

SAYISAL İŞARET İŞLEME LABORATUVARI

DÖNEM PROJESİ

DERS KODU: ELM368

PROJE ADI: OTOMATİK SES TANIMLAMA

PROJE TESLİM TARİHİ: 20.06.2020

PROJE EKİP ÜYELERİ:

ANIL PARLAK, ID = 161024096

ALİ İHSAN GÜLLÜ, ID = 161024013

MÜMTAZ ÖZLEN, ID = 161024044

İçindekiler

ÖNSÖZ	3
ÖZET	4
SEMBOLLER VE KISALTMALAR.....	4
1.GİRİŞ	5
2. NYQUIST ÖRNEKLEME TEOREMİ	6
3. FİLTRELEME	7
IIR filtrelerin FIR filtrelere göre karşılaştırılması.....	7
4. FAST FOURİER TRANSFORM	9
5. KISA ZAMANLI FOURİER DÖNÜŞÜMÜ(SHORT TIME FOURİER TRANSFORM)	10
6. MFCC MEL FREKANS.....	13
7. MFCCs KATSAYILARININ KARŞILAŞTIRILMASI.....	14
Dynamic Time Warping	14
8. YORUMLAR VE DEĞERLENDİRME.....	16
9. KAYNAKLAR	16

ÖNSÖZ

Konuşma tanıma, bir makine veya programın sözcükleri ve tümceleri sözlü olarak tanımlama ve bunları makine tarafından okunabilir bir biçime dönüştürme yeteneğidir. İlkel konuşma tanıma yazılımı sınırlı bir kelime ve deyim sözlüğüne sahiptir ve bunları ancak çok açık bir şekilde konuşulursa tanımlayabilir. Daha sofistike yazılımlar doğal konuşmayı kabul etme yeteneğine sahiptir. Konuşma tanımlama teknolojisi fiziksel engelli bireyler için hayatlarını kolaylaştırıcı olmakla birlikte günümüzde bir çok farklı alanda bu teknoloji kullanılmaktadır. Bu alanlara örnek olarak akıllı evler, çağrı merkezleri, telefon asistan programları gösterilebilir.

Konuşma tanıma, akustik ve dil modelleme yoluyla algoritmalar kullanarak çalışır. Akustik modelleme, dilsel konuşma birimleri ile ses sinyalleri arasındaki ilişkiyi temsil eder; dil modelleme, sesleri benzer sözcükleri ayırt etmeye yardımcı olmak için sözcük dizileriyle eşleştirir. Genellikle gizli Markov modelleri, sistemdeki doğruluğu artırmak için ve konuşmadaki geçici desenleri tanımak için de kullanılır.

Bu proje kapsamında ses kayıtlarının alınması, kayıtlı dosyaların etiketlenmesi ve sınıflandırılması, sınıflandırılan dosyaların özelliklerinin çıkarılması ve daha sonra mikrofondan ses girişi yapılarak, söylenilen kelimenin, hecenin özelliklerinin veri tabanındaki diğer seslerin özellikleriyle karşılaştırılarak metin çıktısı oluşturulması amaçlanmıştır. Kullanıcı birden fazla hece veya kelime söyleyebilir.

ÖZET

Projede ses özelliklerinin çıkartılıp sınıflandırılması işlemi kullanılarak mikrofondan gelen sesin tanımlanması ve yazdırılması üzerinde çalışılmıştır. Proje kapsamında önceden farklı kullanıcıların seslerinin farklı kelimeler söylenmesi sonucu ses dosyalarının kaydedilmesi, kaydedilen ses dosyalarının Fourier Transformatlarının çıkarılması, Mel-Frekans Cepstral Katsayılarının (Mel-Frequency Cepstral Coefficients) çıkarılması, Mel Frekans Cepstral Katsayılarının, Dinamik Zaman Bükme algoritması (Dynamic Time Warping) yardımıyla farklarının hesaplanması ve farkın minimum olduğu etiketli ses dosyasının, etiketinin çıktı olarak yazdırılması uygulanmıştır.

Ses sinyali, herhangi bir sesin iletilmek veya saklanmak için elektromanyetik enerjiye çevrilmiş halidir. Ses, havadaki titreşimin kulak içine teması sonrasında oluşur. Titreşim çok farklı frekanslarda olabilir. Bu titreşim mikrofon vasıtasıyla ses sinyaline çevrilebilir ya da ses sinyalleri hoparlör aracılığıyla titreşime dönüştürülebilir. Her ses işaretinin kendine ait özellikleri mevcuttur. Bu özelliklerden bazıları; frekans spektrumları, spektrogramları, güç yoğunluk fonksiyonları, Mel-Frekans Cepstral Katsayılarıdır.

Projede giriş işareti mikrofon yardımıyla alınacaktır. Program giriş işaretini aldıktan sonra ortamdaki istenmeyen arka plan gürültülerini gidermek ve daha yüksek doğrulukta tahminlerde bulunabilmek için işareti alçak geçiren filtreden geçirip, özelliklerini çıkaracak ve karşılaştırmaları yapacaktır.

SEMBOLLER VE KISALTMALAR

ASR: Automatic Speech Recognition (Otomatik Ses Tanımlama)

MFCCs: Mel-frequency Cepstral Coefficients (MFCCs) (Mel frekans cepstral katsayıları)

DTW: Dynamic Time Warping (Dinamik Zamanda Bükme)

FFT: Fast Fourier Transform

STFT: Short-Time Fourier Transform

dB : Desibel

f: frekans (Hertz)

1.GİRİŞ

Sinyal (işaret), fiziksel değişkenlerin durumu hakkında bilgi taşıyan ve matematiksel olarak fonksiyon (işlev) biçiminde gösterilen kavrama denir. Bir mikrofonun girişindeki basınç bilgisinin zamana göre olan ifadesi $p(t)$ şeklinde gösterilebilirken, çıkışındaki elektriksel sinyali de $x(t)$ olarak ifade edilebilir. Burada $x(t)$ ses sinyalidir ve t zaman değişkenidir. Bu sinyali t ekseninde çizdirmek suretiyle de elektriksel ses bilgisinin grafiğini zamana bağlı elde etmiş oluruz.

