

KARADENİZ TEKNİK ÜNİVERSİTESİ  
FEN BİLİMLERİ ENSTİTÜSÜ

BİLGİSAYAR MÜHENDİSLİĞİ ANABİLİM DALI

SES DOSYALARINDA PASİF DOĞRULAMA YÖNTEMLERİNE DAYALI  
KOPYALA YAPIŞTIR SAHTECİLİĞİ TESPİTİ

YÜKSEK LİSANS TEZİ

Ali HATİPOĞLU

AĞUSTOS 2024

TRABZON



**KARADENİZ TEKNİK ÜNİVERSİTESİ**  
**FEN BİLİMLERİ ENSTİTÜSÜ**

**BİLGİSAYAR MÜHENDİSLİĞİ ANABİLİM DALI**

**SES DOSYALARINDA PASİF DOĞRULAMA YÖNTEMLERİNE DAYALI  
KOPYALA YAPIŞTIR SAHTECİLİĞİ TESPİTİ**

**Ali HATİPOĞLU**

**Karadeniz Teknik Üniversitesi Fen Bilimleri Enstitüsünce**  
**"BİLGİSAYAR YÜKSEK MÜHENDİSİ"**  
**Unvanı Verilmesi İçin Kabul Edilen Tezdir.**

**Tezin Enstitüye Verildiği Tarih : 08 / 08 / 2024**

**Tezin Savunma Tarihi : 23 / 08 / 2024**

**Tez Danışmanı : Dr. Beste ÜSTÜBİOĞLU**

**Trabzon 2024**

## ÖNSÖZ

Siber güvenlik kavramı günümüz teknolojilerinin sosyal ve iş hayatımızda yaygınca kullanılması ile daha da ön plana çıkmaktadır. Sesli yanıt sistemlerinin ve doğrulama sistemlerinin kullanıldığı uygulamalarda ses sinyalinin sahte olup olmadığının tespiti, adli vakalarda delil olarak kabul edilen ses dosyaları üzerinde değişiklik yapıp yapılmadığının tespiti oldukça önemli bir sorundur. “Ses Dosyalarında Pasif Doğrulama Yöntemlerine Dayalı Kopyala Yapıştır Sahteciliği Tespiti” başlıklı bu çalışmada Ses dosyalarında yapılan kopyala-yapıştır sahteciliğinin tespiti için güçlü özellik çıkarım yöntemleri kullanılarak son işlem operasyonlarına dayanıklı, yüksek doğruluk oranına sahip, bir sahtecilik tespit yöntemi geliştirilmiştir.

Yüksek lisans eğitimim ve özellikle de tez çalışması süresince yol gösteren, zaman ayıran değerli danışmanım Dr. Öğr. Üyesi Beste ÜSTÜBİOĞLU’na teşekkürlerimi sunarım. Ayrıca gerek seminer çalışması, gerek tez hazırlama sürecinde bilgi ve tecrübesinden yararlandığım ve katkılarından dolayı Prof. Dr. Öğr. Üyesi Güzin ULUTAŞ’a , eğitim hayatım süresince maddi ve manevi desteklerini hiçbir zaman esirgemeyen, bu günlere gelmemde en büyük paya sahip olan annem Aysel HATİPOĞLU , babam Remzi HATİPOĞLU kardeşlerim Muhammet HATİPOĞLU, Yakup HATİPOĞLU, Özlem SADIKLAR’a ve dostlarıma çok teşekkür eder minnettarlığımı sunarım.

Bu çalışma, Türkiye Bilimsel ve Teknolojik Araştırma Kurumu (TÜBİTAK) tarafından 122E013 numaralı proje ile desteklenmiştir. Projeye verdiği destekten ötürü TÜBİTAK’a teşekkürlerimi sunarım.

Ali HATİPOĞLU

Trabzon 2024

## TEZ ETİK BEYANNAMESİ

Yüksek lisans tezi olarak sunduğum “Ses Dosyalarında Pasif Doğrulama Yöntemlerine Dayalı Kopyala Yapıştır Sahteciliği Tespiti” başlıklı bu çalışmayı baştan sona kadar danışmanım Dr. Öğr. Üyesi Beste Üstübioğlu’nun sorumluluğunda tamamladığımı, verileri/örnekleri kendim topladığımı, deneyleri/analizleri ilgili laboratuvarlarda yaptığımı/yaptırdığımı, başka kaynaklardan aldığım bilgileri metinde ve kaynakçada eksiksiz olarak gösterdiğimi, çalışma sürecinde bilimsel araştırma ve etik kurallara uygun olarak davrandığımı ve aksinin ortaya çıkması durumunda her türlü yasal sonucu kabul ettiğimi beyan ederim. 23/08/2024

Ali HATİPOĞLU

## İÇİNDEKİLER

|                                                                                                                              | <u>Sayfa No</u> |
|------------------------------------------------------------------------------------------------------------------------------|-----------------|
| ÖNSÖZ .....                                                                                                                  | III             |
| TEZ ETİK BEYANNAMESİ.....                                                                                                    | IV              |
| İÇİNDEKİLER.....                                                                                                             | V               |
| ÖZET .....                                                                                                                   | VII             |
| SUMMARY .....                                                                                                                | VIII            |
| ŞEKİLLER DİZİNİ .....                                                                                                        | IX              |
| TABLolar DİZİNİ .....                                                                                                        | X               |
| SEMBOLLER VE KISALTMALAR DİZİNİ .....                                                                                        | XII             |
| 1. GENEL BİLGİLER.....                                                                                                       | 1               |
| 1.1. Giriş .....                                                                                                             | 1               |
| 1.2. Ses Dosyasında Yapılan Sahtecilikler .....                                                                              | 5               |
| 1.3. Tezin Kapsamı.....                                                                                                      | 10              |
| 1.4. Kopyala Yapıştır Sahteciliği Tespiti .....                                                                              | 10              |
| 1.5. Literatür Araştırması.....                                                                                              | 10              |
| 1.5.1. Ses Etkinliği Algılamaya Dayalı Kopyala-Yapıştır Sahteciliği Tespiti<br>Yöntemleri .....                              | 11              |
| 1.5.2. Çerçeve Tabanlı Kopyala-Yapıştır Sahteciliği Tespiti Yöntemleri .....                                                 | 15              |
| 1.5.3. Görüntü Tabanlı Kopyala-Yapıştır Sahteciliği Tespiti Yöntemleri .....                                                 | 16              |
| 1.6. Kopyala-Yapıştır Sahteciliği Tespiti Aşamaları .....                                                                    | 18              |
| 1.6.1. Sesli Segmentlerin Tespiti .....                                                                                      | 18              |
| 1.6.2. Özellik Çıkarma .....                                                                                                 | 21              |
| 1.6.3. Benzerlik Hesaplama .....                                                                                             | 31              |
| 1.6.4. Sahte Segmentlerin İşaretlenmesi .....                                                                                | 33              |
| 2. YAPILAN ÇALIŞMALAR, BULGULAR VE İRDELEME.....                                                                             | 35              |
| 2.1. Veritabanı.....                                                                                                         | 36              |
| 2.2. Kullanılan Ölçütler .....                                                                                               | 37              |
| 2.3. Sesli Kelimelerin Tespiti .....                                                                                         | 38              |
| 2.4. Kelimelerden Özellik Çıkarması .....                                                                                    | 41              |
| 2.4.1. PLP Yöntemi ile Kopyala – Yapıştır Sahteciliği Tespitinin Değerlendirilmesi<br>ve Elde Edilen Deneysel Sonuçlar ..... | 41              |

|                                                                                                                                |    |
|--------------------------------------------------------------------------------------------------------------------------------|----|
| 2.4.2. Rasta-PLP Yöntemi ile Kopyala – Yapıştır Sahteciliği Tespitinin Değerlendirilmesi ve Elde Edilen Deneysel Sonuçlar..... | 44 |
| 2.4.3. Spectral Yöntemi ile Kopyala – Yapıştır Sahteciliği Tespitinin Değerlendirilmesi ve Elde Edilen Deneysel Sonuçlar.....  | 46 |
| 2.4.4. MFCC Yöntemi ile Kopyala – Yapıştır Sahteciliği Tespitinin Değerlendirilmesi ve Elde Edilen Deneysel Sonuçlar.....      | 48 |
| 2.5. Benzerlik Hesaplaması .....                                                                                               | 51 |
| 2.5.1. PLP Yöntemi ile Kopyala – Yapıştır Sahteciliği Tespitinin Değerlendirilmesi ve Elde Edilen Deneysel Sonuçlar .....      | 51 |
| 2.5.2. Rasta-PLP Yöntemi ile Kopyala – Yapıştır Sahteciliği Tespitinin Değerlendirilmesi ve Elde Edilen Deneysel Sonuçlar..... | 55 |
| 2.5.3. Sepctral Yöntemi ile Kopyala – Yapıştır Sahteciliği Tespitinin Değerlendirilmesi ve Elde Edilen Deneysel Sonuçlar.....  | 59 |
| 2.5.4. MFCC Yöntemi ile Kopyala – Yapıştır Sahteciliği Tespitinin Değerlendirilmesi ve Elde Edilen Deneysel Sonuçlar.....      | 63 |
| 2.6. Sahte Segmentlerin Belirlenmesi .....                                                                                     | 67 |
| 2.7. PLP Yönteminin Literatürdeki Yöntemlerle Karşılaştırılması .....                                                          | 69 |
| 3. SONUÇLAR .....                                                                                                              | 74 |
| 4. ÖNERİLER.....                                                                                                               | 77 |
| 5. KAYNAKLAR .....                                                                                                             | 78 |
| ÖZGEÇMİŞ                                                                                                                       |    |

## ÖZET

### SES DOSYALARINDA PASİF DOĞRULAMA YÖNTEMLERİNE DAYALI KOPYALA YAPIŞTIR SAHTECİLİĞİ TESPİTİ

Ali HATİPOĞLU

Karadeniz Teknik Üniversitesi  
Fen Bilimleri Enstitüsü  
Bilgisayar Mühendisliği Anabilim Dalı  
Danışman: Dr. Öğr. Üyesi Beste ÜSTÜBİOĞLU  
2024, 81 Pages

Günümüzde ses dosyalarına yönelik saldırılar, özellikle düzenleme araçlarının sayısındaki artış ve kullanımlarındaki kolaylık sayesinde kötü niyetli kullanıcılar tarafından oldukça fazla yapılabilir hale gelmiştir. Adli ortamlarda delil olarak sunulan ses dosyalarının doğruluklarının ispatlanabilmesi büyük önem arz etmektedir. Sayısal ses doğruluğunu tespit etmek için literatürde aktif ve pasif doğrulama yöntemleri olmak üzere iki ana kategoride yöntemler önerilmiştir. Pasif yöntemlerde, sesin doğrulanması işleminde sesten elde edilen istatistiksel özellikler kullanılarak ses doğrulaması gerçekleştirilmektedir. Tez kapsamında geliştirilen pasif doğrulama yöntemlerine dayalı bu çalışma ile ses dosyalarındaki kopyala yapıştır sahteciliklerinin tespit edilmesi amaçlanmıştır. Girişten alınan ses dosyası önerilen sesli aktivite tespit yöntemi ile segmentlere ayrıştırılmıştır. Elde edilen segmentlerden PLP, Rasta-PLP, MFCC ve Spectral yöntemleri kullanılarak özellik çıkarımı gerçekleştirilmiştir. Çıkarılan özellikler arasında benzerlik hesaplamasında korelasyon, Öklit uzaklığı ve DTW' den yararlanılmıştır. Önerilen sahtecilik lokalizasyon algoritmasıyla da şüpheli ses dosyasındaki sahte segmentler işaretlenmiştir. TIMIT ve Arabic Speech Corpus veri setlerinden üretilen iki ayrı sahte veri tabanı üzerinde yapılan deneysel sonuçlar yöntemin başarısının, literatürdeki çalışmalara göre oldukça yüksek olduğunu göstermektedir.

**Anahtar Kelimeler:** Kopyala – Yapıştır Sahteciliği, Pasif Yöntemler, PLP, Rasta-PLP, MFCC, Spectral Centroid

Master's Thesis

## SUMMARY

### DETECTION OF COPY MOVE FORGERIES IN AUDIO RECORDINGS USING PASSIVE VERIFICATION TECHNIQUES

Ali HATİPOĞLU

Karadeniz Technical University  
The Graduate School of Natural and Applied Sciences  
Computer Engineering Graduate Program  
Supervisor: Dr. Beste ÜSTÜBİOĞLU  
2024, 81 Sayfa

With the increasing number and ease of use of editing tools, attacks on audio files have become quite common, especially by malicious users. Verification of the authenticity of audio recordings presented as evidence in legal proceedings is of paramount importance. In the literature, methods for detecting digital audio authenticity have been proposed, which are generally categorized into two main groups: active and passive verification methods. In passive methods, audio verification is performed using statistical features extracted from the audio, and no additional information is required as in active methods. This study, based on passive verification methods developed within the scope of the thesis, aims to detect copy-move forgeries in audio files. The audio file taken from the input is segmented using the proposed speech activity detection method. Feature extraction was performed from the obtained segments using PLP, Rasta-PLP, MFCC, and Spectral methods. Correlation, Euclidean distance, and DTW were used in the similarity calculation between the extracted features. The proposed forgery localization algorithm also marked the forged segments in the suspicious audio file. Experimental results on two synthetic databases generated from TIMIT and Arabic Speech Corpus datasets demonstrate that the proposed method significantly outperforms previous studies in the literature.

**Keywords:** Copy – Move Forgery, Passive Techniques, PLP, Rasta-PLP, MFCC, Spectral Centroid



## ŞEKİLLER DİZİNİ

### Sayfa No

|           |                                                                                                                                                                          |    |
|-----------|--------------------------------------------------------------------------------------------------------------------------------------------------------------------------|----|
| Şekil 1.  | Ses sinyalinde yapılan sahtecilik tespit yöntemlerinin sınıflandırılması .....                                                                                           | 4  |
| Şekil 2.  | a) Örnek 1 ses sinyali b) Örnek 2 ses sinyalinden belirlenen kelime c) Örnek 1 ile Örnek 2 den alınan kelimenin birleştirilmesi.....                                     | 5  |
| Şekil 3.  | Önceden öğrenilmiş konuşmacı modeli ile farklı ses kaydından yapay ses üretilmesi .....                                                                                  | 8  |
| Şekil 4.  | a) Örnek orijinal ses dosyası sinyali b) Ses sinyalinin sonundaki kelimenin kopyalanarak sinyalin ortasına eklenmesi .....                                               | 9  |
| Şekil 5.  | Ses etkinliği algılamaya dayalı kopyala – yapıştır sahteciliği tespiti akış diyagramı .....                                                                              | 11 |
| Şekil 6.  | Çerçeve tabanlı kopyala – yapıştır sahteciliği tespiti akış diyagramı .....                                                                                              | 15 |
| Şekil 7.  | Görüntü tabanlı kopyala – yapıştır sahteciliği tespiti akış diyagramı .....                                                                                              | 16 |
| Şekil 8.  | Ses sinyalindeki sesli bölgelerin pitch yöntemi ile belirlenmesi .....                                                                                                   | 21 |
| Şekil 9.  | Genel PLP Akış Diyagramı .....                                                                                                                                           | 22 |
| Şekil 10. | Genel MFCC Akış Diyagramı .....                                                                                                                                          | 26 |
| Şekil 11. | Genel SCF ve SCM Akış Diyagramı .....                                                                                                                                    | 28 |
| Şekil 12. | İki Farklı Örnek Sinyalin SCF ve SCM grafiği .....                                                                                                                       | 29 |
| Şekil 13. | İki Farklı Örnek Sinyalin LPC, SCF ve SCM grafiği .....                                                                                                                  | 30 |
| Şekil 14. | Uygulanan kopyala – yapıştır sahteciliği tespiti işlem basamakları .....                                                                                                 | 36 |
| Şekil 15. | Örnek ses sinyalinin frekans genlik değerine göre görüntüsü .....                                                                                                        | 39 |
| Şekil 16. | Örnek ses sinyalinden pitch değerlerine göre tespit edilmiş kelimeler .....                                                                                              | 40 |
| Şekil 17. | Timit veri tabanında uygulanan yöntemlerin F1-skoruna göre ataklı ve ataksız dosyalarda performans karşılaştırılması .....                                               | 50 |
| Şekil 18. | Timit veri tabanında PLP yöntemi ile literatürdeki yöntemlerin ataksız ses sinyalinde sahtecilik tespit etme oranı karşılaştırılması .....                               | 69 |
| Şekil 19. | Timit veri tabanında PLP yöntemi ile literatürdeki yöntemlerin medyan filtreleme uygulanmış ses sinyalinde sahtecilik tespit etme oranı karşılaştırılması .....          | 70 |
| Şekil 20. | Timit veri tabanında PLP yöntemi ile literatürdeki yöntemlerin 20 dB gürültü eklenmiş ses sinyalinde sahtecilik tespit etme oranı karşılaştırılması .....                | 71 |
| Şekil 21. | Timit veri tabanında PLP yöntemi ile literatürdeki yöntemlerin 30 dB gürültü eklenmiş ses sinyalinde sahtecilik tespit etme oranı karşılaştırılması .....                | 72 |
| Şekil 22. | Arabic Speech Corpus veri tabanında PLP yöntemi ile literatürdeki yöntemlerin medyan filtre eklenmiş ses sinyalinde sahtecilik tespit etme oranı karşılaştırılması ..... | 73 |

## TABLÖLAR DİZİNİ

### Sayfa No

|           |                                                                                                                                                                                  |    |
|-----------|----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|----|
| Tablo 1.  | Pitch yöntemi ile sinyalde elde edilen öz vektör değerleri.....                                                                                                                  | 21 |
| Tablo 2.  | Timit veri tabanında orijinal ve veri tabanından oluşturulmuş sahte dosya sayısı.....                                                                                            | 37 |
| Tablo 3.  | Arabic Speech Corpus veri tabanında orijinal ve veri tabanından oluşturulmuş sahte dosya sayısı.....                                                                             | 37 |
| Tablo 4.  | Örnek ses sinyalinden elde edilen segmentlerin hesaplanan pitch değerleri.....                                                                                                   | 39 |
| Tablo 5.  | Timit veri tabanından elde edilen sahte veri tabanında PLP yöntemi özellik çıkarım ve korelasyon hesaplaması ile sahtecilik tespiti performans değerleri...                      | 42 |
| Tablo 6.  | Arabic speech corpus veri tabanından elde edilen sahte veri tabanında PLP yöntemi özellik çıkarım ve korelasyon katsayısı ile sahtecilik tespiti performans değerleri.....       | 43 |
| Tablo 7.  | Timit veri tabanından elde edilen sahte veri tabanında Rasta-PLP yöntemi özellik çıkarım ve korelasyon hesaplaması ile sahtecilik tespiti performans değerleri.....              | 44 |
| Tablo 8.  | Arabic speech corpus veri tabanından elde edilen sahte veri tabanında Rasta-PLP yöntemi özellik çıkarım ve korelasyon katsayısı ile sahtecilik tespiti performans değerleri..... | 45 |
| Tablo 9.  | Timit veri tabanında Spectral yöntemi özellik çıkarım ve korelasyon hesaplaması ile sahtecilik tespiti performans değerleri .....                                                | 46 |
| Tablo 10. | Arabic speech corpus veri tabanından elde edilen sahte veri tabanında Spectral yöntemi özellik çıkarım ve korelasyon katsayısı ile sahtecilik tespiti performans değerleri.....  | 47 |
| Tablo 11. | Timit veri tabanında MFCC yöntemi özellik çıkarım ve korelasyon katsayısı ile sahtecilik tespiti performans değerleri .....                                                      | 48 |
| Tablo 12. | Arabic speech corpus veri tabanından elde edilen sahte veri tabanında MFCC yöntemi özellik çıkarım ve korelasyon katsayısı ile sahtecilik tespiti performans değerleri.....      | 49 |
| Tablo 13. | Timit veri tabanından elde edilen sahte veri tabanında PLP yöntemi özellik çıkarım ve Öklid uzaklığı hesaplaması ile sahtecilik tespiti performans değerleri .....               | 52 |
| Tablo 14. | Timit veri tabanından elde edilen sahte veri tabanında PLP yöntemi özellik çıkarım ve DTW hesaplaması ile sahtecilik tespiti performans değerleri.....                           | 53 |
| Tablo 15. | Arabic speech corpus veri tabanından elde edilen sahte veri tabanında PLP yöntemi özellik çıkarım ve Öklid uzaklığı ile sahtecilik tespiti performans değerleri.....             | 53 |

|                                                                                                                                                                                      |    |
|--------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|----|
| Tablo 16. Arabic speech corpus veri tabanından elde edilen sahte veri tabanında PLP yöntemi özellik çıkarım ve Öklid uzaklığı ile sahtecilik tespiti performans değerleri.....       | 54 |
| Tablo 17. Timit veri tabanından elde edilen sahte veri tabanında Rasta-PLP yöntemi özellik çıkarım ve Öklid uzaklığı hesaplaması ile sahtecilik tespiti performans değerleri.....    | 55 |
| Tablo 18. Timit veri tabanından elde edilen sahte veri tabanında Rasta-PLP yöntemi özellik çıkarım ve DTW hesaplaması ile sahtecilik tespiti performans değerleri .....              | 56 |
| Tablo 19. Arabic speech corpus veri tabanından elde edilen sahte veri tabanında Rasta-PLP yöntemi özellik çıkarım ve Öklid uzaklığı ile sahtecilik tespiti performans değerleri..... | 58 |
| Tablo 20. Arabic speech corpus veri tabanından elde edilen sahte veri tabanında Rasta-PLP yöntemi özellik çıkarım ve DTW ile sahtecilik tespiti performans değerleri .....           | 59 |
| Tablo 21. Timit veri tabanında Spectral yöntemi özellik çıkarım ve Öklid uzaklığı ile sahtecilik tespiti performans değerleri .....                                                  | 60 |
| Tablo 22. Timit veri tabanında Spectral yöntemi özellik çıkarım ve DTW hesaplaması ile sahtecilik tespiti performans değerleri .....                                                 | 61 |
| Tablo 23. Arabic speech corpus veri tabanından elde edilen sahte veri tabanında Spectral yöntemi özellik çıkarım ve Öklid uzaklığı ile sahtecilik tespiti performans değerleri.....  | 62 |
| Tablo 24. Arabic speech corpus veri tabanından elde edilen sahte veri tabanında Spectral yöntemi özellik çıkarım ve DTW ile sahtecilik tespiti performans değerleri .....            | 63 |
| Tablo 25. Timit veri tabanında MFCC yöntemi özellik çıkarım ve Öklid uzaklığı ile sahtecilik tespiti performans değerleri .....                                                      | 64 |
| Tablo 26. Timit veri tabanında MFCC yöntemi özellik çıkarım ve DTW hesaplaması ile sahtecilik tespiti performans değerleri .....                                                     | 65 |
| Tablo 27. Arabic speech corpus veri tabanından elde edilen sahte veri tabanında MFCC yöntemi özellik çıkarım ve Öklid uzaklığı ile sahtecilik tespiti performans değerleri.....      | 66 |
| Tablo 28. Arabic speech corpus veri tabanından elde edilen sahte veri tabanında MFCC yöntemi özellik çıkarım ve DTW ile sahtecilik tespiti performans değerleri .....                | 66 |
| Tablo 29. Örnek ses dosyasından hesaplanan kelimeler arası korelasyon katsayıları.....                                                                                               | 67 |
| Tablo 30. Örnek ses dosyasından hesaplanan kelimeler arası Öklid uzaklığı .....                                                                                                      | 68 |

## SEMBOLLER VE KISALTMALAR DİZİNİ

|           |                                                                                          |
|-----------|------------------------------------------------------------------------------------------|
| CNN       | : Konvolüsyonel sinir ağı (Convolutional Neural Networks)                                |
| dB        | : Desibel                                                                                |
| DCT       | : Ayırık kosinüs dönüşümü (Discrete Cosine Transform)                                    |
| DFT       | : Ayırık fourier dönüşümü (Discrete Fourier Transform)                                   |
| DSSIM     | : Yapısal benzerlik indeksi ölçüsü (Difference Structural Similarity Index Measure)      |
| DTW       | : Dinamik zaman eşleştirme (Dynamic Time Warping)                                        |
| DTC       | : Ayırık zamanlı evrişim (Discrete Time Convulation)                                     |
| FFT       | : Hızlı fourier dönüşümü (Fast Fourier Transform)                                        |
| GRU       | : Kapılı tekrarlayan birim (Gated Recurrent Unit)                                        |
| GTCC      | : Gammatone cepstral katsayıları (Gammatone Cepstral Coefficients)                       |
| LBP       | : Yerel ikilik desen (Local Binary Patterns)                                             |
| LPC       | : Doğrusal tahmin kodlama (Linear Prediction Coding)                                     |
| LSTM      | : Uzun kısa süreli bellek (Long Short-Term Memory)                                       |
| MDCT      | : Değiştirilmiş ayırık kosinüs dönüşümü ( Modified Discrete Cosine Transform)            |
| MFCC      | : Mel frekansı cepstrum katsayısı (Mel Frequency Cepstrum Coefficients)                  |
| NCCF      | : Normalize edilmiş çapraz korelasyon fonksiyonu (Normalized Cross Correlation Function) |
| PLP       | : Algısal doğrusal Tahmin (Perceptual Linear Prediction)                                 |
| Rasta PLP | : Bağıl spektrum algısal doğrusal tahmin (Relative Spectra Perceptual Linear Prediction) |
| RNN       | : Geri beslemeli sinir ağı (Recurrent Neural Network)                                    |
| SCF       | : Spektral merkez nokta frekansı (Spectral Centroid Frequency)                           |
| SCM       | : Spektral merkez nokta büyüklüğü (Spectral Centroid Magnitude)                          |
| SD        | : Küçük sapma (Smallest Deviations)                                                      |
| SVD       | : Tekil değer ayrışımı (Singular Value Decomposition)                                    |
| SVM       | : Destek vektör makinesi (Support Vector Machine)                                        |
| SIFT      | : Ölçek değişmez öznitelik dönüşümü (Scale Invariant Feature Transform)                  |

STFT : Kısa zamanlı hızlı fourier dönüşümünü (Short Time Fast Fourier Transform)  
SURF : Hızlandırılmış sağlam özellikler (Speeded-Up Robust Features)  
TTS : Metinden konuşma üretme (Text To Speech)  
VAD : Ses etkinliği algılama (Voice Activity Detection)



## 1. GENEL BİLGİLER

### 1.1. Giriş

Teknolojinin sağladığı kolaylıklarla sesli iletişim kullanılarak pek çok işimizi uzaktan sesli yanıt sistemleri ile gerçekleştirebilmekteyiz. Sesli yanıt sistemi, bankacılık işlemleri, abonelik ve taahhüt işlemleri, kimlik doğrulama gibi alanlarda yararlanılmaktadır. Bu durum bazı güvenlik sorunları ortaya çıkardığı için genellikle görüşmeler kanunlar çerçevesinde kayıt altına alınır. Taraflar arasında yaşanan anlaşmazlıklarda bu kayıtlar mahkemeye adli dijital delil olarak sunulurlar. Günümüzde mahkemeler için en az fiziksel deliller kadar ses ve görüntü içerikli dijital delillerde kabul edilmekte ve adli olayların sonuçlanmasında anahtar rol oynamaktadırlar.

Birçok kurumsal firma sesli yanıt sistemleri ile vermiş oldukları hizmet üzerinde kullanıcının işlem yapabilmesine olanak sağlıyor. Bankacılık hizmetleri, telekomünikasyon hizmetleri buna örnektir. Sesli tanıma, sesli kimlik doğrulama gibi uygulamalarda sahtecilik yöntemleri kullanılarak oluşturulan sahte seslerle yetkisiz erişim sağlanabilmektedir. Bu durum hukuki suç unsuru olarak kullanılmaktadır.

Diğer taraftan insanlar görüşme esnasında telefonla ses kaydı alarak görüşme kaydını mahkeme için delil olarak kullanmaktadır. Mahkemeye delil olarak sunulan bu ses kayıtlarının doğruluğu da benzer bir hukuki sorun ortaya çıkartmaktadır.

Gelişen teknolojiye paralel olarak yapay zekâ uygulamalarının sağladığı kolaylıklar ve ses düzenleme yazılımları ile ses dosyaları üzerinde kolaylıkla değişiklikler yapılabilmektedir. Gelişmiş bu yöntem ve uygulamalar kullanılarak aslında var olmayan seslerin sahteleri üretilmektedir. Önceden kaydedilmiş farklı ses dosyalarını birleştirerek, tek bir ses dosyasındaki kelimelerin yerlerini değiştirerek ya da deep fake kullanarak yapay ses üretimi veya text to speech ile ses dosyası sahteciliği gerçekleştirilmektedir. Bu yöntemler kullanılarak oluşturulan ses sinyalleri ile kimlik taklidi, dolandırıcılık, bilgi kirliliği, sahte propaganda gibi amaçlarla kanuna aykırı faaliyetler yürütülmektedir.

Ses dosyası sahteciliğinin tespiti için kullanılan yöntemler Şekil 1’ de görüldüğü üzere iki ana sınıfa ayrılmıştır. Bu yöntemler aktif ve pasif yöntemlerdir. Aktif yöntemler; ses dosyası kaydedilirken bazı tekniklerle ses dosyasının içine kaydeden kişi, kayıt zamanı, kaydedilen yer hakkında bilgi eklenmesidir. Orijinal ses sinyaline filigran, dijital imza,

kaydeden kiři ya da cihaz bilgileri eklenebilir. Dosyaya eklenen bu bilgilerle ses dosyasının orijinal olduđu ya da düzenlenip deđiřtirildiđi ile ilgili ipucu vermektedir.

Pasif yöntemde ise ses dosyasına eklenen bu ön bilgiler yoktur. Onun yerine ses sinyalindeki konuşmacının, ortamın ya da mikrofonun akustik özelliklerden, ses dosyasının yapısından, dosya bütünlüğünden yararlanılarak sahtecilik tespit edilmeye çalışılır. Şekil 1’de görüldüğü üzere pasif yöntemler ses sinyali taşıyıcı temelli ve içerik temelli yöntemler olarak ikiye ayrılır (Üstübioğlu vd., 2023).

Taşıyıcı temelli yöntemler bütünlük kontrolü temelli ve dosya içerik temelli yöntemler olarak ikiye ayrılır.

Bütünlük kontrolü temelli yöntemler ses sinyalinin bütünlüğünü kontrol ederek dosyanın sahte olup olmadığını tespit eder. Bu yöntemler zaman damgası temelli yöntem ve hash kodu temelli yöntem olarak ikiye ayrılır. Zaman damgası temelli yöntemde ses dosyasının kayıt tarihi ve saati gibi zaman damgaları kullanılır. Zaman damgası hash fonksiyonları ile işlenerek ses dosyasının bütünlüğünü doğrulayacak şekilde kod üretilir. Hash kodu temelli yöntem içerik bütünlüğünü sağlamak için ses dosyasının sinyal içeriğinden hash algoritması ile sayısal bir imza oluşturur. Bu hash kodu tekil olarak oluşturulduğundan dosya içeriğı deđiřtirildiğinde geçerli hash kodun üretilmesi çok zordur (Qamhan vd., 2021).

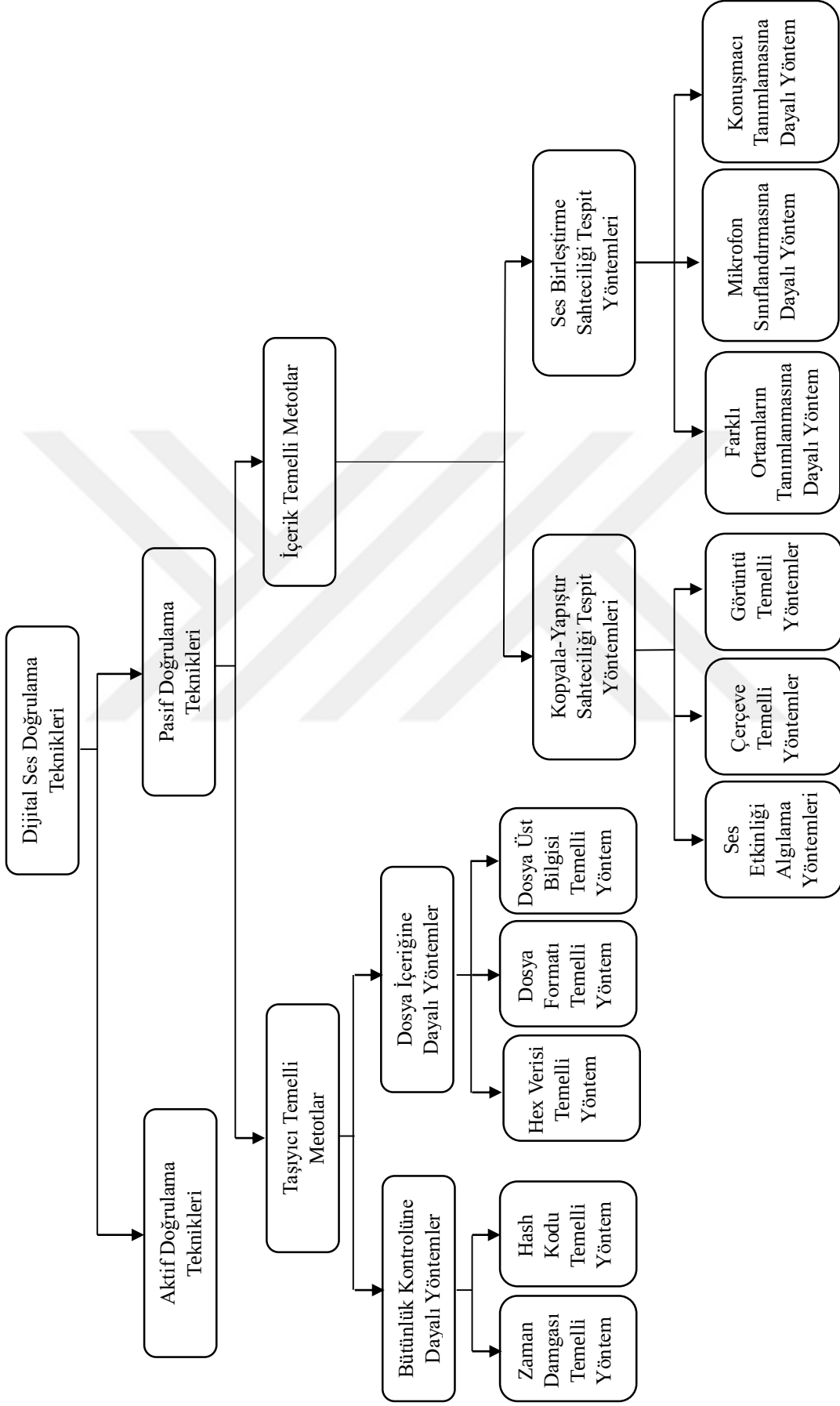
Dosya içeriğı temel alan sahtecilik tespiti çalışmaları hex verisi temelli yöntem, dosya formatı temelli yöntem ve dosya meta üst bilgisi temelli yöntemlerdir. Hex verisi dosyanın ham bayt verisi olduğundan bu yöntem ses sinyalindeki özellikler ya da dosya formatı yerine ham veri ile ilgilenir. Dosyanın hex verisi üzerindeki anormallikleri istatistiksel yöntemlerle hesaplayarak tespit edilir. Farklı olarak örüntü tanıma yöntemi ile de önceden tanımlanmış hex anormallikleri dosya ile kıyaslanarak dosyanın sahte olup olmadığı tespit edilebilir. Dosya formatı temeli yöntemde ses dosyasının formatı analiz edilir. Sahte ses üretici yazılımlardan elde edilen ses dosyası formatları gerçek ses dosyası formatından farklılık gösterirler. Meta üst bilgi verilerine dayalı sahtecilik tespitinde ise dosyanın boyutu, oluşturulma tarihi, format bilgisi gibi bilgilerden yararlanır. Sonradan üretilen sahte ses dosyalarında bu bilgilerde tutarsızlık oluşur. Dosya içeriğine dayalı sahtecilik tespit yöntemleri diđer yöntemlerle birlikte kullanılarak sahtecilik tespiti daha güvenilir hale getirilebilir (Zakariah vd., 2017).

