T.C
SELÇUK ÜNİVERSİTESİ
FEN BİLİMLERİ ENSTİTÜSÜ

YAPAY SİNİR AĞINI KULLANARAK KİŞİYE ÖZEL KOMUT TANIMA

Gediz TATAR

YÜKSEK LİSANS TEZİ

ELEKTRİK-ELEKTRONİK MÜHENDİSLİĞİ A.B.D.

Konya, 2008

T.C
SELÇUK ÜNİVERSİTESİ
FEN BİLİMLERİ ENSTİTÜSÜ

YAPAY SİNİR AĞINI KULLANARAK KİŞİYE ÖZEL KOMUT TANIMA

Gediz TATAR

YÜKSEK LİSANS TEZİ

ELEKTRİK-ELEKTRONİK MÜHENDİSLİĞİ ANABİLİMDALI

Bu tez ..... tarihinde aşağıdaki jüri tarafından oybirliği / oyçokluğu ile kabul edilmiştir.

.....	.....
Yrd. Doç. Dr. Musa AYDIN	Yrd. Doç. Dr. Ömer AYDOĞDU
(Üye)	(Üye)

.....
Öğr. Gör. Dr. S. Sinan GÜLTEKİN
(Danışman)

ÖZET

Yüksek Lisans Tezi

YAPAY SİNİR AĞINI KULLANARAK KİŞİYE ÖZEL KOMUT TANIMA

Gediz TATAR

Selçuk Üniversitesi
Fen Bilimleri Enstitüsü
Elektrik-Elektronik Mühendisliği Anabilim Dalı

Danışman : Öğr.Gör. Dr. S.Sinan GÜLTEKİN

2008. 76 sayfa

Jüri : Yrd. Doç. Dr. Musa AYDIN
Yrd. Doç. Dr. Ömer AYDOĞDU
Öğr. Gör. Dr. S. Sinan GÜLTEKİN

Kaydedilmiş olan bir ses, herhangi bir olayın ortaya çıkarılıp aydınlatılmasında mevcut tek delil olabilir. Böylece şüpheli yada şüpheliler ortaya çıkarılarak işlenen suçun faillerinin bulunmasında kolaylık sağlayabilir. Günümüzde bu tekniğin uygulanmasıyla terör eylemleri, cinayetler, kaçırma, tehdit, şantaj, tecavüz, organize suçlar ve telefonla rahatsız etme olayları aydınlatılabilmektedir.

Ses tanıma ve tanımlama işlemi, bütün işitsel veya görsel duyuların kullanıldığı çok yönlü bir işlemdir. Bu işlem, bilinmeyen bir sesin bir veya daha fazla bilinen sesle tanınması veya elenmesi amacıyla işitsel veya görsel olarak karşılaştırılması şeklinde tanımlanabilir. Seslerin, kendi başına sahip olduğu karakteristikler ve

özellikler yardımıyla çeşitli analiz teknikleri ve yöntemleri uygulanarak diğer seslerden ayırt edilmesi bu tanıma temel teşkil eder.

Bu çalışmada Yapay Sinir Ağları'nın Çok Katlı Perseptron ağ yapısı kullanılarak, ses tanıma problemine uygulanmıştır. Elde edilen sonuçların hata analizi yapılarak, kullanılan öğrenme algoritmalarının bu problem için uygulanabilirliliği ve herhangi bir sistem kontrolünün yapılabilirliliği hedeflenmiştir.

LPC tekniği ile elde edilen ses genlik verileri Yapay Sinir Ağları Modelinin giriş katmanının oluşturmuştur. Çıkış katmanı ise her komut kelimeyi temsil eden ve karşılık gelen 1 ve 0'lardan meydana getirilmiştir. Yapay Sinir Ağları Modeli ile eğitme işlemleri, hedeflenen 1 ve 0 çıkış verileri için en iyi değer elde edilinceye kadar denenmiş ve ağ yapısı belirlenmiştir. Elde edilen sonuçların hata hesapları yapılmış ve herhangi bir sistemin kontrolünü sağlamaya yönelik değerlendirmeler yapılmıştır.

Ayrıca hata hesabı kullanılmadan sadece komutlar ayrı ayrı Yapay Sinir Ağı modelinde eğitilerek konuşmacıların tanınabilmesi sağlanmıştır.

Anahtar Kelimeler : Ses tanıma, Yapay Sinir Ağları

ABSTRACT

MS Thesis

SPECIALLY PERSONALIZED ORDER RECOGNIZATION BY USING ARTIFICAL NEURAL NETWORKS

Gediz TATAR

Selcuk Universty
Gradute School of Natural and Applied Sciences
Department of Electrical-Electronics Engineering

Supervisor : Lecturer Dr. S.Sinan GÜLTEKİN

2008. 76 Pages

Jury : Yrd. Doç. Dr. Musa AYDIN
Yrd. Doç. Dr. Ömer AYDOĞDU
Öğr. Gör. Dr. S. Sinan GÜLTEKİN

A voice recorded into a cassette might be the only evidence to reveal a event. In this way, the suspect or the suspects are found and it can be easier to find the ones who commit crimes. In today's world actions such as terror, murders, hijacking, threat, blackmail, rape, organized crimes and disturbing on the phone can be revealed.

Voice defining is a many sided process in which all auditory and visual senses are used. This process can be defined as auditory and visual comparison to define an

unidentified voice by means of a more identified voice. In essence, by means of the characteristics of the voices, they are distringuished from others by using various analysis techniques and methods which forms a basis to this definition.

In this study, ANN's Multi Layer Perceptron topology has been applied to voice recognition problem. By accomplishing the error analyse of the results, the applicibility of used trainin algotithms and feasibility to any system control are intended.

The voice amplitude data optained by the LPC technique have formed the entrance layer of the Artificial Neural Networks. Whereas; exit layer has been formed by 1's and 0's that represents every command word. The process of training via Artificial Neural Networks has been continued until the best data was optained out of 1's and 0's and finally the network formation was specified. The error calculations of the results were done and evaluations on providing any system control were also done.

Besides, without using error calculation, only by training the commands in Artificial Neural Network model, the recognition of the speakers has been provided.

Key Words : Voice Recognition, Artificial Neural Networks

TEŞEKKÜR

Yüksek Lisans çalışmalarım sırasında özverili yardımlarından dolayı; tez danışmanım Öğr. Gör.Dr. S. Sinan GÜLTEKİN'e , tez çalışmalarım sırasında bana sabırla destek veren nişanlım Çiğdem'e, dayım Kemal YILMAZ'a ve her zaman yanımda hissettiğim aileme sonsuz teşekkürlerimi sunarım.

Gediz TATAR
Ocak 2008, KONYA

İÇİNDEKİLER

ÖZET i
ABSTRACT iii
TEŞEKKÜR v
İÇİNDEKİLER vi
SİMGELER VE KISALTMALAR ix

BÖLÜM 1 1
GİRİŞ 1
1.1.Ses, Konuşma, Konuşmacı Tanıma Uygulamalarında ve Yapay Sinir Ağlarında Yapılan Çalışmalar 2

BÖLÜM 2 5
SESİN ÖZELLİKLERİ 5
2.1. Ses Nedir? 5
2.2. Konuşma 6
2.2.1. Ses Yolunun İşlevi ve Akustik Analiz 7
2.2.2. Ses Sisteminde Uyarım 8
2.3. İşitme 8

BÖLÜM 3 10
SESİN KODLANMASI 10
3.1 Giriş 10
3.2. Kodlayıcı Tipleri 10
3.2.1. Doğrusal Kuantizasyon Teknikleri 10
3.2.2. Vektörel Kuantizasyon Teknikleri 11
3.3. Kanal Kodlama 20
3.3.1. Otomatik Yeniden Gönderme (Automatic Repeat Request - ARQ) 20
3.3.2. Kodlama Oranı 21
3.3.3. Basit Parite Kontrol Kodu 21
3.3.4. Hata Düzeltme 21

3.4 Fourier Analizi 22
3.4.1 Fourier Dönüşümleri 22
3.4.2 Ayrık Fourier Dönüşümleri (DFT) 22
3.4.3. Pencerelenmiş Fourier Dönüşümleri (WFT) 23
3.5 Ses Kodlamanın En Çok Kullanıldığı Alan Ses Tanıma 24
3.5.1. Ses Tanımanın Avantajları ve Sınırlamaları 26
3.6 Ses Kodlamanın Gelişimi 28

BÖLÜM 4 31
YAPAY SİNİR AĞLARININ SES TANIMADA KULLANILMASI 31
4.1. Kavramlar ve Terimler 31
4.2. Yapay Sinir Ağlarının Sınıflandırılması 36
4.2.1. Tek Katmanlı İleri Beslemeli Ağlar 36
4.2.2. Çok Katmanlı İleri Beslemeli Ağlar 37
4.2.3. Geri Dönümlü Ağlar 37
4.2.4. Kafes Yapılı Ağlar 38
4.3. Yapay Sinir Ağlarında Öğrenme 38
4.3.2. Güdümsüz Öğrenme 39
4.4. Çok Katmanlı Perseptron ve Geri Yayılımlı Öğrenme 39
4.4.1. Geri Yayılım Algoritmaları 41
4.4.2 DeltaBarDelta (DBD) 42
4.4.3 Genişletilmiş DeltaBarDelta (Extended DBD) 44
4.4.4 Hızlı Yayılım (HY) 47
4.5.Radyal Tabanlı Yapay (RTY) Sinir Ağı Yapısı 50
4.6. Ses Tanımada Yapay Sinir Ağları 52

BÖLÜM 5 56
YSA İLE SES TANIMA MODELİNİN OLUŞTURULMASI 56
5.1 GİRİŞ 56
5.2 YSA Ses Tanıma Modeli 57
5.3 Hata Analizi 65
5.4 Komutlara Göre Konuşmacı Tanıma 67

5.5. Sonuçların Değerlendirilmesi..... 70

BÖLÜM 6 71
SONUÇLAR 71

BÖLÜM 7 74
KAYNAKLAR 74

SİMGELER VE KISALTMALAR

SİMGELER

y_{in}:	Ses Bölümü Örnekleri
x:	Giriş
y:	Çıkış,
F:	Aktivasyon Fonksiyonunu
P:	Nöron Giriş Sayısı
n:	Çıkış Sayısı
w:	Ağırlık
η:	Öğrenme Katsayısı
μ:	Momentum Katsayısı
w(k):	Bağlantı Ağırlığı
δ(k):	Ağırlık Değişiminin Eğim Bileşeni
α:	Sabit Bir Öğrenme Oranı
κ:	Öğrenme Katsayısı Artma Faktörü
φ:	Öğrenme Katsayısı Azaltma Faktörü
μ(k):	Momentum Hızı
κ_α :	Sabit Öğrenme Hızı Skala Faktörü
γ_α :	Sabit Öğrenme Hızı Üstel Faktörü
κ_μ :	Sabit Momentum Hızı Skala Faktörünü
φ_μ :	Sabit Momentum Hızı Azaltma Faktörü
γ_μ :	Sabit Momentum Hızı Üstel Faktörünü
λ :	Düzeltme Tolerans Parametresi
δ:	Hızlandırma Katsayısı
X:	Çok Boyutlu Giriş Vektörü
C:	Giriş Vektörü ile Aynı Boyutta Merkez Vektörü
σ:	Standart Sapma Değeri
t_i:	Komut Kelimelere Karşılık Gelen Test Çıkış Kodları

t_h: Komut Kelimelere Karşılık Gelen YSA Test Çıkış Kodları

n: Eleman Sayısı

KISALTMALAR

HF: Yüksek Frekans

LPC: Doğrusal Öngörüm Kodlaması

PCM: Darbe Kodu Modülasyonu

DPCM: Diferansiyel Darbe Kodu Modülasyonu

ADPCM: Uyarlanabilir Diferansiyel Darbe Kodu Modülasyonu

CELP: Kod Etkileşilim Doğrusal Önkestirim

AMDF: Ortalama Genlik Farkı Fonksiyonu

DFT: Ayrık Fourier Dönüşümleri

WFT: Pencerelenmiş Fourier Dönüşümleri

Hz: Hertz (frekans birimi)

FFT: Hızlı Fourier Dönüşümleri

LVQ: Öğrenme Vektör Nicelendirmesi Ağı

RBF: Radyal Tabanlı Fonksiyon Ağı

TDNN: Zaman Gecikmeli Yapay Sinir Ağı

dB: Desibel

OGFİ: Ortalama Genlik Farkı İşlevi

SSDC: Sonsuz Darbe Cevaplı

FEC: İleri Hata Düzeltme

ARQ: Otomatik Yeniden Gönderme

FFT: Hızlı Fourier Dönüşümü

HMM: Hidden Markov Modeli

DBD: Delta Bar Delta

HY: Hızlı Yayılım

MTH: Mutlak Toplam Hata

OMTH: Ortalama Mutlak Toplam Hata

BÖLÜM 1

GİRİŞ

Konuşma organlarının düzenli çalışmasıyla, anlamlı sözcükler ve tümceler oluşturmak için ağzımızdan çıkardığımız birimlere ses denir. İnsan ses yolu yaklaşık 1-7 cm uzunluğunda bir ucu kapalı boru şeklinde bir yapıdır. Bu borunun biçimi, ses organlarının devinimiyle değişir. Ağız, burun, boğaz boşluğundaki organlar aracılığıyla ciğerlerden çıkan hava, başka bir deyişle konuşmanın ham öğesi olan soluk, biçimlenir ve ses birimlerine dönüşür. Söz konusu organlar seslerin çıkışını gerçekleştirmek için kımıltılar yaparlar. Salt sesin konuşma sesine dönüşmesinde adı geçen organlar sesi sanki yoğurur, işler, konuşma sesi biçimine sokar. Ses telleri ötümlü seslerde belli aralıklarla açılıp kapanır. Bu açılıp kapanma erkeklerde saniyede ortalama 130 Hz kadınlarda ise 220 Hz frekansla tekrarlanır. Açılıp kapanma frekansı sabit olmayıp çıkarılacak sese ve vurguya göre değişir.

Bu konudaki çalışmalar 2. Dünya Savaşı yıllarına kadar uzanır. Bir suç unsuru olarak ses, aşağı yukarı bütün kriminal olaylarda kullanılmaktadır.

Ses tanıma ve tanımlama işlemi, bütün işitsel ve görsel duyuların kullanıldığı çok yönlü bir işlemdir. Bu işlemi, bilinmeyen bir sesin bir veya daha fazla bilinen sesle tanınması veya elenmesi amacıyla işitsel veya görsel olarak karşılaştırılması şeklinde tanımlanabilir. Bu olayın temel olarak dayandığı ilke; seslerin, kendi başına sahip olduğu karakteristikler ve özellikler yardımıyla çeşitli analiz teknikleri ve yöntemleri uygulanarak diğerlerinden ayırt edilmesidir.

Elde bulunan kayıt içerisinde yer alan konuşmaların sahibini bulmak ya da kimliğini ortaya çıkarmak için birtakım metot ve tekniklerden faydalanılmaktadır. Kullanılan bu teknikleri kısaca işitsel (aural) ve görsel (visual) teknikler denebilir. İşitsel teknik ile konuşma stili, lehçe, kekeleme, konuşma hataları v.b. dilsel

özellikler incelenir. Görsel teknikle ise genel format şekli ve pozisyonu, perde periyot düzeni, enerji dağılımı gibi özellikler incelenir.

"Ses izi" olarak tanımlanan Spektrogram çok yaygın olarak kullanılan konuşmacı tanıma metodudur. Aynı heceye karşılık gelen frekansların zaman içindeki değişimlerini incelemek yoluyla teşhis etme yöntemidir.

Tanımlama konusunda gerekli sayılan kelime sayısı konusunda ulusal bir standart mevcut değildir, doğruluk derecesi ise kaydın özelliğine ve yapılan işlemin başarı derecesine göre değişiklik gösterir.

1.1. Ses, Konuşma, Konuşmacı Tanıma Uygulamalarında ve Yapay Sinir Ağlarında Yapılan Çalışmalar

Yapay zeka ile abonenin ses analiz işleminin abonenin SIMKart aldığı operatörün yetkili servisinde yapılabileceğini öne süren Önder Türkoğlu, "Mobil Haberleşmede Ses Tanıma ve Eğitime Etkileri" isimli çalışması ile ses tanımaya yönelik bir çalışma sunmuştur (Türkoğlu 1998).

Rifat Kurban, Mehmet Tunçkanat ve Şeref Sağıroğlu, "Yapay Sinir Ağları ile Konuşmacı Tanıma" isimli çalışmalarında, konuşmacı tanımada YSA'nı kullanarak Levenberg-Marquardt (Levenberg 1944, Marquardt 1963) öğrenme algoritması ile 6 farklı kişinin yazı-bağımlı ve yazı-bağımsız konuşmalarının tanınmasında bir yaklaşım sunmuşlardır. Bu çalışmalarında %88 ile %94 oranında bir başarı sağlamışlardır (Kurban, Tunçkanat 2003).

Yük.Müh. Pelin Kuş, "Ses Sinyallerinin Düşük Hızda İletimi" isimli yüksek lisans tez çalışmasında NATO standardı olan 2400 bit/s ve 800 bit/s hızındaki ses kodlayıcılarını programlayarak çalıştırmış ve başarımları biri kadın biri erkek konuşmacı tarafından söylenen ve yaklaşık 3'er saniye süren iki farklı cümleyi denemiştir. Sonuç olarak 2400 bit/s hızındaki kodlayıcıyla sesin kalitesinde bozulma ve mekanik bir tını oluşmasına rağmen anlaşılabilir iletişim yapılabildiğini

gözlemiş, 800 bit/s hızındaki kodlayıcıda ise tatminkar bir sonuç alamamıştır (Kuş 1998).

A.Alpaslan Altun ve H.Erdinç Koçer "Güvenlik Alanına Yeni Bir Yaklaşım Biyometrik Sistemler" isimli çalışmalarında biyometrik teknolojisine genel bir bakış yapılmış, geliştirilen biyometrik sistemler anlatılmış ve biyometrik sistemlerin güvenlik alanındaki güvenirliği tartışılmıştır (Altun, Koçer 2003).

Halit Ergezer, Mehmet Dikmen ve Erkan Özdemir "Yapay Sinir Ağları ve Tanıma Sistemleri" isimli çalışmalarında tanıma sistemleri hakkında kısa bir bilgi vererek yapay zeka üzerine yapılabilecek bu alandaki çalışmaların günümüz teknolojinin gelişimine büyük katkı sağlayacağını dile getirmiştir (Ergezer, Dikmen 2003).

"Ses ile Kimlik Doğrulama" isimli Yüksek Lisans Semineri çalışmasında ses ile kimlik doğrulama konusu ele alınmış ve ses sinyalinin üretilişi ve modellenmesi anlatılmıştır. Ayrıca sistem performansının değerlendirilmesi, yanlış onaylama ve yanlış reddetme hataları ele alınmış ve sistem performansını etkileyen faktörler incelenerek mevcut ses tanıma sistemleri, insanlardaki ses tanıma sistemi ile karşılaştırılmıştır (Yaparoğlu 2002).

"Objelerin Bölütlenmiş Görüntüleri Kullanılarak Yapay Sinir Ağlarıyla Tanınması" isimli tez çalışmasında, bir yapay sinir ağı yaklaşımını C++ dilinde simule ederek obje tanıma uygulamaları yapılmıştır ve bu uygulamanın performansı incelenerek sistemin avantajları ve dezavantajları belirlenmeye çalışılmıştır (Demiray 1998).

"Mikroşerit Anten Parametrelerinin Yapay Sinir Ağları ile Analizi ve Tasarımı" isimli doktora tezinde, üç tip mikroşerit antenin (üçgen, dikdörtgen ve daire) karekteristik parametrelerinin hesabı ve bir tip mikroşerit antenin tasarımı YSA kullanılarak yapılmıştır. Bu çalışmada, YSA'ların eğitilmesinde ÇKP ve RTY ağ yapısı üzerinde 18 farklı öğrenme algoritması kullanılmış ve bu algoritmaların performansları karşılaştırılmıştır (Gültekin 2002).

"Speech Emotion Recognition Using Hidden Markov Models" isimli çalışmada hisli konuşma tanıma çalışması yapılmıştır. Bu çalışmada tanıma verimliliğini arttırmak için Gizli Markov Modeli kullanılmıştır (Nwe T.L., Foo s.W. 2003).

"HMM-Based Channel Error Mitigation and its Application to Distributed Speech Recognition" isimli çalışmada, dağıtılmış konuşma tanımı için kanal hatasının azaltılmasını Temel Gizli Markov Modeli kullanılarak yapılması önerilmiş ve HMM'ye baglı teknikler etkin bir şekilde kanal hatalarını azaltabileceği belirtilmiştir (Peinado A.M., Sánchez V. 2003).

"High Quality Speech Transformation Based on Linear Prediction Coding and Pitch Synchronization" isimli çalışmada, konusma transformasyonu metodu olarak pitch eşlemesi ile birlikte Linear Prediction Code (LPC) tekniği tavsiye edilmiştir ve konuşma, farklı ses renkleri ve tonları yada yüksek kalitede konuşma hızı için degiştirilebileceği ifade edilmiştir (Yang D.M. 2003).

"Yapay Sinir Ağları Modeli ile Ses Tanıma", farklı YSA algoritmaların ses tanımadaki performanslarının mukayesesi yapılarak ses tanımaya uygun olan algoritmanın belirlenmesi amaçlanmıştır. Ayrıca, bu çalışma ile YSA'lar ile ses tanıma problemine farklı bir yaklaşım getirilmiştir. Bu yaklaşım, YSA'ında eğitilen ses komut verilerinden elde edilen sonuçlarla, tanıma ve kontrol işlemini gerçekleştirmek için, hata analizi yapılarak elde edilen hata değerlerini kullanmayı hedeflemiştir (Gültekin D.2004).

Yukarıda, ses tanıma, konuşmacı tanıma, biyometrik sistemler, ses ile kimlik doğrulama, hisli konuşma tanıma, LPC Tekniğinin birkaç uygulaması ve YSA'ların hem ses tanıma hemde diğer mühendislik uygulamarı kısaca özetlenerek verilmiştir. Bugüne kadar yapılan çalışmalarda, ses tanıma probleminde Hopfield, Self Organizing Map (SOM), Radial Basis Function (RBF) gibi ağların ve Çok Katmanlı Perseptron (ÇKP) ağlarına ait algoritmaların kullanıldığı belirlenmiştir.

BÖLÜM 2

SESİN ÖZELLİKLERİ

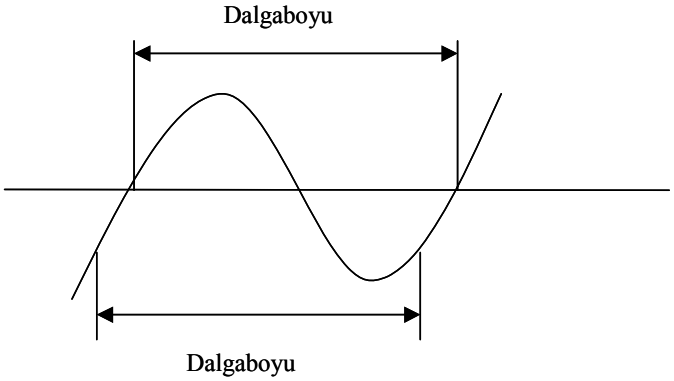
2.1. Ses Nedir?

Ses, insan kulağını etkileyerek işitme duyusu oluşturan hava molekülleri titreşimleri, ya da bunların neden olduğu ufak hava basınç değişimleri gibi, ya da bu fiziksel olayın neden olduğu işitsel izlenim gibi tanımlanır.

Ses fizyolojisi ile ilgili bazı önemli kavramları şöyle açıklanabilir:

Ses Dalgası : Ses sıvı, katı, gaz ortamlarında 20 Hz ile 20 KHz arasındaki insan kulağının algılayabileceği basınç değişiklikleri olarak tanımlanmaktadır. Bu frekans aralığındaki mekanik dalgalar işitme duyumuzu uyardıklarından, bizim için özellikle önemli olan ses dalgalarını oluştururlar.

İnsan kulağına bir ses dalgası geldiğinde kulak ses dalgasındaki basınç değişikliklerini sinirlerdeki itkilere çevirir ve bunlar beyinde duyulan sesler olarak yorumlanır.



Şekil 2.1. Basit bir sinüs dalgası

Buna en basit ses dalgasının sadece bir frekansı ve sabit bir genliği vardır. Buna sinüs dalgası adı verilir. Şekil 2.1'de basit bir sinüs dalgası grafiği görülmektedir.

