Electronic Letters on Science & Engineering 2(2) (2006)
Available online at www.e-lse.org

Speaker Recognition Using Mel-Frequency Cepstrum Coefficients and Artificial Neural Network for Security Systems

Ahmet Küçüker[1] ,Can Yüzkollar[2], Ahmet Sansli[2],Fatih Sen[2]
[1]Sakarya University, Electronic Engineering Department, Esentepe Campus,
[2]Sakarya University, Computer Engineering Department, Esentepe Campus,
Sakarya, Turkey

Abstract: In this study; the sample sounds of the speakers are transfered to computer for identifing the speaker. Study is based on three steps; In first step analog sound that taken is converted to data to be processed and determine own characteristic of each sound. in second step converted sounds called data are used to train neural network. at last step trained neural network is test by the test sounds and obtained istatistical informations.

Keywords: Neurel Network, Speaker Identification, Speech Recognition, Mel Cepstrum Coefficients

Güvenlik Sistemleri Için Mel Frekans Kepstrum Katsayilari ve Yapay Sinir Aglari Kullanilarak Konusmaci Tanima

Özet: Bu çalismada konusmaci ses örnekleri bilgisayara alinarak bu konusmacilarin kimliklerinin belirlenmesi saglanmistir. Çalisma üç asamadan olusmaktadir. Ilk asama konusmaci sesinin dijital ortamda veriye dönüstürülmesi islenebilir hale getirilerek özellik çikartilmasi ikinci asama elimizdeki konusmaci ses örnekleriyle yapay sinir aginin egitilmesi ve son asama da egitilmis sinir agina test verileri göndererek istatistik bilgiler elde edilmesidir.

Anahtar Kelimeler: Sinir Aglari, Konusmaci kimliklendirme, Konusmaci tanima, Mel Kepstrum Katsayilar
Reference to this paper should be made as follows (bu makaleye asagidaki sekilde atifta bulunulmali):
A. Kucuker et al, 'Speaker Recognition Using MFCC and Artificial Neural Network For Security Systems' , Elec Lett Sci Eng , vol. 2 (2) , (2006), 8-15

1 Giris

Otomatik olarak tanimlama veya dogrulama için kullanilan, insana ait, ölçülebilir, essiz fizyolojik ölçüler vardir. Bu ölçülerin kaybolmasi, unutulmasi ve bir baskasi tarafindan kullanilmasi, fiziksel saldirilar disinda pratikte imkânsiz kabul edildigi gibi taklit edilmeleri de çok zordur. Gelistirilen teknolojileri kullanarak her ortamda kisinin el, parmak izi, ses, göz, imza, retina, kulak sekli, DNA, klavyeye basis deseni, koku ve yürüyüs kistaslarin ölçülerek taninmasi mümkündür. [1] Dolayisiyla konusmaci tanimlama bu alanda kullanilabilir nitelik kazanmaktadir. Böyle bir sisteme ihtiyaç duyulmasinin sebebi insanlarin parola hatirlamalarinin gerekliligini ortadan kaldirmak ya da parolanin çalinmasi ihtimaline karsi ek bir tedbir almak da denilebilir. Bu yüzden günden güne ASR(Automatic Speaker Recognition) otomatik konusmaci tanima sistemleri gelistirilmektedir. . Konusmaci tanima sistemlerinin güvenlik disinda da birçok uygulama alani vardir. Konusma tanima sistemleri sayesinde telefonda kredi karti numarasi tuslamak yerine numarayi sesli söyleyebiliyor veya arabada sesli komutlarla bilgiye ulasabiliyoruz. [2]

