

# KONUŞMA TANIMA ALGORİTMALARININ CBS'DE KULLANIMI VE DSA TABANLI BİR KONUŞMA TANIMA UYGULAMASI

Pınar ÇİVİCİOĞLU<sup>1</sup>, A. Emin KARKINLI<sup>2</sup>, Erkan BEŞDOK<sup>3</sup>

<sup>1</sup>Doç. Dr., Erciyes Üniversitesi, Sivil Havacılık Yüksekokulu, Uçak Elektrik-Elektronik Bölümü, [civici@erciyes.edu.tr](mailto:civici@erciyes.edu.tr)

<sup>2</sup>Y. Müh., Arş. Gör., Erciyes Üniversitesi, Mühendislik Fakültesi, Harita Mühendisliği Bölümü, [ekarkinli@erciyes.edu.tr](mailto:ekarkinli@erciyes.edu.tr)

<sup>3</sup>Pof. Dr., Erciyes Üniversitesi, Mühendislik Fakültesi, Harita Mühendisliği Bölümü, [ebesdok@erciyes.edu.tr](mailto:ebesdok@erciyes.edu.tr)

## ÖZET

Ses veri işleme teknolojilerinde sağlanan gelişmelerin bir sonucu olarak, CBS uygulamalarında konuşma tanımının sağlayabileceği faydaları araştırmak gereği ortaya çıkmıştır. Konuşma tanıma konusunda sağlanan başarıların oldukça yeni olması nedeniyle konuşma tanımının CBS'de kullanımı konusunda gerçekleştirilen çalışmalar çok sınırlı sayıdadır. Bu bildiriye, Saklı Markov Modeller (HMM), Dinamik Zaman Bükme (DTW) ve Gaussian Karışım Modellerine (GMM) dayalı olarak, ön-tanımlı bir veri sözlüğünde yer alan ayırık konuşmaların tanıma başarıları incelenmiştir. Elde edilen sonuçlar konuşma tanıma uygulamalarının CBS amaçlı veri toplamada kullanılabileceğini göstermektedir.

**Anahtar Sözcükler:** Konuşma Tanıma, CBS, Diferansiyel Arama Algoritması.

## ABSTRACT

### USING SPEECH RECOGNITION ALGORITHMS IN GIS AND AN APPLICATION OF DSA ON SPEECH RECOGNITION.

As a result of the improvements in Audio/Video processing techniques, investigation of potential application areas of speech recognition and geovideo in GIS applications have become more important. Since speech recognition technology has not completed its improvement, there is a limited number of studies achieved about the use of related technologies for GIS applications. In this paper, isolated speech recognition success of hidden markov models, dynamic time warping, gaussian mixture models and DSA-based artificial neural networks have been investigated by using a predefined codebook on GIS. A geovideo application, which may be a useful tool for data collection in several GIS applications, has also been presented. Experimental results exposed that DSA based speech recognition applications rise the success of the neural networks.

**Keywords:** Multimedia GIS, Speech Recognition, Differential Search Algorithm.

## 1. GİRİŞ

Sesin multimedia-CBS uygulamalarındaki önemi özellikle son yıllarda kavranmaya başlanmıştır. Birçok multimedia CBS kullanıcısı grafik içermeyen veritabanı verilerini okumak yerine dinlemeyi tercih etmektedir. Bu insan-bilgisayar etkileşiminde daha doğal bir iletişim yöntemidir. Modern CBS uygulamalarında grafik veritabanında bulunan konumsal bir nesneye görsel olarak erişebilen her kullanıcı ilgili nesneye ilişkin özelliklere bu nesneye ilişkilendirilmiş ses-komutlarını kullanarak da ulaşabilmelidir. Var olan büyük potansiyelden dolayı konumsal nesnelere hakkında 'konuşma tanıma tabanlı veri toplama' ve mevcut grafik içermeyen veriyi kullanıcıya sesli iletebilmek için gerekli 'doğal ses sentezleme' uygulamaları modern multimedia-CBS geliştirme çalışmalarında çok önemli araştırma alanları haline gelmiştir.

Multimedia-CBS uygulamalarında ses-tabanlı yöntemlerin kullanımı, CBS veritabanından kullanıcıya doğru gerçekleşen tek-modlu veri akışının çok-modlu, interaktif ve daha doğal bir şekilde sağlanması için kullanılabilecek etkin araçlar geliştirilmesine olanak vermektedir. Modern MEMS teknolojilerinin ve Giyilebilir-Hesaplayıcıların ortaya çıkmasıyla birlikte insan ve bilgisayar etkileşiminde ses gibi daha doğal araçların kullanımı yaygınlaşmaya başlanmıştır. Konuşma tanıma insan ve bilgisayarın etkileşiminde kullanılabilecek en doğal araçlardandır. Böylece CBS verilerine yapay olarak sentezlenmiş ses verileriyle erişilebilir veya CBS veritabanına konuşma tanıma ile veri sağlanabilir.

Konuşma sentezleme CBS uygulamalarında özellikle görme kusuru bulunan kişilere, çocuklara, yaşlılara veya stres-altında çalışan ilgili yönetici personele aradıkları bilgiyi CBS veri havuzundan ayıklayarak interaktif bir şekilde ulaştırmak için kullanılabilir. Görme kusuru olan kişilerin navigasyon ihtiyaçlarının karşılanmasında ve birçok modern navigasyon cihazında konuşma-sentezleme teknolojisi oldukça yararlı olacaktır.

Çok-modlu interaktif bir navigasyon sisteminde kullanıcı gitmek istediği yeri veya bulunduğu konuma yakın çevrede yer alan özellikli-tesisleri navigasyon sistemine sesli olarak iletir. Ardından ilgili sistem kullanıcıya gerekli verileri genellikle bir grafiğe eşlik eden sentezlenmiş bir sesle sunar. Konuşma tanıma motoru kullanan bir CBS kullanıcısı, ilgili sistemi sesli olarak ileticeği isim, adres, telefon gibi baskın bir öncül tanımlayıcı kullanarak sorgulayabilir. Bu yöntem insanlar arasındaki iletişim modeline çok yakın doğal bir çözüm olduğundan CBS ve kullanıcı arasındaki etkileşimi de kolaylaştırır.

