

SES DELİLLERİNİN İÇERİK ÖNİNCELEMESİ VE TAHRİFİNİN  
TESPİTİ İÇİN YENİ YÖNTEMLER

ERKAM UZUN

YÜKSEK LİSANS TEZİ  
BİLGİSAYAR MÜHENDİSLİĞİ ANABİLİM DALI

TOBB EKONOMİ VE TEKNOLOJİ ÜNİVERSİTESİ  
FEN BİLİMLERİ ENSTİTÜSÜ

ARALIK 2013

ANKARA

Fen Bilimleri Enstitü onayı

---

Prof. Dr. Necip Camuřcu  
Müdüř

Bu tezin Yüksek Lisans derecesinin tüm gereksinimlerini sağladıđını onaylarım.

---

Doç. Dr. Erdoğan DOĐDU  
Anabilim Dalı Başkanı

ERKAM UZUN tarafından hazırlanan SES DELİLLERİNİN İÇERİK ÖN-  
İNCELEMESİ VE TAHRİFİNİN TESPİTİ İÇİN YENİ YÖNTEMLER adlı bu  
tezin Yüksek Lisans tezi olarak uygun olduđunu onaylarım.

---

Yrd. Doç. Dr. Hüsrev Taha SENCAR  
Tez Danıřmanı

Tez Jüri Üyeleri

Başkan : Yrd. Doç. Dr. Tansel ÖZYER

Üye : Yrd. Doç. Dr. Hüsrev Taha SENCAR

Üye : Doç. Dr. Bülent TAVLI

## TEZ BİLDİRİMİ

Tez içindeki bütün bilgilerin etik davranış ve akademik kurallar çerçevesinde elde edilerek sunulduğunu, ayrıca tez yazım kurallarına uygun olarak hazırlanan bu çalışmada orijinal olmayan her türlü kaynağa eksiksiz atıf yapıldığını bildiririm.

Erkam UZUN

Üniversitesi : TOBB Ekonomi ve Teknoloji Üniversitesi  
Enstitüsü : Fen Bilimleri  
Anabilim Dalı : Bilgisayar Mühendisliği  
Tez Danışmanı : Yrd. Doç. Dr. Hüsrev Taha SENCAR  
Tez Türü ve Tarihi : Yüksek Lisans – Aralık 2013

Erkam UZUN

## SES DELİLLERİNİN İÇERİK ÖNİNCELEMESİ VE TAHRİFİNİN TESPİTİ İÇİN YENİ YÖNTEMLER

### ÖZET

Ses işleme teknolojisindeki gelişmeler ve ses tabanlı uygulamalardaki büyük artış, sayısal seslerden adli kanıt toplama alanında yeni yöntemler geliştirilmesi ihtiyacını da beraberinde getirmiştir. Bu amaçla geliştirilen ilk yöntem ses delillerinin ön incelemesine yöneliktir. Bir veri depolama cihazındaki seslerden konuşma içerikleri diğer seslerden ayırtırmak üzere geliştirilen yöntem, adli bilişim uzmanlarının daha efektif çalışmalarını sağlamayı hedeflemektedir. Diğer faaliyet alanı ise ses tahrifinin tespiti alanındadır. Geliştirilen yöntem ses üstünde çift-sıkıştırma ve sıklıkla kullanılan ses efektlerinin izlerini ayırtarak işlem görmüş veya tahrif edilmiş sesleri tespit etmeyi sağlamaktadır.

**Anahtar Kelimeler:** ses delillerinin ön incelemesi, konuşma ve konuşma dışı seslerin ayrıştırılması, objektif konuşma kalitesi değerlendirme, ses kodlama, ses efektleri, ses tahrif tespiti, ses içerik bütünlüğü.

**University** : **TOBB University of Economics and Technology**  
**Institute** : **Institute of Natural and Applied Sciences**  
**Science Programme** : **Computer Engineering**  
**Supervisor** : **Asst. Prof. Hüsrev Taha SENCAR**  
**Degree Awarded and Date** : **M.Sc. – December 2013**

**Erkam UZUN**

**NEW TECHNIQUES FOR PRELIMINARY ANALYSIS AND  
TAMPER DETECTION OF AUDIO EVIDENCE**

**ABSTRACT**

The advances in the digital audio processing technology and the increasing number of audio applications have led to a need for novel approaches in audio forensics. The first one concerns on the search and organization of the audio evidence. Our primary application focus is the preliminary examination of large amount of audio evidence in a storage device by clustering them into speech and non-speech classes, thereby reducing the work load of a forensic expert. We introduced a new technique that can improve the state-of-the-art in audio content classification. Another area is about authentication of audio signals. Techniques developed towards this goal aim at detecting double compressed audio and identifying traces of commonly used audio effects to ultimately distinguish between unmodified and tampered audio signals.

**Keywords:** preliminary analysis of audio evidence, speech vs non-speech discrimination, objective speech quality assessment, audio encoding, audio effects, audio tamper detection, audio content integrity.

## TEŐEKKÖR

Bu tez kapsamında yapılan alıŐmalarda gÖrüşlerini ve desteęini esirgemeyen tez danıŐmanım Yrd. Do. Dr. Hüsrev Taha SENCAR'a ve bu süreçte benden manevi desteęini esirgemeyen aileme teşekkür ederim.

# İÇİNDEKİLER

<b>1 Giriş</b>	<b>1</b>
<b>2 Objektif Konuşma Kalitesi Ölçütleri</b>	<b>4</b>
2.1 Algısal Ses Kalite Ölçütleri . . . . .	4
2.1.1 Bark Spektral Bozulması . . . . .	4
2.1.2 Güncellenmiş Bark Spektral Bozulması . . . . .	5
2.1.3 Geliştirilmiş Güncellemeli Bark Spektral Bozulması . . . . .	6
2.1.4 Ağırlıklı Eğim Spektral Mesafesi . . . . .	6
2.1.5 Normalize Blok Ölçümü . . . . .	7
2.1.6 Algısal Konuşma/Ses Kalitesi Değerlendirme . . . . .	7
2.2 Spektral Tabanlı Ölçütler . . . . .	7
2.2.1 Itakura-Saito ve COSH Mesafesi . . . . .	7
2.2.2 Kepstral Mesafe Ölçütü . . . . .	8
2.2.3 Kısa Zamanlı Fourier-Radon Dönüşümü . . . . .	8
2.2.4 Spektral Faz (Büyüklüğü) Bozulması . . . . .	8
2.2.5 Log-Olabilirlik Oranı . . . . .	9
2.2.6 Log-Alan Oranı . . . . .	9

2.3	Zaman Alanlı Ölçütler . . . . .	10
2.3.1	Sinyal-Gürültü Oranı . . . . .	10
2.3.2	Czekanowski Uzaklığı . . . . .	10
<b>3</b>	<b>Ses Delillerinde İçerik Önincelemesi</b>	<b>11</b>
3.1	Benzer Çalışmalar . . . . .	13
3.2	İçerik Sınıflandırmada Objektif Konuşma Kalitesi Ölçütlerinin Kullanımı . . . . .	16
3.2.1	Yaygın Ölçütler (CF) . . . . .	19
3.3	Deneyler . . . . .	22
3.3.1	Veri Kümesi . . . . .	22
3.3.2	Testler ve Başarım Sonuçları . . . . .	26
3.4	Sonuçlar . . . . .	35
<b>4</b>	<b>Ses Delilerinde Tahrif Tespiti</b>	<b>37</b>
4.1	Tahrif Yöntemleri . . . . .	38
4.2	Metodoloji . . . . .	39
4.3	Deneyler . . . . .	40
4.3.1	Onaylama Deneyi . . . . .	42
4.3.2	Ortak Sinyal Deneyi . . . . .	44
4.3.3	Farklı Kayıt Deneyi . . . . .	45
4.3.4	Ortak Kaynak Deneyi . . . . .	47
4.3.5	Zaman Alanlı Tahrif Testleri . . . . .	48



4.4	Sonuçlar . . . . .	49
-----	--------------------	----

<b>ÖZGEÇMİŞ</b>		<b>56</b>
-----------------	--	-----------

## ŞEKİLLERİN LİSTESİ

- 3.1 Farklı kodlayıcılar ile kodlanmış veriler üzerinde yapılan genel sınıflandırıcı testlerine ait SQF sonuçları . . . . . 31
- 3.2 Orijinal ve farklı seviyelerde kodlanmış dosyaların homojen dağılımı ile oluşturulmuş eğitim modeli (sağlam genel sınıflandırıcı) üzerinden yapılan sınıflandırma sonuçları . . . . . 32
- 4.1 Ses Tahrifi Tespiti Akış Diyagramı . . . . . 41

## ÇİZELGELERİN LİSTESİ

2.1	Objektif Konuşma Kalitesi Ölçütleri (SQF) . . . . .	5
3.1	Eski Çalışmalar . . . . .	13
3.2	Çizelge-3.1'de Verilen Öznitelikler ve Kısaltmalar . . . . .	17
3.3	Yaygın Ölçütler . . . . .	20
3.4	Konuşma ve Müzik Ayırt Etmeye Yönelik Ortak Alan Testleri . .	26
3.5	Konuşma ve Konuşma Dışı Sesleri Ayırt Etmeye Yönelik Ortak Alan Testleri . . . . .	27
3.6	Alan Dışı Test Sonuçları . . . . .	28
3.7	Genel Sınıflandırıcı Testleri . . . . .	29
3.8	Aşamalı Eğitim (AE) Seti Güncelleme Testleri . . . . .	29
3.9	Ses Efektleri Eklenmiş Dosyaların Sınıflandırılması . . . . .	33
3.10	Veri Kümeleri Üzerinden Yapılan Karşılaştırma Deneyleri . . . . .	34
3.11	Öznitelik Tabanlı Yapılan Karşılaştırma Testleri . . . . .	35
3.12	CF ve SQF için Öznitelik Hesaplama Süreleri . . . . .	35
4.1	Zaman Alanlı Tahrif Yöntemleri . . . . .	39
4.2	Efekt Tabanlı Tahrif Yöntemleri . . . . .	40

4.3	Veri Kümesi . . . . .	41
4.4	İkili Sınıflandırıcı Sonuçları . . . . .	43
4.5	Çok Sınıflı Sınıflandırıcı Sonuçları . . . . .	44
4.6	Test Sinyallerine Uygulanan Rastgele Seçilmiş Tahrif Yöntemleri .	45
4.7	Rastgele Seçilmiş Çoklu Efekt ile Tahrif Edilen Sinyallerin Tespiti	45
4.8	Bir Kişiyeye Ait Ses Sinyalinde Farklı Konuşma Kayıtları Üzerinden Tahrif Tespiti . . . . .	46
4.9	Bir Kişiyeye Ait Ses Sinyalinde Aynı Kaynak Kayıtları Üzerinden Tahrif Tespiti . . . . .	47
4.10	Zaman Alanlı Tahrif Testleri Sonuçları . . . . .	49

# 1. Giriş

Bir araştırma alanı olarak sayısal adli bilişim, dört temel sorunun çözümüne yönelik yeni yöntem ve tekniklerin geliştirilmesini hedeflemektedir. Bunlar:

- delil sayılabilecek verilerin tespiti ve elde edilmesi,
- delillerin düzenlenerek aranabilir ve incelenebilir hale getirilmesi,
- delillerin kaynağının nitelendirilmesi,
- delillerin orijinalliğinin tespitidir.

Günümüze kadar bu araştırma alanında değerlendirilebilecek çalışmalar büyük çoğunlukla iki alanda yoğunlaşmıştır. Bunların ilki ses kaynağının tespiti alanına giren konuşmacı tanıma konusundadır. Diğer alan ise ses verisinin orijinalliğinin tespitine yöneliktir. Bu alanda yapılan çalışmalar çoğunlukla seslerin gürültüden arındırılarak zenginleştirilmesi ve konuşma tanıma ile ses kayıtlarındaki anomalilerin spektral yöntemler kullanılarak tespit edilmesi konuları üzerine olmuştur. Ancak gelişen teknolojiler ve ses tabanlı uygulamalardaki artış, önceden öngörülmemiş yeni problemler ve ihtiyaçlar ortaya çıkarmıştır. Bu çalışma kapsamında yukarıda verilen problemlerden delillerin ön incelemesi ve orijinalliğinin tespitine yönelik yeni yaklaşımlar önerilmiş ve özgün yöntemler geliştirilmiştir.

Bu bağlamda yapılan ilk çalışma delil ön incelemesi alanındadır. Adli bilişim uzmanlarının günlük faaliyetlerinde karşılaştıkları en büyük zorluk yüksek hızla artan veri miktarı ile ilgilidir. Bu durum delil toplama ve analizi işleminin büyük oranda uzman merkezli, manuel yöntemlerle gerçekleşiyor olması gerçeğiyle birleştirildiğinde, bu işlemlerde uzmanlara yardımcı olabilecek otomatik hesaplama

yöntemlerinin geliştirilmesi ihtiyacını daha da acil kılmaktadır. Bu ihtiyaca yönelik olarak geliştirilen yöntem bir veri depolama cihazındaki konuşma içeren seslerin hızlı bir şekilde ayırılmasını sağlamaktadır. Bu amaçla önerilen yaklaşımın yenilikçi yani konuşma sinyallerinin kalitesinin insan algısı faktörlerine bağlı olarak nicelendirilmesi için geliştirilen objektif konuşma kalitesi ölçütlerinin (SQF), konuşma sinyalinin genel manada karakterize edilmesinde kullanılmasıdır. Farklı özelliklere sahip 10 veri kümesi üzerinde yapılan testler, tekniğin konuşma içeren sesleri %99 başarıyla ayırabildiğini gösterdiği gibi yöntemin performansının (farklı veri kümeleri üzerinde de) genellenebilir olduğu da belirlenmiştir. Mevcut çalışmalarla yapılan performans karşılaştırmaları önerilen yöntemin ses içeriğinin sınıflandırılmasında daha yüksek başarı elde ettiğini göstermektedir.

Çalışma kapsamında cevap aranan ikinci problem ise delillerin orijinalliğinin tespitidir. Seslerin kesip-kopyalanmasından, sesin bir kısmının özel bir şekilde işlenmesine kadar geniş bir tanımlı olan ses tahrifini saptamak için yeni bir yöntem önerilmiştir. Yöntemin özü tahrif sırasında uygulanması beklenen ses işleme işlemlerinin bıraktığı izlerin tespit edilmesi üzerine kuruludur. Bu amaçla ses düzenleme için yaygınlaştıran Adobe Audition yazılım aracında da var olan ses efektleri incelenerek ses işlemede sıklıkla kullanılan 21 efekt belirlenmiştir. Bu tür efektler uygulanmış sesleri, hiç bir düzenlemeye tabi tutulmamış, orijinal seslerden ayırabilmek için ise konuşma seslerinin temel niteliklerini tanımlayan objektif konuşma kalitesi ölçütlerinin yanı sıra bir sese ait fiziksel efektleri temsil eden zaman ve frekans alanlı ölçütler de kullanılarak 50 boyutlu bir öznitelik kümesi belirlenmiştir. Bu öznitelikler kullanılarak efektli ve orijinal sesleri temsil eden sınıflandırıcı modelleri oluşturulmuş ve testler yapılmıştır. İlk aşamada her bir efektin yalnız başına uygulandığı durum göz önünde tutularak efekt uygulanmış sesler tespit edilmeye çalışılmıştır. Her bir efektin %100'e yakın bir başarımla ayırt edilebildiğini gösteren test sonuçları, oluşturulan öznitelik kümesinin efekt tespitinde kullanılabileceğini göstermiştir. Birden çok efektin (3-4) uygulanabileceği pratik durumları göz önünde bulunduran tahrif testlerinin tümünde çoklu efekt uygulanmış seslerin ayırımının %90'ın üstünde bir başarımla gerçekleştirilebildiği görülmüştür. Diğer yandan gerçek hayatta karşılaşılabilecek kesme, yer değiştirme, sinyal ekleme gibi zaman alanlı tahrif senaryoları karşısında önerilen yöntemin %85 civarında bir başarımla gösterdiği gözlenmiştir.

Takip eden bölümlerde sırasıyla bu çalışmalara temel oluşturan objektif konuşma kalitesi ölçütleri tanıtılmış ve ilk araştırma konusu olan konuşma ve konuşma dışı seslerin ayırt edilmesi anlatılmıştır. Sonraki bölümlerde ise ses tahrifinin tespitine yönelik yöntem ve deney sonuçları verildikten sonra son bölümde genel bir değerlendirme yapılmıştır.

## 2. Objektif Konuşma Kalitesi Ölçütleri

Bu bölümde ses içeriği sınıflandırılmasında ve ses tahrif tespitinin belirlenmesinde kullanılacak objektif konuşma kalitesi ölçütleri sunulmuştur. Objektif konuşma kalitesi ölçütleri işlenmek üzere verilen orijinal sinyal ( $x(t)$ ) ve bu sinyalden üretilen referans sinyal ( $y(t)$ ) üzerinden hesaplanırlar. Referans sinyal üretilirken verilen orijinal sinyal gürültüden arındırılmaktadır. Bir sinyali gürültüden arındırmak için dalgacık büzülme, bağımsız bileşen analizi, maksimum olabilirlik ve ayrık dalgacık dönüşümü gibi yöntemler kullanılmaktadır. Bu çalışma kapsamında yaygın olarak kullanılan bir dalgacık tabanlı gürültüden arındırma yöntemi kullanılarak referans sinyaller üretilmiştir[15].

Her bir kalite ölçütü 20-100 ms. arasında değişen bölütler üzerinden hesaplanarak bunların ortalaması şeklinde değerlendirilmektedir. Çizelge-2.1' de verilen algısal, spektral ve zaman alanlı toplam 21 adet konuşma kalitesi ölçütlerine dair detaylı açıklama takip eden bölümlerde verilmiştir. Eşitliklerde geçen N toplam bölüt sayısını vermektedir.

