

**YILDIZ TEKNİK ÜNİVERSİTESİ**  
**FEN BİLİMLERİ ENSTİTÜSÜ**

SEYİR HALİNDEKİ GEMİLERİN ÇALIŞMA SESİNDEN TANINMASI

İDRİS AYKUN

FBE Bilgisayar Mühendisliği Anabilim dalında

Hazırlanan

YÜKSEK LİSANS TEZİ

Tez Danışmanı: Yrd. Doç. Dr. M. Elif KARSLIGİL

İSTANBUL, 2007

# İÇİNDEKİLER

	<b>Sayfa</b>
SİMGE LİSTESİ .....	iii
KISALTMA LİSTESİ .....	iv
ŞEKİL LİSTESİ .....	v
ÇİZELGE LİSTESİ .....	vii
ÖNSÖZ .....	viii
ÖZET .....	ix
ABSTRACT .....	x
1. GİRİŞ.....	1
2. SES HAKKINDA GENEL BİLGİ.....	4
2.1 Sesin Tanımı .....	4
2.2 Ses Dalgaları .....	4
2.2.1 Frekans (Sıklık).....	5
2.2.2 Dalga boyu .....	6
2.2.3 Genlik .....	6
2.2.4 Ton .....	7
2.2.5 Tını .....	7
2.2.6 Sesin şiddeti ve desibel ölçeği .....	7
2.3 Sayısal Ses .....	9
2.3.1 Analog-Sayısal dönüşüm .....	9
2.3.2 Örnekleme teoremi .....	9
2.3.3 Kuantalama ve kodlama .....	10
3. SES ÖZELLİKLERİNDEN GEMİ TANIMA SİSTEMİNE GENEL BAKIŞ .....	11
3.1 Ön İşlemler .....	11
3.1.1 Wave formatlı giriş dosyasının okunması .....	12
3.1.2 Sesin filtrelenmesi .....	12

3.2	Özellik Çıkartımı.....	12
3.2.1	Çerçeveleme.....	13
3.2.2	Pencereleme .....	14
3.2.3	Hızlı Fourier Dönüşümü (FFT).....	15
3.2.4	Mel ölçeğine aktarma .....	17
3.2.5	$\Delta$ MFCC ve $\Delta^2$ MFCC ile ek özelliklerin çıkartılması.....	18
3.3	k-Ortalama Yöntemi ile Özelliklerin İndirgenmesi .....	20
3.4	Tanıma Aşaması.....	21
3.4.1	Destek Vektör Makineleri (SVM).....	22
3.4.2	k-En Yakın Komşuluk (k-NN).....	23
3.4.3	Gemi veritabanında olmayan seslerin eşik değeri kullanılarak belirlenmesi .....	24
4.	UYGULAMA .....	27
4.1	Kullanılan Ses Örnekleri .....	27
4.2	Ön İşlemler, Özellik Çıkartımı ve Özelliklerin İndirgenmesi .....	27
4.3	Sınıflandırma.....	29
4.3.1	Sınıflandırmada kullanılacak yöntemin belirlenmesi.....	29
4.3.2	k-NN yönteminin hiyerarşik uygulanması ile başarı oranının arttırılması.....	30
4.3.3	Sınıflandırmayı kolaylaştırmak amacıyla Euclid uzaklığının değiştirilmesi.....	31
4.3.4	Sınıflandırmaya dâhil olmayan seslerin tespiti için eşik değerinin belirlenmesi.....	32
4.3.5	k-NN yöntemi ile sınıflandırma .....	35
4.3.5.1	Sınıflandırmada eş benzerlik oranları oluşması halinde alternatif bir çözüm .....	35
4.4	Tanıma Parametrelerinin Seçimi.....	36
5.	SONUÇLAR .....	39
	KAYNAKLAR.....	40
	EKLER.....	43
	Ek 1 Program Hakkında Genel Bilgi .....	44
	ÖZGEÇMİŞ .....	55

## **SİMGE LİSTESİ**

*dB* Desibel

*Hz* Hertz

*W* Watt

## **KISALTMA LİSTESİ**

<i>DCT</i>	Discrete Cosine Transform – Ayrık Kosinüs Dönüşümü
<i>DFT</i>	Discrete Fourier Transform – Ayrık Fourier Dönüşümü
<i>EER</i>	Equal Error Rate – Eşit Hata Oranı
<i>FAR</i>	False Acceptance Rate – Hatalı Kabul Oranı
<i>FRR</i>	False Rejection Rate – Hatalı Ret Oranı
<i>FFT</i>	Fast Fourier Transform – Hızlı Fourier Dönüşümü
<i>GMM</i>	Gaussian Mixture Model – Gauss Karışım Modeli
<i>ICA</i>	Independent Component Analysis – Bağımsız Bileşen Analizi
<i>IDFT</i>	Inverse Discrete Fourier Transform – Ters Ayrık Fourier Dönüşümü
<i>k-NN</i>	k Nearest Neighborhood – k En Yakın Komşuluk
<i>MFCC</i>	Mel Frequency Cepstral Coefficients – Mel Frekanslı Cepstral Katsayıları
<i>PCA</i>	Principal Component Analysis – Temel Bileşen Analizi
<i>PCM</i>	Pulse Code Modulation – Darbe Kod Modülasyonu
<i>RADAR</i>	Radyo Detecting and Ranging – Radyo Algılama ve Mesafe Tayini
<i>RIFF</i>	Resource Interchange File Format – Kaynak Değişim Dosya Formatı
<i>SVM</i>	Support Vector Machines – Destek Vektör Makineleri
<i>VQ</i>	Vector Quantization – Vektör Kuantalama

## ŞEKİL LİSTESİ

	<b>Sayfa</b>
Şekil 2.1 Ses dalgalarının görüntüsü [Knight, 1960].....	5
Şekil 2.2 Sesin dalga boyu .....	6
Şekil 2.3 Ses dalgasının genliği.....	6
Şekil 2.4 Analog (sürekli) ses işareti .....	9
Şekil 2.5 Sayısal (ayrık) ses işareti.....	10
Şekil 2.6 Sayısal ses işaretinin kuantalanması ve kodlanması .....	10
Şekil 3.1 Sistemin genel yapısı.....	11
Şekil 3.2 Ön işlemlere ait blok diyagram.....	11
Şekil 3.3 Özellik çıkartımına ait blok diyagram.....	12
Şekil 3.4 Ses işaretinin çerçevesi .....	13
Şekil 3.5 Yaygın olarak kullanılan pencere fonksiyonları [National Instruments, 2006] .....	14
Şekil 3.6 Pencereleme etkisi.....	15
Şekil 3.7 Periyodik bir sinyalin fourier analizi ile elde edilmiş sinüs ve kosinüs bileşenleri...	15
Şekil 3.8 Örnek bir sinyalin DFT ile zaman domeninde ayrıştırılması .....	16
Şekil 3.9 On altı noktalı bir sinyal için bit tersleme işlemi.....	17
Şekil 3.10 Mel ölçekli filtre bankası.....	18
Şekil 3.11 $\Delta$ MFCC ve $\Delta^2$ MFCC hesaplanması.....	19
Şekil 3.12 k-ortalama yönteminin uygulanışı [Brazma] .....	20
Şekil 3.13 Sisteme ait genel çalışma yapısı .....	22
Şekil 3.14 Destek vektörlerinin oluşturulması [Burges, 1998] .....	23
Şekil 3.15 k-NN yöntemi ile sınıflandırmaya bir örnek [Claus, 2004] .....	24
Şekil 3.16 Eşik değerinin hatalı ret oranına etkisi .....	25
Şekil 3.17 Eşik değerinin hatalı kabul oranına etkisi .....	25
Şekil 3.18 Eşit hata oranı ile optimum eşik değerinin belirlenmesi .....	26
Şekil 4.1 Tanımda kullanılan sınıflar ve sınıf örnek sayılarını gösteren dendogram .....	27

Şekil 4.2 Giriş sinyali ve ön vurgulama filtresi uygulanmış hali (A_1_1.wav).....	28
Şekil 4.3 MFCC, $\Delta$ MFCC ve $\Delta^2$ MFCC vektörleri .....	37
Şekil 4.4 $\Delta$ MFCC ve $\Delta^2$ MFCC vektörleri ile genişletilmiş özellik vektörü .....	37

## ÇİZELGE LİSTESİ

	<b>Sayfa</b>
Çizelge 2.1 Günlük hayattan bazı sesler ve şiddet değerleri.....	8
Çizelge 2.2 Uzaklık ile ses şiddeti arasındaki ilişki .....	9
Çizelge 4.1 SVM yöntemine ait başarı oranları .....	29
Çizelge 4.2 k-NN yönteminde sınıflandırma başarı oranları .....	30
Çizelge 4.3 k-NN yönteminde alt sınıf başarı oranlarının sınıflara göre dağılımı .....	30
Çizelge 4.4 Hiyerarşik k-NN yönteminde sınıflandırma başarı oranları.....	31
Çizelge 4.5 Hiyerarşik k-NN yönteminde alt sınıf başarı oranlarının sınıflara göre dağılımı..	31
Çizelge 4.6 Farklı eşik seviyesi değerleri için hesaplanan hatalı reddetme oranları (FRR) .....	33
Çizelge 4.7 Farklı eşik seviyesi değerleri için hesaplanan hatalı kabul etme oranları (FAR) ..	34
Çizelge 4.8 k-NN yöntemi ile tanımaya bir örnek (Test dosyası: T_2_4.wav) .....	35
Çizelge 4.9 Tanımda kullanılan özellik sayısının tanıma başarısı ve çalışma süresine etkisi.	38



## **ÖNSÖZ**

Bu çalışmamda da her zaman olduğu gibi benden desteğini esirgemeyen ailem ve arkadaşlarıma, yardım ve yönlendirmelerinden dolayı danışman hocam saygı değer M. Elif KARSLIGİL'e, eğitim hayatımda üzerimde emeği bulunan tüm diğer değerli hocalarıma, çalışmada kullanılan ses bilgilerini sağlayan T.S.K Araştırma Merkez Komutanlığı'na teşekkürlerimi sunarım.

## ÖZET

Günümüzde ulaşım, savunma veya denetim amaçlı olarak RADAR (Radyo Detecting and Ranging – Radyo Algılama ve Mesafe Tayini) kullanılarak kontrol edilmektedir. Radarlar dalga halinde bir sinyal gönderilmesi ve bu sinyalin cisimden yansımaya bakılarak sinyali yansıtan cisim hakkında bilgi sahibi olmamızı sağlarlar. Buradan bir cismin tespiti işlemi için ilk hareketin, tespit merkezi tarafından yapılıyor olduğu anlaşılmaktadır. Bu da tespit edilen cismin de kendisine çarpan dalgayı ve dolayısıyla tespit merkezini algılayabileceğini göstermektedir.

Güvenlik açısından kritik savunma sistemlerinde, izlenen cisme izlenildiğinin fark ettirilmemesi büyük önem taşır. Cisme bir dalga göndermeksizin yaydığı sinyalleri kontrol ederek cisimle ilgili bir tespitte bulunulabilirse önemli bir avantaj sağlanmış olur.

Bu tezde, çalışma gürültüleri kullanılarak deniz taşıtlarını sınıflandıran bir sistem geliştirilmiştir. Deniz taşıtlarına ait motor, pervane, jeneratör gibi çalışma birimlerinin çalışma esnasında ortama yaydıkları sesler incelenerek karakteristikleri belirlenmiştir. Karakteristik özelliklerin elde edilmesinde MFCC (Mel Frequency Cepstral Coefficients – Mel Frekanslı Cepstral Katsayıları) yöntemi kullanılmış, belirlenen özellikler k-ortalama yöntemi ile indirgenerek sınıflandırmaya hazır hale getirilmiştir. Sınıflandırma işleminde SVM (Support Vector Machines – Destek Vektör Makineleri) yöntemi kullanıldığında alt sınıfların tanınmasında %47.61 gibi düşük bir başarı oranı elde edildiğinden; sınıflandırma işleminin, k-NN (k-Nearest Neighbor – k-En Yakın Komşuluk) yöntemi ile yapılmasına karar verilmiştir. Sınıflandırılması istenilen örnek için de özellik çıkartımı ve özelliklerin indirgenmesi işlemlerinin yapılmasının ardından k-NN yöntemi kullanılarak bilinmeyen geminin hangi gemi sınıfına ait olduğu bulunmuştur. Gemi veri tabanında olmayan bir geminin sorgulanması durumu da düşünülerek gemilerin bir sınıfa dâhil edilip edilmeyeceğine karar verilmesi için gemi veritabanındaki sesler ile sınıflandırılmak istenilen gemi sesi arasındaki benzerlik için bir eşik değeri belirlenmiş bu eşik değerinin altında kalan sesler için sınıflandırma yapılmamıştır.

Geliştirilen gemi tanıma sistemi bir sınıfa dâhil olan 42 gemi sesi örneği için ana sınıflandırmada %85.71 alt sınıflandırmada %71.43 oranında başarı sağlamıştır. Seçilen uygun eşik değeri için FRR (False Rejection Rate – Hatalı Ret Oranı) %12 olarak gerçekleşmiştir. Sistem ayrıca bir sınıfa dâhil olmayan, sınıflandırma dışı 26 ses örneği için aynı eşik değeri ile hatalı kabule karşı da test edilmiştir ve bu sesler için FAR (False Acceptance Rate – Hatalı Kabul Etme Oranı) %12 olarak hesaplanmıştır.

## **ABSTRACT**

Nowadays, either for supervision or for defense purposes, the transportation is being controlled by RADAR (Radio Detecting and Ranging). RADAR helps us get information about objects of interest by transmitting electromagnetic radio waves to the target and analyzing the reflecting signals. We understand that in order to determine any given object's information (speed, distance etc) the first action is always taken by the control center. This also tells us that the moving object can also notice the electromagnetic radio waves and hence the control center can also be sensed.

It is of great importance in the defense systems that the object under inspection should not notice being recognized. If we could get information about the target by controlling the waves sent by the target rather than sending any signals to it we could be on the advantageous side.

In this thesis the vehicles were classified by the noise they make while operating. Working units of the marine vehicles (engine, fan, the generator and etc) analyzed and the sounds they diffuse to the environment while operating were characterized. In the characteristics determination the MFCC (Mel Frequency Cepstral Coefficients) method used to determine the features of the vehicles and by the VQ (Vector Quantization) method, the defined features degraded to a specific number. Due to the low success ratio (47.61%) of SVM method for the classification within sub classes, k-NN (k-Nearest Neighbor) method selected for the classification and recognition of the vehicles. A threshold value has been determined to validate if any query occurs for a ship that does not exists on the train set. If a ship sound has got fewer score than the threshold then that ship described as unrecognizable by the system and system does not make any classification for that sound.

The system developed in this study tested with 42 different ships sounds and got 85.71% success ratio for the main classification and 71.43% for the sub classification. For these 42 sounds that subject to one of existing class, FRR (False Rejection Rate) occurred as 12%. System also tested with 26 impostor ship sounds for the false acceptations, FAR (False Acceptance Rate) occurred as 12% too.

## 1. GİRİŞ

Bu tez çalışmasında seyir halindeki bir geminin ve üzerinde çalışan tüm makinelerin (motor, pervane, jeneratör vb.) ortama yaydıkları sesleri değerlendirerek gemi modelini tanıyan bir sistem geliştirilmiştir. Hava gibi az yoğun gaz ortamında sesin yayılma hızı yavaş olacağından ve sesteki hızlı jet gibi hava taşıtlarının olması nedeniyle hava taşıtlarında sistemin kullanılamayacağı açıktır. Deniz taşıtlarının gerek hızlarının yavaş olması gerekse sesin yoğun ortamlarda daha hızlı yayılması sonucunda böyle bir kaynağa bağlı tanıma sisteminin deniz taşıtlarının sınıflandırmasında uygulanabilir olduğundan söz edilebilir.

Ses kaynağı tanıma, nesnelere çıkardıkları seslerden yola çıkarak sınıflandırma ya da kimliğini belirleme işlemidir. Yolda yürürken arkamızdan gelen bir araç korna çaldığında duyduğumuz taşıt gürültüsü ve korna sesini birleştirerek arkamızı dönmeden de yaklaşan taşıtın bir otomobil, kamyon ya da motosiklet olduğunu algılayabiliriz. Benzer şekilde bir arkadaşımız adımızı seslendiğinde onun kim olduğunu sesinden algılayabilir, telefonda daha önce konuştuğumuz bir kişiyi yine sesinden tanıyabilir, radyoda çalan şarkıyı kimin seslendirdiğini kolaylıkla söyleyebiliriz.

Ses kaynağı tanıma alanında insan sesini tanımaya yönelik olarak yapılan çalışmalar bilimsel literatürde büyük ilgi ve kabul görmüştür ve bir ses kaynağının tanımlanmasına örnek olarak verilir [Reynolds, 1995].

Konuşma sesi tanıma sistemlerinde kullanılan özellikler genellikle doğrusal tahmin ve ayrık fourier dönüşümü üzerine kurulmuş kepsstral katsayılarıdır ve bazen bu kepsstrallerin birinci ve ikinci dereceden türevleri de kullanılır. İnsan sesi tanımanın amacı kişi tanıma veya iddia edilen sesin o kişiye ait olup olmadığını anlamak olabilir. Eğer kelimelerin önemi söz konusu ise tanıma işi yazı bağımlı veya bağımsız olabilir. Bu alanda en çok kullanılan istatistiksel yöntem GMM (Gaussian Mixture Model – Gauss Karışım Modeli) yöntemidir [Eronen, 2001].

Konuşma sesi tanıma sistemlerinin performansı, test ortamının değişen akustik durumundan veya konuşma sinyallerine karışan sesler yüzünden düşebilir [Murthy & Beaufays & Heck & Weintraub, 1999][Alonso-Martinez & Faundez-Zanuy, 2000].

Ses ve ses kaynağı tanıma üzerine müzik alanında da müzik çözümlemesi şeklinde birçok çalışma yapılmıştır. Müzik çözümlemesi bir müzik parçasında hangi notanın, hangi enstrüman tarafından ne uzunlukta çalındığının ortaya konulması işlemidir.

Genel olarak birden fazla enstrüman tarafından veya gitar ya da piyano gibi tek bir polifonik enstrümanla çalınan bir müzikte yapılan iş, bir polifonik ses perdesinin izlenmesidir. Bu sestem geriye dönerek, müziği oluşturan seslerin doğru şekilde ayrıştırılabilmesi normal bir insan için zor olan ve engin müzik tecrübesi gerektiren bir iştir.

Buna karşın, müzik çözümlemesi müzik teorisyenleri ve müzisyenler için önemli bir araç olduğundan bu konuda otomatik tanıma işlemi yapabilen araçlara sahip olmak önem taşımaktadır. Polifonik iz takibi işlemiyle ilgili yapılan çalışmalarda kaliteli ses kaynakları (müzikal enstrümanlar) kullanıldığında çözümleme işleminin daha basit bir hale geldiği görülmüştür [Kashino & Murase, 1998].

Ses ve ses kaynağı tanıma üzerine yapılan çalışmaların bir kısmı da taşıtların sınıflandırılması amacıyla gerçekleştirilmiştir.

Wu ve arkadaşları yaptıkları çalışma ile taşıtları sınıflandırmaya çalışmışlardır. Taşıtların sınıfları olarak seçtikleri araba, kamyon ve motosiklet seslerini tanımak için PCA (Principal Component Analysis – Temel Bileşen Analizi) yöntemi ile güç spektrumundan türettikleri özellikleri kullanmışlar ve farklı sınıflardaki sesleri küçük bir veritabanı kullanarak kümeleyebilmişlerdir [Wu & Siegel & Khosla, 1998].

Jarnicki ve arkadaşları tarafından yapılan çalışmada askeri taşıtlar, nakliye araçları ve sivil taşıtların sınıflandırılması yapılmıştır. Sınıflandırma filtre bankası analizi kullanılarak belirlenen sınıflar için %90'ın üzerinde bir başarı oranı ile gerçekleştirilmiştir [Jarnicki & Mazurkiewicz & Maciejewski, 1998].

Taşıtların hareket halinde etrafa yaydıkları seslerin kullanılarak sınıflandırıldığı bir başka çalışma da Case Gaunard tarafından gerçekleştirilmiştir. Gaunard, sistemi yaklaşık 100 ms uzunluğunda kaydedilen araba, kamyon, motosiklet, uçak ve tren sesleri ile eğitmiş, özellik çıkartımı aşamasında MFCC işlemi sonucunda ortaya çıkan 12 kepsral katsayıyı, özelliklerin sınıflandırılmasında da GMM yöntemini kullanmış ve sınıflandırmada %90 gibi yüksek bir başarı oranı elde etmiştir [Gaunard, 1998].

Deniz taşıtlarının sınıflandırılması alanında Mahmut Alkan tarafından yapılan çalışmada taşıtların çalışmaları esnasında etrafa yaymış oldukları gürültüler giriş verisi olarak değerlendirilerek MFCC katsayılarından elde edilen özellikler kullanılarak sınıflandırma işlemi yapılmıştır. [Alkan, 2005].

Bu çalışmada deniz taşıtlarının seyir halinde yaymış oldukları: pervane, jeneratör ve diğer ses çıkartan teçhizatlarının karakteristik özellik taşıyan toplam ses bilgisi kullanılmıştır. Özellik çıkartımında MFCC işlemi sonucunda ortaya çıkan 12 kepsral katsayının yanı sıra bu 12 kepsral katsayının zaman içindeki değişimini gösteren  $\Delta$ MFCC ve  $\Delta^2$ MFCC değişim değerleri de yardımcı özellikler olarak seçilerek sınıflandırmada 36 özelliğin kullanılması sağlanmıştır. Elde edilen özellikler kuantalama ile sabit bir sayıya indirgendikten sonra, sınıflandırma başarısını arttırmak amacıyla k-NN yöntemi hiyerarşik şekilde uygulanarak deniz taşıtlarının sınıflandırması yapılmıştır.

İlerleyen bölümlerde çalışma sesinden gemi tanınmasında izlenen yol detaylı olarak anlatılacak, başarı oranları ve ileride yapılabilecek benzer çalışmalarda kolaylık sağlayabileceği düşünülen bilgilere yer verilecektir.

## 2. SES HAKKINDA GENEL BİLGİ

Bu bölümde sesin tanımı, ses dalgalarının fiziksel özellikleri ve analog sesin sayısal ortama aktarılması ile ilgili bu çalışmaya temel olan bilgiler verilecektir.

### 2.1 Sesin Tanımı

Ses, iletken bir ortamda bir etken aracılığıyla meydana gelen mekanik titreşimlerin dalgalar halinde yayılarak işitme duyumunun sağlanmasıdır.

Ses, hava ya da diğer iletici ortamlarda bir dalga biçiminin ya da hareketinin kulak tarafından algılanmasıdır.

Ses esnek ortamlarda iletilen basınç ya da hızdaki değişim sonucu, kulaktan geçerek üretilen işitsel duydur.

Ses, titreşimlerden oluşan bir enerji türüdür [Wikipedia].

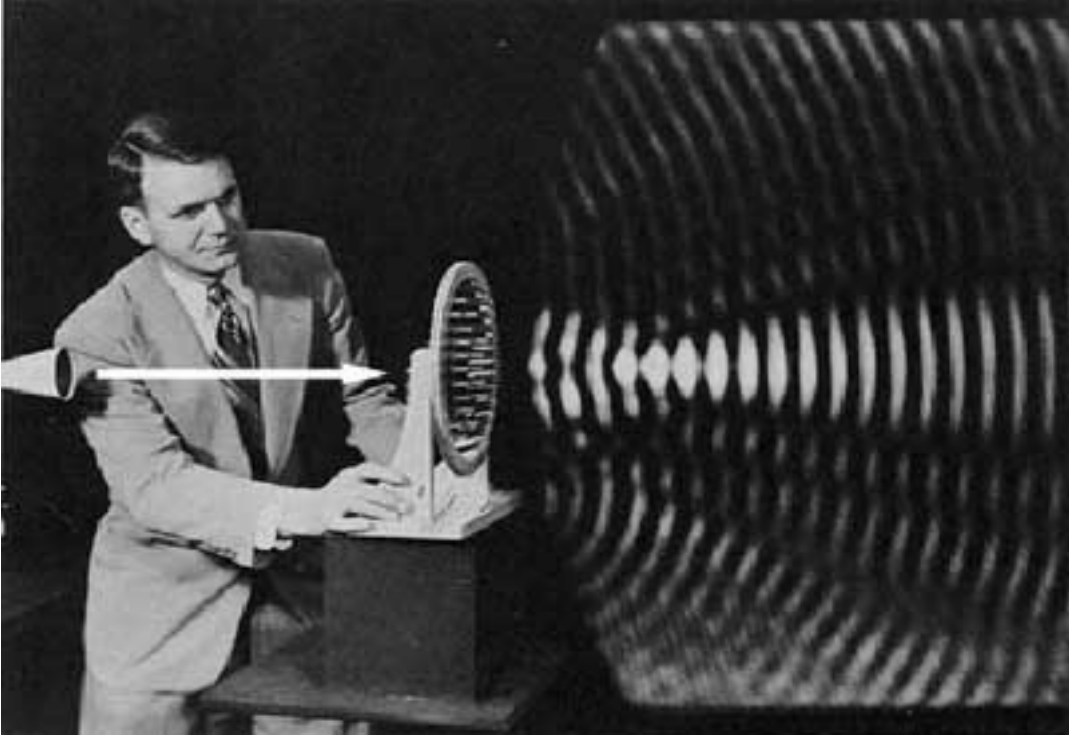
Bu tanımlardan yola çıkarak ses, iletken bir ortamda herhangi bir etken aracılığıyla oluşan titreşimlerin, hareket eden dalgaların oluşturduğu fiziksel bir nicelik olarak tanımlanabilir. Bu tanımlamada sistemin değişebilen parçalarını kaynak, iletim ortamı ve algılayıcı ortam oluşturur.

