İSTANBUL TEKNİK ÜNİVERSİTESİ ★ FEN BİLİMLERİ ENSTİTÜSÜ

SAHNE ANALİZİ İÇİN SES KAYNAĞI TESPİTİ

YÜKSEK LİSANS TEZİ

İsmail İren SALTALI

Bilgisayar Mühendisliği Anabilim Dalı

Bilgisayar Mühendisliği Programı

MAYIS 2015

İSTANBUL TEKNİK ÜNİVERSİTESİ ★ FEN BİLİMLERİ ENSTİTÜSÜ

SAHNE ANALİZİ İÇİN SES KAYNAĞI TESPİTİ

YÜKSEK LİSANS TEZİ

İsmail İren SALTALI
(504121525)

Bilgisayar Mühendisliği Anabilim Dalı

Bilgisayar Mühendisliği Programı

Tez Danışmanı: Yrd.Doç.Dr. Gökhan İNCE

MAYIS 2015

İTÜ, Fen Bilimleri Enstitüsü'nün 504121525 numaralı Yüksek Lisans Öğrencisi **İsmail İren SALTALI**, ilgili yönetmeliklerin belirlediği gerekli tüm şartları yerine getirdikten sonra hazırladığı **"SAHNE ANALİZİ İÇİN SES KAYNAĞI TESPİTİ"** başlıklı tezini aşağıdaki imzaları olan jüri önünde başarı ile sunmuştur.

Tez Danışmanı :	**Yrd.Doç.Dr. Gökhan İNCE** İstanbul Teknik Üniversitesi	.....
Jüri Üyeleri :	**Yrd.Doç.Dr. Yusuf YASLAN** İstanbul Teknik Üniversitesi	.....
	Yrd.Doç.Dr. Sinan KALKAN Ortadoğu Teknik Üniversitesi	.....
		.....

Teslim Tarihi : **04 Mayıs 2015**
Savunma Tarihi : **22 Mayıs 2015**

Babama,

ÖNSÖZ

İstanbul Teknik Üniversitesi Bilimsel Araştırma Projeleri Koordinasyon Birimine (BAP proje no: 37967) verdikleri destek için teşekkür ederim.

Bugüne kadar desteklerini hiçbir zaman esirgemeyen anneme ve babama, hergün yanımda olan ve yıllardır ev arkadaşlığımı da yapan çok sevgili ve sevimli kardeşime ve bundan sonraki yeni çekirdek ailemin ilk üyesi, hayat arkadaşıma sabır, destek ve sevgileri için teşekkür ederim.

Mayıs 2015 İsmail İren SALTALI

İÇİNDEKİLER

Sayfa

ÖNSÖZ ..... **vii**
İÇİNDEKİLER ..... **ix**
KISALTMALAR ..... **xi**
ÇİZELGE LİSTESİ ..... **xiii**
ŞEKİL LİSTESİ ..... **xv**
ÖZET ..... **xvii**
SUMMARY ..... **xix**
1. GİRİŞ **1**
1.1 Tezin Amacı 2
1.2 İlgili Çalışmalar 3
1.3 Uygulama 6
2. SİSTEME GENEL BAKIŞ **9**
2.1 Önişleme 9
2.1.1 Başlangıç ve bitiş ses eşiklerinin tespiti 11
2.1.2 Ayrık Fourier Dönüşümü (DFT) 11
2.2 Öznitelik Çıkarma 12
2.2.1 Spektral band gücü özniteliği 13
2.2.2 Sesin süresi özniteliği 16
2.2.3 Tınısal öznitelikler 16
2.2.3.1 Zamansal sıfır geçiş oranı 17
2.2.3.2 Spektral ağırlık merkezi 17
2.2.3.3 Azalış ve akış 17
2.2.3.4 Mel ölçekli kepstrum katsayıları (MFCC) 17
2.3 Sınıflandırma 18
2.3.1 Destek vektör makinesi (SVM) 18
2.3.2 Çok katmanlı algılayıcı (MLP) 19
3. LABORATUVAR DENEYLERİ **21**
3.1 Donanım ve Yazılım Özellikleri 21
3.2 Nesneler 22
3.3 Deneysel Düzen 23
3.4 Değerlendirme Kriterleri 23
3.5 Deneyler ve Sonuçlar 24
3.5.1 Çarpma olayı ile malzeme sınıflandırma 25
3.5.2 Olay sınıflandırma 26
3.5.3 Çarpma ve düşme olayları ile malzeme sınıflandırma 27

4. ROBOT DENEYLERİ 31
4.1 Robotlar 31
4.2 Deneyler ve Sonuçlar..... 31
4.2.1 Gezgin robot ile deneyler 31
4.2.2 Robotik kol ile deneyler 33
5. SONUÇ VE GELECEK ÇALIŞMALAR..... 35
KAYNAKLAR..... 37
ÖZGEÇMİŞ 41

KISALTMALAR

DFT	**:** Discrete Fourier Transform (Ayrık Fourier Dönüşümü)
FFT	**:** Fast Fourier Transform (Hızlı Fourier Dönüşümü)
SVM	**:** Support Vector Machine (Destek Vektör Makinesi)
MLP	**:** Multi Layer Perceptron (Çok Katmanlı Algılama)
Marsyas	**:** Music Analysis, Retrieval and Synthesis for Audio Signals (Ses Sinyalleri için Müzik Analizi, Öznitelik Çıkarma ve Birleştirme)
ZCR	**:** Zero Crossing Rate (Sıfır Geçiş Oranı)
DCT	**:** Discrete Cosine Transform (Ayrık Kosinüs Dönüşümü)
MFCC	**:** Mel-Frequency Cepstral Coefficients (Mel Ölçekli Kepstrum Katsayıları)
WEKA	**:** Waikato Environment for Knowledge Analysis
ROS	**:** The Robot Operating System (Robot İşletim Sistemi)
LIBSVM	**:** A Library for Support Vector Machines (Destek Vektör Makineleri İçin Kütüphane)
BSD	**:** Berkeley Software Distribution (Berkeley Yazılım Dağıtımı)
SURF	**:** The Speeded-Up Robust Features (Hızlandırılmış Sağlam Özellikler)
RBF	**:** Radial Basis Function (Radyal Temelli Fonksiyon)
İTÜ	**:** İstanbul Teknik Üniversitesi

ÇİZELGE LİSTESİ

Sayfa

Çizelge 2.1: Düşme ve çarpma olaylarında sesin süreleri 16
Çizelge 3.1: Çarpma olayı ile malzeme sınıflandırma..... 25
Çizelge 3.2: Çarpma olayı ile malzeme sınıflandırma deneyinde hata matrisi 26
Çizelge 3.3: Olay sınıflandırma sonuçları 26
Çizelge 3.4: Olay ve nesnenin birlikte sınıflandırıldığı sonuçlar 27
Çizelge 3.5: Çarpma ve düşme olayları ile malzeme sınıflandırma deneyinde hata matrisi..... 28
Çizelge 3.6: Tınısal öznitelikleri ile malzeme sınıflandırma..... 29
Çizelge 4.1: Pioneer 3-AT robotunu kullanarak çarpma olayı ile malzeme sınıflandırma 33
Çizelge 4.2: Robotik kol ile düşürme ve çarpma eylemleri için malzeme tanıma sonuçları..... 34
Çizelge 4.3: Robotik kol ile düşürme ve çarpma eylemleri ayırt etme sonuçları.... 34

ŞEKİL LİSTESİ

Sayfa

Şekil 2.1 : Sisteme genel bakış. 9
Şekil 2.2 : Sesin eşikleme işleminden önceki gösterimi. 9
Şekil 2.3 : Sesin sayısal değerlerinin mutlak değeri alınmış hali. 10
Şekil 2.4 : Sesin mutlak değeri alınmış ve pencerelenmiş gösterimi ve ses eşikleri. 10
Şekil 2.5 : DFT spektrumu. 12
Şekil 2.6 : Tanıma üzerinde FFT pencere boyutu ve FFT vektör sayısının etkisi. 13
Şekil 2.7 : Çarpma etkileşimindeki spektral vektörlerin normalize gücünün ortalaması ve standart sapması. 15
Şekil 2.8 : Düşme etkileşimindeki spektral vektörlerin normalize gücünün ortalaması ve standart sapması. 15
Şekil 3.1 : Deneylerde eğitim ve testlerde kullanılan nesneler. 22
Şekil 3.2 : Hassasiyet ve hatırlama 24
Şekil 4.1 : Pioneer 3-AT bir nesneyi düşürürken 32
Şekil 4.2 : a) Robotik kol bir nesneyi düşürürken, b) Robotik kol bir nesneye vururken. 34

SAHNE ANALİZİ İÇİN SES KAYNAĞI TESPİTİ

ÖZET

Ses işaretlerini yorumlayarak nesneleri ve olayları tanıma kabiliyeti insanın önemli yeteneklerinden biridir. Bu sayede insanlar bulundukları ortam hakkındaki farkındalıklarını artırarak, karar verme aşamasında bu bilgiler ışığında daha doğru hareketlere yönelirler. Örnek olarak, kapalı bir kutu içerisinden gelen cam sesi, kişiyi o kutuyu taşıma esnasında daha dikkatli olmaya sevk eder. Robotlarda ise nesneleri ve olayları çıkardıkları seslere göre sınıflandırma son derece önemli ancak oldukça zor bir görevdir. Bu tez çalışmasında, robotların ortamda tespit ettikleri sesin özelliklerini kullanarak nesneleri ve olayları otomatik olarak sınıflayabilecekleri bir sistem kapsamında ses verilerinin ayırt edici özelliklerini kullanan farklı gözeticili öğrenme yöntemleri kıyaslanmıştır. Çalışmalar kapsamında geliştirilen sistem temel olarak üç aşamadan meydana gelmektedir. Bunlar; önişleme, öznitelik çıkarma ve sınıflandırmadır.

Önişleme, sisteme mikrofon aracılığıyla aktarılan ses sinyalinin öznitelik çıkarma öncesinde gerekli eşikler ve dönüşümlerden geçtiği bölümdür. Eşikleme ses sinyalinin sistem açısından faydalı kısımlarını almak için yapılır. Eşiklemeden sonra ise ses ayrık Fourier dönüşümüne (DFT) tabi tutularak spektral özniteliklerin çıkarımı için hazırlanır. Öznitelik çıkarma işlemlerinde ise önişlemeden gelen sinyal ve verilerden yararlanarak sesin spektral ve tınısal öznitelikleri ile sınıflandırma algoritmalarına uygun öznitelik vektörleri hesaplanmaktadır. Bu vektörler hem eğitim hem de test aşamalarında aynı şekilde oluşturulmaktadır. Çalışmaların sınıflandırma bölümü eğitim ve test aşamalarından meydana gelmektedir. Sınıflandırma, öncelikle kaynağı bilinen sesleri ile sınıflandırma algoritmasının eğitilerek bir model oluşturması daha sonra ise kaynağı bilinmeyen sesin bu model yardımıyla doğru şekilde sınıflandırılmasından oluşmaktadır. Çalışmalarda destek vektör makinesi (SVM) ve çok katmanlı algılayıcı (MLP) olmak üzere iki adet makine öğrenme algoritması kullanılmıştır.

Sistem geliştirme ve seslerin özniteliklerini analiz etme amacıyla tez çalışmaları sırasında bir seri deneyler icra edilmiştir. Yapılan deneyler iki bölümde değerlendirilmektedir. Bunlar; 1) Laboratuvar deneyleri, 2) Robot deneyleri. İlk bölümdeki deneyler daha çok sistemin geliştirilmesi ve sesin öznitelik analizi amacıyla yapılmış, ikinci bölümdeki deneyler ise sistemin robotik sistemler üzerinde çalışma performanslarını ölçmek amacıyla yapılmıştır. Deneylerde kullanılmak üzere bir nesne seti oluşturulmuştur. Nesneler cam, metal, porselen, karton ve plastik gibi farklı malzemelerden imal edilmiş tipik ev eşyalarıdır. Bu nesnelerden ses çıkarmak amacıyla vurma ve düşme olmak üzere iki etki yöntemi kullanılmıştır. Çıkarılan sesler ayrıştırıcı özelliğe sahip öznitelikler açısından incelenmiştir. İncelemeler için seslerin

spektral ve tınısal öznitelikleri çıkarılmıştır. Bu özniteliklerden bazıları sınıflandırma algoritmaları için faydalı olurken, bazılarının ya hiç etkisi olmamış ya da olumsuz etkileri olmuştur.

Gözeticili öğrenme yöntemlerinin önerilen öznitelikler ile malzemeleri sınıflandırma başarımı analiz edildikten sonra sistemin düşme ve çarpma olaylarını ayırt edebilme becerisi de incelenmiştir. Deneyler için iki farklı robot platformu kullanılmıştır: tekerlekli gezgin bir robot ve yedi serbestlik dereceli bir robot kolu. Önerilen sistem, malzemeleri ve olayları sınıflandırmada yüksek sınıflandırma başarımı ile çalışmaktadır. Ayrıca sistem açık kaynak kodlu platformlar üzerinde geliştirilmeye ve farklı uygulamalarda kullanıma uygun şekilde geliştirilmiştir.

SOUND SOURCE IDENTIFICATION FOR SCENE ANALYSIS

SUMMARY

Humans are successful at categorization of objects using auditory signals and outcomes of events. The ability to categorize objects and outcomes of events using auditory signals is rather advanced in humans. Thus, by increasing the awareness of people about their environment, they make more accurate and correct decisions. For example, when a person hears glass-like sounds coming from a closed box, he may decide to carry it more carefully due to risk of fracture risk. The ability to categorize objects and events using auditory signals is extremely important, but it is a difficult task in robots. In this thesis, different supervised learning methods using distinctive features from sound data were compared as part of a system for robots to clasify objects and events automatically using auditory features of environmental sounds. The developed system consists of three fundamental processes: preprocessing, feature extraction, and classification.

A chain of elementary audio processing blocks constitute the preprocessing phase, which sequentially acquires the audio data from the microphone, creates audio windows (frames) and applies threshold for sound detection. Creating audio windows is quite important due to there can be sudden and unrelated changes within received sounds. By framing effects of these changes are suppressed and more robust preprocessing is provided. Thresholding process consists of two seperate thresholds. First one is onset threshold to detect active and useful sound. After that detection offset threshold is becoming active to detect end of this sound. Besides thresholding, in the preprocessing phase discrete Fourier transform is applied to obtain spectral representation of sounds. Fast Fourier transform (FFT) is used for DFT calculation. FFT algorithm is calculating exactly same results with DFT but it has lower complexity. There are there discrete but simultaneous processes are working after thresholding and discrete Fourier transform is application for feature extraction. First one is extracting features from the spectrum of sound, second one is extracting timbral features and third one is measuring length of detected sound. Timbral features consists of time-domain zero-crossings, spectral centroid, rolloff, flux and mel-frequency cepstral coefficients (MFCC) features. These feature extraction processes are applied exactly the same way in both training and testing sessions.