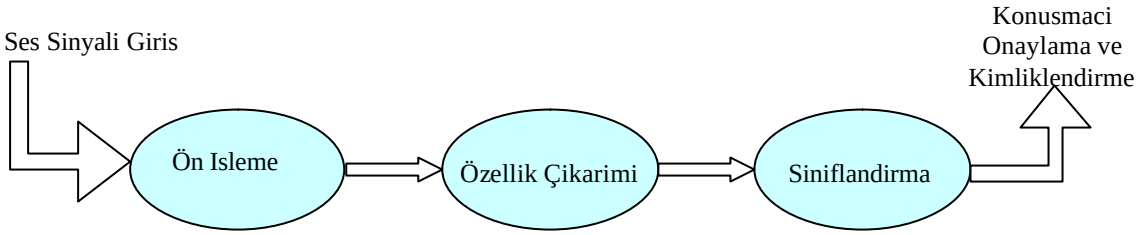
2 Konusmaci Tanima Sistemi

Insanin ses özelliklerini barindiran ses dalgalari mikrofon vasitasiyla elektrige çevrilerek bilgisayara veri olarak alinmaktadir. Birçok arastirmaci istatistiksel özellikler kullanarak konusmaci tanima üzerine çalisma yapmistir [3]. Konusmaci tanima sistemini fonksiyon yönünden iki baslik altinda siniflandirabiliriz. Ilki konusmaciyi onaylama ikincisi konusmaciyi

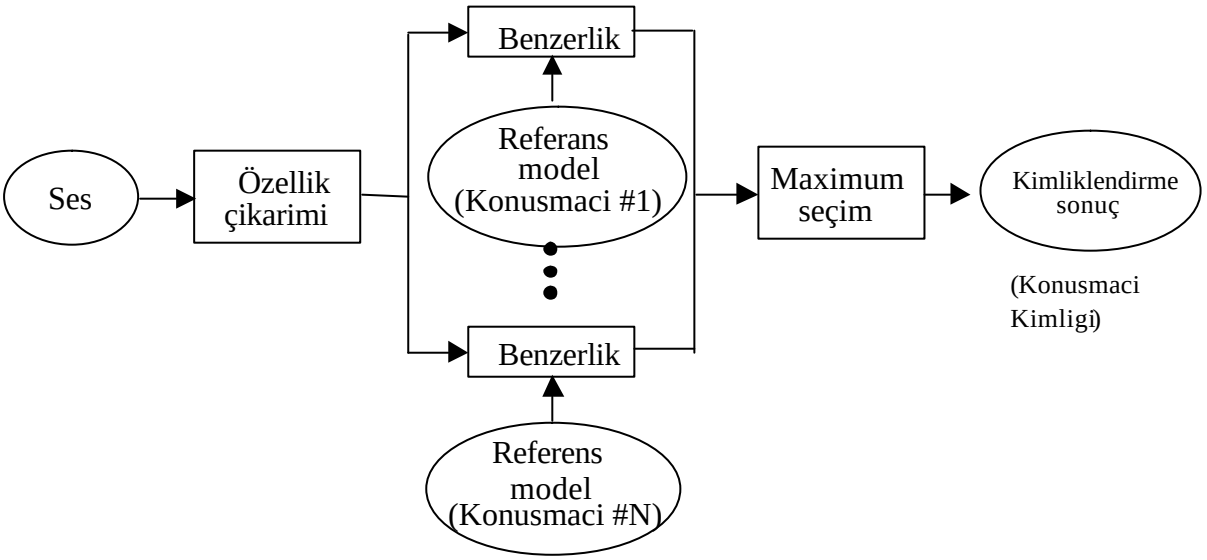
*Corresponding author; Tel.: +(90) 264 2955140 , E-mail: kucuker@sakarya.edu.tr , can@sakarya.edu.tr
ISSN 1305-8614

kimliklendirmedir. [4]. Uygulamamizda Mel Frekans Kepstrum Katsayilari yöntemiyle konusmacinin seslerinden gerekli özellikler çikarilmis, çikarilan bu özelliklere Yapay zeka teknigi uygulanarak konusmacinin taninmasi saglanmistir.

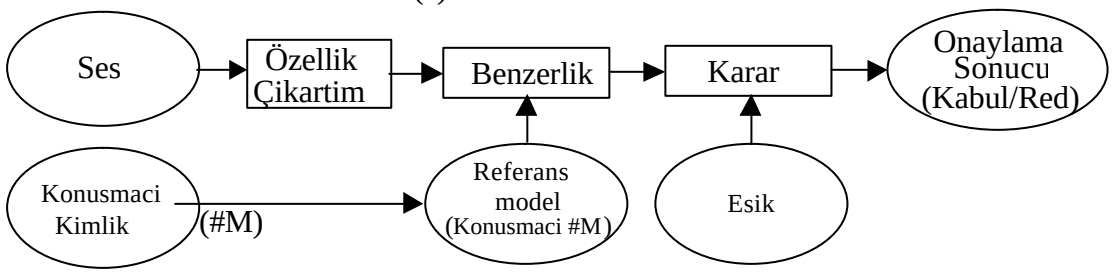
Konusmaci onaylama, konusmacinin kimlik talebini onaylama ya da onaylamama islemidir. Konusmaci onaylamanin iki çesit hata kriteri vardir. Yanlis Kabul(False Accept) ve Yanlis Red (False Reject) Bu iki kriterin kesisimi ise " Esit Hata Orani (Equal Error Rate) " olarak adlandirilir.Uygun bir tanima sistemi için Esit Hata Orani ; Yanlis Kabul ve Yanlis Red arasinda uzlastirici görev yapmalidir.[5] Konusmaci kimliklendirme sunulan insan seslerinin hangi konusmaciya ait oldugunun belirlenmesi islemidir. Temel bir konusmaci tanima sistemi sekilde gösterildigi gibidir. (Sekil 1)