Frekans : İki tepe arasındaki uzaklık dalga boyudur. Bir saniyede gözlenen dalga tepesi sayısına frekans denir. Frekans sesin tizliğini belirler. Saniyedeki çevrim (Cycle Per Second-CPS) veya Hertz (Hz) ile gösterilir. Düşük frekanslar bas sesler, yüksek frekanslar ise yüksek seviyeli seslerdir.

Genlik : Sesin diğer bir karakteristik özelliği genliğidir. Sesler yumuşak veya yüksek olurlar. Bu havada, havayı sıkıştırmak için kullanılan güce bağlı olan bir tür küçük veya büyük basınca karşılık gelir. Ses gücü veya seviyesi için desibel (dB) birimi kullanılmaktadır. Kulağın algılama özelliği ile ilgili yapılan çalışmalarda ses gücünün artması ile hissedilen ses artışının doğrusal olmadığı ve logaritmik bir ses şiddeti ile duyma olduğu anlaşılmıştır. Bu nedenle algılanan ses logaritmik bir büyüklüktür. Haberleşme sistem ve cihazlarının yapısı ve ölçü birimleri de bu sebepten logaritmik olarak düzenlenmiştir.

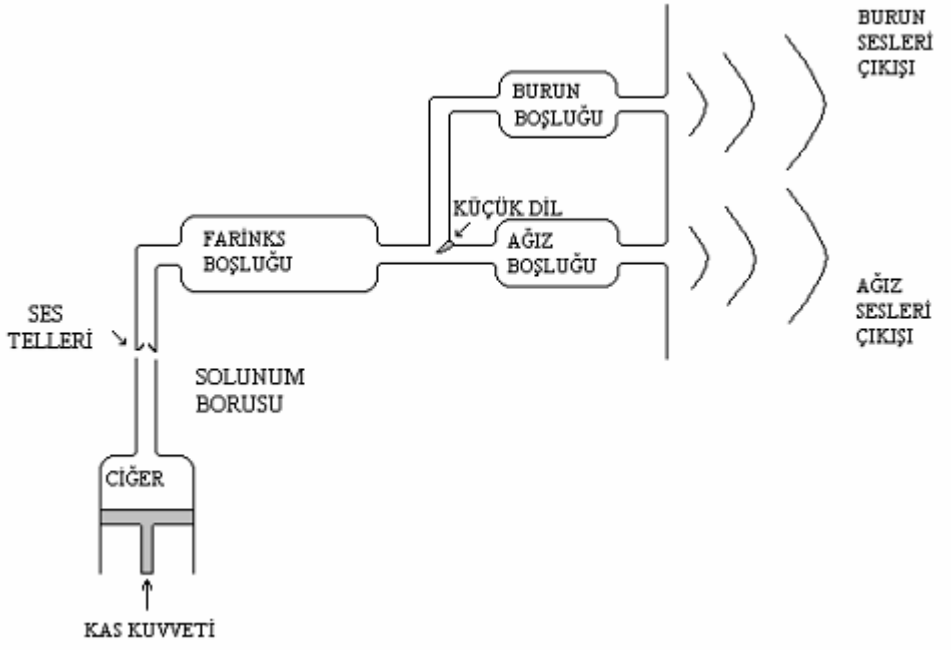
Gürültü : Periyodik olmayan titreşimlerdir. Kulağın, teknik duyumu bakımından sınırları zorlayan ve psikolojik rahatsızlık doğuran seslere gürültüdür.

2.2. Konuşma

Ses dalgası, ses üretim sistemini meydana getiren anatomik yapıların istemli hareketleri sonucunda oluşan, akustik bir basınç dalgasıdır. Bu sistemin ana bölümleri ciğerler, nefes borusu, gırtlak, boğaz, ağız boşluğu ve burun boşluğudur. Teknik terim olarak boğaz ve ağız boşluğu 'ses yolu' olarak tanımlanır. Dolayısıyla ses yolu, gırtlak çıkışından başlayıp, dudaklarda sona erer. Burun yolu ise damaktan başlar burun deliklerinde sona erer. Ses üretimi için kritik olan anatomik yapılar, ses telleri, damak, dil, dişler ve dudaklardır. Ağız iyice açıldığı zaman ağız boşluğunun arka tarafında duran damağın yumuşak uzantısına 'küçük dil' denir. Ses yolunu

oluşturan bu anatomik yapılar, farklı pozisyonlar alarak değişik sesleri oluştururlar (Aydın 2005).

Ses yolunun yapısı Şekil 2.2'de görülmektedir.



Şekil 2.2. Ses yolunun yapısı

2.2.1. Ses Yolunun İşlevi ve Akustik Analiz

İnsan sesinin dalga şeklini incelediğimizde, fiziksel sistem zamana bağlı olarak değiştiği için, dalga şeklinin de zamanla değiştiğini görürüz. Konuşma sesleri, kısa süreler boyunca benzer akustik özellikler gösteren ses parçalarına ayrılabilir.

Ses sinyallerinin zamana bağlı dalga şekillerine bakılarak, sinyal periyotları, yoğunlukları, süreleri ve her bir ses parçasının sınırları tespit edilebilir. Ancak, ardarda gelen sesler birbirlerini etkilerse bu sınırları belirlemek mümkün olmayabilir. İnsanların ses üretme ve sesi algılama sistemlerindeki organların yapılarından kaynaklanan bazı sınırlamalardan dolayı, ses dalga şeklinde değişime sebep olan

yapıların pozisyonlarının kısa zaman aralıklarında değişmediği farzedilebilir (Kuş 1998).

2.2.2. Ses Sisteminde Uyarım

Sesin başlıca özelliği, uyarım şeklidir. İki temel uyarım şekli vardır. Bunlar ötümlü ve ötümsüz uyarımlardır. Bunlardan başka ötümlü uyarım, ötümsüz uyarım ve sessizliğin birleşimi sonucunda oluşan dört tip uyarım daha vardır; karışık, patlamalı, fısıltı ve sessiz uyarım (Kuş 1998).

Ötümlü sesler, havanın nefes borusundan veya ses telleri arasından geçmeye zorlanmasıyla gırtlakta üretilir. Ötümsüz sesler, hava akımının sıkıştırılmış ses yolundan geçmeye zorlanması ile üretilirler. Birbiri ardınca, ötümlü-ötümsüz olarak üretilen seslere 'karışık sesler' denir. Patlamalı seslerde ses yolunun son kısmı kapatılarak içeride basınç oluşması sağlanır. Basınçlı hava, aniden bırakılarak bu sesler üretilir.

2.3. İşitme

İşitme sistemi bizim işitme duyumuzdan sorumludur. Bu sistem akustik ses dalgalarını alır ve onları beyin tarafından yorumlanan sinir kodlarına dönüştürür. İnsanın dış dünya ile ilişki kurmasını sağlayan beş duyu organından biri de kulaktır. Kulak ses titreşimlerini sezer ve sinir uyarılarına çevirerek beyne yollar. Gerçek ses algılaması beyinde olur. Kulak ve beyin arasındaki ilişki, seslerin algılanmasını, işlemlerden geçmesini ve seslerin taklit edilebilecek şekilde öğrenilmesini sağlar. İşitme sistemi dört ana bölümden oluşur: dış kulak, orta kulak, iç kulak ve beyine giden sinir yolları.

İşitmenin gerçekleşebilmesi için;

a)Sesin olması,

b)Sesin kulağa ulaşması,

c)O sesin insan kulağının alabileceği frekans ve şiddet sınırları içinde olması,

d)Sesin kulaktaki dış, orta ve iç bölümleri aşması,

e)Sesin işitme merkezine ulaşması ve merkezce algılanması, gerekmektedir.

BÖLÜM 3

SESİN KODLANMASI

3.1 Giriş

Ses sıkıştırma frekans veya zaman baz alınarak iki farklı metotla incelenebilir. Kullanılan sıkıştırma algoritmasının tipi ihtiyaç duyulan fonksiyonelliğe ve istenen ses kalitesine göre seçilir.

Her iki yöntemde de sıkıştırma, fazlalığın atılması yöntemidir. Ses sıkıştırma da kullanılmayan kısımların atılması hata ya da distorsiyon olarak nitelendirilir.

Pek çok sıkıştırma tekniğinde amaç transfer edilen datanın ve saklama alanının azaltılmasıdır. Pek çok yüksek kaliteli teknik 64 kBit/saniye gibi yüksek değerlerde sıkıştırma yaparken 1'e 24 oranında sıkıştıran teknikler de vardır. Ancak sıkıştırmanın fazla olması, elde edilen datanın gerçek zamanlı olarak işlenebilmesinde ve konuşma tanımada kullanılmasını zorlaştırır.

3.2. Kodlayıcı Tipleri

Analog ses sinyallerini dijital formata çeviren kodlayıcılar dalga ya da ses kodlayıcı olabilirler. İkisi arasındaki fark, biri tamamen insan konuşmasına göre optimize edilmiş, ona göre hazırlanmıştır, diğeri ise tüm ses tipleri için geçerlidir.

3.2.1. Doğrusal Kuantizasyon Teknikleri

3.2.1.1. Darbe Kodu Modülasyonu (Pulse Code Modulation - PCM)

Sesin doğrusal kuantizasyon yöntemi ile dijital formata dönüştürüldüğü en basit

yöntemdir. Temel olarak 8 kHz'de ses sinyalini örnekleyerek kuantize eder. Çıkış akışı yaklaşık 64 kBit/saniyedir. Bu sebeple bu çeşit kodlama gerçek zamanlı sistemlerde, yüksek bant genişliği isteyeceği, hafıza ve kaynak sıkıntısı yaratacağı için pek uygun değildir.

3.2.1.2. Diferansiyel Darbe Kodu Modülasyonu (Differential Pulse Code Modulation - DPCM)

Darbe Kodu Modülasyonuna göre daha etkili bir yöntemdir çünkü ses sinyali içindeki gereksiz kısımları, daha sonra önceki ve sonrakinden örneklenebilecek şekilde atar. Böylece sıkıştırıcının tek yaptığı birbiri ardı sıra gelen örneklerdeki farkı belirtmektir. Çözme işlemi sırasında bu sinyaller yeniden oluşturulur.

3.2.1.3.Uyarlanabilir Diferansiyel Darbe Kodu Modülasyonu (Adaptive Differential Pulse Code Modulation - ADPCM)

Uyarlanabilir Diferansiyel Darbe Kodu Modülasyonu, 32 kBit/saniye gibi oranlarda çok yüksek ses kalitesi sağlayabilir. 16, 24,32 ve 40 kBit/saniyelik bit akış oranlarında çalışacak şekilde standart hale gelmiştir. ADPCM algoritma olarak PCM'den farklıdır çünkü örneklenmiş ses sinyalinin kuantize edilmesinin yerine önkestirilen ve kuantize edilen sinyal arasındaki farkı kuantize eder. İyi bir önkestirimde gerçek sinyal ile tahmini sinyal arasındaki fark çok küçük olacaktır ve bu da daha düşük bit akış hızı anlamına gelecektir. Arkasında çalışan kuantizer tek tip değildir ve farklı sinyal modellerinde kullanılmak üzere optimize edilebilir.

Sinyalin yeniden üretilmesi kuantize edilmiş farkın tahmini sinyale eklenmesiyle bulunur. Bu sayede orijinal sese çok yakın bir sinyal elde edilmiş olur. ADPCM metodu sadece 2:1 gibi çok düşük bir sıkıştırma sunsa da data sıkıştırma metodlarıyla beraber kullanılarak 4:1 oranına ulaşılabilir.

3.2.2. Vektörel Kuantizasyon Teknikleri

Vektörel kuantizasyon, datayı doğrusal yerine vektörel olarak kodladığı için daha yüksek performans sağlar.

3.2.2.1. Kod Etkileşilim Doğrusal Önkestirim (Code Linear Prediction - CELP)

Kod Etkileşilim Doğrusal Önkestirim, insanı bir filtre ve tahrik kaynağından ibaret görerek yüksek sıkıştırma oranlarına ulaşır. Sıkıştırmayı filtrenin kod listelerini yaparak gerçekleştirir. Konuşma sinyali geldikçe filtre kendini insan sesinin karakteristiklerine göre adapte eder. Eğer daha yüksek sıkıştırma istenirse Kod Etkileşilim Doğrusal Önkestirim kodu içindeki duyulamayan frekanslara ait data atılır.

Kod Etkileşilim Doğrusal Önkestirim, domeninde çalışır. 4800 bps ile yaklaşık 13:1 sıkıştırma sağlar. 100 milisaniye civarında bir gecikmesi vardır ve işlem karmaşıklığı 16.5 MIPS'tir (Millions of Instructions Per Second).

Önce gelen sesten öncül parametreler belirlenir. Daha sonra kod listesinde buna uygun, minimum hata veren tahrik modelinin belirlenmesi takip eder. En sonunda da bu modele gelen ses datası uydurularak parametreler üretilir.

Gelen parametreler, kod kitabından uygun parametrelerle birleştirilerek tahrik modeline yerleştirilir. Daha sonra buradan spektral parametreler çıkarılır ve ses yeniden üretilir.

3.2.2.2. VSELP

VSELP, oldukça iyi düzenlenmiş bir kod listesi kullanır. İşlem karmaşıklığını azaltmak ve kanal hatalarında daha doğru sonuç vermesi için kodlar birbirlerine birbirlerinin kuyruğu gibi eklenmiştir. VSELP modeli, 8 kHz'de örnekler. Bunun dışında çalışması Kod Etkileşilim Doğrusal Önkestirime çok benzer.

3.2.2.3. Doğrusal Öngörüm Kodlaması (Linear Predictive Coding - LPC)

Düşük hızda ses kodlayabilmek için öncelikle çok iyi bir ses üretim modeline sahip olmak gerekir. Gönderici, ses çıktısını analiz eder ve model parametrelerini elde eder. Bu parametreler benzer ses dalgası üretmek için sentezlemede kullanılacağı alıcıya gönderilir. Bu yaklaşımın doğru olması için iyi bir kaynak modellemesinin yapılmış olması gerekir. Günümüzde ses üretimi için iyi modeller mevcuttur. Dolayısıyla bu yaklaşım düşük hızda ses kodlamada geniş olarak kullanılmaktadır.

Hangi lisanın kullanıldığına bağlı olmaksızın her insanda ses üretimi için aynı mekanizma kullanılır. Bununla birlikte bu mekanizma, çıktıların durumunu kısıtlayan bir takım fiziksel kurallara sahiptir. Dolayısıyla ses, bir model olarak analiz edilebilir. Bu modelin parametreleri hesaplanarak alıcıya gönderilir. Alıcıda ise bu parametreler kullanılarak ses sinyali tekrar üretilir. Bu analiz/sentez yaklaşımı ilk kez Homer DUDLEY tarafından Bell laboratuarlarında ortaya çıkmıştır. Homer DUDLEY kanal ses kodlayıcılarını da ilk geliştiren kişidir (Dudley 1939).

En basit ses sentezleme modeli şekil 3.1 de görülmektedir.



Şekil 3.1. Ses sentezleme modeli

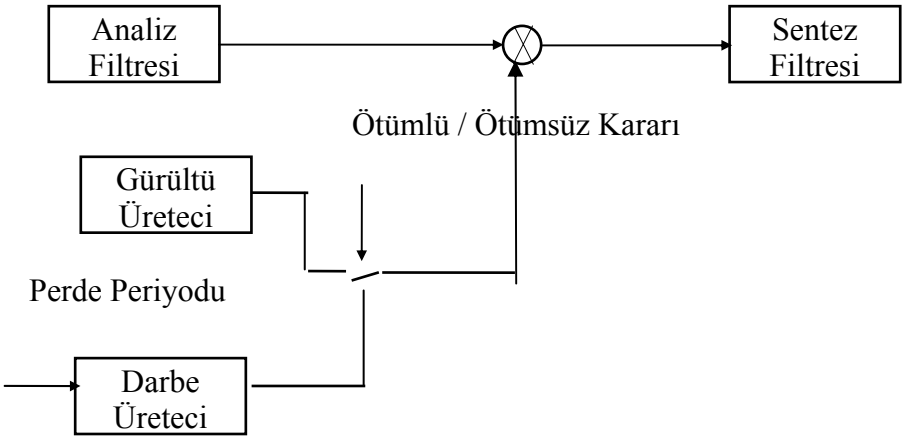
Göndericide ses çerçevelere ayrılır. Her bölüm için ses sinyali, uyarım sinyali ve ses yolu filtresinin parametreleri hesaplanır. Bazı prosedürlerde, uyarım sinyali için bir model alıcıya gönderilir. Uyarım sinyali daha sonra alıcıda sentezlenir ve ses yolunun uyarım sinyali olarak kullanılır. Diğer prosedürlerde ise uyarım sinyalinin kendisi analiz ve sentezde kullanılır.

Kanal Ses Kodlayıcıları; Kanal ses kodlayıcılarında, giriş ses parçalarının her

biri bant geçirgen filtre bankaları kullanılarak analiz edilir. Bu filtreler, analiz filtreleridir. Her filtrenin çıkışı sabit aralıklar için ölçülür ve alıcıya gönderilir. Sayısal kestirimlerde, enerji kestirimi, filtre çıktısının karesinin ortalama değeri olarak ölçülebilir. Analog uygulamada ise zarf dedektörünün örneklenmiş çıktısından kestirim yapılır. Genellikle saniyede 50 kez kestirim yapılır. Filtre çıktısının kestirimine, ötümlü, ötümsüz bölgelerde karar verilir. Gönderici alıcıya aynı zamanda perde periyodu ile ilgili bilgileri de gönderir.

Alıcı ise ses yolu filtresi bant geçirgen filtre grubu ile ifade edilir. Bu filtrelere sentez filtresi denilip, analiz filtresine oldukça benzerdir.

Sesin ötümlü ve ötümsüz olma durumlarına bağlı olarak, sentez filtresinin girişine, periyodik darbe sinyali veya gürültü üreteci bağlanır. Darbe sinyalinin periyodu, analiz bölümünde tahmin edilen perde periyodudur. Giriş, tahmin edilen enerji değerine göre ölçeklendirilir. Sentez blok şeması şöyledir.



Şekil 3.2. Kanal Ses Kodlayıcısı

İlk dört formant ses analiz ve sentezinde yeterlidir. Her formantın bant genişlikleri tahmin edilir. Alıcıda uyarım sinyalleri, formant sıklıkları ve bant genişlikleri ayarı yapan filtrelerden geçirilir. Uyarım sinyalinin öneminin anlaşılması ses kodlamada önemli bir gelişme olmuştur. Ötüm bilgisine böylece daha çok önem

verilmeye başlanmıştır. Ötümlü olarak uyarılacak kanal ses kodlayıcısında, ses önce dar bantlı alçak geçirgen filtreden geçirilir. Alçak geçirgen filtrenin çıkışı örneklenip alıcıya gönderilir. Alıcıda, bu sinyalin, yüksek dereceli harmoniklerini üretilerek, uyarım sinyali olarak kullanılır. Ötümlü uyarımda, hem perde periyodu kestirimi hem de ötümlü/ötümsüz ayrımı yapılmaz. Ötümlü veya ötümsüz ayrımı kesin olarak yapılamayan bölümler oldukça az olduğundan, kalite oldukça iyidir.

3.2.2.3.1. Ötümlü/Ötümsüz Belirlenmesi

Ötümlü sinyaller, diğerlerine göre daha büyük genlik değerine sahiptirler, daha fazla enerji içerirler. Ötümsüz sinyaller ise daha yüksek frekanslara sahiptirler. Ancak ortalamada her iki sinyal tipi de yakın değerler içerirler. Dolayısıyla bir ses sinyali segmentinin Ötümlü/ötümsüz kararını vermek için sinyalin sıfır çizgisini kaç kere kestiğine bakmamız gerekir. Kullanmış olduğumuz algoritmada sinyal önce 1 kHz Low-Pass filtreden geçirilmektedir. Bu sayede elde edilen sinyalin arka plandaki gürültü ile karşılaştırılması sonucu sesin ötümlü olup olmadığına karar verilir. Ancak iki ötümlü segment arasında kalmış olan ötümsüz parçaların arada kaybedilmemesi için komşu segmentlerde bu hesaplamada kullanılır.

3.2.2.3.2. Pitch Periyodunun Belirlenmesi

Analiz aşamasının en hesaba dayalı kısmıdır. Yıllar boyunca pek çok farklı algoritma denenmiştir ancak segment boyu azalıp örnek sayısı azaldıkça ve gürültü miktarı arttıkça bu hesaplamanın doğruluğu da azalmıştır. LPC10 algoritması içinde Ortalama Genlik Farkı Fonksiyonu (AMDF; Average Magnitude Difference Function) kullanılmıştır.

Ortalama Genlik Farkı İşlevi şöyle tanımlanır.

$$OGF\dot{I}(P) = \frac{1}{N} \sum_{i=k0+1}^{N+k0} \left| y_i - y_{i-p} \right| \qquad (3.1)$$

Eğer sinyal P_o ile periyodikse birbirlerinden P_o örnek kadar uzaklıkta yaklaşık aynı değerler alacaklarından, OGFİ'leri en küçük olur. Ötümlü seslerde P=Po değerinde olduğu gibi, ötümsüz seslerde de OGFİ en küçük değeri alır.

OGFİ sadece perde periyodunu bulmada kullanılmaz aynı zamanda ötümlü/ötümsüz ayrımı için de kullanılır. Ötümsüz seslerde çerçeve boyunca hesaplanan OGFİ oldukça düşük değerler alır. Ötümlü seslerde ise OGFİ belirli bir miktarın üzerinde seyreder, sadece periyot değerlerinde belirgin düşüşler gösterir. (K

Perde periyodunu tahmin etmek için tam periyot alanı taranmaz. İnsanların konuşma ve duyma periyotlarına ait aralık incelenir. LPC10 algoritması perde periyodunu 2.5 ve 19.5ms arasında hesaplanmıştır. Yani saniyede 8000 örnekle periyot değeri, 20≤P≤160'dır.

3.2.2.3.3. Ses Borusu Modeli Filtresi

Doğrusal öngörüm kodlamasında ses yolu, doğrusal bir filtre ile modellenir. Göndericide, sesin incelenen bölümü için, hata karesinin ortalamasından analiz ile en uygun filtre katsayıları hesaplanır. Eğer, analiz edilen ses bölümü örnekleri $\{y_{in}\}$ ise; e_n^2 'yi en küçük yapan $\{a_i\}$ 'leri seçmemiz gerekir.

$$e_n^2 = \left(y_n - \sum_{i=1}^{M} a_i\, y_{n-i} - G \in \right)^2 \tag{3.2}$$

Eğer e_n^2 'nin beklenen değeri $\{a_i\}$ katsayılarına göre türevini alırsak, M adet eşitlik elde ederiz.

$$\frac{\delta}{\delta a_j} E\left[y_n - \sum_{i=1}^{M} a_i\, y_{n-i} - G \in_n \right]^2 \tag{3.3}$$

$$\Rightarrow \; -2E\left[\left(y_n - \sum_{i=1}^{M} a_i\, y_{n-i} - G \in_n \right) y_{n-j} \right] = 0 \tag{3.4}$$

$$\Rightarrow \sum_{i=1}^{M} a_i\, E\left[y_{n-i}\, y_{n-j} \right] = E\left[y_{n-i}\, y_{n-j} \right] \tag{3.5}$$

$j \neq 0$ için $E\{\in_n y_{n-j}\} = 0$ olduğundan $E\{y_{n-i}\, y_{n-j}\}$' nin hesaplanması gerekir. Bunu hesaplamak için iki algoritma mevcuttur.

1. Özilinti yaklaşımı
2. Özkovaryans yaklaşımı

Özilinti yaklaşımında $\{y_n\}$'nin durağan oluğu varsayılarak ,

$$E\{y_{n-i}\, y_{n-j}\} = R_{yy}\{|i-j|\} \tag{3.6}$$

Aynı zamanda, analizini yaptığımız ses bölümü dışındaki örneklerin sıfır olduğu farzedilir. Böylece,

$$R_{YY}(k) = \sum_{n=n0+1+k}^{n0+N} y_n\, y_{n-k} \tag{3.7}$$

Bu M adet eşitlik, matris formunda da yazılabilir.

