

Audio Processing

חלק בלתי נפרד מיישומי AI בעולם של Big Data

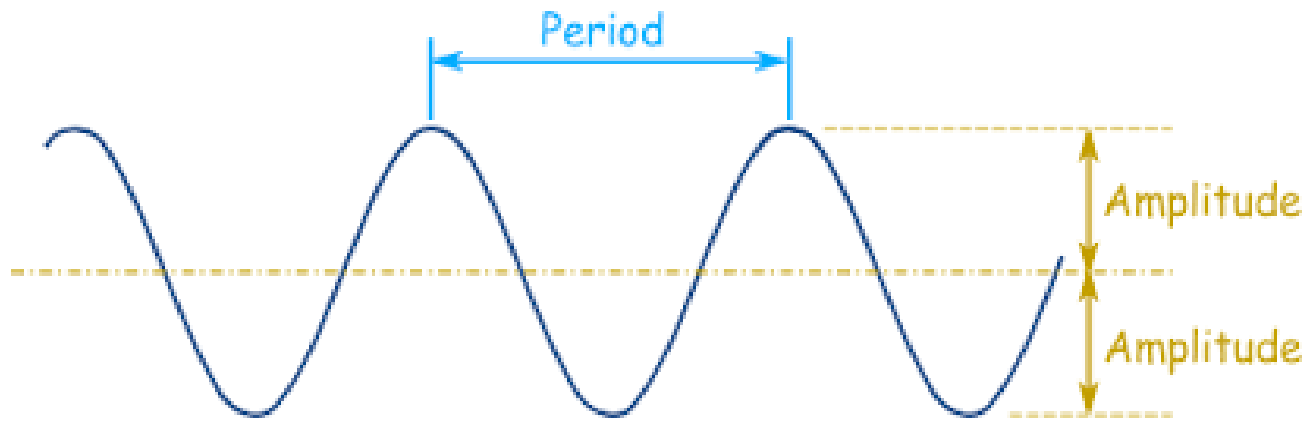
בעידן הדיגיטלי, קול ווידאו הפכו למקורות מידע מרכזיים – ממערכות זיהוי דיבור, דרך מצלמות חכמות ועד ניתוח רגשות ומעקב אחר תנועה.

איך מחשבים "מבינים" קול ווידאו 📺

אילו אתגרים טכנולוגיים קיימים 📺

איך עיבוד של נתונים לא-מובנים משתלב בעולם ה-Big Data 📺

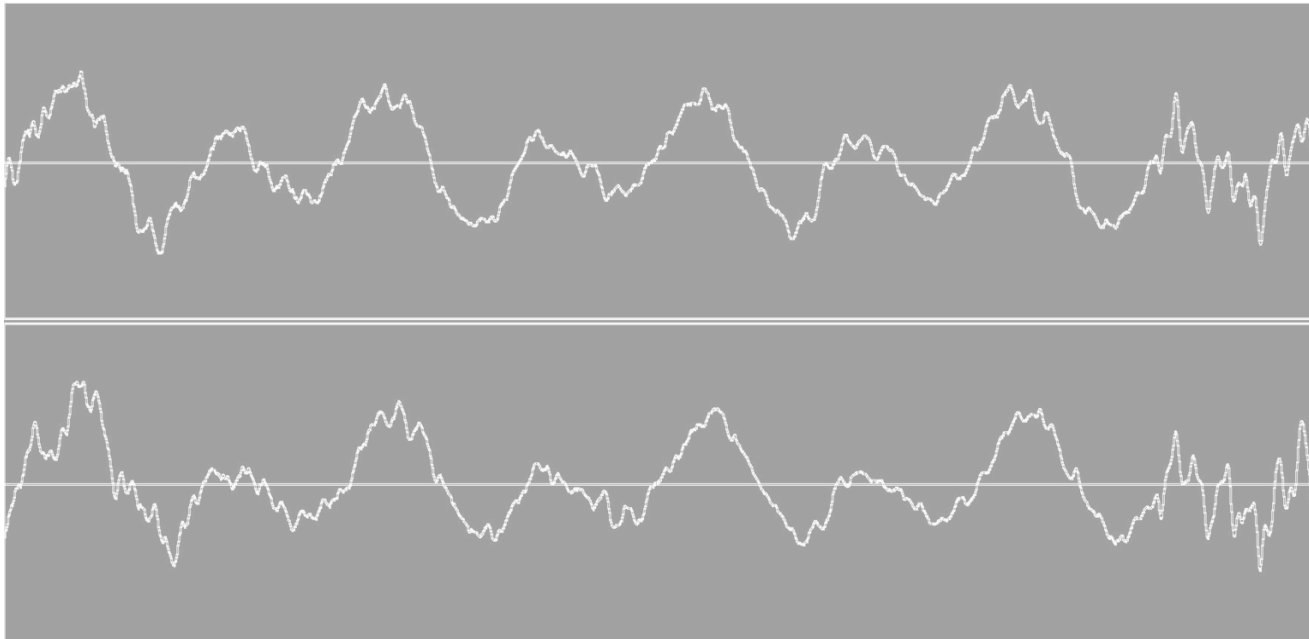
ונראה דוגמאות פרקטיות לעיבוד קול ווידאו באמצעות Python 📺



