

**İSTANBUL TEKNİK ÜNİVERSİTESİ ★ FEN BİLİMLERİ ENSTİTÜSÜ**

**GAUSS KARIŞIM MODELLERİ KULLANILARAK  
SES İMZALARININ SINIFLANDIRILMASI**

**YÜKSEK LİSANS TEZİ  
Müh. Kadir HERKİLOĞLU  
(504031311)**

**Anabilim Dalı : Elektronik ve Haberleşme Mühendisliği**

**Programı : Telekomünikasyon Mühendisliği**

**Tez Danışmanı: Prof. Dr. Bilge GÜNSEL**

**AĞUSTOS 2005**

## **ÖNSÖZ**

Akademik çalışmalarım sırasında her türlü desteğini esirgemeyen değerli hocam Prof. Dr. Bilge GÜNSEL' e ve bugünlere gelmemi sağlayan aileme ve yakınlarıma teşekkürü bir borç bilirim. Ayrıca, üzerimde çok emeği olan ağabeyim Yusuf YASLAN'a ve proje çalışma arkadaşlarım, Samet YASLAN ve Serap KIRBİZ'a teşekkür ederim.

AĞUSTOS 2005

Kadir HERKİLOĞLU

## İÇİNDEKİLER

<b>KISALTMALAR</b>	<b>v</b>
<b>TABLO LİSTESİ</b>	<b>vi</b>
<b>ŞEKİL LİSTESİ</b>	<b>vii</b>
<b>SEMBOL LİSTESİ</b>	<b>viii</b>
<b>ÖZET</b>	<b>ix</b>
<b>SUMMARY</b>	<b>x</b>
<b>1. GİRİŞ</b>	<b>1</b>
<b>1.1 Ses İmzasının Tarihçesi</b>	<b>2</b>
<b>1.2 Ses İmzası Tanıma Sistemleri</b>	<b>3</b>
<b>2. GELİŞTİRİLEN SİSTEMİN GENEL YAPISI VE SES İMZALARININ OLUŞTURULMASI</b>	<b>6</b>
<b>2.1 Ses İmzası Oluşturma</b>	<b>8</b>
2.1.1 Kullanılan pencere ve çerçeveler	9
2.1.2 Öznitelik çıkarma	10
2.1.2.1 Sıfır geçiş sayısı (SGS)	10
2.1.2.2 Kök ortalama karesel enerji (KOKE)	12
2.1.2.3 Kepstral akı (KA)	13
2.1.2.4 İzgesel akı (İA)	15
2.1.2.5 İzgesel kitle merkezi (İKM)	16
2.1.2.6 İzgesel azalma noktası (İAN)	18
2.1.2.7 Mel-Frekansı keppstrum katsayıları (MFKK)	19
2.1.3 Son işleme	19
<b>3. SES İMZASI MODELLEME VE SINIFLANDIRMA</b>	<b>20</b>
<b>3.1 Gauss Karışım Modeli (GKM)</b>	<b>20</b>
3.1.1 Beklenti enbüyükleme (BE) algoritması	21
3.1.2 Gauss karışım modeli için BE algoritması	23
3.1.3 GKM parametrelerinin analizi	25
<b>3.2 Destek Vektör Makinesi (DVM)</b>	<b>25</b>
3.2.1 Doğrusal ayrılabilir durum	25
3.2.2 Doğrusal ayrılamaz durum	27
3.2.3 Doğrusal olmayan durum	28
<b>3.3 Ses İmzası Sınıflandırma</b>	<b>29</b>
<b>4. PERFORMANS TESTLERİ</b>	<b>31</b>
<b>4.1 Tür Ayrımı Testleri</b>	<b>31</b>
<b>4.2 Parça Ayrımı Testleri</b>	<b>35</b>
<b>4.3 Performans Testleri Sonuçları</b>	<b>36</b>
<b>5. GÜRBÜZLÜK TESTLERİ</b>	<b>38</b>
<b>5.1 MP3 Sıkıştırma Atağı</b>	<b>38</b>
<b>5.2 Zamanda Sıkıştırma Atağı</b>	<b>39</b>
<b>5.3 Beyaz Gauss Gürültüsü Atağı</b>	<b>40</b>

<b>5.4 0-10kHz Alt Geçiren Süzgeç ile Filtreleme Atađı</b>	<b>42</b>
<b>6. SONUÇLAR VE TARTIŞMALAR</b>	<b>43</b>
<b>KAYNAKLAR</b>	<b>45</b>
<b>EKLER</b>	<b>48</b>
<b>ÖZGEÇMİŞ</b>	<b>59</b>

## KISALTMALAR

<b>KA</b>	:Kepstral Akı (Cepstral Flux)
<b>BDA</b>	:Bozulma Diskriminant Analizi (Distortion Discriminant Analysis)
<b>BE</b>	:Beklenti Enbüyükleme (Expectation Maximisation)
<b>DVM</b>	:Destek Vektör Makinesi (Support Vector Machine)
<b>GKM</b>	:Gauss Karışım Modeli (Gaussian Mixture Model)
<b>İA</b>	:İzgesel Akı (Spectral Flux)
<b>İDÖ</b>	:İzgesel Düzlük Ölçüsü (Spectral Flatness Measure)
<b>İKM</b>	:İzgesel Kitle Merkezi (Spectral Centroid)
<b>İAN</b>	:İzgesel Azalma Noktası (Spectral Roll Off Point)
<b>KOKE</b>	:Kök Ortalama Karesel Enerji (Root Mean Square Energy)
<b>MFKK</b>	:Mel Frekansı Kepstral Katsayıları (Mel Frequency Cepstral Coefs.)
<b>MPEG</b>	:Moving Pictures Expert Group
<b>ms</b>	:Milisaniye
<b>s</b>	:Saniye
<b>SGS</b>	:Sıfır Geçiş Sayısı (Zero Crossing Rate)
<b>SNR</b>	:İşaret Gürültü Oranı (Signal to Noise Ratio)
<b>YTBA</b>	:Yönlü Temel Bileşen Analizi (Oriented Principle Comp. Analysis)

## TABLO LİSTESİ

	<u>Sayfa No</u>
<b>Tablo 4.1.</b> Tür ayrımı testlerinde kullanılan veri seti, türler ve sınıfları.....	31
<b>Tablo 4.2.</b> Müzik türlerinin GKM ile sınıflandırma sonuçları .....	32
<b>Tablo 4.3.</b> GKM ile yapılan sınıflandırma sonucu oluşan pozitif yanlış alarm oranları.....	32
<b>Tablo 4.4.</b> Müzik türlerinin DVM ile sınıflandırma sonuçları.....	33
<b>Tablo 4.5.</b> DVM ile yapılan sınıflandırma sonucu oluşan pozitif yanlış alarm oranları.....	34
<b>Tablo 4.6.</b> Parça ayrımı testlerinde kullanılan veri seti, türler, sanatçılar, örnek sayıları ve parça adetleri.....	35
<b>Tablo 4.7.</b> GKM ile parça ayrımı pozitif yanlış alarm oranları.....	36
<b>Tablo 5.1.</b> mp3 sıkıştırma atağı müzik türü sınıflandırma başarımı.....	38
<b>Tablo 5.2.</b> mp3 sıkıştırma atağı parça sınıflandırma başarımı.....	38
<b>Tablo 5.3.</b> Sıkıştırılmış verilerin de eğitimde kullanılması sonucu elde edilen müzik türü sınıflandırması başarımları.....	39
<b>Tablo 5.4.</b> Sıkıştırılmış verilerin de eğitimde kullanılması sonucu elde edilen parça sınıflandırması başarımları.....	39
<b>Tablo 5.5.</b> Zamanda sıkıştırma atağı müzik türü sınıflandırması başarımları.....	40
<b>Tablo 5.6.</b> Zamanda sıkıştırma atağı parça sınıflandırması başarımları.....	40
<b>Tablo 5.7.</b> Gauss gürültüsü atağı müzik türü sınıflandırması başarımları.....	40
<b>Tablo 5.8.</b> Gauss gürültüsü atağı parça sınıflandırması başarımları.....	41
<b>Tablo 5.9.</b> Gürültü eklenmiş verilerin de eğitimde kullanılması sonucu elde edilen müzik türü sınıflandırması başarımları.....	41
<b>Tablo 5.10.</b> Gürültü eklenmiş verilerin de eğitimde kullanılması sonucu elde edilen müzik türü sınıflandırması başarımları.....	41

## ŞEKİL LİSTESİ

	<u>Sayfa No</u>
Şekil 2.1 : Ses imzası sistemleri genel yapısını gösterir şema.....	6
Şekil 2.2 : Geliştirilen ses imzası sisteminin blok şeması.....	8
Şekil 2.3 : Alt ve üst çerçevenin şematik gösterimi.....	9
Şekil 2.4 : 67 parçadan çıkarılan ortalama SGS öznelikleri.....	11
Şekil 2.5 : 67 parçadan çıkarılan deęişinti SGS öznelikleri.....	11
Şekil 2.6 : 67 parçadan çıkarılan ortalama KOKE öznelikleri.....	12
Şekil 2.7 : 67 parçadan çıkarılan deęişinti KOKE öznelikleri.....	13
Şekil 2.8 : 67 parçadan çıkarılan ortalama KA öznelikleri.....	14
Şekil 2.9 : 67 parçadan çıkarılan deęişinti KA öznelikleri.....	14
Şekil 2.10 : 67 parçadan çıkarılan ortalama İA öznelikleri.....	15
Şekil 2.11 : 67 parçadan çıkarılan deęişinti İA öznelikleri.....	16
Şekil 2.12 : 67 parçadan çıkarılan ortalama İKM öznelikleri.....	17
Şekil 2.13 : 67 parçadan çıkarılan deęişinti İKM öznelikleri.....	17
Şekil 2.14 : 67 parçadan çıkarılan ortalama İAN öznelikleri.....	18
Şekil 2.15 : 67 parçadan çıkarılan deęişinti İAN öznelikleri.....	18
Şekil 4.1 : Müzik türlerinin GKM ile sınıflandırma sonuçları.....	32
Şekil 4.2 : GKM ile yapılan sınıflandırma sonucu oluşan pozitif yanlış alarm oranları.....	33
Şekil 4.3 : Müzik türlerinin DVM ile sınıflandırma sonuçları.....	34
Şekil 4.4 : DVM ile yapılan sınıflandırma sonucu oluşan pozitif yanlış alarm oranları.....	35

## SEMBOL LİSTESİ

$\Sigma$	: GKM ile eğitilen sistemin Kovaryans Matrisi
$\mu$	: GKM ile eğitilen sistemin her sınıf oluşturduğu ortalama vektörleri
$w$	: GKM ile eğitilen sistemin karışımların çarpan vektörü
$a$	: DVM ile eğitilen sistemin destek vektörleri
$\theta$	: GKM ile eğitilen sistemin parametreler vektörü
$S_{i,tr}$	: Eğitim kümesindeki i numaralı 2 sn'lik ses bölütü
$S_{i,ts}$	: Test kümesindeki j numaralı 2 sn'lik ses bölütü
$x_i$	: i numaralı ses bölütünden oluşturulmuş ses imzası vektörü
$x_j$	: j numaralı ses bölütünden oluşturulmuş ses imzası vektörü
$s_i$	: i numaralı 1024 örnekli ses çerçevesi
$k$	: örnek indisi
$N$	: Bir alt çerçevedeki örnek sayısı (N=1024)
$C_i$	: i numaralı çerçevenin kepsstrum katsayı vektörü
$S_i$	: i numaralı çerçevenin güç izgesi
$d$	: boyut sayısı
$N()$	: Normal dağılım
$A(\theta, \theta^s)$	: Yardımcı İşlev
$Q$	: Gizli Değişken
$\theta^s$	: s. iterasyondaki parametre vektörü
$A_{tr}$	: Eğitim kümesi
$K()$	: Çekirdek işlev



## GAUSS KARIŞIM MODELLERİ KULLANILARAK SES İMZALARININ SINIFLANDIRILMASI

### ÖZET

Bu tez çalışması kapsamında, Gauss karışım modelleri kullanılarak ses imzalarının sınıflandırılması gerçekleştirilmiştir. Geliştirilen sistemde, daha önceden sisteme eğitim için verilen ses verileri kullanılarak, ses imzaları modellenmekte ve sistem girişine verilen 2 saniyelik ses bölütünün hangi parçaya veya hangi müzik türüne ait olduğu yüksek hızla, yüksek doğruluk oranıyla ve düşük yanlış alarm oranıyla bulunabilmektedir. Geliştirilen sistemle literatürde bulunan sistemlerle aynı tanıma performansına ulaşılmıştır.

Tez kapsamında ses imzası modelleme Gauss Karışım Modelleri kullanılarak yapılmış ve imzalar Bayes sınıflandırıcı ile sınıflandırılmıştır. Karşılaştırma açısından testler Destek Vektör Makinesi kullanılarak ses imzalarını öğrenen ve sınıflandıran bir sistem için de tekrarlanmıştır. Testler kapsamında tür ve parça ayrımı performansları incelenmiştir. Bunlara ek olarak sistemin, mp3 sıkıştırma, kanal gürültüsü ekleme, zamanda sıkıştırma ve 10kHz kesim frekanslı alt geçiren süzgeçle filtreleme ataklarına karşı gürbüzlüğü test edilmiş ve sonuçları tablolar ve grafikler halinde sunulmuştur. Buna göre sistem, mp3 sıkıştırma atağında 128kBit'lik kodlamaya değerine kadar gürbüzlüğü koruyabilirken, 40dB'lik işaret/gürültü oranının altındaki oranlarda eklenen Beyaz Gauss Gürültüsüne karşı gürbüzlüğü koruyamamaktadır. Buna karşın zamanda sıkıştırma atağına karşı %4'lük sıkıştırma oranında bile gürbüzlüğü kaybetmemektedir. Zamanda sıkıştırma işlemine karşı gürbüzlük öznitelik-tabanlı sınıflandırma ile ses tanımanın literatürde bulunan eğitimcisziz sistemlere karşı önemli bir üstünlüğüdür. Geliştirilen ses imzası tanıma sisteminin işlemsel karmaşıklığı düşük olup hızlı modelleme ve sınıflandırma yapabildiği gibi arama veritabanının kolaylıkla güncellenmesine olanak tanır. Bu da geniş müzik veritabanlarının kolaylıkla işlenmesine olanak sağlamaktadır.

## **CLASSIFICATION OF AUDIO FINGERPRINTS USING GAUSSIAN MIXTURE MODELS**

### **SUMMARY**

In this thesis, a system that classifies audio fingerprints using Gaussian Mixture Models classifier, is proposed. The system has the ability to decide the music clip and the music type of a 2 seconds long audio segment, fast, with a high accuracy and with low positive false alarm ratio. The proposed system brings an innovation by providing robustness to time compression attacks to which most of the current systems are not robust. The classifier gives the results much faster than ordinary searching algorithms with a high identification percentage. 2 second granularity is the second innovation proposed in this thesis work. This level of granularity is succeeded with keeping the high performance and it is below the MPEG21 standard. The performance tests for identification of audio fingerprints are performed by using GMM classifier. However, in order to compare the results the same tests with the same conditions are repeated by using Support Vector Machine. In these tests, the ability to identify the clip and the music type of the segments is observed. Furthermore, the robustness of the system to mp3 compression, White Gaussian channel noise adding, time compression, and 10kHz low pass filtering attacks is examined. According to the results the proposed system is robust mp3 compression better than 128kBit coding, Gaussian noise above 40dB SNR, and among the best it is robust to time compression even though the ratio is %4. However, it is not robust to 0-10kHz low pass filtering. Beyond these, the proposed system has a low computational complexity and can easily be updated. So it provides fast and automatic process of large audio databases.

## 1. GİRİŞ

Elektronik teknolojisindeki hızlı gelişme veri depolama olanaklarını artırmış bunun sonucunda otomatik olmayan yöntemlerle işlenmesi uzun zaman alan büyük boyutlu ses verileri kullanan çeşitli uygulamaları olanaklı kılmıştır. Bu durum büyük boyutlu ses verisinin otomatik işlenmesi gereksinimini doğurmuştur. Buna bağlı olarak, konu üzerine çeşitli bilimsel araştırmalar yapılmıştır. Otomatik ses bölütleme, veri tabanından geri çağırım, müzik türü sınıflandırması bu çalışmalar arasındadır.

Ses imzası, bir müzik parçasını yegâne tanımlayabilen veri dizisi olarak tanımlanmaktadır. Bu tanımıyla ses imzası, insanlardaki parmak izine benzemektedir. Kavram olarak parmak izi, insanı tanımaya, diğerlerinden ayırt etmeye yetecek bilgiyi içerisinde saklamaktadır. Ancak gözden kaçmamalıdır ki parmak izinden o kişiye ait tüm bilgiler elde edilemez. Saç rengi, boyu, kilosu gibi kişiye ait nitelikler parmak izinde saklı değildir. Aynı şekilde ses imzası da ses klipleriyle ilgili özet ama yeterli bilgiyi tutmaktadır.

Ses verisi çok uzun ve büyük hacimler kaplayan bir veri türü olduğundan işlenmesi oldukça uzun sürmektedir. Bu durumda da küçük bir ses bölütünü uzun bir ses katarı içerisinde aramak veya geniş bir veritabanında aramak çok zahmetli bir iş olmaktadır. Bunu kolaylaştırmanın yolu ise verinin taşıdığı bilgiyi azaltmaksızın veriyi azaltmaktır. Ses imzası ilk olarak bu soruna çözüm olarak önerilmiştir[1].

Tez kapsamında ses bölütleme öznitelikleri ve Gauss Karışım Modeli kullanan ses imzası sistemi geliştirilmiştir. Sistem 2sn'lik bir ses bölütünün ait olduğu müzik klipini doğru ve hızlı olarak bulabildiği gibi parçanın hangi müzik türüne ait olduğunu da belirleyebilmektedir. Geliştirilen sistem temel olarak üç bölümden oluşmaktadır:

- Ses İmzası Oluşturma
- Ses İmzası Modelleme
- Ses İmzası Sınıflandırma

Geliştirilen sistem wav ve mp3 formatlı ses kliplerini kabul etmekte; formattan bağımsız ses imzası çıkarabilmekte ve modelleyebilmektedir. Ayrıca yanlış alarm oranı düşük, doğru ve hızlı sınıflandırma yapabilmektedir.

### 1.1 Ses İmzasının Tarihçesi

Ses imzasının tarihçesi insanlardaki parmak izlerinin her insan için yegâne olmasının fark edilmesiyle başlamaktadır.

Belgelerin imzalanmasında yakın zamana kadar parmak izi kullanılmış ve bu parmak izleri insanları tanıtan imzalar olarak kabul görmüşlerdir. Bunun ilk örneklerini günümüzden 2000 yıl öncesinde, Çin'de görmek mümkündür. İmparatorların, anlaşmaları imzalarırken başparmaklarının ön yüzünü kullandıkları bilinmektedir. Bu da göstermektedir ki parmak izinin her insanda farklılık gösterdiği, o dönemin imparatorları veya onların yaverleri tarafından bilinmekteydi.

19. yüzyılın sonlarına gelindiğinde, Sir Francis Galton insan parmak izlerinin farklılığını kanıtlamıştır. Daha sonrasında Scotland Yard müfettişlerinde Sir Edward Henry insanları parmak izlerinden tanıyabilen bir sistem geliştirmiştir. İnsanların parmak ucu epitel dokusundaki çizgileri çıkaran bu sistem, günümüzde kullanılmakta olan sistemlerin temelini teşkil etmektedir. [9,11]

Günümüze gelindiğinde, teknolojinin ilerlemesiyle, çoğulortam verisinden imza tanımlama üzerine bilimsel ve endüstriyel araştırmalar hızlanmıştır. İlk çalışmalar 1990'ların ortalarına doğru yapılmıştır. Ses verisinin içeriğini, ses verisi üzerindeki sahne geçişlerini bulmaya yönelik çalışmalar[2-5,20] ses imzası çalışmalarına basamak teşkil etmiştir. Sayısal ses damgalamayla devam eden süreçte müzik parçalarının sahiplik haklarının korunması problemine sayısal ses damgalama çözüm olarak sunulmuş, ancak sayısal ses damgalama sistemlerindeki gürbüzlük-duyulmazlık ödünleşmesi, yani gürbüz damgalama sonrası eklenen damganın duyulabilecek düzeye gelmesi, yeni çözümlere yönelme ihtiyacını beraberinde getirmiştir.

Bu ihtiyaca ses imzası sistemleri cevap vermiştir. Ses imzası sistemleri doğrudan ses içeriğini çıkarıp özetlemeye yönelik sistemler olduğundan ses parçasına herhangi bir kod veya bilgi eklenmediğinden ses verisini bozmamaktadır. Ses imzasının tanınmasına yönelik başlatılan çalışmalar iki amaç taşımaktaydı: Ses verisini en iyi

ve etkin biçimde özetleyen ve yegâne olarak tanımlayabilen ses imzasının çıkarılması ve çıkarılan imzanın çok büyük bir veritabanında hızlı ve doğruluk oranı yüksek ama pozitif yanlış alarm oranı düşük olarak aranması.

O dönemden günümüze kadar olan süreçte ses bölütünün özetlenmesinde ve ses parçasını yegâne olarak tanımlayabilen veri dizisinin çıkarılmasında farklı yaklaşımlara rastlanmaktadır. Buna karşın hepsinin ortak amacı, depolanacak verinin boyutunun azaltılması, bir başka deyişle, boyutu yüksek olan ses vektörünün daha düşük boyutlu uzaya iz düşürülmesi, böylece imza karşılaştırma veya sınıflandırma işleminin süresinin azaltılmasıdır.

## 1.2 Ses İmzası Tanıma Sistemleri

Ses imzasını sistemleri temel olarak iki bölümden oluşmaktadır:

- Ses imzası çıkarımı
- Veri tabanında ses imzası arama ve eşleme

Günümüzde yapılan çalışmalar da bu doğrultudadır ve çıkarılan ses imzasının parametrelerinin geliştirilmesine[11-15,17,19] yönelik ya da veri tabanında ses imzası arama algoritmalarının hızlandırılmasına yöneliktir[18,21].