$RA = P$ ve

$$R = \begin{vmatrix} R_{YY}(0) & R_{YY}(1) & R_{YY}(2) \ldots & R_{YY}(M-1) \\ R_{YY}(1) & R_{YY}(0) & R_{YY}(1) \ldots & R_{YY}(M-2) \\ R_{YY}(2) & R_{YY}(1) & R_{YY}(0) \ldots & R_{yy}(M-3) \\ \vdots & \vdots & \vdots & \vdots \\ R_{YY}(M-1) & R_{YY}(M-2) & R_{YY}(M-3) \ldots & R_{YY}(0) \end{vmatrix}$$

$$A = \begin{vmatrix} a_1 \\ a_2 \\ a_3 \\ \vdots \\ a_M \end{vmatrix} \tag{3.8}$$

ve $A = R^{-1} P$ olarak filtre katsayıları bulunur. R sadece simetrik değil, aynı zamanda köşegen elemanları da aynıdır. Mesela, esas köşegende $R_{yy}(0)$ elemanı vardır. Ayrıca, esas köşegenin alt ve üst köşegenlerinde de sadece $R_{yy}(1)$ elemanı vardır. Bu tür matrislere 'Toeplitz Matris' denir. Toeplitz matrislerinin tersini almak için bir takım algoritmalar vardır. Bu algoritmalardan biri Levinson - Durbin algoritmasıdır.

Levinson - Durbin algoritması M inci dereceden bir filtre katsayılarını bulmak için M inci dereceye kadar olan tüm filtrelerin hesaplamalarını yapar. Bununla

birlikte, filtre katsayıları bulunurken, yansıma katsayıları (veya kısmi korelasyon ya da PARCOR katsayıları) olarak bilinen k_i 'ler de üretilir.

Aşağıdaki gösterimlerde filtre dereceleri üst simge olarak yazılmıştır. Yani 5 inci dereceden bir filtrenin katsayısı $\{a_i^5\}$ olarak gösterilir. Algoritma aynı zamanda ortalama hatanın tahmin edilen değerini de hesaplar. E [e_n^2] ortalama hata M inci dereceden filtre kullanılarak E_m ile ifade edilir. Algoritma şöyledir:

1. $E_0 = R_{YY}(0)$, $i = 0$ (3.9)

2. $i = i + 1$ (3.10)

3. $k_i = \left[\sum_{j=1}^{i-1} a_j^{i-1} R_{yy}(i+j+1) - R_{yy}(i) \right] / E_{i-1}$ (3.11)

4. $a_j^i = k_i$ (3.12)

5. $a_j^i = a_j^{i-1} + k_i + a_{i-j}^{i-1}$ $j = 1,2 \ldots i-1$ (3.13)

6. $E_i = \left(1 - k_i^2\right) E_{i-1}$ (3.14)

7. $i < M \quad \Rightarrow$ 2 nci maddeye dön. (3.15)

Ötümlü bir sesin tam olarak yeniden üretilebilmesi için ses yolu filtresinin derecesinin yeterince yüksek olması gerekir. Genellikle on ve ondan fazladır. Eğer filtre SSDC(Sonsuz Darbe Cevaplı) ise katsayılardaki hata, kararsız bir yapıya sebep olabilir. Bu, özellikle doğrusal kestirim kodlaması için gerekli olan yüksek dereceli filtreler için geçerlidir. Filtre katsayılarının alıcıya gönderilebilmesi için nicemlenmesi gerekir. Yani, nicemleme hataları da bu katsayıları etkileyip bir kararsızlığa yol açabilir.

Yansıma katsayılarından filtre katsayıları hesaplanabilir. Yansıma katsayılarının büyüklükleri daima bir'den küçüktür. Bu yüzden kararlı bir yapıdır. Dolayısıyla, $\{a_i\}$ katsayılarını nicemleyip alıcıya göndermektense $\{k_i\}$ katsayılarını nicemleyip gönderilir. Elde edilecek değerler, en fazla bir olduğu için analiz ve

sentez için yüksek dereceden filtreler kullanılabilir. Ses sinyalinin kararlı olmadığını düşünürsek, filtre katsayılarını hesapladığımız eşitlikler de değişir.

$E\{y_{n-i}\ y_{n-j}\}$ hem i , hem de j 'nin işlevidir.

$$C_{ij} = E\{y_{n-i}\ y_{n-j}\} \tag{3.16}$$

$$CA = S$$

$$C = \begin{vmatrix} c_{11} & c_{12} & c_{13} & \cdots & c_{1M} \\ c_{21} & c_{22} & c_{23} & \cdots & c_{2M} \\ \vdots & \vdots & \vdots & & \vdots \\ c_{M1} & c_{M2} & c_{M3} & \cdots & c_{MM} \end{vmatrix}$$

$$S = \begin{vmatrix} c_{10} \\ c_{20} \\ c_{30} \\ \vdots \\ c_{M0} \end{vmatrix} \tag{3.17}$$

$$c_{ij} = \sum_{n=n0+1}^{n0+N} y_{n-i}\ \ y_{n-j} \tag{3.18}$$

Burada, seçilen ses bölümünün dışındaki örneklerin sıfır olduğu kabulü yapılmaktadır. Bunun anlamı, C matrisini oluştururken, önceki ses bölümlerine ait örnekler de kullanılır. Bu metot, 'Kovaryans Metodu' dur.

C matrisi simetriktir ama Toeplitz değildir. Dolayısıyla filtre katsayılarının çözümünde Levinson-Durbin metodunu kullanamayız. Bu durumda eşitlikler, Cholesky Ayrışımı kullanılarak çözülür.

LPC- 10 Algoritması, yansıma katsayılarını hesaplamak için kovaryans metodu kullanılır. Ötümlü/ötümsüz kararını güncelleştirmek içinse, yansıma katsayıları kullanılır. Genellikle ötümlü sesler için ilk iki yansıma katsayısı bire yakındır. Eğer ilk iki yansıma katsayısı küçük değerlere sahipse sesin ötümsüz olduğuna karar verilebilir.

3.3. Kanal Kodlama

Analogdan sayısala çevrilen sinyale kanaldaki bozulmaları telafi etmek üzere birtakım eklemeler yapılır. Bu eklerin gönderilmek istenen mesaj ile ilgili bir ilgisi olmadığı için gereksiz (redundancy) olarak isimlendirilir. Bu ekler sayesinde alınan sinyalin kanaldan geçerken hataya uğradığı anlaşılabilir, hatta bu hatalar düzeltilebilir. Bu amaçla gönderilecek sayısal sinyal üzerinde yapılan işlemlere kanal kodlama işlemi adı verilir.

Kanal kodlamanın 2 temel yolu vardır. Birincisinde eklenen parite bitleri ile gelen kodda hata olup olmadığı anlaşılır. Ancak hata düzeltme yapılamaz. İkincisinde ise eklenen parite bitleri ile kodda hem hata olup olmadığı anlaşılır, hem de hata varsa bu hata hata düzeltme kapasitesi çerçevesinde düzeltilir. Buna İleri Hata Düzeltme (FEC-Forward Error Correction) denir.

3.3.1. Otomatik Yeniden Gönderme (Automatic Repeat Request - ARQ)

Kanal kodlamanın hatayı düzeltmeye değil de sadece tespit etmeye yönelik olduğu sistemlerde mesajı yollayan tarafa mesajın hatalı alındığı ile ilgili uyarı gönderilebilir. Buna otomatik yeniden gönderme metodu denir.

Bu tür sistemlerde alıcı her doğru aldığı mesaj paketi için ACK (Acknowledgement), her hatalı aldığı mesaj paketi için ise NAK (Negative Acknowledgement) gönderir. Böylece hatalı alınan mesajların tekrarı sağlanarak hatasız mesaj alınması zorlanır.

ARQ metodu FEC metoduna göre çok daha basittir. Uygulaması kolaydır. FEC metodunda ise yeniden gönderme ile ilgili herhangi bir sistem tasarımı yapmaya gerek olmaz. FEC kanaldaki ACK-NAK mesajlarının iletilmesindeki gecikmelerden etkilenmez. Ayrıca, eğer kanalda çok fazla hata oluyor ise ARQ metodu çok fazla yeniden yollama yapılmasını gerektirir. Halbuki FEC metodunda böyle bir durum söz konusu olmaz.

3.3.2. Kodlama Oranı

Blok kodlama tekniğinde, k bitten oluşan mesaj kodu n bitten oluşan bir koda dönüştürülür. Yani (n-k) adet parite biti eklenir. Bu koda (n,k) kod denir. Veri bitlerinin, eklenen bitlere (k/n) oranına ise kodlama oranı denir. Kodlama oranı ne kadar küçük olursa hata düzeltme o kadar iyi olur. Ancak o oranda da kullanılacak band genişliği artar.

3.3.3. Basit Parite Kontrol Kodu

Tekli parite kontrol kodu, bir mesaj kodunun bitlerinin modülo 2 aritmetiğine göre toplanarak çıkan sonucun mesaj koduna eklenmesinden ibarettir.

000
001
1010
0011
1100

Alıcı tarafında gelen kodun bitleri (parite kontrol biti dahil) modülo 2 aritmetiğine göre toplanır ve eğer sonuç 0 ise kod hatasız, sonuç 1 ise kod hatalı kararı verilir. Hatalı alınan kodun düzeltilebilmesi mümkün değildir.

3.3.4. Hata Düzeltme

Hata düzeltme için en önemli parametre en küçük Hamming mesafesidir. Hamming mesafesi, iki kodun birbirinden farklı olan elemanlarının sayısıdır.

İkili sistemde bir kodun Hamming mesafesi o kodun 0'dan farklı olan elemanlarının sayısıdır. Örneğin; 10101 için Hamming mesafesi 3'tür.En küçük Hamming mesafesi ise bir kod topluluğunda Hamming mesafesi en küçük olan kodun mesafe değeridir.

10101, 11110, 01110 kodlarını ele aldığımızda bu kodların Hamming mesafeleri sırasıyla 3, 4 ve 3'tür. Bu kod topluluğu için en küçük Hamming mesafesi 3 olur.

Bir kod topluluğunda kaç adet hata düzeltilebileceği en küçük Hamming mesafesine bağlıdır. Düzeltilebilecek hata sayısını t, en küçük Hamming mesafesini d ile ifade edersek; $t = (d - 1) / 2$ olur. Yani önceki paragrafta verilen örnek için düzeltilebilecek hata sayısı 1'dir.

3.4 Fourier Analizi

Fourier'in fonksiyonları sinüs ve kosinüsün bir lineer kombinasyonu olarak temsil etmesi, hem diferansiyel denklemlerin analitik ve sayısal çözümlerinde hem de haberleşme işaretlerinin analizi ve düzeltilmesinde yaygın olarak kullanıldı. Fourier ve dalgacık analizi arasında çok güçlü bağlar vardır .

3.4.1 Fourier Dönüşümleri

Fourier dönüşümünün faydası, zaman tanım kümesindeki bir işaretin frekans içeriğini analiz etme kabiliyetinde yatar. Dönüşüm, ilk olarak tanım kümesi zaman olan bir fonksiyonu, tanım kümesi frekans olan bir fonksiyona çevirmek suretiyle çalışır. O zaman sinyalin frekans içeriği incelenebilir. Çünkü dönüştürülen fonksiyonun Fourier katsayıları, her frekans değerinde sinüs ve kosinüs fonksiyonlarının her birinin katkısını temsil eder. Ters Fourier dönüşümü de, verinin frekans tanım kümesinden zaman tanım kümesine dönüştürülmesini gerçekleştirir.

3.4.2 Ayrık Fourier Dönüşümleri (DFT)

Ayrık Fourier dönüşümü (DFT) fonksiyonun sonlu sayıdaki örnek noktasından Fourier dönüşümünü tahmin eder. Örnek noktaların, diğer zamanlarda işaretin neye benzediğini tipik olarak gösterdiği farzedilir.

Ayrık Fourier dönüşümü (DFT), sürekli Fourier dönüşümünün sahip olduğu simetri özelliklerinin hemen hemen aynısına sahiptir. Ayrıca, ters ayrık Fourier dönüşümü, ayrık Fourier dönüşümü için olan formülü kullanarak kolayca hesaplanır. Çünkü iki formül hemen hemen özdeştir.

3.4.3. Pencerelenmiş Fourier Dönüşümleri (WFT)

Eğer $f(t)$ periyodik olmayan bir işaretse, periyodik fonksiyonlar olan sinüs ve kosinüsün toplamı işareti doğru olarak temsil etmez. Sinyali periyodik yapmak için onu yapay olarak uzatabiliriz. Fakat bu durum uç noktalarda ek süreklilik gerektirir. Pencerelenmiş Fourier dönüşümü (WFT), periyodik olmayan işaret daha iyi nasıl temsil edilir problemine bir çözümdür. Pencerelenmiş Fourier dönüşümü (WFT), tanım kümesi aynı anda zaman ve frekans olan işaretler hakkında bilgi vermek için kullanılabilir.

Pencerelenmiş Fourier dönüşümü (WFT) ile, giriş işareti olan $f(t)$ bölümlere ayrılır ve her bölüm frekans açısından ayrı ayrı analiz edilir. Eğer işaret keskin, sivri uçlu değişimlere sahipse, uç noktalarda kesitler sıfıra yakınsayacak şekilde giriş verisi pencerelenir. Bu pencereleme işlemi, ortasından ziyade aralığın uç noktalarına yakın kısmına daha az önem veren bir ağırlık fonksiyonu aracılığı ile gerçekleştirilir. Pencerenin etkisi, işareti zamanla sınırlamaktır.

3.4.3.1. Fourier Dönüşümleri

Bir fonksiyonu örneklerle yaklaşık olarak temsil etmek ve Fourier integralini ayrık Fourier dönüşümü ile temsil etmek için, mertebesi örnek nokta sayısı olan *n*'e eşit bir matrisi uygulamak gerekir. Bir *n* x *n* matrisi bir vektörle çarpmak, n^2 mertebesinde aritmetik işleme neden olduğu için, örnek nokta sayısı arttığında problem çabucak daha kötü bir hale gelir. Ama eğer örnekler düzgün yerleştirilmişse, o zaman Fourier matrisi birkaç matrisin çarpımı şeklinde çarpanlara ayrılabilir ve sonuçta oluşan bu çarpanlar, toplam nlogn mertebesinde aritmetik işleme neden

olacak şekilde bir vektöre uygulanabilir. Bu Hızlı Fourier Dönüşümü (FFT) diye isimlendirilir.

3.5 Ses Kodlamanın En Çok Kullanıldığı Alan Ses Tanıma

Ses tanıma teknolojisi ile ülkemizde ticari manada ciddi olarak ilgilenen pek fazla firma yoktur. Bu konuda daha çok çeşitli üniversitelerde lisansüstü tezleri şeklinde çalışmalar yapılmaktadır. Yurt dışında ise, özellikle de Amerika'da bu konuda çalışan pek çok firma vardır. Amerika'da 1994 den beri 1250 civarında kuruluş bu konu ile ilgili çalışmalarda bulunmuştur. Bunların 30'a yakını üniversite diğerleri ise ticari ve askeri kuruluşlardır. Bu konu ile ilgili çalışmalar yapan kuruluşların içinde US Army ve US Navy de yer almıştır. Bunların haricinde telefon şirketlerinden meşrubat şirketlerine kadar pek çok kuruluş bu çalışmalara katılmıştır.

Ses tanıma ve doğal dil işleme, Microsoft'un hesaplarına göre DOS'tan Windows'a geçişten sonraki en büyük atılım olacaktır. Onlara göre bu teknoloji normalde cansızmış gibi görünen bir objeyle olan (bilgisayar) ilişkileri köklü biçimde değiştirecektir. Ancak Microsoft'a göre bu teknoloji birden ortaya çıkabilecek bir teknoloji değildir. Sonraki on yıl içinde ortak çalışmalar sonucunda yavaş-yavaş gelişip yerine oturacaktır.

Bu teknolojiyi 4 başlık altında incelemek mümkündür;

Telefonda Ses (konuşma) Tanıma; Komutları anlayan bilgisayarlar ile, telefonla servis veren veya verebilecek olan şirketler için bu konu büyük önem arz etmektedir.

Dikte Ettirme; Herhangi bir sürekli konuşmaya kısıtlı olarak izin veren, mevcut yazılımlar vardır. Örnek olarak "Microsoft Dictation" ve "Dragon Dictate" verilebilir. Bu programların doğruluk oranları %90-95'ler civarındadır. Ancak hala çalışmalar devam etmektedir. Çünkü bu hata oranı 3000 kelimelik bir makalede pek çok boşluk kalmasına sebep olmaktadır.

Konuyu Anlayan Tanıyıcılar; Bu alandaki çalışmalar sadece söylenen kelimeyi anlamayı değil ne demek istediğinizi yani söylediğiniz cümlenin anlamını çıkarmayı hedeflemektedir. Bu hedefe ulaşmak öncelikle uzmanlaşmış uygulamalar yapmayı ve öncelikle sınırları belirli alanlar içinde kalmayı, bunu başardıktan sonra genel kullanıma geçmeyi düşünmektedirler

Doğal Dil Anlama; Bilgisayarlar, sürekli konuşmayı ve diyalogları anlayabildiğinde, bu, teknolojideki büyük bir devrim olacaktır. Henüz bu teknolojinin gelmesine en az 15 yıl varmış gibi gözükmektedir. Ama bu teknoloji yayıldığında insan hayatında büyük değişikliklere neden olacaktır. Bilgisayara yapması istenen işlem normal bir cümle şeklinde söylenecek ve istenen işlem bilgisayar tarafından yerine getirilecektir. Microsoft ve IBM firmaları beraber bu hedefe ulaşmak için çalışmaktadırlar. Özellikle Microsoft, işletim sistemine bu teknolojiyi yerleştirmenin yollarını aramaktadır.

Ses tanıma problemine getirilen farklı çözüm tarzları vardır. Bunlar tanınması gereken konuşmanın kesikli mi yoksa sürekli mi olduğundan etkilenirler. Yani iki konuşmayı da aynı tekniklerle tanımak zordur. Kesikli bir konuşmanın tanınması daha kolaydır ve kelime-kelime yapılması gayet uygundur. Sürekli bir konuşmanın ise kelime-kelime tanınması daha zordur çünkü kelimelerin nerede başlayıp nerede bittiği bilinmemektedir. Dolayısıyla sürekli tanıma genelde fonem bazında yapılmaktadır. Fonem anlam içeren en küçük ses demektir. Yani fonem bir heceden daha kısa bir sestir. Normal bir hecede başlangıç-orta-bitiş olmak üzere hemen-hemen üç fonem bulunur. Ancak fonemleri de birbirinde kesin hatlarla ayırmak pek mümkün değildir. Bu nedenle fonemleri tanıyacak ve temsil edecek çeşitli sistemler geliştirilmiştir. Bunların başında Hidden Markov Modeli (HMM) gelmektedir (Nwe ve ark. 2003, Peinado ve ark. 2003). Ses tanıma problemi kişiye bağımlı, kişiden bağımsız ya da kişiye uyum sağlayan tarzlarda çözülebilir. Ses tanıma probleminde önemli olan diğer bir nokta da tanınacak kelimelerin (kelime haznesi) sayısıdır.

- Küçük kelime dağarcığı – 10-100 kelime
- Orta kelime dağarcığı – 100-1000 kelime

• Geniş kelime dağarcığı – 1000-10000 kelime
• Çok geniş kelime dağarcığı – 10000 ve daha fazla kelime

Günümüz teknolojisinde, sesin sayısal olarak işlenmesi analog işlenmesine göre bir takım üstünlükler taşır. Sayısal işleme, sesin iletimi ve saklanmasında önemli bir boyutta bellek alanı ve bant genişliği tasarrufu sağlar, ses güvenli bir şekilde iletilir ve kriptolanabilir, ayrıca, sesin perde genişliği ve zaman eksenindeki yerleşimini değiştirebilme imkanı verir.

Ses sinyallerinin yüksek frekans (HF) kanallarından, dar bantlı, yüksek güvenilirlikli ve gizli bir şekilde iletimini sağlayabilmek için düşük hızda ses kodlama üzerinde durulmuştur.

Ses haberleşmesi, günümüzde ve gelecekte sadece sivilde değil, stratejik ve taktik askeri uygulamalarda da haberleşmenin ana dalı olarak mevcudiyetini koruyacaktır. Sinyal işleme tümleşik devre teknolojisindeki hızlı gelişmeler ve telsiz, seyyar ve taşınabilir ses haberleşmesi için gittikçe artan talep sonucunda, ses kodlama araştırmaları yüksek bir motivasyon ile sürmektedir.

Seyyar haberleşme ve geniş ölçekli taşınabilir telsiz telefonları, sınırlı radyo tayfını giderek zorladığı için araştırmacılar, düşük güç tüketimli, minyatür boyutlara indirgenebilen teknolojiye sahip ve düşük maliyetli olan düşük bit hızında ve yüksek kalitede ses kodlamaya doğru itilmişlerdir.

Kod-Uyarımlı Doğrusal Tahmin Kodlaması (Linear Prediction Coding - LPC) olarak tanımlanan tekniklerin geliştirilmesi ve araştırılması ile ses işleme oldukça önem kazanmıştır.

3.5.1. Ses Tanımanın Avantajları ve Sınırlamaları

Ses tanıma metodunun avantajları ve dezavantajları bulunur. Hala çözümü olmayan problemler teknikte bazı sınırlamalara neden olur. İlk olarak, teknik

kontrollü ve sessiz ortamda iyi çalışır. Yüksek ses seviyeleri tekniğin faydalı olan avantajlarını bulmayı zorlaştırır.

Eski basit ses tanıyıcılarda içerik düşüncesi yoktur. Bir çıktıdaki sonuçlar sadece bir kelime topluluğudur ve kullanıcı için anlamı vardır. Aynı zamanda tanıyıcı büyük bir sözlük kullanıyorsa sistem tanımada zorlanır. Çünkü sözlüğü genişletmek sistemi daha karışık yapar.

Tanıyıcı daha fazla kullanılabilir işlem gücü gerektirdiği için uygulamanın kalanı bundan olumsuz etkilenir. Bir tanıyıcı için diğer bir problem, araştırmalar en normal sesin kısa kelimeler içerdiğini gösterirken uzun kelimeleri ayırmanın daha kolay olmasıdır.

Ses tanıma konusunda halihazırda uygulamadaki sıkıntı, her insanın konuşma tarzının ve ağzının farklı olması, yani bir standardın olmamasıdır. Tanınabilir bir lehçeye rağmen, bir ses tanıma aracı herkes için çalışmayı garanti etmez. Bazı insanlar, diğerleri en sessiz ortamlarda bile tanıma yapamıyorken, en gürültülü ortamlarda bile bu aracı kullanabilirler.

Bazı konuşma tanıma araçları frekans ölçüsünün tamamını kullanamayan donanıma sahiptir. Bu çok kapsamlı olmayan donanım özellikle yüksek frekanslı girişi kapsamaz ve çıktı sonuçları güvenilir değildir. Erkek sesinin kadın sesine tercih edilme nedeni budur.

Sekizinci ve son sınırlama tanıyıcının kapasite seviyesi ile ilgilidir. Mükemmel şartları olan bir iyi-fonksiyonlu tanıyıcı bile sürekli olarak çeşitli hatalar yapar. Tanıyıcı, hatalı kelimeler duyabilir, söylenen kelimeleri atlayabilir ve kelimeleri yanlış anlayabilir. Kelimenin doğruluk oranı %95'dir. Ama unutmamalıyız bir doğruluk oranının %95 olması istatistiksel bakış açısıyla 8-10 kelimeli cümlelerin yarı zamandan daha fazlasında en az bir hata ile tanınmasıdır. Ses tanıyıcılar için tüm bu kısıtlamaları maddeler halinde özetleyecek olursak:

1. Sessiz, kontrollü ortamda en iyi çalışır.
2. Eski basit ses tanıyıcılar dilsizdir, kelimeleri duygusuz söylerler.
3. Büyük sözlük kullanıldığında karmaşıklık artar.
4. Yüksek miktarda işlem gücü ister.
5. Kısa kelimeleri ayırt etmenin zorlukları vardır.
6. Herkes için çalışma garantisi yoktur.
7. Bazı tanıyıcıların diğerlerine göre daha fazla zahmetli işleme tarzı vardır.
8. En iyi durumda kelime doğruluğu %95'e ulaşır.

