

**T.C
FIRAT ÜNİVERSİTESİ
FEN BİLİMLERİ ENSTİTÜSÜ**

YAPAY ZEKÂ TABANLI KONUŞMA TANIMA SİSTEMİ

YÜKSEK LİSANS TEZİ

Mehmet Bilal ŞAŞMAZ

Anabilim Dalı: Elektronik ve Bilgisayar Eğitimi

Programı: Telekomünikasyon

AĞUSTOS-2011

**T.C.
FIRAT ÜNİVERSİTESİ
FEN BİLİMLERİ ENSTİTÜSÜ**

YAPAY ZEKÂ TABANLI KONUŞMA TANIMA SİSTEMİ

YÜKSEK LİSANS TEZİ

Mehmet Bilal ŞAŞMAZ

(091131105)

Anabilim Dalı: Elektronik ve Bilgisayar Eğitimi

Programı: Telekomünikasyon

Tez Danışmanı: Yrd. Doç. Dr. Davut HANBAY

Tezin Enstitüye Verildiği Tarih: 9 AĞUSTOS 2011

AĞUSTOS-2011

T.C
FIRAT ÜNİVERSİTESİ
FEN BİLİMLERİ ENSTİTÜSÜ

YAPAY ZEKÂ TABANLI KONUŞMA TANIMA SİSTEMİ

YÜKSEK LİSANS TEZİ

Mehmet Bilal ŞAŞMAZ

(091131105)

Tezin Enstitüye Verildiği Tarih: 9 AĞUSTOS 2011

Tezin Savunulduğu Tarih: 25 AĞUSTOS 2011

Tez Danışmanı: Yrd. Doç. Dr. Davut HANBAY (F.Ü)

Diğer Jüri Üyeleri: Doç. Dr. Sami EKİCİ (F.Ü)

Doç. Dr. Engin AVCI (F.Ü)

ÖNSÖZ

Bu çalışmamda engin bilgileriyle yol gösteren, sabrı ve bilgisiyle bana destek olan değerli danışman hocam Sayın Yrd. Doç. Dr. Davut HANBAY'a ,Sayın hocalarım Doç.Dr İbrahim TÜRKOĞLU'na, Doç. Dr. Abdulkadir ŞENGÜR'e, Doç.Dr.Engin AVCI'ya, Doç.Dr.Sami EKİCİ'ye ayrıca tez çalışmam boyunca birlikte hareket ettiğimiz Sayın Eda Rençber'e teşekkürü bir borç bilir, saygılarımı sunarım.

Ayrıca bu süre boyunca bana tahammül eden, hayatımın her anında maddi ve manevi desteklerini devamlı hissettiren anne babama ve değerli eşime sonsuz teşekkürlerimi sunuyorum.

Mehmet Bilal ŞAŞMAZ

ELAZIĞ-2011

İÇİNDEKİLER

	<u>Sayfa No</u>
ÖNSÖZ.....	I
İÇİNDEKİLER.....	II
ÖZET	IV
SUMMARY.....	V
ŞEKİLLER LİSTESİ.....	VI
TABLolar LİSTESİ.....	VII
KISALTMALAR LİSTESİ	VIII
SEMBOLLER LİSTESİ.....	IX
1. GİRİŞ.....	1
1.1. Akustik	1
1.2. İnsanda Konuşma, İşitme	2
1.3. Konuşma Tanıma	5
1.4. Konuşma Tanıma İle İlgili Yapılan Çalışmalar	5
2. KONUŞMA TANIMADA KULLANILAN YÖNTEMLER.....	8
2.1. Saklı Markov Modeli	8
2.2. Dinamik Zaman Bükmesi.....	10
2.3. Vektör Nicemleme (Vector Quantization)	10
3. FONETİK.....	12
3.1. Parçalı Ses Birimleri.....	12
3.1.1. Türkçe’de Parçalı Ses Birimleri.....	12
3.2. Parçalarüstü Ses Birimleri	13
4.KONUŞMA TANIMA SİSTEMLERİ	14
4 .1. Öznitelik Çıkarma Metotları.....	14
4.1.1. MFKK(Mel-Frekans Kepstrum) Katsayıları	14
4.1.1.2. Çerçeveleme.....	15
4.1.1.2. Pencereleme (Windowing)	15
4.1.1.3. Hızlı Fourier Dönüşümü(HFD)	16
4.1.1.4.Mel-Frekansına Çevirme	16
4.1.1.5. Mel Frekansı Kepstrum Katsayıları (MFKK).....	17
4.2. Dalgacık Dönüşümü.....	17

4.2.1. Sürekli Dalgacık Dönüşümü	19
4.2.2. Ayırık Dalgacık Dönüşümü	22
4.2.3. Dalgacık Paket Analizi	23
4.2.4. Entropi Hesaplama Teknikleri	24
4.2.5. Dalgacık(Wavelet) Dönüşümü ile Entropi Hesaplama	27
5. AKILLI SİSTEMLER	29
5.1. Yapay Sinir Ağları	29
5.1.1. Temel Yapay Sinir Ağı Kavramları	29
5.1.2. Yapay Sinir Hücresi (Nöron)	30
5.1.3. Aktivasyon Fonksiyonu	30
5.1.4. Katmanlar	31
5.1.5. Yapay Sinir Ağı Topolojileri	31
5.1.6. İleri Beslemeli Ağlar	32
5.1.7. Geri Dönüşümlü Ağlar	32
5.1.8. Yapay Sinir Ağlarının Eğitilmesi	34
5.1.9. Öğrenme Stratejileri	34
5.1.10. Hata Fonksiyonu	35
5.1.11. Öğrenme Kuralları	36
5.2. Uyarlamalı Sinirsel Bulanık Ağ (USBA) Mimarisi	37
5.2.1. USBA İçin Geri Yayılımlı Öğrenme Algoritması	40
6. YAPAY ZEKÂ TABANLI KONUŞMA TANIMA UYGULAMALARI	46
6.1. MFKK ve Yapay Zekâ ile Konuşma Tanıma	46
6.1.1. MFKK ve YSA ile Konuşma Tanıma	46
6.1.2. MFKK ve USBA ile Konuşma Tanıma	48
6.2. DD Kullanarak Yapay Zekâ Tabanlı Konuşma Tanıma	49
6.2.1. DD ve YSA ile Konuşma Tanıma	49
6.2.2. DD ve USBA ile Konuşma Tanıma	50
7. SONUÇLAR VE TARTIŞMA	52
KAYNAKLAR	54
ÖZGEÇMİŞ	57

ÖZET

Konuşma tanıma günümüz bilgi teknolojilerinde popüler bir konu olmaya başlamıştır. Konuşma tanıma sayesinde kelimelerin dijital ortama transferi diğer yöntemlerden çok daha hızlı ve kusursuz yapılmaktadır. Ayrıca konuşma tanıma ile cihazların sesle kontrolü sağlanabilmektedir.

Bu çalışmada konuşma tanıma sistemlerinde kullanılan genel metotlar hakkında bilgi verilmiştir. En yaygın kullanılan yapay zekâ metotlarından YSA ve UBSA tanıtılmıştır. Gerçeklenen konuşma tanıma sisteminde ilk olarak Dalgacık Dönüşümü ile entropi değerleri ve MFCC hesaplanarak konuşmayı temsil eden özellikler çıkarılmıştır. Daha sonra elde edilen özellik vektörleri kullanılarak yapay zekâ tabanlı konuşma tanıma sistemlerinin eğitimi gerçekleştirilmiştir. Eğitilen modeller daha sonra konuşma saptama başarımı için test edilmiştir. Elde edilen sonuçlar tablolar halinde verilmiştir.

Anahtar kelimeler: konuşma tanıma, yapay sinir ağları, mfcc, entropi, dalgacık dönüşümü

SUMMARY

Speech recognition becomes popular for information technologies in recent years. Translating words to the digital media is being easier thanks to the speech recognition. Most of the electronic devices can be controlled by voices using speech recognition technologies.

In this study, basics methods for speech recognition are described briefly, the most popular artificial intelligence methods NN and ANFIS was also described. At first in the realized systems the features vectors of speech signals based on wavelet transform entropy and mel frequency cepstral coefficients are evaluated. After than these feature vectors are used for training of artificial intelligence based speech recognizer models. The trained models are tested for speech determination. Obtained results are tabulated.

Keywords: Speech recognition, neural network, Adaptive Network Based Fuzzy Inference System, Entropy, Wavelet Transform

ŞEKİLLER LİSTESİ

	<u>Sayfa No</u>
Şekil1.1. Konuşma ve duyma sürecine ilişkin şematik yapı	3
Şekil1.2. Sesoluşumuna ilişkin anatomik yapı	3
Şekil 1.3. Duymaya ilişkin anatomik yapı	4
Şekil 2.1. Ses Tanıma sistemine ilişkin genel blok yapı.....	8
Şekil2.2. Basit bir soldan sağa SMM.....	9
Şekil 4.1. Öznitelik çıkarma	15
Şekil 4.2. Mel yayımlımlı filtre bankası.....	17
Şekil 4.3. Pencerelemiş Fourier dönüşümü için zaman-frekans diyagramı	18
Şekil4.4. Dalgacık dönüşümü için zaman-frekans diyagramı	18
Şekil4.5. Zaman-frekans diyagramı	19
Şekil 4.6. Dalgacığın sinyal üzerinde kaydırılması	20
Şekil 4.7. Ölçeğin değiştirilmesi.....	20
Şekil 4.8. Sinyal spectogramı	21
Şekil 4.9. Sinyalin üç boyutlu gösterimi	22
Şekil 4.10. ADD ayrışım ağacı.....	23
Şekil 4.11. DPA ağaç yapısı	24
Şekil 4.12. Düşük entropi kavramı.....	25
Şekil 4.13. Yüksek entropi kavramı.....	25
Şekil 4.14. Entropi katsayıları çıkarma.....	27
Şekil 5.1. N adet girdisi olan basit bir nöron.....	30
Şekil 5.2. İleri beslemeli yapay sinir ağı topolojisi.....	32
Şekil 5.3. Geri beslemeli yapay sinir ağı topolojisi	33
Şekil 5.4. Bir yapay sinir ağının eğitilmesi işlemi.....	35
Şekil 5.5. 2 girişli 4 kurallı bir UBSA sınıflandırıcı yapısı.....	38
Şekil 6.1. Konuşma sinyalinin hızlı fourier dönüşümü.....	46
Şekil 6.2. HFD'si alınan sinyalden Keptal katsayıların elde edilmesi.....	47

TABLÖLAR LİSTESİ

	<u>Sayfa No</u>
Tablo 3.1. Türkçe için hece yapısı ve ilgili yapılara ilişkin örnekler	13
Tablo 4.1. İşaret işlemede yaygın kullanılan entropi türleri ve denklemleri.....	26
Tablo 6.1. Çok katmanlı yapay sinir ağının yapısı ve eğitim parametreleri	48
Tablo 6.2. MFKK ve YSA ile konuşma tanıma sonuçları	48
Tablo 6.3. USBA modelinin yapısı ve eğitim parametreleri.....	49
Tablo 6.4. MFKK ve USBA ile konuşma tanıma sonuçları.....	49
Tablo 6.5. Uyarlamalı Sinirsel Bulanık Ağ sisteminin parametreleri ve başarı yüzdesi ..	50

KISALTMALAR

VN	: Vektör Nicemleme
LDA	: Linear Discriminant Analysis
YSA	: Yapay Sinir Ağları
MFKK	: Mel Frekans Kepstral Katsayı
SMM	: Saklı Markov Model
ADD	: Ayrık Dalgacık Dönüşümü
DD	: Dalgacık Dönüşümü
DPA	: Dalgacık Paket Ayrıştırma
DZE	: Dinamik Zaman Eğirme
USBA	: Uyarlamalı Sinirsel Bulanık Ağ

SEMBOLLER LİSTESİ

ε	: Öğrenme katsayısı
a	: İşlem elemanının x aktivitesini ayarlayan reel değerli bir sabit
$\mathcal{C}$	: Çıkış
D	: Uzaklık ölçütü
d_i	: i . hücrenin çıkışı
e	: Hata ölçütü
$E(t)$	: Hata fonksiyonu
f	: Hertz biriminden frekans
F	: Eşik Fonksiyonu
m	: Mel frekansı
M	: Kepstrum katsayılarının sayısı
n	: Pencere merkezini belirleyen örnek indis
p, q ve r	: Lineer çıkış parametreleri
R	: Referans (eğitim) şablonu
s	: İşaret
s_i	: İşaretin i . katsayısı
T	: Toplam fonksiyonu
W_i	: Ağırlık katsayıları
$x(t)$	: İşaret
x_i	: Giriş
X_k	: k . filtrenin log-enerjisi
y_i	: i . hücrenin çıkışı
λ	: Gaussian karma modeli
$\mu_i(x)$	: Üyelik derecesi
$\psi(t)$	: Dalgacık

1. GİRİŞ

Teknolojinin gelişiminde canlıların sahip oldukları özelliklerin insan yapımı araç ve gereçlere aktarılması isteği önemli bir yere sahiptir. Ses ile ilgili ilk çalışmalar öncelikle sesin fiziki yapısını anlamaya yöneliktir. Sonraları biyolojik olarak insanda ses oluşumu ve işitmenin fizyolojisi anlaşılmaya çalışılmıştır. Ses oluşumunun yapısının ortaya konması ile yapay konuşma kaynakları modellenmiş ve gerçekleştirilmiştir[1]. Gerçekleştirilen bu sistemler teknolojik gelişimi izleyerek sırayla mekanik, elektromekanik ve elektronik sistemler olarak ortaya çıkmıştır. Günümüzde konuşma sentezleme olarak kendine yer bulan sistemler kişisel bilgisayarlarda da birtakım kısıtlamalar çerçevesinde gerçekleştirilebilen uygulamalar olarak kullanılabilmektedir. Sinyal işleme ve bilgisayar bilimlerindeki gelişmelere paralel olarak konuşma tanıma; yönelik farklı yöntem ve metotlar ortaya konmuştur. Günümüzde konuşma tanıma; otomatik çeviri, komut tanıma, otomobil içi konuşma tanıma, insan-bilgisayar ara yüzü, ev otomasyonu, robotik uygulamaları, bilgisayar temelli dil öğreniminde telaffuz geliştirme, otomatik veri girişi ve benzeri konularda uygulama alanı bulmaktadır[1]. Konuşma tanıma yöntemlerinin gelişimi ile biyometrik sistemlerin bir parçası olarak konuşmacı tanıma ve doğrulama sistemleri geliştirilmiştir. Konuşma tanıma ve doğrulama sistemleri güvenlik uygulamalarında geniş bir kullanım alanı bulmaktadır.

1.1 Akustik

Konuşma ve duymanın temelinde ses dalgası bulunmaktadır. Ses dalgası yapı itibarıyla bir basınç dalgasıdır ve ortamdaki parçacıkların basınç değişimini iletmesi ile yayılır.

Sesin bir ortamda yayılması ile ortama ilişkin basınç, yoğunluk, ortamdaki parçacıkların hızı ve sıcaklık değerlerinde zamansal değişiklikler meydana gelir. Bu değişimlerin sağlaması gereken süreklilik ve sınır koşulları göz önüne alınarak, $\rho(\vec{r}, t)$ basıncın zaman ve yere göre değişimini, $u(\vec{r}, t)$ parçacık hızının zamana ve yere göre değişimini göstermek üzere yoğunluğu ρ olan bir ortamda düzlemsel bir ses dalgası denklem (1.1) ile verilen koşulu sağlaması gerektiği söylenebilir [2].

$$-\frac{\partial \rho}{\partial \chi} = \rho \frac{\partial u}{\partial t} \quad (1.1)$$

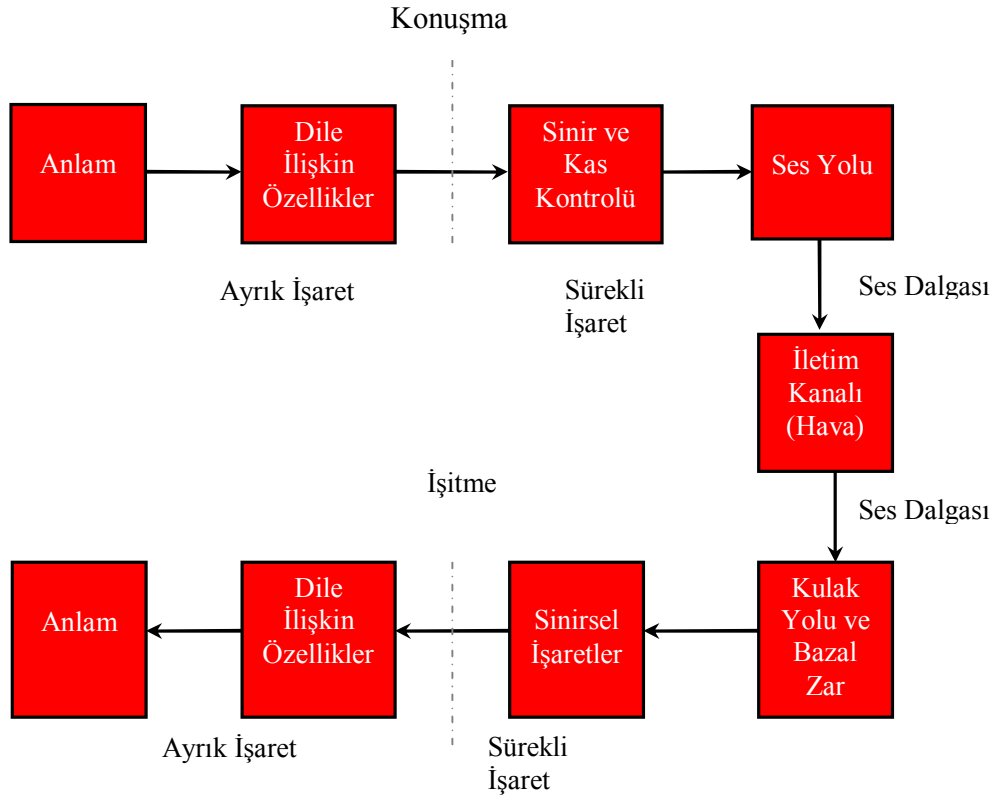
Ses dalgasının yayıldığı ortamın düzgün olmayan bir tüp olarak modellenmesi ile süreklilik ve sınır koşulları yardımıyla insanda ses yoluna ilişkin rezonans frekansları, formant değerleri, anti-rezonans değerleri gibi fiziki özellikler hesaplanabilmektedir. Hesaplanan bu değerler ses tanımada akustik özelliklerin belirlenmesinde temel oluşturmaktadır [2].

