

**BAŐKENT ÜNİVERSİTESİ
FEN BİLİMLERİ ENSTİTÜSÜ**

**İŐİTSEL SAHNELERİN TANINMASI İÇİN ÇEVRESEL SES
ANALİZİ**

SELVER EZGİ KÜÇÜKBAY

**YÜKSEK LİSANS TEZİ
2015**

**İŐİTSEL SAHNELERİN TANINMASI İÇİN ÇEVRESEL SES
ANALİZİ**

**AUDIO EVENT ANALYSIS FOR AUDITORY SCENE
RECOGNITION**

SELVER EZGİ KÜÇÜKBAY

Başkent Üniversitesi
Lisansüstü Eğitim Öğretim ve Sınav Yönetmeliğinin
BİLGİSAYAR Mühendisliğı Anabilim Dalı İçin Öngördüğü
YÜKSEK LİSANS TEZİ
olarak hazırlanmıştır.

2015

“İşitsel Sahnelerin Tanınması İçin Çevresel Ses Analizi” başlıklı bu çalışma, jürimiz tarafından 04/02/2015 tarihinde, **BİLGİSAYAR MÜHENDİSLİĞİ ANABİLİM DALI** 'nda **YÜKSEK LİSANS TEZİ** olarak kabul edilmiştir.

Başkan : Prof. Dr. Buyurman BAYKAL

Üye (Danışman) : Yrd. Doç. Dr. Mustafa SERT

Üye : Prof. Dr. Adnan YAZICI

ONAY

..../02/2015

Prof. Dr. Emin AKATA
Fen Bilimleri Enstitüsü Müdürü

TEŞEKKÜR

Üniversite hayatına adım attığım ilk günden beri sonsuz desteğini hissettiren, bana her zaman güvenen, her zorluğu kolayca aşabileceğime inanan ve beni inandıran danışmanım Sayın Yrd. Doç Dr. Mustafa SERT'e değerli katkılarından dolayı,

Sonsuz sevgisi ve anlayışı ile hayatımı kolaylaştırdığı ve sabırla tüm sorunlarıma ortak olduğu için hayat arkadaşım ve sevgili eşim Serkan KÜÇÜKBAY'a,

Hayatta merhamet ve iyi niyetin her şeyden daha üstün olduğunu öğreten babam Hasan YALNIZ'a, azim ve disiplinle perçinlenmiş başarı kavramını bana aşıl原因 annem İpek YALNIZ'a ve önümde iyi bir rol model olarak beni çalışmaya şevk eden abim Özgür İlhan YALNIZ'a

Akademik başarıları ile bana ışık tutan, bilgi ve çalışmalarını hayranlıkla takip ettiğim, mesleğimi bana sevdiren, bana güvenen, inanan ve her zaman yanımda olan babam Prof. Dr. Hasan KÜÇÜKBAY ve annem Doç. Dr. F. Zehra KÜÇÜKBAY'a, bilgisi ve özgüveni ile bana her zaman destek olarak, tüm stresli anlarımda profesyonelce beni yönlendiren kardeşim Furkan KÜÇÜKBAY'a,

Hayatıma girdiği günden beri tüm sorunlarıma benimle birlikte çözüm bulduğu, beni anladığı, dinlediği ve en önemlisi hayatıma neşe kattığı için canım oda arkadaşım Hatice TOMBUL'a ve diğer tüm dostlarıma,

TEŞEKKÜR EDERİM.

ÖZ

İŞİTSEL SAHNELERİN TANINMASI İÇİN ÇEVRESEL SES ANALİZİ

Selver Ezgi KÜÇÜKBAY

Başkent Üniversitesi Fen Bilimleri Enstitüsü

Bilgisayar Mühendisliği Anabilim Dalı

Ses verileri, içerisinde birçok ses türünü barındırır ve çokluortam uygulamaları için önemli bir kaynaktır. Bu çalışmada ofis ortamından alınmış 16 ayrı ses olayının (*alarm, boğaz temizleme, öksürük, kapı çarpması, çekmece, klavye, anahtar, kapı vurma, gülme, bilgisayar faresi, sayfa çevirme, nesnenin masaya çarpması, telefon, yazıcı, konuşma ve elektrik düğmesi*) analizi ve sezimi sağlanmıştır. Bununla birlikte 10 işitsel sahnenin (*otobüs, kalabalık sokak, ofis ortamı, açık market, park, sessiz sokak, restoran, süpermarket, tren ve tren istasyonu*) ayrı olarak sınıflandırılması gerçekleştirilmiştir. Ayrıca, sezimlenen ses olayları ile veri kümesindeki “*ofis*” işitsel sahnesi arasındaki ilintiler incelenmiştir. Ses örneklerinin içerik tanımlaması için Mel frequency cepstral coefficient (MFCC) özneliği, Destek Vektör Makineleri (DVM) öğrenme algoritması ile birlikte kullanılmıştır. MFCC ve DVM için parametre eniyilemesi yapılmış ve öznelik çıkarımı için uyarlamalı bir frekans analiz yöntemi önerilmiştir.

ANAHTAR SÖZCÜKLER: Ses olay sezimi, işitsel sahne tanıma, ses içerik analizi, MFCC, DVM.

Danışman: Yrd.Doç.Dr. Mustafa SERT, Başkent Üniversitesi, Bilgisayar Mühendisliği Bölümü.

ABSTRACT

AUDIO EVENT ANALYSIS FOR AUDITORY SCENE RECOGNITION

Selver Ezgi KÜÇÜKBAY

Başkent University Institute of Science

Department of Computer Engineering

Audio data contains several sound types and is important source for multimedia applications. In this thesis, we present a system for analysing and detecting 16 distinct audio events namely; *alert, clear throat, cough, door slam, drawer, keyboard, keys, knock, laughter, mouse, pageturn, pen drop, phone, printer, speeh* and *switch* that are collected from office live environments. The recognition of 10 different auditory scenes (*bus, busy street, office, open airmarket, park, quiet street, restaurant, supermarket, tube and tubestation*) is also performed in the study. Moreover, correlations between audio scenes and audio events are investigated. Support Vector Machine (SVM) classifier along with the Mel Frequency Cepstral Coefficient (MFCC) feature are used through the analyses. In addition, we propose an adaptive frequency analysis scheme for feature extraction and perform optimizations for feature representation and classifier design.

KEYWORDS: Audio event detection, audio scene recognition, audio content analysis, MFCC, SVM.

Supervisor: Asst. Prof. Dr. Mustafa SERT, Başkent University, Department of Computer Engineering.

İÇİNDEKİLER LİSTESİ

ÖZ	i
ABSTRACT	ii
İÇİNDEKİLER LİSTESİ	iii
ÇİZELGELER LİSTESİ.....	v
KISALTMALAR.....	vii
1. GİRİŞ.....	1
1.1 Tezin Organizasyonu	3
2. LİTERATÜR TARAMASI	4
3. TEMEL TANIM VE KAVRAMLAR	10
3.1 Ses.....	10
3.2 Ses Olayı	10
3.3 İşitsel Sahne	10
3.4 Ses Özneliği.....	11
3.4.1 Mel ölçekli keprum katsayıları (Mel frequency cepstral coefficients) .	11
3.5 Sınıflandırma Yöntemi	12
3.5.1 Destek vektör makineleri	14
3.6 Çapraz Doğrulama.....	16
3.7 Performans Kriterleri	17
4. SES OLAY SEZİMİ	19
4.1. Öznelik Çıkarımı ve Gösterimi	20
4.2. Sınıflandırıcı Tasarımı.....	24
4.3 Deneysel Çalışmalar ve Değerlendirme.....	25
4.3.1 Ses olayları veri kümesi	26
4.3.2 Değerlendirme yöntemi	27
4.3.3 MFCC katsayı testleri	27
4.3.4 Pencere ve atlama süreleri testleri	28
4.3.5 Kesinlik-Duyarlılık grafikleri	31
5. İŞİTSEL SAHNE TANIMA	34
5.1 Öznelik Çıkarımı ve Gösterimi	34
5.3 Sınıflandırıcı Tasarımı.....	35
5.4 Deneysel Çalışmalar ve Değerlendirme.....	35
5.4.1 İşitsel sahneler veri kümesi	36
5.5 Ses Olayı – İşitsel Sahne İlinti Analizi.....	37

5.5.1 Yöntem.....	38
5.5.2 Deneysel sonuçlar.....	40
6. SONUÇLAR VE TARTIŞMA.....	43
KAYNAKLAR LİSTESİ	47
EKLER LİSTESİ	51

ÇİZELGELER LİSTESİ

Çizelge 3. 1	Hata matrisi.....	17
Çizelge 4. 1	Veri kümesindeki ses olayı sınıfları ve toplam süreleri.....	27
Çizelge 4. 2	Farklı pencere ve atlama süreleri için ölçülen performanslar	29

ŞEKİLLER LİSTESİ

Şekil 3. 1	Genel ses sınıfları	11
Şekil 3. 2	MFCC öznitelik vektörleri çıkarım aşamaları	13
Şekil 3. 3	DVM sınıflandırıcı hiper düzlem	14
Şekil 3. 4	5-Kat çapraz doğrulamanın bir döngüsü [36]	17
Şekil 4. 1	Ses olayı sezimi için önerilen sistemin blok şeması	19
Şekil 4. 2	Her bir ses olayı için baskın frekans değerleri.....	22
Şekil 4. 3	Gerçekleştirilen testler ve akışları	25
Şekil 4. 4	Farklı MFCC katsayıları için elde edilen başarımlar	28
Şekil 4. 5	Varsayılan frekans aralıkları ve sınıflara ait özel frekans aralıkları ile çıkarılan öznitelik vektörleriyle eğitilen sistemin performansı	30
Şekil 4. 6	Standart ve önerilen yöntem başarımları	31
Şekil 4. 7	Alarm, boğaz temizleme, öksürük ve kapı çarpması sınıfları için elde edilen kesinlik-duyarlılık grafiği.....	32
Şekil 4. 8	Çekmece, klavye, anahtar ve kapı vurma sınıfları için elde edilen kesinlik-duyarlılık grafiği	32
Şekil 4. 9	Gülme, bilgisayar faresi, sayfa çevirme ve nesnenin masaya çarpması sınıfları için elde edilen kesinlik-duyarlılık grafiği	33
Şekil 4. 10	Telefon, yazıcı, konuşma ve elektrik düğmesi sınıfları için elde edilen kesinlik-duyarlılık grafiği	33
Şekil 5. 1	İşitsel sahne sınıflarının baskın frekans değerleri	35
Şekil 5. 2	Sistemin çapraz doğrulama karşılaştırması	37
Şekil 5. 3	Ses olaylarının işitsel sahnelere olan katkısı için tasarlanan yapı	39
Şekil 5. 4	Ofis işitsel sahnesi için 30ms çerçeve boyutu ile gerçekleştirilen analizler	40
Şekil 5. 5	Ofis işitsel sahnesi için 90ms çerçeve boyutu ile gerçekleştirilen analizler	41
Şekil 5. 6	Ofis işitsel sahnesi için 990ms çerçeve boyutu ile gerçekleştirilen analizler	41

KISALTMALAR

DVM	Destek Vektör Makineleri
MFCC	Mel-Frequency Cepstral Coefficients
SA	Sinir Ağları
GKM	Gauss Karışım Modeli
SMM	Saklı Markov Modeli
TBA	Temel Bileşen Analizi
RTF	Radyal Taban Fonksiyonu
SSMM	Sıradüzensel Saklı Markov Modeli
YSA	Yapay Sinir Ağları

1. GİRİŞ

Günümüz modern teknolojisinin en önemli hedeflerinden biri insana yakın sistemler geliştirilmesidir. Bu hedef; insanlara daha iyi hizmet verebilmek ve kullanıcı etkileşimlerini güçlendirmek gibi temel gereksinimlerden ileri gelmektedir. Aynı zamanda, kullanıcılar da gün geçtikçe bilgiye olan ihtiyaçlarının daha hızlı giderilmesini istemekte ve dolayısı ile birçok alanda daha pratik yöntemler talep etmektedirler. Gelişen dünyanın isteklerinin cevaplanmasında, son zamanlarda temel olarak çokluortam verileri üzerinde de çalışmalar yoğunluk kazanmıştır. Bu durum, insan hayatına en yakın ve gündelik yaşantıyı en iyi ifade edecek verilerin ses, video gibi çokluortam verileri olduğundan kaynaklanmaktadır. Önceki çalışmalar özellikle çokluortam verilerinden olan video ve resim ile ilgilidir. Fakat bu gibi çalışmalarda temel problem, geliştirilen sistemlerin genelinen görüntü tabanlı olmasından kaynaklanmaktadır. Örneğin, görüntü tabanlı bir robot uygulamasında, görüş mesafesinin azaldığı durumlarda sistem beklenen performansı gerçekleştiremeyebilir. Benzer olarak, çokluortam verisi olan video analizlerinde görüntü tabanlı ilerlemek her zaman en doğru sonuca götürmeyebilir [37]. Bu nedenle, insan doğasının da en aktif reflekslerinden biri olan ses ile sayısal sistemler üzerinde çalışabilmek, araştırmacılar için yeni bir alan haline gelmiştir. İşitsel tabanlı sistemler üzerindeki çalışmalar son yıllarda hızla artmaktadır. Bu alandaki ilk çalışmalar ağırlıklı olarak içerik-tabanlı genel ses sınıflandırma problemini ele almaktadır [22, 31, 38, 39, 40, 41, 42]. Genel ses sınıflandırmada, ses verisi; konuşma, müzik ve çevresel ses gibi ana ses kategorilerine ayrıştırılmaktadır [43]. Bununla birlikte, son beş-on yılda çokluortam olay sezimi (MED – Multimedia Event Detection) adı altındaki çalışmalar giderek artmaktadır [27,28]. Bu çalışmalarda temel amaç; herhangi bir çevrede bulunan ses ve görsel sahnelerdeki belirli olaylarının bulunması olarak ifade edilmektedir. Müzik ve konuşma gibi genel ses kategorileri kendine ait tonal ve örüntüsel karakteristikler içermekteyken, ses olayları için belli bir karakteristik yapıdan söz edilememektedir. Örneğin, sıradüzensel ses sınıflandırmada, bir ses verisi öncelikle sessizlik, müzik, konuşma ve ses olaylarına, daha sonra müzik için pop ve caz gibi alt türlere; konuşma sesleri için kadın ve erkek gibi alt kategorilere inilebilir [3]. Ancak, ses olayları için bu şekilde bir alt sınıflandırma yapmak bu

seslerin çeşitliliğinden dolayı neredeyse imkânsızdır. Ayrıca farklı kaynaklardan çıkan ses olayları insan kulağının bu sesleri benzer algılamasından dolayı ayırt edilemeyebilir ve kendine özgü ses olayları barındıran işitsel sahnelerin sezimlenmesini daha güç olabilir [3].

Ses olaylarının ve buna bağlı işitsel sahnelerin sezimiyle ilgili çalışmalar son yıllarda artmaktadır [1-19, 26, 29-31]. Bu çalışmaların genelinde, bir işitsel sahnenin sezimlenmesi için öncelikle içerdiği ses olayları analiz edilmektedir. Örneğin Muhammad vd. çalışmalarında restoran, kalabalık cadde, alışveriş merkezi, ofis ortamı ve park gibi birçok işitsel sahnenin sezimini gerçekleştirmiştir [4].

İşitsel sahnelerin ve ses olaylarının sezimlenmesi ile ilgili yöntemler, öncelikle bu ses dosyaları için en uygun sınıflandırıcı ve öznitelik vektörünün seçilmesi ile başlamaktadır. Sınıflandırıcı yöntemleri olarak genelde en çok Destek Vektör Makineleri (DVM), Saklı Markov Modelleri (SMM), Sinir Ağları (SA) ve Gauss Karışım Modeli (GKM) yöntemleri kullanılmıştır. Mevcut çalışmaların büyük çoğunluğu, sesleri en iyi ifade edebilecek öznitelik vektörlerini bulmak ve en iyi makine öğrenme yöntemlerini uygulamak üzerine yoğunlaşmaktadır.

Bu tez çalışmasında, ses olaylarının ve ses sahnelerinin içerik-tabanlı sezimi için bir sistem sunulmaktadır. Önerilen sistemde, 10 işitsel sahne (*otobüs, kalabalık sokak, ofis ortamı, açık market, park, sessiz sokak, restoran, süpermarket, tren ve tren istasyonu*) ve *ofis* işitsel sahnesi ile ilintili olabilecek 16 ayrık ses olayı (*alarm, boğaz temizleme, öksürük, kapı çarpması, çekmece, klavye, anahtar, kapı vurma, gülme, bilgisayar faresi, sayfa çevirme, nesnenin masaya çarpması, telefon, yazıcı, konuşma ve elektrik düğmesi*) sınıflandırılmıştır. Ayrıca, *ofis* işitsel sahnesinin içerdiği ses olayları ve bu ses olaylarının *ofis* işitsel sahnesine olan katkıları incelenmiştir. Bu kapsamda, kullanılan öğrenme algoritmasının başarımını arttıracak parametrelerin eniyilemesi gerçekleştirilmiş, sınıflandırıcı ile çalışacak öznitelik vektörü analiz edilerek en uygun değerler bulunmuş ve öznitelik çıkarımında kullanılacak ses sinyaline ait özgün frekans analizleri gerçekleştirilmiştir.

Ses dosyasını ifade edecek öznitelik vektörü olarak Mel Frequency Cepstral Coefficients (MFCCs) kullanılmıştır. Makine öğrenme yöntemi olarak Destek Vektör Makinesi (DVM) seçilmiştir. Benzer çalışmalardan farklı olarak, öznitelik çıkarımında sinyal frekans karakteristiklerini temel alan özgün bir analiz yöntemi önerilmiş ve geleneksel yöntemlerle karşılaştırmalı olarak sonuçları değerlendirilmiştir.

1.1 Tezin Organizasyonu

Bölüm 2'de literatür taraması verilmiştir. Bölüm 3'te ses olayları ve işitsel sahneler için temel tanım ve kavramlar verilmiştir. Ayrıca, kullanılan makine öğrenme algoritması (DVM) ve öznitelik çıkarım yöntemi (MFCC) hakkında temel bilgiler sunulmuştur. 4. bölümde ses olaylarının sezimi ve 5. bölümde işitsel sahnelerin tanınması için uygulanan yöntemler, deneysel çalışmalar ve değerlendirmeler verilmiştir. Ayrıca, seçilen ses olayları ile *ofis* işitsel sahnesi arasındaki ilintiler incelenmiştir. Son bölümde ise değerlendirmeler ve gelecek çalışma planları aktarılmıştır.

2. LİTERATÜR TARAMASI

Ses tanıma konusundaki çalışmalar, ses verisinden öznitelik çıkarımı ve sınıflandırıcı tasarımı konularına yoğunlaşmaktadır. Bu alanda yapılan çalışmalardan öne çıkanlar aşağıda özetlenmiştir.

Eronen vd. [1], çalışmalarını çevresel seslerin tanımlanması sistemi üzerine gerçekleştirmişlerdir. Önerilen sistemde model eğitiminde kullanılmak üzere sınıflandırıcı olarak SMM ve k-NN algoritmaları kullanılmıştır. Veri kümesi olarak gerçek dünya örneklerinden alınan 24 farklı çevresel ses kullanılmışlardır. Daha sonra bu 24 çevresel sesi ortak karakteristiklerine göre 6 üst seviye sınıf olarak düzenlemişlerdir. Üst seviye sınıflar içerisinde dış ortam, motorlu taşıtlar, açık ortam, ev ortamı ve içerisinde kilise, demiryolu ve metro gibi yankılı sesleri içeren işitsel sahneler vardır. Sistemin çevresel sesleri tanıma başarımı %69 iken, bu başarımlar işitsel sahneler için %88 olarak rapor edilmektedir. Test işlemi gerçek kişiler üzerinde gerçekleştirildiğinde, başarımlar sırasıyla %58 ve %82'e düşmektedir.

Ma vd. [2] çalışmalarında çevresel seslerin sezimlenmesi ve sınıflandırılması için uyarlanabilir bir sistem ve bu sistemin deneysel sonuçlarını sunmuşlardır. Aynı zamanda çevresel seslerin analizi ile konum ve kişinin aktivite bilgisi elde edilmesi amaçlanmıştır. Bu şekilde, kullanıcının anlık durumu ile ilgili bilgi almayı hedeflemişlerdir. Çevresel seslerin ses dosyalarından öznitelik çıkarımında MFCC yöntemi kullanılmışlardır, model eğitiminde kullanılacak sınıflandırıcı olarak ise SMM algoritmasını kullanmışlardır. Bunun dışında ses dosyaları gerçek kişilere dinleterek ayrı bir test işlemi gerçekleştirmişlerdir. Çalışmalarında 10 farklı ses kullanılmıştır. Bu çevresel sesler; ofis, derslik, otobüs, şehir içi trafik, tren istasyonu, sahil, bar, çamaşırhane, futbol maçı, sokak (şehir merkezinde). Bunların dışında ek olarak sessizlik eklenmiştir. Doğruluk oranları her sınıf bazında verilmiş ve değerler %75 ile %100 arasında değişmektedir. Genel ortalama ise %92 olarak raporlanmıştır. Hatta birçok sınıf için %100'lük bir doğruluk elde edilmiştir. En kötü doğruluk oranı ise sokak için çıkmıştır. Bu sınıf en çok çamaşırhane ve tren sınıflarıyla karışmaktadır. Kullanılan SMM, yöntemi daha düşük bant genişliği için en uygun değerlere ayarlandığında doğruluk genel ortalama olarak %96 seviyelerine ulaşmıştır. Bu testler kişiler üzerinde

uygulandığında, insanların sınıfları doğru sezimleme oranı %35 seviyesinde kalmıştır. Fakat çalışmalarında da belirttikleri gibi, kullanılan veri kümesi, çok küçük ve veri kümesinin dağılımı düşüktür.

Okuyucu vd. [3], benzer çevresel ses kategorilerinin büyük ölçüde tanımlanmada ve daha iyi bir doğruluk oranı elde edebilmede kullanılmak üzere çevresel sesleri en iyi ifade edebilecek öznelikleri bulmak için detaylı bir öznelik ve sınıflandırıcı analizi gerçekleştirmişlerdir. MPEG-7 ailesi, MFCC, zero crossing rate (ZCR) ve bunların farklı birleşimleri ile oluşturdukları 11 farklı öznelik birleşimi denenmiş ve bu öznelikler model eğitiminde kullanılmak üzere DVM ve HMM sınıflandırıcılarında işlem görmüştür. HMM özel ayarları için farklı durumlar denenmiş ve bunların arasından en uygun olan 5 durumlu HMM modeli kullanılmıştır. DVM için çekirdek ayarı olarak radyal tabanlı fonksiyon (RTF) kullanılmıştır. Sonuçlara göre, en iyi sınıflandırıcı DVM ve en iyi öznelik birleşimi ise Audio Spectrum Flatness (ASF), Audio Spectrum Centroid (ASC), Audio Spectrum Spread (ASS), Audio Harmonicity (AH) özneliklerinden meydana gelen ASFCS-H olarak raporlanmıştır. Performans değerlendirmeleri f-measure üzerinden yapılmıştır. Buna göre bu öznelik birleşiminin ve en yüksek oranın elde edildiği model olan DVM sınıflandırıcısı ile sistemin doğruluk oranı %80 olmaktadır.

Muhammad vd. [4], çevresel seslerin sezimi için MFCC özneliği ve seçilmiş MPEG-7 ses tanımlayıcılarını kullanan bir sistem önermişlerdir. MPEG-7 tanımlayıcıları öncelikle Fisher oranlarına göre sıralanmış ve ilk 30 MPEG-7 tanımlayıcısına Temel Bileşen Analizi (TBA) uygulanmıştır, sonuç olarak bu aileden 13 tane öznelik elde edilmiştir. Daha sonra önerilen sistemin öznelik kümesini oluşturabilmek için bu 13 öznelik MFCC özneliklerine eklenmiştir. Model eğitiminde kullanılmak üzere sınıflandırıcı olarak GMM seçilmiştir. Sistem 10 farklı çevresel ses kullanılarak değerlendirilmiş ve sonuçları 4 farklı kategori için raporlanmıştır. İlk kategoride, sadece MFCC özneliği kullanılarak, ikinci kategoride tüm MPEG-7 ses tanımlayıcı öznelikleri kullanılarak eğitilen sistem üzerinde gerçekleştirilen test işlemleridir. Üçüncü kategoride seçilmiş MPEG-7 ses tanımlayıcısı öznelikleri ile ve son kategoride ise olarak çalışmada önerilen, seçilmiş MPEG-7 ve MFCC birleşiminden meydana gelen öznelik kümesi ile eğitilmiş model dosyası üzerinde gerçekleştirilen test işlemlerinin sonuçlarıdır.

Buna göre tüm sistemin ortalama doğruluk oranları sırasıyla %85,1, %89,5, %90,8 ve %93 olarak sunulmaktadır. Elde edilen başarımlar yüksek (%93) olmasına rağmen, seçilen MPEG-7 öznitelikleri ve MFCC öznitelik birleşimi, öznitelik boyutlarını bir hayli arttıracak için model eğitiminde hesaplama ve zaman karmaşıklığına yol açabilir.

Su vd. [5], çevresel ses ve işitsel sahnelerin sezimi için bir sistem tasarlamıştır. Çalışmalarında Local Discriminant Bases (LDB) adı verilen tekniği kullanmışlardır. Bu teknikte, çevresel sesler için zaman-frekans alt uzaylarının ayırımı tanımlanmaktadır. Çalışmaları değerlendirildiğinde 10 işitsel sahne ve 21 çevresel ses için sırasıyla %74,3 ve %81 doğruluk oranları elde etmişlerdir. Fakat veri kümesi içerisinde arka planı gürültü olan sesler kullanıldığında genel doğruluk yüzdesi %28,6 oranına düşmektedir. Çalışmalarında kullanılan veri kümesinin, internet televizyon ve filmlerden aldıklarını raporlamışlardır, ancak erişilebilir değildir.

Wang vd. [6], çevresel seslerinin sınıflandırılmasını gerçekleştiren bir sistem önermişlerdir. Çalışmalarında 12 farklı çevresel ses kullanmışlardır. Model eğitiminde sınıflandırıcı olarak DVM ve k-NN yöntemlerini uygulamışlardır. Öznitelik seçiminde 3 MPEG-7 ses tanımlayıcısı, ASC, ASS ve ASF öznitelikleri alınmıştır. Sınıflandırma doğruluk oranı genel sistem için %85,1 çıkmıştır.