Bütün bunlardan başka, ses tanıma bir iş yeri için çok büyük potansiyellere sahiptir ve yetersizlikleriyle insanlar için eğitimsel uyum süreci vardır. Çoğu durumda bir ses tanıma aracı çalışmazsa, bu kullanıcının davranışına ve bilgisine bağlıdır. Bu gibi yetersiz bilgiyi önlemek için iki ölçüm vardır. Kullanıcılar sistemi ses ile çalıştırmak için iyi hazırlanmış olmalıdır ve güncel teknolojileri bilmelidirler.

3.6 Ses Kodlamanın Gelişimi

a.Analog ses bilgisini sayısal haberleşme sistemleri üzerinden gönderebilmek için ses bilgisinin sayısal işarete çevrilmesi gerekir. Analog ses bilgisinin sayısal işarete dönüştürülme işlemine ses kodlama adı verilmektedir.

b.Haberleşme amaçlı ses kodlama ilk olarak 1960'larda uzun mesafe telefon sistemlerinde 64 bit/s Pulse Code Modulation (PCM) yöntemiyle gerçekleştirilmiştir.

c.PCM yönteminde analog ses işareti saniyede 8000 defa örneklenir. Örneklenen her parça ise 8 bit sayısal bilgiye dönüştürülür. Yani başka bir deyişle bir saniyelik ses bilgisi 64 Kbit/s hızında sayısal bilgiye dönüştürülür

ç.64 Kbit/s hızı telli sistemlerde uygulanabilirken frekans kaynağının sınırlı olduğu PMR sistemlerine uygulanamaz. Çünkü 64 Kbit/s sayısal veri göndermek için gerekli bant genişliği mevcut ve gelecek PMR sistemlerinde kullanılacak bant genişliğinin çok üstündedir.

d.Çok seviyeli modülasyon teknikleri ile bu hızdaki sayısal bilgiyi dar haberleşme kanallarından geçirmek mümkündür ancak; bu da pahalı ve zahmetli bir yöntemdir.

e.1970'li yıllarda dar bant sistemler için büyük bir gelişme sayılabilecek Continuous Variable Slope Delta (CVSD) Modülasyon tekniği geliştirildi. Bu yöntem ile insan sesinin 32 Kbps örnekleme hızından 12 Kbps hızına kadar açık ve anlaşılır bir şekilde örneklenmesi başarıldı.

f.CVSD yönteminde PCM'dekinden farklı olarak seviyeler değil seviyelerdeki değişmeler 12000 kez (12 kbps örnekleme hızı için) örneklendi. Örneklenen her seviye ise bir bit ile gösterildi. Ses kalitesindeki düşüklükten dolayı daha çok kriptolu sistemlerde kullanılmıştır.

g.Linear Predictive Coding (LPC), yönteminin geliştirilmesi ve bu yöntemin getirdiği kanal hızı ve ses kalitesi özellikleri ile dar bant sistemlerde kullanılabilecek bir kodlama yöntemine erişilmiş oldu.

ğ.LPC kodlama yönteminde sesin fiziksel üretimi, ağız yapısı, nazal boşluklar, dil hareketleri simüle edilerek ses bilgisi sayısallaştırılmaktadır. LPC ses kodlayıcılar, perde sıklığı, genlik, sesli/sessiz sesler gibi bilgileri tanımlamak için parametreler kullanmaktadır.

h.Basit LPC kodlayıcılar ile 2.4 Kbit/s hızında bile anlaşılır ses kodlamak mümkün olmaktadır. Ancak bu yöntemle ses karakterinde bozulma olduğundan sadece askeri sistemlerde kullanılmaktadır.

ı.CELP kodlama tekniği LPC kodlama tekniğinin bir türevidir. Bu teknikte ses bilgisi bir adet uzun süreli ve iki adet de kısa süreli vektör dizisine bölünür ve bu vektör blokları standart bir kod kitabına göre kodlanarak toplanırlar.

i.CELP kodlayıcı değişken oranlı bir kodlayıcıdır. Maksimum oranda/hızda kodlama yaparken 8550 bps de kodlama yapar. Konuşma aralarındaki duraksamalarda, arka plandaki gürültü sadece 800 bps ile kodlanır. Günümüzde standart olarak 4.8 kbps örnekleme hızındaki CELP kodlayıcılar kullanılmaktadır.

j.USA savunma bakanlığı tarafından, daha önce NATO standardı olarak kullanılan LPC-10 ses kodlama algoritması ses kalitesinin düşük olması nedeniyle, yeni ses kodlama algoritması standardı olarak MELP kabul edilmiştir.

k.MELP ses kodlama tekniği de günümüzün en son teknolojileri ile 2.4 kbps örnekleme hızında kaliteli ses kodlaması yapılabilmektedir. MELP, dar bant kodlama tekniğidir ve geleneksel LPC parametrik modeline göre çalışmaktadır. Bu modelden dört ana fark gösterir :

(1) Periyodik darbe dizisi ve gürültü karışımı ile uyartım
(2) Periyodik olmayan darbe dizisi
(3) Darbe dağıtma filtresi
(4) Değişken/Uygunlaştırıcı Spektral Süzme

BÖLÜM 4

YAPAY SİNİR AĞLARININ SES TANIMADA KULLANILMASI

Yapay sinir ağları (YSA) günümüzde bilgi sınıflama ve bilgi yorumlamanın içinde bulunduğu değişik problemlerin çözümünde kullanılmaktadır (Elmas, 2003). Özel olarak ses tanıma problemi ele alındığında özellik vektörü çıkarılmış bir ses sinyalinin tanınması da bu kapsama girmektedir. Ses tanıma sürecinde ses sinyalinin modellenmesi aşamasında yapay sinir ağları kullanabilir.

Yapay sinir ağlarının ses tanımadaki uygulaması üzerine bilgi vermeden önce ses tanıma amacı için YSA teorisinden biraz bahsedelim.

4.1. Kavramlar ve Terimler

YSA biyolojik sinir ağlarından esinlenerek geliştirilmiş, bilgi işlem sistemleridir.

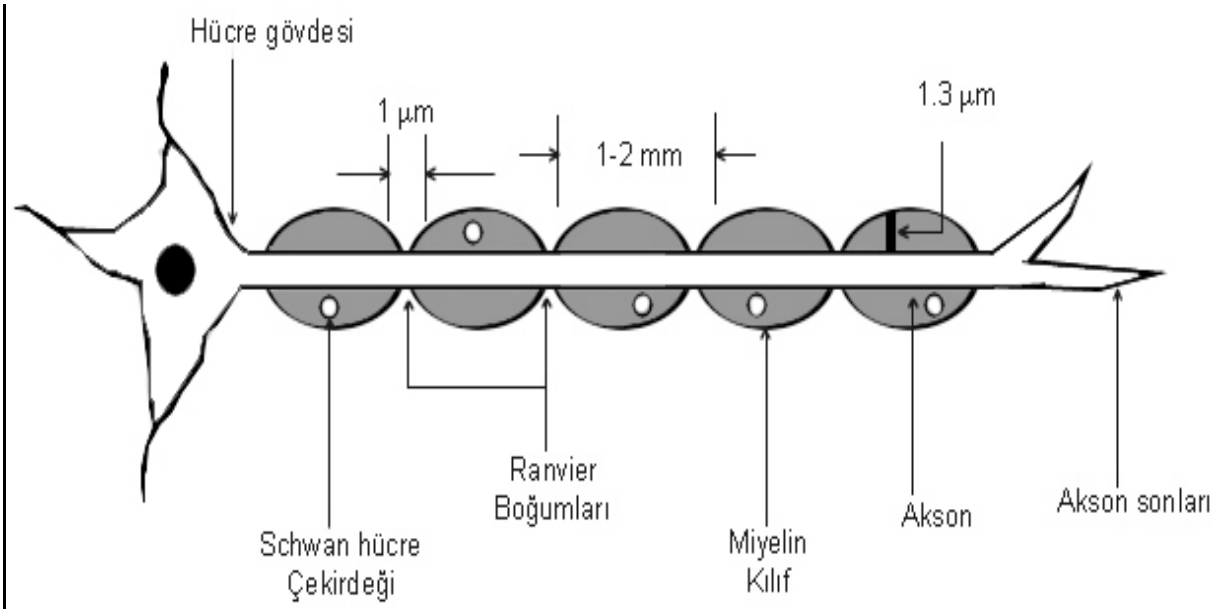
Bir sinir ağı, bilgiyi depolamak için doğal eğilimi olan basit birimlerden oluşan paralel dağıtılmış bir işlemcidir. Bu ağlar sinaptik ağırlıklar olarak bilinen nöronlar arası bağlantı kuvvetlerini, bilgiyi depolamak için kullanır. Benzer olarak YSA da birçok işlemciden oluşur ve farklı biçimlerde ifade edilebilen nümerik verileri taşıyan bağlantı yada ağırlıklarla birbirine bağlı yapılardır.

Biyolojik sinir ağlarında girdi sinyallerini alan, yorumlayan ve uygun çıktıyı ileten temel işlemci nöron olarak adlandırılır. Bir nöron, gövde (cell body), gövdeye giren sinyal alıcıları (dentrit) ve gövdeden çıkan sinyal iletici (akson) olmak üzere üç kısımdan oluşur (Sağıroğlu vd. 2003).

Dentritler, nörona bilgiyi alan ve sayısal olarak birden fazla olabilen yapılardır

ve iç yapıları nöronla aynıdır.

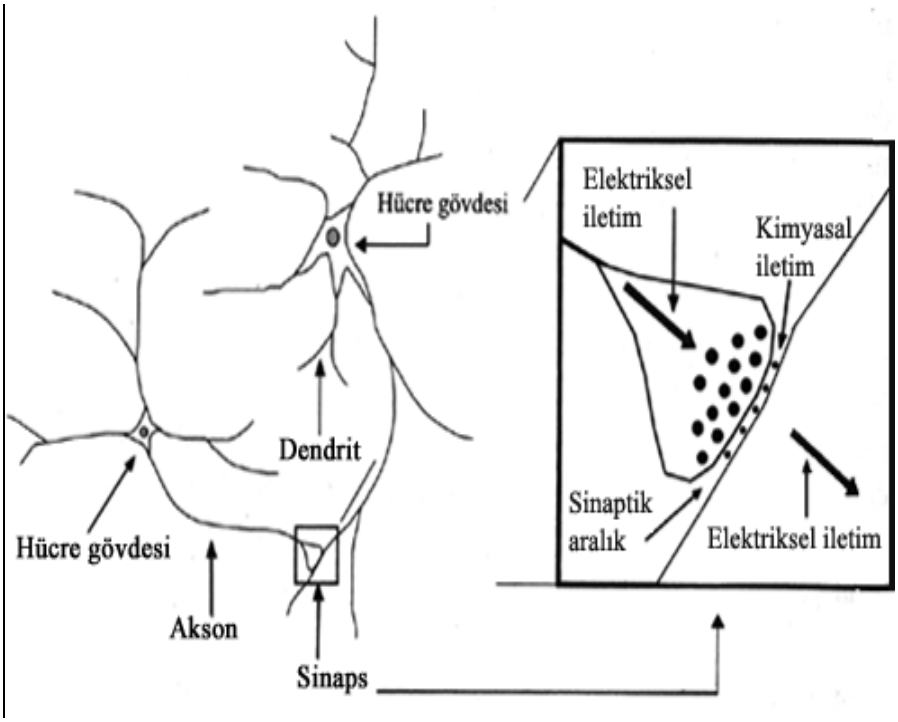
Aksonlar, dentritten aldığı bilgiyi diğer hücrelere aktaran uzantılardır. Uzunlukları birkaç mikrondan, 1-2 metreye kadar değişebilir. Her nöronun yalnızca bir aksonu vardır. Aksonlar akson kesecikleri denilen ve içlerinde çok miktarda vezikül bulunan bir çok yapı ile sonlanırlar. Bu keseciklerde sinaptik geçişte rol alan kimyasal taşıyıcılar (nörotransmitter) bulunur. Aksonlar özel bir örtüye sahip olmalarına göre miyalinli yada miyalinsiz olarak sınıflandırılabilirler. Akson üzerini örten miyalin kılıfın, yalıtım ve darbe hızını arttırmak gibi iki önemli görevi vardır. Aşağıda miyalinli bir nöronun yapısı gösterilmiştir.



Şekil 4.1. Miyalinli bir nöron yapısı

Sinir hücreleri, dış kısmı pozitif, iç kısmı negatif yüklü ve ortasında çift lipid tabakasıyla yalıtılmış, kutuplanmış bir kondansatöre benzer. Bir uyarıcı ile bu kondansatör nötrlenebilir hatta anlık olarak ters kutuplanabilir. Zarın belirli bir bölgesinde 1ms kadar süren bu potansiyel değişiminden sonra zar bu bölgede eski haline dönerken, potansiyel değişikliği zar boyunca yayılmaya devam eder. Akson

boyunca iletilen bilginin temel birim olan bu potansiyel değişimine aksiyon potansiyeli denir.

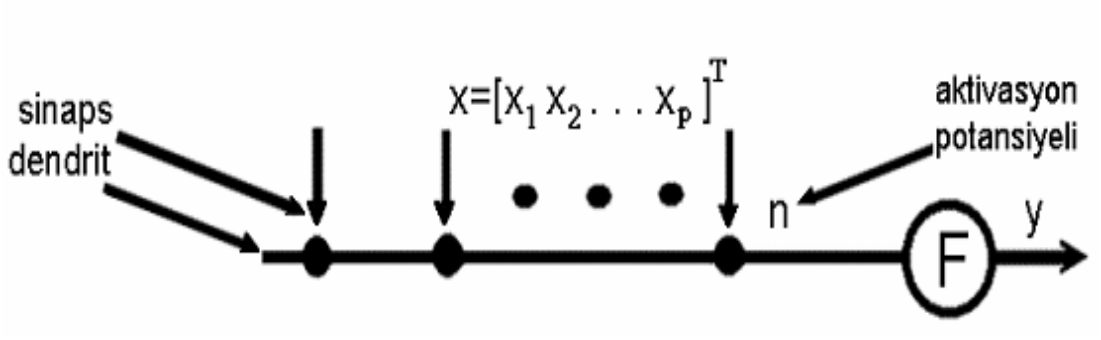


Şekil 4.2. Biyolojik nöron

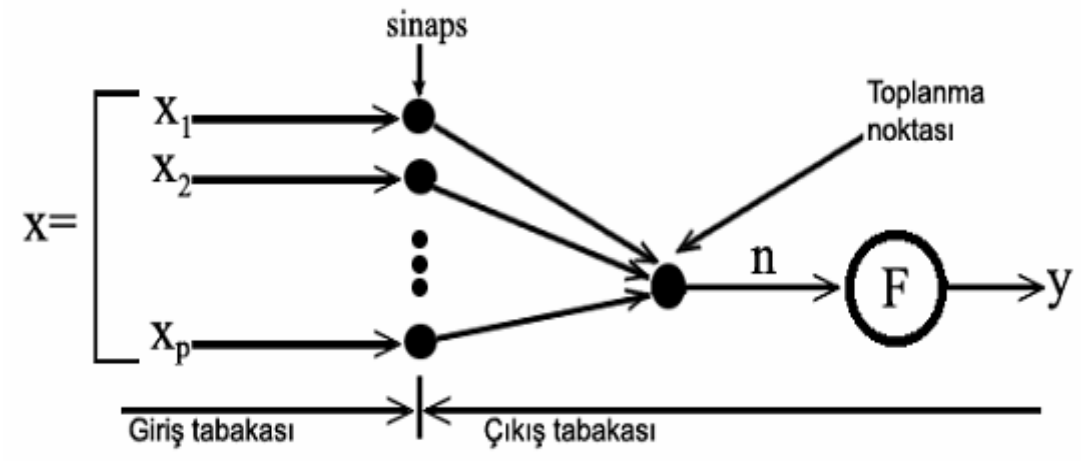
Sinir hücreleri arasında iletişimin gerçekleştiği, yapısal ve fonksiyonel olarak özelleşmiş bölgelere sinaps adı verilir. Mesajı gönderen ve presinaptik hücre olarak adlandırılan hücre ile mesajı alan ve postsinaptik hücre bu kavşak bölgelerinde bir birlerine oldukça yaklaşırlar.

Yapay sinir ağları, biyolojik sinir ağlarından esinlenerek, bir birine bağlı doğrusal ve/veya doğrusal olmayan bir çok işlemci elemandan oluşur. Bir yapay nöron temel olarak girişler, ağırlıklar, toplam fonksiyonu, aktarım fonksiyonu ve çıkış olmak üzere beş kısımdan oluşur.

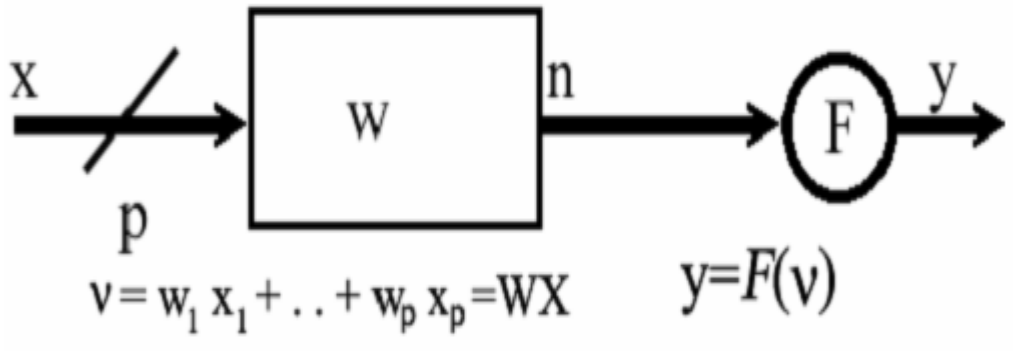
P girişli ve n çıkışlı bir yapay nöronun farklı grafik gösterimleri şöyledir:



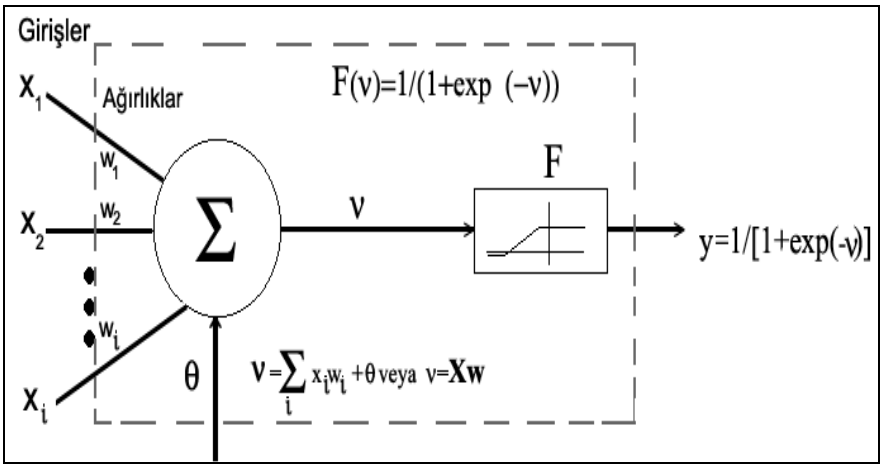
Şekil 4.3. Dentrit gösterimi



Şekil 4.4. İşaret akış gösterimi



Şekil 4.5. Blok diyagram



Şekil 4.6. Ayrıntılı gösterim

Bu gösterimlerde **x** girişleri, **y** çıkışı, **F** aktivasyon fonksiyonunu, **p** nöron giriş sayısını, **n** çıkış sayısını, **w** ise ağırlıkları ifade eder.

Dentrit gösteriminde presinaptik aktiviteleri giriş işaretlerinin p elemanlı sütun vektörü olarak gösterilir. $x = \left[x_1, x_2, ..., x_p\right]^T$ giriş desenlerinin uzayı p boyutludur.

Sinapslar ağırlıklar olarak adlandırılan ayarlanabilir parametreler ile karakterize edilirler. Ağırlıklar, p elemanlı satır vektörü olarak düzenlenir:

$$W = \left[w_1, w_2, ..., w_p\right]$$

İşaret akış gösteriminde, p tane ağırlığı olan bir nöron giriş noktalarının bir katmanı şeklinde düzenlenir. Ağırlıklar, giriş ile toplama noktası arasındaki bağlantılara karşılık gelir.

Sinapslardan ve dentritlerden geçen giriş işaretleri, 'toplam post-sinaptik aktiviteyi tanımlayan' aktivasyon potansiyeli olarak toplanır.

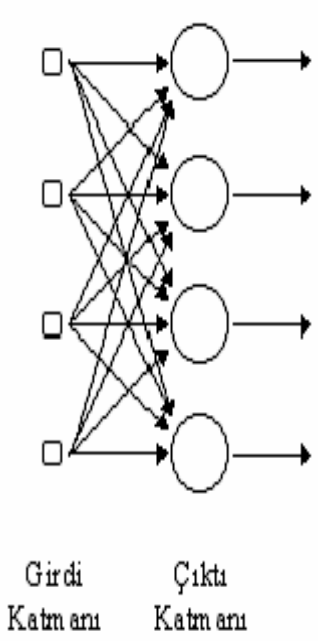
Aktivasyon potansiyeli giriş işaretlerinin ve ağırlıklarının lineer toplamı olarak şekillenmiştir. Yani ağırlıklar ile geçiş vektörleri çarpımıdır:

$$v = \sum_{i=1}^{p} w_i x_i = [w_1 w_2 ... w_p] \begin{bmatrix} x_1 \\ x_2 \\ \\ x_p \end{bmatrix} = WX^T \tag{4.1}$$

4.2. Yapay Sinir Ağlarının Sınıflandırılması

Yapay sinir ağlarında, ağ mimarisini belirleyen önemli etmenlerden biri de öğrenme algoritmasıdır. Genel olarak ağ mimarileri dört ana grupta toplanabilir (Yücetürk 2000). Bu alt-bölümde bu ağ mimarilerinin her biri hakkında kısa bilgi verilecektir.

4.2.1. Tek Katmanlı İleri Beslemeli Ağlar

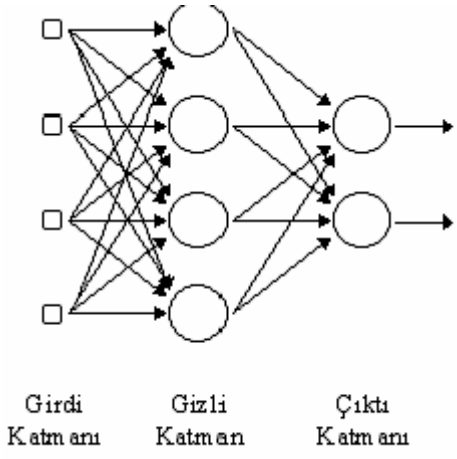


Şekil 4.7. Tek katmanlı ileri beslemeli ağ modeli

Katmanlı modellerdeki en basit ağ tipi olup bir çıktı katmanı ve buna bağlı bir girdi katmanı bulunmaktadır (Şekil 4.7).

4.2.2. Çok Katmanlı İleri Beslemeli Ağlar

Tek katmalı ağlardaki girdi ve çıktı katmanından başka, bir yada daha fazla sayıda gizli katman içeren ağlara çok katmanlı ağ denir. Dış dünya tarafından doğrudan müdahale edilmediği için gizli katman adı verilen katmanda bulunan birimlere de gizli birimler adı verilir (Şekil 4.8).

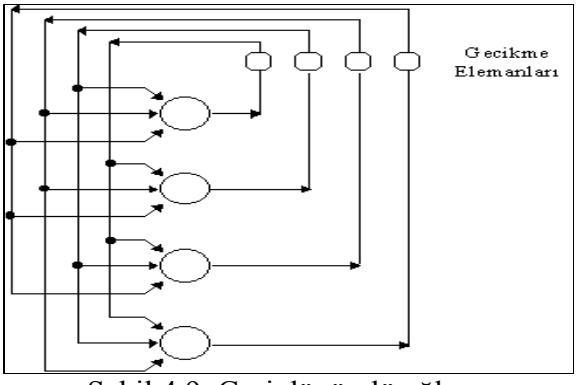


Şekil 4.8. Çok katmanlı ileri beslemeli ağ modeli

Çok katmanlı ağlar tek katmanlı ağlara göre daha karmaşık problemlere çözüm getirebilmektedir. Ancak eğitilmesi zordur.

4.2.3. Geri Dönümlü Ağlar

Geri dönümlü ağların ileri beslemeli ağlardan farkı, bu ağlarda en az bir tane geri-besleme döngüsünün bulunmasıdır (Şekil 4.9).