Ses imzasının taşınması ve sağlaması gereken belirli özellikler bulunmaktadır. Bu özellikleri şu şekilde sıralanabilir:

- **Gürbüzlük** Ses klipi ciddi hasarlar geçirdikten sonra bile tanınabiliyor mu? Bu hasarlara pek çok şey sebep olabilir. Mp3 sıkıştırma, kuantalama hataları, kanal gürültüsü, zamanda sıkıştırma v.b. bunlara örnektir. Ses imzasının bu tip ataklardan etkilenmemesi beklenir.
- **Güvenilirlik** Ses parçaları, veri tabanındaki ses klipleriyle doğru olarak eşleştirilebilmelidirler.
- **İmza Büyüklüğü** Ne kadar ses imzasının saklanacağı ve bunun hafızada ne kadar yer tutacağı bir ölçüsüdür. İmza büyüklüğü aynı zamanda veritabanında hızlı arama yapabilmenin de bir parametresidir. Olabildiğince küçük imzalar elde edilmeye çalışılır.

- **Öge Boyu (Granularite)** Ses klipinin tanınabilmesi için gerekli ses bölütünün gerekliliğinin bir ölçüsüdür. MPEG21 standardına göre öge boyu 5sn'den küçük veya ona eşit olmalıdır.
- **Veri Tabanında Arama Hızı** Çok büyük veri tabanlarında aramalar yapıldığından ses imzası sisteminin etkin ve hızlı arama algoritmasına sahip olması gerekmektedir. Aynı zamanda veritabanının kolay güncellenebilir olması gerekir. Ayrıca arama işleminin işlemsel karmaşıklığının düşük olması, doğru ve isabetli olması kadar önemlidir.[1]

Bu parametrelerin bazıları arasında doğrudan, bazıları arasında dolaylı ödünlüşmeler vardır. Örneğın *İmza Büyüklüğü* arttıkça *Gürbüzlük* artmaktadır çünkü *İmza Büyüklüğü*nün azalması ses parçasından elde edilen bilginin azalması anlamına gelmektedir. Ancak bu durumda *Veri Tabanında Arama Hızı* azalmaktadır. Yapılan bilimsel araştırmalar bu ödünlüşmenin optimizasyonu üzerine yoğunlaşmaktadır.

Yapılan çalışmalarda en büyük farklılık, ses imzasını oluşturacak olan özniteliklerin seçiminde görölmektedir. Cano ve arkadaşları [1], Mel Frekanslı Kepstral Katsayılarını (MFKK) kullanılırken, Eric Allamanche [10] İzgesel Düzlük Ölçüsünü (İDÖ) kullanmaktadır. İDÖ bir ses verisinin gürültüsel mi yoksa tonsal bir veri mi olduğunun bir ölçüsüdür. Kimura[18] her bir bandın enerjisini kendisine temsilci vektör olarak seçerken, Haitma ve Kalker[11] 33 adet bark ölçekli bandlar arasındaki enerji farkını ikili sistemde kodlayarak oluşturdukları kıyıcı kodları kullanmışlardır.

Sukittanon ve Atlas [16] ise İzgesel Kestirimlerin ve dolayısıyla hesaplanan izgesel özniteliklerin güvenilir olamayacağını öne sürmüş ve ses işaretinin zaman uzay değişimini tanımlayabilmek için Modülasyon Frekanslı Analizi'ni çözüm olarak sunmuşlardır. Öznitelik olarak 19 bark ölçekli band geçiren filtrenin modülasyon frekanslı kestirimlerini kullanmışlardır. Burges ve ekibi [12] ise olaya yeni bir boyut getirerek hali hazırda kullanılan özniteliklerin tamamının buluşsal olduğunu ve bu yüzden optimal olamayacağını iddia etmişlerdir. Çözüm olarak Bozulma Diskriminant Analizini (BDA) sunmaktadırlar. Bu analiz esnasında yeni olarak Yönlü Temel Bileşen Analizi (YTBA) yöntemini sunmuşlardır. YTBA işaret değişintisini en büyükleyen ve gürültü gücünü en küçükleyen yönü seçen bir izdüşürme yöntemidir.

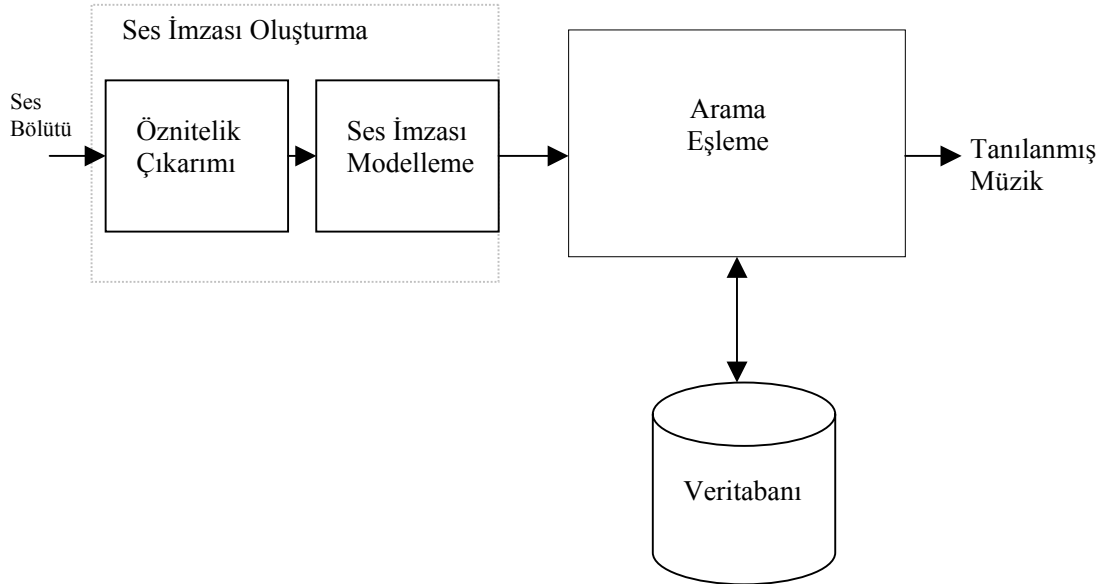
Arama Algoritmalarını hızlandırmaya yönelik çalışmalar[18,21] yapıldığı gibi arama yönteminin güvenilirliğini artıracak çalışmalar da yapılmıştır[10,17,19].

## 2. GELİŞTİRİLEN SİSTEMİN GENEL YAPISI VE SES İMZALARININ OLUŞTURULMASI

Tez kapsamında geliştirilen ses imzası tanıma sistemi temel olarak iki modülden oluşmaktadır.

- Ses İmzası Oluşturma
- Veritabanında Ses İmzası Arama ve Eşleme

Bu modüller ve alt modülleri şekil 2.1 'de görülmektedir.



**Şekil 2.1:** Ses imzası sistemleri genel yapısını gösterir şema.

Ses imzası oluşturma kısmı iki bloktan oluşmaktadır. Öznitelik çıkarımı bloğunda sisteme girilen ses bölütünden, ön işlemlerin ardından kullanılacak öznitelikler hesaplanır. Ardından hesaplanan öznitelikler modelleme işlemine tabi tutularak imza olarak kullanılacak veri dizisi elde edilmiş olur. Ses imzası modellemede, kısa ses bölütleri için uzun ses bölütü içerisinde çıkarılan öznitelikler, beklenen değer, değişim ve türev işlemleriyle özetlenmektedir. Burada ses imzasıyla ilgili, Ses İmzası Boyutu ve Öge Boyu parametreleri belirlenmektedir. Her ikisinde mümkün olduğu



kadar düşük olması beklense de bu noktada asıl belirleyici parametre Gürbüzlük olmaktadır.

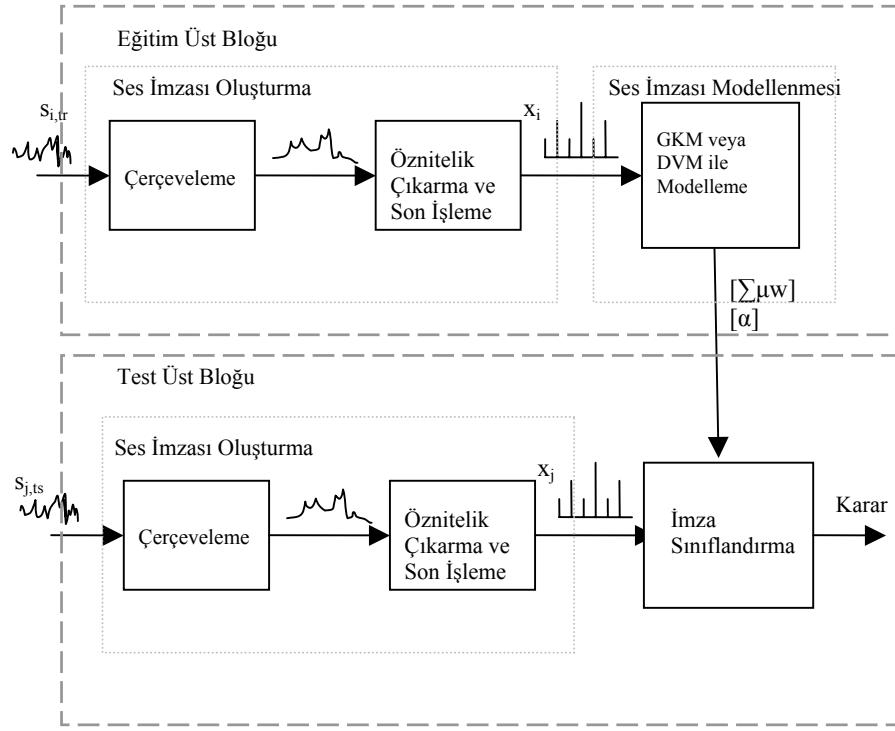
Modellenen ses imzası Veritabanı Ses İmzası Arama ve Eşleme modülüne gönderilmekte ve burada daha önceden veritabanına yüklenmiş ses imzalarıyla karşılaştırma yapılmaktadır. Karşılaştırmada birçok yöntem kullanılmakla birlikte ses imzası vektörleri arası uzaklıklara bakılması bütün sistemlerde ortak olarak yapılmaktadır. Bu noktada kullanılan metrikler ve yöntemler farklılık göstermektedir. Öklid[10] ve Hamming[11] en çok kullanılan uzaklık metrikleri olmakla birlikte kendi benzerlik ölçülerini kullanan sistemlerde vardır. Örneğin [18]'de histogram kesişimi benzerlik ölçüsü kullanılırken, [19]'da Üstel Sözde Norm ölçüsü kullanılmaktadır.

Tez kapsamında geliştirilen ses imzası sistemi, girişinden wav veya mp3 uzantılı dosyalar alabildiği gibi ses verisinin biçiminden bağımsız olarak gerçek zamanlı ses imzası oluşturma ve modelleme yapabilmektedir. Aynı zamanda süratli ve isabetli sınıflandırma yapabilme yeteneğine de sahiptir. Yapılan testler sonucunda sistemin müzik parçası ayrımının dışında o parçanın ait olduğu müzik türünü de yüksek doğruluk oranıyla tanımlayabildiği gözlenmiştir. Şekil 2.2 'de geliştirilen ses imzası sisteminin blok şeması görülmektedir.

Sistem girişine gelen ses bölütü, yükseltilmiş bir kosinüs olan Hamming Penceresi kullanılarak yumuşatıldıktan sonra uygun bir dönüşüm uygulanarak başka uzaylara taşınır, burada gerekli öznitelikler hesaplanır ve belli çerçeveler boyunca özetlenerek Ses İmzası oluşturulur. Bu imzalar oluşturulduktan sonra modellenmek üzere Ses İmzası Modelleme modülüne gönderilir. Tez kapsamında modelleme için Gauss Karışım Modeli (GKM) [23] kullanılmaktadır. Ancak sonuçların karşılaştırılması açısından Destek Vektör Makinesi(DVM) [22] ile de modelleme yapılarak imza sınıflandırılması çalışmaları yapılmıştır.

Şekil 2.2 'den görülebileceği gibi bu sistemde veri tabanı yerine eğitim için sisteme verilen ses bölütlerinden elde edilen parametreler kullanılmaktadır. Ses imzasını veritabanında aramak yerine sınıflandırma işlemi kullanılmaktadır. Böylece geniş bir veritabanında çok hızlı arama yapmak mümkün olabilmektedir. Ayrıca sınıflandırma işlemi Gauss Karışım Modeli Sınıflandırıcısı kullanıldığından veritabanının güncellenmesi, yani yeni sınıfların eklenmesi veya çıkarılması da kolaylaşmaktadır.

Şekildeki  $s_{i,tr}$  sistemin eğitim üst bloğu girişine, sistemin eğitimi için verilen  $i$ . 2 saniye süreli ses bölümünü gösterirken  $s_{j,ts}$  sistemin test üst bloğu girişine verilen  $j$ . 2 saniye süreli sınıfı kararlaştırılacak ses bölümünü temsil etmektedir.  $[\sum\mu w]$ , GKM eğitim parametrelerini gösterirken,  $[\alpha]$  DVM 'nin hesapladığı destek vektörleri göstermektedir.



Şekil 2.2: Geliştirilen ses imzası sisteminin blok şeması

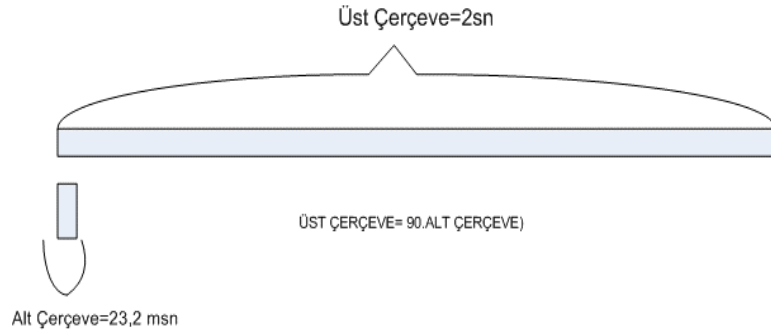
## 2.1 Ses İmzası Oluşturma

Ses İmzasının oluşturulması üç aşamadan oluşmaktadır. İlk olarak sistem girişine gelen wav veya mp3 formatındaki stereo ses verisi bir ön işleme ile 44100 Hz 'lik örnekleme frekansına sahip mono ses işaretine dönüştürüldükten sonra, Hamming veya Kare Pencere kullanılarak alt çerçevelere ayrılmaktadır. Ses verisinin içeriğini çıkarmaya yönelik öznitelikler alt çerçevelerde hesaplanmaktadır. Ardından öz niteliklerin, belirlenmiş olan “Öge Boyu” uzunluğundaki ses işaretleri boyunca, beklenen değeri ve değişintisi hesaplanmakta ve ilgilenilen ses klib için “ses imzası” elde edilmektedir. Son adımda oluşturulan ses imzasının normal dağılımından uzak, ayrıcalıklı değerleri ve öznitelik tanımına göre anlamsız olan değerler son işlemeye tabi tutularak eğitim setinden çıkarılmaktadır. Bu üç aşamanın da işlemsel karmaşıklığı düşük olup ses dosyalarının uzunluğunun 1/9 unda

tamamlanabilmektedir. Böylece geniş veritabanlarının oluşturulması veya sınıflandırmaya uygun hale getirilmesi için çok az bir zaman harcanmaktadır.

### 2.1.1 Kullanılan pencere ve çerçeveler

Ses verisinin durağan olmayan rasgele bir süreç olması nedeniyle, özellikle frekans uzayında yapılacak işlemler genellikle ses verisinin durağan olarak kabul edildiği en çok 40ms'lik alt çerçevelerde yapılmaktadır. Tez kapsamında geliştirilen sistemde 2 çerçeve kullanılmaktadır. Birincisi özniteliklerin hesaplandığı 23,2 ms uzunluğunda ve 44100 Hz 'lik veride 1024 ses örneğine karşı düşen alt çerçevedir. Örnek sayısının 1024 seçilmesinde MPEG Katman 3 uyumluluğu ve hızlı Fourier dönüşümü yapabilmek için 2'nin üsteli olması gerekliliği göz önünde tutulmuştur. İkinci çerçeve, alt çerçevelerde hesaplanan özniteliklerin beklenen değer ve değişintisinin alınarak özetlendiği üst çerçevedir. Üst çerçeve 90 alt çerçeveden oluşmaktadır ve yaklaşık 2sn uzunluğundadır. Alt çerçevede hesaplanan öznitelikler, üst çerçeve boyunca beklenen değer ve değişinti kullanılarak özetlenmektedir. Şekil 2.3 'te kullanılan çerçevelerin şematik gösterimi görülmektedir.



**Şekil 2.3:** Alt ve üst çerçevenin şematik gösterimi

Bunlara ek olarak, frekans uzayında hesaplanacak öznitelikler için sistem girişindeki 1024 örneklilik ses bölütüne Hamming pencereleme uygulanmaktadır. Yükseltilmiş kosinüs olan Hamming penceresi yoluyla, ses verisi yumuşatılarak, alt çerçevedeki ses verisinin ilk ve son bölümündeki örnekler birbirine yaklaştırılmakta, böylece bölüt periyodik olarak düşünüldüğünde işaretin başlangıcı ile sonu arasındaki geçiş yumuşatılarak Fourier dönüşümü sırasında oluşacak anlamsız yüksek frekans bileşenlerinden kurtulmak amaçlanmaktadır. Zaman uzayında hesaplanan öznitelikler için ise birim genlikli Pencere uygulanmaktadır.

### 2.1.2 Öznitelik çıkarma

Ses imzası, ses bölütüne ait içerik veya karakteristiği taşımalıdır. Bu yüzden tez kapsamında geliştirilen ses imzası siteminde ses içeriğini çıkarmaya yönelik öznitelikler kullanılmıştır. Ses içeriğini elde etmek için daha önce otomatik ses bölütleme ve sınıflandırmada kullanılan özniteliklerden yararlanılmıştır[2-4]. Bu öznitelikler şunlardır:

- Sıfır Geçiş Sayısı (SGS)
- Kök Ortalama Karesel Enerji (KOKE)
- Kepstral Akı (KA)
- İzgesel Akı (SFlux)
- İzgesel Kitle Merkezi (İKM)
- İzgesel Azalma Noktası (İAN)
- Mel Frekansı Kepstral Katsayıları (12 adet) (MFKK)

Öznitelikler her bir alt çerçevede hesaplandıktan sonra üst çerçeve boyunca beklenen değeri ve değişintisi hesaplanmaktadır. Öznitelik olarak hesaplanan beklenen değer ve değişinti kullanılmaktadır. Böylece 92160 boyutlu ses verisi 36 boyutlu uzaya iz düşürülmüş olmaktadır, yani 92160 örnekli 2 sn 'lik ses bölütünden 36 boyutlu Ses İmzası oluşturulmaktadır. Bu aşamadan sonra hesaplanan öznitelikler son işlemeye tabi tutulduktan sonra modellenmek üzere, GKM ve DVM sınıflandırıcılarına gönderilmektedir.

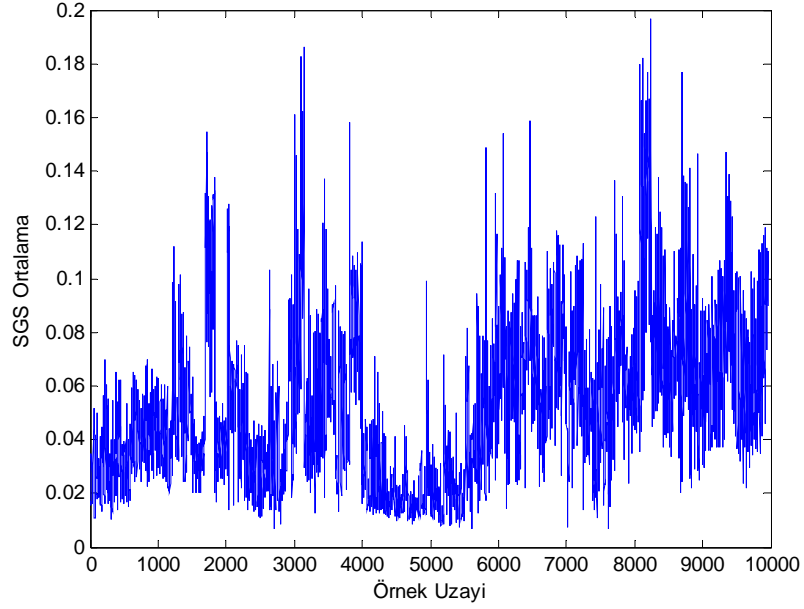
#### 2.1.2.1 Sıfır geçiş sayısı (SGS)

Zaman uzayı sıfır geçiş sayısıdır. Ses işareti güç izgesinin denge noktasını belirlemektedir. Uzun zamanlı ortalamaları farklı müzik türlerinde farklı bölgelerde toplanmaktadır. Aynı zamanda, aynı türden kliplerde de farklı değerler alabilmektedir. Şekil 2.4 ve 2.5 'te bu özneliğin beklenen değer ve değişintilerinin 67 parça ve 7 farklı müzik türü boyunca değişimi gözlenmektedir. (2.1) ve (2.2) 'deki bağıntılar kullanılarak hesaplanmaktadır[2]:

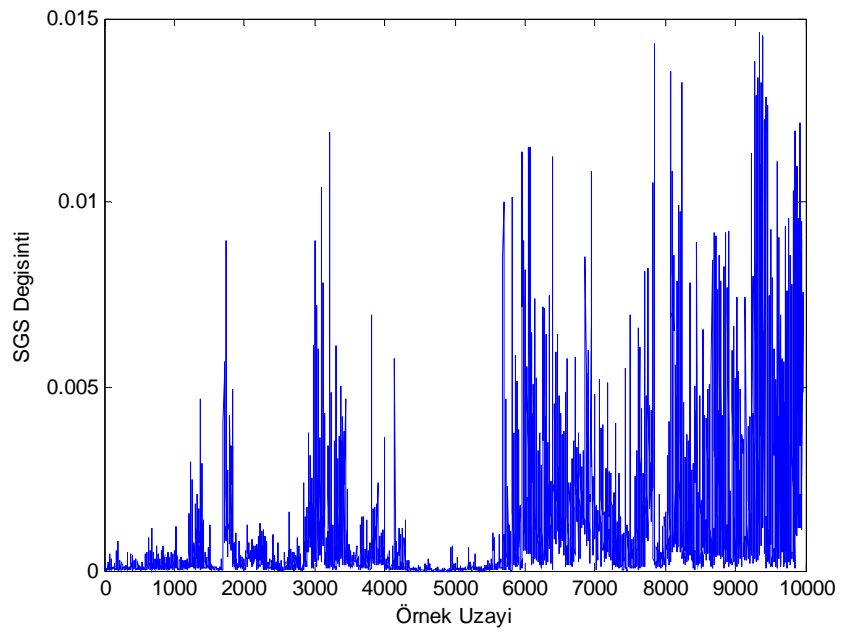
$$SGS = \frac{1}{2} \sum_k |\text{sgn}[s_i(k)] - \text{sgn}[s_i(k-1)]|. \quad (2.1)$$

Buradaki  $\text{sgn}(x) = \begin{cases} 1, & x \geq 0 \\ -1, & x < 0 \end{cases}$  (2.2)

Formüldeki  $s_i$ ,  $i$ . 1024 örnekli ses çerçevesini gösterirken,  $k$  ise örnek indisine karşı düşmektedir.