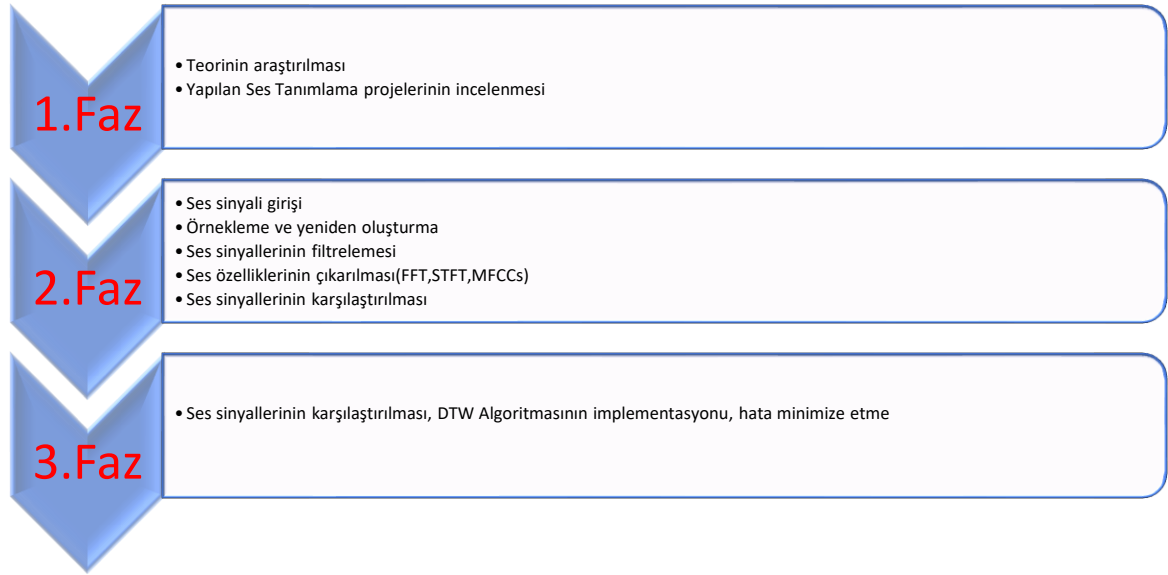
Bir akustik ses dalgasının uzay ve zaman değişkenleri cinsinden kısmi diferansiyel denklem şeklindeki ifadesi, bir konuşmacının sesinin havada oluşturduğu basınç dalgaları ve bu dalgaların bir mikrofon tarafından çevrildiği eşdeğer elektrik voltaj ve akım seviyesinin zamana bağlı ifadesi birer sinyal örneğidir.

Sinyaller temelde ikiye ayrılır

- Analog sinyal: hem bağımsız değişken hem de bağımlı değişken sürekli değerler alabilir.
- Sayısal sinyal: her iki değişken de ayrık değerler alır.

Analog işaretler, taşıdıkları bilgiyi uzay ve zamanın sürekli bir fonksiyonu olarak ifade ederler. İşaretin aldığı değerleri de yine sürekli olarak ifade edilebilir. Örneğin hava basıncı herhangi bir t anında sürekli değer alabilen bir değişkendir. Sayısal işaretler ise taşıdıkları bilgiyi örnek uzayında tutar, ayrık zamanda (zamandan bağımsız) sinyaller belirtirler.

Projede ses sinyali mikrofon yardımıyla alınır. Sürekli zamandaki ses sinyali ayrık zamana taşınır ve örnekler alınır. Örnekleme sonucunda sinyal üzerinde gelen ortamdan kaynaklı gürültü bileşenleri filtreden geçirilerek sinyalin işlenmesi kolaylaştırılır. Filtreden sonra sinyalin özelliklerinin çıkarılması işlemine geçilir. Özellik çıkarma işleminden sonra her bir farklı sinyal etiketlenerek dosya yoluna yazdırılır. Daha sonra program başladığında kullanıcının mikrofona ilettiği ses sinyali(kullanıcının konuşması) tekrar örnekleme tabii tutulur ve filtreden geçirilerek özellikleri çıkartılır. Yeni gelen sinyalin özellikleri, daha önceden etiketlenen sinyallerin özellikleriyle karşılaştırılır ve en yakın ses işareti tespit edilerek yeni giriş işareti etiketlenir. Şekil-1’de projenin zaman çizelgesi verilmiştir.



Şekil 1 Proje Süreç Çizelgesi

2) NYQUIST ÖRNEKLEME TEOREMİ

Nyquist örnekleme teoremi, dijital sinyal işleme alanında, sürekli zaman sinyalleri ve ayrık zaman sinyalleri arasında temel bir köprü görevi gören bir teoremdir. Ayrı bir örnek dizisinin, sonlu bant genişliği sürekli-zaman sinyalinden tüm bilgileri yakalamasına izin veren bir örnekleme hızı için yeterli bir koşul oluşturur. Açıkçası, teorem sadece sonlu bir frekans bölgesinin dışında sıfır olan bir Fourier dönüşümüne sahip bir matematiksel fonksiyon sınıfı için geçerlidir. Bir kişi sürekli bir fonksiyonu ayrık forma indirgediğinde ve sürekli bir fonksiyona geri girdiğinde, sonucun aslına uygunluğunun orijinal numunelerin yoğunluğuna (veya numune hızına) bağlı olmasını bekleriz. Örnekleme teoremi, belirli bir bant genişliğiyle bantla sınırlı olan fonksiyonlar sınıfı için mükemmel uygunluk için yeterli olan bir örnekleme hızı kavramını tanıtır, böylece örnekleme sürecinde hiçbir gerçek bilgi kaybolmaz. Fonksiyon sınıfı için bant genişliği açısından yeterli örnekleme hızını ifade eder. Teorem ayrıca, numunelerden orijinal sürekli zaman fonksiyonunun mükemmel bir şekilde yeniden yapılandırılması için bir formüle yol açar.

Örnekleme, bir sinyalin bir değer dizisine dönüştürülmesi işlemidir. Eğer bir $x(t)$ fonksiyonu B hertz' den daha yüksek frekans içermiyorsa, koordinatları $1 / (2B)$ aralıklı saniye arayla verilir. bir dizi noktada vererek verilir saniye arayla. Bu nedenle, yeterli bir örnekleme hızı, saniyede $2B$ örnekten daha büyük bir şeydir. Aynı şekilde, belirli bir örnekleme hızı f_s , $B < \frac{f_s}{2}$ bir bant sınırı için mükemmel bir yeniden yapılandırma mümkündür. Örneklerin $x(t)$ 'yi temsil etmeleri için örneklerin Nyquist oranını aşmaları gerekir. $\frac{f_s}{2}$ eşiğine Nyquist frekansı denir ve örnekleme ekipmanının(A-D-C çevirici) bir özelliğidir. Düzgün örneklenmiş $x(t)$ 'nin tüm anlamlı frekans bileşenleri Nyquist frekansının altında bulunur. Bu eşitsizliklerle tanımlanan duruma Nyquist kriteri veya bazen Raabe durumu denir. Teorem, sayısallaştırılmış görüntü söz konusu olduğunda boşluk gibi diğer alanların işlevlerine de uygulanabilir. Diğer alanlardaki tek değişiklik, t , f_s ve B 'ye uygulanan ölçü birimleridir.

