

## VQ Yöntemiyle Konuşmacı Cinsiyetinin Belirlenmesi

Vasif V. Nابیev<sup>1</sup>

Ergün Yücesoy<sup>2</sup>

### Özet

Bu çalışmada konuşmacı cinsiyetinin metinden bağımsız olarak belirlenmesi amaçlanmaktadır. Önerilen sistem iki bölümden oluşmaktadır. Birinci bölüm olan eğitim aşamasında deneklerden alınan ses kayıtlarından öznitelik vektörü hesaplanır. Çalışmada öznitelik vektörü olarak MFCC(Mel Frequency Cepstral Coefficients) kullanılmıştır. Elde edilen MFCC öznitelik vektörü VQ (Vector Quantization) yöntemiyle sınıflandırılır ve veritabanına kaydedilerek eğitim aşaması tamamlanır. İkinci bölüm olan test aşamasında konuşmacı cinsiyeti bilinmeyen ses kayıtları giriş olarak alınır ve eğitim aşamasındaki gibi öznitelik vektörü hesaplanır. Elde edilen öznitelik vektörü eğitim veritabanındaki verilerle kıyaslanarak erkek ve bayan sınıflar için ortalama bir uzaklık değeri hesaplanır. Bu uzaklık değerlerinden küçük olanı test verisinin hangi sınıfa ait olduğunu belirtir. Çalışmada TIMIT veritabanı üzerinde çeşitli testler yapılmıştır. Bu testlerden 168 konuşmacının 10'ar cümle söylediği toplam 1680 veriden oluşan test kümesinde yalnızca 34 hatalı karar verilerek %98,80 başarı elde edilmiştir.

**Anahtar kelimeler:** Konuşmacı tanıma, cinsiyet tanıma, vektör niceleme (VQ)

### 1. Giriş

Konuşma insanlar arasındaki iletişimin en kolay ve doğal şeklidir. Konuşma sırasında dinleyiciye yalnızca kelimeler iletilmez. Aynı zamanda konuşmacı hakkında kimlik, yaş, cinsiyet, ruh hali gibi bilgilerde iletilir. İnsanlar arasındaki bu iletişimin bilgisayarla da kurulması için yoğun çalışmalar yapılmaktadır. Özellikle biyometrik sistemlerde sesin kullanımı hem maliyet hem de kullanım kolaylığı açısından avantaj sağlamaktadır. Örneğin ortama yerleştirilen bir mikrofon aracılığıyla kişinin haberi bile olmadan ses kaydı alınabilir ve sistem uygulanabilir. Ayrıca uzaktan erişim kolaylığı da ses biyometrisinin avantajlarından bir diğeridir. Bu çalışmada ses dalgasının içinde barındırdığı kişisel bilgilerden konuşmacı cinsiyetinin otomatik olarak belirlenmesi amaçlanmaktadır. Bu

<sup>1</sup> Prof. Dr. KTÜ, Mühendislik Fakültesi, Bilgisayar Mühendisliği Bölümü, [vasif@ktu.edu.tr](mailto:vasif@ktu.edu.tr)

<sup>2</sup> Öğr. Gör. Ordu Üniversitesi, Ordu Meslek Yüksek Okulu, [yusesoy@hotmail.com](mailto:yusesoy@hotmail.com)

gerçekleştirildiğinde diğer ses işlem uygulamalarında cinsiyete göre modeller oluşturularak başarı artırılabilir.

Üst düzeyde bütün konuşmacı tanıma sistemleri öznitelik çıkarma ve eşleştirme olmak üzere iki bölümden oluşur. Öznitelik çıkarma ses sinyalinden konuşma ve konuşmacıyı temsil edecek en az miktarda bilginin çıkarılması işlemidir. Cinsiyet belirlemede kullanılan iki yaklaşım vardır. Birincisi cinsiyete bağımlı bir öznitelik olan perde frekansının kullanımınıdır (Vergin, Farhat & O'Shaughnessy, 1996; Eskidere & Ertaş, 2007; Nahiye & Yücesoy, 2008). Diğer yaklaşım ise MFCC gibi sesin akustik özelliklerini kullanan genel örüntü tanıma yaklaşımıdır (Harb & Chen, 2003). Perde frekansının belirlenmesi ses kalitesine son derece bağımlı olduğu için her iki yaklaşımın birlikte kullanıldığı çalışmalarda yapılmıştır (Parris & Carey, 1996; Ting, Yingchun & Zhaohui, 2006).