Pasif yöntemlerden içerik temelli yöntemler ise kopyala-yapıştır sahteciliğı tespit yöntemleri ve ses birleřtirme sahteciliğı tespit yöntemleri olarak iki başlıkta incelenir.

Kopyala-yapıştır sahteciliği tespiti için ses aktivitesi tespit temelli, çerçeve temelli ve görüntü temelli yöntemler geliştirilmiştir. Bu yöntemlerde ses aktivitesi tespit yöntemlerinde ses kaydı üzerinde, özellik çıkarım yöntemleri ile sesli ve sessiz bölgeler tespit edilir. Tespit edilen konuşmalar kullanılan yöntemlere göre kelimelere ya da segmentlere bölünür. Bu kelime ya da segmentler birbiri ile benzerliği karşılaştırılarak sahtecilik tespit edilmeye çalışılır. Çerçeve tabanlı yöntemde ses sinyali eşit zaman aralıklarda çerçevelere bölünür ve her çerçeve birbiri ile aralarında benzerlik hesabı yapılır. Benzerlik derecesi değerine göre çerçevelerin kopya olduğuna karar verilir. Görüntü tabanlı yöntemlerde ise ses sinyalinin histogramından iki boyutlu görüntüsü üretilir. Görüntü işleme teknikleri ile elde edilen görsel üzerindeki benzer bölümler tespit edilmeye çalışılır. Pasif yöntemler adli olarak kabul edilebilir derecede sahtecilik tespit yöntemi kabul edilmektedirler.

Ses birleştirme sahteciliği tespit yöntemleri kayıt ortamlarının tanımlanmasına dayalı, mikrofon sınıflandırmasına dayalı ve konuşmacı tanıma dayalı yöntemler kullanılmaktadır. Kayıt ortamının tanımlanmasına dayalı yöntem, farklı ses sinyallerinin kayıt ortamlarını ayırtırmaya yönelik yöntemlerdir. Ses kayıt ortamlarının farklılığına bağlı olarak sinyaldeki ses seviyesi, çevresel gürültü, akustik gibi özellikler farklılık gösterir. Bu temel özelliklerden yararlanılarak ses sahteciliği tespit edilmeye çalışılmıştır. Mikrofon sınıflandırma yönteminde ise ses sinyalinde sesin kaydedildiği mikrofon kendine ait farklı akustik özellikler barındırır. Bu özellikler mel frekanslı cepstrum, formant frekans özellikleri, ses gürültü oranı gibi özelliklerdir. Akustik niteliklerinden yararlanarak sinyalden özellik çıkarımı ile ses kaydındaki tutarsızlıklar tespit edilmeye çalışılır (Qamhan vd., 2021). Konuşmacı tanıma dayalı yöntemlerde ise birleştirilen ses sinyalleri farklı konuşmacıların ses kayıtlarından oluşmaktadır. Sesler benzetilmiş kayıtların birleştirilmesi ile insan işitmesi ile anlaşılmayacak düzeydedir. Ancak farklı konuşmacıların akustik özelliklerinden yararlanarak sinyaldeki farklı konuşmacıların tanınması mümkündür. Mikrofon sınıflandırma yönteminde olduğu gibi akustik özellikler kullanılarak konuşmacı tanıma çalışılır. Bu özelliklerden önceden elde edilen veri seti üzerinde farklı konuşmacılarla model eğitimi yapılır. Bu model ses kaydındaki akustik özelliklerle konuşmacıyı tanıma çalışır. Farklı konuşmacılara ait birleştirilmiş ses kaydında tanımlanan konuşmacılar arasında tutarsızlık gözlenir. Bu tutarsızlıkla ses birleştirme sahteciliği tespit edilmeye çalışılır (Jadhav vd., 2019).





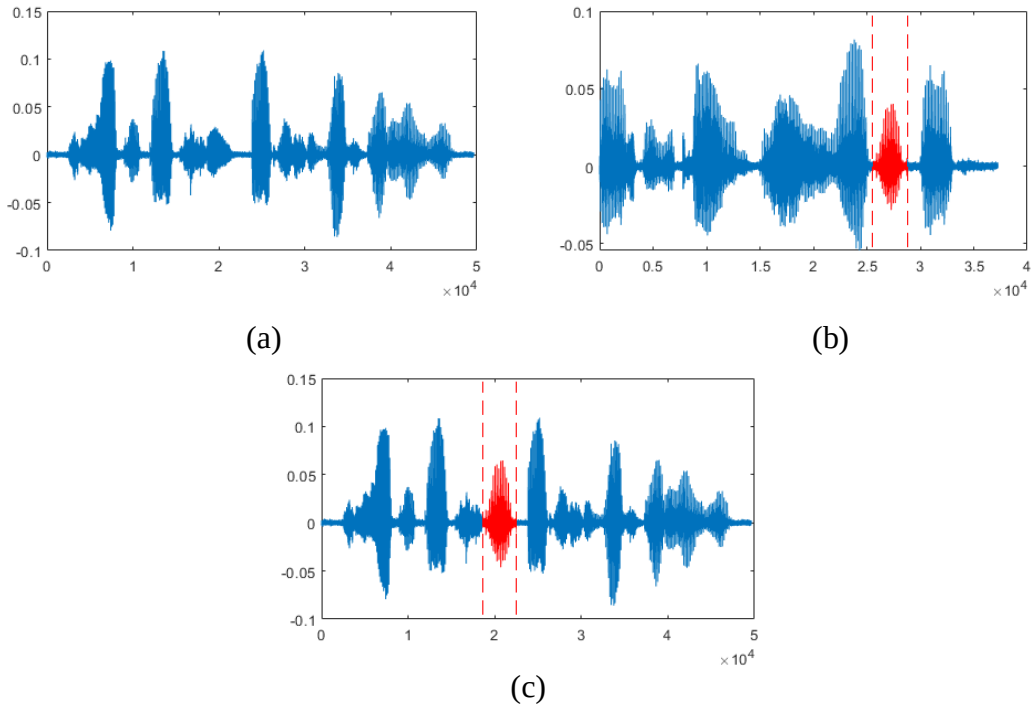
Şekil 1. Ses sinyalinde yapılan sahtecilik tespit yöntemlerinin sınıflandırılması

## 1.2. Ses Dosyasında Yapılan Sahtecilikler

Yazılım dünyasındaki çalışmalarla daha da geliştirilen ses düzenleme uygulamaları ile ses dosyaları üzerinde düzenleme yapmak çok kolaylaşmıştır. Günümüzde profesyonel olarak ses ve kayıt düzenleme uygulamalarına Pro Tools, Adobe Audition, Logic Pro, Ableton Live, Steinberg Cubase, FL Studio örnek olarak verebiliriz. Gelişmiş filtre algoritmaları ile desteklenen bu uygulamalarla ses dosyasında anlaşılması zor derecede düzenlemeler yapılabilir. Çok uzman olmayan birisi bile gelişmiş programların bu özellikleri ile kolaylıkla ses dosyası üzerinde sahtecilik yapabilir.

Yapılan ses dosyası sahtecilikleri kullanılan yöntemlere göre ses birleştirme, yapay zekâ ve derin öğrenme ile metinden konuşma üretme ya da kopyala – yapıştır olarak sınıflandırılabilir.

Ses birleştirme (splicing) sahteciliğinde birden fazla ses dosyasından yararlanılır. Ses dosyalarının tamamı ya da bir kısmı uygun şekilde birleştirilerek sahte ses dosyası elde edilir. Örneğin ilk ses dosyasındaki konuşmada “onaylıyorum” kelimesini kesip ikinci ses dosyasına eklenerek “işlemi onaylıyorum” şeklinde yeniden düzenlenebilir. Farklı zaman ve ortamlarda kaydedilen ses dosyasındaki konuşmalar aynı kişiye ait olsa da temel ses frekansı farklı olacaktır.



Şekil 2. a) Örnek 1 ses sinyali b) Örnek 2 ses sinyalinden belirlenen kelime c) Örnek 1 ile Örnek 2 den alınan kelimenin birleştirilmesi

Şekil 2’de görüldüğü üzere a ses kaydına, b deki ses kaydından bir parça alınarak eklenmiş ve c deki yeni ses sinyali elde edilmiştir.

Sıkıştırma işlemi tabanlı yöntemler: Ses birleştirme sahteciliği tespiti alanında yapılan çalışmalarda, değiştirilmiş ayırık kosinüs dönüşümü (Modified Discrete Cosine Transform - MDCT) katsayılarının istatistiksel özelliklerinin sınıflandırılması sıklıkla kullanılan bir yöntemdir. Liu ve arkadaşları (2010) başta olmak üzere Qiao vd. (2010) ve Luo vd. (2012) gibi araştırmacılar, destek vektör makinesi (Support Vector Machine - SVM) gibi makine öğrenmesi algoritmalarını kullanarak MDCT katsayılarının istatistiksel özelliklerini sınıflandırmış ve başarılı sonuçlar elde etmişlerdir. Hiçsönmez vd. (2013) ise tek ve çift sıkıştırılmış sesleri ayırt etmek amacıyla farklı bir yaklaşım benimsemiş, çözülmüş ve kodlanmış ses dosyalarından elde edilen farklı özelliklerin kombinasyonu ile bir sınıflandırma sistemi önermiştir. Luo vd. (2017) ise derin öğrenme tabanlı bir yaklaşımla, yığılmış oto kodlayıcı (Stacked Autoencoder - SAE) ve evrensel arka plan modeli - Gauss karışım modeli (Universal Background Model—Gaussian Mixture Model / UBM-GMM) gibi modeller kullanarak daha karmaşık ses işleme problemlerine çözüm bulmaya çalışmıştır. Uyarlanabilir çoklu bit hızı (Adaptive Multi-Rate – AMR) ses sıkıştırma standardı üzerinde yapılan deneyler, bu yöntemin performansını değerlendirmek için yaygın olarak kullanılmaktadır.

Mikrofon (kayıtçı) tanımlama tabanlı yöntemler: Cuccovillo vd. (2013) tarafından yapılan çalışmada, ses kayıtları Hanning penceresi kullanılarak çerçevelere ayrılmış, elde edilen çerçevelerden RASTA-MFCC katsayıları çıkarılmış ve bu katsayılar radyal temel fonksiyon (Radial Basis Function - RBF) çekirdekli SVM sınıflandırıcısı ile sınıflandırılmıştır. Benzer şekilde, Bykhovsky (2020), ses sinyallerindeki ENF harmonik genlik katsayılarını kullanarak SVM tabanlı bir sınıflandırma sistemi önermiştir. Kraetzer vd. (2007) ve Buchholz vd. (2009) gibi diğer çalışmalar da, farklı makine öğrenmesi algoritmaları ile ses sahteciliğini tespit etmeye yönelik çeşitli yöntemler sunmuştur.

Çevresel sinyallerin tespiti tabanlı yöntemler : Ali vd. (2017) çalışmasında, insan işitme sisteminin psikoakustik özelliklerini de dikkate alarak daha gelişmiş bir özellik çıkarma yöntemi kullanmıştır. Malik vd. (2010), Zhao vd. (2012) ve Ikram vd. (2010) gibi diğer çalışmalar da, çevresel seslerin sınıflandırılması konusunda farklı yöntemler ve özellik çıkarma teknikleri önererek bu alandaki çalışmalara önemli katkılar sağlamışlardır.

Konuşmacı tanımlama tabanlı yöntemler: Todisco vd. (2017), sabit Q dönüşümü temelli yeni bir özellik olan sabit Q kepsral katsayıları (Constant-Q Cepstral Coefficient –

CQCC's) kullanarak ve bu özellikleri GMM ile sınıflandırarak bu alana önemli katkılar sağlamıştır. Campbell vd. (2007) ve Champod vd. (2000) gibi önceki çalışmalar da konuşmacı tanımlama probleminde benzer yaklaşımlar benimsemişlerdir. Ancak Huang ve Pun (2020), bu yöntemlerin yeniden oynatma saldırılarına karşı yeterince dayanıklı olmadığını belirterek, doğrusal filtre bankası özelliklerini ve derin öğrenme mimarilerini bir araya getiren yeni bir yöntem önermişlerdir. Özellikle ses kayıtlarındaki sessiz bölgelere odaklanarak elde edilen doğrusal filtre bankası özellikleri, DenseNet-BiLSTM mimarisi ile sınıflandırılarak, CQCC ve GMM tabanlı yöntemlere göre daha başarılı sonuçlar elde edilmiştir.

Diğer yöntemler: Zhang vd. (2019), bu alanda yeniden örnekleme yöntemine dayalı farklı bir yaklaşım sunmuşlardır. Kısa zamanlı hızlı fourier dönüşümünü (Short Time Fast Fourier Transform -STFT) ve yerel iki desen (Local Binary Patterns - LBP) gibi özellik çıkarma tekniklerini birleştirerek elde ettikleri özellikleri SVM ile sınıflandırarak başarılı sonuçlar elde etmişlerdir. Derin öğrenme yöntemlerinin bu alandaki etkisi de göz ardı edilemez. Jadhav vd. (2019), konvüsiyonel sinir ağı (Convolutional Neural Networks – CNN) mimarilerini kullanarak ses sinyallerindeki farklılıkları tespit etmek için yenilikçi bir yaklaşım önermişlerdir. Benzer şekilde, Lin vd. (2016), ENF sinyallerinin seviyelerini CNN ile öğrenerek sahteciliği tespit etmeyi amaçlamışlardır. Chen vd. (2018) ise MDCT dönüşümü ve CNN'leri birleştirerek, el ile çıkarılan özelliklere göre daha başarılı sonuçlar elde etmiştir.

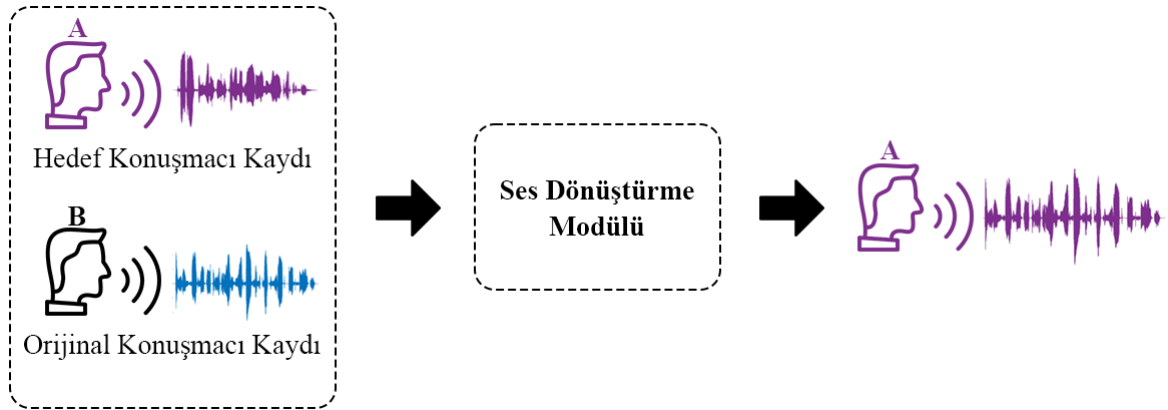
Bu çalışmalar, ses birleştirme sahteciliğinin tespiti için hem geleneksel sinyal işleme tekniklerinin hem de derin öğrenme yöntemlerinin etkinliğini göstermektedir. STFT, LBP, SVM, CNN, MDCT gibi farklı tekniklerin bir araya getirilmesiyle daha güçlü ve güvenilir sistemler geliştirilebileceği görülmektedir. Gelecekte, bu alanda büyük veri kümeleri üzerinde eğitilen daha karmaşık derin öğrenme modelleri ve çok modlu (ses, video vb.) analizlerin birleştirilmesiyle daha da başarılı sonuçlar elde edilebileceği öngörülmektedir.

Yapay zekânın kötüye kullanılmasıyla ortaya çıkan derin sahtekarlık (deep fake) ile metinden konuşmaya dönüştürme ve konuşmadan konuşmaya dönüştürme veya ses dönüştürme olmak üzere iki tür sahtecilik uygulanmaktadır (Mubarak vd., 2023).

Metinden konuşma üretme (Text To Speech - TTS) yönteminde konuşmacının önceden kaydedilmiş farklı ses dosyalarından elde edilen verilerle derin öğrenme gerçekleştirilir. Konuşmacının vurgu ve ses tonu yapay zekâ yöntemleri ile taklit edilerek yapay konuşmacı elde edilir. İstenen ses dosyasını elde etmek için metin dosyasında bulunan

yazılar bu yapay konuşmacıya okutularak aslında hiç olmayan konuşmaya ait yapay ses dosyası üretilmektedir. Bu manipölasyon tekniğı, belirli bir kişiyi taklit eden konuşmaya dönüştürölmek üzere yazılan metinlerin sentezlenmesini içerir. Metinden konuşma üretme, sesli asistanlar, navigasyon sistemleri ve konuşmadan konuşmaya çeviri gibi çeşitli uygulamalarda kullanılan eski bir teknolojidir. Ancak geleneksel TTS yaklaşımları doğal olmaktan uzak olup insan benzeri konuşmayı taklit edemez; konuşma sesleri yapaydır ve kolayca fark edilebilir.

Metinden konuşmaya dönüştürmeden farklı olarak, ses dönüştürme tekniğinde başka bir kişinin, önceden tanımlanmış konuşmacı sesine uyacak şekilde konuşmacının ses özelliklerini değıştirir. Bir konuşmacının ses kaydından elde edilen kelimelere göre aynı konuşmayı önceden öğrenilmiş yapay konuşmacıya okutarak ta sahte ses sinyali üretilmektedir (Almutairi vd., 2022).

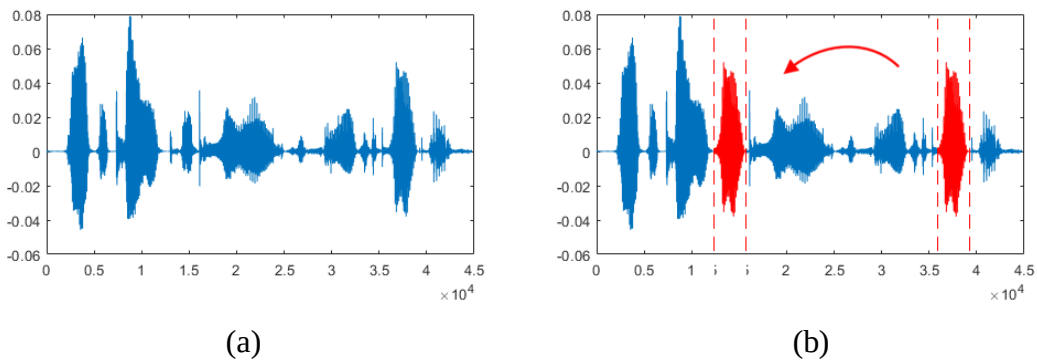


Şekil 3. Önceden öğrenilmiş konuşmacı modeli ile farklı ses kaydından yapay ses üretilmesi

Şekil 3’te görüldüğü gibi A konuşmacısının önceden kaydedilmiş farklı konuşmalarına ait ses kayıtlarından derin öğrenme ile yapay konuşmacı oluşturulur. B konuşmacısı ise A konuşmacısının söylemesini istediğı cümleleri söyleyerek orijinal ses kaydı oluşturur. B konuşmacısından elde edilen orijinal ses kaydı ile ses dönüştürme modölüne sokularak A konuşmacısının öğrenilmiş sesi, tonlaması ve vurgusu ile sahte ses dosyası üretilir. B konuşmacısının ses sinyali yerine metin belgesi de kullanılabilir. Ancak, Resemble.AI ve MURF.AI gibi kullanıcı dostu çevrimiçi ses sentezi araçları tehlikeli olabilir, çünkü hedef kişinin sesini taklit ederek bireyleri aldatmaya yönelik ses içeriğı oluşturmak için kullanılabilir. Bu sahtecilikler, kimlik hırsızlığı ve yanlış bilgi yayma gibi kötü niyetli faaliyetler için kullanılabilir (Mubarak vd., 2023).

Yapay seslerin tespit edilmesine yönelik Galou vd. (2011) gizli Markov modeli (Hidden Markov Model - HMM) tabanlı bir yöntem uygulamıştır. HMM tabanlı bu modelde metin-konuşma sistemlerinin kalitesi algısal olarak mükemmel olmasa da, düşük arka plan gürültüsüne sahip kısa ifadelerde başarılı sonuçlar elde etmişlerdir. Sanchez vd. (2015) sinyalin faz bilgisini kullanarak, GMM temelli ikili sınıflandırma modeli uygulamıştır. Eğitim aşamasında yapay ve doğal seslerden GMM parametrik bir model oluşturulmuştur. Göreceli faz kaymaları (Relative Phase Shifts - RPS) ve MFCC den elde edilen vektörler bu parametrik modelle karşılaştırılarak yapay sesler tespit edilmiştir. Hamza vd. (2022) frekans alanındaki ses sinyallerinin, belirli senaryolarda insanı yanıltma potansiyeli olan derin sahte seslerin tespit ve sınıflandırılmasında faydalı özellikler sağlayabileceği mel frekansı cepstrum katsayısı (Mel Frequency Cepstrum Coefficients – MFCC) dayalı özellik çıkarımı yaparak SVM ve karar ağacına dayalı sınıflandırma gerçekleştirmiştir. Chakravarty vd. (2023) MFCC ve gammatone cepstral katsayıları (Gammatone Cepstral Coefficients - GTCC) özellik çıkarım yöntemlerini hibrit olarak kullanarak GMM, SVM, Uzun Kısa Süreli Bellek (Long Short-Term Memory - LSTM), Kapılı Tekrarlayan Birim (Gated Recurrent Unit - GRU) sınıflandırma yöntemleri ile gelişmiş bir model uygulamıştır.

Kopyala – yapıştır (copy-move) sahteciliğinde ise tek bir ses dosyasındaki farklı bir bölümün kopyalanarak bir başka bölümün üzerine ya da yanına yapıştırılması ile anlamı değiştirilmiş yeni bir ses dosyası elde edilir. Böylece ses dosyasındaki ifadenin anlamsal bütünlüğü değiştirilebilir (Yan vd., 2019). Örneğin kaydedilmiş bir ses dosyasındaki “hesabımdan para çekilmesini onaylamıyorum, krediyi onaylıyorum” cümlesi kelimelere ayrılarak yeniden düzenlenir. Kopyala- yapıştır sahteciliği uygulanmış örnek ses sinyali Şekil 4’de görülmektedir.



Şekil 4. a) Örnek orijinal ses dosyası sinyali b) Ses sinyalinin sonundaki kelimenin kopyalanarak sinyalin ortasına eklenmesi

Ses dosyasında geçen onaylıyorum kelimesi kopyalanarak onaylamıyorum kelimesinin üzerine yapıştırılır ve “hesabımdan para çekilmesini onaylıyorum, krediyi onaylıyorum” olarak değiştirilebilir. Ses dosyaları üzerinde bu şekilde yapılan sahtecilikler kopyala- yapıştır sahteciliği olarak isimlendirilir.

### **1.3. Tezin Kapsamı**

Bu tezin içeriğinde, hukuki delil olarak da kabul gören ses dosyalarında kopyala-yapıştır sahteciliğinin tespitine yönelik pasif yöntemler uygulanmıştır. Kopyala- yapıştır sahteciliğinin pasif yöntemlerle tespitinde kullanılan literatürdeki çalışmalar açıklanmıştır. Genel bilgiler bölümünde uygulanan yöntemler sahtecilik tespit yöntemine göre sınıflandırılarak yöntemler kısaca açıklanmıştır. Yapılan çalışmalar, bulgular ve irdeleme bölümünde ses dosyasındaki sahtecilik tespiti için yaptığımız çalışmalar özetlenmiştir. Kullanılan veri tabanı anlatılarak performans ölçümünde yararlanılan metrikler tanımlanmıştır. Devamında ise çalışmaların sonucunda elde edilen istatistiksel sonuçlar yorumlanarak literatürdeki ses kopyala- yapıştır sahteciliği tespitinde yapılmış çalışmalarla kıyaslanmıştır.

### **1.4. Kopyala Yapıştır Sahteciliği Tespiti**

Ses dosyalarında yapılan sahteciliğin tespiti günümüzde bilişim dünyasındaki gelişmelere paralel olarak önem kazanmıştır. Bu alanda yapılan çalışmalar ses dosyalarında değişiklik yapıp yapılmadığını, ses dosyasının sahte olup olmadığını tespit etmeyi amaçlar.

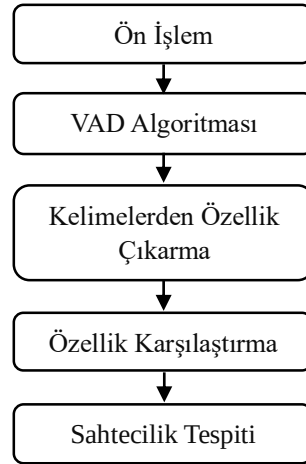
### **1.5. Literatür Araştırması**

Ses dosyasında yapılan sahteciliklerin tespit edilmesi ile ilgili literatürde farklı tekniklerle uygulanan yöntemler önerilmiştir. Yapılan çalışmaları incelediğimizde üç başlık altında toplanabilir. 1- Kopyala yapıştır sahteciliğinin tespiti için ses etkinliği algılama (Voice Activity Detection - VAD) tabanlı sahtecilik tespit yöntemleri, ses sinyalindeki kelimelerin karşılaştırılması ile sahteciliği tespit etmeye çalışmıştır. 2- Çerçeve tabanlı sahtecilik tespit yöntemleri ise sinyali eşit çerçevelere bölerek çerçeveler arasındaki benzerliği tespit etmiştir. 3- Görüntü tabanlı sahtecilik tespit yöntemleri ise, ses sinyalinden

elde edilen görüntüler üzerinde benzerlik tespiti yapmaya çalışır. Örneğin, spektrogram gibi ses sinyalinin görselleştirilmesiyle elde edilen görüntüler, sahteciliğin tespiti için kullanılır. Bu yöntem, sinyalin zamansal ve frekansal özelliklerini görselleştirerek analiz yapmayı mümkün kılar. Aşağıda bu çalışmalar başlıklar halinde detaylandırılmıştır. Önerilen yöntemler özellik çıkarım yöntemine ve benzerlik tespitine göre farklılıklar göstermektedir.

### 1.5.1.Ses Etkinliği Algılamaya Dayalı Kopyala-Yapıştır Sahteciliği Tespiti Yöntemleri

Ses dosyasında kopyala-yapıştır sahteciliğinin tespit edilmesinde ses etkinliği algılamaya dayalı yöntemlerde ses sinyalindeki sesli ve sessiz bölümler tespit edilir. Tespit edilen sesli segmentlerin özellik çıkarımı yapılarak karşılaştırılmaktadır. Bu yöntemlere ses etkinliği algılamaya dayalı yöntemler denir. Ses sinyali başından sonuna kadar frekans değerlerinin özellikleri çıkartılarak konuşmanın geçtiği anlar ses etkinliği algılama algoritması ile işaretlenir. Aktivite olan bölümler “1” ile işaretlenirken konuşma olmayan bölümler ise “0” ile işaretlenir. “1”lerden oluşan işaretlenmiş blokların her biri ses sinyalinde geçen bir kelimeyi ifade etmektedir. VAD tabanlı yöntemlerin genel akış diyagramı Şekil 5’te görülmektedir.



Şekil 5. Ses etkinliği algılamaya dayalı kopyala – yapıştır sahteciliği tespiti akış diyagramı

VAD algoritmasından elde edilen kelimelerin özellik çıkartım algoritmaları ile özellikleri çıkartılarak birbiri ile kıyaslanır. Kelimelerden elde edilen özelliklerin benzerlikleri Pearson korelasyon katsayısı, Öklid uzaklığı, ortalama fark gibi yöntemlerle hesaplanarak kopyala-yapıştır sahteciliği tespit edilmeye çalışılmıştır.



Mannepalli vd. (2019) önerdiği bu yöntemde kelimelerin tespit edilebilmesi için sıfır geçiş oranı ve mel frekansı cepstrum katsayısı (Mel Frequency Cepstrum Coefficients-MFCC) özellik çıkartım yöntemini kullanmıştır. Özellikleri çıkartılan kelimeler dinamik zaman eşleştirme (Dynamic Time Warping – DTW) algoritması ile sahte ve orijinal sinyal arasındaki mesafeleri hesaplanmıştır. Son olarak kelimeler arasındaki benzerlik mesafesine bakarak kopyala-yapıştır sahteciliği olan kelimeler tespit edilmiştir.

Yan vd. (2019) ataklara karşı perde özelliklerinin etkilendiğini düşünerek YAAPT adı verilen güçlü bir perde izleme metodu benimsemişlerdir. Belirlenen sesli kelimelerin perde dizilerinden ilk iki formant özelliğini çıkartarak ses sinyalinin doğrusal tahmin kodlama (Linear Prediction Coding - LPC) yöntemi ile temel bileşenleri belirlenmeye çalışılmıştır. Perde değerlerinin elde edilen özellikleri ile benzerlik tespitinde ise DTW algoritması kullanılmıştır. İki konuşma kelimesinin hesaplanan DTW değeri eşik değerinden küçükse kopyala-yapıştır sahteciliği yapıldığına karar verilmiştir.

Ayrık fourier dönüşümü algoritmasını benimseyen bu yöntemde Huang vd. (2019) ses sinyalini bölümlere ayırarak sesli alanları tespit etmiştir. Buradaki hassasiyet sahteciliğin tespit edilebilmesi için çok önemlidir. Tespit edilen sesli bölümlerin ayrık fourier dönüşümü yöntemi ile özellikleri çıkartılarak sesli bölümlerin özellik dizisi elde edilir. Sesin değiştirilip değiştirilmediği ile ilgili tespit için son adım benzerlik karşılaştırmasıdır. Bu yöntemde zaman kaybını azaltmak için özellik dizisini sıralayan önemli iki adım içerir. Bu sayede sesli bölümleri tüm tespit edilen bölümlerle değil sadece komşu olan bölümlerle benzerlik karşılaştırması yapılır. Komşuların özellikleri karşılaştırılırken Pearson korelasyon katsayısına bakılır. Hesaplanan korelasyon katsayısı eşik değerinden daha büyük çıkarsa bu iki kelimenin kopyala-yapıştır olduğu tespit edilir.

Anh vd. (2019) mahkemede delil olarak kullanılan ses dosyalarının uzun olduğunu ve sahtecilik tespiti için fazla zaman harcadığına dikkat çekmiştir. Çalışmasında ise ses dosyasının genliğini büyüklük olarak kullanarak sahteciliği tespit etmişlerdir. Ses sinyalindeki aynı kişinin aynı kelimeyi söylediği farklı zamanlardaki gerçek kayıta sinyalin genliğinde temel ses frekansları farklılık gösterir ve kolaylıkla ayırt edilebilir.

Sinyaldeki kelimelerin büyüklüklerini tespit etmek için sinyaldeki genliklerin negatif değerlerini tersini alarak pozitif değere döndürmüştür. Böylelikle tüm değerler 0 dan büyük hale gelmiştir. Ses kayıt ortamındaki gürültü değerine göre eşik belirleyerek eşikten büyük olan sinyal bölümlerini kelime olarak belirlediği büyüklük dizilerini elde etmiştir. Elde ettiği büyüklükleri karşılaştırmak için en küçük sapma (Smallest Deviations - SD) yöntemini

kullanmıştır. Kopyalanan kelimelerin sapma değerlerinin 0 dan çok küçük olduğu benzerlik olmayan kelimelerde ise 5,6,12 gibi sonuçlar vermektedir. Elde ettiği karşılaştırma sonuçlarından belirlenen ikinci bir eşik değeri ile kopyala-yapıştır sahteciliği uygulanan kelimeleri tespit etmiştir.

Ses sinyalinin dört farklı özellik çıkarım yöntemiyle elde edilen özellik dizileri kullanılmıştır. Xie vd. (2018) çoklu özellikten karar verme yöntemini önermiştir. Birden fazla özellik kullanarak daha hassas kopyala- yapıştır sahteciliği tespiti yapılabildiğini savunmuştur. Kelimelerin doğru tespit edilebilmesi sahteciliğin tespitinde çok önemli bir adımdır. Gürültüye karşı dayanıklılığını artırmak için ikili eşik değeri kullanılmıştır. Kaba tespit için kullanılan yüksek eşik değerinden büyük olan değerler kelime olarak belirlenmektedir. Belirlenen kelimeler ikinci eşik değeri ile kontrol edilerek eşik değerinden küçük olan bölgeler işaretlenir. Kelimeler bu iki nokta arasında kalan bölgelerdir.

Çoklu karar verme için kelimelerden çıkardığı dört farklı özelliği gammatone, mel frekanslı cepstral katsayıları, perde, dizisel fourier dönüşümü katsayıları yöntemleri ile elde etmiştir. Tespit edilen kelimelerin dört farklı özelliği ayrı ayrı çıkartılmıştır. Çıkartılan özelliklerden gammatone ve dizisel fourier dönüşümü için Pearson korelasyon katsayısı, mel frekanslı cepstral katsayıları ve perde yöntemleri için ortalama fark hesaplanarak yakınlıkları hesaplanmıştır. Hesaplanan değerler C4-5 karar ağacı ile karşılaştırılarak kelime benzerlikleri tespit edilmeye çalışılmıştır.

Imran vd. (2017) kelimelerin tespit edilmesinde tüm ses dosyasını 20 milisaniyelik birbiri ile örtüşmeyen bölümlere ayırmıştır. Her bölümün toplam genliği ayrı ayrı hesaplandıktan sonra toplam genlik dizisinden hesaplanan eşik değeri ile sesli ve sessiz bölümler tespit edilmiştir. Bölümün toplam genlik değeri eşik değerinden büyükse sesli bölüm değilse sessiz bölüm olarak işaretlenmiştir. Sesli olarak işaretlenen bölümler birleştirilerek kelimeler belirlenmiştir.

Tespit edilen kelimelerin herbiri için histogram değeri hesaplanarak kelimelere ait histogram dizisi üretilmiştir. Tespit edilen kelime sayısı büyüklüğünde bir kare matrisle tüm kelimelerin histogramları diğerleri ile karşılaştırılır. Tüm kelimelerin histogram değerleri ile ortalama kare hatası ve enerji oranı hesaplanmaktadır. Sahteciliğin tespit edilmesinde bu iki parametre kullanılmaktadır. Enerji oranı histogram değerlerinin karelerinin toplamı şeklinde hesaplanır. Çıkan sonuç 1 veya çok yakın bir değer ise karşılaştırılan kelimelerin histogramlarının aynı olduğu anlamına gelir. Ayrıca kelimeler arasında hesaplanan ortalama kare hatası 0 veya çok yakın bir değer çıkarsa bu kelimelerin sahte olduğunu gösterir.

Bu yöntemde öncelikle hecelerin tespit edilebilmesi için ilk olarak ses sinyali normalleştirilir. Daha sonra birkaç parçaya bölünerek sesli hecelerin yerleri tespit edilir. Wang vd. (2017) kelimeleri tespit ettikten sonra ayrık kosinüs dönüşümü (Discrete Cosine Transform - DCT) yöntemi ile belirlenen hecelere ait katsayılar hesaplamıştır. Elde edilen katsayılar bir sonraki adımda tekil değer ayrışımı (Singular Value Decomposition - SVD) yöntemi ile hesaplanarak tekil değerler vektörü elde edilir.