Konuşma tanıma küresel konum bilgisi ile zaman domeninde senkronize edilebilirse ellerin ve gözlerin yoğun çalıştığı CBS için sahada veri toplama uygulamalarında etkin bir veri toplama yöntemi geliştirilebilir. Bu bildiride yaygın olarak kullanılan konuşma tanıma algoritmalarının izole edilmiş (ayrık) konuşma tanıma yetenekleri incelenmiş ayrıca çoklu-paralel yapay sinir ağların kullanılarak bir konuşma tanıma yöntemi geliştirilmiştir. Geliştirilen yeni yöntem, ilgili ağların eğitiminden sonra elde edilen çözümlerin  $\pm\%10$  saçılmasıyla elde edilen arama uzayında sub-optimal çözümün Differansiyel Arama Algoritması (DSA) kullanılarak geliştirilmesine dayanmaktadır.

Bu bildirinin geri kalan bölümleri şu şekilde düzenlenmiştir. Bölüm 2’de DSA tanıtılmıştır, Bölüm 3’te Konuşma Tanıma yöntemleri kısaca incelenmiştir. Bölüm 4’te Uygulamalar, Bölüm 5’te ise Sonuçlar verilmiştir.

---

**Algorithm 1: Pseudo-Code of Differential Search Algorithm**


---

```

Input: ObjFun, SizeOfColony, dim, P1, P2, low, up, and maxcycle
Output: globalminimum, globalminimizer
// please see Matlab for 'ones', 'rand', 'randg', 'randi', 'feval' and 'randperm'
1 function algo_ds(SizeOfColony,dim,P1,P2,low,up,maxcycle)
  // INITIALIZATION
2 globalminimum=inf
3 if numel(low)==1 then
4   low=low*ones(1,dim)
5   up=up*ones(1,dim)
6 end
7 superorganism=genpop(SizeOfColony,dim,low,up)
8 y=ObjFun(superorganism)
9 for epk form 1 to maxcycle do
  // SELECTION-I
10  donor=superorganism(randperm(SizeOfColony),:)
11  scale=randg(2*rand)*(rand-rand)
  // MUTATION (Morphogenesis)
12  stopover=superorganism+scale.*(donor-superorganism)
  // P1=P2+'3*rand' has been used in DSA
13  p1=eval(P1)
14  p2=eval(P2)
15  r=rand(SizeOfColony,dim)
16  if rand<rand then
17    if rand<p1 then
18      for i form 1 to SizeOfColony do
19        r(i,:)=r(i,:) < rand
20      end
21    else
22      for i form 1 to SizeOfColony do
23        r(i,randi(dim))=0
24      end
25    end
26  else
27    for i form 1 to SizeOfColony do
28      d=randi(dim,1,[p2*rand])
29      for j form 1 to size(d,2) do
30        r(i,d(j))=0
31      end
32    end
33  end
  r=r>0
  // RECOMBINATION
34  stopover(r)=superorganism(r)
35  stopover=update(stopover,low,up)
36  fit_s=ObjFun(stopover)
37  ind=fit_s<y
  // SELECTION-II
38  if numel(ind)>0 then
39    y(ind)=fit_s(ind)
40    superorganism(ind,:)=stopover(ind,:)
41  end
42  y_indexbest ← min(y) | indexbest ∈ {1, 2, 3, ..., SizeOfColony}
43  if y_indexbest < globalminimum then
44    globalminimum=y_indexbest
45    globalminimizer=superorganism(indexbest,:)
  // Export globalminimum and globalminimizer
46  end
47 end
  // Sub-Functions
48 function pop=genpop(a,b,low,up)
49  pop=ones(a,b)
50  for i form 1 to a do
51    for j form 1 to b do
52      pop(i,j)=rand*(up(j)-low(j))+low(j)
53    end
54  end
55 end
56 function p=update(p,low,up)
57  [SizeOfColony,dim]=size(p)
58  for i form 1 to SizeOfColony do
59    for j form 1 to dim do
60      if p(i,j)<low(j) then
61        p(i,j)=rand*(up(j)-low(j))+low(j)
62      end
63      if p(i,j)>up(j) then
64        p(i,j)=rand*(up(j)-low(j))+low(j)
65      end
66    end
67  end
68 end
69 end

```

---

**Şekil 1** : DSA algoritmasının sembolik-kodu (Çivicioğlu, P., 2012).

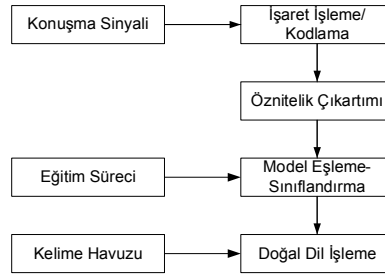
## 2. DİFFERANSİYEL ARAMA ALGORİTMASI (DSA)

Differansiyel Arama Algoritması (DSA) Çivicioğlu, P. Tarafından sayısal optimizasyon problemlerinin çözümü için geliştirilmiş sürü tabanlı bir diferansiyel gelişim algoritmasıdır. DSA'nın problem çözme başarısı çok gelişmiş Differansiyel Gelişim Algoritmalarından (JDE, JADE, SADE, EPSDE gibi) ve CMAES, GSA, ABC, PSO2011 gibi yaygın olarak kullanılan diğer sayısal optimizasyon algoritmalarından daha yüksektir (Çivicioğlu, P., 2012). DSA'nın çok az kontrol parametresi vardır, ayrıca DSA'nın problem çözme başarısı genellikle ilgili kontrol parametrelerinin başlangıç değerlerine aşırı duyarlı değildir. DSA, birçok evrimsel algoritma gibi sabit bir popülasyon büyüklüğü ile çalışır ve popülasyon büyüklüğü genel olarak 5-50 arasında seçilebilir. Bu bildiride DSA, geriyayılım tabanlı çok katmanlı bir yapay sinir ağ yapısının elde ettiği ağırlık ve kutuplama değerlerini sub-optimal olarak geliştirmek üzere kullanılmıştır. DSA'nın genel yapısı Şekil 1'de verilmiştir.