Training and testing sessions constitute classification section of system. Training is part that machine learning algorithm creates a model from traning set to classify test set which will be given later. Machine learning algorithm is trained with labeled sounds and creates a model using information about these sounds. Then in the testing session, machine learning algorithm is used that model to classify test sets. Testing mostly repeated several times in the experiments using same model due to obtain

best results in ideal situtation. In the classification session two widely used machine learning algorithm is selected to use. These are Support Vector Machine (SVM) and Multi-Layer Perceptron (MLP). In the experiments it is shown that at some level MLP became insufficient due to processing time length. From that moment, only SVM algorithm is used.

In the course of this of thesis, series of experiments are performed in order to develop the system and analyze features. Experiments are evaluated in two sections. These are; 1) Laboratory experiments, 2) Robot experiments. In the first section of experiments, the main purpose is to develop a system and analyze the performance od features of sounds. In the second section of experiments, the system is applied to the robots to the measure performance of system in the real situations. A set of objects is collected to be used in the experiments. This set of objects consists of common household objects. Objects are made of five different materials and every material category comprises five different objects. There is five different materials, *glass, porcelain, plastic, metal* and *carton*. A total of 25 different objects are used in the experiments. The material of the objects are used as a class name indicator. Besides the set of objects, two different events are used to generate sounds out of objects. *Drop* and *hit* events are used not only to make sound out of objects but also as a label of sound. By doing that, a pair of labels is obtained. Five different object materials with two different events make 10 different class labels for classification. Received sounds are analyzed to determine the separative features. For analyzing, spectral and timbral features are extracted from detected sounds. Amoung related researches, there is no study using the duration of sound as a feature. In this thesis, it is found that duration of sound can be a distinctive feature between not only materials but also events. On the other hand, timbral features of sounds are useful for many other aplications, but these applications are mostly music related. In this system, another use of timbral features is investigated. Besides spectral and duration features, timbral features extract and add to feature vector to increase the success of classification. Contrary to expectations timbral features did not contribute to the classification process.

For classification two different machine learning algorithms are used. These algorithms are widely accepted by the community to classify the unlabelled sounds. The first method called Support Vector Machine (SVM) is primarily a method that performs classification tasks by constructing hyperplanes in a multidimensional space that separates cases of different class labels. More spesifically, the goal of the SVM is to train a radial basis function and classify labelled sounds. Second learning method is Multi Layer Perceptron (MLP). An MLP algorithm is a feed-forward artificial neural network model that maps sets of input data onto a set of appropriate outputs. An MLP consists of multiple layers of nodes in a directed graph, with each layer fully connected to the next one. Except for the input nodes, each node is a neuron (or processing element) with a nonlinear activation function. MLP utilizes a supervised learning technique called back-propagation for training the network.

The laboratory and robot experiments were performed using a computer equipped with an Intel Core @2.40GHz CPU, 16GB RAM and Sony ECM-C10 microphone. The microphone is used to capture audio signals at a standard 32-bit resolution and 16 kHz sampling rate. The audio processing chain including the traning and test phases runs on HARK, which is an open-sourced robot audition software. After the system

is developed by performing laboratory experiments, system is applied to the robots. Two robots are used in the robot experiments. First one is Pioneer 3-AT, a four-wheel mobile robotic platform. Using this robot only drop sounds are tested. With this test simulation of fault detection is performed. The second robot is a 7-DOF robotic arm. With robotic arm both drop and hitting sounds were recorded and tested. The work demonstrated a feasibility study of a system which is able to discriminate material of objects and events from each other with a high accuracy. The system can be used on different cognitive robots to analyze the environmental changes and detect failures. As a result of all experiments, system proved that detection and classification of objects and events is highly successful by using auditory features.

1. GİRİŞ

Sahne analizi ve sürekli izleme karmaşık ortamlarda bilişsel robotların güvenli görev icrası için çok önemlidir. Bir çok uygulamada, robotların farklı nesneler ile etkileşime girmesi gerekir ve malzemenin önceden bilinmesi robotun temas kuracağı cisme yaklaşım şeklini belirler. Ancak, nesnelerin yapıldığı malzemeler ile ilgili ön bilgi her zaman mevcut olmayabilir veya böyle bir ön bilgiyi edinmek, kaydetmek ve işlemek mümkün olmayabilir. Bu gibi durumlarda, robotların bu nitelikleri belirlemesi için görsel, koku, işitsel veya içsel (proprioseptif[1]) algılayıcı sistemler gerekir. Görme, işitme, koklama, dokunma gibi farklı duyu kipleri malzeme sınıflandırmaları için kullanılabilir [1–5].

Ses tabanlı izleme nesnelerle etkileşim sırasında harici ses verileri yorumlamak için süreçleri içerir. Yayılan seslerin kaynağı temelde nesnelerin titreşimidir. Farklı malzemelerden yapılan nesneler fiziksel yapılarına göre farklı ses özelliklerine sahiptirler. Bu hem sınıflandırmada, hem de sınıflandırma hataları düzeltmede işitsel sahne analizi yapan sistemlere yardımcı olur. Bu nedenle, ses kipi robotların çevre algılaması (environment perception) ve olay sonuçlarını izleyebilmesi (action monitoring) için önemli bir girdi sağlar. Bilgisayarlarla görsel tekniklerinin aksine basit bir mikrofon ile bile 360 derecelik çevresel bir kaplama sağlanabilir. Hatta çok kanallı mikrofon veya birden fazla mikrofon kullanarak ses kaynağının yerini belirlemek bile mümkündür [6].

Bu çalışmada robotların çalıştıkları sahneyi analiz etmesi için bir sistem yaratmayı hedefledik. Robotlar çevreyi analiz ederken, etrafını sürekli olarak izlemesi ve işitsel sinyalleri değerlendirmesi amaçlanmaktadır. Bu sistem ile robotlar sadece nesnenin malzemelerini değil, aynı zamanda meydana gelen olayları ayırt etme yeteneği kazanmaktadır. Deneylerde robotun nesnelerin malzemelerini ayırt etmesi için 25

[1]Propriosepsiyon; eklemlerin boşluktaki pozisyonunu, konumunu, hareketini algılama duyusudur. Proprioseptif bu duyu ile ilgili olan demektir (İngilizce: Proprioceptive).

tipik ev eşyası kullanıldı. Farklı malzeme tipleri nesneler kümesini oluştururken, düşme ve çarpma ise olay olarak nitelendirdirilen sesin çıkmasına sebep olan etkileşimler kümesini oluşturmaktadır. Sesler Pioneer 3-AT gezgin robot ve robotik kol kullanarak düşme ve çarpma olayları neticesinde oluşturuldu. Aynı maddeden yapılmış nesnelerin ses benzerliklerinin yüksek olduğu varsayımından yola çıkılarak öncelikle bu benzelikleri oluşturan özniteliklerin sayısal olarak tespidi ve analizi yapıldı. Yapılan deneyler sonucunda aynı malzemelerdeki ses özniteliklerinin benzerlikleri ile farklı malzemelerdeki ayrışımları ortaya konuldu. Daha sonra ise deneylerde düşme ve çarpma olayları ile oluşturulan sesleri anlık olarak işleyip özniteliklerini çıkaran, bu çıkarılan öznitelikler ile makine öğrenmesi yöntemleri kullanarak bir model oluşturup yeni gelen sesleri sınıflandırabilen bir sistem yaratıldı. Bu sistem önce insan eli ile yapılan deneylerde daha sonra ise robotlarla yapılan deneylerde kullanılarak nesnelerin malzemelerini, düşürme ve çarpma olayları sırasında yüksek bir performans ile ayırt edelebileceği gösterildi.

1.1 Tezin Amacı

Bu tez çalışmasında aşağıda belirtilen hedeflere ulaşılması hedeflenmiştir.

1. Çeşitli malzemedeki cisimlerden elde edilen seslerin incelenerek ayırt edici özniteliklerin araştırılması ve belirlenmesi,
2. Belirli malzemelerden yapılmış cisimlerden ses elde etmek için farklı olaylar kullanmak ve bu yöntemlerle elde edilen seslerin incelenerek olaylar arası ayrıt edici özniteliklerin araştırılması ve belirlenmesi,
3. Belirlenen özniteliklerin makine öğrenmesi algoritmalarında en yüksek sınıflandırma performansı ile çalışması için gerekli optimizasyonların yapılması,
4. Seçilen makine öğrenmesi algoritmaları ve öznitelikler ile canlı testlerde yeterli olacak hızda sınıflandırma yapabilen bir sistemin geliştirilmesi,
5. Tüm ses işleme ve sınıflandırma işlemlerinin canlı olarak robotik sistemler üzerinde test edilmesi ve yüksek malzeme sınıflandırma yüzdesi elde edilmesidir.

1.2 İlgili Çalışmalar

Nesneleri tanımak için bir çok farklı teknik bulunmaktadır. Bunlardan bazıları kamera kullanımı ile [1–3, 7], lazer ile [8, 9], koku ile [5], dokunsal etkileşim ile [3, 4] veya ses öznitelikleri ile [3, 4, 8, 10–12] uygulanmaktadır. Bu tekniklerden ses ve lazer hariç hiç biri 360 derecelik bir kaplama alanı sağlayamaz. Hatta lazer bile kullanılan cihaz ile bağlantılı olarak çoğu zaman 360 dereceye kadar kaplama alanına erişememektedir [8]. Bu nedenle ses veya lazer tekniklerinden birini kullanmak çevre algılaması için fazladan kaplama alanı sağlamaktadır. Bunun yanı sıra ses küresel bir kaplama alanına sahiptir. Sesin hem bu özelliği hem de nesnelerin arkasına dahi yayılabilmesi, sadece görüş alanı ve mesafesinde algılama yapabilen diğer algılama tekniklerine göre daha etkin ve etkili bir yöntem olduğunu gösterir. Yapay yaratılan sesler hariç bir nesnenin titreşimi sebebiyle oluşan seslerin o nesneye ait özellikler barındıracağı açıktır. Sesin bu özelliğinden yararlanarak robotik sistemler üzerinde ses ile görev icra edilen sahnelizi bir çok çalışmaya konu olmuştur [3, 4, 8, 10–12].

İncelenen çalışmalar arasında en az rastlanan nesne sınıflandırma yöntemi koku ile sınıflandırmadır. İsveç Örebro Üniversitesinde yapılan bir çalışmada [5] elektronik burun olarak nitelenen bir seri gaz sensörü ile nesneler tanınmaya ve sınıflandırılmaya çalışılmıştır. Önerilen sistem bir oda içerisinde dolaşan bir gezgin robot üzerinde uygulanmıştır. Gezgin robot hareketi esnasında yüksek seviye ve güçlü kokular yayan nesnelerin kokularını değişik uzaklıklardan analiz etmektedir. 5 farklı koku üzerinde yapılan farklı deneylerde önerilen sistem %90'lık bir başarıma ulaşabilmiştir.

Birçok nesne tanıma uygulaması genel olarak iki farklı makine öğrenmesi yöntemi kullanılmaktadır. (1) nesnelerin makine öğrenmesi yöntemleri kullanılarak sınıflandırıldığı denetimsiz yöntemler [13] (örnek; k-Ortalamalar[2], hiyerarşik kümeleme[3], vb.) (2) nesnelerin etiketlenerek bir tanıma modeli geliştirilmek üzere kullanıldığı ve daha sonra bu model ile yeni gelen farklı nesnelerin tanındığı denetimli yöntemler (örnek; destek vektör Makinesi (SVM), karar ağacı[4]). İncelenen çalışmalarda çoğunlukla denetimli makine öğrenmesi yöntemleri kullanıldığı görülmüştür.

[2]İngilizce: k-Means.
[3]İngilizce: Hierarchical Clustering.
[4]İngilizce: Decision Tree.

Bu yöntemler arasında ise destek vektör makineleri diğer yöntemlere göre daha öne çıkan ve sıklıkla kullanılan bir algoritmadır. Bu çalışmalar ışığında bu tez çalışmasında da denetimli makine öğrenmesi yöntemlerinin kullanılmasının uygun olacağı değerlendirilmiştir.

Sinapov'un çalışmasında [3] 20 farklı kategoriden 100 farklı nesne seçilmiştir. Bu nesneleri tanımlamak için robot nesneler üzerinde kavrama, kaldırma, tutma, sallama, düşürme, dürtme, itme ve basma olmak üzere 9 farklı interaktif davranış kullanılarak oluşan seslerden 6 farklı öznitelik çıkarımında bulunmuştur. Bu öznitelikler robotik kolun ve elin propriyoseptif öznitelikleri, hızlandırılmış sağlam öznitelikler (SURF), işitsel, renk, optik akış özniteliklerdir. Ses öznitelikleri için sadece ayrık Fourier dönüşümü uygulanması ile öznitelik çıkarımı yapılmıştır.

Bazı çalışmalarda çocuklar için uygulanan eğitim yaklaşımları yapay zeka uygulamalarında da kullanılmıştır. Örneğin Schenck'in çalışmasında robota nesneleri öğretebilmek için Montessori testi uygulanmıştır. Montessori eğitimi çoklu yaş gruplarına hitap etmesi, özel eğitim malzemeleri, öğrenci seçimi ile uzun süreli eğitim blokları üzerinde çalışma ve işbirliği özellikleri ile bilinen bir eğitim sistemidir [14]. Çalışmada 7 serbestlik dereceli bir robotik kola sahip insansı robot kullanılmıştır. Öznitelik çıkarma işlemi ve deneyler robotik kol vasıtasıyla 10 faklı hareket ile yapılmıştır. Yapılan deneyler sırasında proprioseptif ve ses öznitelikleri kaydedilmiştir. Nesneleri birbiriyle eşlemek hedefiyle yapılan sistem oldukça başarılı sonuçlar vermiştir. Özellikle ses kutularını eşlemede başarım %100'e ulaşmıştır.

Sinapov'un diğer bir çalışmasında [11] kupa bardak, şişe ve oyuncakları da içeren 5 farklı malzemeden imal edilmiş 25 tipik ev eşyası üzerinde, 5 farklı robotik davranış kullanılarak benzer deneyler yapılmıştır. Aynı zamanda kullanılan bazı eşyalar içine hareket ettiği zaman ses çıkarması için ufak nesneler koyarak daha fazla ayırt edilebilirlik sağlanması amaçlanmıştır. Yapılan deneyler sırasında proprioseptif ve ses öznitelikleri kaydedilmiştir. Ses öznitelikleri diğer birçok çalışmalarda olduğu gibi ayrık Fourier dönüşümü ile çıkartılmıştır. Çalışma sonucunda malzeme tanıma başarımı ortalama olarak %84.06 olmuştur. Aynı testler kullanılan nesnelerin içerisinin boş olup olmamasına göre de sınıflandırmak için icra edilmiştir. Bu testlerde ise 5 farklı robotik davranışın başarım ortalaması %95.78 olarak gerçekleşmiştir.

Ses öznitelikleri kullanarak nesnelerin yanı sıra bulunan ortam ve çevre koşullarını sınıflandırma üzerine de çalışılmıştır [15]. Bu çalışmada da ayrık Fourier dönüşümü hesabında kullanılan FFT algoritması ses öznitelikleri çıkarmada kullanılmıştır. Bunun ile birlikte çalışmalarda, zamansal, istatistiksel ve frekans olmak üzere 3 kategoride 62 farklı öznitelik elde kullanılmıştır. Yaklaşık 160,000 ses örneği üzerinde oluşturulan model ile 10 farklı çevrede yapılan deneylerde ortalama %91.42 tanıma başarımına erilmiştir.

