

```
In [51]: import os
import librosa
import numpy as np
import pandas as pd
import matplotlib.pyplot as plt
import seaborn as sns
import speech_recognition as sr
from sklearn.cluster import KMeans
from sklearn.preprocessing import StandardScaler
```

```
In [52]: #pip install librosa
#pip install SpeechRecognition
```

```
In [53]: # different functions for audio extractions
def extract_audio_features(audio_path):
    y, sr_rate = librosa.load(audio_path, sr=None)
    duration = librosa.get_duration(y=y, sr=sr_rate)
    tempo, _ = librosa.beat.beat_track(y=y, sr=sr_rate)
    pitch = librosa.yin(y, fmin=librosa.note_to_hz('C2'), fmax=librosa.note
    pitch_var = np.var(pitch)
    return {
        "duration": duration,
        "tempo": tempo,
        "pitch_variability": pitch_var,
    }

def transcribe_audio(audio_path):
    recognizer = sr.Recognizer()
    with sr.AudioFile(audio_path) as source:
        audio = recognizer.record(source)
    try:
        return recognizer.recognize_google(audio)
    except sr.UnknownValueError:
        return ""
    except sr.RequestError:
        return ""

def extract_nlp_features(text):
    words = text.split()
    hesitation_words = ["uh", "um"]
    hesitations = sum(text.lower().count(hw) for hw in hesitation_words)
    pauses = text.count("...")
    return {
        "word_count": len(words),
        "pauses": pauses,
        "hesitations": hesitations,
        "speech_rate": len(words) / 60.0, # assuming 1 min length, refine
    }
```

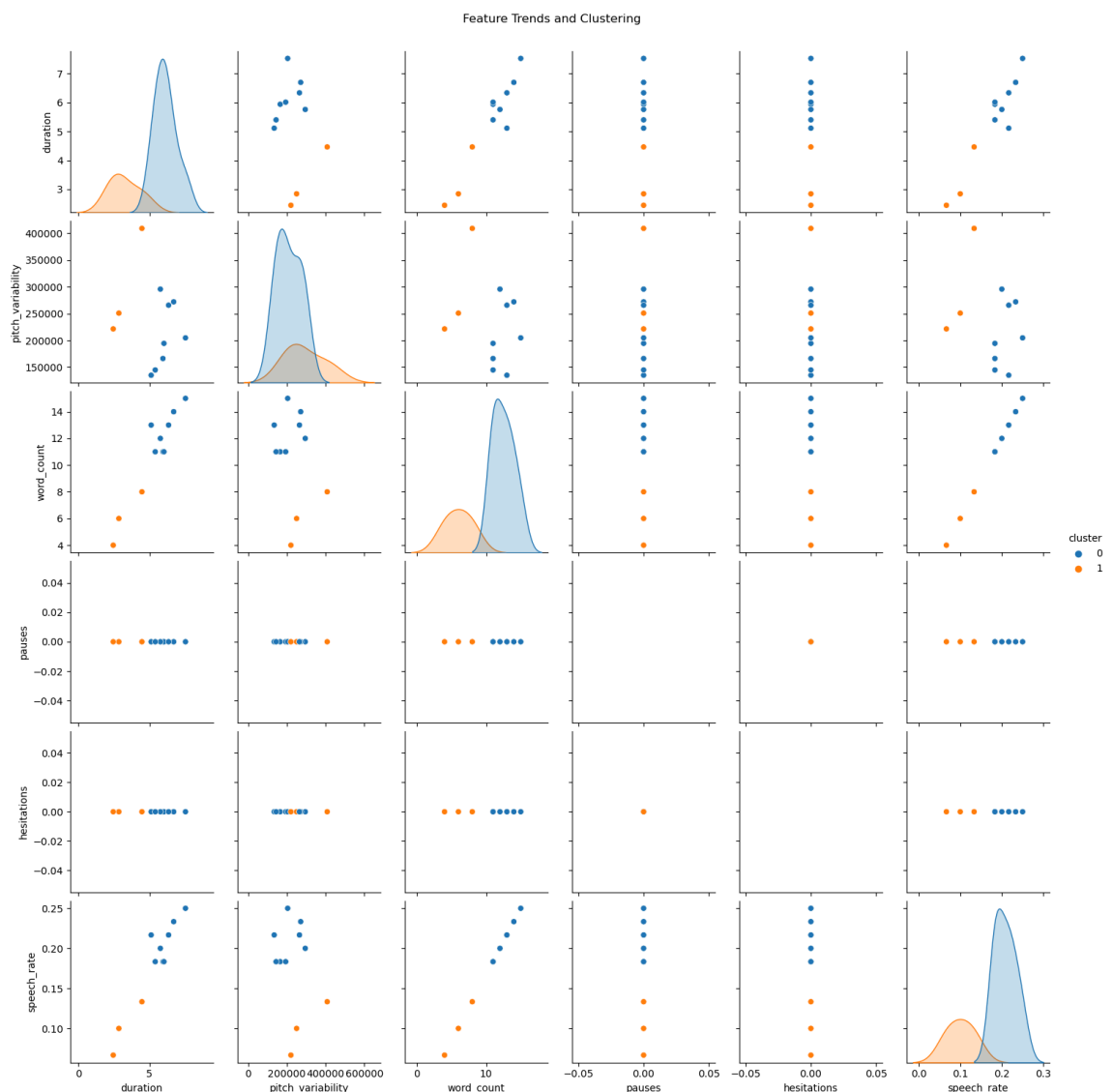
```
In [54]: #processing
audio_folder = "audio_samples/" # Replace with actual path
audio_files = [f for f in os.listdir(audio_folder) if f.endswith('.wav')]

feature_list = []
for file in audio_files:
    path = os.path.join(audio_folder, file)
    audio_feats = extract_audio_features(path)
    text = transcribe_audio(path)
    nlp_feats = extract_nlp_features(text)
    combined = {**audio_feats, **nlp_feats, "file": file}
    feature_list.append(combined)
```

```
In [55]: #data frame creation
df = pd.DataFrame(feature_list)
```

```
In [61]: # Visualizations
sns.pairplot(df, hue="cluster")
plt.suptitle("Feature Trends and Clustering", y=1.02)
plt.show()
```

