T.C.
YILDIZ TEKNİK ÜNİVERSİTESİ
FEN BİLİMLERİ ENSTİTÜSÜ

KONUŞMACI VE KONUŞMA TABANLI KASKAD BİYOMETRİK KONTROL SİSTEMİ TASARIMI

ROAYA SALHALDEN ABDALRAHMAN

YÜKSEK LİSANS TEZİ
ELEKTRONİK VE HABERLEŞME MÜHENDİSLİĞİ ANABİLİM DALI
ELEKTRONİK PROGRAMI

DANIŞMAN
DR. ÖĞR.ÜYESİ NİHAN KAHRAMAN

İSTANBUL, 2018

T.C.

YILDIZ TEKNİK ÜNİVERSİTESİ

FEN BİLİMLERİ ENSTİTÜSÜ

KONUŞMACI VE KONUŞMA TABANLI KASKAD BİYOMETRİK KONTROL SİSTEMİ TASARIMI

Roaya Salhalden ABDALRAHMAN tarafından hazırlanan tez çalışması 18.05.2018 tarihinde aşağıdaki jüri tarafından Yıldız Teknik Üniversitesi Fen Bilimleri Enstitüsü Elektronik ve Haberleşme Mühendisliği Anabilim Dalı'nda **YÜKSEK LİSANS TEZİ** olarak kabul edilmiştir.

Tez Danışmanı

Dr.Öğr.Üyesi Nihan KAHRAMAN

Yıldız Teknik Üniversitesi

Jüri Üyeleri

Dr.Öğr.Üyesi Nihan KAHRAMAN

Yıldız Teknik Üniversitesi _____

Dr.Öğr. Üyesi Bülent BOLAT

Yıldız Teknik Üniversitesi _____

Doç. Dr. Mürvet KIRICI

İstanbul Teknik Üniversitesi _____

ÖNSÖZ

Bize bilim ve bilgi yolunu aydınlatan ve bu görevi nasip edip yerine getirmemize yardımcı olan Tanrı'ya bin şükür.

Bu çalışmanın konusunun belirlenmesinde ve yürütülmesi sırasında değerli bilgilerini ve zamanını benden esirgemeyerek her fırsatta çalışmamla yakından ilgilenen, eleştirileriyle yol gösteren danışman hocam Nihan Kahraman'a teşekkür ve minnetimi özellikle belirtmek isterim. Ayrıca yoğun çalışmaları sırasında zamanını bana ayırıp yüksek lisans tezimi değerlendirdiği ve destek verdiği için Dr. Öğr. Üyesi Bülent Bolat hocama teşekkürü borç bilirim.

Bu çalışmayı, bu varlıktaki en değerlilerime, sevgili anneme ve babama adadım. Tabii ki, sevgili kocamın bana yardım etmesini ve okuma hayatımda yanımda durmasını hiç unutamam bunun için kalbimin derinliklerinde ona teşekkür ediyorum. Küçük kızıma da, Sedra, teşekkür ediyorum, çünkü o benim gücümün ve ilham kaynağımın sebebiydi.

Mayıs, 2018

Roaya ABDALRAHMAN

İÇİNDEKİLER

Sayfa

SİMGE LİSTESİ vi

KISALTMA LİSTESİ vii

ŞEKİL LİSTESİ vii

ÇİZELGE LİSTESİ ix

ÖZET x

ABSTRACT xii

BÖLÜM 1
GİRİŞ 1
1.1 Giriş 1
1.2 Konuşmacı Tanıma 4
1.2.1 Tanımlama ve Doğrulama 4
1.2.2 Metin Bağımlı ve Metin-Bağımsız Konuşmacı Belirleme 6
1.2.3 Açık Set Ve Kapalı Set 7
1.3 Konuşma Tanıma 7
1.3.1 Konuşma Tanıma Türleri 7
1.3.2 Karşılaşılabilecek Güçlükler 8
1.4 Araştırma Amaçları 9
1.5 Tez Organizasyonu 9

BÖLÜM 2

LİTERATÜR TARAMASI 11

2.1 Literatürdeki Çalışmalar 11

BÖLÜM 3

METODOLOJİ 17

3.1 Giriş 17
3.2 Konuşmanın Önemi 18
3.3 Konuşma Kayıt 19
3.4 Ses Sinyali Özelliği 20
3.4.1 Ses ve Sessizlik Tespiti 20
3.4.2 Özellik Çıkarma 20
3.4.2.1 Mel Ölçek Süzgeç Dizisi Enerjilerinin Kepstral Gösteriminin(MFCC) 20
3.4.2.2 Pitch Periyot Yöntemi 21
3.5 Çapraz Korelasyon (CC) 22
3.6 Otokorelasyon (AC) 23
3.7 Sinyal Enerjisi Ve Sinyal Gücü 23

BÖLÜM 4

ÖNERİLEN YÖNTEM VE SONUÇ 28

4.1 Önerilen Yöntem 28
4.2 Konuşmacı Tanıma 28
4.3 Konuşma Tanıma 31

BÖLÜM 5

SONUÇ VE ÖNERİLER 35

KAYNAKLAR 36

EK-A 38

ÖZGEÇMİŞ 61

SİMGE LİSTESİ

Mel	Mel Frekans Dönüşümü
f	Sinyâl'in Frekansı
log	Logaritma
x(n)	Giriş Sinyâl
d_t	Pitch Dedektörleri
r_t	Pitch Dedektörleri
W	Pencerenin Uzunluğu
τ	Pencerenin Uzunluğu Bir Zaman Alanı Hesaplaması Olduğunu Belirtir
E	Sinyal Enerjisi
P	Sinyal Gücü

KISALTMA LİSTESİ

AC	Otokorelasyon
ASI	Otomatik Konuşmacı Tanımlama
ASV	Otomatik Konuşmacı Doğrulama
CC	Çapraz Korelasyon
DCT	Ayrık Kosinüs Dönüşümü
DTW	Dinamik Zaman Eğrisi
FFT	Hızlı Fourier Transformasyon
HMM	Gizli Markov Modeli
IEEE	Institute of Electrical and Electronics Engineers
LCD	Sıvı Kristal Ekran
MF	Medyan Filtresi
MFCC	Mel Ölçek Süzgeç Dizisi Enerjilerinin Kepstral Gösteriminin (MFCC)
SNR	Sinyalin Gürültü Oranını
SR	Konuşmacı Tanıma
SRS	Konuşmacı Tanıma Sistemi
SS	Spektral Çıkarma
TIMIT	Farklı Cinsler ve Lehçelerde Amerikan İngiliz Konuşmacılarının Fonemik Ve Sözcüksel Olarak Konuşmalarını Yaptıkları Konuşma
UART	Evrensel Asenkron Alıcı / Verici
VRS	Ses Tanıma Sistemini
YSA	Yapay Sinir Ağı

ŞEKİL LİSTESİ

Sayfa

Şekil 1.1 Konuşmacı tanımlama sistemi 2
Şekil 1.2 Doğrulama sistemi 3
Şekil 1.3 Konuşmacı tanıma sınıflandırma systemi 4
Şekil 1. 4 Tanımlama ve doğrulama sistemlerinin 5
Şekil 1. 5 Konuşmacı tanımlama ve doğrulama sistemlerinin pratik örnekleri 6
Şekil 3. 1 Önerilen konuşmacı tanımlama yöntemi 18
Şekil 3. 2 Matlab'da ses kayıt akışı 19
Şekil 3. 3 MFCC'nin hesaplanması 21
Şekil 3. 4 Perde işaretlerinin kırmızı ile gösterildiği bir vokal fonem 22
Şekil 3. 5 SRS Algoritması 25
Şekil 3. 6 Ses sinyali yerleştirme 26
Şekil 3. 7 Sessizlik kaldırma algoritması 26
Şekil 3. 8 Özellik çıkartma yöntemini ve konuşmacı tanımasında eşleşen işlemi göstermektedir 27
Şekil 4. 1 Konuşmacı tanıma sisteminin blok diyagramı 29
Şekil 4. 2 Matlab'ta sinyalin girişi (a) orijinal giriş sinyali (b) ve gürültü bastırılmış versiyonu (c) 30
Şekil 4. 3 Konuşma tanıma sisteminin blok diyagramı 31
Şekil 4. 4 Konuşma tanıma için giriş sinyali ve giriş pitch periyotları 32
Şekil 4. 5 Tanıma sisteminin blok diyagramı 34

ÇİZELGE LİSTESİ

Sayfa

Çizelge 4.1 Konuşmacı tanıma için doğru pozitif, doğru negatif, yanlış pozitif ve yanlış negatif sonuçlar 30

Çizelge 4. 2 Konuşma tanıma için doğru pozitif, doğru negatif, yanlış pozitif ve yanlış negatif sonuçlar 33

Çizelge 4. 3 Basamaklı ses tanıma sistemi için doğru pozitif, doğru negatif, yanlış pozitif ve yanlış negatif sonuçlar 34

ÖZET

KONUŞMACI VE KONUŞMA TABANLI KASKAD BİYOMETRİK KONTROL SİSTEMİ TASARIMI

Roaya Salhalden ABDALRAHMAN

Elektronik ve Haberleşme Mühendisliği Anabilim Dalı

Yüksek Lisans Tezi

Tez Danışmanı: Dr.Öğr.Üyesi Nihan KAHRAMAN

Bu tezde biyometrik tanımlama sistemlerinden konuşmacı tanıma ve ardından konuşma tanıma ile kontrol edilebilen bir system için gerekli yazılım önerilmiştir. Çalışmanın ilk aşamasında metin bağımsız konuşmacı tanıma işlemi gerçekleştirilmiş, daha sonrasında ise belirlenen kullanıcıya ait şifre kelimenin tanınması ile birleştirilerek hibrid bir ses tanıma işlemi ortaya konmuştur. Böylece konuşmacı ve konuşma tanıma yöntemlerini birlikte kullanan bir ses biyometrik sistemi sunulmaktadır. Bu basamaklı yöntem, Yanlış Pozitif Oranını (FAR) azaltır ve biyometrik tanıma sisteminin güvenliğini artırır.

Mel ölçek süzgeç dizisi enerjilerinin kepstral gösteriminin (MFCC), konuşmacı tanıma için istenen öznitelik katkısını sağladığı bilindiğinden, çalışmanın ilk aşamasında bu yöntem kullanılmıştır; ancak MFCC'nin gürültüye eğilimli davranışı yüzünden hızlı bir gürültü bastırma yaklaşımı, yani gürültü kapısı, bu noktada kullanılmaktadır. Sinyalin gürültü oranının (SNR) konuşmanın bölümlenmemiş kısımlarında daha yüksek olması nedeniyle, konuşma sinyalinin bu kısımları kısa zaman enerjisi kullanılarak tespit edilir ve kesilir. Bir karenin kısa süreli enerjisi önceden tanımlanmış bir eşikten düşükse, gürültü olarak kabul edilir ve bu bölüm kaldırılır. Elde edilen MFCC değerleri biyometrik şablon olarak kullanılır. Hedef sisteme yeni giriş geldiğinde kullanılan yöntem ise şablon

olarak kaydedilen MFCC'ler ile test verisinin MFCC değerleri arasındaki Euclidian mesafe metriğinin hesaplanması ve belirli bir eşik değerinden büyük olup olmamasına dayanmaktadır. Ses tanıma aşaması ise otokorelasyon ve gelişmiş sinyal işlemeden elde edilen minimum saha frekansına göre yapılır.

Tez çalışmasında, kaskad ses tanıma yöntemi için MATLAB tabanlı bir yazılım sunulmuştur. Böylece ses tanıma özelliğini doğruluk, güvenlik ve penetrasyon zorluğu açısından daha iyi tanımladığı gösterilmiştir. Çalışmada elde edilen verilere göre, tanımlama sisteminin verimliliği yaklaşık % 91,2 olarak elde edilmiştir.

Anahtar Kelimeler: konuşmacı tanıma, mel frekansı cepstral katsayıları, pitch periyot

ABSTRACT

CASCADED BIOMETRIC CONTROL SYSTEM BASED ON SPEAKER AND SPEECH RECOGNITION

Roaya Salhalden ABDALRAHMAN

Department of Electronics and Communications Engineering

MSc. Thesis

Adviser: Asst. Prof. Dr. Nihan KAHRAMAN

In this thesis, a hybrid software is proposed for a system that can be controlled by speaker recognition and then speech recognition in biometric identification systems. In the first phase of the study, text-independent speaker recognition was performed, after that it is combined with the recognition of the password of the defined user, and a hybrid voice recognition process is revealed. Therefore a voice biometric system that uses speech and speech recognition methods together is presented. This concatenated method reduces the False Positive Rate (FAR) and increases the safety of the biometric recognition system.

Mel Frequency Cepstral Coefficient (MFCC) method is used in the first phase of the study, as it is known that it provides the desired attribute contribution for speaker recognition; however, due to the noise-prone behavior of the MFCC, a fast noise suppression approach, the noise gate is used at this point. Since the signal to noise ratio (SNR) is higher in the unpartitioned portions of the speech, these portions of the speech signal are detected using short time energy and interrupted. If the short-term energy of a square is lower than a pre-defined threshold, it is considered noise and this section is removed. The MFCC values obtained are used as biometric templates. The method used when new entry into the target system is based on whether the Euclidian

distance metric between the MFCC values of the test data and the MFCCs recorded as templates is greater than a certain threshold value. The voice recognition phase is based on the minimum field frequency obtained from autocorrelation and advanced signal processing.

In this thesis study, MATLAB-based software for cascade voice recognition method is presented. Thus, it has been shown that voice recognition is better defined in terms of accuracy, security and penetration difficulty. According to the data obtained in the study, the efficiency of the identification system was approximately 91.2%.

Keywords: Speaker recognition, mel frecquency cepstral cofficients, pitch period.

BÖLÜM 1

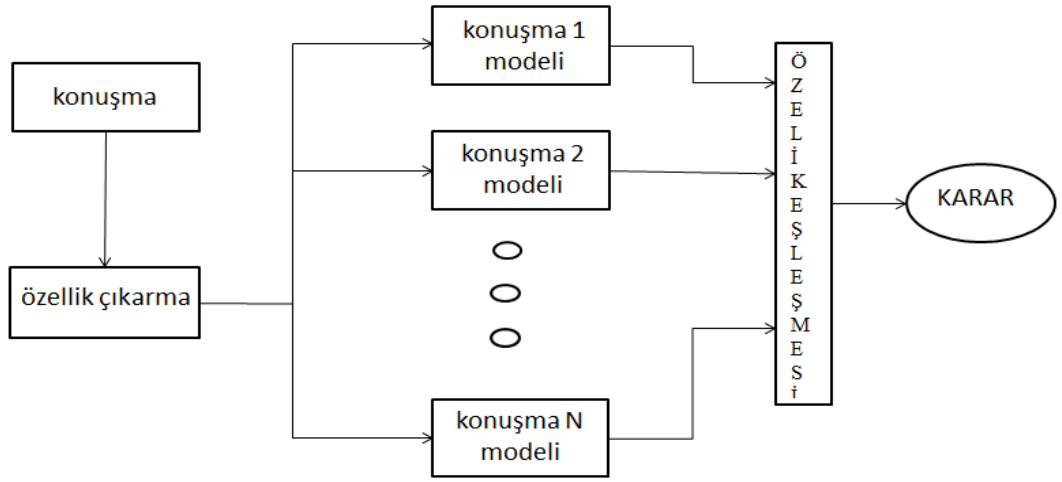
GİRİŞ

1.1 **Giriş**

İnternet ve bilgisayar çağında, son zamanlarda birçok yeni gelişme kaydedilmiştir. Dijital bilgisayar ve endüstriyel otomasyon döneminde, her şey bilgisayarlar vasıtasıyla yapılmakta ve takip edilmektedir Bankacılıktan otomatik nükleer enerji santrallerine kadar, matematik, kontrol teorisi ve imalat açısından inanılmaz ilerleme bulunmaktadır. Bütün bu gelişmeler, teknolojinin ilerlemesine katkıda bulunur. Ancak teknolojilerin ilerlemesiyle birlikte, güvenlik tehditleri bilgi, havaalanı, ev, uluslararası ve ulusal menkul kıymetler gibi çeşitli alanlarda artmıştır. Homeland Security kurumuna'a göre kimlik hırsızlığı yılda 56.6 milyar ABD Doları tutarında olduğu açıklanmıştır. Aynı makaleye göre, uzmanlar birçok olayın algılanmadığını veya bildirilmediğini söylemektedir. Güvenlik tehditlerinin ve dolandırıcılık işlemlerinin artması nedeniyle güvenilir kullanıcı kimlik doğrulama gereksinimi artmış ve bu nedenle biyometrik güvenlik sistemleri ortaya çıkmıştır. Biyometri alanının önemi gittikçe artmaktadır ve popüler işletim sistemlerinin birçoğu yalnızca bu özellikleri içermektedir ve yalnızca işletim sistemleri değil, birkaç popüler sosyal web sitesi farklı kişileri sınıflandırmak için bu özellikleri kullanmaktadır [1].

