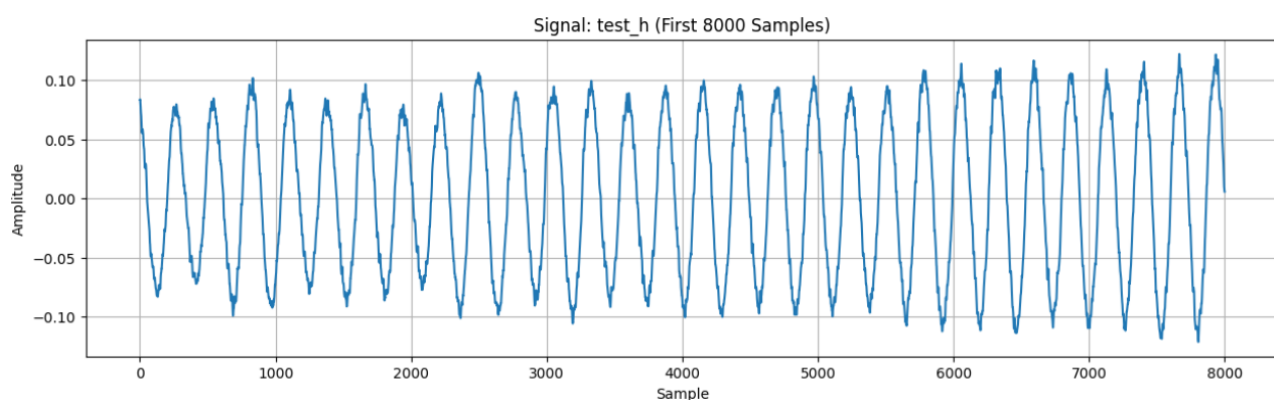
V této sekci jsem pomocí mnou implementované funkce `plot_signals` zobrazila grafy jednotlivých testovacích a referenčních signálů. Grafy jsou omezené na různý počet vzorků (16 000, 8 000, 4 000), aby křivka signálu byla v každém dobře rozpoznatelná.

Níže můžeme vidět jednotlivé grafy.

