

Audio Processing



Big Data או בעולם של Big Data

בעידן הדיגיטלי, קול ווידאו הפכו למקורות מידע מרכזיים – מערכות דיהוי דיבור, דרך מצלמות חכבות ועד ניתוח רגשות ומעקב אחר תנועה.

איך מחשבים "مبינים" קול ווידאו

אילו אלגוריתם טכנולוגיים קיימים

איך עיבוד של נתונים לא-모בנים משתלב בעולם ה-Big Data

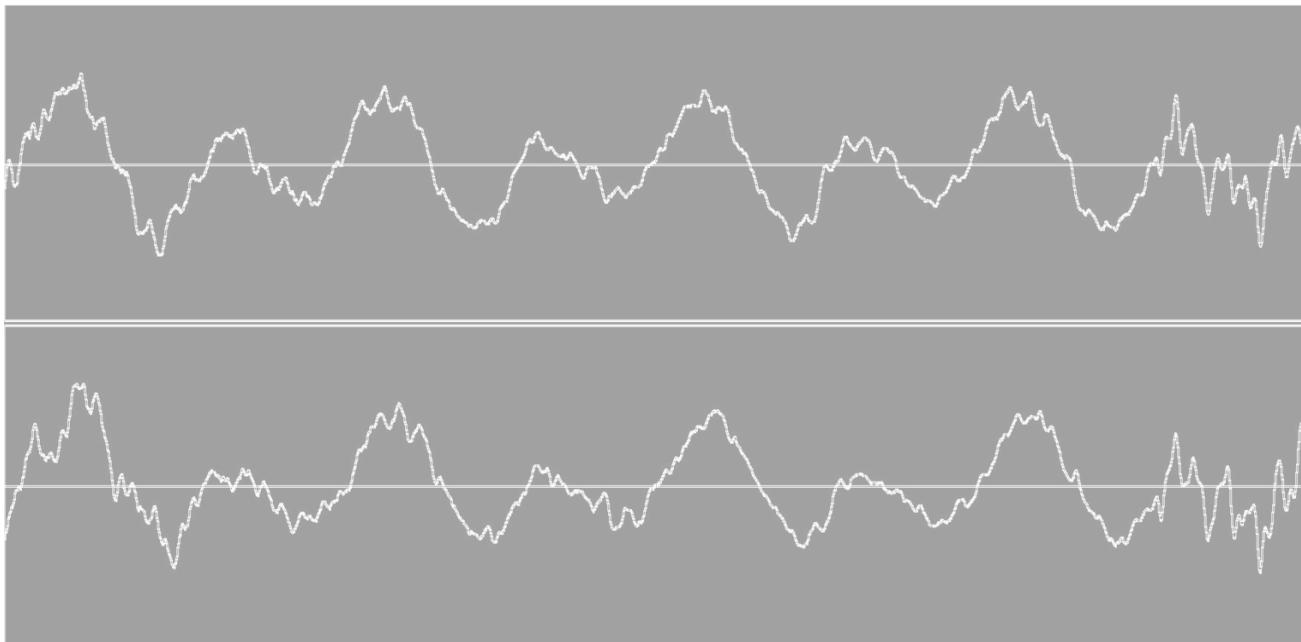
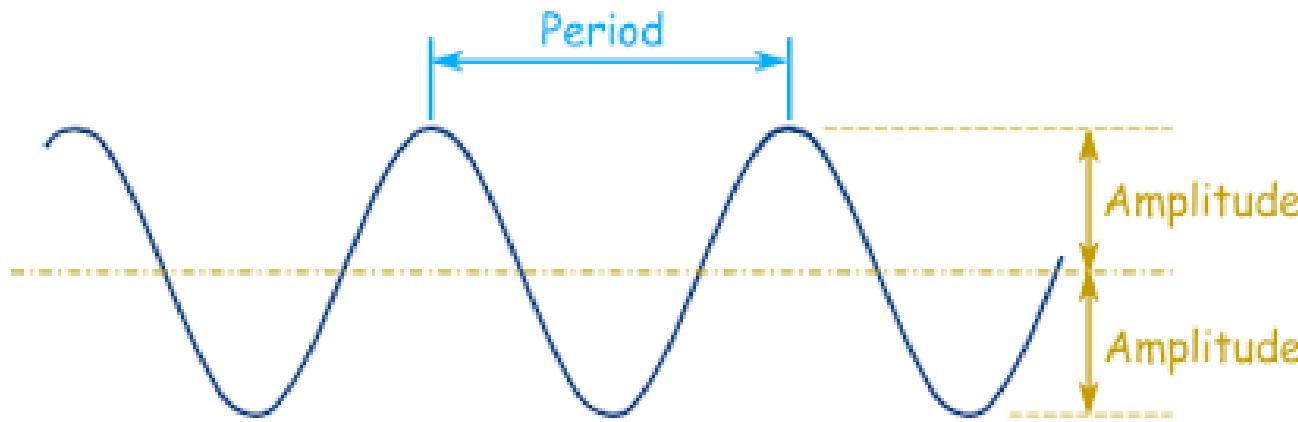
ונראה דוגמאות פרקטיות לעיבוד קול ווידאו באמצעות Python

