



۱. (۱۰٪) [پژوهش]

همانگونه که در کلاس بیان شد، پردازش گفتار در کاربردهای مختلف کاربرد دارد. هدف این سوال مروری بر برخی از کاربردهای دیگر پردازش گفتار است که در اسلایدهای کلاس ارائه نشده است. هر دانشجو دست کم یک موضوع (کاربرد یا روش) را بررسی کند. پس از مطالعه منابع مرتبط (مقاله، پایان نامه، گزارش و ...) از آنها یک گزارش کوتاه تهیه کنید. در این گزارش کوتاه باید به چستی موضوع، روش و مدل بکار بسته شده، دادگان، نتایج اشاره کند (و هرچیز پایه‌ای به فراخور بودن در منبع آن موضوع برگزیده شده). توجه شود که برای [هر] موضوع، دست کم یک منبع معتبر را که در گزارش خود از آن بهره برده‌اید، همراه پاسخ بفرستید.

در انجام این بخش کوشا باشید و دقت به خرج دهید. چه بسا در برگزیدن موضوع پروژه نهایی تان کمک کننده باشد.

۲. (۱۰٪) [آشنایی با یک سیستم بازشناسی گفتار]

در این سوال می‌خواهیم با یک سیستم بازشناسی گفتار آشنا شویم. برای این کار با انتخاب‌های که در تلفن‌های هوشمند خود یا... دارید می‌توانید این بخش را انجام دهید. پیشنهاد ما به Google Speech Service یا سایت SpeechTexter.com است. اگر تلفن‌های هوشمند اپل دارید، این بخش را در یک تلفن هوشمندی که اندرویدی است انجام داده و گزارش دهید.

الف- تلاش کنید تا در شرایطی تقریباً ایده‌آل (نگفتن واژگان سنگین و غریب، نبود نویزهای محلی و...) ۵ جمله با طول حدودی ۱۰ واژه را به سامانه برگزیده شده بگویید تا آن‌ها را برای شما تایپ کند. اصل جمله‌های خود را با آنچه سیستم به شما بر می‌گرداند، مقایسه کنید و در گزارش خود بیاورید. همچنین در مورد چرایی رخداد WER در یک سیستم بازشناسی گفتار، پژوهشی انجام دهید.

ب- خطاهای رخ داده‌شده را برای هر واژه در جمله‌ای که گفته‌اید را با سنج $Word Error Rate$ (WER) گزارش کنید.



یادداشت: *Word Error Rate*, نرخ خطاهای رخ داده‌شده در رونویسی (*transcription*) بدست آمده از یک سیستم بازشناسی گفتار است. این خطاها می‌تواند نادرست نویسی واج‌ها در یک واژه (*S*), درج نادرست واژه‌های نگفته شده (*I*) و واژگان حذف شده (*D*) باشد. بنابراین:

$$WER = \frac{D + I + S}{X}$$

در این رابطه، X شمار واژگان گفته‌شده است.

از آنجایی که با یک نرخ سر و کار داریم، در صورت رخدادن خطا، برای هر واژه نهایتاً عدد یک را در نظر بگیرید.

پ- در تلاشی دیگر، یک جمله بگویید و بعد آن را اصلاح کنید (برای نمونه به سیستم مورد نظر بگویید: "open the firefox, no sorry, the chrome"). رفتار سیستم را برای این حالت گزارش دهید. این کار را با یک چت‌بات (اگر با ویندوز هستید، Cortana و اگر با مک هستید، Siri) نیز انجام دهید (تنها این بخش چت‌بات می‌تواند به زبان انگلیسی باشد) و نتیجه را گزارش دهید.

۳. (۱۵٪) [آشنایی با کتابخانه‌ی numpy]

مطالب کمکی این سوال به همراه تمرین ارسال شده است. در این پوشه وارد فایل `Python_Basics_With_Numpy.ipynb` شوید. وظیفه شما کامل کردن قسمت‌هایی است که به شما گفته شده است.

```
### START CODE HERE ### (≈ 1 line of code)
|
### END CODE HERE ###
```

برای مثال:

We will often specify "(≈ X lines of code)" in the comments to tell you about how much code you need to write. It is just a rough estimate, so don't feel bad if your code is longer or shorter.