Kelimelere ait tekil vektörler sıra ile birbiri ile karşılaştırılarak kopyala-yapıştır sahteciliği tespiti için Öklid uzaklığı hesaplanır. Hesaplanan mesafe fark derecesidir. Fark derecesine göre eşik değeri belirlenir. İki kelime arasında hesaplanan fark derecesi değeri eşikten daha büyük ise bu iki kelime kopyala-yapıştır olarak tespit edilir.

Ses sinyalindeki sesli ve sessiz bölgelerin tespit edilmesinde perde dizileri kullanılmıştır. Perde dizilerini belirlemek için Yan vd. (2015) sağlam bir perde izleme metodu olan YAAPT algoritmasını kullanmıştır. Bu yöntem perde özelliklerinin çıkartılması ve perde dizilerinin karşılaştırılması olarak iki temel aşamadan oluşur.

Tespit edilen kelimelerden YAAPT algoritması ile elde edilen perde özellik değerleri tüm kelimeler için diğer kelimelerle kıyaslanmaktadır. Kıyaslama işlemi için hem Pearson korelasyon katsayısı hem de ortalama fark hesaplanır. Korelasyon değerleri -1 ile 1 aralığında hesaplanır. 1 değeri hesaplanan kelimeler birbiri ile benzerdir. Ortalama fark değerinde ise sahte olan perde dizisi değeri normallere göre çok düşük çıkmaktadır. Sahte sesleri ayırt etmek için uygun bir eşik değeri belirlenir. Eşik değerinden daha küçük ortalama değere sahip kelimeler kopyala-yapıştır sahteciliği olarak belirlenir.

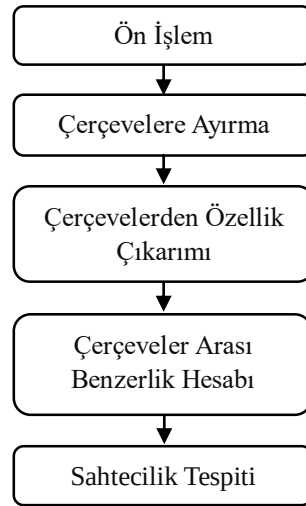
Üstübioğlu vd. (2022) kelimeleri belirledikten sonra benzerlik karşılaştırması için özellik çıkarım yöntemi olarak değiştirilmiş ayrık kosinüs dönüşümü ( Modified Discrete Cosine Transform - MDCT) yöntemini uygulamıştır. Ses aktivitesi tespit etmek için perde dizisi kullanılarak ses dosyasındaki sesli ve sessiz bölümler işaretlenmiştir. Sesli olarak belirlenen bölümler 1 ile sessiz olan bölümler 0 ile işaretlenmiştir. 1 ile işaretlenen bu bölümlere ait perde dizileri ayrı ayrı elde edilerek kelimelere ayrılmıştır.

Kelimeler belirlendikten sonra değiştirilmiş ayrık kosinüs dönüşümü yöntemi ile perde dizilerinden özellik çıkartımı yapılarak kelimelere ait veriler elde edilmiştir. Kelime sayısı kadar  $m \times m$  kare matris oluşturularak tüm kelimelerin diğer kelimelerle olan Öklid uzaklığı hesaplanmıştır. Kelimelere ait elde edilen bu Öklid değerlerinden minimum değeri olan kelimelerin eşik değerinin altında olanların kopyala yapıştır sahteciliği yapıldığı belirlenmiştir.

Akdeniz vd. (2024) sesteki kopyala-yapıştır sahtekarlığını tespit etmek için ilk kez uçtan uca bir sistem tasarlamıştır. Bu sistem, geri beslemeli sinir ağı (Recurrent Neural Network - RNN) ve LSTM modellerine dayanmaktadır. Geliştirdikleri bu sistem üç aşamadan oluşmaktadır. Ses sinyallerinin sessiz ve sesli segmentlerine ayırma. Her bir segmentten Mel frekansı temelli özelliklerin çıkarılması. RNN ve LSTM modelleri kullanılarak sesteki kopyala-yapıştır sahtekarlığının tespit edilmesi.

### 1.5.2. Çerçeve Tabanlı Kopyala-Yapıştır Sahteciliği Tespiti Yöntemleri

Bu yöntem kelimelerin tespit edilmesi yerine ses sinyalini T zaman aralığında çerçevelere bölünerek karşılaştırma temelini benimsemiştir. Xiao vd. (2014) sinyalin eşit T zamanlı periyotlara bölünmesinde T zaman aralığının belirlenmesi yöntem için önemlidir. T zaman aralığı çok küçük olursa çerçeve sayısını artıracığından hesaplama işlemi için yük artıracak ve uygulanabilirliği azalacaktır. T zaman aralığı çok büyük olduğunda ise tekrarlanan çerçevelerin tespit edilemeyecek ve yöntemin sonuçlarının doğruluğunu etkileyecektir. Çerçeve tabanlı kopyala - yapıştır sahteciliği tespiti akış diyagramı Şekil 6'da görülmektedir.



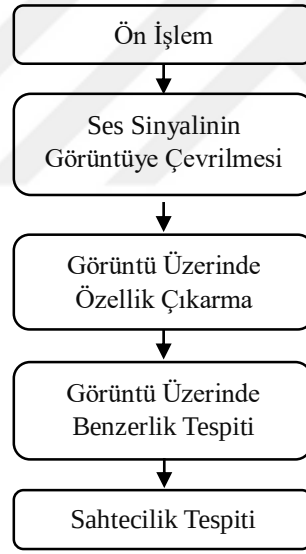
Şekil 6. Çerçeve tabanlı kopyala – yapıştır sahteciliği tespiti akış diyagramı

Yöntemin en önemli adımı ise bölünen çerçevelerin benzerliklerinin tespitidir. Bu yöntemde benzerlik ses dosyasındaki farklı çerçeve arasındaki benzerlik derecesi olarak tanımlanmıştır. İki farklı çerçeve arasındaki benzerlik derecesini hesaplamak için ayrık

zamanlı evrişim (Discrete Time Convulation- DTC) benzerlik fonksiyonu tanımlamıştır. Ses dosyasındaki her çerçeve arasındaki benzerlik hesaplanarak tekrarlanan sahte sesin tespiti için eşik belirlenir. İkili benzerlik hesabı sonucunda farklı olan çerçevelerin benzerlik derecesi 0'a yakın çok küçük bir değer çıkar. Sahte çerçevede ise sonuç maksimum değeri verir ve 1 olarak elde edilir. 1 değerini veren ikili karşılaştırmalar kopyalanmış çerçeve olarak tespit edilir.

### 1.5.3. Görüntü Tabanlı Kopyala-Yapıştır Sahteciliği Tespiti Yöntemleri

Ses sahteciliğinin tespitinde ses sinyali yerine sinyalin görüntüsü üzerinde görüntü işleme teknikleri ile benzerlik tespit eden yöntemlerdir. Bu tekniklerde öncelikle ses sinyalinin frekans değerleri iki boyutlu görsel haline dönüştürülür. Oluşan görüntü üzerinde farklı görüntü işleme teknikleri kullanılarak benzerlik tespit edilmeye çalışılır.



Şekil 7. Görüntü tabanlı kopyala – yapıştır sahteciliği tespiti akış diyagramı

Bu yöntem ses sinyalinin iki boyutlu görüntüsü için zaman-frekans boyutlu Mel spektogramını kullanmaktadır. Bu farkı yaklaşımda Üstübioğlu vd. (2023) elde edilen iki boyutlu ve renkli görüntü geliştirilen konvolüsyonel sinir ağı (Convolutional Neural Networks – CNN) sınıflandırma sistemi ile sahte ve orijinal olarak iki sınıfa ayrılır. Çalışmada görüntü üzerinde veri artırma tekniği olarak Kera's kütüphanesi kullanılarak spektogram görüntülerine uygulanmıştır. Geliştirilen konvüsyonel sinir ağı sistemi spektogram özelliklerini çıkartır ve özellik haritalarını oluşturur. Geliştirilen konvüsyonel

sinir ağı mimarisi bu özellikleri kullanarak sahte ve orijinal sesleri ayırt etmek için bir sınıflandırma gerçekleştirir.

Görüntüye dayalı yöntemler üzerine uygulanan ilk çalışmasında Üstübioğlu konvüsiyonel sinir ağı mimarisi kullanarak ses sinyalinin sahte olup olmadığına karar vermiştir. Ancak Bu yöntem hangi kelimenin kopyala-yapıştır olduğuna dair bilgi üretememektedir. Görüntüye dayalı geliştirdiği diğer yöntemde ise Üstübioğlu vd. (2024) bu soruna çözüm üreterek iki boyutlu mel spektogram görüntülerinden anahtar nokta ve bunların özellik tanımlayıcı vektörü çıkarım yaklaşımını benimsemiştir.

Görüntü üzerinde anahtar noktaları elde etmek için ölçek değişmez öznelik dönüşümü (Scale Invariant Feature Transform –SIFT) yöntemi kullanılmıştır. Anahtar nokta özellikleri çıkartılırken RGB kanallarından ayrı ayrı özellik çıkartılır. Kanallardan elde edilen anahtar noktalara ait özellikler birbirine benzer şekilde örtüşür. Örtüşen değerlerle olası hatalı eşleştirmelerin önüne geçilmiş olur. Yöntemde tespit edilen anahtar noktaların etrafındaki 8x8 lik blok sahte adayı olarak belirlenir ve 8 yöndeki komşuları işaretlenir. Sahte adayı olan bölgenin ortasındaki değerin komşuları ile korelasyon katsayısı hesaplanır. Korelasyon değeri 0,95 ten daha büyük bir değer çıkarsa sahte blok olarak işaretlenir. İşaretlenen bu noktaların bölgesi ikilik görsel olarak belirlenerek kopya ses olan bölgeler spektogram üzerinde işaretlenerek başlangıç ve bitiş süreleri belirtilir. X ekseninde belirtilen bölgeler ses sinyalindeki zamanı belirtir. Böylece tespit edilen kopya sinyalin yeri işaretlenmiş olur.

Ulutaş vd. (2022) konuşma sinyalinin spektrogramı oluşturmak için STFT kullanılmaktadır. Bunun için, ses sinyali önce 30 ms'lik çerçevelere bölüp her çerçeve Hamming penceresi ile çarpılmıştır. Daha sonra, tüm alt çerçevelere FFT uygulanmıştır. Elde edilen görsel üzerinde anahtar noktaları belirlemek için de hızlandırılmış sağlam özellikler (Speeded-Up Robust Features - SURF) algoritması uygulanmıştır. Anahtar noktalar arası benzerlik tespiti için de Öklid uzaklığı hesaplayarak kopyalanmış kelimeleri tespit etmiştir.

Benzer şekilde Üstübioğlu vd. (2024) uyguladığı yöntemde ise ses aktivitesi tespiti yöntemiyle görüntü tabanlı görsel benzerlik yöntemlerini bir arada kullanarak ataklara daha dayanıklı bir çalışma sunmuştur. Konuşmadaki kelimelerin tespit edilmesi için perde dizisine dayalı yöntemini kullanmıştır. Elde edilen kelimelerin sinyalde özelliklerini çıkartmak yerine kelimelerin her birinin iki boyutlu ve renkli mel spektogram görsellerine dönüştürülmüştür. Mel spektogramları elde etmek için kısa zamanlı hızlı fourier dönüşümünü (Short Time Fast Fourier Transform -STFT) kullanılmıştır.

Kelimeler görsele dönüştürüldükten sonra görsellerin birbiri ile benzerliklerini tespit etmek için yapısal benzerlik indeksi ölçüsü (Difference Structural Similarity Index Measure - DSSIM) yöntemi kullanılmıştır. Hesaplanan bu ölçüt 1 ile -1 arasında bir değer çıkar. 1 değeri benzerliği hesaplanan iki görüntünün aynı olduğunu, -1 değeri ise bu görüntülerin çok farklı olduğunu göstermektedir.

Yazıcı vd. (2022) de benzer şekilde STFT yöntemi ile Mel spektoramı elde etmiştir. İkili gradyan modeli (Binary Gadian P - BGP) ile özellik çıkarımı yaparak anahtar nokta yerine spektogram üzerinde blokların benzerliklerini tespit etmeye çalışmıştır.

Üstübioğlu görüntü tabanlı kopyala- yapıştır sahteciliği tespitinde spektogram yerine cochleagram görsellerini kullanarak yeni bir yöntem önermiştir. Pitch tabanlı YAAPT algoritması ile sesli ve sessiz segmentleri belirledikten sonra sesli segmentlerin cochleagram görselleri üretilir. Bir konuşmanın cochleagramı, konuşmanın zaman-frekans görüntüsündeki frekans bileşenlerine karşılık gelir. Önerilen yöntem, cochleagram görüntüler arasındaki benzerliği hesaplamak için SSIM metriğini kullanır. SSIM, bir görüntüden parlaklık, kontrast ve yapı olmak üzere üç özellik çıkarır. Görsel bilgi olarak bu özelliklerin birleşimini kullanır. Segmentlerden elde edilen benzerlik vektöründe en yüksek değer eşik değerine göre kopyala-yapıştır olarak işaretlenir.

## **1.6. Kopyala-Yapıştır Sahteciliği Tespiti Aşamaları**

Ses sinyallerinde yapılan kopyala-yapıştır sahteciliği, özellikle adli süreçlerde ve güvenilirliğin önemli olduğu durumlarda büyük sorunlara yol açabilir. Adli vakalarda delil olarak kullanılan ses dosyalarının üzerinde değişiklik yapılmamış dosya olması gerekir. Sahte ses dosyalarının üzerinde kelimelerin yerleri kopyalanıp başka bir bölgeye yapıştırılarak kopyala-yapıştır sahteciliği uygulanmaktadır. Bu tür sahtecilikleri tespit etmek için çeşitli yöntemler ve teknikler kullanılmaktadır. Kopyala-yapıştır sahteciliği tespit etme aşamaları aşağıda detaylandırılmıştır.

### **1.6.1. Sesli Segmentlerin Tespiti**

Ses sinyalindeki periyodik özellikleri ve tonlamanın belirlenmesi için perde (Pitch) yöntemi kullanılır. Pitch, ses sinyali işlemede çok önemli ve yaygın olarak kullanılan bir özelliktir. Bir konuşma sinyalinde, pitch ses sinyalinin bir ölçüsüdür ve ses titreşiminin

frekansını temsil eder. Bir ses sinyalinin temel frekansının ne kadar yüksek ya da düşük olduğunu belirtir (Yan vd., 2019). Pitch tabanlı kelime tespit algoritmasının adımları aşağıda detaylandırılmıştır.

- Konuşma Sinyali Ön İşlem

Konuşma sinyalinin mutlak değeri ve konuşma sinyalinin karesi. İlk olarak, sinyalin mutlak değerini ve karesini sırasıyla hesaplanır. Daha sonra bu iki tür doğrusal olmayan işlenmiş sinyali perde çıkarımı için kullanılır. Doğrusal olmayan bir işlem kullanmak, spektrumu düzeltmeye ve formantın gücünü azaltmaya yardımcı olabilir. Aynı zamanda, doğrusal olmayan bir işlem toplam ve fark frekansları da oluşturur, bu da eksik bir perdenin kısmen geri yüklenmesi için kullanılabilir.

- Frekans Alanında Perde Adayı Tahmini

Bu adımda, spektral analiz temelli perde adayları çıkarılır. İlk olarak, spektral harmonik korelasyon (SHC) kullanarak doğrusal olmayan konuşma sinyalinin spektrumundan perde adayları çıkarılır. SHC, (1) ile tanımlanır.

$$SHC(t, f) = \sum_{f'=L/2}^{L/2} \prod_{r=1}^{N+1} S(t, rf + f') \quad (1)$$

Burada  $S(t, f)$  t anı ve f frekansındaki büyüklük spektrumudur. L, frekanstaki spektral pencere uzunluğudur. N harmonik sayısı ve f frekanstır. SHC(t, f) her çerçeve için 0 ile 1.0 arasında normalize edilir. SHC(t, f), perdenin tam katlarında yüksek genliğe sahip olacaktır, bu da konuşma sinyalinin frekans bölgesinde perde adaylarını bulmak için faydalıdır.

Temel frekans üzerinde hesaplama yapan pitch algoritmasında NLFER, sesli ve sessiz segmentleri ayırt etmek için kullanılır. Frekans takibini kolaylaştırmak için normalleştirilmiş düşük frekans enerji oranı (NLFER), (2) gibi hesaplanır.

$$NLFER(t) = \frac{\sum_{f=2xF_{0\_min}}^{F_{0\_max}} S(t, f)}{\frac{1}{T} \sum_{t=1}^T \sum_{f=2xF_{0\_min}}^{F_{0\_max}} S(t, f)} \quad (2)$$

Burada N çerçeve sayısıdır.  $F_{0\_min}$  ve  $F_{0\_max}$  perde aralığını belirtir.  $S(t, f)$  t anı ve f frekansındaki konuşma sinyalinin spektrumudur. Sesli çerçeveler yüksek bir NLFER'a, sessiz çerçeveler ise düşük bir NLFER'a sahiptir.



- Zaman Alanında Perde Adayı Tahmini

Bu adımda, hem orijinal sinyallerden hem de doğrusal olmayan sinyallerden zaman bölgesinde perde adayları çıkarılır. Kısa süreli otokorelasyon fonksiyonu, zaman bölgesinde geleneksel bir perde izleme yöntemidir. Burada, normalize edilmiş çapraz korelasyon fonksiyonu (Normalized Cross Correlation Function - NCCF) olarak adlandırılan modifiye edilmiş bir otokorelasyon fonksiyonu kullanılmıştır. Bu fonksiyon (3) ile hesaplanır.

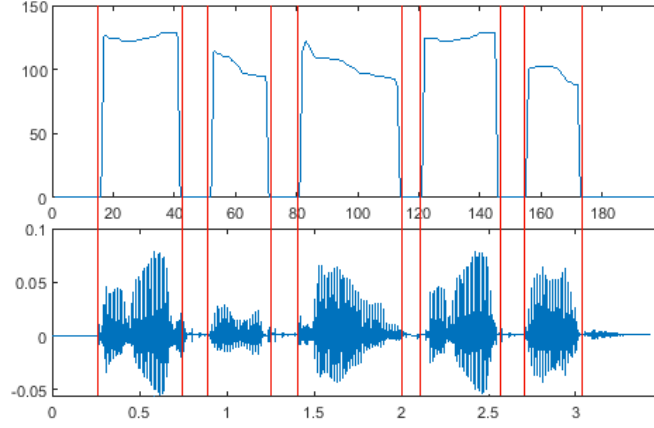
$$NCCF(k) = \frac{1}{\sqrt{e_0 e_k}} \sum_{n=0}^{L-K_{max}} s^2(n) \quad (3)$$

Burada L çerçeve uzunluğunu ifade eder. Kmin ve Kmax, perde arama aralığını kapsamak için kullanılan gecikme değerleridir. NCCF, [-1,1] aralığında kendiliğinden normalize edilir ve NCCF sonucu, perdenin tam katlarına eşit gecikme değerlerinde 1 olacaktır. Yan vd. (2019)'e göre, NCCF geleneksel otokorelasyon fonksiyonundan daha iyidir.

- Dinamik Programlama ile Perde Belirleme

2. ve 3. adımlarda, bir konuşma üretiminin aralığı boyunca perde adayları ve bunlarla ilişkili değerlendirme matrisi elde edildi. En iyi perde adaylarını seçmek için bir geçiş maliyet fonksiyonu ve yerel maliyet fonksiyonu tasarlandı. Son olarak, bu iki maliyet türünü hesaplayıp en aza indirerek en iyi perde adaylarını seçmek için standart bir Viterbi algoritması kullanılmıştır (Zahorian & Hu, 2008).

Copy- Move sahteciliğini tespit edebilmek için ses sinyalindeki konuşma ön işlemde sonra pitch tabanlı özellik çıkarım yöntemi ile sinyal üzerindeki sesli ve sessiz bölgeler olarak belirlenir. Sesli olarak işaretlenen segmentler serisi kelimeleri oluşturur. Şekil 8'de örnek sinyalde pitch yöntemi ile tespit edilmiş kelimeler görülmektedir. Örnek sinyalle tespit edilen kelimeler kırmızı çizgilerle işaretlenmiştir.



Şekil 8. Ses sinyalindeki sesli bölgelerin pitch yöntemi ile belirlenmesi

Pitch ile elde edilen özellik vektöründe ses sinyalinin konuşma olan segmentlerinde 0'dan büyük değerler dönerken konuşma olmayan segmentlerde ise 0 değeri döner. Sessiz olarak 0 ile işaretlenmiş iki indis arası bir kelimenin özellik değerlerini oluşturur. Tablo 1'de örnek bir sinyalin sesli bölümlerin tespit edilmiş vektör değerleri örnek olarak görülmektedir.

Tablo 1. Pitch yöntemi ile sinyalde elde edilen öz vektör değerleri

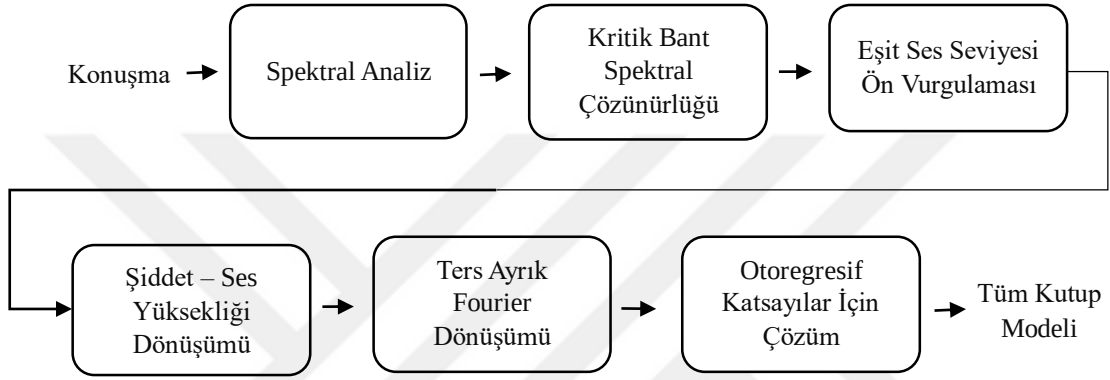
|                |    |    |    |        |        |        |        |        |        |        |    |    |
|----------------|----|----|----|--------|--------|--------|--------|--------|--------|--------|----|----|
| Vektör indisi  | 14 | 15 | 16 | 17     | 18     | 19     | 20     | 21     | 22     | 23     | 24 | 25 |
| Özellik değeri | 0  | 0  | 0  | 125.98 | 125.98 | 124.03 | 124.03 | 124.03 | 124.03 | 122.37 | 0  | 0  |

### 1.6.2.Özellik Çıkarma

Ses sinyalindeki kopyala-yapıştır sahteciliğinin tespitinde izlenen yöntem pitch yöntemi ile belirlenmiş kelimelerin birbiri ile karşılaştırılması ile en benzer kelimelerin tespit edilmesidir. Kelimelerin özelliklerinin çıkartılması kopyala-yapıştır sahteciliğinin tespit edilmesi için önemli bir adımdır. Tespit edilen kelimelerin özellik vektörleri ses tanımda yaygınca kullanılan yöntemler uygulanarak elde edilmiştir. Kopyala - yapıştır sahteciliğinin tespit edilmesinde özellik çıkarımından elde edilen öz vektörler yöntemin dayanıklılığı ve performansı açısından önemlidir. Aşağıda bu yöntemlerden bazıları ele alınarak uygulama esasları detaylandırılmıştır.

- Algısal Doğrusal Tahmin Özellik Çıkarma Yöntemi (Perceptual Linear Prediction – PLP)

PLP yöntemi konuşma tanımada yaygınca kullanılan çok bilindik frekans tabanlı bir özellik çıkarım yöntemidir. PLP yöntemi, insan işitmenin bilinen bir çok özelliğini pratik mühendislik yaklaşımları ile simüle ederek elde edilen işitme benzeri konuşma spektrumunu otoregresif tüm-kutup modele yaklaştırılır (Hermansky, 1990). PLP nin genel akış diyagramı Şekil 9’da görülmektedir. PLP yönteminin aşamaları aşağıda detaylandırılmıştır.



Şekil 9. Genel PLP Akış Diyagramı

### 1. Spektral Analiz

Konuşma sinyali çerçevelerini ağırlıklandırmak için Hamming Penceresi (Hamming Window) kullanılır. Hamming penceresi (4) ile hesaplanır. N çerçeve boyutunu ifade eder.

$$W(n) = 0.54 + 0.46 \cos[2\pi n / (N - 1)] \quad (4)$$

Çerçevenin tipik olarak varsayılan boyutu 20 ms. Ayırık fourier dönüşümü (Discrete Fourier Transform - DFT) çerçevelenmiş konuşma segmentlerini frekans alanına dönüştürür. Burada dönüşüm için hızlı fourier dönüşümü (Fast Fourier Transform - FFT) kullanılır. Kısa süreli konuşma spektrumunun gerçek ve sanal bileşenlerinin kareleri alınır ve toplanarak kısa süreli güç spektrumu (5) de ifade edildiği gibi elde edilir.

$$P(\omega) = Re[S(\omega)]^2 + Im[S(\omega)]^2 \quad (5)$$

## 2. Kritik Bant Spektral Çözünürlüğü

İnsan işitmesinin belirli bir frekanstaki sesleri ayırt etme yeteneğini simüle etmek için kritik bant spektral çözünürlüğü kısa süreli güç spektrumunu (4) kullanılarak bark frekansına dönüştürmek için hesaplanır. Bark ölçeği, insan kulağının algıladığı şekilde frekansları temsil eder. Bu adımda kritik bant filtreleme de uygulanır. Spektrum  $P(\omega)$ , frekans eksenini  $\omega$  boyunca (6) ile bark frekansına bükülür.  $\omega$  açısal frekansı rad/s ile göstermektedir.

$$\Omega(\omega) = 6 \ln \{ \omega / 1200\pi + [(\omega / 1200\pi)^2 + 1]^{0.5} \} \quad (6)$$

Elde edilen bükülmüş spektrum daha sonra, (7) uygulanarak tahmin edilen simüle edilmiş kritik bant maskeleme eğrisinin  $\Psi(\Omega)$  spektrumu ile kıvrılır.

$$\Psi(\Omega) = \begin{cases} 0, & \text{for } \Omega < -1.3 \\ 10^{2.5(\Omega-0.5)}, & \text{for } -1.3 \leq \Omega \leq -0.5 \\ 1, & \text{for } -0.5 \leq \Omega \leq 0.5 \\ 10^{-(\Omega-0.5)}, & \text{for } 0.5 \leq \Omega \leq 2.5 \\ 0, & \text{for } \Omega > 2.5 \end{cases} \quad (7)$$

Ses sinyalinin frekans dizisi  $\Psi(\Omega)$  ve  $P(\omega)$  arasındaki ayrık konvülyasyon (8) ile hesaplanarak kritik bant güç spektrumunu temsil eden bir dizi elde edilir. Nispeten geniş kritik bant maskeleme eğrisi  $\Psi(\Omega)$  ile kısa süreli güç spektrumu  $P(\omega)$  'in çarpıştırılması,  $\theta(\Omega)$ 'in spektral çözünürlüğünü orijinal  $P(\omega)$ 'e kıyasla önemli ölçüde düşürür.

$$\theta(\Omega_i) = \sum_{i=-1.3}^{2.5} P(\Omega - \Omega_i) \cdot \Psi(\Omega) \quad (8)$$

Örnekleme aralığının tam değeri, spektral örneklerin tamamı analiz bandını kapsayacak şekilde seçilir.

## 3. Eşit Ses Seviyesi Ön Vurgulaması

Örneklenen  $\theta[\Omega(\omega)]$ , simüle edilmiş eşit ses yüksekliği eğrisi ile ön vurgulamadan geçirilir.

$$\Xi[\Omega(\omega)] = E(\omega) \cdot \theta[\Omega(\omega)] \quad (9)$$

$E(\omega)$  fonksiyonu, insan işitmesinin farklı frekanslardaki eşit olmayan duyarlılığının bir yaklaşımıdır ve işitme hassasiyetini yaklaşık 40 dB seviyesinde simüle eder. Örneklerin ilk ve son değerleri yakın komşularının değerlerine eşitlenir. Hermansky tarafından önerilen yaklaşım (10) de ifade edilmiştir.

$$E(\omega) = \frac{[(\omega^2 + 56.8 \times 10^6)\omega^4]}{(\omega^2 + 6.3 \times 10^6)^2 \times (\omega^2 + 0.38 \times 10^9)(\omega^2 + 9.58 \times 10^{26})} \quad (10)$$

#### 4. Şiddet-Ses Yüksekliği Dönüşümü

Tüm-kutup modellemesinden hemen önceki son işlem, kübik kök genlik sıkıştırmasıdır. Bu işlem, duyma güç yasası denilen bir yaklaşım olup, sesin şiddeti ile algılanan ses yüksekliği arasındaki doğrusal olmayan ilişkiyi (11) ile simüle eder.

$$\Phi(\Omega) = \Xi[\Omega]^{0.33} \quad (11)$$

Bu işlem, psiko-fiziksel eşit-ses yüksekliği ön vurgusu ile birlikte, kritik bant spektrumlarının spektral genlik değişimini de azaltır. Böylece, sonraki tüm-kutup modellemesi nispeten düşük bir model düzeni kullanılarak yapılabilir.

#### 5. Ters Ayırık Fourier Dönüşümü

Cepstral katsayıları elde etmek için ters ayırık Fourier dönüşümü (IDFT) kullanılır. Logaritmik spektral özelliklerin cepstral analiz ile zaman alanına geri dönüşümünü sağlar. Bu dönüşüm, ses sinyalinin spektral özelliklerinin zaman alanında yeniden yapılandırılmasını ve insan işitme sistemine uygun bir şekilde temsil edilmesini sağlar.

#### 6. Otoregresif Katsayılar İçin Çözüm

PLP analizinin son işleminde, tüm-kutup model, ses sinyalinin sadece kutuplarla temsil edildiği (sıfırları olmayan) bir spektral modeldir.  $\Phi(\Omega)$  tüm-kutup modelinin spektrumu ile yaklaşık olarak temsil edilir. Bu model, ses sinyalinin otokorelasyon fonksiyonuna dayanarak spektral yoğunluğunu tahmin eder. Burada ters DFT, ters FFT'den daha iyi bir seçimdir çünkü yalnızca birkaç öz-korelasyon değerine ihtiyaç vardır.

- Bağlı Spektrum Algısal Doğrusal Öngörülü Özellik Çıkarma Yöntemi (Relative Spectra Perceptual Linear Prediction – Rasta PLP)

PLP hesaplamalarında, görelî spektrum (RASTA) filtresinin konuşma tanıma performansını artırdığı bilinmektedir. RASTA filtresi, sinyaldeki hem yavaş hem de hızlı

değişimleri baskılamak için kullanılır. İnsan işitmesinin yavaş değişen seslere karşı duyarsızlığı, insanların iletişim ortamının frekans karakteristiğindeki yavaş değişimlere fazla dikkat etmemelerinin ya da sürekli arka plan gürültüsünün insan konuşma iletişimini ciddi şekilde bozmamasının sebebini açıklayabilir (Hermansky & Morgan, 1994).

Ancak, insan algısına dair deneysel kanıtlar sınırlı destek sağlasa bile, konuşma sinyalindeki yavaş değişen bileşenlerin baskılanması mühendislik açısından faydalıdır. Konuşma analizini yavaşça değişen veya sabit faktörlere karşı daha az hassas hale getirmek için, PLP konuşma analizinde geleneksel kritik bantlı kısa süreli spektrumu, her frekans kanalının filtre geçişli olduğu ancak sıfır frekansında keskin bir spektral sıfır noktasına sahip bir filtre ile elde edilen spektral tahmin ile değiştirilmiştir. Bu işlem sayesinde her bir frekans kanalındaki sabit veya yavaş değişen bileşenler baskılandığı için, yeni spektral tahmin, kısa süreli spektrumdaki yavaş değişimlere karşı daha az hassas hale gelir.

Bu tekniğin kullanılmasının temel amacı, PLP'yi göreceli spektral metodolojisi sayesinde doğrusal spektral bozulmalara karşı daha dayanıklı hale getirmektir. RASTA-PLP hesaplama adımları aşağıdaki gibidir:

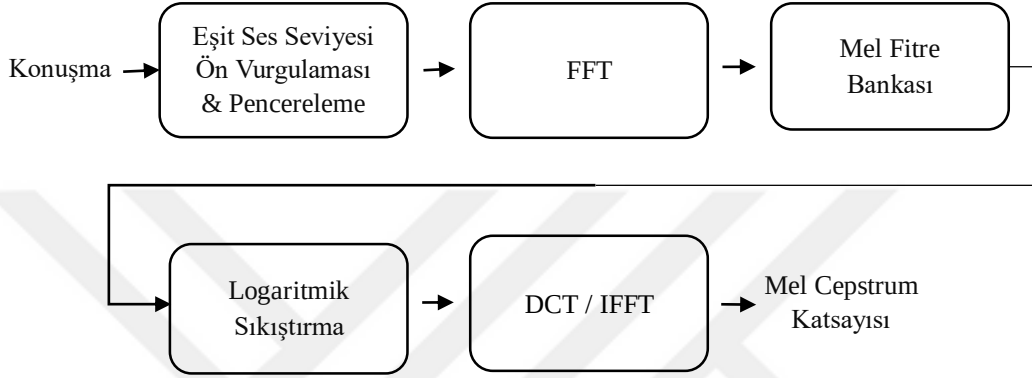
1. PLP'de olduğu gibi Kritik Bant Spektral Çözünürlüğünü hesaplanır.
2. Logaritmik kritik bant spektrumunun zamansal türevini, beş ardışık spektral değerden geçen bir regresyon çizgisi kullanılarak tahmin edilir.
3. Doğrusal olmayan hesaplama bu alanda yapılabilir.
4. Birinci dereceden sonsuz impuls cevabı sistemini kullanarak, logaritmik kritik bant zamansal türevini yeniden entegre edilir.
5. Konvansiyonel PLP ile uyumlu olarak, eşit seslilik eğrisini ekleyin ve 0.33 ile çarparak işitme güç yasası taklit edilir.
6. Bu göreceli logaritmik spektrumun ters logaritmasını alın ve böylece göreceli işitsel spektrumu elde edilir.
7. Konvansiyonel PLP tekniğini takip ederek, bu spektrumun tam kutuplu bir modeli hesaplanır.

- Mel Frekanslı Cepstrum Katsayısı Özellik Çıkarma Yöntemi (Mel Frequency Cepstral Coefficients – MFCC)

Konuşma tanımda kullanılan en yaygın ve baskın özellik çıkarma yöntemi Mel-Frekanslı Cepstrum Katsayılarını (MFCC) hesaplamaktır. MFCC'ler, insan kulağının algılama biçimine dayanan mel ölçeğini kullanan frekans alanında konuşma tanıma için kullanılan en popüler özellik çıkarma tekniklerinden biridir. Frekans alanı özellikleri olarak

değerlendirilen MFCC'ler, zaman alanı özelliklerinden çok daha hassas sonuçlar verir (Dave, 2013).

MFCC, konuşmayı dinlerken insanlar tarafından kullanılan parametrelere benzer parametreleri konuşmadan çıkaran bir ses özelliği çıkarma tekniğidir. Aynı zamanda diğer tüm bilgileri önemsizleştirir. MFCC özellik çıkarımının genel akış diyagramı Şekil 10'da gösterilmiştir. İşlem basamakları aşağıda detaylandırılmıştır.