אורך גל הוא המרחק בין שתי נקודות עוקבות

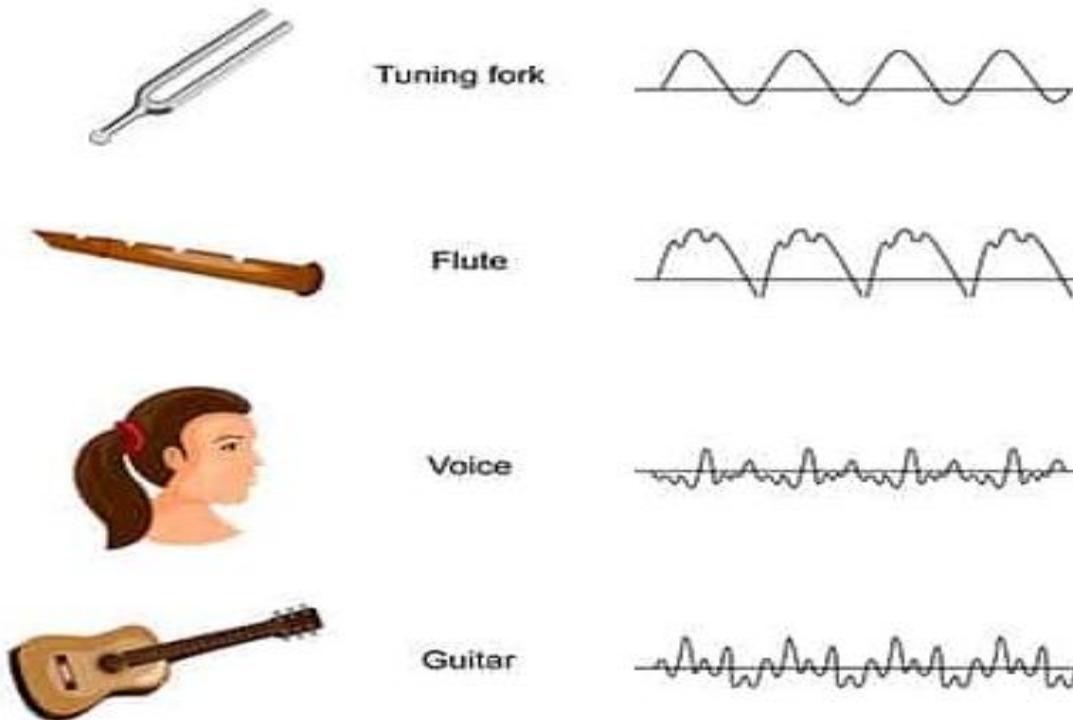
סימטריות, כמו בין שני רבעים (peaks) או שני שפלוֹת (troughs) של גל, והוא נמדד ביחידות של אורך (למשל מטרים).