**Şekil 2.4:** 67 parçadan çıkarılan ortalama SGS öznitelikleri.



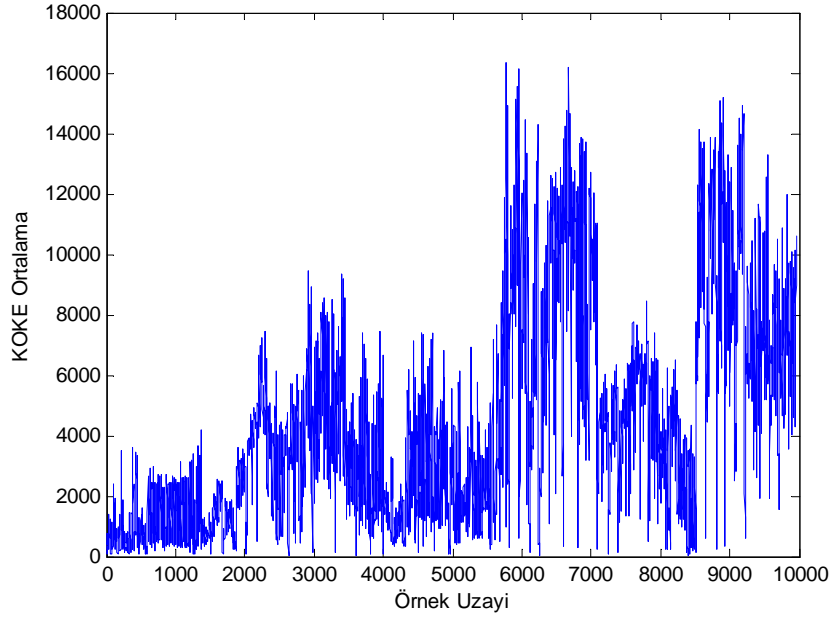
**2.5:** 67 parçadan çıkarılan değişimi SGS öznitelikleri.

### 2.1.2.2 Kök ortalama karesel enerji (KOKE)

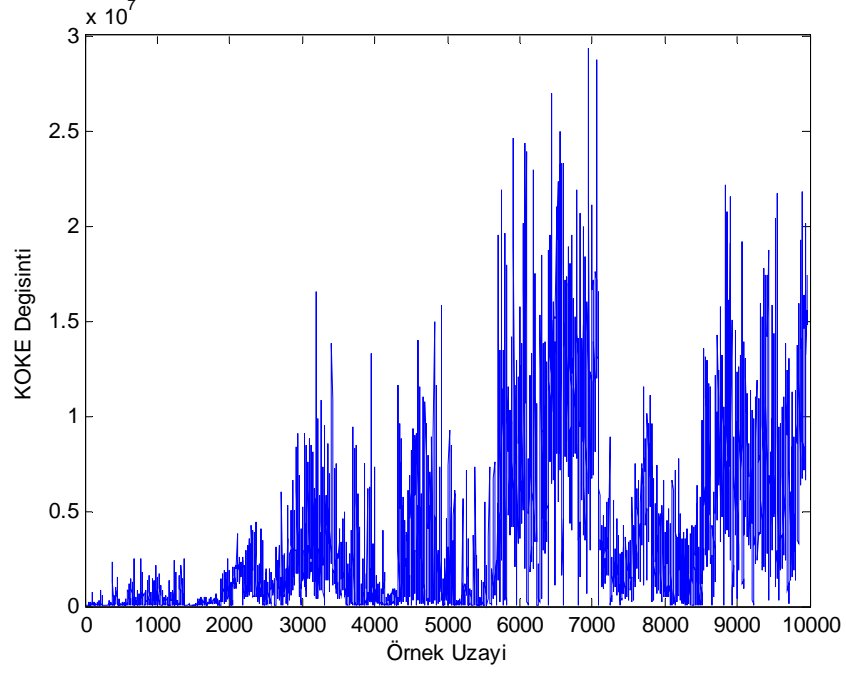
Bir çerçevenin ses seviyesinin bir ölçüsüdür. Bölütlemeye ses deęişimleri anahtar yerlerde olduğundan bu öznelik oldukça önemlidir. Çoğunlukla, farklı müzik türleri de farklı ortalamalar etrafında kümelenebilir. Şekil 2.6 ve 2.7 'de bu özneliğin beklenen deęer ve deęişimlerinin 67 parça ve 7 farklı müzik türü boyunca deęişimi gözlenmektedir. (2.3) 'teki formüle göre hesaplanmaktadır[3]:

$$KOKE = \sqrt{\frac{1}{N} \sum_k s_i(k)^2} \quad (2.3)$$

ki buradaki N, bir çerçevedeki örnek sayısı olup, 1024'tür.



Şekil 2.6: 67 parçadan çıkarılan ortalama KOKE öznelikleri.



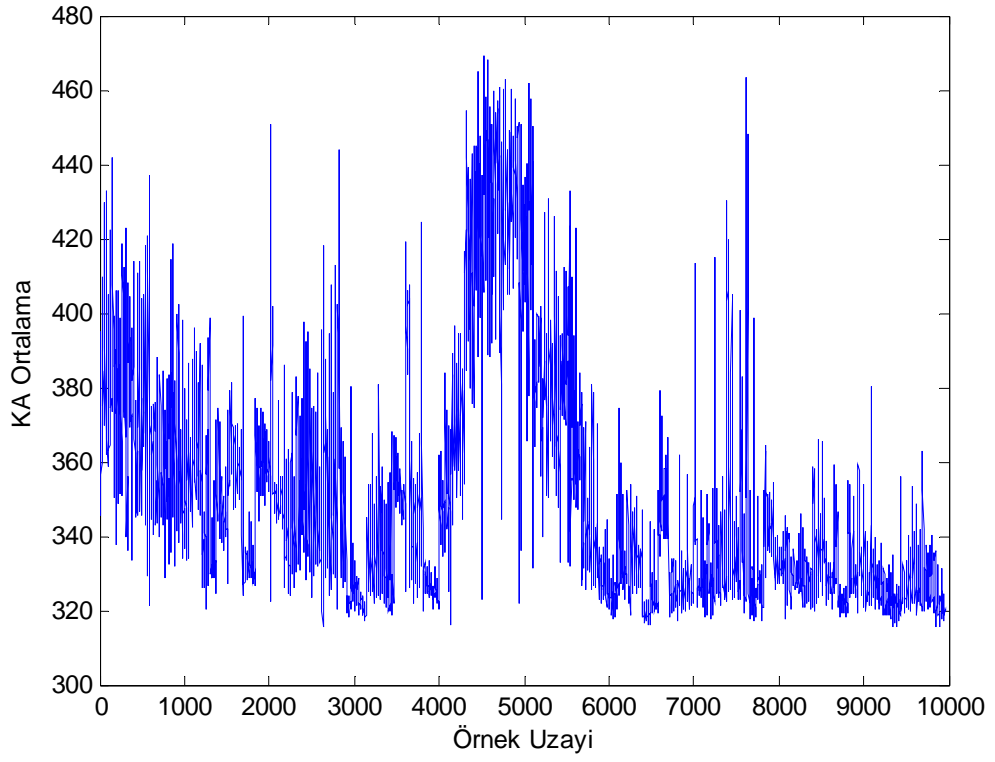
Şekil 2.7: 67 parçadan çıkarılan deęişinti KOKE öznitelikleri.

### 2.1.2.3 Kepstral akı (KA)

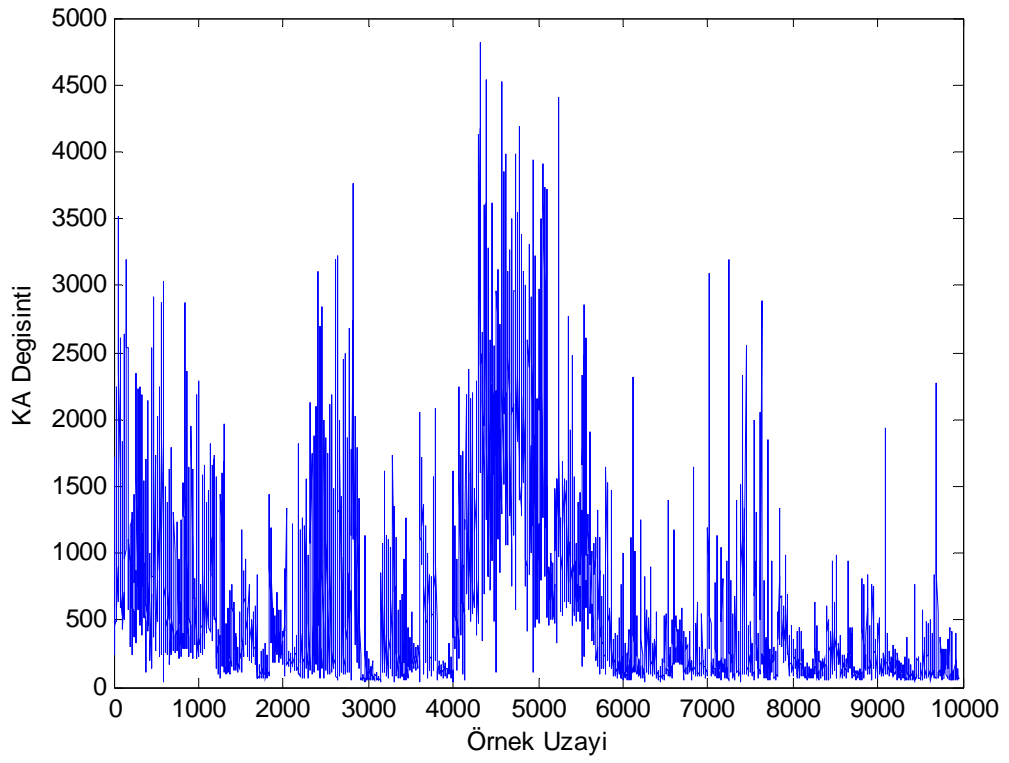
Tanım olarak, erişilen iki komşu çerçevenin kepstrum uzayı farklarının normudur. Sesin enerjisindeki deęişimin bir göstergesi olup bölütlemeye sık kullanılan bir özniteliktir. Uzun dönemli ortalaması ve deęişintisi farklı klipler için farklılık göstermektedir. Şekil 2.8 ve 2.9 ‘da bu özniteliğin beklenen deęer ve deęişintilerinin 67 parça ve 7 farklı müzik türü boyunca deęişimi gözlenmektedir. (2.4) ’teki formüle göre hesaplanmaktadır[4]:

$$KA = \sqrt{\sum_{k=1}^N [C_i(k) - C_{i+1}(k)]^2} = \|C_i - C_{i+1}\| \quad (2.4)$$

Burada,  $C_i$  i numaralı alt çerçevenin kepstrum uzayı katsayılarını göstermektedir.



Şekil 2.8: 67 parçadan çıkarılan ortalama KA öznelikleri.



Şekil 2.9: 67 parçadan çıkarılan değişinti KA öznelikleri.

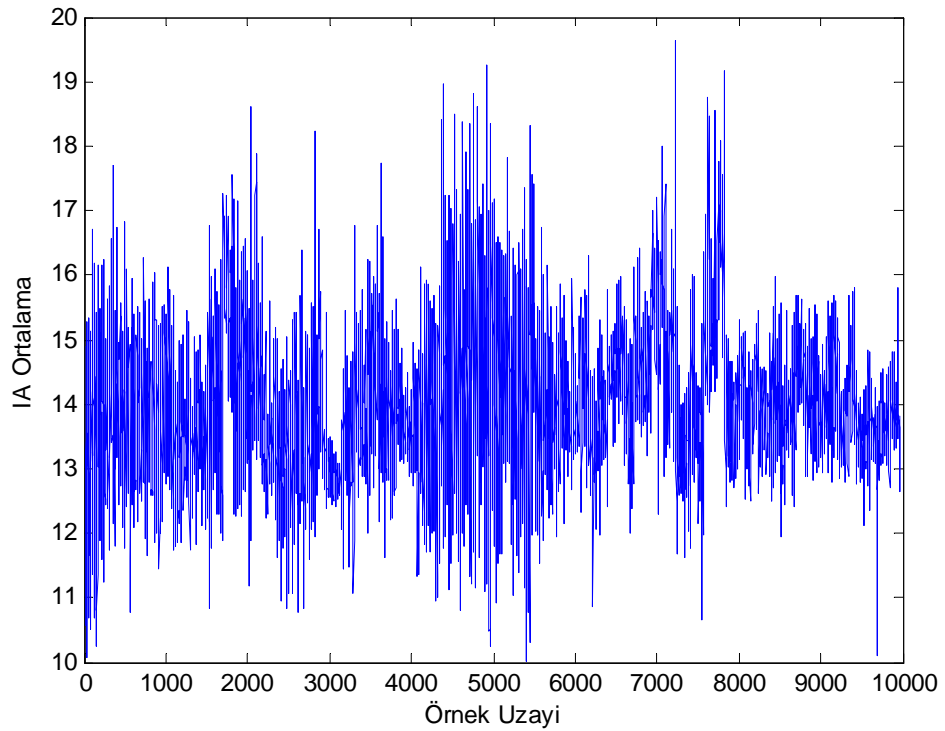


#### 2.1.2.4 İzgesel akı (İA)

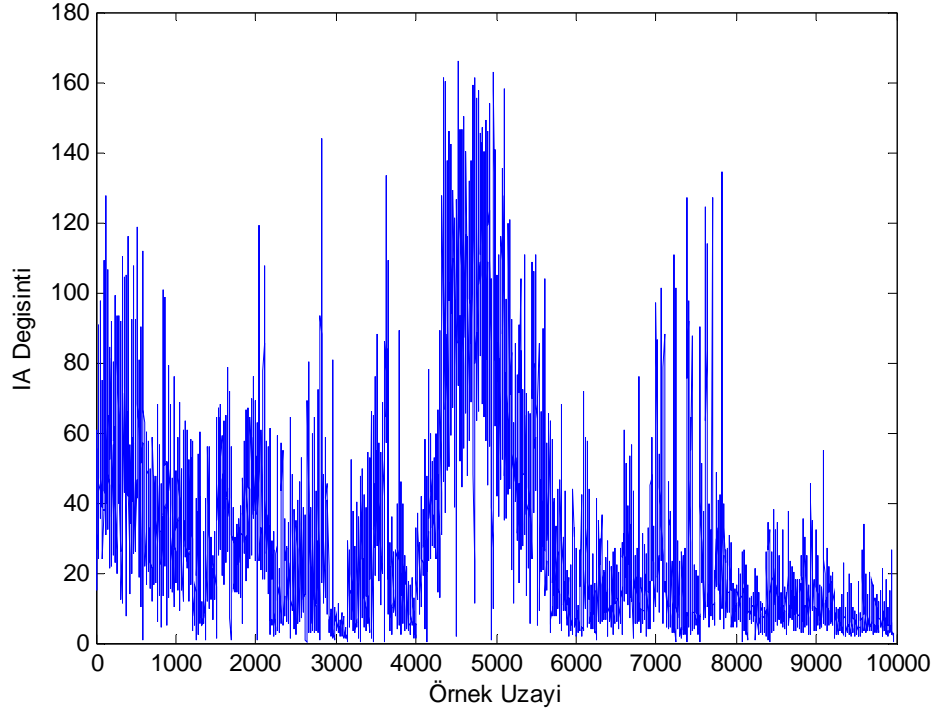
Erişilen iki komşu çerçeve arasındaki izgesel farkın normudur. İzgenin çerçeveler arasındaki değişiminin bir ölçüsü olup, bölütleme için uygun bir öznitelik olmakla beraber, uzun süreli ortalaması ve değişintisi kliplerin ve müzik türlerinin ayırımında yardımcı olmaktadır. Şekil 2.10 ve 2.11 'de bu özneliğin beklenen değer ve değişimlerinin 67 parça ve 7 farklı müzik türü boyunca değişimi gözlenmektedir. (2.5) 'teki formüle göre hesaplanmaktadır[3,5]:

$$IA = \sqrt{\sum_k [S_i(k) - S_{i+1}(k)]^2} = \|S_i - S_{i+1}\| \quad (2.5)$$

Burada,  $S_i$  i numaralı çerçevenin dB ölçekli izgesidir.



Şekil 2.10: 67 parçadan çıkarılan ortalama İA öznitelikleri.



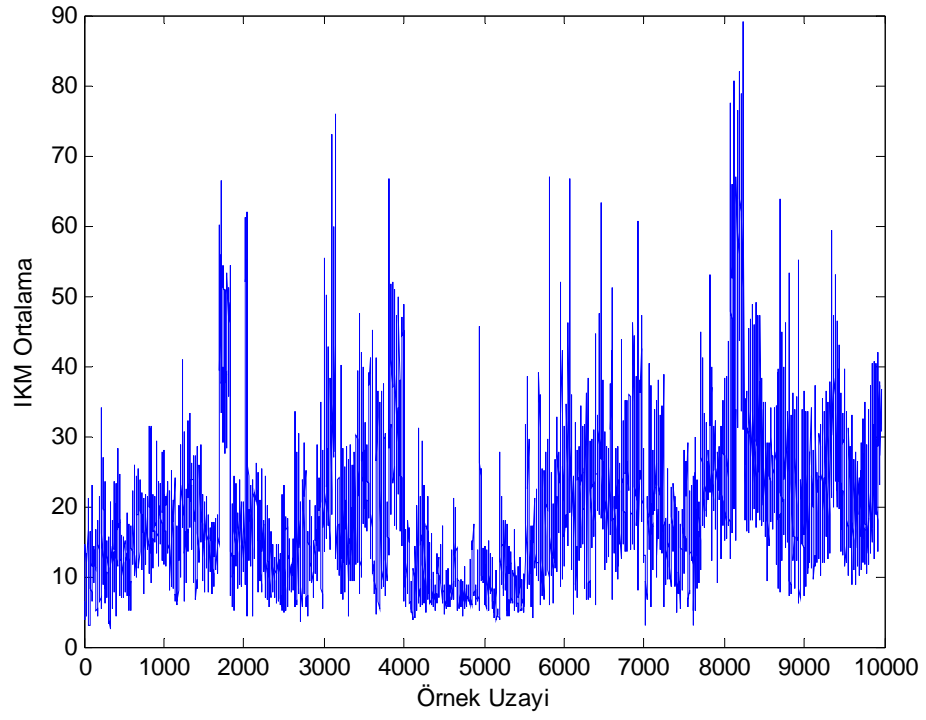
Şekil 2.11: 67 parçadan çıkarılan deęişinti İA öznitelikleri.

### 2.1.2.5 İzgesel kitle merkezi (İKM)

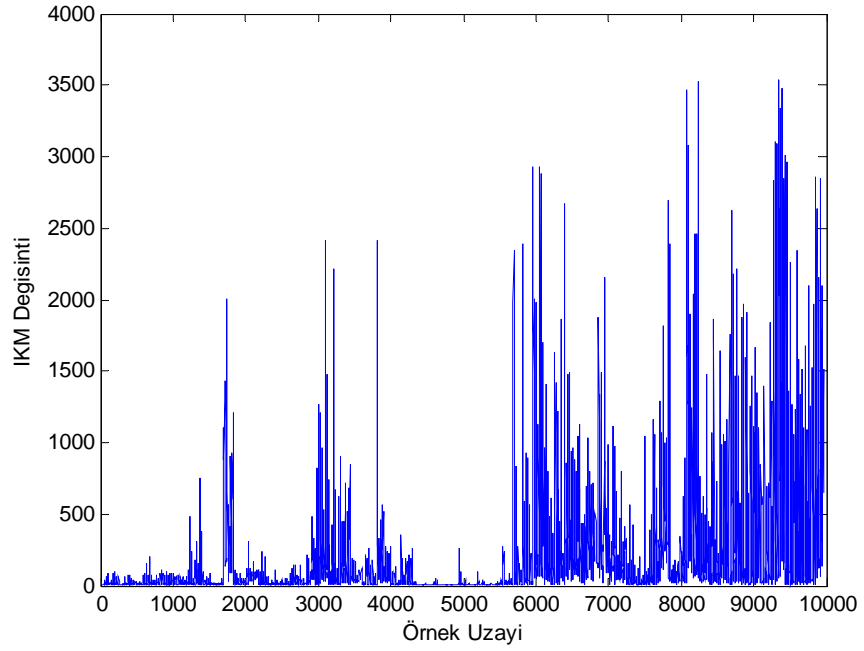
İzgenin denge noktasıdır. Sıfır Geçiş Sayısı ile dorudan ilişkilidir. Şekil 2.12 ve 2.13 'te bu özneliğın beklenen deęer ve deęişintilerinin 67 parça ve 7 farklı müzik türü boyunca deęişimi gözlenmektedir. (2.6) 'daki formüle göre hesaplanmaktadır [3]:

$$İKM = \frac{\sum_k kS_i(k)}{\sum_k S_i(k)} \quad (2.6)$$

Burada  $S_i$ , i numaralı alt çerçevenin dB ölçekli izgesidir.