Chu vd. [7], çevresel seslerin tanımlanması üzerine çalışmışlardır. Ses sinyallerini farklı seviyelere ayrıştırarak, aynı sinyalin farklı temsillerini göstermişlerdir. Böylelikle seslerin daha iyi tanımlanabileceğini raporlarında ifade etmişlerdir. Çalışmada 14 farklı çevresel ses kullanılmıştır. Etkin bir öznitelik elde etmek için Matching Pursuit (MP) tabanlı bir algoritma ile öznitelik seçim işlemlerini gerçekleştirmişlerdir. Daha sonra elde edilen bu öznitelikleri MFCC özniteliği ile birleştirerek daha yüksek doğruluk değerleri elde etmeyi amaçlamışlardır. Model eğitiminde kullanılmak üzere sınıflandırıcı olarak GKM ve k-NN ikilisini seçmişlerdir. Sonuçlarını 3 farklı öznitelik birleşimi ve ayrı ayrı GKM ve k-NN sınıflandırıcılarından aldıkları sonuçlara göre sınıf bazında sunmuşlardır. Diğer özniteliklerden aldıkları sonuçlarla kıyaslandığında, en yüksek doğruluk, MP tabanlı öznitelikler ile MFCC özniteliğinin birleşiminden ortaya çıkan yeni öznitelik kümesi için elde edilmiştir.

Lee vd. [8], çevresel seslerdeki belirli bölgeleri anlamsal olarak sezimleyebilen bir sistem üzerine çalışmışlardır. Veri kümesini oluşturabilmek için YouTube¹ üzerinden 25 farklı kavrama ait 1873 adet video indirilmiştir. Kavramlar farklı birçok sınıflara ayrılmıştır: Aktiviteler (kayak, dans), özel durumlar (doğum günü, mezuniyet), konumlar (sahil, park) veya bir sahnedeki belirli nesnelere (hayvan, bebek, bot v.b.). Her sınıf için ortalama kesinlik değerini kullanarak yapılan değerlendirmeler sonucunda en yüksek başarımlar alkış için elde edilmiştir.

Beritelli ve Grasso [9], arka plan sesleri için bir örüntü tanıma sistemi üzerine çalışmıştır. Çalışmada yedi farklı arka plan sesi kullanılmıştır: Araba, otobüs, inşaat, çöplük, fabrika, ofis ve havuz olmak üzere 7 tanedir. Model eğitimlerinde kullanılmak üzere sınıflandırıcı olarak SA ve öznitelik çıkarıcı olarak ise MFCC kullanmışlardır. Sistem değerlendirmesi yanlış sınıflandırılan ses dosyalarının yüzdesi üzerinden yapılmıştır. Ortalama doğruluk oranı sesin uzunluğuna bağlı olarak %75 ile %95 arasında değişmektedir.

Feki vd. [10], ses akışı analizi temelini kullanarak çevresel seslerin sezimi için bir yapı tasarlamıştır. Çalışmalarında konuşma, müzik, zil tonları, tren, motosiklet, patlama, helikopter, kapı çarpması, köpek havlaması, kuş sesi gibi 20 farklı sınıfı kullanmışlardır. Önerilen sistem 3 aşamadan meydana gelmektedir. İlk aşama ön hazırlık aşaması olup, ses dosyası bölümlere ayrılmış ve sessizlik bölümleri sezimlenmiştir. Sesin karakteristiklerini belirleyebilmek için short-time energy (STE), low short-time energy ratio (LSTER), spectral flux (SF), band periodicity (BP) ve MFCC özniteliklerini kullanmışlardır. İkinci aşamada model eğitimi sırasında sınıflandırıcı olarak sinir ağları (SA), HMM ve DVM kullanılarak, konuşma, müzik ve çevresel sesleri otomatik olarak sınıflandırılmıştır. Son aşamada ise ikili sınıflandırıcıları sarmalayan yeni bir sistem uygulamışlardır. Çalışmalarının sonucunda ses kavramlarını tanımlamada %90'dan daha fazla bir başarı elde edilmiştir.

Güvensan ve Taysi [11], akıllı ev sistemleri için çevresel ses sezimi yapabilen bir sistem önermiştir. Çalışmalarında buzdolabı, çırpıcı, aspiratör, bulaşık makinesi, fön makinesi ve havalandırma gibi ev içerisinde kullanılan cihazların seslerini veri kümesi olarak kullanmışlardır. Model oluşturmada kullanılmak üzere DVM ve k-NN

¹ <https://www.youtube.com/>

sınıflandırıcılarını seçmişlerdir. Çalışmalarının sonuçlarında, en iyi sonucun MFCC özniteliği ile birlikte DVM sınıflandırıcısı kullanıldığında elde edildiğini raporlamışlardır (%98). Fakat bu çalışma, çok kısıtlı bir veri kümesi üzerinde gerçekleştirilmiştir.

Shin vd. [12], insanlar için normal olmayan sağlık durumlarının sezimi için öksürük sesi üzerinde çalışmışlardır. Gerçek zamanlı olarak tasarlanan bu uygulama, akustik bilgiler kullanarak kişinin sağlık durumunu izlemek ve anormal durumları tespit edebilmektedir. Yapay Sinir Ağları (YSA) ve HMM' den oluşan karma bir model çevredeki diğer seslerden öksürük sesini ayırt edebilmek için sınıflandırma sırasında kullanılmıştır. Önerilen bu karma model HMM sınıflandırıcısı ve MFCC özniteliği kullanılarak tasarlanan geleneksel sistemlere göre daha iyi sonuçlar vermiştir.

Schröder vd. [13], çevresel seslerin sezimi için bir sistem önermişlerdir. Çalışmaları 2 katmanlı GMM sınıflandırıcısından oluşmaktadır. Çalışmalarında kullanılmak üzere sundukları veri kümesi *IEEE Challenge on Audio and Acoustic Signal Processing (AASP) Challenge on Detection and Classification of Acoustic Scenes and Events*'den alınmıştır [21]. Ses klibi bazında yapılan değerlendirmelere göre önerilen sistem f-ölçütü olarak %45,17 oranında bir performans göstermiştir [35].

Niessen vd. [14], çevresel seslerin sınıflandırılması için yöntemler önermişlerdir. Model eğitiminde kullanılmak üzere sıradüzensel saklı Markov modeli (SSMM) sınıflandırıcısını kullanmışlardır. Ses dosyalarından öznitelik çıkarımı için ise MFCC, ZCR, SF gibi farklı yöntemler kullanmışlardır. Veri kümesi olarak *AASP Challenge on Detection and Classification of Acoustic Scenes and Events* resmi internet sitesinde erişime açık olarak yayınlanan sentetik olmayan ofis ortamı veri kümesini kullanmışlardır [21]. Ses klibi bazında yapılan değerlendirmeler sonucu, %34,51'lik bir doğruluk sonucuna ulaşmışlardır.

Vuegen vd. [15], çalışmalarını *AASP Challenge on Detection and Classification of Acoustic Scenes and Events* isimli yarışma için çevresel seslerin sezimi üzerine yapmışlardır. Model eğitimi sırasında GMM sınıflandırıcısını kullanmışlar, öznitelik çıkarımı olarak ise MFCC seçmişlerdir. [13] ve [14]'de olduğu gibi klip bazında

verilen deęerlendirme sonucu geliřtirilen sistemin performansı %30,77 olarak raporlanmıřtır.

Çalıřmalar incelendięinde, en önemli zorluęun çevresel seslerin çeřitlilięinden kaynaklanan karar verme zorluęu olduęu görölmektedir. Öyle ki, ses olayı sınıf sayısı arttıkça, problem daha ayırık olmakta ve tanıma başarımı düşmektedir. Bu çalışmada öncelikle, bu problem ele alınmıř ve geliřtirilen yöntemler gerçek hayat verileri üzerinde test edilmiřtir.

3. TEMEL TANIM VE KAVRAMLAR

Bu bölümde, çalışmamızda kullanılan temel tanım, kavram ve yöntemler açıklanmıştır.

3.1 Ses

Ses, atmosferde canlıların işitme organları tarafından algılanabilen periyodik basınç değişimleridir [33]. Her kaynak farklı frekanslarla titreşim yaparak ses üretir. Sesin iletilmesi ve kaydedilebilmesi için bir çevirimden geçerek ses sinyali haline getirilir. Bu çevirimi gerçekleştiren mikrofon gibi çevirici araçlardır. Ses sinyali sesin sayısal sistemlerdeki elektriksel modellenmiş halidir ve insan kulağı tarafından algılanmazlar.

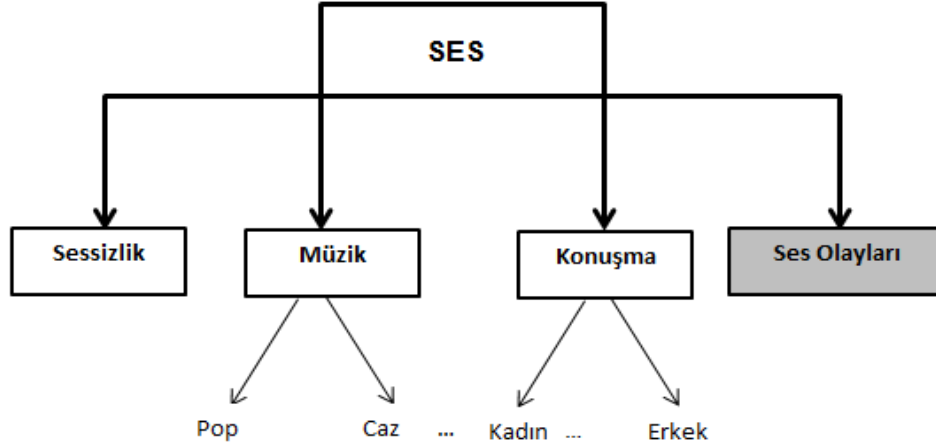
3.2 Ses Olayı

Günlük hayatta duyduğumuz sesler özelliklerine göre sessizlik, müzik, konuşma ve ses olayları olarak alt gruplara ayrılmaktadır. Ses olayları, tanım olarak bir sesin çevresi ve çevreyle ilgili faktörleriyle ilişkilendirilmiş özel bir harmanlamadan meydana gelen sesleri ifade etmektedir. Ses olayları, bazı özelliklerinden dolayı müzik ve konuşma seslerinden farklı ifade edilir. Bu özelliklerden en önemli olanı belirli bir karakteristik yapılarının olmamasıdır. Örneğin stüdyoda kaydedilen müzik, belli notalardan ve nakarat bilgilerinden meydana gelmekteyken, çevresel sesler için böyle bir durum söz konusu değildir. Ses olayları aynı zamanda ortam ve zamana göre çeşitlilik gösterebilirler. Ses olayları frekans düzleminde görülen gürültü benzeri düz yapılarından dolayı konuşma ve müzik gibi seslere göre daha zor sezimlenmektedir [7]. Bir konuşma için ses dosyasının ön planındaki sesler ile uğraşırken, çevresel seslerde arka plandaki seslerle ilgilenilmektedir.

3.3 İşitsel Sahne

İşitsel sahne, bir ses klbinin mantıksal parçasıdır. Bir sahne, belirli bir bağlama ait birçok farklı ortam karakteristiğinden meydana gelmektedir ve bu ortam karakteristikleri bir ses kaynağında bulunan ses olaylarının birleşimi olarak ifade edilir. Bir işitsel sahne içerisinde birden çok ses olayı bulunabilir veya birden çok işitsel sahne içerisinde aynı ses olayını barındırabilir. Örneğin bir sokak sahnesi için içerisinde trafik ve motor sesleriyle ilgili birden çok ses olayı bulunabilir. Bir

işitsel sahneyi ses olaylarının tanımlanmasıyla ilişkilendirmek sahne sezimleme yaklaşımını güçlendirebilir. Buna göre birbirleriyle ilişkili birçok ses olayından meydana gelen bir işitsel sahne için anlamsal çıkarımlar yapılabilmektedir. Bu sebeple bir işitsel sahnenin otomatik sezimi için öncelikle içerdiği ses olaylarının analizini yapmak gerekmektedir.



Şekil 3. 1 Genel ses sınıfları

3.4 Ses Özniteliği

Ses öznitelikleri analiz edilen sinyalden elde edilen ve bu sinyalin karakteristik bilgilerini taşıyan değerler kümesi olarak ifade edilir. Sinyalden elde edilen veriler öznitelik vektörü adı verilen veri tanımlayıcıları tarafından tutulur ve sinyalden anlamsal bilgileri çıkarma işlemine öznitelik çıkarımı adı verilir. Literatürde oldukça fazla öznitelik çıkarma yöntemleri bulunmaktadır. Tez kapsamında sesin karakteristiklerini elde edebilmek için farklı parametre seçenekleri ve çeşitli katsayılar ile MFCC öznitelik çıkarma yöntemi seçilmiştir.

3.4.1 Mel ölçekli kepstrum katsayıları (Mel frequency cepstral coefficients)

Mel ölçekli kepstrum katsayıları (MFCC) ses tanıma uygulamalarındaki başarısının yüksek olmasından dolayı en fazla kullanılan öznitelik çıkarım yöntemlerinden biridir [22]. MFCC insan kulağının algılama şeklini modellemektedir. Ses sinyalinin kısa süreli güç spektrumunu temsil etmektedir. Temeli frekansın doğrusal olmayan

Mel ölçeği üzerinde log güç spektrumunun doğrusal kosinüs dönüşümüne dayanmaktadır. Ses sinyalinden MFCC vektörünü çıkarabilmek için, örnekleme ve pencereleme işlemleri uygulanır. Örnekleme işlemi sırasında verilen sürekli sinyal ayırık hale getirilir. Pencereleme aşaması sırasında ise sinyalin zaman bölgesinde bir fonksiyon ile çarpılır. Bu sayede sinyalin işlenmek istenilen bölümü alınır. MFCC öznitelik çıkarma yönteminde kullanılan dikdörtgen, Barlett, Kaiser, Hamming gibi birçok pencereleme fonksiyonu vardır, fakat en yaygın kullanılan Hamming pencereleme fonksiyonudur. Bu çalışmada da kullanılan fonksiyonun denklemi şu şekilde tanımlanmıştır:

$$w(n)=0.54-0.46 \times \cos\left(\frac{2\pi n}{N-1}\right) \quad (3.1)$$

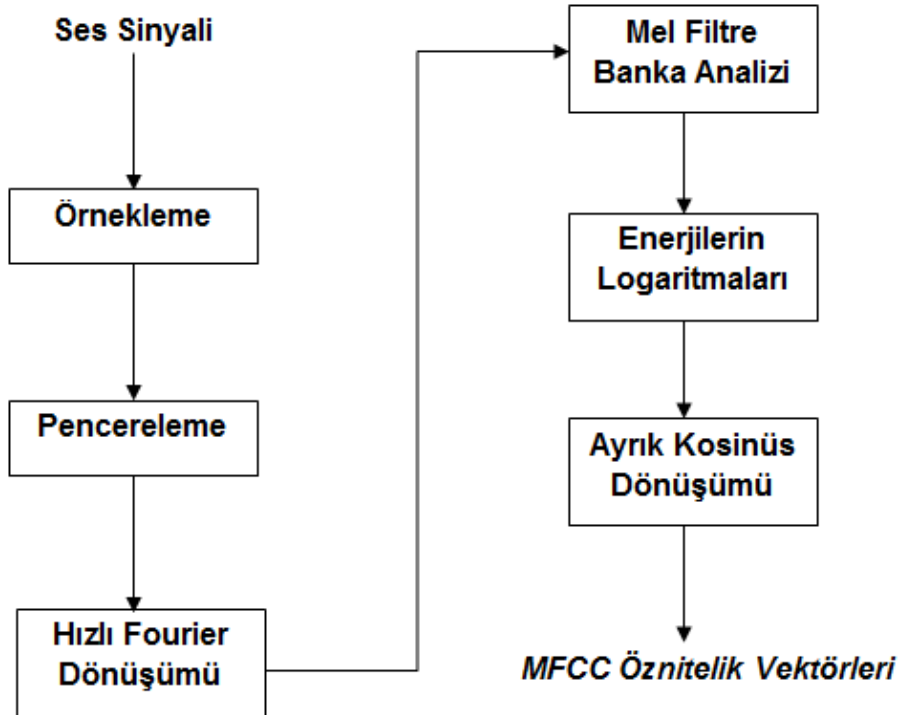
Sinyal işlemede, süzgeç ailesi, verilen bir sinyalin birçok bileşen ayıran bant geçiren filtreler olarak tanımlanır. Bu bileşenlerin her biri tekil bir frekans alt bandını taşır [34]. MFCC öznitelik çıkarımı aşamasında ise hızlı Fourier dönüşümü sonucu elde edilen genlik spektrumu, mel skalasında eşit olarak dağıtılmış ve birbirini %50 oranında kesen üçgen filtre çarpılır. Bu çarpma işlemi sonunda her bir filtrenin altındaki enerji hesaplanır. Bu bantları geçiren filtrelere Mel süzgeç aileleri, yapılan işlem ise Mel süzgeç ailesi analizi denilir. [44] Daha sonra bu aşamada elde edilen N adet enerji değerinin logaritması alınır. Bu logaritma değerlerinin ayırık kosinüs dönüşümü alınarak öznitelik vektörleri elde edilir. MFCC öznitelik çıkarma adımları Şekil 3.2' de gösterilmiştir.

3.5 Sınıflandırma Yöntemi

Sınıflandırma, temel anlamda karar verme amacıyla kullanılan bir işlemdir. Ses sınıflandırma uygulamalarında amaç, bir ses sinyalindeki belirli desenleri karşılık geldikleri sınıflara otomatik olarak yerleştirmektir. Sınıflandırma yöntemleri içerisinde ses uygulamalarına uygun birçok yöntem bulunmaktadır. Bunlar; Saklı Markov Modelleri, En yakın K Komşuluk, Gauss Karışım Modeli ve Destek Vektör Makineleri gibi yöntemlerdir. Bu tez çalışmasında, ses olaylarını ve işitsel sahneleri sezimlemek için kullanılan sınıflandırıcı yöntemi, örüntü tanıma uygulamalarında başarımı nedeniyle DVM olarak seçilmiştir [3, 6, 10, 11].

3.5.1 Destek vektör makineleri

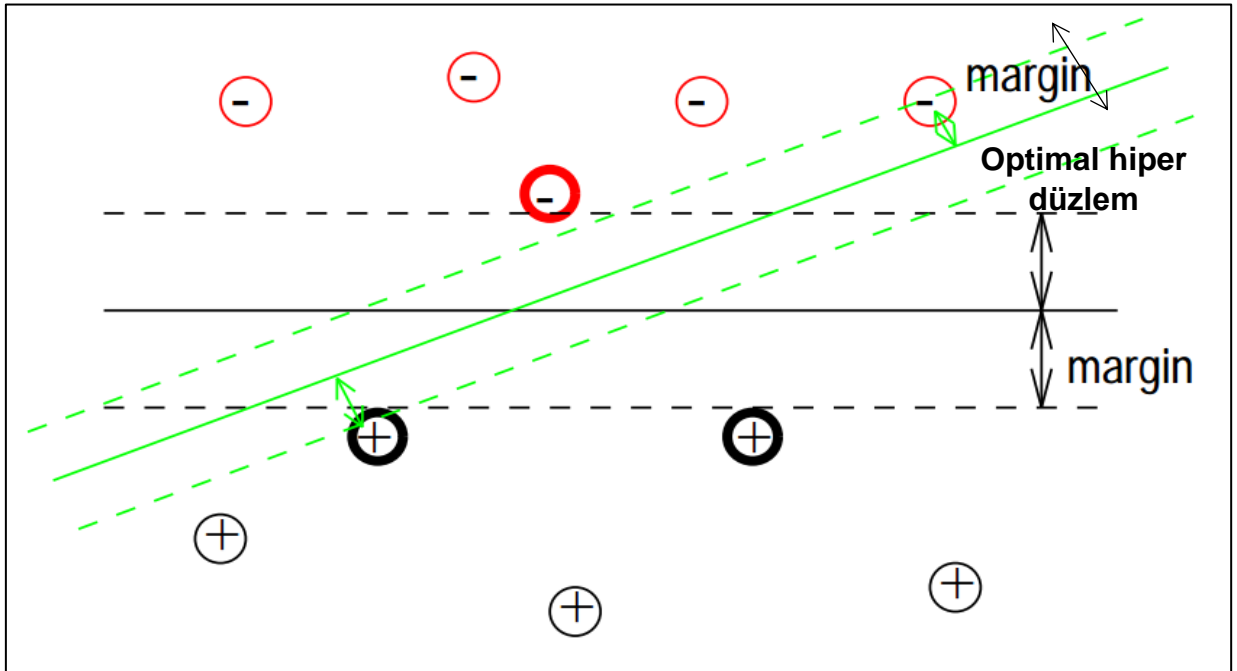
Destek Vektör Makineleri ilk olarak Vapnik ve arkadaşları [20, 25] tarafından iki sınıflı bir hiper düzlem yaratarak sınıflandırmak prensibi üzerine kurulmuştur ve bu hiper düzlemler istatistiksel öğrenme teorisini kullanarak hesaplanmaktadır. Destek vektör makinelerinde temel düşünce ayırt edilebilir örüntüler için en uygun hiper düzlemi bulmak ve ayırt edilemeyen örüntüler için ise orijinal verideki örüntüleri yeni bir uzaya geçirmek için dönüşümler uygulamaktır. DVM'lerdeki destek vektörleri, karar yüzeyine en yakın olan ve sınıflandırılması en zor olan noktalardır. İki boyutlu uzay için bir eksen ayırt etme konusunda yeterli olurken, çok boyutlu uzaylar için hiper düzlem gereklidir. Temel bir tanım vermek gerekirse, n boyutlu bir uzayda n-1 boyutlu bir alt uzay yaratan noktalar kümesine hiper düzlem denir. Örneğin üç boyutlu uzayın belli bir bölgesini temsil eden iki boyutlu bir düzlem üç boyutlu uzayın bir hiper düzlemi olarak ifade edilir. İki sınıflı bir veri kümesinde, eğitim kümesindeki pozitif ve negatif örnekler Şekil 3.3'de görüldüğü gibi bu hiper düzlemi oluşturur.



Şekil 3. 2 MFCC öznitelik vektörleri çıkarım aşamaları

Birçok hiper düzlem yaratılabilir fakat en uygun olanı bularak veri kümesini sınıflandırmak gerekmektedir. En uygun hiper düzlem, hiper düzlem ile en yakın noktaya olan uzaklığı en büyük olandır. En uygun hiper düzlem denklemi aşağıdaki gibidir.

$$W \cdot X + b = 0; W \in R^N \quad b \in R \quad (3.2)$$



Şekil 3. 3 DVM sınıflandırıcı hiper düzlem

W hiper düzleme dik olan vektörü ifade ederken, b sabit bir sayıdır. X ise sınıfı bilinmeyen bir noktayı ifade etmektedir. Doğrusal olarak birbirinden ayrılabilen veriler için karar fonksiyonu ise şu şekilde tanımlanmaktadır:

$$f(x) = w \cdot x + b = \sum_{j=1}^N w_j \cdot x_j + b \quad (3.3)$$

Bu denkleme göre, verilen x noktası pozitif sınıfına ait ise $f(x) \geq 0$, negatif sınıfına ait ise $f(x) < 0$ şartına göre karar verilir. Denklemdaki w n -boyutlu bir vektörü, b ise sabit bir sayıyı ifade etmektedir.

Doğrusal olarak birbirinden ayırt edilemeyen veriler için, öznelik vektörlerini daha büyük boyutlu bir uzaya taşımak için çekirdek fonksiyonları kullanılır. En uygun hiper düzlemi yaratmak amacıyla çekirdek fonksiyonları kullanılmak istenirse karar fonksiyonu denklemini şu şekilde değiştirmektedir:

$$f(x) = w \cdot x + b = \sum_{j=1}^N w_j \cdot x_j \cdot K(x_j, x) + b \quad (3.4)$$

DVM için kullanılan doğrusal, polinom, radyal tabanlı sigmoid çekirdek olarak en yaygın dört seçeneği çalışmalarda kullanılmaktadır. Farklı veri kümeleri için uygun olan çekirdek fonksiyonu seçilerek kullanılabilir. DVM temelinde ikili sınıflandırma yapan bir algoritma olmasına rağmen, birçok problemde kullanabilmek üzere çok sınıflı uygulamalarda da kullanılabilir. Bu tez çalışmasında fonksiyon olarak daha yüksek performansta sonuçlar verdiği için radyal tabanlı çekirdek fonksiyonu kullanılmıştır. Çoklu sınıflandırmada DVM için iki farklı yöntem mevcuttur. Bunlar bire karşı bir ve bire karşı hepsi stratejileridir.