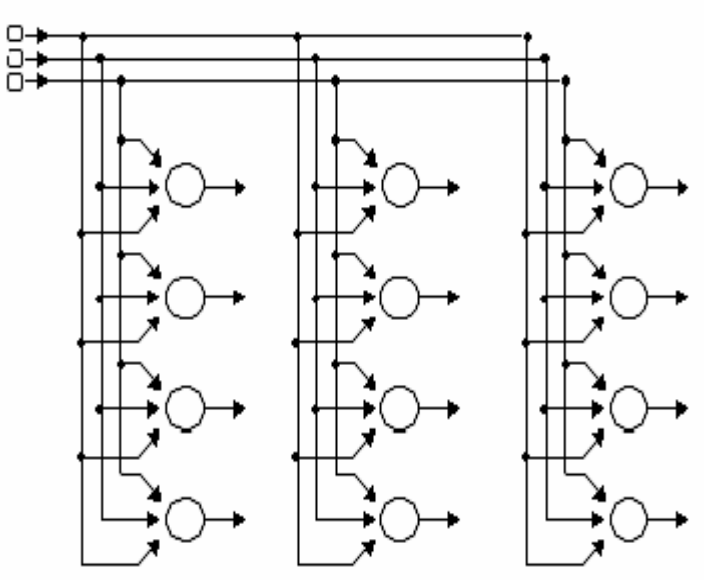


Şekil 4.9. Geri dönümlü ağlar

Bu gibi yapıların bazılarında senkronizasyonun sağlanması için gecikme elemanlarının kullanılması gerekebilmektedir. Bu yapıya sahip ağlarda geri-besleme sinyali aynı katmandaki nöronlara gönderilebileceği gibi, alt katmanlardaki nöronlara da gönderilebilir.

4.2.4. Kafes Yapılı Ağlar

Kafes yapılı ağlar bir, iki veya çok-boyutlu nöron dizinlerinden meydana gelmekte olup girdi katmanı, dizinlerin hepsine girdi bilgisi sağlamaktadır (Şekil 4.10).



Şekil 4.10. Kafes yapılı ağ modeli

4.3. Yapay Sinir Ağlarında Öğrenme

Yapay sinir ağlarında bilgi, nöronlar arasındaki bağlantılar üzerindeki ağırlık değerleri üzerinde tutulur. Bu yüzden yapay sinir ağlarının eğitimi ve öğrenme bu ağırlıklara değerlerin verilmesi ve değiştirilmesi anlamına gelmektedir. Öğrenme yaklaşımı, ağ yapıları için ayırt edici bir özelliktir. En genel anlamda öğrenme,

güdümlü öğrenme ve güdümsüz öğrenme olmak üzere iki grupta toplanabilir (Yücetürk 2000).

4.3.1. Güdümlü öğrenme

Güdümlü öğrenme, yapay sinir ağlarında yaygın olarak kullanılan bir yöntemdir. Eğitim sırasında sisteme bir girdi ve hedef çıktı vektörü çifti verilir. Giriş bilgisinden üretilen çıktı değeri ile hedef çıktı değeri kıyaslanarak ağırlıkların değiştirilmesi için gerekli bilgi hesaplanır. Bu kıyaslamadan elde edilen fark, hata olarak belirlenen değerden küçük oluncaya kadar eğitim devam eder. Hata değeri istenen değerin altına düştüğünde ağırlık değerleri sabitlenir ve eğitim biter.

Yapay sinir ağlarının eğitiminde kullanılan hedef çıktı 1 veya 0 gibi ikili olabileceği gibi bir örüntü de olabilir. Belli bir girdi kümesine karşılık ilgili çıktı kümesini hatırlayacak şekilde eğitilmiş sistemlere çağrışımlı bellek adı verilir. Eğer girdi vektörü ile çıktı vektörü aynı ise buna öz-çağrışımlı bellek, çıktı vektörü farklı ise buna da karşıt-çağrışımlı bellek adı verilir.

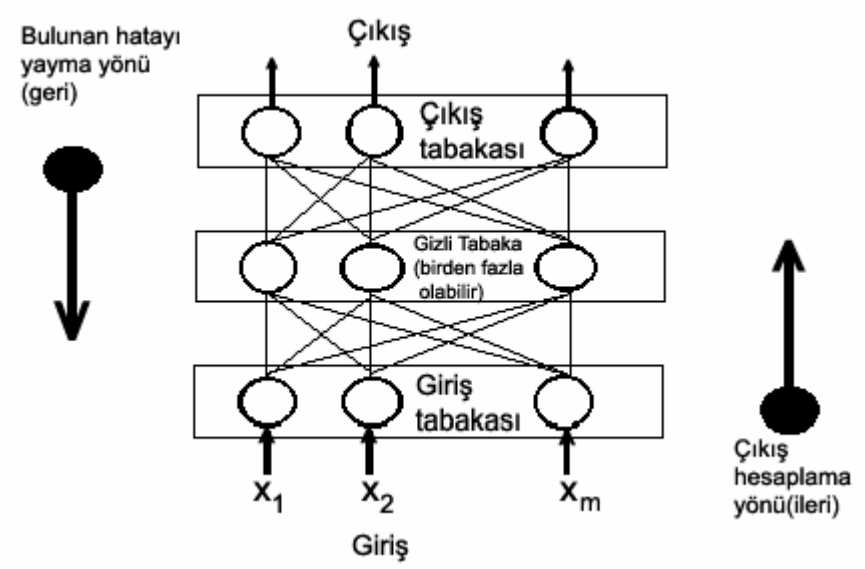
4.3.2. Güdümsüz Öğrenme

Kullanılan girdi vektörlerinin hangi sınıfa ait olduğu verilmiyorsa, bu sınıflama işlemini girdileri birbirleriyle karşılaştırarak sistemin kendisi yerine getiriyorsa, bu tür sistemlerdeki öğrenme yöntemine güdümsüz öğrenme adı verilir. Bu sistemlerde bir grup girdi vektörü sisteme verilir, ancak hedef çıktılar belirtilmez. Sistem girdiler içinde birbirine en çok benzeyenleri gruplar ve her bir grup için farklı bir örüntü tanımlar. Özdüzenlemeli özellik haritaları bu yöntemi kullanarak sınıflama işlemini yerine getirir.

4.4. Çok Katmanlı Perseptron ve Geri Yayılımlı Öğrenme

Bu bölümde ses tanıma sistemimizde faydalandığımız güdümlü öğrenme metotlarından Çok Katmanlı Perseptron (Multi Layer Perceptron-MLP) ve Geri

Yayılımlı Öğrenme (Back propagation) hakkında bilgi vereceğiz. Çok katmanlı perseptronlar, ileri beslemeli, tam bağlantılı, bir girdi katmanı, bir veya birden fazla gizli katman ve bir çıktı katmanından oluşan hiyerarşik yapıya sahip yapay sinir ağı mimarileridir. Böyle bir ağdaki herhangi bir nöronun çıkışı $y_k = f\left(\sum w_k x\right)$ ile hesaplanır. Giriş katında herhangi bir bilgi işleme yapılmaz. Buradaki işlem elemanı sayısı problemin giriş sayısına bağlıdır. Ara katman sayısı deneme yanılma yoluyla bulunur. Çıkış katmanı eleman sayısı ise uygulanan probleme bağlıdır (Yücetürk 2000).



Şekil 4.11. Çok katmanlı perseptron yapısı

Geri yayılımlı öğrenme metodu olarak isimlendirilen eğitim algoritmasının temeli ilk olarak Werbos'un Harvard Üniversitesi'nde verdiği doktora tezinde ileri sürülmüş (Werbos 1974), daha sonra Parker tarafından MIT'nin bir teknik raporunda ele alınmış (Parker 1985) ve Rumelhart ve arkadaşları tarafından da popüler hale getirilip uygulanabilir şekle dönüştürülmüştür.

Daha önce Adaline mimarisi ile kullanılan LMS veya Delta kuralının geliştirilmiş ve çok katmanlı yapıya uyarlanmış hali olan geriyayılımlı öğrenme metodu, çok katmanlı perseptron yapısı üzerinde çok çeşitli problemlerin çözümünde kullanılmaktadır. Burada geri yayılım kavramı, oluşan hatalar için kullanılmaktadır.

Geri yayılımlı öğrenme metodu ile çok katmanlı perseptron ağı üzerinden iki geçiş söz konusudur. İleri besleme aşamasında eğitim için kullanılan girdi, sisteme beslenir ve bunun sonuçları her bir katmanı geçerek çıktı katmanına kadar gelir ve girdiye karşılık bir çıktı elde edilir. Bu aşama sırasında ağ üzerindeki ağırlık değerleri sabit tutulur. İkinci aşama olan geri besleme aşamasında, elde edilen çıktı ile hedef çıktı arasındaki farktan hata sinyali elde edilir ve bu sinyal ağ yapısında geriye doğru yayılır. Bu geri yayılım aşamasında, oluşan hatayı minimuma indirecek şekilde ağırlık değerleri güncellenir.

Bu güdümlü öğrenme işleminin tekrarlanması ile sistem, kendisine beslenen girdi örneklerini öğrenir. Eğitim bittikten sonra sistemin uygulama sürecinde sadece ileri beslemeli ilk aşama yeterli olmaktadır. Bu durumda sistemin eğitilmesi uzun süre almakla beraber, eğitilmiş bir sistemden bilgi alınması çok hızlı olmaktadır.

4.4.1. Geri Yayılım Algoritmaları

Geri yayılım algoritması (Haykin 1994), dereceli azalan ve ÇKP'ları öğretmede çok kullanılan temel bir öğrenme algoritmasıdır. Bu algoritmalarda, i ve j kat nöronları arasındaki ağırlık ifadesi olan $\Delta w_{ji}(k)$ değişikliği hesaplanır. Genel ifade,

$$\Delta w_{ji}(k+1) = \eta \delta_j x_i + \mu \Delta w_{ji}(k) \tag{4.2}$$

olarak verilir. Eşitlik (4.2)'de η öğrenme katsayısı, μ momentum katsayısı ve δ_j ara veya çıkış katındaki herhangi bir j nöronuna ait bir faktördür. Çıkış katı için bu faktör aşağıdaki şekilde verilir.

$$\delta_j = \frac{\partial f}{\partial net_j}\left(y_j^t - y_j\right) \tag{4.3}$$

Burada, $net_j \equiv \Sigma x_i w_{ji}$ ve y_j^t ise j nöronunun hedef çıkışıdır. Ara katlardaki nöronlar için ise bu faktör,

$$\delta_j = \left(\frac{\partial f}{\partial net_j} \right) \sum_q w_{qj} \delta_q \qquad (4.4)$$

olarak verilir. Ara katlardaki nöronlar için herhangi bir hedef çıkış olmadığından, (4.3) nolu eşitlik yerine (4.4) nolu eşitlik kullanılır. Bu duruma bağlı olarak çıkış katından başlayarak δ_j faktörü, bütün katlardaki nöronlar için hesaplanır. (4.2) nolu eşitliğe bağlı olarak bütün bağlantılar için ağırlıkların güncelleştirilmesi böylece gerçekleştirilir.

4.4.1.1 Standart Geri Yayılım (SGY)

Standart geri yayılım algoritması (SGY), denklem (4.3)'deki ifade de momentum katsayısının (μ=0) sıfır olması durumundaki değişikliği şeklinde ifade edilir. Bu ise,

$$\Delta w_{ji}(k+1) = \mu \delta_j x_i \qquad (4.5)$$

denklemi ile gösterilir.

Standart gradiyent düşürme ile öğrenmede öğrenme oranı öğrenme süresi boyunca sabittir. Öğrenme başarısı öğrenme oranı ile yakından ilgilidir. Eğer öğrenme oranı çok büyük ise algoritma salınım yapar ve kararlı davranamaz. Eğer öğrenme oranı çok küçük ise algoritmanın sonuca yakınsaması uzun sürer. Pratikte öğrenme oranının ne olması gerektiğinin belirlenmesi oldukça güçtür ve gerçekten de öğrenme süresi boyunca öğrenme oranı da sürekli değişebilir.

4.4.2 DeltaBarDelta (DBD)

DeltaBarDelta (DBD) çok katmanlı perseptronlarda bağlantı ağırlıklarının yakınsama hızını arttırmak için kullanılan bir sezgisel yaklaşımdır. Deneysel çalışmalar, ağırlık uzayının her boyutunun tüm hata yüzeyi açısından tamamen farklı

olabileceğini göstermiştir. Hata yüzeyindeki değişimleri açıklamak için, özellikle, ağın her bağlantısı kendi öğrenme katsayısına sahip olmalıdır (Jacobs 1988). Bu düşünce, tek ağırlık boyutu için uygun adım büyüklüğü, tüm ağırlık boyutları için uygun olmayabilir. Bununla birlikte, her bir bağlantıya bir öğrenme oranı atanarak ve bu öğrenme oranının zamanla değişmesine izin verirsek, yakınsama zamanını azaltmak için daha çok serbestlik derecesi sağlanmış olur.

İleri beslemeli YSA yapıları çoğu zaman karmaşıktırlar. Ağdaki her bağlantı için en uygun öğrenme katsayıları kümesini belirlemek oldukça zaman alıcı olabilir. Eğimin geçmişteki değerlerini kullanarak, yerel hata yüzeyinin eğriliğini çıkarmak için sezgisellik uygulanabilir. Bir bağlantı için ağırlık değişimlerinin işareti, birkaç ardışık zaman adımları sırayla değiştiği zaman, bağlantı için öğrenme oranı azalmalıdır. Bağlantı ağırlık boyutu hata yüzeyi ile ilgili büyük bir eğriliğe sahiptir. Bağlantı ağırlık değişimleri bir kaç ardışık zaman adımları için aynı işarete sahip olduğundan bağlantı için öğrenme oranı arttırılmalıdır. Ağırlık boyutu hata yüzeyi ile ilgili küçük bir eğriliğe sahiptir ve yüzey önemli bir mesafe için aynı doğrultudaki eğime göre devam eder.

SGY algoritmasında eğim bileşeni aşağıdaki şekilde verilir.

$$\delta(k) = \frac{\partial E(k)}{\partial w(k)} \tag{4.6}$$

Burada E(k), k anındaki hata değerini, w(k) bağlantı ağırlığını ve δ(k) ise ağırlık değişiminin eğim bileşenini göstermektedir. SGY algoritmasında bağlantı ağırlığı,

$$w(k+1) = w(k) + \alpha\delta(k) \tag{4.7}$$

olarak güncelleştirilir. Burada α sabit bir öğrenme oranıdır. DBD öğrenme kuralında, her bağlantı için değişken öğrenme oranı α(k) atanır ve bağlantı ağırlığının güncelleşmesi

$$w(k+1) = w(k) + \alpha(k)\,\delta(k) \tag{4.8}$$

şeklinde yapılır. Jacobs (1988), her bağlantıda öğrenme hızının azalması ve artması için sezgiselliğin gerçekleştirilmesinde, δ(k) eğim bileşeninin ağırlıklı ortalamasını kullanılmıştır. Bu ağırlıklı ortalama $\bar{\delta}(k)$'dır. Özellikle, ağırlıklı ortalama aşağıdaki şekilde gösterilmiştir.

$$\bar{\delta}(k) = (1-\theta)\delta(k) + \theta\,\delta(k-1) \tag{4.9}$$

Burada, θ konveks ağırlık faktörüdür. Sezgisel açıdan, önceki eğim bileşeninin üstel artması ve şu anki eğim bileşeni aynı işaretli ise, öğrenme oranı κ sabiti ile artan ağırlıkla birleştirilir. Mevcut eğim bileşeni üstel ortalamadan farklı işaretli ise, öğrenme oranı mevcut değeri ile orantılı olarak azalır. Bu öğrenme oranının güncelleştirilmesi aşağıdaki eşitlikle tanımlanmıştır.

$$\Delta\alpha(k) = \begin{cases} \kappa & \bar{\delta}(k-1)\,\delta(k) > 0 \\ -\varphi\alpha(k) & \bar{\delta}(k-1)\,\delta(k) < 0 \\ 0 & \text{diğer} \end{cases} \tag{4.10}$$

Burada, κ öğrenme katsayısı artma faktörü, φ öğrenme katsayısı azaltma faktörü, α(k) k anındaki öğrenme oranını göstermektedir. Eşitlik (4.10)'den görülebileceği gibi algoritma, öğrenme katsayılarını lineer olarak artırmakta, fakat geometrik olarak azaltmaktadır.

4.4.3 Genişletilmiş DeltaBarDelta (Extended DBD)

Genişletilmiş DeltaBarDelta (GDBD) öğretme algoritması (NeuralWare 1996), eğitmede belirli bir anda iterasyon genel hatayı hesaplayarak hata düzeltme özelliğine sahiptir. Belirli bir andaki (ya da herhangi bir anda) ölçülen hata, bir önceki anda bulunan hatadan büyük ise ağın ağırlıkları bir önceki ağırlık değerlerini alır. Daha iyi sonuçlar elde etmek için hata düzeltme özelliğine dayanıklılık

(patience) faktörü eklenmiştir. Her peryotdaki (iterasyon) hatayı test etmek yerine daha önce iyileştirilme işlemi yapıldığından hata, n peryotda test edilir.

GDBD'nin DBD'den farkı aşağıda özetlenmiştir.

- Sezgisel momentum kullanılır.
- Ağırlık uzayında büyük atlamalar ortadan kaldırılmıştır.
- Geometrik azaltma büyük (amacı aşan) atlamaları engellemede yeterince hızlıdır.

GDBD'de, her bir bağlantıdaki momentum hızı $\mu(k)$ ve öğrenme hızı $\alpha(k)$ zamanla değişmektedir. DBD ile karşılaştırmak için standart delta kuralı ile verilen momentum için bağlantı ağırlıklarının değiştirilmesi

$$\Delta w(k+1) = \alpha\delta(k) + \mu\Delta w(k) \tag{4.11}$$

değerinin ağırlıklara katkısıyla

$$w(k+1) = w(k) + \Delta w(k+1) \tag{4.12}$$

elde edilebilir. Burada α ve μ sabittir. Bunun yanında GDBD için değişken öğrenme ve değişken momentum hızları aşağıdaki gibi elde edilebilir. Ağırlıklar,

$$\Delta w(k+1) = \alpha(k)\,\delta(k) + \mu(k)\Delta w(k) \tag{4.13}$$

değerinin bir önceki ağırlığa eklenmesiyle

$$w(k+1) = w(k) + \Delta w(k+1) \tag{4.14}$$

bulunur. Burada, $\mu(k)$ k zamanındaki bağlantı momentum hızıdır. Momentum hızı $\mu(k)$ ve öğrenme hızı $\alpha(k)$ aşağıdaki kurallara göre benzer şekilde ayarlanır. İlk önce, $\bar{g}$ nin hesaplanması DBD algoritmasına göre belirlenir.

$$\bar{g}(k) = (1-\theta)\left(\delta(k) + \theta\delta(k-1)\right) \tag{4.15}$$

Burada θ konveks ağırlık faktörüdür. GDBD için öğrenme hızı değişimi:

$$\Delta\alpha(k) = \begin{cases} \kappa_\alpha \exp\left(-\gamma_\alpha \left|\bar{\delta}(k)\right|\right) & \text{eğer } \bar{\delta}(k-1)\,\delta(k) > 0 \\ -\varphi_\alpha\, \alpha(k) & \text{eğer } \bar{\delta}(k-1)\,\delta(k) < 0 \\ 0 & \text{diğer durumda} \end{cases} \tag{4.16}$$

formülünden elde edilir. Burada exp, üstel fonksiyonunu, κ_α, sabit öğrenme hızı skala faktörünü φ_α, sabit öğrenme hızı azaltma faktörü ve γ_α, sabit öğrenme hızı üstel faktörünü göstermektedir.

Benzer şekilde, k zamanındaki momentum hızı değişimi,

$$\Delta\mu(k) = \begin{cases} \kappa_\mu \exp\left(-\gamma_\mu \left|\bar{\delta}(k)\right|\right) & \text{eğer } \bar{\delta}(k-1)\,\delta(k) > 0 \\ -\varphi_\mu\, \mu(k) & \text{eğer } \bar{\delta}(k-1)\,\delta(k) < 0 \\ 0 & \text{diğer durumda} \end{cases} \tag{4.17}$$

ifadesinden bulunabilir. Burada, κ_μ, sabit momentum hızı skala faktörünü φ_μ, sabit momentum hızı azaltma faktörü ve γ_μ, sabit momentum hızı üstel faktörünü göstermektedir.

Şunada dikkat edilmelidir ki öğrenme ve momentum hızları, onların azalması ve artmasını kontrol eden farklı sabitlere sahiptir. $\bar{\delta}(k)\,\delta(k)$'nın işareti sezgisel olarak artma mı yoksa azalma mı olduğunu tespit etmek için kullanılır. Azalmanın ayarlanması DBD'nin kurallarına göre belirlenir. Ancak öğrenme ve momentum oranlarının artırılması $\left|\bar{\delta}(k)\right|$'nın üstel azalan fonksiyonu olarak değiştirilir. Böylece,

küçük eğimli bölgelerde, büyük eğimli bölgelere nazaran daha büyük artırımlar yapılabilecektir. Bu, atlama problemi için kısmi bir çözümdür.

Ağırlık uzayında, osilasyonları ve aşırı atlamaları engellemek için, üst sınırlar her bir bağlantı öğrenme hızları ve momentum hızları üzerinde belirlenir. Matematiksel ifadesi, tüm bağlantılar için aşağıdaki gibidir.

$$\alpha(k) \leq \alpha_{max} \tag{4.18}$$

$$\mu(k) \leq \mu_{max} \tag{4.19}$$

Burada, α_{max} öğrenme katsayısının üst sınırı ve μ_{max} ise momentum katsayısının üst sınırıdır.

Sonuç olarak, GDBD algoritmasına düzeltme özellikli bir hafıza yerleştirilmiştir. Kullanıldığı zaman, eğitme datalarının sunumu her iterasyondan sonra, hata değerlendirilir. Hata, $E(k)$, bir önceki minimum hatadan küçük ise ağırlıkları o anki en iyiler olarak hafızada saklanır. Düzeltme tolerans parametresi λ, düzeltme sürecini kontrol eder. O andaki hata önceki minimum hatayı aşarsa, yani

$$E(k) > E_{min}\lambda \tag{4.20}$$

olursa, tüm bağlantı ağırlıkları, hafızada saklı olan en iyi değerlerle değiştirilir. Ayrıca, öğrenme ve momentum hızları, düzeltmeyi başlatmak için azaltılır.

4.4.4 Hızlı Yayılım (HY)

Çok katlı perseptronların eğitilmesi için kullanılan ve sezgisel bir öğretme algoritması olan hızlı yayılım (Quickprop) algoritması, Scott Fahlman tarafından geliştirilmiştir (Fahlman 1988). Sezgisel algoritmalar, tecrübeye ve eğitilmiş tahmine

dayalı kurallar içermektedir. Bu algoritmalar optimum çözümü garanti edemeyip sadece ona yakın çözümü bulmayı garanti etmektedir.

Problem uzayında mümkün olan en kısa sürede, bir çözüm bulmak için ağırlık uzayında eğimi (gradient) azaltmak gerekebilir. Bunun yerine çözümden fazla uzaklaşmaksızın mümkün olan en büyük adımlarla kısa sürede çözüme ulaşmak arzu edilir. Hata fonksiyonunun eğim derecesi biliniyorsa, güvenli bir şekilde ağırlık uzayında daha büyük adımla çözüme ulaşılabilir. Bu algoritmada iki geleneksel yaklaşım birleştirilerek çözülmüştür. Bunlar;

- Hesaplamanın geçmişteki durumu hakkında bazı sezgilere dayanan (genel veya her bir ağırlık için ayrı) dinamik olarak ağırlıkların ayarlanması.

- Her bir ağırlığa göre hatanın ikinci türevinin belirgin kullanımı

Hızlı yayılım algoritması bir ikinci dereceli metoddan daha sezgiseldir ve çok az da olsa Newtonmetoduna dayanan bu metodda iki kabul kullanılır.

- Her bir ağırlık için, ağırlık hata eğrisi kolları yukarı doğru açık olan bir parabol ile yaklaştırılabilir.

- Hata eğrisinin eğilimindeki değişim, diğer tüm ağırlıkların aynı andaki değişiminden etkilenmez.

