T.C.

YILDIZ TEKNİK ÜNİVERSİTESİ

FEN BİLİMLERİ ENSTİTÜSÜ

DERİN ÖĞRENME KULLANARAK KONUŞMA BÖLÜTLERİNİN TESPİTİ İÇİN OPTİMAL ÖZELLİK PARAMETRE KÜMESİ BELİRLEME

Özlem BATUR DİNLER

DOKTORA TEZİ

Bilgisayar Mühendisliği Anabilim Dalı

Bilgisayar Mühendisliği Programı

Danışman

Prof. Dr. Nizamettin AYDIN

Temmuz, 2020

T.C.

YILDIZ TEKNİK ÜNİVERSİTESİ

FEN BİLİMLERİ ENSTİTÜSÜ

DERİN ÖĞRENME KULLANARAK KONUŞMA BÖLÜTLERİNİN TESPİTİ İÇİN OPTİMAL ÖZELLİK PARAMETRE KÜMESİ BELİRLEME

Özlem BATUR DİNLER tarafından hazırlanan tez çalışması 14.07.2020 tarihinde aşağıdaki jüri tarafından Yıldız Teknik Üniversitesi Fen Bilimleri Enstitüsü Bilgisayar Mühendisliği Anabilim Dalı, Bilgisayar Mühendisliği Programı **DOKTORA TEZİ** olarak kabul edilmiştir.

Prof. Dr. Nizamettin AYDIN
Yıldız Teknik Üniversitesi
Danışman

Jüri Üyeleri

Prof. Dr. Nizamettin AYDIN, Danışman
Yıldız Teknik Üniversitesi _____

Doç. Dr. M. Fatih AMASYALI, Üye
Yıldız Teknik Üniversitesi _____

Dr. Öğr. Üyesi Tolga ENSARİ, Üye
İstanbul Üniversitesi-Cerrahpaşa _____

Prof. Dr. Banu DİRİ, Üye
Yıldız Teknik Üniversitesi _____

Prof. Dr. Fikret S. GÜRGEN, Üye
Boğaziçi Üniversitesi _____

Danışmanım Prof. Dr. Nizamettin AYDIN sorumluluğunda tarafımca hazırlanan Derin Öğrenme Kullanarak Konuşma Bölütlerinin Tespiti İçin Optimal Özellik Parametre Kümesi Belirleme başlıklı çalışmada veri toplama ve veri kullanımında gerekli yasal izinleri aldığımı, diğer kaynaklardan aldığım bilgileri ana metin ve referanslarda eksiksiz gösterdiğimi, araştırma verilerine ve sonuçlarına ilişkin çarpıtma ve/veya sahtecilik yapmadığımı, çalışmam süresince bilimsel araştırma ve etik ilkelerine uygun davrandığımı beyan ederim. Beyanımın aksinin ispatı halinde her türlü yasal sonucu kabul ederim.

Özlem BATUR DİNLER

TEŞEKKÜR

Doktora eğitimim süresince, değerli fikirlerini, yardımlarını ve desteğini hiçbir zaman esirgemeyen kıymetli hocam Prof. Dr. Nizamettin AYDIN'a sonsuz teşekkür ederim.

Çalışmama katkılarından dolayı Türkiye Radyo Televizyon Kurumu'na teşekkürlerimi sunarım.

Hayatım boyunca iyi veya kötü her koşulda elimi bırakmayan ve her türlü fedakârlıktan kaçınmayan sevgili annem Emine BATUR, sevgili babam Halil BATUR ve sevgili kardeşlerim Dr. Öğr. Üyesi Canan BATUR ŞAHİN ve Dr. Miray BATUR'a şükran ve sevgilerimle…

Özlem BATUR DİNLER

İÇİNDEKİLER

SİMGE LİSTESİ

C_t	Hücre durumu
f_t	Unutma kapısı katmanı
f	Frekans skalasından bir değişken
h_t	t.zaman çıktı vektörü
i_t	Giriş kapısı katmanı
k	Komşu Sayısı
m	Mel frekans skalasından bir değişken
o_t	Çıkış katmanı
$P(A\|B)$	Koşullu Olasılık
r_t	Sıfırlama kapısı katmanı
$R_n(\tau)$	Otokorelasyon Fonksiyonu
$w(n)$	Pencereleme Fonksiyonu
$X(w)$	Fourier Dönüşümü /Ayrık Zamanlı Fourier Dönüşümü Sonucu
$x(t)$	Zaman boyutunda işaret
x_t	t.zaman giriş vektörü
$x[n]$	Zaman boyutunda ayrık işaret
$X[k]$	Ayrık Fourier Dönüşüm sonucu
z_t	Güncelleme kapısı katmanı

KISALTMA LİSTESİ

ANN	Yapay Sinir Ağları (Artificial Neural Network)
CNN	Evrişimsel Sinir Ağları (Convolutional Neural Network)
DCT	Ayrık Kosinüs Dönüşümü (Discrete Cosine Transform)
DFT	Ayrık Fourier Dönüşümü (Discrete Fourier Transform)
DL	Derin Öğrenme (Deep Learning)
DTFT	Ayrık Zamanlı Fourier Dönüşümü (Discrete Time Fourier Transform)
FFT	Hızlı Fourier Dönüşümü (Fast Fourier Transform)
GRU	Geçitli Tekrarlayan Birim (Gated Recurrent Unit)
Hz	Hertz
kHz	kiloHertz
k-NN	k -En Yakın Komşu
LPC	Doğrusal Öngörülü Kodlama (Linear Predictive Coding)
MFCC	Mel Frekans Kepstral Katsayıları (Mel Frequency Ceptsral Coefficients)
MLP	Multilayer Perceptron (Çok Katmanlı Algılayıcılar)
ms	Milisaniye
NB	Naive Bayes
PLP	Algısal Doğrusal Öngörü (Perceptual Linear Predictive)
RF	Rastgele Orman (Random Forest)
RNN	Tekrarlayan Sinir Ağları (Recurrent Neural Network)
sn	Saniye
SM	Sınıflandırıcı Metotları
SVM	Destek Vektör Makineleri (Support Vector Machines)
TTS	Metinden Konuşma Sentezleme (Text to Speech)
ZCR	Sıfır Geçiş Sayısı (Zero Crossing Rate)

ŞEKİL LİSTESİ

TABLO LİSTESİ

ÖZET

Derin Öğrenme Kullanarak Konuşma Bölütlerinin Tespiti İçin Optimal Özellik Parametre Kümesi Belirleme

Özlem BATUR DİNLER

Bilgisayar Mühendisliği Anabilim Dalı

Doktora Tezi

Danışman: Prof. Dr. Nizamettin AYDIN

Konuşma, birçok kişisel bilgi içeren bir biyometrik işarettir. İnsan iletişiminin en doğal ve en verimli biçimini temsil etmektedir. Gelişen teknoloji ile birlikte, bu konuşma işaretinden elde edilen bilgiler kullanılarak konuşma tanıma, konuşmacı tanıma, konuşma sentezleme ve konuşma kodlama ve çözme gibi çok çeşitli ses işleme uygulamaları geliştirilmektedir. Günümüzde özellikle güvenlik gerektiren kişisel işlemlerde bu uygulamalar aktif bir rol oynamaktadır. Bu uygulamaların geliştirilmesinde çoğu zaman konuşma bölütlerinin tespit sistemi bir ön işlem olarak kullanıldığından konuşma bölütlerinin doğru tespit edilmesi oldukça önemlidir. Konuşma bölütleme (segmentasyon), bir konuşma işaretini daha küçük akustik birimlere bölme işlemi olarak adlandırılır. Aynı zamanda, konuşma işaretini kelimeler, heceler veya fonemler arasında sınır bulma prosedürü olarak tanımlamak ta mümkündür. Bu tez çalışmasında, sürekli bir konuşma içerisindeki Ünsüz (Consonant), Ünlü (Vowel), ve Sessiz (konuşmanın olmadığı, Silent) (C/V/S) bölgeleri Geçitli Tekrarlayan Birim (Gated Recurrent Unit, GRU) tekrarlayan sinir ağlarına dayalı tahmin edebilen (belirleyebilen) fonem tabanlı bir konuşma tespit sistemi geliştirilmiştir. Bu amaçla, C/V/S konuşma bölütlerinin sınırlarını tanımlamak için 4 farklı pencere uzunluğu, 3 farklı pencereleme yöntemi ve 3 farklı hibrit özellik çıkarım

yöntemi birlikte kullanılarak 6 farklı sınıflandırıcı yöntemi ile test edilmiştir. Böylece çeşitli parametrelerin farklı hibrit özellik çıkarım yöntemleri ile birlikte kullanılmasının C/V/S konuşma bölütlerinin tespit sistemi üzerindeki etkisi incelenmiştir.

Bu çalışmada, Enerji, Sıfır Geçiş Sayısı (ZCR) ve Mel Frekans Kepstral Katsayı (MFCC) temelli bir hibrit özellik çıkarım yöntemi kullanılmıştır. Bu bağlamda, farklı hibrit özellik çıkarım yöntemleri çeşitli parametreler ile birlikte kullanılarak bir ses işaretinin içerisindeki C/V/S konuşma içeren bölütlerin tespitini en iyi modelleyen parametre setinin belirlenmesi amaçlanmıştır. Yapılan uygulamalar sonucunda GRU modelinin, Kürtçe akustik işaretini karakterize etme başarımını arttırdığı gözlenmiştir. Ayrıca, günümüzde, Kürtçe alanında çok az sayıda akademik çalışma yapıldığından dolayı, bu çalışma bu alanda önemli bir katkı yapacaktır.

Anahtar Kelimeler: Veri kümesi, derin öğrenme, ünsüz/ünlü/sessiz, bölütleme, konuşma bölütlerin tespiti.

ABSTRACT

Determining Optimal Feature Parameter Set For Detection Of Speech Segments Using Deep Learning

Özlem BATUR DİNLER

Department of Computer Engineering

Doctor of Philosophy Thesis

Advisor: Prof. Dr. Nizamettin AYDIN

Speech is a biometric sign containing a lot of personal information. It represents the most natural and efficient form of human communication. Along with the developing technology, a wide range of sound processing applications such as speech recognition, speaker recognition, speech synthesis, and speech coding and decoding have been developed using the data obtained from this speech signal. Nowadays, these applications play an active role, especially in personal processes that require security. Since the detection system of speech segments is usually used as a pre-treatment in the development of these applications, it is very important to determine speech segments accurately. The procedure of dividing a speech signal into smaller acoustic units is called speech segmentation. It is also possible to define speech signal segmentation as the procedure of finding boundaries between words, syllables, or phonemes. A phoneme based speech detection system that can predict (detect) Consonant, Vowel, and Silent (no speech) (C/V/S) regions in a continuous speech based on Gated Recurrent Unit (GRU) recurrent neural networks was developed in this thesis study. For this purpose, 4 different window lengths, 3 different windowing methods and 3 different hybrid feature extraction methods were tested together with 6 different classifier methods in order to define the boundaries of C/V/S speech segments. Thus, the effect of the use of various

parameters with different hybrid feature extraction methods on the recognition system of C/V/S speech segments was examined.

A hybrid feature extraction method based on Energy, Zero-Crossing Rate (ZCR), and Mel Frequency Cepstral Coefficient (MFCC) was used in this study. In this proposed method, it was aimed to determine the parameter set that best models the detection of segments containing C/V/S speech within a sound signal by using different hybrid feature extraction methods together with various parameters. As a result of the applications, it was observed that the GRU model increased the performance of characterizing the Kurdish acoustic signal. Furthermore, since there are very few academic studies in the field of Kurdish nowadays, this study will provide a significant contribution to this field.

Keywords: Dataset, deep learning, consonant/vowel/silent, segmentation, speech segment detection.

1
Giriş

1.1 Literatür Özeti

Son yıllarda, çeşitli ses işeleme uygulamalarındaki çoğu çalışmalar geleneksel makine öğrenmesi yöntemlerinin aksine, Tekrarlayan Sinir Ağları (RNN) yöntemlerini ve bu yöntemlerin en yaygın iki türü olan Uzun Kısa Süreli Hafıza (Long Short Term Memory, LSTM) ve GRU modellerini kullanarak geliştirilen yaklaşımlar üzerine ilerleme göstermiştir. Bu yaklaşımların temel motivasyonu, giriş verilerini birbirlerinden bağımsız olarak değil bir zaman serisi içerisinde ağa girişi sağlamasıdır. Böylece ardışıl şekilde gelen girdiler birbirlerine bağlı hesaplamalar ile geçmiş bilgileri dikkate alarak çıktılar üretirler.

Bu alanda yapılacak çalışmalardan bahsedilirse; Graves ve Jaitly [1], RNN yöntemi ile ses işaretlerinin fonetik gösterimini kulanmadan ses işaretlerini doğrudan metine çeviren bir konuşma tanıma sistemi geliştirmişlerdir.

Shewalker vd. [2], çalışmalarında konuşma tanıma sistemi için LSTM ve GRU yöntemlerini Technology Entertainment Design - Laboratoire Informatique de l'Uiversité du Maine (TED - LIUM) konuşma veri kümesine uygulayarak elde edilen performansları karşılaştırmışlardır.

Diğer bir çalışmada Ravanelli vd. [3], konuşma tanıma problemi için GRU' ların basitleştirilmiş bir mimariye sahip doğrultulmuş doğrusal birim aktivasyonları ile hiperbolik tanjantın yerini alan kompakt bir tek kapılı Basitleştirilmiş-GRU (Li-GRU) modeli önermişlerdir. Elde edilen sonuçlar, önerilen modelin eğitim süresini GRU'ya kıyasla %30'dan daha fazla hesaplama karmaşıklığını azalttığını göstermiştir. Mevcut araştırma, Texas Instruments Massachusetts Teknoloji Enstitüsü (TIMIT), DIRHA-English, CHiME-4 ve TED-talk konuşma veri kümelerinden yararlanmıştır.

Bir başka çalışmada, Yuan vd. [4], görsel ve işitsel verilerden konuşma tanıma sisteminin tasarlanması için Yardımcı Kayıp Çok Modlu-GRU (Auxiliary Loss

Multimodal - GRU, ALM-GRU) mimari yapısını önermişlerdir. Önerilen mimaride, AVLetters, AVLetters2, ve AVDigits veri kümeleri kullanılmıştır.

Marolt vd. [5], bir ses dosyasındaki müzik verileri ve ham konuşma işareti verilerini ayırmaya odaklanmışlardır [10]. Kısa ses bölümlerini etiketlemeye yönelik evrişimli derin ağ mimarisi kullanılmıştır. %10'dan daha yüksek bir sınıflandırma doğruluğu iyileştirmesi elde edilmiştir.

Wang vd. [6] ise, GRU içerisindeki kapı aktivasyon işaretlerini analiz etmişlerdir. Aynı zamanda işaretlerin zamansal yapısının fonem bölütlerinin sınırları ile yüksek korelasyona sahip olduğu gözlenmiştir. Geleneksel yaklaşımlara göre elde edilen korelasyonun daha iyi sonuçlar verdiği doğrulanmıştır. Bu çalışmada TIMIT veri kümesi kullanılmıştır.

Chen vd. [7], bir ses dosyasındaki müzikten insan sesinin tespitini GRU yaklaşımı ile gerçekleştirmişlerdir.

Zheng vd. [8], konuşma işaretlerinden duygu tanıma sistemini GRU ve CNN yöntemlerini birlikte ele alarak geliştirmişlerdir.

Graves ve Schmidhuber [9], LSTM ile çerçeve tabanlı fonem sınıflandırma uygulamasını TIMIT veri kümesi üzerinde gerçekleştirmişlerdir.

Bir diğer çalışmada Franke vd. [10], derin çift yönlü LSTM'leri kullanarak ses kayıtlarındaki fonem sınırlarının otomatik tespitini incelemişlerdir. Ses kayıtları İngilizce ve Bantu dili konuşmalarını içeren TIMIT ve BUCKEYE veri setleri üzerinde gerçekleştirilmiştir.

Günümüzde, daha fazla araştırmacı CNN ve RNN modellerini etkili bir şekilde birleştirmişlerdir [11]. Bu modeller performans doğruluğunu artırmak için farklı ses işleme teknolojilerine uygulanmıştır [12-16]. Genel olarak, CNN ve RNN yöntemlerinin birlikte kullanılması yüksek başarıların elde edilmesini sağlamıştır. Bu nedenle bu çalışmada RNN ve CNN modeli birlikte kullanılmıştır.

1.2 Tezin Amacı

Türkiye; geçmişten bugüne farklı etnik grupları (Ermeniler, Rumlar, Boşnaklar, Arnavutlar, Araplar, Kürtler vs.) bir arada barındırdığından, çeşitli dil ailesine mensup farklı dil ve lehçelerin konuşulmasına ev sahipliği yapmıştır. Kürtçe, günlük hayatta Türkçe'den sonra en çok konuşulan dillerden biridir [17]. Bu sebeple, bu çalışmada Türkiye'de yaşayan dil gruplarından Kürtçe dilinin ses özelliği ele alınmıştır. Bu çalışma kapsamının hedefi Kürtçe dilinde fonem tabanlı bir otomatik konuşma tespit sisteminin geliştirilmesidir. Bu hedef doğrultusunda aşağıdaki maddeler amaçlanmıştır.

1. Kürtçenin fonetik özelliklerinin incelenmesi.
2. Kürtçe dilinde fonem tabanlı özgün bir ses veri kümesinin oluşturulması
3. Veri kümesi üzerinde kullanılacak hibrit özellik çıkarım yöntemlerinin performans sonucuna katkısının saptanması.
4. Hibrit özellik çıkarım aşamasında, farklı pencereleme yöntemlerine ve pencere uzunluklarına bağlı parametrelerle üretilecek hibrit özelliklerin performans sonucuna etkisinin belirlenmesi.
5. Eğitim aşamasında Derin Öğrenme (Deep Learning, DL) yöntemlerinden biri olan GRU tekrarlayan sinir ağlarının kullanılmasının başarıma etkisinin incelenmesi.
6. Yapay sinir ağları modeli ve geleneksel sınıflandırıcı yöntemleri ile sınıflandırma başarımının test edilmesi.
7. En iyi performansı verecek konuşma tespit sisteminin parametre setinin (hibrit özellik çıkarım yöntemi, pencereleme tekniği, pencere boyutu ve sınıflandırıcı yönteminin) ortaya konulması.
8. Gelecekte, ses işleme alanında Kürtçe üzerine çalışmalarda bulunacak araştırmacılara katkı sağlanması.

1.3 Hipotez

Konuşma bölütlerinin tespit sistemi, konuşma tanıma, konuşmacı tanıma, dil tanıma, konuşma sentezleme, konuşma kodlama ve çözümleme gibi çeşitli konuşma işleme sistemlerinin temelini oluşturmaktadır [18, 19].

Bir konuşma bölütlerinin tespit sisteminin doğruluğu büyük ölçüde konuşma dilini oluşturan fonem setine ve bu fonemlerin fonetik özelliklerine bağlıdır. Konuşulan lehçe, konuşmacının yaşı ve cinsiyeti ve kelimelerin telaffuzundaki değişiklikler sistemin başarısını etkileyen diğer önemli unsurları oluşturmaktadır. Buna ek olarak, tonlama etkisi veya hece stresi (vurgusu) gibi faktörler nedeniyle konuşma bölütlerinin tespit süreci daha zor hale gelmektedir. Bu bağlamda, bu zorluklarla başa çıkmak için DL yöntemlerinden RNN modelinin özel bir türü olan GRU algoritması kullanılmıştır. Bu algoritmaya alternatif çözüm olarak, GRU tekrarlayan sinir ağlarına göre daha genelleştirilmiş bir tekrarlayan sinir ağı algoritması olan LSTM kullanılabilir. Ancak GRU, standart LSTM algoritmalarına göre daha basit bir yapıya sahip olduğundan karmaşıklık ve işletim maliyeti daha düşüktür ve kullanım yaygınlığı da giderek artmaktadır [20]. Bu nedenlerden dolayı GRU, bu çalışma kapsamında tercih edilmiştir. LSTM tekrarlayan sinir ağları GRU'nun aksine, bilgi akışı kontrolünü uzun süreli ve kısa süreli bellek birimleri üzerinden sağlayabilmektedir. Kontrol mekanizması olarak GRU'da yer alan sıfırlama ve güncelleme kapılarının aksine LSTM'de giriş, çıkış ve unutma kapıları bulunmaktadır.

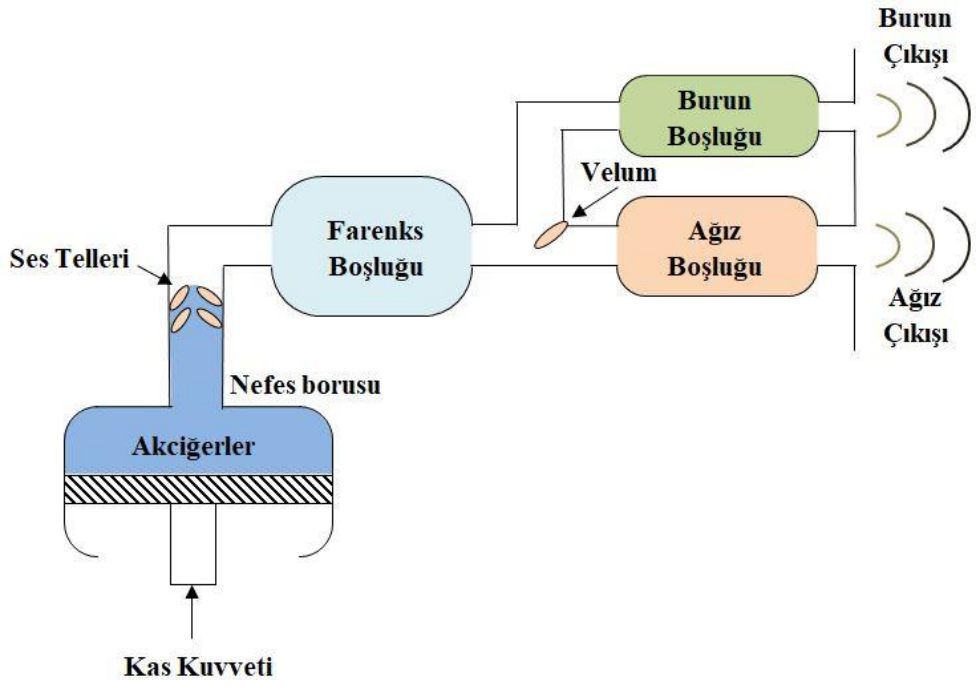
Bu tez çalışması kapsamında, Türkiye'de konuşulan Kürtçe dilinin akustik özelliklerine dayalı bir otomatik C/V/S konuşma bölütlerinin tespit sistemi geliştirilmiştir. Bu amaç doğrultusunda özgün bir Kürtçe veri kümesi oluşturulmuştur. Oluşturulan veri kümesi % 66 eğitim seti ve % 33 test seti olmak üzere ikiye ayrılmıştır. Veri kümesindeki konuşma işaretleri Enerji + ZCR + MFCC, Enerji + ZCR + D-MFCC (Delta MFCC) ve Enerji + ZCR + DD-MFCC (Delta Delta MFCC) ayırt edici özellik parametrelerinin birleşiminden oluşan 3 farklı hibrit özellik çıkarım yöntemi ile elde edilmiştir. Elde edilen her bir hibrit özellik çıkarım yönteminde 20 ms, 25 ms, 30 ms ve 35 ms pencere boyutları ve Hamming, Hanning ve Rectangular pencereleme yöntemleri birer

parametre olarak kullanılarak, bu parametrelere bağlı farklı özellikler ortaya çıkarılmıştır. Daha sonra, bu özelliklerden elde edilen hibrit özellik vektörleri eğitim seti tabanında GRU tekarlayan sinir ağları ile eğitilmiş ve test seti tabanında CNN, Çok Katmanlı Algılayıcılar (Multilayer Perceptron, MLP), Naive Bayes (NB), Rastgele Orman (Random Forest, RF) ve k-En Yakın Komşu (k-Nearest Neighbour, k-NN) metotları ile test edilmiştir. Böylece, bu çalışmanın sonunda farklı parametrelerin değişimlerinden üretilen hibrit özellik vektörlerinin ve farklı sınıflandırma metotlarının Kürtçe C/V/S konuşma tespit sistemi üzerindeki performans sonucuna katkıları incelenmiştir. Bu çalışma, GRU tabanlı Kürtçe C/V/S konuşma bölütlerinin tespiti için en uygun özellik parametre setini belirlemeye yönelik yapılmış kapsamlı bir çalışmadır. Saptanan en uygun parametre özellikleri, Türkiye Radyo Televizyon Kurumu (TRT)'den alınmış haber içerikli Kürtçe konuşma işaretleri ile analiz edilmiştir. Bu çalışmanın GRU tabanlı konuşma bölütlerinin tespiti kapsamındaki diğer çalışmalardan temel farklılıkları, Kürtçe alanında çok az sayıda akademik çalışmanın yapılmış olması, özgün bir veri kümesi örneklemesinin oluşturulması ve Kürtçe konuşma tespit sistemi için en uygun özellik parametrelerinin tespit edilmesidir. Kısacası, bu çalışma ile Kürtçe'nin ses özelliklerine dayalı bir konuşma tespit sistemi geliştirilmiştir.