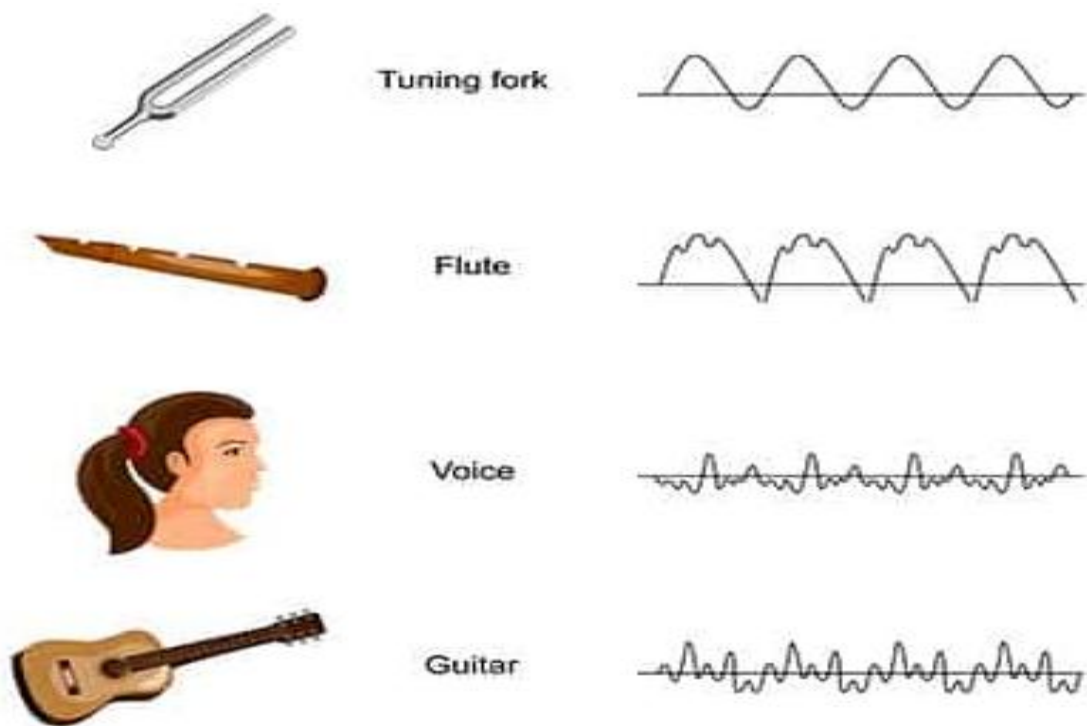
אורך גל הוא המרחק בין שתי נקודות עוקבות סימטריות, כמו בין שני רכסים (peaks) או שני שפלות (troughs) של גל, והוא נמדד ביחידות של **אורך** (למשל מטרים).

