167475

KOCAELİ ÜNİVERSİTESİ * FEN BİLİMLERİ ENSTİTÜSÜ

OTOMOBİL İÇİNDEKİ CİHAZLARIN SESLE KONTROLÜNE YÖNELİK KONUŞMACIDAN BAĞIMSIZ KONUŞMA TANIMA SİSTEMİNİN GERÇEK ZAMANLI UYGULAMASI

YÜKSEK LİSANS TEZİ

Elektronik Öğretmeni Ersoy KELEBEKLER

Anabilim Dalı : Elektronik ve Bilgisayar Eğitimi

Danışman: Yrd. Doç. Dr. Melih İNAL

MAYIS 2006

KOCAELİ ÜNİVERSİTESİ * FEN BİLİMLERİ ENSTİTÜSÜ

OTOMOBİL İÇİNDEKİ CİHAZLARIN SESLE KONTROLÜNE YÖNELİK KONUŞMACIDAN BAĞIMSIZ KONUŞMA TANIMA SİSTEMİNİN GERÇEK ZAMANLI UYGULAMASI

YÜKSEK LİSANS TEZİ

Ersoy KELEBEKLER

Tezin Enstitüye Verildiği Tarih: 17 Mayıs 2006

Tezin Savunulduğu Tarih: 06 Haziran 2006

Tez Danışmanı	Üye	Üye
Yrd. Doç. Dr. Melih İNAL	Prof. Dr. Yılmaz ÇAMURCU	Yrd. Doç. Dr. Serhat YILMAZ
(.....)	(.....)	(.....)

MAYIS 2006

OTOMOBİL İÇİNDEKİ CİHAZLARIN SESLE KONTROLÜNE YÖNELİK KONUŞMACIDAN BAĞIMSIZ KONUŞMA TANIMA SİSTEMİNİN GERÇEK ZAMANLI UYGULAMASI

Ersoy KELEBEKLER

Anahtar Kelimeler: Konuşma, Konuşma Tanıma, Otomobil İçerisinde Sesle Kontrol, Gürültü Bastırma, Gerçek Zamanlı Konuşma Tanıma

Özet: Teknolojinin ilerlemesi ile birlikte insan-makine arasındaki iletişim ihtiyacı günden güne artmaktadır. Bunun bir sonucu olarak makinelerin ses işareti kullanılarak kontrol edilmesi ihtiyacı doğmuştur. Bugüne dek yapılan çalışmalar sonucu başarıyla çalışan pek çok konuşma tanıma sistemi geliştirilmiştir. Yapılan başarılı çalışmalara rağmen konuşma tanıma sistemleri henüz amaçlanan verime ulaşamamıştır.

Otomobil içerisindeki cihazların sesle kontrolü ile ilgili çalışmalar 30 yıla yakın süredir devam etmektedir. Fakat çalışmalar sonucunda oluşturulan sistemler güvenilirlik ve maliyet açısından istenilen düzeyde olmadığı için ticari olarak yaygınlaşmamıştır. Otomobil içerisindeki gürültünün konuşma tanımayı zorlaştırması, sistemlerin verimini ve güvenilirliğini azaltmaktadır. Bu amaçla gerçek zamanlı konuşma tanıma sisteminde kullanılabilecek en iyi gürültü bastırma algoritması araştırılmıştır. Konuşma tanıma sistemlerinin verimini etkileyen diğer bir unsur sistemin konuşmacıya bağımlı veya konuşmacıdan bağımsız olmasıdır. Konuşmacıdan bağımsız konuşma tanıma sistemlerinin gerçekleştirilmesinde sistemin eğitiminde kullanılan konuşmacı sayısı ile sistemin verimi doğru orantılı olarak artar. Oluşturduğumuz konuşmacıdan bağımsız konuşma tanıma sistemi 10 erkek ve 10 bayan konuşmacının telaffuzları kullanılarak eğitilmiştir ve sistemin gerçek zamanlı uygulaması MATLAB 6.5 sürümünde Data Acquisition Toolbox kullanılarak gerçekleştirilmiştir.

REAL TIME APPLICATION OF SPEAKER INDEPENDENT SPEECH RECOGNITION SYSTEM FOR VOICE ACTIVATED CONTROL OF DEVICES IN CAR

Ersoy KELEBEKLER

Keywords: Speech, Speech Recognition, Voice Activated Control in Car, Noise Cancellation, Real Time Speech Recognition

Abstract: Requirement of communication between human and machine has been increasing together with proceeding of technology day by day. Consequently, requirement of voice activated control of machines has arisen. As a result of studies until now, numerous speech recognition systems working successfully have been developed. In spite of performed successfully studies, speech recognition systems haven't achieved level of desired performance, yet.

Studies about voice activated control in car have been carrying on about 30 years. But, end of the studies, performed systems haven't commercially became widespread due to their inadequate reliability and costs. Making difficulties effects on speech recognition of noise in car decrease speech recognition system's performance and reliability. For this aim, optimum noise cancellation algorithm able to use on real time speech recognition system has been sought. Another effect of altering speech recognition system's performance is being dependent speaker or independent speaker of system. For performing of speaker independent speech recognition systems, number of speaker used for training the system is directly proportional with the system's performance. Our performed speaker independent speech recognition system has been trained by using 10 male and 10 female utterance and system's real time application has been performed by using Data Acquisition Toolbox on MATLAB version 6.5

ÖNSÖZ ve TEŞEKKÜR

Konuşma tanıma sistemleri, önemini günden güne arttırmaktadır. Konuşma tanıma uygulamaları günlük yaşamımızın pek çok alanına dahil olmaya başlamıştır. Konuşma tanıma sistemlerinin otomobil içersinde kullanımının yaygınlaşmasıyla sürüş güvenliği ve konforu arttırılacaktır. Konuşma tanıma sistemlerinin araç içerisinde kullanılması çalışmaları uzun yıllardır sürdürülmesine karşın konuşma tanıma sistemlerinin güvenilirliği ve verimi arzulanan seviyeye ulaşmadığından, araç içi konuşma tanıma sistemleri ticari olarak yaygınlaşmamıştır. Bunun bir diğer nedeni araç içerisindeki gürültünün konuşma tanımayı zorlaştırmasıdır. Bu tezde gerçekleştirilen araç içi gürültünün bastırılmasında kullanılabilecek en iyi yöntemin araştırılması çalışmasından elde edilen sonuçların, ileride hızlı donanımlarla gerçekleştirilecek konuşma tanıma sistemlerinde gürültünün etkisini minimize etmek için kullanılması hedeflenmektedir. Ayrıca tez çalışmamız, konuşmacıdan bağımsız konuşma tanıma işleminin gerçek zamanlı uygulamasına bir örnek teşkil etmektedir.

Bu çalışmanın ortaya çıkarılmasında tavsiyeleri ile yol gösteren ve desteğini her zaman yanımda hissettiğim danışmanım Sayın Yrd. Doç. Dr. Melih İNAL'a sonsuz şükran ve saygılarımı sunarım. Ayrıca, çalışmam süresince yardımlarını ve desteklerini esirgemeyen hocalarım Öğr. Gör. Yavuz Selim FATİHOĞLU, Yrd. Doç. Dr. Celal ÇEKEN, arkadaşım Arş. Gör. Kerem KÜÇÜK'e ve tüm hocalarıma, arkadaşlarıma, çok değerli aileme teşekkürlerimi sunarım.

İÇİNDEKİLER

ÖZET ii
ABSTRACT iii
ÖNSÖZ ve TEŞEKKÜR iv
İÇİNDEKİLER v
SİMGELER DİZİNİ ve KISALTMALAR viii
ŞEKİLLER DİZİNİ x
TABLOLAR DİZİNİ xii

BÖLÜM 1. GİRİŞ 1
1.1. Tez Çalışmasının Amacı 2
1.2. Tez Çalışmasını Oluşturan Bölümler 3

BÖLÜM 2. KONUŞMA VE KONUŞMA TANIMA SİSTEMLERİ 5
2.1. Ses Nedir? 5
2.2. Ses Üretiminin Anatomisi ve Fizyolojisi 6
2.3. Seslerin Sınıflandırılması 7
2.4. Konuşma ve Konuşma Üretim Modeli 8
2.5. Konuşma Tanıma Sistemleri 11
2.5.1. Konuşmacı bağımlı sistemler 12
2.5.2. Konuşmacıdan bağımsız sistemler 12
2.5.3. Ayrık kelime tanıma sistemleri 13
2.5.4. Sürekli konuşma tanıma sistemleri 13
2.5.5. Ses tabanlı konuşma tanıma sistemleri 13
2.6. Konuşma Tanıma Sistemlerinin Araç İçerisinde Kullanılması 14
2.7. Konuşma Tanıma Süreci 15
2.7.1. Konuşma işaretinin küçük parçalara ayrılması 16
2.7.2. Konuşma bölgesinin saptanması 17
2.7.2.1. İşaret enerji seviyesi 18

2.7.2.2. Sıfır geçiş sayısı 19
2.7.2.3. Yüksek dereceli spektral saptama 21
2.7.3. Dinamik zaman eğreltimi 22
2.7.4. Özellik çıkartım yöntemleri 23
2.7.4.1. Doğrusal öngörülü kodlama 24
2.7.4.2. Mel-Frekans ölçekli kepstral katsayıların elde edilmesi 27
2.7.5. Euclidean uzaklığı 28