Analog ses sinyalinin örnekleme frekansını Hertz olarak (f_s) 16000Hertz seçtik. Dolayısıyla filtre çıkışındaki ve ses işleme süreçleri (Fourier transformları, MFCCs...) sonrasında ses sinyalinde örtüşme olmasının önüne geçmiş olduk.

3) FİLTRELEME

Sinyal işlemede filtreleme, sinyaldeki bazı istenmeyen bileşenleri kaldıran bir cihaz veya prosestir . Filtreleme, bir sinyal işleme sınıfıdır; filtrelerin tanımlayıcı özelliği, sinyalin bazı yönlerinin tamamen veya kısmen bastırılmasıdır. Çoğu zaman, bu bazı frekansların veya frekans bantlarının çıkarılması anlamına gelir. Filtreler özellikle elektronikte ve ayrıca telekomünikasyon, radyo, televizyon, ses kaydı, radar, kontrol sistemleri, müzik sentezi, görüntü işleme ve bilgisayar grafikleri alanlarında sıkça kullanılmaktadır.

Filtreleri sınıflandırmanın birçok farklı temeli vardır ve bunlar birçok farklı şekilde örtüşür; basit bir hiyerarşik sınıflandırma yoktur. Filtreler şunlar olabilir:

- Lineer veya Non-Lineer
- Zamanda Değişmez veya Zamanda Değişir
- Analog veya Dijital
- IIR(sonsuz dürtü yanıtı) veya FIR(sonlu dürtü yanıtı)

IIR filtrelerin FIR filtrelere göre karşılaştırılması

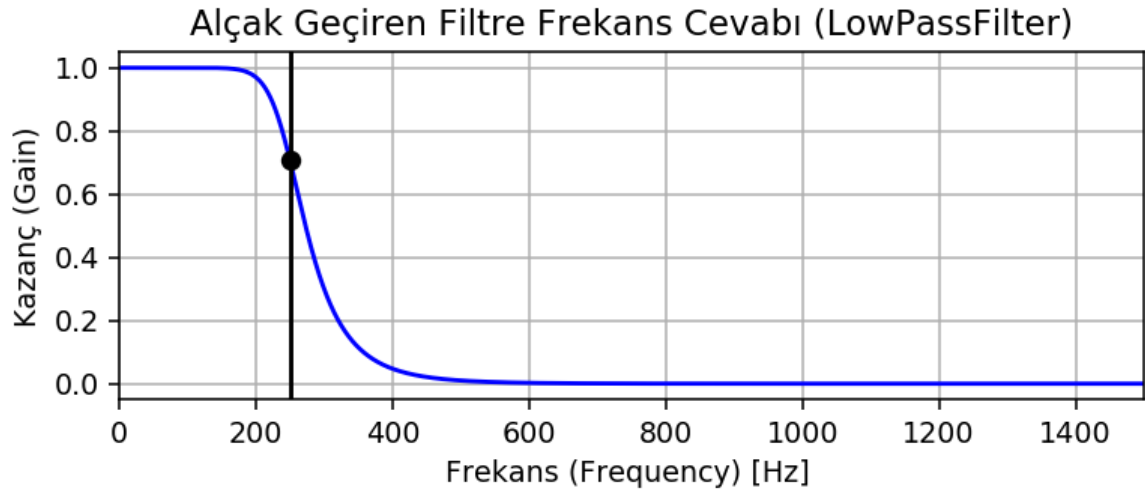
	IIR FİLTRELER	FIR FİLTRELER
Faz (grup gecikmesi)	Kontrol etmesi zordur. Özel bir teknik bulunmamaktadır.	Lineer faz her zaman mümkündür.
Kararlılık (Stability)	Kararsız olabilmektedir.	Her zaman kararlıdır.

Şekil 2 IIR filtreler ve FIR filtrelerin karşılaştırılması

Ses sinyali üzerindeki gürültü bileşenlerinin giderilmesi ve STFT ve MFCCs analizlerinde yüksek doğruluk elde edilmesi amacıyla alçak geçiren filtreden geçirilerek, yüksek frekanslardaki gürültü bileşenleri giderilmiştir. Bu karara sinyalin fourier spektrumunun incelenmesi ve konuşma sesinin düşük frekanslarda bulunması (özellikle 2000-3000Hertz arasında güç spektral yoğunluğunun bulunmasının gözlemlenmesi sonucu karar verilmiştir. FFT konusu işlenirken konuya daha ayrıntılı değinilecektir.

Filtrenin sonsuz dürtü cevabıyla oluşturduk. Filtrenin IIR seçilmesindeki kararımızı etkileyen faktör; IIR filtrelerinin benzer filtreleme işlemlerini gerçekleştirmek için genellikle daha az katsayı gerektirmesi, IIR filtrelerinin daha hızlı çalışması ve daha az bellek alanı gerektirmesidir. Programımız veri setlerini karşılaştırırken yeterince uzun süreye sahiptir. Dolayısıyla filtreleme aşamasından IIR filtre seçerek daha hızlı hesap yapabilmesinin önünü açmış olduk. Filtre spesifikasyonları şu şekildedir:

1. ButterWorth Alçak Geçiren Filtre (BW Low Pass Filter)
2. N, derece(rank) = 6
3. Kesme Frekansı (Cut-Off Frequency) = 250 Hertz

