



**AKILLI EV ALETLERİ İÇİN KONUŞMACI BAĞIMLI AYRIŞIK SÖZCÜK
TANIMA SİSTEMİ**

Adem ÇİÇEK

**YÜKSEK LİSANS TEZİ
ELEKTRİK ELEKTRONİK MÜHENDİSLİĞİ ANABİLİM DALI**

**GAZİ ÜNİVERSİTESİ
FEN BİLİMLERİ ENSTİTÜSÜ**

AĞUSTOS 2015

Adem ÇİÇEK tarafından hazırlanan AKILLI EV ALETLERİ İÇİN KONUŞMACI BAĞIMLI AYRIŞIK SÖZCÜK TANIMA SİSTEMİ” adlı tez çalışması aşağıdaki jüri tarafından OY BİRLİĞİ ile Gazi Üniversitesi Elektrik Elektronik Mühendisliği Anabilim Dalında YÜKSEK LİSANS TEZİ olarak kabul edilmiştir.

Danışman: Doç. Dr. Özgül SALOR DURNA

Elektrik-Elektronik Mühendisliği Anabilim Dalı, Gazi Üniversitesi

Bu tezin, kapsam ve kalite olarak Yüksek Lisans Tezi olduğunu onaylıyorum

.....

Başkan : Doç. Dr. Nursel AKÇAM

Elektrik-Elektronik Mühendisliği Anabilim Dalı, Gazi Üniversitesi

Bu tezin, kapsam ve kalite olarak Yüksek Lisans Tezi olduğunu onaylıyorum

.....

Üye : Yrd. Doç. Dr. Seniha Esen YÜKSEL

Elektrik ve Elektronik Mühendisliği Anabilim Dalı, Hacettepe Üniversitesi

Bu tezin, kapsam ve kalite olarak Yüksek Lisans Tezi olduğunu onaylıyorum

.....

Tez Savunma Tarihi: 04/08/2015

Jüri tarafından kabul edilen bu tezin Yüksek Lisans Tezi olması için gerekli şartları yerine getirdiğini onaylıyorum.

.....
Prof. Dr. Şeref SAĞIROĞLU
Fen Bilimleri Enstitüsü Müdürü

ETİK BEYAN

Gazi Üniversitesi Fen Bilimleri Enstitüsü Tez Yazım Kurallarına uygun olarak hazırladığım bu tez çalışmada;

- Tez içinde sunduğum verileri, bilgileri ve dokümanları akademik ve etik kurallar çerçevesinde elde ettiğimi,
- Tüm bilgi, belge, değerlendirme ve sonuçları bilimsel etik ve ahlak kurallarına uygun olarak sunduğumu,
- Tez çalışmada yararlandığım eserlerin tümüne uygun atıfta bulunarak kaynak gösterdiğimi,
- Kullanılan verilerde herhangi bir değişiklik yapmadığımı,
- Bu tezde sunduğum çalışmanın özgün olduğunu,

bildirir, aksi bir durumda aleyhime doğabilecek tüm hak kayıplarını kabullendiğimi beyan ederim.

Adem ÇİÇEK

04.08.2015

AKILLI EV ALETLERİ İÇİN KONUŞMACI BAĞIMLI AYRIŞIK SÖZCÜK TANIMA SİSTEMİ

(Yüksek Lisans Tezi)

Adem ÇİÇEK

GAZİ ÜNİVERSİTESİ
FEN BİLİMLERİ ENSTİTÜSÜ

Ağustos 2015

ÖZET

Günümüzde oldukça sık karşılaştığımız akıllı ev aletleri, insan hayatını, özellikle yaşlı ve engelli olan kişiler için, oldukça kolaylaştırmaktadır. Bu tez çalışması kapsamında, bir akıllı ev aleti örneği alınarak, bunun üzerinde çalışmak üzere bir konuşmadan komut tanıma sistemi geliştirilmiştir. Sesli komut tanıma sisteminde, akıllı ev aletinin kullanımını sağlayacak 17 kelime içeren bir komut kümesi oluşturulmuştur. Sistemin eğitimi için, MATLAB ile bir kullanıcı arayüzü oluşturulmuş ve bu arayüz vasıtasıyla altısı bay dördü bayan toplam 10 kişiden alınan komutların her biri sekizer defa kaydedilmiş ve bir konuşmacı veritabanı oluşturulmuştur. Öznitelik olarak Mel Frekans Kepstral Katsayıları (MFCC - Mel Frequency Cepstral Coefficients) ve Çizgisel Spektrum Frekansları (LSF – Line Spectral Frequency) kullanılmış ve her konuşmacının bütün komutları beş tekrar için Saklı Markov Modelleri (HMM – Hidden Markov Models) kullanılarak modellenmiştir. Her konuşmacının geriye kalan üç komutu test için kullanılmış ve test sonuçları farklı öznitelikler için ayrı ayrı analiz edilmiştir. Sonuçlar, kelime, kişi ve genel sistem başarımları bazında karşılaştırılmıştır. Sonuç olarak akıllı ev aletleri için kullanıcının kendisinin yapacağı kısa bir eğitim ile yüksek tanıma başarımları ile çalışabilecek Türkçe bir komut tanıma sistemi geliştirilmiştir. 10 konuşmacının bütün komutlarının doğru tanıma yüzdesi MFCC öznitelikleri için %97,65 ve LSF öznitelikleri için %93,33 olarak elde edilmiştir. Akıllı bir ev aleti örneği olarak çamaşır makinası seçilerek, çalışma senaryosu oluşturulmuş ve bu senaryo üzerinden kullanıcı ile sesli iletişim kuran bir diyalog sistemi geliştirilmiştir. Bu sistem yeni bir kullanıcının sistemi kendi sesi için eğitip, akıllı ev aletini sesli komutlarla çalıştırmasını sağlamaktadır.

Bilim Kodu : 905.1.067

Anahtar Kelimeler : Akıllı Ev, Ayrışık Sözcük Tanıma, Çizgisel Spektrum Frekansları, Konuşma Tanıma, Mel Frekans Kepstral Katsayıları, Saklı Markov Modeli, Uzaktan Kontrol

Sayfa Adedi : 60

Danışman : Doç. Dr. Özgül SALOR-DURNA

SPEAKER DEPENDENT ISOLATED WORD RECOGNITION SYSTEM DESIGNED
FOR SMART HOME APPLIANCES

(M. Sc. Thesis)

Adem ÇİÇEK

GAZİ UNIVERSITY

GRADUATE SCHOOL OF NATURAL AND APPLIED SCIENCES

August 2015

ABSTRACT

Smart home appliances, which are encountered quite often nowadays, facilitate human life, considerably, especially for old and disabled people. Within the scope of this thesis, a voice-command recognition system has been developed to work upon a sample home appliance. For the voice-command recognition system, a 17-word command set has been prepared specific to a smart home appliance. For the training phase, commands spoken by ten people, six male and four female, have been recorded eight times for each command, and a voice database has been built with the help of an interface designed using MATLAB graphical user interface unit. Mel Frequency Cepstral Coefficients (MFCC) and Line Spectral Frequencies (LSF) have been used as features and five repetitions of each command of each speaker have been modelled using Hidden Markov Models (HMM). The remaining three repetitions of each speaker have been used for the test phase. Recognition performances have been compared on the basis of word, speaker and general system performance. In conclusion, a Turkish command recognition system, which a new user is able to make operative after a short training period with high recognition performance, has been realised in MATLAB software environment for any smart home appliance. The recognition performance of system evaluated separately for both features and has been found to be 97,65% with MFCCs, and 93,33% with LSFs. A scenario has been built for washing machine operation as the sample smart home appliance, and a dialogue system has been developed based on this scenario. The dialogue system lets a new user train the isolated word recognition system for his/her own voice and run the smart home appliance with high recognition performance.

Science Code : 905.1.067

Key Words : Hidden Markov Models (HMM), Isolated Word Recognition, Line Spectral Frequencies (LSF), Mel Frequency Cepstral Coefficients (MFCC), Smart Home, Speech Recognition, Remote Control

Page Number : 60

Supervisor : Assoc. Prof. Dr. Özgül SALOR-DURNA

TEŐEKKÜR

Çalıőma süreci boyunca yardımlarını hiçbir zaman esirgemeyen kıymetli hocam Doç. Dr. Özgöl SALOR-DURNA'ya,

Konuőma tanıma sistemlerinde en zor işlerden biri olan kayıt alma işleminde vakitlerini ayırarak veri seti oluőturmamda yardımcı olan çalıőma arkadaşlarıma,

Destekleriyle her zaman yanımda olan, beni moralde tutan aileme en içten dileklerle teşekkür ederim.

İÇİNDEKİLER

	Sayfa
ÖZET	iv
ABSTRACT.....	v
TEŞEKKÜR.....	vi
İÇİNDEKİLER	vii
ÇİZELGELERİN LİSTESİ.....	ix
ŞEKİLLERİN LİSTESİ.....	xiv
SİMGELER VE KISALTMALAR.....	xvi
1. GİRİŞ.....	1
2. KONUŞMA TANIMA SİSTEMLERİNİN GENEL YAPISI	7
2.1. Konuşma Tanıma Sistemleri	7
2.1.1. Konuşmacı bağımlı ve bağımsız sistemler.....	7
2.1.2. Sürekli ve ayrışık sistemler	8
2.1.3. Sözcük ve fonem tabanlı sistemler.....	8
2.1.4. Sözcük yakalama sistemleri	8
2.2. Konuşmanın Oluşumu ve Modellenmesi	9
2.3. Öznitelik Çıkarımı.....	11
2.3.1. Mel frekans kepstral katsayıları	11
2.3.2. Doğrusal öngörü katsayıları	12
2.3.3. Çizgisel spektrum katsayıları	13
2.4. Sınıflandırma.....	14
2.4.1. İstatistiksel modeller	14
2.4.2. Şablon modeller	23
3. GELİŞTİRİLEN AYRIŞIK SÖZCÜK TANIMA SİSTEMİ.....	27

Sayfa

3.1. Geliştirilen Konuşma Tanıma Sisteminin Genel Yapısı	27
3.2. Eğitim Aşaması	27
3.2.1. MATLAB GUI ile veritabanının oluşturulması.....	28
3.2.2. MFCC ve LSF yöntemleri ile özniteliklerin çıkarılması.....	30
3.2.3. Saklı markov modeli ile özniteliklerin eğitimi-model oluşturma	31
3.3. Test Aşaması	33
3.3.1. MFCC+HMM ile başarımlar.....	33
3.3.2. LSF+HMM ile başarımlar.....	34
4. AKILLI EV ALETLERİ İÇİN GELİŞTİRİLEN DİYALOG SİSTEMİ	37
4.1. Örnek Bir Ev Aleti – Çamaşır Makinesi ve Kullanılacak Komutlar	37
4.2. Akıllı Ev Aletinin Diyalog Sistemi için Akış Şeması	41
4.3. Akış Diyagramı Çıkarılan Akıllı Ev Aletinin MATLAB GUI’de Dialog Tanıma Sisteminin Gerçekleştirilmesi	42
5. SONUÇ VE ÖNERİLER	43
KAYNAKLAR	45
EKLER	49
EK -1. Konuşma tanıma sistemini test ederken kullanılan, her bir kullanıcıya ait üç test datası için elde edilen test sonuçları	50
ÖZGEÇMİŞ	60

ÇİZELGELERİN LİSTESİ

Çizelge	Sayfa
Çizelge 2.1. Konuşma tanıma sistemlerinin sınıflandırılması	7
Çizelge 2.2. Probleme ait durum geçiş matrisi	20
Çizelge 2.3. Adamın yanında şemsiye bulundurma olasılıkları.....	21
Çizelge 3.1. Durum geçiş matrisinin başlangıç değeri.....	31
Çizelge 3.2. Başlangıç vektörünün başlangıç değeri	32
Çizelge 3.3. DURULA komutuna ait eğitim sonrası durum geçiş matrisleri (Soldaki matris MFCC ile sağdaki LSF ile)	32
Çizelge 3.4. DURDUR komutuna ait eğitim sonrası durum geçiş matrisleri (Soldaki matris MFCC ile sağdaki LSF ile)	32

ŞEKİLLERİN LİSTESİ

Şekil	Sayfa
Şekil 2.1. Konuşmanın oluşumunda etkili vücut yapıları	9
Şekil 2.2. Konuşmanın oluşumu ve aktarımı	10
Şekil 2.3. Genel bir konuşma tanıma sisteminin eğitim aşaması.....	10
Şekil 2.4. MFCC katsayılarının elde edilme basamakları.....	11
Şekil 2.5. Probleme ait model	20
Şekil 2.6. Bir yapay sinir hücresinin çalışması	23
Şekil 2.7. DTW ile benzerlik ölçümü	24
Şekil 2.8. Öklit ile benzerlik ölçümü	24
Şekil 3.1. Veritabanını oluşturan konuşmacıların cinsiyete göre dağılımı	27
Şekil 3.2 Geliştirilen ayrışık kelime tanıma sisteminin eğitim aşaması	28
Şekil 3.3. Veritabanının oluşturulmasında kullanılan kullanıcı arayüzü	29
Şekil 3.4. Kullanıcı tarafından alınan örnek bir DURULA komutuna ait 13x46 MFCC öznitelik matrisi yüzey gösterimi	30
Şekil 3.5. Kullanıcı tarafından alınan örnek bir DURULA komutuna ait 13x46 MFCC öznitelik matrisi yüzey gösterimi	30
Şekil 3.6. Cinsiyete göre MFCC öznitelikleri için tanıma sisteminin performansı	33
Şekil 3.7. Kelime bazında MFCC öznitelikleri için tanıma sisteminin performansı	34
Şekil 3.8. Kişi bazında MFCC öznitelikleri için tanıma sisteminin performansı (Kullanıcı 1-6 : bay, Kullanıcı 7-10: bayan)	34
Şekil 3.9. Cinsiyete göre LSF öznitelikleri için tanıma sisteminin performansı	35
Şekil 3.10. Kelime bazında LSF öznitelikleri için tanıma sisteminin performansı	35
Şekil 3.11. Kişi bazında LSF öznitelikleri için tanıma sisteminin performansı (Kullanıcı 1-6 : bay, Kullanıcı 7-10: bayan).....	36
Şekil 4.1. Geliştirilen ayrışık kelime tanıma tabanlı diyalog sistemi	37

Şekil	Sayfa
Şekil 4.2. (a) Diyalog sisteminin ayrıntılandırılmış akış şeması: YIKAMA komutunun seçilme durumu	38
Şekil 4.2. (a) (devam) Diyalog sisteminin ayrıntılandırılmış akış şeması: YIKAMA komutunun seçilme durumu	39
Şekil 4.2. (b) Diyalog sisteminin ayrıntılandırılmış akış şeması: DURULA komutunun seçilme durumu	40
Şekil 4.2. (c) Diyalog sisteminin ayrıntılandırılmış akış şeması: ÇIKIŞ komutunun seçilme durumu	41
Şekil 4.3. Kullanıcının ev aletini Diyalog Tanıma Sistemi ile kullanabilmesi için tasarlanan arayüz	42

SİMGELER VE KISALTMALAR

Bu çalışmada kullanılmış simgeler ve kısaltmalar, açıklamaları ile birlikte aşağıda sunulmuştur.

Simgeler

Açıklamalar

dB

Desibel

kHz

Kilohertz

Kısaltmalar

Açıklamalar

DCT

Discrete Cosine Transform

DTW

Dynamic Time Warping

FFT

Fast Fourier Transform

FIR

Finite Impulse Response

GUI

Graphical User Interface

HMM

Hidden Markov Model

k-NN

k- Nearest Neighborhood

LBG

Linde Buzo Gray

LPC

Linear Predictive Coefficient

LSF

Line Spectral Frequency

MFCC

Mel Frequency Cepstral Coefficient

PLP

Perceptual Linear Prediction

RASTA

Relative Spectral

SNR

Signal to Noise Ratio

SVM

Support Vector Machine

VQ

Vector Quantization

YSA

Yapay Sinir Ağı

1. GİRİŞ

Gelişen teknolojiyle birlikte insanoğlu, onu anlayan, ihtiyaçlarının giderilmesini kolaylaştıran keşifler istemektedir. Bu keşiflerden en çok dikkat çekenlerden biri konuşma tanıma uygulamalarıdır.

Konuşma tanıma konusuna ilişkin geçmişte mekanik olarak [1] konuşma tanıma sistemi gerçekleştirme çalışmaları yapılmış ayrıca bilim kurgu meselesi olmuştur. Günümüzde ise, bilim kurgu olmaktan çıkarak, sayısal sinyal işleme ve geliştirilen algoritmalar ile yüksek tanıma doğruluğu olan uygulamalar gerçekleştirilmektedir [2]. Bu uygulamalar, adli soruşturmalar, dikte etme, gömülü sistemler, eğitim ve tıp alanı [3] gibi bir çok alanda kullanılmakta, özellikle engelli insanlar için çok faydalı olabilmekte ve onları daha rahat bir hayata kavuşturabilmektedir.

Konuşma tanıma, giriş olarak alınan konuşmanın sinyal işleme tekniklerini kullanarak ilgili metnin tanınmasıdır, yani veri tabanındaki hangi metin modeline karşılık geldiğinin bulunması işlemidir. Konuşma tanıma sistemi ise konuşma tanımayı yapmanın yanında çıkış olarak, tanınan metne bağlı komutu gerçekleştiren uygulamalardır.

Ses tanıma sistemlerini genel olarak konuşmacı açısından konuşmacı bağımlı ve konuşmacı bağımsız; metin açısından ise metne bağımlı ve metne bağımsız; konuşma sürekliliği açısından ise sürekli ve ayrışık şeklinde sınıflandırmak mümkündür [4].

Literatür incelendiğinde konuşma tanımanın; uygulamaların kullanılabilirliğini, doğruluğunu arttırmak için sistemlere entegre edildiği ve bunu yaparken var olan bir konuşma tanıma sisteminin hazır olarak kullanılabildiği [5-6] veyahut uygulama için yeni bir konuşma tanıma sisteminin gerçekleştirilerek başarımın değerlendirildiği veya başarımı arttıracak yöntemler, konuşma tanıma sistemleri için geliştirildiği görülmektedir.

Ozgunduz ve diğerleri, konuşma sinyali gibi akustik bir ses çıkaran iki tip araç için araç tanıma sistemi tasarlamıştır. Bu sistemde, sinyallerin karakteristik özelliklerini elde etmek için Mel Frekans Kepstral Katsayıları (MFCC - Mel Frequency Cepstral Coefficients) kullanılmış ve Vektör Nicemleme (Kuantalama) işlemi ile MFCC'leri indirlemiştir.

Tanıma işleminde ise, k-En Yakın Komşuluk (k-NN - Nearest Neighborhood) ve Destek Vektör Makineleri (SVM - Support Vector Machines) ayrı ayrı kullanılarak, SVM ve k-NN (k=5 için) yöntemleri ile en yüksek tanıma başarımı, ortalama %82 olarak elde etmiştir [7].

Karacı, özilinti fonksiyonu ile öznitelik çıkararak tanıma aşamasında örüntü tanımının kullanıldığı Türkçe için ayrışık bir tanıma sistemini Matlab ortamında gerçekleştirmiştir. Kelime bazında tanıma yüzdeleri en düşük ve en yüksek olarak sırasıyla bay sesi için %80-%100, bayan sesi için ise %70-%80 olarak bulunmuş ve veritabanındaki örnek sayısı arttığında bu başarıların arttığı gösterilmiştir [8].

Hsu ve Yang, bir ortamda bulunan bütün ev aletlerini, ses komutlarını kullanarak, konuşma tanıma mikroişlemcisi ve bluetooth içeren uzaktan bir kontrol mekanizmasıyla kontrol edebilecek ve kendi kendine öğrenen bir uygulama gerçekleştirmiştir [9].

Koç, Türkçe veritabanı TURTEL kullanılarak tasarlanan gürbüz bir ayrışık konuşma tanıma sistemi için akustik öznitelik analizi üzerine çalışma yapmış ve gürültülü ortamlarda RASTA-PLP(Relative Spectral - Perceptual Linear Prediction – Bağıl Spektral Algısal Doğrusal Tahmin) katsayıları ile, PLP ve MFCC için performans araştırmış ve daha yüksek bir performans elde etmek için MFCC, RASTA-PLP karışımı yeni bir öznitelik önermiştir. Sonuç olarak, temiz bir ortam için en iyi sonucu MFCCler vermiş; gürültülü (konvolüsyon gürültüsü olan) bir ortamda RASTA-PLP ile daha iyi performans elde edilmiştir. Önerilen yöntemin tek başına MFCC veya RASTA kullanmaya nazaran daha iyi sonuçlar verdiği gösterilmiştir [10].