C:\Users\91709\anaconda3\Lib\site-packages\seaborn\axisgrid.py:118: UserWarning: The figure layout has changed to tight
self._figure.tight_layout(*args, **kwargs)



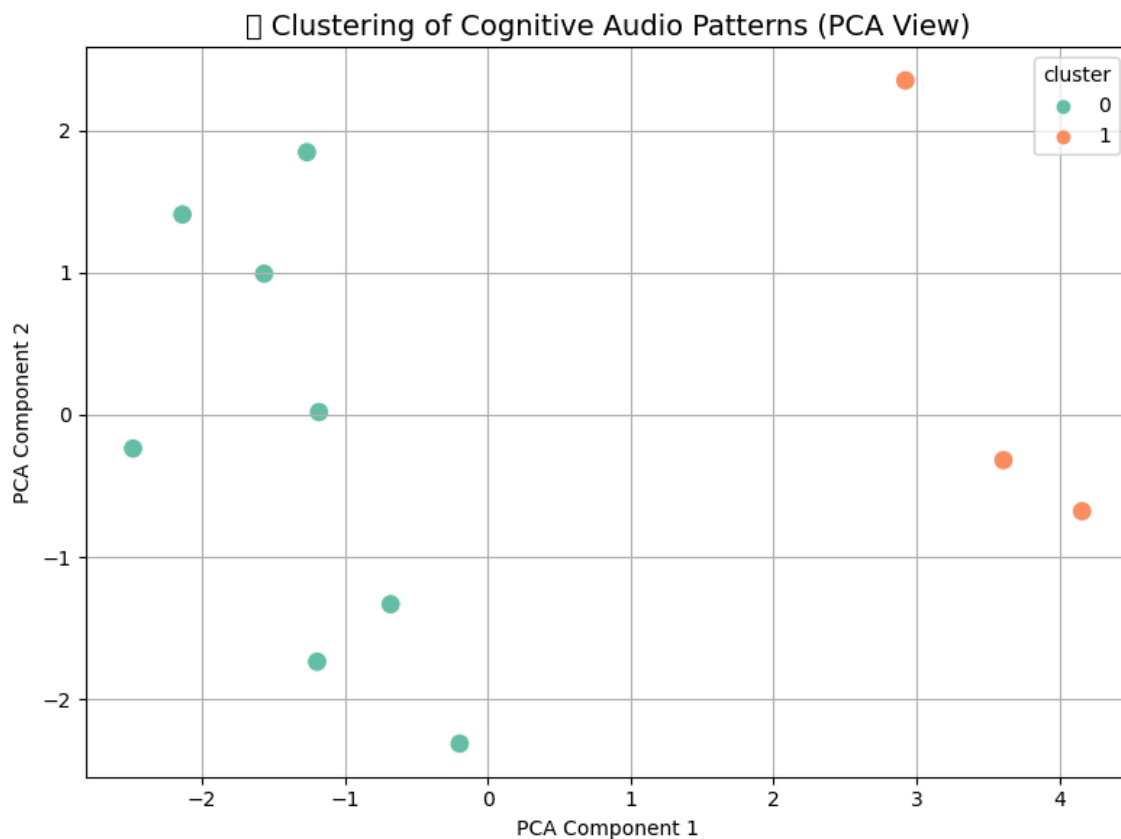
```
In [67]: # Clustering
X = df.drop(columns=["file"])
scaler = StandardScaler()
X_scaled = scaler.fit_transform(X)
kmeans = KMeans(n_clusters=2, random_state=42)
df["cluster"] = kmeans.fit_predict(X_scaled)

# Clustering Plot through PCA
from sklearn.decomposition import PCA

pca = PCA(n_components=2)
X_pca = pca.fit_transform(X_scaled)
df["pca1"] = X_pca[:, 0]
df["pca2"] = X_pca[:, 1]

plt.figure(figsize=(8, 6))
sns.scatterplot(data=df, x="pca1", y="pca2", hue="cluster", palette="Set2",
plt.title("🧠 Clustering of Cognitive Audio Patterns (PCA View)", fontsize=
plt.xlabel("PCA Component 1")
plt.ylabel("PCA Component 2")
plt.grid(True)
plt.tight_layout()
plt.show()
```

```
C:\Users\91709\anaconda3\Lib\site-packages\sklearn\cluster\_kmeans.py:141
2: FutureWarning: The default value of `n_init` will change from 10 to 'au
to' in 1.4. Set the value of `n_init` explicitly to suppress the warning
    super()._check_params_vs_input(X, default_n_init=10)
C:\Users\91709\anaconda3\Lib\site-packages\sklearn\cluster\_kmeans.py:143
6: UserWarning: KMeans is known to have a memory leak on Windows with MKL,
when there are less chunks than available threads. You can avoid it by set
ting the environment variable OMP_NUM_THREADS=1.
    warnings.warn(
C:\Users\91709\AppData\Local\Temp\ipykernel_17008\2623423771.py:22: UserWa
rning: Glyph 129504 (\N{BRAIN}) missing from current font.
    plt.tight_layout()
C:\Users\91709\anaconda3\Lib\site-packages\IPython\core\pylabtools.py:152:
UserWarning: Glyph 129504 (\N{BRAIN}) missing from current font.
