

TÜRKİYE CUMHURİYETİ
YILDIZ TEKNİK ÜNİVERSİTESİ
FEN BİLİMLERİ ENSTİTÜSÜ

KONUŞMA VERİSİNDEN DUYGU DURUM TESPİTİ

Ferda DOYRANLI TAYŞI

YÜKSEK LİSANS TEZİ
Bilgisayar Mühendisliği Bölümü
Bilgisayar Mühendisliği Programı

Danışman
Öğr. Grv. Dr. Ömer Özgür BOZKURT

Nisan, 2019

TÜRKİYE CUMHURİYETİ
YILDIZ TEKNİK ÜNİVERSİTESİ
FEN BİLİMLERİ ENSTİTÜSÜ

KONUŞMA VERİSİNDEN DUYGU DURUM TESPİTİ

Ferda DOYRANLI TAYŞI tarafından hazırlanan tez çalışması 10.04.2019 tarihinde aşağıdaki jüri tarafından Yıldız Teknik Üniversitesi Fen Bilimleri Enstitüsü Bilgisayar Mühendisliği Bölümü Bilgisayar Mühendisliği Programı **YÜKSEK LİSANS TEZİ** olarak kabul edilmiştir.

Öğr. Grv. Dr. Ömer Özgür BOZKURT
Yıldız Teknik Üniversitesi
Danışman

Jüri Üyeleri

Öğr. Grv. Dr. Ömer Özgür BOZKURT, Danışman
Yıldız Teknik Üniversitesi

Dr. Öğr. Üyesi Göksel BİRİCİK, Üye
Yıldız Teknik Üniversitesi

Dr. Öğr. Üyesi Oğuz ATA, Üye
Altınbaş Üniversitesi

Danışmanım Öğr. Grv. Dr. Ömer Özgür BOZKURT sorumluluğunda tarafımca hazırlanan KONUŞMA VERİSİNDEN DUYGU DURUM TESPİTİ 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.

Ferda DOYRANLI TAYŞİ

İmza

TEŞEKKÜR

Öncelikle, yüksek lisans eğitimi ve tez çalışmaları süresince yardımlarını ve zamanını esirgemeyen danışmanım Öğr. Grv. Dr. Ömer Özgür BOZKURT'a, verdiği desteğin yanı sıra bu süreçte motivasyonumu yüksek tutmam için katkılarını esirgemeyen Dr. Öğr. Üyesi Z. Cihan TAYŞİ'ye içtenlikle teşekkürlerimi sunarım.

Aileme, MLP Sağlık Hizmetleri'ndeki yöneticime ve çalışma arkadaşlarıma, tez sürecindeki manevi destekleri, kişisel hayatımdaki ve iş hayatımdaki sorumluklarımı benim yerime üstlenerek tez çalışması için daha fazla zaman bulabilmemi sağladıkları ve bu tezin başarıyla sonuçlanmasını kolaylaştırdıkları için, Enes KURT'a teknik konulardaki yardımları ve manevi desteği için teşekkür ederim.

Ferda DOYRANLI TAYŞİ

İÇİNDEKİLER

SİMGE LİSTESİ	vi
KISALTMA LİSTESİ	vii
ŞEKİL LİSTESİ	ix
TABLO LİSTESİ	x
ÖZET	xi
ABSTRACT	xii
1 GİRİŞ	1
1.1 Literatür Özeti	1
1.2 Tezin Amacı	10
1.3 Hipotez	10
2 MATERYAL VE YÖNTEM	11
2.1 Veri Seti	11
2.1.1 BES	11
2.1.2 eNTARFACE'05	11
2.1.3 HUMAINE	12
2.1.4 Ryerson Multimedia Lab (RML)	12
2.1.5 Danish Emotional Speech	12
2.1.6 Polish Emotional Speech	12
2.1.7 Tez için Oluşturulan Veri Seti (TOVS)	13
2.2 SES KAYITLARININ İŞLENMESİ ve ÖZELLİK ÇIKARIMI	14
2.2.1 Sesin Özellikleri	14
2.2.2 Özellik Çıkarımı	21
2.3 SINIFLANDIRMA	23
2.3.1 Sınıflandırma Yöntemleri	23
2.3.2 Weka ile Sınıflandırma	26
3 DENEYSEL BULGULAR	27

3.1	Kullanılan Ses Özellikleri	27
3.2	Kullanılan Sınıflandırıcılar	28
3.2.1	Naive Bayes	29
3.2.2	Random Forest	30
3.3	Senaryoların Oluşturulması	30
3.3.1	Duygulara Göre Oluşturulan Senaryolar	31
3.3.2	Klip Boyutuna Göre Oluşturulan Senaryolar	31
3.3.3	Cinsiyete Göre Oluşturulan Senaryolar	32
3.3.4	Yaşa Göre Oluşturulan Senaryolar	33
4	SONUÇ ve ÖNERİLER	35
	Referanslar	38

SİMGE LİSTESİ

k	Frekans Değeri
$\log\text{HNR}$	Logarithmic Harmonics-to-Noise Ratio
n	Çerçeve İçerisinde, Üzerinde İşlem Gerçekleştirilen Örneğin Numarası
N	Çerçeve İçerisindeki Toplam Örnek Sayısı
r	Çerçeve
sign	İşaret Fonksiyonu
$\sum$	Toplam
W_k	İlgilenilen Frekans Aralığının Hesaplanabilmesi İçin Kullanılan Ağırlık Değerleri
X_r	Sinyalin Frekans Bölgesindeki Hali

KISALTMA LİSTESİ

ACE	ASCII Compatible Encoding
ACF	Autocorrelation Function
AD	Active Directory
ARFF	Attribute Relation File Format
ANN	Artificial Neural Network
ASR	Automatic Speech Recognition
BES	Berlin Emotional Speech
ECG	Electrocardiogram
F0	Fundamental Frequency
FS	Forward Selection
GMM	Gaussian Mixture Models
HMM	Hidden Markov Model
IBk	Instance Based Learning
ITU-R	International Telecommunication Union Radiocommunication Sector
k-NN	K-Nearest Neighbors
LDC	Linear Discriminant Classifier
LDC	Linguistic Data Consortium
LFPC	Log Frequency Power Coefficients
LPCC	Linear Prediction Cepstral Coefficients
LWL	Locally Weighted Learning
MANN	Modular Artificial Neural Network
MFCC	Mel-Frequency Cepstral Coefficients
MLP	Multilayer Perceptron

MYSA	Modüler Yapay Sinir Ağı
NN	Neural Nets
PCA	Principal Component Analysis
PEAQ	Perceptual Evaluation of Audio Quality
PFS	Promising First Selection
RBK	Radial Basis Kernel
RML	Ryerson Multimedia Lab
RMS	Root Mean Square
SBFS	Sequential Backward Feature Selection
SFFS	Sequential Floating Forward Selection
SFS	Sequential Forward Selection
SVM	Support Vector Machine
TOVS	Tez için Oluşturulan Veri Seti
XML	Extensible Markup Language
YSA	Yapay Sinir Ağı
ZCR	Zero Crossing Rate

ŞEKİL LİSTESİ

Şekil 2.1	Dosya İsimlendirme Yöntemi	13
Şekil 2.2	Sıfırdan Geçiş Sayısı [38]	15
Şekil 2.3	Kısa Zamanlı Enerji [39]	16
Şekil 2.4	Spektral Ağırlık Merkezi [40]	16
Şekil 2.5	Spektral Yuvarlanma [41]	17
Şekil 2.6	Spektral Akı [42]	18
Şekil 2.7	F0 [43]	18
Şekil 2.8	MFCC [44]	19
Şekil 2.9	MFCC [45]	20
Şekil 2.10	Tını [47]	21
Şekil 2.11	Audacity ekran görüntüsü	22

TABLO LİSTESİ

Tablo 2.1	TOVS içerisindeki örneklerin cinsiyet ve duygulara göre dağılımı . .	14
Tablo 3.1	Özellik Seçimi Sonuçları - TOVS içindeki Nötral, Mutlu, Heyecanlı ve Kızgın duyguları için	29
Tablo 3.2	Özellik Seçimi Sonuçları - BES Veri Setindeki Nötral, Mutlu, Heyecanlı ve Kızgın duyguları için	29
Tablo 3.3	Özellik Seçimi Sonuçları - BES Veri Setindeki Nötral, İğrenme, Kızgın ve Üzgün duyguları için	30
Tablo 3.4	Sınıflandırma Sonucundaki Duygu Dağılımı	31
Tablo 3.5	TOVS’de Duygulara Göre Random Forest ve Naive Bayes Sonuçları (%)	31
Tablo 3.6	BES’de Duygulara Göre Random Forest ve Naive Bayes Sonuçları (%)	31
Tablo 3.7	TOVS’de Klip Boyutlarına Göre Random Forest ve Naive Bayes Sonuçları (%)	32
Tablo 3.8	BES’de Klip Boyutlarına Göre Random Forest ve Naive Bayes Sonuçları (%)	32
Tablo 3.9	TOVS’de Cinsiyete Göre Random Forest ve Naive Bayes Sonuçları (%)	33
Tablo 3.10	TOVS’de Yaşa Göre Random Forest ve Naive Bayes Sonuçları (%) .	34

KONUŞMA VERİSİNDEN DUYGU DURUM TESPİTİ

Ferda DOYRANLI TAYŞI

Bilgisayar Mühendisliği Bölümü

Yüksek Lisans Tezi

Danışman: Öğr. Grv. Dr. Ömer Özgür BOZKURT

İnsanın duygusal durumunun değerlendirilmesine yönelik çalışmalar insan-bilgisayar etkileşiminde yaşanan hızlı gelişmeler sonucu önem kazanmıştır. Literatürde duygusal durumun tespit edilmesi için ses, görüntü, beyin dalgaları ve fizyolojik ölçümler gibi farklı veri tipleri üzerinde çalışmalar bulunmaktadır. Bu çalışma kapsamında ses verisi kullanılarak duygusal durum tespitini hedefleyen, Türkçe'ye özgü bir sistem geliştirilmesi amaçlanmıştır. Sistemin oluşturulabilmesi için öncelikle çok sayıda konuşmacının bulunduğu kapsamlı bir veri seti hazırlanmıştır. Ses verisinden duygusal durumun tespit edilmesine yönelik çalışmalarda kullanılan ses özellikleri incelenmiş ve bu özelliklerden Türkçe için uygun olanları tespit edilmeye çalışılmıştır. Son olarak sistem içerisinde kullanılacak olan sınıflandırıcının belirlenmesi için çeşitli testler yapılmıştır.

Anahtar Kelimeler: Duygu tanıma, ses tabanlı tanıma, özellik seçimi, özellik çıkarımı, sınıflandırma

ABSTRACT

SPEECH EMOTION RECOGNITION

Ferda DOYRANLI TAYŞI

Department of Computer Engineering
Master of Science Thesis

Advisor: Instructor Dr. Ömer Özgür BOZKURT

Detecting emotional state of a human has gained importance due to advances in the human computer interaction. There are several research conducted with different types of data such as audio, video, Electrocardiogram(ECG), and physiological in order to detect emotional state of a human. In this study, we focus on audio based emotion recognition and designed a system specifically for Turkish language. In order to create such a system, we created an audio database that contains a wide selection speakers. After that we surveyed the audio features that were used in audio based emotion recognition systems and applied feature selection methods to select most appropriate ones for the Turkish language. Finally we conducted several tests to determine the most suitable classifier for the task.

Keywords: Emotion recognition, mood recognition, audio based recognition, feature selection, feature extraction, classification

YILDIZ TECHNICAL UNIVERSITY
GRADUATE SCHOOL OF NATURAL AND APPLIED SCIENCES

1.1 Literatür Özeti

İnsanın duygusal durumunun değerlendirilmesine yönelik çalışmalar insan-bilgisayar etkileşiminde yaşanan hızlı gelişmeler sonucu önem kazanmıştır. İnsanların duygusal durumlarının belirlenmesi sadece insan-bilgisayar etkileşimi için değil insan-insan etkileşiminde de önemli bir problemdir.

Özellikle çok kısa sürede oldukça fazla miktarda bilginin üretildiği günümüzde bu bilginin aynı hızda işlenmesi ve değerlendirilmesi gerekmektedir. Bu yüzden duygusal durumun tespitine yönelik çalışmalar önem kazanmaktadır. Duygusal durumun değerlendirilmesi metin, görüntü ve ses gibi farklı verilerden bir veya birden fazlası kullanılarak yapılabilmektedir.

Tez çalışması kapsamında ses verisi üzerinden duygusal durumun tespit edilmesine yönelik bir sistem geliştirilmesi hedeflenmektedir. Bu hedef kapsamında duygusal durum tespiti için geliştirilmiş olan ses tabanlı sistemlere dair literatür özeti aşağıda verilmiştir. Duygusal durum tespiti için ses verisi kullananların yanı sıra farklı türde veriler kullanan sistemler de literatür özeti kapsamına dahil edilmiştir.

Duygusal durumun tespit edilmesinin önemli kullanım alanlarından biri çağrı merkezleridir. Çağrı merkezi çalışanlarının yaptığı telefon görüşmeleri, müşteri memnuniyeti adına kayıt altına alınmakta ve sonrasında hem müşteri memnuniyeti açısından hem de çağrı merkezi çalışanlarının performanslarının belirlenmesi adına değerlendirilmektedir. Ancak oluşan verinin miktarı, her çağrının değerlendirilmesini imkansız kılmaktadır. Böyle bir yapının yürütülebilmesi için duygusal durumun tespit ve takip edilebildiği sistemlere ihtiyaç vardır.

[1] çalışmasında yazarlar, De Silva vd. [2] tarafından yapılan çalışmada elde edilen sonuçları geliştirmek üzere, aynı veri seti üzerinde farklı bir yaklaşım geliştirmişlerdir. Çalışma kapsamında kullanılan veri seti ses ve görüntü örnekleri içermektedir. Ses örnekleri küçük çerçevelere bölünerek, her çerçeve için temel frekans (F0) değeri,

en yüksek, ortalama ve en düşük perde deęerleri, standard sapma deęeri, RMS (Root Mean Square - Kare Ortalamalarının Karekökü) enerji deęeri, en yüksek ve en düşük perde türev (derivative of pitch contour) deęerleri hesaplanmaktadır. Çalışma kapsamında kural tabanlı bir tanıma sistemi oluşturulmuştur. De Silva vd. çalışmasında birbiri ile karıştırılabildięi belirtilen duygusal durumlar için ses ve video özelliklerinin ne sırayla kullanılacağına ilişkin kural tabanlı bir yapı oluşturulmuştur. Örneğin üzüntü ve beğenmeme duygularına ait ses verilerinden elde edilen en yüksek ve ortalama perde deęerlerinin, dięer duygusal durumların perde deęerlerinden daha yüksek olduęu tespit edilmiştir. Geliştirilen sistemde öncelikle bu iki özellik kontrol edilmekte, belirlenen bir eşik deęerinin üzerinde olmaları durumunda görüntü özelliklerinde göz çevresinde kırışıklık olup olmadığı kontrol edilerek hangi duygusal durum olduęuna karar verilmektedir.

Petrushin [3] çalışmasında bir çağrı merkezine ulaşan çağrıların ajitasyon ve sakin olarak sınıflandırılarak önceliklendirilmesi için bir sistem geliştirmiştir. Geliştirilen sistemin başarımının ölçülebilmesi için mutluluk (happiness), kızgınlık (anger), üzüntü (sadness), korku (fear) ve normal durumlarını içeren bir veri seti oluşturulmuştur. Veri setinin oluşturulmasında 30 farklı konuşmacının toplam 700 kısa ifadesi kullanılmıştır. Çalışma kapsamında yer alan 23 konuşmacıdan kayıtları deęerlendirerek duygusal durumları tespit etmeleri istendiğinde, mutluluk için %61.4, kızgınlık için %72.2, üzüntü için %68.3, korku için %49.5 ve normal durumu için %66.3 başarımla sınıflandırma yapıldığı tespit edilmiştir. Kişilerin kendi ses kayıtlarındaki duygusal durumu belirlemeleri istendiğinde ise mutluluk için %71.2, kızgınlık için %98.1, üzüntü için %80, korku için %78.8 ve normal durumu için %71.9 başarı sağlanmıştır. Çalışma kapsamında temel frekans (F0), vocal energy, frequency spectral features, formants, and temporal features (speech rate and pausing) özellikleri kullanılarak bir sınıflandırıcı oluşturulmuştur. Birden fazla ANN (Artificial Neural Network - Yapay Sinir Ağı) kullanılarak oluşturulan sınıflandırıcının ortalama başarımı %70 olarak belirtilmiştir.