1.2 İnsanda Konuşma, İşitme

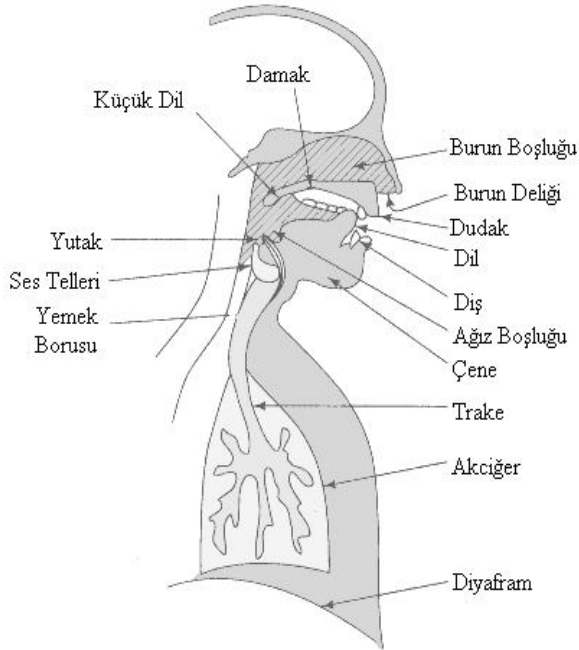
Konuşma tanıma ve konuşma sentezinde insan konuşması ve duyması model olarak alındığı için insanda konuşmanın ve duymanın anlaşılması önemli ve gereklidir.

İnsanda konuşmanın oluşması, sinirsel olarak beyinde anlamsal ifadenin oluşturulmasından sonra, gerekli sinirsel sinyallerin iletimi, ses mekanizmasının uygun şekilde çalıştırılması ile gerçekleştirilir ve oluşan ses dalgası hava yardımıyla iletilir. İşitme ise kulağa ulaşan ses dalgasının işlendikten sonra sinirsel yollarla beyne ulaştırılıp anlamsal birimlere dönüştürülmesi ile gerçekleşir. Şekil 1.1 ile bahsedilen yapının şematik gösterimi verilmiştir.

Ses oluşumu mekanik olarak akciğer, soluk borusu, ses telleri, gırtlak, çene, dil, diş, küçük dil, damak, burun boşluğu ve dudak gibi organların etkisiyle sağlanmaktadır (Şekil 1.2). Bu yapı bir sistem yapısı ile modellenerek konuşmaya ilişkin akustik model elde edilmiştir[2].



Şekil 1.1.Konuşma ve duyma sürecine ilişkin şematik yapı[3]



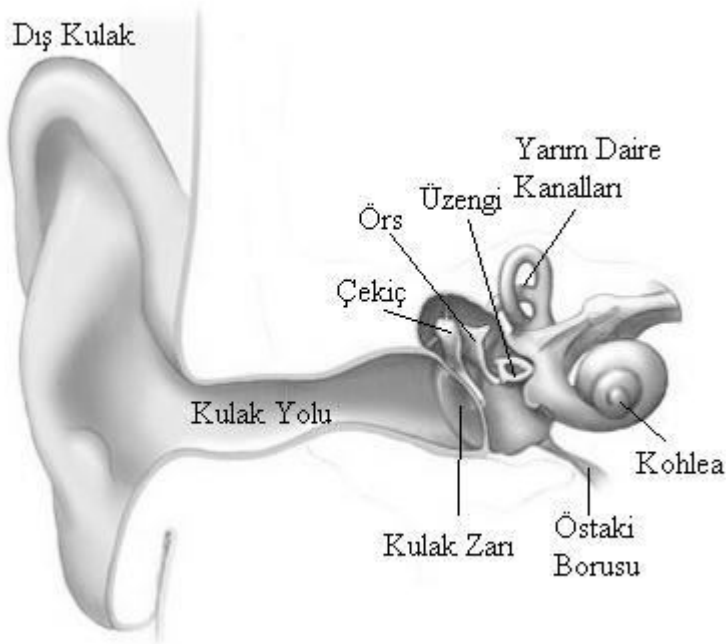
Şekil 1.2. Ses oluşumuna ilişkin anatomik yapı [4]

Akustik teoriye göre ses oluşumu kaynak, filtre ve yayılma ile ilgili transfer fonksiyonlarına sahip alt sistemlerden oluşmaktadır. Buna göre ses ve konuşma ile ilgili:

$$P_r(Z) = S(Z)T(Z)R(Z) \quad (1.2)$$

S: kaynak transfer fonksiyonu, T: ses organları ile ilgili transfer fonksiyonu, R: havada yayılma ile ilgili transfer fonksiyonu olmak üzere P_r : ağızdan r mesafesindeki sesin transfer fonksiyonu olarak elde edilir.

Duyma ise ses üretiminde var olan transfer fonksiyonlarının tersleri ile oluşturulmuş bir sistem olarak düşünülebilir. Hava yolu iletilen ses dalgaları dış kulak yardımıyla kulak yoluna iletilir. Kulak zarında oluşan titreşimler çekiç, örs ve üzengi kemikleri ile duyma organı kohleaya ulaştırılır. Kohleada bulunan bazılar zar üzerindeki kılcak yapılar ve lif yoğunluğunun kohlea boyunca giderek artan yapı göstermesi, kohleanın bir ucunda düşük frekansa, diğer ucunda yüksek frekansa duyarlılığı sağlar. Bazılar zar üzerindeki kılcak korti reseptör hücreleri ile sinirsel işarete dönüştürülen veri duyma siniri ile beyne iletilir. Duymaya ilişkin anatomik yapı Şekil 1.3.'da gösterilmiştir.



Şekil 1.3. Duymaya ilişkin anatomik yapı [1]

1.3 Konuşma Tanıma

Konuşma tanıma günlük hayatta pek çok ortam ve uygulamada kendine yer bulmuş bir konudur. Genel olarak konuşma tanıma ile insan konuşmasının bilgisayar tarafından anlaşılması ve buradan bilgi çıkarımı hedeflenmektedir. Konuşma tanıma problemi birçok farklı boyut ile tanımlanabilmektedir. Konuşma tanıma üzerine yapılan çalışmalar artık günümüzde bahsedilen çok boyutlu problem uzayının belirli bölgelerinde özelleşmektedir. Problem uzayı aşağıda bahsedilen boyutlar ve uç durumlar ile tanımlanabilmektedir.[3] Konuşma modu: ayrık kelime tanıma ↔ sürekli konuşma tanıma

Tanıma birimi: sözcük tabanlı ↔ fonem tabanlı

Konuşma stili: okuma ↔ spontane konuşma

Konuşmacıya bağlılık: konuşmacı bağımsız tanıma ↔ konuşmacı bağımlı tanıma

Sözlük genişliği: 20 kelime ↔ 20000 kelime

Dil modeli: sonlu durum makineleri ↔ doğal dil işleme ile içerik temelli

Sinyal için gürültü oranı: 10 dB ↔ 25 dB

Kullanılan alıcı: gürültü önleyen mikrofon ↔ cep telefonu [2]

Yukarıda bahsedilen problem uzayının bir bölgesini ilgilendiren konuşma tanımaya ilişkin izlenecek adımlar üç aşama olarak incelenebilir; bunlar önışlemler, özellik çıkarma ve sınıflandırma aşamalarıdır. Önışlem aşaması genel olarak ses verisinin çeşitli filtreleme ve ön-vurgulama işlemlerinden geçirilmesi ile ses verisinin nasıl ve hangi çözünürlükte sayısallaştırılacağına ilişkindir. Özellik çıkarması ses verisinde konuşma tanıma yapılabilmesi için ayırt edici özelliklerin çeşitli sinyal işleme teknikleri ile ortaya konmasını amaçlamaktadır. Özellik değerlerinin sınıflandırılması ise ses verisinin ardışıl yapısını göz önüne alarak özellik dizisinden en olası kelime çözümlemesinin elde edilmesini sağlar.

1.4 Konuşma Tanıma İle İlgili Yapılan Çalışmalar

Ses ile ilgili bilimsel çalışmalar tarihsel sırayla ses sentezleme, ses sinyali işleme ve konuşma tanıma üzerine yoğunlaşmıştır. Ses sentezleme için sırasıyla mekanik, elektro-mekanik ve elektronik sistemler kullanılmıştır. Ses sinyali işleme tarihsel sırasıyla zaman boyutunda, frekans boyutunda ve kepsral özelliklerin elde edilmesi ile ilgilenmiştir.

Konuşma tanıma konusunda ise tarihsel olarak şablon karşılaştırma, örüntü tanıma ve istatistikî modellerle ilgilenilmiştir[2].

Son yapılan çalışmalarda konuşma tanımada en çok kullanılan yöntemlerden birisi yapay sinir ağları olmuştur. İlk olarak McCullough ve Pitts tarafından 1943'te ortaya konan yapay sinir ağları, 1980'lerde hata geri yayılım (error back-propagation) metodu bulunana kadar geniş bir kullanım alanı bulamamıştır. Hata geri yayılım metodunun ortaya konması ile yapay sinir ağlarının fonem tanıma, ayrık rakam tanıma gibi işlerde başarılı sonuç verdiği gösterilmiştir [3]. 1990'larda minimum sınıflandırma hatası kavramının benimsenmesi ile ayırıcı eğitim (discriminative training) yöntemi ve Destek Vektör Makinesi (Support Vector Machine) gibi çekirdek tabanlı (kernel based) metotlar konuşma tanımada kullanılmaya başlanmıştır [1,2].

1996'da Hu ve arkadaşları dil için temel birim olarak gördükleri hecelerden ve hece benzeri ses birimlerinden hareketle İngilizce ay isimleri için 29 heceden oluşan tanıma sistemleri ile %84'lük başarı oranı yakalamışlardır. Aynı yıl Boulard, bezer bir çalışmayı Almanca için gerçekleştirmiştir. Hauenstein Saklı Markov Modeli ve yapay sinir ağlarının birlikte kullanıldığı bir sistem tasarlayarak bu yapının sadece bir yöntemin kullanıldığı diğer tasarımlarından daha iyi sonuç verdiğini göstermiştir [2].

Wu ve arkadaşları tarafından yürütülen bir çalışma ile hece ve fon veya fonem bazında hibrit sitemlerinin tanıma başarısını arttırdığı gösterilmiştir[2].

Ganapathiraju, Hamaker, Picone, ve Doddington 2001 yılında geliştirdikleri hece temelli konuşma tanıma çalışmaları ile fon veya fonem temelli sistemlerin başarı oranına ulaşmışlardır.

Günümüzde ticari veya araştırma amaçlı geliştirilen konuşma tanıma sistemleri için elde edilen hata oranları: rakam tanıma için % 0,3, 1000, kelimelik sözlüklü okuma sesi için %3, 20,000, kelimelik sözlüklü okuma sesi için %6, 10,000, kelimelik sözlüklü karşılıklı konuşma için %20, telefon üzerinden 10,000, kelimelik sözlüklü konuşma için %30, seviyelerinde gerçekleşmektedir.

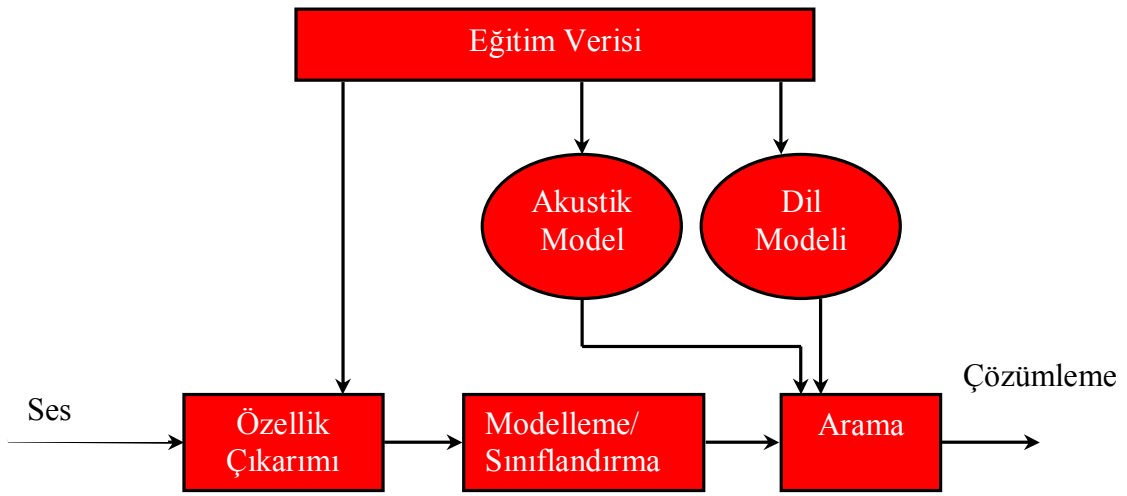
Bu tez çalışmasında fonem temelli bir konuşma tanıma sistemi tasarlanmıştır. Bu amaçla Mel Frekans Kepstral Katsayısı (MFKK), Dalgacık dönüşümü ve entropi değerleri özellik çıkarımı için kullanılmıştır. Elde edilen özellikler YSA ve USBA yapay zekâ metotları ile sınıflandırılarak konuşma tanıma sistemleri gerçekleştirilmiştir.

Bu çalışmanın ilk bölümünde konuşma tanıma ile ilgili temel konular ve yapılan çalışmalar anlatılmıştır.

2. bölümde genel olarak konuşma tanıma sistemlerinde kullanılan sınıflandırıcılar açıklanmıştır.
3. bölümde, Türkçe konuşma tanımada kullanılan parçalı ses birimleri ile genel olarak parçalı ses birimlerinden bahsedilmiştir.
4. bölümde, MFKK ve Dalgacık dönüşümü tabanlı özellik çıkarımı anlatılmıştır.
5. MFKK ve DD tabanlı YSA ve USBA mimarisine ilişkin teorik yapı anlatılmıştır.
6. bölümde, yapılan uygulamalar detaylı olarak anlatılmıştır.
7. bölümde, elde edilen sonuçlar değerlendirilmiştir.

2. KONUŞMA TANIMADA KULLANILAN YÖNTEMLER

Konuşma tanıma problemini iki ana başlık altında; özellik çıkarımı ve özellik vektörlerinin sınıflandırılması olarak ele almak gerekir. Bu şekilde özellik çıkarım yöntemleri ile özelliklerin sınıflandırılması için kullanılan yöntemlerin alan olarak birbirleri ile karışmalarının önüne geçilmiş olur. Konuşma tanıma ile ilgili genel yapı Şekil 2.1.de gösterilmiştir.



Şekil 2.1. Ses Tanıma sistemine ilişkin genel blok yapısı [3]

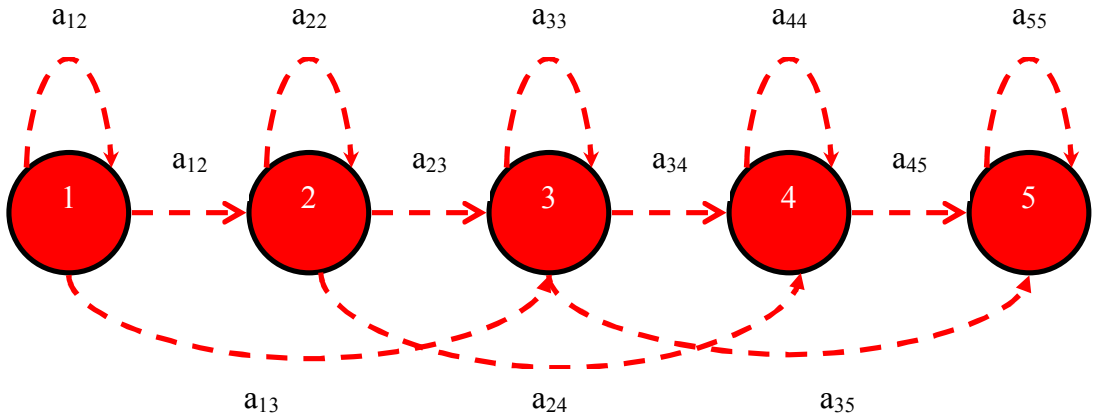
2.1.Saklı Markov Modeli

Bu metottaki temel fikir ses sinyalinin parametrik bir rasgele işlem olarak ifade edilebilmesidir. Saklı Markov Modeli (SMM) iki stokastik süreç içerir. İlk olan Markov süreci zaman ile ilgili değişikliklerde kullanılır ve durumları içeren bir Markov zinciri üretir. Diğer süreç gözlemlenebilir özellik parametreleri veya gözlemler denilen rasgele değişkenler içerir [2].

Aslında her söylem ideal durumda bir SMM'e sahip olmalıdır. Bazen bu mümkün olmaz, bu yüzden kelime düzeyinde SMM' lerimiz olmalıdır. Söylenen kelime ile en uygun kelimeyi eşlemek istediğimiz için, bir SMM bir veritabanındaki tüm kelimeler için en iyisini yapabilmelidir. Ama bu durumda sözlük, zaman kısıtlamalarının üstesinden gelmek için yeterince küçük olmalıdır.

SMM'in yapısı (Şekil 2.2) bir durumlar zincirinden meydana gelir. SMM zinciri üzerindeki her durum kelimenin bir parçasına karşılık gelir. Her durum bir diğerine geçişlerle bağlıdır.

Geçişler, geçiş olasılıklarına (a_{ij}) bağlı olarak durum değiştirmeye imkân verir. Durumlara iliştilen emisyon olasılıkları (b_j) bir özellik vektörünün, referansın belirli bir zaman aralığıyla olan spektral benzerliğini gösterir. Sistem girdisine göre oluşturulan özellik vektörleri dizisine bağlı olarak, model üzerinde birinci durumdan başlayan farklı yollar izlenebilir. Bazı durumların tekrarı veya atlanması kullanıcının konuşma hızındaki değişimlere sistemin adaptasyonunu sağlar. Bir kelimenin tanınabilmesi için referans olarak alınan durumdan itibaren izlenen yolun en son duruma, kabul edilebilir bir olasılıkla ulaşması gereklidir [2]. Bir SMM modeli her anda durumu değişen birimleri olan bir sonlu durum makinesidir. Her t ayrık zaman anında, i durumundan j durumuna geçiş gerçekleşir ve gözlem vektörü o_t yoğunluk vektörü $b_j(o_t)$ ile dışarı verilir. Bundan başka i durumundan j durumuna geçiş aynı zamanda rasgeledir ve a_{ij} yoğunluğu ile olur. Şekil 2.2'de, beş durumlu soldan sağa SMM atlamasız olarak verilmiştir.



Şekil2.2. Basit bir soldan sağa SMM

Her durum bir diğer duruma geçişlerle bağlıdır. Geçişler, geçiş olasılıklarına (a_{ij}) bağlı olarak durum değiştirmeye imkân verir. Bir konuşma modelinin tanınabilmesi için referans olarak alınan durumdan itibaren izlenen yolun en son duruma kadar kabul edilebilir bir olasılıkla ulaşması gereklidir. SMM'ler ile çözülebilecek üç temel problem vardır.

- Model parametreleri verildiği zaman çıkış sırasının olasılığını hesaplamak (Forward algorithm)
- Model parametreleri verildiğinde istenen bir çıkışı sağlayacak gizli durumların sırasını hesaplamak (Viterbi algorithm)
- Verilen bir çıkış sırasına göre durum geçişlerini ve çıkış olasılıklarını hesaplamak (Baum-Welch algorithm)

SMM' ye dayalı konuşma tanıma sistemi işlem basamaklarının değişik adımlarında bu üç algoritma kullanılarak bir sonuca ulaşılır. Modelleme aşamasında forward-backward algoritmaları kullanılarak konuşmacıya ait model parametreleri elde edilir. Karar verme aşamasında ise Viterbi algoritması kullanılarak, oluşturulan model ile var olan model arasındaki benzerlik ölçülür.