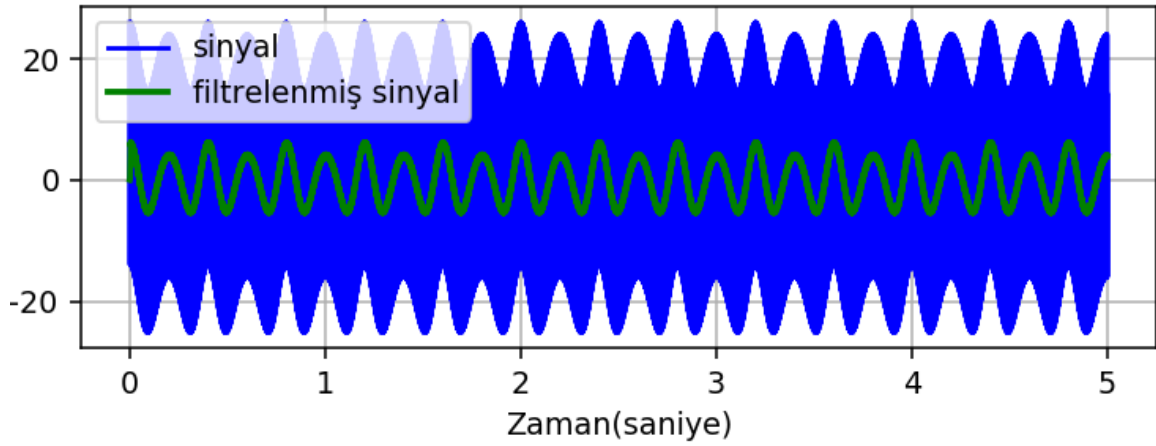


Şekil 3 Tasarlanan Filtrenin Frekans Spektrumu

Bu filtrenin davranışını daha yakından incelemek için filtre girişine

$$5 * \cos(10 * \pi * t) + \cos(15 * \pi * t) + 20 * \cos(1000 * \pi * t)$$

Sinyalini uyguladığımız zaman çıkış şu şekilde olmaktadır:



Şekil 4 Filtre girişine uygulanan sinyalin çıkışta gözlemlenmesi

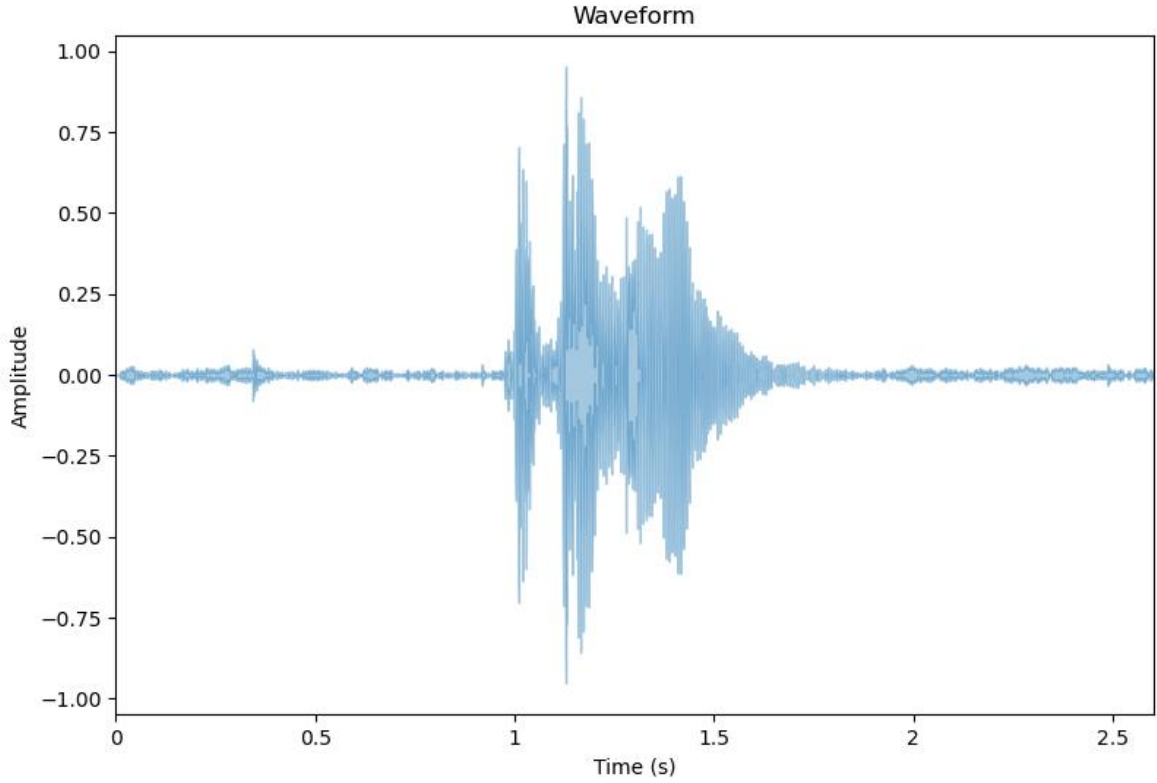
Grafikten de görüldüğü üzere filtre yüksek frekansları süzmekte ve alçak frekansları geçirmektedir.

4) FAST FOURIER TRANSFORM

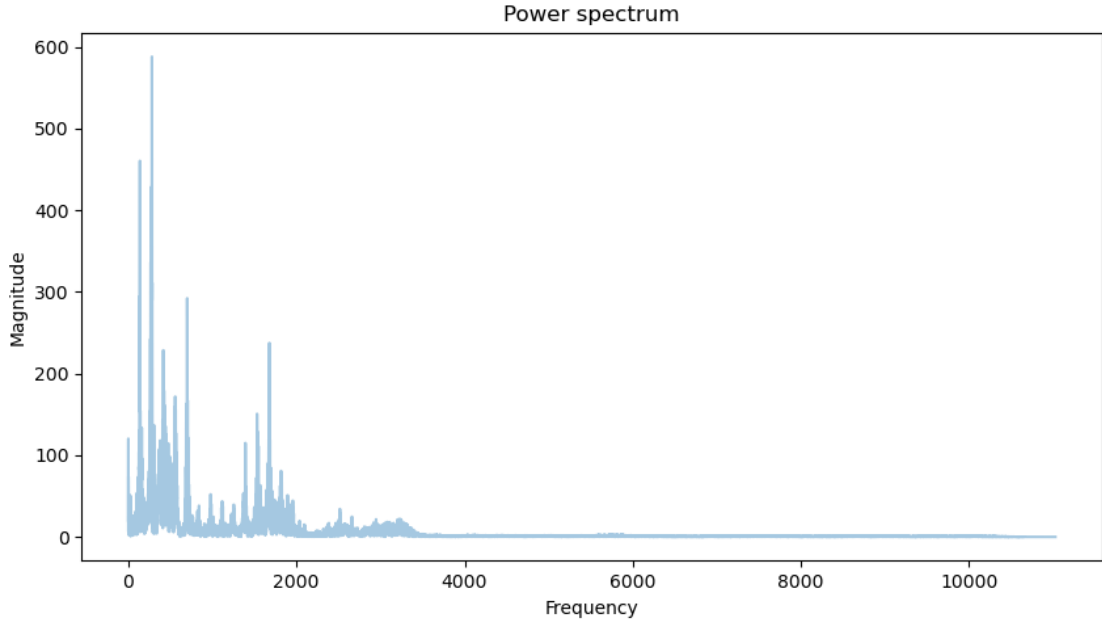
Hızlı Fourier dönüşümü (Fast Fourier Transform-FFT) titreşim analizinde kullanılan, istatistik tabanlı, matematiksel bir işlemdir. Karışık sinyal ayrıştırır ve hangi frekansta ne şiddette bir titreşim olduğunu gösterir. Kısaca FFT sinyallerimizi zaman alanından frekans alanına geçirirken kullandığımız bir işlemdir. FFT tekrarlanmayan sinyalleri dikkate almaz. Karmaşık sinyaller içinde periyodik olanları belirleyip harmonik bileşenlerine ayırır.