### 2.1 Algısal Ses Kalite Ölçütleri

#### 2.1.1 Bark Spektral Bozulması

Bu ölçüt konuşma kalitesinin konuşma sinyaline ait, işitsel hassasiyet olarak tanımlanan, ses yüksekliği ile orantılı olduğu varsayımına dayanmaktadır. Bark



Çizelge 2.1: Objektif Konuşma Kalitesi Ölçütleri (SQF)

Ölçüt Adı	Ölçüt Sayısı	Bölüt Boyutu(ms)
<b>Algısal Ses Kalite Ölçütleri</b>		
Bark Spektral Bozulması	1	60
Güncellenmiş Bark Spektral Bozulması	1	80
Geliştirilmiş Güncellemeli Bark Spektral Bozulması	1	20
Ağırlıklı Eğim Spektral Mesafesi	1	40
Normalize Blok Ölçütü	2	60
Algısal Konuşma/Ses Kalitesi Değerlendirme	2	32
<b>Spektral Tabanlı Ölçütler</b>		
Itakura-Saito ve COSH Mesafesi	2	100
Kepstral Mesafe Ölçütü	3	20
Kısa Zamanlı Fourier-Radon Dönüşümü	1	60
Spektral Faz (Büyüklüğü) Bozulması	2	40/20
Log-Olabilirlik Oranı	1	60
Log-Alan Oranı	1	60
<b>Zaman Alanlı Ölçütler</b>		
Sinyal-Gürültü Oranı	2	20
Czekanowski Uzaklığı	1	40

spektrumu insan kulağına ait, karmaşık sinyallere karşı verilen cevaplarda, doğrusal olmayan genlik ve frekans dönüşümlerini yansıtmaktadır [53]. Bozulma değeri Eşitlik-2.1'de verilen iki Bark spektral vektörü arasındaki Öklid mesafesinin karesinin ortalaması olarak hesaplanır. Eşitlikte geçen  $C$  kritik bant sayısını,  $L_x^j(i)$  ve  $L_y^j(i)$  sırasıyla orijinal ve referans sinyaller için verilen  $i$  kritik bandında bulunan  $j$  bölütüne ait Bark spektrumlarını göstermektedir.

$$\frac{1}{N} \sum_{j=1}^N \left[ \sum_{i=1}^C (L_x^j(i) - L_y^j(i))^2 \right] \quad (2.1)$$

### 2.1.2 Güncellenmiş Bark Spektral Bozulması

Bu ölçüt hesaplanırken bir gürültü maskeleye eşiği kullanılır. Gürültü maskeleye eşiği [56]'daki gibi hesaplanır ve bozulma değeri Eşitlik- 2.2'de verildiği gibi

duyulabilen gürültü seviyesi farklarının ortalaması olarak verilir. Eşitlikte geçen  $M(i)$   $i$  kritik bandına ait bozulma indikatörü olarak kullanılmaktadır.

$$\frac{1}{N} \sum_{j=1}^N \left[ \sum_{i=1}^C M(i) |L_x^j(i) - L_y^j(i)| \right] \quad (2.2)$$

### 2.1.3 Geliştirilmiş Güncellemeli Bark Spektral Bozulması

Gürültü seviyesi farkı hesaplanırken ilk 15 Bark spektral bileşeni dikkate alınır. Burada da yine gürültü maskeleye eşiği kullanılmaktadır. Ayrıca, maskeleye sonrası etkileri dikkate alan yeni bir algısal model kullanımı benimsenmiştir [55]. Bozulma değeri Eşitlik-2.3'de verildiği gibi hesaplanmaktadır. Eşitlikte geçen  $D_{xy}^j(i)$  Bark spektralleri farkını ( $|L_x^j(i) - L_y^j(i)|$ ),  $Th(i)$  ise  $i$  kritik bandına ait  $j$  bölütündeki gürültü maskeleye eşiğini belirtmektedir.

$$\frac{1}{N} \sum_{j=1}^N \left[ \sum_{i=1}^{15} \text{Max}\{D_{xy}^j(i) - Th(i), 0\} D_{xy}^j(i) \right] \quad (2.3)$$

### 2.1.4 Ağırlıklı Eğim Spektral Mesafesi

Bu ölçüt kritik bant filtreleme bankası yoluyla düzeltilmiş ses spektrumlarını karşılaştırarak eğimler arasındaki benzerlik oranını ölçmektedir [31]. Bu ölçüt Eşitlik-2.4'de verildiği gibi 36 kritik frekans bandındaki spektral farkların karekökünün ortalaması olarak hesaplanır. Eşitlikte geçen  $\omega(i)$   $i$  kritik frekans bandına ait spektral maksimumu göstermektedir.

$$\frac{1}{N} \sum_{j=1}^N \left[ \sum_{i=1}^{36} \omega(i) (\Delta E_x(i) - \Delta E_y(i))^2 \right] \quad (2.4)$$

### 2.1.5 Normalized Blok Ölçümü

Burada iki konuşma sinyali arasındaki spektral salınım zaman ve frekans ölçeğinde hesaplanır. Konuşma sinyali öncelikle frekans ölçeğinden Bark spektral ölçeğine alınarak burada doğrusal olmayan sinyal genliği sıkıştırma işlemine tabi tutulur. Bu şekilde konuşma sinyallerine gürültü alanına yakınsayan bir dönüşüm uygulanmış olur. Burada zaman ve frekans tabanlı normalize blok ölçümü şeklinde iki öznelik hesaplanır. Zaman tabanlı ölçümde frekans ölçeğinde dönüşümü yapılmış sinyallerin zaman aralıklarındaki farkı alınırken, frekans tabanlı ölçümde zaman ölçeğinde dönüşümü yapılmış sinyallerin frekans aralıklarındaki farkı hesaplanır [51].

### 2.1.6 Algısal Konuşma/Ses Kalitesi Değerlendirme

Bu ölçüt (ITU-T Recommendation P.862 tarafından) dar bant konuşma kodlayıcıları ve telefon sistemlerini değerlendirmek için önerilmiş bir objektif ölçüttür. Bu ölçüt zaman hizalama, ses gürültüsünün algısal akustik modeli, bozulma işleme, bilişsel modelleme, frekans ve zaman alanında bozulma toplamı gibi birçok bileşeni son bozulma değerini elde etmede kullanır [44].

## 2.2 Spektral Tabanlı Ölçütler

### 2.2.1 Itakura-Saito ve COSH Mesafesi

Itakura-Saito iki sinyale ait doğrusal öngörülü kodlama (LPC) katsayıları arasındaki mesafeyi, Eşitlik-2.5'de verildiği gibi, iki spektral vektörün karelerinin log büyüklük oranları olarak hesaplar [28]. Bunun simetrik versiyonu ise COSH mesafesi olarak hesaplanır [22]. Eşitlikte geçen  $S_x^j(w)$  ve  $S_y^j(w)$  sırasıyla orijinal ve referans sinyallere ait  $j$  bölütündeki güç spektrumlarını göstermektedirler.

$$\frac{1}{N} \sum_{j=1}^N \left[ \int_{-\pi}^{\pi} \left( \log \frac{S_y^j(w)}{S_x^j(w)} + \frac{S_x^j(w)}{S_y^j(w)} - 1 \right) \frac{dw}{2\pi} \right] \quad (2.5)$$

## 2.2.2 Kepstral Mesafe Ölçütü

İki spektral arasındaki benzerliği ölçer ve Eşitlik-2.6'daki gibi iki sinyale ait LPC kepstral katsayıları arasında tanımlanır. Bu uzaklıklar tüm çerçeveler üzerinden hesaplanarak bunların ortalaması, maksimumu ve Eşitlik-2.7'deki gibi ağırlıklı ortalaması şeklinde değerlendirilir.

$$d(c_x, c_y, j) = \left[ (c_x(0) - c_y(0))^2 + 2 \sum_{k=1}^L (c_x(k) - c_y(k))^2 \right]^{1/2} \quad (2.6)$$

Bu eşitlikte bulunan  $d(c_x, c_y, j)$  değeri  $j$  bölütünde hesaplanan  $L$  uzaklığını,  $c_x(k)$  ve  $c_y(k)$  ise kepstral katsayıları göstermektedirler.

$$\left( \sum_{j=1}^N w(j) d(c_x, c_y, j) \right) / \left( \sum_{j=1}^N w(j) \right) \quad (2.7)$$

Burada geçen  $w(j)$ ,  $j$  bölütüne ait ağırlık katsayısıdır.

## 2.2.3 Kısa Zamanlı Fourier-Radon Dönüşümü

Bir sinyalin kısa zamanlı Fourier dönüşümünün (STFT) frekans alanlı izdüşümü doğrudan sinyal büyüklüğünü verirken, zaman alanlı izdüşümü o sinyalin büyüklük spektrumunu verir. Bu ölçüt ise iki sinyal için hesaplanan STFT'lerin Radon dönüşümlerinin farkının karelerinin ortalaması olarak hesaplanır [20].

## 2.2.4 Spektral Faz (Büyüklüğü) Bozulması

Eşitlik-2.8 ve Eşitlik-2.9 orijinal ve gürültüden arındırılmış sinyallerin faz ve faz büyüklüğü bozulmalarındaki farkı vermektedir [41]. ( $\lambda = 0.025$  faz ve büyüklük terimlerini birleştirmek için seçilmiştir.)

$$\frac{1}{N} \sum_{w=1}^N |\theta_x(w) - \theta_y(w)|^2 \quad (2.8)$$

$$\frac{1}{N} \left( \begin{array}{c} \lambda * \sum_{w=1}^N |\theta_x(w) - \theta_y(w)|^2 \\ + \\ (1 - \lambda) * \sum_{w=1}^N ||X(w)| - |Y(w)||^2 \end{array} \right) \quad (2.9)$$

Eşitliklerde geçen  $\omega$  ayrık frekans indeksini göstermekle beraber  $0 \leq w \leq N - 1$  aralığında yer alır. Ayrıca,  $\theta_x(w)$  ve  $\theta_y(w)$  faz spektrumlarını,  $X(w)$  ve  $Y(w)$  büyüklük spektrumlarını ifade etmektedirler.

### 2.2.5 Log-Olabilirlik Oranı

Bu ölçüt konuşma sinyali üretim mekanizmasına ait LPC modellerini kullanır ve Eşitlik-2.10'daki gibi iki sinyale ait LPC katsayılarından hesaplanır [55].

$$\frac{1}{N} \sum_{j=1}^N \left[ \log \left( \frac{\vec{\alpha}_{x,j}^T \bar{R}_{y,j} \vec{\alpha}_{x,j}}{\vec{\alpha}_{y,j}^T \bar{R}_{y,j} \vec{\alpha}_{y,j}} \right) \right] \quad (2.10)$$

Eşitlikte geçen  $\vec{\alpha}_{x,j}$  ve  $\vec{\alpha}_{y,j}$  sırasıyla orijinal ve referans sinyallerine ait  $j$  bölütündeki LPC katsayılarını ifade etmektedirler. Diğer yandan  $\bar{R}_{y,j}$  ise referans sinyaline ait otokorelasyon matrisini ifade etmektedir.

### 2.2.6 Log-Alan Oranı

Bu ölçüt insan sesini farklı kesitlerde kademeli tek tip silindirik tüpler şeklinde modellemeye dayanmaktadır ve katsayıları komşu tüplerin kesit alanları ve LPC katsayıları arasındaki log-alan oranı olarak hesaplanır. Tüm çerçeveler için hesaplanan katsayıların ortalaması öznitelik olarak değerlendirilir [13].

## 2.3 Zaman Alanlı Ölçütler

### 2.3.1 Sinyal-Gürültü Oranı

Orijinal ( $x(t)$ ) ve referans ( $y(t)$ ) sinyallerini zaman alanında örnek bazında karşılaştırır. Bölütsel sinyal-gürültü oranı ise her bir çerçevede yapılan karşılaştırmanın ortalaması olarak hesaplanır.

### 2.3.2 Czekanowski Uzaklığı

Eşitlik-2.11'de verildiği gibi  $x(t)$  ve  $y(t)$  sinyallerini zaman alanında karşılaştıran kolerasyon tabanlı bir ölçüttür.

$$\frac{1}{N} \sum_{j=1}^N \left[ \sum_{i=1}^M \left( 1 - \frac{2 * \min(x_j(i), y_j(i))}{x_j(i) + y_j(i)} \right) \right] \quad (2.11)$$

Burada geçen  $M$ ,  $j$  bölütünün büyüklüğünü,  $x_j(i)$  ve  $y_j(i)$  ise sırasıyla  $i$  bölütündeki orijinal ve referans sinyalleri göstermektedir.

## 3. Ses Delillerinde İerik Önincelemesi

Dijital adli biliřim alanı eřitli dijital araç ve elektronik depolama aygıtlarında bulunan kanıtların toplanması ve analizi ile ilgilenmektedir. Bu tür dijital aygıtların boyutları ve kompleksitesi ne yazık ki sürekli artmaktadır. Bunun bir nedeni yeni format ve araçlardan oluşan dijital ortamın hızla gelişimi olmasına rağmen esas nedeni kanıt içerebilecek veri boyutundaki eksponansiyel artıştır. Dolayısı ile dijital adli biliřim uzmanları bu derece büyük boyutlu veriyi kontrol edebilecek yenilikçi araç ve tekniklere ihtiyaç duymaktadırlar.

Adli biliřim uzmanlarının soruşturma sırasında sıklıkla karşılaştıkları önemli bir zorluk genellikle bilgisayar sistemlerinden ve mobil cihazlardan alınan kanıt niteliğindeki büyük boyutlarda verinin incelenmesi işlemidir. Bir adli biliřim uzmanına on binlerce ses dosyası içeren geniş depolama kapasitesi olan bir aygıtın (bilgisayar, akıllı telefon, tablet, harici sabit disk vs.) verildiği bir davayı göz önüne alalım. Böyle bir aygıtta yürütölen dava ile hiç bir ilgisi olmayan şarkı ve müzik dosyaları olacağı gibi görüşme ve sesli mesaj gibi mahkemeye güçlü delil olarak sunulabilecek ses dosyaları da olabilir. Böyle bir senaryoda adli biliřim uzmanı en kötü durumda verilerin hepsini kontrol edip bunların soruşturmayla alakalı olanlarını tespit etmek için çok fazla zaman harcamak zorunda kalacaktır. Adli biliřim uzmanı yalnız bu eleme işleme sürecinden sonra onaylama, açıklama ve ses tanıma işlemleri için diseksiyon ve analiz sürecine başlayabilecektir. Bu problem incelenen depolama aygıtının kapasitesindeki artışa bağılı olarak daha da kötüleşecektir. Bu yüzden ses içeriğinin otomatik olarak belirlenmesi adli biliřim uzmanlarının günlük iş yükünü önemli derecede azaltacaktır. Söz konusu inceleme bir adli soruşturma kapsamında yapılacağından

bu tarz bir otomasyon soruşturmanın tamamlanması için geçen süreyi kayda değer bir oranda azaltacaktır.

Yukarda anlatılan senaryodan hareket ederek, bu çalışmada, konuşma ve konuşma dışı sesleri ayırt etmek için ses içeriğini güvenli bir şekilde analiz eden yeni bir teknik sunulmuştur. Pratikte ses içeriği konuşma, müzik, konuşma dışı ve müzik dışı ses ve birçok farklı akustik kaynaktan gelen daha kompleks yapıda sesler olmak üzere farklı akustik ses sınıflarından oluşmaktadır. Fakat, müzik dünyada en yaygın kullanılan dijital ürünlerden birisi olduğundan tipik bir kişisel sayısal cihaz yoğun olarak müzik dosyaları içerecektir. Bu yüzden adli bilişim analizi bakış açısından konuşmanın, büyük olasılıkla müzik içeren, diğer akustik türlere göre daha ön planda olduğu ses içeriklerini ayırt edecek bir ikili sınıflandırıcı sistemi gerçekleştirmek daha çekici olacaktır.

Bir adli bilişim ortamında kullanılacak otomatik ses içeriği sınıflandırma sistemi bir dizi temel endişeleri adres etmelidir. Birincisi ve en önemlisi, böyle bir sistem farklı dil ve lehçelerdeki insan sesine ait ortak akustik-fonetik özellikleri yakalamalıdır. Aynı zamanda, böyle bir sistem yetişkin kadın ve erkek sesi hatta çocuk sesi arasındaki farklılıklara karşı hassas olmamalıdır. Ayrıca bu sistem arka plan gürültüsü içeren konuşma sinyalleri karşısında da performansını korumalıdır. Ses dosyaları genel olarak farklı formatlarda sıkıştırılmış olarak saklandığından ve transfer edildiğinden dolayı sistemin performansı farklı oranlarda değişen bit hızlarında yapılan kodlamalar karşısında değişmemelidir. Son olarak sistemin başarısı genel olarak kullanılan ses işleme teknikleri sonucu azalmamalıdır.

Yukardaki endişeleri adres etmek için, bu çalışmada, rutin olarak binlerce konuşma içeren ses dosyasını inceleyen bir adli bilişim uzmanı perspektifinden ses içeriği sınıflandırma işleminde yapılan literatürdeki çalışmalar üzerine inşa ettiğimiz objektif konuşma kalitesi ölçütlerinin kullanımını tanıttık. Bu ölçütler ilk olarak konuşma sinyallerine ait algılanan kalite problemlerinin, bozulmaları tespit ve ölçme yolu ile, teşhisinde ve değerlendirilmesinde kullanılmıştır. Buradaki yaklaşımın altında yatan fikir farklı türdeki ses sinyallerinin yapısal farklılıklarından ötürü konuşma için tasarlanmış ve rafine edilmiş kalite ölçütlerinin sistematik olarak türdeki seslere karşı farklı davranmasıdır. Dolayısı ile bu tür sistematik değişimler içerik sınıflandırmada kullanılabilirler. Bunun için önerilen metot öncelikle verilen sinyal için gürültü bastırma yöntemleri



aracılığı ile bir referans sinyali üretir. Daha sonra verilen sinyal ve üretilen referans sinyalden konuşma tanımlayıcı kalite ölçütleri çıkartılır. Yöntemin başarımı kapsamlı testler ile ölçülmüş ve literatürdeki mevcut çalışmalar ile karşılaştırılmıştır. Sonuçlar yöntemin farklı senaryo ve veri kümelerinde tutarlı bir şekilde çalıştığını göstermektedir. Deneylerde kullanılan veri kümesi genişliği ve uygulanan ses işleme teknikleri kapsamında bu çalışma bu alanda yapılan en kapsamlı çalışma olmuştur.