Ses ve sesin özniteliklerinin kullanıldığı başka bir alan ise sesin kaynağı olan olayı tanıma ve ayırt etmedir. Dennis'in [16] bir çalışmasında sesin spektral öznitelikleri kullanılarak sesin kaynağı olan olaylar tespit edilmiştir. Çalışmada kısa zamanlı Fourier dönüşümüne tutularak sesin spektrogramı çıkarılmakta daha sonra ise bunun üzerinden öznitelikler belirlenmektedir. Deneylerde farklı olaylar ve bunların birbiri üzerinde tanımaya olan etkileri araştırılmıştır.

Araştırmalar sırasında nesneleri tanıma ve sınıflandırma uygulamalarında genellikle ses özniteliklerini elde etmek için tek kanallı bir mikrofon kullanılıdığı görülmektedir [3, 4, 10, 11]. Fakat nesne tanımanın yanı sıra sesin kaynağının konumu tespit edilmek istediğinde ise bir mikrofon dizisi ile bu yapılabilmektedir. Sesin kaynağının bu şekilde tespit edilebilir olması robotik sistemler açısından büyük bir avantaj sağlamaktadır. Özellikler değişken ortamlarda iş yapan veya gezgin robotlar sesin bu özelliğini kullanarak sahne analizi başarımını yüksek ölçüde arttırabilir [17–20].

Çeşitli çalışmalar [3, 4, 10, 11] göstermekte ki nesneleri tanımak ve sınıflandırmak için ses özellikleri ile tanıma oldukça etkili bir yöntemdir. Bu yöntem ile robotlar çevresel farkındalıklarını artırabilir ve görevlerini daha yüksek bir başarı oranı ile yerine getirmeleri sağlanabilir. Robotların etraflarında meydana gelen bir hatayı tespit etmeleri, çevresel dinlemenin ve nesneleri tanımanın getirdiği en büyük kazanımlardan biridir. Bütün bu çalışmalar [3, 4, 10, 11] ses sinyalinden Ayrık Fourier Dönüşümü (DFT) ile özellik çıkarımı yapmakta ve bunu öznitelik vektörü olarak kullanmaktadır.

1.3 Uygulama

Bu çalışmalar kapsamında öncelikle testler ve analiz aşamalarında kulanılacak nesneler ve malzeme çeşitleri belirlendi. Malzemeler özellikle insan kulağınında ayırt edebileceği nesnelerden seçilerek gerçek dünya ile karşılaştırılması amaçlandı. Deneyler için seçilen nesnelerin tamamı insan kulağı tarafından da ayırt edilebilir malzemelerden yapılmıştır. Fakat tez çalışmasına konu olan problemi biraz daha zorlaştırmak ve sistemi daha zor koşullarda denemek adına bazı benzer seslere sahip malzemeler de eklenmiştir. Özellikle cam ve porselen sesleri özellikle insan kulağı için ayırt etmesi zor olan malzemelerdendir.

Sistem geliştirme aşamasına gelmeden öncelikle seslerin analizi yapıldı. Çeşitli malzemelerden elde edilen sesler Matlab[5] ile incelenerek frekans ve zaman alanında ayırt edilebilirliği, olması gereken başlangış ve bitiş eşik değerleri, ortalama ve toplam gücün zamansal değişimi gibi konularda analizler yapıldı. Analizler sonucunda kullanılacak malzemeler, sesin oluşumunda kullanılacak olaylar, kullanılacak başlangıç ve bitiş eşikleri, başlangıç olarak kullanılacak ve incelenecek öznitelikler belirlendi.

Sesin oluşumunda kullanılacak olayların seçiminde iki unsura dikkat edildi. Birincisi, laboratuvar deneylerinde test edilen hareketlerin benzer şekilde robotlar tarafından da uygulanabilir olması. İkinci unsur ise, hareketlerin malzemelerden yeterli seviyede ses çıkarabilecek olması. Örnek vermek gerekirse kullandığımız malzemeler itibariyle nesneyi sıkarak ses çıkarmak mümkün olmayacaktır.

Matlab ile yapılan ses analizinden sonra sistem kurulumuna geçilmiştir. Sistem ROS ile entegre olarak çalışabilen açık kaynak kodlu HARK[6] sistemi üzerinde geliştirilmiştir. Bu şekilde sistem kolaylıkla ROS ile yönetilen robotik sistemlere aktarılabilir halde geliştirilmiştir.

[5]MATLAB Sürüm 2014a, The MathWorks, Inc., Natick, Massachusetts, Amerika Birleşik Devletleri.

[6]http://winnie.kuis.kyoto-u.ac.jp/HARK/

Daha sonra Bölüm 2.2'de ayrıntılı olarak anlatılan öznitelik çıkarma işlemleri gerçekleşmiştir. Bu tezde Ayrık Fourier Dönüşümünün yanı sıra sesin uzunluğu bir öznitelik olarak kullanılmıştır. Sesin uzunluğu özniteliği, düşme ve çarpma seslerini ayırmada büyük fark yaratmıştır. DFT ve ses uzunluğu özelliklerinin yanı sıra sesin tınısal[7] özellikleri de (zamansal sıfır geçiş oranı, spektral ağırlık merkezi, azalış, akış, mel ölçekli kepstrum katsayıları (MFCC)) test edilmiştir. Fakat bu özellikler robotik sistem uygulamasında bir katkı sağlamamıştır.

Tezin sınıflandırma aşamasında iki adet makine öğrenmesi yöntemi kullanılmıştır. Bunlar Destek Vektör Makinesi (SVM) ve Çok Katmanlı Algılayıcıdır (MLP). Laboratuvar deneylerindeki sınıflandırmaların tamamı WEKA ile yapılmıştır. MLP algoritması WEKA programı dahilinde olan makine öğrenmesi algoritmalarındandır. SVM içinse yine WEKA için LIBSVM [21] kütüphanesinin Java ile yazılmış eklentisi kullanılmıştır. Canlı testlerde ise özellikle MLP'nin işlem süresinin uzun olması nedeniyle sadece SVM kullanılmaya başlanmış ve daha hızlı bir erişim sağlamak ve sınıflandırma algoritmasını güçlendirmek için C++ programlama dili ile kodlanmış LIBSVM kütüphanesi sisteme dahil edilmiştir.

Aynı zamanda benzer bir çok çalışamada da [3, 4, 10, 11] robotik kol kullanılmıştır. Bu tezde ise ses ile çevre analizi seyyar robotlara taşınmıştır. Bu önemli uygulama göstermiştir ki bu tür sistemler sadece robotik kollarda değil daha karmaşık robotlar da kullanılabilir.

[7] İngilizce: Timbral.

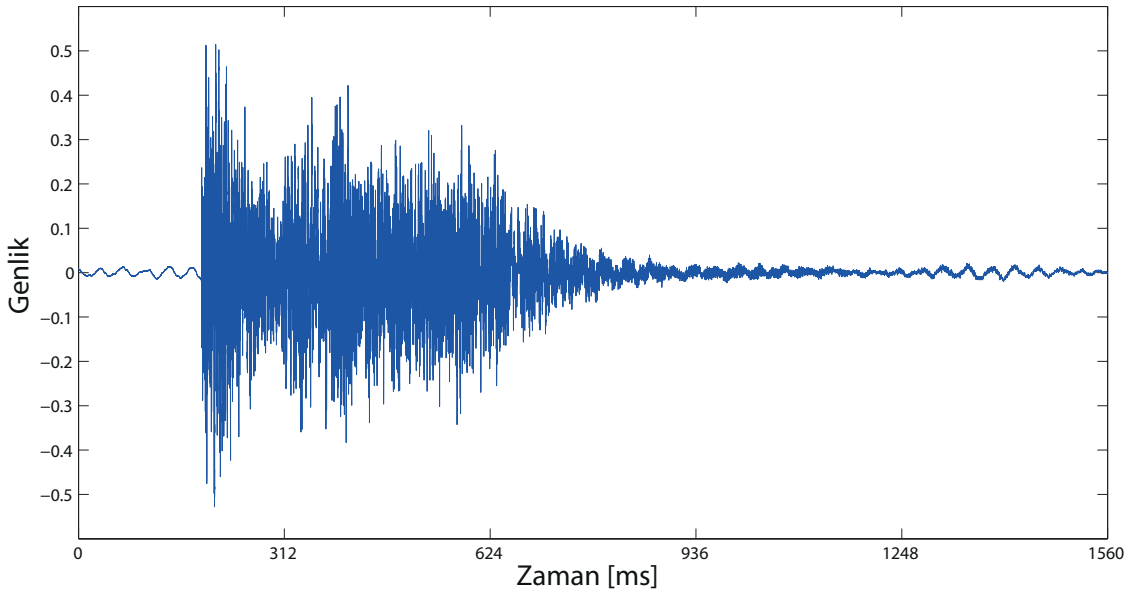
2. SİSTEME GENEL BAKIŞ

Bu bölümde Şekil 2.1'deki akış şemasında gösterilen ses işleme sisteminin temel mimarisi özetlenmektedir. Önerilen sistem üç temel ardışık süreçten oluşmaktadır. Bu süreçler: Önişleme, öznitelik çıkarma ve sınıflandırma şeklindedir.

Şekil 2.1: Sisteme genel bakış.

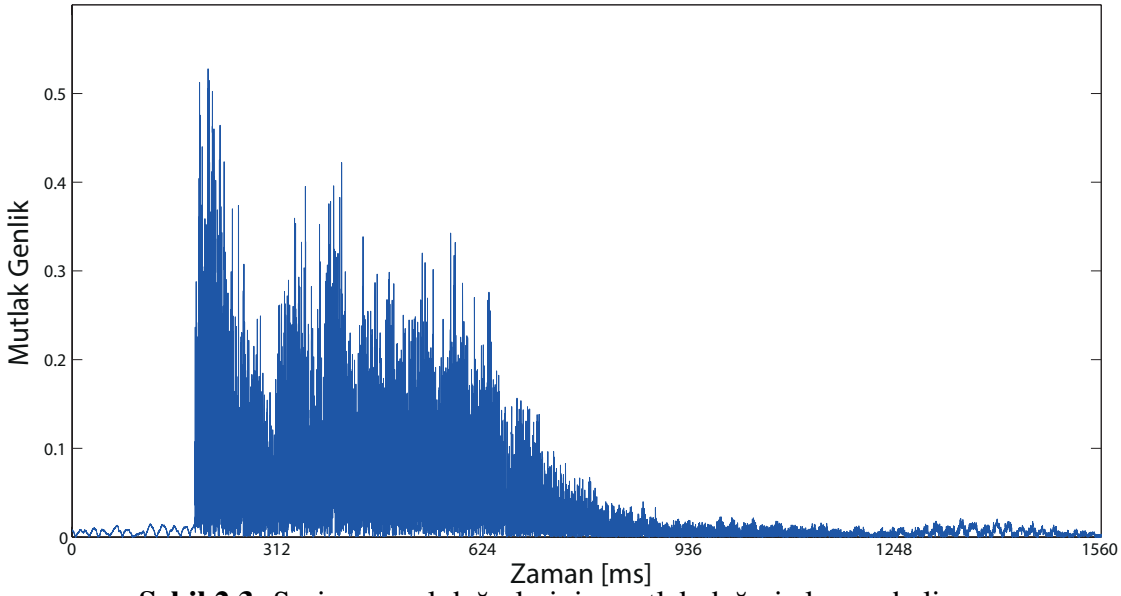
2.1 Önişleme

Bu safhada zamansal alandaki bir sinyal olarak alınan sesin önce başlama ve bitiş eşik değerleri arasında kalan kısmı alınıp daha sonra FFT ve süre tespiti yapılmaktadır.

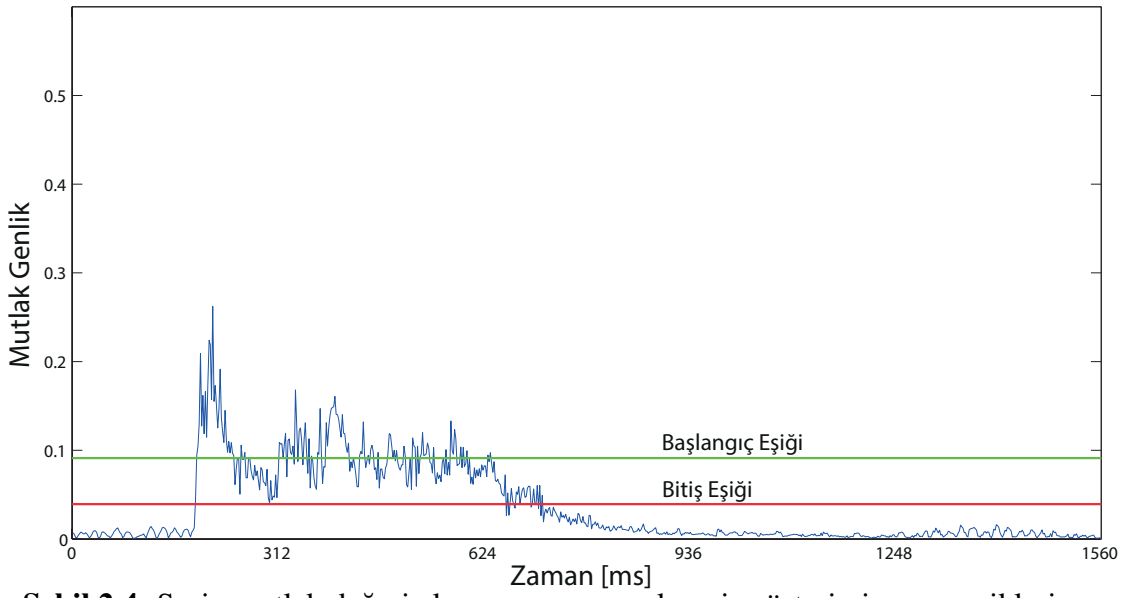


Şekil 2.2: Sesin eşikleme işleminden önceki gösterimi.

Sistem başlangıç anından itibaren sürekli olarak bir dinleme yapmaktadır. Deney yapılan ortamlarda ve özellikle gerçek dünyada mutlak sessizlik mümkün olmadığı ve çoğu zaman ufak başka seslerin ve gürültülerin varlığından dolayı ses bir eşikleme



Şekil 2.3: Sesin sayısal değerlerinin mutlak değeri alınmış hali.



Şekil 2.4: Sesin mutlak değeri alınmış ve pencerelenmiş gösterimi ve ses eşikleri.

işlemine tabi tutulur ve faydalı ses tespit edilir. Sesin eşikleme işleminden önceki durumunun grafiksel gösterimi Şekil 2.2'de görülmektedir. Gösterildiği şekilde elde edilen sesin öncelikle mutlak değeri alınarak negatif değerler pozitife çevrildi. Şekil 2.3'de mutlak değeri hesaplanmış ses gösterilmektedir. Daha sonra ise sese pencereleme yapılarak eşikleme işlemine hazır hale getirildi. Sesin o penceredeki mutlak kuvveti başlangıç eşik değerini aştığı zaman sistem bir ses tespit ettiğini algılar. Bu andan itibaren yine pencereleme yapılarak sesin bitiş eşik değerinin altına inmesi beklenir. Bu sırada gelen tüm sesler dikkate alınır. Şekil 2.4'de görüldüğü

üzere bitiş eşik değeri başlangıç eşik değerinden daha alçak bir değerdir. Bunun sebebi sesin süresini daha doğru tespit edebilmek ve başlangıçta herhangi bir ufak gürültünün eşik değerini aşarak işleme sokulmasının önüne geçmektir. Bu şekilde nesne ile her etkileşim sonucu ortaya çıkan işitsel işaretler birbirinden ayrık olarak temsil edilmektedir.