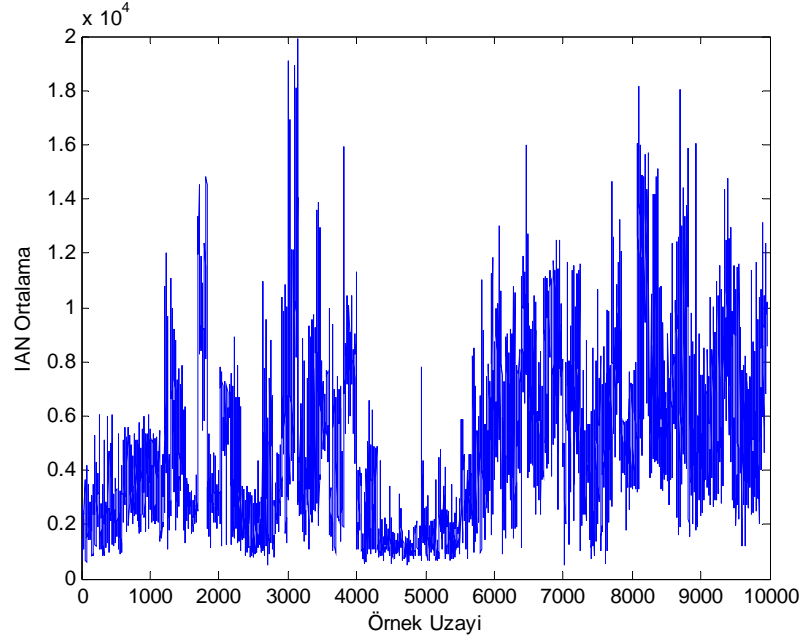
Şekil 2.12: 67 parçadan çıkarılan ortalama İKM öznelikleri.



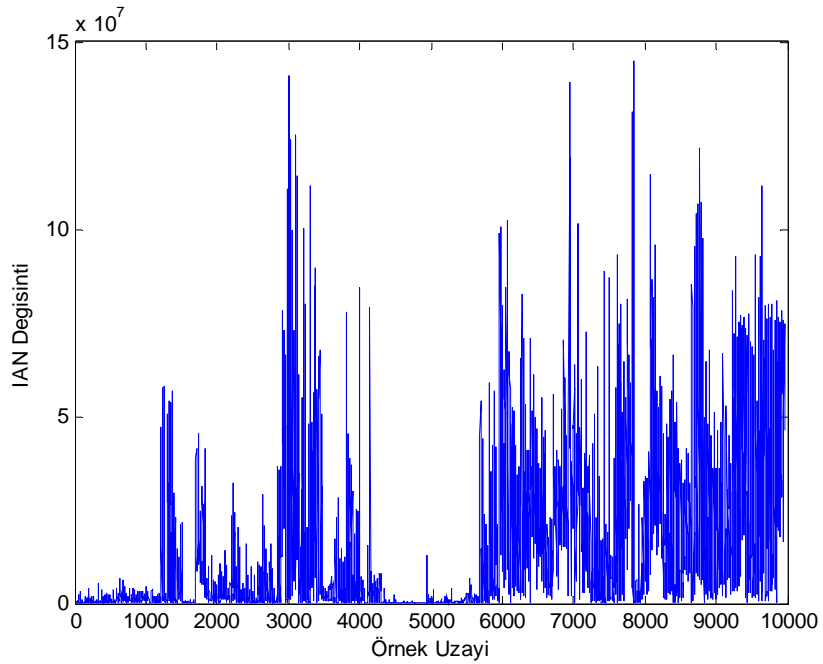
Şekil 2.13: 67 parçadan çıkarılan ortalama İKM öznelikleri.

### 2.1.2.6 İzgesel azalma noktası (İAN)

İzgesel güç dağılımının %95 'lik kısmına denk gelen normalize frekans değeridir. İzgenin dağılımının eğikliğinin bir ifadesidir. İzgesel merkez ve sıfır geçiş sayısı ile doğrudan ilişkilidir[3]. Şekil 2.14 ve 2.15 'te bu özneliğin beklenen değer ve değişimlerinin 67 parça ve 7 farklı müzik türü boyunca değişimi gözlenmektedir.



Şekil 2.14: 67 parçadan çıkarılan ortalama İAN öznelikleri.



Şekil 2.15: 67 parçadan çıkarılan değişimi İAN öznelikleri.

### **2.1.2.7 Mel-Frekansı keprum katsayıları (MFKK)**

Diğer özniteliklerden farklı olarak bu öznitelik 12 boyutludur[5]. Daha çok konuşma ve konuşmacı tanıma uygulamalarında kullanılan bu katsayıların, uzun dönemli ortalamalarının farklı müzik klipleri için farklı değerler etrafında kümелendiği gözlenmiştir. Bu özniteliğe ait ortalama ve değişinti çizimleri Ek-A 'da sunulmakta olup, klipler ve müzik türleri arasındaki geçişler belirgindir.

### **2.1.3 Son işleme**

Öznitelik Çıkarımı işlemi bittikten sonra çıkarılan öznitelikler son işlemeye tabi tutulur. Bu aşamada, özniteliğin tanımı dışına çıkan veya örneklerin yoğunlaştığı bölgenin çok dışına çıkan örnekler elenirler. Bu tip örneklerin oluşma sebebi, müzik klipi içerisinde yer alan sıfır bölgelerdir. Yani ses işaretinin bulunmadığı bölgelerdir. Örneğin bu tip durumlarda İzgesel Akı ve Kepstral Akı değerleri, ortalamalarının çok üzerine çıkmaktadır. İzgesel azalma noktası ise örnekleme frekansının dışında değerler vermektedir. Böylece bu tip bölgeleri ayırt etmek çok kolay hale gelmektedir ve bu örnekler veri setinden çıkarılmaktadırlar.

### 3. SES İMZASI MODELLEME VE SINIFLANDIRMA

Tez kapsamında geliştirilen sistem, 67 parça ve 7 ayrı müzik türünü sınıflandırmak için tasarlanmıştır. Bu amaçla modellemede güçlü sınıflandırıcıların kullanılması planlanmıştır. Ses imzası modelleme için önce Destek Vektör Makinesi kullanılmıştır. DVM 'nin seçiminde güncel çalışmalarda yüksek performans sergilemesi ve bu yüzden popüler olması etkin rol oynamıştır. Ancak tez kapsamında yapılan çalışmalar ses imzalarının dağılımının Gauss Karışım Modeli (GKM) ile daha iyi modellenebileceğini göstermiştir. Her iki modelleme yöntemi de performans testlerinde kullanılmış ve ses imzası modelleme ve sınıflandırmadaki başarımları karşılaştırılmıştır.

#### 3.1 Gauss Karışım Modeli (GKM)

Gauss Karışım Modelinin eğitiminde tüm sınıflar birbirinden bağımsız kabul edilir. O yüzden her bir sınıf için ayrı eğitim yapılır.

Gauss karışım modeli, eğitim veri setini, birbirinden bağımsız birden fazla Gauss dağılımının karışımıyla tanımlayan bir modeldir. Yani eğitim için verilen örneklerin, Gauss dağılımı ile örnek üreten birden fazla bağımsız kaynaktan üretildiği varsayıp, bu kaynaklara ait Gauss parametrelerinin, karışımın olasılık yoğunluk işlevini ençoklayacak şekilde optimizasyonudur. Böylece, veri setinin tek bir dağılımdan üretildiğinin varsayıp, gerekli dağılım parametrelerinin kestirimiyle modelleme yapan sistemlerin yetersiz kaldığı durumlarda bile başarılı sonuçlar verebilmektedir. Bir Gauss dağılımının olasılık yoğunluk işlevi (3.1) 'de tanımlanmıştır:

$$N(\mathbf{x} | \boldsymbol{\mu}, \boldsymbol{\Sigma}) = \frac{\exp\left(-\frac{1}{2}(\mathbf{x} - \boldsymbol{\mu})^t \boldsymbol{\Sigma}^{-1}(\mathbf{x} - \boldsymbol{\mu})\right)}{\sqrt{(2\pi)^d |\boldsymbol{\Sigma}|}} \quad (3.1)$$

$\boldsymbol{\mu}$  ortalama vektörünü gösterirken,  $\boldsymbol{\Sigma}$  ortak deęişinti matrisidir. Birbirinden bağımsız  $K$  adet Gauss kaynağından oluşan Gauss Karışım Modelinin olasılık yoğunluk işlevi ise (3.2) 'de tanımlanmıştır.

$$p(\mathbf{x}|\boldsymbol{\theta}) = \sum_{i=1}^K w_i \cdot \mathcal{N}(\mathbf{x}|\boldsymbol{\mu}_i, \boldsymbol{\Sigma}_i) \quad (3.2)$$

Buradaki  $w_i$  parametreleri Gauss Karışım Modelindeki Gauss kaynaklara ait ağırlıklardır. Bunlar da eğitim sırasında optimize edilecek parametreler arasındadır. Bu katsayılar (3.3) ve (3.4) 'daki kısıtları sağlamak zorundadırlar:

$$\sum_i w_i = 1 \quad (3.3)$$

ve

$$0 \leq w_i \leq 1, \forall i \quad (3.4)$$

Eğitim sırasında, her sınıf için, o sınıfa ait eğitim kümesi örnek vektörleri alınıp, bu örnekler kullanılarak Gauss karışımlarının parametreleri  $\boldsymbol{\theta} = [\boldsymbol{\mu}_1, \boldsymbol{\Sigma}_1, w_1, \boldsymbol{\mu}_2, \boldsymbol{\Sigma}_2, w_2, \dots, \boldsymbol{\mu}_K, \boldsymbol{\Sigma}_K, w_K]$  öğrenilir. Parametrelerin öğrenilmesinde genelde analitik bir yöntem olan En büyük Olabilirlik (Maximum Likelihood) kestirimi kullanılır. Ancak bu yöntem analitik çözümü olmayan durumlarda yetersiz kalmaktadır. Bunun yerine iteratif bir yöntem olan Beklenti Enbüyükleme (Expectation Maximization) yöntemi yaygın olarak kullanılmaktadır. Bu şekilde boyut sayısı artsa bile hızlı ve güvenilir parametre kestirimi yapılabilmektedir[23]. Tez kapsamında eğitim kümesinden ses imzasının modellenmesinde BE algoritması kullanılmaktadır.

### 3.1.1 Beklenti enbüyükleme (BE) algoritması

BE algoritması, bazı elemanları eksik veri setleri için olasılık dağılım parametrelerinin hesaplanmasında kullanılan iteratif bir algoritmadır. BE iteratif özelliğinden dolayı aynı zamanda analitik bir çözüme ulaşamayan uygulamalarda da kullanılabilir.

BE algoritmasıyla,  $p(\mathbf{X}, \mathbf{Y}|\boldsymbol{\theta})$  ve parametre vektörü  $\boldsymbol{\theta}$  tanımlı iken,  $p(\mathbf{X}, \mathbf{Y}|\boldsymbol{\theta})$  olasılığını  $\boldsymbol{\theta}$ 'ya göre en büyükleme amaçlanmaktadır. Burada  $\mathbf{X} = \{\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_T\}$  birbirinden bağımsız örnek vektörler topluluğunu gösterirken,  $\mathbf{Y}$  veri setindeki eksik örnek vektörlerini göstermektedir. Bu durumu geliştirilen sistem için yorumladığımızda

amaç  $\theta=[\mu,\Sigma,w]$  parametre vektörüne göre sistem girişine verilen  $\mathbf{X}$  vektörü için  $p(\mathbf{X},\mathbf{Y}|\theta)$  olasılığını enbüyüklemektir [8].

BE algoritmasının gerçekleşmesinde izlenen yol,  $p(\mathbf{X},\mathbf{Y}|\theta)$  işlevinin en büyüklemesini basitleştirecek bir işlevin belirlenmesi, yani sınıflandırma için sistem girişine verilecek bir örneğin, o sınıftan olabilme olasılığını enbüyükleyecek bir işlevin belirlenmesidir.

Bu nedenle algoritmanın her aşamasında :

**Beklenti (B)- Adımı:** Belirlenen işlevin, o anki parametreler ve örnekler kullanılarak kestirimi

**Enbüyükleme (E)- Adımı:** Verinin ve Gizli Değişkenin dağılımını en büyükleyecek şekilde parametrelerin güncellenmesi gerçekleşir.

Yukarıdaki varsayımlar altında BE algoritmasının Beklenti adımında (3.5) 'deki  $Q$  beklenti işlevini oluşturulmaktadır.

$$Q(\theta, \theta^s) = E_{\mathbf{Y}} \left[ \ln p(\mathbf{X}, \mathbf{Y} | \theta) | \mathbf{X}, \theta^s \right] \quad (3.5)$$

Buradaki  $\theta^s$  dağılım parametreleri için yapılan bir önceki kestirimleri,  $\theta$  ise tüm dağılımı tanımlayan yeni kestirim için tanımlı parametre vektörünü temsil etmektedir.  $Q$  işlevi verinin,  $\theta^s$  'nin o anki kestirimine bağlı olarak bilinmeyen  $\mathbf{Y}$  vektörüne bağlı olabirliğinin beklenen değerini hesaplamakta kullanılmaktadır. BE Enbüyükleme adımında ise  $Q$  işlevi  $s$  adımında öngörülen  $\theta^s$  ya göre en büyüklenir ve  $\arg \max_{\theta} Q(\theta, \theta^s) \rightarrow \theta^{s+1}$  ataması gerçekleştirilir.

BE algoritması bir  $\theta^0$  başlangıç değeriyle başlatılır ve log olasılığın her iterasyonda artacağı kesindir[23]. Ancak, iterasyonu başlatacak bu ilk değer, algoritmanın kestiriminin başarılı olması açısından çok önemlidir. İyi bir başlangıç değeri iyi bir kestirimi sağlarken, tersi bir durum kestirim başarısız olmasına yol açabilmektedir. Çözüm olarak iki yöntem kullanılmaktadır; birinci yöntem başlangıç noktalarını rasgele farklı bölgelerden seçmek iken, diğeri ise öbekleme algoritmalarından yararlanmaktadır [23].



### 3.1.2 Gauss karışım modeli için BE algoritması

BE algoritmasını GKM için yorumlarsak, X eksik veri bulunan veri setini gösterirken, Y ise her  $\mathbf{x}_i$  i numaralı örnek vektörünün hangi kaynak tarafından üretildiği bilgisidir.

Bu noktada yeni bir işlev tanıtalım:

$$z_{i,j} = \begin{cases} 1 & \mathbf{x}_i, j \text{ nolu Gauss kaynağı tarafından üretilmişse} \\ 0 & \text{değilse} \end{cases} \quad (3.6)$$

z işlevini de kullanarak, her bir örneğin ve kaynağın birbirinden bağımsız ve düzgün dağıldığı varsayımı altında  $\boldsymbol{\theta}$  'ya göre birleşik olasılık dağılım işlevi (3.7) ve (3.8) 'de gösterildiği gibi olmaktadır.

$$p(\mathbf{X}, \mathbf{Y} | \boldsymbol{\theta}) = \prod_{i=1}^T \prod_{j=1}^K \mathbf{w}_j^{z_{i,j}} p(\mathbf{x}_i | j, \boldsymbol{\theta})^{z_{i,j}} \quad (3.7)$$

$$\Rightarrow \ln p(\mathbf{X}, \mathbf{Y} | \boldsymbol{\theta}) = \sum_{i=1}^T \sum_{j=1}^K z_{i,j} \ln(\mathbf{w}_j p(\mathbf{x}_i | j, \boldsymbol{\theta})) \quad (3.8)$$

Eşitlikteki T eğitim kümesindeki örnek sayısını belirtirken K karışımdaki Gauss kaynağı sayısını göstermektedir. j ise kaynak numarasını simgelemektedir. Bu aşamadan sonra B adımına geçilir. B adımında log olasılık yoğunluk işlevinin koşullu beklenen değeri, yani  $Q(\boldsymbol{\theta}, \boldsymbol{\theta}^s)$  işlevi hesaplanır. Çarpımları doğal logaritma ile toplamlara çevirerek ve Beklenti işlevinin toplama üzerine dağılımını da kullanarak  $Q(\boldsymbol{\theta}, \boldsymbol{\theta}^s)$  işlevi 3.9 'daki halini alır.

$$\begin{aligned} Q(\boldsymbol{\theta}, \boldsymbol{\theta}^s) &= E[\ln p(\mathbf{X}, \mathbf{Y} | \boldsymbol{\theta}) | \mathbf{X}, \boldsymbol{\theta}^s] \\ &= \sum_{i=1}^T \sum_{j=1}^K E[z_{i,j} | \mathbf{X}, \boldsymbol{\theta}^s] \ln \mathbf{w}_j + E[z_{i,j} | \mathbf{X}, \boldsymbol{\theta}^s] \ln p(\mathbf{x}_i | \boldsymbol{\theta}) \end{aligned} \quad (3.9)$$

[8] de gösterilmektedir ki  $Q(\boldsymbol{\theta}, \boldsymbol{\theta}^s)$  işlevini  $\boldsymbol{\theta}$  'ya göre en büyükmek  $\mathbf{X}$  vektör dizisinin  $p(\mathbf{X} | \boldsymbol{\theta}^{s+1})$  olasılığını artırmakla beraber,  $Q$  'nun en büyük olması, olasılığın da en büyük olmasına karşı düşmektedir[8]. Buna göre E adımında s iterasyon numarasını göstermek üzere  $P(j | \mathbf{x}_i, \boldsymbol{\theta}^s)$  hesaplanırken, M adımında  $Q$  'yu en büyükmek  $\boldsymbol{\theta}$  parametreleri bulunmaktadır. Bunun için (3.10) ifadesi kullanılmaktadır.

$$\begin{aligned}
\frac{\partial Q}{\partial \boldsymbol{\theta}} &= 0 \\
\frac{\partial Q}{\partial \boldsymbol{\mu}_j} &= 0 \rightarrow \hat{\boldsymbol{\mu}}_j \\
\frac{\partial Q}{\partial \boldsymbol{\Sigma}_j} &= 0 \rightarrow \hat{\boldsymbol{\Sigma}}_j \\
\frac{\partial Q}{\partial \mathbf{w}_j} &= 0 \rightarrow \hat{\mathbf{w}}_j
\end{aligned} \tag{3.10}$$

Buradan yola çıkarak,  $\boldsymbol{\theta}$  parametreleri  $\boldsymbol{\mu}$ ,  $\boldsymbol{\Sigma}$  ve  $\mathbf{w}$  (3.11),(3.12) ve (3.13)'deki bağıntılar kullanılarak iteratif bir şekilde hesaplanırlar[8].

$$\hat{\boldsymbol{\mu}}_j = \frac{\sum_{i=1}^T \mathbf{x}_i P(j|\mathbf{x}_i, \boldsymbol{\theta}^s) \cdot \frac{p(\mathbf{x}_i | j, \boldsymbol{\theta})}{p(\mathbf{x}_i | \boldsymbol{\theta})}}{\sum_{i=1}^T P(j|\mathbf{x}_i, \boldsymbol{\theta}^s) \cdot \frac{p(\mathbf{x}_i | j, \boldsymbol{\theta})}{p(\mathbf{x}_i | \boldsymbol{\theta})}} \tag{3.11}$$

$$\hat{\boldsymbol{\Sigma}}_j = \frac{\sum_{i=1}^T (\mathbf{x}_i - \hat{\boldsymbol{\mu}}_j)^2 \cdot P(j|\mathbf{x}_i, \boldsymbol{\theta}^s) \cdot \frac{p(\mathbf{x}_i | j, \boldsymbol{\theta})}{p(\mathbf{x}_i | \boldsymbol{\theta})}}{\sum_{i=1}^T P(j|\mathbf{x}_i, \boldsymbol{\theta}^s) \cdot \frac{p(\mathbf{x}_i | j, \boldsymbol{\theta})}{p(\mathbf{x}_i | \boldsymbol{\theta})}} \tag{3.12}$$

$$\hat{\mathbf{w}}_j = \frac{\sum_{i=1}^T \mathbf{w}_j \cdot P(j|\mathbf{x}_i, \boldsymbol{\theta}^s) \cdot \frac{p(\mathbf{x}_i | j, \boldsymbol{\theta})}{p(\mathbf{x}_i | \boldsymbol{\theta})}}{\sum_{k=1}^K \sum_{i=1}^T \mathbf{w}_k P(k|\mathbf{x}_i, \boldsymbol{\theta}^s) \cdot \frac{p(\mathbf{x}_i | k, \boldsymbol{\theta})}{p(\mathbf{x}_i | \boldsymbol{\theta})}} \tag{3.13}$$

BE ilk koşullara hassas olarak bağımlı iteratif bir yöntemdir. Bu yüzden başlangıç koşulları iyi verilmelidir. Bu hesaplama için tez kapsamında bir öbekleme yöntemi olan k-ortalamar algoritmasından yararlanılmaktadır. Bunun seçiminde k-ortalamar algoritmasının isabetli ve hızlı sonuç vermesi etken olmuştur[7,8,23].