    fig.canvas.print_figure(bytes_io, **kw)
```



```
In [62]: # --- Report Summary ---  
print("Insightful Features:", X.columns.tolist())  
print("Clustering Centers:\n", kmeans.cluster_centers_)
```

Insightful Features: ['duration', 'tempo', 'pitch_variability', 'word_count', 'pauses', 'hesitations', 'speech_rate', 'cluster']

Clustering Centers:

```
[[ 0.5253694  0.34509036 -0.30481506  0.54546788  0.          0.  
  0.54546788 -0.61237244]  
[-1.40098507 -0.92024097  0.81284015 -1.45458102  0.          0.  
 -1.45458102  1.63299316]]
```

```
In [68]: # --- Final Report Summary ---

def interpret_cluster(cluster_id):
    if cluster_id == 0:
        return "● Low Risk: Likely fluent, consistent speech"
    else:
        return "● High Risk: Signs of hesitations, pitch variability, slow

# 1. Show summary of features per audio

print("\n--- Individual Audio Reports ---")
for _, row in df.iterrows():
    print(f"\n📄 File: {row['file']}")
    print(f" Cluster: {row['cluster']} → {interpret_cluster(row['cluster'])}")
    for col in X.columns:
        value = row[col]
        if isinstance(value, (np.ndarray, list)):
            value = value[0]
        print(f" {col}: {value:.2f}")

# 2. Show feature importances (clustering centers)
print("\n--- Clustering Centers ---")
centers_df = pd.DataFrame(kmeans.cluster_centers_, columns=X.columns)
centers_df.index = ['Cluster 0 (Low Risk)', 'Cluster 1 (High Risk)']
print(centers_df)
```

--- Individual Audio Reports ---

File: common_voice_en_42693784.wav

Cluster: 0 → ● Low Risk: Likely fluent, consistent speech
duration: 5.94
tempo: 125.00
pitch_variability: 166203.79
word_count: 11.00
pauses: 0.00
hesitations: 0.00
speech_rate: 0.18
cluster: 0.00
pca1: -0.68
pca2: -1.33

