**T.C.**
**SİİRT ÜNİVERSİTESİ**
**FEN BİLİMLERİ ENSTİTÜSÜ**

# SES SİNYALLERİNDEN YAŞ GRUBU VE CİNSİYET BİLGİSİNİN TAHMİN EDİLMESİ

**YÜKSEK LİSANS**

**Abdulhalık OĞUZ**
**(153111010)**

**Elektrik-Elektronik Mühendisliği Anabilim Dalı**

**Tez Danışmanı: Dr. Öğr. Üyesi Yılmaz KAYA**

**Haziran-2018**
**SİİRT**

## TEZ KABUL VE ONAYI

Abdulhalık Oğuz tarafından hazırlanan "Ses Sinyallerinden Yaş Grubu ve Cinsiyet Bilgisinin Tahmin Edilmesi" adlı tez çalışması 03/04/18 tarihinde aşağıdaki jüri tarafından oybirliği/oyçokluğu ile Siirt Üniversitesi Fen Bilimleri Enstitüsü Elektrik Elektronik Mühendisliği Anabilim Dalı'nda YÜKSEK LİSANS olarak kabul edilmiştir.

| Jüri Üyeleri | İmza |
|---|---|
| **Başkan** Doç.Dr. Necmettin Sezgin | ..... |
| **Danışman** Dr. Öğr. Üyesi Yılmaz KAYA | ..... |
| **Üye** Dr. Öğr. Üyesi Melih KINCAN | ..... |
| **Üye** | ..... |
| **Üye** | ..... |

Yukarıdaki sonucu onaylarım.

Doç. Dr. Fevzi HANSU
Fen Bilimleri Enstitüsü Müdürü

## ÖN SÖZ

Bu çalışma boyunca değerli fikirlerini ve desteklerini benden eksik etmeyen, araştırmalarımın her aşamasında öneri ve yardımlarını esirgemeyen danışman hocam Sayın Dr. Öğr. Üyesi Yılmaz Kaya'ya ve her daim yanımda olan ve beni bu günlere getiren aileme, sevgi ve desteklerini hiç esirgemeyen eşime ve oğluma en derin duygularla teşekkürlerimi sunarım.

Abdulhalık OĞUZ
SİİRT-2018

# İÇİNDEKİLER

**Sayfa**

## TABLOLAR LİSTESİ

**Sayfa**

# ŞEKİLLER LİSTESİ

**Sayfa**

## KISALTMALAR VE SİMGELER LİSTESİ

| Kısaltma | Açıklama |
|---|---|
| **ANN** | : Artificial neural network |
| **DCT** | : Discrete cosine transform |
| **DFT** | : Discrete fourier transform |
| **DT** | : Decision tree |
| **FFT** | : Fast fourier transform |
| **GMM** | : Gaussian mixture model |
| **LDA** | : Linear discriminant analysis |
| **LPC** | : Linear predictive coding |
| **LPCC** | : Linear predictive cepstrum coefficients |
| **LS-SVR** | : Least squares support vector regression |
| **MFCC** | : Mel-frequency cepstrum coefficients |
| **MF&LP** | : MFCC ve LPCC karışım modeli |
| **MLP** | : Multilayer perceptron |
| **NB** | : Naive bayes |
| **PLP** | : Perceptual linear prediction |
| **RASTA** | : Relative spectral transform |
| **SDC** | : Shifted delta cepstral |
| **SVM** | : Support vector machine |
| **UBM** | : Universal background model |
| **VQ** | : Vector quantization |
| **WEKA** | : Waikato environment for knowledge analysis |
| **WPPCA** | : Weighted-pairwise principal components analysis |

# ÖZET

## YÜKSEK LİSANS TEZİ

## SES SİNYALLERİNDEN YAŞ GRUBU VE CİNSİYET BİLGİSİNİN TAHMİN EDİLMESİ

**Abdulhalık OĞUZ**

**Siirt Üniversitesi Fen Bilimleri Enstitüsü**
**Elektrik-Elektronik Mühendisliği Anabilim Dalı**

**Danışman : Dr. Öğr. Üyesi Yılmaz KAYA**

**2018, 56 Sayfa**

Teknolojinin hızla gelişimi, büyük veri teknolojilerinin artışı ve veri depolama ve işleme yöntemleri ile daha fazla meşguliyet; konuşma tanıma sistemlerinin önemini ileri ölçüde artırmıştır. Konuşmacının cinsiyetini ve yaş aralığını belirleyebilmek ise konuşma tabanlı uygulamalarda büyük önem arz etmektedir.

Uygulama alanı olarak çocuk seslerini ayırmaya yönelik çalışmalar; çocuklarda ortaya çıkan disleksi gibi bazı konuşma bozukluklarının tanımlama aşamasında veya çocuklara yönelik geliştirici interaktif oyun programlarında önem kazanmaktadır. Ayrıca yetişkin seslerini ayırmaya yönelik çalışmalarda ise insan kaçırma, tehdit telefonları, yanlış ihbarlar gibi kriminal durumlarda konuşmacının karakteristik özelliklerini daha iyi tanımlamayacak verilere ulaşabilmesi, polis istasyonlarına veya hastanelere gelen aramalarda yaşlı ve çocuk ses profillerine öncelik verilmesi veya müşterilerin daha iyi tanımlanabilmesi gibi durumlarda önem arz etmektedir.

Bu çalışmada, konuşmacılardan alınan ses örneklerinden çeşitli yöntemlerle elde edilen öz niteliklerin kullanılması ile kişilerin cinsiyeti ve yaş grubu tahmin edilmiştir. İlkokul, ortaokul, lise ve üniversite öğrenci gruplarının her birinden 8 erkek ve 8 kız öğrencinin sesi alınmıştır. Bu dört grup için toplamda 64 öğrenciden ses kaydı alınmıştır. Veri seti için bir kısmı Türkçe ‘de birleşim gücü yüksek kelimeler bir kısmı da sık kullanılan rastgele kelimelerden oluşan 32 adet Türkçe kelime seçilmiştir.

Alınan ses örneklerinden öznitelik çıkarımı için literatürde sıkça kullanılan Mel-frekansı kepstral katsayıları (Mel-Frequency Cepstral Coefficients, MFCC) ve Doğrusal öngörüm kepstrum katsayıları (Linear predictive cepstrum coefficients, LPCC) yöntemleri kullanılmıştır. Ayrıca iki öz nitelik vektörünün elemanları beraber alınarak MF&LP karışım modeli denenmiştir. Elde edilen öznitelik vektörleri k en yakın komşu (KNN), yapay sinir ağları (YSA) ve destek vektör makineleri (DVM) gibi makine öğrenmesi yöntemleri kullanılarak sınıflandırılmıştır. Sınıflandırma performansı; yaş grubu tahmini için %96 civarında iken cinsiyet tespiti için %94,6 civarında olmaktadır.

**Anahtar Kelimeler:** Cinsiyet tespiti, lpcc, mfcc, mf&lp, ses sinyali, yaş grubu tahmini

## ABSTRACT

## MS THESIS

## ESTIMATING AGE GROUP AND GENDER INFORMATION FROM SPEECH SIGNALS

**Abdulhalık OĞUZ**

**The Graduate School of Natural and Applied Science of Siirt University**
**The Degree of Master of Science**
**In Electrical-Electronics Engineering**

**Supervisior : Dr. Faculty Member, Yılmaz KAYA**

**2018, 56 Pages**

The rapid development of technology, the increase of large data technologies and the enhancement in occupation of data storage and processing methods has significantly increased the importance of speech recognition systems. The ability to determine the gender and age group of the speaker has great importance in speech-based applications.

Studies considering application areas as distinguishing children voices are gaining significance in the process of detecting speech disorders such as dyslexia that occurs in children, or in improving interactive game programs for children. In addition, the studies have been done to distinguish adult voices can be utilized to access data which can characterize the characteristics of the speaker in criminal situations such as human abduction, threatening telephones and false alarms. It may serve in giving priority to elderly and child voices at police stations or hospital calls, withal it may lead to a better user-profiling the age interval of the customers.

In this study, gender and age category of the speakers has been estimated based on the features extracted by various methods from the speech recording samples. The voices of 8 male and 8 female students were taken from each elementary school, secondary school and high school and university student groups. A total of 64 students' voice recordings were taken from these four groups. For the dataset, 32 Turkish words were chosen, some of which are high-word combinations in Turkish and some of which are frequently used random words.

Mel-Frequency Cepstral Coefficients (MFCC) and Linear Predictive Cepstrum Coefficients (LPCC) methods, which are frequently used in the literature, have been used to extract the features from the speech samples. In addition, the MF&LP mixture model was tested by taking the elements of the two feature vectors together. Obtained feature vectors are classified using machine learning methods such as K nearest neighbors (KNN), Artificial neural networks (ANN), and support vector machines (SVM). Classification performance; for age group estimation is about 96% while for gender detection is around 94.6%.



**Keywords:** Age group estimation, gender detection, lpcc, mfcc, mf&lp, speech signal

# 1. GİRİŞ

Teknolojinin hızla gelişimi, büyük veri teknolojilerinin artışı ve veri depolama ve işleme yöntemleri ile daha fazla meşguliyet; konuşma tanıma sistemlerinin önemini ileri ölçüde artırmıştır. Bu gelişmeler ve ihtiyaçtan ötürü, yaş ve cinsiyet sistemlerinin biyometrik olarak tanınma ihtiyacı ortaya çıkmıştır. Biyometri; tanımlama, erişim kontrolü ve/veya gözetim amaçları için insanların özelliklerini inceleyen bir bilgisayar bilimi dalıdır. Biyometrik tanımlayıcılar; yüz tanıma, DNA, retina ve parmak izi gibi fizyolojik özellikler veya ritim, ses veya yürüyüş gibi davranışsal özellikleri inceler.

Ses dalgası, ses üretim sistemini meydana getiren anatomik yapıların istemli hareketleri sonucunda oluşan akustik bir basınç dalgasıdır. Bu sistemin ana bölümleri ciğerler, nefes borusu, gırtlak, boğaz, ağız boşluğu ve burun boşluğudur. Teknik terim olarak boğaz ve ağız boşluğu ‘ses yolu’ olarak tanımlanır. Dolayısıyla ses yolu, gırtlak çıkışından başlayıp, dudaklarda sona erer (Selen, 1979).

Ses üretimi için kritik olan anatomik yapılar, ses telleri, damak, dil, dişler ve dudaklardır. Ses yolunu oluşturan bu anatomik yapılar farklı pozisyonlar alarak değişik sesleri oluştururlar. Ses üretimi bir akustik filtreleme işlemi olarak düşünülebilir. Akustik filtre, ses üretim yollarının özelliklerini gösterir (Karasartova, 2011).

İnsan sesini kullanarak ve insan bilgisini tanımlayarak insanın bilgisayar etkileşimine etkisini ve verimliliğini arttırmak çok önemlidir. Sadece bir kullanıcının konuştuğu bilgiyi değil, aynı zamanda nasıl konuşulduğunu ve anlamını da bilmek ihtiyacı giderek artmaktadır. Konuşma özelliğini etkileyen çok sayıda parametre vardır. Bunların en bilinenleri cinsiyet, yaş, sağlık, dil, lehçe, aksan, duygusal durum ve konuşmacının konuşurken ki dikkat durumudur (Shivaji ve Ramesh, 2015)

Bu tez, biyometrik tanımlayıcılarda davranışsal özelliklerden biri olan ses konusuna odaklanmaktadır. Konuşma uygulamaları için yaş ve cinsiyet algısı, birçok pratik uygulamaya sahiptir ve insan-bilgisayar etkileşimi veya konuşmacının karakteristik özelliklerini daha iyi tanımlayacak verilere ulaşılabilmek gibi birçok uygulamada yararlı olabilir.

Konuşmacıların; ses özelliklerine etkisinin araştırılması 1950'lerden (Mysak, 1959) beri sürdürülmekte iken, insan sesinden cinsiyeti ve yaşı tahmin etmeye çalışan bilgisayar tabanlı sistemler, 2000'lerin başlarından itibaren bazıları aşağıda belirtilen ciddi çalışmalara konu olmuştur.

Fonetik olarak izole edilmiş kelimeler ile yapılan araştırmalar ile yaş tanımada insan algısı ve makine algısı karşılaştırılmıştır (Schötz, 2001). İnsan duyu mekanizmasını taklit edebilen Mel frekansı kepstrum katsayıları (Mel-frequency cepstrum coefficients, MFCC) kullanılarak otomatik yaş ve cinsiyet tanıma sistemleri için öncü çalışmalar yapılmıştır (Minematsu, 2002). Özellikle yaşlı kullanıcıların özel ihtiyaçlarına cevap verebilmek adına sesin seğirme durumu ve konuşma durumu özellik olarak alınıp, yaşlı erkek, yaşlı olmayan erkek, yaşlı kadın ve yaşlı olmayan kadın kategorilerinde sınıflandırma alanları seçilmiştir (Müller ve ark., 2003).

Yetişkinler ve çocuklar arasında ayrım yapabilme yeteneğine sahip bir konuşma rehberlik sisteminin tasarımı için hem akustik hem de dilbilimsel özellikler konuşma özelliği vektörleri olarak kullanılmış ve Destek vektör makineleri (DVM) tabanlı bir sınıflandırma yöntemi geliştirilmiştir (Nisimura ve ark., 2004). Alman Telekom'u tarafından sağlanan 4000 adet ses verisi analizinde, MFCC öznitelik çıkarım metodu ile konuşmacıların yaşı; çocuk, genç (erkek-kadın), orta yaş (erkek-kadın) ve yaşlı olmak üzere 4 sınıfa ayrılmış ve analiz edilmiştir (Bocklet ve ark., 2008).

İnsan-robot etkileşimi için geliştirilmek istenen cinsiyet ve yaş grubu tanıma sisteminde konuşma özellikleri olarak MFCC ve Doğrusal öngörüm kepstrum katsayıları (Linear predictive cepstrum coefficients, LPCC) kullanılarak, yaş kategorileri çocuk ve yetişkin olan çalışmada başarılı bir şekilde cinsiyet ve yaş ayrımı yapılmıştır (Lee ve Kwak, 2012).

Bu tezin amacı, güvenilir tanıma sonuçları veren bir yaş ve cinsiyet tanıma sistemi oluşturmaktır. Konuşma sinyalinin güvenilir gösteriminin seçimi, öz nitelik vektörlerinin belirlenmesi amaçlanmıştır. Bu tanıma sistemi için kritik sorular; kullanılma nedeni, sistem uygulanırken kullanılacak sınıflandırıcı, gerçek zamanlı ve gürültülü koşullarda gösterilecek başarı performansıdır.

Uygulama alanı olarak çocuk seslerini ayırmaya yönelik çalışmalar; çocuklarda ortaya çıkan disleksi gibi bazı konuşma bozukluklarının tanımlama aşamasında veya çocuklara yönelik geliştirici interaktif oyun programlarında önem kazanmaktadır.

Yetişkin seslerini ayırmaya yönelik çalışmalarda ise insan kaçırma, tehdit telefonları, yanlış ihbarlar gibi kriminal durumlarda konuşmacının karakteristik özelliklerini daha iyi tanımlayacak verilere ulaşabilmesi, polis istasyonlarına veya hastanelere gelen aramalarda yaşlı ve çocuk ses profillerine öncelik verilmesi veya müşterilerin daha iyi tanımlanabilmesi adına çağrı merkezlerine yapılan aramalarda yaş aralığı belirlenebilmesi gibi durumlarda önem arz etmektedir. Ayrıca bu çalışma; çok

özel bir topluluk hakkında cinsiyet ve yaş ile ilgili elde edilmek istenen bilgiler için yardımcı olabilmektedir.

Otomatik yaş ve cinsiyet tanımanın kolay bir iş olmadığını gösteren birçok sebep bulunmaktadır. İlk nedenlerden biri, her bireyin konuşma karakteristiğinin benzersiz olmasıdır. Bir başka zorluk da gürültü faktörüdür. Gürültü konuşmacının sesinden başka bir şey olabilir.

Dilin her konuşmacısı farklıdır. Fark, konuşmacının ses anatomisinden gelmektedir. Bir erkek ve bir kadının konuşma özellikleri cinsiyet açısından çok benzer olabilmekte ve aynı zamanda farklı yaş gruplarından insanlar da yaş sınıflandırması açısından benzer konuşma özelliklerine sahip olabilmektedir. Bu nedenle, iyi tanıma sonuçları elde etmek için ve sistemin doğru olması için sistemin çok sayıda veri üzerinden eğitilmesi gerekir.

Konuşma uygulamaları için yaş ve cinsiyet tanıma uygulamaları; gelişime açık bir araştırma alanıdır. Bu sistemin uygulanmasında bazı girişimler olsa da, özellikle gürültülü ortamlarda iyi sonuçlar beklenmemesi gerekmektedir. Çoğu gerçek zamanlı uygulamalar gürültülü ortamlarda gerçekleştiğinden, sağlam bir yaş ve cinsiyet tanıma sistemine sahip olmak özellikle önemlidir. Birçok araştırmacı bu problemi çözmek için farklı konuşma özelliklerini ve farklı sınıflandırıcıları denemeye devam etmekte fakat mükemmel bir çözüm bulunamamaktadır. Gürültü, gerçek konuşmayı etkileyebilir ve bu yanlış bir sınıflandırmaya yol açabilir. Kalabalık insanların gürültüsü, sokak gürültüsü, araba gürültüsü, restoran gürültüsü veya benzeri gürültüler olabilir. Bu nedenle, güvenilir bir yaş ve cinsiyet tanıma sistemine sahip olmak için, ham konuşma verilerine bazı ön işleme tekniklerinin uygulanarak bu sesin ortadan kaldırılması gerekir (Erokyar, 2014).

Konuşmacının cinsiyetinin ve yaşının tahmin edilmesi olaylarının, birbiri üzerinde hassas bir eğilim göstermesinden dolayı; bu iki konu beraber çalışılmıştır. Yaş ve cinsiyet tanıma sistemi iki bölümden oluşmaktadır. İlk kısım ön işleme ve öz nitelik çıkarma olarak adlandırılmakta iken ikinci kısım ise sınıflandırma olarak adlandırılır. Birinci bölümde ses sinyali, dijital sinyal işleme teknikleri kullanılarak önceden işlenir ve daha sonra işlenen bu sinyalden MFCC ve LPCC gibi metotlar ile sınıflandırmada işe yarayacak özellikler çıkarılır. Daha sonra bu konuşma özellikleri, makine öğrenmesi yöntemlerinden birini eğitmek için kullanılır. Sınıflandırma bölümünde, sesten çıkarılan bu özellik vektörleri karşılaştırılarak en iyi eşleşme bulunur.

Bu tezde ilkokul, ortaokul, lise ve üniversite öğrenci gruplarının her birinden 8 erkek ve 8 kız öğrencinin sesi alınmıştır. İlkokul öğrencilerinin yaş aralığı 9 ve 10, ortaokul öğrencilerinin yaş aralığı 13, 14 ve 15, lise öğrencilerinin yaş aralığı 17 ve 18 ve üniversite öğrencilerinin yaş aralığı 21, 22 ve 23 yaş aralığındadır. Bu dört grup için toplamda 64 öğrenciden ses kaydı alınmıştır. Veri seti için bir kısmı Türkçe 'de birleşim gücü yüksek kelimeler bir kısmı da sık kullanılan rastgele kelimelerden oluşan 32 adet Türkçe kelime seçilmiştir (Keklik, 2011). Seçilen kelimeler şunlardır:

"Adam", "Arkadaşlık", "Barış", "Buğday", "Büyütmek", "Cumartesi", "Çocuk", "Doğmak", "Dürüst", "Düşünmek", "Fabrika", "Geliştirmek", "Hayat", "Hazırlanmak", "Hissetmek", "İnsan", "İnternet", "Kadın", "Mayıs", "Merhaba", "Millet", "Mutlaka", "Papatya", "Para", "Sabah", "Sağlamak", "Sanatçı", "Sevgi", "Sinema", "Tepe", "Umut", "Yarın".

Toplamda 2048 ses örneği toplanmıştır. Bu kayıtların 32'si erkek ve 32'si kız öğrencilere ait olup, her bir cinsiyet için 1024 adet ses örneği, her bir yaş grubu için de 512 adet ses örneği alınmıştır.

Alınan seslerden öz nitelik çıkarımı yapılmıştır. Bunun için sık kullanılan yöntemlerden MFCC ve LPCC öz nitelik çıkartım algoritmaları kullanılmıştır. Ayrıca iki öz nitelik vektörünün elemanları beraber alınarak MF&LP karışım modeli denenmiştir. Sınıflandırma yöntemleri olarak; k en yakın komşu (KNN), YSA ve DVM gibi farklı makine öğrenme yöntemleri kullanılmıştır.

Bu tezin geri kalanı şu şekilde düzenlenmiştir. Bölüm 2'de daha önce yapılan akademik çalışmalardan örnekler verilmiştir. Bölüm 3'te çalışılan materyal ve metot hakkında bilgiler verilmiştir. Bölüm 4'te, sistemin tasarımı belirtilmiş ve elde edilen çeşitli bulgulara ait sonuçlar verilmiştir. Son olarak, Bölüm 5' te tartışma kısmından bahsedilmiş ve gelecekteki olası araştırmalar hakkında bilgiler verilmiştir.

## 2. LİTERATÜR ARAŞTIRMASI

Makine öğrenme yöntemleriyle sesten cinsiyet ve yaş tahmin etme; günümüzde gittikçe yaygınlaşan kullanım alanlarına sahip olup, birçok araştırma grubu tarafından daha önceden farklı metotlarla araştırılıp, analiz edildi. Konuşmacıların konuşma özelliklerine etkisinin araştırılması 1950'lerden (Mysak, 1959) beri sürdürülmekte iken, insan sesinden cinsiyeti ve yaşı tahmin etmeye çalışan gerçek sistemler, 2000'lerin başlarından itibaren ciddi çalışmalara konu olmuştur.

Schötz (2001), farklı yaşlarda (21-30, 61-73) olan 4'ü erkek ve 4'ü kadın her bir konuşmacıdan aldığı 3 İsveç'çe fonetik olarak izole edilmiş uyarıcı toplam 24 kelime ile yaptığı araştırmada bu alanda ilk çalışmalardan birini gerçekleştirmiştir. Algı testinde, 38 dinleyiciye (19 erkek ve 19 kadın, yaş aralığı 14-60) daha önce alınan 24 kelime dinletilerek; ilk etapta alınan ses sahiplerinin yaşlarının tahmin edilmesi istenmiştir. Yaş grubu doğru tahmin edilen kişilerin sayısı %50 ile %92 arasında bulunmuştur.

Minematsu (2002), 12 öğrenci tarafından kulak yolu ile algılanan yaşın otomatik sınıflandırılması için bir teknik önermiştir. Belirlenen öğrenciler tarafından daha önce yaşlı olarak belirlenen 43 kişi ve eşit sayıda yaşlı olmayan olarak değerlendirilen konuşmacılar, Mel frekansı kepstrum katsayıları (Mel-frequency cepstrum coefficients, MFCC) ve bu katsayıların türevinin alınması ile elde edilen yeni öznitelik katsayıları (DMFCC) ve Gaussian karışım modeli (Gaussian mixture model, GMM) ve normal dağılım kullanılarak modellendi. Yaş gruplarını 20-30, 40-50, 60, 70 ve 70'ten büyük olmak üzere 5 farklı grupta tahmin eden bir yaklaşım önerildi. Sınıflandırma için Lineer diskriminant analizi (Linear discriminant analysis, LDA) ve Yapay sinir ağları (Artificial neural network, YSA) kullanıldı. LDA yöntemi ile yaşlılar; %90,9 oranında başarılı sınıflandırıldı. Daha sonra sınıflandırıcıyı iyileştirmek için hesaplanan konuşma hızı ve sesin şiddetindeki endişe (tedirginlik) özellik olarak eklenerek, başarı oranı %95,3'e çıkarıldı.