[4] çalışmasında yazarlar, insan bilgisayar etkileşimi sırasında kaydedilmiş olan diyalogları inceleyerek, olumsuz duygusal durum içerenlerin tespit edilmesine yönelik bir sistem geliştirmişlerdir. Geliştirilen sistemde kızgınlık, hüsrana ve benzeri duygusal durumlar olumsuz olarak kabul edilirken, mutluluk, haz vb. duygusal durumlar olumlu olarak kabul edilmiştir. Sistemde ses sinyalinin temel akustik özellikleri olan perde ve enerji kullanılmıştır. Perde bilgisinin ortalama, medyan, standard sapma, en yüksek ve en düşük deęerleri özellik olarak kullanılırken, enerji için ortalama, medyan, standard sapma, en yüksek ve aralık deęerleri kullanılmıştır. Belirlenen bu özellikler öncelikle PCA (Principal Component Analysis - Temel Bileşenler Analizi) yöntemi kullanılarak farklı bir boyuta indirgenmiş, sonrasında da

PFS (Promising First Selection - Umut Verici İlk Seçim) ve FS (Forward Selection - İleri Seçim) algoritmaları kullanılarak erkek ve kadın konuşmacılar için farklı sayıda özellik seçilmiştir. Sınıflandırıcı olarak LDC (Linear Discriminant Classifier - Doğrusal Ayırıcı Sınıflandırıcı) ve k-NN (K-Nearest Neighbors - K-En Yakın Komşular) yöntemlerinin test edildiği çalışmada, kadın konuşmacılar için yaklaşık %80 başarı, erkek konuşmacılar için ise yaklaşık %76 başarı sağlanmıştır.

[5] çalışmasında Yacoub vd. Interaktif Sesli Yanıt (Intractive Voice Response) sistemleri için özellikle önemli olan kızgınlık durumunun diğer durumlardan ayırt edilmesine yönelik bir sistem oluşturmuşlardır. Sistemde kullanılacak olan ses özelliklerinin ve sınıflandırıcıların belirlenmesi noktasında üzgün (sad), sıkın (bored), mutlu (happy), kızgın (anger) ve nötr durumları değerlendirilmiştir. Sistem içerisinde, Pennsylvania Üniversitesi Linguistic Data Consortium tarafından sağlanan 9 saatlik konuşma verisi kullanılmıştır. Duygusal durumun tespit edilmesi için ses verisi 384 örneklik (yaklaşık 18 ms.) çerçevelere bölünmüş ve her çerçeve için perde ile ilgili 12, enerji ile ilgili 12 ve sesteki duraksamaların yapısı ile ilgili olarak 13 özellik olmak üzere toplam 37 özellik üretilmiştir. Tüm özellikler FS yöntemi kullanılarak kademelendirilmiş ve en yüksek kademeli 19 özellik seçilmiştir. Oluşturulan özellik seti ANN, SVM (Support Vector Machine - Destek Vektör Makinesi), k-NN ve karar ağacı sınıflandırıcıları kullanılarak test edilmiştir. Yapılan testler sonucunda seçilen özelliklerin kızgınlık durumunun diğer durumlardan ayırt edilmesinde %83 başarı sağladığı, ancak tüm duygusal durumların tek tek tespit edilmesi durumunda %8.7 gibi oldukça düşük bir başarı gösterdiği belirtilmektedir.

[6] çalışmasında Nwe vd. duygusal durum tespiti için LFPC (Log Frequency Power Coefficients - Günlük Frekans Güç Katsayıları) özelliklerinin başarımını incelemişlerdir. Özelliklerin başarımı, LPCC (Linear Prediction Cepstral Coefficients - Doğrusal Tahmin Güç Spektrumu Katsayıları) ve MFCC (Mel-Frequency Cepstral Coefficients - Ton-Frekansı Güç Spektrumu Katsayıları) özelliklerinin başarımaları ile karşılaştırılmıştır. Özelliklerin karşılaştırılması için oluşturulan sistemde sınıflandırma için Discrete Hidden Markov Model kullanılmıştır. Sistemin test edilmesi için altısı Burmese dilini, altısı Mandarin dilini konuşan 12 konuşmacıdan alınan ve kızgınlık (anger), tiksime (disgust), korku (fear), neşe (joy), üzüntü (sadness) ve şaşkınlık (surprise) duygusal durumlarının herbiri için 60 ses kaydının bulunduğu toplam 720 kayıttan oluşan bir veri seti kullanılmıştır. Ses verisinin değerlendirilmesinde çerçeveleme yöntemi kullanılmış olup, çerçeve boyutu temel frekansın iki katı olacak şekilde 16 ms olarak seçilmiş ve çerçevelerin 9 ms. uzunluğundaki kısımları örtüştürülmüştür. Elde edilen sonuçlarda LFPC özellik setinin LPCC ve MFCC özellik setlerine göre daha başarılı olduğu ve duyguların sınıflandırılmasında ortalama %78 başarı gösterdiği belirtilmiştir.

[7] çalışmasında Bhatti vd. konuşmacının dilinden bağımsız şekilde duygusal durumun tespit edilebildiği bir sistem geliştirmişlerdir. Geliştirilen sistemin test edilmesinde mutluluk (happiness), üzüntü (sadness), kızgınlık (anger), korku (fear), şaşkınlık (surprise) ve iğrenme (disgust) duygusal durumlarını içeren İngilizce, Çince, Urdu ve Endonezyaca dillerinde kaydedilmiş toplam 500 ifadenin yer aldığı bir veri seti kullanılmıştır. Çalışma kapsamında kullanılmak üzere öncelikle ses sinyalinin perde ve genliği ile ilgili toplam 17 adet özellik belirlenmiş sonrasında çeşitli özellik seçim yöntemleri kullanılarak özellik sayısı 12'ye indirilmiştir. Seçilen 12 özellik MANN (Modular Artificial Neural Network - Modüler Yapay Sinir Ağı) kullanılarak değerlendirildiğinde %83.31 oranında başarımla sağlanmıştır.

[8] çalışmasında Busso vd. duygusal durumun tespit edilmesine yönelik çalışmaları inceleyerek sadece yüz ifadelerinin, sadece ses verisinin ve her iki verinin kullanıldığı sistemlerin avantaj ve dezavantajlarını incelemişlerdir. Sistemlerin karşılaştırılması için kızgınlık (anger), üzüntü (sadness), mutluluk (happiness) ve nötr durumlarını içeren Ses sinyalinin perde ve enerji özelliklerine bağlı 11 özelliği kullanıldığı yapıda %70.9 başarımla sağlanırken, sistemin mutluluk ve kızgınlık durumunu ayırmada yetersiz kaldığı belirtilmiştir.

[9] çalışmasında Lee vd. kızgınlık (anger), mutluluk (happiness), nötr ve üzüntü (sadness) duygusal durumlarının tespit edilebilmesi için fonem tabanlı bir sistem geliştirmişlerdir. Geliştirilen sistem içerisinde MFCC katsayılarının ilk 13 tanesi, bu katsayıların farkları ve hızlanma katsayıları kullanılarak 39 boyutlu bir özellik seti oluşturulmuştur. Özelliklerin çıkarımı işlemi ses sinyallerinin 25 ms uzunluğunda kesilmesi ve 10 ms uzunluğunda örtüştürülmesi ile oluşturulan çerçeveler üzerinde gerçekleştirilmiştir. Yarı profesyonel bir aktrist tarafından söylenen toplam 880 ifadeden oluşan veri seti üzerinde yapılan test işlemlerinde yaklaşık %75 oranında başarımla sağlanmıştır.

[10] çalışmasında duygusal durumun tespit edilmesi için hem ses özelliklerinden hem de dilin özelliklerinden faydalanan bir sistem geliştirilmiştir. Sistemin geliştirilmesi ve test edilmesi için iki farklı veri seti oluşturulmuştur. Geliştirmeler sırasında kullanılan veri setinde Almanca ve İngilizce kelimelerden oluşan ve sadece bir tanesi kadın olan toplam 13 konuşmacı tarafından seslendirilen 2829 ifade yer almaktadır. Sistemin test edilmesi için ise 700 ifadeden oluşan farklı bir veri seti kullanılmıştır. Her iki veri setinde de kızgınlık (anger), iğrenme (disgust), korku (fear), neşe (joy), nötr, üzüntü (sadness) ve şaşkınlık (surprise) duygusal durumları bulunmaktadır. İlk aşamada 200 adet ses özelliğinin oluşturulduğu çalışmada Linear Discriminant Analysis yöntemi kullanılarak özellik vektörünün uzunluğu 33'e indirilmiştir. k-NN, GMM (Gaussian Mixture Models - Gauss Karışım Modelleri), NN (Neural Nets - Sinir Ağları) ve SVM

sınıflandırma algoritmalarının başarımlarının karşılaştırıldığı çalışmada sadece ses özelliklerinin kullanıldığı durumda en iyi başarımları yaklaşık %75 ile MLP yöntemi göstermiştir.

[11] çalışmasında ses analizi ile kızgınlık (anger), mutluluk (happiness), üzüntü (sadness), şaşırma (surprise) ve nötr duygusal durumlarının tespit edilmesine yönelik bir sistem geliştirilmiştir. Sistemin geliştirilmesi aşamasında Danish Emotional Speech veri seti üzerinde çalışmalar yapılmıştır. Sistem kapsamında öncelikle 39 adet farklı özellik oluşturulmuştur. Oluşturulan bu özellik seti içerisinde SFS (Sequential Forward Selection - Sıralı İleri Seçim) yöntemi kullanılarak en iyi beş adet özellik olan F0'ın ikinci türevi, F0'ın birinci türevi, MBE5'in ikinci türevi, MBE4'ün ikinci türevi, ve F1'in birinci türevi seçilmiştir. Duygusal durumların sınıflandırılması için hem HMM (Hidden Markov Model - Gizli Markov Modeli) hem de SVM sınıflandırıcıları kullanılarak yapılan testlerde HMM için %99.5, SVM yöntemi için ise %88.9 başarımlar elde edilmiştir. Çalışma kapsamında MFCC özelliklerinin tek başına sınıflandırma başarımlarına etkileri de değerlendirilmiş ve sadece ilk beş MFCC katsayısı kullanılarak bir deneme yapılmıştır. Bu denemede yaklaşık %60 başarımlar elde edilmiş ve bunun sonucu olarak MFCC özelliklerinin tek başına duygusal durum değerlendirmesinde yeterli olmadığı belirtilmiştir.

[12] çalışmasında Iseli vd. yaş ve cinsiyet gibi unsurların sesin temel özellikleri üzerindeki etkilerini araştırmışlardır. Araştırma kapsamında farklı yaş gruplarından 38 kadın ve 54 erkekte oluşan 92 kişilik bir gruba ait veri seti üzerinde yapılan testlerde F0, open quotient ve spectral tilt özelliklerinin, yaş ve cinsiyet ile doğrudan ilgili olduğu tespit edilmiştir. Ses özelliklerinin çıkarımında ses sinyali 25 ms uzunluğundaki çerçevelere bölünmüş ve çerçeveler her adımda 10 ms. kaydırılmıştır. Ses özelliklerinin çıkarımında Snack Sound Toolkit [13] uygulaması kullanılmıştır.

[14] çalışmasında yazarlar, ses verisinden duygusal durumun belirlenmesine yönelik olarak oluşturulacak sistemlerde kullanılabilecek ses özelliklerinin belirlenmesine yönelik Ensemble random forest to trees model isimli bir yöntem önermişlerdir. Önerilen yöntem ile oluşturulan yeni özellik setinin sistemin sınıflandırma başarımlarını %2 kadar arttırdığı belirtilmiştir. Ancak çalışma kapsamında kullanılan veri setinin detayları, özellik çıkarımında kullanılan yöntemlerin detayları ve sınıflandırıcıların parametreleri ile ilgili yeterli bilgi verilmemiştir.

Clavel vd. [15] çalışmasında, ses verisi kullanılarak korku (fear) duygusal durumunun tespit edilmesini hedeflemişlerdir. Bu kapsamda çeşitli filmlerdeki nötr, korku ve farklı olumsuz duygusal durumların olduğu çeşitli sahnelerden oluşan bir veri seti oluşturulmuştur. Oluşturulan veri setindeki ses verisi, 40 ms.

uzunluğundaki çerçeveler kullanılarak değerlendirilmiştir. Çerçeveler her adımda 30 ms uzunluğundaki kısımları örtüşecek şekilde kaydırılmıştır. Her çerçeve için 534 adet özelliğin hesaplandığı çalışmada sınıflandırma için GMM yönteminden faydalanılmıştır. Önerilen sistem korku durumunun tespit edilmesinde yaklaşık olarak %71 başarımla göstermiştir.

[16] çalışmasında yazarlar İtalyanca için ses verisinden duygusal durumun tespit edilmesine yönelik bir sistem önermişlerdir. Önerilen sistemin test edilebilmesi için öncelikle çeşitli televizyon dizileri ve filmlerinden alınan örneklerden oluşan ve kızgınlık (anger), korku (fear), neşe (joy), üzüntü (sadness) ve nötr duygusal durumlarını içeren 391 örnekten oluşan bir veri seti oluşturulmuştur. Sınıflandırma için literatürdeki diğer çalışmalara benzer şekilde temel frekanslar F0, F1, F2, F3, enerji özellikleri ve MFCC değerlerini içeren bir özellik seti oluşturulmuştur. Bu özelliklerin çeşitli varyasyonlarının kullanılması ile özellik vektörünün boyutu 240'a çıkarılmıştır. Çalışmada duygusal durumun sınıflandırılması için iki farklı sistem önerilmektedir. Birinci sistemde ses örnekleri, özellik çıkarımının ardından bir sınıflandırıcı ile hangi duygusal duruma ait oldukları belirlenirken, ikinci sistemde ses örnekleri üzerinde öncelikle cinsiyet tespiti yapılmakta ve sonrasında ilgili cinsiyete özgü bir duygusal durum sınıflandırıcısı kullanılarak duygusal durum belirlenmektedir. İki kademeli olan sistemin başarımda %5 ila %7 oranında iyileşme sağladığı görülmektedir.

[17] çalışmasında yazarlar konuşma sırasında duygusal durumun tespit edilmesi için diğer çalışmalardan farklı olarak konuşmanın içeriğini kullanan bir sistem önermişlerdir. Önerilen sistemde ses sinyalinin fiziksel ve algısal çeşitli özelliklerini çıkartmak yerine konuşmada geçen kelimelerin bulunması ve bu kelimelerin kullanım şekline göre duygusal durumun tespit edilmesi yöntemi kullanılmıştır. Konuşmadaki kelimelerin anlamlandırılması için hem vektör uzay modeli(vector space modelling) hem de dize çekirdeği(string kernel) yöntemleri ile testler gerçekleştirilmiştir. Yazarlar elde edilen sistemin oldukça düşük başarıma sahip olduğunu belirtirken, bunun temel olarak ASR (Automatic Speech Recognition - Otomatik Konuşma Tanıma) ile elde edilen kelimelerin %30 oranında hatalı olmasına bağlamaktadırlar. Bunun yanında sonuçlar kısmında vektör uzay modeli yönteminin dize çekirdeği yöntemine göre çok daha efektif olduğu da belirtilmektedir.

Yang ve Lugger çalışmalarında[18] sesten duygusal durum çıkarımında kullanılmak üzere yeni özellikler önermişlerdir. Önerilen yeni özellikler, sesin kalitesinin ölçülmesinde kullanılan özelliklerden ve sesin armoni özelliklerinden oluşmaktadır. Yazarlar önerdikleri yeni özelliklerin başarımlarını, sesten duygusal durum çıkarımında çokça kullanılan fiziksel özellikler ve MFCC katsayılarının başarımları ile

karşılaştırmışlardır. Elde edilen sonuçlar, önerilen yeni özelliklerin eski özellikler ile beraber kullanılması durumunda başarımda %2 oranında iyileşme sağladığını göstermektedir.