Takip eden bölümlerde sırası ile önce bu alanda yapılan çalışmalar anlatılmış ve daha sonra önerilen yöntemin bu çalışmalardan farklı yönlerinin üzeri çizilmiştir. Sonraki bölümde ise bu çalışmanın sayısal adli bilişim alanına uygunluğunu gösteren kapsamlı test senaryoları ve elde edilen sonuçlar anlatılmıştır.

### 3.1 Benzer Çalışmalar

Çizelge 3.1: Eski Çalışmalar

Çalışma	Öznitelikler	Veri Türü: Uzunluğu	Veri Erişimi	Kümesi	Sınıflandırma Yöntemi	Başarım (%)
1997 [46]	1,2,3,4,5,6,7,8	S:20 min; M:20 min	Açık [46]		S vs M	94.44
2001 [49]	1,9,10	S:32 min; M:32 min	Açık [49]		S vs M	82.00
2005 [5]	1,2,8,11,12,13	S:20 min; M:20 min	Açık [46]		S vs M	95.08
2006 [7]	2,8,14,15,16,17	S + M:20 hrs	Özel		S vs M	99.10
2007 [39]	1,2,3,6,7,8,18	S:1 hr; M:1 hr	Özel		S vs M	98.17
2007 [45]	1,19,20,21	S:20 min; M:20 min	Açık [46]		S vs M	98.75
2008 [48]	17,20,22,23,24	S:50 min; M:5 hrs	Açık [59]		S vs M	94.25
2008 [52]	8,25,26	S:80 min; M:80 min	Özel		S vs M	98.30
2008 [42]	1,11	S + M:12 hrs	Özel		S vs M	95.68
2009 [34]	1,2,3,6,7,8,11,27,28	S:13 hrs; M:23 hrs	Açık [3]		S vs M	98.60
2011 [54]	1,6,7,8,11,20,21,29,30,31,32	S + M + ES:8 hrs	Özel		3S vs M vs ES	94.20
2011 [19]	2,11	S:83 min; M:80 min	Açık [59]		S vs M	90.59
2011 [10]	1	S:3.9 hrs; M:3.9 hrs	Özel		S vs M	98.50
2011 [9]	1,18,23	S:114 min; M:80 min	Özel		S vs M	97.42
2011 [27]	1,2	S:3.5 hrs; SL:30 min	Özel		S vs NS	95.60
2011 [6]	6,8,11,12	S:100 min; M:102 min; N:102 min	Özel		S vs M vs N	81.90
2012 [23]	1,2,3,6,7,8,31,32	S + M + SL:2 hrs	Açık [49]		3S vs M vs SL	84.17

Seslerin otomatik olarak sınıflandırılması daha önce konuşma algılama, bilgi geri kazanımı, içerik analizi, ses kodlama, otomatik bölütleme ve müzik türü sınıflandırma gibi birçok uygulama alanı kapsamında kullanılmıştır. Esasında bu çalışmalardaki yaklaşımlar birbirlerine büyük oranda benzemektedirler ve sadece iki ana unsurda farklılık gösterirler. Birincisi, farklı türden ses dosyalarını karakterize etmek için kullanılan özniteliklerin seçimidir. Bu amaç için temel

olarak ses sinyallerine ait zamansal ve spektral karakteristikleri yakalamak için zaman, frekans ve zaman-frekans tabanlı analizler ile çeşitli öznitelikler sunulmuştur (bu öznitelikleri kısaca gözden geçirmek ve karşılaştırma yapmak için [46, 12, 38, 54]' e bakınız).Zamansal öznitelikler doğrudan sinyalin kendisinden çıkarılmaktadır ve genellikle kısa çerçeveler halinde hesaplanmaktadır. En yaygın kullanılan zaman tabanlı öznitelikler sıfır geçme oranının ve kısa zamanlı sinyal enerjisinin farkları, ortalamaları veya varyansları hakkındaki istatistiksel bilgiyi içermektedirler.

Spektral öznitelikler ise ses sinyallerinin spektral alandaki formunda yapılan hesaplamalara dayanan tanımlayıcılardır. Bu öznitelikler genellikle Kısa Zamanlı Fourier Dönüşümü (STFT), dalgacık ayrıştırma veya Doğrusal Öngörülü Kodlama (LPC) ve Mel-Frekansları Kepstrum Katsayıları (MFCC) gibi diğer model tabanlı algoritmalar ile hesaplanırlar. Spektral katsayıların yanı sıra bu öznitelikler yaygın olarak centroid, rolloff, flux, varyasyon gibi spektral özellikler ve spektral enerji dağılımı ile ilgili istatistikleri de içermektedirler. Her iki türden öznitelikleri içeren diğer bir kategori ise müzik sinyallerinde olduğu kadar konuşma içeren sinyallerde de önemli bir özellik olan perde karakteristiği ile ilgilidir. Spektral ve zamansal özniteliklerin her ikisinde doruk süresi, perde ayarlama, perde yoğunluğu ve perde izlerinin sürekliliği hakkında bilgi içeren ölçütler içermektedir. Ses içeriği sınıflandırması yapan herbir yaklaşım farklı kombinasyon ve sayıda bu özniteliklerden kullanmaktadır.

Benzer çalışmalara ait diğer ana unsur veriyi tanımlayan öznitelikleri önceden tanımlanmış etiketler ile eşleştiren sınıflandırıcı algoritmalarıdır. Birçok uygulama alanı ile ilgili olduğundan ses içeriği sınıflandırılması alanında yapılan çoğu çalışma konuşma ve müzik sinyallerini ayırt etmeye odaklanmıştır. Diğer yandan bazı çalışmalar konuşma, müzik türleri, sessizlik, gürültü, çevresel ses ve bunların birkaçının birleşmesinden oluşan farklı ses türlerini sınıflandırmaya odaklanmışlardır. Çalışmalarda farklı karmaşıklık derecesinde birçok sınıflandırma düzeni daha iyi sınıflandırma başarımı ve gerçek zamanlı uygulamalar için daha kısa sürede karar verme süresine erişebilmek amacıyla kullanılmıştır. ([54, 19, 23, 39, 45, 35] referansları farklı çalışmalarda kullanılmış farklı sınıflandırma algoritmaları hakkında daha detaylı bilgi vermektedirler.) Sınıflandırma performanslarını değerlendirmek için bu çalışmalar eğitim ve test aşamalarında genellikle kendi uygulamalarına özel kendi veri kümelerini kullanmaktadırlar.

Bu veri kümeleri genellikle yüzlerce veya binlerce kısa süreli farklı türde ses kayıtlarından veya tek bir kaynaktan elde edilmiş uzun bir kayıttan oluşmaktadır.

Özet olarak ses içerik sınıflandırılmasına yönelik başarımları yeni öznetelikler tanımlayarak, daha iyi sınıflandırma algoritmaları kullanılarak veya bunların her ikisini de yaparak arttırılabilir. Fakat tüm sınıflandırma problemlerinde olduğu gibi daha kapsamlı ve geniş öznetelik kümelerinin kullanılması içerik sınıflandırma başarımlarını arttırma kapsamında sınıflandırıcı algoritması seçimini daha iyi özneteliklerin geliştirilmesi yanında daha az önemli kılmaktadır [37]. Bu yüzden yukarıda bahsedilen iki ana unsur birbirinden isole edilebilir ve bir ses içeriği sınıflandırma algoritması kullandığı öznetelikler bakımından karakterize edilebilir. Bu bakış açısıyla Çizelge 3.1 bu alanda yapılan yüksek başarımlı ve geniş veri kümesine sahip çalışmaları kronolojik sırada özetlemektedir (Bu çizelgede verilen öznetelik numaraları ve kısaltmalar Çizelge3.2'de açıklanmıştır). (Dikkat edilmelidir ki [46] ve [49] çalışmalarında kullanılan veri kümeleri oldukça küçük veri kümeleridir, fakat bu çalışmalara ait öznetelikler takip eden çalışmalarda kullanılmıştır. Ayrıca, bunlara ait veri kümeleri açık kaynaklı olduğundan takip eden bazı çalışmalarda performans karşılaştırma testleri için kullanılmıştır.) Görüleceği üzere geniş öznetelik spektrumuna sahip bu çalışmalar test edilen veri kümeleri üzerinde %84-99 arasında değişen başarımları sahiptirler.

Genel olarak ses içeriği sınıflandırmada genel amaçlı bir kıyaslama veri kümesinin olmaması önemli bir problem olarak kabul görmektedir. Sonuç olarak farklı çalışmalara ait sınıflandırma performanslarının niceliksel karşılaştırılması çok zor olmaktadır ve çoğu çalışma öne sürdüğü yöntemi sadece kendi oluşturdukları veri kümeleleri ile elde ettikleri sonuçları rapor etmiştir. Elde edilen performans sonuçlarının genelleştirilmesi ise başka bir tartışma konusudur. Sonuçlar sunulurken çoğu çalışma eğitim ve test sınıfları için aynı veri kümesinden aldıkları örnekleri kullanmıştır. Fakat, multimedya uygulamalarında tek bir kaynaktan alınan örneklerin verilen medya türüne ait tüm çeşitleri kapsamaması ve temsil etmesi mümkün değildir. Bu bakımdan tek bir veri kaynağına bağlı kalmak yapılan analiz her ne kadar aynı veri kümesinde tutarlı ve yüksek başarımlı sonuçlar versede daha önce görmediği bir veri kümesinde zayıf performe edeceğinden çalışmaya önyargı ile yaklaşılmasına yol açacaktır. Bir adli bilişim uygulaması kapsamında düşünüldüğünde ele alınması gereken diğer bir husus öne sürülen sınıflandırıcı yönteminin konuşma sinyallerine ait

dil, lehçe ve konuşma tarzlarındaki farklılıklar ile başa çıkabilmelidir. Aynı zamanda, genel olarak kullanılan ses işleme tekniklerine, sıkıştırma yöntemlerine ve yaygın olarak uygulanan ve sese ait gürültü, perde, zaman, uzay ve tını gibi özellikleri değiştirmeyi amaçlayan ses efektlerine karşı hassasiyeti olmamalıdır. Bu alanda yapılan benzer çalışmalara baktığımızda bu unsurların göz ardı edildiği görülmektedir.

Bu çalışmada çoğunlukla konuşma sinyallerine özel yeni bir öznitelik kümesi tanıtılmıştır. Ayrıca bu ölçütleri önceki çalışmalarda kullanılan iyi tasarlanmış özniteliklerle birleştirerek konuşma içeren seslerin konuşma dışı seslerden ayrıştırılmasındaki performansını karşılaştırdık. Yukarıda anlatılan hususları adres etmek için kendi oluşturduğumuz kapsamlı bir veri kümesinin yanında önceki çalışmalarda kullanılan tüm açık kaynaklı veri kümelerini bu yöntemin başarımını ölçmek için kullandık. Ayrıca yöntemin başarımının farklı formlarda ses işleme tekniklerine karşın tutarlı olduğundan emin olmak için farklı senaryolarda testler yaptık.

## 3.2 İçerik Sınıflandırmada Objektif Konuşma Kalitesi Ölçütlerinin Kullanımı

Ses kalitesi değerlendirilmesi ses işlemeye bağlı algısal bozulmalar kapsamında ses sistemlerinin performansının ölçülmesi için önemli bir role sahiptir [36, 43, 11]. Bir araştırma konusu olarak objektif konuşma kalitesi ölçütleri subjektif dinleme testleri ile iyi korele olmuş objektif değerlendirmeler sunan istatistiksel ve ölçülebilir yöntemlerin gelişmesi ile ilgilidir. Subjektif dinleme testlerinde belirli sayıda eğitilmiş dinleyici kitlesinin ses sinyallerine ilişkin algısal skorları dikkate alınır. Günümüzde objektif ses kalite ölçütleri telefon ağlarının ve ses kodeklerinin kalitesini değerlendirmek ve ölçmek dışında farklı uygulama alanlarında da kullanılmaktadır. Özellikle, bu uygulamalar gürültü bastırma algoritmaları [25], ses kaynağı ayrıştırma algoritmaları [17], konuşma sentezleme algoritmaları [57], ses damgalama algoritmaları [1] ve ses gizli yazı algoritmaları [41] etrafında yoğunlaşmaktadırlar.

Objektif kalite değerlendirme teknikleri konuşma sinyaline veya birçok türde

Çizelge 3.2: Çizelge-3.1'de Verilen Öznitelikler ve Kısaltmalar

Tür	No	Kısaltma	Açıklama
Öznitelik	1	MFCC	Mel-Frekansı Kepstral Katsayıları
	2	ZCR	Sıfır Geçme Oranı
	3	LER	Düşük Enerji Oranı
	4	ED	Olay Yoğunluğu
	5	CM	Kepstrum Büyüklüğü
	6	SC	Spektral Ağırlık Merkezi
	7	SR	Spektral Roll-Off
	8	SF	Spectral Flux
	9	STFTC	Kısa Zamanlı Fourier Dönüşümü Katsayıları
	10	DWTC	Ayrık Dalgacık DÖnüşümü Katsayıları
	11	STE	Kısa Süreli Enerji
	12	V2W	Voice to White
	13	AL	Aktivite Seviyesi
	14	LFR	Düşük Frekans Oranı
	15	HFR	Yüksek Frekans Oranı
	16	LD	Ses Yüksekliği
	17	PD	Periyodiklik Tespiti
	18	LPC	Doğrusal Öngörülü Kodlama
	19	BR	Parlaklık
	20	PT	Perde
	21	SP	Spektral Faz Büyüklüğü
	22	RC	Yansıma Katsayıları
	23	LSF	Hat Spektral Frekansı
	24	FE	Bölüt Enerjisi
	25	ER	Enerji Oranı
	26	HSS	Harmonik Yapı Durağanlığı
	27	AC	Otokorelasyon Katsayıları
	28	SS	Spektral Dağılım
	29	APD	Ortalama Perde Yoğunluğu
	30	TPD	Tonal Güç Yoğunluğu
	31	RMS	Karekök Ortalama Enerji
	32	SE	Spektral Entropi
Veri Kümesi	1	S	Konuşma
	2	M	Müzik
	3	ES	Çevresel Ses
	4	SL	Sessizlik
	5	N	Gürültü

ses ien yksek kalitede sinyale baęlı olması bakımından iki sınıfta kategorize edilmektedirler. Bu iki tr ses arasındaki farklılık daha ok konuřma kodeklerinin dięer tr seslere ait kodlayıcılardan nasıl farklılařtıęı ile alakalıdır. Konuřma sinyalinin periyodik doęası ve sınırlı oranda sahip olduęu ses boyutu rneęin mzik gibi daha kompleks yapıda bir sinyalden daha efektif olarak temsil edilmesine olanak saęlamaktadır. Subjektif dinleme testleri ile desteklenemeyen geniř bir uygulama dizinine baęlı konuřma iřleme alanındaki geliřmeler kalite deęerlendirmesi iin birok objektif teknięin geliřmesine yol amıřtır.

Birok konuřma kalitesi deęerlendirme algoritması verilen bir sinyalin kalitesini deęerlendirmek iin bir referans sinyale ihtiya duymaktadır. Pratikte bu iřlem sinyallerin herbir kk alt paraya ayrılmıř sinyal iftleri arasındaki uzaklıklardan belirli hesaplamaların yapılmasını gerektirmektedir. Bu hesaplamalar zaman, frekans veya algısal alanlı olabilmektedir. Genel bir kalite lt ise alt paralara ait hesaplamaların ortalaması řeklinde bulunmaktadır.

Objektif konuřma kalitesi ltleri genel olarak algısal faktrleri hesaplamalarında nasıl kullandıklarına bakılarak kategorize edilebilirler [36, 11, 25, 32, 40, 14, 33, 30, 8, 24]. Zaman alanlı kalite ltleri doęrudan sinyallerin dalga formu zerinden sinyal-grlt oranı ve koralasyon gibi zellikler ıkardıklarından gereklemesi en kolay ltlerdir. Bu tarz ltlerin bir varyansıda farklı frekans bantlarının greceli nemlerini dikkate alarak bir aęırlık fonksiyonu uygulayan frekans alanlı ltlere geniřletilebilir. Geniř bir lt grubu ise hesaplamalarını iki sinyal arasındaki spektral farklılıklardan yola ıkarak yapmaktadırlar. Burada byklk spektrumlarındaki farklılıkların konuřma kalitesindeki bozulmaları yansıttıęı varsayılmaktadır. Bu ltler genel olarak konuřma sinyallerine ait doęrusal ngrl kodlama (LPC) ve kepstral temsillerine dayanmaktadırlar ve spektral farkları nasıl cezalandırdıklarına gre ayrıřmaktadırlar. Bu grupta yaygın olarak log-olabilirlik oranı, Itakura-Saito mesafesi, Kepstral mesafe gibi ltler kullanılmaktadırlar.