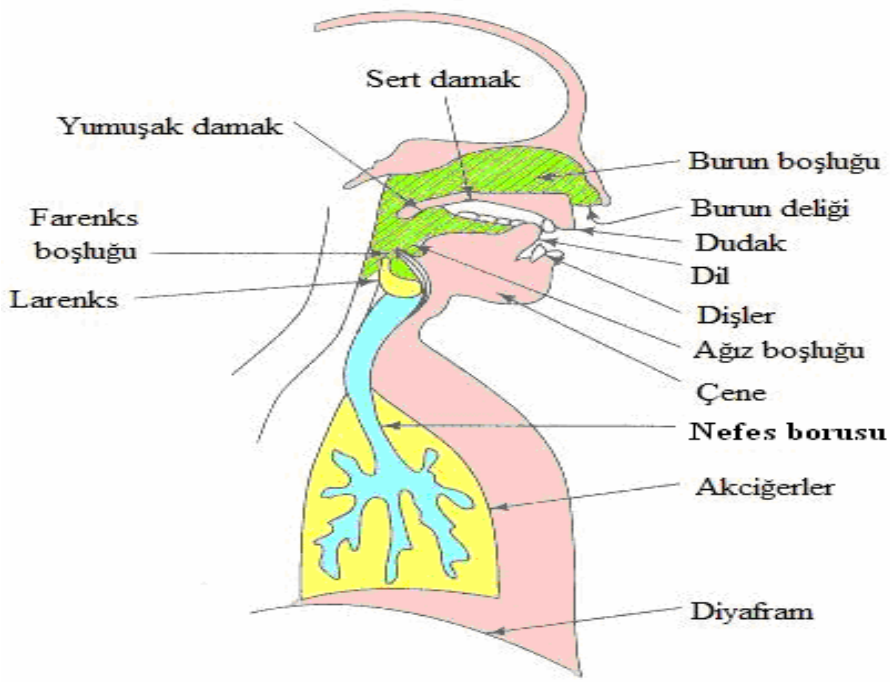
2
İnsanda Konuşma Oluşum Süreci

2.1 Ses Üretim Mekanizması ve Konuşmanın Oluşumu

Fiziksel olarak ses, moleküllerin titreşim nedeniyle hava yoluyla iletilmesi sonucu oluşan boyuna ya da enine bir basınç dalgasıdır. İnsanlarda bu basınç dalgasının ses olarak oluşumu ise, akciğerlerden gelen hava basıncının soluk borusunda bulunan ses tellerinde titreşmesi ve bu titreşimin gırtlak, boğaz, ağız boşluğu ve burun boşluğundan geçmesi ile gerçekleşir [21]. Konuşma oluşum sürecinde, ses üretim sistemi neredeyse sonsuz sayıda ses üretme yeteneğine sahiptir. İnsanda ses üretim mekanizmasının şematik gösterimi Şekil 2.1'de ve ses üretim mekanizmasında etkin rol alan tüm organlar Şekil 2.2'de ayrıntılı olarak gösterilmiştir.



Şekil 2.1 İnsanda Ses Üretim Mekanizması[22].



Şekil 2.2 İnsanda Ses Üretim Mekanizmasında Rol Alan Organlar[22].

Kişilerdeki ses telleri, damak yapısı, diş yapısı, dil ve dudak yapısı gibi önem arz eden anatomik yapıların şekil, boyut ve almış oldukları pozisyon farklılıkları, ses verilerinin karakteristik veri olma özelliğini sağlamıştır. Böylece aynı ataya sahip kardeşlerde dahi, sesler benzerlik gösterse bile kardeşlerdeki ses örüntüleri tamamen birbirinin aynı değildir.

Ses dalgası en çok önem teşkil eden dalga örnekleridir. Ses dalgaları ancak iletici ortamlarda yayılabilirler. Bu sebeple, ses dalgalarının yayılabilme hızı bulundukları ortama göre değişir. Bu bağlamda basınç, sıcaklık ve yoğunluk gibi faktörler ses dalgalarının yayılmasına etkide bulunurlar.

2.2 Ses Üretim Mekanizmasında Rol Oynayan Temel Kavramlar

Ses dalgası fiziksel bir niceliktir. Bir ses dalgasının oluşumunda etkin rol alan en temel kavramlar aşağıdaki gibidir:

2.2.1 Genlik

Bir ses dalgasının, erişebileceği maksimum değer olarak ifade edilir.

2.2.2 Dalga Boyu

Bir ses dalgasının ardışıl iki tepe veya iki çukur noktası arasındaki mesafe olarak tanımlanır. Dalga boyu "λ (lambda)" ile gösterilir.

2.2.3 Periyot

Kendini yenileyerek sürüp giden hareketin bir defa tamamlanması için geçen saniyedir [23]. Periyot " T " ile gösterilir.

2.2.4 Frekans

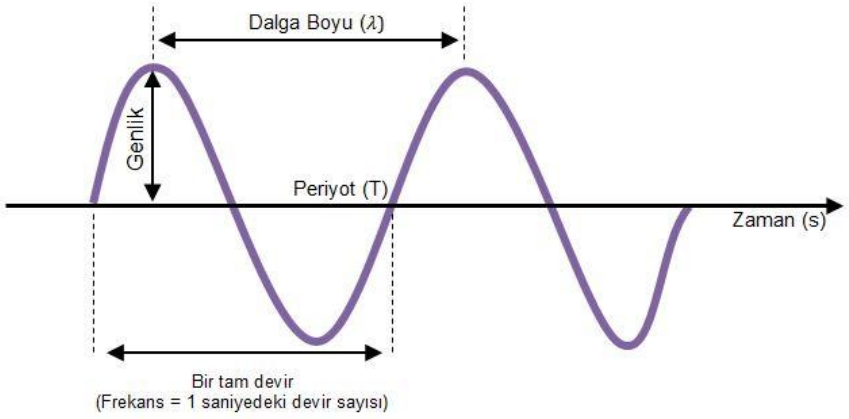
Bir ses dalgasının bir saniyedeki titreşim sayısı olarak tanımlanır. Bir ses dalgasını oluşturan frekans değeri Denklem 2.1 ile hesaplanmaktadır:

$$f = \frac{1}{T} \tag{2.1}$$

Burada, f frekans değerini ve T periyot sayısını belirtmektedir. Frekans "Hertz (Hz)" ile ölçülmektedir.

Bir insanın algılayabileceği (işitebileceği) frekans değerleri 20 Hz ile 20000 Hz değerleri arasındadır. Bir insanın günlük hayattaki konuşmalarının frekans değerleri ise, 500 Hz ile 2000 Hz değerleri arasındadır. Yüksek frekanslı sesler ince (tiz) ses, düşük frekanslı sesler ise kalın (bas) ses olarak tanımlanır. Kadınların frekans değerleri erkeklerin frekans değerlerinden daha yüksektir. Bu sepeble, kadınlar ince sesli ve erkeklerde kalın seslidir.

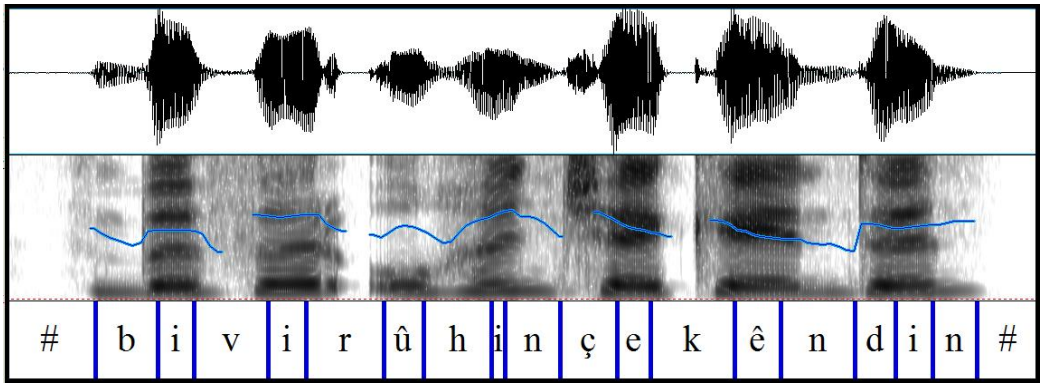
Şekil 2.3'te bir ses dalgasının genliği, dalga boyu, periyotu ve frekansı gösterilmektedir.



Şekil 2.3 Genlik, Dalga Boyu, Periyot ve Frekans Kavramlarının Gösterimi.

2.2.5 Perde

Perde, ses telleri tarafından üretilen titreşimlerin saniye başına düşen sayısına bağlıdır. Frekansların yüksek ve düşük seviyede olması sesin tiz veya pes olarak algılanmasına neden olur. Bu algılama hissi "perde" ya da "temel frekans (F_0)" olarak adlandırılır. Şekil 2.4' te bir konuşma işaretinin perdesi gösterilmiştir. Ortalama temel frekans, tipik olarak erkekler için 85-185 Hz, kadınlar için 165-200 Hz aralığındadır. Kadınlar erkeklerden daha yüksek temel frekanslarla konuşurlar.

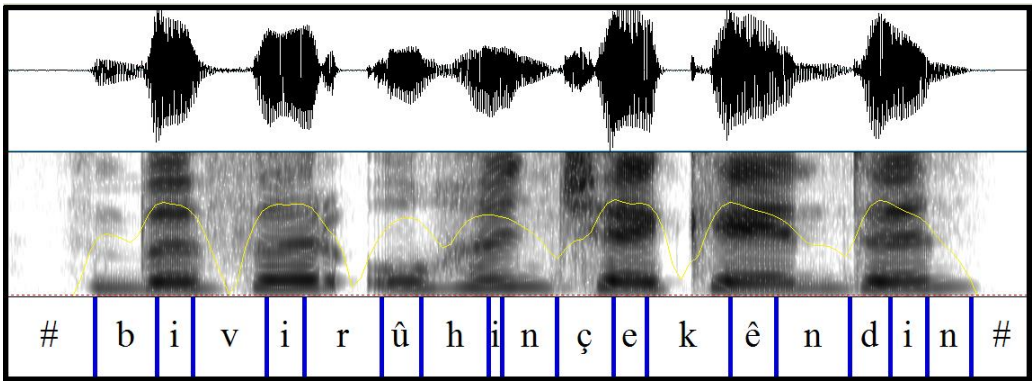


Şekil 2.4 Ses İşaretinde Perde Kavramının Gösterimi.

2.2.6 Şiddet

Sesin genlik değeri ile ilişkili bir kavramdır. Sesin genlik değerinin büyüklüğü ile orantılı olarak değişir. Sesin genlik değeri arttıkça sesin enerjisi ve şiddeti de artmaktadır. Sesin enerjisi ve şiddeti genlik değerinin büyüklüğü ile orantılı olarak değişir. Sesin şiddeti "Desibel (db)" ile ölçülür. Şekil 2.5' te bir konuşma işaretinin şiddeti gösterilmiştir. Ses şiddeti logaritmik olarak büyür. Bu bağlamda, işitme duyusu

doğrusal büyüklükteki değerleri algılamayıp, logaritmik büyüklükteki değerleri algılamaktadır.



Şekil 2.5 Ses İşaretinde Şiddet Kavramının Gösterimi.

2.2.7 Gürültü

Titreşimlerin periyodik bir dalga oluşturamadığı sese denir. Bu sesler sinir bozucu ve rahatsızlık uyandırır [24].

3
Kürtçe'nin Fonetik Özellikleri

3.1 Kürtçe'nin Dil Yapısı

Hint-Avrupa dil ailesi grubundan olan Kürtçe, yapı bakımından çekimli bir dildir. Bu bağlamda, Kürtçe; Fransızca, İngilizce, Rusça ve Almanca gibi Avrupa dilleri ile önemli ölçüde benzerlikler göstermektedir. Fakat Kürtçe Avrupa dilleri arasından en çok Farsçaya benzemektedir. Lakin her iki dilin kelime hazineleri, morfoloji yapıları, fonetik özellikleri ve gramer yapıları tamamen kendilerine özgüdür [25].

Modern Kürt Edebiyatında Kürt dili lehçesi Kurmanci ve Sorani olmak üzere 2 ana dala ayrılır [26]. Fakat Gorani, Zazaki ve Lurri de Kürt dilinin en önemli lehçeleri arasında yer almaktadır. Türkiye'de yaşayan Kürt toplumunun en yaygın konuştuğu lehçe Kurmanci lehçesidir. Bu nedenle bu çalışmada Kürtçe'nin Kurmanci lehçesi incelenmiştir.

Geçmişte, Kürt toplumunun çeşitli nedenlerden dolayı farklı coğrafyalara dağılması Arapça, Latin ve Kiril alfabesinin kullanılmasına neden olmuştur. Kürt alfabesinde harflerin her biri sadece bir sese karşılık gelir ve hiçbir şekilde başka bir sesi ifade etmezler. Bütün sözcükler, söylendikleri gibi yazılır ve okunurken de bütün yazılı harfler telaffuz edilir [26, 27]. Kürtçe dilinde Türkçede olduğu gibi, her hecede mutlaka bir ünlü ses bulunmalıdır ve en küçük hece birimini ünlü fonemler oluşturmaktadır.

3.2 Kürtçe Fonem Seti

Fonem, dilbiliminde konuşmanın en küçük temel birimi olarak ifade edilmektedir. Bir fonem seti ise, bir dildeki olası tüm kelimeleri tanımlamak için gereken minimum simge sayısı olarak tanımlanır [21]. Fonemler, ünlü sesler ve ünsüz sesler olmak üzere 2 farklı grupta sınıflandırılmaktadır. Ünlülerin üretim teorisi; enerji kaynağı, titreşim ve ses yolunda oluşan rezonans frekanslarına bağlıdır. Akciğerler insan sesinin üretim sisteminin enerji kaynağıdır. Akciğerlerden gelen hava basıncının gırtlakta yer alan ses telleri arasında bir dizi süreklilik ve şiddetle dışarı itilmesi bu tellerin titreşmesine sebep olur. Bu titreşimler ünlülerin oluşumuna kaynaklık etmektedir. Aynı zamanda

ünsüzlerden bazıları da ses tellerinin titreşimi ile üretilirler. Ses tellerinin titreşimi sonucu elde edilen bu sesler ötümlü (voiced) sesler olarak adlandırılır. Bu nedenle, tüm ünlüler ile kimi ünsüz sesler ötümlü ses özelliğindedir. Havanın sürtünme ve patlamaya dayalı yarattığı gürültü kaynaklı titreşimler ile elde edilen kimi ünsüzlerde ötümsüz (unvoiced, voiceless) sesler olarak adlandırılır. Ötümlü seslerin sinyali periyodik bir yapı sergilerken, ötümsüz seslerin sinyali periyodik olmayan bir yapı sergilemektedirler. Bu bağlamda ötümlü sesler, genlik-zaman gösteriminde birbirini tekrar eden bir örüntüye sahip iken, ötümsüz seslerin herhangi bir örüntüsü yoktur ve rastgeledir. Bu durum da, ünlü ile ünsüz fonemlerin ayırt ediciliğini güçleştirmektedir.

Latin harflerine dayalı Kürt alfabesi /b/, /c/, /ç/, /d/, /f/, /g/, /h/, /j/, /k/, /l/, /m/, /n/, /p/, /q/, /r/, /s/, /ş/, /t/, /v/, /w/, /x/, /y/ ve /z/ 23 ünsüz fonem ile /a/, /e/, /ê/, /i/, /î/, /o/, /u/ ve /û/ 8 ünlü fonem olmak üzere toplam 31 fonemden oluşmaktadır.

3.2.1 Kürtçe'deki Ünlü Fonemlerin Ses Özellikleri

Ünlü sesler, ses telleri ile elde edilen titreşimlerin ses yolunda salınıma uğratılması sonucu elde edilen fonemlere denir. Ünlü fonemler çıkarılırken, söyleyiş süresine göre; uzun ünlü ve kısa ünlü, dilin almış olduğu pozisyon farklılıklarına göre; ön damak ünlüleri ve arka damak ünlüleri ile dilin damağa olan yakınlığına göre yüksek, orta ve alçak ünlü olmak üzere Tablo 3.1 'de ki gibi 3 farklı grupta sınıflandırılmaktadır [25-27].

Tablo 3.1 Kürtçe Ünlü Fonemlerin Temel Özellikleri [21].

	Uzun Ünlüler			**Kısa Ünlüler**		
	Yüksek	Orta	Alçak	Yüksek	Orta	Alçak
Ön Damak	/î/, /û/	/ê/				/e/
Arka Damak			/a/, /o/	/i/, /u/		

3.2.2 Kürtçe'deki Ünsüz Fonemlerin Ses Özellikleri

Ünsüz sesler, ses yolunun herhangi bir konumunda, akan havaya yaratılan engellerle (tıkamalarla) üretilen fonemlere denir. Bu ünsüz fonemler çıkarılırken, çıkarılış

yerlerine göre 6 dudak ünsüzü, 7 ön damak ve diş ünsüzü, 5 damak ünsüzü ve 5 arka damak ve gırtlak ünsüzü olmak üzere 4 farklı grupta sınıflandırılır [25-27].

Dudak ünsüzleri: /b/, /f/, /m/, /p/, /v/, /w/.

Ön damak ve diş ünsüzleri: /d/, /l/, /n/, /r/, /s/, /t/.

Damak ünsüzleri: /c/, /ç/, /j/, /ş/, /y/ .

Ön damak ve Gırtlak ünsüzleri: /g/, /k/, /h/, /q/, /x.

Tablo 3.2 ‘de Kürtçe dilinde yer alan fonemlerin isimleri, IPA yazımları ve fonetik açıklamaları ayrıntılı olarak verilmiştir.

Tablo 3.2 Kürtçe Fonemler [28].

Kürtçe Fonemler	**IPA**	**Fonetik Açıklama**
a	[ɑ]	Geniş, kalın, düz ünlü
b	[b]	Çift dudaksıl, ötümlü, kapantılı
c	[dʒ]	Damaksıl, ötümlü, yarı kapantılı
ç	[tʃ]	Damaksıl ,ötümsüz, üflemesiz yarı kapantılı
	[tʃ^{h}]	Damaksıl, ötümsüz ,üflemeli yarı kapantılı
d	[d]	Alveo-dental, ötümlü,kapantılı
e	[ae]	Genişe yakın, ince, düz ünlü
	[ɛ]	Orta, geniş, ince, düz ünlü
ê	[e]	Orta, dar, ince, düz ünlü
	[ə]	Orta, merkez, düz ünlü
f	[f]	Labiodental, ötümsüz sürtünmeli
g	[g]	Artdamaksıl ötümlü kapantılı
h	[h]	Gırtlaksıl sürtünmeli
	[ħ]	Boğazsıl sürtünmeli
i	[ɨ]	Kısa, merkez, düz ünlü
î	[i]	Dar, ince, düz ünlü
j	[ʒ]	Post-alveolar ötümlü sürtünmeli
k	[k]	Artdamaksıl ötümsüz üflemesiz kapantı
	[k^{h}]	Artdamaksıl ötümsüz üflemeli kapantı
l	[l]	Alveo-dental yanünsüz
m	[m]	Çiftdudaksıl genizsil kapantı
n	[n]	Alveo-dental genizsil kapantı
o	[o]	Orta dar, kalın, yuvarlak ünlü
p	[p]	Çiftdudaksıl ötümsüz üflemesiz kapantı
	[p^{h}]	Çiftdudaksıl ötümsüz üflemeli kapantı
q	[q]	Uvuler ötümsüz kapantı
r	[r]	Alveolar titrek ünsüz
	[ɾ]	Alveolar çarpmalı
s	[s]	Alveolar ötümsüz sürtünmeli
ş	[ʃ]	Post-alveolar ötümsüz sürtünmeli

Tablo3.2 Kürtçe Fonemler (devamı)

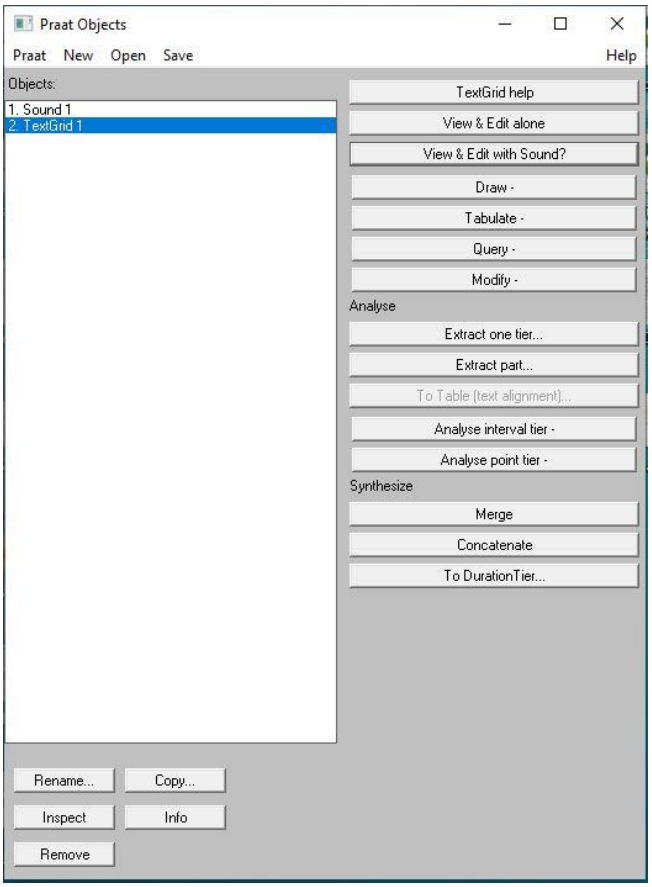
t	[t]	Alveo-dental ötümsüz üflemesiz kapantı
	[t^h]	Alveo-dental ötümsüz üflemeli kapantı
u	[ʊ]	Dara yakın, hafif ortalanmış yuvarlak ünlü
û	[u]	Dar, kalın, yuvarlak ünlü
v	[v]	Labiodental ötümlü sürtünmeli
w	[w]	Çiftdudaksıl aproksimant
x	[x]	Artdamaksıl ötümsüz sürtünmeli
y	[y]	Damaksıl aproksimant
z	[z]	Alveolar ötümlü sürtünmeli

4
Ses Analiz Araçları

Ses inceleme alanında kullanılabilecek sayısız açık kaynak veya ücretli yazılım bulunmaktadır. Praat, Audacity, Wavesurfer, Adobe Audition, Acoustica ve WinPitch ses işleme uygulamalarında en yaygın kullanılan ses analiz araçlarıdır. Bu analiz araçları, basit ama güçlü bir arayüze sahip yazılımlardır.

4.1 Praat

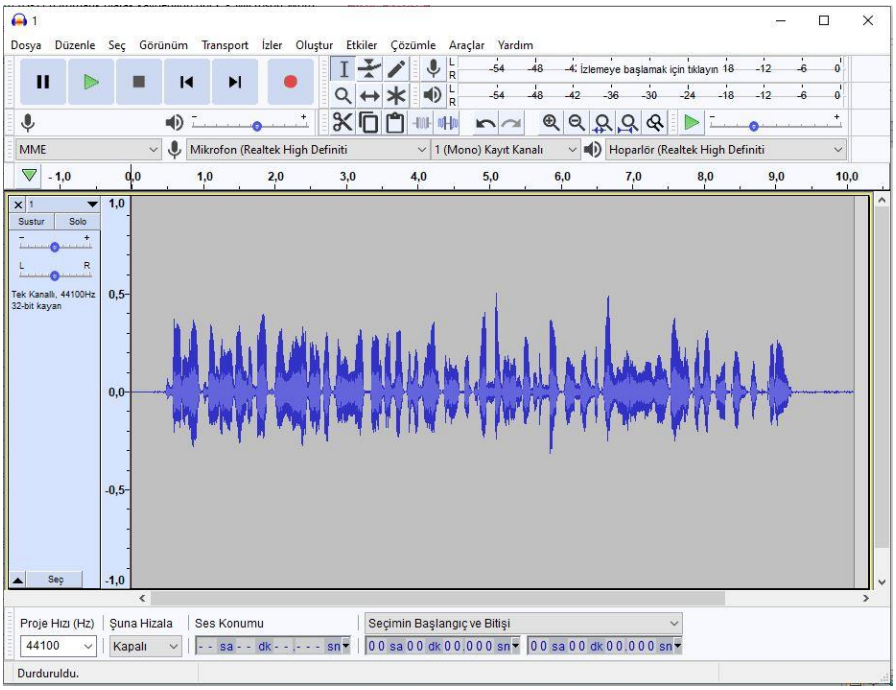
Praat, Amsterdam üniversitesi fonetik bilimler bölümünden Paul Boersma ve David Weenink tarafından tasarlanmış ve sürekli geliştirilen bir yazılım aracıdır. Konuşma işaretlerinin ve ses örneklerinin akustik olarak incelendiği ve değerlendirildiği ücretsiz ve açık kaynak kodlu bir yazılımdır [29]. Praat, özellikle dilbilimcilerin fonetik çalışmalarında kullandıkları bir ses analiz yazılımıdır. PRAAT, kullanıcının eklenti kullanımına izin veren çok esnek ve kullanımı basit ve kolay bir yazılımdır. Praat, Unix, Linux, Mac ve Microsoft Windows (2000, XP, Vista, 7, 8, 10) işletim sistemleri platformlarını desteklemektedir [30]. Şekil 4.1'de Praat yazılım aracının ara yüzü ve işlev menüleri gösterilmiştir. Bu işlev menüleri genel olarak; bir ses dosyasını açmak, oluşturmak, kaydetmek ve ses dosyasına ait işaretin dalga şekli, spektrogram grafiği, formant frekansları, perde ve şiddeti gibi akustik özellikler hakkındaki bilgilerin görsel ve istatistiksel analizlerinin gerçekleştirilmesini sağlamaktadır.



Şekil 4.1 Praat Programının Arayüzü.