Genellikle hızlı yayılım algoritmasının performans testleri diğer tekniklerle karşılaştırıldığında oldukça iyidir. Bu algoritma, özellikle gürültü seviyesi az olan dataya sahip problemlerde, GDBD kadar iyi sonuç vermektedir (Neuralwork 1996). Hızlı yayılım algoritması gürültü seviyesi fazla olan problemlerde çok iyi çalışır ancak gürültü seviyesi yüksek problemlerde GDBD algoritması kullanılmalıdır (Neuralwork 1996).

Bu algoritmada ağırlık hızlandırma ve kırpılması ihmal edilerek k1 den k'ye

kadar ağırlıklardaki değişim,

$$\Delta w(k) = \varepsilon_l L(k) + \alpha_q Q(k) \tag{4.21}$$

formülü ile verilir. ε_l ve α_q sırasıyla, lineer ve quadratik ifadeler için öğrenme katsayılarıdır.

$$L(t) = \begin{cases} h(k) & \text{eğer } h(k)\,h(k-1) \geq 0 \\ 0 & \text{diğer durumlarda} \end{cases} \tag{4.22}$$

ve

$$Q(k) = \begin{cases} \mu_b \Delta w(k-1) & \text{eger } h(k)\left(h(k) - \left(\dfrac{\mu_b}{\mu_b + 1} \right) h(k-1) \right) \geq 0 \\ \Delta q(k) & \text{diger durumlarda} \end{cases} \tag{4.23}$$

eşitlikleri ile verilir. Bu ifadelerde μ_b momentum büyüme faktörü, $h(k) = \partial E / \partial w(k)$ eğimi ve

$$\Delta q(k) = \frac{\Delta w(k) h(k)}{h(k-1) - h(k)} \tag{4.24}$$

ifadesi minimum adım miktarını göstermektedir. Buradan ağırlık fonksiyonunu güncelleştirmek için, delta ağırlık fonksiyonu ve ağırlık hızlandırma katsayısı aşağıdaki denklemde görüldüğü gibi işleme katılır. Ağırlıklar,

$$w(k) = (1-\delta) w(k-1) + \Delta w(k) \tag{4.25}$$

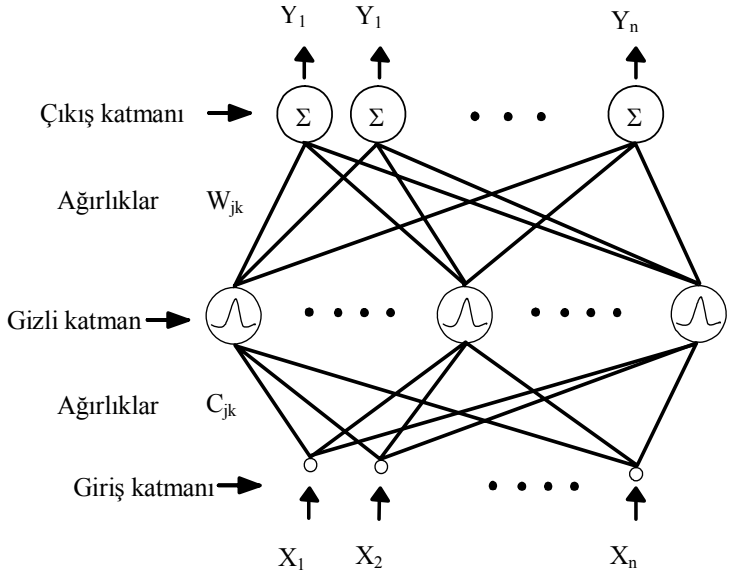
ifadesiyle güncellenir. Burada δ hızlandırma katsayısıdır. Son olarak ağırlık çok küçük ise, 0 alınarak kırpılır. Burada,

$$|w(k)| < \kappa \text{ ise } w(k) = 0 \text{ alınır.} \tag{4.26}$$

burada, κ ağırlık kırpılma faktörüdür.

4.5. Radyal Tabanlı Yapay (RTY) Sinir Ağı Yapısı

RTY sinir ağı yapısı Şekil 4.12'de gösterilmektedir. ÇKP yapısına alternatif bir YSA yapısı olarak bilinir. RTY sinir ağı yapısı, genellikle radyal olarak simetrik arakatman işlemci elemanlarının harici olarak ifade edilmesiyle oluşan bir ağ tipidir (Moody ve Darken 1988, Moody ve Darken 1989). Buna benzer algoritmalar, enterpolasyon ve yakınsama problemleri üzerine çalışan Powell (1987), Broomhead ve Lowe (1989) gibi araştırmacılar tarafından da sunulmuştur.

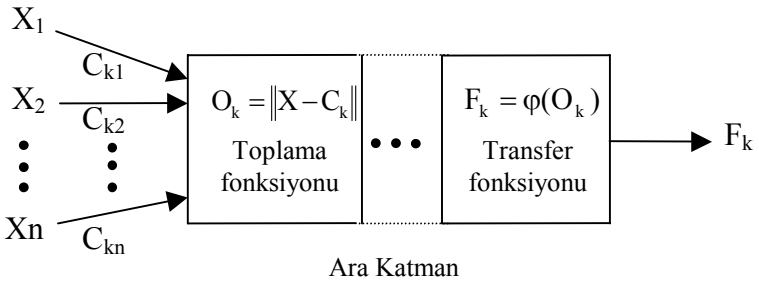


Şekil 4.12 RTY Sinir ağı mimari yapısı.

Bu yapı içerisinde bir işlemci elemanda temel olarak şu üç özellik bulunur.

1. Giriş ve ara katmanlar arasındaki ağırlık vektörü olarak saklanan, giriş uzayında bir merkez vektörü,
2. Giriş vektörünün merkezden ne kadar uzak olduğunu belirlemek için uzaklık ifadesi (bu ifade standart Öklit uzaklığı olarak alınır).
3. Tek değişkenli olan ve uzaklık fonksiyon çıkışını giriş olarak alan işlemci elemanın çıkış değerini belirleyen bir aktivasyon (transfer) fonksiyonu.

Şekil 4.13'de ara katman işlemci eleman yapısı gösterilmektedir.



Şekil 4.13 RTY YSA yapısının ara katman işlemci eleman yapısı.

f(x) transfer fonksiyonu

$$f(x) = \varphi(\|X - C\|) \tag{4.27}$$

ifadesinden bulunur. Daha açık bir ifadeyle ara katman çıkış ifadesi,

$$F_k = e^{\frac{-O_k}{\sigma_k^2}} \tag{4.28}$$

$$O_k = \|X - C_k\| = \sqrt{\sum_{i=1}^{n} (X_i - C_{ki})^2}, \quad k = 1,2,3,\ldots,n \tag{4.29}$$

Bu ifadede X, çok boyutlu giriş vektörünü, C giriş vektörü ile aynı boyutta merkez vektörünü ve σ ise standart sapma değerini gösterir. Ara katman işlemci elemanlarının eğitilmesi için C_k ağırlıkları ile σ_k genişliklerinin belirlenmesi gerekmektedir. Ara katman işlemci eleman fonksiyonlarının merkezi konumlarının (C_k) belirlenmesi için literatürde bazı yöntemler önerilmiştir (Moody ve Darken 1989, Lowe 1989, Nowlan 1989, Chen ve ark. 1991, Leonard ve ark. 1991, 1992, Jang ve ark. 1997).

RTY Yapısı İle ÇKP Yapısı Arasındaki Farklılıklar :

RTY'ları ÇKP ağlarından daha hızlı öğrenir. Fakat artan girişle çok yüksek

boyutlu RTY yapılarına ihtiyaç duyulur. RTY ağ yapısı genellikle sistem modelleme, tahmin ve sınıflandırma problemlerinin çözümü için kullanılır. RTY yapılarının, ÇKP ağlarına göre üstünlükleri aşağıdaki gibidir;

- Geriyayılım algoritması ile eğitilmiş ÇKP yapılarından daha hızlı öğrenir.
- Karar sınırlarının belirlenmesinde ve sınıflandırmada ÇKP yapılarından daha doğru sonuçlar verebilir (Leonard ve ark. 1991).
- RTY'sının ara katmanı ÇKP yapısı ara katmanından daha doğal bir yorumlama yeteneğine sahiptir. RTY'sı ara katman elemanları giriş uzayı için bir yoğunluk fonksiyonu temsil edebilir ve yeni bir giriş vektörünün eğitim vektörü olarak aynı dağılımın parçası olma olasılığını ortaya çıkarabilir (Leonard ve ark. 1991).

Bu yapılarının dezavantajları ise;

- RTY'sının eğitiminin ilk aşamasında merkez vektörünün belirlenmesi için kontrolsüz öğrenme yönteminin kullanılması durumunda önemli diskriminant bilgileri kaybolabilir.
- Regresyon problemleri etkin bir çözüm için sınırlandırılmamış aktivasyon fonksiyonlarına ihtiyaç duyar.
- ÇKP yapıları bir çok yapısal formda olabilir (Hagan ve ark. 1996).
- Örnek sayısı fazla olan çalışmalarda çok sayıda işlemci elemana ihtiyaç duyulur. (Güney ve ark. 2001)

4.6. Ses Tanımada Yapay Sinir Ağları

Ses olduğu takdirde ağın girişi sesin bazı bilgilerini içeren kısa süreli bir spektral gösterim yani özellik vektörü olabilir.YSA'lar herhangi bir eşleme için eğitilebildiğinden ses tanımada ses sinyalinin sınıflandırılması (modellenmesi) aşamasında kullanılabilir. Özellikle bundan sorumludur, çünkü diğer bazı sınıflandırma tekniklerinin karmaşıklık düzeyleri fazladır.

Ses tanımada kullanılabilecek farklı ağ tipleri vardır. Geri beslemeli bir ağda,

ağa verilen bir yapı sınıflayıcı gibi kullanılarak bazı çıktı birimleri elde edilir. Çıktı birimlerinden bazılarının hareketi kısıtlanırken, her sınıf için istenilen çıktı birimi etkin tutulur. Bu sınıflandırma fonksiyonunu daha da açarsak, nesnelerin gözlemlenebilir sınıflardaki olasılık yoğunluk fonksiyonunu tahmin etmek için YSA'lar kullanılmalıdır. Bir ağ böylece SMM sistemi gibi bir probabilistik olayı içine alan bir sistemin bileşeni de olabilir (Keller 1994).

İleri beslemeli veya geriyayılım algoritmaları ile eğitilen ağalara ek olarak sınıflandırmada kullanılabilecek farklı YSA yapıları vardır. Burada bahsedilebilecek iki yaklaşım vardır: Öğrenme Vektör Nicelendirmesi Ağı (Learning Vector Quantization-LVQ) sınıflar arasındaki farkı belirleyebilen iki katmanlı bir ağı eğiten bir algoritmadır (Kohonen 1988). Diğer yaklaşım ise Radyal Tabanlı Fonksiyon Ağıdır (Radial Basis Function-RBF) (Moody ve Darken 1989).

Bundan başka, ses verisi için kompleks veri gösterimlerinin yeni bir türünü oluşturmak için kullanılan YSA yapıları vardır. Bu tür ağlara bir örnek Kohonen ağları olarak bilinen Özörgütlemeli Harita Ağlarıdır (Self Organizing Maps-SOMs) (Kohonen 1989; Kohonen 1990). Bu ağlar Rekabete Dayalı Öğrenme (Competitive Learning) ile giriş verisinin yapısına göre kendilerini organize eder. Ayrıca SOM'lar konuşmanın gösterimi, analiz edilmesi ve karakterize edilmesinde faydalıdır. Örneğin, seslendirme bozukluklarını belirlemek için kullanılırlar.

YSA metotları farklı tipteki yapıları öğrenebilir. Bağlantılardaki ağırlık değerlerini değiştirerek öncelik verdiği yapıyı değiştirebilir. Bu nedenle YSA'daki başarı oranı fazladır, özellikle ses tanıma ve yorumlanmasında etkilidir. Bir YSA sistemi yüksek sayıda birim gerektirir ve eğitimi zordur. Dile ait işlemlerde insan yeteneğinin büyük bölümü beynimizde doğuştan gelen sinirsel bağlantılardan ortaya çıkmıştır. Doğuştan gelme ve öğrenme arasındaki bağlantı için modeller geliştirmek istenmiştir, ancak bu yıllardır makinalarla sağlanamamıştır. YSA sistemleri geliştirilirken ortaya çıkan problem bütün modellerin ağda gizli olmasıdır. Bu da ağda gerçekte ne olduğunu anlamamızı ve böylece ses tanıma işleminin nasıl gerçekleştiğini kavrayabilmemizi zorlaştırır (Holmes 2001).

Günümüze kadar, ses tanıma sistemlerinin büyük bölümünde yapay sinir ağları modelleri başarıyla uygulanmıştır. YSA'daki önemli özellikler şunlardır:

1. YSA'daki öğrenme algoritmaları tanıma için eğitilen ağın ayrı sınıfları arasındaki farkları yapısı gereği görebilir. Gerçekte bu gösterir ki, çeşitli YSA'ların çıktısı girişin verdiği çıktı sınıflarının bir sonraki gelenin olasılıklarını kesin tahmin etmesi olarak yorumlanabilir. Yapay sinir ağları doğru çıktı sınıfı ve alternatif hatalı sınıflar arasındaki farklılığı maksimize ederek eğitilir ve bundan dolayı sınıflandırma hatası minimize edilir.
2. Ağa giriş olarak verilen istatistiksel dağılımın özellikleri hakkında güçlü varsayımlarda bulunmaya gerek yoktur. Aynı zamanda, YSA'lar sınıflandırmayı gerçekleştirmek için çeşitli kısıtları birleştirebildiğine göre, özelliklerin bağımsız olduğunu farzetmek gereksizdir.
3. Girişte birkaç çerçeve bulunmasıyla veya geri besleme bağlantılarını kullanarak, ağ ses çerçeveleri arasındaki içerik bağlılığını modeller.
4. Çoğu YSA sistemleri geleneksel Von Neumann bilgisayar yapısında simule edilmesine rağmen, YSA yapılarının paralel ve kurallı doğası onları kolay donanım uygulamalarına uyumlu yapar.

Ayrışık kelime tanıma ve ses fonem sınıflandırılmasında fon sınırlarının belirlenmesi için ses ön-parçalara ayrılır. YSA sistemleri bu konuda SMM ile karşılaştırılabilir ve bazen SMM'lerden daha iyi sonuçlar verir. YSA'lar aynı zamanda bozuk ses sinyallerindeki gürültüyü azaltan ön uç uygulamaları (front-end applications) için faydalıdır. Bazen, YSA'lar kendi başlarına bazı sürekli ses tanıma görevleri için henüz etkili değildir. Bir bağlantıcı ağı eğitmek amacı ile giriş (özellik vektörleri) ve çıkış (tanınan sınıflar) arasındaki uygunluğu belirlemek gereklidir. Bu bilgi sürekli ses için uygulanamaz olma eğilimindedir, çünkü kelime sınırlarının yerleri ilerlerken genelde bilinmez. Daha genel olan, tekrarlanan yapılar ve Zaman Gecikmeli Yapay Sinir Ağları (Time Delay Neural Network-TDNN) sesin zamanla değişen yapısını yakalar. Bağlantıcı modeller zaman ölçeği değişimlerinin üstesinden kolaylıkla gelemez. Buna karşı SMM'lerin güçlü olan bir yanı spektral parçalama ve

geçici özellikleri olmasıdır. Bundan dolayı, sürekli sesi tanıma ve parçalamada YSA ve SMM birlikte daha verimlidir.

BÖLÜM 5

YSA İLE SES TANIMA MODELİNİN OLUŞTURULMASI

5.1 GİRİŞ

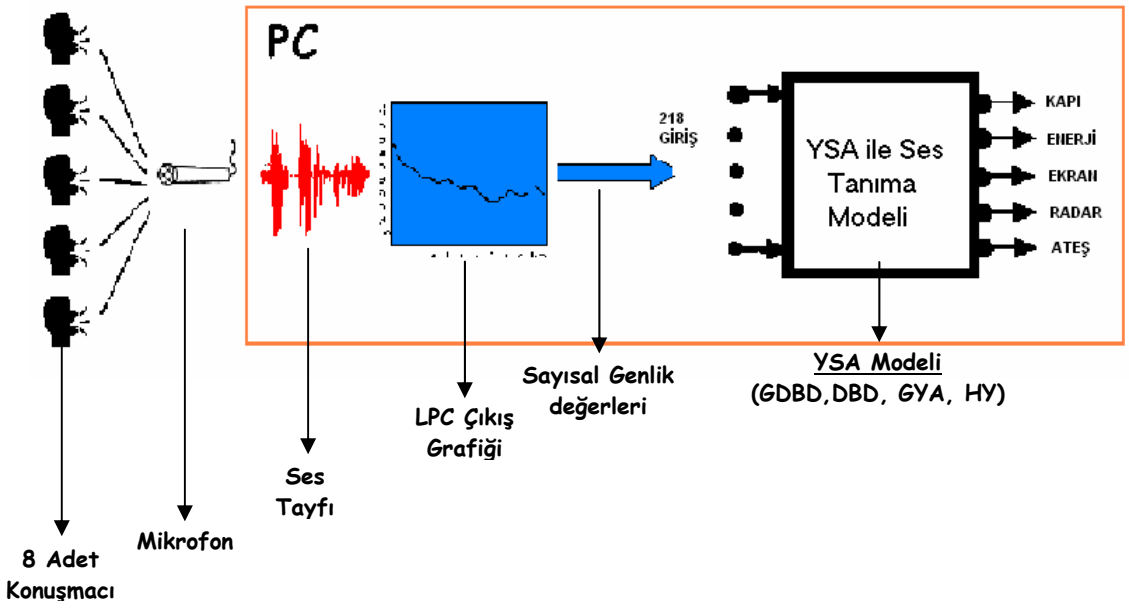
Bu çalışmada amaç insan sesinin bir bilgisayar aracılığıyla tanınmasıdır. Bu tanıma için öncelikle sesin dinlenmesi işlemi gerçekleştirilir. Dinleme işlemi gerçekleştirilirken, ses normal bir mikrofon ve ses kartı aracılığıyla örneklenerek alınır. Dinleme işlemi sürekli aktif olan bir işlemdir. Dinleme sonucunda değeri değiştirilebilen bir eşik seviyesini aşan güçte ve değeri değiştirilebilen bir zaman diliminden daha uzun süreli bir ses sinyali alınırsa ki aradaki kısa kesintiler göz ardı edilmiştir, bu bir kelime olarak algılanır ve değeri belli olan bir zaman dilimine yayılarak (shrink) uzun ve kısa söyleme etkisi ortadan kaldırılır. Daha sonra elde edilen yayılmış ses sinyali (genlik değerleri) her hangi bir kodlama tekniği ile kodlanır ve kodlanmış katsayılar seti oluşturulur.

Katsayılardan oluşan data seti bir giriş katmanı, bir ara katman ve bir çıkış katmanı olmak üzere toplam 3 katmanlı ve ileri beslemeli bir YSA'na giriş olarak verilir. Yapay sinir ağı için çıkış seti olarak, her kelimeyi temsil etmek üzere 1 ve 0'lardan oluşan bir set hazırlanır. Bu set sayesinde hangi kelimenin söylendiği çıkış için oluşturulacak elektronik bir düzen ile anlaşılabilir.. Yapay sinir ağının çıkışında ki koddan hangi kelimenin söylendiğinin anlaşılması için bir tabloya ihtiyaç vardır. Bu tablo eğitme aşamasında eğitici tarafından hazırlanmalı ve bir dosyaya kaydedilmelidir. Daha sonra bu tabloya bakarak hangi kelimenin söylendiği anlaşılabilir ve tanıma işlemi böylece tamamlanmış olur. Ayrıca burada hata hesabını da kullanarak konuşmacıya ait olan komutlar tanınmaya çalışıldı.

Ayrıca beş komuta göre konuşmacıyı ayırt etme işlemi de aynı YSA algoritmasındaki farklı iterasyonlarla sağlanmıştır.

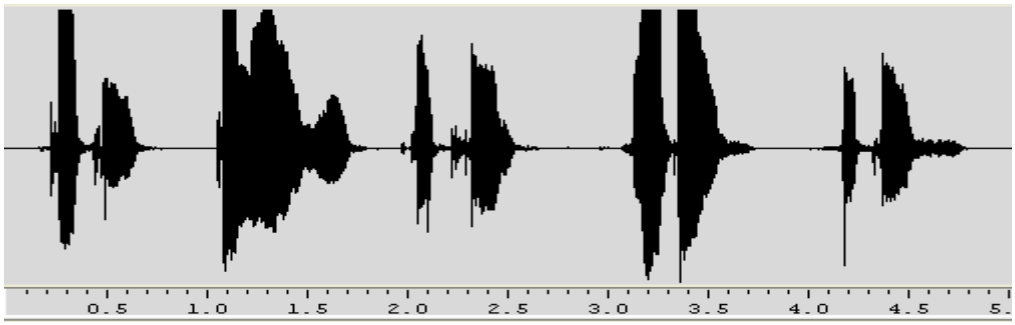
5.2 YSA Ses Tanıma Modeli

Genel hatları yukarıda ifade edilen bu çalışmanın genel bir şekli Şekil 4.1'de gösterilmiştir.

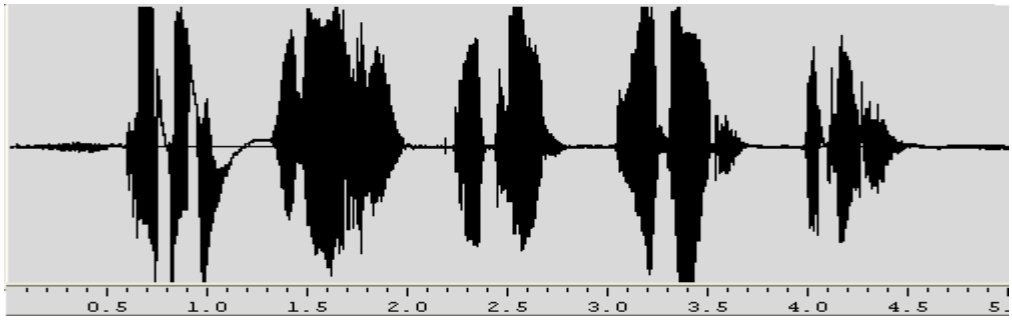


Şekil 5.1 Ses Tanıma modeli

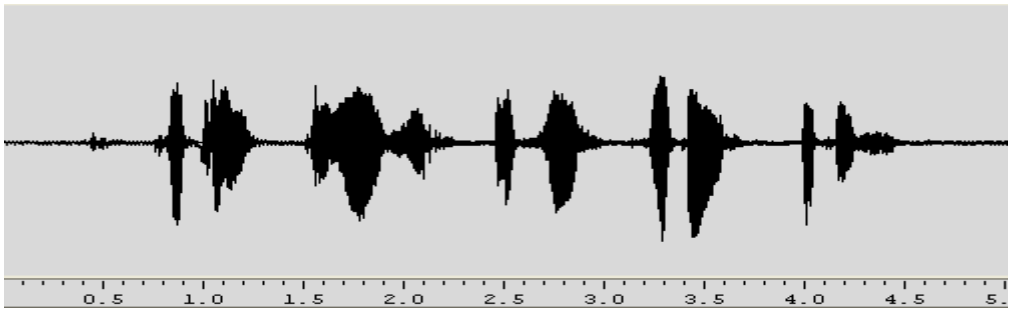
Şekilden de görüleceği gibi bir mikrofon marifeti ile 8 farklı kişiden, beş farklı komut kelime (Ateş, Ekran, Enerji, Kapı, Radar) sesleri üçer defa olmak üzere alınarak AVI uzantılı bir dosyaya kaydedilerek dosyalanmıştır. Örnek olması açısından her hangi 5 kişiden alınan bu ses dosyalarının Şekil 5.2a,b,c,d,e'de ve Şekil 5.3a,b,c,d,e'de sırasıyla dalga tayfları ile frekans–genlik grafikleri verilmiştir.