Sekil-1 Temel Konusmaci Tanima Sistemi Blok diyagrami



(a) Konusmaci Kimliklendirme



(b) Konusmaci Onaylama

Sekil 2 Konusmaci Onaylama ve Kimliklendirme

* Corresponding author; Tel.: +(90) 264 2955140 , E-mail: kucuker@sakarya.edu.tr, can@sakarya.edu.tr
ISSN 1305-8614

Konusmaci Tanima sistemini fonksiyon yönünden iki bölüme ayrildigi gibi method olarak da iki bölüme ayrilmaktadir. Bunlardan ilki *Konusma Metninden Bagimsiz* konusmaci tanima digeri *Konusma Metnine Bagimli* Konusmaci tanima olarak adlandirilir [4]. Çalismamizda ikinci metod kullanilmistir.

3 Özellik Çikarim

Bu çalismada özellik çikarim yöntemlerinden olan mel-frekans Cepstrum katsayilari yönteminden faydalanilmistir. [6 , 7] *Cepstrum* kavrami ilk olarak 1963'de *Bogert, Healy* ve *Tukey* tarafindan kullanilmistir. *Cepstrum*, homomorfik sinyal isleme teknikleri içinde yer alir. [6]. Homomorfik sistemler dogrusal olmayan sistemlerin bir sinifi olarak kabul edilirler. Dogrusal sistemler homomorfik sistemlerin özel bir durumudur. Sesli ifade baglaminda kullanilan homomorfik sistem, *S(f)=V(f).G(f)* seklinde ifade edilebilir. Burada *s(n)* sesli ifadeyi, *v(n)* girtlagi, yani, sesin izledigi yolu, *g(n)* ise asil ses sinyalini, yani ses telleri tarafindan üretilen ve degisime ugramamis ses sinyalini temsil eder. Bu sekilde ifade edildiginde girtlagin etkisini asil sesten homomorfik sinyal isleme teknikleriyle ayirmak mümkün olmaktadir. *Cepstrum* degerlerinin hesaplanmasi için (Es. 1.1) esitligi kullanilir. [7]

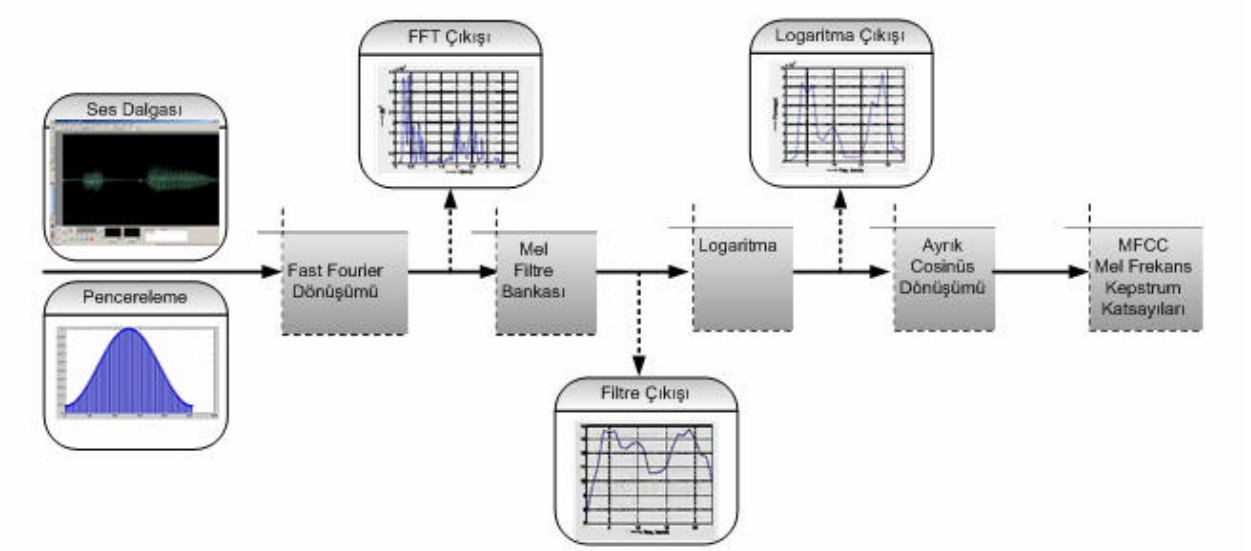
$$c(n) = \frac{1}{N_s} \sum_{k=0}^{N_s-1} \log_{10} \left| S(k) \right| e^{(2p/N_s)kn}, \quad 0 \le n \le N_s - 1 \qquad (1.1)$$

Burada *c(n), n. Cepstrum* olarak adlandirilir. *S(k), Cepstrum* degerinin ait oldugu frekans araligi için alinan *fourier* dönüsümünü belirtir. *Ns* o andaki çerçevenin boyunu gösterir. Burada dikkat edilecegi gibi *c(0)* dogrudan o andaki DFT (**Discrete Fourier Transform**) spektrum degerini gösterir. Sesli ifadenin gürültüden ayirt edilmesi için önce spektrumun logaritmasi alinir ve ters *fourier* dönüsümü yapilir. Bu sekilde belirlenen *Cepstrum* degerleri *fourier* dönüsümünden türetilmis *cepstral k*atsayilar(*fourier transform derivated cepstral coefficients*) olarak adlandirilir Burada hesaplanan *Cepstrum* degeri sesli ifade tanimada kullanilan önemli bilgileri elde etmede etkilidir. *Fourier d*önüsümünde kullanilan frekans *mel* skalasinda örneklenirse elde edilen *Cepstrum* degerleri *Mel Cepstrum* degerleri olarak adlandirilir . [7]

Mel frekans cepstral özellikleri için her pencereye sirasiyla su adimlar uygulanir[5]. (Sekil 3)

- Ayrik Fourier Dönüsümü alinir
- Ayrik Fourier Dönüsümü katsayilari Mel filtre bankasinin genlik frekans cevabina göre agirliklandirilir.
- Logaritmik Enerjileri Hesaplanir
- Ayrik Kosinüs Dönüsümü bulunur

* Corresponding author; Tel.: +(90) 264 2955140 , E-mail: kucuker@sakarya.edu.tr, can@sakarya.edu.tr
ISSN 1305-8614



Sekil 3 Özellik Çikarim Adimlari Semasi

Mel Cepstrum degerlerini hesaplamak için Mel filtre bankasi yöntemi (1.2) formülüyle uygulanir. (Davis and Mermelstein). Burada X_k *mel* skalasinda k. bant geçisli filtrenin uygulanmasiyla elde edilen spektrum degeridir. Mel Filtre bankasinin amaci duyma mekanizmasinin kritik bant filtrelerini simulasyonunu yapmaktir. Filtreler Mel Skalasinda düzenli bir sekilde yerlesen üçgensel filtrelerdir. [7 , 10 , 11]. Üçgensel filtre çikislari Y (i); i = 1; : : : ;M logaritma kullanarak sikistirilmis ve ayrik kosinüs dönüsümü uygulanmistir. [8].

$$melc(n) = \sum_{i=0}^{M} \log Y(i) \cos\left[\frac{pn}{M}(i - \frac{1}{2})\right] \qquad (1.2)$$

Bu fonksiyonun sonucunda elde edilen katsayilar ile sesimizin artik yapay sinir agina girebilecek sekle yani islenebilecek sekle getirildigini söyleyebiliriz.

Özellik çikariminda farkli bir yaklasim olarak Yapay Sinir Aglari da kullanilmaktadir [12, 13]

4- Yapay Sinir Aglari

Yapay sinir aglari (YSA) günümüzde bilgi siniflama ve bilgi yorumlamanin içinde bulundugu degisik problemlerin çözümünde kullanilmaktadir [9]. Özelliklerini elde ettigimiz ses sinyallerinin taninmasi da bu kapsama girmektedir. Çalismamizda genelleme özelligini kullanarak bozulmus seslerin de taninmasi üzerinde durulmustur.

