

<b>ÖZET.....</b>	<b>IV</b>
<b>SUMMARY.....</b>	<b>V</b>
<b>TEŞEKKÜR.....</b>	<b>VI</b>
<b>İÇİNDEKİLER.....</b>	<b>VII</b>
<b>SİMGELER VE KISALTMALAR.....</b>	<b>IX</b>
<b>ŞEKİLLER DİZİNİ.....</b>	<b>X</b>
<b>1. GİRİŞ.....</b>	<b>IX</b>
1.1. Biyometrik.....	IX
1.2. Konuşmacı Tanıma Temelleri.....	X
1.3. Konu İle İlgili Yayınlar Ve Çalışmalar.....	XI
<b>2. TEMEL AKUSTİK VE SİNYAL İŞLEME.....</b>	<b>XII</b>
2.1. Ses.....	XII
2.2. Akustikte Uzay ve Zaman.....	XIII
2.3. Ses Yakalama ve Saklama.....	XIV
2.4. İşitme.....	XV
<b>3. KONUŞMACI TANIMA.....</b>	<b>XVII</b>
3.1. Konuşmacı Saptama.....	XVIII
3.1.1. Kaydolma Aşaması (Enrollment).....	XVIII
3.1.2. Saptama Aşaması (Identification).....	XIX
3.1.3. Karar Verme.....	XIX
3.2. Öznitelik Çıkarma (Feature Extraction).....	XX
3.2.1. Çerçeveleme (Frame Blocking).....	XXI
3.2.2. Pencereleme (Windowing).....	XXII
3.2.3. Fourier Dönüşümü (FFT).....	XXII
3.2.4. Mel Frekans Saptırması (Mel-Frequency Warping).....	XXII
3.2.5. Kepstrum .....	XXIII
3.3. Vektör Kuantumlama.....	XXIV
3.4. Öznitelik Karşılaştırma.....	XXIV
3.5. Öznitelik Karşılaştırma (Önerilen Yöntem).....	XXVI
3.5.1. Yeni Yöntemin Üstünlüğü.....	XXVII
<b>4. SİSTEMİN GERÇEKLEŞTİRİMİ.....</b>	<b>XXIX</b>

4.1. Geliştirilen Matlab Fonksiyonları.....	XXIX
.....	XXXII
4.2. Kullanılan Kod Kütüphaneleri.....	XXXIII
<b>5. TEST SONUÇLARI.....</b>	<b>XXXIV</b>
<b>6. SONUÇ .....</b>	<b>XXXVI</b>
<b>KAYNAKLAR.....</b>	<b>XXXVII</b>
<b>ÖZGEÇMİŞ.....</b>	<b>XLI</b>

# 1. GİRİŞ

En büyük güvenlik tehlikelerinden birisi, yetkileri olan birisini taklit etmek, o kişiymiş gibi davranmaktır. Bu sorunun üstesinden gelmeye çalışan güvenlik hizmetleri saptama (identification) ve doğrulamadır (verification / authentication). Saptama ve doğrulama, kişinin bildiği bir şeyle (şifre, anne kızlık soyadı gibi), sahip olduğu bir şeyle (akıllı kart gibi) ya da kişisel karakteristikleri ile (ses, parmak izi, retina gibi) yapılabilir. Kimlikler çalınabilir, şifreler unutulabilir ya da kırılabilir. İngiltere’de yapılan bir araştırmaya göre [1], her 3 kişiden birisi PIN numaralarını yazmaktadır. Şirketlere, bilgi sistemlerine izinsiz erişimler ve bilgilerin değiştirilmesi her gün gerçekleşmekte ve yıkıcı sonuçlar olabilmektedir. Kredi kartı sahtekarlığı giderek artmakta ve iflaslara yol açmaktadır. Toplumların, ilerleyen teknoloji ile birlikte bilgisayar bağımlılığının arttığı günümüzde, ek güvenlik yöntemlerine ihtiyaç duyulmaktadır.

Konuşan kişiyi tanıma, kişiyi, kendisine özgü olan ses özelliklerinden tanımayı hedeflemektedir. Bankacılık gibi güvenliğin gerekli olduğu alanlarda, sesle aramada, veritabanı erişimi hizmetlerinde, sesli posta iletisi gönderiminde, bilgisayarlara uzaktan erişimde, kriminal laboratuvarlarında, kişileri ses örneklerinden teşhis etmek için kullanılabilir [2].

Bu tezde bilinen bir konuşmacı tanıma algoritması irdelenmiş, yeni bir karşılaştırma algoritması önerilmiştir.

## 1.1. Biyometrik

Günümüz teknolojisinde ek güvenlik ihtiyaçlarına cevap verebilen biyometrik, kişilerin sahip olduğu, özgün fiziksel ve davranışsal özellikleri kullanan bir örüntü eşleme sistemidir [3]. Biyometrik, şifre hatırlama ya da kimlik taşıma gibi ihtiyaçlara gerek duymadan, sadece kişiye ait olan özellikleri kullanır. Bu sistemler kişiye ait bilgileri kaydedip bir veritabanı oluştururlar ve daha sonra her güvenlik

sorgulamasında kullanırlar. En çok bilinen biyometrik teknikleri aşağıdaki gibidir [4]:

DNA Yapısı Analizi

Kulak İzi Tanıma

Yüz Tanıma

İmza Doğrulama

Yürüme Şeklinden Tanıma

El ve Parmak Geometrisi Tanıma

İris Tanıma

Retina Taraması

Koku Tanıma

Parmak İzi Tanıma

Ses Tanıma (Konuşmacı Tanıma)'dır.

## **1.2. Konuşmacı Tanıma Temelleri**

Konuşmacı Tanıma teknolojisinin temelleri 1960'lı yıllarda atılmıştır. O zamandan bugüne, değişik teknik gruplar, bu teknolojinin bugün bir çok uygulamada verimli olarak kullanılabilmesi için çalışma yapmaktadırlar. Konuşma, insanlar arasındaki en doğal ve en temel iletişim aracıdır. Konuşmanın kişi tanımada kullanılmasının avantajları; kullanımı basit olması, retina taraması ya da el taraması

operasyonlarının aksine hareket kısıtlayıcılığının olmaması, uzak mesafelerde de tanıma yapılabilmesi ve de diğer yöntemlere göre daha az maliyetli olmasıdır. Karşılaşılan tipik problemler ise; kanal uyumsuzluğu (kaydolma sırasında kullanılan mikrofon ile tanıma sırasında kullanılanların farklı olması), arka planda mevcut gürültüler, akustik çevre uyumsuzluğudur (kaydın laboratuarda yapılması, tanıma işleminin de ofiste yapılması gibi) [5].

### 1.3. Konu İle İlgili Yayınlar Ve Çalışmalar

Çizelge 1, metin bağımsız konuşmacı tanıma alanındaki çalışmalarını kronolojik sırada göstermektedir. Tablodaki sütunlar sırasıyla, çalışmayı gerçekleştiren kişileri (kaynak), çalışmada kullanılan yöntemi, konuşmacı sayısını (nüfus), başarı oranını, kaydolma aşamasında kullanılan ses örneklerinin uzunluğunu ve saptama aşamasında kullanılan ses örneklerinin uzunluğunu göstermektedir.

Kaynak	Yöntem	Nüfus	Başarı Oranı	Ses Uzunluğu (Kaydolma Aşaması)	Ses Uzunluğu (Saptama Aşaması)
Marvin R. Sambur 1976 [6]	OLP	21	%94	-	-
Markel and Davis 1979 [7]	Uzun Süreli İstatistik	17	%98,05	15dk	39s
R. Schwartz, S. Roucos, and M. Berouti 1982 [8]	Nonparametric pdf	21	%97	10s	2s
Li and Wrench 1983 [9]	Örüntü Eşleme	11	%96	100s	10s
F. Soong, A. Rosenberg, L. Rabiner, B. Juang 1985 [10]	VQ	100	%98	-	-
T. Matsui and S. Furui 1991 [11]	VQ	9	%99	17s	30s
Douglas A.Reynolds and Richard C. Rose 1995 [12]	GMM	49	%80,8	80-100s	15s
Zhong-Xuan Yuan, Bo-Ling Xu, and Chong-Zhi Yu 1999 [13]	MFCC	42	%99,1	-	-
Jin, Q. and Waibel, A. 2000 [14]	LDA	230	%97	30s	15s
M. Selvi,U. Yapanel, F. Gürgen 2002 [15]	WMFCC	93	%98,9	30-45s	2s
Jiuqing Deng, Qixiu Hu 2003 [16]	Örüntü Eşleme	200	%98,5	60s	12s
L. Zhang, B. Zheng and Z.	GA	20	%93,85	90s	3s

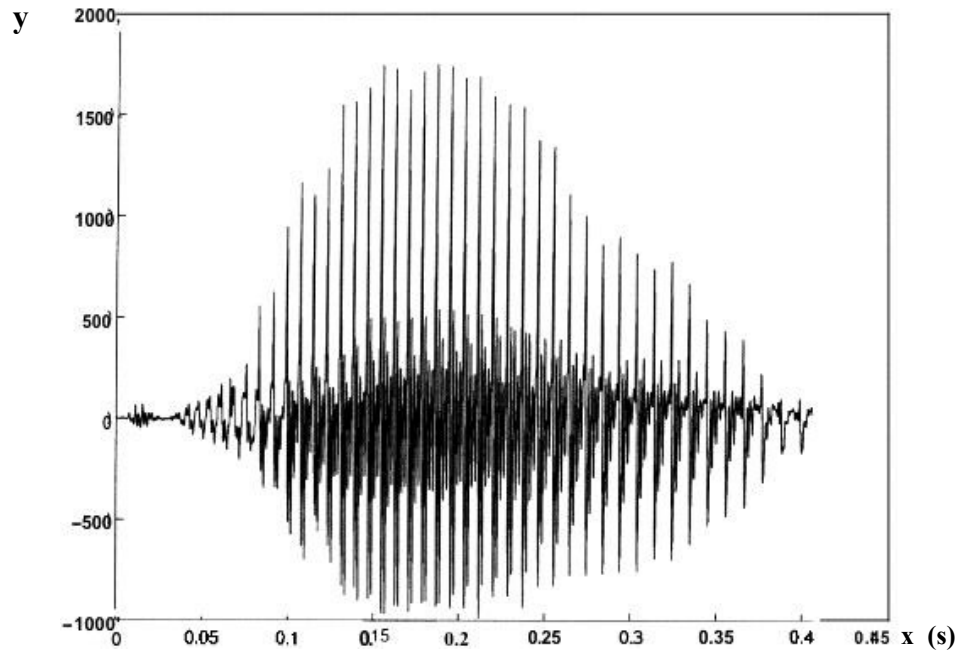
Çizelge 1.1. Konu ile ilgili yapılan çalışmalar

## 2. TEMEL AKUSTİK VE SİNYAL İŞLEME

Ses işleme mekanizmalarını anlamak için öncelikle temel ses ve sinyal işleme bilgisi vermek yerinde olacaktır.

