

MFCC

MEL FREQUENCY CEPSTRAL COEFFICIENTS

Abdullah MEMİŞOĞLU

ÖZET

Bu yazı MFCC tanımını, anlamını, bir ses sinyaline uygulanmasını ve çıktıların yorumlanmasını içermektedir. MFCC, Ses Tanıma ve Müzikten Bilgi Alma Disiplinlerarası Bilimi (Music Information Retrieval) alanlarında çokça kullanılan bir işlemdir. Bir telefona söylenen sayıların otomatik tanınmasından sese tür sınıflandırmalarına ve ses benzerliklerinin ölçümüne kadar birçok uygulama alanına sahip bir işlemdir.

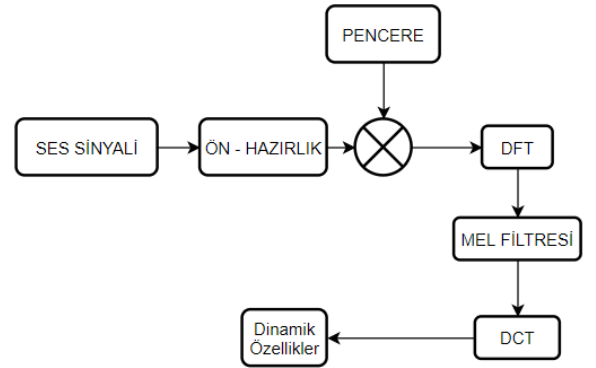
ANAHTAR KELİMELER

MFCC, Mel Frekans Kepstral Katsayıları, Ses Frekansı, Kepstrum, Mel Ölçeği

1. GİRİŞ

MFCC, bir ses sinyali nasıl katsayılarla dökülür? Ve bu katsayılar ne ifade eder? Soruları üzerine ortaya atılmış bir tanımdır. MFCC tanımı itibarıyla ses sinyalinin kısa zamanlı güç spektrumunun Mel Ölçeği üzerindeki ifadesidir. MFCC ismini Mel Frequency Cepstral Coefficients yani Mel-Frekans Kepstral Katsayıları'ndan alır. Kepstrum, bir zaman sinyalinin logaritmasının spektrumu olarak bilinir. Mel Frekans Kepstrumu Kepstrumu(MFK)'ndan farklı olarak, doğrusal eşit aralıklı frekans bantlarından oluşmamaktadır. MFK insan işitme sistemine yaklaşan Mel Ölçeği'nde eşit aralıklı frekans bantları oluşturmaktadır. Bu sayede insan kulağı ile anlaşılabilir farklar oluşturmak mümkündür. İnsan kulağının 300-400 Hz frekansına ait iki sesi ayırt etmesi, 900-1000 Hz frekansına ait iki sesi ayırt etmesinden daha zordur. Ancak aradaki fark eşittir. Mel Ölçeği sayesinde anlaşılabilirliğin artması mümkün olmaktadır.

MFCC Uygulama Adımları



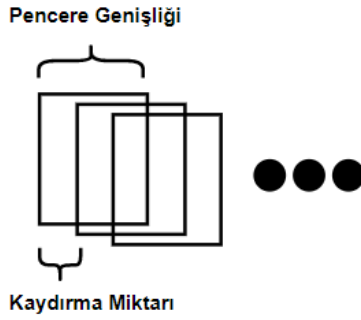
Şekil 1: MFCC Blok Şeması

1. **ÖN-HAZIRLIK:** Bu aşamada daha yüksek frekansları vurgulayan bir filtre oluşturulur. Amacı, yüksek frekans bölgesinde dik bir düşüşe sahip olan seslerin spektrumunu dengelemektir. En yaygın olarak kullanılan ön vurgu filtresi aşağıdaki transfer fonksiyonu ile verilmektedir.

$$H(z) = 1 - bz^{-1}, 0.4 < b < 1.0 \quad (eq.1)$$

Şekil 2: MFCC Blok Şeması

2. **PENCERE:** Bu aşamada sesin analiz edileceği zaman periyodu belirlenerek bir pencere oluşturulur. Örneğin bir konuşmayı analiz etmek için 20 ms (ileri düzey analizlerde 10 ms olarak da belirlenebilir.) olarak belirlenir. Burada iki ayrı kavram bulunmaktadır. Birincisi analiz edilecek alan olan pencere boyutu bir diğeri ise kaydırma miktarıdır. Pencere olarak genel anlamda Hamming ve Hanning pencerelerinin kullanımı yaygındır. Bir pencere boyutunda elde edilen her veriye çerçeve(frame) ismi verilir.



Şekil 2: Pencereleme Sistemi Şeması

3. DFT (Discrete Fourier Transform):

Pencerelemiş her bir çerçevenin ayrık zamanlı fourier dönüşümü alınarak genlik spektrumu elde edilir.

$$X(k) = \sum_{n=0}^{N-1} x(n)e^{-j2\pi nk/N} \quad 0 \leq k \leq N-1 \quad (eq.2)$$

Bu denklemde N, DFT hesaplamak için kullanılacak nokta sayısına karşılık gelmektedir.