Zeng ve diğerleri, ses ile çalışan kablosuz ev aletleri tasarlanmış ve performans değerlendirmesi yapmıştır. Saklı Markov modelleri (HMM – Hidden Markov Models) ile ayrışık ve sınırlı gramer modellenmiş bir konuşma tanıma sistemi gerçekleştirilmiş ve çeşitli gürültü şekilleri ve şiddetleri altında test edilmiştir. Oluşturulan tanıma sisteminin, test konuşmacıları için %85 ve üstünde cümle tanıma başarısı verdiği gösterilmiştir. Kelime tanıma başarısı ise, farklı sinyal gürültü oranı (SNR – signal noise ratio) değerlerinde arka planda beyaz gürültü, müzik ve konuşma olduğu durumlar için verilmiş ve SNR arttıkça doğrusal olarak tanıma doğruluğunun da arttığı ve tanımaya en büyük etkiyi beyaz gürültünün yaptığı belirtilmiştir [11].

Aşıl原因 ve Günel, MATLAB ortamında SVM ile hece tabanlı ayrışık sözcük tanıma sistemi geliştirmiştir. Öznitelik olarak MFCClerin kullanıldığı tanıma sistemi yaklaşık %91 tanıma doğruluğu göstermiştir [12].

Jung ve diğerleri geleneksel bir mikrofon ve mikrofon dizisi algoritması ile konuşma tanıma sistemi performanslarını incelemiş ve tek kanal geleneksel mikrofon için %63 iken diğer yöntemler yaklaşık %93 başarımla elde etmişlerdir [13].

Kelebekler ve İnal MFCC, otokorelasyon analizi ve kovaryans metotları ile öznitelikler elde edip bu örüntüler kullanarak kişi bağımlı tanıma sistemi oluştururken öklit yönteminden faydalanmışlardır. Sistem, bay için %85,33; bayanlar için %91,33 tanıma performansı göstermiştir [14].

Priyadarshani ve diğerleri, öznitelik olarak MFCC kullanılan ve eğitimi DTW ile Sri Lanka yerli dili için ayrışık konuşma tanıma sistemi gerçekleştirmiştir. Ayrıca bu çalışmada sesin zaman içerisinde değişimini değerlendirmek için üç oturum şeklinde bir kez kayıt alındıktan sonra üç ay ve bir sene sonrasında kayıtlar alınarak başarımlar değerlendirilmiştir. İlk oturumda dört konuşmacı için %96,49 ile %99,05 ve ikinci oturumda %92,03 ile %95,35 arasında değişirken üçüncü oturumda tanıma başarımları %88,34 ile %91,66 arasında değişmektedir. Zaman içerisinde ses tellerinin değişimi sebeplerinden dolayı tanıma oranlarında açıkça düşüş görülmektedir [15].

Günel ve diğerleri, ortak vektör yaklaşımı ile sınıflandırma yaparken öznitelik olarak MFCC'leri kullanmış ve eğitim ile test aşamalarında TI-DIGIT veritabanını kullanmıştır. Ses örneklemelerinin ilk on ve yirmi çerçevesinin ortak vektör yaklaşımı yöntemine tabi tutulması durumları için tanıma performansları elde edilmiştir. Kişi bağımsız gerçek zamanlı bir rakam tanıma sistemi gerçekleştirmiştir [16].

Chang ve diğerleri, öznitelik olarak MFCC kullanılan ve eğitimi DTW ile gerçekleştirilen bir konuşma tanıma sistemi için en düşük %90 en yüksek %95 tanıma başarımları elde etmiştir. Bu tanıma sistemi saç kurutma makinesi, çamaşır makinesi, klima, televizyon, mikrodalga, masa lambası ve fan ev aletlerini isimleriyle çalıştırmayı amaçlamıştır. Tanıma oranları gömülü sistemin, konuşularak söylenen bu cihaz isimlerini ne düzeyde tanıdığını göstermektedir [17].

Zeng ve diğeri, fonem tabanlı HMM'ler ile MFCC öznitelikleri kullanılarak kişi bağımsız bir tanıma sistemi gerçekleştirmiştir. Yerli İngilizce konuşan, Doğu Hint aksanı, Endonezya aksanı ve Çin aksanı konuşan kişiler ile sistem farklı SNR değerlerinde çeşitli gürültü ortamları içerisinde test edilmiştir. Sistem, ortam bazlı olarak sinyal gürültü oranı (SNR – signal to noise ratio) = 0 desibel (dB) olduğunda beyaz gürültüde tanıma oranı %40'ın altındadır. Bazı kullanıcılar için %0'lara düşmektedir. SNR artıkça tanıma başarısı artmaktadır. SNR = 25 dB değerinde konuşmacılar için tanıma başarımları %75 ile %94 arasında değişmektedir [18].

Bülbül ve Karacı, özilinti fonksiyonunu kullanarak öklit yöntemi kullanarak, kriter olarak %85 benzerliğe dikkate alan, kişi bağımsız tanıma yapan bir tanıma sistemi ile bayanlar için %70-%90, baylar için %80-%100 arasında tanıma başarımları elde etmiştir [19].

Meena ve diğeri, bulanık mantık ve sinir ağlarını kullanarak yeni bir yöntem ile energy entropisi, kısa zaman enerjisi ve sıfır geçiş özniteliklerinden faydalanarak konuşma tanıma ile cinsiyet sınıflandırmasını gerçekleştirmiştir. Cinsiyet sınıflandırılmasında, bulanık mantık ve sinir ağları çıkış değerlerinin ortalama değeri alınıp belirlenen bir eşik (threshold) değeri kriterine göre karar verilir [20].

Yalçın ve Ülker, görme engelliler için ses analizi ile bir e-posta iletimi yapabileceği c# (csharp) programlama dili kullanılarak bir paket program tasarlamıştır [21].

Iwasawa ve diğeri, konuşmanın ve el kol hareketi davranışlarının tamamlayıcı olduğunu söylemektedir. Bu yüzden insan ve robot arasında doğal bir etkileşimi amaçlamış ve hem konuşma tanıma hem de el kol hareket davranışı tanıma yapan çok modlu bir etkileşim birimi olan, nesne pozisyonunu kullanarak çalışan bir ev aleti hizmet sistemi önermiş ve uygulamasını yapmıştır [22].

Nandyala ve Kumar, dinamik programlama ile eğitim aşaması HMM, YSA ve SVM'lere nazaran çok kısa sürecek DTW algoritmasını gerçekleştirerek öznitelik olarak MFCC'leri kullanıldığı %82 tanıma doğruluğu gösteren bir konuşmacı bağımlı Hintçe ayrışık konuşma tanıma sistemi gerçekleştirmiştir [23].

Soda ve diğeri, içerisinde ev aletlerinin bulunduğu bir ev için Japonca ses ile kontrol edilebilen bir ev ağı sistemi geliştirmiştir. Böyle bir sistemde sadece ses kullanılacağı için etkileşimler yani giriş olarak alınan sesler mekanik, bilgi taşımayan şeyler olabilecektir. Bu noktada bilgi taşıyan girişleri almak için yöntem olarak mikrofona dizileri kullanma, giriş karşılık ne algılandığını ifade eden geribesleme veren web hizmeti şeklinde sanal bir arayüz birimi (virtual agent) tasarlanmıştır. Ses tanıma için Julies konuşma tanıma motoru hazır olarak kullanılmış ve ortalama %64 tanıma performansı elde etmişlerdir. Tanıma doğruluğu komutlara bağlı olarak çokça değişmekte ve düşük doğrulukta olan komutlar araştırıldığında benzer telafuza sahip komutların yanlış tanındığı çıkarımı yapılmaktadır. Örneğin Japonca fan demek olan “senpuuki” sık sık bütün anlamına gelen “zenbu” ile karıştırılmaktadır [24].

Bu tez çalışmasında da Türkçe dilinde akıllı bir ev aleti için konuşmacıya özgü komut tanıma sistemi, yani konuşmacı bağımlı ayrışık sözcük tanıma sistemi geliştirilmiş ve bu sistemin bir ev aleti için MATLAB grafik kullanıcı arayüzü (GUI – Graphical User Interface) kullanılarak uygulaması yapılmıştır. Arayüz yardımı ile kullanıcı, cihazı yüksek performansla kullanabilmesi için kısa bir eğitim veritabanı oluşturacaktır. Bu veritabanı o kişinin ismiyle sisteme kaydedilecektir. Sistemi çalıştırma aşamasında ise ilk önce kullanıcı kendisine ait eğitim dosyasını seçecek ve sonrasında sistem o kişi için çalışmaya diyaloglar ile komut bekleme modunda olacaktır. Hiçbir şey konuşulmadığında veya algılanamadığında sistem, tekrardan diyalog sistemi ile kullanıcıya algılayamadığını ifade edecek ve komut bekleme moduna geçecektir.

2. KONUŞMA TANIMA SİSTEMLERİNİN GENEL YAPISI

2.1. Konuşma Tanıma Sistemleri

Konuşma tanıma sistemlerinde genel olarak tanıma süreci şu şekilde işlemektedir: Gelen metin, öznitelikleri çıkarıldıktan sonra veritabanında eğitilerek, modelleri oluşturulup saklanan metin modellerinden yüksek benzerlik verdiği modelin sınıfından olduğu, sonuca göre tespit edilir veyahut yine sonuca göre algoritmalar ve eşik değerleri yardımıyla sistemde böyle bir kelime olmadığı çıktısı verilir.

Konuşma tanıma sistemlerinin sınıflandırılması Çizelge 2.1’de özetlenmiştir [4]. Ayrıca Sözcük Yakalama Sistemleri de bu sınıflandırmaya eklenebilir.

2.1.1. Konuşmacı bağımlı ve bağımsız sistemler

Konuşmacı bağımlı sistemlerde tek bir konuşmacı için referans şablonları oluşturulur. Farklı konuşmacıları da tanıması için onların da referans şablonları oluşturulmalı ve veritabanı güncellenmelidir [25]. Konuşmacı bağımsız sistemler ise herhangi birinin konuşmasında tanınmanın yapılabildiği sistemlerdir. Bu sistemlerde yüksek başarımlı elde etmek, konuşmacı bağımlı sistemlere göre zordur ve bunun için herkesin ses kayıtlarını almak mümkün olmadığından, genellikle fonem tabanlı yöntemler kullanılarak konuşmacıların tanınması yoluna gidilmektedir.

Çizelge 2.1. Konuşma tanıma sistemlerinin sınıflandırılması

Konuşmacıya Göre	Konuşmacı Bağımlı Sistemler
	Konuşmacı Bağımsız Sistemler
Konuşmaya Göre	Sürekli Sistemler
	Ayrışık Sistemler
Ses Birimine Göre	Sözcük Tabanlı Sistemler
	Fonem Tabanlı Sistemler

2.1.2. Sürekli ve ayrışık sistemler

Sürekli konuşma tanıma sistemleri kelimeler arasında duraksama olmaksızın yapılan konuşmaların tanındığı sistemlerdir. Bu sistemler için kelime sınırlarını tayin etmek, konuşmacının hızı ve yanlış telaffuzlar karşılaşılan zorluklar arasındadır [10].

Ayrışık konuşma tanıma sistemleri kelimelerin arasında net bir duraksamanın olduğu, dolayısıyla kelimeler arasını belirlemenin gerek olmadığı daha önceden oluşturulan modellere göre tanımanın yapıldığı sistemlerdir. Bu sistemler akıllı ev aletleri vb. uygulamalar için kullanılabilir uygun sistemlerdir.

2.1.3. Sözcük ve fonem tabanlı sistemler

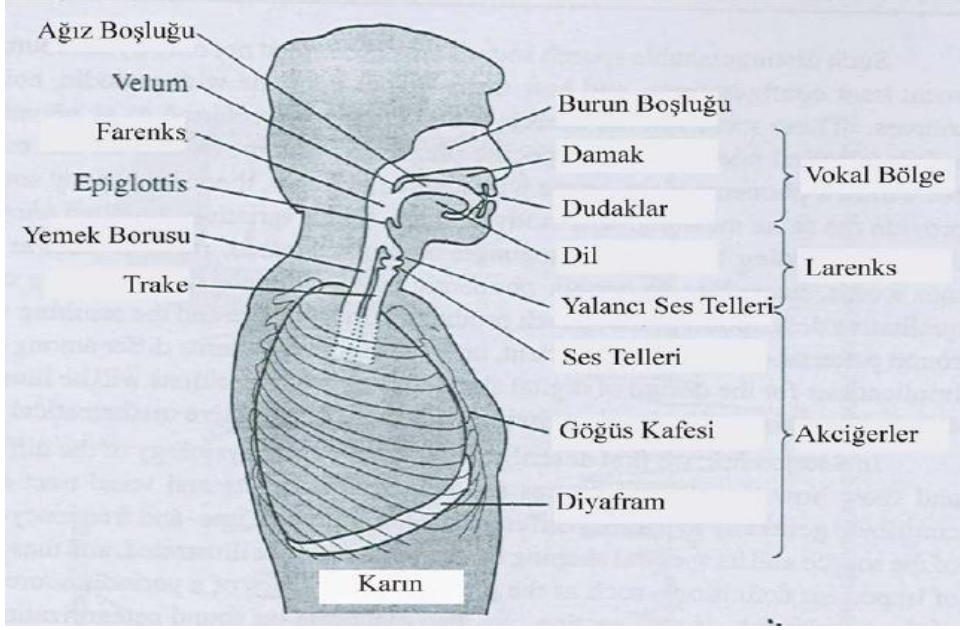
Sözcük tabanlı sistemler, en küçük tanıma birimi olarak sözcüklerin alındığı sistemlerdir. Bu sistemler sözcük kayıtları yapıldığı için fazla bellek ihtiyacı duyarlar [25].

Fonem tabanlı sistemler ise en küçük tanıma birimi olarak fonemleri kullandığı için daha az miktarda bellek işgal ederler, fakat tanıma doğrulukları sözcük tabanlı tanıma sistemlerine göre düşüktür. Bu performans düşüklüğü kelime ve kelime altı yapıların arasındaki geçişlerden kaynaklanmaktadır [10]. Bir tek fonem değil de ikili, üçlü fonem yapıları kullanarak bu performans fonem tabanlı sistemler için arttırılabilmektedir [25].

2.1.4. Sözcük yakalama sistemleri

Literatürde anahtar yakalama sistemleri olarak da geçen sözcük yakalama sistemleri istenilen sözcüğün metin içinden tespit edilmesini sağlayan sistemlerdir.

Bir konuşma sinyalinde bilindiği gibi sessizlik ve konuşma metinleri bulunmaktadır. Bu konuşma metninde ise istenilen metin –anahtar kelime- yönüyle bakıldığında anahtar kelime ve diğer kelimeler bulunmaktadır. Sözcük yakalama sistemlerinde önemli bir şekilde performansı etkileyen bu sözcük dışı kelimelerin de modellenerek sistemin oluşturulmasına ihtiyaç olacaktır. Bu modelleme çöp (garbage) modelleme olarak bilinir [26]. Bu sistemler sözlü basında, firma veya bazın şahısların hakkında geçen metinlerin bulunması gibi özel araştırmalarda kullanılabilir.



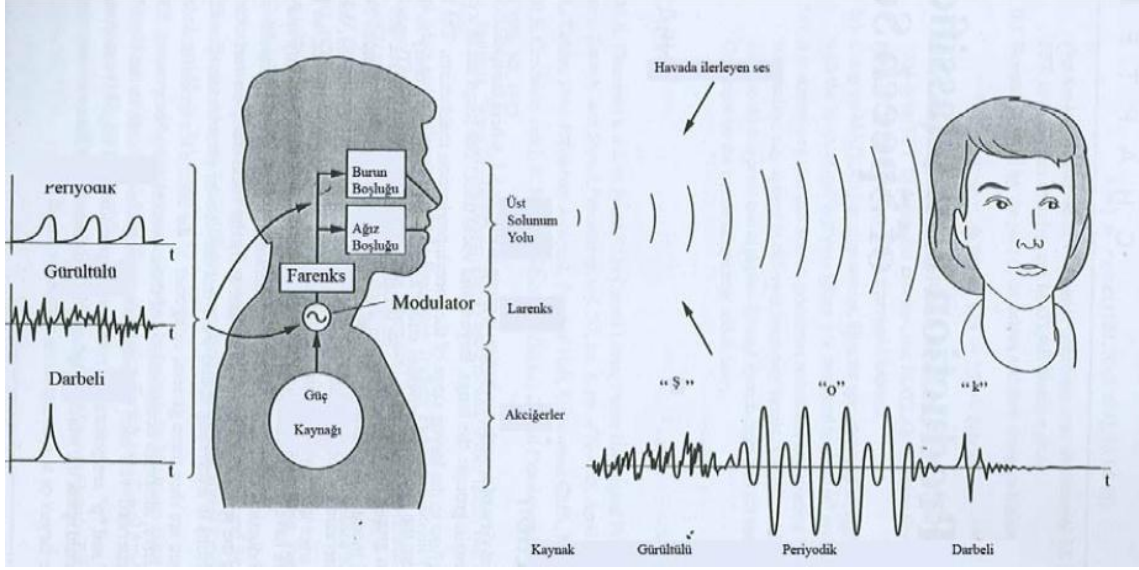
Şekil 2.1. Konuşmanın oluşumunda etkili vücut yapıları [1]

2.2. Konuşmanın Oluşumu ve Modellenmesi

Ses dalgası bir basınçtır ve bu basınç diyaframla başlayıp akciğer, soluk borusu, ses tellerini takiben ağız boşluğu ve son olarak dudaklarla şekillenerek dış ortama gönderilir [27]. Konuşmanın oluşması, beyinde anlamlı ifade oluştuktan [27] sonra Şekil 2.1'deki ses mekanizmasının uygun bir şekilde çalıştırılması sonucu meydana gelir.

Konuşma oluşumu ve karşı tarafa aktarımı da Şekil 2.2'de model olarak gösterilmiştir. Hava gönderen bir güç kaynağının gönderdiği sinyal, ses üretim yolunda (vocal tract) vücut yapıları tarafından değişimlere uğratılarak konuşma sinyali meydana gelir ve karşı tarafa gönderilir [27]. Karşı tarafın anlaması kulak yapısının işleyişi ve beynin anlamlandırması ile olur.

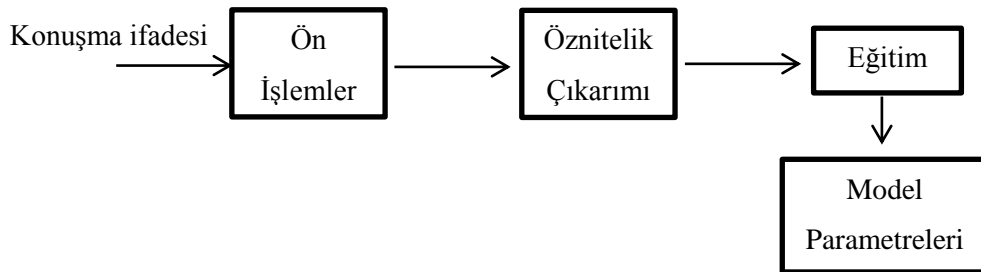
Sessel yapılara göre oluşturulan bir model, akustik model olarak bilinir. Bu modelin yanısıra anlam dikkate alınarak oluşturulan anlamsal ve grammer dikkate alınarak oluşturulan dilsel modeller ile de konuşmalar modellenmektedir. Modeli oluşturulacak konuşma ifadeleri, bir çok farklı şekilde temsil edilmektedir. Bunlara öznitelik vektörleri denilmektedir.



Şekil 2.2. Konuşmanın oluşumu ve aktarımı [1]

$$\bar{W} = \arg_{\max} P(W|A) = \frac{P(W)P(A|W)}{P(A)} \quad (2.1)$$

Eş. 2.1 konuşma tanımada çözülmesi gereken bir problemdir. Bu eşitlikte $P(W|A)$, A konuşması yani gözlem vektörleri bilindiğinde W kelimesinin görülme olasılığı; $P(A)$, kaydedilen A konuşmasının görülme olasılığı, $P(A|W)$ ise W kelimesi olduğu bilindiğinde onun A kelimesinin görülme olasılığı ve son olarak $P(W)$, W kelimesinin görülme olasılığıdır [28]. Sonuç olarak verilen gözlem vektörleri için tanınan W kelimesi hem $P(W)$ dilsel modeline hem de $P(A|W)$ akustik modeline bağlı olmaktadır. Bizim çalışmamızda dilsel model olmadığı için sadece akustik modelleme yapılmaktadır.