### 3.1.3 GKM parametrelerinin analizi

Tüm veri setinin her bir özneliğın kendi içerisinde birim deęişintiye sahip olacak şekilde ölçeklenmesi sonrası elde edilen veri ile sistemin 7 sınıflı tür ayrımı için eğitimi yapılmıştır. Yer sorunu nedeniyle sadece ilk sınıfa ait ilk ortalama vektörü ve kovaryans matrisi EK B 'de sunulmuştur. Sağlıklı öznelikler için kovaryans matrisinin köşegeninde bulunan deęer, bulunduğu sütunun en büyük deęeri olmalıdır. Eğitim sonrası ortaya çıkan kovaryans matrisine bakıldığında birkaç öznelik hariç, özneliklerin çoğunlukla iyi seçildięi görülebilmektedir. Bu konuda en kötü örnek 2 nolu öznelik olan KOKE özneliğidir. Görüldüğü üzere birçok öznelikle ilintilidir. Bunun dışında, beklenildiğı üzere SGS ile İKM ve İAN arasında da ilinti bulunmaktadır. Çünkü butüç öznelik de İzgenin ağırlık merkeziyle ilgilidir. En başarılı öznelikler ise KA, İA ve MFKK'dır.

### 3.2 Destek Vektör Makinesi (DVM)

Destek vektör Makinelerinin yüz sınıflandırma, nesne tanıma, konuşmacı tanıma gibi bir çok uygulamada uygulanması ve başarılı sonuçlar vermesi geliştirilen ses imzası tanıma sisteminde, çalışmalara DVM ile başlanmasının sebebi olmuştur. Bu kısımda DVM 'nin arkasındaki teori anlatılacaktır.

#### 3.2.1 Doğrusal ayrılabilir durum

$\mathbf{x}_i \in R^d$   $d$  boyutlu ses imzası vektörünü göstermek üzere ve  $y_i \in \{-1, +1\}$  eğitim kümesi örneklerinin ikili sınıflandırma için sınıf etiketleri olmak üzere, eğitim kümesi  $\mathbf{A}_{tr} = \{(\mathbf{x}_1, y_1), (\mathbf{x}_2, y_2), \dots, (\mathbf{x}_T, y_T)\}$  bilinmektedir. Bu eğitim kümesi kullanılarak  $\mathbf{a} \cdot \mathbf{x} + b = 0$  ayırıcı düzlemi bu eğitim kümesini aşağıdaki koşulu sağladığı takdirde ayırabilmektedir:

$$\begin{aligned} \mathbf{a} \cdot \mathbf{x}_i + b &> 0 \quad \text{eger} \quad y_i = 1 \\ \mathbf{a} \cdot \mathbf{x}_i + b &< 0 \quad \text{eger} \quad y_i = -1 \end{aligned} \quad (3.14)$$

$\mathbf{a}$  ve  $b$  parametreleri üzerinde ufak bir modifikasyon ile 3.15 'teki ifade elde edilir.

$$y_i (\mathbf{a} \cdot \mathbf{x}_i + b) \geq 1, \forall i \quad (3.15)$$

Optimal ayırıcı hiper-düzlemi bulabilmek için hiper-düzlem ile ona en yakın örnek arasındaki uzaklığın ençoklanmasına ihtiyaç duyulmaktadır.  $D(\mathbf{a},b)$  uzaklık işlevi olmak üzere en yakın örneğin optimal hiper-düzleme olan uzaklığı 3.16 'da belirtilmiştir.

$$D(\mathbf{a},b) = \min_{\{\mathbf{x}_i|y_i=1\}} \frac{\mathbf{a} \cdot \mathbf{x}_i + b}{|\mathbf{a}|} - \max_{\{\mathbf{x}_i|y_i=-1\}} \frac{\mathbf{a} \cdot \mathbf{x}_i + b}{|\mathbf{a}|} \quad (3.16)$$

2.20'den görüleceği üzere uygun min ve max noktalar +1 ve -1 'dir. Bu yüzden en büyüklememiz gereken işlev, aşağıda sunulmaktadır.

$$D(\mathbf{a},b) = \frac{1}{|\mathbf{a}|} - \frac{-1}{|\mathbf{a}|} = \frac{2}{|\mathbf{a}|} \quad (3.17)$$

Bu yüzden problemimiz  $O(\mathbf{a}) = \frac{1}{2}|\mathbf{a}|^2$  işlevini 3.15 'teki kısıtlara göre en küçükleme problemine dönüşmektedir. Optimum DVM sınıflandırıcısı, Denklem (3.18) ile verilen Lagrange maliyet fonksiyonu çözülerek bulunmaktadır:

$$L(\mathbf{a},b,\alpha) = \frac{1}{2}\|\mathbf{a}\|^2 + \sum_{i=1}^T \alpha_i (\mathbf{y}_i [(\mathbf{a} \cdot \mathbf{x}_i) + b]) \quad (3.18)$$

Buradaki Langrange işlevini ençoklayacak  $\mathbf{a},b$  ve  $\alpha$  parametreleri, L işlevinin bu parametrelere göre türevlerinin sıfıra eşitlenmesiyle oluşan denklemin analitik çözümüyle bulunmaktadır. Buradan  $\alpha_i \geq 0$  ve  $\sum_i \alpha_i \mathbf{y}_i = \mathbf{0}$  olmak üzere,

$$W(\alpha) = \sum_i \alpha_i - \frac{1}{2} \sum_{i,j} \alpha_i \alpha_j \mathbf{y}_i \mathbf{y}_j \mathbf{x}_i \mathbf{x}_j \quad (3.19)$$

Lagrange'in Wolf eş işlevi elde edilmektedir. Amaç bu işlevi enküçükleme. Her eğitim vektörü için bir  $\alpha_i$  Lagrange çarpanı bulunmaktadır. DVM eğitimi sonucunda,  $\alpha_i > 0$  'a karşı gelen eğitim vektörleri "destek vektörleri" olarak adlandırılmaktadır. Neticede, geri kalan eğitim vektörleri için  $\alpha_i = 0$  olmaktadır.

### 3.2.2 Doğrusal ayrılabilir durum

Bir önceki bölümde anlatılan problem ve bulunan çözümler eğitim kümesinin doğrusal bir hiper-düzlem tarafından ayrılabilir olduğu durum için anlatılı ve ayrılabilir durumlar için geçersizdir. Bu tip durumlar için 3.14 'teki denklemlere, yanlış sınıflandırmalar için bir ceza parametresi getirilir.

$$\begin{aligned} \mathbf{a} \cdot \mathbf{x}_i + b &\geq 1 - \xi_i & \text{eger } \mathbf{y}_i &= 1 \\ \mathbf{a} \cdot \mathbf{x}_i + b &\leq \xi_i - 1 & \text{eger } \mathbf{y}_i &= -1 \\ \xi_i &\geq 0, \forall i \end{aligned} \quad (3.20)$$

Buna göre her yanlış sınıflandırılan eğitim örneği için bir  $\xi_i > 1$  parametresi oluşacaktır. Bu da hata sayısının  $\sum_i \xi_i$  küçük olması anlamına gelmektedir. Bu yüzden yanlış sınıflandırılan eğitim kümesi elemanları için parametre setine bir de ceza parametresi eklenir. Bu durumda enazlanacak işlev,

$$O(\mathbf{a}) = \frac{1}{2} \|\mathbf{a}\|^2 + C \left( \sum_i \xi_i \right) \quad (3.21)$$

olmaktadır. Bu durumda optimum DVM sınıflandırıcısı, Denklem (3.22) ile verilen Lagrange maliyet fonksiyonu çözülerek bulunmaktadır.

$$L(\mathbf{a}, b, \alpha, \xi, \mu) = \frac{1}{2} \|\mathbf{a}\|^2 + C \sum_{i=1}^T \xi_i - \sum_{i=1}^T \alpha_i (\mathbf{y}_i [\mathbf{a} \cdot \mathbf{x}_i + b] - 1 + \xi_i) - \sum_{i=1}^T \mu_i \xi_i \quad (3.22)$$

Denklem (2.27)' de  $\alpha_i, \mu_i \geq 0$  ( $i=1, \dots, T$ )' ler Denklem (3.20) ile verilen koşulları sağlamak için tanımlanan Lagrange çarpanlarıdır.

Optimum DVM sınıflandırıcısı,  $L(\mathbf{a}, b, \alpha, \xi, \mu)$  fonksiyonunun  $\mathbf{a}, b, \xi, \alpha, \mu$ ' ye göre türevlerinin sıfır olması ve  $\alpha_i, \mu_i \geq 0$  koşulları ile  $L(\mathbf{a}, b, \alpha, \xi, \mu)$ ' nin enbüyüklenmesiyle bulunmaktadır. Buradan  $C \geq \alpha_i \geq 0$  ve  $\sum_i \alpha_i \mathbf{y}_i = \mathbf{0}$  olmak üzere,

$$W(\alpha) = \sum_i \alpha_i - \frac{1}{2} \sum_{i,j} \alpha_i \alpha_j \mathbf{y}_i \mathbf{y}_j \mathbf{x}_i \mathbf{x}_j \quad (3.23)$$

Lagrange'in Wolf eş işlevi elde edilmektedir. Bu problemin çözümünden her eğitim vektörü için bir  $\alpha_i$  Lagrange çarpanı bulunmaktadır.

### 3.2.3 Doğrusal olmayan durum

Gerçekleştirilen ses imzası tanıma sistemleri gibi bir çok sistemde, eğitimde kullanılan örnekler lineer bir hiper-düzlem tarafından ayıramayacak veya doğrusal hiper-düzlemlerle sınıflandırma başarımının düşeceği cinsten örneklerdir. Bu yüzden doğrusal olmayan sınıflandırıcılara ihtiyaç duyulmaktadır. Her ne kadar DVM için yukarıda anlatılan özellikler doğrusal sistemler için geçerli de olsa, DVM örnekleri daha üst uzaylara dönüşümünü yaparak bu sorunun da üstesinden gelebilmektedir. Yani  $\Phi : R^d \mapsto H$  dönüştürme işlevi kullanılarak,  $d$  boyutlu  $R$  uzayındaki örnekler, daha yüksek boyutlu  $H$  uzayına izdüşürülmektedir. Bu noktadan sonra  $H$  uzayında ayırıcı hiper-düzlem aranmaktadır. Bu işlem  $R$  uzayında doğrusal olmayan hiper-düzlem arayışına karşı düşmektedir.

$H$  uzayında ayırıcı bir hiper-düzlem arayışında, eğitim için kullanılacak veri 3.19'deki gibi hep çarpımlar halinde görünecektir. Bu yüzden daha yüksek uzayda sadece  $\Phi(\mathbf{x}_i) \cdot \Phi(\mathbf{x}_j)$  biçimindeki çarpımlarla ilgilenilmektedir. Uzayın boyut sayısı yükseldikçe, bu işlemin işlevsel karmaşıklığı ve hesaplanma süresi çok artmaktadır. Bu problem çözüm olarak,  $K(\mathbf{x}_i, \mathbf{x}_j) = \Phi(\mathbf{x}_i) \cdot \Phi(\mathbf{x}_j)$  olmak üzere  $K()$  Çekirdek İşlevleri sunulmuştur. Her örnek çarpımının olduğu yerde çekirdek işlevi kullanılır ve haritalama işlevi  $\Phi$ 'nin bilinmesine gerek kalmadan çarpım işlemlerinin işlemsel karışıklığından kurtulmuş olunur.

Tez çalışması kapsamında, ses imzalarının doğrusal olmayan doğasından dolayı, Gauss Radyal Baz Fonksiyonu (RBF) çekirdeği kullanılarak doğrusal olmayan bir DVM sınıflandırıcı tasarlanmaktadır. Gauss RBF çekirdeği,  $\gamma$ , evrensel baz fonksiyonu genişliği olmak üzere,

$$K(\mathbf{x}_i, \mathbf{x}_j) = \exp\left(-\gamma \|\mathbf{x}_i - \mathbf{x}_j\|^2\right), \gamma > 0 \quad (3.24)$$

ile tanımlanmaktadır.

Denklem (2.24) ile verilen Wolfe eş fonksiyonunun  $a_i$  ve  $\mu_i$  'ye göre türevlerinin sıfırlanması sonucu optimum DVM sınıflandırıcısı elde edilmektedir. DVM eğitimi sonucunda elde edilen optimum hiper-düzlemin parametreleri:

$$\mathbf{a} = \sum_{i=1}^T \alpha_i y_i \mathbf{x}_i \quad (3.25)$$

$$b = -\frac{1}{2} \sum_{v \in SV} \alpha_v y_v \left[ K(\mathbf{x}_v, \mathbf{x}_{c_1}) + K(\mathbf{x}_v, \mathbf{x}_{c_2}) \right] \quad (3.26)$$

Denklem (2.24)' teki  $y_v \in \{-1, +1\}$ ,  $\alpha_v > 0$  Lagrange çarpanına karşı gelen  $\mathbf{x}_v \in SV$ , destek vektörünün sınıf etiketidir.  $\mathbf{x}_{c_1}$  ve  $\mathbf{x}_{c_2}$  ise, sırasıyla, -1 ve 1 ile etiketlenen sınıflara ait destek vektörleridir.

### 3.3 Ses İmzası Sınıflandırma

Sınıflandırıcıların eğitimi sonrası modellenen ses imzaları sınıflandırma bölümüne aktarılmaktadır. Bu noktadan sonra sistem girişine tanımlanmak için verilen ses bölütlerinden öznitelikler hesaplandıktan sonra 36 boyutlu ses imzaları sınıflandırıcı girişine aktarılır. Sınıflandırıcı, test için girilen ses bölütünün hangi parçaya veya türe ait olduğunu belirler.

GKM 'de eğitimde hesaplanmış parametrelerle o imza vektörünün her bir sınıf içinde olma olasılığı;

$$P(\mathbf{x}_i | C_j), i \in [1, T], j \in [1, M] \quad (3.27)$$

hesaplanmaktadır. Buradaki M sınıf sayısını göstermekte olup, müzik türü sınıflandırma için 7, müzik parça sınıflandırması için 67 'dir. Ses bölütünün etiketi en büyük olasılığı sağlayan sınıfa atanmaktadır.

T, test kümesindeki örnek sayısını göstermektedir.

DVM 'de ise iki çeşit sınıflandırma mevcuttur. Sınıflar arası 2'li sınıflandırmalar ve 1'i diğerlerine karşı sınıflandırmalar. Her iki yöntemde de aslında yapılan ikili sınıflandırmadır. Bu sınıflandırmalarda bir sınıf +1 ile etiketlenirken diğeri -1 ile

etiketlenir ve eğitimi de bu şekilde yapılır. Bu sınıf kümesindeki tüm ikili sınıflar için yapılır ve her bir sınıflandırma işlemi için ayrı parametreler tutulur. Sınıflandırma sırasında ise 2.33 'teki karar işlevinin,  $F(z)$ , işaretine bakılır ve o sınıflandırma için karar verilir. Bütün ikili sınıflandırmalar içerisinde en çok sınıflandırmayı kazanan sınıf, imzanın eşleştirildiği sınıf olarak seçilmektedir[22].

$$F(z) = \text{sgn} \left[ \sum_{i=1}^m y_i a_i K(\mathbf{x}_i, \mathbf{z}) + b \right] \quad (3.28)$$

Bu formülde  $\mathbf{z}$ , sınıflandırıcıya verilen ve sınıfı belirlenmek istenen 36 boyutlu ses imzasını,  $b$  eğitim sırasında hesaplanan hiperdüzlem parametresini,  $m$  destek vektör sayısını,  $K()$  ise RBF çekirdek fonksiyonunu göstermektedir.



#### 4. PERFORMANS TESTLERİ

Performans testleri sistemin tür ayrımı ve parça ayrımındaki başarımını ölçmeye yönelik yapılmıştır. Testlerde, karşılaştırılmak amacıyla, hem GKM için hem de DVM için sonuçlar elde edilmiştir. Bu testler için toplamda 5 saat 44 dakikalık bir müzik veritabanı oluşturulmuştur ve testlerde kullanılmıştır. Kullanılan veri tabanı 7 ayrı müzik türüne ait parçalardan oluşmaktadır. Her müzik türüne ait en az 2 şarkıcı ve her şarkıcıya ait en az 3 parça bulunmaktadır. Tüm veritabanı toplamda 67 şarkıdan oluşmaktadır. Bu da 2 sn 'lik bölütler göz önünde bulundurulduğunda toplamda 9990 örneğe karşı düşmektedir. Testlerde, eğitim kümesi tüm veri içerisinden düzgün dağılımlı olarak seçilmiş ve test kümesi olarak veri setinin tamamı kullanılmıştır. Testlerde ek olarak kullanılan eğitim setinin tüm veri setine oranının performansa olan etkisi de incelenmiş, grafiksel ve tablolar halinde sunulmuştur.

##### 4.1 Tür Ayrımı Testleri

Tür ayrımı testlerinde tez kapsamında geliştirilen imza tanıma sisteminin otomatik müzik türü ayırt etme performansı incelenmiştir. Bu testler sırasında kullanılan veri seti, türler ve sınıfları tablo 4.1 'de gösterilmiştir.

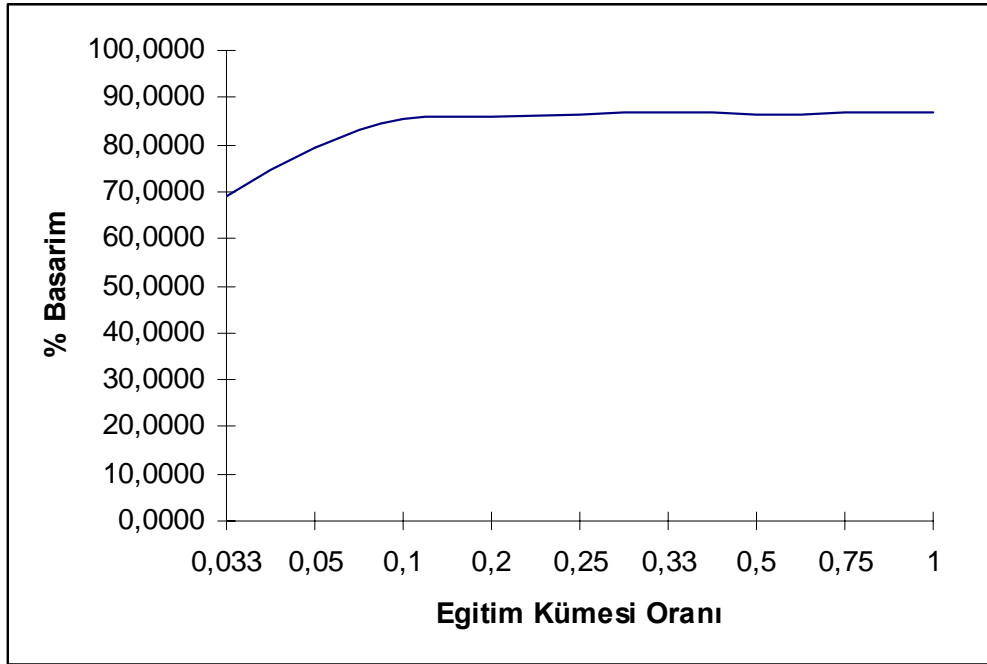
**Tablo 4.1:** Tür ayrımı testlerinde kullanılan veri seti, türler ve sınıfları

Tür	Sınıf No	Örnek Sayısı
Klasik	1	1418
Sufi	2	1481
Free Jazz	3	1537
Yabancı Etnik	4	1480
Yabancı Pop	5	1457
Yerli Etnik	6	1507
Yerli Pop	7	1480

Bu veriler kullanılarak ilk önce GKM ile sonuçlara ulaşılmıştır. Yapılan tür tespiti çalışmalarında alınan performans sonuçları tablo 4.2 'de gösterilmiştir. Şekil 4.1 'de ise bu tablonun grafiksel gösterimi yapılmaktadır.

**Tablo 4.2:** Müzik türlerinin GKM ile sınıflandırma sonuçları

Eğitim Kümesi Oranı	% Başarım	Eğitim Kümesi Oranı	% Başarım
0,033	69,0090	0,33	86,6767
0,05	79,2192	0,5	86,4765
0,1	85,3353	0,75	86,6366
0,2	86,0460	1	86,6567
0,25	86,5966		



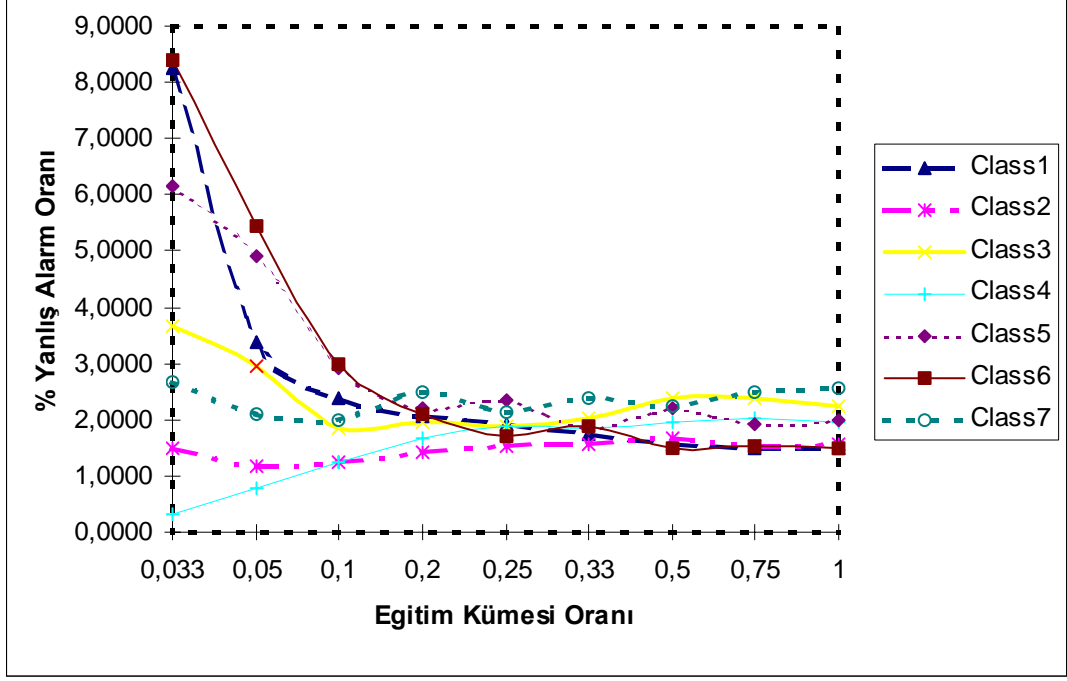
**Şekil 4.1:** Müzik türlerinin GKM ile sınıflandırma sonuçları

Yukarıdaki sınıflandırma işlemleri yüksek bir performansla başarıldığı gibi her bir müzik türüne ait pozitif yanlış alarm oranları da düşüktür. Tablo 4.3 'te her bir tür için, yapılan her sınıflandırmaya ait pozitif yanlış alarm oranları sunulmaktadır. Bunun grafiksel gösterimini ise şekil 4.2 'de görmek mümkündür.