BÖLÜM 3. SES İŞARETİNDE GÜRÜLTÜNÜN BASTIRILMASI 30
3.1. Gürültü Nedir? 30
3.2. Uyarlamalı Gürültü Bastırma 32
3.3. Konuşma İşaretinde Gürültünün Uyarlamalı Süzgeç Kullanılarak Bastırılması 33
3.3.1. En küçük ortalama kareler algoritması (Least Mean Squares - LMS) 36
3.3.2. Normalleştirilmiş en küçük ortalama kareler algoritması (Normalized Least Mean Squares - NLMS) 37
3.3.3. Özyinelemeli en küçük kareler algoritması (Recursive Least Mean Squares -RLS) 38
3.3.4. Süzgeç ağırlıklarının yenilenmesinde genetik algoritmaların kullanılması 39
3.4. Gürültü Bastırma İşleminde Yapay Sinir Ağlarının Kullanılması 41
3.4.1. Çok katmanlı ileri beslemeli ağlar 45
3.4.2. YSA öğrenme kuralları 46
3.5. Gürültü Bastırma İşleminde Uyarlamalı Sinir-Bulanık Çıkartım Sisteminin (Adaptive Neuro-Fuzzy Inference System - ANFIS) Kullanılması 47
3.6. Gürültü Bastırma Çalışmaları 50
3.6.1. NLMS uygulaması 53
3.6.2. RLS uygulaması 54
3.6.3. GA uygulaması 55
3.6.4. YSA uygulaması 58
3.6.5. ANFIS uygulaması 59
3.6.6. Algoritmaların karşılaştırılması 60

BÖLÜM 4. ARAÇ İÇİ CİHAZLARIN SESLE KONTROLÜ 65
4.1. Sesle Kontrol İşleminin Yapısı 65
4.2. Veritabanının Oluşturulması 67
4.3. Özellik Çıkartım Yönteminin Belirlenmesi 70
4.4. Konuşma Tanıma Sisteminin Verimi 71
4.5. Konuşma Tanıma Sisteminin Süzgeçlenmiş Gürültülü Konuşma İşaretleri İçin Veriminin Belirlenmesi 74
4.6. Gerçek Zamanlı Konuşma Tanıma Sistemi 76

BÖLÜM 5. SONUÇLAR VE ÖNERİLER 78

KAYNAKLAR 81
EK A: PROGRAM KODLARI 87
ÖZGEÇMİŞ 92

SİMGELER DİZİNİ ve KISALTMALAR

w	: Ağırlık Vektörü
z^{-1}	: Geri kaydırma operatörü
u	: Giriş vektörü
e	: Hata ifadesi
y	: Çıkış vektörü
d	: İstenen çıkış değeri
E_m	: m. pencerenin enerjisi
Z_m	: m. pencerenin sıfır geçiş sayısı
D	: Başlangıçtan bitişe en kısa yolun maliyeti
ŝ(n)	: Kestirilmiş sinyal
c	: Kepstral katsayılar
d(r,t)	: t ve r vektörlerinin birbirlerine olan uzaklıkları
s.	: saniye
I	: Birim matris
λ	: Unutma faktörü
μ	: Adım boyu
YSA	: Yapay Sinir Ağları
DÖK	: Doğrusal Öngörülü Kodlama
ÇKA	: Çok Katmanlı Ağ
FIR	: Finite Impulse Respond (Sonlu Darbe Yanıtı)
IIR	: Infinite Impulse Respond (Sonsuz Darbe Yanıtı)
MSE	: Mean Square Error (Ortalama Karesel Hata)
LMS	: Least Mean Squares (En Küçük Ortalama Kareler)
NLMS	: Normalized Least Mean Squares (Normalleştirilmiş En Küçük Ortalama Kareler)
RLS	: Recursive Lease Squares (Özyinelemeli En Küçük Kareler)
GA	: Genetik Algoritma
ANFIS	: Adaptive Neuro-Fuzzy Inference System (Uyarlamalı Sinir-Bulanık Çıkartım Sistemi)

SNR	: Signal Noise Ration (Sinyal Gürültü Oranı)
GUI	: Graphic User Interface (Kullanıcı Grafik Arayüzü)
MFCC	: Mel-Frequency Cepstral Coefficient (Mel-Freakans Ölçekli Kepstral Katsayılar)
DFT	: Discrete Fourier Transform (Ayrık Zamanlı Fourier Dönüşümü)
DTC	: Discrete Time Cosine Transform (Ayrık Kosinüs Dönüşümü)
CSLU	: (C)enter for (S)poken (L)anguage (U)nderstanding

ŞEKİLLER DİZİNİ

Şekil 2.1: Ses üretiminin Kaynak-Süzgeç Modelinin evreleri 6
Şekil 2.2: Kaynak-Süzgeç Modelin matematiksel gösterimi 7
Şekil 2.3: Konuşma üretiminin Kaynak-Süzgeç Modeli 8
Şekil 2.4: Türkçe 'a' Sesinin 20 ms'lik dalga şekli 9
Şekil 2.5: Türkçe /a/ ve /f/ seslerinin dalga şekli ve frekans spektrumları 10
Şekil 2.6: Genel Konuşma Tanıma Süreci 15
Şekil 2.7: 'SAHTE' kelimesi için (a) zaman-genlik (b) çerçeve-ortalama enerji şekilleri 18
Şekil 2.8: 'SAHTE' kelimesi için (a) zaman-genlik (b) çerçeve-sıfır geçiş sayısı şekilleri 20
Şekil 2.9: Dinamik zaman eğreltimi için (a) maliyeti hesaplanmış noktalar arasında olası geçişler (b) maliyeti en düşük yol 22
Şekil 2.10: Konuşma işaretinin Doğrusal Öngörü Modeli 25
Şekil 2.11: MFCC elde etme süreci 28
Şekil 3.1: 0,5 sn. uzunluğunda (a) beyaz gürültü işareti ve frekans spektrumu (b)renkli gürültü işareti ve frekans spektrumu 31
Şekil 3.2: Gürültü bastırma işleminin yapısı 34
Şekil 3.3: Uyarlamalı süzgeç yapısı ve yapıya ait kavramlar 35
Şekil 3.4: M. derece FIR süzgeç iç yapısı 35
Şekil 3.5: RLS algoritması işaret akış diyagramı 38
Şekil 3.6: Genetik Algoritma akış diyagramı 40
Şekil 3.7: Biyolojik bir beyin sinir hücresi ve bileşenleri 42
Şekil 3.8: Yapay sinir hücresinin genel yapısı 42
Şekil 3.9: Çok katmanlı ileri beslemeli ağ modeli 45
Şekil 3.10: ANFIS ağının yapısı 49
Şekil 3.11: F1'e ait; (a) temiz konuşma işareti, (b) +5dB, 0dB, -5dB beyaz gürültü bindirilmiş konuşma işaretleri, (c) +5dB, 0dB, -5dB renkli gürültü bindirilmiş konuşma işaretleri 52

Şekil 3.12: Gürültü bastırma işlemi için hazırlanan GUI 53
Şekil 3.13: Beyaz gürültü için NLMS ve RLS algoritmalarının karşılaştırılması 54
Şekil 3.14: Renkli gürültü için NLMS ve RLS algoritmalarının karşılaştırılması 55
Şekil 3.15: NLMS ve RLS algoritmalarının ortalama işlem sürelerinin karşılaştırılması 55
Şekil 3.16: Beyaz gürültü için farklı nesil sayılarında GA verimlerinin karşılaştırılması 56
Şekil 3.17: Renkli gürültü için farklı nesil sayılarında GA verimlerinin karşılaştırılması 57
Şekil 3.18: Farklı nesil sayıları için GA ortalama işlem sürelerinin karşılaştırılması 57
Şekil 3.19: Bir girişli, gizli katmanda on işlem birimi ve bir çıkışa sahip ÇKA yapısı 58
Şekil 3.20: Bir girişli, on kurallı 'Sugeno Bulanık' modeli eşdeğer ANFIS yapısı ... 59
Şekil 3.21: Beyaz gürültü için NLMS, RLS, GA, ÇKA ve ANFIS algoritmalarının karşılaştırılması 61
Şekil 3.22: Renkli gürültü için NLMS, RLS, GA, ÇKA ve ANFIS algoritmalarının karşılaştırılması 61
Şekil 3.23: NLMS, RLS, GA, ÇKA ve ANFIS algoritmalarının ortalama işlem sürelerinin karşılaştırılması 62
Şekil 4.1: Konuşma tanıma işleminin blokları 65
Şekil 4.2: Komut listesinin ağaç yapısı 68
Şekil 4.3: Gerçek zamanlı uygulama için hazırlanan GUI penceresi 77

TABLOLAR DİZİNİ

Tablo 3.1: Biyolojik sinir sistemi bileşenlerinin yapay sinir ağında karşılıkları 43
Tablo 3.2: GA başlangıç parametreleri ve işlemleri 56
Tablo 3.3: ANFIS yapısının özellikleri 60
Tablo 3.4: NLMS, RLS, GA, ÇKA ve ANFIS algoritmalarının MSE sonuçları 63
Tablo 3.5: NLMS, RLS, GA, ÇKA ve ANFIS algoritmalarının işlem süreleri (saniye) 64
Tablo 4.1: Sistemde kullanılan komutların listesi ve komutların işlevleri 69
Tablo 4.2: Yöntemlerin işlem süreleri (saniye) 71
Tablo 4.3: Sistem verimi ölçüm sonuçları (%) 73
Tablo 4.4: Sistemin gürültüden temizlenmiş konuşma işaretleri için verimi (%) 75
Tablo 4.5: Data Acquisition Toolbox özellikleri 76

BÖLÜM 1. GİRİŞ

Dil, sözcükler aracılığıyla fikir alışverişi, yani iletişimdir. Her şey gibi dil de yaşamın ihtiyaçlarından ve zorunluluklarından, topluluk üyeleri arasındaki iletişim ihtiyacından doğmuştur. Teknolojinin gelişmesi, üretilen sistemlerin insanlar tarafından kullanımını kolaylaştırma ihtiyacı, insan-makine iletişiminde ses işaretinin kullanılmasını kaçınılmaz hale getirmiştir.