אמפליטודה היא עוצמת הגל, המיצגת את הגובה של רכס או שפל בגל. היא משקפת את חזק התנודות בגל, וככל שהיא גבוהה יותר, הצליל או האנרגיה של הגל חזקים יותר.

תדירות (Frequency) היא מספר התנודות או המछזרים שתרחשים בזמן נתון, בדרך כלל בשנייה אחת, ונמדד ביחידות של הרץ (Hz). תדירות גבוהה יותר פירושה יותר התנודות לשנייה, בעוד שתדירות נמוכה יותר פירושה פחות התנודות לשנייה.



Timbre (Timbre) הוא מאפיין של הצליל שמאפשר לנו להבחין בין שני מקורות קוליים שונים, אפילו אם הם מנוגנים בתדר ובאפליטודה זהים. זהו "הטעם" או "הגונאים" של הצליל, והוא נגרם מה מגוון והצורה של גלי הקול הנפלטים מהմבשירים השונים, כמו פסנתר וחלוצרה. טימבר תלוי במאפיינים כמו הרמוניות, התקופה וההשתנות של הקול לאורך זמן.



https://www.google.co.il/search?q=lorel+yanny&rlz=1C1CHBD_iwIL1087IL1098&oq=lorel+yann/&gs_lcp=EgZjaHJvbWJyBggAEEUYOTIJCAEQAegKGIAEMgkIAhAAGAoYgAQyCQgDEAAYChiABDIQAQQABgKGIAEMgkIBRAGAoYgAQyCQgGEAAYChiABDIJCacQABgKGIAEVgkICBAAGAoYgAQyCQgJEAAChiABNIBBzg3NNowajeoAgiwAgE&sourc eid=chrome&ie=UTF-3#fpstate=ive&vld=cid:71cbb60,vid:JvKKd32Yw?E,st:0



librosa עם

```
import librosa  
import librosa.display  
from IPython.display import Audio
```

```
file_path = librosa.example('brahms')  
signal, sr = librosa.load(file_path)  
Audio(data=signal, rate=sr)
```

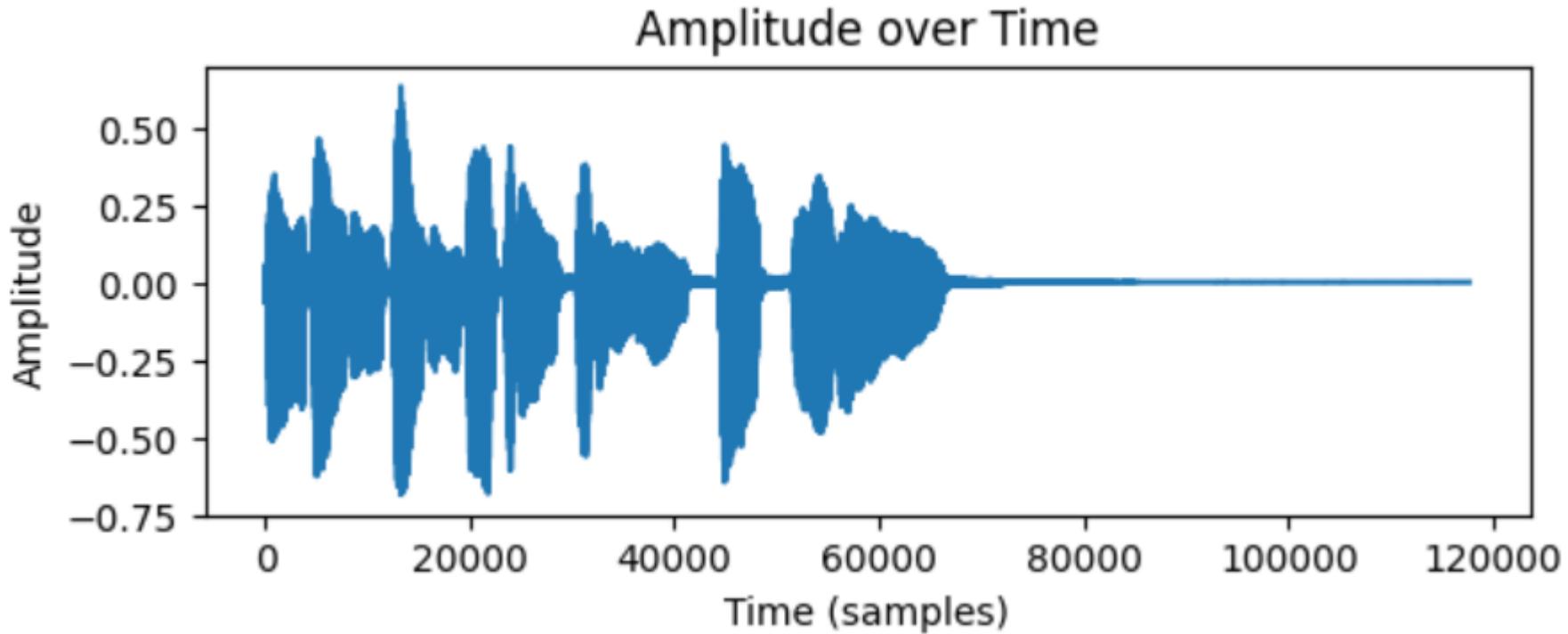
מערך של דגימות #
כמה דגימות נלקחו בשניה #
sr

