Alman Dili Üzerinde Konuşmacı Cinsiyetinin Otomatik Olarak Belirlenmesi

Çiğdem Bakır
Yıldız Teknik Üniversitesi, Bilgisayar Mühendisliği

Geliş Tarihi: 2016-04-05 Kabul Tarihi: 2016-07-09

Öz

Kişi tanıma sistemleri biyometrik verilerin güvenli bir şekilde iletimini, tasarımını, sınıflandırılmasını gerekli kılmaktadır. Ayrıca konuşmacıların cinsiyeti belirlenerek biyometrik ses işlemlerinde daha başarılı sonuçlar elde edilebilir. Bu çalışmada Almanca ses biçim ve özelliklerine bakılarak konuşmacının cinsiyetinin otomatik olarak tanınması için bir sistem tasarlanması amaçlanmıştır. 50 erkek ve 50 kadından Almanca farklı uzunlukta kelime ve cümle olarak yaklaşık 2658 ses örneği alınmıştır. Bu ses örnekleri tek kelime olduğu gibi birden fazla kelime de içermektedir. Ses örneklerinin öznitelikleri MFCC (Mel Frequency Cepstral Coefficients) kullanılarak elde edilmiştir. Elde edilen ses örneklerinin öznitelik vektörleri Saklı Markov Modelleri(Hidden Markov Models-HMM), Dinamik Zaman Bükmesi(Dynamic Time Warping-DTW) ve Gauss Mixture Models (Gauss Karışım Modeli-GMM) yöntemleri ile eğitilmiştir. Test aşamasında ise eğitilen ses örneklerine bakılarak verilen ses örneğinin cinsiyeti belirlenmeye çalışılmıştır. Ayrıca çalışmada kullanılan tüm sınıflandırma algoritmalarının sonuçları ve performansları karşılaştırılmalı olarak sunulmuştur.

Anahtar Kelimeler: Konuşmacı Cinsiyeti Tanıma Sistemi; Dinamik Zaman Bükmesi; Saklı Markov Modeli.

Automatic Determination of the Speaker on the German Language

Çiğdem Bakır
Yıldız Technical University, Computer Engineering

Abstract

Authentication systems necessitate transmission, design and classification of biometric data in a secure manner. Moreover, in voice process of biometric can be obtained successful results by determining gender of speaker. In this study, the aim was to designed system taking German sound forms and properties for automatic recognition gender of speaker. Approximately 2658 German voice samples of words and clauses with differing lengths have been collected from 50 males and 50 females. This voice samples includes more than one word as a word. Features of these voice samples have been obtained using MFCC (Mel Frequency Cepstral Coefficients). Feature vectors of the voice samples obtained have been trained with such methods as Hidden Markov Model, Dynamic Time Warping and Gauss Mixture Model. In the test phase, gender of a given voice sample has been identified taking the trained voice samples into consideration. Results and performances of the algorithms employed in the study for classification have been also demonstrated in a comparative manner.

***Keywords*:** Speaker Gender Recognition System; Hidden Markov Model; Dynamic Time Warping.

1. Giriş

Günümüzde teknolojinin gelişmesiyle birlikte güvenlik sorunları da açığa çıkmıştır. Bu sorunların başında ise güvenlik gelmektedir. Özellikle kimlik belirleme gibi biyometrik sistemler güvenlik konusunun önemli bir bölümünü oluşturmaktadır. Bu sebeple çeşitli suçlara konu olan ses kayıtlarının adli ses incelemeleri gerekmektedir. Özellikle ticari işlemlerde kişilere ait bilgilerin başka kişilere

*Sorumlu Yazar: Adres: Yıldız Teknik Üniversitesi, Elektrik Elektronik Fakültesi Bilgisayar Mühendisliği Bölümü, Davutpaşa Mah. Davutpaşa Caddesi 34220 Esenler- İstanbul, TÜRKİYE, E-mail: cigdem.bakr@gmail.com

Doi: 10.21541/apjes.24088

geçmesinin engellenmesi için birtakım çalışmalar yapılmıştır. El yazısı tanıma, imza tanıma, yüz tanıma, iris tanıma, ses tanıma bu çalışmaların birkaçını oluşturur [6].