4.2 Audacity

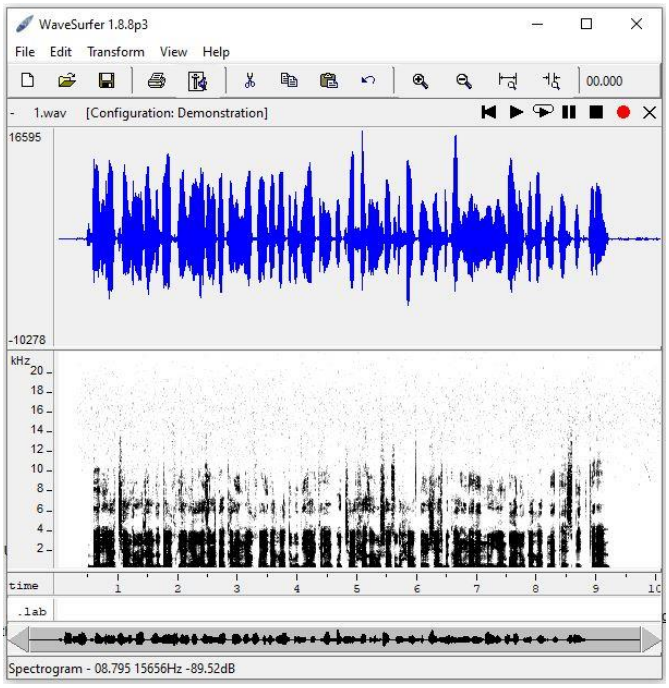
Audacity, Windows, macOS, Linux ve Unix işletim sistemleri için kullanılabilen ücretsiz ve açık kaynaklı bir sayısal ses düzenleyici ve kayıt uygulama yazılımıdır [31]. Audacity, Praat programına kıyasla akustik özellikleri ölçümlendirme veya karakterize edebilme yönüyle daha zayıf, fakat ses kayıt ve düzenleme yönüyle daha güçlü bir yazılımdır. Şekil 4.2'de Audacity yazılım aracının 2.3.3 versiyonunun ara yüzü gösterilmiştir. Audacity yazılımının farklı versiyonları görünüm ve işlevsellik bakımından farklılıklar göstermektedir.



Şekil 4.2 Audacity Programının Arayüzü.

4.3 Wavesurfer

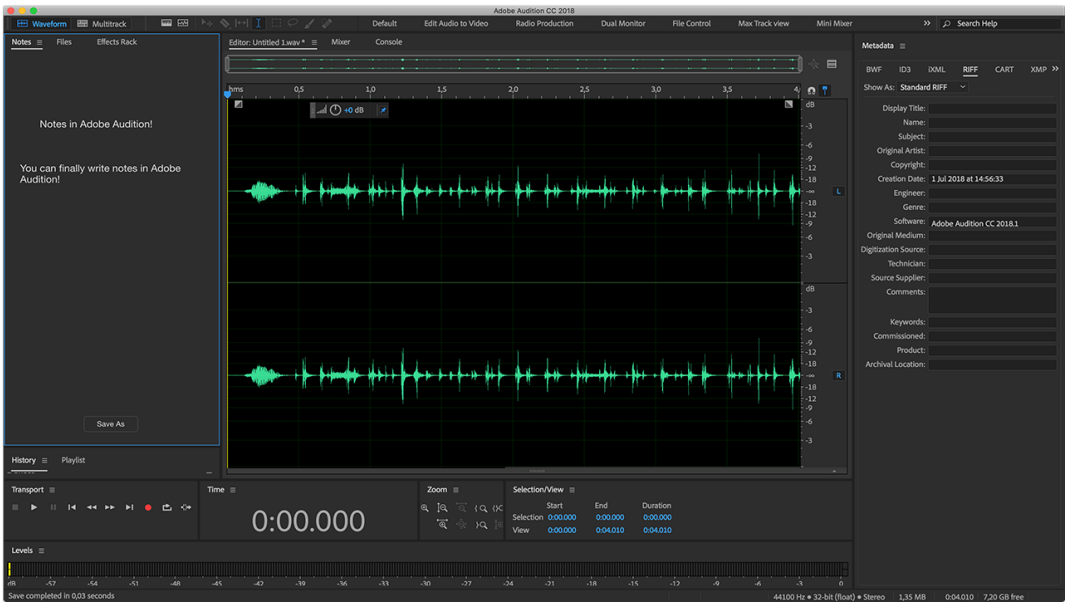
WaveSurfer, ses işaretlerinin görselleştirilmesi için kullanılan açık kaynak bir ses yazılım editörüdür. Sesin analiz ve transkripsiyon işlevi tipik uygulamalarıdır. WaveSurfer, diğer uygulamalara yerleştirilebilir ve eklentilerle de genişletilebilir. Microsoft Windows, Mac OS X, Linux, Solaris, HP-UX, FreeBSD ve IRIX platformlarında çalışmaktadır [32]. Şekil 4.3'te Wavesurfer yazılımının ara yüzü gösterilmiştir.



Şekil 4.3 Wavesurfer Programının Arayüzü.

4.4 Adobe Audition

Adobe Audition yazılımı; ses içerikleri oluşturmak, düzenlemek ve düzeltmek için dalga formu ve spektral görüntüler içeren kapsamlı bir araç setidir [33].Lisanlı kullanımı gerektiren, bu yazılım Windows, Linux ve Mac işletim sistemlerinde çalışmaktadır. Şekil 4.4'te Adobe Audition yazılımının ara yüzü gösterilmiştir.



Şekil 4.4 Adobe Audition Programının Arayüzü.

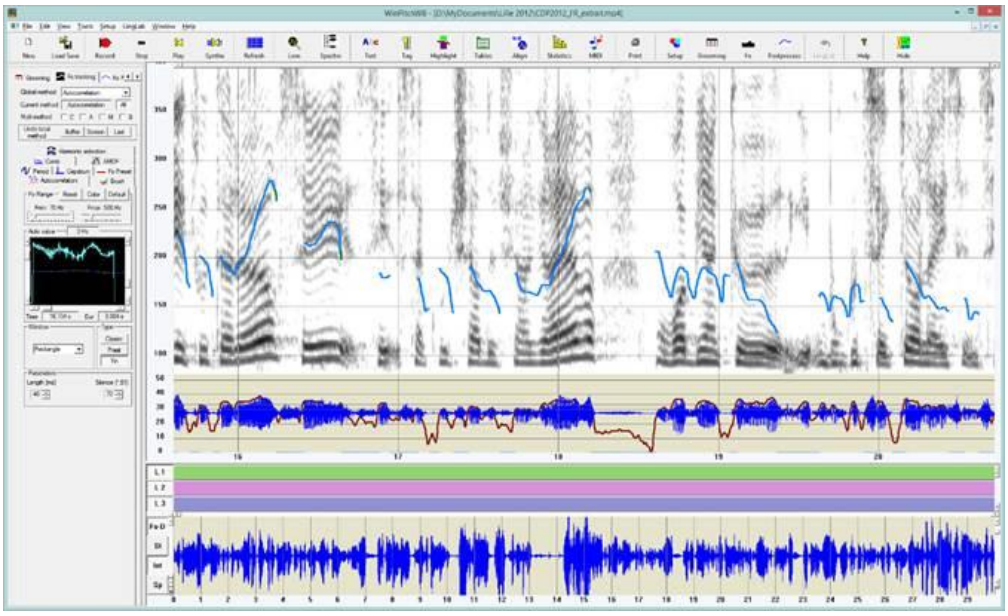
4.5 Acoustica

Acoustica, Mac ve Windows işletim sistemi yüklü sistemlerde kullanılabilen ücretli bir yazılımdır. Bu yazılım ile kayıt restorasyonu, spektogram ve dalgacık analizi gibi işlemler gerçekleştirilebilmektedir [34]. Şekil 4.5'te Acoustica yazılımının ara yüzü gösterilmiştir.

Şekil 4.5 Acoustica Programının Arayüzü.

4.6 WinPitch

WinPitch, prosodik araştırmalar, gerçek zamanlı spektogram analizi, temel frekans ve ses analizi işlevlerini gerçekleştirebilmektedir. Windows ve Mac işletim sistemi yüklü bilgisayarlarda kullanılabilen ücretli bir yazılımdır [34,35]. Şekil 4.6'da WinPitch yazılımının ara yüzü gösterilmiştir.



Şekil 4.6 WinPitch Programının Arayüzü.

Ses işleme alanında kullanılabilecek diğer ses analiz araçlarının performans özelliklerinin karşılaştırılması Şekil 4.7'de gösterilmiştir.

YAZILIM / DEĞERLENDİRME	DC Forensics	STC SIS II	SESTEK	Acu-Exp. Forensic Audio	CEDAR Cambr.	Audacity	PhonEdit	SPS/WASP	Agnitio SIFT	Speech Analyzer	Praat	UCL Enhance	CoolEdit	Acoustica 7	WaveSurfer	Adobe Audition	IKAR Lab	lingWaves	WinPitch 10	TrueRTA	GoldWave	WavePad	Raven	Sound Forge	SoundRuler	SpectraPlus	QE Plug-in	C-T FOENICS
Lisans gereksinimi																												
Ücretli	√	√	√	√	√	-	-	-	√	-	-	-	√	√	-	√	√	√	√	√	√	√	√	√	-	√	√	√
Ücretsiz	-	-	-	-	-	√	√	√	-	√	√	√	-	-	√	-	-	-	-	-	-	-	-	-	√	-	-	-
USB Lisansı (dongle)	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	√	-	-	-	-	-	-	-	-	√	-
Eklenti Desteği	-	-		-	-	√	-	-	-	-	√	-	√	√	√	-	-	-	-	-	√	√	-	√	-	-	√	-
Platform																												
Windows	√	√	√	√	√	√	√	√	√	√	√	√	√	√	√	√	√	√	√	√	√	√	√	√	√	√	√	√
Linux	-	-				√	√	-		-	√	-	-	-	√	√	-	-	-	-	√	-	√	-	√	-	-	-
MAC	-	-				√	-	-		-	√	-	-	√	√	√	-	-	-	-	-	√	√	√	-	-	-	-
Backup	√	√		√		√	-	-		√	-	-	-	√	√	√	√	√	√	√	√	√	√	√	-	√	√	√
Ağ/Bulut'tan dosya aktarımı ve yönetme	-	-			√	-	-	-	√	-	-	-	-	-	-	√	-	-	-	-	-	√	-	-	-	-	-	-
Arayüz kişiselleştirme	√	√	√	√	√	√	√	-	√	-	√	-	√	√	-	√	√	-	-	-	√	√	√	√	-	√	-	-
Önceden kaydedilmiş (preset) yüklenebilir ayarlar	√	-				-	-	-		-	-	-	-	√	-	√	-	-	-	-	√	√	√	√	-	-	√	-
Gelişmiş klavuz ve yardım	√	√	√	√	√	√	-	√	√	√	√	√	√	√	√	√	√	√	√	√	√	√	√	√	√	√	√	√
Raporlama desteği	-	√	√	√	√	-	-	-	√	√	-	-	-	-	-	-	√	-	-	-	-	-	√	-	-	-	√	-

Şekil 4.7 Ses Analiz Araçları ve Özellikleri [36].

5
Konuşma Bölütlerinin Tespiti ve Ses İşleme Uygulamaları

Konuşma bölütleme, bir konuşma işaretini daha küçük akustik birimlere bölme işlemini ifade etmektedir. Bu bağlamda konuşma bölütleme, konuşma işaretindeki ses birimlerinin sınırlarının belirlenmesi işlemidir. Bölütleme işlemi bir konuşma dilinde yer alan kelimelerin, hecelerin veya fonemlerin arasındaki sınırların belirlenmesi için kullanılmaktadır. Bu çalışmada, fonem seviyesinde bir bölütleme işlemi gerçekleştirilmiştir. Bu durumda konuşma işaretlerini oluşturan fonemlerin başlangıç ve bitiş noktaları tahmin edilmiştir [37]. Konuşma bölütleme, konuşma tanıma, konuşmacı tanıma, konuşma sentezleme, konuşma kodlama ve çözümleme uygulamalarında kullanılan bir ön işlem adımıdır. Bu uygulamalar için konuşma bölütlerinin doğru tespit edilmesi büyük bir önem taşımaktadır.

5.1 Konuşma Tanıma

Konuşma tanıma, ses işleme alanında en önemli konular arasında yer almaktadır. Konuşma tanıma sistemi, söylenen kelimeleri ve ifadeleri tanımada ve bunları makineler ve özellikle bilgisayarlar tarafından anlaşılabilecek bir formata dönüştürmede kullanılan en yaygın teknolojilerden biridir. Bu teknolojiler kullanıcının, klavye ve fare gibi araçlara ihtiyaç duymadan doğrudan konuşarak işlem yapabilme olanağını sunar [38]. Günlük hayatta en çok kullanılan uygulamaları engelli yardım uygulamaları, robotik uygulamalar, telefon bankacılığı uygulamaları ve otomatik çağrı cihazı uygulamalarıdır. Özellikle adli suçların tespiti ile güvenlik ve erişim kontrolü gerektiren uygulamalar için de konuşma tanıma sistemleri büyük ölçüde önem kazanmıştır.

Genel bir konuşma tanıma sisteminde, mikrofon ya da telefon aracılığıyla alınmış bir konuşma ifadesinin tanıma işlemini gerçekleştirmek amacıyla ayırtedici özellik (öznitellik) vektörleri çıkarılır ve çıkarılan bu özellik vektörleri eğitilerek bir model oluşturulur. Daha sonra hangi konuşma ifadesi olduğu sorgulanacak test konuşma işaretinin modeli çıkarılır ve daha önce (eğitilmiş) modeller ile karşılaştırılarak eşleşme olup olmadığı tespit edilir [39]. Özetle, bir konuşma tanıma sisteminde girdi olarak ses

işaretinin alınması, sesin işlenmesi ve çıktı olarak ses işaretinin tanınması amaçlanmaktadır. Konuşma tanıma sistemleri, konuşmacıya bağımlı veya konuşmacıdan bağımsız tanıma, sözcük seviyesinde tanıma veya fonem seviyesinde tanıma ve gerçek zamanlı sürekli konuşma tanıma veya ayrık kelime tanıma şeklinde birçok farklı boyut ile ele alınarak değişik şekillerde tasarlanabilmektedir.

5.2 Konuşmacı Tanıma

Konuşmacı tanıma sistemi, kişiye özgü karakteristik bilgilerin veya özelliklerin bulunduğu konuşma örüntülerinden kimin konuştuğunun otomatik olarak saptanmasıdır. Günlük hayatta konuşmacı tanıma sistemleri en çok sesli arama uygulamalarında, telefonla gerçekleştirilen bankacılık ve alışveriş uygulamalarında, veri tabanına erişim uygulamalarında ve adli uygulamalarda kullanılmaktadır. Kısacası, konuşmacı tanıma sistemleri kişisel güvenlik veya kişisel gizlilik gerektiren birçok önemli uygulama alanlarında kullanılmaktadır. Bir konuşmacı tanıma sisteminde gerçekleştirilmesi gereken üç temel işlem vardır. Bu işlemlerin ilki, her bir konuşmacıya ait ses işaretinden ayırt edici özelliklerin (özellik vektörlerinin) çıkarılması işlemi; ikincisi, elde edilen özellik vektörlerinin makine öğrenmesi metotlarıyla eğitilerek modellenmesi ve veri tabanına kayıt edilmesi işlemi; üçüncüsü ise, ”Kim olduğu sorgulanacak kişinin“ modelinin çıkartılması ve veri tabanındaki hangi model ile eşleştiğinin karar verildiği (test edildiği) sınıflandırma işlemidir.

Konuşmacı tanıma sistemi, kullanılan uygulamaya ve tekniğe göre konuşmacı belirleme ve konuşmacı doğrulama olmak üzere iki farklı yol ile gerçekleştirilebilir. Konuşmacı belirlemede, bilinmeyen konuşmacı işaretinin veri tabanındaki konuşmacılardan “hangi kişiye veya kime” ait olduğunun sorgulanması işlemidir. Eğer, sorgulanması istenen kişinin konuşma işareti veri tabanına kayıtlı ise kapalı küme, sorgulanması istenen kişinin konuşma işareti, veri tabanına kayıtlı değil ise açık küme olarak adlandırılır. Konuşmacı doğrulama ise, konuşmacı işaretinin “iddia edilen kişiye” ait olup olmadığının sorgulandığı işlemdir.

Konuşmacı tanıma sistemleri, metine bağımlı (text-dependent) ve metinden bağımsız (text-independent) olarak 2 farklı şekilde gerçekleştirilebilir. Metine bağımlı konuşmacı tanıma sistemlerinde, kişiler önceden belirlenmiş veya tanımlanmış bir metine bağlı kalırken, metinden bağımsız konuşmacı tanıma sistemlerinde ise, kişiler önceden

belirlenmiş veya tanımlanmış herhangi bir metine bağlı değillerdir. Bu bağlamda, metine bağımlı konuşmacı tanıma sistemlerinde, eğitim ve test aşamalarında kullanılan veriler (metinler) tamamen birbirinin aynı iken, metinden bağımsız konuşmacı tanıma sistemlerinde ise, hem eğitim aşamasında hem de test aşamasında kullanılan veriler (metinler) tamamen birbirinden farklıdır.

5.3 Konuşma Sentezleme

Konuşma sentezleme yazılı bir metni konuşma işaretlerine çevriren sistemlerdir. Bu sistemler literatürde metinden konuşma sentezleme (tetx to speech, TTS) olarak adlandırılırlar. Günümüzde görme engelli uygulamalarında, sesli yanıt sistemlerinde, bilgi ve uyarı sistemlerinde kullanılmaktadır. Konuşma sentezleme sistemleri kural tabanlı formant sentezleyiciler, söyleyiş (articulatory) sentezleyiciler ve eklemeli (concatenative) sentezleyiler olmak üzere 3 farklı yaklaşım ile gerçekleştirilebilirler [40, 41]. Birinci yaklaşım sentezleyiciler, konuşma işaretinin doğrusal öngörülü kodlaması (LPC, Linear Predictive Coding) yöntemine dayanmaktadır. Bu yöntemde bir konuşma işaretinin, eski örneklerinin doğrusal birleşimi şeklinde olduğu düşünülüp ses işaretinin karakteristik katsayıları yaklaşık olarak hesaplanır. Elde edilen yaklaşık sonuç ile gerçek değer arasındaki fark (hata) en aza indirilir. İkinci yaklaşım sentezleyiciler, tüm ses birimlerinin (fonemlerin), insan ses üretim mekanizmasında nasıl oluşturulduğunun modellenmesi temeline bağlıdır. Üçüncü yaklaşım sentezleyiciler, bir konuşma işaretini, sesbirim, difon, trifon, seslem vb. gibi önceden kaydedilen ses parçalarını belirli işaret işleme teknikleriyle bir araya getirerek oluşturulur.

Konuşma sentezleme sistemleri metin işleme ve konuşma sentezleme bölümlerinden oluşmaktadır. Metin işleme bölümünde, sisteme girdi olarak verilen bir metin ayrıştırılarak konuşma sentezleme bölümüne hazırlanmaktadır. Ancak sisteme verilen her metni doğru bir şekilde işlemek oldukça karmaşık ve zor bir işlemdir. Konuşma sentezleme bölümünde kullanılan sentezleme yöntemine göre metnin ses işaretine çevrilir.

5.4 Konuşma Kodlama ve Çözümleme

Konuşma kodlama, bir konuşma işaretinin en az kanal kapasitesi, yüksek kalite ve düşük maliyet ile iletiminin gerçekleştirilmesdir. Kısaca konuşma kodlaması, bir konuşma işaretinin geçici olarak daha az bit kullanacak şekilde sıkıştırılabildiği ve

ardından sıkıştırılmış konuşma işaretinin en önemli bilgileri kaybetmeden tekrar açma (çözümleme) işlemidir. Bir konuşma kodlayıcı sisteminin etkin tasarımı, ses kalitesine, bit hızına, iletim maliyetine, kanal kapasitesine ve kodlama ve kod çözmek için gereken işlem süresine (karmaşıklığına) bağlı parametreler ile ilişkilidir.

6
Yöntem ve Araçlar

6.1 Özellik (Öznitelik) Çıkarım Yöntemleri

Ses işleme sistemlerinde, ses işaretlerinin işlenebilmesi için, yapılması gereken en temel işlem, her bir konuşma işaretine ait ses örüntülerinden ayırt edici özelliklerin çıkarılmasıdır. Bir ses işaretini temsil eden özellikler zaman uzayında ve frekans uzayında tanımlanabilmektedir. Günümüzde enerji ve sıfır geçiş sayısı (Zero Crossing Rate, ZCR) zaman uzayında ve Mel Frekansı Kepstrum Katsayıları (Mel Frequency Cepstral Coefficients, MFCC) ve türevleri (Delta MFCC (D-MFCC) ve Delta Delta MFCC (DD-MFCC)) frekans uzayında hesaplanabilen en yaygın özellik çıkarım parametreleridir.

6.1.1 Enerji

Enerji, ses işleme alanında özellik çıkarım yöntemlerinde sıklıkla kullanılan parametrelerden biridir. Fiziksel açıdan enerji, bir konuşma işaretinin herhangi bir zaman diliminde ne kadar işaret bulundurduğunun bir ölçüsünü teşkil etmektedir. Bu bağlamda, bir konuşma işaretinin enerji değeri konuşma işaretini oluşturan genlik değişimleri ile ifade edilmektedir. Bu bakımdan, kısa zaman enerjisi, genlik değişimlerini etkileyen bir etkendir [42-44]. Genel olarak kısa zaman enerjisi (E_n) Denklem 6.1 ile hesaplanmaktadır:

$$E_n = \frac{1}{N} \sum_{m=-\infty}^{\infty} [x(m)w(n-m)]^2 \tag{6.1}$$

Burada, $x(m)$ ayrık zamanlı ses işaretini, n zaman indeksini, $w(n\text{-}m)$ enerjinin hesaplandığı pencereyi ve N pencere uzunluğunu temsil etmektedir.

6.1.2 Sıfır Geçiş Sayısı

Sıfır geçiş sayısı (ZCR), bir ses işaretinin belirli bir zamandaki cebirsel işaretinin değişim sayısını ifade etmektedir. Sıfır geçişlerin tekrar sıklığını gösteren sayı, işaretin

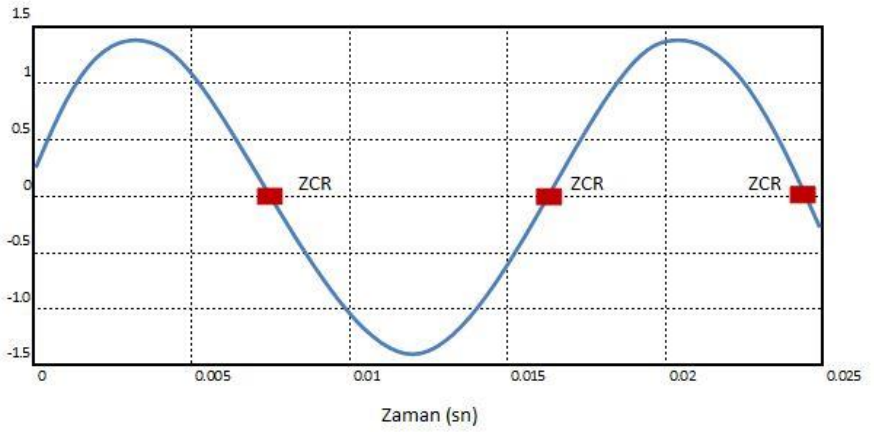
frekans içeriğinin basit bir ölçümünü ifade eder. Ses işaretlerinde ZCR, belli bir zaman aralığında ya da bir çerçeve içerisinde, ses işaretinin genlik değerinin sıfır değerine kaç defa geçtiği ile ölçülmektedir. ZCR değeri ile bir işaretin sıklığı ölçülmektedir. [42-44]. Şekil 6.1'de bir ses işaretinin sıfır geçiş sayısı gösterilmiştir.

Konuşma işaretinin kısa süreli ZCR (Z_n) değeri aşağıdaki denklemler ile ifade edilmektedir:

$$Z_n = \frac{1}{2}\sum_{m=-\infty}^{\infty} |sgn[x(n-m)] - sgn[x(n-m-1)]|w(m) \tag{6.2}$$

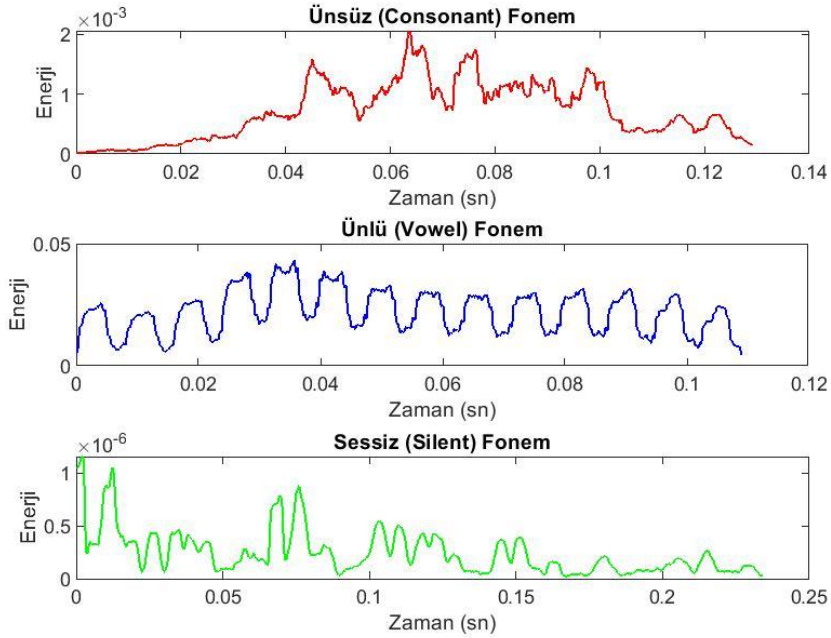
$$sgn[x(n)] = \begin{cases} 1, & x(n) \geq 0 \\ -1, & x(n) < 0 \end{cases} \tag{6.3}$$

Burada, *x* ses işaretini ifade ederken, *w* ise pencere fonksiyonunu ifade etmektedir.

