



Домаћи задаци из Обраде и препознавања говора 2022/2023

Напомена: Приликом предаје домаћих задатака потребно је припремити један збирни извештај за сва три домаћа задатка. Збирни извештај, осим текста задатка, добијених резултата са коментарима и одговорима на сва питања, треба да садржи и кратак преглед релевантне теорије. Домаћи се предаје на MSTeams платформи у секцији Assignments.

Први задатак:

1. Коришћењем комерцијалног микрофона у програмском окружењу МАТЛАБ, снимити говорну секвенцу у дужини од 20-ак секунди. Секвенцу снимити са фреквенцијом одабирања 8 или 10 kHz и она треба да се састоји десетак јасно сегментираних речи.
2. Коришћењем кратковременске енергије и кратковременске брзине проласка кроз нулу извршити одређивање почетка и краја појединих речи. Добијени резултат приказати графички. Преслушати сегментиране делове звучне секвенце и коментарисати добијени резултат. (По жељи се овај поступак може поновити применом Teager енергије).
3. Снимити нову секвенцу од пар речи (богатих самогласницима, рецимо ономотопеја...) и на основу тако снимљене секвенце проценити pitch периоду сопственог гласа. Користити две различите методе па упоредити и коментарисати добијене резултате.

Други задатак:

1. Коришћењем комерцијалног микрофона у програмском окружењу МАТЛАБ, снимити говорну секвенцу у дужини од 20-ак секунди. Секвенцу снимити са фреквенцијом одабирања 8 kHz у шеснаестобитној (default) резолуцији.
2. Испројектовати $\mu = 100$ и $\mu = 500$ компандинг квантизатор са 4, 8 и 12 бита и за њих одредити зависност односа сигнал-шум за различите вредности односа (X_{max}/σ_x). Овај однос мењати променом варијансе корисног сигнала, простим скалирањем почетне снимљене секвенце. Приказати резултате графички.
3. Испројектовати Делта квантизатор за секвенцу из тачке 1. Адекватно подесити параметар Δ тако да се добије што бољи квалитет квантизације. Упоредити облике оригиналног и квантизованог сигнала. Шта се дешава када је корак квантизације Δ превише мали или превише велики? Да ли се хистограм прираштаја може користити за одређивање адекватног параметра Δ ? Пратити квалитет звука и промене у амплитуди за сваки случај.

Трећи задатак:

Снимити базу са 3 изговорене цифре, где је свака цифра изговорена 10 пута од стране истог говорника (30 секвенци у бази).

1. Написати функцију *preprocessing* која прима говорну секвенцу и враћа је након извршене предобраде (сегментација и филтрирање).
2. Имплементирати функцију *feature_extraction* која за прослеђену секвенцу враћа обележја заснована на LPC и/или кепстралним коефицијентима (дозвољено је коришћење уграђених функција уз теоријски опис).
3. Коначна функција *cifer_recognition* треба да покрене код за снимање говорне секвенце, и затим да обележја снимљене секвенце добијена на основу функција из тачака 1. и 2. проследи класификатору по избору. Када класификатор донесе одлуку, исписати је у командном прозору.
4. Успешност класификације тестирати на по 5 новоснимљених секвенци из сваке класе и приказати у облику конфузионе матрице. Такође приказати конфузиону матрицу за тренинг скуп. За сваку од наведених тачака дати сажет преглед теорије на којој се заснива, као и детаљан опис имплементације саме функције. Резултате сваке тачке приказати графички на одабраној секвенци и прокоментарисати утицај избора обележја и класификатора на исход класификације. Издвојити и прокоментарисати примере тачно и погрешно класификованих секвенци.