2.2. Dinamik Zaman Bükmesi

Dinamik zaman bükme yöntemi konuşma tanıma yöntemlerinde sıklıkla kullanılan bir diğer yöntemdir. Bu yöntem daha çok diğer yöntemlerle birlikte kullanılan ve daha çok tanıma işlemlerinin verimliliğini artırmak amacıyla kullanılan bir yöntemdir. Bu yöntemde, konuşma ifadelerini seslendirme süreleri sıkıştırılarak ya da genişletilerek referanslarla karşılaştırılmaları ilkesi kullanılmaktadır [4]. Aynı sözcüğü aynı kullanıcı tekrar seslendirdiğinde bile bir seslendiriliş daha önceki seslendirilişlere benzemeyebilir. Çünkü sözcüğün uzunluğu doğrusal olmayan bir biçimde genişleme ve daralma gösterir[1]. Dinamik zaman bükme yöntemi sözcüğün ya da fonemin sinyalinin, referans şablonu ile aynı zaman aralığında olabilmesi için zaman ekseninde daralma ya da genişleme yapmayı amaçlar. Sözcük tanıma ya da fonem tanıma için genel olarak dinamik zaman eşleştirme yöntemi kullanılmaktadır. Dinamik zaman eşleştirme yönteminde zaman eksenini doğrusal olmayan bir biçimde genişletilip daraltılarak referans şablonu ile tanınacak olan sesli ifade kesiminin başlangıç ve bitiş zamanları çakıştırılmaya çalışılır. Amaç karşılaştırmanın aynı zaman aralıkları için yapılmasını sağlamaktır. Dinamik zaman bükme işlemi, devingen programlama tekniği kullanılarak gerçekleştirilir.

2.3.Vektör Nicemleme

Vektör nicemleme teknikleri veri sıkıştırma ve kodlama için kullanılan yöntemlerdir. Prensipte olarak hem alıcı (encoder) hem de vericide (decoder) mevcut iki tablonun

(look-up table) verilerine, bu verilerin sıra numarası (index) kullanılarak ulaşmayı hedefler[4]. Eğer tablolarda yer alan veriler vektörler ise yöntem vektör nicemleme adını alır. Vektör nicemleme de önemli olan kılavuz kitap (code-book) olarak da adlandırılan tabloların oluşturulmasıdır. Bu tablolar oluşturulur iken nicemlenecek vektörlerin istatistiksel özellikleri kullanılarak karşılaştırılması muhtemel tüm verilerin kılavuz kitapta yer almasına (ya da belli bir hata kriterine göre en az hata ile yer almasına) dikkat edilmektedir. Kılavuz kitap/kitaplar oluşturulurken Lloyd algoritması kullanılır[4]. Bu algoritma, genel olarak, tekrar eden iki basamaktan oluşmaktadır.

Lloyd Algoritması:

- 1) Mevcut tüm vektörlerin “en yakın komşu” kuralına göre merkeziciklerin etrafına kümelenmesi
- 2) Eski merkeziciklerin belli bir hata kriterine göre en az hata verecek şekilde yeniden hesaplanması

3. FONETİK

Konuşmada anlam farkına neden olan en küçük ses birimi fon olarak adlandırılmaktadır. Fonlar tüm diller için ortak bir yapıda değildir. Her dil için kendine özgü ses birimleri mevcuttur. Kimi dillerde ses birimleri doğrudan harflere karşılık gelirken, kimi dillerde ise sesler harflerden farklı olarak simgesel telaffuz birimleri olarak fonemlere karşılık gelmektedir. Konuşma tanımada ses birimleri parçalı ses birimleri ve parçalar üstü ses birimleri olarak iki ana başlık altında incelenir.[1]

3.1 Parçalı Ses Birimleri

Ses birimlerinin fonlara veya fonemlere karşılık düşen yapısı parçalı ses birimleri olarak isimlendirilir. Konuşma tanımada parçalı ses birimleri üzerinden sistemin modellenmesi ile daha geniş sözlüklü tanıma sistemleri tasarlanabilmekte, fakat gerek sözlük genişliğinden gerekse parçalı ses birimlerinin sahip olduğu yüksek standart sapmadan dolayı bu tür sistemlerin tanıma başarısı diğer sistemlere kıyasla daha düşük çıkabilmektedir.

3.1.1 Türkçede Parçalı Ses Birimleri

Türkçe Altay dilleri içerisinde Oğuz grubuna dahildir. Türkçede kullanılan ses birimleri ünlü-ünsüz olma durumuna, ses tellerinin titreşimin, seslerin çıkış yeri ya da sesi çıkaran organ veya seslerin çıkış biçimine göre sınıflandırılmaktadır[1]. Türkçede ünlü sesler {a, e, ı, i, o, ö, u, ü}, ünsüz sesler ise {b, c, ç, d, f, g, h, j, k, l, m, n, p, r, s, ş, t, v, y, z} olarak sınıflandırılmaktadır. Ünsüz sesler, ses tellerinde titreşime sebep olup olmamalarına göre ötümlü sesler {b, d, g, v, z, j, c, l, r, m, n, y} ve ötümsüz sesler {p, t, k, f, s, ş, ç, h} olarak sınıflandırılmaktadır. Ünlü sesler çıkış yerlerine göre dil önü ünlü sesleri {i, ü, e, ö} ve dil arkası ünlü sesleri {ı, u, a, o} olarak ayrıştırılırken ünsüz sesler çıkış yeri ve yardımcı organa göre çift dudak {p, b, m}, alt dudak-üst dişler {f, v}, dil ucu-diş arkası {t, d}, dil ucu- diş eti {s, z, n, r, l, ç, c}, dil önü-sert damak {ş, j, y}, dil arkası-damak {k, g}, ses teli-gırtlak {h} sesleri olmak üzere sınıflandırılmaktadır.

Seslerin çıkış biçimlerine göre sınıflandırılmasında ses organlarının aldıkları durumlar önemlidir. Ünlü sesler çıkış biçimlerine göre dil-damak arası açıklığa göre dar

{i, ü, ı, u}, orta {e, ö, ü}, geniş {a};dudakların biçimine göre düz {i, ı, e, a}, yuvarlak {u, ü, o, ö} olarak sınıflandırılırlar.

Ünsüz sesler çıkış biçimlerine göre ağız ve geniz sesleri olarak sınıflandırılır. Geniz ünsüzleri {m, n} sesleridir. Ağız ünsüzleri ise patlamalı {b, p, d, t, g, k}, sızmalı {v, f, z, s, j, ş, h}, patlamalı sızmalı {ç, ç}, yan ünsüz {l}, çarpmalı ünsüz {r}, yarı ünlü {y} olarak sınıflandırılmaktadır[1].

3.2 Parçalar üstü Ses Birimleri

Parçalar üstü ses birimleri heceler veya sözcüklerdir. Ağızın tek bir hareketiyle çıkartılabilen bir ünlü ya da bir ünlü ile bir veya birkaç ünsüz sesin birleşmesiyle oluşturulan ses modeline hece denir. Hecelerin birleştirilmesinden sözcükler oluşur.

Türkçe için altı çeşit hece yapısından söz edilebilir. Türkçe için altı çeşit hece yapısı ve ilgili yapıya ilişkin bir örnek tablo 3.1 ile verilmiştir [1].

Tablo 3.1. Türkçe için hece yapısı ve ilgili yapılara ilişkin örnekler

Hece Yapısı	Örnek
Ünlü	A
Ünlü ünsüz	Ek
Ünlü ünsüz ünsüz	İlk
Ünsüz ünlü	Ca
Ünsüz ünlü ünsüz	Sel
Ünsüz ünlü ünsüz ünsüz	Sert

4.KONUŞMA TANIMA SİSTEMLERİ

4.1. Öznitelik Çıkarma Metotları

Bir konuşma tanıma uygulaması gerçekleştirirken, ses sinyallerinin tanınabilmesi için öncelikle bu sinyallerin doğru bir şekilde ifade edilmesi gerekmektedir. Bir başka deyişle, tetkik edilen ses sinyalinin içinde barındırdığı ve yalnızca tanınması hedeflenen kelimeye özel unsurlar belirlenmelidir. Daha sonra belirlenen bu unsurlar bir öznitelik vektörü ile ifade edilmesi gerekir[5].

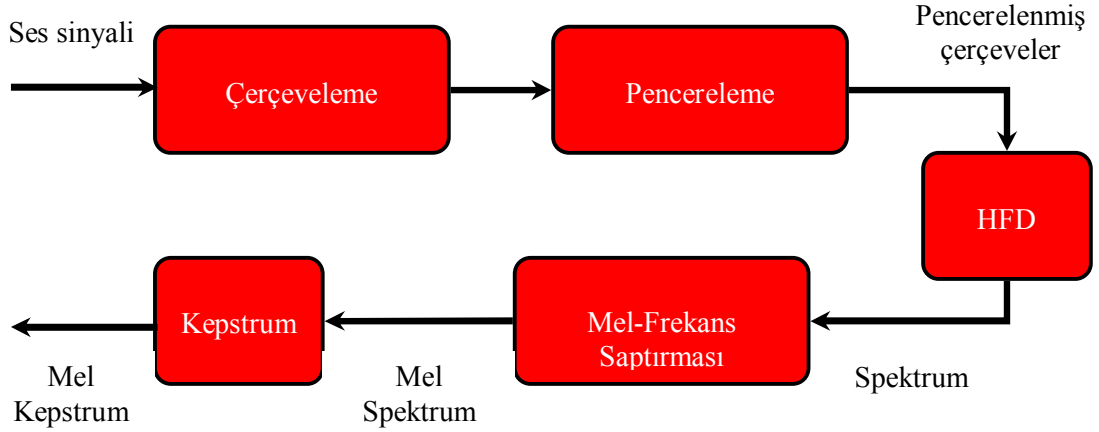
Ses sinyalindeki o sese özel unsurların çıkartılması işlemine öznitelik çıkarma denir.

4.1.1. MFKK(Mel-Frekans Kepstrum Katsayıları)

MFKK katsayıları, konuşma tanıma alanında en çok kullanılan öznitelik çıkarma yöntemlerinden birisidir [6,7]. Öznitelik çıkarma işlemi hem eğitim (kaydolma) hem de test (saptama) aşamasında kullanılır ve aşağıdaki aşamalardan oluşur [8,9].

1. Çerçeveleme (Frame Blocking)
2. Pencereleme (Windowing)
3. Fourier Dönüşümü (FD)
4. Mel-Frekansı Saptırması (Mel-Frequency Warping)
5. Kepstrum (Cepstrum)

Öznitelik vektörlerini çıkarmaktaki gaye, ses örneğindeki, konuşmacıyı tanımlayan sessel özelliklerden ödün vermeden yüklü miktardaki veriyi özetlemektir [10]. Böylece tanıma işlemi kolaylıkla gerçekleşir. Öznitelik çıkarma işlemleri Şekil 4.1’de gösterilmiştir.



Şekil 4.1. Öznitelik çıkarma

4.1.1.1. Çerçeveleme (Frame Blocking)

Sürekli konuşma sinyali N adet konuşma örneği içeren çerçevelere ayrılır ve her komşu çerçeve $M < N$ şartını sağlayacak şekilde oluşturulur. Birinci çerçeve N adet konuşma örneği içerirken ikinci çerçeve birinci çerçeveden M adet örnek sonra başlatılır. Böylece her çerçeve kendisinden bir önceki çerçevenin belli bir kısmını örtmüştür. Örtme işleminin amacı bir çerçeveden diğer çerçeveye geçişi yumuşatmaktır.

4.1.1.2. Pencereleme (Windowing)

İkinci aşamada ise çerçevelenen sinyal pencerelenir. Pencerelemenin amacı çerçevenin başında ve sonunda bulunan süresizlikleri ortadan kaldırmaktır. Böylece öznitelik vektörüne katkı sağlamayacak katsayıların azaltılması amaçlanmaktadır. Bu aşamada en çok kullanılan pencereleme yöntemi Hamming fonksiyonudur. Hamming fonksiyonunun tanımı denklem 4.1 deki gibidir[7].

$$p(t) = \begin{cases} 0,54 + 0,46 \cos 2\pi t / N & -N/2 \leq t \leq N/2 \\ 0, & |t| > N/2 \end{cases} \quad (4.1)$$

N, pencere süresini gösterir.

4.1.1.3. Hızlı Fourier Dönüşümü (HFD)

Bir sonraki aşama, her çerçevede HFD almaktır. Bu dönüşüm Ayrık Fourier Dönüşümünün hızlı (AFD) bir şeklidir ve tanım kümesini zaman domeninden frekans domenine geçirir.

4.1.1.4. Mel-Frekansına Çevirme

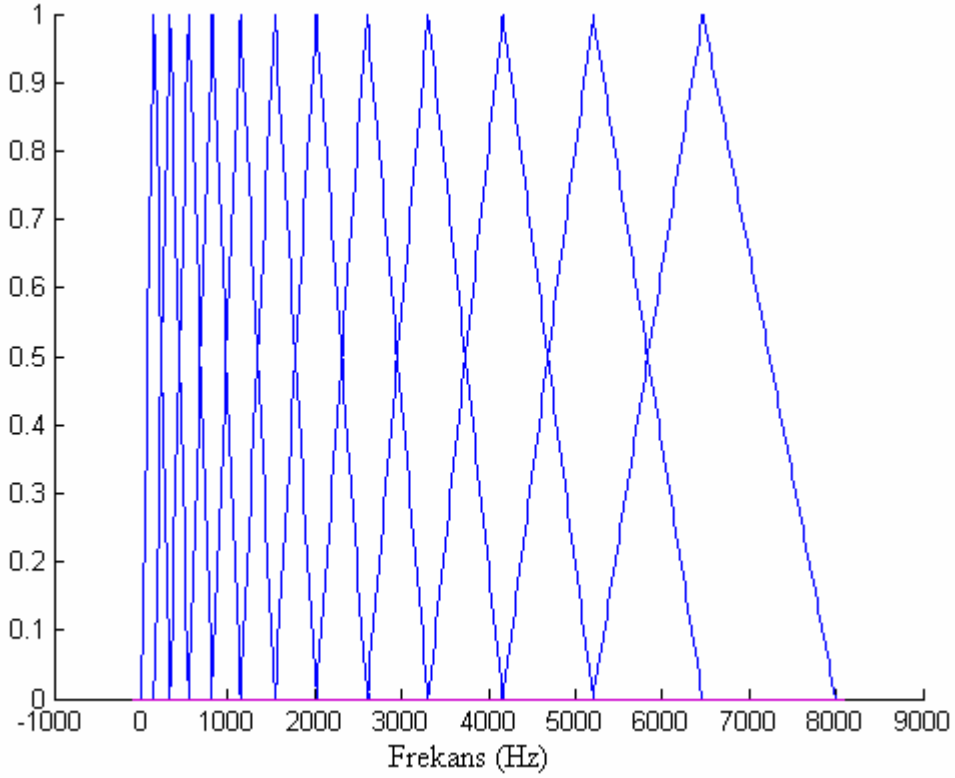
Normal bir insan kulağı frekansları doğrusal olmayan bir şekilde algılar. Araştırmalara göre ölçüler, 1 kHz'e kadar doğrusal olarak, daha yüksek değerlerde ise logaritmiktir olarak artmaktadır [10,11]. İnsan kulağının frekans cevabını karakterize eden bu ölçüye Melodi Ölçüsü denir. Bir frekansı mel-frekansına çevirmek için aşağıdaki formül kullanılır [11]:

$$m(f) = 2595 * \log (1 + f / 700) \quad (4.3)$$

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

Konuşma tanımanın bu aşamasında bant geçiren süzgeç (band-pass filter) kullanılır. Bu filtre bankı üçgen bant geçiren özelliktedir. Bu filtrenin yayılış özelliğinden dolayı band genişliği sabit bir mel frekans aralığına sahiptir. Bu filtre bankı, sinyalin HFD ile frekans domenine çevrilmiş haline uygulanır.

Bu mel dönüştürücü filtre bankındaki her bir filtre ile frekans domeninde bir histogram kutusu (kutuların birbiri üzerine binmesi) oluşturulur. Şekil 4.2'de mel yayımlı filtre bankası gösterilmiştir.



Şekil4.2.Mel yayımlı filtre bankası [12]

4.1.1.5. Mel Frekans Kepstrum Katsayıları (MFKK)

Öznitelik çıkarmanın son aşamasında logaritması alınmış mel spektrumunun frekans domeninden tekrar zaman domenine çevrilmesi gerekmektedir. Bu işlemin sonucunda Mel-Frekans Kepstral Katsayıları (MFKK) elde edilir. Her çerçeve için elde edilen vektörlere de öznitelik vektörleri denir. MFKK aşağıdaki gibi hesaplanır:

$$mfkk_i = \sum_{k=1}^K (X_k) \cos\left[i\left(k - \frac{1}{2}\right)\frac{\pi}{2}\right] \quad i=1,2, \dots, K \quad (4.4.)$$

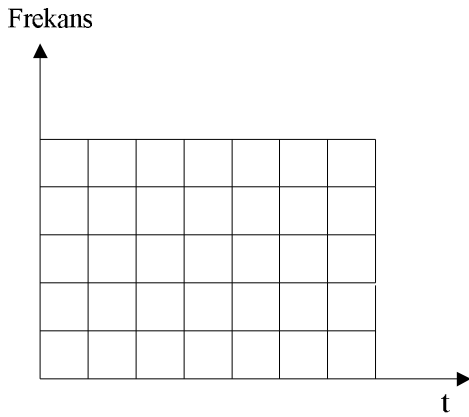
K, kepstrum katsayılarının sayısını, X_k , $i = 1,2,\dots,K$, ise k. filtrenin log-enerji çıktısını göstermektedir.

4.2. Dalgacık Dönüşümü

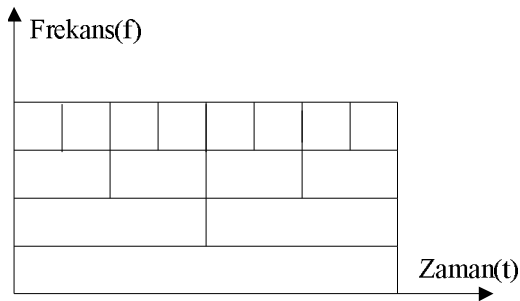
Dalgacık Dönüşümü (DD) ilk defa jeofizik alanında geliştirilmiştir [13]. Daha sonra yapılan çalışmalarla bu teori dahada sağlamlaştırılarak ayırık zamanlı işaretlere uyarlanması yapılmıştır. Dalgacık Paket dönüşümü(DPD), ayırık dalgacık dönüşümünün geliştirilmiş bir hali de Coifman ve Wickerhauser tarafından önerilmiştir. Bugün dalgacıklar ve dalgacık paketleri işaret işlemenin farklı alanlarında kullanılmaktadır [13].

DD durağan olmayan sinyallerin analizini sağlar. Zaman-frekans çözünürlüğü en uygun olacak şekilde elde edilir. Analizi yapılacak fonksiyon istenildiği gibi seçilebilir. DD ve Dalgacık analizleri fourier dönüşümünün genelleştirilmiş halidir.