### 2.1. Ses

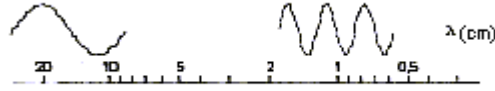
Ses fiziksel olarak dinleyici ile ses kaynağı arasında bulunan sıvı, gaz ya da katı ortamda oluşan bir dizi titreşimlerdir [18]. Ortamın elastiki yapısı sesin kaynaktan dalgalar şeklinde yayılmasına olanak verir. Bir obje titreşime uğradığı zaman enerjisinin bir kısmını çevresine ses olarak verir. Ses vakumlu ortamlarda ilerlemez. Bir ses kaynağının ilerleyişi Şekil 2.1’de görülmektedir. Ses havada basınç farklılıkları yaratarak ilerler. Burada dikkat edilmesi gereken sadece basınç farklılıklarının ilerlediği havanın ilerlemediğidir. Bunu, su yüzeyine atılan bir taşın oluşturduğu dalgalara denk gelen bir cismin yukarı aşağı hareket etmesine benzetebiliriz. Bu nedenledir ki güçlü hoparlörlerin önünde hava akımı oluşmaz. Ses dalgaları 20 derece hava sıcaklığında yaklaşık olarak saniyede 344 metre ilerleyebilir.



Şekil 2.1. Ses Dalgası. (x zamanı, y ise genliği göstermektedir)

### ***Frekans***

İki tepe arasındaki uzaklık dalga boyu, bir saniyede aynı noktadan geçen tepe sayısı da frekanstır. Şekil 2.2’de düşük ve yüksek frekanslı iki ses dalgası görülmektedir. Frekans birimi Hertz’dir.



**Şekil 2.2.** Yüksek ve düşük dalga boyları

### ***Genlik***

Sesin diğer bir özelliği genliğidir. Ses yumuşak ya da gürültülü olabilir. Hava ortamında bu, havanın sıkıştırılma gücüne göre oluşan küçük ya da büyük basınç değişimleridir. Sesin gücünü ölçmek için desibel birimi kullanılır. Desibel birimi fiziksel bir çıkarımdan daha çok insan kulağının duyma şekline göre hesaplanmıştır. İnsan kulağı tüm ses seviyelerini aynı şiddette hissetmez, belli bir seviyeden sonra ses şiddeti logaritmik olarak hissedilir [34]. Bu yüzden desibel birimi logaritmik bir formülle hesaplanmıştır.

## **2.2. Akustikte Uzay ve Zaman**

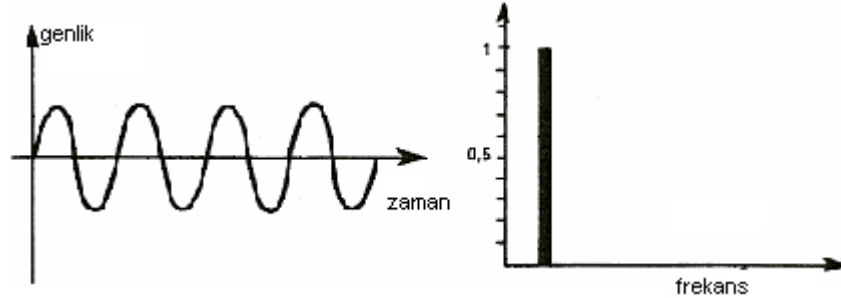
Zaman, akustik biliminde önemli bir rol oynar. Zaman ve uzay arasında sıkı bir ilişki vardır, çünkü ses uzay içerisinde zamanla yol alır. Bu bilgiler göz önüne alınarak akustik sinyaller 3 sınıfa ayrılabilir:

**Periyodik Sinyaller:** Zamanla tekrar eden sinyallerdir.

**Gelişigüzel Sinyaller:** Periyodik olmayan sinyallerdir.

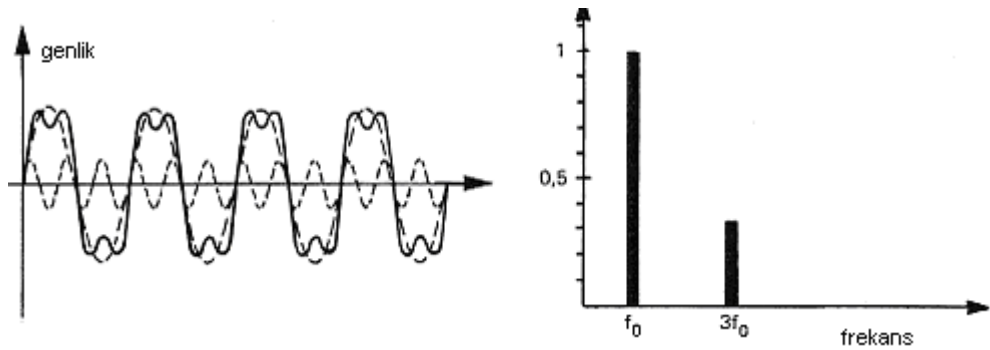
**Darbe (Pulse):** Zamanla tekrar etmeyen ancak belli bir şekli olan sinyallerdir.

Bir sinyalin spektrumu o sinyali oluşturan temel sinyalleri gösterir. Siren ya da ışık gibi kararlı bir sinyali ele alırsak, spektrum zamanla değişmez yani kararlı durumdadır. Şekil 2.3'te bu tür bir sinyale örnek olarak bir sinüs dalgasının zaman-genlik ve frekans-genlik (spektrum) grafikleri gösterilmiştir.



**Şekil 2.3.** Sinüs sinyali ve spektrumu

Her ses kendisini oluşturan basit sinüs sinyalleri ile ifade edilebilir. Seslerin basit sinüs sinyallerine ayrıştırılma işlemi, ilk kez 19. yüzyılda Fransız matematikçi Fourier tarafından gösterilmiştir. Şekil 2.4'de iki sinüs sinyalinin birleşmiş hali görülmektedir. Zaman-genlik gösteriminde belirgin olmamakla birlikte spektrum gösteriminden, bu karmaşık sinyali oluşturan iki sinyal açıkça görülmektedir.



**Şekil 2.4.** Birleşmiş iki sinüs sinyali

### 2.3. Ses Yakalama ve Saklama

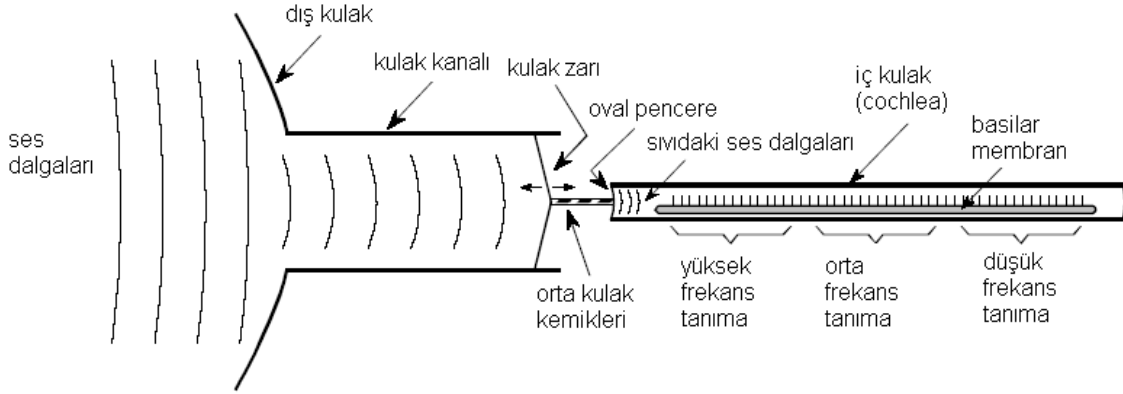
Sesin bilgisayar tarafından işlenebilmesi için, ses sinyallerinin yakalanıp saklanması gerekmektedir. Bu operasyon bize, havadaki basınç değişimlerini (ses dalgaları) bilgisayarın anlayabileceği bir dizi sayıya dönüştürmeyi sağlayacaktır. Bunu yapabilmek için, basınç değişimlerini elektrik akımına çeviren bir mikrofon ve elektrik sinyallerini sayılara dönüştüren bir örnekleme cihazı gerekmektedir. Örnekleme cihazları ADC (Analog To Digital Converter) olarak da bilinirler. Kişisel

bilgisayarlarda, ses kartı bu işlevi yerine getirir. Ses kartının, sayısal noktaları kaydetme hızları örnekleme frekansı olarak bilinir.

Bilgisayar sesin sadece belli anlardaki değerlerini yakalar. Sinyali, belirsizliğe yer vermeyecek şekilde tanımlayabilmek için sinyaldeki en büyük frekans, örnekleme frekansının yarısından büyük olmamalıdır [19], [20]. Aksi takdirde bir sinyal olduğundan farklı görünebilir, buna aliasing denir. Örneklenen sinyalin bu davranışı Nyquist Teoremi ve Shannon Teoremi ile açıklanır [21]. Nyquist frekansı bir sinyaldeki en yüksek frekans değerinin yarısıdır.

## 2.4. İşitme

İnsan kulağı oldukça karmaşık bir yapıya sahiptir. Duyma sistemini daha da karmaşıklaştıran ise iki kulaktan gelen verilerin dünyanın en karmaşık sinir ağı olan beyinde birleştirilmesidir. Bu kısımda işitme hakkında bilinen gerçeklerden bahsedilecektir. İşitme ile ilgili henüz anlayamamış bir çok etki ve olaylar mevcuttur. Şekil 2.5 insan kulağının ana yapısını göstermektedir. Dış kulak iki kısımdan oluşur; kulak çevresi ile birlikte baş kısmına bağlı bulunan kıkırdak yapı ve 0.5 cm yarıçapında başın iç kısmına doğru 3 cm uzayan kulak kanalı. Bu yapılar çevresel sesleri, baş iskeletinin güvenli iç kısımlarında bulunan orta ve iç kulak kısımlarına yönlendirirler. Kulak kanalının sonunda kulak zarı adı verilen ince bir deri tabakası bulunur. Kulak zarına çarpan ses dalgaları zarın titreşmesine neden olur. Orta kulak, bu titreşimleri iç kulağa (cochlea) ileten bir dizi kemik yapısından oluşur. İç kulağa ulaşan ses titreşimleri sinirsel uyarılara dönüşürler. İç kulak içi sıvı dolu olup kabaca 2 mm çapında ve 3 cm uzunluğundadır. Şekilde düz görünmesine rağmen gerçekte bir deniz kabuğunu andıracak şekilde kıvrılmış bir yapıya sahiptir.



