



Konuşmacı tanıma için eğitim algoritmalarının karşılaştırılması

Murat ÇAVDARLI^a, Ömer ESKİDERE^{b*}, Figen ERTAS^a

^aUludağ Üniversitesi, Mühendislik Mimarlık Fakültesi, Elektronik Müh., Bursa, TÜRKİYE

^bUludağ Üniversitesi, Teknik Bilimler Meslek Yüksekokulu, Mekatronik Programı, Bursa, TÜRKİYE

Anahtar Kelimeler

BM
LBG
K-ortalama
Konuşmacı tanıma
Gauss karışım modeli

ÖZET

Bu çalışmada Gauss Karışım Modeli (GKM) temeline dayanan bir konuşmacı tanıma sisteminde eğitim algoritmaları karşılaştırılmaktadır. GKM eğitim parametrelerinin kestiriminde Beklentinin Maksimumlaştırılması (BM) algoritması yaygın olarak kullanılmaktadır. Bu makalede Vektör nicemleme eğitim parametrelerinin kestirimi amacıyla kullanılan k-ortalama ve Linde, Buzo, Gray (LBG) eğitim algoritmaları GKM'ye uygulanmaktadır. TIMIT ve NTIMIT veritabanları kullanılarak BM, k-ortalama ve LBG eğitim algoritmalarının konuşmacı tanıma performansları karşılaştırılmaktadır. Ayrıca model başlangıç değerlerine karşı hassas olan BM ve k-ortalama algoritmalarının veritabanları için ideal başlangıç değerleri belirlenmektedir.

A comparison of training algorithms in speaker identification

Keywords

BM
LBG
k-means
Speaker identification,
Gaussian mixture model

ABSTRACT

In this study, training algorithms are compared in Gaussian mixture model (GMM) based a speaker identification system. The Expectation maximization (EM) algorithm has widely been used to estimation of GMM parameters. In this article, the k-means and LBG are applied to GMM in order to estimate the vector quantization training parameters. The EM, the k-means and LBG training algorithms are tested with TIMIT and NTIMIT databases and are compared speaker identification performance. Furthermore, the EM and k-means algorithms which sensitive against model initialization values are found optimum model initialization values for databases.

* Sorumlu yazar (corresponding author) e-posta: oeskidere@uludag.edu.tr

1. GİRİŞ

Önceden duyduğumuz konuşmaların sonraki karşılaşmalarda kime ait olduklarını rahatlıkla hatırlayabiliriz. Telefonda konuşurken, telefon hattında gürültü olsa bile pek çok zaman karşıdaki kişiyi tanıyabiliriz. Özel olarak konuşan kişinin kimliğini bulmak için kullanılan diğer ipuçları hatalı veya çok belirsiz olduğu durumlarda ses ile konuşan kişiyi tanıma oldukça çok kullanılan bir yöntemdir [1, 2].

İnsan ses üretme mekanizmasının modellenmesi, ses sinyalinin doğasını anlamaya yardımcı olduğu kadar ses üretiminde de doğrudan gereklidir. İnsan algı mekanizmasındaki seçiciliğin modellenmesi, kişinin sesini karakterize eden öznitelikler ile mümkün olabilir [3, 4, 5]. Bu özniteliklerdeki değişimler Gauss karışım modeli (GKM) ile modellenebilir [6].

Farklı konuşmacılar ve farklı diller kendi olasılık yoğunluklarına sahiptirler. Bu olasılık yoğunluklarla konuşmacıların kim olduğunu tespit edebiliriz. Bu yöntemin temel prensibi kişisel akustik niteliklerin olasılık yoğunluk fonksiyonunun birden çok Gauss yoğunluk fonksiyonu kullanılarak gösterimidir. Bütün konuşmacı tanıma sistemleri birbirinden bağımsız iki aşamaya hizmet etmek zorundadır. Bunlardan ilki eğitim aşaması iken ikincisi ise test aşamasıdır.

Eğitim aşamasında tüm kullanıcılar, bir referans modeli oluşturmak için ses örnekleri verir, ikinci aşamada ise giriş sinyali referans modelleri ile karşılaştırılarak saptama yapılır [7]. Konuşmacı tanıma sisteminin blok diyagramı şekil 1'de görülmektedir.

2. GEREÇ VE YÖNTEM

2.1 Konuşmacı Modeli

GKM içindeki Gauss bileşenlerin her biri ile spektral yapı olarak bilinen geniş fonetik sınıflar kolayca karakterize edilir. Bu fonetik sınıflar bazı konuşmacı bağımlı ses yolu yapılarını yansıtır, konuşmacı kimlik modellenmesinde kullanılır [6]. Ayrıca Gauss karışım yoğunluğu, bir konuşmacıdan alınan sözcüklerle gözlemlerin uzun süreli dağılımında düzgün bir yaklaşım sağlamaktadır [8].

Bir Gauss karışım yoğunluğu, M bileşenli yoğunlukların toplamının ağırlıklandırılması olup denklem 1'deki gibi ifade edilir.

$$p(\bar{x} / \lambda) = \sum_{i=1}^M p_i b_i(\bar{x}) \quad (1)$$

Burada her bir konuşmacının GKM'si λ ile gösterilmektedir. \bar{x} , D boyutlu rastgele değişen vektör, $b_i(\bar{x})$, bileşen yoğunlukları ($i = 1, \dots, M$) ve p_i , karışım ağırlıklarıdır. Her bir bileşen için D boyutlu Gauss fonksiyonu denklem 2'de görülmektedir [9].

$$b_i(\bar{x}) = \frac{1}{(2\pi)^{D/2} |\Sigma_i|^{1/2}} \exp\left\{-\frac{1}{2}(\bar{x} - \bar{\mu}_i)' \Sigma_i^{-1} (\bar{x} - \bar{\mu}_i)\right\}, \quad (2)$$

Burada $\bar{\mu}_i$ ortalama vektör ve Σ_i ortak değişinti matrisidir. Karışım ağırlıkları $\sum_{i=1}^M p_i = 1$ şeklinde sınırlandırılır. Gauss karışım modeli, her bileşenin ortalama vektörü, ortak değişinti matrisi ve karışım ağırlık değerleri olarak denklem 3'deki gibi ifade edilmektedir.

$$\lambda = \{p_i, \bar{\mu}_i, \Sigma_i\} \quad i = 1, \dots, M \quad (3)$$

2.2 BM Algoritması

BM algoritması, istatistiksel veri analizi [10], konuşma tanıma [11], gürültünün kaldırılması [12] gibi pek çok alanda kullanılmaktadır. BM algoritmasının geniş kullanılmasının nedeni her bir özyinelemeden sonra benzerlik fonksiyonu artışını garanti edip pek çok karışık kestirim problemleri için güçlü yapıya sahip olmasıdır. BM algoritmasının temelindeki iddia ilk model başlangıcı, yeni model $\bar{\lambda}$, $P(X|\bar{\lambda}) \geq P(X|\lambda)$ olarak kestirilir. Eski model yerine yeni model yerleştirilir bu işlem ve yakınsama süreci eşik değerine ulaşılan kadar devam edilir.