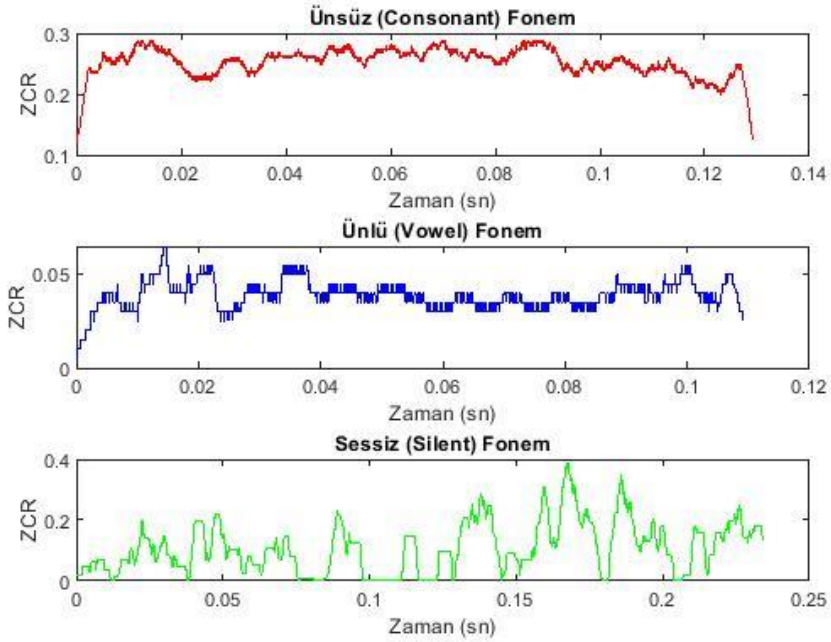


Şekil 6.1 Bir Ses İşaretinin Sıfır Geçişleri.

Akustik olarak ünlü fonemler ünsüz fonemlere göre daha yüksek enerji değeri taşımaktadırlar ve sessiz (konuşmanın olmadığı) fonemler, ünlü ve ünsüz fonemlere göre çok daha düşük (sıfıra yakın) enerji değeri taşımaktadırlar. Buna karşılık, ünlü fonemler ünsüz fonemlere göre daha düşük ZCR değeri taşımaktadırlar ve sessiz (konuşmanın olmadığı) fonemler, ünlü ve ünsüz fonemlere göre çok daha yüksek ZCR değeri taşımaktadırlar. Fakat, her ünlü, ünsüz ve sessiz fonemleri sadece enerji ve/veya ZCR değerlerine göre ayırt etmek mümkün olmamaktadır. Çünkü bu durum tamamen konuşmacıya ve konuşmacının konuşmayı ifade ediş tarzına bağlıdır. Şekil 6.2'de bir ünsüz, ünlü ve sessiz foneme ilişkin örnek enerji dağılımı ve Şekil 6.3'te ise, örnek ZCR dağılımı gösterilmiştir.



Şekil 6.2 Ünsüz, Ünlü ve Sessiz Fonemlere İlişkin Örnek Enerji Ölçümü.



Şekil 6.3 Ünsüz, Ünlü ve Sessiz Fonemlere İlişkin Örnek ZCR Ölçümü.

6.1.3 Mel Frekansı Kepstrum Katsayıları (MFCC)

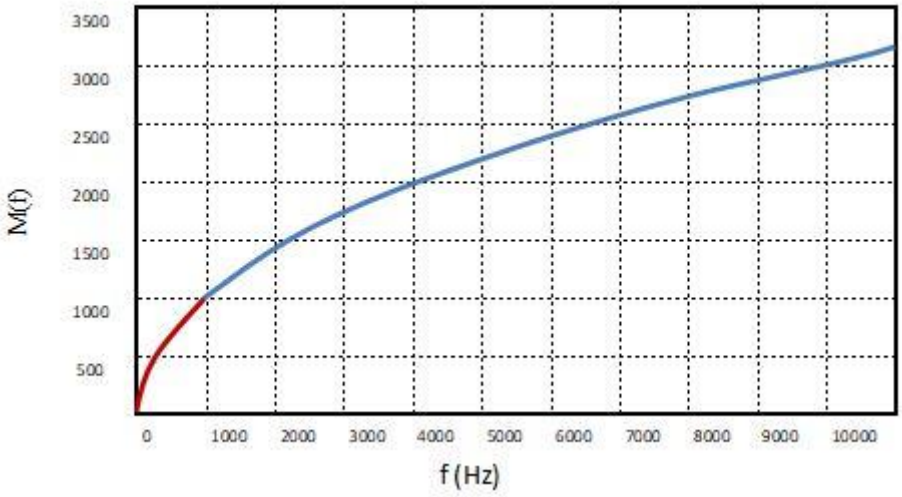
MFCC, ses işleme çalışmalarında başarılı sonuçlar elde edilmiş etkili bir özellik çıkarım parametresidir [45-46]. MFCC insan kulağının duyma özelliğini taklit eden ve Hızlı Fourier Dönüşümü (Fast Fourier Transform, FFT) tabanlı bir sayısal teknik analizdir [47]. FFT, çerçevelenmiş bir konuşma parçasını zaman uzayından frekans uzayına

dönüştürmek için kullanılır. *N* örnekli bir çerçeve için FFT dönüşümü Denklem 6.4 ile ifade edilir.

$$X_n = \sum_{k=0}^{N-1} x_k e^{-2\pi jkn/N} , \quad n = 0, 1, 2, \ldots\ldots, N-1 \tag{6.4}$$

Burada, *x*(*k*) ses işaretinin zaman eksenindeki *k.* değerini, X_n ise ses işaretinin frekans eksenindeki değerini göstermektedir.

İnsanlarda işitme sistemi, 1 kHz e kadar olan frekans değerlerini doğrusal, 1 kHz den yüksek olan frekans değerlerini ise, logaritmik olarak algıladığından "f" Hz birimindeki doğrusal frekans değerlerinin "mel(f)" Mel frekans değerlerine dönüştürülmesi gerekmektedir.



Şekil 6.4 Frekans Mel Dönüşüm Grafiği[48].

Şekil 6.4'te, Hz birimindeki frekans ile Mel frekans dönüşüm arasındaki ilişkinin grafiği gösterilmişir. Bu grafikte, Hz biriminden verilen bir ses işaretinin Mel frekans birimine çevirme işleminin cebirsel eşitliği Denklem 6.5 ile hesaplanmaktadır:

$$M(f) = 2595 \log_{10}\left(1 + \frac{f}{700}\right) \tag{6.5}$$

Burada, *f* Hz birimindeki doğrusal frekansı, *M* (*f*) ise Mel birimindeki frekans değerini ifade etmektedir.

FFT işlemi ile genlik spektrumu hesaplanan işaret, Mel spektrumunu elde etmek için Mel ölçekli bir süzgeç dizisinden geçirilir. Bu süzgeç dizisi, 1 kHz 'e kadar doğrusal ve 1 kHz'den yüksek frekanslarda ise logoritmik olarak yerleştirilmiş üçgen şeklindeki süzgeçlerden oluşmaktadır [48]. Hz biriminde verilen frekans değerlerini Mel ölçekli

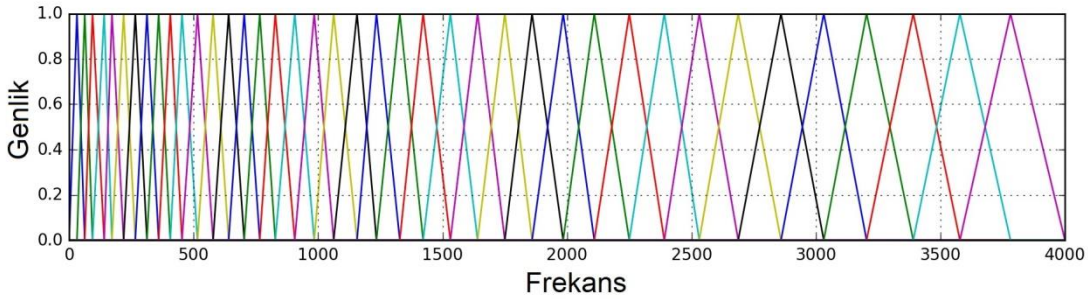
frekans spektrumuna çevirmek için Denklem (6.5) kullanılmaktadır. Şekil 6.5'te Mel-süzgeç dizisi görülmektedir. Böylece, doğrusal frekans değerlerinin insan sesinin algılanan frekansına yakın olan logaritmik (Mel) frekans boyutuna dönüşümü sağlanır. Daha sonra, Mel ölçekli süzgeç dizisinden geçirilen ses işaretinin logaritması alındıktan sonra Ayrık Kosinüs Dönüşümü (Discrete Cosine Transform, DCT) ile tekrar zaman frekansına çevrilerek mel frekansı kepstral katsayıları (MFCC) hesaplanır. Sonuçta, her bir *N* örnekli çerçeve için ayırt edici özellik vektörleri bulunmuş olur. MFCC katsayıları Denklem 6.6 ile hesaplanır:

$$c(n) = \sqrt{\frac{2}{K}} \sum_{i=1}^{K} (\log \tilde{s}_i) Cos\left[n\left(i - \frac{1}{2}\right)\frac{\pi}{K}\right], \quad n = 0, 1, 2, \ldots\ldots, K-1 \qquad (6.6)$$

Burada, *n*, mel frekansı kepstral katsayısı indeksi, $\tilde{s}_i$, mel spektrumu çıkış işaretini, *i*, süzgeç indeksi ve *K* süzgeç katsayısını göstermektedir. *K* (20 veya 40) adet katsayıdan genelde 13'ü MFCC katsayısı olarak kullanılır. MFCC 13x1 özellik (öznitelik) vektörü Denklem 6.7 ile gösterilmektedir.

$$K_{MFCC} = [c_1;\ \ldots;\ c_{13}\ ] \qquad (6.7)$$

Eşitlikte c_1; ……; c_{13} MFCC katsayılarını ifade etmektedir.



Şekil 6.5 Mel-Süzgeç Dizisi

Kepstral katsayıların değişimini daha iyi modellemek için genelde 13 MFCC katsayısının birinci ve ikinci türevleri alınarak Delta MFCC (D-MFCC) ve Delta Delta MFCC (DD-MFCC) katsayıları MFCC katsayılarına eklenmektedir. [47].

6.1.4 Delta MFCC

Konuşma işareti, küçük zaman dilimine ayrılan çerçevelerde sabit olduğundan delta ve delta delta katsayıları, bitişik çerçeveler arasındaki kepstral özellik vektörlerinin değişim hızını ve ivmesini modeler [47]. Deltalar, çerçeveler arasındaki farkın hesaplanması ile elde edilir.

Bir çerçeveden elde edilen 13 boyutlu bir adet MFCC vektörü için; n. vektörün (n-1). vektörden çıkarılmasıyla delta vektörü elde edilir[48].

Her bir çerçeve için 13 boyutlu MFCC katsayısına 13 delta katsayısı eklenerek 26 boyutlu bir D-MFCC özellik vektörü elde edilir. D-MFCC 26x1 özellik vektörü Denklem 6.8 ile gösterilmektedir.

$$K_{D-MFCC} = [c_1; \dots; c_{13}; c_{d1}; \dots; c_{d13}] \tag{6.8}$$

Eşitlikte, c_1 ; ……; c_{13} MFCC katsayılarını ve c_{d1} ; ……; c_{d13} delta katsayılarını ifade etmektedir.

6.1.5 Delta Delta MFCC

Bir çerçeveden elde edilen 13 boyutlu bir adet MFCC vektörü için; n. delta vektöründen (n-1). delta vektörünün çıkarılmasıyla delta delta vektörleri elde edilir[48].

Her bir çerçeve için 26 boyutlu D-MFCC katsayısına 13 delta delta katsayısı eklenerek 39 boyutlu bir DD-MFCC özellik vektörü elde edilir. DD - MFCC 39x1 özellik vektörü Denklem 6.9 ile gösterilmektedir.

$$K_{D\,D-MFCC} = [\, c_1\, ; \dots;\, c_{13}\, ;\, c_{d1}\, ; \dots;\, c_{d13};\, c_{a1};\, \dots;\, c_{a13}\,] \tag{6.9}$$

Eşitlikte c_1 ; ……; c_{13} MFCC katsayılarını, c_{d1} ; ……; c_{d13} delta katsayılarını ve c_{a1}; ……; c_{a13} delta delta katsayılarını ifade etmektedir.

6.2 Pencereleme Teknikleri

Genel bir ifade ile pencereleme, çerçevelenmiş bir işaretin özel bir fonksiyon ile çarpım işlemidir. Denklem 6.10’ de pencereleme işlemi verilmiştir:

$$y(n) = w(n) \times x(n) \tag{6.10}$$

Eşitlikte, $x(n)$ giriş işaretini, $y(n)$ pencerelenmiş işareti ve $w(n)$ ise pencere fonksiyonunu göstermektedir.

Bu adımdaki amaç, her bir çerçevenin başındaki ve sonundaki süreksiz bilgileri ortadan kaldırmaktır [49]. Bu işlem, çeşitli pencereleme teknikleri ile gerçekleştirilebilir. Pencereleme için en yaygın kullanılan teknikler Hamming, Hanning ve Rectangular pencereleme teknikleridir [50-51]. Blackman, Barlett ve Kaiser de kullanılan diğer pencereleme teknikleridir. Bu çalışma kapsamında Hamming, Hanning ve Rectangular pencereleme teknikleri incelenmiştir.

6.2.1 Hamming Pencereleme

Hamming pencereleme fonksiyonu, Denklem 6.11 ile tanımlanmaktadır.

$$w(n) = \begin{cases} 0.54 - 0.46cos\left(\frac{2\pi n}{N-1}\right), & 0 \leq n \leq N-1 \\ 0, & diğer \end{cases} \tag{6.11}$$

Burada, N örnek sayısını (pencere uzunluğunu) ifade etmektedir.

6.2.2 Hanning Pencereleme

Hanning pencereleme fonksiyonu, Denklem 6.12 ile tanımlanmaktadır:

$$w(n) = \begin{cases} \frac{1}{2}\left\{1 - cos\left(\frac{2\pi n}{N-1}\right)\right\}, & 0 \leq n \leq N-1 \\ 0, & diğer \end{cases} \tag{6.12}$$

Burada, N pencere uzunluğunu göstermektedir.

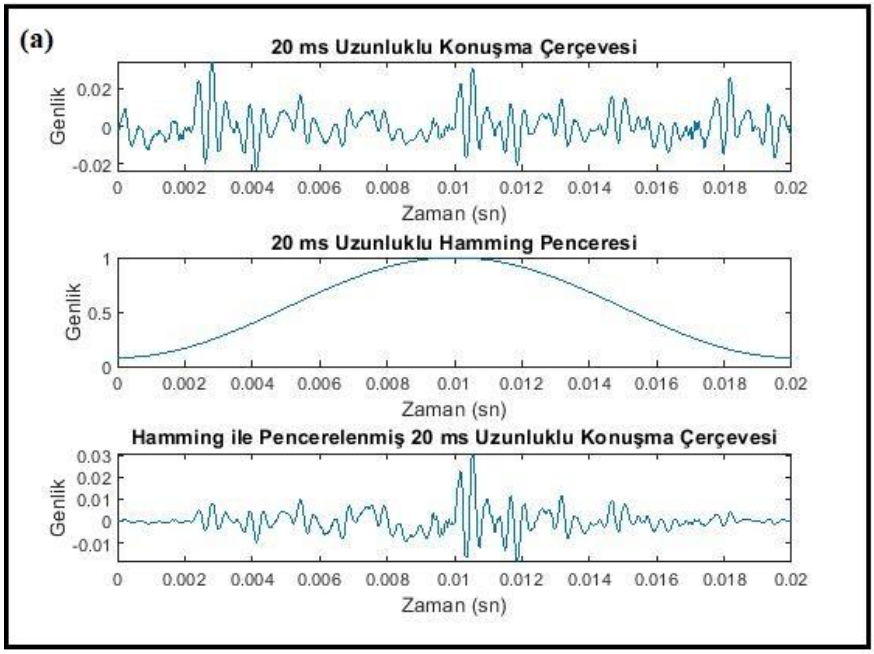
6.2.3 Rectangular Pencereleme

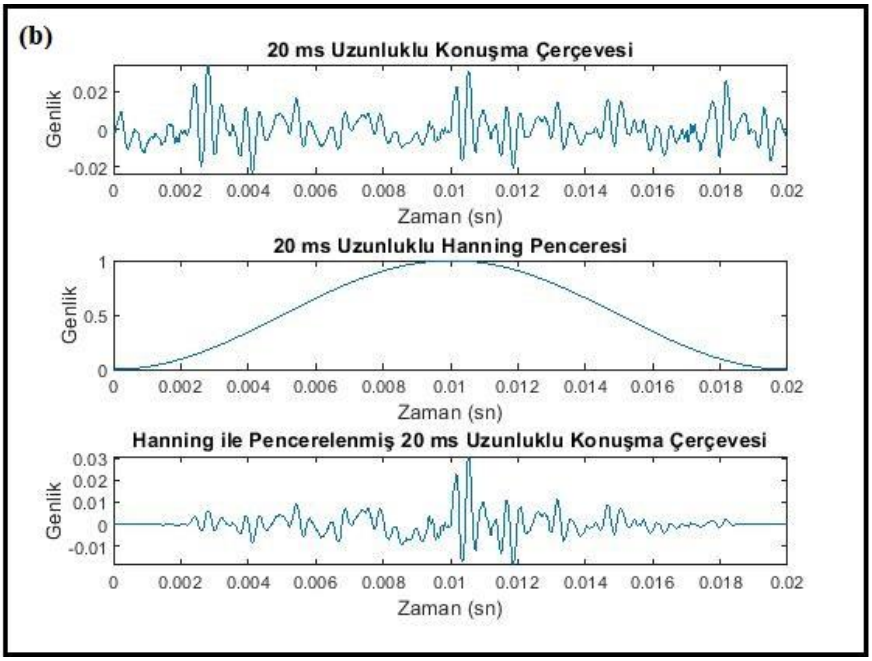
Rectangular (Dikdörtgen) pencereleme fonksiyonu, Denklem 6.13 ile tanımlanmaktadır.

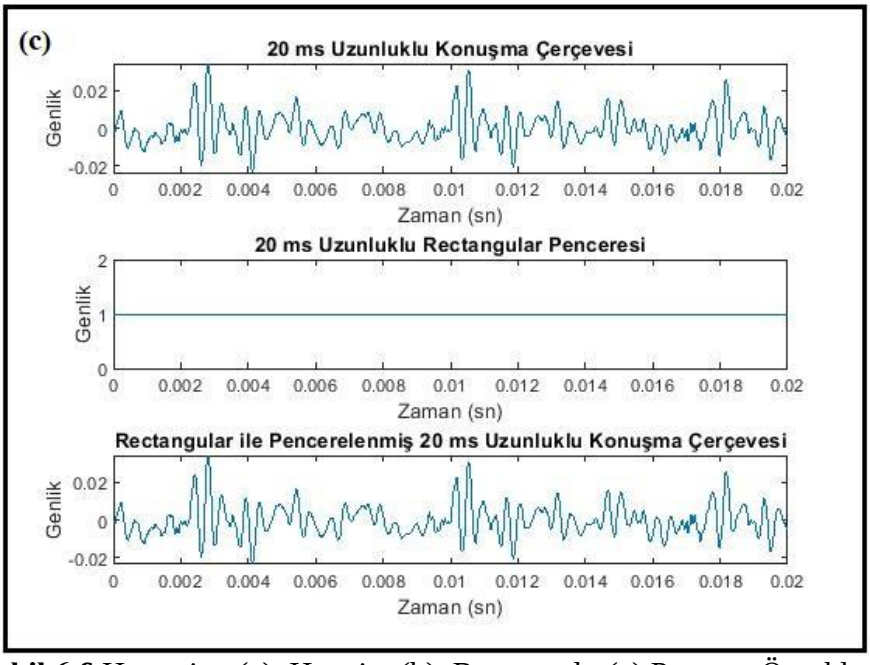
$$w(n) = \begin{cases} 1, & 0 \leq n \leq N-1 \\ 0, & diğer \end{cases} \tag{6.13}$$

Burada, N pencere uzunluğunu belirtmektedir.

Bir kadın konuşmacıdan alınan 20 ms'lik konuşma işareti için Hamming, Hanning ve Rectangular pencereleme örnekleri, Şekil 6.6' da gösterilmektedir.


(a)
20 ms Uzunluklu Konuşma Çerçevesi
Genlik
Zaman (sn)
20 ms Uzunluklu Hamming Penceresi
Genlik
Zaman (sn)
Hamming ile Pencerelenmiş 20 ms Uzunluklu Konuşma Çerçevesi
Genlik
Zaman (sn)


(b)
20 ms Uzunluklu Konuşma Çerçevesi
Genlik
Zaman (sn)
20 ms Uzunluklu Hanning Penceresi
Genlik
Zaman (sn)
Hanning ile Pencerelenmiş 20 ms Uzunluklu Konuşma Çerçevesi
Genlik
Zaman (sn)



Şekil 6.6 Hamming (a), Hanning(b), Rectangular(c) Pencere Örnekleri.

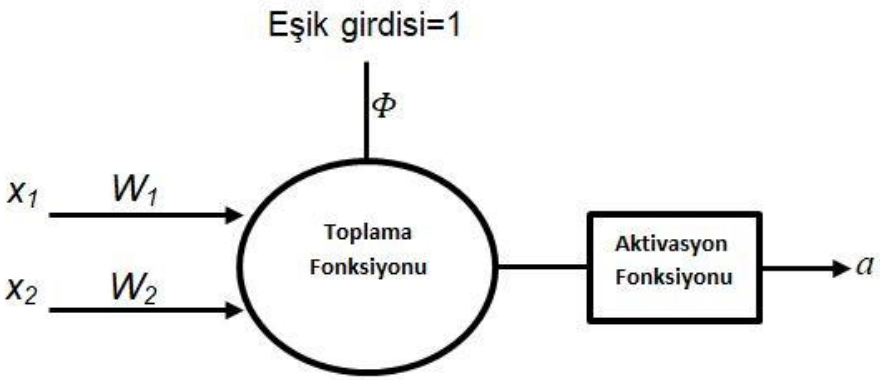
6.3 Yapay Sinir Ağları (ANN)

Yapay Sinir Ağları (Artificial Neural Network, ANN), insan beyninin sinirsel iletimini taklit ederek oluştulmuş bir yapay zeka tekniğidir. ANN ses işleme uygulamalarında örüntü tanıma ve sınıflandırma amacıyla kullanılan etkili bir yöntemdir.

ANN'ler birden fazla düğümün (nöronun) birbirleriyle olan ağırlıklı bağlantılarından öğrenme, genelleme, yeni bilgiler keşfedebilme ve türetebilme amacıyla geliştirilmiş sistemlerdir. Bu sistemlerdeki ağırlıklı bağlantılar çıktı davranışını öğrenebilecek şekilde eğitilerek değiştirilir. Basit bir ANN örneği Şekil 6.7'de gösterilmektedir. Bu ANN biririne bağlı iki katmandan oluşmaktadır. Bunlar, giriş katmanı ve çıkış katmanıdır. Giriş katmanında x_1 *ve* x_2 girdi verilerini ve çıkış katmanında da a çıktı verisini ifade etmektedir. w_1 ve w_2, girdi verilerinin bağlantı ağırlıklarını ifade etmektedir. Toplama fonksiyonu (Σ); nörona gelen net girdiyi hesaplamak için, aktivasyon fonksiyonu (f); nörona gelen net girdiye karşılık üreteceği çıktı değerini belirlemek için kullanılmaktadır. En sık kullanılan toplama fonksiyonu girdi değerleriyle ağırlıkların çarpılıp toplandığı toplam fonksiyonudur.

Bu mimari yapıda giriş katmanında yer alan bütün verilerin ağırlıklı bağlantıları tek bir nöron birimine bağlanarak çıkış katmanındaki çıktıyı hesaplar. Hesaplanan çıktı değerinin sıfır olmaması için bir eşik değeri (Φ) kullanılır ve bu eşik değeri sisteme her zaman 1 olarak atanır [52]. Ağın çıktısı Denklem 6.14 eşitliği ile hesaplanmaktadır:

$$a = f(\sum_{i=1}^{2} W_i\, x_i + \Phi) \tag{6.14}$$



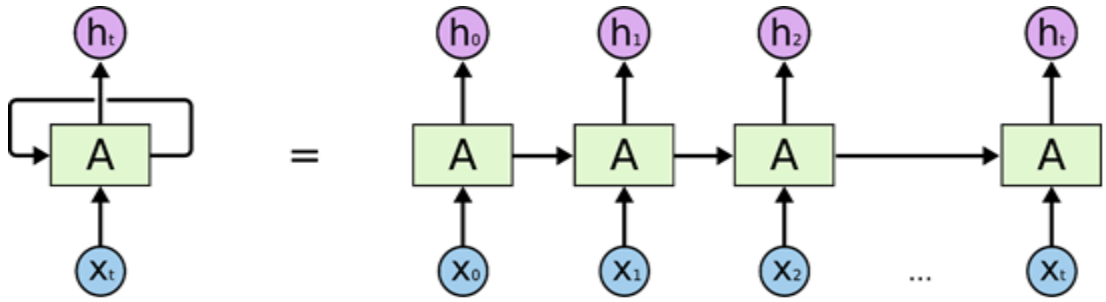
Şekil 6.7 Basit Bir ANN Örneği.