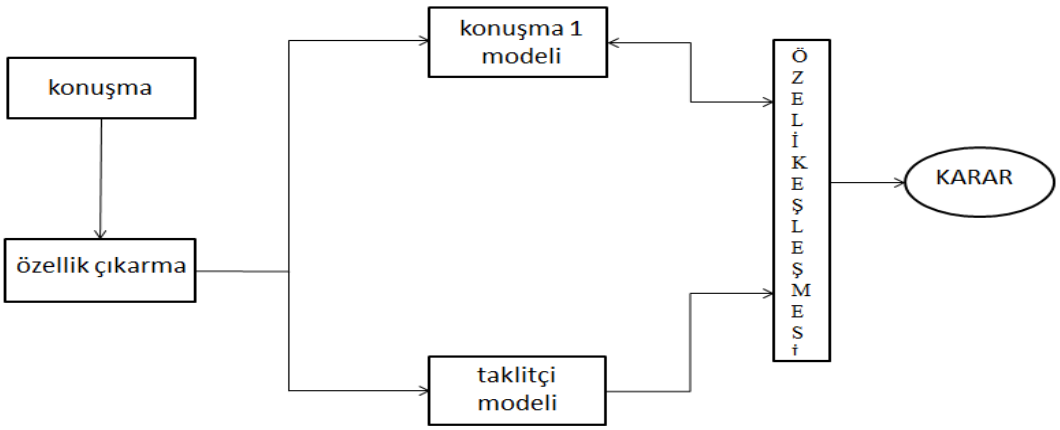
Davranışsal biyometri içerisinde yer alan konuşmacı tanıma, konuşma işaretlerinden kişilerin otomatik olarak tanınması görevidir. Bu teknik konuşmacı kimliğini doğrulamak ve güvenli hizmetlere erişimi kontrol etmek için konuşulan konuşmayı kullanmayı mümkün kılar, yani çevrimiçi işlemler, veritabanı erişim hizmetleri, bilgi hizmetleri, gizli

bilgi alanları için güvenlik kontrolü, bilgisayarlara uzaktan erişim vb. Konuşmacı Tanıma (Speaker Recognition SR), birçok farklı konuşmacıya özgü görevleri kapsayan önemli bir konudur. Reynolds'a (2002) göre, bu görevler metin bağımlı (konuşmacıların belirli bir metni söylemesi beklenir) ve metin bağımsız (konuşmacıların istediği her şeyi konuşabileceği) görevler olarak alt kategorilere ayrılabilir[2]. Metin bağımlı konuşmacı tanımada, tüm konuşmacılar aynı şeyi söylemesi ve konuşmacıların bu şekilde ayırt edilmesi hedeflenir. Bu tezin ana hedefi, metne bağımlı konuşmacı tanıma performansının doğruluğunu artırmaktır. Bu çalışmada doğruluğun arttırılması için literatürde de kullanılan Mel ölçek süzgeç dizisi enerjilerinin kepstral gösterimi (Mel Frequency Cepstral Coefficient, MFCC) algoritmaları denenmiştir. Bunun nedeni, çok popüler oldukları ve doğru sonuçlar vermeleridir. Konuşmacı tanıma (SR), öznitelik çıkarma ve sınıflandırma olmak üzere konuşmacı tanıma ve konuşmacı doğrulama bölümlerine bölünebilir. Temel konuşmacı tanımlama sistemi, Şekil 1.1'deki blok diyagramı ile gösterildiği şekilde sağlar.



Şekil 1.1 Konuşmacı tanımlama sistemi

Konuşmacının belirli özellikleri konuşma verilerinden çıkarılır ve önceden kaydedilmiş şablonlardan oluşturulan konuşmacı modelleriyle karşılaştırılır. Özelliklerin en çok uyan olduğu model meşru konuşmacı olarak seçilir. Çoğu durumda, karşılaştırma sonucunda model olasılık skoru oluşturur ve maksimum olasılık skoru üreten model seçilir.



Şekil 1.2 Doğrulama sistemi

Öte yandan, Şekil 1.2'deki blok diyagramı ile gösterilen konuşmacı doğrulama sistemi, bir konuşmacının kimliğini kabul veya reddeder. Özellikler, konuşma verilerinden çıkarılır ve daha önce kayıtlı verilerden yaratılan meşru konuşmacı modelinin yanı sıra taklitçi bir konuşmacı modeliyle karşılaştırılır. Konuşmacı modelinden üretilen olasılık puanı, taklitçi modelden çıkarılır. Sonuçta elde edilen skor eşik değerinden büyükse, konuşmacı meşru bir konuşmacı olarak kabul edilir. Her iki durumda da bu sistemleri kullanan kişilerin zaten kayıtlı olduğu beklenmektedir. Ayrıca bu sistemler metin bağımlı veya metin bağımsız olabilir. Metne bağımlı sistem, bir konuşmacıyı eğitmek ve test etmek için sabit bir ifade kullanır. Tam tersine, metin bağımsız sistem eğitim ve test amaçlı sabit bir cümle kullanmaz. Güvenlik yanında, konuşmacı tanıma çeşitli uygulamalara sahiptir ve hızla artmaktadır. Konuşmacı tanımanın uygulanabileceği alanlar şunlardır [3]:

1.Giriş Kontrolu

Güvenli fiziksel konumların yanı sıra gizli bilgisayar veritabanlarına bir konuşmacı sesi ile erişilebilir. Erişim, özel ve gizli web sitelerine de verilebilir.

2.Çevrimiçi İşlemler

Banka bilgisine erişmek veya bir öğeyi telefonla satın almak için kullanılan bir geçiş ifadesine ek olarak, konuşma sinyali bir güvenlik katmanı olarak kullanılabilir.

3.Yasa Uygulama

Adli analizde, konuşmacı tanıma sistemleri ek bilgi sağlamak için kullanılabilir. Tutukevinde çağrı izleme cezaevinde otomatik olarak yapılabilir.

4.Konuşma Veri Yönetimi

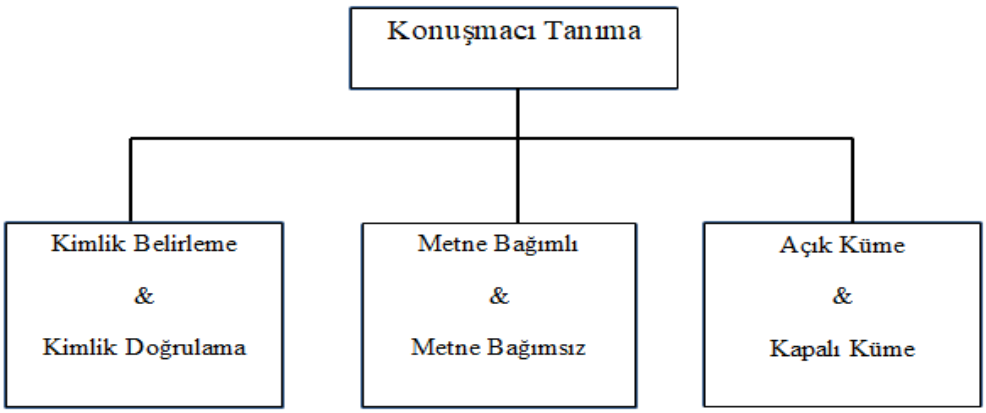
Sesli mesaj hizmetleri, sesli madencilik uygulamaları ve kayıtlı veya canlı bildirim toplantılar konuşmacıları otomatik olarak etiketlemek için konuşmacı tanıma özelliğini kullanabilir.

5.Multimedya ve Kişiselleştirme

Ses parçaları ve müzik otomatik olarak şarkıcı ve parça bilgileri ile etiketlenebilir. Web siteleri ve bilgisayarlar, hizmeti kullanan kişiye göre özelleştirilebilir.

1.2 **Konuşmacı Tanıma**

Konuşmacı tanıma, bir dizi kategoride sınıflandırılabilir. Şekil 1.3'de konuşmacı tanıma sınıflandırmaları görülmektedir.



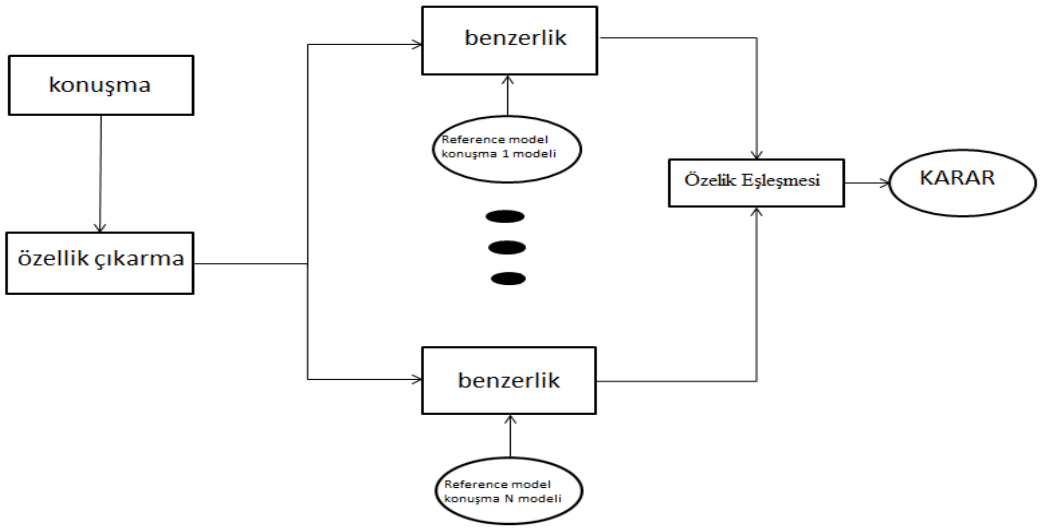
Şekil 1.3 Konuşmacı tanıma sınıflandırma systemi

1.2.1 **Tanımlama ve Doğrulama**

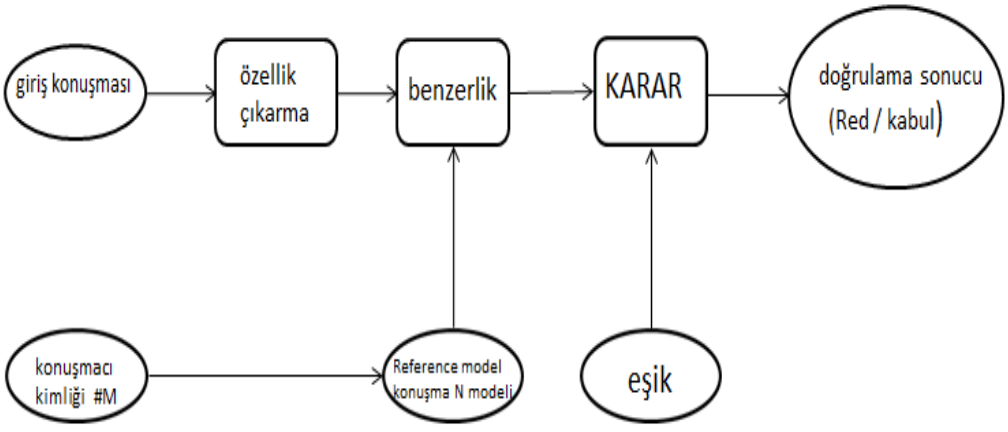
Otomatik konuşmacı tanımlaması ve doğrulaması, genellikle fiziksel konumlara veya bilgisayar sistemlerine yetkisiz erişimi önlemek için en doğal ve ekonomik yöntemler olarak kabul edilir. Detaylandırılacak olursa;

•Konuşmacı Tanımlama: Hangi kayıtlı konuşmacının belirli bir ifade sağladığını belirleme sürecidir.

•Konuşmacı Doğrulama: Bir konuşmacının kimlik talebini kabul etme veya reddetme sürecidir. Şekil 1.4(a) ve Şekil 1.4(b)'de, konuşmacı tanımlama ve doğrulama sistemleri arasındaki temel farklılıkları göstermektedir.

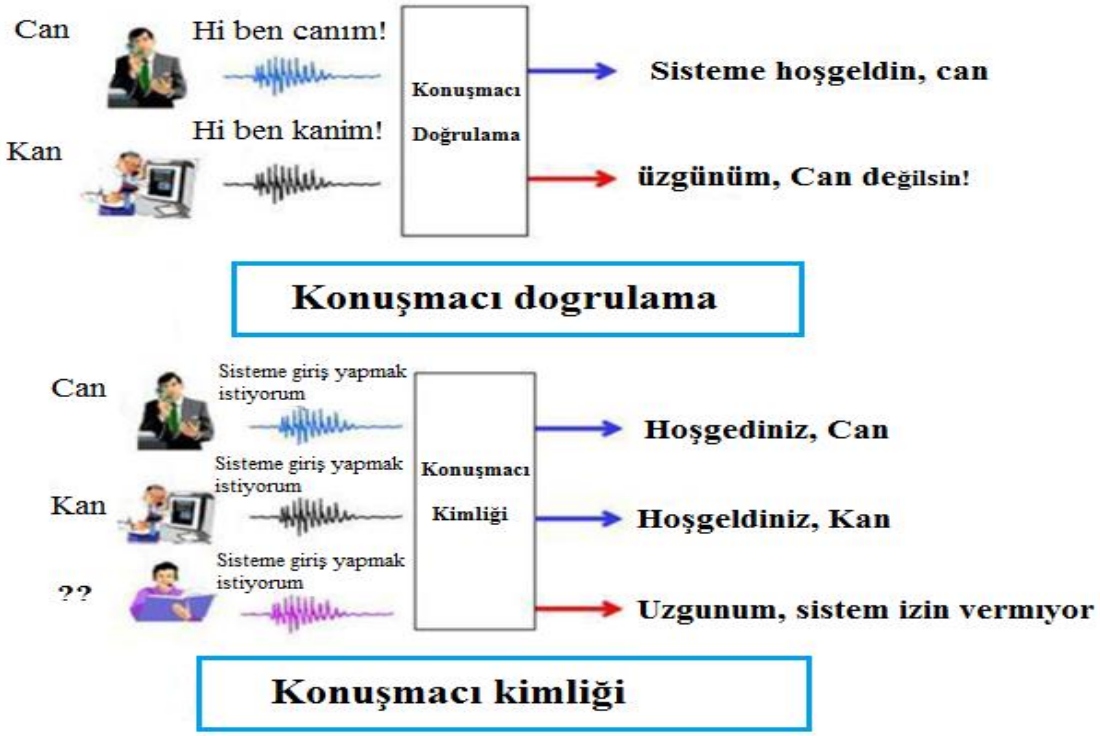


(a) Konuşmacı tanımlama



(b) Konuşmacı doğrulama

Şekil 1.4 Tanımlama ve doğrulama sistemlerinin blok diyagramları



Şekil 1.5 Konuşmacı tanımlama (kimliklendirme) ve doğrulama sistemlerinin pratik örnekleri

Şekil 1.4, her iki sürecin teorik blok diyagramlarını verirken, Şekil 1.5, sistemlerin pratik bir şekilde uygulanmasını göstermektedir.

1.2.2 Metin Bağımlı ve Metin-Bağımsız Konuşmacı Belirleme

Bu kategori, tanımlama işlemi sırasında konuşmacı tarafından verilen metne dayanmaktadır. Şöyle ki;

•Metin Bağımlı: Bu durumda, test ifadesi eğitim aşamasında kullanılan metinle aynıdır. Test konuşmacının sistem hakkında önceden bilgisi bulunmaktadır.

•Metin Bağımsız: Bu durumda, test konuşmacının eğitim aşamasının içeriği hakkında önceden bilgisi yoktur ve herhangi bir şey konuşabilir.

1.2.3 Açık Set Ve Kapalı Set

Bu sınıflandırma kategorisi, bir sistemde mevcut eğitimli konuşmacı setine dayanmaktadır.

•Açık Set: Açık bir sette herhangi bir sayıda eğitimli konuşmacı bulunabilir. Açık bir konuşmacı setimiz var ve konuşmacı sayısı bir taneden daha büyük olabilir.

•Kapalı Set: Kapalı bir sistemde, sisteme kayıtlı yalnızca belirli (sabit) bir kullanıcı vardır.

Bu tez çalışmasında kapalı bir eğitimli konuşmacı seti kullanılmıştır.

1.3 **Konuşma Tanıma**

Konuşma tanıma, mikrofon yardımıyla insan tarafından konuşulan kelimeleri yakalamak ve tanımak için sağlanan bir teknolojidir. Bu yakalanan kelimeler daha sonra konuşma tanıyıcı tarafından tanınır ve sonuçta sistem tanınmış sözcükleri çıkarır. Konuşma tanıma süreci farklı adımlardan oluşur. Bu adımlar aşağıdaki bölümlerde açıklanmaktadır. Konuşma tanıma sürecinde ideal durumda, bir konuşma tanıma operatörü, bir insan tarafından söylenen tüm kelimeleri tanır ancak, pratik olarak bir konuşma tanıma operatörünün performansı, konuşmanın içerik sayısına bağlıdır. Dijital sözlükler, birden çok kullanıcı ve gürültülü ortamda bile günümüzde başarılı sayılabilen sistemlere örnek verilebilir.

1.3.1 Konuşma Tanıma Türleri

Konuşma tanıma sistemleri, sözcükleri ve sahip oldukları sözcüklerin listesini anlama yeteneklerine göre sınıflara ayrılabilir. Aşağıda birkaç konuşma tanıma sınıfı anlatılmıştır:

•İzole Konuşma

Yalıtılmış sözcükler genellikle iki sözcük arasında bir duraklamayı içerir, sadece tek bir kelimeyi kabul ettiği anlamına gelmez, bunun yerine bir seferde bir ifade gerektirir.

•Bağlantılı Konuşma

Bağlı kelimeler veya bağlı konuşma, yalıtılmış konuşmaya benzer, ancak aralarındaki minimum duraklama ile ayrı ifadelere izin verir.

•Sürekli Konuşma

Sürekli konuşma, kullanıcının neredeyse doğal konuşmasına izin verir, aynı zamanda bilgisayar diktesi denir.

•Doğal Konuşma

Temel düzeyde, doğal bir ses olan ve prova edilmeyen bir konuşma olarak düşünülebilir. Kendiliğinden konuşma yeteneğine sahip bir otomatik konuşma sistemi (Automatic Speech Recognition, ASR) sistemi, birlikte çalışılan sözcükler, "ums", "ahs" ve hatta hafif duraksamalar gibi çeşitli doğal konuşma özelliklerini kullanabilmelidir.

