



T.C
KIRŞEHİR AHİ EVRAN ÜNİVERSİTESİ
FEN BİLİMLERİ ENSTİTÜSÜ
İLERİ TEKNOLOJİLER ANABİLİM DALI

SES KOMUTLARI İLE ROBOT KOLU KONTROLÜ

Ozan Fırat ÇIPLAK

YÜKSEK LİSANS TEZİ
İLERİ TEKNOLOJİLER ANABİLİM DALI

KIRŞEHİR / 2021



T.C
KIRŞEHİR AHI EVRAN ÜNİVERSİTESİ
FEN BİLİMLERİ ENSTİTÜSÜ
İLERİ TEKNOLOJİLER ANABİLİM DALI

SES KOMUTLARI İLE ROBOT KOLU KONTROLÜ

Ozan Fırat ÇIPLAK

YÜKSEK LİSANS TEZİ

DANIŞMAN

Dr. Öğr. Üyesi Serkan KESER

KIRŞEHİR / 2021

TEZ BİLDİRİMİ

Tez içindeki bütün bilgilerin etik davranış ve akademik kurallar çerçevesinde elde edilerek sunulduğunu, ayrıca tez yazım kurallarına uygun olarak hazırlanan bu çalışmada bana ait olmayan her türlü ifade bilginin kaynağına eksiksiz atıf yapıldığını bildiririm.

Ozan Fırat ÇIPLAK



20.04.2016 tarihli Resmi Gazete’de yayımlanan Lisansüstü Eğitim ve Öğretim Yönetmeliğinin 9/2 ve 22/2 maddeleri gereğince; Bu Lisansüstü teze, Kırşehir Ahi Evran Üniversitesi’nin abonesi olduğu intihal yazılım programı kullanılarak Fen Bilimleri Enstitüsü’nün belirlemiş olduğu ölçütlere uygun rapor alınmıştır.



ÖNSÖZ

Mühendislik alanında yüksek lisans yapmak bir mühendise Yüksek Mühendis ünvanı kazandırır. Bu nedenle bir yüksek lisans çalışması çok değerlidir. Bana fırsatı sağlayarak yüksek lisans eğitimim boyunca sağlam bilimsel bilgisi ile her türlü desteği sağlayan, bilgi ve deneyimini benimle paylaşan, beni anlayışla yönlendiren ve bun tezin ortaya çıkmasını sağlayan tez danışmanım Dr. Öğr. Üyesi Serak KESER’e minnet duyduğumu özellikle belirtmek isterim. Yüksek lisans eğitimim sırasında ailem her zaman yanımda olan ve beni teşvik eden aileme ve Özlem/Zeynel BAŞIBÜYÜK (Kırşehir Ahi Evran Üniversitesi) çiftine özel teşekkürlerimi iletmek isterim.

Temmuz, 2021

Ozan Fırat ÇIPLAK

İÇİNDEKİLER

ÖNSÖZ	ii
İÇİNDEKİLER.....	iii
ŞEKİLLER LİSTESİ	vi
TABLolar LİSTESİ	vii
SİMGE VE KISALTMALAR	viii
ÖZET	ix
ABSTRACT	xi
1. GİRİŞ	1
1.1. Araştırmanın Önemi	2
2. KAYNAK ARAŞTIRMASI	4
2.1. Konuşma Tanımanın Tanımı ve Avantajları	4
2.1.1. Kullanım Kolaylığı	4
2.1.2. Veri toplama hızı	4
2.1.3. Uzaktan Veri İletme İmkânı	4
2.2. Konuşma Tanıma Çeşitleri	4
2.2.1. Konuşmacıya Göre Konuşma Tanıma	5
2.2.1.1 Konuşmacı Bağımlı Konuşma Tanıma	5
2.2.1.2 Konuşmacı Bağımsız Konuşma Tanıma	6
2.2.1.3 Konuşmacı Bağımlı ve Bağımsız Tanıma Sistemlerinin Değerlendirilmesi	6
2.2.2. Temel Alınan Ses Birimine Göre Konuşma Tanıma	6
2.2.2.1. Fonem Tabanlı Konuşma Tanıma	7
2.2.2.2. Kelime Tabanlı Konuşma Tanıma	7
2.2.3. Sesin Sürekliliğine Göre Konuşma Tanıma	7
2.2.3.1. İzole Konuşma Tanıma	7
2.2.3.2. Bağlı Konuşma Tanıma	8
2.2.3.3. Sürekli Konuşma Tanım	8
2.2.4. Metne Göre Konuşma Tanıma	8
2.2.4.1. Metne Bağımlı Konuşma Tanıma	8
2.2.4.2. Metinden Bağımsız Konuşma Tanıma	8
2.3. Konuşma Tanımanın Aşamaları	9
2.3.1. Konuşma Sinyali Sınırlarının Belirlenmesi	9

2.3.1.1. Kısa süreli enerji (STE)	9
2.3.1.2. Sıfır Geçiş Oranı	10
2.3.2. Konuşma sinyalinin ön işlenmesi	10
2.3.3. Öznitelik vektörlerinin elde edilmesi	10
2.3.3.1. Çerçeveleme ve Pencereleme	12
2.3.3.2. Hızlı Fourier Transformu (HFD)	13
2.3.3.3. Mel filtreleme ve MFK Katsayılarının Elde Edilmesi	13
2.4. Konuşma Tanımda Kullanılan Temel Sınıflandırıcılar	14
2.4.1. Dinamik Zaman Bükmesi (DZB)	15
2.4.2. Yapay Sinir Ağları (YSA)	15
2.4.2.1. YSA'ların Kullanıldığı Alanlar	16
2.4.2.2. Yapılarına Göre Yapay Sinir Ağları	16
2.4.2.3. Yapay Sinir Ağının Yapısı	16
2.4.2.4. Yapay Sinir Ağlarının Eğitilmesi	18
2.4.3. Saklı Markov Modelleri (Hidden Markov Models (HMM))	18
2.4.4. Alt Uzay Sınıflandırıcılar	18
2.4.4.1. Ortak Vektör Yaklaşımı (AOV) ve Ayırt Edici Ortak Vektör Yaklaşımı (AOVY)	18
2.4.4.2. Fisher Doğrusal Ayrım Analizi (FDAA)	20
3. MATERYAL VE YÖNTEM	22
3.1. Yalıtık Kelime Tanıma	22
3.2. Eğitim Veritabanının Oluşturulması	22
3.3. Dil Modelleri	23
3.3.1. Dil Modelli Sınıflama	23
3.3.2. Dil Modelsiz Sınıflama	24
3.4. Sınıflandırıcıların kullanımı	24
3.4.1. Ayırt Edici Ortak Vektör Yaklaşımının Kullanımı	24
3.4.2. FDAA Kullanımı	24
3.4.3. YSA Kullanımı	24
3.5. Robot Kolu Kontrolü	25
3.5.1. Robot Kolu Yapısı	25
3.5.2. Robot Kolu Koordinatları	30
4. BULGULAR	31
4.1. AOVY ve FDAA İçin Bulunan Konuşma Tanıma Oranları	32

4.2. YSA İin Bulunan Konuşma Tanıma Oranları	34
4.3 Kullanılan Sınıflandırıcıların Tanıma Süreleri	36
5. SONUÇ ve ÖNERİLER	36
KAYNAKÇA	38
EKLER	42
ÖZGEÇMİŞ	45



ŞEKİLLER LİSTESİ

Şekil 2.1. Konuşma tanıma türleri	5
Şekil 2.2. Konuşmacı bağımlı ve bağımsız tanıma oranlarının yıllara göre değerleri [11]	6
Şekil 2.3. Konuşma tanıma aşamaları	9
Şekil 2.4. Sinyalin sıfır geçiş ve kısa süreli enerji grafiği.....	10
Şekil 2.5. MFKK katsayılarının elde edilme aşamaları	11
Şekil 2.6. Çerçeveleme örneği	12
Şekil 2.7. Hamming pencere yapısı	13
Şekil 2.8. Frekans-Mel Dönüşümü	14
Şekil 2.9. Mel bank filtre yapısı	14
Şekil 2.10. Yapay sinir ağı katmanları	17
Şekil 2.11. Fischer doğrusal ayrım analizi	20
Şekil 3.1. Kullanılan robot kolu	25
Şekil 3.2. Servo 1	26
Şekil 3.3. Servo 2 ve 3	26
Şekil 3.4. Servo 4	27
Şekil 3.5. Servo 5	28
Şekil 3.6. Servo 6	28
Şekil 3.7. Tanıma için kullanılan robot kol sistemi	30
Şekil 3.8. 24 adet nesnenin konumları	30
Şekil 4.1. AOVY için dil modelsiz komut kümeleri tanıma oranları (Ortalama=%90,73)	32
Şekil 4.2. FDAA için dil modelsiz komut kümeleri tanıma oranları (Ortalama=%89,48)	32
Şekil 4.3. AOVY için dil modellenli komut kümeleri tanıma oranları (Ortalama=%98,3)	33
Şekil 4.4. FDAA için dil modellenli komut kümeleri tanıma oranları (Ortalama=%97,1)	34
Şekil 4.5. YSA için dil modelsiz tanıma oranları (ortalama=%81,9).....	35
Şekil 4.6. YSA için dil modellenli tanıma oranları (ortalama=%91,1).....	36

TABLÖLAR LİSTESİ

Tablo 3.1. Renkli nesnelere ait komut kümesi	23
Tablo 3.2. Tanınan komutlara karşılık gönderilen karakterler	30
Tablo 4.1: AOVY, OVY, FDAA ve YSA için ortalama tanıma oranları.....	31



SİMGE VE KISALTMALAR

Kısaltmalar	Açıklama
AOVY	: Ayıredici Ortak Vektör Yaklaşımı
AOV	: Ortak Vektör Yaklaşımı
AKD	: Ayrık Kosinüs Dönüşümü
ANN	: Artificial Neural Network
DAA	: Doğrusal Ayrım Analizi
DCVA	: Discriminative Common Vector Approach
DVM	: Destek Vektör Makinesi
DZB	: Dinamik Zaman Bükme,
FLDA	: Fisher Linear Discriminant Analysis
FDA	: Fisher Doğrusal Ayrım Analizi
SMM	: Saklı Markov Model
Hz.	: Hertz
HFD	: Hızlı Fourier Dönüşümü
KSE	: Kısa Süreli Enerji
LÖK	: Lineer Öngörü Katsayıları
MFKK	: Mel frekans kepsral katsayıları
MFS	: Mel-Frekansı Saptırması
SGO	: Sıfır geçiş oranı
YSA	: Yapay Sinir Ağları

ÖZET

Yüksek Lisans Tezi SES KOMUTLARI İLE ROBOT KOLU KONTROLÜ

Ozan Fırat ÇIPLAK

Kırşehir Ahi Evran Üniversitesi

Fen Bilimleri Enstitüsü

İleri Teknolojiler Anabilim Dalı

Danışman: Dr. Öğr. Üyesi Serkan KESER

Gün geçtikçe cihazların uzaktan kontrolünü gerçekleştiren tanıma sistemleri büyük önem kazanmaktadır. En çok kullanılan tanıma sistemleri olarak konuşma, yüz ve parmak izi tanıma sistemleri gösterilebilir. Konuşma tanıma sistemleri güvenlik sistemlerinde, cihaz kontrolü sistemlerinde ve dikte ettirme sistemlerinde gerçek zamanlı olarak kullanılabilir. Bu çalışmada konuşma komutlarının gerçek zamanlı olarak tanınması ile robot kolu kontrolü gerçekleştirilmiştir. Konuşma komutlarının tanınması için etkili alt uzay sınıflandırıcıları olan Fisher Doğrusal Ayrım Analizi (FDAA), Ayırteci Ortak Vektör (AOVY) ve Yapay Sinir Ağları (YSA) kullanılmıştır. Bu sınıflandırıcılardan AOVY ilk kez ses tanımada kullanılmıştır. Eğitim seti için, herbiri altı farklı renge sahip dört farklı nesne için toplam 24 adet konuşma cümleleri oluşturulmuştur. Eğitim kümesi konuşmacı bağımlı olarak oluşturulmuştur. Test aşamasında gerçek zamanlı olarak konuşma sinyallerinin tanınması ile robot kolu önceden koordinatları belli olan nesneye yönlendirilmektedir. Bu yönelme işleminin yapılabilmesi için öncelikle bilgisayar yazılım arayüzü ile konuşma komutları gerçek zamanlı olarak tanınmakta ve tanınan komuta göre ilgili veriler RS232 seri iletişim protokolü kullanılarak robot kontrol kartına iletilmektedir. Ardından kontrol kartında herbir nesnenin yerinin bilgisini içeren mikrodenetleyici yardımı ile robotun servo motorları nesne konumuna

dođru y nelmektedir.  alıřma sonucunda AOVY i in dil modeli ortalama konuřma tanıma oranı %98,3 ve dil modelsiz %90,73 olarak bulunmuřtur. FDAA i in dil modelsiz ortalama konuřma tanıma oranı %89,48 ve dil modeli %97,1 olarak bulunmuřtur. YSA i in dil modeli ortalama konuřma tanıma oranı %91,1 ve dil modelsiz %81,9 bulunmuřtur.  alıřma sonucunda AOVY en y ksek ses tanıma oranına sahip olup diđer y ntemlerden daha iyi sonu lar vermiřtir.

Anahtar Kelimeler: Konuřma tanıma, robot kolu kontrol , FDAA, AOVY, YSA



ABSTRACT

Master of Science Thesis

ROBOT ARM CONTROL WITH VOICE COMMANDS

Ozan Fırat ÇIPLAK

Kirsehir Ahi Evran University

Institute of Science

Advanced Technologies Department

Supervisors: Asst. Prof. Dr. Serkan KESER

Recognition systems, which perform remote control of devices, have gained significant importance day by day. Speech, face, and fingerprint recognition systems seem to be the most frequently used recognition systems. Speech recognition systems can be used in real time in security systems, device control systems and dictation systems. In this study, the robot arms controlling by recognizing the real-time speech commands have been done. The Fisher Linear Discriminant Analysis (FLDA) and Discriminative Common Vector Approach (DCVA) of effective subspace classifiers have been used to recognize speech commands. One of these classifiers, DCVA was used for the first time in voice recognition.

For the training set, a total of 24 speech sentences have been created for four different objects which have got six different colors. The training set has been created as speaker-dependent. During the test phase, by recognizing the speech signals in real-time, the robot arm has been directed to the object whose coordinates have been previously determined. In order to perform this direction process, firstly speech commands are recognized in real time by the computer software interface and the relevant data are transmitted to the robot control card using the RS232 serial communication protocol according to the recognized command. Then, with the help of the microcontroller containing the information of the location of each object on the control card, the robot's servo motors are directed towards the object position. As a result of the study, the average speech recognition rate for DCVA with language model was 98.3% and without language model was 90,73%.

For the FLDA, the average speech recognition rate without language model was 89.48% and with language model was 97.1%. For ANN, the average speech recognition rate with language model was 91,1% and without language model 81,9%.

Keywords: Speech recognition, Robot arm controlling, FLDA, DCVA, ANN



1. GİRİŞ

Yerleşik yaşam yaklaşık 11000 yıl önce başladı. İnsanlık uzun süre küçük bilgi birikimleri ile ilerledi ve her küçük yeni bilgi insanlığı bir adım ileriye taşımıştır. Zaman içerisinde bilgi üretimi de ivme kazanmıştır. Rönesans'ın başlattığı aydınlanma çağı bilgi üretimini hızlandırmıştır. Bu bağlamda son 200 yıl özel bir önemdedir. Bilgisayarın bulunması ve kullanımı bilgi üretimi ve paylaşımını ayrıca hızlandırmıştır. Bilgisayar kullanımı özellikle otomasyon alanında inanılmaz bir gelişme sağlamıştır. Otomasyon ve robotik alanında bazı gelişmeler olsa da, bu alan halen emekleme aşamasındadır ve araştırılmayı bekleyen çokça husus vardır.