Basit bir ANN mimari yapısı tek katmanlı algılayıcılar (perceptron) olarak da adlandırılırlar. Tek katmanlı algılayıcılar, sadece doğrusal olan problemlerin çözümü için kullanılırlar ve bu algılayıcılar çok katmanlı algılayıcıların temelini oluşturmaktadırlar.

Derin öğrenme (DL), temeli ANN modeline dayanan 2006 yılında geliştirilen ve günümüzde özellikle ses ve görüntü işleme alanında sıkça kullanılmakta olan bir modeldir. Derin öğrenme yöntemleri ile karmaşık ilişkilerin analizi çok kolay bir şekilde gerçekleştirilir. Mimari olarak çoklu katman yapısında ve bu yapı itibari ile çok sayıda saklı bilgileri barındırma kabiliyetine sahiptir. Derin öğrenme ile model veriler ağdan geçerken ilgili bilgileri otomatik olarak öğrenir ve özetler. "Derin" terimi ağdaki katman sayısını (yani katman sayısı ne kadar fazlaysa ağın derinliğinin de o kadar fazla olduğunu) ifade etmektedir.

6.3.1 Tekrarlayan Sinir Ağları (RNN)

Tekrarlayan Sinir Ağı, sıralı gelen verilerin işlendiği bir ağ sistemidir. Bu sistemler, zamansal bilgileri çok dinamik bir şekilde potansiyel olarak yakalayabilir ve ağın her bir zaman adımı için kullanılacak bağlamsal bilgi miktarına özgürce karar verebilme mekanizmasını sağlar [53]. İnsan beyninin biyolojik tekrarlı sinir ağlarının (bRNN) çalışma prensibinden esinlenilmiştir. RNN içerisindeki her nöron, önceki girdi ile gelen bilgileri koruyarak çıktı üretir. Bu yapılarda bilgi akışının devamı yönlendirilmiş bir döngü içerisindeki matematiksel hesaplamalar ile gerçekleşir. Böylece, RNN geçmiş bilgileri kullanarak tanıma, sınıflandırma, zamansal ilişki ve tahmin işlemlerini gerçekleştirmektedir [54]. RNN ağ sisteminin yapısı Şekil 6.8'de gösterilmiştir. Şekle göre, X*:* giriş değerini, A: RNN sinir ağı hücresini ve h: çıkış değerini belirtmektedir. RNN sinir ağı hücresinden çıkan bir değer tekrar kendisine gelerek bir döngü oluşturmaktadır. Bu döngü sayesinde yeni bilgi, eski (önceki) bilgi ile birlikte değerlendirilmiştir.

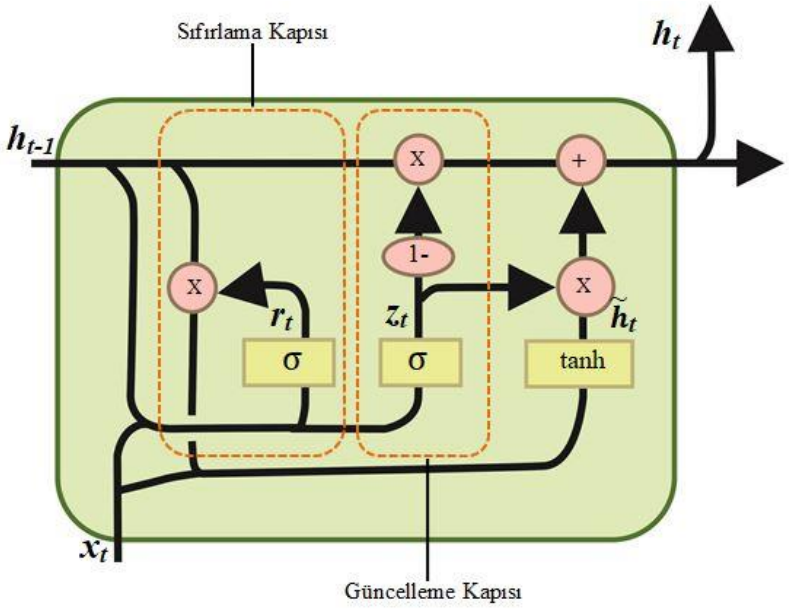


Şekil 6.8 RNN Ağ Modeli[55].

6.3.2 Geçitli Tekrarlayan Birim (GRU) Tekrarlayan Sinir Ağları

Geçitli Tekrarlayan Birim, uzun vadeli bağımlılıkları öğrenebilen bir RNN sinir ağı modeli olarak tanımlanır. Ancak, standart RNN'lerde bütün bilgiler ağ içerisinde tutulduğu için hangi bilgiler ne kadar sürede hatırlanacak ve gereksiz bilgilerin de geçmişte saklanması, gerekli olan bilgilerin hatırlanamaması büyük bir sorun oluşturmaktadır. Bu sorunlar RNN'lerin eğitim sırasında karşılaşılabilecek patlayan ve yok olan gradyan (vanishing gradient) problemleri olarak ortaya çıkmaktadır. Bu sebeple RNN ağ sisteminin özel bir türü olan GRU bu problemlere çözüm getirmek için geliştirilmiştir. GRU, RNN'nin diğer bir özel mimari türü olan LSTM ile karşılaştırılabilir bir performans sunmaktadır. GRU'lar karmaşık LSTM mimari

yapısının daha basit bir tasarımıdır. Bu sepeble GRU model yapısı da, LSTM birimine benzer şekilde, fakat GRU'da ayrı bir bellek hücresi olmadan birim içindeki bilgi akışını modüle eden gizli durum ve karar kapıları (geçitleme mekanizmaları) vardır. Bu bağlamda, GRU'ların hesaplama işlemi LSTM'lere göre daha hızlıdır. GRU model yapısı Şekil 6.9'da gösterilmiştir.



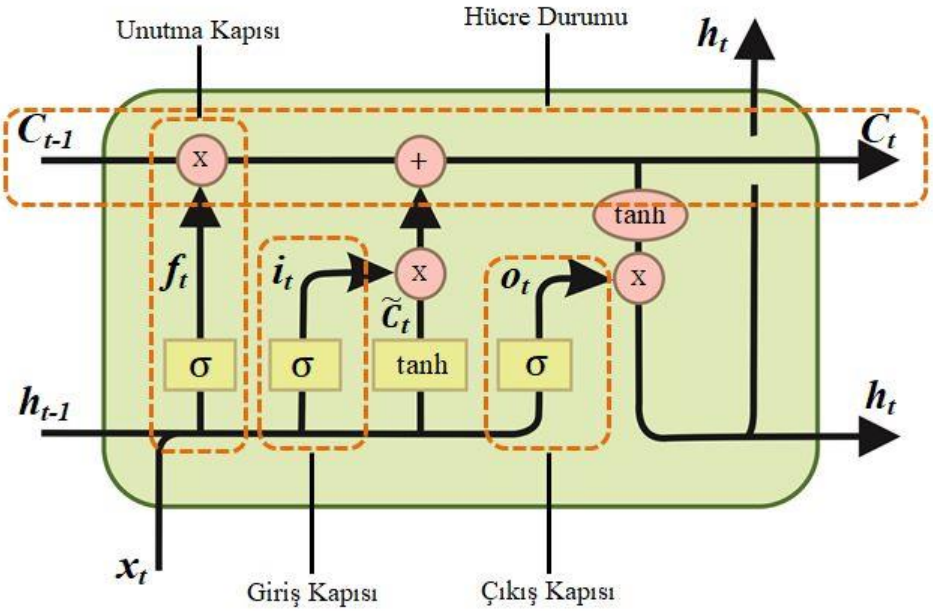
Şekil 6.9 GRU Model Yapısı [56].

GRU sıralı giriş verilerinin uzun süreli bağımlılıklarını veya sıralı giriş verilerini uzun zaman boyunca silmeden bellekte tutma kabiliyetini, GRU birimi içindeki hesaplamalar ile gerçekleştirir. Bu hesplamalarda her bir zaman dilimi için yalnızca bir gizli durum üretilir. Bu gizli durum, gizli durumun ve giriş verilerinin geçtiği geçitleme mekanizmaları ve hesaplamaları nedeniyle hem uzun hem de kısa vadeli bağımlılıkları aynı anda tutabilir. GRU hücresi, yalnızca güncelleme kapısı (update gate) ve sıfırlama kapısı (reset gate) olarak adlandırılan iki kapı içerir. Zaman adımlarındaki bilgi akışının kontrolünü güncelleme kapısı, geçmiş verilerin ne kadarının geçip ne kadarının unutulacağına sıfırlama kapısı karar vermektedir. GRU'daki bu kapılar önemli olan bilgileri korurken önemsiz olan bilgileri seçici olarak filtrelemek için eğitilmiştir. Bu kapılar, giriş ve/veya gizli durum verileri ile çarpılarak 0 ile 1 arasında değerler içeren vektörlerdir. Kapı vektörlerindeki değer sıfıra yakınsa, giriş veya gizli duruma karşılık gelen verilerin önemsiz olduğunu ve bu nedenle sıfır olarak döneceğini gösterir. Öte yandan, kapı vektöründeki değer bir değerine yakınsa gelen verilerin önemli olduğu ve

bu verilerin kullanılacağı anlamına gelir. Bu bağlamda, GRU tekrarlayan sinir ağları sıralı verilerdeki uzun vadeli bağımlılıkları etkili bir şekilde koruyabilmektedir. Kısaca özetlemek gerekirse, bu mimari yapı daha uzun vadeli tahminler yapabilmek için kapı birimlerinin nasıl kullanılması gerektiğini öğrenir [57].

6.3.3 Uzun Kısa Süreli Hafıza (LSTM) Tekrarlayan Sinir Ağları

LSTM, GRU tekrarlayan sinir ağlarının daha geniş bir değişimi olarak düşünülebilir. LSTM model yapısı GRU'lardan farklı olarak bir hafıza hücresi birimine sahiptirler. Bu hafıza hücresi ile önceki zamandan gelen bilgiler bir sonrakine aktarılarak, bilgilerin uzun veya kısa süreli zaman aralığında hatırlanması sağlanmaktadır. LSTM tekrarlayan sinir ağları yapısı, bir hafıza hücresi (memory cell), giriş (input), çıkış (output) ve unutma (forget) karar kapılarından meydana gelmektedir. LSTM model yapısı Şekil 6.10'da gösterilmiştir [58].



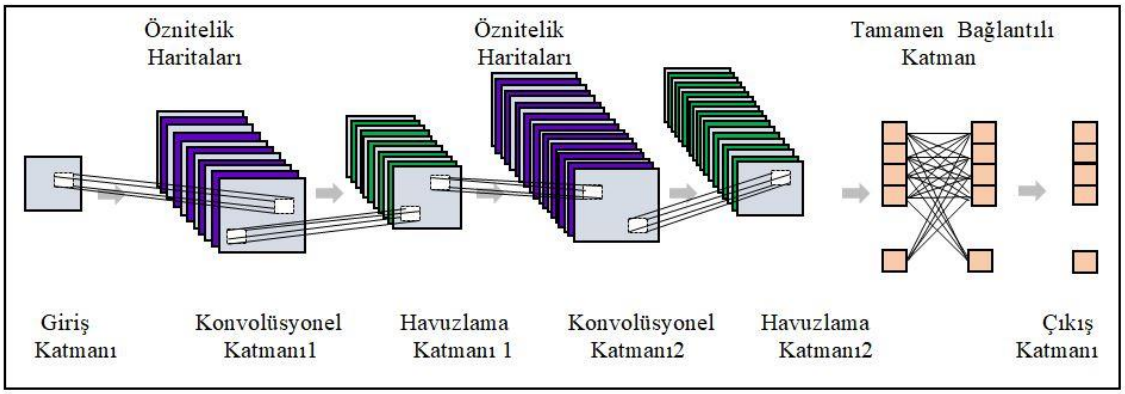
Şekil 6.10 LSTM Model Yapısı[56].

6.3.4 Evrişimsel Sinir Ağları (CNN)

Derin öğrenme yöntemlerinden olan evrişimsel sinir ağları (CNN veya ConvNet), ses işleme, görüntü işleme ve yapay zeka uygulamaları için etkin çözümler sunmaktadır. İleri yönlü bir sinir ağı olan evrişimsel sinir ağları bir MLP türüdür [59]. Katmanlar, önceki katmanın çıktısını girdi olarak kullanmak suretiyle, her bir gizli katman ile düğümler veya nöronlar arasında birbirine bağlanır. Şekil 6.11'de evrişimsel sinir

ağlarının çalışma prensibinin genel yapısı sunulmaktadır. CNN bir giriş katmanı, bir çıkış katmanı ve bunlar arasında bir dizi gizli katmanı içermektedir. Bu gizli katmanlar temelde konvolüsyonel katmanı, havuzlama katmanı ve tamamen bağlantılı katmanlardan oluşmaktadır. Konvolüsyonel katmanında giriş katmanındaki özellik vektörleri filtre kerneli ile konvolüsyon işlemine tabi tutularak öznitelik haritaları elde edilir. Havuzlama katmanında elde edilen her bir öznitelik haritası ayrı ayrı ele alınarak her bir harita komşu değerlerin ortalaması veya maksimum değerinin elde edilmesi ile örneklenmiş öznitelik haritaları üretilir. Havuzlama katmanı ile öznitelik vektörlerinin boyutsal küçülmesi ve ağın daha kısa bir sürede öğrenmesi sağlanır. Son katman olan tamamen bağlantılı katmanda önceki katmanlardan gelen her girişin bu katmandaki tüm nöronlara bağlı çıkışın üretildiği katmandır.

CNN'de her bir katman, giriş verilerindeki özellikleri tespit etmek amacıyla onlarca veya yüzlerce katman üzerinde tekrar tekrar öğrenme gerçekleştirir [60].



Şekil 6.11 CNN Mimari Yapısı [60].

6.4 Sınıflandırıcı Yöntemleri

6.4.1 Naive Bayes

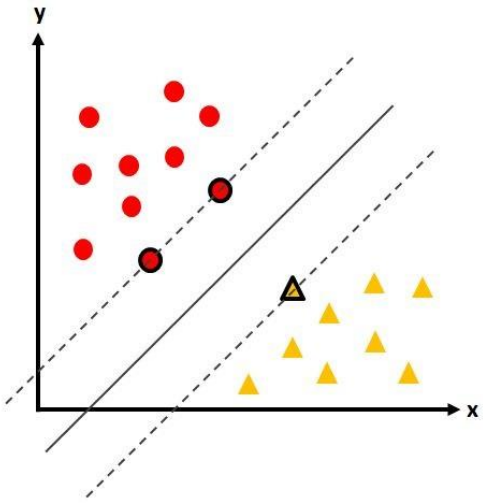
Naive Bayes, Bayes teoremine bağlı olasılık tabanlı bir sınıflandırma algoritmasıdır. Bu algoritmada, test verisinin sınıfı, eğitim seti üzerinde yapılan bir dizi olasılık hesaplamalar ile elde edilen en yüksek değere sahip olan sınıf kümesine dâhil olması ile belirlenir. Büyük veri setleri için kullanışlı ve etkili bir algoritmadır. Bayes teoremi Denklem 6.15 ile ifade edilmektedir.

$$P(A|B) = \frac{P(A|B) \times P(A)}{P(B)} \tag{6.15}$$

- $P(A|B)$: B olayının A olayında olma olasılığı
- $P(B|A)$: A olayının B olayında olma olasılığı
- $P(A)$: A olayının olma olasılığı
- $P(B)$: B olayının olasılığı

6.4.2 Destek Vektör Makinaları

Destek vektör makinaları (Support Vector Machines, SVM), özellikle ses işleme ve görüntü işleme gibi örüntü tanıma uygulamalarında özellik uzayındaki örnekleri ayırabilen etkin bir sınıflandırıcı yöntemidir. SVM yöntemlerinde amaç özellik uzayında yer alan iki farklı sınıfın örnekleri arasındaki en uzak sınırın (hiper düzlemin) bulunmasıdır. Şekil 6.12'de görüldüğü gibi örnekler arasındaki mesafeyi en uygun şekilde ayırabilen bir karar sınırını bulmayı amaçlar. Paralel doğrular üzerinde siyah olarak işaretlenmiş örnekler destek vektörlerini belirtir.

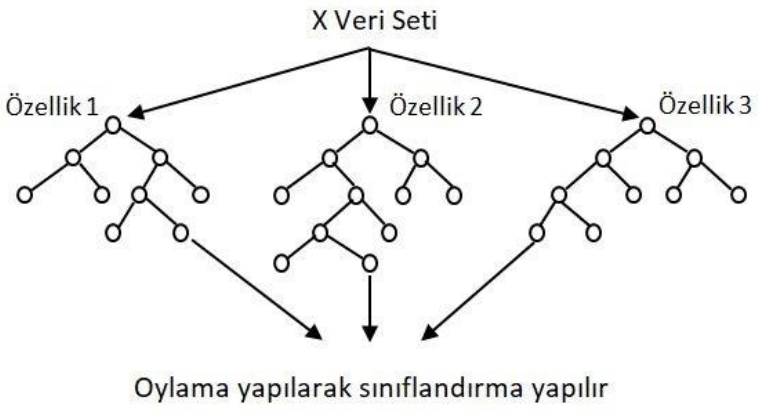


Şekil 6.12 SVM Sınıflandırma Metodu.

6.4.3 Rastgele Orman

Rastgele Orman (Random Forest, RF) yöntemi, hem sınıflandırma hem de regresyon problemlerinin çözümünde sıklıkla kullanılan veri madenciliği modellerinden biridir. Bu yöntemde eğitim rastgele çok sayıda farklı alt setlerin eğitilmesiyle oluşturulan karar ağaçları ile gerçekleştirilir. Bu yöntemde oluşturulan karar ağaçları topluluğu RF olarak

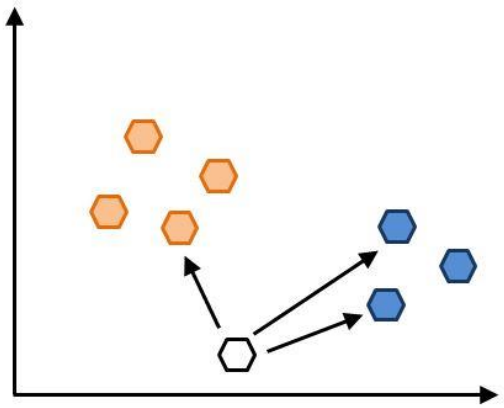
adlandırılmaktadır. RF sınıflandırma modeli Şekil 6.13'de gösterilmiştir. Bu sınıflandırma modelinde sınıfı bilinmeyen bir test örneği en yüksek değerli karar ağacının sınıfına göre atanır. RF modelinin en önemli avantajları aşırı öğrenme/ezberleme (overfitting) ve aykırılık (outlier) problemlerini önlemesidir.



Şekil 6.13 RF Sınıflandırma Metodu.

6.4.4 k-En Yakın Komşu

k-En Yakın Komşu (k-Nearest Neighbour, k-NN) yöntemi, sınıflandırma ve regresyon tahmin problemlerinde kullanılan makine öğrenmesi yöntemlerindendir. Basit ve kolay uygulanabilirliğinin yanı sıra büyük veri setleri için kararlılığın olması bu yöntemin kullanılmasını yaygınlaştırmıştır. Bu yöntem sınıflandırılması istenen yeni bir örneğin daha önceki etiketlenmiş örneklerden hangisine daha çok benzeyeceğinin uzaklık mesafesi ile belirlenmesine dayanan bir sınıflandırma yöntemidir. k-NN algoritmasında en çok kullanılan uzaklık ölçü birimi Öklit mesafesidir. Manhattan ve Minkowski ölçüleri de kullanılmaktadır. Bu yöntemde en yakın kaç tane komşu seçileceği k-NN algoritmasının performansını önemli ölçüde etkilemektedir. Bu nedenle en yakın kaç tane komşu seçileceğine karar veren bir *k* değeri kullanılmaktadır. Şekil 6.14'te *k* değeri 3 seçilerek yeni gelen bir örneğin uzaklık mesafesi olarak kendisine en yakın 3 komşusuna bakarak hangi sınıfa ait olacağına karar verilmektedir.



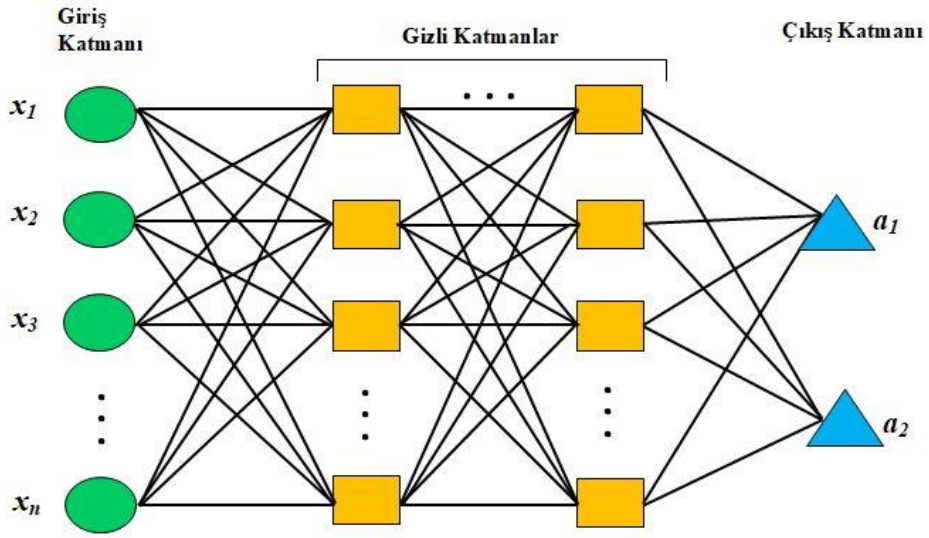
Şekil 6.14 k-NN Sınıflandırma Metodu.

6.4.5 Çok Katmanlı Algılayıcılar (MLP)

MLP, biyolojik sistemlerdeki sinir ağlarına benzer şekilde insan beyninin çalışma modelinden esinlenilmiş ve öğrenme algoritmasına sahip nöronların birbirlerine bağlanması ile oluşan bir yapay sinir ağı yapısıdır. MLP'de, her katmanın tamamen bir sonrakine bağlı olduğu katmanlarda birden fazla karar düğümü yer alır [61].

Şekil 6.15'te genel bir MLP mimarisi gösterilmektedir. Bu mimari birbirine bağlı üç katmandan oluşur. Bunlar, giriş katmanı, belirli sayıda gizli (ara) katman ve çıkış katmanıdır. Giriş katmanı, x_1, x_2, x_3, ..., x_n girdi verilerini alarak belirli ağırlık işlemlerine tabi tutup bir sonraki gizli katmana verileri aktarır. Daha sonra o gizli katman, varsa kendisinden sonraki gizli katmana gelen verileri aktarır. Böylece her bir katmanın çıkışı bir sonraki katmanın girişi olmaktadır. Çıkış katmanı ise, önceki katmanlardan gelen verileri işleyerek ağın a_1 ve a_2 çıktılarını belirler. Bilgi bu katmanlar arasındaki paralel çalışan düğümlerin (nöron birimlerinin) hiyerarşik bağlantıları üzerine yayılır. Bu bağlamda MLP sistemleri doğrusal olmayan fonksiyonları kullanarak karmaşık problemlere başarımı yüksek çözümler sunmaktadır.

Bu sistemlerde, ağın eğitimi ileri doğru yayılım (forward propogation) ve geri doğru yayılım (back propogation) olmak üzere iki aşamada gerçekleştirilmektedir. Birinci aşamada, giriş verileri gizli katmanlardan geçerek çıktı katmanına kadar işlenir. İkinci aşamada ise, çıktı katmanından elde edilen hata ağırlık değerlerine dağıtılarak her iterasyonda hata payının azaltılması beklenir



Şekil 6.15 MLP Mimari Yapısı [60].

MLP sistemlerinde belirli bir problemin çözümüne uygun kullanılacak gizli katman sayısı, nöron birimlerinin sayısı ve bağlantı ağırlıklarının sayısı gibi parametrelerin belirlenmesi zor bir işlemdir. Bu bağlamda, bu parametreler deneme yanılma yoluyla probleme en iyi çözümü (başarımı) sunacak parametlerin elde edilmesi ile belirlenir.

7
Önerilen Model: GRU Tabanlı C/V/S Konuşma Bölütlerinin Tespiti için En Uygun Özellik Parametre Setinin Belirlenmesi

Bu tez çalışmasında, fonem seviyesinde bir konuşma tespit sisteminin derin öğrenme yöntemi ile tasarlanması amaçlanmıştır. Bu amaçla ilk olarak özgün bir veri kümesi oluşturulmuştur. Daha sonra, oluşturulmuş bu özgün veri kümesine bir dizi işlem (önişleme, özellik çıkarma, eğitim ve test) uygulanmıştır. Bu işlemlerin uygulanmasında PRAAT 6.0.49 [62], MATLAB R2018a [63], Keras Kütüphanesi [64], Weka 3.9.3 [65] ve WekaDeeplearning4j [66] yazılım araçları kullanılmıştır.