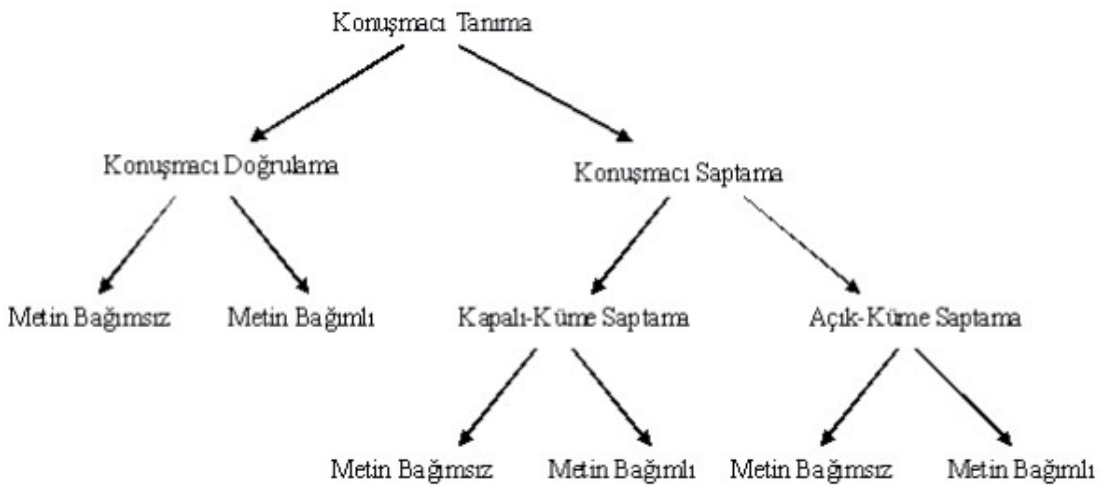
**Şekil 2.5.** İnsan kulağının fonksiyonel şeması [22]

İç kulak basilar membran ismi verilen bir yapıyı barındırır. Bu yapı yaklaşık 12000 sensör hücresinden oluşan iç kulak sinirleridir. Basilar membrandaki her sinir fiberi belirli frekansa karşı hassastır. Kulakta meydana gelen bir diğer bilgi kodlaması da ses verisinin sinirlerden iletimi sırasında oluşur. Sinir hücreleri bilgiyi, etki potansiyeli denilen kısa aralıklı elektrik sinyalleri ile gönderirler. Basilar membrandaki bir sinir hücresi, titreşimin her devri için bir etki potansiyeli oluşturarak ses sinyalini kodlayabilir. Örneğin 200 hertzlik bir ses dalgası bir sinir hücresi tarafından saniyede 200 etki potansiyeli oluşturularak ifade edilebilir [22].

İnsan kulağı duyma aralığı, en düşük duyma seviyesi olan 0 dB ile, ağrı hissedilmeye başlanan 140 dB arasındadır. Normal konuşma yaklaşık olarak 60 dB dir. Kulağın hassasiyeti şaşırtılacak derecededir. Şöyle ki çok zayıf sesler işitilirken, kulak zarı bir molekül çapından daha az bir mesafe kadar titreşim gösterir. Normal ses duyma aralığı 20 hertz ile 20 kHz frekansları arasında kabul edilir. Ancak 1 kHz ile 4kHz arasındaki frekanslara daha duyarlıdır. Örneğin bir dinleyici, 3 kHz frekansındaki bir sesi 0 dB de duyabilirken, 100 hertz frekansındaki bir sesi 40 dB de duyabilmektedir. Yine aynı şekilde bir dinleyici 3 kHz frekansında, %0,3'lük frekans farklılığından oluşan tonları ayırt edebilmektedir. Bu oran 100 hertz de %3 e çıkmaktadır. Bir karşılaştırma olarak bir piyano üzerindeki ardıl tuşlar arasında %6'lık bir frekans farkı vardır [22].

### 3. KONUŞMACI TANIMA

Konuşmacı tanıma, ses dalgalarında bulunan konuşmacıya özgü bilgileri kullanarak, konuşmacılar arasından kimin konuştuğunun saptanmasıdır [2]. Konuşmacı tanıma iki ana bölüme ayrılabilir; konuşmacı doğrulama (speaker verification) ve konuşmacı saptama (speaker identification). Konuşmacı doğrulama, bilinmeyen bir ses örneğinin, iddia edilen kişiye ait olup olmadığının belirlenmesidir [23]. Konuşmacı doğrulama, iletişim hizmetlerinde, bankacılık hizmetlerinde, özel kayıtlara vb. güvenli erişimler sağlamak amacıyla bir çok uygulamada kullanılabilir. Konuşmacı saptama ise bilinmeyen bir ses örneğinin, belli konuşmacıların ses kayıtlarından oluşan bir veritabanı içerisinde hangi kişiye ait olduğunun bulunmasıdır. Konuşmacı tanıma, metin bağımlı ya da metin bağımsız olabilir. Metin bağımlı sistemlerde konuşulan metin sistem tarafından önceden bilinmektedir. Metin bağımsız sistemlerde ise, metin, herhangi bir sözdizimi olabilir [24]. Diğer taraftan, konuşmacı tanıma, açık küme ya da kapalı küme olabilir. Kapalı kümede bilinmeyen ses örneği, veritabanındaki konuşmacılardan birisine aittir. Açık kümede ise ses örneği veritabanındaki konuşmacılardan hiç birisine ait olmayabilir [25]. Dolayısı ile açık küme konuşmacı tanıma sistemlerinde, ret sonucunu da içeren fazladan bir olasılık daha vardır. Şekil 3.1’de konuşmacı tanıma genel sınıflandırma şeması verilmiştir.



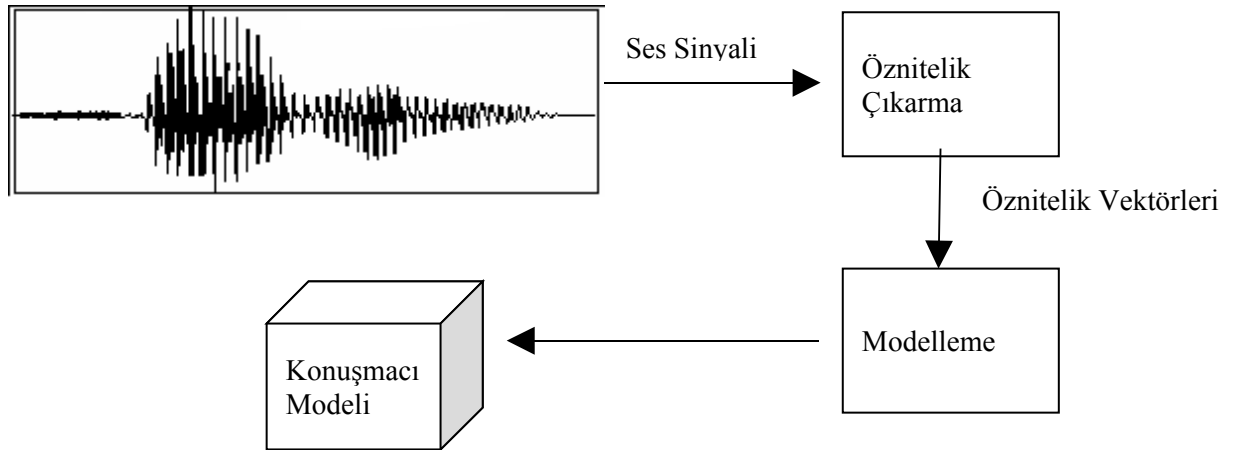
Şekil 3.1. Konuşmacı tanıma sınıflandırma

### 3.1. Konuşmacı Saptama

Konuşmacı saptama iki ana aşamadan oluşur. Kaydolma aşaması (enrollment) ve saptama (identification) aşaması. Kaydolma aşamasında konuşmacıların tüm ses örnekleri alınarak işlenir ve bir veritabanında saklanır. Kaydolma aşamasının amacı her konuşmacı için bir referans modeli oluşturmaktır. Daha sonra bu modeller saptama aşamasında kullanılırlar.

#### 3.1.1. Kaydolma Aşaması (Enrollment)

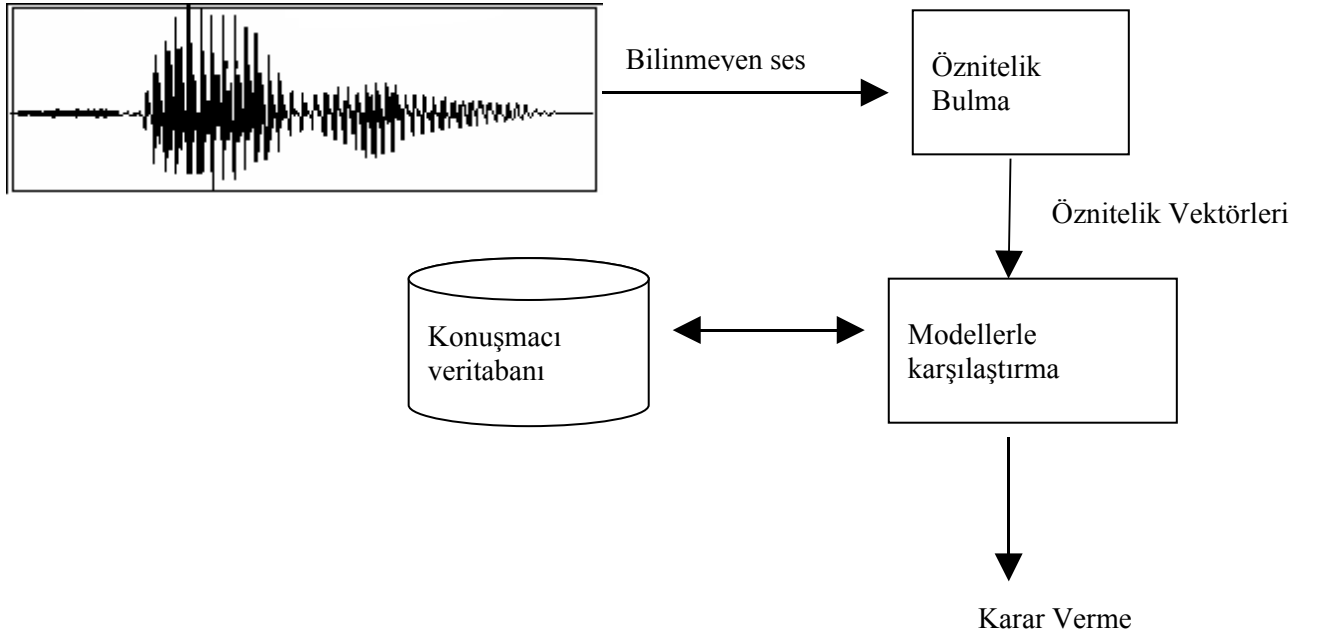
Konuşmacı saptama sisteminin ilk aşaması olan kaydolma aşamasında, konuşmacılardan alınan ses örnekleri çeşitli sinyal işleme teknikleriyle, konuşmacıya özgü olan öznitelik vektörlerini çıkarmak amacıyla işleme tabi tutulurlar. İşlem sonunda her konuşmacı için bir model oluşturulur ve konuşmacı veritabanına kaydedilir. Bu veritabanı daha sonra saptama aşamasında kullanılmak üzere saklanır. Gerçek zamanlı konuşma tanıma sistemlerinde zaman önemli bir unsur olduğundan, tüm sistemi yavaşlatan hesaplamaların mümkünse bu aşamada yapılması tanıma zamanını hızlandıracaktır. Şekil 3.2’de kaydolma aşamasının genel diyagramı verilmiştir.