[19] çalışmasında yazarlar, ses verisinden duygusal durumun tanınmasında kullanılmak üzere, ITU-R (International Telecommunication Union Radiocommunication Sector - Uluslararası Telekomünikasyon Birliği Radyo Haberleşme Sektörü) tarafından Recommendation BS.1387 tanımlanan "PEAQ (Perceptual Evaluation of Audio Quality - Ses Kalitesinin Algısal Olarak Değerlendirilmesi)" modelindeki özelliklere benzer özelliklerin kullanılmasını önermektedirler. BES (Berlin Emotional Speech) veri seti üzerinde yapılan testlerde, yazarlar tarafından önerilen sistemin, genel olarak kullanılan fiziksel ve algısal özelliklere göre %10 daha yüksek başarımla gösterdiği belirtilmiştir.

[20] çalışmasında yazarlar sesten duygusal durumun tespit edilmesinde kullanılmak üzere yeni özellikler önermiştir. Geçiş Tayfi Özellikleri (Modulation Spectral Features) olarak isimlendirilen bu yeni özellikler, ses sinyalinin spectra-temporal ifadesi üzerinden hesaplanmaktadır. BES veri seti üzerinde yapılan testlerde yeni özellik seti, duygusal durum çıkarımında sıkça kullanılan MFCC özellikleri ile karşılaştırılmıştır. Sonuç olarak her iki özellik setinin beraber kullanıldığı durumlarda 7 farklı duygusal durum için başarımların %91.6 olduğu tespit edilmiştir.

[21] çalışmasında duygusal durumun belirlenmesinde literatürdeki mevcut ifade tabanlı yöntemlere alternatif olarak sesli harflere uygulanan vurgunun kullanılması önerilmiştir. Çalışmada BES veri setindeki kayıtlar, vurgulu, vurgusuz ve ünsüzler olarak işaretlenmiştir. Bu işaretleme işleminden sonra ilgili alanlar için fiziksel ve algısal toplam 237 farklı özellik içeren vektörler oluşturulmuştur. Sınıflandırma işlemi için RBK (Radial Basis Kernel - Radyal Basis Çekirdeği) kullanılan SVM sınıflandırıcısından faydalanan çalışmada önerilen özellik seti ile %82.3 başarımla elde edilmiştir.

Vogt ve Adrea [22] çalışmasında ses sinyalinden duygusal durumun tespit edilebilmesi için iki aşamalı bir sınıflandırma sistemi önermişlerdir. Sistem ilk aşamada konuşmacının cinsiyetini tespit etmekte, sonrasında ise ilgili cinsiyete uygun duygusal durum sınıflandırma sistemini kullanarak konuşmacının duygusal durumunu belirlemektedir. Önerilen sistem BES ve SmartKom Mobile verisetleri üzerinde perde, enerji ve MFCC özellikleri ve Naive Bayes sınıflandırıcısı kullanılarak test edilmiştir. Yapılan testlerin sonucunda cinsiyet ayrımını başarılı yapılabilmesi durumunda duygusal durumun tespit edilmesindeki başarımların arttığı tespit edilmiştir.

Sebe vd. [23] İnsan-Bilgisayar etkileşiminde önemli bir problem olan duygusal

durumun tespit edilebilmesi yönelik olarak yaptıkları araştırmada görsel ve işitsel veri kullanarak kişilerin 11 olası duygusal durumdan hangisinde oldukları tespit edilmeye çalışılmaktadır. Çalışma kapsamında genel olarak kullanılan nötr(neutral), mutlu (happy), şaşırma (surprise), kızgınlık (angry), iğrenme (disgust), korku (fear) ve üzgün (sad) duygusal durumlarının yanı sıra ilgi (interest), sıkılma (boredom), karmaşa (confusion) ve hüsran (frustration) durumları da kullanılmıştır. Çalışma kapsamında oluşturulan sistem, konuşmacı bağımlı olup, sadece sistemdeki eğitim seti içerisinde bulunan 24 kadın ve 14 erkekten oluşan 38 kişilik grup için sonuçlar üretebilmektedir. Sistemde ses verisinin 3 temel özelliği kullanılmıştır. Bu özellikler logaritmik enerji, hece oranı (syllable rate) ve sesin perdesidir. Sistemin duygusal durumları tanıyabilme başarısı sadece görüntü özellikleri kullanıldığında yaklaşık %56, sadece ses özellikleri kullanıldığında yaklaşık %45, hem ses hem de görüntü özellikleri kullanıldığında ise %83.56 ila %98.34 arasında değişmektedir.

[24] çalışmasında yazarlar, HMM yöntemini kullanarak 6 temel duygusal durum olan kızgınlık, antipati, korku, mutluluk, üzüntü ve şaşkınlık durumlarının görsel ve işitsel veri kullanılarak tespit edilmesine yönelik bir sistem oluşturmuşlardır. Sistem içerisinde ses verisinin perde özelliğinden faydalanılmıştır. Ses kayıtları 0.01 sn uzunluğunda parçalara bölünerek her parçadaki perde değerleri hesaplanmış ve HMM için girdi olarak kullanılmıştır. İki denekten her duygusal duruma ait 12'şer kayıt alınarak oluşturulan veri seti üzerinde yapılan testlerde önerilen sistemin başarımının %72 civarı olduğu tespit edilmiştir.

[25] çalışmasında yazarlar kişilerin depresyonda olup olmadığının tespit edilmesinde kullanılabilecek bir sistem geliştirmişlerdir. Geliştirilen sistem hem görsel hem de işitsel veri kullanılmaktadır. Sistem içerisinde ses verisine ait toplam 2268 adet özellik hesaplanmaktadır. Bu özellik seti içerisinde sesin fiziksel özellikleri ile ilgili olanları ses şiddeti (loudness), sıfırdan geçiş sayısı (zero crossing rate), 250 – 650 Hz ve 1 kHz – 4 kHz bandlarındaki enerji, spektral yuvarlanma (spectral roll-off) noktaları, spektral akı (spectral flux), entropi (entropy), varyans, çarpıklık (skewness), basıklık (kurtosis), psikoakustik keskinlik (psychoacoustic sharpness), harmoniklik (harmonicity), düzlük (flatness), F0, seğirme (jitter), parıltı (yerel) (shimmer (local)), jitter (delta: "jitter of jitter"), Logaritmik Harmonik-Gürültü Oranı (logarithmic Harmonics-to-Noise Ratio (logHNR)) vb. özellikler bulunmaktadır. Geliştirilen sistem 292 farklı kişiye ait toplam 340 adet video kaydı içeren bir veri seti ile test edilmiştir. Ses sinyali 3 sn ve 20 sn uzunluğunda iki farklı çerçeve boyutu kullanılarak değerlendirilmiştir. Bu değerlendirme işleminin devamlılığının sağlanabilmesi için her iki durumda da çerçeveler 1 sn uzunluğunda kaydırılarak işlemler gerçekleştirilmiştir. Sistemin başarımı, bağımsız kişilerin veri setinde yaptıkları etiketlemelerin, sistemin ürettiği sonuçlarla karşılaştırılması sonucunda elde edilmiştir.

[26] çalışmasında yazarlar nötr(neutral), kızgınlık (angry), ve korku (fear) duygusal durumlarının ses analizi ile tespit edilmesine yönelik bir sistem oluşturmuşlardır. Oluşturulan sistem içerisinde ses sinyali 55ms uzunluğunda çerçevelere bölünerek, her bir çerçeve için karşılıklı bilgi (mutual information), boyut korelasyonu (dimension correlation), entropi korelasyonu (entropy correlation), Shannon entropi, Lempel–Ziv karmaşıklık (complexity) and Hurst üs (exponent) özellikleri çıkartılmıştır. Sonrasında herbir özelliğin ortalama (mean), standart sapma (standard deviation), çarpıklık (skewness) ve basıklık (kurtosis) değerleri hesaplanarak 24 boyutlu bir özellik vektörü oluşturulmuştur. Oluşturulan özellik vektöründeki özelliklerin sınıflandırma başarımı üzerine etkisi Polish Emotional Speech Veri Seti üzerinde SFFS (Sequential Floating Forward Selection - Sıralı Kayan İleri Seçimi) yöntemi kullanılarak belirlenmiş ve özellik vektörünün boyutu 20 özelliğe indirilmiştir. Sınıflandırma işlemi için 3 katmanlı bir YSA kullanılmıştır. Giriş katmanında özellik vektörü ile aynı sayıda, gizli katmanda 50 adet, çıkış katmanında ise 3 adet nöron bulunmaktadır. ANN'nin eğitilmesi için esnek geri yayılım (resilient backpropagation) algoritması kullanılmıştır. Gizli katmandaki nöronlarda hiperbolik teğet (hyperbolic tangents), çıkış nöronlarında ise doğrusal fonksiyon kullanılmıştır. Oluşturulan sistem, Polish Emotional Speech veri setinde %72.28, Berlin Emotional Speech veri setinde %75.4, LDC Emotional Speech veri setinde ise %80.75 başarımlar göstermiştir.

[27] çalışmasında yazarlar kişilerin nötr (neutral), kızgınlık (anger) ve neşe (joy) duygusal durumlarından hangisinde olduğunu tespit etmek üzere kalp ritmi ve ses sinyalini kullanan bir sistem geliştirmişlerdir. Geliştirilen sistemde ses verisi örtüşen çerçevelere bölünerek işlenmiş ve her çerçeve için MFCC değerleri hesaplanmıştır. Elde edilen özellik seti, Weka programında bulunan Naive Bayes, Lazy IB1, RBF Network, Logistic, Ada Boost M1, Bagging ve Random tree sınıflandırıcıları ile test edilmiştir. Eşit sayıda kadın ve erkek içeren 30 kişilik bir gruptan alınan ses örnekleri ile yapılan testlerde farklı sınıflandırıcılar için %37 ile %45 arasında değişen tanıma başarımları elde edilmiştir. Çalışma kapsamında kullanılan MFCC özelliklerinin kaç katsayı (coefficient) içerdiği, sınıflandırma algoritmalarının parametreleri, eğitim seti içerisindeki örneklerin duygusal durumlara göre dağılımı vb. birçok detay belirtilmemiştir.

Lalitha vd. [28] çalışmasında BES veri seti üzerinde MFCC, Cepstrum, ve sentetik olarak genişletilmiş MFCC özellik setleri ve bu üç farklı özellik setinin olası kombinasyonlarının duygusal durumların sınıflandırılmasındaki başarımlarına etkilerini ölçmüştür. Çalışma kapsamında veri seti içerisinde bulunan nötr (neutral), kızgınlık (anger), korku (fear), neşe (joy), üzüntü (sadness), tiksinti (disgust) ve sıkılma (boredom) duygusal durumlarının tamamı değerlendirilmeye alınmıştır. Çalışma kapsamında sınıflandırma algoritması olarak ANN kullanılmıştır. Ancak

alıřma kapsamında zellik setlerinin hesaplanması, sınıflandırma algoritmasına ait parametreler vb. birçok konunun detayları paylaşılmamıştır. alıřma kapsamında elde edilen duygusal durum tanıma başarısının en iyi durumda %85.7 olduėu belirtilmektedir.

1.2 Tezin Amacı

Literatürde yapılan alıřmalarda herhangi bir dil için duygusal durum tespitine yönelik olarak geliştirilen sistemlerin farklı dillerde aynı başarıyı gösteremediėi görölmektedir[29]. Tez kapsamında Türke için sadece ses verisi kullanılarak konuşmacının duygusal durumunun tespit edilebilmesine yönelik bir sistemin geliştirilmesi hedeflenmektedir. Bu kapsamda öncelikle bir radyo programının internet ortamında mevcut olan kayıtlarından faydalanılarak konuşmacıların cinsiyet ve yař bilgisinin de bulunduėu bir veri seti oluşturulmuştur. Yapılan etiketleme alıřması sonucu kızgınlık, üzüntü, heyecan, mutluluk, aresizlik ve nötr durumlarının yer aldığı toplam 2194 adet kayıttan oluşan bir veri seti oluşturulmuştur.

Bu veri seti kullanılarak Türke için yukarıda sayılan 6 farklı duygusal durumun ses verisi üzerinden tespit edilmesine yönelik bir sistemin geliştirilmesi hedeflenmiştir.

1.3 Hipotez

Literatür taraması kapsamında Almanca başta olmak üzere birçok farklı dil için ses verisinden duygusal durumun tespit edilmesine yönelik sistemler geliştirildiėi görölmüştür. Literatürde sadece ses verisi üzerinde duygusal durumun ıkarımına yönelik olarak önerilen sistemler temel olarak iki noktada ayrılmaktadırlar: Bunlardan birincisi kullanılan zellik seti diėeri ise kullanılan sınıflandırma algoritmasıdır. Yine literatür taraması kapsamında farklı diller için benzer zellik setlerinin kullanılması durumunda sistemlerin başarımında deėişiklikler olduėu gözlenmiştir. Bu yüzden de tez kapsamında, kapsamlı bir veri seti üzerinde, literatürde tanımlı olan sesin eřitli fiziksel ve algısal zellikleri kullanılarak Türke'ye özgü bir sistem geliştirilebileceėi düşünölmektedir.

2 MATERYAL VE YÖNTEM

2.1 Veri Seti

Ses analizi ile duygusal durumun tespit edilmesine yönelik olarak farklı dillerde yapılmış çeşitli çalışmalar bulunmaktadır. Bu çalışmalar kapsamında farklı veri setleri oluşturulmuştur. Bu bölümde önceki çalışmalarda kullanılan veri setlerinin oluşturulması, yapıları ve içerikleri hakkındaki detaylı bilgiler alt başlıklar halinde verilmektedir. Bu bilgiler ışığında çalışma kapsamında oluşturulan veri setinin detayları da son kısımda paylaşılmaktadır.

2.1.1 BES

Burkhardt ve arkadaşları tarafından oluşturulan bu veri seti, ana dilleri Almanca olan beş aktris ve beş aktörden oluşan 10 kişinin günlük hayatta kullanılan Almanca 10 farklı cümleyi seslendirmesi ile oluşturulmuştur. Veri seti, 10 cümlenin tüm konuşmacılar tarafından nötral (neutral), kızgınlık (anger), korku (fear), neşe (joy), üzüntü (sadness), tiksinti (disgust) ve sıkılma (boredom) duygusal durumlarını bir veya daha fazla kez yansıtacak şekilde seslendirmesi ile oluşturulmuş toplam 800 kayıt içermektedir. [30]

2.1.2 eNTARFACE'05

eNTARFACE'05 veri seti, 14 farklı milletten toplamda 42 konuşmacının kayıtlarından oluşturulmuştur. 42 konuşmacının %81'i erkek, %19'u kadındır. Kayıtların yapılması iki hafta sürmüştür. Konuşmacılara mutluluk (happiness), üzüntü, şaşkınlık (surprise), kızgınlık, tiksinti, korku duygu durumları için İngilizce hikayeler verilmiş ve konuşmacılar verilen duyguları taklit ederek okumuşlardır. İki hakem duyguların gerçekçi olarak yansıtılıp yansıtılmadığına karar vermiştir. Başarılı olanlar veri setine eklenmiş, başarısız olanlar elenmiştir. [31]

2.1.3 HUMAINE

HUMAINE veri seti duygusal durumun gözlenmesi ve tespit edilmesine yönelik olarak görsel ve işitsel verinin yanında fizyolojik verilerinde tutulduğu bir veri setidir. Veri seti, doğal ve suni olarak iki grupta değerlendirilen farklı veri setlerindeki kayıtlardan oluşturulmuştur. Doğal veri setleri, televizyondaki realite programları, sokak röportajları ve çeşitli tartışma programlarına ait video kayıtlarını içermektedir. Yapay veri setleri içerisinde ise iki kişi arasında önceden belirlenen bir konudaki diyaloglar, sürücülerin farklı koşullar altındaki sürüşlerinin görüntüleri ile fizyolojik (kan basıncı, nefes sıklığı, vucüt sıcaklığı) durumlarına ait kayıtları içeren çeşitli veri setleri bulunmaktadır. [32]

2.1.4 Ryerson Multimedia Lab (RML)

RML veri seti, 8 farklı konuşmacının 6 farklı dilde, kızgınlık, tiksinti, korku, üzüntü, şaşkınlık ve mutluluk duygu durumlarındaki görsel ve işitsel kayıtlarından oluşmaktadır. Toplamda 720 kayıt içermektedir. [33]