Maksimum benzerlik GKM parametreleri, BM algoritmasının özel bir hali olan döngüsel parametre kestirim yolu ile bulunur. GKM parametrelerinin maksimum benzerlik yöntemi ile kestirimi için Baum yaklaşım fonksiyonu kullanılmaktadır [6]. Bu şekilde

Erciyes Üniversitesi Fen Bilimleri Enstitüsü Dergisi 26(2): 143-153 (2010)

denklem 4’de verilen sonsal olasılık parametresi elde edilmektedir.

$$p(i_t = i | \bar{x}_t, \lambda) = \frac{p_i b_i(\bar{x}_t)}{\sum_{k=1}^M p_k b_k(\bar{x}_t)} \quad (4)$$

Burada i . durumun sonsal olasılığı verilmektedir. Bu kestirim denklemi özyinelemeli parametre kestirim prosedürünün temelini oluşturur. Karışım ağırlıkları, $\sum_{i=1}^M \bar{p}_i = 1$ sınırlama şartları altında denklem 5’deki gibi elde edilmektedir.

$$\bar{p}_i = \sum_{t=1}^T p(i_t = i / x_T, \lambda) \quad (5)$$

Burada $p(i_t = i | \bar{x}_t, \lambda)$ denklem 4’ de verilmektedir. Bileşen yoğunluk ortalamaları denklem 6’daki gibidir.

$$\hat{\mu}_i = \frac{\sum_{t=1}^T p(i_t = i / \bar{x}_T, \lambda) \bar{x}_T}{\sum_{t=1}^T p(i / \bar{x}_T, \lambda)} \quad (6)$$

Köşegen ortak değişinti matrisi denklem 7 deki gibidir.

$$\hat{\sigma}_i^2 = \frac{\sum_{t=1}^T p(i_t = i / \bar{x}_T, \lambda) x_T^2}{\sum_{t=1}^T p(i_t = i / x_T, \lambda)} - \bar{\mu}_i^2 \quad (7)$$

Yukarıda tanımlanan denklem (5), (6) ve (7) kullanılarak GKM parametreleri özyinelemeli olarak BM algoritmasından kestirilir. BM algoritması aşağıdaki adımlar ile özetlenebilir [13].

Başlangıç: $\lambda^{(0)} = \{p_i, \bar{\mu}_i, \sigma_i^2\}$ model başlangıç parametreleri belirlenir.

B-Adımı: $\bar{\lambda}$ yeni model parametreleri, denklem 4’de verilen sonsal olasılık parametresi ve λ model parametrelerinin denklem (5), (6) ve (7) da kullanılması ile kestirimi yapılır.

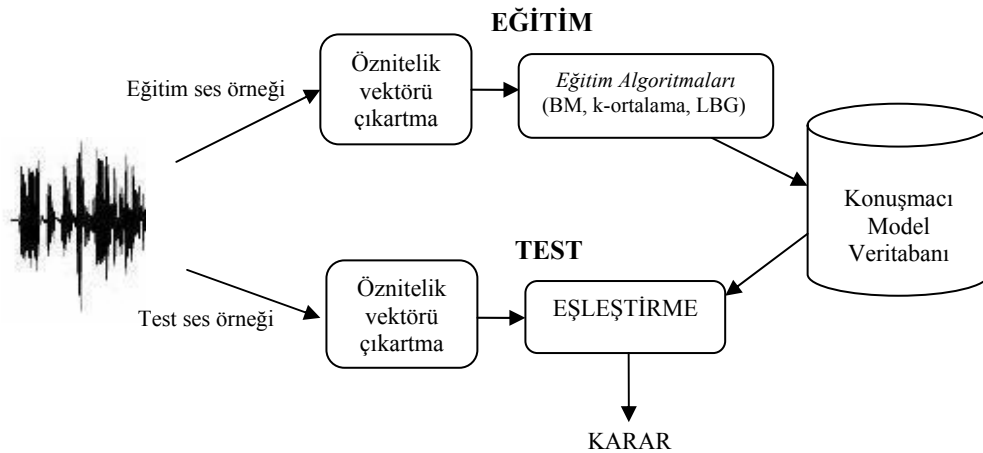
M-adımı: Şu andaki model parametreleri, bulunan yeni model parametreleri ile yer değiştirilir:

Özyineleme: Benzerlik fonksiyonu sabitleştiği bir noktaya kadar B ve M adımlarına devam edilir. Özyinelemeler arasındaki değişim, denklem 8 kullanılarak bulunabilir.

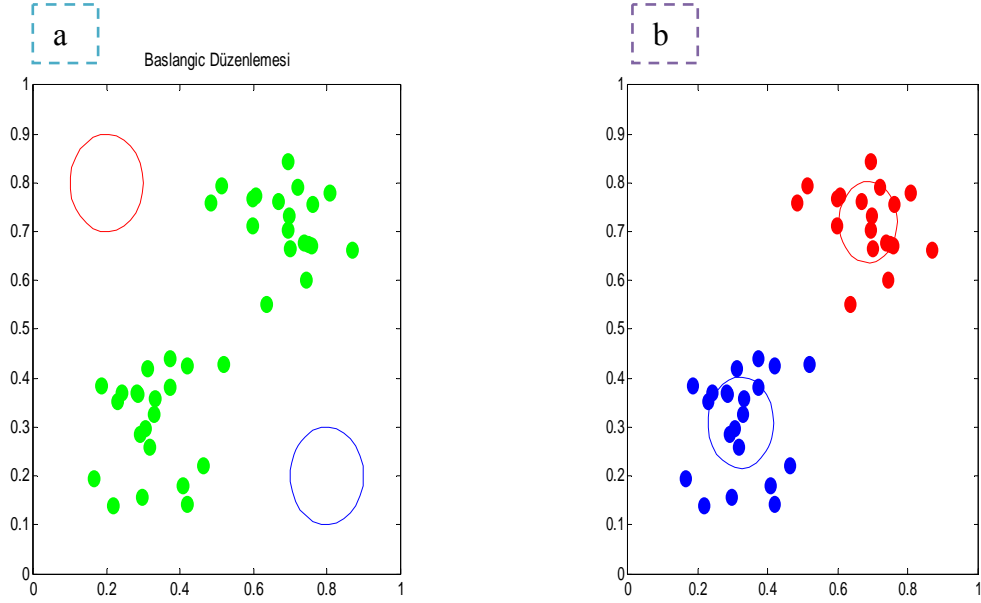
$$\hat{E} - E = -\sum_n \ln \left\{ \frac{p(x / \bar{\lambda})}{p(x / \lambda)} \right\} \quad (8)$$

Kestirilen denklemden elde edilen yeni model parametreleri, benzerlik fonksiyonunda monoton artışı garanti eder, algoritma benzerlik fonksiyonunun sabitleştiği bir noktaya yakınsamayı sağlar [10].

GKM’nin parametre modelleme kapasitesi, modelin keyfi bir şekilde dağılmasına imkân tanır. GKM’nin olasılık temelli yapısı, olasılık ve istatistik teorisinde güçlü teknikler sayesinde pek çok uygulama oluşturmasına imkân verir. Şekil 2’de iki boyutlu düzlemde noktalarla gösterilen verilere BM algoritmasının uygulanması görülmektedir. İki Gauss karışım dağılımı ile 15 özyineleme sonunda ortalamaların yakınsaması görülmektedir.



Şekil 1. Konuşmacı tanıma sisteminin genel yapısı



Şekil 2. Küme sayısı iki için BM algoritması ile ortalamaların (a) başlangıç durumunda (b) özyinelemeler sonunda yakınsaması

2.3 K-ortalama Algoritması

K-ortalama algoritması, kümeleme problemini çözen en basit gözetimsiz öğrenme algoritmalarından biridir [14]. K-ortalama algoritması, adından da anlaşılabilceği gibi öznelik vektörlerini k adet merkezle ifade etmeye çalışan bir yöntemdir. Merkezle ilk değer ataması yapıldıktan sonra merkez değerlerinin güncellenmesi için iki farklı yöntem kullanılır. Birinci yöntemde giriş kümesindeki her bir örneğin hangi merkeze yakın olduğu hesaplanır. Aynı merkeze yakın olan örneklerin ortalaması alınarak merkezin değeri güncellenmiş olur. Durma koşulu sağlanana kadar bu işlem tekrar edilir. İkinci yöntemde giriş kümesinden bir örnek seçilir ve bu örneğin merkezlere olan uzaklığına bakılır. Örneğin en yakın olduğu merkez bulunarak bu merkezin değeri güncellenir. Her bir örnek için bu işlem tekrarlanır. Merkez değeri güncellenirken merkezle örnek arasındaki mesafe değeri her nokta azalan bir öğrenme katsayısıyla çarpılarak kullanılır. Amaç, gerçekleştirilen bölümlenme işlemi sonunda elde edilen kümelerin, küme içi benzerliklerinin maksimum ve kümeler arası benzerliklerinin minimum olmasını sağlamaktır. Küme benzerliği, kümenin ağırlık merkezi olarak kabul edilen bir nesne ile kümedeki diğer nesnelere arasındaki uzaklıkların ortalama değeri ile ölçülmektedir [15]. Bu çalışmada birinci yöntem

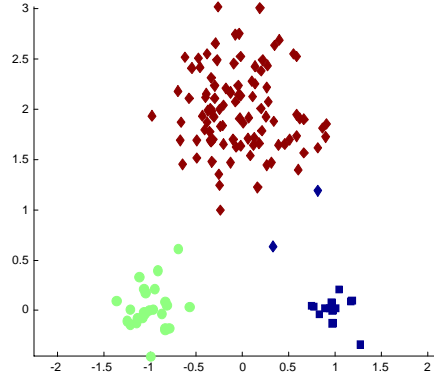
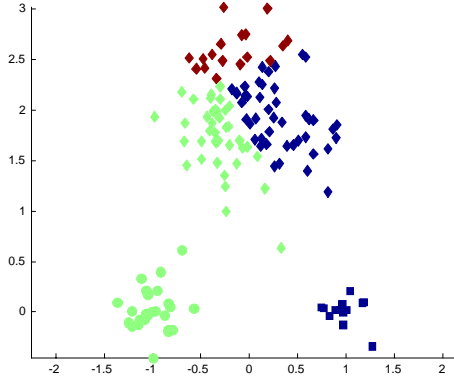
kullanıldı ve bu yöntemde izlenen adımlar şöyledir[5];

1. İlk değer atama: İlk küme merkezleri belirlenir.
2. Giriş kümesindeki her bir örneğin merkez noktalara olan uzaklığı hesaplanır ve elde edilen sonuçlara göre en yakın kümeye yerleştirilir.
3. Oluşan kümelerin yeni merkez noktaları o kümedeki tüm örneklerin ortalaması alınarak güncellenir. Konuşmacı $S_i, i=1,2,\dots,k$ için $\mu_i, x_j \in S_i$ $X = \{\vec{x}_1, \vec{x}_2, \dots, \vec{x}_j\}$ küme vektörlerinin ortalama vektörleri $\{\vec{\mu}_1, \vec{\mu}_2, \dots, \vec{\mu}_i\}$, denklem 9'daki gibi bulunur.

$$\mu_i = \frac{1}{n} \sum_{x_j \in S_i} x_j \quad (9)$$
4. Merkez noktaları değişmeyene kadar bu işlem tekrarlanır. Yeni ortalama değerler bulunur.
5. GKM ağırlık değerleri kümeye atanan ortalama değerlerden, GKM ortak değişinti matrisi ise kümeye atanan öznelik vektörlerin ortak değişintilerinden bulunur [16].

K-ortalama algoritmasında her bir nesnenin merkez noktalara uzaklığını hesaplamak için kullanılan birçok metrik yöntem bulunmaktadır. Bu çalışmada k-ortalama algoritmalarında Öklit uzaklık formülü kullanıldı. Şekil 3 de noktalarla gösterilen iki boyutlu

öz nitelik vektörlerinin, küme sayısı 3 için, 6 özyineleme sonunda elde edilen ortalama değerleri görülmektedir



Şekil 3. Küme sayısı 3 için (a) k-ortalama algoritması ile ortalamaların başlangıç durumu (b) özyinelemeler sonunda yakınsaması

2.4 LBG Algoritması

LBG algoritmasında, uzaydaki bir grup noktanın tek bir noktaya uzaklıkları toplamının minimum olması için o tek nokta grup merkezinde alınır. LBG algoritması birçok karakteristiği k-ortalama algoritmasıyla benzerdir. LBG algoritması daha çok Vektör Nicemleme (VN) amacıyla geliştirilmiştir [4]. VN’de, M sayıdaki vektörü N sayıdaki vektör ile temsil etmektedir. Temsil eden vektörlere kod vektörü denilmektedir. Her kod vektörü temsil ettiği vektör gruplarının merkezinde seçilmektedir. LBG algoritması en çok bilinen ve çalışılan kod kitabı oluşturma algoritmalarından biridir. LBG algoritması öz yinelenmeli bir algoritma olup algoritmanın adımları şunlardır [16, 17];

1. Bütün eğitim vektörlerinin ağırlık merkezi bulunur.
2. Bütün eğitim vektörleri en yakın sınıfa konur. (Başlangıçta tek sınıf olduğundan bütün vektörler o sınıfa konur.)
3. Her sınıfın ağırlık merkezi hesaplanır. Bu ağırlık merkezi o sınıfı temsil edecek kod vektördür.

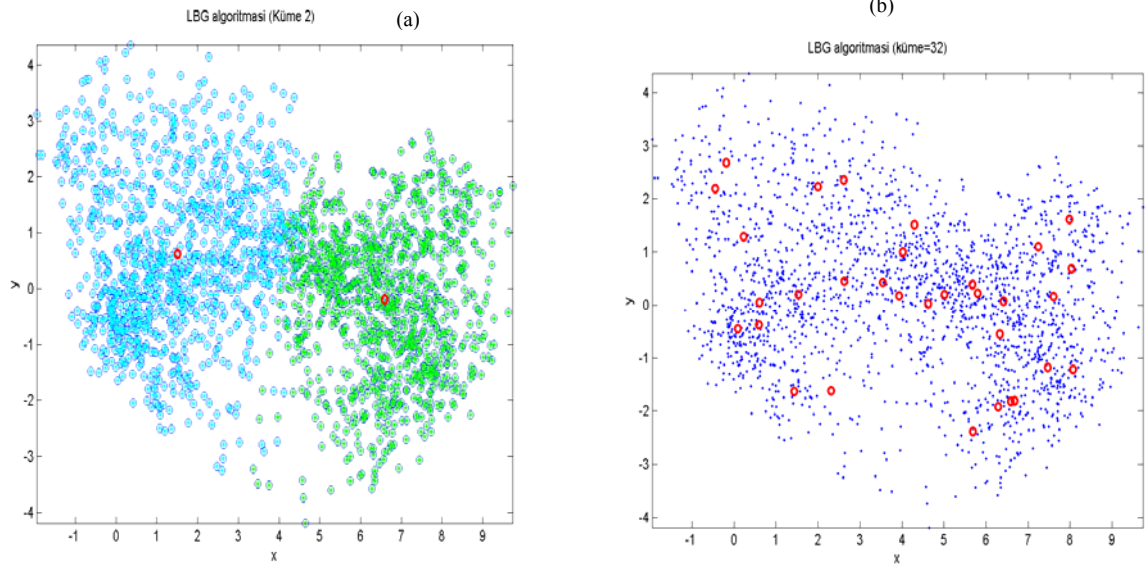
$$\bar{\mu}_i^1 = \frac{1}{n} \sum_{x_j \in S_i} x_j \quad (10)$$

4. Toplam bozulma hesaplanır. Toplam bozulma miktarı ile bir önceki yinelemedeki bozulma miktarı arasındaki fark büyükse 2. adıma dönülür.
5. Kod vektör sayısı yeterli ise özyineleme durdurulur.
6. Her kod vektör ikiye bölünür.

$$\bar{\mu}_1^2 = \bar{\mu}_i^1 + \varepsilon \quad \bar{\mu}_2^2 = \bar{\mu}_i^1 - \varepsilon \quad (11)$$

Bütün vektörlerin en yakın sınıfta olması algoritma sonucunun veya her adımdaki sonucun optimum olmasını garanti etmez. Bu sonuçlar başlangıç sınıf vektörünün ağırlık merkezinin farklı yerde olmasına veya kod vektörlerinin ikiye ayrılması sırasında eklenen veya çıkarılan ε değerine bağlı olarak değişir.

Ortalama değerleri bunduktan sonra, GKM ağırlık değerleri kümeye atanan ortalama değerlerden, ortak değişinti matrisi ise kümeye atanan öz nitelik vektörlerin ortak değişintilerinden bulunur [16]. İki boyutlu düzlemde küme sayısı 32 için LBG algoritmasının yakınsaması şekil 4’de görülmektedir.



Şekil 4. Küme sayısı 32 için LBG algoritması ile ortalamaların (a) başlangıç durumunda (b) özyinelemeler sonunda yakınsaması

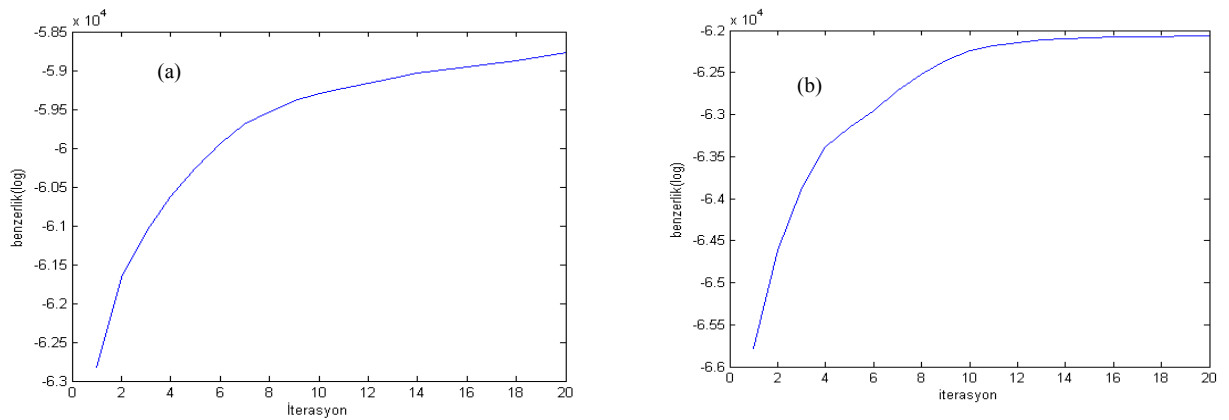
2.5 Model eğitimi aşamasında yapılan ayarlamalar

GKM modelinde, konuşmacıların eğitiminde bazı ayarlamalar yapılması gerekmektedir ki bunlar, GKM konuşmacı model eğitiminde BM algoritmasının yakınsama miktarı, BM ve k-ortalama algoritmalarının model başlangıç değerlerinin ayarı ve eğitim aşamasında model tekilliklerini önlemek için değışinti sınırlanmasıdır.

2.5.1 BM algoritmasının özyineleme sayısı

BM parametre kestirimi, benzerlik fonksiyonunun maksimum olduğu model parametre değerlerinin

bulunmasıdır. BM algoritmasında her bir özyineleme benzerlik fonksiyonunun artışı sağlar. Özyineleme sayısı pratik anlamda benzerlik fonksiyonunun yeterli oranda yakınsayıp yakınsamadığını bulmak için gereklidir. Gauss karışım modeli kullanılarak, bir konuşmacının eğitilmesinde BM algoritmasının 20 özyineleme için benzerlik fonksiyonunun değışimi şekil 5’de görülmektedir. Şekillerde Gauss karışım bileşen sayısı 32 ve 16 alınmaktadır. TIMIT veritabanında her bir konuşmacı, 8 cümle’ye karşılık gelen yaklaşık 2400 öznitelik vektörü kullanılarak eğitilmektedir.



Şekil 5. GKM eğitim için BM algoritmasının benzerlik fonksiyonunun (a) karışım sayısı 32 (b) karışım sayısı 16 için değışimi

Şekil 5’den görüleceği üzere, BM algoritmasına ait benzerlik fonksiyonu, alabileceği maksimum değerin % 90’ına ilk 5 özyineleme içerisinde ulaşmakta ve 15 özyineleme içerisinde belli bir değere yakınsamaktadır. Bu yakınsama, konuşmacı değişimlerinden, model başlangıcı için kullanılan yöntemlerden, karışım bileşen sayısı ve eğitilen veri miktarından bağımsızdır [6]. Şekillerden de görüleceği üzere deneylerde BM özyineleme sayısı 15 alınması yeterlidir.

2.5.2 Model başlangıç değerleri

Gauss karışım modelinin eğitim safhasında başlangıç model değerlerine sahip olması gerekir. BM algoritması başlangıç değerlerinden (λ^0) bağımsız lokal maksimum benzerlik değerlerini garanti eder. Fakat GKM için benzerlik denklemi birkaç lokal maksimum değer ve farklı başlangıç değerine sahip olup, farklı lokal maksimum değere yönelebilir [9]. BM ve k-ortalama eğitim algoritması kullanılarak farklı başlangıç şartlarında konuşmacı tanıma performansı etkilenme oranı bilinmemektedir. Bu sorunun çözümü için deneyler kısmında konuşmacı modeli farklı başlangıç şartlarında eğitilip konuşmacı tanıma oranı ölçülecektir.

2.5.3 Değişinti sınırlanması

Gauss karışım modeli, öznitelik vektörleri kullanılarak eğitilirken değişinti vektörünün belirli bileşenlerinin oldukça küçük genlikte (sıfıra yakın) olduğu gözlenmektedir [6]. Bu durum özellikle Gauss karışım bileşen sayısı büyük olduğu (≥ 32) durumlar için geçerlidir. Gözlenen bu küçük değişintiler modelin benzerlik fonksiyonundaki tekilliklerin sonucudur. Bu durum maksimum benzerlik sınıflandırıcısı kullanıldığı durumlarda konuşmacı modelinde bozulmalar meydana getireceğinden konuşmacı tanıma performansında düşmeler olur. Bu tekillikler eğitim için yeterli sözcük olmadığı durumlarda artar. Aynı durum, model çok fazla karışım bileşen sayısına sahip olduğu durumlarda da gözlenir. Ayrıca tekillikler, telefon veya gürültülü konuşma ortamlarında veri kırılması olduğu durumlarda da oluşabilir [18].

Modelin eğitimi esnasında oluşan model değişinti değerlerinin sıfıra yönelmesini önlemek için sabit değişinti sınırlaması uygulanır. Herhangi bir i . karışım bileşeninin değişinti vektörü, σ_i^2 , olmak üzere değişinti sınırlaması denklem 12’deki gibi ifade edilir.

$$\overline{\sigma_i^2} = \begin{cases} \sigma_i^2 > \sigma_{\min}^2 \Rightarrow \sigma_i^2 \\ \sigma_i^2 \leq \sigma_{\min}^2 \Rightarrow \sigma_{\min}^2 \end{cases} \quad (12)$$

burada σ_{\min}^2 , minimum değişinti değeri olup elde edilen $\overline{\sigma_i^2}$ değişinti değeri her özyineleme için bulunur.

Minimum değişinti değerinin çok yüksek seçilmesi durumunda, bileşen değişintileri aynı σ_{\min}^2 değeri ile sınırlandırılacak, buna bağlı olarak tanıma performansında düşme olacaktır [6]. Bu değer için çok küçük alınması durumunda da değişinti sınırlaması arzulan işlevini göremeyecektir. Bu durumda en ideal minimum değişinti değeri denenecek bulunur. Yapılan deneylerde $\sigma_{\min}^2 = 0.001$ alınması daha iyi model kestirimine imkân tanımaktadır [9].

2.6 Konuşmacıların test edilmesi

Test aşaması konuşmacı test kayıtlarını kullanarak öznitelik vektörlerinin elde edilmesi ve bulunan öznitelik vektörleri ile eğitim aşamasında oluşturulan her bir Gauss karışımının karşılaştırılması sürecidir. Herhangi bir konuşmacının GKM yönteminde sınanması iki aşamadan oluşmaktadır. İlk aşamada söz konusu konuşmacıya ait sınama yöntemlerinin tamamının tanıtılmış tüm konuşmacıların her birine ait olma olasılıkları hesaplanır. İkinci aşamada ise bulunan bu olasılıkların en büyüğüne sahip konuşmacı sınama kümesine sahip konuşmacı olarak belirlenir.

S konuşmacı için $S = \{1, 2, \dots, S\}$ $\lambda_1, \lambda_2, \lambda_3, \dots, \lambda_s$ ler ile temsil edilmiştir. Amaç verilen bir gözlem dizisi $X = \{\bar{x}_1, \bar{x}_2, \bar{x}_3, \dots, \bar{x}_t\}$ için maksimum sonsal olasılığına sahip olan konuşmacı modelini bulmaktır. Bayes karar kuralı kullanılarak denklem 13 yazılabilir [9].

$$\hat{S} = \arg \max_{1 \leq k \leq S} \Pr(\lambda_k / X) = \arg \max_{1 \leq k \leq S} \frac{p(X / \lambda_k) \Pr(\lambda_k)}{p(X)} \quad (13)$$

Tüm konuşmacıların eşit olasılıklı olması $\Pr(\lambda_k) = 1/S$ ve $p(X)$ değerinin de tüm konuşmacılar için aynı olması varsayıldığında denklem 13 basitleştirilirse denklem 14 elde edilir.

$$\hat{S} = \arg \max_{1 \leq k \leq S} \Pr(X / \lambda_k) \quad (14)$$

Denklem 14’ün logaritması alınırsa denklem 15 elde edilir.

$$\hat{S} = \arg \max_{1 \leq k \leq S} \sum_1^T \log p(\bar{x}_i / \lambda_k) \quad (15)$$

Aday konuşmacıya ait test sözcüklerinden elde edilen öznelik vektörleri, denklem 15 e uygulanır. Bu işlem her bir aday konuşmacı için yapılır ve sonuçta maksimum olasılıklı aday konuşmacı denklem 15 kullanılarak bulunur.

3. DENEYSEL SONUÇLAR

3.1 Veritabanları ve Spektral Analiz

Konuşmacı tanıma amacıyla kullanılan ses örnekleri TIMIT ve NTIMIT veritabanlarına aittir. TIMIT veritabanındaki kişilerin ses örneklerinde, akustik gürültü, konuşmacı kayıtlarında oluşan oturumlar arası konuşmacı sesi değişimi ve mikrofondan dolayı seste bozulma gözlenmez. Bu nedenle temiz konuşma veritabanı olarak adlandırılır. NTIMIT veritabanı, TIMIT veritabanındaki cümlelerin karbondan yapıma telefon ahizesi üzerinden bir yerel veya uzun mesafe merkez ofise iletilip ve aynı hat üzerinden tekrar kayıt için geri alınmış halidir [19].

Öznelik vektörleri elde edilmesinde Mel frekansı kepstum katsayıları (MFCC) kullanılmaktadır. MFCC elde edilmesinde çerçevelerin örtüşme oranı 10 ms alınır, çerçevelere Hamming pencereleme uygulanmaktadır. Pencereleyen sesin 512 örnek FFT'si alınır, Slaney tarafından tanımlanan Mel ölçeğe [20], üçgen süzgeç dizilerinden geçirilir. Süzgeçten geçirilen işaretin logaritması alınır ve ayrık kosinüs dönüşümü alınmıştır Her bir çerçeveye karşılık olarak TIMIT veritabanı için 24, NTIMIT veritabanı için 20 boyutlu öznelik vektörleri kullanılmaktadır. NTIMIT veritabanı için konuşmadan sessiz kısımlar atılmaktadır[21].

Tüm deneylerde, TIMIT ve NTIMIT veritabanlarının test dizininden 168 kişinin her birine ait 10 cümleden 8'i (yaklaşık 24 saniye) eğitim için, kalan 2 cümleden her biri (yaklaşık 3 saniye) test için kullanılmaktadır. Gauss karışım sayısı 32 alınır, BM algoritması için özyineleme sayısı 15, k-ortalama algoritması için özyineleme sayısı 100 alınmaktadır. Değişinti sınırlaması olarak $\sigma^2_{\min} = 0.001$ değeri kullanılmaktadır. [13, 22].

3.2 Deneyler

Konuşmacı tanıma için BM, k-ortalama ve LBG algoritmaları karşılaştırılacaktır. Ancak BM ve k-ortalama eğitim algoritmaları, model başlangıç değeri değişiminden önemli oranda etkilenmektedir. Bu nedenle bu iki eğitim algoritması değişik model başlangıç durumları için test edilecektir. Son olarak

bulunan en iyi başlangıç şartlarında BM, k-ortalama ve LBG algoritmaları karşılaştırılacaktır.

1. K-ortalama algoritmasının başlangıç durumlarına göre konuşmacı tanıma başarımları incelenecektir. Tüm konuşmacılar eğitilirken başlangıç verilerine bağlı olarak k-ortalama algoritması ile ortalama değerler bulunmakta, elde edilen sonsal olasılık değerlerine bağlı olarak ağırlık ve ortak değişinti matrisleri elde edilmektedir. Her bir konuşmacı ; $\lambda = \{p_i, \mu_i, \Sigma_i\}$ ve $i = 1, \dots, M$ olmak üzere ağırlık, ortalama ve ortak değişinti matrislerinden oluşan M adet karışım bileşen sayısı kullanılarak modellenmektedir.

K-ortalama algoritması için model başlangıç değerinin belirlenmesinde 4 değişik yöntem kullanılacaktır. Ortalama vektöre en yakın öznelik verileri (OEY), ortalama vektöre en uzak öznelik verileri (OEU), rasgele değişen veriler (RD) ve LBG algoritması başlangıç değeri belirlenmesinde kullanılacaktır. Vektör nicemeleme yöntemiyle k-ortalama algoritmasının, başlangıç değeri belirlenmesinde bu yöntemler yaygın olarak kullanılmaktadır. K-ortalama algoritması için bu başlangıç yöntemleri Gauss karışım modelinde denenecektir. Şekil 6'da iki boyutlu öznelik verileri için k-ortalama algoritmasında kullandığımız başlangıç yöntemleri görülmektedir

NTIMIT ve TIMIT veritabanları için bölüm 3.1'de belirlenen şartlarda, konuşmacıların yalnız k-ortalama algoritması ile eğitilip 4 farklı başlangıç yöntemi için elde edilen konuşmacı tanıma oranları tablo 1'de görülmektedir.

Tablo 1. K-ortalama algoritması başlangıç yöntemlerine göre tanıma oranları (%)

Başlangıç yöntemi	NTIMIT	TIMIT
OEY	71.1	99.7
OEU	67.6	97.9
RD	69	99.1
LBG	69.9	99.4

K-ortalama algoritmasının Gauss karışım modeline uygulanması sonucu her iki veritabanı içinde en iyi tanıma başarımları, OEY başlangıç yöntemi için elde edilmektedir. Başlangıç yöntemi rastgele değişen veriler kullanıldığı durumda elde edilen tanıma başarımlarında NTIMIT veritabanı için yaklaşık 73 puan, TIMIT veritabanı için ise yaklaşık 70.6 puan değişim gözlenmektedir. Bu durum modelin başlangıç durumu değişimlerine karşı hassasiyetini göstermektedir.