Şekil 3.2. Kaydolma

### 3.1.2. Saptama Aşaması (Identification)

Saptama aşamasında, girdi olarak verilen bilinmeyen bir ses örneği işlenip öznitelik vektörleri çıkarılır. Bu aşama kaydolma aşamasındaki ilk işlem ile aynıdır. Daha sonra bu vektörler bir karşılaştırma algoritması yardımıyla (matching algorithm) konuşmacı veritabanında kayıtlı bulunan diğer modellerle karşılaştırılırlar. Bu işlem sonunda her model için bir benzerlik değeri bulunur. Bulunan en benzer model, ses örneğinin sahibi olarak nitelendirilir. Saptama aşaması şekil 3.3'te gösterilmektedir.

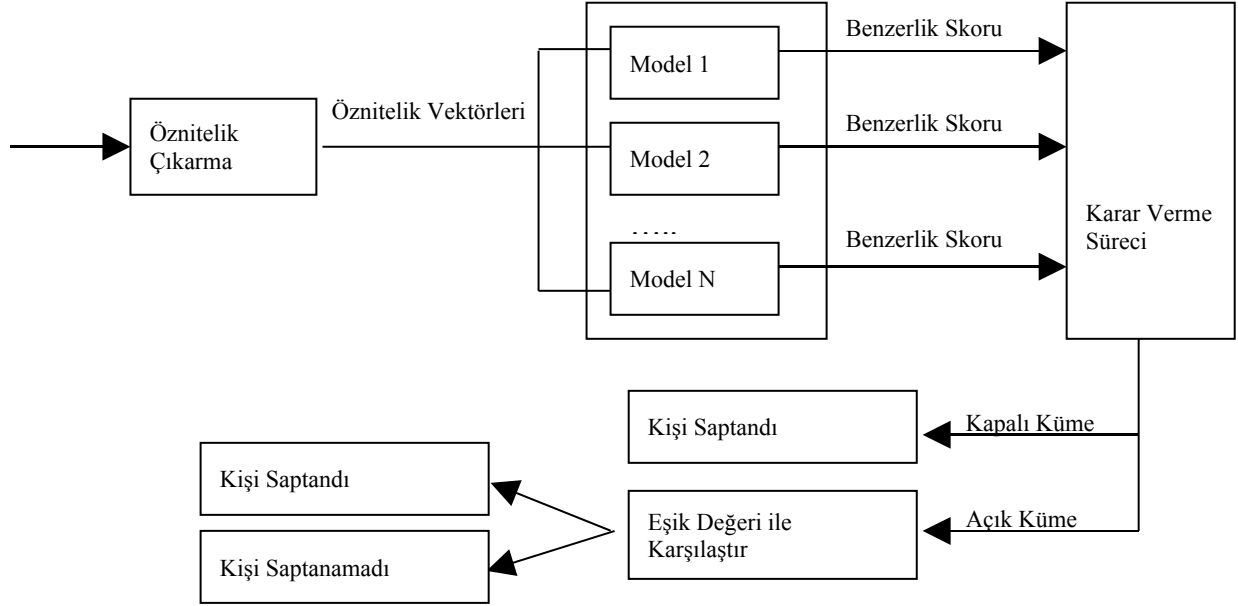


Şekil 3.3. Saptama

### 3.1.3. Karar Verme

Karar verme süreci, tanıma metodunun kapalı küme yada açık küme olmasına göre değişir. Kapalı küme tanıma sisteminde, veritabanındaki bir konuşmacı bilinmeyen ses örneğinin sahibi olarak seçilmek zorundadır. Dolayısı ile en yüksek benzerlik skoru alan konuşmacı modeli seçilmiş olacaktır. Diğer taraftan açık küme konuşmacı saptamada, en yüksek skoru alan model hemen seçilmez, ek bir kontrole daha tabi tutulur. Bu değer daha önce belirlenmiş bir eşik değeri ile karşılaştırılır.

Değer, eşik değerinden küçük ise, ses örneği veritabanındaki hiç bir konuşmacıya ait değildir. Konuşmacı tanıma sisteminin karar verme aşaması ile birlikte genel şeması şekil 3.4'te gösterilmiştir.



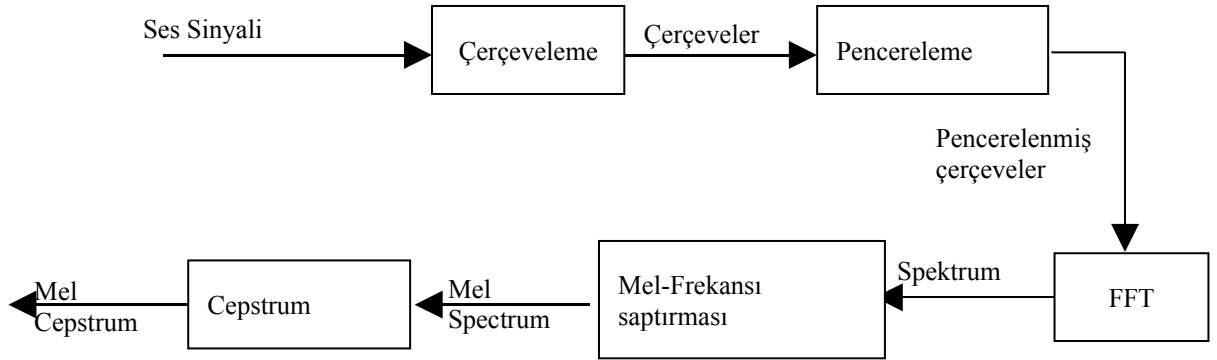
Şekil 3.4. Konuşmacı tanıma karar verme süreci

### 3.2. Öznitelik Çıkarma (Feature Extraction)

Ses sinyalindeki akustik özelliklerin çıkarılmasına öznitelik çıkarma denilir. MFCC, konuşma tanıma alanında en çok kullanılan öznitelik çıkarma yöntemlerinden birisidir [26]-[32]. Öznitelik çıkarma hem kaydolma hem de saptama aşamasında kullanılır ve aşağıdaki aşamalardan oluşur [33], [38]:

1. Çerçeve Bloklama (Frame Blocking)
2. Pencereleme (Windowing)
3. Fourier Transformasyonu (FFT)
4. Mel-Frekansı saptırması (Mel-Frequency Warping)
5. Kepstrum (Cepstrum)

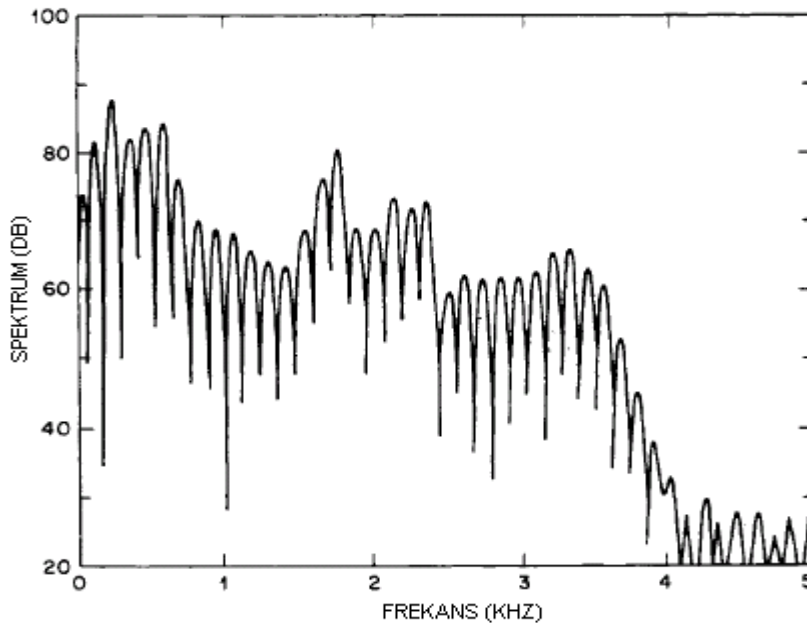
Öznitelik çıkarmanın ana hedefi, ses örneğindeki, konuşmacıyı tanımlayan akustik özelliklerden ödün vermeden yüklü miktardaki veriyi özetlemektir [34]. Bu sayede tanıma işlemi kolaylaşır ve daha verimli olur. Öznitelik çıkarma işlemleri şekil 3.5'te gösterilmiştir.



Şekil 3.5. Öznitelik çıkarma

### 3.2.1. Çerçeveleme (Frame Blocking)

Araştırmalar göstermiştir ki ses sinyali karakteristikleri yeteri kadar küçük bir zaman aralığında kararlı kalmaktadır. Bu nedenle ses sinyalleri kısa zaman aralıklarında işlenmektedir [35], [36]. Sinyaller, genellikle 20 ve 100 milisaniye arasında değişen uzunluklarda çerçevelere bölünürler. Çoğu durumda en etkili zaman aralığı 20-30 ms arasındadır [37]. Kısa zaman aralıklarına bölünen ses dalgasından elde edilen bir spektrum örneği şekil 3.6’da görülmektedir. Her çerçeve kendisinden bir önceki çerçevenin belli bir kısmını örter. Örtme yönteminin amacı bir çerçeveden diğerine geçişim yumuşak olmasını sağlamaktır.



Şekil 3.6. Bir ses dalgasına ait kısa süreli spektrum örneği [37]

### 3.2.2. Pencereleme (Windowing)

İkinci aşama bütün çerçeveleri pencerelemektir. Pencerelemenin amacı çerçeve başında ve sonundaki süreksizlikleri ortadan kaldırmaktır. Bu aşamada en çok kullanılan pencereleme fonksiyonlarından birisi Hamming fonksiyonudur [26], [32], [38]. Hamming fonksiyonu, aşağıdaki gibi tanımlanır [37]:

$$p(t) = \begin{cases} 0,54 + 0,46 \cos 2\pi t / N, & -N/2 \leq t \leq N/2 \\ 0, & |t| > N/2 \end{cases}$$

N, pencere süresi olup tipik olarak 25 ms olarak seçilir [37].

### 3.2.3. Fourier

#### Dönüşümü (FFT)

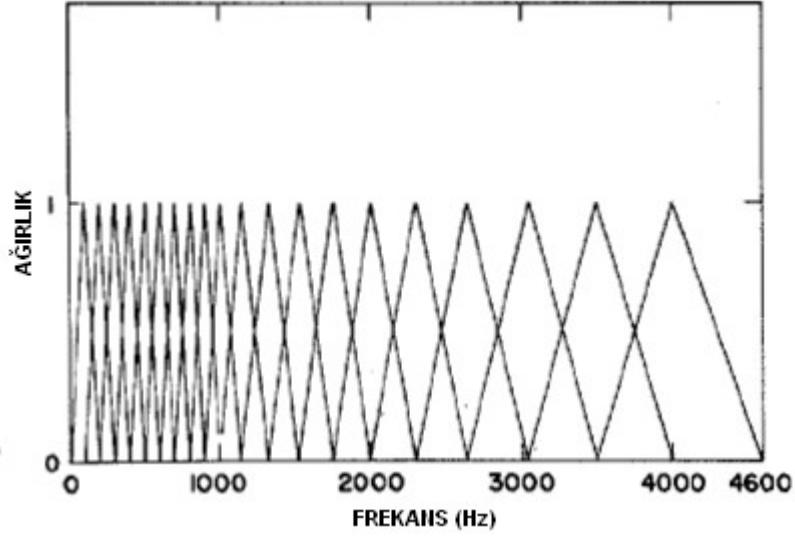
Bir sonraki aşama her çerçevede Hızlı Fourier Dönüşümü almaktır. Bu dönüşüm Ayırık Fourier Dönüşümünün hızlı bir şeklidir ve tanım kümesini zaman uzayından frekans uzayına geçirir.

### 3.2.4. Mel Frekansı Saptırması (Mel-Frequency Warping)

Normal bir insan kulağı frekansları doğrusal olmayan bir şekilde algılar. Araştırmalar göstermiştir ki, ölçüler, 1 kHz e kadar doğrusal, daha yüksek değerlerde ise logaritmiktir [23], [34]. İnsan kulağının frekans yanıtını karakterize eden bu ölçüye Mel-Ölçüsü (Melodi Ölçüsü) de denir. Konuşmacı tanımanın bu aşaması bant geçiren süzgeç (band pass filter) olarak kullanılır. Mel-Ölçüsü diyagramı şekil 3.6'da görüldüğü gibidir. Bir frekansı mel-frekansına çevirmek için aşağıdaki formül kullanılır [23]:

$$m = 2595 \log_{10}(1 + f / 700)$$

$f$ , Hertz biriminden frekansı,  $m$  ise mel frekansını göstermektedir.



Şekil 3.6. Mel-Ölçüsü diyagramı [26]

### 3.2.5. Kepstrum

Öznitelik çıkarmanın son aşaması olarak, her çerçeve ters fourier transformasyona uğratılır ve frekans uzayından tekrar zaman uzayına döndürülür. Bu işlemin sonucu olarak Mel-Frekansı Kepstral Katsayıları (MFCC: Mel-Frequency Cepstral Coefficients) elde edilir. MFCC aşağıdaki gibi hesaplanır [26]:

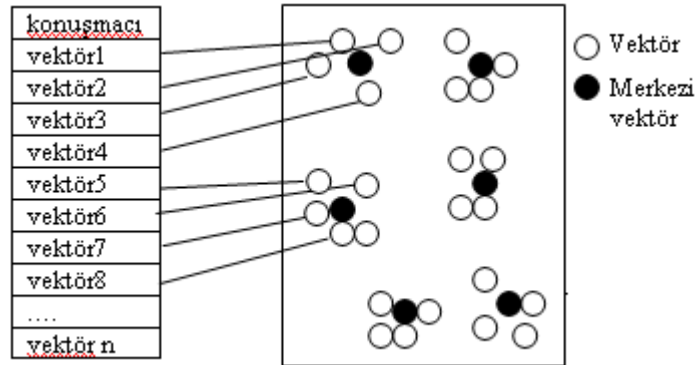
$$MFCC_i = \sum_{k=1}^{20} X_k \cos \left[ i \left( \frac{k-1}{2} \right) \frac{\pi}{20} \right], \quad i = 1, 2, \dots, M$$

$M$ , kepstrum katsayılarının sayısını,  $X_k$ ,  $k = 1, 2, \dots, 20$ , ise  $k$ . filtrenin log-enerji çıktısını göstermektedir. Bu projede MFCC katsayısı 12 olarak seçilmiştir. Her çerçeve için elde edilen 12 boyutlu vektörlere (feature vectors: öznitelik vektörleri) Kod Sözcüğü (Codeword) denilir. Bir konuşmacı için bulunan tüm kod sözcüklerinden oluşan kümeye de Kod Kitabı (Codebook) denilir. Kod kitabı o konuşmacıyı niteleyen bir model olarak ta düşünülebilir. Öznitelik çıkarma işleminin sonucunda, her konuşmacı için, daha sonra saptama aşamasında kullanılmak üzere kod kitapları çıkarılır.

### 3.3. Vektör Kuantumlama

Kepstrum aşamasından sonra elde edilen öznitelik vektörlerinin sayısı çok büyük boyutlardadır. Yüksek boyutlarda veriyi işlemek çok zaman alacağından bu vektörlerin bir yöntemle daha küçük boyutlara indirgenmesi sağlanmalıdır. Vektör Kuantumlama (VQ: Vector Quantization) algoritması bu ihtiyaca cevap verebilen hızlı ve etkili yöntemlerden birisidir [10].

Vektör Kuantumlama, büyük sayıdaki vektörleri belli bir sayıda kümelere indirir. Her kümede merkezi bir vektör bulunur ve bu vektör kümedeki tüm vektörleri temsil eder. Bu sayede n sayıdaki vektör sabit sayıda bir vektör dizisine indirgenmiş olur. Şekil 3.7’de VQ yaklaşımı görülmektedir. Bir konuşmacı, VQ uygulanmadan önce n adet öznitelik vektörü ile temsil edilmektedir. Kümeleme operasyonundan sonra ise, daha önceden belirlenen sayıda merkezi vektörler ile daha uygun bir şekilde temsil edilir.



Şekil 3.7. Bir konuşmacı için vektör kuantumlama

### 3.4. Öznitelik Karşılaştırma

Konuşma tanıma sisteminin son aşaması olan karar verme aşamasında bilinmeyen ses örneğinin öznitelik vektörleri, veritabanındaki her konuşmacı modeli için bulunan vektörlerle karşılaştırılır. Öznitelik karşılaştırma olarak bilinen bu işlem aslında bir örüntü eşleme (pattern matching) problemidir. En iyi bilinen karşılaştırma algoritmalarından birisi, Vektörler arasındaki öklidyen uzaklıklarının bulunmasıdır. Benzerlik ölçüsü iki vektör arasındaki uzaklık ile ters orantılıdır. Şöyle ki iki vektör birbirine ne kadar yakınsa, aralarındaki uzaklık değeri azalır ancak benzerlik değeri

de artar. Öklidyen uzaklığını, konuşma saptama sistemine uyarlırsak, ses örneği için bulunan kod kitabı ile modellere ait olan kod kitapları arasındaki benzerlik bulunurken kullanılacak formül aşağıdaki gibidir:

$$S(C_k, C_x) = \sum_{i=1}^T \frac{1}{d_{\min}(v_i, C_x)} \quad (1)$$

**S(C<sub>k</sub>, C<sub>x</sub>):** İki kod kitabı (C<sub>k</sub>, C<sub>x</sub>) arasındaki benzerlik değeri

**C<sub>x</sub>:** Bilinmeyen ses örneğine ait kod kitabı

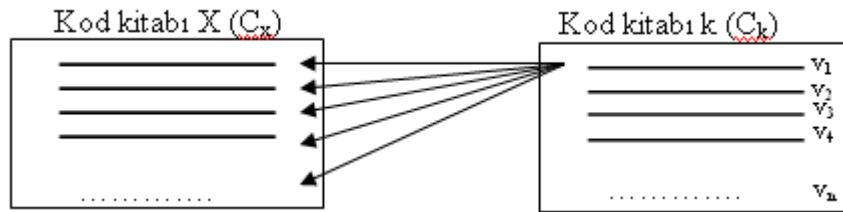
**C<sub>k</sub>:** Konuşmacı veritabanındaki k. kod kitabı

**v<sub>i</sub>:** C<sub>k</sub> kod kitabındaki i. kod vektörü

**d<sub>min</sub> (v<sub>i</sub>, C<sub>x</sub>):** v<sub>i</sub> kod vektörüne C<sub>x</sub> kod kitabındaki en yakın uzaklıkta bulunan kod vektörünün uzaklığı

**T:** Kod kitaplarındaki vektör sayısı (bütün kod kitaplarında aynıdır)

Şekil 3.8'de görüldüğü üzere bilinmeyen ses örneği için bulunan öznelik vektörlerinin oluşturduğu kod kitabına C<sub>x</sub> adı verilmiştir. Kayıtlı modeller arasında bulunan bir kod kitabı ile yapılan karşılaştırma şemasal olarak gösterilmiştir. Benzerlik ölçüsü bulunurken k. kod kitabındaki her vektör için C<sub>x</sub> kod kitabındaki en yakın uzaklıkta bulunan vektör arasındaki uzaklıklar hesaplanır. Bu değerlerin tersleri toplanarak iki kod kitabı arasındaki benzerlik ölçüsü bulunur.



**Şekil 3.8.** Öklidyen uzaklık yöntemi

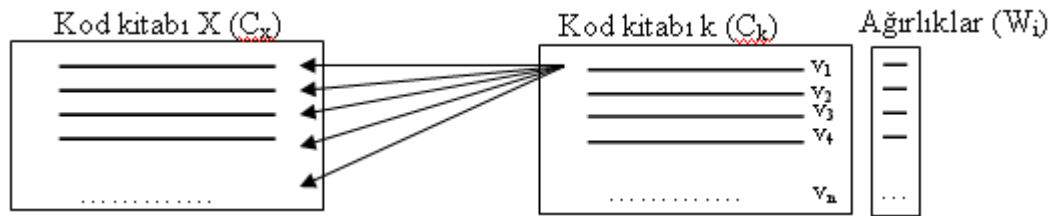
Bu yönteme yapılan bir iyileştirme, Kinnunen ve Franti tarafından önerilen veritabanındaki bilinen modeller arasında bulunan korelasyonların dikkate alındığı ağırlıklı karşılaştırma metodudur [39]. Bu yöntem, diğer kod kitaplarına benzemeyen kod vektörlerinin bir konuşmacıyı tanımlamada daha etkin olacağını vurgular.

Dolayısı ile konuşmacıyı, diğer kod vektörlerinden daha kolay ayırt edici özeliği olan vektörleri vurgulamak için onlara yüksek ağırlık skorları atanır. Ağırlık değerlerinin uygulanması için her kod kitabındaki vektörler için ağırlık değeri bulunur ve kaydolma aşamasında veritabanında saklanır. Bu ağırlık değerleri daha sonra saptama aşamasında, benzerlik skorları hesaplanırken kullanılır. (2) de görüldüğü gibi ağırlıklı karşılaştırma yönteminde, yerel benzerlik skorları, bu skorun hesaplanmasında kullanılmış olan en yakın kod vektörlerine karşılık gelen ağırlık değerleri ile çarpılırlar.

$$S(C_k, C_x) = \sum_{i=1}^T \frac{1}{d_{\min}(v_i, C_x)} * w_i \quad (2)$$

Bu formül (1) ile aynı olup tek fark olarak denkleme ağırlık skoru eklenmiştir.  $w_i$ ,  $v_i$  kod vektörüne  $C_x$  kod kitabındaki en yakın uzaklıkta bulunan kod vektörünün ağırlık değeridir.

Ağırlıklı karşılaştırma yöntemi şekil 3.9'da gösterilmiştir.