1.3.2 Karşılaşılabilecek Güçlükler

Konuşma ve Konuşmacı tanımada, temel işlem konuşma sinyalini kaydetmektir. Böylece ilk işlem dijital sinyal işleme yöntemleri uygulanabilmesi için konuşma sinyali dijital forma dönüştürülür. Ancak bu dönüşüm sırasında çevresel gürültülerin uygun süzgeçler vasıtasıyla süzülmesi gerekmektedir. Ayrıca konuşmacının vurgusu ve dili, belirlenmesi gereken zorlu bir diğer noktadır. Bunların dışında, literatürden aşağıdaki zorluklar eklenebilir:

•Boyut: Kelime bilgisinde kelime türlerinin sayısı.

•Konuşmacı: Belirli bir konuşmacı veya konuşmacıdan bağımsız olarak mı ayarlandı? Konuşmacı özelliklerine ve aksanına uyum var mıdır?

•Akustik Çevre: Gürültü, rakip konuşmacı ler, kanal koşulları (mikrofon, telefon hattı, oda akustiği).

•Stil: Sürekli olarak konuşulmakta mı yoksa konuşma izole mi? Planlanmış monolog veya kendiliğinden konuşma mıdır?

1.4 **Araştırma Amaçları**

Bu proje, birçok konuşmacı arasında belirli bir konuşmacının tanımlanmasını beklemektedir. İlk önce Konuşmacı ve Konuşma Tanıma'yı sorunsuz bir şekilde gerçekleştirmek için, akan hedefler Matlab dijital sinyal işleme toolbox ve programlanabilir fonksiyonlar yardımı ile gerçekleştirilecektir:

•Konuşmacı, konuşma sinyalini, DOT WAV veya DOT MP3 gibi dijital formatta kaydederek Matlab iş istasyonuna girebilir. Tek kanal veya çift kanal gibi kanal tipi, en iyi sinyal sunumu için insan konuşmasının örnekleme oranını da seçebilir.

•Sistemde bir referans sinyali saklanabilir, bu sinyal sonraki aşamalarda karşılaştırmanın gerçekleştirilmesi için bir referans olarak seçilmiştir.

•Hızlı fourier transformasyon (Fast Fourier Transform, FFT) uygulayarak sinyalin frekans sunumunu değerlendirir. Aynı durum, mikrofonun durumu ve çevresiyle kaplı gürültü gibi, aynı sesle ilişkili konuşma sinyalini ve diğer rahatsızlıkları oluşturan bilgiyi üretecektir.

•Sinyalin çıkardığı gürültü bileşenini çıkarmak için dijital filtreler kullanılmalıdır. İki mikrofon kullanılacaktır; ses sinyali diğer mikrofon tarafından kaydedilmelidir;

•Sinyal yerleşiminden sonra; konuşmacının karakterini değerlendirmek için referans veritabanıyla karşılaştırılacaktır.

•Sonuç ekranında, tanıma sonuçları üzerine isim ve atama gibi konuşmacı bilgilerini, yani (SİZİN: RUYA, 26 YAŞ, YÖNETMEN) ya da (SİSTEME ERİŞİM İÇİN YETKİLİ DEĞİLSİNİZ) görülecektir.

1.5 **Tez Organizasyonu**

Bu tez raporu, konuşmacı ve konuşma tanıma algoritmalarının tüm sürecinin gösterileceği altı bölümden oluşmaktadır. Giriş bölümünde, konuşmacı ve konuşma tanımanın anlamları, farklılıkları ve zorlukları konusunda genel bilgiler içermektedir. İkinci bölümde “Literatür Araştırması” literatürde yer alan konuşma ve konuşmacı tanıma ile ilgili makalelerin bir bileşimi, her araştırmanın dezavantajları ve avantajları

ile ayrıntılı olarak gözden geçirilecektir. Üçüncü “Metodoloji” bölümünde önerilen sistem önerilen yöntemler ve materyaller hakkında ayrıntılı bilgi sunulacaktır. Dördüncü “Pratik Model” bölümünde Matlab ve Simulink'teki sistem uygulaması temel teknolojilerle gösterilecektir. Beşinci, “Sonuçlar” bölümünde, bu proje tarafından elde edilen sonuçlar raporlanacak ve yapılabilecek gelecek çalışmalarla ilgili önerilerde bulunulacaktır. Son olarak, bölüm altı “Referanslar”, bu tezde kullanılan referansları ve araştırma katkılarını içerecektir.

BÖLÜM 2

LİTERATÜR TARAMASI

2.1 Literatürdeki Çalışmalar

R Sujirat Attawibulkul, Boonserm Kaewkamnerdpong, Yoshikazu Miyanaga 2017 yılında MFCC tabanlı konuşma tanıma sistemindeki gürültülü konuşma eğitiminin robot destekli otizm tedavisine yönelik gürültü bastırma potansiyeli üzerine çalışmışlardır. Japonca konuşma veri tabanında temiz konuşma eğitimi için, MFCC ile birlikte gürültü azaltma tekniklerinin de kullanılması ile MFCC'den daha yüksek doğruluk ile iyileşmeyi sağlayabildiklerini göstermişlerdir. En sonunda deneysel sonuçlarına göre, gürültü söndürmeli MFCC'nin, temiz konuşma ve gürültülü konuşma ile test edildiğinde MFCC'den daha iyi performans sağladığını göstermişlerdir. Attawibulkul ve arkadaşları, gürültülü konuşma denemelerinde, çevrede bulunan gürültü türü ile performansın daha da artırılabildiğini, yani test ortamında bulunan gürültü ile gürültülü konuşma eğitiminin kullanılması, konuşma tanıma performansının, temiz konuşma eğitiminden daha iyi bir performans göstermesini sağlayabildiğini belirtmişlerdir. MFCC'nin hesaplama süresi gürültü bastırma ile MFCC'den daha düşüktür, ancak bu düşüşün oranı çok büyük değildir. Robot destekli otizm terapisinde sosyal etkileşimi ve iletişimi bütünleştirmeye yönelik olarak, robot gürültülü ortamlarda kullanılabilecek konuşma tanıma yeteneğine sahip olmalıdır[4].

E. Ambikairajah, M.J. Carey, G. Tattersall, 1980 yıllarında yaplıkları çalışma şudur, konuşma işleminde konuşma pitch periyodunun tahmininin önemli olduğunu dile

getirdiler. Gerçek zaman aralığı tespiti, sadece programlanabilir bir işlemci üzerinde veya özel amaçlı bir donanımda uygulanmaya uygun verimli bir algoritmanın seçilmesiyle mümkündür. Sesli konuşmanın pitch periyodunu tespit etmek için periodogram algoritmasının (p.a.) kullanılması önerilmektedir. Bu algoritma aşağıdaki nedenlerden dolayı caziptir: (a) çarpma işlemi yoktur; (b) 16-bit bir bilgisayarda (ör., mikroişlemci) uygulandığında, hesaplama mikroişlemcinin dinamik aralığını aşmadan tamsayı aritmetiğinde yapılabilir; (c) pitch periyodunu makul doğrulukla tahmin etmek için basit bir tekniktir. Periodogram algoritmasını kullanarak konuşma sinyalleri ve sinüzoitlerin analizinin sonuçları sunulmakta ve seslendirilmiş konuşmanın pitch periyodunu tahmin etmenin alternatif bir yöntemi olan ortalama büyüklük farkı fonksiyonu ile karşılaştırmalar yapılmaktadır[5].

Zunjing Wu, Zhigang Cao 2005 yılında çalışmalarında, Standart MFCC analizindeki logaritmik dönüşümün, gürültülü hassasiyetin iyileştirilmesi için kombine bir fonksiyonla değiştirilmesini önermişlerdir. Mel-frequency cepstral katsayısına (MFCC) dayanan konuşmacı tanıma sistemlerinde araştırma yapmışlar ve temiz konuşma için yüksek tanıma oranı sağlayabilmişlerdir, ancak performansın gürültülü ortamlarda önemli ölçüde azaldığını vurgulamışlardır. Gürültüyü daha da bastırmak için spektral çıkarma ve medyan filtresi gibi konuşma geliştirme yöntemleriyle birlikte özellik çıkarma işlemini önermişlerdir. SS (spectral subtraction) ve MF(medianfilter) ile parçalı bir dönüşüm fonksiyonunu birleştiren geliştirilmiş bir MFCC tabanlı özellik tanımlama yöntemi geliştirmişlerdir. Bu yöntemin kullanılmasının nedeni, eğitim ve test arasındaki olumsuzluk nedeniyle tanıma sistemlerinin performansını önemli ölçüde azaltma eğiliminde olmasıdır. TIMIT veritabanı kullanılarak yapılan test sonuçları, temel sistemdeki tanımlama doğruluğunun gürültülü ortamlarda hızla düştüğünü ve geliştirilmiş MFCC tabanlı özellik yönteminin SS yönteminden daha iyi bir şekilde sistem performansını geliştirdiğini göstermektedir[6].

Hairol Nizam Mohd. Shah ve arkadaşları 2014 yılında yaptıkları çalışma şudur, yönetici sesini tanımlamak için bir ses tanıma sistemi tasarlamışlardır. Ses tanımayı kodlamak için MATLAB yazılımını kullanarak yönetici sesi doğrulanabilir. Yönetici, eğitim oturumunda bir kez eğitmek ve saklamak için şifreyi bir kez kullanır. Test oturumunda

kullanıcılar, bir eşleşme olup olmadığını anlamak için şifreyi tekrar girerler. MATLAB simülasyonu kullanılarak, çıktı ya kullanıcının tanınmasını ya da reddedilmesini sağlayabilir. Sistemin test edilmesi sonucunda, belirli kullanıcının sesini ve reddedilen diğer kullanıcıların sesini başarılı bir şekilde tanır [7].

M. S. Likitha, S. R. R. Gupta, K. Hasitha and A. U. Raju 2017 yılında , konuşmanın çeşitli bilgilerden oluşan karmaşık bir sinyal olduğunu belirttiler. İnsan konuşmasından duyulan duygunun tanınması, önemli bir sinyal işleme problemidir. Hoparlör ses sinyalinden çıkarılan verileri kullanarak, konuşmacı duygularının tanındığını fark ettiler. Konuşma duygularının tanınmasında, konuşmacının duygusal durumu konuşmasından çıkarılır. Mel Frekanslı Cepstral Katsayısı (MFCC) tekniği, bir konuşmacının sesinden duyduğu duyguyu tanımak için kullanılır. Onlar tarafından tasarlanan sistem mutlu, üzgün ve öfke duyguları için onaylanmış ve verimlilik yaklaşık % 80 olarak bulunmuştur[8].

M. Sadeghi and H. Marvi 2017 yılında, ses tanıma uygulaması için tanıma doğruluğunu azaltmadan en uygun Mel-Frekans Kesişim Katsayıları (MFCC) özelliklerini elde etmek için çalıştılar. İki olasılık kullanabilen bir algoritma önerdiler, bir sınıflandırıcı olarak nöral ağ algoritmaları ve bir optimizasyon algoritması olarak diferansiyel evrim optimizasyon algoritması sunmuşlardır. Bu algoritmayı MATLAB yazılımında uyguladıktan sonra, şimdiye kadar her bir çerçeve için en az 13 olan MFCC özelliklerinin sayısının, tanıma doğruluğu azaltılmadan, çerçeve başına 5'e düştüğü görülmüştür. Bu iki durumda ses tanıma sisteminin doğruluğu hala 87.5% 'tir, yani konuşma sinyallerini daha az MFCC özelliği ile modellenebileceğini göstermişlerdir[9].

A. Charisma, M. R. Hidayat and Y. B. Zainal 2017 yılında şunu yapmaya çalıştılar, endonezya dilinde üç kişiyi tanımlamak için sistem tasarladılar. Sistem başarı oranlarını karşılaştırarak test edilmiştir. Ses sistemini doğrulamada başarıyı büyük ölçüde etkilemek için toplam karesel hata olarak adlandırılan eşik değerini hesaplamayı önerdiler[10].

S. R. Suralkar ve arkadaşları 2013 yılında, kablosuz iletişim yoluyla konuşmacı tanıma özelliğini kullanan konuşma tanıma otomasyon sistemi üzerine çalışmışlardır. Bu

çalışma, Matlab kodunu kullanarak kablosuz iletişim ve konuşmacı tanıma kullanarak otomasyon sisteminin tasarımını vermektedir. Matlab'ın basit programlama arayüzü, projede konuşma analizi için ideal bir araçtır. Bu otomasyon sistemi, ev aletlerinin yanı sıra endüstride de kullanışlıdır. Bu çalışma, inşa edilen ve uygulanan bir kablosuz otomasyon sisteminin genel tasarımını tartışmaktadır. Konuşma tanıma, Matlab'ın veri tabanında saklanan konuşma komutlarının tanınması üzerine odaklanır ve hoparlörden gelen konuşmacı komutuyla eşleşir. Konuşmacının konuşmasını tanımlamak ve konuşmanın özelliklerini çıkarmak için Mel Frekansı Cepstral Katsayısı (MFCC) algoritması kullanılır. Göreceli olarak ucuz olan düşük güçlü RF ZigBee telsiz kablosuz iletişim modülleri kullanır. Bu otomasyon sistemi, “Işık”, “Fan” vb. Gibi konuşma komutlarını kullanarak bir evdeki veya ofisteki ışıkları, fanları ve diğer elektrikli cihazları kontrol etmeye yöneliktir. Ayrıca, güvenlik büyük bir sorun değilse, o zaman konuşmacı, cihazları konuşmacı tanımlaması olmadan kontrol etmek için kullanılır[11]..

Er. Abhishek Thakur ve arkadaşları 2011 yılında, Matlab kullanarak kontrol sistemi ve konuşmacı tanıma kodunu tasarlayan bir proje geliştirdiler. Matlab’ın basit programlama arayüzü, Konuşma analizi projeleri için ideal bir araçtır. Mevcut proje için genel Matlab programlamasında ve kontrol sisteminin tasarımında tecrübe kazanılmıştır. Temel bir tanıma algoritması Matlab'da bir kural tabanı ile sıralamak ve konuşma ifadesinin önceden tanımlanmış zaman çerçevesine dayanan en olası eşleşmeyi seçmek için yazılmıştır. İlk adım, fs = 8000 Hz olan komut sözcüklerini kaydetmek için zaman çerçevesini tanımlamaktır. Bir sonraki adım, “wavrecord” komutunu kullanarak anahtar kelime örneğini kaydetmek, 0.1 büyüklükte bir ses örneğinin üzerinde bir değer almak, farkı hesaplamak ve “wavwrite” komutunu kullanarak dosyayı saklamaktır. Anahtar kelime prosedürü için diğer örnekleri saklamak, önceki ile aynıdır. İkinci adımda, dosya okunur ve mevcut ses örneği için 0.1 büyüklüğün üzerinde değer alınır. Fark hesaplanır ve daha sonra önceden tanımlanmış zaman çerçevesiyle karşılaştırılan bir değişkende saklanır, eğer eşleşirse çıktı verilir. Ses anahtar sözcüğü kullanılarak uygulanan algoritma ve kontrol sistemi fan hızını, ısıtıcı sıcaklığını ve robot yönünü kontrol eder. Elde edilen deneysel sonuçlara dayanarak, önerilen algoritmanın gerçekten işlevsel olduğunu ve ev aletleri ve endüstriyel

robotların sesli anahtar kelime kontrolünde kullanılabileceğini göstermektedir. Bu çalışmanın ana katkısı, anahtar kelime tanıma ve kontrol sistemi fikrini sunmasıdır. Deneyler ayrıca yaklaşımın anahtar kelime tanıma için iyi olduğunu göstermektedir[13].

P. Yasotha, N. Rajasekaran ve M. Nathiya 2014 yilinda çalışmalarında ses örneğinden ses özelliklerinin çıkarılması için Mel Ölçekli Frekanslı Cepstral Katsayıları (MFCC) kullanılır ve ayrıca HMM, çıkarılan özelliklere dayanarak konuşmacıyı tanımak için kullanılır. Asıl amaç, konuşmacı ilk HMM parametrelerini kullanarak ayıklanan özelliği tanıması ve daha sonra orijinal ses sinyali ile karşılaştırılmasıdır. Simülasyon sonuçları, hesaplama zamanına göre konuşma tanımada bir iyileşme göstermektedir, Burada MATLAB çıkışı, UART kullanarak mikrodenetleyiciye aktarılabilir ve LCD modülünden görüntülenebilir[13].

Abhishek Thakur ve arkadaşları 2014 yılında, matlab programlama arabiriminin Hintçe Anahtar Kelime Tanıma için ideal bir araç olduğunu belirttiler. Özelliğin çıkarılması için, Hintçe Anahtar kelime veritabanı Matlab 7.5 kullanılarak dizayn edilmiştir. Veritabanı sekiz anahtar sözcükten oluşmaktadır. Her anahtar kelime veritabanında on konuşmacı, sekiz erkek konuşmacı ve iki kadın konuşmacı tarafından sekiz komut için toplam 80 örnekte saklanmıştır. MFCC katsayıları ve Dinamik Zaman Çözümü (Dynamic Time Warping (DTW)) biçiminde çıkarılan konuşma sinyalinin özellikleri, eşleştirme teknikleri olarak kullanılmıştır. Bu yaklaşım, özellikleri ayıklamak için bitiş noktası algılama, MFCC kullanarak ifade tespit etmek için tekniği sunar ve test modellerini karşılaştırmak için DTW kullanır. Tanıma sonuçları temiz ve gürültülü test verileri için test edilmiştir. Sistem, temiz veriler için ortalama doğruluk oranı 97.50 % iken, gürültülü veriler için 91,25 % veya daha fazla kabul edilebilir olduğu söylenebilir, çünkü çoğu insan, sisteme on kez veya daha az bir kez bir komutu tekrar etmekten sakınmayacaktır. Sistem, giriş konuşmasını kabul etmek için az miktarda adanmış bellek ve analogdan dijitale dönüştürücü olan ortak mikrodenetleyicilerden biri kullanılarak gerçekleştirilebilir. Sistem çok çeşitli tüketici elektroniğine dahil olmak için hızlı, küçük ve uygun maliyetli olacaktır. Bu nedenle, bu araştırmanın amacı, küçük bir ev aletine sığacak kadar küçük ve gerçek zamanlı olarak çalıştırılabilen küçük, hoparlör bağımlı, izole bir kelime, sınırlı kelime dağarcığı tanıma sistemi geliştirmektir[14].