4. **MEL FİLTRESİ:** DFT aldıktan sonra elde edilen sinyal bir bant geçiren filtreler dizisinden geçerek bir sinyal oluşturulur. Bu filtreler dizisine Mel Filtreleri, oluşan sinyale ise Mel Spektrumu ismi verilir. Mel, insan kulağının algıladığı frekansa dayalı bir ölçü birimidir. Mel ölçeği yaklaşık olarak 1 kHz'in altında doğrusal bir frekans aralığı ve 1 kHz'in üzerinde logaritmik bir aralıktır. Kepstrum ve Mel Frekans Kepstrumu arasındaki fark da buradan gelmektedir. İlgili aralıkların doğrusal olmak yerine logaritmik değişimi ses sinyallerinin analizlerini kolaylaştırır. Normal kepstrumda frekans değerine f, Mel ölçeğindeki değerine de f_{MEL} denilecek olursa ilgili çıkarım eq.3'teki gibi elde edilir.

$$f_{MEL} = 2595 \log_{10} \left(1 + \frac{f}{700} \right) \quad (eq.3)$$

5. DCT (Discrete Cosine Transform):

Dönüştürülen Mel frekans katsayılarına uygulanan DCT, bir dizi kepsral katsayı üretir. DCT'yi hesaplamadan önce, Mel spektrumu genellikle bir log ölçeğinde temsil edilir. Sinyal bilgilerinin çoğu ilk

birkaç MFCC katsayısı ile temsil edildiğinden, sistem sadece yüksek dereceli DCT bileşenlerini yok sayan veya kesen katsayıları çıkararak sağlam hale getirilebilir. MFCC katsayıları aşağıdaki formül ile elde edilir.

$$c(n) = \sum_{m=0}^{M-1} \log_{10}(s(m)) \cos \left(\frac{\pi n(m-0.5)}{M} \right)$$

$n = 0, 1, 2, \dots, C-1$ (eq.4)

Burada $c(n)$ kepsral katsayıları, C ise MFCC sayısını belirlemektedir.

6. **Dinamik Özellikler:** Kepsral katsayılar, yalnızca belirli bir çerçeveden bilgi içerdiklerinden, genellikle statik özellikler olarak adlandırılır. Sinyalin zamansal dinamikleri hakkında ek bilgi, kepsral katsayıların birinci ve ikinci türevlerinin hesaplanmasıyla elde edilir. Birinci dereceden türev delta katsayıları, ikinci dereceden türev ise delta-delta katsayıları olarak adlandırılır. Konuşma tespiti örneği için delta katsayıları konuşma hızı hakkında bilgi verir ve delta-delta katsayıları konuşmanın hızlanmasına benzer bilgiler sağlar.

PYTHON DİLİNDE MFCC

Python dilinde MFCC oluşturmak için **librosa** kütüphanesinden faydalanılmıştır. Öncelikle ses verisinin librosa.load methodu ile örneklenerek kayan noktalı zaman serisine dönüştürülmesi sağlanmalıdır.

```
# Sesi Görüntüleme
plt.figure(figsize=(15,5))
plt.subplot(421)
x,sr = librosa.load(CoughSound)
librosa.display.waveplot(x,sr=sr)
```

Sonrasında librosa.feature.mfcc methodu ile ilgili ses verisinin MFCC katsayılarının alınması ve görselleştirmesi mümkündür. Burada x ses zaman verisine, sr ise ses frekansına karşılık gelmektedir.

```
# MEL FREKANS KATSAYI ÇIKARIMI  
plt.figure(figsize=(15,5))  
plt.subplot(423)  
mfkk=librosa.feature.mfcc(x,sr=sr)  
  
librosa.display.specshow(mfkk,x_axis="s")  
plt.colorbar()
```

KAYNAKÇA

- 1) <https://www.analyticsvidhya.com/blog/2021/06/mfcc-technique-for-speech-recognition/>
- 2) <https://link.springer.com/content/pdf/bbm%3A978-3-319-49220-9%2F1.pdf>
- 3) <https://towardsdatascience.com/how-i-understood-what-features-to-consider-while-training-audio-files-eedfb6e9002b>
- 4) <https://medium.com/prathena/the-dummys-guide-to-mfcc-aceab2450fd>
- 5) <https://medium.com/@tanveer9812/mfccs-made-easy-7ef383006040>
- 6) <https://ijirae.com/volumes/vol1/issue10/27.NVEC10086.pdf>
- 7) <https://medium.com/datarunner/librosa-9729c09ecf7a>
- 8) <https://musicinformationretrieval.com/mfcc.html>
- 9) https://en.wikipedia.org/wiki/Mel-frequency_cepstrum
- 10) https://en.wikipedia.org/wiki/Music_information_retrieval