Şekil 3.9. Ağırlıklı öklidyen uzaklık yöntemi

### 3.5. Öznitelik Karşılaştırma (Önerilen Yöntem)

Ağırlıklı karşılaştırma yöntemine önerilen yeni yöntem birinci yöntemdekinin aksine, bilinmeyen kod kitabındaki vektörlerden model kod kitaplarına olan uzaklıkları kullanır. Benzerlik ölçüsü yeni yöntemde aşağıdaki gibi hesaplanır:

$$S(C_x, C_k) = \sum_{j=1}^T \frac{1}{d_{\min}(v_j, C_k)} * w \quad (3)$$

$S(C_k, C_x)$ : İki kod kitabı ( $C_k, C_x$ ) arasındaki benzerlik değeri

$C_x$ : Bilinmeyen ses örneğine ait kod kitabı

$C_k$ : Konuşmacı veritabanındaki k. kod kitabı

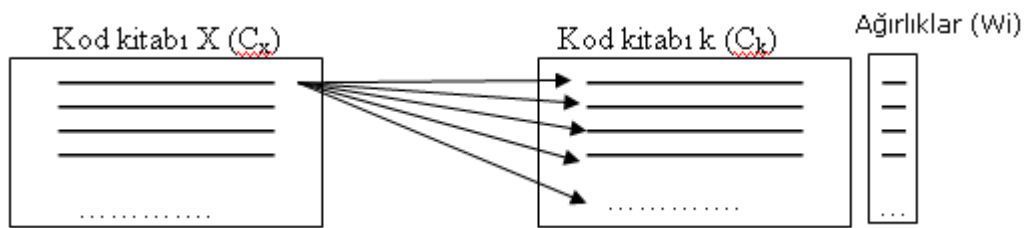
$v_j$ :  $C_x$  kod kitabındaki j. kod vektörü

$d_{\min}(v_j, C_k)$ :  $v_j$  kod vektörüne  $C_k$  kod kitabındaki en yakın uzaklıkta bulunan kod vektörünün uzaklığı

$T$ : Kod kitaplarındaki vektör sayısı

$w$ :  $C_k$  kod kitabındaki  $v_i$  kod vektörüne karşılık gelen ağırlık değeri

Önerilen yeni ağırlıklı karşılaştırma yöntemi şekil 3.10'da gösterilmiştir.

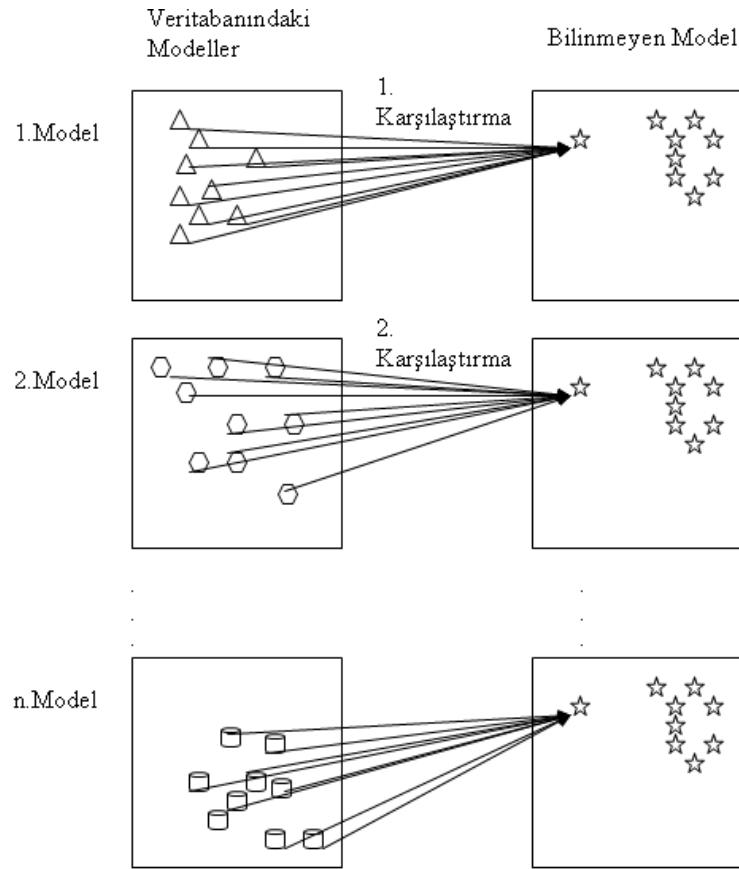


Şekil 3.10. Önerilen yöntem

### 3.5.1. Yeni Yöntemin Üstünlüğü

Kaydolma aşamasında, aynı kullanıcıya ait bir çok ses örneği harmanlanıp o konuşmacıya özgü nitelik vektörleri çıkarılır. Bir konuşmacıya ait ses verisi ne kadar fazla olursa çıkarılacak nitelik vektörlerinin kalitesi de artar. Çok kısa süreli ses örneklerinden, elde edilen öznelik vektörleri konuşmacıyı çok iyi tanımlayamaz. Konuşmacı tanıma ile ilgili bir çok çalışmada, değişik uzunluktaki ses örnekleri ile yapılan testler bu görüşü desteklemektedir [11], [14], [34], [39], [40], [41], [42], [43]. Dolayısı ile kaydolma aşamasında çıkarılan öznelik vektörleri, saptama aşamasında alınan ses örneğinden oluşturulan öznelik vektörlerinden daha tanımlayıcıdır. (2) de görüldüğü üzere, 1. yöntemde, bilinen bir modeldeki kod vektörlerinden bilinmeyen modeldekilere olan minimum uzaklıklar bulunmaktadır. Karşılaştırmada benzerlik skorunu etkileyen en önemli faktör minimum uzaklığı sağlayan vektörün konuşmacıyı yeteri kadar temsil edip edemediğidir. Minimum uzaklığı sağlayan vektör konuşmacıyı iyi temsil etmiyorsa Şekil 1'de örneklenen durum ortaya çıkar.

Bu durumda bulunan uzaklık değerleri gerçeği yansıtmaktan uzaktır, ve veritabanında bulunan her model için bilinmeyen modeldeki hatalı vektör(ler) devreye girecek ve tüm sonuçların hatalı çıkmasına neden olacaktır. Bilinmeyen modeldeki temsil yeteneği az olan bir (ya da bir kaç) vektör tüm sonuçların hatalı çıkmasına neden olur. Yukarıda açıklandığı üzere saptama aşamasındaki ses örneğinden oluşturulan vektörlerin temsil yeteneği kaydolma aşamasında oluşturulan vektörlere göre çok daha azdır. Bu nedenle karşılaştırma, bilinmeyen vektöre doğru değil de referans modeldeki vektörlere doğru yapılmalıdır. Bu yöntem iki açıdan daha avantajlıdır. 1. si, referans modellere doğru yapılan karşılaştırmalarda, minimum uzaklığın bulunacağı vektörler daha tanımlayıcı oldukları için bulunan uzaklık değerleri daha doğru sonuçlar verecektir. 2. si referans modelde minimum uzaklığı sağlayan vektör hatalı olsa bile bu sadece o karşılaştırmayı etkileyecek, sonraki karşılaştırmalar bu karşılaştırmadan bağımsız olarak devam edecektir. Oysa ki 1. yöntemde, bilinmeyen modeldeki hatalı vektörler tüm modelleri aynı anda olumsuz yönde etkiler.



Şekil 3.11. 1. Yöntemde modellerin karşılaştırılması

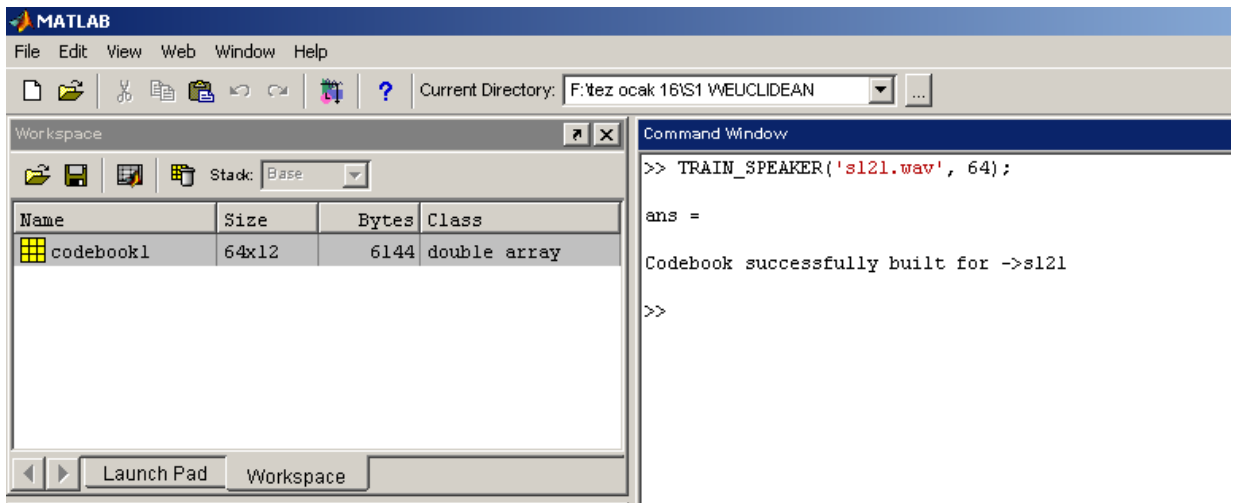
## 4. SİSTEMİN GERÇEKLEŞTİRİMİ

Konuşmacı tanıma sistemi için mevcut bulunan yöntemin test edilmesi ve yeni yöntemin tasarlanıp test edilmesi için Matlab 6 kullanıldı. Matlab, dünya çapında, mühendisler ve bilim adamları tarafından kabul görmüş, yüksek düzey bir hesaplama ve grafikleme aracıdır. Karmaşık matematiksel ve istatistiksel operasyonlar için hazır kütüphaneler barındırmakla birlikte, etkileşimli, matrise dayalı programlama yapısı ile çok kolay ve hızlı bir çalışma ortamı sağlamaktadır.

### 4.1. Geliştirilen Matlab Fonksiyonları

Sistemin gerçekleştirimi için tasarlanıp yazılan matlab fonksiyonları ve açıklamaları aşağıdaki gibidir:

**TRAIN\_SPEAKER.m:** Verilen bir ses örneği için kod kitabı oluşturur. İlk olarak sayısal ses sinyali değerleri hafızaya alınır. Bu değerler girdi olarak mel-kepstrum fonksiyonuna gönderilir ve sonuç olarak mel-kepstrum katsayılarından oluşan kod vektörleri (öznitelik vektörleri) bulunur. Bulunan vektör sayısını belli bir sayıya indirmek için (bu tezde standart olarak 64 kod vektörü kullanılmıştır) LBG Vektör Kuantumlama algoritması kullanılır. Bu fonksiyonun örnek kullanımı ve çıktısı şekil 4.1’de görüldüğü gibidir.