Konuşma tanıma sistemleri diğer pek çok değişik alanda kullanılmaktadır. Bu alanlar özellikle sesli komutlar ile ev içi cihazların kontrolünün sağlandığı akıllı evler, robotların ve taşıtların kontrolü, interaktif sesli cevap sistemleri, sesin yazıya dönüştürüldüğü dikte ettirme ve konuşmacı tanıma gibi birçok uygulama alanında kullanılmaktadırlar [1-4]. Konuşma tanımayı etkileyen pek çok faktör bulunmaktadır. Bu faktörler genel olarak ses sinyallerine eklenen gürültüler, ses kaynağının ses alıcısına olan uzaklığı, seslendirilen kelimelerin yanlış telafuzu, seslendirilen sözcüklerin hangi sınıflayıcılar ile sınıflandırıldığı, kullanılan ses veri tabanı büyüklüğü, kişi bağımlı ya da kişi bağımsız olarak tanıma yapılması olarak sayılabilir. Bu faktörlerden sınıflandırıcılar, ses tanıma oranlarının yüksek olmasında en büyük etkenlerden birini oluşturmaktadır. Sınıflandırıcıların tanıma oranlarına ise veri tabanı büyüklüğü ve kişi bağımlılık etkenleri büyük etki etmektedir. Literatürde en çok kullanılan ses tanıma sınıflayıcıları Dinamik Zaman Bükme (DZB), Saklı Markov Model (SMM), Yapay Sinir Ağları (YSA) veya altyüzay sınıflandırıcılar olarak sayılabilir. Bu sınıflandırıcılar içinde SMM, kişi bağımsız ve gerçek zamanlı konuşma tanıma sistemlerinde yaygın olarak kullanılmaktadır. Dil modellerini kullandığı için yüksek bir başarımlı orana sahiptir. Fakat SMM ile karşılaştırılabilir tanıma oranları veren sınıflandırıcıların kullanımı gün geçtikçe artmaktadır [5-7]. Bir alt uzay sınıflandırma yöntemi olan Ortak Vektör Yaklaşımı (OVY), sınırlı sayıda yalıtık kelime tanıma uygulamalarında yüksek tanıma oranları vermektedir [5, 8,9]. Yapılan çalışmalarda sınırlı sayıda kelime kullanılarak %95'in üzerinde başarımlar elde edilmiştir. Ayrıca OVY sınıflandırma yaparken pek çok sınıflayıcıya göre daha hızlı çalışmaktadır. OVY yöntemini temel alan ve özellikle yüz tanıma uygulamalarında kullanılan Ayırtedici Ortak Vektör Yaklaşımı (AOVY) ilk olarak Çevikalp (2005) tarafından

önerilmiştir [10]. AOVY yaklaşımı yüz tanıma uygulamalarında diğer alt uzay metotları olan Eigenface ve Fisherface sınıflandırıcılarına göre daha iyi sonuçlar vermektedir. Ayrıca tıpkı AOV gibi bu alt uzay metotlarına göre daha hızlı bir hesaplama süresine sahiptir. Fisherface yöntemi benzer şekilde genel olarak yüz tanıma çalışmalarında kullanılan ve doğrusal ayrım analizi (DAA) yöntemini temel alan bir alt uzay sınıflandırıcıdır.

Bu çalışmada yalıtık kelime tanıma işlemi kişi bağımlı olarak AOVY, FDAA ve YSA sınıflandırıcıları ile gerçekleştirilmiştir. Böylece görüntü sinyalleri yerine ses sinyalleri kullanıldığında elde edilecek tanıma başarımlarının incelenmesi amaçlanmıştır. Ses veritabanı sınırlı sayıda (10 adet) sözcükten oluşmaktadır. Eğitim ses veri tabanında bu sözcüklerin her biri için 50 ses sinyali kişi bağımlı olarak bilgisayar ortamında kaydedilmiştir. Ses sinyallerine MFCC uygulanarak öznitelik vektörleri elde edilmiştir. Eğitim aşamasında öncelikle ses veri tabanındaki her sözcüğe ait öznitelik vektörleri AOVY, FDAA ve YSA kullanılarak eğitilmiştir. Test aşamasında bir mikrofon aracılığı ile seslendirilen iki sözcüğün her biri ayrı ayrı bilgisayar ortamında yazılmış bir program aracılığıyla sınıflandırılmıştır. Bilgisayar aracılığıyla tanınan kelimelere göre bilgisayar ile RS-232 seri bağlantıya sahip robot kolu servo motor kontrol kartına sinyaller gönderilerek robot kol istenilen nesneye doğru yönlendirilmiştir. Sınıflandırma sonucunda elde edilen tanıma başarımlarını artırmak için bir çeşit dil modeli de geliştirilmiştir.

1.1. Araştırmanın Amacı ve Önemi

Günümüzde konuşma tanıma ile gerçekleştirilen sistemlerin sayısı gittikçe artmaktadır. Bu sistemler insan hayatını kolaylaştırmak için geliştirilmiştir. Özellikle uzaktan ses komutlarıyla aygıt kontrolü oldukça yaygınlaşmıştır. Konuşma tanıma için SMM, DZB, YSA vb.. gibi pek çok yöntem kullanılmaktadır. Özellikle hızlı tanıma gerçekleştirmek gerçek zamanlı konuşma tanıma sistemleri için çok hayati bir durumdur. Görüntü tanıma alanında kullanılan AOVY bu alanda yüksek tanıma başarısının yanı sıra diğer pek çok yöntemle göre hızlı tanıma gerçekleştirebilmesi ile öne çıkmaktadır. Ancak literatürde AOVY için ses sinyalleri ile ilgili bir çalışma yapılmamış olup, genel olarak ses sinyalleri ile Ortak Vektör Yaklaşımı (AOV) kullanılmıştır. Dolayısı ile ilk kez bu çalışmada ses tanıma için AOVY tanıma oranları incelenmiştir. Diğer bir alt uzay sınıflayıcı yöntemi FDAA, sınıflama hızı bakımından AOVY'den yavaş olmasına rağmen görüntü ve ses tanımada iyi sonuçlar vermektedir. Bu çalışmada, AOVY ve FDAA alt uzay sınıflandırıcılarının başarımları ses tanımada ilk kez

birlikte deęerlendirilmiřtir. Bylece kk veri tabanı ieren konuřma tanıma iřlemleri iin bu iki alt uzay sınıflayıcının kullanılıp kullanılmayacakları test edilmek istenmiřtir. Ayrıca YSA ile de test iřlemi gerekleřtirilerek  sınıflayıcı iin de bir karřılařtırma yapılmıřtır.

Elde edilen sonular neticesinde, gelecekte istenilen komutları yerine getirebilecek mobil robotlardan, engelli insanlara entegre edilmiř robot protezlerin kontrolne, ya da internet zerinden bir cihazın kontrolne kadar pek ok alanda, belirlenen metot (ya da metotlar) ile uygulamalar yapılabilir.

2. KAYNAK ARAŞTIRMASI-KONUŞMA TANIMA

2.1. Konuşma Tanımanın Tanımı ve Avantajları

İnsanların en yaygın olarak kullandığı iletişim aracı konuşmadır. İnsan – bilgisayar iletişimi ve veri girişi açısından, kullanıcının alışkın olduğu en yaygın iletişim aracını, yani doğal konuşma dilini kullanması bilgisayar kullanımını kolaylaştıracaktır. Bunun için bilgisayarın bir mikrofon vb. bir ses algılama cihazı ile insan sesini tanıması gerekmektedir. Konuşma tanıma, konuşmacı tarafından söylenen sözleri tanımlamak için işitsel sinyallerin analizi işlemidir. Konuşma tanıma sistemlerinin avantajları aşağıdaki verilmiştir.

2.1.1. Kullanım Kolaylığı

Konuşma tanıma sistemleri, veri giriş aracı olarak mikrofonu kullanır. Mikrofon günlük hayatta kullandığımız telefon bilgisayar tablet, televizyon gibi pek çok cihazda mevcuttur. Veri kaynağı, insanın alışkın olduğu özel çaba gerektirmeyen konuşma olduğundan veri girişi ve kullanımı oldukça kolaydır.

2.1.2. Veri Toplama Hızı

Konuşma tanıma sistemleri ile veri girişi, konuşma dilini kullandırması bakımından, elle veri girişine oranla daha hızlı olacaktır. Hızlı veri girişi ile veri tabanı daha hızlı oluşturulabilmektedir.

2.1.3. Uzaktan Veri İletme İmkânı

Sesin yayılma özelliğinden dolayı arada mesafeler olsa bile ses tanıma işlemi yapılabilir. Böylece verici ile alıcı arasında fiziksel bir donanım olmadan konuşma tanıma işlemi gerçekleştirilebilir.

2.2. Konuşma Tanıma Çeşitleri

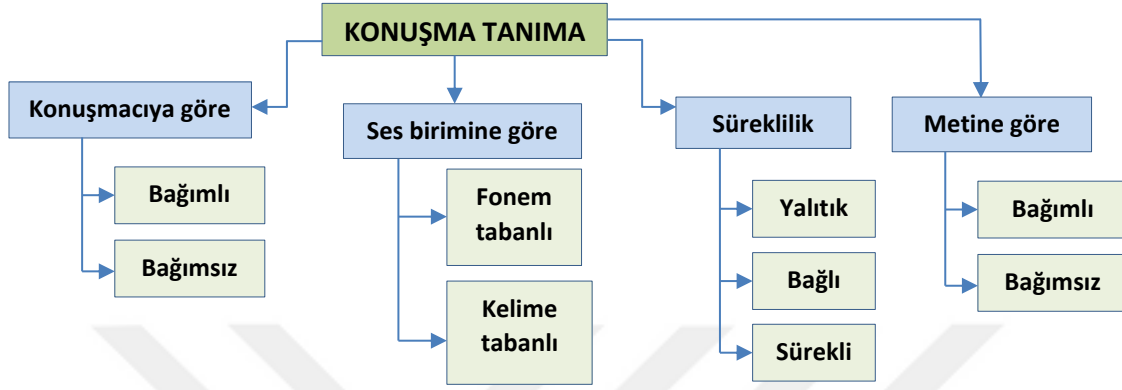
Konuşma tanıma, yeteneğine ve kullanımına bağlı olarak farklı türlere ayrılır [11].

Bu türler:

- Konuşmacı bağımlılığına göre bağımlı ve bağımsız tanıma,
- Temel alınan ses birimine göre fonem (ses birim) tabanlı ve kelime tabanlı tanıma,

- Sesin sürekliliğine göre izole, bağlı ve sürekli tanıma,
- Metin bağımlılığına göre bağımlı ve bağımsız konuşma tanıma olarak sınıflandırılır.

Aşağıdaki **Şekil 2.1.**'de konuşma tanımanın belirtilen türleri görülmektedir.



Şekil 2.1. Konuşma tanıma türleri

2.2.1. Konuşmacıya Göre Konuşma Tanıma

Bu türde sınıflandırma, kişiye bağımlı ve kişiden bağımsız olarak iki şekilde yapılmaktadır.

- Konuşmacı bağımlı
- Konuşmacıdan bağımsız

2.2.1.1. Konuşmacı Bağımlı Konuşma Tanıma

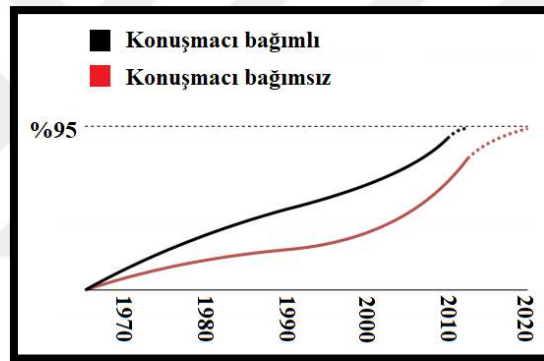
Konuşmacı bağımlı sistemler, belirli kullanıcı ya da kullanıcılar tarafından önceden sisteme tanımlanmış bir kelime ya da kelime grupları ile birlikte tanımlanır. Konuşmacı bağımlı sistemlerde, başka bir konuşmacı sesinin tanınması istenildiğinde, sistem üzerinde kayıtlı olan ve konuşma tanıma için kaynak olarak alınan verilerin güncellenmesi gereklidir [12]. Bu sistemler, yüksek komut sayımına sahiptir ve kelime tanıma için yüksek oranda doğruluk yüzdesi bulunmaktadır. Bu sistemin eksisi yalnızca sistemi eğitmiş olan kişiye doğru bir şekilde tepki vermesidir. Bu tür sistemlerde konuşmacıyı bir veya daha fazla kişi oluşturmaktadır. Tanımlanan her kişinin sesinin tanınması için referans şablonları bulunmalıdır [13].

2.2.1.2. Konuşmacı Bağımsız Konuşma Tanıma

Konuşmacı bağımsız sistemler, konuşmacı bağımlı sistemler gibi bir veri tabanı güncellemesine ihtiyaç duymazlar ve herhangi bir kişinin sesini tanıyabilirler. Konuşmacı bağımsız sistemlerin tanıma performansı, konuşmacı bağımlı sistemlere göre daha düşüktür. Fakat kullanım alanı göz önüne alındığında, zorluğuna rağmen konuşmacı bağımsız sistemler, konuşmacı bağımlı sistemlere göre bir adım öne çıkmaktadır [14].

2.2.1.3. Konuşmacı Bağımlı ve Bağımsız Tanıma Sistemlerinin Değerlendirilmesi

Geçmiş çalışmalar incelendiğinde konuşmacı bağımlı tanıma sistemlerinde başarı oranının konuşmacı bağımsız sistemlere göre daha yüksek seviyelerde olduğu görülmüştür [11]. Ancak Şekil 2.2.'den de görülebileceği gibi son yıllarda konuşmacı bağımlı ve konuşmacı bağımsız sistemlerin tanıma oranları arasındaki fark kapanmaktadır.



Şekil 2.2. Konuşmacı bağımlı ve bağımsız tanıma oranlarının yıllara göre değerleri [11]

2.2.2. Temel Alınan Ses Birimine Göre Konuşma Tanıma

Konuşma tanıma sistemleri temel alınan ses birimine göre iki gruba ayrılır:

- Fonem tabanlı sistemler
- Kelime tabanlı sistemler

Bu ayrım ile sistemin uygulanmasında kullanılan teknikler ve sistemin kullanıldığı alanlar değişir. Fonem tabanlı konuşma tanıma sistemleri, fonemlerin (harf/hece) en küçük ses birimi olarak kabul edildiği sistemlerdir. Kelime tabanlı konuşma tanıma sistemleri ise tanıma için ön görülen en küçük ses birimin kelime kabul edildiği sistemlerdir [13].

2.2.2.1. Fonem Tabanlı Konuşma Tanıma

Konuşmacı tarafından söylenen, birbirinden farklı özellikler içeren anlamlı en küçük ses parçalarına ses birim (fonem) adı verilmektedir. Tanıma için ses birimler ikili, üçlü ya da hece olarak alınabilirler. Fonem tabanlı konuşma tanıma sisteminin en büyük avantajı bir dili oluşturan tüm fonemlerden oluşan bir veri tabanı oluşturularak o dildeki tüm kelimelerin tanınabilmesini sağlamasıdır [15]. Ancak kelimeleri oluşturan ardışık fonemlerin sınırları tam olarak kestirilemediğinden ötürü tek ses birime göre tasarlanan sistemlerde tanıma oranları düşük olabilmektedir. Buna karşın az olan fonem sayısından dolayı, hızlı sonuç üretme olanağı sayesinde, hataları en aza indirme amaçlı güncellemeler kolaylıkla gerçekleştirilmektedir [16].

2.2.2.2. Kelime Tabanlı Konuşma Tanıma

Konuşma tanıma için gerekli olan en küçük unsurun kelime olarak kabul edildiği sistemdir. Uygulama anlamında yüksek tanıma oranı ile birlikte kelime tabanlı ses tanıma sistemlerinde hafıza gereksinimleri fazladır. Az kelime içeren tanıma sistemleri için genel olarak yüksek tanıma oranlarına erişilmektedir.

2.2.3. Sesin Sürekliliğine Göre Konuşma Tanıma

Konuşma tanıma sistemleri sesin sürekliliğine göre üç gruba ayrılır [11]:

- İzole konuşma tanıma
- Bağlı konuşma tanıma
- Sürekli konuşma tanıma.

Bu ayırım ile sistemin uygulanmasında kullanılan teknikler ve sistemin kullanıldığı alanlar değişir.

2.2.3.1. İzole Konuşma Tanıma

İzole yani ayrışık konuşma tanıma sistemi, kısa aralıklarla seslendirilen kelimelerin tanınması işlemidir. İzole kelime tanıma sistemlerinde konuşmacı tarafından seslendirilen sözcükler arasında belirli süre boşluk bırakılmalıdır. Boşluklar arasında seslendirilen kelimeler birbirinden bağımsız olarak sistem tarafından tanınır. Kelimeler gerçek zamanlı olarak tanınabilir.

2.2.3.2. Baęlı Konuşma Tanıma

Baęlı konuşma tanıma sisteminde konuşmacı seslendirdięi sözcükler aralarında kısa boşluklar bırakmalıdır. Bu sistemlerden sonraki evre konuşmacı tarafından seslendirilen sözcüklerin aralarında beklemedięi sürekli ses tanıma sistemidir [17].

2.2.3.3. Sürekli Konuşma Tanıma

Sürekli konuşma tanıma sistemi kelimeler arasında ara verilmeden tanımayı amaçlar. Sürekli konuşma tanıma sistemi içerisinde söylenen kelimenin ne zaman söylendięi ya da ne zaman bitirildięi önemli değildir. Kelimeler gerçek zamanlı olarak tanınırlar. Bu sistem içerisinde, konuşma esnasında ki telaffuzlar ve deęişkenler başlıca sorunlardır [18]. Sürekli konuşma tanımının en büyük avantajı, konuşmacı beklemeden doğal bir biçimde konuşur.

2.2.4. Metne Göre Konuşma Tanıma

Konuşma tanıma sistemleri, metne baęlılığı baz alınarak iki gruba ayrılır:

- Metne baęlı sistemler.
- Metinden bağımsız sistemler.

Fakat metin bağımsız konuşma tanıma sistemleri test aşamasında eğitim setinden türetilen söz dizileri kombinasyonlarını da tahmin edebilmektedir.