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

תדירות (Frequency) היא מספר התנודות או המחזורים שמתרחשים בזמן נתון, בדרך כלל בשנייה אחת, ונמדדת 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 "הגוונים" של הצליל, והוא נגרם מהמגוון והצורה של גלי הקול הנפלטים מהמכשירים השונים, כמו פסנתר וחצוצרה. טימבר תלוי במאפיינים כמו הרמוניות, התקופה והשתנות של הקול לאורך זמן.



https://www.google.com/search?q=lorel+yanny&rlz=1C1CHBD_iwL1087IL1098&oq=lorel+yanny&gs_lcrp=EgZjaHJvbWJyBggAEEUYOTIJCAEQAg3gKGIAEMgkIAhAAGAoYzAQyCQgDEAAAYChiABDICAQQAQBgKGIAEMgkIBRvAGAoYgAQyCQgGEAAAYChiABDIJCACQABgKGIAEMgkICBAAGAoYgAQyCQgJEAAAYChiABNIBBzg3NNowajeoAgiwAgE&sourceid=chrome&ie=UTF-3#fpstate=ive&vld=cid:71cbb60,vid:JvKKd32Yw2E,st:0



עבודה עם librosa

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

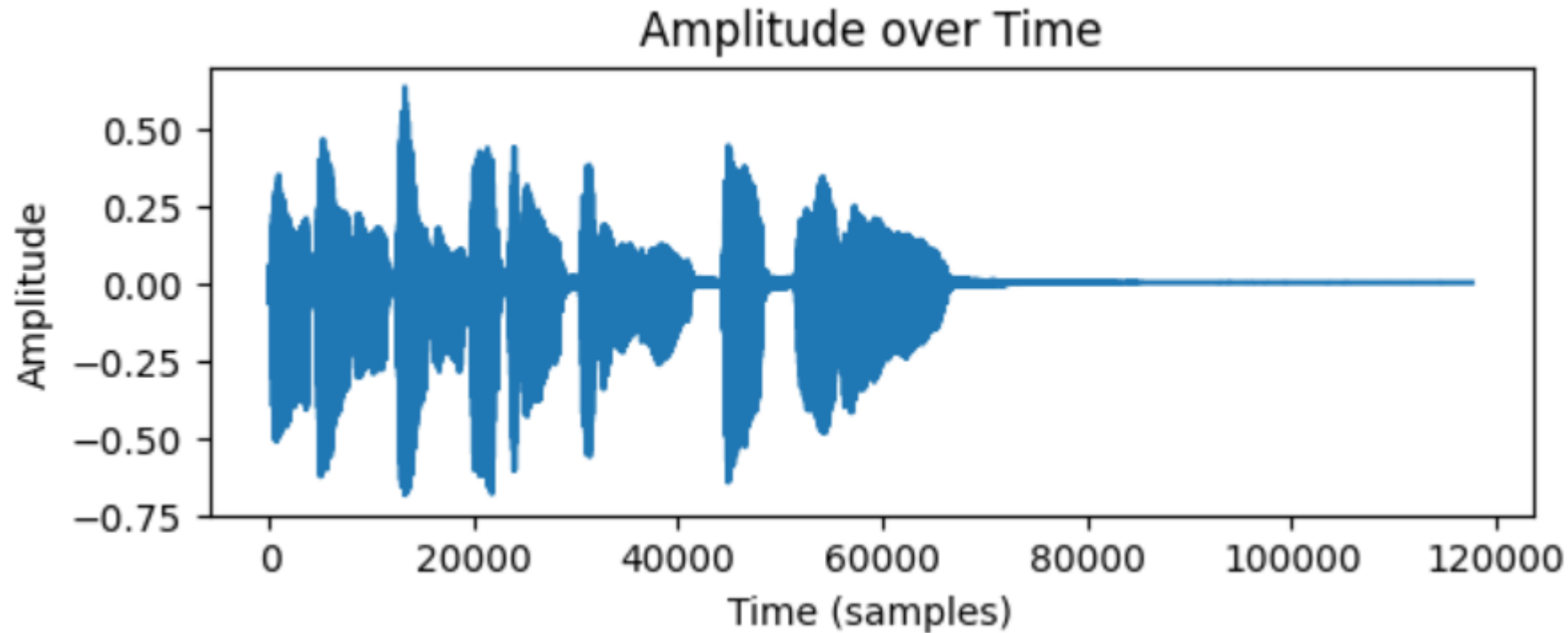
```
file_path = librosa.example('brahms')
signal, sr = librosa.load(file_path)
Audio(data=signal, rate=sr)
```