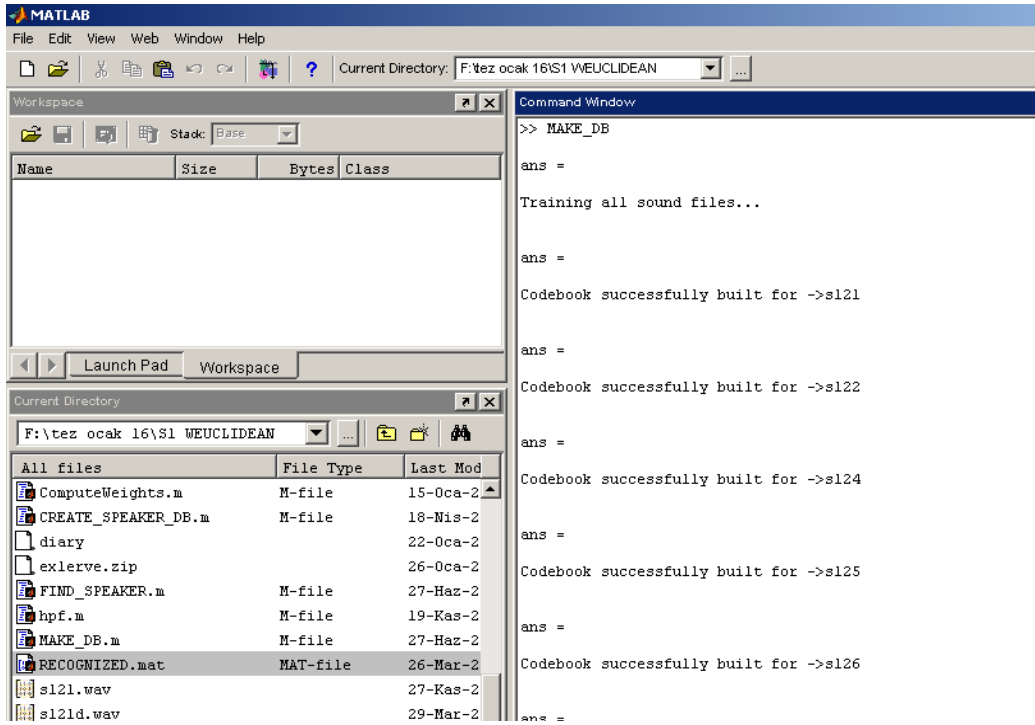


Şekil 4.1. TRAIN\_SPEAKER fonksiyonu kullanımı

Fonksiyon, 1. parametre olarak ses dosyasının ismini alırken 2. parametre olarak ta oluşturulacak öznitelik vektör sayısını alır. Sonuç olarak ta 64 x 12'lik bir matris döndürür.

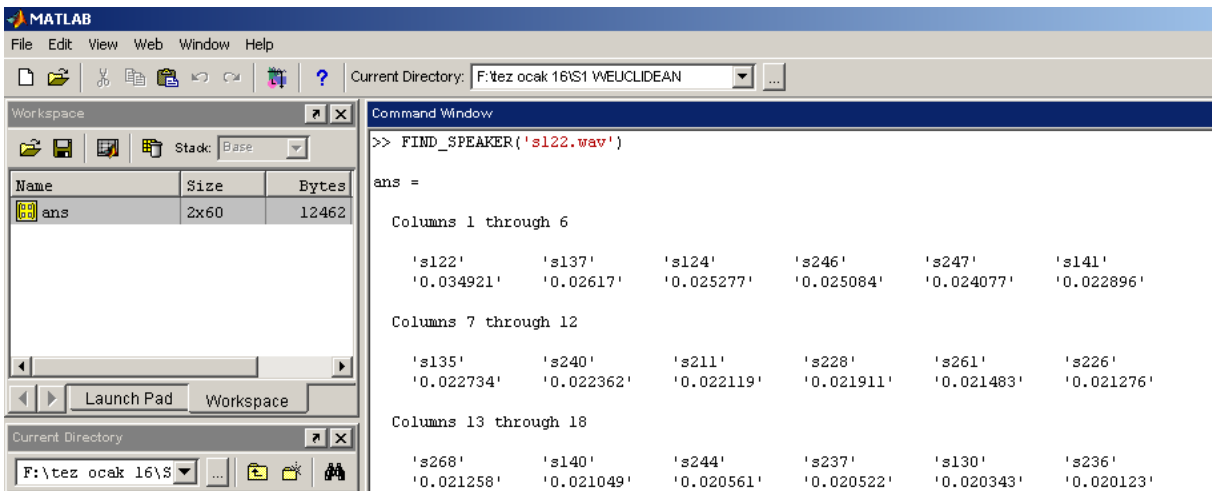
**CREATE\_SPEAKER\_DB.m:** Tüm kullanıcıların iki boyutlu kod kitaplarını olarak 3 boyutlu bir matriste saklar. Her kod kitabı, 64 satırlık ve her satırda 12 MFCC katsayılarını içeren 64 x 12 boyutunda bir matristir. 40 konuşmacı için oluşacak 3 boyutlu matrisin boyutları 64 x 12 x 40 olacaktır. Bu fonksiyon aynı zamanda her kod kitabı için, kod vektörlerinin ağırlık değerlerini içeren bir ağırlık (weight) matrisi oluşturur. Bir kod kitabı için oluşan ağırlık matrisi, kod vektörlerinin sayısı kadar kolona sahiptir ve tek satırdan oluşur. Her kolon bir kod vektörüne karşılık gelen ağırlık değerini içerir. Tüm ses dosyaları için ağırlıklar hesaplanıp bir matrise yerleştirirler. 40 ses dosyası için 40 x 64 boyutunda bir ağırlık matrisi oluşturulacaktır. Bu fonksiyon hiç parametre almaz, bilinen bir klasördeki tüm kod kitaplarını otomatik olarak bulup işler.

**MAKE\_DB.m:** Konuşmacı veritabanının oluşturulması için her ses dosyasının, TRAIN\_SPEAKER fonksiyonu kullanılarak çalıştırılması gerekmektedir. Bu işlemleri elle yapmak uzun zaman alacaktır. İşte bu fonksiyon, işlemi otomatikleştirir ve bir klasördeki tüm ses dosyalarını okuyarak ilgili fonksiyonu çağırır ve kod kitaplarını oluşturur. Hemen ardından oluşan kod kitaplarını tek bir matriste toplamak ve ağırlık matrislerini oluşturmak üzere de CREATE\_SPEAKER\_DB fonksiyonunu çağırır. Dolayısı ile MAKE\_DB fonksiyonu, kaydolma aşamasını tek bir komutla yapmayı mümkün kılar. Bu fonksiyonun çalışma görüntüsü şekil 4.2'de verilmiştir.



Şekil 4.2. MAKE\_DB çalışma görüntüsü

**FIND\_SPEAKER.m:** Verilen bir ses örneğini veritabanındaki modellerle karşılaştırıp her modele olan benzerlik değerlerini bulur ve değerlerin büyüklüklerine göre sıralar. Konuşmacı tanımada en son adım olan saptama işlemini bu fonksiyon gerçekleştirir. Parametre olarak ses dosyasının ismini ve vektör sayısını alır. Şekil 4.3'te bilinmeyen bir ses dosyasının benzerlik değerleri bu fonksiyon kullanılarak bulunmuştur.



Şekil 4.3. FIND\_SPEAKER fonksiyonu ile saptama

**AUTO\_FIND.m:** Konuşma saptama sisteminin başarı oranını ve güvenilirliğini göstermek üzere bir çok test yapılması gerekiyordu. Tüm bu testlerin yapılması için, bir ses dosyası kümesindeki tüm sesler için bu işi otomatikleştiren bir fonksiyon yazıldı. AUTO\_FIND fonksiyonu hiç parametre almadan, bilinen bir klasördeki tüm ses örneklerini alarak FIND\_SPEAKER fonksiyonuna gönderir. Fonksiyon, çıktı olarak, saptanan tüm sesler için RECOGNIZED.mat, saptanamayan sesler için de UNRECOGNIZED.mat isimli bir Matlab dosyası üretir. Fonksiyon, bir ses örneğinin doğru saptanabildiğini, dosyasının ismini, veritabanında en yüksek benzerlik değerine sahip olarak bulunan modelin ismiyle karşılaştırarak bulur. Aynı isimde iseler saptama başarılıdır.

```

MATLAB
File Edit View Web Window Help
Current Directory: F:\tez ocak 16\S1 WEUCLIDEAN

Workspace
Stack: Base
Name      Size      Bytes      Class
ans       2x60      12462      cell array

Command Window
>> AUTO_FIND
ans =
AUTO_FIND STARTED...

ans =
s121.wav is recognized.. drate =29.9504

ans =
s122.wav is recognized.. drate =25.0107

ans =
s124.wav is recognized.. drate =18.0688

ans =
s125.wav is recognized.. drate =18.6605

ans =
s126.wav is recognized.. drate =22.2532

Current Directory
F:\tez ocak 16\S1 WEUCLIDEAN
All files      File Type      Last
40_3.TXT      Text           12-Ar
40_4.TXT      Text           13-Ar
40_5.TXT      Text           13-Ar
40_6.TXT      Text           13-Ar
40_7.TXT      Text           13-Ar
AUTO_FIND.m   M-file        05-0
cell2num.m    M-file        29-0
ComputeWeights.m M-file        15-0

```

Şekil 4.4. AUTO\_FIND fonksiyonu çalışma görüntüsü

Şekil 4.4'te bu fonksiyonun çalışma görüntüsü verilmiştir. Şekilde görülen *drate* değeri 1. benzerlik değeri ile 2. benzerlik değerleri farkının 1. benzerlik değerine oranının yüzdesel ifadesidir. Bu değer, iki sistemin güvenilirliklerinin karşılaştırılması için kullanılır.

## 4.2. Kullanılan Kod Kütüphaneleri

Geliştirilen Matlab fonksiyonlarını, ses işleme ile ilgili bir çok genel fonksiyonu içeren bir araç olan VOICEBOX kütüphanesini kullanmıştır. Bu kütüphane, İngiltere Imperial Koleji, Elektrik Elektronik Mühendisliği Bölümünden Mike Brookes tarafından yazılmış olup 2005 tarihi itibarı ile <http://www.ee.ic.ac.uk/hp/staff/dmb/voicebox/voicebox.html> adresinden GNU lisansı ile indirilebilir.

## 5. TEST SONUÇLARI

Önerilen yeni yöntem uluslararası YOHO veritabanından alınan ses örnekleri ile test edilmiştir. Yoho veritabanı, kaydolma ve saptama aşamaları için ayrı kümeler içermektedir. Testlerde, 40 konuşmacıdan oluşan bir küme kaydolma aşamasında, diğer bir 40 kişilik küme ise saptama aşamasında kullanılmıştır. 40 konuşmacı içindeki tanıma oranları şekil 5.1’de gösterilmektedir. Tanıma oranları ağırlıklı karşılaştırma yöntemi için %92,5, önerilen yöntem için de %97,5 bulunmuştur. S1, ağırlıklı karşılaştırma yöntemini, S2 de yeni yöntemi temsil etmektedir.