Şekil 10. Genel MFCC Akış Diyagramı

#### 1. Eşit Ses Seviyesi Ön Vurgulaması & Pencereleme

Konuşma sinyalindeki yüksek frekanslı bileşenleri güçlendirerek spektral enerjiyi eşitlemek için ön vurgulama yapılır. Ön vurgulamadan sonra konuşma sinyali bir pencere fonksiyonu kullanılarak çerçevelere ayrılır. Daha sonra her zaman dilimi, kenarlardaki kesintileri ortadan kaldırmak için Hamming penceresi ile pencereye alınır. Bu işlem için genellikle uzunluğu 20-30 ms arasında değişen bir Hamming penceresi kullanılır. Uzunluğu  $n$  olan bir Hamming penceresinin filtre katsayıları  $W(n)$  ifade (12)'a göre hesaplanır.

$$W(n) = 0.54 - 0.46 \cos \frac{2\pi n}{N-1}, 0 \leq n \leq N-1 \quad (12)$$

$N$  toplam örnek sayısı,  $n$  ise mevcut örnek sayısını ifade eder.

#### 2. FFT

Pencere uygulamasından sonra, her zaman dilimi için bir FFT hesaplanarak, sinyalin zaman alanındaki frekans bileşenleri çıkarılır. FFT, işleme hızını artırmak için kullanılır.

### 3. Mel Filtre Bankası

Daha sonra fourier dönüştürülmüş çerçeveye logaritmik mel ölçekli filtre uygulanır. Bu ölçek, yaklaşık olarak 1 kHz'e kadar doğrusaldır ve daha yüksek frekanslarda logaritmik hale gelir. Konuşma frekansı ile mel ölçeği arasındaki ilişki (13) ile ifade edilir.

$$Frekans(Mel \text{ Ölçekli}) = [2595 \log 1 + f(Hz)/700] \quad (13)$$

MFCC'lerde, mel ölçekli filtre bankaları kullanılır. Bu filtre bankalarında, yüksek frekanslı filtrelerin bant genişliği düşük frekanslı filtrelerden daha geniştir, ancak zamansal çözünürlükleri aynıdır.

### 4. Logaritmik Sıkıştırma

Verilen bir sinyal yoğunluğunun algılanan gürlüğünü modellemek için filtre çıkışları logaritmik fonksiyon ile sıkıştırılır. Mel Frekans Bankası'ndan elde edilen her bir enerji değeri, logaritmik fonksiyonla dönüştürülür. Bu işlem, büyük enerji değerlerini sıkıştırarak küçük enerji değerleriyle daha karşılaştırılabilir hale getirir.

### 5. DCT

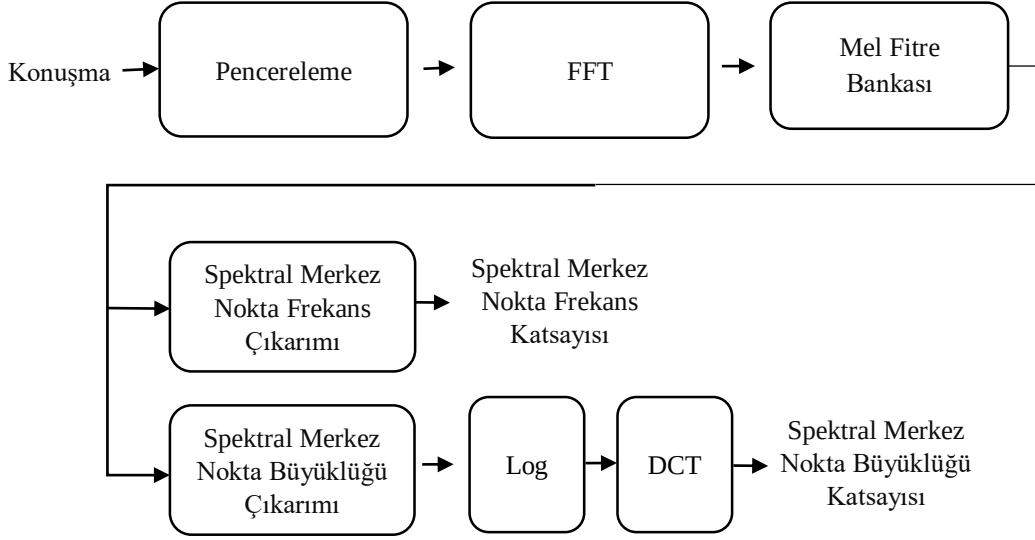
Son adımda, filtre bankasının çıktılarına DCT uygulamaktır. DCT, elde edilen katsayıları önemlerine göre sıralar. 0. katsayı güvenilir olmadığı için hariç tutulur.

Konuşmanın her bir karesi için bir dizi MFCC hesaplanır. Bu katsayı seti, bir akustik vektör olarak adlandırılır ve konuşmanın fonetik açıdan önemli özelliklerini temsil eder. Bu vektör, konuşma tanıma sistemlerinde daha ileri analiz ve işleme için yararlanılır.

- Spektral Merkez Nokta Büyüklüğü, Spektral Merkez Nokta Frekansı Özellik Çıkarma Yöntemi (Spectral Centroid Magnitude , Spectral Centroid Frequency – SCF )

SCM yöntemine göre özellik çıkarım aşamaları Şekil 11'de gösterildiği gibi hesaplanır.  $n \in [0, N-1]$  için  $s[n]$  bir konuşma çerçevesini yani konuşma örneğini,  $S[f]$  ise bu çerçevenin spektrumunu temsil eder. Daha sonra,  $S[f]$ , her biri alt sınır frekansı ( $lk$ ) ve üst sınır frekansı ( $uk$ ) ile tanımlanan  $K$  alt banda bölünür. Filtrenin frekans örneklemeli frekans cevabı  $wk[f]$ 'dir (Kua vd., 2010).





Şekil 11. Genel SCF ve SCM Akış Diyagramı

Spektral Merkez Nokta Frekansı (Spectral Centroid Frequency - SCF), belirli bir alt bant için ağırlıklandırılmış ortalama frekans olarak tanımlanır. Bu ağırlıklandırmada, her bir frekans bileşeninin normalize edilmiş enerjisi kullanılır.

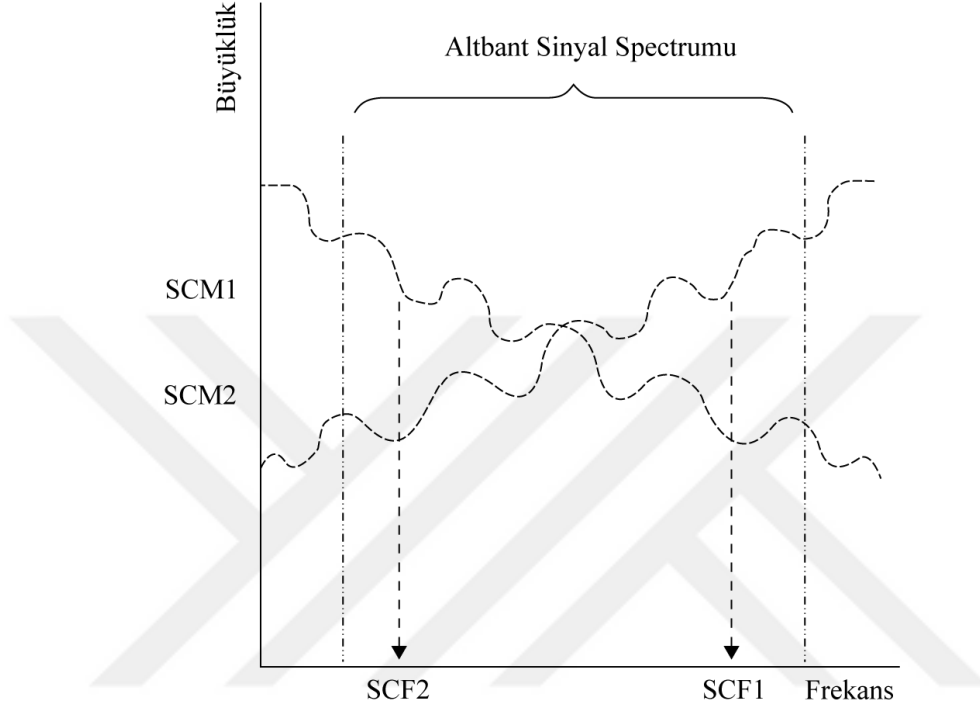
Bu ölçüm, her alt bandın ağırlık merkezini yakaladığı için, komşu alt bantlarda tepe noktası olarak görülen formantların yaklaşık konumunu tespit edebilir. Ancak, alt bandın ağırlık merkezi, özellikle dar bant genişlikleri için ses kaynağının ürettiği harmonik yapı ve ses perdesi frekanslarından da etkilenir. Bu nedenle, SCF özelliği ses perdesi ve harmonik yapıda meydana gelen değişikliklerden etkilenir.  $k$ . alt bant için spektral sentroid frekansı  $F_k$  (14) gibi tanımlanır.

$$F_k = \frac{\sum_{f=1_k}^{U_k} f |S[f]| \omega_k[f]}{\sum_{f=1_k}^{U_k} |S[f]| \omega_k[f]} \quad (14)$$

SCF frekans tabanlı bir özellik olduğundan, ikinci dereceden tüm-kutup FM yönteminin çıkarılmasından esinlenerek bark ölçekli gabor filtre bankasını kullanarak SCF değerleri hesaplanmıştır. Ek olarak, konuşma güç spektrumu ve filtre bankası frekans cevabını daha iyi yakalamak için FFT noktalarının sayısını bir derece artırılmıştır. Bu işlemin SCF performansı üzerinde önemli bir etkisi olduğu görülmüştür.

Spektral Merkez Nokta Büyüklüğü (Spectral Centroid Magnitude - SCM) belirli bir alt banttaki ağırlıklı ortalama büyüklüğü ifade eder. Ağırlıklandırmada ise her frekans

bileşeninin kendi büyüklüğü kullanılır ve (14)'deki gibi hesaplanır. SCM, birinci dereceden bir yaklaşımla, alt banttaki enerji dağılımını yakalar. Bu durum, Şekil 12'de ortalama enerjileri aynı olan iki farklı sinyal üzerinden gösterilmiştir (Kua vd., 2010).



Şekil 12. İki Farklı Örnek Sinyalin SCF ve SCM grafiği

İki farklı örnek alt bant sinyalinin (sürekli çizgi (1) ve kesikli çizgi (2)) SCF ve SCM çıkarımı gösterilmiştir. Her iki sinyalin ortalama enerjileri eşittir. SCM hesaplamasında frekans ağırlıklandırması kullanıldığından, SCM1 değeri SCM2 değerinden daha büyüktür.

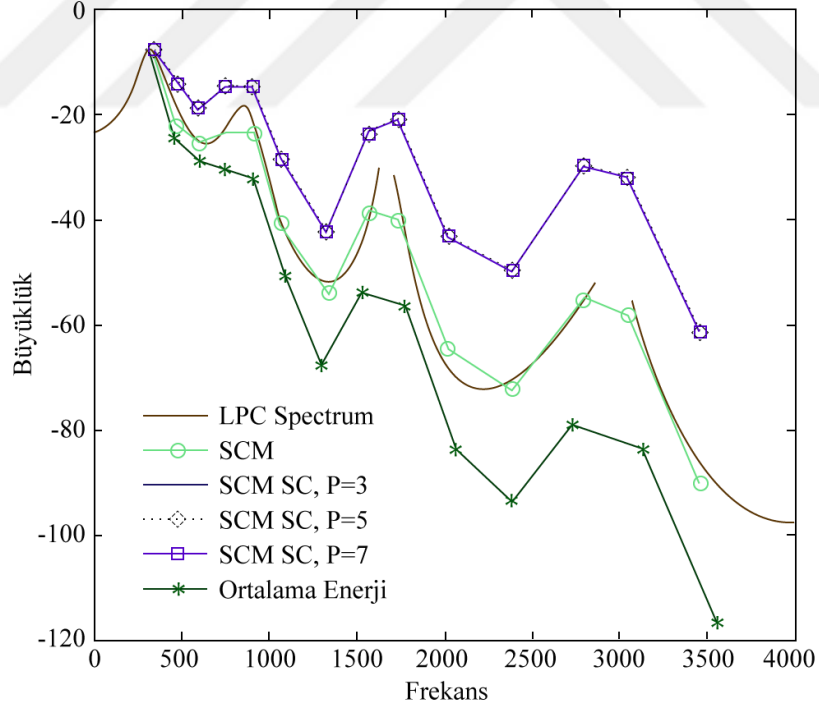
Ağırlıklandırma fonksiyonu nedeniyle, her iki sinyal de farklı SCF ve SCM değerleri ile temsil edilecektir. Ağırlıklandırma fonksiyonunun eğimindeki değişikliğe de dikkat edilmelidir; bu, farklı alt bant genişlikleri için özellik elemanı varyanslarında farklılıklar yaratır. Ortalama enerji, Denklem (15)'de  $f=1$  değerini atayarak hesaplanabilir. Spektral Merkez Nokta Büyüklüğü, spektral merkez frekans konumundaki büyüklük olduğundan, konuşmacı tanımada faydalı olan formant bilgilerini de taşırır.

$$F_k = \frac{\sum_{f=1_k}^{U_k} f |S[f] \omega_k[f]|}{\sum_{f=1_k}^{U_k} f} \quad (15)$$

Denklem (13)'e bakıldığında, paydada kullanılan terim, SCF'den farklı olarak konuşmacıya bağımlı değildir. SCM'nin konuşmacı bağımlılığını artırmak için, her alt banttaki en önemli P adet frekans bileşeni kullanılarak alternatif bir formülasyon önerilebilir. Bu formülasyon (16)'deki gibidir.

$$M_{sc,k} = \frac{\sum_{f'_i \in I_k} f'_i |S[f'_i]| \omega_k[f'_i]}{\sum_{f'_i \in I_k} f'_i} \quad (16)$$

Burada her bir frekans bileşeni için ağırlık hesabı yapılan  $I_k$ ,  $|S[f]| \omega_k[f]$  ifadesinin en büyük P değerine karşılık gelen frekans kümesini temsil eder. Bu alternatif SCM yöntemine, Önemli Bileşenlere Dayalı SCM (SCM\_SC) denir. Şekil 13'te görüldüğü gibi, SCM bir SCF grafiğine karşı çizildiğinde, her alt bandın merkez frekansına karşı çizilen ortalama enerjiye göre LPC spektrumuna daha iyi bir yaklaşım sağlar (Kua vd., 2010).



Şekil 13. İki Farklı Örnek Sinyalin LPC, SCF ve SCM grafiği

### 1.6.3. Benzerlik Hesaplama

Ses kaydından kelimeler tespit edildikten sonra farklı özellik çıkarım yöntemleri ile kelimelere ait özellik vektör değerleri hesaplanır. Tespit edilen bu kelimelerin aralarındaki benzerlik ilişkisine bakılır. İstatistiksel hesaplamalarla sinyalde tespit edilen tüm kelimelerin diğer kelimelerle olan yakınlık değerleri hesaplanır. Bu tez çalışmasında uygulanan 4 farklı özellik çıkarım yönteminden elde edilen kelimelere ait öz vektörler 3 farklı istatistiksel hesaplama yapılarak sınıflandırılmıştır. Bunlar hesaplamalar Öklid, korelasyon ve DTW'dir.

Tespit edilen kelimeler diğer tüm kelimelerle benzerliği hesaplanırken kelime sayısına göre bir matris elde edilir. Bu matrisin örneğin 1- 2 karşılaştırması ile 2-1 karşılaştırması aynı sonucu vereceğinden matrisin köşegen matrisin üstünde kalan üst üçgensel matris elemanları karşılaştırılarak hesaplama yapılmıştır. Böylece işlem süresi yarıya indirilmiştir.

- Korelasyon Katsayısı

Korelasyon analizi, genellikle veri bilimi, istatistik ve sosyal bilimlerde kullanılır. Korelasyon katsayısı iki değişken arasındaki ilişkiyi ölçmek için kullanılan bir istatistiksel ölçüdür. İki değişken arasındaki ilişkinin doğasını anlamak için önemli bir araç olarak bilinir. Değişkenlerin nasıl birbirine bağlı olduğunu görmek ve analiz etmek için önemli bir tekniktir. Korelasyon katsayısı genellikle Pearson korelasyon katsayısı olarak adlandırılır ve istatistiksel literatürde sıkça kullanılır. Pearson korelasyon katsayısı, -1 ile +1 arasında bir değer alır ve denklem (17) ile hesaplanır (Akdeniz & Becerikli, 2021).

$$r = \frac{\sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_{i=1}^n (x_i - \bar{x})^2 + \sum_{i=1}^n (y_i - \bar{y})^2}} \quad (17)$$

Pearson korelasyon katsayısı, iki değişken arasındaki doğrusal ilişkiyi ölçer.  $r = 1$  ise tam pozitif doğrusal ilişki,  $r = -1$  ise tam negatif doğrusal ilişki,  $r = 0$  ise doğrusal bir ilişkinin olmadığını gösterir.  $x_i$  ve  $y_i$  sırasıyla iki ses sinyalinden elde edilen öz vektör değişkenin değerleri,  $\bar{x}$ ,  $\bar{y}$  ise bu öz vektörlerin ortalamaları,  $i$  ise öz vektörlerin eleman sayısıdır. Değerler arasında bir ilişki varsa,  $r$  değeri bu ilişkinin gücünü ve yönünü belirtir. Değişkenler arasında daha güçlü bir ilişki varsa, korelasyon katsayısı +1 veya -1'e daha yakın olacaktır. Örneğin, +0.8 veya -0.8 gibi değerler güçlü bir ilişkiyi ifade ederken, +0.2 veya -0.2 gibi değerler daha zayıf bir ilişkiyi ifade eder.

Korelasyon katsayısı, aynı zamanda iki değişken arasındaki ilişkinin lineer olup olmadığını belirlemek için de kullanılır. Ancak önemli bir nokta, korelasyon katsayısının korelasyonun tüm türlerini ölçmediğidir. Örneğin, iki değişken arasında karmaşık bir ilişki varsa (örneğin, doğrusal olmayan ilişkiler), Pearson korelasyon katsayısı doğru bir ölçüm olmayabilir.

- Öklid Uzaklığı

Öklid, genellikle "Öklid'in mesafesi" veya "Öklidyen mesafesi" olarak bilinen kavramdır. Matematikte, Öklid'in mesafesi iki nokta arasındaki doğrusal mesafeyi ölçmek için kullanılan bir metriktir. İki nokta  $(p_1, p_2 \dots p_n)$ ,  $(q_1, q_2 \dots q_n)$  arasındaki Öklidyen mesafesi (18)'a göre hesaplanır (Thakur & Sahayam, 2013).

$$d = \sqrt{(p_1 - q_1)^2 + (p_2 - q_2)^2 + \dots + (p_n - q_n)^2} \quad (18)$$

Bu formül, iki boyutlu uzayda (örneğin bir düzlemde) iki nokta arasındaki doğrusal uzaklığı belirler. N-boyutlu bir uzayda bir nokta olarak temsil edilir. Bu noktadaki her bir koordinat, verinin bir özelliğini (özniteliğini) ifade eder.

Bilgisayar biliminde sınıflandırma için Öklid uzaklığı, özellikle makine öğrenimi ve veri madenciliği gibi alanlarda kullanılır. Örneğin, bir sınıflandırma probleminde elimizde bir veri kümesi var ve her bir örnek belirli özelliklere sahip (örneğin boy, ağırlık gibi). Bir yeni örneği bu veri kümesindeki diğer örneklerle karşılaştırmak istiyoruz ve onu en yakın komşu (nearest neighbour) yöntemiyle sınıflandırmak istediğimizde Öklid uzaklığı burada önemli bir rol oynar. Yeni örneğin diğer örneklerden ne kadar uzakta olduğunu Öklid uzaklığı kullanarak hesaplayabiliriz. Bu mesafeye göre en yakın komşuları bulabilir ve yeni örneği onların sınıfına atayabiliriz. Öklid uzaklığı bilgisayar biliminde özellikle sınıflandırma ve kümeleme gibi problemlerde, benzerlik ve uzaklık hesaplamalarında kullanılan temel bir kavramdır.

- Dynamic Time Warping

DTW zaman serileri veya diğer zamanla değişen veriler arasında benzerlik ölçmek için kullanılan bir yöntemdir. Genellikle zaman serilerinin zaman ölçeği veya hızı farklı olduğunda veya zaman içindeki küçük varyasyonlar nedeniyle düz bir zaman ölçeğinde karşılaştırılması zor olduğunda kullanılır. Konuşma hizalamasının farklı zamanlamaları, konuşma tanımda mesafe ölçümünün temel bir sorunudur. Özellikle konuşma sinyalleri arasındaki benzerlikleri ölçmek ve konuşma tanıma sistemlerinde kullanılan bir yöntemdir.

DTW, konuşma sinyallerinin zaman içindeki varyasyonlarını hesaba katarak benzerlik ölçmek için idealdir, çünkü konuşma sinyalleri genellikle zaman ölçeği ve hız farklılıkları gösterebilir.

DTW, sadece zaman serilerinin uzunluklarını değil, aynı zamanda şekillerini de karşılaştırır. Bu sayede, zaman serilerinin genel trendleri ve şekilleri arasındaki benzerlikler daha iyi anlaşılır.

Ses sinyalinden elde edilen özellik matrisleri arasındaki en iyi eşleşmeyi bulmak için dinamik programlama yöntemini kullanır. Bu yöntemde, her iki zaman serisinin her bir noktası için bir uzaklık matrisi oluşturulur. Bu matriste, birinci zaman serisinin her bir noktası ile ikinci zaman serisinin her bir noktası arasındaki Öklid uzaklığı hesaplanır. Daha sonra, bu matris üzerinde en düşük maliyetli bir yol bulunur. Bu yol, iki zaman serisi arasındaki en iyi eşleşmeyi temsil eder. DTW'nin temel formülü, (19) gibi ifade edilebilir (Mohan & Babu, 2014).

$$D(i, j) = d(i, j) + \min \begin{cases} D(i-1, j), \\ D(i, j-1), \\ D(i-1, j-1) \end{cases} \quad (19)$$

Formülde yer alan  $D(i, j)$ ; i'nci ve j'nci noktalar arasındaki en düşük kümülatif uzaklığı temsil eder.  $d(i, j)$  ise i'nci ve j'nci noktalar arasındaki Öklid uzaklığını ifade eder. DTW sonucu olarak elde edilen minimum maliyet, iki zaman serisi arasındaki benzerlik ölçüsü olarak yorumlanabilir. Daha düşük maliyet, iki zaman serisinin daha benzer olduğunu, yani daha yakın olduğunu gösterir.

#### 1.6.4. Sahte Segmentlerin İşaretlenmesi

Tespit edilen kelimelere ait öz vektörleri arasında benzerlik hesaplaması yukarıda anlatılan yöntemlere göre hesaplandıktan sonra kullanılan yöntemle göre sahte segmentler belirlenir. Pearson korelasyon katsayısı iki kelimenin benzerlik oranını istatistiksel olarak büyüklüğünü hesaplamak için kullanılır. Kelimeler arasında sıra ile birbiri ile karşılaştırılırken eşik değerine bakılarak iki kelimenin kopya olup olmadığına karar verilir.

İki kelime arasında hesaplanan korelasyon katsayısı büyüdükçe kelimeler arası benzerlik ilişkisi kuvvetlenmektedir. Hesaplanan katsayı değeri küçüldükçe kelimeler arasında benzerlikten söz edemeyiz. Yapılan çalışmalarda uygulanan yöntem ve veri

tabanına göre empirik olarak belirlenen eşik değerine göre kelimelerin kopya olduğuna karar verilir.

PLP özellik çıkarım yöntemi ve yapılan korelasyon benzerlik hesaplaması ile yapılan çalışmalarda eşik değeri empirik olarak 0,95 alınmıştır. Kelimeler arası hesaplanan korelasyon katsayısı 0,95'ten daha büyük bir değerse kelimeler arası benzerlik ilişkisi büyük olduğu için bu kelimeler kopya olarak işaretlenir.

Yukarıda anlatılan Öklid uzaklığı benzerlik hesaplamasında ise kelimeler arası yakınlığa bakılmaktadır. Elde edilen öz vektörler arasındaki yakınlık derecesi hesaplanır. Tüm kelimeler arasındaki yakınlık dereceleri sıra ile kontrol edilir. Kelimeler arası Öklid uzaklığı ne kadar küçükse kelimelerin birbirine benzer olduğu anlaşılır. PLP özellik çıkarım yöntemi ile yapılan çalışmalarda Öklid uzaklığı için empirik olarak eşik değeri 50 olarak belirlenmiştir. Kelimeler arası hesaplanan Öklid uzaklığı 50 den küçük ise kelimelerin kopya olduğuna karar verilerek işaretlenir.

DTW yöntemi ise Öklid uzaklığı gibi kelimeler arası yakınlık derecesine bakılır. DTW zaman alanında frekansın nasıl değiştiğini hesaplar ve iki sinyali birbirine hizalama için kullanılır. Öz vektörleri hesaplanan kelimelerin sırayla birbiri ile olan benzerlikleri DTW ile hesaplanır. Hesaplanan DTW benzerlik değerleri ne kadar küçükse iki kelimenin arasında farklılığın az olduğu yani benzerliğin fazla olduğu anlamına gelir. PLP özellik çıkarım yöntemi ile yapılan çalışmalarda DTW için empirik olarak eşik değeri 100 alınmıştır. Kelimeler arası hesaplanan DTW benzerlik değeri sıra ile kontrol edilir. DTW değeri 100'den küçük ise kelimeler arasında benzerlik ilişkisi olduğuna karar verilir. Benzer olan bu iki kelime kopya olarak işaretlenir.

## 2. YAPILAN ÇALIŞMALAR, BULGULAR VE İRDELEME

Tez kapsamında yapılan araştırmalarda kopyala-yapıştır ses sahteciliğinin tespiti için öncelikle ses sinyalinde bulunan kelimeler başarılı bir şekilde tespit edilerek ayrılmıştır. Sinyal okunduktan sonra pitch yöntemi ile sinyale ait pitch dizisi elde edilir. Ses aktivitesi tespit edilmek için dizi baştan itibaren taranarak 0'dan büyük dizi elemanları işaretlenir. Ses aktivitesi tespit edilen segmentler 1 olarak işaretlenirler. Pitch dizisi baştan taranarak 1 ile işaretlenmiş ardışık segmentler yeni kelimeye ait segment dizisine dahil edilirler. Pitch dizisi taramasında 0 ile karşılaşıldığında kelimenin sonu olarak işaretlenerek yeni kelime dizisine geçilir. Tespit edilen her bir segment dizisi bir kelimeyi temsil eder. Segmentlerin sayısı tespit edilen kelime sayısını vermektedir.

Kelimeler tespit edildikten sonra bu kelimelerin farklı özellik çıkarım yöntemleri ile öz vektörler elde edilmiştir. Tespit edilen kelimelerin özellik çıkarımı için ses tanımada yaygınca kullanılan plp, rasta-plp, mfcc ve spectral yöntemleri kullanılmıştır. Ses sinyalinde kopyalanarak başka bir konuma yapıştırılan kelimelerin öz vektörleri birbirine benzerlik göstermektedir. Aynı kelime orijinal olarak farklı anlarda kaydedilmiş olsa bile kelimelerin temel frekansı farklı olacağından kopyala-yapıştır olan kelimelere göre benzerlik derecesi düşük çıkmaktadır.

Kopyala-yapıştır sahteciliğinin tespit edilmesinde elde edilen öz vektörlerin benzerlik hesaplaması yapılarak her bir kelimenin diğer kelimelerle olan benzerlik ilişkisi tespit edilmeye çalışılmıştır. Bu aşamada kelime öz vektörleri ile korelasyon katsayısı, Öklid uzaklığı ve DTW yöntemleri ile istatistiksel olarak benzerlik hesaplaması yapılmıştır.

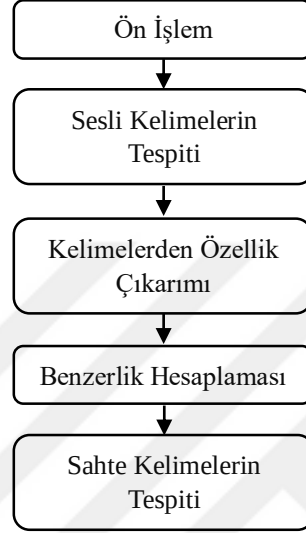
Hesaplanan benzerlik değerleri eşik değerine göre karşılaştırılarak kopyala-yapıştır sahteciliği tespit edilmeye çalışılmıştır. Kullanılan yöntem ve veri tabanına göre eşik değeri farklılık göstermektedir. Bu yüzden yapılan çalışmada veri setine ve kullanılan yöntem ve empirik olarak eşik değerleri kullanılmıştır.

Kopyala-yapıştır sahteciliğini pasif yöntemle tespit etmek için bu tez kapsamında uygulanan yöntemlerin genel akış diyagramı Şekil 14'te verilmiştir. İşlem basamakları alt başlıklarda detaylı şekilde anlatılmıştır.

Bu bölümde kopyala-yapıştır sahtecilik tespiti için tez kapsamında, çalışmalarda kullanılan plp, rasta-plp, mfcc ve spectral yöntemlerinin performans analizleri incelenmiştir. Yöntemlerin başarısını ölçmek için Timit ve Arabic Speech Corpus veri tabanları



kullanılmıştır. Çalışmada başarı ölçümü için kullanılan metrikler anlatılmıştır. Yöntemlerin başarısını değerlendirirken öz vektörler arası benzerlik tespiti için ayrı ayrı korelasyon katsayısı, Öklid uzaklığı ve DTW hesaplaması yapılmış ve deneysel sonuçları aşağıda verilmiştir. Aşağıda kullanılan yöntemlerin deney sonuçlarına göre avantaj ve dezavantajları ayrıntılı olarak ele alınmıştır.



Şekil 14. Uygulanan kopyala – yapıştır sahteciliği tespiti işlem basamakları

## 2.1. Veritabanı

Literatürde sahtecilik tespiti için kullanılan ortak bir veri tabanı bulunmamaktadır. Ses sinyalindeki kopyala-yapıştır sahteciliği tespiti için TIMIT veri tabanı ve Arabic Speech Corpus veri tabanı kullanılmıştır. Bu veri tabanındaki orijinal ses dosyalarından rastgele segment alınarak aynı dosyadaki rastgele başka bir segmentin üzerine yapıştırılarak sahte ses dosyaları oluşturulmuştur. Sahte ses oluşturulurken rastgele alınan segmentlerin süreleri yaklaşık olarak 0.2sn ile 0.6sn arasındadır.

Veri tabanlarından elde edilen sahte ses dosyalarına medyan filtresi uygulanarak, 20 dB ve 30 dB gürültü ekleyerek, 32bit ve 64bit sıkıştırma yapılarak sahteciliğin tespiti zorlaştırılmıştır. Bu atak uygulanmış sahte ses dosyaları ile literatürdeki yöntemlerin ve önerilen yöntemin dayanıklılığı ölçülmüştür.

TIMIT veri tabanında 128 adet orijinal ses dosyasından 368 adet sahte ses elde edilmiştir. Bu sahte seslere ayrı ayrı ataklar uygulanarak medyan filtresi uygulanmış, 20 dB ve 30 dB gürültü eklenmiş, 32 bit ve 64 bit sıkıştırılmış ataklı sahte ses dosyaları elde

edilmiştir. Yöntemlerin performansları bu ataklı dosyalarla da ölçülmüştür. Dosya sayıları tablo 2’te görülmektedir.

Tablo 2. Timit veri tabanında orijinal ve veri tabanından oluşturulmuş sahte dosya sayısı

| Atak              | Ses Dosyası Sayısı | Toplam |
|-------------------|--------------------|--------|
| Orijinal          | 128                | 128    |
| Ataksız           | 368                |        |
| Medyan Filtre     | 368                |        |
| 20 dB Gürültü     | 368                | 2208   |
| 30 dB Gürültü     | 368                |        |
| 32 Bit Sıkıştırma | 368                |        |
| 64 Bit Sıkıştırma | 368                |        |

Arabic Speech Corpus veri tabanında ise 36 adet orijinal ses dosyasından 30 adet sahte ses elde edilmiştir. Bu sahte seslere ayrı ayrı ataklar uygulanarak medyan filtresi uygulanmış, 20 dB ve 30 dB gürültü eklenmiş, 64 bit sıkıştırılmış ataklı sahte ses dosyaları elde edilmiştir. Dosya sayıları tablo 3’te görülmektedir.

Tablo 3. Arabic Speech Corpus veri tabanında orijinal ve veri tabanından oluşturulmuş sahte dosya sayısı

| Sınıf             | Ses Dosyası Sayısı | Toplam |
|-------------------|--------------------|--------|
| Orijinal          | 36                 | 36     |
| Ataksız           | 30                 |        |
| Medyan Filtre     | 30                 |        |
| 20 dB Gürültü     | 30                 | 150    |
| 30 dB Gürültü     | 30                 |        |
| 64 Bit Sıkıştırma | 30                 |        |

Bu tez çalışmasında yöntemlerin performansını ölçmek için deneysel çalışmalar hem 2208 sahte ses içeren Timit veri tabanından oluşturulmuş veri setinde hem de 150 sahte ses içeren Arabic Speech Corpus veri tabanından oluşturulmuş sette test edilmiştir.

## 2.2. Kullanılan Ölçütler

Uygulanan yöntemlerde elde edilen sonuçların başarısını ölçmek için duyarlılık, doğruluk, kesinlik ve F1 skoru hesaplaması yapılmıştır. Ölçüm için kullanılan DP sahte ses dosyalarının sahte olarak bulunan sayıyı ifade ederken YN ise sahte ses dosyalarına hatalı

olarak orijinal tespit edilen sayıyı ifade eder. DN Orijinal dosyaların doğru olarak tespit edilmiş sayısını, YP ise orijinal dosyalara hatalı olarak sahte olarak belirlenen sayıyı ifade eder. Denklem (20)'deki hesaplama ile duyarlılık, denklem (21)'deki hesaplama ile kesinlik, denklem (22)'daki hesaplama ile doğruluk ve denklem (23) ile F-1 skoru hesaplanmıştır.

$$Duyarlılık = \frac{DP}{DP + YN} \quad (20)$$

$$Doğruluk = \frac{DP + DN}{DP + YP + DN + YN} \quad (21)$$

$$Kesinlik = \frac{DP}{DP + YP} \quad (22)$$

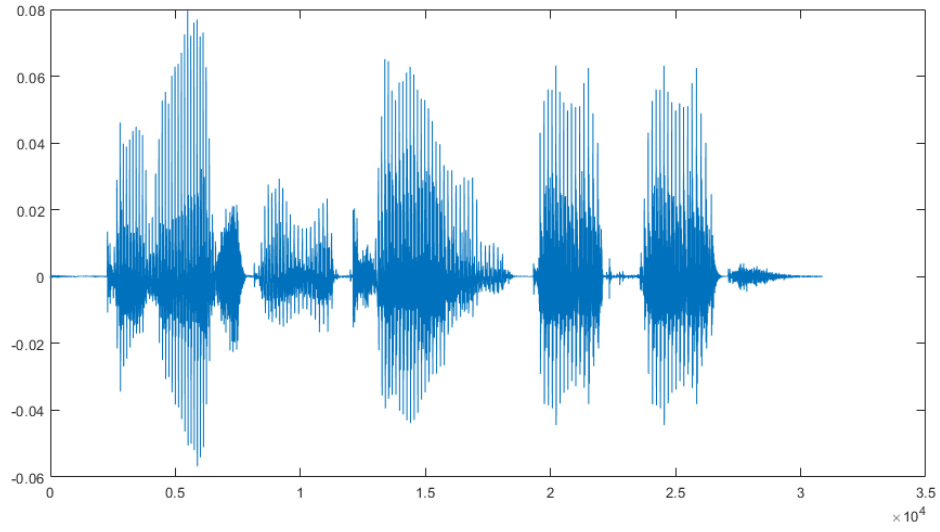
$$F1 - Skoru = \frac{DP}{DP + \frac{1}{2}YP + YN} \quad (23)$$

### 2.3. Sesli Kelimelerin Tespiti

Ses dosyasındaki sahteciliğin tespit edilmesindeki önemli aşamalardan biri de sesli kelimelerin tespit edilmesidir. Sesli kelimelerin tespiti için ses aktivitesi tespiti yöntemi (VAD) uygulanmıştır. Bu tez çalışmasında güçlü bir yöntem olan Pitch yöntemi ile sesteki kelimeler tespit edilmeye çalışılmıştır.