2.1.1 Başlangıç ve bitiş ses eşiklerinin tespiti

Başlangıç ve bitiş ses eşikleri her deney için ayrı olarak deneysel olarak tespit edilmiştir. Tespit için öncelikle ortamın sessizlik anında mikrofon tarafından akratılan sinyal seviyesi en alt seviye olarak kabul edilmiştir. Daha sonra en yüksek ses gücüne sahip olan malzeme olan metal ve en alçak ses gücüne sahip olan malzeme karton ile düşme ve çarpma sesleri oluşturularak incelenmiştir. Başlangıç eşiği bu seslerden her birini tespit edecek fakat hiç bir ortam gürültüsünün geçemeyeceği seviyede bir değer olarak belirlenmiştir. Daha sonra ise bitiş eşik değeri tüm seslerin gürültü ile genlik olarak aynı seviyeye yaklaştığı noktada, gürültü genliğinden yüksek bir değer olarak belirlenmiştir.

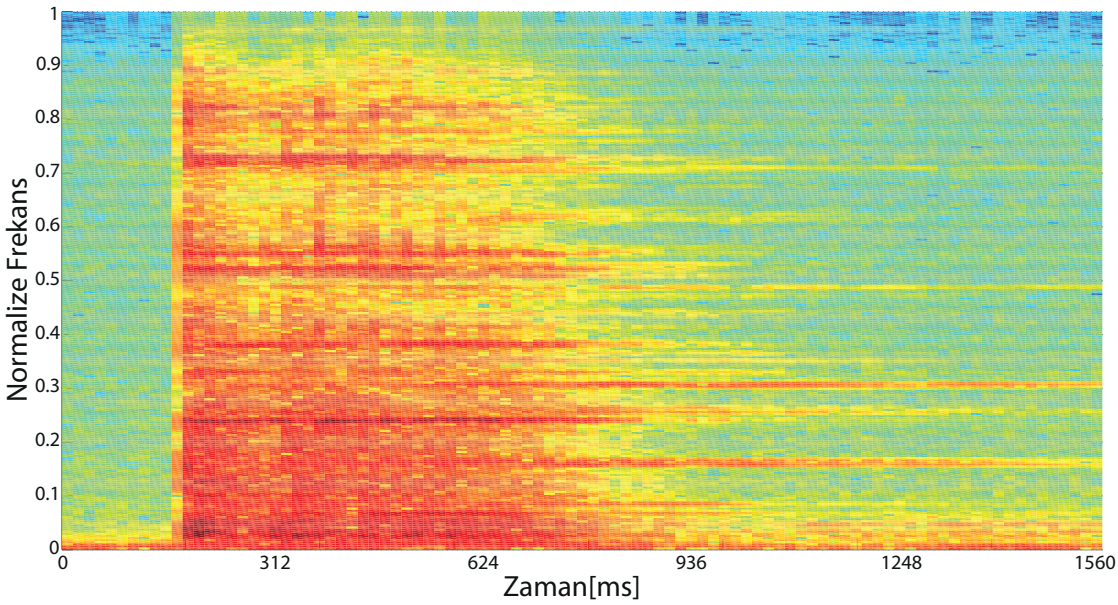
2.1.2 Ayrık Fourier Dönüşümü (DFT)

Eşikleme işlemi ile zamanda ayrıştırılarak sesler Ayrık Fourier Dönüşümü'ne tabi tutulmaktadır. DFT ile sesin zaman alanındaki sinyali eşdeğer frekans alanındaki sinyale dönüştürerek daha anlamlı bir gösterim elde edilmiştir. Aşağıdaki formülde X sinyal zaman serisi x sinyalinin frekans alanındaki gösterimi, k frekans bileşeni, N, x sinyalindeki toplam örnek sayısı, x zaman serisi sinyali, n zaman alanındaki örneğin sıra numarası ve i hayali birim değeridir.

$$X_k = \sum_{n=0}^{N-1} x_n e^{-i\pi k \frac{n}{N}} \tag{2.1}$$

DFT işleminin karmaşıklığı $O(n^2)$'dir. Bu işlem için karmaşıklığı $O(nlogn)$ [22] olan ve DFT ile tamamen aynı sonucu veren Hızlı Fourier Dönüşümü algoritması kullanılmıştır. Şekil 2.5'te, Şekil 2.2'deki ses sinyaline FFT uygulanarak elde edilen spektrum gösterilmektedir. Çalışmada FFT algoritması 512 örneklik bir çerçeve ve

256 örneklik bindirme ile uygulanmıştır. Şekil 2.5'de görüldüğü üzere 0 ile 1 arasına normalize edilmiş frekans değerlerinin güçleri gösterilmektedir. Renkler, kendisine y ekseninde karşılık gelen frekansın x ekseninde karşılık gelen andaki güç seviyesini göstermektedir. Bazı frekansların diğerlerine göre daha yüksek bir güce sahip olduğu ve bu gücü zamanla birlikte diğer frekanslara göre daha uzun süre korudukları gözlenmektedir. Bunun gibi sesin karakteristik olacabileceği düşünülen öznitelikleri sonraki aşamalarda analiz edilmiştir.



Şekil 2.5: DFT spektrumu.

Şekil 2.5'de 0 ila 160. milisaniyeler arasında alçak frekanslarda yüksek enerji bulunmasına rağmen, yüksek frekanslarda enerji miktarının az olduğu görülmektedir. Bunun sebebi Şekil 2.2'deki ses sinyalinin aynı zaman dilimine denk gelen kısımlarında da açıkca görülen belirgin bir sesin olmaması ve çoğunlukla ortam gürültüsünün sürekli olarak varlığını korumasından ileri gelmektedir.

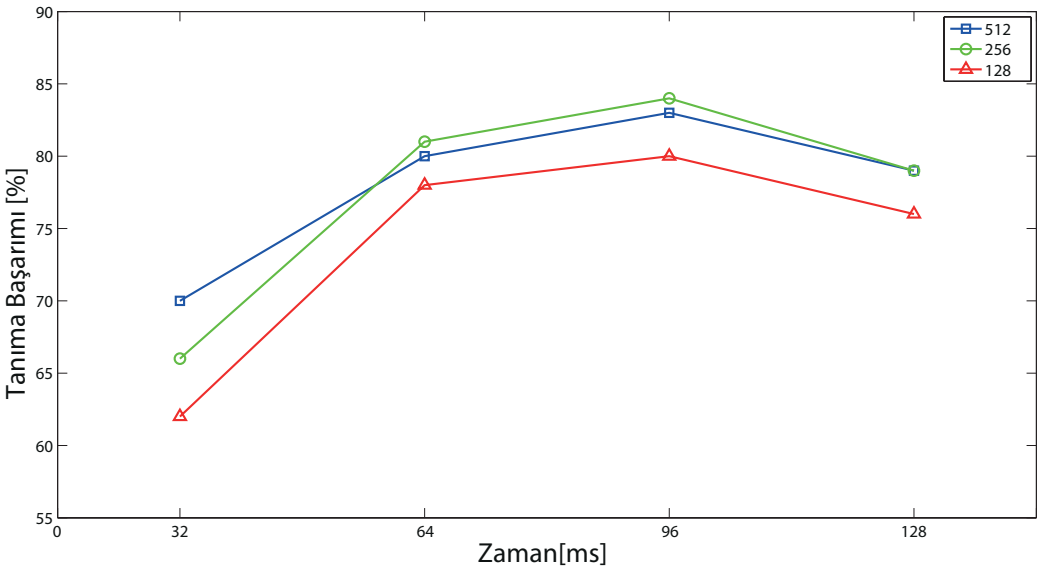
2.2 Öznitelik Çıkarma

Ses sınıflandırmada kullanılmak üzere tez çalışmaları boyunca farklı ses öznitelikleri çıkarılmış ve sistem üzerinde başarımları deneysel olarak araştırılmıştır. İlk olarak literatür araştırmaları sırasında çoklukla uygulanan bir öznitelik olduğu görülen FFT

öznitelikleri ile başlanmıştır. Daha sonra ise ses uzunluğu ve zamansal sıfır geçiş oranı gibi çeşitli ek öznitelikler ile başarımın arttırılması amaçlanmıştır.

2.2.1 Spektral band gücü özniteliği

Algılanan her bir faydalı ses için ayrı ayrı FFT hesaplanarak öznitelik vektörü oluşturulmuştur. Hesaplama için 512 örneklik bir çerçevede 256 örneklik bir bindirme tercih edilmiştir. İdeal çerçeve işleme boyutlarına Şekil 2.6'de görüldüğü üzere sistemin malzeme tanıma yüzdeleri incelenerek karar verilmiştir. Çerçeve büyüklüğü, frekans çözünürlüğü ile doğru oratılı, çıkarılan vektör sayısı ile ters orantılıdır. Referans çalışmalar baz alınarak [3, 10, 11] ve deneysel olarak incelendiğinde 512 örneklik bir çerçeveden doğan spektrum $2^8 + 1 = 257$ frekans bandına ayrılır. 257 adet freakans bandının sesin frekans ekseninde gücünün gösterimi için uygun bir çözünürlük olacağı görülmüştür. Yine algoritmada kullanılan 256 örneklik bindirme ile FFT uygulanan çerçeveler arası zamansal bazda bağın kopmamasına ve her bir yeni çerçevenin bir öncekinden gelen frekans niteliklerini bir kısmını taşıması amaçlanmıştır.



Şekil 2.6: Tanıma üzerinde FFT pencere boyutu ve FFT vektör sayısının etkisi.

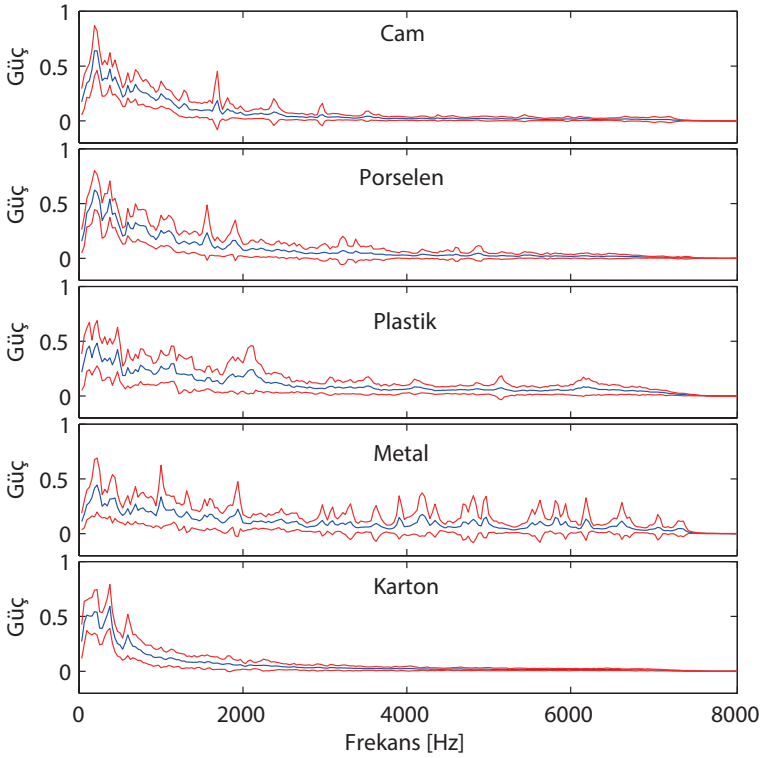
FFT algoritması bahsedilen bu ayarlarla işletildiği zaman 16kHz örnekleme hızı ile kaydedilen bir ses sinyali için 1 saniyelik faydalı ses için 62 adet FFT vektörü hesaplamaktadır. Şekil 2.5'de de görüldüğü gibi sesin genliği zamanla sönümlenmekte ve malzemeden gelen sesin genliği, gürültü ile aynı seviyeye yaklaşarak o malzemee

ait olan niteliklerin bastırılmasına sebep olmaktadır. Ayrıca kapalı ortamlarda sesin yankılanması nedeniyle malzemeden düşme ve çarpma olayları neticesinde oluşan ses çok kısa sürede aynı noktaya tekrar ulaşarak mikrofon üzerinden tespit edilen seste bozulmalara sebep olacaktır. Bu nedenlerle hesaplanan FFT vektörlerinden zamansal olarak önceki çerçevelere ait olan vektörler seçilmelidir. Bu seçilen vektörlerin denk geleceği zaman aralığı hem sesin malzemeye özgü özelliklerini barındıracak kadar uzun, hem de sayılan olumsuz etkenlerden etkilenmeyecek kadar kısa olmalıdır. Bu seçimi yaparken deneysel yöntemler kullanılmış ve FFT vektörünün çıkartıldığı zaman uzunluğunun tanıma üzerindeki etkisi araştırılmıştır. Şekil 2.6'de bu deneylerin sonuçları görülmektedir.