Kendiliğinden ses çıkaran varlıklara doğal ses kaynakları denir. Ses kaynaklarının ses çıkarabilmesi için titreşim gerekir. İnsanların ses çıkarması da titreşimle olur. Gırtlığımızda bulunan ses telleri akciğerlerimizde bulunan hava ile titreşerek ses çıkmasını sağlarlar.

### 2.2 Ses Dalgaları

Dalgalar genel olarak, mekanik ve elektromanyetik dalgalar olmak üzere iki ana gruba ayrılır. Elektromanyetik dalgalar, yayılmak için bir ortama ihtiyaç duymazlar ve boşlukta da yayılabilirler. Mekanik dalgalar ise, enerjilerini aktarabilmek için ortam taneciklerine ihtiyaç duyarlar. Bu yüzden boşlukta (örneğin uzayda) yayılamazlar. Ses dalgaları da mekanik dalgalar olduklarından yayılmak için maddesel bir ortama ihtiyaç duyarlar. Ses, nesnelere titreşiminden meydana gelen ve uygun bir ortam içerisinde (hava, su vb.) bir yerden başka bir yere, sıkışma (compressions) ve genleşmeler (rarefactions) şeklinde ilerleyen bir dalgadır; dolayısıyla ses, bir basınç dalgasıdır.

Bell Telefon laboratuvarlarında 1960 tarihinde çekilen fotoğrafta, özel bir ses merceği ve özel bir görüntüleme yöntemi kullanılarak, resimde solda görülen kornadan çıkan ses dalgalarının görüntüsü elde edilebilmiştir [Knight, 1960] (Şekil 2.1).



Şekil 2.1 Ses dalgalarının görüntüsü [Knight, 1960]

Sesin iletimi sırasında kullanılan ortamın karakteristik özellikleri sesin hızını ve menzilini belirler. Sesin suda yayılmasına etki eden üç ana faktör mevcuttur. Bunlar sıcaklık, yoğunluk (tuzluluk) ve basınç (derinlik) değerleridir. Örneğin %0.1'lik bir tuz oranı artışı ses hızında 1.2 m/sn'lik bir artışa, 30 metrelik derinlik artışı 6 m/sn'lik ses hızı artışına, 1 derecelik sıcaklık artışı da 3m/sn'lik ses hızı artışına neden olur.

Aşağıdaki bölümlerde ses dalgalarının fiziksel özellikleri incelenmiştir.

### 2.2.1 Frekans (Sıklık)

Bir dalganın frekansı, dalganın hava veya başka bir ortam içinden geçerken ortamdaki parçacıklarının ne sıklıkta titreştiğine bağlıdır. Frekans ileri geri titreşimlerin zamana bağlı olarak ölçülmesi ile hesaplanır. Saniyedeki titreşim sayısı özel olarak Hertz birimi ile ifade edilir ve kısaca f ile gösterilir. Bir ses dalgasının bir devri tamamlaması için gerekli zamana ses dalgasının periyodu denir ve T ile gösterilir.

$$T = 1 / f \quad (2.1)$$

Yüksek frekans değerleri için Hertz'in bin katı olan kilohertz (kHz) birimi kullanılır. İnsan kulağının duyabildiği sesler 20 ile 20000 Hz (20kHz) aralığında frekanslara sahip olabilir. Frekans değeri 20 Hz'in altında olan titreşimlere ses altı (infrasonik) titreşimler, 20 kHz'in üzerinde olan titreşimlere sesüstü (ultrasonik) titreşimler denilmektedir [Başaran, 1981].



### 2.2.2 Dalga boyu

Bir dalganın ardışık iki tepe veya iki çukur noktası arasındaki mesafe bize dalga boyunu verir. Bir başka deyişle dalga boyu bir periyotluk süre içinde ses dalgasının kat ettiği yoldur. Dalga boyu  $\lambda$  ile gösterilir [Hassall & Zaveri, 1979].

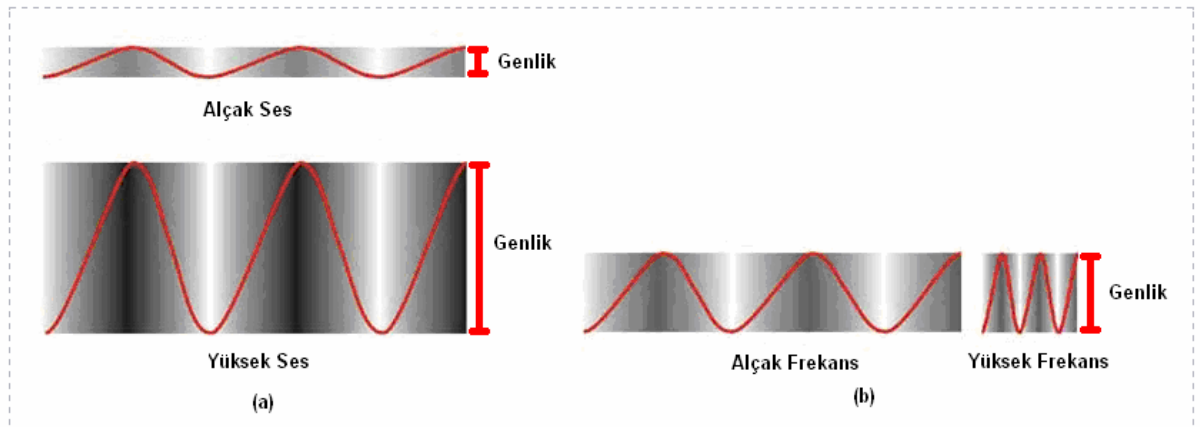
$$\lambda \text{ (Mt)} = V \text{ (340 Mt / sn)} / f \text{ (Hz)} \quad (2.2)$$



Şekil 2.2 Sesin dalga boyu

### 2.2.3 Genlik

Genlik, ses dalgalarının dikey büyüklüğünün bir ölçüsüdür. Ses dalgalarını oluşturan sıkışma ve genleşmeler arasındaki fark, dalgaların genliğini belirler. Ses dalgaları havada veya başka bir ortamda titreşen objeler tarafından üretilir. Örneğin titreştirilen bir gitar teli, yaptığı periyodik salınım hareketi ile hava moleküllerinin belli bir frekansta sıkışmasını ve genleşmesini sağlar. Bu şekilde teldeki enerji havaya iletilmiş olur. Enerjinin miktarı, teldeki titreşim genliğine bağlıdır. Eğer tele fazla enerji yüklenirse, tel daha büyük bir genlikle titreşir. Teldeki titreşim genliği ne kadar fazla ise ortam tanecikleri (örneğin hava molekülleri) tarafından taşınan enerji de o kadar fazladır. Enerji ne kadar fazla ise sesin şiddeti de o kadar büyük olacaktır. Bu ifadeler, titreşen tüm cisimler için geçerlidir.



Şekil 2.3 Ses dalgasının genliği

Şekil 2.3’de koyu renkli bölgeler sıkışmaları, açık renkli bölgeler ise genleşmeleri simgelemektedir. Eğriler ise bu sıkışma ve genleşmelerin iki boyutlu grafiksel temsilleridir. Dikkat edilirse, sıkışma miktarı arttıkça (yüksek seste olduğu gibi) sesin şiddeti de artmaktadır. Aynı frekansa sahip seslerden genliği büyük olan kulağa daha şiddetli olarak ulaşırken (a); aynı genlik değerine sahip seslerden frekansı yüksek olan kulağa frekansı düşük olanla aynı şiddette fakat daha çok sayıda ulaşır (b).

#### 2.2.4 Ton

Müzikte, dia tonik (doğal majör) gamda bir tam aralık olarak tanımlanan ton, belli bir frekansta ve perdede üretilen saf ses anlamında kullanılır. Saf tonlar doğal ortamda fazla karşılaşılmayan ve genellikle müzik aletleri veya ses üreteçleri aracılığıyla üretilen seslerdir. Yüksek frekanslı (yüksek perdeden) sesler tiz, düşük frekanslı (düşük perdeden) sesler pes (bas) olarak algılanılır [Wikipedia].

#### 2.2.5 Tını

İnsan kulağının logaritmik işitmesi nedeniyle kulak tarafından açığa çıkarılan ses dalgalarına harmonik denir. Harmonikler birleşerek tınıyı oluşturur. Tını, sesin rengini ifade eden bir terimdir. Aynı oktavda, aynı notayı (tonu) aynı yoğunlukta ve aynı uzunlukta çalan bir kemanla bir flüt arasındaki temel fark, tını farkıdır. Enstrümanları oluşturan bileşenlerin doğal frekanslarındaki farklılıklar, sonuçta oluşan sesin farklı bir tınıda olmasını sağlar. Bu sayede, farklı müzik aletlerinden çıkan özdeş notaları kolaylıkla ayırt edebiliriz.

#### 2.2.6 Sesin şiddeti ve desibel ölçeği

Şiddet, ses dalgalarının taşıdıkları enerjiye bağlı olarak birim alana uyguladıkları kuvvettir. Birimi genellikle metrekare başına Watt ( $W/m^2$ ) olarak ifade edilir. Sesin şiddeti, ses kaynağına olan uzaklığın karesi ile ters orantılıdır. İnsan kulağı çok düşük ve çok yüksek şiddette sesleri duyabilme yeteneğine sahiptir. İnsan kulağının algılayabileceği en düşük ses şiddeti, eşik şiddet olarak bilinir. Kulağa zarar vermeden işitilebilen en yüksek sesin şiddeti ise, eşik şiddetinin yaklaşık bir milyon katı kadardır. İnsan kulağının şiddet algı aralığı bu kadar geniş olduğundan, şiddet ölçümü için kullanılan ölçek de 10'un katları, yani logaritmik olarak düzenlenmiştir. Buna desibel ölçeği adı verilir. Sıfır desibel mutlak sessizliği değil; işitilemeyecek kadar düşük ses şiddetini (ortalama  $1 \times 10^{-12} W/m^2$ ) gösterir.

Desibel, bir oranı veya göreceli bir değeri gösterir ve bel biriminin 10 katıdır. Alexander Graham Bell'in anısına bel adı verilen birim, iki farklı büyüklüğün oranının logaritması olarak tanımlanmaktadır. Yani 1 bel, birbirlerine oranları 10 olan iki büyüklüğü göstermektedir

(örneğin 200 / 20). Bu oranın çok büyük olmasından dolayı Desibel adı verilen ve oranların logaritmasının 10 katı olarak tanımlanan birim daha yaygın olarak kullanılmaktadır. Bu sayılardan biri bilinen bir sayı olarak alındığından, Desibel; söz konusu bir büyüklüğün ( $P_i$ ) referans büyüklüğe ( $P_{ref}$ ) oranının logaritmasının 10 katıdır (2.3) [Henderson, 1998].

$$dB = 10 \times \log (P_i / P_{ref}) \quad (2.3)$$

dBA ise insan kulağının en çok hassas olduğu orta ve yüksek frekansların özellikle vurgulandığı bir ses değerlendirmesi birimidir. Gürültü azaltması veya kontrolünde çok kullanılan dBA birimi, ses yüksekliğinin öznel değerlendirmesi ile ilişkili bir kavramdır. Eşik şiddetindeki ses sıfır desibeldir ve  $1 \times 10^{-12} \text{ W/m}^2$  değerine eşdeğerdir. 10 kat daha şiddetli ses  $1 \times 10^{-11} \text{ W/m}^2$ ; yani 10 dB iken, 100 kat daha şiddetli ses 20 dB'dir. Çizelge 2.1'de günlük hayatta sıklıkla karşılaştığımız bazı ses kaynakları ve bunların ürettiği seslerin desibel olarak şiddetleri karşılaştırma amacıyla verilmiştir [Henderson, 1998].

Kaynak	Şiddet ( $\text{W/m}^2$ )	dB
Eşik şiddeti	$1 \times 10^{-12}$	0
Yaprak hışırtısı	$1 \times 10^{-11}$	10
Fısıltı	$1 \times 10^{-10}$	20
Normal konuşma	$1 \times 10^{-6}$	60
Caddedeki yoğun trafik	$1 \times 10^{-5}$	70
Elektrik süpürgesi	$1 \times 10^{-4}$	80
Büyük orkestra	$6,3 \times 10^{-3}$	98
Walkmanın en yüksek sesi	$1 \times 10^{-2}$	100
Rock konserinin ön sırası	$1 \times 10^{-1}$	110
Jet uçağının kalkışı	$1 \times 10^2$	140
Kulak zarı hasarı	$1 \times 10^4$	160

Çizelge 2.1 Günlük hayattan bazı sesler ve şiddet değerleri

Ses dalgaları enerjilerini üç boyutlu ortamda taşıırken, kaynaktan uzaklaştıkça ses dalgalarının şiddeti azalır. Artan uzaklıkla birlikte ses dalgalarının şiddetinin azalması ses dalgalarındaki enerjinin daha geniş alanlara yayılmasından kaynaklanır. Ses dalgaları iki boyutlu bir ortamda dairesel olarak yayılır. Enerji korunduğu için enerjinin yayıldığı alan arttıkça güç azalmalıdır. Şiddet ve uzaklık arasında ters-kare ilişkisi vardır. Bu yüzden kaynağa olan uzaklık 2 katına çıktığında şiddet  $\frac{1}{4}$  'üne düşer. Benzer şekilde kaynağa olan uzaklık  $\frac{1}{4}$  'üne düştüğünde şiddet 16 katına çıkar. Uzaklık arttıkça sesin şiddeti, uzaklığın karesi oranında azalır. Uzaklık ile ses şiddeti arasındaki ilişki Çizelge 2.2'de gösterilmiştir [Henderson, 1998].

Uzaklık (m)	Şiddet (birim)
1	160
2	40
3	17.8
4	10

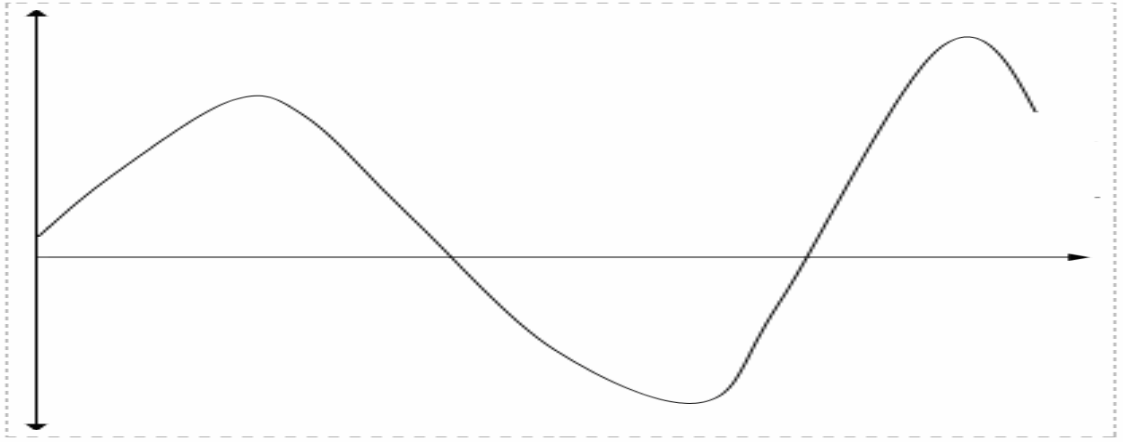
Çizelge 2.2 Uzaklık ile ses şiddeti arasındaki ilişki

## 2.3 Sayısal Ses

Sesin bilgisayar ortamında işlemeye önce, havadaki basınç değişimlerini ifade eden ve süreklilik arz eden analog ses işaretinin, bilgisayarın anlayabileceği ayırık olan sayısal işaretlere çevrilmesi gerekir. Bu nedenle bir sensor ile basınçtaki değişimler elektrik sinyallerine, analog sayısal dönüştürücü ile de bu elektrik sinyalleri sayısal ifadelerle dönüştürülür.

### 2.3.1 Analog-Sayısal dönüşüm

Analog sayısal dönüştürücüler ile sonsuz noktada tanımlı olan elektriksel işaret (Şekil 2.4) ayırık hale getirilmeli yani sonlu sayıda noktada tanımlı olması sağlanmalıdır.



Şekil 2.4 Analog (süreklili) ses işareti

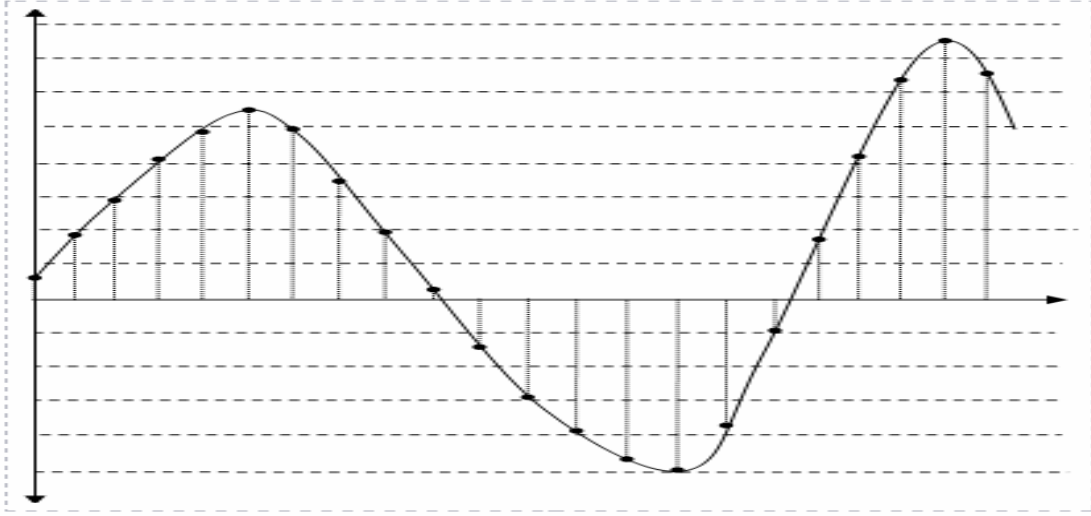
### 2.3.2 Örnekleme teoremi

Nyquist-Shannon örnekleme teoremi, en büyük frekans bileşeni  $f_{\max}$  olan sürekli bir sinyalin, bu en büyük frekansın iki katından daha büyük  $f_s$  frekanslı bir sinyalde zamanda ayırık örneklenecek herhangi bir veri kaybı olmaksızın geri çağrılabilmesini ifade eder (2.4) [Nyquist, 1928] [Shannon, 1949].

$$f_s \geq 2 \times f_{\max}$$

(2.4)

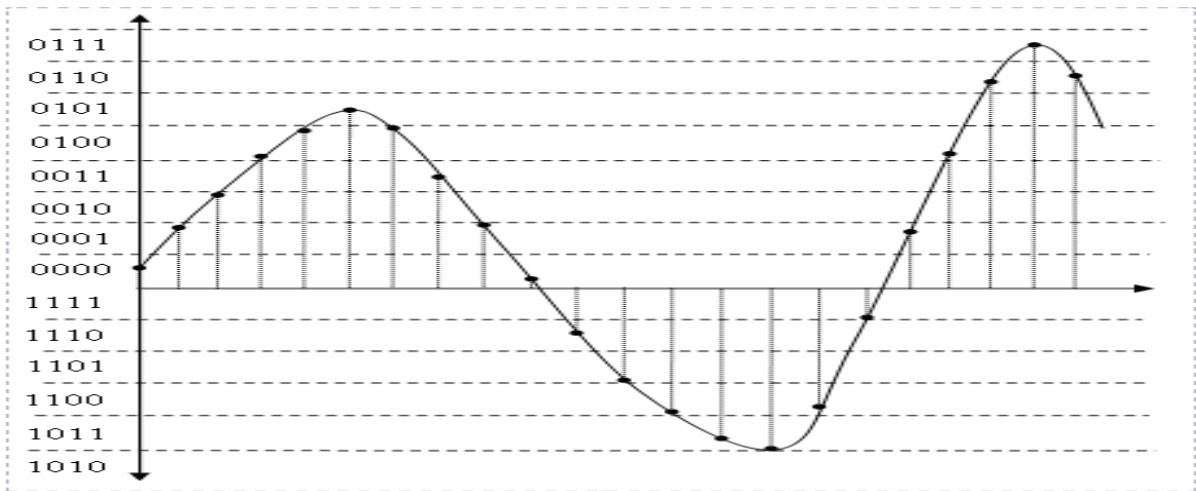
Şekil 2.5’de analog ses işaretinin, örnekleme teoremine uygun aralıklardaki değerleri alınarak ses işareti ayrık hale getirilmiştir.



Şekil 2.5 Sayısal (ayrık) ses işareti

### 2.3.3 Kuantalama ve kodlama

Bir fonksiyonun alabileceği değerler kümesinin elemanları ne kadar fazla olursa, bu değerleri kodlamak için ihtiyaç duyulan depolama alanı da o kadar artar. Örneklenen sinyalin her bir örnekleme anındaki genlik değerinin kaç farklı seviye ile belirleneceğini göstermek için birbirine yakın olan değerler ortak bir sayısal işarete yuvarlanılır. Kuantalama yapılırken ne kadar fazla depolama alanı kullanılırsa, yuvarlamalardan kaynaklanan kuantalama hatası da o kadar düşük olacaktır. PCM (Pulse Code Modulation – Darbe Kod Modülasyonu) kodlamada kuantalama seviyeleri 16 bit ile ifade edilir; bu da  $2^{16} = 65536$  kuantalama seviyesi kullanılabildiğini ifade eder.

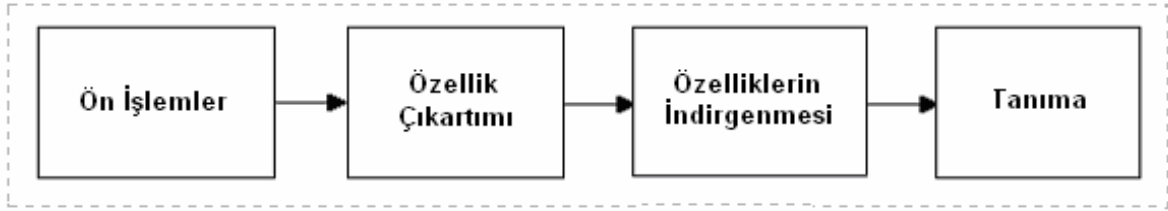


Şekil 2.6 Sayısal ses işaretinin kuantalanması ve kodlanması

### 3. SES ÖZELLİKLERİNDEN GEMİ TANIMA SİSTEMİNE GENEL BAKIŞ

Bu çalışmada deniz taşıtlarının çalışma esnasındaki yaymış oldukları sesler kullanılarak seyir halinde olan deniz taşıtlarının sınıflandırılmaları yapılacaktır. Deniz taşıtlarının hareketleri sırasında yayılan bu toplam ses gemiye ait karakteristik bir özelliktir. Bu nedenle deniz taşıtlarının çalışmaları esnasında yaymış oldukları: pervane, jeneratör ve diğer ses çıkartan teçhizatlarının yaymış oldukları toplam ses kullanılarak taşıtların karakteristik özelliklerinin belirlenmesi ile bu taşıtlar hakkında bir fikir sahibi olunabilmesi amaçlanmıştır.

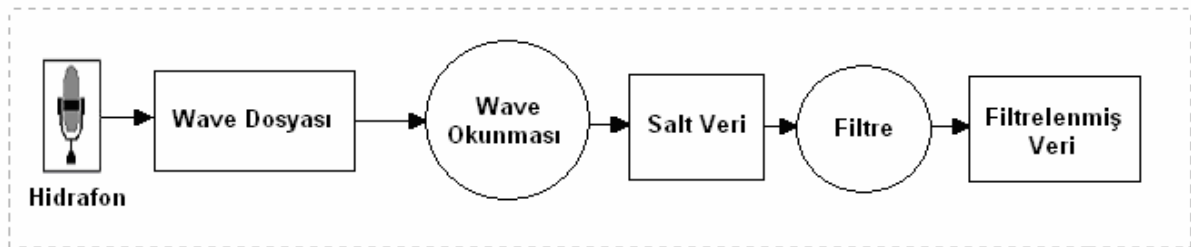
Sistemde, önce bu toplam çalışma sesi kaydedilmiş, kayıt üzerinde iyileştirme çalışmaları yapıldıktan sonra; iyileştirilmiş kayıt üzerinden özellik çıkartımı işlemi yapılmıştır. Ancak çıkartılan özellik sayısı hem fazla yer kapladığından hem de ses verisinin uzunluğuna bağlı olarak değişen boyutta olduğundan bulunan özelliklerin sabit sayıdaki geneli ifade eden kısmı seçilmiş ve son olarak da özelliklerin sınıflandırılmasıyla tanıma işlemi tamamlanmıştır. Sistemin genel yapısı Şekil 3.1’de gösterilmiştir.