DD’de, sinyal sürecinde kaydırılan ölçeklenebilir modüle edilmiş pencereler kullanılır ve her yeni konumda onun spektral davranışı incelenir. Farklı çözünürlüğe sahip sinyallerin frekans-zaman sunumları gerçekleştirilmiş olur. Aslında pencerelenmiş fourier analizinde Şekil 4.3 de görüldüğü gibi zaman frekans seviyeleri için sabit pencereler kullanılır[14].



Şekil 4.3.Pencerelenmiş Fourier dönüşümü için zaman-frekans diyagramı



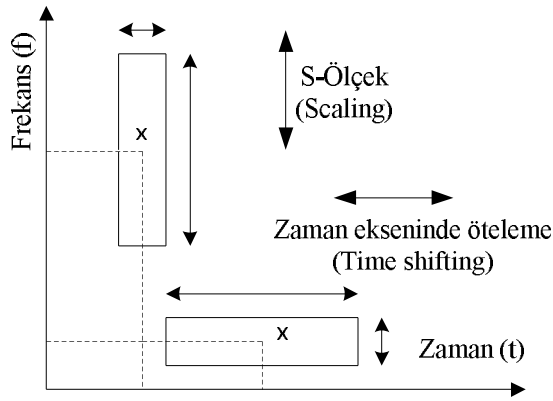
Şekil 4.4 Dalgacık dönüşümü için zaman-frekans diyagramı

Fourier dönüşümünde eşit aralıklarla işlem yapılırken DD alçak frekanslarda geniş zaman aralığında, yüksek frekanslarda ise küçük zaman aralığında işlem yapılmaktadır. Şekil 4.4’de DD için zaman-frekans diyagramı gösterilmiştir. Yani dönüşümünde boyutları değişik pencereler zaman-frekans domeninde kullanılmaktadır [14].

Fourier dönüşümü bize frekans domeninde bilgi verirken zaman domeninde kaybolmalara neden olmaktadır. Bu problemin nedeni frekansların, zaman içerisindeki konumlarıyla, frekans bölgesindeki özelliklerini birleştirebilme yeteneği olmamasıdır. Bunun neticesinde frekans bölgesindeki en ufak değişiklik tüm zaman boyunca

değişikliğe sebep olmaktadır. Fourier Dönüşümünün tersine, DD ana dalgacığın kaydırılmasıyla zaman bölgesinde, ana dalgacığın ölçeklendirilmesiyle de frekans bölgesinde yerinin belirlmesine izin verir. DD’de zaman bölgesinde kaydırma ve frekans domeninde ölçeklendirme işlemi Şekil 4.5’de gösterilmiştir. Ana dalgacığa uygulanan kaydırma ve ölçekleme işlemleri, sinyalin yerel bölgeleri ile dalgacık arasındaki karşılıklı ilişkiyi temsil edecek olan katsayıları belirlemede kullanılır.

Zaman-frekans diyagramı yardımıyla bir sinyalin yüksek ve alçak geçiren bileşenleri açık olarak elde edilir. İki boyutlu diyagramda bir eksen zaman (t) diğer eksen ise frekans (f)’dir.



Şekil4.5 Zaman-frekans diyagramı

Dalgacık analiz metotlarının geleneksel metotlara göre üstünlükleri aşağıdaki gibi sıralanabilir:

- Spektrumun düşük frekans bölümlerinde, DD oldukça süratlidir. Bu özelliklerinden ötürü, DD’ nin durağan olmayan işaretlerin esnek zaman-frekans gösterimlerinin elde edilmesinde etkili bir yöntem olduğu görülmektedir [14].
- Frekans spektrumundaki farklı bölgeler için, daha basit ve daha farklı frekans çözünürlükleri seçebilir.
- Analiz için spektrumdaki birkaç frekans bandı kullanılacaksa, tüm spektrumu hesaplamaya gerek yoktur.

4.2.1.Sürekli Dalgacık Dönüşümü

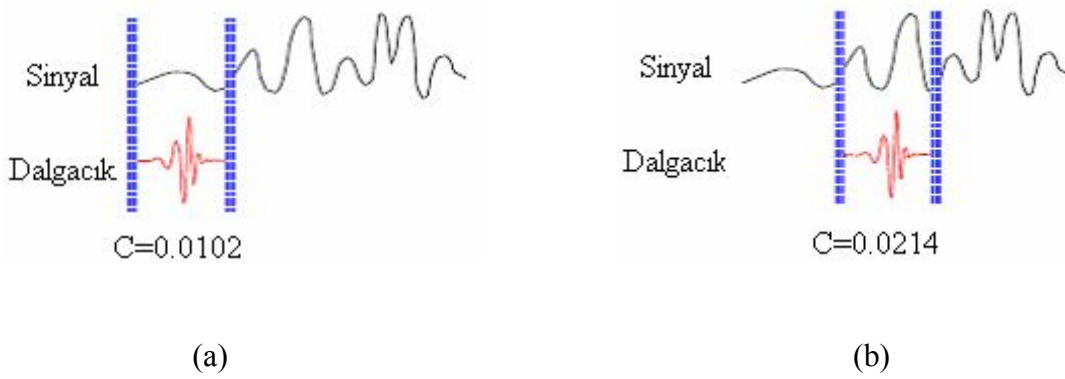
Sürekli dalgacık dönüşümü (SDD), sinyalin toplam zamanı ile dalgacık fonksiyonunun kaydırılmış ve ölçeklenmiş versiyonlarının çarpılması olarak tanımlanır.

Denklem 4.5’ da verilen fonksiyonda SDD görülmektedir. Bu fonksiyonda, $x(t)$ sinyalinin, zamanda kaydırılan ve ölçeklendirilen ψ dalgacığı ile çarpılması görülmektedir. Her ölçeği değiştirme işlemi tespit edilecek frekans değerini belirlemek için, her zamanda kaydırma işlemi de sinyalin farklı bir bölgesini incelemek için kullanılır [13].

$$SDD(a,b) = - \int_{-\infty}^{\infty} x(t) \frac{1}{\sqrt{a}} \psi\left(\frac{t-a}{b}\right) dt \quad (4.5)$$

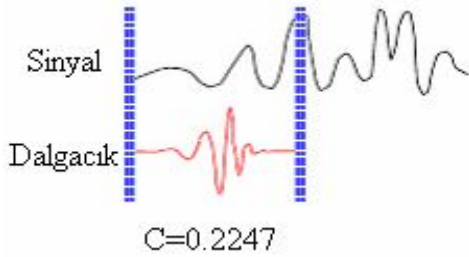
Ölçeği belirleyen her b değerinde dalgacık, zamanı belirleyen a değeri ile zamanda kaydırılır. Bu işlem diğer ölçeklerde de devam edilerek dalgacık katsayıları elde edilir.

Analiz neticesinde oluşan değerlere dalgacık katsayıları denir ve seçilen dalgacık tipi ile orijinal sinyalin 4.6. denklemi sonucundaki üretilen değerleridir. Şekil 4.6 bir sinyale SDD uygulamak için, sinyalin üzerinde dalgacığın kaydırılması, Şekil 4.7 da dalgacığın ölçeklendirilmesi verilmiştir.



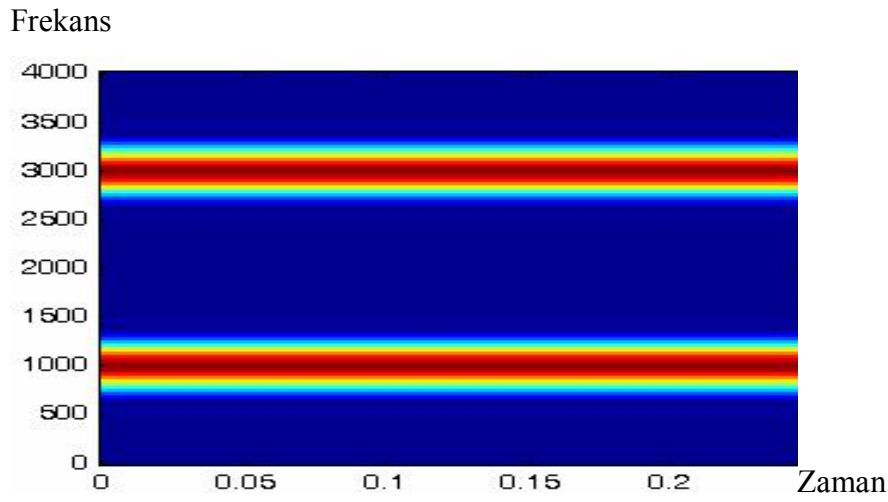
Sekil 4.6. Dalgacığın sinyal üzerinde kaydırılması

(a) Başlangıç durumu (b) Sağa kaydırılmış durum

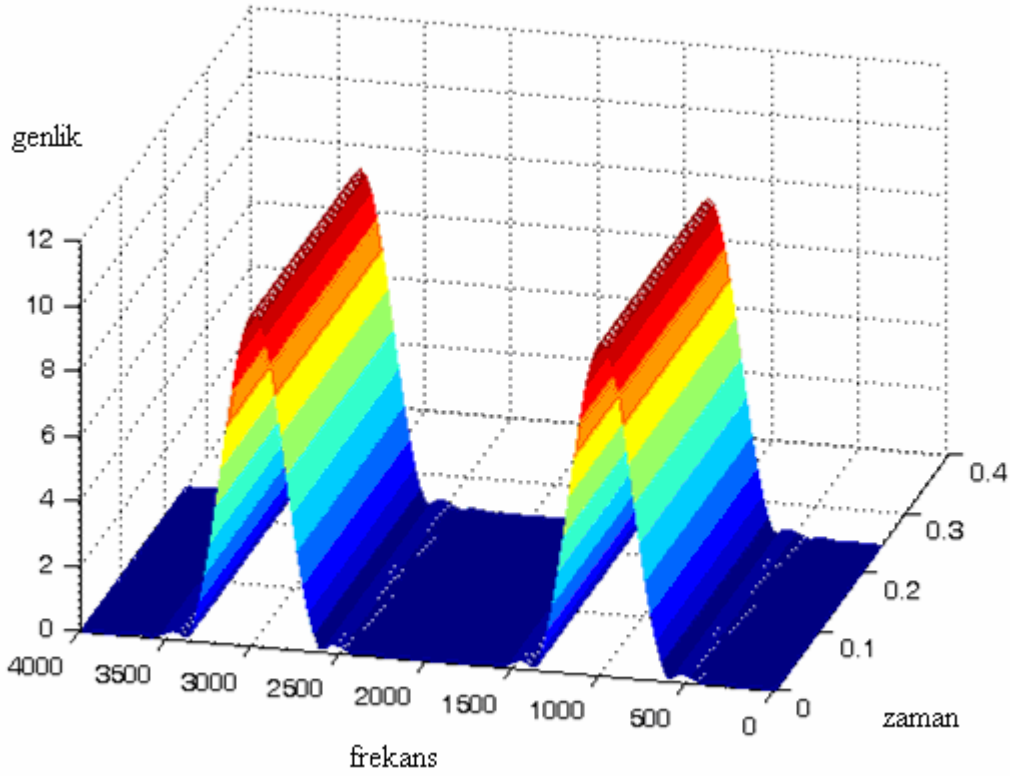


Sekil 4.7. Ölçeğin değiştirilmesi

SDD' nün genliği skolagram olarak adlandırılır. Skolagram sayesinde SDD' nün iki boyutlu veya üç boyutlu grafikleri çizilebilir [14]. 1000 Hz ve 3000 Hz frekans değerlerine sahip iki ayrı sinyalin toplamı sonucunda oluşan sinyalin iki boyutlu spektogramı Şekil 4.8' de verilmiştir. Yatay eksen zaman, düşey eksen de frekans bilgisini vermektedir. Sinyalin genliği ise renklerle belirlenmektedir. Şekil 4.9 aynı sinyalin üç boyutlu grafiğini göstermektedir. Şekillerden de anlaşılacağı üzere, sinyal içerisindeki frekans bileşenleri, zaman ve genlik bilgileri ile birlikte alınabilmektedir. [13]



Şekil 4.8. Sinyal spektogramı



Şekil 4.9. Sinyalin üç boyutlu gösterimi [13]

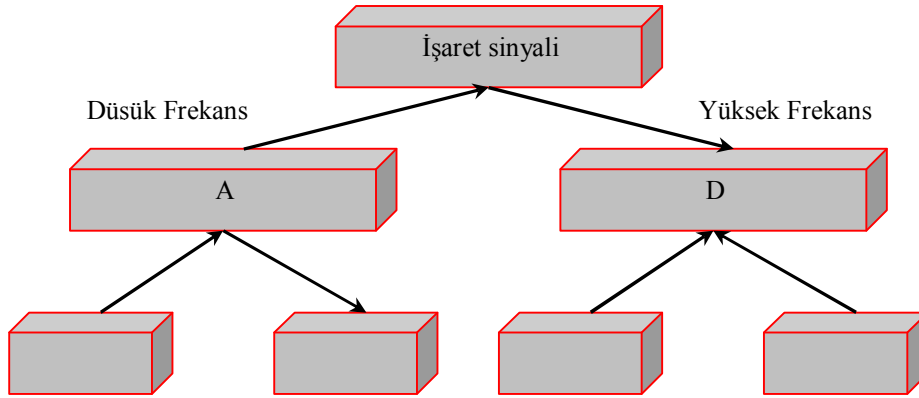
4.2.2. Ayrık Dalgacık Dönüşümü

SDD'nin hesaplama yükü fazladır. Dönüşümü sağlamak için kullandığı bilgi miktarı oldukça büyüktür. Daha etkili bir yol olan Ayrık Dalgacık Dönüşümü (ADD), ana dalgacığın sadece belirli kısımlarında işarete bakar. ADD'yi anlamak ve dalgacıkların yerini görmek için filtre bankası kavramını iyi bir şekilde tanıtmak gerekir. Çünkü filtre bankası, dalgacıkların ayrık eşdeğerlerini oluşturur [13].

Tek bir filtre belirli bir frekans cevabına sahiptir. İşaret sinyali filtreden geçtikten sonra işaretin içindeki bilginin bir bölümü kaybolur. Bu sebeple tek bir filtre, filtrelenmiş işareten tekrar asıl işareti elde etmek için kullanılamaz. Çünkü yitirilen bilgi tekrar elde edilemez. Eğer iki filtre kullanılırsa, biri alçak frekans bilgisini, diğeri ise yüksek frekans bilgisini tutar. Yani işaret içindeki bilginin tamamını içermiş olur. Bunun sonucunda bu iki filtrenin çıkışı asıl işaret sinyalini tekrar elde etmek için birleştirilebilir. Filtre kümesi veya filtre bankası, spektrum bilgisini ayırmak için kullanılarak işareti gittikçe daha ince frekans bantlarına ayrıştırmayı sağlar. Filtrelerin çıkışından elde edilen veri örneklenmelidir. Çünkü işaret, filtre bankasının her bir seviyesinden geçirildiği zaman verinin miktarı iki katına çıkmaktadır.[13]

bileşenlerini tam olarak ayırtırmak yerine, yüksek frekans bileşenlerini daha iyi ayırtırmaya olanak sağlar. Böylece tüm zaman-frekans düzlemi Şekil 4.11’de görüldüğü gibi mümkün olabilecek tüm alt bölümlere, farklı zaman–frekans pencereleri ile bölünür.

DPA’nın sağladığı avantaj, asıl işaretin en uygun ZFG’ sini elde etmek için ayrışımın farklı seviyelerini birleştirmesidir.



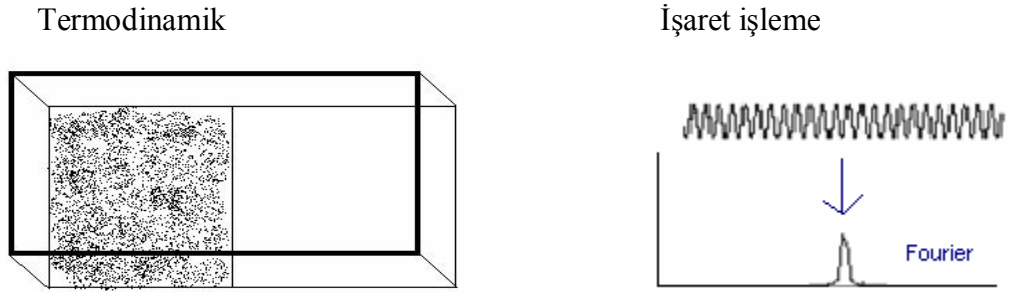
Şekil 4.11. DPA ağaç yapısı

4.2.4. Entropi Hesaplama Teknikleri

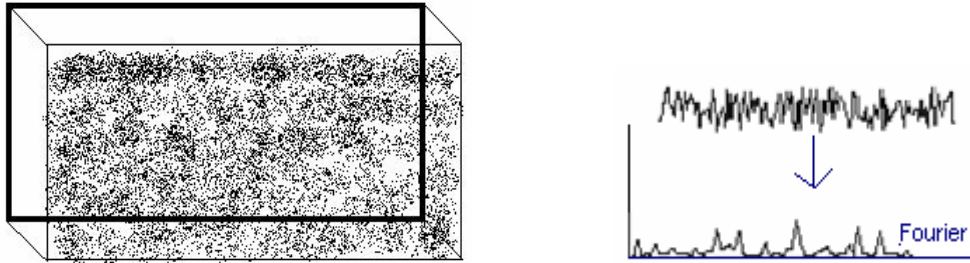
Bir sistemin düzensizliğinin ölçüsü olarak, entropi kavramı tanımlanabilir. Özellikle, termodinamik ve fizikte bir sistemin düzenliliğini ölçmek için en fazla kullanılan bir terimdir. Bu tanımlamada düzensizlik kelimesi ile anlatılmak istenen ifade, sistemin toplam enerjisinin mevcut olan tanecikler arasındaki dağılımıdır. Entropi hesaplama tekniği, durağan olmayan bir sinyalin düzensizlik derecesini ölçmek için kullanılan uygun bir tekniktir [13]. Ayrıca entropi kavramı, bir olayın içerdiği bilginin ortalama miktarını ölçmek amacı ile de kullanılmaktadır [13].

Şekil 4.12 ve Şekil 4.13’de entropi kavramının anlaşılabilmesine yönelik olarak, termodinamik ve işaret işleme alanlarındaki kullanımı karşılaştırmalı olarak gösterilmiştir. Şekil 4.12’den görüleceği gibi gaz molekülleri belirli bir alana sıkıştırılmıştır. Buradaki gaz molekülleri daha düzenli bir yapıda oldukları için düşük entropi değerine sahiptirler. Şekil 4.13’de ise gaz molekülleri sıkıştırıldığı bölgeden serbest bırakılarak tüm kapalı alana yayılmışlardır. Burada ise gaz moleküllerinin düzenli yapısı bozulduğu için gaz molekülleri yüksek entropi değerine sahiptirler.

Termodinamikte kullanılan bu kavram, iletişim teorisinde bir iletideki düzensizliđi veya belirsizliđi ölçmek için yararlanabilecek iyi bir araçtır. İletide işleme açısından Şekil 4.12 ve Şekil 4.13 incelendiğinde, iletide durağan ise frekans bölgesinde dar bir spektrum ile düşük bir entropi değeri elde edilir (Şekil 4.12). İletide durağan deđilse, frekans spektrumu geniş bir banda yayılarak yüksek bir entropi değeri oluşturur[13]. (Şekil 4.13) .