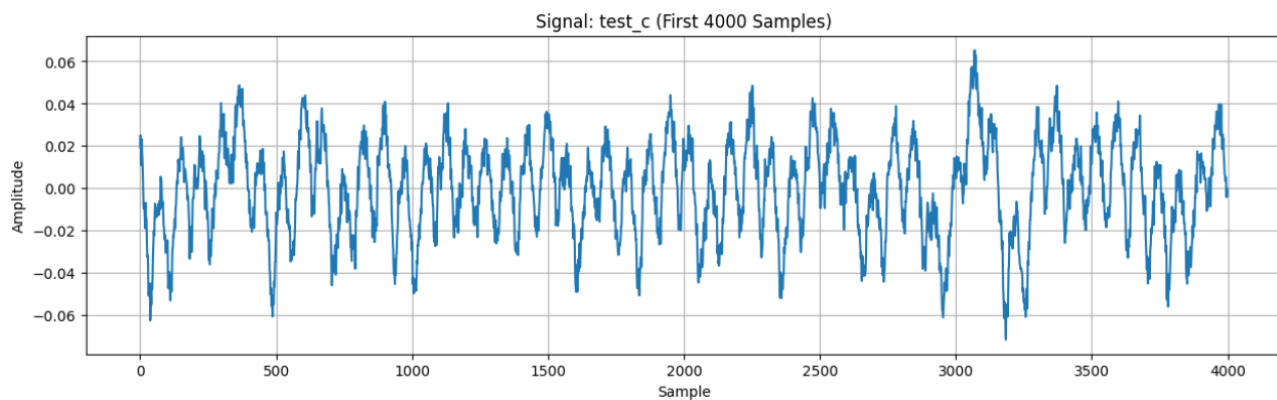
#### 3.2.1 Testovací signály



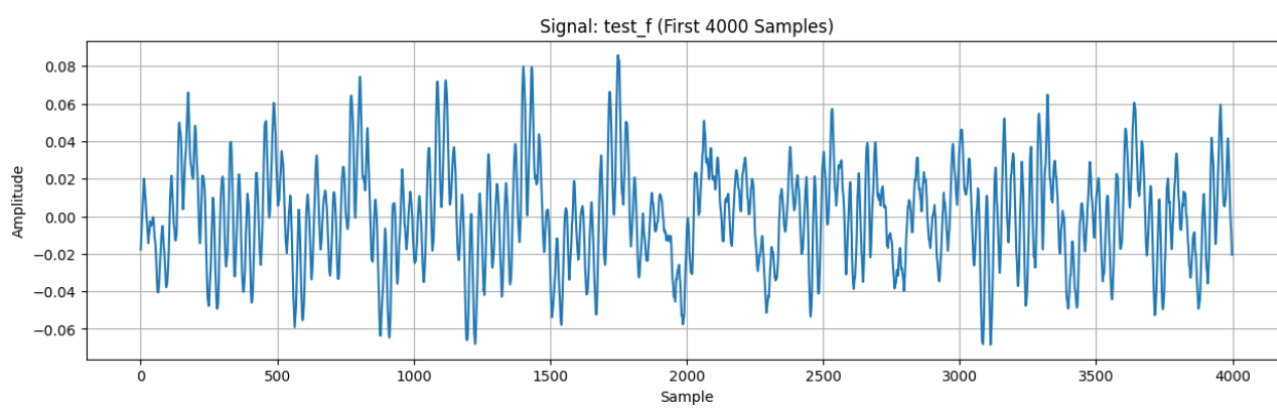
Obrázek 1: test\_g



Obrázek 2: test\_h

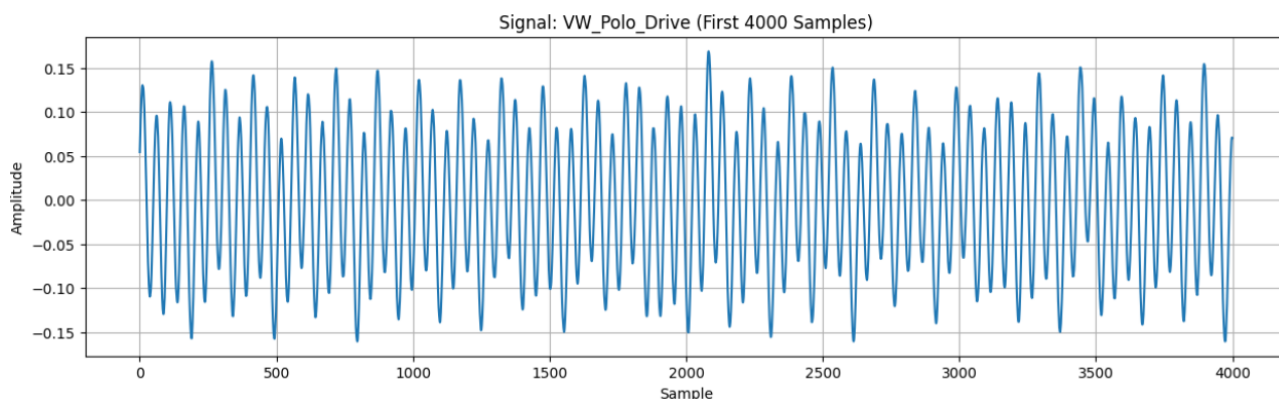


Obrázek 3: test\_c

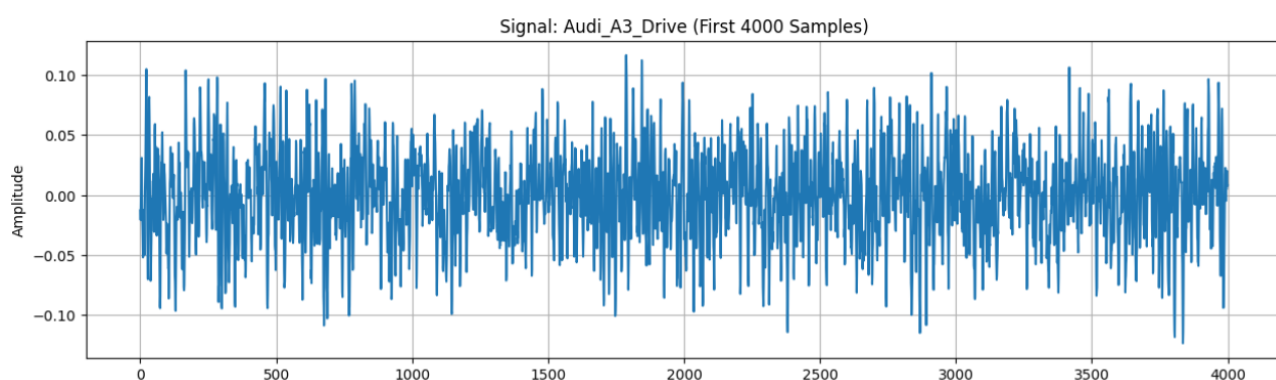


Obrázek 4: test\_f

### 3.2.2 Referenční signály



Obrázek 5: VW\_Polo\_Drive



Obrázek 6: Audi\_A3\_Drive

## 4 Extrakce MFCC příznaků

Pro každý signál se spočítají koeficienty MFCC, *delta* (první derivace) a *delta-delta* (druhá derivace) koeficienty. Pro každý z