2. BM algoritmasının başlangıç değerleri değişimine göre tanıma başarımı incelenecektir. BM algoritmasının başlangıç değerleri bulunurken 6 farklı yöntem test edilmektedir. İlk 4 yöntemde tablo 1’de verilen başlangıç yöntemleri, k-ortalama algoritması için ayrı ayrı başlangıç değeri olarak alınıp k-ortalama algoritması ile bu başlangıç değerlerine bağlı olarak konuşmacılar eğitilmektedir. Elde edilen ortalama değerleri BM algoritmasının başlangıç değeri olarak kullanılmaktadır. Denklem 4’de verilen sonsal olasılık parametreleri hesaplanıp denklem 5,6, 7 de yerine koyularak model parametreleri olarak ağırlık ortalama ve ortak değişinti matrisi elde edilmektedir. Ayrıca BM algoritması başlangıç değerleri olarak RD ve LBG algoritması da test edilmektedir. Tablo 2’de iki farklı veritabanı için bölüm 3’de belirtilen şartlarda, BM eğitim algoritmasının başlangıç yöntemlerine bağlı olarak elde edilen konuşmacı tanıma oranları görülmektedir.

Tablo 2. BM algoritması başlangıç yöntemlerine göre tanıma oranları (%)

Başlangıç yöntemi	NTIMIT	TIMIT
k-ort.+OEY	73.5	99.4
k-ort.+OEU	68.4	98.5
k-ort.+RD	72	99.1
k-ort.+LBG	74.1	99.4
RD	68	99.1
LBG	71.4	99.4

BM algoritması ile konuşmacıların eğitildiği NTIMIT veritabanı için; başlangıç algoritması olarak LBG algoritması ile ortalamalar bulunup bu ortalamaların k-ortalama algoritması tarafından kullanıldığı durumda en yüksek başarımlar elde edilmektedir. TIMIT veritabanı için başlangıç yöntemi değişimi tanıma başarımını çok etkilememektedir. Başlangıç yöntemi rasgele değişen veriler kullanıldığı durumlarda elde edilen tanıma başarımlarında, NTIMIT veritabanı için yaklaşık ∓ 2 puan, TIMIT veritabanı için ise yaklaşık ∓ 0.6 puan değişim gözlenmektedir.

3. Konuşmacı tanıma için BM, K-ortalama ve LBG eğitim algoritmalarının karşılaştırılması yapılacaktır. K-ortalama için deney 1 ve BM için deney 2’de elde edilen en iyi başlangıç durumları kullanılarak üç eğitim algoritması karşılaştırılmaktadır. Tablo 3’de NTIMIT ve TIMIT veritabanları için bölüm 3.1’de belirtilen şartlarda BM, k-ortalama ve LBG algoritmasının konuşmacı tanıma başarımları görülmektedir.

Tablo 3. Üç farklı eğitim algoritmasının tanıma oranları (%)

Eğitim algoritması	NTIMIT	TIMIT
BM	74.1	99.4
k-ort.	71.1	99.7
LBG	67	99.1

Tablo 3’den görüleceği üzere üç eğitim algoritması karşılaştırıldığında, en yüksek başarımları NTIMIT veritabanı için BM algoritması, TIMIT veritabanı için ise k-ortalama algoritması ile elde edilmektedir. K-ortalama algoritması, hata parametrelerinin değerini en az yapmak için büyük kümeleri bölerek mümkün olduğunca birbirinden ayırık ve kendi içinde sıkışık kümeler bulmaya çalışır. LBG algoritması ise, en ideal nicemleme bölgelerinin belirlenmesi için kullanılır. LBG tüm kümeler için her bir özyinelemede ortalamayı hesaplamayabilir. Bu nedenle elde edilen sonuçlardan görüleceği üzere, GKM için LBG algoritması konuşmacı tanımadaki tek başına eğitim algoritması olarak verimli olmamaktadır.

4. SONUÇLAR

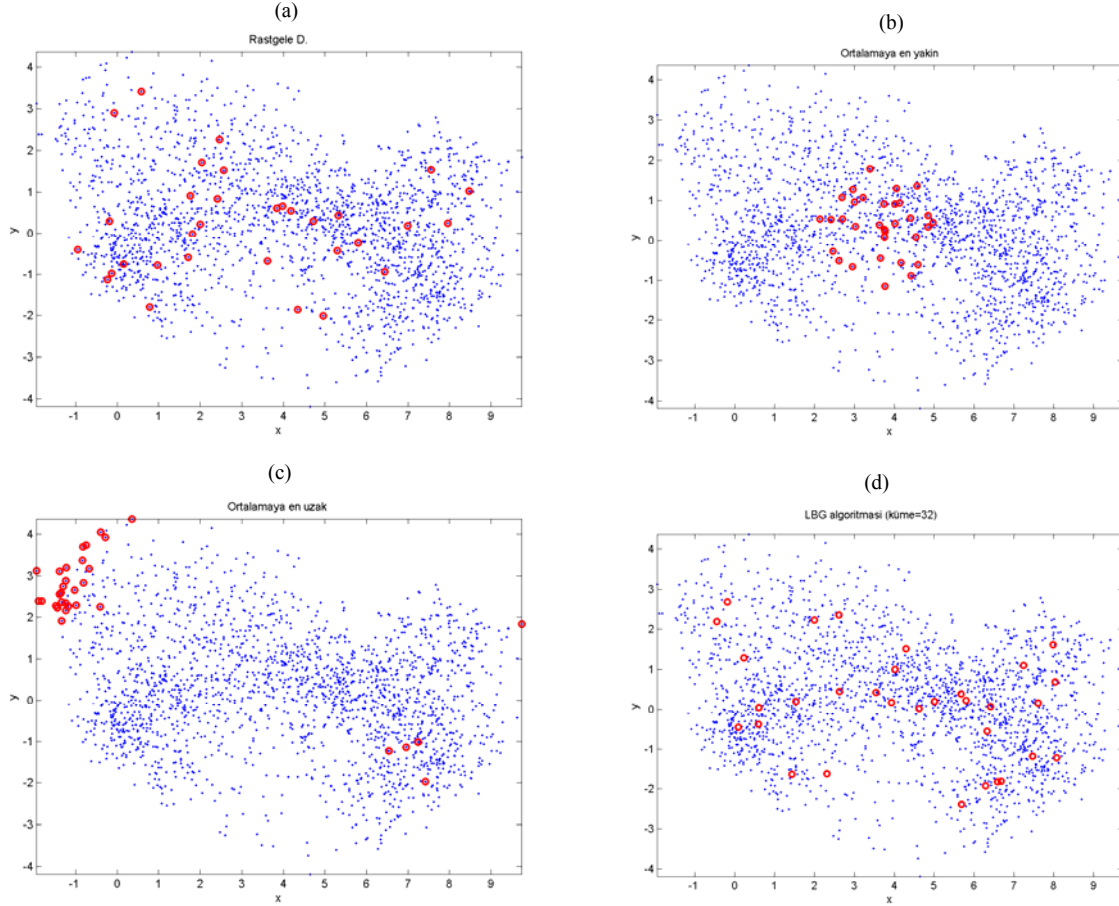
Bu makalede, vektör nicemleme yöntemi ile birlikte kullanılan k-ortalama ve LBG algoritmaları GKM’ye uygulanmıştır. Gauss karışım modeli kullanılarak BM, k-ortalama ve LBG eğitim algoritmalarının konuşmacı tanıma performansları, TIMIT ve NTIMIT veritabanları için karşılaştırılmıştır. BM ve K-ortalama algoritmaları için en iyi başlangıç şartları bulunmuş ve bu algoritmaların başarımları karşılaştırılmıştır. Ayrıca GKM ile konuşmacıların eğitiminde, BM algoritmasının yakınsama miktarı ve eğitim aşamasında model tekliklerini önlemek için değişinti sınırlandırmasının ne olması gerektiği belirlenmiştir.