E. S. Wahyuni 2017 yılında, çalışmasında Hijaiyah denilen Arapça dilini doğrulamak için Mel-Frekanslı Cepstral Katsayılar (MFCC) tabanlı öznitelik ve Yapay Sinir Ağı (YSA) sınıflandırma yöntemini kullanmıştır. Üç harfli (sa, sya, and tsa) bir sistem tasarlamış, ortalama doğruluk sonucu 92.42% 'dir. Tanıma doğruluğu her harf için (sa, sya ve tsa) sırasıyla 92.38% , 93.26% ve 91.63%' tür. Son olarak, konuşma tanımada katkıda bulunan özelliklerin üretilmesinde özellik çıkarma yeteneğinin ve öğrenme sürecindeki sınıflandırma yönteminin yeterliliğinin başarılı konuşma tanımasının iki anahtarı olduğunu belirtmiştir[15]. .

BÖLÜM 3

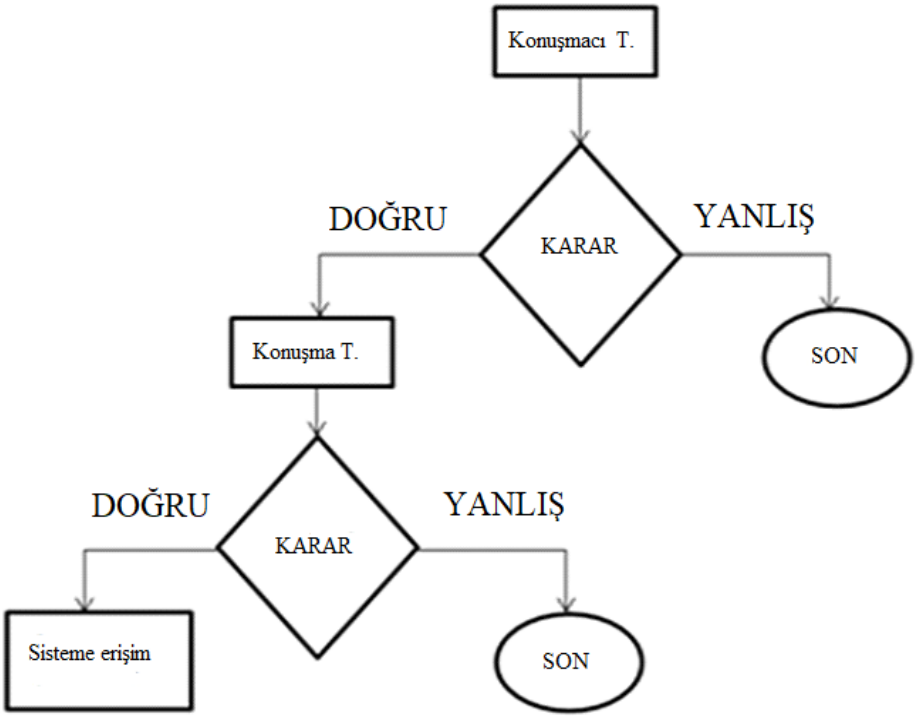
METODOLOJİ

3.1 Giriş

Bireylerin anahtar kelime tabanlı güvenlik sistemlerinde tanımlanmasına yönelik güvenli özelliklerinden ötürü, biyometrik teknolojinin günümüzde daha popüler olduğu kabul edilmektedir. En eski biyometrik tanımlama yöntemleri, parmak izi, göz ve yüz taraması vb. gibi fiziksel özellikleri içerir. Ancak davranış özelliklerini kullanan bir kişiyi tanımanın başka yolları da vardır. Bu ifadedeki en basit ve ucuz teknoloji, ses tanıma sistemleri ile gerçekleştirilebilir. Biyometrik ses tanıma teknolojisi, cep telefonları, ATM'ler ve kişisel bilgisayarlar için özellikle güvenli erişim için konuşma veya hoparlör tanıma üzerinde yoğunlaşmaktadır. Bu tezde bir kişisel bilgisayara güvenli bir şekilde erişmek için hem konuşma hem de konuşmacı tanıma uygulamaları sunulmaktadır. Önerilen tanıma algoritması MATLAB ortamı kullanılarak geliştirilmiştir. Bu sistem, önce bir kişinin kimliğini kendi ses örüntüsüyle doğrular, sonra önceden tanımlanmış kişisel anahtar kelimeyi kontrol eder. Önerilen sistem çoklu biyometrik bir sistem olarak düşünülebilir. Genellikle çoklu biyometrik sistemler birden fazla sensör gerektirir ama önerilen çoklu biyometrik sistemde tek sensör kullanılır. Konuşmacı tanıma, özellikle güvenli bir kişisel bilgisayara giriş yapmak için ve biyometrik tanımlama problemleri için kolay ve ucuz bir çözümdür. Konuşmacı tanıma sisteminde sadece bir mikrofonun olması yeterlidir. Dizüstü bilgisayarlar için standart bir cihaz olan basit bir mikrofon, bir biyometrik yakalama aygıtı gibi davranabilir ve ek donanım gereksinimini önler. Ancak, parmak izi ve iris tanıma gibi son teknoloji yöntemlere ilişkin göreceli

olarak düşük doğruluktan dolayı ses kullanarak biyometrik tanımlama tercih edilmemektedir [16]. Bu çalışmada, doğruluğun arttırılabilmesi için iki adımlı bir tanımlama planı uygulanmıştır. İlk adımda, kişiyi tanımlamak için metin bağımsız konuşmacı tanıma yöntemi uygulanmış; daha sonra konuşma tanıma modülü tarafından doğrulanmıştır. (Şekil 3.1).

Pitch periyod, konuşma tanıma problemi için makul doğrulukta kullanışlı ve basit bir özelliktir [5] ve mel frekanslı cepstral katsayılar (MFCC) konuşma ve konuşmacı tanıma için en çok kullanılan özelliktir. Bununla birlikte, MFCC gürültüye çok duyarlıdır [6] ve gürültü bastırma performansı artırabilir [4].



Şekil 3.1 Önerilen konuşmacı tanımlama yöntemi

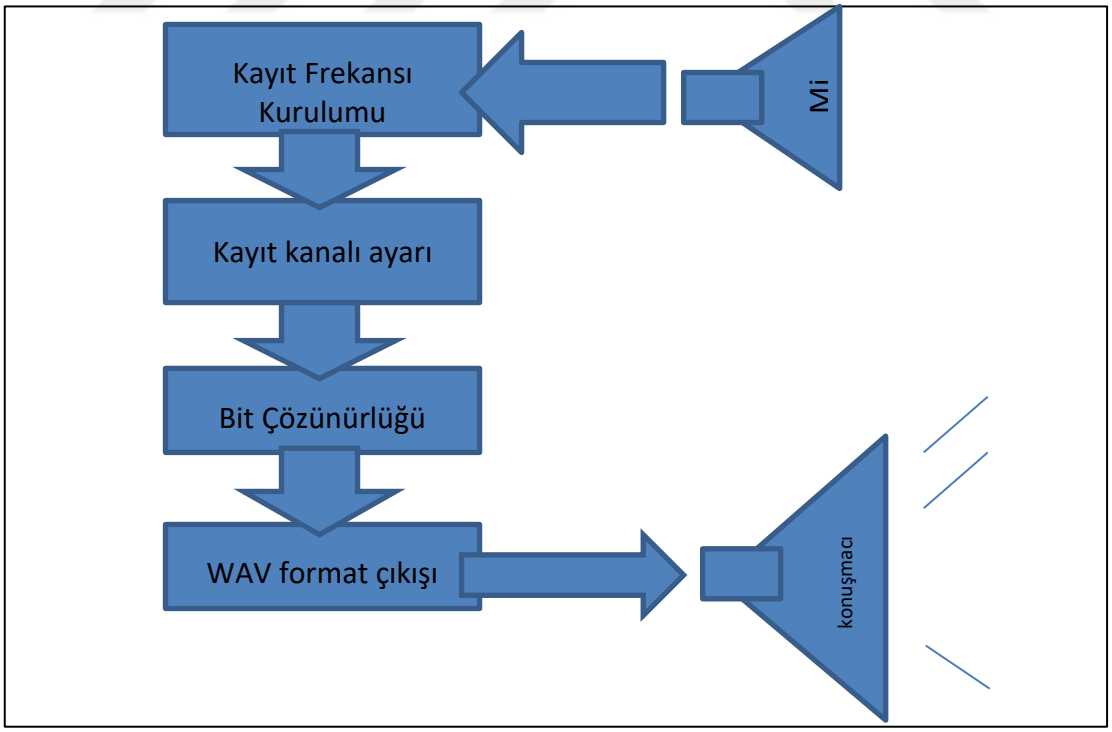
3.2 **Konuşmanın Önemi**

İnsan türü, tüm canlılar arasında ses tonları ve ses sinyalleri üretme yetenekleriyle ayırt edilir. İnsan için konuşma, iletişim kurmayı ve hayatta kalmayı sağladığı için önemli bir

özelliktir. Ses sinyallerini rutin yaşamında kullanan her bireyin kendi ses tonu vardır. Farklı insanlar arasındaki ses değişimleri, vokallerinde farklı frekanslara atfedilmektedir. Frekans, her insana verilen tek özelliktir ve ses tellerini değiştirmek için bir sonuç olarak üretilir. Gitar gibi kordonlu müzik teçhizatı, kordlarının farklı hareketleriyle farklı tonlar üretir. Genellikle gitaristler farklı mesafelerden basarak kordonların uzunluklarını değiştirdiğinde ortaya çıkar. Bununla birlikte, aynı kavram insan ses tellerine uygulanmaktadır, insan sadece seslerini kendi ses frekanslarına göre üretebilir.

3.3 **Konuşma Kayıt**

Kişinin konuşmasını kaydederek mikrofondan Matlab'a sinyal getirilir, bu işlemin yüksek kalitede yapılması şarttır. Böylece istenen kaliteye ulaşmak için birçok normun takip edilmesi gerekmektedir. Kayıt kanalları, tek kanallı veya çift kanallı (Steario veya Mono), kayıt sırasında belirtilen ses sinyalinin sıklığı önemli parametrelerdir. Ses sinyalinin kaydedilmesi için normalde 44100 Hz ve 16 bit çözünürlük kullanılır. Şekil 3.2 Matlab'a veri girişini göstermektedir.



Şekil 3.2 Matlab'da ses kayıt akışı.

3.4 **Ses Sinyali Özelliği**

Bu tür sinyallerin analiz edilmesi için konuşma sinyali analizinin temelleri anlaşılmalıdır.

3.4.1 Ses ve Sessizlik Tespiti

Tipik bir konuşma cümlesinin işareti iki ana bölümden oluşur: biri konuşma bilgisini taşır, diğeri de herhangi bir sözlü bilgi olmadan, konuşmalar arasındaki sessiz bölümlerini içerir, Konuşmanın sözel (bilgilendirici) kısmı iki kategoriye ayrılabilir:

•Sesli konuşma.

•Sessiz konuşma.

Sesli konuşma ağırlıklı olarak ünlü seslerden oluşur. Glottis üzerinden hava zorlayarak üretilir. Ses tellerinin gerginliğinin uygun şekilde ayarlanması, kabloların açılmasına ve kapanmasına ve neredeyse periyodik hava darbelerinin üretilmesine neden olur. Bu bakliyatlar vokal yolunu heyecanlandırır. Psikoakustik deneyleri, bu bölümün konuşmanın bilgisinin çoğunu taşıdığını gösterir ve böylece bir konuşmacının karakterize edilmesi için anahtarları tutar. Sessiz (gürültülü) konuşma bölümleri, vokal yolunun bir noktasında oluşan (genellikle ağız ucuna doğru) bir büzülme yoluyla havayı zorlayıp türbülans üreterek oluşur. Bunları birbirinden ayırt edebilmek, konuşma sinyali analizi için çok önemlidir.

3.4.2 Özellik Çıkarma

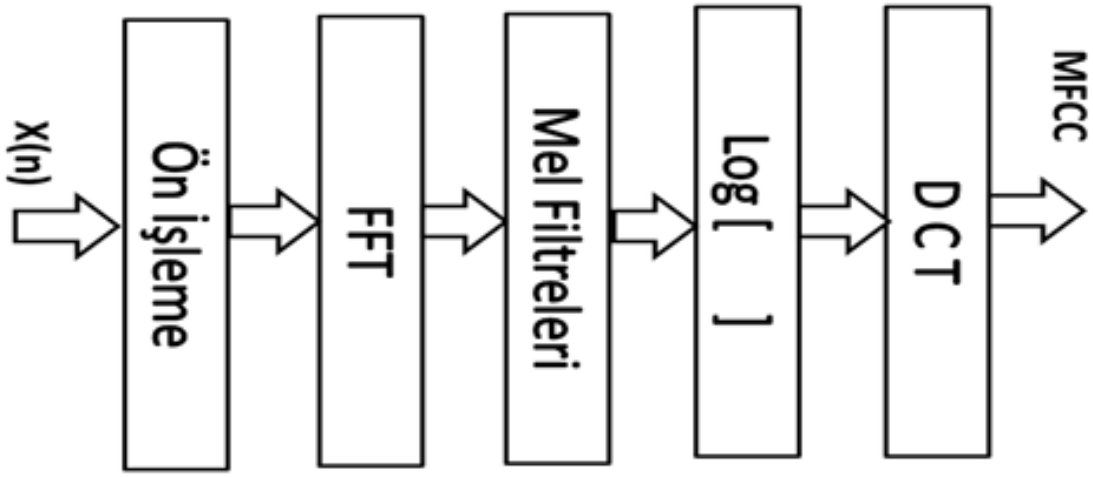
3.4.2.1 **Mel Ölçek Süzgeç Dizisi Enerjilerinin Kepstral Gösteriminin(MFCC)**

Pitch periyodu ve MFCC son zamanlarda konuşma tanıma problemlerinde başarıyla kullanılan iyi bilinen yöntemlerdir [5]. Cepstrum, karmaşık sinyaller üzerinde iki veya daha fazla konvolüsyonlu sinyalden oluşan bir homomorfik dönüşümdür. Konuşma üretim mekanizması, bir uyarma sinyali tarafından sürülen tüm kutup filtresinin bir zaman varyantı olarak modellenebilir. Zaman alanında, konuşma sinyali, uyarımı filtreden geçirerek elde edilir. Ancak hücresel alanda konvolüsyon operatörü toplanır

ve genellikle filtre, septral alanın alt kısmını işgal eder. Bu nedenle, ses bilgisini belirleyen filtre parametreleri, düşük geçişli bir cep telefonu filtresi kullanılarak kolayca belirlenebilir. Mel frekansı kelimesi mel ölçeği anlamına gelir. Mel skalası, oktavları tonlar arasında algısal olarak eşit bir mesafe olarak belirleyen algısal bir ton sistemidir. Literatürde, mel frekans dönüşüm şemaları için çeşitli doğrusallar vardır. Bu çalışmada, mel ölçeği aşağıdaki gibi hesaplanır.

$$\text{Mel} = 2595 * \log(1 + \text{f} / 700) \quad (3.1)$$

MFCC hesaplamanın hesaplama gereksinimleri oldukça yüksektir. Şekil 3.3: bu çalışmada kullanılan MFCC hesaplama sürecinin blok diyagramını göstermektedir.



Şekil 3.3 MFCC'nin hesaplanması

3.4.2.2 **Pitch Periyot Yöntemi**

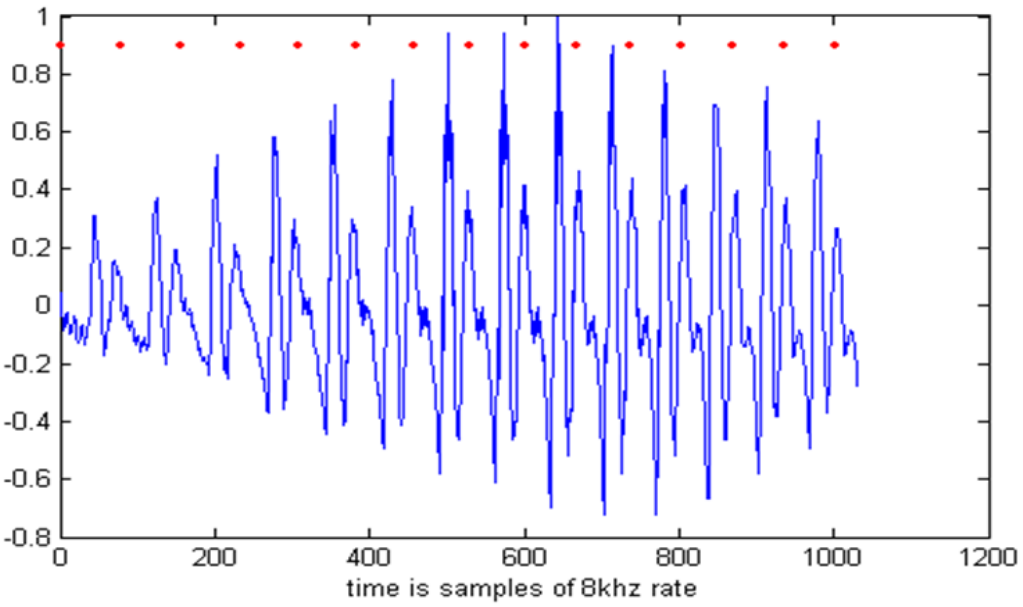
MFCC benzer şekilde, peitch periyodu insanlar için ayırt edici bir özelliktir ve biyometrik tanımlama için kullanılabilir. Konuşmanın dile getirilen kısmı yarı periyodik olarak kabul edilir. Bu periyodik sinyalin kök frekansı, pitch periyodu olarak adlandırılır. Pitch periyodu tespiti için iyi bilinen bir yöntem aşağıdaki gibidir. Belirli bir ardışık iki

konuşma parçası çifti için pitch döngüleri yüksek derecede ilişkilidir. Pitch algılama algoritması bu iki denklemle özetlenir:

$$r_t(\tau) = \sum_{j=1}^{w-\tau} x_j x_{j+r} \quad (3.2)$$

$$d_t(\tau) = \sum_{j=1}^{w-r} \left(x_j - x_{j+r}\right)^2 \quad (3.3)$$

d_t ve $r_t$pitch dedektörleri, W pencerenin uzunluğu ve bunun bir zaman alanı hesaplaması olduğunu belirtir. Şekil 3.4' de pitch işaretlerinin kırmızı ile gösterildiği bir vocal fonem açıklanmaktadı.