2.1.5 Danish Emotional Speech

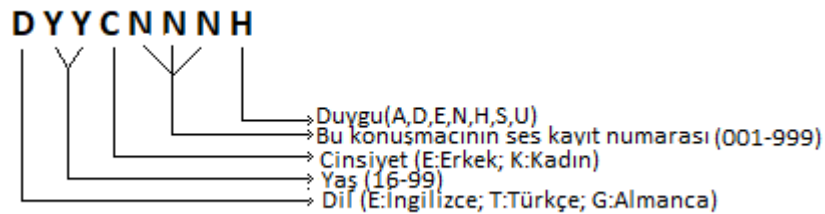
Veri seti, iki kadın iki erkek olmak üzere dört oyuncunun, nötral, şaşkınlık, mutluluk, üzüntü, kızgınlık duygularını canlandırmasından oluşmaktadır. Her oyuncu, iki kelime, dokuz cümle ve iki paragrafı seslendirmektedir. Tüm veri seti yaklaşık 30 dakikalık konuşmadan meydana gelmektedir. 20 dinleyici ile testler yapılmıştır. Şaşkınlık ile mutluluk, nötral ile üzgün arasında karışıklık olsa da %67,3 oranında doğruluk sağlanmıştır. [34]

2.1.6 Polish Emotional Speech

Veri seti, üç grup konuşmacının (profesyonel oyuncu, amatör oyuncu ve amatörler), günlük yaşama ait 10 adet kısa cümleyi, kızgınlık, üzüntü, mutluluk, korku, tiksinti, şaşkınlık ve nötral duygu durumlarında seslendirmesinden oluşmaktadır. Amatör veri seti kayıtları bitince, profesyonel aktörlerin kayıtları ile devam edilmektedir. 2000'den fazla amatör kaydı yapılmıştır. 33 tanesi müzik eğitimi almış, 27 tanesi ise Lehçe bilmeyen 202 dinleyici ile sistem test edilmiştir. Cinsiyet ve müzik eğitiminin sınıflandırmada etkili olduğu görülmekle beraber, kadınların skoru erkeklerinkinden %5 fazladır. Kadınlar tarafından, duygu durumları daha iyi anlaşılırken, erkekler nötral durumları daha iyi saptamıştır. En kötü sonuçlar nefret duygusuna aittir. Nötral ve üzüntü durumlarında da hatalı sınıflandırma olmuştur. [35]

2.1.7 Tez için Oluşturulan Veri Seti (TOVS)

Literatürde farklı diller ile gerçekleştirilmiş duygu çıkarım çalışmalarına bakıldığında her dilin yapısı gereği farklı ses özelliklerinin kullanılması gerektiği tespit edilmiştir. Tez kapsamında hedeflenen Türkçe için duygusal durum analizi yapabilen bir sistemin geliştirilmesidir. Ancak hali hazırda bu tip bir çalışmada kullanılabilecek geniş kapsamlı ve etiketlenmiş bir veri seti bulunmadığı için çalışma kapsamında yeni bir veri seti oluşturulması ihtiyacı ortaya çıkmıştır. Yeni veri setinin oluşturulmasında mümkün olduğunca farklı özelliklere sahip kişilerin ses kayıtlarının yer alması hedeflenmiştir. Bu tip bir veri setinin oluşturulmasında literatürde izlenen üç farklı yöntem bulunmaktadır. Bu yöntemlerden birincisi bireylerden farklı duygusal durumlarına ait ses kayıtlarının toplanmasıdır. Bu tip çalışmanın alacağı zaman ve kişilerden alınması gerekli izinlerin getireceği bürokratik engeller göz önünde bulundurularak veri setinin oluşturulmasında bu yöntem seçilmemiştir. Diğer bir yöntem ise seslendirme sanatçıları veya tiyatrocular gibi deneyimli kişilerden ses kayıtları alınmasıdır. Bu yöntem ile ilgili olarak bir fizibilite çalışması yapılmış ve veri setinin istenilen büyüklüğe ulaşmasının maliyetinin yüksek olacağı tespit edilmiştir. Dolayısıyla tez kapsamında sıradan kişilere ait mevcut ses kayıtlarının etiketlenerek kategorize edilmesi yöntemi kullanılmıştır. Bu kapsamda bir radyo programlarının Internet üzerinde mevcut olan ses kayıtları kullanılmıştır. Radyo programında programa bağlanan kişiler, isimlerini ve yaşlarını belirterek kendilerini tanıttıktan sonra radyo sunucusu ile belirli bir konu hakkındaki fikirlerini paylaşmaktadırlar. Ses kayıtlarının oluşturulmasında ilk olarak radyo programına ait toplam 307 konuşmacının ses kaydı incelenerek, katılımcıların ses kayıtları çıkartılmıştır. Ses kayıtlarının toplam uzunluğu yaklaşık 63.500sn'dir. Katılımcıların ses kayıtları yaş ve cinsiyet özelliklerine göre isimlendirilen dosyalarda tutulmuştur. İkinci adımda ise katılımcıların konuşmaları duygusal durumlarına bağlı olarak alt parçalara bölünmüştür. Dosya isimlendirme formatı Şekil 2.1'de açıklanmaktadır. ¹



Şekil 2.1 Dosya İsimlendirme Yöntemi

Belirtilen bu çalışmaların sonucunda 158 erkek ve 149 kadına ait toplam 307 adet ses kaydı toplanmıştır. Bu kayıtların cinsiyet ve duygusal durumlara göre dağılımı

¹A:Anger (Kızgın), D:Desperate (Umutsuz), E:Excited (Heyecanlı), N:Neutral (Nötr), S:Sad (Üzgün), U:Unknown (Bilinmeyen)

Tablo 2.1’de verilmiştir. Heyecanlı duygu durumunda kayıt sayısı fazla olduğundan, testlerde bu duygu durumuna kaymalar olmuştur. Bu nedenle heyecanlı duygu durumu veri sayısı optimize edilerek testler tekrarlanmıştır.

Tablo 2.1 TOVS içerisindeki örneklerin cinsiyet ve duygulara göre dağılımı

Cinsiyet	Kızgın (Angry)	Umutsuz (Desperate)	Heyecanlı (Excited)	Mutlu (Happy)	Nötr (Neutral)	Üzgün (Sad)
Erkek	106	9	481	159	116	25
Kadın	259	12	522	193	244	68
Toplam	365	21	1003	352	360	93

2.2 SES KAYITLARININ İŞLENMESİ ve ÖZELLİK ÇIKARIMI

2.2.1 Sesin Özellikleri

Ses özellikleri genel olarak iki kategoriye ayrılabilir. Bunlar fiziksel ve algısal özelliklerdir. Fiziksel özellikler, akustik sinyalin ölçülebilir nitelikleriyle ilgilidir ve insani algılara bağlı değildir. Algısal özellikler ise, sesin öznel algısı ile ilgilidir ve işitsel modeller kullanılarak ölçülmelidir. Sesin özellikleri ile ilgili aşağıda verilen bilgiler [37] çalışmasından alınmıştır.

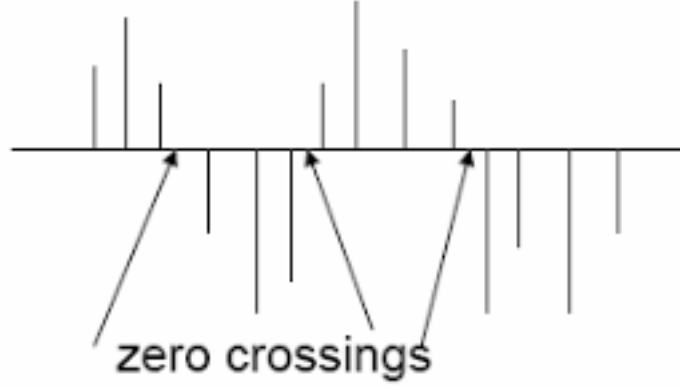
2.2.1.1 Fiziksel Özellikler

Sinyalin geçici ya da tayf niteliklerinin belirli açılardan yakalanan, düşük seviyeli sinyal parametreleridir. Bu özelliklerden bazıları, algısal olarak harekete geçirilse de, fiziksel özellik olarak sınıflandırılırlar. Çünkü, onlar doğrudan dalga şiddeti veya kısa zamanlı tayfi değerler olarak hesaplanır.

2.2.1.1.1 Sıfırdan Geçiş Sayısı Ses sinyalinin sıfır noktasından kaç defa geçtiğinin hesaplanması ile elde edilen bir özelliktir. Sinyalin sıfırdan geçme sayısının hesaplanmasında denklem 2.1 kullanılır. Denklem içerisindeki X_r ses sinyali içerisindeki r . çerçeveyi, N çerçeve içerisindeki toplam örnek sayısını, n ise bu çerçeve içerisinde üzerinde işlem gerçekleştirilen örneğin numarasını belirtir. Bu özelliğin grafik ile gösterimi de Şekil 2.2 ’de verilmiştir.

$$ZCR_r = \frac{1}{2} \sum_{n=1}^N |\text{sign}(x_r(n)) - \text{sign}(x_{r-1} - (n))| \quad (2.1)$$

$$\begin{cases} 1, & x \geq 0; \\ -1, & x < 0 \end{cases} \quad (2.2)$$



Şekil 2.2 Sıfırdan Geçiş Sayısı [38]

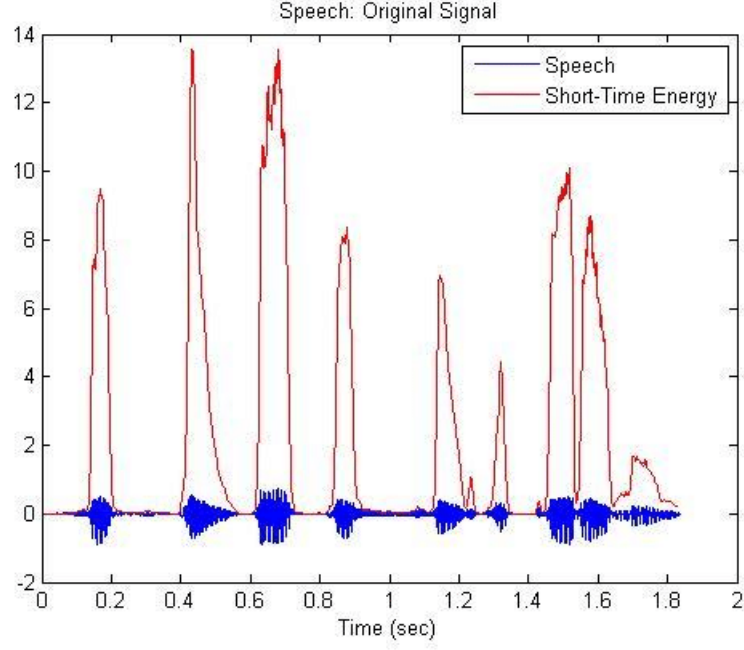
2.2.1.1.2 Kısa Zamanlı Enerji (Short Time Energy) Sinyalin zamansal zarfını temsil eder ve çerçeve içerisindeki genlik değerlerinin karelerinin ortalaması alınarak denklem 2.3 yardımı ile hesaplanır. Sinyal içeriğinin değerlendirilmesinde anlık değerinden ziyade zamana bağlı olarak değişimi kullanılabilir. Bu özelliğin grafik ile gösterimi de Şekil 2.3 'de verilmiştir.

$$E_r = \frac{1}{N} \sum_{n=1}^N |x_r(n)|^2 \quad (2.3)$$

2.2.1.1.3 Bant Seviyesi Enerjisi (Band Level Energy) Sinyalin belirli bir frekans bölgesi içinde sahip olduğu enerjiyi ifade eder ve denklem 2.4 kullanılarak hesaplanır. k frekans değerini, X_r sinyalin frekans bölgesindeki halini, $W[k]$ ise ilgilenilen frekans aralığının hesaplanabilmesi için kullanılan ağırlık değerlerini ifade eder.

Bant seviyesi enerjisindeki ani geçişler sinyalin tayfsal enerji dağılımında veya tınısında bir değişikliği ifade eder. Dağılımın daha efektif olması ve işitme algısına benzerliğini arttırmak için logaritmik olarak hesaplanır.

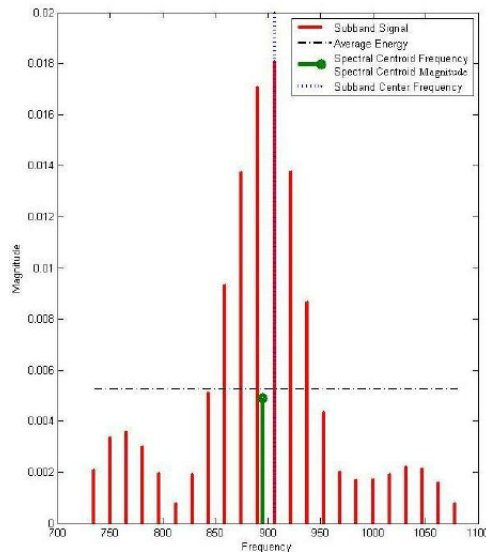
$$E_r = \frac{1}{N} \sum_{k=1}^{N/2} (X_r[k]W[k])^2 \quad (2.4)$$



Şekil 2.3 Kısa Zamanlı Enerji [39]

2.2.1.1.4 Spektral Ağırlık Merkezi (Spectral Centroid) Genlik tayfının ağırlık merkezidir ve ses sinyaline ait bir çerçeve içerisindeki ağırlıklı ortalama frekans değerini ifade eder. Ağırlık değerleri olarak ilgili frekansın çerçeve içerisindeki genlik değeri kullanılır. Bu özelliğin grafik ile gösterimi de Şekil 2.4 'de verilmiştir.

$$C_r = \frac{\sum_{k=1}^{N/2} f[k]|X_r[k]|}{\sum_{k=1}^{N/2} |X_r[k]|} \quad (2.5)$$

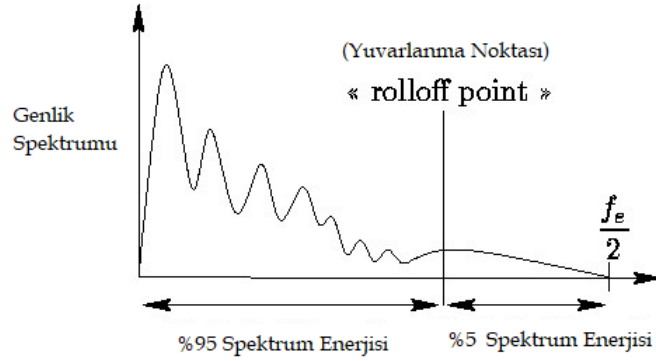


Şekil 2.4 Spektral Ağırlık Merkezi [40]

2.2.1.1.5 Spektral Yuvarlanma Tayf şeklinin kaba bir tanımlayıcısıdır. Kendisinden daha küçük frekanslara ait genlik değerlerinin toplamı, tüm frekanslara ait genlik değerlerinin toplamının %85'inden büyük olan frekans değeri olarak ifade edilir. Bu özelliğin grafik ile gösterimi de Şekil 2.5 'de verilmiştir.

$$R_r = f[k] \quad (2.6)$$

$$\sum_{k=1}^K |X_r[k]| \leq 0,85 \sum_{k=1}^{N/2} |X_r[k]| \quad (2.7)$$

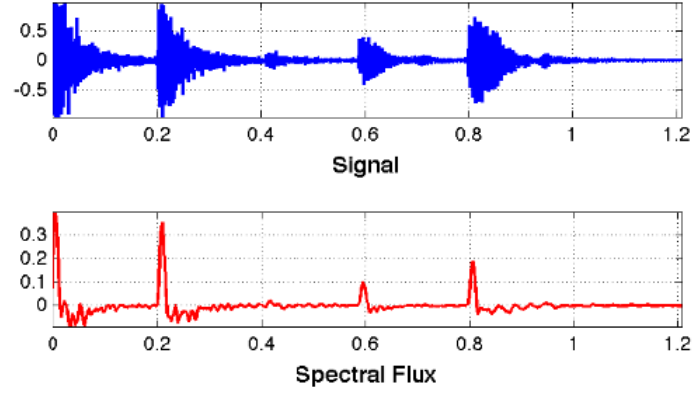


Şekil 2.5 Spektral Yuvarlanma [41]

2.2.1.1.6 Spektral Akı Çerçeveler arasındaki genlik değerlerinin değişimi olarak ifade edilir ve tayf içerisindeki yerel değişimi gösterir. Spektral akı değerindeki yüksek bir değer, genlikte anlık bir değişimi ve dolayısıyla kaydadeğer algısal bir değişim ifade eder. Spektral akı değeri denklem 2.8 kullanılarak hesaplanır. Bu özelliğin grafik ile gösterimi de Şekil 2.6 'da

$$F_r = \sum_{k=1}^{N/2} (|X_r[k]| - |X_{r-1}[k]|)^2 \quad (2.8)$$

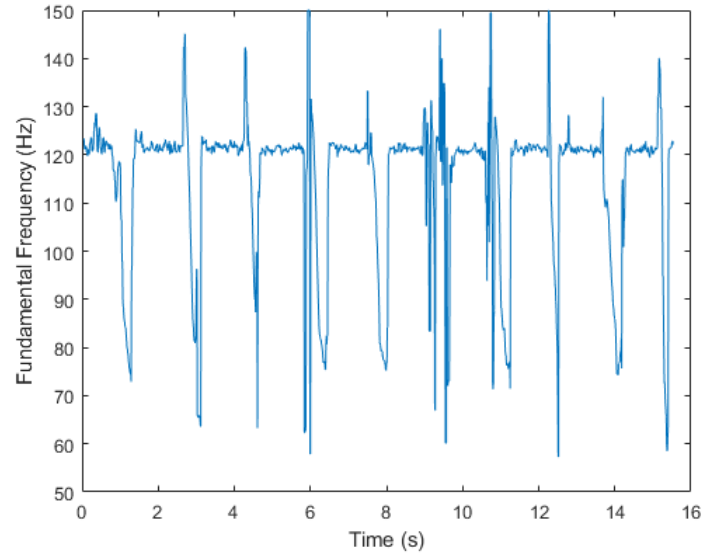
2.2.1.1.7 Temel Frekans (F0) Temel frekans, dalga formunun zaman eksenindeki periyodikliğinin ölçülmesi ile hesaplanır. Ayrıca sinyalin frekans tayfındaki ilk harmoniğinin frekansı kullanılarak da tespit edilebilir.