Şekil 2.3. Genel bir konuşma tanıma sisteminin eğitim aşaması

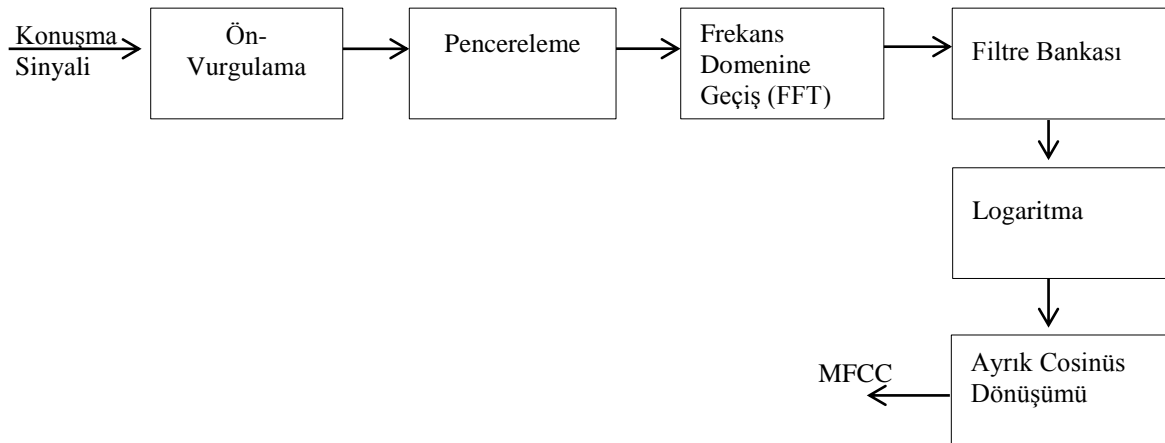
2.3. Öznitelik Çıkarımı

Öznitelik çıkarımı konuşma tanıma sistemlerinde önemli bir aşamadır. Bu aşamada kelimeleri temsil edecek katsayıların bulunması hedeflenmektedir.

Öznitelik çıkarımı genellikle tersinir olmayan kayıplı bir yöntemdir yani özniteliklerden tekrardan orijinal sinyali elde etmek mümkün değildir. Konuşma tanımada bu kayıplı dönüşümü kabul etmeyi zorlayan etkenlerden temel olarak birincisi hesaplamasal karmaşıklık ikincisi de gürbüzlüktür. Parametre sayısını artırarak doğruluğunu artırma işlemi, hesaplamasal karmaşıklığı arttıracak ve nihayetinde de gürbüzlük azalacağından sistemi daha iyi bir sonuca götürmeyecektir. Ayrıca katsayıların artması eğitim aşamasını da uzatacaktır [29].

2.3.1. Mel frekans kepstral katsayıları

İnsan kulağının işitmesinin örnek alındığı öznitelik çıkarma yöntemi, konuşma tanıma sistemleri için en çok kullanılan ve en fazla başarımlı veren algoritmalarından biridir. Şekil 2.4'te MFCC'lerin elde edilme basamakları gösterilmektedir. MFCC'ler insan işitmesini model alan mel skalayı kullanarak öz nitelikleri oluşturur. Tabii ki elde edilme aşamasında bir dizi işlem bloklarında geçmektedir. Başlangıç olarak çerçevelere bölünen konuşma sinyalinin herbir elemanı ön vurgulama filtresinden geçirilir. Bu filtre aslında bir sonlu darbe cevabı (FIR-Finite Impuls Response) bir filtredir ve genellikle Eş. 2.2'deki formül ile ifade edilir.



Şekil 2.4. MFCC'lerin elde edilme basamakları

$$H(z) = 1 - \alpha z^{-1} \quad (2.2)$$

Bu filtrenin amacı sinyal içerisindeki yüksek frekans bileşenlerini vurgulamaktır yani daha baskın hale getirmektir [30]. Buradaki α ifadesi sıfır ile bir arasında değer almakta ve genellikle 0,95 değeri kullanılmaktadır.

Bu işlemden sonra çerçeveye pencereleme işlemi uygulanır. Bunun amacı ise çerçevenin süreksizlik kısmı olan başı ve sonu frekans bölgesinde de spektrumu bozacaktır. Bu etkiyi minimize etmek için çerçeveler pencereleme işleminden geçirilir. Sonraki aşama pencerelenmiş her bir çerçevenin hızlı fourier dönüşümünü (FFT - Fast Fourier Transform) alarak frekans domenine geçme işlemidir. Böylece artık sinyal mel filtre bankasından geçme işlemine hazır olacaktır. Filtre bankasının çıkışının logaritmasının alınmasının sebebi çıkışların istatistiksel yapısının normal dağılıma yaklaşması içindir [27]. Son olarak çerçeve ayrık kosinüs dönüşümünden (DCT - Discrete Cosine Transform) geçirilir. Bu işlemin yapılmasının nedeni sadece bilginin toplandığı bölgeyi alarak boyut azaltmak ve sinyali doğrusal bağımsız olarak ifade etmektir. Sonuçta MFCC'ler öznitelik olarak hazır olacaktır.

2.3.2. Doğrusal öngörü katsayıları

Doğrusal Öngörü Katsayılarının (LPCs - Linear Predictive Coefficients) ortaya çıkmasının temel mantığı konuşulan ifadenin her bir örneği, geçmiş p adet örneklerin belli katsayılar ile çarpılarak doğrusal toplamı şeklinde ifade edilebileceğidir ve bu Eş. 2.3 ile gösterilmektedir. s'_n n . örnek olmak üzere, p adet önceki örneklerin doğrusal kombinasyonu ve $e[n]$ tahmin hata sinyalinin toplamı ile elde edilir.

$$s'_n = \sum_{i=1}^p a_i s_{n-i} + e[n] \quad (2.3)$$

Geçmiş örneklerdeki tahmin edilmeye çalışılan katsayılar yani a_i , LPC'lerdir. Eş. 2.3'ün transfer fonksiyonu da tamamen kutuplardan oluşan (all-pole) bir filtredir ve matematiksel ifadesi Eş. 2.4'teki gibidir.

$$H(z) = \frac{1}{1 - \sum_{i=1}^p a_i z^{-i}} \quad (2.4)$$

Kayıpsız bir borunun transfer fonksiyonu da bu şekilde all-pole bir filtredir ve boğaz yapısı her ne kadar kayıpsız ve silindirler biçiminde olmasa da bu model konuşulan ifadenin örneklerini elde edebilecek bir yaklaşımdır [31].

2.3.3. Çizgisel spektrum katsayıları

Fumitada Itakura tarafından geliştirilen çizgisel spektrum katsayıları (LSPs - Line Spectrum Pairs) doğrusal öngörü katsayıları tabanlı olup konuşma kodlamada çok kullanılmakta ve ve nicemleme gürültüsü açısından da daha az gürültüye sahiptir.

LSF'ler, LPC'ler gibi aynı bilgi içermekte fakat gösterim açısından avantajları bulunmaktadır. Eş. 2.5'ten LSF'ler türetilir. Burada $A(z)$, LPC analiz transfer fonksiyonudur. Filtre derecesi herhangi yeni bir bilgi eklemeksizin $P+1$ ' e genişletilebilir [32].

$$A(z) = 1 - \sum_{i=1}^p a_i z^{-i} \quad (2.5)$$

$$P(z) = A(z) - z^{-(p+1)} A(z^{-1}) \quad (2.6)$$

$$Q(z) = A(z) + z^{-(p+1)} A(z^{-1}) \quad (2.7)$$

$$A(z) = \frac{1}{2} (P(z) + Q(z)) \quad (2.8)$$

$P(z)$ ve $Q(z)$ polinomları, ses tellerinin tamamen açık veya kapalı olduğu zaman ses yolunun kayıpsız bir boru şeklinde gösterimi olarak düşünülebilir [32]. Bu polinomların sıfırları birim çember üzerine dağıldığı için polinomların kökleri e^{jw_i} olmakta ve w_i 'ler de LSF'ler olarak bilinmektedir [33].

2.4. Sınıflandırma

Konuşma sinyalinin oluşturulmasından sonra öznitelikler elde edilmesi ile model oluşturulması aşamasına gelinmektedir. Bu modellemelere göre de sınıflandırma işlemi yapılmaktadır.

2.4.1. İstatistiksel modeller

İstatistiksel modeller karmaşık matematiksel alt yapıya sahip olmakla birlikte güçlü sınıflandırma yöntemleridir. Eğitim aşamasında her bir kelime için istatistiki model oluşturulup test aşamasında gelen konuşmanın hangi modele daha çok uyduğu olasılıksal olarak çıkış verilir.

$$P(Y|\lambda^s) = p(y_1, y_2, \dots, y_l | \lambda^s) \quad s = 1, 2, 3, \dots, N \quad (2.9)$$

Eş. 2.9'da Y kelimeye ait gözlem vektörleri yani öznitelik vektörleri, λ^s ise N adet kelime modelini temsil etmektedir. Bu olasılık saklı markov modelleri, yapay sinir ağları ile çözülebilir [30].

Saklı markov modeli

Saklı markov modelini anlatmadan önce markov zincirinden bahsetmek daha iyi olacaktır. Markov zincirlerinde durumlar bizim çıkışlarımızdır ve bu çıkışlarımız önceki durumlara bağlı olabilmektedir. Birinci dereceden bir markov zincirinde ise sonraki durum sadece bulunan duruma bağlıdır. Bir Markov zinciri için model Eş. 2.10'daki gibi tanımlanır [28].

$$\lambda = \{S, A, \pi\} \quad (2.10)$$

Burada S durum kümesini, A durumlar arası geçiş matrisini, π ise başlangıç durum vektörünü temsil etmektedir. Eş. 2.10'daki parametrelerden geçiş matrisi (a_{ij}), Eş. 2.11-14'teki eşitlikleri sağlamalıdır [28]. q_t , t anındaki durumdur. S_i , i durumunda olduğunu göstermektedir.

$$A = [a_{ij}] \quad (2.11)$$

$$a_{ij} = P((q_t = S_j | q_{t-1} = S_i)), \quad 1 \leq i, j \leq N \quad (2.12)$$

$$a_{ij} \geq 0, \quad \forall i, j \quad (2.13)$$

$$\sum_{j=1}^N a_{ij} = 1, \quad \forall i \quad (2.14)$$

Başlangıç durum vektörü π , Eş. 2.15-17'deki eşitlikleri sağlamalıdır [28].

$$\pi_i = P(q_1 = S_i), \quad 1 \leq i \leq N \quad (2.15)$$

$$\pi_i \geq 0, \quad \forall i \quad (2.16)$$

$$\sum_{i=1}^N \pi_i = 1 \quad (2.17)$$

Saklı markov modellerinde ise durumlar belli yani gözlenebilir değildir ve çıkış olarak sadece gözlem vektörleri vardır. Saklı markov modeli Eş. 2.18'deki gibi tanımlanır [28].

$$\lambda = \{N, M, A, B, \pi\} \quad (2.18)$$

N durum sayısını,, M ayrık gözlem sembolleri sayısı, A durumlar arası geçiş olasılık matrisi, B gözlem sembollerinin olasılık dağılımları ve π başlangıç durum olasılık vektörüdür. B gözlem sembollerinin olasılık dağılımları Eş. 2.19 ile tanımlanmaktadır. t anında j durumunda k sembolünün üretilme yani $O_t = v_k$ olma olasılığı anlamına gelmektedir. Eş. 2.20-21 ile verilen eşitlikleri sağlamalıdır.

$$b_j(k) = P(O_t = v_k | q_t = S_j) \quad 1 \leq j \leq N \quad (2.19)$$

$$b_j(k) \geq 0, \quad \forall j, k \quad 1 \leq k \leq N \quad (2.20)$$

$$\sum_{k=1}^M b_j(k) = 1, \quad \forall j \quad 1 \leq j \leq N \quad (2.21)$$

İleri-geri algoritması

İleri – Geri algoritması (Forward –Backward Algorithm) ile model verildiğinde sıralı gözlem dizisi için olasılık hesaplanır. $P(O|\lambda)$ hesabı yapılır. Bu durumda sıralı gözlem dizilerini veren bütün durum dizileri için toplam olasılık hesabı yapılır. Bulunan olasılık değeri model başarısı olarak değerlendirilir. İleri algoritması o_1, \dots, o_t kısmi sıralı gözlem dizisi için t anında s_i durumunda olma olasılığını verir.

$$\alpha_t(i) = P(o_1 o_2 \dots o_t, q_t = s_i | \lambda) \quad (2.22)$$

Başlama:

$$\alpha_1(i) = P(o_1, q_1 = s_i) = \pi_i b_i(o_1), \quad 1 \leq i \leq N \quad (2.23)$$

İleriye doğru çıkarım:

$$\begin{aligned} \alpha_{t+1}(j) &= P(o_1 o_2 \dots o_{t+1}, q_{t+1} = s_j) = \\ &= \sum_i P(o_1 o_2 \dots o_{t+1}, q_t = s_i, q_{t+1} = s_j) = \\ &= \sum_i P(o_1 o_2 \dots o_t, q_t = s_i) a_{ij} b_j(o_{t+1}) = \\ &= [\sum_i \alpha_t(i) a_{ij}] b_j(o_{t+1}), \quad 1 \leq j \leq N, 1 \leq t \leq T-1 \end{aligned} \quad (2.24)$$

Sonlandırma:

$$P(o_1 o_2 \dots o_T) = \sum_i P(o_1 o_2 \dots o_T, q_T = s_i) = \sum_i \alpha_T(i) \quad (2.25)$$

Geri algoritması t anında s_i durumunda iken o_{t+1}, \dots, o_T kısmi sıralı gözlem dizisinin gerçekleşme olasılığını verir. Aslında iki algortimadan da birkaç işlem sonucunda aynı sonuca ulaşmak mümkündür fakat Baum-Welch algoritması için bu iki algoritmaya ihtiyaç vardır.

$$\beta_t(i) = P(o_{t+1} o_{t+2} \dots o_T | q_t = s_i, \lambda) \quad (2.26)$$

Başlama:

$$\beta_T(i) = 1 \quad 1 \leq i \leq N \quad (2.27)$$

Geriye doğru çıkarım:

$$\begin{aligned} \beta_t(i) &= P(o_{t+1} o_{t+2} \dots o_T | q_t = s_i) = \\ &= \sum_j P(o_{t+1} o_{t+2} \dots o_T, q_{t+1} = s_j | q_t = s_i) = \\ &= \sum_j P(o_{t+2} o_{t+3} \dots o_T | q_{t+1} = s_j) a_{ij} b_j(o_{t+1}) = \\ &= \sum_j \beta_{t+1}(j) a_{ij} b_j(o_{t+1}), \quad 1 \leq i \leq N, 1 \leq t \leq T-1 \end{aligned} \quad (2.28)$$

Sonlandırma:

$$\begin{aligned} P(o_1 o_2 \dots o_T) &= \sum_i P(o_1 o_2 \dots o_T, q_1 = s_i) = \\ &= \sum_i P(o_1 o_2 \dots o_T | q_1 = s_i) P(q_1 = s_i) = \sum_i \beta_1(i) b_i(o_1) \pi_i \end{aligned} \quad (2.29)$$

Viterbi algoritması

Viterbi algoritması mantık olarak ileri – geri algoritmasından farklı olmamakla birlikte bu algoritma ile yine model verildiğinde sıralı gözlem dizisi için maksimum olasılığı veren sıralı durum dizisi bulunur yani hangi sıralı durum dizisi en büyük olasılıkla sıralı gözlem dizisini bize veriyor sorusuna cevap verir. Aranılan şey $P(Q | o_1 o_2 \dots o_T)$ 'i veya $P(Q, o_1 o_2 \dots o_K)$ 'i maksimum yapan $Q = q_1 \dots q_T$ sıralı durum dizisidir.

$$\begin{aligned} \delta_t(j) &= \max P(q_1 \dots q_{t-1}, q_t = s_j, o_1 o_2 \dots o_t) = \\ &= \max_i [a_{ij} b_j(o_t) \max P(q_1 \dots q_{t-1} = s_i, o_1 o_2 \dots o_{t-1})] \end{aligned} \quad (2.30)$$

Burada s_j durumuna en yüksek olasılıkla s_i durumundan ulaşıyorsa s_i bilgisini kaydetmek (backtracking işlemi) gerekecektir.

Başlama:

$$\delta_1(i) = \max P(q_1 = s_i, o_1) = \pi_i b_i(o_1), \quad 1 \leq i \leq N \quad (2.31)$$

İleriye doğru çıkarım:

$$\begin{aligned} \delta_t(j) &= \max P(q_1 \dots q_{t-1}, q_t = s_j, o_1 o_2 \dots o_t) = \\ &= \max_i [a_{ij} b_j(o_t) \max P(q_1 \dots q_{t-1} = s_i, o_1 o_2 \dots o_{t-1})] = \\ &= \max_i [a_{ij} b_j(o_t) \delta_{t-1}(i)], \quad 1 \leq j \leq N, 2 \leq t \leq T \end{aligned} \quad (2.32)$$

$$\psi_t(j) = \arg \max_i [a_{ij} b_j(o_t) \delta_{t-1}(i)] \quad (2.33)$$

Sonlandırma:

$$i_T^* = \arg \max_i [\delta_T(i)] \quad (2.34)$$

Geriye gitme (Backtracking):

$$i_t^* = \psi_t(i_{t+1}^*) \quad (2.35)$$

Baum-welch algoritması

Baum – Welch Algoritması model parametrelerinin kestirimi için saklı markov modelde kullanılır. Kestirimi yapılan model parametreleri başlangıç durum vektörü, durum geçiş matrisi ve gözlem matrisidir (Eş. 2.36-28). λ' , yeni model parametrelerine sahip model olmak üzere, model parametreleri için parametre kestirimi ile $P(O|\lambda') > P(O|\lambda)$ veren parametreler modelin güncel parametreleri olacaktır. Bununla birlikte bu algoritma global bir maksimumu garanti etmez.

$$a_{ij} = P(s_j | s_i) \quad (2.36)$$

$$b_j(k) = P(v_m | s_j) \quad (2.37)$$

$$\pi_i = P(s_i) \quad (2.38)$$

$o_1 o_2 \dots o_T$ gözlem dizisi verildiğinde t anında s_i durumunda t+1 anında s_j durumunda olma olasılığı Eş. 2.39'da gösterilmektedir.

$$\begin{aligned}\xi_t(i,j) &= P(q_t = s_i, q_{t+1} = s_j \mid o_1 o_2 \dots o_T) \\ \xi_t(i,j) &= P(q_t = s_i, q_{t+1} = s_j, o_1 o_2 \dots o_t) / P(o_1 o_2 \dots o_t) = \\ &= \frac{P(q_t = s_i, o_1 o_2 \dots o_t) a_{ij} b_j(o_{t+1}) P(o_{t+2} \dots o_T \mid q_{t+1} = s_j)}{P(o_1 o_2 \dots o_t)}\end{aligned}$$

$$\frac{\alpha_t(i) a_{ij} b_j(o_{t+1}) \beta_{t+1}(j)}{\sum_i \sum_j \alpha_t(i) a_{ij} b_j(o_{t+1}) \beta_{t+1}(j)} \quad (2.39)$$

$o_1 o_2 \dots o_T$ gözlem dizisi verildiğinde t anında s_i durumunda olma olasılığı Eş. 2.40'ta verilmektedir.

$$\gamma_t(i) = P(q_t = s_i \mid o_1 o_2 \dots o_T) \quad (2.40)$$

$$\gamma_t(i) = P(q_t = s_i, o_1 o_2 \dots o_t) / P(o_1 o_2 \dots o_t) = \alpha_t(i) \beta_t(i) / \sum_i \alpha_t(i) \beta_t(i) \quad (2.41)$$

$\sum_t \xi_t(i,j)$: s_i durumundan s_j durumuna geçişlerin beklenen değeri

$\sum_t \gamma_t(i)$: s_i durumunda v_k gözleminin gerçekleşmesinin beklenen değeri, $o_t = v_k$

$\gamma_1(i)$: $t = 1$ anında s_i durumunda beklenen bulunma sıklığı

Yukarıdaki çıkarımlar yardımı ile Eş. 2.42-44'teki güncel model parametreleri elde edilmiş olur.

$$\underline{a}_{ij} = \sum_t \xi_t(i,j) / \sum_t \gamma_t(i) \quad (2.42)$$

$$\underline{b}_j(k) = \sum_t \xi_t(i,j) / \sum_{t=1, o_t = v_k} \gamma_t(i) \quad (2.43)$$

$$\underline{\pi}_i = \gamma_1(i) \quad (2.44)$$

Markov zinciri ve saklı markov modeli için bir örnek

Bu kısımda, markov zinciri ve saklı markov modeli için bir yerdeki günlük hava tahmini ile ilgili bir örnek verilmektedir [34]. İlgili yerde güneşli, yağmurlu ve sisli olmak üzere üç hava tipi bulunmakta ve hava durumu gün içerisinde değişmemektedir.