7.1 Veri Kümesinin Toplanması

Bu çalışmada özgün bir Kürtçe veri kümesi oluşturulmuştur. Veri kümesinde yer alan ses örnekleri Türkiye Radyo Televizyon Kurumunun haber sitesinden alınmıştır. Alınan ses örnekleri eğitim, kültür, sanat, ekonomi, sağlık ve politika konulu farklı haber içeriklerinden oluşmaktadır.

Farklı uzunluktaki ses örnekleri, yaşları 20-45 arasında değişen 4 yetişkin erkek ve 3 yetişkin kadın konuşmacıdan alınmıştır. Elde edilen kayıtlar 16-bit çözünürlükte, tek kanallı 44100 Hz örnekleme frekansıyla "PCM wave" dosya formatındadır. Bu çalışmada, Kürtçe dilindeki sürekli konuşmalardan elde edilen ses örneklerine fonem seviyesinde bölütleme (segmentasyon) işlemi uygulanarak yaklaşık 6819 fonemden oluşan bir veri kümesi hazırlanmıştır. Bu çalışmada kullanılan veri kümesi özelliklerine ait ayrıntılı bilgiler erkek konuşmacılar için Tablo 7.1 ve kadın konuşmacılar için Tablo 7.2' de gösterilmiştir.

Tablo 7.1 Erkek Konuşmacıların Veri Kümesi Özellikleri

Erkek Konuşmacılar	Fonem Sayısı	Ünlü Fonem Sayısı (V)	Ünsüz Fonem Sayısı (C)	Sessiz Fonem Sayısı (S)
Erkek-Konuşmacı1	1.332	583	749	36
Erkek-Konuşmacı2	1.096	469	627	40
Erkek-Konuşmacı3	1.100	462	638	46
Erkek-Konuşmacı4	327	146	181	12
Toplam	3.855	1.660	2.195	134

Tablo 7.2 Kadın Konuşmacıların Veri Kümesi Özellikleri

Kadın Konuşmacılar	Fonem Sayısı	Ünlü Fonem Sayısı (V)	Ünsüz Fonem Sayısı (C)	Sessiz Fonem Sayısı (S)
Kadın-Konuşmacı1	780	326	454	26
Kadın-Konuşmacı2	1.410	615	795	56
Kadın-Konuşmacı3	774	341	433	44
Toplam	2.964	1.282	1682	126

7.2 Ön İşleme

Bu çalışmada, mikrofon sesinden kaynaklı gürültüleri ortadan kaldırmak için, veri kümesini oluşturan ses örnekleri, bir ön işlem aşaması olarak Wiener filtresine tabi tutulmuştur [67]. Wiener filtresi, frekans uzayında gerçekleştirilen bir yöntemdir ve gürültülü bir konuşma işareti $y(n)$ Denklem 7.1 ile ifade edilir:

$$y(n) = x(n) + v(n) \tag{7.1}$$

Burada, $x(n)$ temiz konuşma işaretini, $v(n)$ beyaz Gauss gürültüsünü ve n ayrık zaman

değişkenini tanımlamaktadır.

n zamanındaki temiz konuşma örneği ile öngörülen (tahmini) konuşma örneği arasındaki fark $e_x(n)$ hata işareti Denklem 7.2 ile hesaplanmaktadır:

$$e_x(n) \triangleq x_n - h^T y(n) \tag{7.2}$$

Burada, üst simge *T*, bir vektör veya matrisin transpozunu tanımlamaktadır. *h*, *L* uzunluğunda bir Sonlu Darbe Tepki (Finite Impulse Response, FIR) süzgecini belirtmektedir ve transpoz (h^T) değeri Denklem 7.3 ile ifade edilmektedir:

$$h^T = [h_0, h_1, \dots, h_{L-1}] \tag{7.3}$$

Denklem 7.4 de *y*, $y(n)$ gözlem işaretinin en son örneklerini içeren bir vektördür ve transpoz (y^T) değeri Denklem 7.4 ile ifade edilmektedir:

$$y^T = [y(n), \dots, y(n-1), y(n-L+1)] \tag{7.4}$$

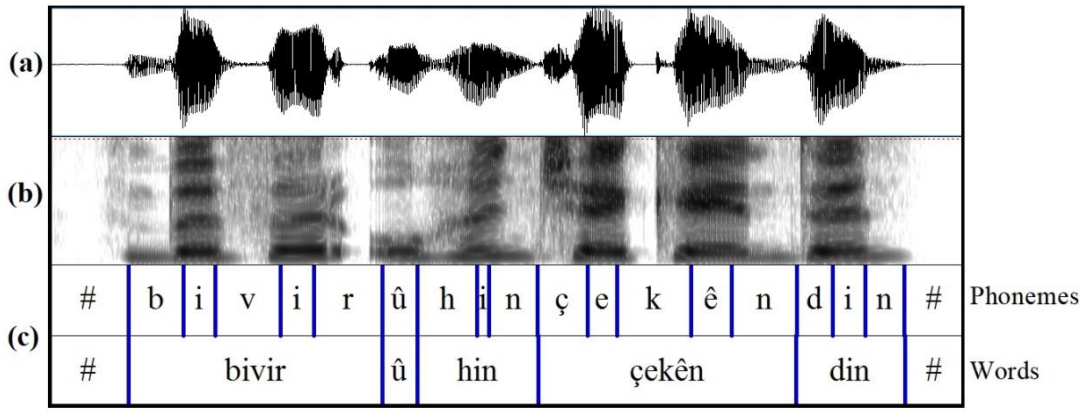
Wiener filtre katsayıları, ortalama karesel hatayı ($e_x^2(n)$) minimum yapan değer ile elde edilir.

7.3 Veri Kümesinin Hazırlanması

Kürtçe dilinde oluşturulmuş veri kümesindeki her bir konuşma işaretine uygun ünlü, ünsüz ve sessiz (C/V/S) konuşma sınıfı atamak için, her bir konuşma cümlesi fonem düzeyinde elle bölütlenmiştir. Elle bölütleme yönteminin, otomatik bölütleme yöntemine kıyasla daha doğru performans gösterdiği düşünülmektedir [68-71]. Bu nedenle, bu çalışmada fonem düzeyinde özgün bir Kürtçe veri kümesi oluşturmak için elle bölütleme yöntemi kullanılmıştır.

Sürekli bir konuşma işaretinin fonem düzeyinde bölütleme işleminin gerçekleştirilmesi için konuşma işaretinin önce kelime düzeyinde bir bölütleme işlemine tabi tutulması gerekmektedir. Bu çalışmada bölütleme işlemi "Praat" ses analiz aracı ile gerçekleştirilmiştir. Bu yazılım aracı ile, her bir konuşmacıya ait sürekli konuşma işaretlerinin spektrogram ve dalga özellikleri görsel ve işitsel olarak dikkatlice incelenmiş ve her bir konuşma işareti C/V/S olarak etiketlenmiştir. Böylece, her bir konuşma işaretinden C/V/S etiketli bir konuşma dosyası oluşturulmuştur.

Şekil 7.1'de Praat programı ile yetişkin bir kadın konuşmacı tarafından Kütçe ifade edilen "Bivir û hinçeken din" örnek cümlesinin dalga şekli, spektrogramı, fonem ve kelime seviyesindeki bölütleme ve etiketleme işlemleri gösterilmiştir.

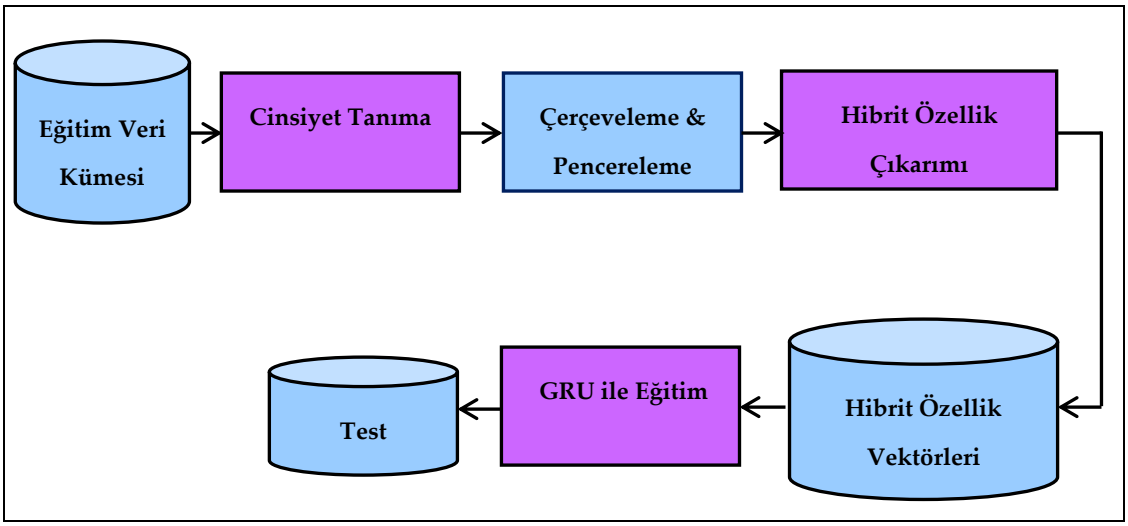


Şekil 7.1 Kürtçe Konuşma İşaretinin Dalga Formu (a), Spektrogramı (b) ve Fonem ve Kelime Düzeyindeki Bölütleme ve Etiketleme İşlemleri (c).

Kayıtlardan elde edilen Kürtçe dilindeki sürekli konuşmaların C/V/S bölütleme ve etiketleme işlemleri tamamlandıktan sonra C/V/S fonem tabanlı bir veri kümesi oluşturulmuştur. Oluşturulan veri kümesi %66 eğitim veri kümesi ve %33 test veri kümesi olarak ikiye ayrılmıştır.

7.4 GRU ile Önerilen C/V/S Konuşma Bölütlerinin Tespitinin Uygulama Adımları

Bu tez çalışması, GRU tekrarlayan sinir ağları tabanında geliştirilmiş öğrenme modeli ile C/V/S konuşma bölgelerinin tespit edilmesi üzerine odaklanmıştır. Aynı zamanda, literatür çalışmalarında güncel ve yaygın bir kullanım alanına sahip olan derin sinir ağları mimarilerinden GRU tekrarlayan sinir ağı modelinin eğitim başarısına etkisi incelenmiştir.



Şekil 7.2 Önerilen Modelin Akış Diyagramı.

GRU ile önerilen mimarinin genel adımları Şekil 7.2'de gösterilmektedir. Bu mimaride eğitim veri kümesi, elle etiketlenmiş C/V/S konuşma bölütlerini içermektedir. Cinsiyet tanıma adımında; etiketli konuşma örneklerinden konuşmacının cinsiyeti belirlenmiştir. Çerçeveleme ve pencereleme adımında; konuşma örnekleri çerçeve olarak adlandırılan 20 ms'lik, 25 ms'lik, 30 ms'lik veya 35 ms'lik küçük zaman aralıklı bloklara bölünmüştür. Daha sonra çerçevelenmiş her bir konuşma işareti Hamming, Hanning veya Rectangular pencereleme fonksiyonlarından geçirilmiştir. Hibrit özellik çıkarımı adımında; pencerelenmiş her bir çerçevenin ayırt edici özellikleri "Energy, ZCR ve MFCC (EZMFCC)", "Energy, ZCR ve Delta MFCC (EZDMFCC)" veya "Energy, ZCR ve Delta Delta MFCC (EZDDMFCC)" özellik çıkarım parametrelerinin birlikte kullanılmasından oluşan hibrit özellik çıkarım tekniği ile elde edilmiştir. Hibrit özellik vektörleri adımında; çerçeve başına 15-, 28- veya 41- boyutlu hibrit özellik vektörleri üretilmiştir. GRU ile eğitim adımında, farklı uzunluktaki her bir çerçevenin (pencrenin) farklı boyutlardaki her bir hibrit özellik vektörü C/V/S örüntüsünü öğrenmek için GRU'ya girdi olarak verilmiştir. Elde edilen sonuçlar; GRU tabanında eğitilmiş hibrit özellik vektörlerinin test veri kümesi tabanında sınıflandırma işleminin gerçekleştirilmesi ile tamamlanmaktadır. Aşağıdaki bölümlerde, önerilen modelin uygulanmasına ilişkin detaylar açıklanmaktadır.

7.4.1 Cinsiyet Tanıma

Bu bölümde, konuşma işaretlerinden konuşmacının cinsiyetinin belirlenmesi için kısa süreli otokorelasyon ve ortalama büyüklük farkı fonksiyonları birlikte kullanılmıştır. Burada, erkek konuşmacıların kadın konuşmacılara kıyasla daha düşük temel frekansa sahip olması özelliğine dayanan bir eşik değeri belirlenmiştir [72].

Kısa süreli otokorelasyon, temel frekans değerini tahmin etmek için kullanılan en yaygın yöntemlerden biridir. Ayrık zaman işareti $x(n)$ için, otokorelasyon fonksiyonu Denklem 7.5 ile tanımlanmaktadır:

$$R_n(\tau) = lim_{N\to\infty} \frac{1}{2N+1} \sum_{n=-N}^{N} x(n) \cdot x(n+\tau) \tag{7.5}$$

Burada, τ gecikme, n, zaman indeksini, N bir çerçeve içerisindeki örneklerin toplam sayısını ifade etmektedir. Otokorelasyon fonksiyonu, bir işaretin kendisinin gecikmeli bir kopyası ile korelasyonunun bir ölçüsüdür. Konuşma işaretinde, kısa süreli otokorelasyon fonksiyonundaki ana tepe normalde perde periyoduna eşit bir gecikme ile gerçekleşir. Bu tepe, bu nedenle tespit edilir ve zaman konumu, giriş konuşmasının perde periyodunu vermektedir[72].

Her bir çerçeveye kısa süreli otokorelasyon işlemi uygulanır ve böylece her çerçeve için bu işlem hesaplandıktan sonra tüm çerçevelerin ortalama değeri hesaplanır.

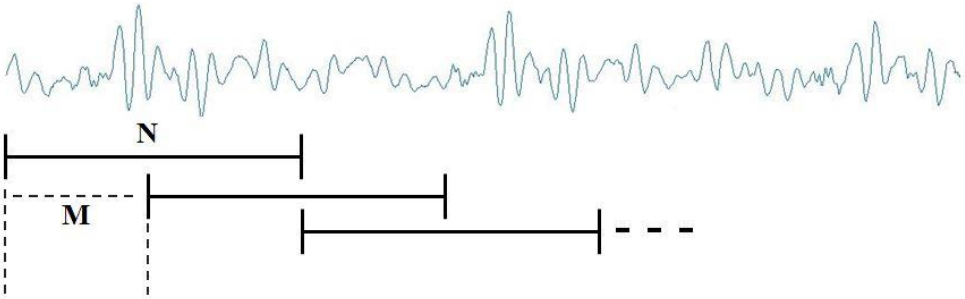
Ortalama Büyüklük Farkı Fonksiyonu (Ortalama Magnitude Difference Function, AMDF), geleneksel olarak kullanılan otokorelasyon tabanlı algoritmalardan biridir [72]. Bu algoritmada, geciken konuşma ile orijinali arasında farlılık sinyali oluşturulur ve her gecikme değerinde mutlak büyüklük alınır. Her bir çerçevede gerçekleştirilen AMDF Denklem 7.6 ile hesaplanmaktadır:

$$AMDF(t) = \frac{1}{L} \sum_{i=1}^{L} |s(i) - s(i-t)| \tag{7.6}$$

Burada, $s(i)$ giriş konuşmasının örneklerini ve L çerçeve uzunluğunu belirtmektedir. Konuşma işaretine ait temel frekans değerleri

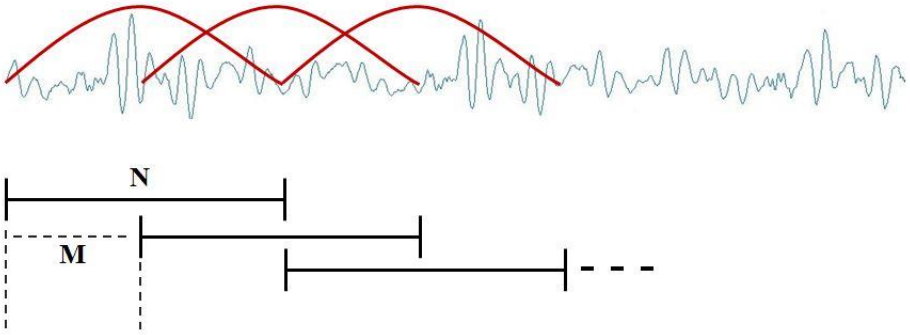
7.4.2 Çerçeveleme ve Pencereleme

Çerçeveleme, konuşma işaretinin küçük zaman diliminde işlenerek konuşma işaretinin daha kararlı durumundaki karakteristik özelliklerinin elde edilmesidir [73]. Şekil 7.3'e göre, verilen konuşma işareti sürekli *N* uzunluklu çerçevelere ayrılır ve ilk çerçeveden sonraki diğer bütün çerçeveler kendilerinden bir önceki çerçeveler ile *N-M* kadar örtüşür. Böylece, bir çerçeveden diğer bir çerçeveye geçiş *N-M* kadar yumuşatılmış olur ve M<N'dir. Bu tez çalışmasında, ses işaretlerinin ayırt edici özellikleri 20, 25, 30 ve 35 ms'lik N uzunluklu küçük zaman dilimleri arasında işlenmiştir. Burada M değeri, N/2 olarak kullanılmıştır.



Şekil 7.3 Bir Konuşma İşaretinin Çerçeveleme İşlemi.

Çerçevelenmiş her bir ses işaretinde, Hamming, Hanning, veya Rectangular pencereleme teknikleri kullanılmıştır. Şekil 7.4'te Hamming pencereleme tekniği gösterilmiştir.



Şekil 7.4 Bir Konuşma İşaretinin Pencereleme işlemi.

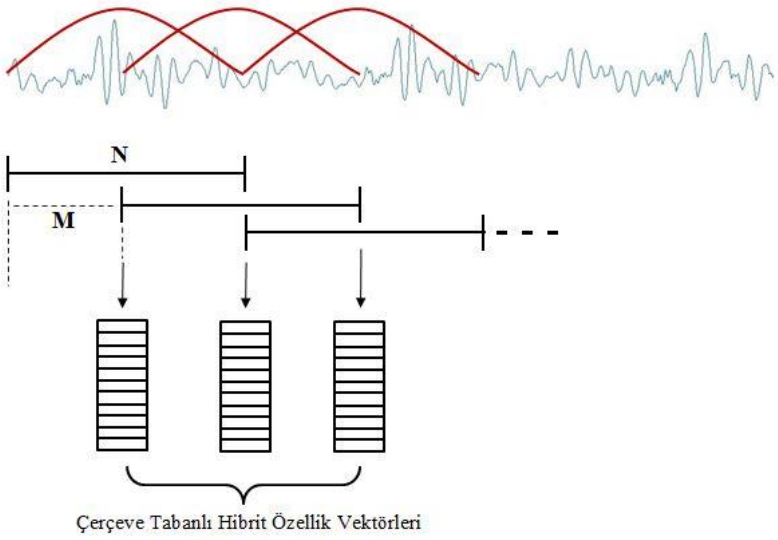
Bu aşamada, farklı uzunluktaki pencerelerin (çerçevelerin) farklı pencereleme teknikleri ile kullanılmasından elde edilen hibrit özelliklerin C/V/S konuşma bölütlerinin tespit edilmesindeki performans sonucuna katkısı gözlenmiştir.

7.4.3 Hibrit Özellik Çıkarımı

Bu çalışmada, C/V/S konuşma bölütlerinin tespitinin başarı oranını arttırmak için, faklı özellik çıkarım parametrelerinin birlikte kullanıldığı bir hibrit özellik çıkarım yöntemi kullanılmıştır. Bu yöntemde ilk olarak, C/V/S konuşma bölütlerini ayırt edebilen enerji ve ZCR özellik çıkarım parametrelerinin beraber değerlendirildiği bir hibrit özellik çıkarım yöntemi kullanılmıştır. Ancak, C/V/S konuşma bölütlerinin tespitinde daha yüksek bir doğruluk performansının elde edilmesi için liteartür çalışmalarında genel kabul görmüş ve başarılı sonuçlar elde edilmiş MFCC özellik çıkarım parametresine dayalı daha etkin bir hibrit özellik çıkarım yöntemi temel alınmıştır. Bu bağlamda, bu çalışma kapsamında, Enerji, ZCR, MFCC (EZMFCC); Enerji, ZCR, Delta MFCC (EZDMFCC) ve Enerji, ZCR, Delta Delta MFCC (EZCRDDMFCC)'den oluşan 3 farklı hibrit özellik çıkarım yöntemi kullanılmıştır. Aynı zamanda sözü edilen bu hibrit özellik çıkarım yöntemleri sırasıyla Hamming, Hanning veya Rectangular pencereleme fonksiyonları ile 20, 25, 30 veya 35 ms pencere (çerçeve) sürelerinden oluşan paramerteler ile birlikte ele alınmıştır. Tüm bu parametreler C/V/S konuşma bölütlerinin ayırt edici özelliklerinin elde edilmesinde kullanılmıştır. Hibrit özellik çıkarım aşamasında, C/V/S konuşma bölütlerinin tespitinde, farklı parametre değerleri (pencere uzunluğu, pencere teknikleri ve hibrit özellik çıkarım yöntemleri) girilerek farklı özellikler elde edilmiş ve bu özelliklerin başarıma etkisi incelemiştir.

7.4.4 Hibrit Özellik Vektörlerinin Elde Edilmesi

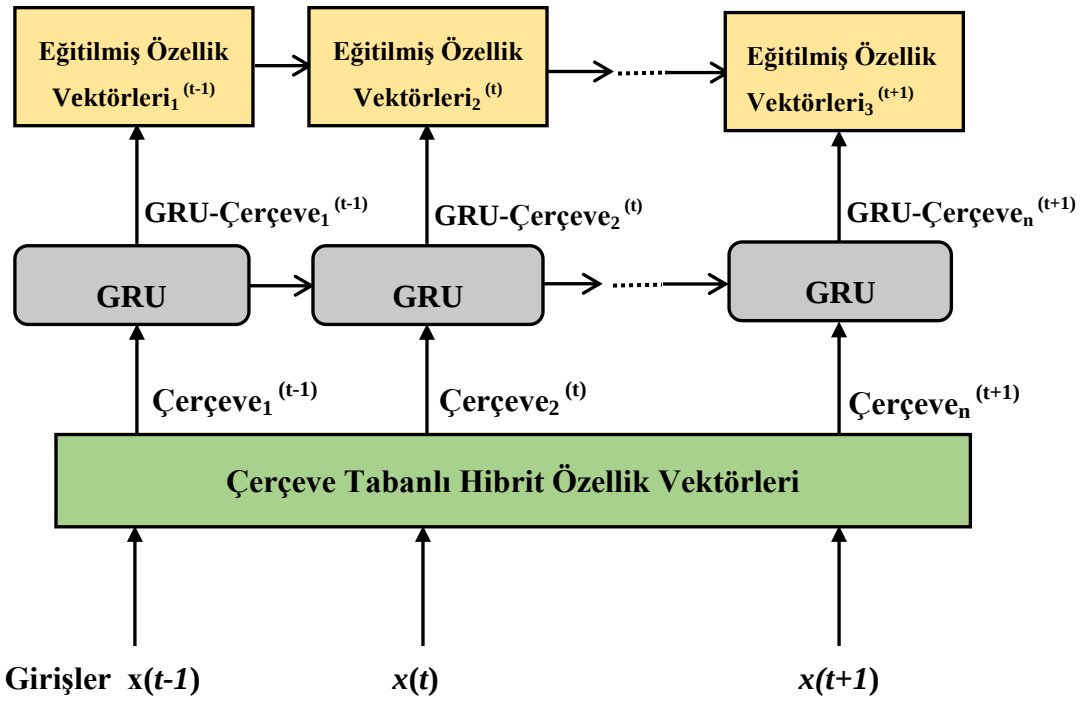
Hibrit özellik çıkarımı aşamasında, ses işaretinin içerisindeki konuşma içeren bölütlerin kategorisini (C/V/S) tespit etmek amacıyla farklı boyutlarda hibrit özellik vektörleri üretilmiştir. Veri kümesi tabanında girdi işareti, farklı boyutlardaki çerçevelere bölündüğü zaman çok sayıda hibrit özellik vektörleri elde edilmektedir. Şekil 7.5'te çerçeve tabanlı hibrit özellik vektörlerinin gösterimi yer almaktadır.



Şekil 7.5 Çerçeve Tabanlı Hibrit Özellik Vektörlerinin Elde Edilmesi.

7.4.5 GRU ile Model Oluşturma ve Eğitim

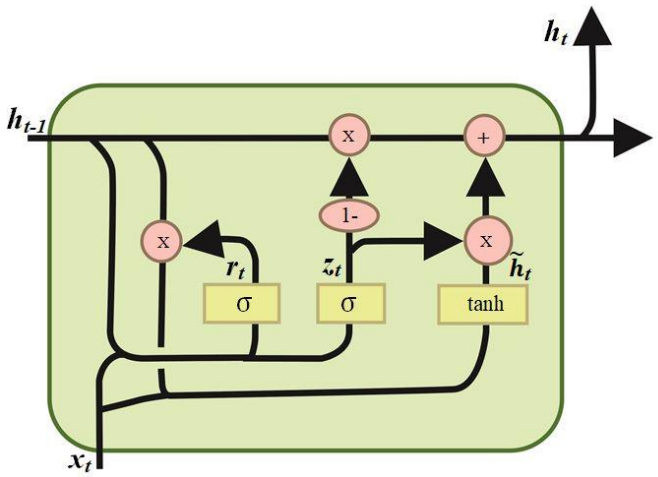
Eğitim seti tabanında üretilen hibrit özellik vektörlerinin GRU modelinde kullanılması, modelin eğitim aşamasını oluşturmaktadır. GRU ile önerilen modelin akış diyagramı Şekil 7.6'da gösterilmiştir.