Ses dosyasının sinyali okunarak dijital frekans dizisi elde edilmiştir. Örnek ses dosyasının frekans görüntüsünü Şekil 15'te görülmektedir.

Sinyal okunduktan sonra yukarıda anlatılan pitch algoritması uygulanarak sinyale ait pitch değerleri hesaplanır. Bu pitch değerleri bir dizi şeklinde elde edilir. Örnek ses dosyasından 191 elemanlı pitch dizisi elde edilmiştir. Diğer bir deyişle hesaplanan pitch değerleri ile sinyalin tamamını temsil eden bir pitch vektör dizisi elde edilir. Örnek ses sinyaline ait pitch dizisi Tablo 4'te görülmektedir. Bu tablodaki 0 değerleri o anda konuşma olmadığı anlamına gelir yani sessiz segment olarak belirlenir. Dizideki 0'dan büyük değerler ise konuşma olan pitch değerini vermektedir. Hesaplanan değerlerin tabloda ondalık kısımları göz ardı edilmiştir.



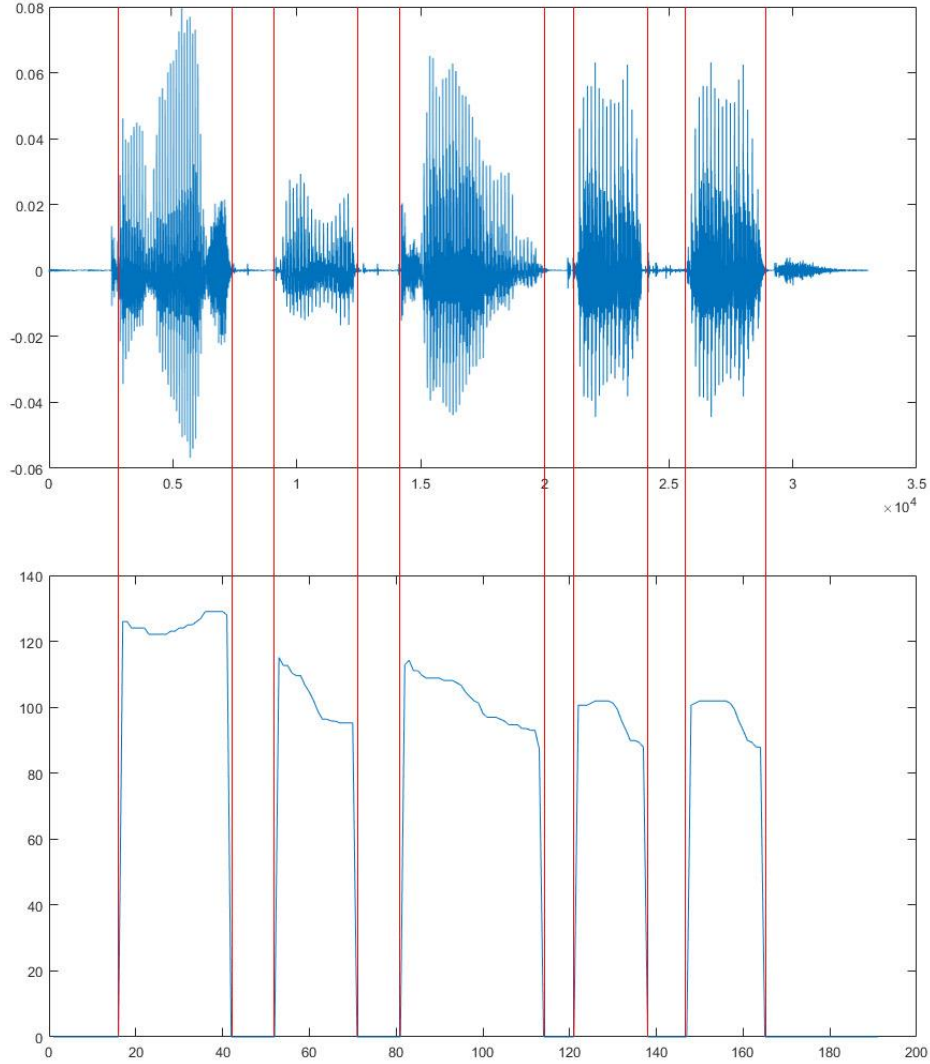
Şekil 15. Örnek ses sinyalinin frekans genlik değerine göre görüntüsü

Tablo 4. Örnek ses sinyalinden elde edilen segmentlerin hesaplanan pitch değerleri

|       |     |     |     |     |     |     |     |     |     |     |     |     |     |     |
|-------|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|
| Indis | 1   | 2   | 3   | 4   | 5   | 6   | 7   | 8   | 9   | 10  | 11  | 12  | 13  | 14  |
| Pitch | 0   | 0   | 0   | 0   | 0   | 0   | 0   | 0   | 0   | 0   | 0   | 0   | 0   | 0   |
| Indis | 15  | 16  | 17  | 18  | 19  | 20  | 21  | 22  | 23  | 24  | 25  | 26  | 27  | 28  |
| Pitch | 0   | 0   | 126 | 126 | 124 | 124 | 124 | 124 | 122 | 122 | 122 | 122 | 122 | 123 |
| Indis | 29  | 30  | 31  | 32  | 33  | 34  | 35  | 36  | 37  | 38  | 39  | 40  | 41  | 42  |
| Pitch | 123 | 124 | 124 | 125 | 125 | 126 | 127 | 129 | 129 | 129 | 129 | 129 | 128 | 0   |
| Indis | 43  | 44  | 45  | 46  | 47  | 48  | 49  | 50  | 51  | 52  | 53  | 54  | 55  | 56  |
| Pitch | 0   | 0   | 0   | 0   | 0   | 0   | 0   | 0   | 0   | 0   | 115 | 113 | 113 | 110 |
| Indis | 57  | 58  | 59  | 60  | 61  | 62  | 63  | 64  | 65  | 66  | 67  | 68  | 69  | 70  |
| Pitch | 110 | 110 | 107 | 105 | 102 | 99  | 96  | 96  | 96  | 96  | 95  | 95  | 95  | 95  |
| Indis | 71  | 72  | 73  | 74  | 75  | 76  | 77  | 78  | 79  | 80  | 81  | 82  | 83  | 84  |
| Pitch | 0   | 0   | 0   | 0   | 0   | 0   | 0   | 0   | 0   | 0   | 0   | 113 | 114 | 111 |
| Indis | 85  | 86  | 87  | 88  | 89  | 90  | 91  | 92  | 93  | 94  | 95  | 96  | 97  | 98  |
| Pitch | 111 | 110 | 109 | 109 | 109 | 109 | 108 | 108 | 108 | 107 | 107 | 105 | 103 | 102 |
| Indis | 99  | 100 | 101 | 102 | 103 | 104 | 105 | 106 | 107 | 108 | 109 | 110 | 111 | 112 |
| Pitch | 101 | 98  | 97  | 97  | 97  | 96  | 96  | 95  | 95  | 95  | 94  | 94  | 93  | 93  |
| Indis | 113 | 114 | 115 | 116 | 117 | 118 | 119 | 120 | 121 | 122 | 123 | 124 | 125 | 126 |
| Pitch | 87  | 0   | 0   | 0   | 0   | 0   | 0   | 0   | 0   | 101 | 101 | 101 | 101 | 102 |
| Indis | 127 | 128 | 129 | 130 | 131 | 132 | 133 | 134 | 135 | 136 | 137 | 138 | 139 | 140 |
| Pitch | 102 | 102 | 102 | 101 | 99  | 96  | 93  | 90  | 90  | 89  | 88  | 0   | 0   | 0   |
| Indis | 141 | 142 | 143 | 144 | 145 | 146 | 147 | 148 | 149 | 150 | 151 | 152 | 153 | 154 |
| Pitch | 0   | 0   | 0   | 0   | 0   | 0   | 0   | 101 | 101 | 102 | 102 | 102 | 102 | 102 |
| Indis | 155 | 156 | 157 | 158 | 159 | 160 | 161 | 162 | 163 | 164 | 165 | 166 | 167 | 168 |
| Pitch | 102 | 102 | 101 | 99  | 96  | 93  | 90  | 89  | 88  | 88  | 0   | 0   | 0   | 0   |
| Indis | 169 | 170 | 171 | 172 | 173 | 174 | 175 | 176 | 177 | 178 | 179 | 180 | 181 | 182 |
| Pitch | 0   | 0   | 0   | 0   | 0   | 0   | 0   | 0   | 0   | 0   | 0   | 0   | 0   | 0   |
| Indis | 183 | 184 | 185 | 186 | 187 | 188 | 189 | 190 | 191 |     |     |     |     |     |
| Pitch | 0   | 0   | 0   | 0   | 0   | 0   | 0   | 0   | 0   |     |     |     |     |     |

Örnek sinyale göre elde edilen pitch dizisi değerleri 1. İndisten 0 ile başlamaktadır. Tespit edilen ilk ses aktivitesi 17. dizi elemanındadır. 1-16. indisler sessiz yani konuşma olmayan segment olarak işaretlenir. Ses aktivitesi 17. İndisten başlayarak ilk sessiz indise kadar yani 42. indise kadar ses aktivitesi tespit edilmiştir. 17-41. indisler sinyalimizde tespit edilen ilk kelimeyi temsil eder. Pitch dizisindeki tüm indisler sıra ile kontrol edilerek işaretlenen indisler kelime dizisine aktarılarak kelimeler belirlenmiş olur. Örnek ses dosyasında belirlenmiş kelimelere ait pitch değerleri Şekil 16'da görülmektedir.

Şekil 16'da görüldüğü üzere sesli bölgelerde hesaplanan pitch değerleri kelimelere ait tepecikleri göstermektedir. Hesaplanan pitch değerlerine göre tespit edilen kelimeler sinyal ile örtüşmektedir. Örnek ses sinyalinde tespit edilen 5 kelime bulunmaktadır.



Şekil 16. Örnek ses sinyalinden pitch değerlerine göre tespit edilmiş kelimeler

## 2.4. Kelimelerden Özellik Çıkarımı

Sinyalden ses aktivitesi tespiti için pitch yöntemi ile tespit edilen kelimelerin sırayla öz vektörleri hesaplanır. Öz vektörleri elde etmek için literatürde LBP, Pitch Similarity gibi farklı yöntemler önerilmiştir. Bu tez çalışmasında ise plp, rasta-plp, mfcc ve spectral özellik çıkarım yöntemleri ile ayrı ayrı özellik çıkararak plp yönteminin başarısı araştırılmıştır. Yukarıda ifade edildiği üzere tespit edilen kelimelerin özellik çıkarım yöntemleri kullanılarak öz vektörleri elde edilir. Yöntemlerin başarısını daha sağlıklı değerlendirebilmek için hem Timit hem de Arabis Speech Corpus veri tabanında test edilmiştir. Kullanılan yöntem ve veri tabanına göre eşik değer yapılan deneysel çalışmalara göre empirik olarak belirlenmiştir.

### 2.4.1. PLP Yöntemi ile Kopyala – Yapıştır Sahteciliği Tespitinin Değerlendirilmesi ve Elde Edilen Deneysel Sonuçlar

PLP yöntemi Hermansky tarafından 1990 yılında ses sinyalindeki çevresel seslerin bastırılarak konuşmaları öne plana çıkaran insan işitmesini taklit eden bir yöntem olarak geliştirmiştir. Ses sinyalindeki kopyala-yapıştır sahteciliğinde uygulayarak performans analizi yapılmıştır. Sahtecilik tespitinde özellik çıkarım aşamasında PLP kullanılarak ses dosyasındaki kelimelere ait öz vektörler elde edilir.

Elde edilen kelimelere ait öz vektörleri korelasyon katsayısı hesaplanarak Timit veri tabanında test edildiğinde Tablo 5'deki performans değerleri elde edilmiştir. Timit veri tabanında test edilen orijinal ses dosyaları, sahte dosyalarının yanında, gürültüye karşı dayanıklılığı ölçmek için medyan filtresi uygulanmış, 20 dB ve 30 dB gürültü eklenmiş, 32 Bit ve 64 Bit sıkıştırılmış sahte dosyalarla da ayrı ayrı yöntemin başarısı test edilmiştir. Kelimeler arası benzerlik tespiti için korelasyon katsayısında eşik değeri olarak 0.95 değeri alınmıştır. Hesaplanan korelasyon katsayısı 0.95 ten küçük olan dosyalarda benzerlik ilişkisi zayıf olduğundan sahtecilik olmadığına karar verilmiştir. 0.95 ve daha büyük olan dosyada ise benzerlik ilişkisi kuvvetli olduğu için muhtemel sahte kelime adayları belirlenmiştir. Ayrıca sahte aday kelimelerin doğru olarak tespit edilip edilmediği de göz önünde bulundurulmuştur.

Ölçülen bu sonuçlara göre PLP yönteminin ataksız ses dosyalarında başarısının yüksek olduğu açıkça görülmektedir. Doğruluk ölçütüne göre performans analizini değerlendirdiğimizde ataksız dosyalarda %98'lik doğruluk sağlamışken en az başarıyı %84

ile 20 dB’lik gürültü eklenmiş dosyalarda sağlamıştır. Atak uygulanmış ses dosyalarında başarının düştüğü tüm metrik türlerinde görülmektedir. Medyan filtresi uygulanmış dosyalarda %2’lik performans düşüşü olurken 30 dB gürültü eklenmiş dosyalarda %1’lik performans kaybı söz konusudur. 32 Bit sıkıştırılmış dosyalarda 0,87 doğruluk oranı ile 20 dB gürültülü dosyalara göre daha iyi bir sonuç olsa da ataksız dosyalara göre %11’lik performans kaybı vardır. 64 Bit sıkıştırılmış dosyalarda %88’lik doğruluk oranı ile % 10 daha az performans göstermiştir.

Tablo 5. Timit veri tabanından elde edilen sahte veri tabanında PLP yöntemi özellik çıkarım ve korelasyon hesaplaması ile sahtecilik tespiti performans değerleri

|                    | Duyarlılık | Doğruluk | Kesinlik | F1 Skor |
|--------------------|------------|----------|----------|---------|
| Ataksız            | 0,99       | 0,98     | 0,99     | 0,99    |
| Medyan Filtrelleme | 0,96       | 0,96     | 0,99     | 0,98    |
| 20 dB Gürültü      | 0,80       | 0,84     | 0,98     | 0,88    |
| 30 dB Gürültü      | 0,97       | 0,97     | 0,99     | 0,98    |
| 32 Bit Sıkıştırma  | 0,84       | 0,87     | 0,98     | 0,90    |
| 64 Bit Sıkıştırma  | 0,85       | 0,88     | 0,98     | 0,91    |

F1 skoru ile değerlendirildiğinde atak uygulanmış ses dosyalarında medyan filtrede % 1’lik performans kaybı olurken 20 dB’lik gürültü eklenmiş ses dosyalarında ise %11’lik performans kaybı olmuştur. 30 dB de ise çok fazla etkilenmediğini ve %1’lik kayıp olduğu görülmektedir. 30 dB’lik gürültü eklenen dosyalar 20 dB gürültü eklenen dosyalara göre daha başarılı bir şekilde tespit edilmektedir. 32 Bit sıkıştırılmış dosyalarda ise %9 daha az başarılı iken 64Bit sıkıştırılmış dosyalar %8 daha az başarılı olduğu gözlenmiştir. Tüm ölçüm oranları değerlendirildiğinde yöntem en çok ataksız dosyalarda en az ise 20 dB gürültü eklenmiş dosyalarda başarı sağladığı tablo 5’de açıkça görülmektedir.

Timit veri tabanı ile yapılan çalışmaları pekiştirmek için farklı bir veri tabanı olan Arabic Speech Corpus ile de çalışılmıştır. Arabic Speech Corpus, Arapça konuşmaların olduğu ses dosyalarını içerir. Arabic Speech Corpus veri tabanı farklı bir fonetik dil gurubu olduğundan araştırmaya farklı bir bakış açısı katmıştır. Tez çalışmasında kullanılan özellik çıkarım yöntemleri ve benzerlik tespit için kullanılan hesaplamalar aynı şekilde bu veri tabanına da ayrı ayrı uygulanmıştır.

Arabic speech corpus veri tabanından elde edilen sahte veri tabanında PLP yönteminin korelasyon katsayısı hesaplanarak uygulanan sahtecilik tespitinden elde edilen sonuçlar

Tablo 6’da görülmektedir. Arabic Speech Corpus veri tabanında korelasyon katsayısı ile benzerlik hesaplamak için eşik değeri olarak 0,97 kullanılmıştır.

Tablo 6. Arabic speech corpus veri tabanından elde edilen sahte veri tabanında PLP yöntemi özellik çıkarım ve korelasyon katsayısı ile sahtecilik tespiti performans değerleri

|                    | Duyarlılık | Doğruluk | Kesinlik | F1 Skor |
|--------------------|------------|----------|----------|---------|
| Ataksız            | 0,97       | 0,91     | 0,85     | 0,91    |
| Medyan Filtrelleme | 0,97       | 0,91     | 0,85     | 0,91    |
| 20 dB Gürültü      | 0,97       | 0,91     | 0,85     | 0,91    |
| 30 dB Gürültü      | 0,97       | 0,91     | 0,85     | 0,91    |
| 64 Bit Sıkıştırma  | 0,93       | 0,89     | 0,85     | 0,89    |

Ataksız, medyan filtreli, 20 dB Gürültü, 30 dB Gürültü eklenmiş dosyalarda duyarlılık, doğruluk, kesinlik ve F1 skorları aynıdır (%97, %91, %85, %91). Bu sonuçlar, ses verilerinin işlenmesinde kullanılan farklı atak yöntemlerine karşı, PLP yönteminin korelasyon katsayısı ile kullanıldığında Arapça konuşmalara sahip dosyaların performansı üzerinde benzer sonuçlar üretmektedir. 64 Bit sıkıştırılmış dosyalarda diğerlerine kıyasla duyarlılık %93 doğruluk %89 ve F1 skor %89 değeri ile biraz düşük sonuç vermiştir, ancak kesinlik %85 değeri diğer durumlarla aynıdır. Ses dosyalarının 64 bit sıkıştırma ile işlenmesinin model performansını hafifçe düşürebileceğini, özellikle duyarlılık ve doğruluğun biraz azaldığını göstermektedir.

Bu sonuçları Timit veri tabanındaki Tablo 5 ile karşılaştırdığımızda ataksız dosyalarda Timit veri tabanı duyarlılık, doğruluk, kesinlik ve F1 skorlarının hepsi daha yüksektir. %98 doğruluk oranı ile Arabic Speech Corpus veri tabanındaki sonuçlara göre %7 daha başarılı bir sonuç elde etmiştir. %99 F1 skorunda ise %8’lik üstünlük sağlamıştır. Timit veri tabanında en düşük performans elde edilen 20 dB gürültü eklenmiş dosyalarda ise Arabic Speech Corpus veri tabanında elde edilen sonuçların %7 daha üstün olduğu görülür. Benzer şekilde %91 F1 skoru ile %3 daha üstündür. Ataklı dosyalarda en iyi sonuç veren 30 dB gürültü eklenmiş dosyalarda Timit veri tabanında %6 daha başarılı sonuç alınmıştır. 64 Bit sıkıştırılmış dosyalarda ise Arabic Speech Corpus veri tabanının başarısı %1 daha iyi sonuç vermiştir. Ses sinyalindeki dil farklılık gösterse de PLP yöntemi korelasyon katsayısı ile uygulanan model genel olarak başarılı sonuçlar üretmiştir.



#### 2.4.2. Rasta-PLP Yöntemi ile Kopyala – Yapıştır Sahteciliği Tespitinin Değerlendirilmesi ve Elde Edilen Deneysel Sonuçlar

Hermansky ve Morgan PLP yöntemini geliştirerek Rasta-PLP'yi uygulamıştır. Geliştirdikleri bu yöntem PLP ile hemen hemen aynı adımları içermektedir. Rasta-PLP spektrumu Rasta filtrelemeden önce Mel ölçeğinde filtreler. Rasta filtreleme, spektrumdaki ani değişiklikleri yumuşatmaya ve konuşma kanalındaki statik renklenmeyi gidermeye yardımcı olur. Cepstrum katsayılarını hesaplamadan önce spektrumdaki değişim oranlarını hesaplar. Bu, konuşmacıların sesindeki farklılıkların etkisini azaltmaya yardımcı olur.

Sahtecilik tespitinde özellik çıkarım aşamasında Rasta-PLP kullanılarak ses dosyasındaki kelimelere ait öz vektörler elde edilir. Elde edilen kelimelere ait öz vektörleri korelasyon katsayısı hesaplanarak Timit veri tabanında test edildiğinde Tablo 7'deki performans değerleri elde edilmiştir. Kelimeler arası benzerlik tespiti için korelasyon katsayısında eşik olarak 0.95 değeri alınmıştır.

Tablo 7. Timit veri tabanından elde edilen sahte veri tabanında Rasta-PLP yöntemi özellik çıkarım ve korelasyon hesaplaması ile sahtecilik tespiti performans değerleri

|                    | Duyarlılık | Doğruluk | Kesinlik | F1 Skor |
|--------------------|------------|----------|----------|---------|
| Ataksız            | 0,92       | 0,90     | 0,94     | 0,93    |
| Medyan Filtrelleme | 0,83       | 0,83     | 0,94     | 0,88    |
| 20 dB Gürültü      | 0,69       | 0,73     | 0,92     | 0,79    |
| 30 dB Gürültü      | 0,87       | 0,86     | 0,94     | 0,90    |
| 32 Bit Sıkıştırma  | 0,75       | 0,77     | 0,93     | 0,83    |
| 64 Bit Sıkıştırma  | 0,79       | 0,80     | 0,93     | 0,86    |

Sonuçlar incelendiğinde en başarılı performansı ataksız dosyalarda %90 doğruluk oranı ile göstermiştir. 0,73 doğruluk oranı ile yine en az başarıyı 20 dB gürültü eklenmiş dosyalarda sağlamıştır. Atak uygulanmış ses dosyalarında en iyi başarıyı 0,86 doğruluk oranı ile 30 dB gürültü eklenmiş dosyalarda göstermiştir. 62 Bit sıkıştırılmış dosyalar %80 doğruluk oranı ile 32 Bit sıkıştırılmış dosyalara göre %3 daha iyi sonuç vermiştir. Medyan filtre uygulanmış dosyalarda ise 30 dB gürültü eklenmiş dosyalara göre %3 performans kaybı söz konusudur. F1 skoru değerlendirildiğinde ise doğruluk oranlarına göre daha iyi gibi görünse de performans başarı sıralaması doğruluk oranı ile aynıdır.

Rasta-PLP yöntemi ile korelasyon katsayısı benzerlik hesabı ile elde edilen sonuçları Tablo 5'deki PLP yöntemi sonuçları ile karşılaştırdığımızda doğruluk oranına göre ataksız dosyalarda %8 daha başarısız sonuç elde edilmiştir. Atak uygulanmış dosyalarda ise medyan

filtre uygulanmış dosyalarda %13, 20 dB gürültü uygulanmış dosyalarda %11, 30 dB gürültü uygulanmış dosyalarda %11, 32 Bit sıkıştırılmış dosyalarda %10, 64 bit sıkıştırılmış dosyalarda ise %8 performans kaybı gözlenmiştir. Performans kaybı tüm dosya türlerinde %8 ile %13 arasında yakın bir kayıp söz konusudur. Tablo 7 ve Tablo 5'deki başarı performansına bakıldığında sıralamanın aynı olduğu açıkça görülmektedir.

Arabic Speech Corpus veri tabanında Rasta-PLP yönteminin uygulanarak korelasyon katsayısı hesabı ile performans değerlendirmesi yapılmıştır. Uygulanan yöntemler birbiri ile kıyaslanarak değerlendirilmiştir. Korelasyon katsayısı ile uygulanmış metot sonuçları Tablo 8'de görülmektedir. Arabic Speech Corpus veri tabanında benzerlik tespiti için eşik değeri olarak 0,90 kullanılmıştır.

Tablo 8. Arabic speech corpus veri tabanından elde edilen sahte veri tabanında Rasta-PLP yöntemi özellik çıkarım ve korelasyon katsayısı ile sahtecilik tespiti performans değerleri

|                    | Duyarlılık | Doğruluk | Kesinlik | F1 Skor |
|--------------------|------------|----------|----------|---------|
| Ataksız            | 0,83       | 0,89     | 0,93     | 0,88    |
| Medyan Filtrelleme | 0,83       | 0,89     | 0,93     | 0,88    |
| 20 dB Gürültü      | 0,77       | 0,86     | 0,92     | 0,84    |
| 30 dB Gürültü      | 0,83       | 0,89     | 0,93     | 0,88    |
| 64 Bit Sıkıştırma  | 0,87       | 0,91     | 0,93     | 0,90    |

Bu veriler, 64-bit sıkıştırılmış dosyaların genel performans açısından diğer test dosyalarına göre üstün olduğunu ve en yüksek başarıyı sağladığını göstermektedir. 20 dB gürültü eklenmiş dosyalar ise, %86 doğruluk oranı ile performans açısından en düşük değeri sergilemiştir. Ataksız, medyan filtre uygulanmış ve 30 dB gürültü eklenmiş dosyalarda %89 doğruluk oranı ile aynı başarıyı elde etmiştir. F1 skoru ile değerlendirdiğimizde de en iyi performans %90 ile 64 bit sıkıştırılmış dosyalarda elde edilmiştir. 20 dB gürültü eklenmiş dosyalarda yine %84 ile %6 64 Bit sıkıştırılmış dosyalara göre daha düşük başarı sağlanmıştır.

Arabic Speech Corpus veri tabanından elde edilen bu sonuçları PLP yöntemi ile kıyasladığımızda 64 Bit sıkıştırılmış dosyalarda Rasta-PLP yönteminin %91 doğruluk oranı ile PLP yöntemine göre %2 daha iyi başarı elde etmiştir. Ataksız dosyalar, medyan filtre uygulanmış ve 30 dB gürültü uygulanmış dosyalarda PLP nin %2, 20 dB gürültü eklenmiş dosyalarda ise %4 başarılı olduğu görülmektedir. F1 skoruna göre değerlendirdiğimizde de %90 ile 64 Bit sıkıştırılmış dosyalarda başarı gösterirken ataksız diğer ataklı dosyalarda PLP yönteminin başarılı olduğu görülmektedir.

Timit veri tabanında uygulanmış Rasta-PLP ve korelasyon katsayısı ile sahtecilik tespiti verileri ile kıyasladığımızda, yöntemin Arabic Speech Corpus veri tabanında medyan filtre uygulanmış dosyalarda doğruluk oranına göre %6, 20 dB gürültü eklenmiş dosyalarda %13, 64 Bit sıkıştırılmış dosyalarda %13 daha üstün başarı göstermiştir. Buna rağmen ataksız dosyalarda %1 ve 30 dB gürültü eklenmiş dosyalarda %3 daha düşük performans elde etmiştir.

#### 2.4.3. Spectral Yöntemi ile Kopyala – Yapıştır Sahteciliği Tespitinin Değerlendirilmesi ve Elde Edilen Deneysel Sonuçlar

Spectral Centroid bir ses sinyalinin spektrumunun ortalama frekansını temsil eden bir ölçüdür. Başka bir deyişle, ses sinyalinin frekans bileşenlerinin ne kadar "ortada" yoğunlaştığını gösterir. Elde edilen kelimelere ait öz vektörleri korelasyon katsayısı hesaplanarak Timit veri tabanında test edildiğinde Tablo 9'daki performans değerleri elde edilmiştir. Korelasyon katsayısı için eşik değeri olarak 0.95 belirlenmiştir.

SCM özellik çıkarım yöntemi ve korelasyon katsayısı ile hesaplanan benzerlik ile elde edilen analiz sonuçlarına göre ataksız dosyalarda %86 doğruluk oranı ile en iyi performansı elde etmiştir. Önceki yöntemlerde olduğu gibi ataklı dosyalarda ise 30 dB gürültü eklenmiş dosyalarda %83 doğruluk oranı sağlamıştır. Aynı şekilde en düşük performansı yine %67 doğruluk oranı ile 20 dB gürültü eklenmiş dosyalarda görülmektedir. Medyan filtre eklenmiş dosyalarda, 20 dB gürültü eklenmiş dosyalara göre %12, 32 Bit sıkıştırılmış dosyalarda %4, 64 Bit sıkıştırılmış dosyalarda ise %7 daha iyi performans göstermiştir. F1 skorunu değerlendirdiğimizde yine %90 ile ataksız dosyalarda en iyi başarıyı gösterirken %74 ile en düşük başarıyı 20 dB gürültü eklenmiş dosyalarda göstermektedir.

Tablo 9. Timit veri tabanında Spectral yöntemi özellik çıkarım ve korelasyon hesaplaması ile sahtecilik tespiti performans değerleri

|                    | Duyarlılık | Doğruluk | Kesinlik | F1 Skor |
|--------------------|------------|----------|----------|---------|
| Ataksız            | 0,89       | 0,86     | 0,92     | 0,90    |
| Medyan Filtrelleme | 0,79       | 0,79     | 0,92     | 0,85    |
| 20 dB Gürültü      | 0,63       | 0,67     | 0,90     | 0,74    |
| 30 dB Gürültü      | 0,84       | 0,83     | 0,92     | 0,88    |
| 32 Bit Sıkıştırma  | 0,68       | 0,71     | 0,90     | 0,77    |
| 64 Bit Sıkıştırma  | 0,73       | 0,74     | 0,91     | 0,81    |

Ataklı dosyalarda 20 dB gürültü eklenmiş dosyalara göre medyan filtre eklenmiş dosyalarda % 11, 30 dB gürültü eklenmiş dosyalarda %14, 32Bit sıkıştırılmış dosyalarda %3, 64 Bit sıkıştırılmış dosyalarda %7 daha iyi sonuç gözlemlenmiştir.

Spectral özellik çıkarım yöntemi sonuçlarını Tablo 5'deki PLP sonuçları ile kıyasladığımızda %86 doğruluk oranı ile en iyi performansı gösterdiği ataksız dosyalarda PLP yönteminde ise %98 doğruluk oranı sağlanmıştır. İki yöntem kıyaslandığında PLP yönteminin %12'lik daha iyi performans başarısı vardır. Spectral yöntemi ile en düşük %67 doğruluk oranı ile 20 dB gürültü eklenmiş dosyalarda performans gösterirken PLP yönteminde de %84 doğruluk oranı ile en düşük doğruluk oranını 20 dB gürültü eklenmiş dosyalarda gözlemlenmiştir. Paralel performans sonuçları elde edilse de PLP yönteminin 20 dB gürültü eklenmiş dosyalarda %17 daha iyi sonuç elde ettiği açıkça görülmektedir. Doğruluk oranına göre PLP yöntemi medyan filtre eklenmiş dosyalarda %17, 30 dB gürültü eklenmiş dosyalarda %14, 32 Bit sıkıştırılmış dosyalarda %16, 64 Bit sıkıştırılmış dosyalarda ise %14 daha üstün başarı sağlamıştır.

Spectral özellik çıkarım yöntemi ile Arabic Speech Corpus veri tabanında yapılan testler bu bölümde ele alınmıştır. Sepecral yöntemini korelasyon katsayısı benzerlik tespiti ile uyguladığımızda elde edilen test sonuçları Tablo 10'da verilmiştir. Bu testte korelasyon katsayısı için eşik değeri olarak 0,95 kullanılmıştır.

Tablo 10. Arabic speech corpus veri tabanından elde edilen sahte veri tabanında Spectral yöntemi özellik çıkarım ve korelasyon katsayısı ile sahtecilik tespiti performans değerleri

|                    | Duyarlılık | Doğruluk | Kesinlik | F1 Skor |
|--------------------|------------|----------|----------|---------|
| Ataksız            | 0,87       | 0,89     | 0,90     | 0,88    |
| Medyan Filtrelleme | 0,90       | 0,91     | 0,90     | 0,90    |
| 20 dB Gürültü      | 0,97       | 0,94     | 0,91     | 0,94    |
| 30 dB Gürültü      | 0,87       | 0,89     | 0,90     | 0,88    |
| 64 Bit Sıkıştırma  | 0,97       | 0,94     | 0,91     | 0,94    |

Spectral özellik çıkarım yöntemi ve korelasyon katsayısı ile benzerlik hesabı ile elde edilen analiz sonuçlarına göre %94 doğruluk oranı ile en iyi performansı 20 dB gürültü eklenmiş dosyalarda elde edilmiştir. Genellikle en kötü sonuçları 20 dB gürültü eklenmiş dosyalarda alırken bu yeni durum dikkat çekicidir. Benzer şekilde 64 Bit sıkıştırılmış dosyalarda da %94 doğruluk oranı ile başarı sağlamıştır. En düşük değerleri %89 doğruluk oranı ile ataksız dosyalarda ve 30 dB gürültü eklenmiş dosyalarda göstermiştir. Medyan filtre

uygulanmış dosyalarda ise %2 daha iyi sonuç elde edilmiştir. F1 skoruna bakıldığında benzer durum burada da görülmektedir. Duyarlılık, kesinlik ve F1 skoruna bakıldığında benzer şekilde en iyi sonuç 20 dB gürültü eklenmiş ve 64 Bit sıkıştırılmış dosyalarda elde edilmiştir.

Arabic Speech Corpus veri tabanında spectral özellik çıkarım yöntemi ve korelasyon katsayısı ile benzerlik hesabı yapan yöntem ile elde ettiğimiz sonuçları PLP özellik çıkarım yöntemi ile kıyasladığımızda ataksız ve 30 dB gürültü eklenmiş dosyalarda %2 lik performans kaybı görülmektedir. 20 dB gürültü eklenmiş dosyalarda %3, 64 Bit sıkıştırılmış dosyalarda %5 daha iyi doğruluk oranı ile üstün olduğu gözlemlenmiştir.

#### 2.4.4. MFCC Yöntemi ile Kopyala – Yapıştır Sahteciliği Tespitinin Değerlendirilmesi ve Elde Edilen Deneysel Sonuçlar

MFCC yöntemi, insan işitme sisteminin frekanslara karşı duyarlılığını taklit ederek çalışır. İnsan kulağı, düşük frekanslardaki değişikliklere yüksek frekanslardaki değişikliklere göre daha az duyarlıdır. MFCC yöntemi de bu özelliği göz önünde bulundurarak, ses sinyalinin Mel ölçeği adı verilen bir frekans ölçeğine dönüştürür. Bu özellik çıkarım yöntemi ile elde edilen öz vektörler korelasyon katsayısı benzerlik hesaplaması ile test edilmiştir. Elde edilen kelimelere ait öz vektörleri korelasyon katsayısı hesaplanarak Timit veri tabanında test edildiğinde Tablo 11'deki performans değerleri elde edilmiştir. Korelasyon katsayısı ile sahtecilik tespitinde eşik değeri 0.95 alınmıştır.