Çizelge 2.2. Probleme ait durum geçiş matrisi

	DURUMLAR	Yarının hava durumu		
		Güneşli	Yağmurlu	Sisli
Bugünün hava durumu	Güneşli (S1)	0,8	0,05	0,15
	Yağmurlu(S2)	0,2	0,6	0,2
	Sisli (S3)	0,2	0,3	0,5

Böyle bir durumda yarın ve sonraki günlerin hava tahminini yapmak için geçmiş n günün hava durumu bilgisinden yararlanmak gereklidir. Eş. 2.45'ten bu sonuç bulunabilir.

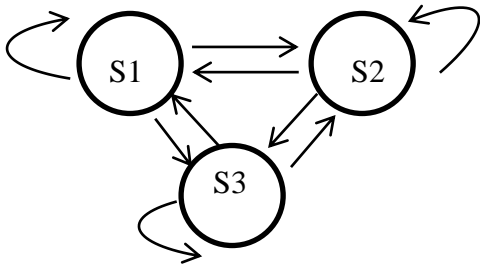
$$P(w_n | w_{n-1}, w_{n-2}, \dots, w_1) \quad (2.45)$$

Elbette n arttıkça toplanması gereken istatistiki bilgi de artacaktır. $n = 5$ için $3^5 = 243$ adet varyasyon incelenmelidir. Bu noktada, problem basitleştirme ile birinci dereceden bir markov zinciri ile çözülebilir ve Eş. 2.45, Eş. 2.46'ya indirgenir.

$$P(w_n | w_{n-1}, w_{n-2}, \dots, w_1) \approx P(w_n | w_{n-1}) \approx P(w_{yarın} | w_{bugün}) \quad (2.46)$$

Soru olarak bugünün güneşli olduğu verildiğinde yarının güneşli ve bir sonraki günün de yağmurlu olma olasılığı sorulduğunda bu birinci dereceden bir markov zinciri ile kolayca çözülebilecektir.

Çizelge 2.2'de, bu probleme ait Şekil 2.5'teki modelin durum geçiş matrisi verilmiştir.



Şekil 2.5. Probleme ait model

Çizelge 2.3. Adamın yanında şemsiye bulundurma olasılıkları

	Gözlem Olasılıkları
Güneşli	0,1
Yağmurlu	0,8
Sisli	0,3

$P(w_3 = Yağmurlu, w_2 = Güneşli, w_1 = Güneşli)$ ifadesi $P(w_3 = Yağmurlu | w_2 = Güneşli, w_1 = Güneşli) * P(w_2 = Güneşli | w_1 = Güneşli)$ olasılığına eşit ve bu ifade de $P(w_3 = Yağmurlu | w_2 = Güneşli) * P(w_2 = Güneşli | w_1 = Güneşli)$ olasılığı olarak yazıldığında sonuç $(0,05)*(0,8) = 0,04$ elde edilir.

Problemi birkaç gündür bir odada kilitli bir adamın kendisine yemek getiren bir adamın elinde şemsiye getirme olasılığına göre hava tahmini yapması durumuna çevirdiğimizde bu kez durumlarımız saklı durumda olacağı ve gözlemlere göre hava tahmini yapabileceğimiz ortaya çıkmaktadır. Böyle bir problemin çözümünde saklı markov modeli kullanılabilir. Burada bilgi olarak durumlar için Çizelge 2.3'teki gözlem olasılıkları yani yemek getiren adamın yanında şemsiye bulundurma olasılıkları vardır. Adam odada kilitli değil iken olan Eş. 2.47 artık mevcut hava adamdan saklı olduğu için Eş. 2.48'deki gibi olacaktır.

$$P(w_1, w_2, \dots, w_n) = \prod_{i=1}^n P(w_i | w_{i-1}) \quad (2.47)$$

$$P(w_1, \dots, w_n | u_1, \dots, u_n) = P(u_1, \dots, u_n | w_1, \dots, w_n) P(w_1, \dots, w_n) / P(u_1, \dots, u_n) \quad (2.48)$$

u_i , adama yemek getiren kişi i gününde şemsiye getirdi ise Evet, getirmedi ise Hayır olacaktır. $P(u_1, \dots, u_n | w_1, \dots, w_n)$ olasılığı bütün i değerleri için $u_i, w_i \neq j$ olmak üzere bütün u_j, w_j lerden bağımsız farzedildiğinde $\prod_{i=1}^n P(u_i | w_i)$ olarak hesaplanabilir. $P(w_1, \dots, w_n)$ olasılığı ilk problemdeki gibidir. $P(u_1, \dots, u_n)$ ise şemsiyeyi yanında bulundurma dizisi olasılığıdır.

Soru olarak bir kişinin odada kilitli olduğu gün hava durumunun güneşli ve yemek getiren kişinin herhangi bir gün şemsiye yanında bulundurma olasılığı 0,5 olduğu farzedilsin. Sonraki gün kilitli odada bulunan kişiye yemek getiren adamın şemsiye ile odaya geldiği biliniyor.

Bu ikinci günün yağmurlu olma olasılığı $P(w_2 = Yağmurlu | w_1 = Güneşli, u_2 = Evet)$ yani $\frac{P(w_2=Yağmurlu, w_1=Güneşli | u_2=Evet)}{P(w_1=Güneşli, u_2=Evet)}$ ifadesi sorulduğunda yapılacak işlemler aşağıdaki gibi olacaktır.

w_1 ve u_2 birbirinden bağımsız olacağı için $P(w_1 = Güneşli, u_2 = Evet)$ ifadesi $P(w_1 = Güneşli)P(u_2 = Evet)$ olarak yazılabilir. Dolayısıyla ifade artık $\frac{P(w_2=Yağmurlu, w_1=Güneşli | u_2=Evet)}{P(w_1=Güneşli)P(u_2=Evet)}$ şekline dönüşecektir.

Bayes kuralından paydaki $P(w_2 = Yağmurlu, w_1 = Güneşli | u_2 = Evet)$ ifadesi $P(u_2 = Evet | w_2 = Yağmurlu, w_1 = Güneşli)P(w_2 = Yağmurlu, w_1 = Güneşli)$ olarak yazılabilir. Markov varsayımı ile $P(u_2 = Evet | w_2 = Yağmurlu, w_1 = Güneşli)$ ifadesi de $P(u_2 = Evet | w_2 = Yağmurlu)$ olasılığına eşit olacaktır.

$P(w_2 = Yağmurlu, w_1 = Güneşli)$ ise $P(A, B) = P(A|B)P(B)$ formülü uygulandığı zaman $P(w_2 = Yağmurlu | w_1 = Güneşli)P(w_1 = Güneşli)$ ifadesine eşit olur.

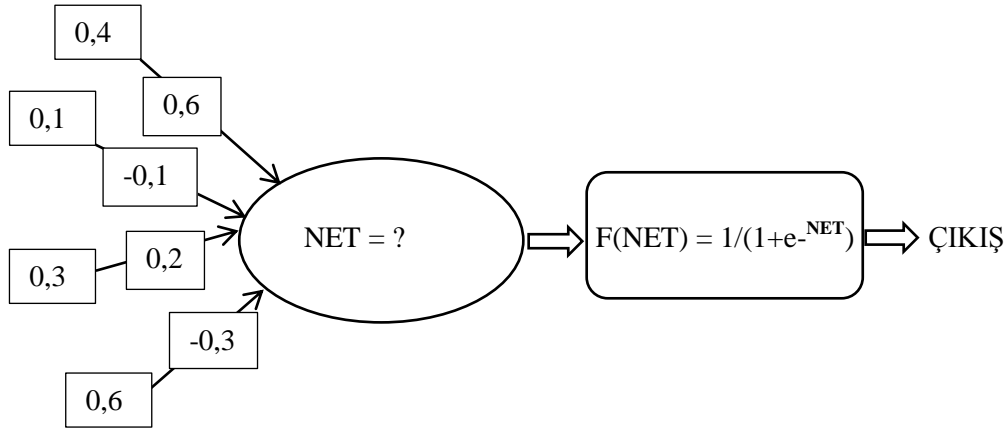
Pay ve paydadaki $P(w_1 = Güneşli)$ olasılıkları da sadeleştirildiğinde $\frac{P(u_2 = Evet | w_2 = Yağmurlu) P(w_2 = Yağmurlu | w_1 = Güneşli)}{P(u_2 = Evet)} = \frac{(0,8)(0,05)}{0,5} = 0,08$

sonucu elde edilir.

Yapay sinir ağları

İnsanoğlunun sinir yapısının modellendiği bir sınıflandırma yöntemidir. Bir sinir ağında giriş ve çıkış katmanları ve ara (gizli katman) bulunmaktadır. Çıkış katmanı dışında herbir katmandaki sinir sonraki katmandaki sinirler ile bağlantılı ve katmanlardaki sinirler arasında kendi içlerinde bir geribesleme veyahut bağlantı bulunmamaktadır [35].

Giriş katmanına gelen girişler yani öznitelikler alınır bağlantı ağırlıklarıyla çarpılır ve gizli katmandaki sinirlerde toplanır sonra da gizli katmandan aynı şekilde gelinir bağlantı ağırlıkları ile çarpılır, çıkış katmanındaki sinirlerde toplanır ve son olarak bir çıkış üretilir. Şekil 2.6'da örnek bir yapay sinir hücre verilmektedir.



Şekil 2.6. Bir yapay sinir hücresinin çalışması [35]

Yapay sinir ağlarının eğitimleri danışmanlı, danışmansız, pekiştirerek öğrenme şeklinde olmaktadır. Danışmanlı öğrenmede sisteme girişler ile birlikte o girişlere karşılık olarak hangi çıkışın verildiğini gösteren çıkışlar da verilmektedir. Danışmansız öğrenme de ise sadece girişler verilir ve sistem girişleri kendi gruplar. Aslında bu öğrenme işlemi sırasında bağlantı katsayılarının tespit edilmesi işlemi yapılır. Bir yapay sinir hücresinin nasıl çalıştığını gösteren bir örnek aşağıda verilmektedir [35].

Sinir hücresine gelen bilgi girdi değerlerinin ilgili bağlantı değerleri ile çarpılması sonucu elde edilir. Çıktı ise fonksiyon ile verilir.

$$NET = 0,4*0,6 + 0,1*(-0,1) + 0,3*0,2 + 0,6*(-0,3) = 0,11$$

$$ÇIKIŞ = 1/(1+e^{-0,11}) = 0,53$$

2.4.2. Şablon modeller

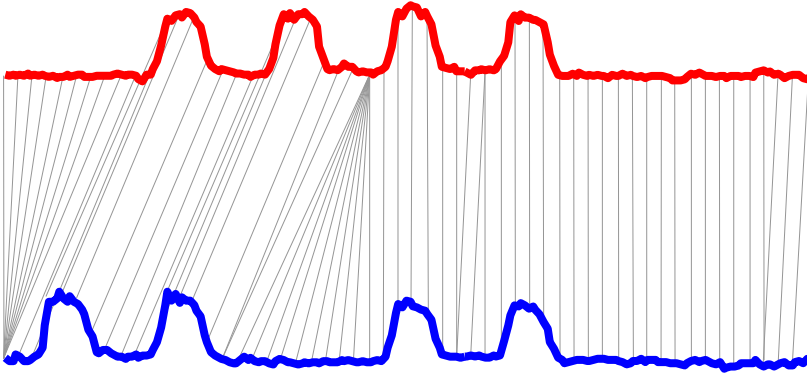
Şablon modellerde test şablonu, veritabanındaki sınıflar yani referans şablonları ile karşılaştırılmaktadır. Tanıma uygulamalarında ilk kullanılmaya başlanan ve en basit sınıflandırma yöntemlerindedir [30].

Veritabanının çok büyük olmadığı uygulamalar için uygundur. Aksi halde veritabanı tarama sürecinde çok zaman alacaktır.

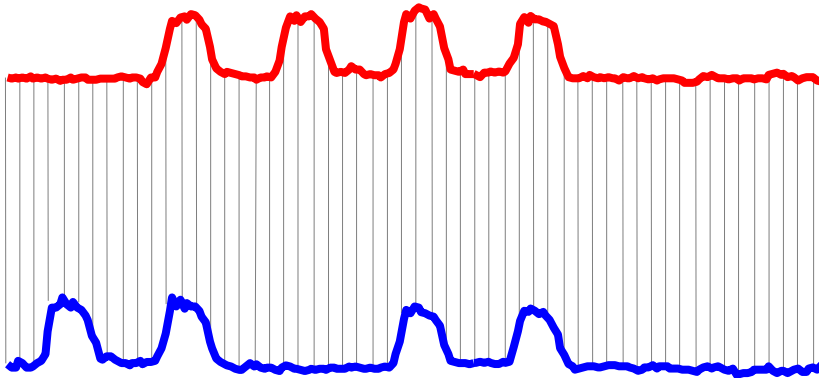
Dinamik zaman eğirme

Zaman bağımlı konuşma sinyallerinin kişiler tarafından seslendirilmesi her zaman aynı olmamaktadır. Farklı hızlarda söylenebilmekte yani bir kelime bazen üç saniye sürerken bazen de beş saniye sürmektedir. Dolayısıyla iki konuşma örüntüsünün bu değişim hızlarını ayarlayan bir mekanizma gereklidir [8]. Dinamik zaman eğirme (DTW – Dynamic Time Warping) algoritması, bu işlemi gerçekleştirmektedir. Şekil 2.7 ve Şekil 2.8’de örnek iki sinyal için öklit ile DTW’nin çalışma mantığı gösterilmiştir. Öklit birebir eşleşmeye bakarken DTW doğrusal olmayan bir ölçüm ile karşılaştırma yapmaktadır [36].

Zaman serileri için çok daha gürbüz bir mesafe ölçüm yöntemi olan DTW algoritması, farklı fazda olsalar bile benzer şekillerin eşleşmesini sağlayabilmektedir [23].



Şekil 2.7. DTW ile benzerlik ölçümü



Şekil 2.8. Öklit ile benzerlik ölçümü

Vektör nicemleme

Vektör nicemleme (VQ – Vector Quantization) ilk önce sıkıştırma algoritması olarak geliştirilmiş fakat daha sonra sınıflandırma algoritması olarak kullanılmıştır [37-38].

Vektör nicemlemede eğitim için verilen N adet M boyutlu özniteliklerden k doğal sayı olmak üzere 2^k adedince kod kitabı oluşturulur. Oluşturulan kod kitapları eğitim verisinden Linde Buzo Gray (LBG) algoritması ile oluşturulan centroid olarak adlandırılan $2^k \times M$ adet küme merkezlerine karşılık gelmektedir [39].

LBG algoritmasının adımları şöyledir [40]:

- Adım 1: İlk önce bütün eğitim setini temsil eden ortalama kod vektörü elde edilir.
- Adım 2: Kod vektörü Eş. 2.47 ve Eş. 2.48'deki işlemler ile kaç adet kod vektörü oluşturulmak isteniyorsa elde edilmiş olur. h eleman pozitif tamsayı C_N , ortalama kod vektörü ve epsilon küçük bit tamsayı, C_h^+ ve C_h^- ortalama kod vektöründen elde edilmiş yeni kod vektörleridir.

$$C_h^- = C_N * (1 - h * \text{epsilon}) \quad (2.47)$$

$$C_h^+ = C_N * (1 + h * \text{epsilon}) \quad (2.48)$$

- Adım 3: Eğitim kümesinin her bir elemanı yukarıdaki centroidler ile mesafe ölçümü yapılarak kümelendirilir.
- Adım 4: Yeni oluşan kümeleneğe göre ağırlık merkezi güncellenir.
- Adım 5: İstenilen sayıda kod vektörü elde edilinceye ve ağırlık merkezlerinde değişme olmayıncaya kadar işlem iteratif bir şekilde algoritma çalışmaya devam eder (Adım 2'ye).

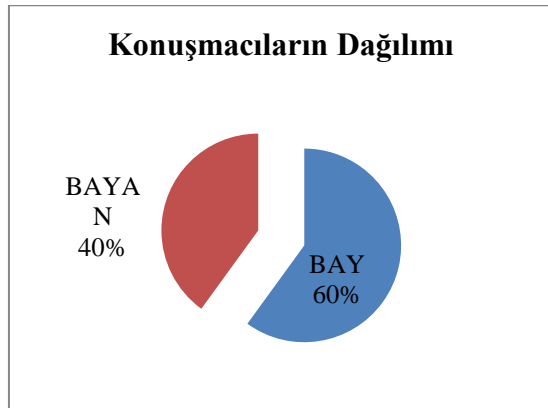
3. GELİŞTİRİLEN AYRIŞIK SÖZCÜK TANIMA SİSTEMİ

3.1. Geliştirilen Konuşma Tanıma Sisteminin Genel Yapısı

Geliştirilen konuşma tanıma sistemi kullanıcı bağımlı ayrışık kelime tanıyan bir sistemdir. Sistemde 17 adet kelime kullanılmaktadır. Kullanılan BAŞLAT, KIRK, HAYIR, ÇIKIŞ, YIKAMA, ELLİ, ÇAMAŞIR, BİTİR, DOKSAN, SICAKLIK, BEYAZ, ALTMİŞ, DURULA, RENKLİ, EVET, DURDUR, OTUZ kelimeleri genel olarak seçilen akıllı ev aleti –çamaşır makinesi- için kullanılabilir komutlardır. Bu komutlar bir önceki cümlede verilen sıralama ile konuşmacılara okutulup sistem eğitilmektedir. Verilen sıralama ile konuşmacılara okutulmasının sebebi konuşmacının sıralı okuma psikolojisine girmesini engellemektir. Sistemde, öznitelik olarak hem MFCC'ler hem de LSF'ler kullanılmaktadır. Her iki öznitelik için sonuçlar ayrı ayrı verilecektir. Kelime modelleri saklı markov modeli (HMM) ile eğitilmekte yani modellere ait parametreler HMM yöntemi ile çıkarılmaktadır. Bir konuşmacıya ait eğitim aşamaları Şekil 4.1'de ayrıntılarıyla gösterilmiştir.

3.2. Eğitim Aşaması

Her komut kelimesi, iki saniyelik kayıtlar alınarak sekizer defa olmak üzere 10 konuşmacıdan alındıktan sonra her kullanıcı için bu sekiz kelimenin beşi eğitim amacıyla kullanılmıştır. Veritabanındaki kullanıcıların cinsiyete göre dağılımı Şekil 3.1'de gösterilmektedir. Konuşmacıların komut kelime kayıtları MATLAB ortamında grafik kullanıcı arayüzü yardımıyla elde edilmiştir.



Şekil 3.1. Veritabanını oluşturan konuşmacıların cinsiyete göre dağılımı



Şekil 3.2. Geliştirilen ayrışık kelime tanıma sisteminin eğitim aşaması

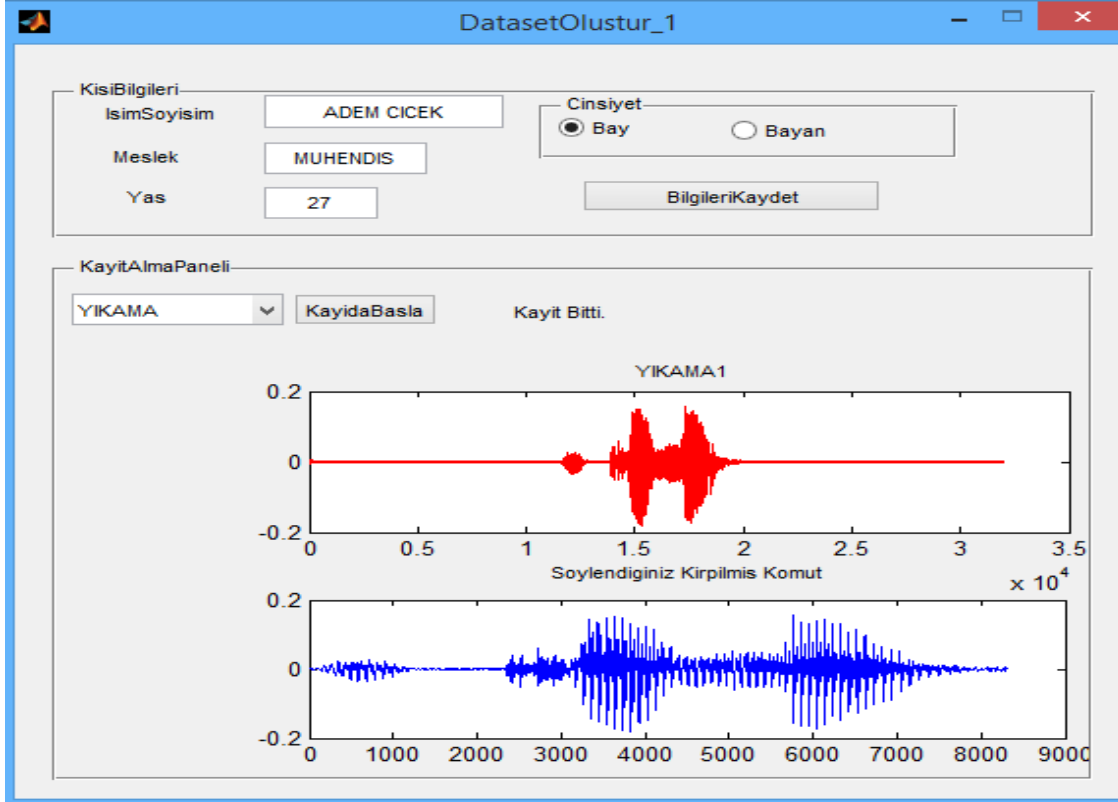
Geliştirilen ayrışık kelime tanıma sisteminin eğitim aşaması Şekil 3.2'deki gibi gerçekleşmektedir. Sonuç olarak bu kısımda konuşma veritabanı oluşturulup model parametreleri elde edilmektedir. Model parametreleri de Şekil 4.1'deki Ayrışık Sözcük Tanıma algoritmasında kullanılmaktadır.