**Tablo 4.3:** GKM ile yapılan sınıflandırma sonucu oluşan pozitif yanlış alarm oranları

Eğitim Kümesi Oranı	1	2	3	4	5	6	7	Toplam
0,033	8,2482	1,5115	3,6737	0,3303	6,1461	8,3984	2,6827	30,9909
0,05	3,3634	1,1712	2,9500	0,7908	4,9249	5,4454	2,1121	20,7578
0,1	2,3824	1,2613	1,8519	1,2312	2,9329	3,0030	2,0020	14,6647
0,2	2,0621	1,4314	1,9720	1,6817	2,1922	2,1121	2,5025	13,9540
0,25	1,9119	1,5215	1,8919	1,9019	2,3323	1,7017	2,1421	13,4033

0,33	1,7317	1,5716	2,0320	1,8418	1,8619	1,9019	2,3824	13,3233
0,5	1,5716	1,6717	2,3724	1,9620	2,2022	1,5115	2,2322	13,5236
0,75	1,5115	1,5315	2,3924	2,0120	1,9119	1,5215	2,4825	13,3633
1	1,5015	1,5616	2,2523	1,9620	2,0020	1,4915	2,5726	13,3435

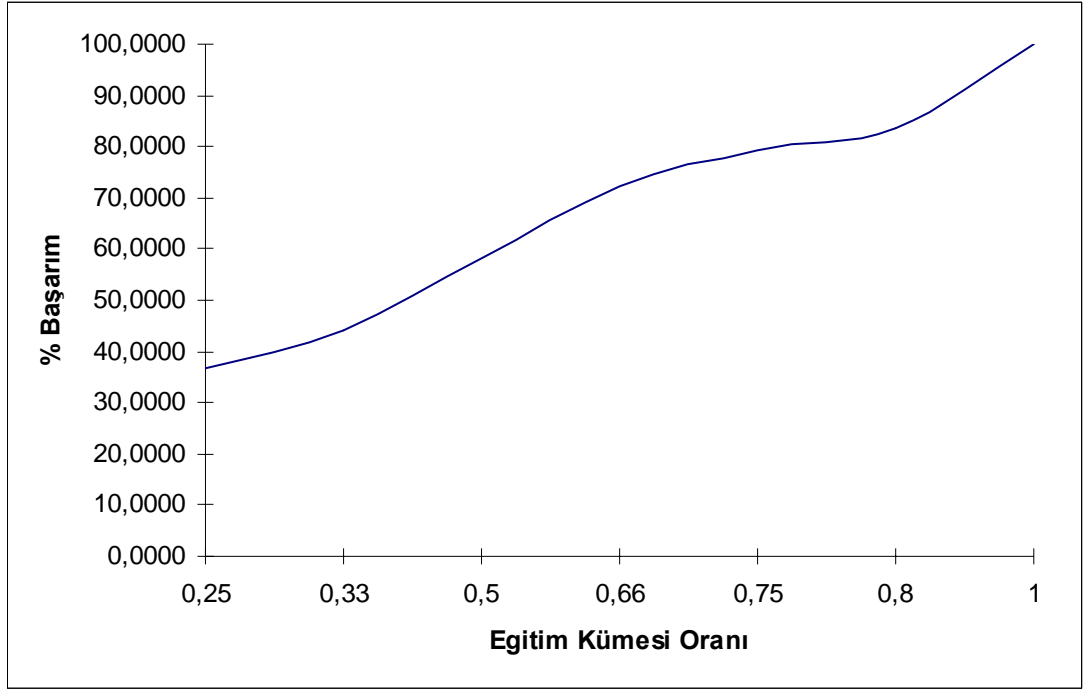


**Şekil 4.2:** GKM ile yapılan sınıflandırma sonucu oluşan pozitif yanlış alarm oranları

Aynı çalışmalar, aynı veri setiyle DVM ile tekrarlanmıştır. Yapılan tür tespiti çalışmalarında alınan performans sonuçları tablo 4.4 'te gösterilmiştir. Şekil 4.3 'te ise bu tablonun grafiksel gösterimi yapılmaktadır.

**Tablo 4.4:** Müzik türlerinin DVM ile sınıflandırma sonuçları

Eğitim Kümesi Oranı	% Başarım
0,25	36,8568
0,33	44,0941
0,5	58,1880
0,66	72,3723
0,75	79,2192
0,8	83,4234
1	100,0000

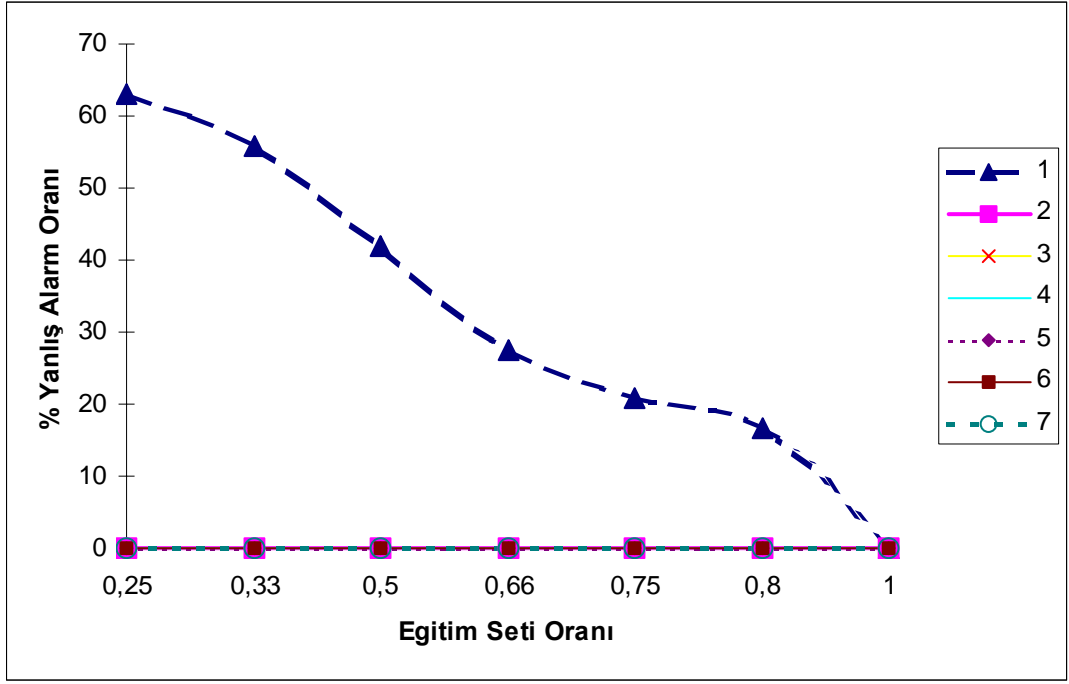


**Şekil 4.3:** Müzik türlerinin DVM ile sınıflandırma sonuçları

Tablo 4.5 'te her bir tür için, yapılan her sınıflandırmaya ait pozitif yanlış alarm oranları sunulmaktadır. Bunun grafiksel gösterimini ise şekil 4.4 'te görmek mümkündür.

**Tablo 4.5:** DVM ile yapılan sınıflandırma sonucu oluşan pozitif yanlış alarm oranları

Eğitim Kümesi Oranı	1	2	3	4	5	6	7
0,25	63,14314	0	0	0	0	0	0
0,33	55,90591	0	0	0	0	0	0
0,5	41,81181	0	0	0	0	0	0
0,66	27,62763	0	0	0	0	0	0
0,75	20,7808	0	0	0	0	0	0
0,8	16,57658	0	0	0	0	0	0
1	0	0	0	0	0	0	0



**Şekil 4.4:** DVM ile yapılan sınıflandırma sonucu oluşan pozitif yanlış alarm oranları

#### 4.2 Parça Ayrımı Testleri

Parça ayrımı testlerinde tez kapsamında geliştirilen imza tanıma sisteminin otomatik müzik türü ayırt etme performansı incelenmiştir. Bu testler sırasında kullanılan veri seti, parçalar ve sınıfları tablo 4.6 'da gösterilmiştir.

**Tablo 4.6:** Parça ayrımı testlerinde kullanılan veri seti, türler, sanatçılar, örnek sayıları ve parça adetleri

Tür	Sanatçı	Örnek Sayısı	Parça Sayısı
Klasik	Mozart	798	5
	Grieg Schumann	620	3
Sufi	Yansimalar	643	4
	Ömer Faruk Tekbilek	838	5
FreeJazz	Erik Trufaz	674	4
	Jan Garbarek	553	3
	Thelenious Monk	310	2
Yabancı Etnik	Anuar Brahem	860	4
	Eleni Karandiru	620	5
Yabancı Pop	Lou Bega	731	6
	Celine Dion	726	5
Yerli Etnik	Selda Bağcan	708	5
	Laço Tayfa	799	5
Yerli Pop	Erkan Güteryüz	717	5
	Pınar Dilşeker	763	6

Toplamda 67 müzik parçasından oluşan veri setinin tamamı eğitimde kullanılarak bir veritabanı oluşturulmuştur. Daha sonra bu verilerin tamamı test için kullanılarak sınıflandırma performansına bakılmıştır. Geliştirilen ses imzası tanıma sistemi GKM ile %94.6369 başarıyla parçaları birbirinden ayırmayı başarmıştır. Tablo 4.7 'de GKM ile yapılmış ses imzası sınıflandırma sonucundaki her bir parçaya ait pozitif yanlış alarm oranları sunulmuştur.

**Tablo 4.7:** GKM ile parça ayrımı pozitif yanlış alarm oranları

Sınıf No	% Yanlış Alarm Oranı	Sınıf No	% Yanlış Alarm Oranı	Sınıf No	% Yanlış Alarm Oranı	Sınıf No	% Yanlış Alarm Oranı
1	0.3415	23	0.1205	45	0	67	0
2	0.1908	24	0	46	0		
3	0.0502	25	0	47	0.0201		
4	0.3415	26	0.0100	48	0.0502		
5	0.1506	27	0.5022	49	0.0603		
6	0.2812	28	0.6026	50	0.0100		
7	0	29	1.0244	51	0		
8	0.0703	30	0.0904	52	0		
9	0.0301	31	0.2511	53	0		
10	0.0100	32	0.2210	54	0		
11	0	33	0.0100	55	0		
12	0	34	0.1707	56	0		
13	0.0100	35	0.0100	57	0		
14	0.0100	36	0.0402	58	0		
15	0.1004	37	0.0502	59	0		
16	0.0703	38	0.0100	60	0.0301		
17	0.0502	39	0	61	0.0100		
18	0.0301	40	0.0603	62	0.0301		
19	0.0803	41	0	63	0		
20	0.0301	42	0	64	0.0502		
21	0.0502	43	0	65	0.0100		
22	0	44	0	66	0.0201		

Aynı işlemler DVM ile tekrarlandığında, sistem tüm test setini %100 doğrulukla tanımayı başarmıştır. Bu yüzden hiçbir sınıfta pozitif yanlış alarm gözlenmemiştir.

### 4.3 Performans Testleri Sonuçları

Yapılan tür ve klip ayrımı performans testlerinde görülmektedir ki GKM, DVM'ye göre eğitim kümesine bağımlılık konusunda kullanıcıya çok daha fazla özgürlük sağlamaktadır. Daha az örnekle iyi bir sistem eğitimini mümkün kılmaktadır. Ancak

eđitim kümesinin tümü bile kullanılsa %100 doğrulukla klip veya sınıf ayrımı sağlanamamakla beraber %94.6369'luk performans da azımsanmamalıdır. DVM 'de test kümesinin tamamı eğitimde kullanıldığında %100 başarıma ulaşılabilir.

Bu başarısına rağmen performans açısından DVM birçok konuda GKM 'nin gerisinde kalmaktadır. Aynı veri setiyle yapılan ölçümlerde görülmüştür ki, GKM 'nin eğitim ve sınıflandırma süreleri DVM 'ninkine oranla çok daha kısadır. Bu da gerçek zamanlı uygulamalarda çok daha fazla kanalın eş zamanlı dinlenerek müzik bölütlerinin doğru eşlenmesine olanak sağlamaktadır. Bu değer DVM 'de 10'lar mertebesinde iken GKM 'de 1000'ler mertebesindedir. Haitzma ve Kalker kendi sistemlerinde ancak 200 kanalı yönetebildiklerinden bahsetmektedirler[11].

Buna ek olarak sisteme veritabanı yapısı açısından bakıldığında GKM DVM 'ye göre çok daha büyük bir esneklik sağlamaktadır. Sunulan sistemde veritabanı yerine eğitim sonucu elde edilen parametreler kullanıldığından yeni bir sınıf veri tabanına eklenmek istendiğinde veya var olan sınıf çıkarılmak istendiğinde veya güncellenmesi gerektiğinde, yeni parametrelerin hesaplanması gerekecektir. Bu işlem GKM 'de sadece yeni sınıf için veya güncellenmesi gereken tür sınıfı için yapılması gerekirken, DVM 'de tüm sistem için yapılır ve bu da çok büyük veri tabanlarında çok fazla zaman kaybettirmektedir.

Yukarıda bahsedilen sonuçlara dayanılarak sisteme yapılacak ataklara karşı sistemin dayanıklılığını ölçecek testlerde, yani gürbüzlük testlerinde, sadece GKM ile elde edilen sonuçlara yer verilmiştir.

## 5. GÜRBÜZLÜK TESTLERİ

Ses imzalarının sağlaması gereken özelliklerden birisi de doğal yada kasti yapılan ataklara karşı sistemin performansını korumasıdır. Buna, sistemin ataklara karşı gürbüzlüğü adı verilir. Bu bölümde tez kapsamında geliştirilen sistemin mp3 sıkıştırma, radyo yayınlarında yapılan zamanda sıkıştırma, kanal gürültüsü ve alt geçiren süzgeç ile filtreleme ataklarına karşı gürbüzlüğü incelenmiştir.

Çalışmalarda performans testlerinde kullanılan veri aynen kullanılmış, bu verilere çeşitli ataklar uygulanıp, karar verme ve eşleme performansları incelenmiştir. Elde edilen başarımlar değerleri, tür sınıflandırmadaki %86,6567 ve parça sınıflandırmadaki %94,6369 'luk ataksız başarımlarla karşılaştırılmıştır.

### 5.1 MP3 Sıkıştırma Atağı

7 farklı müzik türüne ait 67 adet parçaya çeşitli kodlama oranlarında mp3 sıkıştırma işlemi uygulanmıştır. Sıkıştırılmış dosyalar, wav biçimli ses verileriyle eğitilmiş sistemin test üst bloğu girişine uygulanmış ve karar verme ve eşleme performansı incelenip ataksız elde edilen performans ile karşılaştırılmıştır. Sıkıştırılmış dosyalara ait doğru karar verme ve eşleme performansları tablo 5.1 ve tablo 5.2 'de sunulmuştur.

**Tablo 5.1:**mp3 sıkıştırma atağı müzik türü sınıflandırma başarımları

<b>Müzik Türü Sınıflandırma</b>	
Sıkıştırma Oranı	% Başarım
320K	85,5148
192K	85,2269
128K	74,9824
64K	59,4410

**Tablo 5.2:**mp3 sıkıştırma atağı parça sınıflandırma başarımları

<b>Parça Sınıflandırması</b>	
Sıkıştırma Oranı	% Başarım
320K	85,8061



192K	84,7640
128K	60,5357
64K	27,4884

Yukarıdaki iki tablodan da görüleceği üzere; sunulan sistem 128 kBit mp3 sıkıştırılmaya kadar hem müzik türü hem müzik parçası sınıflandırmada performansını korumuştur. Ancak bu değerden sonra elde edilen başarımlar kabul edilebilir değerler değildir. Bu aşamadan itibaren, sıkıştırılmış dosyaların da sistemin eğitiminde kullanılması, karşılaşılan soruna çözüm olarak sunulabilir. Tablo 5.3 ve 5.4 'te sıkıştırılmış verilerin de eğitimde kullanılmasıyla elde edilen sınıflandırma başarımları sunulmuştur.

**Tablo 5.3:**Sıkıştırılmış verilerin de eğitimde kullanılması sonucu elde edilen müzik türü sınıflandırması başarımları

<b>Müzik Türü Sınıflandırma</b>		
Sıkıştırma Oranı	Ataklı veri sınıflandırma % başarımları	Ataksız veri sınıflandırma % başarımları
320K	86,6399	86,7229
192K	86,4144	86,4417
128K	85,1740	85,3470
64K	79,4591	83,2982

**Tablo 5.4:**Sıkıştırılmış verilerin de eğitimde kullanılması sonucu elde edilen parça sınıflandırması başarımları

<b>Parça Sınıflandırması</b>		
Sıkıştırma Oranı	Ataklı veri sınıflandırma % başarımları	Ataksız veri sınıflandırma % başarımları
320K	92,0442	92,4777
192K	91,8486	92,3170
128K	90,9921	91,1520
64K	87,7539	89,9468

Görüleceği üzere, sıkıştırılmış dosyaların da eğitimde kullanılması, sistem performansını radikal biçimde değiştirmektedir.

## 5.2 Zamanda Sıkıştırma Atağı

7 farklı müzik türüne ait 67 adet parçaya çeşitli sıkıştırma oranlarında zaman bölgesi sıkıştırma işlemi uygulanmıştır. Zaman bölgesi sıkıştırma işlemi, dosyanın boyu belirli oranlarda kısaltıldığı halde insan kulağının farkı algılayamamasına dayanmaktadır. Yayın yapan radyo kanalları özellikle reklam müziklerini bu şekilde

zaman bölgesinde sıkıştırıp, zaman kazanabilmektedirler. Sıkıştırılmış dosyalar, normal uzunluktaki wav biçimli ses verileriyle eğitilmiş sistemin test üst bloğu girişine uygulanmış ve karar verme ve eşleme performansı incelenip ataksız elde edilen performans ile karşılaştırılmıştır. Sıkıştırılmış dosyalara ait doğru karar verme ve eşleme performansları tablo 5.5 ve tablo 5.6 'da sunulmuştur.

**Tablo 5.5:** Zamanda sıkıştırma atağı müzik türü sınıflandırması başarımları

<b>Muzik Türü Sınıflandırma</b>	
Sıkıştırma Oranı	% Başarım
1%	85,8381
2%	85,7143
3%	85,9315
4%	85,6812

**Tablo 5.6:** Zamanda sıkıştırma atağı parça sınıflandırma başarımları

<b>Parça Sınıflandırması</b>	
Sıkıştırma Oranı	% Başarım
1%	86,6942
2%	87,2279
3%	87,0665
4%	87,1131

Yukarıdaki tablolardan da görüleceği üzere, tez kapsamında geliştirilen sistem %4'lük zaman bölgesi sıkıştırmada bile performansını korumaktadır. Bu konuda en başarılı çalışmalardan biri olan Haitsma ve Kalker[11] 'in geliştirdikleri ses imzası sisteminde, %2' lik sıkıştırmada bile sistemin başarımı çok düşmektedir. Bu da tez kapsamında geliştirilen sistemin başarısını açıkça ortaya koymaktadır.

### 5.3 Beyaz Gauss Gürültüsü Atağı

Bu testlerde geliştirilen sistemin kanal gürültüsüne olan gürbüzlüğü test edilmiştir. Kanal gürültüsünü temsilen ses verilerine çeşitli işaret/gürültü oranlarında beyaz Gauss gürültüsü eklenmiştir. Gürültülü dosyalardan elde edilen imzaların sınıflandırma performansları incelenmiştir. Sonuçlar tablo 5.7 ve 5.8 'de sunulmuştur.

**Tablo 5.7:** Gauss gürültüsü atağı müzik türü sınıflandırması başarımları

<b>Müzik Türü Sınıflandırma</b>	
SNR	% Başarım
10 dB	12,6324
20 dB	27,4728

30 dB	49,2115
40 dB	72,7033

**Tablo 5.8:** Gauss gürültüsü atağı parça sınıflandırması

<b>Parça Sınıflandırma</b>	
SNR	% Başarım
10 dB	3,7592
20 dB	8,2816
30 dB	30,5423
40 dB	60,5691

Yukarıdaki iki tablodan da görüleceği üzere; sunulan sistem 40 dB SNR değerinden daha düşük değerlerde hem müzik türü hem müzik parçası sınıflandırmada performansını koruyamamaktadır. Bu aşamadan itibaren, gürültü eklenmiş ses verilerinin de sistemin eğitiminde kullanılması, karşılaşılan soruna çözüm olarak sunulabilir. Tablo 5.9 ve 5.10 'da sıkıştırılmış verilerin de eğitimde kullanılmasıyla elde edilen sınıflandırma başarımları sunulmuştur.

