



**T.C.**  
**SELÇUK ÜNİVERSİTESİ**  
**FEN BİLİMLERİ ENSTİTÜSÜ**



**GİZLİ MARKOV MODELLERİNİ**  
**KULLANARAK NEFRET SÖYLEMİ TESPİTİ**

**MOHAMMED QASIM ABBAS MOUSA**

**YÜKSEK LİSANS TEZİ**

**Bilgisayar Mühendisliği Anabilim Dalı**

**Eylül-2021**  
**KONYA**  
**Her Hakkı Saklıdır**

## TEZ KABUL VE ONAYI

MOHAMMED QASIM ABBAS MOUSA tarafından hazırlanan “GİZLİ MARKOV MODELLERİNİ KULLANARAK NEFRET SÖYLEMİ TESPİTİ” adlı tez çalışması 16/09/2021 tarihinde aşağıdaki jüri tarafından oy birliği / ~~oy çokluğu~~ ile Selçuk Üniversitesi Fen Bilimleri Enstitüsü Bilgisayar Mühendisliği Anabilim Dalı’nda YÜKSEK LİSANS TEZİ olarak kabul edilmiştir.

### Jüri Üyeleri

### İmza

#### Başkan

Dr. Öğr. Üyesi Fahrettin HORASAN

.....

#### Danışman

Prof. Dr. Adem Alpaslan ALTUN

.....

#### Üye

Dr. Öğr. Üyesi Murat KÖKLÜ

.....

Yukarıdaki sonucu onaylarım.

Prof. Dr. Sait GEZGİN  
FBE Müdürü

## TEZ BİLDİRİMİ

Bu tezdeki bütün bilgilerin etik davranış ve akademik kurallar çerçevesinde elde edildiğini ve tez yazım kurallarına uygun olarak hazırlanan bu çalışmada bana ait olmayan her türlü ifade ve bilginin kaynağına eksiksiz atıf yapıldığını bildiririm.

## DECLARATION PAGE

I hereby declare that all information in this document has been obtained and presented in accordance with academic rules and ethical conduct. I also declare that, as required by these rules and conduct, I have fully cited and referenced all material and results that are not original to this work.

İmza

MOHAMMED QASIM ABBAS MOUSA

Tarih:

## ÖZET

### YÜKSEK LİSANS TEZİ

## GİZLİ MARKOV MODELLERİNİ KULLANARAK NEFRET SÖYLEMİ TESPİTİ

**MOHAMMED QASIM ABBAS MOUSA**

**Selçuk Üniversitesi Fen Bilimleri Enstitüsü  
Bilgisayar Mühendisliği Anabilim Dalı**

**Danışman: Prof. Dr. Adem Alpaslan ALTUN**

**2021, 38 Sayfa**

**Jüri**

**Prof. Dr. Adem Alpaslan ALTUN  
Dr. Öğr. Üyesi Fahrettin HORASAN  
Dr. Öğr. Üyesi Murat KÖKLÜ**

Konuşma tanıma, teknolojinin ilerlemesiyle birlikte araştırmalara konu olmuş ve çok farklı alanlarda kullanılır hale gelmiştir. Hayatımızın birçok alanında çeşitli etkileşimli konuşmaya dayalı uygulamalar mevcuttur. Konuşma tanıma sistemleri, yazmanın zorlaştığı durumlarda oldukça kullanışlıdır. Bu tez çalışmasında, kütüphaneler üzerinde izole kelime konuşma tanıma uygulamak için Gizli Markov modelinin uygulanması üzerine yapılan çalışmaları sunmaktadır. Ayrıca kelime tanıma sisteminde maksimum verim elde etmek için belirli bir kullanıcı için (kullanıcıya bağlı) kendi kendini seçen kelimeler kümesi kullanılarak sistemi geliştirmek ve eğitmek için Gizli Markov modelinin uygulanması üzerine yapılan çalışmaları sunmaktadır. Her bir kelime ile 15 varyasyon ile eğitim için kullanmak üzere toplam 105 kelime veren 15 farklı şekilde kaydedilen 7 alfabe örneği kullanılmıştır. Bu sistem gerçek dünyada sesli güvenlik sistemi kullanılarak sistem güvenliğinde ve özellikle çocuklar ve engelliler için kullanılabilir.

**Anahtar Kelimeler:** Baum-Welch algoritması, Gizli markov modelleri, Konuşmadan metne dönüştürme, N-gram, Nefret söylemi tespiti, Otomatik ses tanıma.

## **ABSTRACT**

### **MS THESIS**

#### **HATE SPEECH DETECTION USING HIDDEN MARKOV MODELS**

**MOHAMMED QASIM ABBAS MOUSA**

**THE GRADUATE SCHOOL OF NATURAL AND APPLIED SCIENCE OF  
SELÇUK UNIVERSITY  
THE DEGREE OF MASTER OF SCIENCE  
IN COMPUTER ENGINEERING**

**Advisor: Prof. Dr. Adem Alpaslan ALTUN**

**2021, 38 Pages**

#### **Jury**

**Prof. Dr. Adem Alpaslan ALTUN  
Assist. Prof. Dr. Fahrettin HORASAN  
Assist. Prof. Dr. Murat KÖKLÜ**

Speech recognition has been the subject of research with the advancement of technology and has become used in many different fields. There are various interactive speech-based applications in many areas of our lives. Speech recognition systems are very useful in situations where typing becomes difficult. In this thesis, we present the studies on the implementation of the Hidden Markov model to implement isolated word speech recognition on libraries. It also presents the studies on the implementation of the Hidden Markov model to develop and train the system using a set of self-selecting words for a given user (user dependent) to achieve maximum efficiency in the word recognition system. 7 alphabet samples recorded in 15 different ways were used, giving a total of 105 words to use for education, with 15 variations for each word. This system can be used in system security, especially for children and disabled people, by using a voice security system in the real world.

**Keywords:** Baum-Welch algorithm, Hidden markov models, Speech to text conversion, N-gram, Hate speech detection, Automatic voice recognition.

## ÖNSÖZ

Hem kötü hem de iyi zamanlarımda beni destekledikleri için aileme teşekkür ederim. Ayrıca, danışmanım Prof. Dr. Adem Alpaslan ALTUN'un bilgisi, desteği ve sabrından dolayı teşekkür ederim. Son olarak, diğer araştırma görevlisi arkadaşlarıma seminer çalışmam süresince bana moral ve destek oldukları için ve mutlu, huzurlu ve yapıcı bir iş ortamı sağladıkları için teşekkür ederim.

MOHAMMED QASIM ABBAS MOUSA  
KONYA-2021



# İÇİNDEKİLER

<b>ÖZET</b> .....	<b>iv</b>
<b>ABSTRACT</b> .....	<b>v</b>
<b>ÖNSÖZ</b> .....	<b>vi</b>
<b>İÇİNDEKİLER</b> .....	<b>vii</b>
<b>KISALTMALAR</b> .....	<b>viii</b>
<b>1. GİRİŞ</b> .....	<b>1</b>
<b>2. KAYNAK ARAŞTIRMASI</b> .....	<b>3</b>
2.1. Konuşma Tanıma Yaklaşımları .....	7
2.2. Nefret Söylemi Algılama İncelemesi:.....	11
2.2.1. Sözlükler .....	12
2.2.2. Mesafe Ölçütü .....	12
2.2.3. Kelime torbası (BOW).....	12
2.2.4. N-gram .....	13
<b>3. MATERYAL VE YÖNTEM</b> .....	<b>14</b>
3.1. Saklı Markov Modelleri .....	15
3.2. İleri Algoritma .....	15
3.3. Baum-Welch Algoritması .....	16
3.4. Sistem tasarımı.....	19
3.4.1. Öznitelik çıkarma.....	19
3.4.2. Eğitim.....	21
3.4.3. Sınıflandırma .....	21
<b>4. DENEYLER VE TARTIŞMALAR</b> .....	<b>22</b>
<b>5. SONUÇLAR VE ÖNERİLER</b> .....	<b>25</b>
5.1 Sonuçlar .....	25
5.2 Öneriler .....	26
<b>KAYNAKLAR</b> .....	<b>27</b>

## KISALTMALAR

### Kısaltmalar

ASR	: Automatic speech recognition (Otomatik konuşma tanıma).
AVTD	: Audiovisual Gemini Database (Görsel-İşitsel İkizler Veritabanı)
BOW	: Bag of Words (Kelime Torbası)
CL	: Computational Linguistics (Hesaplamalı Dilbilim).
CSJ	: Corpus of Spontaneous Japanese Project
DESA	: Deep Convolutional Neural Networks (Derin evrişimli sinir ağları)
DWT	: Discrete Wavelet Transform (Ayrık Dalgacık Dönüşümü)
EM	: Expectation Maximization (Beklenti maksimizasyonu)
FEL	: Foundation of Endangered Languages (Tehlike Altındaki Diller Vakfı).
FFT	: Fast Fourier Transform (Hızlı fourier dönüşümü).
HCI	: Human Computer Interaction (İnsan Bilgisayar Etkileşimi).
HLT	: Human Language Technologies (İnsan Dili Teknolojileri).
HMM	: Hidden Markov Model (Gizli Markov Modeli).
HSM	: Harmonic/Stochastic Model (Harmonik/Stokastik Model)
IBM	: International Business Machines (Uluslararası İş Makineleri).
IDA	: Institute for Defence Analysis (Savunma Analizi Enstitüsü).
IDT	: Integrated Device Technology (Entegre Cihaz Teknolojisi).
IWR	: Intelligent Word Recognition Systems (Akıllı Kelime Tanıma Sistemleri)
LPC	: Linear Predictive Coding (Doğrusal öngörülü kodlama).
LRBN	: Hidden Regression Bayesian network (Saklı Regresyon Bayes ağı)
MCE	: Minimum Classification error (Minimum Sınıflandırma hatası).
MFCC	: Mel-Frequency Cepstral Coefficients (Mel frekansı cepstral katsayıları).
NLP	: Natural Language Processing (Doğal Dil İşleme)
SIL	: Summer Institute of Linguistics (Yaz Dilbilim Enstitüsü).
SVM	: Destek Vektör Makineleri (Support Vector Machine)
VQ	: Vector Quantification Technique (Vektör Niceleme tekniği)
WPD	: Wavelet Packet Decomposition (Dalgacık Paket Ayrıştırma)

## 1. GİRİŞ

Konuşma analizleri, kişilerin ses verilerinden bilgi çıkarmasına olanak tanır. Duygu analizi, belirli konuşma ürünlerini veya hizmetlerini anlamaya çalışan konuşma analitiğinin kullanımlarından biridir. Bu analiz, temsilcilerin müşterileri telefonda konuşarak müşterileri etkileşim kurmasına ve ortaya çıkabilecek sorunları çözmesine yardımcı olmak için de kullanılabilir. Ses hassasiyetini analiz etmek için kullanılan iki ana yöntem vardır. Birincisi fonolojik modelleme, diğeri ise dilbilimsel modellemedir. Dilbilimsel modelleme, ses kayıtlarının metin dosyalarına çevrilmesini ve metin içeriğine göre analiz sonrası yapılmasını gerektirir. Bu tür modellemede, bazı ortamlarda daha yüksek olasılıkla belirli kelime ve ifadelerin kullanılması kabul edilebilir. Konuşma metinlerinde sık sık belirli kelimeler veya ifadelerle karşılaşılıyorsa, kelime dağarcığının zayıflığından bahsedilebilir. Birkaç araştırmacı, dilbilimsel modellemenin fonemik analiz için iyi sonuç verdiğiine dair güçlü kanıtlar sağlamıştır. Araştırmacılar, araştırmalarında modellerini oluştururken kelime özelliklerine, gramer yapılarına ve anlamsal özelliklere önem vermişlerdir. Ancak, her ses kaydı için iyi bir insan transkripti elde etmek göz korkutucu ve maliyetli olabilir. Ses kayıtlarını hassas bir şekilde metne dönüştürmek için iyi bir otomatik metin çıkarma sistemi oluşturmak zordur. Metindeki en ufak bir hata bile dilbilimsel model için büyük bir fark yaratabilir.

