

**T.C.
YILDIZ TEKNİK ÜNİVERSİTESİ
FEN BİLİMLERİ ENSTİTÜSÜ**

İNSAN SESİNDEN DUYGU TANIMA

CEVAHİR PARLAK

**YÜKSEK LİSANS TEZİ
BİLGİSAYAR MÜHENDİSLİĞİ ANABİLİM DALI**

**DANIŞMAN
DOÇ. DR. BANU DİRİ**

İSTANBUL, 2015

T.C.
YILDIZ TEKNİK ÜNİVERSİTESİ
FEN BİLİMLERİ ENSTİTÜSÜ

İNSAN SESİNDEN DUYGU TANIMA

Cevahir PARLAK tarafından hazırlanan tez çalışması 25.06.2015 tarihinde aşağıdaki jüri tarafından Yıldız Teknik Üniversitesi Fen Bilimleri Enstitüsü Bilgisayar Mühendisliği Anabilim Dalı'nda **YÜKSEK LİSANS TEZİ** olarak kabul edilmiştir.

Tez Danışmanı

Doç. Dr. Banu DİRİ

Yıldız Teknik Üniversitesi

Jüri Üyeleri

Doç. Dr. Banu DİRİ

Yıldız Teknik Üniversitesi

Prof. Dr. Nizamettin AYDIN

Yıldız Teknik Üniversitesi

Prof. Dr. Çiğdem Eroğlu ERDEM

Bahçeşehir Üniversitesi

ÖNSÖZ

Tezimdeki danışmanlığından ve katkılarından dolayı Doç. Dr. Banu DİRİ'ye teşekkürü bir borç bilirim. Onun tavsiyesi ve danışmanlığı olmasaydı böyle bir tez ortaya çıkamazdı.

Jüri üyelerimden Prof. Dr. Nizamettin AYDIN'a verdiği ders, Prof. Dr. Çiğdem Eroğlu ERDEM'e jüri üyeliğinde bulunma nezakatinden dolayı ayrıca teşekkür etmek isterim.

Boğaziçi Üniversitesinden ders aldığım Prof. Dr. Fikret GÜRGEN'e de tezin gelişmesindeki katkılarından dolayı teşekkür ederim.

Yıldız Teknik Üniversitesi'ndeki diğer hocalarıma ve çalışanlara da yüksek lisans eğitimimde sağladıkları nezih ortam nedeniyle minnetlerimi ifade etmek isterim.

Haziran, 2015

Cevahir PARLAK

İÇİNDEKİLER

	Sayfa
KISALTMA LİSTESİ	vii
ŞEKİL LİSTESİ.....	ix
ÇİZELGE LİSTESİ	x
ÖZET	xii
ABSTRACT	xiv
BÖLÜM 1	
GİRİŞ	1
1.1 Literatür Özeti.....	1
1.1.1 Çok Setli Mevcut Çalışmalar	2
1.1.2 Tek Setli Mevcut Çalışmalar	3
1.2 Hipotez	4
1.3 Tezin Bilimsel Katkıları.....	6
1.4 Tezin Önemli Bulguları	6
1.5 Tezin Ana Hatları	7
BÖLÜM 2	
DUYGU NEDİR?.....	8
2.1 Duygunun Tanımı ve Türleri	8
2.2 Duygu İfade Yolları	12
2.2.1 Yüz İfadesi.....	12

2.2.2	Vücut Dili	12
2.2.3	Ses	12
2.2.4	Biyolojik Belirtiler	13
2.2.5	Termal Kızılaltı Sinyaller	13
2.3	Duygu Veritabanları	14
2.3.1	Doğal (Natural/Spontaneous) Duygu Veritabanları.....	18
2.3.2	Yapay (Simulated/Acted) Duygu Veritabanları.....	18
2.3.3	Zorlanmış (Elicited/Induced) Duygu Veritabanları	19
2.4	Duygu Etiketleme.....	19
2.5	Duygu Tanıma Uygulama Alanları.....	20

BÖLÜM 3

İNSAN SESİNDEN DUYGU TANIMA.....	23
3.1 Ses Dosyaları İçin Önışleme	24
3.2 Ses Aktivite Algılaması.....	25
3.3 Öznitelik Çıkarma	25
3.3.1 Temel Frekans (Fundamental Frequency, Pitch)	26
3.3.2 Konuşma Hızı ve Kalitesi (Speaking Rate, Voice Quality)	28
3.3.3 Teager Enerji Operatörü.....	28
3.3.4 Sıfır Geçiş Oranı ve Kısa Zaman Enerji.....	28
3.3.5 Frekans Spektrumu ve Formantlar	29
3.3.6 Mel Frekans Kepstrum Katsayıları	32
3.3.7 Türevler	33
3.3.8 PLP.....	34
3.3.9 Öznitelik Fonksiyonları	34
3.4 Öznitelik Seçimi.....	35
3.5 Sınıflandırma.....	36
3.5.1 kNN.....	36
3.5.2 GMM	37
3.5.3 SVM	37
3.5.4 HMM	37

3.5.5	ANN	37
3.5.5	Karar Ağacı Algoritmaları.....	38
BÖLÜM 4		
DENEYSEL KURULUM.....		39
4.1	EmoSTAR	39
4.2	Berlin Duygu Veritabanı	40
4.3	OPENSIMILE	41
4.4	Weka	44
4.5	Weka Komut Satırı	46
4.6	Deneyin Yapılışı	47
4.6.1	Berlin Duygu Veritabanı Üzerine Çalışmalar	48
4.6.2	EmoSTAR Üzerindeki Çalışmalar.....	52
4.6.3	EmoSTAR + EmoDB Üzerine Çalışmalar	52
4.6.4	EmoSTAR ve EmoDB Arasındaki Çapraz Sınıflandırmalar	53
4.7	Öznitelik Seçme.....	55
BÖLÜM 5		
ÖNERİLEN FORMANT TABANLI FİLTRE BANKALARI		57
5.1	Sesli Harflerin Formantlarına Dayalı Filtre Bankaları ve Yüksek Frekans- Düşük Frekans Bölgesi Enerji Oranı	57
5.2	Duraklamalar	59
5.3	Mutlak Şiddet Değerleri	60
5.4	Yeni Özniteliklerle Deneysel Sonuçlar.....	61
5.5	Eklenen Yeni Özniteliklerle Öznitelik Seçme Uygulamaları.....	64
BÖLÜM 6		
SONUÇ VE ÖNERİLER		68
KAYNAKLAR.....		71
ÖZGEÇMİŞ.....		76

ANN	Artificial Neural Network
ARFF	Attribute Relation File Format
ATM	Automatic Teller Machine
CSV	Comma Separated Values
DC	Direct Current
DCT	Discrete Cosine Transform
DFT	Discrete Fourier Transform
EEG	ElectroEncefaloGram
EmoDB	Berlin Emotional Database
EmoSTAR	Emotional Speech Transcription and Recognition
FACS	Facial Action Coding System
FFT	Fast Fourier Transform
fNIRS	functional Near Infrared Spectroscopy
GMM	Gaussian Mixture Models
HMM	Hidden Markov Model
HTK	Hidden Markov Toolkit
Hz	Hertz
IEEE	Institute of Electrical and Electronics Engineers
KNN	k Nearest Neighbor
LFPC	Log Frequency Power Coefficients
LH	Low High Frequency Ratio
LPC	Linear Predictive Coding
LPCC	Linear Prediction Cepstral Coefficients
MFCC	Mel Frequency Cepstral Coefficients
NN	Neural Network
OpenSmile	Open Speech and Music Interpretation by Large-space Extraction
PLP	Perceptual Linear Prediction
STAFFER	Spectro-Temporal Auditory Formant Filter Bank Energy Ratios
STE	Short Time Energy
SVM	Support Vector Machine
TEO	Teager Energy Operator
VAD	Voice Activity Detection
WOZ	Wizard of OZ

WEKA
ZCR

Waikato Environment for Knowledge Analysis
Zero Crossing Rate

ŞEKİL LİSTESİ

	Sayfa
Şekil 2. 1 İki boyutlu duygu uzayı.....	10
Şekil 2. 2 Üç boyutlu duygu uzayı	11
Şekil 2. 3 Üç boyutlu duygu sınıflandırma	11
Şekil 2. 4 Yankısız oda	15
Şekil 3. 1 Duygu tanıma genel mimarisi	23
Şekil 3. 2 Konuşma sinyali için pencereleme.....	24
Şekil 3. 3 Pencere fonksiyonunun (üstte) bir sinyaldeki (ortada) etkisi (altta)	25
Şekil 3. 4 Nötr konuşma temel frekans çizgisi	27
Şekil 3. 5 Kızgın konuşma temel frekans çizgisi	27
Şekil 3. 6 Sıfır geçiş oranı ve kısa zaman enerji çizgileri	29
Şekil 3. 7 Sekiz kelimesi için formant izleme	30
Şekil 3. 8 U harfi için FFT uygulaması.....	30
Şekil 3. 9 İ harfi için FFT uygulaması	31
Şekil 3. 10 Türkçe sesli harfler için F1 ve F2 formant düzlemi	31
Şekil 3. 11 Mel filtre bankaları	33
Şekil 3. 12 Konuşma sinyali için öznitelik çıkarma	33
Şekil 3. 13 Öznitelik seçimi için filtre yaklaşımı	36
Şekil 3. 14 Öznitelik seçimi için alt uzay seçme yaklaşımı	36
Şekil 3. 15 Yapay sinir ağı	38
Şekil 4. 1 Weka ön işleme penceresi	44
Şekil 4. 2 Weka sınıflandırma penceresi	45
Şekil 4. 3 Weka öznitelik seçme penceresi.....	46
Şekil 4. 4 Weka komut satırı	47
Şekil 5. 1 Sesli harflerin formantlarına dayalı filtre bankaları	58
Şekil 5. 2 Kalın ve ince sesli harflerin formantlarına dayalı yüksek frekans ve	59
frekans filtre bankaları	59
Şekil 5. 3 EmoDB’de kızgın ve nôtür cümlelerdeki yüksek ve düşük frekanslardaki	59
enerji oranları	59
Şekil 5. 4 EmoDB’de kızgın ve üzgün cümlelerdeki duraklama oranları	60
Şekil 5. 5 Bir sinyale ait frekans spektrumun logaritmik (üstte) ve mutlak değerli.....	61
(altta) gösterimi	61

ÇİZELGE LİSTESİ

Sayfa

Çizelge 2. 1	Whissel [11], Plutchnik [12] ve Cowie [13]'nin çalışmalarına göre duygu türleri	11
Çizelge 2. 2	Duygu veritabanları	16
Çizelge 3. 1	Duygu tanımada kullanılan alt düzey öznitelikler	26
Çizelge 3. 2	Duygu tanımada alt düzey özniteliklere uygulanan fonksiyonlar	35
Çizelge 4. 1	EmoSTAR kayıtlarının duygulara göre dağılımı	40
Çizelge 4. 2	EmoDB kayıtlarının duygulara göre dağılımı	40
Çizelge 4. 3	OpenSMILE konfigürasyon dosyalarının öznitelik sayıları	43
Çizelge 4. 4	EmoDB için sınıflandırma başarıları	48
Çizelge 4. 5	EmoDB'de Emobase (988 öznitelik) için SMO hata matrisi	49
Çizelge 4. 6	EmoDB için farklı öznitelik gruplarıyla sınıflandırma sonuçları	49
Çizelge 4. 7	EmoDB'de 532 prozodik öznitelikli Emobase için SMO hata matrisi	50
Çizelge 4. 8	EmoDB'de 6084 MFCC öznitelikli Emo_large için SMO hata matrisi	50
Çizelge 4. 9	EmoDB'de farklı duygu grupları için Emobase sınıflandırma sonuçları	51
Çizelge 4. 10	EmoDB'de farklı duygu grupları için Emo_large sınıflandırma sonuçları	51
Çizelge 4. 11	Emobase Kızgın-Nötr-Mutlu-Üzgün için SMO hata matrisi	51
Çizelge 4. 12	Emobase Nötr-Mutlu-Üzgün için SMO hata matrisi	51
Çizelge 4. 13	Emobase Kızgın-Nötr için SMO hata matrisi	51
Çizelge 4. 14	EmoSTAR'da Kızgın-Nötr-Mutlu-Üzgün için sınıflandırma sonuçları	52
Çizelge 4. 15	EmoSTAR için farklı öznitelik gruplarıyla sınıflandırma sonuçları	52
Çizelge 4. 16	EmoDB'de Kızgın-Nötr-Mutlu-Üzgün için sınıflandırma sonuçları	53
Çizelge 4. 17	EmoSTAR ve EmoDB'nin karıştırılmasıyla elde edilen sonuçlar	53
Çizelge 4. 18	EmoSTAR eğitim EmoDB test seti olmak üzere sınıflandırma sonuçları	53
Çizelge 4. 19	EmoDB eğitim EmoSTAR test seti olmak üzere sınıflandırma sonuçları	54
Çizelge 4. 20	EmoSTAR ve EmoDB ayrı ayrı eğitim ve test seti olmak üzere Emobase sınıflandırma sonuçları	54
Çizelge 4. 21	EmoSTAR ve EmoDB ayrı ayrı eğitim ve test seti olmak üzere Emo_large sınıflandırma sonuçları	54
Çizelge 4. 22	EmoDB 7 duygu için öznitelik seçme sonuçları	55
Çizelge 4. 23	EmoSTAR 4 duygu için öznitelik seçme sonuçları	56
Çizelge 5. 1	EmoDB'de 7 duygu için yeni özniteliklerle sınıflandırma sonuçları	61
Çizelge 5. 2	EmoDB'de 4 duygu için yeni özniteliklerle Kızgın-Nötr-Mutlu-Üzgün sınıflandırma sonuçları	62
Çizelge 5. 3	EmoSTAR'da 4 duygu için yeni özniteliklerle sınıflandırma sonuçları	62

Çizelge 5. 4	EmoSTAR için STAFFER öznitelikleriyle farklı duygu grupları sınıflandırma sonuçları	62
Çizelge 5. 5	EmoDB için STAFFER öznitelikleriyle farklı duygu grupları sınıflandırma sonuçları	63
Çizelge 5. 6	EmoSTAR ve EmoDB için 315 yeni öznitelikle çapraz test sonuçları	63
Çizelge 5. 7	EmoSTAR ve EmoDB için 7 yeni öznitelikle çapraz test sonuçları.....	63
Çizelge 5. 8	EmoSTAR ve EmoDB için Emobase konfigürasyonuna yeni özniteliklerin .. eklenmesiyle elde edilen çapraz test sonuçları	64
Çizelge 5. 9	EmoSTAR ve EmoDB için Emo_large konfigürasyonuna yeni özniteliklerin eklenmesiyle elde edilen çapraz test sonuçları	64
Çizelge 5. 10	EmoDB’de 7 duygu için yeni enerji operatörlerinin öznitelik seçme sıralama sonuçları	65
Çizelge 5. 11	EmoSTAR’da yeni enerji operatörlerinin öznitelik seçme sıralama sonuçları.....	65
Çizelge 5. 12	EmoDB’de 7 duygu için yeni özniteliklerin öznitelik seçme sıralama sonuçları.....	65
Çizelge 5. 13	EmoSTAR’da yeni özniteliklerin öznitelik seçme sıralama sonuçları.....	65
Çizelge 5. 14	EmoDB’de 7 duygu için eklenen yeni özniteliklerle özellik seçme..... sonuçları.....	66
Çizelge 5. 15	EmoSTAR’da 4 duygu için eklenen yeni özniteliklerle özellik seçme sonuçları.....	66
Çizelge 5. 16	EmoDB için Emobase konfigürasyonunda Prozodi+24MFCC+STAFFER öznitelik setiyle elde edilen öznitelik seçme sonuçları.....	67

İNSAN SESİNDEN DUYGU TANIMA

Cevahir PARLAK

Bilgisayar Mühendisliği Anabilim Dalı

Yüksek Lisans Tezi

Tez Danışmanı: Doç. Dr. Banu DİRİ

İnsanlar arasındaki en önemli iletişim aracı konuşmadır. Konuşma ile insanlar birbirlerine sadece düşüncelerini değil duygularını da aktarabilirler. Bunun yanı sıra insan sesi, parmak izi ve iris gibi biyometrik özellikleri de beraberinde taşıyabilmektedir. Konuşma ile karşımızdaki kişinin düşüncesini, duygusunu, cinsiyetini ve yaşını da tahmin edebilmekteyiz. İnsan sesi ve konuşması ile ilgili çalışmalar her ne kadar çok eski olsa da insan sesinden duygu analizi yeni bir araştırma alanıdır. Özellikle son 20 yıl içinde bu alanda çok önemli çalışmalar yapılmıştır. Bu çalışmalar başlangıçta sadece konuşmanın olumlu veya olumsuz olmasıyla ilgilenirken zamanla daha fazla duygu türleri çalışma kapsamına alınmış ve bu amaçla pek çok duygu veritabanı geliştirilmiştir. Bu çalışmada veritabanlarından en popüler olanı Berlin Emotional Database ve kendi veritabanımız olan EmoSTAR kullanılmıştır.

İnsan-robot iletişimde de duygu analizi önemli bir yer tutmaktadır ve gün geçtikçe daha ileri adımlar atılmaktadır. Önceleri konuşma sentezi yoluyla makinelerin insanlara konuşarak cevap vermeleri amaçlanmaktaydı. Bu çalışmalar oldukça başarılı olmuş ve ilk başlarda mekanik gözüken makine sesi artık insan sesine yakın bir doğallığa kavuşmuştur. Makine konuşmasındaki doğallığı daha ileri seviyelere taşımanın bir yoluda makine konuşmasına duygu ilave edebilmektir. Duygu tanıma çalışmaları konuşma ve konuşmacı tanıma ve sentez sistemlerinin geliştirilmesinde de faydalıdır.

Anahtar Kelimeler: Çok setli duygu analizi, duygu madenciliđi, prozodi, ses perdesi, EmoSTAR

EMOTION RECOGNITION FROM THE HUMAN VOICE

Cevahir PARLAK

Department of Computer Engineering

MSc. Thesis

Adviser: Assoc. Prof. Banu DİRİ

Human speech is the most important communication tool between humans. Humans express not only their thoughts with speech but they can also express their feelings. Human voice is a biometric tool like fingerprint or iris and can carry many information such as emotion, gender and even age. Although, the works on human speech date back more than 100 years ago emotion analysis is relatively a new research area. In the last 20 years very extensive researches have been conducted in this field. In the beginning only positive versus negative speech analysis were conducted but as technology matured, many emotion kinds have been added and many emotional databases have been developed for this purpose. In this work, one of the most popular of these databases Berlin Emotional Database was used together with EmoSTAR which is our own emotional speech corpus.

Keywords: Cross-corpus emotion analysis, emotion mining, prosody, pitch, EmoSTAR

YILDIZ TECHNICAL UNIVERSITY

GRADUATE SCHOOL OF NATURAL AND APPLIED SCIENCES

1.1 Literatür Özeti

Duygu tanıma ile ilgili ilk araştırmalara 80'li yılların ortalarından itibaren rastlanmaya başlanmış (Van Bezooijen, 1984; Tolkmitt ve Scherer, 1986) bilgisayar teknolojisindeki gelişmelere paralel olarak 90'lı yılların ortalarından itibaren ivme kazanmıştır. Bu çalışmalar önceleri sadece konuşmanın olumlu veya olumsuz oluşu ile ilgilenmekte iken sonraları teknolojinin gelişmesiyle beraber pek çok duygu türünü ve stres derecesini de kapsama alanı içine almıştır. Zamanla duygu tanımlama için yeni öneriler getirilmiş ve klasik duygu tanımlamanın yanı sıra 2 boyutlu, 3 boyutlu, ayrık ve sürekli duygu tanımları ortaya konmuştur. Bu çalışmaların temel öğelerinden birisi de duygu veritabanlarıdır. Yapay zeka teknolojilerindeki gelişmelerde bu alandaki ilerlemeleri hızlandırmıştır. Bugün gelinen noktada duygu tanımadaki çalışmalarda oldukça iyi sonuçlar elde edilebilmektedir. Duygu tanıma çalışmalarında kullanılan veritabanlarındaki duygu türü ve bu duyguları sınıflandırmak için kullanılan özniteliklerin sayısı araştırmalara göre birbirlerinden çok farklılık göstermektedir. Bazı araştırmalarda hedeflenen duygu sayısı az olduğundan sayılı birkaç öznitelik kullanılırken bazı araştırmalarda ise binlerce özniteliğin kullanılması gerekmektedir. Bu tür çalışmalarda önceleri klasik öznitelikler olarak bilinen prozodik öznitelikler kullanılmakta iken bugün artık prozodi ve spektrum özniteliklerinin yanı sıra bunlardan çeşitli istatistiksel fonksiyonlarla türetilen binlerce öznitelik kullanılmaktadır. Bu çalışmalarda duygu sınıflandırmanın yanı sıra stres sınıflandırma ve duygu yüklü konuşma sentezi de yapılmaktadır. Son zamanlarda sesin akustik öznitelikleriyle

beraber kelime tanıma ve yüz ifadesi tanıma da kullanılmaya başlanmıştır. Sınıflandırmadaki başarı oranları da veritabanına ve duygu sayısına göre %30 ile %90 arasında farklılıklar göstermektedir.

Duygu tanıma çalışmaları genellikle tek bir veritabanı üstünde gerçekleştirilmektedir ve başarı oranlarında limitlere yaklaşmıştır. Bu nedenle bu çalışmada iki farklı veritabanı ile çalışılacak ve bu veritabanlarından biri eğitim seti diğeri test seti olarak kullanılmak suretiyle çapraz testler gerçekleştirilecektir. Bu tür çalışmalar literatürde nispeten daha yeni ve daha az olduğu gibi üstesinden gelinmesi gereken daha fazla problem bulunmaktadır. Çapraz testli çok setli çalışmalarda başarı oranları da tek setli çalışmalara göre oldukça düşük çıkabilmektedir.

1.1.1 Çok Setli Mevcut Çalışmalar

Oflazoğlu ve Yıldırım [1] TURES (Turkish Emotional Speech Database) ve VAM (Vera am Mittag) veri setlerinde setler arası bir çalışma gerçekleştirmiş ve 3 boyutlu duygu uzayında TURES'in eğitim seti olarak kullanılmasıyla 4 duygu (kızgın, nötr, mutlu, üzgün) üstünde pozitiflik, aktivasyon ve baskınlık için sırasıyla %0,28, %0,74 ve %0,71 korelasyon katsayıları elde etmişlerdir. Bu çalışmada openSMILE ile çıkarılan 1532 öznitelik TURES ve VAM veritabanlarında kullanılmıştır. VAM ve TURES veritabanları ayrı ayrı test ve eğitim verisi olarak kullanılmış ve çapraz uygulama yapılmıştır. Sadece TURES veritabanı üstündeki tek setli çalışmada ise 4 duygu (kızgın, nötr, mutlu, üzgün) üstünde Weka LIBSVM sınıflandırma ile %57,5 başarı oranına ulaşılmıştır.

Zhang vd. [2] ABC, AVIC, DES, eINTERFACE, SAL ve VAM setlerinde yaptıkları çapraz testlerde iki boyutlu duygu uzayında aktivitede %61,6 ve pozitiflikte ise %58,4 başarı oranı elde etmişlerdir.

Schüller vd. [3] EmoDB'nin test seti, AVIC, DES, eINTERFACE, SmartKOM ve SUSAS'ın eğitim seti olarak kullanıldığı çalışmalarında %35 ile %70 arasında başarı oranları elde etmişlerdir.

ABC, AVIC, DES, EmoDB, eINTERFACE, SAL, SUSAS ve VAM olmak üzere 8 veri setinin kullanıldığı bir çalışmada [4], VAM üstünde eğitim diğer 7 veri seti üstünde de test yapılarak iki boyutlu duygu uzayında aktivitede %67,7 başarı oranı yakalanmıştır. DES

üstünde eğitim diğer 7 veri seti üstünde test yapılarak pozitiflik boyutunda %54,8 başarı oranı yakalanmıştır.

Eyben vd. [5] SVM (SMO) ile SmartKOM, Aibo, SAL ve VAM setleri üstündeki çalışmada setlerden üçünü eğitim diğerini de test seti olarak kullanarak %53,4 başarı oranı elde etmişlerdir.

[6] da EmoDB, ABC, AVIC, DES, SUSAS, Enterface, SAL, SmartKom ve VAM veritabanlarında 6552 öznitelikten oluşan bir konfigürasyon denenmiş ve EmoDB’de HMM/GMM tabanlı sınıflandırıcı ile %77,1 ile en yüksek, SAL veritabanında ise %32,7 ile en düşük başarı oranı gözlemlenmiştir. Aynı çalışma SVM ile EmoDB’de %85,6 ile en yüksek, SAL’da %30,6 ile en düşük oranı elde etmiştir.

1.1.2 Tek Setli Mevcut Çalışmalar

[7] de SUAS veritabanında 3 düzeyli stres tanıma üstünde çalışılmıştır. TEO tabanlı özniteliklerle stres tespitinde %92,75, ORI veritabanında TEO tabanlı özniteliklerle 5 duygu üstünde %89,17 başarı oranı elde edilmiştir.

[8] de MFCC ve prozodik özniteliklere ilaveten konuşma ritmi ve ses yüksekliği öznitelikleri kullanılarak 487 öznitelikle SVM’de Berlin EmoDB’de 7 duygu üstünde %80,27 başarı oranı elde edilmiştir. Duygular ikili gruplar halinde sınıflandırmaya da tabi tutulmuşlardır.

[9] da doğal bir veritabanı olan FAU Aibo üzerinde 3713 akustik ve 531 lingistik öznitelikle 4 duygu (anne gibi, nötr, empati ve kızgın) üstünde yapılan çalışma sadece akustik özniteliklerle %63,4 ve sadece lingistik özniteliklerle %62,6 başarı oranı ile sonuçlanmıştır. Akustik ve lingistik özniteliklerin beraber kullanılmasıyla ise %65,35 başarı oranı elde edilmiştir.

[10] da BabyEars ve Kismet veritabanları üzerinde tek setli çalışmada sırasıyla %65.40 ve %88.30 başarı oranlarına ulaşılmasına rağmen aynı verisetleri arasındaki çapraz deneylerde BabyEars eğitim, Kismet test olmak üzere %54.40, Kismet eğitim BabyEars test olmak üzere %45,00 başarı oranları elde edilmiştir.

[11] de EmoDB üstünde 12 Mel Filtre bankasının 1. ve 2. farkları alınarak yeni öznitelikler türetilmiş ve LibSVM ile kızgın, nötr, mutlu, üzgün ve korku olmak üzere 5 duygu üstünde %93,75'lik başarı oranı elde edilmiştir. Cinsiyet tabanlı sınıflandırmada ise erkeklerde %94,73, kadınlarda %100 başarı elde edilmiştir.

[12] de Wu vd. zaman-frekans öznitelikleriyle EmoDB üstünde 7 duygu sınıfında %85,6 başarı oranına ulaşmışlardır.

[13] te FAU Aibo veritabanı üstünde farklı öznitelik setlerinin başarısı karşılaştırılmıştır. SVM ve Rastgele Orman sınıflandırıcıları tüm öznitelikler ve indirgenmiş 150 öznitelikle %61,0 ve %61,2 başarı oranı elde etmiştir. Öznitelik gruplarındaki öznitelik sayısı 26 ile 1718 arasında değişmektedir.

[14] te Bozkurt vd. formant pozisyonlarına dayalı Wğırlıklandırılmış Mel Frekans Cepstral Katsayıları (WMFCC) ile yeni öznitelikler türeterek FAU Aibo seti üzerinde duygu tanıma çalışmaları yapmışlardır. Bu çalışmada sınıflandırma Hidden Markov Model (HMM) sınıflandırıcılarla, negatif ve negatif olmayan olmak üzere 2 sınıf, kızgın, empatik, nötr, pozitif ve diğerleri olmak üzere 5 sınıf olarak gerçekleştirilmiştir. Bu çalışmada 2 sınıf üstünde WMFCC, MFCC ve formant öznitelikleri ile %70,55, 5 sınıf üstünde ise WMFCC, MFCC ve Line Spectral Frequency (LSF) öznitelikleri ile %43,59 başarı oranları elde edilmiştir.

[15] de VAM veritabanı üstünde 46 akustik öznitelik kullanılmıştır. Bunların 26'sı MFCC, 6'sı enerji, 9'u temel frekans ve 5'i süre özneliğidir. Bu özniteliklerin duygu tanımda önemleri araştırılmış ve MFCC özniteliklerinin en yüksek önceliğe sahip olduğu temel frekans ve enerjinin de onu takip ettiği görülmüştür. Süre ile ilgili öznitelikler ise en az öneme sahip öznitelikler olarak görülmüştür.