3.2.1. MATLAB GUI ile veritabanının oluşturulması

MATLAB'ın grafik kullanıcı arayüzü ortamında tasarlanan kullanıcıya bir arayüz yardımıyla veritabanı oluşturulmuştur. Bu arayüz Şekil 3.3'te gösterilmektedir. Konuşmacılardan kayıt almadan önce konuşmacıya ait bilgilerin ilgili alanlara girilmesi istenmektedir. Aksi halde program tarafından mesaj kutusu ile ikaz edilmektedir. Bilgilerini girdikten sonra, ses dosyaları ve konuşmacının bilgilerinin de olduğu konuşmacıya ait bir dosya, arayüz programı ile oluşturulmaktadır.

Bu arayüz yardımı ile konuşmacı eğitmek istediği kelimeyi seçmekte ve sonrasında KayıtAl butonuna basılarak kayıt başladığında 'Kayıt Başladı' ve kayıt süresi bittiğinde 'Kayıt Bitti' metni ile bilgilendirme yaparak kaydı alınmaktadır. Bu kayıtlar, 16 kHz örnekleme frekansında alınmıştır. İlgili kelime için kaçınıcı kaydın alındığı hakkında KELİME# şeklinde bilgi verilmektedir. Ayrıca konuşmacı yapmış olduğu kaydı SİL butonu ile silebilmektedir.

YIKAMA komutu için kullanıcıdan ses kaydı alma aşamasını gösteren Şekil 3.3'teki arayüzde, ayrıca kayıt alınan kelime sinyalinin grafiği ve sessizlik kısımlarının çıkarılmış şekliyle, 'soylediginiz kırılmış komut' etiketi altında, sinyalin bu halinin grafiği gösterilmektedir. Şekildeki grafiklerden de görülebileceği gibi yaklaşık olarak ilk 12000 ve 20500. örnekten sonrasının sessizlik yani konuşmanın bulunmadığı kısımdır.



Şekil 3.3. Veritabanının oluşturulmasında kullanılan kullanıcı arayüzü

Kayıt alınan komutların sessizlik kısımları, 10 ms'lik adımlarla alınan her bir 20 ms'lik çerçevelerin enerji eşikleri ve sıfır geçiş sayıları baz alınarak MATLAB ortamında yazılan program yardımıyla tespit edilmiştir.

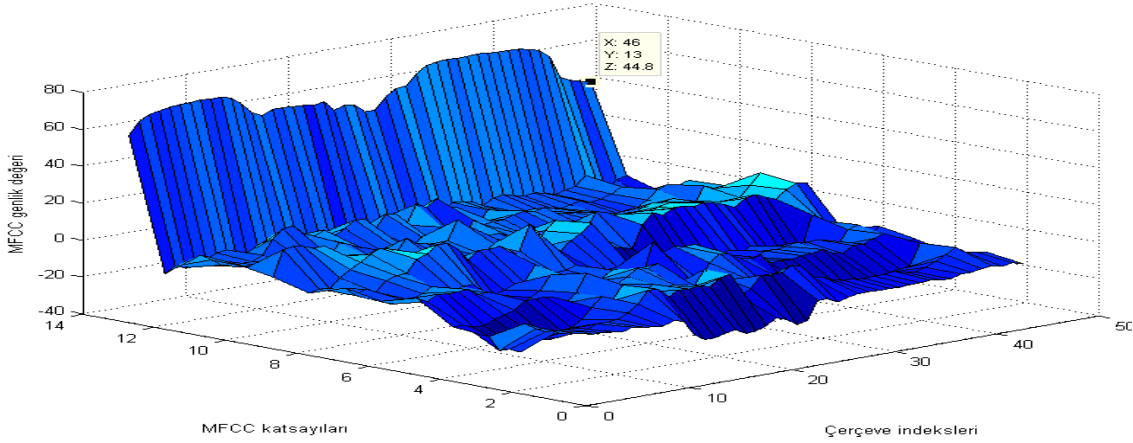
Sessiz kısımları bulma algoritması iki aşamalı olarak çalışmaktadır. İlk önce pencere enerjileri göz önünde bulundurularak alt ve üst enerji eşiği belirlenmektedir. Belirlenen üst enerji ile çerçeveler karşılaştırılmakta ve bu eşiği aşan ilk çerçeveden de geriye doğru gidilerek alt enerji eşiğinden daha düşük olan çerçeveden bir sonraki çerçeve, başlangıç çerçevesi olarak kabul edilmektedir. Aynı şekilde sinyalin sonundan başlanarak başlangıç çerçevesi bulmak için kullanılan yöntem, bitiş çerçevesini bulmak için kullanılır.

İkinci adım olarak da başlangıç çerçevesinden sesin başlangıcına doğru gelerek çerçeveler için sıfır geçiş sayıları karşılaştırılır. Deneysel olarak belirlenen sıfır geçiş eşik değeri, çerçeveler için yine deneysel olarak belirlenen ardışık olarak bu eşik değerini kaç defa geçtiğini gösteren değerden büyük ise başlangıç çerçevesi artık burası olarak değiştirilir. Aynı şekilde, komut sinyalinin sonundan başlanarak başlangıç çerçevesi güncelleştirmek için kullanılan yöntem ile bitiş çerçevesi güncelleştirilmektedir. Son olarak da başlangıç

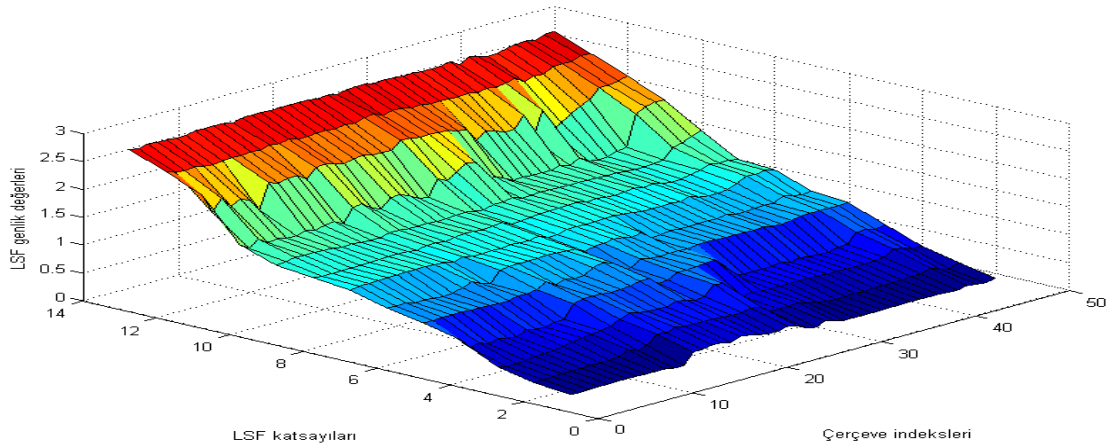
çerçevesinden bitiş çerçevesine kadar olan çerçeveler alınarak komut sinyalinin sessizlik kısımları çıkarılmış olmaktadır.

3.2.2. MFCC ve LSF yöntemleri ile özneliklerin çıkarılması

Her bir komut örnek sinyali, 10 ms'lik örtüşme olacak şekilde 20 ms'lik çerçevelere ayrılmış ve her bir çerçeve pencereleme işleminden geçirilerek MFCC ve LSF yöntemleri kullanılarak öznelikler çıkarılmıştır. 16 kHz örnekleme frekansı için pencere başına 320 adet örnek düşmektedir. 20 ms'lik çerçeveler alınmasının nedeni ses sinyalinin bu süre için istatistiki özelliklerinin durağan olmasındadır. MFCC ve LSF için özneliklerin çıkartılması 13 adet katsayı ile gerçekleştirilmiştir.



Şekil 3.4. Kullanıcı tarafından alınan örnek bir DURULA komutuna ait 13x46 MFCC öznelik matrisi yüzey gösterimi



Şekil 3.5. Kullanıcı tarafından alınan örnek bir DURULA komutuna ait 13x46 MFCC öznelik matrisi yüzey gösterimi

Herbir komut kelimesi bir sınıf olarak düşünölmüş ve o sınıfa ait ses dosyalarının tek tek her çerçevesi için N çerçeve sayısı olmak üzere, $13 \times N$ şeklinde MFCC ve LSF hücrelerine aktarılmıştır. Örnek olarak Şekil 3.4 ve 3.5'te DURULA komutuna ait MFCC ve LSF öznitelik matrislerinin yüzey grafikleri gösterilmektedir.

3.2.3. Saklı markov modeli ile özniteliklerin eğitimi-model oluşturma

Öznitelikler çıkarıldıktan sonra saklı markov modeli (HMM) ile sınıflar için, yani örnek dosyaları oluşturulan herbir komut kelimesi için, model oluşturulmuş ve eğitim yapılmıştır.

Herbir sınıf için altı durum kullanılmış ve durumlar için model olarak konuşma sinyali, sadece bir sonraki duruma geçişin olduğu ileriye doğru HMM modeline uygundur.

Saklı markov modeli ile eğitim yapılacağı için herbir sınıf ait A durum geçiş matrisi, π başlangıç vektörü; ayrıca gözlem vektörleri gaussian normal dağılım olarak düşünöldüğünden, μ ortalama vektörü ve σ varyans vektörü elde edilmiştir. A matris ve π vektörleri için alınan başlangıç değerleri, Çizelge 3.1 ve 3.2'de verilmektedir. σ ve μ başlangıç değerleri ise, gözlem vektörleri elemanları durumlara dağıılmakta ve sonrasında dosya örnekleri üzerinden ortalama alınarak bulunmaktadır. Son olarak Baum-Welch algoritması ile sistemin model parametreleri elde edilmektedir. Model parametrelerinin Baum-Welch algoritması ile nasıl elde edildiği ikinci bölümdeki istatistiksel modeller başlığı altında anlatılmaktadır.

Çizelge 3.1. Durum geçiş matrisinin başlangıç değeri

(1,1)					(1,6)
0,5	0,5	0	0	0	0
0	0,5	0,5	0	0	0
0	0	0,5	0,5	0	0
0	0	0	0,5	0,5	0
0	0	0	0	0,5	0,5
0	0	0	0	0	1
(6,1)					(6,6)

Çizelge 3.2. Başlangıç vektörünün başlangıç değeri

1	0	0	0	0	0
(1,1)					(1,6)

Çizelge 3.3. DURULA komutuna ait eğitim sonrası durum geçiş matrisleri (Soldaki matris MFCC ile sağdaki LSF ile)

(1,1)		(1,6)	(1,1)		(1,6)
0,64	0,36	0	0	0	0
0	0,93	0,07	0	0	0
0	0	0,91	0,09	0	0
0	0	0	0,65	0,35	0
0	0	0	0	0,87	0,13
0	0	0	0	0	1
(6,1)		(6,6)	(6,1)		(6,6)

Çizelge 3.3 ve 3.4’de DURULA ve DURDUR komut sınıflarına ait eğitim sonrasında elde edilen model parametrelerinden biri olan durum geçiş matrisleri verilmiştir. Durum modelimiz, ileriye doğru ve altı durumlu olduğu için, ilk indis bulunulan durumu, ikinci indis ise sonraki durumu göstermek üzere, sadece (1,1), (1,2), (2,2), (2,3), (3,3), (3,4), (4,4), (4,5), (5,5), (5,6) ve (6,6) indislerine ait durum geçişleri olmaktadır. Görüldüğü gibi durum geçiş matris değerleri bu iki kelime için yaklaşık değerlerdir ki bu beklenen bir durumdur. Çünkü her iki kelime, kelime altı benzer yapılara sahiptir.

Çizelge 3.4. DURDUR komutuna ait eğitim sonrası durum geçiş matrisleri (Soldaki matris MFCC ile sağdaki LSF ile)

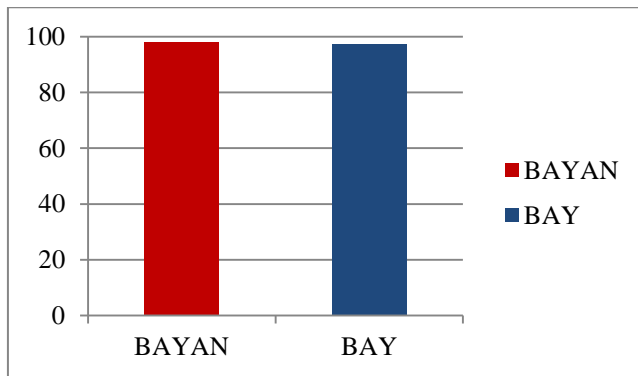
(1,1)		(1,6)	(1,1)		(1,6)
0,77	0,23	0	0	0	0
0	0,90	0,1	0	0	0
0	0	0,91	0,09	0	0
0	0	0	0,75	0,25	0
0	0	0	0	0,90	0,10
0	0	0	0	0	1
(6,1)		(6,6)	(6,1)		(6,6)

3.3. Test Aşaması

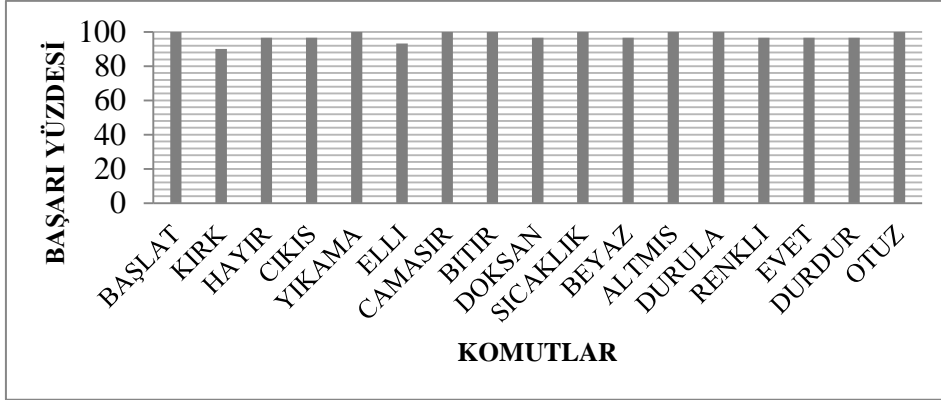
Test aşaması her bir konuşmacıdan alınan sekiz komut kelimesi örneğinin üç adedi ile yapılmıştır. Burada, gelen test dosyası daha önceden oluşturulan sınıf modellerinin hangisine en çok uyuyorsa yani hangi model, maksimum olasılık değerini üretiyorsa o sınıfa ait olduğu kararına varılır. Bu işlem hem MFCC hem de LSF öznelikleri için yapılarak ortalama sistem performansı elde edilmiştir.

3.3.1. MFCC+HMM ile başarımlar

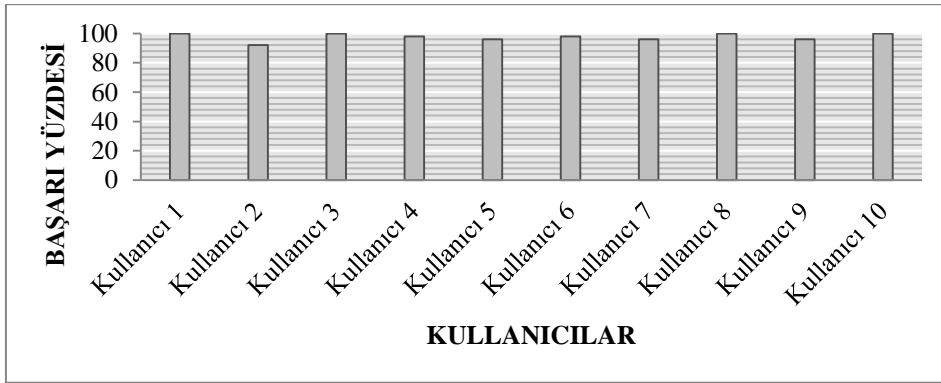
Öznelik olarak MFCC kullanıldığında, 10 konuşmacının 17 kelimeyi tanıma performansları ortalama olarak %97,65 olarak elde edilmiştir. Her bir konuşmacının her sözcüğünün üçer tekrarının ayrı ayrı tanıma başarımları Ekler kısmında ayrıntılı olarak verilmiştir. Bay ve bayan olarak ayrı ayrı değerlendirildiğinde ise tanıma başarımlarının baylar için %97,39 olduğu, bayanlar için ise %98,04 olduğu gözlenmiştir (Şekil 3.6). Bu sonuca bakılarak, geliştirilen sistemin bay ve bayan konuşmacılar için yaklaşık olarak aynı performansta çalıştığı söylenebilir. Sistemin başarımları her bir kelime için de ayrı ayrı değerlendirilmiştir ve sonuçlar Şekil 3.7’de görüldüğü gibidir. En düşük kelime tanıma performansı KIRK komutu için %90 ve ELLI komutu için %93,33 olarak elde edilmiştir. Buradan da seçilen kelimelerin tanıma başarımlarının sistemin çalışmasına olumsuz etkisinin olmayacağı düşünülmüştür. Sistemin performansı kullanıcı bazında değerlendirildiğinde ise en düşük tanıma performansının %92,16 ile bay bir kullanıcı ile elde edildiği görülmüştür (Şekil 3.8).



Şekil 3.6. Cinsiyete göre MFCC öznelikleri ile tanıma sisteminin performansı



Şekil 3.7. Kelime bazında MFCC öznitelikleri için tanıma sisteminin performansı



Şekil 3.8. Kişi bazında MFCC öznitelikleri için tanıma sisteminin performansı (Kullanıcı 1-6 : bay, Kullanıcı 7-10: bayan)

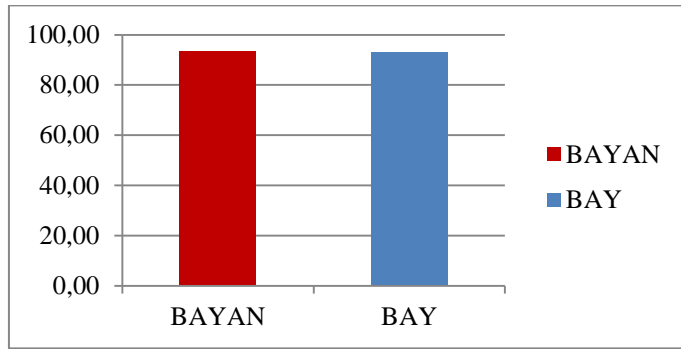
Ekler kısmında kullanıcıların söylediği üç adet test kelimelere karşılık tanıma sisteminin öznitelik olarak MFCC kullandığı durumda ne cevap verdiği her bir kullanıcı için verilmektedir (EK -1).