Müller ve ark. (2003), özellikle yaşlı kullanıcıların özel ihtiyaçlarına cevap verebilmek için, yaş ve cinsiyet tanıma sistemini inceledi. Sesin seğirme durumu ve konuşma durumu özellik olarak alınırken, yaşlı erkek, yaşlı olmayan erkek, yaşlı kadın ve yaşlı olmayan kadın olan sınıflandırma için dört kategori seçildi. Sınıflandırma işlemi için YSA, k en yakın komşu (k-nearest neighbors, KNN), Naif Bayes (Naive bayes, NB) ve destek vektör makineleri (Support vector machine, DVM) kullanıldı.

Cinsiyet sınıflandırmada başarı oranı ortalama %80,5 iken yaşlı olan ve yaşlı olmayan diye ayrım yapılan yaş gruplandırmasında ortalama başarı oranı %91,5 bulundu.

Nisimura ve ark. (2004), yetişkinler ve çocuklar arasında ayrım yapabilme yeteneğine sahip bir konuşma rehberlik sistemini araştırdı. Hem akustik hem de dilbilimsel özellikler konuşma özelliği vektörleri olarak kullanılmış, DVM tabanlı bir yöntem geliştirilmiş ve %92,4 başarı oranı elde edilmiştir.

Schötz (2007), daha önce yaptığı çalışmayı geliştirerek; konuşma oranı, ses basıncı, temel frekans (F0) ve yaş tanımada konuşma üretim mekanizmasının yaşlanması gibi akustik konuşma özelliklerini incelemiştir. Bununla birlikte, bu akustik özelliklerin karmaşık ilişkili olduğu ve çeşitli faktörlerden etkilendiği tespit edilmiştir. Kadın ve erkek yaşının farklılığının etkisi, konuşmacının fizyolojik sağlığının iyi ve kötü olması, gerçek yaş ile algılanan yaş durumu ve ayrıca farklı konuşma türleri arasında ayrımlar buna örnek olarak verilmiştir.

Metze ve ark. (2007) yaş ve cinsiyet tanıma için telefon uygulamalarında dört farklı yaklaşım incelendi. Aynı veri seti üzerinde insanlar ve onların duyma ve ayırt etme sistemleri arasında bir karşılaştırma yapıldı. Bu yaklaşımlar; paralel bir telefon tanıyıcı, çeşitli özellikleri birleştiren dinamik bir Bayes ağı, lineer tahmin analizi yaklaşımı ve son olarak MFCC yöntemine dayalı yaş ve cinsiyet ayrımı için GMM kullanılması idi. İlk yaklaşımları olan paralel telefon tanıyıcısı; cinsiyeti ayırt etme açısından insanlar kadar iyi olsa da; kısa seslerde doğruluk oranının yetersiz kaldığı görüldü.

Kim ve ark. (2007) tarafından olası bir ev robotu hizmeti için yaş ve cinsiyet sınıflandırma sistemi tasarlandı. MFCC öz nitelik vektörü olarak ve GMM, Çok katmanlı algılayıcı (Multilayer perceptron, MLP) ve YSA ise sınıflandırıcı olarak kullanıldı. Cinsiyet tanıma için sınıflandırıcı GMM seçildiğinde başarı oranı %94,9 ve yaş tanıma için başarı oranı doğruluğu%94,6 bulundu. Sesteki seğirmeyi de özellik vektörü olarak eklediklerinde ve sınıflandırma aracı olarak YSA kullandıklarında, cinsiyet tanıma oranı %81,09'a düşerken, yaş tanıma oranı %96,57'ye yükseldi.

Bocklet ve ark. (2008) yaptığı çalışmada ise Alman Telekom'u tarafından sağlanan 4000 adet ses verisi analizinde, MFCC öznitelik çıkarım metodu ile beraber GMM ve DVM sınıflandırma metotları kullanılarak konuşmacıların yaşı; çocuk, genç (erkek-kadın), orta yaş (erkek-kadın) ve yaşlı olmak üzere 4 sınıfa ayrıldı ve ortalama %49 oranında başarı sağlandı.

Sedaaghi (2009) yaş ve cinsiyet ayrıştırma için; iki farklı ses veri tabanı kullanmıştır. Bunlardan ilki Danimarka duygu ses veri tabanı DES ve diğeri İngiliz dili ses veri tabanı ELSDSR'dir. DVM, KNN ve GMM sınıflandırma metotları olarak kullanılırken; sırasıyla cinsiyet tanıma ve yaş tanıma için en yüksek başarı %95 ve %88 oranında bulundu.

GMM/DVM karışım modeli ile oluşturulan tanıma sistemi; Feld ve ark. (2010) tarafından araba-insan etkileşiminde kullanılmak üzere otomatik konuşmacı yaşı ve cinsiyeti tanıma şeklinde araştırıldı. 15 yaşından küçükler için ve 54 yaşından büyükler için cinsiyet ayırt etmeksizin, 15-24 ve 25-54 yaş aralığındakiler için ise cinsiyet sınıflandırması da dâhil yaş grupları araştırıldı.

Nguyen ve ark. (2010) tarafından yaş, cinsiyet ve aksan tanıma sistemi önerildi. Sınıflandırıcılar olarak Vektör kuantizasyonu (Vector quantization, VQ), GMM ve DVM kullanıldı. Sistem, 108 konuşmacı ve her konuşmacı için 200 ifade içeren Avustralya konuşma veri tabanında test edildi. Sınıflandırma doğruluk oranları %97,96 ile %98,68 arasında bulundu.

Kısa ve uzun süreli akustik ve ahenkli (ses tınısı, gür seslilik vb. ) özellikler kullanılarak eğitilen dört farklı modelin birleştirilmesiyle, yaş ve cinsiyet sınıflandırma sistemi Meinedo ve Trancoso (2010) tarafından 4 farklı ses veri tabanında incelenmiş ve tasarlanan karışım modeli %51,2 başarı göstermiştir.

Bocklet ve ark. (2010), daha önceki araştırmalarını geliştirerek; çoklu sistemlere ve onların birleşimlerine dayanan bir yaş ve cinsiyet tanıma sistemi geliştirdiler. Tamamı telefon konuşma kayıtlarından oluşan ve 772 konuşmacıdan yaklaşık 70.000 sözcük içeren bir ses veri tabanı kullanıldı. Yaş grubu olarak çocuk, genç, yetişkin ve yaşlı yaş gruplarına ayrılan seslerden; spektral özellikler, ahenk özellikleri ve gırtlak özellikleri çıkarıldı. En iyi yaş ve cinsiyet sınıflandırması sistemi olarak GMM ve Genel arka plan modeli (Universal background model, UBM) kullanılmıştır. Bu sistem için sınıflandırma doğruluğu%42,4 bulundu.

Dobry ve ark. (2011) tarafından yaşın regresyon veri tabanı yardımı ile belirlendiği çalışmaların yanı sıra ağırlıklı ikili temel bileşenler analizi (Weighted-pairwise principal components analysis, WPPCA) adında süper vektör olarak adlandırılan yeni bir konuşma boyutunu indirgeme yöntemi de önerildi. Bu yöntem iki farklı görevde test edildi. İlk görev yaş grubu sınıflandırması iken, ikinci görev kesin yaş tahminiydi. Bu boyut düşürme yöntemi ile kesin yaş tahmini aşamasında daha hızlı ve iyi sonuçlar alındı.

Bahari ve Hamme tarafından (2012) Belçika'da yapılan çalışmada ise 18-35, 36-45 ve 46-81 yaşları arası 555 konuşmacının sesi yine aynı aralıkta En küçük kareler destek vektör regresyonu (Least squares support vector regression, LS-SVR) yöntemi ile toplamda %48'e yakın bir başarı oranı ile tahmin edildi.

Lee ve Kwak (2012), insan-robot etkileşimi için cinsiyet ve yaş grubu tanıma sistemini geliştirdi. Konuşma özellikleri olarak MFCC ve Doğrusal öngörüm kepstrum katsayıları (Linear predictive cepstrum coefficients, LPCC) ve sınıflandırıcılar olarak DVM ve Karar ağacı (Decision tree, DT) kullandılar. Yaş kategorileri çocuk ve yetişkin olan çalışmada, 7 erkek ve 7 kadın sesi kullanıldı. Cinsiyet için %93,16 ve yaş grubu için %91,39 başarı oranları elde edildi.

Erokyar (2014) tarafından yapılan çalışmada İngiliz dili ses veri tabanı ELSDSR veri tabanı kullanılarak yaş ve cinsiyet tanıma sistemi tasarlanmıştır. Öz nitelik olarak MFCC ve gürültülü verilerde daha güvenilir sonuçlar veren ve MFCC kullanılarak elde edilen Shifted delta cepstral (SDC) alındı. Son olarak, daha iyi tanıma oranları elde etmek adına sesten elde edilen perde (pitch) ve MFCC birleşimi kullanıldı. Tasarlanan karışım sistemi, ELSDSR konuşma veri setinde %64,2 oranında genel tanıma başarısı göstermiştir.

Mirhassani ve ark. (2014) Malezya'da yaptığı çalışmada; 7-12 yaş arası 360 farklı çocuğun yaşını tahmin etmek için ses veri tabanını sesteki ünlü yapısına göre altı gruba ayıran “böl ve yönet” stratejisini önermiştir. Bu strateji; işlem verilerinin karmaşık dağılımında azalmaya ve dolayısı ile sınıflandırıcının öğrenme performansının artmasına neden olmuştur. Aynı zamanda ses içindeki ünlüler yaş tahmini için daha güvenilir bilgiler içermektedir. MFCC kullanılarak elde edilen öz nitelikler, ileri beslemeli YSA sınıflandırıcının çıktılarının bulanık veri füzyonu ile birleştirilmesi sonucu genel karara varmak için kullanılmış ve yaş tahmini doğruluğunda %53.33'e varan bir iyileşme ortaya çıkarmıştır.

Waller ve ark. (2015), konuşmacının konuşma hızının, doğal konuşma oranının daha hızlı veya daha yavaş konuşma oranları ile karşılaştırılması ile konuşmacı yaşının tahmininin nasıl etkilendiğini incelemeyi amaçlamıştır. Yapılan deneylerde; dinleyicilere, üç farklı konuşmacı yaş grubundan (genç, orta yaşlı ve yaşlı yetişkinler) sesli okuma örnekleri sunulmuştur. Konuşma hızının normalden daha hızlı olduğu seslerde konuşmacılar daha genç ve konuşma hızının normalden daha yavaş olduğu seslerde de daha yaşlı olarak tahmin edildi. Bu konuşma hızı etkisinin, daha genç (20-25 yaş) konuşmacılarla karşılaştırıldığında daha yaşlı (60-65 yaş) konuşmacılar için biraz

daha önemli olduğu görülmüştür. Bu da konuşma hızının, daha yaşlı konuşmacılar için; yaş aralığını anlama işareti olarak daha büyük önem kazandığı anlamına gelmektedir. Bu model, dinleyicilerin kendiliğinden (rastgele) konuşma yaşlarını tahmin ettiği başka bir deneyde ise daha belirgin şekilde ortaya çıkmıştır.

Shivaji ve Ramesh (2015) konuşmacı sesinden yaş ve duygu tahmini elde etmek için yaptığı çalışmada 12 yaşından küçükler çocuk, 13-30 yaş arası genç, 30-50 yaş arası yetişkin ve 50 yaşından büyükler yaşlı olarak alınmıştır. MFCC katsayılarının Temel bileşenler analizi (Principal component analysis, PCA) yöntemi ile boyutu indirgenerek ve GMM süper vektörleri de öz nitelik olarak alınarak DVM yardımı ile sınıflandırılmıştır.

Yücesoy ve Nabiyev (2016) tarafından yapılan çalışmada, açık ve kapalı mekânlarda cep telefonu ve karasal bağlantılarla yapılan telefon konuşmalarının veri tabanı olarak kullanıldığı sistemde konuşmacıların cinsiyetlerine göre üç (erkek, bayan, çocuk), yaşlarına göre dört (çocuk, genç yetişkin ve yaşlı) ve her iki özelliğine göre ise yedi sınıfa ayrılması amaçlanmıştır. Bu amaçla konuşmaların yalnızca sesli bölümlerinden elde edilen MFCC katsayıları ile oluşturulan GMM modelleri süper vektörlere dönüştürülerek DVM sınıflandırıcısına uygulanmıştır. Çalışmada konuşmaların ses içeren bölümlerinin belirlenmesinde sinyalin enerji özelliği kullanılırken GMM modellerinin eğitiminde ise geniş bir veri tabanı ile eğitilen UBM uyarlanması yaklaşımı tercih edilmiştir. Çalışmada ayrıca farklı sayıda bileşenle oluşturulan GMM modelleri farklı uzunluklu konuşmalarla test edilerek GMM bileşen sayısı ve konuşma süresinin yaş ve cinsiyet tespiti üzerindeki etkisi de araştırılmıştır. Yapılan çalışmada en yüksek başarı oranları cinsiyet kategorisinde %92,42, yaş kategorisinde %60,10 ve yaş & cinsiyet kategorisinde ise %60,02 olarak ölçülmüştür.

## 3. MATERYAL VE METOT

Ses sinyali analog olup, işitilebilir bir dalgadır. Ses kartının bir sesi kullanması için öncelikle sesin analog formdan dijital forma dönüştürülmesi gerekir. Analog sinyalin dijital forma dönüştürülmesi sinyalin kodlanarak bir saniye içerisinde birçok kere örneklenmesi ve ses dalgasının yüksekliğinin kaydedilmesi ile olur.

Kulaklarımızın duyma bant genişliği için maksimum oran 22 kHz sınırı civarında kabul edilir. Bunun iki katı olan ve CD'lerin örnekleme miktarı olarak kullanılan saniyede 44.100 örnek sayısı teoride idealdir. Daha yüksek bir örnekleme sayısı ses ile ilgili daha fazla verinin saklanması dolayısı ile daha iyi bir veriye sahip olmak demektir. Bu çalışmada ses sinyalleri düşük sayılabilecek bir örnekleme oranı olan 16 kHz ile örneklenmiştir. Bunun nedeni gerçek zamanlı konuşma ortamına yakın bir frekansta ve daha düşük kalitede en iyi sonuçları elde etmeye çalışmaktır.

Ses ile ilgili önemli bilgilerden bazıları da frekans ve genlik bilgileridir. Sesin 1 saniyede oluşturduğu titreşim sayısına frekans denir. Sesin frekansı arttıkça incelir. Düşük frekanslı sesler ise kalın seslerdir. Örneğin; kızların sesi erkeklere göre daha incedir. Bu kızların sesinin frekansının daha yüksek olduğu anlamına gelir. Ses genliği frekanstan farklıdır. Frekansta sesin tonu önemliyken genlik açısından sesin şiddetinin önemi vardır. Örneğin; trenin sesi otomobilin sesinden daha şiddetlidir. Bu da trenin sesinin genliğinin daha fazla olduğu anlamına gelir.

### 3.1. Materyal

Çalışma ile ilgili konuşmacılar ilkokul, ortaokul, lise ve üniversite öğrenci gruplarından seçilmiştir. Çalışmada Waveform audio file format (wav) formatına uygun bir ses kayıt cihazı ile konuşmacılardan işlenmemiş ses verileri toplanmıştır. Mp3 ve diğer dosya biçimlerinin aksine wav olarak kaydedilmiş ses örnekleri sadece sayısallaştırılmış sesler olup, sese herhangi bir sıkıştırma işlemi uygulanmamıştır. Ses kayıt işlemi; normal koşullarda ve konuşmacıya yakın mesafede 16 kHz örnekleme oranı ile yapılmıştır. Ses örnekleri için ilkokul, ortaokul, lise ve üniversite öğrenci gruplarının her birinden 8 erkek ve 8 kız öğrencinin sesi alınmıştır. Bu dört grup için toplamda 64 öğrenciden ses kaydı alınmıştır.

Veri seti için bir kısmı Türkçe ‘de birleşim gücü yüksek kelimeler bir kısmı da sık kullanılan rastgele kelimelerden oluşan 32 adet Türkçe kelime seçilmiştir (Keklik, 2011). Seçilen kelimeler şunlardır:

"Adam", "Arkadaşlık", "Barış", "Buğday", "Büyütmek", "Cumartesi", "Çocuk", "Doğmak", "Dürüst", "Düşünmek", "Fabrika", "Geliştirmek", "Hayat", "Hazırlanmak", "Hissetmek", "İnsan", "İnternet", "Kadın", "Mayıs", "Merhaba", "Millet", "Mutlaka", "Papatya", "Para", "Sabah", "Sağlamak", "Sanatçı", "Sevgi", "Sinema", "Tepe", "Umut", "Yarın".

Ses kayıtları alınan öğrencilerin dağılımı aşağıdaki tabloda (Tablo 3.1.) verilmiştir. İlkokul öğrencilerinin yaş aralığı 9 ve 10, ortaokul öğrencilerinin yaş aralığı 13, 14 ve 15, lise öğrencilerinin yaş aralığı 17 ve 18 ve üniversite öğrencilerinin yaş aralığı 21, 22 ve 23 yaş aralığındadır.

**Tablo 3. 1. Ses kayıtları alınan öğrencilerin dağılımı**

| | İlkokul | Ortaokul | Lise | Üniversite | Toplam |
|---|---|---|---|---|---|
| Erkek | 8 | 8 | 8 | 8 | 32 |
| Kız | 8 | 8 | 8 | 8 | 32 |
| Toplam | 16 | 16 | 16 | 16 | 64 |

Her bir yaş grubunda 8 erkek ve 8 kız öğrenci seslendirme yapmıştır. Her bir öğrencinin 32 kelimeyi seslendirmesi ile toplamda 2048 ses örneği elde edilmiştir. Bunların yarısı erkek öğrencilere kalan yarısı da kız öğrencilere aittir. Her bir yaş grubundan da 512 ses örneği elde edilmiştir. Seslendirilen kelimelerden en uzun süreli olanı 2,31 saniye, en kısa süreli olanı da 0,36 saniye sürmektedir.

## 3.2. Metot

### 3.2.1. Özellik çıkarım yöntemleri

Ses sinyalini parametrik olarak temsil etmek için kullanılabilecek çeşitli yöntemler vardır (Rabiner ve Juand, 1993). Bu yöntemler ses sinyaline ilişkin, kısa zaman enerjisi, sıfır geçiş hızları, seviye geçiş hızları ve diğer ilgili parametreleri içerebilir. Ses sinyalinin parametrik olarak temsilinde en önemli olanı Kısa zamanlı spektral analizdir.

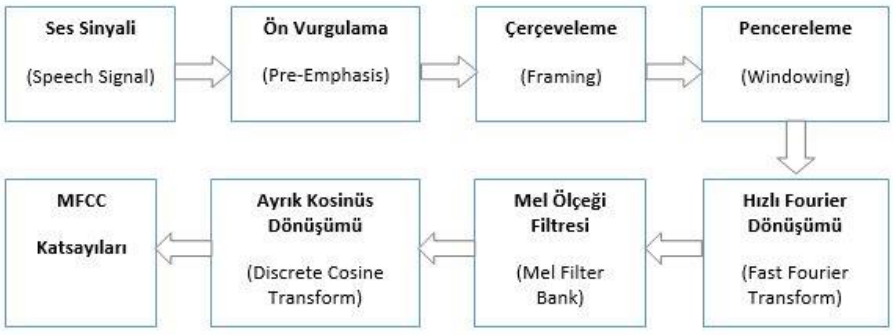
Ses sinyalinin parametrik olarak temsil edilmesi, yani ses sinyaline ilişkin özelliklerin çıkarımı için geliştirilmiş yöntemlerden en çok kullanılanları aşağıda belirtilmiştir:

- Mel Frekansı Kepstrum Katsayıları (MFCC)
- Doğrusal Öngörüm Kepstrum Katsayıları (LPCC)
- Algısal Doğrusal Öngörüm (PLP)
- Göreceli Spektrum Yöntemi (RASTA)

Bu çalışmada ise ses sinyallerinden anlamlı veriler elde etmeyi sağlayacak öznitelik çıkarımı işlemi için yukarıdaki algoritmalardan; Mel ölçeğine uygulanmış konuşmaya ait güç spektrumunu kullanan Mel frekansı kepstrum katsayıları (Mel-frequency cepstrum coefficients, MFCC) ve yine sıkça kullanılmakta olan Doğrusal öngörüm kepstrum katsayıları (Linear predictive cepstrum coefficients, LPCC) algoritmaları kullanılmış, diğer yöntemler ise kısaca açıklanmıştır (Kaushal ve Mistry, 2016).

### 3.2.1.1. Mel-frequency cepstrum coefficients

Davis ve Mermelstein tarafından 1980 yılında tanıtılan Mel frekansı kepstrum katsayıları (MFCC), insan duyu mekanizmasını taklit edebildiği için bir nevi ses sinyallerinin parmak izini çıkartabilen bir algoritmaya sahiptir. Aşağıda bir ses sinyalinden MFCC katsayılarının elde edilme aşaması gösterilmiştir (Chandra ve ark., 2014).




**Şekil 3. 1. MFCC katsayılarının elde edilmesi aşamasının blok diyagramı**

#### 3.2.1.1.1. Ön vurgulama

İlk aşamada ön vurgulama işlemi gerçekleşmektedir. İnsanların gırtlakta ses üretimi sırasında bastırdığı yüksek frekanslı kısmın telafi edilmesi adına ses sinyaline

filtre uygulanır ve yüksek frekanslı kısmın enerjisi elde edilmek istenen akustik bilgilere uygun hale getirilir (Picone, 1993). Yaygın olarak kullanılan ön vurgu fitresi (3.1) bağıntısıyla gerçekleştirilir.

$$Y[n] = x[n] - a * x[n-1], a \approx (0{,}95 - 0{,}97) \quad (3.1)$$

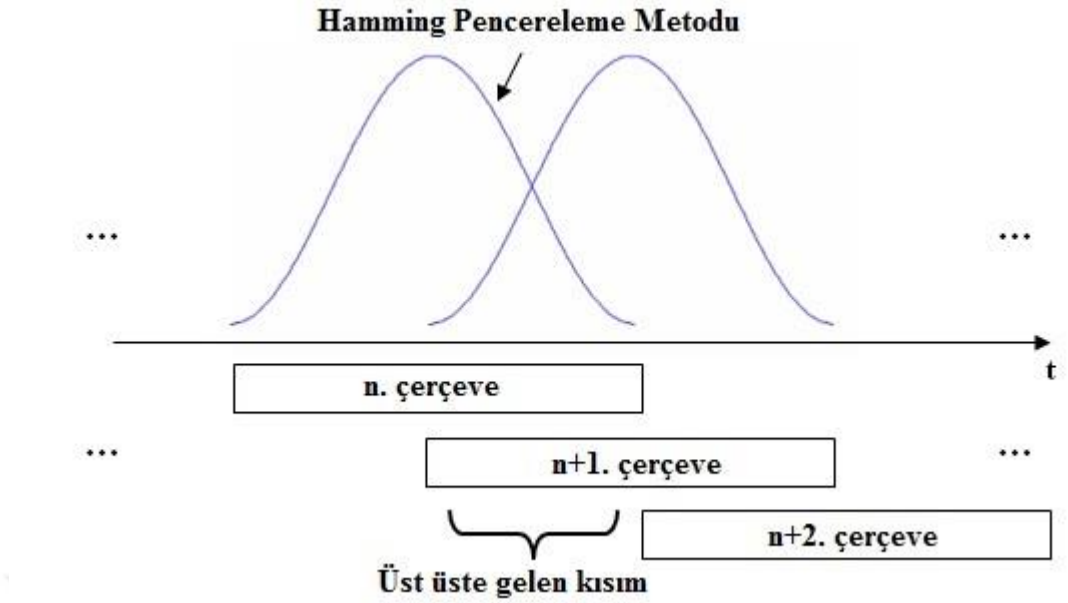
Bu çalışmada ön vurgulama katsayısı $a$ = 0,97 olarak seçilmiştir. Değişken "n"; ayrık zaman sinyalinin sadece "zaman" indeksidir. X ise konuşmadaki giriş sinyalidir.