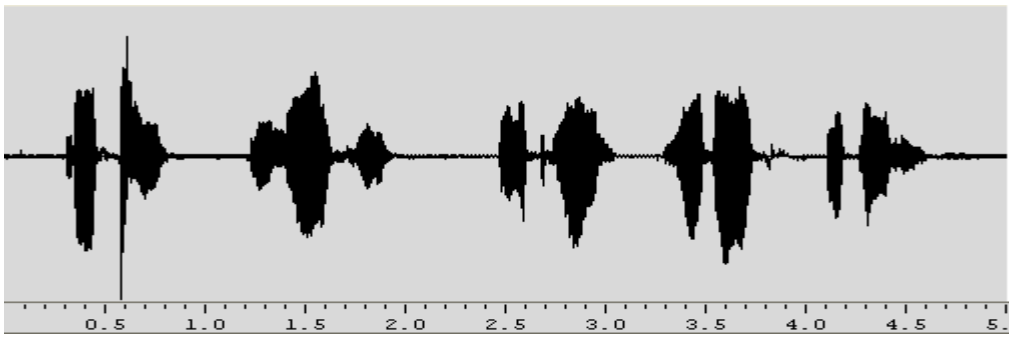
Şekil 5.2a 1. Konuşmacıya ait Kapı, Enerji, Ekran, Radar, Ateş kelimelerinin dalga tayfı



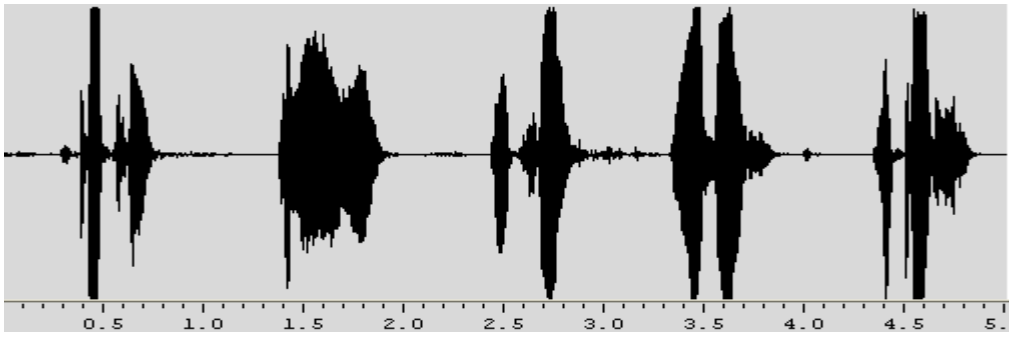
Şekil 5.2b 2. Konuşmacıya ait Kapı, Enerji, Ekran, Radar, Ateş kelimelerinin dalga tayfı



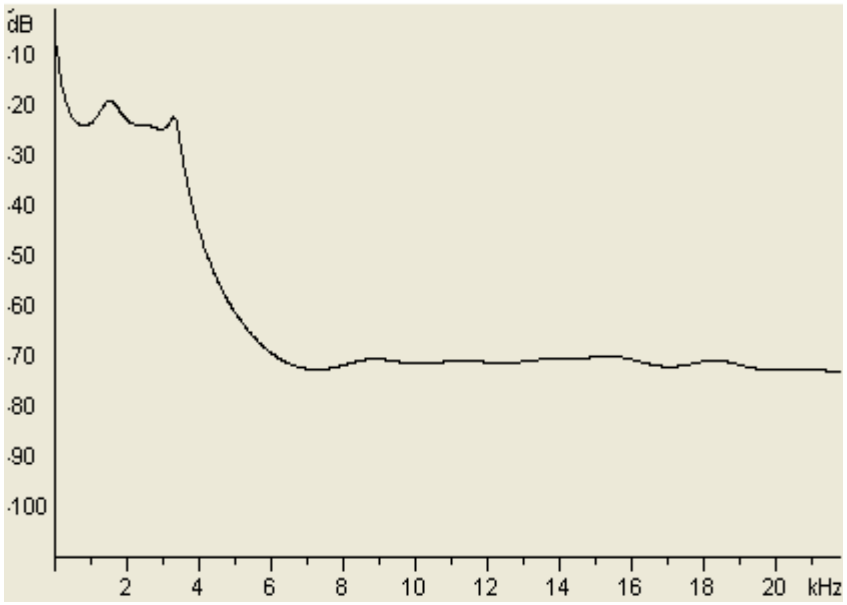
Şekil 5.2c 3. Konuşmacıya ait Kapı, Enerji, Ekran, Radar, Ateş kelimelerinin dalga tayfı



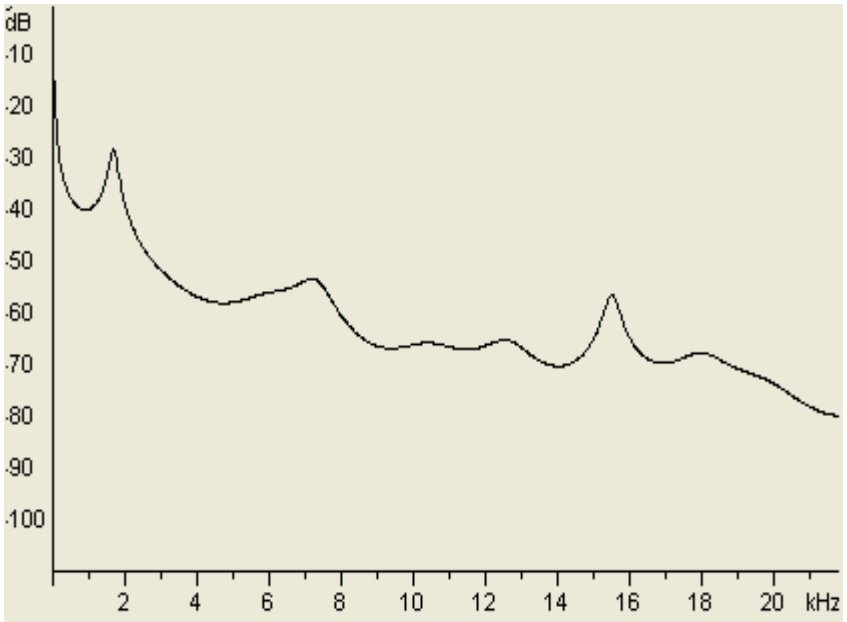
Şekil 5.2d 4. Konuşmacıya ait Kapı, Enerji, Ekran, Radar, Ateş kelimelerinin dalga tayfı



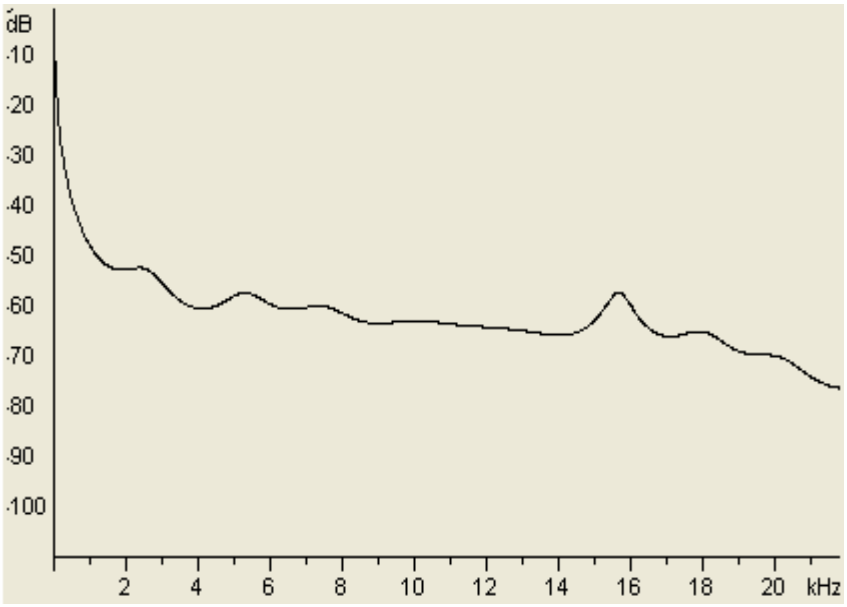
Şekil 5.2e 5. Konuşmacıya ait Kapı, Enerji, Ekran, Radar, Ateş kelimelerinin dalga tayfı



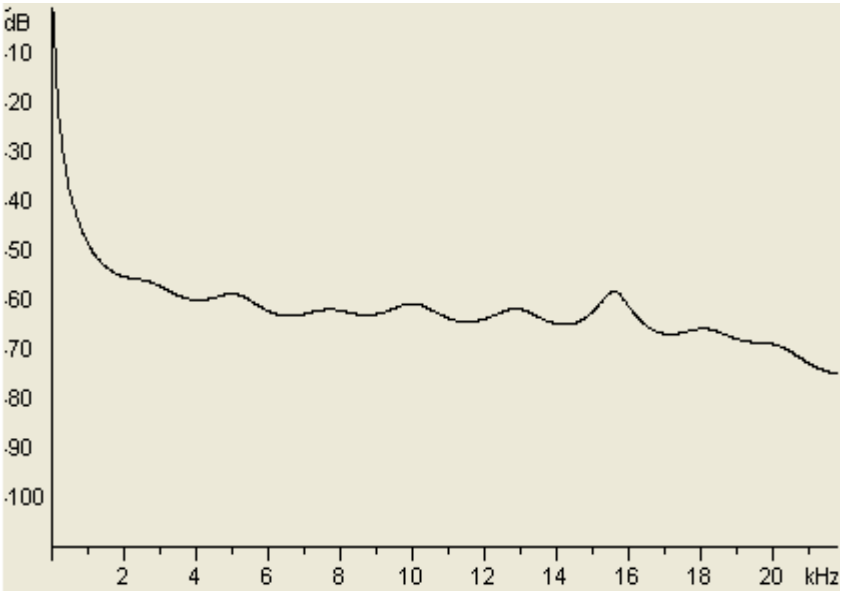
Şekil 5.3a 1. Konuşmacıya ait Kapı, Enerji, Ekran, Radar, Ateş kelimelerinin LPC Tekniği ile elde edilen frekans – genlik grafiği



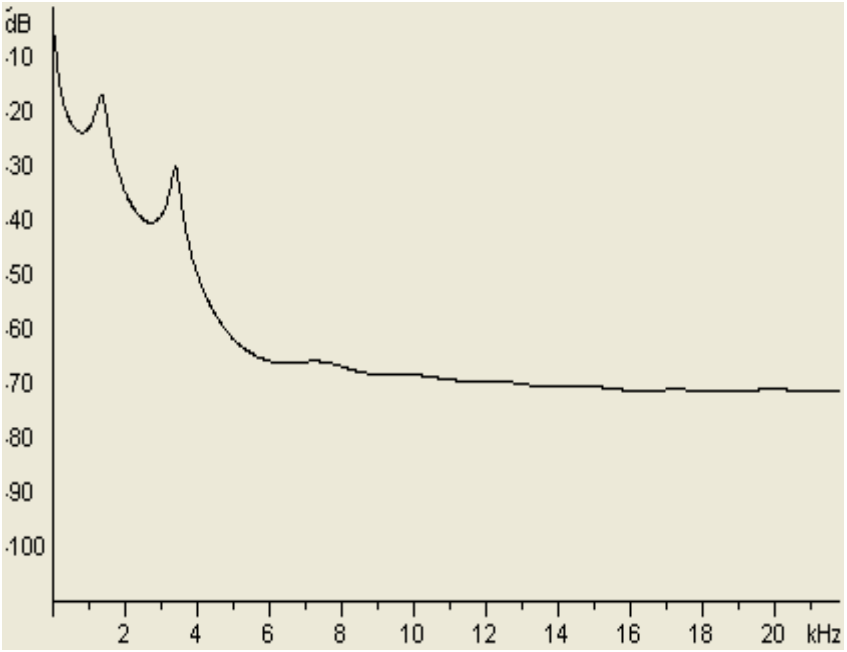
Şekil 5.3b 2. Konuşmacıya ait Kapı, Enerji, Ekran, Radar, Ateş kelimelerinin LPC Tekniği ile elde edilen frekans – genlik grafiği



Şekil 5.3c 3. Konuşmacıya ait Kapı, Enerji, Ekran, Radar, Ateş kelimelerinin LPC Tekniği ile elde edilen frekans – genlik grafiği



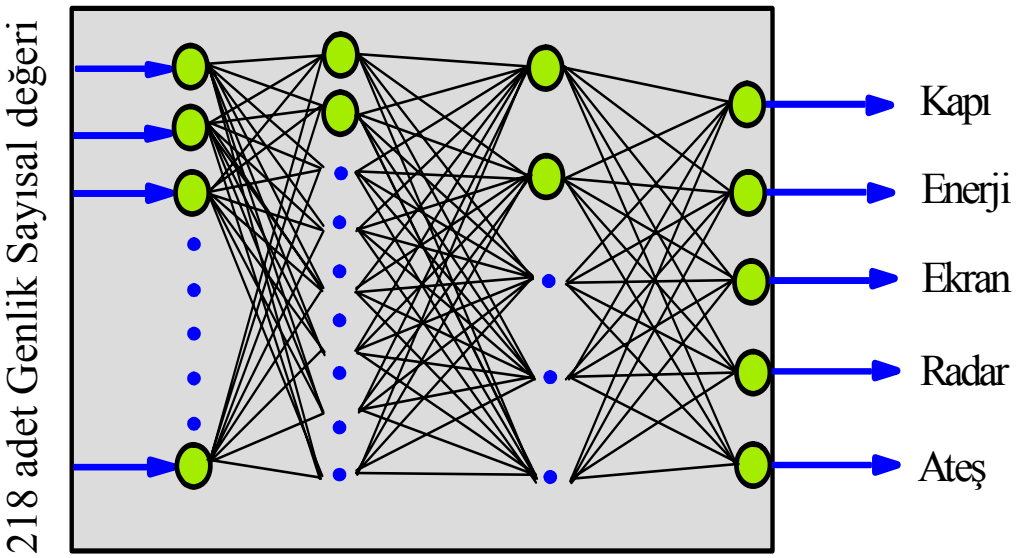
Şekil 5.3d Konuşmacıya ait Kapı, Enerji, Ekran, Radar, Ateş kelimelerinin LPC Tekniği ile elde edilen frekans – genlik grafiği



Şekil 5.3e 5. Konuşmacıya ait Kapı, Enerji, Ekran, Radar, Ateş kelimelerinin LPC Tekniği ile elde edilen frekans – genlik grafiği

Şekil 5.2a,b,c,d'deki ses örnekleri "wavesurfer" (WAVESURFER) isimli paket program kullanılarak Şekil 5.3a,b,c,d de gösterilen LPC Tekniği ile kelime kelime ve 0.5 sn'lik eşit sürelerde örneklenerek 10.7Hz 22039.2Hz aralığında genlik sayısal değerleri elde edilmiştir. Elde edilen bu komut kelimelerin 312.2Hz 4984.9Hz (218 adet örnekleme) aralığındaki frekanslara karşılık gelen genlik sayısal değerleri dikkate alınmış 4984.9 Hz üstü (280 adet örnekleme) frekanslara karşılık gelen genlik sayısal değerler dikkate alınmamıştır. Bunun nedeni insan kulağının algıladığı ötümlü, ötümsüz ve arka plan seslerin yaklaşık 300Hz5kHz aralığında olduğudur.

Daha sonra Şekil 5.4'de gösterildiği gibi 218 girişli ve her çıkışı bir komuta karşılık gelecek şekilde oluşturulan 5 çıkışlı Çok Katmanlı Perseptron yapısına ait bir YSA modeli oluşturulmuştur.



Şekil 5.4 218 girişli 5 çıkışlı iki ara katmanlı YSA modeli

6x2x5 = 70 adet (her altı kişi için ikişerli beş adet komut) çıkış komutu için 1 ve 0 ‘lardan oluşan çıkış modeli oluşturulmuştur. Çıkış komutlarını temsilen her hangi bir değer verilebilmesine karşın, çıkışların 1 ve 0 ‘larla oluşturulmasının nedeni, çıkıştan alınacak YSA değerlerinin elektronik bir düzen kurularak rahatlıkla kullanabilmesine yöneliktir. Bu sayede bu komutlar bir işlemci marifetiyle, kumanda ve kontrol amaçlı kullanılabilecektir. YSA modellerinde 3. Bölümde açıklanan Çok Katmanlı Perseptron’lara ait Genişletilmiş DeltaBarDelta (GDBD) Algoritması kullanılmış ve elde edilen eğitim sonuçları Tablo 1’de verilmiştir.

Tablo 1 6 Konuşmacıya ait YSA modeline verilen çıkışlar ve GDBD ile eğitilen YSA çıkışları

	Komut Kelimelere Karşılık Gelen Çıkış Kodları					YSA Modellerinden elde edilen GDBD Çıkışları				Eğitim	Komutlar
1. Konuşmacı	1	0	0	0	0	**1.049112**	0.018187	0.004912	0.013946	0.036381	**Ateş**
	1	0	0	0	0	**0.988432**	0.003988	0.001442	0.003135	0.007499	**Ateş**
	0	1	0	0	0	0.011538	**0.972612**	0.010325	0.019463	0.002682	**Ekran**
	0	1	0	0	0	0.024717	**0.996634**	0.002171	0.002054	0.00988	**Ekran**
	0	0	1	0	0	0.031435	0.018358	**0.997279**	0.015821	0.008679	**Enerji**
	0	0	1	0	0	0.028855	0.05742	**0.970267**	0.032583	0.03727	**Enerji**
	0	0	0	1	0	0.016213	0.012268	0.008859	**1.029526**	0.021753	**Kapı**
	0	0	0	1	0	0.00166	0.009036	0.002638	**0.987195**	0.015151	**Kapı**
	0	0	0	0	1	0.004897	0.008318	0.006606	0.011432	**0.994752**	**Radar**
	0	0	0	0	1	0.006637	0.017723	0.020897	0.083001	**1.01155**	**Radar**
2. Konuşmacı	1	0	0	0	0	**0.998917**	0.009799	0.017816	0.023257	0.008269	**Ateş**
	1	0	0	0	0	**0.970059**	0.02645	0.019343	0.007253	0.018881	**Ateş**
	0	1	0	0	0	0.010034	**1.028394**	0.010998	0.010026	0.01138	**Ekran**
	0	1	0	0	0	0.012031	**0.970289**	0.02096	0.002615	0.027736	**Ekran**
	0	0	1	0	0	0.001202	0.005721	**0.991945**	0.002486	0.012913	**Enerji**
	0	0	1	0	0	0.026878	0.031863	**1.009249**	0.002563	0.03205	**Enerji**
	0	0	0	1	0	0.028221	0.011764	0.034423	**0.962416**	0.01008	**Kapı**
	0	0	0	1	0	0.013338	0.014172	0.002293	**1.01445**	0.012936	**Kapı**
	0	0	0	0	1	0.005167	0.016851	0.002884	0.025648	**1.000085**	**Radar**
	0	0	0	0	1	0.035438	0.010332	0.01143	0.010584	**0.997264**	**Radar**
3. Konuşmacı	1	0	0	0	0	**0.996138**	0.020762	0.010746	0.07461	0.043397	**Ateş**
	1	0	0	0	0	**0.951855**	0.049893	0.016448	0.020358	0.046527	**Ateş**
	0	1	0	0	0	0.03565	**1.064084**	0.087597	0.010175	0.044507	**Ekran**
	0	1	0	0	0	0.026258	**0.97745**	0.045428	0.051192	0.023012	**Ekran**
	0	0	1	0	0	0.048061	0.031056	**1.028218**	0.000043	0.043694	**Enerji**
	0	0	1	0	0	0.03204	0.013502	**0.979402**	0.00417	0.031784	**Enerji**
	0	0	0	1	0	0.013351	0.008038	0.03992	**1.009563**	0.035908	**Kapı**
	0	0	0	1	0	0.02361	0.004502	0.020817	**0.993027**	0.001687	**Kapı**
	0	0	0	0	1	0.011203	0.060753	0.015148	0.007035	**0.999646**	**Radar**
	0	0	0	0	1	0.051682	0.019304	0.045576	0.000001	**0.970192**	**Radar**
4. Konuşmacı	1	0	0	0	0	**1.018331**	0.016634	0.040229	0.032205	0.021129	**Ateş**
	1	0	0	0	0	**1.001589**	0.012777	0.028919	0.019187	0.015053	**Ateş**
	0	1	0	0	0	0.022173	**0.989991**	0.012132	0.026236	0.016233	**Ekran**
	0	1	0	0	0	0.000138	**1.020538**	0.0212	0.019496	0.00873	**Ekran**
	0	0	1	0	0	0.000036	0.009686	**0.99817**	0.022111	0.014614	**Enerji**
	0	0	1	0	0	0.001403	0.007772	**0.983052**	0.023781	0.016427	**Enerji**
	0	0	0	1	0	0.013745	0.014799	0.005445	**1.002108**	0.019624	**Kapı**
	0	0	0	1	0	0.022788	0.028437	0.037425	**0.980949**	0.029115	**Kapı**
	0	0	0	0	1	0.007017	0.016467	0.002317	0.019344	**1.002474**	**Radar**
	0	0	0	0	1	0.032129	0.007826	0.00995	0.015901	**0.98982**	**Radar**

5. Konuşmacı	1	0	0	0	0	**0.990306**	0.03883	0.015397	0.018023	0.011755	**Kapı**
	1	0	0	0	0	**1.054759**	0.044474	0.006137	0.01836	0.034086	**Kapı**
	0	1	0	0	0	0.005001	**0.998947**	0.014649	0.02311	0.008278	**Enerji**
	0	1	0	0	0	0.008342	**0.992584**	0.022558	0.007701	0.021947	**Enerji**
	0	0	1	0	0	0.012058	0.0337	**1.012555**	0.014696	0.005681	**Ekran**
	0	0	1	0	0	0.000218	0.009839	**0.994376**	0.000443	0.002248	**Ekran**
	0	0	0	1	0	0.01273	0.012525	0.028494	**0.98877**	0.001931	**Radar**
	0	0	0	1	0	0.008886	0.0066	0.005682	**1.011645**	0.009036	**Radar**
	0	0	0	0	1	0.004391	0.001583	0.020635	0.012561	**0.974152**	**Ateş**
	0	0	0	0	1	0.002517	0.008598	0.011183	0.003225	**1.002876**	**Ateş**
6. Konuşmacı	1	0	0	0	0	**0.994347**	0.007923	0.020546	0.025813	0.006466	**Kapı**
	1	0	0	0	0	**1.006017**	0.002938	0.013618	0.000417	0.001876	**Kapı**
	0	1	0	0	0	0.000786	**0.975552**	0.018173	0.014327	0.003757	**Enerji**
	0	1	0	0	0	0.005675	**1.001249**	0.001589	0.006802	0.007572	**Enerji**
	0	0	1	0	0	0.009625	0.006522	**0.988883**	0.001305	0.023907	**Ekran**
	0	0	1	0	0	0.044653	0.024239	**0.994488**	0.002358	0.004375	**Ekran**
	0	0	0	1	0	0.027077	0.002953	0.003664	**0.9953**	0.004252	**Radar**
	0	0	0	1	0	0.007582	0.002017	0.012013	**1.003701**	0.001188	**Radar**
	0	0	0	0	1	0.015241	0.000792	0.007098	0.014469	**0.996039**	**Ateş**
	0	0	0	0	1	0.033917	0.006399	0.007314	0.004585	**1.006773**	**Ateş**

Tablo 2'de, Tablo 1'de eğitim sonuçları verilen ilk altı kişiye ait eğitim sonuçlarıyla 7. kişiye ait verilerinin test sonuçları görülmektedir.

Tablo 2 7. Konuşmacıya ait Test Çıkışları

1	0	0	0	0	**0.00491**	0.07751	0.16184	0.022291	0.964703
0	1	0	0	0	0.11871	**0.12438**	0.12421	0.11459	1.124935
0	0	1	0	0	0.11761	0.07464	**0.115772**	0.10285	0.477109
0	0	0	1	0	0.04103	0.03764	0.07278	0.997682	0.130737
0	0	0	0	1	0.06313	0.09422	0.01253	0.05572	1.086958

Tablo 3'de, Tablo 1'de eğitim sonuçları verilen ilk altı kişiye ait eğitim sonuçlarıyla 8. kişiye ait verilerinin test sonuçları görülmektedir.

Tablo 3 8. Konuşmacıya ait Test Çıkışları

1	0	0	0	0	**0.095763**	0.03799	0.12492	0.01959	0.89354
0	1	0	0	0	0.131247	**0.010641**	0.746484	0.018422	0.501576
0	0	1	0	0	0.148587	0.031691	**0.573558**	0.028472	0.544249
0	0	0	1	0	0.028572	0.042427	0.060684	1.043222	0.0557
0	0	0	0	1	0.023975	0.073362	0.091462	0.10371	1.058276

Tablo 4'de, Tablo 1'de eğitim sonuçları verilen ilk altı kişiye ait eğitim sonuçlarıyla 5. kişiye ait olan fakat eğitime tabi tutulmamış verilerinin test sonuçları görülmektedir.