❖ **Bire Karşı Bir:** Bu yöntemde her bir sınıf çifti için bir DVM modeli oluşturulur. Bu durumda fazla sınıflı uygulamalarda çok fazla model dosyası olduğundan zaman ve bellek problemleri ortaya çıkmaktadır. n sınıflı bir uygulama için oluşturulacak toplam model sayısı m 3.5'da verilen denkleme göre hesaplanmaktadır. Bu sayıda model dosyasının sistem tarafından öğrenilmesi gerekmektedir.

$$m = \frac{n * (n - 1)}{2} \quad (3.5)$$

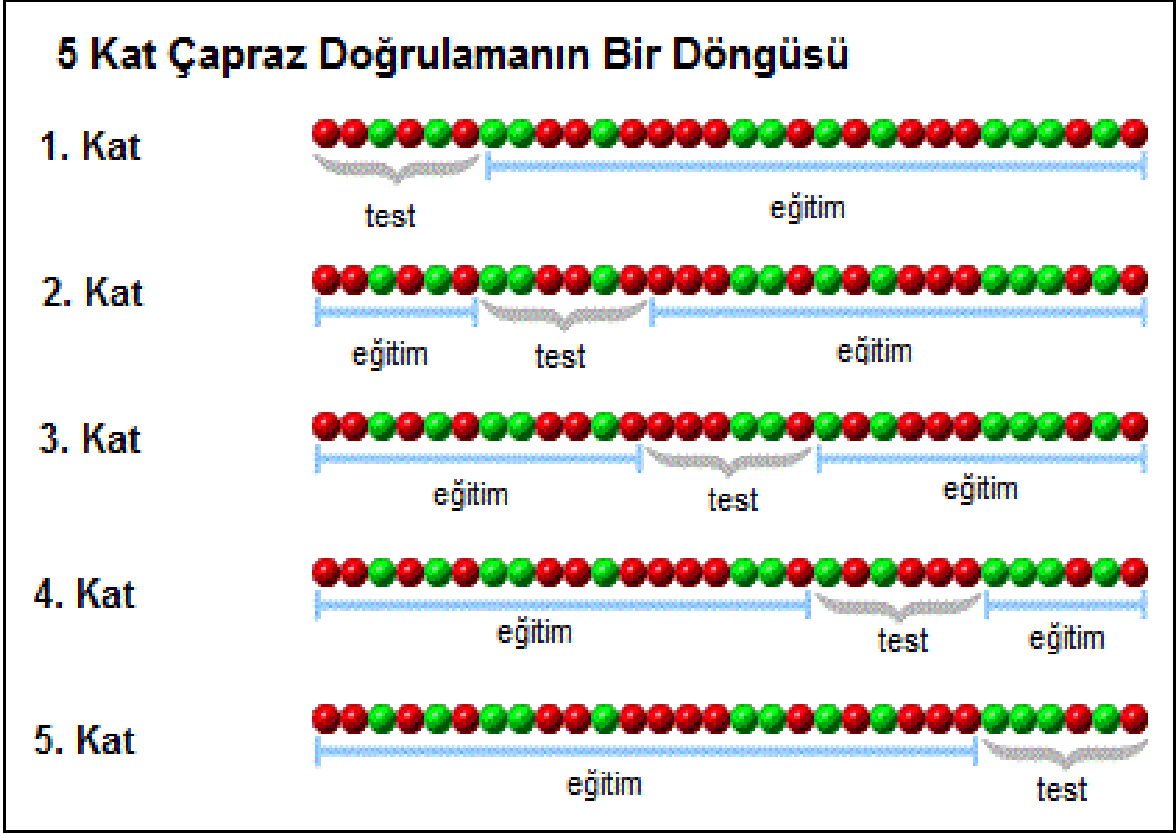
❖ **Bire Karşı Hepsi:** Bire karşı hepsi yönteminde, sınıf bazında DVM modeli yaratılır. Burada amaç, her bir modelde bir sınıfın örnekleri tanınır ve geri kalan sınıflardan ayırt edilir. N sınıflı bir uygulama için, n adet model yaratılır. Test aşamasında, olasılıklar üzerinden maksimum olana atama yapılır. Her iki yöntem farklı çalışmalarda sıklıkla kullanılmaktadır. İki strateji arasında belirgin derecede farklar gözlenmediğinden ve oluşturduğu model dosyasının az olması sebebiyle zaman ve bellek maliyet performansını arttıracığından seçilen yöntem bire karşı hepsi yöntemidir.

3.6 Çapraz Doğrulama

Çapraz doğrulama var olan tüm veri kümesini kullanarak sonuçları değerlendirmeyi ve karşılaştırmayı sağlayan istatistiksel bir yöntemdir. Veri kümesi verilen bir n değerine göre rastgele n parçaya bölünür. Her defasında 1 grup test için ayrılır, geri kalan grup model eğitimi sırasında kullanılır. Bir önceki aşamada test için kullanılan grup bir sonraki aşamada model eğitime dahil olurken, model eğitimi aşamasında kullanılan gruplarda sırasıyla test işlemine dahil edilir. Bu süreç n defa tekrar eder ve sonuçlar her tekrardan elde edilen performans ölçütlerinin ortalaması şeklinde ifade edilir. Örneğin N=5 seçilirse, çapraz doğrulamada tüm veri kümesi 5 parçaya bölünür. 5 parçadan 4 tanesi eğitim için 1 tanesi test verisi olarak alınır ve model eğitimi gerçekleştirilir. Bu sayede her bir küme hem eğitim, hem test aşamasında kullanılmış olur. Tez kapsamında kullanılan veri kümesinin değerlendirme aşamasında çapraz doğrulama kullanılmak istenildiğinden, n=5 seçilmiştir. Veri kümesi üzerinde 5-kat çapraz doğrulama yapılmıştır.

3.7 Performans Kriterleri

Tez çalışması kapsamında değerlendirmeler, doğruluk, kesinlik, duyarlılık ve f-ölçütü şeklinde sunulmuştur. Testler sonucunda ulaşılan bilgiler hata matrisi ile ifade edilebilir. Hata matrisinin yapısı Çizelge 3.1' de verilmiştir. Hata matrisine göre, satırlar test kümesindeki örneklerin gerçek sınıflarını, sütunlar ise tahmin edilen değerleri ifade eder.



Şekil 3. 4 5-Kat çapraz doğrulamanın bir döngüsü [36]

Çizelge 3. 1 Hata matrisi

Tahmin Edilen Gerçek	Pozitif Sınıf	Negatif Sınıf
Pozitif Sınıf	TP	FN
Negatif Sınıf	FP	TN

TP (Doğru Kabul – True Positive): Sistemin tahmin ettiği sınıf pozitif ve gerçekte de pozitif sınıfında olan örneklerin sayısını ifade eder.

FN (Yanlış Red– False Negative): Sistemin tahmin ettiği sınıf negatif ama gerçekte pozitif sınıfına ait olan örneklerin sayısını ifade eder.

FP (Yanlış Kabul – False Positive): Sistemin tahmin ettiği sınıf pozitif ama gerçekte negatif sınıfına ait olan örneklerin sayısını ifade eder.

TN (Doğru Red – True Negative): Sistemin tahmin ettiği sınıf negatif ve gerçekte de negatif sınıfında olan örneklerin sayısını ifade eder.

Bu tanımlar ile başarımlar ölçütleri hesaplanır. Buna göre; doğruluk tüm örnekler içerisinde doğru sınıflandırılan pozitif ve negatif sınıfına ait olan örneklerin yüzdelerini, kesinlik sistemin yaptığı pozitif sınıf tahminlerinin kaçının gerçekte pozitif sınıfına ait olduğunun yüzdesini, duyarlılık aslında pozitif sınıfa ait olan örneklerin kaçının doğru tahmin edildiğinin yüzdesini göstermektedir. F ölçütü ise kesinlik ve duyarlılık değerlerinin harmonik ortalamasını ifade etmektedir.

Başarımlar ölçütleri denklem şeklinde ifade edilirse:

$$Doğruluk = \frac{TP + TN}{TP + FP + TN + FN} \quad (3.6)$$

$$Kesinlik = \frac{TP}{TP + FP} \quad (3.7)$$

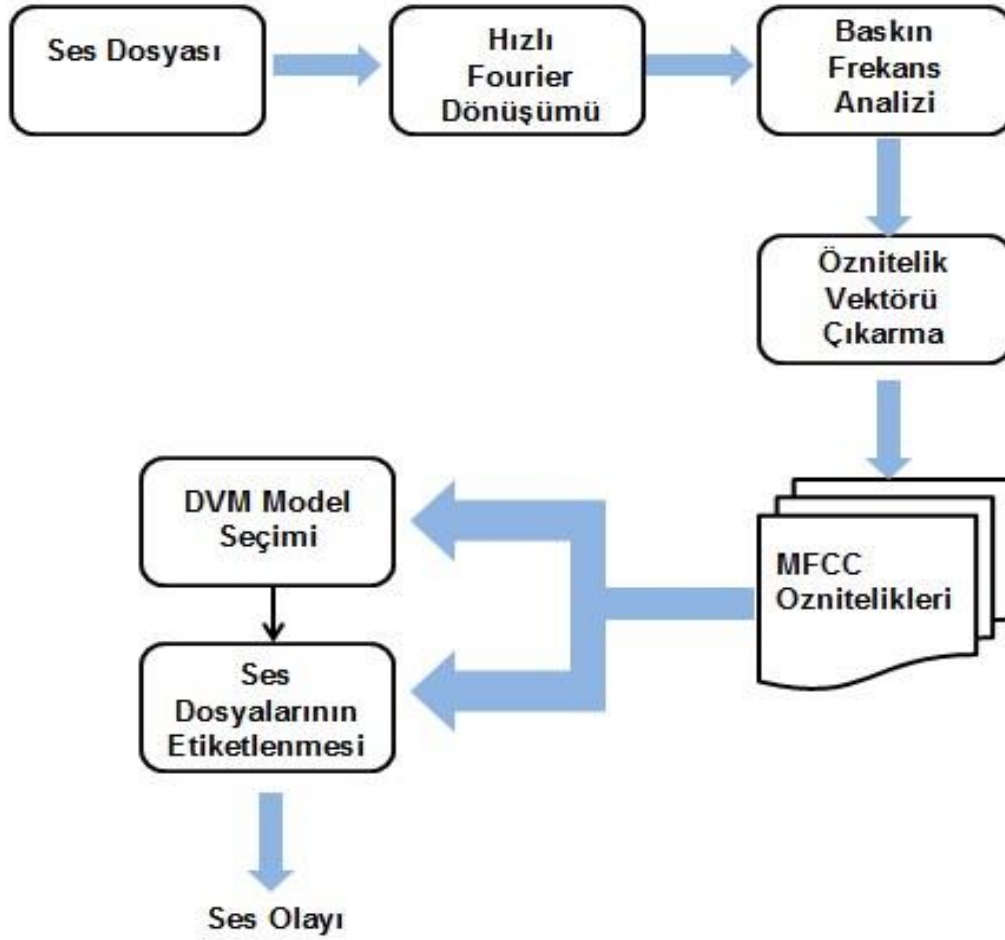
$$Duyarlılık = \frac{TP}{TP + FN} \quad (3.8)$$

$$F - \text{Ölçütü} = \frac{2 * Kesinlik * Duyarlılık}{Kesinlik + Duyarlılık} \quad (3.9)$$

4. SES OLAY SEZİMİ

Tez çalışmasının ana amaçlarından birincisi, verilen bir ses klibinin hangi ses olayına ait olduğunun sezimlenmesidir. Bu bölümde, problemin çözümü için önerdiğimiz yöntemler sunulmaktadır.

Önerilen sistem, Hızlı Fourier Dönüşümü, Baskın Frekans Analizi, Öznitelik Vektörlerini Çıkarma, MFCC Öznitelikleri, DVM Model Seçimi, Ses Dosyalarının Etiketlenmesi olarak altı ana aşamadan oluşmaktadır. Bu aşamaları gösteren sistemin genel blok diyagramı Şekil 4.1 'de verilmiştir.



Şekil 4. 1 Ses olayı sezimi için önerilen sistemin blok şeması

4.1. Öznitelik Çıkarımı ve Gösterimi

Her bir ses olayı farklı karakteristiğe sahip olduğundan farklı bilgiler taşır. Diğer bir deyişle, ses olayları farklı frekans bileşenlerinden oluşabilir. Konuyla ilgili çalışmalar incelendiğinde, gürbüz öznitelik çıkarım yöntemleriyle uygulanan yöntemler ses olaylarını sezimleme konusunda iyi sonuçlar vermektedir [4, 6, 10]. Ancak, öznitelik çıkarımı aşamasında çoğunlukla standart parametre ve değerler kullanılmaktadır. Örneğin, MFCC öznitelik çıkarımı aşamasında, katsayılar belirli bir frekans aralığından (alt ve üst sınır) elde edilmektedir. Fakat her bir ses dosyası için aynı frekans aralıklarını kullanmak, bu aralığın uygun olmadığı ses klipleri için önemli bilgilerin kaybolmasına yol açabilir. Bu durumun, kullanılan makine öğrenme algoritmalarının başarımlarını olumsuz yönde etkilemesi kaçınılmazdır. Tezin özgün katkılarından birisi de, bu aşamada uygulanan, uyarlanabilir öznitelik çıkarım yöntemidir.

Tezde önerilen öznitelik çıkarım yöntemi, veri kümesindeki her bir ses klbinin frekans spektrumundaki baskın frekansların bulunması ve elde edilen frekans değerlerinin MFCC öznitelik çıkarım aşamasında kullanılması prensibine dayanır.

Bu çalışmada, MFCC öznitelik çıkarımı için [45]' de verilen gerçekleştirim kullanılmıştır. Gerçekleştirim platformu olarak MATLAB² seçilmiştir. MFCC öznitelik vektörlerini oluşturabilmek için, analiz edilecek girdi ses dosyasının alt ve üst frekans sınırlarına ihtiyaç duyulmaktadır. Bu değerler standart MFCC öznitelik çıkarım yönteminde, alt frekans değeri için 300 Hz, üst frekans değeri için ise 3700 Hz olarak tanımlanmaktadır. Varsayılan değerler ile işlem yapılması durumunda, her ses olayının farklı frekans karakteristikleri olabileceğinden, önemli bilgiler kaybedilebilir. Analizlerde kullanılacak üst frekans sınırı, Nyquist Teoremine [46] göre belirlenebilir. Örneğin, örneklem hızı 44,1 kHz olan bir ses dosyası içerisindeki en yüksek frekans, Nyquist teoremine göre 22050 Hz olabilmektedir. Ancak, analizlerde kullanılacak alt frekans sınırı her bir ses klbi için farklı olabilmektedir. Bu sınırın doğru seçilmesi, elde edilecek özniteliklerin gürbüzlüğü açısından önemlidir. Bu problemin çözümü için, sınıf bazında baskın frekans analizi gerçekleştiriyoruz. Bu çalışmada baskın frekans, bir ses klbi içerisinde genliği en büyük olan frekans olarak tanımlanmıştır. Buna göre, sınıf bazındaki

² <http://www.mathworks.com/products/matlab/>

baskın frekanslar (4.1)'de önerdiğimiz denklem ile hesaplanmaktadır. E_i i. ses sınıfını temsil etmek üzere, $f_{dominant}(E_i)$, i. sınıfın baskın frekansını vermektedir. Denkleme göre, bir sınıfın baskın frekansı, o sınıftaki her bir ses klipinin baskın frekanslarının aritmetik ortalaması olarak tanımlıdır:

$$f_{dominant}(E_i) = \sum_{(k=0) \in E_i}^{N-1} f_k(\text{idx}(\max(|F_k|))) \quad (4.1)$$

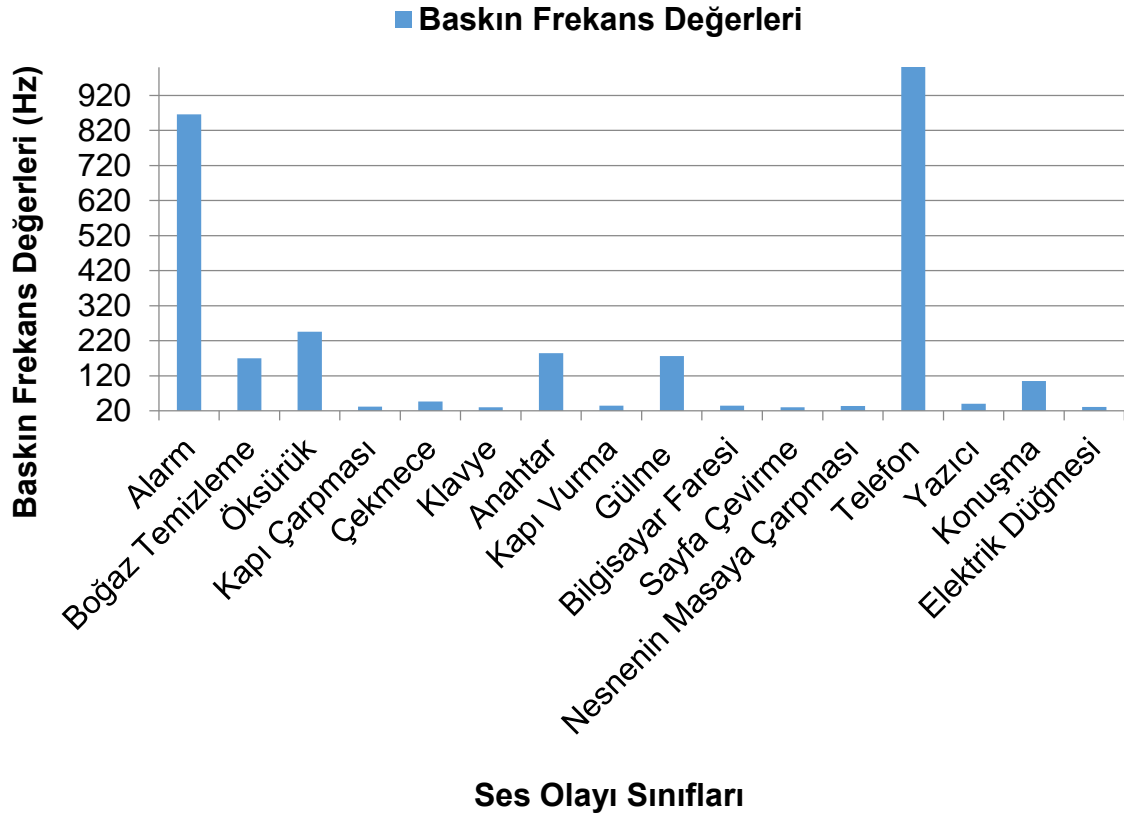
Burada N, i. ses olayı sınıfının toplam klip sayısı olmak üzere; F_k , k. ses klipinin Fourier dönüşümünü, $\text{idx}(y)$, y'nin indis numarasını, $f_k(z)$ ise k. ses klipinin z indisindeki frekans değerini temsil etmektedir. Her sınıf için hesaplanan $f_{dominant}$ değeri, MFCC öznitelik çıkarımı aşamasında alt frekans sınırı olarak kullanılır.

Öncelikle, verilen bir ses olayı sınıfındaki bütün ses klipleri frekans düzlemine dönüştürülür. Buradan, ses klipi içerisinde genliği en büyük olan frekans hesaplanır. Bu hesaplama yapılırken, en büyük genliğin olduğu indis bulunup, bu indisteki frekans değeri alınır. Bu frekans değeri, baskın frekans değeri olarak adlandırılır. Sınıf bazında bir aralık belirlemek için, sınıftaki örneklerin baskın değerlerinin ortalaması alınır. Böylelikle, hesaplanan bu baskın frekans değeri, MFCC öznitelik çıkarımı aşamasında kullanılacak alt frekans değeri olarak alınır. Üst frekans değeri tüm sınıflar için aynı ve 22050Hz olarak belirlenmiştir.

Çalışılan veri kümesi için elde edilen baskın frekans değerleri Şekil 4.2'de verilmiştir.

MFCC öznitelik vektörlerinin çıkarımı aşamasında, farklı MFCC katsayıları ve farklı pencere ve atlama süreleri kullanılarak en uygun öznitelik değerlerini bulmaya yönelik ampirik değerlendirmeler yapılmıştır. Ayrıca, pencere analizlerinde, farklı katsayı sayılarının başarıma etkisi incelenmiştir. Bu amaçla, varsayılan değer olarak kullanılan 13-katsayılı MFCC'ye ek olarak 20, 48 ve 128 katsayılı MFCC öznitelikleri çıkarılmış ve sonuçları değerlendirilmiştir. Değerlendirme sonuçları Bölüm 4.3.3'de sunulmaktadır. Deneysel sonuçlara göre, en yüksek başarımla 20-katsayılı MFCC ile elde edildiğinden, öznitelik çıkarımında bu değer kullanılmıştır. Bir sonraki aşamada, MFCC özniteliğini çıkarım aşaması için en uygun pencere ve atlama boyutlarını bulabilmek için ampirik analizler yapılmıştır. Pencere süresi, doğru bir öznitelik çıkarım aşaması için önemli parametrelerden biridir. Pencere,

belirli uzunluktaki ses bölümlerini ifade eder. Tüm sinyali aynı anda incelemek yerine, veriyi anlamlı küçük parçalara ayırmak suretiyle sinyal içerisindeki değişimleri kolayca vurgular. Seçilen pencere süresi, probleme göre farklılık gösterebilir. Atlama süresi, ses sinyali içerisinde seçilen pencere süresi ile birlikte kullanılır ve analiz edilecek bir sonraki pencerenin nereden başlayacağını tanımlar. Diğer bir deyişle, pencere boyutlarının ne kadar atlama ile ilerleyeceğini gösterir. Eğer, atlama süresi pencere süresinden büyük olursa, ses sinyali içerisinde bazı bilgiler kaybedilebilir. Bu nedenle, her bir pencere süresinde bilgi kaybını engellemek adına örtüşmelerin olması gerekmektedir. Bu nedenle, farklı pencere (10ms, 20ms, 30ms, 50ms, 100ms) ve atlama (5ms, 10 ms, 15ms, 20ms, 25ms, 30ms, 50ms, 100ms) süreleri için testler yapılmış ve sınıflandırma başarımı en yüksek olan değerlere karar verilmiştir. Yapılan deneysel çalışmalara göre en uygun değerler pencere süresi (t_w) için 30ms, atlama süresi (t_h) için ise 10ms olarak belirlenmiştir.



Şekil 4. 2 Her bir ses olayı için baskın frekans değerleri

Öznitelik çıkarımı aşamasında kullanılacak pencere ve atlama süreleri, verilen bir ses klibinden çıkarılacak öznitelik vektörlerinin boyutunu tanımlar: n bir ses klibindeki analiz pencere sayısı ve m , 1 pencereden çıkarılacak MFCC katsayı sayısını temsil etmek üzere, bu ses klibinin öznitelik vektörünün (F) boyutu ($n \times m$) şeklide tanımlıdır. Örneğin; 30 saniyelik bir ses klibinden, $t_w = 30ms$ ve $t_h = 10ms$ 20-katsayılı MFCC özniteliği çıkarılırsa, n değeri aşağıdaki denklem kullanarak hesaplanır:

$$n = t/t_h \quad (4.2)$$

Burada t , ses klibinin milisaniye cinsinden süresini ifade etmektedir. Yukarıda bahsedilen örnek için $n = (30 \times 1000) / 10ms = 3000$ olarak hesaplanır. O halde 30 saniyelik bir ses dosyasının öznitelik vektörünün (F) boyutu 3000×20 olmaktadır. Bu gösterimde, F 'nin her satırı, bir analiz penceresinin 20-katsayılı MFCC özniteliğini göstermektedir.

Son olarak, klip bazında karar vermek için, boyutu ($n \times m$) olan öznitelik vektörü F , boyutu ($1 \times m$) olan F' vektörüne dönüştürülür. F' vektörü, F vektörünün her sütun aritmetik ortalaması alınarak hesaplanmıştır.

$$F = \begin{bmatrix} F_{1,1} & \cdots & F_{1,m} \\ \vdots & \ddots & \vdots \\ F_{n,1} & \cdots & F_{n,m} \end{bmatrix}_{n \times m} \quad (4.3)$$



$$F' = [F_{1,1} \quad \cdots \quad F_{1,m}]_{1 \times m} \quad (4.4)$$

Model eğitimi ve test işlemleri sırasında öznitelik vektörü olarak F' (4.4) kullanılmaktadır.

4.2. Sınıflandırıcı Tasarımı

Sınıflandırıcı yöntemi olarak, literatürdeki yaygın kullanımı ve örüntü tanıma problemlerindeki başarımı nedeniyle DVM öğrenme yöntemi seçilmiştir. DVM yönteminin veri kümesine uygulanması için MATLAB ortamında LIBSVM Kütüphanesi [23] kullanılmıştır. DVM yöntemi temelde ikili (+/-) sınıflandırma algoritması olduğundan ve bu çalışmada 16 sınıftan oluşan çoklu-sınıflandırma problemi ele alındığından, DVM algoritması bire-karşı-hepsi stratejisi kullanılarak çoklu sınıflandırma problemine uyarlanmıştır. Bu durumda, 16 farklı ses olayı sınıfı için, eğitim aşamasında 16 model oluşturulmuştur. Model eğitimleri sırasında, eğitilen sınıfa ait örnekler pozitif olarak işaretlenirken, geri kalan sınıflara ait örneklerin tamamı negatif olarak işaretlenmiştir. Doğrusal olmayan sınıflandırma problemlerinde (soft marjin) DVM algoritmasının iki farklı parametresinin hata maliyeti (C) ve gama (γ) eniyilemesi için ampirik testler gerçekleştirilmiştir. Karşılaştırmalı sonuçlara göre en iyi sonuçları veren hata maliyeti (C) ve gamma değeri (γ) değerleri kullanılmıştır. Bu parametrelerin anlamları aşağıda verilmektedir.

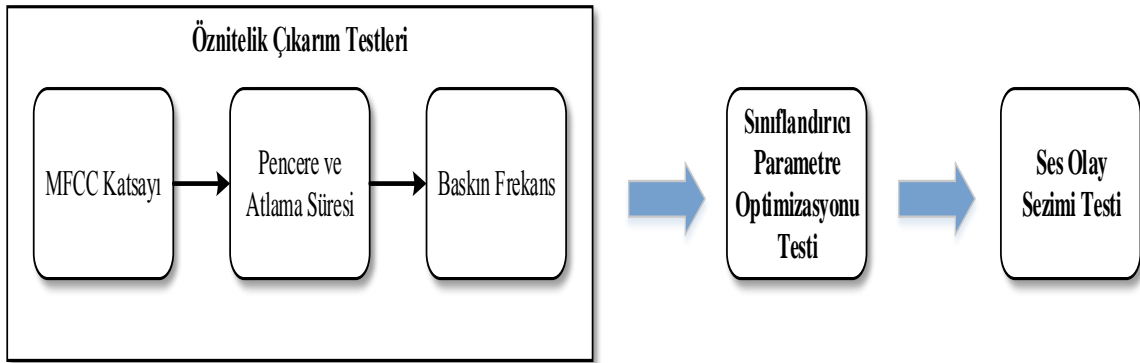
Hata Maliyeti (C): Hata maliyeti marjin maksimizasyonu ile deneme hatası minimizasyonu arasındaki değiş tokuşu belirler. Yani C sabiti yanlış örneklerin maliyetteki ağırlığını belirtmektedir. Yüksek alınan C değeri için, yüksek hata beklentisi var demektir. Böylelikle DVM oldukça az yanlış karar vermeye çalışacaktır. Bu bağlamda da test aşamasında verilerin doğru sınıflandırılması hedeflenir. Doğru sınıflandırma oranının artması demek sınıflandırıcı karmaşıklığının düşmesi demektir. C değerinin düşük alınmasında durumlar tam tersidir. Bu durumda daha az örnek yanlış sınıflandırılacak, dolayısıyla sınıflandırıcı karmaşıklığı artacaktır.