## 3. KONUŞMA TANIMA

Konuşma tanıma, insan ve bilgisayar arasında iletişim kurmanın en doğal yollarından birisidir ve sözcük-ses sinyallerinin bilgisayar tarafından anlaşılabilmesini amaçlar. Uzun süredir, konuşma veya duyma güçlüğü çeken insanların sosyal-yaşam ve iletişim konforlarını arttırmak için metinden sese geçiş sağlayan (Text-to-Speech) veya sesi metne dönüştürebilen (Speech-to-Text) sistemler üzerinde araştırmalar yapılmaktadır.

Temel olarak ses, dinleyici ile ses kaynağı arasında bulunan sıvı, gaz ya da katı ortamda oluşan bir dizi titreşimle oluşur. Bir obje titreştiğinde sahip olduğu enerjinin bir bölümünü ses olarak yayar. Ses vakumlu ortamlarda ilerleyemez. Ses havada basınç farklılıkları yaratır; yani duyma organı aslında sadece basınç farklılıklarını algılar. Ses hızı  $20^0$  hava sıcaklığında yaklaşık 344 metre/sn'dir. Bilgisayar aracılığıyla ses tanıma karşılaşılan zorluklar ses tanımanın kompleks doğasından kaynaklanır. Sinyal işleme, akustik, örnek tanıma, haberleşme ve bilgi teorisi, dilbilim, psikoloji ve bilgisayar bilimi ses tanıma içinde kullanılan disiplinlerdir. Başarılı bir ses tanıma sisteminin gerçekleştirilmesi için bu disiplinlere ait kapsamlı bilgilere sahip olunması gerekmektedir. Uzun yıllardır üzerinde çalışılan ses tanıma yöntemlerinin temel amacı ses sinyalinin bilgisayar tarafından anlamlı karşılıklara dönüştürülmesidir. Ses işlemede en büyük hedef bilgisayar kullanıcılarına klavye kullanmadan sadece sesle kullandıkları veritabanlarına veri beslemesi yapabilmelerini sağlamaktır. Veritabanları CBS ve CBS-destek sistemlerinin temel bileşenlerindedir. Ses CBS veritabanlarını beslemek, sorgulamak veya erişim-güvenliğini sağlamak için kullanılabilir. Şekil 2'de bir konuşma tanıma süreci şema üzerinde açıklanmıştır.



Şekil 2: Konuşma tanıma teknik-uygulama süreci.

Uygulamada konuşma tanıma yazılımları kullanılacakları teknik alana göre tasarlanmaktadır. Konuşma tanıma karşılaşılan sorunların çözümü için değişik yöntemler geliştirilmiştir. Konuşma tanıma motorları genellikle aşağıda verilen üç temel sistemden birisine uyarlar;

- Ayırık sözcük tanıma sistemleri (isolated word recognition systems),
- Sözcük yakalama sistemleri (word spotting systems, speech-minning),
- Sürekli konuşma tanıma sistemleri (continous speech recognition system).

Ayırık sözcük tanıma ya da yalıtılmış konuşma tanıma sistemleri literatüründe sunulan konuşma tanıma uygulamalarında en yüksek başarıyı veren sistemlerdir. Bu sistemlerde sözcükler arasında kısa duraklar bulunur. Hedef, sözcük tanımadır; ayırık ses parçacıklarının tanınması amaçlanmaz. Konuşma tanıma çözümü zor olan noktalardan birisi de sözcüklerin başlangıç ve bitiş noktalarının belirlenmesidir. Ayırık sözcük tanıma sözcüklerin birbirinden bağımsız telaffuz edilmesi tanımayı kolaylaştırır.

CBS amaçlı konumsal veri toplama çalışmaları genellikle esnek bir kod-kitabına bağlı kalınarak yapılır. İlgili kod kitabı nesnelere hakkında kullanılacak tanımları içerir; 'kat adedi', 'yol genişliği', 'toprak yol', 'sosyal tesis', 'okul', 'yasak bölge', 'ağaç tipi', 'yapı tipi', 'ruhsat tarihi', 'zaman', 'x konumu', 'zemin tipi' gibi. Esnek bir kod-kitabı kullanmak CBS verilerinin standardizasyonunu kolaylaştırır. Dolayısıyla ilgili CBS nesnelere ilişkin sesli karşılıklarının metin eşlenikleri kod kitabından elde edilebilir.

Sözcük yakalama sistemleri, sürekli-konuşma içinde belirli bir kelimenin varlığını araştırır. Bu tür uygulamalarda en başarılı sonucu veren yöntem 'dinamik zaman esnetme' olarak bilinen yöntemdir. Bu yöntemde sözcük tanıma sistemlerinde tanınacak her bir kelime bir şablon tarafından gösterilir. Kelimeler sadece şablonlar tarafından bilindiğinden dolayı, tanımayı sesin bitiş noktalarından bağımsız kılmak önemlidir. Tanıma işlemi, aranan şablonun sesli ifade içinde çakıştığı bir örüntü arama biçiminde gerçekleşir. Sürekli konuşma tanıma, 'sürekli sesli ifade tanıma sistemleri', 'bağlı sözcük tanıma sistemleri' ve 'karşılıklı konuşma tanıma sistemleri' olmak üzere üç gruba ayrılır. Sürekli ya da ayrık olmalarının dışında konuşma tanıma sistemleri konuşmacıya bağımlılığına göre de ikiye ayrılır:

- Konuşmacıya bağımlı konuşma tanıma sistemleri (Speaker Dependent).
- Konuşmacıdan bağımsız konuşma tanıma sistemleri (Speaker Independent).