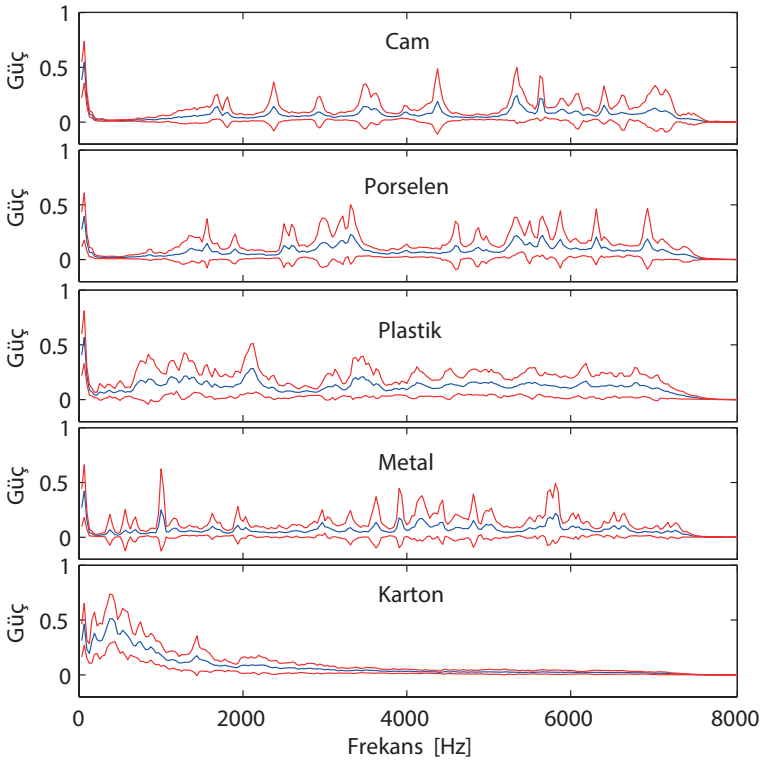
Deneyler sonucunda faydalı sese ait ilk 96ms'nin en yüksek tanıma yüzdesine sahip olduğu tespit edilmiştir. Bu nedenle 512 örneklik bir çerçevede 256 örneklik bir bindirme değerleri ile sesin başladığı andan itibaren hesaplanan ilk 5 adet FFT vektörü öznitelik belirlemek için seçilmiştir. Seçilen 5 vektörün zamansal doğrultuda ortalamaları alındıktan sonra aşağıdaki formülde gösterildiği gibi maksimum güç ile normalize edilerek o malzeme sesine ait öznitelik vektörü elde edilmiştir. Formülde F_{norm} normalize edilmiş yeni öznitelik vektörünü, k FFT vektöründeki nokta numarası, N vektörün uzunluğu, F_{max} vektörün en büyük değerini, $F[k]$ k noktasına denk gelen genlik değeridir.

$$F_{norm} = \frac{\sum_{k=1}^{N} kF[k]}{F_{max}} \tag{2.2}$$

Şekil 2.7 ve 2.8'de beş farklı nesne kategorisinin FFT vektörlerinin ortalamaları ve standart sapmaları gösterilmektedir. Bu vektörler her bir kategori için toplanmış 5 adet nesne ile beş kez aynı şekilde etkileşime girmesi ile oluşturulmuş bir veritabanından elde edilmiştir (Bölüm 3.2). Şekil 2.7'de görüldüğü üzere özellikle metal nesneler dağınık olarak bir çok frekansta yüksek bir genliğe sahiptir. Bunun aksine karton nesnelerin alçak frekanslardaki genlik değerleri yüksek iken, yüksek frekanslarda genlik değeri neredeyse sıfır seviyesinde bulunmaktadır. Ayrıca Şekil 2.7 ile 2.8 karşılaştırıldığında düşme etkileşimi ile elde edilen seslerde genlik değeri frekans olarak daha dağınık ve toplamda daha yüksek bir genliğe sahiptirler. Bunun nedeni



Şekil 2.7: Çarpma etkileşimindeki spektral vektörlerin normalize gücünün ortalaması ve standart sapması.



Şekil 2.8: Düşme etkileşimindeki spektral vektörlerin normalize gücünün ortalaması ve standart sapması.

düşme olayı sırasında sesin birden çok kez farklı noktalardan yer ile temas ederek yüksek genlikte farklı sesler çıkarmasıdır.

2.2.2 Sesin süresi özniteliği

Çizelge 2.1: Düşme ve çarpma olaylarında sesin süreleri

Malzeme	**Çarpma** [ms]	**Düşme** [ms]
Cam	182±97	579±395
Porselen	218±66	957±1270
Plastik	208±57	625±395
Metal	328±165	1000±1310
Karton	260±67	468±123

Sese hızlı Fourier dönüşümü uygulanması ve oluşan vektörlerin analizi sayesinde her bir malzemenin etkileşim sonrasında farklı sürelerde titreşimlerini koruyarak azalan genlikte ses çıkarmaya devam ettikleri gözlemlenmiştir. malzemeler arası bu farklılığın sesleri birbirinden ayırmada yararlı olabilecek bir öznitelik olduğu değerlendirilmiştir. İkinci bir öznitelik olarak düşünülen sesin süresi öznitelik çıkarma algoritmasına dahil edilmiştir. Sesin başlangıç algılama eşiğini geçtiği (onset) ve bitiş algılama eşiğinden düştüğü (offset) zamanlar arasındaki süre ölçülerek, bu süre öznitelik vektörüne eklenmiştir. Çizelge 2.1'de farklı nesneler ve olayların deneylerde kaydedilen ses süreleri görülmektedir. Aynı olay neticesinde farklı malzemelerden çıkan ses süreleri incelendiğinde metal nesnelerden çıkan sesin daha yüksek sürelerde devam ettiği açıkça görülmektedir. Fakat malzemeler arası ses süresi farklılığından ziyade olaylar arası ses süreleri farklılığı daha belirgindir. Bu farklılığın olumlu katkı sağladığı olayları ayırt etme deneyleri Bölüm 3.5'de incelenmiştir.

2.2.3 Tınısal öznitelikler

Tınısal öznitelikler genellikle müzik alanında sıklıkla kullanımaktadır [23–25]. Fakat nesnelerin sesleri üzerinde kullanımının örnekleri bulunmamaktadır. Tınısal öznitelikler ses üzerinden çıkarılan öznitelikler olması dolayısıyla bu çalışmasına dahil edilmiştir.

2.2.3.1 Zamansal sıfır geçiş oranı

Sıfır geçiş oranı bir sinyalin zamansal düzlemde ilerlerken genlik değerinin işaret değiştirme sayısının sinyal uzunluğuna oranıdır. Yani sinyal değerinin artıdan eksi değere veya tam tersi geçişleri ile ölçülür. Genellikle ritimsel ve özellikle müzik seslerini sınıflandırmada kullanılmaktadır [26]. Aşağıdaki formülde sıfır geçiş oranının hesaplanması gösterilmiştir. Formülde s sinyali, T sinyal uzunluğunu göstermekte, $sgn[x(n)]$ fonksiyonu ise sinyalin sıfır geçisi için 1, diğer zamanlar 0 değerini almaktadır.

$$zcr = \frac{1}{T-1} \sum_{t=1}^{T-1} sgn[s_{t-1} s_t] \tag{2.3}$$

$$sgn[x(n)] = \begin{cases} 1, & x(n) < 0 \\ 0, & x(n) \geq 0 \end{cases} \tag{2.4}$$

2.2.3.2 Spektral ağırlık merkezi

Spektral ağırlık merkezi, sesin spektrumundaki her bir bandın ortalaması olarak değerlendirilebilir. FFT uygulanmış sesin ayrı ayrı her bir vektörünün genlik değerleri ile aşağıdaki formül kullanarak hesaplanır. Formülde k FFT vektöründeki nokta numarası, N vektörün uzunluğu, $F[k]$ k noktasına denk gelen genlik değeridir [27].

$$Spektral\ Ağırlık\ Merkezi = \frac{\sum_{k=1}^{N} kF[k]}{\sum_{k=1}^{N} F[k]} \tag{2.5}$$

2.2.3.3 Azalış ve akış

Spektral bir öznitelik olan azalış ses sinyali spektrumunundaki eğim değerini ifade etmektedir. Akış ise ses işaretinin spektrumunundaki şekil değişimlerinin sayısal değerini temsil etmektedir [28].

2.2.3.4 Mel ölçekli kepstrum katsayıları (MFCC)

Mel ölçekli kepstrum katsayıları temel olarak, sese FFT uygulamadan sonra her bir Mel frekansındaki gücün logaritması alınarak hesaplanır [29]. Daha sonra Mel frekansı logaritmik güç değerlerine ayrık kosinüs dönüşümü (Discrete Cosine Transform-DCT) uygulanarak MFCC değerleri bulunur. MFCC öznitelikleri genellikle insan sesi ve

konuşma tanıma uygulamalarında kullanılmaktadır. MFCC hesaplamasının temel aşaması FFT olduğundan aynı pencereleme ve bindirme değerleri ile FFT'de olduğu gibi 1 saniyelik ses için 62 adet öznitelik vektörü oluşturmaktadır. Uygulama safhasında elde edilen MFCC değerleri yine 5 ardışık vektörün ortalaması alınarak sınıflandırma amacıyla kullanılmıştır.

2.3 Sınıflandırma

Sınıflandırma iki aşamada gerçekleşmektedir. İlk aşamada sınıfları bilinen sesler ile kulanılan makine öğrenmesi algoritmasına uygun model oluşturulur. İkinci aşamada ise sınıfları bilinmeyen sesler bu model kullanılarak sınıflandırılır. Etiketsiz sesleri sınıflandırmak için yaygın olarak kullanımı kabul görmüş iki farklı makine öğrenme algoritmasının bu tezde sınıflandırıcı tasarımında yer alması uygun görülmüştür. Destek Vektör Makinesi (SVM) ve Çok Katmanlı Algılama (MLP) yöntemleri diğer sınıflandıcılar arasında öne çıkmaktadır. Destek vektör makinesi ile çok katmanlı algılama yöntemleri aralarında karşılaştırıldığında ise SVM hem ortalama başarım olarak hem de işlem hızı olarak MLP'ye göre oldukça üstün olduğunu göstermektedir [30]. SVM'nin daha iyi olduğu konusunda çalışmalar olsa dahi MLP'nin de oldukça yaygın olarak kullanılan bir yöntem olmasından dolayı bu tez çalışmasına dahil edilmiştir. Her iki öğrenme yöntemi de aynı öznitelik vektörünü giriş olarak kullanacak şekilde tasarlanmıştır. Testler sırasında n-kat çapraz doğrulama (n-fold cross validation) yöntemi kullanılmıştır. Çapraz doğrulama yöntemi gerçek ortam teslerinden önce etiketi bilinen sesler ile sistemin performansını ölçmek amaçlıdır. Yöntem, eğitim setini n alt kümeye ayırarak 1 alt kümesini test seti olarak $n-1$ kümeyi de eğtim seti için kullanarak işletilmektedir.

2.3.1 Destek vektör makinesi (SVM)

Optimizasyona dayalı sınıflandırma teknikleri arasında yer alan Destek Vektör Makinesi (SVM), veri madenciliği sınıflandırma problemlerinde kullanılan bir yöntemdir. Bu yöntem sınıflandırmayı doğrusal veya doğrusal olmayan bir fonksiyon yardımıyla yerine getirir. Destek vektör makinesi yöntemi, veriyi birbirinden ayırmak için en uygun fonksiyonun tahmin edilmesi esasına dayanmaktadır.

Destek Vektör Makineleri öğrenme, sınıflandırma, kümeleme, yoğunluk tahmini ve son olarak da verilerden regresyon kuralları üretmek için kullanılan bir eğitme algoritmasıdır. SVM'nin teorik temelleri ilk olarak V.Vapnik tarafından 1960'lı yıllarda atıldıktan sonra ilk olarak 1995 yıllında sınıflandırmada kullanılmıştır. Vapnik'in teorisi eğitim kümelerindeki hata ile VC(Vapnik-Chervonenkis) boyutuna göre ifade edilen hipotez uzayının karmaşıklığının her ikisini de küçükleyen çözümün bulunduğunu göstermektedir [31].

Bu sınıflar arasında birçok doğrunun çizilebilme ihtimali vardır. Hata toleransını en düşük hale getirmek için çizilebilecek doğrular arasında her iki sınıfa en uzak olacak şekilde çizilen doğru bulunmalıdır.

Bu tez çalışmasında SVM tipi olarak nu-SVM diye de bilinen SVM Tip 2 kullanıldı. SVM Tip 2, hata fonksiyonunun en aza indirgenmesi temelli bir SVM tipidir. Ayrıca çekirdek fonksiyonu olarak radyal temelli fonksiyon (Radial Basis Function-RBF) kullanılmıştır.

2.3.2 Çok katmanlı algılayıcı (MLP)

Çok katmanlı algılama (MLP), ileri beslemeli, çok katmanlı ve doğrusal olmayan ayrıştırma yapabilen bir yapay sinir ağı türüdür. MLP, bir giriş katmanı, bir çıkış katmanı ve bu iki katman arasında yer alan bir veya birden fazla gizli katmandan oluşur.

MLP'nin çalışma yapısını iki bölümde incelemek mümkündür. Birincisi ileri yönde gerçekleşen iletim, diğeri ise hatanın geri yayılması algoritmasıdır. Ağa bir giriş uygulandığında girişler bir önceki katmandan bir sonraki katmana doğru çıkış katmanına kadar iletilir. Bu ileri yöndeki iletimdir. Eğitme aşamasında uygulanan giriş sinyali çıkış katmanına ulaştığında, elde edilen çıkış ile istenen çıkış arasındaki fark, yani hata sinyali geri yayılım ile giriş katmanına doğru sırasıyla iletilir [32].

3. LABORATUVAR DENEYLERİ

Bu bölümde, önerilen ses işleme mimarisinin performansının yanı sıra seçilen özniteliklerin etkinliğini test etmek için gerçekleştirilen deneyler anlatılmaktadır. Deneyler robotik sistemlerden bağımsız olarak icra edilmiştir.

3.1 Donanım ve Yazılım Özellikleri

Deneyler Intel Core @2.40GHz CPU, 16GB RAM kullanan bir bilgisayar üzerinde çalışan Ubuntu işletim sistemi ve Sony ECM-C10 mikrofon ile yapılmıştır. Mikrofon ses sinyallerini 32-bit çözünürlük ve 16kHz örnekleme oranı ile kaydetmekte kullanılmıştır.

Eğitim ve test safhalarında kullanılan ve eklenen ses işleme modülleri robotların işitmesine hizmet eden açık kaynak kodlu HARK sistemi kullanılarak yapılmıştır. Laboratuvar deneylerindeki sınıflandırmaların tamamı WEKA ile yapılmıştır. MLP algoritması WEKA programı dahilde olan makine öğrenmesi algoritmalarındandır. SVM içinse yine WEKA için LIBSVM kütüphanesinin Java ile yazılmış eklentisi kullanılmıştır.

Marsyas, genellikle müzik seslerini işlemede kullanılan, George Tzanetakis tarafından geliştirilmiş açık kaynak kodlu bir ses işleme kütüphanesidir[1]. Bu tezde marsyas kütüphanesi kullanılarak kaydedilen faydalı seslerin tınısal öznitelikleri çıkartılmıştır. Daha sonra tınısal öznitelikler FFT ve sesin süresi öznitelikleri ile birlikte deneylerde kullanılmıştır. Marsyas öznitelikleri olarak zamansal sıfır geçiş oranı, spektral ağırlık merkezi, azalış, akış ve Mel Ölçekli Kepstrum Katsayıları (MFCC) ile deneyler icra edilmiştir.