Tablo 11. Timit veri tabanında MFCC yöntemi özellik çıkarım ve korelasyon katsayısı ile sahtecilik tespiti performans değerleri

|                    | Duyarlılık | Doğruluk | Kesinlik | F1 Skor |
|--------------------|------------|----------|----------|---------|
| Ataksız            | 0,98       | 0,97     | 0,98     | 0,98    |
| Medyan Filtrelleme | 0,93       | 0,93     | 0,97     | 0,95    |
| 20 dB Gürültü      | 0,80       | 0,83     | 0,97     | 0,88    |
| 30 dB Gürültü      | 0,96       | 0,96     | 0,98     | 0,97    |
| 32 Bit Sıkıştırma  | 0,84       | 0,86     | 0,97     | 0,90    |
| 64 Bit Sıkıştırma  | 0,85       | 0,87     | 0,97     | 0,90    |

Tablo 11'deki test sonuçları doğruluk oranına göre değerlendirildiğinde %97 doğruluk oranı ile en yüksek ataksız dosyalarda elde edilmiştir. En düşük başarıyı %83 ile 20 dB gürültü eklenmiş dosyalarda görülmektedir. %14 daha düşük bir başarı olsa da gürültü

eklenmiş dosyaların önceki değerlendirmelerine göre oldukça yüksek bir başarı elde etmiştir. Ataklı dosyalarda %1 daha düşük başarı ile 30 dB gürültü eklenmiş dosyalarda çok yakın değer elde edilmiştir. Medyan filtresi uygulanmış dosyalarda %4 lük bir kayıp, 32 Bit sıkıştırılmış dosyalarda %11, 64 Bit sıkıştırılmış dosyalarda ise %10 başarı düşmüştür. Korelasyon katsayısı ile elde edilen sonuçları PLP yöntemi ile karşılaştırdığımızda MFCC yönteminde %97 doğruluk oranı ile ataksız dosyalarda elde edilen en iyi performans PLP yöntemi %98 doğruluk oranı ile %1 daha iyidir. En düşük performans gösterilen 20 dB gürültü eklenmiş dosyalarda PLP yönteminde %84 doğruluk oranı elde edilmişken MFCC yöntemi %1 daha düşük performans sağlamıştır. Tablo 5'teki sonuçlar dikkate alındığında PLP yönteminin 30 dB gürültü eklenmiş dosyalarda, 32 Bit sıkıştırılmış dosyalarda ve 64 Bit sıkıştırılmış dosyalarda % 1 daha başarılı olduğu görülmektedir. Medyan filtresi eklenmiş dosyalarda ise %3 daha iyi performans elde edilmiştir.

F1 skoru değerlendirilerek incelediğimizde de 20 dB gürültü eklenmiş dosyalarda %88 ve 32 Bit sıkıştırılmış dosyalarda %90 f1 skoru ile her iki yöntemde de aynı başarı sağlanmıştır. Buna rağmen PLP yönteminin medyan filtre eklenmiş dosyalarda %3, 30 dB gürültü eklenmiş dosyalarda ve 64 Bit sıkıştırılmış dosyalarda %1 daha üstün olduğu görülmektedir.

Mel ölçeğini kullanan MFCC yönteminin performans ölçümlerini Arabic Speech Corpus veri tabanında uygulayarak hem yöntemler arası hem de Timit veri tabanı ile karşılaştırılmıştır. MFCC yöntemini korelasyon katsayısı benzerlik hesabı ile uygulanan yöntem performans ölçümleri Tablo 12'de görülmektedir.

Tablo 12. Arabic speech corpus veri tabanından elde edilen sahte veri tabanında MFCC yöntemi özellik çıkarım ve korelasyon katsayısı ile sahtecilik tespiti performans değerleri

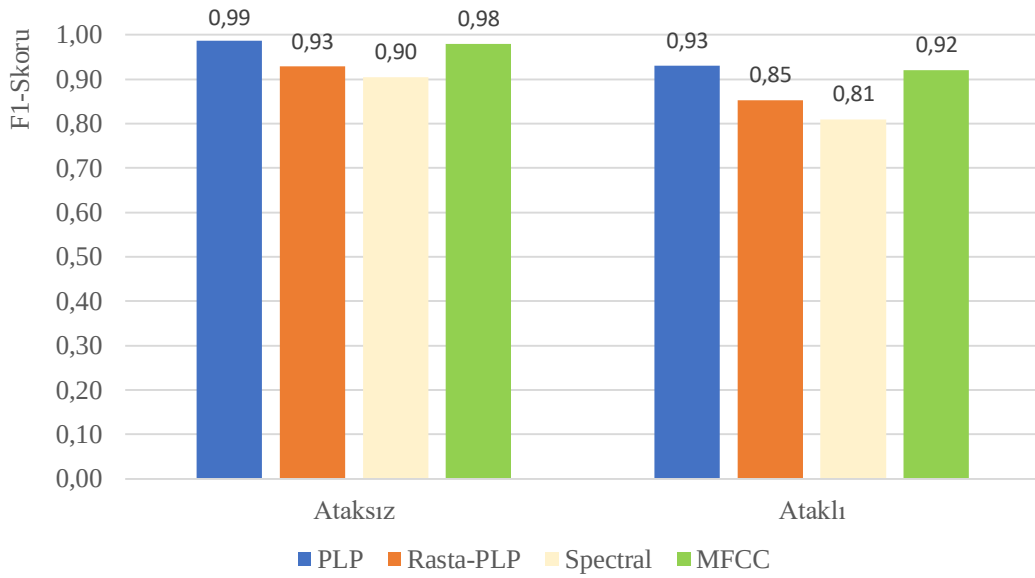
|                    | Duyarlılık | Doğruluk | Kesinlik | F1 Skor |
|--------------------|------------|----------|----------|---------|
| Ataksız            | 0,93       | 0,89     | 0,85     | 0,89    |
| Medyan Filtrelleme | 0,97       | 0,91     | 0,85     | 0,91    |
| 20 dB Gürültü      | 0,97       | 0,91     | 0,85     | 0,91    |
| 30 dB Gürültü      | 0,97       | 0,91     | 0,85     | 0,91    |
| 64 Bit Sıkıştırma  | 0,90       | 0,88     | 0,84     | 0,87    |

Tablo 12'deki veriler değerlendirildiğinde MFCC özellik çıkarım yönteminde en iyi sonu. Medyan filtre uygulanmış, 20 dB ve 30 dB gürültü eklenmiş dosyalarda %91 doğruluk oranı ile eşit olarak hesaplanmıştır. Tüm metrik değerlerine bakıldığında hepsinde eşit ve diğer dosyalara göre üstün sonuç sağladıkları görülmektedir. En düşük sonucu %88 doğruluk

oranı ile 64 Bit sıkıştırılmış dosyalardan elde edilmiştir. Ataksız dosyalarda ise %89 doğruluk oranı ile 64 Bit sıkıştırılmış dosyalara göre %1 daha iyi sonuç sağlamıştır.

MFCC yönteminin korelasyon benzerlik hesabı verilerini PLP yöntemi ile kıyasladığımızda yakın sonuçlar elde edildiği görülmektedir. Medyan filtresi eklenmiş, 20 dB ve 30 dB gürültü eklenmiş dosyalarda tüm metriklerde aynı sonuç elde edilmiştir. Ataksız dosyalarda %2, 64 Bit sıkıştırılmış dosyalarda %1 performans kaybı gözlenmiştir. Aynı performans kıyaslamasını Timit veri tabanındaki ile değerlendirdiğimizde benzer sonucun elde edildiği görülmektedir. Timit veri tabanında medyan filtresi eklenmiş dosyalarda MFCC yöntemi %3 performans kaybı görülürken diğer dosya türlerinde %1 kayıp olduğu görülmektedir. Bu performans değerleri ile farklı veri tabanlarında da yöntemlerin başarısının örtüştüğünü gösterir.

Uygulanan yöntemlerin ataklı ses dosyalarındaki performansını karşılaştırmak için F1-skoru değerlerinin her yöntem için ayrı ayrı ortalaması hesaplanmıştır. Yöntemlerin ataksız dosyalar ve elde edilen ataklı dosyaların ortalama verilerine göre karşılaştırması Şekil X'de görülmektedir.



Şekil 17. Timit veri tabanında uygulanan yöntemlerin F1-skoruna göre ataklı ve ataksız dosyalarda performans karşılaştırılması

Şekildeki sonuçlara göre ataksız dosyalarda en iyi performansı %99 F1-skoru ile PLP yöntemi göstermiştir. %98 F1-skoru ile en yakın performansı MFCC yöntemi göstermiştir. En düşük performansı %90 F1- skoru ile Spectral yöntemi uygulanan model gösterirken,

Rasta-PLP yöntemi uygulanan model %3 daha iyi performans göstermiştir. Ataklı dosyalardan elde edilen ortalama F1-skor değerlerine göre %93 ile PLP yönteminin uygulanan modellerde en iyi yöntem olduğu görülmektedir. Ataksız dosyalara göre ataklı dosyalarda %6 düşük performans göstermiştir. Ataksız dosyalarda olduğu gibi en yakın performansı %1 farkla MFCC yöntemi ile sağlanmıştır. Aynı şekilde en düşük performansı %81 ile Spectral yönteminde elde edilirken, Rasta-PLP yöntemi %85 F1-skoru ile daha iyi performans göstermiştir. Timit veri tabanında ataklı ve ataksız dosyalardaki performans analizine göre PLP yönteminin çalışılan yöntemlere göre en başarılı yöntem olduğu gözlemlenmiştir.

## 2.5. Benzerlik Hesaplaması

Tüm özellik çıkarım yöntemlerinde korelasyon katsayısı ile benzerlik hesabı yaparak kelimeler arası benzerlik tespit edilmeye çalışılmıştır. Yöntemlerin benzerlik hesaplamasında korelasyon yerine Öklid uzaklığı ve DTW hesaplanarak performansı yeniden test edilmiştir.

### 2.5.1. PLP Yöntemi ile Kopyala – Yapıştır Sahteciliği Tespitinin Değerlendirilmesi ve Elde Edilen DeneySEL Sonuçlar

Bu tez çalışmasında önerilen PLP özellik çıkarım yöntemi ile benzerlik hesaplama yöntemi olarak Öklid uzaklığı kullanılmıştır. Tablo 13’de Öklid uzaklığı performans sonuçları görülmektedir.

Hesaplanan F1 skoruna göre Öklid uzaklığı aynı öz nitelik vektörlerine uygulanmış olmasına rağmen eşik değerinin daha ortak kümede buluşmasından dolayı doğru tespit sayısı azaldığı gözlemlenmiştir. Eşik değeri olarak 50 belirlenmiştir. 50 den daha küçük Öklid uzaklığı hesaplanan kelimeler arasında yakınlık ilişkisi olduğundan bu kelimelerin kopyalanmış aday kelime olduğu varsayılmıştır. Kelimeler arasında hesaplanan Öklid uzaklığının en küçük değeri 50 den büyük ise bu dosyada kopyalanmış bir kelime olmadığı varsayılmıştır.

Ataksız dosyalarda %86 doğruluk oranı ile en iyi performansı sağlamasına rağmen korelasyon katsayısı performansına bakılarak Tablo 5 ile kıyaslandığında %12 daha düşük bir başarı göstermiştir. En az başarıyı yine 20 dB gürültü eklenmiş dosyalarda gösterirken 32 Bit sıkıştırılmış dosyalarda da %17 performans kaybı ile %70 başarı gösterirken 64 Bit



sıkıştırılmış dosyalarda ise %16 performans kaybı gözlemlenmiştir. %82 doğruluk oranı ile ataklı dosyalarda en fazla başarıyı 30 dB gürültü eklenmiş dosyalarda sağlamıştır.

Tablo 13. Timit veri tabanından elde edilen sahte veri tabanında PLP yöntemi özellik çıkarım ve Öklid uzaklığı hesaplaması ile sahtecilik tespiti performans değerleri

|                    | Duyarlılık | Doğruluk | Kesinlik | F1 Skor |
|--------------------|------------|----------|----------|---------|
| Ataksız            | 0,85       | 0,86     | 0,96     | 0,90    |
| Medyan Filtrelleme | 0,74       | 0,78     | 0,95     | 0,83    |
| 20 dB Gürültü      | 0,64       | 0,70     | 0,95     | 0,76    |
| 30 dB Gürültü      | 0,79       | 0,82     | 0,96     | 0,87    |
| 32 Bit Sıkıştırma  | 0,63       | 0,70     | 0,95     | 0,75    |
| 64 Bit Sıkıştırma  | 0,65       | 0,72     | 0,95     | 0,77    |

Ataksız dosyalarda yine %90 F1 skoru ile en başarılı olduğu görülmesine rağmen aynı dosyalarda korelasyon katsayılarına göre %9 daha az başarı sağladığı görülmektedir. PLP yöntemi uygulanmış korelasyon katsayısı benzerlik tespiti ile kıyaslandığında medyan filtre eklenmiş dosyalara % 15 performans kaybı söz konusudur. % 75 F1 skoru ile 32 Bit sıkıştırılmış dosyalarda en az başarıyı sağlamıştır. Ataklı dosyalarda Öklid uzaklığı ile yapılan benzerlik hesabı, korelasyon katsayısı hesabına göre %11 ve daha fazla performans kaybı gözlenmiştir.

Aynı performans benzerlik hesaplamasında DTW hesaplanarak yeniden test edilmiştir. Tablo 14’de DTW performans sonuçları görülmektedir. Eşik değeri olarak 100 belirlenmiştir. Bu eşikten daha düşük DTW skoru alan kelimeler, birbirine çok yakın ve muhtemelen kopyalanmış kabul edilir. Eğer iki kelime arasındaki DTW mesafesi 100’den büyükse, bu kelimeler arasında benzerlik bulunmadığı ve dolayısıyla kopyalanma ihtimalinin olmadığı varsayılır.

DTW mesafesi hesaplanarak PLP yöntemi ile kopyala-yapıştır sahtecilik tespiti sonuçlarına göre en çok başarıyı ataksız dosyalarda %87 doğruluk oranı ile sağlamıştır. %68’lik başarı ile en az 20 dB filtre eklenmiş dosyalarda ve 32 Bit sıkıştırılmış dosyalarda sağlamıştır. 30 dB gürültü eklenmiş dosyalar yine ataksız dosyalara yakın başarı göstermesine rağmen %5’lik performans düşüşü görülmektedir.

F1 skoruna bakıldığında en iyi performansı %90 ile yine ataksız dosyalarda gösterirken, %73 ile 20 dB gürültü eklenmiş dosyalarda ve 32 Bit sıkıştırılmış dosyalarda en düşük performansı göstermiştir.

Tablo 14. Timit veri tabanından elde edilen sahte veri tabanında PLP yöntemi özellik çıkarım ve DTW hesaplaması ile sahtecilik tespiti performans değerleri

|                    | Duyarlılık | Doğruluk | Kesinlik | F1 Skor |
|--------------------|------------|----------|----------|---------|
| Ataksız            | 0,84       | 0,87     | 0,97     | 0,90    |
| Medyan Filtrelleme | 0,69       | 0,75     | 0,97     | 0,80    |
| 20 dB Gürültü      | 0,58       | 0,68     | 0,96     | 0,73    |
| 30 dB Gürültü      | 0,78       | 0,82     | 0,97     | 0,87    |
| 32 Bit Sıkıştırma  | 0,59       | 0,68     | 0,96     | 0,73    |
| 64 Bit Sıkıştırma  | 0,65       | 0,73     | 0,97     | 0,78    |

Tablo 14 ve Tablo 13 karşılaştırıldığında ataksız dosyalarda ve 64 Bit sıkıştırılmış dosyalarda DTW hesaplaması doğruluk oranına göre %1 daha iyi sonuç sağlarken, 30 dB gürültü eklenmiş dosyalarda aynı performansı göstermiştir. 20 dB gürültü eklenmiş dosyalarda ve 32 Bit sıkıştırılmış dosyalarda ise %2 performans kaybı vardır. Her iki benzerlik hesabında yakın değerler elde edildiği görülmektedir.

Tablo 4 ile karşılaştırıldığında ise doğruluk oranlarını kıyasladığımızda ataksız dosyalarda %11, Medyan filtre eklenmiş dosyalarda %21, 20 dB gürültü eklenmiş dosyalarda %16, 30 dB gürültü eklenmiş dosyalarda %15, 32 Bit sıkıştırılmış dosyalarda %19 64 Bit sıkıştırılmış dosyalarda ise %15 performans düşmüştür.

PLP özellik çıkarım yöntemini Öklid uzaklığı ile uygulanan model Arabic speech corpus veri tabanından elde edilen sahte veri tabanında Tablo 15'deki sonuçları üretmiştir. Benzerlik tespiti için eşik değeri olarak Arabic Speech Corpus veri tabanında 1000 olarak belirlenmiştir.

Tablo 15. Arabic speech corpus veri tabanından elde edilen sahte veri tabanında PLP yöntemi özellik çıkarım ve Öklid uzaklığı ile sahtecilik tespiti performans değerleri

|                    | Duyarlılık | Doğruluk | Kesinlik | F1 Skor |
|--------------------|------------|----------|----------|---------|
| Ataksız            | 0,73       | 0,82     | 0,85     | 0,79    |
| Medyan Filtrelleme | 0,70       | 0,80     | 0,84     | 0,76    |
| 20 dB Gürültü      | 0,70       | 0,80     | 0,84     | 0,76    |
| 30 dB Gürültü      | 0,70       | 0,80     | 0,84     | 0,76    |
| 64 Bit Sıkıştırma  | 0,70       | 0,80     | 0,84     | 0,76    |

Öklid uzaklığı ile uygulanan PLP modeli, ataksız dosyalarda %82 doğruluk oranı ile en yüksek performansı göstermiştir. Ancak, bu performans, Tablo 6'da yer alan korelasyon katsayıları ile karşılaştırıldığında %9 oranında daha düşük bir başarı düzeyini gösterir.

Gürültü eklenmiş dosyalar arasında %80 doğruluk oranı ile tümünde aynı başarı gözlemlenirken korelasyon ile benzerlik hesabına göre %11 daha kötü sonuç sağlamıştır. 64Bit sıkıştırılmış dosyalarda bu kayıp oranı %9 dur.

Ataksız dosyalarda modelin en yüksek başarısı %79 F1 skoru ile belirlenmiştir. Ancak, bu dosyalarda korelasyon katsayılarıyla yapılan performans değerlendirmeleri, %12 oranında daha düşük bir başarıyı gösterir. PLP yönteminin uygulandığı ve Öklid uzaklığı ile benzerlik tespiti yapılan ataklı dosyalarda %76 F1 skoru ile tümünde aynı başarıyı göstermiştir. Korelasyon katsayısı ile uygulanan PLP modeline göre %15 daha kötü sonuç vermiştir.

Tablo 13'deki Timit veri tabanı ile kıyasladığımızda, Arabic Speech Corpus veri tabanındaki %82 doğruluk oranı ile %4 daha düşük performans göstermiştir. Ataklı dosyalarda en iyi performans gösterdiği 30 dB gürültü eklenmiş dosyalarda ise %2 performans kaybı vardır. Buna rağmen ataklı dosyaların tümünde gösterdiği %80 doğruluk oranı ile medyan filtre eklenmiş dosyalarda %2, 20 dB gürültü eklenmiş dosyalarda %10 ve 64 Bit sıkıştırılmış dosyalarda da %8 daha iyi sonuç üretmiştir.

PLP yöntemini DTW benzerlik hesabı ile uyguladığımızda Tablo 16'daki sonuçlar elde edilmiştir. Benzerlik tespitinde DTW hesaplaması için eşik değeri olarak 100 belirlenmiştir.

Tablo 16. Arabic speech corpus veri tabanından elde edilen sahte veri tabanında PLP yöntemi özellik çıkarım ve Öklid uzaklığı ile sahtecilik tespiti performans değerleri

|                    | Duyarlılık | Doğruluk | Kesinlik | F1 Skor |
|--------------------|------------|----------|----------|---------|
| Ataksız            | 0,70       | 0,80     | 0,84     | 0,76    |
| Medyan Filtrelleme | 0,73       | 0,82     | 0,85     | 0,79    |
| 20 dB Gürültü      | 0,73       | 0,82     | 0,85     | 0,79    |
| 30 dB Gürültü      | 0,73       | 0,82     | 0,85     | 0,79    |
| 64 Bit Sıkıştırma  | 0,70       | 0,80     | 0,84     | 0,76    |

DTW mesafesi hesaplanarak PLP yöntemi ile kopyala-yapıştır sahtecilik tespiti sonuçlarına göre en çok başarıyı %82 doğruluk oranı ile medyan filtre uygulanmış, 20 dB ve 30 dB gürültü eklenmiş dosyalarda sağlamıştır. Metodun başarısı %80'lik doğruluk oranı ile ataksız dosyalarda ve 64 Bit sıkıştırılmış dosyalarda görülmüştür.

F1 skoruna bakıldığında en iyi performansı %79 ile yine ataksız dosyalarda gösterirken, %76 ile ataksız dosyalarda ve 64 Bit sıkıştırılmış dosyalarda en düşük performansı göstermiştir. F1 skorunun da doğruluk oranına paralel sonuç ürettiği görülmektedir.

Tablo 16 ve Tablo 15 karşılaştırıldığında ataksız dosyalarda ve 64 Bit sıkıştırılmış dosyalarda DTW hesaplaması doğruluk oranına göre %2 daha iyi sonuç sağlarken, medyan filtresi uygulanmış, 20 dB ve 30 dB gürültü eklenmiş dosyalarda ise DTW yönteminin %2 daha iyi olduğu görülmektedir.

Tablo 6 ile yapılan karşılaştırma sonucunda, doğruluk oranlarında ataksız dosyalarda %11 oranında, medyan filtre eklenmiş dosyalarda ise 20 dB ve 30 dB gürültü eklenmiş dosyalarda %9 oranında bir performans düşüşü gözlemlenmiştir. F1 skorları açısından değerlendirildiğinde, Arabic Speech Corpus veri tabanında korelasyon hesabı ile uygulanan PLP yönteminin ataksız dosyalarda %15, medyan filtre eklenmiş dosyalarda %12, 20 dB ve 30 dB gürültü eklenmiş dosyalarda %12, ve 64-bit sıkıştırılmış dosyalarda %13 oranında daha yüksek başarı sağladığı belirlenmiştir. Bu sonuçlar, korelasyon hesabı ile PLP yönteminin genel olarak daha başarılı bir metot olduğunu açıkça göstermektedir.

### 2.5.2. Rasta-PLP Yöntemi ile Kopyala – Yapıştır Sahteciliği Tespitinin Değerlendirilmesi ve Elde Edilen DeneySEL Sonuçlar

Rasta-PLP yöntemini Öklid uzaklığı hesaplayarak benzerlik tespiti uygulanarak Tablo 17'deki sonuçlar elde edilmiştir. Rasta-PLP yönteminde Öklid uzaklığı için eşik olarak 0.1 belirlenmiştir.

Tablo 17. Timit veri tabanından elde edilen sahte veri tabanında Rasta-PLP yöntemi özellik çıkarım ve Öklid uzaklığı hesaplaması ile sahtecilik tespiti performans değerleri

|                    | Duyarlılık | Doğruluk | Kesinlik | F1 Skor |
|--------------------|------------|----------|----------|---------|
| Ataksız            | 0,79       | 0,83     | 0,97     | 0,87    |
| Medyan Filtrelleme | 0,55       | 0,65     | 0,96     | 0,70    |
| 20 dB Gürültü      | 0,57       | 0,66     | 0,96     | 0,71    |
| 30 dB Gürültü      | 0,77       | 0,81     | 0,97     | 0,86    |
| 32 Bit Sıkıştırma  | 0,46       | 0,58     | 0,95     | 0,62    |
| 64 Bit Sıkıştırma  | 0,59       | 0,68     | 0,96     | 0,73    |

Öklid uzaklığı ile Rasta-PLP çalışması sonuçlarına baktığımızda en başarılı performansı %83 ile ataksız sahte dosyalarda göstermiştir. En az başarıyı ise 32 Bit sıkıştırılmış dosyalarda %58 ile gözlenmiştir. Ataksız dosyalar, 30 dB gürültülü dosyalara göre %2 daha iyi bir sonuç elde edilirken, %68 doğruluk oranı ile 64 bit sıkıştırılmış dosyalara göre %15 daha iyi sonuç elde edilmiştir. 20 dB gürültülü dosyalar doğruluk

oranına göre önceki analizlerde %57 ile en kötü sonucu verirken bu testte 32 Bit sıkıştırılmış dosyalara göre %8 daha iyi performans sağlamıştır.

F1 skoruna bakıldığında da performans sıralaması doğruluk oranı ile örtüşmektedir. Ataklı dosyalarda %87 F1 skoru ile en iyi başarıyı ataksız dosyalarda gösterirken Tablo 7'deki korelasyon katsayısı ile yapılan çalışmaya göre performans kaybı yaşanmamıştır. Ancak ataksız dosyalarda %6, medyan filtre uygulanmış dosyalarda %18, 20 dB gürültü eklenmiş dosyalarda %8, 30 dB gürültülü eklenmiş dosyalarda %4, 32 Bit sıkıştırılmış dosyalarda ise %21, 64 Bit sıkıştırılmış dosyalarda ise %13 performans düşmüştür. Burada %21 ile 32 Bit sıkıştırılmış dosyalardaki performans kaybı dikkati çekmektedir.

Rasta-PLP Öklid uzaklığı sonuçlarını Tablo 13'deki PLP sonuçları ile karşılaştırdığımızda ataksız dosyalarda doğruluk oranında %3 lük kayıp görülmektedir. Medyan filtre uygulanmış dosyalarda %13, 20 dB gürültü uygulanmış dosyalarda %4, 32 Bit sıkıştırılmış dosyalarda %12, 64 Bit sıkıştırılmış dosyalarda ise %4 performans kaybı vardır. Buna rağmen 30 dB gürültü eklenmiş dosyalarda %4 daha iyi sonuç elde edilmiştir. En fazla performans kaybı medyan filtre uygulanmış dosyalarda gerçekleşmiştir.

Rasta-PLP yönteminden elde edilen öz vektörlere benzerlik hesaplamasında DTW yöntemi ile yeniden test edilmiştir. Ses işlemede, farklı uzunluklarda olan iki sinyal arasındaki benzerliği ölçmek için kullanılan bir algoritmadır. Sesin hız değişimi ve gürültüye dayanıklı olması yöntemin avantajıdır. Tablo 18'de DTW performans sonuçları görülmektedir. Eşik değeri olarak 0.2 belirlenmiştir.

Tablo 18. Timit veri tabanından elde edilen sahte veri tabanında Rasta-PLP yöntemi özellik çıkarım ve DTW hesaplaması ile sahtecilik tespiti performans değerleri

|                    | Duyarlılık | Doğruluk | Kesinlik | F1 Skor |
|--------------------|------------|----------|----------|---------|
| Ataksız            | 0,74       | 0,79     | 0,96     | 0,84    |
| Medyan Filtrelleme | 0,39       | 0,53     | 0,93     | 0,55    |
| 20 dB Gürültü      | 0,43       | 0,55     | 0,93     | 0,59    |
| 30 dB Gürültü      | 0,63       | 0,70     | 0,95     | 0,76    |
| 32 Bit Sıkıştırma  | 0,37       | 0,51     | 0,93     | 0,53    |
| 64 Bit Sıkıştırma  | 0,50       | 0,61     | 0,94     | 0,66    |

DTW ile Rasta-PLP çalışması sonuçlarına baktığımızda en başarılı performansı %79 doğruluk oranı ile ataksız sahte dosyalarda göstermiştir. Buna rağmen Öklid uzaklığına göre %3 lük, korelasyon katsayısına göre %11'lik performans düşüklüğü görülmüştür. %51 doğruluk oranı ile en düşük performansı 32 Bit sıkıştırma ile göstermiştir. Tablo 7 ve Tablo

17’de en düşük performansı yine 32 Bit sıkıştırılmış dosyalarda gözlemlenmesine rağmen Öklid uzaklığına göre %7’lik, korelasyon katsayısına göre %26’lık performans düşüklüğü görülmüştür. Ataklı dosyalarda %70 doğruluk oranı ile en iyi başarıyı 30 dB gürültü eklenmiş dosyalarda gerçekleştirirken 20 dB gürültü eklenmiş dosyalarda %15 daha düşük başarı elde edilmiştir. 64 Bit sıkıştırılmış dosyalarda, 32 Bit sıkıştırılmış dosyalara göre %10 daha iyi sonuç sağlamasına rağmen Öklid uzaklığına göre %7 lik, korelasyon katsayısına göre %19 luk performans düşüşü gözlenmiştir.

F1 skoru değerlendirildiğinde %84 ile yine ataksız dosyaların başarısı önde görülmektedir. Ataklı dosyalarda ise 30 dB gürültü eklenmiş dosyaların %76’lık F1 skoru ile önde olduğu %53 ile en düşük sonucun 32 Bit sıkıştırılmış dosyalarda olduğu tespit edilmiştir. 64 Bit sıkıştırılmış dosyalarda 32 Bit sıkıştırılmış dosyalara göre %13 daha iyi performans gösterirken, 20 dB gürültü eklenmiş dosyalar 30 dB gürültü eklenmiş dosyalara göre %17 daha düşük F1 skoru hesaplanmıştır. Tablo 7 ve Tablo 17 ile karşılaştırdığımızda DTW hesaplaması ile ataksız dosyalarda elde edilen en iyi %84’lük F1 skoru Öklid uzaklığına göre %3, korelasyon katsayısına göre %9 daha kötü performans sergilemiştir. DTW hesaplamasında en düşük 32 Bit sıkıştırılmış dosyalarda %53 olarak gerçekleştirirken, Öklid uzaklığına göre %9, korelasyon katsayısına göre %30 performans kaybı görülmektedir.

Rasta-PLP’nin DTW hesaplamasını PLP ile kıyasladığımızda ise PLP yönteminin başarısı öne çıkmaktadır. Ataksız dosyalarda PLP %90 F1 skoru ile %6 daha başarılıdır. Meydan filtresi uygulanmış ataklı dosyalarda ise Rasta-PLP %25 daha düşük performans gösterirken 20 dB gürültü eklenmiş dosyalarda %14, 30 dB gürültü eklenmiş dosyalarda %11, 32 Bit sıkıştırılmış dosyalarda %20, 64 Bit sıkıştırılmış dosyalarda ise %11 başarısı düşük çıkmıştır.

Rasta-PLP özellik çıkarım yöntemi ile Öklid uzaklığı hesaplayarak uygulanan sahtecilik tespiti yöntemi Arabic Speech Corpus veri tabanındaki test sonuçları Tablo 19’da görülmektedir. Öklid uzaklığı ile kelimeler arası benzerlik tespiti için eşik değeri olarak 0,092 kullanılmıştır.

Rasta-PLP ve Öklid uzaklığı ile benzerlik hesabı en iyi doğruluk oranını %86 ile ataksız dosyalarda sağlarken en düşük doğruluk oranını %77 ile 20 dB gürültü eklenmiş ve 64 Bit sıkıştırılmış dosyalarda görülmüştür. 20 dB gürültüye göre 30 dB gürültü eklenmiş dosyalarda %5 daha iyi sonuç elde edilirken medyan filtre eklenmiş dosyalarda ise %8 daha iyi performans sağlanmıştır.

Tablo 19. Arabic speech corpus veri tabanından elde edilen sahte veri tabanında Rasta-PLP yöntemi özellik çıkarım ve Öklid uzaklığı ile sahtecilik tespiti performans değerleri

|                    | Duyarlılık | Doğruluk | Kesinlik | F1 Skor |
|--------------------|------------|----------|----------|---------|
| Ataksız            | 0,80       | 0,86     | 0,89     | 0,84    |
| Medyan Filtrelleme | 0,77       | 0,85     | 0,88     | 0,82    |
| 20 dB Gürültü      | 0,60       | 0,77     | 0,86     | 0,71    |
| 30 dB Gürültü      | 0,70       | 0,82     | 0,88     | 0,78    |
| 64 Bit Sıkıştırma  | 0,60       | 0,77     | 0,86     | 0,71    |

F1 skoru ile de değerlendirdiğimizde en iyi sonucu %84 ile yine ataksız dosyalarda sağlanmış ve en düşük sonuç %13 düşüşle 20 dB gürültü eklenmiş ve 64 Bit sıkıştırılmış dosyalarda görülmüştür. 30 dB gürültü eklenmiş dosyalarda ise ataklı dosyalarda en iyi sonucu gösteren medyan filtre eklenmiş dosyalara göre %4 daha düşük performans göstermiştir.

Arabic Speech Corpus veri tabanındaki Rasta-PLP Öklid uzaklığı sonuçlarını korelasyon katsayısı hesaplaması ile kıyasladığımızda ataksız dosyalardaki test sonuçları %86 doğruluk oranı ile korelasyon hesabına göre %3'lük düşüş görülür. Medyan filtre uygulanmış dosyalarda %4 düşüş gözlenirken, en düşük doğruluk oranı elde edilen 20 dB gürültü eklenmiş dosyalarda %9, 30 dB gürültü eklenmiş dosyalarda ise %7 performans kaybı görülmüştür. En fazla performans kaybını ise %14 düşüşle 64 Bit sıkıştırılmış dosyalarda gözlenmiştir. F1 skoruna göre değerlendirdiğimizde ataksız dosyalarda, Öklid uzaklığı yöntemi %84 başarı ile korelasyon yöntemine göre %4 daha düşük performans göstermiştir. Medyan filtresi uygulanan dosyalarda bu fark %6'ya yükselirken, 20 dB ve 30 dB gürültü eklenen dosyalarda sırasıyla %13 ve %10'luk bir düşüş gözlemlenmiştir. En büyük performans kaybı ise %19 ile 64 bit sıkıştırılmış dosyalarda yaşanmıştır.

Rasta-PLP özellik çıkarım yöntemini DTW benzerlik tespiti hesaplaması ile uygulandığında yöntemin kopyala- yapıştır sahteciliği tespitinde elde ettiği sonuçlar Tablo 20'de görülmektedir. DTW benzerlik hesabında eşik değeri olarak 0,23 kullanılmıştır.

Ataksız ve medyan filtreli dosyalar, diğer test koşullarına kıyasla daha yüksek duyarlılık, doğruluk, kesinlik ve F1 skorları sunmaktadır. 20 dB gürültü eklenmiş dosyalar, %73 doğruluk oranı ile en düşük performansı sergilerken, 30 dB gürültü eklenmiş dosyalar ve 64-bit sıkıştırılmış dosyalar benzer performans seviyelerine ulaşmıştır, ancak bunlar hâlâ ataksız ve medyan filtreli dosyalardan doğruluk oranına göre %9 daha düşük sonuç göstermektedir.

Tablo 20. Arabic speech corpus veri tabanından elde edilen sahte veri tabanında Rasta-PLP yöntemi özellik çıkarım ve DTW ile sahtecilik tespiti performans değerleri

|                    | Duyarlılık | Doğruluk | Kesinlik | F1 Skor |
|--------------------|------------|----------|----------|---------|
| Ataksız            | 0,80       | 0,83     | 0,83     | 0,81    |
| Medyan Filtrelleme | 0,80       | 0,83     | 0,83     | 0,81    |
| 20 dB Gürültü      | 0,57       | 0,73     | 0,77     | 0,65    |
| 30 dB Gürültü      | 0,60       | 0,74     | 0,78     | 0,68    |
| 64 Bit Sıkıştırma  | 0,60       | 0,74     | 0,78     | 0,68    |

F1 skoruna bakıldığında en iyi performans ataksız ve medyan filtreli dosyalarda gözlemlenmiştir. Bu dosyalar, modelin yüksek doğruluk ve duyarlılık ile dengeli bir şekilde performans sergilediğini gösterir. Gürültü ve sıkıştırma eklenen dosyalar, modelin F1 skorunda belirgin bir düşüşe neden olmuştur, özellikle 20 dB gürültü eklenmiş dosyalar %65 ile en düşük F1 skoruna sahip olup, performansın ciddi şekilde etkilendiğini ortaya koymaktadır.

Rasta- PLP yöntemi DTW hesabını, korelasyon katsayısı hesabı ile kıyasladığımızda en iyi sonucu %83 doğruluk oranı ile ataksız dosyalarda elde etmiş olsa bile korelasyon katsayısı ile benzerlik hesabına göre %6 performans düşmüştür. Aynı durum medyan filtresi eklenmiş dosyalarda da görülmektedir. En düşük performans elde edilen 20 dB gürültü eklenmiş dosyalarda ise %13 daha kötü sonuç elde edilmiştir. 30 dB gürültü eklenmiş dosyalarda %15, 64 Bit sıkıştırılmış dosyalarda ise %17 düşüş görülmektedir. DTW benzerlik hesabı ile uygulanan yöntemde ciddi derecede performans kaybı yaşanmıştır.