Konuşmacı bağımlı konuşma tanıma sistemlerinin konuşma tanımada sağladıkları başarı daha yüksektir. Konuşma tanımada algılanan tüm ses örneklerinin boyutlarının birbirinden farklı olması nedeniyle tüm örneklerin aynı boyutta temsil edilmesi problemiyle karşılaşılır. Ayrıca ses dosyalarının boyutu, sesin işlenebilmesi için olması gerekenden çok daha büyük boyutta olabilir veya ses dosyasının içeriği sözkonusu uygulama için anlamsız birçok veri taşıyor olabilir. Bu nedenle ses dosyalarını standart hale getirebilmek için bir öznitelik çıkartma operatörü kullanılmalıdır. Öznitelik çıkartma işlemi ses dosyasından sözkonusu uygulama amaçları için işe yarayan ve istenilen özelliklerinin ayıklanmasını amaçlar. Öznitelik çıkartma işlemi konuşmacıların konuşma stilleri arasındaki analitik farkı belirtmelidir.

Bitişken bir dil olan Türkçe'de, sözcüklerin sonlarına ard-arda ekler konularak yüzlerce farklı sözcük oluşturulabilir. Türkçe eklentili bir dil yapısına sahip olduğundan metinden-konuşmaya geçişte yaşanan zorluklara benzer teknik güçlükler konuşmadan-metne geçerken de ortaya çıkar. Bu nedenle Türkçe gibi eklentili dillerde konuşma tanıma araştırmaları genellikle bir uzmanlık alanında seçilen sınırlı sayıda terminolojiyi tanıyabilen sistemler üzerinde yoğunlaşmıştır.

Literatürde Konuşma tanıma problemi genellikle 'özellik çıkarımı' ve 'özellik vektörlerinin sınıflandırılması' olmak üzere iki temel başlık altında incelenir. Ses paterninin sınıflandırılabilmesi için konuşmanın özniteliklerinin ayıklanmış olması gerekir. Bu süreç özellik çıkartımı olarak bilinir. Konuşma tanımada yüksek doğruluk elde etmek amaçlanıyor ise uygun öznitelik parametrelerinin elde edilmesine odaklanılmak gerekir. Çok farklı öznitelik çıkartımı algoritmaları bulunmaktadır. En sık kullanılan özellik çıkarım yöntemleri Fourier güç katsayılarının hesaplanması, Mel kepsral katsayıların hesaplanması, sinyal enerji seviyesinin hesaplanması, sıfır geçiş sayısının hesaplanması, tepe değerlerin hesaplanması, tepe değere ulaşma sayısının belirlenmesi, öz-benzerlik katsayılarının hesaplanması ve doğrusal öngörülü kodlamadır.

Konuşma tanıma problemi özetle bir patern tanıma problemidir. Konuşma tanıma verilen bir kelime-egitim veritabanına ait özellik vektörlerini kullanarak algılanan bir kelime-test ses verisinin hangi kelime olduğunu bulmaya dayanır. Konuşma tanımada özellik vektörlerinin sınıflandırılmasında Gizli Markov Modeli, Dinamik Zaman Bükmesi ve Yapay Sinir Ağları yöntemlerinin kullanımı daha çok tercih edilmektedir. Konuşma tanıma uygulamalarında Mel ölçekli kepsrum katsayıları (MFCC) günümüzde en çok kullanılan ses özneliğidir. MFCC, doğrusal öngörü analizi (LPC)'ye gerek kalmadan hesaplanabilmektedir. MFCC LPC'nin aksine insan kulağının duyma yönteminden esinlenmiştir. MFCC bir dizi süzgeç kullanılarak elde edildiğinden dolayı gürültülü seslerde LPC katsayılarına göre daha iyi tanıma performansı sağlar. İlgili bilimsel literatüre göre, MFCC hem konuşma tanıma hem de konuşmacı tanıma uygulamalarında başarılı sonuçlar vermektedir. MFCC, insan kulağının kritik işitme bant genişliği ve frekansındaki değişimlerin, düşük frekansta doğrusal süzgeçler kullanılarak ve yüksek frekanslarda da logaritmik süzgeçler kullanılarak modellenmesini amaçlar. Böylece önemli konuşma öznelikleri ayıklanabilir. Bahsedilen eşik-ölçekleme süreci, mel frekansı ölçeği olarak bilinmektedir; 1000 Hz ve altı sesler doğrusal frekans bölgesi, 1000 Hz ve üstü seslerse logaritmik frekans bölgesi olarak kabul edilir.

MFCC spektrum katsayıları kayan noktalı sayılar olduğundan konuşma spektrumunun kepsral gösterimi, işaretin pencereleme analizi ile verilen lokal spektral özelliklerini çok iyi ifade edebilmektedir.

MFCC hesaplanırken öncelikle, sürekli konuşma işareti, N örnekten oluşan çerçevelere ayrılır. Ardından ses sinyaline sayısal işaret işleme uygulamalarında standart bir yöntem olarak kullanılan pencereleme işlemi uygulanır. Pencereleme işleminin amacı işarettaki süresizlikleri minimum hale getirmektir. Bu işlemin ardından hızlı Fourier dönüşümü kullanılarak N örnekten oluşan zaman domenindeki işaret, frekans domenine dönüştürülür. Frekans domenine dönüştürülen işarete Mel frekans içeriğine göre tasarlanmış filtre dizisi uygulanarak sonucun logaritması

alınır. Ardından da, logaritmik mel spektrumundan ayrık kosinüs dönüşümü kullanılarak zaman domenine geçilir. Böylece mel frekansı keppstrum katsayıları elde edilmiş olur.