Şekil 3.4 Pitch işaretlerinin kırmızı ile gösterildiği bir vokal fonem

3.5 **Çapraz Korelasyon (CC)**

Aynı kelime için kaydedilen konuşma sinyallerinin aynı olduğunu varsayarsak, iki kaydedilmiş konuşma sinyalinin spektrumları da benzerdir. İki benzer spektrumun çapraz korelasyonunu yaparken ve çapraz korelasyonu çizerken, çapraz korelasyon çapraz korelasyon tanımına göre simetrik olmalıdır. Konuşma tanıma karşılaştırması için kaydedilen iki frekans spektrumunun çapraz korelasyonunu hesapladıktan sonra, çapraz korelasyonun maksimum değerinin konumunu bulmak ve maksimum değer pozisyonuna kalan değerleri eksiltmek için değerleri maksimum değer konumuna

doğru kullanmalı, daha sonra, bu farkın mutlak değerini almalı ve bu mutlak değerin ortalama kare hatasını bulmalıyız. İki sinyalin çapraz korelasyon simetrisi, her iki sinyalin eşleşme seviyesini gösterir. Hatta, daha fazla simetrik olan çapraz korelasyon, daha küçük olan ortalama kare hatasının değeridir. Eğitilmiş kelimelerin ve hedef kelimelerin ortalama kare hatalarını karşılaştırarak sistem, farklı gecikmelerdeki çapraz korelasyon farklılıklarının minimum ortalama kare hatasına dayanarak test sinyali ile daha iyi eşleşen eğitim kelimesine karar verir.

3.6 **Otokorelasyon (AC)**

Otokorelasyon, farklı sinyaller ile korelasyon yerine, sinyal ve kendisi için çapraz korelasyon hesaplanması olarak ifade edilebilir. Otokorelasyon algoritması, sinyalin kendi kendine nasıl korunduğunu ölçmeyi amaçlamaktadır. Böylece, otokorelasyonlar arasındaki minimum farkı bulmak için eğitim sinyali otokorelasyonları karşılaştırılmıştır. Sistem, otokorelasyon algoritmasını kullanarak pitch periyodu özelliğini hesaplayacaktır.

3.7 **Sinyal Enerjisi Ve Sinyal Gücü**

Sinyal enerjisi ve sinyal gücü terimleri, bir sinyali karakterize etmek için kullanılır. Onlar aslında enerji ve güç ölçüleri değildir. Sinyal enerjisi ve gücü, karmaşık değerleri alan sinyaller dahil olmak üzere, herhangi bir sinyale (x (t)) işaret eder.

χ (t) sinyalinin E sinyal enerjisi :

$$E = \int_{-\infty}^{\infty} |x(t)|^2 . dt \tag{3.4}$$

P sinyal gücü :

$$P = \lim_{T\to\infty} \frac{1}{2T} \int_{-T}^{T} |x(t)|^2 \, . dt \tag{3.5}$$

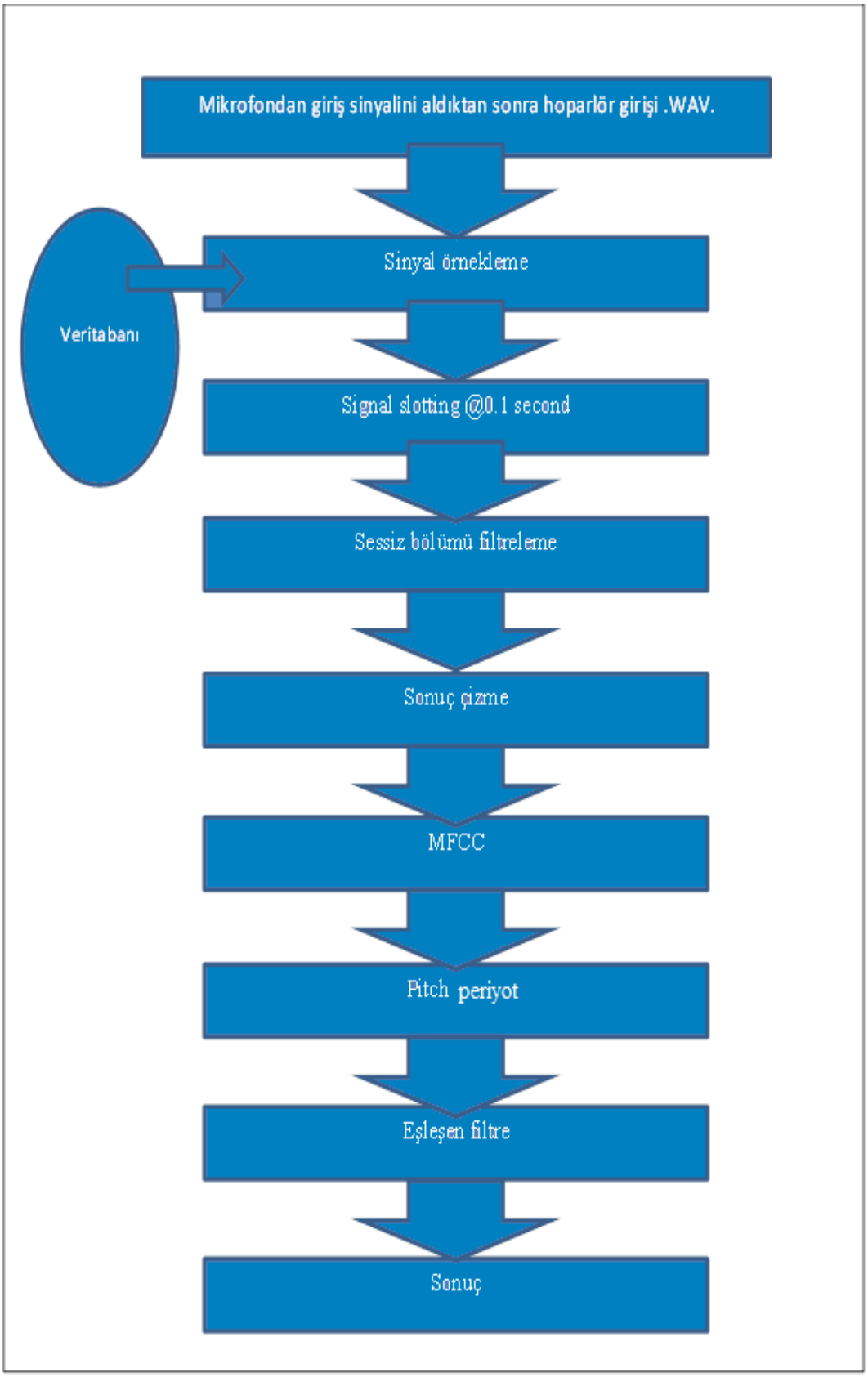
$0 < E < \infty$ ise, o zaman sinyal x (t) bir enerji sinyali olarak adlandırılır.

Ancak, bu durumu karşılayamayan sinyaller vardır. Bu sinyaller enerji değil de güç sinyali olarak kabul edilir ve bu sinyal güç sinyali olarak adlandırılır. Bir enerji

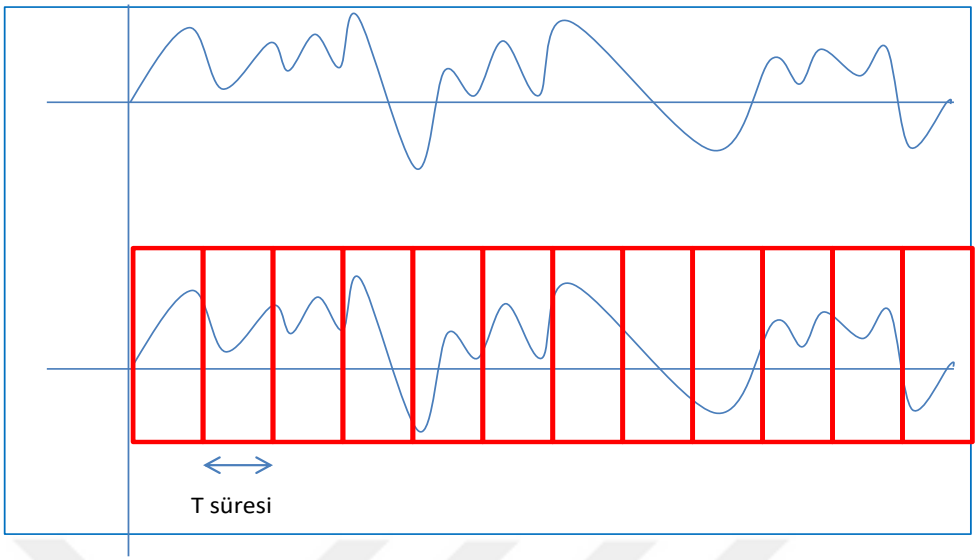
sinyalinin gücünün sıfır $0 < P < \infty$ ($P = 0$) olduğu ve bir güç sinyali için enerjinin sonsuz olduğu unutulmamalıdır.Bazı sinyaller ne enerji ne de güç sinyalleridir.

3.8 **Algoritma Yapısı**

Aşağıdaki şekilde gösterildiği gibi, konuşmacı, mikrofon kullanarak sesini Matlab'a kaydedebilir ve bundan sonra sistem analog sinyali dijital formata dönüştürecektir. Örnekleme, sinyalin dijital forma dönüştürülmesinden önce sistemde meydana gelen ilk süreçtir. Dolayısıyla konuşma analog sinyali frekansa göre yuvalar hâlinde oluşacak ve böylece dijital sinyal işleme uygulanabilecektir. Şekil 3.5 konuşma tanıma sisteminin tüm sürecini tasvir etmektedir. Buradaki sinyal, şekil 3.6'deki benzer genişlikteki zaman dilimlerine bölünür.

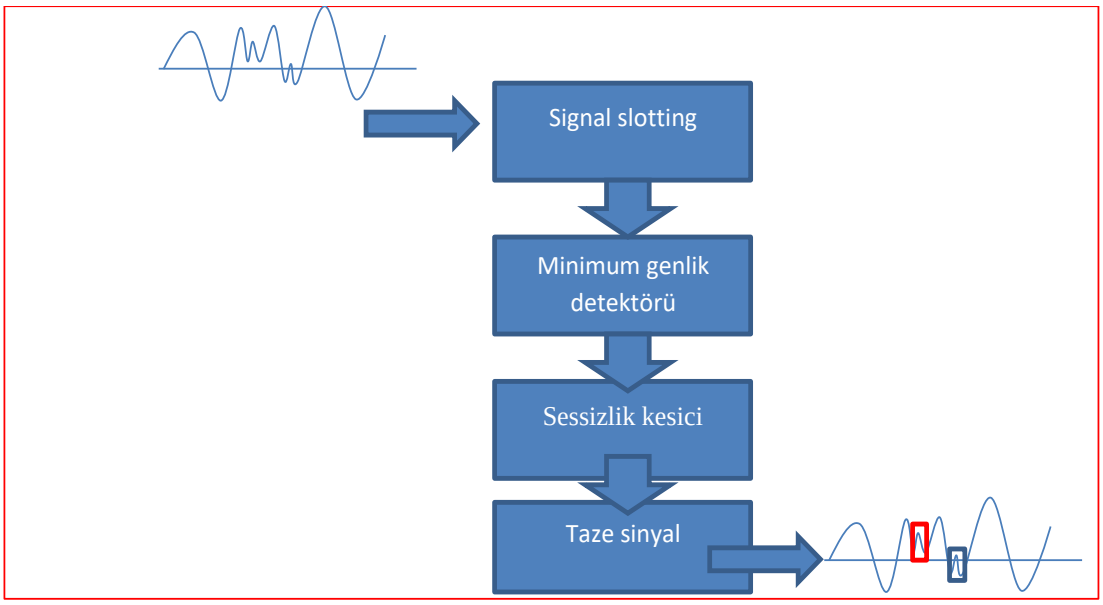


Şekil 3.5 SRS algoritması.



Şekil 3.6 Ses sinyali yerleştirme.

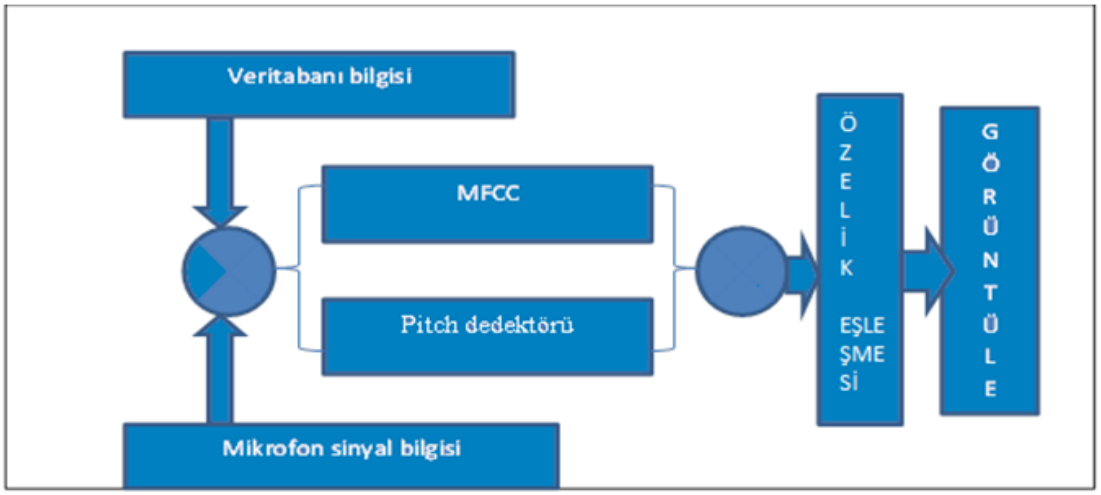
Bu yuvalarda, her bir yuvada sinyal gücü amplitüdü ölçülür ve minimum sinyal seviyesi belirlenir. Bu seviye, sinyal işleminin bir sonraki aşamasına referans girişi olarak kullanılır. Bu işlem esnasında sinyal parçalarının sesli ve sessiz kısımları ayırt edilebilir. Sinyalden çıkarılacak bu sürecin sonunda dile getirilen konuşmanın sesli kısmı kalmaya devam edecektir. Şekil 3.7 sessizlik giderme işlemini göstermektedir.



Şekil 3.7 Sessizlik kaldırma algoritması.

Şekilde görüldüğü gibi, programa giren sinyal gürültü ve sessizlik iptali için kullanılır. Şekilde olduğu gibi, tüm sinyal noktalarının karşılaştırılmasından sonra her slotta en düşük amplitüd algılanır ve minimum sayida nokta tespit edilir ve bir sessizlik olarak kabul edilir. Bu sessizlik, tespit edilen seviyeye (sinyal örnekleri arasındaki en düşük genlik) bakarak bütün yuvalardan iptal edilir. Konuşmacı tanımasında, sadece MFCC hesaplanır. Bu bölümde pitch periyodunu belirlememiz gerekmiyor ama konuşma tanımada, her ikisinin de pitch periyodu ve MFCC hesaplaması gerekir. Pitch frekansı, en iyi sinyal gücünün meydana geldiği minimum frekanstır. Bu frekanstaki her ses sinyaline özgüdür ve konuşmacı tanımlama amacı için kullanılır.

Yukarıdaki işlemde, her iki sinyal de veritabanındaki sinyallere ve kaydedilen test sinyaline uygulanır. Böylece, eşleşen filtreyle karşılaştırmak için her iki sonuç da kullanılabilir. Filtre konfigürasyonları, veri tabanı sinyal istatistiklerine göre ayarlanır, daha sonra test sinyal, eşleşen filtreye ikinci giriş olarak girilir. Konuşmacı tanımada, her iki girişten gelen pitch frekansları ve MFCC'ler karşılaştırılır.Bu nedenle, veritabanından gelen minimum frekansın frekansı ve MFCC özelıkleri, mikrofonun (test sinyali) aynısına benzerse, sadece mesaj, ekranda görünecek ve "erişime izin verilir" mesajı gösterilecektir. Aksi halde pitch frekansları eşit değilse, program bunu "giriş reddedildi" şeklinde gösterecektir. Şekil 3.8, özellik çıkartma yöntemini ve konuşmacı tanımasında eşleşen işlemi göstermektedir.



Şekil 3.8 Özellik çıkartma yöntemini ve konuşmacı tanımasında eşleşen işlemi göstermektedir.