Öznitelik eşleştirme ise bilinmeyen kullanıcının ses girişinden elde edilen özniteliklerin bilinen konuşmacı kümesiyle karşılaştırılması işlemidir. Karşılaştırılma sırasında kullanılan metne göre sistem metne bağımlı ve metinden bağımsız olarak ikiye ayrılır. Metne bağımlı sistemlerde eğitim ve test aşamasında aynı metin kullanılırken, metin bağımsız sistemlerde farklı metin kullanılır. Uygulanan yöntemlerde metne bağımlı ve metinden bağımsız olarak ikiye ayrılır. Günümüzde DTW ve HMM metne bağımlı (Matsui & Funii, 1992; Silverman & Morgan, 1990; Doddington, 1985), VQ ve GMM ise metinden bağımsız sistemlerde kullanılan yöntemlerdendir (Rosenberg & Soong, 1987; Reynolds & Rose, 1995).

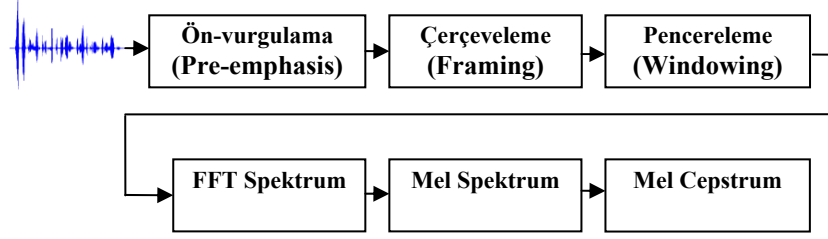
Bu çalışmada insan kulağının sesi algılamasını model alan MFCC özniteliği vektör nicelme (VQ) yöntemiyle birlikte kullanılmıştır. MFCC öznitelik vektörünün belirlenmesi bölüm 2 de, öznitelik vektörünün VQ yöntemiyle cinsiyete göre sınıflandırılması bölüm 3 de anlatılmıştır. Konuşmacı cinsiyetinin belirlenmesi bölüm 4 de, test ve deneysel sonuçlar da bölüm 5 de verilmiştir.

## 2. Öznitelik Çıkarma

Konuşma sinyali yalnızca konuşma bilgisi içermez. Aynı zamanda kimlik, yaş, cinsiyet, ruh hali gibi konuşmacıya bağımlı bilgilerde konuşma sinyalinden çıkarılabilir (Naik, 1990). Öznitelik çıkarma konuşma sinyalinden bu bilgileri temsil eden az sayıda parametrenin belirlenmesi işlemidir. Bu aşama tanıma sistemleri için son derece önemlidir ve başarıyı doğrudan etkiler. Konuşma sinyali yavaş değişen bir sinyaldir ve sözde durağan olarak isimlendirilir. Bu nedenle çoğu analiz yöntemleri konuşma sinyalinin durağan akustik özelliklere sahip olduğu kısa parçalar boyunca uygulanır. Konuşma sinyalini parametrik olarak temsil eden LPC, MFCC, PLP gibi birçok yöntem vardır (Campbell, 1997; Deller, Hansen & Proakis, 2000; Rabiner & Juang, 1993). Bu çalışmada MFCC özniteliği kullanılmıştır.

MFCC özniteliği insan kulağının frekans algılamasını model alan ve en çok tercih edilen yöntemlerden birisidir. Yapılan araştırmalar sonunda insanın frekans içeriklerini 1Khz ye

kadar lineer 1Khz'nin üzerinde ise logaritmik olarak algıladığı görülmüştür (Deller, Hansen & Proakis, 2000; Rabiner & Juang, 1993). MFCC mel ölçeği olarak isimlendirilen bu ölçeğe göre konuşmayı parametreleştiren bir yöntemdir. Yöntemin blok diyagramı Şekil 1'de gösterilmiştir.



Şekil 1.MFCC öznelik vektörünün çıkarılması

### 2.1. Ön vurgulama

Yüksek frekanslı sesli bölgelerde ses üretim sisteminin yapısından kaynaklanan bir sönümlenme oluşur. Bu nedenle ön vurgulama olarak isimlendirilen ve yüksek frekanslı bölgeleri güçlendiren bir filtreleme uygulanır (Picone, 1993). Ön vurgulamanın amacı ani düşüşlerin olduğu sesli bölgelerin spektrumunu dengelemektir. Ön vurgulama aynı zamanda bazı ağızsız etkileri de ortadan kaldırır. Yaygın olarak kullanılan ön vurgu fitresi (1) bağıntısıyla gerçekleştirilir.