těchto typů koeficientů se vypočítá **průměr** a **směrodatná odchylka**. Všechny tyto informace se spojí do jednoho vektoru příznaků pro každý signál (`mfcc_features`).

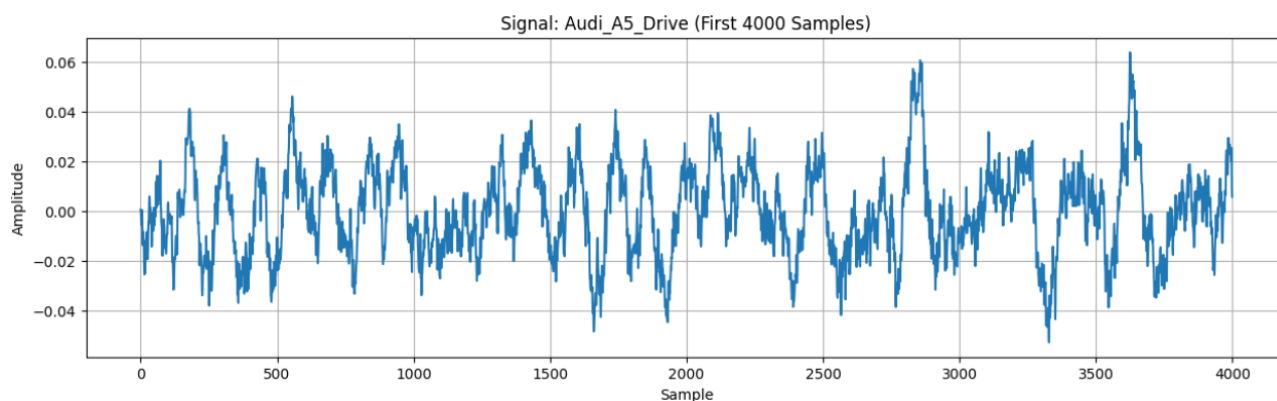
### 4.1 Výpočet vzdáleností a korelací

Pro každou dvojici (testovací, referenční) signálů se spočítá *eukleidovská vzdálenost* a *Pearsonův korelační koeficient* mezi jejich příznakovými vektory. Tyto výpočty jsou realizovány pomocí nainportovaných knihoven, především knihoven `sklearn.metrics.pairwise` a `scipy.stats`.