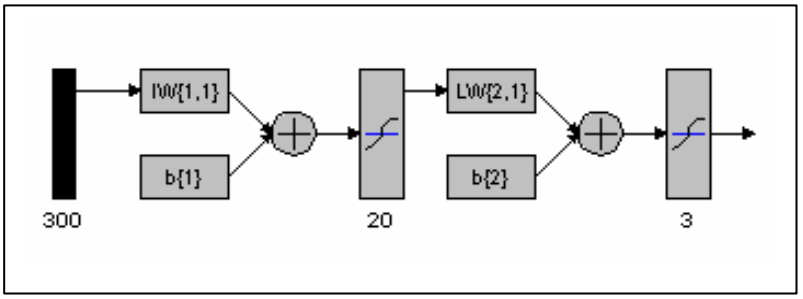
Yapay sinir aglari söyle siniflandirilmaktadir:

- Tek katmanli ileri beslemeli aglar : Katmanli modellerdeki en basit ag tipi olup bir çikti katmani ve buna bagli bir girdi katmani bulunmaktadir
- Çok katmanli ileri beslemeli aglar : Çalismamizda bu yöntemi seçmemizin sebebi tek katmanli aglara göre daha karmasik problemlere çözüm getirebilmesidir. Dezavantaji ise Tek katmanli ileri beslemeli aglara göre egitimi daha uzun sürmektedir [10]

Tek katmali aglardaki girdi ve çikti katmanindan baska, bir yâda daha fazla sayida gizli katman içeren aglara çok katmanli ag denir. Dis dünya tarafindan dogrudan müdahale

* Corresponding author; Tel.: +(90) 264 2955140 , E-mail: kucuker@sakarya.edu.tr, can@sakarya.edu.tr
ISSN 1305-8614

edilmedigi için gizli katman adi verilen katmanda bulunan birimlere de gizli birimler adi verilir (Sekil 4).



Sekil 4 Çok katmanli ileri beslemeli ag modeli

Geri yayilimli ögrenme metodu olarak isimlendirilen egitim algoritmasinin temeli ilk olarak Werbos'un Harvard Üniversitesi'nde verdigi doktora tezinde ileri sürülmüs [11], daha sonra Parker tarafindan MIT'nin bir teknik raporunda ele alinmis [12] ve Rumelhart ve arkadaslari tarafindan da popüler hale getirilip uygulanabilir sekle dönüstürülmüstür.

Ileri besleme asamasinda egitim için kullanilan girdi, sisteme beslenir ve bunun sonuçlari her bir katmani geçerek çikti katmanina kadar gelir ve girdiye karsilik bir çikti elde edilir. Bu asama sirasinda ag üzerindeki agirlik degerleri sabit tutulur.

Ikinci asama olan geri besleme asamasinda, elde edilen çikti ile hedef çikti arasindaki farktan hata sinyali elde edilir ve bu sinyal ag yapisinda geriye dogru yayilir. Bu geri yayilim asamasinda, olusan hatayi minimuma indirecek sekilde agirlik degerleri güncellenir.[13] Bu durumda sistemin egitilmesi uzun süre almakla beraber, egitilmis bir sistemden bilgi alinmasi çok hizli olmaktadir

5 Uygulama

Uygulamada konusmaciyi tespit etmek için kapi kelimesi kullanilmistir. Kapi kelimesi çok basit bir kelime gibi görünmesine ragmen YSA nin tanimasi açisindan büyük zorluk içermektedir. Fonetik olarak söylenisi rahat oldugu için herkes birbirine yakin sekilde telafuz etmektedir. Bu sebeple bu kelimeyi taniyabilen bir sistem diger kelimeleri daha rahat ögrenecektir. Örnegin Sicak kelimesi gibi bir kelime kullanilmis olsaydi S harfi ve K harfinin sonda kullanilmasi kisiden kisiye daha çok farklilik gösterdiginden tanima islemi çok daha kolaylasmis olacakti. [14]

Seçtigimiz bu kelime için her bir konusmacidan 10 örnek alinarak toplam 30 ses kaydi kullanilmistir. Alinan veriler Matlab ortaminda *voicebox toolbox* kullanilarak MFCC fonksiyonu ile katsayilar elde edilmistir [15]. Elde edilen 25 satir 12 sütunlu bu öznitelik vektörlerinin yapay sinir agina girebilmesi için gereken matris dönüsümleri yapilmis ve 300*30 luk egitim seti (aas1 olarak adlandirildi) hazirlanmistir. Daha sonra her bir konusmacidan 5 er test sesi kaydedilerek gerekli matrisler elde edilmistir.