3.3.2. LSF+HMM ile başarıım

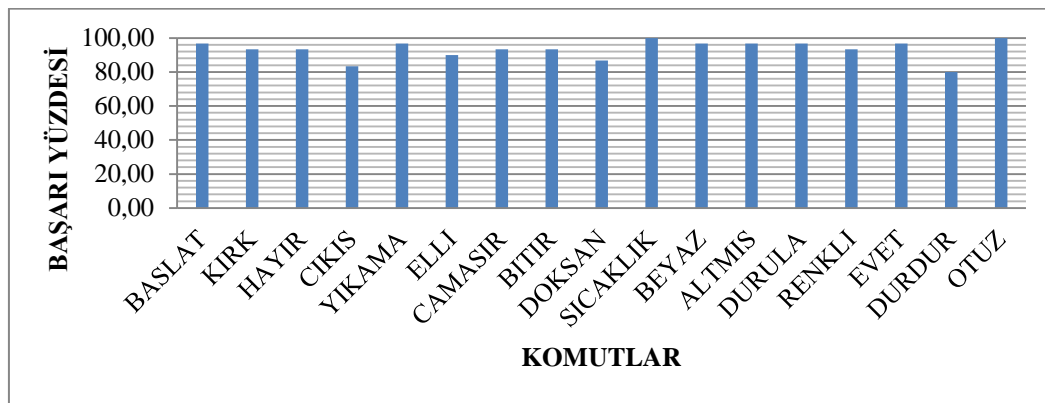
Öznitelik olarak LSF kullanıldığında, 10 konuşmacının 17 kelimeyi tanıma performansları ortalama olarak %93,33 olarak elde edilmiştir. Her bir konuşmacının her sözcüğünün üçer tekrarının ayrı ayrı tanıma başarımı Ekler kısmında ayrıntılı olarak verilmiştir. Bay ve bayan olarak ayrı ayrı değerlendirildiğinde ise tanıma başarımlarının baylar için %93,14 olduğu, bayanlar için ise %93,63 olduğu gözlenmiştir (Şekil 3.9). Bu sonuca bakılarak, LSF için de geliştirilen sistemin bay ve bayan konuşmacılar için yaklaşık olarak aynı performansta çalıştığı söylenebilir. Ancak LSF'ler kullanıldığında, MFCC'lerden farklı olarak, bayan konuşmacıların kelime tanıma oranının bay konuşmacılara göre daha yüksek

olduğu gözlenmiştir. Sistemin başarımı her bir kelime için de ayrı ayrı değerlendirilmiştir ve sonuçlar Şekil 3.10'da görüldüğü gibidir.

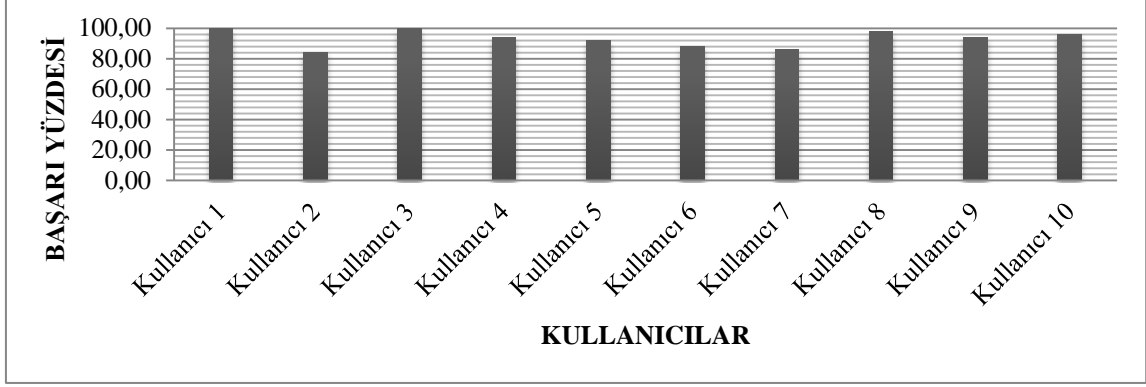
En düşük kelime tanıma performansı DURDUR komutu için %80 ve ÇIKIS komutu için %83,33 olarak elde edilmiştir. LSF'ler kullanıldığı zaman, MFCC'lerin kullanımından farklı olarak, kelime ve kullanıcı başına performansın %90'nın altında düştüğü görülmüştür. Bu durumun sistemin çalışmasına olumsuz etkileyeceği düşünülerek, son sistemin öznitelik olarak MFCC kullanmasının uygun olacağı kararına varılmıştır. Sistemin performansı kullanıcı bazında değerlendirildiğinde ise en düşük tanıma performansının da aynı şekilde MFCC'de olduğu gibi bay bir kullanıcı olan Kullanıcı_2 için %84,31 ile elde edildiği görülmüştür (Şekil 3.11).



Şekil 3.9. Cinsiyete göre LSF öznitelikleri için tanıma sisteminin performansı



Şekil 3.10. Kelime bazında LSF öznitelikleri için tanıma sisteminin performansı



Şekil 3.11. Kişi bazında LSF öznelikleri için tanıma sisteminin performansı (Kullanıcı 1-6 : bay, Kullanıcı 7-10: bayan)

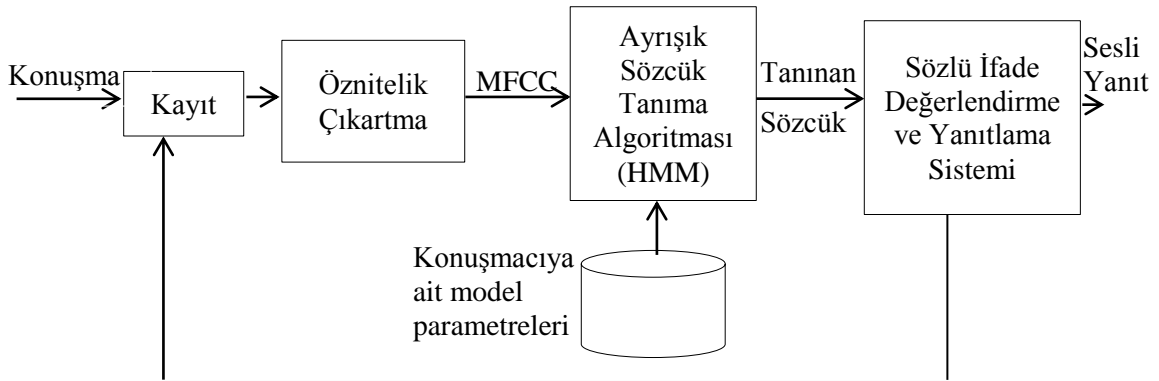
Ekler kısmında kullanıcıların söylediği üç adet test kelimelere karşılık tanıma sisteminin öznelik olarak LSF kullandığı durumda ne cevap verdiği her bir kullanıcı için verilmektedir (EK-1).

4. AKILLI EV ALETLERİ İÇİN GELİŞTİRİLEN DİYALOG SİSTEMİ

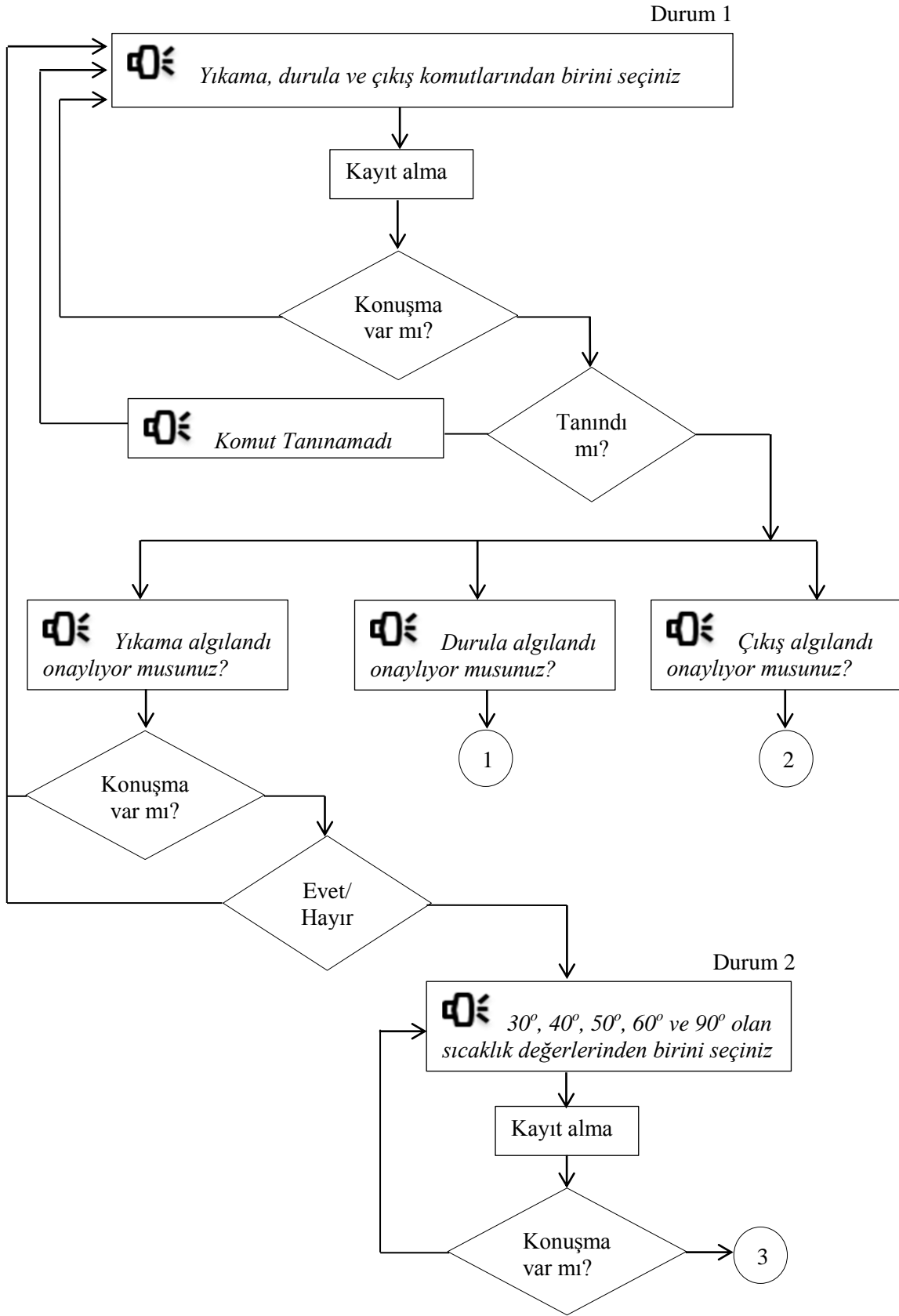
Geliştirilen konuşma tanıma sisteminin blok şeması Şekil 4.1’de verilmiştir. Konuşma, yani tanınması beklenen ayrışık sözcük, kayıt alındıktan sonra ona ait öznitelikler çıkartılmaktadır. Öznitelikler, sistem performansının daha yüksek olduğu önceki bölümde belirlendiği MFCC’ler olarak seçilmiştir. Daha sonra HMM tabanlı ayrışık sözcük tanıma algoritması kullanılarak tanınan sözcük belirlenmekte ve bu sözcük “Sözlü İfade Değerlendirme ve Yanıt Sistemi” tarafından değerlendirilerek kullanıcıya gerekli yanıtlar sesli olarak verilmektedir. “Sözlü İfade Değerlendirme ve Yanıt Sistemi”, sesli yanıt ürettikten sonra, buna karşılık kullanıcı tarafından seslendirilen komut kelimeleri algılamak üzere tekrar konuşma tanıma bloklarına Şekil 4.1’de gösterildiği gibi geri dönmektedir. Geliştirilen diyalog sisteminin ayrıntıları bundan sonraki bölümlerde anlatılacaktır.

4.1. Örnek Bir Ev Aleti – Çamaşır Makinesi ve Kullanılacak Komutlar

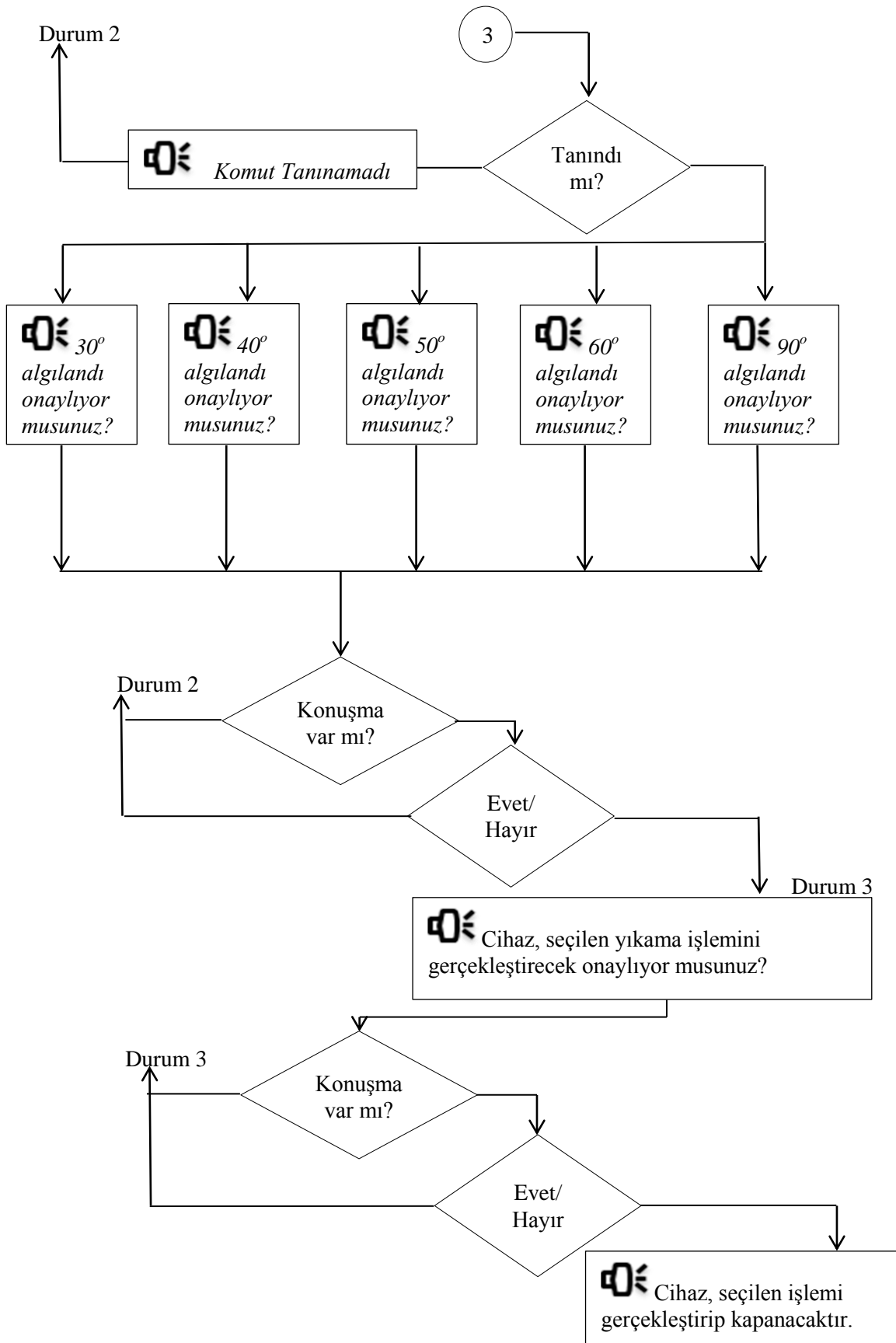
Geliştirilen konuşma tanıma sistemi için örnek bir ev aleti olarak çamaşır makinesi düşünülmüştür. Bu seçimin nedeni, diğer ev aletlerine oranla çalışmasında daha fazla programların, seçimlerin olmasıdır. Bu cihaz için kullanılacak 10 adet komut belirlenmiştir. Bu komut kelimeleri YIKAMA, DURULA, ÇIKIŞ, OTUZ, KIRK, ELLİ, ALTMİŞ, DOKSAN, EVET ve HAYIR olarak seçilmiştir.



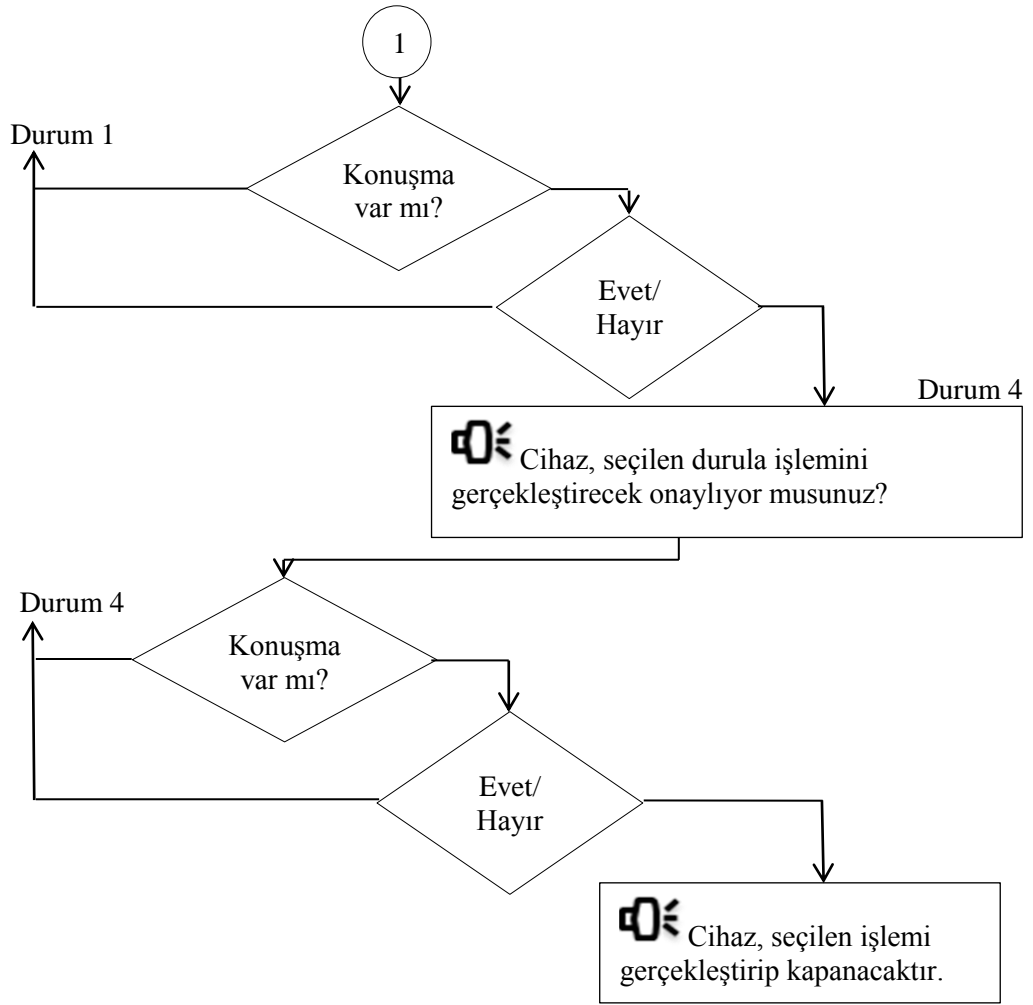
Şekil 4.1. Geliştirilen ayrışık kelime tanıma tabanlı diyalog sistemi