Objektif kalite ltlerinin subjektif konuřma kalitesi deęerlendirmelerini tam olarak yansıtmaları gerektięinden en yaygın ve geliřmiř objektif konuřma kalitesi ltleri insal iřitsel modellerini kendi tasarımları ile birleřtirirler. Bu ltler ilk olarak konuřma sinyalini bark spektrumları ve ses ykseklięi gibi psikoakustik modelleri hesaba katan algısal alanlara dnřtrrler. Daha sonra bir idrak

modülü bu algısal girdileri kalitedeki farklılıkları belirlemek için yorumlar. Bu alanda yaygın olarak kullanılan ölçütler Bark spektralleri ve birçok varyasyonlarını, spektral eğim mesafesini, normalize blok ölçümlerini (MNBs) ve standart algısal konuşma kalitesi ölçütünün (PSQM) yanı sıra algısal konuşma ve ses kalitesi değerlendirme ölçütlerini (PESQ, PEAQ) içermektedir.

Sonuç olarak objektif konuşma kalite ölçütleri geniş bir algısal konuşma nitelik kümesine sahip olduğundan konuşma içeren seslerin sınıflandırmasında da potansiyel olarak kullanılabilir. Fakat, bahsedildiği gibi bu ölçütler bir referans sinyale ihtiyaç duymaktadırlar. Göz önünde bulundurulmuş ses delillilerinin ön incelemesinde böyle bir referans sinyal söz konusu olmadığından yapay bir referans sinyal üretilmesi gerekmektedir. Her ne kadar intruzif olmayan kalite değerlendirme ölçütü sunulmuş olsa da [18, 21, 16], bunlar genel olarak iletişim ağlarından gelen konuşma kalitesini ölçmek için kullanılmaktadır. Bu tarz ölçütlerin tasarımı genel olarak ağ hatalarını göz önüne alan bozulmalar gibi özel bozulma modellerine dayanmaktadırlar. Dolayısıyla bunlar genel bir bozulma modeli için kullanılamazlar.

Bu bağlamda konuşma sinyallerinin diğer sinyallerden ayrıştırılması için objektif konuşma kalitesi ölçütleri bir referans sinyale ihtiyaç duymaktadır. Gizli yazı araştırmaları kapsamında ses kalitesinin değerlendirilmesi için referans sinyal üretme işlemi daha önce başarıyla kullanılmıştır [41]. Önerilen yöntemin anahtar noktası objektif konuşma kalite ölçütlerinin konuşma sinyalleri göz önünde bulundurulmuş olarak geliştirilmiş olması ve bunun sonucu olarak da konuşma dışı sinyaller üzerinde tutarsız değerler üreteceği ve bu değerlerin konuşma sinyallerine ait değerlerden sistematik farklılıklar göstereceğidir.

### 3.2.1 Yaygın Ölçütler (CF)

Çalışmada literatürde kullanılan ve doğrudan verilen sinyalden çıkarılan zaman ve spektral alanlı yaygın ölçütler (CF), konuşma kalite ölçütleri ile karşılaştırma yapmak için kullanılmıştır. Yaygın ölçütler doğrudan ses dalgasına yönelik basit hesaplamalar içeren özniteliklerin yanında daha gelişmiş işitsel modelleri baz alan ve sese ait dinamizm, ritim, tını, perde ve tonalite gibi karakteristik özellikleri de hesaba kata öznitelikleri içermektedir. Yaygın ölçütler ayrıca

hızlı Fourier dönüşümü ve farklı filtreleme bankaları tabanlı öznitelikleri de içermektedir. Çizelge-3.3'de çalışmada kullanılan yaygın ölçütler tanımlanmıştır. Aşağıda kısaca açıklanan bu ölçütler hakkında daha detaylı bilgi [46, 34, 54, 23, 50]'de bulunmaktadır.

Çizelge 3.3: Yaygın Ölçütler

Ölçüt Adı	Ölçüt Sayısı	Fiziksel Özellik
Karekök Ortalama Enerji	1	Dinamizm
Düşük Enerji Oranı	1	Dinamizm
Sıfır Geçme Oranı	1	Tını
Spektral Roll-Off	1	Tını
Parlaklık	1	Tını
Modalite	1	Tonalite
Spektral Ağırlık Merkezi	6	Tonalite
Olay Yoğunluğu	1	Ritim
Sinyal Tempo	1	Ritim
Darbe Netliği	1	Ritim
Inharmonicity	1	Perde
Mel-Frekansı Cepstral Katsayıları	13	Tını

### 3.2.1.1 Karekök Ortalama Enerji

Sinyale ait genel enerji, sinyalin zaman alanında her bir noktadaki karesinin ortalamasının karekökü alınarak hesaplanabilir.

### 3.2.1.2 Düşük Enerji Oranı

Ortalama enerjiden düşük enerjiye sahip çerçevelerin oranıdır. Bu oran enerji eğrisindeki karşılaştırmalı çerçevelerin tespit edilmesine yardım eder.

### 3.2.1.3 Sıfır Geçme Oranı

Sinyalin zaman alanında işaret değiştirme oranıdır.



#### **3.2.1.4 Spektral Roll-Off**

Toplam enerjinin önemli bir bölümü sinyalin yüksek frekans içeren bir oranı altındadır. Bu oran %85 ile %95 arasında değişmektedir.

#### **3.2.1.5 Parlaklık**

Kesim frekansının üzerinde kalan sinyalin enerjisidir. Parlaklık 0-1 arası bir sayı olarak ifade edilmektedir.

#### **3.2.1.6 Modalite**

Sinyalin majör veya minör olmasını -1 ve +1 arasında puanlayarak belirler. +1'e giden değerler daha majör, -1'e giden değerler daha minör olduğunu gösterir.

#### **3.2.1.7 Spektral Ağırlık Merkezi**

Perdeler arasındaki enerji dağılımlarından 6 boyutlu bir vektör hesaplar.

#### **3.2.1.8 Olay Yoğunluğu**

Saniyede sergilenen notaların sayısı gibi olayların ortalama sıklığını tahmin eder.

#### **3.2.1.9 Sinyal Tempo**

Bir müzik sinyalindeki dakikada geçen vuruş sesi olarak tanımlanabilir.

#### **3.2.1.10 Darbe Netliği**

Vuruş sesinin gücüne bakarak ritmik netliği tahmin eder.



konusmacının birbirine karışan konuşma sinyallerini içermektedir. Konuşma kümesi ayrıca yaygın olarak kullanılan ses işleme tekniklerine maruz kalmış sinyallerin yanı sıra farklı dil, lehçe, ve karakterde konuşma sinyalleri içerebilirler. Konuşma dışı sesler (NSA) ise bir çok türde sözlü ve sözsüz müzik, sadece sözlü enstrümentsiz müzik, çevresel sesler (araba, uçak, trafik, yangın, yağmur, kalabalık, alkış, kapı, hayvan, doğa sesi..gibi), tek kelimelik konuşma ve müzik dışı sesler, melodi vs. gibi ses sinyalleri içermektedir. Aşağıda kullanılan veri kümelerine ait detaylı bilgi yer almaktadır.

### **3.3.1.1 NSA-I**

- Kaynak: 640 adet müzik CD'sinden toplanmıştır
- Toplam Uzunluk: 5.56 saat
- Örnek Sayısı X Örnek Uzunluğu: 4000 X 5 sn.
- Format / Örnekleme Frekansı: WAV / 44.1 Khz
- Açıklama: 14 farklı türden alınmış farklı müzik parçaları

### **3.3.1.2 NSA-II**

- Kaynak: Marsyas [49]
- Toplam Uzunluk: 8.33 saat
- Örnek Sayısı X Örnek Uzunluğu: 1000 X 30 sn.
- Format / Örnekleme Frekansı: WAV / 22.0 Khz
- Açıklama: 10 farklı türden alınmış farklı müzik parçaları

### **3.3.1.3 NSA-III**

- Kaynak: Açık kaynak ses veri tabanı [2]
- Toplam Uzunluk: 30 dk.
- Örnek Sayısı X Örnek Uzunluğu: 500 X (1-14) sn.
- Format / Örnekleme Frekansı: WAV / 44.1 Khz
- Açıklama: Konuşma ve müzik dışı 12 farklı kategoriye ait sesler

#### 3.3.1.4 S-I

- Kaynak: VoxForge [4]
- Toplam Uzunluk: 7.78 saat
- Örnek Sayısı X Örnek Uzunluğu: 4000 X (5-13) sn.
- Format / Örnekleme Frekansı: WAV / 16.0 Khz
- Açıklama: Farklı kişiler tarafından farklı ortam ve donanımlar ile kaydedilmiş farklı içerikteki İngilizce konuşma kayıtları

#### 3.3.1.5 S-II

- Kaynak: NTimit [59]
- Toplam Uzunluk: 2.05 saat
- Örnek Sayısı X Örnek Uzunluğu: 2400 X (5-300) sn.
- Format / Örnekleme Frekansı: WAV / 16.0 Khz
- Açıklama: Farklı kişiler tarafından telefon ağı üzerinden kaydedilmiş arka plan gürültüsü içeren farklı içerikli İngilizce konuşma sinyalleri

#### 3.3.1.6 S-III

- Kaynak: VoxForge [4]
- Toplam Uzunluk: 56.2 dk.
- Örnek Sayısı X Örnek Uzunluğu: 160 X (5-68) sn.
- Format / Örnekleme Frekansı: WAV / 8-48 Khz
- Açıklama: İbranice, Almanca, Yunanca, İspanyolca, Fransızca, İtalyanca, Hollandaca, Portekizce, Rusça ve Hintçe dillerinde konuşma sinyalleri

#### 3.3.1.7 S-IV

- Kaynak: International Dialects of English Archive [3]
- Toplam Uzunluk: 58.56 saat

- Örnek Sayısı X Örnek Uzunluğu: 880 X (0.5-11) dk.
- Format / Örnekleme Frekansı: MP3 / Değişken Bit Hızı
- Açıklama: 91 farklı lehçede İngilizce kayıtlar

### 3.3.1.8 S-V

- Kaynak: YouTube
- Toplam Uzunluk: 2.94 saat
- Örnek Sayısı X Örnek Uzunluğu: 2120 X 5 sn.
- Format / Örnekleme Frekansı: AAC / 44.1 Khz
- Açıklama: 5 farklı konuşmacıya ait Türkçe kayıtlar

### 3.3.1.9 C-I

- Kaynak: Scheirer and Slaney [46]
- Toplam Uzunluk: 30 dk.
- Örnek Sayısı X Örnek Uzunluğu: 120 X 15 sn.
- Format / Örnekleme Frekansı: WAV / 22.0 Khz
- Açıklama: Radyo yayını üzerinden kaydedilmiş konuşma ve müzik sinyalleri

### 3.3.1.10 C-II

- Kaynak: Marsyas [49]
- Toplam Uzunluk: 64 dk.
- Örnek Sayısı X Örnek Uzunluğu: 128 X 30 sn.
- Format / Örnekleme Frekansı: WAV / 22.0 Khz
- Açıklama: Televizyon yayını üzerinden kaydedilmiş konuşma ve müzik sinyalleri

## 3.3.2 Testler ve Başarım Sonuçları

### 3.3.2.1 Ortak Alan Testleri

Bu bölümde önceki çalışmaların hemen hepsinde yapılan deney senaryosu kapsamında alınmış test sonuçları yer almaktadır. Ortak alan testlerinde eğitim ve test verileri aynı veri kümesi üzerinden seçilen farklı örnekleri içermektedir. Buda eğitim sınıfının test sınıfındaki verilerle karakteristik olarak aynı olduğunda erişilebilecek en yüksek başarıyı elde etmeyi sağlamaktadır. Bu amaçla NSA-I, NSA-III ve S-I veri kümeleri üzerinden farklı testler yapılmıştır.

İlk olarak önerilen yöntemin konuşma ve müzik ayırımı üzerindeki başarımı test edilmiştir. Bunun için NSA-I ve S-I veri kümeleri kullanılmıştır. NSA-I veri kümesinde bulunan 14 farklı türe ait müzik dosyaları eğitim ve test sınıflarına homojen ve eşit şekilde dağıtılmıştır. Bu test yöntemin örnekleme frekansına olan bağımlılığını ölçmek için sabit 16.0 Khz, 44.1 Khz ve karışık örnekleme frekanslarında tekrarlanmıştır. Bu testlere ait sonuçlar Çizelge-3.4'de verilmiştir. Daha sonra önerilen yöntemin konuşma ve müzik içermeyen konuşma dışı sinyalleri ayırt etmesindeki başarımı ölçmek için NSA-III ve S-I üzerinden sınıflandırma yapılmıştır. Son olarak konuşma ve konuşma dışı sinyalleri ayırt etmek için NSA-I, NSA-III ve S-I veri kümeleri üzerinden testler yapılmıştır. Bu deneylerin sonuçları Çizelge-3.5'de yer almaktadır. Yapılan tüm testlerde eğitim ve test kümelerine seçilen veri kümelerinin örtüşmeyen %50'lik bölümleri dağıtılmıştır. Çizelge-3.4 ve 3.5 de verilen sonuçlara göre aynı veri kümeleri üzerinden yapılan sınıflandırma işleminde CF ve SQF %97-99 arasında yüksek başarımla çalışmaktadır.

Çizelge 3.4: Konuşma ve Müzik Ayırt Etmeye Yönelik Ortak Alan Testleri

Örnekleme Frekans	CF (%)		SQF (%)		AF (%)	
	S	NSA	S	NSA	S	NSA
16 KHz	99.40	98.95	99.20	99.30	99.95	99.45
44.1 KHz	99.35	98.40	99.30	99.05	99.85	99.55
16-44.1 KHz	99.50	97.95	99.00	98.95	99.95	99.45

Çizelge 3.5: Konuşma ve Konuşma Dışı Sesleri Ayırt Etmeye Yönelik Ortak Alan Testleri

Veri Kümeleri	CF (%)		SQF (%)		AF (%)	
	S	NSA	S	NSA	S	NSA
NSA <sub>III</sub> vs S <sub>I</sub>	97.88	96.15	98.31	96.58	99.15	99.15
NSA <sub>I</sub> &NSA <sub>III</sub> vs S <sub>I</sub>	99.79	94.24	99.15	96.80	99.58	97.65

### 3.3.2.2 Alan Dışı Testler

Alan dışı testlerde eğitim ve test kümelerine ait veriler tamamen farklı veri kümelerinden seçilmiştir. Bu deney senaryosu sınıflandırıcının farklı veri kümeleri üzerindeki başarımını ölçmeye olanak sağlamaktadır. Bizim amacımız SQF'nin farklı dil ve lehçelere ait yapıları içermeyen bir eğitim kümesi üzerinden bu tarz verileri sınıflandırma başarımını ölçmektir. Bu amaçla ortak alan testlerinde kullanılan sınıflandırıcı modellerinden birisi (NSA-I ve S-I ile hazırlanan 16-44.1 Khz örnekleme frekansındaki model) seçilerek NSA-II, S-II, S-III ve S-IV sınıflandırılmıştır. Bu sınıflandırmalara ait sonuçlar Çizelge-3.6'de yer almaktadırlar.

Sonuçlardan görüldüğü üzere SQF en kötü sınıflandırmayı çok dilli veri kümesi (S-III) üzerinde yaparak %78.62-99.55 arasında bir başarıım gösterirken, CF sadece müzik veri kümesini (NSA-II) başarılı bir şekilde sınıflandırmış ve çok lehçeli veri kümesinde (S-IV) makul bir başarıım göstermiştir, fakat diğer veri kümeleri üzerinde başarısız olmuştur. Bunun sonucu olarak tüm özniteliklerle (AF) yapılan sınıflandırmada da yine bu veri kümeleri üzerinde başarısız olunmuştur. Ayrıca, AF S-III üzerinde CF'ye nispeten biraz daha kötü sınıflandırma yapması öznitelik seçimini veya boyut düşürmeyi gerektirmektedir. Genel olarak bakıldığında SQF her ne kadar alan dışı testlerde CF'ye nazaran daha yüksek başarıım gösterse de, konuşma içeriğinin tespitinde, daha yüksek başarımlar elde etmek için eğitim modelinde veri çeşitliliği gerekmektedir.

### 3.3.2.3 Eğitim Kümesini Genişletme

Bu bölümde daha tutarlı bir sınıflandırıcı modeli geliştirmek için iki yaklaşım önerilmektedir. Birincisi, eğitim modelini mümkün olduğunca fazla çeşitte veri

Çizelge 3.6: Alan Dışı Test Sonuçları

Öznitelik Kümesi	Accuracy (%)			
	NSA <sub>II</sub>	S <sub>II</sub>	S <sub>III</sub>	S <sub>IV</sub>
CF	97.50	08.53	55.35	81.50
SQF	99.40	95.44	78.62	99.55
AF	100.0	27.20	50.90	99.77

ile oluşturmak. İkincisi, test verisinden küçük miktarda bir veriyi, manüel olarak etiketleyip, mevcut eğitim modeline katarak yeni bir eğitim modeli oluşturmak. Takip eden testlerde bu yaklaşımlara ait sonuçlar elde edilmiştir.

**Genel Sınıflandırıcı Tasarımı** Bu bölümde her bir veri kümesi belli oranlarda eğitim ve test alt sınıflarına bölünmüş ve bu alt sınıflar birleştirilerek tek bir sınıflandırıcı ve test kümesi oluşturulmuştur. Böylelikle her seferinde bir ortak alan testi hazırlanmıştır. Her bir test senaryosunda veri kümelerinden eğitim için %10 ile %90 arasında bir oranda veri seçilmiş kalan verilerde test için ayrılmıştır. Böylelikle kabul edilebilir bir sınıflandırma için her bir veri kümesinden eğitim için seçilmesi gereken örnek oranı tespit edilmeye çalışılmıştır. Çizelge-3.7’de verilen sonuçlar CF ve SQF ile yapılan sınıflandırmada %10 ile %90 arası veri seçiminde %95-99 arası bir başarımla elde edilmiştir. Aynı zamanda, beklendiği gibi, tüm test senaryolarında AF her iki öznitelik kümesinden de daha iyi sonuç vermiştir. Sonuçlar, ayrıca, veri kümelerindeki örnekleme frekansı, sıkıştırma oranları ve uzunluk gibi özelliklerde oluşacak farklılıkların sınıflandırma başarımlarında önemli bir etkiye sahip olmadığını göstermektedir. Son olarak, her bir veri kümesinden %10-30 arasında yapılacak bir veri seçimi, eğitim modelinde yeterli miktarda çeşitliliği sağlayarak %96 üzerinde bir sınıflandırma başarımla göstermek için uygundur.