Şekil 4.12. Düşükte entropi kavramı



Şekil 4.13. Yüksek entropi kavramı

Entropi ile ilgili ilk çalışmalar, termodinamik alanında 19. yüzyıl da yapılmıştır. 1940'lı yıllarda Shannon tarafından haberleşme kanallarının doluluđunu tahmin etmek ve haberleşme iletidelerinin içerdiiđi bilgi miktarını ölçmek için entropi kavramı kullanılmıştır [13]. Son yıllarda ise entropi kavramı iletide işleme alanında da yaygın olarak kullanılır hale gelmiştir. Yapılan çalışmalardan bazıları; entegre devre teknolojisinde karbon üzerindeki elektron enerji kaybını bulmak [13], JPEG görüntü kodlamada oluşabilecek gürültüyü tespit etmek [13], EEG iletidelerinin spektral

entropilerini kullanarak anestezi derinliğini tahmin etmek [13], biyomedikal işaretlerinin dalgacık dönüşüm uzayından istatistiksel bilgilerini elde ederek işaretin gürültüsüz şeklini kestirmek [15] ve biyomedikal görüntülerin yeniden yapılandırılmasında Norm entropi metodu kullanarak yüksek kalitede MEG görüntüleri elde etmek [15] sayılabilir.

İşaret işlemede yaygın olarak kullanılan entropi hesaplama çeşitleri; Shannon, Eşik (Threshold), Logaritmik Enerji, Norm ve Sure metotlarıdır. Bunun dışında farklı entropi hesaplama teknikleri de bulunmaktadır. Aşağıdaki entropi yöntemlerinde; s işareti, s_i ise işaretin i . katsayısını göstermektedir. Tablo 4.1’de işaret işlemede kullanılan entropi türleri ve denklemleri verilmiştir.

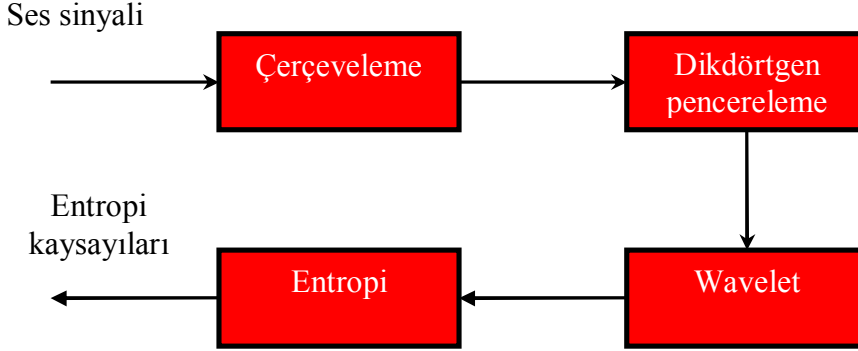
Tablo 4.1. İşaret işlemede yaygın kullanılan entropi türleri ve denklemleri

Entropi Türleri	Entropi Denklemleri
1. Shannon Entropi	$E(s) = -\sum_i s_i^2 \cdot \log_2(s_i^2)$ ve $\log(0)=0$ kabul edilerek
2. Norm Entropi	$E(s) = \sum_i s_i ^p$ ve $1 \leq p < 2$
3. Logaritmik Enerji	$E(s) = \sum_i \log_2(s_i^2)$ ve $\log(0)=0$ kabul edilerek $E(s) = \sum_i E(s_i)$ ε Pozitif bir eşik değeri olup
4. Eşik Entropi	$ s_i > \varepsilon \Rightarrow E(s_i) = 1$ ve $ s_i \leq \varepsilon \Rightarrow E(s_i) = 0$
5. Sure Entropi	$ s_i \leq \varepsilon \Rightarrow E(s_i) = \sum_i \min(s_i^2, \varepsilon^2)$ burada ε pozitif bir eşik değeri

Belirtilen entropi hesaplama tekniklerinin, işaret işlemedeki kesin kullanım alanları belirli olmayıp, uygulamalara göre başarımları değişebilmektedir. Fakat durağan olmayan işaretler için zamana bağımlı entropi hesaplaması daha kullanışlı olacaktır. Bu sebeple işaret katsayıları s_i ’lerin zamana bağımlı olması gerekir [13].

4.2.5.Dalgacık (Wavelet) Dönüşümü ile Entropi Hesaplama

DD ile entropi katsayıları hesaplama işlemi hem kaydolma hem de saptama aşamasında kullanılır. Entropi katsayılarının elde edilme aşamaları Şekil 4.14’de verilmiştir.



Şekil 4.14. Entropi katsayıları çıkarma

Konuşma ve konuşmacı tanıma sistemlerinde kullanılan spektral ölçümleri iki bölüme ayırabiliriz. Birincisi güç, sinyalin sahip olduğu kaba spektral ölçümleri, ikincisi ise spektral genlik, spektrumdaki belirli frekans aralıkları üzerindeki güç ölçümüdür [13]. Konuşma ve konuşmacı tanıma sistemlerinin parametre kümesinde bu iki ölçüm yer alır. Konuşma tanıma sistemlerinde güç ölçülerinin kullanımı denklem 4.6’da verilmiştir.

$$P(n) = \frac{1}{N_s} \sum_{m=0}^{N_s-1} (w(m)s(n - \frac{N_s}{2} + m)),$$

(4.6)

N gücü hesaplanacak örnek sayısını, $w(m)$ ağırlık fonksiyonunu, $s(n)$ sinyali ve n pencerenin merkezini belirleyen örnek indisini gösterir. Pek çok tanıma sistemi gücü direk olarak kullanmak yerine insan duyma sistemini taklit ettiği için denklem 4.7’yi kullanır.

$$P_{db}(n) = \log_{10}(P(n) * 10)$$

(4.7)

Güç, konuşma ve konuşmacı tanıma sistemlerinde bulunan diğer parametreler gibi çerçeve temelinde hesaplanır. Çerçeve süresi T_f , parametreler kümesinin geçerli olduğu

zaman uzunluđu olarak tanımlanır. Çerçeve periyodu ise benzer şekilde arka arkaya yapılan parametre hesaplamaları arasında geçen süredir. Çerçeve oranı ise saniyede hesaplanan çerçevelerin sayısıdır (Hz). Aynı zamanda önemli olan bir kavram da gücün hesaplandığı aralıktır. Toplamın hesaplanacağı örnek sayısı olan N_s pencere süresi olarak bilinir. Pencere süresi ile çerçeve süresi birlikte güç hesabındaki sinyalin değişimini izlerler.

Ağırlık fonksiyonu, pencere fonksiyonu olarak ta adlandırılır. Pencere fonksiyonunda amaç eldeki sinyali belli aralıklarla sınırlamaya çalışmaktır. Örneğin aralık içerisinde sabit bir değeri alan ve aralık dışında da sıfır olan fonksiyon dikdörtgen pencere fonksiyonudur. Herhangi bir fonksiyon veya sinyal bu pencere ile çarpıldığında aralık dışındaki değeri sıfır olur. Sesli ifade verilerinin önişleme aşamasında kullanılan pencereler aşağıda yer almaktadır.

- Dikdörtgen Pencere
- Barlett Pencere
- Hanning Pencere
- Hamming Pencere
- Blackman Pencere
- Kaiser Pencere.

5. AKILLI SİSTEMLER

5.1. Yapay Sinir Ağları

Yapay Sinir Ağları (YSA) kavramının temeli, biyolojik sinir sistemine dayanmaktadır. Yapay sinir ağları, insan beynindeki sinir hücrelerinin çalışmasını taklit eder. İlk sinir hücresi (nöron), 1943’de nöropsikolog Warren McCulloch ve mantıkçı Walter Pitts tarafından, üretilmiştir [16]. Tek katmanlı bir algılayıcı olan, McCulloch ve Pitts (MCP) modeli yapay sinir ağlarının temelini oluşturur. Minsky ve Papert tarafından yazılan ve tek katmanlı algılayıcıların kısıtlılıklarını anlatan kitap pek çok araştırmacının bu konuya ilgilerini yitirmelerine sebep olmuştur [17]. 1980li yılların başında hata geriye yayma yönteminin bulunmasına kadar, çok az araştırmacı çalışmalarına devam etmiştir. Hopfield tarafından 1982 yılında yapılan çalışmalardan sonra, yapay sinir ağları uygulamaları tekrar ortaya çıkmaya başlamış ve o zamandan bu yana endüstriyel, ticari ve bilimsel uygulamalarda başarı ile kullanılmıştır [16].

Bir yapay sinir ağı, birbirine hiyerarşik olarak bağlı ve birbiri ile ağırlıklı bağlantılar vasıtası ile haberleşen bir grup işleme biriminden (nöron) oluşur. Bu bağlantıların ağırlıkları, ön bilgiler kullanılarak ayarlanabilir veya belirli bir öğrenme kuralına göre değişebilecek şekilde eğitilerek belirlenir. Bu durum, Rumelhart [18] tarafından yapılan tanımlamada şu şekilde ifade edilmiştir: “yapay sinir ağlarında öğrenme problemi ağıın istenilen işlemi yapmasına olanak verecek bağlantı güçlerini bulmaktan ibarettir”. Yapay sinir ağları, gerçek değerli, ayrık değerli ve vektör değerli fonksiyonların öğrenilmesinde genel ve pratik bir yöntem sağlamaktadır. Bu çalışmada, performans değerlendirme alanında önemli bilgiler sağlayabilecek, işlem ve haberleşme sürelerine ait verilerin tahminde kullanılacak yapay sinir ağı modelleri önerilmektedir.

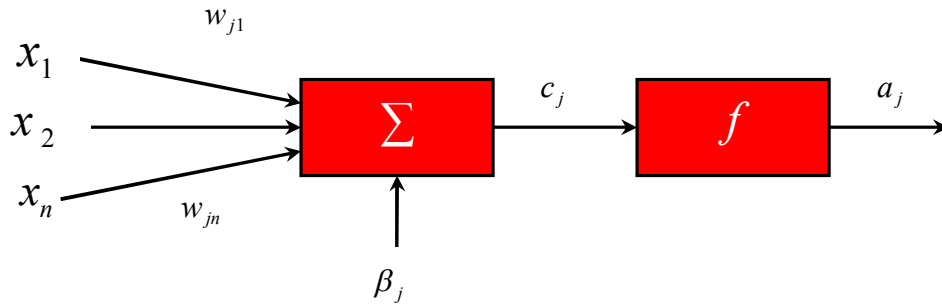
5.1.1 Temel Yapay Sinir Ağı Kavramları

Yapay sinir ağları, bilgiyi insan beynine benzer bir şekilde işler ve örnek yolu ile öğrenir. Bir yapay sinir ağı, çok sayıda birbiri ile bağlı, belirli bir problemi çözmek için paralel çalışan işlem elemanlarından (nöron) oluşur. Yapay sinir ağı modelleri, kullanılan topoloji, öğrenme stratejisi (öğretmenli, öğretmensiz vb.) ve öğrenme algoritması ile tanımlanır [19].

5.1.2 Yapay Sinir Hücresi (Nöron)

Yapay sinir hücresi, komşularından veya dış kaynaklardan çeşitli girdileri alıp bunları diğer birimlere yayan ve belirli bir çıktıyı hesaplamak için kullanan işleme birimidir. Şekil 5.1’de $x_1, x_2, \dots, x_n$ şeklinde n adet girdisi olan bir yapay sinir hücresi gösterilmiştir. Girişlerin her biri $w_{j1}, w_{j2}, \dots, w_{jn}$ ağırlıklarındadır. Toplama fonksiyonu Σ , hücreye gelen net girdiyi hesaplar. Ağırlıklı toplama fonksiyonu kullanıldığında, a nöronunun çıktısı şu şekilde ifade edilebilir[16].

$$a_j = f(c_j) \quad (5.1)$$



Şekil 5.1. n adet girdisi olan basit bir nöron

Bu durumda C_j denklemi denklem 5.2. deki gibidir.

$$c_j = \sum_{i=1}^n w_{ji} \prod_{n=1}^m x_{in} + \beta_j \quad (5.2)$$

5.1.3 Aktivasyon Fonksiyonu

Her bir yapay sinir hücresi, girdilerinin adedi dışında f aktivasyon fonksiyonu ile tanımlanır. Bu fonksiyon, hücreye gelen girdiyi işleyerek, hücrenin bu girdiye karşılık üreteceği çıktıyı belirler. Toplama fonksiyonunda olduğu gibi, aktivasyon fonksiyonu olarak da farklı fonksiyonlar kullanılmaktadır. Ağdaki işlem elemanlarının tamamının aynı aktivasyon fonksiyonunu kullanması da şart değildir.

5.1.4 Katmanlar

Temel olarak, tüm yapay sinir ağıları benzer bir yapıya sahiptir. Bu yapı içinde, bazı hücreler dış dünyadan girdileri alırken diğerleri dışarıya çıktıları vermekle görevlidir. Yapay sinir ağıları, rasgele bağlanabilecek bir grup hücreden ibaret değildir. Bir yapı oluşturmanın en kolay yolu bir grup elemandan oluşan katmanlar oluşturulmasıdır. Bu katmanlar arasındaki bağlantılar, toplama ve aktivasyon fonksiyonları işleyen bir yapay sinir ağını meydana getirir. Sadece tek katmandan oluşan, kullanışlı sinir ağıları var olmakla birlikte, çoğu uygulama için sinir hücreleri, 3 katman halinde ve her katman içinde paralel olarak bir araya gelerek ağı oluştururlar [16].

Bir yapay sinir ağında, en az bir girdi katmanı, sıfır veya daha fazla gizli (ara) katman ve bir çıktı katmanı bulunur. Girdi katmanı dışındaki tüm katmanlar, sinir hücreleri veya nöronlardan oluşur.[16]

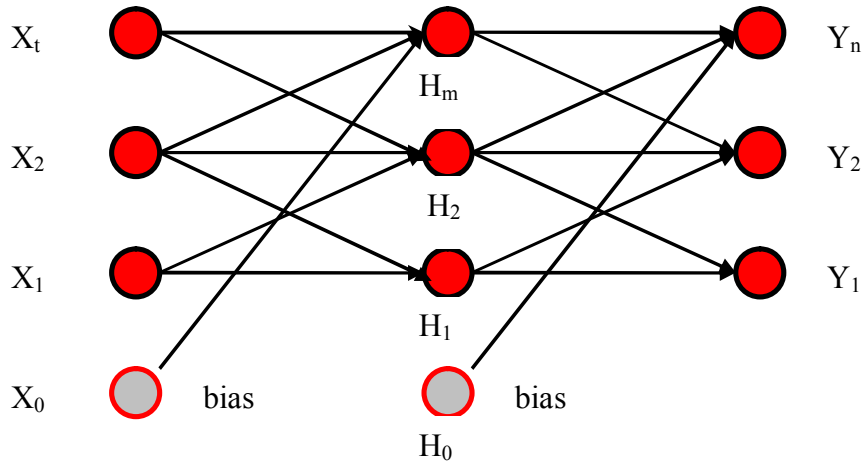
Girdi katmanı, verileri ya girdi dosyalarından ya da gerçek zamanlı uygulamalardan alır. Çıktı katmanı ise bilgiyi dış dünyaya, ikinci bir bilgisayar sürecine veya elektromekanik kontrol sistemlerinde olduğu gibi diğer cihazlara gönderir. Bu iki katman arasında pek çok gizli katman olabilir, bu ara katmanlar çeşitli şekillerde bağlanmış pek çok sinir hücresinden oluşur. Yapay sinir ağının girdi katmanının boyutu, sahip olunan girdi örneklerinin boyutuna eşittir. İstenen çıkışların sayısı ise, çıktı katmanındaki nöronların sayısını belirler. Genel olarak, yapay sinir ağındaki gizli katmanların sayısı ve her gizli katmandaki nöron sayısı ağın karmaşık fonksiyonları benzetme oranını belirler. Ancak bu, karmaşık ağların her zaman daha başarılı olacağı anlamına gelmez. Ağın temsil gücü ile hesaba katacağı gürültü arasında bir uyum vardır. Ağ ne kadar karmaşık hale gelirse, gürültüye karşı da o kadar hassas olur ve temeldeki fonksiyon ile birlikte girdilerdeki gürültünün de öğrenilmesi kolaylaşır [16].

5.1.5 Yapay Sinir Ağı Topolojileri

Tüm yapay sinir ağıları; nöronlar, bağlantılar, toplama ve aktivasyon fonksiyonu kavramlarına dayandığı için, farklı tipteki ağların topolojileri bile birbirine benzer. Çeşitlemelerin çoğu, farklı öğrenme kuralları ve bu kuralların ağın tipik topolojisini değiştirme biçimi ile ortaya çıkar. Yapay sinir ağı topolojileri, ileri beslemeli ve geri beslemeli (geri dönüşümlü) ağlar olmak üzere, bağlantıların şekillerine göre, iki ana grupta toplanabilir. Bu çalışmada ileri beslemeli topoloji modeline sahip yapay sinir ağı modeli kullanılmıştır [18].

5.1.6. İleri Beslemeli Ağlar

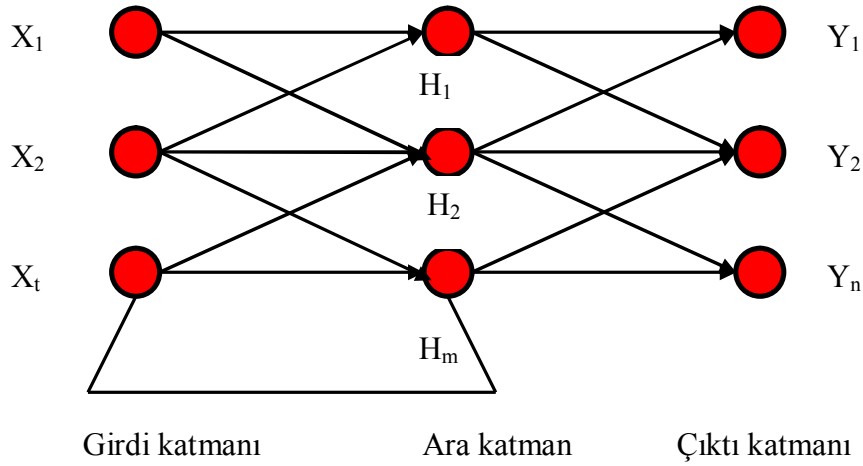
İleri beslemeli yapay sinir ağlarında, giriş birimlerinden çıkış birimlerine, veri akışı tamamen ileri doğrudur. Birimlerin çıkışlarından, aynı katmanda veya önceki katmanlarda yer alan nöronların girişlerine bağlantı yapılmasına izin verilmez. Bu topoloji, yapay sinir ağına problemle ilgili tüm bilgilerin bir kerede giriş olarak verilebileceği, sonucun (çıktının) tamamen o andaki girdi değerlerine bağlı olduğu durumlarda kullanılır. Şekil 5.2’de İleri beslemeli yapay sinir ağı topolojisi verilmiştir.