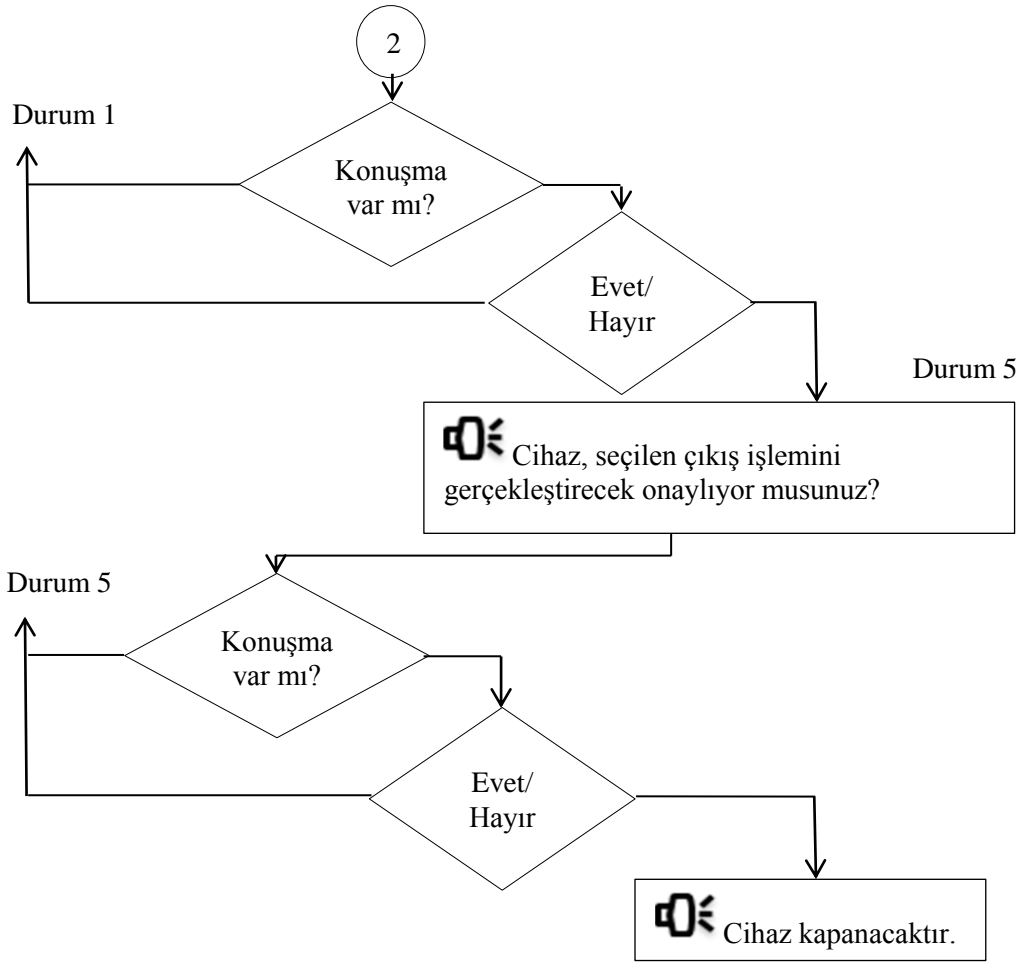
Şekil 4.2. (a) Diyalog sisteminin ayrıntılandırılmış akış şeması: YIKAMA komutunun seçilme durumu



Şekil 4.2. (a) (devam) Diyalog sisteminin ayrıntılandırılmış akış şeması: YIKAMA komutunun seçilme durumu



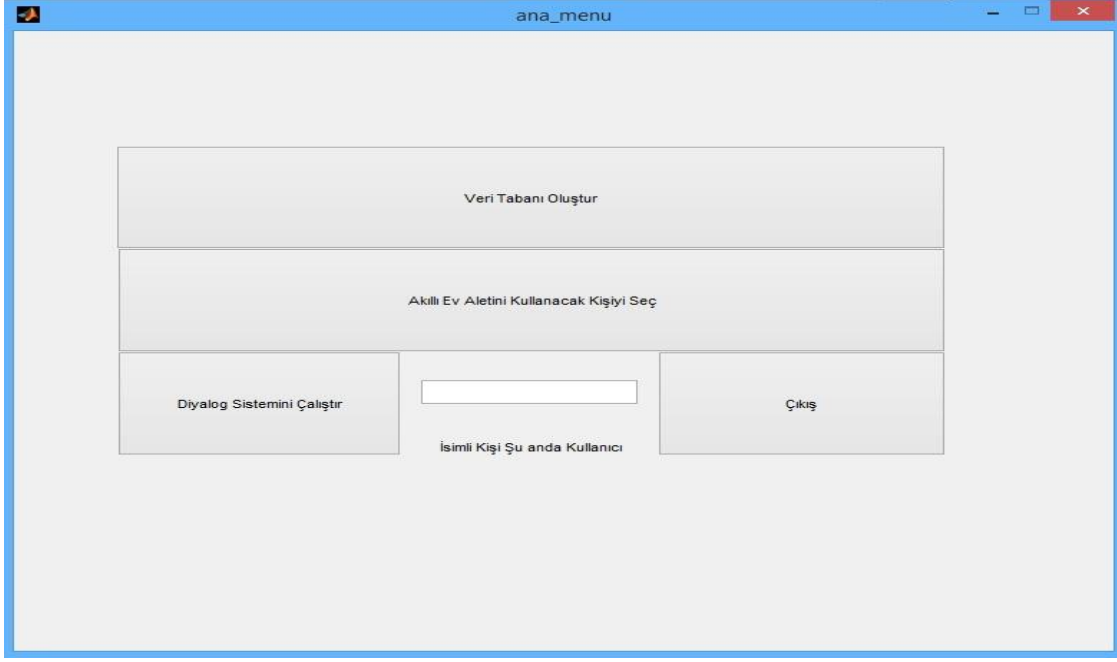
Şekil 4.2. (b) Diyalog sisteminin ayrıntılandırılmış akış şeması: DURULA komutunun seçilme durumu



Şekil 4.2. (c) Diyalog sisteminin ayrıntılandırılmış akış şeması: ÇIKIŞ komutunun seçilme durumu

4.2. Akıllı Ev Aletinin Diyalog Sistemi için Akış Şeması

Akıllı ev aletinin çalışması, ev aletinde alıcının ve üzerinde bir vericinin bulunduğu kablosuz bir kumandayla kontrol edileceği bir sistem olarak tasarım yapılmıştır. MATLAB GUI (Graphical User Interface) ve üzerindeki butonlar, kablosuz bir kumandayı temsil edecek şekilde düşünülmüştür. Gerçekte kişi kendisi kumanda üzerinden eğitim için gerekli veritabanını oluşturacaktır. Sonrasında eğitim için kendi oluşturduğu veritabanını seçerek sistemi eğitecek ve artık sistem kullanıcının kendisi için çalışmaya hazır hale gelecektir. Şekil 4.1’de verilen diyalog sisteminin akıllı ev aleti için ayrıntılandırılmış akış şeması Şekil 4.2’de verilmiştir.



Şekil 4.3. Kullanıcının ev aletini Diyalog Tanıma Sistemi ile kullanabilmesi için tasarlanan arayüz

4.3. Akış Diyagramı Çıkarılan Akıllı Ev Aletinin MATLAB GUI'de Dialog Tanıma Sisteminin Gerçekleştirilmesi

Hangi konuşmacı kullanıcı olacak ise ona ait eğitim ses klasörü seçilerek eğitim işlemi tamamlanır ve sistem kullanıma hazır hale gelir. Dialog Tanıma Sistemin'de kullanıcıya neleri seçmesi gerektiği bilgilendirilir ve tanınan kelimenin söylediği kelime olup olmadığına dair evet/hayır şeklinde teyit alınır. Teyit için kullanılan kelimeler için de yine tanıma sistemi çalışarak işleme devam edilir. Şekil 4.3'te oluşturulan arayüzün ekran görüntüsü verilmiştir.

Aynı arayüz kullanılarak, eğitim veritabanının oluşturulması da mümkündür. Bu işlem için Şekil 4.3'te görülen *Veri Tabanı Oluştur* butonunun seçilmesi gerekmektedir. Bu şekilde, yeni bir kullanıcı da kendi sesini sisteme tanıtır, eğitimini sağlayabilir ve daha yeni konuşmacı için de diyalog sistemi çalıştırılabilir.

5. SONUÇ VE ÖNERİLER

Son zamanlarda çok sık karşılaştığımız akıllı ev aletleri, insan hayatını, özellikle yaşlı ve engelli olan kişiler için, oldukça kolaylaştırmaktadır. Bu tez çalışması kapsamında, ilk aşamada akıllı bir ev aleti -çamaşır makinesi- örneği için, bunun üzerinde çalışmak üzere bir sesten komut tanıma sistemi geliştirilmiştir. Geliştirilen bu tanıma sisteminde, MATLAB ortamında saklı markov model (HMMs – Hidden Markov Models) algoritmaları kullanılmıştır. Örnek olarak alınan akıllı ev aletinin çalışmasında kullanılabilecek 17 adet olası komut sözcüğü belirlenmiş ve bu olası komut kelimeleri üzerinden sistemin sınıflandırma yüzdesi verilmiştir. Sistemin eğitimi için, MATLAB grafik kullanıcı arayüzü birimi kullanılarak tasarlanan bir arayüz yardımı ile, altısı bay dördü bayan toplam 10 kişiden alınan komutların her biri sekizer defa kaydedilmiş ve bir konuşmacı veritabanı oluşturulmuştur. Öznitelik olarak Mel Frekans Kepstral Katsayıları (MFCC - Mel Frequency Cepstral Coeffiecints) ve Çizgisel Spektrum Frekansları (LSF – Line Spectral Frequency) kullanılmış ve her konuşmacının bütün komutları beş tekrar için saklı Markov modelleri (HMMs – Hidden Markov Models) kullanılarak modellenmiştir. Her konuşmacının geriye kalan üç komutu test için kullanılmış ve test sonuçları farklı öznitelikler için ayrı ayrı analiz edilmiştir. Sonuçlar, kelime, kişi ve genel sistem başarımı bazında karşılaştırılmıştır. Sonuç olarak akıllı ev aletleri için kullanıcının kendisinin yapacağı kısa bir eğitim ile yüksek tanıma başarımı ile çalışabilecek Türkçe bir komut tanıma sistemi MATLAB yazılım ortamında gerçekleştirilmiştir. 10 konuşmacının bütün komutlarının doğru tanıma yüzdesi MFCC öznitelikleri için %97,65 ve LSF öznitelikleri için %93,33 olarak gerçekleştirilmiştir.

İkinci aşamada, akıllı ev aleti için bir diyalog sistemi oluşturulmuştur. Bir önceki aşamada, MFCC öznitelikleri daha iyi performans sergilediği için, bu aşamada doğrudan MFCC'ler ile ayrışık sözcük tanıma sistemi çalıştırılmaktadır. Diyalog sisteminin çalışma senaryosu oluşturulmuş ve bu senaryo üzerinden kullanıcı ile sesli iletişim kuran bir kullanıcı arayüzü geliştirilmiştir. Bu arayüz yeni bir kullanıcının sistemi kendi sesi için eğitip, akıllı ev aleti ile iletişime geçerek sesli komutlarını iletmesini sağlamaktadır. Diyalog sisteminin komutları oluşturulurken, sistemin kesintisiz çalışabilmesi için, komut sözcüklerinin ayrı ayrı tanıma yüzdelerinin yüksek olması gerekmektedir. Tanınması zor olan sözcükler seçildiğinde, bu sözcükleri seslendiren kullanıcının komutunu algılayamayan ya da yanlış algılayan sistem, o noktada takılacak ya da başa dönecek ve söz konusu komut bir türlü

gerçekleştirilemeyecektir. Böyle bir durumun oluşmaması için, tanınma başarımı düşük olan sözcüklerin komut olarak seçilmemesi sağlanabilir. Bu durum, tasarlanan sistemde “SİSTEM” komutu için ortaya çıkmıştır. Tanınma oranı %90’ın altında olarak tespit edildiği için, bu sözcük komut listesinden çıkartılmıştır. Bu sözcüğün, diğer sözcüklerden ayrılan tarafı, yabancı kökenli bir sözcük olmasıdır. Konuşma tanıma sisteminin bu sözcük için düşük başarımla sergilemesinin sebebi, yabancı kökenli sözcüklerin konuşmacılar tarafından doğru ve her seferinde tutarlı bir şekilde seslendirilememesi olabilir.

Sistemin tanıma performansı; düşük olan kelimelerin olması kullanıcı, ortam ve ses kaydı alınan cihazların kalitesi gibi parametrelerden etkilenmektedir. Kullanıcının performans düşüklüğüne etkisi eğitim kümesine verdiği örneklerin test için söyledikleriyle tutarlı olmaması olmuştur. Kelimelerin yapısından ortaya çıkan, kullanıcılar tarafından zor söyleniş tanıma performansını etkilemiştir. Örneğin KIRK, CIKIS komutu, yapısında patlamalı yapılar bulunduran bir kelimedir. ELLI ve EVET kelimeleri için ise e harfinin açık ve kapalı söylenişleri tanımayı zorlaştırmıştır. Kelime içinde yanyana sessiz harflerin bulunması da söylemeyi zorlaştırdığı için tanımayı da zorlaştırmaktadır. Ayrıca aynı kelime altı yapılar içeren kelimeler yakın model parametrelerine sahip olacağı için bu durumda sistemin yanlış tanınmasına sebep olmaktadır.

Diyalog tanıma sistemi ise sınırlı sayıda komut içerecek şekilde örnek bir ev alet için tasarlanmıştır. Komut kümesi kullanılacak akıllı ev aletine göre daha da genişletilebilir ya da daraltılabilir. Bu sistem kullanıcı bağımlı olduğundan dolayı kullanıcının kendisine ait olmayan eğitim kümesi için cihazı çalıştırdığında performans oldukça düşecektir. Gerçi olan sistemde bu durumun, kimin eğitim verileri ile sistem eğitildiğini gösteren bir yazı ile önüne geçilmiştir. Başka bir yöntem olarak böyle bir durumu ortadan kaldırmak veya başka kişiler tarafından kullanılmasını engellemek için sisteme konuşmacı tanıma algoritmaları eklenebilir.

Ayrıca sisteme ek olarak yük çıkartmamak amacıyla gelen konuşma sözcüklerinin de kısa ve ayrışık olmasından dolayı dinamik zaman eğirme işlemi yapılmamıştır. Daha yüksek performans için dinamik zaman eğirme işleminden sonra sözcüklerin eğitime alınması bir yöntem olarak önerilebilir. Sistemin gürbüzlüğü açısından ise RASTA – PLP gibi gürültüye karşı daha dirençli öznelikler kullanılabilir.

KAYNAKLAR

1. Demirci, M. D. (2005). *Bilgisayar Destekli Ses Tanıma Sistemi*, Yüksek Lisans Tezi, İstanbul Üniversitesi Fen Bilimleri Enstitüsü, İstanbul.
2. Yloni, D. (2012). DSP Based Speech Operated Home Appliances Using Zero Crossing Features. *Signal Processing: An International Journal (SPIJ)*, -(2), 44-55.
3. Yalçın, N. (2008). Konuşma Tanıma Teorisi ve Teknikleri¹, *Kastamonu Eğitim Dergisi*, 16(1), 249-266.
4. Furui, S. (1989). *Digital Speech Processing, Synthesis, and Recognition*, New York: Marcel Dekker.
5. Erbad, A. ve Krasic, C. (2007). A Hybrid Framework to Program Smart Enviroments, *IEEE Forth International Conference on Innovations in Information Technology*, -(-), 228-232.
6. Tan, B. ve Li, J. (2011). A Speech Remote Control Remote System Realization Based on Computer Telecommunication Integration, *IEEE Third International Workshop on Intelligent Systems and Applications*, -(-), 1-3.
7. Ozgunduz, E., Turkmen, H. I., Senturk, T. ve Karsligil, M.E. (2010). Vehicle identification using acoustic and seismic signals, *IEEE Signal Processing and Communications Applications Conference*, -(-), 941-944.
8. Karacı, A. (2006). *Bilgisayar Ortamında Sesli İfadeler Tanıma*, Yüksek Lisans Tezi, Gazi Üniversitesi Fen Bilimleri Enstitüsü, Ankara.
9. Hsu, C. and Yang, S. (2008). Constructing Intelligent Living Space Controlling System with Bluetooth and Speech-Recognition, *IEEE Eighth International Conference on Intelligent Systems Design and Applications*, -(-), 666-671.
10. Koç, A. (2002). *Acoustic Feature Analysis for Robust Speech Recognition*, MSc Thesis, Boğaziçi University Institute for Graduate Studies in Science and Engineering, İstanbul.
11. Zeng, X., Fapojuwo, A. O. and Davies, R. J. (2006). Design and Performance Evaluation of Voice Activated Wireless Home Devices, *IEEE Transactions on Consumer Electronics*, 52(3), 983-989.
12. Aşlıyan, R., Günel, K. (2009). Destek Vektör Makineleri Yöntemiyle Türkçe Konuşma Tanıma Sistemi, *Akademik Bilşim '09*, -(-),-
13. Jung, S., Sung, K., Park, M., Kang, E., Hwang, W., Won, J., Lee, W. and Han, S. (2013). A study on robust control of mobile robot by voice command, *IEEE 13th International Conference on Control, Automation and Systems*, -(-), 657-659.

14. Kelebekler, E. ve İnal, M. (2008), Otomobil İçindeki Cihazların Sesle Kontrolüne Yönelik Konuşma Tanıma Sisteminin Gerçek Zamanlı Laboratuar Uygulaması, *Gazi Üniversitesi Politeknik Dergisi*, 11(2), 109-114.
15. Priyadarshani, P. G. N., Dias, N. G. J. and Punchihewa, A. (2012). Dynamic Time Warping Based Speech Recognition for Isolated Sinhala Words, *IEEE 55th International Midwest Symposium on Circuits and Systems*, -(-), 892-895.
16. Günal, S., Edizkan, R. ve Barkana, A. (2003). Ortak Vektör Yaklaşımı Yöntemiyle Gerçek Zamanlı Rakam Tanıyıcı Tasarımı, 11th Sinyal İşleme ve İletişim Uygulamaları Kurultayı, -(-), 308-311.
17. Chang, C., Zhou, Z., Lin, S., Wang, J. and Wang, J. (2012). Intelligent Appliance Control Using A Low-Cost Embedded Speech Recognizer, *IEEE Eighth International Conference on Computing and Networking Technology*, -(-), 311-314.
18. Zeng, X., Fapojuwo, A. O. and Davies, R. J. (2006) Design and Performance Evaluation of Voice Activated Wireless Home Devices, *IEEE Transactions on Consumer Electronics*, 52(3), 983-989.
19. Bülbül, H. İ., Karacı, A. (2007). Bilgisayar Ortamında Sesli Komutları Tanıma: Örüntü Tanıma Yöntemi, *Kastamonu Eğitim Dergisi*, 15(1), 45-62.
20. Meena, K., Subramaniam, K. and Gomathy, M. (2013). Gender Classification in Speech Recognition Using Fuzzy Logic and Neural Network, *The International Arab Journal of Information Technology*, 10(5), 477-485.
21. Yalçın, N., Ülker, Ü. (2011). Görme Engelliler İçin Ses Analizi ile E-posta İletimi, *Gazi Üniversitesi Bilişim Teknolojileri Dergisi*, 4(3), 37-46.
22. Iwasawa, M., Yamaguchi, T. and Fujimoto, Y. (2010). Home Appliance Service System by Using an Object Position and Multimodel Interaction with Communication Robot, *IEEE International Symposium on Micro-NanoMechatronics and Human Science*, -(-), 459-464.
23. Nandyala, S. P. ve Kumar, T. K. (2010). Real Time Isolated Word Speech Recognition System for Human Computer Interaction, *International Journal of Computer Applications*, 12(2), 1-7.
24. Soda, S., Nakamura, M., Matsumoto, S., Izumi, S., Kawaguchi, H. and Yoshimoto, M. (2012). Implementing Virtual Agent as an Interface for Smart Home Voice Control, *IEEE 19th Asia-Pacific SoftwareEngineering Conference*, 1(-), 342-345.
25. Tunalı, V. (2005). *A speaker Dependent, Large Vocabulary, Isolated Word Speech Recognition System For Turkish*, MSc Thesis, Marmara University Institute for Graduate Studies in Pure and Applied Sciences, İstanbul.
26. Duran, Ş. (2001). *Keyword Spotting Using Hidden Markov Models*, MSc Thesis, Boğaziçi University Institute for Graduate Studies in Science and Engineering, İstanbul.