Şekil 2.6 Spektral Akı [42]

Müzik enstrümanları ve insan sesi için temel frekansın tespit edilmesi, sinyalin her periyodundaki değişimler ve temel frekansın diğer harmoniklere göre zayıf olma ihtimalinden dolayı oldukça zordur.

Zaman ekseninde periyodikliğin bulunmasında sinyalin özilinti fonksiyonu (autocorrelation function, ACF) ile Denklem 2.9'a göre hesaplanır. Bu özelliğin grafik ile gösterimi de Şekil 2.7 'da verilmiştir.

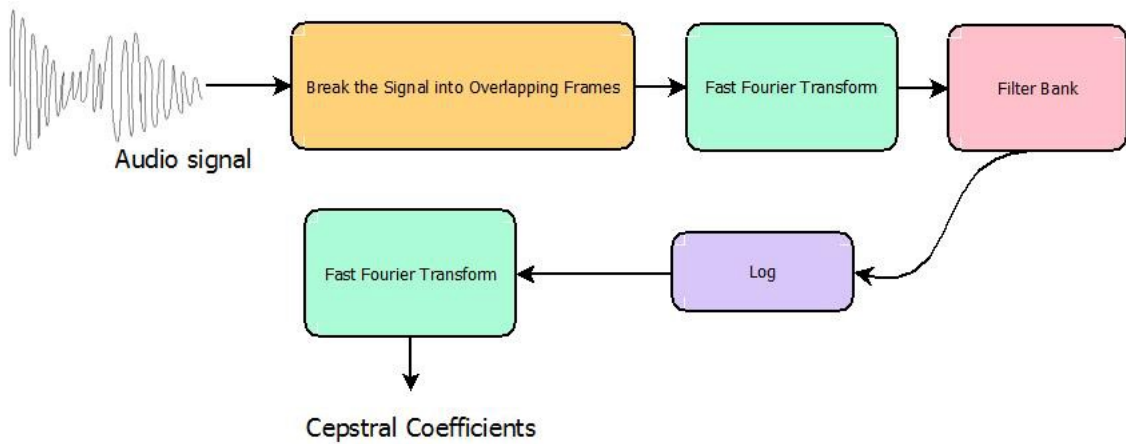


Şekil 2.7 F0 [43]

$$R_r = \frac{1}{N} \sum_{n=0}^{N-1} (x_r[n]x_r[n+r]) \quad (2.9)$$

2.2.1.1.8 Mel Frekanslı Güç Spektrumu Katsayıları (Mel-Frequency Cepstral Coefficients, MFCC) MFCC, kısa zaman spektrumu zarfının kompakt bir sunumunu sağlayan algısal olarak motive edilmiş özelliklerdir.

Cepstral, spectral (tayfsal) kelimesindeki spec kısmının tersten yazılması sonucu elde edilmiş bir kelimedir. Bunun nedeni cepstrum'un tayfsal bantlardaki değişimin oranını ifade etmesidir. Herhangi bir sinyalin frekans tayfına açılması durumunda, sinyal içerisindeki periyodik bileşenlerin frekans tayfında keskin tepe noktaları olarak görülecektir. Şekil 2.8'de bir sinyalin zaman tayfındaki hali ve FFT kullanılarak cepstral katsayıların elde edilişi modellenmiştir.



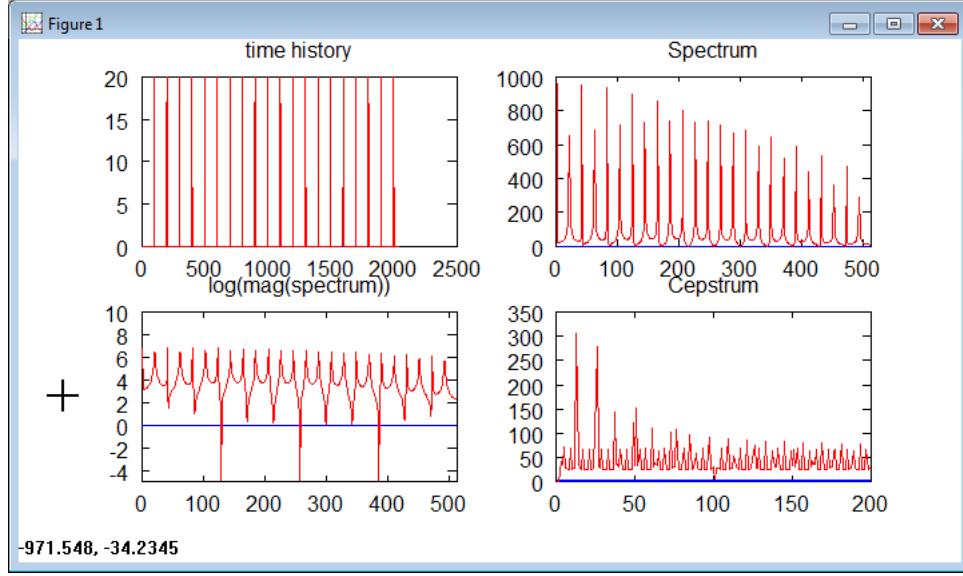
Şekil 2.8 MFCC [44]

Şekil 2.8'de MFCC yöntemindeki adımlar görülmektedir: Öncelikle ses sinyali birbiri üzerine örtüşen küçük pencerelere ayrılır. Pencerelerin örtüştürülmesinin nedeni ses sinyali frekans tayfına dönüştürüldüğünde devamlılığı sağlayabilmektir. Her pencere içerisindeki örneklerle FFT uygulanarak frekans tayfına dönüşüm yapılır. Bu yüzden pencere boyutunun ikinin üssü olması tercih edilir. Elde edilen FFT katsayılarına üçgen yapıdaki Mel filtreleri uygulanır.

FFT dönüşümü sonrasında elde edilen katsayı değerlerinin logaritması alınır. Sonrasında ayrık kosinüs dönüşümü (DCT - discrete cosine transform) uygulanır. Elde edilen değerler cepstrum olarak ifade edilen MFCC katsayılarıdır. Bu katsayılar ne zaman ne de frekans tayfında değildirler. Bu yüzden de cepstrum olarak isimlendirilmişlerdir. Şekil 2.9'da bir sinyale MFCC adımları sonucunda elde edilen ara ve sonuç değerler gösterilmektedir.

2.2.1.2 Algısal Özellikler

Ses verisinden konuşmacı tanıma, sesin algısal özelliklerine bağlı olarak gerçekleştirilir. Algısal özellikler bölümlleme ve sınıflandırma işlemlerinde kullanılacak



Şekil 2.9 MFCC [45]

temel alternatiflerdir. Sesin uyandırdığı psikolojik hissiyat kabaca, ses şiddeti (loudness), perde (pitch) ve tını (timbre) olarak kategorize edilebilir.

Ses şiddeti ve perde özellikleri artan bir genlik ölçeğinde sıralanabilen yapıdadırlar. Ancak tını çeşitli bileşenlerden oluşan çok boyutlu bir yapıdır ve aynı ses şiddeti ve perdedeki farklı seslerin ayırt edilmesini sağlar.

İşitme sisteminin sayısal bir modeli kullanılarak ses dalgasının kısa süreli algısal özellikleri hesaplanabilir. Özellikle ses şiddeti, perde ve özelliklerin zaman içerisindeki dalgalanmaları oldukça yaygın olarak kullanılan algısal özelliklerdir.

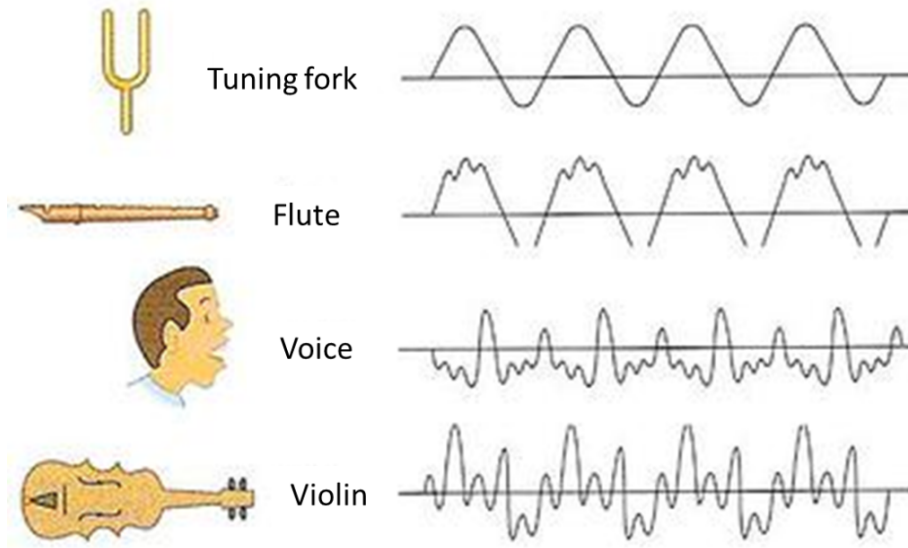
2.2.1.2.1 Ses Şiddeti Ses şiddeti, ses sinyalinin gücünün algılanmasıdır. Beklendiği üzere sesin yoğunluğu ile ilgili olmakla beraber sesin sürekliliğine ve tayfına da bağlıdır. Psikolojik açıdan, hissedilen ses şiddeti, sesin işitsel sinirlerde oluşturduğu toplam aktivitesi ile belirlenir. Ses şiddeti ile sesin yoğunluğu arasında doğrusal olmayan bir ilişki vardır. [46] çalışmasında, Fado(Portekiz halk müziği) şarkıları söyleyen şarkıcıların sesleri incelenmiştir. Ses şiddetinin, kadınlara oranla erkeklerde, profesyonellere oranla amatörlerde, yaşlılara göre gençlerde daha yüksek olduğu gözlemlenmiştir.

2.2.1.2.2 Perde Perde, algısal bir özellik olmasına rağmen, temel frekansın fiziksel niteliği ile yakından ilgilidir. Öznel perde değişiklikleri F_0 logaritması ile ilgilidir, bu nedenle müzikte sabit perde değişimi, sabit temel frekans oranını ifade eder. Fado şarkıcılarındaki ses perdesinin kadınlar için alçak, erkekler için yüksek olduğu

gözlemlenmiştir. Ayrıca kadınlarda keskin, erkeklerde ise derin olarak skorlanmıştır. Gençlere göre yaşlılarda, amatörler göre de Fado şarkıcılarında ses perdesi daha derin olarak skorlanmıştır. [46]

2.2.1.2.3 Tını Tını farklı ses kaynaklarını ayırt etmemizi sağlayan ses özelliğidir. Örneğin farklı müzik enstrümanlarından çıkan aynı şiddetteki aynı notanın hangi müzik aletine ait olduğunu anlayabilmemizi sağlayan özelliktir. Benzer şekilde aynı sözleri söyleyen farklı kişileri de tanımamızı sağlayan kişilerin seslerinin farklı tınlara sahip olmasıdır.

Ses sinyali farklı frekanslardaki titreşimlerden oluşur. Çoğu zaman ses sinyali birden fazla frekans içerir. Ses sinyali içerisindeki en düşük frekans temel frekans olarak isimlendirilir. Bu değer aynı zamanda sesin perdesini ifade eder. Geri kalan frekanslar ise üst ton veya harmonik olarak isimlendirilir. Bu harmoniklerin oluşturduğu karışım sayesinde ses kaynağını ayırt edebiliriz. Şekli 2.10'da aynı temel frekansa sahip olan farklı ses kaynaklarının oluşturduğu ses sinyali verilmiştir.



Şekil 2.10 Tını [47]

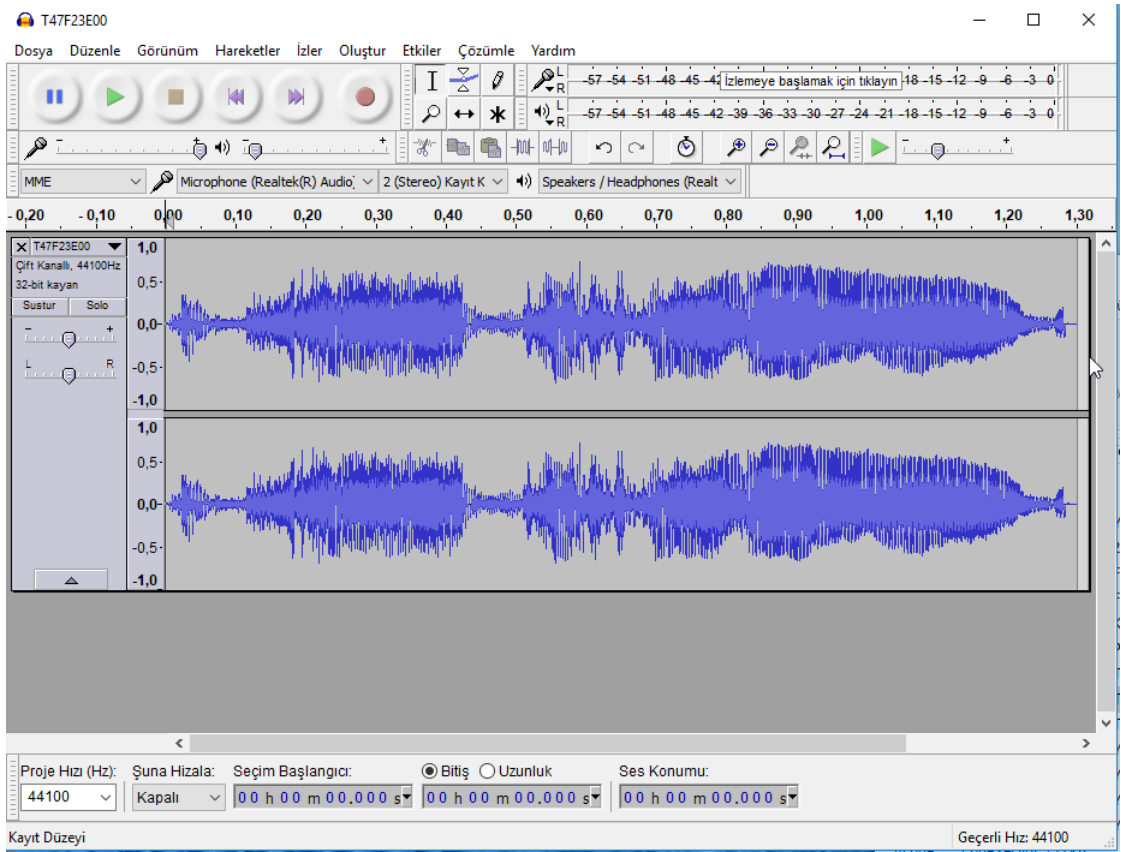
2.2.2 Özellik Çıkarımı

Bu bölümde, ses kayıtlarının sınıflandırılabilmesi için kullanılacak özelliklerin belirlenmesi ve bu özelliklerin kayıtlardan çıkartılabilmesi için yapılan işlemler açıklanmaktadır.