Gama (γ): Sezgisel olarak, gama parametreleri kullanılan RTF genliğini kontrol ederek DVM'nin genelleştirme yeteneğini kontrol etmektedir. Düşük alınan gama değeri, düşük etki ve yüksek varyans verirken, yüksek seçilen bu değer yüksek etki ve düşük varyans vermektedir. Gama değeri hiper düzlemde karmaşıklığın derecesini kısıtlar.

Verilen bir problemde, hangi C ve ν değerlerinin en iyi olduğu daha önceden tahmin edilemez. Bu nedenle parametrelerin ayarlanarak en iyilemesi gerekmektedir. Genel olarak bu parametrelerin en iyilemesi gerçekleştirilirken Grid arama algoritması kullanılır [23]. Bu algoritma, öğrenme algoritması için belirlenen alt küme parametreleri ile detaylı bir arama kurar. Grid arama algoritması veri kümesi üzerinde çapraz doğrulama gibi performans ölçütleri ile birlikte gerçekleştirilir. Grid arama algoritması çalıştırılırken, veri kümesindeki tüm verileri ve değerlendirme işleminin çapraz doğrulama değerini Veri kümesi değerlendirme aşamasında 5- kat çapraz doğrulama kullanıldığından, Grid arama algoritması gerçekleştirildiğinde çıkan en uygun parametre ikilisi $C=0,03125$ ve $\nu = 0,007825$ değerlerini almaktadır.

4.3 Deneysel Çalışmalar ve Değerlendirme

Tez kapsamında yapılan çalışmalardan ilki olan ses olay sezimi için uyguladığımız yöntemler test edilmiş ve sonuçları değerlendirilmiştir. Gerçekleştirilen testler üç ana başlık altında toplanmaktadır: Öznitelik çıkarım testleri, sınıflandırıcı parametre optimizasyonu ve ses olayı sezimi (Şekil 4.3).



Şekil 4. 3 Gerçekleştirilen testler ve akışları

Kullanılan veri kümesi ve gerçekleştirilen testlerin sonuçları izleyen bölümlerde sunulmuştur.

4.3.1 Ses olayları veri kümesi

Literatürde müzik ve konuşma alanında erişime açık çeşitli veri kümeleri bulunmasına rağmen, ses olayları için çok daha az sayıda güvenilir kaynak bulunmaktadır. Bu tez çalışmasında DVM model eğitimi ve test aşamalarında erişime açık bir veri kümesi olan *IEEE AASP Challenge Detection and Classification of Acoustic Scenes and Events* isimli veri kümesi kullanılmıştır [21]. Bu veri kümesi içerisindeki veriler üç gruba ayrılmaktadır. Bunlar; sentetik olarak oluşturulmuş ses olayları, gerçek ortamlarda oluşturulmuş ses olayları ve gerçek ortamlarda oluşturulmuş işitsel sahne veri kümeleridir. Çalışmamızda, tasarladığımız sistemin gürbüzlüğünü değerlendirebilmek amacıyla, gerçek ortamlardan elde edilmiş ses olayları (ofis – gerçek ortam) kullanılmıştır. Seçilen veri kümesi, ofis ortamında kaydedilmiş günlük hayata ait ses olaylarını içermektedir. Bu ses olayları; *alarm sesi, boğaz temizleme sesi, öksürük, kapının çarpma sesi, çekmece klavye, anahtar, kapıyı çalma sesi, gülme, bilgisayar faresinin çıkardığı ses, sayfaların çevrildiğinde çıkardığı ses, bir nesnenin masaya düşerken çıkardığı ses, telefon, yazıcı, konuşma ve elektrik düğmelerinin açılıp kapanma seslerinden oluşmaktadır*. Bu veri kümesinde var olan 16 farklı ses olayları tezin ilerleyen kısımlarında *sınıflar* olarak isimlendirilmiştir. Her bir sınıf 20 örnek içermektedir ve her örneğin süreleri farklıdır. Sınıfların toplam süreleri Çizelge 4.1’de verilmiştir. Verilen tüm örnekler işlenmemiş (ham) formattadır ve örneklem hızı 44,1 kHz, çift kanal ve her bir örnek 24 bit derinlik ile temsil edilmektedir.

4.3.2 Değerlendirme yöntemi

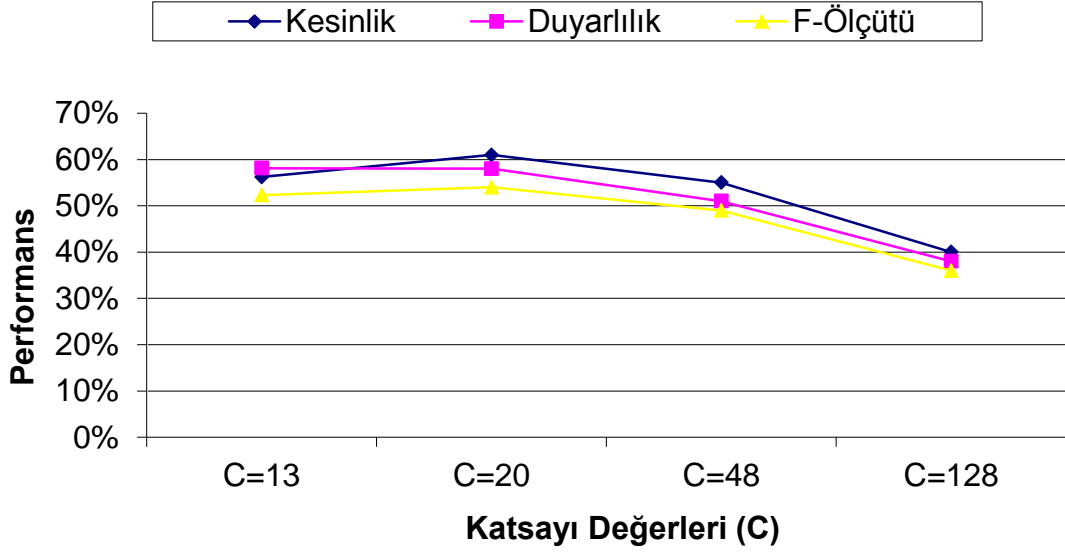
Veri kümesi üzerinde yapılan testler sonucunda bir karar vermek gerektiğinde, 2 seçenek ortaya çıkmaktadır. İlk seçenek, her bir ses klibinin tüm pencereleri için herhangi bir ses olayına karar verme, ikinci seçenekte ise bir ses klibinin tamamı için bir sınıfa karar vermektir. İlk seçenek düşünüldüğünde, seçilen t_w değerine göre, her bir pencere için sistemin o parçaya bir etiketleme yapması gerekecektir. Kullandığımız veri kümesindeki her bir ses olayı klipi, sadece tek bir sınıfa ait olduğundan (başka ses olayları içermediğinden) klip-bazlı etiketleme yöntemine karar verilmiştir. Yani, her bir ses klipi için sistem test edilen klibe bir ses olayı ataması yapacaktır.

Çizelge 4. 1 Veri kümesindeki ses olayı sınıfları ve toplam süreleri

Sınıf İsmi	Süresi
Alarm	40 sn
Boğaz Temizleme	23 sn
Öksürük	23 sn
Kapı Çarpması	44 sn
Çekmece	33 sn
Klavye	1 dk 16 sn
Anahtar	41 sn
Kapı Vurma	26 sn
Gülme	30 sn
Bilgisayar Faresi	29 sn
Sayfa Çevirme	1 dk 03 sn
Nesnenin Masaya Çarpması	16 sn
Telefon	3 dk 05 sn
Yazıcı	7 dk 01 sn
Konuşma	1 dk
Elektrik Düğmesi	10 sn
TOPLAM	18 dk 49 sn

4.3.3 MFCC katsayı testleri

Bu testte, analiz penceresinden çıkarılan MFCC katsayıların sayısının başarıma etkisi incelenmiştir. Varsayılan değer olan 13-katsayılı MFCC'ye ek olarak 20, 48 ve 128 katsayılı MFCC öznitelikleri çıkarılarak sınıflandırma gerçekleştirilmiştir. Elde edilen sonuçlar Şekil 4.4'de gösterilmektedir. Test veri kümesinde en yüksek kesinlik, duyarlılık ve F-ölçütü 20-katsayılı MFCC için elde edildiğinden, sonraki aşamalarda bu değer kullanılmıştır. Farklı MFCC katsayıları için yapılan sınıflandırmaların hata matrisleri A.1, A.2, A.3 ve A.4 olarak verilmiştir.



Şekil 4. 4 Farklı MFCC katsayıları için elde edilen başarımlar

4.3.4 Pencere ve atlama süreleri testleri

Bu testte, farklı t_w ve t_h değerlerinin ses olayı tanıma başarımına olan etkileri incelenmiştir. t_w seçiminde, literatürde yaygın olarak kullanılan değer aralıkları (10ms, 20ms, 30ms, 50ms, 100ms); t_h seçiminde ise, pencerelerin iki farklı örtüşme durumu ve örtüşmeme durumları dikkate alınmıştır.

Gerçekleştirilen testler sonucunda elde edilen sonuçlar Çizelge 4.2 'de sunulmaktadır. Her bir durum için gerçekleştirilen 13 testin hata matrisleri A.5 – A.17 olarak verilmiştir. Çizelge 4.2'de görüldüğü üzere, $t_w = 10\text{ms}$ ve $t_h = 5\text{ms}$ için en iyi performans değerini vermektedir. Ancak, bu değerler $t_w = 30\text{ms}$ ve $t_h = 10\text{ms}$ alınarak gerçekleştirilen testlerin sonuçlarıyla kıyaslandığında, aralarında anlamlı bir farkın olmadığı görülmektedir. Küçük pencere/atlama sürelerinin getireceği bellek/hesaplama karmaşıklılığı baz alınarak, çalışmada kullanılacak analiz pencere ve atlama süreleri sırasıyla $t_w = 30\text{ms}$ ve $t_h = 10\text{ms}$ olarak seçilmiştir.

Çizelge 4. 2 Farklı pencere ve atlama süreleri için ölçülen performanslar

t_w (ms), t_h (ms)	Kesinlik (%)	Duyarlılık (%)	F-ölçütü (%)
10,5	63	58	56
10,10	60	58	56
20,10	59	57	55
20,20	59	57	55
30,10	62	58	55
30,15	56	55	52
30,30	56	56	53
50,10	53	55	51
50,25	57	56	53
50,50	57	55	53
100,10	59	58	55
100,50	56	56	53
100,100	57	57	54

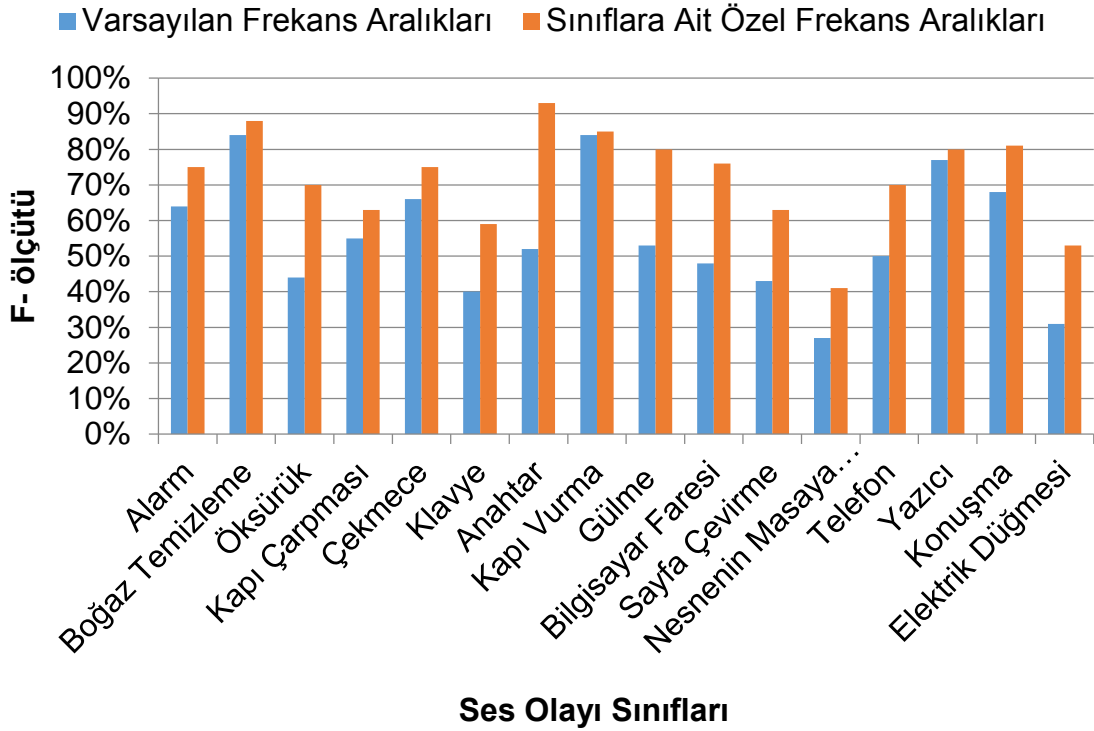
4.3.4 Baskın frekans testleri

Bu bölümde, MFCC öznitelikleri, standart ve baskın frekans analizi yöntemi ile çıkarıldığında elde edilen başarımlar test edilmiştir. MFCC öznitelik çıkarımında varsayılan alt ve üst frekans sınırları (300Hz ve 3700Hz) kullanılarak eniyilenmiş DVM ile eğitilen model test edildiğinde elde edilen hata matrisi A.18'de verilmiştir. Hata matrisinden de okunacağı üzere klavye, elektrik düğmesi ve nesnenin masaya çarpması sınıfları sayfa çevirme ve anahtar sınıfları ile büyük oranda karışmaktadır. Bunun nedeni, sabit frekans aralığında (300Hz-3700Hz) çıkarılan MFCC özniteliklerinin tüm ses kliplerini aynı oranda karakterize edemiyor olması olarak açıklanabilir.

MFCC öznitelikleri, bu çalışmada önerilen baskın frekans analizi yöntemi ile çıkarıldığında elde edilen hata matrisleri A.19'da sunulmuştur. Bu hata matrisinden de görüldüğü üzere, A.18'de büyük oranda karışan öksürük, elektrik sınıflarındaki örneklerin doğru tanımlama performansının (sırasıyla +6, +5) arttığı gözlenmektedir. Böylelikle, diğer sınıflarla karışma durumunda azalma meydana gelmektedir. Varsayılan ve önerilen yöntemle elde edilen genel sınıflandırma

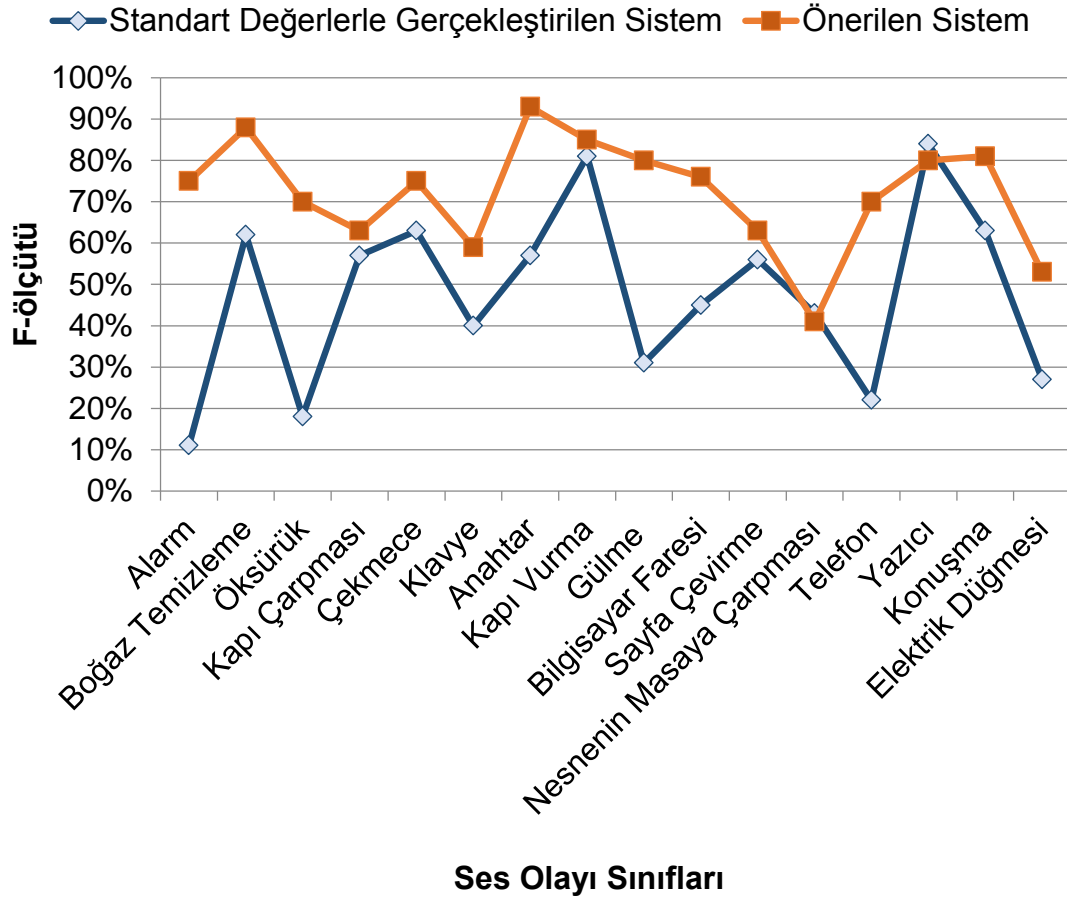
başarımı Şekil 4.5' de gösterilmiştir. Buna göre, tüm sınıfların tanıma başarımında anlamlı artışlar olmuştur. En küçük artış, %4 oranı ile yazıcı sınıfında, en büyük artış ise %41 oranı ile anahtar sınıfında kaydedilmiştir.

Çalışmada önerilen yöntemlerle, genel başarımda F-ölçütü değeri ile %17 oranında bir artış elde edilmiştir. Aynı zamanda, %72'lik F-ölçütü skoru ile veri kümesi üzerinde klip bazında değerlendirme türünde literatürdeki en iyi başarımların (%27) elde edilmiştir [35].



Şekil 4. 5 Varsayılan frekans aralıkları ve sınıflara ait özel frekans aralıkları ile çıkarılan öznelitekt vektörleriyle eğitilen sistemin performansı

Sonuç olarak, önerilen ve standart olarak kullanılan yöntemlerin performansları karşılaştırmalı olarak Şekil 4.6' de sunulmuştur. Buna göre önerilen yöntem, *yazıcı* ve *nesnenin masaya çarpması* sınıfları haricinde diğer tüm sınıflarda standart yöntemlerin üzerinde bir başarımlar vermiştir.

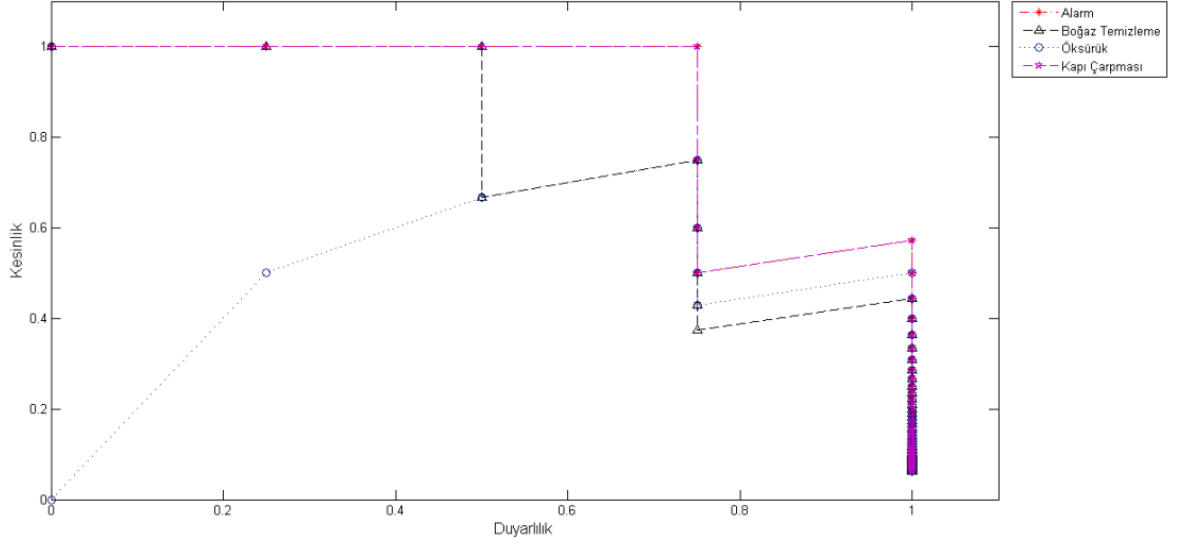


Şekil 4. 6 Standart ve önerilen yöntem başarımları

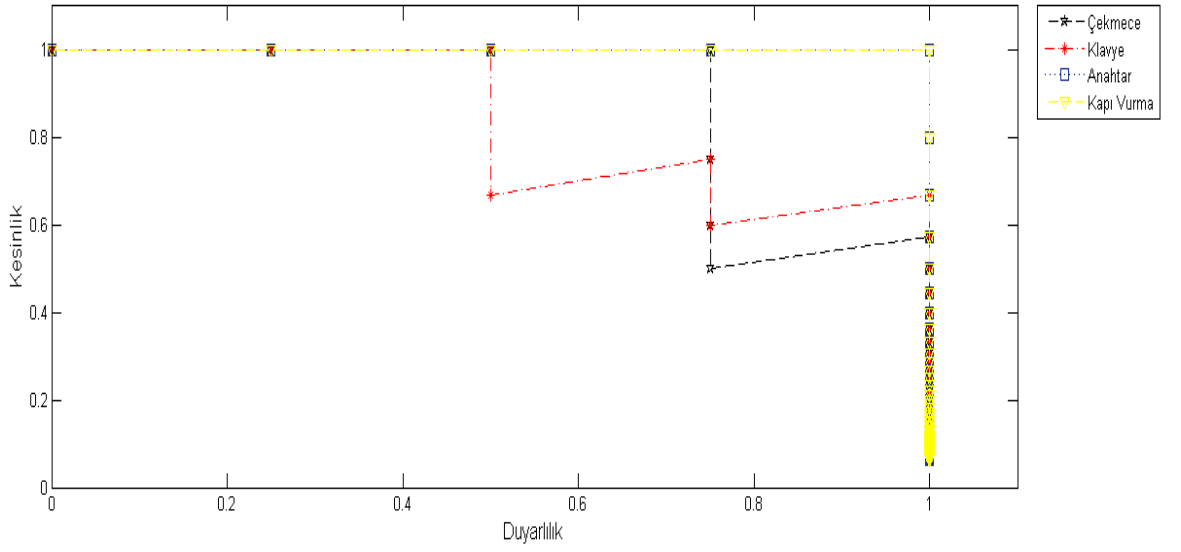
4.3.5 Kesinlik-Duyarlılık grafikleri

Kesinlik-duyarlılık grafiği, bilgi geri erişim sistemlerinde, arama stratejilerinin değerlendirilmesi aşamasında yaygın olarak kullanılan ölçütlerden birisidir. Bu çalışma sonuçlarının, ilgili alanlardaki (bilgi geri erişim vb.) olası kullanımlarını değerlendirebilmek amacıyla, 16 aylık ses olayı sınıflarının kesinlik-duyarlılık (precision-recall) grafikleri verilmiştir (Şekil 4.7 – 4.10).

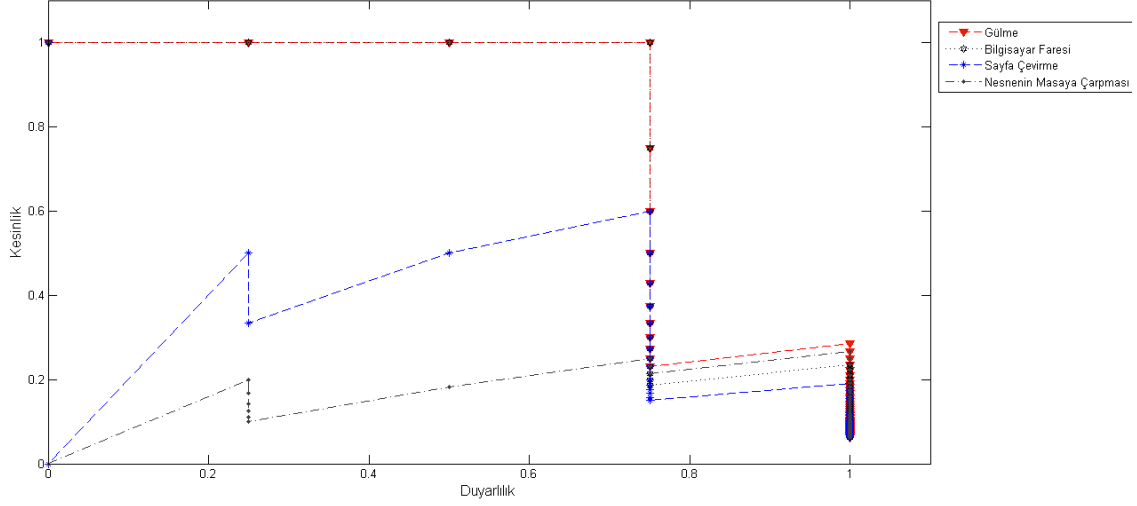
Ses olayı sınıfları fazla sayıda olduğundan, sonuçlar gruplar halinde dört farklı grafikte sunulmuştur. Kesinlik-duyarlılık grafikleri bilgi geri erişim sistemlerinin performans metriği olarak adlandırılmaktadır.