TIMIT veritabanı için en yüksek başarımlar k-ortalama algoritması ile elde edilmektedir. Bu veritabanı için BM algoritması ile k-ortalama algoritması arasında tanıma başarımında önemli bir fark bulunmamaktadır. Bununla birlikte k-ortalama algoritmasının hesaplama işlem yükü BM algoritmasına göre önemli oranda düşüktür [16]. GKM’de TIMIT gibi gürültü içermeyen temiz veritabanları için k-ortalama algoritması, hem yapılan işlem azlığı hem de tanıma başarımı yönünden tercih edilebilir.

NTIMIT veritabanında en yüksek başarımlar tablo 3’den görüleceği üzere BM algoritması ile elde edilmektedir. BM algoritması ile k-ortalama algoritması arasında tanıma başarımında 3 puanlık fark bulunmaktadır. NTIMIT gibi telefon iletim hatlarının etkilerini içeren veritabanları için BM algoritması tanıma başarımı yönünden uygun yöntemdir.

LBG algoritması tek başına konuşmacıların eğitiminde düşük sonuçlar vermektedir. Ancak BM eğitim algoritması ile konuşmacılar eğitilirken, başlangıç değerleri belirlenmesinde k-ortalama algoritması ile birlikte LBG algoritması kullanıldığında, tablo 2'den

görüreceği üzere en iyi başarımlar elde edilmektedir. LBG algoritması model başlangıç değeri belirlenmesinde tercih edilebilir.



Şekil 6. Küme sayısı 32 için k-ortalama algoritması ile (a) rastgele değişen (b) Ortalamaya en yakın (c) ortalamaya en uzak (d) LBG algoritmasına göre bulunan başlangıç değerleri

5. KAYNAKLAR

1. Reynolds, D.A., An Overview of Automatic Speaker Recognition Technology ICASSP, 4072-4076, 2002.
2. Ganchev, T., Speaker Recognition, PhD thesis, Dept. of Electrical and Computer Engineering, University of Patras, Greece, 61-82, 2005.
3. Kinnunen, T., Spectral Features for Automatic Text-independent Speaker Recognition, Ph.D. thesis, University of Joensuu, Department of Computer Science 49-115, 2003.
4. Rabiner, L. R. and B. H. Juang., Fundamentals of Speech Recognition. Prentice Hall, Englewood Cliffs, 1993.
5. Furui, S., Digital Speech Processing, Synthesis, and Recognition. M. Dekker Inc., 1989.

Erciyes Üniversitesi Fen Bilimleri Enstitüsü Dergisi 26(2): 143-153 (2010)

6. Reynolds, D.A., A Gaussian Mixture Modeling Approach to Text Independent Speaker Identification. Ph.D., thesis, Georgia Inst. of Technology, 1992.
7. Ertaş, F., Fundamentals of Speaker Recognition. Journal of Engineering Sciences, No. 2-3, Pamukkale, 85-193, 2000.
8. Bhattacharyya, S. T. Srikanthan, P. Krishnamurthy, Ideal GMM Parameters & Posterior Log Likelihood for Speaker Verification, Proceedings of the IEEE Signal Processing Society Workshop, USA. ISBN: 0-7803-7196-8, 471-480, 2001.
9. Reynolds, D. A., R. C. Rose., Robust Text-Independent Speaker Identification Using Gaussian Mixture Speaker Models. IEEE Trans. Speech Audio Processing, 3. 72-83, 1995.
10. McLachlan, G., Mixture Models. New York, NY: Marcel Dekker, 1 st. ed., 1988.
11. Rabiner, L. R., A tutorial on Hidden Markov Models and Selected Applications in Speech Recognition. Proceedings of IEEE, vol. 77, no. 2, 257-286, 1989.
12. Feder, M. Weinstein and A. Oppenheim, A New Class Of Sequential and Adaptive Algorithms with Application to Noise Cancellation. in Proceedings of the International Conference on Acoustics, Speech and Signal Processing, 557-560, 1988.
13. Eskidere Ö., İstatiksel Modelleme ile Konuşmacı Tanıma. Doktora tezi, Uludağ Üniversitesi, 2007.
14. J.B. McQueen, Some Methods of Classification and Analysis of Multivariate Observations, Proc. 5th Berkeley Symp. Mathematic Statist. Probability, 281-296. Univ. of California, Berkeley, USA, 1967.
15. Duda R. O., Hart P. E. and Stork D.G., Pattern Classification 2nd Edition, Wiley-Interscience Publication, USA. 2001.
16. G. Singth, A.Panda, S. Bhattacharvya, T. Srikanthan, Vector Quantization Techniques for GMM Based Speaker verification, ICASSP, 65-69, 2003.
17. Linde, Y., A. Buzo., R. M. Gray, An Algorithm for Vector Quantization, IEEE Trans. Communications, Vol. 28, No. 1, 84-95, 1980.
18. Shannon, B., J. Kuldip. and K. K. Paliwal., A Comparative Study of Filter Bank Spacing for Speech Recognition. Microelectronic Engineering Research Conference, 2003.
19. Campbell, J. P. and A. D. Reynolds., Corpora for the Evaluation of Speaker Recognition Systems. IEEE Trans. Speech Audio Processing, 829-832, 1999.
20. Slaney, M., Auditory Toolbox: A MATLAB Toolbox for Auditory Modeling Work Technical Report, Interval Research Corporation, 29-32, 1998.
21. Aliaa, A. Y., A. S. Ebada, W. H. El Behaidy, Development of Automatic Speaker Identification System. 21st National Radio Science Conference, 2004.
22. Eskidere, Ö. Ertaş, F., Gauss karışım modeli ile konuşmacı tanımda Parametre değerlerinin seçimi, SİU, Eskişehir, 2007.