Şekil 5.2 İleri beslemeli yapay sinir ağı topolojisi

5.1.7. Geri Beslemeli Ağlar

Geri beslemeli ağlarda, önceki girişler ile ilgili bilgiler, ara veya çıkış katmanındaki nöronlardan geriye doğru yapılan bağlantılar aracılığıyla tekrarlanır ve girdilerle birleşir. (Şekil 5.3) Geri beslemeli ağlar, ağına verilecek girişlerin sırasının önemli olduğu, önceki girişlerin kaydının bir şekilde saklanıp yeni girişler ile etkileşime sokulması istenen durumlarda kullanılır. Dinamik sistemlerin modellenmesinde ve öğrenilmesinde geri dönüşümlerin olması, özellikle zaman gecikmelerini dikkate almak için önemlidir.



Şekil 5.3 Geri beslemeli yapay sinir ağı topolojisi

Geri dönüşümlü ağlar, tam geri dönüşümlü ve kısmi geri dönüşümlü olmak üzere ikiye ayrılır.

Tam geri dönüşümlü ağlarda, hepsi eğitilebilir olan, ileriye ve geriye doğru rasgele bağlantılar vardır. Bazı durumlarda, birimlerin aktivasyon değerleri bir dinlenme sürecine girer ve ağ artık aktivasyonların değişmediği kararlı bir duruma geçer. Ancak, ağın aynı cevapları veren durumları tekrarlayarak ya da her seferinde farklı bir çıktı üreterek kararlı bir duruma gelememesi de söz konusudur. Bağlantıların ağırlıklarına bazı sınırlamalar getirilerek ağın kararlı duruma geçmesi sağlanabilir. Tam geri dönüşümlü ağlar, daha çok en uygun şekle sokma problemlerinde (optimizasyon), çağrışımsal bellek gibi kullanılır.

Kısmi geri dönüşümlü ağlarda, ağın işlem elemanlarına ek olarak, ara katman elemanlarının geçmiş durumlarını hatırlamak için kullanılan içerik elemanları bulunmaktadır. Geri besleme sadece içerik elemanları üzerinde yapılır ve bu bağlantılar eğitilemezler. Ancak ağın çıktıları, hem önceki durumlara hem de o anki durumlara bağlı olarak oluşmaktadır. Geçmiş durumları hatırlayabilmeleri nedeni ile bu ağların dinamik bir belleğe sahip olma özelliği vardır. İleri beslemeli ağların aksine, geri dönüşümlü ağların dinamik özellikleri önemlidir. Kısmi geri dönüşümlü yapay sinir ağları sınıfında iki önemli mimari vardır. Elman [20] ara katmandaki nöronlardan, ek girdi olarak kullanılacak bir grup içerik elemanlarına geri besleme yapılmasını önermiştir. Jordan [20] ise, çıkış elemanlarından içerik elemanlarına geri besleme yapan bir ağ tanımlamıştır. Kısmi geri dönüşümlü ağlar, popüler geriye yayma algoritmasının

da kullanılmasına olanak verdiğinden, ileri beslemeli ağların basitliği ile tam geri dönüşümlü ağların karmaşıklığı arasında bir noktayı temsil eder.

5.1.8. Yapay Sinir Ağlarının Eğitilmesi

Yapay sinir ağı modellerinin gücü, bağlantı ağırlıklarının ayarlanma şekliyle bire bir ilişkilidir. Bu ağırlıkların, belirli bir veri kümesine göre, ayarlanması süreci ağı eğitilmesi olarak tanımlanır. Bu şekilde kullanılan veri kümesi ise eğitim kümesi olarak adlandırılır. Eğitimin temel amacı, ağı eğitimi kümesinde var olan birtakım modelleri öğrenecek şekilde bağlantı ağırlıklarının ayarlanmasıdır. Bu şekilde ayarlanan ağ, gelecekteki durumları (bilinmeyen verileri) tahmin etme yeteneğine sahip olacaktır. Yapay sinir ağlarının eğitiminde, öğretmenli, öğretmensiz, destekleyici veya karma stratejiler kullanılır.

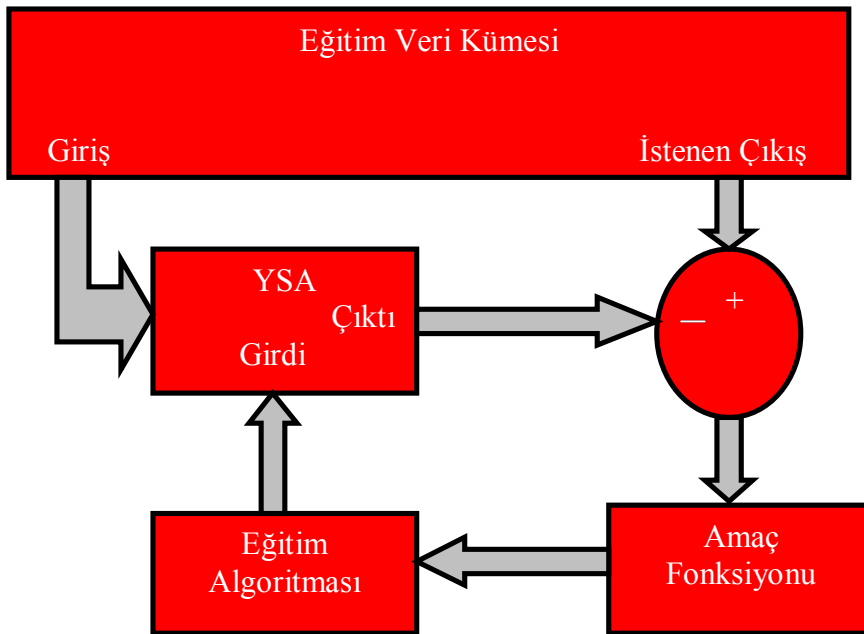
5.1.9. Öğrenme Stratejileri

Öğretmenli öğrenmede, eğitim kümesi olarak hem girdiler hem de beklenen çıktılar ağı gösterilir. Ağ, girdileri işledikten sonra bulduğu çıktıları istenen çıktılar ile karşılaştırır. Hatalar, sistemde geri yayılarak ağı kontrol eden bağlantı ağırlıkları yeniden ayarlanır. Bir yapay sinir ağının eğitimi sırasında aynı veriler defalarca işlenerek bağlantı ağırlıkları düzeltilir. Öğretmenli öğrenme, pek çok yapay sinir ağı modelinde en yaygın olarak kullanılan stratejidir. Öğretmensiz öğrenmede ise, eğitim kümesi olarak girdiler ağı gösterilir ancak istenen çıktılar verilmez. Sistem, verilen girdileri gruplamak ya da ilişkilendirmek için hangi özellikleri kullanacağına kendisi karar verir. Bu strateji genellikle, kendini örgütleme ya da uyarılma olarak isimlendirilir. Öğretmensiz öğrenme alanında önde gelen araştırmacılardan biri olarak, Kohonen ağının geliştiricisi, Tuevo Kohonen sayılır [17]. Kohonen ağı, tek katmanlı ve pek çok bağlantı içeren, kendini örgütleyen bir ağ modelidir.

Destekleyici öğrenme stratejisinde, öğrenen sisteme bir öğretmen yardımcı olur. Öğretmen, sisteme çıktıları göstermek yerine, sistemin kendi kendine çıktı üretmesini bekleyerek bu çıktıların doğru veya yanlış olduğunu gösteren bir sinyal üretir. Bu stratejiyi kullanan sistemlere örnek olarak öğrenmeli vektör kuantization (ÖVK) ağı gösterilebilir [18].

Kısmen öğretmenli, kısmen öğretmensiz olarak öğrenme yapan ağlar karma stratejili olarak sınıflandırılabilir. Bu tür ağlara örnek olarak radyal tabanlı ağlar (RTA) ve olasılık tabanlı ağlar (OTA) verilebilir.

Bir yapay sinir ağının, öğretmenli öğrenme stratejisi ile, nasıl eğitildiği Şekil 5.4’de gösterilmiştir. Eğitim kümesindeki veriler, girdi ve hedef olmak üzere iki kısımdan oluşur. Başlangıçta, ağ bağlantılarının ağırlıkları rasgele atanabilir. İlk örnek verinin girdi kısmı ağı gösterildiğinde, ağ bu başlangıç ağırlıklarına göre bir çıktı hesaplar. Bu hesaplanan çıktı, örnek verinin hedef çıktısı ile kıyaslanır ve aradaki fark seçilen bir algoritma tarafından ağırlıkları düzeltmek için kullanılır. Bu süreç, verilen x girdilerine karşılık hedeflenen t çıktıları veren bir $f(x)$ fonksiyonunun bulunmasına benzetilebilir.



Şekil 5.4 Bir yapay sinir ağının eğitilmesi işlemi

Yapay sinir ağlarının eğitiminde, tüm eğitim verilerinin bir kerede ağı gösterildiği toptan (batch) ve her devirde bir örneğin ağı gösterildiği artımlı (incremental) olmak üzere iki tarz vardır [18]. Aynı eğitim kümesinin tekrar işlenmek üzere ağı verilme sayısı ise devir sayısı (epoch) olarak adlandırılır.

5.1.10 Hata Fonksiyonu

Bir yapay sinir ağını eğitmek ve ne kadar başarılı olduğunu ölçebilmek için bir hata fonksiyonu tanımlanmalıdır. Ağ bağlantılarının ağırlıkları, bu hata fonksiyonunu

azaltacak yönde ayarlanmalıdır. En yaygın olarak kullanılan hata fonksiyonlarından biri olan ortalama karesel hata (okh) aşağıdaki gibi ifade edilebilir:

$$okh = \frac{1}{Q} \sum_{k=1}^Q e(k)^2 = \frac{1}{Q} \sum_{k=1}^Q (t(k) - a(k))^2 \quad (5.3)$$

p_Q ağırlık girdileri ve t_Q buna karşılık beklenen hedef çıktı olmak üzere, eğitim $[p_1.t_1].[p_2.t_2] \dots [p_Q.t_Q]$ şeklinde verilmiş olsun. Her girdi ağırlık gösterildiğinde, ağırlık hesapladığı Q_a çıktısı, hedef Q_t çıktısı ile kıyaslanır. Hata, hedef çıktı t ile ağırlık çıktısı arasındaki fark olarak hesaplanır. Bundan sonra ağırlık bağlantılarının ağırlıkları bu farkların

karelerinin toplamının ortalamasını küçültecek şekilde ayarlanır.

5.1.11. Öğrenme Kuralları

Yapay sinir ağlarının öğrenmesinde pek çok kural kullanılmakla birlikte, bunların çoğu, en iyi bilinen ve en eski öğrenme kuralı olan, Hebb kuralının çeşitlemeleridir [18]. Öğrenme kurallarının en önemlilerinden bir kısmı aşağıda açıklanmıştır.

Hebb Kuralı: Bu kural ilk olarak Donald Hebb tarafından 1949 yılında yazdığı kitabında tanımlanmıştır. Bu kurala göre eğer bir nöron başka bir nörondan bir bilgi alırsa ve bu nöronların her ikisi de aktif ise, yani matematiksel olarak aynı işareti taşıyor ise, bu iki nöron arasındaki bağlantı güçlendirilmelidir.

Hopfield Kuralı: Hebb kuralına benzer ancak, kuvvetlendirme veya zayıflatmanın miktarı da belirlenir. Bu kural, her iki hücre de aktif veya her iki hücre de pasif ise bağlantı ağırlığının öğrenme katsayısı (oranı) kadar artırılması ya da azaltılmasını ister. Öğrenme katsayısı genellikle kullanıcı tarafından belirlenen 0–1 arasında sabit bir değerdir.

Delta Kuralı: Bu kural, Hebb kuralının biraz daha geliştirilmiş ve oldukça yaygın olarak kullanılan şeklidir. Delta kuralı, işlem biriminin çıktısı ile hedeflenen çıktı değeri arasındaki farkı (delta) azaltacak yönde, giriş bağlantılarının ağırlıklarının sürekli değiştirilmesi fikrine dayanır. Bu kural, bağlantı ağırlıklarını, ağırlık ortalama karesel hata değerini düşürecek şekilde değiştirir. Aynı zamanda, Widrow-Hoff öğrenme kuralı veya en küçük ortalama kareler (LMS) öğrenme kuralı olarak da adlandırılır. Delta kuralının çalışmasında, çıktı katmanındaki delta hatası, aktivasyon fonksiyonunun türevi ile dönüştürüldükten sonra, bir önceki katmandaki giriş bağlantı ağırlıklarının

ayarlanmasında kullanılır. Diğer bir deyişle, her seferinde bir katmana olmak üzere, hata önceki katmanlara geri yayılır ve ilk katmana ulaşılan dek bu geri yayma devam eder. İleri beslemeli, geri yayımlı yapay sinir ağları, ismini, bu şekilde hata terimi hesaplanmasından almıştır. Delta kuralı kullanılırken, ağın istenen doğruluk noktasına yakınsayabilmesi için, giriş verilerinin rasgele olmasına dikkat edilmelidir[18].

Eğim Düşümü Kuralı: Delta kuralına benzeyen eğim düşümü yönteminde de, ağırlıklara uygulanmadan önce, delta hatasının değiştirilmesinde aktivasyon fonksiyonunun türevi kullanılır. Ancak bu kuralda, ağırlıklara uygulanacak öğrenme katsayısına bir orantısal sabit [18] daha eklenir. Bu kural kullanıldığında kararlı duruma yakınsamanın gerçekleşmesi uzun sürmektedir. Ancak, çıkış katmanına daha yakın olan katmanlar için öğrenme katsayısının giriş katmanına yakın olanlardan daha küçük olacak şekilde seçilmesi halinde, yakınsamanın daha hızlı gerçekleştiği de gözlenmiştir [18].

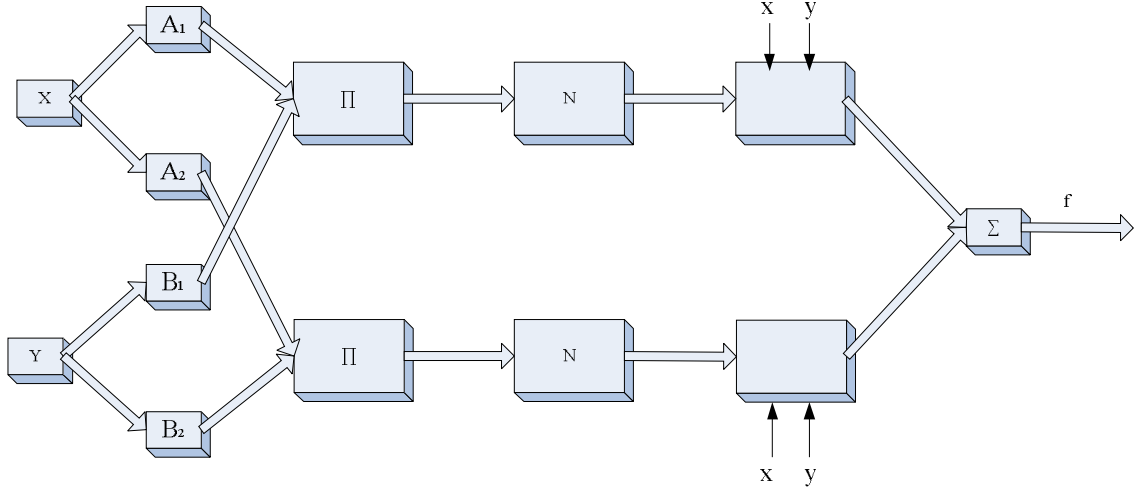
Kohonen Kuralı: Teuvo Kohonen bu kuralın geliştirilmesinde, biyolojik sistemlerin öğrenmesinden esinlenmiştir. Bu yöntemde, işlem elemanları ağırlıklarını değiştirmek ve öğrenmek için birbirleri ile yarışır. En büyük çıktıyı üreten hücre, kazanan hücre olarak yakınındaki hücrelerden daha kuvvetli hale gelir ve hem kazanan hücrelerin, hem de komşusu olan diğer hücrelerin ağırlıklarını değiştirebilir. Komşuluğun sınırları eğitim süresince değişim gösterebilir, genellikle büyük bir komşuluk ile başlanarak eğitim süreci ilerledikçe komşuluk sınırı daraltılır. Bu yöntem, verilerin istatistiksel ve topolojik modellemesinde kullanılır ve kendini örgütleyen yapılar veya topolojiler olarak da adlandırılırlar.[18].

5.2. Uyarlamalı Sinirsel Bulanık Ağ (USBA) Mimarisi

Sinirsel bulanık sistemler, yapay sinir ağlarının paralel hesaplayabilme ve öğrenme kabiliyeti ile bulanık mantığın uzman bilgisini kullanarak sonuçlar çıkarabilme özelliklerinin birleşiminden oluşur. Böylece sinirsel bulanık sistemler sayesinde YSA daha anlaşılır bir hale gelir [17].

USBA, ele alınan problem için oluşturulan yapıya göre olası tüm kuralları atayabilmekte ya da kuralların veriler yardımıyla uzman tarafından atanmasına imkân vermektedir. USBA'nın kural oluşturabilmesi veya kural oluşturulmasına olanak sağlaması uzman görüşlerinden faydalanmış olduğu anlamına gelmektedir. USBA 5

katmandan meydana gelir. Bu sistem Şekil 5.5’te gösterilmiştir. USBA yapısındaki her katmana ait düğüm işlevleri ve katmanların işleyişi sırasıyla şöyledir. [13]



Şekil 5.5. 2 girişli 4 kurallı bir USBA sınıflandırıcı yapısı

Bu USBA yapısı yukarıda da görüldüğü gibi 5 katmandan meydana gelmiştir. Burada x ve y giriş, f ise çıkış olarak alınırsa temel kural yapısı şu şekilde yazılabilir:

$$\text{Eğer } x \text{ } A_1 \text{ ve } y \text{ } B_1 \text{ ise } f_1 = p_1x + q_1y + r_1 \quad (5.4)$$

Burada p_i , q_i ve r_i lineer çıkış parametrelerini göstermektedir. İki girişli ve bir çıkışlı olan bir USBA’ nın temel yapısı Şekil 5.5’te görülmektedir. Bu yapıda, 5 katman ve 4 adet eğer-ise kuralı kullanılmıştır.