2.2.2.1 Audacity ile seslerin parçalara ayrılması

Audacity açık kaynak kodlu ve birçok işletim sistemine uyumlu bir ses editörü ve kaydedicisidir. Bu program sayesinde ses dosyaları istenilen boyutlarda bölünüp, kaydedilebilir. Ekran görüntüsü Şekil 2.11’de verilmiştir.[48]

TOVS’yi oluşturmak için bir araya getirilen ses kayıtları radyo programlarından alınmıştır ve iki ya da daha çok kişinin farklı duygularda olabilen karşılıklı konuşmalarını içermektedir. Ses kayıtlarının konuşmacı bazında ayrıştırılması, arada konuşma içermeyen kısımların temizlenmesi ve ardından duygu bazında parçalanması, son olarak da bu kayıtların etiketlenmesi gerekmektedir. Özellik çıkarımı için ses kayıtları üzerinde yapılması zorunlu bu işlemler için Audacity uygulamasından faydalanılmıştır. Uygulama ile ses kaydı dinlenirken, konuşma parçaları başlangıç ve bitişlerinden kesilerek ayrı dosyalar olarak kaydedilmiş, daha sonra bu ses dosyaları içerisinde farklı duygular tespit edilirse, yeniden parçalanması yoluna gidilmiştir.



Şekil 2.11 Audacity ekran görüntüsü

2.2.2.2 JAudio ile Özellik Çıkarımı

JAudio, özellik çıkarımı için geliştirilen bir uygulama çatısıdır. Kendi içerisindeki algoritmalar sayesinde sesin 27 özelliğinin çıkarımını yapmaktadır. Bu özelliklerden

bazıları: Zero Crossing Rate, MFCC, Spectral Flux vb'dir. Çıktıları ACE XML veya ARFF olarak üretebilmektedir. [49]

Audacity ile parçalanıp etiketlenmiş olan ses dosyaları, 0,5sn, 1sn, 1,5sn ve 2sn lik bölümlere ayrılarak JAudio'ya verilmiştir. Ses kayıtları için JAudio'da bulunan 22 temel özelliğin ve bunlarından alt özelliklerinden oluşan 204 özelliğin değerleri elde edilerek veritabanına yazılmıştır. Daha sonra sınıflandırma için bu değerler belirli senaryolar dahilinde seçilerek .arff dosyaları üretilmiştir.

2.3 SINIFLANDIRMA

2.3.1 Sınıflandırma Yöntemleri

Literatürde çok sayıda sınıflandırma yöntemi bulunmaktadır. Bu bölümde, tez çalışması için hazırlanan veri seti üzerinde yapılan testlerde daha iyi sınıflama başarımı sağlayan yöntemler ile başarımı karşılaştırma amacıyla literatürde yaygın kullanılan yöntemler hakkında bilgi verilmiştir.

2.3.1.1 Bayes Ağı (Bayes Net)

Bayes ağları, verilerden ve / veya uzman görüşünden modeller oluşturmak için kullanılabilecek bir olasılıksal grafik modeldir. Tahmin, anomali tespiti, teşhis, otomatik bakış açısı, muhakeme, zaman serileri tahmini ve belirsizlik altında karar verme dahil olmak üzere çok çeşitli görevler için kullanılabilirler. Bayes ağları, olasılık dağılımlarından inşa edilirler ve aynı zamanda tahmin ve anomali saptama, akıl yürütme ve tanılama, belirsizlik ve zaman serileri tahmini altında karar verme için olasılık yasalarını kullanırlar.[50]

Bayes Ağı yöntemi çeşitli arama algoritmaları ve kalite ölçüm araçları kullanmaktadır. [51] Bayes Ağı, sınıflandırmasına ait ağ yapıları, durumlara bağlı olasılık dağılımları gibi, temel sınıfta çeşitli veri yapıları sunmaktadır. Bu temel sınıf aynı zamanda Bayes Ağı öğrenme algoritmalarına ortak teşkil edecek K2 ya da B gibi ortamları da sunmaktadır. Bayes Ağı, ağın duruma bağlı olasılık tablosunun tahminlenmesi için 4 çeşit algoritma kullanır. Arama işlemi, K2 veya TAN algoritmaları ile yapılmakta ya da daha gelişmiş yöntem ve algoritmalar kullanılabilmektedir. Bu yöntem ya da algoritmalar temel olarak tepe - tırmanış tekniği (hill-climbing), benzetimli tavlama (simulated annealing), tabu arama (tabu search) ve genetik algoritmaları kullanmaktadır. Tercihen arama hızı Aktif Rehber (Active Directory - AD) ağaçları kullanılarak arttırılabilir. Bununla birlikte, ağın yapısını öğrenebilmek için durumsal bağımsızlık testlerini kullanan 2 algoritma bulunmaktadır. Buna alternatif olarak

network yapısı XML dosyasından da yüklenebilmektedir. [52]

2.3.1.2 Karar Tablosu

Karar tablosu (Decision Table), hem test hem de gereksinim yönetiminde kullanmak için mükemmel bir araçtır. Temel olarak, karmaşık iş kurallarıyla uğraşırken gereksinimleri formüle etmek için yapılandırılmış bir alıştırmaadır. Karar tabloları karmaşık mantığı modellemek için kullanılır. Tüm olası koşul kombinasyonlarının göz önüne alındığını ve şartların gözden kaçırılması durumunda bunu görmenin kolay olduğunu görmeyi kolaylaştırabilirler. [53]

Weka’da ise Karar Tablosu - Basit karar tablosunu oluşturmak ve kullanmak için kullanılan sınıflandırıcılardandır. [54] Karar tablosu, en iyi ilk arama işlemi ile özelliklere ait alt kümeleri değerlendirir ve bunun için çarpaz onaylama kullanır. Karar tablosunda bir seçenek, tablonun içermediği her bir örneğe ait sınıfı belirlemek için en yakın komşu metodunu kullanır. Böylelikle tablonun genel büyük kısmını referans almak yerine, benzer özellikteki kümelere odaklanır. [52]

2.3.1.3 Gizli Markov Modeli

Gizli Markov Modeli (Hidden Markov Model-HMM), bir dizi gözlemlere ait olasılık dağılımlarını temsil eden bir araçtır. Gizli Markov Modeli; modellenmekte olan sistemin Markov Süreci ile gözlemlenmemiş durumlarını içeren istatistiksel bir modeldir. Bu model en basit dinamik Bayes ağı olarak gösterilebilir. [55]

2.3.1.4 IBk

Örnek tabanlı öğrenme (Instance based learning), spesifik örnekleri kullanarak sınıflandırma tahmininde bulunur. Instance based learning algoritmaları, belirli örneklerden türetilen bir dizi soyutlamayı sürdürmezler. Bu yaklaşım, geniş depolama ihtiyacı olan en yakın komşu algoritmasını geliştirir. [56]

2.3.1.5 J48

J48 ile, her bir kuralı takip eden parantez içerisinde yer alan numaralar, bu numaradan sonra gelen yanlış sınıflandırılan bir kural varsa eğer bunlara ait örneklerin sayısını verir. J48’in ek özellikleri; eksik değerlerin muhasebesi, karar ağaçları sürekli özellik değer aralıkları, kuralların çıkarılması, vb. dir. [57]

2.3.1.6 k-Nearest Neighbors

KNN, girdileri kategorilere ayıran basit bir makine öğrenme algoritmasıdır. KNN, el yazısı mektupların tanımlanmasında, gen açıklaması analizlerinde kullanılabilir. Örneğin, hangi genin hangi karakteri belirlediğinin tespitinde kullanılır. KNN Parametrik Olmayan (Non-Parametric) özelliği ile, olasılık dağılımına ait herhangi bir tahminleme yapmaz. Girdi parametreleri belli olmayan uygulamalar için bu oldukça faydalıdır. Bu özellik parametrik algoritmalara göre daha sağlam bir yapı sunmaktadır. Bunun tam tersinde yer alan parametrik makine öğrenme algoritmaları, parametrik olmayanlara göre biraz hata üretme özelliği taşımaktadır. Bu hata, girdi olasılıklarının hesaba alınmasının, karar verme üzerindeki etkisinden kaynaklanmaktadır. [58]

2.3.1.7 Locally Weighted Learning (LWL)

LWL, Naive Bayes'in bölgesel olarak ağırlaştırılmış versiyonudur. Deneysel sonuçlar bu yöntemin standard Naive Bayes'e göre başarımı nadiren düşürdüğünü, genellikle arttırdığı göstermektedir. Bu yöntemin, Naive Bayes'i iyileştirmeye çalışan diğer yöntemlere göre en büyük avantajı kavramsal ve sayısal olarak basitliğidir. [59]

2.3.1.8 Naive Bayes

Naive Bayes, olasılıksal bilgiyi öğrenmek, kullanmak ve sunmak için anlam bilimi ile beraber basit bir yaklaşım sağlamaktadır. Bu yöntem, sınıf bilgisi içeren eğitim örneklerinin ve test örneklerinin sınıfını doğru tahmin etmek için tasarlanmıştır. Bayesian ağının özel bir biçimi olarak görebildiğimiz bu sınıflayıcı, dayandığı varsayımlar nedeniyle "naive" olarak isimlendirilmiştir. Bu varsayımlar: özellikle, öngörücü özelliklerin, verilen sınıf şartlarından bağımsız olduğunu ve gizli özelliklerin de tahmin sürecini etkilemediğini varsayar. [60]

2.3.1.9 Random Forest

Random Forest, ağaç öngörücülerin bir kombinasyonudur. Öyle ki, her ağaç bağımsız olarak örneklenen bir rastgele vektörün değerlerine bağlıdır ve tüm ağaçlar için aynı dağılım geçerlidir. Ormandaki ağaçların sayısı arttıkça, genelleme hatası bir sınır değere yaklaşır. Ağaç sınıflayıcı ormanın genelleme hatası, ormandaki ağaçların bireysel gücüne ve aralarındaki korelasyona bağlıdır. Ormandaki ağaç sayısı belirli bir sayıya ulaştığında, en popüler sınıf için oylama yapılır. Buna da "random forest" denir. [61]

2.3.2 Weka ile Sınıflandırma

Waikato Environment for Knowledge Analysis (Bilgi Analizi için Waikato Ortamı) ibaresinin kısaltması olan Weka, adını Yeni Zelenda'ya özgü bir kuş türü olan Weka (veya Whoden)'dan almaktadır. Yeni Zelenda'daki Waikato Üniversitesinde, Java programlama dili ile yazılmış bir makine öğrenmesi ve veri madenciliği uygulamasıdır. Verinin çeşitli algoritmalar aracılığı ile sınıflandırılmasına imkan sağlar. Veri analizi ve modelleme tahmini için algoritmaların gerçekleştirimlerini ve bu işlemleri kolayca yapabilmek için kullanışlı arayüzler içermektedir. Veri madenciliği ve makine öğrenmesi ile ilgili de çok çeşitli algoritmalar barındırmaktadır. Açık kaynak kodlu ve ücretsiz erişilebilen bir uygulamadır. Ayrıca tüm platformlarda çalışabilmektedir. [62]

3

DENEYSEL BULGULAR

2.1. Bölümde anlatıldığı şekilde veri seti oluşturulduktan sonra, toplanan ses kayıtları Audacity programı ile içerdikleri duygulara göre parçalara ayrılarak etiketlenmiştir. Etiketlemede 6 duygu kullanılmıştır. Bunlar, Nötral (Neutral), Heyecanlı (Excited), Mutlu (Happy), Kızgın (Angry), Üzgün (Sad), Umutsuz (Desperate) duygularıdır.

Ses kayıtlarının tümü parçalara ayrılp etiketlendikten sonra, JAudio programı kullanılarak etiketlenen ses kayıtlarının özellik çıkarımı yapılmış, ardından her bir ses kaydına ilişkin elde edilen özellik değerleri, diğer bilgiler ile birlikte veritabanına kaydedilmiştir.

3.1 Kullanılan Ses Özellikleri

JAudio programı ile özellik çıkarımı yapmak için, 22 özellik ve onların türevleri olmak üzere sesin elde edilebilen özelliklerinin tamamı(204) kullanılmıştır. 22 temel özellik aşağıda listelenmiştir.

- Spectral Centroid
- Spectral Rolloff Point
- Spectral Flux
- Compactness
- Spectral Variability
- Root Mean Square
- Fraction Of Low Energy Windows
- Zero Crossings
- Strongest Beat

- Beat Sum
- Strength Of Strongest Beat
- Strongest Frequency Via Zero Crossings
- Strongest Frequency Via Spectral Centroid
- Strongest Frequency Via FFT Maximum
- MFCC
- LPC
- Method of Moments
- Partial Based Spectral Centroid
- Partial Based Spectral Flux
- Peak Based Spectral Smoothness
- Relative Difference Function
- Area Method of Moments

Bu özelliklerin hangilerinin kullanılacağına karar verebilmek için Weka ile veri seti üzerinde analizler yapılmıştır. Weka tarafından sağlanan özellik seçimi (Feature selection) yöntemlerinden Correlation-based Feature Selection Subset Evaluation(CfsSubsetEval), Information Gain Attribute Evaluator(InfoGainAttributeEval), PrincipalComponent yöntemleri ile indirgenen özellik kümeleri ve 204 özelliğin tümü kullanılarak yapılan sınıflandırma başarımları kıyaslanmıştır. Özellikler üzerinde kısıtlama yapıldığında, tüm özelliklerin kullanımına göre sınıflandırma başarımında bir artış gözlenmediğinden 204 özelliğin tamamının kullanılmasına karar verilmiştir. Özellik seçimi için Weka üzerinde kullanılan sınıflandırıcı ve özellik seçimi filtreleri ile kısıtlanan özellikler ile elde edilen sınıflandırma başarımları karşılaştırmaları Tablo 3.1, 3.2 ve 3.3’de gösterilmiştir.

3.2 Kullanılan Sınıflandırıcılar

Tablolarda yer alan 6 sınıflandırıcı ile yapılan testlerin sonuçları değerlendirilerek, en iyi başarımları sergileyen Random Forest sınıflandırıcısı ve literatürde en yaygın yöntem olmasının yanı sıra BES veri setinde ikinci en iyi sonuçları sunan Naive Bayes yöntemi

Tablo 3.1 Özellik Seçimi Sonuçları - TOVS içindeki Nötral, Mutlu, Heyecanlı ve Kızgın duyguları için

	Cfs Subset Eval	InfoGain Attribute Eval(150)	InfoGain Attribute Eval(50)	Principal Component (150)	Principal Component (50)	Filtresiz (50)
Naive Bayes	36,65	30,49	28,63	47,38	48,55	30,59
J48	66,99	67,10	65,24	63,63	63,99	66,97
Random Forest	75,92	75,24	72,33	72,59	72,89	74,91
Bayes Net	51,35	48,09	43,48	53,73	53,66	48,25
Ibk	74,92	76,60	71,17	72,40	72,20	76,77
Decision Table	53,16	53,24	52,41	51,52	51,53	53,09

Tablo 3.2 Özellik Seçimi Sonuçları - BES Veri Setindeki Nötral, Mutlu, Heyecanlı ve Kızgın duyguları için

	Cfs Subset Eval	InfoGain Attribute Eval(150)	InfoGain Attribute Eval(50)	Principal Component (150)	Principal Component (50)	Filtresiz (50)
Naive Bayes	69,38	67,49	62,83	66,52	68,58	67,26
J48	63,70	62,45	62,53	59,40	60,54	62,35
Random Forest	72,90	75,34	70,48	70,70	71,86	74,33
Bayes Net	69,87	68,22	64,52	62,04	62,04	68,22
Ibk	66,10	66,28	65,64	57,76	59,20	64,70
Decision Table	61,81	61,76	62,45	59,69	59,69	61,85

ile ilerlenmesi uygun görülmüştür. Random Forest haricindeki algoritmalarından hiçbirisi her iki veri seti için de eşdeğer seviyede başarımlar gösteremediğinden, değerlendirme için tercih edilmemiştir.