signal # מערך של דגימות
sr # כמה דגימות נלקחו בשנייה

השמעה של מקטע

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

```
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, כשציר ה- X מייצג את הזמן, וציר ה- Y מייצג את התדרים.

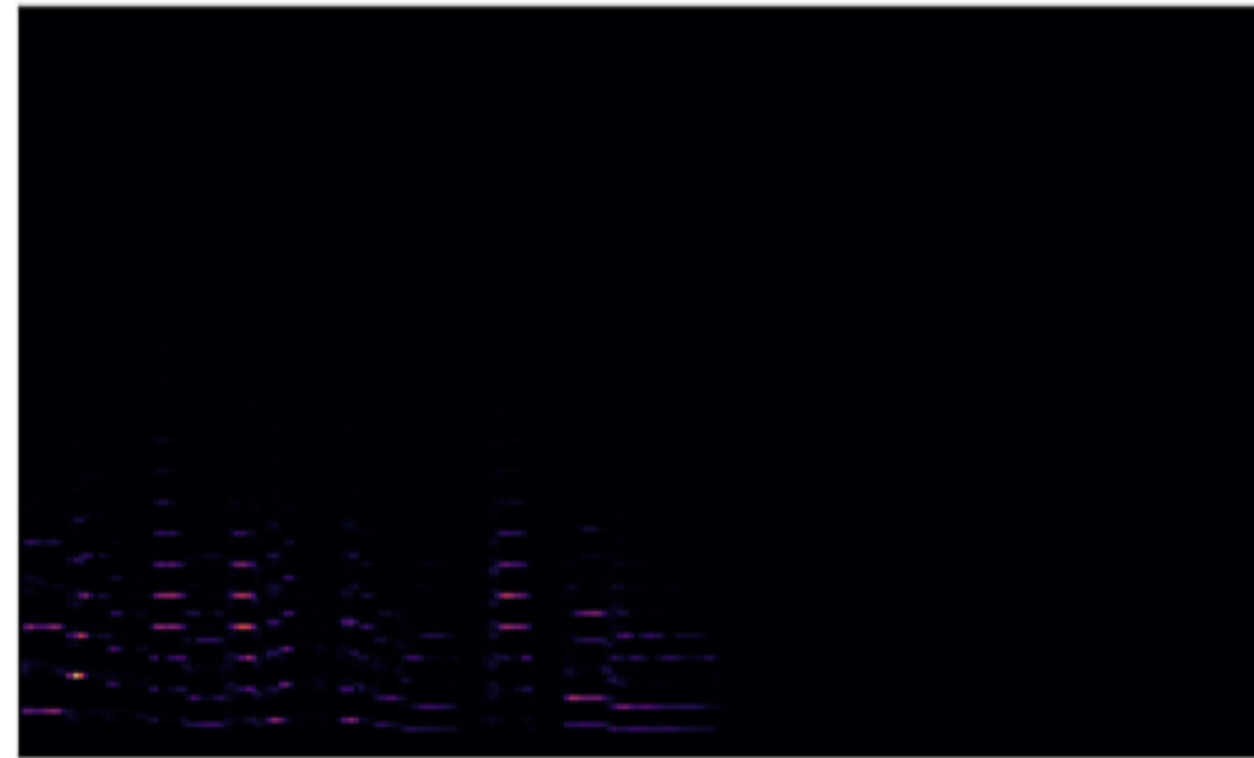
הצבעים בספקטוגרמה מייצגים את

האמפליטודה (החוזק) של כל תדר בזמן נתון – צבעים מציינים עוצמת. ספקטוגרמות משמשות לניתוח אודיו, זיהוי דיבור, וניתוח מוזיקה.

```
#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 האות. 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.



Mel Spectrogram

• **Y-axis** מציין את **מדרג ה-Mel**, במקום את התדרים.

מחקה את איך שבני אדם, תופסים את הצלילים. הוא מבוסס

על רעיון שהתפיסה האנושית לא שווה לאורך כל טווח

התדרים – כלומר, אנחנו שמים יותר תשומת לב לתדרים

נמוכים יותר רגישים להם.