1.2 Hipotez

Tezin amacı yeni bir duygu veritabanı sunmak ve insan sesinin akustik özniteliklerine dayanarak konuşmadaki duyguyu konuşmacı ve dilden bağımsız olarak tanımlamaktır. Ancak, tek veri seti üstündeki duygu tanıma çalışmalarının hemen hemen doyuma ulaştığı göz önünde bulundurularak tezimizde iki veri seti arasında çapraz testlere ağırlık verilecek ve yeni öznitelikler sunulacaktır. Bununla beraber duygu tanımanın

sadece ses öznitelikleriyle yapılamayacağı bilinmelidir. Zira insanlar duygularını çok farklı şekillerde ifade edebilirler. Duygu tespitinde içinde bulunan ortam, kişinin tavırları, yüz ifadesi, konuşmanın içeriği ve cümle içinde kullandığı kelimelerin anlamları da mutlaka göz önünde bulundurulması gereken faktörlerdir. Oscar töreninde ödül aldıktan sonra Oscar heykelciği elinde iken konuşan ve ağlayan bir sanatçının sevinçten ağladığını sadece sesinin akustik öznitelikleriyle anlayabilmemiz imkansızdır.

Biz insanlar için karşımızdaki kişinin duygusunu anlayabilmek her ne kadar basit ve olağan bir iş gibi görünse de bu işi bir makineye yaptırmaya kalktığımızda olayın ne kadar karmaşık olduğu ortaya çıkmaktadır. İnsanlar kendilerindeki örüntü tanıma yeteneğinin ne kadar eşsiz olduğunun farkında bile değildirler. Duygu tanıma oldukça karışık bir örüntü tanıma problemidir ve pek çok faktörü içinde barındırmaktadır. Bu faktörler arasında kişilerin içinde bulunduğu ortam, kişinin hareketleri, yüz ifadesi, konuştuğu sözcükler ve onların anlamları, sesinin akustik öznitelikleri sayılabilir. Bu çalışmanın kapsamının darlığı nedeniyle bu faktörlerin hepsinin ele alınması mümkün değildir ve sadece sesin akustik öznitelikleriyle ilgilenecektir.

İlk adım olarak insanlar arasındaki konuşmalardan elde edilmiş güvenilir bir veritabanına ihtiyaç duyacağız. Veritabanımız ses dosyalarından oluşmakta ve her dosya belli bir duygu için etiketlenmektedir.

İkinci adım ses dosyalarından duygu analizi için kullanılabilecek sağlam ve güvenilir öznitelikleri çıkarmaktır. Bu özniteliklerin seçimi uygulamamızın başarısında büyük öneme sahiptir. Zira bir sonraki adımdaki yapay zeka teknikleri bu özniteliklere dayanarak karar üreteceklerdir. Öznitelik çıkarma için açık kaynaklı, hızlı ve güvenilir bir program olan openSMILE kullanılacaktır.

Üçüncü adımda elde edilen özniteliklere dayanarak duyguların sınıflandırılması gerçekleştirilecektir. Bu amaçla açık kaynak kodlu Weka aracından yararlanılacaktır. WEKA'nın openSMILE ile uyumlu olması bizi elde edilen verilerin tekrar formatlanması zahmetinden de kurtaracaktır. Doğru bir şekilde kurulumu yapıp çalıştırıldıktan sonra WEKA bize sınıflandırmayla ilgili her türlü bilgiyi zahmetsizce sağlayacaktır.

Son adımda ise mevcut özniteliklere yeni öznitelikler ekleyerek sınıflandırma performanslarını gözlemleyeceğiz.

Ancak şunu belirtmekte fayda vardır ki konuşma tanıma ve türevleri ile ilgili çalışmaların şu anda dünya üstündeki mevcut en karmaşık teknoloji olduğu göz önünde bulundurularak ve bu çalışmanın kapsamının darlığı nedeniyle bu konularda fazla derinlemesine incelemelerde bulunmak mümkün olmadığından insan konuşmasından duygu analizi tanıtıldıktan sonra bunların uygulama alanlarından bahsedilecek ve okuyucuya konunun genel kapsam ve derinliği ile ileriki çalışmaların odak noktalarını kavrayacak bir sezgisellik kazandırılmaya çalışılarak büyük resmi görmesi sağlanacaktır.

1.3 Tezin Bilimsel Katkıları

Bu çalışmada önceki çalışmalara ek olarak yeni bir duygu veritabanı ortaya konacaktır. Bunun yanı sıra sınırlı sayıda öznitelik kullanarak yeterli başarı elde edilebileceği gösterilecek ve önceki çalışmalarla kıyaslaması yapılacaktır. Bugüne kadar yapılan çalışmalarda henüz kullanılmamış sesli harflerin formantlarına dayalı filtre bankaları ve enerji operatörleri öznitelik çıkarmada kullanılacak ve performansa olan katkısı gözlemlenecektir. Yeni enerji operatörü hesaplanırken yerleşmiş geleneklere cesur bir meydan okumayla logaritmik şiddet yerine mutlak şiddet değerleri kullanılacaktır. Çalışma iki farklı veritabanı üstünde ayrı ayrı ve veritabanları karıştırılarak 988 ve 6669 özniteliği kapsayan konfigürasyonlarla ve üç ayrı sınıflandırıcı ile gerçekleştirilecektir. Bu veritabanları üstünde öznitelik seçme konusu da ayrıntılı olarak incelenecektir.

1.4 Tezin Önemli Bulguları

Bu tezde önerilen yeni öznitelikler sınıflandırıcıların neredeyse tamamında başarı oranını olumlu etkilemiştir. Öznitelik sayıları 988 ve 6669 olan 2 ayrı konfigürasyon test edilmiş ve öznitelik sayısının artışının başarı oranlarına bazen pozitif bazen de negatif etkisi olduğu tespit edilmiştir. 6669 öznitelikle yapılan uygulamalar beklenen performanstan uzak kalmıştır. Bunun yanı sıra EmoSTAR'daki hata matrisinde Doğru-Pozitif oranları her duygu için birbirine daha yakın, EmoDB'de ise mutlu ve kızgın sınıflandırmada diğer duygulara göre oldukça fazla hata görülmüştür. Veritabanlarından birinin eğitim diğerinin test seti olarak kullanıldığı çapraz testlerde EmoSTAR'ın eğitim seti olarak kullanılmasının EmoDB'nin eğitim seti olarak kullanılmasına göre daha iyi sonuçlar verdiği görülmüştür. Çapraz testlerdeki nispeten

düşük başarı oranı daha büyük çaplı veritabanlarına ihtiyacı ortaya koymaktadır. Öznitelik seçiciler oldukça iyi performans sergilemişlerdir.

1.5 Tezin Ana Hatları

Bu tezin anahatları aşağıdaki gibi özetlenebilir.

Birinci bölüm literatür özeti ve problemin tanımını, tezi amacını, bilimsel katkılarını ve tezin anahatlarını sunmaktadır.

İkinci bölümde duygunun tanımı üstünde durulacak, bugüne kadar üretilen duygu veritabanları tanıtılacak ve duygu tanımanın uygulama alanlarından bahsedilecektir.

Üçüncü bölüm insan sesinden duygu tanıma çalışmalarında izlenen yol üzerine kuruludur. Bu amaçla kullanılan öznitelikler ve yöntemler irdelenecektir.

Dördüncü bölüm deneysel kurulumun tanıtılmasını amaçlamaktadır. Deneylerimizde kullanılan veritabanları Berlin Duygu Veritabanı ve bizim oluşturduğumuz EmoSTAR tanıtılacak, öznitelik çıkarma için kullanılan openSMILE ve sınıflandırma için kullanılan Weka aracından kısaca bahsedilecektir.

Beşinci bölüm önerilen yeni özniteliklerin tanıtılması ve veritabanları üstündeki uygulamasına ayrılmıştır.

Altıncı ve son bölüm ise sonuçlar ve ileriki çalışmaları kapsamaktadır.

BÖLÜM 2

DUYGU NEDİR?

Duygu tanıma üzerine çalışmalara geçmeden önce duygu tanımı, sınıflandırması, duygu veritabanları ve duygu tanımanın uygulama alanları üstünde durmamız faydalı olacaktır.

2.1 Duygunun Tanımı ve Duygu Türleri

Psikoloji ve felsefede duygu temel olarak psikofizyolojik ifadeler, biyolojik reaksiyonlar ve zihinsel durumlarla karakterize edilen sübjektif vicdani bir deneyim olarak tarif edilir. Alternatif bir tarif özel bir fizyolojik aktiviteyle bağlantılı pozitif veya negatif bir deneyim olarak verilebilir. Duygu üstündeki araştırmalar psikoloji, nöroloji, endokrinoloji, tarih, sosyoloji ve bilgisayar bilimlerindeki çalışmalarla son 20 yıl içinde büyük ivme kazanmıştır. Araştırmalar önceleri daha çok yüz ifadesi üstünde yoğunlaşmış olmasına rağmen sonraları ses, vücut dili, biyolojik belirtiler ve termal kızılötesi yöntemlerde ortaya çıkmış ve bunların bazıları beraber kullanılmıştır (ses ve yüz ifadesi gibi).

Duygu sınıflandırma konusunda üç temel görüş bulunmaktadır. Birinci görüş duyguları kategorik olarak sınıflandırır ve duyguların ayrık ve birbirlerinden belirgin olarak farklı olgular olduğunu iddia eder. Ayrık duygu düşüncesinin öncülerinden Paul Ekman bazı duyguların evrensel olarak kültürden bağımsız olarak tanınabileceğini iddia eder. Ekman 6 temel duygu tanımlamıştır: Kızgın, tiksinti, korku, mutlu, üzgün ve şaşkın.

İkinci görüş ise duyguların 2 veya 3 boyutlu olarak gruplandırılabilirliğini ileri sürer (Şekil 2.1 ve Şekil 2.2). Bu görüşün öncülerinden Robert Plutchnik her ne kadar

duyguların biyolojik olarak yönlendirildiği konusunda Ekman ile aynı görüşte olsa da duygu dairesi denen ve 8 temel duyguyu pozitif ve negatif olarak gruplayan bir teori ileri sürmüştür: Neşeye karşı üzgün (joy-sadness), kızgınlığa karşı korku (anger-fear), güvene karşı tiksinti (trust-disgust) ve şaşkınlığa karşı beklenti (surprise-anticipation). Bazı temel duygular birbirleriyle karışarak daha karmaşık duyguları oluşturabilirler. Temel RGB renklerinin karışarak diğer renkleri oluşturması gibi temel duygularda karışarak tüm duygu spektrumunu oluşturabilirler. Örnek olarak kızgınlık ve tiksinti birleşerek aşağılama duygusunu oluşturabilir. Bradley [16], Lang vd. [17], Osgood vd. [18], Russel ve Mehrabian'ın [19] çalışmaları duyguları 2 ve 3 boyutlu olarak sınıflandırmıştır. İki boyutlu sınıflandırma sadece valans ve aktivite olarak gösterilirken 3 boyutlu sınıflandırma bunlara kontrolü de ekler.

- Valans : Pozitif veya negatif
- Aktivite (Activation/Arousal) : Harekete hazır veya değil
- Kontrol (Dominance/Potency/Power) : Baskın veya uysal anlamlarını (kontrollü veya kontrolsüz) taşımaktadır.

Boyutlu yaklaşım avantajlarına rağmen bazı eleştiriler de almıştır. Birinci olarak duygu boyutlarını 2 veya 3'e indirmek aşırı ve gereksiz bir çabadır ve bilgi kaybına neden olabilir. İkinci olarak bazı duygular birbirinin içine girebilir ve hatta bazıları 3 boyutlu uzayın dışında kalabilir. Üçüncü olarak kafa karışıklığı (confusion) gibi bazı duyguların duygu ile ilgili durumlarının yerleri 3B uzayda henüz belli değildir [7], [20]. Bugün duyguların sınıflandırılması konusunda psikologlar arasında hala bir uzlaşma sağlanabilmiş değildir. Russel ve Carroll (Russel & Carroll, 1999) mutluluk ve üzümlük gibi duyguların birbirlerinin zıttı olduğunu söyler ve bir arada bulunamayacağını söylerken Cacioppo ve Berntson (Cacioppo & Berntson, 1994) ise olumlu ve olumsuz duyguların ayrılabilirliğini, mutluluk ve üzümlük gibi karışık duyguların ise birlikte varılabilirliğini iddia eder. Duygu etiketleme konusunda da araştırmacılar arasında tam bir uzlaşma yoktur ve aynı duygu için farklı isimler kullanılmaktadır. (anger - rage, happiness - joy - elation) [21].

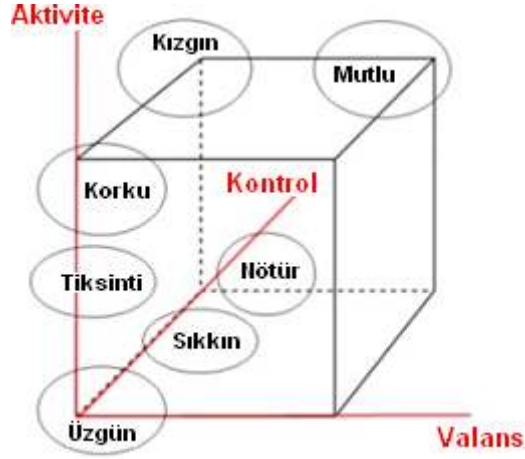
Üçüncü görüş Appraisal teorisi ise duygunun farklı kişilerde farklı yanıtlara neden olan olayların değerlendirilmesinden ortaya çıktığını söyler. Bu tanım boyutlu tanımın

geliştirilmiş halidir ve 3 boyuta yeni boyutlar ilave eder ve boyut sayısını 16'ya kadar yükseltir ancak, araştırmacılar açısından oldukça zorlu bir tanımdır [20]. Appraisal teorisine göre duygu etrafımızdaki önemli olayları veya kişileri değerlendirme yollarıdır. Zarif bir temel sağlar ve bu çatının üstüne kurulacak takip teknikleri geliştirmek amacıyla pilot çalışmalar yapılmıştır. Kategorilerinin çoğunda güvenilirlik sorundur zira doğal duygu verilerinde birden çok kavram eşzamanlı olarak değerlendirilmelidir. Ancak bunlara rağmen Appraisal tabanlı tanımlar dünya çapında kabul görmüştür [22], [23].

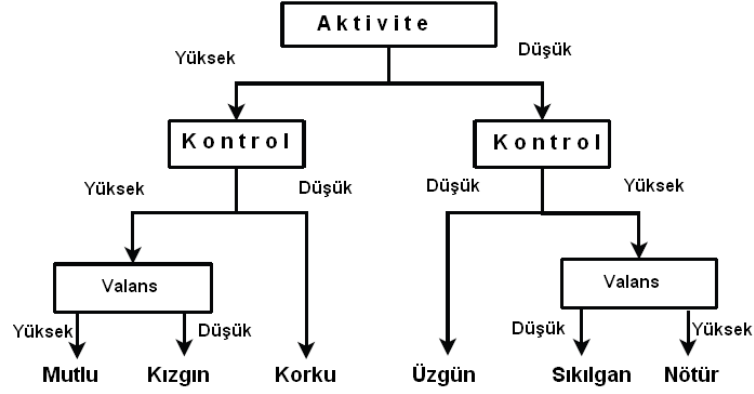
Bu tanımlar arasında duygu tanımda kullanımı en kolay olan ve en sık rastlanan kategorik tanımdır zira duyguları belli temel sınıflara ayırarak sınıflandırıcıların işlerini kolaylaştırmaktadır ancak, son zamanlarda 2 ve 3 boyutlu tanımlar üstünde geliştirilen veritabanları ve çalışmalarda bulunmaktadır. Literatürde duygular pek çok alt sınıfa ayrılmıştır ve bunlardan biri Whissell [24], Plutchik [25] ve Cowie [26]'nin çalışmalarına göre Çizelge 2.1'de verilmektedir [7].



Şekil 2. 1 İki boyutlu duygu uzayı [20]



Şekil 2. 2 Üç boyutlu duygu uzayı [20]



Şekil 2. 3 Üç boyutlu duygu sınıflandırma [27]

Çizelge 2. 1 Whissel [24], Plutchnik [25] ve Cowie [26]'nin çalışmalarına göre duygu türleri

Kabul Edici	Sıkılğan	Haset	Paylaşan	Planlı	Kendine Hakim	Üzgün
Maceracı	Sersemlemiş	Hiddetli	Eleştiren	Hoşnut	Dingin	Alaycı
Şefkatli	Acı	Umutlu	Meraklı	Kendini Düşünen	Çekingen	Tatmin
Korkmuş	Övünçlü	Terk Edilmiş	Cesur	Ümitsiz	Sosyal	Korkan
Saldırgan	Canı Sıkkin	Öfkeli	Meydan Okuyan	Düşman	Kederli	Küçümseyici
Uzlaşmacı	Sakin	Cömert	Zevkli	Aşağılanmış	İnatçı	Ürkek
Büyüleyici	Dikkatli	Sevinçli	İlgi Bekleyen	Sabırsız	Boyun Eğen	Toleranslı
Karışık	Neşeli	Mahzun	Bunalmış	İtici	Şaşkın	Güvenen
Eğlenceli	İğrenme	İhtiraslı	Ümitsiz	Kararsız	Şüpheli	Şefkatsiz
Kızgın	İlgisiz	Kederli	Uzlaşmaz	İçerlemiş	Sempatik	Emin Değil
Can Sıkıcı	İtaatsiz	Sızlanan	Hayal Kırıklığı	Meraklı	Anlaşılmaz	Paylaşmayan
Karşıt	Memnuniyetsiz	Suçlu	Cesareti Kırık	İlgili	Kavgacı	Dost Değil
Beklentili	Tatminsiz	Mutlu	Sinirli	Tahammülsüz	Hazır	Mutsuz
Endişeli	Güvensiz	Çaresiz	İtaatkar	Tahrik Edilmiş	Yenilikçi	Yenilikçi Değil
Duyarsız	İstekli	Tedirgin	Boyun Eğen	Kıskanç	Pervasız	Asempatik
Kaygılı	Coşkulu	Ümitli	Nefretli	Neşeli	Asi	Tereddütlü
Utangaç	Heyecanlı	Kafası Karışık	Panik	Gönülsüz	Reddedici	Kindar
Şaşırtıcı	Utanmış	Aşağılayıcı	Sabırlı	Yalnız	Vicdan Azabı	Uyanık
Özenli	İçi Boş	Hoşnut	Dalgın	Uysal	Gücenmiş	Hayretli
Ürkmüş	Girişken	Aksi	Kafası Dağınık	Küçük düşmüş	Karşı Gelen	Kaygılı

2.2 Duygu İfade Yolları

İnsanlar duygularını pek çok yolla ifade ederler. Bunlar yüz ifadesi, vücut dili, konuşma, içerik, biyolojik belirtiler olarak sayılabilir. İnsanların gerçek hayatta yaptıklarının aksine bu alandaki ilk çalışmaların en önemli eksikliklerinden biri sadece belli bir modalite üstünde durmaktı (sadece ses veya sadece yüz ifadesi). Son zamanlarda yapılan araştırmalar ise birden fazla modaliteyi göz önüne alarak değerlendirme yapmaya çalışmaktadırlar (hem ses hem yüz ifadesi) [28].

2.2.1 Yüz İfadesi

Ekman vd. değişik duygularla ilgili fotoğraflar üstünde yaptıkları araştırmalar 6 temel duygunun (mutlu, üzgün, korku, şaşkın, tiksinti ve kızgın) evrensel olarak kültür ve dilden bağımsız olarak tanınabileceğini göstermiştir. Ekman ve Friesen duygularla paralingistik iletişim arasındaki bağlantıyı gösterebilmek için Yüz Hareketleri Kodlama Sistemini geliştirmişlerdir (Facial Action Coding System, FACS) (Ekman & Friesen, 1978) [28].

2.2.2 Vücut Dili

Sosyal psikoloji üstündeki çalışmalar duygu tanımada insanların vücut hareketlerinin önemine dikkat çekmiştir (Hadjikhani & De Gelder, 2003). Ancak bu konudaki çalışmalar yüz ifadesi kadar ilgi çekmemiştir. Laban'ın vücut hareketlerinin etiketlenmesine ilişkin (Laban & Ullmann, 1988) çalışmalarına rağmen FACS gibi bir sistem geliştirilememiştir ve vücut dili henüz daha keşfedilmemiş bir araştırma alanı olarak varlığını sürdürmektedir [28].

2.2.3 Ses

Konuşma insanlar arasında fikir, gözlem ve duyguları paylaşmanın önemli bir aracıdır. Ancak konuşmanın sadece içeriğine bakıp söyleniş tarzını görmezden gelerek verilmek istenen mesajı tam olarak doğru algılayamayız. Konuşmadaki duygu ifade eden kelimeler dilden dile değişiklik gösterir ve bu konuda bir genelleme yapmak zordur. Ancak paralingistik duygu ipuçları (bazı kelimeleri uzatma, kısaltma veya vurgulama) dilden bağımsız kabul edilebilir ve bu konudaki psikolojik ve psikolingistik araştırmalar

pek çok akustik duygu özniteliğini ortaya koymuştur. Bunlar arasında temel frekans, enerji, ritim ve süre gibi prozodi ile ilgili öznitelikler sayılabilir.

2.2.4 Biyolojik Belirtiler

Fonksiyonel Yakın Kızılaltı Spektroskopi (functional Near Infrared Spectroscopy, fNIRS) ile ölçülen beyin sinyalleri, kafa derisinden ölçülen elektroensefalogram (EEG) ve çevresel sinyaller, kalp ritmi ile ilgili aktiviteler, derideki terleme ve benzeri faaliyetler insanların duygularından kaynaklanan aktivitelere örnek olarak gösterilebilir [28].

2.2.5 Termal Kızılaltı Sinyaller

Nöropsikoloji ve fizyoloji üstündeki bazı araştırmalar insan vücut ısı ile duygusu arasında bağlantı olduğunu göstermektedir. Nakayama vd. maymunların yüzlerindeki ısı değişimleri stresli ve korkutucu anlarda izleyen bir araştırma gerçekleştirmişlerdir. Bu çalışmalar yüzdeki bir ısı azalmanın nötr duygu halinden negatif duygu haline geçişi gösterdiğini ortaya koymuştur (Nakayama vd. 2005). Vianna ve Carrive ise fareler üstünde korkutucu durumlardaki ısı değişimleri izlemişlerdir (Vianna ve Carrive, 2005). Bu çalışmalarda vücudun belli bölgelerinde (göz, kafa ve sırt) ısı artışı olurken, kuyruk ve pençelerde eşzamanlı ısı azalması olduğu gözlemlenmiştir. Biyolojik belirtilerin aksine kızılaltı termal belirtiler insan vücuduna temas olmadan da tespit edilebilir. (Pavlidis vd., 2001) ve (Tsiamyrtzis vd., 2007) tarafından yapılan çalışmalar orbital kaslardaki kan akışındaki artış ile insanlardaki endişe ve stres arasında bir ilişki olduğunu göstermiştir. Benzer şekilde Puri vd. stres halinde alın damarlarındaki kan akışında artış gözlemlemişlerdir [28].

Duygu tanıma birden fazla modaliteyi kullanma düşüncesi 1998'den sonralara kalmıştır (Riseberg vd., 1998). İlk çalışmalar sesli ve görsel verilerin kaynaştırılması ile ilgili idi. Bu konudaki çalışmalar birden fazla modalitenin duygu tanımadaki başarı oranını önemli oranda arttırabileceğini göstermiştir. Picard'ın (Picard vd., 2001) fizyolojik sinyallerden duygu tanıma çalışmaları ortaya çıkmıştır. Ambady ve Rosenthal'ın çalışmaları duygu tanıma yüz ifadesi ve vücut dilinin en etkili ipuçları olduğunu göstermiştir (Ambady ve Rosenthal, 1992). Vücut dili üstündeki çalışmalar ise ancak 2003 yılından sonra gelişmeye başlamıştır (Camurri vd., 2003). Yüz ifadesi ve

vücut dilini birleştiren çalışmalara örnek olarak (Güneş ve Piccardi, 2007; Karpouzis vd., 2007; Martin vd., 2006) gösterilebilir. Termal ve beyin sinyalleriyle ilgili çalışmalarda mevcuttur (Nakasone vd., 2005; Takahashi, 2004; Pun vd., 2006; Puri vd., 2005; Savran vd., 2006; Takahashi, 2004; Tsiamyrtzis vd., 2007) [28].

2.3 Duygu Veritabanları

Duygu tanıma çalışmalarının en önemli öğelerinden biri de hiç şüphesiz duygu veritabanlarıdır. Duygu veritabanlarının elde edilmesi oldukça zahmetli bir iştir. Çünkü pek çok duygu türü vardır ve bunları doğal halleriyle elde edebilmek oldukça ciddi bir çaba gerektirmektedir. Bu veritabanları sesli, görsel veya sesli-görsel olarak hazırlanmaktadır. Duygu veritabanları oluşturulurken duygunun kategorik tanımının kullanılması yaygındır. Bu tür veritabanları doğal, yapay veya zorlanmış duygulardan elde edilir. Doğal olarak elde edilen duygu verileri araştırmalar için ideal bir kaynaktır ancak, bunların elde edilmesi hem çok zor ve zaman alıcı hem de beraberinde etik ve yasal haklar gibi bazı sakıncalar taşımaktadır. Bugüne kadar yapılan çalışmalar yapay olarak elde edilen duygularında doğal duygular kadar gerçekçi akustik özelliklere sahip olabileceğini göstermektedir. Duygu veritabanı geliştirme açısından her ne kadar oldukça fazla sayıda çalışma var ise de bu çalışmalardan elde edilen veritabanlarının güvenilirlikleri ayrı bir sorun teşkil etmektedir. Her şeyden önce bu çalışmalarda üzerinde anlaşılmış ortak bir duygu seti olmadığı gibi ortak bir dilde yoktur. Tek kişi ile hazırlanan bir veritabanının bu alandaki çalışmalara yeterli bir katkı sağlaması beklenemez. Veritabanlarını kullanmadan önce bu açıdan bir değerlendirme yapmak faydalı olacaktır. Ortak bir dil olmayışı da yeterli bir değerlendirme yapma imkanı vermemesi açısından ayrı bir problem teşkil etmektedir. Esasen kategorik duygu tanıma çalışmaları dilden bağımsız kabul edilebilir. Çünkü kategorik sınıflandırma temel duygu türlerinin evrensel olduğunu kabul eder (Ekman vd., 2003) ve her dil ve kültürde benzer olduğunu söyler. Duygu veritabanlarında rol alan konuşmacı sayısı da oldukça değişkendir. Bazı veritabanlarında sadece 1 kişi bütün duyguları taklit etmeye çalışırken bazılarında ise yüzlerce konuşmacı bulunmaktadır. Kullanılan cümlelerde ayrı bir önem taşımaktadır. Zira bu cümlelerin anlamlarının ilişkili oldukları duygu ile bağlantısı olmalıdır. Bütün duygu çeşitleri için aynı cümlelerin kullanılması da ayrı bir tezat teşkil

etmektedir. “Bugün hava çok güzel” cümlesini üzgün veya korkmuş bir şekilde söylemenin söyleyen aktör ne kadar profesyonel rol yaparsa yapsın gerçekçi olduğu iddia edilemez. Konuşmanın içeriği de duygu tanımada oldukça önemli bir faktördür ve bu nedenle yapay duygu veritabanları ile ilgili çalışmalara son zamanlarda eleştiri okları yönlendirilmeye başlanmıştır. İnsanlar duygu tanıyabilmek için pek çok faktörü kullanırlar. Konuşan kişinin kim olduğu, neden ve nerede konuştuğu, karşısındaki insanlarla arasındaki bağlar ve daha pek çok sayılabilecek faktör bunlar arasındadır. Ancak yapay duygular üstündeki çalışmalarda kısmen bu alanda belli bir amaca hizmet etmekte ve bu alanda gelişmelerin önünü açmaktadır. Duygu ifadesi cinsiyete göre farklılık gösterebileceğinden hem kadın hem erkek konuşmacılardan faydalanılmalıdır. Veritabanlarındaki duygu türü ve doğallık farklılığı bunlar üzerinde yapılan çalışmalarında birbiriyle kıyaslanabilmesini zorlaştırmaktadır. Çağrı merkezleri ve TV programlarından elde edilen verilerin halka sunulması etik bazı sorunları da beraberinde getirmektedir. Bu veritabanları yankısız (anechoic) laboratuvarlarda kaydedildiğinden mühendisler tarafından etiketleme zorluğu ortaya çıkmaktadır. Yankısız oda (Şekil 2.4), ses veya elektromanyetik dalgaların yansımaları tamamen emebilen odalardır. Bu odalar dışarıdan gelebilecek gürültülere karşı da izole edilmişlerdir. Böylece tamamen sessiz bir ortam elde edilebilir ve çalışmalar dış etkenlerden etkilenmeden doğru sonuçlar üretebilirler [29].



Şekil 2. 4 Yankısız oda [30]

Veritabanlarının bazıları halka açık olmasına rağmen bazıları ücretli veya halka kapalıdır. Bütün bu sayılan zorluklardan sonra çalışmalarımız sayıları az da olsa ücretsiz yapay duygu veritabanları üzerinde olacaktır. Çizelge 2.2’de bugüne kadar üretilen

duygu veritabanları gösterilmektedir. Bu veritabanlarından bazıları sadece ses verilerinden oluşmakta bazıları ise sesin yanında video ve bazı biyolojik sinyallerde içermektedir.

Çizelge 2.2 Duygu veritabanları [31], [32], [33]

Ad	Dil	# Kişi	Duygular	Tür
Abelin ve Allwood (2000)	İsveç	1	Kz, Kor, M, Ü, Süp, Tik, Bsk, Ut	YD
Alpert vd. (2001)	İng.	22	Dep, N	DD
Alter vd. (2000)	Alman.	1	Kz, M, N	YD
Ambrus (2000), Interface	İng, Sloven	8	Kz, Tik, Kor, N, Süp	YD
Amir vd. (2000)	İbrani	40	Kz, Tik, Kor, Nş, Ü	DD
Ang vd. (2002)	İng	-	Kz, Şaş, N, Kor, Yorgun	DD
Banse ve Scherer (1996)	Alman	12	Kz, M, Ü	YD
Batliner vd. (2004)	Alman, İng	51	Kz, Sık, Nş, Süp	ZD
Bulut vd. (2002)	İng	1	Kz, M, N, Ü	YD
Burkhardt ve Sendlmeier (2000)	Alman	10	Kz, Kor, Nş, N, Ü, Sık, Tik	YD
Caldognetto vd. (2004)	İtalyan	1	Kz, Tik, Kor, Nş, Ü, Süp	YD
Choukri (2003), Groningen	Hollanda	238	-	YD
Chuang ve Wu (2002)	Çin	2	Kz, AntiPa, M, Kor, Süp, Ü	YD
Clavel vd. (2004)	İng	18	N, Kor	YD
Cole (2005), Kids' Speech	İng	780	-	DD
Cowie ve Douglas-Cowie (1996), Belfast Structured	İng	40	Kz, Kor, M, N, Ü	DD
Douglas-Cowie vd. (2003), Belfast Natural	İng	125	-	YD
Edgington (1997)	İng	1	Kz, Sık, Kor, M, N, Ü	YD
Engberg ve Hansen (1996), DES	Danimarka	4	Kz, M, N, Ü, Süp	YD
Fernandez ve Picard (2003)	İng	4	N, Stres	DD
Fischer (1999), Verbmobil	Alman	58	Kz, Dep, N	DD
France vd. (2000)	İng	70	Dep, N	DD
Gonzalez (1999)	İng,İspanyol	-	Dep, N	ZD
Grimm () VAM	Alman	47	Val, Akt, Kontrol	DD
Hansen (1996), SUSAS	İng	32	Kz, Ld Etki, Stres, Görev Stresi	DD, DY
Hansen (1996), SUSC-0	İng	18	N, Stres	A-Stres
Hansen (1996), SUSC-1	İng	20	N, Stres	P-Stres
Hansen (1996), DLP	İng	15	N, Stres	C-Stres
Hansen (1996), DCIEM	İng	-	N, Uyku	ZD
Heuft vd. (1996)	Alman	3	Kz, Kor, Nş, Ü	YD
Iida vd. (2000), ESC	Japon	2	Kz, Nş, Ü	ZD
Iriondo vd. (2000)	İspanyol	8	Kz, Nş, Ü, Süp	YD
Kawanami vd. (2003)	Japon	2	Kz, M, N, Ü	YD
Lathoud, Odobez, Perez (2003) AVIC	İng	21		DD
Lee ve Narayanan (2005)	İng	-	Neg, Poz	DD

Çizelge 2.2 Duygu veritabanları [31], [32], [33] (devamı)

Lieberman (2005), Emotional Prosody	İng	-	End, Kz, M, N, Pan, Ü, Süp	YD
Linnankoski vd. (2005)	İng	13	Dik, Kz, Kor, Ü	ZD
Lloyd (1999)	İng	1	Stres	YD
Makarova ve Petrushin (2002), RUSSLANA	Rus	61	Kz, M, Süp, Ü, Kor, N	YD
Martin (2006) EnterFACE	İng	42		YD
Martins vd. (1998), BDFALA	Portekiz	10	Kz, Tik, M, Küçümseme, N, Ü	YD
McMahon vd. (2003), RESTEIA	İng	29	Kz, Şok, Stres	ZD
Montanari vd. (2004)	İng	15	-	DD
Montero vd. (1999), SES	İspanyol	1	Kz, Tik, M, Ü	YD
Mozziconacci ve Hermes (1997)	Hollanda	3	Kz, Sık, Kor, Nş, Küçümseme, N, Ü	YD
Niimi vd. (2001)	Japon	1	Kz, Nş, Ü	YD
Nordstr vd. (2004)	İsveç	1	M, N	YD
Nwe vd. (2003)	Çin	12	Kz, kor, Tik, Nş	YD
Parlak, Diri (2013) EmoSTAR	İng, Türkçe	-	Kz, N, M, Ü	YD, DD
Pereira (2000)	İng	2	Kz, M, N, Ü	YD
Petrushin (1999)	İng	30	Kz, Kor, M, N, Ü	YD
Polzin ve Waibel (2000)	İng	-	Kz, Kor, N, Ü	DD
Polzin ve Waibel (1998)	İng	5	Kz, Kor, M, N, Ü	YD
Rahurkar ve Hansen (2002), SOQ	İng	6	5 Stres Seviyesi	YD
Scherer (2000b) Lost Luggage	Değişik	109	K, M, Aldırmaz, Ü, Stres	DD
Scherer (2000a)	Alman	4	Kz, Tik, kor, Nş, Ü	DD
Scherer vd. (2002)	İng, Alman	100	2 Görev Stresi, 2 Stres	YD
Schiel vd. (2002), SmartKom	Alman	45	Kz, Tatminsiz, N	DD
Schröder and Grice (2003)	Alman	1	Yum, Modal, Bağırın	DD
Schröder (2000)	Alman	6	Kz, Sık, Tik, Endişe	YD
Schüller (2007) ABC	Alman	8	Sal, Nş, Sarhoş, Kz, N, Yor	YD
Slaney ve McRoberts (2003), Babyyears	İng	12	Onay, Dik, Yasakçı	YD
Stibbard (2000), Leeds	İng	-	-	DD
Tato (2002), AIBO	Alman	14	Kz, Sık, M, N, Ü	DD, ZD
Tolkmitt ve Scherer (1986)	Alman	60	Stres	ZD
Yıldırım ve Çağlar, TURES (2013)	Türkçe	-		YD
Wendt ve Scheich (2002), Magdeburger	Alman	2	Kz, Tik, Kor, M, Ü	YD
Yıldırım vd. (2004)	İng	1	Kz, M, N, Ü	YD
Yu vd. (2001)	Çin	-	Kz, M, N, Ü	YD
Yuan (2002)	Çin	9	Kz, Kor, Nş, N, Ü	ZD

Kz: Kızgın, N: Nötr, Ü: Üzgün, Tik: Tiksinti, Kor: Korku, M: Mutlu, Sık: Sıkın, Nş: Neşeli, Bsk: Baskın, Sal: Saldırgan, Yor: Yorgun, Pan: Panik, Süp: Sürpriz, End: Endişe, Ut: Utangaç, Dep: Depresyon, Akt: Aktivasyon, Dik: Dikkat, Val: Valans, Antipa: Antipati, Neg: Negatif, Poz: Pozitif, Ld: Lombard, A-Stres: Gerçek (Actual) Stres, P-Stres : Fiziksel (Physical) Stres, C-Stres : Sezgisel (Cognitive) Stres, YD: Yapay Duygu, DD: Doğal Duygu, ZD: Zorlanmış Duygu, İng : İngilizce.

2.3.1 Doğal Duygu (Natural/Spontaneous) Veritabanları

Tamamen günlük hayatta kendiliğinden meydana gelen olaylar esnasında elde edilen duygulardır. Elde edilmeleri oldukça zordur ve etik, yasal haklar ve teknik zorluklar (sensörlerin yerleştirilmesi, gürültü, aydınlatma vb.) gibi pek çok sorunu da beraberinde taşımaktadır. Karşılıklı konuşmalar ve etkileşimlerden elde edilebilirler ancak bir duyguyu bir olaya bağlayabilmek pek kolay değildir. Zira duygu oldukça sübjektif bir olgudur ve çelişkili olabilir. Bu nedenle bu tür veriler ya insan-insan etkileşiminden ya da insan-makine etkileşiminden elde edilir. İnsan-makine etkileşiminde makine bir insan tarafından yönlendirilebilir ki bu tür uygulamalar Oz Büyücüsü (Wizard of Oz, WOZ) senaryosu olarak da bilinir. Yapay duygu ile doğal duygu arasında yüz ifadesi açısından ortaya çıkan farklar üstüne Cohn ve arkadaşlarının oldukça geniş kapsamlı çalışmaları bulunmaktadır (Affect Analysis Group, 2008) [28].

2.3.2 Yapay Duygu (Simulated/Acted) Veritabanları

Bir profesyonel tarafından verilen bir cümlelin belli bir duyguyu yansıtacak şekilde okunmasıdır. Yapay duygular duygu tanıma çalışmalarında en güvenilir ve en çok kullanılan kaynaklardır. Zira elde edilmeleri nispeten kolaydır ve doğal duygulardaki etik ve yasal haklar gibi sorunları da taşımazlar. Ancak şunu belirtmekte de fayda vardır ki yapay duyguların doğal ortamdaki duygularla aynı özelliklere sahip olup olamayacakları tartışmalıdır. Wissel (1989) yapay duyguların doğal duygulardan zamanlama ve kullanılan kelimeler açısından farklı olabileceği sonucuna varmıştır. Ekman (2005)'da yapay duygulardaki yüz ifadelerinin doğal duygulardan farklı olacağını söyler. Yüz ifadeleri için geçerli olan bu düşünceler sesli ifadeler içinde geçerlidir. Yapay duygular laboratuvarlarda yankısız odalarda kaydedilmektedirler. Genel olarak yapay duyguların doğal duygulara göre daha abartılı olduğu söylenebilir ve bu nedenle taklit duygular üstünde çalışan bir sistem doğal duygular üstünde yeterli başarıyı gösteremez. Bu dezavantajlarına rağmen yapay duygu veritabanları araştırmalarda çok yoğun kullanılmaktadır. Bu sayede veritabanı tasarımı üstünde yüksek kontrol sağlanabilmekte, cümleler tekrar tekrar okunabilmektedir ve gürültüsüz yüksek kaliteli veritabanları oluşturulabilmektedir. Bu kayıtlar kişi çağrı merkezindeki bir otomatik konuşma sistemiyle konuşurken veya bir insan tarafından seslendirilen (Wizard of OZ,

WOZ) (Fischer, 1999) bir makine ile konuşurken de yapılabilir [31]. Bu konudaki yaygın kullanımlar araç kullanımında stres algılama (Fernandez ve Picard, 2003), öğretici diyaloglar (Litman ve Forbes, 2003; Liscombe, 2005), bilgi sistemleri, insan-robot iletişimi ve sanal ajanlarla etkileşimdir [34].

2.3.3 Zorlanmış Duygu (Elicited/Induced) Veritabanları

Zorlanmış duygular doğal veya yapay değildir. Genel olarak profesyonel olmayan kişilerin profesyonel kişileri taklit etmesiyle elde edilen duygu verileri zorlanmış duyguya örnek olarak verilebilir. Zorlanmış duygular bazen kişilere ilaç verilerek de elde edilebilir. Bir kişi ilaç verilerek stresli bir duygu içine sokulabilir ve konuşması takip edilebilir [7].

Senaryo yaklaşımı (Scherer vd., 1991; Nagae, 1998), bir ilk paragraf ekleme (Fairbanks ve Pronovost, 1939; Davitz, 1964) ve diyalog yaklaşımı (Ito, 1986) gibi zorlama duygu teknikleri literatürde önerilmiştir ancak bu yaklaşımlarla sınırlı sayıda konuşma araştırılabilir. Bu nedenle bazı araştırmalarda metin yaklaşımı yöntemi kullanılır ve her kayıttan önce hedeflenen duyguyu elde etmeye yardımcı olacak bilgiler verilir veya sohbet yapılır. Örnek olarak mutluluk için neşeli bir diyalog, kızgınlık için ceza kesen trafik polisi gibi yaklaşımlar kullanılabilir [9].

2.4 Duygu Etiketleme

Duygu veritabanları için ses kayıtları elde edildikten sonra bu kayıtların belirli duygular için etiketlenmesi gereklidir. Bu sayede verilerin bazıları sınıflandırıcılar için eğitim verisi olarak kullanılmak üzere öznitelik çıkarma işlemine tabi tutulacaktır. Etiketleme işlemi bizzat konuşmacı tarafından yapılabileceği gibi tarafsız bir dinleyici tarafından da yapılabilir. Zorlanmış duygularda ayrıca etiketlenirler. Sadece ses kayıtlarından yapılan duygu etiketlemenin pek çok sorunları da beraberinde getirdiği ortadadır. Zira video görüntüsü olmadan sadece sese bakarak duygu etiketlemeye çalışmak oldukça sübjektif kararlar verilmesine sebep olabilir. İnsanlar karşılarındaki kişinin duygularını sadece konuşmalarındaki tonlara göre değil içinde bulundukları ortama, vücut diline, yüz ifadelerine göre de değerlendirirler. Bu faktörler olmadan duygu etiketlemeye

çalışmak oldukça sıkıntılı bir süreci de beraberinde getirmektedir. Bütün bu faktörleri içine alan bir araştırma bu tezin kapsamının dışındadır.

2.5 Duygu Tanıma Uygulama Alanları

Özellikle son yirmi yılda bilgisayar ve yapay zeka uygulamalarındaki gelişmeler insan konuşması üstündeki çalışmalara ve bunlara yönelik uygulamalara büyük ivme kazandırmıştır. Duygu tanıma çalışmaları başta insan-makine etkileşimi olmak üzere aşağıda belirtilen pek çok alanda önemli rol oynamaktadır.

- **Akıllı Öğrenme Sistemleri**

Öğrenen kişinin duygu seviyesini belli bir düzeyde tutarak öğrenmesini kolaylaştırmaya çalışan sistemlerdir [32].

- **Yalan Makinesi**

Kişilerin yalan söylediğini tespit için kullanılır ve suçlu takibi gibi adli alanlarda kullanılır [32].

- **Bankacılık**

ATM makinelerinin insanlarla konuşarak iletişim kurmasını bir adım daha ileri götürerek bu makinelerin stres, kızgınlık ve aldatma gibi duyguları tespit etmesi amacıyla kullanılır. Ayrıca ATM'nin verdiği sesli cevapların duygu içermesi konuşmanın daha doğal olmasını sağlar [32].

- **Araç İçi Sistemler**

Araç içi sistemlerde duygu tanıma uygulamaları sürücünün o anki duygu halini tespit ederek hatalı sürüş yapmasını engellemeye çalışır [32].

- **Sanal Diyalog Sistemlerinde Prozodi**

İnsan-makine etkileşiminde prozodi kullanımı şaşkınlık ile diğer duyguların ayrılmasında kullanıldığında kelime tanımaya dayalı dil modellerinden daha iyi sonuçlar verebilmektedir. Ginevra Castellano'nun [35] bir çalışmasında, iCats adlı robot çocuklarla satranç oynarken oyundaki pozisyona göre duygusal yüz ve sözcük ifadeleri kullanmakta ve böylece oyuncunun oyuna olan ilgisini arttırabilmektedir. Yapılan

arařtırmalar [36] ses sıkıřtırmanın duygu tanımda performansı olumsuz etkileyebileceđini göstermektedir. Bu nedenle özellikle halk güvenliđi ile ilgili alanlarda güvenlik birimleri ile yapılan konuřmalarda ses sıkıřtırma kullanımı dođru deđildir.

- **Çađrı Merkezleri**

Çađrı merkezleri görevlilerle müşteriler arasında diyalogların en çok yařandığı yerlerdir ve özellikle kızgın, mutlu, üzgün ve nötür duygu ayırımı buralarda önemlidir. Çađrı merkezlerindeki binlerce konuřmanın insan eliyle ayıklanması oldukça zahmetli ve zaman alıcı bir olaydır. Buralarda otomatik duygu tanıma uygulamalarıyla olumsuz içerik geçen konuřmalar tespit edilerek bunların nedenleri araştırılır ve müşteri memnun olmadıysa sorun çözölmeye çalışılır. Çađrı merkezi görevlisi eksik bilgi vererek müşteriyi ikna edemiyorsa görevli bir eğitim sürecine tabi tutulabilir.

- **Sesli Mesaj Sınıflandırma**

Sesli mesajlar içerdikleri duygulara göre sınıflandırılarak gerekli cevaplar otomatik olarak verilebilir.

- **Bilgisayar Oyunları**

Bilgisayar oyunlarında zorluk derecesi oyuncunun duygusuna göre otomatik ayarlanabilir. Eğer oyuncu kızgın ise zor seviye, gevşemişse kolay seviye seçilebilir. Aradaki duygularda da ara seviyeler seçilebilir.

- **Ses Terapisi**

İnsanlardaki konuřma bozukluklarını tedaviye çalışan ses terapistleri hastanın duygu halini anlayarak sorununun ne olduđunu tespit edebilirler. Alzheimer, depresyon, otizm ve řizofreni tespiti gibi konularda konuřma sinyalindeki akustik özelliklerin kullanımı önerilmiştir [7].

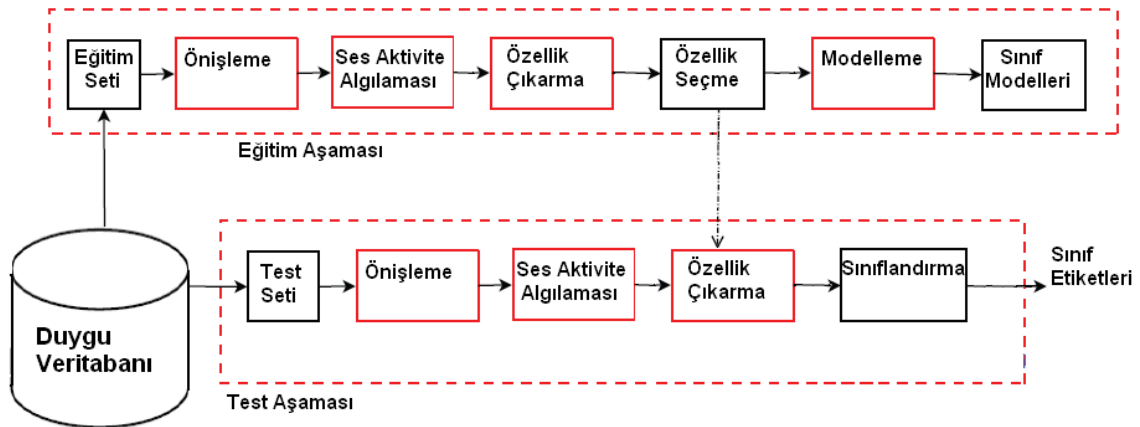
- **Robotlar**

Robotlar hastaneler, süpermarketler, evler ve ofislerde insanlarla iletişim kurarlar. Karřılarındaki kişilerin duygu halini iyi tanıyabilen bir robot daha arkadařça bir diyalog ortamı oluşturabilir. Bunun yanı sıra robotun konuřması da duygu içerdığı takdirde

daha dođal bir konuřma ile karřısındaki kiřilere hitap edebilir ve konuřmanın sıkıcı olmasının önüne geilmiř olunur [32].

İNSAN SESİNDEN DUYGU TANIMA

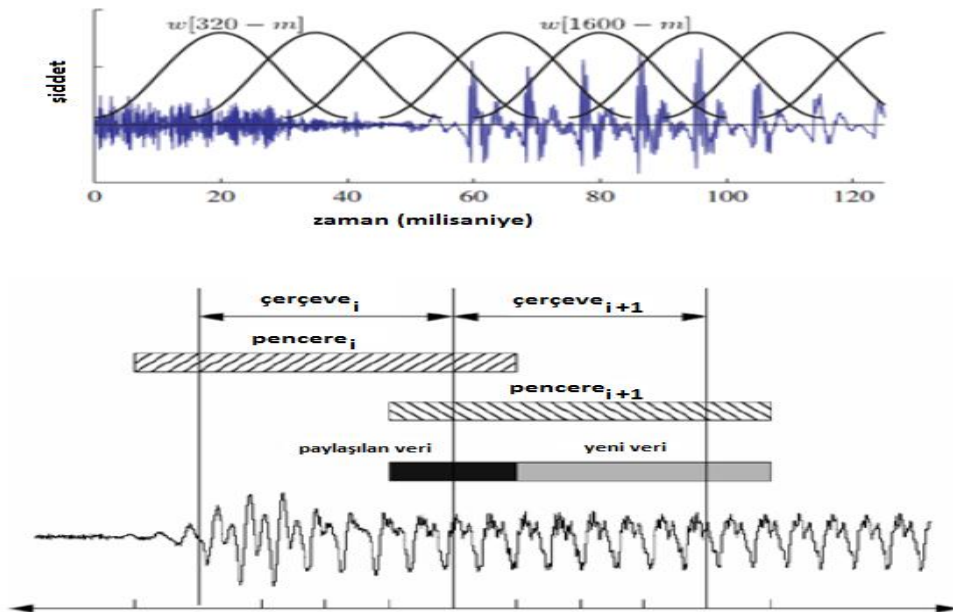
Bu bölüm insan konuşmasındaki duygu ile ilgili olan akustik özelliklere ve bu alandaki mevcut çalışmalara ayrılmıştır. İnsan konuşması ile ilgili çalışmalar her ne kadar 50 yıl kadar öncelere uzansa da bu çalışmalarda daha çok konuşma ve konuşmacı tanıma üzerine yoğunlaşmıştır. İnsan konuşmasından duygu çıkarımı ile ilgili araştırmalar son 20 yıl içinde büyük bir ivme kazanmıştır. Bu çalışmalarda izlenen yol Şekil 3.1’de gösterilmektedir. Bu akış şemasında ilk adım olarak duygu veritabanlarındaki ses kayıtlarından duygu tanımada yararlanılabilecek akustik öznitelikler çıkartılmakta, bu öznitelikler içinden bazıları seçilmekte ve son aşama olarak da sınıflandırma yapılmaktadır. Akustik öznitelikler en genel halleriyle özetlenecek olursa prozodik öznitelikler (temel frekans, enerji, ritim), frekans spektrum öznitelikleri (Mel, MFCC, PLP, Formantlar) ve bunların istatistiksel bazı fonksiyonlarıdır (ortalama değer, maksimum, minimum, çarpıklık, basıklık, birinci delta, ikinci delta). Prozodik ve frekans öznitelikleri alt düzey öznitelikler (Low Level Descriptors, LLD) olarak da adlandırılır.



Şekil 3. 1 Duygu tanıma genel mimarisi

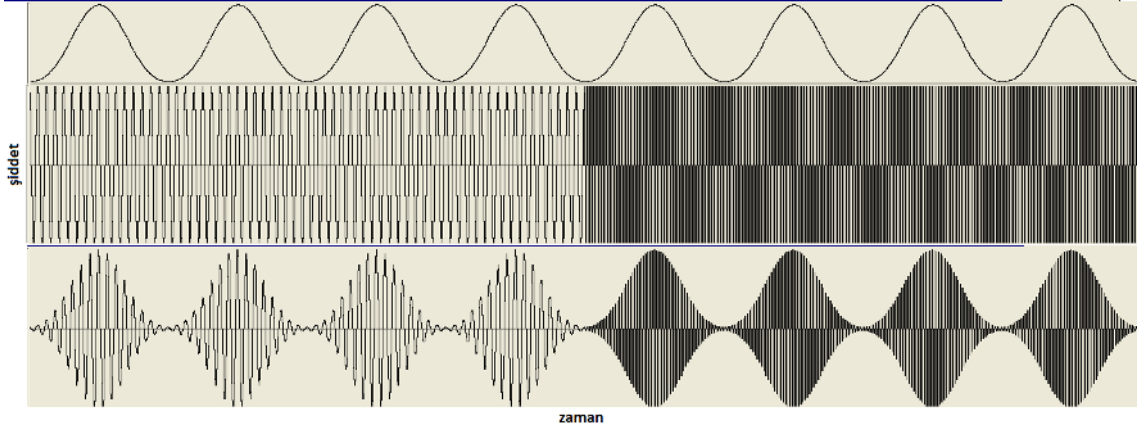
3.1 Ses Dosyaları İçin Önişleme

Duygu veritabanları hazırlanırken kullanılan mikrofon gibi cihazlar ve bu veritabanlarındaki konuşma verilerinin kaydedildiği ortam gürültü ve doğru akım (Direct Current, DC) gibi bazı istenmeyen özelliklere sahip olabilir. Bu tür özellikler hem öznitelik çıkarmada hatalı sonuçlara sebep olurlar hem de sınıflandırıcıların performansını olumsuz etkilerler. Örnek olarak eğer sinyalin içinde DC değeri varsa zamansal özellikler arasındaki önemli bir öznitelik olan Sıfır Geçiş Oranı hatalı hesaplanacaktır. Bu nedenle ses verilerini bir önişlemeden geçirerek veriler arasındaki duygu tanıma ile ilgili olmayan farkların giderilmesi gereklidir. Bir sinyaldeki DC değeri, sinyalin kendisinden sinyalin ortalaması çıkarılarak elenebilir. Gürültü önleme için çeşitli filtreler kullanılabilir ve ayrıca konuşma kayıtlarının sinyal şiddetinin veritabanındaki tüm ses dosyaları için yaklaşık belli seviyede olması sağlanabilir. Konuşma sinyali işlenmek amacıyla belli zaman çerçevelerine (20, 30 milisaniye gibi) bölünür ve üst üste bindirmeli pencere (10, 15 milisaniye gibi) uygulanır. Bu zaman dilimlerinden elde edilen özniteliklerden öznitelik vektörleri oluşturulur ve sınıflandırıcılarda kullanılır. En sık kullanılan pencereler Hamming ve Blackman pencereleridir. Şekil 3.2’de bir sinyale pencere fonksiyonunun üst üste bindirmeli olarak uygulanışı, Şekil 3.3’te ise pencere fonksiyonunun bir sinyal üstündeki etkisi gösterilmektedir.



Şekil 3. 2 Konuşma sinyali için pencereleme [37], [38]

Telefon veya televizyon kanalıyla ses kaydı yaparken ses sıkıştırması yapıp yapılmadığına dikkat edilmelidir. Bazen daha az veri iletmek amacıyla ses verisinin 300 Hz'den düşük ve 3.400 Hz'den büyük olan frekans bölgesi silinmektedir. Bu tür veriler üzerinde temel frekans ve diğer alt düzey öznelik algoritmalarının yanılgıları büyük bir olasılıktır.



Şekil 3. 3 Pencere fonksiyonunun (üstte) bir sinyaldeki (ortada) etkisi (altta)

3.2 Ses Aktivite Algılaması (Voice Activity Detection)

Ses dosyalarından öznelik çıkarma esnasında öncelikle konuşmanın başlangıç ve sonunun belirlenmesi gereklidir. Bu amaçla konuşmanın başındaki ve sonundaki sessiz kısımlar çıkarılmalıdır. Bunun yanı sıra duygu tanımda öznelik çıkarımı için konuşma verisi içindeki sesli kısımların belirlenmesi gereklidir. Duygu tanıma ile ilgili öznelikler genellikle sesli bölgelerden elde edilir.

3.3 Öznelik Çıkarma

İnsan konuşmasından duygu tanıma çalışmalarının dayanağı herhangi bir duygu halinin oluşturduğu fizyolojik etkilerin insanların seslerinde bazı değişikliklere yol açmasıdır. Erkeklerin sesi kızgınlık anlarında kadınlardan daha fazla enerji düzeylerine ulaşır. Yapılan bazı araştırmalarda erkeklerin kızgınlıklarını düşük konuşma hızıyla, kadınların ise aksine yüksek konuşma hızıyla ifade ettiği ortaya konmuştur (Heuft vd., 1996; lida vd., 2000). Bu nedenle duygu tanıma çalışmalarının en önemli adımlarından birisi konuşma verisinden hedeflenen duygularla ilgili en çok bilgi içeren özneliklerin çıkartılmasıdır. Bu amaçla ses dosyaları üzerinde önışleme, zamansal, spektral bilgilerin

çıkarılması ve bazı istatistiksel fonksiyonların bu alt düzey özniteliklere uygulanması işlemleri yapılmaktadır.