### 2.5.3. Sepctral Yöntemi ile Kopyala – Yapıştır Sahteciliği Tespitinin Değerlendirilmesi ve Elde Edilen Deneysel Sonuçlar

Bu özellik çıkarım yöntemi ile elde edilen öz vektörler korelasyon katsayısındaki performansını değerlendirmek için Öklid uzaklığı ve DTW benzerlik hesaplaması ile ayrı ayrı test edilerek kıyaslanmıştır. Spectral özellik çıkarım yöntemi ve Öklid uzaklığı ile benzerlik hesaplanarak elde edilen performans değerleri Tablo 21’de görülmektedir. Sahtecilik tespiti için Öklit mesafesi kullanılırken eşik değeri 110 olarak belirlenmiştir. Burada elde edilen ölçümlere göre benzer şekilde en iyi başarıyı %91 doğruluk oranı ile ataksız dosyalarda elde ederken, en düşük başarıyı %73 ile 20 dB gürültü eklenmiş dosyalarda elde edilmiştir. Ataklı dosyalardan %89 doğruluk oranı ile en iyi performansı 30



dB gürültü eklenmiş dosyalarda sağlamıştır. 64 Bit sıkıştırılmış dosyalardaki başarı %82 doğruluk oranı ile 32 Bit sıkıştırılmış dosyalara göre %6 daha iyidir.

Tablo 21. Timit veri tabanından elde edilen sahte veri tabanında Spectral yöntemi özellik çıkarım ve Öklid uzaklığı ile sahtecilik tespiti performans değerleri

|                    | Duyarlılık | Doğruluk | Kesinlik | F1 Skor |
|--------------------|------------|----------|----------|---------|
| Ataksız            | 0,89       | 0,91     | 0,98     | 0,93    |
| Medyan Filtrelleme | 0,74       | 0,79     | 0,97     | 0,84    |
| 20 dB Gürültü      | 0,66       | 0,73     | 0,97     | 0,79    |
| 30 dB Gürültü      | 0,87       | 0,89     | 0,98     | 0,92    |
| 32 Bit Sıkıştırma  | 0,70       | 0,76     | 0,97     | 0,81    |
| 64 Bit Sıkıştırma  | 0,78       | 0,82     | 0,97     | 0,87    |

F1 skoru değerlendirildiğinde ise benzer sonuçlar dikkat çekmektedir. %93 F1 skoru ile en iyi performans ataksız dosyalarda görülürken %79 F1 skoru ile en düşük performans yine 20 dB gürültü eklenmiş dosyalarda görülmüştür. Ataklı dosyalarda ise en iyi sonucu %92 F1 skoru ile 30 dB gürültü eklenmiş dosyalarda göstermiştir.

Timit veri tabanındaki Öklid uzaklığı ile benzerlik hesaplanmış spectral özellik çıkarımını yine Öklid uzaklığı uygulanmış PLP yöntemi ile kıyasladığımızda farklılık olarak spectral özellik çıkarım yönteminin daha başarılı olduğu görülmüştür. Ataksız dosyalarda %91 doğruluk oranı ile PLP ye göre %5 daha üstün olduğu görülmektedir. En düşük başarı gösterilen 20 dB gürültü eklenmiş dosyalarda ise %3'lük üstünlük sağlamıştır. Ataklı dosyalarda en iyi başarı gösterilen 30 dB gürültü eklenmiş dosyalarda da %7'lik üstünlük sağlamıştır. Medyan filtre uygulanmış dosyalarda %1, 32 Bit sıkıştırılmış dosyalarda %6, 64 Bit sıkıştırılmış dosyalarda %10 daha başarılı olduğu görülmüştür.

Spectral özellik çıkarım yöntemi ve DTW ile benzerlik hesaplanarak elde edilen performans değerleri Tablo 22'de görülmektedir. Eşik değeri olarak 188 belirlenmiştir. DTW yöntemini Öklid uzaklığına göre kıyasladığımızda %95 doğruluk oranı ile en iyi başarı gösterilen ataksız dosyalarda Öklid uzaklığı ile yapılan hesaplamada %91 olarak gerçekleşmiştir. DTW yöntemi %4 daha iyi performans göstermiştir. En düşük başarı gösterilen 20 dB gürültü eklenmiş dosyalarda %72 doğruluk oranı ile Öklid uzaklığına göre %1, medyan filtresi eklenmiş dosyalarda %2 daha az başarı sağlamıştır. Buna rağmen 30 dB gürültü eklenmiş dosyalarda %1, 32 Bit sıkıştırılmış dosyalarda %1 daha iyi, 64 Bit sıkıştırılmış dosyalarda ise aynı doğruluk oranını elde etmiştir.

F1 skoru değerlerine bakıldığında da ataksız dosyalarda %3, 30 dB gürültü eklenmiş dosyalarda %1, 32 Bit sıkıştırılmış dosyalarda %1 daha iyi performans sağlarken 64 Bit sıkıştırılmış dosyalarda %87 F1 skoru ile aynı performansı göstermiştir. Medyan filtresi uygulanmış dosyalarda ve 20 dB gürültü eklenmiş dosyalarda %1 daha düşük performans göstermiştir. F1 skorunun doğruluk oranı ile paralel sonuç ürettiği açıkça görülmektedir.

Tablo 22. Timit veri tabanından elde edilen sahte veri tabanında Spectral yöntemi özellik çıkarım ve DTW hesaplaması ile sahtecilik tespiti performans değerleri

|                    | Duyarlılık | Doğruluk | Kesinlik | F1 Skor |
|--------------------|------------|----------|----------|---------|
| Ataksız            | 0,97       | 0,95     | 0,96     | 0,96    |
| Medyan Filtrelleme | 0,73       | 0,77     | 0,94     | 0,83    |
| 20 dB Gürültü      | 0,67       | 0,72     | 0,94     | 0,78    |
| 30 dB Gürültü      | 0,90       | 0,90     | 0,95     | 0,93    |
| 32 Bit Sıkıştırma  | 0,73       | 0,77     | 0,94     | 0,82    |
| 64 Bit Sıkıştırma  | 0,81       | 0,82     | 0,95     | 0,87    |

Spectral özellik çıkarım yönteminde DTW benzerlik hesabı analizi ile PLP özellik çıkarım yöntemi Tablo 9 ile karşılaştırıldığında ataksız dosyalarda spectral özellik çıkarımının %95 doğruluk oranı ile PLP yöntemine göre %8 daha başarılı olduğu görülmektedir. %72 doğruluk oranı ile en düşük başarıyı gösterdiği 20 dB gürültü eklenmiş dosyalarda PLP yönteminde %68 doğruluk oranı elde edilmiştir. Ataklı dosyalarda en başarılı sonuç elde edilen 30 dB gürültü eklenmiş dosyalarda ise spectral yöntemi %8 daha başarılıdır. Spectral yöntemin medyan filtre uygulanmış dosyalarda %2, 32 Bit sıkıştırılmış ve 64 Bit sıkıştırılmış dosyalarda %9 daha başarılı sonuç ürettiği tespit edilmiştir.

Spectral yöntemini Öklid uzaklığı ile benzerlik hesaplayarak uyguladığımız yöntemin Arabic Speech Corpus veri tabanındaki sonuçlarını Tablo 23'te görülmektedir. Öklid uzaklığı benzerlik hesabında eşik değeri olarak 110 kullanılmıştır.

Öklid uzaklığı ile benzerlik hesaplaması verileri incelendiğinde en iyi performansı ataksız, medyan filtresi uygulanmış ve 20 dB gürültü eklenmiş dosyalarda tüm metriklerde aynı sonuç elde edilmiştir. %92 doğruluk oranı ile ataksız dosyalarda korelasyon katsayısı hesabına göre %3 daha başarılıdır. Korelasyon katsayısına göre medyan filtre uygulanmış dosyalarda ise %1'lik, 30 dB gürültü eklenmiş dosyalarda %91 doğruluk oranı ile %2 performans artışı söz konusudur. Bu üstünlüğe rağmen 20 dB gürültü eklenmiş dosyalarda %2 ve 64 Bit sıkıştırılmış dosyalarda %3 performans düşüşü gözlemlenmiştir. Ataksız

dosyalarda Öklid uzaklığı benzerlik hesaplaması en iyi performansı göstermektedir. Diğer ataklı dosyalarda benzer performans görülmektedir.

Tablo 23. Arabic speech corpus veri tabanından elde edilen sahte veri tabanında Spectral yöntemli özellik çıkarım ve Öklid uzaklığı ile sahtecilik tespiti performans değerleri

|                    | Duyarlılık | Doğruluk | Kesinlik | F1 Skor |
|--------------------|------------|----------|----------|---------|
| Ataksız            | 0,97       | 0,92     | 0,88     | 0,92    |
| Medyan Filtrelleme | 0,97       | 0,92     | 0,88     | 0,92    |
| 20 dB Gürültü      | 0,97       | 0,92     | 0,88     | 0,92    |
| 30 dB Gürültü      | 0,93       | 0,91     | 0,88     | 0,90    |
| 64 Bit Sıkıştırma  | 0,93       | 0,91     | 0,88     | 0,90    |

Arabic Speech Corpus veri tabanındaki PLP özellik çıkarım yöntemi ile spectral özellik çıkarım yönteminin Öklid uzaklığı ile benzerlik hesabı sonuçlarını kıyasladığımızda tüm ataklı ve ataksız dosyalarda spectral yönteminin daha başarılı sonuç ürettiği görülmektedir. Ataksız dosyalarda %92 doğruluk oranı ile PLP yöntemine göre %10 daha üstün sonuç elde etmiştir. Medyan filtre uygulanmış dosyalarda ve 20 dB gürültü eklenmiş dosyalarda %12, 30 dB gürültü eklenmiş dosyalarda ve 64 Bit sıkıştırılmış dosyalarda %11 performans artışı gözlemlenmiştir. Aynı kıyaslamanın Timit veri tabanında da benzer değerlendirme sonucu verdiği görülmektedir. Spectral yönteminin Öklid uzaklığı ile PLP yöntemine göre performans artışı olduğu her iki veri tabanında da görülmektedir.

Spectral özellik çıkarım yöntemini benzerlik hesaplaması için DTW yöntemi uygulayarak elde edilen sonuçlar Tablo 24’de görülmektedir. DTW benzerlik hesabı için eşik değeri olarak 188 değeri belirlenmiştir. Genel olarak tüm metriklerde en iyi performans medyan filtre uygulanmış dosyalarda gözlenmiştir. %91 doğruluk oranı ile en yakın performans %1 kayıp ile ataksız dosyalarda gözlenmiştir. En düşük performans değeri %80 doğruluk oranı ile 20 dB gürültü eklenmiş dosyalarda medyan filtre uygulanmış dosyalara göre %12 kayıpla elde edilmiştir. 30 dB gürültü eklenmiş dosyalarda %3 daha düşük performans elde edilirken 64 Bit sıkıştırılmış dosyalarda ise performans kaybı %6 olarak ölçülmüştür. DTW yöntemi ile yapılan benzerlik hesaplaması, Öklid uzaklığı ile kıyaslandığında belirgin bir performans düşüşü göstermiştir. Tüm metrikler incelendiğinde en iyi sonucu medyan filtresi uygulanmış dosyalarda Öklid uzaklığı hesaplamasında da %92 doğruluk oranı ile benzer sonuç elde edilmiştir. Ataksız dosyalarda %1 performans kaybı görülürken en düşük performans değeri elde edilen 20 dB gürültü eklenmiş dosyalarda %12

düşüşle ciddi performans kaybı görülmüştür. 30 dB gürültü eklenmiş dosyalarda %2 kayıp yaşanırken, 64 Bit sıkıştırılmış dosyalarda ise %5 performans düşüklüğü yaşanmıştır.

Tablo 24. Arabic speech corpus veri tabanından elde edilen sahte veri tabanında Spectral yöntemi özellik çıkarım ve DTW ile sahtecilik tespiti performans değerleri

|                    | Duyarlılık | Doğruluk | Keskinlik | F1 Skor |
|--------------------|------------|----------|-----------|---------|
| Ataksız            | 0,93       | 0,91     | 0,88      | 0,90    |
| Medyan Filtrelleme | 0,97       | 0,92     | 0,94      | 0,95    |
| 20 dB Gürültü      | 0,70       | 0,80     | 0,95      | 0,81    |
| 30 dB Gürültü      | 0,90       | 0,89     | 0,87      | 0,89    |
| 64 Bit Sıkıştırma  | 0,83       | 0,86     | 0,86      | 0,85    |

Arabic Speech Corpus veri tabanında elde ettiğimiz sonuçları aynı yöntemle Timit veri tabanından elde ettiğimiz sonuçlarla kıyasladığımızda Tablo 22'deki Timit veri tabanındaki sonuçlarda %95 doğruluk oranı ile en iyi sonucu ataksız dosyalarda görülürken %4 daha başarılı sonuç alınmıştır. Timit veri tabanında diğer başarılı sonuç %90 doğruluk oranı ile 30 dB gürültü eklenmiş dosyalardan %1 daha iyi olduğu gözlenmiştir. Arabic Speech Corpus veri tabanında ise en iyi sonucu %92 doğruluk oranı ile medyan filtre eklenmiş dosyalarda görülürken Timit veri tabanına göre %15 daha iyi sonuç elde edilmiştir. Arabic Speech Corpus veri tabanında 20 dB gürültü eklenmiş dosyalarda %8, 64 Bit sıkıştırılmış dosyalarda ise %4 daha iyi sonuç elde edilmiştir.

#### 2.5.4.MFCC Yöntemi ile Kopyala – Yapıştır Sahteciliği Tespitinin Değerlendirilmesi ve Elde Edilen Deneysel Sonuçlar

İnsan kulağı, farklı frekanslardaki seslere farklı duyarlılıklara sahiptir ve MFCC bu duyarlılığı matematiksel bir modelle temsil eder. MFCC özellik çıkarım yöntemi ve Öklid uzaklığı ile benzerlik hesaplanarak elde edilen performans değerleri Tablo 25'de görülmektedir. Benzerlik hesabı ile sahtecilik tespiti için eşik 5 olarak belirlenmiştir. Öklid uzaklığını, korelasyon katsayısına göre Tablo 11 ile kıyasladığımızda %97 doğruluk oranı ile en iyi başarı gösterilen ataksız dosyalarda korelasyon katsayısı ile yapılan hesaplamada eşit olarak gerçekleşmiştir. Ancak Medyan filtre uygulanmış dosyalarda %91, 20 dB gürültü eklenmiş dosyalarda %81 doğruluk oranı ile, Öklid uzaklığında %2 başarı düşüklüğü görülmektedir. 30 dB gürültü eklenmiş, 32Bit sıkıştırılmış ve 64 Bit sıkıştırılmış dosyalarda ise performans kaybı %1 dir.

Tablo 25. Timit veri tabanından elde edilen sahte veri tabanında MFCC yöntemi özellik çıkarım ve Öklid uzaklığı ile sahtecilik tespiti performans değerleri

|                    | Duyarlılık | Doğruluk | Kesinlik | F1 Skor |
|--------------------|------------|----------|----------|---------|
| Ataksız            | 0,98       | 0,97     | 0,98     | 0,98    |
| Medyan Filtrelleme | 0,90       | 0,91     | 0,97     | 0,93    |
| 20 dB Gürültü      | 0,77       | 0,81     | 0,97     | 0,86    |
| 30 dB Gürültü      | 0,96       | 0,95     | 0,98     | 0,97    |
| 32 Bit Sıkıştırma  | 0,82       | 0,85     | 0,97     | 0,89    |
| 64 Bit Sıkıştırma  | 0,84       | 0,86     | 0,97     | 0,90    |

F1 skoru ile değerlendirdiğimizde ise ataksız dosyalarda, 30 dB gürültü eklenmiş dosyalarda ve 64 Bit sıkıştırılmış dosyalarda aynı skor elde edilirken medyan filtre uygulanmış ve 20 dB gürültü eklenmiş dosyalarda %2, 32 Bit sıkıştırılmış dosyalarda %1 daha düşük başarı elde etmiştir. Her iki benzerlik hesaplama yönteminde de yakın değerler çıksa da Öklid uzaklığı hesabının daha az başarılı olduğu görülmektedir.

Öklid uzaklığı hesaplaması ile MFCC yöntemi Tablo 25 ve PLP yöntemi Tablo 13'teki doğruluk oranına göre kıyaslandığında MFCC yönteminin en iyi performans sağladığı ataksız dosyalardaki %97'lik doğruluk oranı ile PLP yöntemine göre %11 daha başarılıdır. Aynı şekilde en az başarı sağlanan 20 dB gürültü eklenmiş dosyalarda elde edilen %81 doğruluk oranı ile yine %11 başarılıdır. MFCC yönteminin Öklid uzaklığı hesaplamasında Medyan filtre uygulanmış, 30 dB gürültü eklenmiş dosyalarda %13, 32Bit sıkıştırılmış dosyalar %15, 64 Bit sıkıştırılmış dosyalarda ise %14 da üstün olduğu açıkça görülmektedir.

MFCC özellik çıkarım yöntemi ve DTW ile benzerlik hesaplanarak elde edilen performans değerleri Tablo 26'da görülmektedir. Eşik değeri olarak 35 kullanılmıştır. DTW ile benzerlik hesaplanarak elde edilen en iyi sonuç %90 doğruluk oranı ile yine ataksız sahte dosyalarda olmuştur. Ancak Öklid uzaklığı ile yapılan aynı teste %97 doğruluk oranına göre göre %7 daha düşük sonuç elde edilmiştir. En düşük performans elde edilen 20 dB gürültü eklenmiş dosyalarda %69 doğruluk oranı ile Öklid uzaklığına göre %12 daha düşük performans sağlamıştır. Ataklı dosyalarda %87 ile 30 dB gürültü eklenmiş dosyalarda en iyi performansı sağlarken Öklid uzaklığına göre %8 daha düşük doğruluk oranı elde etmiştir. Medyan filtre uygulanmış dosyalarda %78 doğruluk oranı ile %13, 30 dB gürültü eklenmiş dosyalarda %87 doğruluk oranı ile %8, 32 Bit sıkıştırılmış dosyalarda %74 doğruluk oranı ile %11 ve 64 Bit sıkıştırılmış dosyalarda ise %79 doğruluk oranı ile %7 daha düşük performans göstermiştir.

Tablo 26. Timit veri tabanından elde edilen sahte veri tabanında MFCC yöntemi özellik çıkarım ve DTW hesaplaması ile sahtecilik tespiti performans değerleri

|                    | Duyarlılık | Doğruluk | Kesinlik | F1 Skor |
|--------------------|------------|----------|----------|---------|
| Ataksız            | 0,90       | 0,90     | 0,96     | 0,93    |
| Medyan Filtrelleme | 0,74       | 0,78     | 0,95     | 0,83    |
| 20 dB Gürültü      | 0,62       | 0,69     | 0,94     | 0,75    |
| 30 dB Gürültü      | 0,87       | 0,87     | 0,96     | 0,91    |
| 32 Bit Sıkıştırma  | 0,69       | 0,74     | 0,95     | 0,80    |
| 64 Bit Sıkıştırma  | 0,76       | 0,79     | 0,95     | 0,84    |

DTW ile benzerlik hesaplamasını PLP de uygulanmış Tablo 14'deki DTW sonuçları ile kıyasladığımızda MFCC yönteminin ataksız dosyalarda sağladığı % 90 doğruluk oranı ile PLP yönteminde uygulanan DTW hesaplamasına göre % 3 daha başarılıdır. En düşük performans gösterdiği 20 dB gürültü eklenmiş dosyalarda %69 doğruluk oranı ile de %1 daha başarılıdır. Medyan filtre uygulanmış dosyalarda %3, 30 dB gürültü eklenmiş dosyalarda %5, 32 Bit sıkıştırılmış dosyalarda %6 ve 64 Bit sıkıştırılmış dosyalarda ise %6 daha üstün olduğu görülmektedir. En çok %6 ile sıkıştırılmış dosyalarda performans artışı gözlemlenmiştir. F1 skoruna göre karşılaştırdığımızda da MFCC yöntemi %93 performans değeri ile PLP ye göre %3 daha başarılıdır. %75 F1 skoru ile en düşük başarı elde edilen 20 dB gürültü eklenmiş dosyalarda %2 daha üstün olurken medyan filtresi uygulanmış dosyalarda %3 performans artışı gözlemlenmiştir. Ataklı dosyalardan %91 ile en iyi sonucu 30 dB gürültü eklenmiş dosyalarda elde ederken PLP yöntemine göre %4 daha başarılıdır. MFCC özellik çıkarım yöntemi 32 Bit sıkıştırılmış dosyalarda %7, 64Bit sıkıştırılmış dosyalarda ise %6 performans artışı göstermiştir.

Benzerlik hesabında Öklid uzaklığı kullanılarak Arabic Speech Corpus veri tabanında uygulanan MFCC yöntemi performans değerleri Tablo 27'de görülmektedir. Öklid uzaklığı benzerlik hesaplamasında eşik olarak 13 değeri kullanılmıştır. Elde edilen performans değerlerine baktığımızda Tüm dosya türlerinde hemen hemen aynı performans değeri ölçülmüştür. Ataksız, medyan filtre uygulanmış, 20 dB ve 30 dB gürültü uygulanmış dosyalarda tüm metriklerde aynı değerler elde edilmiştir. 64 Bit sıkıştırılmış dosyalarda ise %91 doğruluk oranı ile %1 daha düşük başarı elde edilmiştir.

Korelasyon hesabı ile uygulanan MFCC yöntemi ile doğruluk oranına göre kıyasladığımızda ataksız dosyalarda ve 64 Bit sıkıştırılmış dosyalarda %3'lük performans

iyileşmesi görülürken diğer ataklı dosyalarda ise %1'lik performans artışı görülmektedir. Diğer metriklere bakıldığında da benzer sonuçlar görülmektedir.

Tablo 27. Arabic speech corpus veri tabanından elde edilen sahte veri tabanında MFCC yöntemi özellik çıkarım ve Öklid uzaklığı ile sahtecilik tespiti performans değerleri

|                    | Duyarlılık | Doğruluk | Kesinlik | F1 Skor |
|--------------------|------------|----------|----------|---------|
| Ataksız            | 0,97       | 0,92     | 0,88     | 0,92    |
| Medyan Filtrelleme | 0,97       | 0,92     | 0,88     | 0,92    |
| 20 dB Gürültü      | 0,97       | 0,92     | 0,88     | 0,92    |
| 30 dB Gürültü      | 0,97       | 0,92     | 0,88     | 0,92    |
| 64 Bit Sıkıştırma  | 0,93       | 0,91     | 0,88     | 0,90    |

MFCC özellik çıkarım yöntemi DTW benzerlik hesabı ile uygulandığında elde edilen performans sonuçları Tablo 28'de görülmektedir. Tüm dosyalar değerlendirildiğinde çok yakın performans değerleri elde edilmiştir. Doğruluk oranına göre ataksız, medya filtre uygulanmış ve 20 dB gürültü eklenmiş dosyalarda %83 doğruluk oranı sağlanmıştır. Öklid uzaklığı benzerlik hesabına göre %9'luk bir performans düşüşü görülmektedir. 30 dB gürültü eklenmiş dosyalarda bu kayıp %10 olarak ölçülmüştür.

Tablo 28. Arabic speech corpus veri tabanından elde edilen sahte veri tabanında MFCC yöntemi özellik çıkarım ve DTW ile sahtecilik tespiti performans değerleri

|                    | Duyarlılık | Doğruluk | Kesinlik | F1 Skor |
|--------------------|------------|----------|----------|---------|
| Ataksız            | 0,87       | 0,83     | 0,79     | 0,83    |
| Medyan Filtrelleme | 0,87       | 0,83     | 0,79     | 0,83    |
| 20 dB Gürültü      | 0,87       | 0,83     | 0,79     | 0,83    |
| 30 dB Gürültü      | 0,83       | 0,82     | 0,78     | 0,81    |
| 64 Bit Sıkıştırma  | 0,83       | 0,82     | 0,78     | 0,81    |

PLP yönteminde uygulanan DTW yöntemi ile kıyasladığımızda benzer şekilde tüm dosya sınıflarında performans düşüşü görülmektedir. PLP yönteminde doğruluk oranına göre en fazla performans kaybı %11 ile ataksız dosyalarda gözlemlenirken MFCC yönteminde ise %9 düşüşle 30 dB gürültü eklenmiş dosyalarda görülmektedir.

Timit veri tabanındaki performans değerlendirmeleri ile kıyasladığımızda MFCC yöntemi ile DTW benzerlik hesabı yönteminde performans düşüşü benzer şekilde görülmektedir. Ölçüm sonuçlarına göre yöntemlerin farklı veri tabanlarındaki başarısı örtüşmektedir.

## 2.6. Sahte Segmentlerin Belirlenmesi

Kelimelerden elde edilen öz vektörlerle yapılan benzerlik hesabına göre ses sinyalinde kopyala-yapıştır sahteciliği yapıp yapılmadığına karar verilir. Hesaplanan yakınlık değerlerine sıra ile bakılarak Öklid ve DTW hesaplamasında en düşük olan, korelasyon katsayısı hesabında ise en yüksek olan ilişki tespit edilmeye çalışılır. Bu tespit yapılabilmesi için ses sinyalinin kayıt ortamı değerlendirilerek empirik olarak eşik değeri belirlenir. Bu eşik değerleri araştırmada kullanılan farklı veri tabanlarında farklı olacak şekilde belirlenmiştir.

Pearson korelasyon katsayısı hesabı yapılarak benzerlik tespitinde kelimeler arasındaki benzerlik oranı ne kadar büyük olduğuna bakılır. Kelimeler arasındaki hesaplanan maksimum korelasyon değeri eşik değerinden küçük olan ses sinyallerinin orijinal olduğuna karar verilir. Korelasyon değerinin büyük olması kelimeler arasındaki benzerlik ilişkisinin de büyük olduğu anlamına gelir. Eşik değerinden daha büyük bir değer elde edilmişse en büyük değeri veren kelimeler birbirinin kopyası olduğu kabul edilir ve kopyala-yapıştır sahteciliği yapıldığına karar verilir.

Tablo 29’da kelimeler arasında hesaplanan korelasyon değerleri görülmektedir. Timit veri tabanında sahte dosyalarda eşik değeri 0,95 olarak belirlenmiştir. Bu eşığe göre örnek ses dosyasında 0,95 ten daha büyük bir korelasyon değeri olduğundan bu dosyada kopyala-yapıştır sahteciliği yapıldığına ve hesaplanan maksimum korelasyon katsayısı değeri 0.999998344548311 ile tespit edilen 1. ve 4. kelimelerin birbirinin kopyası olduğuna karar verilir.

Tablo 29. Örnek ses dosyasından hesaplanan kelimeler arası korelasyon katsayıları

| Kelime Konum | Kelime Konum | Korelasyon Katsayısı |
|--------------|--------------|----------------------|
| 1            | 2            | 0.362601306482568    |
| 1            | 3            | 0.521046764895369    |
| 1            | 4            | 0.999998344548311    |
| 1            | 5            | 0.859336985394727    |
| 2            | 4            | 0.319496242534206    |
| 2            | 5            | 0.480169177145391    |
| 3            | 4            | 0.535838159614849    |
| 4            | 5            | 0.859295133394910    |



Kelimelerden elde edilen segmentlerin dizi boyutuna bakarak ve kelimelerden elde edilen özellik vektörlerinin boyutlarına bakarak kelimelerin sahte adayı olup olmadığına karar verilir. Sahte adayı kelimelerin dizi boyutunun yakın olması beklenir. Ayrıca segment boyutu küçük olan kelimeler de sahte adayı değildir. Aralarında bir benzerlik olmadığı düşünülen bu kelimeler arasındaki korelasyon hesabı yapılmamıştır. Bu ayırım hesaplamada kaynak maliyetini azaltır. Bu yüzden tablo 29’da görülen değerler benzerlik olabilecek kelimeler arasında hesaplama yapılmış korelasyon değerleridir.

Öklid uzaklığına bakılırken kelimeler arasındaki ilişkinin birbirine ne kadar yakın olduğuna bakılır. Kelimeler arasındaki hesaplanan minimum benzerlik mesafesi eşik değerinden büyük olan ses sinyallerinin orijinal olduğuna karar verilir. Eşik değerinin altında bir değer elde edilmişse minimum değeri veren kelimeler birbirinin kopyası olduğu kabul edilir ve kopyala-yapıştır sahteciliği yapıldığına karar verilir. DTW’ de Öklid uzaklığı gibi minimum değerleri ele alınırken eşik değeri metoda göre daha farklı seçilmiştir. Yine eşik değerinin altındaki minimum değeri veren kelimelerde kopyala-yapıştır sahteciliği yapıldığına karar verilir.

Tablo 30’da kelimeler arasında hesaplanan Öklid uzaklığı görülmektedir. Timit veri tabanında sahte dosyalarda eşik değeri 50 olarak belirlenmiştir. Bu eşığe göre örnek ses dosyasında 50 den daha küçük bir Öklid değeri olduğundan bu dosyada kopyala-yapıştır sahteciliği yapıldığına ve hesaplanan minimum Öklid uzaklığı değeri 0.779527170444566 ile tespit edilen 1. ve 4. kelimelerin birbirinin kopyası olduğuna karar verilir.

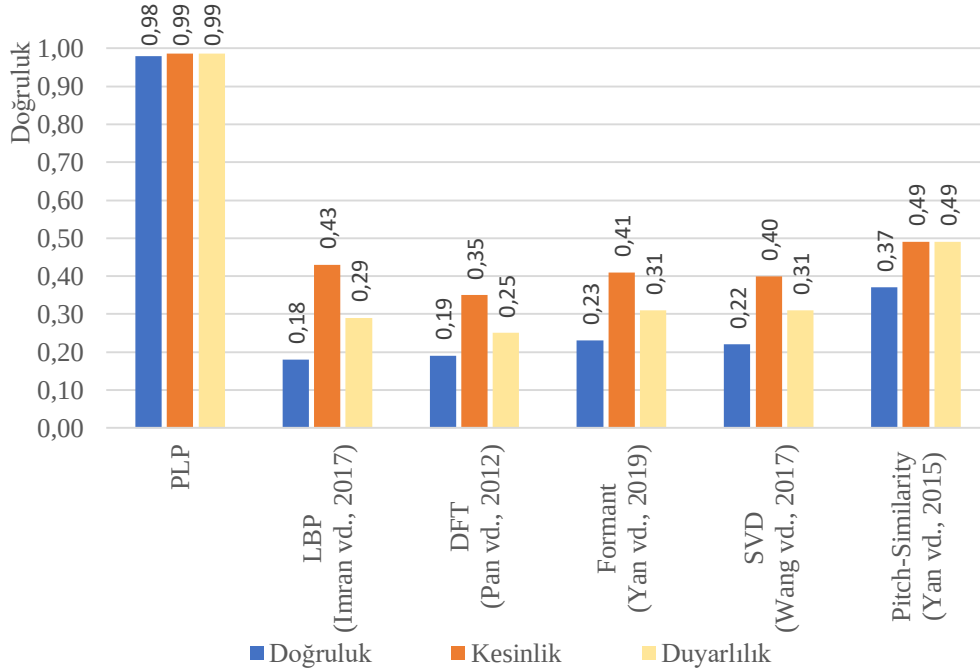
Tablo 30. Örnek ses dosyasından hesaplanan kelimeler arası Öklid uzaklığı

| Kelime Konum | Kelime Konum | Öklid uzaklığı    |
|--------------|--------------|-------------------|
| 1            | 2            | 329.690625136514  |
| 1            | 3            | 400.747680608963  |
| 1            | 4            | 0.779527170444566 |
| 1            | 5            | 703.597220335252  |
| 2            | 3            | 276.459681417101  |
| 2            | 4            | 329.174045160619  |
| 2            | 5            | 762.797714884074  |
| 3            | 4            | 400.195288980408  |
| 3            | 5            | 945.161199474450  |
| 4            | 5            | 703.978487984438  |

## 2.7. PLP Yönteminin Literatürdeki Yöntemlerle Karşılaştırılması

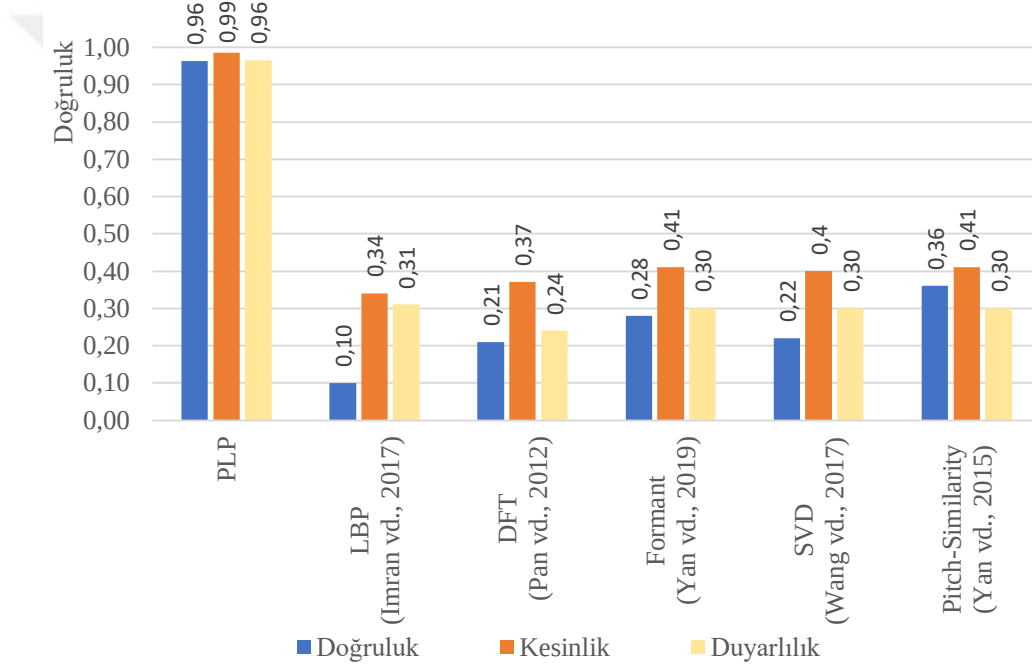
Tez kapsamında yapılan araştırmada genel olarak PLP özellik çıkarım yönteminin korelasyon katsayısı benzerlik hesabı ile kullanılan metot daha üstün olduğu görülmüştür. Önerilmiş olan PLP yönteminin başarısını analiz etmek için literatürde önceden uygulanmış farklı yöntemlerle karşılaştırılmıştır. Karşılaştırma yapılan bu yöntemler LBP (Imran vd., 2017), DFT (Pan vd., 2012), Formant (Yan vd., 2019), SVD (Wang vd., 2017) ve Perde-Benzeri (Pitch-Similarity) (Yan vd., 2015) yöntemleridir. Timit veri tabanındaki ataksız ve ataklı tüm performans analiz sonuçları doğruluk, duyarlılık, kesinlik metrikleri diğer yöntemlerle kıyaslanarak başarısı aşağıda irdelenmiştir (Üstübioğlu vd., 2022).

Şekil 17’de görüldüğü üzere en yüksek doğruluk, kesinlik ve duyarlılık değerlerine sahip yöntem olarak PLP görünüyor. Bu, PLP’nin ataksız dosyalarda kopyala – yapıştır ses sahteciliğini tespit etmede oldukça etkili olduğunu göstermektedir. Önerilen yöntem %98 doğruluk oranı ile başarı sağlamışken diğer yöntemlerin başarısı oldukça düşük kalmıştır. Karşılaştırılan yöntemlerden Pitch benzeri yöntem daha yakın olduğu görülse de istendik başarıdan oldukça uzak duruyor.