File: common_voice_en_42693841.wav

Cluster: 0 → ● Low Risk: Likely fluent, consistent speech
duration: 5.11
tempo: 133.93
pitch_variability: 135030.95
word_count: 13.00
pauses: 0.00
hesitations: 0.00
speech_rate: 0.22
cluster: 0.00
pca1: -1.19
pca2: -1.74

File: common_voice_en_42693855.wav

Cluster: 0 → ● Low Risk: Likely fluent, consistent speech
duration: 6.01
tempo: 187.50
pitch_variability: 194436.77
word_count: 11.00
pauses: 0.00
hesitations: 0.00
speech_rate: 0.18
cluster: 0.00
pca1: -1.18
pca2: 0.02

File: common_voice_en_42693856.wav

Cluster: 1 → ● High Risk: Signs of hesitations, pitch variability, slower pace
duration: 2.84
tempo: 101.35
pitch_variability: 250812.11
word_count: 6.00
pauses: 0.00
hesitations: 0.00
speech_rate: 0.10
cluster: 1.00
pca1: 3.61
pca2: -0.32

File: common_voice_en_42693864.wav

Cluster: 0 → ● Low Risk: Likely fluent, consistent speech
duration: 5.76
tempo: 197.37
pitch_variability: 295576.21

word_count: 12.00
pauses: 0.00
hesitations: 0.00
speech_rate: 0.20
cluster: 0.00
pca1: -1.26
pca2: 1.85

File: common_voice_en_42693865.wav

Cluster: 0 → ● Low Risk: Likely fluent, consistent speech
duration: 6.70
tempo: 178.57
pitch_variability: 271875.29
word_count: 14.00
pauses: 0.00
hesitations: 0.00
speech_rate: 0.23
cluster: 0.00
pca1: -2.13
pca2: 1.41

File: common_voice_en_42693871.wav

Cluster: 1 → ● High Risk: Signs of hesitations, pitch variability, slower pace
duration: 4.46
tempo: 83.33
pitch_variability: 408759.55
word_count: 8.00
pauses: 0.00
hesitations: 0.00
speech_rate: 0.13
cluster: 1.00
pca1: 2.92
pca2: 2.35

File: common_voice_en_42693876.wav

Cluster: 0 → ● Low Risk: Likely fluent, consistent speech
duration: 7.52
tempo: 129.31
pitch_variability: 204714.16
word_count: 15.00
pauses: 0.00
hesitations: 0.00
speech_rate: 0.25
cluster: 0.00
pca1: -2.48
pca2: -0.24

File: common_voice_en_42693883.wav

Cluster: 0 → ● Low Risk: Likely fluent, consistent speech
duration: 5.40
tempo: 85.23
pitch_variability: 144855.21
word_count: 11.00
pauses: 0.00
hesitations: 0.00
speech_rate: 0.18
cluster: 0.00
pca1: -0.20
pca2: -2.31

File: common_voice_en_42693884.wav

Cluster: 0 → ● Low Risk: Likely fluent, consistent speech
 duration: 6.34
 tempo: 163.04
 pitch_variability: 265565.28
 word_count: 13.00
 pauses: 0.00
 hesitations: 0.00
 speech_rate: 0.22
 cluster: 0.00
 pca1: -1.56
 pca2: 0.99

File: common_voice_en_42693885.wav

Cluster: 1 → ● High Risk: Signs of hesitations, pitch variability, slower pace
 duration: 2.45
 tempo: 120.97
 pitch_variability: 221391.81
 word_count: 4.00
 pauses: 0.00
 hesitations: 0.00
 speech_rate: 0.07
 cluster: 1.00
 pca1: 4.16
 pca2: -0.68

--- Clustering Centers ---

	duration	tempo	pitch_variability	word_count
Cluster 0 (Low Risk)	0.525369	0.345090	-0.304815	0.545468
Cluster 1 (High Risk)	-1.400985	-0.920241	0.812840	-1.454581

	pauses	hesitations	speech_rate	cluster	pca
Cluster 0 (Low Risk)	0.0	0.0	0.545468	-0.612372	-0.58738
Cluster 1 (High Risk)	0.0	0.0	-1.454581	1.632993	1.56637

	pca2
Cluster 0 (Low Risk)	-0.117867
Cluster 1 (High Risk)	0.314312

In [70]:

In []: