



MUSTAFA KEMAL ÜNİVERSİTESİ
FEN BİLİMLERİ ENSTİTÜSÜ
ELEKTRİK-ELEKTRONİK MÜHENDİSLİĞİ
ANABİLİM DALI

TÜRKÇE KONUŞMA SİNYALİNDEN
DUYGU TANIMA

ÇAĞLAR OFLAZOĞLU

YÜKSEK LİSANS TEZİ

Antakya/HATAY

ARALIK-2011



MUSTAFA KEMAL ÜNİVERSİTESİ
FEN BİLİMLERİ ENSTİTÜSÜ
ELEKTRİK-ELEKTRONİK MÜHENDİSLİĞİ
ANABİLİM DALI

TÜRKÇE KONUŞMA SİNYALİNDEN
DUYGU TANIMA

ÇAĞLAR OFLAZOĞLU

YÜKSEK LİSANS TEZİ

Antakya/HATAY

ARALIK-2011

MUSTAFA KEMAL ÜNİVERSİTESİ
FEN BİLİMLERİ ENSTİTÜSÜ

TÜRKÇE KONUŞMA SİNYALİNDEN
DUYGU TANIMA

ÇAĞLAR OFLAZOĞLU
YÜKSEK LİSANS TEZİ

ELEKTRİK-ELEKTRONİK MÜHENDİSLİĞİ ANABİLİM DALI

Yrd. Doç. Dr. Serdar YILDIRIM danışmanlığında hazırlanan bu tez 30/12/2011 tarihinde aşağıdaki jüri üyeleri tarafından oybirliği ile kabul edilmiştir.

Yrd.Doç.Dr. Serdar YILDIRIM
Başkan

Yrd.Doç.Dr. Ersin ÖZDEMİR
Üye

Yrd.Doç.Dr. Yakup KUTLU
Üye

Bu tez Enstitümüz Elektrik-Elektronik Mühendisliği Anabilim Dalında hazırlanmıştır.

Kod No:

Prof. Dr. Hüseyin GÖZÜBENLİ
Enstitü Müdürü

Bu çalışma TÜBİTAK tarafından desteklenmiştir.

Proje No: 109E243

Not: Bu tezde kullanılan özgün ve başka kaynaktan yapılan bildirişlerin, çizelge, şekil ve fotoğrafların, kaynak gösterilmeden kullanımı, 5846 sayılı Fikir ve Sanat Eserleri Kanunundaki hükümlere tabidir.

İÇİNDEKİLER

	Sayfa
ÖZET.....	III
ABSTRACT.....	IV
SİMGELER VE KISALTMALAR DİZİNİ	V
ÇİZELGELER DİZİNİ	VI
ŞEKİLLER DİZİNİ.....	VIII
1. GİRİŞ.....	1
2. ÖNCEKİ ÇALIŞMALAR.....	3
2.1. Giriş.....	3
2.2. Literatürdeki Veritabanları.....	3
2.3. Yapılan Çalışmanın Kapsamı.....	6
3. MATERYAL VE YÖNTEM	7
3.1. Materyal	7
3.1.1. Türkçe Duygusal Konuşma Veritabanının Oluşturulması.....	7
3.2. Yöntem.....	10
3.2.1. Veritabanının Duygu İçeriklerinin Belirlenmesi	10
3.2.2. Türkçe Konuşmanın Akustik Özelliklerinin Değerlendirilmesi.....	14
3.2.3. Duygu Tanıma ve Duygu Boyutlarının Kestirimi	15
3.2.3.1. Kategorik Duygu Tanıma.....	15
3.2.3.2. Duygu Boyutlarının Kestirimi.....	16
3.2.4. Duygu Tanımda ve Duygu Boyutlarının Kestirimde Öznitelik Seçimi ve Sınıflandırma Yöntemleri	17
3.2.4.1. Öznitelik Seçimi.....	18
3.2.4.1.1. İlinti Tabanlı Öznitelik Seçimi	18
3.2.4.1.2. Bilgi Kazancı Öznitelik Değerlendirme Yöntemi	19
3.2.4.1.3. RELIEF-F Öznitelik Değerlendirme Yöntemi	20
3.2.4.2. Sınıflandırma ve Kestirim Yöntemleri.....	20
3.2.4.2.1. Naive Bayes	20
3.2.4.2.2. Bayes Ağı.....	21
3.2.4.2.3. Destek Vektör Makineleri.....	22
3.2.4.2.4. Destek Vektör Regresyonu	22
3.2.4.3. Kategorik Modelin Başarım Değerlendirmesi	23
3.2.4.4. Duygu Boyutlarını Kestirim Modelinde Başarım Değerlendirmesi	25
4. ARAŞTIRMA BULGULARI VE TARTIŞMA.....	26
4.1. Türkçe Duygusal Konuşma Veritabanının Değerlendirilmesi	26

4.1.1. Konuşma Sinyallerinin Duygu İçeriklerinin Belirlenmesi	26
4.1.1.1. Kategorik Değerlendirme Sonuçları	26
4.1.1.2. 3-Boyutlu Duygu Uzayı Değerlendirme Sonuçları.....	28
4.1.2. Kategorik Sınıfların ve 3-Boyutlu Duygu Uzayı Değerlendirmelerinin Birlikte Analizi	30
4.2. Türkçe Duygusal Konuşmanın Akustik Özelliklerinin İncelenmesi.....	32
4.2.1. Temel Frekans İncelemesi	33
4.2.2. Formant Frekansları Analizi	36
4.2.3. Ünlü Harflerin Duygu Durumuna Göre Sürelerinin Analizi	42
4.3. Türkçe Konuşmadan Duygu Tanıma ve Duygu Boyutlarının Kestirimi	44
4.3.1. Duygu Tanıma ve Duygu Boyutları Kestiriminde Kullanılan Akustik Öznitelikler	44
4.3.2. Kategorik Modelin Başarım Değerlendirme Sonuçları	45
4.3.3. Duygu Boyutları Kestirim Sonuçları	49
5. SONUÇ VE ÖNERİLER	51
KAYNAKLAR	52
TEŞEKKÜR.....	55
ÖZGEÇMİŞ	56

ÖZET**TÜRKÇE KONUŞMA SİNYALİNDEN DUYGU TANIMA**

İnsan-bilgisayar etkileşim teknolojisinde, kullanıcı ve bilgisayar arasındaki etkileşimi daha doğal hale getirecek arayüzlerin tasarımı önemli bir konu haline gelmiştir. Etkileşim sırasında bu tür arayüzlerin uygulanmasına yönelik anahtar adımlardan biri de kullanıcının duygusal durumlarının belirlenmesidir. Bu çalışmada, Türkçe konuşma sinyalinden duygu tanıma konusu araştırılmıştır. Bu amaçla, 5304 adet konuşma sinyalini içeren Türkçe duygusal konuşma veritabanı oluşturulmuştur. Oluşturulan veritabanı konuşma sinyallerinin hem kategorik hem de 3-boyutlu duygu uzayı değerlendirmesini içermektedir. Türkçe konuşmanın akustik parametrelerinin duygulardan nasıl etkilendiği hakkında bilgi elde edebilmek için; temel frekans, formant frekansları ve ünlü harf süreleri gibi konuşma prosodisiyle ilgili bir grup akustik parametre, duygunun fonksiyonu olarak detaylı bir şekilde analiz edilmiştir. Bu çalışmada özellikle konuşma sinyalinden *negatif-negatif değil* ve *kızgın-kızgın değil* ikili sınıflandırma problemleri ile temel duygu boyutlarının (değerlik, aktivasyon ve baskınlık) kestirimi üzerine odaklanılmıştır. Konuşma sinyalinden otomatik duygu tanıma için birçok makine öğrenme ve öznitelik seçim algoritmaları incelenmiştir.

2011, 54 sayfa

Anahtar kelimeler: Duygusal konuşma veritabanı, duygusal konuşmanın akustik özellikleri, duygu tanıma, temel duygu boyutları kestirimi

ABSTRACT**EMOTION RECOGNITION FROM TURKISH SPEECH SIGNAL**

An emerging trend in human–computer interaction technology is to design spoken interfaces that facilitate more natural interaction between a user and a computer. Being able to detect the user’s affective state during interaction is one of the key steps toward implementing such interfaces. In this study, emotion recognition from Turkish speech signal is explored. For this purpose, first a Turkish emotional speech database which includes 5304 speech signals is constructed. The database includes both categorical and 3-dimensional emotional space ratings of speech signals. In order to obtain knowledge on how acoustic parameters of Turkish speech are affected from emotions, a set of acoustic parameters related to speech prosody such as fundamental frequency, formant frequencies and vowel durations were analyzed in detail as a function of emotion. This study particularly focuses on two binary classification problems, recognizing *negative* vs. *non-negative* emotions and recognizing *angry* vs. *non-angry* emotions, and the problem of estimating emotion primitives (valence, activation, and dominance) from speech signals. Several machine learning and feature selection algorithms were explored for automatic emotion recognition from speech.

2011, 54 pages

Anahtar kelimeler: Emotional speech databases, acoustic features of emotional speech, emotion recognition, emotion primitives estimation

SİMGELER VE KISALTMALAR DİZİNİ

F_i	F-Ölçütü
κ	Kappa Katsayısı
p_i	Keskinlik Oranı
r_i	Doğruluk Oranı
A	Aktivasyon (Activation) Düzlemi
B	Baskınlık (Dominance) Düzlemi
CC	İlinti Katsayısı
D	Değerlik (Valence) Düzlemi
DVM	Destek Vektör Makinaları
DVR	Destek Vektör Regresyonu
MAE	Ortalama Mutlak Hata
MFKK	Mel-Frekansı Spektral Katsayıları
RMSE	Ortalama Karesel Hatanın Karekökü
UA	Ağırlıksız Toplam Doğruluk
WA	Ağırlıklı Toplam Doğruluk

ÇİZELGELER DİZİNİ

	Sayfa
Çizelge 3.1. Ses işaretlerinin toplam sürelerinin cinsiyete göre dağılımları.....	9
Çizelge 3.2. İkili duygu kategorilerinde yapılan kabuller.....	16
Çizelge 3.3. Örnek bir sınıf dağılımı.....	23
Çizelge 3.4. A ve B örnek sınıflarına ilişkin değerlendirme sonuçları	24
Çizelge 4.1. Konuşma sinyallerinin duygu kategorilerine göre dağılımları (Kategorik Değerlendirme-1).....	27
Çizelge 4.2. Konuşma sinyallerinin duygu kategorilerine göre dağılımları (Kategorik Değerlendirme-2).....	27
Çizelge 4.3. Kappa katsayıları	28
Çizelge 4.4. Değerlendirme testini yapan değerlendiriciler arasındaki ortalama standart sapma ve ortalama korelasyon sonuçları. D: değerlik, A: aktivasyon, B: baskınlık.....	29
Çizelge 4.5. Duygu sınıflarının değerlik, aktivasyon ve baskınlık boyutlarındaki ortalama ve standart sapma değerleri.....	31
Çizelge 4.6. Duygu sınıflarının değerlik, aktivasyon ve baskınlık boyutlarındaki ortalama ve standart sapma değerleri.....	32
Çizelge 4.7. Farklı duygu kategorileri için ortalama ve standart sapma (std) F0 değerleri	34
Çizelge 4.8. Erkek ünlü seslerin ortalama F0 değerlerinin çoklu karşılaştırma sonuçları	36
Çizelge 4.9. Kadın ünlü seslerin ortalama F0 frekans değerlerinin çoklu karşılaştırma sonuçları.....	36
Çizelge 4.10. Erkek ve kadın konuşmacılarda ünlü harflerin duygu kategorilerine göre ortalama formant frekansları ve standart sapmaları.....	39
Çizelge 4.11. Erkek ünlü seslerin ortalama F1 formant frekans değerlerinin çoklu karşılaştırma sonuçları	40
Çizelge 4.12. Kadın ünlü seslerin ortalama F1 formant frekans değerlerinin çoklu karşılaştırma sonuçları	40
Çizelge 4.13. Erkek ünlü seslerin ortalama F2 formant frekans değerlerinin çoklu karşılaştırma sonuçları	40
Çizelge 4.14. Kadın ünlü seslerin ortalama F2 formant frekans değerlerinin çoklu karşılaştırma sonuçları	41
Çizelge 4.15. Erkek ünlü seslerin ortalama F3 formant frekans değerlerinin çoklu karşılaştırma sonuçları	41
Çizelge 4.16. Kadın ünlü seslerin ortalama F3 formant frekans değerlerinin çoklu karşılaştırma sonuçları	41
Çizelge 4.17. Farklı duygu kategorileri için ünlü seslerin sürelerinin ortalama ve standart sapma (std) değerleri	42

Çizelge 4.18. Ünlü harflerin duygusal kategorilere göre ortalama süreleri	43
Çizelge 4.19. Erkek ünlü seslerin ortalama sürelerinin çoklu karşılaştırma sonuçları ...	43
Çizelge 4.20. Kadın ünlü sürelerinin çoklu karşılaştırma sonuçları	44
Çizelge 4.21. CFS yöntemiyle 10-kat çapraz doğrulama işlemi yapılarak gerçekleştirilen öznitelik seçimi sonuçları	46
Çizelge 4.22. Relief-F ve Infogain öznitelik değerlendirme yöntemleriyle seçilen en iyi ilk 20 öznitelik (Sıralama en iyiden başlanarak yapılmıştır)	47
Çizelge 4.23. Negatif-Negatif Değil sınıflandırma için toplam doğruluk sonuçları. UA: ağırlıksız toplam doğruluk, WA: ağırlıklı toplam doğruluk	47
Çizelge 4.24. Negatif-Negatif Değil sınıflandırma problemi her bir sınıf için doğruluk oranı, keskinlik, duyarlılık ve F-ölçütü sonuçları	48
Çizelge 4.25. Kızgın-Kızgın Değil sınıflandırma için toplam doğruluk (overall accuracy) sonuçları. UA: ağırlıksız toplam doğruluk, WA: ağırlıklı toplam doğruluk	48
Çizelge 4.26. Kızgın-Kızgın Değil sınıflandırma problemi her bir sınıf için doğruluk oranı, keskinlik, duyarlılık ve F-ölçütü sonuçları	49
Çizelge 4.27. Destek vektör regresyon sonuçları	50

ŞEKİLLER DİZİNİ

	Sayfa
Şekil 3.1 Türkçe duygusal konuşma veritabanı oluşturulma aşamaları.....	8
Şekil 3.2 Web tabanlı kategorik değerlendirme sistemi.....	11
Şekil 3.3 Kategorik değerlendirmeler için C# yazılımı	12
Şekil 3.4 Web tabanlı 3-boyutlu değerlendirme sistemi	12
Şekil 3.5 3-boyutlu duygu uzayı değerlendirmesi için C# yazılımı.....	13
Şekil 3.6. Akustik özelliklerin analizine ilişkin blok diyagramı	14
Şekil 3.7 Duygu boyutlarını kestirim modeli.....	17
Şekil 3.8 Duygu tanıma sistemi için blok diyagramı	17
Şekil 4.1 3-Boyutlu duygu uzayında değerlendirme (1. satır değerlik, 2. satır aktivasyon, 3. satır baskınlık).....	29
Şekil 4.2 Temel duygu kategorileri (Nötr, Üzgün, Mutlu, Kızgın) için değerlik, aktivasyon ve baskınlık sonuçlarının dağılımları	30
Şekil 4.3 Mutlu (*, yeşil), Üzgün (o, mavi), Kızgın (x, kırmızı) ve Nötr (+, sarı) duygu sınıflarının duygu uzayındaki dağılımları (Kategorik değerlendirme-1 için)....	31
Şekil 4.4 Kategorik değerlendirme-2 (Pozitif, Nötr, Negatif) için değerlik, aktivasyon ve baskınlık sonuçlarının dağılımları.....	31
Şekil 4.5 Pozitif (*, yeşil), Nötr (+, sarı) ve Negatif (x, kırmızı) duygu sınıflarının duygu uzayındaki dağılımları (Kategorik değerlendirme-2 için)	32
Şekil 4.6 Ses işaretinin WaveSurfer programındaki görünüşü	33
Şekil 4.7 Duygu kategorilerinin F0 dağılımlarının kutu grafiği olarak gösterimleri (a) Erkek (b) Kadın.....	35
Şekil 4.8 Ünlü harflerin ortalama temel frekansları, (a) erkek (b) kadın.....	35
Şekil 4.9 Erkek konuşmacılara ((a), (c) ve (e) grafikleri) ve kadın konuşmacılara ((b), (d) ve (f) grafikleri) ait ünlü harflerin duygu kategorilerine göre ortalama formant frekans grafikleri	38
Şekil 4.10 Ünlü harflerin duygusal kategorilere göre ortalama süreleri, (a) erkek konuşmacılar, (b) kadın konuşmacılar	43

1. GİRİŞ

Teknolojik alanda yapılan atılımlar sayesinde kurum ve şirketlerin sunduğu interaktif hizmetler günlük hayatımızın bir parçası haline gelmiştir. Günümüzde şirketler konuşma tanıma ve sentezleme teknolojilerini interaktif sistemlerine entegre ederek müşteri memnuniyetini arttırmayı hedeflemektedirler. Bu gelişmelere paralel olarak interaktif sesli yanıt sistemleri gibi bilgisayar-insan iletişim uygulamalarını kullanıcıların ihtiyaçlarına daha doğal ve etkin şekilde cevap verebilecek sistemler haline getirmek yolunda duygu tanıma ve sentezleme sistemleri geliştirilmesi için çalışmalar yoğunlaşmıştır. Literatürde Türkçe konuşma sinyalinden duygu tanıma üzerine yapılan çalışmalar sınırlı düzeydedir. TÜBİTAK Ulusal Genç Araştırmacı Kariyer Geliştirme Programı tarafından desteklenen 109E243 nolu ve “*Türkçe Konuşma İçin Gerçek Zamanlı Duygu Tanıma Sistemi Tasarımı*” isimli proje kapsamında hazırlanan bu tez çalışmasında Türkçe konuşma sinyalinden duygu tanıma sistemi tasarımı ele alınmıştır.

Tez kapsamında konuşma sinyalinden duygu tanıma sistemi tasarımı üç aşamada gerçekleştirilmiştir. İlk aşamada Türkçe duygusal konuşma sinyalleri elde edilip duygusal içerikleri belirlenerek Türkçe duygusal konuşma veritabanı oluşturulmuştur. Duygu tanıma sisteminin başarısı herhangi bir örüntü tanıma probleminde olduğu gibi, sistemin modellemesinde kullanılan veri setine bağlıdır. Tanıma sisteminin duyguların akustik modellerini ayırt etmeyi öğrenebilmesi için farklı duygu içeriklerine sahip çok sayıda konuşma sinyalinin ihtiyaç vardır (Douglas-Cowie ve ark., 2003). Buradaki en büyük zorluk, çok sayıda konuşma sinyalinin elde edilerek duygu içeriklerinin belirlenmesidir. Yapılan tez kapsamında 55 adet Türkçe filmde 5304 tane konuşma sinyali elde edilmiş ve çok sayıda kişi tarafından dinleme testleri yapılarak duygu içerikleri belirlenmiştir. Konuşma sinyallerinin duygu içerikleri belirlenirken hem kategorik sınıflandırma hem de 3-boyutlu duygu uzayı yaklaşımları kullanılmıştır.

Tezin ikinci aşamasında akustik özniteliklerin çıkarımı ve seçimi gerçekleştirilmiştir. Duygu tanıma çalışmalarında konuşma sinyalinden elde edilecek özniteliklerin çıkarımı ve seçimi oluşturulacak olan modelin performansı açısından oldukça önemlidir. Akustik özniteliklerin elde edilmesinde temel frekans (F0), enerji, sıfır geçiş oranı, sesli olma olasılığı ve Mel-frekansı kepsral katsayıları (MFKK) gibi konuşma sinyalinin prosodik ve spektral karakteristiklerinden yararlanılmıştır. Bu akustik parametrelerden sözlü

ifade düzeyinde çıkarılan istatistikler akustik öznitelik olarak kullanılmıştır. Çıkarılan akustik özniteliklerin hepsi duyguları ayırt etme açısından uygun olmayabilir. Bu nedenle sınıflandırma performansını artırmak ve aynı zamanda hesaplama karmaşıklığını ve süresini azaltmak için farklı öznitelik seçimi algoritmaları uygulanarak duygu tanıma problemine en uygun öznitelikler seçilmeye çalışılmıştır.

Son aşamada ise farklı makine öğrenme algoritmaları, konuşma sinyalinden duygu tanıma probleminin çözümü için kullanılmış ve performansları ayrıntılı olarak karşılaştırılmıştır. Çalışma kapsamında konuşma sinyallerinin duygu tanıma problemi hem kategorik sınıflandırma hem de duygu boyutları kestirimi olarak ele alınmıştır.

Ayrıca tez kapsamında duygunun Türkçedeki seslerin akustik özellikleri üzerine etkileri istatistiksel analizler yardımıyla incelenmiştir. Yapılan analizlerle duyguya ilişkin bilginin konuşma işareti içerisinde ne şekilde kodlandığı bilgisinin ayrıntılı analiziyle duygu tanıma çalışmalarının yanı sıra duygu sentezleme çalışmalarında kullanılmak üzere önemli bilgiler elde etmemize yardımcı olacaktır.

Tüm bu çalışmalar ile gelecekte gerçek zamanlı duygu tanıma sistemi tasarımı konusunda önemli bilgiler elde edilmesi hedeflenmiştir.

2. ÖNCEKİ ÇALIŞMALAR

2.1. Giriş

Bu çalışma kapsamında ele aldığımız konulardan ilki Türkçe Duygusal Konuşma Veritabanının oluşturulması üzerinedir. Literatürde Türkçe duygusal konuşma üzerine oluşturulmuş kapsamlı bir veritabanı bulunmamaktadır. Duygusal konuşma veritabanı oluşturulurken doğallık, kapsam ve içerik konularına dikkat etmek gerekmektedir. Veritabanının hem araştırmada hem de tanıma sistemini geliştirmede kullanılabilmesi için olabildiğince doğal, kapsamda geniş ve gerçek durumları yansıtır olmalıdır. Oluşturulacak veritabanı hedef dilin karakteristik özelliklerini kapsamalı, konuşmacı sayısı konuşmacıdan bağımsız duygu tanıma sistemi oluşturmak için yeterli sayıda olmalı ve sistemin kullanım yerine uygun duygu sınıflarını içerisinde barındırmalıdır.

2.2. Literatürdeki Veritabanları

Son yıllarda duygusal konuşma kayıtları toplanmasına yönelik yapılan çalışmalar büyük ölçüde artmıştır. Mevcut veritabanlarının kapsamlı değerlendirmelerine (Douglas-Cowie ve ark., 2003; Ververidis ve Kotropoulos, 2003; Schuller ve ark., 2009a) ve HUMAINE topluluğunun çalışmalarında (Humaine, 2011) yer verilmiştir. Bu veritabanlarının bir kısmını profesyonel aktörlerin önceden belirlenmiş ve metinsel içerik olarak genelde duygu içermeyen belirli sayıdaki cümleleri belirli duygu sınıflarında (*mutlu, şaşkın, üzgün, kızgın, korku, nötr vb.*) seslendirmeleriyle toplanan ses kayıtları oluşturmaktadır (Ververidis ve Kotropoulos, 2003). Bu tip veritabanlarına Berlin Duygusal Konuşma Veritabanı (Burkhardt ve ark., 2005) ve Danimarkaca Duygusal Konuşma Veritabanı (Engbert ve Hansen, 2007) örnek olarak verilebilir.

Konuşma sinyalinde verilmeye çalışılan duygunun gerçekliği ve doğallığının kabul edilebilir bir seviyede olması önemlidir. Son yıllarda yapılan çalışmalarda daha doğal veriler elde etmek için izlenen yollar arasında bir senaryo çerçevesinde duyguyu tetikleyecek ortamların yaratılarak konuşma kayıtlarının toplanması da bulunmaktadır. Bu tür veritabanlarına örnek olarak Sensitive Artificial Listener (SAL) (Douglas-Cowie ve ark., 2007), Airplane Behaviour Corpus (ABC) (Schuller ve ark., 2007), Speech Under Simulated and Actual Stress (SUSAS) (Hansen ve Bou-Ghazale., 1997), Audiovisual Interest Corpus (AVIC) (Schuller ve ark., 2009b), SmartKom (Steininger

ve ark., 2002) ve FAU AIBO (Batliner ve ark., 2006) verilebilir. Örneğin, FAU AIBO (Batliner ve ark., 2006) veritabanı 51 çocuğun Sony AIBO robotla konuşmalarının WoZ tekniğiyle kayıt edilmesi ve daha *sonra neşeli, şaşkın, çaresiz, tedirgin, öfkeli, sıkkın ve nötr* gibi duygu sınıflarında etiketlenilmesiyle oluşturulmuş bir veritabanıdır.

Veritabanı oluşturmada izlenen yollardan biri de acil merkez arama kayıtları, müşteri hizmetleri arama kayıtları, terapi seans kayıtları gibi doğal konuşma kayıtlarının kullanılmasıdır. Vidrascu ve Devillers (2005), France ve ark. (2003), Capital Bank Service and Stock Exchange çağrı Merkezi ses kayıtları (Devillers ve Vasilescu, 2004) ve Geneva Airport Lost Luggage çağrı merkezi ses kayıtları (Scherer ve Ceschi, 2000) veritabanlarını buna örnek olarak verebiliriz.

Bir diğer duygusal veritabanı oluşturma yöntemi ise filmler, TV programları ve dizilerden alınmış görüntü ve ses kayıtlarının değerlendirilmesidir. Alman TV talk show programı Vera-Am-Mittag (VAM) (Grimm ve ark., 2008) kayıtları kullanılarak oluşturulmuş duygusal veritabanı bu yöntemle örnek olarak verilebilir. Bu veritabanı TV programına katılan 47 davetliden kayıt edilmiş 946 konuşma sinyalinde oluşmaktadır. Bu veritabanında konuşma sinyalleri 3-boyutlu duygu düzleminde Self-Assessment Manikin (SAM) (Bradley ve Lang, 1994) kullanılarak değerlendirmeye tabi tutulmuştur. Polzin ve Waibel (2000) çalışmasında filmlerden çıkardığı 1586 sinirli, 1076 üzgün ve 2991 nötr konuşma sinyalini kullanmıştır. Belfast Naturalistic Veritabanı (Douglas-Cowie ve ark., 2000) televizyon sohbet programları ve röportajlardan oluşturulmuş geniş duygu içeriğine sahip bir veritabanı olarak bu türe girmektedir. Korku tipi duyguların toplandığı veritabanı SAFE (Situation Analysis in Fictional and Emotional Corpus) (Clavel ve ark., 2006) 30 adet korku, drama, psikolojik ve gerilim türü filmlerden çıkarılan yaklaşık 7 saatlik konuşma kaydını kapsamaktadır.

Tez çalışmasında oluşturulan veritabanında orijinal dili Türkçe olan dizi ve filmlerden yararlanılmıştır. Konuşma işaretlerinin doğal ortamdan alınmayıp profesyonel aktör olarak nitelendirebilecek konuşmacıların seslendirdiği film ve dizilerden elde edilmesi sebebiyle oluşturulan veritabanını yarı-doğal/simule bir veritabanı olarak nitelendirebiliriz.

Veritabanı kapsam olarak ele alındığında içerdiği duygu kategorileri bakımından da literatürdeki örneklerinin incelenmesi gerekmektedir. Oluşturulan veritabanı kategorik olarak *Mutlu, Üzgün, Kızgın, Nötr, Korku, Şaşkın* ve *Diğer* olmak üzere 7 farklı duygu

kategorisinin yanı sıra üst seviye duygu kategorilerinden *Pozitif*, *Nötr* ve *Negatif* ile de değerlendiriciler tarafından sınıflandırılmıştır. Veritabanında bu kategorik sınıflandırmalara ek olarak duygu boyutu düzeyinde bir sınıflandırma ile de değerlendiriciler tarafından değerlendirilmiştir. Literatürde Vera-Am-Mittag veritabanında bulunan konuşma işaretleri değerlendiriciler tarafından duygu boyutu düzeyinde sınıflandırılarak, Berlin Duygusal Konuşma Veritabanı (Burkhardt ve ark., 2005) temel duygu kategori (Kızgın, Nötr, Korku, Sıkın, Mutlu, Üzgün, İğrenme) ile sınıflandırılmıştır.

Tez çalışmasında ele alınan konulardan bir diğeri de duygusal konuşma işaretlerinin akustik analizidir. Literatürde Meral ve ark. (2006) yapmış oldukları çalışmada oluşturmuş oldukları veritabanında bulunan konuşma işaretlerinin duygu kategorilerine göre temel sıklık eğrilerini incelenmişlerdir. Tez çalışmamızda konuşma işaretlerimizin duygu kategorilerinin ünlü/sesli harflere göre temel frekans, formant frekansları ve sürelerle ilişkin değişimleri incelenmiştir

Tez çalışmasında ele alınan son konu ise duygu kategorileri tanınması ve boyutlarının kestirimi üzerinedir. Literatürde yapılan çalışmalar incelendiğinde kategorik sınıflandırma üzerine yapılan çalışmaların oldukça fazla olduğu görülmektedir (Lee ve Narayanan, 2005; Schuller ve ark., 2010, Yıldırım ve ark., 2011). İkili kategorik sınıflandırma çalışmaları üzerine Lee ve Narayanan (2005) yaptıkları çalışmada havayolları müşteri hizmetleri konuşma kayıtlarını kullanarak *negatif- negatif değil* sınıflandırma çalışması yapılmış ve %80 başarı elde etmişlerdir. Çalışmalarında akustik bilginin yanı sıra dilbilimsel ve diyalog bilgisini de kullanmışlardır. Clavel ve ark. (2008) gözetleme sistemleri için *korku-tipi* duygu tanıma üzerine çalışma yapmışlar ve %70 başarı elde etmişlerdir. Polzehl ve ark. (2011) interaktif sesli sistemlere yönelik *kızgınlık* tanıma çalışması yapmışlardır.

Diğer bir yöntem olan duygu boyutlarının kestirimi üzerindeki çalışmalar, psikoloji bilimindeki gelişmelerle birlikte dikkat çekici seviyede artmıştır (Grimm ve ark., 2007; Wu ve ark., 2010; Espinosa ve ark., 2010). Tez çalışmasında ikili duygu kategorileri ile duygu boyutları kestirimi üzerine çalışmalar yapılmıştır.

2.3. Yapılan Çalışmanın Kapsamı

Bu tez çalışması TÜBİTAK Ulusal Genç Araştırmacı Kariyer Geliştirme Programı tarafından desteklenen 109E243 nolu ve “Türkçe Konuşma İçin Gerçek Zamanlı Duygu Tanıma Sistemi Tasarımı” isimli proje kapsamında hazırlanmış olup, tezin içeriği aşağıdaki şekilde özetlenmiştir.

Tezin birinci bölümünde konuya giriş yapılmış, tezin amacı ve önemi hakkında bilgiler verilmiştir.

Tezin ikinci bölümünde önceden yapılmış çalışmalar hakkında bilgiler verilmiş ve tez kapsamı konusunda değerlendirme yapılmıştır.

Tezin üçüncü bölümünde materyal ve yöntemler hakkındaki bilgilere yer verilmiştir. Materyal bölümünde Türkçe duygusal konuşma veritabanının kapsamı ve oluşturulması konuları ele alınmıştır. Yöntem bölümünde ise veritabanının duygusal içeriklerinin belirlenme yöntemleri, akustik özelliklerin değerlendirilmesi için kullanılan yöntemler ve duygu tanıma ve duygu boyutlarının kestiriminde öznitelik seçimi ve sınıflandırma yöntemlerine değinilmiştir.

Tezin dördüncü bölümünde duygusal konuşma veritabanı için duygusal içeriklerin belirlenmesi ve kategorik sınıflar ile 3-boyutlu duygu uzayı değerlendirmelerinin birlikte analizi, Türkçe duygusal konuşmanın akustik analizi ve Türkçe konuşmadan duygu tanıma ve duygu boyutlarının kestirimi üzerine yapılan çalışmalara ilişkin sonuçlar verilmiştir.

Tezin son bölümünde ise yapılan çalışmalara ilişkin genel sonuçlar verilmiş ve öneriler sunulmuştur.

3. MATERYAL VE YÖNTEM

3.1. Materyal

Bu çalışmada orijinal dili Türkçe olan film ve dizilerden çıkarılan konuşma diyaloglarından yararlanılmıştır. Filmlerin komedi, dram, korku, gerilim, macera, romantik ve aksiyon gibi çeşitli türlerden olmasına dikkat edilerek farklı duygu sınıflarında konuşma sinyalleri elde edilmeye çalışılmıştır.

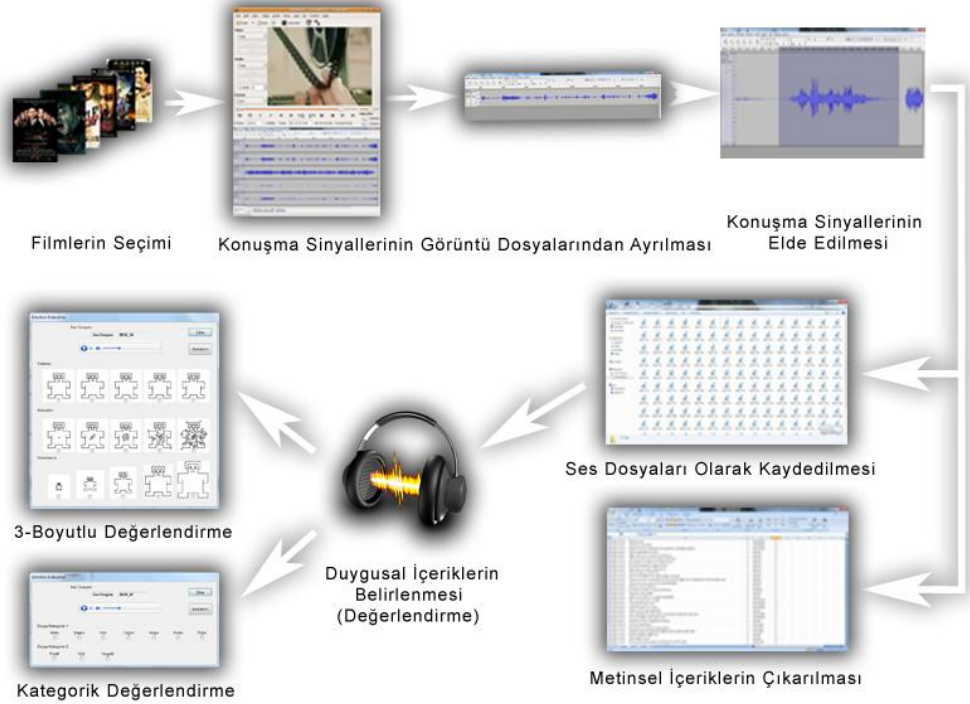
İlk olarak filmlere ait video dosyalarında bulunan ses kanalları görüntü kanalından ayrılarak farklı dosyalara kayıt edilmiştir. Daha sonra altı kanaldan oluşan AC-3 formatındaki ses dosyasından konuşma bilgisinin bulunduğu kanal 48 kHz örnekleme frekansı ve 16 bit çözünürlükte diğer kanallardan ayrılarak yeni bir ses dosyası olarak kaydedilmiştir. Bir sonraki aşamada ses dosyasında bulunan konuşma diyalogları cümle cümle ayrılarak ayrı ses dosyaları olarak kaydedilmiştir. Bu işlem sırasında konuşma sinyallerine ait metinsel içerikler ve konuşmacıların cinsiyet bilgileri de daha sonraki çalışmalarda kullanılmak üzere kayıt altına alınmıştır. Sonuç olarak 55 tane filmde 5304 adet konuşma dosyası veritabanına eklenmiştir. Bu ses dosyalarından 3699 tanesi erkek konuşmacılara ve 1605 tanesi de bayan konuşmacılara aittir.

3.1.1. Türkçe Duygusal Konuşma Veritabanının Oluşturulması

Literatürde yapılan çalışmalar incelendiğinde duygusal içerikli konuşmaların toplanması için 4 farklı seçeneğin olduğunu söylenebilir. Bu seçenekleri;

1. Profesyonel aktör ve aktrislerin önceden belirlenmiş ve içerik olarak genelde duygu içermeyen belirli sayıdaki cümlelerin belirli duygu sınıflarında seslendirmeleriyle toplanan ses kayıtlarının kullanımı
2. Bir senaryo çerçevesinde duyguyu tetikleyecek ortamların oluşturularak konuşmaların toplanması, WoZ yöntemi kullanılarak toplanan insan-makine iletişim kayıtlarının kullanımı
3. Acil merkez arama kayıtları, müşteri hizmetleri arama kayıtları, terapi seans kayıtları gibi doğal konuşma kayıtlarının kullanımı
4. Filmler, TV tartışma programları, röportaj ve dizilerin görüntü ve ses kayıtlarının kullanımı olarak sıralanabilir.

Toplanan verilerin kabul edilebilir gerçeklik/doğallık seviyesinde olması tanıma sisteminin başarısı için önemlidir. Gerçek hayat ortamında veri toplamadaki zorluklar göz önüne alındığında filmler, TV tartışma programları, röportajlar ve dizlerden elde edilecek konuşma kayıtlarının hedeflenen doğallık düzeyinde ve yine hedeflenen duygu sınıfları için yeteri sayıda konuşma sinyali elde etmek için uygun bir veri toplama yöntemi olduğu düşünülmektedir. Yapılan çalışmada Türkçe olarak seslendirilmiş olan film ve dizilerinden çıkarılmış olan konuşma diyaloglarından yararlanılmıştır. Filmlerin *Komedi*, *Dram*, *Korku*, *Gerilim*, *Macera*, *Romantik* ve *Aksiyon* gibi çeşitli türlerden olmasına dikkat ederek farklı duygu sınıflarında konuşma sinyalleri elde edilmiştir (Oflazoglu ve Yıldırım, 2011).



Şekil 3.1 Türkçe duygusal konuşma veritabanı oluşturulma aşamaları

Filmlerin tamamı orijinal DVD olarak temin edilerek işlenmiştir. Seçilen DVD'lerin Dolby Digital 5.1 (AC-3) ses formatını içeriyor olmasına özellikle dikkat edilmiştir. Dolby Digital 5.1 (AC-3) ses formatının en önemli özelliği en az 5 adet ses kanalını

içermesidir. Bu formatta konuşma işaretlerinin, ortam sesleri, arka fon müzikleri vb. seslerden farklı bir kanalda bulunuyor olması konuşma işaretlerinin ayrıştırılmasında kolaylık sağlamaktadır.

İlk olarak DVD'lerde bölümlerle ayrılmış halde bulunan film sahneleri bütünlüğü sağlamak amacıyla VOB (Video **O**bject) dosyalarına dönüştürülerek kaydedilirler. VOB dosya formatı görüntü işareti ile birlikte mevcut olan tüm ses kanallarını da içinde barındırmaktadır. Elde edilen VOB uzantılı video dosyalarında bulunan ses kanalları görüntü kanalından ayrılarak AC-3 formatında kaydedilir. Son olarak altı kanaldan oluşan AC-3 formatındaki ses dosyasından konuşma bilgisinin bulunduğu kanal 48 kHz örnekleme frekansı ve 16 bit çözünürlükte diğer kanallardan ayrılarak yeni bir ses dosyası olarak kaydedilir.

Filmlere ait ses dosyaları dinlenerek konuşma diyalogları cümle cümle ayrı ses dosyaları olarak kaydedilir. Bu işlem sırasında sonraki çalışmalarda kullanılmak üzere dosya isimleri, metinsel içerikleri ve konuşmacıların cinsiyetleri başka bir dosyaya kaydedilmektedir. Şiveli, ortam gürültüsünün fazla olduğu ve söylenen metnin anlaşılmasında zorluk çekilen konuşma kayıtlarının bilgileri de dosyada belirtilmektedir. Film ses dosyalarından konuşma diyaloglarının elde edilmesi aşaması veri toplamanın en zorlu ve zaman alıcı kısımlarından birisini oluşturmaktadır.

Sonuç olarak 55 adet film ve diziye ilişkin toplam 5304 adet konuşma ses dosyası veritabanına eklenmiştir. Bu ses dosyalarının 3699 tanesi erkek, 1605 tanesi ise bayan konuşmacılara aittir. Konuşma dosyalarının toplam sürelerinin cinsiyetlere göre dağılımı Çizelge 3.1.'de verilmiştir. Elde edilen konuşma işaretlerinin ortalama süresi 2.34 saniyedir.

Çizelge 3.1. Ses işaretlerinin toplam sürelerinin cinsiyete göre dağılımları

Cinsiyet	Toplam Konuşma Sinyali Süreleri (dk)
Erkek	138.06
Bayan	59.73
Toplam	197.79

3.2. Yöntem

3.2.1. Veritabanının Duygu İçeriklerinin Belirlenmesi

Literatürde konuşma sinyallerinin duygusal içeriklerinin belirlenmesinde iki farklı yaklaşım ön plana çıkmaktadır. Bunlardan en yaygın ve klasik olanı, konuşma sinyallerindeki duygu içeriğinin ayrık kategorilere (üzgün, kızgın, mutlu vb.) sınıflandırılmasıdır. Literatürde bulunan birçok çalışmada bu klasik yaklaşım kullanılmıştır (Burkhardt ve ark., 2005; Engbert ve ark., 2007). Tez çalışması kapsamında, literatürdeki çalışmalar da göz önüne alınarak Türkçe film ve dizilerden elde edilen konuşma kayıtlarının mutlu, şaşkın, üzgün, kızgın, korku, nötr ve diğer olmak üzere 7 kategoride etiketlenmesi yapılmıştır. Diğer seçeneği duygu sınıflarından herhangi birisine girmeyen konuşma işaretleri için seçenek olarak verilmiştir. Konuşma sinyallerinin aynı zamanda daha üst seviye duygu tanımları olan *pozitif*, *negatif* ve *nötr* kategorilerinde de etiketlenmesi yapılmıştır.

İkinci yaklaşım ise duygunun çok boyutlu duygu uzayında bir nokta olduğu varsayımından yola çıkarak konuşma sinyalinin duygusal içeriğinin sayılarla ifade edilmesine dayanmaktadır. Kategorik sınıflandırmada konuşma sinyallerinin hangi duygu sınıfına ait olduğu öznel bir konudur. Bazı konuşma sinyalleri herhangi bir duygu kategorisine girmeyebilir, ya da konuşma sinyali birden fazla duyguyu içerisinde barındırabilir. Son yıllarda psikoloji biliminde duygu üzerine yapılan çalışmalar duyguların çok boyutlu bir uzayda birer nokta veya bölge olarak tanımlanabileceğini göstermiştir (Schlosberg, 1954; Russell, 2003). Çok boyutlu değerlendirme, kategorik sınıflandırma yaklaşımının aksine daha fazla genellenebilirliğe sahiptir. Kategorik çalışmalara göre daha az sayıda çalışma olmasına rağmen son yıllarda çok boyutlu yaklaşım kullanılarak yapılan çalışmalar dikkat çekici seviyede artmıştır. Değerlik (valence) ve aktivasyon (activation) boyutlarının çok geniş bir aralıkta duygu çeşitliliğini kapsadığı bilindiği için en yaygın olarak kullanılan duygu düzlemleridir. Değerlik, pozitiften negatife doğru bir seviyeyi temsil ederken aktivasyon düşükten (sakin) yükseğe (heyecanlı) uyarım seviyesini göstermektedir (Whissel, 1989). Baskınlık (dominance) ise konuşmacının kuvvetini zayıftan güçlüye değerlendiren bir boyuttur.

Yapılan çalışmada veritabanımız hem kategorik hem de çok boyutlu duygu uzayı ile değerlendirilmiştir. Veritabanının oldukça fazla sayıda konuşma sinyali içermesi sebebiyle değerlendirme işlemleri tek oturumda gerçekleştirilememiş ve bu nedenle de yardımcı yazılımlara ihtiyaç duyulmuştur. Bu yazılımlardan ilki web tabanlı olup, değerlendiricilerin internet üzerinden web sayfasına bağlanarak değerlendirmelerini yapmalarına imkân vermektedir. Sürekli internet bağlantısı olanağı bulunmayan değerlendiriciler için C# programı kullanılarak bir yazılım hazırlanmıştır. Bu iki sistem birbirleriyle uyumludur. Bu şekilde değerlendiricilerin istedikleri platformlarda dinleme ve değerlendirme yapmaları sağlanmıştır.

Değerlendirme işlemi iki farklı adımda gerçekleştirilmiştir. İlk adımda değerlendirici iki farklı kategorik sınıflandırma formunu kullanarak *mutlu*, *şaşkın*, *üzüntü*, *kızgın*, *korku*, *nötr* ve *diğer* 7 farklı duygu kategorisinde ve daha üst seviye duygu tanımları olan *pozitif*, *negatif* ve *nötr* kategorilerinde değerlendirilmiştir.

Ses Dosyasının Adı: MO11-55

Duygu Kategorisi-1:



Mutlu
 Şaşkın
 Nötr
 Üzgün
 Kızgın
 Korku
 Diğer

Duygu Kategorisi-2:

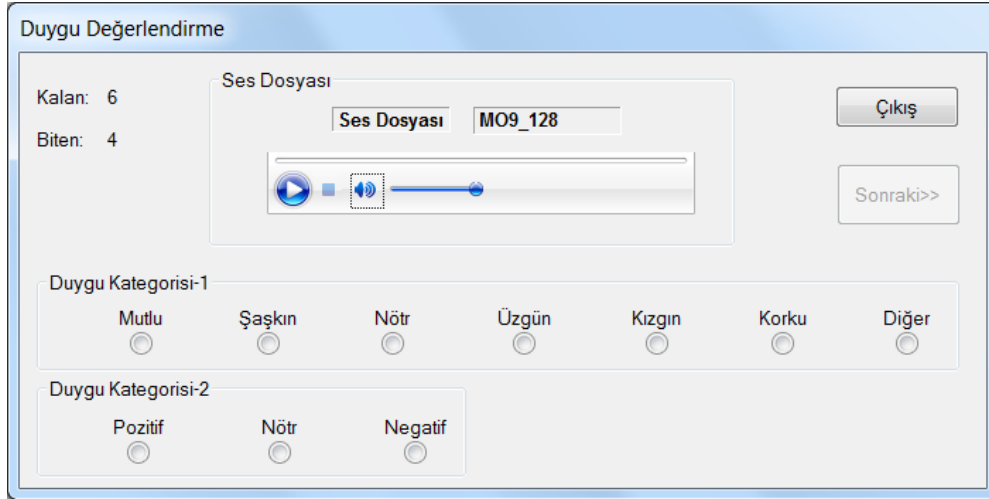
Pozitif
 Nötr
 Negatif

Onay:

Bu Çalışma Tübitak 3501 Kariyer Projesi Kapsamında Yapılmaktadır.(Proje No:109E243)

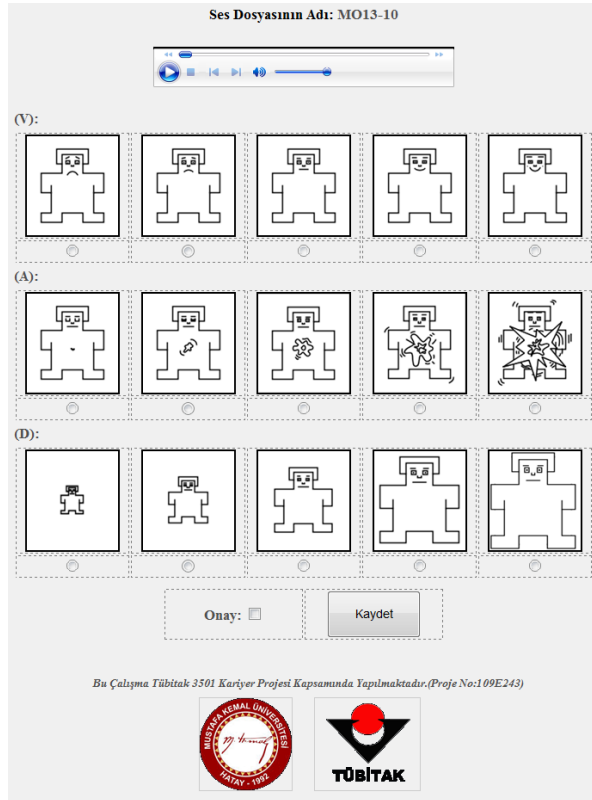



Şekil 3.2 Web tabanlı kategorik değerlendirme sistemi

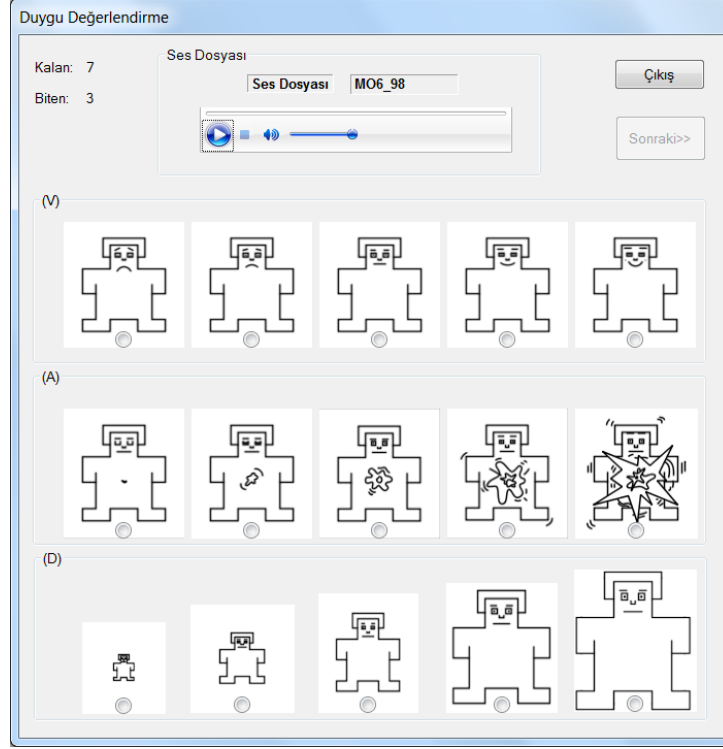


Şekil 3.3 Kategorik değerlendirmeler için C# yazılımı

Değerlendirme işleminde ikinci adım ise konuşma sinyallerinin 3-boyutlu duygu uzayında değerlendirilmesidir. Bu aşamada Bradley ve Lang'ın (1994) geliştirdiği Self-Assessment Manikin (SAM) görselleri kullanılmıştır. Önceki adıma benzer şekilde değerlendiriciler bu işlem için web tabanlı ya da C# yazılımı tercih edilebilmektedir.



Şekil 3.4 Web tabanlı 3-boyutlu değerlendirme sistemi



Şekil 3.5 3-boyutlu duygu uzayı değerlendirmesi için C# yazılımı

Konuşma sinyalinde bulunan duygunun belirlenmesi (her boyuttaki değeri) için dinleme testini yapan değerlendiricilerden elde edilen sonuçların birlikte değerlendirilmesi gerekmektedir. Birçok değerlendiriciden gelen sonuçları birleştirmek için En Büyük Olabilirlik Kestirimi (EBOK) yöntemi kullanılmıştır. Burada amaç çok sayıda dinleyiciden elde edilen değerleri kullanarak konuşma sinyalinde verilmeye çalışılan gerçek duyguyu tahmin etmektir.

Dinleme testini yapan değerlendiricilerden elde edilen sonuçlar;

$$\hat{x}_{n,k} = (\hat{x}_{n,k}^{(D)}, \hat{x}_{n,k}^{(A)}, \hat{x}_{n,k}^{(B)})^T \quad (3.1.)$$

şeklinde olsun. Burada D, A ve B sırasıyla değerlik, aktivasyon ve baskınlığı, n konuşma sinyali indeksini ($1 \leq n \leq N$), k dinleyici ($1 \leq k \leq K$) indeksini göstermektedir.

En Büyük Olabilirlik Kestirimi gerçek değer en iyi tahmini bulmak için hata kareler ortalamasını minimize eder. Burada $i \in (D, A, B)$ dir.

$$E \left\{ \left| \hat{x}_n^{(i)} - x_n^{(i)} \right|^2 \right\} \rightarrow \min \quad (3.2.)$$

Böylece bu yöntemle kestirilen değer (3.3.) denklemi kullanılarak bulunur.

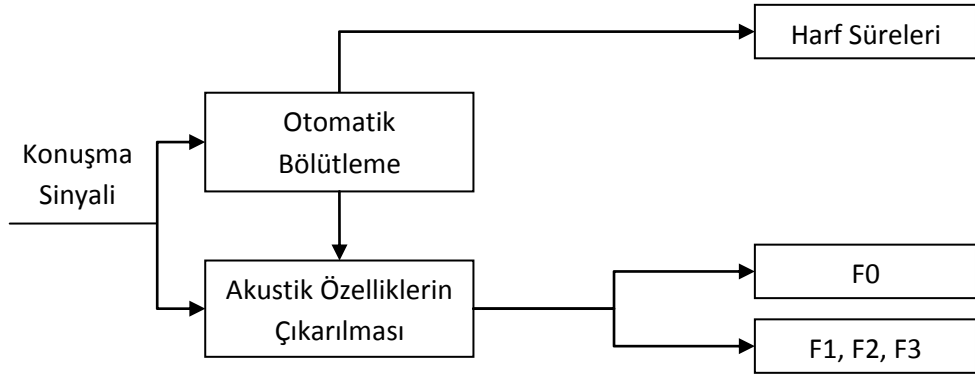
$$\hat{x}_n^{EBOK,(i)} = \frac{1}{K} \sum_{k=1}^K \hat{x}_{n,k}^{(i)} \quad (3.3.)$$

Bu da dinleyicilerden elde edilen değerlerinin ortalamasını almak demektir.

Değerlendirme testini gerçekleştirmiş olan değerlendiricilerin 3-boyutlu duygu uzayında ne kadar uyumlu olduklarını ölçmek amacıyla standart sapmaları (3.4.) denklemi kullanılarak hesaplanmıştır.

$$d_n^{(i)} = \sqrt{\frac{1}{K-1} \sum_{k=1}^K (\hat{x}_{n,k}^{(i)} - \hat{x}_n^{EBOK,(i)})^2} \quad (3.4.)$$

3.2.2. Türkçe Konuşmanın Akustik Özelliklerinin Değerlendirilmesi



Şekil 3.6. Akustik özelliklerin analizine ilişkin blok diyagramı

Yapılan çalışmada duygu kategorilerine ilişkin konuşma sinyallerinden elde edilen özellikler ile konuşma içerisinde geçen sesli/ünlü harflerin birbirleri ile olan ilişkilerini incelemek için istatistiksel yöntemlerden yararlanılmıştır. Sesli harflerin yerlerinin tespiti/bölütleme için HTK (Young ve ark., 2009) programının güdümlü-hizalama (forced-alignment) tekniği kullanılmıştır. Konuşma sinyalinden çıkarılan temel frekans

ve formant frekanslarına ilişkin özelliklerin elde edilmesi sırasında önceki adımda elde ettiğimiz sesli/ünlü harflerin sürelerinden yararlanılmıştır. Bu adımda çıkarılan özelliklerin duygu üzerindeki etkisini incelemek için ise ANOVA analizi ve çoklu karşılaştırma testleri uygulanmıştır. Yapılan çalışmalara ilişkin değerlendirmeler Bölüm 4.2.'de verilmiştir.

3.2.3. Duygu Tanıma ve Duygu Boyutlarının Kestirimi

3.2.3.1. Kategorik Duygu Tanıma

Teknolojik alanda yapılan atılımlar sayesinde kurum ve şirketlerin sunduğu sesli yanıt hizmetleri günlük hayatımızın bir parçası haline gelmiştir. Günümüzde şirketler konuşma tanıma ve sentezleme teknolojilerini interaktif çağrı merkezi sistemlerine bütünleştirerek müşteri memnuniyetini arttırmayı hedeflemektedirler. Bu gelişmelere paralel olarak interaktif sesli yanıt sistemleri gibi bilgisayar-insan iletişim uygulamalarını kullanıcıların ihtiyaçlarına daha doğal ve etkin şekilde cevap verebilecek sistemler haline getirmek yolunda duygu tanıma sistemleri geliştirilmesi için çalışmalar yoğunlaşmıştır. Özellikle kullanıcıyla otomatik sistemin etkileşimi sırasında müşterinin kızdığı, sinirlendiği durumlar için sistemin diyalog stratejisini değiştirerek duruma çözüm üretmesi mümkün olabilecektir (Metze vd., 2009). Belirli duygu sınıfı tanıma yönelik çalışmalar son yıllarda artmıştır. Literatürde akustik bilginin yanı sıra dilbilimsel ve diyalog bilgisini de kullanarak havayolları müşteri hizmetleri konuşma kayıtları ile *negatif- negatif değil* sınıflandırma çalışması yapan Lee ve Narayanan (2005), yaptıkları bu çalışmada %80 başarı elde etmişlerdir. Clavel ve ark. (2008)'nin gözetleme sistemleri için *korku-tipi* duygu tanıma üzerine yaptıkları bir çalışmada ise %70 başarı elde edilmiştir. Polzehl ve ark. (2011) interaktif sesli sistemlere yönelik *kızgınlık* tanıma çalışması yapmışlardır. Çalışmalarında akustik ve dilbilimsel bilgi kullanarak farklı veritabanları için %75 ile %79 arasında değişen doğruluk oranlarına ulaşmışlardır. Tez çalışmasında, literatürdeki çalışmalar da dikkate alınarak Türkçe duygusal veritabanı için *negatif-negatif değil* ve *kızgın-kızgın değil* ikili-sınıflandırma problemleri üzerine odaklanılmıştır.

Çizelge 3.2. İkili duygu kategorilerinde yapılan kabuller

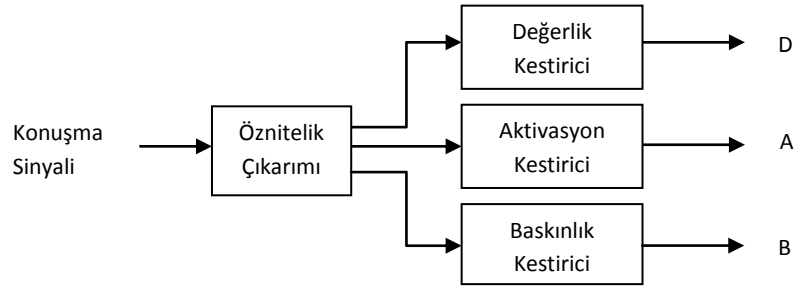
İkili Kategoriler	Kızgın Değil	Kızgın	Negatif Değil	Negatif
İçerikleri	Nötr, Üzgün, Mutlu	Kızgın	Nötr, Mutlu	Üzgün, Kızgın

Yapılan çalışmalarla ilgili değerlendirme sonuçları Bölüm 4.3.'te verilmiştir.

3.2.3.2. Duygu Boyutlarının Kestirimi

Konuşma sinyalinden duygu tanıma üzerine yapılan birçok çalışmada, duygular belirli sayıdaki duygu kategorisine sınıflandırılmıştır (Lee ve Narayanan, 2005; Schuller ve ark., 2010; Yıldırım ve ark., 2011). Son yıllarda psikoloji biliminde duygu üzerine yapılan çalışmalar duyguların çok boyutlu bir uzayda birer nokta veya bölge olarak tanımlanabileceğini göstermiştir (Russell, 2003). Çok boyutlu duygu uzayı yaklaşımı, kategorik sınıflandırmanın aksine daha fazla genellenebilirliğe sahiptir. Kategorik çalışmalara göre daha az sayıda çalışma olmasına rağmen son yıllarda çok boyutlu yaklaşım kullanılarak yapılan çalışmalar dikkat çekici seviyede artmıştır (Grimm ve ark., 2007; Wu ve ark., 2010; Espinosa ve ark., 2010). Değerlik ve aktivasyon boyutlarının çok geniş bir aralıkta duygu çeşitliliğini kapsadığı bilindiği için en yaygın olarak kullanılan duygu boyutlarıdır. *Değerlik*, pozitiften negatife doğru bir seviyeyi temsil ederken, *aktivasyon* düşükten (sakin) yükseğe (heyecanlı) uyarım seviyesini göstermektedir. *Baskınlık* ise konuşmacının hissedilen kuvvetini zayıftan güçlüye gösteren bir boyuttur.

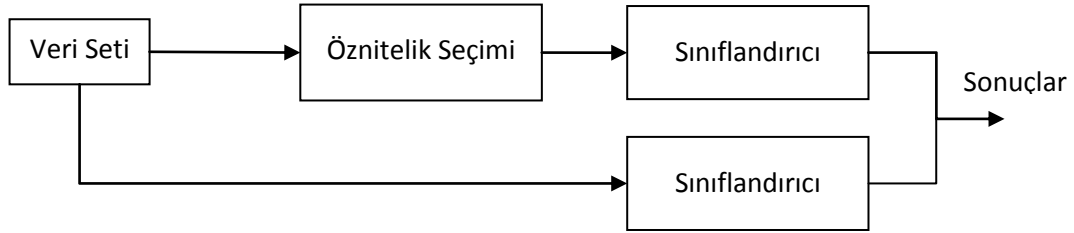
Duygu boyutlarının kestirimi üzerine yapılan çalışmalarda Şekil 3.7'de verilen model kullanılmıştır. Sisteme gelen konuşma işaretinden elde edilen öznitelikler her bir duygu boyutu için önceden modellenmiş olan kestiricilere girerek boyut değeri hesaplanmaktadır. Bu modelde her bir duygu boyutunun kestirimi için destek vektör regresyon yöntemi kullanılmıştır (Vapnik ve ark., 1997).



Şekil 3.7 Duygu boyutlarını kestirim modeli

Yapılan çalışmalarla ilgili değerlendirme sonuçları Bölüm 4.3.'de verilmiştir.

3.2.4. Duygu Tanımda ve Duygu Boyutlarının Kestirimde Öznitelik Seçimi ve Sınıflandırma Yöntemleri



Şekil 3.8 Duygu tanıma sistemi için blok diyagramı

Duygu tanıma ve duygu boyutlarının kestirimi konularında kullanılacak olan özniteliklerin tespiti ve seçilecek sınıflandırma yöntemi sistem başarısı açısından oldukça önemlidir. Kullanılacak olan sınıflandırıcının, sınıflandırılmak istenen veriyi yüksek başarıyla ayırabilmesi için seçilecek özniteliklerin sınıflar arasındaki farkları ortaya koyacak nitelikte olmalıdır. Sınıflandırılmak istenen duygu kategorilerini en iyi şekilde ayırabilen özniteliklerin seçilmesi ve bunlara uygun bir sınıflandırıcı kullanılarak model karmaşıklığı azaltılabileceği gibi aynı zamanda sınıflandırma performansı da artırılabilir. Literatürde yapılan benzer çalışmalar incelenerek öznitelik seçim yöntemlerinden; *ilinti tabanlı öznitelik seçimi (CFS)* (Hall, 1999), *bilgi kazancı öznitelik değerlendirme (InfoGain)* (Quinlan, 1986) ve *RELIEF-F öznitelik değerlendirme* (Kononenko, 1994) yöntemleri çalışmamızda tercih edilmiştir. Elde edilen öznitelik vektörlerinin sınıflandırılması işlemi için probleme uygun bir sınıflandırıcı seçilmesi gereklidir. Çalışmada, literatürde tercih edilen sınıflandırma

yöntemlerinden Naive Bayes (John ve Langley, 1995), Bayesian Network (Bouckaert, 1995) ve Destek Vektör Makineleri (Hearst, 1998) üzerinde yoğunlaşmış ve başarımları kıyaslanmıştır. Duygu boyutlarının kestirimi üzerine yapılan çalışmada ise sınıflandırıcı olarak Destek Vektör Regresyonu (Vapnik ve ark., 1997) kullanılmıştır.

3.2.4.1. Öznitelik Seçimi

Öznitelik seçimi sınıflandırma algoritmasının performansını artırmak, hesaplama karmaşıklığını ve süresini azaltmak amacıyla makine öğrenme alanında sıklıkla başvurulan bir yöntemdir. Genel olarak öznitelik seçimi yöntemlerini kapsayıcılar (wrappers) ve süzgeçler (filters) olarak iki geniş kategoriye ayırmak mümkündür. Kapsayıcılar öğrenme algoritmasının performansını maksimize eden öznitelikleri bulmaya çalışırlar. İyi performans verdikleri literatürde kanıtlandıysa da çok yavaş çalışmaktadırlar. Süzgeç yöntemleri ise kullanılacak öğrenme algoritmalarından bağımsızdır ve kapsayıcılara göre çok daha hızlı çalışmaktadırlar. Süzgeç yöntemleri verideki tutarlılık, öznitelikler arası korelasyon veya öznitelikle sınıf arasındaki ilişkiye göre seçim yapmaktadır. Bu çalışmada, *ilinti tabanlı öznitelik seçimi (CFS)* (Hall, 1999), *bilgi kazancı öznitelik değerlendirme (InfoGain)* (Quinlan, 1986) ve *RELIEF-F öznitelik değerlendirme* (Kononenko, 1994) olmak üzere üç farklı süzgeç yöntemi kullanılmıştır.

3.2.4.1.1. İlinti Tabanlı Öznitelik Seçimi

CFS (Correlation Based Feature Selection) yöntemi öznitelikleri tek tek puanlamak yerine öznitelik alt kümelerinin değerini puanlayarak öznitelik seçimi yapar (Hall, 1999). Bu yöntem “iyi bir öznitelik altkümesi sınıfla yüksek ilintili ve aynı zamanda da birbirleriyle ilintisiz öznitelikleri içerir” hipotezine dayanmaktadır. Öznitelik altkümeleri (3.5.) ile verilen sezgisel değer kullanılarak puanlanır.

$$Merit_S = \frac{k\bar{r}_{cf}}{\sqrt{k + k(k-1)\bar{r}_{ff}}} \quad (3.5.)$$

Burada $Merit_S$ k tane öznitelik içeren S öznitelik altkümelerinin sezgisel değeri (heuristic merit), \bar{r}_{cf} ortalama sınıf-öznitelik ilintisi ve \bar{r}_{ff} ise öznitelikler arası ilintidir.

CFS öznitelik-sınıf ve öznitelik-öznitelik ilintilerini simetrik belirsizlik kullanarak hesaplar ve daha sonra öznitelik altküme uzayında arama yapar. Simetrik belirsizlik öznitelikler arasındaki ve öznitelik-sınıf arasındaki ilişkililik derecesini kestirmede kullanılmaktadır.

$$\text{Simetrik Belirsizlik} = 2.0 * \left[\frac{\text{gain}}{H(C) - H(Y)} \right] \quad (3.6.)$$

$$\text{gain} = H(C) + H(Y) - H(Y, C)$$

Burada $H(Y)$ Y özneliğinin entropisini, $H(C)$ C sınıfının entropisini ve gain ise öznitelik ve sınıf arasındaki bilgi kazancını ifade etmektedir (3.6.).

CFS yöntemi öznitelik altküme uzayını tararken boş öznitelik kümesinden başlayarak ileri en iyi ilk arama sezgisel arama stratejisini kullanır. Ardışık beş tane tam olarak genişletilmiş artık iyileştirme sağlamayan alt küme, aramayı durdurma ölçütü olarak kullanılır.

3.2.4.1.2. Bilgi Kazancı Öznelik Değerlendirme Yöntemi

Bu yöntem özneliğin değerini (önemini) sınıfa göre bilgi kazancını ölçerek belirler. C sınıf önsel entropisi ile sonsal entropisi (özneliğin Y değeri biliniyorken ki entropisi) arasındaki fark bilgi kazancıdır.

$$\text{Infogain} = H(C) - H(C|Y)$$

$$\text{Infogain} = - \sum_C P(C) \log_2 P(C) - \sum_Y \left(-P(Y) \times \sum_C P(C|X) \log_2 P(C|X) \right) \quad (3.7.)$$

Burada $H(C)$ C sınıf önsel entropisini, $H(C|Y)$ C sınıf sonsal entropisini ifade etmektedir (3.7.). Bu yöntemde her bir özneliğin bilgi kazancı hesaplanır ve büyükten küçüğe sıralanır. Büyük kazanç değeri sınıfları ayırmada daha güçlü oluğu anlamına gelmektedir. Bu yöntem öznelikleri ayrı ayrı değerlendirdiğinden öznitelikler arasındaki artıklık ilişkisi hakkında bilgi vermemektedir.

3.2.4.1.3. RELIEF-F Öznitelik Değerlendirme Yöntemi

Aynı sınıfa ait bir en yakın komşu bir isabet (hit), H, farklı sınıfa ait bir en yakın komşu ise ıskı (miss), M olarak adlandırılmaktadır. Algoritmada farklı sınıflardan bir tane en yakın ıskı M bulmak yerine, algoritma her farklı sınıf için bir en yakın ıskı $M(C)$ bulmakta ve bunların tahminlerinin $W[f]$ güncellenmesi için yaptıkları katkıların ortalamalarını almaktadır.

$$W(f) = W(f) - \frac{diff(f, R, H)}{m} + \sum_{C \neq sınıf(R)} [P(C) \times diff(f, R, M(C))]/m \quad (3.8.)$$

$$diff(f, E_1, E_2) = \frac{|value(f, E_1) - value(f, E_2)|}{\max(f) - \min(f)}$$

Buradaki asıl fikir algoritmanın, özniteliklerin sınıf çiftlerini, hangi iki sınıfın birbirine en yakın olduğuna bakılmaksızın, birbirinden ayırt edebilme yeteneğini değerlendirebilmesidir (3.8.) (Kononenko, 1994).

3.2.4.2. Sınıflandırma ve Kestirim Yöntemleri

Konuşma sinyalinin ikili duygu kategorilerinin kestiriminde sınıflandırma yöntemi olarak Naive Bayes, Bayesian Network ve Destek Vektör Makineleri tercih edilmiştir.

3.2.4.2.1. Naive Bayes

Naive Bayes olasılıksal bir sınıflandırma yöntemidir. Öznitelik vektörünün boyutu çok yüksek olduğunda tercih edilen bir sınıflandırıcıdır (John ve Langley, 1995). Eğitim setinde bulunan her bir örnek, $X \in R^n$, n boyutlu uzayda tanımlı ve veri setimizde m adet sınıf bulunuyor olsun, $C_i \in R^1$, $i = 1, \dots, m$. Naive bayes sınıflandırıcı X örneğinin C_i sınıfına ait olma olasılığını bayes teoremini kullanarak hesaplar (3.9.).

$$P(C_i|X) = P(X|C_i)P(C_i) \quad (3.9.)$$

$P(C_i)$, sınıf önsel olasılığıdır. Özniteliklerin birbirinden bağımsız olduğunu kabul ederek $P(C_i|X)$, (3.10.) şekilde hesaplanır.

$$P(C_i|X) = \prod_{k=1}^n P(x_k|C_i) \quad (3.10.)$$

Yöntem karar verirken son olasılığı en büyütme (maximum a posterior, MAP) yöntemini kullanır. $P(C_i|X)$ olasılığını maksimize eden sınıf test örneğinin sınıfı olarak seçilir (3.11.).

$$\operatorname{argmax}_c P(C = c) \prod_{k=1}^n P(X_k = x_k|C = c) \quad (3.11.)$$

3.2.4.2.2. Bayes Ağı

Verinin kesin olmadığı ya da kısmen ulaşılmaz olduğu durumlarda modelleme için çok etkilidir. Yapı olarak bu ağlar bir dizi değişken, değişkenleri birbirine bağlayan grafiksel yapı (yönlü graf) ve bir dizi koşullu olasılıktan oluşur (Bouckaert, 1995). Sınıflandırma işlemi, öznelik değişkenleri adı verilen $x = x_1, \dots, x_n$ değişken seti verilmişken, sınıf değişkeni adı verilen $y = x_0$ değişkeninin sınıflandırılmasından oluşmaktadır. Bir sınıflandırıcı $h : x \rightarrow y$, x 'in bir örneğini y değerine eşleyen bir fonksiyondur. Sınıflandırıcı, (x, y) üzerindeki örneklerden oluşan D veri seti ile eğitilir. Eğitim işlemi, U üzerindeki D veri seti için uygun bir Bayes ağı bulmaktan ibarettir. $U = \{x_1, \dots, x_n\}$, $n \geq 1$ olacak şekilde U 'nun değişkenler seti olduğunu varsayalım. Bir değişkenler seti U üzerindeki Bayes ağı B bir ağ yapısı B_S olarak adlandırılır. B_S , U üzerindeki yönlü çevrimsiz çizge (Directed Acyclic Graph) ve bir dizi olasılık tablosundan $B_P = \{p(u|pa(u)) | u \in U\}$ oluşmaktadır. Olasılık tablolarındaki $pa(u)$, B_S içerisindeki u değişkeninin üstlerinin setidir. Bir Bayes ağı bir olasılık dağılımını, $P(U) = \prod_{u \in U} p(u|pa(u))$, ifade eder. Bir Bayes ağını sınıflandırıcı olarak kullanmak için, Bayes ağı tarafından ifade edilen $P(U)$ dağılımı kullanılarak $\operatorname{argmax}_y P(y|x)$, hesaplanır (3.12.).

$$P(y|x) = P(U)/P(x)$$

$$P(U) = \prod_{u \in U} p(u|pa(u)) \quad (3.12.)$$

3.2.4.2.3. Destek Vektör Makineleri

DVM yönteminde amaç veri kümesini doğrusal olmayan çekirdek fonksiyon kullanarak daha yüksek bir boyuta dönüştürüp sınıfları ayıracak hiper düzlem oluşturmaktır (Hearst, 1998).

$$f(x) = \text{sign}\left(\sum_i \alpha_i K(x, x_i) + b\right) \quad (3.13.)$$

k Karar verme, doğrusal olmayan karar verme fonksiyonu (3.13.) kullanılarak yapılır. Burada α_i 'ler Lagrange çarpanları, b bias terim ve $K(x_i, x_j)$ çekirdek fonksiyonudur. Uygun çekirdek fonksiyonun seçimi sistemin başarısında önemlidir.

$$K(x_i, x_j) = e^{-\gamma \|x_i - x_j\|^2} \quad (3.14.)$$

Çalışmada radial bias (RBF) çekirdek fonksiyonu (3.14.) kullanılmış ve çekirdek parametresi γ deneysel olarak belirlenmiştir.

3.2.4.2.4. Destek Vektör Regresyonu

$x_i \in R^n$ in öznitelik vektörü ve $z_i \in R^1$ hedef çıktı olduğu $\{(x_1, z_1), \dots, (x_l, z_l)\}$ eğitim verisini düşünelim. $C > 0$ ve $\epsilon > 0$ verilmiş parametreler olmak üzere destek vektör regresyon formu:

$$\min_{w, b, \xi, \xi^*} \frac{1}{2} \mathbf{w}^T \mathbf{w} + C \sum_{i=1}^l \xi_i + C \sum_{i=1}^l \xi_i^*$$

$$\mathbf{w}^T \phi(x_i) + b - z_i \leq \epsilon + \xi_i \quad (3.15.)$$

$$z_i - \mathbf{w}^T \phi(x_i) - b \leq \epsilon + \xi_i^*$$

$$\xi_i, \xi_i^* \geq 0, i = 1, \dots, l$$

Destek vektör regresyonu (3.15.)'te verilen ikili optimizasyon problemini çözmeye çalışmaktadır (Vapnik ve ark., 1997).

$$\min_{\alpha, \alpha^*} \frac{1}{2} (\alpha - \alpha^*)^T Q (\alpha - \alpha^*) + \epsilon \sum_{i=1}^l (\alpha - \alpha^*) + \sum_{i=1}^l z_i (\alpha - \alpha^*)$$

$$e^T (\alpha - \alpha^*) = 0 \quad (3.16.)$$

$$0 \leq (\alpha - \alpha^*) \leq C, i = 1, \dots, l$$

$$Q_{ij} = K(x_i, x_j) \equiv \phi(x_i)^T \phi(x_j)$$

$\phi(x_i)$ çekirdek fonksiyonu x_i vektörünü daha yüksek uzaya haritalar ve $K(x_i, x_j)$ ise kernel fonksiyonudur. (3.16.) çözüldükten sonra, yaklaşım fonksiyonu (3.17.) olur.

$$\sum_{i=1}^l (-\alpha + \alpha^*) K(x_i, x_j) + b \quad (3.17.)$$

3.2.4.3. Kategorik Modelin Başarım Değerlendirmesi

Kategorik sınıflandırma çalışmalarında başarı değerlendirme modelin doğru olarak sınıflandırdığı veriler ile toplam verilerin birbirilerine oranı şeklinde hesaplanmaktadır. Fakat sınıflandırma yapılacak verilerin sınıflara ilişkin dağılımları arasındaki fark büyük ise bu yöntem ile elde edilen sonuçlar yanıltıcı olmaktadır.

Çizelge 3.3. Örnek bir sınıf dağılımı

	A	B	Toplam
A	<u>90</u>	0	90
B	5	<u>5</u>	10
Toplam	95	5	100

A sınıfına ait tahmin başarısı %100 olurken, B sınıfına ait başarı %50'dir. Toplam başarı ise %95 olmuştur. Fakat bu sonuç yanıltıcıdır. Çünkü sınıflara ilişkin örnek

sayıları eşit değildir. Yapılan çalışmalarda kullanılan veri setinde benzer bir durum mevcuttur. Bu nedenle sonuçların sağlıklı olarak analizi için sınıflandırma deneylerinin başarı değerlendirilmesinde, her bir sınıf için keskinlik, duyarlılık ve F-ölçütü hesaplanmıştır. Sınıflandırıcının genel başarısı ise ağırlıksız toplam doğruluk (UA) ve ağırlıklı toplam doğruluk (WA) oranları hesaplanarak belirlenmiştir. C_i sınıfının keskinliği, p_i , sınıflandırıcının C_i sınıfından doğru olarak bulduğu örnek sayısının sınıflandırıcı tarafından C_i olarak sınıflandırılmış toplam örnek sayısına bölünmesiyle elde edilir. C_i sınıfının duyarlılığı, r_i , ise sınıflandırıcının C_i sınıfından doğru olarak bulduğu örnek sayısının C_i sınıfına ait toplam örnek sayısına oranıdır. F-ölçütü duyarlılık ve keskinliğin birlikte değerlendirilmesini sağlayan bir ölçüttür ve duyarlılıkla keskinliğin harmonik ortalamasıdır (3.18.).

$$F_i = \frac{2 \times p_i \times r_i}{p_i + r_i} \quad (3.18.)$$

Ağırlıklı toplam doğruluk (WA), doğru bulunmuş örnek sayısının toplam örnek sayısına bölünmesiyle bulunur. Ağırlıksız toplam doğruluk (UA) ise sınıfların duyarlılık değerlerinin aritmetik ortalamasıdır.

Çizelge 3.3.'te verilen tabloya ilişkin sonuçlar (3.19.) gibi olacaktır.

$$\begin{aligned} r_A &= \frac{90}{90} = 1 & p_A &= \frac{90}{95} = 0,947 & WA &= \frac{95}{100} = 0,95 \\ r_B &= \frac{5}{10} = 0,5 & p_B &= \frac{5}{5} = 1 & UA &= \frac{1,5}{2} = 0,75 \end{aligned} \quad (3.19.)$$

Çizelge 3.4. A ve B örnek sınıflarına ilişkin değerlendirme sonuçları

Sınıflar	r_i	p_i	F_i	WA	UA
A	1	0.947	0,972	0,95	0,75
B	0.5	1	0,667		

Çizelge 3.4.'te verilen tabloda örnek sınıflara ilişkin sınıflandırma sonuçları verilmiştir. Ağırlıksız toplam doğruluğun örnek sınıflar için daha doğru bir sonuç verdiği görülmektedir.

Çalışmalarda yapılan sınıflandırma işleminde 10-kat çapraz doğrulama işlemi kullanılmıştır. Veriler rastgele olarak 10 eşit parçaya bölünür ve her defasında bir parça test için geriye kalan 9 parçada eğitim için kullanılır. Böylece, eğitim-test işlemi 10 kere tekrarlanarak bütün veri test edilmiş olur.

3.2.4.4. Duygu Boyutlarını Kestirim Modelinde Başarım Değerlendirmesi

Konuşma sinyalinden çıkardığımız akustik ve spektral özelliklerin duygu boyutları değerlerini kestirmedeki performansları ilinti katsayısı (CC), karekök ortalama karesel hata (RMSE) (3.20.) ve ortalama mutlak hata (MAE) (3.21.) istatistikleri kullanarak yapılmıştır. RMSE ve MAE aşağıdaki şekilde hesaplanmaktadır.

$$RMSE = \sqrt{\frac{1}{N} \sum_{i=1}^N (\hat{z}_i - z_i)^2} \quad (3.20.)$$

$$MAE = \frac{1}{N} \sum_{i=1}^N |\hat{z}_i - z_i| \quad (3.21.)$$

$z_i \in R^1$ hedef çıktı (duygu boyutu değeri), \hat{z}_i yöntem tarafından kestirilen değer ve N ise toplam veri sayısıdır.

4. ARAŞTIRMA BULGULARI VE TARTIŞMA

4.1. Türkçe Duygusal Konuşma Veritabanının Değerlendirilmesi

4.1.1. Konuşma Sinyallerinin Duygu İçeriklerinin Belirlenmesi

Duygu tanıma problemi üzerine yapılan çalışmalarda, veritabanları oluşturulurken konuşma sinyallerinin duygu içeriklerinin belirlenmesinde iki farklı yaklaşım ön plana çıkmaktadır. Bunlardan en yaygın ve klasik olanı, duygu tanıma problemini çoklu sınıflandırma problemi olarak değerlendirip, sinyalleri *üzgün, kızgın, mutlu* gibi duygu sınıflarında ya da *negatif, negatif değil* gibi daha üst düzey ayrık kategorilerde sınıflandırmaya tabii tutmaktır. Kategorik sınıflandırmada konuşma sinyallerinin hangi duygu sınıfı ile sınıflandırılacağı öznel bir konudur. Bazı konuşma sinyalleri herhangi bir duygu kategorisine girmeyebilirken başka konuşma sinyalleri de birden fazla duyguyu içerisinde barındırabilmektedir. Bu nedenle son yıllarda psikoloji biliminde duygu belirleme üzerine yapılan çalışmalarla duyguların çok boyutlu bir uzayda birer nokta olarak temsil edilebileceği gösterilmiştir. Bu yöntem kategorik sınıflandırma yaklaşımına göre daha fazla duyguyu temsil edebilecek niteliktedir. Duygu uzayındaki her nokta farklı bir duyguyu temsil etmektedir. Bu nedenle tanımlanabilecek duygu sayısı konusunda herhangi bir sınırlama yoktur. Bu nedenle yapılan çalışmada veritabanımız hem kategorik hem de çok boyutlu duygu uzayı ile değerlendirilmiştir. Konuşma sinyalleri; 11 tane bayan ve 16 tane de erkek olmak üzere 27 kişi tarafından hem kategorik hem de 3-boyutlu duygu uzayında değerlendirilmiştir.

4.1.1.1. Kategorik Değerlendirme Sonuçları

Literatürde bulunan çalışmaları da göz önünde bulundurularak, film ve dizilerden elde edilen konuşma işaretleri Kategorik Değerlendirme-1 olarak adlandırdığımız *mutlu, şaşkın, üzgün, kızgın, korku, nötr* ve *diğer* olmak üzere 7 temel duygu kategorisinde ve aynı zamanda Kategorik Değerlendirme-2 olarak adlandırdığımız daha üst seviye duygu tanımları olan *pozitif, negatif* ve *nötr* kategorilerinde değerlendirmeye tabii tutulmuştur. Kategorik Değerlendirme-1'de mevcut duygu sınıflarından herhangi birisine girmediği düşünülen konuşma diyaloglarını sınıflandırmak için *diğer* seçeneği değerlendiricilere seçenek olarak verilmiştir.

Kategorik Değerlendirme-1 ve Kategorik Değerlendirme-2 için değerlendiricilerden elde edilen değerlendirme sonuçlarından çoğunluğun hemfikir olduğu (en çok sayıda dinleyici tarafından seçilen) duygu kategorisi o konuşma diyalogu için duygu sınıfı olarak kabul edilmiştir. Belirlenen duygu sınıflarının kategorilere dağılımı Çizelge 4.1. ve Çizelge 4.2.'de verilmiştir.

Çizelge 4.1. Konuşma sinyallerinin duygu kategorilerine göre dağılımları (Kategorik Değerlendirme-1)

Duygu Sınıfı	Bayan	Erkek	Toplam
Nötr	527	1628	2155
Üzgün	288	282	570
Mutlu	175	231	406
Şaşkın	143	236	379
Kızgın	346	1047	1393
Korku	37	59	96
Diğer	34	93	127
Toplam	1550	3576	5126

Çizelge 4.2. Konuşma sinyallerinin duygu kategorilerine göre dağılımları (Kategorik Değerlendirme-2)

Duygu Sınıfı	Bayan	Erkek	Toplam
Pozitif	272	395	667
Nötr	507	1600	2107
Negatif	783	1597	2380
Toplam	1562	3592	5154

Dinleme testlerini yapan değerlendiricilerin kategorik değerlendirme sonuçlarının birbiriyle ne kadar uyumlu olduklarını ölçmek amacıyla Kappa katsayısından (Fleiss,

1971) yararlanılmıştır. Kappa istatistiği öznel testlerdeki kategorik değişkenler arasındaki uyumu ölçmek için kullanılır. Kappa katsayısı, κ , şu şekilde hesaplanır.

$$\kappa = \frac{P(A) - P(E)}{1 - P(E)} \quad (4.1.)$$

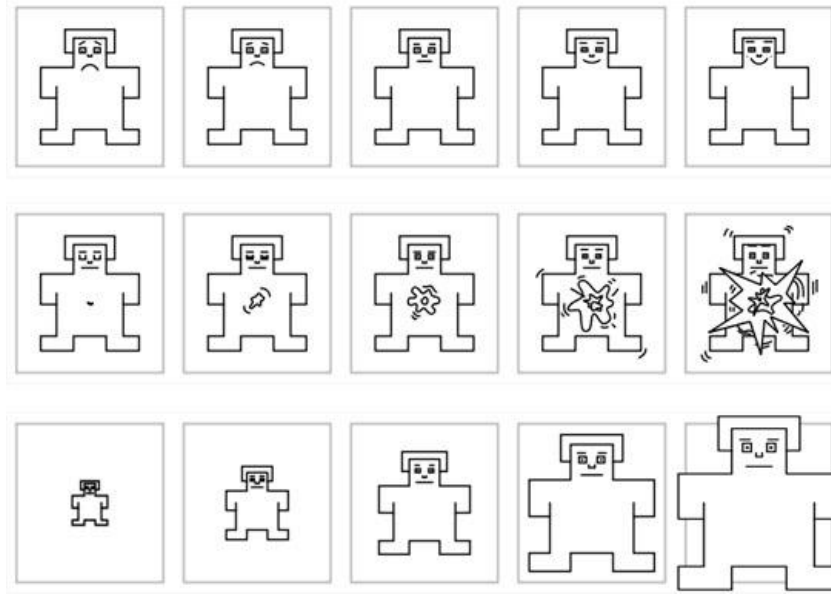
Burada $P(A)$, K tane kodlayıcının kaç defa uyumlu olduğu oranı ve $P(E)$ ise K tane kodlayıcının şans eseri uyumlu olma oranıdır. Eğer kodlayıcılar arasında tam bir uyum varsa $\kappa = 1$, eğer kodlayıcılar arasında hiçbir kodlamada uyum yoksa $\kappa = 0$ çıkacaktır. Çizelge 4.1. (Kategorik Değerlendirme-1) ve Çizelge 4.2.'deki (Kategorik Değerlendirme-2) tablolarda oluşturulan veritabanına ilişkin duygu kategorilerinin dağılımı verilmiştir. Veritabanındaki konuşma işaretlerinin 27 değerlendirici (11'i bayan 16'sı erkek) tarafından değerlendirilmesi ile elde edilen sonuçların değerlendirici uyumları Kappa istatistiği (4.1.) ile tespit edilmiş ve sonuç Çizelge 4.3.'te verilmiştir. Tablodaki sonuçlara göre değerlendiricilerin orta düzeyde uyumlu oldukları sonucuna varılabilir.

Çizelge 4.3. Kappa katsayıları

Değerlendirme	Kappa
Kategorik Değerlendirme-1	0.314
Kategorik Değerlendirme-2	0.277

4.1.1.2. 3-Boyutlu Duygu Uzayı Değerlendirme Sonuçları

Yapılan çalışmada duygunun 3-boyutlu uzayda belirlenmesinde Bradley ve Lang'ın (1994) geliştirdiği Self-Assessment Manikin (SAM) görselleri kullanılmıştır. Değerlendirme testini yapan değerlendiricilerden dinledikleri ses dosyalarını değerlik, aktivasyon ve baskınlık boyutları için değerleri $[-1,1]$ aralığında olan 5-adımlı skalada tanımlamaları istenmiştir. Değerlik negatiften (-1) pozitif (+1) doğru bir aralıkta, aktivasyon sakin/durgunluktan (-1) heyecanlı (+1) olmaya doğru bir aralıkta, baskınlık ise zayıflıktan (-1) güçlü (+1) olmaya doğru olan bir aralıkta olmak üzere verilmiştir (Şekil 4.1).



Şekil 4.1 3-Boyutlu duygu uzayında değerlendirme (1. satır değerlik, 2. satır aktivasyon, 3. satır baskınlık)

Bu değerlendirmede 27 değerlendiricinin tüm konuşma işaretlerine (5304 tane) ilişkin 3-boyutlu duygu kategorisi sonuçları arasındaki uyum incelenmiştir. Uyum Pearson korelasyon katsayıları ile belirlenmeye çalışılmıştır (3.4.). Değerlendirme sonuçlarına ilişkin ortalama standart sapmalar ve ortalama Pearson korelasyon katsayıları Çizelge 4.4.'te verilmiştir.

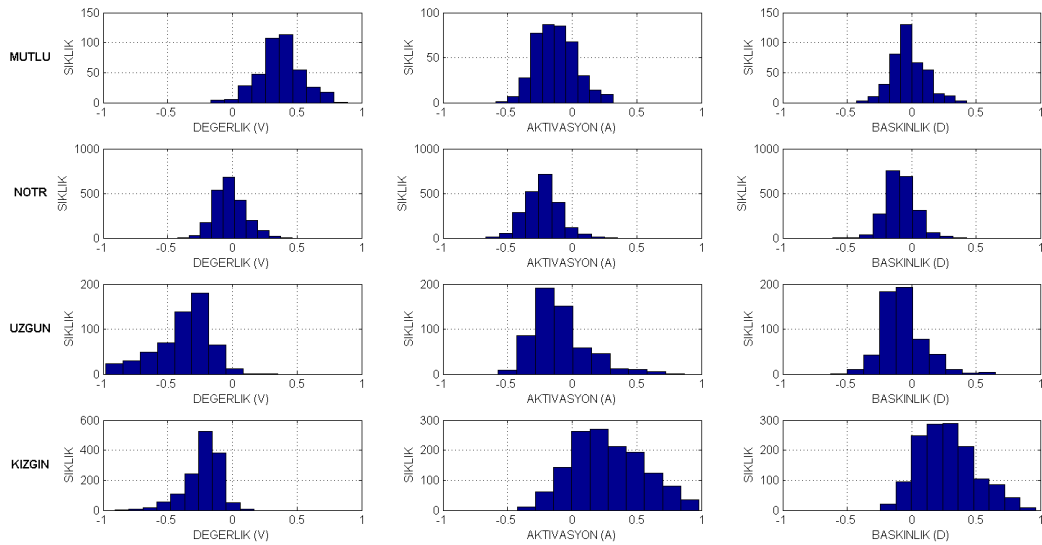
Çizelge 4.4. Değerlendirme testini yapan değerlendiriciler arasındaki ortalama standart sapma ve ortalama korelasyon sonuçları. D: değerlik, A: aktivasyon, B: baskınlık

Ortalama Standart Sapma			Ortalama Korelasyon		
D	A	B	D	A	B
0.28	0.43	0.37	0.62	0.57	0.53

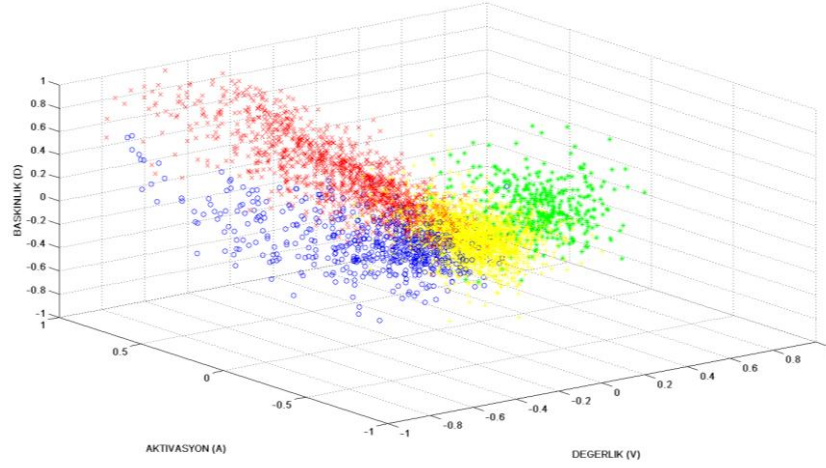
Düşük ortalama standart sapma değerlendiricilerin yüksek oranda uyumlu olduklarını gösterirken, yüksek ortalama korelasyon değerleri değerlendiricilerin daha uyumlu olduklarını göstermektedir. Çizelge 4.4.'te verilen veriler incelendiğinde dinleyiciler arasında her duygu boyutu için değerlendirmelerde kabul edilebilir seviyede bir uyum olduğu gözlenmektedir. Yine tabloda verilen sonuçlardan anlaşılacağı üzere dinleyicilerin *değerlik* boyutunda diğer boyutlara göre daha uyumlu sonuçlar verdiği söylenebilir.

4.1.2. Kategorik Sınıfların ve 3-Boyutlu Duygu Uzayı Değerlendirmelerinin Birlikte Analizi

Oluşturduğumuz veritabanında konuşma sinyallerinin hem kategorik hem de 3-boyutlu duygu uzayında duygusal içeriklerinin bulunması, duygu kategorileri ve bunların duygu uzayındaki yerleri arasındaki ilişkiyi incelememizi sağlamaktadır. Her bir temel duygu sınıfı (mutlu, nötr, kızgın ve üzgün) için duygu uzayı dağılımları (histogram olarak) Şekil 4.2’de verilmiştir. Şekil 4.3’teki grafikte ise bu duygu sınıflarının 3-boyutlu duygu uzayındaki dağılımları birlikte gösterilmiştir. Duygu sınıflarının değerlik, aktivasyon ve baskınlık düzlemlerindeki ortalama değerleri ve standart sapmaları da Çizelge 4.5.’te verilmiştir. Şekil 4.2, Şekil 4.3 ve Çizelge 4.5.’te verilen sonuçlar incelendiğinde veritabanındaki konuşma sinyallerinden kızgın duygu kategorisinde değerlendirilenler orta düzeyde negatif, heyecanlı ve güçlü; üzgün sınıfında olanlar çok negatif ve mutlu sınıfında olanlar çok pozitif olarak bulunmuştur. Elde edilen sonuçlar literatürdeki diğer çalışmalarla benzerlik göstermektedir (Grimm, 2006).



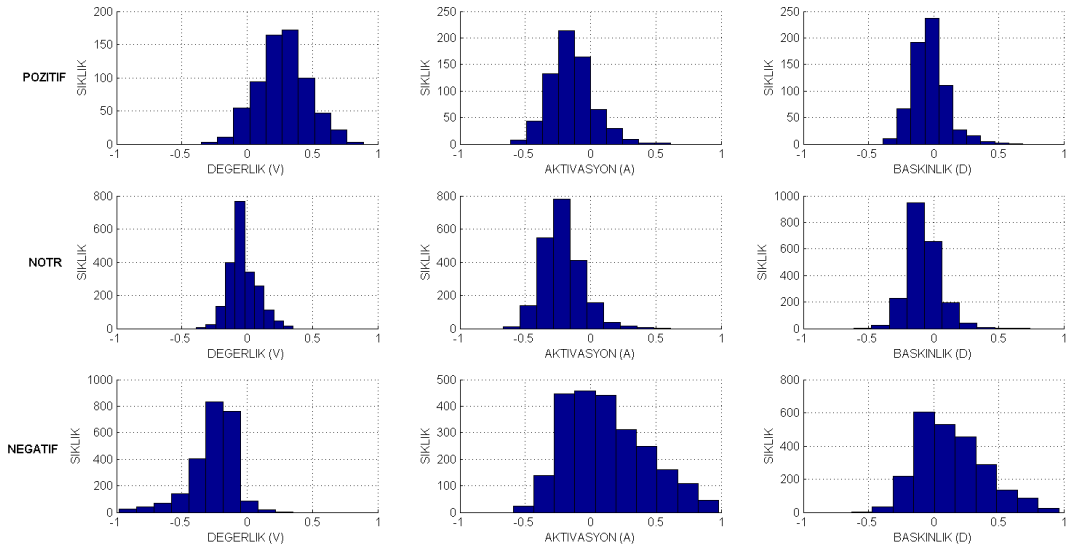
Şekil 4.2 Temel duygu kategorileri (Nötr, Üzgün, Mutlu, Kızgın) için değerlik, aktivasyon ve baskınlık sonuçlarının dağılımları



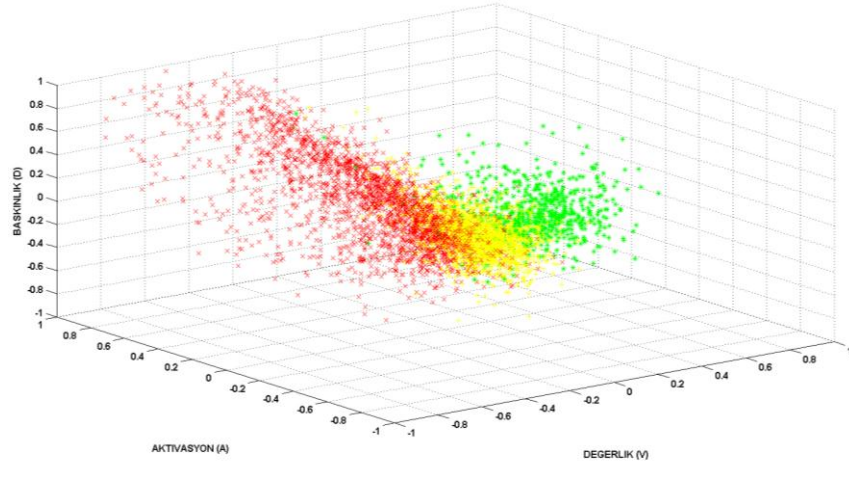
Şekil 4.3 Mutlu (*, yeşil), Üzgün (o, mavi), Kızgın (x, kırmızı) ve Nötr (+, sarı) duygu sınıflarının duygu uzayındaki dağılımları (Kategorik değerlendirme-1 için)

Çizelge 4.5. Duygu sınıflarının değerlik, aktivasyon ve baskınlık boyutlarındaki ortalama ve standart sapma değerleri

Duygu Sınıfı	Değerlik	Aktivasyon	Baskınlık
Mutlu	0.37±0.16	-0.13±0.16	-0.02±0.13
Nötr	-0.02±0.12	-0.23±0.14	-0.08±0.11
Üzgün	-0.38±0.21	-0.10±0.22	-0.07±0.17
Kızgın	-0.23±0.14	0.27±0.28	0.28±0.22



Şekil 4.4 Kategorik değerlendirme-2 (Pozitif, Nötr, Negatif) için değerlik, aktivasyon ve baskınlık sonuçlarının dağılımları



Şekil 4.5 Pozitif (*, yeşil), Nötr (+, sarı) ve Negatif (x, kırmızı) duygu sınıflarının duygu uzayındaki dağılımları (Kategorik değerlendirme-2 için)

Çizelge 4.6. Duygu sınıflarının değerlik, aktivasyon ve baskınlık boyutlarındaki ortalama ve standart sapma değerleri

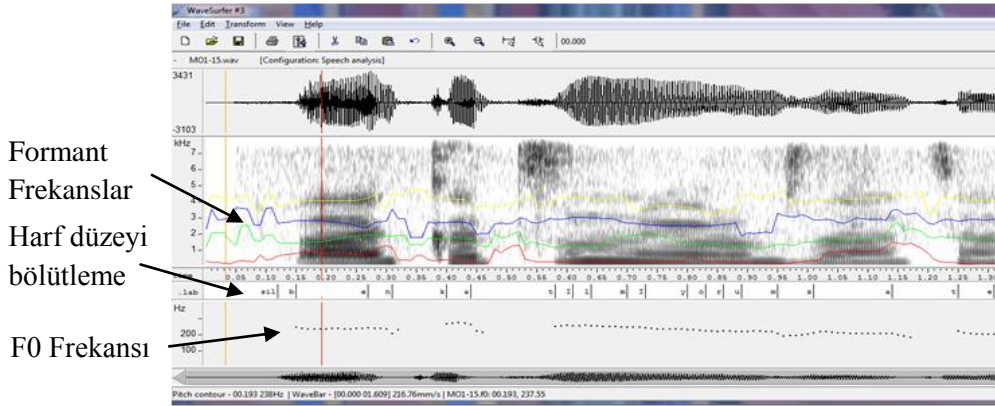
Duygu Sınıfı	Değerlik	Aktivasyon	Baskınlık
Pozitif	0.28±0.19	-0.15±0.17	-0.03±0.13
Nötr	-0.03±0.1	-0.22±0.15	-0.08±0.13
Negatif	-0.26±0.17	0.12±0.31	0.14±0.26

Benzer analiz; pozitif, negatif ve nötr duygu kategorileri için de yapılarak sonuçlar Şekil 4.4, Şekil 4.5 ve Çizelge 4.6.'da verilmiştir. Sonuçlara göre toplanan veritabanında pozitif duygu kategorisi orta seviyede pozitif ve makul seviyede sakin, negatif duygu kategorisinde olduğu belirtilenlerin ise orta seviyede negatif, heyecanlı ve güçlü oldukları söylenebilir.

4.2. Türkçe Duygusal Konuşmanın Akustik Özelliklerinin İncelenmesi

Duygunun insanlar arası etkileşimde önemli bir rol üstlendiği bilinmektedir. İnsanlar kendi aralarındaki iletişim sırasında birçok bilgi kanalını kullanırlar. Konuşma, vücut hareketleri ve yüz ifadeleri iletişim kanallarının en önemlileridir. Konuşmada ne söylediğimiz kadar nasıl söylediğimiz de iletişimin başarısında rol oynamaktadır. Konuşma prosodisi konuşmacının yaşı, cinsiyeti, aksanı gibi bilgileri barındırması yanında duygu durumu hakkında da bilgi içermektedir. Konuşmanın vurgusu, tonlaması, temposu gibi özellikleri duygu durumundan etkilenmekte ve değişikliğe

uğramaktadır. Duygusal konuşma sırasında konuşmanın akustik özelliklerinde meydana gelen değişikliklerin ayrıntılı analizi, duygu tanıma çalışmaları yanı sıra özellikle duygusal konuşma sentezleme çalışmalarında kullanılmak üzere önemli bilgiler elde etmemizi sağlayacaktır. Bu çalışmada, duygunun Türkçedeki seslerin akustik özellikleri üzerine etkileri istatistiksel analizler yardımıyla incelenmiştir. Özellikle Türkçedeki ünlü seslerin temel frekansı, formant frekansları ve süre bilgileri ele alınmıştır. Oluşturulan Türkçe duygusal konuşma veritabanındaki (Oflazoglu ve Yıldırım, 2011) konuşma sinyallerinin harf düzeyinde bölütlenmeleri (harflerin konuşma işaretleri içerisindeki yerleri) HTK (Young ve ark., 2009) programı yardımıyla güdümlü-hizalama (forced-alignment) tekniği kullanılarak elde edilmiştir. Güdümlü hizalama tekniği, tüm veritabanı kullanılarak fonem tabanlı Saklı Markov modellerinin eğitilmesi ve harflerin sinyal içerisindeki başlangıç ve bitiş zamanlarının otomatik olarak elde edilmesidir. Sesin akustik özellikleri cinsiyete bağlı olarak değişiklikler göstermektedir. Örneğin kadın sesinin ortalama temel frekansı erkeklerinkine göre daha yüksektir. Bu nedenle çalışmada analizler kadın ve erkek konuşmacılar için ayrı ayrı yapılmıştır.



Şekil 4.6 Ses işaretinin WaveSurfer programındaki görünüşü

4.2.1. Temel Frekans İncelemesi

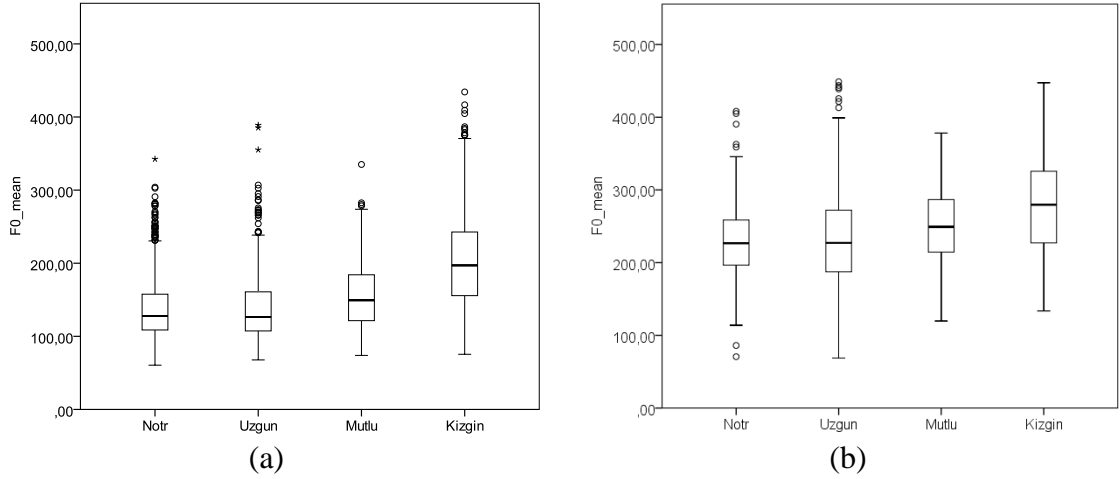
Temel frekans ünlü seslerde ses tellerinin titreşim frekansıdır ve sesin kalınlığı/inceliğiyle orantılıdır. Oluşturulan Türkçe duygusal konuşma veritabanında

bulunan konuşma sinyallerinin F0 frekans eğrileri (contours) ESPS programının *get_f0* fonksiyonu (Talkin, 1995) kullanılarak elde edilmiştir. Ham değerler 3 noktalı kayan ortalama yöntemiyle filtreleme işleminden geçirilerek otokorelasyon yönteminden kaynaklanan F0 hesaplama hataları azaltılmıştır. Konuşma sinyallerine ait ortalama, minimum, maksimum ve ortanca değer gibi cümle düzeyinde istatistiklerin hesaplanmasında filtrelenmiş F0 değerleri kullanılmıştır. Ünlü harflerin F0 değerleri, otomatik bölütleme sonucu bulunan başlangıç ve bitiş zamanları kullanılarak F0 eğrilerinden bulunan aralıklardaki değerlerin ortalaması alınarak hesaplanmıştır. Duygunun F0 üzerine etkisi ANOVA analizi ve çoklu karşılaştırma testleriyle belirlenmeye çalışılmıştır.

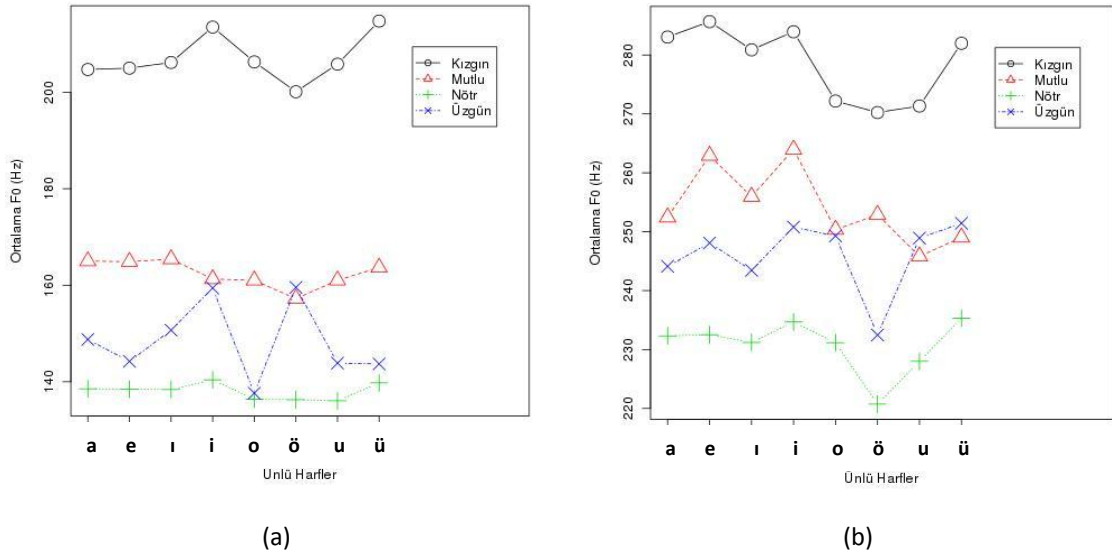
Yapılan ANOVA analizleri, duygunun hem erkek hem de kadın konuşmacılarda F0 üzerindeki etkisinin istatistiksel olarak önemli olduğunu göstermektedir (erkek: $F(3,3169)=391,288$, $p<0.001$, kadın: $F(3,1322)=54,368$, $p<0.001$). Farklı duygu kategorileri için ortalama ve standart sapma F0 değerleri Çizelge 4.7.'de verilmiştir. Ayrıca duyguların F0 dağılımlarını karşılaştırmak amacıyla veriler kutu grafiği olarak Şekil 4.7'de verilmiştir. Kutu grafiği gösterimde orta çizgi medyan, kutunun alt sınırı verinin en düşük %25'lik kısmını ve üst sınırı ise en yüksek %25'lik kısmını sınırlar.

Çizelge 4.7. Farklı duygu kategorileri için ortalama ve standart sapma (std) F0 değerleri

	Erkek		Kadın	
	Ortalama (Hz)	Std (Hz)	Ortalama (Hz)	Std (Hz)
Nötr	136	39	228	48
Üzgün	144	56	240	69
Mutlu	158	48	251	51
Kızgın	203	63	279	64



Şekil 4.7 Duygu kategorilerinin F0 dağılımlarının kutu grafiği olarak gösterimleri (a) Erkek (b) Kadın



Şekil 4.8 Ünlü harflerin ortalama temel frekansları, (a) erkek (b) kadın

Çizelge 4.7.'de verilen değerlerden de anlaşılacağı üzere, kızgın, mutlu ve üzgün konuşmalar nötr konuşmalara göre daha yüksek F0 değerine sahiptir. Duygu sınıflarının ortalama F0 değerleri arasındaki farkların istatistiksel olarak anlamlı olup olmadığını analiz etmek için çoklu karşılaştırma testleri uygulanmıştır. Sonuçlar, erkek konuşmacılarda nötr ve üzgün, kadın konuşmacılarda ise kızgın ve mutlu hariç, ortalama F0 değerleri arasındaki farkların istatistiksel olarak anlamlı olduğunu göstermektedir (Tukey test, $p < 0.05$). Erkeklerde kızgın konuşmaya ait ortalama F0 değerleri, kadınlarda ise hem üzgün hem de kızgın konuşmalara ait ortalama F0

değerleri diğer duygu sınıflarınkilerle karşılaştırıldığında daha çok değişkenlik göstermektedir (Şekil 4.8).

Ünlü seslerin her bir duygu sınıfındaki ortalama F0 değerleri Şekil 4.8'deki grafikte verilmiştir. Hem erkek hem de kadın konuşmacıların kızgın ünlü seslerinin ortalama F0 değerleri diğer duygu sınıflarına göre daha yüksek iken nötr seslerin ortalama F0 değerlerinin düşük olduğu görülmektedir.

Çizelge 4.8. Erkek ünlü seslerin ortalama F0 değerlerinin çoklu karşılaştırma sonuçları

	Nötr	Üzgün	Mutlu	Kızgın
Nötr	-	A,İ,İ	A,E,İ,İ,O,U,Ü	A,E,İ,İ,O,Ö,U,Ü
Üzgün	A,İ,İ	-	A,E,İ,O,U	A,E,İ,İ,O,Ö,U,Ü
Mutlu	A,E,İ,İ,O,U,Ü	A,E,İ,O,U	-	A,E,İ,İ,O,Ö,U,Ü
Kızgın	A,E,İ,İ,O,Ö,U,Ü	A,E,İ,İ,O,Ö,U,Ü	A,E,İ,İ,O,Ö,U,Ü	-

Çizelge 4.9. Kadın ünlü seslerin ortalama F0 frekans değerlerinin çoklu karşılaştırma sonuçları

	Nötr	Üzgün	Mutlu	Kızgın
Nötr	-	A,E,İ,U	A,E,İ,İ,Ö	A,E,İ,İ,O,Ö,U,Ü
Üzgün	A,E,İ,U	-	E	A,E,İ,İ,O,Ö,U
Mutlu	A,E,İ,İ,Ö	E	-	A,E,İ,O,U
Kızgın	A,E,İ,İ,O,Ö,U,Ü	A,E,İ,İ,O,Ö,U	A,E,İ,O,U	-

Çoklu karşılaştırma test sonuçları Çizelge 4.8. ve Çizelge 4.9.'da erkek ve kadın sesleri için özetlenmiştir. Tablolarda ortalama F0 değerleri istatistiksel olarak anlamlı olan ünlü sesler verilmiştir. Tablolardan açıkça görüleceği üzere birçok ünlü sesli harfe ilişkin ortalama temel frekans değeri konuşma sinyalindeki duyguya göre farklılık göstermektedir.

4.2.2. Formant Frekansları Analizi

Formantlar ses yolu filtresinin rezonans frekanslarıdır ve ünlü seslerle ilgili en önemli akustik parametrelerdendir. Türkçede ünlüler çene açıklığına, dudak biçimine ve dilin devimine göre sınıflandırılmaktadır. Örneğin /a/ sesini çıkarırken çene açık, dudak düz

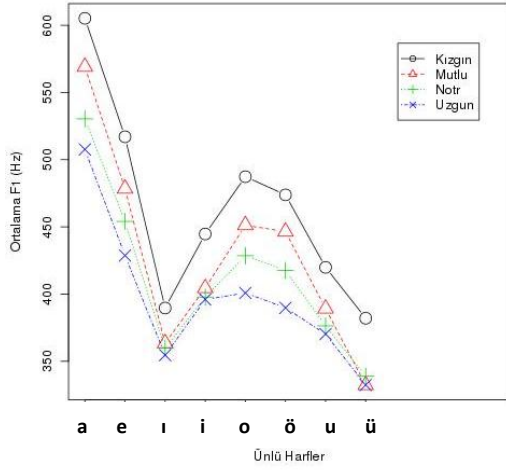
ve dil devimi arkadadır. Formant frekansları seslendirilen ünlüye ve konuşmacıya bağlı olarak değişmektedir. Formant frekansları ve bant genişlikleri duygu tanıma çalışmalarında öznitelik olarak kullanılmıştır (Lee ve Narayanan, 2005; Bozkurt ve ark., 2011). Yapılan analiz çalışmalarında duygunun ünlü seslerin ilk üç formant frekansı üzerindeki etkileri incelenmiştir. Goudbeek ve ark. (2009) duygunun özellikle ilk iki formant frekansını etkilediklerini belirtmişlerdir. Formant frekanslarının çıkarılmasında Snack ses analizi programı kullanılmıştır (Sjolander, 2004).

Ünlü harflerin F1, F2, F3 değerleri otomatik bölütleme sonucu bulunan başlangıç ve bitiş zamanları kullanılarak bulunan aralıktaki değerlerin ortalaması alınarak elde edilmiştir.

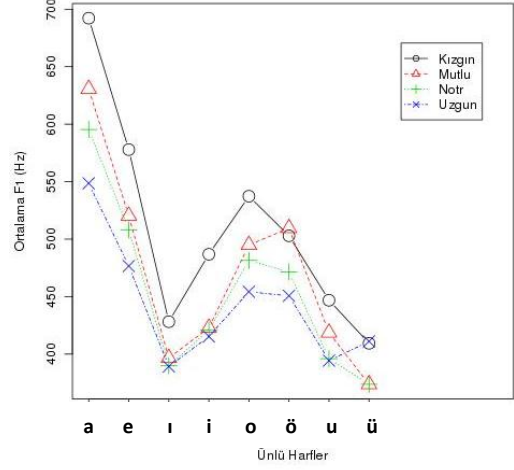
Ünlü seslerin farklı duygu sınıfları için ilk üç formant frekanslarının ortalama grafikleri erkek ve kadın konuşmacılar için Şekil 4.9'da verilmiştir. Grafiklerde ayrıca formant frekanslarının ortalama ve standart sapma (tabloda parantez içinde) değerleri erkek ve kadın konuşmacılar olarak Çizelge 4.10.'da özetlenmiştir. Yapılan iki faktörlü ANOVA analizi, hem erkek hem de kadın konuşmacılarda duygunun formant frekans değerleri üzerinde etkisi olduğunu ve aynı zamanda duygu ile ünlü sesleri arasında etkileşim olduğunu göstermektedir.

Erkek konuşmacılar: F1 formant frekansı iki faktörlü ANOVA $F=9.26$, $p<0.001$, F2 formant frekansı iki faktörlü ANOVA $F=3.2$, $p<0.001$, F3 formant frekansı iki faktörlü ANOVA $F=2.9$, $p<0.001$.

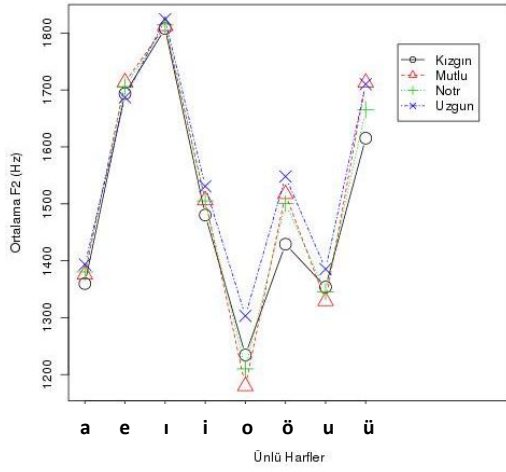
Kadın konuşmacılar: F1 formant frekansı iki faktörlü ANOVA $F=7.15$, $p<0.001$, F2 formant frekansı iki faktörlü ANOVA $F=2.35$, $p<0.001$, F3 formant frekansı iki faktörlü ANOVA $F=1.69$, $p<0.05$.



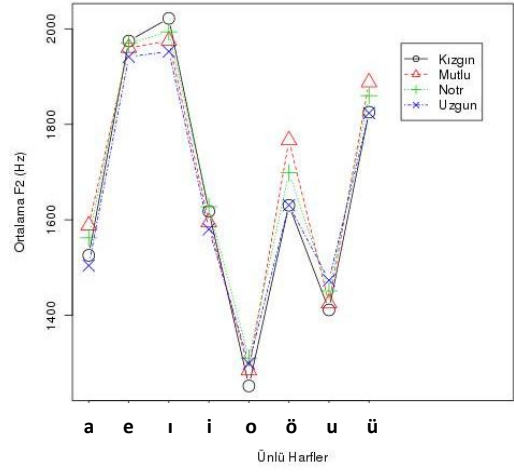
(a)



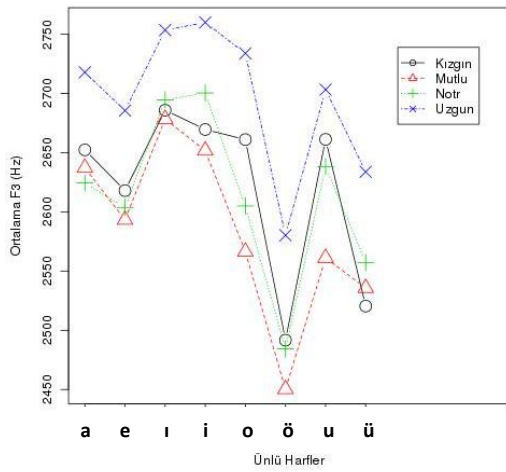
(b)



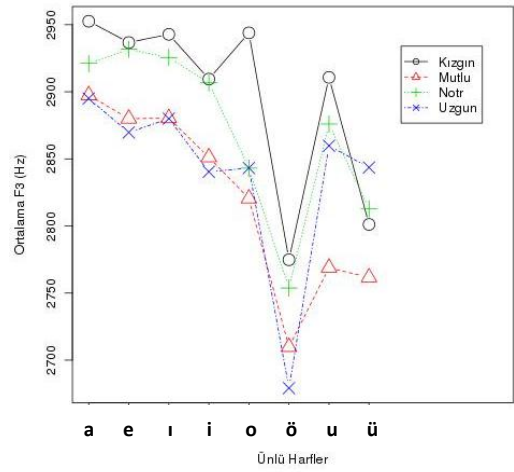
(c)



(d)



(e)



(f)

Şekil 4.9 Erkek konuşmacılara ((a), (c) ve (e) grafikleri) ve kadın konuşmacılara ((b), (d) ve (f) grafikleri) ait ünlü harflerin duygu kategorilerine göre ortalama formant frekans grafikleri

Çizelge 4.10. Erkek ve kadın konuşmacılarda ünlü harflerin duygu kategorilerine göre ortalama formant frekansları ve standart sapmaları

Harf	Duygu	Erkek			Kadın		
		F1	F2	F3	F1	F2	F3
A	Nötr	522 (132)	1373 (256)	2608 (325)	588 (189)	1555 (295)	2908 (325)
	Üzgün	491 (167)	1377 (318)	2698 (364)	538 (201)	1485 (289)	2878 (327)
	Mutlu	559 (125)	1371 (216)	2624 (302)	624 (195)	1583 (292)	2888 (320)
	Kızgın	601 (143)	1356 (201)	2642 (318)	686 (189)	1522 (279)	2939 (358)
E	Nötr	447 (117)	1692 (228)	2588 (289)	502 (168)	1959 (360)	2919 (280)
	Üzgün	417 (125)	1675 (247)	2672 (352)	467 (162)	1928 (356)	2856 (274)
	Mutlu	474 (119)	1706 (233)	2583 (278)	510 (166)	1947 (372)	2864 (292)
	Kızgın	512 (129)	1684 (235)	2606 (297)	574 (188)	1963 (355)	2923 (311)
İ	Nötr	371 (126)	1463 (332)	2645 (359)	396 (122)	1585 (365)	2861 (326)
	Üzgün	356 (152)	1482 (355)	2695 (398)	390 (133)	1547 (328)	2806 (299)
	Mutlu	387 (126)	1471 (291)	2589 (338)	411 (133)	1568 (362)	2815 (358)
	Kızgın	421 (143)	1452 (299)	2626 (340)	455 (164)	1591 (342)	2878 (367)
ı	Nötr	339 (103)	1793 (299)	2664 (357)	372 (122)	1969 (410)	2894 (300)
	Üzgün	326 (118)	1802 (295)	2721 (370)	364 (114)	1924 (398)	2847 (300)
	Mutlu	348 (103)	1796 (266)	2661 (348)	377 (122)	1951 (398)	2848 (322)
	Kızgın	373 (121)	1786 (293)	2653 (344)	410 (129)	1986 (427)	2910 (324)
O	Nötr	422 (109)	1202 (368)	2588 (354)	473 (127)	1305 (352)	2827 (305)
	Üzgün	390 (117)	1290 (482)	2724 (408)	440 (128)	1304 (352)	2828 (316)
	Mutlu	446 (98)	1177 (280)	2556 (363)	488 (123)	1277 (304)	2818 (337)
	Kızgın	482 (127)	1228 (397)	2647 (352)	525 (152)	1248 (316)	2926 (337)
Ö	Nötr	414 (74)	1501 (199)	2476 (338)	469 (99)	1704 (228)	2750 (230)
	Üzgün	379 (70)	1531 (294)	2572 (407)	444 (121)	1628 (282)	2675 (221)
	Mutlu	445 (102)	1521 (129)	2444 (330)	508 (109)	1776 (232)	2706 (208)
	Kızgın	472 (97)	1429 (178)	2483 (356)	500 (120)	1621 (236)	2760 (253)
U	Nötr	354 (119)	1314 (392)	2600 (365)	373 (112)	1419 (369)	2819 (340)
	Üzgün	326 (120)	1339 (405)	2640 (380)	364 (101)	1435 (376)	2811 (343)
	Mutlu	377 (134)	1300 (309)	2517 (353)	397 (114)	1407 (320)	2725 (329)
	Kızgın	405 (151)	1334 (407)	2628 (362)	433 (146)	1389 (376)	2889 (366)
Ü	Nötr	325 (78)	1642 (245)	2532 (334)	359 (90)	1842 (310)	2777 (276)
	Üzgün	302 (66)	1682 (283)	2610 (384)	377 (102)	1795 (358)	2799 (355)
	Mutlu	320 (83)	1682 (268)	2506 (318)	353 (86)	1850 (349)	2745 (324)
	Kızgın	367 (129)	1598 (260)	2490 (350)	395 (108)	1806 (276)	2773 (255)

Özellikle birinci formant frekansının duygudan etkilendiği gözlenmektedir (Şekil 4.9 (a) ve (b)). Şekil 4.9 (a) ve (b) genel olarak incelendiğinde, ünlü seslerin F1 değerlerinin kızgın ve mutlu duyguda nötr duyguya karşılaştırıldığında arttığı; üzgünde ise azaldığı gözlenmektedir. Ünlü seslerin F1 değerlerinde Kızgın>Mutlu>Nötr>Üzgün şeklinde bir sıralama açıkça görülmektedir. Sonuçlara göre ünlü seslerinin F2 formant frekansları üzerinde duygunun etkisi üzerine bir çıkarım yapmak mümkün görünmemektedir. Şekil 4.9 (e)'den anlaşılacağı üzere, erkek konuşmacılarda üzgün duyguda ünlü seslerin F3 formant frekansları diğer duygularla karşılaştırıldığında daha yüksektir. Benzer eğilim kadın konuşmacılarda gözlenmemektedir.

Çizelge 4.11. Erkek ünlü seslerin ortalama F1 formant frekans değerlerinin çoklu karşılaştırma sonuçları

	Nötr	Üzgün	Mutlu	Kızgın
Nötr	-	A,E,O,Ö,U	A,E	A,E,I,İ,O,Ö,U,Ü
Üzgün	A,E,O,Ö,U	-	A,E,İ,O,Ö,U	A,E,I,İ,O,Ö,U,Ü
Mutlu	A,E	A,E,İ,O,Ö,U	-	A,E,I,İ,O,U,Ü
Kızgın	A,E,I,İ,O,Ö,U,Ü	A,E,I,İ,O,Ö,U,Ü	A,E,I,İ,O,U,Ü	-

Çizelge 4.12. Kadın ünlü seslerin ortalama F1 formant frekans değerlerinin çoklu karşılaştırma sonuçları

	Nötr	Üzgün	Mutlu	Kızgın
Nötr	-	A,E,O	A	A,E,I,İ,O,U,Ü
Üzgün	A,E,O	-	A,E,O,Ö	A,E,I,İ,O,Ö,U
Mutlu	A	A,E,O,Ö	-	A,E,I,İ,O,U
Kızgın	A,E,I,İ,O,U,Ü	A,E,I,İ,O,Ö,U	A,E,I,İ,O,U	-

Çizelge 4.13. Erkek ünlü seslerin ortalama F2 formant frekans değerlerinin çoklu karşılaştırma sonuçları

	Nötr	Üzgün	Mutlu	Kızgın
Nötr	-	O		A,Ö
Üzgün	O	-	O	Ö,Ü
Mutlu		O	-	Ö,Ü
Kızgın	A,Ö	Ö,Ü	Ö,Ü	-

Çizelge 4.14. Kadın ünlü seslerin ortalama F2 formant frekans değerlerinin çoklu karşılaştırma sonuçları

	Nötr	Üzgün	Mutlu	Kızgın
Nötr	-	A		A
Üzgün	A	-	A,Ö	
Mutlu		A,Ö	-	A,Ö
Kızgın	A		A,Ö	-

Çizelge 4.15. Erkek ünlü seslerin ortalama F3 formant frekans değerlerinin çoklu karşılaştırma sonuçları

	Nötr	Üzgün	Mutlu	Kızgın
Nötr	-	A,E,İ,O	U	A,O
Üzgün	A,E,İ,O	-	A,E,İ,İ,O,U	A,E,İ,İ,O,Ü
Mutlu	U	A,E,İ,İ,O,U	-	O,U
Kızgın	A,O	A,E,İ,İ,O,Ü	O,U	-

Çizelge 4.16. Kadın ünlü seslerin ortalama F3 formant frekans değerlerinin çoklu karşılaştırma sonuçları

	Nötr	Üzgün	Mutlu	Kızgın
Nötr	-	E,İ	E	O
Üzgün	E,İ	-		A,E,İ,O
Mutlu	E		-	A,E,İ,O,U
Kızgın	O	A,E,İ,O	A,E,İ,O,U	-

Çoklu karşılaştırma test sonuçları Çizelge 4.11. - Çizelge 4.16. arasındaki tablolarda erkek ve kadın sesleri için özetlenmiştir. Tablolarda duygu sınıfları arasında ortalama F1, F2 ve F3 formant frekans farkları istatistiksel olarak anlamlı olan ünlü sesler verilmiştir. Tablolarda verilen sonuçlar konuşmacıların vermeye çalıştıkları duyguya göre ünlü sesleri çıkarırken dilin pozisyonunda, çene ve dudak hareketlerinde değişiklik yaptıklarını göstermektedir.

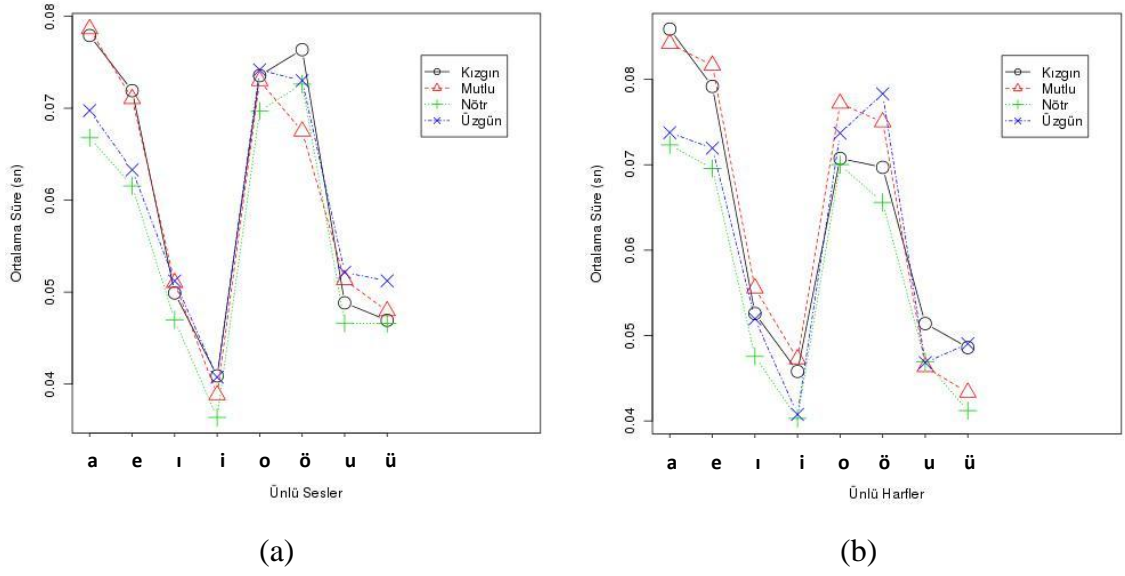
4.2.3. Ünlü Harflerin Duygu Durumuna Göre Sürelerinin Analizi

Süre, duygu üzerine literatürde yapılan çalışmalarda formant frekansları gibi seslerin ayırt edici özelliklerinden birisi kabul edilmektedir. Yapılan bu çalışmada konuşmacının duygu durumunun ünlü seslerin süresi üzerindeki etkileri istatistiksel olarak incelenmiştir. Çizelge 4.17.'de farklı duygu kategorileri için ünlü seslerin sürelerinin ortalama ve standart sapma (std) değerleri verilmiştir. İlk olarak duygu sınıflarının ortalama süre değerlerinin istatistiksel olarak farklı olup olmadıklarını test etmek amacıyla tek-yönlü varyans analizi (one-way ANOVA) Tukey çoklu karşılaştırma testiyle birlikte uygulanmıştır. Analiz sonuçlarına göre, duygu hem erkek hem de kadın konuşmacılarda ortalama ünlü ses süreleri üzerinde istatistiksel olarak etkilidir (Kadın sesleri: $F=43.913$, $p<0.001$, erkek sesleri: $F= 88.932$, $p<0.001$). Çoklu karşılaştırma sonucunda kızgın ve mutlu duyguların ortalama süreleri arasındaki farkın istatistiksel olarak anlamlı olmadığı, diğer bütün ikili karşılaştırmalarda ise ortalama sürelerin istatistiksel olarak farklı olduğu görülmüştür.

Çizelge 4.17. Farklı duygu kategorileri için ünlü seslerin sürelerinin ortalama ve standart sapma (std) değerleri

	Erkek		Kadın	
	Ortalama (ms)	Std (ms)	Ortalama (ms)	Std (ms)
Nötr	56.55	36.68	60.02	41.56
Üzgün	59.82	40.90	62.66	46.35
Mutlu	63.16	47.72	69.77	50.19
Kızgın	63.51	45.60	69.78	51.41

Ünlü seslerin her bir duygu sınıfındaki ortalama süre değerleri Şekil 4.10'daki grafiklerde verilmiştir. Şekil 4.10 (a) ve (b) incelendiğinde açık ünlüler olan /a/,/e/,/o/,/ö/ harflerinin ortalama sürelerinin, kapalı ünlüler olan /i/,/ı/,/u/,/ü/ harflerinkine göre hem erkek hem de kadın konuşmacılar için daha uzun olduğu gözlenmektedir. Ünlü harflerin duygusal kategorilere göre ortalama süreleri Çizelge 4.18.'de verilmiştir.



Şekil 4.10 Ünlü harflerin duygusal kategorilere göre ortalama süreleri, (a) erkek konuşmacılar, (b) kadın konuşmacılar

Çizelge 4.18. Ünlü harflerin duygusal kategorilere göre ortalama süreleri

	Erkek				Kadın			
	Nötr	Üzgün	Mutlu	Kızgın	Nötr	Üzgün	Mutlu	Kızgın
a	66.79	69.74	78.62	77.91	72.32	73.76	84.19	85.89
e	61.55	63.30	71.03	71.91	69.57	71.94	81.64	79.18
i	46.99	51.21	51.02	49.91	47.55	51.96	55.55	52.59
ı	36.36	40.76	38.82	40.89	40.30	40.77	47.25	45.81
o	69.70	74.17	72.98	73.55	70.02	73.71	77.18	70.75
ö	72.62	73.00	67.50	76.37	65.59	78.31	75.00	69.70
u	46.64	52.15	51.34	48.84	46.95	46.83	46.35	51.40
ü	46.58	51.23	47.95	46.94	41.21	49.05	43.33	48.60

Çizelge 4.19. Erkek ünlü seslerin ortalama sürelerinin çoklu karşılaştırma sonuçları

	Nötr	Üzgün	Mutlu	Kızgın
Nötr	-	I	A,E	A,E,I,İ
Üzgün	I	-	A,E	A,E
Mutlu	A,E	A,E	-	
Kızgın	A,E,I,İ	A,E		-

Çizelge 4.20. Kadın ünlü sürelerinin çoklu karşılaştırma sonuçları

	Nötr	Üzgün	Mutlu	Kızgın
Nötr	-		A,E,I,İ	A,E,I,İ
Üzgün		-	A,E	A,E
Mutlu	A,E,I,İ	A,E	-	
Kızgın	A,E,I,İ	A,E		-

Duygunun ünlü seslerin süreleri üzerine etkilerini incelemek ve süreler arasındaki farkların istatistiksel olarak anlamlı olup olmadığını belirlemek için Tukey çoklu karşılaştırma testi uygulanmıştır. Çizelge 4.19. ve Çizelge 4.20.'de duygu sınıfları arasında ortalama süre farkları istatistiksel olarak anlamlı olan ünlü sesler verilmiştir. Tablolar incelendiğinde hem kadın konuşmacılar hem de erkek konuşmacılar için özellikle /a/ ve /e/ seslerinin sürelerinin nötr ve üzgünle karşılaştırıldığında istatistiksel olarak daha uzun olduğu görülmektedir.

4.3. Türkçe Konuşmadan Duygu Tanıma ve Duygu Boyutlarının Kestirimi

4.3.1. Duygu Tanıma ve Duygu Boyutları Kestiriminde Kullanılan Akustik Öznitelikler

Temel frekans (F0-perde) eğrisi, enerji eğrisi, sıfır geçiş oranları (ZCR) ve sesli olma olasılıkları (VoProb) prosodik bilgi olarak, spektral özellik olarak ise konuşma tanıma çalışmalarında yaygın olarak kullanılan mel frekansı spektral katsayılarından (MFKK) yararlanılmıştır (İlk 12 MFKK ve birinci türevleri). Yapılan çalışmada bu akustik parametrelerden sözlü ifade düzeyindeki OpenSMILE (Eyben, 2010) yazılımı ile çıkardığımız 384 tane istatistik, akustik öznitelik olarak duygu kategorileri ve boyutlarının kestiriminde kullanılmıştır. Akustik parametrelerin birinci türevlerine ait istatistikler de hesaplanarak akustik öznitelik setine eklenmiştir. Kullanılan özniteliklerin listesi aşağıdaki gibidir.

Temel Frekans (F0) eğrisi (24 adet): ortalama, maksimum, minimum, standart sapma, skewness, kurtosis, range, maksimum değer yeri (frame olarak), minimum değer yeri (frame olarak) ve lineer regresyon katsayıları (linregc1, linregc2 ve linregerrQ)

Enerji eğrisi (24 adet): ortalama, maksimum, minimum, standart sapma, skewness, kurtosis, range, maksimum değerin yeri (frame olarak), minimum değerin yeri (frame olarak) ve lineer regresyon katsayıları (linregc1, linregc2 ve linregerrQ)

MFKK (1-12) (288 adet): ortalama, maksimum, minimum, standart sapma, skewness, kurtosis, range, maksimum değerin yeri (frame olarak), minimum değerin yeri (frame olarak) ve lineer regresyon katsayıları (linregc1, linregc2 ve linregerrQ)

Sıfır Geçiş oranı (ZCR) (24 adet): ortalama, maksimum, minimum, standart sapma, skewness, kurtosis, range, maksimum değerin yeri (frame olarak), minimum değerin yeri (frame olarak) ve lineer regresyon katsayıları (linregc1, linregc2 ve linregerrQ)

Voicing Probability (VoProb) (24 adet): ortalama, maksimum, minimum, standart sapma, skewness, kurtosis, range, maksimum değerin yeri (frame olarak), minimum değerin yeri (frame olarak) ve lineer regresyon katsayıları (linregc1, linregc2 ve linregerrQ)

4.3.2. Kategorik Modelin Başarım Değerlendirme Sonuçları

Her sınıflandırma probleminde olduğu gibi sınıflandırma için en uygun öznitelikleri belirlemek sınıflandırıcının performansını doğrudan etkilemektedir. Kullanılan öznitelik sayısını azaltmak sistem performansını artıracak gibi öznitelik çıkarımından modelin oluşturulmasına kadar birçok aşamada hesaplama karmaşıklığını ve süresini de azaltacaktır. Ayrıca, öznitelik değerlendirme yöntemleri sayesinde hangi akustik parametrelerin duygu sınıflandırmasında önemli olduğu hakkında da bilgi sahibi olunabilecektir. Bu amaçla konuşma sinyallerinden çıkardığımız 384 adet akustik özniteliği sayısını CFS, RELIF-F ve InfoGain öznitelik seçme ve değerlendirme yöntemlerini uygulanarak kullanılan öznitelik sayısı ve sınıflandırma gücü en fazla olan akustik parametreler belirlenmeye çalışıldı.

İlk olarak öznitelik altkümesi seçme yöntemi olan CFS veri setlerine 10-kat çapraz doğrulama işlemi kullanılarak uygulandı. Bütün çapraz doğrulama işlemlerinde de seçilen özniteliklerin listesi Çizelge 4.21.'de verilmiştir. CFS, öznitelikleri tek tek değil öznitelik alt kümelerini değerlendirdiğinden önem sıralaması yapmamıza imkân vermemektedir. Bu yöntem özniteliklerden birbirleriyle ilintisiz ve aynı zamanda sınıfla yüksek ilintili olanlar kümesini belirlemeye çalıştığından, özniteliklerin birbirlerinden

bağımsız olduğu kabullenmesi yapan Naive Bayes gibi sınıflandırıcıların performansını artırması beklenmektedir.

Çizelge 4.21. CFS yöntemiyle 10-kat çapraz doğrulama işlemi yapılarak gerçekleştirilen öznelik seçimi sonuçları

	Negatif-Negatif Değil	Kızgın-Kızgın Değil
F0	min, mean, median, std, range, reg_coef1	min, mean, median, std, range, reg_coef2
F0_de	min, max, median, std, reg_coef1, reg_coef2	min, max, median, std, reg_coef1, reg_coef2
Energy	mean, linregc2, linregerrQ, stddev,	mean, linregc2, linregerrQ, stddev,
Energy_de	min, lingregc2, stddev, kurtosis	lingregc2, stddev, kurtosis
MFCC	[1]_max, [1]_linregc2, [1]_linregerrQ, [1]_stddev, [2]_max, [2]_min, [2]_mean, [2]_lingregc, [3]_mean, [6]_max, [10]_min, [12]_mean, [12]_stddev	[1]_max, [1]_linregc2, [1]_linregerrQ, [1]_stddev, [2]_maxPOS, [2]_min, [2]_mean, [3]_mean, [6]_max, [12]_mean, [12]_stddev
MFCC_de	[1]_kurtosis, [2]_min, [2]_linregc2, [6]_max, [12]_min	[1]_linregc1, [1]_linregc2, [1]_kurtosis, [2]_min, [2]_linregc2, [6]_max, [12]_min

CFS yöntemi kullanılarak *kızgın-kızgın değil* sınıflandırması için öznelik sayısı 67'ye *negatif-negatif değil* için 55'e indirilmiştir.

İkinci olarak özneliklerin duygu sınıflarını ayırımındaki önem sıralarını belirlemek için RELIF-F ve InfoGain yöntemlerini kullandık. Her iki sınıflandırma problemi için bu yöntemler kullanılarak yapılan sıralamalar (ilk 20) Çizelge 4.22.'de verilmiştir. Hem *kızgın-kızgın değil* hem de *negatif-negatif değil* sınıflandırma için F0 ve enerji öznelikleri listenin en üstünde yer almaktadır. Kızgınlık, korku gibi yoğun (kuvvetli) duyguların nötr konuşma sinyaline göre daha yüksek perde ve enerjiyle modüle edildiği göz önüne alındığında sıralamanın akustik analizlerle uyumlu olduğunu söyleyebiliriz. Sınıflandırmada çalışmasında sıralamadaki ilk 100 öznelik kullanılmıştır.

Çizelge 4.22. Relief-F ve Infogain öznelik değerlendirme yöntemleriyle seçilen en iyi ilk 20 öznelik (Sıralama en iyiden başlanarak yapılmıştır)

Kızgın-Kızgın Değil		Negatif- Negatif Değil	
RELIEF-F	InfoGain	RELIEF-F	InfoGain
Energy_mean	F0_mean	F0_max	F0_mean
F0_max	Energy_mean	F0_mean	F0_median
F0_mean	F0_median	Energy_mean	Energy_mean
Energy_de_stddev	F0_de_std	F0_median	F0_max
F0_median	Energy_stddev	mfcc[2]_mean	F0_de_std
Energy_stddev	F0_max	Energy_de_stddev	Energy_stddev
mfcc[2]_mean	Energy_linregerrQ	Energy_stddev	Energy_linregerrQ
mfcc[2]_min	Energy_de_stddev	mfcc[2]_min	Energy_de_stddev
F0_range	Energy_de_linregerrQ	F0_range	mfcc[2]_mean
Energy_max	Energy_de_linregc2	mfcc[2]_linregc2	F0_std
mfcc[2]_max	F0_std	mfcc[1]_stddev	F0_range
mfcc[1]_stddev	F0_range	mfcc[2]_max	Energy_de_linregerrQ
Energy_range	mfcc[2]_mean	Energy_linregc2	Energy_de_linregc2
mfcc[6]_mean	Energy_linregc2	Energy_max	F0_reg_coef1
Energy_linregc2	mfcc[2]_min	Energy_range	mfcc[2]_min
F0_min	F0_de_reg_coef1	F0_reg_coef1	Energy_linregc2
mfcc[2]_linregc2	F0_reg_coef1	mfcc[11]_linregc2	F0_de_range
mfcc[9]_mean	F0_de_range	F0_min	F0_de_reg_coef1
mfcc[11]_linregc2	Energy_range	mfcc_de[5]_min	F0_de_min
mfcc[8]_mean	Energy_max	mfcc[6]_mean	Energy_range

Çizelge 4.23. Negatif-Negatif Değil sınıflandırma için toplam doğruluk sonuçları. UA: ağırlıksız toplam doğruluk, WA: ağırlıklı toplam doğruluk

	CFS		InfoGain		RELIEF-F		TÜMÜ	
	UA	WA	UA	WA	UA	WA	UA	WA
BayesNet	0,75	0,76	0,74	0,74	0,74	0,74	0,74	0,74
NaiveBayes	0,74	0,76	0,72	0,74	0,73	0,74	0,73	0,74
DVM	0,76	0,79	0,76	0,78	0,77	0,79	0,76	0,78
Temel Referans	0,50	0,61	0,50	0,61	0,50	0,61	0,50	0,61

Çizelge 4.24. Negatif-Negatif Değil sınıflandırma problemi her bir sınıf için doğruluk oranı, keskinlik, duyarlılık ve F-ölçütü sonuçları

		CFS			InfoGain			RELIEF-F			Tümü		
		r_i	p_i	F_i	r_i	p_i	F_i	r_i	p_i	F_i	r_i	p_i	F_i
Bayes Net	Negatif Değil	0,76	0,82	0,79	0,74	0,82	0,78	0,74	0,81	0,77	0,75	0,82	0,78
	Negatif	0,74	0,67	0,70	0,74	0,65	0,69	0,73	0,64	0,68	0,74	0,65	0,69
Naive Bayes	Negatif Değil	0,82	0,79	0,80	0,80	0,78	0,79	0,78	0,79	0,79	0,79	0,78	0,79
	Negatif	0,66	0,70	0,68	0,64	0,67	0,66	0,68	0,66	0,67	0,66	0,67	0,66
DVM	Negatif Değil	0,89	0,79	0,84	0,88	0,79	0,83	0,87	0,80	0,83	0,88	0,79	0,83
	Negatif	0,63	0,78	0,70	0,63	0,77	0,70	0,66	0,76	0,70	0,64	0,78	0,70

Negatif-Negatif değil sınıflandırma sonuçları ağırlıksız toplam, doğruluk ve ağırlıklı toplam doğruluk şeklinde Çizelge 4.23.'te verilmiştir. Ayrıca her bir sınıfa ait keskinlik, duyarlılık ve F-ölçütü sonuçları Çizelge 4.24.'te sunulmuştur. Öznitelik sayısı azaltılarak sınıflandırıcı performanslarında çok büyük artış sağlanmadıysa da, tüm öznitelikler kullanılarak elde edilen sonuçlardan daha iyi veya aynı sonuçlar elde edilerek başarı sağlanmıştır. Daha az sayıda öznitelik kullanmak sınıflandırma modelinin karmaşıklığını ve öznitelik çıkarım süresini azaltacağından özellikle gerçek-zamanlı çalışacak sistemlerin performansını doğrudan etkileyecektir. Yapılan çalışmada, DVM kullanılarak %79 ağırlıklı toplam doğruluk ve %77 ağırlıksız toplam doğruluk sınıflandırma performansı elde edilmiştir.

Çizelge 4.25. Kızgın-Kızgın Değil sınıflandırma için toplam doğruluk (overall accuracy) sonuçları. UA: ağırlıksız toplam doğruluk, WA: ağırlıklı toplam doğruluk

	CFS		InfoGain		RELIEF-F		TÜMÜ	
	UA	WA	UA	WA	UA	WA	UA	WA
BayesNet	0,75	0,75	0,73	0,73	0,72	0,72	0,73	0,73
NaiveBayes	0,72	0,72	0,72	0,71	0,71	0,71	0,71	0,71
DVM	0,73	0,78	0,74	0,79	0,75	0,79	0,75	0,79
Temel Referans	0,50	0,64	0,50	0,64	0,50	0,64	0,50	0,64

Çizelge 4.26. Kızgın-Kızgın Değil sınıflandırma problemi her bir sınıf için doğruluk oranı, keskinlik, duyarlılık ve F-ölçütü sonuçları

		CFS			InfoGain			RELIEF-F			Tümü		
		r_i	p_i	F_i	r_i	p_i	F_i	r_i	p_i	F_i	r_i	p_i	F_i
Bayes Net	Kızgın Değil	0,75	0,84	0,79	0,73	0,83	0,78	0,73	0,82	0,77	0,73	0,83	0,78
	Kızgın	0,74	0,61	0,67	0,73	0,59	0,65	0,71	0,59	0,64	0,73	0,59	0,65
Naive Bayes	Kızgın Değil	0,80	0,80	0,80	0,79	0,80	0,79	0,76	0,80	0,78	0,78	0,80	0,79
	Kızgın	0,63	0,63	0,63	0,63	0,62	0,62	0,65	0,59	0,62	0,63	0,61	0,62
DVM	Kızgın Değil	0,90	0,79	0,84	0,91	0,80	0,85	0,91	0,80	0,85	0,89	0,80	0,84
	Kızgın	0,56	0,76	0,64	0,57	0,77	0,65	0,58	0,78	0,66	0,60	0,74	0,66

Kızgın-Kızgın değil sınıflandırma sonuçları ağırlıksız toplam doğruluk ve ağırlıklı toplam doğruluk şeklinde Çizelge 4.25.'te verilmiştir. Ayrıca her bir sınıfa ait keskinlik, duyarlılık ve F-ölçütü sonuçları Çizelge 4.26.'da sunulmuştur. Çizelge 4.25. ve Çizelge 4.26.'dan anlaşılacağı üzere öznitelik sayısını azaltmak sınıflandırıcı performanslarında düşüşe sebep olmamaktadır. Yapılan çalışmada en iyi performans %79 ağırlıklı toplam doğruluk ve %75 ağırlıksız toplam doğruluk oranlarıyla DVM sınıflandırıcı ve RELIEF-F yöntemiyle azaltılan öznitelikler kullanıldığında elde edilmiştir. Literatürdeki benzer çalışmalarda sadece akustik bilgiyle elde edilen performans oranları %60 ile %75 aralığında değişmektedir.

4.3.3. Duygu Boyutları Kestirim Sonuçları

Konuşma sinyallerinden çıkardığımız akustik özniteliklerin duygu boyutlarını kestirmede ki performansları Çizelge 4.27.'de verilmiştir. Kestirim işleminde 10-kat çapraz doğrulama işlemi kullanılmıştır. Bu yöntemde veriler rastgele olarak 10 eşit parçaya bölünür ve her defasında bir parça test için geriye kalan 9 parça da eğitim için kullanılır. Böylece, eğitim-test işlemi 10 kere tekrarlanarak bütün veri seti test edilmiş olur.

Çizelge 4.27. Destek vektör regresyon sonuçları

	Aktivasyon			Değerlik			Baskınlık		
	CC	MAE	RMSE	CC	MAE	RMSE	CC	MAE	RMSE
F0	0,56	0,18	0,24	0,29	0,16	0,22	0,51	0,14	0,19
MFCC	0,65	0,17	0,22	0,24	0,17	0,23	0,64	0,13	0,17
Energy	0,59	0,18	0,23	0,16	0,16	0,23	0,62	0,13	0,17
ZCR	0,48	0,19	0,25	0,12	0,16	0,23	0,49	0,15	0,19
voiceProb	0,38	0,20	0,26	0,11	0,17	0,23	0,35	0,16	0,21
Hepsi	0.74	0.15	0.19	0.34	0.16	0.22	0.74	0.12	0.15

Çizelge 4.27.'deki sonuçlar incelendiğinde, aktivasyon ve baskınlık değerlerini kestirimde MFCC daha yüksek kestirim gücüne sahipken değerlilik için F0'ın daha yüksek kestirim gücüne sahip olduğu görülmektedir. Yüksek CC ve aynı zamanda düşük MAE ve RMSE değerleri iyi kestirim performansı anlamına gelir. Enerji aktivasyon ve baskınlığın kestiriminde F0'a oranla daha iyi performans göstermektedir. Sonuçlar genel olarak değerlendirildiğinde, akustik özneliklerin duygu boyutları kestirimindeki önemleri şu şekilde sıralanabilir. Değerlilik için F0-MFCC-Enerji-ZCR-voiceProb, aktivasyon için MFCC-Enerji-F0-ZCR-voiceProb, baskınlık için MFCC-Enerji-F0-ZCR-voiceProb. Bütün boyutlar için en iyi sonuçların bütün akustik öznelikler kullanıldığında elde edildiği de görülmektedir. Yine tabloda verilen sonuçlar incelendiğinde akustik özneliklerin aktivasyon ve baskınlığı kestirimde değeri kestirime göre çok daha başarılı olduğu gözlenmektedir.

5. SONUÇ VE ÖNERİLER

Tez kapsamında elde edilen sonuçlar incelendiğinde oluşturulan veritabanının oldukça fazla sayıda konuşmacı barındırması ve konuşma işaretlerinin farklı kayıt ortamlarından alınmış olmasına rağmen değerlendirme ve kestirim sonuçlarındaki başarısının oldukça iyi bir seviyede olduğu görülebilmektedir.

Gerçekleştirilen akustik analiz sonuçları doğrultusunda çeşitli özniteliklerin, duygu kategorileri ile ünlü/sesli harflerle olan ilişkileri istatistiksel olarak ispatlanmış ve duyguların konuşma işareti üzerindeki etkileri ve nasıl kodlandığı gibi konulara açıklık getirilmeye çalışılmıştır.

Yapılan diğer bir çalışmada öznitelik seçimi yöntemlerinin kullanımı ile öznitelik vektörleri ve duygu kategorileri kestiriminde öznitelik seçiminin performansı artırıcı yönde bir etki yaptığını göstermiştir. Ayrıca bu bölümde sınıflandırıcıların performansları da test edilmiştir.

Yapılan tüm bu çalışmalar sonucunda duygunun konuşma sinyali üzerindeki etkileri, öznitelik seçimi ve sınıflandırıcı tasarımı konularına açıklık getirilerek gerçek zamanlı duygu tanıma sistemi konusuna alt yapı hazırlanmıştır. Gelecekte yapılacak çalışmalarda mevcut veritabanının kullanılmasının yanı sıra elde edilen sonuçlar doğrultusunda Türkçe konuşma için birçok duygu kategorisini gerçek zamanlı olarak analiz edebilen ve tanıyabilen çalışmalar gerçekleştirilebilecektir. Ayrıca elde edilen analiz sonuçları ışığında duygusal konuşma sentezleme üzerine de çalışmalar gerçekleştirmek mümkün hale getirilmiştir.

KAYNAKLAR

- Batliner, A., Steidl, S., Schuller, B., Seppi, D., Laskowski, K., Vogt, T., Devillers, L., Vidrascu, L., Amir, N., Kessous, L., and Aharonson, V., 2006. Combining Efforts for Improving Automatic Classification of Emotional User States. **IS-LTC 2006**, 240-245, Ljubljana.
- Bouckaert, R., R., 1995. Bayesian Belief Networks: from Construction to Inference, **Ph.D. thesis**. University of Utrecht.
- Bozkurt, E., Eroglu Erdem, C., Erdem T., and Erzin, E., 2011. Formant Position based Weighted Spectral Features for Emotion Recognition. **Speech Communication**, 53:1186-1197.
- Bradley, M., Lang, P., 1994. Measuring emotion: The selfassessment manikin and the semantic differential. **Journal of Behavioral Therapy and Experimental Psychiatry**, 25:49-59.
- Burkhardt, F., Paeschke, A., Rolfes, M., Sendlmeier, W., and Weiss, B., 2005. A Database of German Emotional Speech. **Interspeech**, 1517-1520, Lisbon.
- Clavel C., Vasilescu, I., Devillers, L., Ehrette, T., Richard, G., 2006. Safe corpus: fear-type emotions detection for surveillance application. **LREC**, 1099- 1104, Genoa.
- Clavel C., Vasilescu, I., Devillers, T., Richard, L., Ehrette, 2008. Fear-type emotion recognition for future audio-based surveillance systems. **Speech Communication**, 50(6):487-503.
- Devillers, L., Vasilescu I., 2004. Reliability of Lexical and Prosodic Cues in two Real-life Spoken Dialog Corpora. **LREC**.
- Douglas-Cowie, E., Cowie, R., Schroeder, M., 2000. A new emotion database: considerations, sources and scope. **ISCA ITRW on Speech and Emotion**, 39-44, Belfast.
- Douglas-Cowie, E., Campbell, N., Cowie, R., and Roach, P., 2003. Emotional speech: Towards a new generation of databases. **Speech Communication**, 40(1-2):33-60.
- Douglas-Cowie, E., Cowie, R., Sneddon, I., Cox, C., Lowry, O., McRorie, M., Martin, J.C., Devillers, L., Abrilan, S., Batliner, A., Amir, N., and Karpousis, K., 2007. The HUMAINE Database: Addressing the Collection and Annotation of Naturalistic and Induced Emotional Data. **Affective Computing and Intelligent Interaction**.(A. Paiva, R. Prada, and R. W. Picard, Editör). 488-500, Berlin.
- Engbert, I., S. ve Hansen A.V., 2007. Documentation of the danish emotional speech database (DES). **Center for PersonKommunikation, Tech. Rep., Denmark**.
- Espinosa, H.P., Garcia, C.A.R., and Pineda, L.V., 2010. Features selection for primitives estimation on emotional speech. **IEEE ICASSP**, 5138–5141.
- Eyben F., Wöllmer, M., Schuller, B., 2010. openSMILE - The Munich Versatile and Fast Open-Source Audio Feature Extractor. **ACM Multimedia (MM)**, 25-29, Italy.
- Fleiss, J.L., 1971. Measuring nominal scale agreement among many raters. **Psychological Bulletin**, 76(5):378-382.
- France, D., Shiavi, R., Silverman, S., Silverman, M., Wilkes, D., 2003. Acoustical properties of speech as indicators of depression and suicidal risks. **IEEE Trans. Biomedical Eng.**, 47 (7):829-837.
- Goudbeek, M., Goldman, J.P., Scherer, K.R., 2009. Emotion dimensions and formant position, **INTERSPEECH-2009**, 1575-1578.

- Grimm, M., Mower, E., Kroschel, K., and Narayanan, S., 2006. Combining categorical and primitives-based emotion recognition. **European Signal Processing Conference (EUSIPCO)**.
- Grimm, M., Kroschel, K., Mower, E., Narayanan, S., 2007. Primitives-based evaluation and estimation of emotions in speech. **Speech Communication** 49(1):787-800.
- Grimm, M., Kroschel, K., and Narayanan, S., 2008. The Vera am Mittag German Audio-Visual Emotional Speech Database. **ICME**, 865-868, Germany.
- Hall, M., A., 1999. Correlation-based feature selection for machine learning, **Doctoral Dissertation, The University of Waikato, Department of Computer Science**.
- Hansen J. and Bou-Ghazale S., 1997. Getting started with susas: A speech under simulated and actual stress database. **EUROSPEECH-97**, 4:1743-1746, Greece.
- Hearst, M.A., 1998. Trends and controversies: Support vector machines. **IEEE Intelligent Systems**, 13(4):18–28.
- Humaine, 2011. <http://emotion-research.net/wiki/Databases>.
- John, G.H., Langley, P., 1995. Estimating Continuous Distributions in Bayesian Classifiers. **Eleventh Conference on Uncertainty in Artificial Intelligence**.338-345, San Mateo.
- Kononenko, I., 1994. Estimating Attributes: Analysis and Extensions of RELIEF. **European Conference on Machine Learning**, 171-182.
- Lee, C.M., and Narayanan, S., 2005. Toward detecting emotions in spoken dialogs. **IEEE Transactions on Speech and Audio Processing**, 13(2):293-303.
- Meral, H.M., Ekenel, H.K., Özsoy A.S., 2006. Türkçede Duygu Çözümlemesi. **Ç.Ü. Türkoloji Araştırmaları Merkezi**.
- Metze, F., Englert, R., Bub, U., Burkhardt, F. and Stegmann, J., 2009. Getting closer: tailored human-computer speech dialog. **Universal Access in the Information Society**, 8(2):97-108.
- Oflazoglu, C., Yildirim, S., 2011. Turkish emotional speech database, **Signal Processing and Communications Applications (SIU)**, 2011 IEEE 19th Conference on, 1153-1156, Antalya.
- Polzehl, T., Schmitt, A., Metze, F., and Wagner, M., 2011. Anger recognition in speech using acoustic and linguistic cues, in **Speech Communication**, 53(9-10):1198-1209.
- Polzin, T.S., Waibel, A., 2000. Emotion-sensitive humancomputer interfaces. **ISCA ITRW on Speech and Emotion**, 201-206, Belfast.
- Quinlan, R.R., 1986. Induction of decision trees. **Machine Learning**, 1:81–106.
- Russell, J.A., 2003. Core affect and the psychological construction of emotion. **Psychological Review**, 110:145- 172.
- Scherer, K., Ceschi, G., 2000. Studying affective communication in the airport: The case of lost baggage claims. **Personality and Social Psychological Bulletin** 26(93): 327- 339.
- Schlosberg, H., 1954. Three dimensions of emotion, **Psychological Review**, 61(2):81-88.
- Schuller, B., Wimmer, M., Arsic, D., Rigoll, G., and Radig, B., 2007. Audiovisual behavior modeling by combined feature spaces. **ICASSP 2007**, II:733-736, USA
- Schuller, B., Vlasenko, F., Eyben, B., Rigoll, G., and Wendemuth, A., 2009a. Acoustic Emotion Recognition: A Benchmark Comparison of Performances. **Automatic Speech Recognition and Understanding Workshop (ASRU 2009)**, 552-557, Italy.

- Schuller, B., Müller, R., Eyben, F., Gast, J., Hörnler, B., Wöllmer, M., Rigoll, G., Höthker, A., and Konosu, H., 2009b. Being Bored? Recognising Natural Interest by Extensive Audiovisual Integration for Real- LifeApplication. **Image and Vision Computing Journal (IMAVIS)**.
- Schuller, B., Vlasenko, B., Eyben, F., Wöllmer, M., Stuhlsatz, A., Wendemuth, A., and Rigoll, G., 2010. Cross-Corpus Acoustic Emotion Recognition: Variances and Strategies. **IEEE Transactions on Affective Computing**, 1(2):119-131.
- Sjolander, K., 2004. Snack sound toolkit, KTH Stockholm. <http://www.speech.kth.se/snack>, Sweden.
- Steininger, S., Schiel, F., Dioubina, O., and Raubold S., 2002. Development of user-state conventions for the multimodal corpus in smartkom. **Workshop on Multimodal Resources and Multimodal Systems Evaluation**, 33-37, Las Palmas.
- Talkin, D., 1995. A robust algorithm for pitch tracking (RAPT).(W. B. Kleijn and K. K. Paliwal, Editör). **Speech Coding and Synthesis,Elsevier Science**, 497-518.
- Ververidis, D. and Kotropoulos, C., 2003. A review of emotional speech databases. **PCI 2003, 9th Panhellenic Conference on Informatics**, 560-574, Greece.
- Vapnik, V., Golowich, S., Smolo, A., 1997. Support Vector Method for Function Approximation, Regression Estimation and Signal Processing. (M. Mozer, M. Jordan, T. Petsche, Editör), Neural Information Processing Systems, **MIT Press**, 9, Cambridge, MA.
- Vidrascu L. and Devillers L., 2005. Detection of real-life emotions in call centers. **Eurospeech**, 1841-1844, Lisbon.
- Wu, D., Parsons, T.D., Narayanan, S.S., 2010. Acoustic Feature Analysis in Speech Emotion Primitives Estimation. **Proceedings of Interspeech**, 785-788.
- Whissel, C.M., 1989. The dictionary of affect in language, Emotion: theory, research, and experience.(R. Plutchik, H. Kellerman, Editör) **Academic Press**, 113-131, New York.
- Yıldırım, S., Narayanan, S., Potamianos, A., 2011. Detecting emotional state of a child in a conversational computer game. **Computer Speech & Language**, Vol. 25(1):29-44.
- Young, S., Evermann, G., Gales, M., Hain, T., Kershaw, D., Liu, X., Moore, G., Odell, J., Ollason, D., Povey, D., Valtchev, V., and Woodland, P., 2009. **The HTK Book(for HTK Version 3.4)**, Cambridge University Engineering Depart.

TEŞEKKÜR

Tez çalışmam sırasında bana araştırma olanağı sağlayan ve her türlü maddi ve manevi yardımı benden esirgemeyen çok değerli danışman hocam sayın Yrd. Doç. Dr. Serdar YILDIRIM'a sonsuz teşekkür ederim.

Yüksek lisans öğrenimim boyunca teknik konularda bana destek veren ve her türlü konuda fikirlerini benden esirgemeyen sayın Yrd. Doç. Dr. Esen YILDIRIM'a, sayın Yrd. Doç. Dr. Ersin ÖZDEMİR'e ve arkadaşlarım Arş. Gör. Ozan ÇAKMAK'a, Arş. Gör. Ekin ARSLAN'a, Sinem UZUN'a, Fırat DUMAN'a ve ayrıca tüm Elektrik-Elektronik/Bilgisayar mühendisliği bölümü öğretim üyelerine en içten teşekkürlerimi sunarım.

Tez çalışmalarım sırasında maddi ve manevi desteğini esirgemeyen, hayatımın her aşamasında bana destek olan aileme ve arkadaşım Demet DOKSÖZ'e teşekkürlerimi sunarım.

ÖZGEÇMİŞ

1987 yılında Antakya’da doğdum. İlk ve orta öğrenimimi Antakya’da tamamladım. 2005 yılında Elazığ Fırat Üniversitesi Elektrik-Elektronik Mühendisliği bölümünü kazandım ve mühendislik eğitimimi 2009 yılında aynı üniversitede tamamladım. 2009 yılının güz döneminde Mustafa Kemal Üniversitesi Fen Bilimleri Enstitüsü Elektrik-Elektronik Mühendisliği Anabilim Dalında yüksek lisans eğitimime başladım. 2009-2010 yılı güz döneminde Hatay Mustafa Kemal Üniversitesi Harbiye Meslek Yüksek Okulunda öğretim görevlisi olarak derslere girdim. 2010 yılının Mart ayında TÜBİTAK destekli olan “Türkçe Konuşma İçin Gerçek Zamanlı Duygu Tanıma Sistemi Tasarımı” adlı projeye burslu öğrenci olarak dahil oldum ve sinyal işleme üzerine çalışmalar gerçekleştirmeye başladım. Projeye başlamamızın ilk yılının sonunda IEEE 19. Sinyal İşleme ve İletişim Uygulamaları Kurultayına “Türkçe Duygusal Konuşma Veritabanı” isimli bildiri ile katıldım. TÜBİTAK destekli olan projeye burslu öğrenci olarak devam etmekteyim.