Konuşma tanıma sistemleri birçok alanda günlük yaşamımıza yavaş yavaş girmeye başlamışlardır ve gelecekte pek çok alanda da yaşamımızın bir parçası olmaları kaçınılmazdır. Konuşma tanıma alanındaki çalışmalardan bazıları; Web üzerinde dolaşımı sağlayabilecek, klavyeye dokunmadan e-posta veya metin yazmaya imkan sağlayacak, programları açma kapamayı sağlayabilecek, haber veya bilgi içerikli internet sayfalarının görme özürlü insanlar tarafından takip edilebilmesine imkan sağlayabilecek uygulamaları içeren insan-bilgisayar iletişimi (Frui, 2000; Thong et al, 2000; Davis, 2001), konuşma tanıma sistemlerinin eğitim alanında kullanılması (Wald, 2004), robotların sesle kontrolü (Li and Okamura, 2003), çocukların dil gelişiminde otomatik dil tanıma sistemlerinin kullanılması (Russel et al, 1996), hayatı kolaylaştırmayı amaçlayan ev içerisinde ki sistemlerin veya verimi arttırmayı amaçlayan çalışma alanında ki sistemlerin sesle kontrolü ve tez çalışmamızda amaçlanan araç içerisindeki cihazların sesle kontrolü olarak gösterilebilir.

Konuşma tanıma alanındaki çalışmalar 1960'larda başlamış ve 1980'lerde yüksek hızlı işaret işleyicilerinin üretilmesiyle hız kazanarak birçok sistemin içerisine girmiştir. 50 yıla yakın süredir devam eden çalışmalar henüz insan-makine iletişimini ses işareti kullanarak gerçekleştirme hedefine tam anlamıyla ulaştıramamıştır. Bunun nedenleri olarak; ses işaretinin çok dinamik bir yapıya sahip olması, özelliklerinin kişiden kişiye farklılıklar göstermesi, konuşma tanıma sürecinde insan doğasında geçirilen evrelerle ilgili daha fazla bilgiye ihtiyaç duyulması sıralanabilir. Konuşma tanıma işlemi çok zor bir görev olmasına rağmen, bugün geldiğimiz noktada

konuşma tanıma sistemleri pek çok uygulamanın bazı kademelerinde kullanılabilmektedir.

1.1. Tez Çalışmasının Amacı

Tez çalışmamızda araç içindeki cihazların sesle kontrolüne yönelik konuşma tanıma sisteminin gerçek zamanlı uygulaması ve konuşma tanıma veriminde belirleyici bir etken olan araç içerisindeki motor gürültüsünün bastırılmasında en iyi yöntemin belirlenmesi amaçlanmaktadır.

1980'lerde yüksek hızlı işaret işleyicilerinin üretilmesiyle birlikte konuşma tanıma sistemlerinin araç içerisinde kullanılmasına yönelik çalışmalar başlamıştır. Bu alanda ticari ve akademik olarak birçok çalışmalar yapılmıştır ve yapılmaya devam etmektedir (Forlines et al, 2005; Cavedon et al, 2004; Weng et al, 2005; Hansen et al, 2001; Watika, 2004; Yapanel et al, 2002; Westphal and Waibel, 1999). Günümüzde birçok araba üreticisi kendi ürettikleri araçlarda konuşma tanıma, sesle yönlendirme, komut ve kontrol sistemlerinin kullanılacağını ifade etmişlerdir. Fakat bu konu ile ilgili güvenilirlik ve maliyet üzerindeki teknolojik çalışmalar sonuçlandırılamamıştır ve hala devam etmektedir (Heisterkamp, 2001).

Konuşma tanıma sistemlerinin araç içerisinde kullanılması ile sürücünün hareket halinde dikkatini dağıtabilecek; ses sistemi, pencere ve klima ayarları gibi araç içi bileşenlerin kontrolü ses ile yapılarak sürücünün dikkatini tamamıyla yola vermesi; dolayısıyla kazaların azaltılması ve araç içi konforun arttırılması amaçlanmaktadır. Araç içinde sesle kontrol konusundaki çalışmalar uzun süredir devam etmesine karşın konuşma tanımadaki zorluklar çerçevesinde ve otomobil içerisindeki gürültünün konuşma tanıma işlemini daha da zorlaştırması gerçekleştirilen sistemlerin amaçlanan düzeyde başarıya ulaşmasını engellemiştir. Tez çalışmamızın temel amacı olan araç içindeki cihazların sesle kontrolüne yönelik konuşma tanıma sisteminin gerçek zamanlı uygulaması işleminde araç içindeki gürültünün bastırılması işlemi başlı başına bir konu olarak karşımıza çıkmaktadır. Bu amaçla

gürültü bastırma algoritmaları incelenmiş ve gerçek zamanlı uygulamaya yönelik en iyi yöntem belirlenmeye çalışılmıştır.

1.2. Tez Çalışmasını Oluşturan Bölümler

Bölüm 1'de tez çalışmasının amacı ve tez çalışmasını oluşturan bölümler hakkında bilgi verilmiştir.

Bölüm 2'de konuşma ve konuşmacı tanıma sistemleri hakkında genel bilgiler verilmiştir. Konuşma üretiminin fizyolojik yapısı, bu yapıya göre geliştirilmiş Kaynak-Süzgeç Modeli, seslerin sınıflandırılması anlatılmıştır. Daha sonra konuşma tanıma sistemleri hakkında genel bilgi verilerek konuşma tanıma sistemlerinin alt sınıfları açıklanmıştır. Konuşma tanıma sistemlerinin geliştirilmesi aşamasında konuşma veri tabanlarının oluşturulmasından ve mevcut veritabanlarından söz edilmiştir. Bölüm 2'nin sonunda tez çalışmasında gerçekleştirilen gerçek zamanlı konuşma tanıma sürecinde gerçekleştirilen işlemler; konuşma bölgesi saptama, Dinamik Zaman Eğreltimi (Dynamic Time Warping-DTW), özellik çıkartım yöntemleri, özellik vektörlerinin birbirlerinden uzaklıklarını belirlemek için kullanılan Euclidean uzaklığı sırasıyla açıklanmıştır.

Bölüm 3'de araç içerisindeki gürültünün bastırılması için en iyi yöntemin belirlenmesi amacıyla ses işaretinde gürültünün bastırılması işlemi açıklanmıştır. Gürültünün tanımı, gürültü çeşitleri ve gürültü hakkındaki diğer kavramlar açıklandıktan sonra uyarlamalı gürültü bastırma işlemi verilmiştir. Uyarlamalı gürültü bastırma yöntemlerinden En Küçük Ortalama Kareler algoritması (LMS), Normalleştirilmiş En Küçük Kareler algoritması (NLMS), Özyinelemeli En Küçük Kareler algoritması ile birlikte Genetik algoritmalar (GA), Yapay Sinir Ağları (YSA) ve Uyarlamalı Sinir-Bulanık Çıkarım Sistemi (ANFIS) gibi yapay zeka uygulaması ile hesaplama yöntemleri açıklanmıştır. Konuşma verisi olarak Center for Spoken Language Understanding (CSLU) Konuşmacı Doğrulama Versiyon 1.1 veritabanından 5 bayan ve 5 erkek konuşmacının ad-soyad ve yaşadıkları şehri telaffuz ettikleri yaklaşık olarak 2–5 sn. arasında süren cümleler kullanılmıştır. Her

konuşma verisine +5dB, 0dB, -5dB oranında gürültü bindirilerek algoritmaların gürültüyü bastırma verimleri ve gerçek zamanlı uygulanabilme olasılıkları incelenmiştir. Elde edilen sonuçlar karşılaştırılmalı olarak verilmiştir ve yöntemler içerisinde gerçek zamanlı uygulama için kullanılabilecek en iyi yöntem belirlenmiştir.

Bölüm 4'de araç içi cihazların sesle kontrolü işlemi açıklanmıştır. Sesle kontrol işleminin yapısı ve konuşma tanıma sürecinde gerçekleştirilen her bir işlem açıklanmıştır. Sistemde kullanılan kelimelerin belirlenmesi, veritabanın oluşturulması ve kullanılan komutların sistemdeki işlevleri verilmiştir. Özilinti (Autocorrelation), Ortak Değişinti (Covariance) ve Mel-Frekans Ölçekli Kepstral Katsayılar (Mel-Frequency Cepstral Coefficient-MFCC) özellik çıkartım yöntemleri açıklanmıştır ve bu yöntemlerden gerçek zamanlı bir sistemde kullanıma en uygun yöntem belirlenmiştir. Veritabanının oluşturulması aşamasında 10 erkek ve 10 bayan konuşmacıdan sistemde kullanılacak kelimeleri beşer kez telaffuz etmeleri istenmiştir. Bu telaffuzlardan üçü sistemin eğitiminde ve diğer ikisi sistemin veriminin test edilmesi işleminde kullanılmıştır. Sistemin verimi; erkeklerin telaffuzlarıyla eğitilen sistemin erkekler için, bayanların telaffuzlarıyla eğitilen sistemin bayanlar için ve iki grubun telaffuzlarıyla eğitilen sistemin iki grup için testi olmak üzere üç grup için belirlenmiştir. Bölüm 3'de gerçekleştirilen çalışmalar sonucunda belirlenen gerçek zamanlı bir uygulamada kullanıma en uygun süzgeçleme yöntemi ile Bölüm 4'te oluşturulan sistem birleştirilerek; gürültülü konuşma işaretleri için sistemin verimi belirlenmiştir. Bölüm 4'ün sonunda ise gerçek zamanlı konuşma tanıma uygulaması, bu işlem için hazırlanmış GUI penceresi ve kullanılan Matlab Data Acquisition Toolbox yapısının özellikleri verilmiştir.

Bölüm 5 'de tez çalışmasından elde edilen sonuçlar ve öneriler yer almaktadır.

BÖLÜM 2. KONUŞMA VE KONUŞMA TANIMA SİSTEMLERİ

2.1. Ses Nedir?

Ses, bir nesnenin hareket ettirilmesiyle oluşan, kulağımız ile algılayabildiğimiz titreşimler olarak tanımlanabilir. Sesin oluşabilmesi, titreşimlerin hava ile temasına bağlıdır. Atmosferde oluşan ses dalgalarının kulağımız tarafından toplanmasından, beyindeki merkezlerde karakter ve anlam olarak algılanmasına kadar olan süreç işitme olarak adlandırılır. Ses fizyolojisi ile ilgili bazı temel kavramlar aşağıda verilmektedir:

Ses Frekansı: Bir saniyede oluşan titreşimlerin sayısına ses frekansı denir. Frekans birimi Hertz (Hz)'dir. Ses işaretlerinin oluşumu sırasında bir saniyede meydana gelen titreşim sayısı sesin frekansını verir. Saniyede meydana gelen titreşim sayısı arttıkça yani sesin frekansı arttıkça ses tizleşir (incelir). Düşük frekanslı sesler ise pes (kalın) sesleri oluşturur. İnsan kulağı 20 Hz ile 20 KHz arasındaki sesleri işitebilir. Saniyede titreşimi 20 Hz'den az olan seslere subsonik, 20 KHz üzerinde olanlara ise ultrasonik sesler denir.

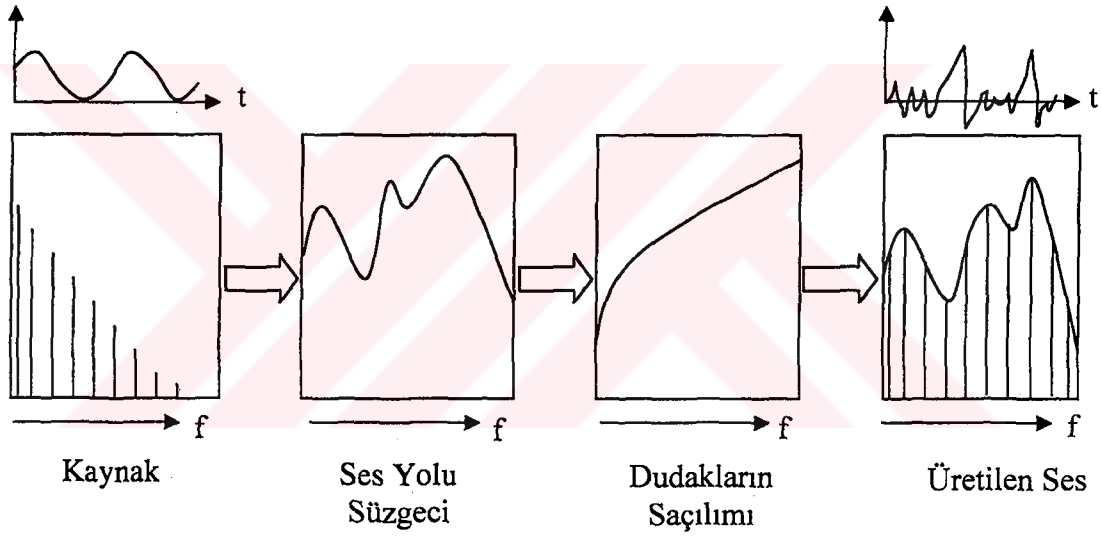
Ses Şiddeti: Sesin kulak tarafından duyulan seviyesine sessin şiddeti denir. Ses şiddeti birimi desibel (dB)'dir. Desibel gerçek anlamda bir birim olmayıp, sessin gücünün insan kulağının duyabileceği en silik sesin gücüne oranının logaritmasıdır. Aşağıda bazı seslerin şiddetleri için örnekler verilmiştir:

Fısıltı sesi	30 dB
Normal Konuşma sesi	60 dB
Kamyon, Çim Biçme Makinesi	90 dB
Rock konseri, Elektrikli Testere	120 dB
Roket Fırlatıcısı sesi	180 dB

Gürültü: İnsan kulağının algılayabilme sınırları içerisindeki her türlü rahatsız edici ses gürültü olarak adlandırılır. Gürültü düzeyi desibel ile verilir.

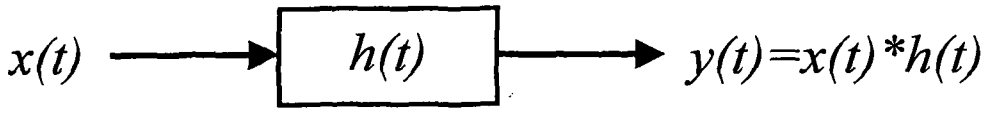
2.2. Ses Üretiminin Anatomisi ve Fizyolojisi

Konuşma, akciğerlerden itilen havanın ses telleri, dil, dişler, dudaklar ve burun gibi eklem noktalarından geçerken farklı şekiller almasıyla üretilir. Hava akciğer tarafından basınçla nefes borusuna ve ses tellerinin başlama noktasına (glottis) itilir. Ses telleri gergin ise bir osilatör gibi titreşir, serbest halde ise titreşmez ve havanın basitçe geçmesine izin verir. Yutak boşluğundan geçen hava akımı, ağız boşluğundan geçerken dil, dişler ve dudaklar tarafından şekillendirilerek konuşma işaretine dönüştürülür. Bu süreç Şekil 2.1'de gösterilmiştir.



Şekil 2.1: Ses üretiminin Kaynak-Süzgeç Modelinin evreleri

Ses Üretimi yaygın bir şekilde iki aşamalı bir süreç olarak tanımlanır (Koreman, 1996). Birinci aşamada ses başlatılır, ikinci aşamada ise süzgeçlenir. Bu evreler arasındaki fark Kaynak-Süzgeç Modelinin kaynağını oluşturur. Modelin temel varsayımı akciğerlerden itilen havanın (ses kaynağının) ses yolunda bulunan ses telleri, dil ve dişler tarafından süzgeçlenmesidir. Sonuçta elde edilen ses dudaklar tarafından saçılımla (radiation) dış ortama gönderilir. Kaynak-Süzgeç Modelin matematiksel gösterimi Şekil 2.2'de verilmiştir. Burada $x(t)$ giriş işaretini, $h(t)$ süzgecin birim darbe yanıtını ve $y(t)$ ikisinin konvalüsyon çarpımından elde edilen çıkışı temsil etmektedir.



Şekil 2.2: Kaynak-Süzgeç Modelin matematiksel gösterimi

Model; kaynak ve süzgecin birbirinden bağımsız olduğunu kabul eder. Son yıllardaki çalışmalar kaynak ve ses yolu arasında bir etkileşim olduğunu göstermesine rağmen Kaynak-Süzgeç Modeli insan sesi tanımlamalarında genel olarak hala kullanılmaktadır.

2.3. Seslerin Sınıflandırılması

Akciğerlerden itilen havanın ses yolu boyunca ses telleri, dil, dişler ve dudaklar gibi farklı eklem noktalarından etkilenmesiyle farklı sesler üretilir. Sesler, ses yolu boyunca etkilendikleri eklem noktalarına göre sınıflandırılırlar:

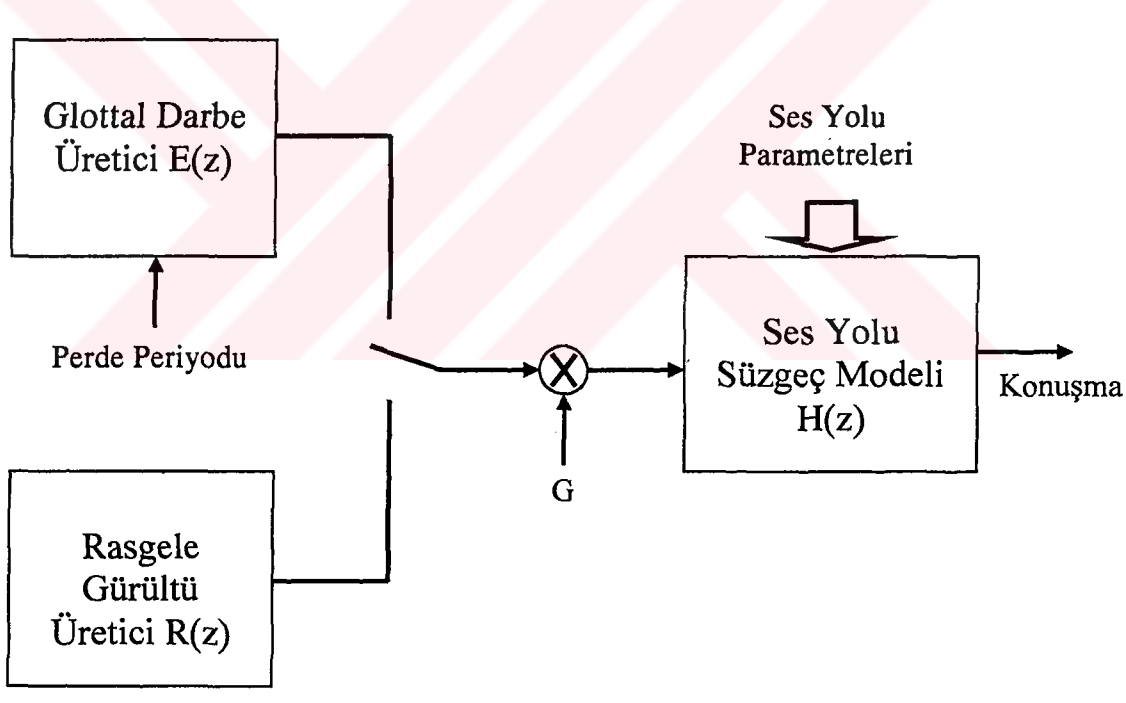
- Ötümlü (Voiced) Sesler: Akciğerlerden itilen hava eğer gergin ses telleri üzerinden geçiyorsa, ses telleri bir osilatör gibi titreşirler ve ötümlü seslerin oluşmasına neden olurlar. Türkçedeki 8 sesli ses (a, e, i, vb.) ve bazı sessiz sesler (b, l, m, vb.) ötümlü seslere örnektir. Sesli sesler oluşurken ses tellerinde meydana gelen ve yaklaşık olarak periyodik olan titreşimin frekansı; ses perdesi (pitch) olarak adlandırılır. Ötümlü sesler için ses yolunun biçimi, tını oluşmasına neden olur; yani ses tellerinin titreşmesiyle meydana gelen sesler birden fazla frekans bileşenine sahiptirler. Ses yolu tını frekansları formants olarak adlandırılırlar.

- Ötümsüz (Unvoiced) Sesler: Akciğerlerden itilen hava eğer serbest ses telleri üzerinden geçiyorsa, ses telleri titreşmez ve sessiz olarak adlandırılan sesler oluşur. Sürtünmeli sesler (s, f, z, vb.) ve nefesli sesler (h) ötümsüz seslere örnektir. Periyodik olmayan ötümsüz sesler üretilirken, sürtünmeli ötümsüz seslerde; ağızda, nefesli ötümsüz seslerde; ses tellerinin başlama noktasında (glottis) türbülanslı hava akımı meydana gelir. Ötümsüz sesleri, konuşma içinde ses olmayan bölgeden ayırt etmek çok zordur. Çünkü düşük enerji seviyelerine ve periyodik olmayan yüksek frekans bileşenlerine sahiptirler.

- Patlamalı (Plosive) Sesler: Akciğerden basınçla itilen hava ses yolunda bazı noktaların daralarak kapanması sonucu sıkıştırılır. Basınçla sıkıştırılmış havanın aniden bırakılmasıyla geçici uyarımlar meydana gelir ve patlamalı sesler oluşur. Patlamalı sesler ses tellerinin uyarımıyla veya ses telleri uyarımı olmadan meydana gelebilir. Ötümlü patlamalı sesler (b, d, g, vb.) ve ötümsüz patlamalı sesler (p, t, k, vb.) patlamalı seslere örnektir.

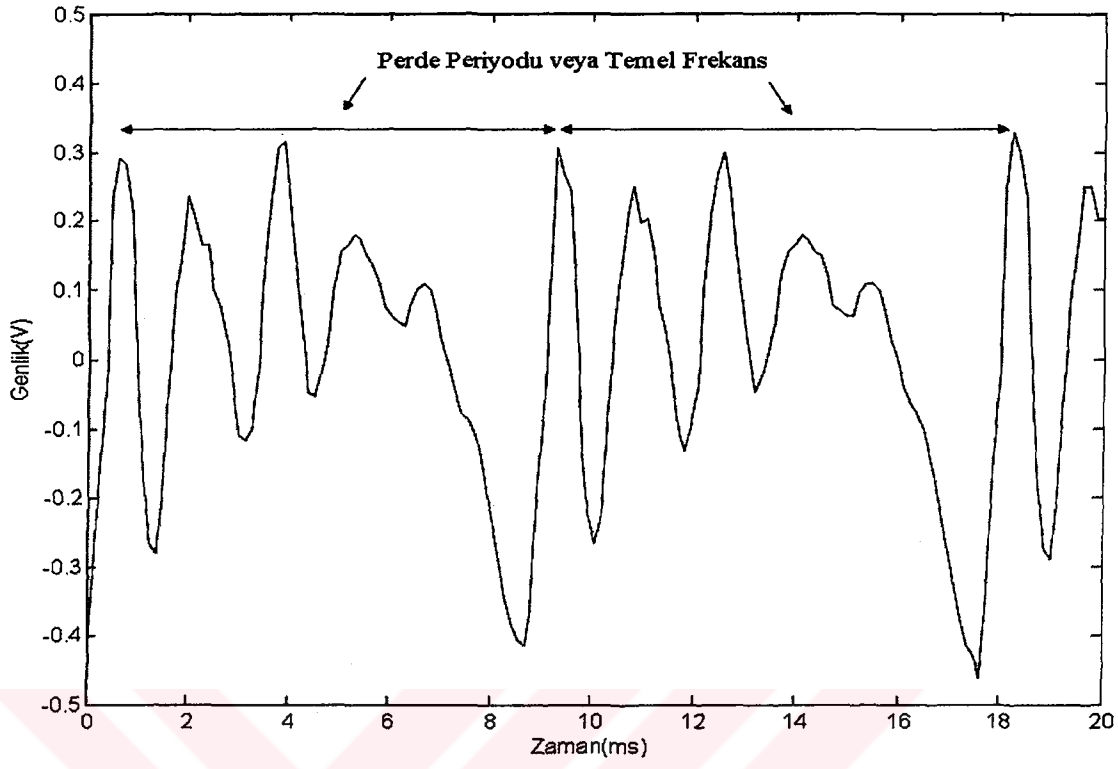
2.4. Konuşma ve Konuşma Üretim Modeli

Ses yolunu ve ses üretim yapısını güçlü bir şekilde temsil eden Kaynak-Süzgeç Modeli Şekil 2.3'te verilmiştir. Bazı seslerin (özellikle sessiz sesler) tam doğru olmasa da, bu model uyarım kaynağının ve ses yolunun yakın bir temsilidir.



Şekil 2.3: Konuşma üretiminin Kaynak-Süzgeç Modeli

Kaynak-Süzgeç modeli bir anahtarla kontrol edilen iki ayrı kaynak farz eder. Birinci kaynak ötümlü seslerin üretimi için kullanılan glottal darbe üreticidir. Ötümlü sesler üretilirken ses tellerinin titreşmesiyle yaklaşık periyodik bir işaret üretilir.