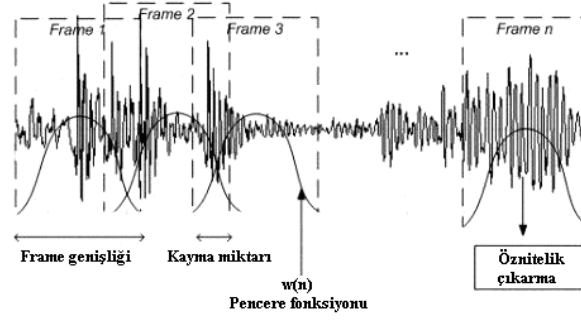
$$Y[n]=x[n]-a*x[n-1], \quad a \approx (0,95-0,97) \quad (1)$$

Bu çalışmada  $a=0,97$  olarak seçilmiştir.

### 2.2. Çerçeveleme ve pencereleme

Bütün ses analiz yöntemlerinde olduğu gibi MFCC yöntemi de sesin durağan akustik özellikler sahip olduğu kabul edilen kısa parçalar boyunca uygulanır (Deller, Hansen & Proakis, 2000; Rabiner & Juang, 1993). Bu parçalar genellikle 20-30ms olarak seçilir ve 10-15ms lik kayma miktarlarında sinyal boyunca hareket ettirilerek sinyalin tümüne uygulanır (Şekil 2). Böylece her çerçeve kendinden önceki çerçevenin bir bölümünü içerisinde barındırır. Elde edilen analiz pencerelerinin sınırlarındaki süreksizliği azaltmak için bir pencere fonksiyonu uygulanır. Ses uygulamalarında genellikle (2) bağıntısıyla verilen hamming penceresi tercih edilir.

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

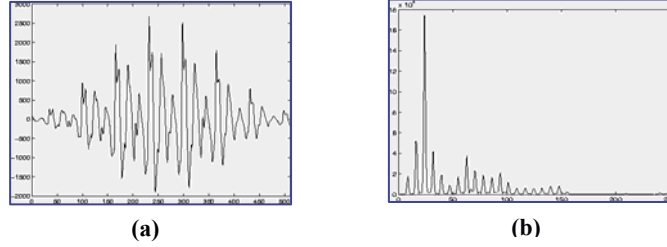


Şekil 2. Pencere fonksiyonuna tabi tutulmuş çerçevenilmiş ses sinyali

### 2.3. FFT ve Mel spektrum

Analiz pencerelerine bölünen konuşma sinyali ayrık fourier dönüşümüyle frekans bileşenlerine ayrılır. (3) bağıntısıyla gerçekleştirilen ayrık fourier dönüşümü karmaşık konuşma sinyalini ayrıştırarak hangi frekansta ne şiddette bir titreşim olduğunu hesaplar. Bu dönüşüm için genellikle hızlı bir algoritmanın uygulandığı FFT yöntemi tercih edilir. Şekil 3 de bir ses sinyali ve FFT sonucu elde edilen frekans bileşenleri gösterilmektedir.

$$X_n = \sum_{k=0}^{N-1} x_k e^{-2\pi jkn/N}, n=0,1,2,\dots,N-1 \quad (3)$$

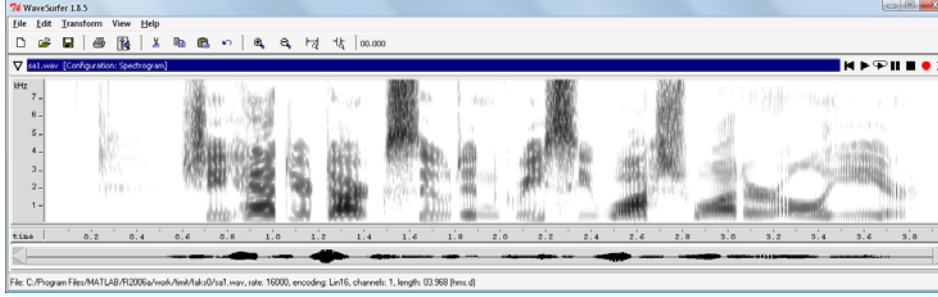


Şekil 3. Ses sinyalinin (a) Zaman uzayı, sn ve (b) Frekans uzayında, Hz görünümü

Tüm analiz pencereleri FFT yöntemiyle frekans bileşenlerine dönüştürülerek spektogram olarak adlandırılan gösterim elde edilir. Spektogram; yatay eksenin zaman, dikey eksenin frekans ve keşişim noktasının ise o frekanstaki genliği belirtecek şekilde bir renkle ifade edildiği üç boyutlu bir gösterim şeklidir. Spektogram gösterimi için genellikle gri renk tonları kullanılarak konuşma sinyali içindeki baskın frekanslar (formant frekansları) koyu hatlarla temsil edilir. Şekil 4 de bir bayan konuşmacının seslendirdiği bir konuşma sinyali ve spektogramı gösterilmiştir.