BÖLÜM 4

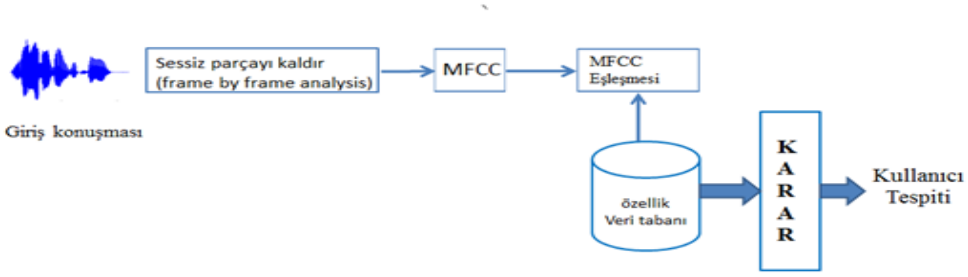
ÖNERİLEN YÖNTEM VE SONUÇ

4.1 Önerilen Yöntem

Bu çalışmada hem metne bağımlı hem de bağımsız konuşmacı tanıma yöntemlerini kullanan bir konuşmacı tanımlama sistemi önerilmiştir. Önerilen sistem ilk önce metinden bağımsız bir tanıma yöntemi kullanarak kullanıcıyı belirler. Kullanıcının kimliği belirlendikten sonra, sistem, kullanıcıdan önceden tanımlanmış bir şifre belirlemesini bekler. Hem parola hem de kullanıcı kimliği eşleşiyorsa, sistem kullanıcıyı kabul eder. Çalışmada, bu yöntemler birleştirilerek sistemin yanlış kabul oranını kısıtlamak amaçlanmaktadır.

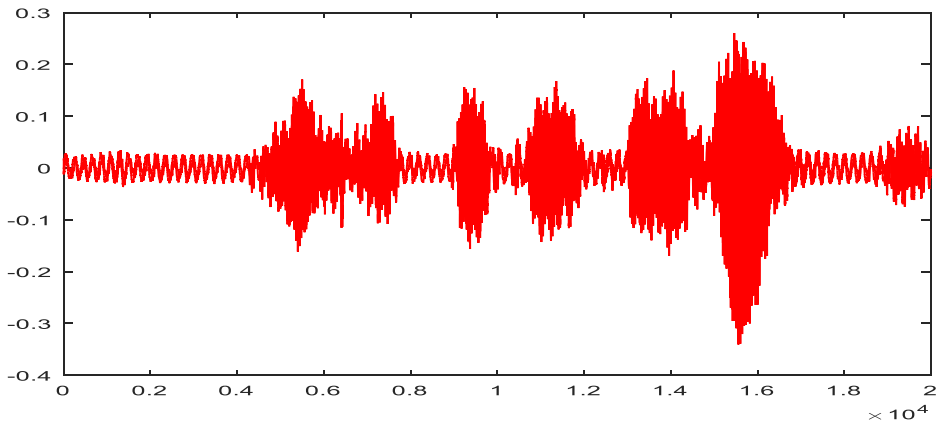
4.2 Konuşmacı Tanıma

Şekil 4'de konuşmacı tanıma sisteminin blok diyagramı görülmektedir. Kullanıcının sesi yakalandıktan sonra, konuşma verisi MFCC'yi hesaplamak için 100ms'lik uzunluklara bölünür. MFCC'nin gürültüye eğilimli olduğu bilinmektedir. Bu nedenle hızlı bir gürültü bastırma yaklaşımı bu noktada kullanılmaktadır. Sinyalin gürültü oranının (SNR) konuşmanın bölümlenmemiş kısımlarında daha yüksek olması nedeniyle, konuşma sinyalinin bu kısımları kısa zaman enerjisi kullanılarak tespit edilir ve kesilir. Bir ses bölütünün kısa süreli enerjisi önceden tanımlanmış bir eşikten düşükse, gürültü olarak kabul edilir ve bu bölüm kaldırılır. Şekil 4.2'de orijinal giriş sinyalini ve gürültüsünün bastırılmış versiyonunu göstermektedir.

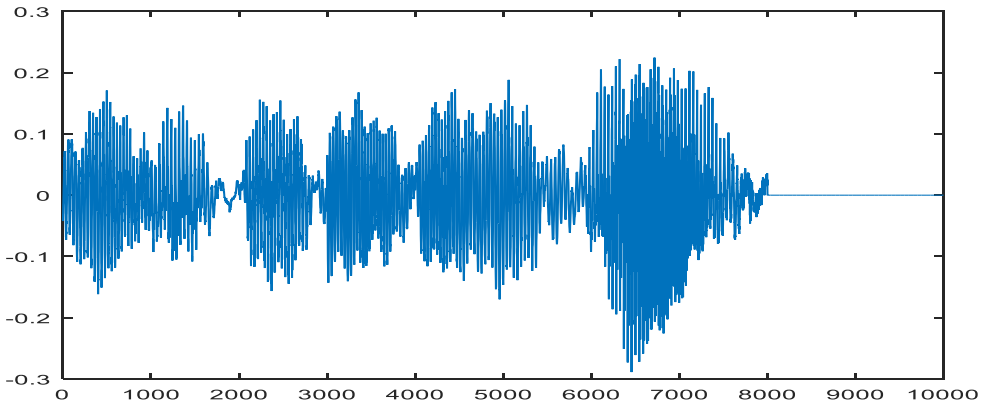


Şekil 4.1 Konuşmacı tanıma sisteminin blok diyagramı.

(a)



(b)



(c)

Şekil 4.2 Matlab'ta sinyalin girişi (a) orijinal giriş sinyali (b) ve gürültü bastırılmış versiyonu (c)

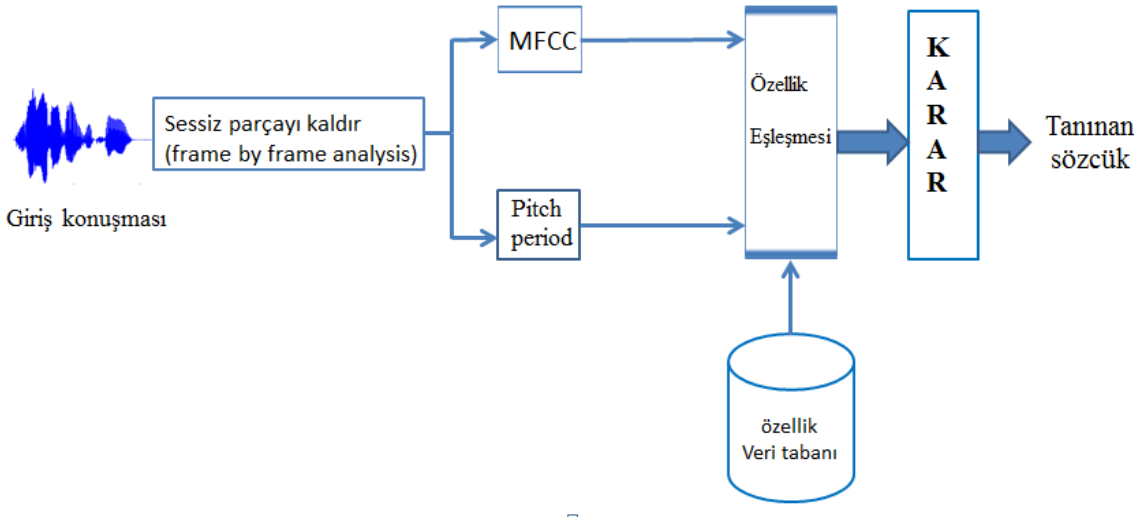
Gürültü bastırma işleminden sonra 20 tane MFCC'ler, denklem 3.1 kullanılarak hesaplanır. Konuşmacı tanıma aşamasında, her biri için 50 sözcük kullanarak sisteme iki kullanıcı kaydedilir. MFCC'ler biyometrik şablonlar olarak kullanılır. Bilinmeyen bir ifade sisteme uygulandığında, bu ifadenin MFCC'leri hesaplanır .Bunların ve kullanıcıların şablonlarından birinin arasındaki Euclidian mesafesi bir eşikten daha küçükse, sistem bilinmeyen ifadenin kimliğine karar verir. Tüm kayıtlar 10 KHz ve 16 bit çözünürlük örnekleme frekansına sahip bir stereo PC mikrofon tarafından yakalanır. İki farklı konuşmacı için simülasyon sonuçları Çizelge 4.1'de verilmiştir.

Çizelge 4.1 Konuşmacı Tanıma için Doğru Pozitif, Doğru Negatif, Yanlış Pozitif ve Yanlış Negatif Sonuçlar

DURUM	POZİTİF	NEGATİF
DOĞRU	**0.78**	**0.12**
YANLIŞ	**0.02**	**0.08**

4.3 Konuşma Tanıma

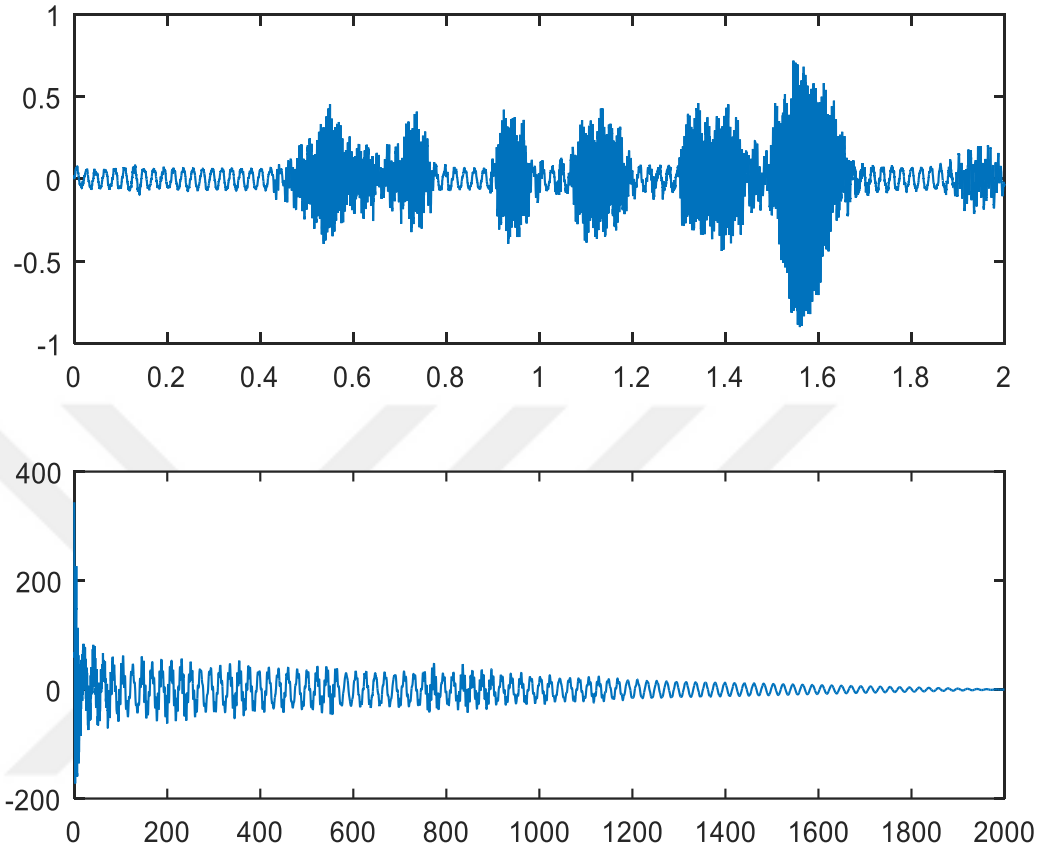
Sistem konuşma tanıma aşamasını belirledikten sonra, özellik olarak her iki pitch periyodu ve MFCC kullanılır. İlk aşamaya ek olarak, belirli bir sinyalin pitch periyodu denk. 2. ve 3. konuşma tanıma süreci, hem veritabanındaki hem de kaydedilen test sinyalindeki sinyale uygulanır, böylece her ikisi de eşleşen filtreyle karşılaştırmak için kullanılabilir. Filtre konfigürasyonları veritabanı sinyal istatistiklerine göre ayarlanır ve test işareti daha sonra eşleşen filtreye ikinci giriş olarak girilir. Sistem MATLAB programı tarafından uygulanır. Ses tanıma şeması, otokorelasyon ve gelişmiş sinyal işlemeden elde edilen minimum saha frekansına göre yapılır. Şekil 4.3 uygulanan konuşma tanıma sistemini göstermektedir.



Şekil 4.3: Konuşma tanıma sisteminin blok diyagramı

Bu aşamada MFCC konuşmacı tanıma gibi kullanılmıştır. Ayrıca, özellikleri karşılaştırmak için pitch periyodu algoritması da kullanılmıştır. Sinyalin pitch frekansları, sinyalde maksimum çıkış gücünün yerini gösterir. Bu frekanslar sonraki işlem için saklanır. Referans girişinden ve test girişinden gelen frekans frekansları, ses tanıma sisteminin (VRS) son aşamasında eşleştirilir. Eşleştirme işleminde, pitch frekansları arasındaki mesafe karşılaştırması kullanılır, bundan sonra deneme ve hata

yöntemiyle bir eşik 12 olarak belirlenir. Şekil 4.4'deki ilk rakam giriş sinyalini, ikincisi belirtilen giriş sinyalinin pitch periyotlarını göstermektedir.

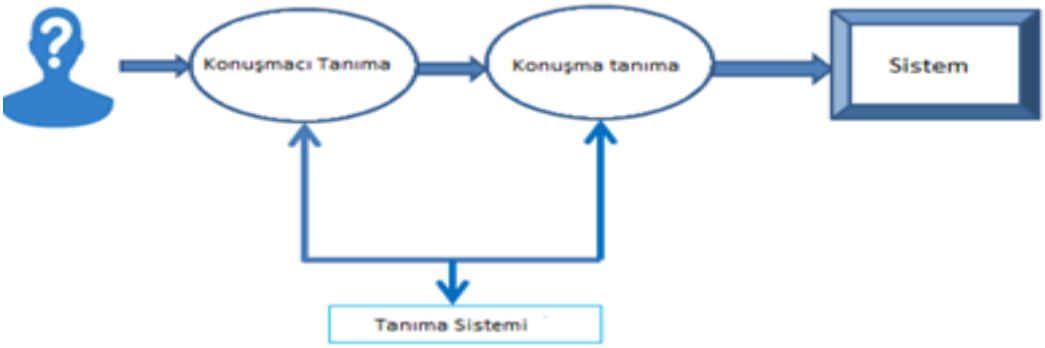


Şekil 4.4 Konuşma tanıma için giriş sinyali ve giriş pitch periyotları

Burada, konuşma tanıma için iki farklı özellik çıkarma yöntemi kullanılır. Her iki Pitch periyodu ve MFCC için sonuçlar Çizelge 4.2'de karşılaştırılmıştır. MFCC'nin ses tanımada daha başarılı olduğu görülmektedir. Her iki metodolojinin birleştirilmesi de bu probleme uygulanmaktadır ve bu birleşme tek tek kullanımdan daha iyi olduğu görülmektedir. Çizelge 4.3, basamaklı ses tanıma sistemi için sonuçları göstermektedir. Bu sistem, doğruluk, güvenlik ve penetrasyon zorluğu açısından daha iyi tanımlanmıştır. Bu çalışmada tanıma sistemi, şekil (4.5) 'de gösterildiği gibi iki bölümden oluşmaktadır.

Çizalge 4.2 Konuşma Tanıma için Doğru Pozitif, Doğru Negatif, Yanlış Pozitif ve Yanlış Negatif Sonuçlar

DURUM	POZİTİF	NEGATIVE
DOĞRU (Pitch periyodu)	**0.63**	**0.09**
YANLIŞ(Pitch periyodu)	**0.09**	**0.09**
DOĞRU (MFCC)	**0.66**	**0.19**
YANLIŞ (MFCC)	**0.08**	**0.07**
DOĞRU (Pitch periyodu + MFCC)	**0.77**	**0.052**
YANLIŞ (Pitch periyodu + MFCC)	**0.048**	**0.048**



Şekil (4.5) Tanıma sisteminin blok diyagramı

Çizalge 4.3 Basamaklı Ses Tanıma Sistemi için Doğru Pozitif, Doğru Negatif, Yanlış Pozitif ve Yanlış Negatif Sonuçlar

DURUM	POZİTİF	NEGATİF
DOĞRU	**0.862**	**0.13**
YANLIŞ	**0.03**	**0.05**

BÖLÜM 5

SONUÇ VE ÖNERİLER

Bu çalışmada biyometrik erişim kontrol sisteminde kullanmak için basamaklı bir hibrit sistem önerilmektedir. Bu iki bağlantılı teknikte, konuşmacı tanıma (MFCC Yönetimini kullanarak) ve konuşma tanıma (pitch-period ve MFCC Yönetimlerini kullanarak) vardır. Başlangıçta, kullanıcı, konuşmacı tanımayı sistemin ilk aşaması olarak geçmelidir. Sistem tanımlama sonucu yeterli değilse, kullanıcı kabul edilmez ve yazılım durur (yani sistem durur ve kullanıcı kişiyede kişisel bilgisayara girmesine izin vermez(. Ama eğer kullanıcı belirlenirse sistem devam eder ve kullanıcıdan şifre ister. Şifre eşleştirildiğinde, kullanıcı, kişisel bilgisayar sistemine giriş yapabilir. Eşleştirme işleminde, pitch frekansları arasındaki mesafe karşılaştırması kullanılır, bundan sonra deneme ve hata yöntemiyle eşik değeri 12 olarak belirlenir. Bu basamaklı prosedür, Yanlış Pozitif Oranı'nı azaltır ve biyometrik tanıma sisteminin güvenliğini artırır. Konuşmacı tanıma aşamasında, her bir konuşmacı için 50 sözcük kullanarak sisteme iki kullanıcı kaydedilir. Tüm ses kayıtları 10 KHz ve 16 bit çözünürlük örnekleme frekansına sahip bir stereo PC mikrofon tarafından yakalanır. Bu sistem, doğruluk, güvenlik ve penetrasyon zorluğu açısından daha iyi tanımlanmıştır. Konuşmacı tanımlama verimliliği yaklaşık %86'dır, Konuşma tanımlama verimliliği yaklaşık %82'dır ama birleştirerek tanımlama sisteminin verimliliği yaklaşık %91,2'ye kadar yüksektir. Bu sonuçlara göre hibrid ses tanıma işleminin başarısının arttırıldığı söylenebilir. Gelecek çalışmalarımızda kişi sayısı arttırılarak, en az 20 kişi ve her kişi için 100 sözcük kullanarak çalışmanın genişletilmesi planlanmaktadır.