Akustik modellemede, ses verilerinin akustik özellikleri kullanılır. Bu özellikler arasında bilgisayar yazılımı ile kolayca hesaplanabilen perde, yoğunluk, konuşma hızı gibi özellikler bulunur. Birleştirildiğinde, bu özellikler duyguların bazı temel göstergelerini sağlayabilir. Fonemik özelliklerin birçok ticari veya psikolojik analizde kullanılması, bunların duygu analizinde de kullanılmasının bir başka nedenidir. Bununla birlikte, ses modellemeyle ilgili sorun, ses kayıtlarının kalitesinin nihai sonuç üzerinde güçlü bir etkiye sahip olmasıdır. Ses verilerinin akustik özellikleri kullanıldığı için, ses kalitesi, sonuçta iyi desenler yaratmada başarısız olacak bu akustik özellikler için doğru değerler elde etme yeteneğini büyük ölçüde etkileyebilir. Diğeri sorun, gerçek dünyanın herhangi bir yerinde her zaman rastgele arka plan gürültüsü olmasıdır. Model

eđitim verileri bu rastgele sesleri ayırt edemeyebilir, bu nedenle canlı sohbet sırasında duyarlılık analizi gerektiğinde üstesinden gelinmesi gereken bir zorluktur.

Ancak bu yöntem, uygulama kolaylığı ve etkinliği nedeniyle daha fazla arařtırmayı ve incelemeyi hak etmektedir.

Birçok ticari řirket, müşterilerden gelen aramalar için çağrı merkezi sistemlerini kullanır. Bu merkezlerde, genellikle santral personeli gelen aramalara yanıt verir ve müşterileri ilgili kişilere yönlendirir. Son derece yoğun çağrı trafiđi nedeniyle müşteriler çok uzun süre bekler. Buna çözüm olarak santral personeli sayısının artırılmasına ihtiyaç vardır. Bu durum firmalara maliyet açısından ek yükler getirmektedir. Merkez çalışanlarının mesai saatlerinden kaynaklanan yorgunluk ve stres gibi etkenler nedeniyle aramanın yanlış yönlendirilmesi gibi sorunlar da olasıdır (Afify ve Siohan, 2004).

Bilindiđi gibi diller iki kısma ayrılır; İnsanların kullandığı makine dili ve doğal dildir. Bilgisayarların, insanların dillerini anlamak ve onlarla iletişim kurmak için doğal dil işlemeyi kullanması gerekir. Kısacası, bilgisayarların doğal dilleri işleme sürecidir. Bu, insan-bilgisayar etkileşimi bilimi (HCI) çerçevesinde görülebilir veya hesaplamalı dilbilim (CL) olarak kabul edilebilir. Burada üzerinde arařtırma yapılan konu hesaplamalı dil işleme durumudur (Davis ve ark., 1952).

## 2. KAYNAK ARAŞTIRMASI

Ses tanıma ile ilgili ilk çalışmalar yirminci yüzyılın başlarında yapılmıştır. İlk ticari ürün, 1920'lerde piyasaya sürülen Radio Rex adlı bir ses tanıma oyunuydu. 1938'de Bell Laboratuvarı'ndaki araştırmacılar bir konuşma sentez makinesi geliştirdiler ve bunu New York'taki Dünya Fuarı'nda sundular (Sen ve ark., 2019).

Otomatik konuşma tanıma (ASR) için geliştirilen sistemlerdeki araştırmacılar daha çok fonemik algılara odaklanmışlardır.

1952'de, her sayının / sayının sesli kısımları sırasında tahmin edilen biçimlendirici frekanslar kullanılarak tek konuşmacıdan izole edilmiş bir sayı tanıma sistemi geliştirmişlerdir (Davis ve ark., 1952).

1959'da MIT Lincoln Lab'da yapılan bir çalışmada, spektral sesli harf rezonansı ölçülmüş ve 10 sesli harf tanıma aracı konuşmacıdan bağımsız olarak geliştirilmiştir (Hunt ve ark., 1959).

1960'larda ve 1970'lerde Japon araştırmacılar ses ve konuşma tanıma alanı üzerine çalışma yapmışlardır. Bilgisayarlar yeterince hızlı olmadığından, araştırmacılar sistemlerinin bir parçası olarak özel amaçlı bir yonga tasarladılar. Tokyo'da Nagata ve ark. H / W sistem çipini kullanarak konuşma tanıma gerçekleştirdiler (Nagata ve Morimoto, 1994). Başka bir çalışmada, Kyoto Üniversitesi'nden Sakai, H / W ses tanıma çipini geliştirmişlerdir (Sakai ve Doshita, 1962). 1963'te Nagata ve NEC Labs'taki meslektaşları bir sayı / sayı tanıma aracı geliştirerek uzun ve verimli bir araştırmaya öncülük etmişlerdir (Furui, 2018).

Özel şirketler de sağlam bir analiz üzerinde çalışmalar gerçekleştirmişlerdir. IBM araştırmacıları, geniş kelime dağarcığında ses tanıma üzerine çalışmışlardır. AT&T Bell Laboratuvarlarında, araştırmacılar konuşmacıdan bağımsız konuşma tanıma testlerini başlatmışlardır (Rabiner, 1989). Doğal olmayan konuşma tanıma için sözcüksel modelde sözcükleri temsil etmek için gereken farklı yapıların sayısını bulmak için çok sayıda kümeleme algoritması kullanılmıştır.

Carnegie Mellon Üniversitesi'nde geliştirilen Harphy sistemi, 1011 kelimenin kelime verilerini yüksek doğrulukla tanır. İlk defa, hesaplamaları etkili bir şekilde azaltmak ve en yakın eşleşen dizileri tanımlamak için sonlu bir ızgara kullanılmıştır.

1980'lerin başlarında, arařtırmaların odak noktası el yazısı sözcüklerin fonemik tanınması üzerine idi. 1980 yılında, Gizli Markov Modeli (HMM) yaklaşımı geliştirilen ana teknolojilerden biridir. IBM, Savunma Analizi Enstitüsü (IDA) ve Dragon sistemler, HMM üzerine arařtırma yapmış olmalarına rağmen 1980'lerin ortalarına kadar bu çalışmalar devam etmiştir (Arora ve Singh, 2012). Ses tanıma sorunlarına bir çözüm olarak sinir ağıları, 1980'lerin sonunda yeniden tanıtılan bir başka teknolojidir.

En önemli paradigma kayması, istatistiksel yöntemlerin tanıtılması olmuştur. Gizli bir Markov modeli (HMM) kullanan rastgele manipülasyon, özellikle 1970'lerin başlarında ses tanımaya büyük katkılarda bulunmuştur. 30 yıl sonra bile bu metodoloji kullanılmaya devam etmektedir. 1980'lerde konuşma tanıma arařtırmalarında şablon temelli yaklaşımlar öne çıkmıştır (Rabiner, 1989).

Sadeliğine rağmen N-gram'ın dil paradigmasının dikkat çekici derecede güçlü olduğu kanıtlanmıştır. Günümüzde, daha pratik konuşma tanıma sistemleri istatistiksel ve yapay zekaya yönelik yaklaşımlara dayanmaktadır ve günümüzde daha fazla gelişme ve sonuç elde edilmiştir. Model tabanlı spektral tahmin algoritması 1993 yılında geliştirilmiştir (Erell ve Weintraub, 1993). 1994 yılında Moshe J. Lasri, harflerin ve sayıların fonogramını incelemiş ve konuşma tanımayı geliştirmişlerdir (Walsh ve Horenczyk, 2001). HMM konuşma tanıma sistemine bakarsak, nöron tabanlı ağ eğitiminde önemli bir yenilik olan Destek Vektör ağının kullanımına dayanmaktadır (Dedman, 2009). Nam Soo Kim ve ark., HMM'ye dayalı sağlam bir olasılık çıktı dağılımını tahmin etmek için çeşitli yöntemler önermişlerdir (Kurian, 2014).

2000 yılında, Juang ve Furui geleneksel olarak Bayes'in çerçevesini çizmişler ve deneysel tanıma hatası en aza indirilerek bir optimizasyon problemini ele almışlardır (Juang ve Furui, 2000). Bu çalışmanın amacı, verilen verilere en uygun olandan ziyade en az tanıma hatası olan bir tanıyıcı tasarlamaktır. Hata minimizasyonu için kullanılan teknikler minimum sınıflandırma hatası ve maksimum karşılıklı bilgidir. Bu teknikler, konuşma tanıma performansına maksimum olabilirliğe dayalı bir yaklaşımı beraberinde getirmiştir. Ses modeli ile telaffuzu arasındaki akustik uyumsuzluğu azaltmak için maksimum olabilirlik stokastik eşleştirme yaklaşımı önerilmiştir (Aydın, 2008).

1999'da gürültülü bir ortamda ses tanıma için işitsel modele yeni bir yaklaşım önerilmiştir (Ostendorf ve ark., 1996). Bu yaklaşım, diğer modellerle karşılaştırıldığında sayısal olarak etkilidir. 2004 yılında, değişken Bayes kestirim tekniği geliştirilmiştir (Afify ve Siohan, 2004). Bu teknikte tahmin, parametrelerin sonraki dağılımına dayanmaktadır. Giuseppe Ricardi ve ark., ASR'de uyarlanabilir bir öğrenme problemini

çözmek için bir teknik geliştirmişlerdir (Zhang ve ark., 2017). Ayrıca 2005 yılında, performansı iyileştirmek için geliştirilen sürekli ve kelime dağarcığı çapında ses tanıma sisteminde önemli ilerleme kaydedilmiştir (Afify ve Siohan, 2004; Furui, 2018).

2005 yılında Furui, ünlü spektral rezonansını incelemek için bir çalışma yürütmüş ve telaffuz analizlerini gerçekleştirmiştir (Hasan ve ark., 2019). Japonya'da 5 yıllık ulusal bir proje olan Corpus of Spontaneous Japanese (CSJ) Project, Furui tarafından hayata geçirilmiştir (Nakatani, 2019). Bu proje için kullanılan veritabanı yaklaşık 7 milyon kelimedenden oluşmakta ve bu da 700 saatlik konuşmaya denk gelmektedir. Bu projede kullanılan teknikler, fonemik ve otomatik sesbirim özetlemenin yanı sıra fonemik modelleme, cümle sınırlarının tespiti, konuşma modelleme ve dil modeli koşullandırmadır (Juang ve Furui, 2000).

Konuşma doğrulama teknolojisi, özellikle otomatik konuşma için konuşma tanıma sistemlerinin sağlamlığını artırmak için araştırılmıştır (Lei ve ark., 2018). İnsanlar birbirleriyle konuştuklarında sesle iletişim oluştururlar. İletişim gürültülü bir ortamda gerçekleştiğinde, ses tanıma daha az başarılıdır. Araştırmacılar görsel yüz bilgisinin ses tanımada, özellikle dudak hareketinde kullanımını incelemişlerdir. Sonuçlar her iki bilgi formunun da özellikle gürültülü bir ortamda yalnızca görsel veya işitsel bilgiler kullanılarak daha iyi performans elde edildiğini göstermiştir.

Temsil açısından HMM tabanlı bir tanıma yaklaşımı, konuşma karakteristiğini geliştirmek için frekans eğilimli spektroskopi bilgisini kullanarak konuşma sinyallerini tanımaya yönelik bir yaklaşım sunar. Frekans spektral bilgisi, Mel frekans spektrumunun geleneksel konuşma tanıma yaklaşımına entegre edilmiştir. Frekans eğimi yaklaşımı, konuşma sinyalinin belirli bir hassasiyeti ile frekans kontrolünden yararlanır, bu da çözünürlük özelliklerinin karışmasına yol açan tanıma saçaklarıyla sonuçlanır. Çözünürlük frekans ayrımı, HMM tabanlı bir konuşma tanıma sistemi için bir eşleme yaklaşımıdır. Simülasyon sonuçları, konuşma tanıma sisteminin öğrenme doğruluğu ve öğrenme doğruluğu açısından konuşma tanıma kalitesi ölçümlerinde iyileşme göstermektedir.

(Ashraf ve ark., 2010), Urduca'da otomatik bir konuşma tanıma sistemi geliştirmek için bir yaklaşım geliştirmişlerdir. Önerilen sistem, mevcut sistemi iyileştirmek için istatistiksel bir yaklaşım olan Gizli Markov Modelini kullanan Sphinx4 adlı konuşma tanıma için açık kaynaklı bir çerçeveye dayanmaktadır. Bu çalışmada iki ayrı kelimedenden oluşan bir sistem tanıtılmıştır. Ayrıca, bu araştırma çalışmasının, orta ve

büyük ölçekli kelime hazinesi için Urduca Konuşma Tanıma Sisteminin geliştirilmesi için temel oluşturacağını öne sürmüşlerdir.

Poşul tarafından 2011 yılında yapılan tez çalışmasında, yönlendirme algoritmasına dayalı etkileşimli bir sesli yanıt sistemi geliştirilmiştir. Sistem müşteri puanına göre derecelendirilmektedir (Poşul, 2011).

Ranjan ve Dubey (2016), Methylan (Hintçe) dilinin izole edilmiş kelimelerinin otomatik olarak tanınması üzerine bir çalışma gerçekleştirmişlerdir. İzole edilmiş Akıllı Kelime Tanıma Sistemleri (IWR), bir kişinin bir mikrofon veya başka bir cihaz aracılığıyla söylediği kelimeyi tanır. Ranjan ve Dubey, telefon güvenliğinde ve çevrimiçi araştırmacılarda kullanılan bankacılık güvenlik çerçevelerini veya metilonek lehçedeki kelimeleri nasıl araştırıp yöneteceklerini keşfetmişlerdir. Konuşma sinyalleri, Mel-Frekansı Kepstral Katsayıları (MFCC) kullanılarak Maithili dil sözcüklerinden lehçe vektör düzenlemesine dönüştürülür. IWR, gizli bir Markov modeli (HMM) kullanılarak elde edilen eleman vektörleri üzerine çizilir. HMM tabanlı lehçe modelleri ve kelime fonemik modelleri, metilasyon sesbirimindeki ünlülerin her birini kapsayan Maithili kelimelerinin sınıflandırılmasındaki hata oranını beş katlı bir doğrulama işlemi kullanarak hesaplamak için kullanılmıştır (Kaur ve Kumar, 2018; Ranjan, 2021).

Gowsalya ve Sridevi (2019), Derin evrişimli sinir ağlarına (DESA) ve basit tekrarlayan birime (BTB) dayalı bir insan sesinden duyguları çıkarmak için bir yöntem önermişlerdir (Gowsalya ve Sridevi, 2019). İlk olarak logaritmik spektrum grafikleri, sabit, delta ve delta ile modifiye edilmiş akustik özelliklerden çıkarılır. Her bir ifadenin spektral eğiminin logaritmasının üç kanalı, DESA girdisi olarak zaman ekseninde farklı bölümlere ayrılmıştır. Daha sonra, ImageNet veri kümesiyle önceden eğitilmiş olan AlexNet, her bir segmentteki bu özellikleri ince ayar yapmak için öğrenmek için kullanılır. BTB modeli, edinilen segment seviyesinin özelliklerini birleştirir. Son olarak, SoftMax sınıflandırıcı, insan sesinden duygu çıkarma türlerini tanımlamak için kullanılır. EMO-DB ve CASIA'daki deneysel sonuçlar, modellerin konuşmaya dahil olan duyguları etkili bir şekilde tanıyabildiğini ve başka bir özellik türüne göre sınıflandırıcılardan daha iyi performans gösterdiğini göstermektedir.

(Lei ve ark., 2018), Saklı Regresyon Bayes ağını (LRBN) kullanarak uçtan uca konuşma tanıma modeli önermişlerdir. LRBN, kompakt bir yapıya ve hızlı parametre öğrenmeye sahiptir. Evrişimli sinir ağlarına kıyasla, daha basit, daha anlaşılır bir yapıya ve öğrenilmesi gereken daha az parametreye sahiptir. Deneysel sonuçlar, Tibet polikrom konuşma tanıma ve LRBN için LRBN / çift yönlü hibrit uzun menzilli bellek

bağlantısının geçici sınıflandırma yapısının özelliğinin çok dilli konuşma grupları arasında ayırım yapmaya yardımcı olduğunu göstermektedir.

2019 yılında Akın ve ark. tarafından yapılan tez çalışmasında, gömülü donanım sistemi üzerinde nokta seviyesinde konsolidasyon yöntemi kullanılarak ses tanıma işlemi gerçekleştirilmiştir. Bu çalışmada Görsel-İşitsel İkizler Veritabanı (AVTD) kullanılmıştır. Bu veri setinin seçilmesinin amacı, ses tanıma algoritmalarının benzer özelliklere sahip ikizler üzerindeki başarı oranlarını incelemektir (Akın ve ark., 2019).

## 2.1. Konuşma Tanıma Yaklaşımları

Konuşma tanımaya yönelik ilk yaklaşımlar, konuşma seslerini bulmaya ve bu seslere uygun etiketleri sağlamaya dayanıyordu. Bu, konuşma dilinde sonlu, ayırt edici fonetik birimlerin (fonemler) var olduğunu ve bu birimlerin zaman içinde konuşma sinyalinde kendini gösteren bir dizi akustik özellik ile geniş bir şekilde karakterize edildiğini varsayan akustik-fonetik yaklaşımın temelidir (Vanajakshi ve Mathivanan, 2017; Agarwal ve ark., 2018).

Fonetik birimlerin akustik özellikleri hem hoparlörlerle hem de komşu seslerle koartikülasyon etkisi olarak adlandırılan oldukça değişken olsa da akustik-fonetik yaklaşımda değişkenliği yöneten kuralların basit olduğu ve bir makine tarafından kolayca öğrenilebileceği varsayılır. Bu yaklaşıma göre, konuşmanın mesaj taşıyan bileşenleri, doğallık, sesli-sessiz sınıflandırma gibi ikili akustik özellikleri ve biçimlendirme yerleri, yüksek ve düşük frekansların oranı gibi sürekli özelliklerin belirlenmesi ile açıkça ortaya çıkarılır. Ticari uygulamalar için, bu yaklaşım uygun bir platform sağlamamıştır. Bu yaklaşım şu sırayla uygulanır: Spektral analiz, Özellik tespiti, Segmentasyon ve Etiketleme, Geçerli kelimeyi tanıma. Ayrıca sözcük sözlüğüne erişmek için dil kısıtlamaları uygulanır.

Şablon Temelli Yaklaşım için, bir prototipik konuşma kalıpları koleksiyonu, aday sözcüklerin sözlüğünü temsil eden referans örüntüleri olarak depolanır. Bilinmeyen bir sözlü ifade, bu referans şablonlarının her biri ile eşleştirilir ve en iyi eşleşen modelin bir kategorisi seçilir. Genellikle her kelime için şablon oluşturulur.

Bunun avantajı, fonemler gibi daha küçük akustik olarak daha deęişken birimlerin bölütlenmesi veya sınıflandırılmasından kaynaklanan hataların önlenemesidir. Sonuç olarak, her kelimenin kendi tam referans şablonuna sahip olması gerekir. Kelime haznesi boyutu arttıkça, şablon hazırlama ve eşleştirme maliyetli olmanın yanı sıra pratik de olmaz.

Makine tarafından konuşma tanıma, insan konuşma sinyalinin herhangi bir insan müdahalesi olmaksızın makine tarafından otomatik olarak metin biçimine dönüştürülmesi olarak tanımlanabilir. Sangwan ve ark. (2017) bu çalışmalarında, konuşma tanıma için DWT (Ayrık Dalgacık Dönüşümü) ve WPD (Dalgacık Paket Ayrıştırma) kullanan iki özellik çıkarma teknięi araştırmışlardır. Bu çalışmada, birincisi Ayrık Dalgacık Dönüşümüne dayalı, ikincisi Dalgacık Paket Ayrıştırmasına dayalı iki konuşma tanıyıcının ve dört sınıflandırıcının karşılaştırılması yapılmıştır. Önerilen yöntem, on basamak ve iki yüz konuşmacıdan oluşan bir veritabanı için uygulanarak 2000 konuşma örneğinden oluşan bir veritabanı haline getirilmiştir. Sonuçlar, ilgili konuşma tanıyıcıların doğruluk oranını sunar (Sangwan ve ark., 2017).

Banos ve ark. (2008), çalışmalarında, istatistiksel parametrik HMM tabanlı konuşma sentezi için konuşma modelleme üzerine yeni bir yaklaşımın ön sonuçlarını sunmuşlardır. Önerilen sistem esnek bir adım-asen kron harmonik/stokastik modele (HSM) dayanmaktadır. Konuşma, iki bileşenin üst üste binmesi olarak modellenmiştir: harmonik bileşen ve stokastik veya periyodik olmayan bileşen. Sentez modelinin adım eşzamansız olması, HMM tabanlı bir sentez sistemine doğrudan entegrasyona izin verir. HMM tabanlı konuşma sentezi sağlayan açık kaynaklı bir yazılım araç seti olan HTS kullanmışlardır. Önerilen HSM yöntemi, aynı konfigürasyon ve veri tabanına sahip HTS temel sistemi ile karşılaştırılmıştır. Bir dizi farklı deney yapılmıştır. Sonuçlar, yüksek kalitede sentezlenmiş sözcelere ulaşıldığını göstermektedir. Sentetik sesin kalitesi ve orijinal sese benzerlik açısından iki sistemi karşılaştıran küçük bir algı testi yapılmıştır. HSM, kalite testinde HTS temel sisteminden daha iyi performans gösterdiği tespit edilmiştir: HSM %53, HTS %35.3 ve kararsız %11.7. Orijinal sese benzerlikle ilgili olarak, HSM-HTS'den biraz daha iyi performans göstermiştir: HSM %35.3, HTS %29.4 ve kararsız %35.3 (Adiga ve Prasanna, 2019).

Deng ve ark. (2003), temiz konuşmayı bozan durağan olmayan akustik gürültünün yinelemeli tahmini için yeni bir algoritma ve sağlam konuşma tanıma için gürültü normalleştirilmiş SPLICE'in konuşma özellięi geliştirme çerçevesinde

algoritmanın başarılı bir uygulamasını önermişlerdir. Gürültü tahmin algoritması, kepsral alandaki akustik ortamın doğrusal olmayan bir modelini kullanır. Algoritmanın merkezinde, söz konusu doğrusal olmayan duruma parçalı doğrusal yaklaşımı geliştiren ve daha sonra gürültü tahmini için doğruluğunu artıran yenilikçi yinelemeli stokastik yaklaşım tekniği yer alır. Deneysel sonuçlar, daha önceki stokastik yaklaşım tekniğinin geliştirilmesinde yeni tanıtılan yinelemelerin kullanılmasının hayati önemini göstermiş ve gürültü tahmin algoritmasının performansının algoritmada gömülü olan unutma faktörüne duyarlılığını göstermiştir (Deng ve O'Shaughnessy, 2018).

Bilgiye dayalı yaklaşım konusunda bu yaklaşım konuşma tanıma sürecini, bir kişinin konuşmayı görselleştirme, analiz etme ve bir dizi ölçülmüş akustik özelliğe dayalı olarak karakterize etmede zekayı uygulama şekline göre mekanize etmeye odaklanır. Yapay Zeka yaklaşımı, akustik fonetik yaklaşım ve örüntü tanıma yaklaşımının bir melezidir. Hem akustik fonetik hem de şablon tabanlı yaklaşım, kendi başlarına, insan konuşma işlemlerine ilişkin önemli bilgiler keşfetmede başarısız olurlar. Sonuç olarak, hata analizi ve bilgiye dayalı sistem geliştirme ilerleme kaydedememiştir.

Geleneksel bilgiye dayalı yaklaşımda, üretim kuralları, deneysel dil bilgisinden veya konuşma spektrogramının gözlemlerinden, buluşsal olarak geliştirilir. Bu yaklaşımda temel darboğaz, fonetik sınıflandırmaların yapılabileceği uygun üretim kurallarını formüle etmek için dilbilimci, fonetikçi veya spektrogram okuma uzmanından bilginin edinilmesi ve sınıflandırılmasıdır. Samouelian (1994) tarafından önerilen sistemin iki ana avantajı vardır. İlk olarak, fonem sınıflandırmasına yönelik veriye dayalı yaklaşımı kullanır, böylece geniş bir konuşma veritabanını kullanarak konuşmacılar arası ve konuşmacılar arası konuşma değişkenliği sorununu çözmeye çalışır. İkincisi, herhangi bir özellik kombinasyonunu (parametrik veya akustik-fonetik) kullanarak karar ağaçları oluşturma yeteneğine sahiptir (Irtza ve Hussain, 2016). Bu araştırmanın amacı, geleneksel akustik-fonetik, konuşmaya özgü veya spektral tabanlı özellikler veya parametreler dahil olmak üzere çeşitli özellik kombinasyonlarını entegre edebilen ve geniş bir konuşma veritabanı kullanarak üretim kurallarını oluşturma sürecini otomatikleştirebilen esnek bir otomatik konuşma tanıma sistemi geliştirmektir (Ghai ve Singh, 2012).

Bağlantısal Yaklaşım için, bu yaklaşım bilginin temsiline ve bilgi kaynaklarının entegrasyonuna odaklanır. Bağlantısal konuşma modeli, konuşma tanımadaki en genç gelişmedir. Bağlantıcı modellerde, bilgi veya kısıtlamalar, tek tek

birimler, kurallar veya prosedürler halinde kodlanmak yerine birçok basit hesaplama birimine dağıtılır. Belirsizlik aynı zamanda, tek bir birimin olasılıkları veya olasılık yoğunluk fonksiyonları olarak değil, birçok birimdeki faaliyet modeline göre modellenir. Hesaplama birimleri doğası gereği basittir ve bilgi, bağlantılı işlem öğeleri arasındaki bağlantılarda ve etkileşimlerde yatmaktadır. Bu birimlerin ağları tarafından gerçekleştirilen hesaplama tarzı, sinir sistemindeki hesaplama tarzına benzer. Bağlantısal öğrenme, bir işleme öğeleri ağını optimize etmeyi veya düzenlemeyi amaçlar. Altta yatan işlem öğesinin basitliği ve tekdüzeliği, bağlantı modellerini donanım uygulaması için çekici kılar, bu da bir ağın çalışmasının verimli bir şekilde simüle edilmesini sağlar. Bağlantısallık, makul bir biliş modeli olarak büyük umut vaat ediyor gibi görünmektedir. Örneğin, Savage ve diğerleri tarafından yeni bir teknik tasarlanmıştır (Pawar ve ark., 2018). Burada her VQ'nun çıkışına bir YSA bloğu eklenmiştir. Her bir kelime için her bir VQ temsilinin global uzaklıkları, YSA'ya beslenmiştir. Bu teknik, tek başına VQ ile elde edilen sonuçlara kıyasla tanıma oranının artmasına neden olmuştur. VQ ve YSA'nın bir arada uygulanmasıyla %100 tanıma oranı elde etmek harikadır ancak tanıma oranındaki artış önemli değildir. Kuantizasyon hatasından muzdarip geleneksel VQ/HMM'yi geliştirmek için, dağıtılmış vektör nicelme tekniği (DVQ), Ayrık HMM tabanlı izole edilmiş kelime ASR sistemini geliştirmek için Debyeche ve diğerleri tarafından tasarlanmıştır (Juang ve Furui, 2000). Bu tekniğin hibrit uygulaması, K-ortalama DVQ ve NN-DVQ'yu içerir ve bu ikisinden NN-DVQ, hata azaltma açısından cesaret verici sonuçlar vermiştir. Hatulan ve ark., Filipin dili için bir girişimde bulunmuş ve ANN/HMM yaklaşımını kullanarak bir konuşmadan metne dönüştürücü geliştirmişlerdir. Ağları belirli bir hedefe eğitmek için ileri beslemeli YSA kullanmışlar ve çıktılar, en olası fonem sırasını tahmin etmek için HMM'ye girdi olarak beslenmiştir (Chan ve ark., 2007).

Destek Vektör Makineleri (SVM) yaklaşımı için, SVM, ayırt edici bir yaklaşım kullanan örüntü tanıma için en gelişmiş güçlü sınıflandırıcılardan biridir. Örnekler ve sınıflandırıcı sınırı arasındaki optimize edilmiş kenar boşluğu, görünmeyen kalıpların genelleştirilmesine yardımcı olur. SVM'ler, veri sınıflandırması için doğrusal ve doğrusal olmayan ayırma hiper düzlemlerini kullanır. Ancak, SVM'ler yalnızca sabit uzunluktaki veri vektörlerini sınıflandırabildiğinden, bu yöntem değişken uzunluklu veri sınıflandırmasını içeren görevlere kolaylıkla uygulanamaz. Değişken uzunluklu veriler, SVM'ler kullanılmadan önce sabit uzunluklu vektörlere dönüştürülmelidir. Maksimum

marj uydurma işlevlerine sahip geliştirilmiş bir doğrusal sınıflandırıcıdır. Bu uydurma işlevi, sınıflandırıcının daha iyi geliştirilmesine yardımcı olan bir düzenleme sağlar. SVM, az sayıda özellik kullanarak model karmaşıklığını kontrol etmek yerine, modelinin VC boyutlarını kontrol ederek model karmaşıklığını kontrol eder. Bu yöntem, boyutsallıktan bağımsızdır ve çok sayıda doğrusal olmayan özelliğin inşasına izin veren ve daha sonra eğitim sırasında uyarlanabilir özellik seçimi gerçekleştiren çok büyük boyutlu uzayların alanlarını kullanabilir. Tüm doğrusal olmayan unsurları özelliklere kaydırarak SVM, VC boyutlarının bilindiği doğrusal modeli kullanabilir (Juang ve Furui, 2000). Sendra ve ark., çerçeve düzeyinde kararlar almak için SVM'yi ve tanınan kelimelerin zincirini elde etmek için Token Geçiş algoritmasını uygulayarak saf SVM tabanlı sürekli konuşma tanıyıcı üzerinde çalıştılar. Token Geçiş Modeli, bir cümledeki kelime sayısı hakkındaki belirsizliği yönetmek için sürekli konuşma tanıma amaçlı Viterbi algoritmasının bir uzantısıdır. Deneylerden elde edilen sonuçlar, küçük bir veritabanı ile tanıma doğruluğunun SVM'ler ile arttığı, ancak büyük veritabanı ile aynı sonucun büyük hesaplama çabası pahasına elde edildiği sonucuna varmıştır (Padmanabhan ve Johnson Premkumar, 2015).

## 2.2. Nefret Söylemi Algılama İncelemesi:

İnternet ve sosyal ağların özellikle iletişim amaçlı kullanımı son yıllarda önemli ölçüde artmıştır. Bu büyüme aynı zamanda daha agresif iletişimin benimsenmesine neden olmaktadır. Bu nedenle, hükümetlerin ve sosyal ağ platformlarının bu tür iletişimi tespit edecek araçlara sahip olması önemlidir. Çevrimiçi olarak yayınlanan nefret söylemini tespit etmek büyük bir sorun olarak karşımıza çıkmaktadır.

İlk olarak “Nefret Söylemi Nedir?” sorusunu inceleyerek bunun teorik yönleri analiz edilmelidir: Kavramın farklı tanımları arasında ayırım yapmak, belirli nefret söylemi vakalarını analiz etmek, nefret söylemini diğer yakın kavramlarla ilişkilendirmek, nefret söyleminin nasıl geliştiğini ve nefret söyleminin ana hedeflerinin kimler olduğunu görmek gibi.

Metnin bir bölümünde nefret söylemi olup olmadığına karar vermek, insanlar için bile doğrusal değildir. Bu nedenle, makinelerin nefret söylemini tanımlamasını

istiyorsak, bu görevi kolaylaştırmak için bu kavramı net bir şekilde tanımlamak önemlidir (Ross ve ark., 2017). Nefret söylemi genel olarak, yayan, kışkırtan, teşvik eden veya belirli bir gruba karşı nefreti, ayrımcılığı veya düşmanlığı haklı çıkarmak üzere genel bir hoşgörüsüzlük iklimine katkıda bulunma ve bunu da verilen gruplara yönelik saldırıları daha olası hale getirme olarak tanımlanabilir (Delgado ve Stefancic, 2014).

Bu çalışmada metin madenciliğinde halihazırda bilinen stratejiler, nefret söyleminin otomatik olarak tespit edilmesi sorununa uyarlanmaya çalışılmıştır.

### 2.2.1. Sözlükler

Metin madenciliğindeki bir strateji, sözlüklerin kullanılmasıdır. Bu yaklaşım, metinde aranan ve sayılan kelimelerin bir listesini (sözlük) oluşturmayı içerir. Bu frekanslar doğrudan özellik olarak veya puanları hesaplamak için kullanılabilir. Nefret söyleminin tespiti durumunda bu, (Kim ve ark., 1999) 'dan toplanan İçerik kelimeleri (hakaret ve küfür, tepki kelimeleri, kişisel zamirler gibi) kullanılarak yapılmıştır.

### 2.2.2. Mesafe Ölçütü

Bazı araştırmalar, metin mesajlarında rahatsız edici kelimelerin kasıtlı bir yazım hatasıyla, genellikle tek bir karakter ikamesi ile karartılmasının mümkün olduğuna işaret etmiştir (Biswas, 2016). Bu terimlere örnek olarak "@ss", "sh1t", "nagger" veya "joo" gibi sesteş sözcükler verilebilir. Levenshtein mesafesi, yani bir dizgiyi diğerine dönüştürmek için gereken minimum düzenleme sayısı bu amaç için kullanılabilir (Anggraini ve Tursina, 2019). Mesafe ölçüsü, sözlüğe dayalı yaklaşımları tamamlamak için kullanılabilir.

### 2.2.3. Kelime torbası (BOW)

Doğal Dil İşleme (NLP) için makine öğrenimi gerçekleştirirken, ortak ilk adım, bir torba kelime veya bir torba n-gram gibi özellikleri çıkarmaktır. Kolaylık sağlamak için hem kelime torbasına hem de n-gram torbasına kelime torbası (BOW) olarak değinilebilir. Daha sonra bu özellikler, makine öğrenimi yöntemlerinin eğitim ve sınıflandırma için özellikleri işleyebilmesi için tek-sıcak vektörler olarak temsil edilirler

(Shao ve ark., 2018). Bu durumda, sözlüklerde olduğu gibi önceden tanımlanmış bir kelime grubu yerine eğitim verilerinde yer alan kelimelere dayalı olarak bir külliyat oluşturulur. Tüm kelimeleri topladıktan sonra, her birinin frekansı bir sınıflandırıcıyı eğitmek için bir özellik olarak kullanılır. Bu tür yaklaşımların dezavantajları, kelime dizisinin ve ayrıca sözdizimsel ve anlamsal içeriğinin göz ardı edilmesidir. Bu nedenle, kelimelerin farklı bağlamlarda kullanılması yanlış sınıflandırmaya yol açabilir. Bu sınırlamanın üstesinden gelmek için N-gramlar benimsenebilir.

#### 2.2.4. N-gram

N-gram modelleme, dil modelleme ve doğal dil işleme alanlarında kullanılan popüler bir özellik tanımlama ve analiz yaklaşımıdır (Ahmed ve ark., 2017). N-gram, uzunluğu N olan bitişik bir öge dizisidir. Bir dizi sözcük, bayt, hece veya karakter olabilir. Metin sınıflandırmasında en çok kullanılan N-gram modelleri, kelime tabanlı ve karakter tabanlı N-gramlardır. Araştırmacılar çalışmalarında, belgenin içeriğini temsil etmek ve belgeyi sınıflandırmak ve özellikler oluşturmak için kelime tabanlı N-gram kullanmıştır. Sahte ve dürüst haber makalelerini ayırt etmek için basit bir N-gram tabanlı sınıflandırıcı geliştirilmiştir. Buradaki fikir, sahte ve doğru haber makalelerini temsil etmek için eğitim verilerinden çeşitli N-gram frekans profilleri oluşturmaktır. Sözcüklere dayalı birkaç temel N-gram özneliği kullanılmış ve N-gram uzunluğunun farklı sınıflandırma algoritmalarının doğruluğu üzerindeki etkisi incelenmiştir.

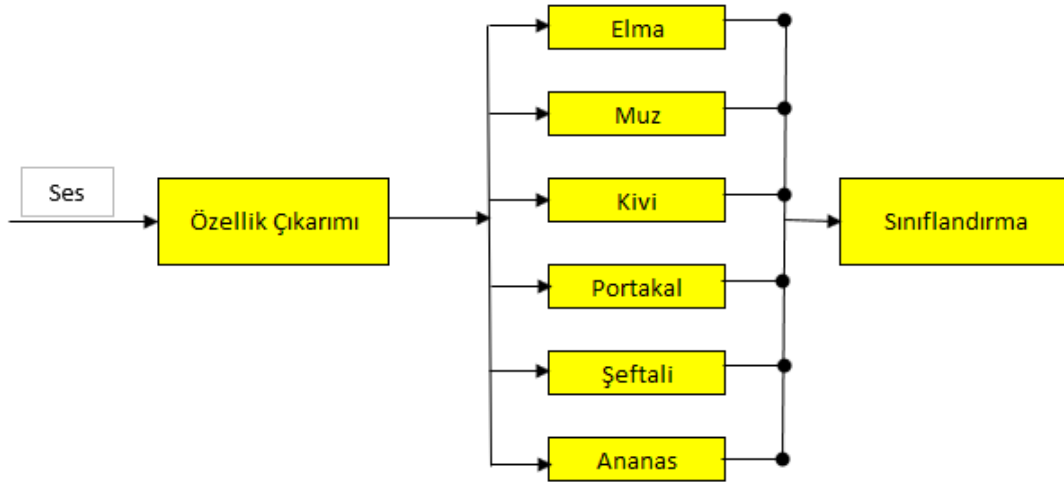
Nefret söylemi otomatik tespitinde ve ilgili görevlerde en çok kullanılan tekniklerden biridir (Alfina ve ark., 2017). En yaygın N-gram yaklaşımı, sıralı kelimelerin N boyutuna sahip listelerde birleştirilmesinden oluşur. Bu durumda amaç, N boyutundaki tüm ifadeleri numaralandırmak ve tüm oluşumları saymaktır. Bu, sınıflandırıcıların performansının iyileştirilmesine izin verir, çünkü bir dereceye kadar her kelimenin bağlamını içerir. Kelime kullanmak yerine, karakter veya heceli N-gram kullanmak da mümkündür. Bu yaklaşım, kelimelerin kullanıldığı zamanki gibi yazım değişikliklerine karşı çok duyarlı değildir. Karakter N-gram özelliklerinin, istismar edici dil tespitinin spesifik problemi için belirteç N-gram özelliklerinden daha öngörücü olduğu kanıtlanmıştır. N-gram özelliklerinin, nefret söyleminin otomatik olarak algılanması sorununda genellikle yüksek öngörülebilir olduğunu, ancak diğerleriyle birleştirildiğinde daha iyi performans gösterdiği açıktır.

### 3. MATERYAL VE YÖNTEM

Ses tanıma, özellik son yıllarda üzerinde birçok çalışmanın yapıldığı zorlu bir alandır. Bu alandaki en başarılı çalışmalara imza atanlardan biri de, 1989'daki Markov modelleri ile Rabiner olmuştur (Rabiner, 1989) .

İyi çalışan ve genel özelliklere sahip bir ses tanıma programı, başta çocuklar, okuma-yazma bilmeyenler ve engelli insanlar olmak üzere, herkese yönelik daha verimli bir iletişim sağlayacaktır. Bir ses tanıma programı, ardıl ya da simultane çeviri olarak tanımlanan sesli çeviriler için de bir alt sistem teşkil edebilir.

Bu tez kapsamında uygulanan ses tanıma sisteminde seslerin tanınması için her kelimeye yönelik olarak Saklı Markov Modeli (HMM) kullanılmıştır. Modeller, etiketli veri setleriyle eğitilir ve sınıflandırma, özelliklerin her modele aktarılması ve ardından en iyi eşleşmenin seçilmesi ile gerçekleştirilir (Şekil 3.1).



Şekil 3.1: Sistemin akış şeması. Ses sinyalinden çıkarılan özellikler her bir kelime modeline aktarılır ve en iyi eşleşme seçilir.

### 3.1. Saklı Markov Modelleri

Bu tez çalışmasında, Saklı Markov Modelleri temel alınmış olup, uygulanan en önemli iki algoritma detaylı şekilde açıklanmıştır.

Saklı modelde, gözlemlenebilir çıktının çok değişkenli bir Gauss dağılımı tarafından üretildiği varsayılır. Bir başka ifadeyle, her durum için bir ortalama vektör ve kovaryans matrisi vardır. Ayrıca, durum geçiş olasılıklarının, saklı Markov zinciri homojen olacak şekilde zamandan bağımsız olduğu varsayılmıştır.

Bu bölümde, bu tez çalışmasında kullanılan bir saklı Markov modelini açıklamak için gerekli gösterim sunulmuştur.

Toplam  $N$  sayıda durum olsun. Geçiş olasılığı  $A$  matrisindeki bir ASSR ögesi,  $s$  durumundan  $st$  durumuna geçiş olasılığını belirtir ve zincirin  $s$  durumunda başlama olasılığı  $\pi_s$ 'dir.  $s$  durumundan gözlemlenebilir çıktıyı modelleyen çok değişkenli Gauss dağılımı için ortalama vektör ve kovaryans matrisi sırasıyla  $\mu_s$  ve  $\Sigma_s$ 'dir. Bir  $o$  gözlemi için  $bs(o)$ ,  $s$  durumunun çok değişkenli Gauss dağılımının  $o$  değerlerinde olasılık yoğunluğunu belirtir. Yerine göre saklı Markov modelini  $\lambda = \{A, \pi, \mu, \Sigma\}$  olarak tanımlayan parametreleri de gösterilir (Bishop, 2006).

### 3.2. İleri Algoritma

Belirli bir model için Denklem 3.1'de gösterildiği üzere bir  $o_1, \dots, o_T$  gözleminin olasılık yoğunluğu hesaplanır. Bu, ses sinyalini büyük olasılıkla oluşturan modeli (örneğin kelime) seçmek için kullanılır.

Tekrarlamalı yapı, sorunu tüm  $s_T$  için  $f(\mathbf{o}_1, \dots, \mathbf{o}_T, s_T; \lambda)$ 'e ihtiyaç duymaktan, tüm  $s_{T-1}$ 'ler için  $f(\mathbf{o}_1, \dots, \mathbf{o}_{T-1}, s_{T-1}; \lambda)$ 'ye ihtiyaç duymaya indirgediğimizde ortaya çıkar. Gösterimi daha anlaşılır hale getirmek için, ileri değişken Denklem 3.2'de formüle edilmiştir.

$$\begin{aligned}
f(o_1, \dots, o_T; \lambda) &= \sum_{s_T} f(o_1, \dots, o_T; \lambda) \\
&= \sum_{s_T} f(o_T | o_1, \dots, o_{T-1}, s_T; \lambda) f(o_1, \dots, o_{T-1}, s_T; \lambda) \\
&= \sum_{s_T} b_{s_T}(o_T) \sum_{s_{T-1}} f(o_1, \dots, o_{T-1}, s_{T-1}; \lambda) \\
&= \sum_{s_T} b_{s_T}(o_T) \sum_{s_{T-1}} f(s_T | o_1, \dots, o_{T-1}, s_{T-1}; \lambda) f(o_1, \dots, o_{T-1}, s_{T-1}; \lambda) \\
&= \sum_{s_T} b_{s_T}(o_T) \sum_{s_{T-1}} a_{s_{T-1} s_T} f(o_1, \dots, o_{T-1}, s_{T-1}; \lambda)
\end{aligned} \tag{3.1}$$

$$\begin{aligned}
\alpha_1(s) &= f(o_1, S_1 = s; \lambda) = b_s(o_1) \pi_s \\
\alpha_t(s) &= f(o_1, \dots, o_t, S_1 = s; \lambda) = b_s(o_t) \sum_{s'} a_{s' s} \alpha_{t-1}(s')
\end{aligned} \tag{3.2}$$

Elde edilen çözümü Denklem 3.3'teki gibi ifade etmek mümkündür:

$$f(o_1, \dots, o_T; \lambda) = \sum_s \alpha_T(s) \tag{3.3}$$

Yukarıdan aşağıya (zamanda geriye doğru) basit şekilde uygulandığında, üstel olarak tekrarlamalı yapı nedeniyle bu çok da fayda sağlamayacaktır. Bununla birlikte, bu basit algoritma, ileri değişkenleri aşağıdan yukarıya (zamanda ileriye doğru) hesaplandığında dinamik programlama kullanılarak verimli bir varyanta kolayca dönüştürülebilir.  $\alpha_t(s)$  değerini tüm  $s$  durumları için, öncelikle  $t = 1$  için ve sonra  $T$ 'ye kadar olacak şekilde hesaplanır. Bu şekilde, önceki zaman adımındaki tüm ileri değişkenler, gerektiğinde kolayca kullanılabilir.

### 3.3. Baum-Welch Algoritması

Gözlem olasılığını en üst düzeye çıkaran  $\lambda$  parametreleri bulunmak istendiğinde bu, saklı Markov modelini ses sinyalleri ile eğitmek için kullanılır.

Baum-Welch algoritması, başlatma değerlerinden yerel olarak en uygun çözüme yakınsayan yinelemeli bir beklenti maksimizasyonu algoritmasıdır (Elharati ve ark., 2020).

$M$  adımında, parametreler, Denklem 3.4'te gösterildiği gibi, sezgisel şekilde güncellenir:

$$\begin{aligned}
 \pi_s &:= \bar{\pi}_s = \frac{t=1\_de\_s\_durumunda\_beklenen\_iterasyon\_sayısı}{t=1\_de\_beklenen\_iterasyon\_sayısı} \\
 a_{ss} &:= \bar{a}_{ss} = \frac{s\_den\_s\_ne\_beklenen\_geçiş\_sayısı}{s\_den\_beklenen\_geçiş\_sayısı} \\
 \mu_s &:= \bar{\mu}_s = s\_durumunda\_beklenen\_gözlem \\
 \sum_s &:= \sum_s = s\_durumunda\_beklenen\_ko\_var\_yans
 \end{aligned} \tag{3.4}$$

Dolayısıyla E-adımı, sabit bir  $\lambda$  için bu beklentilerin hesaplanmasından oluşur. Burada,  $V_s^{(t)}$ ,  $t$  adımında  $s$  durumundan geçiş olayını ve  $V_{s,s'}^{(t)}$ ,  $t$ 'de  $s$ 'den  $s'$  'ye geçiş olayını gösterir. Daha sonra bu beklentiler, Denklem 3.5 ile gösterge fonksiyonları ve beklentinin doğrusallığı kullanılarak hesaplanır.

$$\begin{aligned}
 \bar{\pi}_s &= E\{1[V_s^{(1)}]\} = P(V_s^{(1)}) \\
 \bar{a}_{ss} &= \frac{E\left\{\sum_t 1[V_{s,s'}^{(t)}]\right\}}{E\left\{\sum_t 1[V_s^{(t)}]\right\}} = \frac{\sum_t P(V_{s,s'}^{(t)})}{\sum_t P(V_s^{(t)})} \\
 \bar{\mu}_s &= \frac{E\left\{\sum_t 1[V_s^{(t)}] o_t\right\}}{E\left\{\sum_t 1[V_s^{(t)}]\right\}} = \frac{\sum_t P(V_s^{(t)}) o_t}{\sum_t P(V_s^{(t)})} \\
 \sum_s &= \frac{E\left\{\sum_t 1[V_s^{(t)}] \left(o_t o_t^T - \overline{\mu_s \mu_s^T}\right)\right\}}{E\left\{\sum_t 1[V_s^{(t)}]\right\}} = \frac{\sum_t P(V_s^{(t)}) o_t o_t^T}{\sum_t P(V_s^{(t)})} - \overline{\mu_s \mu_s^T}
 \end{aligned} \tag{3.5}$$

Yukarıda verilen denklemlerde italik olmayan T değeri, transpozu ifade eder ve zamansal bir durumla ilgili değildir. Bu olasılıkları hesaplayabilmek için, önceden tanımlanan ileri değişkene çok benzeyen geriye dönük değişkeni Denklem 3.6 ile ifade edilir.

$$\beta_T(s) \equiv 1$$

$$\beta_T(s) \equiv f(o_{t+1}, \dots, o_T | S_t = s; \lambda) = \sum_{s'} a_{ss'} b_s(o_{t+1}) \beta_{t+1}(s') \quad (3.6)$$

Geriye dönük değişken, öncelikle son zaman adımı için hesaplanır ve daha sonra dinamik programlama ile uygulandığında zamanda geriye doğru hesaplanır (ileri değişken için ayrıntılı olarak açıklanan prosedürün tersi). Ardından olasılıkları Rabiner tarafından kullanılan sembollerle yeniden adlandırıp Denklem 3.7'de gösterildiği gibi ileri ve geri değişkenlerle ifade edilir:

$$\begin{aligned} \gamma_t(s) &\equiv P(V_s^{(t)}) = P(S_t = s | o_1, \dots, o_T; \lambda) \\ &= \frac{f(o_1, \dots, o_T | S_t = s) P(S_t = s)}{f(o_1, \dots, o_T)} \\ &= \frac{f(o_1, \dots, o_t | S_t = s) f(o_{t+1}, \dots, o_T | S_t = s)}{f(o_1, \dots, o_T)} \\ &= \frac{\alpha_t(s) \beta(s)}{f(o_1, \dots, o_T)} \end{aligned} \quad (3.7)$$

$$\begin{aligned} \xi_t(s, s') &\equiv P(V_{s, s'}^{(t)}) = P(S_t = s, S_{t+1} = s' | o_1, \dots, o_T; \lambda) \\ &= \frac{\alpha_t(s) b_s(o_{t+1}) a_{ss'} \beta_{t+1}(s')}{f(o_1, \dots, o_T)} \end{aligned}$$

Son olarak, aşağıdaki ifadeler elde edilir (Denklem 3.8):

$$\begin{aligned}
\bar{\pi}_s &= \gamma_1(s) \\
\bar{a}_{ss'} &= \frac{\sum_t \xi_t(s, s')}{\sum_t \gamma_t(s)} \\
\bar{\mu}_s &= \frac{\sum_t \gamma_t(s) o_t}{\sum_t \gamma_t(s)} \\
\bar{\Sigma}_s &= \frac{\sum_t \gamma_t(s) o_t o_t^T}{\sum_t \gamma_t(s)} - \bar{\mu}_s \bar{\mu}_s^T
\end{aligned} \tag{3.8}$$

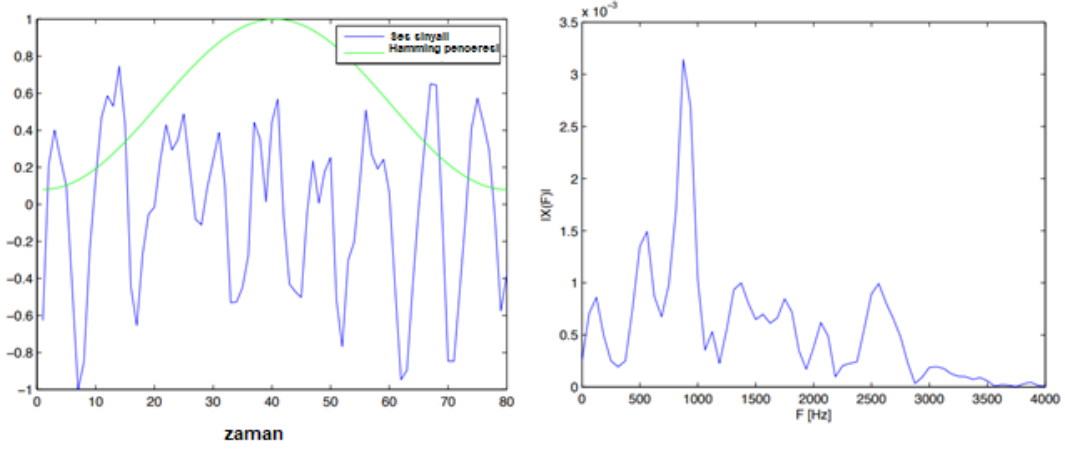
E-adımında,  $\lambda$  parametreleri sabitlenirken tüm  $s$ ,  $s'$  ve  $t$  için  $\gamma_t(s)$  ve  $\xi_t(s, s')$  hesaplamasına indirgenir ve ardından M-adımı, E-adımında yapılan hesaplamalar kullanılarak  $\lambda$  güncellenir. Bu işlem, tatmin edici sonuç elde edilinceye kadar devam ettirilir (Huang ve ark., 2001).

### 3.4. Sistem tasarımı

#### 3.4.1. Öznitelik çıkarma

Kaynak konuşma 8000 Hz'de örneklenmiş ve 16 bit ile kuantalanmıştır. Sinyal, 10 ms'lik konuşmaya karşılık gelen 80 örnekten oluşan kısa çerçevelere bölünür. Çerçeveler, her iki tarafta 20 örnek olacak şekilde üst üste binmektedir. Buradaki amaç, boğazın nispeten sınırlı esnekliği nedeniyle konuşmanın bu kısa süre boyunca durağanlığa yakın olmasıdır. Özellikler frekans alanından seçilir, ancak hızlı Fourier dönüşümünü kullanarak bu noktaya varmadan önce, sinyalin çerçevelenmesinden kaynaklanan spektral sızıntıyı azaltmak için bir Hamming penceresi ile çarpılır (Şekil 3.2).

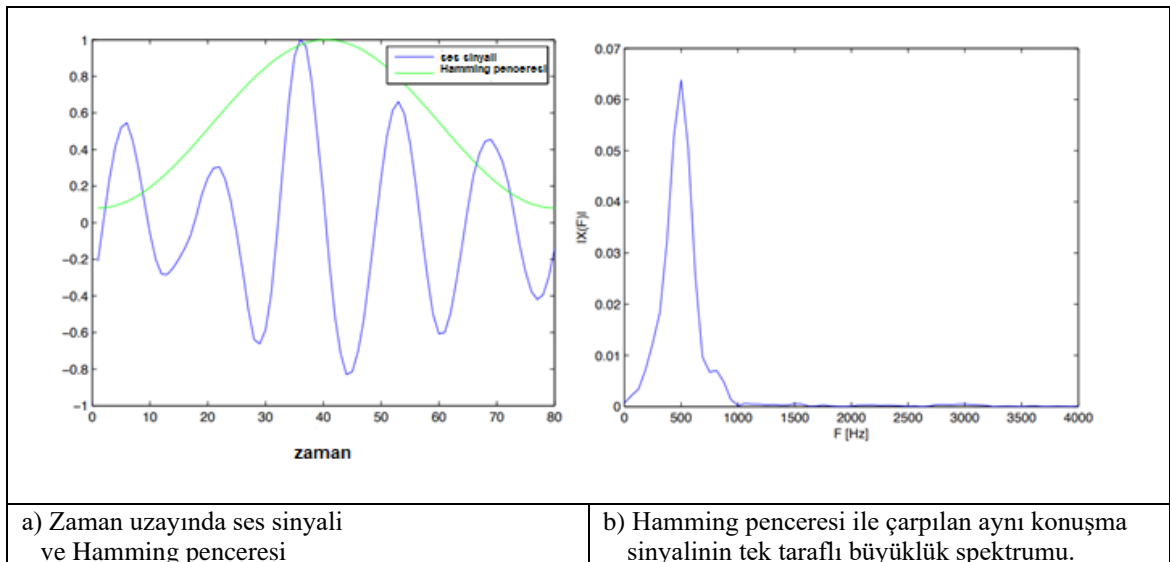
Tek taraflı büyüklük spektrumundan  $D$  en büyük yerel maksimumlar, her çerçeve için özellikler olarak seçilir ve  $D$ , daha sonra incelenecek olan sistemin önemli bir parametresidir.



a) Zaman uzayında ses sinyali ve Hamming penceresi

b) Hamming penceresi ile çarpılan aynı konuşma sinyalinin tek taraflı büyüklük spektrumu.

Şekil 3.2: Konuşma sinyalinin sessiz bir kısmına ait 80 örneklilik ses sinyali çerçevesi. “s” gibi sessiz bir ses daha gürültülüdür ve sesli konuşmaya göre daha yüksek frekanslar içerir.



a) Zaman uzayında ses sinyali ve Hamming penceresi

b) Hamming penceresi ile çarpılan aynı konuşma sinyalinin tek taraflı büyüklük spektrumu.

Şekil 3: Bir konuşma sinyalinin sesli bir kısmına ait 80 örneklilik ses sinyali çerçevesi

### 3.4.2. Eğitim

Eğitim, hem denetimli hem de denetimsiz tekniklerin bir kombinasyonudur. Halihazırda sınıflandırılmış konuşma sinyalleriyle kelime başına bir gizli Markov modeli eğitilir. Her modeldeki farklı durumların sayısı önemli bir seçimdir. Buradaki amaç, her durumun kelimedeki bir fonemi temsil etmesi gerektiğidir. Bununla birlikte Gaussianların kümelenmesi denetimsizdir ve Baum-Welch algoritması için kullanılan başlangıç değerlerine bağlı olacaktır.

Bu tez çalışmasında, başlangıç değerleri olarak  $A$  ve  $\pi$  için tamamen rastgele tahminler kullanılmıştır.  $\Sigma_s$  için, eğitim verilerinin çapraz kovaryans matrisi tüm durumlar için kullanılmıştır. Her durum için, rasgele bir eğitim veri noktası  $\mu_s$  olarak seçilmiştir. Her kelime için eğitim örnekleri bir araya getirilmiş ve Baum-Welch algoritması, 15 iterasyon olarak çalıştırılmıştır.

### 3.4.3. Sınıflandırma

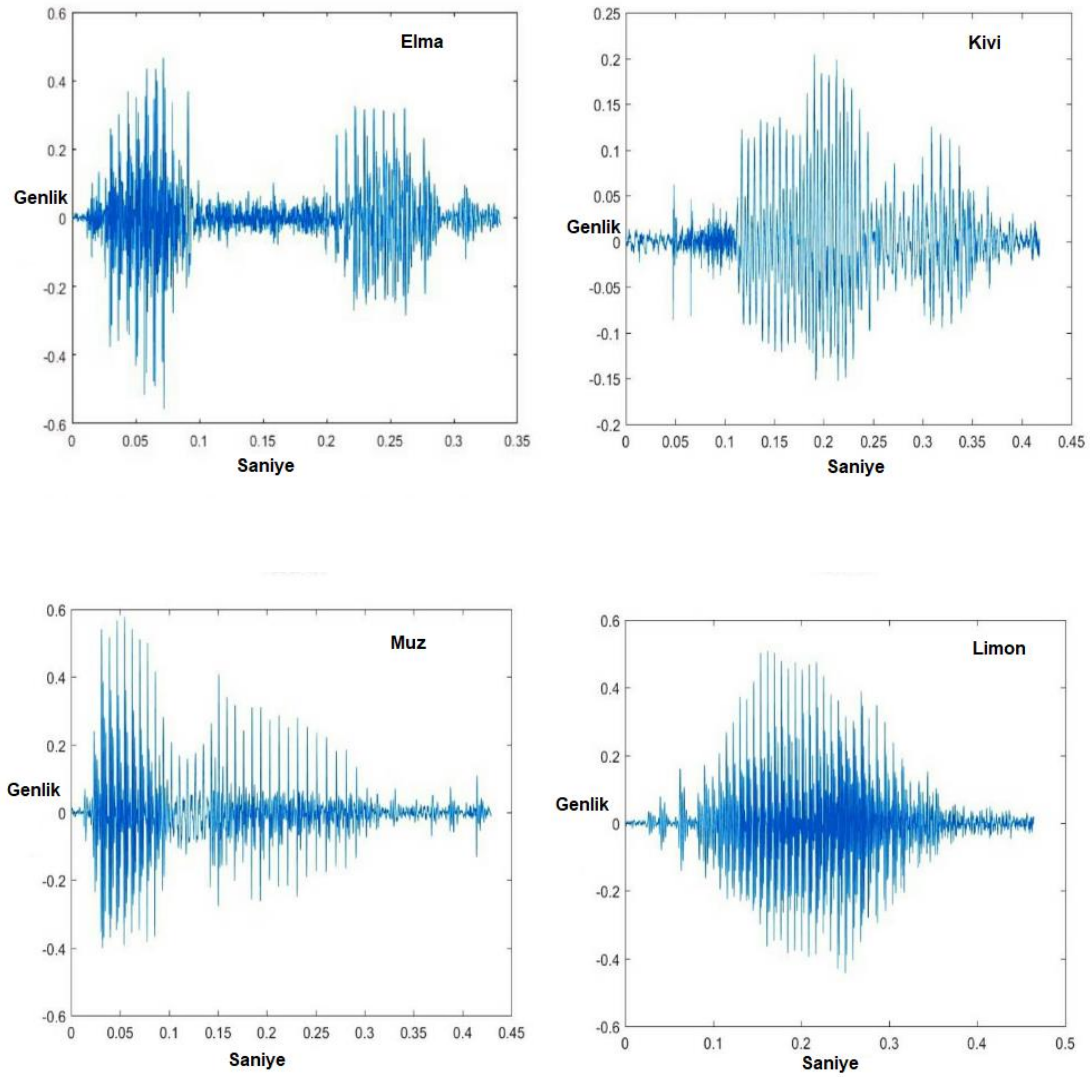
$\lambda_i$  ifadesinin,  $i$  kelimesi için parametre kümesini gösterdiğini varsayalım. Yapılan bir gözlem  $o_1, \dots, o_T$  ile sunulduğunda, seçim aşağıdaki Denklem (4.1) şeklinde elde edilir.

$$\text{Tahmin edilen kelime} = \arg \max_i f(o_1, \dots, o_T; \lambda_i) \quad (4.1)$$

Bununla birlikte  $f(o_1, \dots, o_T; \lambda_i)$  ifadesinin tam olarak ileriye dönük algoritmanın hesapladığı ifade olduğu bilinmelidir.

#### 4. DENEYLER VE TARTIŞMALAR

Bu çalışmada, ses veritabanı oluşturmak amacıyla kelimelerin her biri için, seslendirene ait 15 telaffuz biçimi farklı zamanlarda kaydedilmiştir. Sistemin performansı, 105 ifadeden oluşan kaydedilmiş veri seti üzerinde beş kat çapraz doğrulama ile ölçülmüştür. Şekil 4.1, seslendirilen bazı kelimelerin zamana bağlı genlik bilgilerini göstermektedir.



Şekil 4.1. Bazı örnek ses verilerinin zaman-genlik bilgileri

Ham ses verileri, mikrofon kullanılarak kaydedilmiştir ve kütüphaneler kullanılarak 8000 Hz'lik bir frekansa dönüştürülmüştür.

$$F_s = 8000;$$

$$\text{Çerçeve boyutu} = 80;$$

$$\text{Örtüşme} = 20; D = 6$$

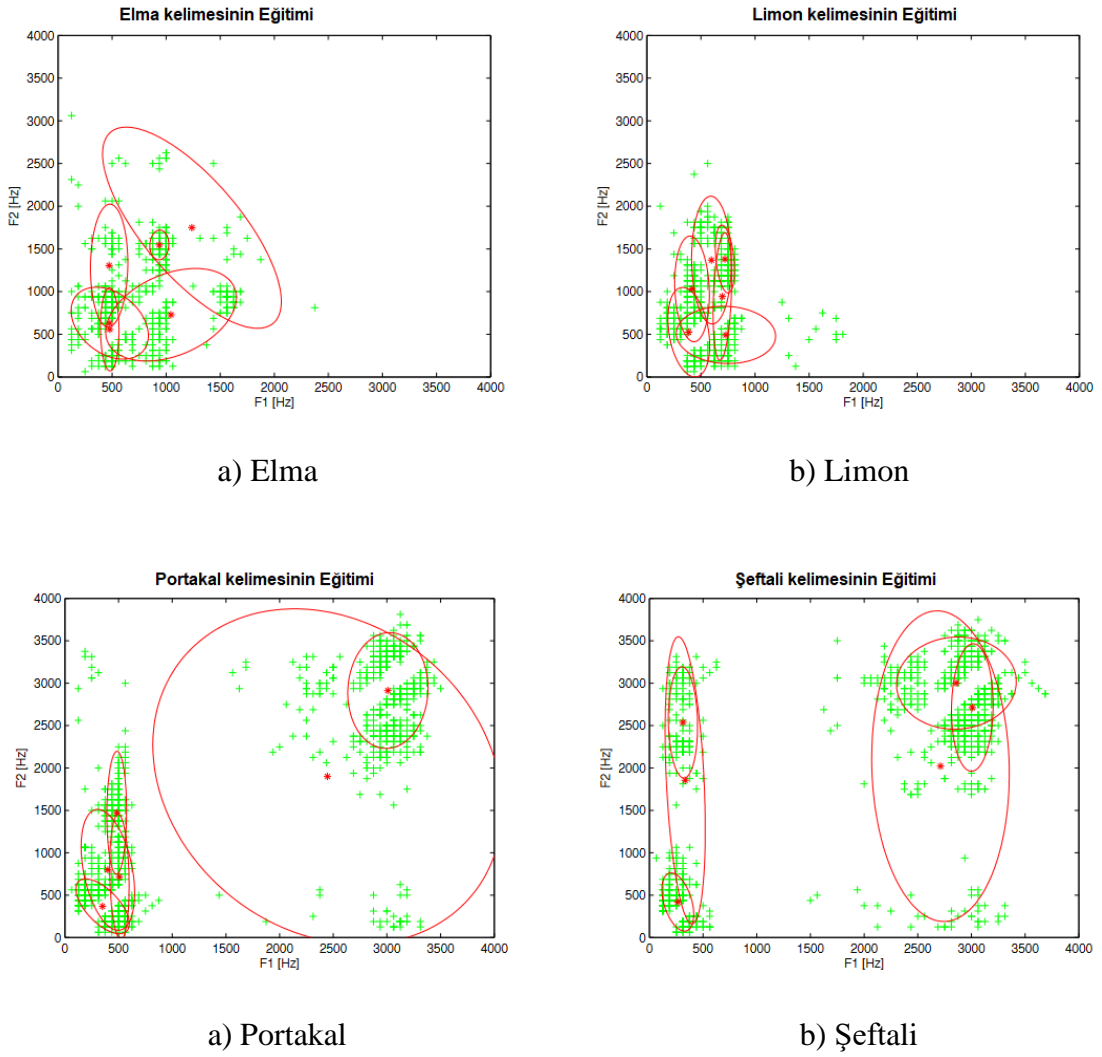
$$y = \text{dalga\_kaydı}(\text{Süre} * F_s, F_s);$$

Yedi kelimeye ait seslerin her biri, her kelime 15 kez olmak üzere kaydedilmiştir. Ardından, özellik çıkarımı için seslerin her biri için frekans zamanı elde edilmiştir. Hem denetimli hem de denetimsiz öğrenmenin elde edilen ses veri kümesi için karşılıklarını belirten etiket tabanlı öğrenme uygulanmıştır. Her konuşma sinyali için her durumun o sesin fenomeninden başka bir şey olmayan bir durum eğitilmiştir. Gauss kümelemesi denetimsizdir ve Baum-Welch algoritmasının başlangıç değerlerine dayanır. İstatistiksel özellikleri kesinlikle takip eden A matrisi ve  $\pi$ 'nin başlangıç değerleri rastgele oluşturulmuştur.  $\Sigma_s$ , gerçekleştirilen tüm yinelemelerde tanımlandığı gibi köşegen kovaryans matrisidir. Başlangıç için Baum-Welch algoritması için 15 yinelemenin etkili olduğu tespit edilmiştir.

Her bir  $i$  kelimesinin kümesi için parametre,  $\theta_1, \dots, \theta_T$ 'den gözlem durumuna sahip olan  $\lambda_i$  olarak belirtilir. Buradan kelime bölümünün  $\arg \max f(\theta_1, \dots, \theta_T; \lambda_i)$  fonksiyonu, ileri algoritma tarafından verilir.

Şekil 4.2'de farklı kelimelere ait uygulanan Baum-Welch algoritmasının 10 yinelemesinden sonra yerleştirilmiş Gauss dağılım fonksiyonları gösterilmiştir. Her birinde bir Gauss fonksiyonu bulunan toplam altı durum bulunmaktadır. Şekilde en baskın iki frekans (özellik) gösterilmiştir. Her bir yeşil artı, bir eğitim konuşma sinyaline ait bir çerçeveyi temsil etmektedir. Yıldızlar, her Gauss fonksiyonunun ortalamasıdır ve elipsler ise %75 güven aralığını gösterir. “şeftali” ve “turuncu” gibi sessiz fonemler içeren sözcüklerde, “elma” ve “limon” gibi içermeyen sözcüklere kıyasla daha yüksek frekanslara sahip olduğu görülmektedir.

Yapılan deneyler, en önemli iki parametrenin saklı durumların sayısı (N) ve her çerçeveden çıkarılan frekans sayısı (D) olduğunu göstermiştir. Bu nedenle, çapraz doğrulama bu parametreler için farklı değerlerle yürütülmüş olup, elde edilen sonuçlar Tablo 4.1'de gösterilmiştir.



Şekil 4.2. Baum-Welch algoritmasının 10 yinelemesinden sonra yerleştirilmiş Gauss Dağılım İşlevleri.

Tablo 4.1: Saklı durumların sayısı için farklı değerlerle beş kat çapraz doğrulama için yanlış sınıflandırma oranları ( $N$ ) ve her çerçeveden çıkarılan frekans sayısı ( $D$ ). Her beş kat çapraz doğrulama prosedürü, 2.9 GHz Intel Core i5 işlemcisinde, 105 ifade için yaklaşık 7 dakika sürer.

$N \setminus D$	2	3	4	5	6	7	8
2			21.9%		8.6%		
3	21.0%	15.2%	9.5%	12.4%	1.9%	14.3%	5.7%
4	16.2%	11.4%	8.6%	5.7%	3.8%	6.7%	4.8%
5	13.3%	8.6%	9.5%	4.8%	2.9%	5.7%	4.8%
6	12.4%	10.5%	3.8%	5.7%	7.6%	6.7%	10.5%
7	15.2%	12.4%	6.7%	10.5%	7.6%	2.9%	8.6%
8			12.4%			5.7%	

## 5. SONUÇLAR VE ÖNERİLER

### 5.1 Sonuçlar

Sonuçlar, özellikle öznitelik çıkarma aşamasında alınan basit yaklaşıma kıyasla oldukça iyidir. Mel frekansı sepstral katsayıları gibi daha gelişmiş özellikler dikkate alınmıştır, ancak elde edilen düşük yanlış sınıflandırma oranları nedeniyle basit frekanslara karar verilmiştir.

Baum-Welch algoritması için eğitim tekrarlamalarının sayısı artırılmıştır ve adımlar arasındaki olasılık farkına bir eşik belirlenmiştir. Bununla birlikte, pratikteki faydasının çok düşük seviyelerde olduğu görülmüş, yürütme süresi ve yanlış sınıflandırma oranının yalnızca yineleme sayısını 15'e sabitlemeye göre kayda değer bir gelişme göstermediği tespit edilmiştir. Yürütme süresinin önemli bir gelişme göstermemesinin nedeni, yürütme süresinin önemli bir kısmının eğitimden ziyade özellik çıkarma sırasında kullanılmış olmasından kaynaklanmaktadır.

Öte yandan,  $N$  çok küçük olduğunda, zamansal bilgi kaybı nedeniyle “ananas” kelimesi olarak yanlış sınıflandırılan birçok “elma” kelimesi elde edilmiştir ve bunun tersi de geçerlidir. Bir diğer önemli parametre, her çerçevedeki örnek sayısıdır. Çerçeve çok küçükse, anlamlı özellikleri seçmek zorlaşmaktadır. Çok büyük seçildiğinde, bu defada zamansal bilgiler kaybolmaktadır. Ancak, zaman kısıtlamaları nedeniyle, bu tez çalışmasında 80 örnek test edilmiştir.

Eğitim örneklerinin birleştirilmesi, sınıflandırma için gerekli olmayan “son durum” dan “başlangıç durumuna” geçiş olasılığını eğitir. Rabiner, çoklu eğitim örnekleri için değiştirilmiş bir Baum-Welch algoritması verir, burada birleştirme gerekli değildir, ancak bu tez çalışmasında birleştirme ile daha iyi sonuçlara ulaşıldığı görülmüştür.

## 5.2 Öneriler

Bu sistemin farklı seslendirmeler ile eğitilip test edilmesi durumunda iyi performans gösterip göstermeyeceği araştırılmalıdır. Bu araştırmanın nedeni ise, özellikle farklı cinsiyetten konuşanlar için farklı seslerin farklı frekans özelliklerine sahip olmasıdır.

Bu çalışmada, izole konuşma tanıma sistemi uygulanmış ve test edilmiştir. Çapraz doğrulama, tek bir konuşmacı için iyi sonuçlar vermiştir. İki belirgin uzantı, birkaç konuşmacı için daha iyi desteklenir ve sürekli konuşma için destek sağlar. Birincisine doğru ilk adım, daha fazla ve daha sağlam özellikler olacaktır. İkincisi için en basit yaklaşım, muhtemelen kelime sınırlarını tespit etmek ve ardından izole edilmiş bir kelime tanıyıcı ile devam etmektir.

## KAYNAKLAR

- Adiga, N. ve Prasanna, S., 2019, Acoustic features modelling for statistical parametric speech synthesis: a review, *IETE Technical Review*, 36 (2), 130-149.
- Afify, M. ve Siohan, O., 2004, Sequential estimation with optimal forgetting for robust speech recognition, *IEEE Transactions on speech and audio processing*, 12 (1), 19-26.
- Agarwal, G., Maheshkar, V., Maheshkar, S. ve Gupta, S., 2018, Recognition of emotions of speech and mood of music: a review, *International Conference on Wireless Intelligent and Distributed Environment for Communication*, 181-197.
- Ahmed, H., Traore, I. ve Saad, S., 2017, Detection of online fake news using n-gram analysis and machine learning techniques, *International conference on intelligent, secure, and dependable systems in distributed and cloud environments*, 127-138.
- Akın, Ü., Yüksel, Ö., Taşcı, E. ve Tuğluoğlu, N., 2019, Fabrication of a New Hybrid Coronene/n-Si Structure by Using Spin Coating Technique and its Photoresponse and Admittance Spectroscopy Studies, *Silicon*, 1-7.
- Alfina, I., Mulia, R., Fanany, M. I. ve Ekanata, Y., 2017, Hate speech detection in the Indonesian language: A dataset and preliminary study, *2017 International Conference on Advanced Computer Science and Information Systems (ICACISIS)*, 233-238.
- Anggraini, N. ve Tursina, M. J., 2019, Sentiment analysis of school zoning system on Youtube social media using the K-nearest neighbor with levenshtein distance algorithm, *2019 7th International Conference on Cyber and IT Service Management (CITSM)*, 1-4.
- Arora, S. J. ve Singh, R. P., 2012, Automatic speech recognition: a review, *International Journal of Computer Applications*, 60 (9).
- Ashraf, J., Iqbal, N., Khattak, N. S. ve Zaidi, A. M., 2010, Speaker independent Urdu speech recognition using HMM, *2010 The 7th International Conference on Informatics and Systems (INFOS)*, 1-5.
- Aydın, İ., 2008, Otomobillerde Kullanılan İzolasyon Malzemelerinin Emme Katsayılarının Ölçülmesi ve Uygun Malzeme Kalınlığının Belirlenmesi, *Yüksek Lisans Tezi, Gazi Üniversitesi, Ankara*.
- Bishop, C. M., 2006, Pattern recognition and machine learning, springer, p.
- Biswas, A., 2016, Performance Enhancement of Automatic Speech Recognition (ASR) Using Robust Wavelet-Based Feature Extraction Techniques.
- Chan, A. J. L., Hatulan, R. J. F., Hilario Jr, A. ve Lim, J. K. T., 2007, Speech to text converter for Filipino language using hybrid artificial neural network/Hidden Markov Model.
- Davis, K. H., Biddulph, R. ve Balashek, S., 1952, Automatic recognition of spoken digits, *The Journal of the Acoustical Society of America*, 24 (6), 637-642.
- Dedman, M., 2009, The Origins & Development of the European Union 1945-2008: A History of European Integration, Routledge, p.
- Delgado, R. ve Stefancic, J., 2014, Hate speech in cyberspace, *Wake Forest L. Rev.*, 49, 319.
- Deng, L. ve O'Shaughnessy, D., 2018, Speech processing: a dynamic and optimization-oriented approach, CRC Press, p.

- Elharati, H. A., Alshaari, M. ve Kępuska, V. Z., 2020, Arabic speech recognition system based on MFCC and HMMs, *Journal of Computer and Communications*, 8 (3), 28-34.
- Erell, A. ve Weintraub, M., 1993, Filterbank-energy estimation using mixture and Markov models for recognition of noisy speech, *IEEE Transactions on speech and audio processing*, 1 (1), 68-76.
- Furui, S., 2018, Digital Speech Processing, Synthesis, and Recognition: Synthesis, and Recognition.
- Ghai, W. ve Singh, N., 2012, Literature review on automatic speech recognition, *International Journal of Computer Applications*, 41 (8).
- Gowsalya, K. ve Sridevi, P., 2019, 'Prediction of fruits and flowers using image analysis techniques', *International Research Journal of Engineering and Technology*, 6 (2).
- Hasan, I. A., Ahmed, M. A. ve Dawood, H. H., 2019, Security Technology for Realistic Measurements with a Distinction on the Speech Recognition in Arabic and English Language Using LMS, Spectral subtraction and A/D Conversions Techniques, *IOP Conference Series: Materials Science and Engineering*, 042023.
- Huang, X., Acero, A., Hon, H.-W. ve Reddy, R., 2001, Spoken language processing: A guide to theory, algorithm, and system development, Prentice hall PTR, p.
- Hunt, J., Ewing, T. N., La-Forge, R. ve Gilbert, W. M., 1959, An integrated approach to research on therapeutic counseling with samples of results, *Journal of Counseling Psychology*, 6 (1), 46.
- Irtza, S. ve Hussain, S., 2016, An Efficient Algorithm To Collect Minimal Speech Corpora, *Pakistan Journal of Engineering and Applied Sciences*.
- Juang, B.-H. ve Furui, S., 2000, Automatic recognition and understanding of spoken language-a first step toward natural human-machine communication, *Proceedings of the IEEE*, 88 (8), 1142-1165.
- Kaur, H. ve Kumar, M., 2018, A comprehensive survey on word recognition for non-Indic and Indic scripts, *Pattern Analysis and Applications*, 21 (4), 897-929.
- Kim, D.-S., Lee, S.-Y. ve Kil, R. M., 1999, Auditory processing of speech signals for robust speech recognition in real-world noisy environments, *IEEE Transactions on speech and audio processing*, 7 (1), 55-69.
- Kurian, C., 2014, A review on technological development of automatic speech recognition, *International Journal of Soft Computing and Engineering*, 4 (4), 80-86.
- Lei, T., Jia, X., Zhang, Y., He, L., Meng, H. ve Nandi, A. K., 2018, Significantly fast and robust fuzzy c-means clustering algorithm based on morphological reconstruction and membership filtering, *IEEE Transactions on Fuzzy Systems*, 26 (5), 3027-3041.
- Nagata, M. ve Morimoto, T., 1994, First steps towards statistical modeling of dialogue to predict the speech act type of the next utterance, *Speech communication*, 15 (3-4), 193-203.
- Nakatani, T., 2019, Improving transformer-based end-to-end speech recognition with connectionist temporal classification and language model integration, *Proc. Interspeech 2019*.
- Ostendorf, M., Digalakis, V. V. ve Kimball, O. A., 1996, From HMM's to segment models: A unified view of stochastic modeling for speech recognition, *IEEE Transactions on speech and audio processing*, 4 (5), 360-378.

- Padmanabhan, J. ve Johnson Premkumar, M. J., 2015, Machine learning in automatic speech recognition: A survey, *IETE Technical Review*, 32 (4), 240-251.
- Pawar, R. V., Jalnekar, R. M. ve Chitode, J. S., 2018, Review of various stages in speaker recognition system, performance measures and recognition toolkits, *Analog Integrated Circuits and Signal Processing*, 94 (2), 247-257.
- Poşul, A., 2011, Telefon bankacılığında interaktif sesli yanıt sistemleri ve beceri bazlı yönlendirme algoritması, 82.
- Rabiner, L. R., 1989, A tutorial on hidden Markov models and selected applications in speech recognition, *Proceedings of the IEEE*, 77 (2), 257-286.
- Ranjan, R., 2021, Speaker Recognition and Performance Comparison based on Machine Learning, *Turkish Journal of Computer and Mathematics Education (TURCOMAT)*, 12 (14), 2297-2306.
- Ross, B., Rist, M., Carbonell, G., Cabrera, B., Kurowsky, N. ve Wojatzki, M., 2017, Measuring the reliability of hate speech annotations: The case of the european refugee crisis, *arXiv preprint arXiv:1701.08118*.
- Sakai, T. ve Doshita, S., 1962, An Automatic Recognition System of Speech Sounds, *音声科学研究= Studia phonologica*, 2, 83-95.
- Sangwan, P., Sheoran, D. ve Bhardwaj, S., 2017, Speech Recognition using Wavelet based Feature Extraction Techniques, *Global Journal of Enterprise Information System*, 9 (2), 23-27.
- Sen, S., Dutta, A. ve Dey, N., 2019, Audio processing and speech recognition: concepts, techniques and research overviews, Springer, p.
- Shao, Y., Taylor, S., Marshall, N., Morioka, C. ve Zeng-Treitler, Q., 2018, Clinical text classification with word embedding features vs. bag-of-words features, *2018 IEEE International Conference on Big Data (Big Data)*, 2874-2878.
- Vanajakshi, P. ve Mathivanan, M., 2017, A detailed survey on large vocabulary continuous speech recognition techniques, *2017 International Conference on Computer Communication and Informatics (ICCCI)*, 1-7.
- Walsh, S. D. ve Horenczyk, G., 2001, Gendered patterns of experience in social and cultural transition: The case of English-speaking immigrants in Israel, *Sex Roles*, 45 (7), 501-528.
- Zhang, Z., Cummins, N. ve Schuller, B., 2017, Advanced data exploitation in speech analysis: An overview, *IEEE Signal Processing Magazine*, 34 (4), 107-129.