Şekil 7.6 GRU ile Önerilen Modelin Akış Diyagramı.

Şekil 7.6'da, GRU bloklarının giriş dizisi $x = (x_1; : : : ; x_t)$ ile ifade edilmektedir. Eğitim seti içerisinde yer alan her bir çerçevenin hibrit özellik vektörleri x_t ile ifade edilmektedir.

Şekil 7.7'de GRU biriminin hücre yapısı gösterilmiştir. x_t, t anındaki mevcut girdi değerini, h_{t-1} önceki gizli durum değerini ve h_t mevcut zamandaki gizli durumu ve çıktıyı ifade etmektedir. r_t ve z_t, t anındaki sıfırlama kapısını (reset gate) ve güncelleme kapısını (update gate) belirtmektedir.



Şekil 7.7 GRU Hücre Mimarisi [73].

Şekil 7.8, C/V/S konuşma bölütlerinin GRU blokları tabanında modellenmesini göstermektedir. Her bir t anında j. GRU bloğunun aktivasyon sonucunda ürettiği çıktı ${h_t}^j$ olarak belirtilmektedir. GRU'nun t zaman adımı için aktivasyonu ${h_t}^j$, önceki zaman adımı aktivasyonu ${h_{t-1}}^j$ ile aday aktivasyonu ${\tilde{h}_t}^j$ arasındaki doğrusal tahmini Denklem 7.8 ve 7.9'da temsil edilmektedir.

$$ {\tilde{h}_t}^j = tanh\left(W \cdot \left[{r_t}^j * {h_{t-1}}^j, x_t\right]\right), \tag{7.8} $$

$$ {h_t}^j = \left(1 - {z_t}^j\right) * {h_{t-1}}^j + {z_t}^j * {\widetilde{h}_t}^j \tag{7.9} $$

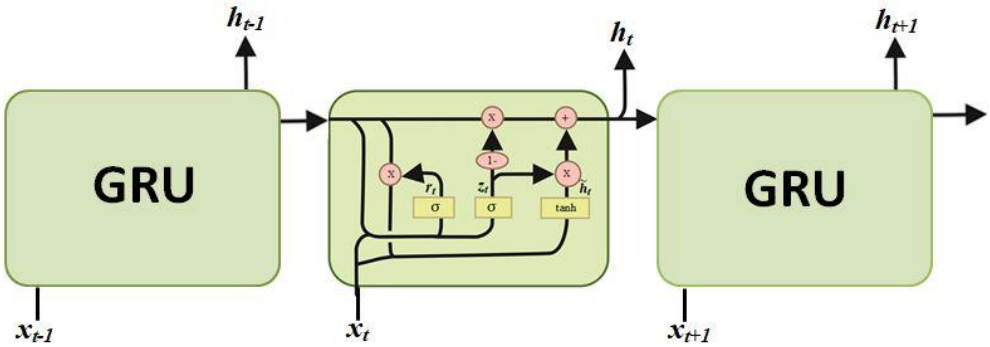
Burada, ${r_t}^j$ resetleme ve ${z_t}^j$ güncelleme kapı vektörlerini simgelemektedir. "*" işareti iki vektör arasındaki birimsel çarpımı ifade etmektedir.

GRU, geçmiş zaman adımından bir sonraki zaman adımına aktaracak bilgileri güncelleme kapısı (z_t) ile ve modelden geçmiş bilgilerin ne kadarının geçip ne kadarının unutulacağına sıfırlama kapısı (r_t) ile karar vermektedir. Çıktı olarak hangi verilerin aktarılması gerektiğine karar veren kapı vektörleri Denklem 7.10 ve 7.11 ile hesaplanmaktadır.

$$z_t{}^j = \sigma(W_Z x_t + U_Z h_{t-1)})^j\,, \tag{7.10}$$

$$r_t{}^j = \sigma(W_r x_t + U_r h_{t-1})^j \tag{7.11}$$

Burada, W_z ve U_Z güncelleme kapısı için ve W_r *ve* U_r sıfırlama kapısı için ağırlık vektörlerini belirtmektedir. Hangi bilgilerin güncellenip resetlendiği ile ilgili karar σ (sigmoid) aktivasyon fonksiyonu aracılığı ile sağlanmaktadır.



Şekil 7.8 GRU Blokları ile C/V/S Konuşma Bölütlerinin Modellenmesi.

7.4.6 Test

Test aşamasında, GRU ile eğitilmiş hibrit özellik vektörlerinin test veri kümesi tabanında sınıflandırılmaları CNN, MLP, NB, SVM, RF ve k-NN sınıflandırıcıları ile gerçekleştirilmiştir.

8
Deneysel Sonuçlar ve Performans Analizi

Bu tez kapsamında Kürtçe konuşma işaretlerinden oluşan özgün bir veri kümesi kullanılmıştır. Elde edilen deneysel sonuçlar ve performans analizi, veri kümesinin % 66 eğitim seti ve % 33 test seti olarak bölünmesiyle elde edilmiştir. Bu çalışmada üç farklı hibrit özellik çıkarım yöntemi kullanılmıştır. İlk kullanılan hibrit özellik çıkarım yöntemi EZMFCC, ikinci kullanılan hibrit özellik çıkarım yöntemi EZDMFCC ve son kullanılan hibrit özellik çıkarım yöntemi EZDDMFCC'dir. Yukarıda belirtilen hibrit özellik çıkarım yöntemleri, Hamming, Hanning veya Rectangular pencereleme tekniklerinin 20, 25, 30 veya 35 ms pencere boyutları için elde edilmiştir. GRU ile eğitilmiş hibrit özellik vektörleri C/V/S konuşma bölütlerinin tespitini amacıyla farklı sınıflandırıcı yöntemleri ile test edilmiştir. Test sonuçlarının elde edilmesinde CNN, MLP, NB, SVM, RF ve k-NN sınıflandırıcıları kullanılmıştır. k-NN sınıflandırıcısının '*k*' parametresi için 3 değeri kullanılmıştır. MLP sınıflandırıcısının performans değeri, gizli katman parametre değerinin 30 alınması ile elde edilmiştir. CNN için, 20 ve 100 filtrelerden oluşan iki konvolüsyonel katmanlı, süzgeç boyutu 5, adım sayısı 2 ve aktivasyon fonksiyonu olarak ReLU kullanılmıştır. 64 birimli iki katmanlı, 0.5 bırakma oranı (dropout), 55 eğitim tur sayısı (epoch), 32 batch boyutu (size), öğrenim oranı (learning rate) 0,001, aktivasyon fonksiyonu ReLU ve Adam optimizasyon parametreleri ile öğrenim modeli geliştirilmiştir. Girdi ile çıktı katmanının ilgili nöronu arasında kayıp fonksiyon olarak Ortalama Hata Kareleri (mean squared error - MSE) ölçütü uygulanmıştır. Bu çalışmada önerilen modelin aşırı öğrenme (Overfit) problemini önlemek için her bir epoch periyodunda eğitim örnekleri rastgele karıştırılmıştır. Ayrıca modele doğruluk oranı önceden belirlenen miktarda tekrarlandığında erken durdurma kullanılmıştır. Bu yaklaşımlar modelin aşırı öğrenmesini azaltmada önemli derecede etkili olmuştur. Bu değişkenler çok sayıda denemenin sonucunda en iyi sonuçları verenler arasından deneme yanılma yöntemiyle seçilmiştir.

Tablo 8.1-8.3'de erkek konuşmacılar için ve Tablo 8.4–8.6'da kadın konuşmacılar için önerilen GRU tabanlı eğitim modelinin CNN, MLP, NB, SVM, RF ve k-NN sınıflandırıcı yöntemleri ile elde edilen doğruluk performans sonuçları yer almaktadır. Doğruluk performanslarının elde edilmesinde Weka programı kullanılmıştır. Doğru tespit edilen C/V/S konuşma bölütlerinin sayısı performans ölçüsü olarak kabul edilmiş ve doğruluk fonksiyonu Denklem 8.1'de belirtilen formül ile hesaplanmıştır:

$$Doğruluk = \frac{TP + TN}{TP + FP + FN + TN} \tag{8.1}$$

Burada, TP, TN, FP ve FN sırasıyla gerçek pozitif, gerçek negatif, yanlış pozitif ve yanlış negatif sayısını temsil etmektedir.

8.1 GRU Tabanlı Eğitim Modeli ile Hibrit Özellik Çıkarım Yöntemlerinin Analiz Sonuçları

Erkek ve kadın konuşmacılar için Tablo 8.1-8.6'da sunulan sınıflandırıcıların doğruluk performans sonuçlarına göre, EZDDMFCC hibrit özellik çıkarım yönteminin tüm pencereleme teknikleri, pencere (çerçeve) boyutları ve sınıflandırıcı yöntemlerinde en yüksek doğruluk performans sonucunun elde edildiği görülmüştür. Bu durum, EZDDMFCC'de mevcut olan MFCC özellik çıkarım yönteminin çift türevinin bir sonucu olarak, yeni özelliklerin sayısı nedeniyle olabilir. EZDDMFCC hibrit özellik vektörleri EZMFCC ve EZDMFCC hibrit özellik vektörlerine kıyasla daha büyük boyuttadır. Ancak, Tablo 8.8'de görüldüğü gibi EZDDMFCC ile doğruluk performans sonuçlarının elde edilmesi için geçen hesaplama süresi daha uzun olmaktadır.

8.2 GRU Tabanlı Eğitim Modeli ile Pencere Uzunluklarının Analiz Sonuçları

20 ms, 25 ms, 30 ms ve 35 ms pencere uzunluklarının en yüksek sınıflandırıcı doğruluk performansları erkek konuşmacılar için sırasıyla %97.60, %96.38, %96.60 ve %97.12 iken, kadın konuşmacılar için sırasıyla %95.30, %95.09, %95.03 ve %95.63 olmaktadır. Elde edilen sonuçlara göre, pencere uzunluklarının sınıflandırıcı doğruluk performansı üzerindeki etkisi önemsizdir. Ancak, 20 ms'lik pencere uzunluklarının, diğer pencere uzunluklarına kıyasla göreceli olarak daha başarılı sonuçlar verdiği gözlenmiştir.

8.3 GRU Tabanlı Eğitim Modeli ile Pencereleme Tekniklerinin Analiz Sonuçları

Hamming, Hanning ve Rectangular pencereleme tekniklerinin türüne dayanan en yüksek sınıflandırıcı doğruluk performansları erkek konuşmacılar için sırasıyla %97.25, %97.04 ve %97.60 iken, kadın konuşmacılar için sırasıyla için %95.10, %95.63 ve %95.30 olmaktadır. Elde edilen sonuçlar, üç farklı pencereleme türü arasında dikkate değer bir doğruluk performans farkı oluşturmadığını ortaya koymuştur.

8.4 GRU Tabanlı Eğitim Modeli ile Sınıflandırıcı Yöntemlerinin Analiz Sonuçları

Doğruluk performans sonuçlarına bakıldığında, CNN sınıflandırıcısının diğer sınıflandırıcılara göre daha yüksek doğruluk performansını (erkek konuşmacılar için %97.60 ve kadın konuşmacılar için %95.63) verdiği görülmüştür.

Tablo 8.1 Erkek Konuşmacılar için GRU-EZMFCC Tabanlı Sınıflandırıcı Doğruluk Performans Sonuçları.

Pencereleme		**Sınıflandırma Yöntemleri ve Performans Değerleri (%)**					
Teknikleri	**Süresi (ms)**	**CNN**	**MLP**	**NB**	**SVM**	**RF**	**k-NN**
Hamming	20	92.60	84.57	74.37	83.14	85.56	84.94
	25	92.11	84.19	74.63	82.91	86.36	85.42
	30	92.04	84.54	73.80	82.93	86.41	84.08
	35	92.59	84.12	74.66	83.36	86.37	84.43

Tablo 8.1 Erkek Konuşmacılar için GRU-EZMFCC Tabanlı Sınıflandırıcı Doğruluk Performans Sonuçları (devamı)

Hanning	20	92.59	84.40	74.51	83.16	85.41	84.50
	25	92.40	84.38	74.61	83.14	86.10	84.89
	30	92.18	85.11	74.48	83.20	86.36	84.72
	35	92.80	85.38	74.51	83.38	87.10	83.42
Rectangular	20	92.05	85.24	73.80	83.30	85.63	84.49
	25	92.24	85.26	74.70	83.38	86.25	83.70
	30	92.44	85.41	74.01	83.66	86.43	83.64
	35	92.85	84.32	74.12	83.33	86.56	83.42

Tablo 8.2 Erkek Konuşmacılar için GRU-EZDMFCC Tabanlı Sınıflandırıcı Doğruluk Performans Sonuçları.

Pencereleme		**Sınıflandırma Yöntemleri ve Performans Değerleri (%)**					
Teknikleri	**Süresi (ms)**	**CNN**	**MLP**	**NB**	**SVM**	**RF**	**k-NN**
Hamming	20	93.98	88.52	77.76	85.20	89.23	86.71
	25	94.71	88.37	78.49	85.44	88.28	85.01
	30	94.44	88.51	77.24	86.15	88.38	84.58
	35	94.00	87.65	77.71	86.03	89.14	84.17

Tablo 8.2 Erkek Konuşmacılar için GRU-EZDMFCC Tabanlı Sınıflandırıcı Doğruluk Performans Sonuçları (devamı)

Hanning	20	93.33	88.08	77.75	86.39	88.35	86.19
	25	94.00	87.72	78.22	86.30	88.89	87.03
	30	94.55	88.63	77.29	85.01	88.09	86.79
	35	94.21	87.70	77.01	85.47	88.23	86.29
Rectangular	20	94.61	86.97	77.35	85.55	88.82	85.18
	25	93.74	88.05	77.73	86.19	88.43	84.54
	30	93.79	88.13	77.08	84.62	87.62	84.61
	35	94.18	87.23	77.55	84.03	88.12	85.31

Tablo 8.3 Erkek Konuşmacılar için GRU-EZDDMFCC Tabanlı Sınıflandırıcı Doğruluk Performans Sonuçları.

Pencereleme		**Sınıflandırma Yöntemleri ve Performans Değerleri (%)**					
Teknikleri	**Süresi (ms)**	**CNN**	**MLP**	**NB**	**SVM**	**RF**	**k-NN**
Hamming	20	97.25	91.14	82.20	89.06	93.53	90.38
	25	95.77	92.05	82.28	90.32	92.14	90.44
	30	96.53	92.36	81.30	89.80	93.19	90.87
	35	97.12	91.28	82.45	89.29	92.70	90.10

Tablo 8.3 Erkek Konuşmacılar için GRU-EZDDMFCC Tabanlı Sınıflandırıcı Doğruluk Performans Sonuçları (devamı)

Hanning	20	97.04	92.65	81.80	89.81	93.74	90.49
	25	96.18	92.76	81.65	89.63	93.55	91.52
	30	96.60	92.13	80.28	89.04	92.00	90.24
	35	96.69	91.05	80.30	90.13	91.87	89.64
Rectangular	20	97.60	91.94	82.71	88.88	92.32	91.12
	25	96.38	92.42	81.73	89.37	93.15	90.23
	30	96.22	92.08	81.45	88.17	92.03	90.61
	35	95.98	92.32	81.38	88.85	92.32	90.33

Tablo 8.4 Kadın Konuşmacılar için GRU-EZMFCC Tabanlı Sınıflandırıcı Doğruluk Performans Sonuçları.

Pencereleme		**Sınıflandırma Yöntemleri ve Performans Değerleri (%)**					
Teknikleri	**Süresi (ms)**	**CNN**	**MLP**	**NB**	**SVM**	**RF**	**k-NN**
Hamming	20	91.61	83.86	71.01	81.57	85.08	84.35
	25	91.11	82.34	72.11	81.43	84.97	83.22
	30	91.20	83.18	72.76	80.91	84.13	84.43
	35	91.14	82.12	72.93	80.79	84.11	83.07

Tablo 8.4 Kadın Konuşmacılar için GRU-EZMFCC Tabanlı Sınıflandırıcı Doğruluk Performans Sonuçları (devamı)

Hanning	20	91.00	82.10	71.58	80.13	84.76	84.32
	25	91.34	82.70	71.19	80.75	86.30	85.14
	30	91.66	82.03	72.77	81.21	85.14	84.22
	35	91.31	82.10	71.91	80.97	86.02	83.36
Rectangular	20	91.22	82.16	71.02	80.82	86.12	85.54
	25	91.01	82.30	71.46	80.63	85.00	85.47
	30	91.41	82.57	72.18	81.02	84.22	84.72
	35	91.60	83.70	72.29	80.94	84.16	83.71

Tablo 8.5 Kadın Konuşmacılar için GRU-EZDMFCC Tabanlı Sınıflandırıcı Doğruluk Performans Sonuçları.

Pencereleme		**Sınıflandırma Yöntemleri ve Performans Değerleri (%)**					
Teknikleri	**Süresi (ms)**	**CNN**	**MLP**	**NB**	**SVM**	**RF**	**k-NN**
Hamming	20	93.14	85.79	73.28	84.54	88.29	87.40
	25	93.37	87.31	73.44	85.63	88.41	85.64
	30	92.15	86.76	74.12	84.52	88.70	86.81
	35	93.50	87.61	74.59	84.25	89.57	86.26

Tablo 8.5 Kadın Konuşmacılar için GRU-EZDMFCC Tabanlı Sınıflandırıcı Doğruluk Performans Sonuçları (devamı)

Hanning	20	93.01	85.28	74.25	84.31	90.45	87.52
	25	93.30	86.39	74.02	84.29	88.07	86.89
	30	93.08	87.20	73.20	84.51	88.42	86.18
	35	92.51	87.28	73.57	84.44	89.90	86.87
Rectangular	20	92.26	87.11	74.18	84.25	88.57	87.83
	25	93.18	86.04	74.15	84.34	90.20	87.63
	30	93.88	86.57	74.82	84.56	89.27	87.60
	35	93.52	86.13	74.64	85.32	88.82	86.47

Tablo 8.6 Kadın Konuşmacılar için GRU-EZDDMFCC Tabanlı Sınıflandırıcı Doğruluk Performans Sonuçları.

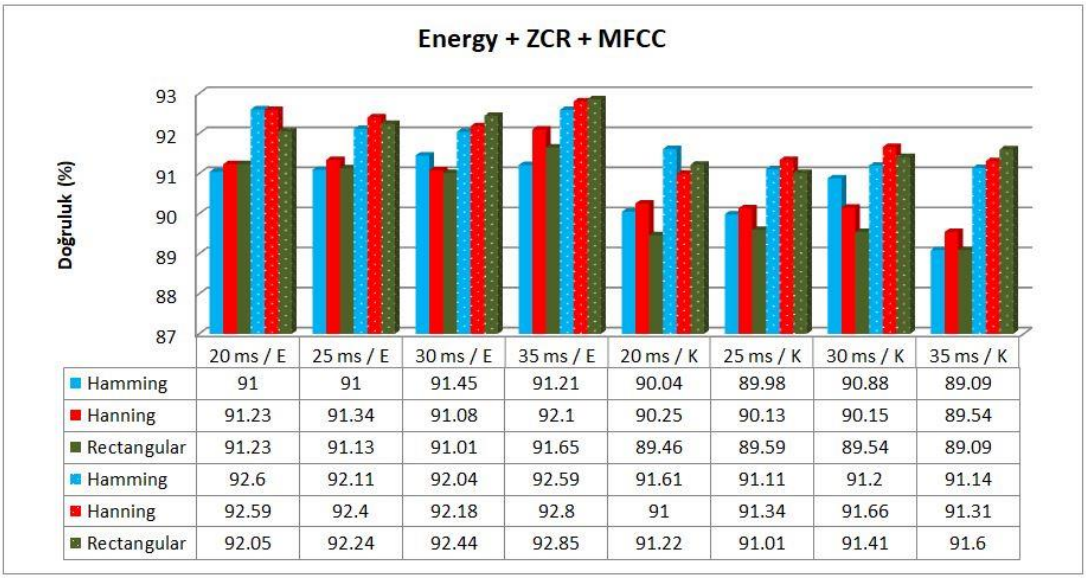
Pencereleme		**Sınıflandırma Yöntemleri ve Performans Değerleri (%)**					
Teknikleri	**Süresi (ms)**	**CNN**	**MLP**	**NB**	**SVM**	**RF**	**k-NN**
Hamming	20	95.10	91.30	75.51	86.29	92.20	91.16
	25	95.09	90.61	78.84	86.17	92.36	91.27
	30	95.03	90.87	77.21	86.94	91.44	91.35
	35	94.67	90.12	77.54	86.75	91.78	91.41

Tablo 8.6 Kadın Konuşmacılar için GRU-EZDDMFCC Tabanlı Sınıflandırıcı Doğruluk Performans Sonuçları (devamı)

Hanning	20	94.82	89.48	78.31	86.01	92.12	91.64
	25	94.31	91.85	76.33	87.20	91.05	92.30
	30	94.93	91.70	77.82	86.76	92.80	91.54
	35	95.63	90.62	77.77	86.10	92.44	91.30
Rectangular	20	95.30	91.47	78.60	87.27	92.30	91.30
	25	94.16	90.56	79.80	87.05	92.49	92.20
	30	94.14	91.57	78.01	87.37	91.33	91.73
	35	95.05	91.13	77.12	87.53	92.70	92.17

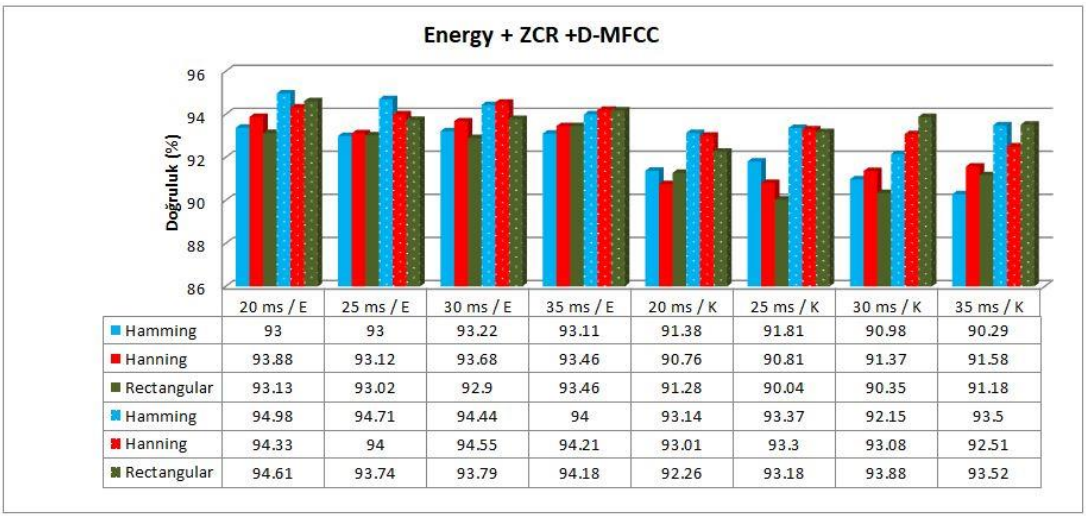
Şekil 8.1-8.3'te, GRU'suz ve GRU'lu (önerilen) eğitim modelinin CNN sınıflandırıcı yöntemi ile elde edilmiş doğruluk performans sonuçları gösterilmektedir. CNN sınıflandırıcı yöntemleri, diğer sınıflandırıcı yöntemlerine göre, nispeten daha yüksek bir performans sağladığından CNN sınıflandırıcı yöntemi örneklendirilmiştir. CNN sınıflandırıcı yöntemi ile elde edilen sonuçlar incelendiğinde, hem erkek hem de kadın konuşmacılar için en yüksek doğruluk performans sonuçlarının önerilen GRU'lu tabanlı eğitim modeli ile ve EZDDMFCC hibrit özellik çıkarım yönteminin kullanılması ile elde edildiği görülmüştür. Erkek konuşmacılar için 20 ms uzlunluklu Rectangular pencereleme yöntemi; kadın konuşmacılar için 35 ms uzunluklu Hanning pencereleme yöntemi ile en yüksek doğruluk performans sonuçları elde edilmiştir (Şekil 8.3). Erkek ve kadın konuşmacılar için önerilen GRU tabanlı eğitim modeli ile C/V/S konuşma bölütlerinin tespiti için saptanan en uygun özellik parametre seti Tablo 8.7'de özetlenmiştir. Erkek konuşmacıların performans sonuçlarının kadın konuşmacıların performans sonuçlarına göre daha yüksek doğruluk performansı elde ettiği gözlenmiştir. Performanslardaki farkın iki sınıftaki konuşmacı sayısının dengesizliğinden kaynaklanmadığı değerlendirilmektedir. Bu durum kadın konuşmacıların konuşma

stilindeki farklılık nedeniyle ortaya çıkabilmektedir. Sonuç olarak, yeni özelliklerin sayısı arttıkça, kadın ve erkek konuşmacılar için başarı oranının arttığı gözlenmiştir.