KAYNAKLAR

[1] "Homeland Security Advisory System." Journal, [online], (Date), Available http://www.dhs.gov/xinfoshare/programs/Copy_of_press_release_0046.shtm.

[2] F. Bimbot and L. Mathan. "Text-Free Speaker Recognition Using an Arithmetic-Harmonic Sphericity Measure",Third European Conference on Speech Communication and Technology, 1993.

[3] L. E. Baum and T. Petrie, (1966) "Statistical Inference for Probabilistic of Finite State Markov Chains," The Annals of Mathematical Statistics, vol. 3 no. 6, pp. 1554-1563, 1966

[4] R Sujirat Attawibulkul; (2017) Boonserm Kaewkamnerdpong; Yoshikazu Miyanaga, " Noisy speech training in MFCC-based speech recognition with noise suppression toward robot assisted autism therapy," IEEE Biomedical Engineering International Conference (BMEiCON).

[5] E. Ambikairajah ; M.J. Carey ; G. Tattersall, (1980) "Estimating the pitch period of voicedspeech," IET Electronics Letters 16(12), June 5 1980.

[6] [6] Zunjing Wu ; Zhigang Cao, (2005) "Improved MFCC-based feature for robust speaker identification" TUP Tsinghua Science and Technology 10(2), April 2005.

[7] Hairol Nizam Mohd. Shah, Mohd. Zamzuri Ab Rashid, Mohd. Fairus Abdollah,Muhammad Nizam Kamarudin, Chow Kok Lin and Zalina Kamis, (2014) "Biometric Voice Recognition in Security System", Indian Journal of Science and Technology, 7(2):104–112, February 2014.

[8] M. S. Likitha, S. R. R. Gupta, K. Hasitha and A. U. Raju, (2017)"Speech based human emotion recognition using MFCC," 2017 International Conference on Wireless Communications, Signal Processing and Networking (WiSPNET), Chennai, 2257-2260.

[9] M. Sadeghi and H. Marvi,(2017) "Optimal MFCC features extraction by differential evolution algorithm for speaker recognition," 2017 3rd Iranian

Conference on Intelligent Systems and Signal Processing (ICSPIS), Shahrood, Iran, 169-173.

[10] A. Charisma, M. R. Hidayat and Y. B. Zainal, (2017) "Speaker recognition using mel-frequency cepstrum coefficients and sum square error," 2017 3rd International Conference on Wireless and Telematics (ICWT), Palembang, 160-163.

[11] S. R. Suralkar1, Amol C.Wani2, Prabhakar V. Mhadse, (2013) “Speech Recognized Automation System Using Speaker Identification through Wireless Communication”, IOSR Journal of Electronics and Communication Engineering (IOSR-JECE) e-ISSN: 2278-2834,p- ISSN: 2278-8735. 6(1): 11-18 May. - Jun. 2013, www.iosrjournals.org.

[12] Er. Abhishek Thakur, Assistant Prof. Neeru Singla, “Design of Matlab-Based Automatic Speaker Recognition and Control System”, Er. Abhishek Thakur* et al. / (IJAEST) INTERNATIONAL JOURNAL OF ADVANCED ENGINEERING SCIENCES AND TECHNOLOGIES 8(1):100 – 106

[13] P. Yasotha1 N. Rajasekaran2 M. Nathiya3,(2014) “Automatic Spectral Analysis for Speech to Text Conversion with Embedded Execution”, IJSRD - International Journal for Scientific Research & Development 2(09) :2321-0613, ISSN (online).

[14] Abhishek Thakur1, Rajesh Kumar2, Amandeep Bath3, (2017) Jitender Sharma4, “Automatic Control of Instruments Using Efficient Speech Recognition Algorithm”, IJEEE, 1(1) : 1694-2426 (March 2014).

[15] E. S. Wahyuni, (2017) "Arabic speech recognition using MFCC feature extraction and ANN classification," 2017 2nd International conferences on Information Technology, Information Systems and Electrical Engineering (ICITISEE), 22-25 , Yogyakarta.

[16] Elshamy E. M., Hussein A. I. et al. (2017). Patricia S. Abril and Robert Plant. 2017. Secure VoIP System Based on Biometric Voice Authentication and Nested Digital Cryptosystem using Chaotic Baker’s map and Arnold's Cat Map. Int. Conf. on Computer and Applications, 140-146.