**Aşamalı Eğitim Seti Güncelleme** Önceki testler eğitim modeli test kümesindeki veri çeşitliliğini kapsadıkça başarımların artacağını göstermiştir. Fakat, gerçek hayatta sınıflandırmak için verilecek tüm verileri kapsayan bir eğitim modeli oluşturmak oldukça zordur. Bu zorluğu aşmak için ses kanıtlarına yönelik ön inceleme işleminde uygulanabilecek bir işlem de temel bir eğitim modelini aşamalı olarak test kümesinden seçilen ufak boyutta bir alt küme ile güncellemektir.



Çizelge 3.7: Genel Sınıflandırıcı Testleri

Eğitim Kümesi Oranı	CF (%)		SQF (%)		AF (%)	
	S	NSA	S	NSA	S	NSA
10%	98.08	94.37	96.55	92.26	98.80	97.05
30%	99.33	97.19	96.71	96.04	98.71	98.61
50%	98.68	97.52	97.15	96.27	99.11	98.15
70%	99.11	97.87	98.85	97.54	99.07	98.71
90%	99.24	99.33	99.36	98.66	99.62	99.50

Pratikte bir depolama aygıtından alınacak ses örnekleri genellikle dil, içerik türü, kodlama türü, kaynak ve konuşmacı gibi ortak özelliklere sahiptirler. Dolayısı ile test kümesinden seçilip manüel olarak etiketlenerek mevcut eğitim modeline katılan veriler daha yüksek başarımda bir sınıflandırma yapılmasını sağlayacaklardır.

Aşamalı eğitimin katkılarını ölçmek amacı ile alan dışı test senaryoları kurulmuştur. Genel sınıflandırıcı testlerinden elde ettiğimiz bilgiye göre herhangi bir veri kümesinden eğitime katılacak %10'luk bir alt küme sınıflandırma başarımlarını %95 gibi makul bir seviyeye çıkarmaktadır. Bu amaçla NSA-I ve S-I veri kümeleri temel eğitim sınıfı olarak seçilmiş ve alan dışı testlerde başarısız olan S-II ve S-III kümelerinden %10'luk rastgele bir altküme seçilerek eğitim kümesi güncellenip tekrar eğitilmiştir. Oluşturulan güncellenmiş eğitim modeli üzerinden test kümelerinin kalan verileri sınıflandırılmıştır. Bu sınıflandırmaya ait Çizelge-3.8'de yer alan sonuçlara göre tüm test senaryolarında sınıflandırma başarımları %96 üzerine çıkmaktadır.

Çizelge 3.8: Aşamalı Eğitim (AE) Seti Güncelleme Testleri

Öznitelik Kümesi	S-II Başarımları (%)		S-III Başarımları (%)	
	AE Öncesi	AE Sonrası	AE Öncesi	AE Sonrası
CF	08.53	97.91	55.35	97.78
SQF	95.44	99.96	78.62	96.30
AF	27.20	99.29	50.90	97.04

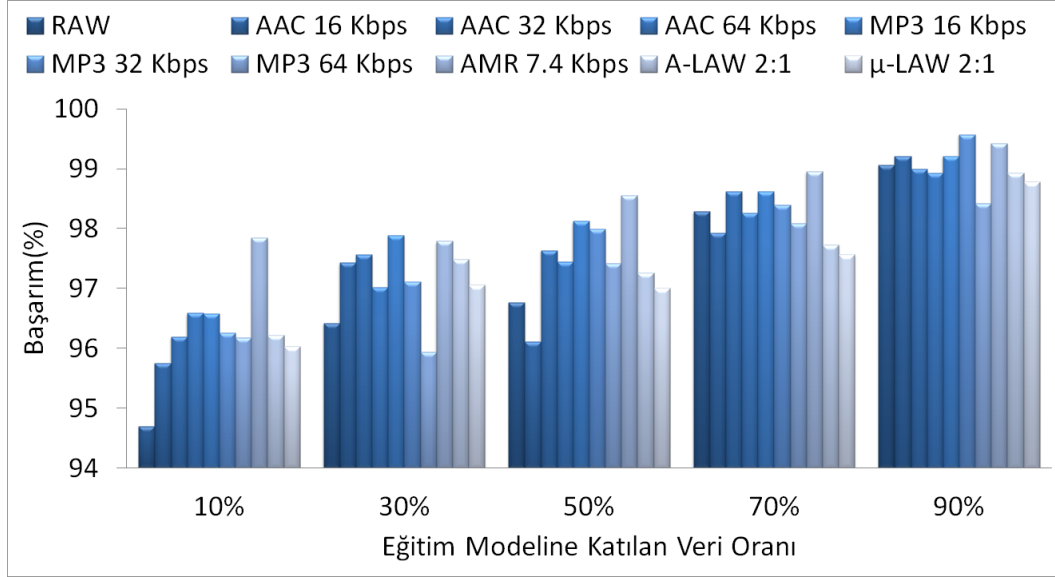
### 3.3.2.4 Sağlamlık Testleri

Önceki bölümlerde anlatılan testlerin büyük bir bölümü her hangi bir kalite kaybına uğramamış veya hiçbir ses işleme yöntemine maruz kalmamış ham veriler ile yapılmıştır. Bu bölümde yaygın olarak yapılan ses işleme işlemlerinin önerilen yöntem üzerindeki başarısı ölçülmektedir. Bu amaçla takip eden bölümlerde kodlanmış ve farklı ses efektleri eklenmiş sesler ile yapılan test sonuçları sunulmuştur.

**Ses Kodlama Testleri** Genel olarak ses dosyaları, daha verimli depolama ve veri transferi yapabilmek için, farklı kodlayıcılar ile sıkıştırılmış halde saklanırlar. Bu amaçla ses sıkıştırmanın yöntem üzerindeki etkisini ölçmek için yaygın olarak kullanılan kodlayıcılar ile kodlanmış veriler üzerinden farklı testler yapılmıştır. Bu kodlayıcılar arasında iOS tarafından kullanılan Advanced Audio Coding (AAC), MPEG 1/2 Audio Layer III (MP3), Android tarafından kullanılan Adaptive Multi-Rate (AMR ya da AMR-NB) ve telefon ağında kullanılan A-law ve  $\mu$ -law yer almaktadır. Tüm bu kodlayıcılar 7.4 – 64 Kbps aralığında düşük bit hızında kayıplı sıkıştırma yapmaktadırlar.

İlk olarak genel sınıflandırıcı testindeki aynı senaryo her bir kodlama yöntemi için ayrı ayrı çalıştırılmıştır. İlk olarak herbir veri kümesi dokuz farklı kodlama seviyesinde kodlanıp tekrar kod çözümü yapılmıştır (Deneylerde S-V veri kümesi, daha önce bilinmeyen kodlayıcılar ile kodlanmış olabileceğinden, kullanılmamıştır). Her bir kodlama türü için veri kümelerinin %10-90'ı eğitim için seçilmiş, kalanlar ise test edilmiştir. Şekil-3.1'de önerilen yöntemin eğitim ve test sınıflarının aynı kodlayıcılar ile kodlandığı, farklı eğitim oranlarında yapılan testlerde elde ettiği sonuçlar yer almaktadır. Bu sonuçlara göre tüm durumlara göre veri kümelerinin en az %10'u bile eğitilse sınıflandırma başarımı %94.7 üzerinde kalmaktadır.

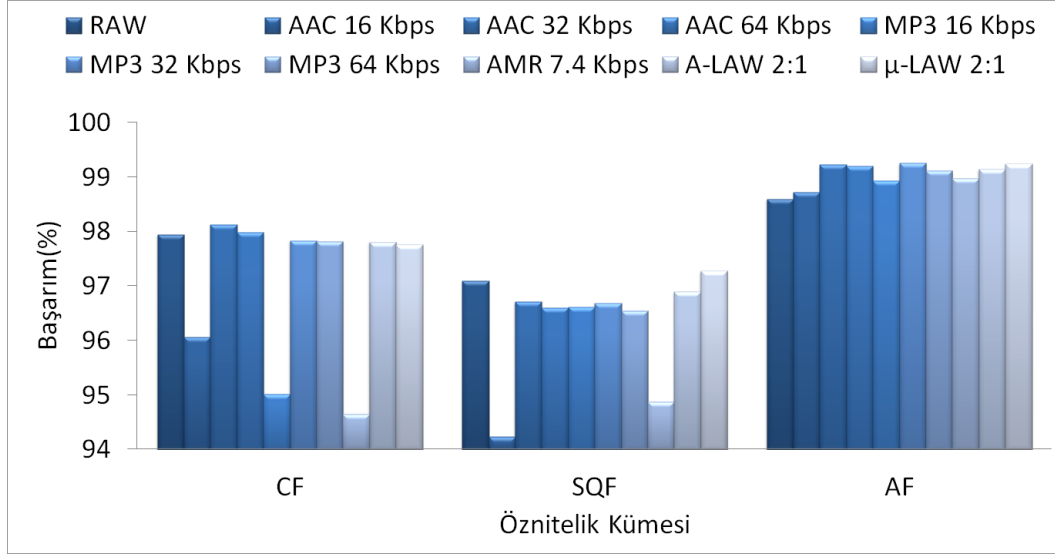
Fakat, çoğu durumda verilen bir sinyalin kodlayıcı türünü bilmek mümkün değildir. Bu amaçla ikinci kodlama test senaryosunda aynı veri kümeleri ile özel bir kodlayıcıya bağımlı olmayan bir sınıflandırıcı oluşturulmuştur. Bunun için her bir veri kümesi 9 eşit parçaya bölünerek her biri ayrı kodlama seviyesinde kodlanmış ve kod çözümü yapılmıştır. Orijinal verilerle birlikte bu 10 alt



Şekil 3.1: Farklı kodlayıcılar ile kodlanmış veriler üzerinde yapılan genel sınıflandırıcı testlerine ait SQF sonuçları

kümedeki her bir veri setinin yarısı test için ayrılmış ve her bir kodlama seviyesindeki veriler birleştirilerek 10 farklı test sınıfı oluşturulmuştur. Eğitim için ayrılan her bir veri kümesinin orijinal hali ve dokuz farklı kodlanmış hali homojen olarak seçilerek bir eğitim sınıfı oluşturulmuştur. Yani, eğitim sınıfında bir ses dosyasının bir orijinal bir de 9 türden birisi ile kodlanmış hali bulunmaktadır. Oluşturulan test kümeleri bu sınıflandırıcı üzerinden test edilmiştir ve sonuçları Şekil-3.2’de verilmiştir. Sonuçlara bakıldığında ortalama sınıflandırma oranlarının CF, SQF ve AF için sırasıyla %94.7, %94.2 ve %98.6 olduğu görülmektedir. Bu sonuçlara göre seçilen özneliliklerin herhangi bir kodlama seviyesinden etkilenmediği görülmektedir ve bu sınıflandırıcıdan bundan sonraki bölümlerde *sağlam genel sınıflandırıcı* adıyla bahsedilecektir.

**Ses Efektlerinin Uygulanması** Sınıflandırma için verilen bir ses sinyali farklı ses efektleri uygulanarak değiştirilmiş olabilir. Burada bizim amacımız önerilen yöntemin bu tarz efektlere maruz kalmış ses sinyallerini sınıflandırmadaki başarımını ölçmektir. Bir ses sinyaline ait efektler sinyal işleme tekniklerinden yola çıkarak genlik, zaman ve perde, tını, gecikme ve yankı, filtreleme, restorasyon ve çok boyutluluk gibi sınıflara ayrılırlar [58], [50]. Bu sınıflara ait 18 adet ses efekti S-V veri kümesine uygulanarak toplamda 19 adet Türkçe içerikli kodlama geçmişi bilinmeyen bir test kümesi oluşturulmuştur. Bu test kümeleri bir önceki



Şekil 3.2: Orijinal ve farklı seviyelerde kodlanmış dosyaların homojen dağılımı ile oluşturulmuş eğitim modeli (sağlam genel sınıflandırıcı) üzerinden yapılan sınıflandırma sonuçları

bölümde oluşturulan sağlam genel sınıflandırıcı üzerinden sınıflandırılarak bir alan dışı sınıflandırma senaryosu gerçekleştirilmiştir. Çizelge-3.9'de etkisiz ve etkili sinyallere ait sınıflandırma sonuçları yer almaktadır.

Test sonuçlarından takip eden önemli çıkarımları yapmak mümkündür. İlk olarak efekt eklenmemiş verilerin sınıflandırma başarımına bakıldığında, sağlam genel sınıflandırıcının önerilen yöntem ile, eğitim modelinde daha önce bilmediği kodlama hızında ve Türkçe içerikli sesler olmamasına rağmen, alan dışı sınıflandırma işleminde %99 oranında bir başarımla elde edildiği görülmektedir. Daha da önemlisi, çoğu ses efekti bu başarımla altına düşülmesine neden olmamıştır. SQF tabanlı sınıflandırma işleminde farklı efektler için %83-100 arasında ve %95.8 ortalama ile doğru sınıflandırma yapılmıştır. Fakat, CF tabanlı model kullanıldığında etkili verilerin doğru sınıflandırılmasında en az %8 seviyesinde ve ortalama %78.7 oranında bir başarımla elde edilerek orijinal verilerin sınıflandırılmasına nazaran önemli bir düşüş yaşandığı gözlenmektedir. Bunun sebebi ise CF'de bulunan bazı özniteliklerin uygulanan ses efektlerinden doğrudan etkilenmesidir.

Çizelge 3.9: Ses Efektleri Eklenmiş Dosyaların Sınıflandırılması

Efekt Grubu	Ses Efekti	Başarım (%)		
		CF	SQF	AF
Orijinal		95.91	99.29	99.51
Genlik	Yükseltme	55.77	98.18	91.63
	Sınırlama	70.63	92.75	97.03
	Soldurma	96.82	100.0	98.99
	Zarflama	95.85	99.91	99.61
Zaman ve Perde	Perde Kaydırma	97.07	99.76	99.85
	Perde Düşürme	80.53	99.88	92.07
	Perde Yükseltme	87.23	99.42	98.38
Tını	Çarpınma	94.16	99.68	99.70
	Flanger	83.16	98.38	96.07
Gecikme ve Yankı	Eko	79.27	90.86	92.88
	Gecikme	74.69	97.76	93.91
	Yankılanma	84.78	94.96	96.44
Filtreleme	Çentik Filtreleme	07.51	97.31	59.08
	Hızlı Filtreleme	56.06	83.05	83.93
	Alçak Geçiren Filtre	87.53	88.18	99.55
Restorasyon	Uyarlanabilir Gürültü Azaltma	93.90	99.10	99.36
	Histogram Azaltma	85.35	84.30	95.50
Çok Boyutlu	Ekolayzır, Yankı, Genişletme, Yükseltme	68.81	97.13	91.19

### 3.3.2.5 Performans Karşılaştırma Testleri

Önerilen yöntemin sınıflandırma performansını literatürde yapılan yöntemlerle karşılaştırmak için veri kümesi bölümünde tanıtılan tüm açık kaynak veri kümeleri kullanılmıştır. Yöntemin başarımını tam anlamıyla diğer çalışmalarla karşılaştırmak için ilgili çalışmanın test senaryosuna göre eğitim ve test kümeleri oluşturulmuştur. Çizelge-3.10'da farklı metotlar ile beş farklı veri kümesi üzerinden yapılan karşılaştırmalar ve bu çalışmalarda kullanılan eğitim-test sınıfı veri oranı dağılımı yer almaktadır. Bu sonuçlara göre önerilen yöntem tüm veri kümeleri üzerinde aynı senaryo kapsamında diğer çalışmaların her birinden daha iyi sonuç vermektedir.

Çizelge 3.10: Veri Kümeleri Üzerinden Yapılan Karşılaştırma Deneyleri

Veri Kümesi	Yöntem	Eğitim-Test Oranı	En İyi Başarım (%)
$C_I$	[46]	90-10	94.44
	[45]	90-10	98.75
	[5]	90-10	95.08
	Önerilen Yöntem	90-10	100.0
$C_{II}$	[49]	90-10	82.00
	Önerilen Yöntem	90-10	98.68
$NSA_{II}$	[23]	-	86.83
	Önerilen Yöntem	50-50	99.50
$S_{II}$	[19]	-	90.59
	[48]	99.5-0.5	94.25
	Önerilen Yöntem	90-10	100.0
$S_{IV}$	[34]	99.4-0.6	98.60
	Önerilen Yöntem	90-10	100.0

Diğer yandan, literatürde yapılan tüm çalışmaların veri kümelerine erişim olmadığından kendi veri kümeleri üzerinden sonuç bildirmiş çalışmalar ile öznitelik tabanlı bir karşılaştırma yapılmaya çalışılmıştır. Bu bağlamda, kendi veri kümesini kullanan çalışmalara ait özniteliklerin hangi seviyede CF içinde temsil edildiği bulunmuştur. Çizelge-3.11’de her bir çalışmaya ait kaç adet özniteliğin CF içinde bulunduğu ve bu çalışmalarda rapor edilen en iyi ortalama sınıflandırma başarısı ile önerilen yöntem (SQF) ile Bölüm-3.3.2.4’de oluşturulan *sağlam genel sınıflandırıcı*’ya ait tüm kodlayıcılar ile kodlanmış veriler üzerinden elde edilen ortalama sınıflandırma başarısı karşılaştırılmıştır. Bu yaklaşım her ne kadar birebir karşılaştırmayı sağlamasa da SQF’nin sınıflandırma gücünü göstermektedir. SQF için verilen sonuçlarda test için kullanılan verilerin kapsamı ve yapısı göz önünde bulundurulduğunda en kötü durumda dahi yöntemin şu ana kadar yapılan çalışmalar kadar veya daha yüksek başarımlar sağladığı görülmektedir.