השמעה של מקטע

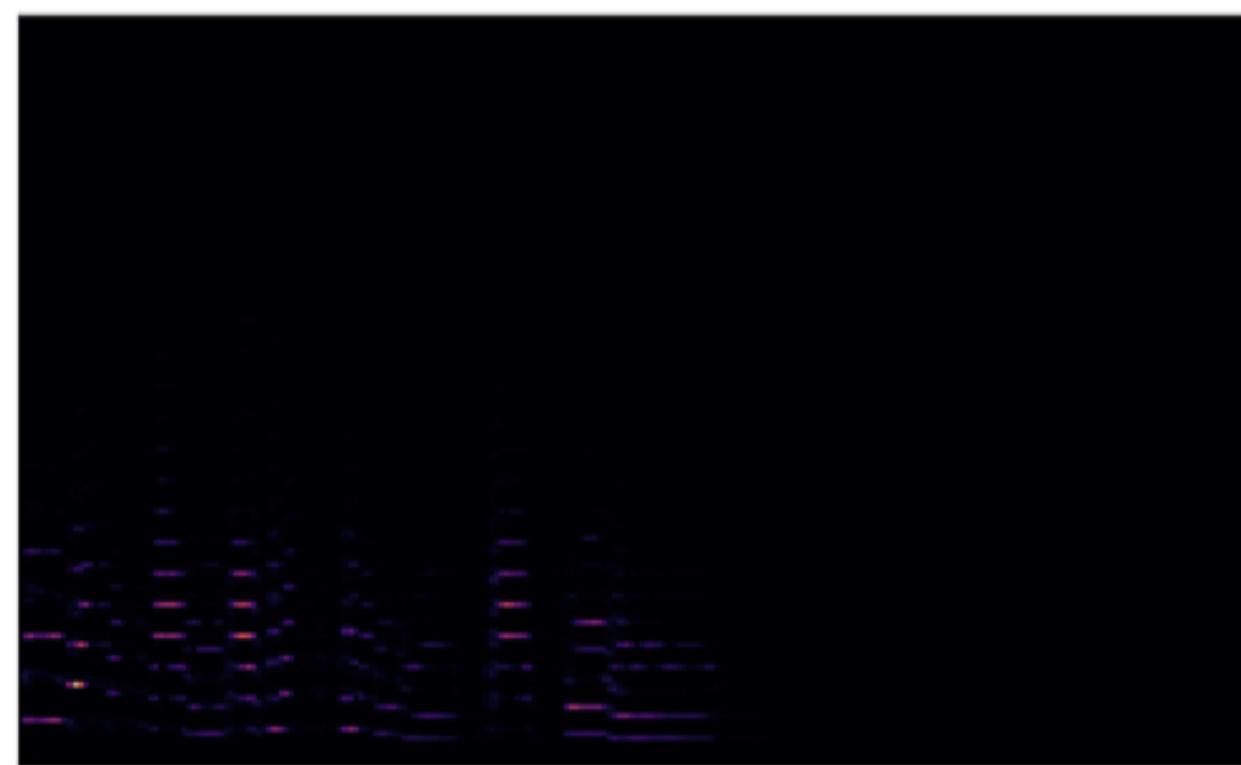
```
file_path = librosa.example('trumpet')  
signal, sr = librosa.load(file_path)  
Audio(data=signal, rate=sr)
```