Şekil 3.1 Sistemin genel yapısı

#### 3.1 Ön İşlemler

Sesin dış ortamdan alınarak, özellik çıkartımında kullanılmaya hazır hale getirilmesine kadar yapılan işlemlerin tamamı ön işlemler olarak nitelendirilmiş ve blok diyagramı Şekil 3.2’de gösterilmiştir.



Şekil 3.2 Ön işlemlere ait blok diyagram

Taşıtlarının sınıflandırılmasında kullanılacak olan sesin kaydedilmesinde, basınçlı ve tuzlu sularda ses kaydetmeye imkân sağlayan özel bir mikrofon türü olan hidrafonlar kullanılmıştır.

Hidrafonlardaki analogdan sayısal dönüşümlerin ardından ses, bilgisayar ortamında wave formatında saklanmaktadır.

### 3.1.1 Wave formatlı giriş dosyasının okunması

Wave dosya formatı Microsoft ve IBM tarafından ortaya konulmuş RIFF'in (Resource Interchange File Format – Kaynak Değişim Dosya Formatı) bir alt setidir. Wave dosyası, dosyanın wave formatı olduğunu gösteren ana kısım, wave dosyasının anlamlandırılması ile ilgili bilgi veren başlık kısmı ve ses bilgisini içerisinde barındıran veri kısmı olmak üzere üç ana bölümden oluşur [EBU-UER, 2001]. Wave dosyası, wave başlığı kısmında verilen bilgilerin ışığında okunularak içerisinde kodlanmış olan sayısal ses bilgisine erişilir.

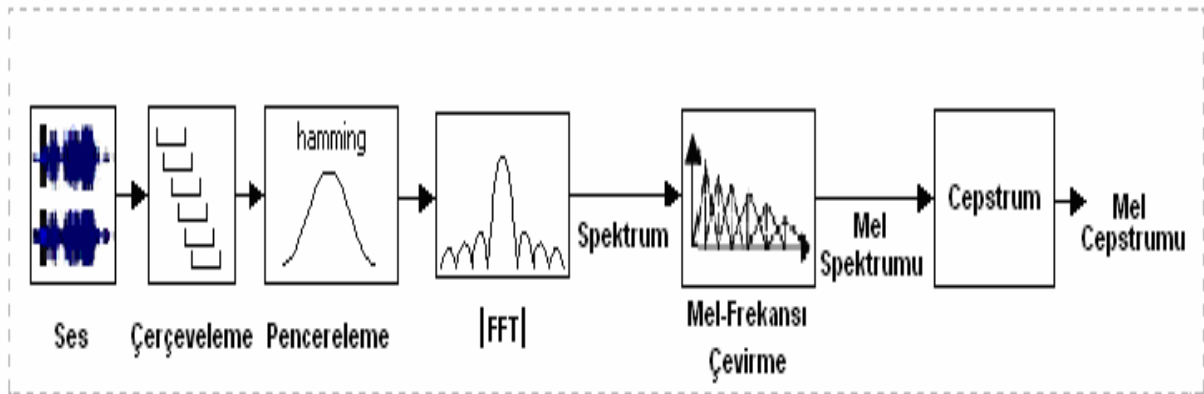
### 3.1.2 Sesin filtrelenmesi

Hidrafonlar aracılığı ile sesin kaydedilmesi sırasında kaydedilmek istenen ana sesin üzerinde parazit etkisi sonucunda oluşan istenmeyen gürültüler de kaydedilebilir. Kullanılan ön vurgulama (pre-emphasis) filtresi ile hem özellik çıkarmada kullanılacak olan yüksek frekanstaki sesler güçlendirilmiş hem de işlem sırasında (t) anındaki ses değerinin anlamlandırılmasında (t-1) anındaki ses verisi de kullanıldığından eğilime uygun olmayan, kısa süreli olarak var olan, parazit nedenli anlık yükselme ve düşmeler de engellenmiştir (3.1) [Robinson, 1995].

$$y[n] = x[n] + \alpha * x[n-1], \quad \alpha = -0.97 \quad (3.1)$$

## 3.2 Özellik Çıkartımı

Özellik çıkartımı, ön işlemler sonucunda elde edilen veriden tanıma aşamasında kullanılmak üzere sese ait karakteristik bilginin seçilmesi işlemidir. Özellik çıkartımı sırasında yapılan işlemler Şekil 3.3'de gösterilmiştir.



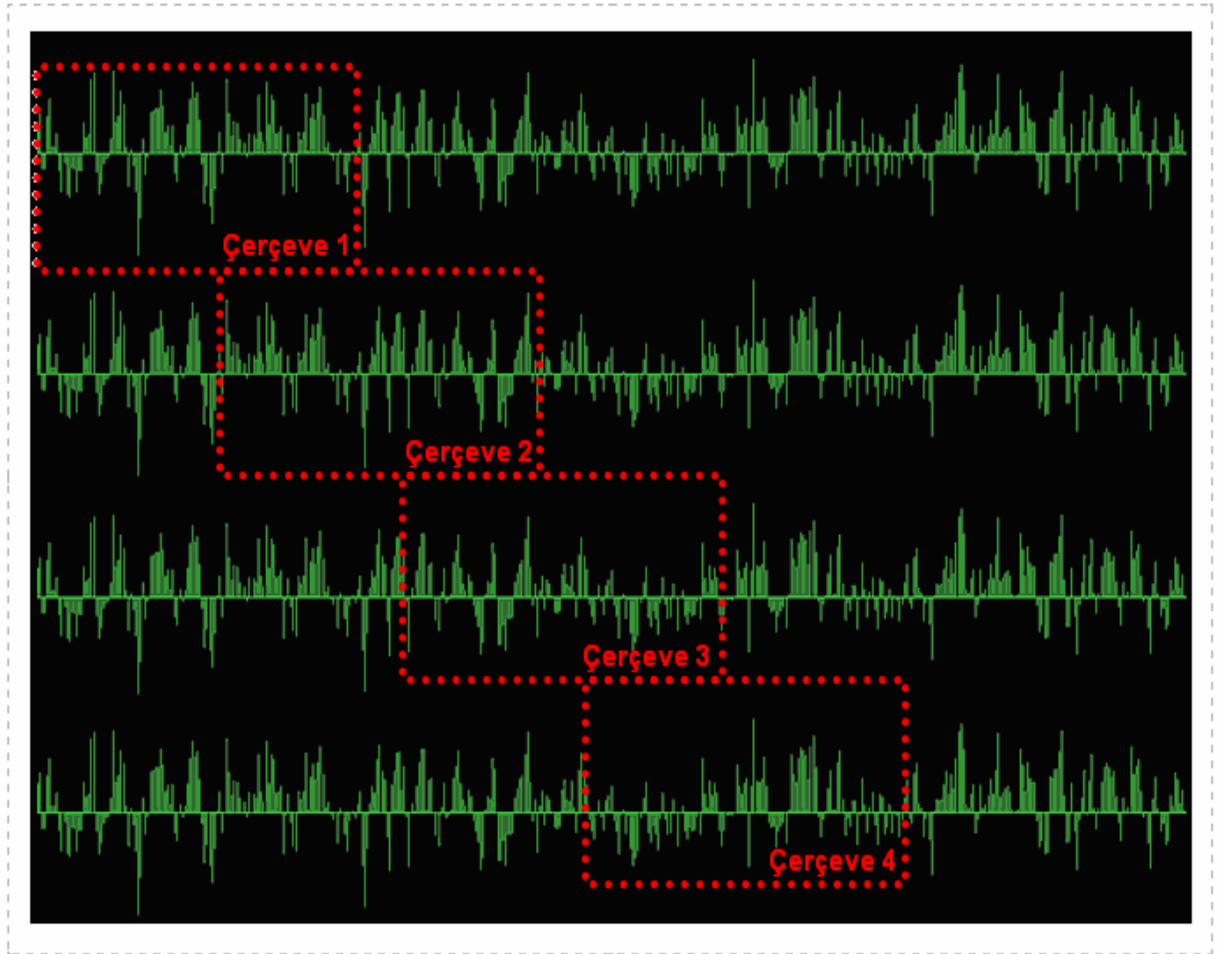
Şekil 3.3 Özellik çıkartımına ait blok diyagram

### 3.2.1 Çerçeveleme

Ses işareti doğası nedeniyle parazit etkiler dışında zamanla yavaş değişimler gösterir. Bu nedenle ses işareti 20-30 milisaniyelik zaman dilimlerinde incelenilerek sese ait akustik karakterler yeterli bir şekilde ortaya çıkartılabilir. Büyük bir yapının, küçük parçalar halinde ele alınarak analizinin yapılması işlemi kısa süreli analiz olarak adlandırılır.

Bu aşamada kısa süreli analizlerin yapılabilmesi için sürekli ses sinyali  $N$  örnekten oluşan çerçevelere ve  $M < N$  olmak üzere komşu  $M$  örnekten oluşan çerçevelere bölünür [Davis & Mermelstein, 1980]. İlk çerçeve  $N$  örnekten oluşurken ikinci çerçeve ilk çerçeveden  $M$  örnek sonra başlar ve ilk çerçevenin  $N - M$  elemanı kadar üzerine biner (Şekil 3.4). Buna sinyalin bir kısmının farklı çerçevelerde örtüşmesi denilir.

Çerçevelerin örtüştürülmesi, başka bir deyişle üst üste bindirilmesi, çerçeveleme işleminde çerçevenin bitiş sınırında yer alan ses bilgisi ile bir sonraki çerçevenin başlangıç kısmında yer alan ses bilgisi arasındaki geçişte ortaya çıkan değişim bilgisinin kaybedilmemesi için gereklidir.



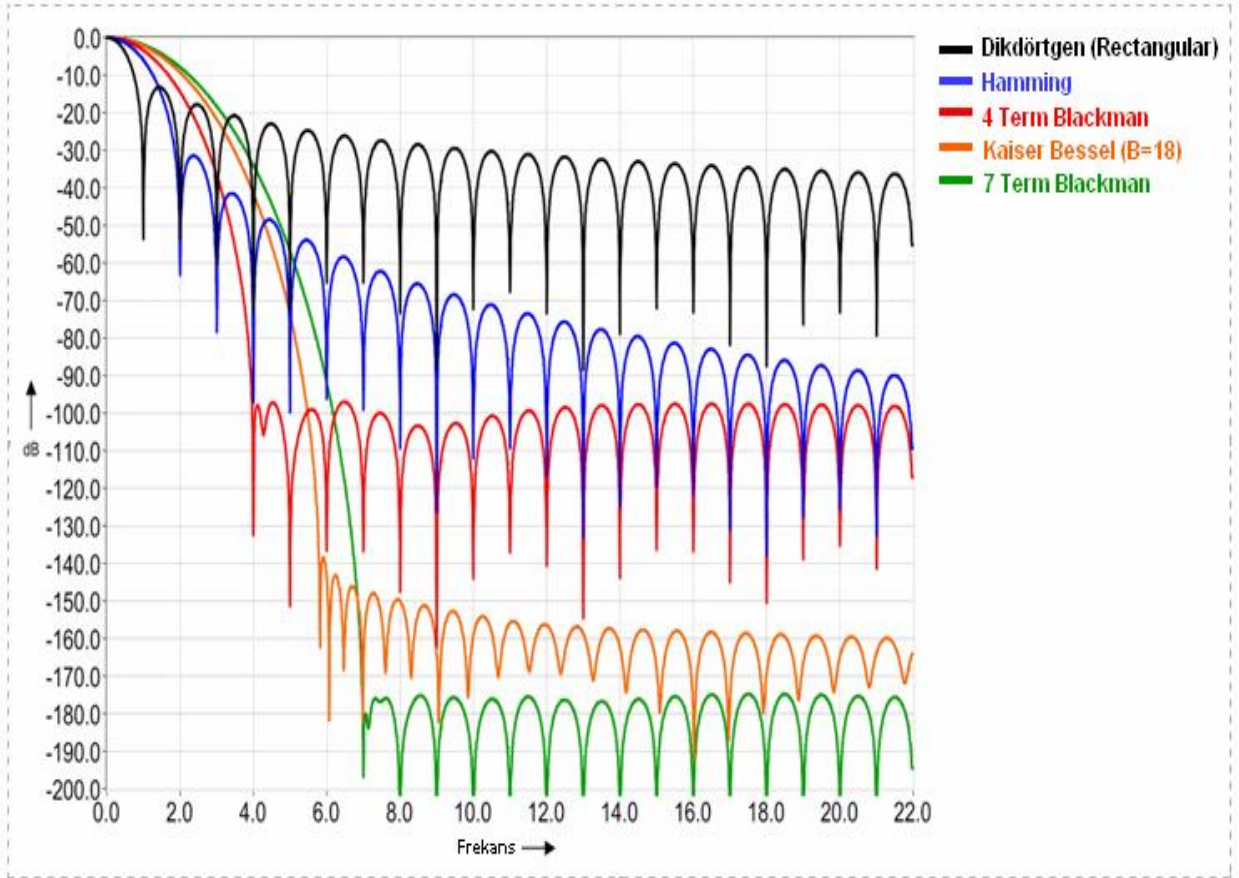
Şekil 3.4 Ses işaretinin çerçeveleme işlemi



### 3.2.2 Pencereleme

Pencereleme işlemi sürekli olan sesin hangi noktasından itibaren işlenmeye başlayacağını ve buna bağlı olarak ortaya çıkacak olan farklı domen dönüşümlerinin kötü etkisini ortadan kaldırmak amacıyla yapılır. Pencereleme işlemine başlanılmadan önce kullanılacak olan pencere fonksiyonu seçilmelidir. Pencere fonksiyonu, sınırlı süreli bir zaman aralığında tanımlanmış, sıfırdan farklı değer alan özel bir matematiksel fonksiyondur. Yaygın olarak kullanılan pencere fonksiyonları dikdörtgen (rectangular), Hamming, Blackman ve Kaiser – Bessel pencereleridir [National Instruments, 2006] (Şekil 3.5). Pencereleme işleminde ses işaretinin her bir elemanı  $x[n]$ , pencere fonksiyonunun her bir elemanı  $w[n]$  ile çarpılarak pencerelemiş  $y[n]$  sinyali elde edilir (3.2).

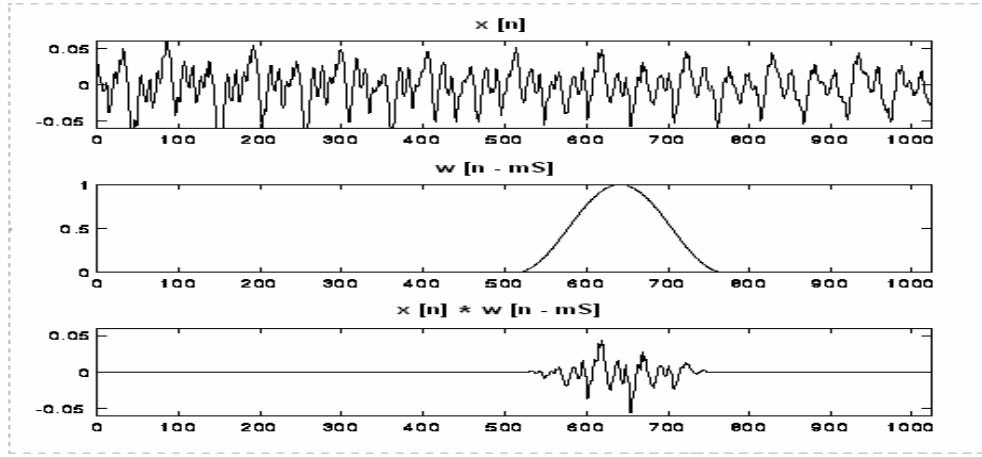
$$y[n] = x[n] * w[n] \quad (3.2)$$



Şekil 3.5 Yaygın olarak kullanılan pencere fonksiyonları [National Instruments, 2006]

Uygulamada daha önceki çalışmalarda başarılı sonuçlar elde edilmesi nedeniyle hamming penceresi (3.3) kullanılmıştır [Rabiner & Juang, 1993].

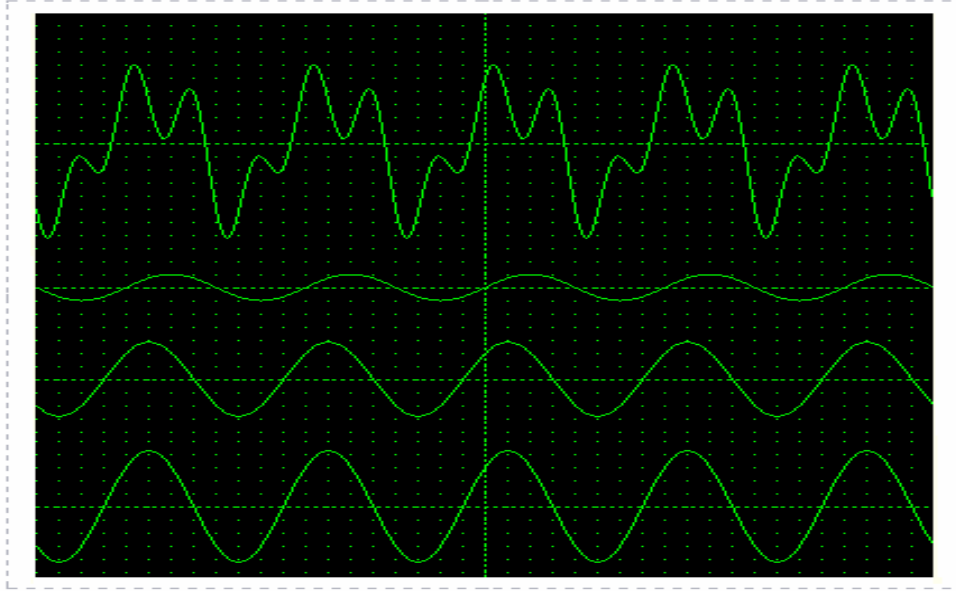
$$w(n) = 0.54 - 0.46 \cos((2\pi n) / (N - 1)) \quad N: \text{Pencere Uzunluğu} \quad (3.3)$$



Şekil 3.6 Pencereleme etkisi

### 3.2.3 Hızlı Fourier Dönüşümü (FFT)

Fourier dönüşümü (transformasyonu) 1822 yılında Fransız matematikçi Jean Baptiste Joseph Fourier tarafından *Théorie analytique de la chaleur* adlı kitabıyla bilim dünyasına takdim edilmiştir [Fourier, 1822]. Fourier'e göre periyodik olan tüm sinyaller, farklı genlik, frekans ve fazlardaki kosinüs ve sinüs dalgalarının temel bileşenlerinin toplamı olarak ifade edilebilir (Şekil 3.7).



Şekil 3.7 Periyodik bir sinyalin fourier analizi ile elde edilmiş sinüs ve kosinüs bileşenleri

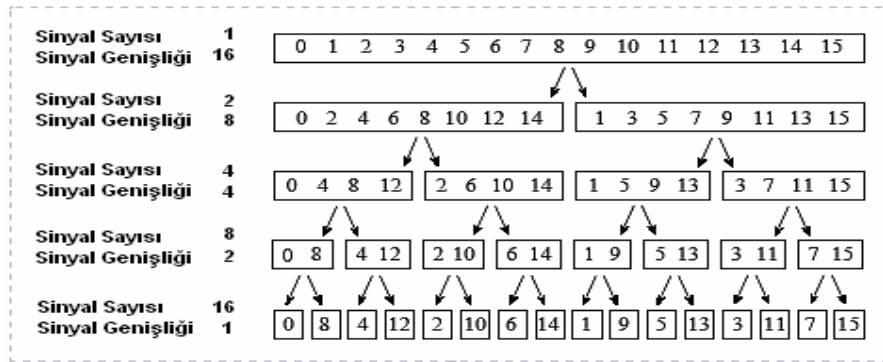
Fourier analizinin sürekli işaretler için genel formülü:

$$X(f) = \int_{-\infty}^{\infty} x(t) e^{-i2\pi ft} dt \quad (3.4)$$

Sinyal işlemede ayırık zamanda çalışabilmek için sürekli işaret üzerinden örnekleme yapılarak işaret sonlu hale dönüştürülür ve DFT (Discrete Fourier Transform – Ayırık Fourier Dönüşümü) uygulanır (3.5).

$$X_k = \sum_{n=0}^{N-1} x_n e^{-\frac{2\pi i}{N}kn} \quad k = 0, \dots, N-1 \quad (3.5)$$

DFT, zaman domenindeki N nokta sinyalini, her biri tek noktadan oluşan N zaman aralığı sinyaline ayrıştırarak çalışır. İkinci adım zaman domenindeki N sinyale karşılık gelen N frekans spektrasını hesap etmektir. Son olarak da elde edilen N spektra tek frekanslı spektrum olarak sentezlenilir [Cooley & Tukey, 1965].



Şekil 3.8 Örnek bir sinyalin DFT ile zaman domeninde ayrıştırılması

Şekil 3.8’de DFT’de kullanılan örnek bir sinyalin zaman domeninde ayrıştırılması gösterilmiştir. Bu örnekte 16 noktalı bir sinyal 4 farklı evrede ayrıştırılmıştır. İlk evrede 16 noktalı sinyal, 8 noktadan oluşan iki ayrı sinyale ayrıştırılmış, ikinci evrede ise 8 noktalı sinyallerin her biri 4 noktalı 4 sinyale ayrıştırılmıştır. Bu prosedür tek noktalı N tane sinyal elde edilinceye kadar devam eder. Sinyalin her ikiye bölünmesinde biniştirilmiş ayrıştırma (interlaced decomposition) kullanılır, yani sinyal tek ve çift numaralı örneklemlerine ayrıştırılır. DFT işleminde zaman domenindeki bu ayrıştırma işlemi, her bir noktanın örnek sıra numarasının ikili sistemdeki ifadesinin tersten okunması şeklinde, yani bit tersleme (bit reverse) algoritması kullanılarak yapılır. Ayrıştırma  $\log_2 N$  kadar evre gerektirmektedir (16 ( $2^4$ ) noktalı sinyal için 4 evre, 512 ( $2^7$ ) noktalı sinyal için 7 evre, 4096 ( $2^{12}$ ) noktalı sinyal için 12 evre gibi). Ayrıştırma esasen sinyaldeki örneklerin sıralarının yeniden düzenlenmesinden ibarettir.

Söz edilen yeniden düzenleme Şekil 3.9’da gösterilmiştir. Sol tarafta, orijinal sinyalin örnek numaraları, ikili sistemdeki karşılıkları ile birlikte listelenmiştir. Sağ tarafta ise yeniden düzenlenmiş örnekler yine ikili sistemdeki karşılıkları ile birlikte listelenmiştir.

Örnek Numaraları Normal Sırada		Örnek Numaraları Terslenmiş Bit Sırasında	
Onlu	İkili	Onlu	İkili
0	0000	0	0000
1	0001	8	1000
2	0010	4	0100
3	0011	12	1100
4	0100	2	0010
5	0101	10	1010
6	0110	6	0100
7	0111	14	1110
8	1000	1	0001
9	1001	9	1001
10	1010	5	0101
11	1011	13	1101
12	1100	3	0011
13	1101	11	1011
14	1110	7	0111
15	1111	15	1111

Şekil 3.9 On altı noktalı bir sinyal için bit tersleme işlemi

DFT algoritmasında bir sonraki adım zaman domenindeki 1 noktalı sinyalin frekans spektrasını bulmaktır. Bu işlem DFT'nin en kolay aşamasıdır, çünkü 1 noktalı sinyalin frekans spektrası kendisine eşdeğerdir.

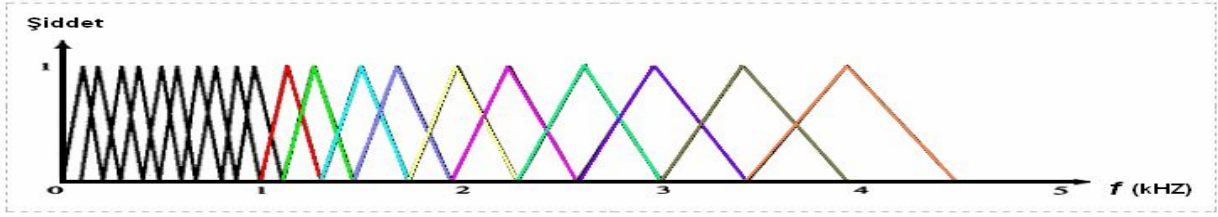
DFT de yapılacak son işlem N frekans spektrasını zaman domeninde ayrıştırmasının yapıldığı sıralamanın tam tersi sıralamada birleştirmektir. Bu aşamada ikili ters çevirme algoritması işe yaramaz ve evre evre geriye doğru ilerlemek durumunda kalırız. İlk evrede 16 frekans spektrası (her biri 1 nokta) 8 frekans spektrasına (her biri 2 nokta) sentezlenir. İkinci evrede, 8 frekans spektrası (her biri 2 nokta) 4 frekans spektrasına (her biri 4 nokta) sentezlenir ve böylece devam edilerek en son evrede 16 nokta frekans spektrumlu DFT çıktısı meydana getirilmiş olunur.