### 3.3.2.6 Hesaplama Zamanı Karşılaştırması

Çizelge-3.12’de farklı uzunluklarda ve 256 Kbps veri hızındaki 100’er sesden SQF ve CF öznitelik kümelerinin çıkarılmasına ilişkin ortalama hesaplama zamanı karşılaştırılması yer almaktadır. Hesaplamanın yapıldığı bilgisayar 3.4 GHz Intel(R) i7-2600 işlemci ve 16 GB RAM taşımakta olup 64 Bit Windows 7

Çizelge 3.11: Öznitelik Tabanlı Yapılan Karşılaştırma Testleri

Yöntem	CF'nin İçerdiği Ölçütlerin Tüm Ölçütlere Oranı	Eğitim-Test Oranı (%)	Başarım (%)
2006 [7]	5/6	-	99.10
2007 [39]	5/7	-	98.17
2008 [42]	1/2	25-75	95.68
2011 [54]	10/18	37.5-50	94.20
2011 [10]	1/1	-	98.50
2011 [9]	1/3	20.7-79.3	97.42
2011 [27]	2/2	-	95.60
2011 [6]	1/4	36-50	81.90
Önerilen Yöntem		50-50	99.31

işletim sistemine sahiptir. Sonuçlardan da görüleceği gibi önerilen yöntemin hesaplama maliyeti yaygın yöntemlerinkine göre daha fazladır. Bunun nedeni önerilen yöntemde öncelikle gürültü giderme algoritmasının çalıştırılması ve özniteliklerin her iki sinyali de kullanılarak çıkarılmasıdır. Bu durum gerçek zamanlı uygulamalar için bir eksiklik sayılabilir; ancak, öznitelik çıkarımı yüksek oranda paralelleştirilebildiğinden ve adli bilişim kapsamında çevrim dışı işlemler kabul edilebilir olduğundan, bizim çalışmamızda bu noksanlık çok büyük bir öneme sahip değildir.

Çizelge 3.12: CF ve SQF için Öznitelik Hesaplama Süreleri

Veri Boyutu	3 sn	5 sn	7 sn	15 sn
CF Hesaplama Süresi (sn)	1.11	1.40	1.54	2.25
SQF Hesaplama Süresi (sn)	2.19	4.12	4.49	9.83

### 3.4 Sonuçlar

Bu çalışmada objektif konuşma kalite ölçütleri kullanılarak konuşma içeriğinin otomatik tespiti yapılmıştır. Burada öne sürülen yöntem bir depolama cihazından adli kanıt toplama işlemi için bir ön inceleme yaparak kanıt niteliği taşıyan konuşma verilerini diğer ses verilerinden ayırmaya çalışmaktadır. Bu yaklaşımın özü objektif konuşma kalite ölçütlerinin dair insan algısını göz önünde bulundurur bir biçimde konuşma sinyallerini modelleyebilmesine dayanmaktadır. Bu

bağlamda önerilen yöntem kullanılarak 88 saatlik geniş ve çeşitli bir veri kümesi üzerinde farklı senaryolarda testler yapılarak literatürde var olan çalışmalar ile kapsamlı bir karşılaştırma yapılmıştır.

Sonuçlar göstermektedir ki SQF veri kümelerinin %10-30'luk kısımları ile bile eğitilse, kalan verileri sınıflandırma başarımı %96-99 arasında olmaktadır. Daha da önemlisi, alan dışı testler SQF'nin CF'ye nazaran daha genel bir sınıflandırıcı olduğunu göstermektedir. Bu testlerin bir parçası olarak, önerilen yöntem en kötü durumda %78 başarımla gösterirken bu oran CF ile %55 seviyesinde kalmıştır. Aşamalı eğitim sınıfı güncelleme işlemi sonuçlarına bakıldığında herhangi bir veri kümesine ait %10'luk bir alt veri seti manüel olarak tasnif edilip eğitime katıldığında ise güncel eğitici ortak alan testlerindeki kadar iyi sonuç verdiği görülmektedir. Testler ayrıca sınıflandırma başarımının örnekleme frekansı veya kodlama geçmişine bağlı olmadığını göstermektedir. Ek olarak, ses efektleri ile bozulmuş ses sinyallerinin sınıflandırılmasında CF sınıflandırma işlemi başarısız olurken SQF birkaç ses efekti dışında diğer ses işleme tekniklerinden etkilenmemiştir. Bu sonuçlar kapsamında SQF histogram azaltma ve çabuk filtreleme işlemlerinde %83-84 sınıflandırma başarımı göstererek alçak geçiren filtreleme bileşenlerine bağımlı olduğunu göstermiştir.



## 4. Ses Delilerinde Tahrif Tespiti

Bir adli bilişim soruşturması sırasında uzmanlar geçerli ve değerli kanıt bulabilmek için büyük miktarlarda multimedya verisi ile karşı karşıya kalmaktadırlar. Konuşma sinyalleri bu tür soruşturmalar için önemli bir delil niteliği taşımaktadırlar. Fakat, bu sinyalleri kanıt olarak kullanabilmek için verilerin orijinalliği korunmalıdır. Bir konuşma sinyali için orijinallik içeriğin kaydedildiği şekilde korunması demektir. Bir adli soruşturma için önemli bir kısmı tahrif edilerek içeriği değiştirilmiş bir ses sinyali kanıt olarak sunulabilir. Diğer yandan ise üzerinde hiçbir oynama yapılmamış bir sesli kanıtın orijinalliği inkar edilebilir. Sonuç olarak tüm durumlarda kanıt olarak kullanılacak bir sesli sinyalin orijinalliği doğrulanmalıdır. Diğer yandan, bir sesli kanıtta içerik bütünlüğü hayati öneme sahip olduğundan ses içeriğinde yapılan en ufak bir oynama bile elde edilen kanıtın yanlış tarafın lehine kullanılmasına yol açabilir.

Günümüzde gelişmiş araçlar sayesinde insan kulağının algılayamayacağı yüksek kalitede modifikasyonlar çok kolay bir şekilde konuşma sinyallerine uygulanabilmektedir. Bu tarz araçlar konuşmacının şahsında [47] bir tahrif yapabildiği gibi konuşma içeriğinde de değişimlere imkan tanımaktadırlar [29]. Dolayısıyla sesli verinin geçerliliğinin sadece bir insan tarafından yapılması eksik veya yanlış bir sonuca varılmasına sebep olabilir. Bu yüzden, adli bilişim uzmanları bir yardımcı veya önışleme mekanizması olarak tahrif tespiti algoritmalarını kullanmalıdırlar.

Yukarıdaki senaryoda anlatılan problemde yola çıkarak bu çalışmada verilen bir konuşma sinyalinin küçük bir kısmında veya tamamında değişim olup olmadığını tespit eden yeni bir ses tahrifi tespiti algoritması sunulmuştur. Genel olarak uzun bir konuşma sinyalinin belirli bölümleri delil niteliği taşımaktadır. Bu

tür şüpheli bölümlerde uzmanlar tarafından kolayca belirlenebilmektedir. Diğer yandan, verilen kısa süreli bir sinyalin tamamı da delil niteliği taşıyabilir. Sonuç olarak şüpheli olarak nitelendirilen bir bölümün orijinalliği kesin olan sinyaller ile içerik bütünlüğü taşıyıp taşımadığı sorgulanması gereken konudur. Bu çalışmada sunulan yöntem orijinalliği kesin olan sinyaller üzerinden çıkardığı eğitim modeli ile şüpheli olan bölümlerin orijinalliğini sorgulamaktadır.

Bir konuşma sinyali dalga boyu, perde, tını, konumlama ve zamanlama gibi algısal özellikler üzerinde temel farklılıklara sahiptir. Bu algısal özelliklerin herbiri yoğunluk, temel frekans, spektral yapı, başlangıç/dengeleme zamanı ve binoral işitmede faz farklılıkları gibi ölçülebilir fiziksel özelliklerle güçlü bir korelasyona sahiptir [26]. Genel olarak bir tahrif işleminde konuşma sinyallerine bu algısal özellikler kapsamında sınıflandırılmış ses efektleri uygulanmaktadır [58, 50]. Diğer yandan, ölçekleme, sinyal silme, sinyal ekleme, sinyal değiştirme, yeniden örnekleme, yeniden niceleme, filtreleme, eko bozma, tonal modifikasyon, genlik modifikasyonu ve sinyal restorasyonu gibi kötü niyetli ataklar konuşma sinyallerine ait yukarıda bahsedilen algısal özellikler üzerinde değişime sebep olmaktadır. Dolayısı ile bu algısal özelliklerle ilintili ölçülebilir fiziksel özellikler bu tür değişimlerden etkilenmektedirler. Bu çalışmada sunulan yöntem ile bu fiziksel ve algısal özellikler üzerinde oluşacak değişimleri ölçebilen öznel özellikler kullanılarak ses tahrifi tespit edilmektedir. Bu öznel özellikler Bölüm-2’de tanımlanan objektif ses kalite ölçütlerinin yanı sıra Bölüm-3’de yaygın ölçütler olarak tanımlanan ve ses sinyallerine ait farklı fiziksel özellikleri ölçen geniş bir kümeyi kapsamaktadır.

## 4.1 Tahrif Yöntemleri

Önerilen yöntemi test etmek için farklı senaryolarda tahrif işlemleri gerçekleştirilmiştir. Bunlardan bazıları doğrudan sinyal üzerinde tahrif işlemi uygulanırken bazıları da sinyalin sahip olduğu efektleri bozmaya yönelik değişiklikler yapmaktadır.

#### 4.1.0.7 Zaman Alanlı Tahrif Yöntemleri

Bu yöntemler kapsamında konuşma sinyaline ait zaman alanlı örnekler üzerinde oynamalar yapılır. Çizelge-4.1'de gösterilen zaman alanlı tahrif yöntemleri ile sinyale ait kısımlar kendi aralarında yer değiştirilebilir, sinyalin bir parçası silinebilir, konuşma sinyaline farklı bir sinyal eklenebilir veya konuşma sinyalinin bir parçası başka bir konuşma sinyalinin aynı boyuttaki bir parçası ile değiştirilebilir.

Çizelge 4.1: Zaman Alanlı Tahrif Yöntemleri

Name	Tampering Scenario
T <sub>1</sub>	Sinyal alt parçalarının kendi içinde değiştirilmesi
T <sub>2</sub>	Sinyale ait bir parçanın silinmesi
T <sub>3</sub>	Konuşma sinyaline yeni bir parça eklenmesi
T <sub>4</sub>	Aynı konuşmacıya ait iki farklı sinyal parçasının yer değiştirilmesi

#### 4.1.0.8 Efekt Tabanlı Tahrif Yöntemleri

Literatürde bir ses sinyaline ait efektler sinyal işleme tekniklerinden yola çıkarak genlik, zaman ve perde, tını, gecikme ve yankı, filtreleme, restorasyon ve çok boyutluluk gibi sınıflara ayrılmıştır [58], [50]. Yöntemde bir ses sinyaline uygulanabilecek geniş spektrumlu bir efekt uzayı kullanılmıştır. Bu efektler Çizelge-4.2'deki gibi sınıflandırılmıştır.

## 4.2 Metodoloji

Yöntemde tahrif incelemesi yapılacak bir sinyalin aynı karakteristik özelliklere sahip herhangi bir işleme tabi tutulmadığından emin olunan eğitim sinyallerine ihtiyaç duyulmaktadır. Bu eğitim sinyali farklı yöntemlerle elde edilebilir. Bu yöntemleri şu şekilde özetleyebiliriz. İlk olarak, eğer verilen ses sinyali uzun bir kayıttan oluşuyorsa ve bu sinyalin belirli bir bölgesinde tahrif olup olmadığı araştırılıyorsa, bu sinyal üzerindeki içerik olarak önem arz etmeyen bölgeler eğitim sinyali olarak seçilir. Diğer yandan, eğer verilen sinyalin tamamında tahrif olup olmadığı araştırılıyorsa bu durumda ses sinyalinin kayıt edildiği cihaza ulaşım

Çizelge 4.2: Efekt Tabanlı Tahrif Yöntemleri

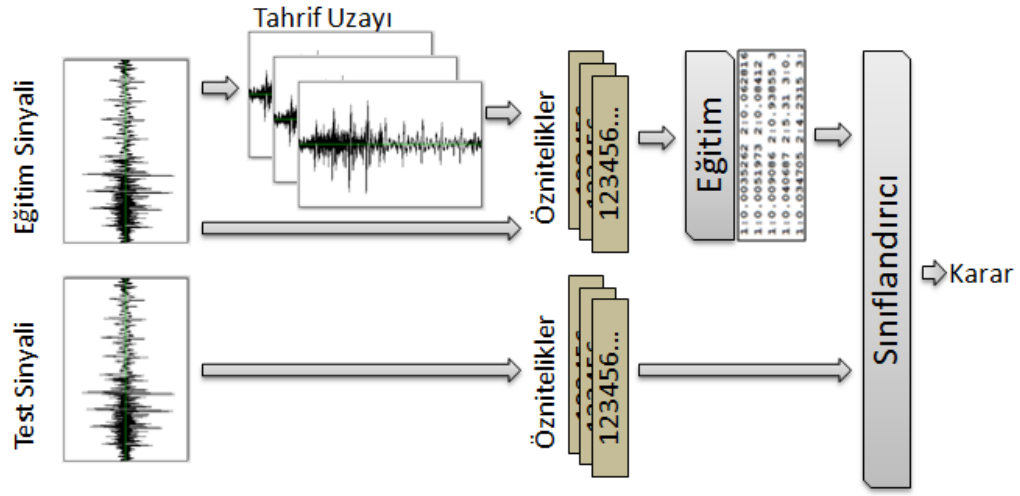
Efekt Grubu	Efekt No	Efekt Açıklaması
Genlik	E <sub>1</sub>	Yükseltme
	E <sub>2</sub>	Sınırlama
	E <sub>3</sub>	Soldurma
	E <sub>4</sub>	Zarflama
Zaman ve Perde	E <sub>5</sub>	Perde Kaydırma
	E <sub>6</sub>	Perde Düşürme
	E <sub>7</sub>	Perde Yükseltme
Tını	E <sub>8</sub>	Çarpınma
	E <sub>9</sub>	Flanger
Gecikme ve Yankı	E <sub>10</sub>	Eko
	E <sub>11</sub>	Gecikme
	E <sub>12</sub>	Yankılanma
Filtreleme	E <sub>13</sub>	Çentik Filtreleme
	E <sub>14</sub>	Hızlı Fourier Dönüşümü Filtrelemesi
	E <sub>15</sub>	Çabuk Filtreleme
	E <sub>16</sub>	Alçak Geçiren Filtre
	E <sub>17</sub>	Yüksek Geçiren Filtre
Restorasyon	E <sub>18</sub>	Uyarlanabilir Gürültü Azaltma
	E <sub>19</sub>	Histogram Azaltma
Çok Boyutlu	E <sub>20</sub>	Sıkıştırma, Filtreleme, Bozulma
	E <sub>21</sub>	Ekolayzır, Yankı, Genişletme, Yükseltme

mümkünse bu cihaz ile aynı konuşmacı tarafından bir eğitim sinyali kayıt edilir. Eğer kayıt cihazına erişim mümkün değil ise konuşmacıya ait farklı ses kayıtları eğitim sinyali olarak seçilir.

Eğitim sinyali oluşturulduktan sonra Çizelge-4.2’de verilen her bir tahrif efekti ile eğitim sinyali tahrif edilir. Daha sonra, incelenecek sinyallere ve tüm eğitim sinyallerine ait öznitelikler çıkartılır. Eğitim sinyallerine ait öznitelikler ile oluşturulan sınıflandırıcı üzerinden verilen sinyallerde tahrif olup olmadığı tespit edilir. Yönteme ait akış diyagramı Şekil-4.1’ de verilmiştir.

### 4.3 Deneyler

**Veri Kümesi** Öncelikle belirlenen özniteliklerin yöntemde kullanılan efektlerin her birini ayırt etme başarımını ölçmek için 1000 adet farklı içerikte ve farklı



Şekil 4.1: Ses Tahrifi Tespiti Akış Diyagramı

konuşmacılar tarafından kaydedilmiş İngilizce ses sinyalleri kullanılmıştır. Daha sonra, yöntem eğitim sinyallerini elde etme açısından farklı şekilde oluşturulmuş veri kümeleri üzerinden test edilmiştir. Bu amaçla ilk olarak İnternet üzerinden beş farklı konuşmacıya ait Türkçe kayıtlar elde edilmiştir. Diğer bir senaryonun gereği olarak bu beş konuşmacıdan birine ait farklı beş kayıt daha elde edilmiştir. Bir diğer senaryo gereği olarak belli bir ses kayıt cihazı ile farklı ortamlarda üç farklı içerikte ses kaydı yapılmıştır. Elde edilen tüm eğitim ve test verileri 5 sn.'lik parçalara bölünerek bu parçalar üzerinden öznitelik çıkarımı yapılmıştır. Çizelge-4.3 de veri kümelerine ait ayrıntılı bilgi verilmiştir.