FFT yöntemiyle elde edilen frekans bileşenleri mel ölçeği olarak bilinen ve insanın frekans algılamasını temsil eden ölçeğe dönüştürülerek mel-spektrumu elde edilir. Mel

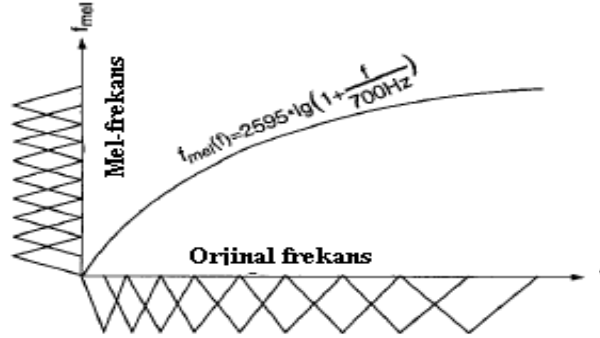
ölçeği 1Khz'ye kadar lineer 1Khz'nin üzerinde logaritmik karakteristiğe sahiptir ve (4) bağıntısıyla temsil edilir.



Şekil 4. WaveSurfer programıyla elde edilmiş bir spektrogram görüntüsü

Dönüşüm işleminde bant genişliği mel ölçeğine göre lineer olarak değişen üçgen filtreler kullanılır. Genellikle filtre katsayısı olarak 20 ile 30 arasında bir değer seçilir. Şekil 5'de mel-fitre kümesinin oluşturulması grafiksel olarak gösterilmiştir.

$$\text{Mel}(f)=2595*\log(1+f / 700) \quad (4)$$



Şekil 5. Mel-fitre kümesi

#### 2.4. Mel cepstrum

Son aşamada sinyalin genlik spektrumu mel ölçeğinde eşit aralıklarla dağılmış ve birbirleriyle %50 oranında kesişen N adet band geçiren üçgen filtreyle çarpılır (Campbell, 1997; Deller, Hansen & Proakis, 2000). Çarpım sonucunda her bir filtre içinde kalan enerjinin logaritması hesaplanır. Mel spektrum katsayılarının logaritması gerçek sayılar olduğu için (5) bağıntısıyla verilen ayrık kosinüs dönüşümü kullanılarak tekrar zaman bölgesine geçilir. Sonuçta elde edilen katsayılar mel-frekanslı cepstrum katsayısı (MFCC) olara isimlendirilir.

$$\tilde{c}_n = \sum_{k=1}^K (\log \tilde{S}_k) \cos \left[ n \left( k - \frac{1}{2} \right) \frac{\pi}{K} \right], n = 1, 2, \dots, K \quad (5)$$

Burada  $\tilde{S}_k$ ,  $k=1,2,\dots,K$  mel spektrum katsayılarıdır. Dönüşüm sonucu elde edilen ilk bileşen  $\tilde{c}_0$  ortalama logaritmik enerjiyi temsil ettiği için genellikle öznelik vektöründen çıkarılır. Bu çalışmada öznelik vektörü olarak ilk 20 MFCC katsayısı kullanılmıştır.

### 3. Sınıflandırma

Konuşmacı özneliklerinin ses sinyalinden çıkarılmasından sonra sınıflandırma aşamasına geçilir. Sınıflandırmada amaç konuşma sinyalinden elde edilen öznelik vektör uzayını belirli sayıda alt bölgeye ayırmaktır. Bu aşama genellikle iki bölüme oluşur; modelleme ve eşleştirme. Modelleme konuşma örneklerinden elde edilen öznelikler temelinde erkek ve bayan için bir model oluşturularak kaydedilmesi işlemidir. Eşleştirme ise model ile bilinmeyen konuşmacı örneklerinden elde edilen öznelikler arasındaki benzerliği belirten bir eşleştirme skorunun hesaplanması işidir.