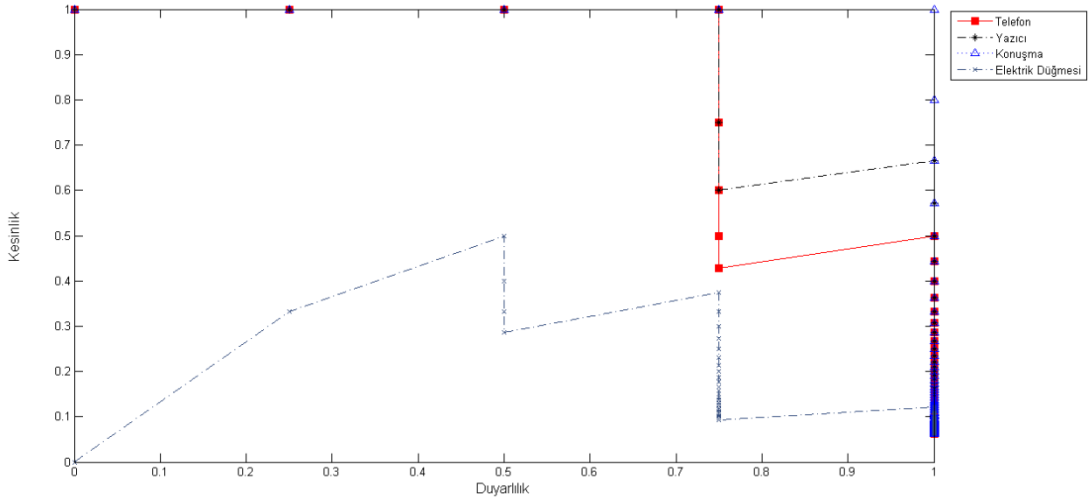
Şekil 4. 7 Alarm, boğaz temizleme, öksürük ve kapı çarpması sınıfları için elde edilen keskinlik-duyarlılık grafiği



Şekil 4. 8 Çekmece, klavye, anahtar ve kapı vurma sınıfları için elde edilen keskinlik-duyarlılık grafiği



Şekil 4. 9 Gülme, bilgisayar faresi, sayfa çevirme ve nesnenin masaya çarpması sınıfları için elde edilen kesinlik-duyarlılık grafiği



Şekil 4. 10 Telefon, yazıcı, konuşma ve elektrik düğmesi sınıfları için elde edilen kesinlik-duyarlılık grafiği

Elde edilen sonuçlara göre *elektrik düğmesi*, *nesnenin masaya çarpması* ve *sayfa çevirme* sesleri haricindeki diğer 13 sınıfın başarılı sonuçlar verdiği değerlendirilmektedir.

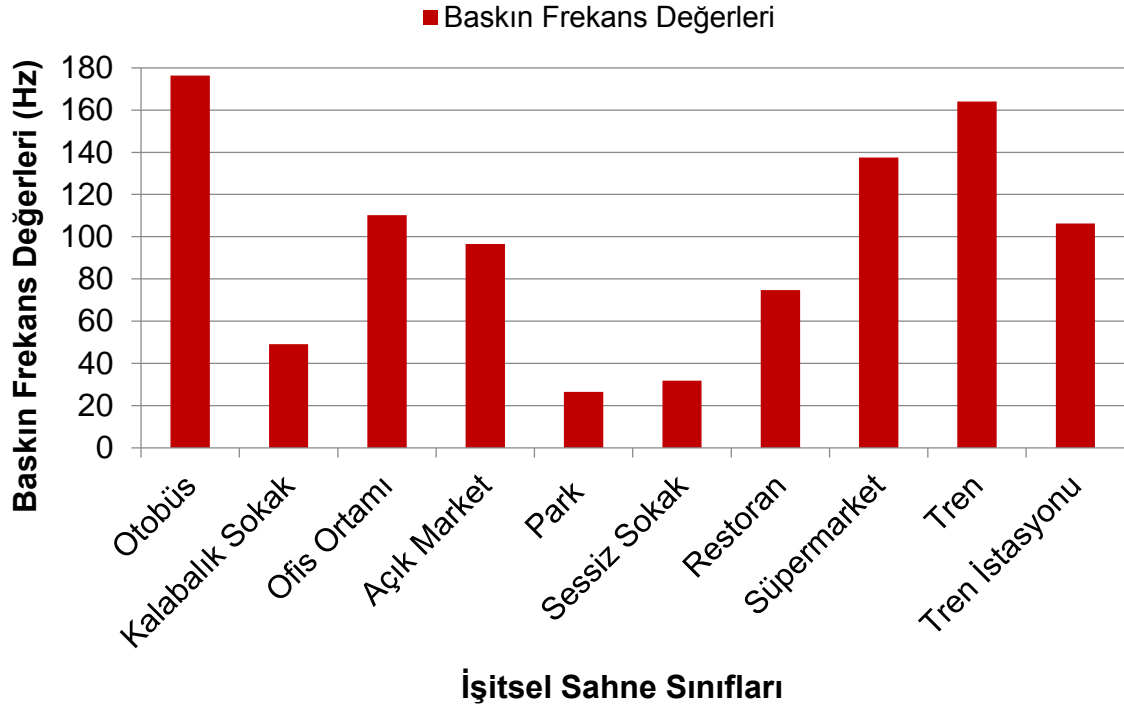
5. İŞİTSEL SAHNE TANIMA

Tez çalışmasının ikinci kısmında, ses olay sezimi için önerilen yöntemleri kullanarak, işitsel sahnelerin ayrık olarak sınıflandırılması hedeflenmiştir.

5.1 Öznitelik Çıkarımı ve Gösterimi

İşitsel sahneler içeriğinde ses olaylarını barındırdığından, aynı ses olayları gibi farklı karakteristikteki seslerin bir araya gelmesiyle meydana gelirler. Her bir ses olayı farklı karakteristiğe sahip olduğundan farklı bilgiler taşır. İşitsel sahnelerin tanınması için, Bölüm 4'deki yöntemler, elde edilen en iyi parametre değerleri ile kullanılmıştır. Pencere analizleri $t_w = 30\text{ms}$ ve $t_h = 10\text{ ms}$ kullanılarak gerçekleştirilmiştir. 13, 20, 48 ve 128 katsayılı MFCC özniteliklerinden en yüksek başarımla elde edilen 20-katsayılı MFCC kullanılmıştır. Ses olaylarında olduğu gibi, işitsel sahnelerdeki sınıflar için baskın frekans değerleri (4.1) denklemindeki gibi hesaplanarak alt frekans değerleri elde edilmiştir. Bu veri kümesindeki örneklerin örneklem hızı 44,1 kHz olduğundan üst frekans değeri Nyquist Teoremine göre 22050 Hz olarak alınmıştır. Toplamda 10 farklı işitsel sahne için 10 frekans aralığı belirlenmiş ve MFCC öznitelik çıkarım aşamasında bu değerler kullanılmıştır. İşitsel sahne sınıfları ve bu sınıflara ait baskın frekans değerleri Şekil 5.1 'de gösterilmiştir.

MFCC özniteliği çıkarılırken her bir ses dosyası için vektör uzunluğu hesaplaması 4. Bölümde verilen (4.3) ve (4.4) denklemlerine göre yapılmıştır. Bu şekilde $n \times m$ boyutlarındaki öznitelik vektörleri $1 \times m$ boyutlarına dönüştürülerek bu formatlar üzerinden model eğitimi ve test işlemi gerçekleştirilmiştir.



Şekil 5. 1 İşitsel sahne sınıflarının baskın frekans değerleri

5.3 Sınıflandırıcı Tasarımı

Sınıflandırıcı yöntemi olarak veri kümesine uygun parametre ayarlaması yapılarak DVM öğrenme yöntemi seçilmiştir. Yöntemler, LIBSVM kütüphanesi [23] ile MATLAB ortamında uygulanmıştır. Veri kümesindeki işitsel sahneler 10 sınıftan oluştuğu için, DVM algoritması ses olayı sezimi aşamasında olduğu gibi bire-karşı-hepsi stratejisi kullanılarak çoklu sınıflandırma problemine uygun hale getirilmiştir. Bu durumda 10 farklı işitsel sahne için, eğitim aşamasında 10 DVM modeli oluşturulmuştur. DVM algoritması hata maliyet (C) ve gama değerleri (γ) için grid arama algoritması kullanılarak optimize edilmiştir. Veri kümesini değerlendirebilmek için 5 katlı çapraz doğrulama yöntemi uygulanmıştır.

5.4 Deneysel Çalışmalar ve Değerlendirme

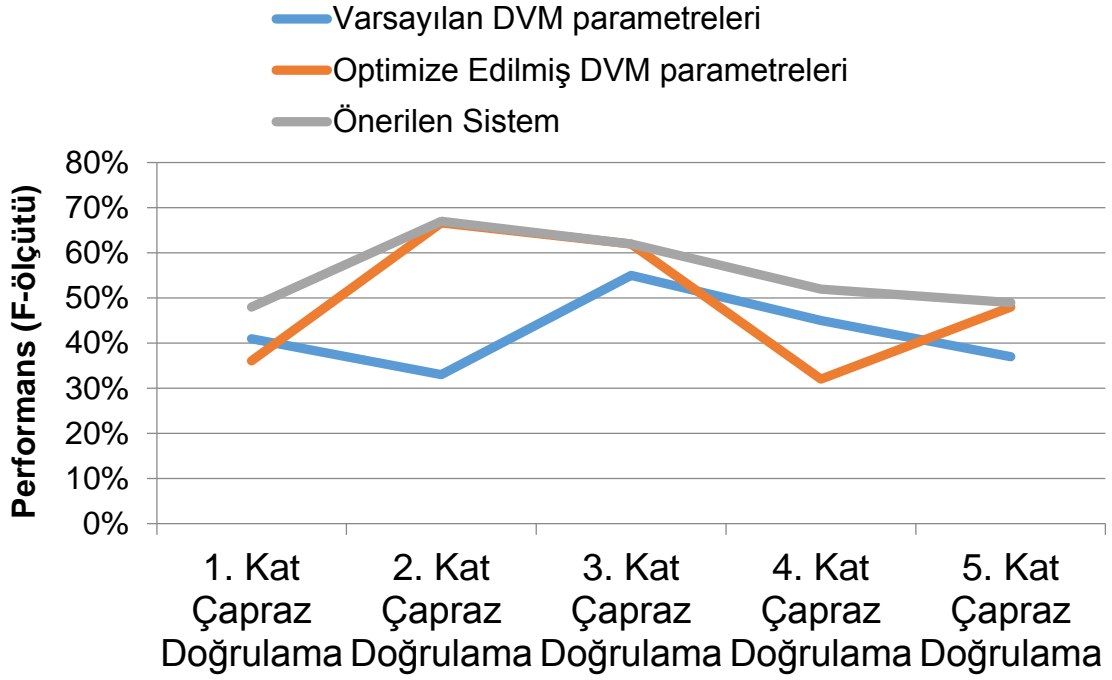
Tez çalışmasının ikinci aşaması olan işitsel sahne tanımlama problemi için, ses olay sezimi adımlarında uygulanan parametre en iyilemesi yapılmış, ses dosyaları için uyarlanabilir öznitelik çıkarım yöntemi gerçekleştirilmiştir. Frekans analizi yapılan ses dosyaları için baskın frekans değerleri bulunmuş ve sınıf bazında ortalama kullanılarak her sınıf için özel frekans aralıkları hesaplanmıştır. Bu

değerler elde edildikten sonra uygun parametrelerle optimize edilen MFCC öznitelik çıkarım aşamasında kullanılmıştır. Daha sonra bu vektörler kullanılarak model eğitimi ve test işlemi gerçekleştirilmiştir.

5.4.1 İşitsel sahneler veri kümesi

İşitsel sahne tanıma konusunda yapılan çalışmalarda farklı veri kümeleri kullanılmaktadır. Bunun önemli bir nedeni, genel erişime ve değerlendirmeye açık veri kümelerinin azlığıdır. Bu noktada önemli olan, sistemin daha gürbüz olabilmesi için, seçilen veri kümesinin günlük hayattan gerçek örneklerden alınmış olmasıdır. Bu kapsamda, erişime açık bir paylaşım ortamı olan *IEEE AASP Challenge Detection and Classification of Acoustic Scenes and Events* isimli veri kümesi kullanılmıştır [24]. Verilen bu veri kümesinde içerisinde veriler üç alt kategoriye ayrılmıştır. Bunlar; sentetik olarak oluşturulmuş ve ofis ortamından alınmış ses olayları, günlük hayattan ofis ortamından gerçek örnekler alınmış ses olayları ve gerçek ortamlardan alınmış işitsel sahne veri kümeleridir. Bu bağlamda işitsel sahnelerin tanımlanması için, bu kaynaktan paylaşılan işitsel sahne sınıflandırma alt kategorisinde bulunan veri kümesi seçilmiştir. Seçilen veri kümesi, *otobüs, kalabalık sokak, ofis ortamı, açık market (pazar), park, sessiz sokak, restoran, süpermarket, tren ve tren istasyonu* sahnelerinden oluşmaktadır. Her sınıf içerisinde 30 saniye uzunluğunda 10 örnek bulunmaktadır. Toplamda 10 sınıf için, 100 örnek vardır. Veri kümesinin toplam süresi yaklaşık 50 dakikadır. Verilen tüm örnekler işlenmemiş (ham) formattadır ve örneklem hızı 44,1 kHz, çift kanal ve her bir örnek 24 bit ile ifade edilmiştir.

Şekil 5.2' de, varsayılan DVM parametreleri, eniyilenmiş DVM parametreleri ve önerilen sistem için elde edilen sonuçlar karşılaştırmalı olarak verilmiştir. Varsayılan parametrelerle uygulanan DVM eğitimiyle gerçekleştirilen testlerin hata matrisleri B.1'de, uygun değerlerle en iyileştirilmiş DVM eğitimiyle gerçekleştirilen testlerin hata matrisleri B.2'de verilmiştir. Önerilen sistemin hata matrisleri ise B.3'de sunulmaktadır.



Çapraz Doğrulama Adımları

Şekil 5. 2 Sistemlerin çapraz doğrulama karşılaştırması

Şekil 5.2'ye göre, varsayılan değerler kullanılarak yapılan öznelik çıkarım ve sınıflandırıcı uygulama yöntemlerine göre, önerilen sistemin, genel tanımlama performansını arttırdığı görülmektedir. İşitsel sahneler, içlerinde farklı karakteristikte sesleri bulduran anlamsal ses bölümleridir. Bu sahneler, ses olaylarının birleşimi ile meydana gelmektedir. Bu sebeple tıpkı ses olayları gibi, belirli frekans değerlerine sahiptir. Bu amaçla bu ses dosyaları içinde öznelik çıkarım aşamasında kullanılmak üzere frekans analizleri gerçekleştirilerek uygun frekans aralıklarını bulmak önemlidir. Varsayılan yöntemler uygulandığında sistemin F-ölçütü başarımları %42 iken, önerilen yöntem ile bu oran %59'a çıkmaktadır.

5.5 Ses Olayı – İşitsel Sahne İlinti Analizi

Tez çalışmasının son aşamasında, ses olayları ile işitsel sahneler arasındaki olası ilintiler incelenmiştir. Daha önceki bölümlerde de açıklandığı üzere, işitsel sahneler içerisinde çeşitli ses olaylarını barındırmaktadır. Bu nedenle, bir ses klibindeki ses

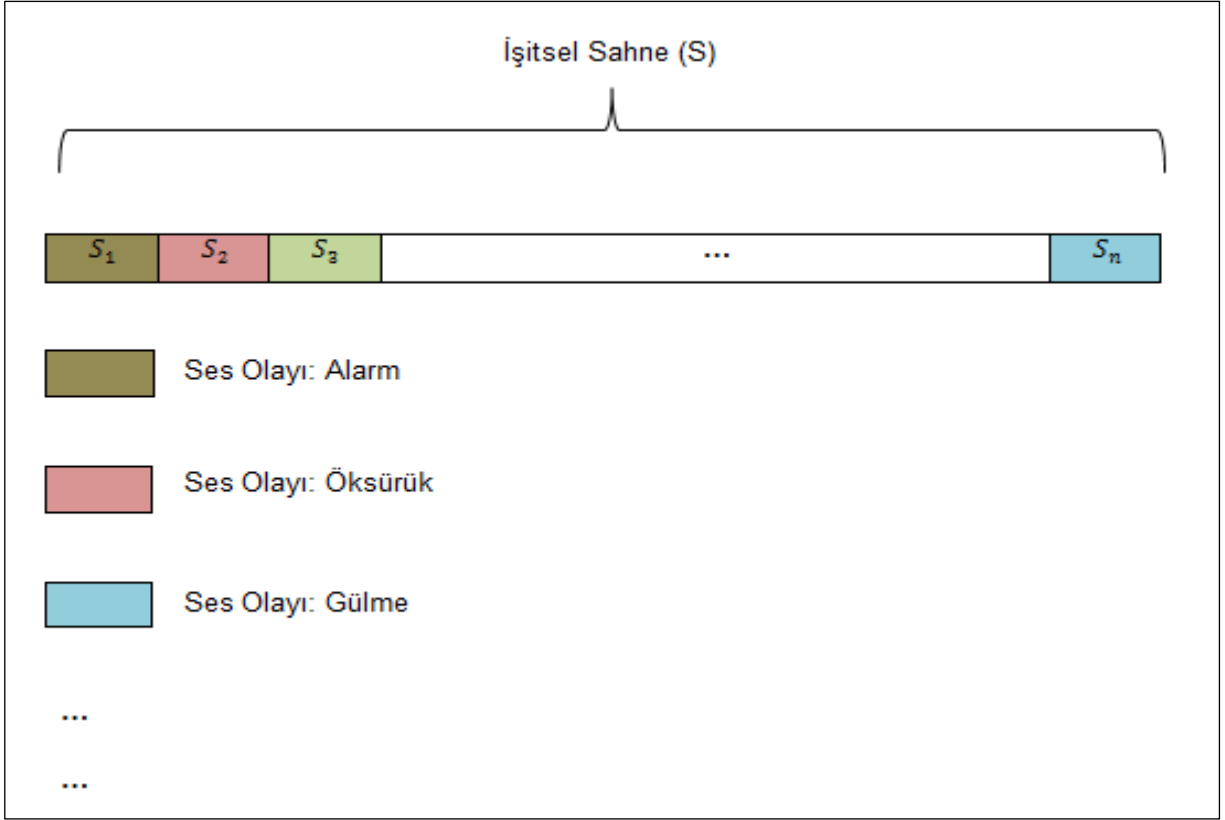
olaylarının sezimlenmesi o ses dosyasının ortam bilgisi ile ilgili ipuçları vermektedir. Örneğin, içerisinde kalabalık insan sesi, çatal kaşık sesleri sezimlenen bir ortamın restoran sahnesi olma olasılığı yüksektir. Bu bağlamda, ses olaylarının sezimi için kullanılan veri kümesi, ofis ortamı içerisinden alınmış 16 ayrı ses olayını barındırmaktadır (*alarm, boğaz temizleme, öksürük, kapı çarpması, çekmece, klavye, anahtar, kapı vurma, gülme, bilgisayar faresi, sayfa çevirme, nesnenin masaya çarpması, telefon, yazıcı, konuşma, elektrik düğmesi*). Aynı zamanda, veri kümesindeki işitsel sahneler arasında *ofis* ortamı bulunmaktadır. Bu iki durum göz önüne alınarak, ofis işitsel sahnesi için, ofis ortamından alınmış 16 ayrı ses olayı ile ofis işitsel sahnesi arasındaki olası ilintilerin incelenmesi ve sonuçlarının değerlendirilmesi amaçlanmıştır.

5.5.1 Yöntem

Ses olaylarının işitsel sahnelere olan katkıları (ilintileri) incelenirken, veri kümesindeki işitsel sahne örneklerinin her biri için klip bazlı karar verme yöntemi yerine, çerçeve bazlı karar verme yönteminin daha uygun olacağı değerlendirilmiştir. Buna göre, incelenecek her bir ses klipi, belli bir uzunluktaki çerçevelere bölünmüş ve her çerçeve için, sistemin o çerçevenin ait olduğu bir ses olayını etiketlemesi sağlanmıştır. Girdi olarak verilen bir işitsel sahnenin, seçilecek çerçeve uzunluğuna göre, birbirleriyle örtüşmeyen zamansal bölütlerinin bir atama işlemi ile değerlendirilmesi sağlanmaktadır. Böylece, bir işitsel sahne içerisinde geçen ses olaylarının bilinmesi ile, ilgili sahne hakkında yorum yapılabilmesi mümkün olacaktır. Çalışmada, çerçeve uzunlukları farklı analizler gerçekleştirmek amacıyla sırasıyla 30ms, 90ms ve 990ms olarak belirlenmiştir. Tasarlanan yapının şeması Şekil 5.3'de gösterilmiştir. Bu analizler gerçekleştirilirken, öznitelik vektörü olarak MFCC ve sınıflandırıcı yöntemi olarak DVM kullanılmıştır.

Buna göre, MFCC öznitelik çıkarım aşamasında verilen pencere süresi, girdi olarak verilen ses klipi için zamansal bir bölüt olmaktadır ve öznitelik vektörlerinin her bir satırı ses dosyasının verilen pencere süresi boyutunda bir analizdir. Daha önceki çalışmalarda, en uygun değer olarak pencere süresi 30ms seçildiğinden, MFCC öznitelik vektöründeki her bir satır 30ms çerçeve boyutundaki analizini göstermektedir. Katkıların inceleneceği senaryolarda, ilk aşama olarak 30ms çerçeve boyutundaki analizler için, girdi olarak verilen ses klibinden çıkartılan

MFCC vektörünün her bir satırına, bir ses olayı sezimi işlemi gerçekleştirilmektedir. 30ms uzunluğundaki çerçeveler için gerçekleştirilen analizler sonucu elde edilen hata matrisi C.1’de verilmiştir.



Şekil 5. 3 Ses olaylarının işitsel sahnelere olan katkısı için tasarlanan yapı

Sonraki aşamada, önce 90ms boyutlarındaki çerçevelere ve son olarak 990ms boyutlarındaki çerçevelere ses olayı etiketlemesi yapılarak analizler tamamlanmıştır. Böylece seçilen çerçeve boyutunun ses olaylarının işitsel sahnelere katkısındaki önemi incelenmiştir.

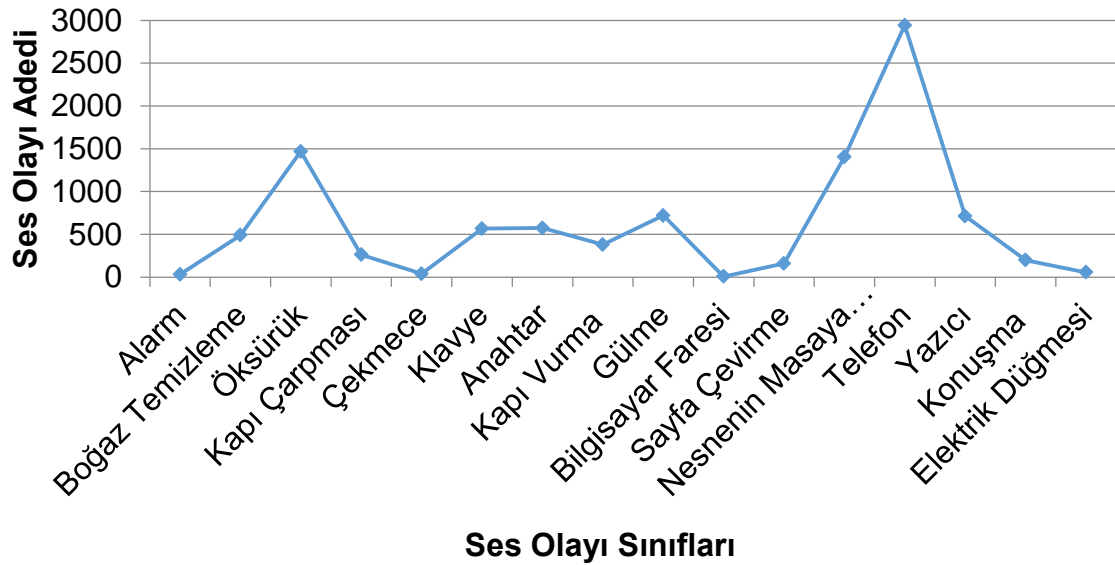
$$F = \begin{bmatrix} F_{1,1} & \dots & F_{1,20} \\ \vdots & \ddots & \vdots \\ F_{3000,1} & \dots & F_{3000,20} \end{bmatrix} \quad (5.1)$$

$$F' = \begin{bmatrix} \frac{1}{N} \sum_{i=1}^N F_{i,1} & \dots & \dots \\ \vdots & \ddots & \vdots \\ \dots & \dots & \dots \end{bmatrix} \quad (5.2)$$

90ms ve 990ms analizlerin gerçekleştirilmesi için, çerçeve boyutlarının bu şekilde ayarlanması gereklidir. MFCC öznitelik çıkarım aşamasında 30mslik pencereler ile analiz gerçekleştirildiğinden, vektörlerin her bir satırı 30ms temsil etmektedir. Bu nedenle, örneğin 90ms'lik çerçeve analizi için, vektörün her bir 3 satırı bir örnek gibi düşünülüp analiz edilmesi gerekmektedir. Bunun için (5.1)'de verilen öznitelik matrisi, (5.2)' de verilen şekliyle veri kümesine uygulanmış ve yeni öznitelik vektörlerine dönüşüm yapılmıştır. (5.2)'de verilen denklemde N değeri, $t_w/30ms$ ile hesaplanmaktadır. Örneğin, 90 ms'lik bir etiketleme için, her bir 3 satır alınmalıdır. Her iki duruma ait etiketlenen ses olaylarını gösteren bilgiler C.2 ve Tablo C.3'de sunulmuştur.

5.5.2 Deneysel sonuçlar

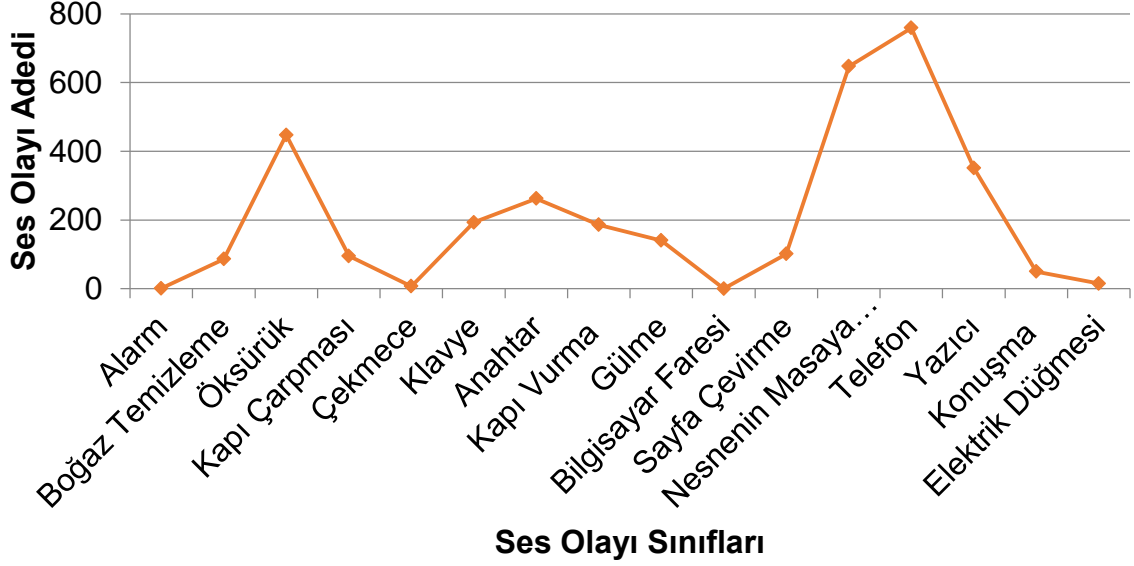
Ses olaylarının işitsel sahnelere olan katkısını bulmak için yapılan analizlerin sonuçları bu bölümde sunulmuştur.