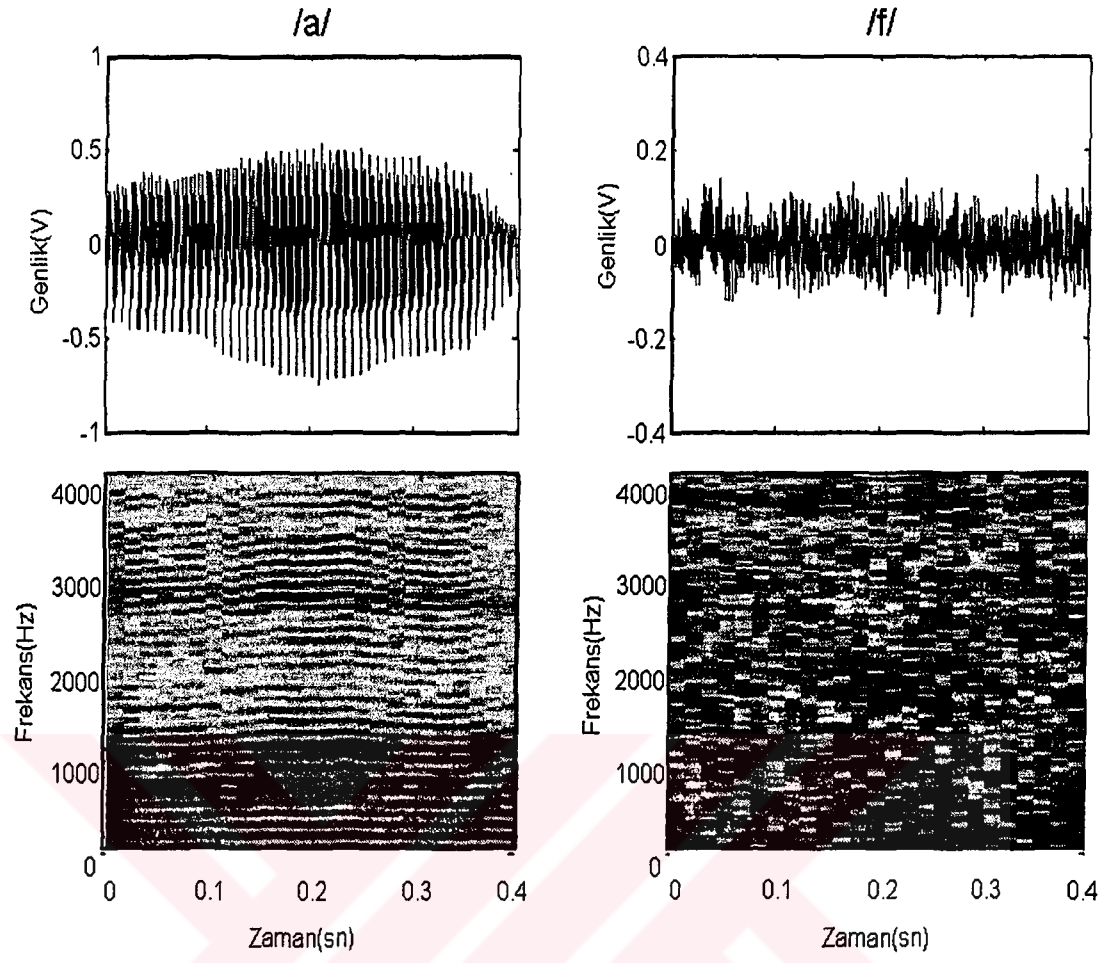


Şekil 2.4: Türkçe 'a' Sesinin 20 ms'lik dalga şekli

Ötümlü sesler için sesin perde periyodu veya temel frekansı olarak adlandırılan darbe eğitim periyodu glottal darbe üretici tarafından üretilir. Şekil 2.4'te verilen 'a' telaffuzunun perde periyodu yaklaşık olarak 9ms ve temel frekansı yaklaşık olarak 112 Hz'dir.

Kaynak-Süzgeç modelinde ikinci kaynak ise gürültü karakteristiği taşıyan ötümsüz seslerin üretimi için rasgele türbülans ve basınçla patlamalı dalga şekli üretmektedir. Ötümsüz seslerin frekans spektrumu, ötümlü seslere göre daha dağınıktır. Bu durum ötümsüz seslerin geniş banda sahip olduğunu göstermektedir. Şekil 2.5'te karşılaştırma amacıyla ötümlü 'a' ile ötümsüz 'f' seslerinin dalga şekilleri ve frekans spektrumları verilmiştir. Ötümlü 'a' sesi 1000 Hz'in altında ve 3000 Hz civarında frekans bileşenlerine sahip iken 'f' sesi ise, tüm ötümsüz seslerin ortak karakteristik özelliği olan, çok dağınık ve geniş bir frekans aralığına sahiptir.

Kaynak-Süzgeç modelinde kaynaktan alınan işaret, insan ses yolunu oluşturan konuşma eklemlerinin dinamik doğasını, matematiksel olarak ifade eden zamanla değişen bir süzgeç ile süzgeçlenir ve çıkışta konuşma işareti elde edilir.



Şekil 2.5: Türkçe /a/ ve /f/ seslerinin dalga şekli ve frekans spektrumları

Ses yolunun şeklinin değişme evresi, modelde ses yolu parametrelerinin değiştirilmesiyle temsil edilmiştir. Ses yolu parametreleri (katsayılar) telaffuzun doğasına bağlı olarak 5 – 20 ms arasında periyotlarla süzgeç ile birleştirilir. Ses yolu, şekli zaman içerisinde değişebilen bir tüpe benzediğinden hava akımının geçişi sırasında formant olarak adlandırılan belirli frekanslarda tını meydana getirmektedir. Ses yolu süzgecinin frekans yanıtındaki tepeler, formant frekanslarını göstermektedirler. Temel frekansı ve formant frekanslarını belirleyen ses yolu uzunluğu ve kalınlığı kişiden kişiye göre değiştiği için temel frekans ve formant frekansları konuşmacı saptama sistemlerinde kullanılırlar.

Ses yolu parametreleri süzgeç modeli ve Kaynak-Süzgeç Modelinin şekli konuşma üretme sistemindeki termal kayıpları veya dudaklardaki saçılma kaybı gibi bazı detayları basitlik için göz ardı etmiştir.

Konuşma esnasında ses yolunun şekli sürekli değiştiğinden, konuşma işareti durağan olmayan, değişken ve çok dinamik yapıya sahip bir işarettir. Ancak konuşma üretimi sırasında ses yolunun kısa zaman aralıkları boyunca sabit kaldığı kabul edilebilir. Böylelikle ses yolu süzgeç modeli için üretilen parametreler bu kısa zaman aralıklarında sabit olarak düşünülür. Konuşma tanıma uygulamalarında, ses yolu, 5-20 ms arasında ki zaman aralıklarında sabit kalan doğrusal bir süzgeç olarak modellenir. Bu model bazı sınırlamaları olmasına rağmen günümüzde kullanılmaya devam etmektedir.

2.5. Konuşma Tanıma Sistemleri

Dil, sözcükler aracılığıyla fikir alışverişi, kısacası iletişimdir. Her şey gibi dil de yaşamın ihtiyaçlarından ve zorunluluklarından, topluluğun üyeleri arasındaki iletişim ihtiyacından doğmuştur. Teknolojinin gelişmesi, üretilen sistemlerin insanlar tarafında kullanımını kolaylaştırma ihtiyacı, insan-makine iletişiminde ses işaretinin kullanılmasını kaçınılmaz hale getirmiştir. Bu yönde çalışmalar 1960'larda başlamış ve 1980'lerde yüksek hızlı işaret işleyicilerinin üretilmesiyle hız kazanarak birçok sistemin içerisine girmiştir. 50 yıla yakın süredir devam eden çalışmalar henüz insan-makine iletişimini ses işareti kullanarak gerçekleştirme hedefine tam anlamıyla ulaştıramamıştır. Bunun nedenleri olarak; ses işaretinin çok dinamik bir yapıya sahip olması, özelliklerinin kişiden kişiye farklılıklar göstermesi, konuşma tanıma sürecinde insan doğasında geçirilen evrelerle ilgili daha fazla bilgiye ihtiyaç duyulması sıralanabilir.

Başarılı bir konuşma tanıma sisteminin aşağıdaki özelliklere sahip olması beklenir(Rabiner, 1990):

- Sistemin kullanımı kolay olmalı, kullanıcıya gerçekten bir yarar sağlayabilmeli.
- Kullanım alanına göre, tatminkar bir tanıma oranına sahip olmalı
- Maliyeti düşük olmalı

İnsan-makine arasında ses ile iletişim, hedeflenenin henüz çok gerisinde olmasına karşın yıllar içerisinde başarıyla çalışan birçok konuşma tanıma yöntemleri geliştirilmiştir. Konuşma tanıma sistemleri üç ölçüte göre değerlendirilir. Birincisi sistemin konuşmacıya bağlı olup olmadığı ölçütüdür. İkinci ölçüt konuşma şeklinin nasıl olduğudur. Bir konuşma tanıma yöntemi; birbirinden ayrıştırılmış sözcükler, birleşik sözcükler veya sürekli konuşma üzerinde tanıma işlemini gerçekleştirebilir. Üçüncü ölçüt ise sistemin tanıyabildiği sözcük sayısıdır. Sistem konuşmacıdan bağımsız hale getirildikçe, ayrık sözcük tanıma işleminden başlayarak devam eden konuşma tanımaya doğru yaklaştıkça ve tanıyabileceği sözcük sayısı arttıkça, sistemi gerçeklemek zorlaşmaktadır.

2.5.1. Konuşmacı bağımlı sistemler

Konuşmacıya bağımlı sistemler daha önceden sisteme tanıtılmış kişiler için konuşma tanıma işlemini gerçekleştirirler. Konuşmacı bağımlı sistemlerin konuşulanı doğru olarak anlayabilmesi için önceden konuşmanın sisteme tanıtılması gerekmektedir. Bu sistemlerin temel aldığı varsayım bir kişinin aynı kelimeyi veya kelime grubunu her defasında birbirini tutan ses ve tempoda söyleyecekleridir. Bu amaçla sisteme dahil edilen konuşmacılardan birkaç kez kayıt alınır ve sistemin eğitimi gerçekleştirilir. Konuşma tanımada veri tabanı oluşturma süreci eğitim olarak adlandırılır. Konuşmacı bağımlı sistemler genellikle sisteme kayıtlı konuşmacı için daha kesin tanıma, sistem dışındaki bir konuşmacı için daha düşük doğrulukta tanıma yaparlar. Bu sistemlerin doğru tanıma yapma verimleri oldukça iyidir.