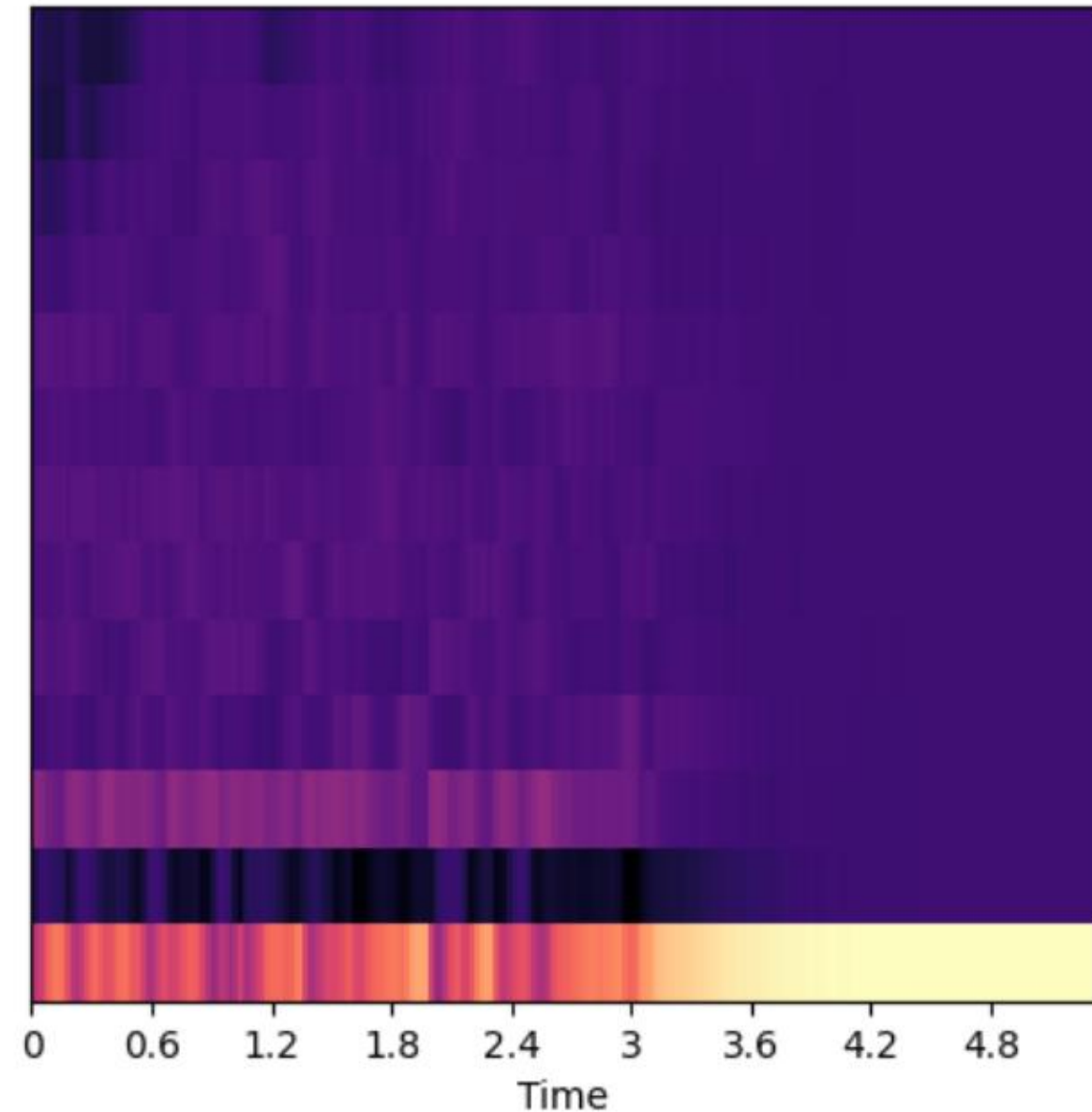
• **X-axis** עדיין מציין את הזמן.