#### 3.2.1.1.2. Çerçeveleme ve pencereleme

Çerçeveleme kısmında ses sinyalinin çok kısa zaman aralıklarında işlenmesi ve bir çerçeveden diğerine geçisin yumuşak olması sağlanmaktadır. Çoğu durumda belirlenen en etkili zaman aralığı 20-30 milisaniye arasındadır. Böylece her çerçeve kendinden önceki çerçevenin bir bölümünü içerisinde barındırır (Deller ve Hansen, 2000; Rabiner ve Juand, 1993). Bu çalışmada ise 25 milisaniye olarak alınmıştır.

Çerçeveleme işlemi sonunda her bir çerçevenin öncesi ve sonrasında oluşabilecek süreksizliği ve kaymaları önlemek adına pencereleme fonksiyonu uygulanır (bkz. Şekil 3.2). Konuşmacı tanıma sisteminde başarımı en yüksek pencereleme fonksiyonunu bulmak için Hamming, Hanning, Blackman, Gauss ve dikdörtgen pencereleme fonksiyonları çerçevelere uygulanmaktadır. Ve bu çalışmada da ses uygulamalarında güçlü bir şekilde kullanılan Hamming pencereleme yöntemi uygulanmış ve bu yöntem (3.2) bağıntısıyla gerçekleştirilmiştir. Ardışık pencereler arasındaki süre 10 milisaniye alınmıştır.

$$w(n) = 0.54 - 0.46 * \cos\left(\frac{2\pi n}{N-1}\right), 0 \leq n \leq N-1 \quad (3.2)$$




**Şekil 3. 2. Hamming pencere fonksiyonuna tabi tutulmuş çerçevelenmiş ses sinyali**

### 3.2.1.1.3. Hızlı fourier dönüşümü ve mel spektrumu

Hızlı fourier dönüşümü (Fast fourier transform, FFT) karışık sinyal yumaklarını ayrıştırır ve hangi frekansta ne şiddette bir titreşim olduğunu gösterir. FFT ile her bir çerçevenin zaman bölgesinden frekans bölgesine çevrimi gerçekleştirilir. Böylece her bir çerçevenin frekans tepkisinin büyüklüğü ölçülmüş olur. FFT, Ayrık fourier dönüşümü (Discrete fourier transform, DFT) uygulamak için hızlı bir algoritmadır. FFT, DFT hesaplanması için etkili ve ekonomik bir algoritmadır. N örnekli bir set için; FFT, aşağıdaki bağıntıyla tanımlanabilir.

$$X_n = \sum_{k=0}^{N-1} x_k e^{-2\pi kn/N}, \quad n = 0, 1, 2 \ldots, N-1 \tag{3.3}$$

Elde edilen frekans tepkileri Mel filter bank adlı filtre yardımı ile mel ölçeğine dönüştürülür. Dönüşüm işleminde bant genişliği mel ölçeğine göre lineer olarak değişen üçgen benzeri filtreler kullanılır. Genellikle filtre katsayısı olarak 20 ile 30 arasında bir değer seçilir. Bu tezde üçgen filtre katsayısı 22 adet seçilmiştir. Farklı ses ve ses uzunlukları için bu sayı değişebilmektedir. Mel ölçeği, insan kulağının algısal özelliğini taklit edecek şekilde tasarlanmış olup, 1 kHz'e kadar doğrusal, 1 kHz'den 8 kHz'e kadar ise logaritmik özellik gösterir (Chandra ve ark., 2014). Bu ölçek (3.4) bağıntısıyla gösterilir.

$$Mel(f) = 2595 * \log(1 + f/700) \tag{3.4}$$

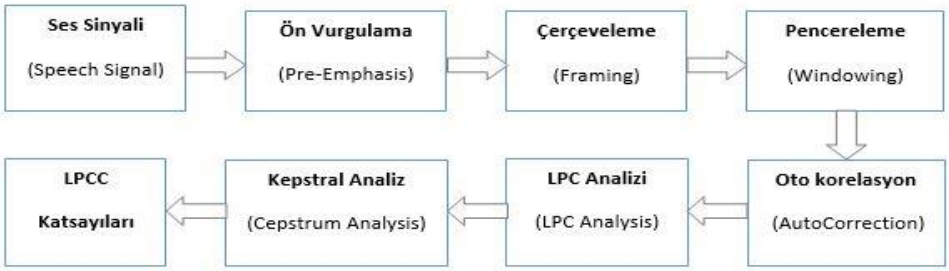
#### 3.2.1.1.4. Ayrık kosinüs dönüşümü ve kepstrum katsayılarının elde edilmesi

Hızlı fourier dönüşümü yolu ile frekans bölgesine çevrilen sinyalleri Ayrık Kosinüs Dönüşümü (Discrete cosine transform, DCT) yardımı ile tekrar zaman benzeri bölgeye çevrilerek sinyalin logaritmik spektrumu olan kepstrumlar elde edilir. Böylelikle sinyalin tahmini spektrumunun logaritmasının ters Fourier dönüşümünün alınması ile kepstrum elde edilir. Böylece elde edilen katsayılar mel-frekanslı kepstrum katsayısı yani MFCC olarak adlandırılır.

### 3.2.1.2. Linear predictive cepstrum coefficients

Doğrusal öngörüm kepstrum katsayıları algoritması (Linear predictive cepstrum coefficients, LPCC); temelde Doğrusal ön kestirim kodlama (Linear predictive coding, LPC) metodu ile elde edilen katsayıların Fourier dönüşümü ile kepstral katsayılara dönüştürülmesi prensibine dayanmaktadır. Dolayısıyla bu yöntemde ilk olarak LPC katsayılarının elde edilmesi ve ardından diğer matematiksel işlemlerin yapılması gerekmektedir.

LPCC algoritmasının aşamaları (Şekil 3.3) ön vurgulama, çerçeveleme, pencereleme, oto korelasyon analizi, LPC analizi ve kepstral analiz hesaplamasıdır (Këpuska ve Elharati, 2015).




**Şekil 3. 3. LPCC katsayılarının elde edilmesi aşamasının blok diyagramı**

#### 3.2.1.2.1. Ön vurgulama

Ön vurgulama aşaması MFCC hesaplamasında kullanılan prensiplerle aynıdır (Bkz. Bölüm 3.2.1.1.1).

#### 3.2.1.2.2. Çerçeveleme ve pencereleme

Çerçeveleme ve pencereleme aşamaları MFCC hesaplamasında kullanılan prensiplerle aynıdır (Bkz. Bölüm 3.2.1.1.2).

#### 3.2.1.2.3. Oto korelasyon

Pencerelenmiş sinyalin her bir çerçevesine oto korelasyon analizi uygulanır. Oto korelasyon;

$r_1(m) = \sum_{n=0}^{N-1-m} \widehat{x_1}(n)..\widehat{x_1}(n+m)\ ,\ \ m = 0, 1, 2\ .\ .\ ., p$ (3.5)

şeklindeki bağıntıyla tanımlanır (3.5). Oto korelasyon analizinin bir artısı da, sıfırıncı oto korelasyonun ilgili çerçevenin enerjisini tanımlamasıdır. Bir çerçevenin enerjisi ses tanıma sistemleri için önemli bir parametredir.

#### 3.2.1.2.4. LPC analizi

Bu bölümde LPC parametre kümesi; her bir çerçeveye ait p+1 oto korelasyondan elde edilir. Oto korelasyon analizinden LPC analizine geçişte Durbin metodu vb. yöntemler kullanılabilir. Durbin algoritmasındaki amaç doğrusal öngörü filtresi katsayıları ile ilgili öngörü hata değişim oranının yinelemeli olarak bulunmasıdır. LPC analizine geçişte oto korelasyon analizinin yerine kovaryans analizi de kullanılabilir. Sonuç olarak elde edilen LPC parametreleri, $a_m$ LPC katsayılarından oluşmaktadır (Rabiner ve Juand, 1993).

#### 3.2.1.2.5. Kepstral analiz

Bu aşamada Fourier dönüşümü ile LPC parametre kümesinden LPCC'ye dönüşüm gerçekleştirilir. Böylece elde edilen kepstral katsayıların gürültü vb. gibi çeşitli etkenlere duyarlılığını minimize edilmiş olur. Bu aşamada kullanılan bağıntı (3.6) aşağıda verilmiştir.

$c_1 = a_1$

$c_n = a_n + \sum_{k=1}^{n-1} \left(1 - \frac{k}{n}\right) a_k . c_{n-k}\ \ ,\ \ 1 < n \leq p$ (3.6)

Böylece elde edilen katsayılar Doğrusal öngörüm kepstrum katsayıları algoritması (Linear predictive cepstrum coefficients, LPCC); olarak adlandırılır. Elde edilmek istenen her bir LPCC katsayısı için öncesinde 2 katı LPC katsayısı elde

edilmiştir. Bu oran sabit olmayıp, eldeki kümeye göre değiştirilebilir. Yani 40 LPCC katsayısı için; öncesinde 80 LPC katsayısı elde edilmiş ve bu 80 katsayıdan da 40 LPCC katsayı için dönüşüm yapılmıştır.

#### 3.2.1.3. Perceptual linear prediction

Algısal doğrusal öngörüm (Perceptual linear prediction, PLP); LPC'nin bir türevi olup, ilk olarak (Hermansky, 1990) tarafından ortaya atılmıştır. PLP'de de, LPC'de olduğu gibi bir dizi parametre hesaplanmaktadır. PLP parametreleri, DFT (Ayrık Fourier Dönüşümü) ve Doğrusal öngörüm tekniklerinin birleştirilmesi ile hesaplanır. Bu yöntemdeki temel fikir, insan kulağının işitme aralığında, fiziki özelliklerinden türetilen bazı karakteristikleri dikkate almasıdır. 800 Hz değerinden daha düşük frekanslarda duyma miktarı frekansla birlikte düşer. İnsan kulağı daha çok duyma frekans aralığının ortasındaki frekanslara duyarlıdır. Bu sorunu çözmek için birçok çalışma yapılmıştır. Bu çalışmalardan biri de bulunan Doğrusal öngörüm katsayılarının mel skalasına uyarlanması olmuştur. Bir başka yaklaşım da Doğrusal öngörü tekniği uygulamadan önce sesli ifadenin güç spektrumunun alınmasıdır.

#### 3.2.1.4. Relative spectral transform

Göreceli spektrum yöntemi (Relative Spectral Transform, RASTA); sesli ifade içindeki çevresel etkilerin, yani gürültünün, modellenmesine dayalı bir sesli ifade modelleme yöntemi olarak kullanılır. Yukarıda belirtilen PLP yöntemi üzerine gürültü modelleme tekniği eklenerek elde edilen bir yöntemdir. RASTA yönteminin dayandığı temel; insan kulağının sesli ifadeyi algılamasının daha önceki seslerden önemli derecede etkilendiğidir. Yani algılama şu andaki ses ile önceki ses arasındaki spektral farka bağlıdır. Bu durumda insan kulağı yavaş değişen seslere daha az duyarlıdır denebilir. Yapılan sesli ifade çözümlemesinin yavaş değişen seslere daha az duyarlı yapılması insan kulağının bu özelliğinin de modellenmesini sağlar. Bunu yapmak için daha önce belirtilen PLP yönteminde kullanılan filtreleme yönteminde değişiklikler yapılmıştır. Kullanılan filtreler spektral sıfır değeri keskinleştirilmiş, yani sıfır frekans düzeyine aniden inen filtrelerle değiştirilmiştir. Böylece frekanslardaki yavaş değişimlerin etkisi azaltılmıştır (Hermansky ve Morgan, 1994).

### 3.2.2. Sınıflandırma Metotları

Konuşmacı özniteliklerinin ses sinyallerinden çıkarılmasından sonra sınıflandırma aşamasına geçilir. Sınıflandırmada amaç konuşma sinyalinden elde edilen öznitelik vektör uzayını belirli sayıda alt bölgeye ayırmaktır. Günümüzde ses işlemede kullanılan değişik sınıflandırma teknikleri vardır. Bu tezde; k en yakın komşu (k-nearest neighbors, KNN), destek vektör makineleri (DVM)(Support vector machine, SVM) ve Yapay sinir ağları (YSA)(Artificial neural network, ANN) gibi farklı makine öğrenme yöntemleri sınıflandırma için kullanılmış ve karşılaştırılmıştır. Sınıflandırma işlemi için WEKA yazılımı kullanılmıştır ( Witten ve Frank, 2005).

#### 3.2.2.1. K en yakın komşu

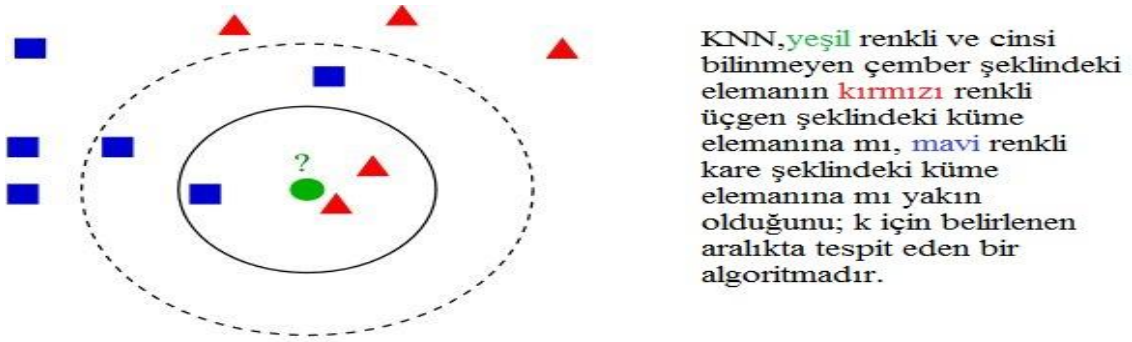
K en yakın komşu (K-nearest neighbors, KNN), sınıflama ve regresyonda kullanılan ve parametrik olmayan bir algoritmadır (Altman, 1992). Başarı oranı yüksek ama tembel bir algoritmadır. Buna göre sınıflandırma sırasında çıkarılan özelliklerden sınıflandırılmak istenen yeni bireyin daha önceki bireylerden k tanesine yakınlığına bakılmasıdır. Örneğin k = 3 için yeni bir eleman sınıflandırılmak istendiğinde; eski pozisyonda sınıflandırılmış elemanlardan en yakın 3 tanesi alınır. Bu elemanlar hangi sınıfa dâhil ise, yeni eleman da o sınıfa dâhil edilir. Elemanlar arasındaki mesafe hesabında genelde Öklid, Manhattan veya Minkowski mesafe hesaplama formüllerinden biri kullanılır. Sırasıyla mesafe hesaplaması bağıntıları aşağıda belirtilmiştir.

$$\text{Öklid} \gg \sqrt{\sum_{i=1}^{k}(x_i - y_i)^2} \tag{3.7}$$

$$Manhattan \gg \sum_{i=1}^{k}|x_i - y_i| \tag{3.8}$$

$$Minkowski \gg (\sum_{i=1}^{k}(|x_i - y_i|)^q)^{\frac{1}{q}} \tag{3.9}$$

KNN yöntemine göre aşağıdaki şekildeki (Şekil 3.4.) gibi sınıflandırılması istenen yeni bir üyenin geldiğini düşünülürse:




**Şekil 3. 4. KNN algoritması gereği karşılaştırılacak nesnelerin temsili**

Yeni gelen üyenin (yeşil renkli küçük çember) hangi gruba üye olduğunu belirleme adına $k = 3$ alınırsa; yeni üyenin en yakınında bulunan 3 cisimden 2 sinin kırmızı renkli üye, diğerinin mavi renkli üye olmasından ve çoğunluğun kırmızı renkli üye olmasından dolayı, yeni gelen üye kırmızı olarak sınıflandırılır. Ama eğer yeni gelen üyenin hangi gruba üye olduğunu belirleme adına $k = 5$ alınırsa; yeni üyenin en yakınında bulunan 5 cisimden 3 ünün mavi renkli üye, diğer 2 sinin kırmızı renkli üye olmasından ve çoğunluğun mavi renkli üye olmasından dolayı, yeni gelen üye mavi olarak sınıflandırılır. Görüldüğü üzere $k$ değeri için belirlenen değere bağlı olarak, sonuç değişebilmektedir (Jivani ve ark., 2016). Bu tezde elemanlar arasındaki yakınlık mesafe hesabında Öklid uzaklığı kullanılmıştır.

### 3.2.2.2. Destek vektör makineleri

Destek Vektör Makineleri (DVM), istatistiksel öğrenme teoremine dayanan örüntü sınıflandırma yöntemidir (Vapnik, 1995). DVM, herhangi bir sınıflandırma ya da regresyon problemini, bir karesel programlama problemine dönüştürerek yerel çözümlere takılmadan çözebilmektedir. DVM'de amaç, veri kümesini mümkün olduğu kadar iyi sınıflandıran en uygun ayırıcı düzlemin bulunmasıdır. Yani iki sınıf arasındaki uzaklığın maksimum (en büyük) olduğu durumun bulunması amaçlanmaktadır. Bu amaç, doğrusal olmayan örnek uzayının doğrusal olarak ayrılabileceği yüksek boyuta aktarıldıktan sonra, farklı örnekler arasındaki en büyük sınırın bulunmasıyla gerçekleştirilir.

DVM'ler, doğrusal ve doğrusal olmayan DVM olmak üzere ikiye ayrılır. Doğrusal DVM'nin yapısındaki çekirdek fonksiyonu sadece giriş uzayının bir ürünüdür ve doğrusal olmayan çekirdek fonksiyonlarının kullanılmasına gerek yoktur. Doğrusal

olmayan DVM'de ise, probleme uygun doğrusal olmayan bir çekirdek fonksiyonunun seçilmesine ihtiyaç vardır.

Doğrusal DVM'ler, doğrusal olarak ayrılabilen ve doğrusal olarak ayrılamayan DVM'ler olarak incelenmektedir. Doğrusal olarak ayrılabilme durumunda, eğitim için kullanılacak $N$ elemandan oluşan bir veri kümesi $\{x_i, y_i\}, i = 1,2,..,N,$ olarak tanımlandığında, $y_i \in \{-1, +1\}$ etiket değerlerini ve $d$ boyutlu $x_i \in \mathcal{R}^d$ özellik vektörünü temsil etmek üzere, bu iki sınıfı temsil eden örnekler doğrudan bir ayırıcı düzlem ile ayrılabilmelidir. DVM'nin amacı, verilen veri kümesini, tanımlanan etiketlere göre bir alt düzlemle ayırıp, aynı sınıfa ait bütün veri noktalarını alt düzlemin aynı tarafında bırakmaktır. Ayırıcı düzlem üzerindeki herhangi bir $x$ noktası, $w$ ayırıcı düzlemin normali ve $|b|/||w||$ ayırıcı düzlemin orijine dik uzaklığı olduğu kabul edildiğinde, DVM algoritması;

$$f(x) = w^T.x + b = 0 \tag{3.10}$$

denklemi ile tanımlanan en uygun ayırıcı düzlemi bulmaya çalışır. Bunun için eğitim kümesinin aşağıdaki bağıntıyı sağlaması gerekmektedir:

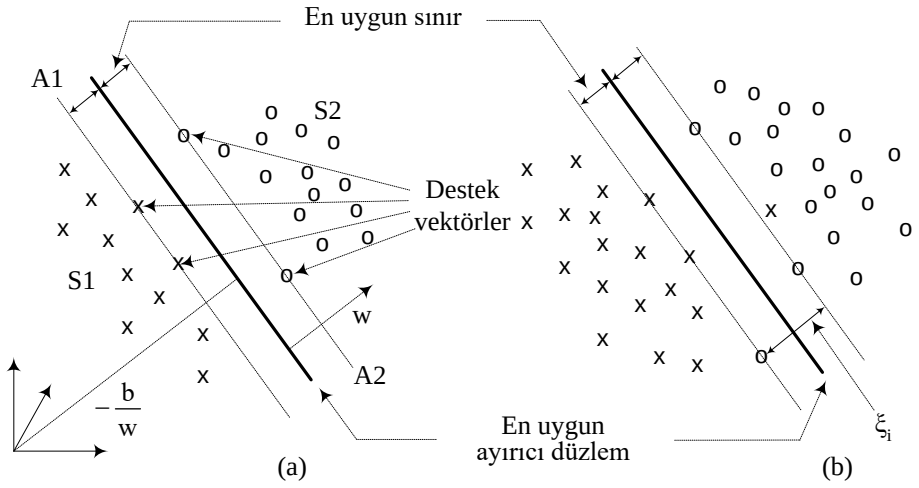
$$y_i = +1\ için,\ w^T.x_i + b \geq +1 \tag{3.11}$$

$$y_i = -1\ için,\ w^T.x_i + b \leq -1 \tag{3.12}$$

Bu eşitsizlikler bir arada ifade edildiğinde;

$$y_i(w^T.x_i + b) - 1 \geq 0 \qquad i = 0,1,\ldots N \tag{3.13}$$

elde edilir. Burada en uygun ayırıcı düzlemi bulmak için $w\ ve\ b$ değerleri hesaplanmalıdır. Doğrusal ayrılabilme durumuna ilişkin temsili gösterim Şekil 3.5'in (a) kısmında verilmiştir.




**Şekil 3. 5. (a) Doğrusal olarak ayrılabilme durumu (b) Doğrusal olarak ayrılamama durumu**

Denklem (3.11)'un S1 sınıfını ayıran A1 ayırıcı düzlemini oluşturan eşitsizlik, Denklem (3.12) eşitsizliğinin ise aynı şekilde S2 sınıfını ayıran A2 ayırıcı düzlemini oluşturan eşitsizlik olduğu kabul edildiğinde; A1 ayırıcı düzleminin orijine dik uzaklığı $1 - |b|/||w||$ ve A2 ayırıcı düzleminin orijine uzaklığı $|-1-b|/||w||$ olacaktır. Bu iki ayırıcı düzlemin en uygun ayırıcı düzleme uzaklıkları ise $|1|/||w||$ kadardır. Bir başka deyişle, iki örnek kümesi arasındaki uzaklık A1 ve A2 ayırıcı düzlemlerinin birbirlerine paralel olmalarından dolayı $|2|/||w||$ kadardır. Burada A1 ve A2 ayırıcı düzlemleri arasında eğitim verilerine ait hiçbir örnek bulunmadığına dikkat edilmelidir. Bu iki ayırıcı düzlem arasındaki en büyük uzaklık ise $||w||$ değerinin en aza indirgenmesiyle bulunabilir. DVM yöntemiyle yapılmaya çalışılan, bu iki ayırıcı düzlemin arasındaki uzaklığın (sınırın) en büyük olmasını sağlamaktır. Bu iki düzlem arasında en büyük sınırın bulunması

$$\min \frac{1}{2} ||w||^2 \tag{3.14}$$

$$y_i(w.x_i + b) - 1 \geq 0 \tag{3.15}$$

ile ifade edilir. Burada Denklem (3.14) çözülecek problemi, Denklem (3.15) ise problemin çözümü sırasında kullanılan koşulu ifade eder. Ayrıca, bu ifade ikinci dereceden bir en uygun şekle sokma problemidir. Problemin çözümü için Lagrange fonksiyonu uygulanabilir. Lagrange fonksiyonunun uygulanmasının iki sebebi vardır. Birincisi Lagrange çarpanlarının hesaplanması daha kolaydır. İkincisi ise problemin doğrusal olmayan durum için de genelleştirilmesi daha uygundur (Vapnik, 1995). Problemin Lagrange fonksiyonu ise,

$$L_P = \frac{1}{2}||w||^2 - \sum_{i=1}^{N} a_i y_i (w^T.x_i + b) + \sum_{i=1}^{N} a_i \tag{3.16}$$

şeklindedir. Bu formülde $a_i \geq 0$ değerleri pozitif Lagrange çarpanları olarak adlandırılır. Ancak Denklem (3.16)'de ifade edilen formülasyonun çözülmesi oldukça karmaşıktır. Çözümün bulunması için Denklem (3.16), Karush-Kuhn-Tucker (KKT) şartları kullanılarak ikili probleme dönüştürülmelidir. Bu problem için, KKT şartlarına bağlı çözüm;

$$L_P = \sum_i a_i - \frac{1}{2}\sum_{i,j} a_i a_j y_i y_j {x_i}^T x_j \tag{3.17}$$

ifadesi ile elde edilmiş olur. Denklem (3.17)'daki ifadenin çözümü $a_i \geq 0$ koşulları altında ikinci dereceden optimizasyon (en iyileme) problemi ile gerçekleştirilir. Burada dikkat edilirse, her eğitim örneği için bir tane Lagrange çarpanının olduğu görülür. Çözümde elde edilen Lagrange çarpanlarının büyük çoğunluğunun değeri sıfır olacaktır.

Geriye kalan $a_i \geq 0$ değerli $x_i$ örnekleri destek vektörlerdir ve A1 veya A2 ayırıcı düzlemlerinin üzerinde yer alırlar. Lagrange çarpanı sıfır olan örnekler ise A1 veya A2 ayırıcı düzlemlerinin arka taraflarında kalan örneklerdir.

Eğer, örnekler doğrusal olarak tamamen ayrılabilir durumda değilse, problemin çözümü için pozitif zayıflık değişkenleri $\xi_i = 1, 2, \ldots, N$ kullanılır. Bu duruma ait en uygun ayırıcı düzlemin temsili gösterimi Şekil 3.5'in (b) kısmında verilmiştir. Bu duruma göre Denklem (3.11) ve (3.12)'deki koşullar $\xi_i \geq 0$ olmak üzere zayıflık değişkenleri ile yeniden tanımlanacak olursa;

$$y_i = +1 \; için, \; w^T.x_i + b \geq +1 - \xi_i \tag{3.18}$$

$$y_i = -1 \; için, \; w^T.x_i + b \leq -1 + \xi_i \tag{3.19}$$

şeklinde denklemler elde edilecektir. $\xi_i = 0$ olması durumunda $x_i$ örneği doğru sınıflandırılmış, $\xi_i \geq 1$ ise yanlış sınıflandırılmış demektir.

Doğrusal olarak ayrılamama durumunda, eğitim verisi içindeki olası her duruma karşı bir çözüm üretilmesini engellemek için sisteme bir $C$ düzenleme parametresi eklenir. Aynı zamanda bu parametre, Lagrange çarpanlarının alabilecekleri en büyük değeri de göstermektedir. Bu şekilde Lagrange çarpanlarının $0 \leq a_i \leq C$ aralığında kalması sağlanmaktadır. $C$ düzenleme parametresi, DVM'nin eğitim aşamasında belirlemesi gereken parametrelerden biridir. Bu doğrultuda Lagrange formasyonu Denklem (3.21)'deki gibi yeniden düzenlenir:

$$L_P = \frac{1}{2}||w||^2 - C\sum_i \xi_i - \sum_i a_i \left\{y_i(w^T.x_i + b) - 1 + \xi_i\right\} - \sum_i \mu_i a_i \tag{3.20}$$

Burada $\mu_i$, zayıflık değişkenlerinin ($\xi_i$) pozitif değerde kalmasını sağlamak için kullanılmış bir Lagrange parametresidir. Bu Lagrange fonksiyonunun çözümü için de KKT şartları uygulanırsa;

$$L_d = \sum_i a_i - \frac{1}{2}\sum_{i,j} a_i a_j y_i y_j x_i{}^T x_j \tag{3.21}$$

elde edilir. Burada, $0 < a_i < C$ aralığında yer alan ve Lagrange çarpanlarına karşılık gelen $x_i$ değerleri destek vektörleri temsil eder.

Doğrusal olmayan problemlerde çözümü bulmanın yolu, çekirdek fonksiyonları ile örneklerin öncelikle daha yüksek boyutlu ve doğrusal olarak ayrılabilecekleri bir uzaya taşınıp çözümün bu yeni uzayda aranmasıdır. Bu durum $\Phi: \Re^d \mapsto H$ olmak üzere, $d$ boyutlu özellik uzayını bir $H$ Öklid uzayına taşıyan $\Phi$ fonksiyonunun olduğu

düşünülerek gerçekleştirilebilir. Böylece DVM'nin eğitim algoritması, $H$ uzayındaki verilerin

$$\Phi(x_i).\Phi(x_j) = K(x_i, x_j) \quad (3.22)$$

şeklindeki iç çarpımlarına bağlı olacaktır. Burada $K$ çekirdek fonksiyonunu temsil eder. Böylece sınıflandırıcı, eğitimden sonra herhangi bilinmeyen bir $x$ örneği Denklem $(3.23)$'deki karar fonksiyonuyla belirlenebilir.

$$f(x) = \sum_{i=1}^{N} a_i y_i \Phi(x_i).\Phi(x) + b = \sum_{i=1}^{N} a_i\, y_i K(x_i, x) + b \quad (3.23)$$

Bu fonksiyonda $N$ destek vektörlerin sayısını, $x_i$ ise destek vektörleri belirtir. Literatürde farklı alanlarda çekirdek fonksiyonlarının sınıflandırma başarımını değerlendirmek üzerine bazı çalışmalar yapılmıştır. Bu çalışmalarda genellikle radyal tabanlı çekirdek fonksiyonları kullanılmasıyla yüksek sınıflandırma başarımının elde edilebileceği vurgulanmıştır (Schölkopf ve ark., 1997). Çünkü radyal tabanlı çekirdek fonksiyonları; seçilen parametre aralıklarına bağlı olarak, hem sigmoid çekirdek fonksiyonunun hem de doğrusal çekirdek fonksiyonunun özelliklerini gösterebilmektedir (Keerthi ve Lin, 2003). Bu tezde; sıkça kullanılan doğrusal, polinom, radyal ve sigmoid çekirdek fonksiyonları içerisinde daha yüksek performans veren radyal tabanlı çekirdek fonksiyonu kullanılmıştır. Weka'da bulunan LibSVM sınıflandırma modeli kullanılmıştır.

#### 3.2.2.3. Yapay sinir ağları

Yapay sinir ağları **(**YSA), insan beynini model alan, nöron olarak adlandırılan basit işlem elemanlarından meydana gelen, doğrusal olmayan ve yüksek karmaşıklığa sahip bir bilgi işleme sistemidir. YSA evrensel bir tanımı olmamakla beraber çoğu bilim adamının "birçok küçük bilgi işleme biriminin bir araya gelmesiyle oluşturulan ağlar" olduğu üzerinde uzlaştıkları yapılardır (Dede, 2008). Ancak YSA, neredeyse her yayında farklı bir özelliği ya da uygulaması öne çıkarılarak yeniden tanımlanmaktadır. YSA algoritmasının temelleri, McCulloch ve Pitts'in 1943 yılında yayınladıkları makale ile atılmıştır. Ancak 1969 yılında algılayıcının doğrusal olmayan problemlerin çözümünde yetersiz olduğunu ispatlayan çalışmalar bu ilerlemenin önünü bir süreliğine kapatmıştır (Minsky ve Papert, 1969). 1985'li yıllara kadar durağan bir dönem geçiren YSA çalışmaları; 1986 yılında Rumelhart ve arkadaşlarının geri yayılım algoritmasını (back propagation) geliştirmesi ile doğrusal olmayan problemlere de yanıt vermeye ve dolayısıyla daha fazla akademisyenin ilgisini çekmeye başlamıştır (Rumelhart ve ark.,

1986). 1985'ten günümüze kadar yapılan çalışmalar; YSA uygulamalarındaki çeşitliliği arttırmış ve literatüre pek çok kaynak kazandırmıştır.

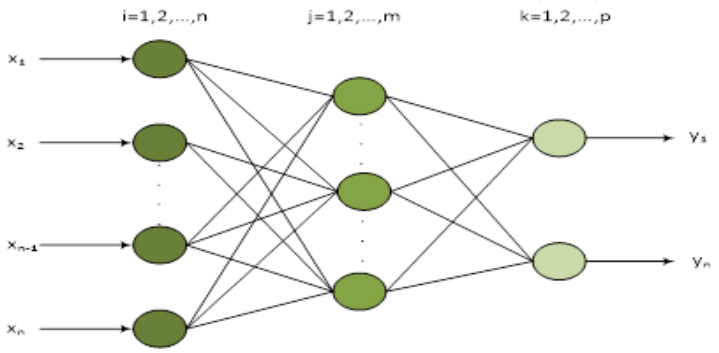
YSA yapıları, katmanlar halinde birlesen nöronlardan oluşmaktadır. Çeşitli ağırlık ve fonksiyonlarla bir araya gelen nöronların oluşturduğu bu katmanlar da farklı yapılardadır. Genel olarak bir YSA uygulamasında üç tip katman bulunur. Bunlar giriş katmanı, saklı katman(lar) ve çıkış katmanıdır.

Giriş katmanı, dış dünyadan gelen bilgilerin alındığı ve ağa sevk edildiği katmandır. Bu katmanda bilgi işleme yapılmamaktadır.

Saklı katman(lar), girdi katmanı ile çıktı katmanı arasında yer alır. Giriş katmanından gelen bilgiler, saklı katmanlar boyunca çeşitli algoritmalarla işlenerek çıkış katmanına gönderilir. Bir YSA algoritmasında, gerçekleştirilmek istenen uygulamanın niteliğine göre bir veya daha fazla saklı katman bulunabilir. Saklı katmanlar çeşitli kaynaklarda gizli katman veya ara katman olarak da adlandırılmaktadır.

Çıkış katmanı, saklı katman(lar)dan gelen bilgileri işleyerek YSA uygulamasının çıkışını oluşturan katmandır.

Basit bir YSA alt modeli olan Çok katmanlı algılayıcı (Multilayer perceptron, MLP) Şekil 3.6'daki gibi gösterilmiştir. Bu üç katmanın her birinde bulunan sinir hücreleri ve bunları birbirine bağlayan ağırlıklardır. Çember biçiminde gösterilenler sinir hücrelerini, hücreleri birbirine bağlayan çizgiler ise ağırlıkları göstermektedir. Bir yapay sinir ağındaki en önemli unsurlardan biri de sinir hücrelerinin birbirlerine veri aktarmalarını sağlayan bağlantılardır. Herhangi bir hücreden, diğer bir hücreye bilgi ileten bir bağlantı aynı zamanda bir ağırlık değerine sahiptir.




**Şekil 3. 6. Örnek bir YSA mimarisi**

Bir yapay sinir hücresinin matematiksel ifadesi,

$y_i = F\big(G(x)\big) = F\big(\sum_{i=1}^{n} w_{ij}\, x_j - Q_i\big);\ x_i = (x_1, x_2, \ldots\ldots x_n)$ (3.24)

biçiminde yazılabilir. Denklem (3.23)'de $x = (x_1, x_2, \ldots\ldots x_n)$ işlenmek üzere gelen girdi değişkenleridir. $w = (x_1, x_2, \ldots\ldots x_n)$ ise ağırlıklar olup bir sinir hücresine gelen bilginin önemini ve hücre üzerindeki etkisini gösterir(Chokmani ve ark., 2008 ). Ağırlıkların öğrenme süresince değerleri değişebilir. $Q_i$, eşik değerini belirtir. *F(.)* aktivasyon fonksiyonudur. Bu fonksiyona gelen *G(.)* girdiyi işleyerek çıktıyı üreten fonksiyondur. Sigmoid, tanjant sigmoid, sin, radial basis vb. gibi farklı aktivasyon fonksiyonları bulunmaktadır. Bir YSA uygulamasındaki hücrelerin tümü aynı veya farklı aktivasyon fonksiyonuna sahip olabilir. Hangi aktivasyon fonksiyonun kullanılacağına, kullanıcının denemeleri sonucunda karar verilir.

Bu tezde Weka'da bulunan ve bir YSA alt modeli olan Çok katmanlı algılayıcı modeli kullanılmıştır. Geri yayılma algoritması için Momentum oranı 0.2, öğrenme oranı 0.3 ve çıkarılan özellik sayısı ile sınıflandırılan sınıf oranının toplamının yarısı kadar gizli (saklı) katman kullanılmıştır. Örneğin cinsiyet sınıflandırılmasında; 80 adet özellik çıkarılmış ve 2 sınıf için sınıflandırılma istendiğinden gizli katman hücre sayısı 41 olmuştur.

### 3. 3. Performans Ölçütleri

Model başarısı değerlendirildiğinde bazı kavramlar bilinmelidir. Bunlar; hata oranı, kesinlik, duyarlılık ve f-ölçütüdür. Modelin başarısı doğru sınıflandırılan verilerin sayıları ile yanlış sınıflandırılan verilerin sayıları baz alınarak ölçülür.

Sınıflandırma işlemi sona erdiğinde elde edilen sonuçlar karışıklık matrisinde belirtilir. Matriste yer alan satır verileri; test setindeki verilerin gerçek adetlerini, sütunlar ise tahminleri gösterir.

**Tablo 3. 2. Performans ölçütleri**

| Gerçek Sistemde / Tahmini Sistemde | **Sınıf 1** | **Sınıf 0** |
|---|---|---|
| **Sınıf 1** | DP (Doğru Pozitif) | YP (Yanlış Pozitif) |
| **Sınıf 0** | YN (Yanlış Negatif) | DN (Doğru Negatif) |

**Doğru pozitif (True positive):** Tamamıyla doğru sınıflandırılan örneklerin sayısını belirtir.

**Doğru negatif (True negative):** Diğer sınıflara ait tamamıyla doğru sınıflandırılan örnek sayısı.

**Yanlış pozitif (False positive):** Yanlış sınıflandırılan örneklerin sayısı.

**Yanlış Negatif (False negative):** Diğer sınıflara ait yanlış sınıflandırılan örnek sayısı.

### 3.3.1. Doğruluk (Accuracy) - Hata Oranı (Error rate)

Model başarımının ölçülmesinde kullanılan en popüler ve basit yöntem, modele ait doğruluk oranıdır. Doğru sınıflandırılmış örnek sayısının (TP +TN), toplam örnek sayısına (TP+TN+FP+FN) oranıdır. Denklem 3.25 'te görüldüğü üzere:

$$\text{Doğruluk} = \frac{\text{TP+TN}}{\text{TP+FP+FN+TN}} \tag{3.25}$$

Hata oranı ise bu değerin 1'e tamlayanıdır. Diğer bir ifadeyle yanlış sınıflandırılmış örnek sayısının (FP+FN), toplam örnek sayısına (TP+TN+FP+FN) oranıdır.

$$\text{Hata oranı} = \frac{\text{FP+FN}}{\text{TP+FP+FN+TN}} \tag{3.26}$$

Veya

$$\text{Hata oranı} = 1 - \text{Doğruluk} \tag{3.27}$$

şeklinde ifade edilebilir.

### 3.3.2. Kesinlik (Precision)

Kesinlik, sınıfı 1 olarak tahmin edilmiş Doğru Pozitif örnek sayısının, sınıfı 1 olarak tahmin edilmiş tüm örnek sayısına oranıdır.

$$\text{Kesinlik} = \frac{\text{TP}}{\text{TP+FP}} \tag{3.28}$$

### 3.3.3. Hatırlama (Recall)

Gerçek değeri pozitif olup pozitif değere sınıflandırılan sayısının, gerçek değeri pozitif olanların tümüne oranıdır.

$$\text{Recall (r)} = \text{TP} / (\text{TP} + \text{FN}) \tag{3.29}$$

### 3.3.4. Duyarlılık (Sensitivity)

Doğru sınıflandırılmış pozitif örnek sayısının toplam pozitif örnek sayısına oranıdır.

$$\text{Duyarlılık} = \text{TP} / (\text{TP} + \text{FN}) \quad (3.30)$$

### 3.3.5. F- ölçütü (F-measure)

Kesinlik ve duyarlılık ölçütleri tek başına anlamlı bir karşılaştırma sonucu çıkarmamıza yeterli değildir. Her iki ölçütü beraber değerlendirmek daha doğru sonuçlar verir. Bunun için f-ölçütü tanımlanmıştır. F-ölçütü, kesinlik ve duyarlılığın harmonik ortalamasıdır.

$$\text{F} - \text{Ölçütü} = \frac{2 * \text{TP}}{2 * \text{TP} + \text{FP} + \text{FN}} \quad (3.31)$$

## 3.4. Önerilen Metot Diyagramları

Bu tezde; konuşmacılardan alınan seslerden çeşitli yöntemlerle elde edilen öz niteliklerin kullanılması ile kişilerin cinsiyeti ve yaş grubu tahmin edilmiştir. Çalışmanın diyagramı Şekil 3.7'de verilmiştir. Her blokta yapılan işlemler kısaca özetlenmiştir.


**Şekil 3. 7. Önerilen çalışma diyagramı**

Blok 1: Bu kısımda ses örnekleri toplanmıştır. Bunun için ilkokul, ortaokul, lise ve üniversite öğrenci gruplarının her birinden 8 erkek ve 8 kız öğrencinin sesi alınmıştır. Bu dört grup için toplamda 64 öğrenciden ses kaydı alınmıştır. Veri seti için bir kısmı Türkçe 'de birleşim gücü yüksek kelimeler bir kısmı da sık kullanılan rastgele kelimelerden oluşan 32 adet Türkçe kelime seçilmiştir. Ses kayıt işlemi; normal koşullarda ve konuşmacıya yakın mesafede 16 kHz örnekleme oranı ile yapılmıştır.

Blok 2: Ön işleme bloğudur. Elde edilen seslerin düzenleme işlemlerinin yapıldığı bloktur. Bunun için Matlab, Goldwave digital audio editor ve Wavepad sound editör programları kullanılmıştır.

Blok 3: Bu blok içerisinde seslerden öz nitelik çıkarımı yapılmıştır. Bunun için sık kullanılan yöntemlerden MFCC, LPCC ve eşit sayıda MFCC ve LPCC katsayılarının beraber alınması ile MF&LP (MFCC&LPCC) başlıklı öz nitelik çıkarım yöntemleri kullanılmıştır. Seslerin işlenmesi ve öz nitelik çıkartımı Matlab dijital sinyal işleme araçları yardımı ile gerçekleştirilmiştir.

Blok 4: Bu blokta elde edilen katsayılar kullanılarak farklı makine öğrenmesi yöntemleri ile sınıflandırma işlemi gerçekleştirilmiştir. Elde edilen özniteliklerin sağlıklı bir sonuç verebilmesi için öznitelik matrisimiz rastgele karıştırılmış ve Z-score normalizasyonuna tabi tutulmuştur (Herve, 2010). Z-score; bütün veri yığınlarındaki birimlerin, ortak bir birim aralığına yığılmasını sağlar. Aynı zamanda standart sapma ve ortalamayı baz alarak birbirinden farklı ölçü birimlerinin karşılaştırmasında kullanılır. Sınıflandırma işlemi 10-kat çapraz geçerlilik testine göre gerçekleştirilmiştir. Bu çalışmada k en yakın komşu (KNN), yapay sinir ağları (YSA) ve destek vektör makineleri (DVM) gibi farklı makine öğrenme yöntemleri sınıflandırma için kullanılmıştır.

Blok 5: Sonuçların paylaşıldığı bloktur. Farklı yaş gruplarındaki öğrenciler arasından alınan ses örneklerinin analizi yapılmış ve makine öğrenmesi yöntemleri ile sınıflandırma yapılmıştır.

# 4. BULGULAR

Bu çalışmada sık kullanılan MFCC ve LPCC öz nitelik çıkarımı yöntemleri kullanılmıştır. Aynı zamanda çıkarımı yapılan eşit sayıda MFCC ve LPCC katsayılarının beraber alınması ile MF&LP (MFCC&LPCC) başlıklı bir öz nitelik çıkarımı yöntemi denenmiştir. Örneğin; 20 katsayılı MF&LP vektörü; 10 adet MFCC ve 10 adet LPCC vektörlerinin birleşiminden meydana gelmektedir. Elde edilen özniteliklerin sağlıklı bir sonuç verebilmesi için öznitelik matrisimiz rastgele karıştırılmış ve z-score normalizasyonuna tabi tutulmuştur. Sınıflandırma işlemi 10-kat çapraz geçerlilik testine göre gerçekleştirilmiştir. Bu çalışmada k en yakın komşu (KNN), yapay sinir ağları (YSA) ve destek vektör makineleri (DVM) gibi farklı makine öğrenme yöntemleri sınıflandırma için kullanılmıştır.

Bu çalışmada elde edilen cinsiyet ve yaş sınıflandırmasına ait bulguların yanı sıra seçilen kelimelerin kendi içinde sınıflandırılmasına ait sonuçlar da aşağıda ayrı başlıklar altında verilmiştir.

## 4.1. Kelimelere Göre Sınıflandırma

Bu çalışmanın yapılmak istenmesinin nedeni; seçilen her bir kelimenin ayrıştırma özelliğine bakılarak, test ve karşılaştırma aşamasında bu verileri kullanabilmektir. Veri seti için aşağıdaki kelimeler seçilmiştir:

"Adam", "Arkadaşlık", "Barış", "Buğday", "Büyütmek", "Cumartesi", "Çocuk", "Doğmak", "Dürüst", "Düşünmek", "Fabrika", "Geliştirmek", "Hayat", "Hazırlanmak", "Hissetmek", "İnsan", "İnternet", "Kadın", "Mayıs", "Merhaba", "Millet", "Mutlaka", "Papatya", "Para", "Sabah", "Sağlamak", "Sanatçı", "Sevgi", "Sinema", "Tepe", "Umut", "Yarın".

Seslendirilen her bir kelimenin ayrıştırma özelliğine öz nitelik vektörleri yardımı ile bakılmıştır. Bunun için seslendirilen her bir kelime birer sınıf olarak alınmıştır. Böylece sınıflandırılacak 32 adet kelime sınıfı olmuştur. 64 öğrenciden bu kelimeler alındığı için 2048 adet ses örneğinin içinde başarı oranı denenmiştir. Yapılan denemeler sonucunda MFCC yöntemi ile elde edilen 40 katsayının; öz nitelik çıkarımı için uygun olduğu bulunmuştur. Böylece 2048*40 boyutunda bir matris elde edilmiş ve matrisin son sütununa ilgili sesin 32 kelime grubu içinde alfabetik olarak sıralaması; sınıf kodu

olarak atanmıştır. DVM (destek vektör makineleri) ile yapılan sınıflandırılma sonucu 2048 kelime içerisinde doğru sınıflandırılan kelime 627 adet olup, toplamda %30,6 oranında başarı sağlanmıştır. Elde edilen karışıklık matrisi ve performans oranları sırasıyla (Tablo 4.1) ve (Tablo 4.2) aşağıda verilmiştir.