Günümüzde ses işlemede kullanılan değişik sınıflandırma teknikleri vardır. Bu çalışmada yüksek başarı oranı ve kolay uygulanabilme özelliğinden ötürü VQ yöntemi kullanılmıştır.

#### 3.1. Vektör niceleme (VQ)

Vektör niceleme geniş bir vektör uzayından sınırlı sayıda bölgeye dönüşüm gerçekleştiren bir tür veri sıkıştırma yöntemidir. Bu yöntemle her bir konuşmacıdan elde edilen öznelik vektörü M adet bölgeye gruplandırılarak konuşmacı modeli oluşturulur. Her bir bölge kod kelimesi olarak isimlendirilen merkez noktasiyla temsil edilir. Kod kelimeleri ise kod kitabını oluşturur. N adet eğitim vektörü kümesini M adet ( $M < N$ ) kod kitabı vektörüne sıkıştırmada kullanılan yöntemlerden birisi LBG algoritmasıdır (Linde, Bum & Gray, 1980). Bu algoritma aşağıdaki özyinelemeli prosedür takip edilerek gerçekleştirilir.

1. Bir boyutlu kod kitabı oluştur: Bu tüm eğitim vektör kümesinin merkezidir.
2. O anki kod kitabı  $\mathbf{y}_n$ 'nin boyutunu aşağıdaki kurala göre iki kat büyüt.

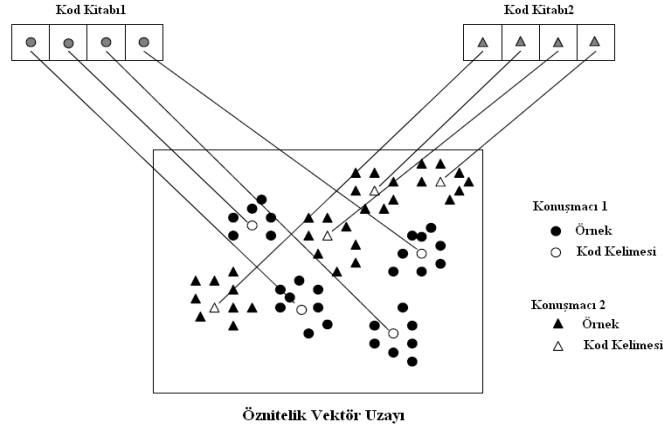
$$\mathbf{y}_n^+ = \mathbf{y}_n(1 + \varepsilon)$$

$$\mathbf{y}_n^- = \mathbf{y}_n(1 - \varepsilon)$$

$\varepsilon$  ise bölümlenme parametresi ( $\varepsilon \approx 0.01-0.05$ )

3. En yakın komşu araması: Her bir eğitim vektörü için o anki kod kitabında en yakın kod kelimesini bul ve o vektörü uygun hücreye ata.
4. Merkez güncelleme: O hücreye atanan eğitim vektörlerinin merkezini kullanarak her bir hücredeki kod kelimesini güncelle.
5. Ortalama mesafe belli bir eşik değerinin altına düşene kadar adım 3 ve 4'ü tekrarla.
6. Kod kitabı boyu M oluşturulana kadar adım 2, 3 ve 4'ü tekrarla

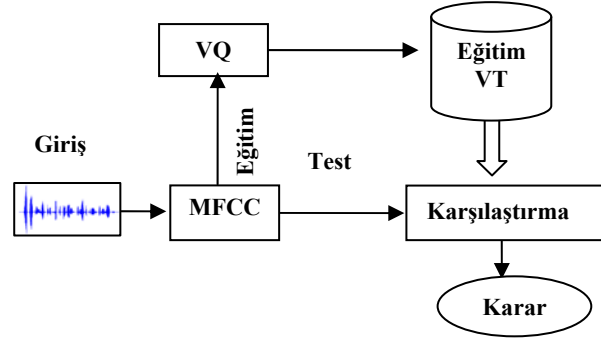
İki konuşmacıya ait iki boyutlu öznelik uzayının VQ yöntemiyle sınıflandırılması Şekil 6'de gösterilmiştir.