Sinir agi asagidaki Matlab kodlariyla olusturulmustur. [18, 20]

```
net = newff(minmax(aas1),[20 3],{'tansig' 'tansig'},'trainscg');
```

* Corresponding author; Tel.: +(90) 264 2955140 , E-mail: kucuker@sakarya.edu.tr, can@sakarya.edu.tr
ISSN 1305-8614

Egitim kodlari [18, 20]

```
net.trainParam.epochs = 1250;
net.trainParam.goal = 1e-003;
net.trainParam.min_grad = 1e-07;
net.trainParam.time = Inf;
net.trainParam.show = 1;
net = train(net,aas1,cikisss);
```

Elimizdeki test seslerinin kime ait oldugunu belirlemek için asagidaki kod kullanilmistir.

```
sonuç=sim(net,ses1)
```

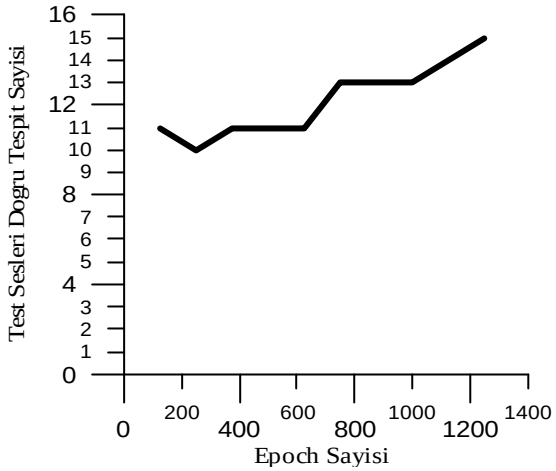
6 Sonuç

Egitimli sinir agina gönderilen test seslerine verilen cevaplar kaydedilerek ilgili dogruluk oranlarinin epochs sayisina göre dagilimi asagida gösterilmistir. (Tablo 1)

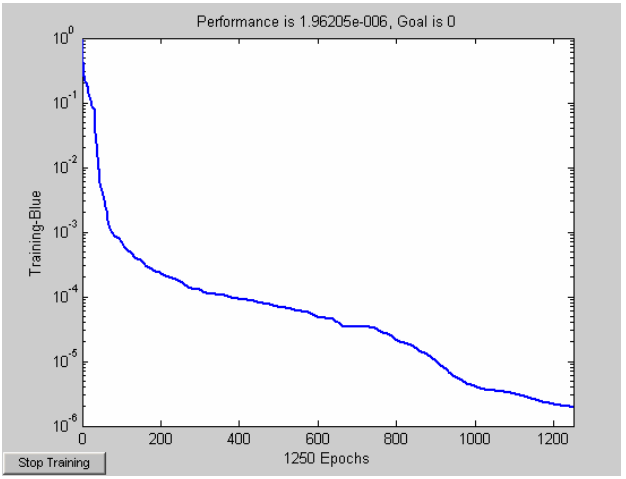
Tablo 1. Dogruluk oranlarinin epochs sayisina göre dagilimi

Epoch Sayisi / Test Sesleri	125	250	375	500	625	750	875	1000	1125	1250
Ak1	+	+	+	+	+	+	+	+	+	+
Ak2	+	+	+	+	+	+	+	+	+	+
Ak3	+	+	+	+	+	+	+	+	+	+
Ak4	+	+	+	+	+	+	+	+	+	+
Ak5	+	+	+	+	+	+	+	+	+	+
As1	+	+	-	-	-	-	-	-	+	+
As2	+	+	+	+	+	+	+	+	+	+
As3	+	+	+	+	+	+	+	+	+	+
As4	+	+	+	+	+	+	+	+	+	+
As5	-	-	-	-	-	-	-	-	-	+
C1	-	-	-	-	-	+	+	+	+	+
C2	-	-	+	+	+	+	+	+	+	+
C3	-	+	-	-	-	+	+	+	+	+
C4	+	+	+	+	+	+	+	+	+	+
C5	+	+	+	+	+	+	+	+	+	+