אמפליטודה לאורך זמן

```
plt.plot(signal)
plt.title('Amplitude over Time')
plt.xlabel('Time (samples)')
plt.ylabel('Amplitude')
plt.show()
```



```
#spectrogram  
plt.subplot(2,1,2)  
sgram = librosa.stft(signal)  
librosa.display.specshow(sgram)  
plt.show()
```



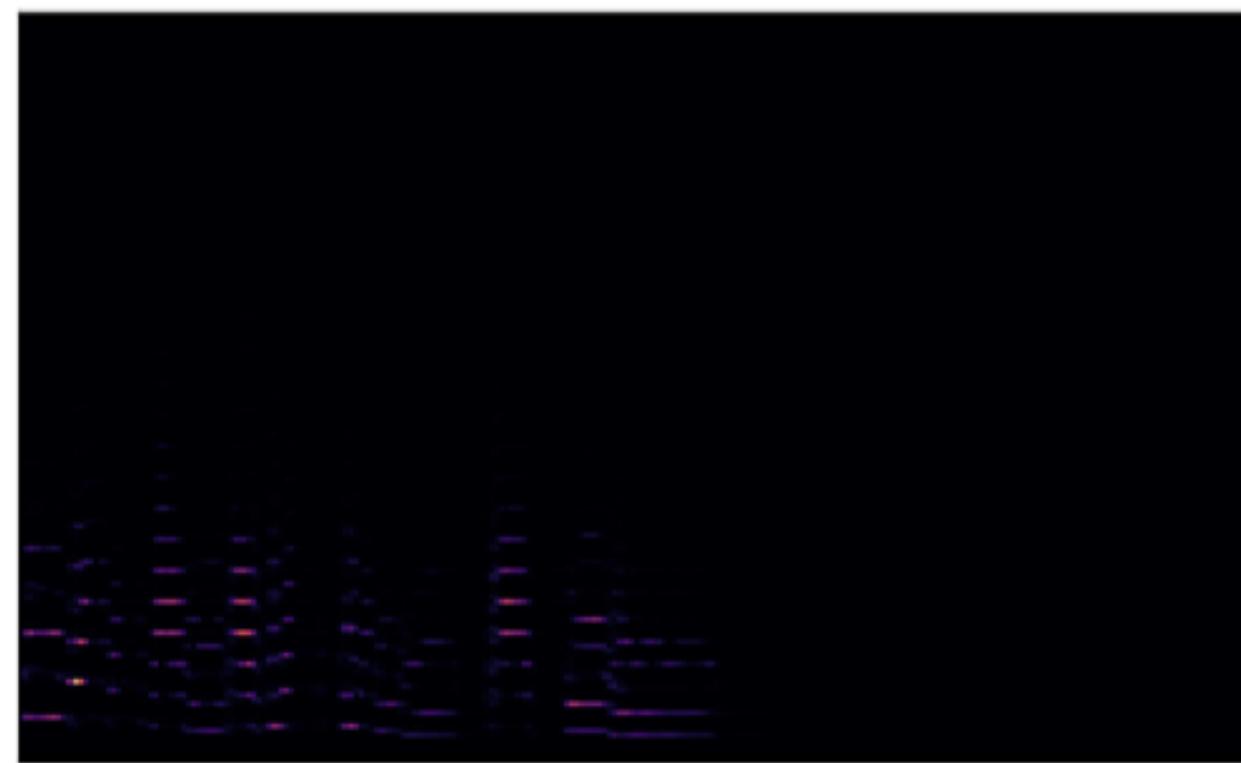
היא ייצוג חזותי של תדרים (frequency) של אות קול לאורך זמן. היא מציגה את התדרים השונים שמרכיבים את הצליל, בשציר ה- X מייצג את הזמן, וציר ה- Y מייצג את התדרים. הצבעים בסקטוגרמה מייצגים את האמפליטודה (החזק) של כל תדר בזמן נתון – צבעים מציינים עוצמת. סקטוגרמות משמשות לניתוח אודיו, זיהוי דיבור, וביתוח מוזיקה.