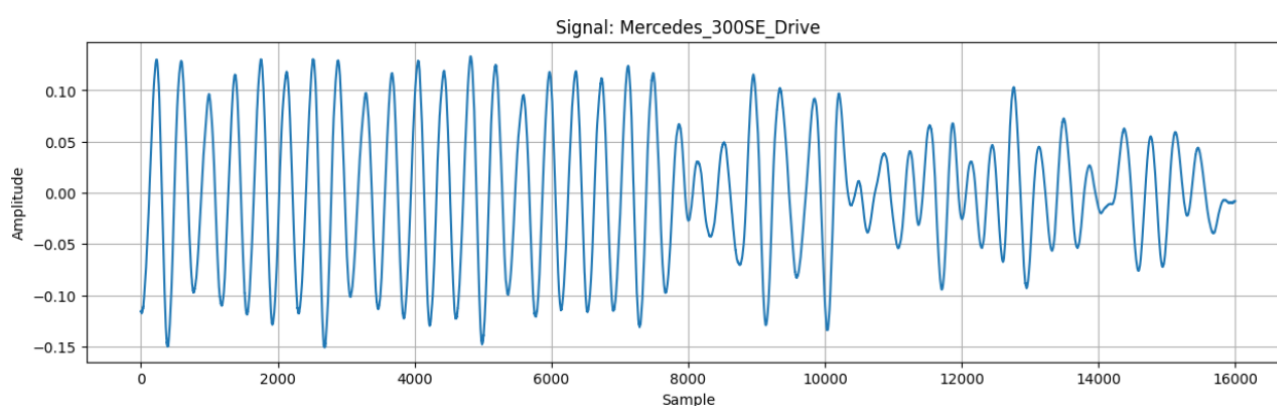
### 4.2 Klasifikace signálů

Klasifikace testovacích signálů se provádí na základě porovnání příznakových vektorů testovacích a referenčních signálů. Proces klasifikace probíhá v několika krocích:

1. Porovnání příznakových vektorů  
Každý testovací signál je porovnán s všemi referenčními signály pomocí eukleidovské vzdálenosti a Pearsonova korelačního koeficientu.
2. Kritéria klasifikace  
Pokud vzdálenost mezi testovacím a referenčním signálem je  $\leq 200$  a korelační koeficient  $\geq 0.95$ , pak je signál považován za definitivní shodu, tedy `definite match`.