Exercise: Set test to "Hello World" in the cell below to print "Hello World" and run the two cells below.

```
>
|
| ### START CODE HERE ### (≈ 1 line of code)
| |
| ### END CODE HERE ###
|
| Python
|
| print("test: " + test)
| print(test)
```



و شما باید آن را بدین گونه پر کنید:

```
Exercise: Set test to "Hello World" in the cell below to print "Hello World" and run the two cells below.
+ Code + Markdown

## START CODE HERE ## (= 1 line of code)
test = "Hello World"
## END CODE HERE ##

[1] ✓ 0.0s Python

print("test: " + test)
print(test)

[2] ✓ 0.0s Python
... test: Hello World
Hello World
```

توجه کنید که اگر مثلاً متغیر خروجی تابع z است حتماً شما هم باید آخرین مقدار نهایی را در z ذخیره کنید.

۴. (۶۵٪) [پایه‌سازی: سنتز گفتار فارسی]

در این سوال می‌خواهیم یک برنامه بنویسید که قادر باشد اعداد صفر تا یک میلیون را سنتز کند. الف) ورودی این برنامه یک عدد (به صورت عددی یا حروفی) است (مانند ۱۴۰۹ یا دویست و دو هزار و ده) و خروجی آن فایل صوتی معادل است. برای این کار قطعاتی از کلمات لازم را ضبط کرده (نرخ نمونه‌برداری 16 KHz، چندی‌سازی 16 Bits، به صورت مونو و فرمت Wav) و با اتصال آنها به هم عدد درخواستی را بسازید. تعداد قطعات ضبط شما چند تا است؟ علاوه بر ارسال کد و قطعات ضبط شده، فایل‌های خروجی برنامه خود را برای دو عدد ۱۴۰۹ و دویست و دو هزار و ده نیز ارسال کنید.

ب) سیستم سنتز گفتار خود را به صورتی تغییر دهید که مبتنی بر واج باشد. برای این کار همه واج‌های فارسی را که در جدول زیر آورده شده، هر کدام را در یک فایل، ضبط کنید (نرخ نمونه‌برداری 16 KHz، چندی‌سازی 16 Bits، به صورت مونو و فرمت Wav). قسمت‌های سکوت ابتدا و انتهای هر کدام از فایل‌ها را به صورت دستی حذف کنید. با پشت سر هم قرار دادن این واج‌ها، سیگنال صوتی دو عدد ۱۴۰۹ و دویست و دو هزار و ده را تولید کنید و به عنوان خروجی ارسال کنید. کیفیت این دو فایل چطور است؟! مزیت این روش این است که علاوه بر اعداد، می‌توانید هر کلمه دیگری در زبان فارسی را سنتز کنید، برای تست این موضوع فایل صوتی سه عبارت/کلمه "دانشگاه تهران"، "به کجا چنین شتابان" و اسم و فامیل خودتان را تولید کنید و ارسال کنید. برای افزایش قابل فهم بودن صداهای تولیدی در این حالت چه روشی را پیشنهاد می‌کنید؟



همه واجهای ضبط شده و صداهاى تولیدی اعداد و کلمات را به همراه پاسخ این تمرین ارسال کنید.

#	IPA	Char	Code	Farsi Letter	Phonetic Description
1	I	i	105	ای	high front unrounded
2	e	e	101	اِ	mid front unrounded
3	a	a	97	آ	low front unrounded
4	u	u	117	او	high back rounded
5	o	o	111	اُ	mid back rounded
6	/	/	47	آ	low back rounded
7	p	p	112	پ	unvoiced bilabial plosive
8	b	b	98	ب	voiced bilabial plosive
9	t	t	116	ت، ط	unvoiced dental plosive
10	d	d	100	د	voiced dental plosive
11	k	k	107	ک	unvoiced velar plosive
12	g	g	103	گ	voiced velar plosive
13	q	q	113	ق، غ	voiced uvular plosive
14]]	93	أ، ؤ، ع	glottal stop
15	\$	\$	36	چ	unvoiced alveopalatal affricate closure
16	,	,	44	ج	voiced alveopalatal affricate
17	f	f	102	ف	unvoiced labiodental fricative
18	v	v	118	و	voiced labiodental fricative
19	s	s	115	س، ث، ص	unvoiced alveolar fricative
20	z	z	122	ز، ذ، ظ، ض	voiced alveolar fricative
21	.	.	46	ش	unvoiced alveopalatal fricative
22	[[91	ژ	voiced alveopalatal fricative
23	x	x	120	خ	unvoiced uvular fricative
24	h	h	104	ح، ه	unvoiced glottal fricative
25	l	l	108	ل	lateral alveolar
26	r	r	114	ر	trill alveolar
27	m	m	109	م	nasal bilabial
28	n	n	110	ن	nasal alveolar
29	y	y	121	ی	approximant palatal