Almanca Hint Avrupa dillerinin Cermen koluna bağlıdır. Dünyada yaklaşık 120 milyon kişi Almanca konuşmaktadır. Ayrıca Almanya uluslar arası anlamda ekonomide, ticarette, sanayide ve diğer birçok alanda önemli bir konuma sahiptir. Bu sebeple Alman dili çok yaygın bir şekilde kullanılmaktadır. Ancak dilin bu kadar yaygınlaşması bu alanda biyometrik verilerde güvenlik problemini ortaya çıkarmaktadır. Bu da güvenli, hızlı bir otomatik ses ve konuşmacı tanıma ihtiyacını gerekli kılmaktadır.
Alman dilinin özelliklerine bakacak olursak; Almanca kök ve eklerden oluşur ve çekimli diller grubuna girer. Latin harfleri kullanılır ve alfabesinde 29 harf bulunur. Almancada her ismin önüne tanımlık getirilir. Telaffuz açısından kelimeler yazıldığı gibi okunur. Ayrıca oluşan çeşitli ses kaymaları ve tonlamalarıyla diğer dillerden ayrılır.

Ses ve konuşmacı tanımak için çeşitli çalışmalar yapılmıştır. Ancak bu çalışmaların çok az kısmında ses sinyaline bağlı olarak cinsiyet belirleme çalışması yapılmıştır. Ses tanıma sistemleri konuşmacıdan bağımsız ve konuşmacıya bağımlı olmak üzere iki kısımda incelenir. Eğitim ve test aşamasında kullanılan ses kaydı konuşmacı ile aynı ise konuşmacıya bağımlı değil ise konuşmacıdan bağımsızdır.

Perry ve çalışma arkadaşları yaşları 4 ile 16 arasında değişen 10 erkek ve 10 kız konuşmacının akustik ses özelliklerine bakarak cinsiyetini belirlemeye çalışmışlardır. Kaynaşmayan iki ünlünün formant frekansı ve temel frekansı ile konuşmacıların yaş grupları arasındaki farklılıkları gözlemlemişlerdir. Formant frekansı sesin karakteristik özellilerini, yaş farklılıklarını ve cinsiyet farklılıkları arasındaki ilişkileri yansıtmıştır. Ancak çalışma İngilizce veritabanı üzerinde gerçekleştirilmiştir[7].

Parris ve Carey konuşmacının cinsiyet belirlemek için metinden bağımsız bir çalışma gerçekleştirmişlerdir. Konuşmacının akustik ses özellikleri ve perde frekansını birleştirerek yeni bir yöntem geliştirmeye çalışmışlardır. OGI veritabanı üzerinde LDA(Linear Discriminant Analysis) özellik çıkarımı yöntemi ile verileri normalize etmişlerdir ve Saklı Markov Modeli ile eğitmişlerdir[11].

Ses ve konuşmacı tanımak için çeşitli çalışmalar yapılmıştır. Jie-Fu ve çalışma arkadaşları yaşları 25 ile 35 arasında değişen 7 erkek ve 5 kadından Çince olarak ses örnekleri alınmıştır[1]. Bu ses örnekleri tonları, ünlüler, ünsüzler ve heceler belirlenmeye çalışılarak sesin kime ait olduğu bulunmaya çalışılmıştır. Ses örnekleri dört frekans bandına bölünmüştür ve her bir frekans bandı analiz edilmiştir. Ancak bu çalışma çok büyük veriler için test edilmemiştir. Ayrıca İngilizce olan benzerlikleri dikkate alınarak gerçekleştirildiğinden istenilen başarıyı tam olarak sağlayamamıştır.

Tokuda ve çalışma arkadaşları Saklı Markov Modeli kullanarak İngilizce konuşma sentezi sistemi geliştirmişlerdir[2]. Bu sistem konuşmacı tanıma için geliştirilmiştir ve sesin özelliğini değiştiren yapıyı belirler. Ancak yapılan çalışmada sentezlenen sesin karakteristik özelliği oldukça düşüktür.

Reynold ve çalışma arkadaşları konuşmacı tanıma sistemlerinin performansını arttırmak amacıyla SuperSID projesini uygulamışladır[3]. Bu projenin amacı konuşmacı sistemlerini geliştirmek, doğruluğunu arttırmak için en uygun özellikleri kullanmaktır. Ancak bu çalışma sesin akustik özelliklerini ve gürültü giderimini tam olarak gerçekleştirmemiştir.

Reynolds ve çalışma arkadaşları Gaussian M,xture Model (GMM) yöntemi kullanılarak konuşmacı tanıma ve doğrulama sistemini gerçeklemeye çalışmışlardır[8]. Sistemde konuşmacı doğrulama olasılık dağılımlarına göre belirlenmeye çalışılmıştır. Bu olasılık dağılımları için farklı 11 hipotez oluşturulmuştur. Çalışmada kullanılan veriler telefon konuşmalarından alınmıştır.

Konuşma haberleşmede önemli bir yere sahiptir. Bu sebeple ses tanıma çalışması yapılmıştır. Ayrıca bu çalışmada güvenlik açığıyla ilgili ses tanıma problemini çözmek için bir simülasyon gerçekleştirilmiştir.Bu simülasyon Matlab programında kodlanarak yapılmıştır. Çünkü Matlab paket programında kod yazılımını sağlayacak hazır toollar mevcuttur.Ancak ses veritabanını oluştururken birtakım zorluklarla karşı karşıya gelinmiştir. Bunların başında farklı kişiler tarafından kelimelerin farklı hızlarda ve farklı telaffuzlarda söylenmesidir. Bunun yanı sıra ses verilerini kaydederken ortamda ve seste meydana gelen gürültünün olması, tonlama etkisi, hece vurgusu gibi sebepler ses tanıma işlemini zorlaştırmaktadır[9].

2. Materyal ve Yöntem

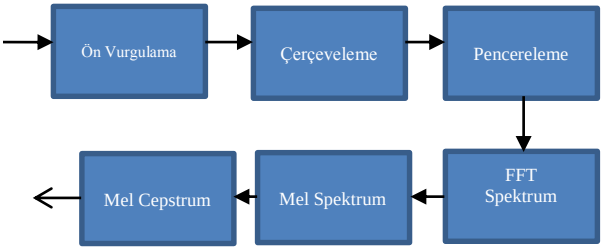
Alman dili ekonomide, sanayide ve ticarette yaygın bir şekilde kullanılır. Bu sebeple bu çalışmada Alman dili üzerinde incelemeler yapılmıştır.

Yapılan çalışmada erkek ve kadınlardan alınan Almanca ses örneklerinden oluşan özgün bir veritabanı üzerinde gerçekleştirilmiştir. Çünkü Almanya ile başta sanayi, ekonomi, ticaret olmak üzere birçok alanda ilişkiler içindeyiz. Bu sebeple Alman dili günümüzde çok önemli bir yere sahiptir. Literatürde yapılan çalışmalar geliştirerek daha önce hiçbir çalışmada kullanılmayan veritabanı üzerinde otomatik olarak gelen kişinin cinsiyeti belirlenmesi açısından önemlidir.

Bu ses örnekleri MFCC ile çeşitli özellik vektörlerine ayrılarak eğitilir. MFCC ile elde edilen özellik vektörü cinsiyet bilgisiyle birlikte veritabanına aktarılır. İkinci aşamada ise kaydedilen ses sinyallerinin özellik vektörlerini DTW, HMM ve GMM gibi sınıflandırma algoritmaları ile eğitilir. Sistem eğitildikten sonra test verisi ile eğitim verisindeki ses sinyallerine bakılarak konuşmacının cinsiyetine karar verilir. Ayrıca konuşma sinyallerinden elde edilen özellik vektörleri DTW, HMM ve GMM ile eğitilerek konuşmacının cinsiyetini tanımadaki sınıflandırma başarısı MFCC-1, MFCC-3 ve MFCC-5 ve MFCC-9 özellik vektörleri için ayrı ayrı hesaplanarak yöntemlerin başarısı karşılaştırılmalı olarak sunulmuştur.

2.1. MFCC (Mel Frequency Cepstral Coefficients)

MFCC(Mel Frequency Cepstral Coefficients) ses işlemede kullanılan bir özellik çıkarımı yöntemidir. Ses verisini alt kümelerine bölerek önemli bilgileri ve özellikleri çıkarmak için kullanılır. MFCC özellik çıkarımı tekniğinin adımları Şekil 1'de gösterilmektedir[10].



Şekil 1. MFCC özellik çıkarımının adımları

MFCC özellik çıkarımı yönteminde iki filtre kullanılır. İlk filtre 1000 Hz'in altındaki frekans değerlerinde lineer; diğer filtre ise 1000 Hz'in üstünde logaritmik dağılıma sahiptir. Ön vurgulama aşaması MFCC özellik vektörünün elde edilmesindeki ilk aşamadır. Bu aşamada yüksek frekansa sahip ses sinyalleri bir filtreden geçirilir. Böylelikle yüksek frekanstta sesin enerjisi arttırılır. Ses sinyalleri analogtur. Çerçeveleme aşamasında ses sinyalleri 20 ile 40 ms arasında küçük çerçevelere bölünerek analogdan sayısala dönüştürülür. N adet çerçeveleye bölünür.

Pencereleme aşamasında ise ses sinyali kaydırılarak hareket ettirilir. Böylelikle kullanılan pencereleme ile gelecek çerçevenin en yakın frekans çizgileri birleştirilir. Pencere türü, genişliği ve kayma miktarı bu aşamada belirlenir. N adet çerçevenin her biri zaman uzayından frekans uzayına FFT(Hızlı Fourier Dönüşümü) ile aktarılır. Frekans uzayında ses sinyallerinin spektral özellikleri gösterilir. Bu spektral özelliklerin ağırlıkları toplamı hesaplanarak mel spektrum elde edilir. Bu mel spektrum üçgen dalgalardan oluşur ve bir dizi süzgeçten geçirilerek oluşturulur. Mel spektrum iki komşu filtrenin frekansını düşürürerek gürültüyü azaltır. Sinyalin logaritması mel spektrum aşamasında alınarak sinyal frekans uzayından tekrar zaman uzayına aktarılır. DCT(Discrete Cosine Transform) kullanılarak zaman uzayında mel frekans cepstrum katsayıları elde edilir.

2.2. Dinamik Zaman Bükmesi (Dynamic Time Warping)

Dinamik Zaman Bükmesi (DTW) zamanla değişen iki konuşma sinyali arasındaki benzerliği hesaplayarak verilen ses sinyalinin hangi konuşmacıya ait olduğunu bulur. Bu yöntem ile iki sinyal arasındaki en optimal zaman eğrisi bulunur. Ayrıca DTW ses, video gibi birçok veri analizinde kullanılan doğrusal bir yöntemdir.

İki sinyal arasındaki yani referans ve test sinyali aynı zaman aralığında olabilmesi için seslendirmeyi zaman içinde yayarak ya da daraltarak birbirine yaklaştırmaya çalışır. Çeşitli benzerlik ölçütleri kullanarak iki sinyal arasındaki benzerlik bulunur. Çünkü aynı sözcük bir seslendirmede uzun diğer bir seslendirmede kısa olabilir. Kısacası; aynı kişinin farklı zamanlarda sözcük seslendirmesi ile ya da kişiden kişiye değişen seslendirmeler zaman içinde farklılıklar gösterebilir. Bu yöntem dinamik programlama ile elde edilir. Benzerlik matrisi bulunduktan sonra maliyet matrisi bulunur ve geriye doğru gidilerek(minimum uzaklık) iki sinyal arasındaki eşleşme bulunmuş olunur.

$$Q= q_1, q_2,\ldots, q_i,\ldots, q_n \quad (1)$$
$$C= c_1, c_2,\ldots, c_j,\ldots, c_m \quad (2)$$

Eşitlik 1 ve eşitlik 2'deki Q ve C iki farklı konuşma sinyalini; **n** ve **m** ise bu konuşma sinyallerinin uzunluklarını gösterir[4]. Bu durumda Q ve C sinyali

arasındaki benzerlik oranı eşitlik 3'deki gibi Öklid uzunluğu kullanılarak hesaplanılır[4].

$$d(q_i,c_j) = (q_i,c_j)^2 \quad (3)$$

Q ve C için (i,j) matrisi oluşturulur. Bu matris kullanılarak birikmiş uzaklıklar matrisi hesaplanılır. D birikmiş maliyet matrisini gösterir ve özyinelemeli olarak hesaplanılır. D(1,1)=d(1,1) başlangıç uzaklığı ile başlar.

$$D(i,j)=\min[D(i-1,j-1),D(i-1,j),D(i,j-1)]+d(i,j) \quad (4)$$

2.3. Gauss Karışım Modeli (Gauss Mixture Model)

Gauss Karışım Modeli bir ya da daha fazla ses sinyalinin gauss dağılımının ağırlık birleşimine dayanan istatiksel bir yöntemdir. Gauss yoğunluğunun ağırlıklı birleşiminin toplamı eşitlik 5'de gösterilmiştir[12].

$$p(x|\lambda) = \sum_{i=1}^{M} p_i\, b_i(x) \quad (5)$$

x özellik vektörünü, **D** boyutlu rastgele vektörü $\boldsymbol{b_i}$ (x), i = 1,…,M yoğunluk bileşenlerini, $\boldsymbol{p_i}$ karışım ağırlığını gösterir.Bu modelin parametreleri Expectation Maximization(EM) algoritması ile bulunur. Eğitim verisindeki tüm sınıflar biribirinden bağımsız Gauss yoğunluk fonksiyonu ile ifade edilir. Karışımı belirleyecek en optimum yoğunluk bileşenleri bulunur. p(x|λ) ifadesini maksimum yapacak gauss model parametreleri bulmak için eşitlik 6 kullanılır[12].

$$p(x|\lambda) = \prod_{t=1}^{T} p(x_t\, I\lambda) \quad (6)$$

2.4. Saklı Markov Modeli (Hidden Markov Model)

Saklı Markov Modelleriyle(HMM) ilgili geçmişten günümüze birçok alanda çalışma yapılmıştır. HMM yüz tanıma, konuşma tanıma, ses tanıma, el yazısı tanıma, vücut hareketleri tanımada, biyoinformatikte, gen tahmininde, kripto-analizde, protein yapısı ve dizilişinde, DNA diziliminde, örüntü tanımada (pattern recognation) yaygın bir şekilde kullanılmıştır. Saklı Markov Modelinde sisteme giriş olarak mevcut durumlar verildiğinde oluşabilecek gelecek durumlar tahmin edilmeye çalışılır. Her çalıştırıldığında farklı bir çıkış değeri üretildiğinden HMM stokastik bir süreçtir. Ayrıca markov modellerinde sistem olasılık dağılımına bağlı olarak kendi durumundan başka bir duruma geçebilir ya da aynı durumda kalabilir..Durumda meydana gelen olasılıklar geçiş olasılıkları olarak adlandırılır. HMM normal markov modelinden farklı olarak durumlar gözlemci tarafından görülmez. Ancak duruma bağlı olan geçişler görülebilir. SMM konuşmacı tanıma sistemlerinde aşağıdaki adımlardan oluşur[5].

Q tane durum olmak üzere S={ S_1, S_2,…, S_Q } oluşan konuşma sinyallerinin mevcut durumunu gösterir.

Başlangıç durum olasılıkları t ayrık zamanda belirlenir. (π= { P_r (S_i |t=0, S_i ε S})

Geçiş olasılıkları mevcut durumlara göre hesaplanılır. (a_{ij} = {P_r (S_j t zamanda t|S_i zamanda t-1), S_i ε S, S_j ε S))
F yani gözlemlenen özellik sayısı belirlenir.

Böylelikle konuşma sinyalinin olasılık dağılımları hesaplanmış olunur.(b_x= {b_i (x)= P_r (x(S_i), S_iε S, x ε F})

Oluşan HMM λ =(a,b,π) ile gösterilir.

Ardışık olarak gelen kısa ses ifadelerini birlikte ele alarak ardına gelebilecek bu kısa kelimeler için bir model oluşturur ve bu modelden yararlanarak uzun süreli sesli ifade kesimlerinin tanınmasını sağlar. Amaç gözlenebilen durumlara karşı gizli olan durumları tahmin etmektir. Durumlar, durumlar arası geçişler ve gözlemler vardır. Ses tanımada gözlemler özellik vektörünü yani ses sinyalini oluşturur. Durumlar ise temel olarak alınan ses ifadelerinin birimine denk gelir. Ses tanımada bu durumda amaç saklı olan durum dizisini gözlemlerden yararlanılarak bulmaya çalışır. Her fonem için ayrı bir model tanımı yapılır. Her fonem bir model olarak düşünülürse ard arda gelen fonem zincirleri bu modellerin ard arda sıralanması ile modellenir. Bu durumda her bir fonemin son durumunda bir sonraki fonemin ilk durumuna bir geçişi söz konusudur. Böylece ses tanıma HMM ile gerçekleştirilmiş olunur.

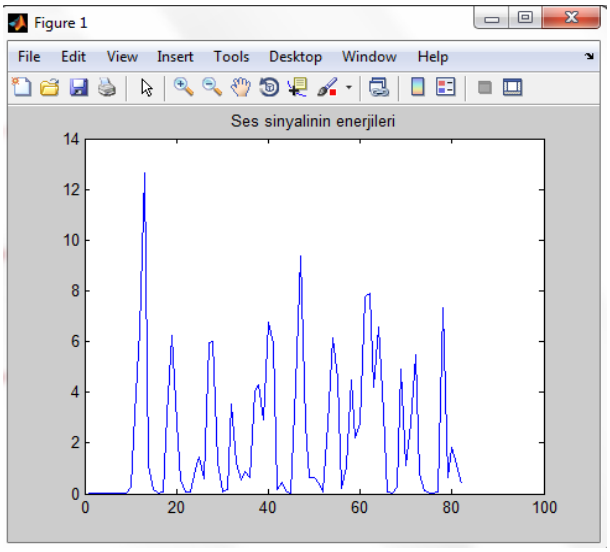
3. Deneysel Çalışma

Bu çalışmada özgün ve gerçek bir Almanca veritabanı kullanılmıştır. Bu veritabanına kişilerin adı soyadı, yaşı, konuşmaları ve cinsiyetleri eklenmiştir. Ses bileşenlerinin MFCC özellik çıkarımı yöntemiyle farklı sayıda özellik vektörleri çıkarılmıştır. Sonraki aşamalarında elde özellik vektörleri kullanılarak GMM, HMM ve DTW yöntemleri kullanılarak ses örnekleri eğitilerek test edilmiştir. Tablo 1'de kaydedilen ses örneklerinin özellikleri gösterilmiştir.

Tablo 1. Kullanılan Özgün Veritabanının Özellikleri

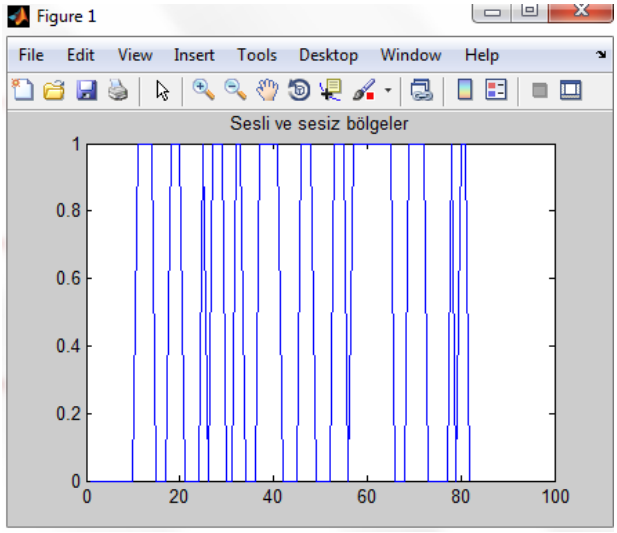
Yaş Aralığı	Konuşmacı Sayısı	
	Erkek	Kadın
18-25 arası konuşmacılar	15	19
26-40 arası konuşmacılar	23	16
41 ve üstü konuşmacılar	12	15

Tablo 2'de MFCC-9 özellik vektörü alınarak elde edilen konuşma örneklerinin GMM, HMM ve DTW için başarı oranları verilmiştir. Kullanılan kelime sayısı arttıkça konuşmacının cinsiyetini tanıma başarısı kullanılan tüm tekniklerde artmıştır. Saklı Markov Modeli diğer tüm tekniklere göre daha başarılı sonuçlar vermiştir. Bu çalışmada özgün ve gerçek bir Almanca veritabanı kullanılmıştır. Bu veritabanına kişilerin adı soyadı, yaşı, konuşmaları ve cinsiyetleri eklenmiştir. Bu veritabanı başta çeşitli üniversitelerdeki eğitim gören öğrenciler olmak üzere çeşitli kurumlarda çalışanlar vs. gibi kişilerden farklı kelime türleri ve çeşitli uzunluklarda oluşan cümleler söyletilerek oluşturulmuştur. Bu veritabanı daha önceden literatürde kullanılmayan ve çeşitli yaşlarda farklı özelliklere sahip kişilerden alınarak elde edilmiştir.

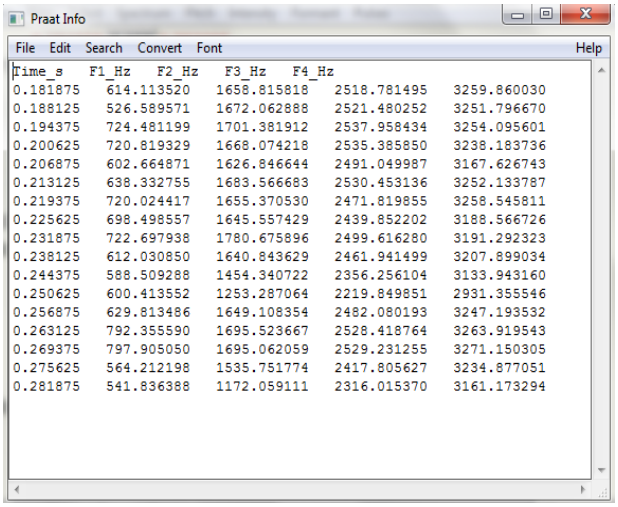


Şekil 2- Ses kaydının enerjisi

Ses bileşenlerinin MFCC özellik çıkarımı yöntemiyle farklı sayıda özellik vektörleri çıkarılmıştır. Sonraki aşamalarında elde özellik vektörleri kullanılarak GMM, SMM ve DTW yöntemleri kullanılarak ses örnekleri eğitilmiştir. Test aşamasında ise eldeki test örneğinin kullanılan yöntemlere erkek ya da kadın olduğu belirlenmeye çalışılmıştır. Ayrıca kullanılan tüm yöntemlerin başarısı hesaplanarak karşılaştırılmalı olarak sunulmuştur. Şekil 2'de örnek bir ses kaydının enerjisi verilmiştir. Şekil 3'de örnek bir ses kaydına ait elde edilen sesli ve sessiz bölgeler Matlab programında belirlenmiştir. Şekil 4'te verilen örnek ses kaydının Praat programında elde edilen formant analizine göre ilk üç formantı Şekil 4'te verilmiştir. Sesin frekansına bakılarak SMM için model oluşturulabilir. Böylelikle gelen ses için bir tahmin belirlenebilir. Formant analizine göre F1 ve F2 formatları orantılı bir şekilde değişiyor. F1 formatı F2 formatından hem kadın sesi için hem de erkek sesi için düşük çıkmaktadır. Kadın sesi F1 ve F2 formatları daha yüksektir.



Şekil 3- Ses kaydında bulunan sesli ve sessiz bölgeler



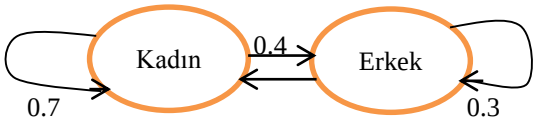
Şekil 4- Formant analizi

Tablo 2'de başarısı yüksek bulunan Saklı Markov Modeli için kullanılan geçiş olasılığının parametreleri verilmiştir. Şekil 5'de SMM geçiş oranının diyagramı verilmiştir. Bu geçiş olasılıkları kullanılarak SMM için bir sistem tasarlanmıştır. SMM modeli için eğitim aşamasında her bir durum için gauss karışım ağırlığı, sayısı gibi parametreler kullanılarak olasılık dağılımı

elde edilmiştir. Ayrıca bu çalışmada SMM için viterbi algoritması kullanılmıştır. Viterbi algoritması gözlemlere bakarak saklı durumları bulur. Bütün olasılıklar toplamaz. Her bir durumda ses vektörüne en benzeyen durumu alır. Bu sebeple diğer SMM algoritmalarına göre daha başarılı sonuçlar verir.

Tablo 2. SMM kullanılan geçiş parametreleri

Durumlar	**Kadın**	**Erkek**
Kadın	0.7	0.4
Erkek	0.3	0.6



0.6

Şekil 5- SMM geçiş diyagramı

Tablo 3'de MFCC-1, MFCC-3, MFCC-5 ve MFCC-9 özellik vektörü alınarak elde edilen konuşma örneklerinin GMM, HMM ve DTW için başarı oranları verilmiştir. Kullanılan kelime sayısı arttıkça konuşmacının cinsiyetini tanıma başarısı kullanılan tüm tekniklerde artmıştır.

Saklı Markov Modeli diğer tüm tekniklere göre daha başarılı sonuçlar vermiştir. Kadınların perde frekansı 120-200 Hz, erkeklerde ise 60-120 Hz arasındadır. Perde frekansından dolayı kadınlarda cinsiyeti belirleme erkeklere göre daha başarılıdır.

4. Sonuçlar

Günümüzde güvenlik ve birçok sebepten dolayı ses tanıma büyük bir önem taşır. Bu çalışmada Alman dili kullanılarak elde edilen özgün veritabanı üzerinde konuşmacının cinsiyetini belirlemeye yönelik bir sistem geliştirilmiştir. Çalışmada kullanılan yöntemlerin sınıflandırma başarısı erkek ve kadınlarda ayrı ayrı hesaplanmıştır ve sonuçlar karşılaştırılmalı bir şekilde sunulmuştur. Sonuçlara bakıldığında; HMM yöntemi diğer sınıflandırma yöntemlerine göre daha başarılı sonuçlar vermiştir. Ayrıca kadınlarda konuşmacı cinsiyeti tanıma sistemi erkeklere göre daha başarılıdır. MFCC 9 özellik çıkarımı 1, 3 ve 5 özellik vektörü kullanılarak elde edilen sonuçlara göre başarısı oldukça yüksektir. Bu çalışma sadece tek kelimelik seslerde değil birden fazla kelime kullanılarak elde edilen cümlelerde de başarılı sonuçlar vermiştir. Ayrıca bu çalışmada sınırlı sayıda kelime kullanılmamıştır. Gelecekte bu yapılan çalışma baz alınarak hibrit bir sistem tasarlanması amaçlanmıştır. Bu sistemde başta Saklı Markov Modeli olmak üzere kullanılan tüm yöntemler geliştirilerek sistemin başarısının arttırılması hedeflenmektedir. Ayrıca sisteme giren kullanıcıların sesinin kesilmesi ya da tam anlamıyla duyulmaması gibi çeşitli sebeplerden dolayı bozuntuya uğrayan seslerin tanınması için de çeşitli çalışmalar yapılacaktır.

Tablo 3. Konuşmacının cinsiyetini tanımada kullanılan yöntemlerin başarısı

Kullanılan Öznitelik Vektörü	**GMM (%)**		**HMM (%)**		**DTW (%)**	
	Kadın	**Erkek**	**Kadın**	**Erkek**	**Kadın**	**Erkek**
MFCC-1	65.15	62.24	71.63	69.63	69.63	67.35
MFCC-3	72.26	65.57	76.35	75.33	76.01	71.93
MFCC-5	78.75	76.33	84.99	82.37	79.22	75.54
MFCC-9	84.26	82.34	98.34	97.02	87.37	86.33

5. Kaynaklar

[1] Q.Jie-Fu, F. Gang, F. Zeng and R. Shannon etc., "Importance of tonal envelope cues in Chinese speech recognition", The Journal of the Acoustical Societct of America, vol.104, no.1, pp.505-510, 1998.

[2] K. Tokuda , H. Zen and A. Black, "An HMM-Based Speech Synthesis System Applied to English", Proc.of 2002 IEEE SSW, pp.227-230, 2012.

[3] D.Reynold , W. Andrews and J.Campbell etc., "The SuperSID Project: Exploiting High-Level Information for High-Accuracy Speaker Recognition", In.Proc. ICASSP, Hong Kong, pp.784-787, 2003.

[4] L.Muda and M.Began, (2010). "Voice Recognition Algorithms using Mel Frequency Cepstral Coefficient (MFCC) and Dynamic Time Warping (DTW) Techniques", Journal Computing, vol.2, issue 3,pp.138-143, ISBN 2151-9617, 2010.

[5] E. Trentin and M. Gori, "A survey of hybrid ANN/HMM models for automatic speech recognition", Elsevier Neurocomputing 37, pp.91-126, 2001.

[6] S.Oh and C.Suen, "A class-modular feed forward neural network for handwriting recognition",

Pattern Recognition, vol.35, issue 1, pp.229-244, 2002.
[7] L.Theodore , N.Ralph and H.Daniel, " The acoustic bases for gender identification from children's voices", J. Acoust. Soc. Am. 109 (6), pp.2988-2998, 2001.
[8] D.Reynolds , T.Quatieri and R.Dunn, "Speaker Verification using Adapted Gaussian Mixture Models", Digital Signal Processing 10, pp.19-41, 2000.
[9] W.Gevaert , G.Tsenov and V.Mladenov, "Neural networks used for speech recognition", Journal of Automatic Control, vol.20, pp.1-7, 2010.
[10] L. Muda, M. Begam and I.Elamvazuthi, " Voice Recognition Algorithms using Mel Frequency Cepstral Coefficient (MFCC) and Dynamic Time Warping (DTW) Techniques", Jornal of Computing, vol.2, issue 3, pp.138-143, ISSN 2151-9617, 2010.
[11] E. Parris, and M.Carey, "Language Independent Gender Identification", Acoustics, Speech, and Signal Processing, 1996. ICASSP-96. Conference Proceedings., 1996 IEEE International Conference on, vol.2, pp.685-688, 1996.
[12] D.A. Reynolds and R.C. Rose, "Robust Text-Independent Speaker Identification Using Gaussian Mixture Speaker Models", IEEE Trans. Speech Audio Proc., 3, (1), pp. 72–83, 1995.