Şekil 5. 4 Ofis işitsel sahnesi için 30ms çerçeve boyutu ile gerçekleştirilen analizler

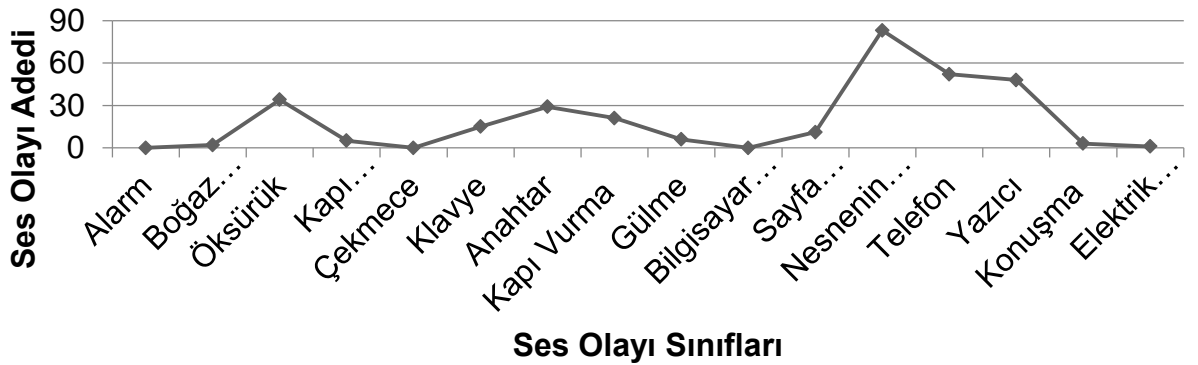
Şekil 5.4 de verilen sonuçlara göre, 30ms çerçeve boyutu kullanılarak analiz edilen ofis işitsel sahnesine ait 10 klip için en fazla katkının telefon ses olayına ait olduğu görülmektedir. Bunun yanında *öksürük*, *kapı vurma*, *gülme* gibi ses olaylarının da

katkıları vardır. En az katkıyı veren ses olayları ise; alarm, çekmece, bilgisayar faresi ve elektrik düğmesi olarak kaydedilmiştir.



Şekil 5. 5 Ofis işitsel sahnesi için 90ms çerçeve boyutu ile gerçekleştirilen analizler

Şekil 5.5'de verilen sonuçlar Şekil 5.4'de verilen sonuçlarla benzerlik göstermektedir. Burada, 90ms çerçeve boyutu seçildiği için, bir ses klipi içerisinde 30ms çerçeve analizine göre daha az sayıda çerçeve seçildiğinden, tek farkın ses olayları adetlerinin her sınıf için azalması olarak söylenebilir.



Şekil 5. 6 Ofis işitsel sahnesi için 990ms çerçeve boyutu ile gerçekleştirilen analizler

Şekil 5.6'da sonuçlar 990ms çerçeve boyutu kullanılarak yapılan analizlerin sonuçlarını göstermektedir. Buna göre, en yüksek katkıyı *nesnenin masaya çarpması* ses olayı sınıfının yaptığı görülmektedir. Diğer ses olaylarının katkıları, 30ms ve 90ms çerçeve boyutları kullanılarak yapılan analizlerle benzer oranlardadır. Gerçekleştirilen analiz sonucunda, bir işitsel sahneyi oluşturan ses olayları bilinirse, o işitsel sahnenin tanınmasının mümkün olabileceği değerlendirilmektedir.

6. SONUÇLAR VE TARTIŞMA

Bu tez çalışmasında, ses kliplerinin içerisinde bulunan ses olaylarının sezimlenmesi, işitsel sahnelerin tanınması ve ses olaylarının işitsel sahnelere (olası) katkılarının anlaşılması amacıyla yöntemler geliştirilmiş ve sonuçları değerlendirilmiştir. Bu kapsamda, öznelik çıkarım yöntemleri içerisinde, konuşma tanıma uygulamalarındaki başarımından dolayı MFCC özneliği kullanılmıştır. Tanıma başarımlarını artırabilmek amacıyla, farklı pencere ve atlama süreleri için analizler gerçekleştirilmiş ve sınıflandırıcı eniyilemesi yapılmıştır. Ses olaylarının farklı spektrumlarını karakterize edebilmek amacıyla, uyarlanabilir bir MFCC öznelik çıkarım yöntemi önerilmiştir. Örüntü tanıma uygulamalarında başarılı bir performans gösteren DVM sınıflandırma algoritması veri kümesine uygun parametre değerleri ayarlanarak çalışmalar tamamlanmıştır.

Çalışmanın ses olayı sezimi aşamasında, 16 (on altı) ayrık ses olayı (*alarm, boğaz temizleme, öksürük, kapı çarpması, çekmece, klavye, anahtar, kapı vurma, gülme, bilgisayar faresi, sayfa çevirme, nesnenin masaya çarpması, telefon, yazıcı, konuşma ve elektrik düğmesi*) için analizler yapılmış ve ses olaylarının sınıflandırılması sağlanmıştır. Kullanılan veri kümesi erişime açık bir paylaşım yeri bulunan *IEEE AASP Challenge Detection and Classification of Acoustic Scenes and Events* isimli yarışmanın alt kategorilerinden biri olan *Office Live Environment* isimli veri kümesinden alınmıştır [21]. Deneysel çalışmalarda MFCC öznelik çıkarım yöntemi, en uygun pencere ve atlama süresi ve veri kümesinde en iyi sonuç veren katsayı seçilerek uygulanmıştır. MFCC öznelik çıkarım algoritması standart yöntemlerle uygulandığında, farklı seslerin farklı frekans spektrumuna sahip olabilmesi nedeniyle, tanıma performansı düşmektir. Bu nedenle, öncelikle her bir ses dosyası için ses sinyalindeki baskın frekans değerleri bulunmuş, MFCC öznelik vektörleri çıkarılırken bu frekans değerleri kullanılarak, uyarlanabilir bir çıkarım süreci gerçekleştirilmiştir. Sınıflandırıcı eniyilemesi aşamasında, Grid arama algoritması uygulanarak veri kümesi için en uygun parametre değerleri ayarlanmıştır. Sınıflandırıcı tasarımı aşamasında yapılan parametre eniyilemesinin, kullanılan öğrenme algoritmasının başarımını önemli ölçüde artırdığı değerlendirilmiştir. Bu çalışmada önerilen yöntemler kullanıldığında, ses olayı sezimleme başarımının F-ölçütü %72 iken, standart yöntemlerle (varsayılan MFCC ve DVM parametreleri ile her ses klipinin sabit frekans aralığında analiz

edilmesi) bu deęer %48 olmaktadır. Elde edilen %24 oranındaki artış, sistemin performansı açısından önemli bir kazanım olarak yorumlanmaktadır. Aynı veri kümesini kullanan dięer alıřmalarla [13-19] kıyaslandığında, sistemin tanıma başarımının literatürü %27 oranında artırdığı söylenebilir.

Sonuçlarda dikkat çeken bir husus, alıřmamızda önerilen dayanak (baseline) sistemin sınıflandırma başarımının (%48), literatürde aynı veri kümesi için elde edilen en yüksek başarımdan (%45,17) daha yüksek olmasıdır. Bu başarımlar dikkat çekicidir ve kullanılan ses öznelik ve sınıflandırıcı kombinasyonu ile açıklanabilir. Literatürde bu kombinasyonu kullanan ve bilginiz dahilindeki tek alıřmada [19] elde edilen başarımlar %19.2 olarak rapor edilmektedir. Detaylı incelendiğinde, öznelik ıkarım ve DVM yöntemlerinin standart olarak uygulandığı ve sınıflandırıcı ekirdek fonksiyonu olarak “doęrusal ekirdek” kullanıldığı anlaşılmaktadır. Doęrusal ekirdek fonksiyonun ok-sınıflı sınıflandırma problemlerinde etkin sonuçlar vermediği bilinmektedir [47]. Bu durum, tez alıřmamızda incelenmiş ve testleri yapılan doęrusal, polinomsal ve radyal tabanlı ekirdek fonksiyonları arasından en yüksek başarımları (%28,63, %33,29, %22,87) radyal tabanlı ekirdek fonksiyonu vermiştir. Bu sonuç, DVM eğitimi esnasında kullanılan ekirdek fonksiyonun başarımlara önemli katkısının olduğu görüşünü desteklemektedir. Sonuç olarak, alıřmamızda kullanılan MFCC ve DVM kombinasyonunun, IEEE AASP veri kümesi üzerinde etkin sonuçlar verdiği ıkarımı yapılabilir.

Bu kombinasyona ek olarak, alıřmamızda önerilen analiz yöntemleri (uyarlamalı öznelik ıkarım yöntemi, farklı pencere analizleri ve sınıflandırıcı parametre eniyileme), %48’lik dayanak sistem başarımlarımızı %72 F-ölçütü deęerine ulařmıştır. Bu başarımların büyük oranda, özgül olarak önerdiğimiz uyarlamalı öznelik ıkarım yöntemi ile sağlandığı deęerlendirilmektedir. Bu yöntemde, analiz edilen tüm ses sınıfları için, bu ses sınıflarındaki her bir ses klibinin frekans karakteristiği MFCC öznelik ıkarımına dahil edilmiştir. Temel amacımız, ses sınıflarının frekans spektrumundaki en baskın (genliđi en yüksek) frekans bileřenin tespit edilmesi ve bu frekans deęerinin MFCC öznelik ıkarım işleminde, alt frekans sınırı olarak kullanılmasıdır. Somut bir örnek vermek gerekirse; bir ses klibinin spektrumunun [10Hz, 100Hz] aralığında olduğunu varsayalım. Bu aralıktaki en baskın (genliđi en yüksek) frekans bileřeninin 30Hz, ikinci en baskın frekans

bileşenin 60Hz, üçüncü en baskın frekans bileşenin 20Hz vb. olduğu varsayılırsa, biz spektrumdaki genliği en yüksek frekans bileşenini (30Hz) seçiyoruz ve MFCC katsayılarını $[30\text{Hz}, f_{\text{Nyquist}}]$ aralığından elde ediyoruz. Başka bir yaklaşım, ikinci en baskın frekansı (60Hz) seçebilir ki, bu frekans bileşeni spektrumda en baskın frekansın (30Hz) sağında kalır; üçüncü en baskın frekans bileşeni (20Hz) seçilirse, bu durumda seçilen frekans bileşeni en baskın frekansın (30Hz) solunda kalır. İkinci ve üçüncü senaryolarda MFCC öznitelikleri sırasıyla $[60\text{Hz}, f_{\text{Nyquist}}]$ ve $[20\text{Hz}, f_{\text{Nyquist}}]$ aralıklarından çıkarılmış olur. Bu yaklaşımların başarımı şu şekilde etkileyeceği düşünülmektedir: Eğer, ikinci, üçüncü, vb. baskın frekanslar çoğunlukla en baskın frekansın solundaki frekanslardan oluşuyorsa, en baskın frekansın solundaki frekansları analizlere dahil etmek, sağındaki frekanslardan oluşuyorsa da sağında frekansları analizlere dahil etmek başarımı olumlu yönde etkileyecektir; aksi durumlarda anlamlı bir katkısı öngörülmemektedir. Bununla birlikte, diğer yöntemlerin analizlere getireceği karmaşıklık açıktır.

Dayanak sistemin zaman karmaşıklığını öznitelik çıkarım süresi ($t_{\text{feature_extraction}}$), öznitelik boyut indirgeme süresi ($t_{\text{dimension_reduction}}$) ve doğrulama süresi ($t_{\text{validation}}$) belirlemektedir. Sistem testleri, 32-bit Windows 7 platformunda Intel Core2Duo P8400 2,26Ghz ve 4 GB belleğe sahip bir kişisel bilgisayar üzerinde gerçekleştirilmiştir. Buna göre, verikümesi üzerindeki $t_{\text{feature_extraction}}=409,946839$ saniye, $t_{\text{dimension_reduction}}=319,0181$ saniye ve $t_{\text{validation}}=17.643607$ saniye olarak ölçülmüştür. Çalışmada önerilen uyarlamalı öznitelik çıkarım yöntemi, her bir ses klibi için Fourier dönüşümü gerektirmektedir. Bu dönüşüm için geçen süre t_{Fourier} ve verikümesindeki örnek sayısı n ile temsil edilirse; ek maliyet $C = n \times t_{\text{Fourier}}$, sistemin eğitim maliyetine sabit olarak etki etmektedir. Test aşamasında bu etki $C = 1 \times t_{\text{Fourier}}$ olarak tanımlıdır.

İşitsel sahnelerin tanınması aşamasında, *IEEE AASP Challenge Detection and Classification of Acoustic Scenes and Events* isimli yarışmanın alt kategorilerinden biri olan *Scene Classification* için oluşturulan ses dosyaları veri kümesi olarak kullanılmıştır [24]. Bu veri kümesindeki 10 ayrı işitsel sahne için (*otobüs, kalabalık sokak, ofis ortamı, açık market, park, sessiz sokak, restoran, süpermarket, tren ve tren istasyonu*) ses olay sezimi sırasında yapılan analizlerden elde edilen en iyi değerler kullanılarak öznitelik çıkarım işlemi gerçekleştirilmiş ve sınıflandırılmıştır.

MFCC öznitelik çıkarım aşamasında, bu sınıftaki ses dosyalarının her biri için, tıpkı ses olay sezimi aşamasında yapıldığı gibi frekans analizi gerçekleştirilmiş ve özel frekans aralıkları ile öznitelik vektörleri oluşturulmuştur. DVM sınıflandırıcısı Grid arama algoritması ile veri kümesine en uygun olan parametrelerle yeniden uyarlanmıştır ve model oluşumunda eniyileştirilmiş DVM kullanılmıştır. Yapılan testler sonucu, F-ölçütü için sistemin %59'luk bir performans sergilediği görülmüştür. Bu oran, sistem varsayılan MFCC ve DVM parametreleri ve sabit frekans aralığı kullanılarak çıkarılan öznitelik vektörleri ile test edildiğinde %42 olarak kaydedilmektedir. Veri kümesi üzerindeki deneyler, önerilen sistemin, varsayılan parametre değerleri ile geliştirilen sistemlere kıyasla, daha yüksek başarımlar elde ettiğini göstermektedir.

Tez çalışmasının hedeflerinden birisi de, ses olaylarının işitsel sahnelere olan katkılarının incelenmesidir. Bu amaçla, işitsel sahnelerin içerdikleri ses olaylarına göre, çerçeve tabanlı bir ses olayı etiketleme yöntemi geliştirilmiştir. Ses olayı sezimi aşamasında kullanılan veri kümesi günlük hayatta ofis ortamından alınan ses örneklerinden oluşmaktadır. Bu nedenle, ses olayları ile işitsel sahne ilintilerinin incelenmesi için ofis sahnesi tercih edilmiştir. Bunun için, verilen bir işitsel sahne, birbirleriyle örtüşmeyen zamansal bölütlere ayrılmış ve bu bölütlerin her biri için ses-olay etiketlemesi yapılmıştır. Zamansal bölüt süreleri, kullanılan veri kümesini de baz alarak 30ms, 90ms ve 990ms olacak şekilde tanımlanmıştır. Böylece, farklı öge boyutlarının analizlere etkisi değerlendirilmiştir. Elde edilen sonuçlara göre, 30ms pencere süresi ile analiz yapıldığında, 90 ve 990mslik pencere boyutuna kıyasla daha hassas atamalar gerçekleştirmiştir. Deneysel çalışmalara göre, ses olaylarının, işitsel sahnelerle farklı oranlarda ilintilerinin olduğu ve bu ilintilerin kullanılması ile işitsel sahne tanıma uygulamalarının geliştirilebileceği değerlendirilmektedir.

Gelecek çalışmalar arasında, ses olayları ile işitsel sahneler arasındaki ilintilerin kural tabanlı bir yaklaşımla tespit edilmesi bulunmaktadır. Buna göre, işitsel sahneler içerisinde var olan ses olayları için bir oylama işlemi yaparak, işitsel sahnenin hangi olay ile ilgili olabileceği konusunda karar verilebilir. Bununla birlikte, hangi ses olayının hangi işitsel sahneyi ne kadar etkilediği ile ilgili istatistiksel bilgilerin analizi yapılabilir.

KAYNAKLAR LİSTESİ

- [1] Eronen A. J., Peltonen V. T., Tuomi J.T., Klapuri A.P., Fagerlund S., Sorsa T., Lorho G. and Huopaniemi J., "Audio-based context recognition," IEEE Transactions on Audio, Speech, and Language Processing, vol.14, no.1, s.321–329, 2006.
- [2] Ma L., Milner B. and Smith D., "Acoustic environment classification," ACM Trans. Speech Lang. Process, vol.3, no.2, s.1–22, 2006.
- [3] Okuyucu C., Sert M. ve Yazıcı A., "Audio Feature and Classifier Analysis for Efficient Recognition of Environmental Sounds," IEEE International Symposium on Multimedia (ISM), s.125–132, 2013.
- [4] Muhammad G., Alotaibi Y. A., Alsulaiman M. and Huda M.N., "Environment Recognition Using Selected MPEG-7 Audio Features and Mel-Frequency Cepstral Coefficients," Fifth International Conference on Digital Telecommunications (ICDT), s.11–16, 2010.
- [5] Su F., Yang L., Lu T. ve Wang G., "Environmental sound classification for scene recognition using local discriminant bases and HMM," In Proceedings of the 19th ACM international conference on Multimedia (MM '11), s.1389–1392, 2011.
- [6] Wang J.-C., Wang J.-F., He K. W., and Hsu C.-S., "Environmental sound classification using hybrid svm/knn classifier and mpeg-7 audio low-level descriptor," in Int. Joint Conf. on Neural Networks(IJCNN'06), s. 1731–1735, 2006.
- [7] Chu S., Narayanan S., and Kuo C.C.J., "Environmental sound recognition with time-frequency audio features," In IEEE Transactions on Audio, Speech and Language Processing, vol.17, no.6, s.1142-1158, 2009.
- [8] Lee K., Ellis D. P. W.,and Loui A. C., "Detecting local semantic concepts in environmental sounds using Markov model based clustering," IEEE International Conference on Acoustic, Speech and Signal Processing, s. 2278-2281, 2010.
- [9] Beritelli F.and Grasso R., "A pattern recognition system for environmental sound classification based on MFCCs and neural networks," IEEE 2nd International Conference on Signal Processing and Communication Systems(ICSPCS 2008),s.1-4, 2008.
- [10] Feki I., Ben Ammar I. and Alimi A.M., "Audio stream analysis for environmental sound classification," IEEE International Conference on Multimedia Computing and Systems(ICMCS'11), s. 1-6, 2011
- [11] Güvensan M.A. and Taysi Z.C., "Environmental sound classification for recognition of house appliances," IEEE 18th Signal Processing and Communications Applications Conference (SIU 2010), s. 431-434, 2010.
- [12] Shin S.H., Hashimoto T. and Hatano S., "Automatic detection system for cough sounds as a symptom of abnormal health condition," IEEE

- Transactions on Information Technology in Biomedicine, vol.13, no.4, s.486-493, 2009.
- [13] Schröder J., Moritz N., Schadler M. R., Cauchi B., Adiloglu K., Anemuller J., Doclo S., Kollmeier B. and Goetze S., "Acoustic event detection using signal enhancement and spectro-temporal feature extraction," IEEE AASP Challenge on Detection and Classification of Acoustic Scenes and Events, s.1-3, 2013.
 - [14] Niessen M. E., Van Kasteren T. L. M. and Merentitis A., "Hierarchical sound event detection" IEEE AASP Challenge on Detection and Classification of Acoustic Scenes and Events, s.1-3, 2013.
 - [15] Vuegen L., Broeck B. V. D., Karsmakers P., Gemmeke J.F., Vanrumste B. and Hamme H.V., "An MFCC-GMM approach for event detection and classification," IEEE Workshop on Applications of Signal Processing to Audio and Acoustics (WASPAA), s.1–3, 2013.
 - [16] Chauhan S., Phadke S. and Sherland C., "Event detection and classification," IEEE Workshop on Applications of Signal Processing to Audio and Acoustics (WASPAA), s.1–3, 2013.
 - [17] Diment A., Heittola T. and Virtanen T., "Sound event detection for Office live and office synthetic aasp challenge," IEEE AASP Challenge on Detection and Classification of Acoustic Scenes and Events, s.1-3, 2013.
 - [18] Gemmeke J. F., Vuegen L., Vanrumste B. and Hamme H. V., "An exemplar-based NMF approach for audio event detection," IEEE AASP Challenge on Detection and Classification of Acoustic Scenes and Events, s.1-3, 2013.
 - [19] Nogueira W., Roma G. and Herrera P., "Automatic event classification using front end single channel noise reduction, MFCC features and a support vector machine classifier," IEEE AASP Challenge on Detection and Classification of Acoustic Scenes and Events, s.1-2, 2013.
 - [20] Vapnik V, The support vector method of function estimation. Nonlinear Modeling: Advanced Black-Box Techniques, s.55:86, 1998.
 - [21] Giannoulis D., Benetos E., Stowell D. and Plumbley M. D., "IEEE AASP Challenge on Detection and Classification of Acoustic Scenes and Events - Development Dataset for Event Detection Task, subtask 1- OL," IEEE Workshop on Applications of Signal Processing to Audio and Acoustics (WASPAA), s.1–4, 2013.
 - [22] Chen L., Gunduz S. and Ozsu M. T., "Mixed type audio classification with support vector machine," IEEE International Conference on Multimedia and Expo (ICME), s.781–784, 2006.
 - [23] Chang C. C. and Lin C.J., "LIBSVM: A library for support vector machines," ACM Transactions on Intelligent Systems and Technology, vol.2, no.3, s.1-27, 2011.
 - [24] Giannoulis D., Benetos E., Stowell D., and Plumbley M. D., "IEEE AASP Challenge on Detection and Classification of Acoustic Scenes and Events - Public Dataset for Scene Classification Task", Queen Mary University of London, 2012.

- [25] Smola A. J. and Schölkopf B., "A tutorial on support vector regression", NeuroCOLT2 Technical Report Series, NC-TR-1998-030, October, 1998.
- [26] Sundaram H. and Chang S.F., "Audio scene segmentation using multiple features, models and time scales," IEEE International Conference on Acoustic, Speech and Signal Processing (ICASSP'00), s. 2441-2444, 2000.
- [27] Oh S., McCloskey S., Kim I., Vahdat A., Cannons K. J., Hajimirsadeghi H., Mori G., Perera A.G.A, Pandey M. and Corso J. J., "Multimedia event detection with multimodal feature fusion and temporal concept localization," Machine Vision and Applications, vol.25, no.1, s.49-69, 2014.
- [28] Lan Z.-Z., Bao L., Yu S.-I., Liu W. and Hauptmann A. G., "Multimedia classification and event detection using double fusion," Multimedia Tools and Applications, vol.71, no.1, s.333-347, 2014.
- [29] Yang Y., Nie F., Xu D., Luo J., Zhuang Y. and Yunhe P., "A Multimedia Retrieval Framework Based on Semi-Supervised Ranking and Relevance Feedback," IEEE Transactions on Pattern Analysis and Machine Intelligence, vol.34, no.4, s.723-742, 2012.
- [30] Zha Z.-J., Wang M., Zheng Y.-T., Yang Y., Hong R. and Chua T.-S., "Interactive Video Indexing With Statistical Active Learning," IEEE Transactions on Multimedia, vol.14, no.1, s.17-27, 2012.
- [31] Doğan E., Sert M., and Yazıcı A., "Content-based classification and segmentation of mixed-type audio by using mpeg-7 features," In Proceedings of IEEE International Conference on Advances in Multimedia (MMEDIA'09), s.152-157, 2009.
- [32] Cai R., Lu L., Zhang H.J., and Cai L.H., "Highlight sound effects detection in audio stream," IEEE International Conference on Multimedia and Expo (ICME), s.III-37, 2003.
- [33] Lu G.I, "Multimedia database management systems," Artech House, 1999.
- [34] Mertin A., "Signal Analysis: wavelets, filter banks, time-frequency transforms and applications," Wiley & Sons Ltd., 1999.
- [35] Giannoulis D. and et al., "Detection and classification of acoustic scenes and events: an IEEE AASP challenge," In IEEE Workshop on Applications of Signal Processing to Audio and Acoustic, s.1-4, 2013
- [36] Mitchell T., "Machine Learning", McGraw hill, 1997.
- [37] Snoek C. G. M. and Worring M., "Multimodal video indexing: a review of the state-of-the-art", Multimedia Tools and Applications, vol.25, no.1, s.5-25, 2005.
- [38] Zhang T. and Kuo C. -C., "Hierarchical classification of audio data for archiving and retrieving," In International Conference on Acoustic, Speech and Signal Processing (ICASSP), s.3001-3004, 1999.
- [39] Kim H. -G., and et al., "Audio classification based on mpeg-7 spectral basis representations," IEEE Tran. on Cir. And Syst. For Video Tech., vol.14, no.5, s.716-725, 2004.