Şekil 6. İki konuşmacı için vektör niceleme

#### 4. Otomatik Cinsiyet Belirleme

Bu çalışmada konuşmacı cinsiyetini otomatik olarak belirleyen metinden bağımsız bir sistem önerilmiştir. Sistem üç aşamadan oluşur. Birinci aşamada cinsiyeti bilinen konuşmacıların ses örneklerinden MFCC öznelik vektörü hesaplanır. Çalışmada öznelik vektörü olarak ilk 20 MFCC katsayısı kullanılmıştır. İkinci aşamada VQ yöntemi kullanılarak her bir konuşmacı için bir kod kitabı oluşturulur. Elde edilen kod kitabı cinsiyet bilgisiyle beraber veritabanına kaydedilerek eğitim aşaması tamamlanmış olur. Son aşamada ise bilinmeyen konuşmacı sesinden elde edilen öznelik vektörü veritabanındaki konuşmacı modelleriyle karşılaştırılır. Karşılaştırma sonucunda elde edilen uzaklık ölçütü cinsiyete göre gruplandırılarak erkek ve bayan sınıflara ilişkin birer ortalama uzaklık değeri elde edilir. Bu değerden küçük olanı test verisinin hangi sınıfa ait olduğunu belirtecektir. Çalışmada uzaklık değerlendirmesi olarak Euclid ölçütü kullanılmıştır. Önerilen sistemin genel yapısı Şekil7'de gösterilmiştir.



Şekil 7. Sistemin genel yapısı

### 5. Test ve Deneysel Sonuçlar

Önerilen sistem üzerinde iki farklı veritabanı kullanılarak test yapılmıştır. İlk olarak 56 bayan 112 erkek 168 kişinin 10'ar farklı cümlesinden oluşan TIMIT veritabanı kullanılmıştır. Sistemin eğitiminde ise test aşamasında kullanılmayan 8 erkek 8 bayanın seslendirdikleri bir cümle kullanılmıştır. TIMIT veritabanı 2 ile 5 sn arasında değişen uzunlukta İngilizce cümlelerden oluşmaktadır. Elde edilen sonuçlar tablo1 de sunulmuştur.

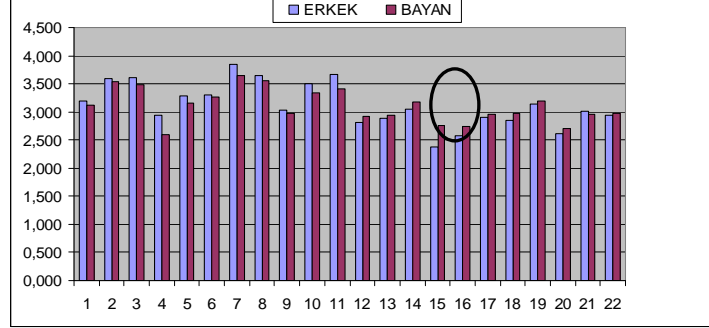
Tablo1. TIMIT veritabanı test sonuçları

<b>Eğitim Kümesi</b>	<b>Kişi sayısı:</b> 16 kişi (8E, 8B)
	<b>Örnek sayısı:</b> 16
<b>Test Kümesi</b>	<b>Kişi sayısı:</b> 168 kişi (56B, 112E)
	<b>Örnek sayısı:</b> 168x10=1680
<b>Kod kitabı boyu</b>	64
<b>Doğru karar sayısı</b>	1646
<b>Hatalı karar sayısı</b>	34
<b>Başarı oranı</b>	%98

Yapılan incelemelerde hatalı karar verilen örnekler için erkek ve bayan sınıflara olan uzaklığın birbirine çok yakın olduğu görülmüştür. Bu örneklerin ikinci bir kıstasa göre değerlendirilerek sonuçların birleştirilmesi başarı oranını arttıracaktır.



İkinci test Boğaziçi üniversitesi tarafından hazırlanan Türkçe bir veritabanı üzerinde yapılmıştır. Bu veritabanı “Dışarıda kar yağıyor, Sınavdan yetmiş aldım, Galatasaray maçı iki sıfır kazandı,..” gibi 11 farklı cümlenin 4 erkek 7 bayan 11 kişi tarafından seslendirilmesiyle oluşturulmuştur. Sistem 11 konuşmacının “Beni çok şaşırttı” cümlesi kullanılarak eğitilmiştir. Test aşamasında ise 11 konuşmacının kalan 10 ar cümlesi kullanılmıştır. Sonuçta 110 örnek arasında yalnızca 1 hatalı cinsiyet kararı verilmiştir. Test sırasında bir erkek ve bir bayan için elde edilen uzaklık değerleri şekil5’de gösterilmiştir.