Obrázek 7: Audi\_A5\_Drive



Obrázek 8: Mercedes\_300SE\_Drive

### 3. Identifikace nejlepší shody

Pro každý testovací signál se identifikuje nejbližší referenční signál na základě kombinace nejmenší vzdálenosti a nejvyšší korelace.

### 4. Označení `no match`

Pokud žádný referenční signál nesplňuje kritéria, je testovací signál označen jako `no match`.

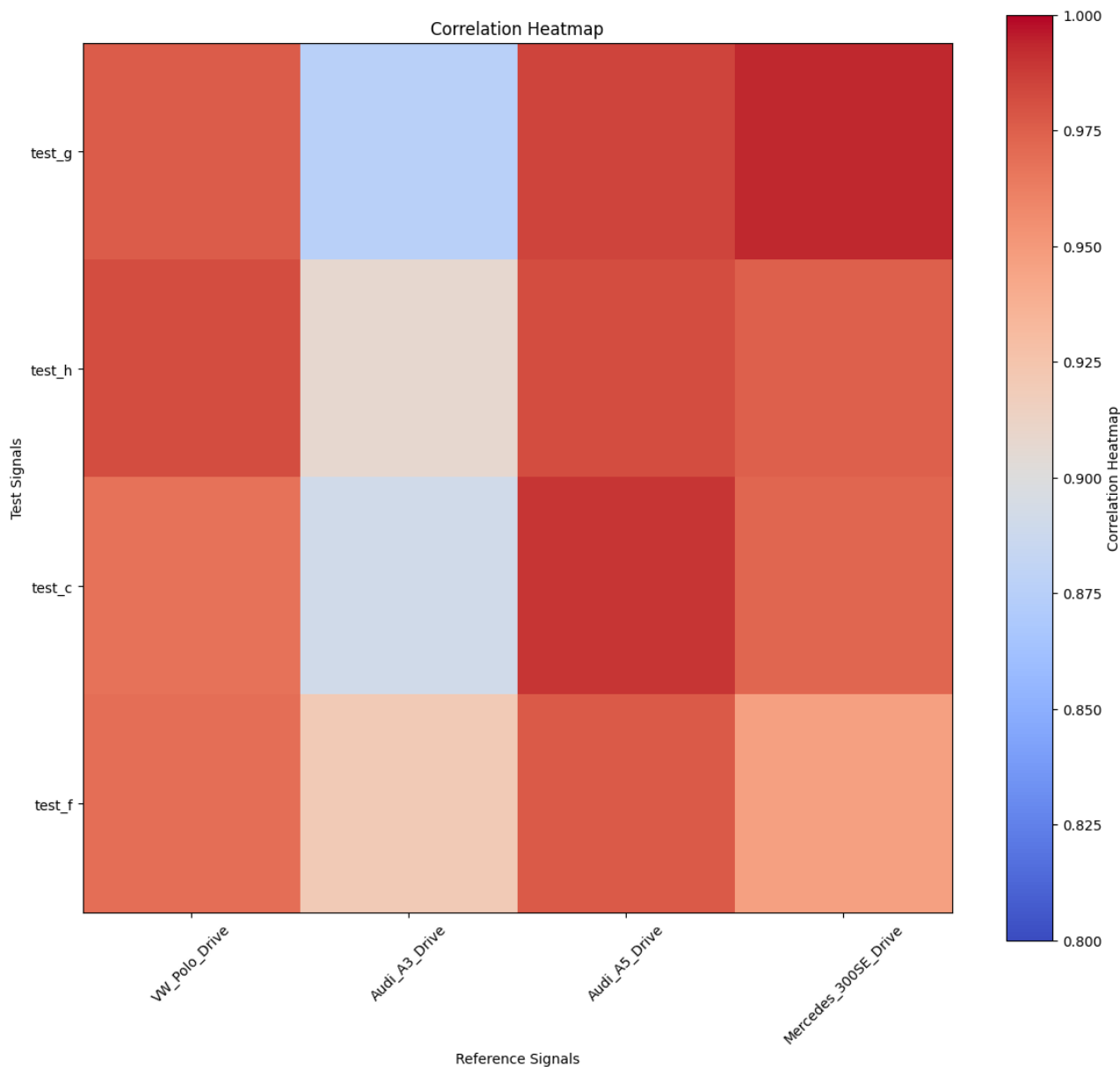
Celý proces klasifikace je navíc optimalizován tak, aby se zabránilo duplicitnímu přiřazení jednoho referenčního signálu k více testovacím signálům. Tento krok zajišťuje, že každý referenční signál může být použit jen jednou.