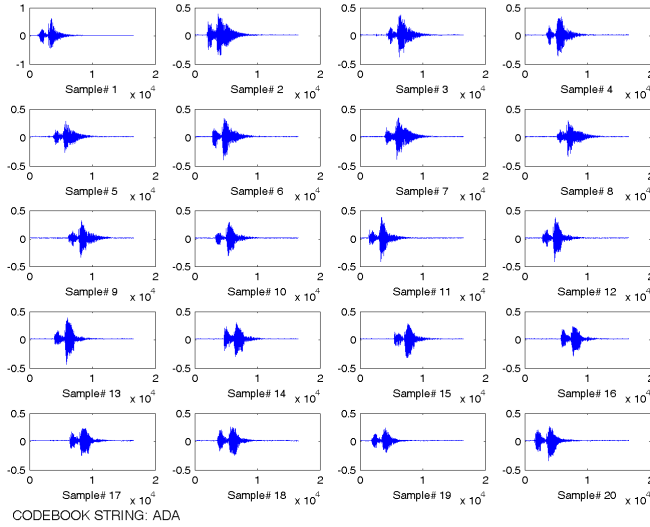
Konuşma tanımda önışlemler sırasında belirli boyutlarda pencereler halinde düzenlenmiş zaman boyutundaki ses verisi spektral analiz ile frekans özellikleri ortaya konacak şekilde frekans boyutuna dönüştürülür. Frekans domeninde elde edilen genlik değerleri, kullanılan pencereleme fonksiyonu boyutunda elde edilir. Özellik vektörünün bu boyutu ile kullanılması yerine belirli bantlara düşen ortalama genlik değerleri alınarak hem özellik vektörü boyutu azaltılır hem de belirli bir filtreleme sağlandığından işaret geliştirilmiş olur.

Bu bildiriye gerçekleştirilen çalışmalarda ses verisinin özneliklerini elde etmek için MFCC katsayıları kullanılmıştır. Kullanılan çerçeveoranı 1/ f olduğundan tüm MFCC vektörleri sabit uzunlukta elde edilmiştir. Gerçekleştirilen deneylerde kullanılan MFCC vektörlerinin uzunluğu 13 ve 32 arasında değişmektedir. Bu bildiriye konuşma verilerinden sesli ve sessiz bölgelerin ayrıştırılmasında Eşitlik 1' de verilen Kısa-Sürekli Enerji (Short-time Energy) değeri kullanılmıştır.

$$E_n = \sum_{m=-\infty}^{\infty} (x[m]w[\hat{n}-m])^2 = \sum_{m=-\infty}^{\infty} x^2[m]w^2[\hat{n}-m] \quad (1)$$

Bu bildirinin hazırlaması için kullanılan kod-kitabı 400 kelimedenden oluşmaktadır ve içeriği Çizelge 1'de verilmiştir.

Derlenen verilerden bir örnek Şekil 3'te görselleştirilmiştir.



Şekil 3 : Derlenen orijinal ses verilerinin grafik olarak görselleştirilmesi.

Çizelge 1 : 400 kelime içeren kod kitabı.

|                    |                 |                 |                  |                |                 |                |
|--------------------|-----------------|-----------------|------------------|----------------|-----------------|----------------|
| 1: ADA             | 61: BÖLÜNÜŞ     | 121: EKSKAVATÖR | 181: İŞ          | 241: MALİYET   | 301: PLENT      | 361: TEHLİKELİ |
| 2: AÇIK            | 62: BRÜKNER     | 122: EKSPRES    | 182: İŞARET      | 242: MALZEME   | 302: POLİMER    | 362: TEKERLEK  |
| 3: AFUYMAN         | 63: BUZ         | 123: ELEVASYON  | 183: İŞARETİLEME | 243: MANİVELA  | 303: PÜSKÜRTME  | 363: TEKERLEK  |
| 4: AGREGA          | 64: BÜZ         | 124: EMLAK      | 184: İŞÇİ        | 244: MANSAP    | 304: RADYE      | 364: TEKNİK    |
| 5: AĞIR TAŞIT      | 65: CAM         | 125: EMNİYET    | 185: İŞGAL       | 245: MARŞ      | 305: RAKİM      | 365: TEKYON    |
| 6: AĞIRLIK         | 66: ÇAKIL       | 126: EMNİYETSİZ | 186: İŞLETME     | 246: MEKİK     | 306: RAKORTMAN  | 366: TEMEL     |
| 7: AKARYAKIT       | 67: ÇAMUR       | 127: EMÜLSİYON  | 187: İZ          | 247: MEMBA     | 307: RAMPA      | 367: TERSYOL   |
| 8: AKILLI          | 68: ÇARPMA      | 128: ENKESİT    | 188: İZİN        | 248: MENFEZ    | 308: RAPOR      | 368: TIR       |
| 9: ALAN            | 69: ÇED         | 129: ESNEK      | 189: JEODEZİ     | 249: MERMER    | 309: RASTER     | 369: TIRMANMA  |
| 10: ALIŞTIRMA      | 70: ÇEKİCİ      | 130: ETMEK      | 190: KADEME      | 250: MESAFESİ  | 310: REFÜJ      | 370: TON       |
| 11: ALIYNMAN       | 71: ÇELİK       | 131: ETRİYE     | 191: KALEM       | 251: MESLEKİ   | 311: REĞLAJ     | 371: TRAFİK    |
| 12: ALTGEÇİT       | 72: ÇERÇEVE     | 132: ETÜT       | 192: KALIP       | 252: MESNET    | 312: RIJİT      | 372: TRANSİT   |
| 13: ALTTEMEL       | 73: ÇEVRE       | 133: FAUNA      | 193: KALİTE      | 253: METRE     | 313: RODMIKS    | 373: TRANVAY   |
| 14: AMPATMAN       | 74: ÇİĞ         | 134: FERAG      | 194: KAMA        | 254: MICIR     | 314: RÖGAR      | 374: TREN      |
| 15: ANAKIRIŞ       | 75: ÇİZGİ       | 135: FEZEYAN    | 195: KAMYON      | 255: MİKTAR    | 315: RÖMORK     | 375: TÜNEL     |
| 16: APLİKASYON     | 76: ÇİZGİSEL    | 136: GABARİ     | 196: KAMYONET    | 256: MİNİBÜS   | 316: RÖPER      | 376: ULAŞIM    |
| 17: ARABA          | 77: ÇOK         | 137: GAGA       | 197: KANTAR      | 257: MOLOZ     | 317: RÜZGAR     | 377: UYARI     |
| 18: ARAÇ           | 78: ÇÖKME       | 138: GAR        | 198: KAPALI      | 258: MOTOR     | 318: SADEME     | 378: VEKTÖR    |
| 19: ARAZÖZ         | 79: ÇÖRTEN      | 139: GARAJ      | 199: KAPLAMA     | 259: MOTORİN   | 319: SAĞ        | 379: VİBRASYON |
| 20: ARIYAT MALZEME | 80: ÇÜRÜK       | 140: GECE       | 200: KAR         | 260: MOZETA    | 320: SARKMA     | 380: VIDA      |
| 21: ASFALT         | 81: ÇÜRÜTME     | 141: GEÇ        | 201: KARAYOLU    | 261: MUAYENE   | 321: SEHİM      | 381: YAĞIŞ     |
| 22: ASMA           | 82: DAĞ         | 142: GEÇİCİ     | 202: KASA        | 262: MURÇ      | 322: SERVİS     | 382: YAĞMUR    |
| 23: AŞTAR          | 83: DAİRESEL    | 143: GEÇİŞ      | 203: KATKI       | 263: MÜCADELE  | 323: SEVİYE     | 383: YANDAN    |
| 24: AŞINDIRICI     | 84: DAMPERLİ    | 144: GEÇİT      | 204: KAŞAK       | 264: MÜCAVİR   | 324: SEVK       | 384: YARALI    |
| 25: AŞINMA         | 85: DEBİ        | 145: GERGİ      | 205: KAYGAN      | 265: MÜHENDİS  | 325: SEYYAR     | 385: YARMA     |
| 26: AŞINMIŞ        | 86: DEBRÜAJ     | 146: GİDO       | 206: KAZA        | 266: MÜLKİYET  | 326: SIKIŞIKLIK | 386: YASAK     |
| 27: AYIRICI        | 87: DEBUŞE      | 147: GİRİŞİ     | 207: KAZI        | 267: MÜTEAHHİT | 327: SINIR      | 387: YASTIK    |
| 28: AYIRMA ŞERİDİ  | 88: DEFORMASYON | 148: GÖRÜŞ      | 208: KAZICI      | 268: NAKLİYE   | 328: SINIRI     | 388: YAVAŞLA   |
| 29: AZ             | 89: DELİCİ      | 149: HIZ SINIRI | 209: KAZIK       | 269: NOKTA     | 329: SİVİLİŞMA  | 389: YAYA      |
| 30: AZAMI          | 90: DEMİRYOLU   | 150: GRADASYON  | 210: KAZIMA      | 270: OKUL      | 330: SİLİNDİR   | 390: YAZ       |