1.Katman: Bulanıklaştırma katmanı olarak ifade edilir. Giriş değerlerini bulanık kümelerle ayırmak için Jang’ın USBA modeli, üyelik fonksiyonu şekil olarak genelleştirilmiş Bell aktivasyon fonksiyonunu kullanır. Burada, her bir düğümün çıkışı, giriş değerleri ile kullanılan üyelik fonksiyonuna bağlı olan üyelik derecelerinden oluşmaktadır ve 2. katmandan elde edilen üyelik dereceleri $\mu_{Ai}(x)$ ve $\mu_{Bi}(y)$ şeklinde gösterilir. Bu katmanda bulunan hücre sayısı, iki giriş ve bu iki girişin her birine iki üyelik fonksiyonu tanımlandığına göre dört adettir ($i=4$ ’dür). Buna göre,

$$O_{1,i}=\mu_{Ai}(x), i=1,2 \quad O_{1,i}=\mu_{Bi-2}(y), i=3,4 \text{ için} \quad (5.5)$$

Burada x ve y girişleri temsil eder. Bu katmanın sahip olduğu çıkış, kuralların varsayım (eğer) kısımlarının üyelik fonksiyonlarına olan üyelik dereceleridir. Kullanılan üyelik fonksiyonu Denklem 5.6. de verilmiştir. Buradan da görüldüğü üzere üyelik fonksiyonu olarak Gauss üyelik fonksiyonu kullanılmıştır.

$$\mu_{Ai}(x), \mu_{Bi-2}(y) = \exp(-(x_i - c_i)/(a_i))^2 \quad (5.6)$$

2. Katman: Kural katmanıdır. Bu katmandaki her bir düğüm, Sugeno bulanık mantık çıkarım sistemine göre oluşturulan kuralları ve sayısını gösterir. Her bir kural düğümünün çıkışı μ_i , 1. katmandan gelen üyelik derecelerinin çarpımını oluşturur (Denklem 5.7).

$$O_{2,i}=w_i=\mu_{Ai}(x)*\mu_{Bi}(y), \quad i=1,2 \quad (5.7)$$

3. Katman: Normalizasyon katmanıdır. Bu katmandaki her bir düğüm, kural katmanından gelen tüm düğümleri giriş değeri olarak kabul edip her bir kuralın normalleştirilmiş ateşleme seviyesini hesaplar (Denklem 5.8).

$$O_{3,i}=\bar{w}_i=w_i/(w_1+w_2), \quad i=1,2 \quad (5.8)$$

4. Katman: Bu katmanda normalize edilmiş olan her bir kural kendine ait çıkış fonksiyonu ile çarpılır (Denklem 5.9).

$$O_{4,i}=\bar{w}_i * f_i=w_i*(p_i x + q_i y + r_i) \quad (5.9)$$

Buradaki p, q ve r lineer parametreleri sonuç parametreleri olarak adlandırılır.

5. Katman: Toplam katmanı olarak adlandırılır. Bu katmanda sadece bir düğüm bulunur. Bu düğüm Σ ile etiketlenmiştir. Burada, 5.katmandaki her bir düğümün çıkış değeri toplanır ve sonuçta, USBA sisteminin gerçek değeri elde edilir (Denklem 5.10).

$$O_{5,i}=\text{toplam çıkış} = \sum_i \bar{w}_i f_i = \frac{\sum_i w_i f_i}{\sum_i w_i} \quad (5.10)$$

5.2.1. USBA İçin Geri Yayılımlı Öğrenme Algoritması

Geri yayılımlı öğrenme algoritmasında ağın çıkışından elde edilen hata değeri giriş katmanına kadar geriye doğru yansıtılarak gerekli ağırlık değişkenlerinin ayarlanması gerçekleştirilir [15]. Burada amaç e hata ölçütünü öğrenme işlemi sonunda bütün giriş örnekleri için sıfıra götürmektir. Ağın çıkışında bulunan hata değeri, istenen değer d ile gerçek çıkış f arasındaki farktır ve Denklem 5.11. deki şekilde ifade edilir,

$$e = d - f \quad (5.11)$$

Bu hata değerine en küçük kareler yöntemi uygulanarak hata ölçütü olan E değeri Denklem 5.12. da verildiği gibi hesaplanır.

$$E = \frac{1}{2} e^2 \quad (5.12)$$

Hata ölçütü kullanıldığında hata geriye doğru katman katman aşağıdaki gibi gösterilir.

5. Katman: Bu katmanda herhangi bir ağırlık değeri ayarlaması yapılmaz, sadece çıkışta bulunan hatanın Denklem 5.13. deki gibi katman çıkışına yansması hesaplanır,

$$\delta^5 = -\frac{\partial E}{\partial f} = -\left[\frac{\partial E}{\partial e} \frac{\partial e}{\partial f} \right] = e \quad (5.13)$$

4. Katman: Bu katmanda p_i , q_i ve r_i gibi sonuç değişkenlerinin ayarlanabilmesi için

gerekli hesaplamalar aşağıdaki gibi gerçekleştirilir.

$$\Delta p_i = -\frac{\partial E}{\partial p_i} = \left[-\frac{\partial E}{\partial f} \right] \left[\frac{\partial f}{\partial p_i} \right] \quad (5.14)$$

$$f = \sum_i \bar{\mu}_i (p_i x + q_i y + r_i) \quad i = 1, 2 \quad (5.15)$$

$$\frac{\partial f}{\partial p_i} = \bar{\mu}_i x \quad (5.16)$$

buradan,

$$\Delta p_i = \delta^5 \bar{\mu}_i x \quad (5.17)$$

Aynı işlemler diğer sonuç değişkenlerine de uygulanarak değişimler hesaplanır. Buna göre,

$$\Delta q_i = \delta^5 \bar{\mu}_i y \quad (5.18)$$

$$\Delta r_i = \delta^5 \bar{\mu}_i \quad (5.19)$$

olur.

Elde edilen bu ifadeler öğrenme oranı katılarak üçgen (delta) kuralı uygulanır. $n = \{p, q, r\}$ olarak alınırsa bu ifadeler tek bir eşitlik olarak aşağıdaki gibi ifade edilir.

$$n_i(k+1) = n_i(k) + \eta_n \Delta n_i(k+1) \quad (5.20)$$

3. Katman: Bu katmanda çıkış katmanından yansıyan hata değeri hesaplanır. Herhangi bir ağırlık hesaplaması ise yapılmaz.

$$\delta_i^3 = -\frac{\partial E}{\partial \bar{\mu}_i} = \left[-\frac{\partial E}{\partial f} \right] \left[\frac{\partial f}{\partial \bar{\mu}_i} \right] \quad i = 1,2 \quad (5.21)$$

$$f = \sum_i \bar{\mu}_i f_i \quad (5.22)$$

$$\frac{\partial f}{\partial \bar{\mu}_i} = f_i \quad (5.23)$$

$$\delta_i^3 = \delta^4 f_i \quad (5.24)$$

2. Katman: Bu katmanda çıkış katmanından yansıyan hata değeri hesaplanır. Herhangi bir ağırlık hesaplaması ise yapılmaz.

$$\delta_i^2 = -\frac{\partial E}{\partial \mu_i} = \left[-\frac{\partial E}{\partial f} \frac{\partial f}{\partial \bar{\mu}_i} \right] \left[\frac{\partial \bar{\mu}_i}{\partial \mu_i} \right] \quad i = 1,2 \quad (5.25)$$

$$\bar{\mu}_i = \frac{\mu_i}{\sum_i \mu_i}, \text{ bu ifade de } a = \sum_i \mu_i \text{ olarak alınır,}$$

$$\frac{\partial \bar{\mu}_i}{\partial \mu_i} = \left[\frac{a - \mu_i}{a^2} \right] \quad (5.26)$$

olarak bulunur. Bu ifade Denklem 5.27. de yerine konularak aşağıdaki gibi son durum elde edilir.

$$\delta_i^2 = \delta_i^3 \left[\frac{a - \mu_i}{a^2} \right] \quad (5.27)$$

1. Katman: Bu katmanda hem hata değeri hem de giriş üyelik işlevlerine ait değişkenlerin ayarlanması sağlanır.

$$\delta_i^1 = -\frac{\partial E}{\partial \mu_{Ai}} = -\left[\frac{\partial E}{\partial f} \frac{\partial f}{\partial \bar{\mu}_i} \frac{\partial \bar{\mu}_i}{\partial \mu_i} \right] \left[\frac{\partial \mu_i}{\partial \mu_{Ai}} \right] \quad i = 1,2 \quad (5.28)$$

$$\frac{\partial \mu_i}{\partial \mu_{Ai}} = \mu_{Bi} \quad (5.29)$$

$$\delta_i^1 = \delta_i^2 \mu_{Bi} \quad (5.30)$$

$$\delta_{i+2}^1 = -\frac{\partial E}{\partial \mu_{Bi}} = -\left[\frac{\partial E}{\partial f} \frac{\partial f}{\partial \bar{\mu}_i} \frac{\partial \bar{\mu}_i}{\partial \mu_i} \right] \left[\frac{\partial \mu_i}{\partial \mu_{Bi}} \right] \quad (5.31)$$

$$\frac{\partial \mu_i}{\partial \mu_{Bi}} = \mu_{Ai} \quad (5.32)$$

$$\delta_{i+2}^1 = \delta_i^2 \mu_{Ai} \quad (5.33)$$

$$\mu_{Ai} = \exp \left[-\left(\frac{x - m_i}{\sigma_i} \right)^2 \right] \quad v_i = -\left[\frac{x - m_i}{\sigma_i} \right]^2 \quad \mu_{Ai} = \exp[v_i] \quad (5.34)$$

Çan eğrisi şekilli üyelik işlevlerinin orta noktalarındaki değişim aşağıdaki gibi hesaplanır.

$$\Delta m_i = -\frac{\partial E}{\partial m_i} = -\left[\frac{\partial E}{\partial f} \frac{\partial f}{\partial \bar{\mu}_i} \frac{\partial \bar{\mu}_i}{\partial \mu_i} \frac{\partial \mu_i}{\partial \mu_{Ai}} \right] \left[\frac{\partial \mu_{Ai}}{\partial v_i} \frac{\partial v_i}{\partial m_i} \right] \quad i = 1,2 \quad (5.35)$$

$$\frac{\partial \mu_{Ai}}{\partial v_i} = \exp[v_i] = \mu_{Ai} \quad (5.36)$$

$$\frac{\partial v_i}{\partial m_i} = 2 \left[\frac{x - m_i}{\sigma_i^2} \right] \quad (5.37)$$

$$\Delta m_i = 2 \left[\frac{x - m_i}{\sigma_i^2} \right] \delta_i^1 \mu_{Ai} \quad \delta_i^1 = \delta_i^2 \mu_{Bi} \quad (5.38)$$

$$\Delta m_i = 2 \left[\frac{x - m_i}{\sigma_i^2} \right] \delta_i^2 \mu_{Ai} \mu_{Bi} \quad \mu_i = \mu_{Ai} \mu_{Bi} \quad (5.39)$$

$$\Delta m_i = 2 \left[\frac{x - m_i}{\sigma_i^2} \right] \delta_i^2 \mu_i \quad (5.40)$$

Aynı şekilde Δm_{i+2} aşağıdaki denklem ile elde edilir,

$$\Delta m_{i+2} = 2 \left[\frac{y - m_{i+2}}{\sigma_{i+2}^2} \right] \delta_i^2 \mu_i \quad (5.41)$$

Çan eğrisi şekilli üyelik işlevlerinin standart sapmalarındaki değişim aşağıdaki gibi hesaplanarak bulunur.

$$\Delta \sigma_i = - \frac{\partial E}{\partial \sigma_i} = - \left[\frac{\partial E}{\partial f} \frac{\partial f}{\partial \bar{\mu}_i} \frac{\partial \bar{\mu}_i}{\partial \mu_i} \frac{\partial \mu_i}{\partial \mu_{Ai}} \right] \left[\frac{\partial \mu_{Ai}}{\partial v_i} \frac{\partial v_i}{\partial \sigma_i} \right] \quad i = 1, 2 \quad (5.42)$$

$$\frac{\partial v_i}{\partial \sigma_i} = 2 \left[\frac{(x - m_i)^2}{\sigma_i^3} \right] \quad (5.43)$$

$$\Delta \sigma_i = 2 \left[\frac{(x - m_i)^2}{\sigma_i^3} \right] \delta_i^2 \mu_i \quad (5.44)$$

Aynı şekilde $\Delta \sigma_{i+2}$ aşağıdaki denklem ile elde edilir.

$$\Delta \sigma_{i+2} = 2 \left[\frac{(y - m_{i+2})^2}{\sigma_{i+2}^3} \right] \delta_i^2 \mu_i \quad (5.45)$$

Elde edilen bu değişimlere üçgen kuralı uygulanarak üyelik işlevine ait değişkenlerin bulunması için aşağıdaki denklemler kullanılır.

$$m_i(k+1) = m_i(k) + \eta_m \Delta m_i(k+1) \quad i = 1, 2, 3, 4 \quad (5.46)$$

$$\sigma_i(k+1) = \sigma_i(k) + \eta_m \Delta \sigma_i(k+1) \quad i = 1, 2, 3, 4 \quad (5.47)$$

Burada toplam 8 adet deęişken ayarlanır[13].

6. YAPAY ZEKÂ TABANLI KONUŞMA TANIMA UYGULAMALARI

Bu bölümde yapılan çalışmalar detaylandırılarak anlatılmıştır. Veri tabanı olarak bizim tarafımızdan hazırlanan Türkçe sesli harflerden oluşan veritabanı kullanılmıştır. Her kişiden her sesli harf için 4 örnek alınmış olup örnek alınan kişi sayısı 6'dır. Dolayısıyla $6 \times 4 \times 8 = 192$ data girdisi olan bir veritabanı kullanılmıştır. Yapılan çalışmalar matlab programı kullanılarak gerçekleştirilmiştir. Yapılan çalışmalar iki aşamadan oluşmaktadır. Birinci aşama eğitim aşaması, ikinci aşama ise test aşamasıdır. Eğitim aşamasında, kaydolma (enrollment) ve saptama (identification) işlemleri gerçekleştirilmiştir. Kaydolma aşamasında konuşmacıların tüm ses örnekleri alınarak işlenecek ve bir veritabanında saklanmıştır. Kaydolma aşamasının amacı, her sesli harf için bir referans modeli oluşturmaktır. Daha sonra bu modeller saptama aşamasında kullanılmıştır.

Test aşaması ise, tanıma metodunun kapalı küme ya da açık küme olmasına göre değişir. Çalışmamızda kapalı küme tanıma sistemi kullanılmıştır. Kapalı küme tanıma sisteminde, veritabanındaki bir ses bilinen ses örneğinin sahibi olarak seçilmek zorundadır. Dolayısı ile en yüksek benzerlik skoru alan ses modeli seçilmiş olacaktır.

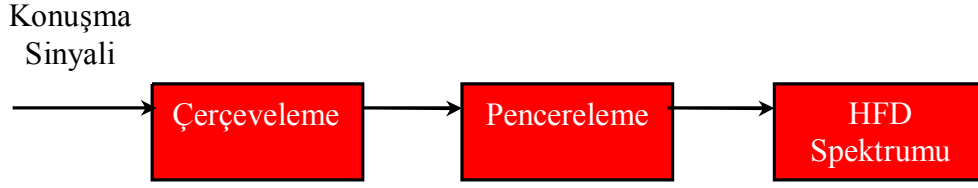
YSA ve USBA MATLAB paket programı kullanılarak eğitilmiştir. Eğitilen ağ yapısı konuşmacılardan alınmış eğitim verilerine göre test edilmiş ve konuşma tanımadaki başarı oranları tespit edilerek karşılaştırılmıştır.

6.1. MFKK ve Yapay Zekâ ile Konuşma Tanıma

6.1.1. MFKK ve YSA ile Konuşma Tanıma

Öznitelik vektörlerinin bulunması sürecinde 6 kişinin söylediği 32 ses kullanılmıştır. Bunun için önce konuşmacılara ait konuşma sinyalleri 30 ms'lik çerçevelere ayrılmıştır. Pencereleme işleminde Hamming pencere kullanılmıştır.

Pencerelemiş konuşma sinyal çerçevelerinin HFD'si alınarak konuşma sinyali frekans bölgesine taşınmıştır. Bir sonraki adımda HFD spektrumundan mel spektrumuna geçilmiştir. Bunun için HFD spektrumunun genlik değerlerinin mutlak kare değerleri hesaplanarak güç spektrumu oluşturulmuştur.



Şekil 6.1.Konuşma sinyalinin hızlı fourier dönüşümü

Mutlak genlik değerlerinin kareleri hesaplanan HFD noktaları, band geçiren üçgen dalgalardan oluşan mel-filtre bankasından geçirilmiştir. Kullanılan mel filtre bankası, 13 adet doğrusal olarak yerleştirilen filtreyi takip eden 27 adet logaritmik filtreden meydana gelmiştir.

Mel spektrumun elde edilmesinden sonra mel kepstral katsayıları elde edilmiştir. Bunun için mel filtre bankası çıkışının logaritması alınarak kesik kosinüs dönüşümü ile zaman bölgesine geçiş yapılır. Oluşturulan bu katsayılar mel-ölçeklendirilmiş HFD temelli kepstral katsayılardır. AKD kullanmanın sebeplerinden biri AKD uygulamasının THFD'ye göre karmaşık aritmetik işlemler gerektirmemesidir. Bütün bu aşamalardan sonra YSA için 13 adet MFKK katsayısı elde edilmiştir.



Şekil 6.2.HFD'si alınan sinyalden Kepstal katsayıların elde edilmesi

Tamamı Türkçe sesli harflerden oluşan 192 adet datanın öznitelik vektörleri kullanılarak eğitim dizisi oluşturulmuştur. İlk aşama eğitim safhasıdır. Çalışmamız metine bağımlı olduğu için özellik çıkarım aşamasında elde edilen $6 \times 32 \times 13$ adet özellik vektörü yapay sinir ağları ile akıllı bir sınıflandırma yapılarak eğitilmiştir. Kullanılan çok katmanlı algılayıcı yapılı yapay sinir ağının eğitim parametreleri ve yapısı Tablo 6.1'de gösterilmektedir.

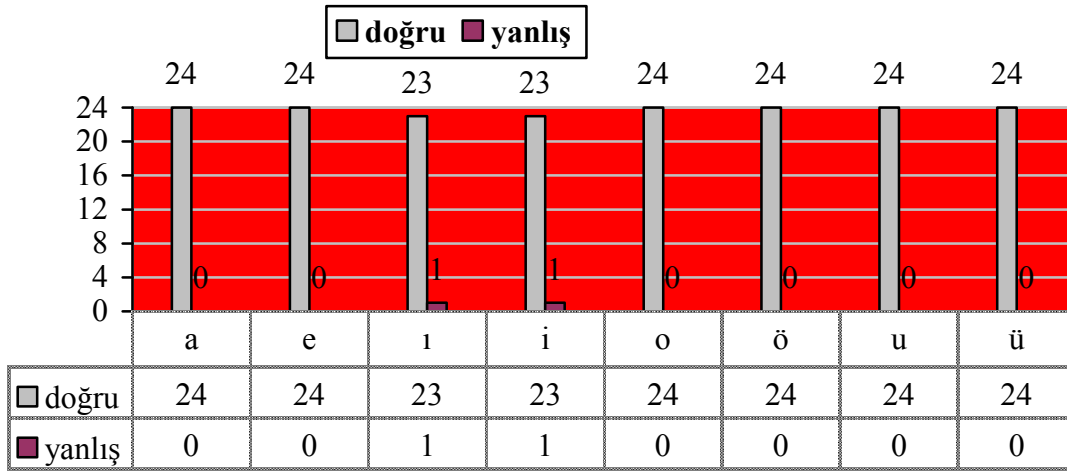
Tablo 6.1. Çok katmanlı yapay sinir ağının yapısı ve eğitim parametreleri

Katman sayısı	3
Aktivasyon fonksiyonları	Logaritmik sigmoid,tanjant sigmoid,linear
Öğrenme algoritması	Geriye yayılım algoritması
Toplam karesel hata	0.000000000000000000000001

Gizli katman sayısı, gizli katmanlardaki hücre sayısı, öğrenme oranı ile moment sabitinin değeri ve aktivasyon fonksiyonu, birkaç farklı denemeden sonra en iyi performansı verebilecek şekilde seçilmiştir.