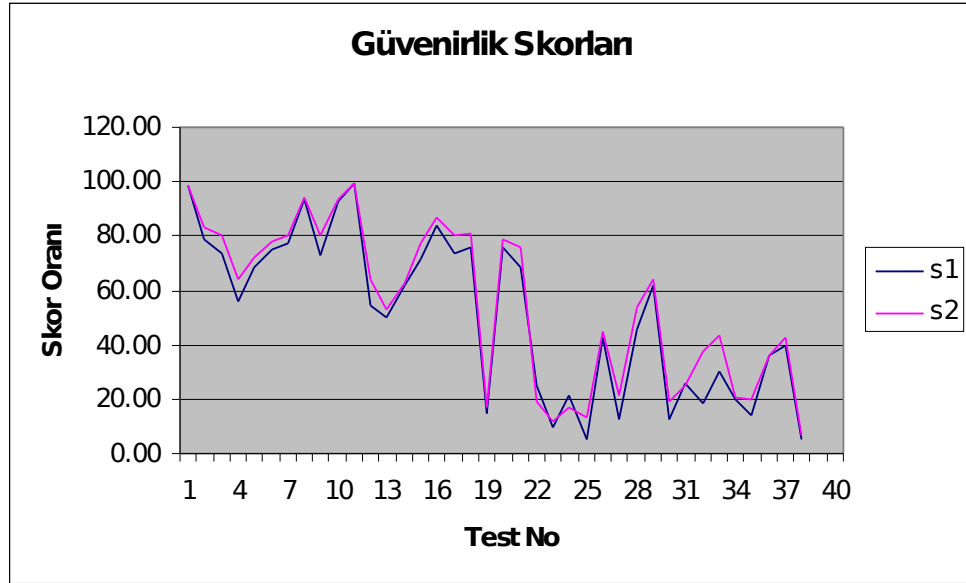
Konuşmacı Sayısı	Tanıma Oranları S1	Tanıma Oranları S2
40	37 (%92,5)	39 (%97,5)

**Şekil 5.1.** Test sonuçları

Yöntemlerin güvenilirliklerini karşılaştırmak için de doğru saptanan konuşmacı modelleri ile en iyi 2. benzerlik değerine sahip konuşmacıların benzerlikleri arasında bir güvenilirlik skoru hesaplanmıştır. Güvenirlik skoru (4) te belirtildiği gibi hesaplanır. sb1 en iyi benzerlik skorunu, sb2 de 2. en iyi benzerlik skorunu temsil etmektedir.

$$S_c = (S_{b1} - S_{b2}) / S_{b1} \quad (4)$$

İki yöntem için de güvenilirlik skorları şekil 5.2’de gösterilmiştir. S1, 1. yöntemin güvenilirlik skorlarını, S2 ise 2. yöntemin güvenilirlik skorlarını temsil etmektedir. 1. Yöntemin ortalama güvenilirlik skoru 48,52, 2. yöntemin skoru ise 52,46 olarak bulunmuştur. Test sonuçları göstermektedir ki önerilen yeni yöntem doğru konuşmacıyı ortalama olarak %7,5 oranında daha iyi saptamaktadır.



**Şekil 5.2** Güvenirlik skorları

## 6. SONUÇ

Bu tezde, ses ve dijital sinyal işleme temelleri ile birlikte, konuşma tanıma konusu işlenmiş, bu alanda yapılmış bir çalışma incelenerek yeni bir yöntem sunulmuştur. Uluslar arası konuşmacı veritabanı olan YOHO sisteminden 40 kişilik konuşma kümesi ile yapılan testler göstermiştir ki önerilen yeni algoritma gerek tanıma gerek güvenilirlik açısından önceki algoritmaya göre daha iyi sonuçlar üretmektedir.

## KAYNAKLAR

- [1] E. Newman, "The Biometric Report," SJB Services, UK, 1995
- [2] G. Doddington, "Speaker Recognition-Identifying People by their Voices," Proc. IEEE, vol. 73, no. 11, pp. 1651-1664, 1985
- [3] S. Furnell and N. Clarke, "Biometrics - No Silver Bullets," Computer Fraud & Security, pp. 9-14, 2005
- [4] A. K. Jain, et al., "Biometric recognition: security and privacy concerns," IEEE Security & Privacy Magazine, vol. 1, no: 2, pp. 33-42, 2003
- [5] R. J. Mammone, et al., "Robust speaker recognition: a feature-based approach," IEEE Signal Processing Magazine, vol. 13, no. 5, pp. 58, 1996
- [6] M. Sambur, "Speaker recognition using orthogonal linear prediction," IEEE Trans. Acoust., Speech, Signal Processing, vol. 24, no. 4, pp. 283-289, 1976
- [7] J. D. Markel and S. B. Davis, "Text-independent speaker recognition from a large linguistically unconstrained time-spaced database," IEEE Trans. Acoust., Speech, Signal Processing, vol. ASSP-27, no. 1, pp. 74-82, 1979
- [8] R. Schwartz, et al., "The application of probability density estimation to text independent speaker identification," IEEE Int. Conf. Acoustics, Speech, and Signal Processing, vol. 7, pp. 1649-1652, 1982
- [9] K. P. Li and E. H. Wrench, Jr., "Text-independent speaker recognition with short utterances," IEEE Int. Conf. Acoustics, Speech, and Signal Processing, vol. 8, pp. 555-558, 1983
- [10] F. Soong, et al., "A vector quantization approach to speaker recognition," IEEE Trans. Acoust., Speech, Signal Processing, vol. 10, pp. 387-390, 1985
- [11] T. Matsui and S. Furui, "A text-independent speaker recognition method robust against utterance variations," IEEE Int. Conf. Acoustics, Speech, and Signal Processing, pp. 377-380, 1991
- [12] D.A. Reynolds and R.C. Rose, "Robust text independent speaker identification using gaussian mixture speaker models," IEEE Transactions on Speech and Audio Processing, vol. 3 no.1 pp. 72-83, 1995
- [13] Zhong-Xuan Yuan, et al., "Binary quantization of feature vectors for robust text-independent speaker identification," IEEE Transactions on Speech and Audio Processing, vol. 7, no. 1, pp. 70-78, 1999

- [14] Jin, Q., Waibel, A., "Application of lda to speaker recognition," International Conference on Spoken Language Processing, vol. 2, pp. 250-253, 2000
- [15] M. Selvi, ve ark., "Use of Wide Band MFCC and Wide Band RCC's for Robust Speaker Identification," Reg. Signal Proc. Conf., 2002
- [16] Jiuqing Deng, Qixiu Hu, "Open set text-independent speaker recognition based on set-score pattern classification," IEEE Int. Conf. Acoustics, Speech, and Signal Processing, vol. 2, no. 2, pp. 73-6, 2003
- [17] L. Zhang, et al., "Codebook design using genetic algorithm and its application to speaker identification," Electronics Letters, vol. 41, no. 10, pp. 619-620, 2005
- [18] P. R. Cook, "Sound Production and Modeling," IEEE Computer Graphics and Applications, vol. 22, no. 4, pp. 23-27, 2002
- [19] C. E. Shannon, "Communication in the presence of noise," IEEE, vol. 72, no. 9, pp. 1192-1201, 1984
- [20] H. Nyquist, "Certain topics in telegraph transmission theory," Proceedings of the IEEE, vol. 90, no. 2, pp. 280-305, 2002
- [21] E. Meijering, "A chronology of interpolation: from ancient astronomy to modern signal and image processing," IEEE vol. 90, no. 3, pp. 319-342, 2002
- [22] S. Smith, The Scientist and Engineer's Guide to Digital Signal Processing, California Technical Publishing, San Diego, 1997
- [23] Thomas F. Quatieri, Discrete-Time Speech Signal Processing: Principles and Practice, Prentice Hall, MIT Lincoln Laboratory, 2002
- [24] Lan Wang, et al., "Capture interspeaker information with a neural network for speaker identification," IEEE Neural Networks, vol. 13, no. 2, pp. 436-445, 2002
- [25] E. Erzin, ve ark., "Multimodal Speaker Identification Using an Adaptive Classifier Cascade Based on Modality Reliability," IEEE Multimedia, vol. 7, no. 5, pp. 840-852, 2005
- [26] S. Davis, P. Mermelstein, "Comparison of parametric representations for monosyllabic word recognition in continuously spoken sentences," IEEE Acoustics, Speech, and Signal Processing, vol. 28, no. 4, pp. 357-366, 1980
- [27] S. Furui, "Cepstral analysis technique for automatic speaker verification," IEEE Acoustics, Speech, and Signal Processing, vol. 29, no. 2, pp. 254-272, 1981

- [28] Heinz Hertlein, et al., "Pass Phrase Based Speaker Recognition for Authentication," BIOSIG, pp. 71-80, 2003
- [29] Jialong He, et al., "A new codebook training algorithm for VQ-based speaker recognition," IEEE Int. Conf. Acoustics, Speech, and Signal Processing, vol. 2, pp. 1091-1094, 1997
- [30] T. E. F. Filho, et al., "Learning vector quantization in text-independent automatic speaker recognition," Vth Brazilian Symposium on Neural Networks, pp. 135-139, 1998
- [31] Li Liu, et al., "Signal modeling for speaker identification," IEEE Int. Conf. Acoustics, Speech, and Signal Processing, vol. 2, pp. 665-668, 1996
- [32] Ningping Fan, Justinian P. Rosca, "Enhanced VQ-Based Algorithms for Speech Independent Speaker Identification," Audio and Video based Biometric Person Authentication, pp. 470-477, 2003
- [33] T. Kinnunen, et al. "Real-Time Speaker Identification and Verification," IEEE Transactions on Speech and Audio Processing vol. PP, no. 99, pp. 1-12, 2005
- [34] H. Gish and M. Schmit, "Text-Independent Speaker Identification," IEEE Signal Processing Magazine, vol. 11, pp. 18-32, 1994
- [35] R. W. Schafer and L. R. Rabiner, "Digital representations of speech signals," IEEE, vol. 63, pp. 662-677, Apr. 1975
- [36] M. R. Schroeder and B. S. Atal, "Generalised short-time power spectra and autocorrelation functions," J. Acoust. Soc. Amer., vol. 34, pp. 1679-1683, 1962
- [37] B. S. Atal, "Automatic recognition of speakers from their voices," IEEE, vol. 64, no. 4, pp. 460-475, 1976
- [38] S. Molau, et al., "Computing Mel-frequency cepstral coefficients on the power spectrum," IEEE Int. Conf. Acoustics, Speech, and Signal Processing, vol. 1, pp. 73-76, 2001
- [39] T. Kinnunen and P. Fränti, "Speaker Discriminative Weighting Method for VQ-Based Speaker Identification," AVBPA, pp. 150-156, 2001
- [40] L. P. Heck, M. Weintraub, "Handset-dependent background models for robust text-independent speaker recognition," IEEE Int. Conf. Acoustics, Speech, and Signal Processing, vol.2, pp. 1071-1074, 1997
- [41] Um Ig-Tae, et al., "Comparison of clustering methods for MLP-based speaker verification," International Conference on Pattern Recognition, vol. 2, pp. 475-478, 2000

- [42] R. Vergin and D. O'Shaughnessy, "On the use of some divergence measures in speaker recognition," IEEE Int. Conf. Acoustics, Speech, and Signal Processing, vol. 1, pp. 309-312, 1999
- [43] Larry P. Heck, et al., "Robustness to telephone handset distortion in speaker recognition by discriminative feature design," Speech Communication, vol. 31, no. 2-3, pp. 181-192, 2000

## ÖZGEÇMİŞ

Ozan MUT 02.03.1979 yılında Tarsus/İÇEL de doğdu. 1996'da Cengiz Topel Lisesi Matematik Bölümünü bitirdikten sonra 1997'de girdiği Marmara Üniversitesi Bilgisayar Mühendisliği Bölümünden 2002 yılında mezun oldu. Aynı yıl Gebze Yüksek Teknoloji Enstitüsü Bilgisayar Mühendisliği Bölümüne Yüksek Lisans öğrencisi olarak girdi.