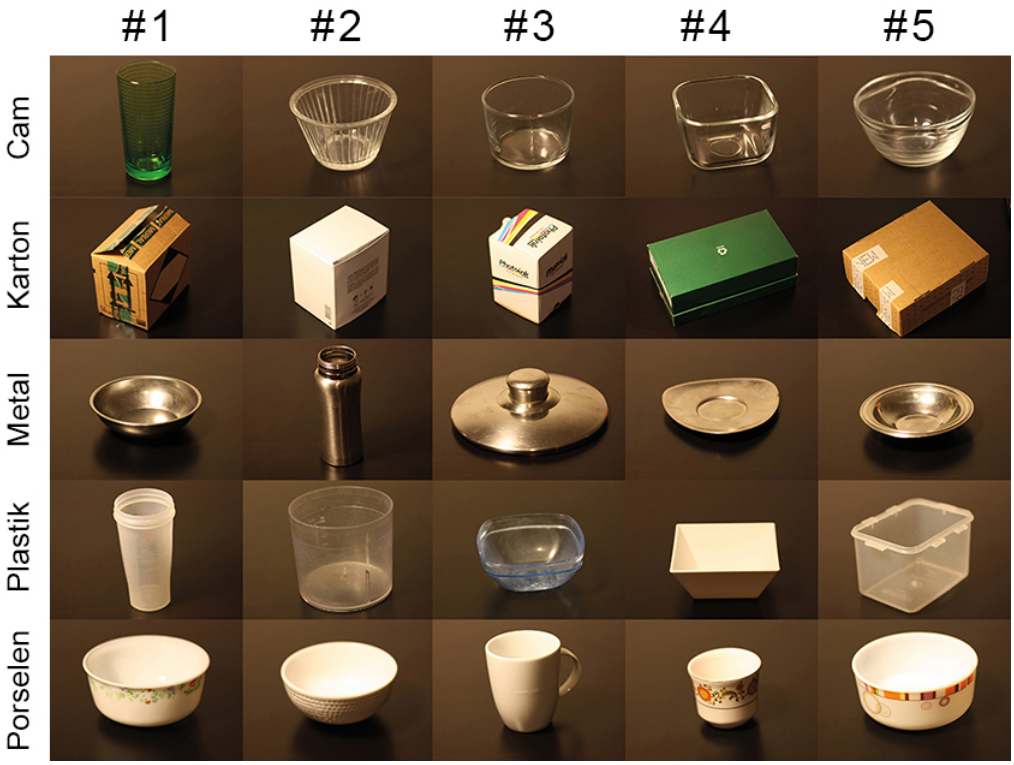
Tüm laboratuvar testleri uygulaması için Java ile kodlanmış bir makine öğrenmesi metodları koleksiyonu olan WEKA[2] kullanılmıştır. Canlı testlerde ise özellikle

[1] http://marsyas.info.
[2] http://www.cs.waikato.ac.nz/ml/weka.

MLP'nin işlem süresinin uzun olması nedeniyle sadece SVM kullanılmaya başlanmış ve daha hızlı bir erişim sağlamak ve sınıflandırma algoritmasını güçlendirmek için C++ ile kodlanmış LIBSVM [21] kütüphanesi sisteme dahil edilmiştir.

3.2 Nesneler

Deneyler için bir robotun ev ortamında etkileşime geçebileceği tipik mutfak eşyaları seçilmiştir. Tipik ev eşyaları seçmekteki ana sebep gerçek dünyada kullanım ideali için geliştirilen robotların, çalışma ortamlarında sıklıkla karşılaşabilecekleri nesne kümesine yakın örnekler ve şartlarla çalışmaktır. Nesneler beş farklı malzemeden imal edilmiş ve her bir malzemeden beş adet farklı eşya kullanılmıştır. Şekil 3.1'te gösterildiği gibi deneylerde toplam 25 farklı ev eşyası kullanılmıştır. Şekil 3.1'deki her bir satırda aynı malzemeden beş farklı eşya bulunmaktadır. Bunlar sırayla cam, karton, metal, plastik ve porselen malzemeden yapılmış nesnelerdir.



Şekil 3.1: Deneylerde eğitim ve testlerde kullanılan nesneler.

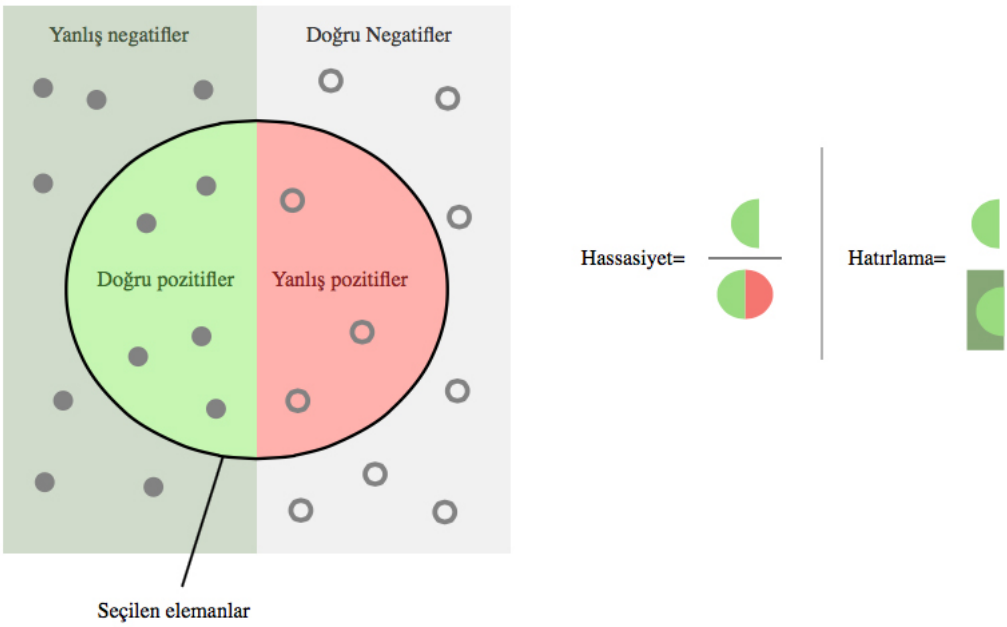
3.3 Deneysel Düzen

Laboratuvar deneyleri asıl olarak sistem geliştirme deneyleri olarak düşünülebilir, çünkü bu deneyler sırasında robotik sistemler kullanılmamış ve ortam gürültüleri minimumda tutularak sistemin en yüksek seviyede başarılı olması amaçlanmıştır. Böylece seçilen özniteliklerin başarıma etkisi üzerinde araştırma yapılarak sistemin en iyi performansı sağlaması amaçlanmıştır. Deneyler için iki farklı ses kümesi oluşturulmuştur: 1) 25 farklı ev eşyasına başka bir eşya ile vurarak (metal bıçak) 2) Eşyaları 10 cm yüksekten ahşap bir masaya düşürerek. Bu iki yöntem olaysal farklılık olarak ayrılmıştır. Birincisi *çarpma* olayı, ikinci ise *düşme* olayıdır. Farklı olayları kullanmaktaki amaç robotların icra ettikleri görevler sırasında karşılaşabilecekleri durumları simule etmektir. Ayrıca testler sırasında sadece sesi oluşturulan farklı malzemeler değil aynı zamanda olaylar da araştırılmış ve sınıflandırılmıştır.

Her olay içinde 25 ev eşyası kullanılarak her bir nesneden 5 adet örnek ses kaydedilmiştir. Örnekler oluşturulurken çarpma eylemi sırasında nesnelerin farklı noktalarına vurulmasına ve düşürme esnasında nesnelerin farklı ilk açılardan bırakılmasına özen gösterilmiştir. Böylece sabit bir ses değil, nesnenin farklı noktalarından farklı sesler elde ederek ses uzayının genişletilmesi amaçlanmıştır. Son olarak tüm sesler nesnelerin malzemelerine ve ses oluşturma eylemine (çarpma ve düşme) göre ayrı kategorilerde gruplanmıştır.

3.4 Değerlendirme Kriterleri

Değerlendirme için makine öğrenme yöntemlerinde sıklıkla kullanılan hassasiyet ve hatırlama değerleri baz alınmıştır. Hassasiyet, seçilen elemanların toplam sayısının doğru seçilen elemanların toplam sayısına oranıdır. Hatırlama, seçilen elemanların toplam sayısının seçilmesi bekelen elemanların toplam sayısına oranıdır [33]. Şekil 3.2'de gösterildiği şekilde hesaplanmaktadır.



Şekil 3.2: Hassasiyet ve hatırlama

3.5 Deneyler ve Sonuçlar

Laboratuvar deneyleri kapsamında dört farklı grupta testler icra edilmiştir. Bunlar: 1) Çarpma olayı ile malzeme sınıflandırma, 2) Olay sınıflandırma, 3) Çarpma ve düşme olayları ile malzeme sınıflandırma. İlk deney grubu sırasında öznitelik olarak sadece FFT vektörü ve ses süresi kullanılmıştır. Bunu takiben sisteme daha sonra tınısal öznitelikleri de eklenerek daha iyi bir performans elde edilmesi amaçlanmıştır. Burada elde edilen öznitelikler ile ikinci ve üçüncü deneylerin ideal sonuçları elde edilmiştir.

Deneyler için iki farklı **test kümesi** oluşturulmuştur. *Açık küme*'nin elemanları aynı malzemelerden yapılmış, ancak eğitim kümesinde bulunmayan ve her bir malzeme sınıfı için iki farklı nesneden oluşmaktadır. *Kapalı küme* ise eğitim kümesi ile tamamen aynı nesnelerden oluşmaktadır.

Buna ek olarak iki farklı **test yöntemi** kullanılmıştır. İlki eğitim kümesi kullanılarak gerçeklenen *n-kat çapraz doğrulama* yöntemi ve ikincisi sesin gerçek zamanda işlendiği *canlı test*'tir.

3.5.1 Çarpma olayı ile malzeme sınıflandırma

Çarpma olayı ile malzeme sınıflandırma deneyleri iki farklı test kümesi ve iki farklı test metodu ile yapılmıştır. Çapraz doğrulama metodu eğitim kümesinin $1/n$ elemanını test için ayırarak sınıflandırma yapan bir yöntem olduğu için kullandığı test kümesi kapalı küme olarak değerlendirilmiştir. Bu yöntemde kat sayısı (n) 10 olarak uygulanmıştır.

Bu bölümdeki deneylerdeki eğitim kümesi için Bölüm 3.2'de bahsedilen 25 nesne ile çarpma olayı dahilinde toplam 125 farklı ses kaydedilmiş ve işlenmiştir. Seslerden sadece FFT ve sesin süresi özniteliği çıkartılmıştır.

Çizelge 3.1: Çarpma olayı ile malzeme sınıflandırma

Sınıflandırıcı	Test Metodu	Test Seti	Hassasiyet	Hatırlama
SVM	Çapraz Doğrulama	Kapalı Set	0,893	0,891
MLP	Çapraz Doğrulama	Kapalı Set	0,867	0,868
SVM	Canlı Test	Açık Set	**0,872**	**0,870**
MLP	Canlı Test	Açık Set	0,810	0,807
SVM	Canlı Test	Kapalı Set	**0,905**	**0,903**
MLP	Canlı Test	Kapalı Set	0,830	0,828

Laboratuvarda gerçeklenen deneylerin birinci bölümünün sonuçları Çizelge 3.1'de sunulmuştur. Çizelgedeki kalın yazılı olanlar başarımda esas alınan değerleri göstermektedir. Sonuçlardaki genel eğilim SVM'nin hem hassasiyet hem de hatırlama değerlerinde MLP'den daha iyi bir sonuç verdiği yönündedir. Bütün metodlar ve test setleri için SVM 0,87'in üzerinde performans göstermiştir. Ayrıca bu testler ile bu bölümde kullanılan özniteliklerin sınıflandırmadaki performanslarının oldukça yüksek olduğu görülmektedir. Bunun yanı sıra test setlerinin sınıflandırmadaki etkisi beklenilen yönde olmuştur. Kapalı setler daha yüksek hassasiyet ve hatırlama oranlarına sahipken, açık setlerde hafif bir düşüş gözlemlenmiştir.

Çarpma olayı ile malzeme sınıflandırma deneylerinden sonra malzeme ve olayların birbiriyle karışma oranlarını tespit etmek için hata matrisleri incelenmiştir. Bu inceleme ile amaç eğer herhangi iki olay veya malzeme arasında belirgin ölçüde karışma mevcut ise bu iki sınıfın birbirinden ayrıştırılabilmesi için özel öznitelikler

bulmaktır. Çizelge 3.2'de görüldüğü üzere sınıflar arasında belirgin bir şekilde birbiriyle karışan bir ikili sınıf veya grup bulunmamaktadır. Çizelge 3.5'de 1'den 5'e kadar olan sayılar 5 farklı malzemeyi temsil eden sınıfları göstermektedir. Bu sınıflar; 1 cam, 2 porselen, 3 plastik, 4 metal, 5 karton.

Çizelge 3.2: Çarpma olayı ile malzeme sınıflandırma deneyinde hata matrisi

		Gerçek Sınıf				
		1	**2**	**3**	**4**	**5**
Sınıflandırıcı Tahmini	1	22	2	1	1	1
	2	1	20	1	1	1
	3	1	2	22	0	0
	4	1	1	1	22	0
	5	0	0	0	1	23

3.5.2 Olay sınıflandırma

Olay sınıflandırma deneyleri için çarpma olayı ile kaydedilen seslere düşme olayının sesleri de eklenmiştir. Böylece toplam ses örneği sayısı ikiye katlanarak 250'ye çıkmıştır. Testler sadece kapalı setler ile icra edilmiştir. Sınıflandırıcı olarak SVM ve MLP algoritmaları kullanılmıştır.

Çizelge 3.3: Olay sınıflandırma sonuçları

Sınıflandırıcı	**Test Metodu**	**Hassasiyet**	**Hatırlama**
SVM	Çapraz Doğrulama	0,962	0,962
MLP	Çapraz Doğrulama	0,958	0,958
SVM	Canlı Test	**0,876**	**0,871**
MLP	Canlı Test	0,861	0,857

Bu deneylerdeki amaç hazırladığımız sistemin sadece malzeme sınıflandırmada değil aynı zamanda olay sınıflandırmada da kullanılabileceğini deneyimlemektir. Yeni test kümesi ile yapılan deneyin sonuçlarını içeren Çizelge 3.3 sistemin olay sınıflandırma üzerindeki performansını göstermektedir. Toplamda iki sınıf içeren deneylerde şans faktörü %50 iken, çapraz doğrulama sonuçlarının %96 olması sistemin çok iyi bir şekilde gerçek dünyada da olay ayırmada başarım sağlayacağını göstermiştir. Bu

öngörüyle bu testler aynı sistem ve test kümesi ile canlı testler icra edilmiştir. Canlı testlerde elde edilen %87 civarındaki olay sınıflandırma başarımı oldukça başarılı bir sonuçtur. Çapraz doğrulama ve canlı teslerin her ikisinde de SVM algoritması MLP'ye göre yüksek başarım elde etmiştir.

3.5.3 Çarpma ve düşme olayları ile malzeme sınıflandırma

Malzemeden bağımsız olarak yapılan olay sınıflandırma deneylerinden sonra sistemin malzeme ve olayı aynı anda sınıflandırmadaki başarımı bu bölümde test edilmiştir. Bu deneylerde de ilk iki bölümde kullanılan öznitelik vektörleri aynen kullanılmıştır. Olay ilişkili nesne sınıflandırma sonuçları Çizelge 3.4'de gösterilmiştir. Olay ve nesne sınıfları birbirleriyle eşlenerek değerlendirildiği için deneylerde toplamda 10 sınıf bulunmaktadır. Sınıf sayısının artışıyla birlikte sistemin performansında da önemli bir düşüş olduğu gözlemlenmiştir. Buna rağmen canlı teslerdeki yaklaşık %60 doğru tanıma oranı 10 sınıf için şans faktörünün %10 olduğu düşünülürse, iyi bir başarı olarak değerlendirilebilir.