3.2.1 Naive Bayes

İlerleyen sayfalarda verilen tablolarda ve yorumlarında görüleceği üzere, Naive Bayes başarımlarını Random Forest'a göre TOVS'de üstünlük sergileyememiş; BES'de bazı senaryolarda az da olsa daha iyi sınıflandırma yapabilmektedir. Naive Bayes için en iyi sistem başarımları %79,73 olarak hesaplanmıştır. Bu sonuç, orta yaşlı kadın

Tablo 3.3 Özellik Seçimi Sonuçları - BES Veri Setindeki Nötral, İğrenme, Kızgın ve Üzgün duyguları için

	Cfs Subset Eval	InfoGain Attribute Eval(150)	InfoGain Attribute Eval(50)	Principal Compo- nent (150)	Principal Compo- nent (50)	Filtresiz (50)
Naive Bayes	74,47	68,82	67,96	72,36	72,13	69,85
J48	66,12	65,35	65,36	63,05	64,23	65,74
Random Forest	81,01	81,04	76,72	76,66	77,60	80,62
Bayes Net	73,46	70,32	67,40	66,27	66,27	70,29
Ibk	76,92	76,89	69,47	64,70	65,03	75,21
Decision Table	58,31	57,55	58,91	58,85	58,85	57,58

konuşmacıların, 2 sn'lik klipleri ve Nötr, Mutlu, Kızgın duyguları ile çalışıldığında elde edilmiştir. Tüm senaryolar için Navie Bayes ile elde edilen sonuçlar, bir sonraki bölümde verilmiştir.

3.2.2 Random Forest

Random Forest için en iyi sistem başarısı %78,31 olarak hesaplanmıştır. Bu sonuç, orta yaşlı kadın konuşmacıların, 2 sn'lik klipleri ve Nötr, Mutlu, Kızgın duyguları ile çalışıldığında elde edilmiştir. Tüm senaryolar için Random Forest ile elde edilen sonuçlar, bir sonraki bölümde verilmiştir.

3.3 Senaryoların Oluşturulması

Farklı veri kümelerinde sistem başarımını sınavabilmek için, duygulara, klip boyutlarına, cinsiyete ve yaşa göre senaryolar oluşturulup, bunların sonuçları analiz edilmiştir. Bu senaryoların oluşturulmasında 2.1. Bölümde verilen değerler dikkate alınmıştır.

Tablo 3.4'te görüleceği üzere Nötral olan kayıtların büyük bir kısmı Kızgın duygu durumuna, Kızgın olan kayıtların büyük bir kısmı da Nötral duygu durumuna kaymıştır. Mutlu olarak etiketlenmiş kayıtlarda ise hem Nötral hem Kızgın duygu durumuna kayma gözlenmiştir. 3 duygu için Random Forest başarısı %71,01 olarak hesaplanmıştır. Bu sonuç, doğru olarak etiketlenen kayıt sayısının, toplam kayıt sayısına bölünmesi ile elde edilmiştir. Tablo 3.4'de gösterilmiştir.

Tablo 3.4 Sınıflandırma Sonucundaki Duygu Dağılımı

a	b	c	classified as
1460	34	603	a = N
165	290	218	b = H
1099	137	3776	c = A
Accuracy of RandomForest:			%71.01

3.3.1 Duygulara Göre Oluşturulan Senaryolar

TOVS’de ve BES’de kayıt sayısı en fazla olan duygular, iki veri setinde ortak olan duygular ve iki veri setindeki tüm duygular ile senaryolar oluşturulmuş ve sınıflandırma başarımlar karşılaştırılmıştır. İki veri setinde Nötr, Kızgın, Mutlu duyguları ortak olup; Umutsuz, Heyecanlı, Üzgün duyguları TOVS’de, Korku, Tiksinti ve Sıkılma duyguları BES’de yer almaktadır. TOVS’de 3 duygu için, BES’de 4 duygu için Random Forest’in %71,01 ve %69,91’lik sonuç ile en iyi başarıyı sağladığı gözlemlenmiştir. Tüm sonuçlar Tablo 3.5 ve 3.6’de gösterilmiştir.

TOVS’de 3 duygu için en iyi sonuç alındığı, duygu sayısı arttıkça sınıflandırma başarımının zayıfladığı görülmektedir. Bunun nedeninin, konuşmacı bazında değişken sayıda duygu etiketli kayıt olması ve duyguların etiketlenmesindeki hatalar olabileceği düşünülmektedir.

Tablo 3.5 TOVS’de Duygulara Göre Random Forest ve Naive Bayes Sonuçları (%)

Duygulara Göre (TOVS)	3Duygu(Nötr, Mutlu, Kızgın)	4Duygu(Nötr, Heyecanlı,) Mutlu, Kızgın	Tüm Duygular(Nötr, Heyecanlı, Mutlu, Kızgın, Umutsuz, Üzgün,Bilinmeyen)
Naive Bayes	61,13	46,19	36,60
Random Forest	71,01	56,96	55,02

Tablo 3.6 BES’de Duygulara Göre Random Forest ve Naive Bayes Sonuçları (%)

Duygulara Göre (BES)	3Duygu(Nötr, Mutlu, Kızgın)	4Duygu(Nötr, Tiksinişmiş,) Mutlu, Kızgın	Tüm Duygular(Nötr, Tiksinişmiş, Mutlu, Kızgın, Korkmuş, Üzgün,Sıkılmış)
Naive Bayes	66,67	68,58	45,19
Random Forest	62,22	69,91	40,00

3.3.2 Klip Boyutuna Göre Oluşturulan Senaryolar

Her bir konuşmacının ses kayıtları duyguya göre ayrıştırıldıktan sonra, 1 ve 1,5 ve 2 sn’lik kliplere parçalanmıştır. Her bir klip boyu için her duygu durumu senaryosu tekrarlanarak klip boyunun sınıflandırma başarımına etkisi incelenmiştir. TOVS’de 3 duygu için, BES’de 4 duygu için 1,5 sn’lik ses kayıtlarında, Random Forest’in %72,16

ve %73,83'lük sonuç ile en iyi başarıyı sağladığı gözlemlenmiştir. Tüm sonuçlar Tablo 3.7 ve 3.8'de gösterilmiştir.

Klip boyunun başarıma etkisi, veri setindeki ses kayıtlarının standart özelliklerde olup olmamasına bağlıdır. TOVS ses kayıtları değişken boyutlarda olduğundan, en uygun klip boyunun belirlenmesi mümkün görülmemektedir. Öte yandan standart boyutta kayıt içeren BES için 1,5 sn'lik klipler optimum sonucu vermektedir.

Tablo 3.7 TOVS'de Klip Boyutlarına Göre Random Forest ve Naive Bayes Sonuçları (%)

Duygulara Göre (TOVS)	3Duygu(Nötr, Mutlu, Kızgın)	4Duygu(Nötr, Heyecanlı,) Mutlu, Kızgın	Tüm Duygular(Nötr, Heyecanlı, Mutlu, Kızgın, Umutsuz, Üzgün,Bilinmeyen)
Naive Bayes			
1000ms	61,13	46,19	36,60
1500ms	61,43	46,16	35,84
2000ms	65,45	48,95	41,62
Random Forest			
1000ms	71,01	56,96	55,02
1500ms	72,16	57,56	55,12
2000ms	72,09	52,87	49,93

Tablo 3.8 BES'de Klip Boyutlarına Göre Random Forest ve Naive Bayes Sonuçları (%)

Duygulara Göre (BES)	3Duygu(Nötr, Mutlu, Kızgın)	4Duygu(Nötr, Tiksinmiş,) Mutlu, Kızgın	Tüm Duygular(Nötr, Tiksinmiş, Mutlu, Kızgın, Korkmuş, Üzgün,Sıkılmış)
Naive Bayes			
1000ms	66,67	68,58	45,19
1500ms	70,97	71,96	48,10
2000ms	48,54	39,84	22,74
Random Forest			
1000ms	62,22	69,91	40,00
1500ms	65,32	73,83	43,94
2000ms	29,61	39,84	22,74

3.3.3 Cinsiyete Göre Oluşturulan Senaryolar

Cinsiyetin ekisini değerlendirebilmek için, veriler kadın ve erkek olarak bölünmüş, bu bölümler üzerinde ayrı ayrı eğitim ve test veri seti oluşturulup, cinsiyetler için klip boyutu ve duygu durumu senaryoları tekrarlanmıştır. TOVS'de kadın konuşmacıların, 3 duygu için olan, 2 sn'lik ses kayıtlarında, Random Forest'in %68,44'lük sonuç ile en iyi başarıyı sağladığı gözlemlenmiştir. Tüm sonuçlar Tablo 3.9'de gösterilmiştir.

Verilerin cinsiyete göre ayrılmasının sınıflandırma başarımını olumsuz etkilemesinin, veri setindeki kayıtların duygu-cinsiyet dağılımının dengesiz olmasından kaynaklandığı değerlendirilmektedir.

Tablo 3.9 TOVS’de Cinsiyete Göre Random Forest ve Naive Bayes Sonuçları (%)

Cinsiyete Göre (TOVS)	3Duygu(Nötr, Mutlu, Kızgın)		4Duygu(Nötr, Heyecanlı,) Mutlu, Kızgın		Tüm Duygular(Nötr, Heyecanlı, Mutlu, Kızgın, Umutsuz, Sad,Bilinmeyen)	
	Kadın	Erkek	Kadın	Erkek	Kadın	Erkek
Naive Bayes						
1000ms	59,71	48,87	47,51	30,29	39,51	25,60
1500ms	61,32	48,94	47,38	29,74	39,89	26,10
2000ms	63,31	49,17	50,46	29,84	41,40	25,10
Random Forest						
1000ms	65,76	54,24	51,31	35,32	47,63	32,94
1500ms	67,83	55,50	51,35	36,57	47,18	33,62
2000ms	68,44	55,40	51,85	36,66	48,55	33,00

3.3.4 Yaşa Göre Oluşturulan Senaryolar

Veri seti, genç(15-30), orta(31-50) ve yaşlı(50+) olarak 3 gruba ayrılmıştır. Cinsiyete ve yaşa göre eğitim ve test verileri oluşturulup klip boyutu ve duygu durumu senaryoları tekrarlanmıştır. Burada, gençlerin duygularının daha zor anlaşıldığı görülmektedir. TOVS’de orta yaşlı kadın konuşmacıların, 3 duygu için olan, 2 sn’lik ses kayıtlarında, Naive Bayes’in %79,73’lük sonuç ile en iyi başarıyı sağladığı gözlemlenmiştir. Yaşlı konuşmacı sayısı kadınlarda 2 olduğu için test edilememiştir. Erkeklerde ise 15 konuşmacı ile testler yapılmış ve konuşmacı sayısı azlığından dolayı çok düşük sonuçlar elde edilmiştir. Tüm sonuçlar Tablo 3.10’de gösterilmiştir.

Tabloda da görüleceği üzere, veri setinde yer alan ses kaydı sayısına bağımlı olmak kaydıyla, TOVS’de en iyi sonuçların orta yaş kadın ve genç erkek kayıtlarından elde edildiği görülmektedir. Her yaş grubunda farklı başarımların elde edilmesi, veri setinde yer alan ses kayıtlarının niteliği ve etiketleme hataları kaynaklı olabilir.

Tablo 3.10 TOVS’de Yaşa Göre Random Forest ve Naive Bayes Sonuçları (%)

Yaşa Göre(TOVS)	Naive Bayes			Random Forest		
3Duygu(Nötr, Mutlu, Kızgın)	1000 ms	1500 ms	2000 ms	1000 ms	1500 ms	2000 ms
Kadın						
Genç	51,69	54,65	51,41	56,66	61,24	56,69
Orta	75,06	77,68	79,73	77,20	77,42	78,31
Erkek						
Genç	60,50	60,70	61,88	60,50	60,70	56,35
Orta	70,61	70,82	71,25	72,81	70,82	71,25
Yaşlı	5,67	5,82	5,44	5,67	5,82	5,44
4Duygu(Nötr, Heyecanlı, Mutlu, Kızgın)						
Kadın						
Genç	34,67	36,65	34,56	36,80	36,33	37,47
Orta	60,68	66,02	65,97	64,32	64,13	64,92
Erkek						
Genç	47,71	47,55	49,16	47,71	47,55	49,16
Orta	40,17	40,30	40,00	37,17	40,30	40,00
Yaşlı	4,33	4,45	4,19	4,33	4,45	4,19
Tüm Duygular (Nötr,Heyecanlı, Mutlu, Kızgın Umutsuz, Üzgün Bilinmeyen)						
Kadın						
Genç	26,61	28,24	27,27	35,09	31,94	33,33
Orta	50,28	52,74	56,62	59,37	58,95	60,90
Erkek						
Genç	46,79	46,60	48,34	46,79	46,60	48,34
Orta	38,03	39,16	38,91	38,35	39,16	38,91
Yaşlı	4,33	4,45	4,19	4,33	4,45	4,19

4 SONUÇ ve ÖNERİLER

Bu tez çalışmasında ses sinyallerinin işlenerek, insanların duygu durumlarının tahmin edilmesi üzerine çalışma yapılmıştır. Veri seti olarak hazır bir veri seti kullanılmamış, radyo programından toplanan konuşmalardan veri seti oluşturulmuştur.

Bu özelliklerin hangilerinin kullanılacağına karar verebilmek için Weka ile veri seti üzerinde analizler yapılmıştır. Weka tarafından sağlanan özellik seçimi yöntemlerinden Correlation-based Feature Selection Subset Evaluation, Information Gain Attribute Evaluator, Principal Component yöntemleri ile indirgenen özellik kümeleri ve 204 özelliğin tümü kullanılarak yapılan sınıflandırma başarımları kıyaslanmıştır. Özellikler üzerinde kısıtlama yapıldığında, tüm özelliklerin kullanımına göre sınıflandırma başarımında bir artış gözlenmediğinden 204 özelliğin tamamının kullanılmasına karar verilmiştir.

Oluşturulan veri setinde etiketlenen 6 duygu durumu (Nötral (Neutral), Heyecanlı (Excited), Mutlu (Happy), Kızgın (Angry), Üzgün (Sad), Umutsuz (Desperate)) olmasına rağmen, Üzgün, Umutsuz duygu durumlarından çok az sayıda veri olduğu için, Nötral, Heyecanlı, Mutlu, Kızgın duygu durumları ile testler yapılmıştır.

Özellik seçimi için yapılan testlerde 6 sınıflandırma algoritması kullanılmış olup, bunlar, Naive Bayes, J48, Random Forest, Bayes Net, Ibk, Decision Table'dır. En iyi sonuçları Random Forest ile elde edildiğinden, literatürde de en yaygın Naive Bayes kullanıldığından, bu iki sınıflandırıcı ile testlerin yapılmasına karar verilmiştir.

Oluşturulan veri seti üzerindeki sistem başarımını karşılaştırmak amacıyla BES veri seti ile de testler/senaryolar tekrarlanmıştır. TOVS'de de BES'de de Random Forest'in çoğu senaryoda daha iyi sonuçlar ürettiği gözlenmiştir. Naive Bayes başarımı Random Forest'a göre TOVS'de üstünlük sergileyememiş; BES'de bazı senaryolarda az da olsa daha iyi sınıflandırma yapabilmıştır.

Naive Bayes için en iyi sistem başarısı %79,73 olarak hesaplanmıştır. Bu sonuç, orta yaşlı kadın konuşmacıların, 2 sn'lik klipleri ve Nötr, Mutlu, Kızgın duyguları ile

çalışıldığında elde edilmiştir.

Random Forest için en iyi sistem başarısı %78,31 olarak hesaplanmıştır. Bu sonuç, orta yaşlı kadın konuşmacıların, 2 sn'lik klipleri ve Nötr, Mutlu, Kızgın duyguları ile çalışıldığında elde edilmiştir.

Farklı veri kümelerinde başarıyı görebilmek adına, duygulara, klip boyutlarına, cinsiyete ve yaşa göre senaryolar oluşturulmuştur.