**Tablo 4. 1. 32 kelime için sınıflandırma sonucu elde edilen karışıklık matrisi(K=Kelime)**

| | K1 | K2 | K3 | K4 | K5 | K6 | K7 | K8 | K9 | K10 | K11 | K12 | K13 | K14 | K15 | K16 | K17 | K18 | K19 | K20 | K21 | K22 | K23 | K24 | K25 | K26 | K27 | K28 | K29 | K30 | K31 | K32 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| K1 | 17 | 0 | 0 | 0 | 0 | 0 | 0 | 9 | 0 | 0 | 3 | 0 | 5 | 9 | 0 | 0 | 0 | 0 | 0 | 7 | 0 | 0 | 3 | 4 | 1 | 4 | 1 | 0 | 1 | 0 | 0 | 0 |
| K2 | 0 | 23 | 1 | 2 | 0 | 0 | 1 | 0 | 0 | 0 | 2 | 0 | 11 | 1 | 0 | 0 | 0 | 8 | 1 | 0 | 0 | 0 | 2 | 1 | 0 | 0 | 8 | 0 | 1 | 2 | 0 | 0 |
| K3 | 0 | 5 | 22 | 2 | 0 | 1 | 5 | 0 | 3 | 1 | 1 | 4 | 0 | 0 | 4 | 0 | 2 | 0 | 4 | 0 | 2 | 0 | 0 | 0 | 0 | 0 | 2 | 2 | 0 | 1 | 0 | 3 |
| K4 | 2 | 1 | 0 | 23 | 0 | 1 | 2 | 4 | 2 | 0 | 5 | 0 | 1 | 2 | 0 | 2 | 0 | 2 | 0 | 0 | 0 | 6 | 1 | 0 | 0 | 0 | 1 | 0 | 4 | 0 | 2 | 3 |
| K5 | 0 | 0 | 0 | 0 | 27 | 3 | 0 | 1 | 3 | 10 | 0 | 7 | 0 | 0 | 0 | 0 | 3 | 0 | 1 | 0 | 3 | 0 | 0 | 0 | 0 | 0 | 0 | 2 | 0 | 2 | 1 | 1 |
| K6 | 0 | 1 | 0 | 1 | 1 | 28 | 4 | 0 | 1 | 1 | 0 | 1 | 0 | 0 | 2 | 7 | 2 | 0 | 5 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 2 | 1 | 5 | 1 | 0 | 0 |
| K7 | 0 | 0 | 6 | 1 | 0 | 4 | 34 | 0 | 3 | 0 | 0 | 2 | 0 | 0 | 0 | 2 | 0 | 1 | 2 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 5 | 0 | 2 | 0 | 2 | 0 |
| K8 | 7 | 1 | 0 | 4 | 0 | 0 | 0 | 24 | 0 | 0 | 1 | 0 | 0 | 5 | 0 | 0 | 1 | 1 | 0 | 2 | 0 | 4 | 2 | 0 | 0 | 4 | 0 | 0 | 2 | 0 | 6 | 0 |
| K9 | 0 | 0 | 0 | 1 | 3 | 1 | 5 | 0 | 32 | 2 | 0 | 4 | 0 | 0 | 2 | 1 | 0 | 0 | 2 | 0 | 6 | 0 | 0 | 1 | 0 | 0 | 0 | 1 | 0 | 2 | 1 | 0 |
| K10 | 0 | 0 | 0 | 0 | 9 | 5 | 0 | 0 | 2 | 17 | 0 | 11 | 0 | 0 | 9 | 0 | 1 | 0 | 2 | 0 | 5 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 1 | 0 | 1 |
| K11 | 5 | 2 | 0 | 4 | 0 | 0 | 0 | 1 | 0 | 0 | 5 | 0 | 4 | 5 | 0 | 0 | 0 | 0 | 0 | 6 | 0 | 8 | 9 | 2 | 0 | 3 | 0 | 0 | 7 | 0 | 0 | 3 |
| K12 | 0 | 0 | 2 | 0 | 9 | 3 | 0 | 0 | 3 | 8 | 0 | 10 | 0 | 0 | 7 | 0 | 2 | 0 | 1 | 0 | 8 | 0 | 0 | 0 | 0 | 0 | 0 | 10 | 0 | 0 | 0 | 1 |
| K13 | 1 | 6 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 2 | 0 | 17 | 0 | 0 | 0 | 3 | 1 | 1 | 2 | 0 | 0 | 5 | 11 | 0 | 0 | 7 | 0 | 2 | 0 | 0 | 6 |
| K14 | 9 | 3 | 0 | 3 | 0 | 0 | 1 | 1 | 0 | 0 | 5 | 0 | 2 | 5 | 0 | 3 | 1 | 6 | 0 | 1 | 0 | 9 | 1 | 1 | 0 | 7 | 0 | 0 | 3 | 0 | 0 | 3 |
| K15 | 0 | 0 | 2 | 0 | 0 | 3 | 1 | 0 | 1 | 11 | 0 | 8 | 0 | 0 | 18 | 1 | 0 | 0 | 1 | 0 | 6 | 0 | 0 | 0 | 0 | 0 | 2 | 3 | 4 | 2 | 0 | 1 |
| K16 | 1 | 0 | 1 | 3 | 0 | 6 | 1 | 1 | 0 | 0 | 1 | 0 | 0 | 3 | 0 | 16 | 4 | 2 | 0 | 0 | 0 | 3 | 0 | 0 | 0 | 1 | 2 | 0 | 14 | 0 | 4 | 1 |
| K17 | 0 | 0 | 0 | 1 | 1 | 6 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 2 | 22 | 0 | 5 | 1 | 10 | 0 | 0 | 0 | 0 | 0 | 1 | 1 | 2 | 5 | 0 | 5 |
| K18 | 1 | 7 | 0 | 6 | 0 | 0 | 1 | 0 | 0 | 0 | 1 | 0 | 3 | 4 | 0 | 4 | 1 | 22 | 0 | 1 | 0 | 2 | 0 | 0 | 0 | 0 | 4 | 0 | 1 | 3 | 0 | 3 |
| K19 | 0 | 0 | 1 | 1 | 1 | 6 | 4 | 0 | 3 | 1 | 0 | 2 | 0 | 0 | 1 | 0 | 8 | 4 | 17 | 0 | 4 | 1 | 0 | 0 | 0 | 0 | 5 | 0 | 2 | 2 | 0 | 1 |
| K20 | 2 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 0 | 4 | 0 | 3 | 2 | 0 | 0 | 1 | 1 | 0 | 14 | 0 | 7 | 13 | 10 | 0 | 2 | 0 | 0 | 1 | 0 | 1 | 2 |
| K21 | 0 | 0 | 0 | 0 | 2 | 1 | 0 | 0 | 3 | 4 | 0 | 8 | 0 | 2 | 8 | 0 | 6 | 0 | 1 | 0 | 17 | 0 | 0 | 0 | 0 | 0 | 1 | 3 | 0 | 4 | 1 | 3 |
| K22 | 1 | 1 | 0 | 4 | 0 | 0 | 0 | 3 | 0 | 0 | 11 | 0 | 1 | 4 | 0 | 3 | 0 | 2 | 0 | 2 | 0 | 11 | 6 | 1 | 3 | 4 | 1 | 0 | 4 | 0 | 0 | 2 |
| K23 | 2 | 1 | 0 | 1 | 0 | 0 | 0 | 1 | 0 | 0 | 4 | 0 | 7 | 1 | 0 | 1 | 1 | 1 | 0 | 6 | 0 | 7 | 11 | 3 | 8 | 2 | 1 | 0 | 3 | 0 | 1 | 2 |
| K24 | 4 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 6 | 1 | 0 | 0 | 0 | 3 | 0 | 10 | 0 | 1 | 1 | 29 | 3 | 0 | 0 | 0 | 2 | 1 | 0 | 1 |
| K25 | 4 | 1 | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 0 | 1 | 0 | 2 | 2 | 0 | 2 | 0 | 0 | 0 | 3 | 0 | 6 | 7 | 2 | 13 | 15 | 2 | 0 | 2 | 0 | 0 | 1 |
| K26 | 7 | 1 | 0 | 0 | 0 | 0 | 0 | 8 | 0 | 0 | 1 | 0 | 3 | 6 | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 3 | 2 | 2 | 18 | 8 | 1 | 0 | 2 | 0 | 0 | 1 |
| K27 | 0 | 8 | 7 | 0 | 0 | 0 | 4 | 0 | 0 | 0 | 0 | 0 | 4 | 0 | 1 | 3 | 3 | 1 | 2 | 0 | 0 | 1 | 3 | 0 | 1 | 0 | 16 | 0 | 7 | 1 | 0 | 2 |
| K28 | 0 | 0 | 3 | 0 | 2 | 1 | 2 | 0 | 6 | 4 | 1 | 9 | 0 | 0 | 4 | 0 | 1 | 0 | 0 | 0 | 10 | 0 | 0 | 0 | 0 | 0 | 2 | 16 | 1 | 0 | 1 | 1 |
| K29 | 2 | 0 | 0 | 3 | 0 | 4 | 0 | 1 | 1 | 0 | 1 | 0 | 1 | 1 | 0 | 11 | 1 | 1 | 1 | 1 | 0 | 2 | 0 | 0 | 2 | 0 | 3 | 0 | 22 | 1 | 3 | 2 |
| K30 | 0 | 0 | 1 | 0 | 1 | 0 | 0 | 0 | 2 | 0 | 0 | 0 | 1 | 0 | 2 | 0 | 3 | 1 | 0 | 0 | 6 | 0 | 1 | 0 | 0 | 0 | 0 | 1 | 0 | 42 | 1 | 2 |
| K31 | 1 | 0 | 1 | 3 | 1 | 4 | 0 | 5 | 2 | 0 | 0 | 0 | 0 | 0 | 0 | 5 | 0 | 0 | 2 | 0 | 2 | 1 | 0 | 0 | 0 | 0 | 1 | 2 | 2 | 0 | 32 | 0 |
| K32 | 0 | 0 | 2 | 4 | 0 | 1 | 1 | 0 | 0 | 0 | 2 | 0 | 4 | 0 | 1 | 0 | 5 | 6 | 2 | 1 | 1 | 3 | 1 | 5 | 0 | 1 | 2 | 0 | 1 | 4 | 0 | 17 |

**Tablo 4. 2. 32 kelime için sınıflandırma sonucu elde edilen performans sonuçları**

| Sınıflar | Doğru pozitiflerin oranı(TP Rate) | Yanlış negatiflerin oranı(FP Rate ) | Kesinlik (Precision) | Duyarlılık (Recall) | F-ölçütü (F-Measure) |
|---|---|---|---|---|---|
| K1 | 0,266 | 0,025 | 0,258 | 0,266 | 0,262 |
| K2 | 0,359 | 0,02 | 0,371 | 0,359 | 0,365 |
| K3 | 0,344 | 0,014 | 0,449 | 0,344 | 0,389 |
| K4 | 0,359 | 0,022 | 0,343 | 0,359 | 0,351 |
| K5 | 0,422 | 0,015 | 0,474 | 0,422 | 0,446 |
| K6 | 0,438 | 0,025 | 0,359 | 0,438 | 0,394 |
| K7 | 0,531 | 0,016 | 0,515 | 0,531 | 0,523 |
| K8 | 0,375 | 0,019 | 0,387 | 0,375 | 0,381 |
| K9 | 0,5 | 0,018 | 0,478 | 0,5 | 0,489 |
| K10 | 0,266 | 0,021 | 0,288 | 0,266 | 0,276 |
| K11 | 0,078 | 0,024 | 0,096 | 0,078 | 0,086 |
| K12 | 0,156 | 0,028 | 0,152 | 0,156 | 0,154 |
| K13 | 0,266 | 0,029 | 0,227 | 0,266 | 0,245 |
| K14 | 0,078 | 0,024 | 0,094 | 0,078 | 0,085 |
| K15 | 0,281 | 0,021 | 0,3 | 0,281 | 0,29 |
| K16 | 0,25 | 0,024 | 0,254 | 0,25 | 0,252 |
| K17 | 0,344 | 0,025 | 0,31 | 0,344 | 0,326 |
| K18 | 0,344 | 0,021 | 0,349 | 0,344 | 0,346 |
| K19 | 0,266 | 0,017 | 0,34 | 0,266 | 0,298 |
| K20 | 0,219 | 0,022 | 0,241 | 0,219 | 0,23 |
| K21 | 0,266 | 0,032 | 0,21 | 0,266 | 0,234 |
| K22 | 0,172 | 0,032 | 0,147 | 0,172 | 0,158 |
| K23 | 0,172 | 0,029 | 0,162 | 0,172 | 0,167 |
| K24 | 0,453 | 0,022 | 0,403 | 0,453 | 0,426 |
| K25 | 0,203 | 0,018 | 0,265 | 0,203 | 0,23 |
| K26 | 0,125 | 0,022 | 0,157 | 0,125 | 0,139 |
| K27 | 0,25 | 0,027 | 0,229 | 0,25 | 0,239 |
| K28 | 0,25 | 0,014 | 0,372 | 0,25 | 0,299 |
| K29 | 0,344 | 0,038 | 0,227 | 0,344 | 0,273 |
| K30 | **0,656** | 0,016 | 0,568 | 0,656 | 0,609 |
| K31 | 0,5 | 0,012 | 0,571 | 0,5 | 0,533 |
| K32 | 0,266 | 0,026 | 0,25 | 0,266 | 0,258 |
| **Ortalama** | **0,306** | **0,022** | **0,308** | **0,306** | **0,305** |

Elde edilen sonuçlar ışığında; 64 defa seslendirilen her bir kelimenin içinde; en yüksek tanınma oranına sahip kelimeler 42 defa ile K30 (“Tepe”), 34 defa ile K7

(“Çocuk”) ve 32 defa ile K9 (“Dürüst”) ve K31 (“Umut”) kelimeleri olmuştur. En düşük tanınma oranına sahip kelimeler ise 5 defa ile K11 (“Fabrika”) ve K14 (“Hazırlanmak”), 8 defa ile K26 (“Sağlamak”) ve 11 defa ile K22 (“Mutlaka”) ve K23 (“Papatya”) kelimeleri olmuştur.

Böylece bu araştırmada kullanılan kelimelerin içindeki en yüksek tanınma oranına sahip kelimelerin, cinsiyet ve yaş aralığını sınıflandırma aşamasında ayrıştırıcı özelliğe sahip olduğu düşünülebilir. Aşağıda cinsiyet ve yaş sınıflandırmasında bahsedilecek akustik görüntüleme yöntemlerinde; bu kısımda elde edilen kelime bilgileri, karşılaştırma işlemlerinde kullanılacaktır.

### 4.2. Cinsiyete Göre Sınıflandırma

Bu çalışmada toplamda 64 öğrenciden ses kaydı alınmıştır. Her bir öğrenciden 32 kelime için eşit sayıda ses alınmıştır. Toplamda 2048 ses örneği toplanmıştır. Bu kayıtların 32’si erkek ve 32’si kız öğrencilere ait olup, her bir cinsiyet için 1024 adet ses örneği vardır.

Erkeklerin yaş ortalaması 15,84 olup, kızların yaş ortalaması ise 15,68’dir. Erkeklere ait ses kayıtlarının toplam süresi 830,85 saniye (yaklaşık 13,84 dakika) olup, bu süre kızlar için 904,72 saniyedir (yaklaşık 15,07 dakika). Toplamda; erkeklere ve kızlara ait 1735,58 saniye (yaklaşık 28,91 dakika) uzunluğunda ses kaydı analiz edilmiştir.

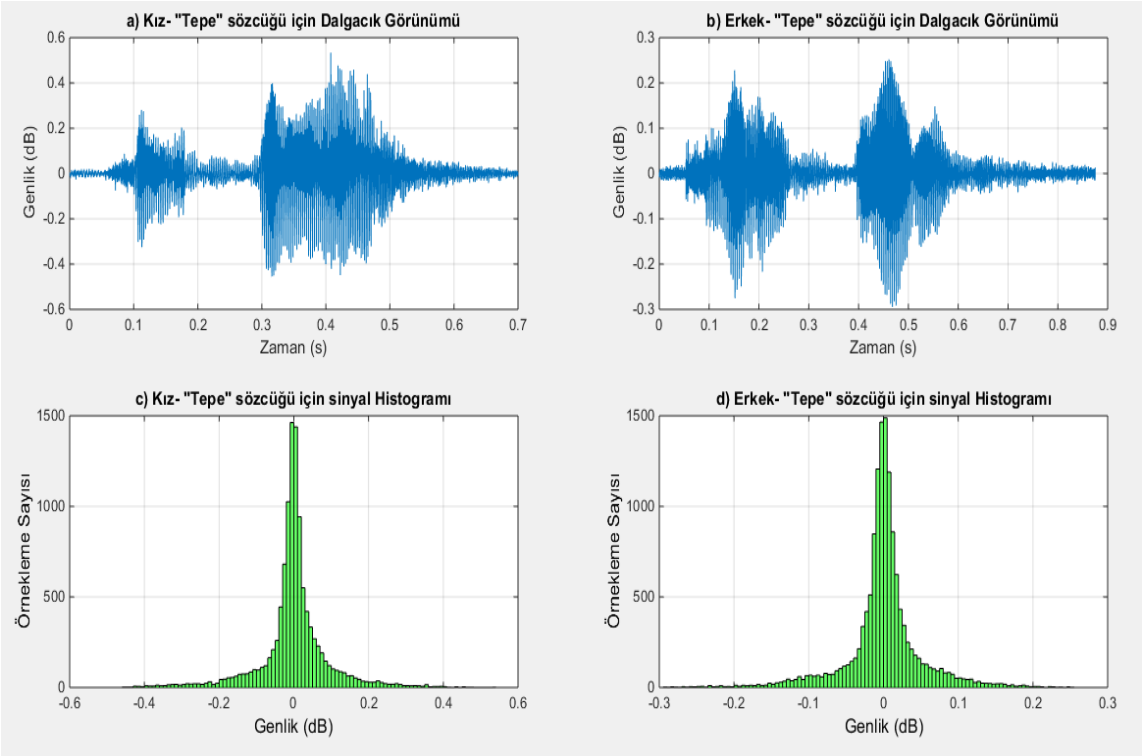
Erkekler için en kısa ses kaydının süresi 0,36 saniye, en uzun ses kaydının süresi 1,72 saniye ve her bir kaydın ortalama süresi 0,81 saniyedir. Bu süre kızlar için sırasıyla 0,37, 2,31 ve 0,88 saniyedir. Cinsiyetlere göre ses kayıt uzunluğunun istatiksel dağılımı aşağıdaki tabloda (Tablo 4.3) verilmiştir.

**Tablo 4. 3. Cinsiyetlere göre ses kayıt sürelerinin istatiksel bilgisi**

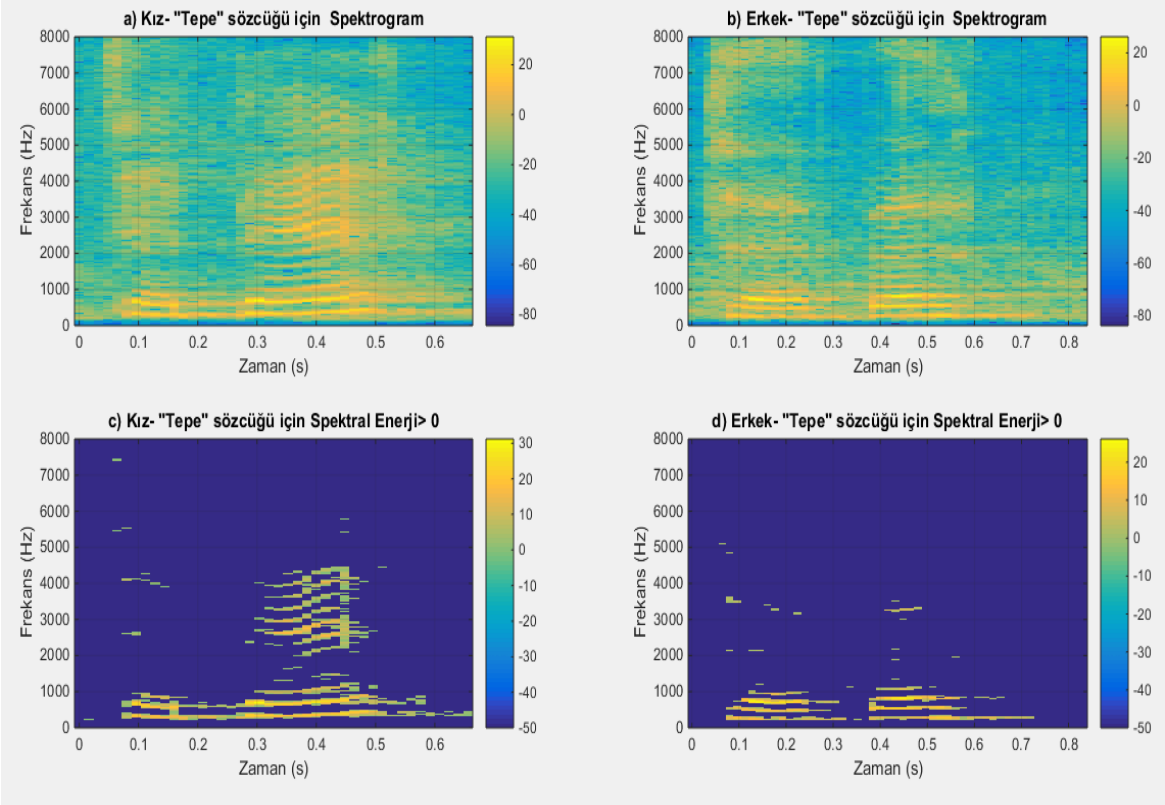
| | Ses Kayıtlarının Toplam Süresi (s) | En Kısa Ses Kaydının Süresi (s) | En Uzun Ses Kaydının Süresi (s) | Ses Kayıtlarının Ortalama Süresi (s) |
|---|---|---|---|---|
| Erkek | 830,85 | 0,36 | 1,72 | 0,81 |
| Kız | 904,72 | 0,37 | 2,31 | 0,88 |

### 4.2.1. Cinsiyete göre akustik karşılaştırma

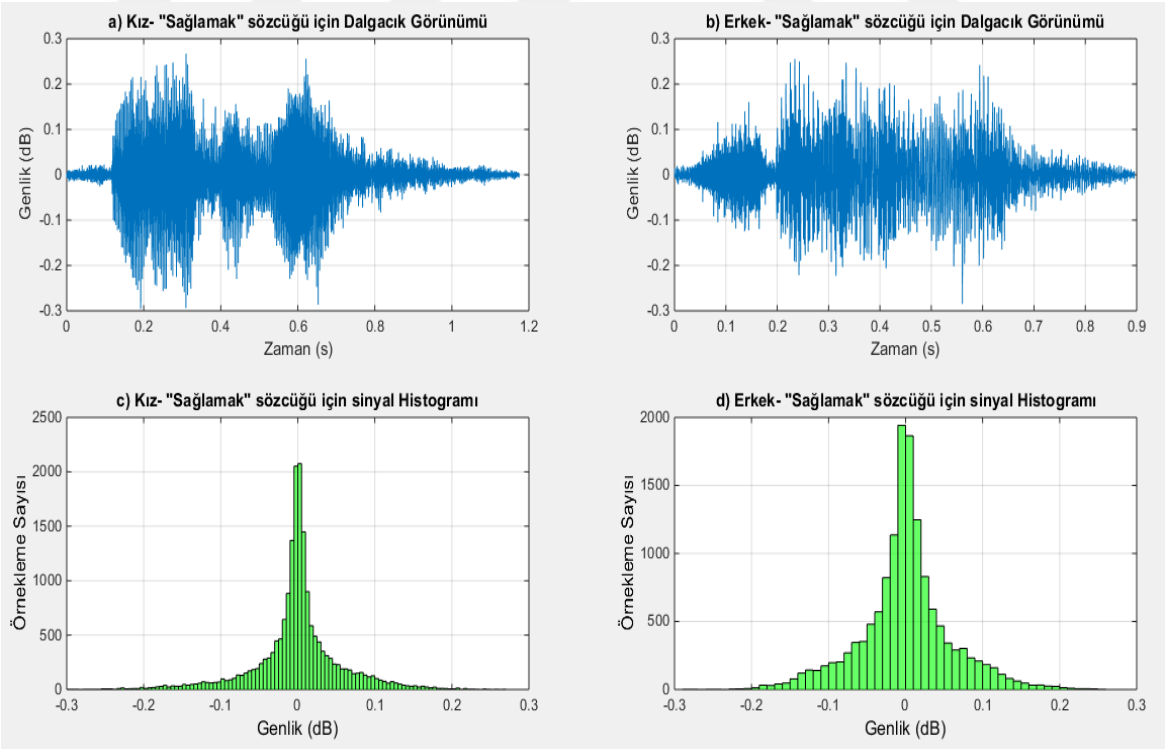
Bölüm 4.1 de belirtilen ve 32 adet kelimenin kendi içinde sınıflandırılması sonucu; sınıflandırmada en çok tanınan kelimelerden biri olan “Tepe” ve sınıflandırmada en az tanınan kelimelerden biri olan “Sağlamak” kelimeleri için Matlab programı yardımı ile elde edilen akustik görüntüleme yöntemlerinden olan sesin dalga görünümü (osilogram) ve sinyal histogramı ve sesin süre, frekans ve şiddet özelliklerini bir arada inceleyen spektrogram ve sese ait enerjinin daha net görünümü adına spektral enerjinin sıfırdan küçük kısımlarının silindiği gösterimler aşağıda (Şekil 4.1, Şekil 4.2, Şekil 4.3 ve Şekil 4.4) verilmiştir.