Hızlı Fourier dönüşümleri mühendislik, müzik, bilim ve matematik uygulamalarında yaygın olarak kullanılmaktadır. Temel fikirler 1965'te popülerleştirildi, ancak bazı algoritmalar 1805 kadar erken türetilmişti. IEEE Bilim ve Mühendislik Dergisi Bilgisayar Bilimi dergisi tarafından 20. Yüzyılın İlk 10 Algoritmasına dahil edildi.

Örnek bir ses sinyali olan, mikrofona söylenen “tekerleme” kelimesinin fft sinin alınması sonucu güç spektral yoğunluk grafiği görülmektedir.



Şekil 5 Zaman(time) domaininde giriş sinyali



Şekil 6 Sinyalin Frekans domaininde güç spektrumu

Frekansların 2000Hertz-2500Hertz bandında yoğunlaştığı ve daha sonraki bileşenlerin zayıf olduğu görülmektedir. Biz de bu grafiği yorumlayarak alçak geçiren filtrenin kesim frekansını belirledik.

5) KISA ZAMANLI FOURİER DÖNÜŞÜMÜ(SHORT TIME FOURİER TRANSFORM)

Kısa süreli Fourier dönüşümü (STFT), pencereli bir sinyalin Fourier dönüşümlerinin bir dizisidir. STFT, bir sinyalin frekans bileşenlerinin zaman içinde değiştiği durumlar için zamana göre ayarlanmış Frekans bilgisi sağlarken standart Fourier dönüşümü, tüm sinyal zaman aralığı boyunca ortalama frekans bilgilerini sağlar.

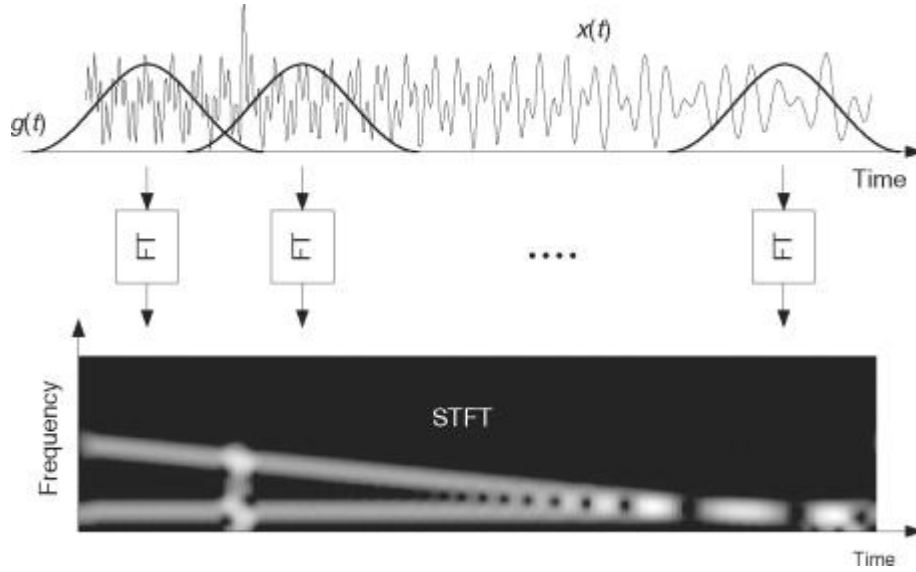
STFT çiftleri şu şekilde ifade edilir:

$$X_{STFT}[m, n] = \sum_{k=0}^{L-1} x[k]g[k - m]e^{-\frac{j2\pi nk}{L}}$$

$$x[k] = \sum_m \sum_n X_{STFT}[m, n]g[k - m]e^{j2\pi nk/L}$$

burada $x[k]$ bir sinyali ve $g[k]$ ise bir L-nokta pencere fonksiyonunu belirtir. $x[k]$ 'nin STFT'si, ve