Şekil 8. Bir bayan bir erkek konuşmacı için uzaklık değerleri

Buradaki ilk 11 örnek bayan konuşmacıya, ikinci 11 örnek ise erkek konuşmacıya aittir. Şekilden de görüldüğü gibi erkek konuşmacının 10. örneği için hatalı karar verilmiştir.

Karşılaştırma sırasında en küçük uzaklığa sahip kayıtlar göz önüne alınarak sistemin konuşmacı tanıma başarısı da incelenmiştir. Sonuçta 110 örneğin 95 inde konuşmacı kimliği doğru olarak saptanmıştır.,

## Gender Identification of the Speaker Using VQ Method

### Extended Abstract

Speaking is the easiest and natural form of communication between people. Intensive studies are made in order to provide this communication via computers between people. The systems using voice biometric technology are attracting attention especially in the angle of cost and usage. When compared with the other biometric systems the application is much more practical. For example by using a microphone placed in the environment voice record can be obtained even without notifying the user and the system can be applied. Moreover the remote access facility is one of the other advantages of voice biometry. In this study, it is aimed to automatically determine the gender of the speaker through the speech waves which include personal information. If the speaker gender can be determined while composing models according to the gender information, the success of voice recognition systems can be increased in an important degree. Generally all the speaker recognition systems are composed of two parts which are feature extraction and matching. Feature extraction is the procedure in which the least information presenting the speech and the speaker is determined through voice signal. There are different features used in voice applications such as LPC, MFCC and PLP. In this study as a feature vector MFCC is used. Feature matching is the procedure in which the features derived from unknown speakers and known speaker group are compared. According to the text used in comparison the system is divided to two parts that are text dependent and text independent. While the same text is used in text dependent systems, different texts are used in independent text systems. Nowadays, DTW and HMM are text dependent, VQ and GMM are text independent matching methods. In this study due to the high success ratio and simple application features VQ approach is used.

In this study a system which determines the speaker gender automatically and text independent is proposed. The proposed system is composed of two levels that are training and testing. In the training level MFCC feature vector is calculated by speaker gender known voice records. MFCC feature vector models the frequency perception of human ear and is one of the most preferred methods. As in all the voice analysis methods, MFCC method is also applied through the short parts which are accepted as having stable voice properties. These parts generally are chosen as 20-30 ms and while moving 10-15 ms shifting amounts they are applied to the whole signal. A window function is applied in order to decrease the discontinuity that are at the edges of derived analysis windows. In voice applications generally hamming window is preferred. Following the windowing procedure the signal is taken to the frequency space by FFT method. The derived FFT spectrum is converted to mel-spectrum by the scale which models human frequency perception and is called as mel-scala. Mel-scala has a linear characteristics up to 1Khz and a logarithmic characteristics over 1 KHz. For converting procedure triangle filters are used of which the

---

band width differs linearly due to the mel-scala. Generally as the filter coefficient a value is chosen between 20 and 30. In the last stage, the logarithm of mel spectrum is taken and we back to time domain. The coefficients derived at the end are called MFCC. The MFCC features derived for each speaker are converted to a smaller vector space by using VQ method. VQ is the transformation to limited numbers of subspaces from a wide vector space. Each subspace is presented with a centre point which is named as code word. Code words constitutes code book. One of the methods which is used to compress N number training vector group to M number (M < N) code book vector is LBG algorithm. This algorithm is realized as following recursive procedure:

1. Design a 1-vector codebook; this is the centroid of the entire set of training vectors
2. Double the size of the codebook by splitting each current codebook  $y_n$  according to the rule

$$\mathbf{y}_n^+ = \mathbf{y}_n(1 + \varepsilon)$$

$$\mathbf{y}_n^- = \mathbf{y}_n(1 - \varepsilon)$$

where  $n$  varies from 1 to the current size of the codebook, and  $\varepsilon$  is a splitting parameter ( $\varepsilon \approx 0.01-0.05$ )

3. Nearest-Neighbor Search: for each training vector, find the codeword in the current codebook that is closest, and assign that vector to the corresponding cell
4. Centroid Update: update the codeword in each cell using the centroid of the training vectors assigned to that cell.
5. Iteration 1: repeat steps 3 and 4 until the average distance falls below a preset threshold
6. Iteration 2: repeat steps 2, 3 and 4 until a codebook size of  $M$  is designed.