**Tablo 5.9:** Gürültü eklenmiş verilerin de eğitimde kullanılması sonucu elde edilen müzik türü sınıflandırması başarımları

<b>Müzik Türü Sınıflandırma</b>		
SNR	Ataklı veri sınıflandırma % başarımları	Ataksız veri sınıflandırma % başarımları
10 dB	69,1320	85,1562
20 dB	77,0296	84,4130
30 dB	84,0777	85,1160
40 dB	85,2947	85,7286

**Tablo 5.10:** Gürültü eklenmiş verilerin de eğitimde kullanılması sonucu elde edilen müzik türü sınıflandırması başarımları

<b>Parça Sınıflandırma</b>		
SNR	Ataklı veri sınıflandırma % başarımları	Ataksız veri sınıflandırma % başarımları
10 dB	94,3867	91,8148
20 dB	95,6504	91,6943
30 dB	96,7138	91,8248
40 dB	95,8740	92,4174

Görüleceği üzere, gürültü eklenmiş sıkıştırılmış dosyaların da eğitimde kullanılması, sistem performansını radikal biçimde değiştirmektedir. Bu şekilde 10dB

işaret/gürültü oranında bile ses imzası sınıflandırma yapılabilmektedir. Parça sınıflandırmadaki performansın ataksız durumdaki performansın üzerine çıkması da dikkat çekicidir.

#### **5.4 0-10kHz Alt Geçiren Süzgeç ile Filtreleme Atağı**

7 farklı müzik türüne ait 67 adet parça, 10 kHzlik alt geçiren süzgeçten geçirildikten sonra, normal bantlı wav biçimli ses verileriyle eğitilmiş sistemin test üst bloğu girişine uygulanmış ve karar verme ve eşleme performansı incelenip ataksız elde edilen performans ile karşılaştırılmıştır. Filtrelenmiş dosyalarla yapılan müzik türü sınıflandırmasında %44.4412'lik bir başarı gösterilirken parça sınıflandırmasında %17.4069 'luk başarı elde edilmiştir. Her ikisi de göstermektedir ki geliştirilen sistem, üst frekans bileşenlerinin yok edildiği bu tip bir atağa karşı performansını koruyamamaktadır.

## 6. SONUÇLAR VE TARTIŞMALAR

Tez kapsamında Gauss Karışım Modeli kullanarak ses imzası sınıflandırma yapabilen ses imzası tanıma sistemi geliştirilmiştir. Geliştirilen sistem, girişine wav ve mp3 dosyalarını kabul edip, formattan bağımsız imza çıkarımı işlemini, işlemsel karmaşıklığı düşük olduğundan gerçek zamandan çok daha kısa sürede tamamlayabilmektedir. Ayrıca, veritabanında ses imzası arama işlemini, modellenen parametreleri kullanarak sınıflandırma yoluyla gerçekleştirdiğinden çok hızlı ve isabetli sonuç vererek, geniş veri setlerinin otomatik işlenmesine olanak sağlamaktadır.

İmzaların oluşturulmasında kullanılan öznitelikler ses bölütünün içeriğini çıkarmaya yönelik olup, ses parçasını yegâne tanımlayabilmektedirler. Ayrıca işlemsel karmaşıklıkları düşük olup, gerçek zamanın çok altında hesaplanabilmektedirler. Bunun yanı sıra öge boyu parametresi 2 saniyedir ve MPEG21 standardı olan 5sn sınırından oldukça düşüktür. Buna karşın yüksek ses imzası tanıma ve eşleme başarımına ulaşma imkânı sağlanmaktadır.

Performans testlerinde DVM ile de sınıflandırma işlemleri tekrarlanmış ve GKM ile performansları karşılaştırılmıştır. Parça ayırımında DVM daha üstün bir performans gösterirken, tür ayırımında GKM 'in üstünlüğü göze çarpmaktadır. Ancak, sisteme yeni bir parça eklenmesi gerektiğinde veya var olan bir parçanın veritabanından çıkarılması gerektiğinde GKM DVM 'ye göre çok daha kullanışlı bir sınıflandırıcı olarak göze çarpmaktadır. Ses imzalarının modellenmesinde ve sınıflandırmasındaki hızları karşılaştırıldığında da GKM 'in DVM 'ye büyük üstünlük kurduğu gözlenmiştir. GKM 'nin bu hızı, hali hazırda yapılmış olan birçok sistemden daha çok kanalı aynı anda yönlendirebilecek kapasiteye sahiptir.

Tez kapsamında ataklara karşı sistemin gürbüzlüğü de incelenmiştir. Buna göre geliştirilen sistem, en başarılı olarak zamanda sıkıştırma atağına karşı koymuştur. %4 'lük sıkıştırma oranına rağmen performansını korumayı başarmıştır. Mp3 sıkıştırmaya 128 kBitlik kodlamaya kadar karşı koyabilen sistem, Gauss gürültüsü atağına ise ancak 40dB SNR değerine kadar başarılı sonuçlar verebilmektedir. Ancak

her iki durumda da sistemin eğitiminde ataklı dosyalar da kullanıldığı takdirde sistem başarımı radikal olarak yükselmekte ve 10 dB SNR ve 64kBitlik kodlamada bile tatmin edici sonuçlar verebilmektedir. Filtreleme atağında ise sistem başarımını koruyamamış ve çok düşük başarımlar elde edilmiştir. Sistemin eğitiminde kullanılan tüm veriler öncelikle bir alt geçiren filtreden geçirilirse, bu sorunun da üstesinden gelinecektir. Çünkü bu atak üç önemli özneliği direkt olarak etkilemektedir. Bunlar; SGS, İKM ve İAN 'dir.

## KAYNAKLAR

- [1] **Cano, P., Batlle, E., Kalker, T., Haitsma, J.**, 2002. A Review of Algorithms for Audio Fingerprinting, *in Proc. of IEEE Workshop of Multimedia and Signal Processing*, pp. 169-173.
- [2] **Zhang, T., Kuo, C.C.J.**, 2001. Audio Content Analysis for Online Audiovisual Data Segmentation and Classification, *IEEE Transactions on Speech and Audio Processing*, Vol. 9, No. 4, pp. 441-457.
- [3] **Tzanetakis, G., Cook, P.**, 1999. Multifeature Audio Segmentation for Browsing and Annotation, *in Proc. 1999 IEEE Workshop on Applications of Signal Processing to Audio and Acoustics*, pp. W99-1-4.
- [4] **Sundaram, H. , Chang, S.F. ,** 2000. Audio Scene Segmentation Using Multiple Features, Models and Time Scales, *ICASSP 2000, International Conference in Acoustics, Speech and Signal Processing*.
- [5] **Lu, L., Li, S.Z., Zhang, H.J.**, 2001. Content-Based Audio Segmentation Using Support Vector Machines, *IEEE ICME*, pp. 191.
- [6] **Raynolds, D. A., Quatieri, T.F., Dunn, R.B.**, 2000. Speaker Verification Using Adapted Gaussian Mixture Models, *Digital Signal Processing*, pp. 19-41.
- [7] **Marques, J., Moreno, J.P.**, 1999. A Study of Musical Instrument Classification Using Gaussian Mixture Models and Support Vector Machines, *Cambridge Research Laboratory Technical Report Series, CRL 99/4*.
- [8] **Bengio, S.**, 2005. Gaussian Mixture Models , *Statistical Machine Learning Lecture Notes*, IDIAP Research Institute, pp. 1-25.
- [9] **Haitsma, J.A.**, 2002. Audio Fingerprinting, Nat. Lab. Unclassified Report, Netherland.
- [10] **Allamanche E., Herre J., Hellmuth O., Bernhard F.B., Cremer M.**, 2002. Content-based Identification of Audio Material Using MPEG-7 Low

Level description, in *Proc. Of the Int. Symp. Of Music Information Retrieval*.

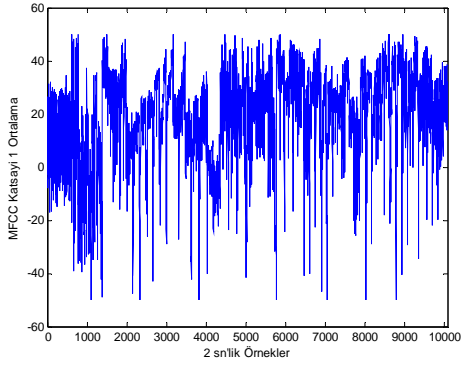
- [11] **Haitsma, J.A., Kalker, T.**, 2002. A Highly Robust Audio Fingerprinting System, *International Symposium on Musical Information Retrieval (ISMIR2002)*, pp. 144-148.
- [12] **Burges, C., Platt, J., Jana, S.**, 2002. Extracting Noise-Robust Features From Audio Data, in *Proc. of ICASSP2002*.
- [13] **Battle, A., Massip, J., Guaus, E.**, 2002. Automatic Song Identification in Noisy Broadcast Audio, *Proceedings of IASTED International Conference on Signal and Image Processing 2002*.
- [14] **Haitsma, J., Kalker, T., Oostveen, J.**, 2001. Robust Audio Hashing for Content Identification, in *Proc. of CBMI'01*.
- [15] **Kurth, F., Ribbrock, A., Clausen, M.**, 2002. Identification of Highly Distorted Audio Material for Querying Large Scale Databases, in *Proc. AES 112th Int. Conv.*
- [16] **Sukittanon, S., Atlas, L.**, 2002. Modulation Frequency Features for Audio Identification, in *Proc. of ICASSP*.
- [17] **Özer, H., Sankur, B., Memon, N.**, 2004. Robust Audio Hashing for Audio Identification, *EUSIPCO'2004: European Conf. on Signal Processing*, pp. 2091-2094.
- [18] **Kimura, A., Kashino, K., Kurozumi, T., Murase, H.**, 2001. Very Quick Audio Searching: Introducing Global Prunning to the Time-Series Active Search, in *Proc. of Int. Conf. On Computational Intelligence and Multimedia Applications*.
- [19] **Mihçak, K.**, 2004. A perceptual Audio Hashing Algorithm, in *4th Workshop on Information Hiding*.
- [20] **Schreir, E.**, 1997. Construction and Evaluation of Robust Multifeature Speech/Music Discriminator, in *Proc. of ICASSP 97*.



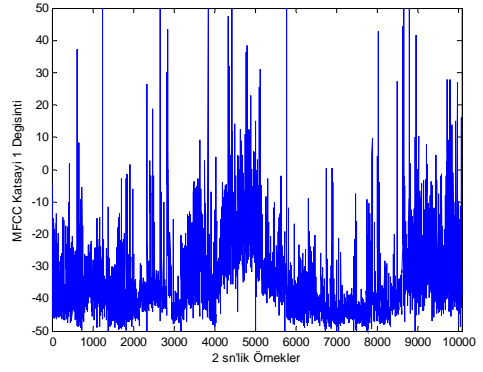
- [21] **Reiss, J., Aucouturier, J.J., Sandler, M.**, 2001. Efficient Multidimensional Searching Routines for Musical Information Retrieval, *International Symposium on Musical Information Retrieval (ISMIR2001)*.
- [22] **Chang, C.C., Lin, C.J.**, LIBSVM A Library for Support Vector Machines, <http://www.csie.ntu.edu.tw/~cjlin/libsvm/>
- [23] **Paalanen, P., Kamarainen, J., Kalviainen, H.**, GMMBAYES - Bayesian Classifier and Gaussian Mixture Model Methods, <http://www.it.lut.fi/project/gmmbayes/>

## EKLER

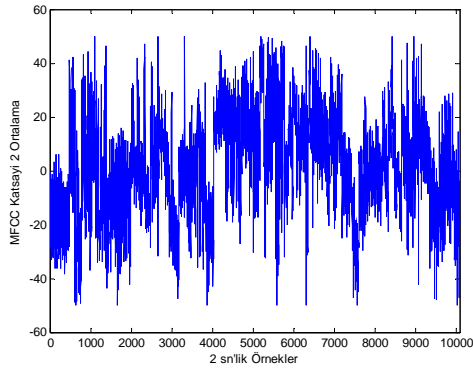
### Ek\_A 67 parçadan çıkarılan ortalama ve deęişinti MFKK öznelikleri