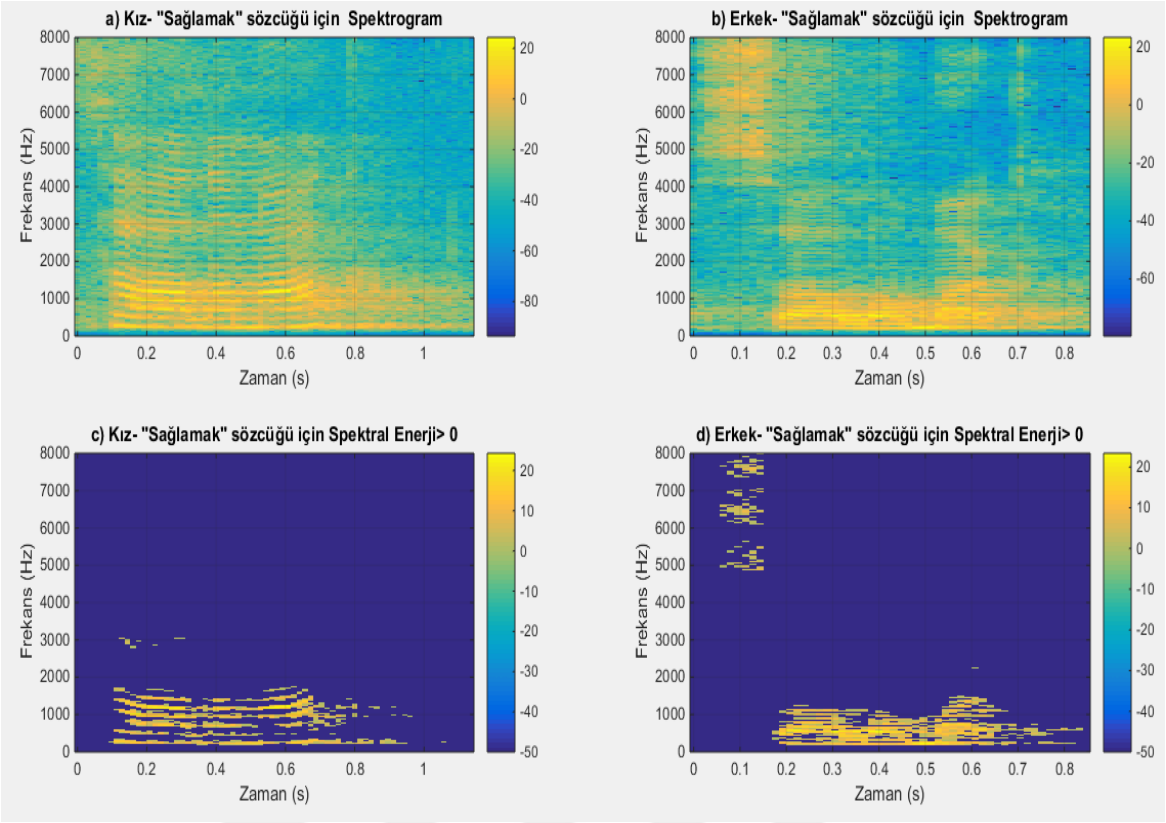

**Şekil 4. 1. “Tepe” Sözcüğü için; (a) Kız öğrenciye ait sesin dalga görünümü (b) Erkek öğrenciye ait sesin dalga görünümü (c) Kız öğrenciye ait sesin histogramı (d) Erkek öğrenciye ait sesin histogramı**




**Şekil 4. 2. "Tepe" sözcüğü için; (a) Kız öğrenci için sesin spektrogramı (b) Erkek öğrenci için sesin spektrogramı (c) Kız öğrenci için spektral enerji>0 (d) Erkek öğrenci için spektral enerji>0**




**Şekil 4. 3. "Sağlamak" Sözcüğü için; (a) Kız öğrenciye ait sesin dalga görünümü (b) Erkek öğrenciye ait sesin dalga görünümü (c )Kız öğrenciye ait sesin histogramı (d) Erkek öğrenciye ait sesin histogramı**




**Şekil 4. 4. "Sağlamak" sözcüğü için; (a) Kız öğrenci için sesin spektrogramı (b) Erkek öğrenci için sesin spektrogramı (c) Kız öğrenci için spektral enerji>0 (d) Erkek öğrenci için spektral enerji>0**

Şekil 4.1 ve Şekil 4.2'de 9 yaşında biri erkek biri kız öğrenciye ait olan ve kelime grupları içerisinde tanınma frekansı yüksek olan "Tepe" sözcüğüne ait iki ses kaydının verileri görsel akustik karşılaştırmalar şeklinde verilmiştir. Şekil 4.3 ve Şekil 4.4'de ise 18 yaşında biri erkek biri kız öğrenciye ait olan ve kelime grupları içerisinde tanınma frekansı düşük olan "Sağlamak" sözcüğüne ait iki ses kaydının verileri görsel akustik karşılaştırmalar şeklinde verilmiştir.

Şekil 4.1 (a), (b) ve Şekil 4.3 (a), (b) kısımlarında seslerin dalgacık görünümü; sesin genliği ve zaman ekseninde karşılaştırılmış olup, Şekil 4.1 (c), (d) ve Şekil 4.3 (c), (d) kısımlarında ise sesin örnekleme sayısı ve genlik yoğunluğu histogram şeklinde verilmiştir. Şekil 4.2 (a), (b) ve Şekil 4.4 (a) (b) kısımlarında frekans-zaman ekseninde sese ait şiddetin spektrogramı görsel olarak verilmiştir. Bu görüntüler sinyale Hızlı fourier dönüşümü (Fast fourier transform, FFT) uygulanarak elde edilmiştir. Şekil 4.2 (c), (d) ve Şekil 4.4 (c), (d) kısımlarında ise karşılaştırılan iki sesin daha net anlaşılabilmesi adına spektral enerjinin sıfırdan küçük kısımları silinmiş ve görsel olarak verilmiştir.

### 4.2.2. Cinsiyete göre sınıflandırmada başarı oranları

Bu kısımda MFCC, LPCC ve MF&LP karışımı için farklı katsayılarda k en yakın komşu (KNN), destek vektör makineleri (DVM) ve yapay sinir ağları (YSA) ile sınıflandırma yapılmıştır. Yukarıdaki üç öznitelik çıkarımı katsayısının; bahsedilen üç sınıflandırma metodu ile sınıflandırılması sonucu elde edilen karışıklık matrisleri ve performans oranları aşağıda verilmiştir.

MFCC yöntemi için farklı katsayılarla yapılan denemeler sonucu elde edilen başarı oranları aşağıda (Tablo 4.4) verilmiştir.

**Tablo 4. 4.Cinsiyet ayrıştırmada MFCC katsayıları için başarı oranları**

| KATSAYI ADEDİ | KNN (%) | DVM (%) | YSA (%) |
|---|---|---|---|
| 10 | 79,1 | 83 | 75,4 |
| 20 | 90,9 | 91,6 | 86,6 |
| **30** | 91,1 | **92,3** | 87,6 |
| 40 | 90,9 | 92 | 88,5 |
| 50 | 91,5 | 92 | 87,4 |
| 60 | 90,8 | 92,1 | 89 |
| 70 | 91,3 | 92,2 | 87,7 |
| 80 | 91,4 | 91,9 | 89,1 |
| 90 | 91,2 | 92,2 | 88,5 |
| 100 | 90,7 | 91,9 | 88,3 |

Yapılan denemeler sonucu 30 MFCC katsayısı üzeri için hem başarı oranı düşmüş, hem de hesaplama maliyeti artmıştır. En yüksek sınıflandırma başarı oranı %92,3 olarak gözlenmiştir. Böylece; 2048 ses örneği içerisinde 1890 adet ses doğru sınıflandırılmıştır. 30 MFCC katsayı kullanılarak bu başarı elde edilmiştir. Sınıflandırma için başarı oranı en yüksek metot DVM olmuştur. 30 MFCC katsayısının sınıflandırılması sonucu elde edilen karışıklık matrisi (Tablo 4.5) ve performans oranı (Tablo 4.6) aşağıda verilmiştir.

**Tablo 4. 5. Cinsiyet- MFCC ile elde edilen karışıklık matrisi (C1: Erkek, C2: Kadın)**

| Aktivite | C1 | C2 |
|---|---|---|
| C1 | **948** | 76 |
| C2 | 82 | **942** |

**Tablo 4. 6. Cinsiyet- MFCC ile elde edilen performans sonuçları**

| Sınıflar | Doğru pozitiflerin oranı(TP Rate) | Yanlış negatiflerin oranı(FP Rate ) | Kesinlik (Precision) | Duyarlılık (Recall) | F-ölçütü (F-Measure) |
|---|---|---|---|---|---|
| C1 | 0,926 | 0,080 | 0,920 | 0,926 | 0,923 |
| C2 | 0,920 | 0,074 | 0,925 | 0,920 | 0,923 |
| **Ortalama** | **0,923** | **0,077** | **0,923** | **0,923** | **0,923** |

LPCC yöntemi için farklı katsayılarla yapılan denemeler sonucu elde edilen başarı oranları aşağıda (Tablo 4.7) verilmiştir.

**Tablo 4. 7. Cinsiyet ayrıştırmada LPCC katsayıları için başarı oranları**

| KATSAYI ADEDİ | KNN (%) | DVM (%) | YSA (%) |
|---|---|---|---|
| 10 | 72,8 | 76 | 71,5 |
| 20 | 83,7 | 84,3 | 77,5 |
| 30 | 87,5 | 88,2 | 82,3 |
| 40 | 89 | 88,8 | 83,3 |
| 50 | 91 | 90,4 | 85,6 |
| 60 | 93 | 92,3 | 87 |
| 70 | 93,4 | 93,1 | 88,9 |
| **80** | **94,4** | 93,7 | 90,1 |
| 90 | 94,3 | 94,4 | 90,2 |
| 100 | 93,9 | 94,1 | 91,7 |

Yapılan denemeler sonucu en yüksek sınıflandırma başarı oranı %94,4 olarak gözlenmiştir. Böylece; 2048 ses örneği içerisinde 1933 adet ses doğru sınıflandırılmıştır. 80 LPCC katsayısı kullanılarak bu başarı elde edilmiştir. Sınıflandırma için başarı oranı en yüksek metot KNN olmuştur. 80 LPCC katsayısının sınıflandırılması sonucu elde edilen karışıklık matrisi (Tablo 4.8) ve performans oranı (Tablo 4.9) aşağıda verilmiştir.

**Tablo 4. 8. Cinsiyet- LPCC ile elde edilen karışıklık matrisi**

| Aktivite | C1 | C2 |
|---|---|---|
| C1 | **975** | 49 |
| C2 | 66 | **958** |

**Tablo 4. 9. Cinsiyet- LPCC ile elde edilen performans sonuçları**

| Sınıflar | Doğru pozitiflerin oranı(TP Rate) | Yanlış negatiflerin oranı(FP Rate ) | Kesinlik (Precision) | Duyarlılık (Recall) | F-ölçütü (F-Measure) |
|---|---|---|---|---|---|
| C1 | 0,952 | 0,064 | 0,937 | 0,952 | 0,944 |
| C2 | 0,936 | 0,048 | 0,951 | 0,936 | 0,943 |
| **Ortalama** | **0,944** | **0,056** | **0,944** | **0,944** | **0,944** |

MF&LP öz nitelik karışımı yöntemi için farklı katsayılarla yapılan denemeler sonucu elde edilen başarı oranları aşağıda (Tablo 4.10) verilmiştir.

**Tablo 4. 10. Cinsiyet ayrıştırmada MF&LP katsayıları için başarı oranları**

| TOPLAM KATSAYI: MF&LP | KNN (%) | DVM (%) | YSA (%) |
|---|---|---|---|
| 10 | 69,1 | 74 | 68,2 |
| 20 | 83,1 | 85,1 | 78 |
| 30 | 91 | 91,6 | 87,1 |
| 40 | 91,7 | 92,4 | 88,4 |
| 50 | 93 | 93,2 | 89,8 |
| 60 | 93,7 | 93 | 89,4 |
| 70 | 93,5 | 93,8 | 91,4 |
| **80** | **94,6** | 93,9 | 90,4 |
| 90 | 94,4 | 94,5 | 92,3 |
| 100 | 94 | 94,4 | 92,8 |

Yapılan denemeler sonucu en yüksek sınıflandırma başarı oranı %94,6 olarak gözlenmiştir. Böylece; 2048 ses örneği içerisinde 1937 adet ses doğru sınıflandırılmıştır. 80 MF&LP (40 adet MFCC ve 40 adet LPCC) katsayısı kullanılarak bu başarı elde edilmiştir. Sınıflandırma için başarı oranı en yüksek metot KNN olmuştur. 80 MF&LP katsayısının sınıflandırılması sonucu elde edilen karışıklık matrisi (Tablo 4.11) ve performans oranı (Tablo 4.12) aşağıda verilmiştir.

**Tablo 4. 11. Cinsiyet-MF&LP ile elde edilen karışıklık matrisi**

| Aktivite | C1 | C2 |
|---|---|---|
| C1 | **958** | 66 |
| C2 | 45 | **979** |

**Tablo 4. 12. Cinsiyet- MF&LP ile elde edilen performans sonuçları**

| Sınıflar | Doğru pozitiflerin oranı(TP Rate) | Yanlış negatiflerin oranı(FP Rate ) | Kesinlik (Precision) | Duyarlılık (Recall) | F-ölçütü (F-Measure) |
|---|---|---|---|---|---|
| C1 | 0,936 | 0,044 | 0,955 | 0,936 | 0,945 |
| C2 | 0,956 | 0,064 | 0,937 | 0,956 | 0,946 |
| **Ortalama** | **0,946** | **0,054** | **0,946** | **0,946** | **0,946** |

Aşağıda (Tablo 4.13) tüm özellik çıkarım yöntemleri için elde edilen başarı oranı ve başarılı olunan sınıflandırma metodunun yanı sıra Matlab'da öz nitelik çıkarımı için geçen zaman ve hazırlanan veri setlerinin Weka yardımı ile sınıflandırılırken aldığı zaman özetlenmiştir.

**Tablo 4. 13. Cinsiyet için elde edilen en başarılı performans/zaman sonuçları**

| Özellik Çıkarım Metodu | Katsayı Adedi | Sınıflandırma Yöntemi | Başarı Oranı (%) | Öz Nitelik Çıkarımı İçin Geçen Zaman (saniye) | Sınıflandırma İçin Geçen Zaman (saniye) |
|---|---|---|---|---|---|
| MFCC | 30 | DVM | 92,3 | 45,2 | 3,6 |
| LPCC | 80 | KNN | 94,4 | 96,2 | 8,1 |
| MF&LP | 80 | KNN | 94,6 | 101,4 | 7,8 |

Kullanılan sınıflandırma yöntemleri içinde en başarılı sonuçlar; MFCC için DVM, LPCC ve MF&LP için KNN öğrenme yöntemi olmuştur. Sınıflandırma için kullanılan YSA ise hem başarı oranı hem de zaman maliyeti açısından düşük performans göstermiştir.

Matlab'da öz nitelik çıkarımı (shuffle, normalizasyon ve veri seti haline getirme süreleri de dâhil) için geçen zaman; bu çalışmanın her aşamasının hazırlandığı bir diz üstü bilgisayarda hesaplanmıştır (i7-4510u işlemcili, 8 GB ddr3 ram kapasiteli). Her bir öz nitelik çıkartım metodunun belirlenen katsayıları için; zaman maliyeti hesaplaması üç defa test edilmiş ve ortalaması alınarak yukarıdaki tabloya eklenmiştir. Hazırlanan veri setlerinin Weka yardımı ile modellenirken (10-kat çapraz geçerlilik testi süresi de dâhil) aldığı süreler de yukarıda belirtilmiştir. Toplam zaman maliyeti ve %90 başarı eşiğini en az katsayı ile başarma açısından MFCC daha başarılı olmuştur.

Netice itibari ile cinsiyet sınıflandırmada en yüksek başarı oranı %94,6 ile MF&LP öz nitelik çıkarımı yönteminde görülmektedir. Bunun için 80 katsayı

kullanılmıştır. Sonrasında LPCC için %94,4 ve MFCC için %92,3 başarı oranı gözlemlenmiş ve sırasıyla 80 ve 30 katsayı kullanılmıştır.

### 4.3. Yaş Grubuna Göre Sınıflandırma

Bu çalışmada ses örnekleri için ilkokul, ortaokul, lise ve üniversite öğrenci gruplarının her birinden 8 erkek ve 8 kız öğrencinin sesi alınmıştır. Bu dört grup için toplamda 64 öğrenciden ses kaydı alınmıştır. Her bir yaş grubu için 512 adet ses örneği alınmış olup, toplamda 2048 ses örneği toplanmıştır.

İlkokul öğrencilerinin yaş aralığı 9 ve 10, ortaokul öğrencilerinin yaş aralığı 13, 14 ve 15, lise öğrencilerinin yaş aralığı 17 ve 18 ve üniversite öğrencilerinin yaş aralığı 21, 22 ve 23 yaş aralığındadır. İlkokul öğrencileri için yaş ortalaması 9.62, ortaokul yaş ortalaması 13.75, lise yaş ortalaması 17.625 ve üniversite için 22.06 yaş ortalaması bulunmaktadır.

İlkokul yaş grubuna ait ses kayıtlarının toplam süresi 475,58 saniye (yaklaşık 7,92 dakika), ortaokul yaş grubuna ait ses kayıtlarının toplam süresi 402,29 saniye (yaklaşık 6,71 dakika), lise yaş grubuna ait ses kayıtlarının toplam süresi 461,10 saniye (yaklaşık 7,68 dakika) ve üniversite yaş grubuna ait ses kayıtlarının toplam süresi 396,59 saniye (yaklaşık 6,61 dakika) olup, toplamda tüm yaş grupları için 1735,58 saniye (yaklaşık 28,91 dakika) uzunluğunda ses kaydı analiz edilmiştir.

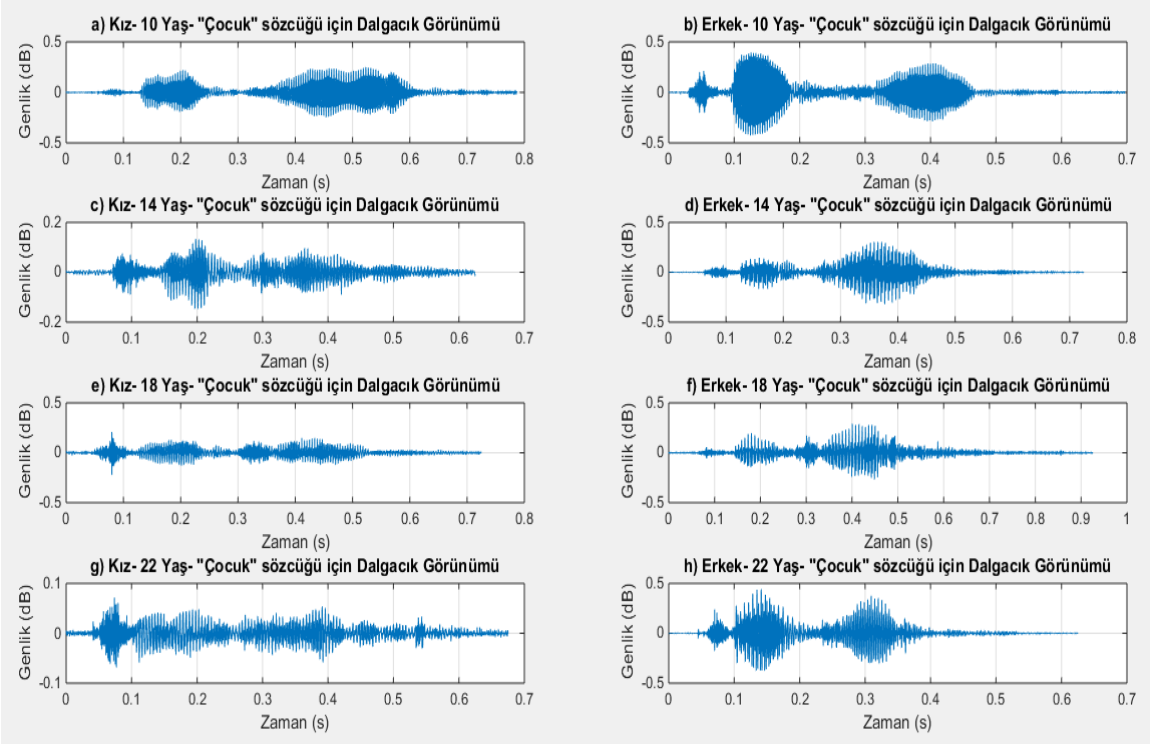
İlkokul yaş grubuna ait en kısa ses kaydının süresi 0,406 saniye, en uzun ses kaydının süresi 2,315 saniye ve her bir kaydın ortalama süresi 0,928 saniyedir. Ortaokul yaş grubuna ait en kısa ses kaydının süresi 0,36 saniye, en uzun ses kaydının süresi 1,225 saniye ve her bir kaydın ortalama süresi 0,785 saniyedir. Lise yaş grubuna ait en kısa ses kaydının süresi 0,475 saniye, en uzun ses kaydının süresi 1,468 saniye ve her bir kaydın ortalama süresi 0,9 saniyedir. Üniversite yaş grubuna ait en kısa ses kaydının süresi 0,425 saniye, en uzun ses kaydının süresi 1,225 saniye ve her bir kaydın ortalama süresi 0,774 saniyedir. Yaş gruplarına göre ses kayıt uzunluğunun istatiksel dağılımı aşağıdaki tabloda (Tablo 4.14) verilmiştir.