2.2.4.1. Metne Baęımlı Konuşma Tanıma

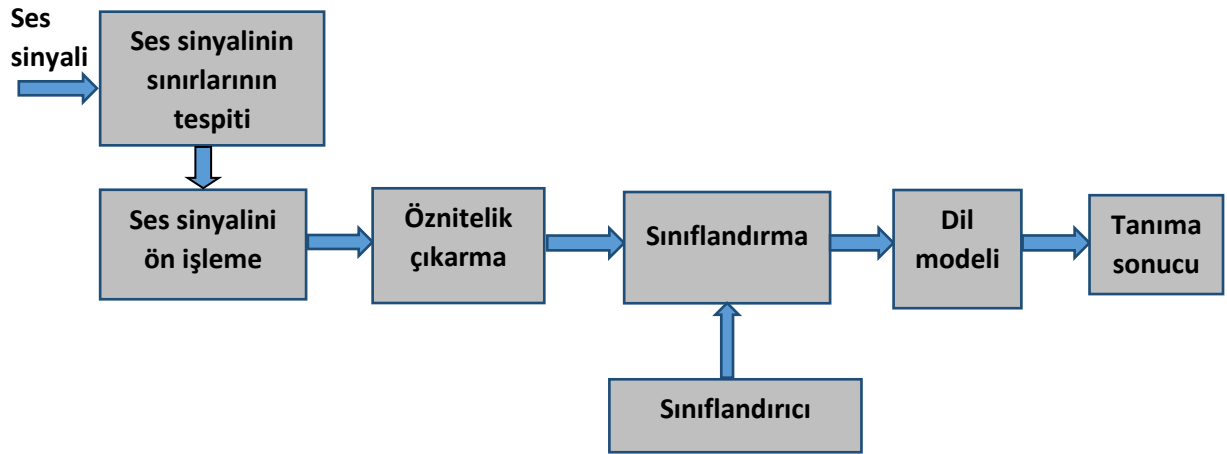
Metne dayalı konuşma tanıma sistemlerinde test aşaması da eğitim verisi gibi aynı metin kütüphanesine baęlıdır. Bu tanıma modelinde, sistem eğitim aşamasında kullanılan kelimelerin farklı seslendirilmeleri ile test edilirler [17].

2.2.4.2. Metinden Bağımsız Konuşma Tanıma

Metinden bağımsız konuşma tanıma sistemleri, modelin eğitimi esnasında kullanılan sözcüklerin dışındaki kelimelerde yanıt verebilmektedir. Bu sistemlerde örneğin sistem, “kırk” ve “dört” kelimelerini tanıyorsa, o halde “kırk dört” kelimesini de tanıyabilmelidir [14].

2.3. Konuşma Tanımanın Aşamaları

Konuşma tanımda öncelikle eğitim ses sinyalleri oluşturulmalı ve kaydedilmelidir. Ardından kaydedilen ses sinyallerinin ön işlenmesi ve öz niteliklerinin çıkarılıp kaydedilmesi gerekir. Böylece eğitim aşaması bitmiş olur. Test aşamasında ise seslendirilen sözcükler için benzer biçimde öz nitelik vektörleri elde edilir ve bir sınıflandırıcı vasıtası ile eğitim kümesindeki en uygun öz niteliğe göre sınıflama gerçekleştirilir. Aşağıdaki Şekil 2.3’de konuşma tanımanın aşamaları verilmiştir.



Şekil 2.3. Konuşma tanıma aşamaları

2.3.1. Konuşma Sinyali Sınırlarının Belirlenmesi

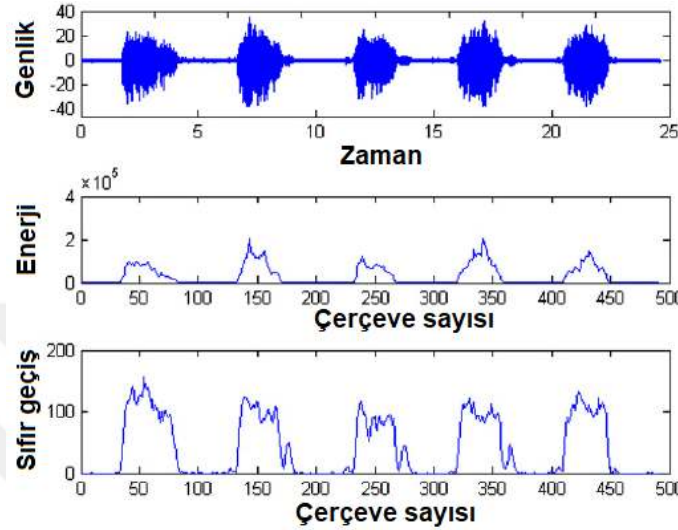
Temel prensip olarak konuşma tanımda ses sinyalinin gereksiz kısımlarının alınmayıp sadece kelime sinyalinin alınması gerekir. Bunun için ses sinyalinin yalnızca gürültülü bölümlerinden belirlenen eşik değeri ile karşılaştırılmasıyla sesli (kelime) bölümü ve gürültülü bölümü tespit edilebilir. Bunun için sıfır geçiş oranı (SGO) ve kısa süreli enerji (KSE) değerleri birlikte kullanılarak başarılı biçimde sesin sınırlarının tespiti gerçekleştirilebilmektedir [19].

2.3.1.1. Kısa Süreli Enerji (KSE)

Konuşma sinyallerinde zamana bağlı olarak konuşmanın genliği değişiklik göstermektedir. Konuşmanın olduğu bölgelerde KSE, konuşmanın olmadığı bölgelere göre daha yüksektir. Bu sayede belli bir enerji eşik değeri kullanarak kelime sinyalinin sınırları belirlenebilir [20]. Genel olarak 20ms gibi küçük pencere süreleri kullanılarak hesaplanırlar.

2.3.1.2. Sıfır Geçiş Oranı

Sıfır geçiş oranı, sinyalin bir çerçevesi boyunca değiştirdiği işaret oranıdır. Diğer bir deyişle, bir çerçeve boyunca sinyalin genlik değerlerinin kaç kez pozitiften negatife (ya da negatiften pozitifte) değiştiğiyle alakalıdır. Bu değişim sayısı çerçeve uzunluğuna bölünerek sıfır geçiş oranı bulunur [20].



Şekil 2.4. Sinyalin sıfır geçiş ve kısa süreli enerji grafiği

2.3.2. Konuşma sinyalinin ön işlenmesi

Özellikle ses sinyalinin yüksek frekanslı kısımlarında gürültünün daha fazla olduğu bilinmektedir. Bu gürültü etkisini azaltmak için ses sinyalinin yüksek frekanslarını bastıran bir filtre uygulanır. Zaman düzleminde bu işlem aşağıdaki gibi gösterilebilir [15];

$$y[n] = x[n] - \alpha x[n], \quad (2.1)$$

Burada $0,95 < \alpha < 0,99$, $x[n]$ ses sinyali, $y[n]$ ise filtre uygulanmış çıkış sinyalidir.

2.3.3. Öznitelik Vektörlerinin Elde Edilmesi

Ses sinyalindeki akustik özelliklerin çıkarılmasına öznitelik çıkarma denilir. Böylece, ses sinyalindeki konuşmacıyı tanımlayan akustik özellikler belirlenerek ses sinyali uzunluğundan daha düşük boyutlu öznitelik vektörleri elde edilir. Bu sayede tanıma işlemi hesaplama açısından kolaylaşır ve daha verimli olur. Literatürde ses tanıma için öznitelik vektörleri

olarak Mel frekans kepsral katsayıları (MFKK) katsayıları yaygın olarak kullanılmaktadır [21,22]. MFKK katsayıları aşağıdaki adımlar ile elde edilmektedir.

Öznitelik çıkarma aşağıdaki aşamalardan oluşur:

- Ses sinyalinin ön vurgulanması (işlenmesi)
- Pencereleme (Windowing) ve çerçeveleme (Framing)
- Hızlı Fourier Dönüşümü (HFD)
- Mel-Frekansı Saptırması (MFS)

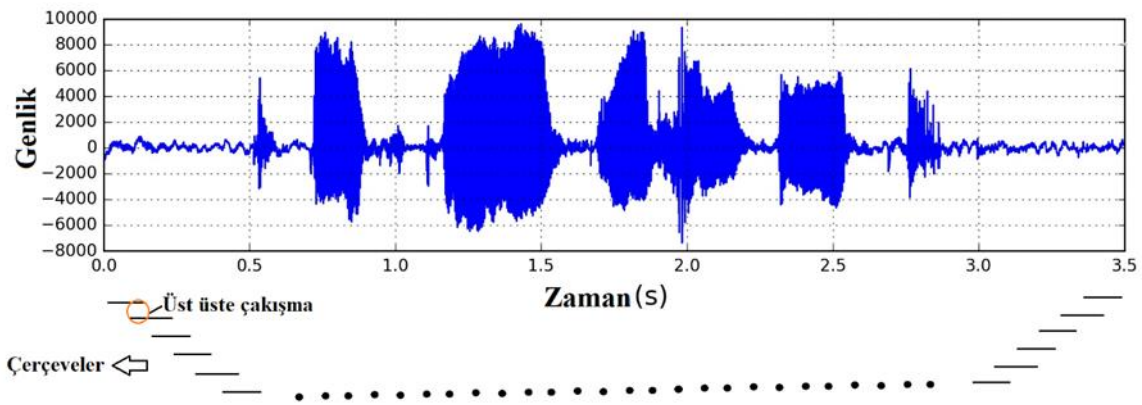
Aşağıdaki Şekil 2.5’de MFCC katsayılarının elde edilme aşamaları verilmiştir. İlk adım olan ses sinyallerinin ön işlenmesi adımı önceden bahsedildiği gibi gerçekleştirilir. Ardından çerçeveleme ve pencereleme adımı gerçekleştirilir. Sonra her bir çerçeveye HFD uygulanır. Son olarakta MFS uygulanır ve ters Ayırık Kosinüs Dönüşümü (AKD) ile zaman ekseninde MFKK katsayıları elde edilir.



Şekil 2.5. MFKK katsayılarının elde edilme aşamaları

2.3.3.1. Çerçeveleme ve Pencereleme

Yapılan araştırmalar ses sinyalinin özellikleri sadece küçük zaman aralıklarında kararlı kaldığını göstermektedir [23]. Bu nedenle ses sinyallerinin kısa zaman aralıklarında işlenmesi gerekmektedir. Sinyaller genellikle en etkili oldukları zaman aralıkları olan 20-30 milisaniye arasındaki uzunluklarda çerçevelere bölünürler. Her çerçeve kendisinden bir önceki çerçevenin belli bir kısmını örter. Örtme yönteminin asıl amacı bir çerçeveden diğerine geçişin daha yumuşak olmasını sağlamaktır.

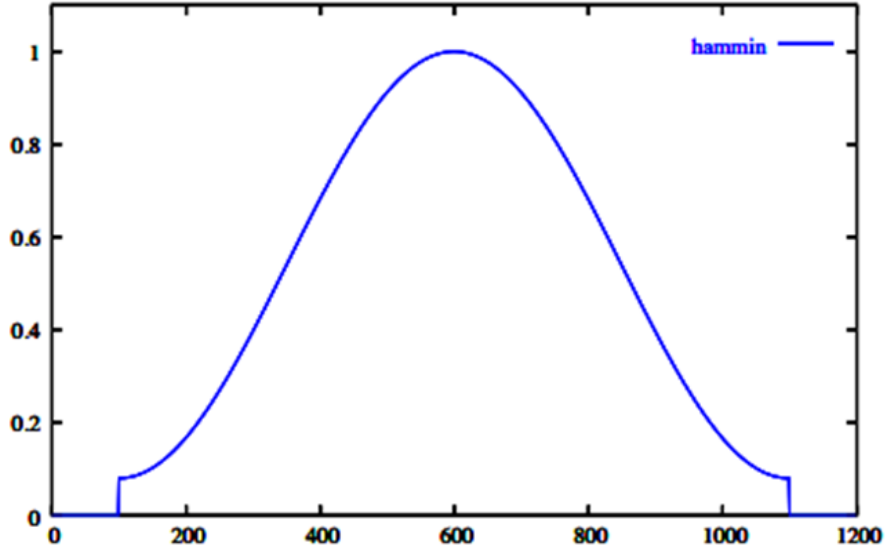


Şekil 2.6. Çerçeveleme örneği

Ses sinyalinin çerçevenmesinden sonra her çerçeveye pencereleme işlemi uygulanır. Bu işlem her çerçeveye uygulanarak çerçeve başı ve sonundaki süreksizliklerin ortadan kaldırılması ve spektral bozulmanın engellenmesi sağlanır. Genel olarak en çok kullanılan pencereleme fonksiyonlarından birisi Hamming fonksiyonudur [24].

$$w[n] = (1 - \alpha) - \alpha \cos\left(\frac{2\pi n}{L-1}\right), \quad (2.2)$$

Burada L pencere uzunluğu ve $\alpha = 0,46164$ değerini alır. Aşağıdaki şekil 7’de Hamming pencerenin yapısı verilmiştir.



Şekil 2.7. Hamming pencere yapısı

Kullanılan diğer pencereleme fonksiyonları şu şekilde verilebilir:

- Barlett Pencere,
- Hanning Pencere,
- Blackman Pencere,
- Kaiser Pencere.

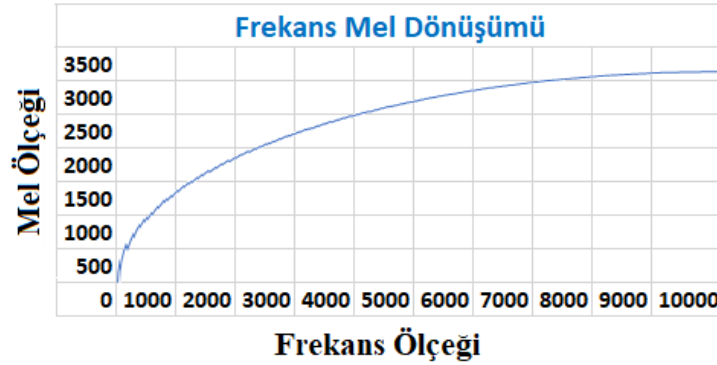
2.3.3.2. Hızlı Fourier Transformu (HFD)

Pencerelemeden sonra ses sinyalinin her çerçevesi için kısa zaman- Fourier dönüşümü (Short Time Fourier Transform) uygulanarak frekans düzlemindeki genlik değerleri elde edilir. Bu genlik değerleri insan duyusunun iyi algıladığı 0-10 kHz frekans değer aralığı için bulunur.

2.3.3.3. Mel filtreleme ve MFK Katsayılarının Elde Edilmesi

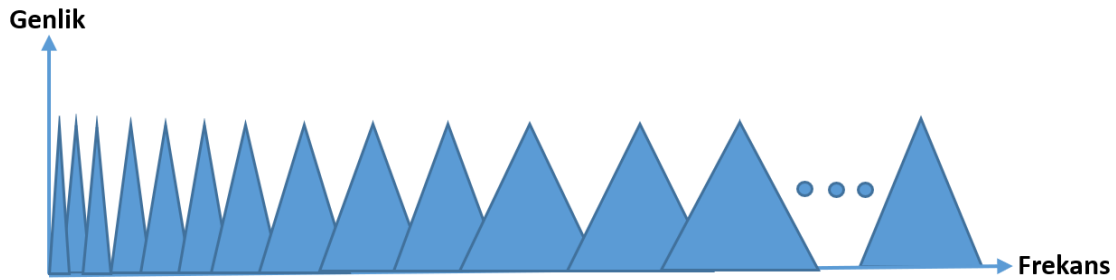
Ses verisinin spektrum özelliğini en iyi şekilde temsil edecek parametreleri bulmak için pek çok yöntem bulunmaktadır. Lineer öngörü katsayıları (LÖK) ve MFKK bu yöntemlere örnek gösterilebilir. İnsan duyusu ses sinyalinin tüm frekanslarını tam olarak doğrusal olarak algılayamaz. İnsan duyusu 1000 Hz'e kadar doğrusal, 1000 Hz'in üstündeki frekansları ise logaritmik olarak algılar [25]. Buradan yola çıkarak çeşitli frekans ölçekleri geliştirilmiştir.

Aşağıdaki Şekil 2.8’de yatay ekseninde klasik frekans düzlemi, dikey ekseninde ise bu frekanslara karşılık gelen Mel frekans değerleri gösterilmiştir.



Şekil 2.8. Frekans-Mel Dönüşümü

Elde edilen HFD spektrumuna aşağıdaki Şekil 2.9’da verilen Mel-Frekans değerleri ile bağıntılı üçgensel bant geçiren filtre bankasından oluşan filtre uygulanır.



Şekil 2.9. Mel bank filtre yapısı

Filtrelerin logaritmik enerji değerleri üzerine ters AKD dönüşümü uygulanarak MFKK katsayıları hesaplanır. MFKK ses tanıma alanında en çok kullanılan öznelik çıkarma yöntemlerinden birisidir. Çalışmalarda genel olarak MFKK katsayıları her çerçeve için 12 adet veya biraz daha fazla seçilir. Her çerçeve için elde edilen bu vektörlere öznelik vektörleri denilir.

