Journal of the Faculty of Engineering and Architecture of Gazi University 31:3 (2016) 501-509

Mühendislik Mimarlık Fakültesi Dergisi
Journal of The Faculty of Engineering and Architecture of Gazi University

Elektronik / Online ISSN : 1304 - 4915
Basılı / Printed ISSN : 1300 - 1884

Konuşmacı yaş ve cinsiyetinin GKM süpervektörlerine dayalı bir DVM sınıflandırıcısı ile belirlenmesi

Ergün Yücesoy[1*], Vasif V. Nabiyev[2]

[1]Ordu Üniversitesi, Teknik Bilimler Meslek Yüksekokulu, 52200, Ordu, Türkiye
[2]Karadeniz Teknik Üniversitesi, Bilgisayar Mühendisliği Bölümü, 61080, Trabzon, Türkiye

Ö N E Ç I K A N L A R

- Konuşmacıları yaş ve cinsiyetlerine göre sınıflandıran yeni bir sistem önerilmiştir
- 954 kişinin 47 saatlik telefon konuşmaları kullanılmıştır
- Önerilen sistem için en uygun bileşen sayısı ve konuşma süresi belirlenmiştir

Makale Bilgileri

Geliş: 26.11.2014
Kabul: 24.05.2016

DOI:
10.17341/gummfd.71595

Anahtar Kelimeler:
Yaş ve cinsiyet tanıma, gauss karışım modeli, gauss karışım modeli süpervektörleri, destek vektör makinesi

ÖZET

Bu çalışmada konuşmacıları yaş ve/veya cinsiyet özelliklerine göre otomatik olarak sınıflandıran bir sistem önerilmiştir. Açık ve kapalı mekanlarda cep telefonu ve karasal bağlantılarla yapılan telefon konuşmalarının giriş olarak kullanıldığı bu 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 GKM modelleri süpervektörlere dönüştürülerek DVM sınıflandırıcısına uygulanmıştır. Çalışmada konuşmaların ses içeren bölümlerinin belirlenmesinde sinyalin enerji özelliği kullanılırken GKM modellerinin eğitiminde ise geniş bir veritabanı ile eğitilen genel arka plan modelinin (GAM) uyarlanması yaklaşımı tercih edilmiştir. Çalışmada ayrıca farklı sayıda bileşenle oluşturulan GKM modelleri farklı uzunluklu konuşmalarla test edilerek GKM bileşen sayısı ve konuşma süresinin yaş ve cinsiyet tespiti üzerindeki etkisi de araştırılmıştır. Yapılan testlerde en yüksek sınıflandırma başarıları 16 saniyelik konuşmaların 64 bileşenli GKM'lerle modellenmesi sonucunda elde edilmiştir. Bu oranlar cinsiyet kategorisinde %92,42, yaş kategorisinde %60,10 ve yaş&cinsiyet kategorisinde ise %60,02 olarak ölçülmüştür.

Determination of a speaker's age and gender with an SVM classifier based on GMM supervectors

H I G H L I G H T S

- A new system for classifying speakers according to their age and genders is proposed
- 47-hour phone conversations from 954 speakers are used
- Optimum number of components and speech duration for proposed system are determined

Article Info

Received: 26.11.2014
Accepted: 24.05.2016

DOI
10.17341/gummfd.71595

Keywords:
Age and gender recognition, gaussian mixture model, gaussian mixture model supervectors, support vector machine

ABSTRACT

In this study, a system classifying speakers according to their age and/or genders is proposed. In this system phone conversations including mobile calls that took place indoor or outdoor are used as inputs. It is aimed to classify the speakers according to their genders into three classes as male, female and child, according to their ages into four classes as child, youth, adult and senior, and finally according to both gender and age into seven classes. For this aim, GMM models that are created with MFCC coefficients obtained by the voiced parts of the conversations are transformed into supervectors. These supervectors are applied to SVM classifier. Signal energy is used for determining the voiced parts of conversations. For the training of GMM models, the adaptation approach of UBM is preferred. Also, by testing GMM models that are created with different number of components and different length conversations, the impact of GMM components number and speech duration on the age and gender identification is investigated. At the end of these tests, the highest classification success rates are obtained by modeling 16-second speeches with 64-component GMMs. The rates obtained from these tests are measured as 92.42% for gender category, 60.10% for age category and 60.02% for age&gender category.

* Sorumlu Yazar/Corresponding author: yucesoye@odu.edu.tr / Tel: +90 452 233 4865

1. GİRİŞ (INTRODUCTION)

İnsanlar arasındaki en doğal ve en etkili iletişim şekli olan konuşma, sadece seslendirilen kelimenin anlam bilgisini içermez. O aynı zamanda konuşmacı hakkında kimlik, yaş, cinsiyet ve psikolojik durum gibi paralinguistik bilgiler de içerir. Konuşmayla iletilen bu bilgilerden iletişim sırasında değişik şekillerde faydalanılır. Örneğin; bir telefon konuşmasında konuşmacının cinsiyeti çoğunlukla anlaşılır ve hitap şekli ona göre belirlenir. Paralinguistik bilgiler aynı zamanda insan-bilgisayar iletişiminde de kullanılabilir. Örneğin; etkileşimli sesli yanıt sistemlerinde konuşmacının yaş ve cinsiyeti otomatik olarak tespit edilerek bu bilgiler istatistiksel amaçlarla veya bekleme kuyruğunda konuşmacıya uygun reklam veya müzik seçiminde kullanılabilir. Benzer şekilde konuşmacıların ruhsal durumu tespit edilerek kızgın müşteriler önceden belirlenmiş operatörlere yönlendirilebilir ve böylece müşteri memnuniyeti arttırılabilir. Ayrıca altyazılı TV yayınlarında yazı rengi konuşmacının cinsiyetine göre değiştirilerek, işitme güçlüğü çeken insanların yazının hangi konuşmacıya ait olduğunu anlamalarına yardımcı olunabilir. Bu uygulamalar dışında cinsiyet bilgisi konuşmacı ve konuşma tanıma sistemlerinde ön bilgi olarak da kullanılır. Cinsiyet ön bilgisi konuşmacı tanıma sistemlerinde araştırma uzayını aynı cinsiyetli konuşmacılarla sınırlandırarak; konuşma tanıma sistemlerinde ise cinsiyet bağımlı modellerin tanımlanmasına imkan sunarak performans artışı sağlar. Cinsiyet sınıflandırma sistemlerinin çoğunda yalnızca yetişkin konuşmacılar kullanılmasına rağmen özellikle çocuk istismarı videolarının tespiti gibi uygulamalar için çocuk seslerinin de sınıflandırılması son derece önemlidir. Ancak ergenlik öncesinde çocuk seslerinden cinsiyet ayrımı yapılamadığı için çocuk sesleri genellikle tek sınıfta değerlendirilir. Günümüzde sesten yaş ve cinsiyet sınıflandırma konusunda genellikle üç kategori kullanılmaktadır [1]. Bunlardan cinsiyet kategorisinde konuşmacılar çocuk, erkek ve bayan olmak üzere üç gruba; yaş kategorisinde ise çocuk, genç, yetişkin ve yaşlı olarak dört gruba ayrılmaktadır. Bu iki kategorinin birleşimi olan son kategoride ise konuşmacılar çocuk, genç erkek, genç bayan, yetişkin erkek, yetişkin bayan, yaşlı erkek ve yaşlı bayan olarak yedi gruba ayrılmaktadır. Kişinin yaş ve cinsiyetinin belirlenmesi konusunda farklı biyometrik özellikler kullanılmaktadır [2, 3]. Bu özelliklerden olan ses biyometrisine ilgi düşük maliyeti, veri girişindeki kolaylığı ve uzaktan erişim avantajlarından dolayı son yıllarda artmıştır. Kişilerin ses kayıtlarına göre erkek ve bayan olarak sınıflandırıldığı Hu ve arkadaşlarının [4] çalışmasında perde frekansı ve GKM sınıflandırıcısından oluşan iki seviyeli bir sınıflandırıcı kullanılarak TIDIGITS veri kümesi üzerinde %98,6 başarı sağlanmıştır. Djemili ve arkadaşlarının MFCC öznitelikleri ve dört farklı sınıflandırıcı (GMM, MLP, VQ ve LVQ) kullandıkları [5] çalışmasında ise IViE veri kümesindeki konuşmacılar cinsiyetlerine göre %96,4 başarıyla sınıflandırılmıştır. Cinsiyet sınıflandırma konusunda perde frekansı ve MFCC özniteliklerinin birlikte kullanıldığı çalışmalar da vardır. Bu çalışmalardan Gaikwad ve arkadaşlarının [6] çalışmasında %95, Phoophuangpairoj ve arkadaşlarının [7] çalışmasında ise %98,92 başarı sağlanmıştır. Metze ve arkadaşlarının telefon konuşmaları üzerinde yaptıkları çalışmada konuşmacıları 7 sınıfa ayıran 4 farklı yaklaşım, insan performansıyla karşılaştırılmıştır [8]. Bu yaklaşımların Speech Dat II veritabanı üzerindeki başarıları sırasıyla %54, %40, %27 ve %42 olarak ölçülmüştür. Ming Li ve arkadaşlarının yaptığı çalışmada ise akustik seviyede beş farklı yöntem skor seviyesinde birleştirilerek cinsiyet kategorisinde %88,2, yaş kategorisinde %52,8 ve yaş&cinsiyet kategorisinde ise %51,1 başarı sağlanmıştır [9]. Bir diğer araştırmada, cinsiyet sınıflandırıcı ve yaş regresyonu birleştirilerek yaş&cinsiyet kategorisinde başarı %45'den %50,7'ye çıkarılmıştır [10]. Meinedo ve Isabel'in [11] çalışmasında ise üç farklı sınıflandırıcı (SVM, MLP ve GMM-UBM) doğrusal lojistik analizi ile birleştirilerek cinsiyet kategorisinde %83,1, yaş kategorisinde ise %51,2 başarı sağlanmıştır. Konuşmacıların cinsiyetlerine göre iki, yaşlarına göre ise üç sınıfa ayrıldığı Oscal ve arkadaşlarının çalışmasında uyarım yoğunluklarına göre yüksek ve düşük olarak ayrılan sesler için farklı öznitelikler seçilerek cinsiyet kategorisinde %98,9, yaş kategorisinde ise %71,6 başarı sağlanmıştır [12]. Yaşları 4 ile 8 arasında değişen 134 çocuk ve 18 yetişkin konuşmacının yaş gruplarına göre sınıflandırıldığı çalışmada ise Neuro-fuzzy sınıflandırıcı kullanılarak %85 sınıflandırma başarısı sağlanmıştır [13].

Bu çalışmada GKM'nin genelleyici gücü ile DVM'nin ayırıcı özelliğini birleştiren GKM süpervektörlerine dayalı DVM yaklaşımı konuşmacıların yaş ve cinsiyet özelliklerine göre sınıflandırılması problemine uygulanmıştır. Farklı oturumlarda kaydedilen telefon konuşmalarından elde edilen MFCC katsayılarının öznitelik olarak kullanıldığı çalışmada yaş ve cinsiyet özellikleri ayrı ayrı ve birlikte ele alınarak üç kategoride testler yapılmıştır. Bu kategorilerin ilki olan cinsiyet kategorisinde konuşmacılar erkek, bayan ve çocuk olarak üç sınıfa ayrılırken, ikinci kategori olan yaş kategorisinde konuşmacılar çocuk, genç, yetişkin ve yaşlı olarak dört sınıfta ayrılmıştır. Bu iki kategorinin birleşimi olan yaş&cinsiyet kategorisinde ise konuşmacılar çocuk, genç erkek, genç bayan, yetişkin erkek, yetişkin bayan, yaşlı erkek ve yaşlı bayan olarak yedi sınıfa ayrılmıştır. Çalışmada ayrıca konuşma süresinin ve GKM bileşen sayısının yaş ve cinsiyet tespitine etkisi de ayrıntılı olarak incelenerek en uygun konuşma süresi ve bileşen sayısı tespit edilmeye çalışılmıştır. Makalenin kalan kısmının konu akışı şöyledir; çalışmada kullanılan veritabanı ve sınıf tanımlamaları Bölüm 2'de, öznitelik vektörü olarak kullanılan MFCC katsayılarının çıkarılma prosedürü Bölüm 3'te anlatılmıştır. Bölüm 4'te GKM modeli, Bölüm 5'te DVM sınıflandırıcısı, Bölüm 6'da ise GKM-DVM sistemi tanıtılmıştır. Çalışma Bölüm 7'de sunulan deneysel bulgular ve Bölüm 8'de sunulan sonuçlar bölümüyle sonlandırılmıştır.

2. VERİTABANI VE SINIF TANIMLAMALARI (DATABASE AND CLASS DEFINITIONS)

Çalışmada önerilen yaklaşım aGender veritabanı [1] kullanılarak değerlendirilmiştir. Bu veritabanı, konuşmacıların yaş ve cinsiyetlerini tespit eden sistemlerin geliştirilmesine yardımcı olmak amacıyla "InterSpeech 2010 Paralinguistic Challenge" organizasyonu tarafından oluşturulmuştur. aGender veritabanı 954 Alman konuşmacının 6 oturumda gerçekleştirdiği 65364 telefon görüşmesinden oluşmaktadır. Uzunlukları 1 ile 6 saniye arasında değişen sabit ve değişken içerikli ifadelerin kullanıldığı bu görüşmelerin toplam süresi yaklaşık 47 saattir. Birer gün arayla yapılan oturumlarda aramalar cep telefonu ve karasal bağlantılarla iç ve dış mekanlarda yapılarak 8000 Hz 16 bit PCM formatında kaydedilmiştir. Eğitim, geliştirme ve test olmak üzere 3 bölümden oluşan aGender veritabanında konuşmacılar çocuk, genç, yetişkin ve yaşlı olarak 4 yaş grubuna ayrılmıştır. Bu yaş gruplarının seçiminde yaşla oluşan fizyolojik değişimlerden ziyade uygulama kaynaklı ihtiyaçlar etkili olmuştur. Çocuk hariç diğer konuşmacılar ise cinsiyetlerine göre erkek ve bayan olarak ayrılarak, 7 yaş-cinsiyet sınıfında tanımlanmıştır (Tablo 1). Bu konuşmacılardan eğitim ve geliştirme bölümündekiler yaş-cinsiyet tanıma sistemlerinin geliştirilmesinde, test bölümündekiler ise sonuçların farklı çalışmalarla kıyaslanmasında kullanılır.

Tablo 1. aGender veritabanında tanımlı sınıflar ve bu sınıflardaki konuşmacı ve kayıt sayıları
(Classes defined in aGender database and speaker and record numbers in these classes)

Sınıf ID	Yaş Grubu	Yaş	Cinsiyet	#Eğitim	#Geliştirme
1	Çocuk	7-14	X	68/4406	38/2396
2	Genç	15-24	Erkek	63/4638	36/2722
3	Genç	15-24	Bayan	55/4019	33/2170
4	Yetişkin	25-54	Erkek	69/4573	44/3361
5	Yetişkin	25-54	Bayan	66/4417	41/2512
6	Yaşlı	55-80	Erkek	72/4924	51/3561
7	Yaşlı	55-80	Bayan	78/5549	56/3826

3. ÖZNİTELİK ÇIKARMA (FEATURE EXTRACTION)

Bütün ses tanıma sistemlerinin ilk aşamasında ses sinyali konuşma ve konuşmacı özelliklerini temsil eden daha az değişkenliğe ve daha fazla ayırıcı özelliğe sahip parametrik değerlere dönüştürülür [14]. Öznitelik olarak isimlendirilen bu parametrik değerlerin ses sinyalinden çıkarılmasında kullanılan değişik yöntemler vardır. Kısa süreli analiz yöntemleri olarak da isimlendirilen bu yöntemler ses sinyalinin durağan olarak kabul edilen kısa parçaları üzerinde uygulanır. Süreç sonucunda her analiz parçasından elde edilen skaler değerler birleştirilerek öznitelik vektörü oluşturulur. Bu çalışmada öznitelik vektörü olarak Mel Frekansı Kepstrum Katsayıları kullanılmıştır. Bu yöntemin aşamaları aşağıda sunulmuştur.

3.1. Ön-Vurgulama (Pre-emphasis)

Ses sinyalindeki frekansla azalan enerjiyi dengelemek için kullanılan ön vurgulama; yüksek frekanslı bölümleri güçlendiren bir filtreleme işlemidir. Dudak ve bazı gırtlaksal etkilerin atılmasını da sağlayan bu işlem için genellikle Eş. 1 ifadesiyle temsil edilen FIR filtresi kullanılır.

$$H(z) = 1 - \alpha z^{-1} \;\; 0{,}9 \leq \alpha \leq 1 \tag{1}$$

3.2. Çerçeve Bloklama ve Pencereleme (Frame Blocking and Windowing)

Zamanla yavaş değişen bir sinyal olan konuşma sinyali yeterince kısa periyotlarda incelendiğinde yaklaşık durağan akustik özelliklere sahip olduğu görülür [14]. Bu nedenle konuşma sinyalleri genellikle kısa süreli parçalara (20-30 ms) bölünerek analiz edilir. Çerçeve olarak isimlendirilen bu parçaların sınırlarındaki bilgi kaybını önlemek için çerçeveler kendinden önceki çerçevenin bir kısmıyla (10-20 ms) çakıştırılır. Sinyalin sınırlarında oluşan süreksizlik ise çerçevelere bir pencere fonksiyonu uygulanarak azaltılır. Bu işlem için genellikle Hamming penceresi kullanılır [15].

3.3. Spektral Analiz (Spectral Analysis)

Konuşma sinyali pencerelendikten sonra zaman uzayındaki örnekler Eş. 2 ile temsil edilen ayrık Fourier dönüşümü kullanılarak frekans uzayına taşınır.

$$X(k) = \sum_{n=0}^{N-1} x(n) e^{-j2\pi nk/N} \;, 0 \leq k \leq N \tag{2}$$

3.4. Mel-Frekanslı Kaydırma (Mel-Frequency Wrapping)

Volkmann ve Newman yaptıkları deneylerde insanların sesin frekans içeriğini Mel ölçeği olarak isimlendirilen doğrusal olmayan bir ölçeğe göre algıladıklarını ortaya koymuşlardır [14]. Yaklaşık 1000 Hz'ye kadar doğrusal; üzerinde ise logaritmik olarak değişen Mel ölçeği genellikle Eş. 3 ifadesiyle temsil edilir.

$$Mel(f) = 2595 \log_{10}\left(1 + \frac{f}{700}\right) \tag{3}$$

Bu ifadedeki f gerçek frekansı $Mel(f)$ ise algılanan frekansı temsil etmektedir. Doğrusal olmayan bu frekans algılamasının MFCC özniteliklerine dahil edilmesi için sesin genlik spektrumu Mel ölçeğine göre eşit aralıklarla yerleştirilmiş bir filtre kümesiyle saptırılır. Bu filtre kümesi alt sınırı kendinden önceki, üst sınırı ise kendinden sonraki filtrenin merkez frekansına gelecek şekilde yerleştirilen üçgen filtrelerden oluşur. Sinyalin Mel spektrumu Eş. 4 ifadesiyle hesaplanır.

$$S(m) = \sum_{k=0}^{\frac{N}{2}-1} |X(k)|^2 |H_m(k)| \;\; 1 \leq m \leq M \tag{4}$$

Burada M filtre sayısını, $H_m(k)$ ise k'ıncı filtrenin m'inci çıkışa katkısını ifade eder.

3.5. Logaritmik Sıkıştırma (Logarithmic Compression)

Bu aşamada sesin gürlüğünün veya sessizliğinin filtre çıkışlarına etkisini azaltmak ve insanın doğrusal olmayan genlik hassasiyetini modellemek için filtre çıkışlarının logaritması alınır. Bu işlem Eş. 5 ifadesiyle temsil edilir.

$$S_{ln}(m) = ln(S(m)) \;\; 1 \leq m \leq M \tag{5}$$

3.6. Ayrık Kosinüs Dönüşümü (Discrete Cosine Transform-DCT)

Algoritmanın son aşamasında filtre çıkışlarına ayrık kosinüs dönüşümü uygulanarak yavaş değişen spektral bilgilerin hızlı değişenlerden ayrıldığı kepstral uzaya geçilir. DCT sonucunda elde edilen vektörün sadece düşük dereceli katsayıları ile MFCC öznitelik vektörü oluşturulur. p kepstrum derecesi olmak üzere $k.$'ıncı MFCC katsayısı Eş. 6 ifadesi ile hesaplanır.

$$C_k = \sqrt{\frac{2}{M}} \sum_{m=1}^{M} S_{ln}(m) \cos\left(\frac{\pi k(m-0{,}5)}{M}\right) \;\; 1 \leq k \leq p \tag{6}$$

4. GAUSS KARIŞIM MODELİ (GAUSSIAN MIXTURE MODEL)

İki veya daha fazla olasılık yoğunluk fonksiyonunun konveks kombinasyonu olan karışım modelleri, özellikle karmaşık verileri modellemede kullanılan güçlü ve esnek bir araçtır. İstatistiksel veri analizi ve makine öğrenmesi gibi birçok alanda kullanılan Gauss karışım modeli (GKM), Eş. 7 ifadesi ile temsil edilir [16].

$$p(x_t|\lambda) = \sum_{i=1}^{M} w_i p_i(x_t) \tag{7}$$

Burada M bileşen sayısını, x_t D-boyutlu bir vektörü, w_i karışım ağırlıklarını, p_i ise bileşen yoğunluklarını temsil etmektedir. Her bileşen ise ortalama vektörü μ_i ve kovaryans matrisi $\sum_i$ olan D-değişkenli bir Gauss fonksiyonudur ve Eş. 8 ifadesi ile temsil edilir.

$$p_i(x_t) = \frac{1}{(2\pi)^{D/2}\sqrt{|\sum_i|}} e^{-\frac{1}{2}(x_t-\mu_i)^T \sum_i^{-1}(x_t-\mu_i)} \tag{8}$$

Tüm bileşenlerin karışım ağırlıkları, ortalama vektörleri ve kovaryans matrisleri Eş. 9 ifadesi ile temsil edilerek GKM modeli oluşturulur.

$$\lambda = \{w_i, \mu_i, \sum_i\} \;\; i = 1, \ldots, M \tag{9}$$

4.1. Parametre Tahmini (Parameter Estimation)

GKM parametrelerinin tahmini için kullanılan değişik yöntemler vardır. Bu yöntemlerden en güncel ve iyi kurgulanmış olanı en çok olabilirlik (ML) kestirimidir. ML kestiriminin amacı verilen eğitim verisinden GKM'nin olabilirliğini maksimum yapan model parametrelerini bulmaktır. T adet eğitim vektörü için GKM olabilirliği Eş. 10 şeklinde yazılabilir.

$$p(X|\lambda) = \prod_{t=1}^{T} p(x_t|\lambda) \tag{10}$$

Bu ifade, λ parametresinin doğrusal olmayan bir fonksiyonudur ve direkt maksimizasyonu mümkün değildir. Bu nedenle ML parametre kestirimi beklenti maksimizasyonu (EM) olarak isimlendirilen tekrarlamalı bir algoritmayla yapılır [17]. Beklenti ve maksimizasyon olmak üzere iki aşamadan oluşan EM algoritması rastgele veya bazı sezgisel yöntemlerle belirlenen bir λ modelinden $p(X|\bar{\lambda}) \geq p(X|\lambda)$ şartını sağlayan $\bar{\lambda}$ modelini tahmin etmeye çalışır. Belirlenen $\bar{\lambda}$ modeli bir sonraki aşamada başlangıç modeli olarak kullanılarak aynı işlemler tekrarlanır. Süreç belirli bir tekrar sayısına ulaşana kadar ya da olabilirlik fonksiyonundaki değişim belirli bir değerin altına düşene kadar tekrarlanarak en uygun parametre tahmini yapılır.

4.2. Genel Arkaplan Modeli (Universal Background Model)

Genel arkaplan modeli (GAM) konuşmacıdan bağımsız özniteliklerin dağılımını temsil eden yüksek bileşenli bir GKM modelidir [18]. GAM'ın eğitiminde kullanılan farklı yaklaşımlar vardır. Bunlardan en basit olanı tüm verilerin basitçe birleştirilerek tek bir GAM'ın EM algoritmasıyla eğitilmesi yaklaşımıdır. Bu yaklaşımda birleştirilen verilerde tüm alt grupların dengeli olmasına dikkat edilmelidir. Aksi halde sonuç model, baskın gruba meyilli olacaktır. Özellikle dengesiz veri dağılımın olduğu durumlarda kullanılan diğer bir yaklaşımda her alt grup için ayrı bir GAM eğitilip daha sonra bu modeller birleştirilir. Bu çalışmada birinci yaklaşım tercih edilmiştir.

4.3. MAP Uyarlaması (Maximum A Posteriori Adaptation)

Özellikle yüksek bileşenli GKM'lerin eğitiminde konuşmacıların kısa süreli verileri yeterli olmayacaktır. Bu durumda model parametrelerinin tahmini için doğrudan EM algoritmasının kullanımı yerine GAM parametrelerinin yeni verilerle güncellenmesi fikrine dayanan uyarlama yaklaşımı kullanılır. Literatürde konuşmacı modellerinin uyarlanması için kullanılan değişik yöntemler vardır [19]. Bunlardan MAP (Maximum A Posteriori) uyarlaması EM algoritmasına benzeyen iki aşamalı bir tahmin algoritmasıdır. MAP algoritmanın ilk aşamasında GAM'daki her bileşen için Eş. 11-14 istatistikleri hesaplanır [18].

$$\Pr(i|x_t) = \frac{w_i p_i(x_t)}{\sum_{j=1}^{M} w_j p_j(x_t)} \tag{11}$$

$$n_i = \sum_{t=1}^{T} \Pr(i|x_t) \tag{12}$$

$$E_i(\mathrm{x}) = \frac{1}{\mathrm{n_i}} \sum_{\mathrm{t=1}}^{\mathrm{T}} \Pr(i|x_t)\, x_t \tag{13}$$

$$E_i(\mathrm{x}^2) = \frac{1}{n_i} \sum_{\mathrm{t=1}}^{\mathrm{T}} \Pr(i|x_t)\, {x_t}^2 \tag{14}$$

Algoritmanın son aşamasında ise GAM parametreleri bu yeni istatistiklerle güncelleştirilerek uyarlanmış GKM parametreleri elde edilir (Eş. 15-17).

$$\widehat{w}_i = \left[\frac{\alpha_i^w n_i}{T} + (1-\alpha_i^w)w_i\right]\gamma \quad (15)$$

$$\hat{\mu}_i = \alpha_i^m E_i(x) + (1-\alpha_i^m)\mu_i \quad (16)$$

$$\widehat{\sigma^2}_i = \alpha_i^v E_i(x^2) + (1-\alpha_i^v)(\sigma_i{}^2 + \mu_i{}^2) - \hat{\mu}_i{}^2 \quad (17)$$

Buradaki $\{\alpha_i^w, \alpha_i^m, \alpha_i^v\}$ katsayıları sırasıyla ağırlık, ortama ve varyans parametrelerinin eski ve yeni tahminleri arasındaki dengeyi belirleyen uyarlama katsayılarıdır. γ ise tüm uyarlanmış karışım ağırlıklarının toplamını 1 yapan bir ölçek faktörüdür.

4.4. GKM Süpervektörü (GMM Supervector)

Öznitelik vektörlerine dönüştürülen konuşma sinyalleri bir sınıflandırıcıya giriş olarak verilerek konuşmacılar yaş, cinsiyet, psikolojik durum gibi farklı özelliklerine göre sınıflandırılabilir. Bu çalışmada kullanılan DVM'nin de dahil olduğu sınıflandırıcıların çoğu sabit uzunluklu vektörleri giriş olarak kabul eder. Bu nedenle boyutu konuşma süresine göre değişen öznitelik vektörlerinin sabit uzunluklu bir vektöre dönüştürülmesi gerekir. Bu amaçla kullanılan GKM süpervektörü, GAM'dan uyarlanarak oluşturulan GKM bileşenlerinin ortalamaları uç uca eklenerek oluşturulur [20].

5. DESTEK VEKTÖR MAKİNELERİ (SUPPORT VECTOR MACHINE)

Destek vektör makineleri (DVM) iki sınıf arasında ayırıcı bir hiper düzlem bulmaya çalışan güçlü bir sınıflandırıcıdır ve güncel birçok uygulamada yaygın olarak kullanılmaktadır [21]. Doğrusal olarak sınıflandırılabilen bir veri kümesi için ayırıcı hiper düzlem sayısı birden fazla olabilir. Fakat bunlardan bir tanesi iki sınıf arasındaki mesafeyi maksimum yapar ve optimum hiper düzlem olarak isimlendirilir. DVM'nin amacı da maksimum marjine sahip bu hiper düzlemi bulmaktır. DVM sınıflandırıcısı bir çekirdek fonksiyonu kullanılarak doğrusal olmayan sınıflandırma problemlerine de uygulanabilir. x gözlem vektörü olmak üzere iki sınıflı bir DVM Eş. 18 ifadesiyle tanımlanır.

$$f(x) = \sum_{i=1}^{N} y_i \alpha_i K(x, x_i) + b \quad (18)$$

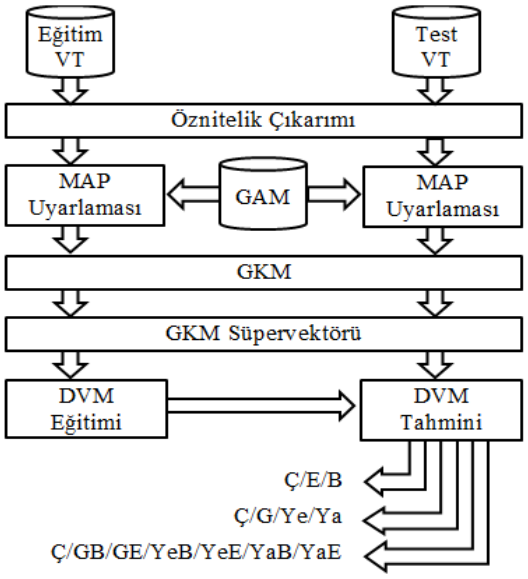
Burada x_i destek vektörlerini, $y_i \in \{1, -1\}$ bu vektörlere karşılık gelen sınıf etiketlerini α_i ve b ise eğitim sırasında belirlenen diğer parametreleri temsil etmektedir. $K(x, y)$ çekirdek fonksiyonu olup giriş vektörünü yüksek boyutlu bir uzaya taşımak için kullanılır. Böylece giriş uzayında doğrusal olarak sınıflandırılamayan veriler genişletilmiş uzayda doğrusal olarak sınıflandırılabilir. Çekirdek fonksiyonu olarak kullanılan çeşitli alternatifler vardır. Bu alternatiflerden Eş. 19 ifadesi ile tanımlanan RBF fonksiyonu giriş vektörünü yüksek boyutlu uzaya doğrusal olmayan bir haritalama ile dönüştürdüğü için sınıf etiketleri ile öznitelikler arasındaki ilişkinin doğrusal olmadığı durumlar için uygundur [22].

$$K(x_i, y_j) = exp\left(-\left\|x_i - y_j\right\|^2 / 2\sigma^2\right) \quad (19)$$

Bu nedenle bu çalışmada da RBF çekirdeği tercih edilmiştir. Burada, σ çekirdek genişliği olup eğitim aşamasında belirlenmesi gereken bir parametredir. Çekirdek fonksiyonunun kullanılmasına rağmen bazı veri kümelerinin doğrusal olarak sınıflandırılması mümkün olmayabilir. Bu veri kümeleri bazı örneklerin marjini ihlal etmesine izin verilerek sınıflandırılabilir. Bu durumda yanlış sınıflandırma anlamına gelen marjin ihlallerinin minimizasyonu ile marjin genişliğinin maksimizasyonu arasında bir denge kurulması gerekir. Bu denge, eğitim aşamasında belirlenen ve yanlış sınıflandırılan noktaları cezalandıran bir parametre (C) ile sağlanır. Çalışmada en uygun RBF parametrelerin belirlenmesi amacıyla ızgara tarama yöntemi kullanılmıştır. Bu yöntemle σ parametresi -5 ile 15, C ise -15 ile 3 arasında değiştirilerek parametre optimizasyonu yapılmıştır. DVM aslında iki sınıflı bir sınıflandırıcı olmasına rağmen çeşitli yöntemler kullanılarak çok sınıflı problemlere de uygulanmaktadır [23]. Bu yöntemlerden biri olan ve bu çalışmada da tercih edilen bire-karşı-bir yaklaşımda tüm sınıfların olası çiftleri için bir DVM sınıflandırıcısı tanımlanır. Tahmin aşamasında ise eğitilen DVM'ler test verileri ile test edilerek en yüksek oy alan sınıf test verisinin sınıfı olarak belirlenir.

6. GKM-DVM SÜPERVEKTÖRÜNE DAYALI SINIFLANDIRMA SİSTEMİ (GMM-SVM SUPERVECTOR- BASED CLASSIFICATION SYSTEM)

GKM'nin genelleyici gücü ile DVM'nin ayırıcı özelliklerini birleştiren GKM-DVM süpervektör yaklaşımı ilk olarak Campbell tarafından konuşmacı doğrulama sisteminde kullanılmıştır [20].



Şekil 1. Önerilen sistemin blok diyagramı
(Block diagram of the proposed system)

Tablo 2. GAM verilerin sınıf içi dağılımı (In-class distribution of UBM data)

Yaş-cinsiyet Sınıfı	Konuşmacı Sayısı	Konuşma Sayısı	Toplam Süre (Saniye)
Çocuk	5	231	662
Genç Erkek	5	330	835
Genç Bayan	5	338	820
Yetişkin Erkek	5	314	806
Yetişkin Bayan	5	346	874
Yaşlı Erkek	5	312	838
Yaşlı Bayan	5	379	1038

Tablo 3. Cinsiyet sınıflandırma başarısı (Gender Classifier Success)

Süre	GKM16	GKM32	GKM64	GKM128	GKM256
2 s	86,96	87,24	87,06	87,89	88,34
4 s	88,83	89,18	90,05	89,91	90,23
8 s	89,30	90,88	91,61	91,29	90,11
16 s	90,62	91,47	***92,42***	91,39	90,28
32 s	90,95	92,06	92,15	90,08	84,16

Konuşmacı doğrulamada sağlanan yüksek başarım yönteme olan ilgiyi arttırmış ve konuşmacıların yaş ve cinsiyetlerine göre sınıflandırılmasının da dahil olduğu birçok çalışmada bu yaklaşım kullanılmaya başlanmıştır. Ancak yapılan çalışmalarda konuşma süresi ve GKM bileşen sayısının yaş ve cinsiyet sınıflandırma üzerindeki etkisi incelenmemiştir. Çalışmamızda bu etkiler incelenerek en uygun bileşen sayısı ve konuşma süresi belirlenmeye çalışılmıştır. Bu amaçla geliştirilen sisteminin blok diyagramı Şekil 1'de gösterilmiştir.

7. SONUÇLAR VE TARTIŞMALAR
(RESULTS AND DISCUSSIONS)

Konuşmacıların yaş ve cinsiyet özelliklerine göre 3 kategoride sınıflandırılması amaçlanan bu çalışmada aGender veritabanı kullanılarak çeşitli deneyler yapılmıştır. Bu deneylerde konuşmacıların ses karakteristiklerini temsil etmek için MFCC katsayılarından oluşan bir vektörü kullanılmıştır. Çalışmamızda konuşmaların yalnızca ses içeren bölümlerinden elde edilen öznitelikler kullanılmıştır. Bu amaçla 20 ms genişliğindeki bir Hamming penceresinin 10 ms kaydırılmasıyla çerçevelere bölünen ses sinyalinden konuşma içermeyen kısımlar enerjiye dayalı olarak çıkarılmıştır. Daha sonra kalan çerçevelerden elde edilen katsayılar birleştirilerek öznitelik vektörü oluşturulur. Çalışmamızda her bir çerçeveden elde edilen 13 MFCC katsayısı ve onların birinci ve ikinci türevlerinden oluşan toplam 39 katsayı öznitelik olarak kullanılmıştır. Çalışmamızda GAM'ın eğitimi için aGender veritabanın eğitim bölümünden rastgele seçilen 35 konuşmacının 2250 konuşması, DVM sınıflandırıcısının eğitimi için ise kalan 436 konuşmacının 30277 konuşması kullanılmıştır. GAM'ın eğitiminde kullanılan verilerin seçiminde her yaş ve cinsiyet grubundan eşit sayıda konuşmacının seçilmesine dikkat edilmiştir. Tablo 2'de bu verilerin sınıf içi dağılımları verilmiştir. Eğitilen DVM sınıflandırıcısı ise geliştirme bölümündeki 299 konuşmacının 20549 konuşması ile test edilerek bu konuşmaların sınıf etiketleri tahmin edilmiştir. Bu işlemler sırasında her konuşmacının seslendirdiği konuşmalar bütün olarak değerlendirilerek sabit uzunluklu parçalara bölünmüştür. Elde edilen sabit uzunluklu konuşmalar ise 16, 32, 64,128 ve 256 bileşenli GKM'lerle modellenerek konuşma süresi ve bileşen sayısının başarıya etkisi incelenmiştir. Çalışmada sınıflandırıcının başarısı sınıf etiketi doğru tahmin edilen örnek sayısı toplam örnek sayısına bölünerek hesaplanmıştır.

7.1. Cinsiyet Sınıflandırma Deneyleri
(Gender Classification Experiments)

Konuşmacıların erkek, bayan ve çocuk olmak üzere üç grupta sınıflandırıldığı cinsiyet kategorisinde elde edilen sonuçlar Tablo 3'de verilmiştir. Bu sonuçlardan konuşma süresi ve GKM bileşen sayısındaki artışın sınıflandırma başarısını arttırdığını ancak belirli bir değerden sonraki artışın başarı üzerinde olumlu bir etkisinin olmadığı görülmektedir. En yüksek sınıflandırma başarısının elde edildiği 16 s'lik konuşmaların 64 bileşenli GKM'lerle modellendiği cinsiyet sınıflandırıcısı için karışıklık matrisi Tablo 4'te verilmiştir.

Tablo 4. Cinsiyet sınıflandırıcı için karışıklık matrisi (Ç:Çocuk, B:Bayan, E:Erkek)
(Confusion Matrix for the Gender Classifier)

	Ç	B	E
Ç	64,44	28,85	6,71
B	4,18	95,64	0,18
E	0,63	1,89	97,48

Bu sonuçlardan hatalı kararların çoğunun çocuk konuşmacıların bayan olarak sınıflandırılmasından

Tablo 5. Yaş sınıflandırma başarısı (Age Classifier Success)

Süre	GKM16	GKM32	GKM64	GKM128	GKM256
2 s	49,34	50,15	50,21	50,68	50,66
4 s	51,99	53,86	54,38	54,05	53,41
8 s	54,74	57,05	57,10	56,69	49,52
16 s	57,97	58,82	*60,10*	56,78	50,89

kaynaklandığı, yetişkin konuşmacılar arasındaki geçişin ise oldukça düşük olduğu görülmektedir. Çocuk ile bayan sınıfları arasındaki yüksek karışıklık temel frekans özelliklerindeki yakınlıkla ilişkilendirilebilir. Normal bir konuşmada çocuk ile bayan konuşmacıların ortalama temel frekans özelliğinde bir çakışma bölgesi oluşurken (208-334hz) çocuk ile erkek konuşmacılar aralıklarında böyle bir çakışma bölgesi yoktur [24]. Bu durum çalışmada elde edilen sonuçlarla benzerlik göstermektedir.

7.2. Yaş Sınıflandırma Deneyleri (Age Classification Experiments)

Konuşmacıların çocuk, genç, yetişkin ve yaşlı olarak 4 grupta sınıflandırıldığı yaş kategorisinde sadece sınıf etiketleri değiştirilerek diğer tüm işlemler cinsiyet kategorisindeki gibi uygulanmıştır. Farklı konuşma süresi ve GKM bileşen sayısı için oluşturulan yaş sınıflandırıcısının sınıflandırma başarısı Tablo 5'te verilmiştir. Bu sonuçlardan konuşma süresi ve GKM bileşen sayısının cinsiyet modeli üzerindeki etkisinin yaş modelinde de geçerli olduğu görülmektedir. En yüksek sınıflandırma başarısının sağlandığı 16 s'lik konuşmaların 64 bileşenli GKM'lerle modellendiği durum için karışıklık matrisi Tablo 6'de verilmiştir. Karışıklık matrisinden çocuk ve yaşlı konuşmacıların genç ve yetişkin konuşmacılara kıyasla daha iyi sınıflandırıldığı, en düşük sınıflandırma başarısının ise yetişkin yaş grubunda olduğu görülmektedir. Bu durum sesin ergenlik ve yaşlılık döneminde daha fazla değiştiği bilgisiyle benzerlik göstermektedir [25].

Tablo 6. Yaş sınıflandırıcı için karışıklık matrisi (Ç:Çocuk, G:Genç, Ye:Yetişkin, Ya:Yaşlı)
(Confusion Matrix for the Age Classifier)

	Ç	G	Ye	Ya
Ç	67,79	16,77	10,07	5,37
G	9,40	60,90	18,80	10,90
Ye	1,58	27,44	45,11	25,87
Ya	1,36	7,26	23,58	67,80

7.3. Yaş&Cinsiyet Sınıflandırma Deneyleri (Age-Gender Classification Experiments)

Konuşmacıların yaş ve cinsiyet özelliklerine göre 7 sınıfa ayrıldığı yaş&cinsiyet kategorisinde elde edilen sonuçlar Tablo 7'de verilmiştir. Yaş&cinsiyet kategorisinde de en yüksek sınıflandırma başarısı 16 s'lik konuşmaların 64 bileşenli GKM'lerle modellenmesi sonucunda elde edilmiştir. Bu durum için karışıklık matrisi Tablo 8'de verilmiştir.

Tablo 8. Yaş&cinsiyet sınıflandırıcı için karışıklık matrisi (Ç:Çocuk, GB:Genç bayan, GE:Genç erkek, YeB:Yetişkin bayan, YeE:Yetişkin erkek, YaB:Yaşlı bayan, YaE:Yaşlı erkek) (Confusion matrix for the age&gender classifier)

	Ç	GB	GE	YeB	YeE	YaB	YaE
Ç	69,2	11,4	4,7	6	2	6,7	0
GB	13,5	65,5	0	12,5	0	7,9	0,6
GE	1,7	0	51,5	1,7	28,7	4,3	12,1
YeB	4,4	24,1	0	44,6	0	26,9	0
YeE	0	0	25,2	0	51,1	0,7	23
YaB	2,2	6,5	0,4	25	0,5	65,4	0
YaE	1,8	0	7,1	0	21,4	1,8	67,9

Tablo 8'deki sonuçlar cinsiyet ve yaş kategorisinde elde edilen sonuçlarla benzerlik göstermektedir. Cinsiyet kategorisinde çocuk ve bayan sınıfları arasında oluşan yüksek karışıklık yaş&cinsiyet kategorisinde bayan yaş gruplarına dağılmıştır. Benzer şekilde yetişkin yaş grubunda görülen düşük başarının yaş&cinsiyet kategorisine etkisi yetişkin erkek ve yetişkin bayan sınıflarında görülmüştür. Ayrıca erkek ve bayan sınıfları arasındaki en yüksek karışıklığın genç erkek ile yaşlı bayan sınıfı arasında olması insan algılamasıyla benzerlik göstermektedir. Sonuçlar aynı veritabanının kullanıldığı [9] çalışmasıyla karşılaştırıldığında önerilen sistemin cinsiyet, yaş ve yaş&cinsiyet kategorilerinde sırasıyla %4, %7 ve %9 başarı artışı sağladığı görülmektedir.

Tablo 7. Yaş & cinsiyet sınıflandırma başarısı (Age&Gender Classifier Success)

Süre	GKM16	GKM32	GKM64	GKM128	GKM256
2 s	47,35	48,18	48,98	50,42	49,91
4 s	51,22	52,69	53,54	53,70	53,13
8 s	52,70	54,47	56,96	55,92	49,47
16 s	55,84	57,97	*60,02*	56,44	49,10
32 s	55,80	57,62	59,88	53,99	39,81

8. SONUÇLAR (CONCLUSIONS)

Bu çalışmada konuşmacıları yaş ve cinsiyet özelliklerine göre sınıflandıran GKM süpervektörlerine dayalı bir DVM sınıflandırıcısı önerilmiştir. aGender veritabanı kullanılarak geliştirilen sistemde konuşmacıların ses karakteristiklerini temsil etmek için MFCC katsayılarından oluşan 39 elemanlı bir vektör kullanılmıştır. Geliştirilen sistem 5 farklı konuşma süresi ve 5 farklı GKM ile test edilerek bu parametrelerin sınıflandırma başarısına etkisi araştırılmıştır. Bu testler sonucunda konuşma süresi arttıkça sınıflandırma başarısının arttığı ancak 16 s'den sonraki süre artışının başarı üzerinde önemli bir etkisinin olmadığı görülmektedir. Benzer şekilde GKM bileşen sayısının da sınıflandırma başarısını etkilediği görülmektedir. Bileşen sayısı arttıkça sınıflandırma başarısı artmakta ancak belirli bir değerden sonraki artışın sonuç üzerinde olumlu bir etkisi olmamaktadır. Yapılan testler sonucunda elde edilen başarı oranları ve model büyüklükleri göz önünde bulundurulduğunda yaş&cinsiyet sınıflandırıcısı için ideal konuşma süresinin 16 s, GKM bileşen sayısının ise 64 olduğu değerlendirilmiştir. Bu süre ve model büyüklüğü kullanılarak yapılan testlerde cinsiyet kategorisinde %92,42, yaş kategorisinde %60,10 ve yaş&cinsiyet kategorisinde %60,02 sınıflandırma başarısı sağlanmıştır. Gelecek çalışmalarda spektral özniteliklere ilave olarak F0, jitter, shimmer ve harmonik gürültü oranı gibi prosodik özniteliklerin yaş ve cinsiyet tanıma problemine etkisi araştırılabilir.

KAYNAKLAR (REFERENCES)

1. Schuller B., Steidl S., Batliner A., Burkhardt F., Devillers L., Müller C.A., Narayanan S.S., The INTERSPEECH 2010 paralinguistic challenge, INTERSPEECH, Makuhari-Japan, 2795-2798, 26-30 September, 2010.
2. Ceyhan E.B., Sağıroğlu Ş., Akyıl E., Gender classification based on ann with using fingerprint feature vectors, Journal of the Faculty of Engineering & Architecture of Gazi University, 29 (1), 201-207, 2014.
3. Chao W.L., Liu J.Z., Ding J.J., Facial age estimation based on label-sensitive learning and age-oriented regression, Pattern Recognition, 46 (3), 628-41, 2013.
4. Hu Y., Wu D., Nucci A., Pitch based gender identification with two stage classification, Security and Communication Networks, 5 (2), 211-225, 2012.
5. Djemili R., Bourouba R., Korba M.C.A., A speech signal based gender identification system using four classifiers, 2012 IEEE International Conference on Multimedia Computing and Systems (ICMCS), Tangier, 184-187, 10-12 May, 2012.
6. Gaikwad S., Gawali B., Mehrotra S.C., Gender Identıfıcatıon Usıng Svm Wıth Combınatıon Of Mfcc, Advances in Computational Research, 4 (1), 69-73, 2012.
7. Phoophuangpairoj R., Phongsuphap S., Two-Stage Gender Identification Using Pitch Frequencies, MFCCs and HMMs, 2015 IEEE International Conference on Systems, Man, and Cybernetics (SMC), Kowloon , 2879-2884, 9-12 Oct., 2015.
8. Metze F., Ajmera J., Englert R., Bub U., Burkhardt F., Stegmann J., Muller C., Huber R., Andrassy B., Bauer J.G., Comparison of four approaches to age and gender recognition for telephone applications, 2007 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), Honolulu-HI, IV-1089-IV-1092, 15-20 April, 2007.
9. Li M., Han K.J., Narayanan S., Automatic speaker age and gender recognition using acoustic and prosodic level information fusion, Computer Speech & Language, 27 (1), 151-167, 2013.
10. Van Heerden C., Barnard E., Davel M., van der Walt C., van Dyk E., Feld M., Müller C., Combining regression and classification methods for improving automatic speaker age recognition, 2010 IEEE International Conference on Acoustics Speech and Signal Processing (ICASSP), Dallas-TX, 5174-5177, 14-19 March, 2010.
11. Meinedo H., Trancoso I., Age and gender classification using fusion of acoustic and prosodic features, INTERSPEECH, Makuhari-Japan, 2818-2821, 26-30 September, 2010.
12. Chen O.T.C., Gu J.J., Improved gender/age recognition system using arousal-selection and feature-selection schemes, 2015 IEEE International Conference on Digital Signal Processing (DSP), Singapore, 148-152, 21-24 July, 2015.
13. Gautam S., Singh L., Developmental pattern analysis and age prediction by extracting speech features and applying various classification techniques, 2015 International Conference on Computing, Communication & Automation (ICCCA), Noida, 83-87, 15-16 May, 2015.
14. Deller J.R., Hanse J.H.L. ve Proakis J.G., Discrete-Time Processing of Speech Signals, IEEE Press, Piscataway, N.J., A.B.D, 2000.
15. Furui S., Digital Speech Processing, Synthesis and Recognition, Marcel Dekker, New York, A.B.D, 2001.
16. Reynolds D.A., Rose R.C., Robust text-independent speaker identification using Gaussian mixture speaker models, IEEE Transactions on Speech and Audio Processing, 3 (1), 72-83, 1995.
17. Dempster A.P., Laird N.M., Rubin D.B., Maximum likelihood from incomplete data via the EM algorithm, Journal of the royal statistical society. Series B (methodological), 1-38, 1977.
18. Reynolds D.A., Quatieri T.F., Dunn R.B., Speaker verification using adapted Gaussian mixture models, Digital signal processing, 10 (1), 19-41, 2000.
19. Ferras M., Leung C.C., Barras C., Gauvain J.L., Comparison of speaker adaptation methods as feature extraction for SVM-based speaker recognition, IEEE Transactions on Audio, Speech and Language Processing, 18 (6), 1366-1378, 2010.
20. Campbell W.M., Sturim D.E., Reynolds D.A., Support vector machines using GMM supervectors for speaker

verification, Signal Processing Letters, IEEE, 13 (5), 308-311, 2006.

21. Takcı H., Diagnosis of Breast Cancer by the Help of Centroid Based Classifiers, Journal of the Faculty of Engineering & Architecture of Gazi University, 31 (2), 323-330, 2016.

22. Hsu C.W., Chang C.C., Lin C.J., A practical guide to support vector classification, Department of Computer Science and Information Engineering, National Taiwan University, 2003.

23. Hsu C.W., Lin C.J., A comparison of methods for multiclass support vector machines, IEEE Transactions on Neural Networks, 13 (2), 415-425, 2002.

24. Williamson G., Human communication: A linguistic introduction (2nd edition), Billingham, UK, 2006.

25. Stathopoulos E.T., Huber J.E., Sussman J.E., Changes in Acoustic Characteristics of the Voice Across the Life Span: Measures From Individuals 4–93 Years of Age, Journal of Speech, Language, and Hearing Research, 54 (4), 1011-1021, 2011.