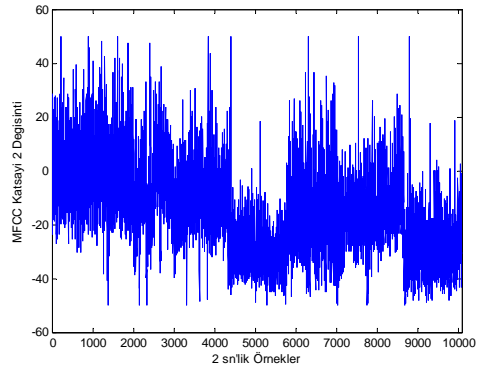
(a) MFKK Katsayı 1 Ortalama



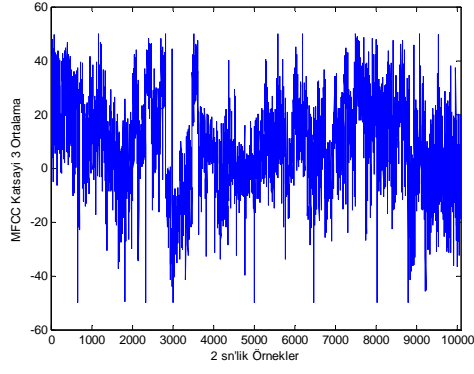
(b) MFKK Katsayı 1 Deęişinti



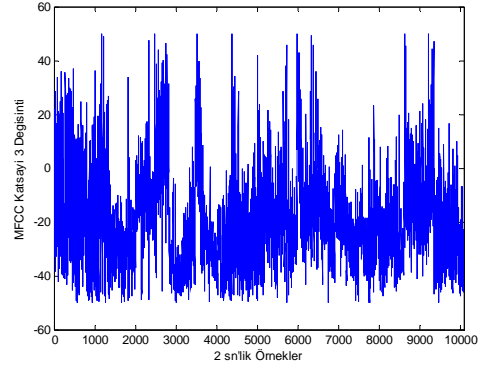
(c) MFKK Katsayı 2 Ortalama



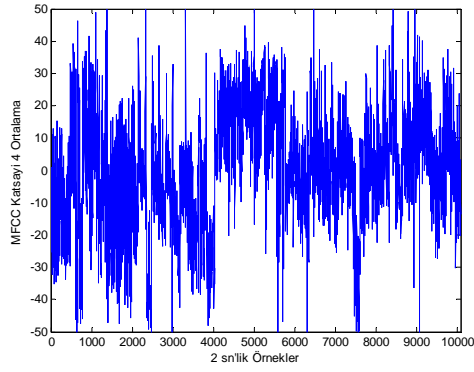
(d) MFKK Katsayı 2 Deęişinti



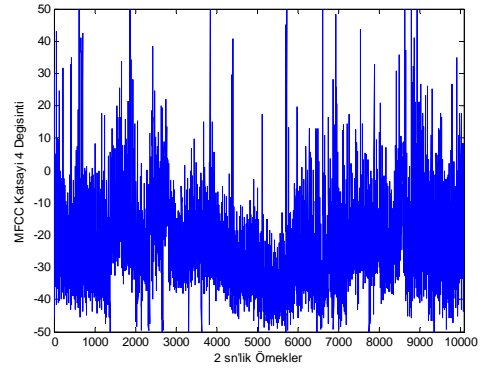
(e) MFKK Katsayı 3 Ortalama



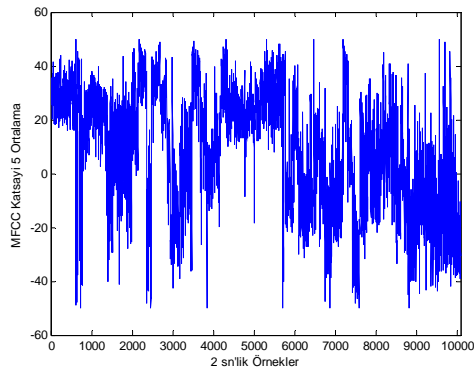
(f) MFKK Katsayı 3 Değişimi



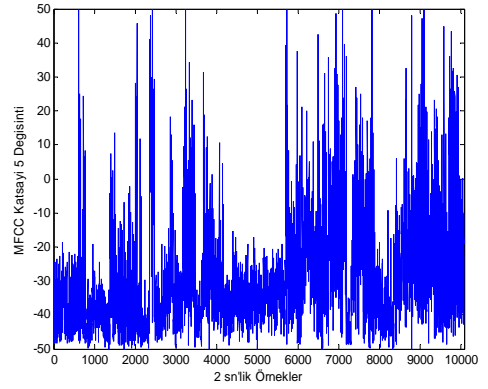
(g) MFKK Katsayı 4 Ortalama



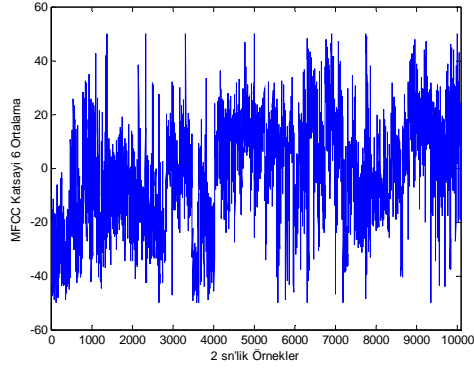
(h) MFKK Katsayı 4 Değişimi



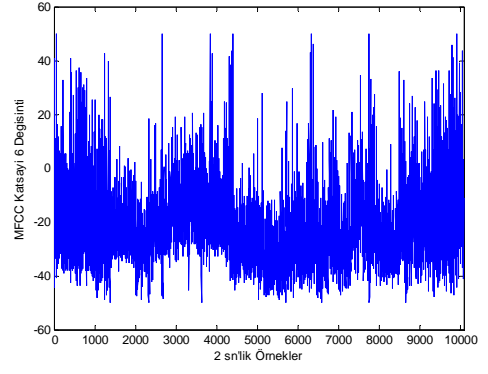
(i) MFKK Katsayı 5 Ortalama



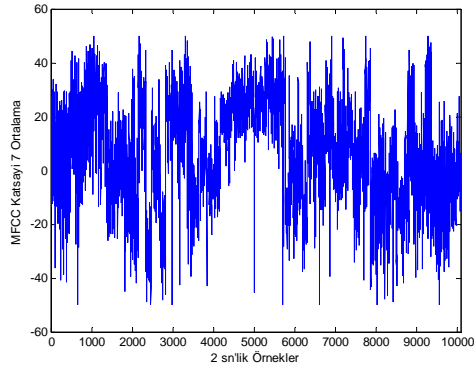
(j) MFKK Katsayı 5 Değişimi



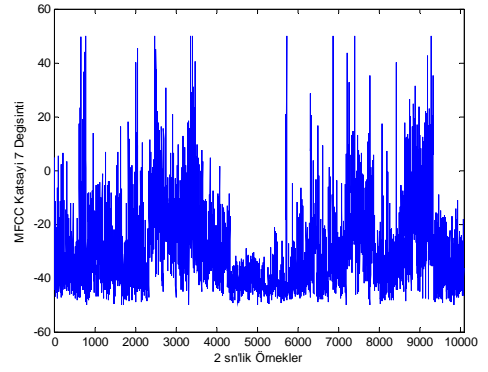
(k) MFKK Katsayı 6 Ortalama



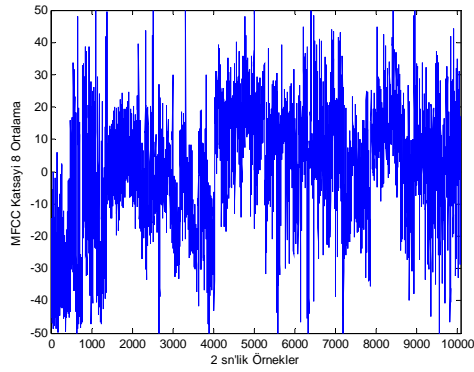
(l) MFKK Katsayı 6 Değişimi



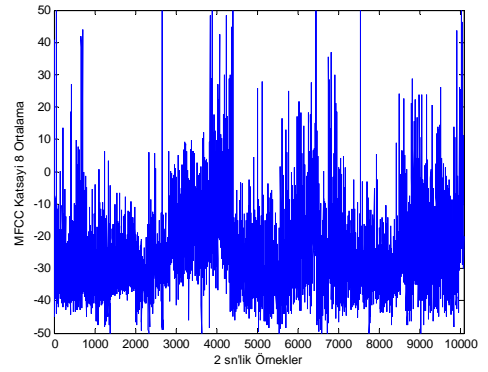
(m) MFKK Katsayı 7 Ortalama



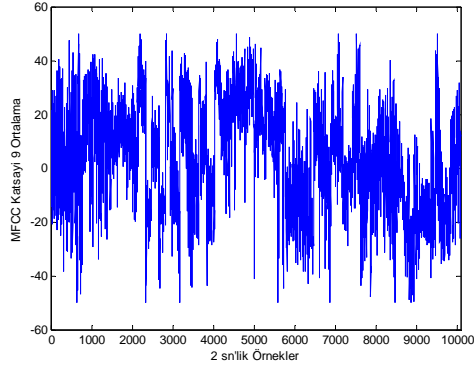
(n) MFKK Katsayı 7 Değişimi



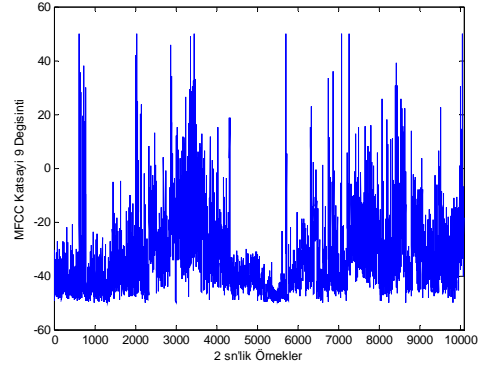
(o) MFKK Katsayı 8 Ortalama



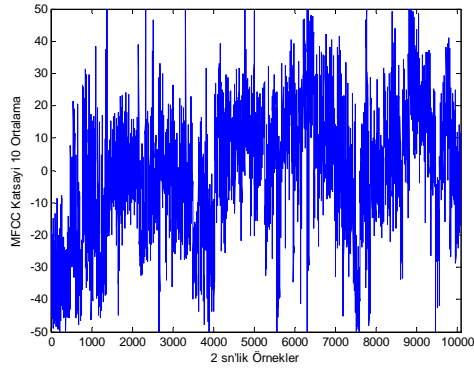
(p) MFKK Katsayı 8 Değişimi



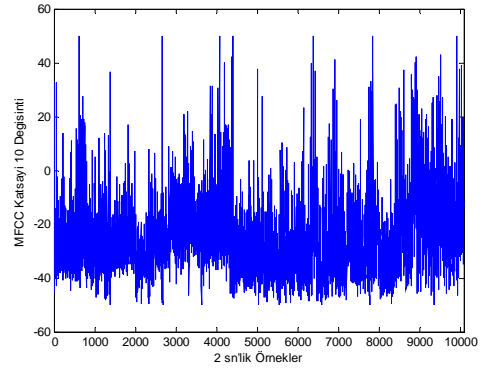
(q) MFKK Katsayı 9 Ortalama



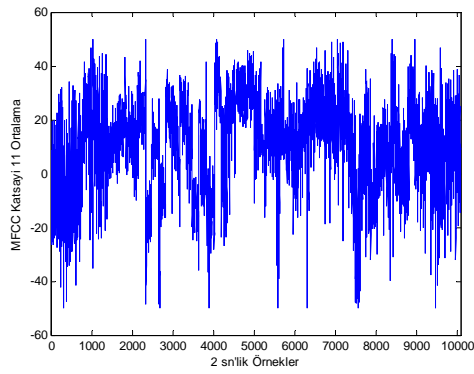
(r) MFKK Katsayı 9 Değişimi



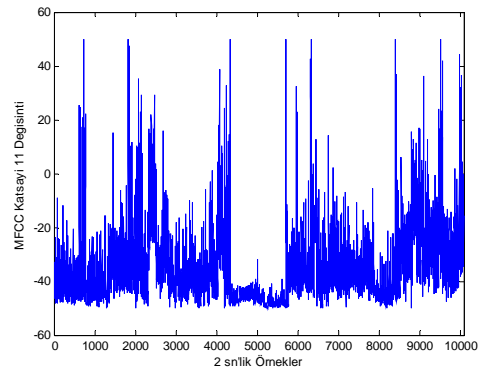
(s) MFKK Katsayı 10 Ortalama



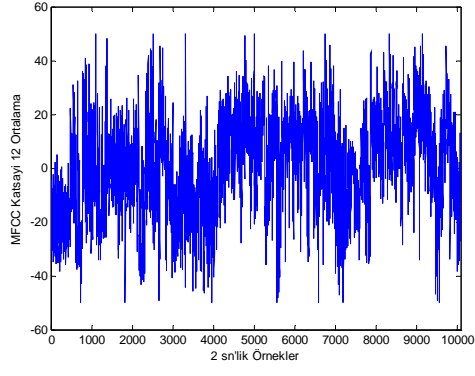
(t) MFKK Katsayı 10 Değişimi



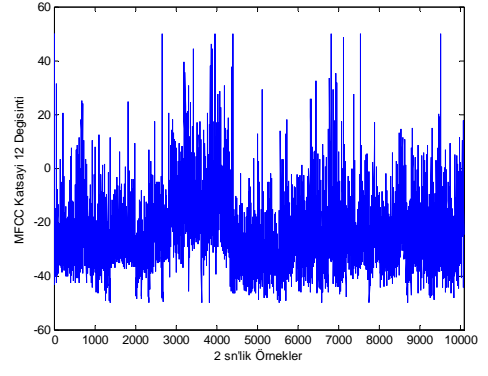
(u) MFKK Katsayı 11 Ortalama



(v) MFKK Katsayı 11 Değişimi



(x) MFKK Katsayı 12 Ortalama



(y) MFKK Katsayı 12 Değişimi



	6	4	26	88	2	1	6	26	6	08	38	42	68	09	22	96	15	50	25	68	81	98	48	3	06	37	3	43	26	74	66	32	1	71	7	48				
	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-		
<b>8</b>	0,369	0,178	0,329	0,139	-0,077	0,364	0,642	1,1815	0,671	1,0511	0,119	1,1503	1,175	0,316	1,119	0,406	1,106	0,068	-0,028	0,503	0,226	0,038	0,013	0,013	0,014	0,016	0,008	-0,048	-0,004	0,138	0,043	0,115	0,216	0,039	0,017	0,015	0,012			
<b>9</b>	0,2915	0,152	0,292	0,155	0,297	0,155	0,546	0,671	0,851	0,774	0,012	0,987	0,122	0,348	0,868	0,499	0,829	0,024	0,033	0,496	0,203	0,028	0,069	0,012	0,006	0,011	0,009	0,017	0,068	0,078	0,010	0,006	-0,043	0,011	0,002	0,037	0,004	0,005		
<b>10</b>	0,38165	0,18593	0,34973	0,13073	0,03822	0,39877	0,65084	1,10511	0,77342	1,10191	0,13484	1,11815	0,16888	1,11766	0,37045	1,02452	0,07076	-0,00648	0,04439	0,52932	0,2602	0,03053	0,0165	0,0233	0,00394	0,0106	0,00062	-0,00626	-0,00062	0,10948	0,03898	0,05013	0,02084	0,01410	0,01737	0,00813	0,00813			
<b>11</b>	0,0917	-0,092	0,0577	0,0503	-0,012	0,1038	0,0838	0,1119	0,0134	0,1348	0,2324	0,1161	0,0317	0,0229	0,0830	0,0550	0,0039	0,0533	0,00506	0,00099	0,0166	0,0711	0,0221	0,0181	0,0540	0,0254	0,0067	0,01313	0,07028	0,09517	0,0766	0,0913	0,0707	0,02848	0,1567	0,2298	0,0801			
<b>12</b>	0,38631	0,19571	0,41845	0,20914	-0,049	0,4381	0,3942	1,12203	0,98797	1,11415	0,1161	1,16386	0,2491	1,4077	1,7041	0,7042	1,0336	0,0408	0,0672	0,2929	0,0264	0,0412	0,0317	0,0084	0,0312	0,0867	0,0792	0,0177	0,01039	0,01071	0,0455	0,1286	0,0124	0,0807	0,0033	0,0574	0,0284	0,0584		
<b>13</b>	0,12123	0,01788	0,1809	0,04532	0,10262	0,06118	0,15682	0,15372	0,12612	0,1688	0,0317	0,2491	0,2438	0,1399	0,1295	0,2420	0,0764	0,0096	0,0106	0,1386	0,1917	0,0163	0,0271	0,1008	0,0849	0,0114	0,0205	0,00611	0,00863	0,01059	0,0518	0,0253	0,0707	0,0203	0,0070	0,0398	0,0593	0,0593		
<b>14</b>	0,40219	0,0249	0,37859	0,2091	0,0441	0,4177	0,0923	1,1818	0,0917	1,1729	0,0252	1,1303	1,1353	0,3978	1,3877	0,6038	1,0478	0,0428	0,0470	0,6556	0,2520	0,0254	0,0095	0,0091	0,0091	0,0805	0,0503	0,0079	0,0057	0,0057	0,0578	0,0517	0,01717	0,0062	0,0076	0,0037	0,0078	0,0049	0,008	
<b>15</b>	0,19775	0,07123	0,22531	0,0094	0,1721	0,2024	0,019	0,3135	0,3404	0,3730	0,0877	0,4095	0,1778	0,3987	0,3020	0,3313	0,2825	0,0408	0,0196	0,2515	0,2111	0,0230	0,0793	0,1685	0,0843	0,0234	0,0002	0,0002	0,0005	0,0005	0,0005	0,0005	0,0005	0,0005	0,0005	0,0005	0,0005	0,0005	0,0005	0,0005
<b>16</b>	-0,0	-0,0	0,33	0,0	-0,0	0,0	0,11	0,0	0,02	0,0	0,41	0,12	0,38	0,30	0,29	0,54	0,27	0,0	0,0	0,58	0,22	0,0	0,0	0,24	0,0	0,0	0,02	0,0	0,0	0,07	0,0	0,01	0,0	0,01	0,0	0,01	0,0	0,0	0,0	



	34 89 1	17 93	30 9	19 25 9	39 79 5	35 44 1	65 96	19	86 05 9	53	08 50 1	69	20 3	31	20 3	1	99 2	98	02 88 5	03 94 2	46 8	07 4	01 81 9	05 78 6	86 6	10 57 6	69 80 9	10 97	03 45 6	04 04	78 56	01 45 6	29 15	02 96 4	66 16	03 21 8		
<b>17</b>	- 0, 32 9	- 0, 86 3	0, 70 3	0, 09 3	0, 84 3	0, 71 9	0, 15 5	0, 02 6	0, 80 9	0, 45 21	0, 05 8	0, 70 4	0, 24 7	0, 60 5	0, 33 7	0, 54 2	0, 54 9	0, 45 5	0, 02 4	0, 02 8	0, 37 4	0, 28 1	0, 01 8	0, 01 82	0, 25 16	0, 08 3	0, 34 5	0, 02 76	0, 02 63	0, 01 1	0, 09 08	0, 00 7	0, 06 43	0, 00 58	0, 00 23	0, 00 56	0, 01 14	
<b>18</b>	- 0, 75 2	- 0, 19 6	0, 03 6	0, 18 9	0, 41 6	0, 37 6	0, 69 7	1, 18 06	0, 82 5	1, 07 64	0, 00 3	1, 36 59	0, 07 88	1, 28 97	1, 27 4	1, 45 5	1, 33 64	0, 04 3	0, 04 18	0, 04 28	0, 21 5	0, 10 53	0, 10 3	0, 20 5	0, 10 2	0, 07 5	0, 00 36	0, 09 6	0, 07 1	0, 00 26	0, 04 3	0, 12 7	0, 10 2	0, 07 7	0, 07 95	0, 07 19	0, 05 95	0, 04 2
<b>19</b>	0, 05 27	0, 01 95	0, 04 2	0, 01 2	0, 04 61	0, 05 55	0, 04 25	0, 06 8	0, 02 48	- 0, 06 8	- 0, 05 6	- 0, 03 5	- 0, 02 88	- 0, 04 8	- 0, 04 9	- 0, 02 4	- 0, 02 3	0, 03 79	0, 00 33	0, 00 56	0, 04 42	- 0, 02 05	0, 01 63	0, 05 9	0, 00 82	0, 01 5	0, 02 9	0, 02 34	0, 04 96	0, 04 88	0, 04 63	0, 05 15	0, 09 84	0, 06 86	0, 07 72	0, 04 77		
<b>20</b>	0, 01 66	0, 01 55	0, 02 7	0, 00 06	0, 01 62	0, 02 34	0, 03 68	0, 04 28	0, 03 35	0, 04 9	0, 00 3	0, 04 2	0, 01 6	0, 04 7	0, 01 5	0, 03 2	0, 02 8	0, 00 18	0, 00 33	0, 00 39	0, 02 8	0, 01 3	0, 01 34	0, 02 39	0, 00 2	0, 00 98	0, 03 03	0, 01 6	0, 01 03	0, 01 19	0, 01 58	0, 00 98	0, 01 76	0, 01 42	0, 01 26	0, 01 48	0, 01 26	0, 01 37
<b>21</b>	- 0, 78 3	- 0, 10 4	0, 44 2	0, 08 1	0, 21 4	0, 25 6	0, 37 9	0, 50 3	0, 49 3	0, 52 2	0, 01 36	0, 07 3	0, 13 1	0, 65 5	0, 25 8	0, 58 4	0, 37 4	0, 04 28	0, 02 5	0, 04 8	0, 02 9	0, 03 3	0, 08 3	0, 00 39	0, 04 39	0, 04 64	0, 03 8	0, 03 22	- 0, 04 1	0, 05 45	0, 02 3	0, 04 6	0, 06 3	0, 07 6	0, 07 3	0, 07 3	0, 04 3	0, 07 3
<b>22</b>	0, 20 98	0, 04 41	0, 42 70	0, 30 38	0, 16 10	0, 20 39	0, 15 98	0, 22 86	0, 20 88	0, 26 02	0, 07 11	0, 29 64	0, 19 17	0, 25 20	0, 21 11	0, 22 07	0, 28 13	0, 21 59	- 0, 01 05	0, 01 70	0, 02 89	0, 02 81	0, 02 42	- 0, 17 09	0, 15 27	0, 06 45	0, 03 87	0, 05 68	0, 09 50	0, 09 60	0, 00 36	0, 11 73	0, 14 14	0, 13 40	0, 01 14	- 0, 10 43	0, 10 14	0, 10 43
<b>23</b>	0, 03 14	0, 00 57	0, 02 23	0, 00 77	0, 02 56	0, 03 10	0, 02 20	0, 03 13	0, 02 02	0, 03 53	0, 02 21	0, 02 12	0, 01 63	0, 02 54	0, 01 30	0, 01 81	0, 01 90	- 0, 02 53	0, 01 62	0, 00 30	0, 02 20	0, 01 42	0, 02 04	0, 00 67	0, 00 28	0, 01 70	0, 01 28	0, 02 15	0, 02 22	0, 02 06	0, 01 99	0, 01 87	0, 04 44	0, 02 87	0, 01 44	0, 04 87	0, 02 63	0, 03 77
<b>24</b>	0, 11 70	0, 03 06	0, 08 75	0, 02 83	0, 08 16	0, 12 26	0, 10 75	0, 14 65	0, 06 17	0, 15 65	0, 12 81	0, 04 22	0, 02 71	0, 09 95	0, 07 93	0, 05 78	- 0, 10 82	0, 05 46	0, 01 09	0, 01 39	0, 08 09	- 0, 02 04	0, 02 67	0, 00 39	0, 00 84	0, 03 83	0, 02 94	0, 10 48	0, 08 61	0, 09 14	0, 07 50	0, 07 22	0, 00 13	0, 00 31	0, 17 34	0, 07 08	0, 08 2	0, 08 94

<b>25</b>	0,11 36 4	0,04 32 7	0,17 05 5	0,02 15 54	0,11 34 6	0,11 67 5	0,16 06 7	0,16 37 4	0,28 99 9	0,22 35 1	0,01 25 5	0,31 67 2	0,10 08 5	0,29 15 9	0,16 85 8	0,24 86 6	0,00 25 5	0,20 96 82	0,00 92 2	0,02 45 39	0,17 27 6	0,00 28 44	0,01 00 73	0,00 82 67	0,20 80 1	-	0,06 25 7	0,31 48 6	0,14 46 9	0,19 89 9	0,10 06 5	0,29 09 4	0,09 99 4	0,18 03 4	0,05 14 82	0,15 21 6
<b>26</b>	0,07 06 72	0,01 49 21	0,11 81 6	0,08 63 2	0,04 99 5	0,05 27 9	0,04 79 37	0,08 12 5	0,00 33 45	0,03 40 8	0,05 36 67	0,08 49 4	0,08 46 29	0,09 43 1	0,08 57 36	0,10 00 6	0,08 98 3	0,10 47 2	0,01 31 5	0,00 65 98	0,04 45 64	0,15 70 5	0,00 84 8	0,02 80 5	0,20 00 1	0,00 87 99	0,10 94 3	0,49 76 1	0,01 95 26	0,40 15 4	0,00 53 84	0,45 85 4	0,05 46 4	0,32 67 4	0,08 99 8	0,39 34 5
<b>27</b>	0,18 06 8	0,15 65 1	0,18 61 2	0,21 17 6	0,17 27 6	0,20 05 5	0,47 69 3	0,56 73 1	0,66 09 8	0,61 94 6	0,02 52 13	0,79 98 7	0,07 14 2	0,80 20 9	0,23 58 4	0,69 80 9	0,34 44 5	0,67 16 5	0,02 45 9	0,03 37 03	0,40 21 8	0,06 87 2	0,01 28 54	0,03 83 53	0,00 06 82	0,10 94 3	0,88 94 6	0,01 80 65	0,05 31 1	0,05 05 95	0,05 05 18	0,08 82 4	0,06 57 19	0,06 08 2	0,04 58 11	0,04 47 5
<b>28</b>	0,02 54 03	-	0,00 02 15	0,03 39 17	0,00 05 53	0,01 32 7	0,02 43 3	0,11 04 3	0,00 17 26	0,00 92 3	0,01 02 28	0,01 12 7	0,05 06 11	0,02 92 19	0,02 10 62	0,00 57 76	0,02 01 34	0,03 31 39	0,03 68 7	0,01 15 27	0,03 25 74	0,31 49 7	0,49 01 1	0,01 64 7	0,09 69 5	0,47 50 1	0,07 14 02	0,64 75 1	0,12 63 7	0,45 22 5	0,06 06 99	0,06 08 99	0,04 58 11	0,04 47 5		
<b>29</b>	0,08 26 87	0,01 17 46	-	0,02 05 26	0,05 47 98	0,07 08 07	0,06 15 26	0,09 09 04	0,02 68 2	0,06 87 9	0,07 95 2	0,01 31 6	0,01 86 7	0,03 32 7	0,00 45 6	0,03 19 63	0,02 09 6	0,04 01 96	0,00 43 19	0,05 95 03	0,01 04 22	0,05 02 58	0,10 22 7	0,14 48 6	0,01 48 26	0,05 31 1	0,09 24 5	0,00 10 76	0,12 11 78	0,11 45 3	0,23 22 4	0,14 77 1	0,16 67 7	0,09 08 4		
<b>30</b>	0,07 71 94	0,02 16 5	0,08 37 1	0,04 36 9	0,03 63 69	0,07 56 91	0,05 51 74	0,13 96 8	0,01 78 9	0,10 94 8	0,09 17 3	0,04 55 8	0,05 59 3	0,05 17 4	0,05 04 04	0,00 97 1	0,01 97 88	0,00 04 7	0,01 05 1	0,00 98 1	0,01 49 88	0,05 58 7	0,01 60 63	0,06 14 06	0,08 46 9	0,19 40 4	0,05 18 95	0,47 50 1	0,00 44 78	0,53 10 7	0,00 60 89	0,00 77 8	0,19 50 6	0,16 33 9	0,45 88 2	
<b>31</b>	0,06 43 65	0,00 44 35	0,00 25 01	0,01 44 4	0,03 12 38	0,06 00 15	0,00 79 66	0,04 15 3	0,06 43 7	0,13 89 8	0,12 86 1	0,02 18 3	0,05 78 59	0,00 20 73	0,07 78 56	0,09 47 08	0,00 59 26	0,04 45 63	0,00 60 45	0,02 36 8	0,00 99 64	0,01 61 89	0,09 89 9	0,10 53 84	0,00 80 4	0,05 14 02	0,07 45 3	0,12 10 89	0,00 95 4	0,21 35 2	0,11 78 2	0,26 88 1	0,14 88 2	0,20 84 2	0,09 56 85	
<b>32</b>	0,08 78 67	0,01 16 39	0,09 48 9	0,06 46 3	0,03 05 29	0,06 98 32	0,01 65 1	0,11 85 42	0,05 11 3	0,07 01 4	0,07 24 5	0,01 53 3	0,01 17 62	0,00 91 2	0,00 45 6	0,00 80 7	0,04 90 3	0,05 09 15	0,01 46 76	0,04 43 3	0,11 73 3	0,01 87 01	0,07 50 47	0,29 06 5	0,45 85 4	0,08 82 19	0,64 75 1	0,11 22 4	0,58 60 8	0,11 35 2	0,87 69 5	0,21 55 5	0,21 63 26	0,14 62 2	0,63 08 5	
<b>33</b>	0,20 77 5	0,04 33 84	-	0,13 06 64	0,20 11 5	0,14 20 4	0,00 21 1	0,00 02 03	0,20 28 18	0,00 20 84	0,08 77 48	0,00 03 88	0,07 02 07	0,00 03 76	0,01 29 51	0,06 07 15	0,00 58 43	0,00 12 39	0,01 48 84	0,00 06 42	0,00 09 09	0,04 01 14	0,11 07 99	0,29 45 4	0,45 85 4	0,08 82 19	0,64 75 1	0,11 22 4	0,58 60 8	0,11 35 2	0,87 69 5	0,21 55 5	0,21 25 2	0,00 31 5	0,00 43 5	



<b>Öznitelik No</b>	<b>Ortalama</b>	<b>Öznitelik No</b>	<b>Ortalama</b>
1	1,3477	19	0,10225
2	0,33374	20	0,059815
3	11,875	21	1,2867
4	13,276	22	1,4384
5	1,4205	23	0,062491
6	1,0719	24	0,12239
7	5,2437	25	1,9739
8	5,3237	26	3,7872
9	6,5939	27	1,8468
10	1,5985	28	2,4835
11	2,5989	29	0,73276
12	0,95158	30	2,3735
13	3,6893	31	0,77088
14	0,28257	32	2,4636
15	3,014	33	0,78175
16	-0,25446	34	2,1283
17	2,5798	35	0,75381
18	-0,5469	36	2,235

## ÖZGEÇMİŞ

Kadir Herkilođlu 1999 yılında Ankara Fen Lisesi'nden, 2003 yılında İstanbul Teknik Üniversitesi Elektronik ve Haberleşme Mühendisliği bölümünden mezun olmuştur. 2003 yılından bu yana İstanbul Teknik Üniversitesi Fen Bilimleri Enstitüsü Telekomünikasyon mühendisliği programında yüksek lisans yapmaktadır. Örüntü tanıma, sayısal ses ve imge damgalama ve ses imzası konuları üzerinde çalışmakta, 1 adet yurt içi bildirisi bulunmaktadır. Öğrenimi esnasında Mayıs 2003'ten Kasım 2003'e kadar 6 ay süreyle Prof. Dr. Bilge GÜNSEL yöneticiliğinde, İTÜ Elektrik Elektronik Fakültesi Çođulortam Sinyal İşleme ve Örüntü Tanıma Laboratuvarında gerçekleştirilen Sayısal Ses Damgalama projesinde çalışmıştır. Şu an HAVELSAN A.Ş. bünyesinde Yazılım Test Mühendisi olarak çalışmaktadır.