2.4. Konuşma Tanımda Kullanılan Temel Sınıflandırıcılar

Yukarıda belirtildiği gibi eğitim aşamasında her konuşma sinyali sınıfına ait öznelikler kaydedildikten sonra test aşamasına geçilir. Test aşamasında test sinyaline ait öznelik vektörü ile eğitim kümesindeki her sınıfa ait öznelikler bir sınıflandırıcı ile karşılaştırılarak

test sinyali bir sınıfa atanır. Konuşma tanımada kullanılan belli başlı sınıflandırıcılar aşağıdaki gibidir.

- Yapay Sinir Ağları (YSA),
- Dinamik Zaman Bükme (DZE),
- Alt uzay sınıflandırıcılar (FDAA, OVY, AOVY vb...),
- Destek Vektör Makinesi (DVM)
- Saklı Markov Modelleri (SMM),

2.4.1. Dinamik Zaman Bükmesi (DZB)

Dinamik zaman bükmesi hız açısından değişiklik gösterebilen iki zaman aralığı arasındaki benzerliği ölçmek için oluşturulan bir algoritmadır [26]. Ses tanıma alanında DZB farklı ses hızlarıyla baş etmek için otomatik konuşma tanıma uygulamalarında kullanılmaktadır. Belli kelimenin seslendirilmesinde kelime uzun ya da kısa olarak seslendirilebilir. Fakat DZB algoritması ile bu seslendirmeler belli zaman aralığında yayılarak veya daraltılarak birbirlerine yaklaştırılır. Böylece sistem üzerinde kayıtlı bulunan kelime şablonuyla seslendirilen kelimenin zamanının örtüşerek karşılaştırılması yapılabilir [27].

2.4.2. Yapay Sinir Ağları (YSA)

Yapay sinir ağları (YSA) bilgiyi sınıflandırma ve yorumlama ihtiyacının doğduğu problemlerin çözümlerinde kullanılır. Ayrıca ses tanıma esnasında özellik vektörleri çıkarılan ses sinyallerini tanıma işlemi, YSA kullanılarak gerçekleştirilebilir. YSA biyolojik olarak sinirsel ağlardan esinlenerek öğrenme sürecinin matematiksel olarak modellenmesini içeren bir yöntemdir. Bu yüzden, çalışmalar ilk olarak beyni oluşturan nöronların modellenmesi ve bilgisayar sistemlerinde uygulanması ile başlamış, daha sonraları bilgisayar sistemlerinin gelişmesiyle birçok alanda kullanılır hale gelmiştir [28].

YSA'lar eğitim esnasında kullanılan sayısal bilgilerden makina öğrenmesi gerçekleştirebilirler. Böylece olayları öğrenerek benzer veriler karşısında doğru sonuçlar verebilirler. Örüntü (pattern) ilişkilendirme ve sınıflandırma yapabilirler. Kendi kendine öğrenebilme ve organize etme yetenekleri vardır ve kendi kendilerini eğitebilirler [29].

2.4.2.1. YSA'ların Kullanıldığı Alanlar

Yapay sinir ağları; sınıflandırma, modelleme ve tahmin uygulamaları olmak üzere, pek çok alanda kullanılmaktadır. Bunlardan başlıcaları;

- Sınıflandırma,
- İlişkilendirme veya örüntü eşleştirme,
- Sinyal filtreleme,
- Veri sıkıştırma,
- Örüntü tanıma,
- Sinyal işleme,
- Optimizasyon,
- Kontrol

uygulamaları gösterilebilir.

2.4.2.2. Yapılarına Göre Yapay Sinir Ağları

Yapay sinir ağları içerdiği nöronların birbirine bağlantı şekline göre ileri ve geri beslemeli olarak ikiye ayrılır.

a) İleri Beslemeli Ağlar: İleri beslemeli ağlarda nöronlar girişten çıkışa doğru düzenli katmanlar şeklindedir. Bir katmandan sadece kendinden sonraki katmanlara bağ bulunmaktadır. Yapay sinir ağına gelen bilgiler giriş katmanına daha sonra sırasıyla ara katmanlardan ve çıkış katmanından işlenerek geçer ve daha sonra çıkışa aktarılır [29].

b) Geri Beslemeli Yapay Sinir Ağları: Geri beslemeli yapay sinir ağlarında ileri beslemeli olanların aksine bir hücrenin çıktısı sadece kendinden sonra gelen hücrenin katmanına girdi olarak verilmez. Kendinden önceki katmanda veya kendi katmanında bulunan herhangi bir hücreye de girdi olarak bağlanabilir [29].

2.4.2.3. Yapay Sinir Ağıнын Yapısı

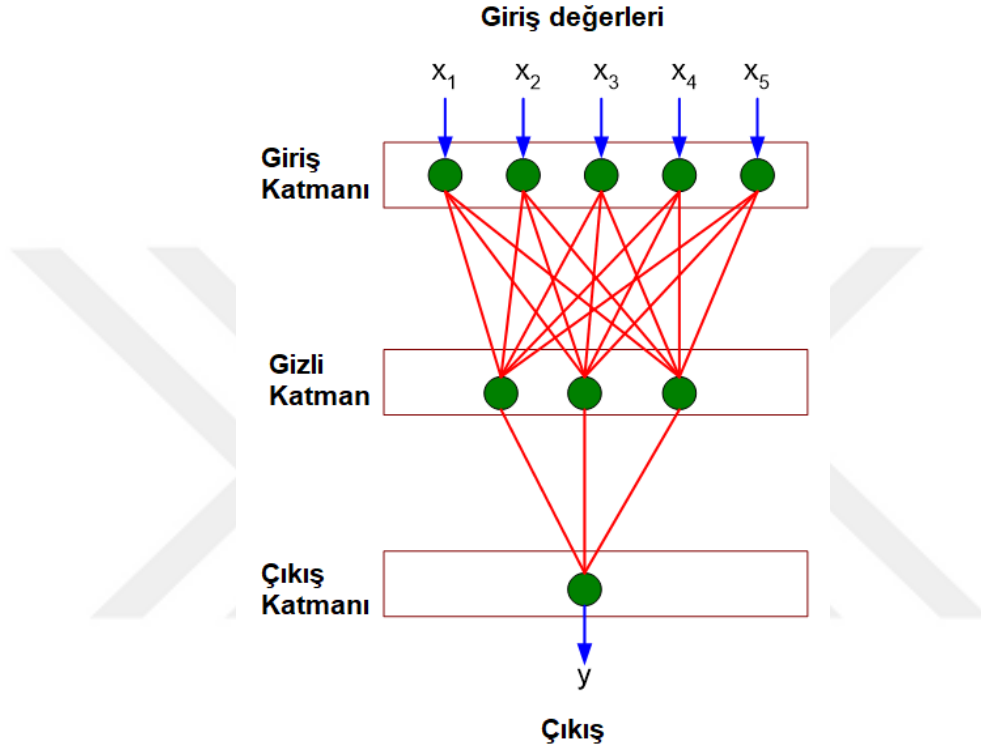
Biyolojik nöronlarda olduğu gibi yapay nöronların da giriş sinyallerini aldıkları, bu sinyalleri toplayıp işledikleri ve çıktıları ilettikleri bölümleri bulunmaktadır [29]. Bir yapay sinir hücresi beş bölümden oluşmaktadır;

- Girdiler
- Ağırlıklar
- Toplama Fonksiyonu (Birleştirme Fonksiyonu)

- Aktivasyon fonksiyonu
- Çıktılar

Ayrıca yapay sinir ağı üç ana katmanda incelenir;

- Giriş Katmanı,
- Ara (Gizli) Katmanlar ve
- Çıkış Katmanı.



Şekil 2.10. Yapay sinir ağı katmanları

- Giriş Katmanı:** Yapay sinir ağına veri girdilerinin geldiği katmandır.
- Ara Katman:** Giriş katmanından çıkan bilgiler ara katmana gelir. Ara katman sayısı ağlara göre değişebilir, yani ara katmandaki nöron sayıları giriş ve çıkış sayısından bağımsızdır. Ara katmanlardaki nöronların sayısının artması hesaplama karmaşıklığını arttırabilmektedir. Ancak böyle bir ağ yapısı yapay sinir ağının daha karmaşık problemlerin çözümünde de kullanılabilmesini sağlayabilmektedir.
- Çıkış Katmanı:** Ara katmanlardan gelen bilgileri işleyerek ağın çıktılarını üreten katmandır. Geri beslemeli ağlarda bu katmanda üretilen çıktı kullanılarak ağın yeni ağırlık değerleri hesaplanır.

2.4.2.4. Yapay Sinir Ağlarının Eğitilmesi

Yapay sinir ağlarının kullandığı bazı öğrenme algoritmalarıyla hata miktarı azaltılıp gerçek çıkışa yaklaşılmaya çalışılır. Bu hata azaltma süresince yapay sinir ağının ağırlıkları yenilenmektedir. Böylece ağırlıklar her bir döngüde yenilenerek hata en aza indirgenmeye çalışılır. Bu amaca ulaşmanın ölçütü de yine dışarıdan verilen verilerin tutarlılığıdır. Ağırlıkların sürekli yenilenerek istenilen sonuca ulaşılan kadar geçen zamana öğrenme adı verilir [30]. Yapay sinir ağı öğrendikten sonra daha önce verilmeyen girişler verilip, sinir ağı çıkışıyla gerçek çıkış değerleri incelenir. Eğer yeni verilen örnekler de doğru biçimde sonuçlar üretiyorsa sinir ağının problemi düzgün şekilde öğrendiği söylenebilmektedir.

2.4.3. Saklı Markov Modelleri (SMM)

Saklı Markov modeli, gözlem dizilerinin bilinmesine karşın temel durum dizileri bilinmediğinden saklı ibaresi ile “Saklı Markov Model” olarak adlandırılır [27]. HMM ile istatistiksel süreç verileri çok iyi biçimde tanımlanabilir. HMM ile geçmişten günümüze birçok alanda çalışma yapılmıştır [31,32]. HMM, konuşma tanımda, yüz tanımda, vücut hareketleri tanımda, el yazısı tanımda, kriptanalizde, protein yapısı ve DNA dizilimlerinde, yaygın olarak kullanılmıştır [33]. Ayrıca Markov modeller, sistem olasılıksal dağılıma bağlı olarak kendi durumundan başka bir duruma geçebilir veya aynı durumda kalabilir. Markov modellerde, bulunan durumdan meydana gelen olasılıklar, geçiş olasılıkları olarak isimlendirilir. Bununla birlikte HMM’de duruma bağlı olan geçişler görülebilir. SMM özellikle fonem tabanlı, kişi bağımsız, sürekli konuşma tanımda yüksek başarımlıdır [34].

2.4.4. Alt Uzak Sınıflandırıcılar

2.4.4.1. Ortak Vektör Yaklaşımı (AOV) ve Ayırt Edici Ortak Vektör Yaklaşımı (AOVY)

Ortak vektör yaklaşımı (OVY) ses tanıma ve görüntü tanıma uygulamalarında kullanılan bir alt uzak sınıflama yöntemidir [19,5]. Bu yöntem ile her sınıfa ait değişmez özellikleri taşıyan bir vektör elde edilir ve bu vektör “ortak vektör” olarak isimlendirilir [5]. Ayırt edici Ortak Vektör Yaklaşımı (AOVY) ise elde edilen ortak vektörlerin birbirlerine göre dağılımlarını en büyükleme için iz düşüm vektör kümesi kullanılmaktadır. Eğer eğitim kümesinde her biri k örnek olan c farklı sınıf varsa bu durumda eğitim setinde toplam $m=kc$ adet örnek olacaktır. m

ses komut sınıfına ait vektör sayısını, n ise her bir vektör boyutunu göstermek üzere, OVY hem yeterli veri durumu ($m \geq n$), hem de yetersiz veri durumları ($m < n$) için uygulanabilir [19]. Aynı durum AOVY içinde geçerlidir. Sınıfı i olan r 'inci sinyal örneğini n -boyutlu uzayda x_j^i ile gösterirsek, sınıflar içi dağılım matrisi S_w aşağıdaki gibi verilir:

$$S_w = \sum_{i=1}^c \sum_{r=1}^k \left((x_r^i - \mu_i)(x_r^i - \mu_i)^T \right) \quad (2.3)$$

Burada, μ_i i 'nci sınıfa ait ortalama vektörü göstermektedir. Farklılık alt uzayı B ve farksızlık alt uzayı $B^\perp$ olmak üzere birbirine dik iki alt uzaya ayrılır. Farksızlık alt uzayı $B^\perp$, S_w matrisinin sıfır öz değerlerine karşılık gelen öz vektörler tarafından gerilir. P ve $\bar{P}$ matrisleri sırasıyla B ve $B^\perp$ uzaylarının iz düşüm matrisleri olarak alınırsa, eğitim setindeki örneklerin $B^\perp$ uzayındaki iz düşümleri aşağıdaki gibi olacaktır.

$$x_{com}^i = x_r^i - Px_r^i = \bar{P}x_r^i, r=1,2,\dots,N, i=1,2,\dots,c \quad (2.4)$$

Aşağıdaki eşitlik 2.5'de belirtilen S_{com} ortak vektörlere ait saçılım matrisi olup, aşağıdaki gibi bulunur,

$$S_{com} = \sum_{i=1}^c (x_{com}^i - \mu_{com})(x_{com}^i - \mu_{com})^T \quad (2.5)$$

Bu eşitlikte μ_{com} ortak vektörlere ait ortalama vektörünü ifade etmektedir. Bu durumda S_{com} matrisinin sıfırdan farklı öz değerlerine karşılık gelen öz vektörler, en uygun iz düşüm vektörlerini verir. Bu vektörlerin sayısı $r \leq c - 1$, S_{com} matrisinin kertesine (rankına) eşittir. En uygun iz düşüm vektörleri üzerindeki iz düşüm katsayılarından oluşan öznitelik vektörleri aynı olup aşağıdaki gibi bulunur;

$$\Omega_i = [< x_m^i, w_i > \dots < x_m^i, w_r >] \quad (2.6)$$

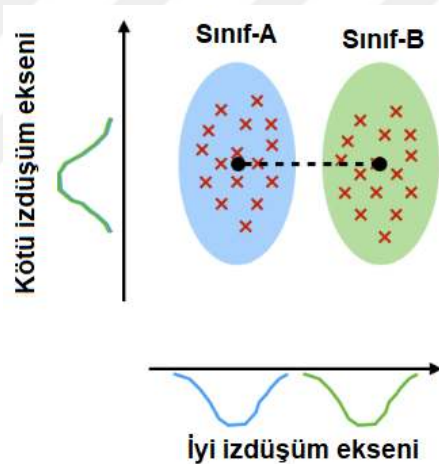
Bu vektörler, “ayırt edici ortak vektörler” olarak adlandırılır. Test aşamasında ses sinyallerinin ayırt edebilmesi için öncelikle bu test sinyaline ait öznitelik vektörleri aşağıdaki eşitlikle bulunur:

$$\Omega_{test} = W^T x_{test} \quad (2.7)$$

Daha sonra Ω_{test} ile eğitim setindeki sınıflara ait ayırt edici ortak vektörlerin arasındaki Öklid uzaklığına bakılır. Test ses sinyali, en küçük uzaklığı veren sınıfa atanır. Ω_{test} her sınıf için tek bir öznelik vektörü ile karşılaştırıldığından tanıma oldukça hızlı gerçekleştirilebilmektedir.

2.4.4.2. Fisher Doğrusal Ayrım Analizi (FDA)

Bu yöntem iz düşüm yönleri, aynı sınıfa ait örneklerin birbirine olan uzaklıklarını en büyükleyecek şekilde belirlenir [35]. FDA ile elde edilen alt uzay, gerek sınıflandırma gerekse boyut indirgeme amaçlı olarak kullanılabilir. Bu yöntem ile en çok sınıf sayısı-1 boyutlu alt uzay dönüşümü elde edilir. Dönüşmüş koordinat sisteminin eksenleri, ayırmama önemine göre sıralanabilir. Sınıflandırma kuralı olarak ‘en yakın sınıf ortalaması’ kullanıldığında, doğrusal karar sınırları elde edilir. Aşağıda A ve B sınıfları için iyi ve kötü iz düşürülme biçimleri Şekil 2.11’de verilmiştir.