- [40] Lu L., Li S. Z., and Zhang H. –J., “Content-based audio segmentation using support vector machines,” In IEEE International Conference on Multimedia and Expo, s. 749-752, 2001.
- [41] Kiranyaz S. and et al., “A generic audio classification and segmentation approach for multimedia indexing and retrieval,” In IEEE Transaction on Audio, Speech, and Language Processing, vol.14, no.3, s.1062-1081, 2006.
- [42] Meinedo H. and Neto J., “Audio segmentation, classification and clustering in broadcast news task,” In IEEE International Conference on Acoustic, Speech and Signal Processing (ICASSP), s. 5-8, 2003.
- [43] Doğan E., Sert M. and Yazıcı A., “A flexible and scalable audio information retrieval system for mixed-type audio signals,” In International Journal of Intelligent Systems (Wiley), vol.26, s.952-970, 2011.
- [44] Gelegin I. ve Bolat B., “Ayrık kelime tabanlı bir konuşma tanıma sistemiyle bilgisayar kontrolü,” TMMOB Elektrik Mühendisleri Odası Elektrik-Elektronik ve Bilgisayar Sempozyumu, 2011.
- [45] Ellis D., “Reproducing the feature outputs of common programs using Matlab and melfcc.m”, <http://labrosa.ee.columbia.edu/matlab/rastamat/mfccs.html>, 2005, Son Erişim Tarihi: 28.01.2015.
- [46] Burg J., “The science of digital media,” Prentice-Hall, Inc., 2007.
- [47] Mainkandan J. and Venkataramini B., “Evolution of multiclass support vector machine classifiers using optimum threshold-based pruning technique,” Singal Processing, IET, vol.5, no.5, s. 506-513, 2011.

EKLER LİSTESİ

EK A:	SES OLAYLARI TESPİT ETME SONUÇLARININ HATA MATRİSLERİ.....	53
Çizelge A. 1	13 Katsayılı MFCC öznitelik vektörü kullanılarak DVM ile eğitilensistemin hata matrisi	53
Çizelge A. 2	20 Katsayılı MFCC öznitelik vektörü kullanılarak DVM ile eğitilen sistemin hata matrisi	54
Çizelge A. 3	48 Katsayılı MFCC öznitelik vektörü kullanılarak DVM ile eğitilen sistemin hata matrisi	55
Çizelge A. 4	128 Katsayılı MFCC öznitelik vektörü kullanılarak DVM ile eğitilen sistemin hata matrisi	56
Çizelge A. 5	Pencere süresi 10ms, atlama süresi 5 ms alınarak 20 katsayılı MFCC öznitelik vektörü kullanılarak DVM ile eğitilen sistemin hata matrisi	57
Çizelge A. 6	Pencere süresi 10ms, atlama süresi 10 ms alınarak 20 katsayılı MFCC öznitelik vektörü kullanılarak DVM ile eğitilen sistemin hata matrisi	58
Çizelge A. 7	Pencere süresi 20ms, atlama süresi 10 ms alınarak 20 katsayılı MFCC öznitelik vektörü kullanılarak DVM ile eğitilen sistemin hata matrisi	59
Çizelge A. 8	Pencere süresi 20ms, atlama süresi 20ms alınarak 20 katsayılı MFCC öznitelik vektörü kullanılarak DVM ile eğitilen sistemin hata matrisi	60
Çizelge A. 9	Pencere süresi 30ms, atlama süresi 10ms alınarak 20 katsayılı MFCC öznitelik vektörü kullanılarak DVM ile eğitilen sistemin hata matrisi	61
Çizelge A. 10	Pencere süresi 30ms, atlama süresi 15ms alınarak 20 katsayılı MFCC öznitelik vektörü kullanılarak DVM ile eğitilen sistemin hata matrisi	62
Çizelge A. 11	Pencere süresi 30ms, atlama süresi 30ms alınarak 20 katsayılı MFCC öznitelik vektörü kullanılarak DVM ile eğitilen sistemin hata matrisi	63
Çizelge A. 12	Pencere süresi 50ms, atlama süresi 10ms alınarak 20 katsayılı MFCC öznitelik vektörü kullanılarak DVM ile eğitilen sistemin hata matrisi	64
Çizelge A. 13	Pencere süresi 50ms, atlama süresi 25ms alınarak 20 katsayılı MFCC öznitelik vektörü kullanılarak DVM ile eğitilen sistemin hata matrisi	65
Çizelge A. 14	Pencere süresi 50ms, atlama süresi 50ms alınarak 20 katsayılı MFCC öznitelik vektörü kullanılarak DVM ile eğitilen sistemin hata matrisi	66

Çizelge A. 15	Pencere süresi 100ms, atlama süresi 10ms alınarak 20 katsayılı MFCC öznitelik vektörü kullanılarak DVM ile eğitilen sistemin hata matrisi	67
Çizelge A. 16	Pencere süresi 100ms, atlama süresi 50ms alınarak 20 katsayılı MFCC öznitelik vektörü kullanılarak DVM ile eğitilen sistemin hata matrisi	68
Çizelge A. 17	Pencere süresi 100ms, atlama süresi 100ms alınarak 20 katsayılı MFCC öznitelik vektörü kullanılarak DVM ile eğitilen sistemin hata matrisi	69
Çizelge A. 18	MFCC öznitelik çıkarımında varsayılan alt ve üst frekans değerleri kullanılarak DVM ile eğitilen sistemin hata matrisi	70
Çizelge A. 19	MFCC özniteliği, baskın frekans değeri kullanılarak çıkarıldığında DVM ile eğitilen sistemin hata matrisi	71

EK B: İŞİTSEL SAHNE TANIMA SONUÇLARININ HATA MATRİSLERİ.....72

Çizelge B. 1	Varsayılan DVM parametreleri kullanılarak DVM ile eğitilen model ile test edilen sistemin hata matrisi	72
Çizelge B. 2	En uygun değeri ayarlanmış MFCC öznitelik vektörleri ve en iyilenmiş DVM parametreleri kullanılarak DVM ile eğitilen model ile test edilen sistemin hata matrisi	73
Çizelge B. 3	Önerilen sistemin hata matrisi	74

EK C: SES OLAYLARININ İŞİTSEL SAHNELERE OLAN KATKILARI HATA MATRİSLERİ.....75

Çizelge C. 1	Ses dosyası içerisindeki 30ms lik çerçeveleri ses olayları ile etiketlendiğinde, her dosyanın içerisindeki ses olaylarının sayıları	75
Çizelge C. 2	Ses dosyası içerisindeki 90ms lik çerçeveleri ses olayları ile etiketlendiğinde, her dosyanın içerisindeki ses olaylarının sayıları	76
Çizelge C. 3	Ses dosyası içerisindeki 990ms lik çerçeveleri ses olayları ile etiketlendiğinde, her dosyanın içerisindeki ses olaylarının sayıları	77

EK A: SES OLAYLARI TESPİT ETME SONUÇLARININ HATA MATRİSLERİ

Çizelge A. 1 13 Katsayılı MFCC öznelik vektörü kullanılarak DVM ile eğitilen sistemin hata matrisi

	Alarm	Boğaz Temizleme	Öksürük	Kapı Çarpması	Çekmece	Klavye	Anahtar	Kapı Vurma	Gülme	Bilgisayar Faresi	Sayfa Çevirme	Nesnenin Masaya Çarpması	Telefon	Yazıcı	Konuşma	Elektrik Düğmesi
Alarm	15	0	0	0	0	0	0	0	1	0	0	0	2	0	1	1
Boğaz Temizleme	0	18	0	0	0	0	0	0	0	1	0	0	0	0	1	0
Öksürük	2	1	5	0	3	2	1	0	3	1	2	0	0	0	0	0
Kapı Çarpması	0	0	0	10	3	0	0	1	0	1	0	1	0	4	0	0
Çekmece	0	0	0	5	12	1	0	0	0	0	1	0	0	0	1	0
Klavye	0	0	0	0	1	10	3	0	0	1	3	0	0	0	0	2
Anahtar	0	0	0	1	0	1	13	0	0	0	3	1	0	0	0	1
Kapı Vurma	0	0	0	0	0	0	0	18	0	2	0	0	0	0	0	0
Gülme	2	0	4	0	0	0	0	0	10	0	0	0	1	1	2	0
Bilgisayar Faresi	1	2	0	0	1	0	1	1	0	12	0	0	0	0	2	0
Sayfa Çevirme	0	0	0	0	2	1	1	0	0	1	15	0	0	0	0	0
Nesnenin Masaya Çarpması	0	0	1	3	3	1	4	0	0	0	3	3	0	2	0	0
Telefon	3	0	0	0	0	0	1	1	1	1	0	2	9	0	1	1
Yazıcı	0	0	0	1	0	0	1	0	0	0	0	0	0	18	0	0
Konuşma	0	1	0	0	0	0	1	0	1	1	0	0	1	0	15	0
Elektrik Düğmesi	1	1	0	0	0	4	2	0	0	3	4	1	1	0	0	3

Çizelge A. 2 20 Katsayılı MFCC öznelik vektörü kullanılarak DVM ile eğitilen sistemin hata matrisi

	Alarm	Boğaz Temizleme	Öksürük	Kapı Çarpması	Çekmece	Klavye	Anahtar	Kapı Vurma	Gülme	Bilgisayar Faresi	Sayfa Çevirme	Nesnenin Masaya Çarpması	Telefon	Yazıcı	Konuşma	Elektrik Düğmesi
Alarm	16	0	0	0	0	0	0	0	0	0	0	0	2	0	1	1
Boğaz Temizleme	0	17	0	0	1	0	0	0	0	1	0	0	0	0	1	0
Öksürük	1	1	7	0	2	1	2	0	4	0	2	0	0	0	0	0
Kapı Çarpması	0	0	0	12	3	0	0	1	0	0	1	0	0	3	0	0
Çekmece	0	0	0	4	14	0	0	0	0	0	1	0	0	0	1	0
Klavye	0	0	0	1	1	8	3	0	0	2	5	0	0	0	0	0
Anahtar	0	0	0	1	0	3	11	0	0	0	3	2	0	0	0	0
Kapı Vurma	0	0	0	0	0	0	0	18	0	2	0	0	0	0	0	0
Gülme	0	0	2	0	1	0	1	0	11	1	0	0	1	1	2	0
Bilgisayar Faresi	1	0	0	0	0	1	0	2	0	10	3	0	0	0	3	0
Sayfa Çevirme	0	0	0	0	1	4	2	0	0	0	13	0	0	0	0	0
Nesnenin Masaya Çarpması	0	0	1	3	2	2	3	0	0	1	4	3	0	1	0	0
Telefon	6	0	0	0	0	0	0	1	1	1	0	1	9	0	1	0
Yazıcı	0	0	0	2	1	0	0	1	0	0	0	0	0	16	0	0
Konuşma	1	1	0	0	0	0	1	0	2	0	0	0	0	0	15	0
Elektrik Düğmesi	0	2	0	0	0	3	3	0	0	2	4	2	0	0	0	4

Çizelge A. 3 48 Katsayılı MFCC öznitelik vektörü kullanılarak DVM ile eğitilen sistemin hata matrisi

	Alarm	Boğaz Temizleme	Öksürük	Kapı Çarpması	Çekmece	Klavye	Anahtar	Kapı Vurma	Gülme	Bilgisayar Faresi	Sayfa Çevirme	Nesnenin Masaya Çarpması	Telefon	Yazıcı	Konuşma	Elektrik Düğmesi
Alarm	14	0	0	0	0	0	0	0	1	1	0	0	3	0	0	1
Boğaz Temizleme	4	12	1	0	1	0	0	0	0	1	0	0	0	0	1	0
Öksürük	4	2	8	0	1	2	0	0	2	0	0	0	1	0	0	0
Kapı Çarpması	0	0	1	5	6	0	0	1	1	1	0	0	0	3	2	0
Çekmece	0	0	1	1	13	0	1	0	0	0	1	1	0	1	1	0
Klavye	0	0	2	1	2	5	2	0	0	2	3	0	0	0	0	3
Anahtar	0	0	0	1	1	0	9	0	0	0	5	3	0	0	0	1
Kapı Vurma	0	0	0	0	0	1	0	16	1	1	0	0	0	0	1	0
Gülme	1	2	3	0	0	0	0	0	12	0	0	0	1	1	0	0
Bilgisayar Faresi	0	1	0	0	0	1	0	1	0	13	0	0	1	0	1	2
Sayfa Çevirme	0	0	0	0	1	3	1	0	0	2	8	1	0	0	0	4
Nesnenin Masaya Çarpması	1	0	1	3	3	1	3	0	0	1	1	5	0	1	0	0
Telefon	4	0	0	0	0	0	0	0	3	1	0	1	11	0	0	0
Yazıcı	1	0	0	1	2	0	0	1	0	0	0	0	0	15	0	0
Konuşma	2	2	0	2	0	0	1	0	1	0	0	0	1	0	11	0
Elektrik Düğmesi	4	0	0	0	0	2	3	0	1	1	3	1	0	0	0	5

Çizelge A. 4 128 Katsayılı MFCC öznitelik vektörü kullanılarak DVM ile eğitilen sistemin hata matrisi

	Alarm	Boğaz Temizleme	Öksürük	Kapı Çarpması	Çekmece	Klavye	Anahtar	Kapı Vurma	Gülme	Bilgisayar Faresi	Sayfa Çevirme	Nesnenin Masaya Çarpması	Telefon	Yazıcı	Konuşma	Elektrik Düğmesi
Alarm	0	0	0	0	0	0	0	0	7	0	0	0	11	0	1	1
Boğaz Temizleme	0	3	1	0	0	0	0	0	5	1	0	0	9	0	1	0
Öksürük	0	0	0	1	1	1	1	0	6	0	1	0	9	0	0	0
Kapı Çarpması	0	0	1	6	3	1	1	1	1	0	1	2	0	2	1	0
Çekmece	0	0	0	3	12	0	1	0	0	1	2	0	1	0	0	0
Klavye	0	1	4	1	1	7	2	0	0	1	1	0	1	0	0	1
Anahtar	0	0	0	2	0	1	9	0	0	0	3	1	1	0	0	3
Kapı Vurma	0	0	0	0	0	0	0	15	2	0	0	0	3	0	0	0
Gülme	0	1	0	0	0	0	0	0	11	0	1	0	7	0	0	0
Bilgisayar Faresi	0	0	1	0	1	1	0	1	2	12	1	0	0	0	1	0
Sayfa Çevirme	0	0	2	0	0	3	2	0	1	0	9	1	0	0	0	2
Nesnenin Masaya Çarpması	0	0	1	4	3	0	2	0	4	0	1	0	4	0	0	1
Telefon	0	0	1	0	0	0	0	0	7	1	0	0	11	0	0	0
Yazıcı	0	0	0	2	1	0	0	1	2	0	0	0	1	13	0	0
Konuşma	0	1	0	2	0	0	1	0	0	1	0	0	4	0	11	0
Elektrik Düğmesi	1	2	0	0	0	2	3	0	3	0	3	1	3	0	0	2

Çizelge A. 5 Pencere süresi 10ms, atlama süresi 5 ms alınarak 20 katsayılı MFCC öznitelik vektörü kullanılarak DVM ile eğitilen sistemin hata matrisi

	Alarm	Boğaz Temizleme	Öksürük	Kapı Çarpması	Çekmece	Klavye	Anahtar	Kapı Vurma	Gülme	Bilgisayar Faresi	Sayfa Çevirme	Nesnenin Masaya Çarpması	Telefon	Yazıcı	Konuşma	Elektrik Düğmesi
Alarm	15	0	0	0	0	0	0	0	0	0	0	0	3	0	1	1
Boğaz Temizleme	0	18	0	0	1	0	0	0	0	1	0	0	0	0	0	0
Öksürük	0	2	8	1	4	2	0	0	0	0	2	1	0	0	0	0
Kapı Çarpması	0	0	0	11	2	1	0	1	0	0	1	0	0	4	0	0
Çekmece	0	0	0	3	15	1	0	0	0	0	1	0	0	0	0	0
Klavye	0	0	0	0	1	8	3	0	0	3	5	0	0	0	0	0
Anahtar	0	0	0	0	0	3	12	0	0	0	3	1	0	1	0	0
Kapı Vurma	0	0	0	0	0	0	0	18	0	2	0	0	0	0	0	0
Gülme	1	1	3	0	1	0	0	0	8	1	0	0	1	1	3	0
Bilgisayar Faresi	0	1	0	0	0	1	1	1	0	10	2	0	0	0	3	1
Sayfa Çevirme	0	0	0	0	2	2	2	0	0	1	13	0	0	0	0	0
Nesnenin Masaya Çarpması	0	0	1	2	4	0	2	0	1	1	4	4	0	1	0	0
Telefon	5	0	0	0	1	0	0	2	1	0	0	0	10	0	1	0
Yazıcı	0	0	1	1	0	0	1	0	0	0	0	0	0	17	0	0
Konuşma	0	2	0	0	0	0	1	0	1	1	0	0	0	0	15	0
Elektrik Düğmesi	1	1	0	0	0	3	2	0	0	3	4	2	0	0	0	4

Çizelge A. 6 Pencere süresi 10ms, atlama süresi 10 ms alınarak 20 katsayılı MFCC öznelik vektörü kullanılarak DVM ile eğitilen sistemin hata matrisi

	Alarm	Boğaz Temizleme	Öksürük	Kapı Çarpması	Çekmece	Klavye	Anahtar	Kapı Vurma	Gülme	Bilgisayar Faresi	Sayfa Çevirme	Nesnenin Masaya Çarpması	Telefon	Yazıcı	Konuşma	Elektrik Düğmesi
Alarm	14	0	0	0	0	0	0	0	1	0	0	0	3	0	1	1
Boğaz Temizleme	0	15	0	0	2	0	0	0	0	1	0	0	1	0	1	0
Öksürük	0	2	8	0	2	1	1	0	3	1	1	1	0	0	0	0
Kapı Çarpması	0	0	0	11	3	0	0	1	0	1	1	0	0	3	0	0
Çekmece	0	0	0	3	16	0	0	0	0	0	1	0	0	0	0	0
Klavye	0	0	0	1	0	10	3	0	0	1	4	0	0	0	0	1
Anahtar	0	0	0	1	0	1	13	0	0	0	3	2	0	0	0	0
Kapı Vurma	0	0	0	0	0	0	0	17	0	2	0	1	0	0	0	0
Gülme	0	0	4	0	0	0	0	0	9	1	0	1	2	1	2	0
Bilgisayar Faresi	1	2	0	1	0	2	1	1	0	9	1	0	0	0	2	0
Sayfa Çevirme	0	0	0	0	3	1	2	0	0	0	13	1	0	0	0	0
Nesnenin Masaya Çarpması	0	0	0	2	3	0	3	0	0	1	4	5	0	2	0	0
Telefon	7	0	0	0	0	0	0	2	1	0	0	2	7	0	1	0
Yazıcı	0	0	0	1	0	0	1	0	0	0	0	0	0	18	0	0
Konuşma	0	1	0	0	1	0	0	0	0	1	0	0	0	0	17	0
Elektrik Düğmesi	1	1	0	1	0	3	2	0	0	2	3	1	0	0	1	5

Çizelge A. 7 Pencere süresi 20ms, atlama süresi 10 ms alınarak 20 katsayılı MFCC öznitelik vektörü kullanılarak DVM ile eğitilen sistemin hata matrisi

	Alarm	Boğaz Temizleme	Öksürük	Kapı Çarpması	Çekmece	Klavye	Anahtar	Kapı Vurma	Gülme	Bilgisayar Faresi	Sayfa Çevirme	Nesnenin Masaya Çarpması	Telefon	Yazıcı	Konuşma	Elektrik Düğmesi
Alarm	14	0	0	0	0	0	0	0	0	0	0	1	2	0	1	2
Boğaz Temizleme	0	20	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Öksürük	2	1	5	0	3	1	0	0	4	1	2	1	0	0	0	0
Kapı Çarpması	0	0	0	10	4	1	0	1	0	0	0	1	0	3	0	0
Çekmece	0	0	0	6	11	1	0	0	0	0	1	0	0	0	1	0
Klavye	0	0	0	1	0	8	3	0	0	2	4	0	0	0	0	2
Anahtar	0	0	0	1	0	1	12	0	0	0	4	1	0	0	0	1
Kapı Vurma	0	0	0	0	0	1	0	18	0	1	0	0	0	0	0	0
Gülme	1	0	4	0	1	0	0	0	9	0	0	1	1	1	2	0
Bilgisayar Faresi	0	1	0	0	1	0	0	1	0	10	3	0	0	0	2	2
Sayfa Çevirme	0	0	0	0	1	1	1	0	0	0	17	0	0	0	0	0
Nesnenin Masaya Çarpması	0	0	0	3	3	1	2	0	0	1	4	5	0	1	0	0
Telefon	2	0	0	0	1	0	0	2	1	0	0	2	10	0	1	1
Yazıcı	0	0	0	1	0	0	1	0	0	0	0	0	0	18	0	0
Konuşma	0	1	0	0	0	0	1	0	1	1	0	0	1	0	15	0
Elektrik Düğmesi	2	1	0	0	0	3	4	0	0	3	4	1	1	0	0	1

Çizelge A. 8 Pencere süresi 20ms, atlama süresi 20ms alınarak 20 katsayılı MFCC öznitelik vektörü kullanılarak DVM ile eğitilen sistemin hata matrisi

	Alarm	Boğaz Temizleme	Öksürük	Kapı Çarpması	Çekmece	Klavye	Anahtar	Kapı Vurma	Gülme	Bilgisayar Faresi	Sayfa Çevirme	Nesnenin Masaya Çarpması	Telefon	Yazıcı	Konuşma	Elektrik Düğmesi
Alarm	15	0	0	0	0	0	0	0	0	0	0	1	2	0	1	1
Boğaz Temizleme	0	19	0	0	0	0	0	0	1	0	0	0	0	0	0	0
Öksürük	2	4	5	0	1	1	0	0	3	1	2	1	0	0	0	0
Kapı Çarpması	0	0	0	9	4	1	0	1	0	0	0	2	0	3	0	0
Çekmece	0	0	0	6	11	1	0	0	0	0	1	0	0	0	1	0
Klavye	0	0	0	1	0	8	3	0	0	2	6	0	0	0	0	0
Anahtar	0	0	0	1	0	2	11	0	0	0	4	1	0	0	0	1
Kapı Vurma	0	0	0	0	0	1	0	18	0	1	0	0	0	0	0	0
Gülme	1	1	3	0	0	0	0	0	10	0	0	1	1	1	2	0
Bilgisayar Faresi	1	1	0	0	1	0	0	1	0	12	2	0	0	0	2	0
Sayfa Çevirme	0	0	0	0	1	1	2	0	0	1	15	0	0	0	0	0
Nesnenin Masaya Çarpması	0	0	1	3	2	2	3	0	0	1	5	2	0	1	0	0
Telefon	2	0	0	0	1	0	0	2	1	0	0	1	12	0	1	0
Yazıcı	0	0	0	1	1	0	0	0	0	0	0	0	0	18	0	0
Konuşma	0	1	0	0	0	0	1	0	1	1	0	0	0	0	15	1
Elektrik Düğmesi	2	1	0	0	0	4	3	0	0	3	4	1	1	0	0	1

Çizelge A. 9 Pencere süresi 30ms, atlama süresi 10ms alınarak 20 katsayılı MFCC öznitelik vektörü kullanılarak DVM ile eğitilen sistemin hata matrisi

	Alarm	Boğaz Temizleme	Öksürük	Kapı Çarpması	Çekmece	Klavye	Anahtar	Kapı Vurma	Gülme	Bilgisayar Faresi	Sayfa Çevirme	Nesnenin Masaya Çarpması	Telefon	Yazıcı	Konuşma	Elektrik Düğmesi
Alarm	15	0	0	0	0	0	0	0	0	0	0	0	3	0	1	1
Boğaz Temizleme	0	17	0	0	1	0	0	0	0	1	0	0	0	0	1	0
Öksürük	1	1	7	0	2	1	2	0	4	0	2	0	0	0	0	0
Kapı Çarpması	0	0	0	12	3	0	0	1	0	0	1	0	0	3	0	0
Çekmece	0	0	0	3	15	0	0	0	0	0	1	0	0	0	1	0
Klavye	0	0	0	1	0	9	3	0	0	2	5	0	0	0	0	0
Anahtar	0	0	0	1	0	3	12	0	0	0	3	1	0	0	0	0
Kapı Vurma	0	0	0	0	0	0	0	18	0	2	0	0	0	0	0	0
Gülme	0	0	3	0	1	0	1	0	10	1	0	0	1	1	2	0
Bilgisayar Faresi	1	0	0	0	0	1	0	2	0	10	3	0	0	0	3	0
Sayfa Çevirme	0	0	0	0	1	5	2	0	0	1	11	0	0	0	0	0
Nesnenin Masaya Çarpması	0	0	1	3	1	2	3	0	0	1	4	4	0	1	0	0
Telefon	6	0	0	0	0	0	0	1	1	1	0	1	9	0	1	0
Yazıcı	0	0	0	2	1	0	0	1	0	0	0	0	0	16	0	0
Konuşma	1	1	0	0	0	0	1	0	2	0	0	0	0	0	15	0
Elektrik Düğmesi	2	1	0	0	0	3	3	0	0	2	3	2	0	0	0	4

Çizelge A. 10 Pencere süresi 30ms, atlama süresi 15ms alınarak 20 katsayılı MFCC öznitelik vektörü kullanılarak DVM ile eğitilen sistemin hata matrisi

	Alarm	Boğaz Temizleme	Öksürük	Kapı Çarpması	Çekmece	Klavye	Anahtar	Kapı Vurma	Gülme	Bilgisayar Faresi	Sayfa Çevirme	Nesnenin Masaya Çarpması	Telefon	Yazıcı	Konuşma	Elektrik Düğmesi
Alarm	15	0	0	0	0	0	0	0	0	0	0	0	3	0	1	1
Boğaz Temizleme	0	19	0	0	1	0	0	0	0	0	0	0	0	0	0	0
Öksürük	1	1	7	0	2	1	2	0	3	0	2	0	0	0	0	1
Kapı Çarpması	0	0	0	11	4	0	0	1	0	0	1	0	0	3	0	0
Çekmece	0	0	0	6	12	0	0	0	0	0	1	0	0	0	1	0
Klavye	0	0	0	1	1	5	3	0	0	3	6	0	0	0	0	1
Anahtar	0	0	0	1	0	1	13	0	0	0	3	2	0	0	0	0
Kapı Vurma	0	0	0	0	0	0	0	18	0	2	0	0	0	0	0	0
Gülme	1	1	2	1	1	0	0	0	9	1	0	0	1	1	2	0
Bilgisayar Faresi	0	2	0	0	0	1	1	2	0	6	2	0	0	0	3	3
Sayfa Çevirme	0	0	0	0	1	1	2	0	0	0	16	0	0	0	0	0
Nesnenin Masaya Çarpması	0	0	1	2	3	0	3	0	0	1	4	4	0	2	0	0
Telefon	4	0	0	0	0	0	0	1	2	1	0	3	7	0	2	0
Yazıcı	0	0	0	2	1	0	0	0	0	0	0	0	0	17	0	0
Konuşma	0	1	0	0	0	1	1	0	2	0	0	0	1	0	14	0
Elektrik Düğmesi	1	1	0	0	0	0	4	0	0	5	5	1	0	0	0	3