At the end of VQ method a code book is composed for each speaker and the training level is completed. In the testing level, the voice records of the speaker whose gender is tried to be determined are used as entries. As in the training level for each test entry MFCC feature vector and VQ code book is calculated. The code books which are from the training level and test level are compared and the speaker gender is determined. During the comparison euclid distance metrics are used and for men and women classes an average distance value is calculated. The smaller one of these two average distance values indicates the class of the test data. The proposed system is tested by TIMIT database and the results shown in Table 1 are derived.

**Table1.** Test results

<b>Training set:</b>	<b>Number of Persons:</b> 16 kişi (8M , 8F)
----------------------	---

---

	<b>Sample number:</b> 16
	<b>Number of Persons:</b> 168 kişi (56F, 112M)
<b>Testing set:</b>	<b>Sample number:</b> 168x10=1680
<b>Codebook size:</b>	64
<b>Correct decision:</b>	1646
<b>Incorrect decision:</b>	34
<b>Success rate:</b>	%98

---

The system is trained with 16 records in which 8 male and 8 female speaks the same sentence. In the testing level 10 different sentences which are spoken by 56 female and 112 male are used. In the total of 1680 test data only 34 incorrect decisions are made and 98% success is achieved.

#### Kaynaklar/References

- Campbell, J. P. (1997). Speaker recognition: a tutorial, *Proceedings of the IEEE*, 85(9), 1437-1462.
- Deller, J. R., Hansen, J. H. L., & Proakis, J. G. (2000). *Discrete-time processing of speech signals*, IEEE Press, Piscataway, N.J.
- Doddington G. R. (1985). Speaker recognition-identification people by their voices, *Proceedings of the IEEE*, 73(11), 1651-1664.
- Eskidere, Ö. & Ertaş, F. (2007). Perde frekansının konuşmacı tanıma etkisi, *15<sup>th</sup> Signal Processing and Communication Applications Conference (IEEE SİU)*, Anadolu University, Eskişehir (pp. 11-13)
- Harb, H. & Chen, L. (2003). Gender identification using a general audio classifier, *IEEE International Conference on Multimedia and Expo*, Baltimore, Maryland, pp.733-736.
- Linde, Y., Bum, A., & Gray, R. M. (1980). An algorithm for vector quantizer design, *IEEE Transactions on Communications*, 28(1), 84-95.
- Matsui, T. & Funii S. (1992). Comparison of text-dependent speaker recognition methods using VQ-distortion and discrete/continuous HMMs, *IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP'92)*, Atlanta, Georgia, pp. 157 -160.
- Nabiyev, V. V., & Yücesoy, E. (2008). Konuşmacı cinsiyetinin temel frekansa göre belirlenmesi, *Çankaya Üniversitesi 1. Mühendislik ve Teknoloji Sempozyumu*, Çankaya Üniversitesi, Ankara, pp. 33-41.
- Naik, J. M. (1990). Speaker verification: a tutorial, *IEEE Communications Magazine*, 28(1), 42-48.
-

- 
- Parris, E. S., & Carey, M. J. (1996). Language independent gender identification, *IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, Atlanta, Georgia, pp. 685-688.
- Picone, J. (1993). Signal modeling techniques in speech recognition, *Proceedings of the IEEE*, 81(9), 1215-1247.
- Rabiner, L., & Juang, B.H. (1993). *Fundamentals of Speech Recognition*, Englewood Cliffs (N.J.), Prentice Hall Signal Processing Series.
- Reynolds D., & Rose R. (1995). Robust text-independent speaker identification using Gaussian mixture speaker models, *IEEE Transactions on Speech and Audio Processing*, 3(1),72-83.
- Rosenberg A. E., & Soong F. K. (1987). Evaluation of a vector quantization talker recognition system in text independent and text dependent modes, *IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP'86)*, Murray Hill, NJ, pp.143-157.
- Silverman, H. F., & Morgan, D. P. (1990). The application of dynamic programming to connected speech recognition, *IEEE ASSP Magazine*, pp. 7-25.
- Ting, H., Yingchun, Y., & Zhaohui, W. (2006). Combining MFCC and Pitch to Enhance the Performance of the Gender Recognition, *IEEE 8<sup>th</sup> International Conference on Signal Processing*, pp. 16-20.
- Vergin, R., Farhat A., & O'Shaughnessy D. (1996). Robust gender-dependent acoustic-phonetic modelling in continuous speech recognition based on a new automatic male/female classification, *Proceedings of the International Conference on Spoken Language Processing (ICSLP)*, Philadelphia, pp. 1081-1084.
-