### 3.2.4 Mel ölçeğine aktarma

Yapılan çalışmalar, insan kulağının frekans içeriklerini algılamasının lineer bir ölçekte gerçekleşmediğini göstermektedir. Örneğin bir sesin frekansı iki katına çıkartıldığında ya da yarıya düşürüldüğünde bu değişim kolaylıkla fark edilemezken, aynı sesin frekansı on katına çıkartılır ya da onda birine düşürülürse insan kulağı bu değişimi daha kolay fark edebilir. Bu yüzden f frekanslı her ses için “mel” ölçeği denilen yeni bir ölçek hesaplanır. Mel-Frekansı 1000 Hz'in altında lineer, 1000 Hz'in üstünde logaritmik dağılıma sahip bir ölçektir [Rabiner & Juang, 1993][Davis & Mermelstein, 1980]. Bir referans noktası olarak alınan 1 kHz'lik ses, insan duyma eşiğinin 40 dB üstünde olmakla birlikte 1000 mel olarak tanımlanır. Hz türünden değeri bilinen bir frekansın mel eşdeğerini bulmak için özel bir formül kullanılır (3.6) [Deller & Hansen & Proakis, 2000].

$$\text{mel}(f) = 2595 \times \log_{10}(1 + f/700) \quad (3.6)$$

Bu dönüşüm formülünü gerçekleyebilmek için mel ölçeğinde düzenli olarak yerleştirilmiş üçgen bant geçiren, aralıkları ve bant genişliği sabit mel frekansı aralığına bağlı bir filtre bankası kullanılmalıdır (Şekil 3.10).



Şekil 3.10 Mel ölçekli filtre bankası

Bu filtre bankasına uygulanması gereken giriş DFT ile frekans domenine taşınmış olan sinyalin güç spektrumudur. DFT işlemi sonucunda frekans domenine taşınmış olan ve karmaşık sayılarla ifade edilen sinyalin güç spektrumu, sinyalin mutlak değerinin alınması yani sinyali ifade eden karmaşık sayıların modüllerinin bulunması yoluyla hesaplanılır (3.7).

$$z = a + bi ; |z| = \sqrt{a^2 + b^2} \quad (3.7)$$

Son aşamada, logaritması alınmış mel spektrumunun frekans domeninden tekrar zaman domenine çevrilmesiyle MFCC bulunmuş olur. Mel Keprstral Katsayıları, dolayısıyla onların logaritmaları alınmış halleri de gerçek sayılardan oluştuğu için işareti frekans domeninden zaman domenine geri döndürebilmek için IDFT (Inverse Discrete Fourier Transform – Ayrık Ters Fourier Dönüşümü) yerine DCT (Discrete Cosine Transform – Ayrık Kosinüs Dönüşümü) yeterli olur (3.8) [Koishida & Tokuday, 1997].

$$c[n] = \sum_{i=0}^N \text{Mel}[i] \cos(\pi n (i - 0.5) / N) ; 0 \leq n < N \quad (3.8)$$

### 3.2.5 $\Delta$ MFCC ve $\Delta^2$ MFCC ile ek özelliklerin çıkartılması

Özellik çıkartma amacıyla yapılan MFCC işlemi sonucunda her bir çerçeve için 13 karakteristik özelliğin yer aldığı bir özellik vektörü oluşturulur. Bu vektörün ilk elemanı ve aynı zamanda giriş sinyalinin ortalama değeri olan  $C[k]_0$  enerji olarak adlandırılır ve sinyal hakkında çok az bilgi içeren bu bileşen tanıma aşamasında çoğunlukla kullanılmaz.

MFCC sonucunda 12 karakteristik özelliği olan vektörler oluşturulur, tanımda yüksek başarılar elde etmek isteniliyorsa işaretin akışı boyunca bu özellik vektörlerinin zaman içerisindeki değişimleri de ayrı birer özellik gibi kullanılabilir. Bu işlem MFCC'nin birinci

türevini almaktır. Ortaya çıkan yeni vektöre  $\Delta$ MFCC (Delta MFCC) vektörü adı verilir (3.9). Bu vektör MFCC vektörüne ilave edilerek 16 karakteristik özellik taşıyan genişletilmiş bir özellik vektörü elde edilir.

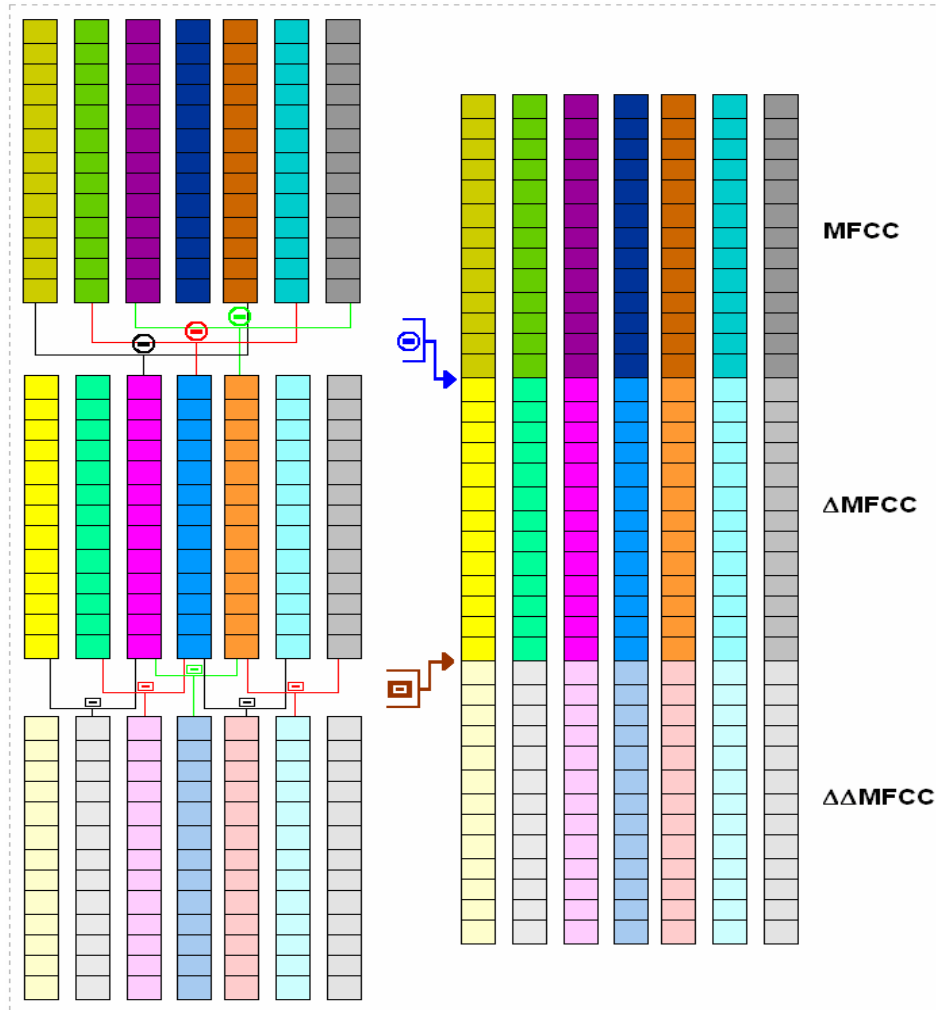
$$\Delta C [ i ] = C [ i ] \quad ; i = 0, 1, N - 1, N \quad (3.9)$$

$$\Delta C [ i ] = C [ i + 2 ] - C [ i - 2 ] \quad ; 1 < i < N - 1$$

Tanımda daha da yüksek bir başarı elde etmek istenilirse  $\Delta$ MFCC vektörü üzerindeki değişimler hesaplanılarak  $\Delta^2$ MFCC (Delta – Delta MFCC) vektörü elde edilebilir (3.10).  $\Delta$ MFCC ve  $\Delta^2$ MFCC işlemleri sonucunda başlangıçta 12 si kullanılabilir 13 karakteristik taşıyan özellik vektörü, 36 sı kullanılabilir 39 karakteristik taşıyan genişletilmiş bir vektöre dönüştürülmüş olur (Şekil 3.11).

$$\Delta^2 C [ i ] = \Delta C [ i ] \quad ; i = 0, N \quad (3.10)$$

$$\Delta^2 C [ i ] = \Delta C [ i + 1 ] - \Delta C [ i - 1 ] \quad ; 0 < i < N$$



Şekil 3.11  $\Delta$ MFCC ve  $\Delta^2$ MFCC hesaplanması

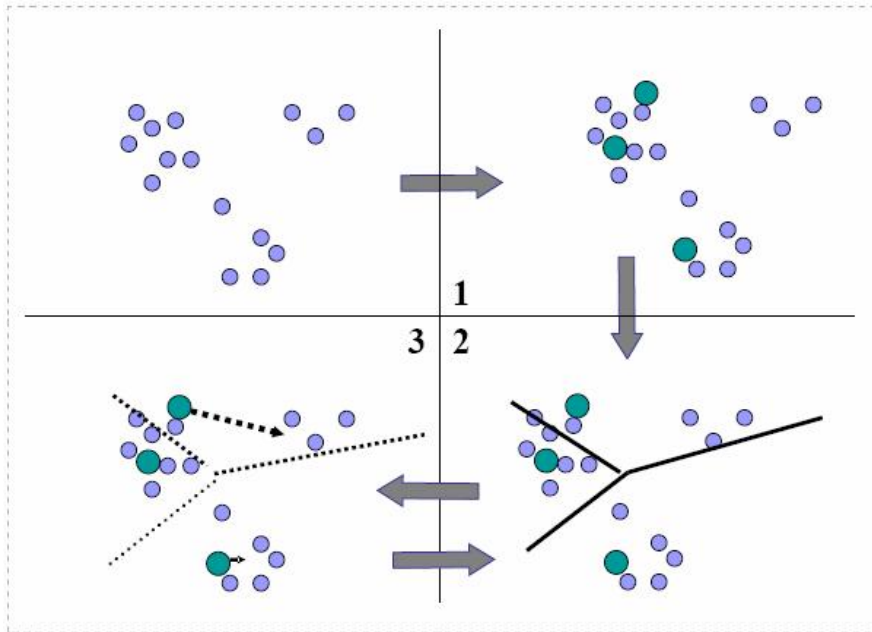
### 3.3 k-Ortalama Yöntemi ile Özelliklerin İndirgenmesi

Özellik çıkartma işlemi sonucunda sistemin girişine uygulanan giriş sinyalinin uzunluğuna bağlı olarak değişen sayıda ve her biri  $\Delta^2$ MFCC işleminin sonucunda elde edilmiş 39 karakteristik genişliğinde farklı özellik vektörleri oluşturulmuş olur. Hem tanıma işleminin hızlı olmasını sağlamak amacıyla hem de girişe farklı uzunlukta sinyallerin uygulanabileceği ihtimali göz önünde bulundurularak değişken olan bu özellik vektörü sayısı değişken olmayan belirli bir sayıda sabitlenmelidir. Bir başka deyişle elde edilmiş olan değişken sayıdaki  $x$  adet özellik vektörü, bu sinyalin tamamını temsil eden  $k$  adet özellik vektörüne indirgenmelidir.

Belirsiz sayıdaki özellik vektöründen  $k$  adet özellik vektörünün seçilmesi işlemi, vektör kuantalama (vector quantization) olarak adlandırılır [Quatieri, 2002]. Çalışmada bu amaçla yaygın olarak kullanılmakta olan  $k$ -ortalama ( $k$ -means) yöntemi uygulanmıştır [Jelinek, 1998].

$k$ -ortalama yöntemi ile mevcut veri  $k$  adet özellikle temsil edilmek isteniliyorsa:

1. Rasgele seçilen  $k$  adet özellik vektörü  $k$  sınıfın merkezi olarak belirlenir.
2. Mevcut özellik vektörleri bu  $k$  adet merkez özellik vektöründen hangisine en yakınsa o sınıfa dâhil edilir.
3. Oluşan sınıfların yeni merkezleri o sınıftaki özellik vektörlerinin ortalama değeri ile değiştirilir.
4. Merkez özellik vektörleri değişmeye kadar 2. ve 3. adımlar tekrarlanır.



Şekil 3.12  $k$ -ortalama yönteminin uygulanışı [Brazma]

Özellik indirgemede kullanılan k-ortalama yöntemine ait yarı kodlama aşağıda verilmiştir.

```

0:  $M_0(k) = k \text{AdetBaslangicSinifMerkezNoktasıBelirle()}$ 
1: repeat
2:   VerileriEnYakinMekezliSinifaAta()
3:    $M(k) = \text{SiniflarIcinMerkezNoktalariniHesapla()}$ 
4: until  $M_T(k) = M_{T-1}(k)$ 

```

k-ortalama yönteminin ikinci adımında bahsedilen vektörler arası uzaklığı hesaplamak için birden fazla yöntem kullanılabilir. Mesafe ölçümünde kullanılan metriklerin kullanışlı olan genel hali Minskowky mesafesidir (3.11) [Hervé, 2007].

$$d_{\lambda}(x,y) = \left\{ \sum (x_i - y_i)^{\lambda} \right\}^{1/\lambda} \quad (3.11)$$

Denklem 3.11 yardımıyla hesaplanan mesafe;

$\lambda = 2$  olduğunda Euclid mesafesine (3.12);  $\lambda = \infty$  olduğunda en büyük fark ( $x_i - y_i$ ) toplamda baskın olduğundan Chebyshev mesafesine (3.13) son olarak  $\lambda = 1$  olduğunda ise Manhattan (City Block) ya da diğer bir deyişle mutlak değer mesafesine eşit olur (3.14).

$$d_e(x,y) = \sqrt{\sum (x_i - y_i)^2} \quad (3.12)$$

$$d_c(x,y) = \max |x_i - y_i| \quad (3.13)$$

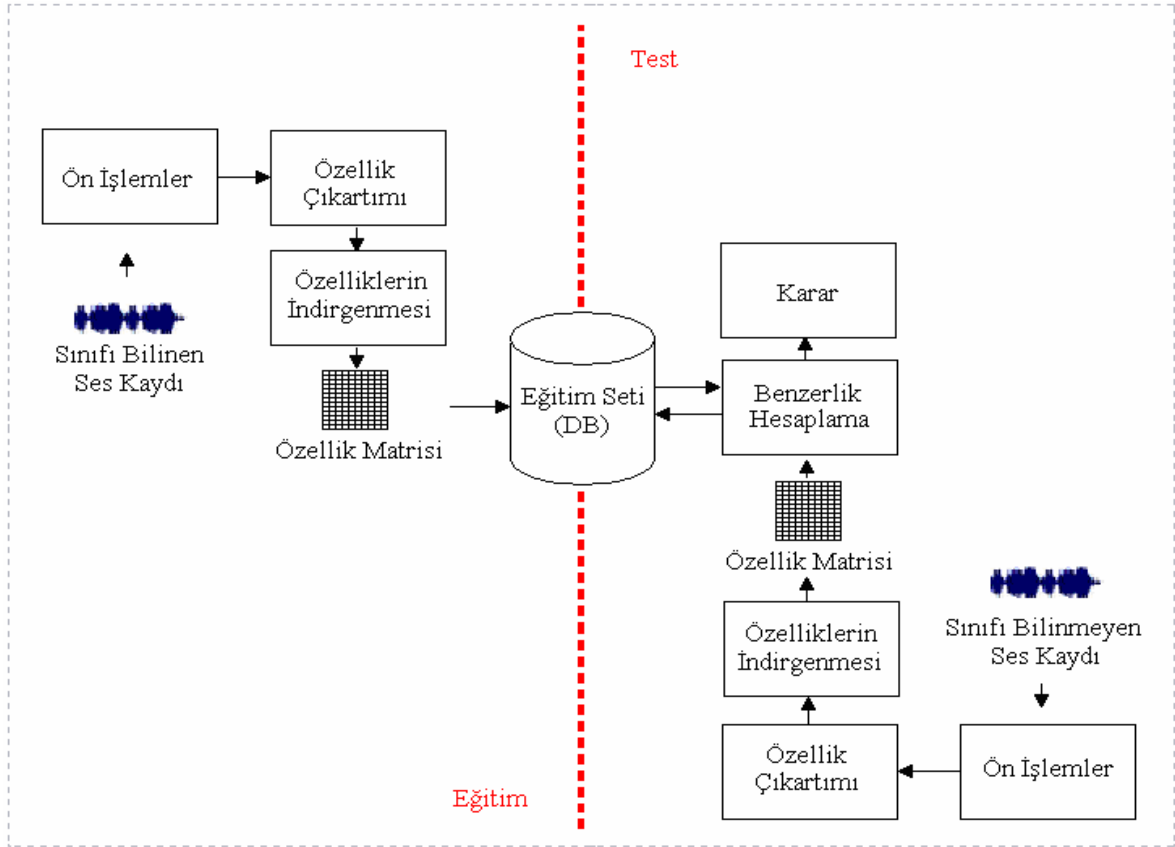
$$d_M(x,y) = \sum |x_i - y_i| \quad (3.14)$$

### 3.4 Tanıma Aşaması

$\Delta^2$ MFCC sonucunda elde edilen sinyalin uzunluğuna bağlı olarak farklı sayıda değişen özellik vektörlerinin vektör kuantalama yoluyla belirli sayıya indirgenmesiyle 39 karakteristik taşıyan n adet özellik vektörü elde edilmiştir. Bir başka deyişle  $[n \times 39]$  boyutunda özellik matrisleri oluşturulmuştur.

Eğitim işlemi sırasında oluşturulan özellik matrisleri bir veri tabanında depolanarak eğitim setini oluştururlar. Eğitim setinin oluşturulmasının ardından, hangi sınıfa dâhil olduğu öğrenilmek istenilen ses kaydı için de özellik matrisi oluşturulur. Eğitim seti özellik matrisleri ile test kaydına ait özellik matrisi arasında benzerlik hesabı yapılarak sınıfı belli olmayan ses kaydının hangi sınıfa daha çok benzediği belirlenir (Şekil 3.13).





Şekil 3.13 Sisteme ait genel çalışma yapısı

Test kaydının eğitim setindeki örneklerden hangisine daha çok benzediğine karar verilmesi için birden fazla yöntem kullanılabilir. Bunlardan en çok kullanılan iki tanesi SVM ve k-NN yöntemleridir.

### 3.4.1 Destek Vektör Makineleri (SVM)

Destek Vektör Makineleri yöntemi, sınıflandırma ve doğrusal olmayan fonksiyon yaklaşımı problemlerinin çözümü için 1992'de Vapnik tarafından önerilen eğitici bir öğrenme algoritmasıdır. Son yıllarda daha yaygın olarak kullanılmaya başlanan Destek Vektör Makineleri, yazı tanıma, nesne tanıma, ses tanıma, yüz tanıma gibi örüntü tanıma uygulamalarında kullanılmıştır [Burges, 1998].

Destek vektör makineleri temelde doğrusal olarak ayırt edilebilen iki sınıf problemlerin çözümünden yola çıkarak doğrusal olarak ayırt edilemeyen veya çoklu sınıf problemlerin çözümü için genelleştirilebilir.

Pozitif örnekleri negatif örneklerden ayıran bir hiper düzlemimiz olduğunu varsayarsak, iki sınıf probleminde doğrusal sınıflandırıcı hiper düzlemin normal vektörü  $w$  ve ofset değeri  $b$  ile tanımlanır. Karar sınırı  $w^T x + b = 0$  doğrusudur. Bu hiper düzlemin ayırdığı yarı

uzaylardan her biri bir sınıf belirtir. Bu durumda (3.15) ile verilen koşulların gerçekleşmesi gerekir:

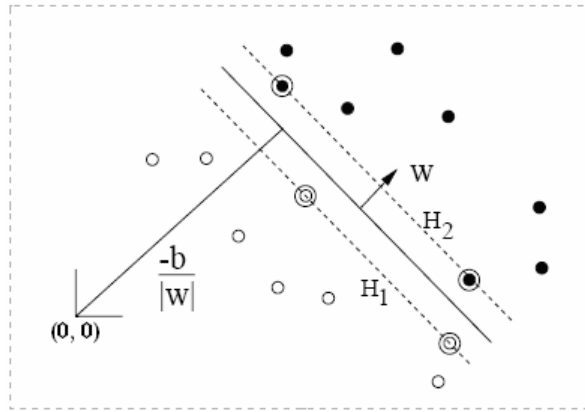
$$\begin{aligned} w^T x_i + b &\geq 1 & y_i &= 1 \\ w^T x_i + b &\leq -1 & y_i &= -1 \end{aligned} \quad (3.15)$$

Sınıf etiketi (3.15) ile verilen eşitsizlikler genelleştirilerek tek bir eşitsizlik olarak da ifade edilebilir (3.16) [Müler & Mika & Ratsch & Tsuda & Schölkopf, 2001].

$$y_i * (w^T x_i + b) \geq 1 \quad (3.16)$$

Doğrusal sınıflandırıcının sınırı, hiper düzlem ile öğrenme verileri arasındaki en kısa uzaklık olarak tanımlanır. Ayırt etme yüzeyine en yakın veriler destek vektörleri olarak adlandırılır ve karar sınırı sadece destek vektörleri ile belirlenir. Destek vektörlerine temel olan noktalar eğitim setindeki elemanların değişmesi durumunda yeniden hesaplanılır.

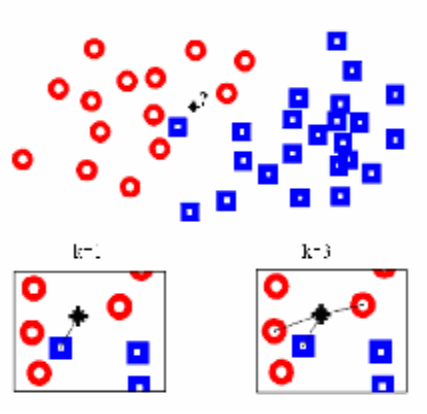
Örnek bir veri seti için oluşan destek vektörleri (destek vektörüne temel olan eğitim seti elemanları daire içerisinde gösterilmiştir), H1 ve H2 hiper düzlemleri ve iki hiper düzlemi ayıran normal vektörü Şekil 3.14'de gösterilmiştir [Burges, 1998]. Destek vektör makineleri algoritması en büyük toleransa sahip ayırıcı hiper düzlem ile sınıflandırma yaparak öğrenme hatasını en aza indirmeye çalışır [Özkaya & Kaya & Gürgen, 2005].



Şekil 3.14 Destek vektörlerinin oluşturulması [Burges, 1998]

### 3.4.2 k-En Yakın Komşuluk (k-NN)

k-NN yöntemi ile sınıflandırma yapılırken test seti elemanının özelliklerinin, tüm eğitim seti elemanlarının özelliklerine olan uzaklığı Bölüm 3.3'te bahsedilen mesafe hesaplarından herhangi birisi ile hesaplanır. Hesaplanan mesafeler küçükten büyüğe doğru sıralanılır ve test seti elemanına en çok benzeyen k eğitim seti elemanı içerisinde çoğunluğun hangi sınıfa dâhil olduğuna bakılarak test verisinin hangi sınıfa daha çok benzediği belirlenir.



Şekil 3.15 k-NN yöntemi ile sınıflandırmaya bir örnek [Claus, 2004]

Şekil 3.15’de örnek bir eğitim seti için dışarıdan gelen bir örneğin  $k=1$  ve  $k=3$  değerleri alınarak sınıflandırılması gösterilmiştir. k-NN yönteminde  $k$  değeri 1 seçildiğinde mavi gruba dahil edilen örnek,  $k$  değeri 3 olarak seçildiğinde kırmızı gruba dahil edilmiştir [Claus, 2004].

X sınıfı öğrenilmek istenilen örneği göstermek üzere; k-NN yöntemine ait yarı kodlama aşağıda verilmiştir.

```

0: eğitimSeti = {(x1, c1), ..., (xN, cN)}
1: FOR EACH trn IN eğitimSeti
2:   Mesafeler.ekle (Mesafe(trn.x, X))
3:   Sınıflar.ekle (trn.c)
4: NEXT
5: IndisSırala (Mesafeler, Sınıflar)
6: CogunlukSınıfıSec (Sınıflar, K)

```

### 3.4.3 Gemi veritabanında olmayan seslerin eşik değeri kullanılarak belirlenmesi

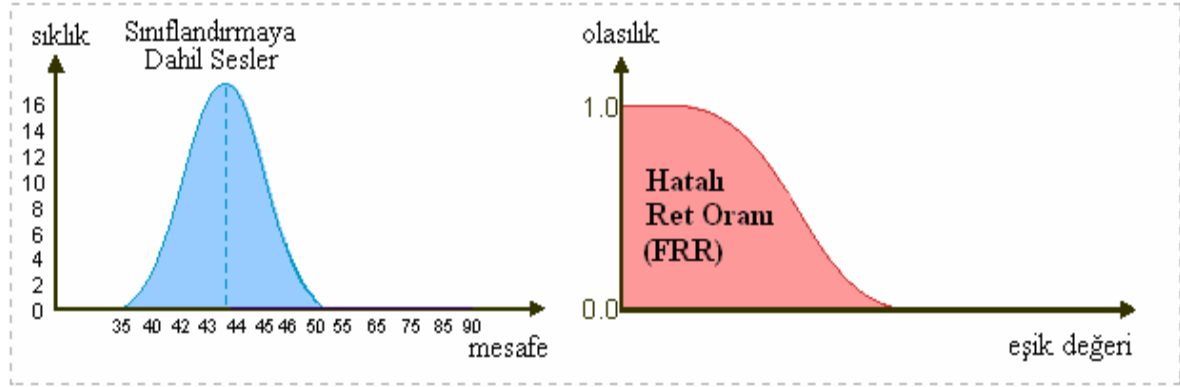
Teorik olarak, sistemin tanıdığı sesler test edilirken hesaplanılan mesafe, her zaman için eğitim setindeki sınıflardan birine dâhil olmayan sesler için hesaplanılan mesafeden daha düşük olmalıdır. Bu doğru olursa bir eşik değeri belirlenerek, bu eşik değerinin üstündeki değerler sistemde tanımlı olan sesler ve eşik değerini geçemeyen sesler de sistemde tanımlı olmayan sesler olarak tek bir eşik değeri ile sınıflandırılabilir.