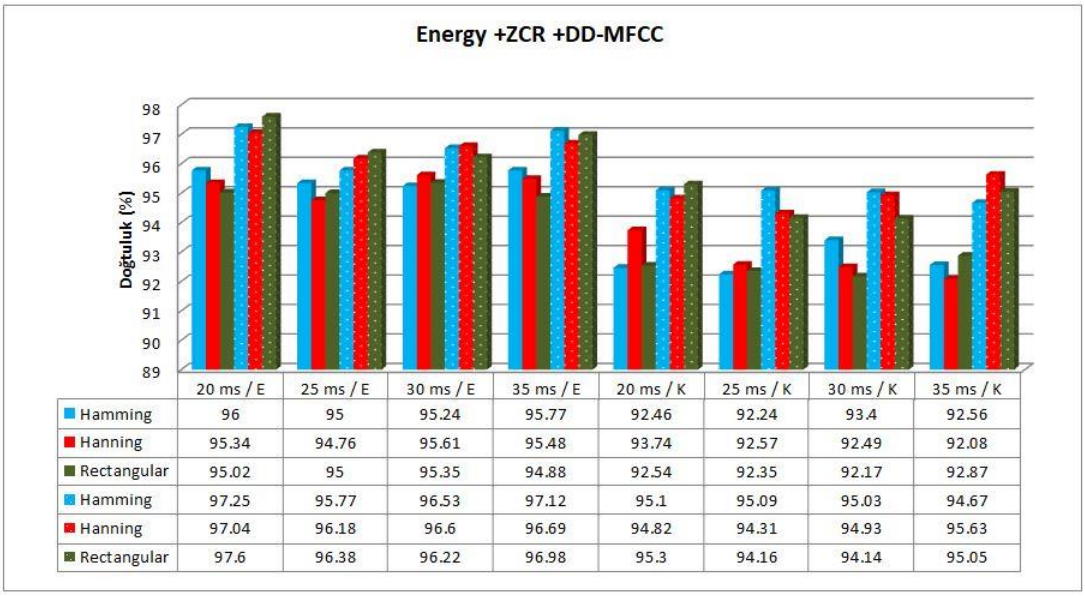
	20 ms / E	25 ms / E	30 ms / E	35 ms / E	20 ms / K	25 ms / K	30 ms / K	35 ms / K
Hamming	91	91	91.45	91.21	90.04	89.98	90.88	89.09
Hanning	91.23	91.34	91.08	92.1	90.25	90.13	90.15	89.54
Rectangular	91.23	91.13	91.01	91.65	89.46	89.59	89.54	89.09
Hamming	92.6	92.11	92.04	92.59	91.61	91.11	91.2	91.14
Hanning	92.59	92.4	92.18	92.8	91	91.34	91.66	91.31
Rectangular	92.05	92.24	92.44	92.85	91.22	91.01	91.41	91.6

Şekil 8.1 Erkek (E) ve Kadın (K) Konuşmacılar için GRU'suz (noktasız çubuklar) ve Önerilen GRU Tabanlı Eğitim Modeli (noktalı çubuklar) ile EZMFCC'nin CNN Sınıflandırıcı Performans Doğruluğu.



	20 ms / E	25 ms / E	30 ms / E	35 ms / E	20 ms / K	25 ms / K	30 ms / K	35 ms / K
Hamming	93	93	93.22	93.11	91.38	91.81	90.98	90.29
Hanning	93.88	93.12	93.68	93.46	90.76	90.81	91.37	91.58
Rectangular	93.13	93.02	92.9	93.46	91.28	90.04	90.35	91.18
Hamming	94.98	94.71	94.44	94	93.14	93.37	92.15	93.5
Hanning	94.33	94	94.55	94.21	93.01	93.3	93.08	92.51
Rectangular	94.61	93.74	93.79	94.18	92.26	93.18	93.88	93.52

Şekil 8.2 Erkek (E) ve Kadın (K) Konuşmacılar için GRU'suz (noktasız çubuklar) Önerilen GRU Tabanlı Eğitim Modeli (noktalı çubuklar) ile EZDMFCC'nin CNN Sınıflandırıcı Performans Doğruluğu.



Şekil 8.3 Erkek (E) ve Kadın (K) Konuşmacılar için GRU'suz (noktasız çubuklar) ve Önerilen GRU Tabanlı Eğitim Modeli (noktalı çubuklar) ile EZDDMFCC'nin CNN Sınıflandırıcı Performans Doğruluğu.

Tablo 8.7 Önerilen GRU Tabanlı Eğitim Modeli ile C/V/S Konuşma Bölütlerinin Tespiti için Saptanan En Uygun Özellik Parametre Seti

Özellik Parametre Seti	**Konuşmacıların Cinsiyeti**	
	Erkek	**Kadın**
Çerçeve Boyutu	20 ms	35 ms
Pencereleme Tekniği	Rectangular	Hanning
Hibrit Özellik Çıkarım Yöntemi	EZDDMFCC	
Sınıflandırma Metodu	CNN	

Uygulamalar, Intel Core i3-2367M işlemci, 4 GB bellek boyutu ve 1.40 GHz saat hızına sahip bir PC üzerinde gerçekleştirilmiştir. Tablo 8.8'de, erkek konuşmacıların önerilen GRU tabanlı eğitim modeline dayalı hibrit özellik çıkarım yöntemlerinin, 20 ms pencere (çerçeve) uzunluklu Hamming pencereleme yöntemi ile gerçekleştirilmiş uygulamaların

hesaplama süresi (karmaşıklığı) yer almaktadır. Bu bağlamda; hesaplama karmaşıklığı/doğruluk performans sonucu aşağıdaki gibi özetlenebilir:

- EZMFCC < EZDMFCC
- EZDMFCC < EZDDMFCC

Tablo 8.8 Erkek Konuşmacıların Önerilen GRU Tabanlı Eğitim Modeline Dayalı 20 ms Uzunluklu Hamming Pencere ile Elde Edilen Hesaplama Karmaşıklığı (dakika)

Sınıflandırma Metotları	**EZMFCC**	**EZDMFCC**	**EZDDMFCC**
CNN	0.42497	0.72882	1.12760
MLP	0.42075	0.72102	1.09690
NB	0.42005	0.72015	1.08012
SVM	0.42013	0.72025	1.08052
RF	0.42083	0.72127	1.08147
k-NN	0.42000	0.72000	1.08000

Bu tez çalışmasında, C/V/S konuşma bölütlerinin tespitinin doğruluk sınıflandırıcı performasını hesaplayabilen en uygun özellik parametre setinin belirlenmesi için kullanılan çeşitli özelliklerin parametre kombinasyonları ve modelleri Tablo 8.9'da özetlenmiştir.

Tablo 8.9 C/V/S Konuşma Bölütlerinin Tepiti için Kullanılan Özellik Parametre Seti ve Modelleri

Önerilen GRU Tabanlı Eğitim Modeli													
Erkek ve Kadın													
Hibrit Özellikler	**SM**	**Hamming**				**Hanning**				**Rectangular**			
		20	**25**	**30**	**35**	**20**	**25**	**30**	**35**	**20**	**25**	**30**	**35**
EZMFCC	CNN	√	√	√	√	√	√	√	√	√	√	√	√
	MLP	√	√	√	√	√	√	√	√	√	√	√	√
	NB	√	√	√	√	√	√	√	√	√	√	√	√
	SVM	√	√	√	√	√	√	√	√	√	√	√	√
	RF	√	√	√	√	√	√	√	√	√	√	√	√
	k-NN	√	√	√	√	√	√	√	√	√	√	√	√
EZDMFCC	CNN	√	√	√	√	√	√	√	√	√	√	√	√
	MLP	√	√	√	√	√	√	√	√	√	√	√	√
	NB	√	√	√	√	√	√	√	√	√	√	√	√
	SVM	√	√	√	√	√	√	√	√	√	√	√	√
	RF	√	√	√	√	√	√	√	√	√	√	√	√
	k-NN	√	√	√	√	√	√	√	√	√	√	√	√
EZDDMFCC	CNN	√	√	√	√	√	√	√	√	√	√	√	√
	MLP	√	√	√	√	√	√	√	√	√	√	√	√

Tablo 8.9 C/V/S Konuşma Tepiti için Kullanılan Özellik Parametre Seti ve Modelleri (devamı)

EZDDMFCC	NB	√	√	√	√	√	√	√	√	√	√	√	√
	SVM	√	√	√	√	√	√	√	√	√	√	√	√
	RF	√	√	√	√	√	√	√	√	√	√	√	√
	k-NN	√	√	√	√	√	√	√	√	√	√	√	√
GRU'suz Eğitim Modeli													
Erkek ve Kadın													
Hibrit Özellikler	**SM**	**Hamming**				**Hanning**				**Rectangular**			
		20	**25**	**30**	**35**	**20**	**25**	**30**	**35**	**20**	**25**	**30**	**35**
EZMFCC	CNN	√	√	√	√	√	√	√	√	√	√	√	√
EZDMFCC		√	√	√	√	√	√	√	√	√	√	√	√
EZDDMFCC		√	√	√	√	√	√	√	√	√	√	√	√

9
Sonuç ve Öneriler

Geçmişten günümüze konuşma bölütlerinin tespit edilmesi ile ilgili çeşitli çalışmalar yapılmıştır. Bu çalışmalar, özellikle çeşitli üniversiteler ve kurumlar tarafından farklı dil ve lehçelerde oluşturulan hazır veri kümeleri üzerinde gerçekleştirilmiştir. Bu tez çalışmasında, Kürtçe dilinde özgün bir veri kümesi oluşturulmuştur. Oluşturulan veri kümesi içerisindeki, sürekli konuşma ifadelerinden GRU tekrarlayan sinir ağlarına dayalı C/V/S konuşma bölütlerinin otomatik olarak tespit edilmesi ele alınmıştır. Bu amaçla, Enerji +ZCR +MFCC, Enerji +ZCR +D-MFCC ve Enerji +ZCR +DD-MFCC hibrit özellik çıkarım yöntemleri; 20 ms, 25 ms, 30 ms ve 35 ms pencere (çerçeve) uzunlukları; Hamming, Hanning ve Rectangular penecereleme teknikleri önemli özelliklerin elde edilmesi için parametre olarak kullanılmıştır. Daha sonra, bu parametrelerden elde edilen hibrit özellik vektörleri CNN, MLP, NB, SVM, RF ve k-NN sınıflandırıcı yöntemleri ile test edilerek C/V/S konuşma bölütlerinin tespit edilmesindeki başarısına etkisi gözlenmiştir. Böylece, bu çalışmanın sonunda, GRU tekrarlayan sinir ağlarına dayalı Kürtçe C/V/S konuşma bölütlerinin tespiti için etkin rol alan parametreler belirlenmiştir.

Elde edilen bulgular, GRU tekararlayan sinir ağlarının Kürtçe dilinde konuşma tespiti için en uygun özellik parametre setlerinin bulunmasına yönelik umut verici sonuçlar verdiğini ortaya koymuştur. Çalışmada, özellik parametrelerinin boyutu arttıkça, kadın ve erkek konuşmacılar için başarı oranının arttığı görülmüştür. Elde edilen sonuçlar, CNN derin öğrenme sınıflandırıcısının MLP ve geleneksel sınıflandırıcılara göre daha başarılı olduklarını göstermektedir. Elde edilen analiz sonuçlarına göre, erkek ve kadın konuşmacılarda hibrit özellik çıkarım yöntemlerini oluşturan özellik çıkarım parametreleri ve sınıflandırıcı yöntemleri seçiminin Kürtçe C/V/S konuşma tespit sisteminin başarımını etkileyen iki önemli etken olduğu gözlenmiştir. Bunula birlikte, çerçeve uzunluğunun (pencereleme süresinin) ve pencereleme yöntemlerinin seçiminin sistemin başarımına fazla bir etkisi olmadığı görülmüştür. Gelecekteki çalışmalarda,

fonem seviyesinde oluşturulan veri kümesinin genişletilmesi ve farklı derin öğrenme algoritmalarının başarım sonuçları araştırılacaktır.

Kaynakça

[1] A. Graves and N. Jaitly, “Towards end-to-end speech recognition with recurrent neural networks,” ICML, pp. 1764 - 1772, 2014.

[2] A. Shewalkar, D.Nyavanandi, S.A. Ludwig, “Performance evaluation of deep neural networks applied to speech recognition: RNN, LSTM and GRU,” JAISCR, vol. 9, pp.235–245, 2019.

[3] M. Ravanelli et al., “Light gated recurrent units for speech recognition.” IEEE Transactions on Emerging Topics in Computational Intelligence, vol. 2, no. 2, pp. 92-102, 2017.

[4] Y. Yuan et al., “Auxiliary loss multimodal GRU model in audiovisual speech recognition,” IEEE Access, vol.6, pp. 5573-5583, 2018.

[5] M. Marolt et al., “Automatic segmentation of ethnomusicological field recordings,” Appl. Sci, vol.9, no.3, pp.1-12, 2019.

[6] Y. Hsuan Wang et al., “Gate activation signal analysis for gated recurrent neural networks and its correlation with phoneme boundaries,” INTERSPEECH, pp. 20-24, 2017.

[7] Z. Chen et al., “Practical singing voice detection system based on GRU-RNN,” CSMT, vol. 568, pp.15-25, 2019.

[8] C. Zheng et al., “An ensemble model for multi-level speech emotion recognition,” Appl. Sci, vol. 10 no. 1, pp. 1-20, 2019.

[9] A. Graves and J. Schmidhuber, “Framewise phoneme classification with bidirectional LSTM and other neural network architectures,” Neural Networks, vol. 18, no. 5–6, pp. 602–610, 2005.

[10] J. Franke, M. Muller and F. Hamlaoui, S. Waibel, “Phoneme boundary detection using deep bidirectional LSTMs,” In Proceedings of the Speech Communication, 12. ITG Symposium, pp. 77–381, 2016.

[11] I. Goodfellow, Y. Bengio and A. Courville, “Deep Learning,” MIT press: Cambridge, Massachusetts, London, England, 2016.

[12] X. Ma, Z. Wu, J. Jia, M. Xu, H. Meng, and L. Cai, “Study on feature subspace of archetypal emotions for speech emotion recognition,” arXiv 2016, arXiv:1602.05875.

[13] C. Li, et al., “Deepspeaker: An end-to-end neural speaker embedding system. Learning,” arXiv 2017, arXiv:1705.02304.

[14] J. Zhao, X. Mao, and L.Chen, “Speech emotion recognition using deep 1D&2D CNN LSTM networks,” Biomed. Signal Process. Control, vol. 47, pp. 312–323, 2019.

[15] D. Wang, X. Wang, and S. LV, “End – to end Mandarin speech recognition combining CNN and BLSTM,” Symmetry, vol. 11, pp. 1–19, 2019.

[16] G. Keren and B. Schuller,“Convolutional RNN: An enhanced model for extracting features from sequential data,” arXiv 2016, arXiv:1602.05875.

[17] https://tr.wikipedia.org/wiki/T%C3%BCrkiye%27de_konu%C5%9Fulan_diller

[18] Y.E. Tetik ve B.Bolat, “Gürültülü ortamlarda konuşma tespiti için yeni bir öznitelik çıkarım yöntemi,” Elektrik-Elektronik ve Bilgisayar Sempozyomu, pp.86-89, 2011.

[19] T. M. Nazmy, M. E. Gadallah, and A.A. Abdelhamid, “A novel method for Arabic consonant/vowel segmentation using wavelet transform,” IJICIS, vol.5, pp.353- 364, 2005.

[20] S. Hochreiter and J. Schmidhuber, “Long short term memory,” Neural Comput., vol. 9, pp.1735-1780, 1997.

[21] Ö. Batur Dinler and F. Karabiber, “Formant analysis of vowels in Kurdish language,” in Proceedings of the 25th Signal Processing Communications Applications Conference, pp.1-4, 2017.

[22] O. Eray, “Destek Vektör Makineleri ile ses tanıma uygulaması,” Pamukkale Üniversitesi, Yüksek Lisans Tezi, 2008.

[23] L. Güner, and İ. Ergenç, “Sesin doğası ve oluşumu (The nature of sound),” pp. 1-24.

[24] C. Yüzkollar, “Yapay sinir ağları kullanılarak paramak izi ve konuşmacı tanıma,” Sakarya Üniversitesi, Yüksek Lisans Tezi, 2007.

[25] M.Ciwan, “Kürtçe Dilbilgisi”, ISBN:9188054-160.

[26] E.D.B, Khan and R. Lescot, “Kürtçe Grameri”, Institut Kurde de Paris, Paris, Fransa, pp. 1-13, 1990.

[27] W.M, Thackston, “Kurmanji kurdish-A reference grammar with selected readings,” Cambridge, Mass: Harvard University, pp. 1-90, 2006. Available online: htpp://bibpurl.oclc.org/web/36880 (accessed on 25 December 2019).

[28] S. Gündoğdu, “Remarks on vowels and consonants in Kurmanji,” J. Soc. Sci. Muş Alparslan, vol. 4, pp. 1-14, 2016.

[29] P. V. Lieshout, “Praat short tutorial,” PhD, University of Toronto, 2003.

[30] https://en.wikipedia.org/wiki/Praat.

[31] https://en.wikipedia.org/wiki/Audacity_(audio_editor)

[32] https://en.wikipedia.org/wiki/WaveSurfer

[33] https://www.adobe.com/tr/products/audition.html

[34] Y. Korkmaz, “Türkçe’deki ünlü harflerin formant frekans değerlerine dayalı adli aksan analizi gerçekleştirimi,” Fırat Üniversitesi, Fen Bilimleri Enstitüsü, Yüksek Lisans Tezi, Elazığ, 2018.

[35] https://www.winpitch.com/

[36] Y. Korkmaz, A. Boyacı,“Adli bilişim açısından ses incelemeleri,” Science and Eng. J of Fırat University, vol. 30, no. 1, pp. 329-343, 2018.

[37] A. E. Sakran et al., “A review: Automatic speech segmentation,” IJCSMC, vol. 4, pp. 308-315, 2017.

[38] A. Yayla et al., “Circuit analysis application interface by using speech recognition technology,” Jret, vol. 5, no. 19, pp. 169–17, 2016.

[39] M. A. Karadaş, “Bilişim teknolojisi (BT) sınıflarında konuşma tanıma teknolojisi ile sınıf otomasyonu,” Gazi Üniversitesi, Bilişim Enstitsü, Yüksek Lisans Tezi, Ankara, 2014.

[40] İ. SEL, “Türkçe metinler için hece tabanlı metinden konuşma sentezleme sistemi,” Fırat Üniversitesi, Fen Bilimleri Enstitsü, Yüksek Lisans Tezi, Elazığ , 2013.

[41] U. Ayaz, “Text to speech synthesis for Turkish using a DSP,” Dokuz Eylül University, Graduate School of Natural and Applied Sciences, M. Sc Thesis, İzmir, 2016.

[42] T. Zhang and C. C. Kuo, “Hierarchical classification of audio data for archiving and retrieving,” ICASSP99, pp .3001-3004, 1999.

[43] G. Hemakumar and P. Punitha, “Automatic segmentation of Kannada speech signal into syllable and sub-words: noised and noiseless signals,” International Journal of Scientific & Engineering Research, vol. 5, no. 1, pp.1707-1711, 2014.

[44] M. Kalamani et al., “Hybrid speech segmentation algorithm for continuous speech recognition,” International Journal on Applications of Information and Communication Engineering, vol. 1, no.1, pp. 39- 46, 2015.

[45] M. Sidiq et al.,”Design and implementation of voice command using MFCC and MMs method,” ICoICT, pp. 375-380.

[46] M.A. Hossan et al., “A novel approach for MFCC feature extraction,” IEEE, pp.1-5, 2011.

[47] E. Yücesoy, and V.V. Nabiyev, “Comparison of MFCC, LPCC and PLP features for the determination of a speaker’s gender,” Signal Processing and Communications Applications Conference, pp. 321-324, 2014.

[48] G. Tumak ,”Saklı markov model tabanlı müzik parçası tanıma sistemi”, Yüksek Lisans Tezi, Yıldız Teknik Üniversitesi, Fen Bilimleri Enstitüsü, 2009.

[49] N. Aydin, and H.S. Markus, “Optimization of processing parameters for the analysis and detection of embolic signals,” Eur. J. Ultrasound, vol. 12, pp. 69–79, 2000.

[50] F.J. Harris, “On the use of windows for harmonic analysis with discrete Fourier transform,” Proc. IEEE, vol. 66, pp. 51–83, 1978.

[51] P.L. Chitra, and R. Aparna “Performance analysis of windowing techniques in automatic speech signal segmentation,” Indian J. Sci. Technol., vol. 8, pp. 1–29.2015.

[52] A.Koç, “Acoustic feature analysis for robust speech recognition”, Master Thesis, Boğaziçi Üniversitesi, Fen Bilimleri Enstitüsü, 2002.

[53] M.K. Baygün, “Türkçe komutları tanıyan ses tanıma sistemi geliştirilmesi”, Pamukkale Üniversitesi, Fen Bilimleri Enstitüsü, Yüksek Lisans Tezi, Denizli, 2006.

[54] A. Graves and N. Jaitly, “Towards end-to-end speech recognition with recurrent neural networks,” ICML, 2014, pp. 1764–1772, 2014.

[55] C. Batur Şahin, “Bilgi çıkarım teknikleri ile gen-kanser ilişkilerinin araştırılması,” Yıldız Teknik Üniversitesi, Fen Bilimleri Enstitsü, Doktora Tezi, İstanbul, 2019.

[56] https://data-flair.training/blogs/tensorflow-recurrent-neural-network/

[57] https://towardsdatascience.com/illustrated-guide-to-lstms-and-gru-s-a-step-by-step-explanation-44e9eb85bf21

[58] R. Rana, “Gated recurrent unit (GRU) for emotion classification from noisy speech,” arXiv 2016, arXiv:1612.07778v1.

[59] A. Shewalkar, D.Nyavanandi, S.A. Ludwig, “Performance evaluation of deep neural networks applied to speech recognition: RNN, LSTM and GRU,” JAISCR, vol. 9, pp.235–245, 2019.

[60] A.M. Başbuğ, “Ses olayı tanıma ve akustik sahne geri getirimi”, Başkent Üniversitesi, Fen Bilimleri Enstitüsü, Yüksek Lisans Tezi, 2019.

[61] B.C. Feltes, B.L. Grisci, J.F. Poloni, and M. Dorn, “Perspectives and applications of machine learning for evolutionary developmental biology,” Mol. Omics, vol. 14, pp. 289-306, 2018.

[62] P. Misra, and A. Giri, “Review of System Identification Using Neural Network Techniques,” Int. J. Electr. Electron. Data Commun, vol. 2, pp. 13–16, 2014.

[63] P. Boersma, “Praat, a system for doing phonetics by computer,” Glot International 5:9/10, 341-345.

[64] https://www.mathworks.com/products/matlab.html

[65] P.W.D. Charles, “Project Title,” GitHub repository, 2013, https://github.com/charlespwd/project-title

[66] E.Frank, M. A. Hall, and I. H. Witten,The WEKA Workbench. Online Appendix for “Data Mining: Practical Machine Learning Tools and Techniques,” Morgan Kaufmann, Fourth Edition, 2016.

[67] S. Lang, F. Bravo-Marquez, C. Beckham, M. Hall, and E. Frank, WekaDeeplearning4j: a Deep Learning Package for Weka based on DeepLearning4j,” In Knowledge-Based Systems, vol. 178, pp. 48-50, 2019. DOI:10.1016/j.knosys.2019.04.013 (author version)

[68] J. Chen, J. Benesty, Y. Huang, and S. Doclo, “New insights into the noise reduction wiener filter, ” IEEE Trans. Audio Speech Lang. Process, vol. 14, pp. 1218–1234, 2005.

[69] P. Cosi, D. Falavigna, M. Omologo, “A preliminary statistical evaluation of manual and automatic segmentation discrepancies,” In Proceedings of the European Conference on Speech Communication and Technology (EUROSPEECH), pp. 693–696, 1991.

[70] S.J. Cox, R. Brady, P. Jackson, “Techniques for accurate automatic annotationof speech waveforms, ” In Proceedings of ICSLP, pp. 1947–1950, 1998.

[71] A. Ljolje, J. Hirschberg, J.P.H. Van Santen, “Automatic Speech Segmentation for Concatenative Inventory Selection; Progress in Speech Synthesis,” Springer, pp. 305–311, 1997.

[72] N. Jain, D. Kaushik, “ Gender voice recognition through speech analysis with higher accuracy, ” In Proceedings of the 8th International Conference on Advance Computing and Communication Technology, pp. 1-5, 2014.

[73] S. Karasartova, “Metinden bağımsız konuşmacı tanıma sistemlerinin incelenmesi ve gerçekleştirilmesi,” Ankara Üniversitesi, Yüksek Lisans.

Tezden Üretilmiş Yayınlar

İletişim Bilgisi: baturrozlem@gmail.com.tr

Konferans Bildirileri

1. Ö. Batur Dinler and F. Karabiber, " Formant analysis of vowels in Kurdish language," in Proceedings of the 25th Signal Processing Communications Applications Conference-SIU, 2017, pp. 1-4.
2. Ö. Batur Dinler and N. AYDIN, "Kurdish digit recognition with Artificial Neural Network," International Engineering and Science Symposium-IESS, pp.760-767 , 2019.

Uluslararası Makaleler

1. Ö. Batur Dinler and N. AYDIN, " Kurdish recognition system digit," The Online Journal of Science and Technology, vol. 8, no. 1, pp. 101-105, 2018.
2. Ö. Batur Dinler and N. AYDIN, " Extraction of the acoustic features of semi-vowels in the Kurdish language," The Online Journal of Science and Technology, vol. 8, no. 2, pp. 79-83, 2018.
3. Ö. Batur Dinler and N. AYDIN, " An optimal feature parameter set based on Gated Recurrent Unit Recurrent Neural Networks for speech segment detection", Applied Sciences, vol. 10, no. 4, pp.1-23, 2020, Doi: 10.3390/app10041273.