ספקטוגרמה

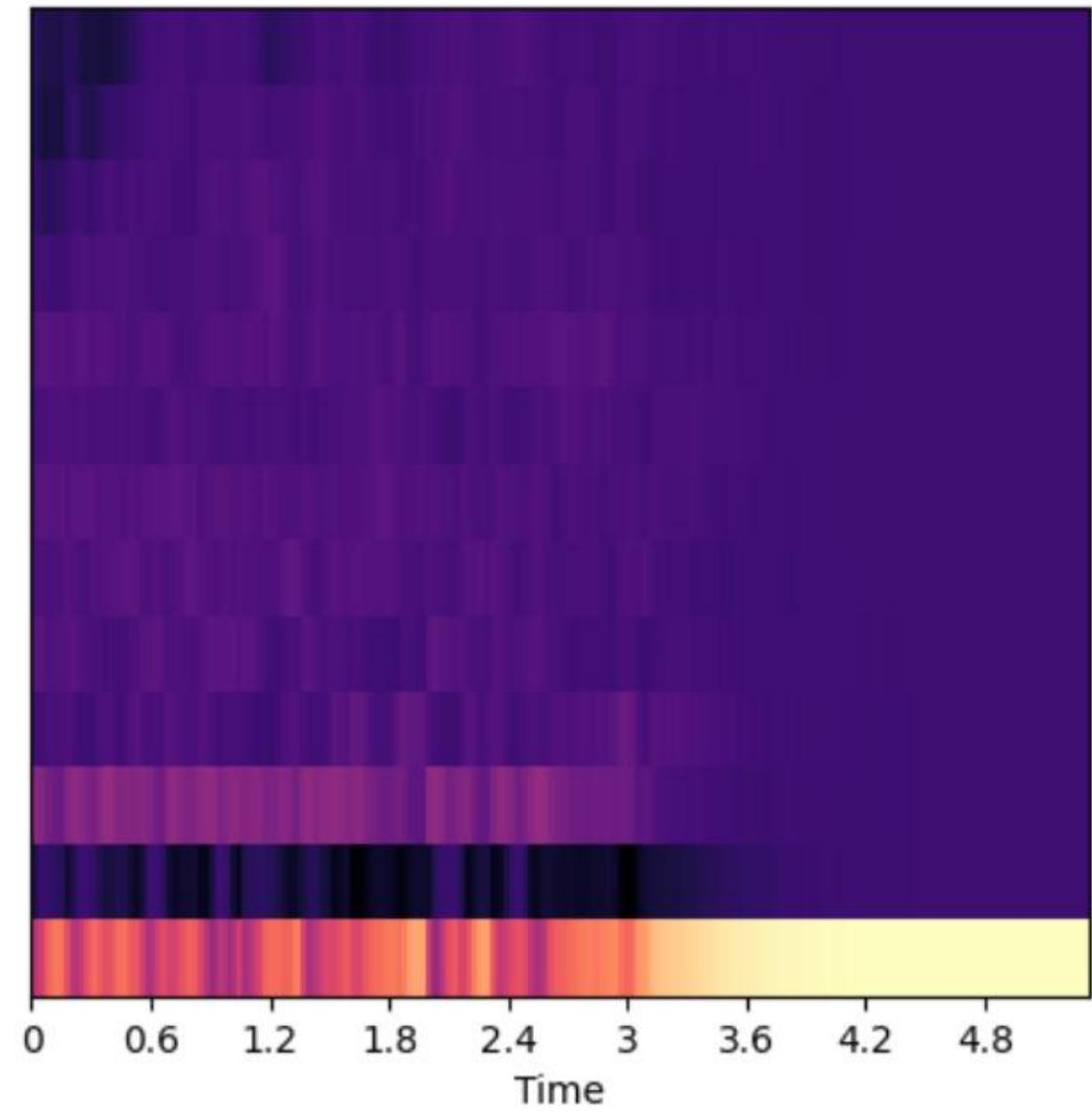
```
#spectrogram  
plt.subplot(2,1,2)  
sgram = librosa.stft(signal)  
librosa.display.specshow(sgram)  
plt.show()
```