|               |                  |                |                |                 |                    |                 |
|---------------|------------------|----------------|----------------|-----------------|--------------------|-----------------|
| 31: BAĞLANTI  | 91: DEPO         | 151: GREYDER   | 211: KAZIYICI  | 271: OLUK       | 331: SINYALİZASYON | 391: YERİ       |
| 32: BAKIM     | 92: DEPREM       | 152: GROBETON  | 212: KEMER     | 272: ONDULASYON | 332: SIYAH         | 392: YERLEŞİM   |
| 33: BANKET    | 93: DERİVASYON   | 153: GUSE      | 213: KENAR     | 273: ORTA       | 333: SOL           | 393: YETERLİLİK |
| 34: BARBAKAN  | 94: DERZ         | 154: GÜÇLÜ     | 214: KENT      | 274: ORTALAMA   | 334: SON           | 394: YOK        |
| 35: BASIT     | 95: DEVAMLİ      | 155: GÜNDÜZ    | 215: KESİKLİ   | 275: OTO        | 335: SONDAJ        | 395: YOL        |
| 36: BAŞLANGIÇ | 96: DEVER        | 156: GÜVENLİ   | 216: KESİM     | 276: OTOBÜS     | 336: SÖMEL         | 396: YÖNETİCİ   |
| 37: BELGESİ   | 97: DIKKAT       | 157: GÜVENSİZ  | 217: KESKİ     | 277: OTOMOBİL   | 337: SİRAL         | 397: YÜK        |
| 38: BENZİN    | 98: DİLASYON     | 158: HACİM     | 218: KESON     | 278: OTOYOL     | 338: STABİLİTE     | 398: YÜZEY      |
| 39: BETON     | 99: DİNGİL       | 159: HAFİF     | 219: KILOGRAM  | 279: OYUK       | 339: STABİLİZE     | 399: ZAYIF      |
| 40: BEYAN     | 100: DİREKSİYON  | 160: HARC      | 220: KILOMETRE | 280: ÖLÇME      | 340: STANDART      | 400: ZORUNLU    |
| 41: BEZEME    | 101: DİSTRİBÜTÖR | 161: HAREKET   | 221: KIRMIZI   | 281: ÖN         | 341: SU            |                 |
| 42: BIÇAK     | 102: DOLDURMA    | 162: HAREKETLİ | 222: KIŞ       | 282: ÖRME       | 342: SÜRTÜNME      |                 |
| 43: BİLET     | 103: DOLGU       | 163: HARİTA    | 223: KİLİTLİ   | 283: PAFTA      | 343: ŞANTİYE       |                 |
| 44: BİLGİ     | 104: DOLU        | 164: HAVA      | 224: KİRİŞ     | 284: PALET      | 344: ŞARTNAME      |                 |
| 45: BİNDER    | 105: DOZER       | 165: HEMZEMİN  | 224: AĞAÇ      | 285: PALYE      | 345: ŞAŞI          |                 |
| 46: BİNEK     | 106: DÖNEL       | 166: HENDEK    | 226: KOMPLEKS  | 286: PARAFİN    | 346: ŞEF           |                 |
| 47: BİRİM     | 107: DÖNÜŞ       | 167: HEYELAN   | 227: KONSOL    | 287: PARALEL    | 347: ŞERİT         |                 |
| 48: BİSİKLET  | 108: DÖŞEME      | 168: HIZ       | 228: KONTROL   | 288: PARK       | 348: ŞEV           |                 |
| 49: BİTİŞ     | 109: DRENAJ      | 169: HIZLAN    | 229: KORİDOR   | 289: PARLAMA    | 349: ŞUBE          |                 |
| 50: BİTÜM     | 110: DUR         | 170: İDARE     | 230: KORKULUK  | 290: PARSEL     | 350: TABAKA        |                 |
| 51: BOMBE     | 111: DURAK       | 171: İFRAZ     | 231: KORNIŞ    | 291: PAS        | 351: TAHSİS        |                 |
| 52: BORDÜR    | 112: DURAKLAMA   | 172: İHALE     | 232: KOROZYON  | 292: PATLAYICI  | 352: TAKOGRAF      |                 |
| 53: BORU      | 113: DUVAR       | 173: İKTİSAP   | 233: KÖT       | 293: PERE       | 353: TALİ          |                 |
| 54: BOSAJ     | 114: DÜŞEY       | 174: İLERİ     | 234: KÖPRÜ     | 294: PERİYODİK  | 354: TALVEG        |                 |
| 55: BOŞ       | 115: DÜZELTME    | 175: İMAR      | 235: KÖŞE      | 295: PEYZAJ     | 355: TANDEM        |                 |
| 56: BOŞALTMA  | 116: DÜZEY       | 176: İNSAAT    | 236: KUM       | 296: PLAK       | 356: TANKER        |                 |
| 57: BOYA      | 117: EĞİM        | 177: İNŞAAT    | 237: KURP      | 297: PLAKA      | 357: TAŞ           |                 |
| 58: BOYKESİT  | 118: EĞİTİM      | 178: İSKELE    | 238: KUSUR     | 298: PLAN       | 358: TAŞI          |                 |
| 59: BOYUNA    | 119: EHLİYET     | 179: İSTASYON  | 239: MAKİNA    | 299: PLANKOTE   | 359: TAŞIT         |                 |
| 60: BOZUK     | 120: EKİP        | 180: İSTİKŞAF  | 240: MALÇ      | 300: PLATFORM   | 360: TEHLİKE       |                 |