Çizelge A. 11 Pencere süresi 30ms, atlama süresi 30ms alınarak 20 katsayılı MFCC öznitelik vektörü kullanılarak DVM ile eğitilen sistemin hata matrisi

	Alarm	Boğaz Temizleme	Öksürük	Kapı Çarpması	Çekmece	Klavye	Anahtar	Kapı Vurma	Gülme	Bilgisayar Faresi	Sayfa Çevirme	Nesnenin Masaya Çarpması	Telefon	Yazıcı	Konuşma	Elektrik Düğmesi
Alarm	15	0	0	0	0	0	0	0	0	0	0	0	3	0	1	1
Boğaz Temizleme	0	18	0	0	2	0	0	0	0	0	0	0	0	0	0	0
Öksürük	2	2	6	0	1	2	1	0	3	0	2	0	1	0	0	0
Kapı Çarpması	0	0	0	11	3	0	0	1	0	1	0	1	0	3	0	0
Çekmece	0	0	0	5	14	0	0	0	0	0	1	0	0	0	0	0
Klavye	1	0	0	1	0	8	2	0	0	1	5	1	0	0	0	1
Anahtar	0	0	0	1	0	1	12	0	0	0	5	1	0	0	0	0
Kapı Vurma	0	0	0	0	0	1	0	18	0	1	0	0	0	0	0	0
Gülme	1	2	1	1	0	0	0	0	8	1	0	0	2	1	3	0
Bilgisayar Faresi	1	1	0	0	1	1	0	1	0	8	2	0	0	1	2	2
Sayfa Çevirme	0	0	0	0	0	3	2	0	0	1	14	0	0	0	0	0
Nesnenin Masaya Çarpması	0	0	1	2	3	0	2	0	0	1	4	5	0	2	0	0
Telefon	6	0	0	0	0	0	0	1	1	1	0	2	7	0	1	1
Yazıcı	0	0	0	1	0	0	0	0	0	0	0	0	0	19	0	0
Konuşma	0	1	0	0	0	0	1	0	1	0	0	0	1	0	15	1
Elektrik Düğmesi	2	2	1	0	0	3	3	0	0	4	2	1	0	0	0	2

Çizelge A. 12 Pencere süresi 50ms, atlama süresi 10ms alınarak 20 katsayılı MFCC öznitelik vektörü kullanılarak DVM ile eğitilen sistemin hata matrisi

	Alarm	Boğaz Temizleme	Öksürük	Kapı Çarpması	Çekmece	Klavye	Anahtar	Kapı Vurma	Gülme	Bilgisayar Faresi	Sayfa Çevirme	Nesnenin Masaya Çarpması	Telefon	Yazıcı	Konuşma	Elektrik Düğmesi
Alarm	9	0	0	0	0	0	0	0	1	0	0	1	7	0	1	1
Boğaz Temizleme	0	19	0	0	0	0	0	0	0	0	1	0	0	0	0	0
Öksürük	1	1	8	0	3	2	0	0	1	0	2	1	1	0	0	0
Kapı Çarpması	0	0	0	12	3	0	0	1	0	0	0	1	0	3	0	0
Çekmece	0	0	0	5	11	1	0	0	0	0	1	1	0	0	1	0
Klavye	0	0	0	0	1	8	4	0	0	1	5	0	0	0	0	1
Anahtar	0	0	0	1	0	3	10	0	0	0	4	2	0	0	0	0
Kapı Vurma	0	0	0	0	0	0	0	18	0	2	0	0	0	0	0	0
Gülme	3	0	3	0	1	0	1	0	7	1	0	0	1	1	2	0
Bilgisayar Faresi	1	2	0	0	0	2	1	0	0	9	2	0	0	0	3	0
Sayfa Çevirme	0	0	0	0	0	1	2	0	0	1	15	0	0	0	0	1
Nesnenin Masaya Çarpması	0	0	0	3	3	0	2	0	0	0	4	5	1	2	0	0
Telefon	2	0	0	0	0	0	0	1	1	1	0	4	11	0	0	0
Yazıcı	0	0	0	1	1	0	0	0	0	0	0	0	0	18	0	0
Konuşma	0	1	0	0	0	0	0	0	2	0	1	0	1	0	15	0
Elektrik Düğmesi	1	2	0	0	0	5	3	0	0	3	2	2	2	0	0	0

Çizelge A. 13 Pencere süresi 50ms, atlama süresi 25ms alınarak 20 katsayılı MFCC öznitelik vektörü kullanılarak DVM ile eğitilen sistemin hata matrisi

	Alarm	Boğaz Temizleme	Öksürük	Kapı Çarpması	Çekmece	Klavye	Anahtar	Kapı Vurma	Gülme	Bilgisayar Faresi	Sayfa Çevirme	Nesnenin Masaya Çarpması	Telefon	Yazıcı	Konuşma	Elektrik Düğmesi
Alarm	15	0	0	0	0	0	0	0	0	0	0	0	3	0	1	1
Boğaz Temizleme	0	15	1	0	1	0	0	0	0	0	0	0	0	0	2	1
Öksürük	1	0	6	0	3	1	0	0	5	0	2	1	0	1	0	0
Kapı Çarpması	0	0	0	9	3	1	0	1	0	0	2	0	0	3	1	0
Çekmece	0	0	0	6	12	0	0	0	0	0	1	0	0	0	1	0
Klavye	0	0	0	1	1	9	2	0	0	1	4	1	0	0	0	1
Anahtar	0	0	0	1	0	2	12	0	0	0	3	2	0	0	0	0
Kapı Vurma	0	0	0	0	0	0	0	18	0	2	0	0	0	0	0	0
Gülme	1	0	2	1	1	0	0	0	9	1	0	0	2	1	1	1
Bilgisayar Faresi	0	0	0	0	0	1	0	1	0	12	2	0	0	1	2	1
Sayfa Çevirme	0	0	0	0	0	2	2	0	0	0	16	0	0	0	0	0
Nesnenin Masaya Çarpması	0	0	1	3	4	2	2	0	0	0	3	4	0	1	0	0
Telefon	8	0	0	0	1	0	0	1	0	1	0	1	6	0	2	0
Yazıcı	0	0	0	2	0	0	1	0	0	0	0	0	0	17	0	0
Konuşma	1	1	0	0	0	1	1	0	2	0	0	0	0	0	14	0
Elektrik Düğmesi	0	2	0	0	0	2	2	0	0	2	6	1	1	0	0	4

Çizelge A. 14 Pencere süresi 50ms, atlama süresi 50ms alınarak 20 katsayılı MFCC öznitelik vektörü kullanılarak DVM ile eğitilen sistemin hata matrisi

	Alarm	Boğaz Temizleme	Öksürük	Kapı Çarpması	Çekmece	Klavye	Anahtar	Kapı Vurma	Gülme	Bilgisayar Faresi	Sayfa Çevirme	Nesnenin Masaya Çarpması	Telefon	Yazıcı	Konuşma	Elektrik Düğmesi
Alarm	15	0	0	0	0	0	0	0	0	0	0	0	3	0	1	1
Boğaz Temizleme	0	16	0	0	2	0	0	0	0	0	0	0	0	0	2	0
Öksürük	1	3	7	0	1	1	0	0	2	0	2	1	2	0	0	0
Kapı Çarpması	0	0	0	11	3	0	0	1	0	1	1	0	0	3	0	0
Çekmece	0	0	0	5	12	0	1	0	0	0	1	0	0	0	1	0
Klavye	1	0	0	0	1	9	3	0	0	1	3	0	0	0	0	2
Anahtar	0	0	0	1	0	4	11	0	0	0	3	1	0	0	0	0
Kapı Vurma	0	0	0	0	0	0	0	18	0	2	0	0	0	0	0	0
Gülme	3	1	5	0	0	0	0	0	6	1	0	0	1	1	2	0
Bilgisayar Faresi	1	0	0	0	1	2	1	1	0	11	1	0	0	0	2	0
Sayfa Çevirme	0	0	0	0	1	2	2	0	0	1	13	0	0	0	0	1
Nesnenin Masaya Çarpması	0	0	1	4	2	1	2	0	0	1	2	5	0	1	0	1
Telefon	6	0	0	0	0	0	0	1	1	1	0	3	7	0	1	0
Yazıcı	0	0	1	1	1	0	0	0	0	0	0	0	0	17	0	0
Konuşma	2	1	0	0	0	0	1	0	1	0	0	0	0	0	14	1
Elektrik Düğmesi	2	2	1	0	0	1	4	0	0	3	0	2	0	0	1	4

Çizelge A. 15 Pencere süresi 100ms, atlama süresi 10ms alınarak 20 katsayılı MFCC öznitelik vektörü kullanılarak DVM ile eğitilen sistemin hata matrisi

	Alarm	Boğaz Temizleme	Öksürük	Kapı Çarpması	Çekmece	Klavye	Anahtar	Kapı Vurma	Gülme	Bilgisayar Faresi	Sayfa Çevirme	Nesnenin Masaya Çarpması	Telefon	Yazıcı	Konuşma	Elektrik Düğmesi
Alarm	16	0	0	0	0	0	0	0	0	0	0	0	2	0	1	1
Boğaz Temizleme	0	17	0	0	0	0	0	0	0	1	0	0	0	0	1	1
Öksürük	2	1	7	0	2	2	0	0	2	0	2	1	1	0	0	0
Kapı Çarpması	0	0	0	11	3	0	0	1	0	0	0	1	0	4	0	0
Çekmece	0	0	0	5	12	1	0	0	0	0	1	0	0	1	0	0
Klavye	0	0	0	1	0	6	3	0	0	2	6	1	0	0	0	1
Anahtar	0	0	0	1	0	2	12	0	0	0	2	2	0	1	0	0
Kapı Vurma	0	0	0	0	0	0	0	18	0	2	0	0	0	0	0	0
Gülme	3	2	4	0	0	0	0	0	4	1	0	0	2	1	2	1
Bilgisayar Faresi	1	0	0	0	0	1	1	1	0	13	0	0	0	0	3	0
Sayfa Çevirme	0	0	0	0	0	1	2	0	0	1	16	0	0	0	0	0
Nesnenin Masaya Çarpması	0	0	1	3	2	1	2	0	0	1	2	7	0	1	0	0
Telefon	6	0	0	0	0	0	0	1	1	1	0	2	8	0	1	0
Yazıcı	0	0	1	1	1	0	0	0	0	0	0	0	0	17	0	0
Konuşma	1	1	0	0	0	0	1	0	2	0	0	0	0	0	15	0
Elektrik Düğmesi	3	1	0	0	0	2	1	0	0	4	1	1	0	0	1	6

Çizelge A. 16 Pencere süresi 100ms, atlama süresi 50ms alınarak 20 katsayılı MFCC öznitelik vektörü kullanılarak DVM ile eğitilen sistemin hata matrisi

	Alarm	Boğaz Temizleme	Öksürük	Kapı Çarpması	Çekmece	Klavye	Anahtar	Kapı Vurma	Gülme	Bilgisayar Faresi	Sayfa Çevirme	Nesnenin Masaya Çarpması	Telefon	Yazıcı	Konuşma	Elektrik Düğmesi
Alarm	16	0	0	0	0	0	0	0	0	1	0	0	2	0	0	1
Boğaz Temizleme	0	15	1	0	1	0	0	0	0	0	0	0	0	0	3	0
Öksürük	2	0	7	0	2	2	0	0	3	0	2	1	1	0	0	0
Kapı Çarpması	0	0	0	11	3	0	0	1	0	0	1	0	0	3	1	0
Çekmece	0	0	0	5	13	0	0	0	0	0	1	1	0	0	0	0
Klavye	0	0	0	1	1	5	3	0	0	1	8	1	0	0	0	0
Anahtar	0	0	0	1	1	2	10	0	0	0	4	1	0	1	0	0
Kapı Vurma	0	0	0	0	0	0	0	18	0	2	0	0	0	0	0	0
Gülme	3	0	4	0	0	0	0	0	7	1	0	0	2	1	2	0
Bilgisayar Faresi	1	0	0	0	0	2	1	0	0	13	1	0	0	0	2	0
Sayfa Çevirme	0	0	0	0	0	2	2	0	0	0	16	0	0	0	0	0
Nesnenin Masaya Çarpması	0	0	1	3	2	1	1	0	0	0	4	6	0	2	0	0
Telefon	4	0	1	0	0	0	0	1	2	1	0	2	9	0	0	0
Yazıcı	0	0	0	2	1	0	0	0	0	0	0	0	0	17	0	0
Konuşma	1	1	0	0	0	0	1	0	2	0	0	0	0	0	15	0
Elektrik Düğmesi	2	3	0	0	0	1	3	0	0	4	4	1	0	0	0	2

Çizelge A. 17 Pencere süresi 100ms, atlama süresi 100ms alınarak 20 katsayılı MFCC öznitelik vektörü kullanılarak DVM ile eğitilen sistemin hata matrisi

	Alarm	Boğaz Temizleme	Öksürük	Kapı Çarpması	Çekmece	Klavye	Anahtar	Kapı Vurma	Gülme	Bilgisayar Faresi	Sayfa Çevirme	Nesnenin Masaya Çarpması	Telefon	Yazıcı	Konuşma	Elektrik Düğmesi
Alarm	15	0	0	0	0	0	0	0	0	0	0	1	2	0	1	1
Boğaz Temizleme	0	19	1	0	0	0	0	0	0	0	0	0	0	0	0	0
Öksürük	2	1	6	0	1	2	0	0	4	0	2	1	0	1	0	0
Kapı Çarpması	0	0	0	10	4	0	0	1	0	0	1	0	0	3	1	0
Çekmece	0	0	0	5	13	0	0	0	0	0	1	0	0	0	1	0
Klavye	0	0	0	1	0	7	4	0	0	1	6	1	0	0	0	0
Anahtar	0	0	0	1	1	2	10	0	0	1	3	2	0	0	0	0
Kapı Vurma	0	0	0	0	0	0	0	18	0	2	0	0	0	0	0	0
Gülme	1	1	4	0	0	0	0	0	9	1	0	0	3	1	0	0
Bilgisayar Faresi	1	1	0	0	0	1	1	0	0	14	0	0	0	0	2	0
Sayfa Çevirme	0	0	0	0	1	1	2	0	0	0	16	0	0	0	0	0
Nesnenin Masaya Çarpması	0	0	1	2	3	1	3	0	0	0	3	4	0	2	0	1
Telefon	5	0	0	0	0	0	0	1	2	1	0	3	7	0	0	1
Yazıcı	0	0	0	1	1	0	0	0	0	1	0	0	0	16	1	0
Konuşma	1	1	0	1	0	1	1	0	2	1	0	0	0	0	12	0
Elektrik Düğmesi	2	1	0	0	0	2	1	0	0	3	5	1	0	0	0	5

Çizelge A. 18 MFCC öznitelik çıkarımında varsayılan alt ve üst frekans değerleri kullanılarak DVM ile eğitilen sistemin hata matrisi

	Alarm	Boğaz Temizleme	Öksürük	Kapı Çarpması	Çekmece	Klavye	Anahtar	Kapı Vurma	Gülme	Bilgisayar Faresi	Sayfa Çevirme	Nesnenin Masaya Çarpması	Telefon	Yazıcı	Konuşma	Elektrik Düğmesi
Alarm	15	0	0	0	0	0	0	0	0	0	0	0	3	0	1	1
Boğaz Temizleme	0	17	0	0	1	0	0	0	0	1	0	0	0	0	1	0
Öksürük	1	1	7	0	2	1	2	0	4	0	2	0	0	0	0	0
Kapı Çarpması	0	0	0	12	3	0	0	1	0	0	1	0	0	3	0	0
Çekmece	0	0	0	3	15	0	0	0	0	0	1	0	0	0	1	0
Klavye	0	0	0	1	0	9	3	0	0	2	5	0	0	0	0	0
Anahtar	0	0	0	1	0	3	12	0	0	0	3	1	0	0	0	0
Kapı Vurma	0	0	0	0	0	0	0	18	0	2	0	0	0	0	0	0
Gülme	0	0	3	0	1	0	1	0	10	1	0	0	1	1	2	0
Bilgisayar Faresi	1	0	0	0	0	1	0	2	0	10	3	0	0	0	3	0
Sayfa Çevirme	0	0	0	0	1	5	2	0	0	1	11	0	0	0	0	0
Nesnenin Masaya Çarpması	0	0	1	3	1	2	3	0	0	1	4	4	0	1	0	0
Telefon	6	0	0	0	0	0	0	1	1	1	0	1	9	0	1	0
Yazıcı	0	0	0	2	1	0	0	1	0	0	0	0	0	16	0	0
Konuşma	1	1	0	0	0	0	1	0	2	0	0	0	0	0	15	0
Elektrik Düğmesi	2	1	0	0	0	3	3	0	0	2	3	2	0	0	0	4

Çizelge A. 19 MFCC özniteliği, baskın frekans değeri kullanılarak çıkarıldığında DVM ile eğitilen sistemin hata matrisi.

	Alarm	Boğaz Temizleme	Öksürük	Kapı Çarpması	Çekmece	Klavye	Anahtar	Kapı Vurma	Gülme	Bilgisayar Faresi	Sayfa Çevirme	Nesnenin Masaya Çarpması	Telefon	Yazıcı	Konuşma	Elektrik Düğmesi
Alarm	13	0	0	0	0	0	0	0	1	0	0	0	4	0	2	0
Boğaz Temizleme	0	18	1	0	0	0	0	0	0	0	0	0	0	0	1	0
Öksürük	0	0	13	1	0	0	2	0	0	0	0	2	1	1	0	0
Kapı Çarpması	0	0	0	12	4	0	0	0	0	0	1	0	0	3	0	0
Çekmece	0	0	0	2	17	0	0	0	0	0	0	0	0	0	1	0
Klavye	0	0	0	1	0	11	0	0	0	1	4	1	0	0	0	2
Anahtar	0	0	0	0	0	0	19	0	0	0	0	1	0	0	0	0
Kapı Vurma	0	0	0	1	0	0	0	17	0	2	0	0	0	0	0	0
Gülme	0	1	1	0	0	0	0	0	15	0	0	0	1	1	1	0
Bilgisayar Faresi	0	0	0	1	0	0	0	1	0	15	2	0	0	0	0	1
Sayfa Çevirme	0	0	0	0	1	3	0	0	0	0	15	0	0	0	0	1
Nesnenin Masaya Çarpması	0	0	1	0	1	1	0	0	0	1	4	7	0	3	1	1
Telefon	1	1	0	0	1	0	0	1	0	0	0	0	14	0	2	0
Yazıcı	0	0	0	0	1	0	0	1	0	0	0	0	0	18	0	0
Konuşma	0	0	0	0	1	0	0	0	0	0	0	0	0	0	19	0
Elektrik Düğmesi	0	1	0	0	0	3	0	0	1	1	2	3	0	0	0	9

EK B: İŞİTSEL SAHNE TANIMA SONUÇLARININ HATA MATRİSLERİ

Çizelge B. 1 Varsayılan DVM parametreleri kullanılarak eğitilen model ile test edilen sistemin hata matrisi

	Otobüs	Kalabalık Sokak	Ofis Ortamı	Açık Market	Park	Sessiz Sokak	Restoran	Süpermarket	Tren	Tren İstasyonu
Otobüs	7	0	1	0	0	0	1	1	0	0
Kalabalık Sokak	0	10	0	0	0	0	0	0	0	0
Ofis Ortamı	0	2	6	0	1	1	0	0	0	0
Açık Market	2	1	0	7	0	0	0	0	0	0
Park	0	3	3	0	1	1	1	0	0	1
Sessiz Sokak	0	0	1	1	2	6	0	0	0	0
Restoran	3	2	0	1	0	0	0	1	1	2
Süpermarket	3	0	1	1	1	0	1	2	1	0
Tren	0	5	1	1	0	0	0	0	2	1
Tren İstasyonu	0	1	0	1	1	0	0	0	1	6

Çizelge B. 2 En uygun değeri ayarlanmış MFCC öznitelik vektörleri ve optimize edilmiş DVM parametreleri kullanılarak eğitilen model ile test edilen sistemin hata matrisi

	Otobüs	Kalabalık Sokak	Ofis Ortamı	Açık Market	Park	Sessiz Sokak	Restoran	Süpermarket	Tren	Tren İstasyonu
Otobüs	6	0	0	0	0	0	2	1	1	0
Kalabalık Sokak	0	9	0	0	0	0	1	0	0	0
Ofis Ortamı	0	0	8	0	1	1	0	0	0	0
Açık Market	2	1	0	6	0	0	0	1	0	0
Park	0	0	2	1	3	3	0	1	0	0
Sessiz Sokak	0	0	1	0	1	7	0	1	0	0
Restoran	1	1	0	2	0	0	2	2	2	0
Süpermarket	2	0	1	2	0	1	1	1	0	2
Tren	0	1	0	0	1	0	2	0	5	1
Tren İstasyonu	0	1	0	1	0	0	1	2	1	4

Çizelge B. 3 Önerilen sistemin hata matrisi

	Otobüs	Kalabalık Sokak	Ofis Ortamı	Açık Market	Park	Sessiz Sokak	Restoran	Süpermarket	Tren	Tren İstasyonu
Otobüs	7	0	0	0	0	0	0	2	1	0
Kalabalık Sokak	0	10	0	0	0	0	0	0	0	0
Ofis Ortamı	0	0	8	0	1	1	0	0	0	0
Açık Market	0	1	0	6	0	0	1	2	0	0
Park	0	0	2	1	3	4	0	0	0	0
Sessiz Sokak	0	0	1	1	2	6	0	0	0	0
Restoran	0	1	0	2	0	0	4	2	0	1
Süpermarket	2	0	1	2	0	0	0	3	0	2
Tren	1	1	0	0	0	0	1	0	6	1
Tren İstasyonu	0	1	0	1	0	0	0	1	0	7

EK C: SES OLAYLARININ İŞİTSEL SAHNELERE OLAN KATKILARI HATA MATRİSLERİ

Çizelge C. 1 Ses dosyası içerisindeki 30ms lik çerçeveleri ses olayları ile etiketlendiğinde, her dosyanın içerisindeki ses olaylarının sayısı

	Alarm	Boğaz Temizleme	Öksürük	Kapı Çarpması	Çekmece	Klavye	Anahtar	Kapı Vurma	Gülme	Bilgisayar Faresi	Sayfa Çevirme	Objenin Masaya Çarpması	Telefon	Yazıcı	Konuşma	Elektrik Düşmesi
Ofis01	4	17	165	0	1	77	0	0	91	0	0	80	262	295	7	1
Ofis02	3	102	59	2	0	39	475	0	89	1	75	13	11	4	127	0
Ofis03	0	10	108	1	2	0	0	50	26	0	0	116	608	76	0	3
Ofis04	8	17	218	1	6	103	1	2	239	0	0	128	137	134	6	0
Ofis05	0	1	76	0	0	0	0	2	6	0	0	3	901	11	0	0
Ofis06	7	48	88	2	0	0	0	34	26	0	0	356	354	39	0	46
Ofis07	4	36	147	31	5	204	74	2	64	0	0	281	65	81	6	0
Ofis08	0	42	400	3	20	1	0	4	31	0	0	235	207	57	0	0
Ofis09	4	62	97	221	4	142	24	2	68	5	84	168	56	6	51	6
Ofis10	1	155	109	0	0	0	0	282	80	0	0	24	341	8	0	0

Çizelge C. 2 Ses dosyası içerisindeki 90ms lik çerçeveleri ses olayları ile etiketlendiğinde, her dosyanın içerisindeki ses olaylarının sayıları

	Alarm	Boğaz Temizleme	Öksürük	Kapı Çarpması	Çekmece	Klavye	Anahtar	Kapı Vurma	Gülme	Bilgisayar Faresi	Sayfa Çevirme	Objenin Masaya Çarpması	Telefon	Yazıcı	Konuşma	Elektrik Düğmesi
Ofis01	0	1	45	0	0	24	0	0	12	0	0	19	49	181	3	0
Ofis02	0	13	4	0	0	7	231	0	13	0	19	5	3	1	38	0
Ofis03	0	1	25	0	0	0	0	23	1	0	0	64	186	34	0	0
Ofis04	0	1	85	0	1	38	0	1	70	0	0	45	23	69	1	0
Ofis05	0	0	26	0	0	0	0	2	1	0	0	3	292	10	0	0
Ofis06	1	6	16	0	0	0	0	16	4	0	0	188	72	16	0	15
Ofis07	0	5	29	11	1	73	30	1	10	0	0	134	13	25	2	0
Ofis08	0	4	159	0	3	0	0	2	2	0	0	115	40	9	0	0
Ofis09	0	10	14	84	2	51	1	0	7	0	82	67	7	3	6	0
Ofis10	0	45	44	0	0	0	0	141	20	0	0	7	74	3	0	0

Çizelge C. 3 Ses dosyası içerisindeki 990ms lik çerçeveleri ses olayları ile etiketlendiğinde, her dosyanın içerisindeki ses olaylarının sayıları

	Alarm	Boğaz Temizleme	Öksürük	Kapı Çarpması	Çekmece	Klavye	Anahtar	Kapı Vurma	Gülme	Bilgisayar Faresi	Sayfa Çevirme	Objenin Masaya Çarpması	Telefon	Yazıcı	Konuşma	Elektrik Düğmesi
Ofis01	0	0	2	0	0	0	0	0	0	0	0	0	3	25	1	0
Ofis02	0	0	1	0	0	0	28	0	0	0	0	0	0	0	2	0
Ofis03	0	0	0	0	0	0	0	1	0	0	0	11	16	3	0	0
Ofis04	0	0	6	0	0	4	0	0	4	0	0	1	0	16	0	0
Ofis05	0	0	2	0	0	0	0	0	0	0	0	0	29	0	0	0
Ofis06	0	0	0	0	0	0	0	1	1	0	0	26	2	0	0	1
Ofis07	0	0	0	0	0	8	1	0	0	0	0	19	0	3	0	0
Ofis08	0	0	16	0	0	0	0	0	0	0	0	14	0	1	0	0
Ofis09	0	0	0	5	0	3	0	0	0	0	11	12	0	0	0	0
Ofis10	0	2	7	0	0	0	0	19	1	0	0	0	2	0	0	0