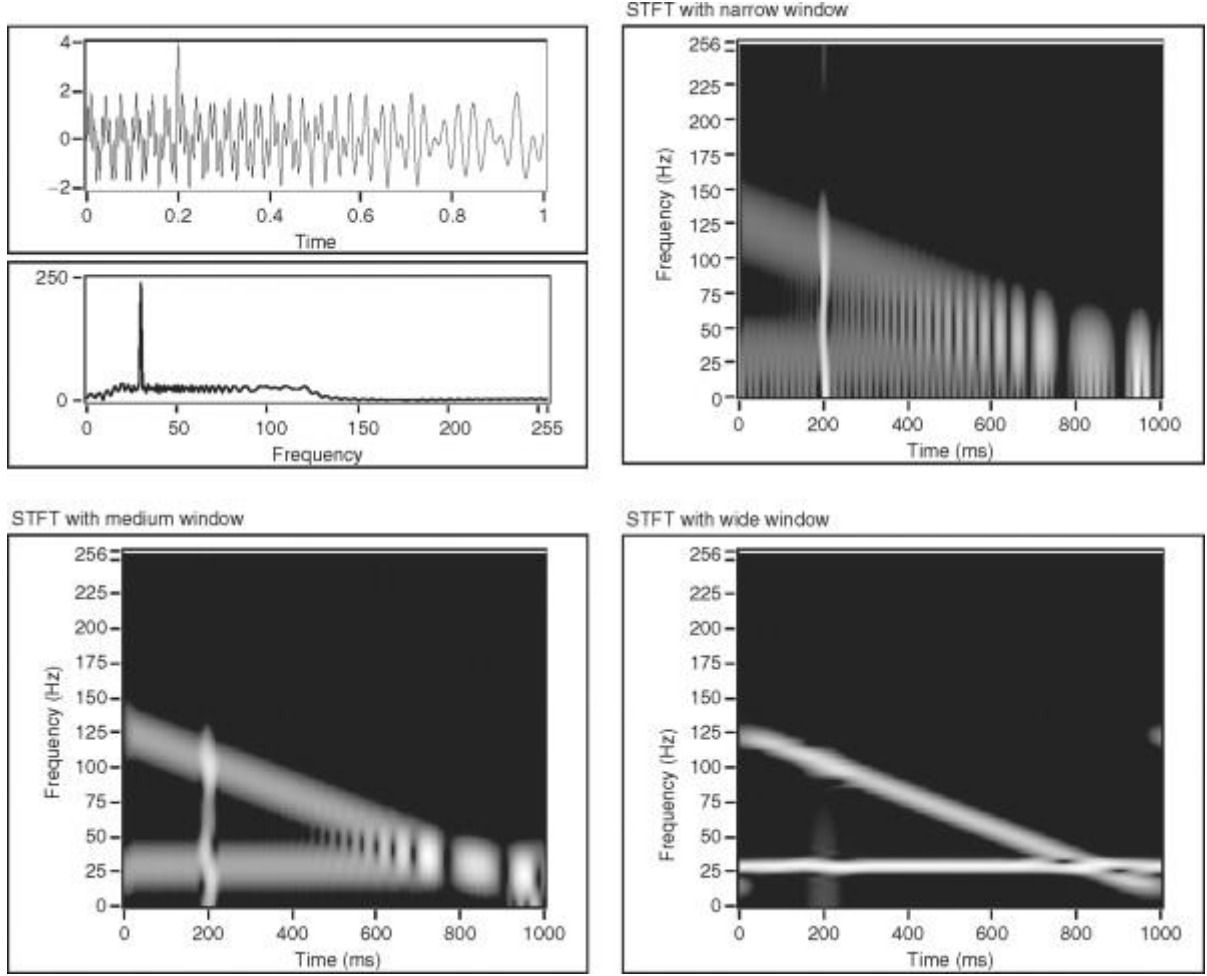
$x[k]g[k - m]$ 'nın Fourier dönüşümü olarak yorumlanabilir. Aşağıdaki şekilde, pencereli bir sinyalin Fourier dönüşümlerini alarak STFT hesaplamasını göstermektedir.



Şekil 7 Pencerelenen sinyalin STFT dönüşümü (zaman-frekans) gösterimi

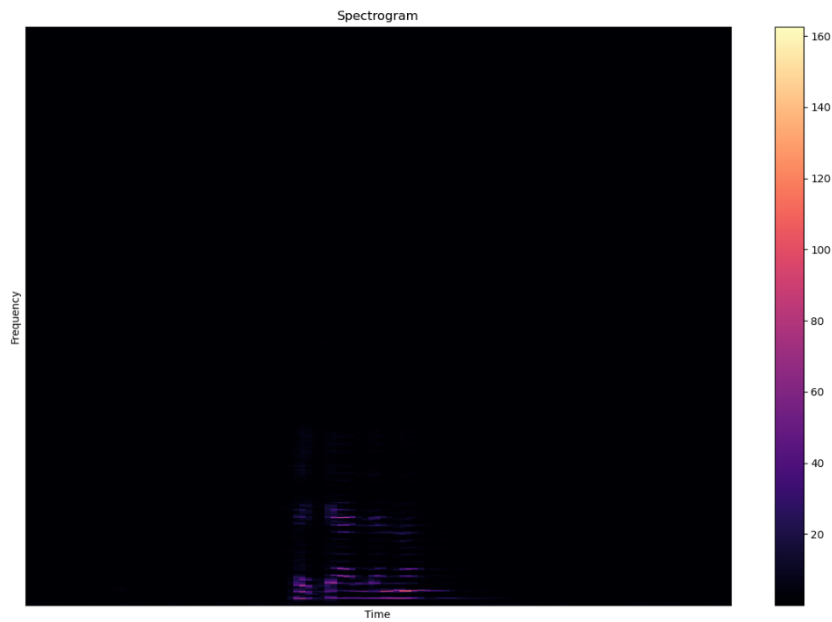
STFT'de zaman ve frekans çözünürlüğü arasında bir denge vardır. Diğer bir deyişle, dar genişlikli bir pencere zaman alanında daha iyi bir çözünürlük sağlarsa da, frekans alanında zayıf bir çözünürlük üretir ve bunun tersi de geçerlidir. STFT'nin görselleştirilmesi genellikle zaman içinde STFT büyüklüğünün yoğunluk grafiği olan spektrogramı ile gerçekleştirilir. Farklı zaman-frekans çözünürlüklerini gösteren üç spektrogram aşağıdaki şekilde gösterilmektedir.

Konuşma Sinyalleri için STFT kullanılmasının sebebi, sesli harfler gibi sabit durum sesleri, doğrusal bir sistemin periyodik olarak uyarılması ile üretilir. Konuşma spektrumu, uyarım spektrumunun ve vokal kanalı frekans yanıtının ürünüdür. Konuşma zamanla değişen bir sinyaldir. İhtiyaç zamanla değişen özellikleri yansıtmak için daha sofistike analiz- hece oranlarında (~ 10 kez/saniye) değişiklikler meydana gelir- 10-30 msn'lik sabit zaman aralıklarında, çoğu konuşma sinyalinin özellikleri nispeten sabittir.

