VA #7 Comparing Countries English Vocalization

Group 1: Nina Nguyen, Melissa McClure, Nolan Orloff, Louie Wong

1, Metadata exploration

In [97]:
```
import pandas as pd
import numpy as np
```

In [77]:
```
df = pd.read_csv('./speech-accent-archive/speakers_all.csv')
```

In [78]:
```
import pandas as pd
```

In [79]:
```
df = df[df['file_missing?']==False].iloc[:, :8]
df.head()
```

Out[79]:

	age	age_onset	birthplace	filename	native_language	sex	speakerid	country
32	27.0	9.0	virginia, south africa	afrikaans1	afrikaans	female	1	south africa
33	40.0	5.0	pretoria, south africa	afrikaans2	afrikaans	male	2	south africa
34	43.0	4.0	pretoria, transvaal, south africa	afrikaans3	afrikaans	male	418	south africa
35	26.0	8.0	pretoria, south africa	afrikaans4	afrikaans	male	1159	south africa
36	19.0	6.0	cape town, south africa	afrikaans5	afrikaans	male	1432	south africa

2, Audio file exploration

In [65]:
```
%%bash
cd speech-accent-archive/recordings/recordings
ls -l | head -5
```

```
total 1865360
-rw-rw-r--@ 1 louiewhw  staff   333530 Sep 21 16:16 afrikaans1.mp3
-rw-rw-r--@ 1 louiewhw  staff   352756 Sep 21 16:16 afrikaans2.mp3
-rw-rw-r--@ 1 louiewhw  staff   431332 Sep 21 16:16 afrikaans3.mp3
-rw-rw-r--@ 1 louiewhw  staff   376998 Sep 21 16:16 afrikaans4.mp3
```

In [88]:
```
import os
import glob
import urllib
import scipy.io.wavfile
import pydub

audiolist = glob.glob("./speech-accent-archive/recordings/recordings/*.mp3")
audiolist[:5][0].split('/')[4].split('.')[0]
```

Out[88]:
```
'kikongo1'
```

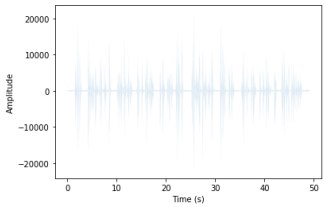
2a. mp3 -> wav

In [91]:
```
for i in audiolist:
    filename = i.split('/')[4].split('.')[0]
    sound = pydub.AudioSegment.from_mp3(i)
    sound = sound.export("wav/" + filename + ".wav", format="wav")
```

2b. Sample plotting

In [115]:
```
import matplotlib.pyplot as plt
import plotly.express as ex
rate,audData=scipy.io.wavfile.read("wav/kikongo2.wav")
time = np.arange(0, float(audData.shape[0]), 1) / rate
plt.plot(time, audData, linewidth=0.01, alpha=0.7)
plt.xlabel('Time (s)')
plt.ylabel('Amplitude')

plt.show()
```



In [ ]:
```
```