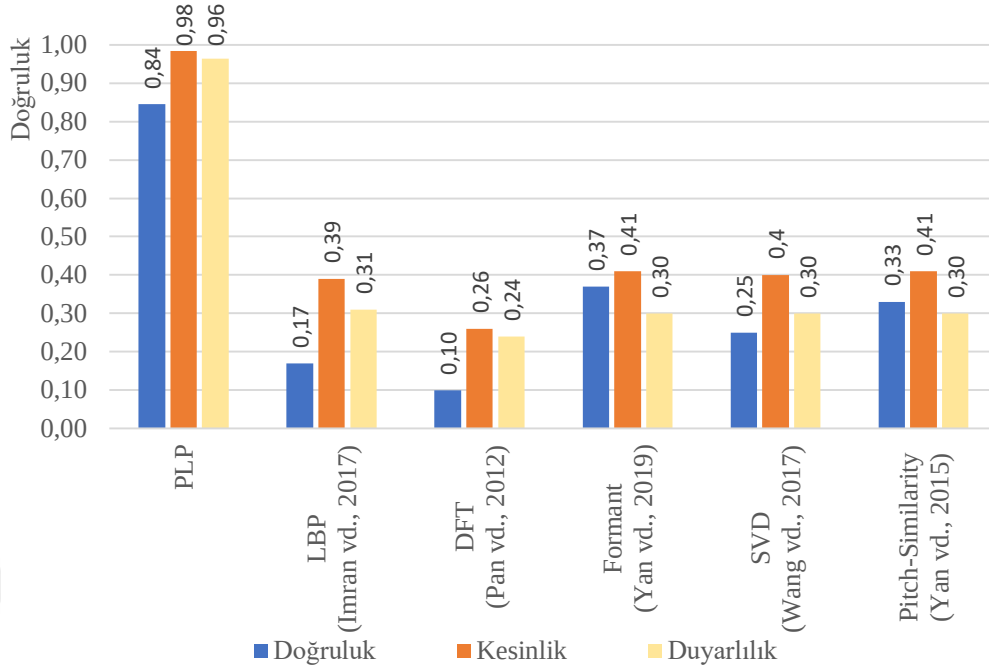
Şekil 18. Timit veri tabanında PLP yöntemi ile literatürdeki yöntemlerin ataksız ses sinyalinde sahtecilik tespit etme oranı karşılaştırılması

Önerilen yöntemin ataksız sinyallerdeki başarısı ataklı sinyallerde de açıkça görülmektedir. Şekil 18’de medyan filtresi uygulanmış ses dosyaları üzerindeki performansı görülmektedir. Medyan filtrede bir miktar performans kaybı olsa da önerilen yöntem diğer yöntemlere göre oldukça başarılıdır. Tüm metriklerde (doğruluk, kesinlik ve duyarlılık) en yüksek değerleri elde eden yöntem PLP’dir. PLP’nin ses dosyası kopyala-yapıştır sahteciliği tespitinde medyan filtresi uygulanmış ataklı dosyalarda da daha etkili yöntem olduğunu görülmektedir. Pitch benzeri yöntem doğruluk ve kesinlik değerlerinde LBP, DFT, formant tabanlı ve SVD’ye göre daha iyi performans gösterse de %36 doğruluk oranı ile PLP’nin oldukça gerisindedir.



Şekil 19. Timit veri tabanında PLP yöntemi ile literatürdeki yöntemlerin medyan filtreleme uygulanmış ses sinyalinde sahtecilik tespit etme oranı karşılaştırılması

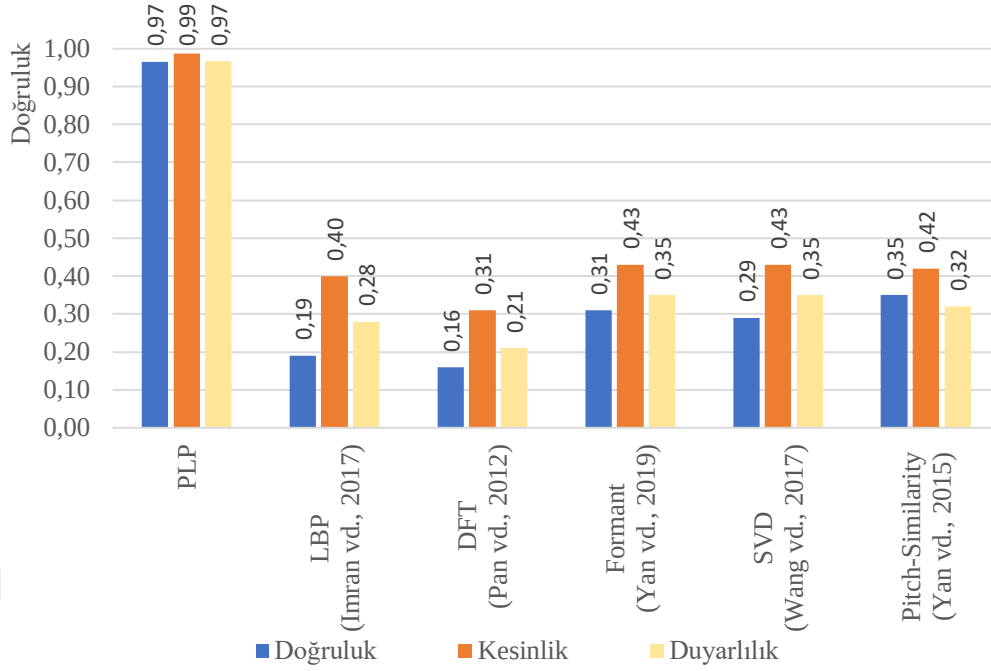
PLP yönteminin en düşük başarıyı gösterdiği 20 dB’lik gaussian gürültüsü uygulanmış dosyalardaki sonuçları şekil 19’da görülmektedir. Önerilen yöntemin performansı düşmüş olmasına rağmen en yakın doğruluk oranı ile pitch benzeri yöntemle göre %47 daha başarılı olduğu görülmektedir. Literatürdeki diğer yöntemler arasında formant tabanlı yöntemin %37 ile en başarılı yöntem olduğu dikkat çekmiştir.



Şekil 20. Timit veri tabanında PLP yöntemi ile literatürdeki yöntemlerin 20 dB gürültü eklenmiş ses sinyalinde sahtecilik tespit etme oranı karşılaştırılması

Şekil 20 de ise 30 dB’lik gaussian gürültüsü uygulanmış dosyalardaki sonuçlar görülmektedir. 30 dB’deki performansı 20 dB’e göre çok daha iyidir. Önerilen yöntem doğruluk oranına göre %13 daha başarılı olurken diğer yöntemlerde gözle görülür bir performans değişikliği olmamıştır. 30 dB gürültü eklenmiş dosyalarda genel olarak tüm yöntemlerde performans artışı gözlenirken formant tabanlı yöntemin performansı doğruluk oranına göre %6 düşmüştür.

30 dB gürültü seviyesi, PLP’nin doğruluk değerini artırmış ve diğer metriklerde (kesinlik ve duyarlılık) %1 farkla daha iyi sonuç vermiştir. Bu, PLP’nin 30 dB gürültü seviyesinde 20 dB gürültü seviyesine göre daha iyi performans sağladığını göstermektedir.



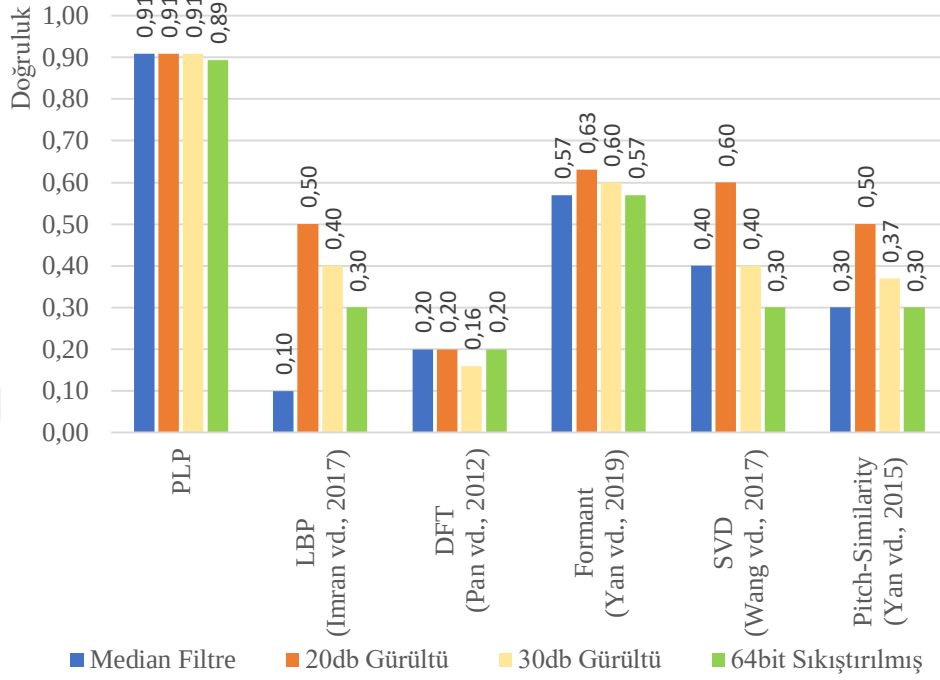
Şekil 21. Timit veri tabanında PLP yöntemi ile literatürdeki yöntemlerin 30 dB gürültü eklenmiş ses sinyalinde sahtecilik tespit etme oranı karşılaştırılması

Arabic Speech Corpus veri tabanından elde edilmiş sahte ses dosyaları ile yapılan çalışmaların sonuçları literatürde önceden önerilmiş yöntemlerin başarısı ile aşağıda karşılaştırılmıştır. Ataklı sahte ses dosyalarında yapılan test sonuçları ile literatürde önerilen diğer yöntemlerin doğruluk oranına göre karşılaştırması Şekil 21’de görülmektedir (Üstübioğlu vd., 2023). Orijinal dosyalara medyan filtresi, 20 dB gürültü, 30 dB gürültü ve 64 bit sıkıştırma uygulanarak elde edilen sahte dosyalarda yapılmış sonuçlara göre analiz edilmiştir.

Sahte dosyalarda meydan filtresi eklenerek elde edilen dosyalar üzerinde yapılan çalışmada PLP yöntemi %91 doğruluk oranı ile en başarılı yöntem olarak yine öndedir. En yakın başarıyı göstermiş olan formant tabanlı yöntemle göre %34 daha üstün başarı sağlamıştır. Diğer yöntemler SVD %40, pitch benzeri yöntemle %30, DFT %20 ve LBP yöntemi %10 doğruluk oranı ile performans sağlamıştır. PLP yönteminin %91 gibi yüksek doğruluk oranı ile güvenilir olduğu görülmektedir.

20 dB gürültü eklenmiş dosyalarda ise literatürdeki diğer yöntemlerin daha iyi performans gösterdiği şekilde görülmektedir. Buna rağmen literatürdeki yöntemlerden en yüksek %63 doğruluk oranı ile formant tabanlı yöntem olsa da çalışmada önerilen PLP yöntemi %91 doğruluk oranı ile daha üstün başarı göstermiştir. 20 dB gürültü eklenmiş sahte

dosyalarda tüm yöntemler performans artışı sağlarken DFT yöntemi performans artışı sağlamamıştır.



Şekil 22. Arabic Speech Corpus veri tabanında PLP yöntemi ile literatürdeki yöntemlerin medyan filtre eklenmiş ses sinyalinde sahtecilik tespit etme oranı karşılaştırılması

30 dB gürültü eklenmiş sahte dosyalarda literatürde önerilmiş yöntemlerin hepsinde başarının düştüğü görülürken. PLP yöntemi 20 dB gürültüye göre 30 dB de aynı performansı sağlamıştır. Formant tabanlı yöntem literatürde önerilen yöntemlere göre %60 doğruluk oranı ile en iyi sonucu elde etmiş olsa bile PLP yönteminin %31 farkla oldukça gerisinde kalmıştır.

Tüm yöntemlerde performans düşüklüğü gözlenen 64 Bit sıkıştırılmış dosyalarda %2 performans kaybıyla yine en iyi başarıyı PLP yönteminin sağladığı görülmektedir. Literatürdeki yöntemlerden en iyi başarıyı %57 doğruluk oranı ile formant tabanlı yöntem sağlarken %30 doğruluk oranı ile pitch benzeri yöntem, SVD ve LBP yöntemleri aynı performansı göstermiştir. En kötü performans ise %20 doğruluk oranı ile DFT yöntemi göstermiştir.

### 3. SONUÇLAR

Bu tez kapsamında ses dosyası kopyala-yapıştır sahtecilik tespiti için farklı özellik çıkarım yöntemleri test edilerek karşılaştırılmıştır. Ses tanıma sistemlerinde yaygınca kullanılan bu yöntemlerin avantajlarını kullanarak kopyala-yapıştır sahteciliği tespitinde de fayda sağlayıp sağlamadığı araştırılmıştır. Timit ve Arabic Speech Corpus ses veri tabanı dosyaları ile çalışılmış ve yöntemlerin farklı dillerdeki, farklı kayıt ortamlarındaki başarısı ölçülmüştür.

Timit veri tabanı orijinal dosyalar ve orijinal dosyalardan türetilmiş ataksız, medyan filtre uygulanmış, 20 dB ve 30 dB gürültü eklenmiş, 32 Bit ve 64 Bit sıkıştırma yapılmış sahte toplam 2236 dosyalardan oluşmaktadır. Arabic Speech Corpus veri tabanı ise orijinal, ataksız, medyan filtre uygulanmış, 20 dB ve 30 dB gürültü eklenmiş, 64 Bit sıkıştırma uygulanmış sahte toplam 186 dosyadan oluşmaktadır.

Özellik çıkarım yöntemlerini uygularken farklı benzerlik hesaplama yöntemlerini de kullanılarak hangi yöntemin kopyala-yapıştır sahtecilik tespitinde daha verimli olduğu araştırılmıştır.

Timit veri tabanında ya uygulanan PLP özellik çıkarım yöntemi, korelasyon katsayısı, Öklid uzaklığı ve DTW benzerlik hesabı ile ayrı ayrı test edilmiştir. Elde edilen sonuçlara göre ataksız dosyalarda %98 doğruluk oranı ile korelasyon katsayısı yönteminin Öklid uzaklığına göre %12 DTW yöntemine göre ise %11 daha üstün başarı göstermiştir. Korelasyon katsayısı ile en düşük performans gösterilen 20 dB gürültü eklenmiş dosyalarda %84 başarı sağlamıştır. Buna karşın Öklid uzaklığından %14, DTW yönteminden %16 daha iyi sonuç elde etmiştir. Diğer metriklere göre de PLP yönteminin korelasyon katsayısı ile daha üstün olduğu görülmektedir.

Arabic Speech Corpus veri tabanında yapılan çalışmada ise ataksız dosyalarda PLP yönteminin korelasyon katsayısı ile elde ettiği başarı %91 doğruluk oranı ile Öklid uzaklığına göre %9, DTW hesaplamasına göre de %11 daha başarılıdır. Elde edilen sonuçlar Timit veri tabanı test sonuçları ile benzerlik göstermektedir.

Rasta-PLP uygulanmış yöntemde konuşma dışındaki çevresel sesleri ve ani ses değişikliklerini bastırarak insan işitmesini taklit eden bir algoritma kullanılmıştır. Rasta-PLP özellik çıkarım yöntemini korelasyon katsayısı benzerlik hesabı ile ataksız dosyalarda %90 doğruluk oranı ile en iyi sonucu vermiştir. Ataksız dosyalardaki bu değer Öklid uzaklığına

göre %7, DTW yöntemine göre %11 daha üstün sonuç vermiştir. En düşük sonucu korelasyon hesabı ile 20 dB gürültü eklenmiş dosyalarda elde ederken Öklid uzaklığında ve DTW hesabında ise medyan filtre eklenmiş dosyalarda en düşük sonuçlar elde edilmiştir. Buna rağmen 20 dB gürültü eklenmiş dosyalarda %73 doğruluk oranı ile Öklid uzaklığına göre %7, DTW ye göre %18 daha iyi sonuç elde etmiştir.

Arabic Speech Corpus veri tabanındaki sonuçlara baktığımızda da ataksız dosyalarda %89 başarı ile korelasyon katsayısı ile benzerlik hesabı Öklid uzaklığına göre %3, DTW hesabına göre %6 daha üstündür. Rasta-PLP yönteminde de korelasyon katsayısı benzerlik hesabı daha üstün sonuç üretmiştir.

Araştırmamızda uyguladığımız diğer bir yöntem spectral ile uyguladığımız benzerlik hesabı yöntemlerinde ise ataksız dosyalarda %95 doğruluk oranı ile DTW hesaplamasının en iyi başarıyı gösterdiği görülmektedir. Öklid hesabına göre %4 korelasyon hesabına göre %9 daha üstün başarı sağlamıştır. En düşük performans elde edilen 20 dB gürültü eklenmiş dosyalarda %72 ile Öklid uzaklığına göre %1 düşüş görülse de diğer metrikler incelendiğinde bu kayıp göz ardı edilmiştir. Korelasyon hesabına göre de %5 daha iyi sonuç verdiği görülmektedir.

Arabic Speech Corpus veri tabanındaki performansa bakıldığında Öklid uzaklığı ile uygulanan yöntemin daha başarılı olduğu görülmektedir. Ataksız dosyalarda %92 doğruluk oranı ile DTW hesabına göre %1, korelasyon hesabına göre %3 daha başarılıdır. Burada dikkat çeken durum diğer yöntemlerin ölçümlerinde en düşük değeri 20 dB gürültü eklenmiş dosyalarda elde ederken Arabic Speech Corpus veri tabanında spectral özellik çıkarım ve korelasyon katsayısı ile benzerlik hesabı %94 doğruluk oranı ile en iyi performansı göstermiştir. Diğer tüm dosyalar değerlendirildiğinde spectral ile DTW hesabı kullanıldığında genel olarak daha iyi sonuç vermektedir.

MFCC özellik çıkarım yönteminden elde ettiğimiz sonuçlara göre korelasyon katsayısı ile benzerlik hesabında ataksız dosyalarda %97 doğruluk oranı ile en iyi performansı göstermiştir. Test sonuçlarına göre Öklid uzaklığı ile yapılan benzerlik hesabı da aynı sonuç verse de korelasyon katsayısı ile benzerlik hesabı medyan filtresi uygulanmış dosyalarda ve 20 dB gürültü eklenmiş dosyalarda %2, 30 dB gürültü eklenmiş, 32 Bit ve 64 Bit sıkıştırılmış dosyalarda ise %1 daha iyi sonuç vermiştir.

Arabic Speech Corpus veri tabanındaki performansa bakıldığında Öklid uzaklığı ile uygulanan yöntemin korelasyon katsayısına göre ataksız dosyalarda ve 64 Bit sıkıştırılmış dosyalarda %3, medyan filtre, 20 dB ve 30 dB gürültü eklenmiş dosyalarda ise %1'lik



üstünlük sağlamıştır. Genel olarak MFCC yöntemi korelasyon katsayısı ve Öklid uzaklığı ile elde edilen sonuçlar birbirine yakın ve benzer başarı oranlarına sahiptir.

Uygulanan PLP, Rasta-PLP, Spectral ve MFCC yöntemlerini kıyasladığımızda en üstün sonucu PLP yönteminin korelasyon katsayısı ile benzerlik hesabından elde edilmiştir. Ataksız dosyalarda %98 doğruluk oranı ile en iyi performansı göstermişken en yakın başarıyı %97 ile MFCC yönteminin korelasyon katsayısı ile uygulanmış yöntemde görülmüştür. %91 doğruluk oranı ile Spectral özellik çıkarımının en iyi sonucu aldığı DTW benzerlik hesabı kullanılan yöntemdir. PLP yöntemine göre %7 daha düşük performans göstermiştir. En düşük performansı ise %89 doğruluk oranı ile Rasta-PLP yöntemi korelasyon katsayısı ile sağlamıştır. PLP ye göre %9 daha kötü sonuç elde edilmiştir.

PLP yöntemini literatürdeki diğer yöntemlerle kıyasladığımızda ise En yüksek doğruluk, kesinlik ve duyarlılık değerlerine sahip, bu da PLP'nin kıyaslanan yöntemlere göre daha güvenilir yöntem olduğunu göstermektedir. 20 dB gürültü eklenmiş dosyalarda performansta hafif bir düşüş olmasına rağmen, kesinlik ve duyarlılık değerleri neredeyse aynı kalmıştır. Bu, PLP'nin gürültüye karşı dayanıklılığını gösterir.

Pitch benzeri yöntem çok başarılı olmasa da PLP'den sonra en iyi performansı sergileyen yöntemdir, ancak gürültü seviyesinden etkilendiği gözlemlenmektedir. LBP ve DFT yöntemleri de düşük performans sergilemektedir ve gürültü eklenmesi durumunda performansları daha da düşmektedir. Bu yöntemler sahtecilik tespitinde PLP'ye göre yeterince etkili değildir.

#### 4. ÖNERİLER

Tez çalışmasında ses dosyalarındaki kopyala-yapıştır sahteciliğini tespit etmek için farklı özellik çıkarım yöntemlerinin başarıları irdelenmiştir. Öneri olarak uygulamada veri tabanına göre manuel olarak belirlenen eşik değerlerinin otomatik belirleyen algoritmalarla desteklenebilir. Benzerlik tespitinde kullanılan sınıflandırma yöntemleri farklı ataklarda farklı üstünlük gösterdiğinden hibrit bir model geliştirilebilir. Gürültü ataklarında daha başarılı sonuç üreten özellik çıkarım ve sınıflandırma yöntemi uygulanabilir.

Kopyala-yapıştır sahtecilik tespitinde güvenilir sonuçlar için PLP yöntemi tercih edilebilir. Yüksek doğruluk, kesinlik ve duyarlılık değerleri, bu yöntemin sahtecilik tespitinde etkili bir çözüm olduğunu göstermektedir.

## 5. KAYNAKLAR

- Akdeniz, F., & Becerikli, Y. (2024). Recurrent Neural Network and Long Short-Term Memory Models for Audio Copy-Move Forgery Detection: a Comprehensive Study. *The Journal of Supercomputing*, 80, 17575–17605.
- Akdeniz, F., & Becerikli, Y. (2021). Detection of Copy-Move Forgery in Audio Signal with Mel Frequency and Delta-Mel Frequency Cepstrum Coefficients. *2021 Innovations in Intelligent Systems and Applications Conference (ASYU)*, (s. 1-6). Elazığ.
- Ali, Z., Imran, M., & Alsulaiman, M. (2017). An Automatic Digital Audio Authentication/Forensics System. *IEEE Access*, 5, 2994-3007.
- Almutairi, Z., & Elgibreen, H. (2022). A Review of Modern Audio Deepfake Detection Methods. *Algorithms*, 15(155), 1-20.
- Anh, N. T., Hang, H. T., & Chen, G. (2019). One Approach In The Time Domain In Detecting Copy Move Of Speech Recordings With The Similar Magnitude. *International Journal of Engineering and Applied Sciences (IJEAS)*, 6(4), 11-13.
- Buchholz, R., Kraetzer, C., & Ditt, J. (2009). Microphone Classification Using Fourier Coefficients. *11th International Workshop on Information Hiding* (s. 235-246). Darmstadt: Springer.
- Bykhovsky, D. (2020). Recording Device Identification by ENF Harmonics Power Analysis. *Forensic Science International*, 307, 1-3.
- Campbell, W. M., Campbell, J. P., Gleason, T. P., Reynolds, D. A., & Shen, W. (2007). Speaker Verification Using Support Vector Machines and High-Level Features. *IEEE Transactions on Audio, Speech, and Language Processing*, 15(7), 2085-2094.
- Chakravarty, N., & Dua, M. (2023). Data Augmentation And Hybrid Feature Amalgamation To Detect Audio Deep Fake Attacks. *Physica Scripta*, 98(9), 1-19.
- Champod, C., & Meuwly, D. (2000). The Inference of Identity in Forensic Speaker Recognition. *Speech Communication*, 31(2), 193-203.
- Chen, B., Luo, W., & Luo, D. (2018). Identification of Audio Processing Operations Based on Convolutional Neural Network. *IH and MMSec 2018 - Proceedings of the 6th ACM Workshop on Information Hiding and Multimedia Security* (s. 73-77). IEEE.
- Cuccovillo, L., Mann, S., Tagliasacchi, M., & Aichroth, P. (2013). Audio Tampering Detection via Microphone Classification. *2013 IEEE 15th International Workshop on Multimedia Signal Processing (MMSP)* (s. 177-182). Pula: IEEE.
- Dave, N. (2013, Temmuz). Feature Extraction Methods LPC, PLP and MFCC. *International Journal For Advance Research In Engineering And Technology*, 1(6), 1-6.
- Galou, G., & Chollet, G. (2011). Synthetic Voice Forgery in the Forensic Context. *Forensic Speech and Audio Analysis Working Group (ENFSI-FSAAWG)*, (s. 1-3). Rome.

- Hamza, A., Javed, A. R., Iqbal, F., Kryvinska, N., Almadhor, A. S., Jalil, Z., & Borghol, R. (2022). Deepfake Audio Detection via MFCC Features Using Machine Learning. *IEEE Access*, 10, 134018-134028.
- Hermansky, H. (1990, Nisan). Perceptual Linear Predictive (PLP) Analysis of Speech. *Acoustical Society of America*, 87(4), 1738-1752.
- Hermansky, H., & Morgan, N. (1994). RASTA Processing of Speech. *IEEE Transactions on Speech and Audio Processing*, 2(4), 578-589.
- Hiçsönmez, S., Uzun, E., & Sencar, H. T. (2013). Methods for Identifying Traces of Compression in Audio. *1st International Conference on Communications, Signal Processing, and their Applications (ICCSPA)*, (s. 1-6). Sharjah.
- Huang, L., & Pun, C.-M. (2020). Audio Replay Spoof Attack Detection by Joint Segment-Based Linear Filter Bank Feature Extraction and Attention-Enhanced DenseNet-BiLSTM Network. *IEEE/ACM Transactions on Audio, Speech, and Language Processing*, 28, 1813-1825.
- Huang, X., Liu, Z., Lu, W., Liu, H., & Xiang, S. (2019, Nisan). Fast and Effective Copy-Move Detection. *International Journal of Digital Crime and Forensics*, 11(2), 47-62.
- Ikram, S., & Malik, H. (2010). Digital Audio Forensics Using Background Noise. *2010 IEEE International Conference on Multimedia and Expo* (s. 106-110). Singapore: IEEE.
- Imran, M., Ali, Z., Bakhsh, S. T., & Akram, S. (2017, Temmuz). Blind Detection of Copy-Move Forgery in Digital Audio Forensics. *IEEE Access*, 5, 12843-12855.
- Jadhav, S., Patole, R., & Rege, P. (2019). Audio Splicing Detection using Convolutional Neural Network. *2019 10th International Conference on Computing, Communication and Networking Technologies (ICCCNT)* (s. 1-5). Kanpur: IEEE.
- Kraetzer, C., Oermann, A., Dittmann, J., & Lang, A. (2007). Digital Audio Forensics: A First Practical Evaluation on Microphone and Environment Classification. *Proceedings of the 9th Workshop on Multimedia & Security*, (s. 63-74). Dallas.
- Kua, J. K., Thiruvaran, T., Nosratighods, M., Ambikairajah, E., & Epps, J. (2010). Investigation of Spectral Centroid Magnitude and Frequency for Speaker. *The Speaker and Language Recognition Workshop* (s. 34-39). Brno: Odyssey.
- Lin, X., Liu, J., & Kang, X. (2016). Audio Recapture Detection With. *IEEE Transactions on Multimedia*, 18(8), 1480-1487.
- Liu, Q., Sung, A. H., & Qiao, M. (2010). Detection of Double MP3 Compression. *Cognitive Computation*, 2, 291-296.
- Luo, D., Luo, W., Yang, R., & Huang, J. (2012). Compression History Identification For Digital Audio Signal. *IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, (s. 1733-1736). Kyoto.
- Luo, D., Yang, R., Li, B., & Huang, J. (2017). Detection of Double Compressed AMR Audio. *IEEE Transactions On Information Forensics And Security*, 12(2), 432-444.

- Malik, H., & Farid, H. (2010). Audio Forensics From Acoustic Reverberation. *2010 IEEE International Conference on Acoustics, Speech and Signal Processing* (s. 1710-1713). Dallas: IEEE.
- Mannepalli, K. P., Krishna, P. V., Krishna, K. V., & Krishna, K. R. (2019, Aralık). Copy and Move Detection in Audio Recordings using Dynamic Time Warping Algorithm. *International Journal of Innovative Technology and Exploring Engineering (IJITEE)*, 9(2), 2278-3075.
- Mohan, B. J., & Babu, R. N. (2014). Speech Recognition Using MFCC and DTW. *2014 International Conference on Advances in Electrical Engineering (ICAEE)*, 1366, s. 1-4. Vellore.
- Mubarak, R., Alsboui, T., Alshaikh, O., Inuwa-Dutse, I., Khan, S., & Parkinson, S. (2023). A Survey on the Detection and Impacts of Deepfakes in Visual, Audio, and Textual Formats. *IEEE Access*, 11, 144497-144529.
- Pan, X., Zhang, X., & Lyu, S. (2012). Detecting Splicing in Digital Audios Using Local Noise Level Estimation. *2012 IEEE International Conference on Acoustics*, (s. 1841-1844). Kyoto.
- Qamhan, M. A., Altaheri, H., Meftah, A. H., Muhammad, G., & Alotaibi, Y. A. (2021). Digital Audio Forensics: Microphone and Environment Classification Using Deep Learning. *IEEE Access*, 9, 62719-62733.
- Qiao, M., Sung, A. H., & Liu, Q. (2010). Revealing Real Quality of Double Compressed MP3 Audio. *Proceedings of the 18th ACM international conference on Multimedia*, (s. 1011 - 1014). New York.
- Sanchez, J., Saratxaga, I., Hernáez, I., & Nav, E. (2015). Toward a Universal Synthetic Speech Spoofing. *IEEE Transactions on Information Forensics and Security*, 10(4), 810-820.
- Thakur, A. S., & Sahayam, N. (2013). Speech Recognition Using Euclidean Distance. *International Journal of Emerging Technology and Advanced Engineering*, 3(3), 587-590.
- Todisco, M., Delgado, H., & Evans, N. (2017). Constant Q cepstral coefficients: A Spoofing Countermeasure For Automatic Speaker Verification. *Computer Speech & Language*, 45, 516-535.
- Ulutaş, G., Tahaoğlu, G., & Üstübioğlu, B. (2022). Forge Audio Detection Using Keypoint Features on Mel Spectrograms. *2022 45th International Conference on Telecommunications and Signal Processing (TSP)*, 413-416.
- Üstübioğlu, A., Üstübioğlu, B., & Ulutaş, G. (2023). Mel spectrogram-based audio forgery detection using CNN. *Signal, Image and Video Processing*, 17, 2211-2219.
- Üstübioğlu, B. (2024). An Attack-Independent Audio Forgery Detection Technique Based on Cochleagram Images of Segments With Dynamic Threshold. *IEEE Access*, 12, 82660-82675.

- Üstübioğlu, B., & Üstübioğlu, A. (2024). Audio Forgery Detection Method Based On Visual Word. *Niğde Ömer Halisdemir University Journal of Engineering Sciences*, 13(1), 350-358.
- Üstübioğlu, B., Küçükuşurlu, B., & Ulutaş, G. (2022). Robust Copy-Move Detection in Digital Audio Forensics Based on Pitch and Modified Discrete Cosine Transform. *Multimedia Tools and Applications*, 81(19), 27149–27185.
- Üstübioğlu, B., Tahaoğlu, G., Ulutaş, G., Üstübioğlu, A., & Kılıç, M. (2024). Audio Forgery Detection And Localization With Super-Resolution Spectrogram and Keypoint-Based Clustering Approach. *The Journal of Supercomputing*, 80, 486–518.
- Üstübioğlu, B., Tahaoğlu, G., & Ulutaş, G. (2023). Detection of audio copy-move-forgery with novel feature matching on. *Expert Systems With Applications*(213), 1-17.
- Wang, F., Li, C., & Tian, L. (2017). An Algorithm of Detecting Audio Copy-Move Forgery Based on DCT and SVD. *IEEE 17th International Conference on Communication Technology (ICCT)*, (s. 1652-1657). Chengdu.
- Xiao, J.-n., Jia, Y.-z., Fu, E.-d., Huang, Z., Li, Y., & Shi, S.-p. (2014). Audio Authenticity: Duplicated Audio Segment Detection In Waveform Audio File. *Journal of Shanghai Jiaotong University (Science)*, 19, 392-397.
- Xie, Z., Lu, W., Liu, X., Xue, Y., & Yeung, Y. (2018, Kasım). Copy-Move Detection Of Digital Audio Based On Multi-Feature Decision. *Journal of Information Security and Applications*, 43, 37-46.
- Yan, Q., Rui, Y., & Huang, J. (2015). Copy-Move Detection of Audio Recording With Pitch Similarity. *IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, (s. 1782-1786). South Brisbane.
- Yan, Q., Yang, R., & Huang, J. (2019, Eylül). Robust Copy–Move Detection of Speech Recording Using Similarities of Pitch and Formant. *IEEE Transactions On Information Forensics And Security*, 14(9), 2331-2341.
- Yazıcı, S., Üstübioğlu, B., Kılıç, M., & Ulutaş, G. (2022). Block-Based Forgery Detection with Binary Gradient Model. *2022 15th International Conference on Information Security and Cryptography (ISCTURKEY)*, (s. 38-43). Ankara.
- Zahorian, S. A., & Hu, H. (2008). A Spectral/Temporal Method for Robust Fundamental Frequency Tracking. *The Journal of the Acoustical Society of America*, 123(6).
- Zakariah, M., Khan, M. K., & Malik, H. (2017). Digital Multimedia Audio Forensics: Past, Present and Future. *Multimed Tools App*, 77, 1009–1040.
- Zhang, Y., Dai, S., Song, W., Zhang, L., & Li, D. (2019). Exposing Speech Resampling Manipulation by Local Texture Analysis on Spectrogram Images. 9(1), 1-16.
- Zhao, H., & Malik, H. (2012). Audio Forensics Using Acoustic Environment Traces. *2012 IEEE Statistical Signal Processing Workshop* (s. 373-376). Ann Arbor: IEEE.

## ÖZGEÇMİŞ

Ali HATİPOĞLU, İlkokulu 100. Yıl İlköğretim Okulu'nda, ortaokulu Prof. Dr. İhsan KOZ İlköğretim Okulu'nda tamamladıktan sonra Trabzon Anadolu Meslek Lisesi Bilgisayar Bölümü'nden 1998 yılında birincilikle mezun oldu. Lisans eğitimini Gazi Üniversitesi Bilgisayar Eğitimi Bölümü'nden 2003 yılında mezun olarak tamamladı. 2004 yılında Milli Eğitim Bakanlığı Araklı Cumhuriyet İlköğretim Okulu'nda bilişim teknolojileri öğretmeni olarak göreve başladı. 2022 yılında Karadeniz Teknik Üniversitesi, Fen Bilimleri Enstitüsü, Bilgisayar Mühendisliği Anabilim dalında tezli yüksek lisans programına başladı. Halen Trabzon Mesleki ve Teknik Anadolu Lisesi'nde Bilişim Teknolojileri Alanı'nda teknik öğretmen ve laboratuvar şefi olarak görev yapmaktadır.

Hatipoğlu, A., ÜSTÜBİOĞLU, B., Üstübioğlu, A., & ULUTAŞ, G., (2024). Detecting Audio Forgery with Rasta -PLP. *32nd IEEE Conference on Signal Processing and Communications Applications Conference (SIU)*, (s. 1-4). Mersin.