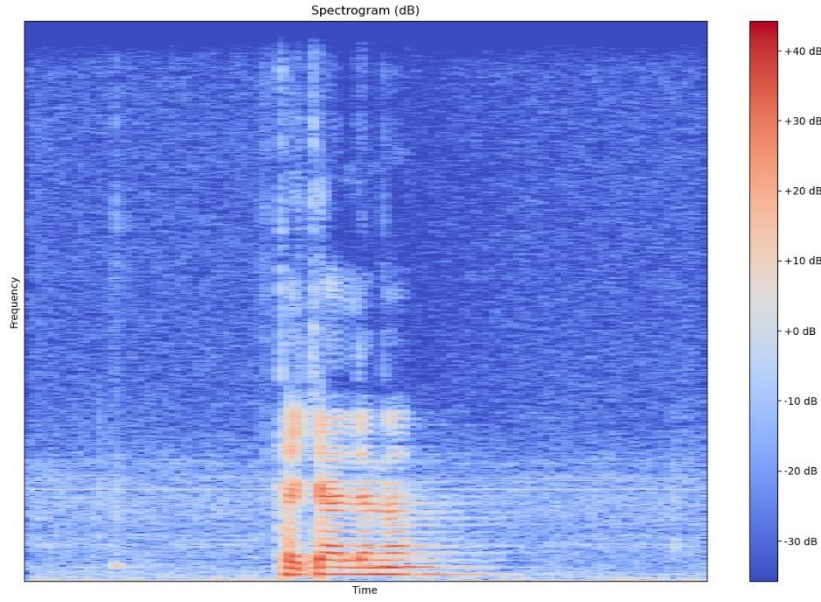


Şekil 8 Farklı zaman ve frekans değerlerine sahip işaretlerin STFT spektrogramları

Örnek olarak yukarıda zaman(time) ve fft domaininde verilen “tekerleme” kelimesinin ses çıktısının STFT spektrogramı şu şekildedir.



Şekil 9 Sinyalin STFT Spektrogramı



Şekil 10 dB ölçeğinde STFT spektrogramı

6) MFCC MEL FREKANS SPEKTRUMU

Ses işlemede, Mel-frekanslı Kepstrum (MFC), bir sesin kısa süreli güç spektrumunun, doğrusal olmayan bir frekans ölçeğindeki bir log güç spektrumunun doğrusal bir kosinüs dönüşümüne dayanan bir temsildir. Mel frekanslı kepstral katsayıları (MFCC'ler) toplu olarak bir MFC oluşturan katsayılardır. Bunlar, ses klibinin bir tür kepstral gösteriminden türetilir. Kepstrum ve Mel-Frekanslı Kepstrum arasındaki fark, MFC'de, frekans bantlarının, insan işitsel sisteminin tepkisini, normal Kepstrumda kullanılan doğrusal aralıklı frekans bantlarından daha yakın olan ölçeğinde eşit aralıklı olmasıdır. Bu frekans eğrilmesi, örneğin ses sıkıştırmasında sesin daha iyi temsil edilmesine izin verebilir. MFCC'ler genellikle aşağıdaki gibi türetilir:

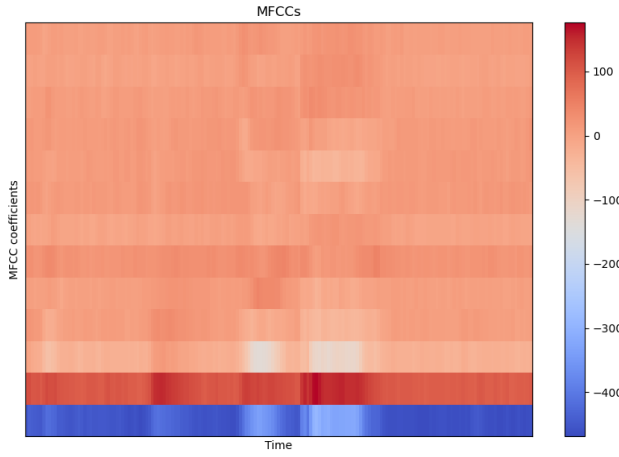
- Bir sinyalin Fourier dönüşümünü (pencereli bir alıntı) alın.
- Üçgen üst üste binen pencereler kullanılarak yukarıda elde edilen spektrumun güçlerini mel ölçeğine eşleyin.
- Mel frekanslarının her birindeki güç kütüklerini alın.
- Bir sinyalmış gibi, mel log güçleri listesinin ayırık kosinüs dönüşümünü alın.
- MFCC'ler, elde edilen spektrumun genlikleridir.

Bu işlemde farklılıklar olabilir, örneğin: ölçeği eşlemek için kullanılan pencerelerin şekli veya aralarındaki farklılıklar, [3] veya "delta" ve "delta-delta" (birinci ve ikinci gibi dinamik özelliklerin eklenmesi) doğrudan kare-kare farkı katsayıları. 2000'li yılların başlarında Avrupa Telekomünikasyon Standartları Enstitüsü, cep telefonlarında kullanılacak standartlaştırılmış bir MFCC algoritması tanımladı.

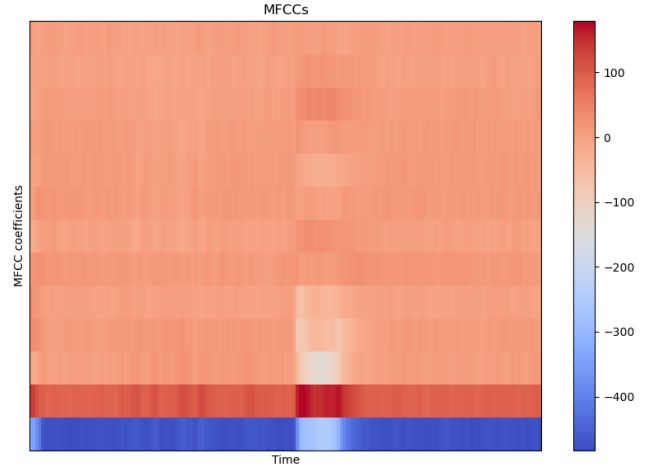
Veritabanındaki ve sonradan mikrofonla kullanıcı tarafından giriş yapılan ses sinyallerinin mel frekans spektrumlarının alınması sonucunda mel frekans kepstrum katsayılarının karşılaştırılması

yapılmıştır. Karşılaştırma temelde; mel frekans kepstrum katsayılarının arasındaki farkları sırasıyla hesaplamakta ve en yakın masafede olan MFCCs özelliğine sahip etiketli dosyayı tespit etmektedir. Aşağıdaki grafiklerde 'a' hecesinin kullanıcılar tarafından mikrofona söylenilmesi sonucu kaydedilen ses dosyalarının MFCCs katsayılarının gösterimi verilmiştir.

Şekil1'de 'a' hecesinin daha önceden bir kadın tarafından söylenip kaydedilmesi ve bu ses sinyalinin MFCCs katsayılarının zamana bağlı değişimi verilmiştir. Şekil:2'de ise bir kullanıcının yine 'a' harfini söylemesi sonucu bu ses sinyalinin MFCCs katsayılarının çıkarılması ve grafikte zamana bağlı gösterimi verilmiştir.