Eğitim aşamasından sonra test aşamasına geçilmiştir. Konuşmacıların eğitimde kullandıkları kelimeler, gürültü eklenerek test aşamasında kullanılmıştır. Gürültülü YSA için doğru konuşma tanıma gerçekleşmezken, gürültü eklenmeden test edildiğinde ise YSA için doğru konuşma tanıma oranı %98,95 olmuştur. Tablo 6.2’de bu metoda ilişkin harf bazında tanıma oranları verilmiştir

Tablo 6.2.MFKK ve YSA ile konuşma tanıma sonuçları



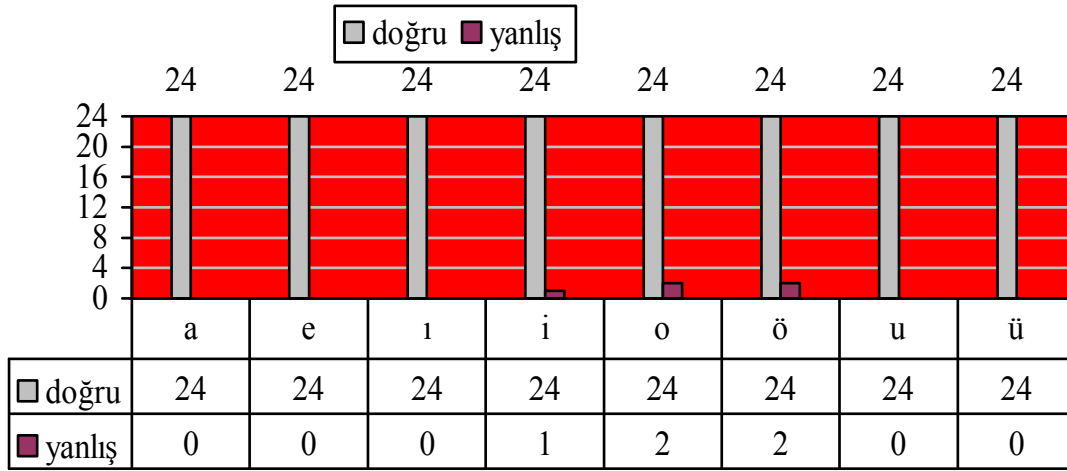
6.1.2. MFKK ve USBA ile Konuşma Tanıma

Burada, özellik çıkarım safhasında elde edilen 13x192 özellik vektörü kullanılarak uyarlamalı sinirsel bulanık ağ ile akıllı bir tanıma yapılmıştır. Kullanılan USBA tabanlı çıkarım sisteminin eğitim parametreleri ve yapısı Tablo6.3’de gösterilmektedir.

Tablo 6.3. USBA modelinin yapısı ve eğitim parametreleri

Katman sayısı	5
Üyelik fonksiyonları	Gbellmf
Eğitim parametreleri Öğrenme kuralı	Melez Öğrenme Algoritması (Doğrusal olmayan giriş parametreleri için Geri yayılım ve doğrusal çıkış parametreleri için En küçük kareler yöntemlerinin birleşimleri)

MFKK katsayıları eğitildikten sonra test aşamasına geçilmiştir. Konuşmacıların eğitimde kullandıkları kelimeler, gürültü eklenerek test edilmiş, fakat USBA için arzu edilen bir başarımla konuşma tanıma gerçekleşmemiştir. Gürültü eklenmeden test edildiğinde ise USBA için doğru tanıma oranı %97,4 olmuştur. Tablo 6.4’de bu metoda ilişkin harf bazında tanıma oranları verilmiştir.

Tablo 6.4. MFKK ve USBA ile konuşma tanıma sonuçları

6.2. DD Kullanarak Yapay Zekâ Tabanlı Konuşma Tanıma

6.2.1. DD ve YSA ile Konuşma Tanıma

Öznitelik vektörlerinin bulunması sürecinde oluşturduğumuz veri tabanından 6 konuşmacının söylediği 32 kelime kullanılmıştır. Bunun için önce seslere ait konuşma sinyalleri çerçevelenmiş ardından dikdörtgen pencere kullanılarak pencerelenmiştir. Bu sayede sinyal bu pencere ile çarpıldığında aralık dışındaki değerleri sıfır olmuştur.

Daha sonra pencerelenen sinyale dalgacık dönüşümü uygulanmıştır. Böylece daha esnek zaman-frekans gösterimleri elde edilerek her seviye için ayrışım filtreleri ölçeklendirilmiştir. Dönüşüm sonucunda ise işaret ve dalgacık birbirleriyle iyi eşleştiğinden işaret ve dalgacık arasındaki korelasyonun da yüksek olduğu görülmüştür.

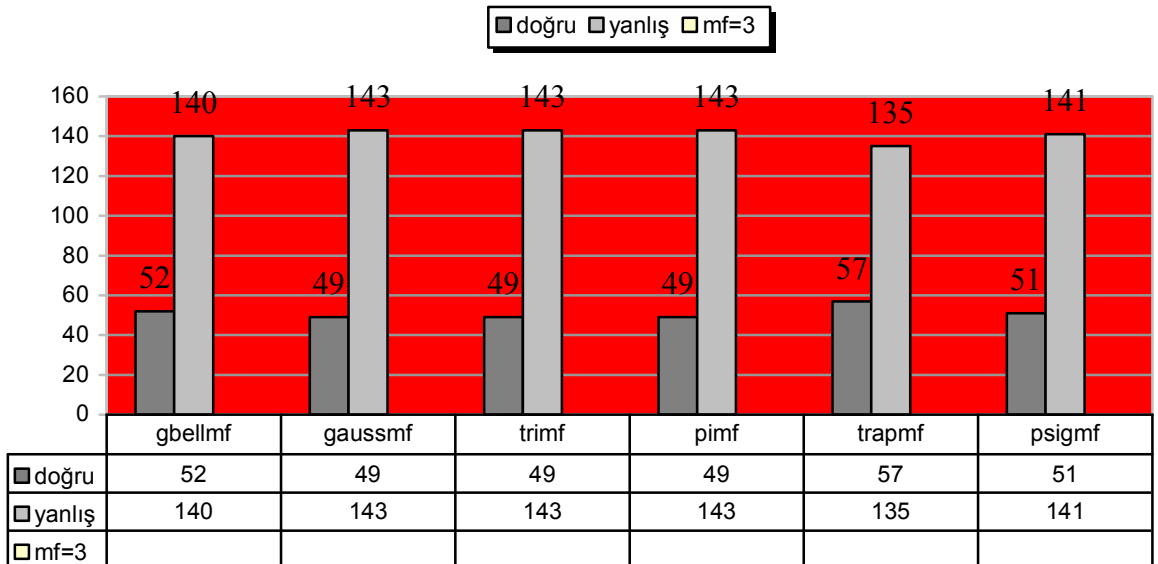
Uygulamamızda Symlet dalgacık türü kullanılarak entropi katsayıları hesaplanmıştır. Entropi tipi olarak ta Shannon kullanılmış ve sonuçta 4 tane entropi katsayısı elde edilmiştir.

Özellik çıkarım safhasında elde edilen 4x192 özellik vektörü YSA ile eğitilmiştir. Test aşamasında ise 1, 5, 10, -1, -5, -10 dB değerinde farklı gürültü değerleri eklenerek test edilmiş fakat arzu edilen başarı elde edilememiştir.

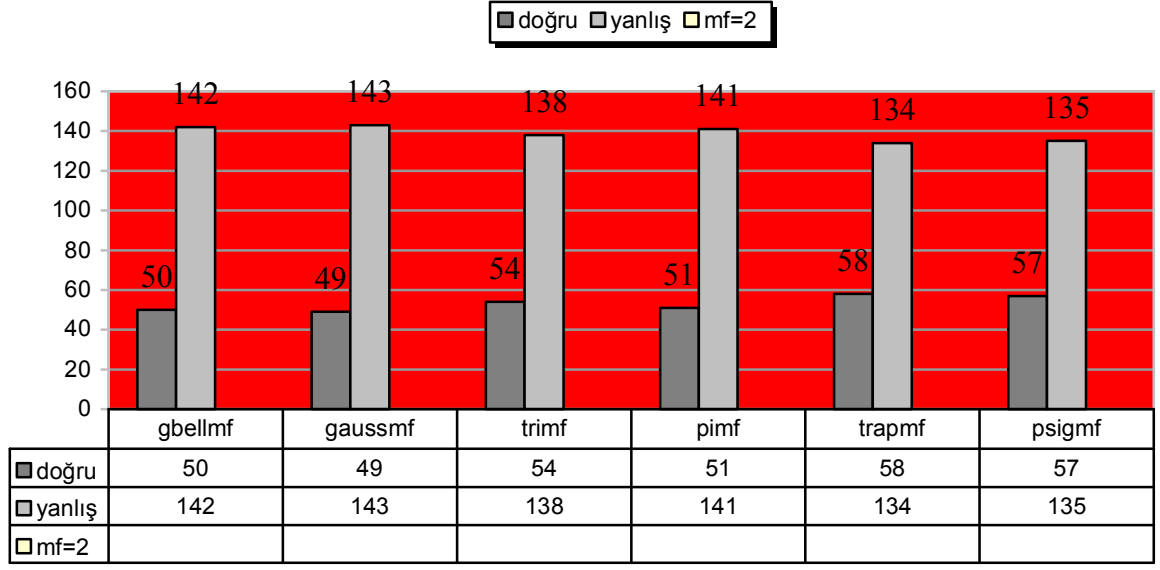
6.2.2. DD ve USBA ile Konuşma Tanıma

Aynı entropi katsayıları USBA ile akıllı bir sınıflandırma yapılarak eğitilmiştir. Eğitim aşamasında elde ettiğimiz katsayılar öncelikle üyelik fonksiyon sayısı 2 ve 3 olan farklı üyelikler ile tek tek eğitilmiştir. Test aşamasında yine aynı fonksiyonlarla test edilmiş farklı oranlarda başarı elde edilmiştir. Kullanılan USBA sisteminde başarı oranı en fazla %29,7 olmuştur. Yapılan çalışmada kullanılan eğitim parametreleri ve başarı yüzdeleri Tablo 6.5.'te gösterilmektedir.

Tablo 6.5. Uyarlamalı sinirsel bulanık ağ sisteminin parametreleri ve başarı yüzdesi



Tablo 6.5' in devamı



7. SONUÇLAR VE TARTIŞMA

Bu tezde, konuşma tanıma sistemlerinin genel yapısı incelenmiş, metine bağımlı ve metinden bağımsız veya kişiye bağımlı ve kişiden bağımsız sistemler için kullanılmakta olan yöntemler özelliklerine göre sınıflandırılmış ve her biri ayrı ayrı incelenmiştir.

Çalışmadaki ilk hedef, geliştirilecek uygulamanın konuşma doğrulama sistemi olarak tasarlanmasıdır. Daha sonraki aşama da ise metin bağımlı ve metin bağımsız sistem arasında bir tercih yapılması gerekmiştir. Gerek daha kolay gerçekleştirilebileceği gerekse de daha fazla başarımlı sağlayacağı düşüncesiyle metin bağımlı bir sistem tasarımı tercih edilmiştir. Metin bağımlı konuşmacı doğrulama uygulamasının gerçekleşmesi sırasında her bir işlem basamağı incelenmiştir.

Öznitelik vektörlerinin belirlenmesinde MFKK yöntemi ile birlikte, kıyaslama yapmak amacıyla DD uygulanmıştır. DD' yi kullanmadaki amaç sinyalin önileme aşamasındaki gürültüleri temizlemektir. Daha sonra DD ile elde edilen veri dizilerinin entropi değerleri hesaplanmıştır. Modelleme aşamasında ise YSA ve USBA gibi yapay zekâ metotları ile modelleme yapılmıştır.

Yapay zekâ tabanlı modelleme çalışmalarında yapılan ölçümler ve hesaplamalar sonucunda elde edilen veriler herhangi bir ön işleme tabi tutulmadan kullanıldığı için sistemin akıllı modelini oluştururken çok sayıda örnek sayısına ihtiyaç duyulmaktadır. Çok sayıda örnek kullanıldığında ise hesaplama süresinin ve hafıza gereksinimin artması ve eğitilen sistemin sistem davranışını ezberlemesi gibi bazı sorunlar ortaya çıkmıştır. Giriş veri dizisine birden çok perspektiften bakılarak elde edilen her bir veri-işaret bileşeninin entropi değerleri hesaplanmıştır.

MFKK katsayıları elde edilerek yapılan YSA ve UBSA tabanlı modelleme çalışmalarında, gürültünün tanımayı az denemeyecek kadar etkilediği görülmüştür. Sonuçta istenilen test başarımlını gösterememiştir.

DD ile elde edilen entropi katsayıları kullanılarak yapılan YSA ve USB A tabanlı modelleme çalışmalarında ise gürültünün YSA akıllı metodu kullanıldığında tanımayı gerçekleştiremediği, USB A akıllı metodu kullanıldığında ise gürültünün az denemeyecek kadar tanımayı etkilediği görülmüştür. Tanıma başarımını etkileyen faktörler olarak entropi türü, dalgacık tipi, dalgacığın seviyesi gibi parametreler olduğu tespit edilmiştir. Aynı zamanda veri dizisi oluştururken kullanılan konuşmacı sayısı ile konuşulan kelime sayısının artışı tanıma başarısını etkileyerek sistem performansını düşürmüştür. Konuşma veritabanımızdaki 6 kişilik konuşma kümesi ile yapılan testler göstermiştir ki önerilen MFKK ve YSA tabanlı tanıma sistemi gerek tanıma gerek güvenilirlik açısından önceki algoritmalara göre daha iyi sonuçlar vermiştir.

Bundan sonraki yapılacak olan çalışmalarda, kullanılacak ve geliştirilecek algoritmalarda veri alma, özellik çıkarma ve sınıflama işlemlerini daha hızlı yapabilecek aynı zamanda gürültünün ya da çevresel şartların tanıma başarımını etkilemeyeceği, güvenilirli farklı yöntemler geliştirilebilir.

KAYNAKLAR

- [1] **Mengüşoğlu, E., (1999)**, Bir Türkçe Sesli İfade Tanıma Sisteminin Kural Tabanlı Tasarımı ve Gerçekleştirimi, Yüksek Mühendislik Tezi, Hacettepe Üniversitesi.
- [2] **Uslu, E., (2007)**, Gizli Markov Modeli İle Geniş Sözlüklü Sürekli Konuşma Tanıma, Yüksek Lisans Tezi, Yıldız Teknik Üniversitesi.
- [3] **Aygün O. 2006**,Konuşmacı tanıma sistemlerinde dalgacık dönüşümü , Yüksek Lisans Tezi, Selçuk Üniversitesi
- [4] **Artuner, H., 1994**, “Bir Türkçe Fonem Kümeleme Sistemi Tasarımı ve Gerçekleştirimi”,Doktora Tezi, Hacettepe Üniversitesi Fen Bilimleri Enstitüsü.
- [5] İnternet1:http://www.emo.org.tr/ekler/b613c6dda6b52b5_ek.pdf
erişim tarihi:Nisan 2011
- [6] **S. Davis, P. Mermelstein, 1980** "Comparison of parametric representations formonosyllabic word recognition in continuously spoken sentences," IEEE Acoustics, Speech, and Signal Processing, vol. 28, no. 4, pp. 357-366,
- [7] **Ningping Fan, Justinian P. Rosca, 2003**, "Enhanced VQ-Based Algorithms for Speech Independent Speaker Identification," Audio and Video based Biometric Person Authentication, pp.
- [8] **T. Kinnunen,2005**, et al. "Real-Time Speaker Identification and Verification," IEEE Transactions on Speech and Audio Processing vol. PP, no. 99, pp. 1-12

- [9] **S. Molau, et al. 2001**, "Computing Mel-frequency cepstral coefficients on the power spectrum," IEEE Int. Conf. Acoustics, Speech, and Signal Processing, vol. 1, pp. 73-76,
- [10] **H. Gish and M. Schmit, 1994**, "Text-Independent Speaker Identification," IEEE Signal Processing Magazine, vol. 11, pp. 18-32,
- [11] **Thomas F.,2002**, Quatieri, Discrete-Time Speech Signal Processing: Principles and Practice, Prentice Hall, MIT Lincoln Laboratory.
- [12] İnternet2:http://courses.cit.cornell.edu/ece576/FinalProjects/f2008/pae26_jsc59/pae26_jsc59/index.html.erişim tarihi:Nisan 2011
- [13] **Türkoğlu, İ., 2002**. Durağan Olmayan İşaretler İçin Zaman-Frekans Entropilerine Dayalı Akıllı Örüntü Tanıma, *Doktora Tezi*, Fırat Üniversitesi Fen Bilimleri Enstitüsü, , Elazığ.
- [14] **Keeton, P. I. J. and Schlindwein, F. S., 1997**, Application of wavelets in Doppler ultrasound,Sensor Review, MCB University Press, .
- [15] **Toprak İnayet,2007**, EEG sinyallerinin dalgacık dönüşümü ve yapay sinir ağları ile analizi,Yüksek lisans tezi,Süleyman DEMİREL Üniversitesi.
- [16] **Principe, J. C., Euliano, N. R. and Lefebvre, W. C., 2000**, Neural and Adaptive Systems. John Wiley & Sons, New York, 656p
- [17]**Elmas, Ç., 2007**. Yapay Zekâ Uygulamaları, Seçkin Yayınları, Ankara.
- [18] **Toprak İnayet,2007**, EEG sinyallerinin dalgacık dönüşümü ve yapay sinir ağları ile analizi,Yüksek lisans tezi,Süleyman DEMİREL Üniversitesi.
- [19] **Shannon, C. E., 1948**, A mathematical theory of communication, Bell System Technology Journal.

[20]**Elmas, Ç., 2003**, Bulanık Mantık Denetleyiciler, Seçkin Yayınları, Ankara.

ÖZGEÇMİŞ

Mehmet Bilal ŞAŞMAZ

Fırat Üniversitesi
Teknik Eğitim Fakültesi
Elektronik ve Bilgisayar Eğitimi Bölümü
Elazığ

Tel: 424 – 2334883
E.posta: sasmazbilal@hotmail.com

Mehmet Bilal ŞAŞMAZ, 1984 yılında Malatya’ da doğmuştur. İlk, orta ve lise öğrenimini sırasıyla, Rahmi akıncı İlköğretim okulu, Şehit Kemal Özalper Anadolu Meslek Lisesinde tamamlamıştır. 2005 yılında, Fırat Üniversitesi Teknik Eğitim Fakültesi Elektronik Bilgisayar Eğitimi Elektronik Öğretmenliği bölümüne girmeye hak kazanmıştır. 2009 yılında lisans eğitimini tamamladıktan sonra aynı yıl F.Ü Fen Bilimleri Enstitüsüne bağlı Elektronik Bilgisayar Eğitimi Anabilim Dalı, Telekomünikasyon Programında yüksek lisans öğrenimine başlamıştır. Mezun olduğu tarih olan 22 Haziran 2009’dan bu güne kadar Malatya Asansör Doğalgaz Ltd.Şti bünyesinde Elektronik ve Bilgisayar teknisyenliği görevini yürütmektedir. Yabancı dili İngilizcedir.