Çizelge 3.4: Olay ve nesnenin birlikte sınıflandırıldığı sonuçlar

Sınıflandırıcı	Test Metodu	Hassasiyet	Hatırlama
SVM	Çapraz Doğrulama	0,755	0,755
MLP	Çapraz Doğrulama	0,755	0,747
SVM	Canlı Test	**0,601**	**0,599**
MLP	Canlı Test	0,620	0,519

Sadece çarpma olayı ile yapılan deneylerde olduğu gibi çarpma ve düşme olayları ile malzeme sınıflandırma deneylerinden sonra da malzeme ve olayların birbiriyle karışma oranlarını tespit etmek için hata matrisleri incelenmiştir. Çizelge 3.5'de görüldüğü üzere sınıflar arasında belirgin bir şekilde birbiriyle karışan bir ikili sınıf veya grup bulunmamaktadır. Çizelge 3.5'de 1'den 10'a kadar olan sayılar 5 farklı malzeme ile 2 farklı olayın eşleşmesi ile elde edilen sınıfları göstermektedir. Bu sınıflar; 1 cam-çarpma, 2 porselen-çarpma, 3 plastik-çarpma, 4 metal-çarpma, 5 karton-çarpma, 6 cam-düşme, 7 porselen-düşme, 8 plastik-düşme, 9 metal-düşme, 10 karton-düşme.

Çizelge 3.5: Çarpma ve düşme olayları ile malzeme sınıflandırma deneyinde hata matrisi

		Gerçek Sınıf									
		1	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**
Sınıflandırıcı Tahmini	**1**	17	1	1	1	1	1	1	1	0	0
	2	1	20	0	0	1	1	1	1	2	0
	3	2	0	19	1	1	1	0	0	0	1
	4	0	1	1	19	1	0	1	1	1	1
	5	1	1	0	0	19	1	2	1	0	0
	6	0	0	1	1	1	19	0	0	3	0
	7	1	0	0	2	0	1	18	1	1	1
	8	2	1	0	0	0	1	2	20	0	1
	9	1	0	3	0	0	0	0	0	17	0
	10	0	1	0	1	1	0	0	0	1	21

FFT değerleri ve sesin süresi öznitelikleri ile yapılan deneylerden sonra başarımın daha da arttırılması amacıyla yeni öznitelikler test edilmiştir. Literatür araştırması sırasında nesne tanıma gibi uygulamalarda kullanımına daha önce rastlanmayan fakat müzik uygulamalarında sıklıkla kullanılan öznitelikler nesne ve olay tanıma amacıyla incelenerek deneyimlenmiştir. Bu amaçla öznitelik çıkarma algoritmaları ses tespitinin yapıldığı eşikleme işleminin hemen ardından çalışacak şekilde uygulanmıştır.

Tınısal özniteliklerinde eklenmesinin ardından sınıflandırmada kullanılan öznitelik vektörü 191 değerlik bir uzunluğa ulaşmıştır. Deneyler sırasında MLP algoritmasının eğitim modeli oluşturma süresi genel itibariyle 10 dakikanın üzerine çıkmaya başlamıştır. Sistemin başarılı olması halinde bile MLP algoritmasının çevrimiçi öğrenme becerisine sahip olmasını istediğimiz robotik sisteme ve canlı testlere uygulanamayacağı değerlendirilerek bu algoritmanın kullanımına şimdilik son verilmiştir.

Tınısal özniteliklerinin analizi için yapılan testlerin tamamı SVM makine öğrenme algoritması ve 10-kat çapraz doğrulama metodu kullanılarak yapılmıştır. Bu deneylerde de çarpma ve düşme olayları ile malzeme sınıflandırma testlerinde olduğu gibi 25 farklı nesneden 2 farklı olay ile toplamda 250 adet ses kullanılmıştır. Bu özniteliklerin başarımı daha ayrıntlı inceleyebilmek için bütün öznitelikler aynı anda kullanıldığı gibi farklı kombinasyonlarda da kullanılarak tanıma olayı üzerindeki

Çizelge 3.6: Tınısal öznitelikleri ile malzeme sınıflandırma.

Öznitelik/Testler	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**
FFT	✓	✓	✓	✓	✓	✓	✗	✗
Ses Süresi	✓	✓	✓	✓	✓	✓	✓	✗
Sıfır Geçiş Oranı	✗	✓	✓	✓	✓	✓	✓	✓
Azalış	✗	✗	✓	✓	✓	✓	✓	✓
Spektral Ağırlık Merkezi	✗	✗	✗	✓	✓	✓	✓	✓
Akış	✗	✗	✗	✗	✓	✓	✓	✓
MFCC	✗	✗	✗	✗	✗	✓	✓	✓
Tanıma Yüzdesi(%)	**75.5**	74.0	73.5	74.1	74.7	69.5	41.3	39.0

etkisi incelenmiştir. Bu kapsamda yapılan yedi farklı deneyin sonuçları ve öznitelik vektörünün içerikleri Çizelge 3.6'de gösterilmiştir. Çizelgedeki ilk deney Bölüm 3.5.3'deki deneyi göstermektedir. En üst satırda bulunan rakamlar yapılan deneylerin numaralarını, ✗ ve ✓ işaretleri ise o işarete satır başında karşılık gelen özniteliğin yine o işarete sütun başında karşılık gelen deneyde kullanılıp kullanılmadığını göstermektedir.

Çizelge 3.6'te görüldüğü üzere Marsyas özniteliklerinin sistemin tanıma başarımı üzerinde olumlu katkısı olmamış, aksine SVM algoritması üzerinde bozucu etkiye sahip olmuştur. Bu testler sonucunda anlaşılmıştır ki tınısal öznitelikler ile tasarlanan sistem için uygun öznitelikler değildir. Dolayısıyla bir sonraki bölümde bahsedilecek olan robotik sistemlerle yapılan deneylerde kullanılmamıştır.

4. ROBOT DENEYLERİ

Bu bölümde, robotları da içeren deneyler insanların gün içerisinde kullandığı üniversite koridorlarında, gerçek ortam koşullarında yapılmıştır. Yöntem iki farklı robota yazılımsal olarak entegre edilmiş ve iki gerçekçi robot senaryosu kullanılarak test edilmiştir.

4.1 Robotlar

Robot deneyleri iki farklı robot ile icra edilmiştir. Bu robotlar Pioneer 3-AT[1] gezgin robot ve robotik koldur. Pioneer 3-AT çok yönlü kullanım alanına sahip 4 tekerlekli bir araştırma robotudur. Bu robotla yapılan deneyler robotun kabiliyeti gereği sadece nesnelerin devrilerek çıkardıkları ses üzerinden icra edilmiştir. Deneylerde kullanılan robotik kol 7 serbestlik derecesine sahiptir. Robotik kol ile ise çarpma ve düşme deneyleri ayrı ayrı icra edilebilmiştir.

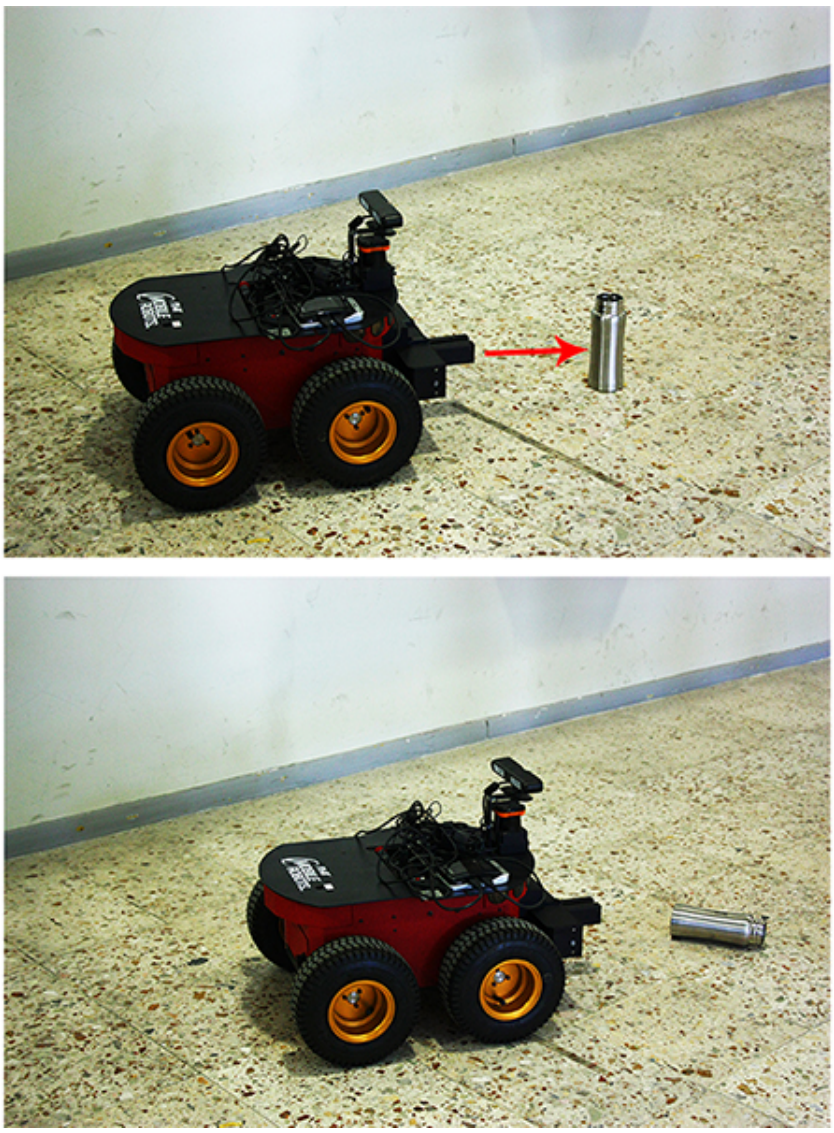
4.2 Deneyler ve Sonuçlar

Robotlar ile deneyler ve sonuçları bu tezin asıl amacı ve gerçek dünya uygulama safhasıdır. Elde edilen sonuçlar sistemin robotlar üzerinde uygulanabilirliğini gösterecektir. Aynı zamanda laboratuvar deneylerinde rastlanmayan değişik ortam şartlarının sistem üzerinde etkisinide gözlemlenebilecektir. Sistemin iki farklı robot üzerinde deneyimlenmesi, sistemin hem farklı alan uygulamalarındaki performansını gözlemlemek hem de kolay uyarlanabilir olduğunu göstermek açısından önemlidir.

4.2.1 Gezgin robot ile deneyler

Gezgin robot Pioneer 3-AT ile yapılan deneylerde, robotun ilerlerken çarptığı bir nesnenin yere düşmesiyle çıkan ses kullanılarak bir sınıflandırma yapılmıştır (Şekil 4.1). Deneylerde 3 farklı malzemeden (metal, plastik ve cam) yapılmış nesneler

[1] www.mobilerobots.com/ResearchRobots/P3AT.aspx



Şekil 4.1: Pioneer 3-AT bir nesneyi düşürürken

kullanılmıştır. Gezgin robotlar ile yapılan deneylerde laboratuvar deneylerine göre daha az nesne kullanılmasının çeşitli sebepleri vardır. Bunun başlıcaları; bazı nesnelerin yapıları gereği devrilmeye müsait olmamaları, karton'dan yapılan nesnelerin devrilme ile ortam gürültüsünün üstünde ses çıkaramaması, dolayısıyla ses tespitinin yapılamaması ve bazı nesnelerin kırılma ihtimallerinin yüksek olmasıdır.

SVM sınıflandırıcısının laboratuvar deneylerinde daha yüksek sonuç vermesinden dolayı varsayılan sınıflandırıcı olarak SVM kullanılmıştır. Deneyler için belirlenen eşik değerleri gezgin robotun motor gürültüsünün ve ortam gürültüsünün yüksek olması sebebiyle laboratuvar deneylerine göre daha yüksek bir değere ayarlanmıştır. Çizelge 4.1'de sunulan deney sonuçlarında görüldüğü üzere çapraz doğrulama

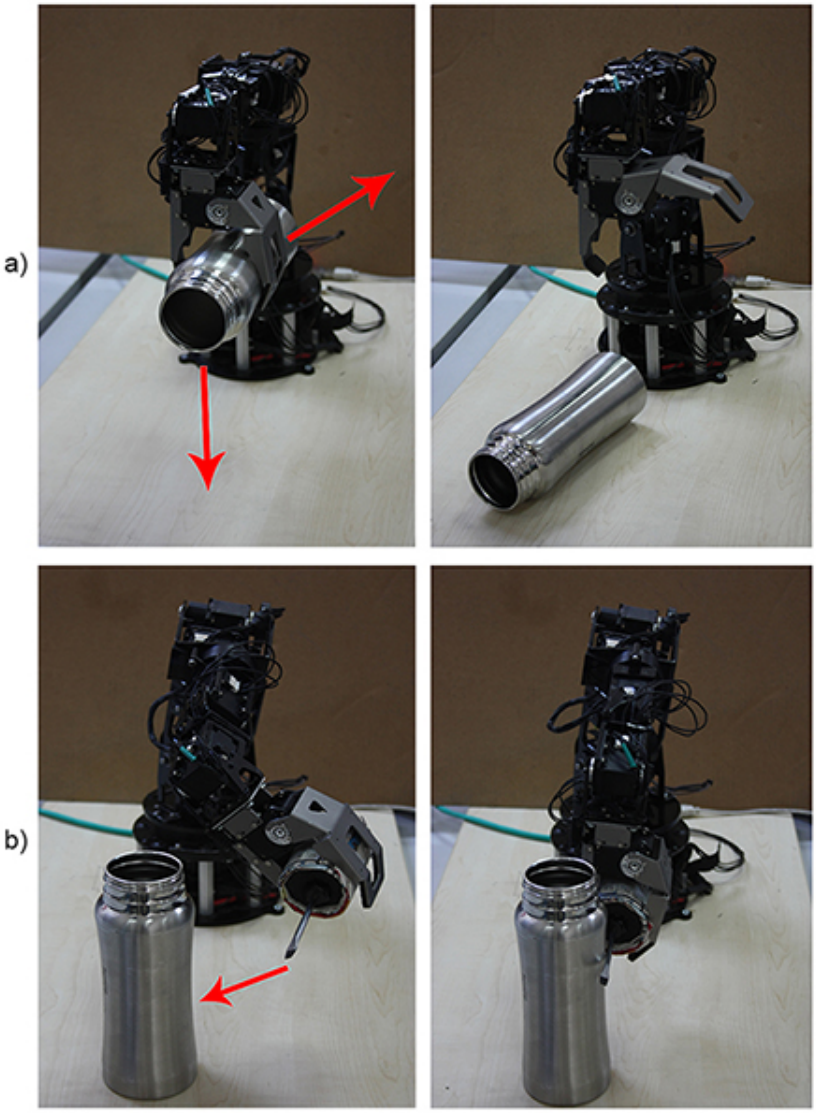
methodu ile yapılan testlerde tanıma başarımı yüksek iken canlı testlerde belirgin bir düşüş gözlemlenmiştir. Bunun nedeni yapılan deney ortamının günlük gürültüleri barındıran üniversite koridoları olması ve eğitim için kaydedilen seslerle test seslerin zamansal farklılığından kaynaklı ortamdaki gürültünün farklı olmasıdır. Aynı zamanda Pioneer 3-AT motorlarının oldukça gürültülü olmasının da test sonuçlarına olumsuz etkisi olmuştur.