Şekil 12 Giriş sinyali 'a', kadın sesi



Şekil 11 Giriş sinyali 'a', erkek sesi

7) MFCCs KATSAYILARININ KARŞILAŞTIRILMASI

İki farklı MFCCs katsayıları alındıktan sonra karşılaştırmanın yapılması DTW Dinamik Zamanda Bükme(Dynamic Time Warping) algoritması yardımıyla yapılmıştır.

Dynamic Time Warping

Zaman serisi analizinde, Dinamik Zamanda Bükme (DTW), hızda değişebilen iki zamansal dizi arasındaki benzerliği ölçmek için kullanılan algoritmalarından biridir. Örneğin, bir kişi diğerinden daha hızlı yürüyor olsa da veya bir gözlem sırasında hızlanmalar ve yavaşlamalar olsa bile, yürüyüşteki benzerlikler DTW kullanılarak tespit edilebilir. DTW, video, ses ve grafik verilerinin zamansal dizilerine uygulanmıştır.

Doğrusal bir diziye dönüştürülebilen herhangi bir veri DTW ile analiz edilebilir. İyi bilinen bir uygulama, farklı konuşma hızlarıyla başa çıkmak için otomatik konuşma tanıma olmuştur. Diğer uygulamalar arasında hoparlör tanıma ve çevrimiçi imza tanıma sayılabilir. Kısaca DTW belirli iki kısıtlama ve kurallarla belirli iki dizi (örneğin zaman serileri) arasındaki optimum eşleşmeyi hesaplayan bir yöntemdir:

- İlk dizideki her indeks, diğer diziden bir veya daha fazla indeks ile eşleştirilmelidir ve bunun tersi de geçerlidir.
- İlk dizideki ilk indeks diğer dizideki ilk indeks ile eşleşmelidir
- İlk dizideki son dizin, diğer dizideki son dizinle eşleştirilmelidir

- İlk diziden indekslerin diğer sekanstan endekslere eşlenmesi monoton olarak artmalı ve bunun tersi de olmalıdır, örneğin, $j > i$ ilk dizini indekslerse, diğer dizinde $i > k$ durumunda indis olmamalıdır. iki endeks olmamalı $i > j$ endeksi k endeksi ile eşleşecek şekilde diğer dizideki $i > k$ 'dir. Böylece, indeks i , indeks j ile ve indeks j , indeks k ile eşleşir.

Optimal eşleşme, tüm kısıtlamaları ve kuralları karşılayan ve maliyetin mutlak farklılıkların toplamı olarak hesaplandığı minimum değerler içeren eşleşme ile gösterilir. Diziler, zaman boyutundaki belirli doğrusal olmayan varyasyonlardan bağımsız olarak benzerliklerinin bir ölçüsünü belirlemek için zaman boyutunda doğrusal olmayan bir şekilde "bükülür". Bu dizi hizalama yöntemi genellikle zaman serisi sınıflandırmasında kullanılır. DTW, verilen iki sekans arasındaki mesafe benzeri bir miktarı ölçmesine rağmen, üçgen eşitsizliğinin tutulacağını garanti etmez. Bir önceki örnekte verilen iki farklı kişinin aynı heceyi söylemesi sonucu çıkarılan MFCCs katsayılarının farkının DTW algoritması ile hesaplanması sonucu terminal çıktısı şu şekildedir:

The normalized distance between the two : 11377.29193716495

Şekil 13 iki sinyalin MFCCs katsayı farkları, program terminal çıktısı

Veri setlerinin giriş (input)'a karşılık karşılaştırılması; dtw algoritma çıktıların arasındaki farkın minimum olduğu etiketli ses dosyasının etiketinin çıktı verilmesine yol açar. Örnek sırasıyla kullanıcının 'a', 'a', 'e' hecelerini mikrofona söylemesi sonucu program terminal çıktısı şu şekildedir:

```
*** recording
* done recording
Your voice has succesfully recorded!
Processing. Please wait...
a
a
e
```

Şekil 14 Ses kaydı ve analizi, program terminal çıktısı

8) YORUMLAR VE DEĞERLENDİRME

Sesler dörder saniye aralıklarla alınarak program tarafından tanımlanabilmiş ve metin çıktısı sağlanabilmiştir. Yalnızca veri setlerinin çok olmamasından kaynaklı yanlış tespit etmeler olabilmektedir. Bu yöntem sesin tanımlanarak metne aktarılmasını sağladığından fiziksel engelli bireylerin hayatlarını kolaylaştırmada veya bir ortamdaki konuşmacıların seslerinin işlenmesinde, müziklerin metne aktarılmasının kolaylaştırılmasında kullanılabilir.

Kayıtlı veri tabanındaki seslerin büyük bir çoğunluğu dijital ortamdaki seslerini ilettiğinden, mikrofona doğrudan konuşmadığından dolayı gürültüler mevcut olmakla birlikte gürültülerin bir kısmı filtrelenerek ses tanımlamanın yüksek doğrulukta olması sağlanmıştır. Ayrıca filtrenin de daha hassas spesifikasyonlarda seçilmesi durumunda ses tanımlamanın doğruluk değerinin artması mümkündür. Program bir dakikalık bir sürede çalışmaktadır. Çalışma süresinin kısalması ve hızlı tespit için veri setlerinin derin öğrenmeyle eğitilmesi ve test edilmeleri gerekmektedir.

9) KAYNAKLAR

- <https://www.minidsp.com/applications/dsp-basics/fir-vs-iir-filtering>
- <https://ocw.mit.edu/courses/electrical-engineering-and-computer-science/6-341-discrete-time-signal-processing-fall-2005/lecture-notes/lec08.pdf>
- <https://www.sciencedirect.com/topics/engineering/short-time-fourier-transform>
- Min Xu; et al. (2004). "HMM-based audio keyword generation" (PDF). In Kiyoharu Aizawa; Yuichi Nakamura; Shin'ichi Satoh (eds.). Advances in Multimedia Information Processing – PCM 2004: 5th Pacific Rim Conference on Multimedia. Springer. ISBN 978-3-540-23985-7. Archived from the original (PDF) on 2007-05-10.
- P. Mermelstein (1976), "Distance measures for speech recognition, psychological and instrumental," in Pattern Recognition and Artificial Intelligence, C. H. Chen, Ed., pp. 374–388. Academic, New York.
- <https://community.sw.siemens.com/s/article/introduction-to-filters-fir-versus-iir>
- <https://ocw.mit.edu/courses/electrical-engineering-and-computer-science/6-345-automatic-speech-recognition-spring-2003/lecture-notes/>
- <http://isle.illinois.edu/~hasegawa/notes/>