2.5.2. Konuşmacıdan bağımsız sistemler

Konuşmacıdan bağımsız konuşma tanıma sistemleri, konuşmacıya bağımlı sistemlerin aksine herhangi bir ön eğitime gerek duymaksızın her konuşmacı tarafından kullanılabilirler. Birçok konuşmacıdan alınan konuşma örnekleri için ortak özellikler çıkarılır ve bu özellikler karşılaştırılarak tanıma işlemi gerçekleştirilir. Bu tür sistemler eğitimin mümkün olmadığı uygulamalarda tercih edilirler. Konuşmacı bağımlı sistemlere göre verimleri daha düşüktür.

2.5.3. Ayrık kelime tanıma sistemleri

Ayrık kelime tanıyıcılar konuşma bölgesini saptamak amacıyla telaffuzun her iki tarafında sessizliğe gereksinim duyarlar. Bu bir seferde sadece bir kelime üzerinde tanıma yapılabileceği anlamına gelmemektedir. Çünkü doğal konuşma sırasında kelimeler arasında her zaman duraksama olmamaktadır. Böylelikle arada duraksama olmadan birkaç kelimenin oluşturduğu bir grup için tanıma işlemi yapılabilmektedir. Sistem kelimelerin başını veya sonunu bulmakla uğraşmaz, tüm telaffuzu bir bütün olarak alır. Telaffuz edilen kelime veya kelime grubu modeldeki kelimeler ile karşılaştırılır ve en yüksek yakınlıktaki kelime yada kelimeler saptanır. En basit konuşma tanıma şekillerinden biridir ve yüksek düzeyde işlemci hızına ihtiyaç duyarlar.

2.5.4. Sürekli konuşma tanıma sistemleri

Sürekli konuşma tanıma sistemleri doğal konuşmanın tanınmasını amaçlamaktadır. Bu işlemi gerçekleştirmek gerçekten çok zordur. Doğal konuşma içersinde var olan telaffuz farklılıkları, vurguların koşullara ve konuşmacılara göre farklı noktalarda yapılması, birçok noktada kelimeler arasında duraksama olmadan konuşmanın gerçekleştirilmesi gibi nedenler sürekli konuşma tanımayı zorlaştırmaktadır. Ayrıca bu konuşma tanıma sistemi yüksek miktarda işlemci hızına ve bellek kullanımına ihtiyaç duyar.

2.5.5. Ses tabanlı konuşma tanıma sistemleri

Ses tabanlı konuşma tanıma sistemleri; konuşulan kelimeyi ses olarak adlandırılan çok küçük birimlere böler ve tanıma işlemini bu sesler üzerinden gerçekleştirir. Kelime tanıma sistemleriyle karşılaştırıldığında ses tabanlı tanıma sistemleri çok daha fazla sayıda kelimeyi içeren anlama kapasitelerine sahiptirler. Ses tabanlı konuşma tanıma sistemi gerçekleştirebilmek için dil yapısının çok iyi bir şekilde modellenmesi gerekmektedir.

2.6. Konuşma Tanıma Sistemlerinin Araç İçerisinde Kullanılması

Konuşma tanıma çalışmalarının temel amacını oluşturan insan-makine iletişiminin bir kolu olarak insan-araç iletişimi veya araç içerisindeki cihazların sesle kontrolü gösterilebilir. Günümüzde birçok araba üreticisi kendi ürettikleri araçlarda konuşma tanıma, sesle yönlendirme, komut ve kontrol sistemlerinin kullanılacağını ifade etmişlerdir. Fakat bu konu ile ilgili güvenilirlik ve maliyet üzerindeki teknolojik çalışmalar hala devam etmektedir (Heisterkamp, 2001).

Bu konu üzerindeki çalışmalar uzun süredir devam etmesine karşın yukarıda belirtilen konuşma tanımadaki zorluklar çerçevesinde, otomobil içerisindeki gürültünün konuşma tanıma işlemini daha da zorlaştırması gerçekleştirilen sistemlerin amaçlanan düzeyde başarıya ulaşmasını engellemiş ve bu sistemler ticari olarak piyasaya sürülememiştir.

Konuşma tanıma, gürültü azaltma, konuşmacı tanıma gibi araç içinde gerçekleştirilebilecek uygulamalar amacıyla her dil grubu için farklı veri tabanları oluşturma çalışmaları yapılmaktadır (Wang et al, 2005). Bu çalışmaların bazıları aşağıda sıralanmaktadır:

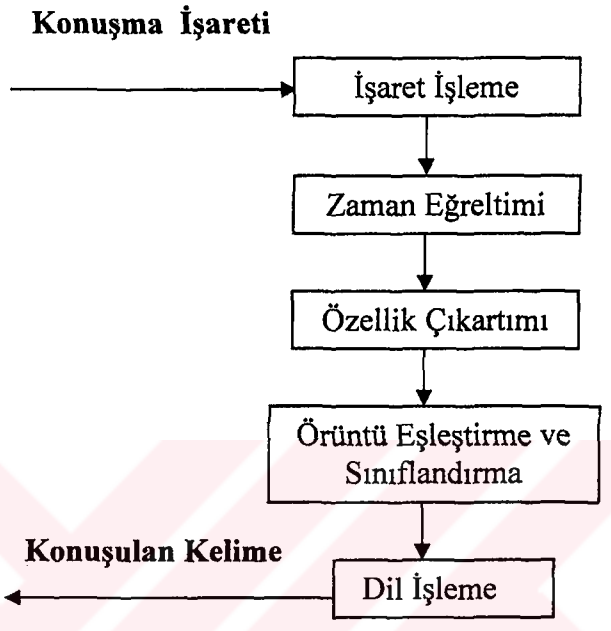
- SpeechDat-Car Veritabanı: İngilizce, Danca (Danimarka dili), Fransızca, Almanca, Fince, Flamanca, İtalyanca, Yunanca ve İspanyolca gibi Avrupa dillerini içermektedir. Bu çalışmanın ortakları arasında BMW, FIAT, Renault, SEAT-Volkswagen gibi araç üretim devlerinin yanı sıra Bosch, Nokia, Alcatel, Sonofon gibi iletişim alanında çalışan şirketler ve birçok üniversite bulunmaktadır (Heuvel et al, 1999).

- MoTiV Veritabanı: Philips, Siemens, Bosch ve Volkswagen şirketlerinin Almanca için ortaklaşa oluşturdukları veritabınıdır.

- CIAIR-HCC Veritabı: Yamaha, Fujitsu, Hitachi, Sanyo gibi şirketlerin ortaklığında Nagoya Üniversitesi tarafından Japonca için oluşturulan bir veritabanıdır.

2.7. Konuşma Tanıma Süreci

Konuşma tanıma süreci kullanılan yönteme göre küçük farklılıklar gösterebilmesine rağmen genel bir konuşma tanıma süreci Şekil 2.6'da gösterildiği şekilde gerçekleşmektedir.



Şekil 2.6: Genel Konuşma Tanıma Süreci

Mikrofon aracılığıyla alınan analog konuşma işareti, bir ön yükselteç ile kuvvetlendirilir ve daha sonra analog-dijital dönüştürücü ile sayısallaştırılarak sistemin girişine verilir. Konuşma tanıma süreci aşağıda belirtilen aşamalar sonucu gerçekleşir:

1. İşaret işleme sürecinin amacı sayısallaştırılmış konuşma işaretini, ortam gürültüsünden, konuşmacıdan kaynaklanan genlik değişimlerinden veya iletim ortamından kaynaklanan bozulmalardan bağımsız olarak üretmektir. Ayrık kelime tanıma sistemleri için gerekli olan kelimenin sessiz bölgeden ayrıştırılarak yalın bir şekilde elde edilmesi bu süreçte gerçekleştirilir.

2. Zaman eğreltimi, işaret işleme süreci sonunda elde edilen sayısal konuşma işareti ile sistemde kayıtlı olan referans işaretinin örnek sayılarının eşitlenmesi sürecidir. Bu süreçte dinamik 'zaman eğreltimi işlemi en çok tercih edilen yöntemdir. Zaman eğreltimi sürecine sınıflandırma sürecinde kullanılacak yönteme göre

gerek duyulmayabilir. Böyle durumlarda sistem işaret işleme sürecinden doğrudan özellik çıkartımı sürecine geçebilir.

3. Özellik çıkartım süreci, ses işaretinin ayırt edici bilgilerini ve özelliklerini içeren sayısal değerler elde etme sürecidir.

4. Örüntü eşleştirme ve sınıflandırma, özellik vektörleri elde edilmiş konuşmanın veri tabanında kayıtlı seslerin özellik vektörleri ile karşılaştırılması ve en yakın eşleşmenin saptanması sürecidir. Bu süreçte kullanılan bazı yöntemler aşağıda belirtilmiştir;

   - Euclidean Uzaklığı
   - Itakura Uzaklığı
   - Vektör Nicemleme
   - Yapay Sinir Ağları
   - Genetik Algoritmalar
   - Bulanık Mantık
   - Bayes Sınıflandırıcı
   - Yapay Ağaç Modeli

5. Dil işleme süreci, kullanılan konuşma tanıma yöntemine veya sınıflandırma yöntemine göre gereksinim duyulan bir süreçtir. Örneğin ses tabanlı konuşma tanıma yöntemlerinde seslerin dil modeline uygun bir şekilde birleştirilmesi işlemi bu süreç içinde gerçekleştirilir.