Çizelge 4.1: Pioneer 3-AT robotunu kullanarak çarpma olayı ile malzeme sınıflandırma

Test Metodu	Hassasiyet	Hatırlama
Çapraz Doğrulama	0.887	0.882
Canlı Test	**0.663**	**0.651**

4.2.2 Robotik kol ile deneyler

Robotik kol ise hem çarpma hem de düşme olaylarını uygulamak için kullanılmıştır. Düşürme hareketi için robotik kol kıskacı ile nesneyi tutup 10cm yükseliğe çıkararak zemine bırakmıştır (Şekil: 4.2a). Çarpma hareketi için ise robotik kol tuttuğu metal tornavida ile nesneye vurmuştur (Şekil: 4.2b). Çizelge 4.2'da robotik kol ile yapılan testlerde düşürme ve çarpma eylemleri için malzeme tanıma sonuçları sunulmuştur. Çapraz doğrulama yöntemi ile yapılan testlerde başarımın gezgin robot ile yapılan testlere göre düşük olduğu gözlense de bunda robotik kol ile yapılan deneylerin 5 kategoride 25 nesnenin tamamının kullanılmasının etkili olduğu göz önünde bulundurulmalıdır. Canlı testlerde ise gezgin robotun aksine robotik kolun daha sessiz çalışması ve testlerin koridorlar yerine laboratuvar ortamına yakın olan kapalı kapılar ardında yapılması sebebiyle gezgin robotun elde ettiği başarımdan daha yüksek bir tanıma yüzdesi elde edilmiştir. Çizelge 4.3'de ise robotik kol ile eylem tanıma testlerinin sonuçları sunulmaktadır. Bu testlerde sistem yüksek oranla eylemleri ayırt edebilmiştir.



Şekil 4.2: a) Robotik kol bir nesneyi düşürürken, b) Robotik kol bir nesneye vururken.

Çizelge 4.2: Robotik kol ile düşürme ve çarpma eylemleri için malzeme tanıma sonuçları

Test Metodu	Hassasiyet	Hatırlama
Çapraz Doğrulama	0.803	0.797
Canlı Test	**0.732**	**0.729**

Çizelge 4.3: Robotik kol ile düşürme ve çarpma eylemleri ayırt etme sonuçları

Test Metodu	Hassasiyet	Hatırlama
Çapraz Doğrulama	0.844	0.842
Canlı Test	**0.814**	**0.803**

5. SONUÇ VE GELECEK ÇALIŞMALAR

Nesneleri kategorilere göre sınıflandırmak akıllı sistemler için insan ortamında bir ön koşuldur. Bu tez çalışmasında, nesnelerden elde edilen sesler ile nesnelerin yapıldığı malzemeyi ve oluşan sesin kaynağı olan olayları sınıflandırmak için bir sistem önerilmiştir. Sistem ROS ile entegre olarak çalışabilen açık kaynak kodlu HARK sistemi üzerinde geliştirilmiştir. Spektral güç ve seslerin süresini içeren öznitelik vektörü kullanılan bu sistemin bir SVM tabanlı sınıflandırıcı ile birleştiğinde, nesnelerin malzemesinin yanı sıra etkileşim türünün de tanındığı gösterilmiştir. Olay sınıflandırma olarak isimlendirdiğimiz bu deneyler seçilen öznitelikler ile oldukça başarılı olduğu ispatlamıştır.

Tez çalışması sırasında başarılı olan kadar, olmayan öznitelikler de tespit edilmiştir. Çalışmalarda özellikle tınısal öznitelikleri olarak nitelendirilen zamansal sıfır geçiş oranı, spektral ağırlık merkezi, azalış, akış ve Mel Ölçekli Kepstrum Katsayıları (MFCC) öznitelikleri bazı referans çalışmalarda, özellikle müzik tabanlı çalışmalarda başarılı olsa da nesnelerin karakteristik sesleri alanında başarılı olmadığı gösterilmiştir.

Çalışmalar sonucunda hazırlanan sistem yine açık kaynak kodlu olarak bulunmaktadır. Ayrıca BSD lisansı ile dağıtılan LIBSVM kütüphanesi nihai olarak sistemin sınıflandırma algoritması olarak sisteme entegre edilmiştir. Dolayısıyla sistem sınıflandırma özellikleri dahil tamamıyla kullanıma ve geliştirmeye hazır haldedir.

Sonuç olarak, önerilen sistem robotik sistemlerin işlem yaptıkları sahneyi analiz etmelerini sağlamaktadır. Deney sonuçlarında görüldüğü üzere sistem ses ile sahne analizi açısından yüksek başarımlar ile çalışmaktadır. Ayrıca çalışmaların ses özelliklerini işlemeye dayanıyor olması robotik sistemlerde çok önemli olan görüş alanı sorunundan bağımsız olarak, küresel ve göz görüş alanı dışında kalan yerleri de kapsayarak yüksek farkındalıklı bir sistem yaratılmasını sağlamıştır. Önerilen sistem robotların sürekli dinleme ile icra ettikleri görevler esnasında odaklandıkları nokta dışında kalan kısımlarda da meydana gelen olayları takip edebilme ve tepki

verebilme yeteneği sağlar. Böylece robotlar sahnesel farkındalıklarını artırabilir, odak noktalarında olmayan noktalarda meydana gelen olayların değerlendirmesini yapabilir ve karar verme mekanizmaları için ekstra yararlı bilgi sağlayarak daha başarılı görevler icra edilmesini sağlayabilir.

Bu tez çalışması neticesinde elde edilen veriler ve tecrübeler ile kurulan sistem bir kaç farklı yönde geliştirilerek yeni çalışmalar yapılabilir. Bunların birincisi, nesne ve malzeme kümesinin genişletilmesidir. Bu çalışmada kullanılan 5 farklı malzemeden yapılmış nesneler daha farklı malzemelerden daha geniş bir yelpazede incelenerek gerçek dünyada daha güçlü ve daha dayanıklı bir sistem geliştirilebilir. Nesne ve malzeme kümesini genişletmenin yanı sıra olay kümesinin genişletilmesi de aynı yönde faydalı olacaktır.

Önerilen sisteme yeni özellikler katmak da gelecek çalışmaların hedefi olabilir. Özellikle bir mikrofon dizi kullanarak sistemi ses kaynağı konumunun tespit edebilir hale getirilmesi faydalı bir geliştirme olacaktır. Özellikle robotik platformlar için geliştirilen sistemin bu ek özellik ile birlikte bulunduğu robotik araçları daha yetenekli kılacaktır.

KAYNAKLAR

[1] **Yildiz, P.**, **Karapinar, S. ve Sariel-Talay, S.**, 2013. Learning Guided Symbolic Planning for Cognitive Robots, IEEE International Conference on Robotics and Automation (ICRA), Autonomous Learning Workshop, Karlsruhe, Almanya.

[2] **Altan, D. ve Sarıel-Talay, S.** Probabilistic Failure Isolation for Cognitive Robots, Proceedings of the 27th International FLAIRS Conference, s.370–375.

[3] **Sinapov, J.**, **Schenck, C.**, **Staley, K.**, **Sukhoy, V. ve Stoytchev, A.**, 2014. Grounding semantic categories in behavioral interactions: Experiments with 100 objects, *Robotics and Autonomous Systems*, **62(5)**, 632 – 645, Özel sayı Semantic Perception, Mapping and Exploration.

[4] **C. Schenck, A.S.**, 2012. The Object Pairing and Matching Task: Toward Montessori Tests for Robots, In Proceedings of the Humanoids 2012 Workshop on Developmental Robotics, Osaka, Japonya.

[5] **Loutfi, A.**, **Broxvall, M.**, **Coradeschi, S. ve Karlsson, L.**, 2005. Object recognition: A new application for smelling robots, *Robotics and Autonomous Systems*, **52(4)**, 272 – 289.

[6] **Valin, J.M.**, **Michaud, F.**, **Rouat, J. ve Létourneau, D.**, 2003. Robust Sound Source Localization Using a Microphone Array on a Mobile Robot, in Proceedings International Conference on Intelligent Robots and Systems, s.1228–1233.

[7] **Sharan, L.**, **Liu, C.**, **Rosenholtz, R. ve Adelson, E.H.**, 2013. Recognizing materials using perceptually inspired features, *International Journal of Computer Vision*, **108(3)**, 348–371.

[8] **L. Calmes, H. Wagner, S.S.G.L.**, 2007. Combining Sound-Localization and Laser-based Object Recognition, AAAI Spring Symposium, Stanford, CA, Amerika Birleşik Devletleri, s.1–6.

[9] **Hao Wu, Guohui Tian, X.W. ve Lu, F.**, 2014. Four-Leg Object Recognition for Service Robot Based on Top-hat Transformation, International Journal of Control and Automation, cilt 7, s.311–322.

[10] **J. Sinapov, A.S.**, 2009. From Acoustic Object Recognition to Object Categorization by a Humanoid Robot, In Proceedings of the RSS Workshop: Mobile Manipulation in Human Environments, Seattle, WA, Amerika Birleşik Devletleri.

[11] **Sinapov, J. ve Stoytchev, A.**, 2011. Object category recognition by a humanoid robot using behavior-grounded relational learning, 2011 IEEE International Conference on Robotics and Automation (ICRA), s.184–190.

[12] **Valin, J.M.**, **Michaud, F.**, **Rouat, J. ve Létourneau, D.**, 2003. Robust Sound Source Localization Using a Microphone Array on a Mobile Robot, in Proceedings International Conference on Intelligent Robots and Systems, s.1228–1233.

[13] **Weber, M.**, **Welling, M. ve Perona, P.**, 2000. Unsupervised Learning of Models for Recognition, Computer Vision - ECCV 2000, cilt1842 of *Lecture Notes in Computer Science*, Springer Berlin Heidelberg, s.18–32.

[14] **Lillard, A.** *Science News*.

[15] **Delgado-Contreras, J.R.**, **Garćıa-Vázquez, J.P.**, **Brena, R.F.**, **Galván-Tejada, C.E. ve Galván-Tejada, J.I.**, 2014. Feature Selection for Place Classification through Environmental Sounds, *Procedia Computer Science*, **37(0)**, 40 – 47, the 5th International Conference on Emerging Ubiquitous Systems and Pervasive Networks (EUSPN-2014)/ The 4th International Conference on Current and Future Trends of Information and Communication Technologies in Healthcare (ICTH 2014)/ Affiliated Workshops.

[16] **Dennis, J.**, **Tran, H. ve Chng, E.**, 2013. Overlapping sound event recognition using local spectrogram features and the generalised hough transform, *Pattern Recognition Letters*, **34(9)**, 1085 – 1093.

[17] **Argentieri, S.**, **Danès, P. ve Souères, P.**, 2015. A survey on sound source localization in robotics: From binaural to array processing methods, *Computer Speech & Language*, **(0)**, –.

[18] **Belloch, J.A.**, **Gonzalez, A.**, **Vidal, A.M. ve Cobos, M.**, 2015. On the performance of multi-GPU-based expert systems for acoustic localization involving massive microphone arrays, *Expert Systems with Applications*, **42(13)**, 5607 – 5620.

[19] **Chen, Z.M.**, **Zhu, H.C. ve Peng, M.**, 2015. Identification and localization of the sources of cyclostationary sound fields, *Applied Acoustics*, **87(0)**, 64 – 71.

[20] **Dehkordi, M.B.**, **Abutalebi, H.R. ve Taban, M.R.**, 2013. Sound source localization using compressive sensing-based feature extraction and spatial sparsity, *Digital Signal Processing*, **23(4)**, 1239 – 1246.

[21] **Chang, C.C. ve Lin, C.J.**, 2011. LIBSVM: A library for support vector machines, *ACM Transactions on Intelligent Systems and Technology*, **2**, 27:1–27:27.

[22] **Cooley, J.W. ve Tukey, J.W.**, 1965. An algorithm for the machine calculation of complex Fourier series, *Mathematics of Computation*, **19**, 297–301.

[23] **Sargent, G.**, **Hanna, P. ve Nicolas, H.**, 2014. Segmentation of music video streams in music pieces through audio-visual analysis, Acoustics, Speech and Signal Processing (ICASSP), 2014 IEEE International Conference on, s.724–728.

[24] **Tsunoo, E.**, **Tzanetakis, G.**, **Ono, N. ve Sagayama, S.**, 2009. Audio genre classification using percussive pattern clustering combined with timbral features, Multimedia and Expo, 2009. ICME 2009. IEEE International Conference on, s.382–385.

[25] **Marolt, M.**, 2005. Audio Melody Extraction Based on Timbral Similarity of Melodic Fragments, Computer as a Tool, 2005. EUROCON 2005.The International Conference on, cilt 2, s.1288–1291.

[26] **Gouyon, F.**, **Pachet, F. ve Delerue, O.**, 2000. On the use of zero-crossing rate for an application of classification of percussive sounds, Proc. COST G-6 Conf. Digital Audio Effects (DAFX-00), Verona, İtalya, s.147–152.

[27] **Schubert, E.**, **Wolfe, J. ve Tarnopolsky, A.**, 2004. Spectral centroid and timbre in complex, multiple instrumental textures, Proc. of the 8th International Conference on Music Perception & Cognition (ICMPC), Evanston, s.654–657.

[28] **Kos, M.**, **Kačič, Z. ve Vlaj, D.**, 2013. Acoustic classification and segmentation using modified spectral roll-off and variance-based features, *Digital Signal Processing*, **23(2)**, 659 – 674.

[29] **Zheng, F.**, **Zhang, G. ve Song, Z.**, 2001. Comparison of different implementations of MFCC, *Journal of Computer Science and Technology*, **16(6)**, 582–589.

[30] **Zanaty, E.**, 2012. Support Vector Machines (SVMs) versus Multilayer Perception (MLP) in data classification, *Egyptian Informatics Journal*, **13(3)**, 177 – 183.

[31] **Vapnik, V.N.**, 1998. Statistical Learning Theory, Wiley, 1. sürüm.

[32] **Bernard, I.**, 1993. Multilayer perceptron and uppercase handwritten characters recognition, Document Analysis and Recognition, 1993., Proceedings of the Second International Conference on, s.935–938.

[33] **Ting, K.**, 2010. Precision and Recall, **C. Sammut ve G. Webb**, (düzenleyenler), Encyclopedia of Machine Learning, Springer US, s.781–781.

ÖZGEÇMİŞ

Ad Soyad: İsmail İren SALTALI

Doğum Yeri ve Tarihi: 03.07.1987

Adres: Balmumcu Mah. Palanga Cad. Deniz Subay Loj. Oruçreis Apt. 70H/2 Beşiktaş/İstanbul

E-Posta: saltalii@itu.edu.tr

Lisans: Bilgisayar Mühendisliği, Deniz Harp Okulu, 2009

Y. Lisans: Bilgisayar Mühendisliği, İstanbul Teknik Üniversitesi, 2015

TEZDEN TÜRETİLEN YAYINLAR/SUNUMLAR

- **SALTALI İ.**, İNCE G., SARIEL S., 2014: İşitsel Olay İzleme İle Sahne Analizi *1. Türkiye Otonom Robotlar Konferansı 2014 (TORK)*, 6-7 Kasım 2014, Ankara, Türkiye
- **SALTALI İ.**, İNCE G., SARIEL S., 2015: Sahne Analizi İçin Ses Kaynağı Tespiti *23. IEEE Sinyal İşleme ve İletişim Uygulamaları Kurultayı (SIU)*, 16-19 Mayıs, 2015, Malatya, Türkiye.