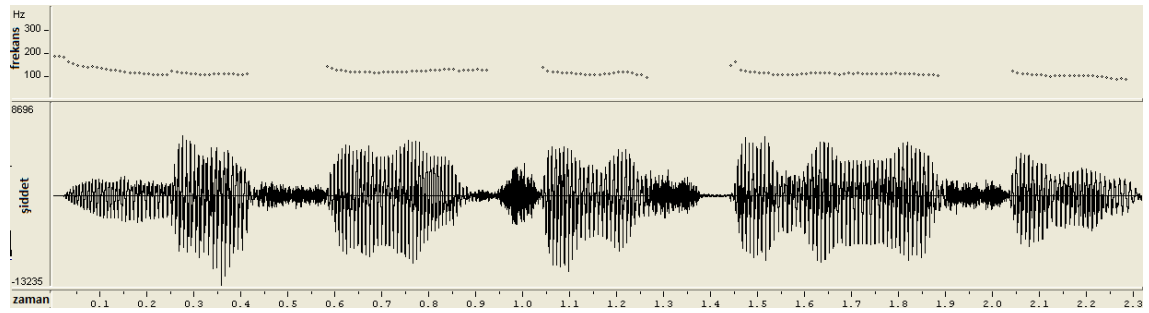
Çizelge 3.1 Duygu tanımda kullanılan alt düzey öznitelikler

Sinyal Enerjisi	1
MFCC (0-12)	13
Mel Filtre Bankaları (0-25)	26
Sıfır Geçiş Oranı	1
Seslilik Olasılığı	1
F0	1
F0 Kontur Zarfı	1
Spektral Enerjiler (0-250, 0-650, 250-650, 1000-4000, 3010-9123 Hz)	5
Spektral Düşüş (25, 50, 75, 90)	4
Spektral Akış	1
Spektral Sentroid	1
Spektral Maksimum Noktası	1
Spektral Minimum Noktası	1

3.3.1 Temel Frekans (Fundamental Frequency, Pitch)

Duygu tanımda kullanılan en önemli akustik özniteliklerden biri temel frekans veya diğer adıyla ses perdesidir ve f_0 olarak gösterilir. Ses tellerinin titreşmesi ile oluşur ve kadın, erkek ve çocuklarda farklı aralıklarda değerler alır. Erkeklerde 80-200 Hz, kadınlarda 150-300 ve çocuklarda ise 200-400 arasında değişir. Kadın sesi ile erkek sesi arasındaki temel farklardan biridir. Kadınlarda temel frekans daha yüksek olduğundan sesleri daha incedir. Frekans spektrumundaki en düşük harmoniktir ve diğer harmonikler f_0 'ın tam katları olarak sıralanırlar. Kişilerin içinde bulundukları duygu hali fizyolojilerine de yansır ve konuşmaları bundan etkilenir. İnsan konuşması oldukça karmaşık bir olgudur ve akciğerlerden çıkan havanın ses tellerinden geçerek ağız ve burun yoluyla dışarı çıkmasıyla oluşur. Bu yolculuk boyunca ağız, burun, dil, dudaklar, dişler ve çene değişik pozisyonlarda bulunarak bazı frekansları geçiren ve bazılarını geçirmeyen bir nevi filtre gibi davranır ve birbirinden farklı seslerin oluşmasını sağlar. Duygu haline göre temel frekansta da önemli değişiklikler meydana gelir. Şekil 3.4 ve 3.5’de kızgın ve nötr bir konuşmadaki temel frekansın zaman içindeki değişimi gösterilmektedir. Görüldüğü gibi nötr konuşmada temel frekans zaman içinde hemen hemen sabit bir değer etrafında gezerken kızgın konuşmada ise oldukça yüksek dalgalanmalar oluşturmaktadır. Temel frekans tespiti oldukça geniş araştırılmış bir

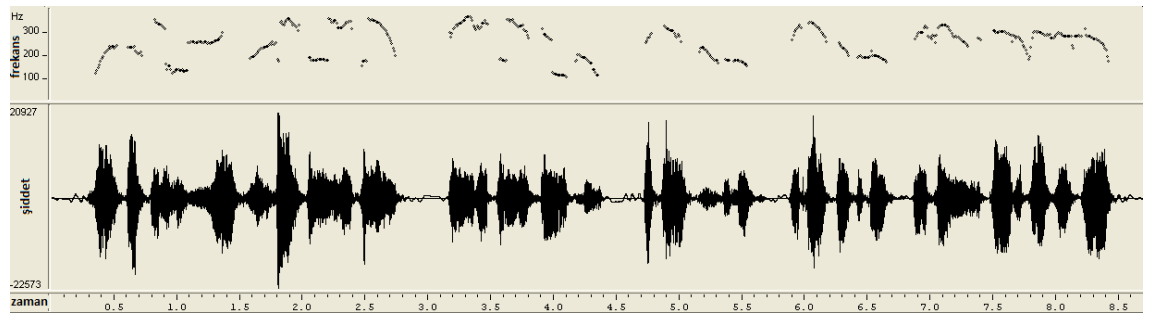
konudur ve pek çok metot bulunmaktadır. Bu metotlar konuşma sinyalindeki sesli bölgelerden temel frekansı tespit etmeye çalışırlar. Temel frekans tespiti hataya meyillidir ve sınıflandırıcıların performansını olumsuz etkileyebilir (Batliner, 2007). Ancak, FAU Aibo veritabanı üzerinde yapılan çalışmalar (Steidl vd., 2008) bu hataların mevcut teknikler sayesinde oldukça azaltılabildiğini göstermiştir [34]. Literatürde uzatma-kısaltma (jitter) ve vurgulama (shimmer) olarak adlandırılan özellikler temel frekans ile ilgilidir ve jitter temel frekanstaki frekans değişimini, shimmer ise temel frekanstaki şiddet değişimini ifade etmek amacıyla kullanılmaktadır.



Şekil 3. 4 Nötr konuşma temel frekans çizgisi

Temel Frekans Tespiti İçin Kullanılan Metotlar

- Otokorelasyon Metodu
- Harmonik Katsayılar Metodu
- Yin Metodu
- Dalgacık Dönüşümü [31]



Şekil 3. 5 Kızgın konuşma temel frekans çizgisi

3.3.2 Konuşma Hızı ve Kalitesi (Speaking Rate, Voice Quality)

Konuşma esnasında duygu haline göre bazı kelimeler uzatılır veya kısaltılır. Konuşma içindeki sessiz ve sesli harflerin oranları (Noise to Harmonic Ratio) ve aralardaki

duraklamalar farklılık gösterir. Benzer şekilde temel frekansta ve enerjide dalgalanmalar meydana gelmektedir. Konuşma kalitesi ise oldukça karmaşık ve tespit edilmesi güç bir özelliktir. Konuşma içindeki bazı kelimelerin uzatılması, duraksamalar konuşma kalitesine etki eden faktörler olarak sayılabilir.

3.3.3 Teager Enerji Operatörü

Teager tarafından başlatılan doğrusal olmayan ve çok kaynaklı konuşma modelleme yaklaşımları çok değişik duygu öznitelikleri çıkarmanın önünü açmıştır [7], [31]. Bazı araştırmacılara göre [39], [40], [41] kızgınlık veya stres anında sesin büyük kısmının temel frekansta titreşen laminar akış (ses yolu içinde sesli harfler tarafından oluşturulan akış, sessiz harfler turbulent akış oluşturur) tarafından üretilmesine rağmen hızlı hava akımının oluşturduğu girdaplar temel frekanstan farklı sinyallerinde oluşmasına neden olurlar.

Kaiser [40] konuşma için anlık enerji operatörünü Teager Enerji Operatörü olarak ayrık zaman düzleminde Denklem 3.1'deki gibi tanımlamıştır:

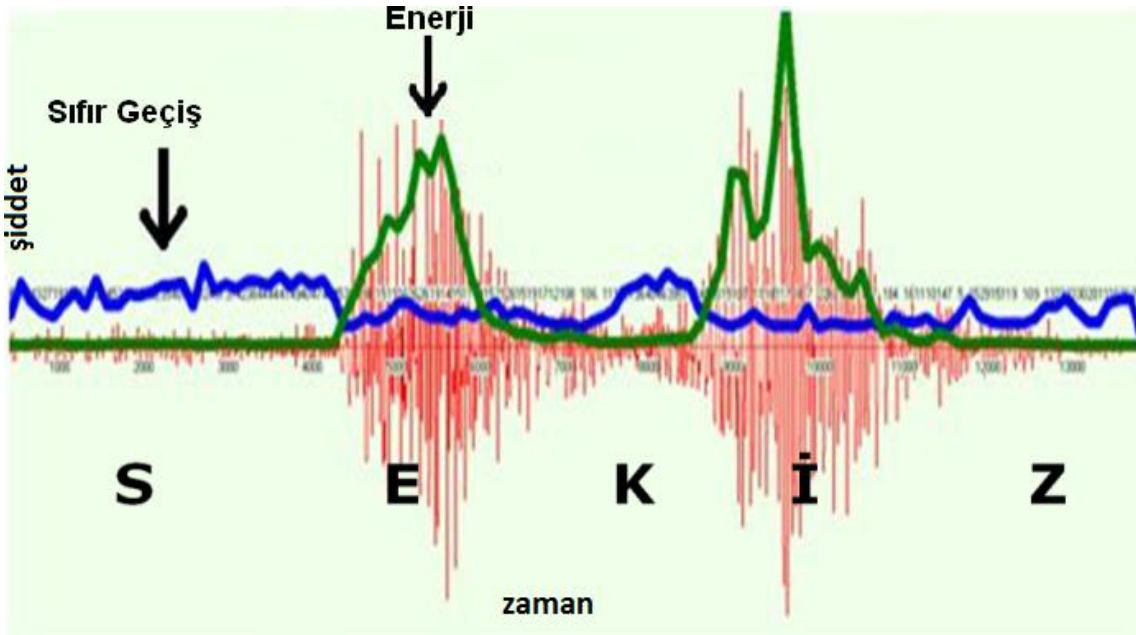
$$\Psi(x[n]) = x^2[n] - x[n+1] \times x[n-1] \quad (3.1)$$

TEO'nun bir avantajı sinyalin harmonik yapısındaki anlık değişimleri takip etmesidir. TEO tabanlı özniteliklerin kullanımı kızgın, nötr, mutlu ve üzgün duygu tespitinde f0 ve MFCC'den daha iyi sonuçlar vermektedir. Nwe vd. [42] SUSAS veritabanında 5 düzeyli stres tespitinde %86, Torabi vd. [43] TEO, temel frekans ve LFPC öznitelikleri ile SUSAS veritabanında yapay stres tespitinde %89 başarı oranı elde etmiştir.

3.3.4 Sıfır Geçiş Oranı ve Kısa Zaman Enerji

Sıfır Geçiş Oranı (Zero Crossing Rate, ZCR) ve Kısa Zaman Enerji (Short Time Energy, STE) özniteliklerini beraber incelemekte fayda vardır. Bu iki öznitelik genellikle sesli ve sessiz harflerin ayırt edilmesinde kullanılırlar. Sessiz harflerde yüksek frekanslı bileşenler daha baskın olduğundan ZCR sesli harflere göre daha büyük değerler alır ancak yüksek frekanslarda sesin şiddeti azaldığından STE değeri azalmaktadır. Şekil 3.6'dan da görülebileceği gibi bu iki öznitelik sesli ve sessiz harfler boyunca birbirlerine

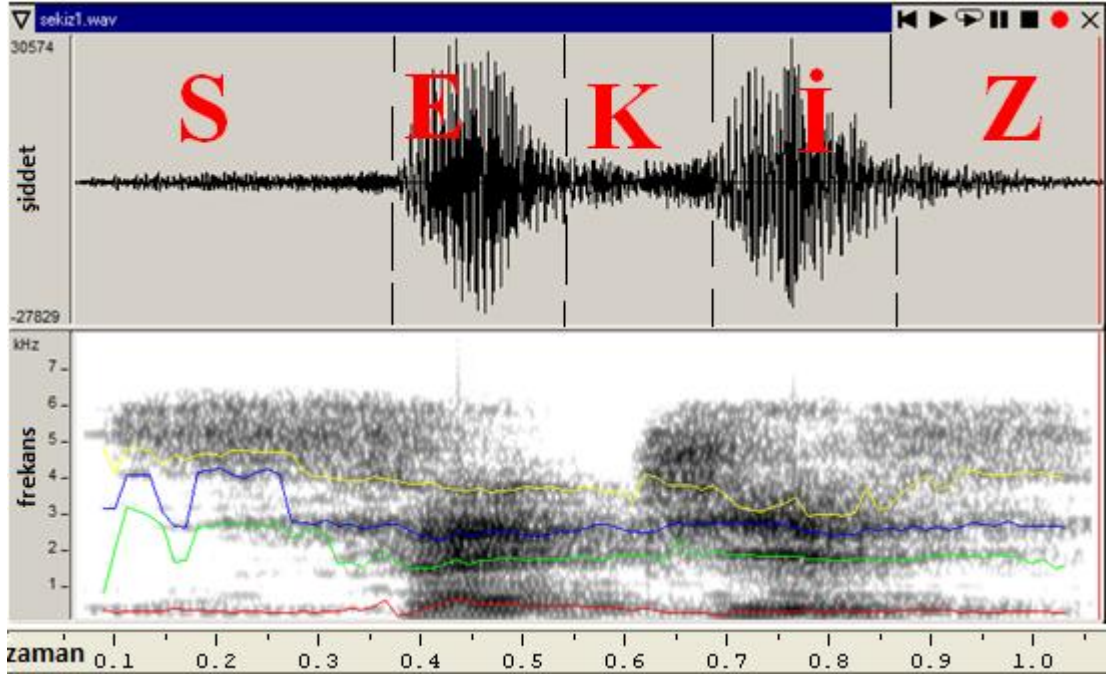
zıt deęerler almaktadırlar. Sesli sessiz harf ayırımı ve kelime sonu bulma konuşma tanımadaki önemli problemlerdendir.



Şekil 3. 6 Sıfır geçiş oranı ve kısa zaman enerji çizgileri

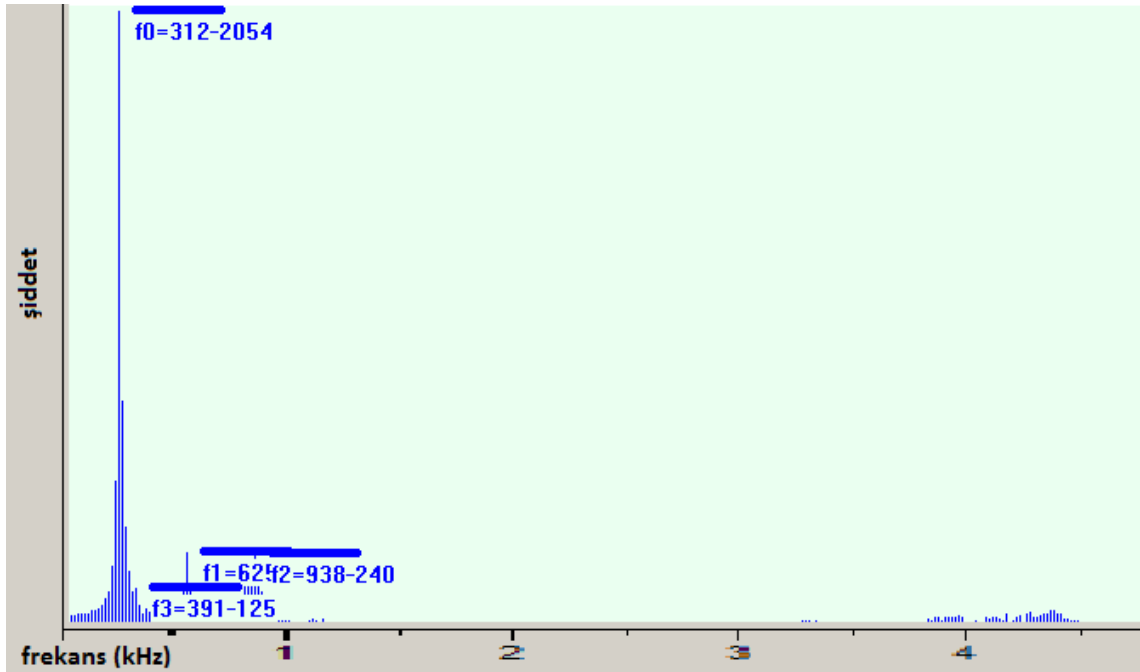
3.3.5 Frekans Spektrumu ve Formantlar

Konuşma sinyali belli zaman aralıklarıyla çerçevelere bölündükten sonra her bir çerçeve için frekans ve şiddet öznitelikleri elde edilir. Bu özniteliklerin kendileri ve zamana göre değişimleri duygu tanıma için önemli ipuçları taşıyabilmektedir. Frekans spektrumu sinyali oluşturan formantlardan oluşmaktadır. Frekans özniteliklerinde bu formantlardan 4-10 tanesi kullanılabilir. Formantların elde edilmesinde LPC (Linear Predictive Coding) ve FFT (Fast Fourier Transform) yöntemi yaygın olarak kullanılmaktadır. Hızlı Fourier dönüşümü formantların tespitinde kullanılmasına rağmen frekans spektrumunu daha iyi ifade eden Mel Filtre Bankaları gibi metotlarda kullanılmaktadır. Şekil 3.7'de sekiz kelimesi için spektrogram ve formant çizgileri gösterilmektedir.

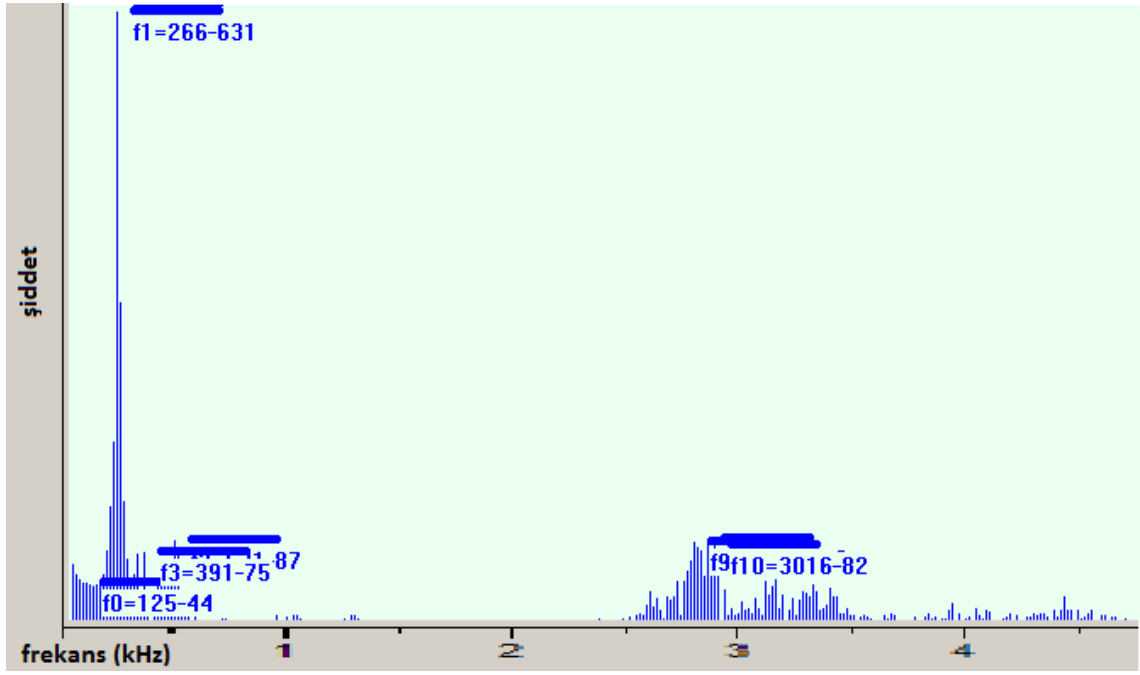


Şekil 3. 7 Sekiz kelimesi için formant izleme

Şekil 3.8 ve 3.9’da U ve İ harflerinin FFT uygulanarak elde edilen frekans spektrumu gösterilmektedir. U harfi için f_0 312 Hertz frekans ve 2054 şiddet değerlerine sahiptir.

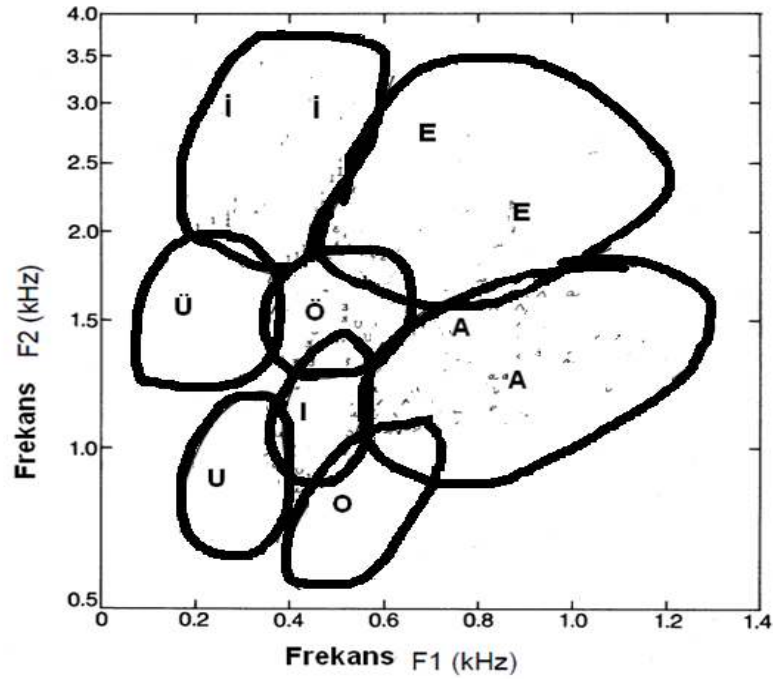


Şekil 3. 8 U harfi için FFT uygulaması



Şekil 3. 9 İ harfi için FFT uygulaması

Her ne kadar sesli harfler frekanslarla birbirinden tamamen ayırt edilebilse de sessiz harfler sadece frekanslarla ayrılamaz. Tamamen aynı frekanslarla birbirinden çok farklı sessiz harfleri oluşturabilmenin mümkün olduğu göz ardı edilmemelidir. Şekil 3.10 Türkçe sesli harf F1 ve F2 düzlemini göstermektedir.



Şekil 3. 10 Türkçe sesli harfler için F1 ve F2 formant düzlemi

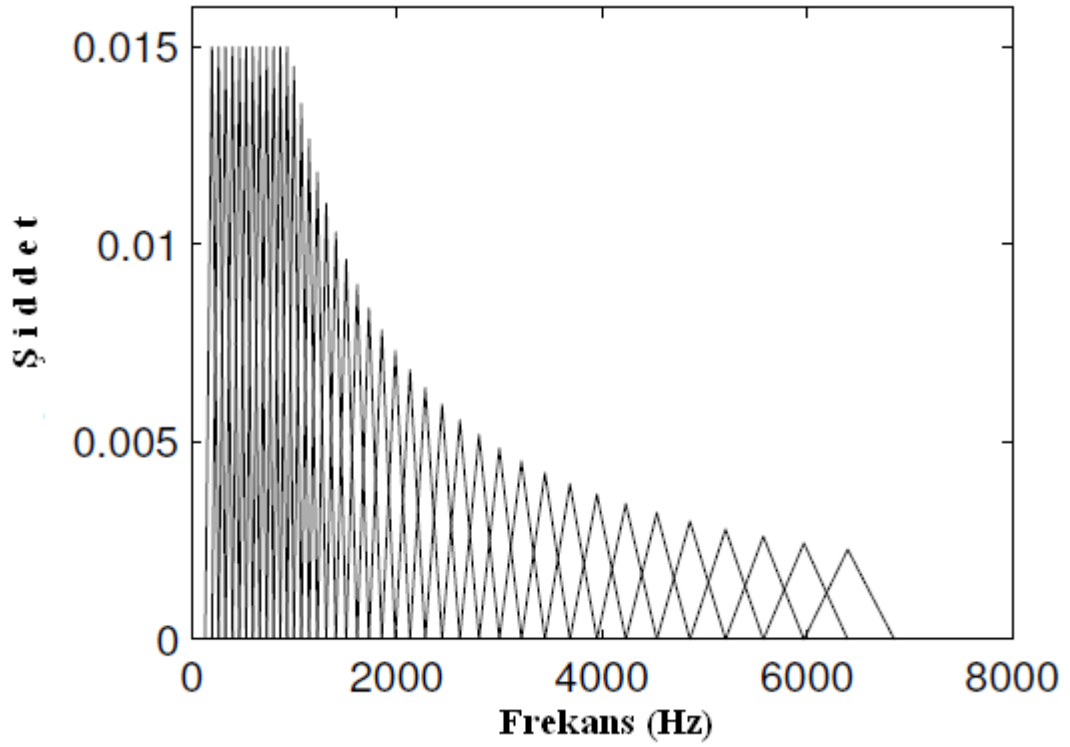
3.3.6 Mel Frekans Kepstrum Katsayıları

Mel Frekans Kepstrum Katsayıları (Mel Frequency Cepstrum Coefficients, MFCC) insan kulağının özelliklerini daha iyi ifade edebilmek amacıyla geliştirilmiştir. İnsan kulağı yüksek frekanslara karşı daha duyarlıdır. Özellikle 3.000 Hz civarındaki frekanslar insan kulağında daha baskın hissedilirler. Öyle ki 3.000 Hz'deki bir frekans bileşeni şiddeti 10 kat daha fazla olan 100 Hz'deki bir frekans bileşenini bastırabilir. Kepstrum spektrumun tersinden türetilmiş bir terimdir ve birimi frekansın tersi olan kafrens'dir (quefrequency). Kepstrumun avantajı gürültüden daha az etkilemesi ve spektrumdaki periyodikliği ve değişikliği daha iyi yakalayabilmesidir. MFCC, insan kulağının davranışını taklit eden ve FFT tabanlı olarak hesaplanan bir sayısal analiz tekniğidir. İnsan duyma sisteminden esinlenilerek tasarlanmış olan Mel frekans ölçeğine (Şekil 3.11) göre hesaplanmaktadır. Şekil 3.12'de konuşma tanıma uygulamalarında izlenen yol gösterilmektedir. MFCC analizi ile elde edilen sayılar MFCC katsayıları (MFCC) olarak adlandırılır. Mel birimi doğrusal bir frekans eksenine göre değil, insan kulağının hangi sesi nasıl algıladığına göre geliştirilmiştir. Mel ölçeği ile frekans ölçeği arasındaki dönüşüm Denklem 3.2'de verilen eşitlikle sağlanmaktadır.

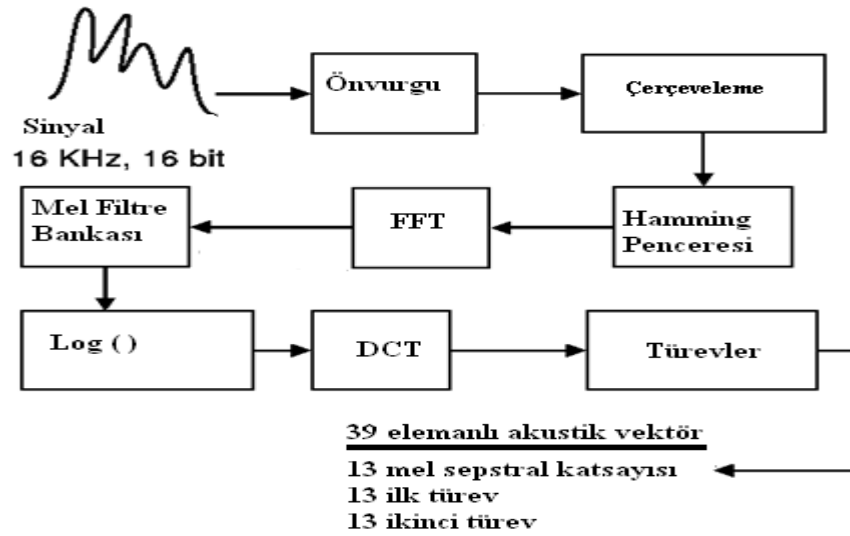
$$F_{mel}(f_{Hertz}) = 2595 \log_{10}\left(1 + \frac{f}{700}\right) \quad (3.2)$$

Buna göre Mel ölçeği 1.000 Hz'den düşük frekanslar için doğrusal, 1.000 Hz'den yüksek frekanslar için ise logaritmik değerlerde dağılım gösterecektir. Bu algısal spektrumu uygulamanın bir yolu Mel ölçeğine göre dağılım gösterecek filtreler tasarlamaktır. Söz konusu filtreler sabit Mel frekans aralıklarıyla dizilirler. MFCC katsayılarını hesaplamak için her çerçeveye aşağıdaki adımlar uygulanır:

- Hızlı Fourier Dönüşümü (Fast Fourier Transform, FFT).
- Hızlı Fourier Dönüşümü katsayıları Mel filtre bankasının genlik frekans cevabına göre ağırlıklandırılır.
- Her filtre bankasının logaritmik enerjileri hesaplanır.
- Ayırık Kosinüs Dönüşümü bulunur.
- Katsayılardan genellikle 2 ile 13 arasındakiler kullanılır.