- **Y-axis**(הציר האנכי): מייצג את ה- **Frequency** (תדרים) של אות. תדרים הם כמו גלים, שמייצגים את הצליל ב- Hertz (Hz).

- **X-axis**(הציר האופקי): מייצג את ה- **Time** (זמן), כלומר מציג את השינויים בתדרים לאורך זמן.



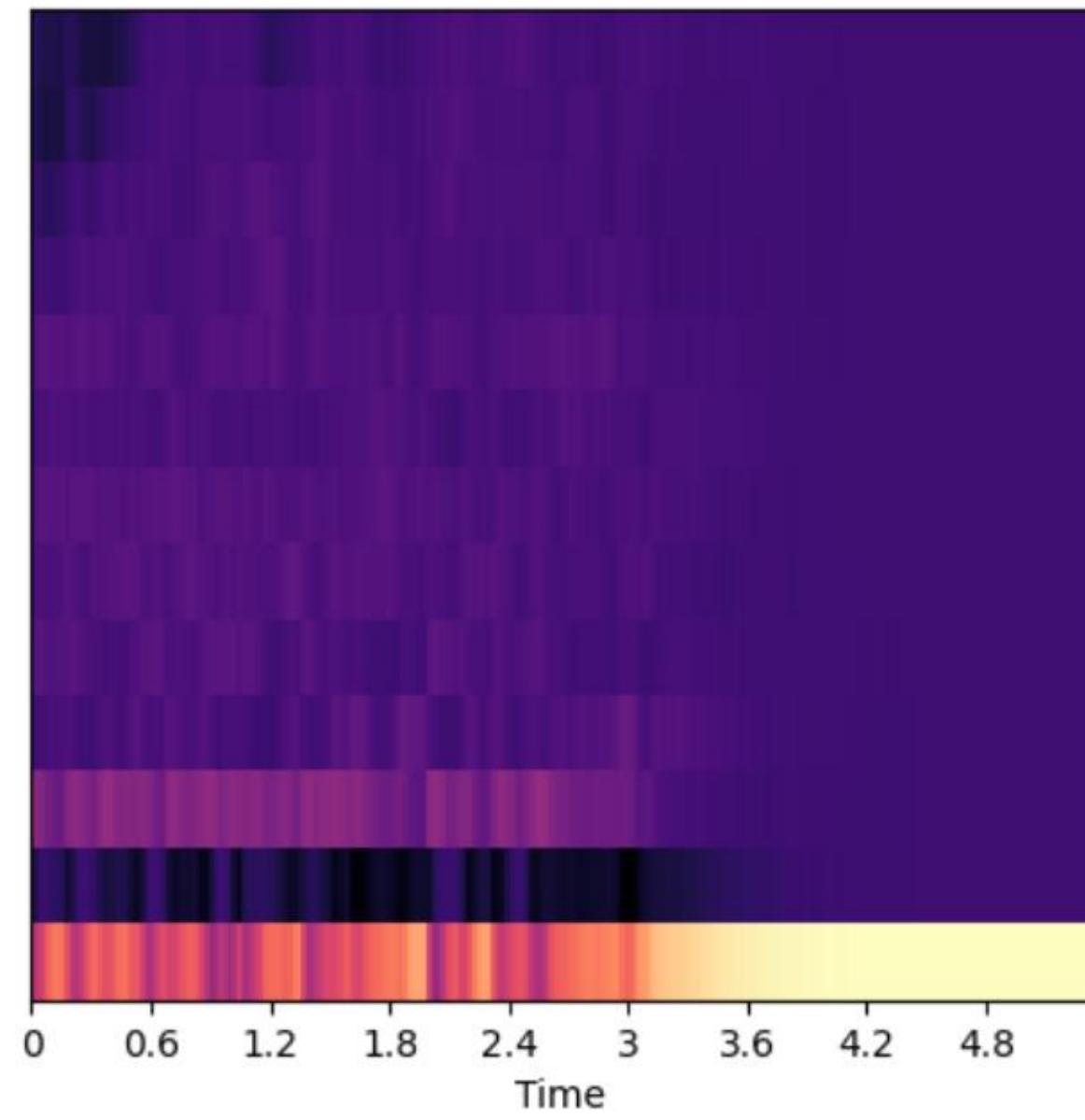
MFCC



Mel Spectrogram

- **Y-axis** – מציין את מדרג ה- **Mel** במקום את התדרים. מהקה את איר שבני אדם, תופסים את הצלילים. הוא מבוע על רעיון שהתפיסה האנושית לא שווה לאורך כל טווח התדרים – ככלומר, אנחנו שמים יותר תשומת לב לתדרים נמוכים ויתר רגישים להם.
- **X-axis** – עדין מציין את הזמן.
- **צבעים** מציינים את החזק של הצליל על פי **(dB) מדרג לוגריתמי** שמשיע להמחיש את החזק של הצלילים בצורה שמתאימה יותר לאוזן האנושית מתאר את החזק על פי המידה בה הצלילים משתנים.

MFCC



```
librosa.display.specshow(mfccs,  
                          sr=sr,  
                          x_axis='time',  
                          cmap= 'magma_r')  
  
plt.colorbar()  
plt.title('MFCC')  
plt.show()
```



כדי להגדיל את הורסתיליות של הנתונים ניתן להשתמש באוגמנטציה