**Tablo 4. 14. Yaş gruplarına göre ses kayıt sürelerinin istatiksel bilgisi**

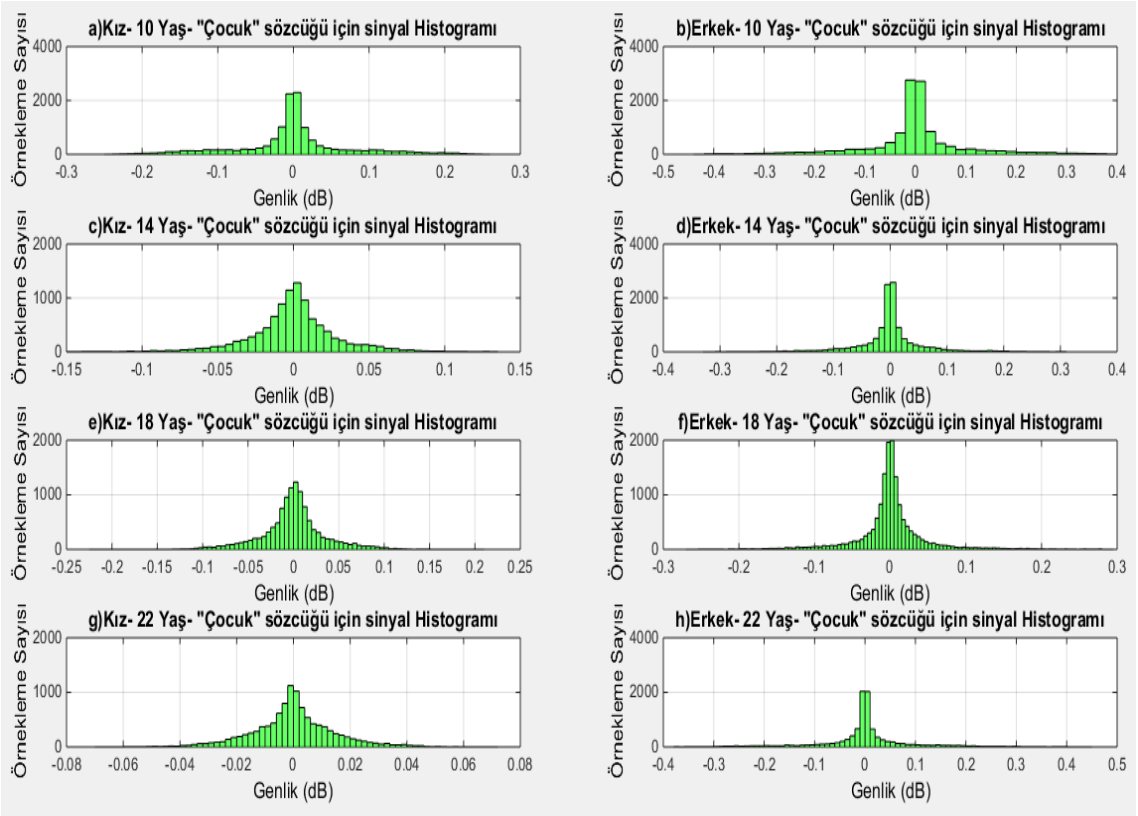
| | Ses Kayıtlarının Toplam Süresi (s) | En Kısa Ses Kaydının Süresi (s) | En Uzun Ses Kaydının Süresi (s) | Ses Kayıtlarının Ortalama Süresi (s) |
|---|---|---|---|---|
| 4. Sınıf | 475,58 | 0,4 | 2,31 | 0,92 |
| 8. Sınıf | 402,29 | 0,36 | 1,22 | 0,78 |
| 12. Sınıf | 461,10 | 0,47 | 1,46 | 0,9 |
| Üniversite | 396,59 | 0,42 | 1,22 | 0,77 |

### 4.3.1. Yaş grubuna göre akustik karşılaştırma

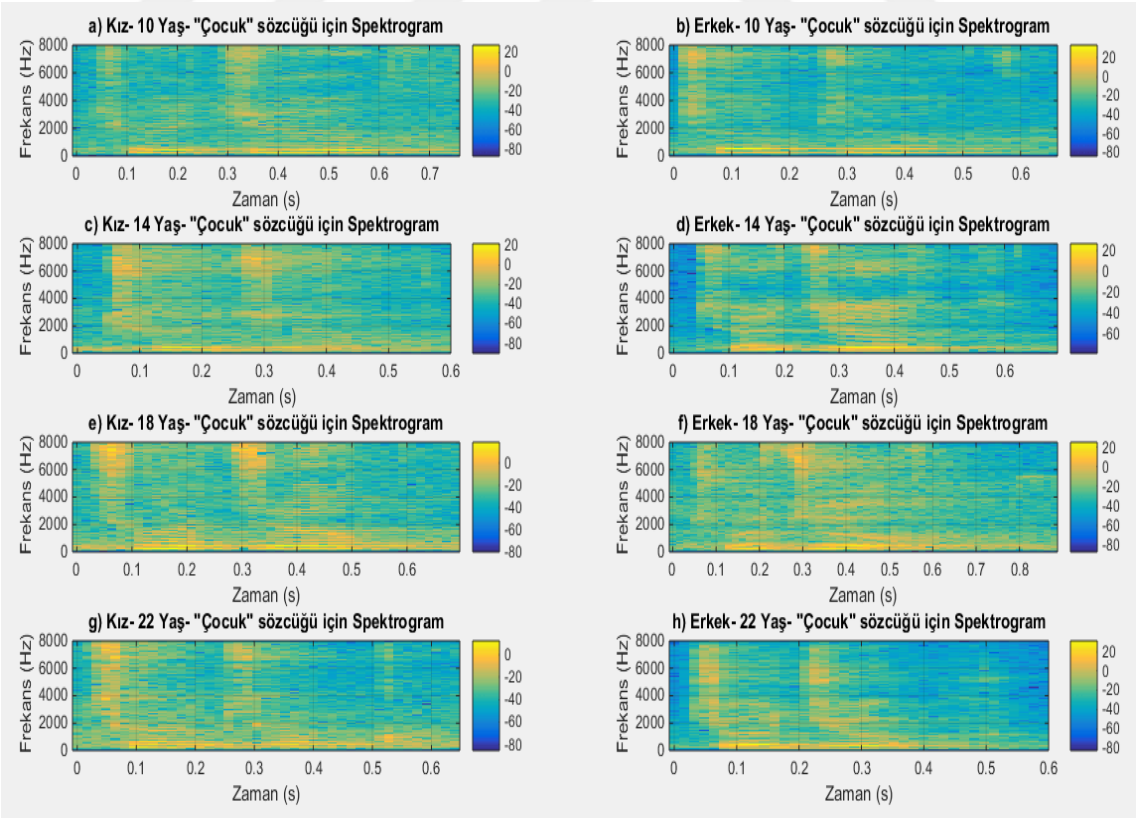
Bölüm 4.1 de belirtilen ve 32 adet kelimenin kendi içinde sınıflandırılması sonucu; sınıflandırmada en çok tanınan kelimelerden biri olan "Çocuk" ve sınıflandırmada en az tanınan kelimelerden biri olan "Hazırlanmak" kelimeleri için Matlab programı yardımı ile elde edilen akustik görüntüleme yöntemlerinden olan sesin dalga görünümü (osilogram) (Şekil 4.5 ve Şekil 4.9), ve sinyal histogramı (Şekil 4.6 ve Şekil 4.10) ve sesin süre, frekans ve şiddet özelliklerini bir arada inceleyen spektrogram (Şekil 4.7 ve Şekil 4.11) ve sese ait enerjinin daha net görünümü adına spektral enerjinin sıfırdan küçük kısımlarının silindiği gösterimler (Şekil 4.8 ve Şekil 4.12) aşağıda verilmiştir.



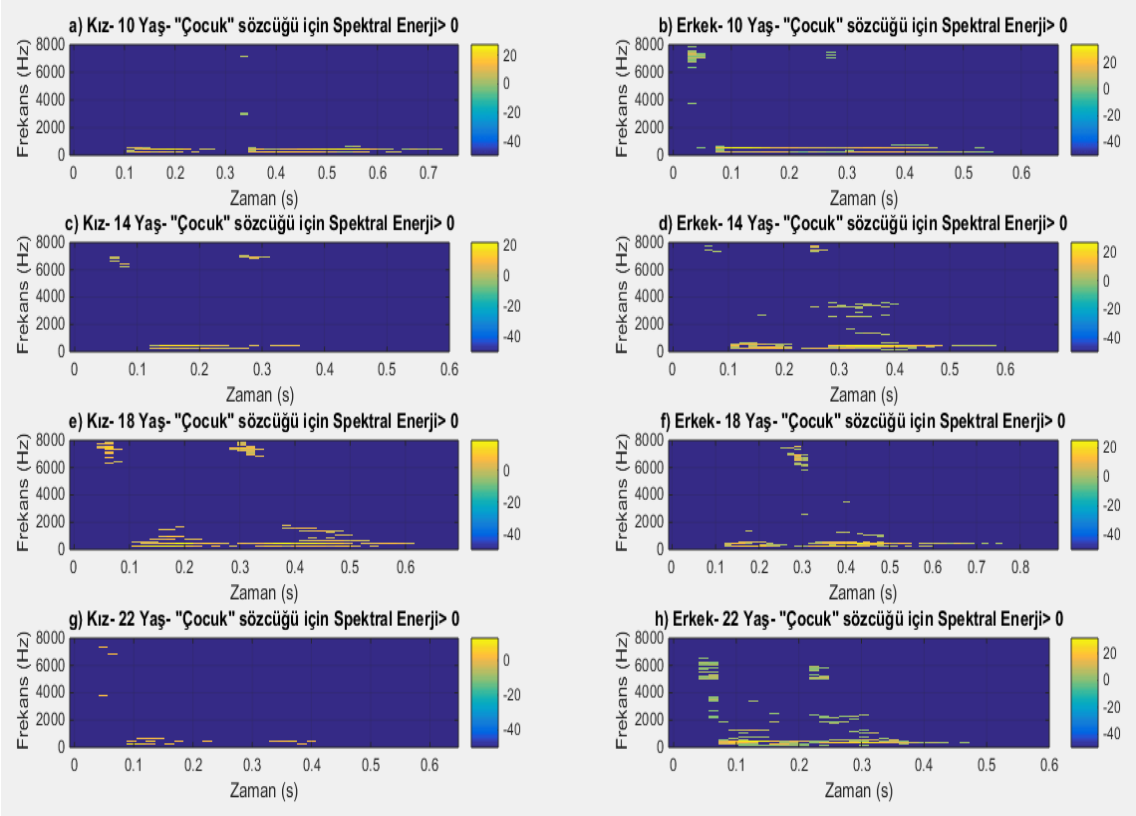

**Şekil 4. 5. "Çocuk" sözcüğü için 4 farklı yaş grubundan 8 kişiye ait sesin dalga görünümü**



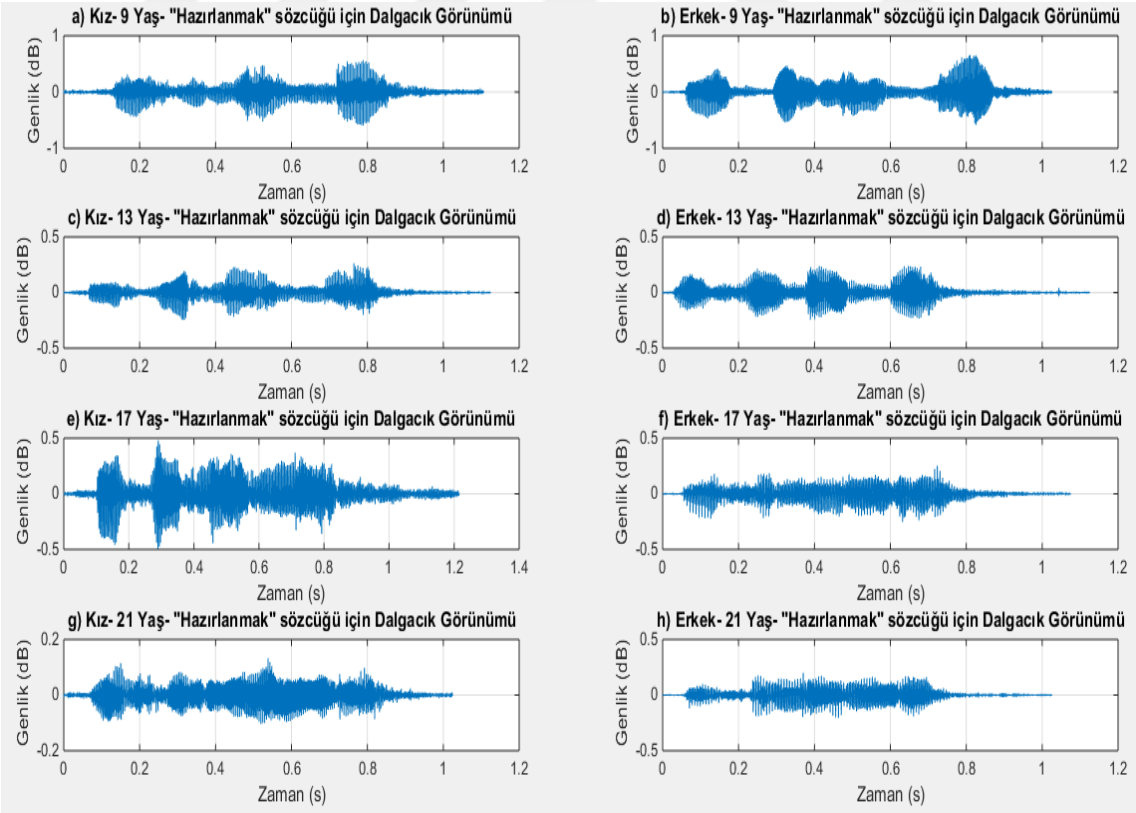

**Şekil 4. 6. "Çocuk" sözcüğü için 4 farklı yaş grubundan 8 kişiye ait sesin histogramı**



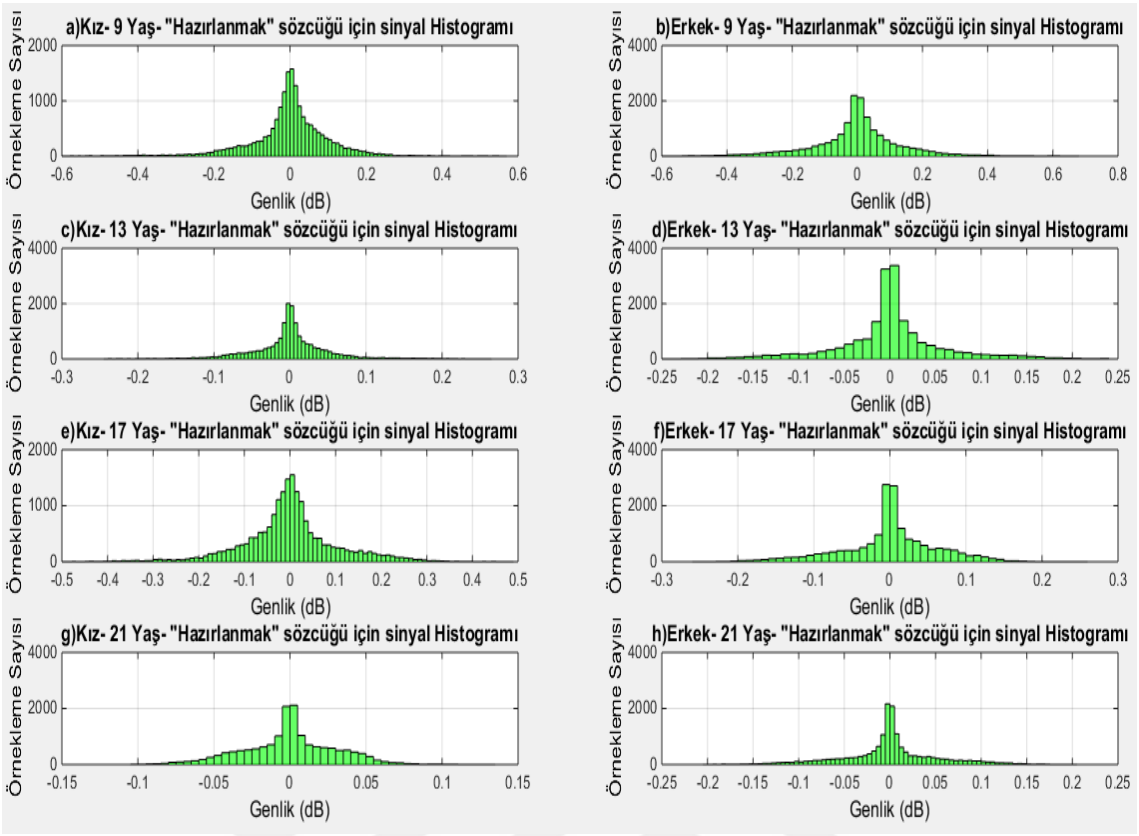

**Şekil 4. 7. "Çocuk" sözcüğü için 4 farklı yaş grubundan 8 kişiye ait sesin spektrogramı**



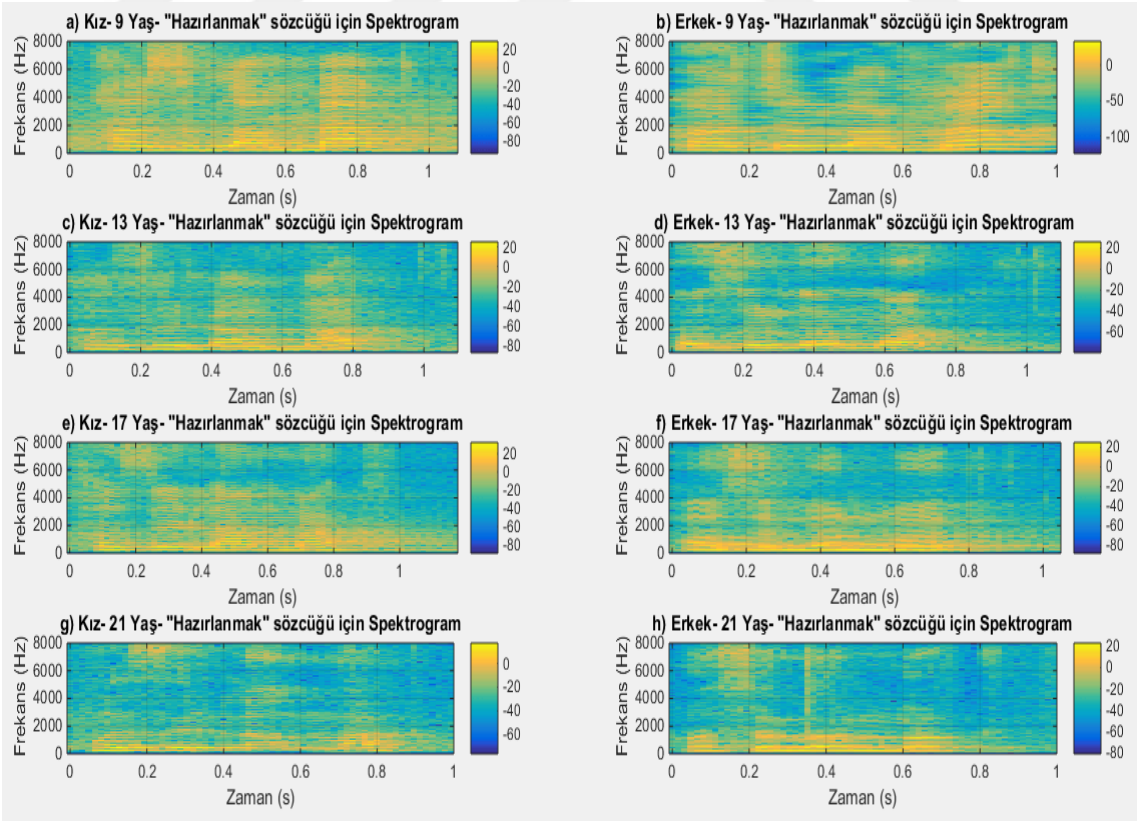

**Şekil 4. 8. "Çocuk" sözcüğü için 4 farklı yaş grubundan 8 kişiye ait spektral enerji>0 gösterimi**



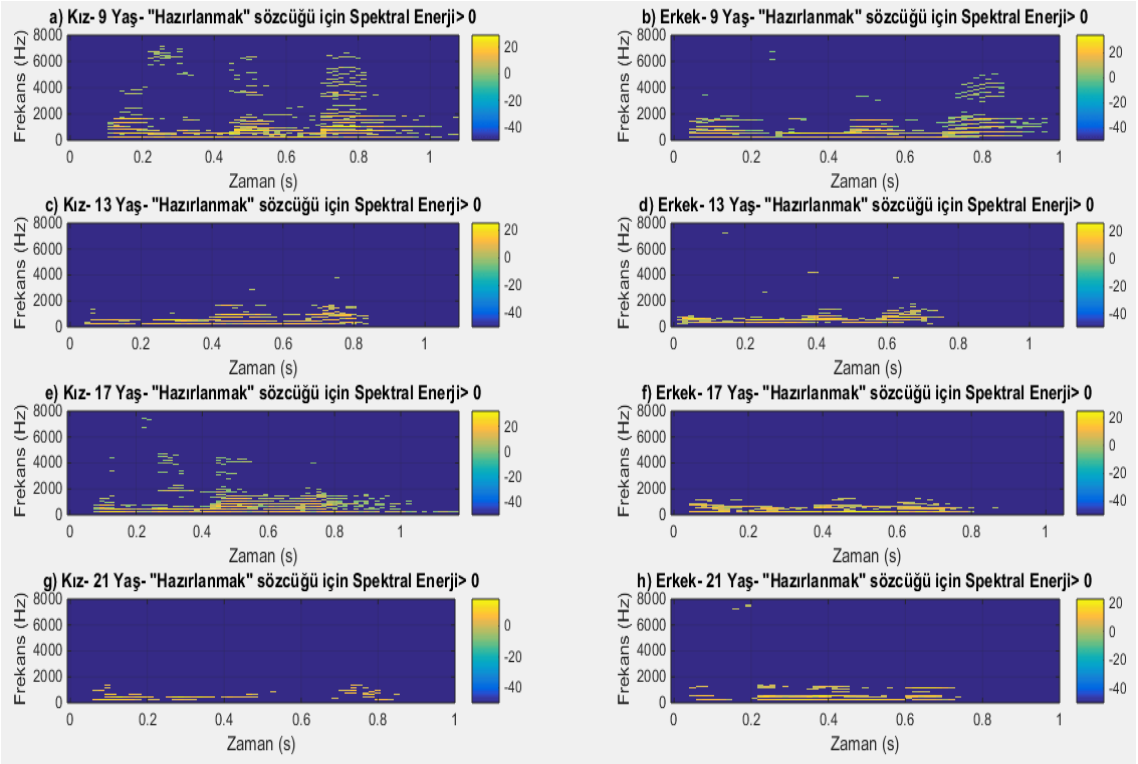

**Şekil 4. 9. "Hazırlanmak" sözcüğü için 4 farklı yaş grubundan 8 kişiye ait sesin dalga görünümü**




**Şekil 4. 10. "Hazırlanmak" sözcüğü için 4 farklı yaş grubundan 8 kişiye ait sesin histogramı**




**Şekil 4. 11. "Hazırlanmak" sözcüğü için 4 farklı yaş grubundan 8 kişiye ait sesin spektrogramı**




**Şekil 4. 12. "Hazırlanmak" sözcüğü için 4 farklı yaş grubundan 8 kişiye ait spektral enerji>0 gösterimi**

Şekil 4.5, Şekil 4.6, Şekil 4.7 ve Şekil 4.8 içerisinde; 10, 14, 18 ve 22 yaşlarında her biri 1 erkek ve 1 kız öğrenciye ait olan ve kelime grupları içerisinde tanınma frekansı yüksek olan "Çocuk" sözcüğüne ait sekiz ses kaydının verileri görsel akustik karşılaştırmalar şeklinde verilmiştir. Şekil 4.9, Şekil 4.10, Şekil 4.11 ve Şekil 4.12 içerisinde; 9, 13, 17 ve 21 yaşlarında her biri 1 erkek ve 1 kız öğrenciye ait olan ve kelime grupları içerisinde tanınma frekansı düşük olan "Hazırlanmak" sözcüğüne ait sekiz ses kaydının verileri görsel akustik karşılaştırmalar şeklinde verilmiştir

Şekil 4.5 ve Şekil 4.9 kısımlarında sekiz ses için dalgacık görünümü; sesin genliği ve zaman ekseninde karşılaştırılmış olup, Şekil 4.6 ve Şekil 4.10 kısımlarında ise sekiz ses için sesin örnekleme sayısı ve genlik yoğunluğu histogram şeklinde verilmiştir. Şekil 4.7 ve Şekil 4.11 kısımlarında frekans-zaman ekseninde sekiz sese ait şiddetin spektrogramı görsel olarak verilmiştir. Bu görüntüler sinyale Hızlı Fourier dönüşümü (Fast Fourier transform, FFT) uygulanarak elde edilmiştir. Şekil 4.8 ve Şekil 4.12 kısımlarında ise karşılaştırılan iki sesin daha net anlaşılabilmesi adına spektral enerjinin sıfırdan küçük kısımları silinmiş ve görsel olarak verilmiştir.

### 4.2.2. Yaş grubuna göre sınıflandırmada başarı oranları

Bu kısımda MFCC, LPCC ve MF&LP karışımı için farklı katsayılarda k en yakın komşu (KNN), yapay sinir ağları (YSA) ve destek vektör makineleri (DVM) ile sınıflandırma yapılmıştır. Yukarıdaki üç öznitelik çıkarımı katsayısının; bahsedilen üç sınıflandırma metodu ile sınıflandırılması sonucu elde edilen karışıklık matrisleri ve performans oranları aşağıda verilmiştir.

MFCC yöntemi için farklı katsayılarla yapılan denemeler sonucu elde edilen başarı oranları aşağıda (Tablo 4.15) verilmiştir.