Şekil 3. 11 Mel filtre bankaları



Şekil 3. 12 Konuşma sinyali için öznitelik çıkarma

3.3.7 Türevler

Fark alma işlemi (delta), öznitelik vektörlerinin herhangi bir biçimde farklarının alınmasıdır. Fark vektörleri sesli ifadenin kısa süreli değişimlerinin açığa çıkmasını

sağlayan öznitelik vektörleridir. Fark alma işlemi sonucu elde edilen vektörler zaman boyutunda öznitelik vektörlerinin benzerliğini açığa çıkarır (Denklem 3.3).

$$\Delta_i(t) = \frac{\sum_{n=1}^{n_0} n \times (S_i(t+n) - S_i(t-n))}{2 \times \sum_{n=1}^{n_0} n^2} \quad (3.3)$$

Bu diferansiyel denklemin sonucunda bir delta parametresi elde edilir. Tipik n_0 değeri 2 olarak seçilmektedir. Aynı şekilde ikinci seviyeden bir türev aldığımızda ise delta-delta parametresini elde ederiz. Delta-delta parametreleri aynı formülün delta parametrelerine uygulanmasıyla elde edilir. Bazı uygulamalar üçüncü seviye fark alma işlemini de kullanırlar.

3.3.8 PLP

PLP (Perceptual Linear Prediction) yöntemi, DFT (Ayrık Fourier Dönüşümü) ve LP tekniklerinin birleştirilmesi ile sesli ifade parametrelerinin hesaplanmasıdır. Bu yöntem insan kulağının duyma sistemini LP yönteminden daha iyi modellemeye yöneliktir. LP tekniğinde sesli ifade modellenirken tüm frekanslardaki sesler eşdeğer tutulmaktadır. Bu durum insan kulağıyla uyumlu değildir. 800 Hz değerinden daha düşük frekanslarda duyma miktarı frekansla birlikte düşer. İnsan kulağı daha çok duyma frekans aralığının ortasındaki 3.000 Hz civarındaki frekanslara duyarlıdır. Bu sorunu çözmek için birçok çalışma yapılmıştır. Bu çalışmalardan biri de bulunan LP katsayılarının Mel ölçeğine uyarlanması olmuştur. Bir başka yaklaşım da LP tekniği uygulamadan önce sesli ifadenin güç spektrumunun alınmasıdır. PLP yöntemi de bu yaklaşımı kullanmaktadır.

3.3.9 Öznitelik Fonksiyonları

Duygu tanıma ile ilgili alt seviye öznitelikler (Low Level Descriptors) elde edildikten sonra bunlara ait bazı fonksiyonların elde edilmesi ve sınıflandırmada kullanılması başarı oranlarında olumlu bir etki yapmaktadır. Bu konudaki ilk çalışmalarda kullanılan öznitelikler sadece alt seviye özniteliklerle özellikle prozodik özniteliklerle sınırlıyken gelişen teknoloji ve yapay zeka çalışmalarındaki ilerlemeler sayesinde çok daha fazla özelliğin kullanılması mümkün olmaktadır. Bugün bu tür çalışmalarda kullanılan öznitelik sayısı binlerle ifade edilmektedir. Fonksiyonlar Çizelge 3.2’de gösterilmiştir.

Duygu tanıma çalışmalarında kullanılan öznitelikler öznitelik seçme algoritmalarıyla otomatik olarak seçilebileceği gibi uzman seçimine dayalı yöntemlerde bu tür çalışmalarda rol alabilmektedir. Uzman görüşüne dayalı yöntemler makine gücü ile tespit edilmesi zor olan Teager Enerji, filtre bankalarının uzun-zaman zamansal zarfları veya ses yoluyla ilgili sesli harflerin uzaması veya daralması gibi özelliklerin kullanılmasında etkili olabilir. Uzman görüşüne dayalı yöntemlerin ileriki yıllarda yapılacak çalışmalarda akustik özniteliklerin hesaplanmasında başvurulan bir akım olması beklenmektedir [9], [34].

Çizelge 3. 2 Duygu tanımda alt düzey özniteliklere uygulanan fonksiyonlar

Maksimum ve Minimum Değerler ve Göreceli Yerleri	4
Maksimum ve Minimum Değerler-Aritmetik Ortalama	2
Maksimum ve Minimum Değerlerin Aralığı	1
Aritmetik, Geometrik ve Kuadratik Ortalama	3
Mutlak ve Sıfırdan Farklı Değerlerin Aritmetik, Geometrik ve Kuadratik Ortalaması	6
Sıfırdan Farklı Değerler	1
Dörttebirler ve Dörttebirler Arası Menzili	6
Yüzdelikler ve Yüzdelikler Arası Menzili	2
Standart Sapma, Varyans, Basıklık, Çarpıklık, Sentroid	5
Sıfır Geçiş Oranı	1
Tepe Sayısı, Tepeler Arası Ortalama Uzaklık, Tepelerin Aritmetik Ortalaması, Tepelerin Aritmetik Ortalaması-Genel Aritmetik Ortalama	4
Lineer Regresyon Katsayıları ve Karşılık Gelen Yaklaşık Hata	4
Kuadratik Regresyon Katsayıları ve Karşılık Gelen Yaklaşık Hata	5
Ayrık Kosinüs Dönüşümü (DCT)	6

3.4 Öznitelik Seçimi

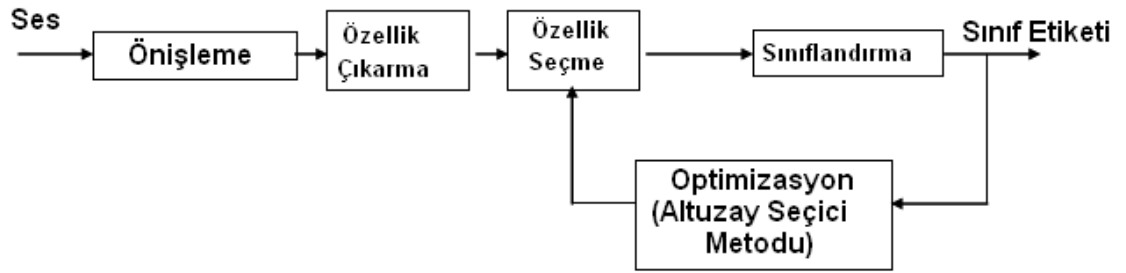
Araştırmalar duygu tanıma çalışmalarında bilgi kazancı olmayan özniteliklerin kullanılmasının performansı olumsuz etkilediğini göstermektedir. Bu nedenle öznitelik seçimi hem performansı arttırmakta hem de hesaplama maliyetini azaltmak açısından önem kazanmaktadır. Genel olarak duygu tanımda öznitelik seçiminde filtreler ve altküme seçiciler (wrapper) kullanılmaktadır. Filtre metodu sınıflandırma sonuçlarını göz önüne almadan sınıflar arası ayırımın artmasına sınıf içi ayırımın ise eşzamanlı olarak azalmasına bakarak optimal çözümü bulmaya çalışır. Alt küme seçiciler ise test edilen öznitelik dışındaki öznitelikleri sabit tutarak test edilen özniteliğin sınıflandırma sonuçlarına etkisine bakarak en yüksek doğruluk oranını veren öznitelikleri seçmeye

çalışır. Filtre metodu altküme seçme metoduna göre daha başarısızdır ancak altküme seçme metodu hesaplama maliyeti açısından daha pahalıdır [7].

Temel Bileşen Analizi (Principal Component Analysis) ve Doğrusal Ayırt Edici Analizi (Linear Discriminant Analysis) filtre metotlarına örnek olarak verilebilir. Sıralı İleri Seçim (Sequential Forward Selection) ve Sıralı Kayan İleri Seçim (Sequential Floating Forward Selection) metotları da altküme seçici (wrapper) metotlara örnektir.



Şekil 3. 13 Öznitelik seçimi için filtre yaklaşımı



Şekil 3. 14 Öznitelik seçimi için alt uzay seçme yaklaşımı

3.5 Sınıflandırma

Bu bölümde duygu tanıma çalışmalarında kullanılan sınıflandırma yöntemlerinden kısaca bahsedilecektir.

3.5.1 kNN

Sınıflandırma problemlerinde çok sık kullanılan bir yöntemdir. Özellikle sınıfların dağılımı konusunda bir ön bilgi yoksa Öklidyen uzaklık kullanılarak sınıflar arası uzaklık hesaplanır ve örneğe en yakın sınıf atanır. Etiketli verileri kullanarak uzaklık hesapladığından kNN eğitici bir sınıflandırma yöntemidir ve gürültüden fazla etkilenmez. kNN metodunun dezavantajı optimum komşu sayısının seçiminin başarı oranını doğrudan etkilemesidir. Hesaplama maliyeti yüksektir ve uzaklık seçiminde hangi özniteliklerin kullanılacağı belirsizdir. Bu yöntemden elde edilen sonuçlar

problemin doğrusal olmayan (nonlinear) yapısını ortaya koymakta ve daha gelişmiş metotların gerekliliğini göstermektedir.

Shami vd. [44] kNN metodunu SVM ve Adaboost karar ağaçları ile dört veritabanını karıştırarak karşılaştırmış ve sınıflandırıcıların benzer sonuçlar verdiğini göstermiştir.

3.5.2 GMM

İstatistiksel bir öznitelik modelleme ve sınıflandırma yöntemidir. Bir sınıfa ait örneklerin dağılımı Gaussian bir fonksiyon olmasa bile birkaç Gaussian fonksiyonun karışımı şeklinde gösterilebilmektedir. Bu tür modellerde eleman sayısı arttırılarak istenen hassasiyette bir sınıflandırma sağlanabilir. Karışım katsayıları Expectation Maximization algoritması ile hesaplanır. Karar Maximum Likelihood modeline göre verilir.

3.5.3 SVM

SVM ikili sınıflandırma problemlerinde son yıllarda oldukça yaygın kullanılan bir metottur. Çok boyutlu uzayda hiper düzlem oluşturulur ve bu hiper düzlemler iki sınıfı birbirinden ayırt etmek için kullanılır. İyi bir hiper düzlem her iki sınıfa da mümkün olduğunca uzak kalmalıdır. SVM metotları başarı açısından GMM metotlarından biraz daha iyi sonuçlar verse de hesaplama maliyetleri açısından pahalıdırlar.

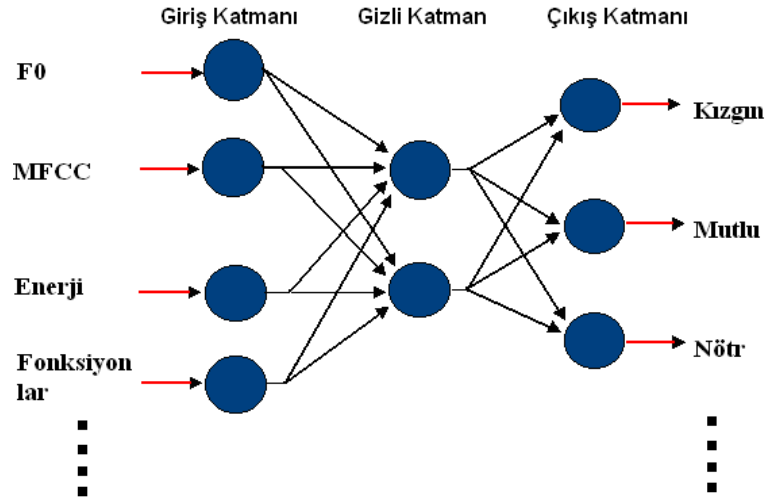
3.5.4 HMM

İstatistiksel bir Markov modelidir. Mevcut durum bir önceki durumlara bağlıdır. Markov zinciri bir sınıfın özelliklerinin zamansal değişimini ve yapısını gösterir. Maksimum Likelihood sınıflandırma kullanır. Mevcut örüntü modellerini kullanarak bilinmeyen giriş örüntüsünün olasılıklarını bulur ve en yüksek olasılıklı örüntüyü seçer. Nwe vd. [45] HMM kullanarak 6 duyguyu LFPC (Short Time Log Frequency Power Coefficients) ile sınıflandırmaya çalışmış ve %80'lik bir başarı oranına ulaşmıştır.

3.5.5 ANN

Yapay Sinir Ağları bir giriş verisi ile çıkışlar arasındaki ilişkiyi bulmaya çalışan katmanlı öğrenme makineleridir. Bir öğrenme algoritması ile eğitilirler. Giriş, gizli ve çıkış katmanlarından oluşurlar ve bu katmanlarda değişik sayılarda düğümler bulunur. Gizli

düğümünün her biri için eğitim algoritması tarafından optimal bir ağırlık değeri bulunur. Duygu tanımadaki eğitim verisinin azlığı göz önüne alındığında GMM metoduna göre daha avantajlıdır.



Şekil 3. 15 Yapay sinir ağı

3.5.6 Karar Ağacı Algoritmaları

Bu tür algoritmalar karar ağaçlarıyla çalışırlar. Karar ağaçlarındaki yapraklar sınıfları veya sınıflandırma sonuçlarını, düğümler ise bu sınıflandırmayı üreten ve birbirini takip eden özniteliklerin birleşimini temsil eder [46].

DENEYSEL KURULUM

Bu çalışma EmoSTAR ve EmoDB (Berlin Duygu Veritabanı) duygu veritabanları üzerinde gerçekleştirilmiştir. Bu alandaki çalışmalardan farklı olarak makine öğrenmesi teknikleri ile bir veritabanından elde edilen modeller, diğer veritabanında test edilecektir. Bu tür veritabanları arası çalışmalar literatürde seyrek olarak bulunmaktadır. Öznitelik çıkarma işlemi için openSMILE aracı kullanılmaktadır. Çıkarılan özniteliklerden farklı öznitelik seçme yöntemleri kullanılarak öznitelik seçimi yapılacak ve Weka aracında sınıflandırma amacıyla kullanılacaktır. Bu amaçla, ilerleyen bölümlerde EmoSTAR, EmoDB, openSMILE ve Weka'nın kısa bir tanıtımı yapılmıştır.

4.1 EmoSTAR

EmoSTAR veritabanı televizyon programlarından, filmlerden ve internetten elde edilen İngilizce ve Türkçe ses kayıtlarıyla bizim tarafımızdan oluşturulmuş yeni bir veritabanıdır. EmoSTAR içerisindeki konuşma verileri doğal ve yapay duygu özellikleri taşımaktadır. Nötr veriler ise haber kanallarından elde edilen veriler olduğundan doğal kabul edilebilirler. Kızgın veriler genellikle filmlerden alınmıştır ve yapay olarak kabul edilebilirler. Mutlu veriler ise Oscar, Golden Globe, MTV ve American Music Awards gibi ödül törenlerinde ödül alan sanatçıların yaptığı konuşmalardan derlendiği için doğal duygu içermektedir. Üzgün veriler ise internetteki videolardan ve filmlerden alınmış olup çoğunluğu doğal bir kısmı da yapay duygu içermektedir. Bu haliyle EmoSTAR doğala oldukça yakın özelliklere sahiptir. Kayıtlar 16.000 Hz'de örneklendirilmiş, 16bit mono formatta MS Wave dosyalarından oluşmaktadır. Konuşma kayıtları için 16.000 Hz'in üstünde örnekleme frekansları kullanmanın ne konuşma

tanıma ne de duygu tanıma açısından bir avantajı yoktur. EmoSTAR'daki kayıtlar EmoDB'ye göre daha uzun cümlelerden oluşmaktadır ve her kayıt farklı bir cümledir. Bu haliyle EmoSTAR EmoDB'den daha heterojen bir yapıya sahiptir. Kayıt uzunlukları 2,2 ile 14,5 saniye arasında değişmektedir. EmoSTAR'daki veri sayısı Çizelge 4.1'de gösterilmektedir.

Çizelge 4. 1 EmoSTAR kayıtlarının duygulara göre dağılımı (İ: İngilizce, T: Türkçe)

	Kızgın	Nötr	Mutlu	Üzgün
Erkek	33 İ 30 T	35 İ 34 T	45 İ	12 İ
Kadın	40 İ	37 İ 20 T	37 İ	51 İ 19 T
Toplam=393	103	126	82	82

4.2 Berlin Duygu Veritabanı (Berlin Emotional Database, EmoDB)

Berlin Duygu Veritabanı Berlin Teknik Üniversitesi'nde Felix Burkhardt vd. tarafından geliştirilmiş yapay duygu içeren bir veritabanıdır [47]. Veritabanı yankısız bir odada 5 erkek 5 kadın olmak üzere 10 profesyonel aktör tarafından 7 farklı duygu için seslendirilen 10 farklı cümlelerin ses kayıtlarından oluşmaktadır. Toplam kayıt sayısı 535'tir, kayıtların uzunlukları 1,2 ile 9 saniye arasında değişmektedir. Ses kayıtları 16.000 Hz'de örneklendirilmiş, 16bit mono formatta MS Wav dosyalarından oluşmaktadır ve duygulara göre dağılımı Çizelge 4.2'de verilmektedir.

Çizelge 4. 2 EmoDB kayıtlarının duygulara göre dağılımı

	Kızgın	Nötr	Mutlu	Üzgün	Sıkkin	Tiksinti	Korku
Erkek	60	39	27	25	35	11	36
Kadın	67	40	44	37	46	35	33
Toplam=535	127	79	71	62	81	46	69

EmoDB'de duyguların kategorik sınıflandırılması kullanılmıştır ve duygular yapay (acted/simulated) olarak üretilmişlerdir. Ses kayıt işlemleri için 40 aktör seçilmiş ve bunların kayıtları 3 uzman dinleyici tarafından duygu türlerine göre değerlendirilerek aralarından duygu seslendirmede en başarılı olan 10 aktör seçilmiştir. Aktörler stüdyoda kendilerine sunulan cümleyi seslendirmeden önce seçilen duyguya göre motive edilmişlerdir. Üzgün duygu için bir yakınlarını kaybetmiş, mutlu duygu için büyük bir para kazanmış gibi motivasyonlar uygulanabilmektedir.

EmoDB için genel olarak bir değerlendirme yapıldığı zaman bu tür araştırmalarda kullanılmak üzere oldukça iyi tasarlanmış bir veritabanıdır. Konuşmacı sayısının diğer veritabanlarına göre nispeten fazla olması ve konuşmacıların yarısının kadın yarısının erkek olması ve temel 7 duyguyu da kapsamı EmoDB'nin üstün yönlerindendir. Literatürdeki bazı veritabanları sadece tek bir kişi tarafından seslendirilmiştir. Bazı veritabanlarında ise duygu sayısı azdır. Bunların yanı sıra EmoDB veritabanındaki kayıtlar dikkatle incelendiği takdirde kendilerine atfedilen duyguyu da oldukça iyi ifade ettikleri görülmektedir. Ancak, EmoDB'nin zayıf yönleri de bulunmaktadır. Öncelikle duygu ifadeleri oldukça homojen gözükmemektedir. Örnek olarak üzgün cümleler incelendiği takdirde hepsinin farklı konuşmacılar tarafından seslendirilmiş olmasına rağmen söyleniş tarzları birbirlerine oldukça benzemektedir. Üzgün cümlelerin tamamı kısık bir ses tonuyla duraklayarak söylenmiştir. Bu cümlelerdeki temel frekans incelendiği takdirde hepsinde düşük bir değerde azalan bir çizgi takip ettiği görülmektedir. Bu durum gerçek hayatta rastlanan durumları tam olarak temsil etmemektedir. EmoDB'nin gerçek hayattaki duygu ifade tarzlarını ne kadar yansıttığı incelenmesi gereken bir konudur.

4.3 OPENSIMILE

Duygu tanıma çalışmalarının en önemli adımlarından birisi de hiç şüphesiz ki öznelilik çıkarma adıımıdır. Öznelilik çıkarma için kullanılan fonksiyonlar oldukça karmaşık olduğundan bunların herkes tarafından ayrı ayrı üretilmesi oldukça zahmetli olmaktadır. Bu amaçla yapılan bazı çalışmalar ücretsiz olarak araştırmacıların kullanımına sunulmuş ve ortak bir platform oluşturulması ve yapılan çalışmaların birbirini destekleyerek ilerlemesi amaçlanmıştır. Konuşma tanıma uygulamaları için kullanılan uygulamalar arasında Open Speech and Music Interpretation by Large Space Extraction (OpenSMILE) [48], Hidden Markov Toolkit (HTK) [49], PRAAT [50], Matlab Auditory Toolbox [51], Sphinx [52] sayılabilir.

OpenSMILE konuşma tanıma ve sinyal işleme için geliştirilmiş bir uygulamadır ve elde edilen öznelilikler konuşma tanımada olduğu gibi duygu tanımada da kullanılabilir. OpenSmile öncelikli olarak ses verileri için geliştirilmiş olsa da kullanılan öznelilikler fizyolojik ve görsel sinyallerde ve diğer fiziksel alıcılardan elde

edilen sinyallerin işlenmesinde de kullanılabilir. OpenSMILE C++ ile yazılmıştır ve Linux, Windows ve MacOS platformlarında çalışabilir. Diğer uygulamalarla uyumluluk sağlamak açısından openSMILE PCM Wave, CSV, ARFF, HTK parametre dosyaları ve ikili matris dosya formatını desteklemektedir. PortAudio kütüphanesi kullanılarak platform bağımsız gerçek zamanlı ses kayıt ve öznitelik çıkarma işlemleri yapılabilir. Açık kaynak kodlu gnuplot yazılımı ile elde edilen özniteliklerle ilgili her türlü grafik gösterimler elde edilebilir.

OpenSMILE Linux, Unix ve Windows sürümleri ücretsiz olarak indirilebilir. Bu çalışmada Windows sürümü kullanılmıştır. IDE klasörü içindeki kaynak kodu Visual Studio 2005 veya 2008 ile derlenebileceği gibi hazır çalıştırılabilir dosya kullanılarak öznitelik çıkarma işlemi komut satırından yapılabilir. Windows sürümünde Smilextract.exe çalıştırılabilir dosyasını komut satırından kullanabilmek için Windows cmd.exe komut satırı dosyası Smilextract.exe dosyasının olduğu klasöre kopyalanmalıdır. Cmd.exe çalıştırıldığı takdirde gelen komut satırı ekranına

Smilextract -h

yazılması halinde openSMILE kullanımı ile ilgili parametreler ekrana gelmektedir. openSMILE ile hangi özniteliklerin çıkarılacağı konfigürasyon dosyaları (.conf uzantılı) ile belirlenir. Hali hazırda mevcut pek çok konfigürasyon dosyası araştırmacıların kullanımına sunulmuştur. Örnek olarak komut satırına aşağıdaki satır girilebilir.

SMILExtract -C config/demo/demo1_energy.conf -I wav_samples/speech01.wav -O speech01.energy.csv

Bu komutun anlamı wav_samples klasöründe bulunan speech01.wav ses dosyası için config/demo klasöründeki demo1_energy.conf dosyasında bulunan özniteliklerin çıkartılması ve bu özniteliklere ait değerlerin speech01.energy.csv dosyasına yazılmasıdır.

İkinci bir ses dosyası için işlem yapılması halinde ikinci ses dosyasının öznitelikleri ilk dosyanın sonuna eklenecektir. Öznitelik çıkarmada kullanılacak konfigürasyon dosyaları ilgili komut satırı parametreleri kullanılarak openSMILE tarafından üretilebileceği gibi hazır konfigürasyon dosyaları da kullanılabilir. Öznitelik konfigürasyon dosyaları config klasörü içinde bulunmaktadır ve öznitelik sayıları Çizelge 4.3’de verilmiştir.

Çizelge 4. 3 OpenSMILE konfigürasyon dosyalarının öznitelik sayıları

Öznitelik Dosyası	Öznitelik Sayısı
Emobase.conf	988 (26 LLD + 26 delta)*19 fonksiyon
Emo_large.conf	6669 (57 LLD + 57 delta+ 57 delta-delta)*39 fonksiyon

MFCC, PLP öznitelikleri, prozodi (temel frekans, sesli ve sessiz harflerin oranları, sessiz kısımlara bağlı öznitelikler) ve müzik öznitelikleri için de ayrı konfigürasyon dosyaları bulunmaktadır. Bu sayede sadece MFCC, PLP ve prozodik öznitelikler kullanılarak öznitelik dosyaları elde edilebilmektedir. Bu dosyalar daha sonra sınıflandırma amacıyla kullanılarak sadece MFCC, PLP veya prozodik özniteliklerle elde edilen başarıların karşılaştırılması yapılabilmektedir. MFCC ve PLP konfigürasyon dosyaları çıktılarını, HTK parametre dosyası türünde, Prozodi konfigürasyon dosyaları çıktılarını ise CSV dosya formatında vermektedir. Çıktı dosya türünü arff olarak değiştirmek için dosyalarda bazı değişikliklerin yapılması gereklidir.

Aşağıda gösterilen komut input.wav dosyasına ait emobase.conf dosyasında belirtilen 988 özniteliği çıkartarak elde edilen değerleri output.arff dosyasına yazmaktadır. Bu komut sonunda elde edilen arff dosyasında sınıf alanında “unknown” verisi bulunmaktadır.

```
SMILExtract -C config/emobase.conf -I input.wav -O output.arff
```

İstenilen sınıf değerlerinin sınıf alanına ve örnek değerlerinin etiketine yazılabilmesi için aşağıdaki komut satırı kullanılabilir.

```
SMILExtract -C config/emobase.conf -I inputN.wav -O output.arff -instname  
inputN -classes {anger, fear, disgust} -classlabel anger
```

Bu komutla arff dosyasındaki en son öznitelik satırı

```
@attribute emotion {anger, fear, disgust}
```

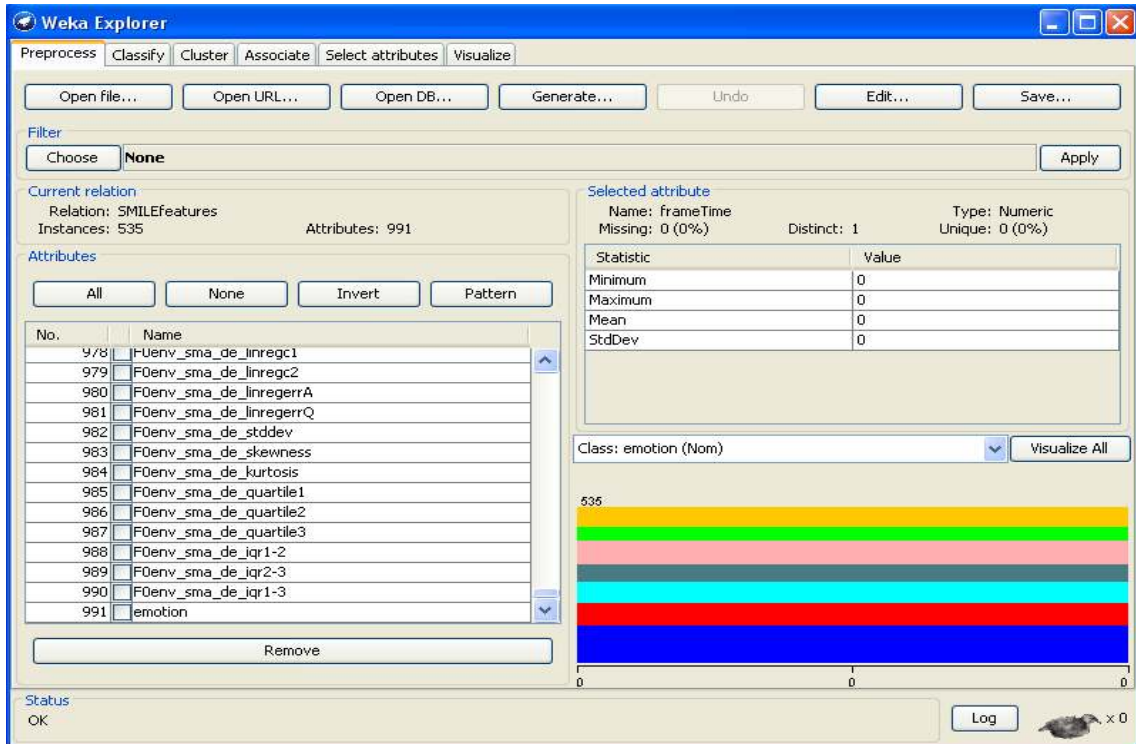
olarak yazılacak ve her öznitelik verisinin en son değeri olarak “anger” etiketi yazılacaktır. Sınıf etiketleri için {A, F, D} şeklinde kısaltmalarda kullanılabilir. İkinci bir ses dosyasını işlemek için aynı komut satırı ikinci dosya ismi ile tekrar çalıştırılmalıdır. Bu durumda elde edilen değerler arff dosyasının son satırından itibaren eklenecektir. Bu nedenle arff dosyalarını yeniden elde etmek için önceki dosyanın silinmesi

gereklidir. OpenSMILE ile çıkarılan öznitelikler ve fonksiyonlarla ilgili daha ayrıntılı bilgi için [47] ve [48]'e başvurulabilir.

