

تمرین سوم پیاده سازی پردازش گفتار

بازشناسی ارقام مجزا

۱) اعداد صفر تا نه را ضبط نمایید طوری که هر عدد دو بار به تنهایی گفته شود. به این ترتیب بیست فایل گفتاری خواهید داشت. یکبار تلفظ اعداد صفر تا نه (۱۰ فایل) را به عنوان مجموعه مرجع و بار دوم تلفظ آنها (۱۰ فایل دیگر) را به عنوان مجموعه تست در نظر بگیرید. بازشناسی را در شرایط ساینز قاب ۳۰ میلی ثانیه و گام قاب بندی ۱۰ میلی ثانیه با پنجره همینگ و با استفاده از ویژگیهای زیر و فاصله کپسترال انجام دهید و نتایج بازشناسی را در جدولی مانند ضمیمه ذخیره کنید. کد نوشته شده و جدول ضمیمه را به همراه کد اپلود نمایید.

(الف) ۱۲ ویژگی کپستروم حقیقی

(ب) ۱۲ ضریب مل کپستروم

(ج) ۱۲ ضریب مل کپستروم و انرژی

(د) ۱۲ ضریب مل کپستروم و انرژی و مشتقات اول آنها (بعد بردار ویژگی: ۲۶)

(ه) ۱۲ ضریب مل کپستروم و انرژی و مشتقات اول و دوم آنها (بعد بردار ویژگی: ۳۹)

(و) ویژگیهای استخراجی از آخرین بلوک ترنسفورمر مدلهای Wav2vec2.0 و HuBERT که در راستای زمان میانگیری شده باشند. در این حالت قاب بندی توسط خود مدل انجام می شود و ورودی مدل کل فایل صوتی است (بعد بردار ویژگی نهایی ۷۶۸ خواهد بود).

درصد بازشناسی	نام ویژگی

موفق باشید

ناصر شریف