Şekil 2.11. İki sınıfın iyi ve kötü biçimde iz düşürülmesi

FDA, Doğrusal Ayrım Analizinden (DAA) türetilmiş bir alt uzay sınıflama metodudur. Bu metot kullanılarak sınıflar arası ve sınıflar içi dağılım oranını en büyükleyen bir dik vektör kümesi (W) bulunur. Bu iz düşürülme biçimi Şekil 2.11’deki iyi iz düşürülmeye denk gelmektedir. Burada sınıflar içi dağılım matrisi aşağıdaki gibi bulunur,

$$S_W = \sum_{i=1}^C \sum_{m=1}^N (x_m^i - \mu_i)(x_m^i - \mu_i)^T \quad (2.8)$$

Sınıflar arası dağılım matrisi ise aşağıdaki gibidir,

$$S_B = \sum_{i=1}^c N(\mu_i - \mu)(\mu_i - \mu)^T \quad (2.9)$$

Burada N sınıfı oluşturan toplam örnek sayısıdır. μ_i sınıf ortalaması ve μ tüm sınıfların ortalamasıdır. En uygun taban vektörleri ($\mathbf{W}_{\text{opt}}$) aşağıdaki gibi bulunur.

$$\mathbf{W}_{\text{opt}} = \underset{\mathbf{W}}{\text{argmax}} \frac{|\mathbf{W}^T \mathbf{S}_B \mathbf{W}|}{|\mathbf{W}^T \mathbf{S}_W \mathbf{W}|} = [\mathbf{w}_1 \ \mathbf{w}_2 \ \dots \ \mathbf{w}_m] \quad (2.10)$$

Burada $m=c-1$ olmaktadır. Aşağıdaki gibi belirtilen denklemde $\mathbf{S}_W^{-1} \mathbf{S}_B$ çarpım sonucunda oluşan matrisin en büyük öz değerlerine karşılık gelen $c-1$ adet öz vektör en uygun taban vektörünü ($\mathbf{W}_{\text{opt}}$) vermektedir.

$$\mathbf{S}_B \mathbf{w}_i = \lambda_i \mathbf{S}_W \mathbf{w}_i, \quad i=1,2,\dots,m \quad (2.11)$$

Tanıma problemlerinde, sınıf içi dağılım matrisi $\mathbf{S}_W$ 'nin genellikle tekil olması problemiyle karşılaşılır. Bu problemin üstesinden gelmek için FDA yöntemi ses sinyali kümesini daha düşük boyutlu bir uzaya iz düşürür. Bunun için öncelikle PCA kullanılarak özellik uzayı boyutu $N - c$ 'ye düşürülür. Ardından $N-c$ boyutlu uzaya standart FDA uygulanarak boyutu $c - 1$ 'e düşürülür ve denklem $\mathbf{x}$ uygulanarak $\mathbf{W}_{\text{opt}}$ matrisi bulunur. Eğitim kümesindeki her sınıfa ait öznitelik vektörleri $\mathbf{W}_{\text{opt}}$ kullanılarak optimum uzaya iz düşürülür. Test aşamasında da test sinyali $\mathbf{W}_{\text{opt}}$ kullanılarak iz düşürülüp $\mathbf{\Omega}_{\text{test}}$ bulunur. Ardından $\mathbf{\Omega}_{\text{test}}$ ile eğitim aşamasında sınıflara ait iz düşürülmüş vektörler arasındaki en küçük öklit mesafeyi veren sınıfa atama yapılır.

Bir sınıflandırıcının başarımını ölçmek için sınıflandırılacak data'lara ait belirli sayıda örnek içeren 'veri tabanı' öncelikle 'eğitim' ve 'test' kümesi şeklinde ikiye ayrılmalıdır. Sonra eğitim kümesindeki öznitelikler kullanılarak elde edilen karar ya da sınıflandırma kuralı, test kümesine uygulanır sınıflandırma başarısı hesaplanır.

3. MATERYAL VE YÖNTEM

Bu çalışmada yalıtık kelime tanıma işlemi kişi bağımlı olarak AOVY, FDAA ve YSA sınıflandırıcıları ile gerçekleştirilmiştir. Ses veritabanı sınırlı sayıda (10 adet) sözcükten oluşmaktadır. Eğitim ses veri tabanında bu sözcüklerin herbiri için 50 ses sinyali kişi bağımlı olarak bilgisayar ortamında kaydedilmiştir. Ses sinyallerine MFKK uygulanarak öznitelik vektörleri elde edilmiştir. Eğitim aşamasında öncelikle ses veri tabanındaki her sözcüğe ait öznitelik vektörleri AOVY, FDAA ve YSA kullanılarak eğitilmiştir. Test aşamasında bir mikrofon aracılığı ile seslendirilen iki sözcüğün her biri ayrı ayrı bilgisayar ortamında yazılmış bir program aracılığıyla sınıflandırılmıştır. Bilgisayar aracılığıyla tanınan kelimelere göre bilgisayar ile RS-232 seri bağlantıya sahip robot kolu servo motor kontrol kartına sinyaller gönderilerek robot kol istenilen nesneye doğru yöneltilmiştir. Sınıflandırma sonucunda elde edilen tanıma başarımını artırmak için bir çeşit dil modeli de geliştirilmiştir.

3.1. Yalıtık Kelime Tanıma

Çalışmada kişi bağımlı ve yalıtık kelime tanıma gerçekleştirilmiştir. Konuşma tanıma için kullanılan yöntemler için öncelikle eğitim aşaması gerçekleştirilmiştir. Bunun için ilk olarak ses veri tabanı oluşturulmuş, ses veri tabanından her sınıf için öznitelik vektörlerinin elde edilmiş ve bu öznitelik vektörleri kullanılarak AOVY, FDAA ve YSA için eğitim işlemi yapılmıştır.

3.2. Eğitim Veritabanının Oluşturulması

Öncelikle kişi bağımlı olarak 10 farklı kelime seslendirilerek bilgisayar ortamına kaydedilmiştir. Bu 10 farklı kelimedenden dördü “küp”, “prizma”, “silindir”, “küre” gibi şekilleri içerirken diğer altı kelime bu şekillere ait renklerden (beyaz, kırmızı, mavi, siyah, sarı, yeşil) oluşmaktadır. Robot kol tarafından ses tanıma ile algılanacak her bir şekle ait 6 renk oluşturulmuştur. Böylece toplam 24 farklı ikili sözcükten oluşan komut yapısı oluşmaktadır. Bu komutlar aşağıdaki Tablo.1’de verilmiştir.

Tablo 3.1. Renkli nesnelere ait komut kümesi

Renk-Nesne	Renk-Nesne	Renk-Nesne	Renk-Nesne	Renk-Nesne	Renk-Nesne
Kırmızı silindir	Beyaz silindir	Mavi silindir	Siyah silindir	Sarı silindir	Yeşil silindir
Kırmızı küre	Beyaz küre	Mavi küre	Siyah küre	Sarı küre	Yeşil küre
Kırmızı prizma	Beyaz prizma	Mavi prizma	Siyah prizma	Sarı prizma	Yeşil prizma
Kırmızı küp	Beyaz küp	Mavi küp	Siyah küp	Sarı küp	Yeşil küp

Eğitim ses veri tabanındaki ses kayıtları her kelime sınıfı için 16 kHz’de örneklenmiştir ve her örnek 16 bit’lidir. Öznitelik vektörleri 40 ms’lik çerçeveler üzerinden elde edilmiştir. Çerçeveler arası %50 üst üste bindirme yapılmıştır. Her bir ses sinyali için 32 çerçeve bulunmuştur. Bu çerçevelerin herbiri 13 MFKK katsayısı ve bir çerçeve enerjisinden oluşan 14 parametreyle temsil edilmektedir. Ayrıca delta ve delta-delta katsayıları da alınarak her çerçeve için toplam 42 özellik vektörü elde edilmiştir. Böylece bir sınıf için her biri 1344 uzunluklu 50 adet öznitelik vektörü elde edilmiştir. Eğitim veri kümesi bu öznitelik vektörlerin her sınıf için birleştirilmesi ile oluşturulmuştur.

3.3. Dil Modelleri

Çalışmada eğitim aşamasında 10 sınıf için öznitelik vektörleri elde edildikten sonra test aşamasında sınıflandırma için iki farklı yöntem kullanılmıştır. Bu yöntemler dil modeli sınıflama ve dil modelsiz sınıflama olarak ikiye ayrılır. Tanıma aşaması bir mikrofon aracılığı ile gerçek zamanlı olarak gerçekleştirilmiştir.

3.3.1. Dil Modeli Sınıflama

Dil modeli sınıflamada komut seti incelendiğinde ilk sözcüklerin renk, ikinci sözcüklerin de nesne olması sebebi ile test aşamasında sınırları tespit edilen ilk sözcük sinyali 6 adet renk sınıfında, ikinci sözcük ise 4 adet olan nesne sınıfında aranmaktadır. Bu durumda bir sözcük 10 sınıftan daha az bir sınıf için tanındığından daha yüksek tanıma oranlarına erişilebilecektir.

3.3.2. Dil Modelsiz Sınıflama

Dil modelsiz sınıflamada ise sınırları belirlenen her sözcük 10 sınıf için tanınmaktadır. Bu durumda dil modeli tanımaya göre daha düşük tanıma oranları oluşabilecektir.

3.4. Sınıflandırıcıların Kullanımı

3.4.1. Ayırt Edici Ortak Vektör Yaklaşımının Kullanımı

Üç adet sınıflandırıcıdan AOVY sınıflandırıcısı yeterli veri durumu ve yetersiz veri durumuna göre farklı tanıma performansları gösteren bir sınıflandırıcıdır. Çalışmada her bir öznitelik vektörü boyutu (1344), tüm sınıflardaki toplam vektör sayısından (500) büyük olduğu için yetersiz veri durumu gerçekleşmektedir. Bu durumda fark ve farksızlık alt uzayları birbirinden ayrılabilen ve tanıma oranları yeterli veri durumuna göre daha iyi çıkmaktadır. Öncelikle 10 adet sınıfın her biri için 1344 boyutlu bir adet ortak vektör bulunur. Ardından W_{opt} kullanılarak bu ortak vektörlerden 9 boyutlu ayırt edici ortak vektörler bulunur. Test sinyali x_{test} , W_{opt} ile iz düşürülüp 10 sınıf için en küçük öklit mesafeyi veren sınıfa atama yapılır.

3.4.2. FDAA Kullanımı

Fisher doğrusal ayırım analizi için öncelikle eğitim aşamasında tüm sınıfları birbirinden maksimum biçimde uzaklaştıran birim dik vektör seti W_{opt} elde edilir. Ardından eğitim sınıfındaki her 50 adetlik öznitelik vektörleri W_{opt} ile optimum alt uzaya iz düşürülür. Sonuçta 10 sınıf (c) olduğundan her öznitelik vektörü 1344 boyuttan 9 ($=c-1$) boyuta düşürülür. Test öznitelik vektöründe Sw ile iz düşürülerek 9 boyutlu öz niteliğe düşürülür ve toplam 500 adet (50 vektör*10 sınıf) 9 boyutlu vektörler ile öklit uzaklığı bakılır. En küçük uzaklığı veren sınıfa atama yapılır.

3.4.3. YSA Kullanımı

YSA için eğitim aşamasında her sınıfa ait 50 adet öz nitelik vektörüne, sınıfına ait bir etiket atanmış ve giriş, ara katman ve tek nörondan oluşan çıkış katmanlı bir yapay sinir ağı oluşturulmuştur. Bu ağ eğitim için Ölçekli eşlenik gradyan geri yayılımı (Scaled conjugate gradient back propagation) kullanılmaktadır. Öznitelik vektörleri boyutu çok büyük olduğu için (1344) çalışmada Temel Bileşen Analizi (TBA) kullanılarak vektör boyutları sırası ile 10, 20,

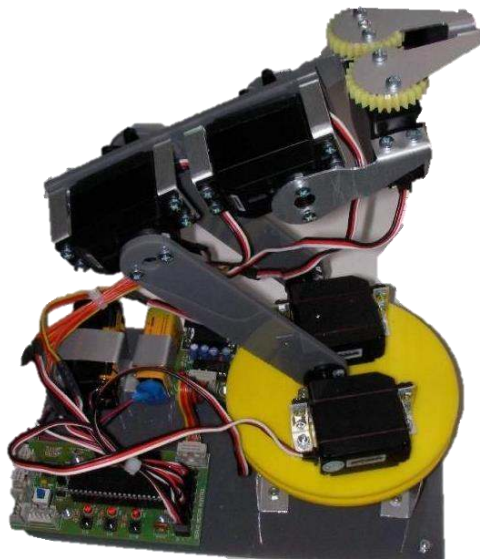
40, 60 ve 80'e düşürülmüş ve giriş katmanı nöron sayıları da bu vektör uzunluklarına eşit alınmıştır. Eğitim tamamlandıktan sonra test sinyaline benzer biçimde TBA ile boyut düşürme gerçekleştirilir. Bu test sinyali ağına girişine verilir ve çıkışta bir etiket değeri elde edilir. Bu değer hangi sınıfın etiket değerine yakınsa o sınıfa atama yapılır.

3.5. Robot Kolu Kontrolü

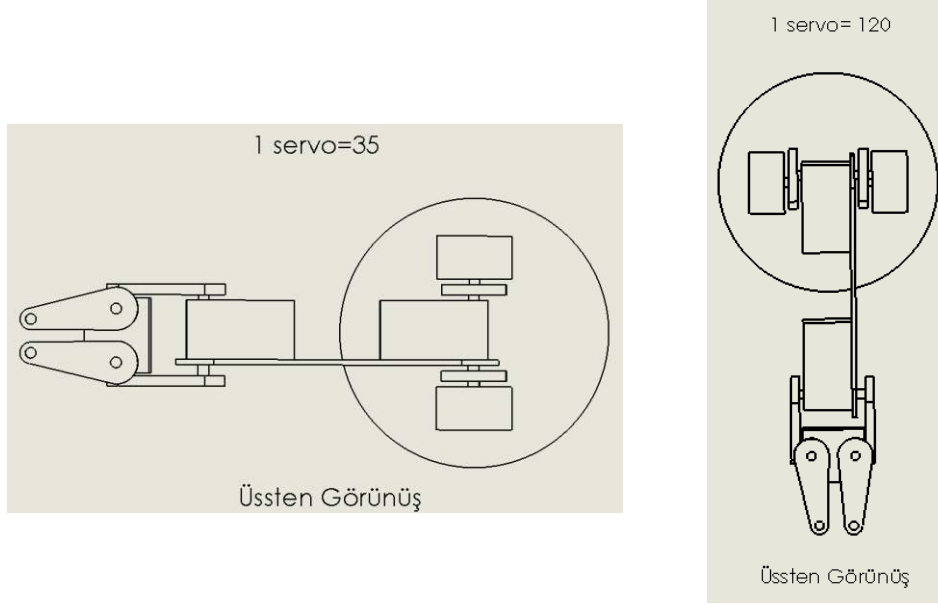
Test aşamasında ses sinyalleri mikrofon aracılığı ile bilgisayara aktarılmakta ve komutları oluşturan her bir kelime sinyalinin başlangıç ve bitiş noktaları çerçevelerindeki sıfır geçiş sayısı ve enerji seviyelerine göre bulunmaktadır. Sınırlar bulunduktan sonra her kelime sinyali için MFKK ile öznelik katsayıları elde edilmektedir. Ardından bu öznelik vektörleri için AOVY ya da FDAA kullanılarak en olası sınıfa atama yapılmaktadır. Tanıma sonucunda bulunan komuta göre oluşturulan bir sinyal RS232 seri haberleşme kablosu üzerinden servo motor kontrol kartına iletilmektedir. Gelen sinyal tanınan komuta karşılık gelen nesnenin koordinat bilgisini içermektedir. Böylece robot kolu istenilen nesneye doğru yönelerek nesneyi kavramakta ve tanıma süreci son bulmaktadır. Başka bir komut geldiğinde benzer süreçler tekrarlanmaktadır.

3.5.1. Robot Kolu Yapısı

Robot kolu aşağıdaki Şekil 3.1'deki gibi verilmiştir. Şekil 3.1'den de görülebileceği gibi seri haberleşme ile gelen sinyalleri işleyen bir kontrol kartına ve hareketi sağlayan servo motorlara sahiptir.