Saklı Markov modelleri (HMM) konuşma tanıma sık kullanılır (Huang, X. D., ve Jack, M.A., 1988, Huang, X. D., Jack, M.A., et al., 1988). Saklı Markov modelinde durumlar markov modellerde olduğunun aksine doğrudan gözlenemez (Huo, Q., ve Lee, C.H., 1998). Gözlem dizilerine çeşitli en-iyileme teknikleri uygulanarak en yüksek olasılıklı durum dizisi elde edilir. SMM'lerin konuşma tanıma yaygın olarak kullanılması 1980'lerden sonra yaygınlaşmıştır. Konuşma işaretinin istatistiksel özellikleri zamanla değişir; dolayısıyla konuşma işareti ergodik değildir ve uzun gözlem süreleri için durağan olmayan karakterdedir. Buna karşın konuşma verilerinin 10ms-30ms için durağan karakterde oldukları varsayılır ve zaman-ekseninde kısa süreli ses verilerinden değişmez öznelikler elde edilebilir. Genellikle ön görülen kısa konuşma süreleri zaman-ekseninde önceden öngörülen uzunluklarda örtüşen pencereler kullanılarak sağlanır. Kod kitabında bulunan 400 kelime, farklı zamanlarda 20'şer defa tekrarlanarak derlendiğinden toplam  $400 \times 20 = 8000$  konuşma örneği bulunmaktadır.

HMM tabanlı konuşma tanıma motorunun alıştırılması (eğitim anlamında) için 400 kelimeye ait kayıtların ilk 15'i kullanılmıştır. Her kelime için kalan 5 ses örnekleme HMM tabanlı konuşma tanıma motorunun testi için kullanılmıştır. HMM kullanılarak gerçekleştirilen konuşma tanıma testlerinde kod-kitabında bulunan 400 kelimeye ait 5'er tekrardan oluşan 2000 test verisinden 1974 tanesi doğru tanınarak %98.70 başarı sağlanmıştır. Bu başarı oranı ilgili bilimsel literatüre göre kullanılan kelime ve model sayısının büyüklüğü göz önüne alındığında oldukça yüksektir.

Dinamik zaman bükmesi (DTW) algoritması (Irwin, M.J., 1988, Issa, K. ve Nagahashi, H., 2008, Jang, M., Han, M.S., ve değerleri, 2011) Kruskal ve Liberman tarafından (Kruskal, J.B., ve Liberman, M. 1983) 1983 yılında iki eğri arasındaki mesafenin belirlenebilmesi için kullanılabilecek yeni bir yöntem olarak geliştirilmiştir. DTW iki eğri arasında ortak patternleri belirleyebilme yeteneğine sahiptir. DTW farklı telefüzlarla seslendirilen aynı kelimeyi ilgili ses verilerini kullanarak kolayca eşleyebilir. DTW iki ses verisini karşılaştırırken ilgili ses sinyallerini zaman domeninde genişletir veya daraltır. DTW ile gerçekleştirilen testlerde kullanılan 400 kelimeye ait 5'er tekrardan oluşan 2000 test verisinden 1961 tanesi doğru tanınarak %98.05 başarı sağlanmıştır. Bu başarı oranı ilgili bilimsel literatüre göre kullanılan kelime ve model sayısının büyüklüğü göz önüne alındığında oldukça yüksektir.