Çizelge 4.3: Veri Kümesi

Veri Kümesi	Toplam Uzunluk	Örnek Sayısı	Format/ Örn.Fr.	Açıklama
$S_{vl}$ -Onaylama Seti	84 dk.	1000	WAV/ 16.0 kHz	1000 farklı konuşmacı ve içerik
$S_{si}$ -Ortak Sinyal	297 dk.	3564	AAC/ 44.1 kHz	5 farklı konuşmacı ve içerik
$S_{sp}$ -Farklı Kayıt	167 dk.	2004	AAC/ 44.1 kHz	1 konuşmacı 6 farklı kaynak, ortam ve içerik
$S_{sd}$ -Ortak Kaynak	47 dk.	564	WAV/ 44.1 kHz	1 konuşmacı aynı kaynak, 3 farklı ortam ve içerik

## 4.3.1 Onaylama Deneyi

### 4.3.1.1 İkili Sınıflandırıcılar

Bu bölümde Çizelge-4.2 de verilen her bir efektin seçilen öznelikler kullanılarak oluşturulan ikili sınıflandırıcılar ile hangi başarımla tespit edilebildiği incelenmektedir. İlk olarak, ses sinyalleri genellikle kodlanmış halde bulunduğu için Onaylama Seti'ne ait tüm konuşma sinyalleri 30 – 48 kbps arasında değişken bit oranlarında kodlanmış ve daha sonra kod çözümü yapılmıştır. Elde edilen bu Onaylama Seti'ne ait tüm dosylara yöntemde kullanılan efektler uygulanmış ve elde edilen 22 kümenin ilk yarısı eğitim, ikinci yarısı test için ayrılmıştır. Daha sonra tahrif edilmemiş eğitim kümesi ve tahrif edilmiş her bir eğitim kümesi ile 21 farklı ikili sınıflandırıcı oluşturulmuş. Aynı şekilde, tahrif edilmemiş test kümesi ve ilgili tahrif edilmiş test kümesi ilgili ikili sınıflandırıcı üzerinden test edilmiştir. Yani, eğitim ve test kümelerindeki efektler aynıdır.

Çizelge-4.4' de her bir efekt için ayrı ayrı tasarlanan sınıflandırıcıların o efekti tespit etme başarımları verilmektedir. Sonuçlara göre önerilen öznelikler ile her bir efekt ortalama %99.92 doğruluk oranı ve %0.16 oranında standart sapma ile tespit edebilmektedir. Ancak gerçek hayatta incelenmek üzere verilen bir sinyalin hangi ikili sınıflandırıcıya sorulacağı bilinemeyecektir. Bu durumda tüm efektlerin tek bir sınıf altında birleştirilmesi yoluyla işlem görmüş ve orijinal sesleri ayırt edebilen ikili bir sınıflandırıcı oluşturulabileceği gibi benzer efektlerin bir arada toplandığı çok sınıflı bir sınıflandırıcı da oluşturulabilir. Bunların da ötesinde ses içeriğinden ve kayıt işleminden kaynaklanabilecek değişimleri doğru modelleyebilmek için genel bir sınıflandırıcı oluşturmak yerine, orijinal sinyallere ait ses özelliklerinin de model oluşturulmasında kullanılması tahrif tepiti başarımının artırılmasında önemli bir yer tutacaktır. Aşağıda verilen deneyler bu doğrultuda, farklı senaryolar gözetilerek yapılmıştır.

### 4.3.1.2 Çok Sınıflı Sınıflandırıcı

Bu bölümde onaylama seti üzerinden yapılan çok sınıflı bir sınıflandırıcıya ait sonuçlar sunulmaktadır. Verilen bir sinyal için tahrif araştırması yapılacağı zaman bunu hangi ikili sınıflandırıcıya soracağımızı bilemeyiz. Bu yüzden

Çizelge 4.4: İkili Sınıflandırıcı Sonuçları

Effect #	Tespit Başarımı (%)	
	Tahrif Edilmiş	Orijinal
E <sub>1</sub>	100.0	100.0
E <sub>2</sub>	100.0	100.0
E <sub>3</sub>	100.0	100.0
E <sub>4</sub>	99.80	100.0
E <sub>5</sub>	99.60	100.0
E <sub>6</sub>	100.0	100.0
E <sub>7</sub>	99.60	100.0
E <sub>8</sub>	100.0	100.0
E <sub>9</sub>	99.80	100.0
E <sub>10</sub>	100.0	100.0
E <sub>11</sub>	100.0	100.0
E <sub>12</sub>	99.40	100.0
E <sub>13</sub>	99.80	100.0
E <sub>14</sub>	99.80	100.0
E <sub>15</sub>	100.0	100.0
E <sub>16</sub>	100.0	100.0
E <sub>17</sub>	100.0	100.0
E <sub>18</sub>	99.40	100.0
E <sub>19</sub>	100.0	100.0
E <sub>20</sub>	99.80	100.0
E <sub>21</sub>	99.80	100.0

oluşturacağımız sınıflandırıcıda tüm etkiler ile tahrif edilmiş örnekler temsil edilmelidir. Bu amaçla önceki deneyde elde edilen tüm efek uygulanmış verilerden, her bir efektin homojen olarak temsil edildiği toplam 7 adet efekt sınıfı ve bir de orijinal sınıf oluşturulmuştur. Daha sonra herbir sınıfın %50'si eğitim için seçilerek 8 sınıflı bir model oluşturulmuş ve kalan veriler bu model üzerinden sınıflandırılmıştır.

Çizelge-4.5' de bu sınıflandırma işlemine ait doğruluk matrisi yer almaktadır. Sonuçlara göre verilen bir sinyalin tahrif edilip edilmediği ve tahrif edildiyse hangi gruba ait tahrif efektleri kullanıldığı ortalama %89.25 doğruluk oranı ve %8.32 oranında standart sapma ile tespit edilebilmiştir. Daha da önemlisi eğitim modeline benzer parametrelere ait orijinal bir sinyal verildiğinde %100 başarımla bu sinyalde tahrif işlemi olmadığı tespit edilmektedir. Bu sonuçlardan yola çıkarak, amacımızın sadece tahrif olup olmadığını ve tahrif yönteminin önemsiz

olduğunu varsayarsak, tüm tahrif efektlerinin temsil edildiği bir tahrif grubu ve orijinal verilerden oluşan bir eğitim modelinin tespit işleminde yüksek başarıyı göstereceği gözlenmiştir.

Çizelge 4.5: Çok Sınıflı Sınıflandırıcı Sonuçları

	Genlik	Zaman - Perde	Tını	Gecikme-Yankı	Filtre	Restorasyon	Çok Boyutlu	Orijinal
Genlik	<b>87.8</b>	2.6	3.2	1.6	0.6	3.8	0.4	0
Zaman-Perde	3.0	<b>83.8</b>	1.2	3.2	0.6	4.4	3.8	0
Tını	7.4	2.2	<b>80.8</b>	3.0	1.0	5.6	0	0
Gecikme-Yankı	2.3	1.1	1.3	<b>94.0</b>	0.2	0.9	0.2	0
Filtre	1.6	0.2	0.4	0.4	<b>97.0</b>	0.4	0	0
Restorasyon	7.8	5.6	5.2	4.4	0.4	<b>76.6</b>	0	0
Çok Boyutlu	2.4	1.6	0.4	1.2	0	0.4	<b>94.0</b>	0
Orijinal	0	0	0	0	0	0	0	<b>100</b>

### 4.3.2 Ortak Sinyal Deneyi

Bu deney senaryosunda verilen bir sinyalin belirli bir kısmında işlem yapılmış yapılmadığı önerilen yöntem ile test edilmektedir. Bu senaryoda eğitim sinyalleri incelenecek uzun konuşma sinyalinin tahrif işlemine tabi tutulmayacak ve tahrif edilmediğinden emin olunan kısımlarından seçilmektedir. Bu amaçla Ortak Sinyal veri setinde bulunan 5 farklı konuşmacıya ait ses sinyallerinin her birinin %10'u test için seçilmiş ve kalan bölgeler eğitim sinyali olarak ayrılmıştır. Eğitim için ayrılan sinyaller yöntemde belirtilen her bir tahrif efekti ile bozularak tüm bu sinyallere ait öznelilikler çıkarılmıştır. Orijinal ve tahrif edilmiş sinyallere ait öznelilikler her bir tahrif sınıfına homojen biçimde dağılacak şekilde bir eğitim modeli oluşturulmuştur. Diğer yandan, her bir test sinyalinin tahrif edilmemiş hali ile üç farklı şekilde rastgele seçilmiş farklı efektlerle tahrif edilmiş halleri test verisi olarak hazırlanmıştır ve bunlara ait öznelilikler çıkarılmıştır. Çizelge-4.6'de her bir test sinyaline uygulanan rasgele efekt seçimleri yer almaktadır. Üretilen test verileri önerilen yöntemle göre hazırlanan eğitim modeli üzerinden sınıflandırılmıştır.

Sınıflandırma sonuçları Çizelge-4.7 daki gibidir. Bu sonuçlara göre belirli bir kısımdan şüphe edilen bir sinyalin ilgili bölgesinde tahrif işlemi yapılmış



Çizelge 4.6: Test Sinyallerine Uygulanan Rastgele Seçilmiş Tahrif Yöntemleri

Konuşmacı	Tahrif-1	Tahrif-2	Tahrif-3
$S_{si-1}$	$E_{16} + E_{10} + E_{19}$	$E_{13} + E_{21} + E_{12}$	$E_{20} + E_5 + E_{14}$
$S_{si-2}$	$E_{10} + E_{13} + E_{12}$	$E_{20} + E_{15} + E_4$	$E_2 + E_{16} + E_4$
$S_{si-3}$	$E_9 + E_{13} + E_{21}$	$E_{16} + E_8 + E_4$	$E_{21} + E_{12} + E_{19}$
$S_{si-4}$	$E_{17} + E_{21} + E_{19}$	$E_2 + E_{16} + E_{15}$	$E_5 + E_{14} + E_{21}$
$S_{si-5}$	$E_{10} + E_5 + E_{12}$	$E_{21} + E_8 + E_{11}$	$E_{18} + E_9 + E_7$

yapılmadığı ortalama %93.34 doğruluk oranı ve %7.41 oranında standart sapma ile tespit edilmektedir.

Çizelge 4.7: Rastgele Seçilmiş Çoklu Efekt ile Tahrif Edilen Sinyallerin Tespiti

Konuşmacı	Tespit Başarımı (%)			
	Orijinal	Tahrif-1	Tahrif-2	Tahrif-3
$S_{si-1}$	89.47	100.0	100.0	100.0
$S_{si-2}$	91.67	100.0	100.0	100.0
$S_{si-3}$	81.25	100.0	75.00	95.31
$S_{si-4}$	81.52	90.32	96.74	100.0
$S_{si-5}$	90.38	90.57	90.38	94.23

### 4.3.3 Farklı Kayıt Deneyi

Bu deney senaryosunda verilen sinyalin tamamında tahrif tespiti yapılacağı durum incelenmiştir. Verilen sinyalden eğitim sinyali çıkarmanın mümkün olmadığı durumda eğitim sinyalleri verilen ses kaydı ile benzerlik gösteren farklı kayıtlardan çıkarılmaktadır. Toplanacak bu farklı kayıtlar incelenecek sinyal ile kayıt ortamı, kayıt cihazı, konuşmacı, kayıt zamanı vs. gibi yönlerden olabildiğince çok ortak karakteristik taşınmalıdır. Fakat en kötü durumda bu konuşmacıya ait eski tarihli kaynağı bilinmeyen eğitim sinyallerinin toplanması söz konusu olabilir. Bu bölümde Farklı Kayıt veri kümesi eğitim sinyali olarak kullanılmıştır. Farklı Kayıt veri kümesi, Ortak Sinyal veri kümesindeki  $S_{si-2}$  konuşmacısına ait internet ortamından temin edilmiş 5 farklı ses kaydını içermektedir. Bu durumda  $S_{si-2}$  konuşmacısına ait ses kaydının orijinali, her bir efektle tahrif edilmiş halleri ve Ortak Sinyal Deneyi'nde Tahrif 1-2-3 yöntemlerindeki birden fazla efektte uğramış halleri test sinyali olarak kullanılmıştır. Eğitim ve test sinyalleri için ayrılan dosya

boyutlarını incelediğimizden yine buradaki veri setinin %90'ı eğitim %10'u test için kullanılmıştır.

Çizelge 4.8: Bir Kişiye Ait Ses Sinyalinde Farklı Konuşma Kayıtları Üzerinden Tahrif Tespiti

Efekt	Tespit Oranı (%)	Efekt	Tespit Oranı (%)
Orijinal			84.55
E <sub>1</sub>	99.52	E <sub>13</sub>	98.10
E <sub>2</sub>	96.19	E <sub>14</sub>	100.0
E <sub>3</sub>	97.62	E <sub>15</sub>	99.52
E <sub>4</sub>	66.19	E <sub>16</sub>	100.0
E <sub>5</sub>	84.76	E <sub>17</sub>	100.0
E <sub>6</sub>	100.0	E <sub>18</sub>	66.19
E <sub>7</sub>	100.0	E <sub>19</sub>	99.52
E <sub>8</sub>	80.48	E <sub>20</sub>	100.0
E <sub>9</sub>	95.71	E <sub>21</sub>	100.0
E <sub>10</sub>	99.05	E <sub>10</sub> +E <sub>13</sub> +E <sub>12</sub>	100.0
E <sub>11</sub>	99.52	E <sub>20</sub> +E <sub>15</sub> +E <sub>4</sub>	100.0
E <sub>12</sub>	96.19	E <sub>2</sub> +E <sub>16</sub> +E <sub>4</sub>	100.0

Çizelge-4.8'da farklı kayıt deneylerine ait sonuçlar yer almaktadır. Buna göre verilen bir sinyalde bir veya birden fazla tahrif işlemi yapıldığı veya herhangi bir işlem yapılmadığı ortalama %96.67 doğruluk oranı ve %9.97 oranında standart sapma ile konuşmacıya ait farklı kayıtlardan yola çıkılarak tespit edilmektedir. Sonuçlarda orijinal yani elimizdeki efekt uzayından herhangi biri ile tahrif edilmemiş dosyaların tespit edilme olasılığı %52.44, genel başarımda %76 olarak elde edilmektedir. Burada eğitim ve test verilerine ait geçmiş bilgileri bilemediğimiz için sinyallere ait kayıt ortamı ve kayıt cihazı gibi ortak yönlerin benzerliği hakkında bilgi sahibi değiliz. Fakat, eğitim verisi için kullanılan 5 farklı sinyalden test verisine kayıt ortamı olarak en çok benzeyen iki tanesi seçilerek yeni bir sınıflandırıcı hazırlandığında ortalama %94.55 doğruluk oranı ve %10.10 oranında standart sapma ile tahrif tespiti yapılmaktadır ve orijinal sinyalin tespit olasılığı %84.55'e yükselmektedir. Sonuç olarak sağlıklı bir eğitim modeli seçildiğinde verilen bir ses dosyasında tahrif işlemi yapıp yapılmadığı %94 üzerinde bir başarı oranı ile tespit edilmektedir.

#### 4.3.4 Ortak Kaynak Deneyi

Bir önceki deney senaryosunda tahrif tespiti için verilen bir sinyale ait eğitim modeli, sinyale ait kayıt cihazına erişim mümkün olmadığından, konuşmacıya ait farklı kaynaklar ile kaydedilmiş sinyallerden seçilmekteydi. Bu bölümde ise incelenmek üzere verilen sinyale ait kayıt cihazına erişimin mümkün olduğu varsayılmaktadır. Bu durumda, eğitim sinyali oluşturmak için konuşmacıya aynı cihaz ile ses kaydı yaptırılmaktadır. Bu deney senaryosu 7 dk. uzunluğunda test ve 20'ser dk. uzunluğunda farklı ortamlarda kaydedilmiş iki adet eğitim sinyali içeren Ortak Kaynak veri kümesi üzerinde gerçekleştirilmiştir. Test kümesi için ayrılan ses kaydı bir önceki deney senaryosundaki tek ve çoklu efektlerle tahrif edilerek 25 farklı test sınıfı oluşturulmuştur ve bu sınıflar aynı kayıt ile elde edilen ve önerilen yöntemle göre hazırlanan eğitim modeli üzerinden sınıflandırılmıştır.

Çizelge 4.9: Bir Kişiyeye Ait Ses Sinyalinde Aynı Kaynak Kayıtları Üzerinden Tahrif Tespiti

Efekt	Başarım (%)	Efekt	Başarım (%)
Orijinal			96.00
E <sub>1</sub>	88.00	E <sub>13</sub>	100.0
E <sub>2</sub>	97.33	E <sub>14</sub>	100.0
E <sub>3</sub>	93.33	E <sub>15</sub>	98.67
E <sub>4</sub>	78.67	E <sub>16</sub>	100.0
E <sub>5</sub>	85.33	E <sub>17</sub>	100.0
E <sub>6</sub>	100.0	E <sub>18</sub>	100.0
E <sub>7</sub>	94.67	E <sub>19</sub>	100.0
E <sub>8</sub>	96.00	E <sub>20</sub>	100.0
E <sub>9</sub>	98.67	E <sub>21</sub>	100.0
E <sub>10</sub>	97.33	E <sub>10</sub> +E <sub>13</sub> +E <sub>12</sub>	100.0
E <sub>11</sub>	93.33	E <sub>20</sub> +E <sub>15</sub> +E <sub>4</sub>	100.0
E <sub>12</sub>	100.0	E <sub>2</sub> +E <sub>16</sub> +E <sub>4</sub>	100.0

Çizelge-4.9'de ortak kaynak deneylerine ait sonuçlar yer almaktadır. Sonuçlara bakıldığında verilen bir sinyale ait kayıt cihazına erişim mümkün ise önerilen yöntemle göre bu sinyalde ortalama %96.69 doğruluk oranı ve %5.43 oranında standart sapma ile bir veya birden fazla tahrif işlemi yapıldığı veya herhangi bir işleme tabi tutulmadığı tespit edilebilmektedir.