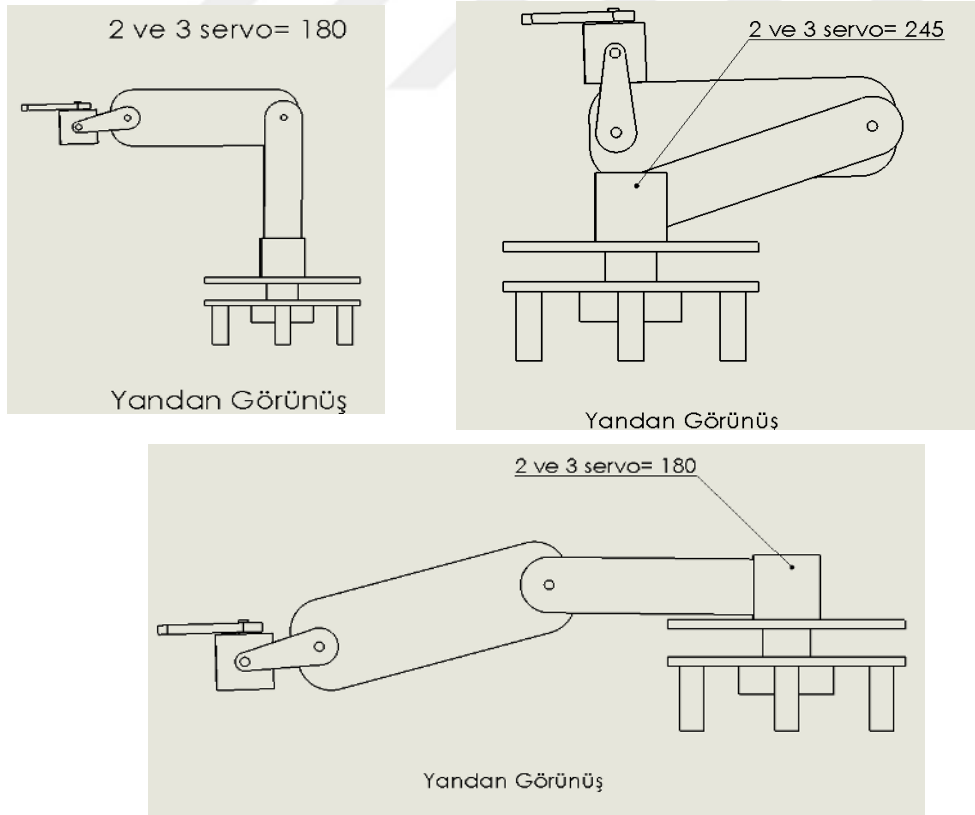


Şekil 3.1 Kullanılan robot kolu



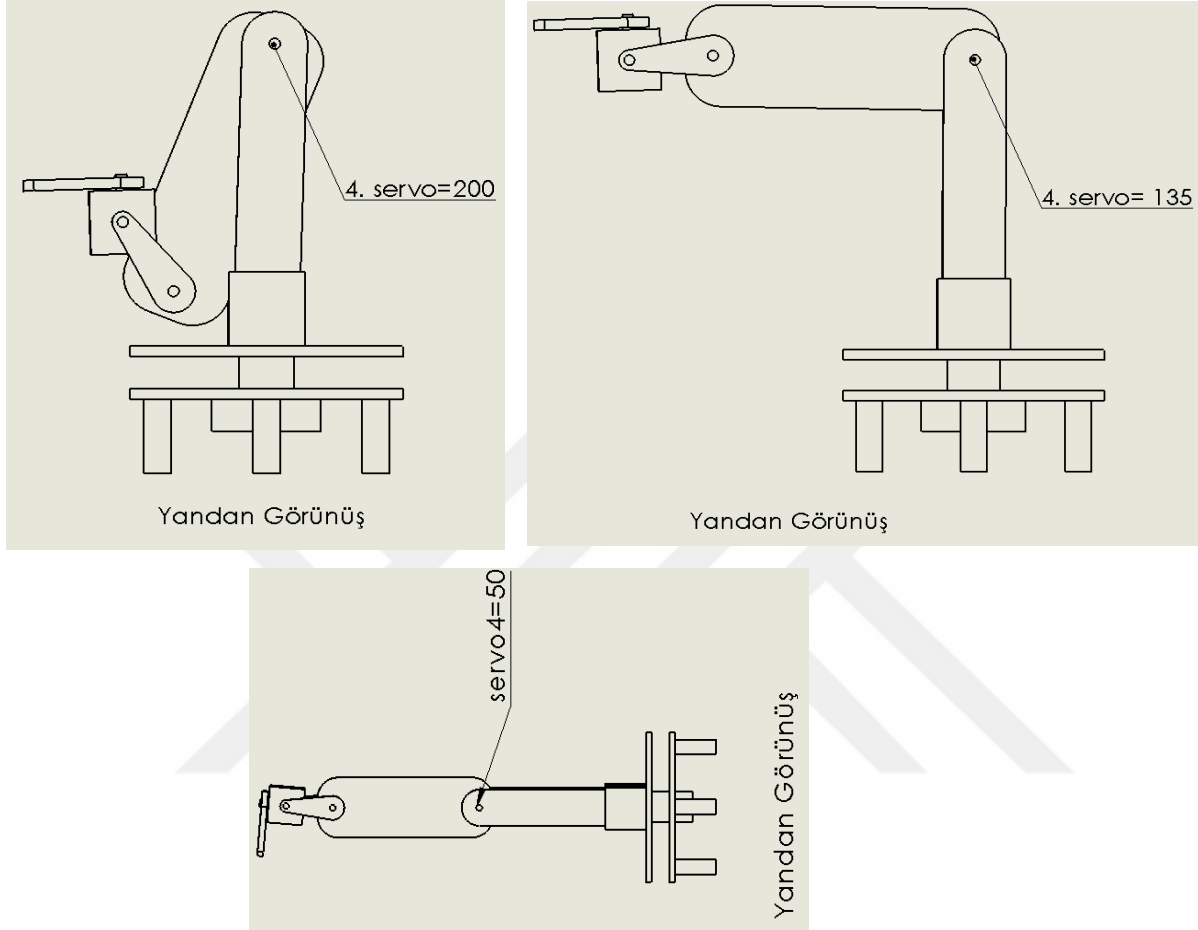
Şekil 3.2. Servo 1 konumları

Robot kolundaki 1 numaralı servo motor şekildeki gibi robot kolunu döndürmektedir. 1 numaralı servo 35 iken robot kol en sağ tarafa dönmekte, 120 iken orta konuma gelmektedir.



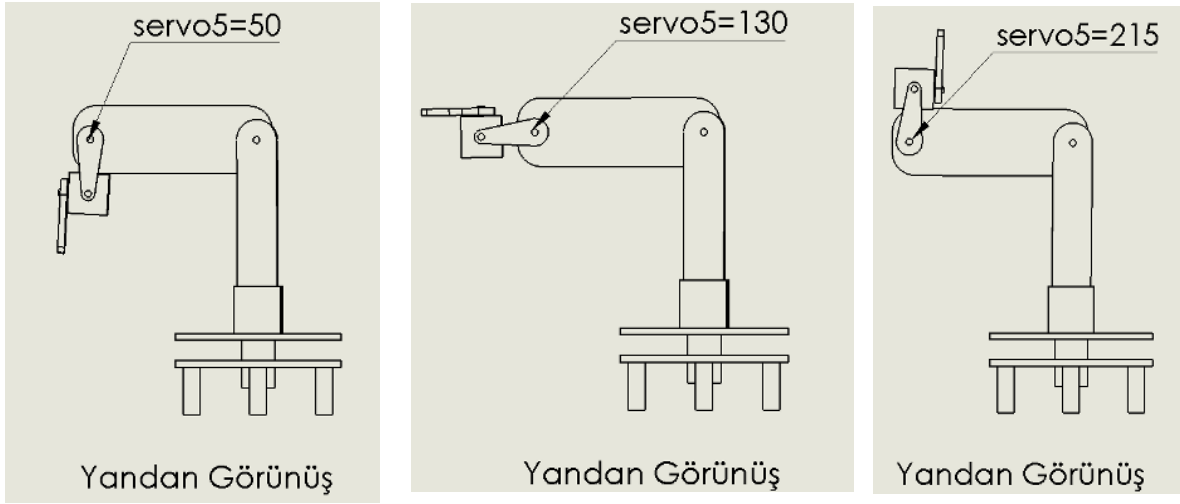
Şekil 3.3. Servo 2 ve 3 konumları

Robot kolundaki 2 ve 3 numaralı servolar birbirleri ile paralel çalışmaktadırlar. Şekilde 2 ve 3 numaralı servonun ileri, geri ve orta konumları görülmektedir. 2 ve 3 numaralı servo 45 iken robot kollunun alt kısmı geriye konumlanır. 2 ve 3 numaralı servo 180 iken robot kolunun alt kısmı ortaya, 90 iken robot kollunun alt kısmı ileriye konumlamaktadır.

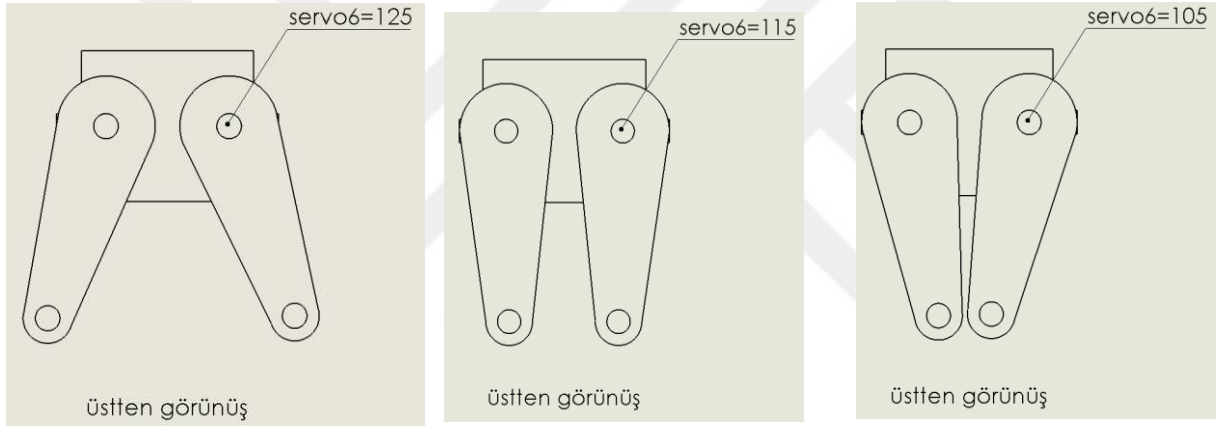


Şekil 3.4. Servo 4 konumları

Robot kolundaki 4 numaralı servo motor Şekil 3.4. deki gibi robot kolunu hareket ettirmektedir. Çalışmada 4 numaralı servo 200 iken robot kolunun üst kol kısmı aşağı doğru, 135 iken robot ortaya ve 50 iken yukarı doğru konumlanır. Robot kolundaki 5 numaralı servo motor şekildeki gibi robot kolunu hareket ettirmektedir. Bu servo 45 iken robot kolunun el kısmı aşağı doğru, 130 iken ortaya doğru ve 215 iken robot kolunun el kısmı yukarıya doğru konumlanır.



Şekil 3.5. Servo 5 konumları



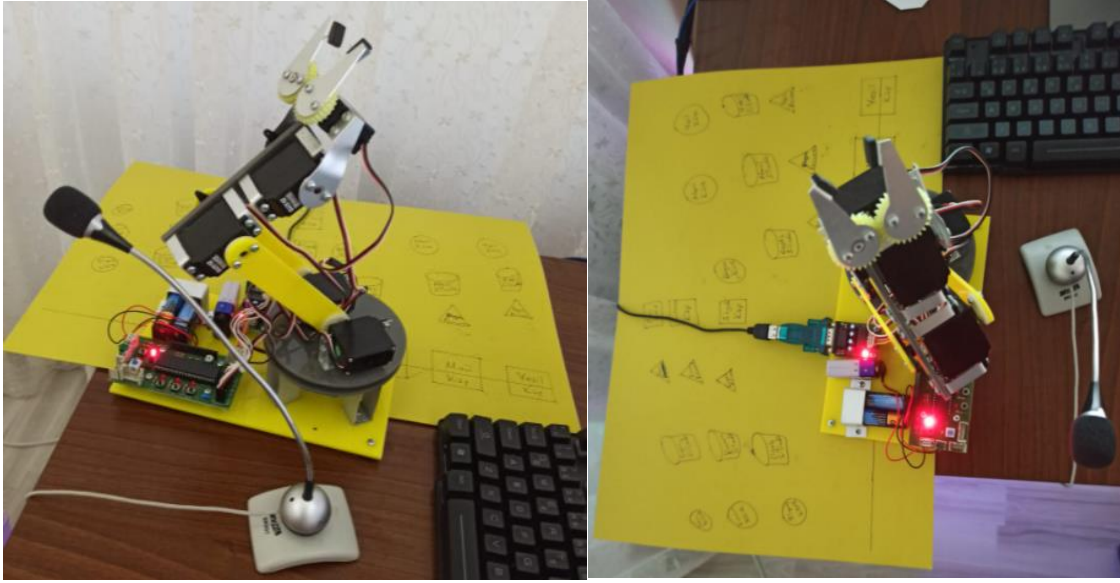
Şekil 3.6. Servo 6 konumlar

6 numaralı servo motorun hareketi Şekil 3.6.'daki gibi robot kolunu hareket ettirmektedir. 6 numaralı servo 125 iken robot kolunun tutamak kısmı tam açık konumlanır iken 115 olduğunda normal konumlanır. 105 iken robot kolunun tutamak kısmı kapalı konumdadır.

Tablo 3.2: Tanınan komutlara karşılık gönderilen karakterler

Tanınan komut	Gönderilen karakter
'beyaz küp'	'a'
'beyaz prizma'	'b'
'beyaz silindir'	'c'
'beyaz küre'	'd'
'kırmızı küp'	'e'
'kırmızı prizma'	'f'
'kırmızı silindir'	'g'
'kırmızı küre'	'h'
'mavi küp'	'i'
'mavi prizma'	'j'
'mavi silindir'	'k'
'mavi küre'	'l'
'sarı küp'	'm'
'sarı prizma'	'n'
'sarı silindir'	'o'
'sarı küre'	'p'
'yeşil küp'	'r'
'yeşil prizma'	's'
'yeşil silindir'	't'
'yeşil küre'	'u'
'siyah küp'	'v'
'siyah prizma'	'w'
'siyah silindir'	'y'
'siyah küre'	'z'

Aşağıdaki şekil 3.7’de komut tanımada kullanılan robot kol sistemi verilmiştir. Bu sistemde görüldüğü üzere ses komutlarını algılamak için bir mikrofon, bir robot kol, tanınan sese ait sinyallerin seri port aracılığı ile alındığı ve servo motor kontrol devresine bağlı bir usb kablo ve nesnelerin pozisyonunu içeren bir koordinat sistemi karton aracılığı ile gerçekleştirilmiştir.

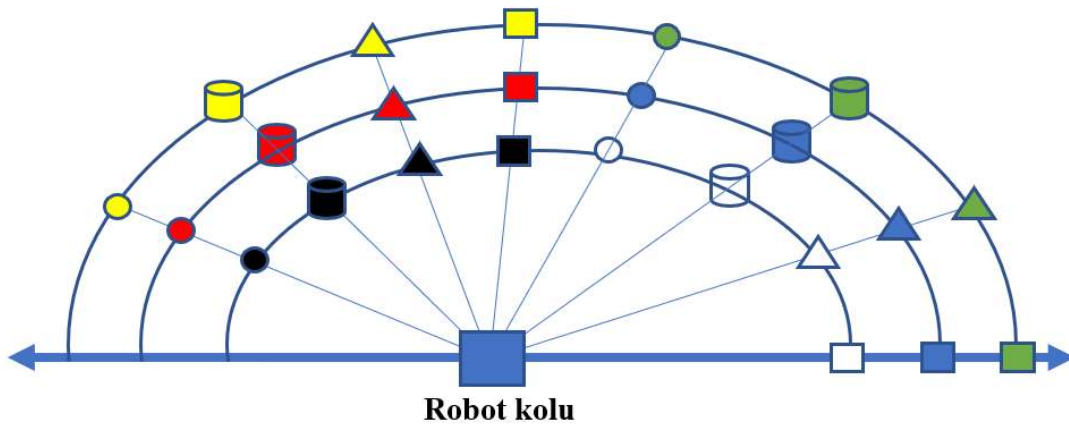


Şekil 3.7. Tanıma için kullanılan robot kol sistemi

3.5.2. Robot Kolu Koordinatları

Bu çalışmada robot kolunun bilgisayardan gönderilen veriler ile nesnelere ait önceden tanımlanmış bazı konumlara gelmesi sağlanmıştır. Bilgisayardan tanınan ses komutlarına karşılık gelen karakterler bilgisayarın seri portundan bir kablo aracılığı ile servo motor kontrol devresine gönderilmektedir.

Tanınan karakter RS-232 seri haberleşme kablosu ile robot kontrol kartına iletilir ve Robot Şekil 3.2’de belirtilen ilgili konuma gittikten sonra alma işlemini gerçekleştirir.



Şekil 3.8. 24 adet nesnenin konumları

4. BULGULAR

Kişi bağımlı kelime tanıma çalışmada her ikili kelime komutu 200 defa gerçek zamanlı olarak seslendirilerek test işlemi gerçekleştirilmiştir. Bu çalışmada AOVY yetersiz veri durumu için ses komut tanıma yapılmaktadır. Çalışmada öncelikle eğitim setindeki her sınıfa ait 50 ses verisi için 10 katlamalı çapraz geçerleme (10 fold cross validation) yapılarak AOVY, OVY, FDAA ve YSA için test işlemi gerçekleştirilmiştir. Bu sonuçlar Tablo 4.1’de verilmiştir. YSA için 40 boyutlu öznitelik vektörlerine göre 40 nörona sahip bir giriş katmanı, 2 adet olan ve ilki 30, ikincisi 1 nörona sahip ara katmanlar ve bir çıkış katmanı olmak üzere 4 katmanlı yapay sinir ağı modeli oluşturulmuştur. Bu ağı eğitimi için Levenberg-Marquardt algoritması kullanılmış, öğrenme oranı $\mu=0.001$ ve iterasyon sayısı 1000 olarak seçilmiştir.