Gaussian Mixture Modeller (GMM) (Martin, R., C. Hoelper, et al. 2001) konuşma tanıma uygulamalarında yaygın olarak kullanılan bir diğer algoritma grubudur. GMM kullanılarak gerçekleştirilen testlerde bir test örneğine karşılık önceden hesaplanmış toplam  $200 \times 15 = 3000$  GMM modeli kullanılmıştır. GMM model parametreleri 200 epok için ve 12 Gaussian model için çözülmüştür. Test sırasında bir test kelimesinin tanınması için gerekli süre eğitim için kullanılan epok sayısından ve Gaussian model sayısından doğrudan etkilenir. Bu bölümde gerçekleştirilen testlerde bir tanıma süreci Matlab platformunda normal dağılıma göre %95 olasılıkla [11.945 18.309] sn arasında değişmektedir. GMM ile gerçekleştirilen testlerde kullanılan 200 kelimeye ait 5'er tekrardan oluşan 1000 test verisinden 821 tanesi doğru tanınmıştır; dolayısıyla konuşma tanıma başarısı %82.10'dur. Bu başarı oranı ilgili bilimsel literatüre göre kullanılan kelime ve model sayısının büyüklüğü göz önüne alındığında kabul edilebilir düzeydedir. Kelimelerin %50.50'si %100 başarı ile tanınmıştır. %23.50'si %80, %15.50'si %60, %8'i %0.40, %1.50'si %20 doğrulukla tanınmıştır. Tüm test verilerinin sadece %1'i tanınmamıştır.

Geri beslemeli eğitim algoritmasının kullanıldığı paralel-çok katmanlı yapay sinir ağ yapıları ile konuşma tanıma uygulaması gerçekleştirilirken kod kitabında yer alan 400 kelime'nin tamamı kullanılmıştır. Ses verileri 32x30x10 mimarisine sahip 40 alt gruba ayrılmış yapay sinir ağlarına paralel olarak sunulmuştur. Sonuçta en büyük tanıma başarısını sağlayan ağ veriyi tanımış olmaktadır. Ağların bir veriyi yanlışlıkla kazanan hepsini alır kuralı gereği kendileri için bloke edilen kodlardan birine yakınsamaları için bir eşik değeri kullanılmıştır. Bu bildiriye kullanılan ilgili eşik değeri 0.80-0.99 için tanıma başarısı %77.83-%86.02 arasında değişmiştir. Bununla birlikte yapay sinir ağları ile konuşma tanıma eğitim aşaması diğer algoritmalara göre aşırı uzundur. Elde edilen tanıma başarılarını arttırmak için ilgili ağlar DSA kullanılarak tekrar sub-optime edilmiştir. Bu durum ağların yanlış sınıflandırdığı (yanlışlıkla kendisine bloke edilen kelimelerden birine yakınsattığı) kelimelerin tanıma başarısını önemli ölçüde arttırmıştır; DSA'nın kullanımı ardından tanıma başarısı ortalama %2.77 artmıştır. Bu durumda elde edilen tanıma başarıları %79.98-%88.40 olarak elde edilmiştir.

#### 4. SONUÇLAR

Bu bildiriye Konuşma tanımanın CBS'te kullanılma potansiyelinden bahsedilmiş ve temel konuşma tanıma algoritmalarının başarıları geniş bir sözlük kullanılarak konuşmacıya bağımlı ayrık konuşma tanıma yaklaşımına göre ayrıntılı olarak incelenmiştir. Elde edilen tanıma sonuçları, konuşma tanımanın CBS'te ileride ciddi uygulama alanları bulacağını göstermektedir. Kod-Kitabının oluşturulmasında kullanılan terimler Ulaştırma Bakanlığınca hazırlanan Ulaştırma ve Haberleşme Terimleri sözlüğünün Karayolları Terimleri bölümünde yer alan terimler arasından derlenmiştir (Bakanlık 2012). Kod kitabında yer alan terimlerin kombinasyonu ile farklı niyetler ifade edilebilmektedir.

#### Teşekkür

Bu çalışma, 110Y309 nolu TÜBİTAK projesi kapsamında desteklenmiştir.

#### KAYNAKLAR

- Çivicioğlu, P.**, 2012, Transforming geocentric cartesian coordinates to geodetic coordinates by using differential search algorithm, *Computers & Geosciences*, 46, 229-247.
- Huang, X. D., Jack, M.A., et al.**, 1988, Parameter Reestimation In Semicontinuous Hidden Markov Modeling Of Speech With Feedback To Vector Quantization Codebook, *Electronics Letters*, 24 (22), 1375-1377.
- Huang, X. D., ve Jack, M.A.**, 1988, Performance Comparison Between Semicontinuous And Discrete Hidden Markov-Models Of Speech, *Electronics Letters*, 24 (3), 149-150.
- Huo, Q., ve Lee, C.H.**, 1998, On-line adaptive learning of the correlated continuous density hidden Markov models for speech recognition, *IEEE Transactions on Speech and Audio Processing*, 6(4): 386-397.
- Irwin, M.J.**, 1988, A Digit Pipelined Dynamic Time Warp Processor, *IEEE Transactions on Acoustics Speech and Signal Processing*, 36(9): 1412-1422.
- Issa, K. ve Nagahashi, H.**, 2008, Detection and classification of the spliced yarn joint using vector quantization and dynamic time warping, *Journal of the Textile Institute*, 99 (4), 347-358.
- Jang, M., Han, M.S., ve değerleri**, 2011, Dynamic Time Warping-Based K-Means Clustering for Accelerometer-Based Handwriting Recognition, *Developing Concepts in Applied Intelligence*, 363: 21-26.
- Kruskal, J.B., ve Liberman, M.** 1999, The symmetric time-warping problem: From continuous to discrete. *Time Warps, String Edits, and Macromolecules: The Theory and Practice of Sequence Comparison*, 1983.
- Martin, R., C. Hoelper, et al.** 2001, Estimation of missing LSF parameters using Gaussian Mixture Models. 2001 Ieee International Conference on Acoustics, Speech, and Signal Processing, Vols I-Vi, Proceedings: Vol I: Speech Processing 1; Vol II: Speech Processing 2 Ind Technol Track Design & Implementation of Signal Processing Systems Neuralnetworks for Signal Processing; Vol III: Image & Multidimensional Signal Processing Multimedia Signal Processing: 729-732.