TOVS'de ve BES'de kayıt sayısı en fazla olan duygular, iki veri setinde ortak olan duygular ve iki veri setindeki tüm duygular ile senaryolar oluşturulmuş ve sınıflandırma başarımlar karşılaştırılmıştır. TOVS'de 3 duygu için, BES'de 4 duygu için Random Forest'in %71,01 ve %69,91'lik sonuç ile en iyi başarıyı sağladığı gözlemlenmiştir.

Her bir konuşmacının ses kayıtları duyguya göre parçalandıktan sonra, 1 ve 1,5 ve 2 sn'lik kliplere parçalanmıştır. Her bir klip boyu için her duygu durumu senaryosu tekrarlanarak klip boyunun sınıflandırma başarımına etkisi incelenmiştir. TOVS'de 3 duygu için, BES'de 4 duygu için 1,5 sn'lik ses kayıtlarında, Random Forest'in %72,16 ve %73,83'lük sonuç ile en iyi başarıyı sağladığı gözlemlenmiştir.

Cinsiyetin etkisini değerlendirebilmek için, veriler kadın ve erkek olarak bölünmüş, bu bölümler üzerinde ayrı ayrı eğitim ve test veri seti oluşturulup, cinsiyetler için klip boyutu ve duygu durumu senaryoları tekrarlanmıştır. TOVS'da kadın konuşmacıların, 3 duygu için olan, 2 sn'lik ses kayıtlarında, Random Forest'in %68,44'lük sonuç ile en iyi başarıyı sağladığı gözlemlenmiştir.

Veri seti, genç(15-30), orta(31-50) ve yaşlı(50+) olarak 3 gruba ayrılmıştır. Cinsiyete ve yaşa göre eğitim ve test verileri oluşturulup klip boyutu ve duygu durumu senaryoları tekrarlanmıştır. Burada, gençlerin duygularının daha zor anlaşıldığı görülmektedir. TOVS'da orta yaşlı kadın konuşmacıların, 3 duygu için olan, 2 sn'lik ses kayıtlarında, Naive Bayes'in %79,73'lük sonuç ile en iyi başarıyı sağladığı gözlemlenmiştir. Yaşlı konuşmacı sayısı kadınlarda 2 olduğu için test edilememiştir. Erkeklerde ise 15 konuşmacı ile testler yapılmış ve konuşmacı sayısı azlığına bağlı olarak tatmin edici sonuçlar elde edilememiştir.

Türkçe için oluşturulmuş bir ses kaydı veri seti olmadığından, hipotezin doğrulanması için veri seti oluşturulması gereği ortaya çıkmıştır. Oluşturulan veri seti, konuşmacıların doğal duygularını yakalayabilmek ve telefon ortamındaki ses kayıtlarından faydalanabilmek için radyo programı kayıtlarından elde edilmiştir. Bu ses kayıtlarında her hangi bir standart ve dengeli dağılım bulunmadığından,

oluřturulan eęitim ve test veri setlerinin sistem bařarımını doęru ltę řphelidir. Ses kayıtlarındaki duygu durum etiketlemelerinin ne kadar doęru yapıldıęının belirlenmesi de sistem bařarımının yorumlanması iin bir ihtiya olarak grlmektedir.

Sistemin eksik ynlerinin ortaya konabilmesi ve geliřtirilebilmesi iin ncelikle, daha geniř, dengeli daęıtılmıř ve oklu oylama ile etiketlenmiř verilenden hazırlanmıř bir veri seti ihtiyaı bulunmaktadır.

- [1] L. S. Chen, T. S. Huang, T. Miyasato, and R. Nakatsu, "Multimodal human emotion/expression recognition," in *Automatic Face and Gesture Recognition, 1998. Proceedings. Third IEEE International Conference on*, IEEE, 1998, pp. 366–371.
- [2] L. C. De Silva, T. Miyasato, and R. Nakatsu, "Facial emotion recognition using multi-modal information," in *Information, Communications and Signal Processing, 1997. ICICS., Proceedings of 1997 International Conference on*, IEEE, vol. 1, 1997, pp. 397–401.
- [3] V. A. Petrushin, "Emotion recognition in speech signal: Experimental study, development, and application," in *Sixth International Conference on Spoken Language Processing*, 2000.
- [4] C. M. Lee, S. Narayanan, and R. Pieraccini, "Recognition of negative emotions from the speech signal," in *Automatic Speech Recognition and Understanding, 2001. ASRU'01. IEEE Workshop on*, IEEE, 2001, pp. 240–243.
- [5] S. Yacoub, S. Simske, X. Lin, and J. Burns, "Recognition of emotions in interactive voice response systems," in *Eighth European conference on speech communication and technology*, 2003.
- [6] T. L. Nwe, S. W. Foo, and L. C. De Silva, "Speech emotion recognition using hidden markov models," *Speech communication*, vol. 41, no. 4, pp. 603–623, 2003.
- [7] M. W. Bhatti, Y. Wang, and L. Guan, "A neural network approach for human emotion recognition in speech," in *Circuits and Systems, 2004. ISCAS'04. Proceedings of the 2004 International Symposium on*, IEEE, vol. 2, 2004, pp. II–181.
- [8] C. Busso, Z. Deng, S. Yildirim, M. Bulut, C. M. Lee, A. Kazemzadeh, S. Lee, U. Neumann, and S. Narayanan, "Analysis of emotion recognition using facial expressions, speech and multimodal information," in *Proceedings of the 6th international conference on Multimodal interfaces*, ACM, 2004, pp. 205–211.
- [9] C. M. Lee, S. Yildirim, M. Bulut, A. Kazemzadeh, C. Busso, Z. Deng, S. Lee, and S. Narayanan, "Emotion recognition based on phoneme classes," in *Eighth International Conference on Spoken Language Processing*, 2004.
- [10] B. Schuller, G. Rigoll, and M. Lang, "Speech emotion recognition combining acoustic features and linguistic information in a hybrid support vector machine-belief network architecture," in *Acoustics, Speech, and Signal Processing, 2004. Proceedings. (ICASSP'04). IEEE International Conference on*, IEEE, vol. 1, 2004, pp. I–577.

- [11] Y.-L. Lin and G. Wei, "Speech emotion recognition based on hmm and svm," in *Machine Learning and Cybernetics, 2005. Proceedings of 2005 International Conference on*, IEEE, vol. 8, 2005, pp. 4898–4901.
- [12] M. Iseli, Y.-L. Shue, and A. Alwan, "Age-and gender-dependent analysis of voice source characteristics," in *Acoustics, Speech and Signal Processing, 2006. ICASSP 2006 Proceedings. 2006 IEEE International Conference on*, IEEE, vol. 1, 2006, pp. I–I.
- [13] K. DJOLANDER, "The snack sound toolkit," <http://www.speech.kth.se/snack>, 2004.
- [14] J. Rong, G. Li, and Y.-P. P. Chen, "Acoustic feature selection for automatic emotion recognition from speech," *Information processing & management*, vol. 45, no. 3, pp. 315–328, 2009.
- [15] C. Clavel, I. Vasilescu, L. Devillers, G. Richard, and T. Ehrette, "Fear-type emotion recognition for future audio-based surveillance systems," *Speech Communication*, vol. 50, no. 6, pp. 487–503, 2008.
- [16] F. Archetti, G. Arosio, E. Fersini, and E. Messina, "Audio-based emotion recognition for advanced automatic retrieval in judicial domain," in *1st Internat. Conf. on ICT Solutions for Justice*, 2008.
- [17] B. Schuller, A. Batliner, S. Steidl, and D. Seppi, "Emotion recognition from speech: Putting asr in the loop," in *Acoustics, Speech and Signal Processing, 2009. ICASSP 2009. IEEE International Conference on*, IEEE, 2009, pp. 4585–4588.
- [18] B. Yang and M. Lugger, "Emotion recognition from speech signals using new harmony features," *signal processing*, vol. 90, no. 5, pp. 1415–1423, 2010.
- [19] C. Sezgin, B. Gunsel, and G. Kurt, "A new perceptual feature set for audio emotion recognition," Tech. Rep.
- [20] S. Wu, T. H. Falk, and W.-Y. Chan, "Automatic speech emotion recognition using modulation spectral features," *Speech communication*, vol. 53, no. 5, pp. 768–785, 2011.
- [21] D. Bitouk, R. Verma, and A. Nenkova, "Class-level spectral features for emotion recognition," *Speech communication*, vol. 52, no. 7-8, pp. 613–625, 2010.
- [22] T. Vogt and E. Andre, "Improving automatic emotion recognition from speech via gender differentiation," in *Proc. Language Resources and Evaluation Conference (LREC 2006), Genoa*, 2006.
- [23] N. Sebe, I. Cohen, T. Gevers, and T. S. Huang, "Emotion recognition based on joint visual and audio cues," in *Pattern Recognition, 2006. ICPR 2006. 18th International Conference on*, IEEE, vol. 1, 2006, pp. 1136–1139.
- [24] L. C. De Silva and P. C. Ng, "Bimodal emotion recognition," in *Automatic Face and Gesture Recognition, 2000. Proceedings. Fourth IEEE International Conference on*, IEEE, 2000, pp. 332–335.
- [25] M. Valstar, B. Schuller, K. Smith, F. Eyben, B. Jiang, S. Bilakhia, S. Schnieder, R. Cowie, and M. Pantic, "Avec 2013: The continuous audio/visual emotion and depression recognition challenge," in *Proceedings of the 3rd ACM international workshop on Audio/visual emotion challenge*, ACM, 2013, pp. 3–10.

- [26] P. Henriquez, J. B. Alonso, M. A. Ferrer, C. M. Travieso, and J. R. Orozco-Arroyave, "Nonlinear dynamics characterization of emotional speech," *Neurocomputing*, vol. 132, pp. 126–135, 2014.
- [27] A. Davletcharova, S. Sugathan, B. Abraham, and A. P. James, "Detection and analysis of emotion from speech signals," *Procedia Computer Science*, vol. 58, pp. 91–96, 2015.
- [28] S. Lalitha, D. Geyasruti, R. Narayanan, and M. Shravani, "Emotion detection using mfcc and cepstrum features," *Procedia Computer Science*, vol. 70, pp. 29–35, 2015.
- [29] M. Pell, L. Monetta, S. Paulmann, and S. Kotz, "Recognizing emotions in a foreign language," *Journal of Nonverbal Behavior*, vol. 33, pp. 107–120, Jun. 2009. DOI: 10.1007/s10919-008-0065-7.
- [30] F. Burkhardt, A. Paeschke, M. Rolfes, W. Sendlmeier, and B. Weiss, "A database of german emotional speech," in *in Proceedings of Interspeech, Lissabon*, 2005, pp. 1517–1520.
- [31] O. Martin, I. Kotsia, B. Macq, and I. Pitas, "The enterface'05 audio-visual emotion database," in *Data Engineering Workshops, 2006. Proceedings. 22nd International Conference on*, IEEE, 2006, pp. 8–8.
- [32] E. Douglas-Cowie, R. Cowie, I. Sneddon, C. Cox, O. Lowry, M. Mcrorie, J.-C. Martin, L. Devillers, S. Abrilian, A. Batliner, *et al.*, "The humane database: Addressing the collection and annotation of naturalistic and induced emotional data," in *International conference on affective computing and intelligent interaction*, Springer, 2007, pp. 488–500.
- [33] Y. Wang, L. Guan, and A. N. Venetsanopoulos, "Kernel cross-modal factor analysis for information fusion with application to bimodal emotion recognition," *IEEE Transactions on Multimedia*, vol. 14, no. 3, pp. 597–607, 2012.
- [34] I. S. Engberg, A. V. Hansen, O. Andersen, and P. Dalsgaard, "Design, recording and verification of a danish emotional speech database," in *Fifth European Conference on Speech Communication and Technology*, 1997.
- [35] P. Staroniewicz and W. Majewski, "Polish emotional speech database—recording and preliminary validation," in *Cross-Modal Analysis of Speech, Gestures, Gaze and Facial Expressions*, Springer, 2009, pp. 42–49.
- [36] Show radyo, <http://www.showradyo.com.tr/>, Accessed: 2018-11-16.
- [37] P. Rao, "Audio signal processing," in *Speech, Audio, Image and Biomedical Signal Processing using Neural Networks*, Springer, 2008, pp. 169–189.
- [38] Zero crossing rate grafik, https://www.asee.org/documents/zones/zone1/2008/student/ASEE12008_0044_paper.pdf/, Accessed: 2019-02-05.
- [39] S. Mache, "Development of text-to-speech synthesizer for pali language," *IOSR Journal of Computer Engineering (IOSR-JCE)*, vol. 18, pp. 35–42, May 2016. DOI: 10.9790/0661-1803013542.
- [40] J. Kua, T. Thiruvaran, H. Nosrati, A. E, and J. Epps, "Investigation of spectral centroid magnitude and frequency for speaker recognition," Jun. 2010.

- [41] *Spectralrolloff*, https://www.researchgate.net/figure/Definition-du-Spectral-Rolloff-Point_fig5_278631244/, Accessed: 2019-03-13.
- [42] *Spectralflux*, http://www.nyu.edu/classes/bello/ACA_files/3-novelty.pdf/, Accessed: 2019-03-13.
- [43] *F0*, <https://www.mathworks.com/help/audio/ref/pitch.html/>, Accessed: 2019-03-13.
- [44] *Mfcc1*, *1*dWnJn5LLS0j8St53ACwqSg.jpeg*, Accessed: 2019-03-18.
- [45] *Mfcc2*, *1*pmjqw39CpC9FHb7o4CmK0A.png*, Accessed: 2019-03-18.
- [46] *Perceptualfeaturesofsingingvoice*, <https://comum.rcaap.pt/bitstream/10400.26/10625/1/2015\%20Audio-perceptual\%20features\%20of\%20the\%20singing\%20voice.pdf>, Accessed: 2019-02-21.
- [47] *Tini*, *timbre - of - the - instruments - 1024x665 . png*, Accessed: 2019-03-18.
- [48] *Audacity*, <https://sourceforge.net/projects/audacity/>, Accessed: 2018-11-14.
- [49] C. McKay, I. Fujinaga, and P. Depalle, “Jaudio: A feature extraction library,” in *Proceedings of the International Conference on Music Information Retrieval*, 2005, pp. 600–3.
- [50] *Bayesnet*, <https://www.bayesserver.com/docs/introduction/bayesian-network>, Accessed: 2019-03-25.
- [51] *Wekabayesnet*, <http://weka.sourceforge.net/doc.dev/weka/classifiers/bayes/BayesNet.html>, Accessed: 2019-03-18.
- [52] *Wekaclassifier*, https://www.cs.waikato.ac.nz/mlweka/Witten_et_al_2016_appendix.pdf, Accessed: 2019-03-18.
- [53] *Decisiontable*, <https://reqtest.com/requirements-blog/a-guide-to-using-decision-tables/>, Accessed: 2019-03-25.
- [54] *Wekadecisiontable*, <http://weka.sourceforge.net/doc.dev/weka/classifiers/rules/DecisionTable.html>, Accessed: 2019-03-18.
- [55] *Wekahmm*, <http://mlg.eng.cam.ac.uk/zoubin/papers/ijprai.pdf>, Accessed: 2019-03-18.
- [56] D. W. Aha, D. Kibler, and M. K. Albert, “Instance-based learning algorithms,” *Machine learning*, vol. 6, no. 1, pp. 37–66, 1991.
- [57] *Wekaj48*, <http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.678.9273&rep=rep1&type=pdf>, Accessed: 2019-03-18.
- [58] *Wekaknn*, <https://brilliant.org/wiki/k-nearest-neighbors/>, Accessed: 2019-03-18.
- [59] E. Frank, M. Hall, and B. Pfahringer, “Locally weighted naive bayes,” in *Proceedings of the Nineteenth conference on Uncertainty in Artificial Intelligence*, Morgan Kaufmann Publishers Inc., 2002, pp. 249–256.
- [60] G. H. John and P. Langley, “Estimating continuous distributions in bayesian classifiers,” in *Proceedings of the Eleventh conference on Uncertainty in artificial intelligence*, Morgan Kaufmann Publishers Inc., 1995, pp. 338–345.

- [61] L. Breiman, "Random forests," *Machine learning*, vol. 45, no. 1, pp. 5–32, 2001.
- [62] S. Singhal and M. Jena, "A study on weka tool for data preprocessing, classification and clustering," *International Journal of Innovative technology and exploring engineering (IJITEE)*, vol. 2, no. 6, pp. 250–253, 2013.