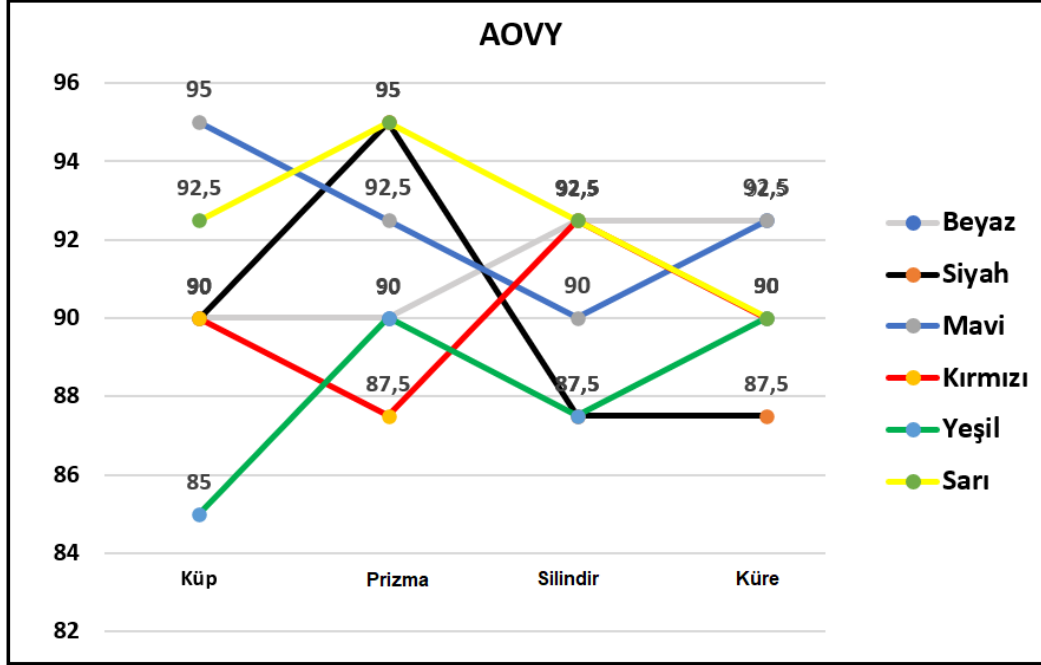
Tablo 4.1: AOVY, OVY, FDAA ve YSA için ortalama tanıma oranları

AOVY	AOV	FDAA	YSA
%99,2	%99,2	%97,6	%71,6

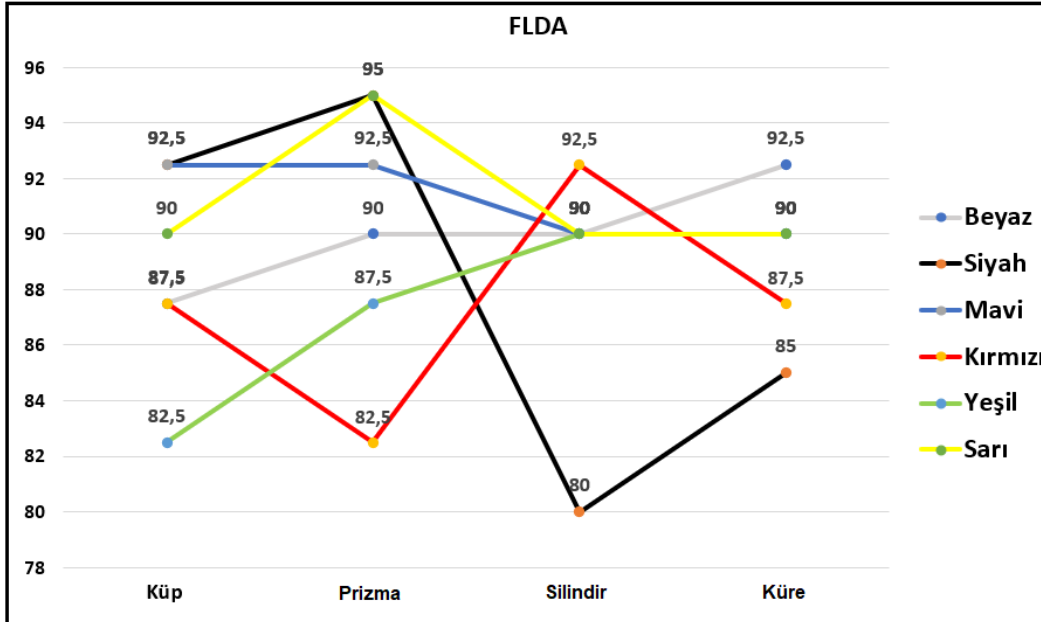
Tablo 4.1’den görülebileceği gibi AOVY ve OVY yetersiz veri durumu için en yüksek ve eşit tanıma oranına sahiptir. AOVY ortak vektörlerin dağılımlarını birbirinden uzaklaştıran bir algoritmaya sahiptir [10]. Ancak eğitim kümesindeki sinyallere ait sınıf dağılımları birbirinden yeterince uzak ise AOVY ve OVY birbirine eşit çıkabilmektedir. AOVY’nin OVY’den bir üstünlüğü ise, test aşamasında sadece sınıf sayısı-1 (9) kadar bir vektörü karşılaştırmasıdır. Ancak OVY, eğitim kümesinde kullanılan bir örneğe ait öz nitelik vektörü boyutunu (1344) karşılaştırmaktadır. Böylece AOVY, OVY’den daha kısa hesaplama süresi vermektedir. Bu avantajlarından dolayı bu çalışmada sadece AOVY ile gerçek zamanlı komut tanıma işlemi gerçekleştirilmiştir.

4.1. AOVY ve FDAA İçin Bulunan Konuşma Tanıma Oranları

Aşağıdaki Şekil 4.1’de ve Şekil 4.2’de dil modeli kullanmadan gerçekleştirilen testlerin sonuçları sırası ile AOVY ve FDAA için verilmiştir.

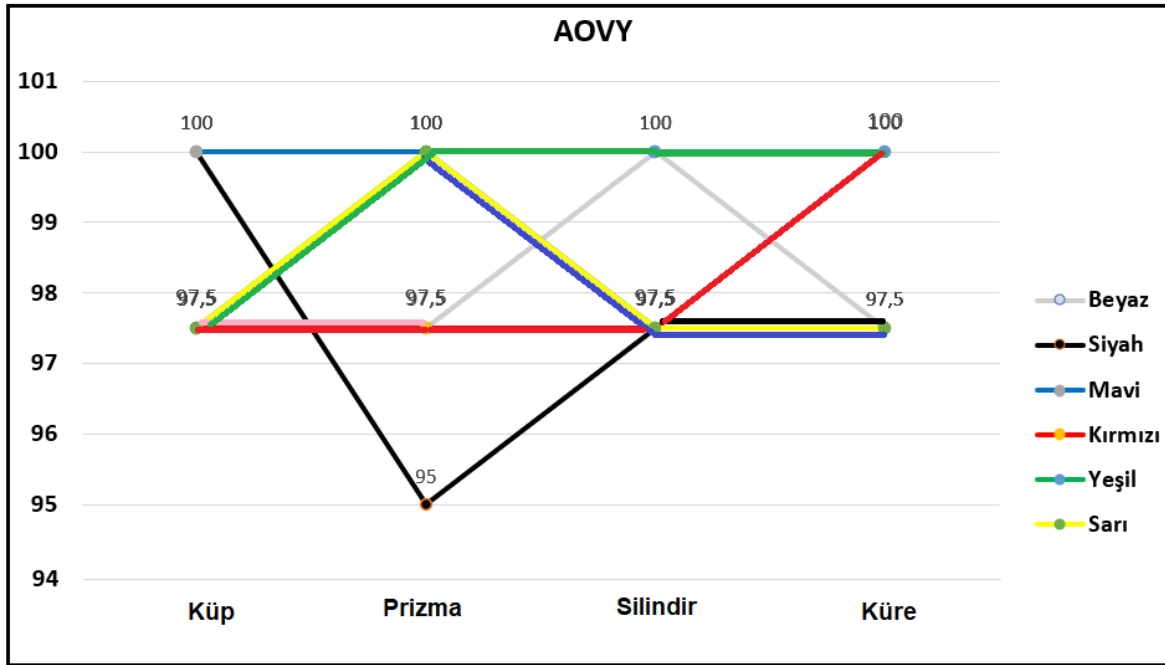


Şekil 4.1. AOVY için dil modelsiz komut kümeleri tanıma oranları (Ortalama=%90,73)

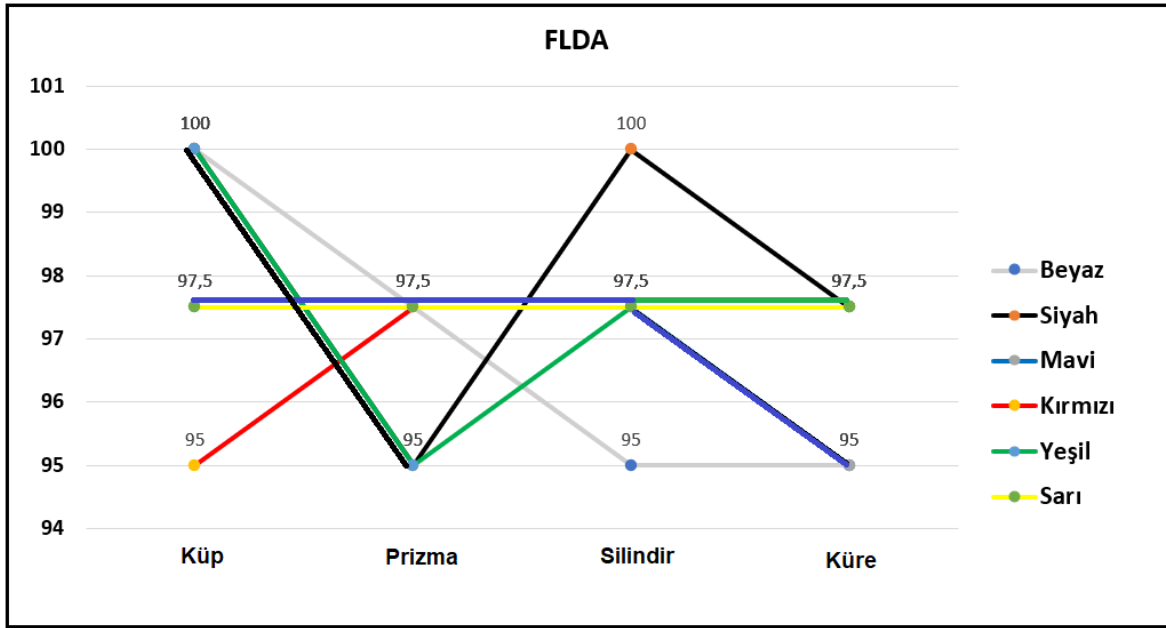


Şekil 4.2. FDAA için dil modelsiz komut kümeleri tanıma oranları (Ortalama=%89,48)

Şekil 4.1 ve Şekil 4.2’de dil modeli kullanılmadan AOVY ve FDAA için 24 farklı komut için bulunmuş ortalama tanıma oranları verilmiştir. 24 komutun tamamının ortalama tanıma oranları ise AOVY ve FDAA için sırası ile %90,73 ve %89,48 olarak bulunmuştur. Her ikili komutun kelimeleri ayrı ayrı 10 adet sınıf içerisinde en yakın sınıfa atanmıştır. Ancak bir komutta her zaman önce renk bilgisi sonra nesne bilgisi seslendirilmektedir. Bu yüzden ilk komutun 6 farklı renk içinden ve ikinci komutun 4 farklı nesne içinden seçilmesini içeren dil modelli bir çalışma daha gerçekleştirilmiştir. Bu şekilde her sözcük kendi sınıf kümesi içerisinde sınıflandırılacağı için daha iyi tanıma başarımları elde edilebilecektir. Aşağıdaki Şekil 4.3 ve Şekil 4.4’de dil modeli kullanılarak gerçekleştirilen testlerin sonuçları sırası ile AOVY ve FDAA için verilmiştir.



Şekil 4.3. AOVY için dil modelli komut kümeleri tanıma oranları (Ortalama=%98,3)



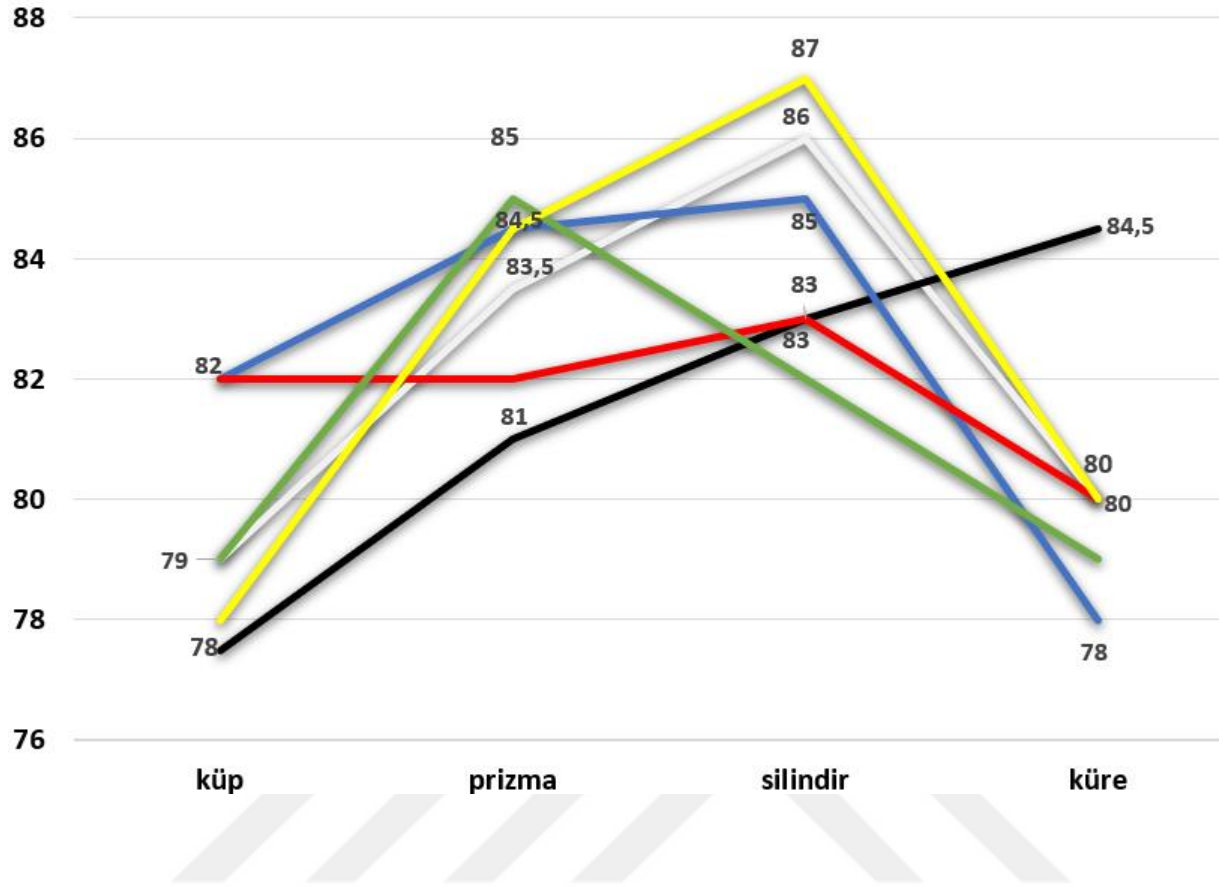
Şekil 4.4. FDAA için dil modelli komut kümeleri tanıma oranları (Ortalama=%97,1)

Şekil 4.3 ve Şekil 4.4'ten görülebileceği gibi dil modelli çalışma ile dil modelsiz çalışmaya kıyasla daha iyi tanıma oranları elde edilmiştir.

