Univerzita Jana Evangelisty Purkyně, Ústí nad Labem

Počítačové zpracování signálu KI/PZS

Seminární práce

# Detekce a přiřazení slov v záznamu řeči

|  |  |
| --- | --- |
| ZS 2023/24 | Tomáš Ulrich  Osobní číslo: F21281 |

## Zadání

Ve zdrojovém souboru *Signal1.txt* a na Obrázku 1 najdete záznam řeči převedený na signál, resp. časovou řadu. Pomocí metod analýzy signálu v časové oblasti, frekvenční oblasti nebo jejich kombinací identifikujte jednotlivá slova v záznamu. Vybírejte z následujícího seznamu slov.

*time, prepare, solution, make, mistake, no, the, probable, long, lecture, method, disaster, fail, work, advice, idea, succeed, easy, is, for, give.*

Vámi navržený algoritmus vyzkoušejte na záznamech *Signal2.txt* a *Signal3.txt* a identifikujte slova i v těchto dvou časových řadách. V případě potřeby algoritmus dále vylepšete. Kromě metod probraných při hodinách lze pro identifikaci jednotlivých slov využít například některé další funkce, například Hammingova funkce nebo Hilbertova transformace, případně jakoukoli další metodu, kterou uznáte za vhodnou, vyjma metod založených na strojovém učení.

Vzorkovací frekvence signálu je ve všech případech 22050 Hz.

Obsah obrázku snímek obrazovky, řada/pruh, Vykreslený graf

Popis byl vytvořen automaticky

Obrázek 1: Záznam řeči ze souboru *Signal1.txt*.

## Postup

### Příprava referenčních slov

Náš seznam slov převedeme na hlasovou stopu a uložíme je do souborů formátu WAV. Tím vytvoříme slovník zvuků. Pro převod textu do hlasového záznamu jsme využili knihovnu pyttsx3.

### Zpracování signálu

Načtený signál trimujeme – odstranění tichých částí signálu s amplitudou nižší než 50 Hz. Tyto části se nahradili nulovými hodnotami. Dále jsme signál rozdělili do oken, která obsahují hlasovou aktivitu. Signál jsme tak rozdělili na jednotlivá slova.

### Porovnání oken s referenčními slovy

Postupně procházíme všechny části textových dat a snažíme se rozpoznat obsah každé části signálu pomocí porovnání s referenčními slovy. Následně se provedlo rozdělení do oken, která obsahují hlasovou aktivitu. Tím jsem dosáhl rozdělení signálu na jednotlivá slova.

Pro každý upravený úsek textových dat a pro každé slovo provedeme extrakci hlasových charakteristik. Využíváme knihovnu **librosa**, která automaticky vypočítává Mel-frekvenční kepstrální koeficienty, spektrální centroid a chromu pro oba zvukové signály.

Poté provádíme výpočet spektrální vzdálenosti mezi hlasovými charakteristikami daných slov a charakteristikami hlasu v daném okně signálu, a to pomocí knihovny **scipy**. To zahrnuje výpočet vzdáleností mezi kepstrálními koeficienty, spektrálním centroidem a chromou obou signálů.

#### Mel-frekvenční spektrogram

Grafická reprezentace ukazující distribuci energie signálů v různých frekvenčních pásmech v lidském sluchovém rozsahu. Využívá mel-škály, což je transformace frekvencí do mel-frekvenčního prostoru, lépe odpovídajícího vnímání lidského sluchu.

#### Mel-frekvenční kepstrální koeficienty

Hodnoty odvozené z mel-frekvenčního spektogramu, představující krátkodobou spektrální energii hlasového signálu. Což je užitečné při analýze krátkých časových úseků zvukového materiálu.

#### Spektrální centroid

Míra používaná v digitálním zpracování signálů k charakterizaci spektra. Ukazuje, kde se nachází hmotnostní střed spektra. Někdy je nazýván střední hodnota spektrální hmotnosti.

#### Chromatogram

Chroma je distribuce energie v různých tónech nezávislých na jejich výšce.