### 4.3.5 Zaman Alanlı Tahrif Testleri

Yukarıda yapılan testler ve sonuçları önerilen yöntemin konuşma sinyallerine ait, fiziksel ölçütlerle ifade edilebilen, efektler üzerinde yapılan manipülasyonların tespit edebilebildiğini açık bir şekilde göstermektedir. Fakat gerçek hayatta en çok karşılaşılabilecek tahrif yöntemleri Çizelge-4.1'deki gibidir. Zaman alanlı bu tahrif senaryolarında yapılan değişimler ile, her ne kadar konuşma sinyallerine ait ses efektlerinde bir manipülasyon yapılmassa da, ses sinyalleri eğitim sınıfında bulunan bu 21 efektten en az biri üzerinde oynama yapılmış gibi davranmaktadırlar. Örneğin, bir konuşma sinyaline ait alt bölütlerin yerlerini değiştirmek verilen efekt gruplarından gecikme ve eko gibi bir etki oluşturmaktadır. Yani, böyle bir değişim konuşma sinyaline ait gecikme ve eko ile ilintili fiziksel ve algısal özelliği bozmaktadır. Bir diğer örnek olarak, bir konuşmacıya ait iki farklı konuşmanın belirli bölümlerinin yer değiştirilmesi sinyal arka planındaki eko ve sinyal segmentlerinin genliklerinde tutarsızlığa neden olacaktır. Dolayısı ile zaman alanlı tahrif yöntemlerini de tespit etmek için bir önceki test senaryolarında geliştirilen eğitim modellerinden birisini kullanmak makul olacaktır ve kabul edilebilir bir başarıyı gösterecektir.

Bu bağlamda, öncelikle ortak kaynak deneylerinde kullanılan  $S_{sd}$  veri kümesine ait test sınıfına zaman alanlı tahrif yöntemlerini aşağıdaki gibi uyguladık:

- $T_1$ -Karıştırma: Beş saniyelik bir sinyalin üç alt bölümü kendi arasında yer değiştirilmiştir.
- $T_2$ -Kesme: Beş saniyelik bir sinyalin iki saniyelik konuşma içeren kısmı silinmiştir.
- $T_3$ -Ekleme: Aynı kişi tarafından aynı kayıt cihazı ile farklı ortamda kaydedilmiş iki saniyelik bir bölüt ekleme.
- $T_4$ -Yer Değiştirme: Aynı kişi tarafından aynı kayıt cihazı ile farklı ortamlarda kaydedilmiş iki farklı konuşmanın iki saniyelik alt bölütlerinin yerleri değiştirilmiştir.

Oluşturulan bu tahrif edilmiş test sınıfları bir önceki bölümde oluşturulan ortak

kaynak sınıflandırıcısına ait eğitim modeli üzerinden test edilmiştir. Çizelge-4.10'de sunulan sonuçlara bakıldığında önerilen yöntemin gerçek hayatta karşılaşılabilecek bir tahrif senaryosunda %82.67 oranında bir başarımla çalıştığı görülmektedir. Bu tür oynamalar doğrudan sinyalin dalga formunda yapılsa bile bu tür manipülasyonlar yüksek bir başarımla tespit edilmektedirler.

Çizelge 4.10: Zaman Alanlı Tahrif Testleri Sonuçları

Yöntem	Açıklama	Tespit Başarımı(%)
T <sub>1</sub>	Karıştırma	69.33
T <sub>2</sub>	Kesme	100.0
T <sub>3</sub>	Ekleme	82.67
T <sub>4</sub>	Yer Değiştirme	78.67

## 4.4 Sonuçlar

Bu çalışmada verilen bir sinyali herhangi bir ses işleme tekniği uygulanarak tahrif edilip edilmediğini bulan öznelik tabanlı bir yaklaşım sunulmuştur. Bir ses sinyali üzerinde oynama yapıldığında ses özelliklerinde bazı değişimler meydana gelmesi kaçınılmazdır. Önerilen yöntem bu farklılıkları belirlenen geniş bir tahrif efekt kümesi ile modelleyerek, verilen sinyalin orijinalliği hakkında yorum yapmayı olanaklı kılmaktadır.

Test sonuçlarına göre önerilen yöntem ile her bir tahrif yöntemi orijinal verilerle oluşturulmuş ikili sınıflandırıcılar ile ortalama %99, çok sınıflı sınıflandırıcılar ile ortalama %89 oranında bir başarımla tespit edilmiştir. Buradaki test verilerinin eğitim modeli ile aynı parametrelere sahip olmasından yola çıkarak, verilen bir sinyalde tahrif işlemi tespit etmek için bu sinyalde ortak parametreler taşıyan bir eğitim modeli oluşturmak için üç farklı yaklaşım test edilmiştir. Bu yaklaşımlar verilen test sinyalinden eğitim modeli oluşturma, konuşmacıya ait farklı kayıtlardan eğitim modeli oluşturma ve aynı kayıt cihazı ile konuşmacıya ait eğitim modeli oluşturma olarak değerlendirilmiştir. Bu yaklaşımlar kapsamında yapılan test sonuçlarına göre verilen bir sinyalde %93-97 arasında bir başarımla tahrif tespiti yapılabilmektedir.

# KAYNAKLAR

- [1] *Objective perceptual audio quality measurement methods. Broadcast Technology no.35, Combined Issue Autumn 2008-Winter 2009*, 2009.
- [2] Free sound effects @ONLINE: [www.freesoundeffects.com](http://www.freesoundeffects.com), 2013.
- [3] The international dialects of english archive @ONLINE: <http://web.ku.edu/idea/index.htm>, may 2013.
- [4] Voxforge open source speech corpus @ONLINE: <http://voxforge.org/>, may 2013.
- [5] E. Alexandre-Cortizo, M. Rosa-Zurera, and F. Lopez-Ferrer. Application of fisher linear discriminant analysis to speech/music classification. In *EUROCON 2005.*, volume 2, pages 1666–1669. IEEE, 2005.
- [6] L. Álvarez, C. Llerena, and E. Alexandre. Application of neural networks to speech/music/noise classification in digital hearing aids. In *ISCGAV-11.*, pages 97–102. WSEAS, 2011.
- [7] J. G. A. Barbedo and A. Lopes. A robust and computationally efficient speech/music discriminator. *Journal of the AES*, 54(7/8):571–588, 2006.
- [8] T. P. Barnwell. Objective measures for speech quality testing. *J. Acoust. Soc. Am*, 66(6):1658–1663, 1979.
- [9] M. Barthet, S. Hargreaves, and M. Sandler. Speech/music discrimination in audio podcast using structural segmentation and timbre recognition. In *Exploring Music Contents*, pages 138–162. Springer, 2011.
- [10] H. Beigi. Audio source classification using speaker recognition techniques. *World Wide Web*, 2011.

- [11] D. Campbell, E. Jones, and M. Glavin. Audio quality assessment techniques—a review, and recent developments. *Signal Processing*, 89(8):1489–1500, 2009.
- [12] M. J. Carey, E. S. Parris, and H. Lloyd-Thomas. A comparison of features for speech, music discrimination. In *ICASSP-99.*, volume 1, pages 149–152. IEEE, 1999.
- [13] D. Chow and W. H. Abdulla. Speaker identification based on log area ratio and gaussian mixture models in narrow-band speech. In *PRICAI 2004: Trends in Artificial Intelligence*, pages 901–908. Springer, 2004.
- [14] S. Dimolitsas. Objective speech distortion measures and their relevance to speech quality assessments. In *IPCSV*, volume 136, pages 317–324. IET, 1989.
- [15] D. L. Donoho and I. M. Johnstone. Ideal denoising in an orthonormal basis chosen from a library of bases. *Comptes Rendus Acad. Sci., Ser. I*, 319, 1994.
- [16] R. K. Dubey and A. Kumar. Non-intrusive speech quality assessment using several combinations of auditory features. *IJST*, 16(1):89–101, 2013.
- [17] V. Emiya, E. Vincent, N. Harlander, and V. Hohmann. Subjective and objective quality assessment of audio source separation. *TASLP*, 19(7):2046–2057, 2011.
- [18] T. H. Falk and W.-Y. Chan. A non-intrusive quality measure of dereverberated speech. In *Proc. IWAENC*, 2008.
- [19] A. Ghosal, B. C. Dhara, and S. K. Saha. Speech/music classification using empirical mode decomposition. In *EAIT-11.*, pages 49–52. IEEE, 2011.
- [20] R. Gonzalez. Radon-based audio classification features. In *ICME-12.*, pages 556–561. IEEE, 2012.
- [21] V. Grancharov, D. Y. Zhao, J. Lindblom, and W. B. Kleijn. Low-complexity, nonintrusive speech quality assessment. *TASLP*, 14(6):1948–1956, 2006.
- [22] A. Gray Jr and J. Markel. Distance measures for speech processing. *ITASS*, 24(5):380–391, 1976.

- [23] M. A. Haque and J.-M. Kim. An analysis of content-based classification of audio signals using a fuzzy c-means algorithm. *Multimed Tools Appl*, pages 1–16, 2012.
- [24] Y. Hu and P. Loizou. Evaluation of objective measures for speech enhancement. In *Proc. Interspeech*, pages 1447–1450, 2006.
- [25] Y. Hu and P. C. Loizou. Evaluation of objective quality measures for speech enhancement. *TASLP*, 16(1):229–238, 2008.
- [26] X. Huang, A. Acero, H.-W. Hon, et al. *Spoken language processing*, volume 15. Prentice Hall PTR New Jersey, 2001.
- [27] M. Huijbregts and F. De Jong. Robust speech/non-speech classification in heterogeneous multimedia content. *Speech Communication*, 53(2):143–153, 2011.
- [28] F. Itakura and S. Saito. Analysis synthesis telephony based on the maximum likelihood method. In *ICA-68.*, volume 17. pp. C17–C20, 1968.
- [29] A. Kain and M. W. Macon. Design and evaluation of a voice conversion algorithm based on spectral envelope mapping and residual prediction. In *Acoustics, Speech, and Signal Processing, 2001. Proceedings.(ICASSP'01). 2001 IEEE International Conference on*, volume 2, pages 813–816. IEEE, 2001.
- [30] N. Kitawaki. Quality assessment of coded speech. *Advances in speech signal processing*, pages 357–385, 1992.
- [31] D. Klatt. Prediction of perceived phonetic distance from critical-band spectra: A first step. In *ICASSP-82.*, volume 7, pages 1278–1281. IEEE, 1982.
- [32] K. Kondo. *Subjective Quality Measurement of Speech: Its Evaluation, Estimation and Applications*. Springer Verlag, 2012.
- [33] R. Kubichek, D. Atkinson, and A. Webster. Advances in objective voice quality assessment. In *GLOBECOM-91.*, pages 1765–1770. IEEE, 1991.
- [34] Y. Lavner and D. Ruinskiy. A decision-tree-based algorithm for speech/music classification and segmentation. *EURASIP JASMP*, 2009:2, 2009.



- [35] C. Lim and J.-H. Chang. Adaptive kernel function of svm for improving speech/music classification of 3gpp2 smv. *ETRI Journal*, 33(6), 2011.
- [36] P. C. Loizou. Speech quality assessment. In *Multimedia Analysis, Processing and Communications*, pages 623–654. Springer, 2011.
- [37] C. D. Manning, P. Raghavan, and H. Schütze. *Introduction to information retrieval*, volume 1. Cambridge University Press Cambridge, 2008.
- [38] M. F. McKinney, J. Breebaart, et al. Features for audio and music classification. In *Proc. ISMIR*, volume 3, pages 151–158, 2003.
- [39] J. Muñoz-Expósito, S. García-Galán, N. Ruiz-Reyes, and P. Vera-Candeas. Adaptive network-based fuzzy inference system vs. other classification algorithms for warped lpc-based speech/music discrimination. *Eng Appl Artif Intel*, 20(6):783–793, 2007.
- [40] N. Nocerino, F. Soong, L. Rabiner, and D. Klatt. Comparative study of several distortion measures for speech recognition. *Speech Communication*, 4(4):317–331, 1985.
- [41] H. Ozer, I. Avcibas, B. Sankur, and N. D. Memon. Steganalysis of audio based on audio quality metrics. In *Electronic Imaging 2003*, pages 55–66. SPIE, 2003.
- [42] A. Pikrakis, T. Giannakopoulos, and S. Theodoridis. A speech/music discriminator of radio recordings based on dynamic programming and bayesian networks. *Trans. Multimedia*, 10(5):846–857, 2008.
- [43] S. R. Quackenbush, T. P. Barnwell, and M. A. Clements. *Objective measures of speech quality*. Prentice Hall Englewood Cliffs, NJ, 1988.
- [44] A. W. Rix, J. G. Beerends, M. P. Hollier, and A. P. Hekstra. Perceptual evaluation of speech quality (pesq)-a new method for speech quality assessment of telephone networks and codecs. In *Acoustics, Speech, and Signal Processing, 2001. Proceedings.(ICASSP'01). 2001 IEEE International Conference on*, volume 2, pages 749–752. IEEE, 2001.
- [45] S. O. Sadjadi, S. Ahadi, and O. Hazrati. Unsupervised speech/music classification using one-class support vector machines. In *ICICS-07.*, pages 1–5. IEEE, 2007.

- [46] E. Scheirer and M. Slaney. Construction and evaluation of a robust multifeature speech/music discriminator. In *ICASSP-97.*, volume 2, pages 1331–1334. IEEE, 1997.
- [47] M. Slaney, M. Covell, and B. Lassiter. Automatic audio morphing. In *Acoustics, Speech, and Signal Processing, 1996. ICASSP-96. Conference Proceedings., 1996 IEEE International Conference on*, volume 2, pages 1001–1004. IEEE, 1996.
- [48] J.-H. Song, K.-H. Lee, J.-H. Chang, J. K. Kim, and N. S. Kim. Analysis and improvement of speech/music classification for 3gpp2 smv based on gmm. *Signal Process. Lett.*, 15:103–106, 2008.
- [49] G. Tzanetakis, G. Essl, and P. Cook. Audio analysis using the discrete wavelet transform. In *Proc. AMTA*, 2001.
- [50] V. Verfaillie, U. Zolzer, and D. Arfib. Adaptive digital audio effects (a-dafx): A new class of sound transformations. *TASLP*, 14(5):1817–1831, 2006.
- [51] S. Voran. Objective estimation of perceived speech quality. i. development of the measuring normalizing block technique. *Trans. Speech Audio Processing*, 7(4):371–382, 1999.
- [52] J. Wang, Q. Wu, H. Deng, and Q. Yan. Real-time speech/music classification with a hierarchical oblique decision tree. In *ICASSP-08.*, pages 2033–2036. IEEE, 2008.
- [53] S. Wang, A. Sekey, and A. Gersho. An objective measure for predicting subjective quality of speech coders. *J. Select. Areas Commun.*, 10(5):819–829, 1992.
- [54] L. Xie, Z.-H. Fu, W. Feng, and Y. Luo. Pitch-density-based features and an svm binary tree approach for multi-class audio classification in broadcast news. *Multimedia systems*, 17(2):101–112, 2011.
- [55] W. Yang. *Enhanced modified bark spectral distortion (EMBSD): An objective speech quality measure based on audible distortion and cognition model*. PhD thesis, Temple University, 1999.

- [56] W. Yang, M. Dixon, and R. Yantorno. A modified bark spectral distortion measure which uses noise masking threshold. In *Speech Coding For Telecomm P*, pages 55–56. IEEE, 1997.
- [57] H. Zen, K. Tokuda, and A. W. Black. Statistical parametric speech synthesis. *Speech Communication*, 51(11):1039–1064, 2009.
- [58] U. Zölzer, X. Amatriain, and J. Wiley. *DAFX: digital audio effects*, volume 1. Wiley Online Library, 2002.
- [59] V. Zue, S. Seneff, and J. Glass. Speech database development at mit: Timit and beyond. *Speech Communication*, 9(4):351–356, 1990.

# ÖZGEÇMİŞ

## Kişisel Bilgiler

Soyadı, Adı : UZUN, Erkam  
Uyruğu : T.C.  
Doğum tarihi ve yeri : 26.02.1988 Üsküdar  
Medeni hali : Evli  
Telefon : +905544244914  
e-mail : euzun@etu.edu.tr

## Eğitim

Derece	Eğitim Birimi	Mezuniyet Tarihi
Y. Lisans	TOBB Ekonomi ve Teknoloji Üniversitesi	2013
Lisans	TOBB Ekonomi ve Teknoloji Üniversitesi	2011

## İş Deneyimi

Yıl	Yer	Görev
2011-2013	TOBB Ekonomi ve Teknoloji Üniversitesi	Yüksek Lisans

## Yabancı Dil

İngilizce (Çok iyi)

## Yayımlar

E .Uzun and H. T. Sencar, "Classification of Audios Containing Speech and Music", Signal Processing and Communications Applications Conference (SIU), 2012 20th.

S. Hicsonmez, E. Uzun, H. T. Sencar, "Methods for Identifying Traces of Compression in Audio", International Conference on Communications, Signal

Processing, and Their Applications (ICCSPA), 2013 1st.

S. Hicsonmez, E. Uzun, H. T. Sencar, "Ses Üzerindeki Kodlama İzlerinin Tespiti İçin Yöntemler" Signal Processing and Communications Applications Conference (SIU), 2013 21st.

E. Uzun, A. Aksac, O. Ozturk, H.E. Kiziloz, D. Incebacak, B. Tavli, K. Bicakci, "Network Lifetime Maximization and Localized Routing Tradeoff in Wireless Sensor Networks" Signal Processing and Communications Applications Conference (SIU), 2013 21st.

Cotuk, H.; Bicakci, K.; Tavli, B.; Uzun, E., "The Impact of Transmission Power Control Strategies on Lifetime of Wireless Sensor Networks," Computers, IEEE Transactions on , vol.PP, no.99, pp.1,1, 0

A.Aksac, E.Uzun and T.Ozyer, "A Real Time Traffic Simulator Utilizing an Adaptive Fuzzy Inference Mechanism by Tuning Fuzzy Parameters", Volume 36, Number 3 (2012), 698-720.