4.4 Weka

OpenSMILE aracının en kullanışlı özelliklerinden biri de öznitelik çıkarımı adımından elde edilen veri dosyalarının HTK, Weka [53], [54] gibi diğer popüler sınıflandırma programlarıyla uyumlu olmasıdır. Bu çalışmada OpenSMILE ile elde edilen arff dosyaları sınıflandırma amacıyla işlenmek üzere Weka aracında kullanılacaktır. Weka, Waikato Üniversitesi tarafından geliştirilmiştir ve içerisinde veri madenciliği ve sayısız makine öğrenmesi tekniğini barındırmaktadır.

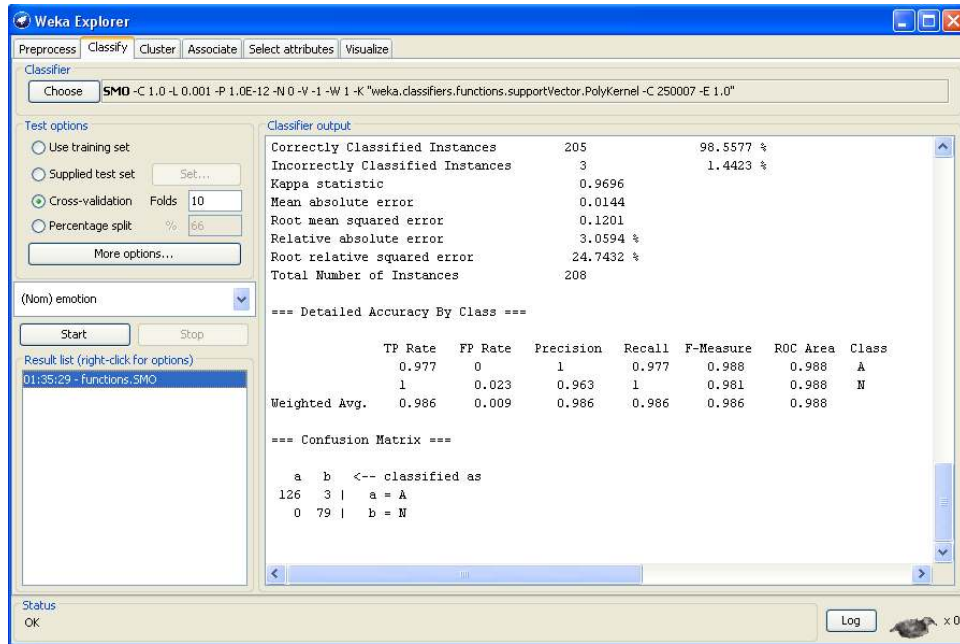
Şekil 4.1'de emo_base.conf dosyasındaki özniteliklerin openSMILE ile çıkarılmasıyla elde edilen emo_base.arff dosyasının Weka Explorer Preprocess sekmesinde OpenFile ile açılmış hali görülmektedir. Dosyada 991 öznitelik bulunmaktadır ancak name, frame_time ve emotion etiketleri duygu sınıflandırma ile ilgili öznitelik değildir ve gerçek öznitelik sayısı 988'dir. Veritabanındaki toplam 535 örnek sayısı ekranda gösterilmektedir. Örneklerin duygulara göre dağılımı hem grafik hem de tablo olarak verilmektedir.



Şekil 4. 1 Weka ön işleme penceresi

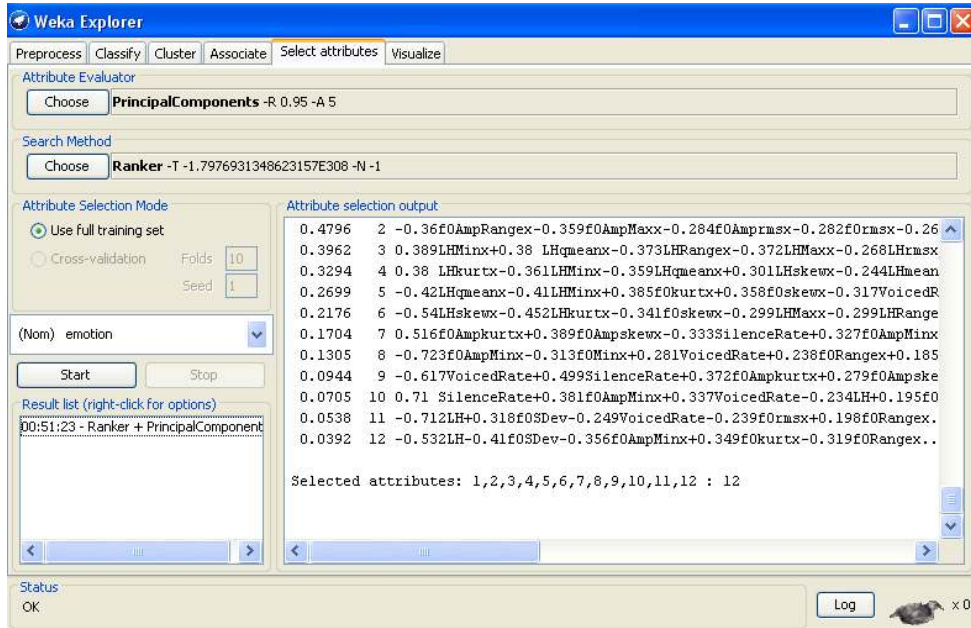
Weka’da LibSVM kullanımı için libsvm.jar dosyasının bulunduğu klasör CLASSPATH sistem değişkeninde tanımlanmış olmalıdır. Bu amaçla Denetim Masası, Sistem, Gelişmiş, Ortam Değişkenleri sekmesinden sistem değişkenleri CLASSPATH seçilir ve değer olarak *C:\Program Files\Weka-3-6\libsvm-3.18\java\libsvm.jar;C:\Program Files\Weka-3-6\libsvm-3.18\windows\libsvm.dll* değerleri girilir.

Veritabanımıza ait öznitelik dosyasını girdikten sonra yapmamız gereken bu öznitelik dosyasındaki değerlere göre sınıflandırma yapmaktır. Bu amaçla Classify sekmesinden Choose düğmesi ile istediğimiz sınıflandırıcıyı seçmekteyiz (Şekil 4.2).



Şekil 4. 2 Weka sınıflandırma penceresi

Weka’da Temel Bileşen Analizi (Principal Component Analysis) veya diğer pek çok yöntemle öznitelik seçmekte mümkündür. Bu amaçla Select Attributes sekmesinden Choose düğmesi ile PrincipalComponents seçilir ve Ranker arama metodu otomatik olarak eşleştirilir. PrincipalComponents üstüne tıklayarak bazı değişkenleri ayarlamak mümkündür. Burada değiştirilmesi gereken değişken “variancecovered” değişkenidir ve 0.95 olarak seçilmiştir. Seçilen öznitelik sayısı fazla ise bunları Preprocessing penceresinden tek tek girmek yerine Weka CLI kullanılabilir. Weka’nın Tools menüsünden seçilen Arff Viewer arff dosyaları üstünde öznitelik silme, örnek silme gibi işlemleri yapabilmektedir (Şekil 4.3).



Şekil 4. 3 Weka öz nitelik seçme penceresi

Weka ile ilgili daha ayrıntılı bilgi için Weka'yla beraber gelen el kitabına müracaat edilebilir.

4.5 Weka Komut Satırı

Bazen Weka'ya komut satırı ile bazı işlemler yaptırmak faydalı olabilir. Bu amaçla Weka komut satırı (CLI, Command Line Interpreter) (Şekil 4.4) kullanılır.

Örnek:

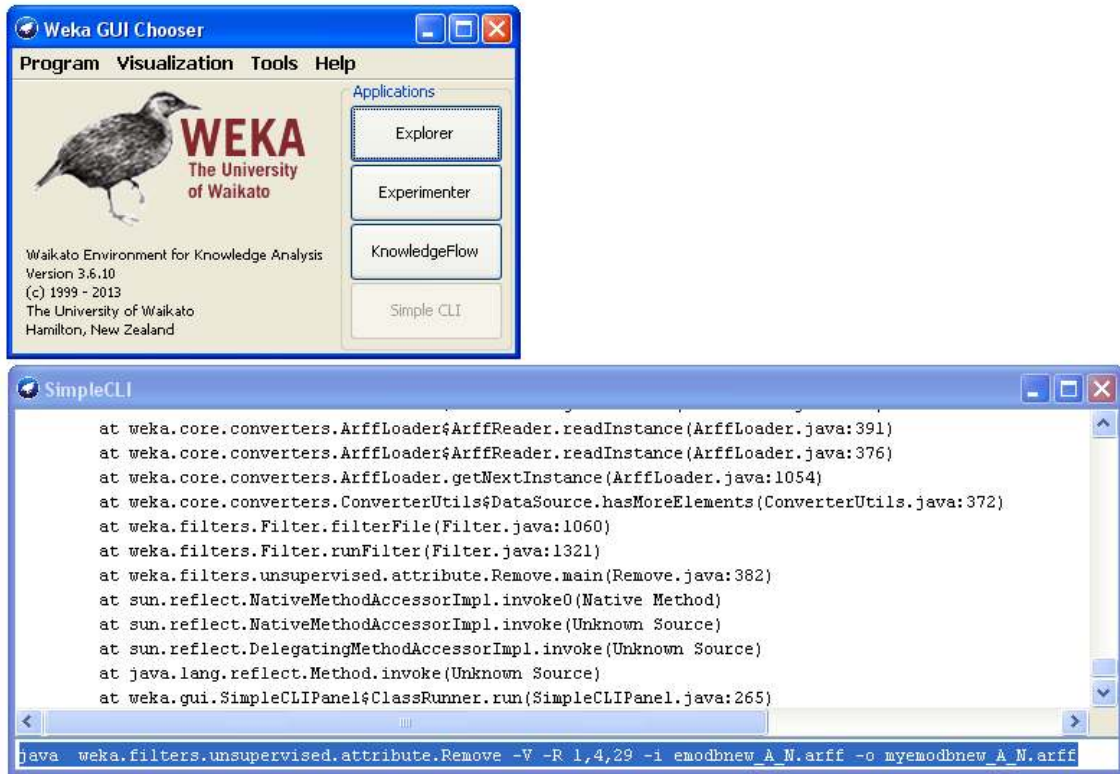
```
java weka.filters.unsupervised.attribute.Remove -V -R 1-4,29,45 -i emodb1.arff -o emodb2.arff
```

Bu komut emodb1.arff dosyasındaki 1, 2, 3, 4, 29 ve 45 nolu öz nitelikler hariç tüm öz nitelikleri siler ve yeni dosyayı emodb2.arff olarak kaydeder. Kaynak dosya Weka.jar ile aynı klasörde değilse tam yol belirtilmelidir. Komutlar CTRL+V ile CLI komut satırı metin kutusuna Kopyala-Yapıştır yapılabilir. Bu komuttan Öz nitelik Seçme amacıyla yararlanmak mümkündür. Öz nitelik seçiciler öz niteliklerin numaralarını verdiklerinden bu numaralar üstteki komuta Kopyala-Yapıştır ile girilerek sadece seçili öz niteliklerden oluşan bir öz nitelik dosyası oluşturulabilir. Böylece, bu öz nitelikleri Explorer penceresinden tek tek seçme veya komut satırına tek tek girme zorluğundan kurtulmak

mümkündür. Aynı işlem daha basit olarak öznitelik seçme işleminden sonra Save Reduced Data menü komutundan da yapılabilir.

```
java weka.classifiers.trees.J48 -t emodb1.arff -T myemodb1.arff -U -p 1 > Results.arff
```

Bu komut ile emodb1.arff dosyası J48 algoritması için eğitim seti olarak myemodb1.arff dosyası ise test seti olarak kullanılır ve sonuçlar Results.arff dosyasına yazılır.



Şekil 4. 4 Weka komut satırı

4.6 Deneyin Yapılışı

Deneyimizdeki ilk adım openSMILE aracı ile veritabanlarımızdaki ses kayıtlarının özniteliklerinin çıkarılmasıdır. Öznitelik çıkarma amacıyla openSMILE ile birlikte gelen 2 adet konfigürasyon dosyası kullanılmaktadır. Bunlar daha önce tanıtılan 988 öznitelik içeren Emobase ve 6669 öznitelik içeren Emo_large dosyalarıdır. Elde edilen özniteliklerin arff dosyası olarak çıktısı alınacak ve bu dosyalar ile Weka’da sınıflandırma yapılacaktır. Öznitelik dosyalarında “name” özniteliği string olarak bulunmaktadır ve Weka’daki sınıflandırıcılar string değerlerle çalışmadığından arff dosyalarını Preprocess düğmesi ile yükledikten sonra “name” özniteliği Remove

düğmesiyle kaldırılmalı veya Tools menüsü altındaki Weka Arff Editörünü kullanarak Delete Attributes menüsüyle “name” özniteliği arff dosyalarından silinmelidir.

Beşinci bölümde ise mevcut öznitelik dosyalarına önerilen yeni öznitelikler ilave edilecek ve performansa olan katkıları gözlemlenecektir. Sınıflandırmalar aksi belirtilmedikçe 10 kat çapraz geçerlemeyle varsayılan değerlerle gerçekleştirilecek ve sonuçlar ağırlıklandırılmış değerler olarak verilecektir.

4.6.1 Berlin Duygu Veritabanı Üzerine Çalışmalar

Berlin Duygu Veritabanı (EmoDB) üzerinde çalıştığımız ilk veritabanıdır. EmoDB üzerinden çıkarılan öznitelikler ile Weka içerisinde yer alan Naive Bayes, SMO ve Bagging sınıflandırıcılarından alınan sonuçlar Çizelge 4.4’te sunulmuştur.

Çizelge 4. 4 EmoDB için sınıflandırma başarıları

	Emobase (988)	Emo_large (6669)
NaiveBayes	53,45	70,09
SMO	87,28	86,54
Bagging	67,28	71,40

Çizelge 4.4’den görüldüğü gibi en başarılı sınıflandırma bir Destek Vektör sınıflandırma yöntemi olan SMO’dur. Bu çalışmada SMO’da polinom kernel kullanılmıştır. Öznitelik sayısının 988’den 6669’a çıkması durumunda SMO sınıflandırıcının başarısında az da olsa bir düşüş meydana gelmiştir. EmoDB’de 988 öznitelik içeren Emobase konfigürasyonu için SMO hata matrisi ve Doğru-Pozitif (DP) oranları Çizelge 4.5’te verilmiştir. Hata matrisleri hangi duygularda daha fazla hata olduğunu görmemizi sağlaması açısından faydalıdır. Çizelge 4.5’te görüldüğü gibi en fazla hata oranı Mutlu sınıfının Kızgın olarak etiketlenmesinde ortaya çıkmaktadır. Bu durum EmoDB’deki hemen hemen tüm sınıflandırmalarda geçerlidir. Çizelgelerde Kızgın, Nötr, Mutlu, Üzgün, Sıkın, Tiksinti, Korku sırasıyla A, N, H, S, B, D, F olarak kısaltılmıştır.

Çizelge 4. 5 EmoDB’de Emobase (988 öznitelik) için SMO hata matrisi

	A	N	H	S	B	D	F	DP
A	118	0	5	0	0	1	3	0,92
N	0	75	0	1	2	0	1	0,94
H	14	0	52	0	0	1	4	0,73
S	0	1	0	57	4	0	0	0,91
B	0	4	0	5	69	3	0	0,85
D	2	2	1	0	2	37	2	0,80
F	3	1	3	2	0	1	59	0,85

NaiveBayes, SMO ve Bagging sınıflandırıcılarının performansını farklı öznitelik gruplarıyla değerlendirmek amacıyla öznitelik dosyalarımızı prozodi ve MFCC tabanlı olmak üzere ikiye ayırdık. Bu sayede sınıflandırıcının sadece prozodik özniteliklerle ve sadece MFCC öznitelikleriyle elde ettiği başarı oranlarını ayrı ayrı tespit ettik. MFCC öznitelikleri prozodik özniteliklere göre daha yüksek başarı oranlarını yakalayabilmektedir. Ancak Emo_large konfigürasyonunda prozodik öznitelik sayısının MFCC öznitelik sayısına göre oldukça az olduğu gözden kaçırılmamalıdır. Çizelge 4.6’da MFCC+Prozodi-Delta tüm özniteliklerden delta ve delta-delta özniteliklerinin çıkarılması ile elde edilen sonuçlara yer verilmiştir. Parantez içindeki sayılar sırasıyla Emobase ve Emo_large için öznitelik sayılarını göstermektedir.

Çizelge 4. 6 EmoDB için farklı öznitelik gruplarıyla sınıflandırma sonuçları

EmoDB	Emobase			Emo_large		
	NB	SMO	Bag	NB	SMO	Bag
Tüm Öznitelikler (988, 6669)	53,45	87,28	67,28	70,09	86,54	71,40
MFCC+Prozodi-Delta (494, 2223)	68,22	84,48	65,23	65,42	86,72	68,59
MFCC (456, 6084)	66,54	80,74	63,55	69,34	87,85	72,14
Prozodi (532, 585)	42,80	83,55	64,48	59,25	76,07	68,03

Sadece 532 adet prozodik özniteliğin kullanıldığı Emobase konfigürasyonu için SMO sınıflandırıcı hata matrisi ve Doğru-Pozitif oranları (DP) Çizelge 4. 7’de gösterilmektedir. En yüksek hata oranı mutlu ve kızgın örneklerde karşımıza çıkmaktadır. Diğer örneklerdeki hatalı sınıflandırmalar farklı duygulara dağılmıştır.

Çizelge 4. 7 EmoDB’de 532 prozodik öznitelikli Emobase için SMO hata matrisi

	A	N	H	S	B	D	F	DP
A	111	0	11	0	1	2	2	0,87
N	0	74	0	0	5	0	0	0,93
H	15	2	48	1	0	1	4	0,67
S	0	3	0	56	3	0	0	0,90
B	0	8	0	4	66	3	0	0,81
D	3	4	0	2	1	35	1	0,76
F	7	1	2	1	0	1	57	0,82

EmoDB’de sadece 6084 adet MFCC özniteliğinin kullanıldığı Emo_large konfigürasyonu için SMO sınıflandırıcı hata matrisi ve Doğru-Pozitif (DP) oranları Çizelge 4. 8’de gösterilmektedir.

Çizelge 4. 8 EmoDB’de 6084 MFCC öznitelikli Emo_large için SMO hata matrisi

	A	N	H	S	B	D	F	DP
A	119	0	5	0	0	0	3	0.93
N	0	73	0	2	3	0	1	0.92
H	14	0	53	0	1	0	3	0.74
S	0	1	0	57	4	0	0	0.91
B	0	2	0	5	74	0	0	0.91
D	0	2	1	1	2	37	3	0.80
F	3	4	4	0	0	1	57	0.82

Duygu tanıma çalışmalarında araştırılan duygu sayısı başarı oranını önemli ölçüde etkilemektedir. Duygu sayısı azaldıkça sınıflandırıcıların işleri daha kolaylaşmakta ve daha yüksek başarı ile sınıflandırma yapabilmektedirler. Bu amaçla öznitelik dosyaları Kızgın-Nötr-Mutlu-Üzgün, Kızgın-Nötr, Mutlu-Nötr-Üzgün olmak üzere 3 değişik sınıflandırmaya tabi tutulmuşlardır. Çizelgelerden (4.9 ve 4.10) rahatça görülebileceği gibi duygu sayısı azaldıkça sınıflandırıcıların başarı oranı da artmaktadır. Emobase konfigürasyonunda Kızgın-Nötr-Mutlu-Üzgün (A-N-H-S) için başarı oranı %92,33 iken Nötr-Mutlu-Üzgün sınıflandırıcıda bu oran %98,58, Kızgın-Nötr sınıflandırıcı da ise %99,51 olmaktadır. Benzer şekilde Emo_large konfigürasyonunda da Kızgın-Nötr-Mutlu-Üzgün (A-N-H-S) için başarı oranı %90,26 iken Nötr-Mutlu-Üzgün sınıflandırıcıda bu oran %96,70, Kızgın-Nötr sınıflandırıcı da ise %98,54 olmaktadır.

Çizelge 4. 9 EmoDB’de farklı duygu grupları için Emobase sınıflandırma sonuçları

Emobase	A-N-H-S	N-H-S	A-N
NB	80,82	91,98	98,05
SMO	92,33	98,58	99,51
Bag	81,41	86,79	95,14

Çizelge 4. 10 EmoDB’de farklı duygu grupları için Emo_large sınıflandırma sonuçları

Emo_large	A-N-H-S	N-H-S	A-N
NB	84,66	94,33	98,05
SMO	90,26	96,70	98,54
Bag	80,53	90,09	96,60

Emobase konfigürasyonunda Kızgın-Nötr-Mutlu-Üzgün için SMO hata matrisi Çizelge 4.11’de gösterilmektedir. SMO sınıflandırıcı 4 sınıflı uygulamada en fazla hatayı mutlu ve kızgın verileri ayırırken yapmaktadır.

Çizelge 4. 11 Emobase Kızgın-Nötr-Mutlu-Üzgün için SMO hata matrisi

	A	N	H	S	DP
A	118	1	8	0	0,92
N	0	77	1	1	0,97
H	13	0	57	1	0,80
S	0	1	0	61	0,98

Emo_large konfigürasyonunda Kızgın-Nötr-Mutlu ve Kızgın-Nötr için SMO hata matrisi Çizelge 4.12 ve Çizelge 4.13’te gösterilmektedir. Sınıf sayısı azaldıkça başarı oranları artmaktadır.

Çizelge 4. 12 Emobase Nötr-Mutlu-Üzgün için SMO hata matrisi

	N	H	S	DP
N	78	0	1	0,98
H	0	70	1	0,98
S	1	0	61	0,98

Çizelge 4. 13 Emobase Kızgın-Nötr için SMO hata matrisi

	A	N	DP
A	126	1	0,99
N	0	79	1

4.6.2 EmoSTAR Üzerindeki Çalışmalar

Oluşturduğumuz EmoSTAR veritabanı üzerinde de benzer çalışmalar yapılmış ve kızgın-nötr-mutlu-üzgün için Çizelge 4. 14'teki sonuçlar elde edilmiştir. SMO EmoSTAR veritabanında da en başarılı sınıflandırıcı olmuştur. Naive Bayes ve Bagging birbirlerine yakın performanslar göstermektedirler.

Çizelge 4. 14 EmoSTAR'da Kızgın-Nötr-Mutlu-Üzgün için sınıflandırma sonuçları

	Emobase (988)	Emo_large (6669)
NB	83,71	83,96
SMO	96,18	96,94
Bag	84,47	87,78

EmoSTAR'da farklı öznitelik gruplarıyla elde edilen sonuçlar Çizelge 4.15'de gösterilmektedir. En yüksek başarı oranı Emo_large konfigürasyonu ile elde edilmiştir. Bu konfigürasyonda delta ve delta-delta özniteliklerinin çıkarılması ile öznitelik sayısı 2223'e inmesine rağmen başarı oranı değişmemiştir.

Çizelge 4. 15 EmoSTAR için farklı öznitelik gruplarıyla sınıflandırma sonuçları

EmoSTAR	Emobase			Emo_large		
	NB	SMO	Bag	NB	SMO	Bag
Tüm Öznitelikler (988, 6669)	83,71	96,18	84,47	83,96	96,94	87,78
MFCC+Prozodi-Delta (494, 2223)	85,49	95,67	86,25	83,46	96,94	88,29
MFCC (456, 6084)	79,64	93,89	82,18	82,95	96,69	87,27
Prozodi (532, 585)	76,33	93,38	84,73	70,99	82,44	78,37

Parantez içindeki sayılar sırasıyla Emobase ve Emo_large konfigürasyon dosyalarındaki öznitelik sayılarını göstermektedir.

4.6.3 EmoSTAR + EmoDB Üzerine Çalışmalar

Duygu tanıma üzerine yapılan çalışmalar genel olarak belli bir veritabanı üstünde yapılmaktadır. Veritabanlarının birbirine karıştırıldığı çalışmaların sayısı oldukça azdır. Ender olarak bazı çalışmalarda, veritabanlarından biri eğitim diğeri de test seti olarak kullanılmaktadır. Bu bölümdeki çalışmada, EmoSTAR ve EmoDB ayrı ayrı eğitim seti ve test seti olarak kullanılmıştır. Bu amaçla öncelikle EmoDB üstündeki kızgın, nötr, mutlu ve üzgün verilerden oluşan 339 verinin sonuçları Çizelge 4.16'da verilmiştir.

Çizelge 4. 16 EmoDB’de Kızgın-Nötr-Mutlu-Üzgün için sınıflandırma sonuçları

EmoDB	Emobase (988)	Emo_large (6669)
NB	80,82	84,66
SMO	92,33	90,26
Bag	81,41	80,53

EmoSTAR ve EmoDB’nin kızgın, nötr, mutlu ve üzgün verilerinin karıştırılmasıyla elde edilen sonuçlar Çizelge 4.17’de verilmiştir. Toplam veri sayısı EmoSTAR 393 ve EmoDB 339 olmak üzere 732’dir. Verilerin karıştırılması başarı oranlarını EmoDB’ye göre arttırmış ancak EmoSTAR’a göre düşürmüştür.

Çizelge 4. 17 EmoSTAR ve EmoDB’nin karıştırılmasıyla elde edilen sonuçlar

EmoSTAR+EmoDB	Emobase (988)	Emo_large (6669)
NB	63,52	79,64
SMO	90,71	90,98
Bag	81,69	82,92

4.6.4 EmoSTAR ve EmoDB Arasındaki Çapraz Sınıflandırmalar

Veritabanları arasında çapraz bir sınıflandırma gerçekleştirmek amacıyla elimizdeki veritabanlarından birini eğitim diğerini de test amacıyla kullandık. EmoSTAR mevcut haliyle sadece 4 duyguyu içerdiğinden çapraz sorgulama amacıyla EmoDB’nin de sadece 4 duyguyu içeren örnekleri kullanılmıştır. İlk aşamada EmoSTAR eğitim seti olarak kullanılmış olup sonuçlar Çizelge 4.18’de gösterilmektedir.

Çizelge 4. 18 EmoSTAR eğitim EmoDB test seti olmak üzere sınıflandırma sonuçları

EmoSTAR Eğitim EmoDB Test	Emobase (988)	Emo_large (6669)
NB	41,88	45,13
SMO	47,19	62,83
Bag	40,41	34,51

İkinci aşamada EmoDB eğitim seti EmoSTAR test seti olarak kullanılmıştır. Sonuçlar Çizelge 4.19’da gösterilmektedir.

Çizelge 4. 19 EmoDB eğitim EmoSTAR test seti olmak üzere sınıflandırma sonuçları

EmoDB Eğitim EmoSTAR Test	Emobase (988)	Emo_large (6669)
NB	40,96	42,74
SMO	44,27	44,02
Bag	35,36	32,82

Çizelgelerden de anlaşılacağı gibi veri setlerinin ayrı ayrı veya karıştırılarak kullanılmasında çok yüksek başarı oranları elde edilebilmesine rağmen çapraz sorgulama sonuçlarında %50'lere varan dramatik düşüşler meydana gelmektedir. Bu da duygu veritabanlarının gerçekliklerinin, doğallıklarının ve sınıflandırıcıların aşırı öğrenmelerinin ciddi sorgulamalara tabi tutulması gerektiğini göstermektedir. Daha ayrıntılı bir inceleme yapmak amacıyla farklı duygu grupları için çapraz test uygulanmış ve sonuçlar Çizelge 4.20 ve 21'de verilmiştir.

Çizelge 4. 20 EmoSTAR ve EmoDB ayrı ayrı eğitim ve test seti olmak üzere Emobase sınıflandırma sonuçları

	EmoSTAR Eğitim EmoDB Test Emobase (988)			EmoDB Eğitim EmoSTAR Test Emobase (988)		
	NB	SMO	Bag	NB	SMO	Bag
Kızgın-Nötr	55,82	74,75	74,27	51,09	65,50	55,89
Mutlu-Nötr-Üzgün	51,88	51,41	62,73	39,65	54,13	53,44
Kızgın-Nötr- Mutlu-Üzgün	41,88	47,19	40,41	40,96	44,27	35,36

Çizelge 4. 21 EmoSTAR ve EmoDB ayrı ayrı eğitim ve test seti olmak üzere Emo_large sınıflandırma sonuçları

	EmoSTAR Eğitim EmoDB Test Emo_large (6669)			EmoDB Eğitim EmoSTAR Test Emo_large (6669)		
	NB	SMO	Bag	NB	SMO	Bag
Kızgın-Nötr	71,84	91,26	75,24	53,27	72,05	49,78
Mutlu-Nötr-Üzgün	45,75	52,83	53,77	48,96	52,75	33,44
Kızgın-Nötr- Mutlu-Üzgün	45,13	62,83	34,51	42,74	44,02	32,82

Emobase konfigürasyonunun SMO ile Mutlu-Nötr-Üzgün ve Emo_large konfigürasyonunun Naive Bayes ile Mutlu-Nötr-Üzgün sınıflandırması dışındaki tüm sınıflandırmalarda EmoSTAR'ın Eğitim seti olarak kullanılması EmoDB'nin Eğitim seti olarak kullanılmasına göre daha başarılı sonuçlar vermiştir.