• **צבעים** מציינים את החוזק של הצליל על פי **dDecibels**

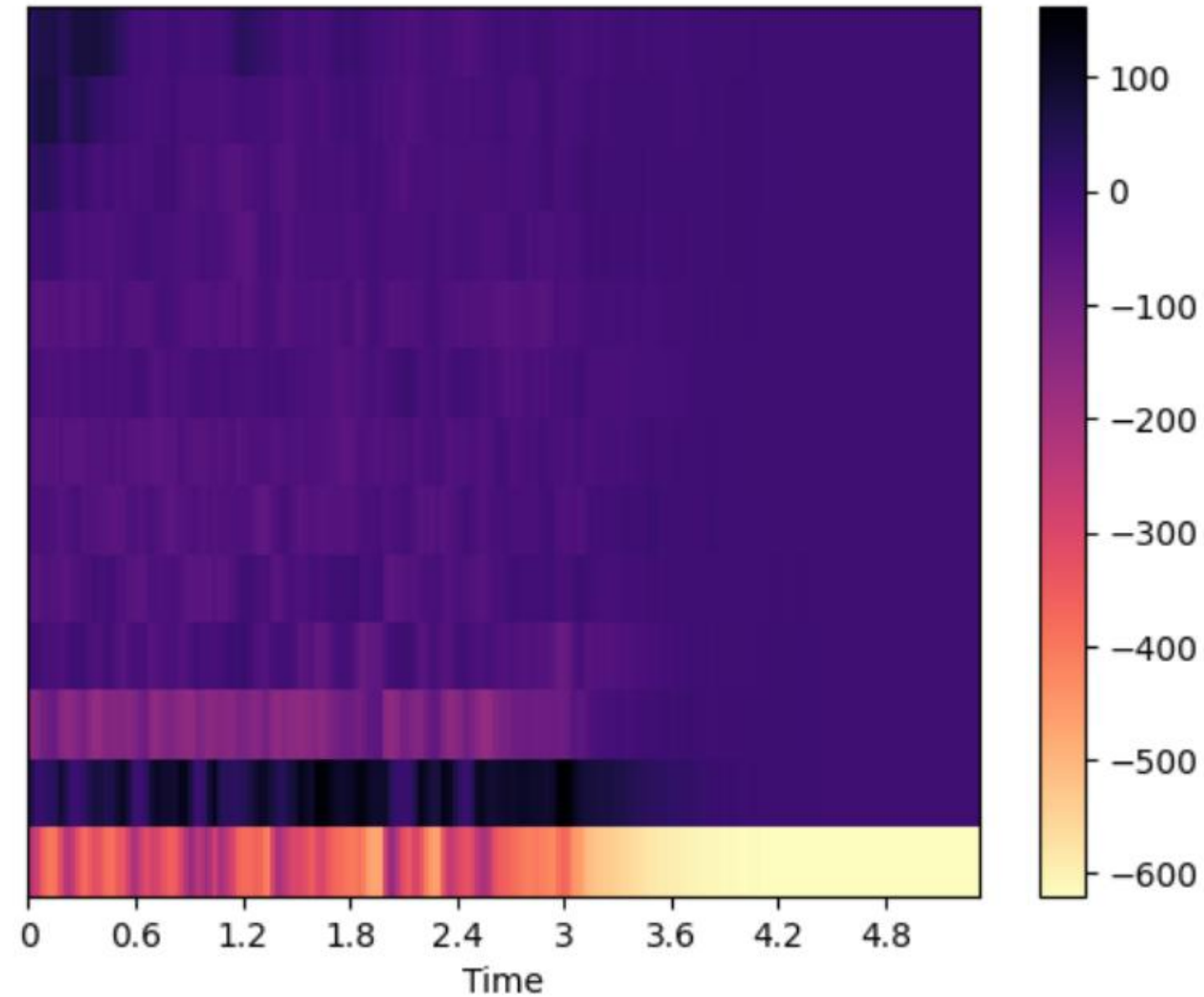
(dB). **מדרג לוגריתמי** שמסייע להמחיש את החוזק של

הצלילים בצורה שמתאימה יותר לאוזן האנושית מתאר את

החוזק 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 הוורסטיילות 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 ה־אודיו, המטרה היא להשתמש במודלים חכמים כדי לנתח ולהפיק תובנות מקבצי אודיו. מה שעושים בפועל הוא להמיר את האודיו למייצגים שניתן לעבוד איתם, כמו **Spectrograms** ו־**Mel Spectrograms** שנעשים באמצעות המרת האודיו לתמונה. המודלים העמוקים כמו **CNNs** (רשתות עצביות קונבולוציוניות) מסוגלים לזהות תבניות ויזואליות בתמונות הללו ולהתאים אותן למשימות שונות, כמו זיהוי דיבור, זיהוי מוזיקה, סיווג קולות, זיהוי אובייקטים באודיו, או שיפור איכות האודיו. השימוש במאפיינים כמו **MFCCs** (Mel-Frequency Cepstral Coefficients) מצמצם את המידע לכדי מאפיינים שמייצגים את הדיבור או הצלילים בצורה טובה יותר לעיבוד עם מודלים עמוקים.