* Corresponding author; Tel.: +(90) 264 2955140 , E-mail: kucuker@sakarya.edu.tr, can@sakarya.edu.tr
ISSN 1305-8614



Sekil 5 – Tablo 1 Degerlerinin Grafigi



Sekil 6 – Egitim Performans Grafigi

Egitim sesleri test sesi olarak kullanildiginda tanima islemi 125 epochs tan itibaren tüm sesler için dogru sonuç vermistir.

7 Tartisma ve Öneriler

Bu çalismada üç konusmacinin ses örnekleri bilgisayara alinarak Mel Frekans Kepstrum Katsayilari yöntemiyle konusmacinin seslerinden gerekli özellikler çikarilmis, çikarilan bu özelliklere Yapay zeka teknigi uygulanarak konusmacinin taninmasi saglanmistir
Tanima sonuçlarina bakildiginda basarim oraninin % 100 olmasi bu sistemin kullanilabilir bir sistem oldugunu ortaya koymustur. Fakat ses taklidi düsünüldügünde bu sistemin parmak izi tanima vb. sistemlerle birlikte kullanimi sistemin kullanilabilirlik oranini artiracaktir.

Referanslar

[1] Simon Liu, Mark Silverman. "A Practical Guide to Biometric Security Technology," IT Professional, vol. 03, no. 1, pp. 27-32, January/February, 2001.

[2] Hemphill, C. T., Agarwal, R., Muthusamy, Y. K., and Gong, Y. "Voice-Driven Information Access in the Automobile", IEEE Vehicular Technology Society News,August, 8-11, 2000

[3] FURUI S., "Comparison of Speaker Recognition Methods Using Statistical Features and Dynamic Features", IEEE Transaction on ASSP, Vol. 29,No. 3, pp. 342-350, June 1981.

[4] INAL M., FATIHOGLU Y. S. " Self Organizing Map And Associative Memory Model Hybrid Classifier For Speaker Recognition" Kocaeli 2002

[5] FARRELL, K. R., MAMMONE, R., J., and ASSALEH, K., T., Jan. 1994. Speaker Recognition Using Neural Networks and Conventional Classifiers, IEEE Trans. On Speech and Audio Processing, Vol.2, No.1,part II.

[6] A.V.Oppenheim and R.W.Schafer, Discrete-Time Signal Processing, Prentice Hall, Englewood Cliffs, NJ, 1989

[7] MENGÜSOGLU, Erhan " Rule Based Design And Implementation Of A Speech Recognition System For Turkish Language" 1999

[8] HUANG, X., Acero, A., and Hon, H.-W. Spoken Language Processing: a Guide to Theory, Algorithm, and System Development. Prentice-Hall, New Jersey, 2001.

* Corresponding author; Tel.: +(90) 264 2955140 , E-mail: kucuker@sakarya.edu.tr, can@sakarya.edu.tr
ISSN 1305-8614

[9] ELMAS, Ç., 2003, "Yapay Sinir Aglari (Kuram,Mimari,Uygulama)", Seçkin Yayinlari.

[10] AYDIN Ö. "Yapay Sinir Aglarini Kullanarak Bir Ses Tanima Sistemi Gelistirilmesi" Yüksek Lisans Tezi, Edirne 2005

[11] WERBOS, P. J., 1974, "Beyond Regression: New Tools for Prediction and analysis in the behavioral sciences

[12] PARKER, D. B., 1985, "Learning-logic: Casting the Cortex of the Human Brain in Silicon", Technical Report TR-47, Center for Computational Research in Economics and Management Science, MIT, Cambridge, MA.

[13] TEMURTAS Dr.F. "Yapay Sinir Aglari Ders notlari" SAÜ 2005

[14] KUCUKER Dr.P. "Fonetik Ders Notlari" SAÜ 2004

[15]. MATLAB® Documentation (2002) Neural Network Toolbox Help, Version 7.0, Release

* Corresponding author; Tel.: +(90) 264 2955140 , E-mail: kucuker@sakarya.edu.tr, can@sakarya.edu.tr