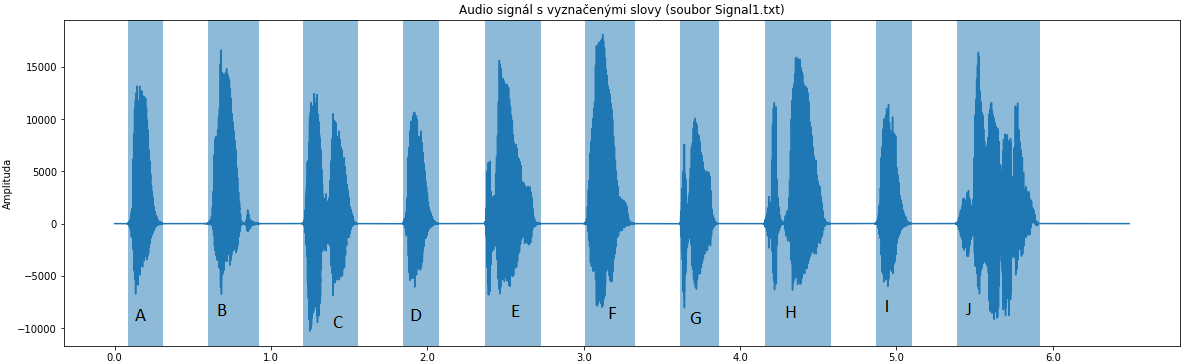
### Statistiky slov

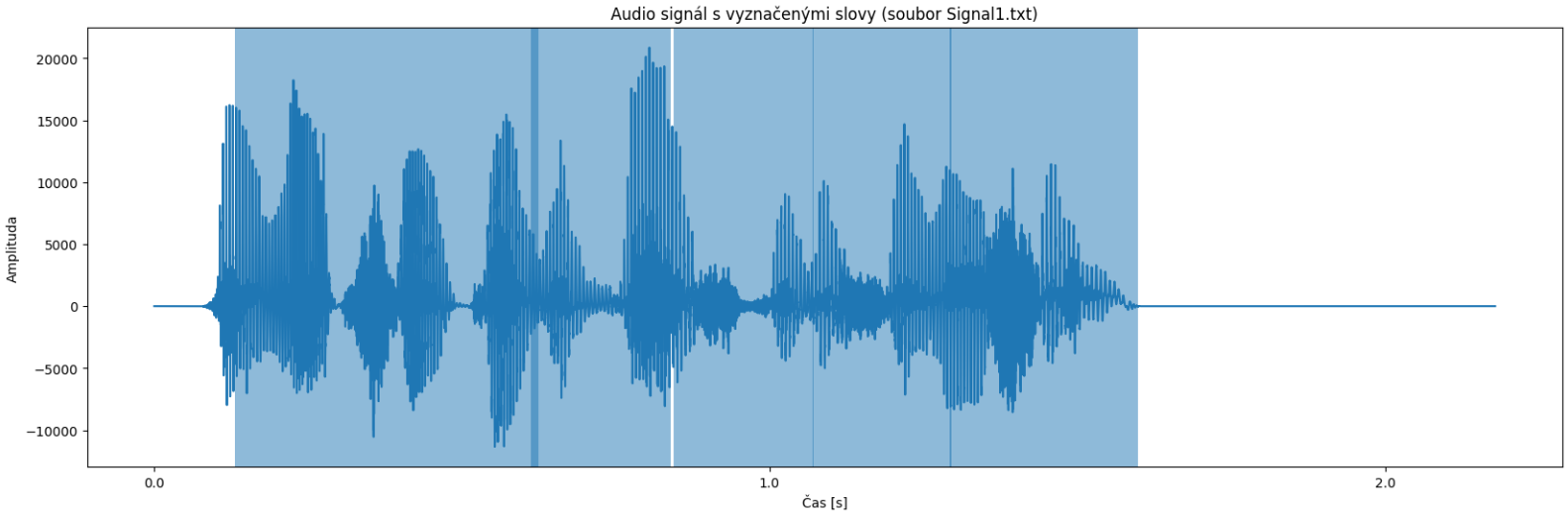
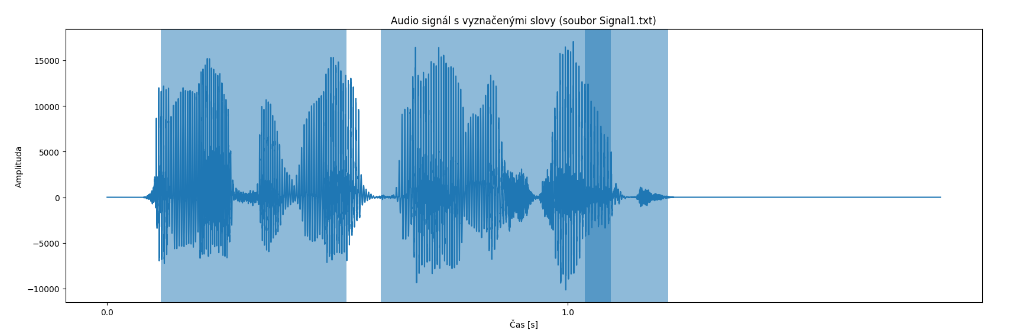
Nakonec budeme vypočítávat pravděpodobnosti výskytu jednotlivých slov v datech na základě podobnosti jejich hlasových charakteristik. Tím získáme informace o nejpravděpodobnějších slovech v hlasových datech.