Bu varsayım gerçek hayattaki tanıma sistemleri için her zaman geçerli değildir. Bazı durumlarda eğitim setindeki sınıflardan birine dâhil olmayan sesler için hesaplanılan mesafe, sistemin tanıdığı sesler test edilirken hesaplanılan mesafeden daha düşük olabilmektedir. Bu yüzden gerçek şudur ki, hangi sınıflandırma yöntemi seçilmiş olursa olsun istenilmeyen sınıflandırma hataları her zaman olacaktır.

Eşik değeri, eğitim seti özellik matrisi ile test seti özellik matrisi arasındaki mesafe farkının düşük olduğu bir seviye olarak belirlenirse, sistem tarafından tanınmayan sınıfların hatalı

olarak bir sınıfa dâhil edilmesiyle oluşan hatalar en aza indirgenmiş olacaktır. Buna karşılık gerçekte sistemde tanımlı sınıflardan birine dâhil olan seslerin sistem tarafından tanınmaması olarak ortaya çıkan hatalar ise artacaktır (Şekil 3.16).

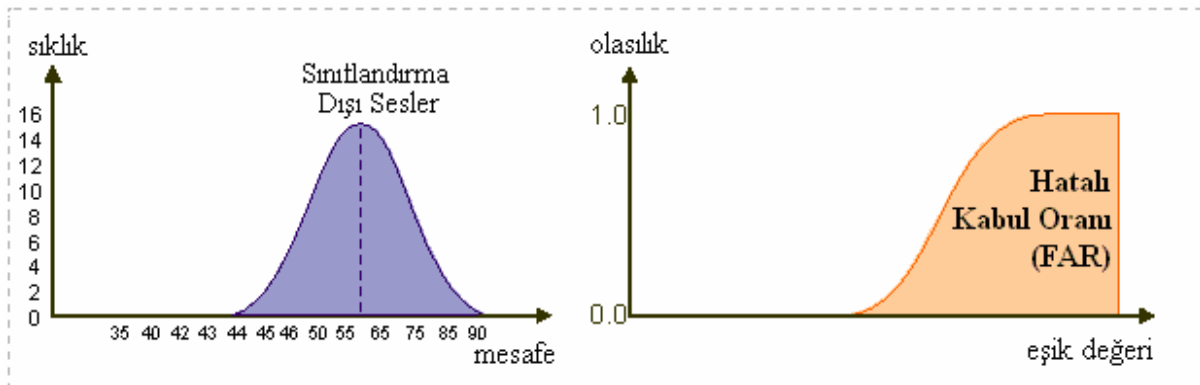
Bilinmeyen bir örneğin mevcut sınıflardan birine ait olmasına rağmen, eşik değerinin yüksekliği nedeniyle tanınmaması durumuna hatalı reddetme (FR – False Rejection), tanıma işlemi süresince hatalı ret ile sonuçlanan karşılaştırma sayısının, tüm karşılaştırma sayısına oranına hatalı ret oranı (FRR – False Rejection Rate) denilir [Bio ID].



Şekil 3.16 Eşik değerinin hatalı ret oranına etkisi

Eşik seviyesi değeri mesafe farkı maksimum olacak şekilde belirlenirse sistemde tanımlı sınıflardan birine dâhil olan seslerin sistem tarafından tanınmaması minimum düzeyde olurken, sistem tarafından tanınmayan sınıfların hatalı olarak bir sınıfa dâhil edilmesi maksimum düzeyde olacaktır (Şekil 3.17).

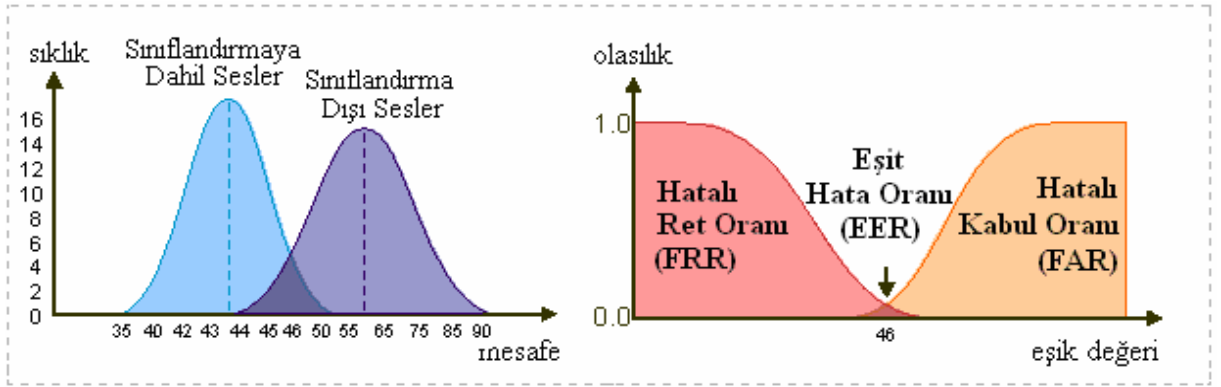
Sınıflandırmaya dâhil olmayan seslerin, eşik değerinin düşük tutulması nedeniyle hatalı olarak bir sınıfa dâhil edilmelerine hatalı kabul etme (FA – False Acceptance), tanıma işlemi süresince hatalı kabul ile sonuçlanan karşılaştırma sayısının, tüm karşılaştırma sayısına oranına hatalı kabul oranı (FAR – False Acceptance Rate) denilir [Bio ID].



Şekil 3.17 Eşik değerinin hatalı kabul oranına etkisi

Test seti ses kaydı ile eğitim seti ses kaydı arasında karşılaştırma yapılırken, test edilen ses kaydı eğitim setindeki tüm seslerle tek tek karşılaştırılır. Tanıma için yapılan karşılaştırmalarda sınıflandırmaya dâhil olan ve olmayan sesler için hesaplanan mesafelerinin bazen örtüşebildiği görülmektedir. Bu nedenle belirli bir bant aralığında, sınıflandırmaya dâhil olan bazı sesler reddedilirken, sınıflandırma dışı olan bazı sesler hatalı olarak bir sınıfa dâhil edilirler. Bu bant aralığında hatalı ret oranının, hatalı kabul oranına eşit olduğu, yani eşit hata oranının (EER – Equal Error Rate) oluştuğu nokta en uygun eşik değeri olarak belirlenir [Bio ID] (Şekil 3.18).

Mesafe hesabı yapılırken özellik matrisleri arasındaki Euclid mesafesi hesaplanır ve buna göre eğitim seti ve test setindeki iki örnek arasında hesaplanan toplam mesafe: eşik değerinin üstündeyse sınıflandırma işlemi yapılmaz, ses sistem dışı olarak nitelenilir; eşik değerinin altındaysa bir sınıflandırma yapılır. Ses kaydının sınıflandırmaya dâhil edilip edilmeyeceği için yapılan bu işleme geçерleme (validation) denilir.



Şekil 3.18 Eşit hata oranı ile optimum eşik değerinin belirlenmesi

## 4. UYGULAMA

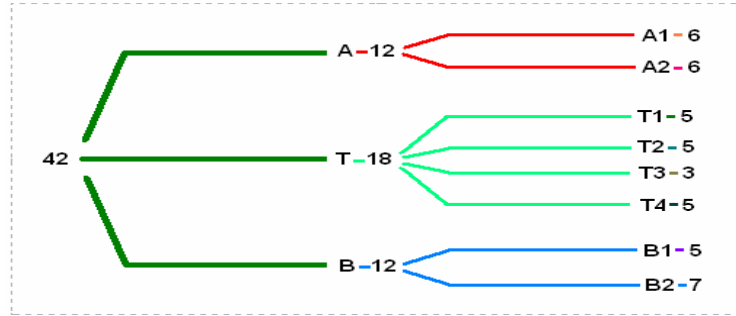
Bu çalışmada deniz taşıtlarının seyir halinde yaymış oldukları: pervane, jeneratör ve diğer ses çıkartan teçhizatlarının karakteristik özellik taşıyan toplam ses bilgisi kullanılarak deniz taşıtlarının sınıflandırılmaları yapılmıştır.

Sistemin çalışma performansını değerlendirmek için deniz taşıtlarına ait 3 ana sınıfa bağlı toplam 8 alt sınıfa ait her biri 30 sn uzunluğunda olan toplam 42 ses örneği kullanılmıştır. Eşik seviyesinin belirlenmesi ve sistemin hatalı kabul etme oranının (FAR) hesaplanması için de yine her biri 30 sn uzunluğunda olan 26 adet sınıflandırma dışı ses örneği kullanılmıştır.

### 4.1 Kullanılan Ses Örnekleri

Çalışmada gemiler seyir halindeyken denizden hidrafonlar aracılığıyla alınarak wave formatında kaydedilmiş ses örnekleri kullanılmıştır. Gemilerin sınıflandırılmasında iki aşamalı hiyerarşik bir yapı kullanılmıştır. Benzer özelliklerdeki gemiler bir ana sınıf altında değerlendirilmiş, daha sonra her bir sınıfta bulunan gemilerin modelini belirlemek amacıyla ikinci bir alt sınıflandırma yapılmıştır.

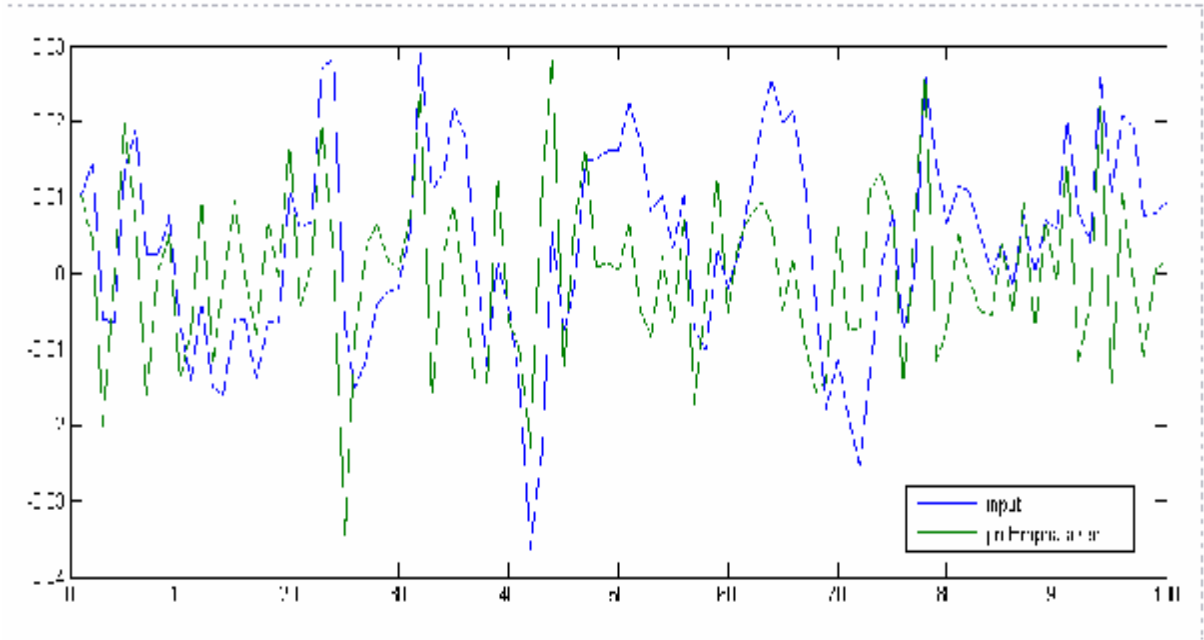
Uygulamada üç ana sınıf (A, B, T) ve bunlara bağlı sekiz alt sınıfa (A1, A2, B1, B2, T1, T2, T3, T4) ait 42 farklı örnek gemi sesi kullanılmıştır. Şekil 4.1’de kullanılan seslerin sınıflara dağılımını gösteren dendogram verilmiştir.



Şekil 4.1 Tanımda kullanılan sınıflar ve sınıf örnek sayılarını gösteren dendogram

### 4.2 Ön İşlemler, Özellik Çıkartımı ve Özelliklerin İndirgenmesi

Sesin kaydedilmesi sırasında kaydedilmek istenen ana sesin yanında parazit etkisi sonucunda oluşan istenmeyen gürültüler de kaydedilebildiğinden, kullanılan ön vurgulama (pre-emphasis) filtresi ile hem özellik çıkarmada kullanılacak olan yüksek frekanstaki sesler güçlendirilmiş hem de işlem sırasında (t) anındaki ses değerinin anlamlandırılmasında (t-1) anındaki ses verisi de kullanıldığından trende uygun olmayan, kısa süreli olarak var olan, parazit nedenli anlık yükselme ve düşmeler de engellenmiştir (Şekil 4.2).



Şekil 4.2 Giriş sinyali ve ön vurgulama filtresi uygulanmış hali (A\_1\_1.wav)

Ses üzerinde yapılan ön işlemlerin ardından, ses işareti özellik çıkartımında kullanılabilir bir hale gelmiş olur. Sonraki aşamada yapılan işlem kulağın sesleri algılaması ve ayırt etmesini taklit etmek amacıyla yapılan MFCC işlemidir.

Çalışmada yaklaşık 30sn uzunluğunda olan ses örnekleri kullanılmıştır. Analog ses sinyallerin örnekleme hızı 16000Hz ile yapıldığından her bir ses dosyası için yaklaşık 480000 (=30X16000) örnek elde edilmiş ve çerçeve adım değeri 125 (=16000/128) olarak hesaplanmıştır (4.1).

$$\text{çerçeve adımı} = \text{örnekleme değeri} / \text{çerçeve hızı} \quad (4.1)$$

Çerçeve genişliği 256 olarak seçilerek ardışık çerçeveler arası yaklaşık %49'luk bir örtüşme sağlanmış ve çerçeve sayısı yaklaşık 3838 olarak hesaplanmıştır (4.2).

$$\text{çerçeve sayısı} = (\text{örnek sayısı} - \text{çerçeve genişliği}) / \text{çerçeve adımı} \quad (4.2)$$

Çerçevelerin elde edilmesinin ardından mel ölçekli filtre bankasını elde edebilmek için 0-1 kHz aralığını kapsayan 13 lineer filtre ve 1-6.4 kHz aralığını kapsayan 27 logaritmik filtre olmak üzere 40 filtre elde edilmiştir. MFCC işlemi için kepsral katsayı adeti 13 olarak seçilmiş ve MFCC işlemi sonucunda giriş sinyalinin uzunluğuna bağlı olarak ~3800-3900 aralığında değişen her biri 13 karakteristik içeren özellik vektörleri elde edilmiştir.  $\Delta$ MFCC ve  $\Delta^2$ MFCC işlemleri sonucunda başlangıçta 12 si kullanılabilir 13 karakteristik taşıyan özellik vektörü, 36 sını kullanılabilir 39 karakteristik taşıyan genişletilmiş bir vektöre dönüştürülmüştür.

Özellik çıkartımının tamamlanmasının ardından elde edilen [ $\sim 3800 \times 39$ ] boyutundaki özellik matrisleri, k-ortalama yöntemi ile [ $128 \times 39$ ] boyutlu sınıflandırmaya hazır özellik matrisleri haline getirilmiştir.

### 4.3 Sınıflandırma

Bundan sonra yapılması gereken işlem gelen bir ses kaydının eğitim setindeki ses kayıtlarına olan mesafesine bakılarak bir sınıflandırma yapılmasıdır.

#### 4.3.1 Sınıflandırmada kullanılacak yöntemin belirlenmesi

Bölüm 3.4 de sınıflandırmada kullanılan iki popüler yöntem olan k-NN ve SVM den bahsedilmiştir. SVM yöntemi k-NN yöntemine göre daha iyi performans sağlarken; gerek eğitim seti için gerek sınıfı öğretilen ses kaydı için destek vektörlerinin oluşturulması işlem yükü nedeniyle uzun zaman almaktadır.

SVM yönteminin ana sınıflar ve alt sınıflar için başarısını hesaplamak için libsvm\* yazılımının 2,83 sürümü kullanılmış, hesaplanan başarı oranları Çizelge 4.1’de verilmiştir.

Özellik Sayısı	Ana Sınıf Tanıma Başarısı	Alt Sınıf Tanıma Başarısı
128	%90.47	%47.61

Çizelge 4.1 SVM yöntemine ait başarı oranları

Çizelge 4.1 incelendiğinde SVM yönteminin ana sınıfları tanımda iyi sayılabilecek bir başarı oranı yakaladığı görülmektedir. Destek vektörü oluşturmada kullanılacak eğitim setinin alt sınıflar düzeyinde verilerek SVM yöntemi ile başarı oranı hesaplandığında ise alt sınıf tanıma başarısı %47.61 gibi düşük bir oranda kalmaktadır.

k-NN yönteminde test edilen özellik matrisinin, eğitim setinde yer alan tüm özellik matrislerine (one against to rest) olan uzaklıkları Bölüm 3.3’te bahsedilen mesafe hesaplarından herhangi birisi ile hesaplanır ve bu uzaklıklar küçükten büyüğe doğru sıralanılır. En çok benzeyen k matrisin içerisinde çoğunluğun hangi sınıfa dâhil olduğuna bakılarak test verisinin hangi sınıfa daha çok benzediği belirlenir.

\* *libsvm* yazılımı <http://www.csie.ntu.edu.tw/~cjlin/libsvm> adresinden ücretsiz olarak temin edilebilir.



k-NN yönteminde ana sınıfların tanınmasında SVM yönteminde olduğu kadar yüksek bir oran elde edilemese de buna yakın bir başarı oranı elde edileceği açıktır. Ancak SVM yönteminde %47.61 olan alt sınıf tanıma başarı oranının k-NN yönteminde daha üst değerlerde çıkma ihtimali de oldukça yüksektir.

k-NN yöntemi ile yapılan benzerlik hesabının SVM yönteminde olduğu gibi karmaşık işlemler içermemesinden dolayı k-NN ile başarı oranlarının hesaplanılmasında özel bir yazılım kütüphanesi kullanılmamış ilgili akış doğrudan seçilen programlama dilinde kodlanmış, hesaplanan başarı oranları Çizelge 4.2 ve Çizelge 4.3’de verilmiştir.

<b>Özellik Sayısı</b>	<b>Ana Sınıf Tanıma Başarısı</b>	<b>Alt Sınıf Tanıma Başarısı</b>
128	%85.71	%69.04

Çizelge 4.2 k-NN yönteminde sınıflandırma başarı oranları

<b>Özellik Sayısı</b>	<b>Alt Sınıf</b>	<b>Alt Sınıf Örnek Sayısı</b>	<b>Alt Sınıf Tanıma Başarısı</b>
128	A	12	%66.66
128	B	12	%50.00
128	T	18	%83.33

Çizelge 4.3 k-NN yönteminde alt sınıf başarı oranlarının sınıflara göre dağılımı

Çizelge 4.2 incelendiğinde k-NN yönteminde alt sınıf tanıma başarı oranının SVM yöntemine göre yaklaşık %50 daha iyi sonuç verdiği görülmektedir.

k-NN yönteminde yapılan işlemlerin işlem yükü getirmeyen dolayısıyla çok zaman almayan basit yapıda işlemler olmaları ve k-NN yönteminde eğitim setine yeni bir kayıt eklenmesi durumunda bunun SVM yönteminde olduğu gibi diğer sınıfların durumunu etkilememesi gibi nedenlerden dolayı uygulamada k-NN yönteminin kullanılmasına ancak alt sınıf tanıma başarı oranını arttırmaya yönelik iyileştirme çalışmalarının yapılmasına karar verilmiştir.

#### **4.3.2 k-NN yönteminin hiyerarşik uygulanması ile başarı oranının arttırılması**

Uygulamada tanınmak istenilen ses kayıtları belirli bir hiyerarşiye göre ana sınıflar ve bu ana sınıflara bağlı alt sınıflardan birisine dâhil olan ses kayıtlarıdır. Sınıflandırmada kullanılacak olan k-NN yönteminde de bu hiyerarşik yapı uygulanırsa tanıma alanı daraltılmış olacağından sistemin sınıflandırma başarısının bu işlem sonucunda olumlu yönde etkileneceği kesindir. Buna göre öncelikle k-NN yöntemi ile ses kaydının hangi ana sınıfa dâhil olduğu

belirlenir, daha sonra da sesin hangi alt sınıfa dâhil olduğunu belirlemek amacıyla sadece birinci adımda belirlenmiş olan ana sınıfa bağlı alt sınıf örneklerini içeren dinamik bir eğitim seti üzerinde yeniden k-NN çalıştırılarak benzerlik hesabı yapılır. Sınıfı öğrenilmek istenilen ses kaydının bir sınıfa dâhil olup olmadığı Şekil 3.13’de gösterilen test işlemi akış diyagramında gösterilen adımlar izlenilerek belirlenir.

<b>Özellik Sayısı</b>	<b>Ana Sınıf Tanıma Başarısı</b>	<b>Alt Sınıf Tanıma Başarısı</b>
128	%85.71	%71.43

Çizelge 4.4 Hiyerarşik k-NN yönteminde sınıflandırma başarı oranları

<b>Özellik Sayısı</b>	<b>Alt Sınıf</b>	<b>Alt Sınıf Örnek Sayısı</b>	<b>Alt Sınıf Tanıma Başarısı</b>
128	A	12	%75.00
128	B	12	%58.33
128	T	18	%77.78

Çizelge 4.5 Hiyerarşik k-NN yönteminde alt sınıf başarı oranlarının sınıflara göre dağılımı

Çizelge 4.4 ve Çizelge 4.5 birlikte incelendiğinde k-NN yönteminin hiyerarşik uygulamasının başarıyı arttırdığı görülmektedir. Alt sınıflarda yer alan örnek sayısı arttıkça hiyerarşik uygulamanın başarı oranına olumlu etkisi daha da artacaktır.

### 4.3.3 Sınıflandırmayı kolaylaştırmak amacıyla Euclid uzaklığının değiştirilmesi

Yapılan bu değişiklikle eğitim seti ile test seti arasında birbirine benzer olan ses örnekleri etkilenmeden, birbirinden uzak olan ses örnekleri arasındaki mesafe daha da arttırılarak sınıflandırmanın daha kolay yapılması sağlanmıştır.

Yeni uzaklık hesabında öncelikle uygulamada elde edilmek istenilen hassasiyete göre değişebilen bir tolerans yüzdesi değişkeni (T) belirlenmiştir (Uygulama için %10). Daha sonra eğitim seti ile test seti ses örnekleri arası toplam mesafe hesabı yapılırken, eğitim seti özellik matrisindeki karşılaştırma noktası baz olarak alınmış ve test setinde karşılık gelen nokta, bu referans noktanın  $\pm \%T$ 'lik bir tolerans değeri uzaklığında ise Euclid mesafesi olduğu gibi ilave edilmiş;  $\pm \%T$ 'lik bir tolerans değerinden daha uzak olan noktalar için hesaplanan Euclid mesafesi daha önceden hesaplanmış olan bir kuvvetlendirme çarpanı ile çarpılmıştır (4.3). Böylelikle yakın olan noktalar etkilenmeden, uzak olan noktalar arasındaki mesafe daha da arttırılarak sınıflandırmanın kolaylaştırılması sağlanmıştır.

T: Tolerans Yüzdesi (%) ( $0 < T < 100$ ), KC: Kuvvetlendirme Çarpanı,

$T_{\text{Min}} = (100 - T) / 100$ ,  $T_{\text{Max}} = (100 + T) / 100$ ,  $KC = 1 + T_{\text{Min}}$  olmak üzere

$$|\text{mesafe}| = (X_i - Y_i)^2 \quad (X_i * T_{\text{Min}}) < Y_i < (X_i * T_{\text{Max}})$$

$$|\text{mesafe}| = ((X_i - Y_i) * KC)^2 \quad \left\{ \begin{array}{l} Y_i < (X_i * T_{\text{Min}}) \\ Y_i > (X_i * T_{\text{Max}}) \end{array} \right. \quad (4.3)$$

#### 4.3.4 Sınıflandırmaya dâhil olmayan seslerin tespiti için eşik değerinin belirlenmesi

Ses kaydının eğitim setine olan uzaklığı (4.3) ile belirlenmiş olan kurala göre hesaplanmış ve bu mesafe sonuçları üzerinde sınıflandırma işlemi yapılmıştır. Bu işlem, test edilen her ses kaydının eğitim sesindeki uzaklığın minimum olduğu sınıfa benzetilmesine neden olur.