```
# פונקציות לאוגמנטציה
def add_noise(signal, noise_level=0.005):
    noise = np.random.randn(len(signal))
    return signal + noise_level * noise

def change_speed(signal, sr, rate):
    return pyrb.time_stretch(signal, sr, rate)

def shift_pitch(signal, sr, n_steps):
    return librosa.effects.pitch_shift(signal, sr=sr, n_steps=n_steps)
```

Deep Learning- בתחום האודיו, המטרה היא להשתמש במודלים חכמים כדי לנתח ולהפיק תובנות מקבצי אודיו. מה שעושים בפועל הוא להמיר את האודיו למיצגים שנייתן לעבוד איתם, כמו **Spectrograms**, **Mel Spectrograms** ו-**CNNs** (רשתות עצביות קונבולוציוניות) מסוגלים לזהות תבניות ויזואליות בתמונות הילו ולהתאים לתמונה. המודלים העמוקים כמו **MFCCs** (Mel-Frequency Cepstral Coefficients) מיצמצם את המידע לבחינה מאפיינית שמייצגים את הדיבור או אותן לשונות, כמו זיהוי דיבור, זיהוי מוזיקה, זיהוי קולות, סיוג אובייקטים באודיו, או שיפור איכות האודיו. השימוש במאפיינים כמו **MFCCs** מחליף בצורה טובה יותר לעיבוד עם מודלים עמוקים.