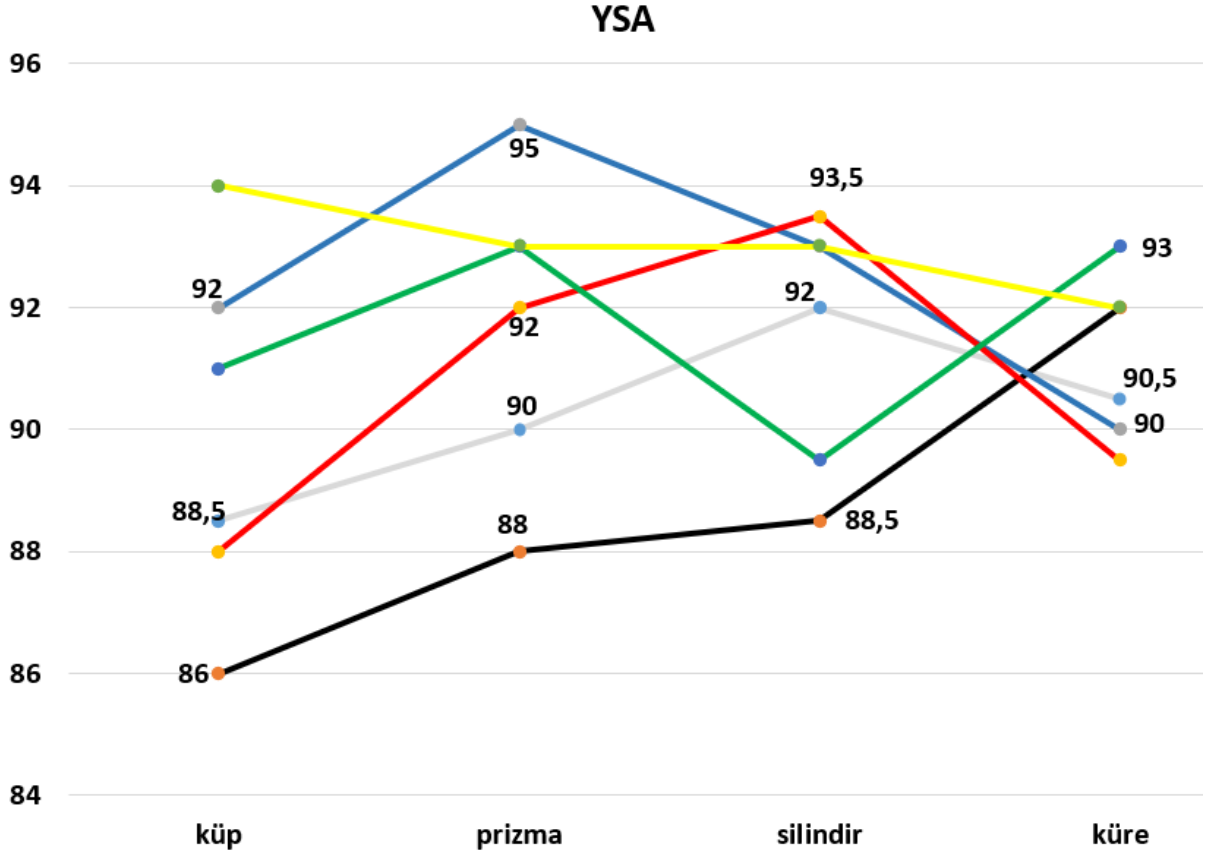
4.2. YSA İçin Bulunan Konuşma Tanıma Oranları

Çalışmada PCA ile 1344 boyutlu MFCC öznitelik vektörleri sırası ile 10,20,40,60 ve 80 boyuta indirgenmiş ve herbiri için test işlemi gerçekleştirilmiştir. En iyi sonucu 40 boyutlu öznitelik boyutu için elde edilmiştir. Aşağıdaki şekiller için 40 boyutlu öznitelik vektörlerine göre 40 nörona sahip bir giriş katmanı, 2 adet ilki 30, ikincisi 1 nörona sahip ara katmanlar ve bir çıkış katmanı olmak üzere 4 katmanlı yapay sinir ağı modeli oluşturulmuştur. Bu ağ eğitim için Levenberg-Marquardt algoritması kullanılmaktadır. Ayrıca öğrenme oranı $\mu=0.001$ ve iterasyon sayısı 1000 olarak seçilmiştir. Aşağıdaki Şekil 4.5 ve Şekil 4.6'da YSA için bulunan sırası ile dil modelsiz ve dil modelli sonuçlar verilmiştir. 24 komut için dil modelsiz ve dil modelli tanıma ortalama tanıma oranları sırası ile %81,9 ve %91,1 olarak bulunmuştur.

YSA dil modelsiz tanıma oranları



Şekil 4.5. YSA için dil modelsiz tanıma oranları (ortalama=%81,9)



Şekil 4.6. YSA için dil modellenli tanıma oranları (ortalama=%91,1)

4.3 Kullanılan Sınıflandırıcıların Tanıma Süreleri

Çalışmada 24 ayrı komutun her biri için 200 adet test gerçekleştirilmiştir. Bu 3 sınıflayıcının sınıflama süreleri ortalama olarak bulunmuştur. Literatürde AOVY sınıf sayısı kadar (10 adet) vektör ile karşılaştırma yapar, ancak test sinyalinin yüksek boyutlu (1344x9) izdüşüm matrisi ile çarpılması gerektiğinden hesaplama süresinin ortalama 0,00162 sn olduğu görülmüştür. FDAA ise eğitim kümesindeki toplam örnek sayısına (500 adet) göre sınıflama yaptığından ve benzer şekilde test sinyali yüksek boyutlu bir izdüşüm matrisi ile çarpım gerektiğinden AOVY'ye göre bir miktar daha uzun hesaplama süresine sahip olduğu görülmüştür (0,0035 sn). YSA ise bu iki sınıflayıcıya göre daha uzun sürede (0.0113 sn) hesaplama yapmıştır.

5. SONUÇ ve ÖNERİLER

Bu çalışmada sınıflandırma için etkin bir biçimde kullanılan alt uzay sınıflandırıcılar olan AOVY ve FDAA'nın yanı sıra YSA'da kullanarak üç farklı sınıflandırıcı ile gerçek zamanlı ve kişi bağımlı konuşma tanıma uygulaması gerçekleştirilmiştir. Tanıma sonucu elde edilen bilgi seri haberleşme ile bilgisayara iletilmiş ve servo motor kontrolü sağlanarak robot kolu istenilen koordinatlara yönlendirilmiştir. İlk önce eğitim kümesindeki 10 adet sınıf için çapraz geçişleme ile AOVY, OVY, FDAA ve YSA sınıflayıcıları test edilmiştir. AOVY ve OVY en yüksek tanıma oranlarını (%99,2) vermiştir. Bu eşit tanıma oranı, eğitim kümesindeki sinyallere ait sınıf dağılımları birbirinden yeterince uzak olduğunda oluşabilmektedir. Çalışma dil modelsiz ve dil modellenmiş olarak iki biçimde gerçekleştirilmiştir. Dil modelsiz çalışmada 24 komut kümesinin tümü için AOVY ve FDAA sırası ile %90,73 ve %89,48 ortalama tanıma oranlarına erişmiştir. YSA ise ortalama %81,9 tanıma oranına sahiptir. Dil modellenmiş çalışmada ise AOVY ve FDAA için sırası ile %98,3 ve %97,1 ortalama tanıma oranlarına erişilmiştir. Dil modellenmiş için YSA ise ortalama %91,1 tanıma oranına sahiptir. Sonuçlar incelendiğinde özellikle dil modellenmiş tanımda AOVY ve FDAA'nın birbirine yakın ve oldukça iyi sonuçlar verdiği görülmektedir. YSA ise bu iki alt uzay sınıflayıcıdan daha düşük sonuçlar vermiştir.

KAYNAKÇA

- [1] Filho, G. L., & Moir, T. J. (2010). From science fiction to science fact: a smart-house interface using speech technology and a photo-realistic avatar. *International journal of computer applications in technology*, 39(1-3), 32-39.
- [2] Anggraeni, D., Sanjaya, W. S. M., Nurasyidiek, M. Y. S., & Munawwaroh, M. (2018). the implementation of speech recognition using mel-frequency cepstrum coefficients (MFCC) and support vector machine (SVM) method based on python to control robot arm. In *IOP Conference Series: Materials Science and Engineering* (Vol. 288, No. 1, p. 012042). IOP Publishing.
- [3] Soujanya, M., & Kumar, S. (2010, August). Personalized IVR system in contact center. In *2010 International Conference on Electronics and Information Engineering* (Vol. 1, pp. V1-453). IEEE.
- [4] Furui, S., Kikuchi, T., Shinnaka, Y., & Hori, C. (2004). Speech-to-text and speech-to-speech summarization of spontaneous speech. *IEEE Transactions on Speech and Audio Processing*, 12(4), 401-408.
- [5] Gülmezoglu, M. B., Dzhafarov, V., Keskin, M., & Barkana, A. (1999). A novel approach to isolated word recognition. *IEEE Transactions on Speech and Audio Processing*, 7(6), 620-628.
- [6] Gorin, A., & Mammone, R. J. (1994). Introduction to the special issue on neural networks for speech processing. *IEEE Transactions on Speech and Audio Processing*, 2(1), 113-114.
- [7] Abdulla, W. H., Chow, D., & Sin, G. (2003, October). Cross-words reference template for DTW-based speech recognition systems. In *TENCON 2003. Conference on convergent technologies for Asia-Pacific region* (Vol. 4, pp. 1576-1579). IEEE.
- [8] Gunal, S., & Edizkan, R. (2007, July). Use of novel feature extraction technique with subspace classifiers for speech recognition. In *IEEE International Conference on Pervasive Services* (pp. 80-83). IEEE.

- [9] Gülmezoğlu, M. B., Dzharfarov, V., Edizkan, R., & Barkana, A. (2007). The common vector approach and its comparison with other subspace methods in case of sufficient data. *Computer Speech & Language*, 21(2), 266-281.
- [10] Cevikalp, H., Neamtu, M., Wilkes, M., & Barkana, A. (2005). Discriminative common vectors for face recognition. *IEEE Transactions on pattern analysis and machine intelligence*, 27(1), 4-13.
- [11] Çakır, M. Y. (2017). Gerçek zamanlı yüksek kalitede ses tanıma (Master's thesis, İstanbul Sabahattin Zaim Üniversitesi, Fen Bilimleri Enstitüsü, Bilgisayar Mühendisliği Anabilim Dalı).
- [12] Baygün, M. K. (2006). *Türkçe Komutları Tanıyan Ses Tanıma Sistemi Geliştirilmesi Yüksek Lisans Tezi*. Denizli: Pamukkale Üniversitesi.
- [13] Gelegin, İ., & Bolat, B. (2011). *Ayrık Kelime Tabanlı Bir Konuşma Tanıma Sistemiyle Bilgisayar Kontrolü*. Elazığ: Elektrik-Elektronik ve Bilgisayar Sempozyumu.
- [14] Dede, G. (2008). *Yapay Sinir Ağları ile Konuşma Tanıma Yüksek Lisans Tezi*. 76 Ankara: Ankara Üniversitesi Fen Bilimleri Enstitüsü
- [15] Salor, Ö., Pellom, B., Çiloğlu, T., Hacıoğlu, K. and Demirekler, M., 2002, New Corpora and Tools for Turkish Speech Research, *Sinyal İşleme ve İletişim Uygulamaları Kurultayı*, s: 476-481.
- [16] Mengüşoğlu, E. (1999). *Bir Türkçe Sesli İfade Tanıma Sisteminin Kural Tabanlı Tasarımı ve Gerçekleştirimi*. Ankara: Hacettepe Üniversitesi Fen Bilimleri Enstitüsü, s. 14-16, 22-26.
- [17] Ghai, W., & Singh, N. (2012, March). Literature Review on Automatic Speech Recognition. *International Journal of Computer Applications* (0975-8887), 41(8).
- [18] Aydın, Ö. (2005). *Yapay sinir ağlarını kullanarak bir ses tanıma sistemi geliştirilmesi Yüksek Lisans Tezi*. Edirne: Trakya Üniversitesi Fen Bilimleri Enstitüsü

- [19] Keser, S., & Edizkan, R. (2009, April). Phonem-based isolated Turkish word recognition with subspace classifier. In 2009 IEEE 17th Signal Processing and Communications Applications Conference (pp. 93-96). IEEE.
- [20] Giannakopoulos, T., & Pirkakis, A. (2014). Introduction to Audio Analysis: a MATLAB® approach. Academic Press.
- [21] Tiwari, V. (2010). MFCC and its applications in speaker recognition. International journal on emerging technologies, 1(1), 19-22.
- [22] Han, W., Chan, C. F., Choy, C. S., & Pun, K. P. (2006, May). An efficient MFCC extraction method in speech recognition. In 2006 IEEE international symposium on circuits and systems (pp. 4-pp). IEEE.
- [23] Lokesh, S., & Devi, M. R. (2019). Speech recognition system using enhanced mel frequency cepstral coefficient with windowing and framing method. Cluster Computing, 22(5), 11669-11679.
- [24] Bhatnagar, A. C., Sharma, R. L., & Kumar, R. (2012). Analysis of Hamming window using advance peak windowing method. International Journal of Scientific Research Engineering & Technology, 1(4), 015-020.
- [25] Karasartova, S. Y., & İLK, H. G. (2011). Metinden bağımsız konuşmacı tanıma sistemlerinin incelenmesi ve gerçekleştirilmesi (Doctoral dissertation, Ankara Üniversitesi Fen Bilimleri Enstitüsü Elektronik Mühendisliği Anabilim Dalı).
- [26] Silva, D. F., & Batista, G. E. (2016). Speeding Up All-Pairwise Dynamic Time Warping Matrix Calculation. *Proceedings of the 2016 SIAM International Conference on Data Mining*. PRDT16.
- [27] Rabiner, L., Juang, B., 'Fundamentals of Speech Recognition', ISBN: 0-13-015157-2, 1993.
- [28] YILMAZ, B. (2019). Maliyet Fonksiyonun Belirlenmesinde Yapay Sinir Ağı Modellerinin Kullanımı. Muhasebe ve Finansman Dergisi, 329-344.
- [29] Öztemel, Ercan. Yapay sinir ağları. Papatya, 2012.

- [30] KELESOGLU, Ö. SİLİS DUMANI KATKILI BETONLARIN ÇARPMA DAYANIMININ YAPAY SİNİR AĞI İLE BELİRLENMESİ. Engineering Sciences, 3(1), 30-38.
- [31] Matsui, T. ve Furui, S. (1994) Comparison of Text-Independent Speaker Recognition Methods Using VQdistortion and Discrete/Continuous HMMs, IEEE Trans. On Speech and Audio Processing, 456-459
- [32] Yu. K, Mason. J. ve Oglesby, J. (1995) Speaker recognition using Hidden Markov Models, dynamic time warping and vector quantisation, IEE Proc.-Vis. Image Signal Process., 142(5), 313-318
- [33] Özcan, G. (2015). Saklı Markov Modelleri ve uygulamaları. Akademik Bilişim, 15, 1-10.
- [34] Khelifa, M. O., Elhadj, Y. M., Abdellah, Y., & Belkasmi, M. (2017). Constructing accurate and robust HMM/GMM models for an Arabic speech recognition system. International Journal of Speech Technology, 20(4), 937-949.
- [35] Erbayram, T. (2020). Boyut indirgeme tekniklerinin sınıflandırma perormanslarının karşılaştırılması (Master's thesis, Necmettin Erbakan Üniversitesi Fen Bilimleri Enstitüsü).

EKLER

ses_fisher.m

```
clear all
clc
load flda.mat
sinif_etiket={'0','1','2','3','4','5','6','7','8','9'};
%Y:Sınıflara ait etiket vektörü
Y=[];Y1=[];Y2=[];
for i=1:10
    k=sinif_etiket{i};
    for jj=1:20
        Y1{jj}=k;
    end
    Y=[Y Y1];
end

aa={'beyaz','kırmızı','küp','küre','mavi','prizma','sarı','silindir','siyah','yeşil'};
sozluk={'beyaz küp','beyaz prizma','beyaz silindir','beyaz küre',...
        'kırmızı küp','kırmızı prizma','kırmızı silindir','kırmızı küre',...
        'mavi küp','mavi prizma','mavi silindir','mavi küre',...
        'sarı küp','sarı prizma','sarı silindir','sarı küre',...
        'yeşil küp','yeşil prizma','yeşil silindir','yeşil küre',...
        'siyah küp','siyah prizma','siyah silindir','siyah küre'};

gonder={'a','b','c','d','e','f','g','h','i','j','k','l','m','n','o','p','r','s','t','u','v','w','y','z'};

Fs = 16000 ;
nBits = 16 ;
nChannels =1;
ID = -1;
recObj = audiorecorder(Fs,nBits,nChannels,ID);
disp('Start speaking.')
recordblocking(recObj,4);
disp('End of Recording.');
```

```

play(recObj);
y = getaudiodata(recObj);
audiowrite('ses.wav',y,Fs);
[segments, fs] = detectVoiced('ses.wav',2);
kk=[];
gg=[];
for i=1:length(segments)
audioIn=segments{i};
audioIn=audioIn-mean(audioIn);
[coeffs] = mfcc(audioIn,fs,'WindowLength',640,"OverlapLength",320);
r=[coeffs]';a=size(r,2);
if a>=20&&a<46
    l=a-20;ll=floor(l/2);
    rt=r(:,ll+2:1+ll+20);
elseif(a<20)
    b=20-a;v=zeros(42,b+1);
    rt=[r(:,2:end) v];
end
rx=rt(:)';
test=rx;te=test*Vpca;tee=te*Wf;
pc=fld-tee;bb=[];
for ii=1:200
    bb(ii)=norm(pc(ii,:));
end
[n1, n2]=min(bb); m=Y{n2};
n22=str2num(m);
gg(i)=n22+1;
end
%Seri porttan RC232 servo motor kartına veri gönderiliyor
serialport = serial('COM1');
set(serialport,'BaudRate',19200);
b=[aa{gg(1)} ' ' aa{gg(2)}];
for i=1:24

```

```
    r=sozluk{i};  
    if strcmp(r,b)  
        fopen(serialport);  
        fprintf (serialport,gonder{i});  
        fclose(serialport);  
    end  
end
```



ÖZGEÇMİŞ

Kişisel Bilgiler	
Adı Soyadı	Ozan Fırat ÇIPLAK
Doğum Yeri	Malatya
Doğum Tarihi	
Uyruğu	TC.
Telefon	
E-posta Adresi	
Eğitim Bilgileri	
Lisans	
Üniversite	
Fakülte	Mühendislik Fakültesi
Bölümü	Makine Mühendisliği
Mezuniyet Yılı	2017
Makale ve Bildiriler	
<i>Uluslararası Konferans ve Sempozyumlar</i> Ozan Fırat ÇIPLAK, Serkan KESER, Robot Arm Controlling With Real-time Speech Recognition Using Subspace Based Classifiers, ISPEC 10th International Conference on Engineering & natural sciences , pp. 247-256, Siirt University, Siirt, TURKEY, 16-18 May 2021.	