Hiç bir sınıfa dâhil olmayan seslerin ayırt edilebilmesi için bir eşik değeri belirlenmelidir. Bir ses kaydı test edilirken hesaplanan mesafe bu eşik değerinin üzerinde ise ses kaydı sınıf dışı olarak nitelendirilmeli, mesafenin eşik değerinin altında olması durumunda sınıflandırma işlemi yapılarak test edilen kaydın eğitim setindeki sınıflardan hangisine en çok benzediği belirlenmelidir. Uygulamada, sınıflandırmaya dâhil olan sesler kullanıldığında farklı eşik seviyesi değerleri için ortaya çıkan hatalı reddetme sayıları (FR) ve buna bağlı olarak hesaplanan hatalı reddetme oranları (FRR) Çizelge 4.6 ve Şekil 3.16'de; sınıflandırmaya dâhil olmayan sesler kullanıldığında farklı eşik değerleri için ortaya çıkan hatalı kabul etme sayıları (FA) ve hesaplanan hatalı kabul etme oranları (FAR) Çizelge 4.7 ve Şekil 3.17'da verilmiştir.



		TST	TRN	MIN MSF	
		X01	A_2_1	51,16435	40
		X02	T_2_5	44,38451	41
		X03	T_2_5	57,90647	42
		X04	B_2_6	54,15482	43
		X05	B_1_2	52,06133	44
		X06	T_2_2	53,26621	45
		X07	T_2_1	53,46310	46
		X08	T_2_1	49,61331	47
		X09	T_2_1	53,05933	48
		X10	B_2_2	42,58184	49
		X11	T_2_1	71,87355	50
		X12	T_2_1	61,42778	51
		X13	T_2_1	52,22209	52
		X14	A_2_1	53,01792	53
		Y01	T_4_5	47,28734	54
		Y02	B_1_5	60,76048	55
		Y03	B_1_5	56,66561	56
		Y04	T_2_1	60,91515	57
		Y05	T_2_1	78,24757	58
		Y06	T_2_1	92,34562	59
		Y07	T_2_1	56,54871	60
		Y08	T_2_1	86,01927	61
		Y09	B_1_5	68,21929	62
		Y10	B_2_3	52,05630	63
		Y11	T_2_2	44,54668	64
		Y12	T_4_2	49,24635	65
					66
					67
					68
					69
					70
					71
					72
					73
					74
					75
					76
					77
					78
					79
					80
					81
					82
					83
					84
					85
					86
					87
					88
					89
					90
					91
					92
					93

EŞİK SEVİYESİ DEĞERLERİ

Çizelge 4.7 Farklı eşik seviyesi değerleri için hesaplanan hatalı kabul etme oranları (FAR)

Çizelge 4.6 ve Çizelge 4.7 birlikte incelendiğinde tanıma için yapılan karşılaştırmalarda sınıflandırmaya dâhil olan ve olmayan sesler için hesaplanan mesafelerinin bazen örtüşebildiği görülmektedir. Bu nedenle belirli bir bant aralığında, sınıflandırmaya dâhil olan bazı sesler reddedilirken, sınıflandırma dışı olan bazı sesler hatalı olarak bir sınıfa dâhil edilirler. Bu bant aralığında hatalı ret oranının, hatalı kabul oranına eşit olduğu, eşit hata oranının (EER – Equal Error Rate) oluştuğu nokta yani en uygun eşik değeri olarak belirlenir. Uygulamada optimum eşik değeri 46, FAR=FRR=0,12 olarak hesaplanmıştır (Şekil 3.18).

#### 4.3.5 k-NN yöntemi ile sınıflandırma

k-NN yöntemi ile sınıflandırma yapılırken test seti elemanının, eğitim seti elemanlarına olan uzaklıkları (4.3) ile belirlenmiş olan kurala göre hesaplanır. Hesaplanan mesafeler küçükten büyüğe doğru sıralanılır ve test seti elemanına en çok benzeyen k eğitim seti elemanı içerisinde çoğunluğun hangi sınıfa dâhil olduğuna bakılarak test verisinin hangi sınıfa daha çok benzediği belirlenir. k-NN hesabında alt sınıflardaki ortalama örnek sayısının 5 olması nedeniyle k = 3 olarak alınmıştır. Eğitim seti sınıflarındaki ortalama örnek sayısının daha fazla olduğu durumlarda k sayısı artırılabilir.

##### 4.3.5.1 Sınıflandırmada eş benzerlik oranları oluşması halinde alternatif bir çözüm

k-NN hesabı yapılırken farklı sınıflara eşit sayıda yakınlık oluşması ihtimalini engellemek için k sayısı tek olarak seçilir fakat buna rağmen yine de farklı sınıflara eş oranda benzerlik söz konusu olabilir. Konuyu uygulama sırasında karşılaşılan bir senaryo üzerinden açıklamak amacıyla sınıfı bilinmeyen bir örneğin eğitim seti örneklerine uzaklığı Çizelge 4.8’de görüldüğü gibi elde edilmiştir.

Uzaklık	Eğitim Seti Dosyası	Eğitim Seti Alt Sınıfı
21.564	T_2_3.wav	T2
22.001	T_2_2.wav	T2
22.806	T_1_5.wav	T1
23.113	T_1_2.wav	T1
23.202	T_4_3.wav	T4
23.495	T_1_3.wav	T1
...	...	...

Çizelge 4.8 k-NN yöntemi ile tanıma bir örnek (Test dosyası: T\_2\_4.wav)

Test edilen T\_2\_4.wav dosyasının özellik matrisinin, eğitim setinde yer alan özellik matrislerine olan uzaklıkları sıralanarak k=5 alındığında sesin 2 kez T2, 2 kez T1 ve 1 kez de T4 alt sınıfına benzediği görülmektedir. Bu durumda T2 ile T1 alt sınıflarına eş oranda

benzerlik söz konusu olmaktadır. Bu eşitliği bozmak amacıyla k değeri bir azaltılır; yeni durumda k=4 olur, yeniden k en yakın komşuluğa bakıldığında test sesinin 2 kez T2 ve 2 kez de T1 alt sınıfına benzediği görülür. Eşitlik halen bozulmadığından k değeri yeniden bir azaltılır. Son durumda k=3 olur; bu durumda test sesinin 2 kez T2 ve 1 kez T1 alt sınıfına benzediği görülür, eşitlik bozulmuş olduğundan tanıma kararı bu son oranlara göre verilerek işlem sonlandırılır.

Yani k komşuluğa göre tanıma işleminde en çok benzeyen sınıflar arasında eş oranlar söz konusu olduğunda eşitlik bozulana kadar k değeri bir azaltılır. En kötü ihtimalle k=1 olana kadar k değerinin azaltılmasına devam edilecektir ve k=1 durumunda test sesine en yakın mesafede olan eğitim seti elemanının sınıfı, test sesine en çok benzeyen sınıf olarak seçilecektir. Bu işleme ait yarı kodlama aşağıda verilmiştir.

```

0: k = başlangıç değeri (k-NN)
1: hata = 1
2: while hata = 1
3:   KNNHesapla()
4:   if ((k > 1) & (benzerlikOrani (Sinif1) = benzerlikOrani (Sinif2)))
5:     k = k - 1
6:   else
7:     hata = 0
8:   end if
9: wend

```

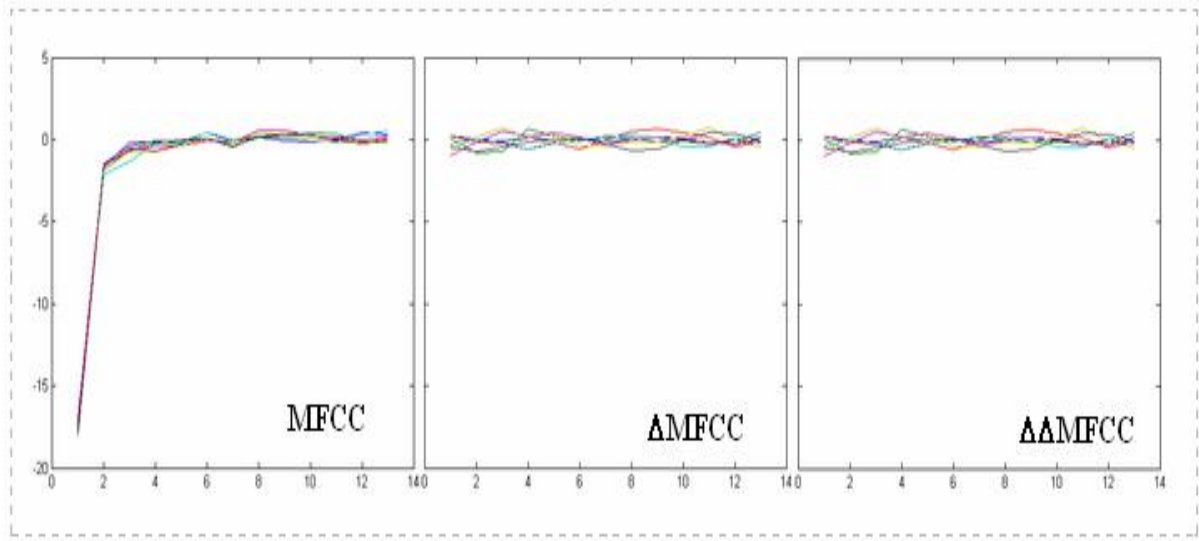
#### 4.4 Tanıma Parametrelerinin Seçimi

Uygulamada özelliklerin çıkartılması aşamasında kullanılacak olan karakteristik sayısı, FFT boyutu, özelliklerin indirgenmesi sırasında belirlenecek olan özellik vektörü sayısı ve tanıma aşamasında test edilen örneğin hangi sınıfa dâhil olduğunu belirlemede en yakın kaç komşuya bakılacağı gibi parametreler uygulama içerisinde parametreler ekranından değiştirilebilir.

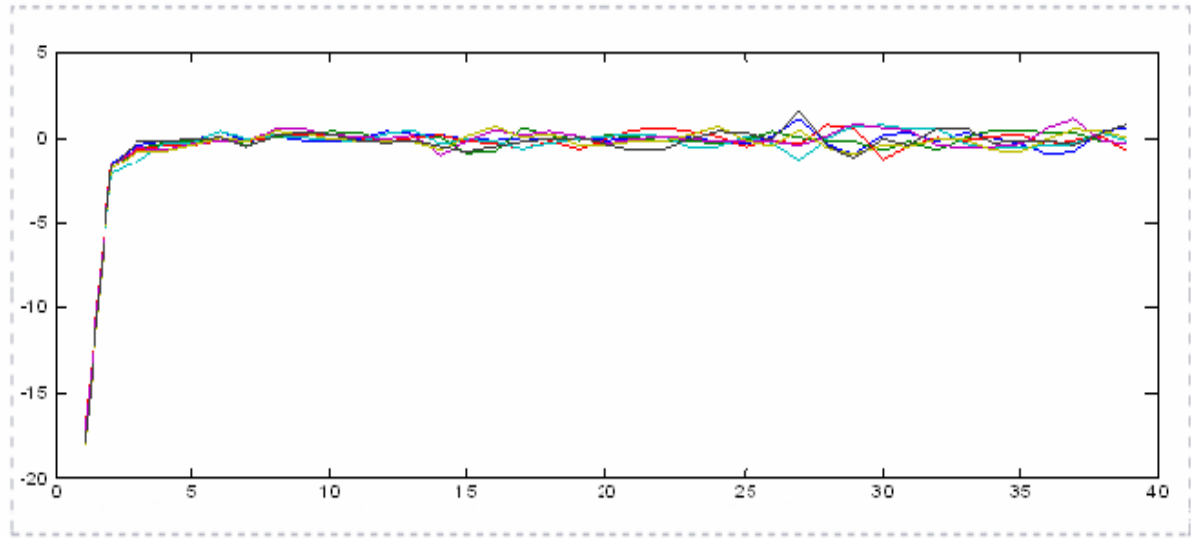
Özellik çıkartmada seçenekler: MFCC, DMFCC ve DDMFCC'dir. Her üç yöntemde de önce MFCC kullanılır. MFCC sonucunda 13 karakteristik özelliği olan vektörler oluşturulur.

Tanımda yüksek başarılar elde etmek isteniliyorsa işaretin akışı boyunca özellik vektörlerinin zaman içerisindeki değişimleri de ayrı birer özellik gibi kullanılabilir. Bu işlem MFCC'nin birinci türevini almaktır. Ortaya çıkan yeni vektöre  $\Delta$ MFCC – DMFCC (Delta MFCC) vektörü adı verilir. Bu vektör MFCC vektörüne ilave edilerek 16 karakteristik özellik taşıyan genişletilmiş bir özellik vektörü elde edilir. Tanımda daha da yüksek bir başarı elde etmek istenilirse  $\Delta$ MFCC vektörü üzerindeki değişimler hesaplanılarak  $\Delta^2$ MFCC (Delta – Delta MFCC) vektörü elde edilebilir.  $\Delta$ MFCC ve  $\Delta^2$ MFCC işlemleri sonucunda başlangıçta

12 si kullanılabilir 13 karakteristik taşıyan özellik vektörü, 36 sı kullanılabilir 39 karakteristik taşıyan genişletilmiş bir vektöre dönüştürülmüş olur.



Şekil 4.3 MFCC,  $\Delta$ MFCC ve  $\Delta^2$ MFCC vektörleri



Şekil 4.4  $\Delta$ MFCC ve  $\Delta^2$ MFCC vektörleri ile genişletilmiş özellik vektörü

MFCC işlemi içerisinde çerçeveleme yapılırken çerçeve boyutu 512, her bir çerçevelemede çevreye dâhil edilen yeni örnek sayısı 256 seçilerek %50 oranında bir çerçeve örtüşmesi sağlanmıştır. Bu oran ne kadar yükseltirse MFCC ile belirlenen karakteristik özellikler o kadar iyi olur; buna karşılık işlem yükü artacağından çalışma hızında düşüş gözlenilir.

MFCC,  $\Delta$ MFCC ve  $\Delta^2$ MFCC işlemlerinin ardından vektör kuantalama yapılarak eğitim setini oluşturacak ve test setinde kullanılacak olan özellik matrisleri oluşturulur. Vektör kuantalamada seçeceğimiz özellik sayısı ideal demet sayısına ulaşana kadar artırıldıkça



başarı oranı da artacak, aynı zamanda çalışma süresi de uzayacaktır. Fakat seçilen özellik sayısı ideal demet sayısından daha fazla olursa yapılan gereksiz demetlemelerden dolayı başarı oranı düşecek, özelliklerin saklanacağı eğitim setinin depolama maliyeti artacak ve ortaya çıkacak olan işlem yükünden dolayı sistemin cevap verme zamanı (response time) da düşmüş olacaktır.

Vektör kuantalamada k-ortalama işlemi sırasında özellik vektörlerinin merkez özellik vektörlerine olan uzaklığı Euclid mesafe denklemine (3.12) göre hesaplanmıştır. Vektör kuantalamada kullanılacak özellik sayısı için alınan farklı değerlere bağlı olarak değişen başarı oranları ve çalışma zamanları Çizelge 4.9'da gösterilmiştir.

<b>Özellik Sayısı</b>	<b>Ana Sınıf Tanıma Başarısı (hk-NN, k =3)</b>	<b>Alt Sınıf Tanıma Başarısı (hk-NN, k =3)</b>	<b>Çalışma Süresi (sn)</b>
32	% 88.10	% 61.90	0.50
64	% 90.48	% 69.04	0.75
128	% 85.74	% 71.43	1.25
256	% 80.95	% 57.14	1.63
512	% 88.10	% 59.52	4.63

Çizelge 4.9 Tanımda kullanılan özellik sayısının tanıma başarısı ve çalışma süresine etkisi

## 5. SONUÇLAR

Bu çalışmada deniz taşıtlarının seyir halinde yaymış oldukları: pervane, jeneratör ve diğer ses çıkartan teçhizatlarının karakteristik özellik taşıyan toplam ses bilgisi kullanılarak deniz taşıtlarının sınıflandırılmaları yapılmıştır.

Sınıflandırma işlemi için ses örneklerine parazitten arındırma gibi iyileştirme ön işlemleri yapılmıştır. Ön işlemlerin ardından MFCC ile özellik çıkartımı yapılmış, MFCC ile bulunan karakteristik özellikler, k-ortalama yöntemi ile sabit bir sayıya indirgenmiştir. İndirgenen özellikler üzerinde k-en yakın komşuluk yöntemi kullanılarak ses örneklerinin sınıflandırılması sağlanmıştır.

Yapılan çalışma daha önceden bu alanda ses tanıma, ses ile sınıflandırma gibi çalışmaların yapılmadığı deniz taşıtlarının sınıflandırılması gibi farklı bir alanda kullanılmış ve en uygun parametreler (özellik sayısı: 128, k-NN k değeri:3) seçildiğinde ana sınıf tanımada %85.71, alt sınıf tanımada ise %71.43 lük başarı oranları elde edilmiştir.

Başarı oranları mükemmel sayılabilecek derecede yüksek olmasa da sadece insan kulağıyla sınıflandırmanın yapılabildiği, otomasyona yönelik hiçbir çalışmanın olmadığı bir uygulama alanı için sınıflandırmada görevli kişilerin kararında destek olabilen iyi sayılabilecek başarı oranlarıdır.

Sistemde sınıflarda yer alan örnek sayıları artırıldığında, k-NN hesabında, k değerleri için daha büyük değerler kullanmak mümkün olacağından buna bağlı olarak başarı oranı da yükselecektir. Ses örneklerinin alınması sırasında ön işlemlerde yapılan parazit arındırma işlemine gerek bırakmayacak şekilde ses kaydı mümkün olduğunda parazit temizleme işlemi ile ses özelliklerinin bir kısmının yumuşatılmasının da önüne geçileceğinden başarı oranı daha da artabilir.

Uygulamanın gerçek zamanlı bir sistem için hızı ve başarısı yeterlidir. Ancak denizin aşırı dalgalı veya fırtınalı olması, ses kaydı esnasında bir başka deniz taşıtına ait ses bilgisinin de kaydedilmesi gibi farklı durum senaryolarına yönelik olarak; birden fazla noktadan ses kaydı alınarak ICA (Independent Component Analysis – Bağımsız Bileşen Analizi) gibi yöntemler ile sadece tanımda kullanılmak istenilen salt ses elde edildikten sonra bu sesin giriş verisi olarak kullanılmasıyla farklı ortam koşullarına karşı sağlam bir sistem elde edilmiş olur.

**KAYNAKLAR**

Alkan, Mahmut (2005), "WSSRS Using Mel Frequency Cepstral Coefficients", Deniz Harp Okulu, Yüksek Lisans Tezi

Alonso-Martinez, C. & Faundez-Zanuy, M. (2000), "Speaker Identification in Mismatch Training and Test Conditions", In Proc. of the IEEE Int. Conf. on Acoustics, Speech and Signal Processing

Başaran, İ. Eren (1981), "Ses Frekans Tekniği", M.E.B. Yayınları, ISBN: 975-11-1140-4

Bio ID (www), "About FAR, FRR and EER", [www.bioid.com/sdk/docs/About\\_EER.htm](http://www.bioid.com/sdk/docs/About_EER.htm)

Brazma, Alvis (Lecture), "Introduction to Microarray Data Analysis and Gene Networks", European Bioinformatics Institute

Burges, J.C. (1998), "A Tutorial on Support Vector Machines for Pattern Recognition", Data Mining and Knowledge Discovery 2, S.121-167

Claus, David (2004), "Nearest Neighbour: Condensing and Editing"

Cooley, James W. & Tukey, John W. (1965), "Mathematics of Computation", Vol. 19, No. 90, S.297-301

Davis, S.B. & Mermelstein, P. (1980), "Comparison of Parametric Representations for Monosyllabic Word Recognition in Continuously Spoken Sentences", IEEE Trans. on ASSP 28, S.357-366.

Deller, Jr. John R. & Hansen, John H.L. & Proakis, John G. (2000), "Discrete-Time Processing of Speech Signals". IEEE Press, ISBN 978-0780353862

EBU-UER (2001), "BWF a Format for Audio Files in Broadcasting", Ver. 1, European Broadcasting Union, Appendix A

Eronen, Antti (2001), "Automatic Musical Instrument Recognition", S.10

Fourier, Joseph (1822), "Théorie analytique de la chaleur", ISBN: 2-87647-046-2

Gaunard, Case (1998), "Automatic Noise Recognition"

Hassall, J.R. & Zaveri, K. (1979), "Acoustic Noise Measurements", Bruel and Kjaer

Henderson, Tom (1998), "The Physics Classroom - Lesson 2: Sound Properties and Their Perception, Intensity and the Decibel Scale"

Hervé, Abdi (2007), "Distance, Encyclopedia of Measurement and Statistics"

Jarnicki, J. & Mazurkiewicz, J. & Maciejewski, H. (1998), "Mobile Object Recognition Based on Acoustic Information". In Proceedings of the 24th Ann. Conf. of the IEEE Industrial Electronics Society, IECON'98, Vol. 3, S.1564-1569

Jelinek, Frederick (1998), "Statistical Methods for Speech Recognition", MIT Press, ISBN 978-0-262-10066-3, S.11

Kashino, K. & Murase, H. (1998), "Music Recognition Using Note Transition Context"

Knight, David C. (1960), "The First Book of Sound: A Basic Guide to the Science of Acoustics", Franklin Watts, Inc. New York, S.80

Koishida, Kazuhito & Tokuday, Keiichi (1997), "Efficient Encoding of Mel-generalized Cepstrum for CELP Coders", Tokyo Institute of Technology

Murthy, H.A. & Beaufays, F. & Heck, L.P. & Weintraub, M. (1999), "Robust Text-Independent Speaker Recognition over Telephone Channels". IEEE Trans. on Acoustics, Speech and Signal Proc., Vol. 7, No. 5, S.554-568.

Müller, K.R. & Mika, S. & Ratsch, G. & Tsuda, K. & Schölkopf, B. (2001), "An Introduction to Kernel Based Learning Algorithms", IEEE Transaction on Neural Networks, Vol. 12, No. 2

National Instruments (2006), "Software Filtering: Windowing - General Analog Concepts", NI Analog Resource Center, Tutorial 5357

Nyquist, H. (1928), "Certain Topics in Telegraph Transmission Theory", Trans. AIEE, Vol. 47, S.617-644

Özkaya, A. Uyar & Kaya, M. Emre & Gürgen, Fikret (2005), "Destek Vektör Makineleri Kullanılarak Aritmi Sınıflandırması", S.13

Quatieri, Thomas F. (2002), "Discrete Time Speech Signal Processing Principles and Practise", Prentice Hall, ISBN 013242942X , S.718-719

Rabiner, L.R. & Juang, B.H. (1993), "Fundamental of Speech Recognition", Prentice Hall, ISBN: 0-13-015157-2

Reynolds, Douglas A. (1995), "Automatic Speaker Recognition Using Gaussian Mixture Speaker Models", The Lincoln Laboratory Journal, Vol. 8, No. 2

Robinson, Tony (1995), "Speech Analysis"

Shannon, C.E. (1949), "Communication in the Presence of Noise", Proc. Institute of Radio Engineers, Vol. 37, No.1, S.10-21

Wikipedia (www), "Ses", <http://tr.wikipedia.org/wiki/Ses>

Wu, H. & Siegel, M. & Khosla, P. (1998), "Vehicle Sound Signature Recognition by Frequency Vector Principal Component Analysis"

**EKLER**

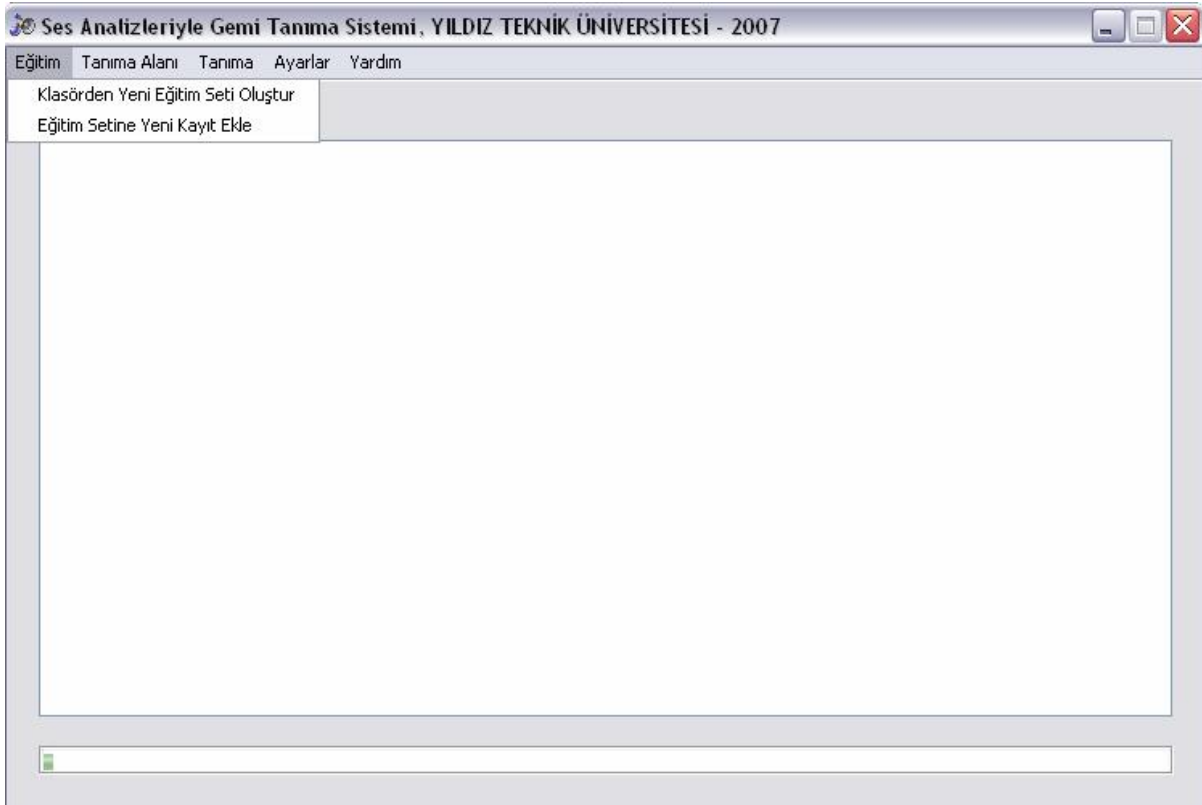
Ek 1 Program Kullanım Kılavuzu

## Ek 1 Program Hakkında Genel Bilgi

Programda eğitim, tanıma alanı, tanıma, ayarlar ve yardım olmak üzere 5 ana modül bulunmaktadır.