EK-A

```
clc;
clear;
dsms=zeros(5,20);
freqS= 10000; % this is  Sampling Frequency
twindow = hamming(4000); % to smooth the speech signal we used Hamming window
m= [twindow ; zeros(6000,1)];
freq = (1:10000);
mfcc(freq) = 2595 * log(1 + freq / 700); % Linear to Mel frequency scale conversion
triangular = triang(100);
window1 = [triangular ; zeros(9900,1)]; % Defining overlapping triangular windows for
window2 = [zeros(50,1) ; triangular ; zeros(9850,1)]; % frequency domain analysis
window3 = [zeros(100,1) ; triangular ; zeros(9800,1)];
window4 = [zeros(150,1) ; triangular ; zeros(9750,1)];
window5 = [zeros(200,1) ; triangular ; zeros(9700,1)];
window6 = [zeros(250,1) ; triangular ; zeros(9650,1)];
window7 = [zeros(300,1) ; triangular ; zeros(9600,1)];
window8 = [zeros(350,1) ; triangular ; zeros(9550,1)];
window9 = [zeros(400,1) ; triangular ; zeros(9500,1)];
window10 = [zeros(450,1) ; triangular ; zeros(9450,1)];
window11 = [zeros(500,1) ; triangular ; zeros(9400,1)];
window12 = [zeros(550,1) ; triangular ; zeros(9350,1)];
window13 = [zeros(600,1) ; triangular ; zeros(9300,1)];
window14 = [zeros(650,1) ; triangular ; zeros(9250,1)];
```

```
window15 = [zeros(700,1) ; triangular ; zeros(9200,1)];
window16 = [zeros(750,1) ; triangular ; zeros(9150,1)];
window17 = [zeros(800,1) ; triangular ; zeros(9100,1)];
window18 = [zeros(850,1) ; triangular ; zeros(9050,1)];
window19 = [zeros(900,1) ; triangular ; zeros(9000,1)];
window20 = [zeros(950,1) ; triangular ; zeros(8950,1)];
% record the database containts 2 seconds of speech
number_of_words = 5;
for count_data = 1 : number_of_words
 trai=input('Press 1 to stat train signal recording');
sample=audiorecorder(freqS,16,2);
disp('start trian signal recording ');
recordblocking(sample, 2);
disp('end recording ..');
gadt = getaudiodata(sample);
filename = 'train.wav';
audiowrite(filename,gadt,freqS);
art=audioread('train.wav');
figure (1);
plot(art,'r');title('train signal in database');
art=audioread('train.wav');
sample=art;
fduration= 0.1;
flenght=fduration*freqS;
NS=length(sample);
numfms=floor(NS/flenght);
newsgmant=zeros(NS,1);
count1=0;
for R=1 :numfms
%extract the frame of speech
fme = sample(( R-1)*flenght+1 :flenght*R);
maxvalue1=max(fme);
```

```
if(maxvalue1 >0.14)
count1=count1+1;
newsgmant( (count1-1)*flenght +1 : flenght*count1)=fme;

end
end
newsgmant(10000:end)=[];
figure (2);plot(newsgmant);
title( 'train signal without silent component');
FFT1 = fft(newsgmant);% Transform to frequency domain
figure(100);plot(FFT1);
melw = abs(FFT1(floor(mfcc(freq)))); % Mel warping
melw = melw./ max(melw);
figure(3);plot(melw);
sp1 = melw.* window1;
sp2 = melw.* window2;
sp3 = melw.* window3;
sp4 = melw.* window4;
sp5 = melw.* window5;
sp6 = melw.* window6;
sp7 = melw.* window7;
sp8 = melw.* window8;
sp9 = melw.* window9;
sp10 = melw.* window10;
sp11 = melw.* window11;
sp12 = melw.* window12;
sp13 = melw.* window13;
sp14 = melw.* window14;
sp15 = melw.* window15;
sp16 = melw.* window16;
sp17 = melw.* window17;
```

```
sp18 = melw.* window18;
sp19 = melw.* window19;
sp20 = melw.* window20;
sms1 = sum(sp1.^ 2); % Determine the energy of the signal within each window
sms2 = sum(sp2.^ 2); % by summing square of the magnitude of the spectrum
sms3 = sum(sp3.^ 2);
sms4 = sum(sp4.^ 2);
sms5 = sum(sp5.^ 2);
sms6 = sum(sp6.^ 2);
sms7 = sum(sp7.^ 2);
sms8 = sum(sp8.^ 2);
sms9 = sum(sp9.^ 2);
sms10= sum(sp10.^2);
sms11 = sum(sp11.^ 2);
sms12 = sum(sp12.^ 2);
sms13 = sum(sp13.^ 2);
sms14 = sum(sp14.^ 2);
sms15 = sum(sp15.^ 2);
sms16 = sum(sp16.^ 2);
sms17 = sum(sp17.^ 2);
sms18 = sum(sp18.^ 2);
sms19 = sum(sp19.^ 2);
sms20 = sum(sp20.^ 2);
sms = [sms1, sms2, sms3, sms4, sms5, sms6, sms7, sms8, sms9, sms10, sms11, sms12,
        sms13, sms14, sms15, sms16, sms17,sms18, sms19, sms20];
sms = log(sms);
figure(4);plot(sms);
dsms = dct(sms); % Determine DCT of Log of the spectrum energies
figure(5);plot(dsms);
dataset_trainspeaker1(count_data,:)=dsms;
end
freqS=10000;
```

```
window = hamming(4000); %  to smooth the speech signal we usedHamming window
m = [window ; zeros(6000,1)];
freq = (1:10000);
mfcc(freq) = 2595 * log(1 + freq / 700); % Linear to Mel frequency scale conversion
triangular = triang(100);
window1 = [triangular ; zeros(9900,1)]; % Defining overlapping triangular windows for
window2 = [zeros(50,1) ; triangular ; zeros(9850,1)]; % frequency domain analysis
window3 = [zeros(100,1) ; triangular ; zeros(9800,1)];
window4 = [zeros(150,1) ; triangular ; zeros(9750,1)];
window5 = [zeros(200,1) ; triangular ; zeros(9700,1)];
window6 = [zeros(250,1) ; triangular ; zeros(9650,1)];
window7 = [zeros(300,1) ; triangular ; zeros(9600,1)];
window8 = [zeros(350,1) ; triangular ; zeros(9550,1)];
window9 = [zeros(400,1) ; triangular ; zeros(9500,1)];
window10 = [zeros(450,1) ; triangular ; zeros(9450,1)];
window11 = [zeros(500,1) ; triangular ; zeros(9400,1)];
window12 = [zeros(550,1) ; triangular ; zeros(9350,1)];
window13 = [zeros(600,1) ; triangular ; zeros(9300,1)];
window14 = [zeros(650,1) ; triangular ; zeros(9250,1)];
window15 = [zeros(700,1) ; triangular ; zeros(9200,1)];
window16 = [zeros(750,1) ; triangular ; zeros(9150,1)];
window17 = [zeros(800,1) ; triangular ; zeros(9100,1)];
window18 = [zeros(850,1) ; triangular ; zeros(9050,1)];
window19 = [zeros(900,1) ; triangular ; zeros(9000,1)];
window20 = [zeros(950,1) ; triangular ; zeros(8950,1)];
test=input('Press 2 to start test signal recording');
ii=audiorecorder(freqS,16,2);
disp(' start test signal recording ');
recordblocking(ii, 2);
disp('end recording');
gad = getaudiodata(ii);
filename = 'test.wav';
```

```
audiowrite(filename,gad,freqS);
gg=audioread('test.wav');
figure (6);
plot(gg);title('test signal')
freqd= 0.1;
newfreq=freqd*freqS;
NNn=length(gg);
NFn=floor(NNn/newfreq);
NZs=zeros(NNn,1);
count2=0;
for RR=1 :NFn
%extract the frame of speech
frame = gg(( RR-1)*newfreq+1 :newfreq*RR);
max_value=max(frame);

if(max_value >0.14)
count2=count2+1;
NZs( (count2-1)*newfreq +1 : newfreq*count2)=frame;

end
end
figure(7);plot(NZs);
title('test signal with out silent part');
NZs(10000:end)=[];
 figure(8);plot(NZs);
FFT = fft(NZs); % Transform to frequency domain
Melw = abs(FFT(floor(mfcc(freq)))); % Mel warping
Melw = Melw./ max(Melw);
figure(9);plot(melw);
sp1 = Melw.* window1;
sp2 = Melw.* window2;
sp3 = Melw.* window3;
```

```
sp4 = Melw.* window4;
sp5 = Melw.* window5;
sp6 = Melw.* window6;
sp7 = Melw.* window7;
sp8 = Melw.* window8;
sp9 = Melw.* window9;
sp10 = Melw.* window10;
sp11 = Melw.* window11;
sp12 = Melw.* window12;
sp13 = Melw.* window13;
sp14 = Melw.* window14;
sp15 = Melw.* window15;
sp16 = Melw.* window16;
sp17 = Melw.* window17;
sp18 = Melw.* window18;
sp19 = Melw.* window19;
sp20 = Melw.* window20;
sms1 = sum(sp1.^ 2);
sms2 = sum(sp2.^ 2);
sms3 = sum(sp3.^ 2);
sms4 = sum(sp4.^ 2);
sms5 = sum(sp5.^ 2);
sms6 = sum(sp6.^ 2);
sms7 = sum(sp7.^ 2);
sms8 = sum(sp8.^ 2);
sms9 = sum(sp9.^ 2);
sms10 = sum(sp10.^ 2); % Determine the energy of the signal within each window
sms11 = sum(sp11.^ 2); % by summing square of the magnitude of the spectrum
sms12 = sum(sp12.^ 2);
sms13 = sum(sp13.^ 2);
sms14 = sum(sp14.^ 2);
sms15 = sum(sp15.^ 2);
```

```
sms16 = sum(sp16.^ 2);
sms17 = sum(sp17.^ 2);
sms18 = sum(sp18.^ 2);
sms19 = sum(sp19.^ 2);
sms20 = sum(sp20.^ 2);
sms = [sms1, sms2, sms3, sms4, sms5, sms6, sms7, sms8, sms9, sms10, sms11, sms12,
        sms13, sms14, sms15, sms16, sms17,sms18, sms19, sms20];
sms = log(sms);
figure(10);plot(sms);
dsp = dct(sms);% Determine DCT of Log of the spectrum energies
%dsp =dsp';
figure(11);plot(dsp);
load('dataset_trainspeaker1.mat');
for hh=1:2
  for j=1:20
dsms(j) = dataset_trainspeaker1(hh,j);
  end
 MSE=(sum((dsms - dsp).^ 2)) / 20;
if MSE<2

fprintf('\n\nYou are the same user\n\n');

%match=imshow('sedra.jpg'); % “Access sedra” is output if within threshold
break;
else
fprintf('\n\nYou are not a same user\n\n');
%notmatch=imshow('123.jpg');% “Access 123” is output in case of a failure (the
        speaker is not recognized)
End
Fs= 10000; % this is  Sampling Frequency
twindow = hamming(4000); % to smooth the speech signal we used Hamming window
m= [twindow ; zeros(6000,1)];
freq = (1:10000);
```

```
mfcc(freq) = 2595 * log(1 + freq / 700); % Linear to Mel frequency scale conversion
triangular = triang(100);
window1 = [triangular ; zeros(9900,1)]; % Defining overlapping triangular windows for
window2 = [zeros(50,1) ; triangular ; zeros(9850,1)]; % frequency domain analysis
window3 = [zeros(100,1) ; triangular ; zeros(9800,1)];
window4 = [zeros(150,1) ; triangular ; zeros(9750,1)];
window5 = [zeros(200,1) ; triangular ; zeros(9700,1)];
window6 = [zeros(250,1) ; triangular ; zeros(9650,1)];
window7 = [zeros(300,1) ; triangular ; zeros(9600,1)];
window8 = [zeros(350,1) ; triangular ; zeros(9550,1)];
window9 = [zeros(400,1) ; triangular ; zeros(9500,1)];
window10 = [zeros(450,1) ; triangular ; zeros(9450,1)];
window11 = [zeros(500,1) ; triangular ; zeros(9400,1)];
window12 = [zeros(550,1) ; triangular ; zeros(9350,1)];
window13 = [zeros(600,1) ; triangular ; zeros(9300,1)];
window14 = [zeros(650,1) ; triangular ; zeros(9250,1)];
window15 = [zeros(700,1) ; triangular ; zeros(9200,1)];
window16 = [zeros(750,1) ; triangular ; zeros(9150,1)];
window17 = [zeros(800,1) ; triangular ; zeros(9100,1)];
window18 = [zeros(850,1) ; triangular ; zeros(9050,1)];
window19 = [zeros(900,1) ; triangular ; zeros(9000,1)];
window20 = [zeros(950,1) ; triangular ; zeros(8950,1)];
% record the database containts 2 seconds of speech
number_of_words = 5;
for count_data = 1 : number_of_words
   Fs=10000;
   %enter the word(writing)
   word = input('write the word:','s');
   % enter the word (speech)
   choice=input('Press 1 for sample recording:');
   i=audiorecorder(Fs,16,2);
   disp('start recording first signal ..');
```

```
  recordblocking(i, 2);
  disp('end recording ..');
  y = getaudiodata(i);
  filename = 'w.wav';
  audiowrite(filename,y,Fs);
  %file will be stored as kkk.wav
  % to plot your input
figure (1);
plot(s1,'r');
%    title (' referance signal in database');
  %you can remove the unvoiced components from the signal (frame by frame
  %analysis)
  s1=audioread('w.wav');
  x=s1;
  frame_duration= 0.1;
  frame_len=frame_duration*Fs;
  N=length(x);
  num_frames=floor(N/frame_len);
  new_segmant=zeros(N,1);
  count=0;
  for k=1 :num_frames

    %extract the frame of speech
    frame = x(( k-1)*frame_len+1 :frame_len*k);
    max_val=max(frame);

    if(max_val >0.14)
      count=count+1;
      new_segmant( (count-1)*frame_len +1 : frame_len*count)=frame;

    end
  end
```

```
   new_segmant(10000:end)=[];

figure (2);plot( new_segmant);
title( 'train signal without silent component');
FFT1 = fft( new_segmant);% Transform to frequency domain
melw = abs(FFT1(floor(mfcc(freq)))); % Mel warping
melw = melw./ max(melw);
figure(3);plot(melw);
sp1 = melw.* window1;
sp2 = melw.* window2;
sp3 = melw.* window3;
sp4 = melw.* window4;
sp5 = melw.* window5;
sp6 = melw.* window6;
sp7 = melw.* window7;
sp8 = melw.* window8;
sp9 = melw.* window9;
sp10 = melw.* window10;
sp11 = melw.* window11;
sp12 = melw.* window12;
sp13 = melw.* window13;
sp14 = melw.* window14;
sp15 = melw.* window15;
sp16 = melw.* window16;
sp17 = melw.* window17;
sp18 = melw.* window18;
sp19 = melw.* window19;
sp20 = melw.* window20;
sms1 = sum(sp1.^ 2); % Determine the energy of the signal within each window
sms2 = sum(sp2.^ 2); % by summing square of the magnitude of the spectrum
sms3 = sum(sp3.^ 2);
sms4 = sum(sp4.^ 2);
```

```
sms5 = sum(sp5.^ 2);
sms6 = sum(sp6.^ 2);
sms7 = sum(sp7.^ 2);
sms8 = sum(sp8.^ 2);
sms9 = sum(sp9.^ 2);
sms10= sum(sp10.^2);
sms11 = sum(sp11.^ 2);
sms12 = sum(sp12.^ 2);
sms13 = sum(sp13.^ 2);
sms14 = sum(sp14.^ 2);
sms15 = sum(sp15.^ 2);
sms16 = sum(sp16.^ 2);
sms17 = sum(sp17.^ 2);
sms18 = sum(sp18.^ 2);
sms19 = sum(sp19.^ 2);
sms20 = sum(sp20.^ 2);
sms = [sms1, sms2, sms3, sms4, sms5, sms6, sms7, sms8, sms9, sms10, sms11, sms12,
        sms13, sms14, sms15, sms16, sms17,sms18, sms19, sms20];
sms = log(sms);
figure(4);plot(sms);
dsms = dct(sms); % Determine DCT of Log of the spectrum energies
A=dsms;
A = A.';
y=audioread('w.wav');
Fs=10000;
max_value=max(abs(y));
y=y/max_value;
t=(1/Fs:1/Fs:(length(y)/Fs));
figure(4);
subplot(2,1,1);
plot(t,y);
sum1=0;autocorrelation=0;
```

```
for l=0:(length(y)-1)
sum1=0;
for u=1:(length(y)-l)
s=y(u)*y(u+l);
sum1=sum1+s;
end
autocor(l+1)=sum1;
end
kk=(1/Fs:1/Fs:(length(autocor)/Fs))*1000;
subplot(2,1,2);
plot(kk,autocor);

auto=autocor(21:160);
max1=0;
for uu=1:140
if(auto(uu)>max1)
max1=auto(uu);
sample_no=uu;
end
end
pitch_period_To=(20+sample_no)*(1/Fs);
Fo=1/pitch_period_To;
 dataset_train(count_data,1)= Fo;
 dataset_train1(count_data,:)= A;
    dataset_train_class1(count_data,:)= string(word);
end
Fs= 10000
window = hamming(4000); %  to smooth the speech signal we usedHamming window
m = [window ; zeros(6000,1)];
freq = (1:10000);
mfcc(freq) = 2595 * log(1 + freq / 700); % Linear to Mel frequency scale conversion
triangular = triang(100);
```

```
window1 = [triangular ; zeros(9900,1)]; % Defining overlapping triangular windows for
window2 = [zeros(50,1) ; triangular ; zeros(9850,1)]; % frequency domain analysis
window3 = [zeros(100,1) ; triangular ; zeros(9800,1)];
window4 = [zeros(150,1) ; triangular ; zeros(9750,1)];
window5 = [zeros(200,1) ; triangular ; zeros(9700,1)];
window6 = [zeros(250,1) ; triangular ; zeros(9650,1)];
window7 = [zeros(300,1) ; triangular ; zeros(9600,1)];
window8 = [zeros(350,1) ; triangular ; zeros(9550,1)];
window9 = [zeros(400,1) ; triangular ; zeros(9500,1)];
window10 = [zeros(450,1) ; triangular ; zeros(9450,1)];
window11 = [zeros(500,1) ; triangular ; zeros(9400,1)];
window12 = [zeros(550,1) ; triangular ; zeros(9350,1)];
window13 = [zeros(600,1) ; triangular ; zeros(9300,1)];
window14 = [zeros(650,1) ; triangular ; zeros(9250,1)];
window15 = [zeros(700,1) ; triangular ; zeros(9200,1)];
window16 = [zeros(750,1) ; triangular ; zeros(9150,1)];
window17 = [zeros(800,1) ; triangular ; zeros(9100,1)];
window18 = [zeros(850,1) ; triangular ; zeros(9050,1)];
window19 = [zeros(900,1) ; triangular ; zeros(9000,1)];
window20 = [zeros(950,1) ; triangular ; zeros(8950,1)];
number_of_words = 100;
for count_data = 1 : number_of_words
  Fs=10000;
test=input('Press 2 to start test signal recording');
ii=audiorecorder(Fs,16,2);
disp(' start test signal recording ');
recordblocking(ii, 2);
disp('end recording');
gad = getaudiodata(ii);
filename = 'test.wav';
audiowrite(filename,gad,Fs);
gg=audioread('test.wav');
```

```
figure (6);
plot(gg);title('test signal')
freqd= 0.1;
newfreq=freqd*Fs;
NNn=length(gg);
NFn=floor(NNn/newfreq);
NZs=zeros(NNn,1);
count2=0;
for RR=1 :NFn
%extract the frame of speech
frame = gg(( RR-1)*newfreq+1 :newfreq*RR);
max_value=max(frame);

if(max_value >0.14)
count2=count2+1;
NZs( (count2-1)*newfreq +1 : newfreq*count2)=frame;

end
end
figure(7);plot(NZs);
title('test signal with out silent part');
NZs(10000:end)=[];
 figure(8);plot(NZs);
FFT = fft(NZs); % Transform to frequency domain
Melw = abs(FFT(floor(mfcc(freq)))); % Mel warping
Melw = Melw./ max(Melw);
figure(9);plot(Melw);
sp1 = Melw.* window1;
sp2 = Melw.* window2;
sp3 = Melw.* window3;
sp4 = Melw.* window4;
sp5 = Melw.* window5;
```

```
sp6 = Melw.* window6;
sp7 = Melw.* window7;
sp8 = Melw.* window8;
sp9 = Melw.* window9;
sp10 = Melw.* window10;
sp11 = Melw.* window11;
sp12 = Melw.* window12;
sp13 = Melw.* window13;
sp14 = Melw.* window14;
sp15 = Melw.* window15;
sp16 = Melw.* window16;
sp17 = Melw.* window17;
sp18 = Melw.* window18;
sp19 = Melw.* window19;
sp20 = Melw.* window20;
sms1 = sum(sp1.^ 2);
sms2 = sum(sp2.^ 2);
sms3 = sum(sp3.^ 2);
sms4 = sum(sp4.^ 2);
sms5 = sum(sp5.^ 2);
sms6 = sum(sp6.^ 2);
sms7 = sum(sp7.^ 2);
sms8 = sum(sp8.^ 2);
sms9 = sum(sp9.^ 2);
sms10 = sum(sp10.^ 2); % Determine the energy of the signal within each window
sms11 = sum(sp11.^ 2); % by summing square of the magnitude of the spectrum
sms12 = sum(sp12.^ 2);
sms13 = sum(sp13.^ 2);
sms14 = sum(sp14.^ 2);
sms15 = sum(sp15.^ 2);
sms16 = sum(sp16.^ 2);
sms17 = sum(sp17.^ 2);
```

```
sms18 = sum(sp18.^ 2);
sms19 = sum(sp19.^ 2);
sms20 = sum(sp20.^ 2);
sms = [sms1, sms2, sms3, sms4, sms5, sms6, sms7, sms8, sms9, sms10, sms11, sms12,
       sms13, sms14, sms15, sms16, sms17,sms18, sms19, sms20];
sms = log(sms);
figure(10);plot(sms);
dsp = dct(sms);% Determine DCT of Log of the spectrum energies
figure(11);plot(dsp);
bc=audioread('test.wav');
Fs=10000;
max_value2=max(abs(bc));
bc=bc/max_value2;
t2=(1/Fs:1/Fs:(length(bc)/Fs));
subplot(2,1,1);
plot(t2,bc);
sum2=0;autocorrelation2=0;
for L=0:(length(bc)-1)
sum2=0;
for u2=1:(length(bc)-L)
s=bc(u2)*bc(u2+L);
sum2=sum2+s;
end
autocor(L+1)=sum2;
end
kkl=(1/Fs:1/Fs:(length(autocor)/Fs))*1000;
%figure(6);
%subplot(2,1,2);
%plot(kkl,autocor);

auto2=autocor(21:160);
max2=0;
```

```
for uu2=1:140
if(auto2(uu2)>max2)
max2=auto2(uu2);
sample_na=uu2;
end
end
pitch_period_T1=(20+sample_na)*(1/Fs);
F1=1/pitch_period_T1;
dataset_traintt(count_data,1)= F1;
 dataset_train1tt(count_data,:)= dsp;
   % dataset_train_class1tt(count_data,:)= string(word);
End
window = hamming(4000); %  to smooth the speech signal we usedHamming window
m = [window ; zeros(6000,1)];
freq = (1:10000);
mfcc(freq) = 2595 * log(1 + freq / 700); % Linear to Mel frequency scale conversion
triangular = triang(100);
window1 = [triangular ; zeros(9900,1)]; % Defining overlapping triangular windows for
window2 = [zeros(50,1) ; triangular ; zeros(9850,1)]; % frequency domain analysis
window3 = [zeros(100,1) ; triangular ; zeros(9800,1)];
window4 = [zeros(150,1) ; triangular ; zeros(9750,1)];
window5 = [zeros(200,1) ; triangular ; zeros(9700,1)];
window6 = [zeros(250,1) ; triangular ; zeros(9650,1)];
window7 = [zeros(300,1) ; triangular ; zeros(9600,1)];
window8 = [zeros(350,1) ; triangular ; zeros(9550,1)];
window9 = [zeros(400,1) ; triangular ; zeros(9500,1)];
window10 = [zeros(450,1) ; triangular ; zeros(9450,1)];
window11 = [zeros(500,1) ; triangular ; zeros(9400,1)];
window12 = [zeros(550,1) ; triangular ; zeros(9350,1)];
window13 = [zeros(600,1) ; triangular ; zeros(9300,1)];
window14 = [zeros(650,1) ; triangular ; zeros(9250,1)];
window15 = [zeros(700,1) ; triangular ; zeros(9200,1)];
```

```
window16 = [zeros(750,1) ; triangular ; zeros(9150,1)];
window17 = [zeros(800,1) ; triangular ; zeros(9100,1)];
window18 = [zeros(850,1) ; triangular ; zeros(9050,1)];
window19 = [zeros(900,1) ; triangular ; zeros(9000,1)];
window20 = [zeros(950,1) ; triangular ; zeros(8950,1)];
test=input('Press 2 to start test signal recording');
ii=audiorecorder(freqS,16,2);
disp(' start test signal recording ');
recordblocking(ii, 2);
disp('end recording');
gad = getaudiodata(ii);
filename = 'test.wav';
audiowrite(filename,gad,freqS);
gg=audioread('test.wav');
%figure (6);
plot(gg);title('test signal')
freqd= 0.1;
newfreq=freqd*freqS;
NNn=length(gg);
NFn=floor(NNn/newfreq);
NZs=zeros(NNn,1);
count2=0;
for RR=1 :NFn
%extract the frame of speech
frame = gg(( RR-1)*newfreq+1 :newfreq*RR);
max_value=max(frame);

if(max_value >0.14)
count2=count2+1;
NZs( (count2-1)*newfreq +1 : newfreq*count2)=frame;

end
```

```
end
figure(7);plot(NZs);
title('test signal with out silent part');
NZs(10000:end)=[];
 %figure(8);plot(NZs);
FFT = fft(NZs); % Transform to frequency domain
Melw = abs(FFT(floor(mfcc(freq)))); % Mel warping
Melw = Melw./ max(Melw);
%figure(9);plot(Melw);
sp1 = Melw.* window1;
sp2 = Melw.* window2;
sp3 = Melw.* window3;
sp4 = Melw.* window4;
sp5 = Melw.* window5;
sp6 = Melw.* window6;
sp7 = Melw.* window7;
sp8 = Melw.* window8;
sp9 = Melw.* window9;
sp10 = Melw.* window10;
sp11 = Melw.* window11;
sp12 = Melw.* window12;
sp13 = Melw.* window13;
sp14 = Melw.* window14;
sp15 = Melw.* window15;
sp16 = Melw.* window16;
sp17 = Melw.* window17;
sp18 = Melw.* window18;
sp19 = Melw.* window19;
sp20 = Melw.* window20;
sms1 = sum(sp1.^ 2);
sms2 = sum(sp2.^ 2);
sms3 = sum(sp3.^ 2);
```

```
sms4 = sum(sp4.^ 2);
sms5 = sum(sp5.^ 2);
sms6 = sum(sp6.^ 2);
sms7 = sum(sp7.^ 2);
sms8 = sum(sp8.^ 2);
sms9 = sum(sp9.^ 2);
sms10 = sum(sp10.^ 2); % Determine the energy of the signal within each window
sms11 = sum(sp11.^ 2); % by summing square of the magnitude of the spectrum
sms12 = sum(sp12.^ 2);
sms13 = sum(sp13.^ 2);
sms14 = sum(sp14.^ 2);
sms15 = sum(sp15.^ 2);
sms16 = sum(sp16.^ 2);
sms17 = sum(sp17.^ 2);
sms18 = sum(sp18.^ 2);
sms19 = sum(sp19.^ 2);
sms20 = sum(sp20.^ 2);
sms = [sms1, sms2, sms3, sms4, sms5, sms6, sms7, sms8, sms9, sms10, sms11, sms12,
        sms13, sms14, sms15, sms16, sms17,sms18, sms19, sms20];
sms = log(sms);
%figure(10);plot(sms);
dsp = dct(sms);% Determine DCT of Log of the spectrum energies
%figure(11);plot(dsp);
load('dataset_trainspeaker1.mat');
%load('dataset_trainspeakerclass.mat');
for hh=1:2
  for j=1:20
dsms(j) = dataset_trainspeaker1(hh,j);
%dsp = dataset_train1tt(hh,j);
  end

MSE=(sum((dsms - dsp).^ 2)) / 20; % Determine the Mean squared error
```

```
if MSE<1.5
fprintf('\n\nYou are the same user\n\n');
match=imshow('sedra.jpg'); % “Access sedra” is output if within threshold
break
else
fprintf('\n\nYou are not a same user\n\n');
notmatch=imshow('123.jpg');% “Access 123” is output in case of a failure (the speaker
        is not recognized)
end
end
if MSE<1.5
   load('dataset_train.mat');
load('dataset_train1.mat');
load('dataset_traintt.mat');
load('dataset_train1tt.mat');
%ww=load('dataset_train_class1.mat');
for t=1:5
FF = dataset_train(t);
F1= dataset_traintt(t);
r1=abs(F1-FF);
for j=1:20
A = dataset_train1(t,j);
dsp = dataset_train1tt(t,j);
end
 A = A.';
 MSE=(sum((A - dsp).^ 2)) / 20;
 if r1<=12&&MSE<1.5
disp('the imput is matching with database');
else
disp(' invalid input - not matching-');
 end
end
```

```
else
end

load('dataset_train.mat');
load('dataset_train1.mat');
load('dataset_traintt.mat');
load('dataset_train1tt.mat');
%ww=load('dataset_train_class1.mat');
for t=1:5
FF = dataset_train(t);
F1= dataset_traintt(t);
r1=abs(F1-FF);
for j=1:20
A = dataset_train1(t,j);
dsp = dataset_train1tt(t,j);
end
 A = A.';
 MSE=(sum((A - dsp).^ 2)) / 20;
 if r1<=12&&MSE<1.5
disp('the imput is matching with database');
else
disp(' invalid input - not matching-');
 end
end
end
```

ÖZGEÇMİŞ

KİŞİSEL BİLGİLER

Adı Soyadı : Roaya Salhalden ABDALRAHMAN

Doğum Tarihi ve Yeri : 13/11/1991 , IRAK/KERKÜK

Yabancı Dili : Arabca/İngilizce

E-posta : rouyasalihy@gmail.com

ÖĞRENİM DURUMU

Derece	**Alan**	**Okul/Üniversite**	**Mezuniyet Yılı**
Y. Lisans			
Lisans	Electronic and Control	Karkuk Technical Unv.	2013
Lise	General	Karkuk Preparator high school	2009

İŞ TECRÜBESİ

Yıl	**Firma/Kurum**	**Görevi**
2014	Karkuk Technical Unv.	Assistant Professor

YAYINLARI

Bildiri

1. Roaya Salhalden A. Abdalrahman, Bülent Bolat , Nihan Kahraman, (2018) “A cascaded voice biometric system”, Procedia Computer Science,(131): 1223–1228 .