4.7 Öznitelik Seçme

Sınıflandırma problemlerinin bilinen sorunlarından biri de aşırı sayıda ve gereksiz öznitelik kullanmanın başarı oranını olumsuz etkilemesidir. Bu çalışmada buraya kadar yapılan incelemelerde öznitelik seçimi yapılmamıştır. Öznitelik sayısının 988’den 6669’a çıkması Naive Bayes ve Bagging’de başarı oranına olumlu etki yaparken, SMO sınıflandırıcıda olumsuz etki yapmıştır. Bu sorunu çözmek amacıyla öznitelik seçme yoluna gidilmiştir. Bu bölümde Weka’da mevcut öznitelik seçme algoritmalarını veritabanlarımıza uygulayıp başarı oranlarını gözlemledik ve hangi özniteliklerin sınıflandırmada daha etkin rol oynadığını tespit ettik. Çizelge 4.22’de EmoDB için CfsSubSet Evaluator ve Linear Forward Selection, ChiSquared Attribute Evaluator ve Ranker, Information Gain Attribute Evaluator ve Ranker öznitelik seçme algoritmalarının 7 duygu için uygulanmasının sonuçları gösterilmektedir. ChiSqu+Rank ile Emobase konfigürasyonunda ise 988 öznitelikten 715 adedi seçilerek %88,03 başarı oranı elde edilmiştir. Bu oran tüm özniteliklerle elde edilen %87,28 başarı oranından daha yüksek olduğu gibi Emo_large konfigürasyonunda 6669 öznitelikle elde edilen %86,54 başarı oranının da üzerindedir.

Çizelge 4. 22 EmoDB’de 7 duygu için öznitelik seçme sonuçları

EmoDB	Emobase (988) ¹			Emo_large (6669) ¹		
	NB (53,45) ²	SMO (87,28) ²	Bag (67,28) ²	NB (70,09) ²	SMO (86,54) ²	Bag (71,40) ²
CfsSub+LFS (136,141) ³	78,50	85,79	70,65	78,50	87,28	75,70
InfoG+Rank (508,4407) ³	75,88	87,10	69,34	71,77	87,28	74,01
ChiSqu+Rank (715,5416) ³	75,14	88,03	68,59	71,21	87,85	72,71

Çizelge 4,23’te EmoSTAR için CfsSubSet Evaluator ve Linear Forward Selection, ChiSquared Attribute Evaluator ve Ranker, Information Gain Attribute Evaluator ve Ranker öznitelik seçme algoritmalarının sonuçları gösterilmektedir. En yüksek başarı oranı InfoG+Rank ve ChiSqu+Rank öznitelik seçiciyle Emo_large konfigürasyonunda

¹ Tüm öznitelik sayısı

² Özniteliklerin tamamı kullanılarak elde edilen sonuçlar

³ Sırasıyla Emobase ve Emo_large için seçilen öznitelik sayıları

elde edilen %97,45 oranıdır ve bu oran bu veritabanında 4 duygu için ulaşılan en yüksek başarı oranıdır.

Çizelge 4.23 EmoSTAR 4 duygu için öznitelik seçme sonuçları

EmoSTAR	Emobase (988) ¹			Emo_large (6669) ¹		
	NB (83,71) ²	SMO (96,18) ²	Bag (84,47) ²	NB (83,96) ²	SMO (96,94) ²	Bag (87,78) ²
CfsSub+LFS (111,138) ³	89,82	96,43	85,75	90,83	96,43	90,33
InfoG+Rank (837,5637) ³	83,96	96,69	85,24	83,20	97,45	88,80
ChiSqu+Rank (837,5637) ³	83,96	96,69	85,24	82,69	97,45	88,80

Genel olarak MFCC tabanlı öznitelikler prozodik özniteliklerden daha üst sıralarda seçilmektedir. Bu durum [15]'deki bulgularla uyum içindedir. CfsSubSet Evaluator öznitelik seçici diğerlerine kıyasla çok daha az öznitelikle oldukça yüksek başarı oranları elde edebilmektedir.

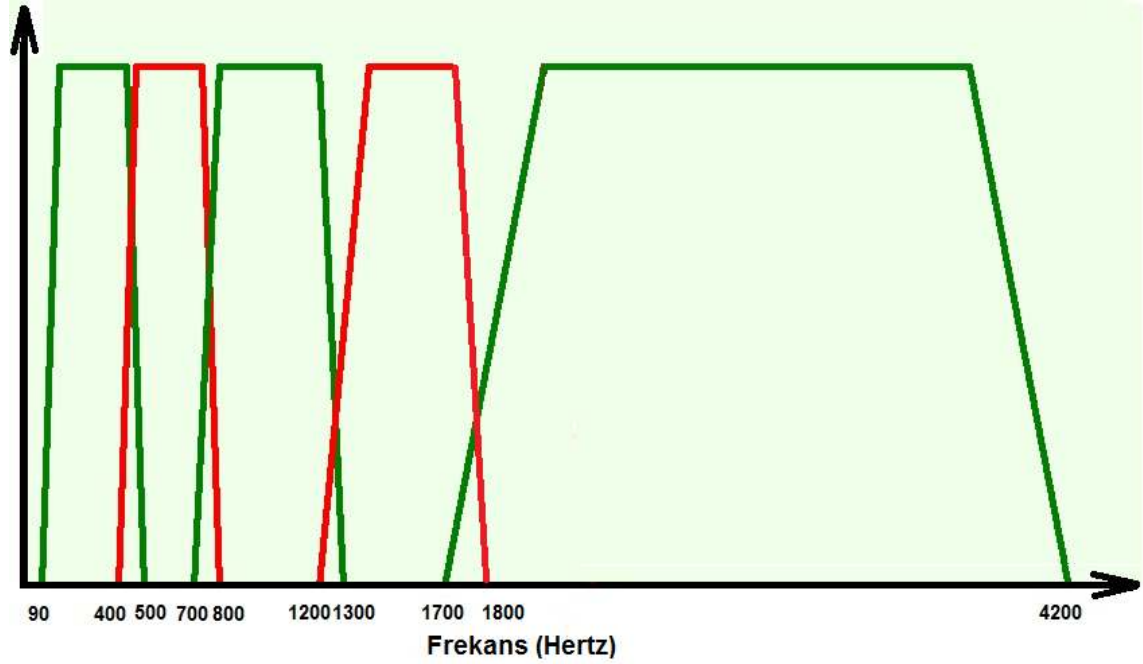
ÖNERİLEN FORMANT TABANLI FİLTRE BANKALARI

Bu çalışmada konuşma sinyalindeki sesli harflerin formantlarına dayalı yeni bir filtre bankası, yüksek frekans bölgesindeki toplam enerjinin düşük frekans bölgesindeki toplam enerjiye oranı, konuşma içindeki duraklamaların oranı yeni öznitelikler (Spectro-Temporal Auditory Formant Filter Bank Energy Ratios - STAFFER) olarak sunulmuştur. Bunun yanı sıra mevcut öznitelikler içinde bulunan fakat logaritmik olarak hesaplanan temel frekans şiddet değeri de yerleşik geleneklere cesurca bir meydan okumayla mutlak değer olarak hesaplanarak yeni bir öznitelik olarak kullanılmıştır. Bu sayede logaritma ile bastırılan şiddet değerleri arasındaki farkın daha belirgin olması ve sınıflandırıcıların bu farkları daha kolay görmesi amaçlanmaktadır. Her ne kadar insan kulağı logaritmik bir frekans analizörü gibi çalışsa da kendimizi doğayı olduğu gibi taklit etmek zorunda hissetmemeliyiz. Uçakların kanat çırparak uçamayışı gibi. Temel frekansın sinyal boyunca izlediği ortalama değerde bu özniteliklere eklenecektir. Maksimum, minimum, aralık, aritmetik ortalama, kuadratik ortalama, geometrik ortalama, standart sapma, çarpıklık, basıklık, delta ve delta-delta fonksiyonlarının uygulanmasıyla 315 kısa zamanlı formant filtre bankası özniteliği elde edilmektedir.

5.1 Sesli Harflerin Formantlarına Dayalı Filtre Bankaları ve Yüksek Frekans-Düşük Frekans Bölgesi Enerji Oranı

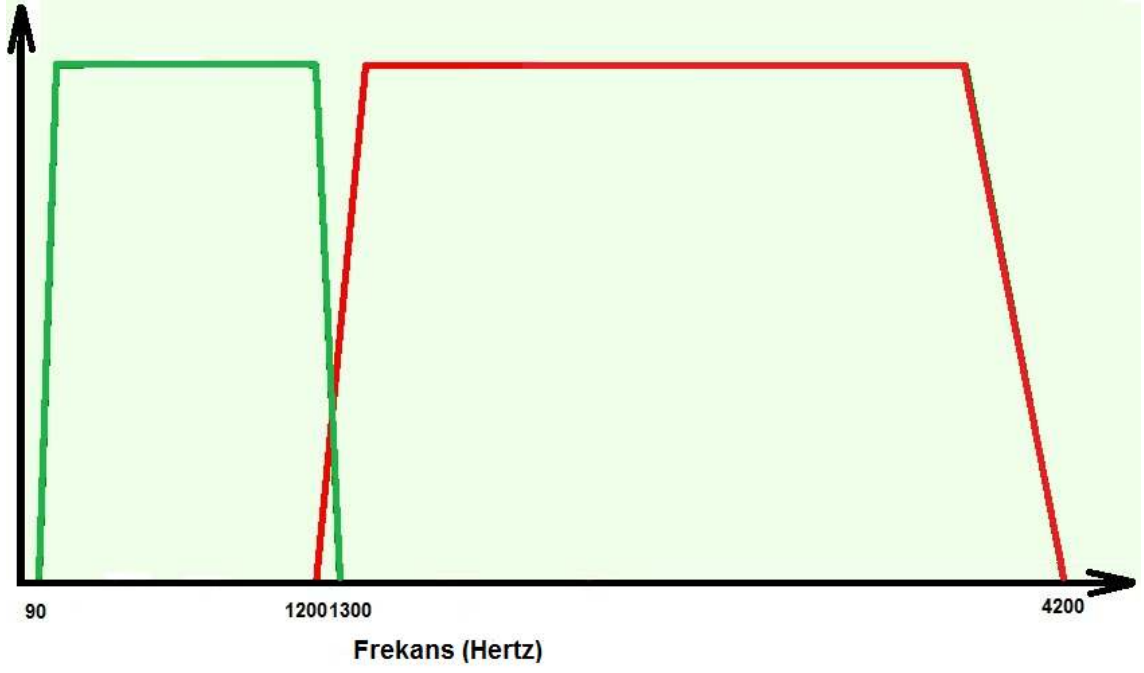
Konuşma sinyalindeki sesli harflerin belli frekanslarda formantlara sahip oldukları bilinmektedir. Sesli harfler bu formantlar sayesinde birbirlerinden tamamen ayırt edilebilmektedirler. Sessiz harflerde ise sadece formantları tanımak yeterli olmamakta sesin oluşum şeklinin bilinmesi gerekmektedir. Bugüne kadar Mel Ölçeği, Bark Ölçeği

gibi deęişik filtre bankaları önerilmiş olmakla birlikte sesli harflerin formantlarına dayalı bir sistem henüz önerilmemiştir. Duygu tanımada prozodik özniteliklerin büyük etkisi ve bu özniteliklerin sesli harflere dayalı olduęu göz önüne alındığında bu tür bir filtre tasarımının önemli bir gelişme sağlaması beklenebilir. Şekil 5.1 ve 5.2’de bu çalışmada kullanılan filtre bankaları gösterilmektedir.

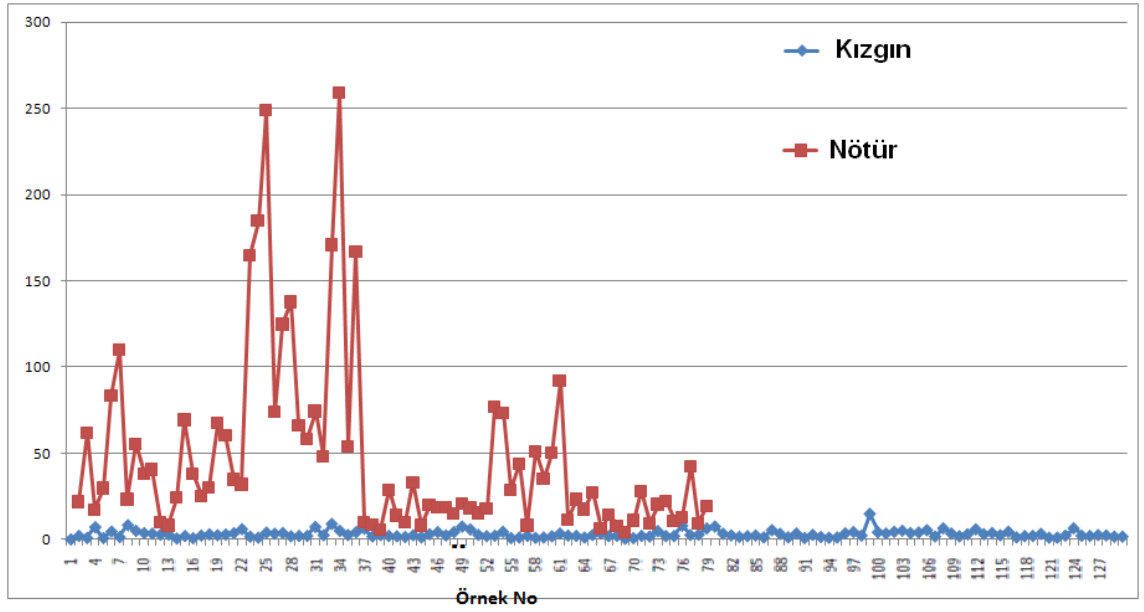


Şekil 5. 1 Sesli harflerin formantlarına dayalı filtre bankaları

Bilindięi gibi insan kulaęı lineer ve logaritmik frekans bölgelerine dayalı bir frekans analizörü gibi çalışmaktadır. Lineer bölge sınırı genel olarak 1.000 Hz olarak kabul edilmekle birlikte kesin bir sınır belirli deęildir. Bu iki bölgenin bir dięer özellięi de kalın sesli harflerle ince sesli harflerin ayrıldığı bölge olmasıdır. Kalın sesli harflerin (a, ı, o, u) formantları lineer bölgede olmasına raęmen ince sesli harflerin (e, i, ö, ü) logaritmik bölgede de formantları bulunmaktadır. Bu çalışma kapsamında yapılan incelemeler bu iki bölgenin sınırının 1.100 Hz ile 1.300 Hz arasında olduğunu göstermektedir. Bu çalışmada sınır deęer olarak 1.200 Hz alınmıştır. Yüksek frekans–düşük frekans bölgesi enerji oranı bilhassa kızgın, mutlu, nötr, üzgün ayrımında oldukça belirgin bir özniteliktir. Şekil 5.2’de yüksek frekans ve düşük frekans enerji bölgeleri gösterilmektedir. Bu bölgeler aynı zamanda insan duyma sisteminin lineer ve logaritmik bölgelerine karşılık gelmektedirler.



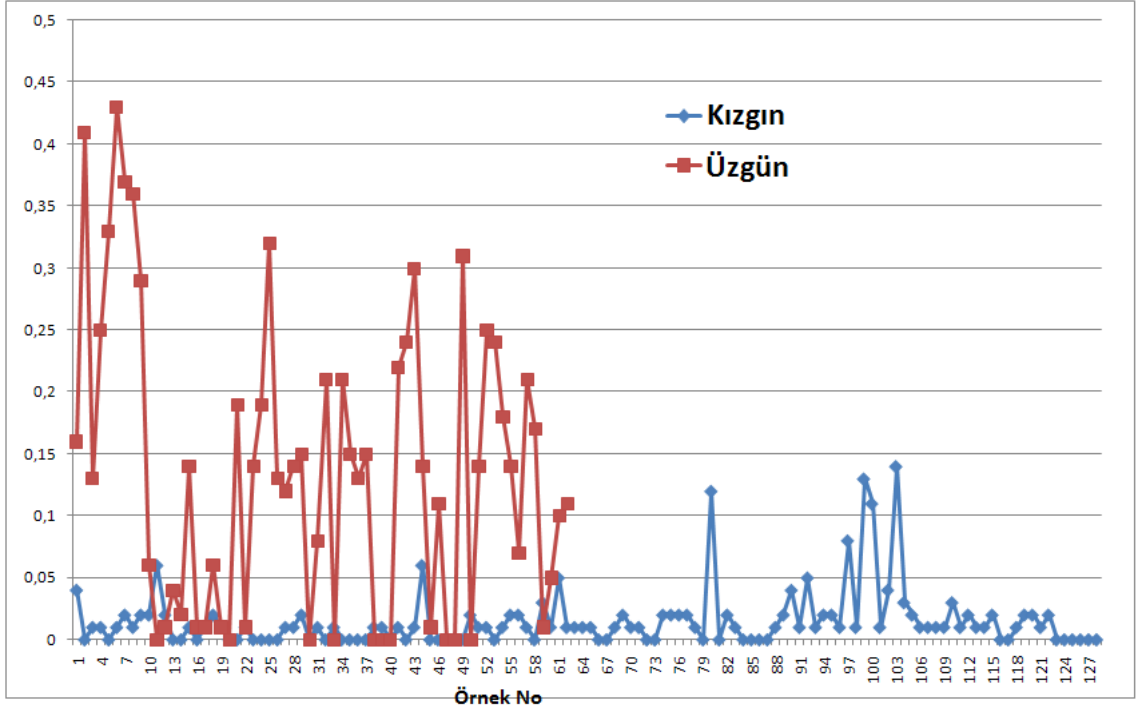
Şekil 5. 2 Kalın sesli ve ince sesli harflerin formantlarına dayalı yüksek frekans ve düşük frekans filtre bankaları



Şekil 5. 3 EmoDB'de kızgın ve nötr cümlelerdeki yüksek ve düşük frekanslardaki enerji oranları

5.2 Duraklamalar

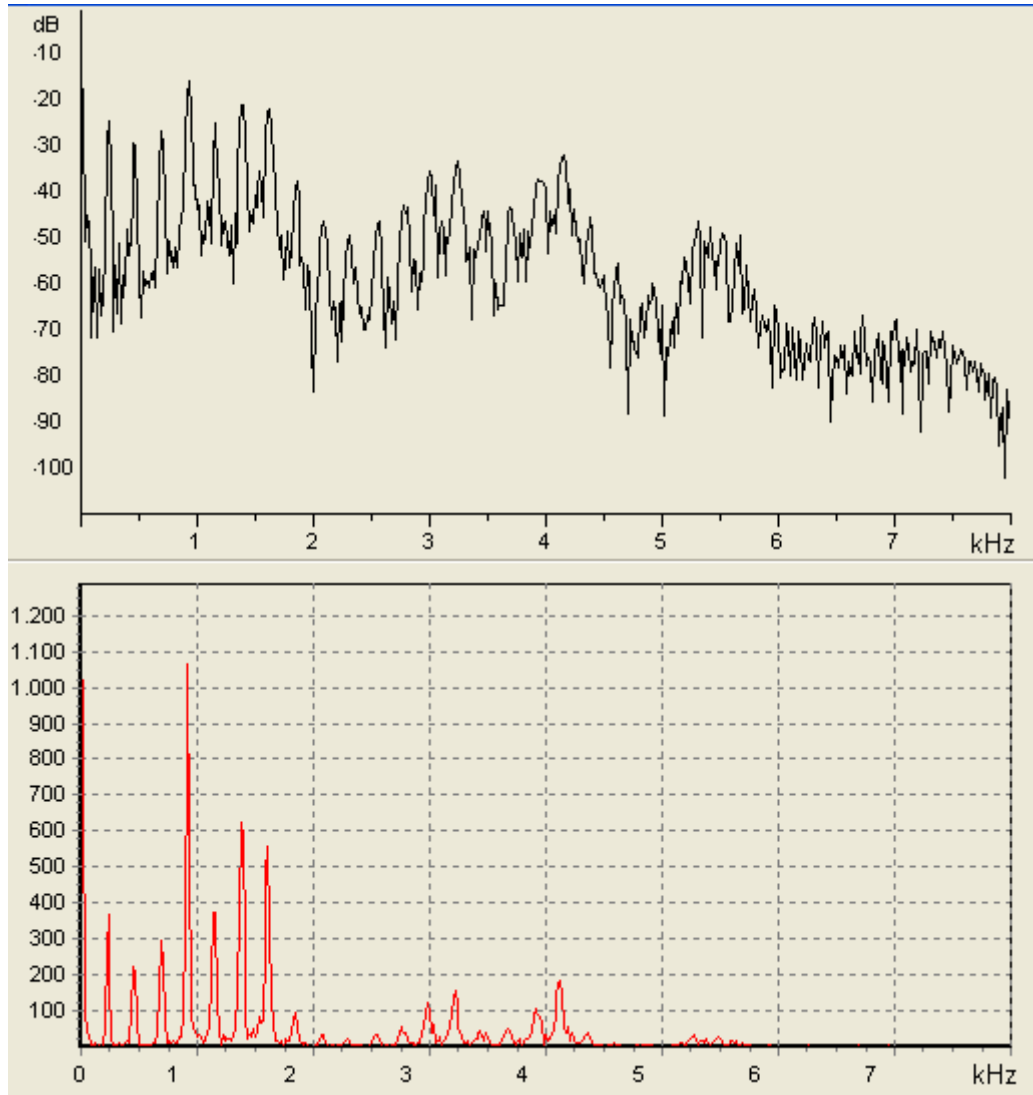
Özellikle üzgün duygularda kendini göstermektedir. Kelimeler arasındaki sessiz geçen kısımlardır. Şekil 5.4'te EmoDB'deki kızgın ve üzgün cümlelerdeki duraklamaların süresinin cümlenin toplam süresine olan oranları gösterilmektedir.



Şekil 5. 4 EmoDB’de kızgın ve üzgün cümlelerdeki duraklama oranları

5.3 Mutlak Şiddet Değerleri

Konuşma tanıma ile ilgili çalışmalarda insan kulağının logaritmik bir frekans analizörü gibi çalışmasından hareketle ses sinyallerindeki şiddet değerlerinin logaritması kullanılır. Ancak logaritmik değerler şiddet değerlerini fazlaca bastırdığından aradaki farkların kaybolmasına veya azalmasına neden olabilmektedir. Bu çalışmada bu gelenekten farklı olarak şiddet değerlerinin mutlak değerleri de kullanılmıştır. Şekil 5.5’te bir sinyale ait logaritmik (üstte) ve mutlak değer (altta) FFT sonuçları gösterilmektedir.



Şekil 5. 5 Bir sinyale ait frekans spektrumunun logaritmik (üstte) ve mutlak değerli (atta) gösterimi

5.4 Yeni Özniteliklerle Deneysel Sonuçlar

EmoDB’de 7 duygu için Emobase ve Emo_large konfigürasyonuna 315 yeni öznitelik ekleyerek elde edilen sonuçlar Çizelge 5.1’de verilmiştir. Eklenen yeni öznitelikler Emo_large-Bagging dışındaki tüm sınıflandırıcılarda daha iyi sonuç vermiştir.

Çizelge 5.1 EmoDB’de 7 duygu için yeni özniteliklerle sınıflandırma sonuçları

EmoDB Tümü	Emobase (988)	Emobase+STAFFER (988+315)	Emo_large (6669)	Emo_large+STAFFER (6669+315)
NB	53,45	70,84	70,09	70,28
SMO	87,28	88,22	86,54	86,54
Bag	67,28	69,34	71,40	69,90

EmoDB için kızgın, nötr, mutlu ve üzgün olmak üzere 4 duyguya ait Emobase ve Emo_large konfigürasyonuna 315 yeni öznitelik ekleyerek elde edilen sonuçlar Çizelge 5.2’de verilmiştir.

Çizelge 5.2 EmoDB’de 4 duygu için yeni özniteliklerle kızgın, nötr, mutlu, üzgün sınıflandırma sonuçları

EmoDB A-N-H-S	Emobase (988)	Emobase+STAFFER (988+315)	Emo_large (6669)	Emo_large+STAFFER (6669+315)
NB	80,82	84,36	84,66	84,36
SMO	92,33	91,74	90,26	91,44
Bag	81,41	80,82	80,53	81,71

EmoSTAR için kızgın, nötr, mutlu ve üzgün olmak üzere 4 duyguya ait Emobase ve Emo_large konfigürasyonuna 315 yeni öznitelik ekleyerek elde edilen sonuçlar Çizelge 5.3’de verilmiştir. Tüm sınıflandırıcılar yeni öznitelikler ile daha iyi sonuç üretmiştir.

Çizelge 5.3 EmoSTAR’da 4 duygu için yeni özniteliklerle sınıflandırma sonuçları

EmoSTAR	Emobase (988)	Emobase+STAFFER (988+315)	Emo_large (6669)	Emo_large+STAFFER (6669+315)
NB	83,71	89,82	83,96	85,24
SMO	96,18	96,69	96,94	97,45
Bagging	84,47	88,04	87,78	88,80

EmoSTAR için farklı duygu gruplarına ait sadece STAFFER öznitelikleri ile elde edilen sonuçlar Çizelge 5.4’de verilmiştir. Sadece STAFFER öznitelikleri kullanılarak kızgın-nötr sınıflandırıcı da %98,69 başarı oranı elde edilebilmiştir.

Çizelge 5.4 EmoSTAR için STAFFER öznitelikleri ile farklı duygu grupları sınıflandırma sonuçları

EmoSTAR	NB	SMO	Bag
Kızgın-Nötr	94,32	98,69	93,01
Mutlu-Nötr-Üzgün	74,13	90,00	83,10
Kızgın-Nötr-Mutlu-Üzgün	69,97	88,29	80,91

EmoDB’de farklı duygular üstünde STAFFER öznitelikleri ile elde edilen sonuçlar Çizelge 5.5’de verilmiştir. A-N-H-S ayırımında A-H ve N-H hata oranı yüksek olmasına rağmen N-H-S ayırımında hata oranı düşüktür. Tümü için nötr ve sıkın ayırımında da hata oranı yüksek çıkmaktadır.

Çizelge 5.5 EmoDB için STAFFER öznitelikleriyle farklı duygu grupları sınıflandırma sonuçları

EmoDB	NB	SMO	Bag
Kızgın-Nötr	97,57	99,02	96,11
Mutlu-Nötr-Üzgün	82,07	92,45	90,09
Kızgın-Nötr-Mutlu-Üzgün	77,28	85,25	81,41
7 Duygu	54,95	70,65	63,92

EmoSTAR’da farklı duygular üstündeki yeni özniteliklerle veritabanları arası EmoSTAR ve EmoDB’nin ayrı ayrı eğitim ve test seti olarak kullanılmasıyla elde edilen çapraz sınıflandırma sonuçları Çizelge 5.6, Çizelge 5.7, Çizelge 5.8 ve Çizelge 5.9 ’da verilmiştir. Çizelge 5.7’den görüldüğü gibi sadece 7 yeni öznitelikle elde edilen çapraz test sonuçları SMO-Emo_large-Kızgın-Nötr-Mutlu-Üzgün sınıflandırıcı hariç diğer tüm öznitelik gruplarının sonuçlarından daha iyidir. Kızgın-nötr sınıflandırıcıda 7 öznitelikle çapraz testte EmoSTAR eğitim, EmoDB test seti olması durumunda %96,11 başarı oranı elde edilebilmiştir. Bu sonuçlar yeni özniteliklerin veritabanları arasındaki değişikliklere karşı daha dirençli olduğunu göstermektedir.