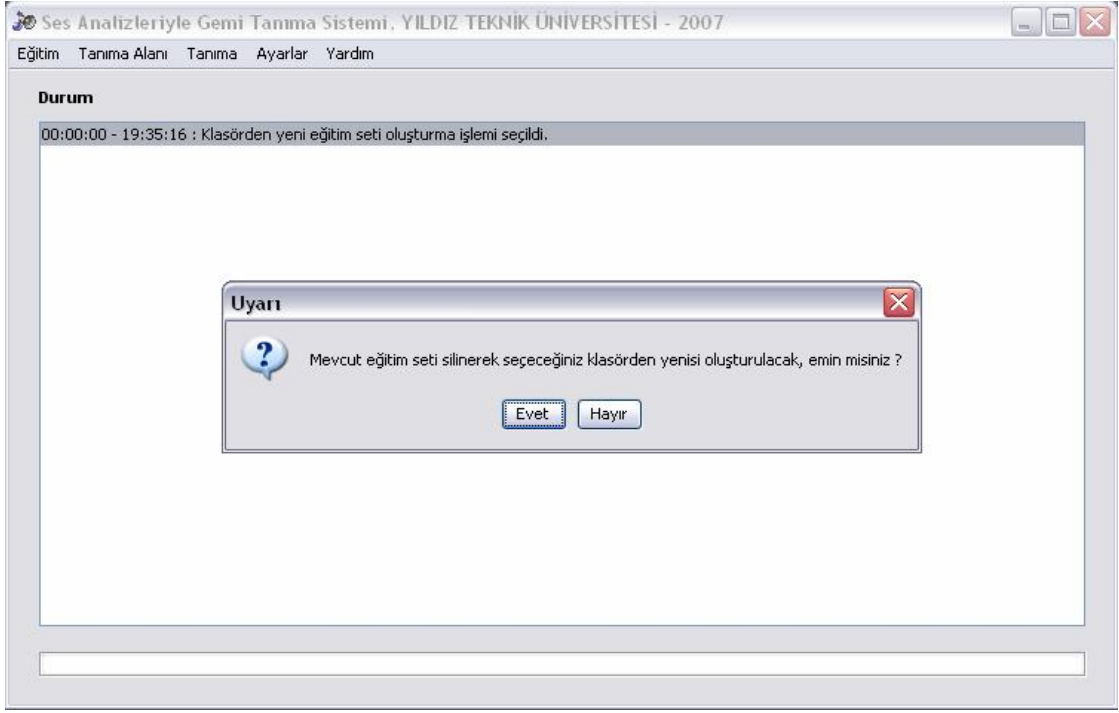
### Eğitim

Program ilk çalıştırıldığında hangi sınıf ve alt sınıfa dâhil olduğu bilinen seslerden eğitim seti yaratılmalıdır. Bunun için wave formatındaki ses dosyaları ana sınıf ve alt sınıf hiyerarşisine uygun olarak klasörlerde depolanılır ve « Eğitim » menüsünden (Şekil Ek 1.1) « Klasörden Yeni Eğitim Seti Oluştur » işlemi seçilir.



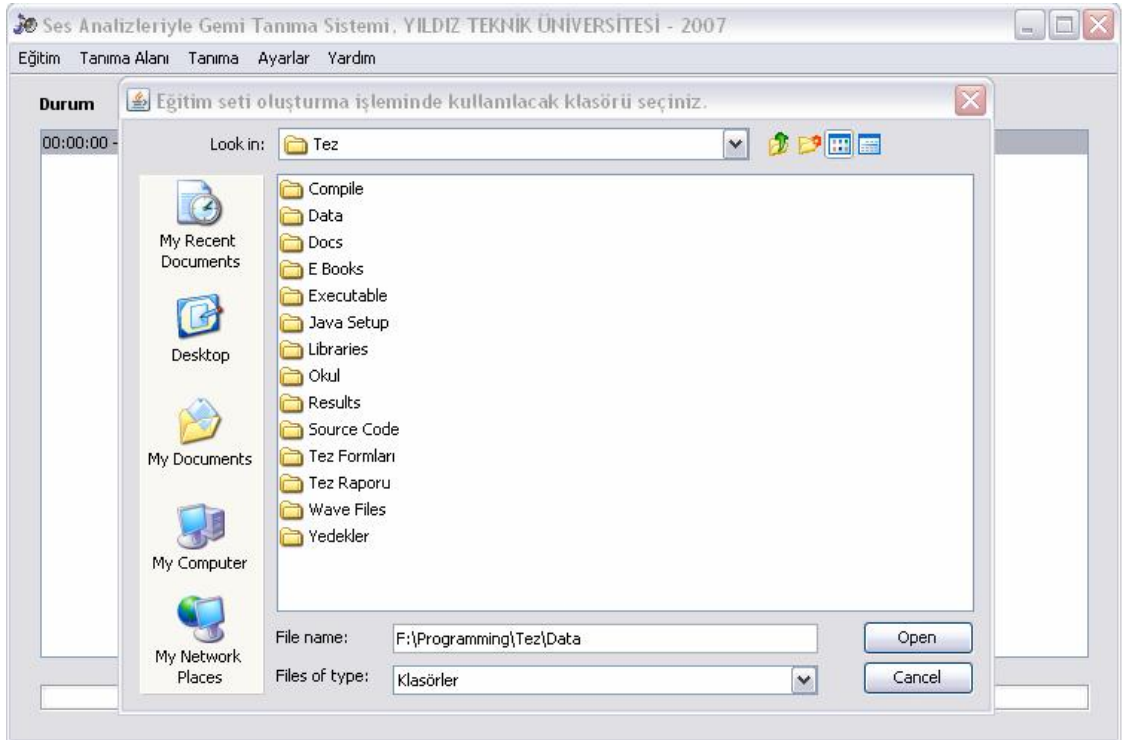
Şekil Ek 1.1 Eğitim Menüsü

Bu seçimin ardından eğer daha önceden oluşturulmuş bir eğitim seti varsa devam edilmesi durumunda bunun silinerek yenisinin oluşturulacağına dair uyarı mesajı görüntülenilir (Şekil Ek 1.2).



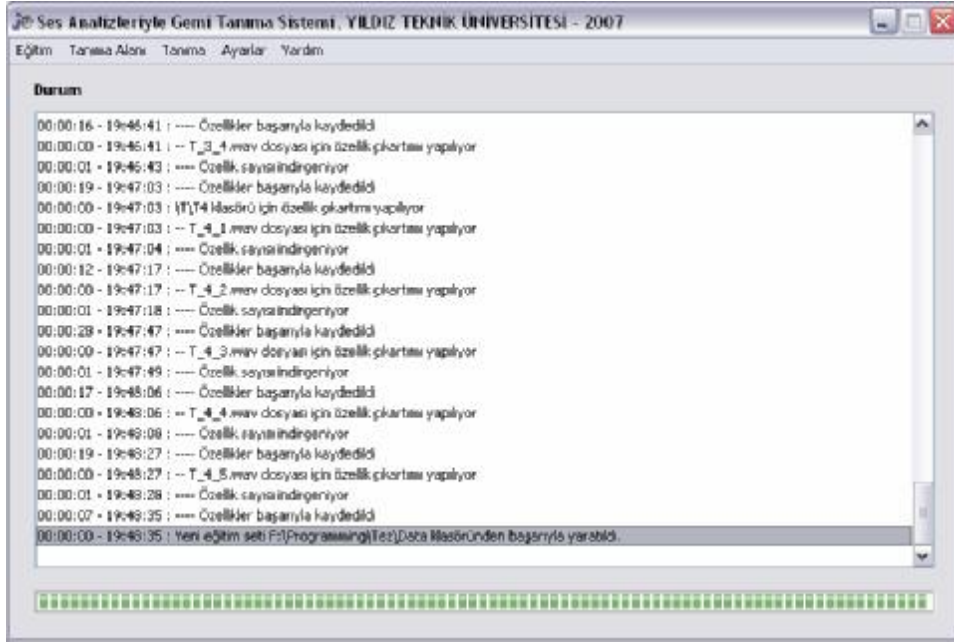
Şekil Ek 1.2 Yeni eğitim seti oluşturulması durumunda mevcut eğitim silineceğini bildiren ve onay veya ret bekleyen uyarı mesajı

Uyarı mesajının onaylanmasının ardından sistem; ana sınıflar, bunlara bağlı alt sınıflar ve ses dosyalarının yer aldığı yeni eğitim setinin oluşturulacağı kök klasörün seçilmesini ister (Şekil Ek 1.3).



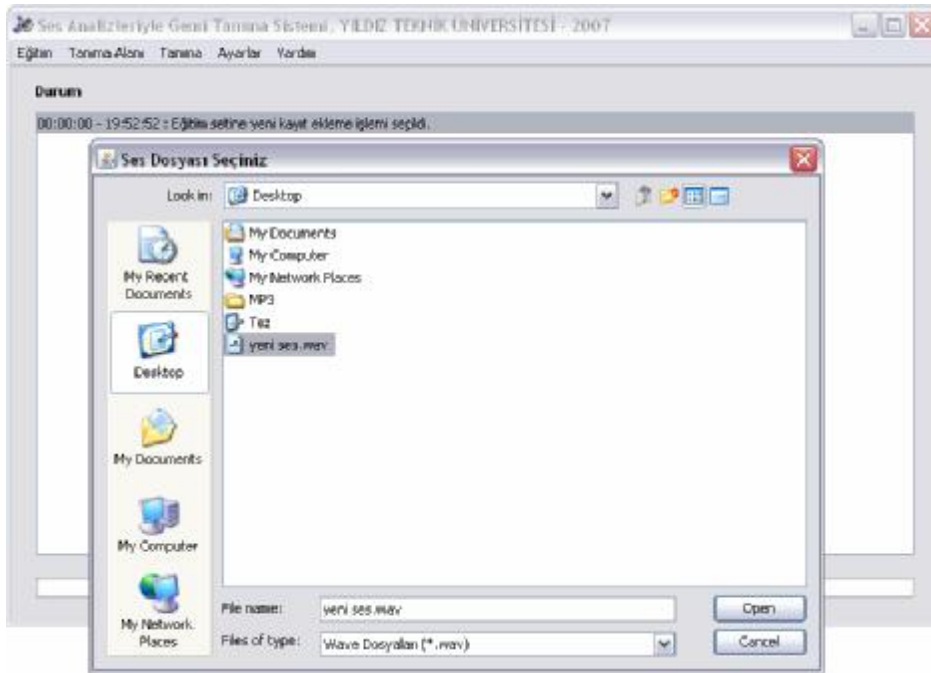
Şekil Ek 1.3 Eğitim seti oluşturmada kullanılacak klasörün seçim ekranı





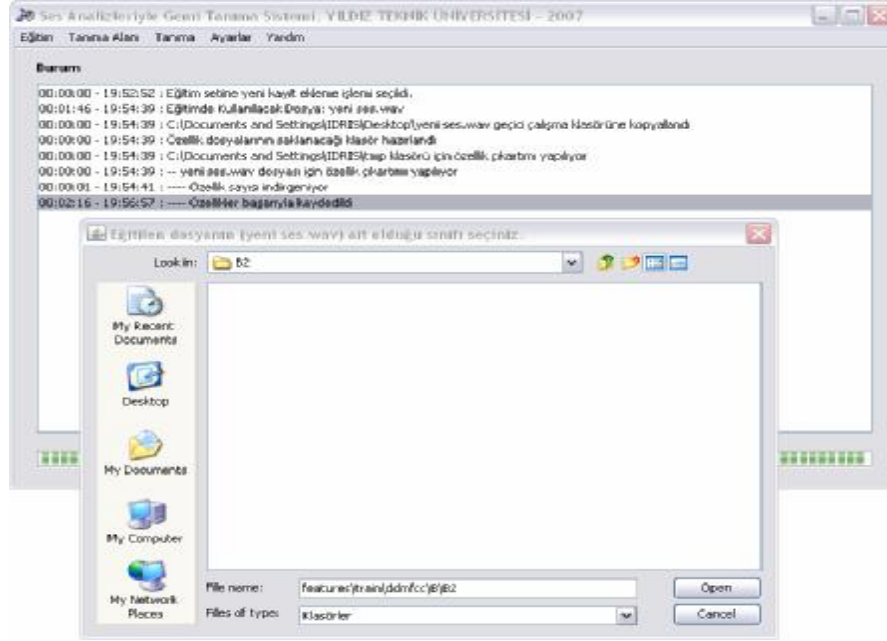
Şekil Ek 1.4 Eğitim seti oluşum sürecini gösteren ekran

Eğitim seti oluşturulduktan sonra zaman içerisinde eğitim setine yeni kayıt ekleme ihtiyacı duyulabilir bu noktada eğitimde kullanılmakta olan tüm sesler için özellik çıkartımına gerek yoktur. « Eğitim » menüsünden (Şekil Ek 1.1) « Eğitim Setine Yeni Kayıt Ekle » işlemi seçilir. Bu seçimin ardından sistem, eğitim setine eklenilmek istenilen dosyanın seçilmesini ister (Şekil Ek 1.5).



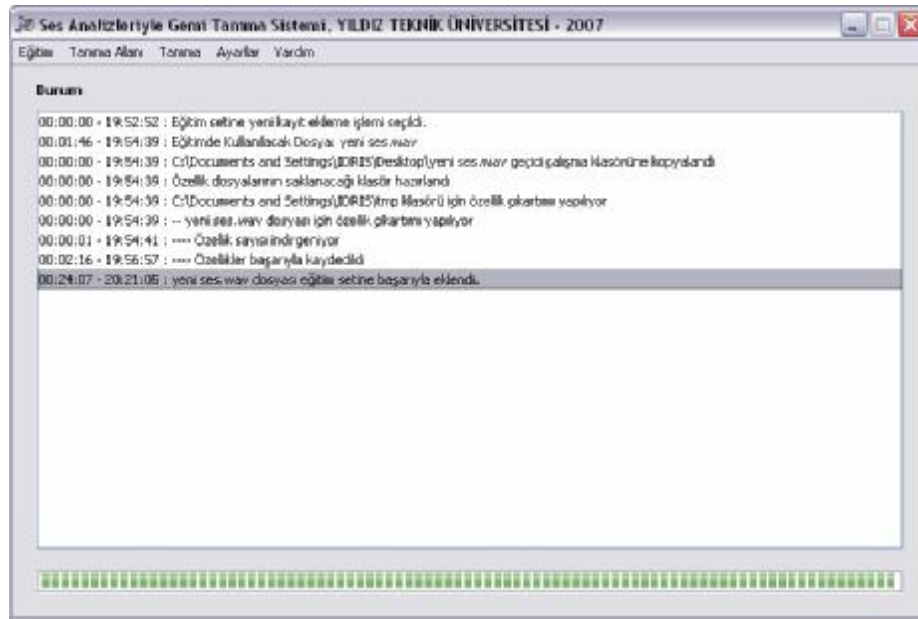
Şekil Ek 1.5 Eğitim setine eklenilmesi istenilen ses kaydının seçim ekranı

Eđitim setine eklenilmesi istenilen ses kaydının seřiminin tamamlanmasının ardından sistem, sese ait karakteristik özellikleri ıkartır ve eđitilen dosyanın hangi sınıfa dâhil edileceđinin seřiminin yapılmasını ister (Şekil Ek 1.6). (Bu noktada yeni bir sınıfa ait yeni bir ses kaydı ekleniliyorsa yeni sınıf adı ile yeni bir klasör oluşturularak, eđitilen ses dosyası bu klasörün altına kaydedilir.)



Şekil Ek 1.6 Eđitilen ses dosyasının dâhil olduđu sınıfın seřim ekranı

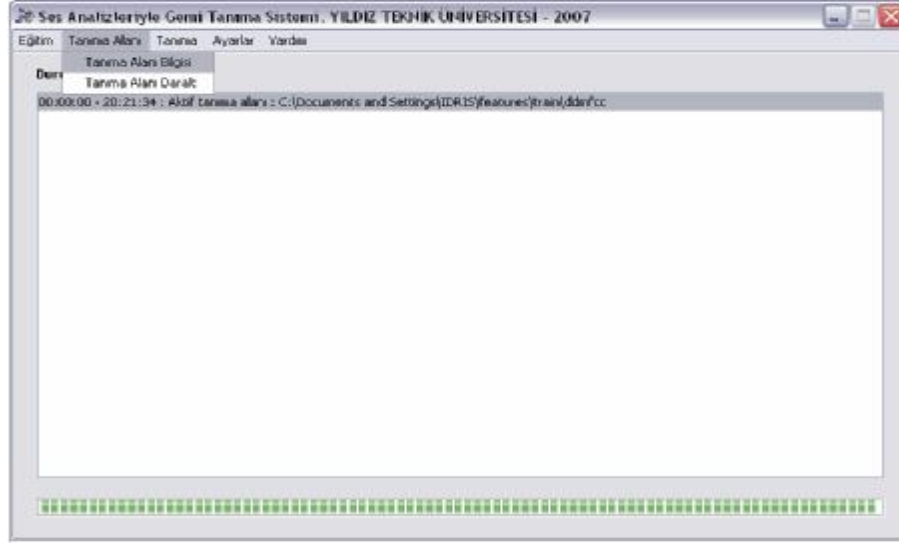
İlgili seřimin yapılmasının ardından ıkartılan özellikler seřilen sınıfın altında yeni bir isimle saklanır (Şekil Ek 1.7).



Şekil Ek 1.7 Yeni ses kaydının eđitim setine başarıyla eklendiđini gösteren ekran

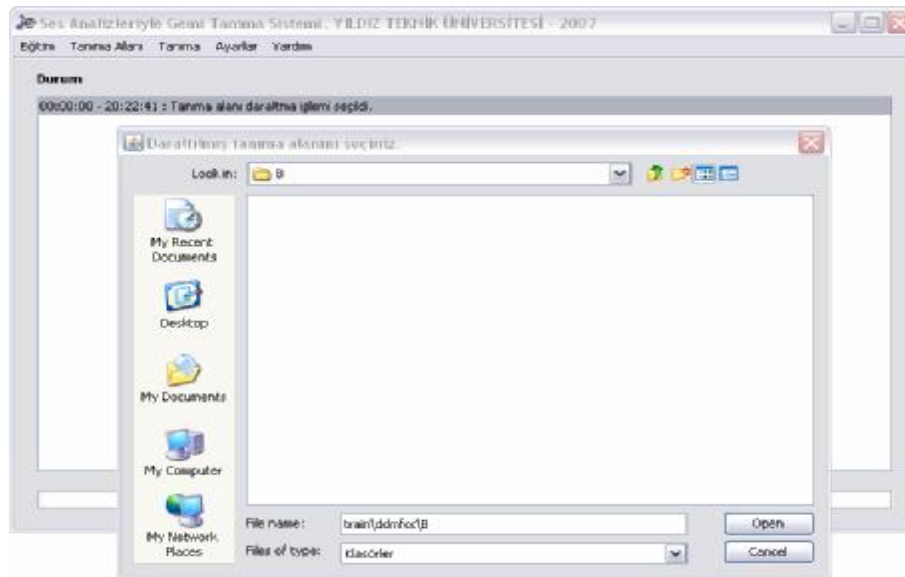
## Tanım Alanı

Program çalışması sırasında hangi sınıfa benzediği merak edilen ses kaydını tüm sınıflarda yer alan örneklerle karşılaştırır ve buna göre bir karar verir. Tanım alanı menüsü kullanılarak programda geçerli olan tanım alanı bilgisi görüntülenebilir (Şekil Ek 1.8).



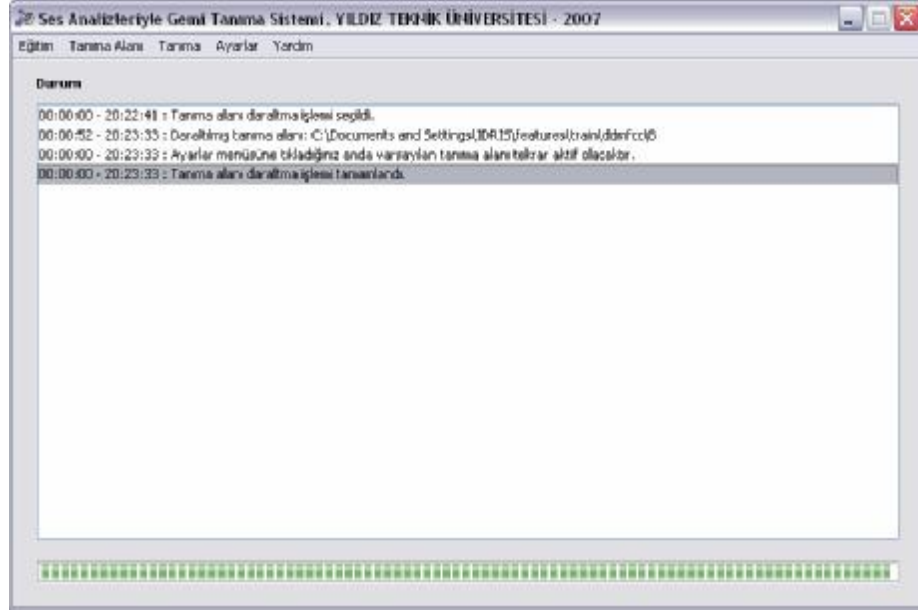
Şekil Ek 1.8 Aktif tanım alanı bilgisi görüntüleme ekranı

Tanım işlemi sırasında test edilen dosyanın belirli bir sınıfın içerisinde yer alan alt sınıflara ne kadar benzediği merak edilirse tanımın sadece bu sınıf ve altında yapılmasını sağlamak amacıyla « Tanım Alanı » menüsünden « Tanım Alanı Daraltma » işlemi seçilir. Bu seçimin ardından sistem, bundan sonra tanımda kullanılacak olan sınırlanmış eğitim setinin seçilmesini ister (Şekil Ek 1.9).



Şekil Ek 1.9 Daraltılmış tanım alanı seçim ekranı

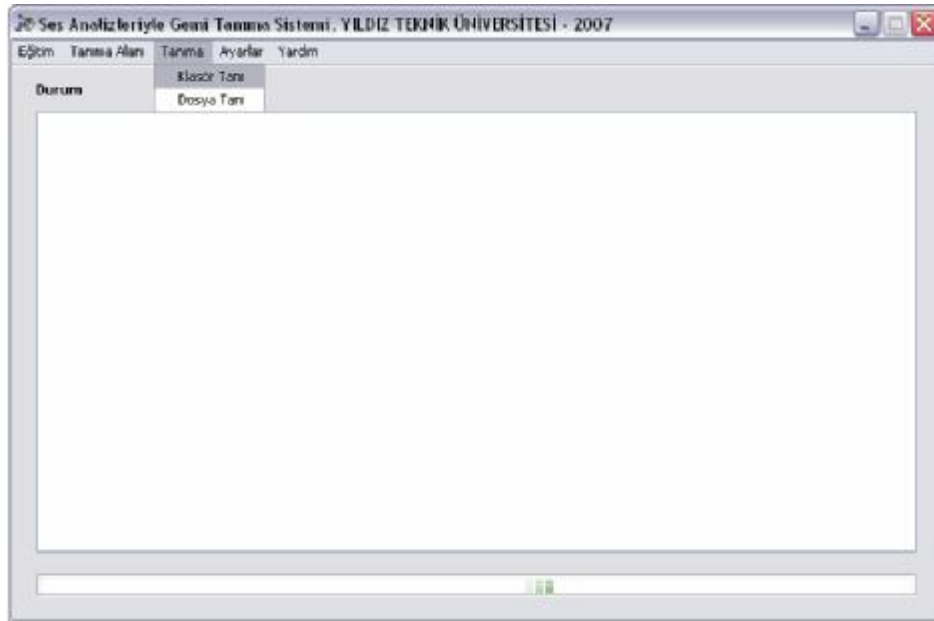
Tanıma alanı « Ayarlar » menüsüne tıklanıldığında ya da programın bir sonraki açılışında yeniden tüm eğitim setini kapsayacak şekilde değiştirilecektir (Şekil Ek 1.10).