27. Uslu, E. (2007). *Gizli Markov Markov Modeli ile Geniş Sözlüklü Sürekli Konuşma Tanıma*, Yüksek Lisans Tezi, Yıldız Teknik Üniversitesi Fen Bilimleri Enstitüsü, İstanbul.
28. Rabiner, L. R. and Juang BH. (1993). *Fundamentals of Speech Recognition*, New Jersey: Prentice Hall.
29. İnternet: Klautau, A. The MFCC. URL: <http://www.webcitation.org/query?url=http%3A%2F%2Fwww.cic.unb.br%2F%7EElamar%2Fte073%2FAulas%2Fmfcc.pdf&date=2015-08-22>, Son Erişim Tarihi: 22.08.2015.
30. Hanilçı, C. (2007). *Konuşmacı Tanıma Yöntemlerinin Karşılaştırmalı Analizi*, Yüksek Lisans Tezi, Uludağ Üniversitesi Fen Bilimleri Enstitüsü, Bursa.
31. Mengüşoğlu, E. (1999). *Bir Türkçe Sesli İfade Tanıma Sisteminin Kural Tabanlı Tasarımı ve Gerçekleştirimi*, Yüksek Lisans Tezi, Hacettepe Üniversitesi Fen Bilimleri Enstitüsü, Ankara.
32. Pellom, B. L. (1998). *Enhancement, Segmentation, and Synthesis of Speech with Application to Robust Speaker Recognition*, PhD Dissertation, Duke University, Durham.
33. Salor, Ö. (2005). *Voice Transformation and Development of Related Speech Analysis Tools for Turkish*, PhD Dissertation, Middle East Technical University Graduate School of Natural And Applied Sciences, Ankara.
34. İnternet: Fosler-Lusser, E. Markov Models and Hidden Markov Models: A Brief Tutorial. URL: <http://www.webcitation.org/query?url=http%3A%2F%2Fdi.ubi.pt%2F%7Ejpaulo%2Fcompetence%2Ftutorials%2Fhmm-tutorial-1.pdf&date=2015-08-22>, Son Erişim Tarihi: 22.08.2015.
35. Çetin, E. (2011). *Yapay Zeka Uygulamaları (Yapay Sinir Ağları – Bulanık Mantık – Genetik Algoritma)*, Ankara: Seçkin Yayıncılık.
36. İnternet: Herrera, R H; Baan M . Automated seismic-to-well ties. URL: http://www.webcitation.org/query?url=http%3A%2F%2Fwww.slideshare.net%2FHenry_cu%2Feage2012-dtw-hvdb&date=2015-08-22, Son Erişim Tarihi: 22.08.2015.
37. Hanilci, C. and Ertas, F. (2009). Principal component based classification for text-independent speaker identification, *IEEE Fifth International Conference on Soft Computing, Computing with Words and Perceptions in System Analysis Decision and Control*, (-), 1-4.
38. Soong, F. K., Rosenberg, A. E., Rabiner, L. R. and Juang, B. H. (1985). A vector quantization approach to speaker recognition, *IEEE International Conference on Acoustics, Speech and Signal Processing*, 10(-), 387-390.

39. Linde, Y., Buzo, A. and Gray, R. M. (1980). An Algorithm for Vector Quantizer Design, *IEEE Transactions on Communications*, 28(1), 84-95.
40. Uzunçarşılı, M. (2005). *Vektör Nicemleme Tekniklerine Dayalı Konuşmacı Tanıma Algoritmalarının İncelenmesi*, Yüksek Lisans Tezi, Ankara Üniversitesi Fen Bilimleri Enstitüsü, Ankara.

EKLER

EK-1. Konuşma tanıma sistemini test ederken kullanılan, her bir kullanıcıya ait üç test datası için elde edilen test sonuçları

Çizelge 1.1. Kullanıcı_1'e ait eğitim sonrasında kendi test verileri ile konuşma tanıma sisteminin çıktıları (Cinsiyet: BAY)

KISALTMA KOMUTLAR	İlgili Komutların Test için Kullanılan Örnekleri (MFCC ile)																
	BASLAT	KIRK	HAYIR	CIKIS	YIKAMA	ELLI	CAMASIR	BITIR	DOKSAN	SICAKLIK	BEYAZ	ALTMIS	DURULA	RENKLI	EVET	DURDUR	OTUZ
BT	3																
KK		3															
HR			3														
CS				3													
YA					3												
EI						3											
CR							3										
BR								3									
DN									3								
SK										3							
BZ											3						
AS												3					
DA													3				
RI														3			
ET															3		
DR																3	
OZ																	3

KISALTMA KOMUTLAR	İlgili Komutların Test için Kullanılan Örnekleri (LSF ile)																
	BASLAT	KIRK	HAYIR	CIKIS	YIKAMA	ELLI	CAMASIR	BITIR	DOKSAN	SICAKLIK	BEYAZ	ALTMIS	DURULA	RENKLI	EVET	DURDUR	OTUZ
BT	3																
KK		3															
HR			3														
CS				3													
YA					3												
EI						3											
CR							3										
BR								3									
DN									3								
SK										3							
BZ											3						
AS												3					
DA													3				
RI														3			
ET															3		
DR																3	
OZ																	3

SAYI	Test örneğinin SAYI kadar doğru tanındığını gösterir
SAYI	Test örneğinin SAYI kadar yanlış tanındığını gösterir
KISALTMA KOMUTLAR: Satırdaki komutların kısaltmasıdır	

EK-1. (devam) Konuşma tanıma sistemini test ederken kullanılan, her bir kullanıcıya ait üç test datası için elde edilen test sonuçları

Çizelge 1.2. Kullanıcı_2'ye ait eğitim sonrasında kendi test verileri ile konuşma tanıma sisteminin çıktıları (Cinsiyet: BAY)

KISALTIMA KOMUTLAR	İlgili Komutların Test için Kullanılan Örnekleri (MFCC ile)																
	BASLAT	KIRK	HAYIR	CIKIS	YIKAMA	ELLI	CAMASIR	BITIR	DOKSAN	SICAKLIK	BEYAZ	ALTMIS	DURULA	RENKLI	EVET	DURDUR	OTUZ
BT	3																
KK		0			2					1							
HR			3														
CS				3													
YA					3												
EI						3											
CR							3										
BR								3									
DN	1								2								
SK										3							
BZ											3						
AS												3					
DA													3				
RI														3			
ET															3		
DR																3	
OZ																	3

KISALTIMA KOMUTLAR	İlgili Komutların Test için Kullanılan Örnekleri (LSF ile)																
	BASLAT	KIRK	HAYIR	CIKIS	YIKAMA	ELLI	CAMASIR	BITIR	DOKSAN	SICAKLIK	BEYAZ	ALTMIS	DURULA	RENKLI	EVET	DURDUR	OTUZ
BT	3																
KK		2					1										
HR			2				1										
CS				0			3										
YA					3												
EI						2								1			
CR							3										
BR								3									
DN									2								1
SK										3							
BZ											3						
AS												3					
DA													3				
RI														3			
ET															3		
DR								1								2	
OZ																	3

SAYI	Test örneğinin SAYI kadar doğru tanındığını gösterir
SAYI	Test örneğinin SAYI kadar yanlış tanındığını gösterir

EK-1. (devam) Konuşma tanıma sistemini test ederken kullanılan, her bir kullanıcıya ait üç test datası için elde edilen test sonuçları

Çizelge 1.3. Kullanıcı_3'e ait eğitim sonrasında kendi test verileri ile konuşma tanıma sisteminin çıktıkları (Cinsiyet: BAY)

KISALTIMA KOMUTLAR	İlgili Komutların Test için Kullanılan Örnekleri (MFCC ile)																
	BASLAT	KIRK	HAYIR	CIKIS	YIKAMA	ELLI	CAMASIR	BITIR	DOKSAN	SICAKLIK	BEYAZ	ALTMIS	DURULA	RENKLI	EVET	DURDUR	OTUZ
BT	3																
KK		3															
HR			3														
CS				3													
YA					3												
EI						3											
CR							3										
BR								3									
DN									3								
SK										3							
BZ											3						
AS												3					
DA													3				
RI														3			
ET															3		
DR																3	
OZ																	3

KISALTIMA KOMUTLAR	İlgili Komutların Test için Kullanılan Örnekleri (LSF ile)																
	BASLAT	KIRK	HAYIR	CIKIS	YIKAMA	ELLI	CAMASIR	BITIR	DOKSAN	SICAKLIK	BEYAZ	ALTMIS	DURULA	RENKLI	EVET	DURDUR	OTUZ
BT	3																
KK		3															
HR			3														
CS				3													
YA					3												
EI						3											
CR							3										
BR								3									
DN									3								
SK										3							
BZ											3						
AS												3					
DA													3				
RI														3			
ET															3		
DR																3	
OZ																	3

SAYI	Test örneğinin SAYI kadar doğru tanındığını gösterir
SAYI	Test örneğinin SAYI kadar yanlış tanındığını gösterir

EK-1. (devam)Konuşma tanıma sistemini test ederken kullanılan, her bir kullanıcıya ait üç test datası için elde edilen test sonuçları

Çizelge 1.4. Kullanıcı_4'e ait eğitim sonrasında kendi test verileri ile konuşma tanıma sisteminin çıktıkları (Cinsiyet: BAY)

KISALTIMA KOMUTLAR	İlgili Komutların Test için Kullanılan Örnekleri (MFCC ile)																
	BASLAT	KIRK	HAYIR	CIKIS	YIKAMA	ELLI	CAMASIR	BITIR	DOKSAN	SICAKLIK	BEYAZ	ALTMIS	DURULA	RENKLI	EVET	DURDUR	OTUZ
BT	3																
KK		3															
HR			3														
CS				3													
YA					3												
EI						3											
CR							3										
BR								3									
DN									3								
SK										3							
BZ										1	2						
AS												3					
DA													3				
RI														3			
ET															3		
DR																3	
OZ																	3

KISALTIMA KOMUTLAR	İlgili Komutların Test için Kullanılan Örnekleri (LSF ile)																
	BASLAT	KIRK	HAYIR	CIKIS	YIKAMA	ELLI	CAMASIR	BITIR	DOKSAN	SICAKLIK	BEYAZ	ALTMIS	DURULA	RENKLI	EVET	DURDUR	OTUZ
BT	3																
KK		3															
HR			3														
CS				3													
YA					3												
EI						3											
CR				1			2										
BR								3									
DN									2	1							
SK										3							
BZ										1	2						
AS												3					
DA													3				
RI														3			
ET															3		
DR																3	
OZ																	3

SAYI	Test örneğinin SAYI kadar doğru tanındığını gösterir
SAYI	Test örneğinin SAYI kadar yanlış tanındığını gösterir

EK-1. (devam) Konuşma tanıma sistemini test ederken kullanılan, her bir kullanıcıya ait üç test datası için elde edilen test sonuçları

Çizelge 1.5. Kullanıcı_5'e ait eğitim sonrasında kendi test verileri ile konuşma tanıma sisteminin çıktıkları (Cinsiyet: BAY)

KISALTIMA KOMUTLAR	İlgili Komutların Test için Kullanılan Örnekleri (MFCC ile)																
	BASLAT	KIRK	HAYIR	CIKIS	YIKAMA	ELLI	CAMASIR	BITIR	DOKSAN	SICAKLIK	BEYAZ	ALTMIS	DURULA	RENKLI	EVET	DURDUR	OTUZ
BT	3																
KK		3															
HR			3														
CS				3													
YA					3												
EI						2								1			
CR							3										
BR								3									
DN									3								
SK										3							
BZ											3						
AS												3					
DA													3				
RI														3			
ET															3		
DR										1						2	
OZ																	3

KISALTIMA KOMUTLAR	İlgili Komutların Test için Kullanılan Örnekleri (LSF ile)																
	BASLAT	KIRK	HAYIR	CIKIS	YIKAMA	ELLI	CAMASIR	BITIR	DOKSAN	SICAKLIK	BEYAZ	ALTMIS	DURULA	RENKLI	EVET	DURDUR	OTUZ
BT	3																
KK		3															
HR			3														
CS				3													
YA					3												
EI						2								1			
CR							2			1							
BR								3									
DN									3								
SK										3							
BZ											3						
AS										1		2					
DA													3				
RI														3			
ET															3		
DR													1			2	
OZ																	3

SAYI	Test örneğinin SAYI kadar doğru tanındığını gösterir
SAYI	Test örneğinin SAYI kadar yanlış tanındığını gösterir

EK-1. (devam) Konuşma tanıma sistemini test ederken kullanılan, her bir kullanıcıya ait üç test datası için elde edilen test sonuçları

Çizelge 1.6. Kullanıcı_6'ya ait eğitim sonrasında kendi test verileri ile konuşma tanıma sisteminin çıktıkları (Cinsiyet: BAY)

KISALTIMA KOMUTLAR	İlgili Komutların Test için Kullanılan Örnekleri (MFCC ile)																
	BASLAT	KIRK	HAYIR	CIKIS	YIKAMA	ELLI	CAMASIR	BITIR	DOKSAN	SICAKLIK	BEYAZ	ALTMIS	DURULA	RENKLI	EVET	DURDUR	OTUZ
BT	3																
KK		3															
HR			3														
CS				3													
YA					3												
EI						2								1			
CR							3										
BR								3									
DN									3								
SK										3							
BZ											3						
AS												3					
DA													3				
RI														3			
ET															3		
DR																3	
OZ																	3

KISALTIMA KOMUTLAR	İlgili Komutların Test için Kullanılan Örnekleri (LSF ile)																
	BASLAT	KIRK	HAYIR	CIKIS	YIKAMA	ELLI	CAMASIR	BITIR	DOKSAN	SICAKLIK	BEYAZ	ALTMIS	DURULA	RENKLI	EVET	DURDUR	OTUZ
BT	2									1							
KK		3															
HR			3														
CS				3													
YA					3												
EI						2				1							
CR							3										
BR				1				2									
DN									3								
SK										3							
BZ											3						
AS												3					
DA													3				
RI										1				2			
ET															3		
DR																3	
OZ																	3

SAYI	Test örneğinin SAYI kadar doğru tanındığını gösterir
SAYI	Test örneğinin SAYI kadar yanlış tanındığını gösterir

EK-1. (devam) Konuşma tanıma sistemini test ederken kullanılan, her bir kullanıcıya ait üç test datası için elde edilen test sonuçları

Çizelge 1.7. Kullanıcı_7'ye ait eğitim sonrasında kendi test verileri ile konuşma tanıma sisteminin çıktıkları (Cinsiyet: BAYAN)

KISALTIMA KOMUTLAR	İlgili Komutların Test için Kullanılan Örnekleri (MFCC ile)																
	BASLAT	KIRK	HAYIR	CIKIS	YIKAMA	ELLI	CAMASIR	BITIR	DOKSAN	SICAKLIK	BEYAZ	ALTMIS	DURULA	RENKLI	EVET	DURDUR	OTUZ
BT	3																
KK		3															
HR			2									1					
CS				3													
YA					3												
EI						3											
CR							3										
BR								3									
DN									3								
SK										3							
BZ											3						
AS												3					
DA													3				
RI														3			
ET						1									2		
DR																3	
OZ																	3

KISALTIMA KOMUTLAR	İlgili Komutların Test için Kullanılan Örnekleri (LSF ile)																
	BASLAT	KIRK	HAYIR	CIKIS	YIKAMA	ELLI	CAMASIR	BITIR	DOKSAN	SICAKLIK	BEYAZ	ALTMIS	DURULA	RENKLI	EVET	DURDUR	OTUZ
BT	3																
KK	1	2															
HR			2	1													
CS				2								1					
YA					3												
EI						3											
CR							3										
BR								3									
DN									3								
SK										3							
BZ											3						
AS												3					
DA												1	2				
RI														3			
ET						1									2		
DR	1											1				1	
OZ																	3

SAYI	Test örneğinin SAYI kadar doğru tanındığını gösterir
SAYI	Test örneğinin SAYI kadar yanlış tanındığını gösterir

EK-1. (devam) Konuşma tanıma sistemini test ederken kullanılan, her bir kullanıcıya ait üç test datası için elde edilen test sonuçları

Çizelge 1.8. Kullanıcı_8'e ait eğitim sonrasında kendi test verileri ile konuşma tanıma sisteminin çıktıkları (Cinsiyet: BAYAN)

KISALTIMA KOMUTLAR	İlgili Komutların Test için Kullanılan Örnekleri (MFCC ile)																
	BASLAT	KIRK	HAYIR	CIKIS	YIKAMA	ELLI	CAMASIR	BITIR	DOKSAN	SICAKLIK	BEYAZ	ALTMIS	DURULA	RENKLI	EVET	DURDUR	OTUZ
BT	3																
KK		3															
HR			3														
CS				3													
YA					3												
EI						3											
CR							3										
BR								3									
DN									3								
SK										3							
BZ											3						
AS												3					
DA													3				
RI														3			
ET															3		
DR																3	
OZ																	3

KISALTIMA KOMUTLAR	İlgili Komutların Test için Kullanılan Örnekleri (LSF ile)																
	BASLAT	KIRK	HAYIR	CIKIS	YIKAMA	ELLI	CAMASIR	BITIR	DOKSAN	SICAKLIK	BEYAZ	ALTMIS	DURULA	RENKLI	EVET	DURDUR	OTUZ
BT	3																
KK		3															
HR			3														
CS				3													
YA					3												
EI						3											
CR							3										
BR								3									
DN									2	1							
SK										3							
BZ											3						
AS												3					
DA													3				
RI														3			
ET															3		
DR																3	
OZ																	3

SAYI	Test örneğinin SAYI kadar doğru tanındığını gösterir
SAYI	Test örneğinin SAYI kadar yanlış tanındığını gösterir

EK-1. (devam) Konuşma tanıma sistemini test ederken kullanılan, her bir kullanıcıya ait üç test datası için elde edilen test sonuçları

Çizelge 1.9. Kullanıcı_9'a ait eğitim sonrasında kendi test verileri ile konuşma tanıma sisteminin çıktıkları (Cinsiyet: BAYAN)

KISALTIMA KOMUTLAR	İlgili Komutların Test için Kullanılan Örnekleri (MFCC ile)																
	BASLAT	KIRK	HAYIR	CIKIS	YIKAMA	ELLI	CAMASIR	BITIR	DOKSAN	SICAKLIK	BEYAZ	ALTMIS	DURULA	RENKLI	EVET	DURDUR	OTUZ
BT	3																
KK		3															
HR			3														
CS				2				1									
YA					3												
EI						3											
CR							3										
BR								3									
DN									3								
SK										3							
BZ											3						
AS												3					
DA													3				
RI								1						2			
ET															3		
DR																3	
OZ																	3

KISALTIMA KOMUTLAR	İlgili Komutların Test için Kullanılan Örnekleri (LSF ile)																
	BASLAT	KIRK	HAYIR	CIKIS	YIKAMA	ELLI	CAMASIR	BITIR	DOKSAN	SICAKLIK	BEYAZ	ALTMIS	DURULA	RENKLI	EVET	DURDUR	OTUZ
BT	3																
KK		3															
HR			3														
CS				3													
YA					2					1							
EI						3											
CR							3										
BR								2						1			
DN									3								
SK										3							
BZ											3						
AS												3					
DA													3				
RI										1				2			
ET															3		
DR																3	
OZ																	3

SAYI	Test örneğinin SAYI kadar doğru tanındığını gösterir
SAYI	Test örneğinin SAYI kadar yanlış tanındığını gösterir

EK-1. (devam) Konuşma tanıma sistemini test ederken kullanılan, her bir kullanıcıya ait üç test datası için elde edilen test sonuçları

Çizelge 1.10. Kullanıcı_10'e ait eğitim sonrasında kendi test verileri ile konuşma tanıma sisteminin çıktıkları (Cinsiyet: BAYAN)

KISALTIMA KOMUTLAR	İlgili Komutların Test için Kullanılan Örnekleri (MFCC ile)																
	BASLAT	KIRK	HAYIR	CIKIS	YIKAMA	ELLI	CAMASIR	BITIR	DOKSAN	SICAKLIK	BEYAZ	ALTMIS	DURULA	RENKLI	EVET	DURDUR	OTUZ
BT	3																
KK		3															
HR			3														
CS				3													
YA					3												
EI						3											
CR							3										
BR								3									
DN									3								
SK										3							
BZ											3						
AS												3					
DA													3				
RI														3			
ET															3		
DR																3	
OZ																	3

KISALTIMA KOMUTLAR	İlgili Komutların Test için Kullanılan Örnekleri (LSF ile)																
	BASLAT	KIRK	HAYIR	CIKIS	YIKAMA	ELLI	CAMASIR	BITIR	DOKSAN	SICAKLIK	BEYAZ	ALTMIS	DURULA	RENKLI	EVET	DURDUR	OTUZ
BT	3																
KK		3															
HR			3														
CS				3													
YA					2					1							
EI						3											
CR							3										
BR								2						1			
DN									3								
SK										3							
BZ											3						
AS												3					
DA													3				
RI														3			
ET															3		
DR																3	
OZ																	3

SAYI	Test örneğinin SAYI kadar doğru tanındığını gösterir
SAYI	Test örneğinin SAYI kadar yanlış tanındığını gösterir

ÖZGEÇMİŞ



Kişisel Bilgiler

Soyadı, adı : ÇİÇEK, Adem
Uyruğu : T.C.
Doğum tarihi ve yeri : 01.01.1988, Bursa
Medeni hali : Evli
Telefon : 0 (376) 226 23 23- 8258
Faks :
e-mail : adem.cicek@gazi.edu.tr

Eğitim

Derece	Eğitim Birimi	Mezuniyet tarihi
Lisans	Uludağ Üniversitesi/ Elektronik Mühendisliği	2010
Lise	Bursa Ahmet Hamdi Gökbayrak Anadolu Öğretmen Lisesi	2006

İş Deneyimi

Yıl	Yer	Görev
2012-Halen	Çankırı Karatekin Üniversitesi	Araştırma Görevlisi

Yabancı Dil

İngilizce

Yayımlar

Hobiler

Elektronik, Kitap okuma, Seyahat



GAZİ GELECEKTİR..