Çizelge 5.6 EmoSTAR ve EmoDB için 315 yeni öznitelikle çapraz test sonuçları

	EmoSTAR Eğitim EmoDB Test-315 STAFFER			EmoDB Eğitim EmoSTAR Test- 315 STAFFER		
	NB	SMO	Bag	NB	SMO	Bag
Kızgın-Nötr (Emo_large-91,26) ¹	78,64	89,80	89,32	75,10	80,34	69,86
Mutlu-Nötr-Üzgün (Emobase-62,73) ¹	27,35	62,73	50,47	39,31	43,79	26,89
Kızgın-Nötr-Mutlu-Üzgün (Emo_large-62,83) ¹	40,11	63,12	51,32	36,89	39,18	41,98

Çizelge 5.7 EmoSTAR ve EmoDB için 7 yeni öznitelikle çapraz test sonuçları

	EmoSTAR Eğitim EmoDB Test-7 STAFFER			EmoDB Eğitim EmoSTAR Test-7 STAFFER		
	NB	SMO	Bag	NB	SMO	Bag
Kızgın-Nötr (Emo_large-91,26) ¹	96,11	75,72	78,15	74,23	74,67	76,41
Mutlu-Nötr-Üzgün (Emobase-62,73) ¹	61,79	66,50	49,52	48,27	54,48	49,65
Kızgın-Nötr-Mutlu-Üzgün (Emo_large-62,83) ¹	58,11	49,26	49,26	46,56	46,81	49,10

¹ Çapraz testlerde bu aşamaya kadar elde edilen en iyi sonuç

Çizelge 5.8 EmoSTAR ve EmoDB için Emobase konfigürasyonuna yeni özniteliklerin eklenmesiyle elde edilen çapraz test sonuçları

	EmoSTAR Eğitim EmoDB Test Emobase+STAFFER (988+315)			EmoDB Eğitim EmoSTAR Test Emobase+STAFFER (988+315)		
	NB	SMO	Bag	NB	SMO	Bag
Kızgın-Nötr (Emo_large-91,26) ¹	75,72	90,77	88,83	58,07	71,17	58,95
Mutlu-Nötr-Üzgün (Emobase-62,73) ¹	58,49	44,81	55,18	46,55	54,48	51,37
Kızgın-Nötr-Mutlu-Üzgün (Emo_large-62,83) ¹	59,58	56,63	41,29	43,25	46,31	42,23

Çizelge 5.9 EmoSTAR ve EmoDB için Emo_large konfigürasyonuna yeni özniteliklerin eklenmesiyle elde edilen çapraz test sonuçları

	EmoSTAR Eğitim EmoDB Test Emolarge+STAFFER (6669+315)			EmoDB Eğitim EmoSTAR Test Emolarge+STAFFER (6669+315)		
	NB	SMO	Bag	NB	SMO	Bag
Kızgın-Nötr (Emo_large-91,26) ¹	71,84	92,23	75,72	53,27	70,74	49,78
Mutlu-Nötr-Üzgün (Emobase-62,73) ¹	45,75	49,05	58,01	48,96	51,72	33,44
Kızgın-Nötr-Mutlu-Üzgün (Emo_large-62,83) ¹	45,13	60,76	34,51	42,74	42,74	33,58

5.5 Eklenen Yeni Özniteliklerle Öznitelik Seçme Uygulamaları

Dördüncü bölümde EmoSTAR ve EmoDB üzerinde öznitelik seçme algoritmalarını uygulamış ve başarılı sonuçlar elde etmiştik. Bu bölümde de aynı algoritmaları eklenen yeni özniteliklerle deneyerek performanslarını ve yeni özniteliklerin seçilen öznitelikler arasındaki sıralamasını gözlemleyeceğiz. Çizelge 5.10 ve Çizelge 5.11’de EmoDB ve EmoSTAR’da uygulanan öznitelik seçme algoritmalarının önerilen yeni enerji operatörlerine (Low/High Energy Ratio (LH), LH2, LH3) verdiği sıralama gösterilmektedir. Özellikle EmoDB’de yeni enerji operatörleri en yüksek öncelikli öznitelik olarak karşımıza çıkmaktadır. EmoSTAR’da ise sıralama değerleri her ne kadar daha düşük olsa da öznitelik sayısı göz önüne alındığı takdirde oldukça iyi kabul edilebilir. Parantez içindeki sayılar toplam öznitelik sayısını göstermektedir.

¹ Çapraz testlerde bu aşamaya kadar elde edilen en iyi sonuç

Çizelge 5.10 EmoDB’de 7 duygu için yeni enerji operatörlerinin öznitelik seçme sıralama sonuçları

EmoDB	Emobase+STAFFER (988+315)			Emo_large+STAFFER (6669+315)		
	LH	LH2	LH3	LH	LH2	LH3
InfoG+Rank	2	5	55	25	60	250
ChiSqu+Rank	4	17	46	37	87	227

Çizelge 5.11 EmoSTAR’da yeni enerji operatörlerinin öznitelik seçme sıralama sonuçları

EmoSTAR	Emobase+STAFFER (988+315)			Emo_large+STAFFER (6669+315)		
	LH	LH2	LH3	LH	LH2	LH3
InfoG+Rank	145	166	547	513	627	2870
ChiSqu+Rank	131	178	614	447	740	3143

F0 aritmetik ortalama ve f0 standart sapma öznitelikleri openSMILE tarafından hesaplanmasına rağmen yeni öznitelik grubunda farklı bir yöntemle tekrar hesaplanmıştır. Çizelge 5.11 ve Çizelge 5.12’de iki değişik yöntemle elde edilen değerlerin öznitelik seçicilerde elde ettiği sıralama sonuçları gösterilmektedir.

Çizelge 5.12 EmoDB’de 7 duygu için yeni özniteliklerin öznitelik seçme sıralama sonuçları

EmoDB	Emobase+STAFFER (988+315)			Emo_large+STAFFER (6669+315)		
	f0mean ¹	f0SDev ²	SRate ³	f0mean	f0SDev	SRate
InfoG+Rank	50-(145) ⁴	74-(128) ⁴	284	240-(1115) ⁴	308-(1102) ⁴	1742
ChiSqu+Rank	79-(187) ⁴	77-(148) ⁴	130	342-(1485) ⁴	332-(1451) ⁴	652

Çizelge 5.13 EmoSTAR’da yeni özniteliklerin öznitelik seçme sıralama sonuçları

EmoSTAR	Emobase+STAFFER (988+315)			Emo_large+STAFFER (6669+315)		
	f0mean	f0Sdev	SRate	f0mean	f0Sdev	SRate
InfoG+Rank	445-(554) ⁴	258-(295) ⁴	432	2309-(4974) ⁴	1151-(3448) ⁴	2227
ChiSqu+Rank	429-(541) ⁴	273-(353) ⁴	338	2152-(4908) ⁴	1313-(3743) ⁴	1691

OpenSMILE f0 ortalama değerini hesaplarken f0’ın sıfır olduğu sessiz kısımları da hesaba katmaktadır. Bu nedenle kızgın kadın konuşmalarında dahi ortalama değerler 140 çıkabilmektedir ki bunun pek de doğru bir yaklaşım olduğu söylenemez.

¹ Temel Frekans Aritmetik Ortalama

² Temel Frekans Standart Sapma

³ Sessizlik Oranı (Silence Rate)

⁴ openSMILE özniteliklerine ait sıralama değerleri

Maksimum f0 olarak konfigürasyon dosyalarında belirlenen 500 Hz değeri de düşük gözükmemektedir. Çizelge 5.14'ten görüldüğü gibi EmoDB'de Information Gain ve ChiSquare öznitelik seçici Ranker arama yöntemiyle kullanıldığında seçilen 997 öznitelikle Emobase+STAFFER konfigürasyonunda %89,71 ile en yüksek başarı oranı yakalanmıştır ki bu oran bugüne kadar EmoDB üstünde elde edilmiş en yüksek başarı oranıdır. CfsSubsetEvaluator öznitelik seçici Linear Forward Selection metoduyla sadece 130 öznitelikle %88,78 başarı oranını yakalayabilmiştir.

Çizelge 5.14 EmoDB'de 7 duygu için eklenen yeni özniteliklerle öznitelik seçme sonuçları

EmoDB	Emobase+STAFFER (988+315) ¹			Emo_large+STAFFER(6669+315) ¹		
	NB (75,88) ²	SMO (88,03) ²	Bag (69,71) ²	NB (77,00) ²	SMO (87,10) ²	Bag (74,01) ²
CfsSub+LFS (153,130) ³	79,43	87,10	70,46	81,49	88,78	74,76
InfoG+Rank (997,5698) ³	75,32	89,71	69,53	71,02	88,59	72,71
ChiSqu+Rank (997,5698) ³	75,32	89,71	69,34	71,02	88,59	72,52

Çizelge 5.15'te EmoSTAR'daki öznitelik seçme metotları incelendiği takdirde öznitelik seçme algoritmaları yeni öznitelikleri kullanarak tüm özniteliklerle elde edilen başarı oranlarını geçmeyi başarmışlardır.

Çizelge 5.15 EmoSTAR'da 4 duygu için eklenen yeni özniteliklerle öznitelik seçme sonuçları

EmoSTAR	Emobase+STAFFER (988+315) ¹			Emo_large+STAFFER (6669+315) ¹		
	NB (87,27) ²	SMO (96,69) ²	Bag (86,76) ²	NB (87,78) ²	SMO (97,45) ²	Bag (88,80) ²
CfsSub+LFS (135,127) ³	93,12	96,94	87,78	92,62	97,96	89,31
InfoG+Rank (1132,5932) ³	90,33	96,94	87,02	85,24	97,45	89,05
ChiSqu+Rank (1132,5932) ³	90,33	96,94	86,51	84,73	97,45	88,80

CfsSubsetEvaluator öznitelik seçici Linear Forward Selection metoduyla sadece 127 öznitelikle %97,96 başarı oranını yakalayabilmiştir. Bu oran 6669 öznitelikli Emo_large öznitelik setiyle elde edilen %97,45 başarı oranının dahi üzerindedir. Genel olarak

¹ Tüm öznitelik sayısı

² Tüm öznitelikler ile elde edilen en yüksek başarı oranları

³ Sırasıyla Emobase ve Emo_large için seçilen öznitelik sayıları

değerlendirildiği takdirde yeni özniteliklerin kullanıldığı konfigürasyonlar orijinal konfigürasyonlara göre hem tüm özniteliklerin kullanılması hem de öznitelik seçicilerin kullanılması durumlarında daha iyi sonuçlar vermektedirler. Öznitelik seçiciler eklenen yeni özniteliklerle orijinal özniteliklere göre EmoDB (Çizelge 4.22 ve Çizelge 5.14) ve EmoSTAR'da (Çizelge 4.23 ve Çizelge 5.15) tüm sınıflandırıcılarda daha yüksek başarı oranını yakalayabilmektedir.

SONUÇ VE ÖNERİLER

Duygu insanları diğer varlıklardan ve makinelerden ayıran en belirgin özelliklerdendir ve insanlar arasındaki iletişimde çok önemli bir yer tutmaktadır. Duygu tanıma biz insanlar için sıradan bir iş olarak gözükmese de, makinelerde uygulanmak istenince olayın karmaşıklığı ortaya çıkmaktadır. Son yıllarda insan-makine arasındaki iletişimde büyük ilerlemeler kaydedilmiştir. Ancak, makinelerin ürettiği mekanik sesler insan-makine etkileşiminin itici olmasına sebep olmaktadır. Makinelere duygu yüklü konuşmayı öğretmek ve karşısındaki kişinin duygu halini anlayabilmesini sağlamak bu yeni etkileşimin daha çekici olmasını sağlayacaktır. Duygu tanıma çalışmalarının en önemli öğelerinden biri duygu veritabanlarıdır. Bu çalışmada yeni bir duygu veritabanı sunulmuş ve mevcut veritabanlarından biriyle çapraz testler gerçekleştirilmiştir. Duygu tanıma çalışmalarında tek setli uygulamalarda oldukça yüksek başarı oranlarına ulaşılabilmesine rağmen çok setli çalışmalarda başarı oranları düşük çıkmaktadır. Bu sorunun çözümü daha geniş veritabanları geliştirmek ve daha iyi özellik kümeleri üretebilmektir. Geliştirdiğimiz veritabanı daha doğal ve heterojen özelliklere sahiptir. Bu nedenle çapraz testlerde her iki veritabanında daha başarılı sonuçlar üretebilmiştir. Veri setlerinde Naive Bayes, SVM_SMO ve Bagging sınıflandırıcılar 988 ve 6669 özellik içeren iki farklı özellik konfigürasyonu ile uygulanmıştır. SVM-SMO tek setli uygulamalarda çok başarılı olmasına rağmen iki setli çapraz testlerde başarı oranları en fazla kayba uğrayan sınıflandırıcı olmaktadır. Özellik sayısındaki artış başarı oranlarına her zaman olumlu etki yapmamıştır. Özellik sayısının 988'den 6669'a çıkmasına rağmen başarı oranlarında beklenen artış gerçekleşmediği gibi aksine özellik seçiciler 988

özellik içeren konfigürasyonu ile her iki veritabanında daha iyi özellik kümeleri üretmeyi başarmışlardır.

Çalışmamızın diğer bir unsuru da özellik seçicilerle daha iyi özellik kümeleri üretebilmektir. Özellik seçiciler oldukça güçlü bir performans ortaya koymuş ve tüm sınıflandırmalarda ve her iki veritabanında daha az özellik sayısı ile orijinal setten daha yüksek başarı oranlarına ulaşmayı başarmışlardır.

Bu çalışmada mevcut özelliklere zamansal-spektrum özellikleri ve bunlara ait istatistiksel fonksiyonlarla elde edilen özellikler eklenmiş ve performans artışı sağlanmıştır. Önerilen özellikler yerleşmiş geleneklere aykırı olarak spektrumdaki logaritmik şiddet değerlerini değil mutlak şiddet değerlerini kullanmışlardır. Bu şekilde spektral bölgelerdeki, logaritma fonksiyonu ile bastırılan farkların sınıflandırıcılar tarafından daha kolay görülmesi ve ayırt edilebilmesi amaçlanmıştır. Bilindiği gibi insan kulağı 1.000 Hz'e kadar doğrusal sonrasında ise logaritmik bir frekans analiz cihazı gibi çalışmaktadır. Yeni özellikler spektrumun doğrusal ve logaritmik bölgelerindeki enerjinin oranını ele almaktadır. Önerilen yeni özellikler bütün sınıflandırmalarda ve her iki veritabanında daha iyi sonuçlar üretmeyi başarmışlardır. Özellikle çok setli çapraz testlerde yeni özellikler veritabanları arasındaki değişikliklere karşı daha dirençli gözükmemektedir. Öyle ki yeni özelliklerin sadece 7'si kullanılarak diğer konfigürasyonlardan daha başarılı sonuçlar alınabilmektedir.

Özellik seçiciler yeni özelliklerle uygulandığında 1303 özellik içeren Emobase+yeni konfigürasyonunda 6984 özellik içeren Emo_large+yeni konfigürasyonundan daha iyi sonuçlar üretebilmiştir. Eklenen yeni özellikler özellik seçme performanslarında da artış sağlamıştır. Özellikle EmoDB veritabanında yeni özellikler özellik seçiciler tarafından en üst sıralarda seçilmektedirler. Temel frekans ortalama değeri ve standart sapma değeri openSMILE ile elde edilmesine rağmen farklı bir yöntemle hesaplanarak önerilen yeni özelliklerin içine dahil edilmiş ve özellik seçicilerde sıralamaya tabi tutulmuşlardır. Yeni yöntemle hesaplanan temel frekans ortalama değeri ve standart sapma değeri özellik seçicilerde daha öncelikli özellik olarak seçilmektedirler. Bu açıdan bakıldığında openSMILE tarafından elde edilen diğer değerlerin de bir incelemeye tabi tutulması faydalı gözükmemektedir.

Tek setli alıřmalarda en bařarılı sınıflandırıcı SMO, en yavaş sınıflandırıcı Bagging, bařarı oranı en düşük sınıflandırıcı Naive Bayes olmuřtur. ok setli apraz testlerde SMO tek setli bařarı oranlarına gre en fazla dřüş gsteren sınıflandırıcı olmaktadır.

apraz testlerde daha bařarılı sonular almak iin daha geniř ve doęal duygu veritabanlarına ihtiya duyulmaktadır. İleriki alıřmalarda deneyler daha byk veritabanları ile sesin yanında yz ifadesi ve kelime tanıma kullanılarak yapılabilir. zellik seicilerle elde edilen zellik gruplarının apraz testlerdeki performansları da ayrı bir arařtırma konusudur.

KAYNAKLAR

- [1] Oflazoglu, C. ve Yildirim, S., (2013). "Recognizing Emotion from Turkish Speech Using Acoustic Features", EURASIP Journal on Audio, Speech and Music Processing, 2013(1):1-11.
- [2] Zhang, Z., Weninger, F., Wöllmer, M. ve Schuller, B., (2011). "Unsupervised Learning in Cross-Corpus Acoustic Emotion Recognition", IEEE Workshop on Automatic Speech Recognition and Understanding (ASRU), December 2011, Waikoloa, Hawaii, 523-528.
- [3] Schuller, B., Vlasenko, B., Eyben, F., Wollmer, M., Stuhlsatz, A., Wendemuth, A. ve Rigoll, G., (2010). "Cross-corpus Acoustic Emotion Recognition: Variances and Strategies". Proceedings of IEEE Trans. on Affective Computing 1(2):119–131.
- [4] Schuller, B., Zhang, Z., Weninger, F. ve Rigoll, G., (2011). "Selecting Training Data for Cross-Corpus Speech Emotion Recognition: Prototypicality vs. Generalization", Proceedings of 2011 Afeka-AVIOs Speech Processing Conference, Telaviv, Israel.
- [5] Eyben, F., Batliner, A., Schuller, B., Seppi, D. ve Steidl, S., (2010). "Cross-Corpus Classification of Realistic Emotions – Some Pilot Experiments", Proc. 7th Intern. Conf. on Language Resources and Evaluation (LREC 2010), Valletta, ELRA, 77-82.
- [6] Schuller, B., Vlasenko, B., Eyben, F., Rigoll, G. ve Wendemuth, A., (2009). "Acoustic Emotion Recognition: A Benchmark Comparison of Performances", Proceedings of IEEE Workshop on Automatic Speech Recognition & Understanding, ASRU 2009, Merano, 552-557.
- [7] He, L., (2010). Stress and Emotion Recognition in Natural Speech in the Work and Family Environments, PhD Thesis, RMIT University.
- [8] Bhargava, M. ve Polzehl, T., (2012). "Improving Automatic Emotion Recognition from Speech Using Rhythm and Temporal Feature", arXiv preprint arXiv:1303.1761.
- [9] Batliner, A., Steidl, S., Schüller, B., Seppi, D., Vogt, T., Wagner, J., Devillers, L., Vidrascu, L., Aharonson, V., Kessous, L. ve Amir, N., (2011). "Whodunnit

- Searching for the Most Important Feature Types Signalling Emotion-Related User States in Speech”, *Computer Speech & Language*, 25(1):4-28.
- [10] Shami, M. ve Verhelst, W., “An Evaluation of the Robustness of Existing Supervised Machine Learning Approaches to the Classification of Emotions in Speech”, *Speech Communication*, 49(3):201-212.
 - [11] Chavhan, Y., Dhore, M. L. ve Yesaware, P., (2010). “Speech Emotion Recognition Using Support Vector Machine”, *International Journal of Computer Applications*, 1(20):6-9.
 - [12] Wu, S., Falk T.H. ve Chan W., (2010). "Automatic Speech Emotion Recognition Using Modulation Spectral Features", *Speech Communication*, 53(5):768-785.
 - [13] Schüller, B., Batliner, A., Seppi, D., Steidl, S., Vogt, T., Wagner, J., Devillers, L., Vidrascu, L., Amir, N., Kessous L. ve Aharonson, V., (2007). “The Relevance of Feature Type for the Automatic Recognition of Emotional User States: Low Level Descriptors and Functionals”, *INTERSPEECH 2007*, Antwerp, 2253-2256.
 - [14] Bozkurt, E., Erzin, E., Erdem, C. E. ve Erdem, A. T., (2011). "Formant Position Based Weighted Spectral Features For Emotion Recognition", *Speech Communication*, 53(9):1186-1197.
 - [15] Wu, D., Parsons, D.T. ve Narayanan, S.S., (2010). “Acoustic Feature Analysis in Speech Emotion Primitives Estimation”, *INTERSPEECH 2010*, 785-788.
 - [16] Bradley M., (1994). “Emotional Memory: A Dimensional Analysis”, *Emotions: Essays on Emotion Theory*, 97-134.
 - [17] Lang, P.J., (1993). “The Network Model of Emotion: Motivational Connections”, *Advances in Social Cognition*, 6:109-133.
 - [18] Osgood C., Suci G. ve Tannenbaum P. H., (1957). *The Measurement of Meaning*, Urbana, University of Illinois Press, 195:36.
 - [19] Russell J. ve Mehrabian A., (1977). “Evidence for a Three-Factor Theory of Emotions”, *Journal of Research in Personality*, 11(3):273–294.
 - [20] Mena, M.E., (2012). *Emotion Recognition From Speech Signals*, MSc Thesis, University Of Ljubljana.
 - [21] Wikipedia, en.wikipedia.org/wiki/Emotion, 01 Mayıs 2014.
 - [22] Devillers, L., Cowie, R., Martin, J.-C., Douglas-Cowie, E., Abrilian, S. ve McRorie, M., (2006). “Real Life Emotions in French and English TV Video Clips: An Integrated Annotation Protocol Combining Continuous and Discrete Approaches”, *5th International Conference on Language Resources and Evaluation (LREC 2006)*, Genoa, Italy, 22.
 - [23] Cowie, E.D., Cowie, R., Sneddon, I., Cox, C., Lowry, O., McRorie, M., Martin, J. C., Devillers, L., Abrilian, S., Batliner, A., Amir, N. ve Karpouzis K., (2007). “The HUMAINE Database: Addressing the Collection and Annotation of Naturalistic and Induced Emotional Data”, *Affective Computing and Intelligent Interaction, ACII 2007*, 488–500.

- [24] Whissel C.M., (1989). The Dictionary of Affect in Language, Emotion: Theory, Research and Experience: 4, The Measurement of Emotions, R. Plutchik ve H. Kellerman, Academic Press, New York, NY.
- [25] Plutchik R., (1994). The Psychology and Biology of Emotion, Harper Collins College Publishers.
- [26] Cowie, R., Douglas-Cowie, E., Tsapatsoulis, N., Votsis, G., Kollias, S., Fellenz, W. ve Taylor J.G., (2001). "Emotion Recognition in Human-Computer Interaction", IEEE Signal Processing Magazine, 18(1):32-80.
- [27] Pradier, M. F., (2011). Emotion Recognition from Speech Signals and Perception of Music, PhD Thesis, Universität Stuttgart.
- [28] Güneş, H., Piccardi, M. ve Pantic, M., (2008). From the Lab to the Real World: Affect Recognition Using Multiple Cues and Modalities, InTech Education and Publishing, Vienna, Austria, 185-218.
- [29] Wikipedia, en.wikipedia.org/wiki/anechoic_chamber, 01 Mayıs 2014.
- [30] The Sound Blog, acousticengineering.files.wordpress.com/2013/07/anechoic-6.jpg, 01 Mayıs 2014.
- [31] Ververidis, D. ve Kotropoulos, C., (2006). "Emotional Speech Recognition: Resources, Features and Methods", Speech Communication 48(9):1162–1181.
- [32] Ramakrishnan, S., (2012). "Recognition of Emotion from Speech: A Review", Speech Enhancement, Modeling and Recognition—Algorithms and Applications, 121.
- [33] Haq, S. ve Jackson, P.J.B., (2011). "Multimodal Emotion Recognition", Machine Audition: Principles, Algorithms and Systems, IGI Global, Hershey, 398-423.
- [34] Schuller, B., Batliner, A., Steidl, S. ve Seppi, D., (2011). "Recognising Realistic Emotions and Affect in Speech: State Of The Art and Lessons Learnt From The First Challenge," Speech Communication, 53(9), 1062–1087.
- [35] Castellano G., Leite I., Pereira, A., Martinho, C., Paiva, A. ve Mc Owan, P.W., (2009, September). "It's All in the Game: Towards an Affect Sensitive and Context Aware Game Companion", IEEE 3rd International Conference on Affective Computing and Intelligent Interaction and Workshops, 2009. ACII 2009, 1-8.
- [36] Voran S., (2008). "Listener Detection of Talker Stress in Low-Rate Coded Speech", IEEE International Conference on Acoustic, Speech and Signal Processing, 2008, ICASSP 2008, 4813-4816.
- [37] Rabiner, L. R. ve Schafer, R. W., (2007). "Introduction to Digital Speech Processing", Foundations and Trends in Signal Processing, 1(1):1-194.
- [38] Picone, J., (1996). Fundamentals Of Speech Recognition: A Short Course, Institute for Signal and Information Processing, Mississippi State University.

- [39] Zhou G., Hansen J.H.L. ve Kaiser J.F., (2001). "Nonlinear Feature Based Classification of Speech Under Stress", IEEE Transactions on Speech and Audio Processing, 9(3):201-216.
- [40] Kaiser J., (1983). "Some Observations on Vocal Tract Operation from a Fluid Flow Point of View", Vocal Fold Physiology: Biomechanics, Acoustics, and Phonatory Control, 358-386.
- [41] Cairns D.A. ve Hansen J.H.L., (1994). "Nonlinear Analysis and Classification of Speech Under Stressed Conditions", The Journal of Acoustical Society of America, 96(6):3392-3400.
- [42] Nwe T.L., Foo S.W. ve De Silva L.C., (2003). "Classification of Stress in Speech Using Linear and Nonlinear Features", IEEE International Conference on Acoustics, Speech, and Signal Processing, ICASSP 2003, 2:9-12.
- [43] Torabi S., AlmasGanj F. ve Mohammadian A., (2008). "Semi-Supervised Classification of Speaker's Psychological Stress", IEEE Biomedical Engineering Conference, CIBEC 2008, Cairo, 1-4.
- [44] Shami M. ve Verhelst W., (2007). "An Evaluation of the Robustness of Existing Supervised Machine Learning Approaches to the Classification of Emotions in Speech", Speech Communication, 49(3):201-212.
- [45] Nwe T.L., Foo S.W. ve De Silva L.C., (2003). "Speech Emotion Recognition Using Hidden Markov Models". Speech Communication, 41(4):603-623.
- [46] Alpaydin, E., (2010). Introduction to Machine Learning, MIT press, 2010.
- [47] Burkhardt, F., Paeschke, A., Rolfes, M., Sendlmeier, W.F. ve Weiss, B., (2005). "A Database of German Emotional Speech", INTERSPEECH 2005, 5:1517-1520.
- [48] Eyben, F., Wöllmer, M. ve Schuller, B., (2009). "openSMILE - The Munich Versatile and Fast Open-Source Audio Feature Extractor", Proceedings of International Conference on Multimedia, ACM, October 2010, Florence, Italy, 1459-1462.
- [49] Young, S., Everman, M. G., Hain, T., Kershaw, D., Liu, X., Odell, J., Ollason, D., Povey, D., Valtchev, V. ve Woodland, P., (2006). The HTK Book (v3.4), Cambridge University Press.
- [50] Boersma, P. and Weenink, D., (2005). Praat: Doing Phonetics by Computer (v. 4.3.14). <http://www.praat.org/>, 2015.
- [51] Fernandez, R., A Computational Model for the Automatic Recognition of Affect in Speech. PhD Thesis, MIT Media Arts and Science, Feb. 2004.
- [52] Walker, W., Lanere, P., Kwok, P., Raj, B., Singh, R., Gouvea, E., Wolf, P. ve Woelfel, J., (2004). Sphinx-4: A Flexible Open Source Framework for Speech Recognition, SMLI TR-2004-139.
- [53] Bouckaert, R. R., Frank E., Hall, M., Kirkby, R., Reutemann, P., Seewald, A. ve Scuse, D., (2013). WEKA Manual, University of Waikato, Hamilton, New Zealand.

ÖZGEÇMİŞ

KİŞİSEL BİLGİLER

Adı Soyadı : Cevahir PARLAK
Doğum Tarihi ve Yeri : 01.06.1968 OF
Yabancı Dili : İngilizce
E-posta : cevahir.parlak@boun.edu.tr

ÖĞRENİM DURUMU

Derece	Alan	Okul/Üniversite	Mezuniyet Yılı
Y. Lisans	Bilgisayar Müh	Yıldız Teknik Üni.	-
Lisans	Bilgisayar Müh	Sakarya Üniversitesi	2012
Lise	Fen	Kazım İşmen Lisesi	1986

YAYINLARI

Makale

1. Parlak, C., Diri, B., (2014). “Farklı Veri Setleri Arasında Duygu Tanıma Çalışması”, DEÜ Mühendislik Fakültesi Mühendislik Bilimleri Dergisi, 16(48):21-29.

Bildiri

1. Parlak, C. ve Diri, B.,(2013). “İnsan Sesinden Duygu Çıkarma”, IEEE 21. Sinyal İşleme Uygulamaları Kurultayı, (SIU), 2013, Kıbrıs.
2. Parlak, C., (2013). “Green Cellular Networks”, IEEE 4th International Conference on Energy Aware Computing Systems and Applicaitons, (ICEAC), 2013, İstanbul, 158-163.
3. Parlak, C., Bilgin, G., (2014). “Hiperspektral Görüntü Sıkıştırma Yöntemlerinin Karşılaştırılması”, IEEE 22. Sinyal İşleme Uygulamaları Kurultayı, (SIU), 2014, Trabzon, 200-203.
4. Parlak, C., Diri, B., Gürgen, F., (2014) “A Cross-Corpus Experiment in Speech Emotion Recognition”, Proceedings of the 2nd International Workshop on Speech, Language and Audio in Multimedia (SLAM), 2014, Malezya, 58-61.
5. Parlak, C., Diri, B., (2014). “Duygu Tanımadaki Farklı Veri Setleri Arasında Bir Çalışma”, Akıllı Sistemlerde Yenilikler ve Uygulamalar, (ASYU), 2014, İzmir, 31-34.