**Tablo 4. 15. Yaş grubuna göre ayrıştırmada MFCC katsayıları için başarı oranları**

| KATSAYI ADEDİ | KNN (%) | DVM (%) | YSA (%) |
|---|---|---|---|
| 10 | 68,5 | 73,4 | 66 |
| 20 | 85,8 | 89 | 77,4 |
| 30 | 86,7 | 89,1 | 77,6 |
| **40** | 87,1 | **90,1** | 79,2 |
| 50 | 87,5 | 89,8 | 80,9 |
| 60 | 87 | 89,4 | 82,3 |
| 70 | 87 | 89,7 | 83,2 |
| 80 | 86,9 | 89,5 | 83,5 |
| 90 | 87,5 | 89,2 | 84,9 |
| 100 | 87 | 89,8 | 83,8 |

Yapılan denemeler sonucu en yüksek sınıflandırma başarı oranı %90,1 olarak gözlenmiştir. Böylece; 2048 ses örneği içerisinde 1845 adet ses doğru sınıflandırılmıştır. 40 MFCC katsayı kullanılarak bu başarı elde edilmiştir. Sınıflandırma için başarı oranı en yüksek metot DVM olmuştur. 40 MFCC katsayısının sınıflandırılması sonucu elde edilen karışıklık matrisi (Tablo 4.16) ve performans oranı (Tablo 4.17) aşağıda verilmiştir.

**Tablo 4. 16. Yaş grubu- MFCC ile elde edilen karışıklık matrisi**
(Y1: İlkokul, Y2: Ortaokul, Y3: Lise, Y4: Üniversite)

| Aktivite | Y1 | Y2 | Y3 | Y4 |
|---|---|---|---|---|
| Y1 | **480** | 11 | 16 | 5 |
| Y2 | 6 | **440** | 29 | 37 |
| Y3 | 8 | 32 | **454** | 18 |
| Y4 | 3 | 28 | 10 | **471** |

**Tablo 4. 17. Yaş grubu- MFCC ile elde edilen performans sonuçları**

| Sınıflar | Doğru pozitiflerin oranı(TP Rate) | Yanlış negatiflerin oranı(FP Rate ) | Kesinlik (Precision) | Duyarlılık (Recall) | F-ölçütü (F-Measure) |
|---|---|---|---|---|---|
| Y1 | 0,938 | 0,011 | 0,966 | 0,938 | 0,951 |
| Y2 | 0,859 | 0,046 | 0,861 | 0,859 | 0,86 |
| Y3 | 0,887 | 0,036 | 0,892 | 0,887 | 0,889 |
| Y4 | 0,92 | 0,039 | 0,887 | 0,92 | 0,903 |
| **Ortalama** | **0,901** | **0,033** | **0,901** | **0,901** | **0,901** |

LPCC yöntemi için farklı katsayılarla yapılan denemeler sonucu elde edilen başarı oranları aşağıda (Tablo 4.18) verilmiştir.

**Tablo 4. 18. Yaş grubuna göre ayrıştırmada LPCC katsayıları için başarı oranları**

| KATSAYI ADEDİ | KNN (%) | DVM (%) | YSA (%) |
|---|---|---|---|
| 10 | 59 | 61,9 | 51 |
| 20 | 76,3 | 77,8 | 62,6 |
| 30 | 83,1 | 84,2 | 70,4 |
| 40 | 86,9 | 88,2 | 77,4 |
| 50 | 91,3 | 90,5 | 84,1 |
| 60 | 93,3 | 92,7 | 86,3 |
| 70 | 93,9 | 93,9 | 88,4 |
| **80** | 93,7 | **94,3** | 89,3 |
| 90 | 94 | 93,9 | 90,3 |
| 100 | 94,1 | 94,3 | 91,2 |

Yapılan denemeler sonucu en yüksek sınıflandırma başarı oranı %94,3 olarak gözlenmiştir. Böylece; 2048 ses örneği içerisinde 1931 adet ses doğru sınıflandırılmıştır. 80 LPCC katsayısı kullanılarak bu başarı elde edilmiştir. Sınıflandırma için başarı oranı en yüksek metot DVM olmuştur. 80 LPCC katsayısının sınıflandırılması sonucu elde edilen karışıklık matrisi (Tablo 4.19) ve performans oranı (Tablo 4.20) aşağıda verilmiştir.

**Tablo 4. 19. Yaş grubu- LPCC ile elde edilen karışıklık matrisi**
(Y1: İlkokul, Y2: Ortaokul, Y3: Lise, Y4: Üniversite)

| Aktivite | Y1 | Y2 | Y3 | Y4 |
|---|---|---|---|---|
| Y1 | **490** | 11 | 3 | 8 |
| Y2 | 6 | **489** | 12 | 5 |
| Y3 | 8 | 18 | **479** | 7 |
| Y4 | 11 | 12 | 16 | **473** |

**Tablo 4. 20. Yaş grubu- LPCC ile elde edilen performans sonuçları**

| Sınıflar | Doğru pozitiflerin oranı(TP Rate) | Yanlış negatiflerin oranı(FP Rate ) | Kesinlik (Precision) | Duyarlılık (Recall) | F-ölçütü (F-Measure) |
|---|---|---|---|---|---|
| Y1 | 0,957 | 0,016 | 0,951 | 0,957 | 0,954 |
| Y2 | 0,955 | 0,027 | 0,923 | 0,955 | 0,939 |
| Y3 | 0,936 | 0,02 | 0,939 | 0,936 | 0,937 |
| Y4 | 0,924 | 0,013 | 0,959 | 0,924 | 0,941 |
| **Ortalama** | **0,943** | **0,019** | **0,943** | **0,943** | **0,943** |

MF&LP öz nitelik karışımı yöntemi için farklı katsayılarla yapılan denemeler sonucu elde edilen başarı oranları aşağıda (Tablo 4.21) verilmiştir.

**Tablo 4. 21. Yaş grubuna göre ayrıştırmada MF&LP katsayıları için başarı oranları**

| TOPLAM KATSAYI: MF&LP | KNN (%) | DVM (%) | YSA (%) |
|---|---|---|---|
| 10 | 57 | 65,4 | 60,4 |
| 20 | 75,9 | 79,4 | 71,8 |
| 30 | 84 | 89,5 | 82 |
| 40 | 88,1 | 92,6 | 86,1 |
| 50 | 90,5 | 92,7 | 88,5 |
| 60 | 91,5 | 94 | 91 |
| 70 | 92,4 | 94,2 | 91,6 |
| 80 | 93,2 | 94,4 | 91,8 |
| **90** | 94,1 | **96** | 93,3 |
| 100 | 93,8 | 95,8 | 93,1 |

Yapılan denemeler sonucu en yüksek sınıflandırma başarı oranı %96 olarak gözlenmiştir. Böylece; 2048 ses örneği içerisinde 1967 adet ses doğru sınıflandırılmıştır. 90 MF&LP (45 adet MFCC ve 45 adet LPCC) katsayısı kullanılarak

bu başarı elde edilmiştir. Sınıflandırma için başarı oranı en yüksek metot DVM olmuştur. 90 MF&LP katsayısının sınıflandırılması sonucu elde edilen karışıklık matrisi (Tablo 4.22) ve performans oranı (Tablo 4.23) aşağıda verilmiştir.

**Tablo 4. 22. Yaş grubu- MF&LP ile elde edilen karışıklık matrisi**

| Aktivite | Y1 | Y2 | Y3 | Y4 |
|---|---|---|---|---|
| Y1 | **489** | 8 | 12 | 3 |
| Y2 | 5 | **490** | 10 | 7 |
| Y3 | 1 | 12 | **493** | 6 |
| Y4 | 2 | 9 | 6 | **495** |

**Tablo 4. 23. Yaş grubu- MF&LP ile elde edilen performans sonuçları**

| Sınıflar | Doğru pozitiflerin oranı(TP Rate) | Yanlış negatiflerin oranı(FP Rate ) | Kesinlik (Precision) | Duyarlılık (Recall) | F-ölçütü (F-Measure) |
|---|---|---|---|---|---|
| Y1 | 0,955 | 0,005 | 0,984 | 0,955 | 0,969 |
| Y2 | 0,957 | 0,019 | 0,944 | 0,957 | 0,951 |
| Y3 | 0,963 | 0,018 | 0,946 | 0,963 | 0,955 |
| Y4 | 0,967 | 0,01 | 0,969 | 0,967 | 0,968 |
| **Ortalama** | **0,96** | **0,013** | **0,961** | **0,96** | **0,961** |

Aşağıda (Tablo 4.24) tüm özellik çıkarım yöntemleri için elde edilen başarı oranı ve başarılı olunan sınıflandırma metodunun yanı sıra Matlab ’da öz nitelik çıkarımı için geçen zaman ve hazırlanan veri setlerinin Weka yardımı ile sınıflandırılırken aldığı zaman özetlenmiştir.

**Tablo 4. 24. Yaş grubu için elde edilen en başarılı performans/zaman sonuçları**

| Özellik Çıkarım Metodu | Katsayı Adedi | Sınıflandırma Yöntemi | Başarı Oranı (%) | Öz Nitelik Çıkarımı İçin Geçen Zaman (saniye) | Sınıflandırma İçin Geçen Zaman (saniye) |
|---|---|---|---|---|---|
| MFCC | 40 | DVM | 90,1 | 41,3 | 5,9 |
| LPCC | 80 | DVM | 94,3 | 92,3 | 11,4 |
| MF&LP | 90 | DVM | 96 | 114,2 | 11,8 |

Matlab’da öz nitelik çıkarımı (shuffle, normalizasyon ve veri seti haline getirme süreleri de dâhil) için geçen süreler de yukarıda belirtilmiştir.

Kullanılan sınıflandırma yöntemleri içinde en başarılı sonuçlar her seferinde DVM öğrenme yöntemi ile olmuştur. Sonuç itibari ile yaş grubu sınıflandırmada en yüksek başarı oranı %96 ile MF&LP öz nitelik çıkarımı yönteminde görülmektedir. Bunun için 90 katsayı kullanılmıştır. Sonrasında LPCC için %94,3 ve MFCC için %90,1 başarı oranı gözlemlenmiş ve sırasıyla 80 ve 40 katsayı kullanılmıştır.

## 5. TARTIŞMA

Bu tezin amacı, konuşma uygulamaları için cinsiyet ve yaş tanıma sistemi oluşturmaktır. Bu tür bir sistemi inşa etmek için gerekli adımlar açıklanmıştır. Bu adımlar arasında ön sinyal işleme teknikleri, konuşma özelliği çıkarma teknikleri ve son olarak sınıflandırma algoritması yer almaktadır. Bu çalışma için üç farklı öz nitelik çıkartım metodu önerilmiştir. İlk metot ile MFCC denenmiş, ikinci modelde konuşma özellikleri olarak LPCC alınmış ve son modelde MFCC ve LPCC özellikleri bir araya getirilerek oluşturulan MF&LP karışım modeli önerilmiştir. Sistemin performansını ölçmek için farklı katsayılar ve KNN, DVM ve YSA sınıflandırıcıları ile çeşitli testler yapılmıştır.

Cinsiyet tahmini için en yüksek başarı oranı %94,6 ile MF&LP karışımı öz nitelik çıkarımı yönteminde görülmüştür. Bunun için 80 katsayı kullanılmıştır. Sınıflandırma için başarı oranı en yüksek metot KNN olmuştur. Sonrasında LPCC için KNN sınıflandırma metodu ile %94,4 ve MFCC için DVM sınıflandırma metodu %92,3 başarı oranı gözlemlenmiş ve sırasıyla 80 ve 30 katsayı kullanılmıştır.

Yaş tahmini için kullanılan sınıflandırma yöntemleri içinde en başarılı sonuçlar her seferinde DVM sınıflandırma yöntemi ile olmuştur. Sonuç itibari ile yaş grubu sınıflandırmada en yüksek başarı oranı %96 ile MF&LP öz nitelik çıkarımı yönteminde görülmüştür. Bunun için 90 katsayı kullanılmıştır. Sonrasında LPCC için %94,3 ve MFCC için %90,1 başarı oranı gözlemlenmiş ve sırasıyla 80 ve 40 katsayı kullanılmıştır.

Test sonuçlarına göre MF&LP karışım modelinin konuşma özelliği olarak hem cinsiyet hem de yaş tahmini için sırasıyla %94,6 ve %96 başarı oranları ile en iyi performansı gösterdiği görülmüştür. Ayrıca genel olarak zorluk derecesi daha yüksek olan çocuk seslerinde de; cinsiyet ayrımı başarılı bir şekilde gerçekleştirilmiştir.

Bu tez çalışmasını önemli kılan detaylardan biri veri setinin özgün ve tamamen Türkçe olmasıdır. Ayrıca seçilen yaş aralığının küçük ve birbirine yakın olması sınıflandırma ve başarı oranı için bir olumsuzluk gösterebilmesine rağmen elde edilen başarı oranı önemlidir. Bir diğeri de; ses kaydetme frekansının normal koşulların altında bir örnekleme oranı olan 16 kHz ile örneklenmesi ve daha düşük kalitede en iyi sonuçların elde edilmeye çalışılmasıdır.

Analiz edilen ses kategorileri ve ses veri setleri her çalışmada farklı olduğu için bu çalışmadaki başarı oranlarını diğer çalışmalarla kıyaslamak aldatıcı sonuçlar verebilir. Çalışmalar kendi içinde incelendiğinde literatür taramasında bir çok çalışmada cinsiyet bilgisi başarı oranı yaş grubu tespiti başarı oranından daha yüksek bulunmaktadır. Yukarıda verilen örnekler içinde bu duruma aykırı çalışmalar Müller ve ark. (2003) ve Kim ve ark. (2007) için geçerlidir. Bu çalışmada yaş için elde edilen sınıflandırma başarı oranı da cinsiyet tespitinden daha başarılı sonuçlar vererek yukarıdaki duruma bir emsal teşkil etmektedir.

Elde edilen performans değerleri başarılı olarak kabul edilebilir. Ancak, önerilen yöntemin hâlihazırdaki performansını artırmak için bazı ek adımlar uygulanabilir. Her şeyden önce, ses veri tabanı boyutu daha kapsamlı hale getirilebilir. Böylece eğitim verilerinin boyutu çok daha büyük olduğundan, sistem performansı daha uygulanabilir hale getirilebilir. Boyut indirgeme yöntemleri daha iyi analiz edilip, eğitim aşamalarında işleme hızını artıracak metotlar denenmeli ve geliştirilmelidir.

Ayrıca mevcut uygulamada en küçük konuşmacının yaşı 9, en büyük konuşmacının yaşı ise 23 iken; ses veri tabanı daha sağlam, etiketli ve dengeli bir veri tabanı ile daha geniş bir yaş aralığı için sınıflandırma yapılabilir, böylece sistem daha efektif ve uygulanabilir hale getirebilir. Bunun için genellikle kullanılan yabancı ses veri tabanı setleri ve bireysel çabalarla elde edilen ve toplanan veri setleri yerine daha geniş ve Türkçe diline ait hem metinsel hem de rastgele konuşma verileri içeren yerel bir ses külliyatı (corpus) ihtiyacı mevcuttur.

## KAYNAKLAR


Selen, N., 1979. Söyleyiş sesbilimi, akustik sesbilim ve Türkiye Türkçesi, Türk Dil Kurum Yayınları.

Karasartova, S., 2011. Metinden bağımsız konuşmacı tanıma sistemlerinin incelenmesi ve gerçekleştirilmesi, Yüksek Lisans Tezi, Ankara Üniversitesi Fen Bilimleri Enstitüsü, Ankara.

Shivaji, J.C., Ramesh, K., 2015. Automatic speaker age estimation and gender dependent emotion recognition. International Journal of Computer Applications. 117. 5-10. 10.5120/20644-3383.

Mysak, E.D., 1959. Pitch and duration characteristics of older males. Journal of Speech, Language, and Hearing Research, vol. 2, pp. 46–54.

Schötz, S., 2001. A perceptual study of speaker age. Working Papers, Lund University, Dept. of Linguistics and Phonetics; Vol. 49.

Minematsu, N., 2002. Automatic estimation of one's age with his/her speech based upon acoustic modeling techniques of speakers. Proc. ICASSP, pp.137–140.

Müller, C.A., Wittig, F., Baus, J., 2003. Exploiting speech for recognizing elderly users to respond to their special needs. INTERSPEECH.

Nisimura, R., Lee, A., Saruwatari, H., Shikano, N., 2004. Public speech-oriented guidance system with adult and child discrimination capability. Proc. ICASSP2004, vol. 1, pp. 433-436.

Bocklet, T., Maier, A., Nöth, E., 2008. Age determination of children in preschool and primary school age with GMM-based super vectors and support vector machines/regression. Text, Speech and Dialogue.

Lee, M. W., Kwak, K. C., 2012. Performance comparison of gender and age group recognition for human-robot interaction. International Journal of Advanced Computer Science & Applications.

Erokyar, H., 2014. Age and gender recognition for speech applications based on support vector machines. Graduate Theses and Dissertations.

Keklik, S., 2011. Türkçe'de on bir yaşına kadar çocuklara öğretilmesi gereken, birleşim gücü yüksek ilk bin kelime, Odu Sosyal Bilimler Araştırmaları Dergisi,Cilt: 2 Sayı: 4

Schötz, S., 2007. Acoustic analysis of adult speaker age. Speaker Classification I. Springer Berlin Heidelberg.

Metze, F., Ajmera, J., Englert, R., Bub, U., 2007. Comparison of four approaches to age and gender recognition for telephone applications. Acoustics, Speech and Signal Processing.

Kim, H., Bae, K., Yoon, H., 2007. Age and gender classification for a home-robot service. Proc. 16th IEEE International Symposium on Robot and Human Interactive Communication, pp. 122–126.

Sedaaghi, M. H., 2009. A comparative study of gender and age classification in speech signals. Iranian Journal of Electrical & Electronic Engineering.

Feld, M., Burkhardt, F., Müller, C.A., 2010. Automatic speaker age and gender recognition in the car for tailoring dialog and mobile services. INTERSPEECH.

Nguyen, P., Tran, D., Huang, X., Sharma, D., 2010. Automatic classification of speaker characteristics. Communications and Electronics ICCE.

Meinedo, H., Trancoso, I,. 2010. Age and gender classification using fusion of acoustic and prosodic features. Proc. INTERSPEECH, pp. 2818-2821.

Bocklet, T., Stemmer, G., Zeissler, V., Nöth, E., 2010. Age and gender recognition based on multiple systems early vs. late fusion. In Proceedings of Interspeech, pp. 2830–2833.

Dobry, G., Hecht, R. M., Avigal, M., Zigel, Y. 2011. Super vector dimension reduction for efficient speaker age estimation based on the acoustic speech signal. Audio, Speech, and Language Processing.

Bahari, M., Hamme, V., 2011. Speaker age estimation and gender detection based on supervised non-negative matrix factorization. 1 - 6. 10.1109/BIOMS.

Mirhassani, S. M., Zourmand, A., Ting, H. N., 2014. Age estimation based on children's voice: a fuzzy-based decision fusion strategy. The Scientific World Journal, vol. 2014, Article ID 534064, 9 pages, 2014. doi:10.1155/2014/534064

Waller, S., Eriksson, M., Sörqvist, P., 2015. Can you hear my age? Influences of speech rate and speech spontaneity on estimation of speaker age. Front. Psychol. 6:978

Yücesoy, E., Nabiyev, V., 2016. Konuşmacı yaş ve cinsiyetinin Gkm süper vektörlerine dayalı bir Dvm sınıflandırıcısı ile belirlenmesi. Gazi Üniversitesi Mühendislik-Mimarlık Fakültesi Dergisi. 31. 10.17341/gummfd.71595.

Rabiner, L.R., Juand, B. H., 1993. Fundamentals of Speech Recognition, Prenticehall, Englewood Cliffs, N.J., ISBN: 0-13-015157-2.

Kaushal, K., Mistry, S. S., 2016. Comparative study of feature extraction techniques for speech recognition system. IJIRSET, Vol.5, Issue 10

Chandra, E., Manikandan, K., Sivasankar, M., 2014. A proportional study on feature extraction method in automatic speech recognition system, IJIRSET, Vol. 2, Issue 1, pp. 772-775

Picone, J., 1993. Signal modeling techniques in speech recognition, Proceedings of the IEEE, 819, 1215–1247.

Deller, R., Hansen, P., 2000. Discrete-time processing of speech signals, IEEE Press, Piscataway, N.J.

Këpuska, V., Elharati, H., 2015. Robust speech recognition system using conventional and hybrid features of mfcc, lpcc, plp, rasta-plp and hidden markov model classifier in noisy conditions, Journal of Computer and Communications, 3, 1-9.

Hermansky, H., 1990. Perceptual linear predictive coding analysis of speech, Journal of the Acoustic Society of America, 87 4: 1738-1752.

Hermansky, H., Morgan, N., 1994. RASTA processing of speech. IEEE Transactions on Speech and Audio Processing, vol. 2, no. 4, pp. 578-589.

Witten, I. H., Frank, E., 2005. Data Mining: Practical machine learning tools and techniques. Morgan Kaufmann.

Altman, N. S., 1992. An introduction to kernel and nearest-neighbor nonparametric regression, The American Statistician, 46:3, 175-185

Jivani, A. G., Shah, K., Koul, S., Naik, V., 2016. The adept K-nearest neighbour algorithm - an optimization to the conventional K-nearest neighbour algorithm, TMLAI, vol 4, issue 1

Vapnik, V., 1995. The nature of statistical learning theory, Springer-Verlag, New York.

Schölkopf, B., Sung, K., Burges, C., Girosi, F., Niyogi, P., Poggio, T., Vapnik, V., 1997. Comparing support vector machines with Gaussian kernels to radial basis function classifiers, IEEE Trans. on Signal Processing, 4511, 2758–2765.

Keerthi, S. S., Lin, C., 2003, Asymptotic behaviors of support vector machines with gaussian kernel, Neural Computation, 157, 1667–1689.

Dede, G., 2008. Yapay sinir ağları ile konuşma tanıma, Yayımlanmamış yüksek lisans tezi, Ankara üniversitesi

Minsky, M., Papert, S., 1969. Perceptrons: An introduction to computational geometry. MIT press expanded edition, Cambridge.

Rumelhart, D. E., Hinton, G. E., Williams, R. J., 1986. Learning representations by back propagating errors. Nature, vol. 323, pp. 533-536.

Chokmani, K., Hamilton, S., Ghedira, M. H., Gingras, H., 2008. Comparison of ice-affected streamflow estimates computed using artificial neural networks and multiple regression techniques, Journal of Hydrology ,349,383– 396.

Herve, A., 2010. Normalizing Data,Thousand Oaks, CA: Sage.

## ÖZGEÇMİŞ

## KİŞİSEL BİLGİLER


| | |
|---|---|
| **Adı Soyadı** | : Abdulhalık OĞUZ |
| **Doğum Yeri ve Tarihi** | : Batman - 12.02.1990 |
| **Telefon** | : 90 (488) 215 04 08 /205 |
| **E-posta** | : ahalikoguz@hotmail.com |


## EĞİTİM

| **Derece** | | **Adı, İlçe, İl** | **Bitirme Yılı** |
|---|---|---|---|
| Lise | : | Batman YDAL Lisesi | 2007 |
| Üniversite | : | Çukurova Üniversitesi, Mühendislik Mimarlık Fakültesi, Bilgisayar Mühendisliği | 2012 |

## İŞ DENEYİMLERİ

| **Yıl** | **Kurum** | **Görevi** |
|---|---|---|
| 2012-2014 | Tarım Bakanlığı TGAE, Ankara | Müh. |
| 2014-2014 | TPAO, Ankara | Müh. |
| 2014-Devam Ediyor | TEİAŞ 16. Bölge, Batman | Müh. |

**YABANCI DİLLER** : İngilizce

## AKADEMİK ÇALIŞMALAR

**Bildiri** A. Oğuz, Y. Kaya, (2017). Gender and Age Group Estimation Based On Speech Signals. IATS' 17 8th International Advanced Technologies Symposium