## Výsledky

Začali jsme zpracování signálu pro každý soubor. Na grafech 1-3 jsou vizualizovány jednotlivé signály rozdělené do oken odpovídajících jednotlivým slovům.

**Graf 1**: soubor Signal1.txt





**Graf 2**: soubor Signal2.txt

E

D

C

B

A

C

B

A

**Graf 3**: soubor Signal3.txt

V tabulkách níže jsou uvedeny pravděpodobnosti výskytu slov (Pro každé okno jsou vybrány tři nejpravděpodobnější slova dle jejich spektrálních vzdáleností):

|  |  |  |  |
| --- | --- | --- | --- |
| **Tabulka 1**: výsledky pravděpodobností slov pro jednotlivá okna signálu Signal1.txt | | | |
| **slovo** | **Pravděpodobnost [%]** | **slovo** | **Pravděpodobnost [%]** |
| **Blok A:** | | **Blok B:** | |
| *The* | 80.8 | *Work* | 66.9 |
| *No* | 29.2 | *Make* | 35.1 |
| *Easy* | 18.7 | *No* | 17.7 |
| **Blok C:** | | **Blok D:** | |
| Easy | 85.0 | The | 67.0 |
| Idea | 25.2 | No | 31.7 |
| Make | 20.6 | Long | 18.7 |
| **Blok E:** | | **Blok F:** | |
| Time | 63.9 | No | 41.7 |
| Give | 23.9 | Work | 28.4 |
| Easy | 23.6 | Make | 24.5 |
| **Blok G:** | | **Blok H:** | |
| The | 38.9 | Long | 31.9 |
| No | 27.0 | Idea | 31.0 |
| Time | 26.9 | Prepare | 27.4 |
| **Blok I:** | | **Blok J:** | |
| The | 66.5 | Solution | 36.6 |
| No | 33.9 | For | 27.8 |
| Long | 18.0 | Fail | 27.3 |

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **Tabulka 2**: výsledky pravděpodobností slov pro jednotlivá okna signálu Signal2.txt | | | | | |
| **slovo** | **Prav.** | **slovo** | **Prav.** | **slovo** | **Prav.** |
| **Blok A:** | | **Blok B:** | | **Blok C:** | |
| Idea | 40.2 | Advice | 35.1 | The | 33.9 |
| Long | 28.1 | Method | 28.2 | Easy | 29.3 |
| Probable | 25.3 | Prepare | 27.7 | No | 27.4 |

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **Tabulka 3**: výsledky pravděpodobností slov pro jednotlivá okna signálu Signal3.txt | | | | | |
| **slovo** | **Prav.** | **slovo** | **Prav.** | **slovo** | **Prav.** |
| **Blok A:** | | **Blok B:** | | **Blok C:** | |
| Prepare | 30.8 | The | 35.8 | Easy | 33.0 |
| Lecture | 30.3 | No | 31.6 | The | 30.9 |
| Method | 28.9 | Long | 24.8 | No | 26.8 |
| **Blok D:** | | **Blok E:** | |  |  |
| Easy | 31.4 | Easy | 47.2 |  |  |
| The | 30.6 | Idea | 27.6 |  |  |
| No | 28.1 | Make | 23.5 |  |  |

## Závěr

Můžeme si všimnout, že soubor Signal1.txt byl úspěšně zpracován díky výrazným mezerám mezi slovy. U ostatních souborů jsme doplatili na nedostatek algoritmu dynamickém výběru oken v signálu. Díky tomu, že na sebe slova těsně navazují a vzájemně se překrývají, tak algoritmus není schopen vytvořit optimální výběr oken.

S řešením pomohl kolega Alex Schönfelder, tato práce vznikla díky jeho výpomoci. Proto jsem použil stejnou tabulku, jelikož jsou hodnoty stejné (Pouze jsem ji graficky upravil).