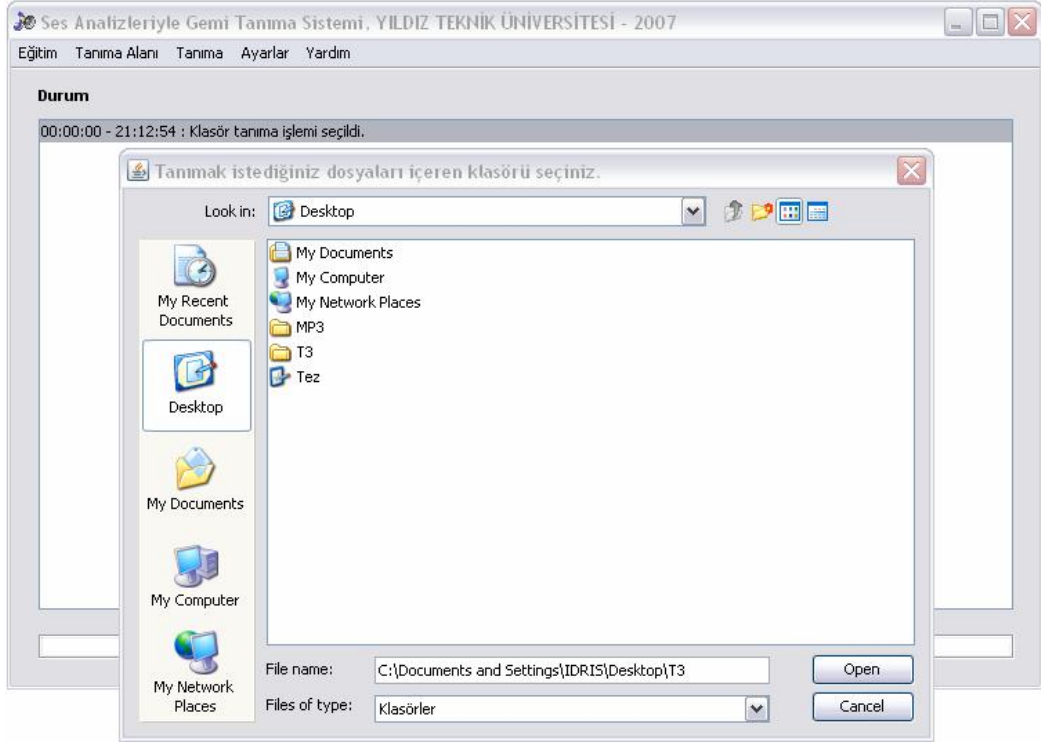
Şekil Ek 1.10 Tanıma alanı daraltma işlemi bilgi ekranı

## Tanıma

Eğitim seti oluşturulduktan sonra sistem, dışarıdan verilen ve hangi sınıfa dâhil olduğu merak edilen seslerin eğitim setindeki mevcut sınıflardan hangisine daha çok benzediği bilgisini verebilir. « Tanıma » menüsünden (Şekil Ek 1.11) « Klasör Tanıma » işlemi seçilirse sistem, tanımda kullanılacak klasörün seçilmesini ister (Şekil Ek 1.12).

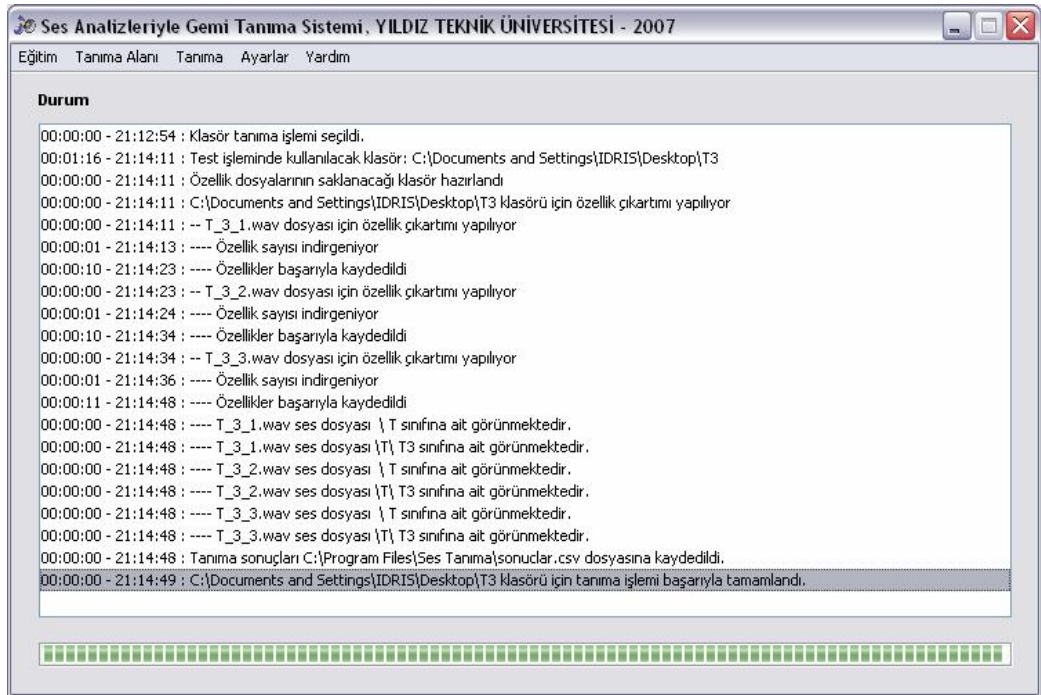


Şekil Ek 1.11 Tanıma menüsü



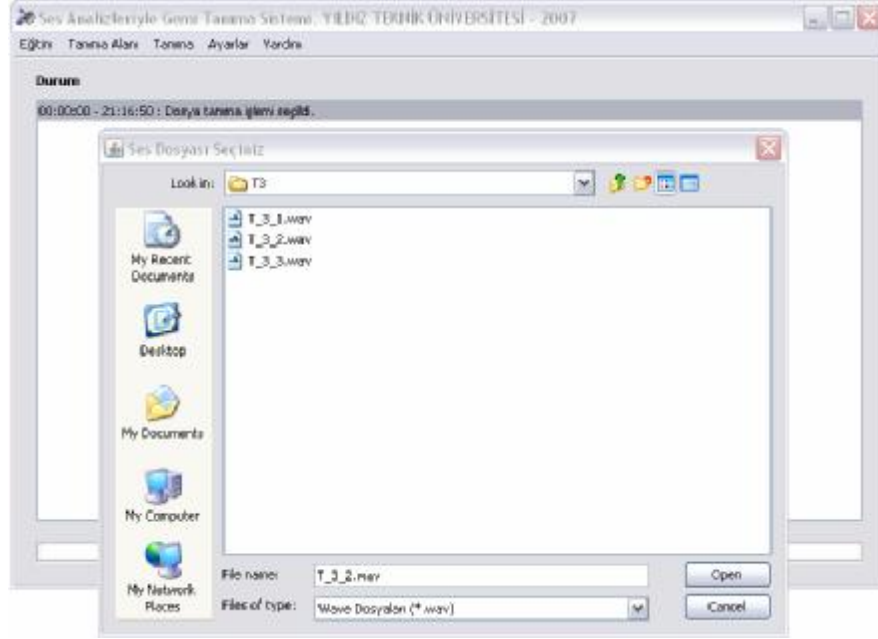
Şekil Ek 1.12 Tanıma işleminde kullanılacak klasör seçim ekranı

Tanıma işleminde kullanılacak klasörün seçiminin ardından, bu klasör ve altında yer alan klasörlerde bulunan tüm dosyalar için tanıma işlemini gerçekleştirir. Tanıma işleminin sonucu ekranda özet olarak kullanıcıya sunulur (Şekil Ek 1.13). Ayrıca benzerlik oranlarının da yer aldığı detaylı bilgi programın kurulu olduğu klasörde csv formatında dosyalar olarak saklanır.



Şekil Ek 1.13 Klasör tanıma işlemi sonuç ekranı

Tanıma işlemi sırasında sadece bir ses kaydının hangi sınıfa benzediği merak ediliyorsa «Tanıma » menüsünden (Şekil Ek 1.11) « Dosya Tanıma » işlemi seçilir. Bu seçimin ardından sistem, tanıma işleminde kullanılacak ses kaydının seçilmesini ister (Şekil Ek 1.14).



Şekil Ek 1.14 Dosya tanımada kullanılacak ses kaydının seçim ekranı

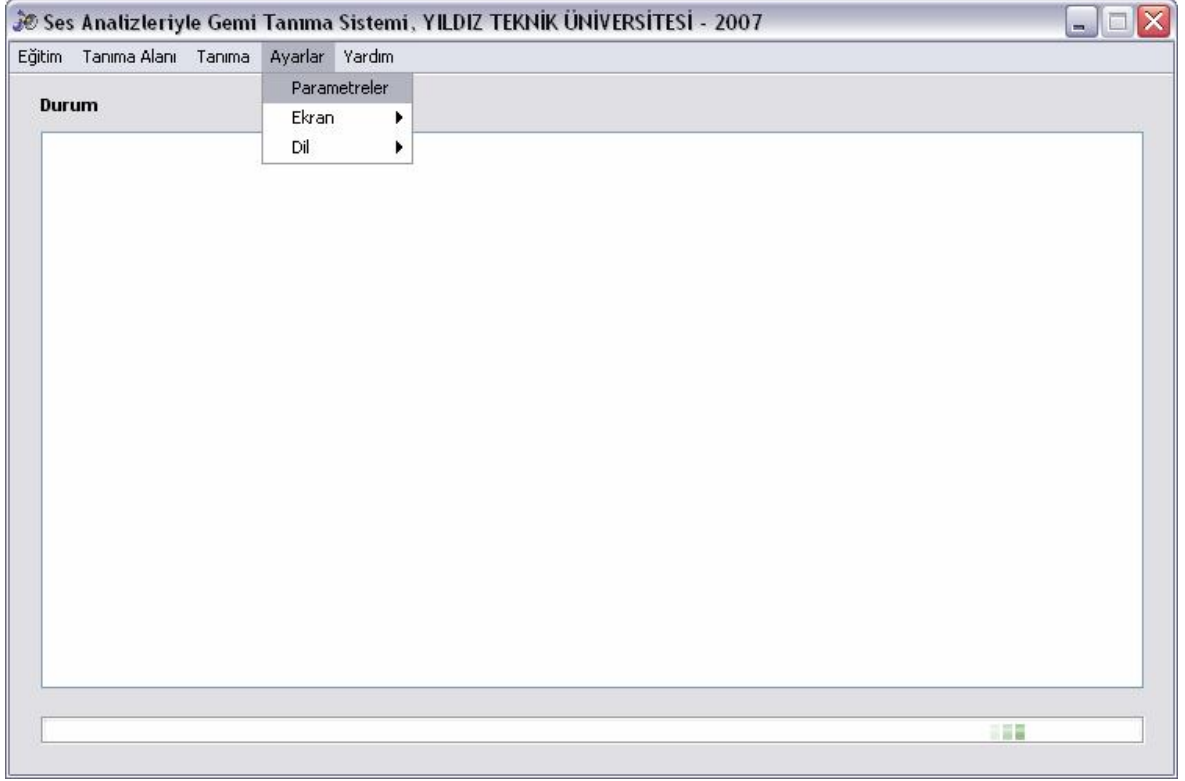
Tanıma işleminde kullanılacak ses kaydının seçiminin ardından, seçili dosya için tanıma işlemi gerçekleştirilir. Tanıma işleminin sonucu ekranda özet olarak kullanıcıya sunulur (Şekil Ek 1.15). Ayrıca benzerlik oranlarının da yer aldığı detaylı bilgi programın kurulu olduğu klasörde csv formatında dosyalar olarak saklanılır.



Şekil Ek 1.15 Dosya tanıma işlemi sonuç ekranı

## Ayarlar

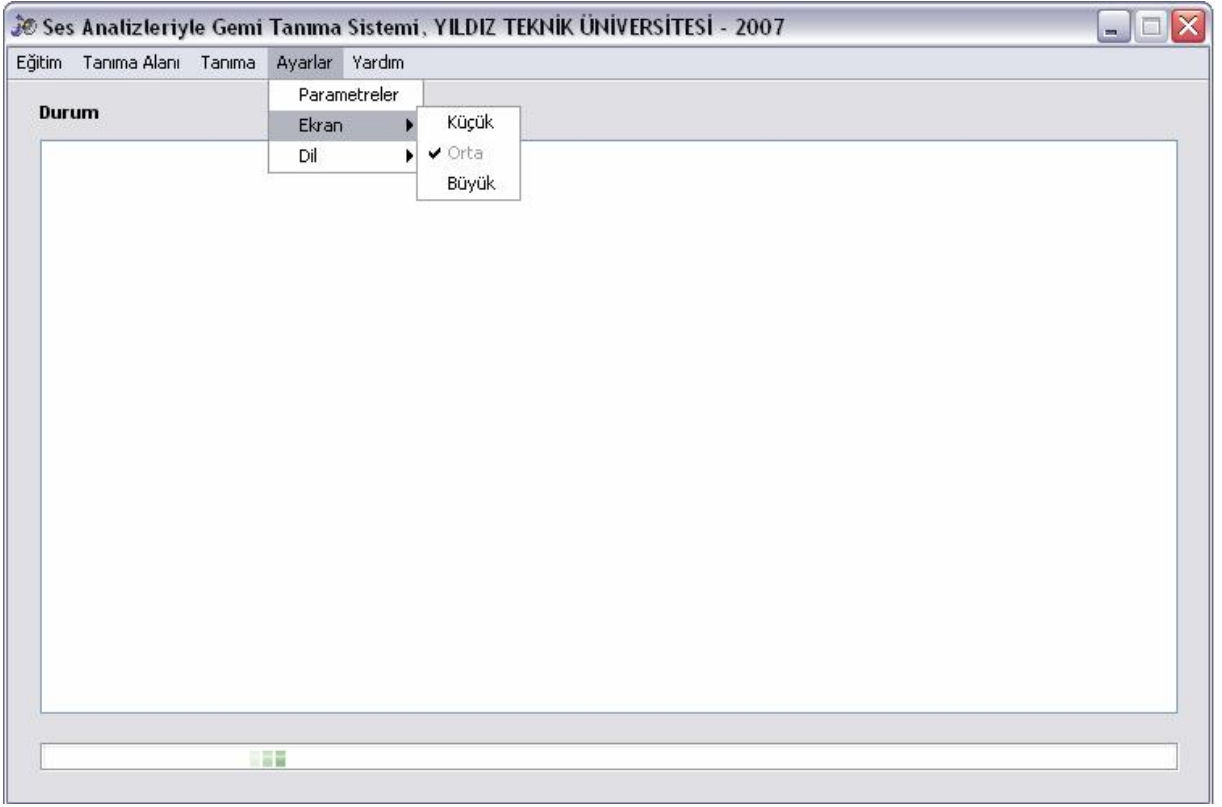
« Ayarlar » menüsü (Şekil Ek 1.16) kullanılarak eğitim ve tanıma işlemleri sırasında sistemin işleyişini değiştirmek amacıyla kullanılacak olan parametreler (Şekil Ek 1.17) ile programın kullanıcı arayüzünü etkileyen uygulama penceresinin boyutlandırılması (Şekil Ek 1.18) ve kullanım dilinin değiştirilmesi (Şekil Ek 1.19) işlemleri yapılır.



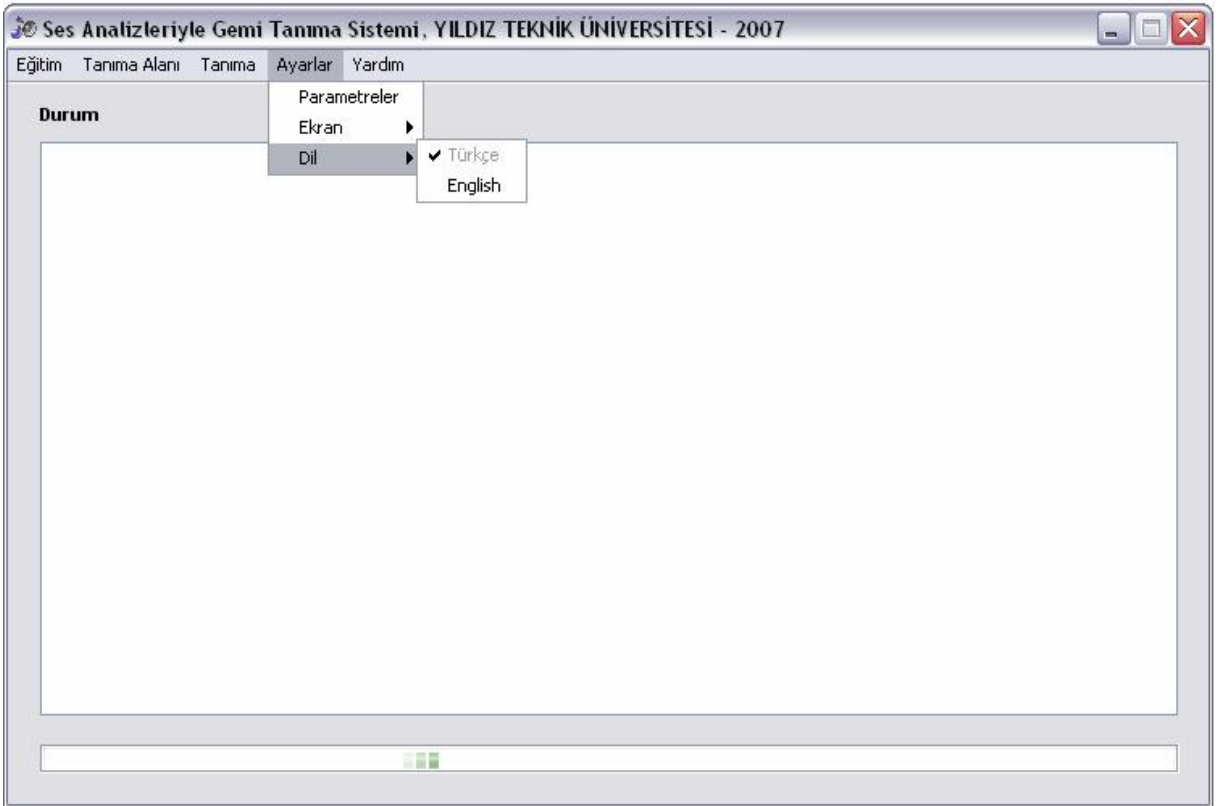
Şekil Ek 1.16 Ayarlar menüsü



Şekil Ek 1.17 Parametreler ekranı



Şekil Ek 1.18 Ekran boyutlandırma menüsü

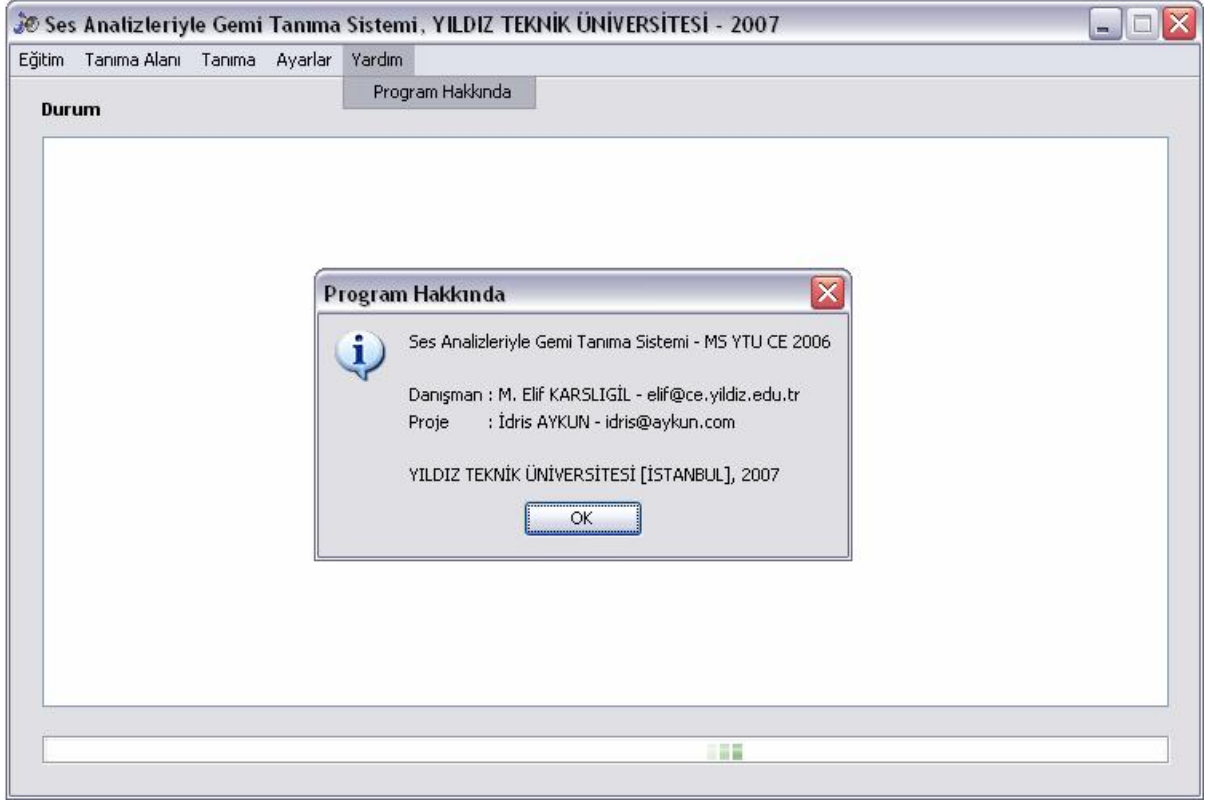


Şekil Ek 1.19 Dil seçimi menüsü



## Yardım

Yardım menüsünde program hakkında özet bilgiye yer almaktadır (Şekil Ek 1.20).



Şekil Ek 1.20 Program bilgi ekranı

**ÖZGEÇMİŞ**

Adı Soyadı İdris AYKUN

Doğum Tarihi 11.09.1981

Doğum Yeri İstanbul

E Posta [idris@aykun.com](mailto:idris@aykun.com)

Web Adresi <http://www.aykun.com>

Lisans Yıldız Teknik Üniversitesi, Bilgisayar Mühendisliği (İstanbul, 2004)

Lise Sakıp Sabancı Anadolu Lisesi, Fen-Matematik (İstanbul, 1999)

Bitirme Projesi (Lisans) Tek Enstrümanla Çalınan Müzik Parçalarının Notalarının Bilgisayar Ortamında Otomatik Olarak Bulunması  
Yıldız Teknik Üniversitesi - İstanbul, 2003

**Proje Yürütücüsü:** Asist. Prof. M. Elif KARSLIĞİL, [elif@ce.yildiz.edu.tr](mailto:elif@ce.yildiz.edu.tr)

**Proje Grubu:** İdris AYKUN (99011054)

**Proje Tanımı:** Sayısal işaret işleme ile bir müzik dosyasında çalınan notaların tespit edilmesi (DSP, Fourier Transform ...)

**Proje Özeti**

Bu projede, bir enstrümanla çalınan müzik seslerini işleyerek notalarını belirleyen bir sistem geliştirilmiştir. Bu yolla üzerinde işlem yapılan müzik dosyasının notaları bilgisayar ortamında gösterilmesi amaçlanmış bu kapsamda spektrum analizi yaparak üzerinde tanıma işleminin gerçekleştirileceği müziğe ait spektrogram oluşturularak müzik notaları belirlenmiştir.

Projede her müzik notasının frekans alanındaki görünümünün saklanması yerine (bu yöntem mühendislik ve depolama maliyetine neden olmaktadır) sadece seçilen bir oktava ait notaların frekans değerleri saklanmış ve girilen frekans değerini ait olduğu oktav ve nota değerine karşı düşüren fonksiyon geliştirilmiştir.

Yazılım geliştirme aşamasında programlama dili olarak python programlama dili kullanılmış, üzerinde çalışılacak geliştirme ortamı olarak da Red Hat Linux işletim sistemi tercih edilmiştir; bu geliştirilen programın sadece UNIX türevleri üzerinde çalışabileceği şeklinde düşünülmemelidir. Geliştirilen yazılım kros-platfrom olarak çalışabilen taşınabilir işletim sisteminizden bağımsız bir yazılımdır.

Geliştirilen yazılım yardımıyla amatör müzisyenler notalarını bilmedikleri fakat cd veya kasetine sahip oldukları müzik parçalarının notalarına istedikleri anda kolaylıkla ulaşabilmelerini sağlayacağı gibi profesyonel müzisyenleri de müzik notalarını kağıtlardan takip etme zahmetinden kurtarabilir.