Tablo 4 5. Konuşmacıya ait Test Çıkışları

1	0	0	0	0	**0.611527**	0.09805	0.118299	0.179942	0.047471
0	1	0	0	0	0.048282	1.103828	0.11349	0.06115	0.12483
0	0	1	0	0	0.133063	1.122348	**0.1243**	0.02647	0.12448
0	0	0	1	0	0.01706	0.02223	0.01501	1.02373	0.003699
0	0	0	0	1	0.018073	0.120775	0.07387	0.02182	1.027492

Tablo 5 4. Konuşmacıya ait Test Çıkışları

1	0	0	0	0	0.98095	0.014286	0.067446	0.004039	0.025788
0	1	0	0	0	0.097993	**0.243701**	0.97647	0.120222	0.120538
0	0	1	0	0	0.080964	0.069972	1.074993	0.004663	0.107766
0	0	0	1	0	0.038234	0.093513	0.144772	0.937159	0.047401
0	0	0	0	1	0.036511	0.091539	0.011401	0.011522	1.074376

Tablo 1, Tablo 2 Tablo 3 Tablo 4 ve Tablo 5 de görüldüğü gibi YSA modeli için ilk 6 konuşmacıya ait 5 komut bilgisinin bulunduğu 2 sesin eğitilerek, sadece giriş bilgisi verilen 4. konuşmacı 5. konuşmacı 7. konuşmacı ve 8. konuşmacı ait 1 ve 0'lara yakınlığı denenmiş ve bu kişilere ait komut bilgisi test olarak üretilmiştir. Buna göre eğitim için oluşturulan set 60 (2x6x5) adet genlik değerinden meydana gelmektedir. Buna karşın test için oluşturulan set ise sadece 4. 5. 7. ve 8. kişiye ait ateş, ekran, enerji, kapı, radar kelimelerinin oluşturduğu 5 (1x5) temsil değerlerinden meydana gelmektedir.

5.3 Hata Analizi

Tablo 1'de verilen GDBD, modeli ile eğitilerek üretilen 4., 5., 7., 8. konuşmacılara ait test çıkış değerleri (Ateş, Ekran, Enerji, Kapı ve Radar), Algoritmaların performanslarının değerlendirilmesine yönelik, Mutlak Toplam Hata

(MTH), Ortalama Mutlak Toplam Hata (OMTH) analizleri Tablo 6'de verilen Hata ifadeleri ile hesaplanarak sonuçlar Tablo 7'de verilmiştir.

Tablo 6 Hata Analizinde kullanılan Hatalar ve İfadeleri

Mutlak Toplam Hata, $$MTH = \sum_{n=1}^{n} |t_i - t_h| \quad (5.1)$$

Ortalama Mutlak Toplam Hata, $$OMTH = \frac{1}{n} \sum_{n=1}^{n} |t_i - t_h| \quad (5.2)$$

t_i = Komut Kelimelere Karşılık Gelen Test Çıkış Kodları
t_h = Komut Kelimelere Karşılık Gelen YSA Test Çıkış Kodları
n = Eleman Sayısı

Tablo 7 GDBD Algoritmasının hata değerleri

YSA	**4. Konuşmacı**	**5. Konuşmacı**	**7. Konuşmacı**	**8. Konuşmacı**
Mutlak Toplam Hata (MTH)	3.153	4.158	7.091	6.128
Ortalama Mutlak Toplam Hata (OMTH)	0.631	0.832	1.418	1.225

Meydana getirilen YSA Modellerinde GDBD Algoritmasını iterasyon sayıları, ağ yapıları, Eğitim / Test RMS hata değerleri ve Eğitim süreleri çok sayıda denemeler sonucunda elde edilerek Tablo 8'de verilmiştir. Bu değerler çok sayıda deneme ile en iyi ağ yapıları bulunarak sonlandırılmıştır. Fakat değerlendirilecek

olan ses sayılarının fazlalıklarından etkilenmemesi için her olasılık düşünülerek değerlendirilmiştir.

Tablo 8 GDBD Algoritmasının İterasyon Sayıları ve Ağ yapıları

YSA	İterasyon Sayısı	Ağ Yapısı		RMS Hata	
		1.Ara Katman	2.Ara Katman	Eğitim	Test
4. Konuşmacı	1.000.000	10 nöron	10 nöron	0.0336	0.2056
5. Konuşmacı	1.000.000	10 nöron	10 nöron	0.0336	0.2694
7. Konuşmacı	1.000.000	10 nöron	10 nöron	0.0353	0.3798
8. Konuşmacı	1.000.000	10 nöron	10 nöron	0.0344	0.3192

Nöron modellerinin eğitiminde kullanılan tüm ağırlıklar, başlangıçta düzgün olarak dağıtılmış rasgele değerlerden oluşmuştur. İşlemci elemanı çıkışı bir sigmoid fonksiyonundan geçirilerek YSA çıkışı alınmıştır.

Bu çalışmada kullanılan Algoritmaların performanslarını önemli ölçüde etkileyen parametreler en iyi değeri alabilmek adına her bir denemede değiştirilerek sonuçlar değerlendirilmiştir.

5.4 Komutlara Göre Konuşmacı Tanıma

Bu kısımda verilen diğer bir çalışmada ilk 6 kişiden alınan beş farklı komut kelime (Ateş, Ekran, Enerji, Kapı, Radar) sesleri, komutlar ayrı ayrı incelenerek konuşmacıları tanıma işlemi gerçekleştirilmiştir. Altı konuşmacının ilk iki ATEŞ komutu eğitime tabi tutulmuş ve örnek olarak 1. konuşmacının eğitime tabi tutulmayan üçüncü ATEŞ komutu teste tabi tutularak kişiler tanınmıştır. Aynı işlem 5 komut içinde ayrı ayrı tekrarlanmış ve sonuçları aşağıda verilmiştir.

Tablo 9 ATEŞ komutunun 6 konuşmacıdaki tanınma testi

1	0	0	0	0	0	**0.768233**	0.734165	0	0.000015	0.000012	0.000982
0	1	0	0	0	0	0.163956	**0.83665**	0.005701	0	0	0
0	0	1	0	0	0	0.071764	0	**0.991379**	0	0.142406	0.00001
0	0	0	1	0	0	0	0	0.000011	**0.999833**	0.019791	0
0	0	0	0	1	0	0	0.192849	0	0.000344	**0.874199**	0.000058
0	0	0	0	0	1	0	0	0	0	0.000372	**0.999985**

Tablo 10 EKRAN komutunun 6 konuşmacıdaki tanınma testi

1	0	0	0	0	0	**0.999879**	0.000026	0.000247	0	0	0.000068
0	1	0	0	0	0	0.026103	**0.912281**	0	0.000007	0	0
0	0	1	0	0	0	0.000024	0.000018	**0.999947**	0	0.000084	0
0	0	0	1	0	0	0	0.000008	0	**0.993336**	0.00965	0.000007
0	0	0	0	1	0	0	0.00002	0	0.00017	**0.998004**	0.000015
0	0	0	0	0	1	0.000007	0	0	0	0.000008	**0.999983**

Tablo 11 ENERJİ komutunun 6 konuşmacıdaki tanınma testi

1	0	0	0	0	0	**0.999712**	0.000039	0.000053	0	0	0.000046
0	1	0	0	0	0	0	**0.806798**	0.000027	0	0.129618	0
0	0	1	0	0	0	0.000074	0.000007	**0.999889**	0	0.000017	0.000009
0	0	0	1	0	0	0	0	0	**0.999948**	0.000085	0.000018
0	0	0	0	1	0	0	0	0.000361	0.00004	**0.99998**	0
0	0	0	0	0	1	0.00012	0	0.000614	0	0	**0.999795**

Tablo 12 KAPI komutunun 6 konuşmacıdaki tanınma testi

1	0	0	0	0	0	**0.999972**	0.000027	0.000132	0	0.000047	0.000016
0	1	0	0	0	0	0	**0.999942**	0.00001	0.000198	0.000008	0
0	0	1	0	0	0	0.000011	0.000022	**0.987437**	0.000027	0.000052	0
0	0	0	1	0	0	0	0.000172	0.000052	**0.999823**	0.000063	0.000009
0	0	0	0	1	0	0	0.000012	0	0.000757	**0.999958**	0.00002
0	0	0	0	0	1	0.002366	0	0	0	0.000113	**0.993072**

Tablo 13 RADAR komutunun 6 konuşmacıdaki tanınma testi

1	0	0	0	0	0	**0.998052**	0.00106	0	0	0	0.000022
0	1	0	0	0	0	0.000028	**0.999927**	0	0	0.00007	0.000068
0	0	1	0	0	0	0	0	**0.999797**	0	0.000638	0
0	0	0	1	0	0	0.00282	0	0	**0.914**	0.035356	0.000035
0	0	0	0	1	0	0.000011	0	0.000009	0	**0.999822**	0.000269
0	0	0	0	0	1	0	0	0	0.21312	0.000326	**1**

Yukarıdaki tablolarda sol taraftaki 1 ler 6 tane konuşmacıyı ifade etmektedir. Görüldüğü gibi her komut konuşmacıları ayırt etmekte oldukça iyi bir başarı göstermektedir. Bu çalışmaya ait GDBD algoritmasının iterasyon sayıları ve ağ yapıları tablo 14 de verilmiştir.

Tablo 14 GDBD Algoritmasının İterasyon Sayıları ve Ağ yapıları

YSA	**İterasyon Sayısı**	**Ağ Yapısı**	
		1.Ara Katman	**2.Ara Katman**
ATEŞ	1.500.000	12 nöron	0 nöron
EKRAN	1.500.000	12 nöron	0 nöron
ENERJİ	1.500.000	12 nöron	0 nöron
RADAR	1.500.000	12 nöron	0 nöron
KAPI	1.500.000	12 nöron	0 nöron

5.5. Sonuçların Değerlendirilmesi

Tablo 1 den görüldüğü gibi 6 konuşmacı için 2 şer defa alınan komutların GDBD de eğitilmesiyle elde edilen sonuçların hem "1" değerine hemde "0" değerine oldukça yakın sonuçlardır. Bu sonuçlardan konuşmacıların ses genlik değerlerinin kodlanarak seslerinin kolayca tanınabileceği anlaşılmaktadır.

Tabloda "1" ve "0" değerlerine olan yakınlıkların dışında tanıma işlemi test çıkışlarındaki mutlak toplam hata ve ortalama mutlak toplam hata oranları kıyaslanarak bir çözüme ulaşılmıştır. Tablodan da görüldüğü gibi eğitime tabi tutulan 6 kişinin içlerinden herhangi bir 2 konuşmacının (4. ve 5. kişilerin) tekrar sesleri alınarak test çıkışları incelenmiş ve mutlak değerleri hesaplanmıştır. Eğitime tabi tutulan kişi sesleri olduğu için, Ortalama Mutlak Toplam Hata değerleri her ikisinde de "1" in altında çıkmıştır. Fakat eğitime tabi tutulmayan diğer bir 2 ses olan 7. ve 8. sesler tekrara bu eğitim setine tabi tutulduğunda Ortalama Mutlak Toplam Hata değerlerinin hep "1" den yukarıda olduğu anlaşılmaktadır.

Bu çalışmada kullanılan ilk altı kişinin 2 si bayan ve 4 ü erkektir. Daha önce yapılan birçok ses tanıma işleminde ki en büyük sorunlardan birinin bayan sesindeki frekans yüksekliğinden dolayı tanınmasıydı ama bu sorun belli aralıklardaki frekansları alarak yaptığımız için bu sorunda ortadan kalkmıştır ve her seste olduğu gibi bayan sesinde de aynı ölçüde iyi sonuçlar elde edilmiştir.

Tablo 9, tablo 10, tablo 11, tablo 12 ve tablo 13 de ise eğitime ayrı ayrı tabi tutulan komutlardan faydalanarak konuşmacıların tanınması işlemi yapılmıştır. Eğitime tabi tutulan bu altı konuşmacının dışında kim olursa olsun bu komutlardan hangisini söylerse söylesin değerler "1" e ve gerekli yerlerde "0" a yakın olmayacağı için çıkış vermeyecektir. Fakat vurgulu harflerde veya türkçe karakterlerde, "ateş" komutunda ki "ş" harfi gibi komutlarda, azda olsa tanınma oranı düşmektedir.

BÖLÜM 6

SONUÇLAR

Günümüz teknolojisinde, sesin sayısal olarak işlenmesi analog işlenmesine göre bir takım üstünlükler taşır. Analog teknikler, ses işaretinin eşit süreli çerçevelere bölünmesi ve her çerçevenin zaman ve/veya frekans ekseninde dilimlere ayrılarak bu dilimlerin bir kripto algoritmasıyla karıştırılması prensibine dayanmaktadır. Karıştırılmış halde iletilen dilimler almaç biriminde yerlerine oturtularak orjinal ses işareti çerçevesi elde edilir. Bu yöntemlerde çerçeve uzunluğu yeterince uzun tutulmadığı için, kelimelerin kendini ele vermesi ya da iyi eğitilmiş bir kulakla çözülebilmesi mümkün olmaktadır. Aynı zamanda senkronizasyon için analog işaretlerin arasına sıkıştırılan sayısal işaretler kısa tutulmaları gerektiğinden düşük RF seviyelerinde kesilmeler ve hatalı çözümlemelerle karşılaşılmaktadır.

Bu tez çalışmasında, insan beyninin yapay olarak benzetimi çabalarıyla ortaya çıkan öğrenme, uyarlanabilirlik, genelleştirilebilme, az bilgi ihtiyaçları, paralel işlem yapabilme, işlem sürelerinin hızlı olması gibi özellikleriyle son yıllarda popüler hale gelen YSA hakkında geniş bilgi verilmiş ve 8 farklı kişiden 5 farklı komut kelime 3'er defa (farklı tonlarda) alınarak kaydedilmiş ve kaydedilen bu seslerin data örnekleri wavesurfer programı ile LPC tekniği kullanılarak komut kelimelerin 0.5 sn'lik sürelerde frekansgenlik parametreleri elde edilmiştir. Elde edilen bu değerler hazırlanan YSA modellerinden olan GDBD algoritması ile eğitilmiş ve son derece başarılı sonuçlar elde edilmiştir.

Diğer taraftan bu çalışmada gerçekleştirilen YSA modeli ile üretilen sonuç daha çok bir komut yerine hata hesabıyla bir komut zinciri kişiye göre tanınmış oldu. Ses tanıma konusunda yapılan çalışmalarda tanınacak olan kelimenin kısa olmasının yanı sıra çok uzun olanında tanınması bir o kadar zor olduğu vurgulanmıştır. Ama bu çalışmada bu sorunda hata analiziyle ortadan kaldırılmıştır. Bu sistem bir veya birden fazla motor, bir otomasyon sistemi, güvenlik amaçlı herhangi bir cihazın sadece bir

grup insan tarafından kontrolü, robot veya robot bir kola komutlar gönderilerek çeşitli amaçlarda işlevlerin yapılabilmesi gibi işlemlerdir.

Aynı zamanda bu eğitim, 7 kişi içinde yapılmış ve yaklaşık aynı sonuçlar alınmıştır yani kişi sayısıyla biraz oynansa da işlemlerdeki sonuçlar çok değişmemekte olduğu görülmektedir.

Diğer bir çalışma olan komutlara göre konuşmacı tanıma işleminin çok iyi sonuçlar verdiği görülmektedir. Bu sayede seçilen belli sayıdaki kullanıcının istenilen aletleri veya düzenekleri kullanması sağlanmış olacaktır. Daha birçok yerde bu çalışma sayesinde kişiye özel kullanım sağlanabileceği alanlar mevcuttur.

Ses kaydı aşamasında en önemli etken mikrofondur. Normal bir mikrofon hiçbir zaman için yeterli olmayacaktır. Daha gürültüsüz ve doğru ses kaydı daha doğru tanımayı sağlayacaktır. Yapay sinir ağları giriş ve çıkış sayıları az olan ve çok karmaşık olmayan matematiksel fonksiyonlarda oldukça başarılı bir öğrenme gösterirler. Fonksiyonlar karmaşıklaştıkça gizli katmandaki düğüm sayılarının arttırılması gerekir. Ancak çok fazla düğüm aynı zamanda öğrenmeyi yavaşlatmaktadır. Çok karmaşık fonksiyonlarda 3 katmanlı ileri beslemeli YSA yeterli olmayabilir, bu tür fonksiyonlarda hem ileri hem de geri beslemeli ve daha çok katmanlı YSA'lar kullanılmalıdır. Giriş veya çıkış sayılarının çok fazla artması YSA'nın öğrenmesini yavaşlatarak, öğrenme oranının makul bir zaman içinde düşük kalmasına sebep olurlar.

Bugünün ihtiyaçlarının karşılanabilmesi ve örneğin yaygın biçimde kullanılan hücresel telefonlar ve elektronik ses taklit sistemlerin çok yakında karşımıza çıkartacağı son derece zorlayıcı gerçek şartların üstesinden gelebilmek için yeni ve daha iyi tekniklerin geliştirilmesi gerekecektir. Bunun yapılabilmesi için ise bu alandaki araştırma ve geliştirme faaliyetleri için ortak bir yön oluşturulması gerekecektir. Bu alanda mühendisleri, özellikle adli alanda uzman dil bilimcileri ve bilgisayar uzmanlarını bir araya getiren konular arası araştırmalara acil ihtiyaç vardır.

Problemin bilimsel ve teknolojik boyutları düşünüldüğünde, ancak uluslararası işbirliği ve çabaların paylaşımının ümit verici bir perspektif olduğu görülecektir.

Ayrıca, tez bireysel görüşümü yansıtmaktadır, Türk Silahlı Kuvvetlerinin görüşünü yansıtmamaktadır.

BÖLÜM 7

KAYNAKLAR

- Altun A.A., Koçer H.E. "Güvenlik Alanına Yeni Bir Yaklaşım Biyometrik Sistemler" International XII. Turkish Symposium on Artificial Intelligence and Neural Networks – TAINN. 2003
- Aydın Ö. "Yapay Sinir Ağlarını Kullanarak Bir Ses Tanıma Sistemi Geliştirilmesi" Trakya Üniversitesi Fen Bilimleri Enstitüsü Yüksek Lisans Tezi 2005.
- Chen S., Cowan C.F.N. ve Grant P.M., "Orthogonal Least Squares Learning Algorithm for Radial Basis Function Networks," IEEE Transactions on Neural Networks, 302-309, 1991.
- Dudley, H. "The Vocoder". Bell Labs, Record 18, pp. 122–126, December 1939. Reprinted in: Schafer, R. W. and Markel, J. D. Speech Analysis. IEEE Press, 1979.
- Demiray A. "Objelerin Bölütlenmiş Görüntüleri Kullanılarak Yapay Sinir Ağlarıyla Tanınması" Anadolu Üniversitesi Fen Bilimleri Enstitüsü Elektrik-Elektronik Mühendisliği Anabilim Dalı Doktora Tezi 1998.
- Ergezer H., Dikmen M. ve Özdemir E. "Yapay Sinir Ağları Ve Tanıma Sistemleri". PİVOLKA, 2(6), 14-17. (2003).
- Fahlman S.E., "An Emprical Study of Learning Speed in Backpropagation Networks," Technical Report CMUCS88162, Carnegie Mellon University, 1988.
- Güney K., Gültekin S.S. ve Sağıroğlu Ş., "Design of Circular Microstrip Antennas Using Artificial Neural Networks." Proc. of 10th Turkish Symposium on Artificial Intelligence and Neural Networks (TAINN'2001), Gazimagusa, Turkish Republic of Northern Cyprus, s. 12-21, June 2001.
- Gültekin S.S., "Mikroşerit Anten Parametrelerinin Yapay Sinir Ağları ile Analizi ve Tasarımı", Selçuk Üniversitesi Fen Bilimleri Enstitüsü Elektrik Elektronik Müh. ABD., Doktora Tezi, Konya, 2002.
- Gültekin D., "Yapay Sinir Ağları Modeli ile Ses Tanıma", Selçuk Üniversitesi Fen Bilimleri Enstitüsü Bilgisayar Müh. ABD., Yüksek Lisans Tezi, Konya, 2004.

- Hopfield J.J., “Neural Networks and Physical Systems with Emergent Collective Computational Abilities,” In: Proc. National Academy of Sciences, USA, April, 79, s. 2554-2558, 1982.
- Haykin S., “Neural Networks, A Comprehensive Foundation”, Macmillan College Publishing Comp. Inc., 1994.
- Hagan M.T., Demuth H.B. ve Beale M., “Neural Network Design”, Boston PWS Publishing Company, 1996.
- Holmes, J., “Speech Synthesis and Recognition”, Taylor and Francis, 2001.
- Jacobs R.A., “Increased Rate of Convergence Through Learning Rate Adaptation,” Neural Networks, 1, s. 295-307, 1988.
- Jang J.S.R., Sun C.T. ve Mizutani E., “NeuroFuzzy and Soft ComputingA Computational Approach to Learning and Machine Intelligence”, PrenticeHall, Inc., 1997.
- Kohonen T., “Statistical Patern Recognition with Neural Networks:” Benchmark Studies, Porceedings of The Second Annual IEEE International Conference on Neural Networks, 1, 1988.
- Kohonen T., “SelfOrganisation and Associative Memory (3rd ed)”, SpRingerVerlag, Berlin, 1989.
- Kuş P. “Ses Sinyallerinin Düşük Hızda İletimi” Hacettepe Üniversitesi Fen Bilimleri Enstitüsü Yüksek Lisans Tezi. Ocak 1998.
- Kohonen, T., “The Self-Organizing Map”, Proceedings of IEEE International Conference on Neural Networks, San Diego,1990.
- Keller, E., “Fundamentals of Speech Synthesis and Speech Recognition”, John Wiley&Sons, 1994.
- Kurban R., Tunçkanat M. Ve Sağıroğlu Ş. “Elektronik-Haberleşme-Bilgisayar Ulusal Proje Aranıyor'03 Öğrenci Sempozyumu” proje 2003, 28-29 Nisan 2003.
- Leonard J.A., Kramer M.A., “Radial Basis Functions for Classifying Process Faults,” IEEE Control Systems, April, 11(3), s.31-38, 1991.
- Levenberg K., “A Method For the Solution of Certain Nonlinear Problems in Least Squares,” Quart.Appl. Math., 2, s. 164-168, 1944.
- Lowe D., “Adaptive Radial Basis Function Nonlinearities and The Problem of

Generalization," In Proceedings of The First IEEE International Conference on Artificial Neural Networks, London, 6, s.171-175, 1989.

- Marquardt D.W., "An Algorithm For LeastSquares Estimation Of Nonlinear Parameters," J. Soc. Ind. Appl. Math., 11, s. 431-441, 1963.
- Moody J. ve Darken C., "Fastlearning in networks of locallytuned processing units," Neural Computation, 1, 281-294, 1989.
- Moody J. ve Darken C., "Learning with Localized Receptive Fields," Proceedings of the 1988 Connectionist Models Summer School, San Mateo CA., Carnegie Mellon University, Morgan Kaufman, 1988.
- NeuralWare Handbook, Neural Computing: A Technology Handbook for Professional II/PLUS and Neural Works Explorer, Pittsburgh. USA., 1996.
- Nowlan S.J., "Maximum Likelihood Competitive Learning, Advances in Neural Information Processing Systems," San Matko, CA, 2, s. 574-582, 1989.
- Nwe T.L.; Foo S.W.; De Silva L.C. "Speech Emotion Recognition Using Hidden Markov Models" Elsevier Science, Speech Communication, vol. 41, iss. 4, pp. 603-623 (21) November 2003.
- Parker, D. B., "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, 1985.
- Peinado A.M., Sánchez V., Pérez-Córdoba J.L. ve De La Torre Á. "HMM-Based Channel Error Mitigation and its Application to Distributed Speech Recognition". Speech Communication, vol. 41/4, pp. 549-561, November 2003
- Türkoğlu, Ö., "Sayısal Hücresel Haberleşme" EMO İstanbul Şube Bülteni,Sayı:14, sayfa:42-28, 1998.
- Wavesurfer., About the Centre for Speech. Technologyhttp://www. Speech.kth. sectt/about. shtml" sectt/about. Shtml"
- Yang D.M. "High Quality Speech Transformation Based on Linear Prediction Coding and Pitch Synchronization" Ethesys, Communication Engineering, Master's Thesis, 2003.
- Yaparoğlu E.T. "Ses ile Kimlik Doğrulama" Gazi Üniversitesi Fen Bilimleri Enstitüsü Yüksek Lisans Semineri 2002.