## 4.3 Výstup

Program generuje seznam testovacích signálů, které byly označeny jako `definite match` a `no match`. Dále také generuje dva grafy pro lepší vizualizaci a to graf *euklidovských vzdáleností* a *koleračního koeficientu*.

## 5 Výsledky

V rámci výsledků program kromě samostatného přiřazení generuje i dva výše zmíněné grafy pro lepší vizualizaci.



Obrázek 9: Graf korelačního koeficientu pro jednotlivé signály

Samotný výsledek přiřazování jednotlivých nahrávek vypadá následovně.

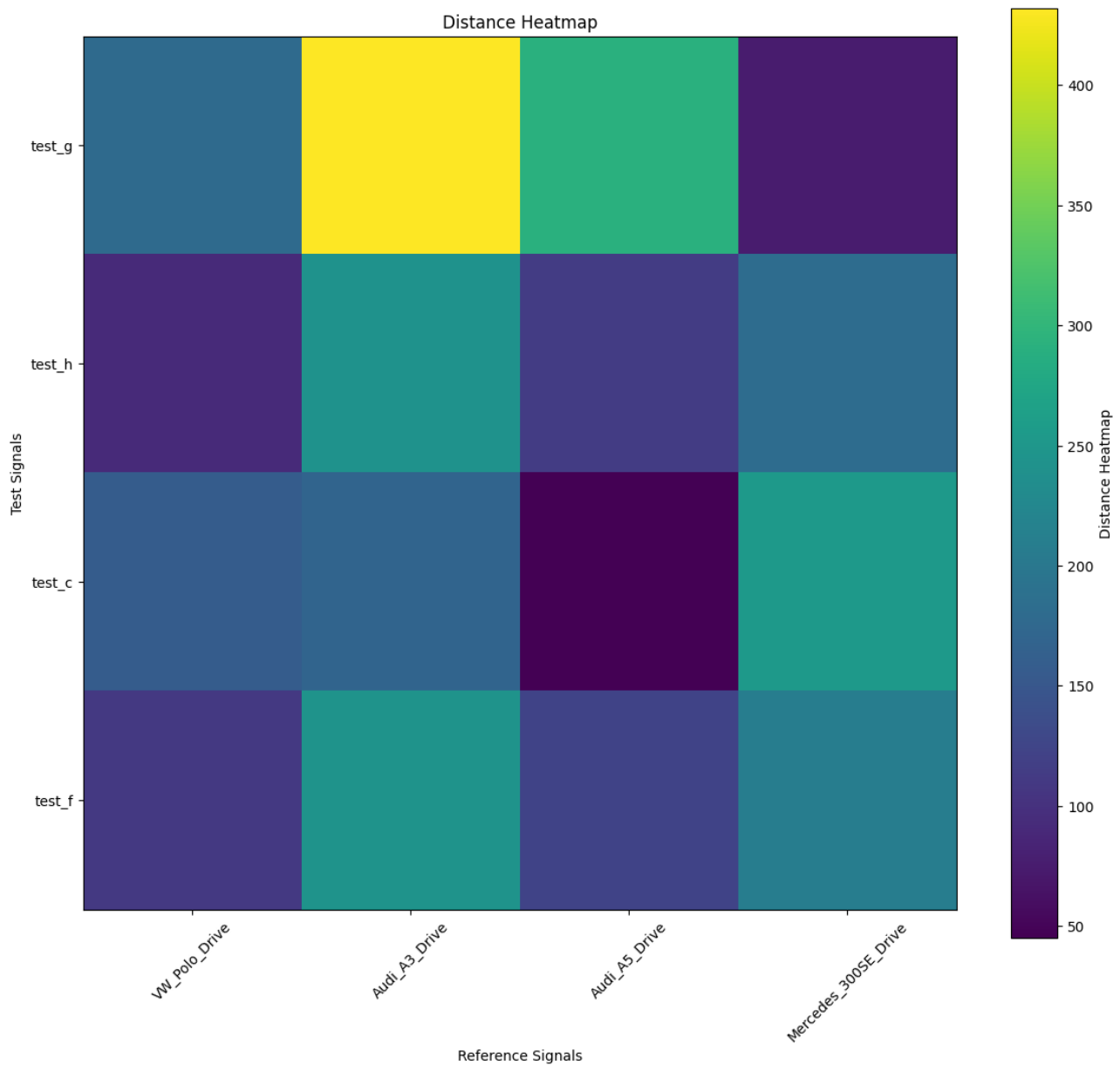


## 6 Závěr

Výsledky analýzy signálu s programem se výrazně liší od mého původního předpokladu vyvozeného z poslechu a pouhého pohledu na vzhled křivky signálů.

## Zdroje

```
@misc{Wikipedia_2024b, url={https://cs.wikipedia.org/wiki/Korelace_(zpracov%C3%A1n%C3%AD_sigu%C3%BDlu)}, title = Korelace (zpracování signálu), journal = Wikipedia, publisher = Wikimedia Foundation, year = 2024, month = Nov}
@miscDataMListic2022, title = MelFrequencyCepstralCoefficients(MFCC)Explained, url = https://www.youtube.com/watch?v=SJo7vPgRlBQ, journal = YouTube, publisher = YouTube, author = DataMListic, year = 2022, month = Nov}
@miscWikipedia2024, url = https://en.wikipedia.org/wiki/Mel-frequency_cepstrum, journal = Wikipedia, publisher = Wikimedia Foundation, year = 2024, month = Nov}
@miscPyTorch2024, title = Mel-frequencycepstralcoefficients(MFCC)forspeechrecognition, url = https://www.geeksforgeeks.org/mel-frequency-cepstral-coefficients-mfcc-for-speech-recognition/, journal = GeeksforGeeks, publisher = GeeksforGeeks, author = PyTorch, year = 2024, month = Jun}
```



Obrázek 10: Graf euklidovské vzdálenosti pro jednotlivé signály

```

----- DEFINITE MATCHES -----
Test signal 'test_g' is definitely matched to reference 'Mercedes_300SE_Drive'
Test signal 'test_c' is definitely matched to reference 'Audi_A5_Drive'
Test signal 'test_h' is definitely matched to reference 'VW_Polo_Drive'
Test signal 'test_f' has no match

```

Obrázek 11: Výsledek