2.7.1. Konuşma işaretinin küçük parçalara ayrılması

Ses işareti çok dinamik bir yapıya sahiptir. Çok küçük zaman aralıklarında hem genliği hem de sıklığı (frekansı) değişebilmektedir. Ses veya konuşma tanıma işleminin gerçekleştirilebilmesi için ses işaretinin durağan olarak kabul edilebildiği 5–20 ms'lik parçalar şeklinde analiz edilmesi gerekmektedir. Ötümlü sesler bu zaman aralılarında yaklaşık olarak periyodik kabul edilirler. Ses işaretlerinin küçük zaman aralıklarına bölünmesine çerçeveleme denir. 5–20 ms'lik küçük parçalara ayrılan konuşma işaretinin her paçasını sabit katsayılar ile çarpmak pencereleme

olarak adlandırılır. Konuşma tanıma sistemlerinde genellikle ses bilgilerini istatistiksel olarak orta noktada toplamayı amaçlayan Hamming veya Hanning pencere kullanılır. Bunların dışında Bartlett, Barthanning, Blackman, Blacmanharris, Bohmanwin, Chebwin, Gauiswin, Kaiser, Dikdörtgen, Üçgen pencereler kullanılabilir. Aşağıdaki denklemlerde sırasıyla Hamming ve Hanning pencerelerin matematiksel ifadeleri verilmiştir:

$$w(k+1) = 0.54 - 0.46\cos\left(2\pi\frac{k}{n-1}\right), \quad k = 0,\ldots,n\text{-}1 \tag{2.1}$$

$$w(k) = 0.5 - 0.5\cos\left(2\pi\frac{k}{n+1}\right), \quad k = 1,\ldots,n \tag{2.2}$$

Pencere geçişlerinde meydana gelebilecek bilgi kayıplarını önlemek amacıyla her pencere kendinden önceki pencerenin yarısını veya üçte birini kapsayacak bir noktadan başlatılır. Bu işlem örtüşme (overlap) olarak adlandırılır.

2.7.2. Konuşma bölgesinin saptanması

Doğal konuşma sırasında çoğunlukla kelimeler arasında herhangi bir duraksama olmaz. Bu durum doğal dil işlemede kelimelerin başlangıç ve bitiş noktalarını belirmede zorluklara neden olur. Ayrık kelime tanıma sistemlerinde güvenilir ve sağlam bir şekilde kelime tanıma gerçekleştirilebilmesi için kelimenin başlangıç-bitiş noktalarının doğru olarak saptanması çok önemlidir (Lamel et al, 1981). Başlangıç-bitiş noktalarının belirlenmesi için genellikle üç yöntem kullanılmaktadır (Rabiner and Sambur, 1975):

1. İşaret enerji seviyesi
2. Sıfır geçiş sayısı
3. Yüksek dereceli spektral saptama

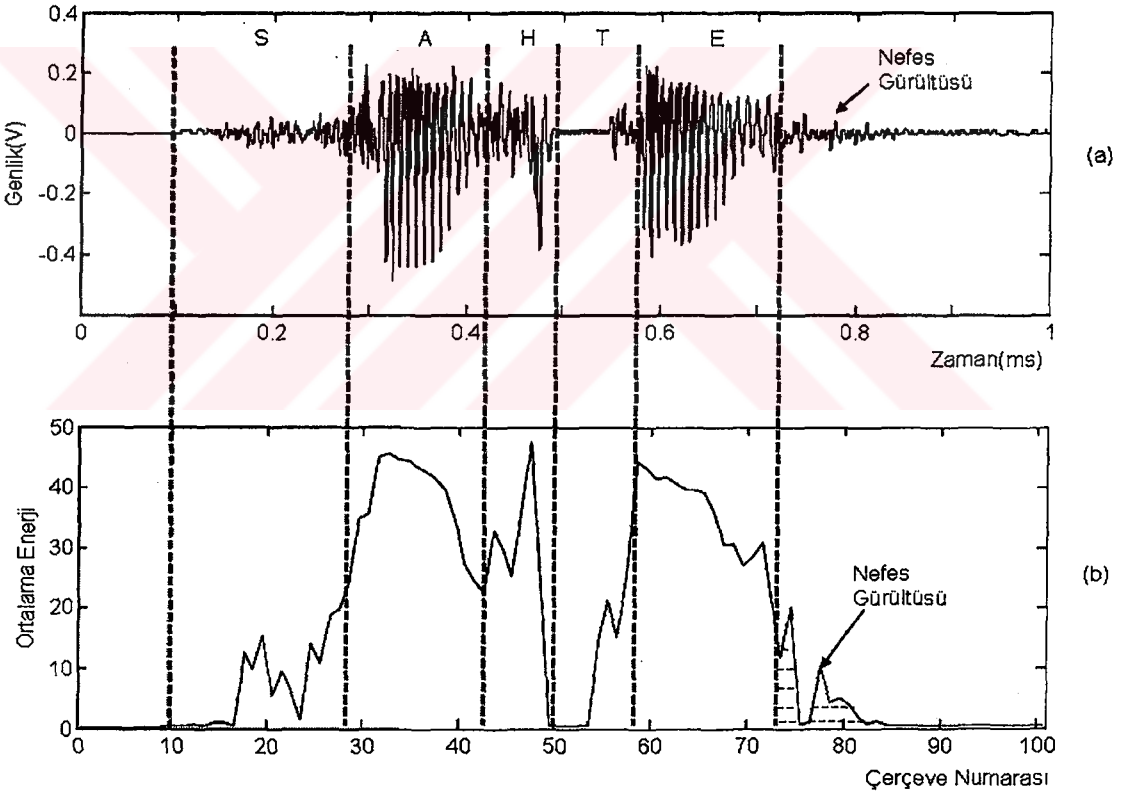
Başlangıç-bitiş noktalarının belirlenmesindeki başlıca zorluklar, başlangıç-bitiş noktalarının işaretteki kısmi değişimlere bağlı olması ve bu değişimlere bağlı olarak yapılacak karar verme hataları, bulunan başlangıç noktalarının otomatik olarak düzeltilmesinin zorluğu ve gürültü olarak sıralanabilir (Türk ve diğ., 1999).

2.7.2.1. İşaret enerji seviyesi

Konuşmanın sessizlikten ayrıştırılması ve kelimenin başlangıç-bitiş noktalarının belirlenmesinde kullanılan ses işaretinin en ayırt edici özelliği enerjisidir. Ses işaretinin genliği zamana bağlı olarak değişmektedir. Pencereleme sonucunda kısa zamanlı parçalara dönüştürülen ses işareti için m. pencerenin ortalama enerjisi Eşitlik (2.3) kullanılarak hesaplanır.

$$E_m = \frac{1}{N}\sum_{n=0}^{N-1}|x_n|^2 , \qquad (2.3)$$

Burada N bir penceredeki örnek sayısını ifade etmektedir.



Şekil 2.7: 'SAHTE' kelimesi için (a) zaman-genlik (b) çerçeve-ortalama enerji şekilleri

Şekil 2.7'de 'sahte' kelimesinin (a)'da dalga şekli, (b)'de 20 ms'lik pencereleri için ortalama enerjileri verilmiştir. Şekilde görüldüğü gibi konuşmanın olmadığı bölümlerde enerjinin sıfır veya ortamdaki gürültüden kaynaklanan sıfıra yakın çok

küçük bir değer olduğu, konuşmanın başladığı noktadan itibaren enerji seviyesinin yükseldiği ve konuşmanın bitmesiyle birlikte enerji seviyesinin tekrar düştüğü görülmektedir.

Ses işaretlerinde bitiş noktasını belirlemek başlangıç noktasını belirlemekten daha zordur. Çünkü konuşma sonlanmasına rağmen akciğerlerden itilen hava dış ortama çıkmaya devam etmektedir ve nefes gürültüsü olarak adlandırılan belli bir enerjiye sahip bu hava basıncı mikrofonun titreşerek ses olarak algılanmasına neden olmaktadır. Konuşma işaretinde ötümsüz sesler, ötümlü seslere oranla oldukça düşük enerji ortalamalarına sahiptirler. Şekil 2.7'de içinde ötümsüz 's' harfi, patlamalı ötümsüz 't' harfi, nefesli ötümsüz 'h' harfi, ünlü ötümlü 'a' ve 'e' harflerinden oluşmuş 'sahte' kelimesinin ortalama enerji şeklinden ötümlü ve ötümsüz sesler arasındaki enerji seviyeleri farkı açıkça görülmektedir. Ötümsüz seslerin düşük enerji seviyelerine sahip olmalarından dolayı sadece ortalama enerji seviyelerine bakarak ötümsüz seslerle başlayan veya biten kelimelerin başlangıç-bitiş noktalarını belirlemek oldukça zordur.

2.7.2.2. Sıfır geçiş sayısı

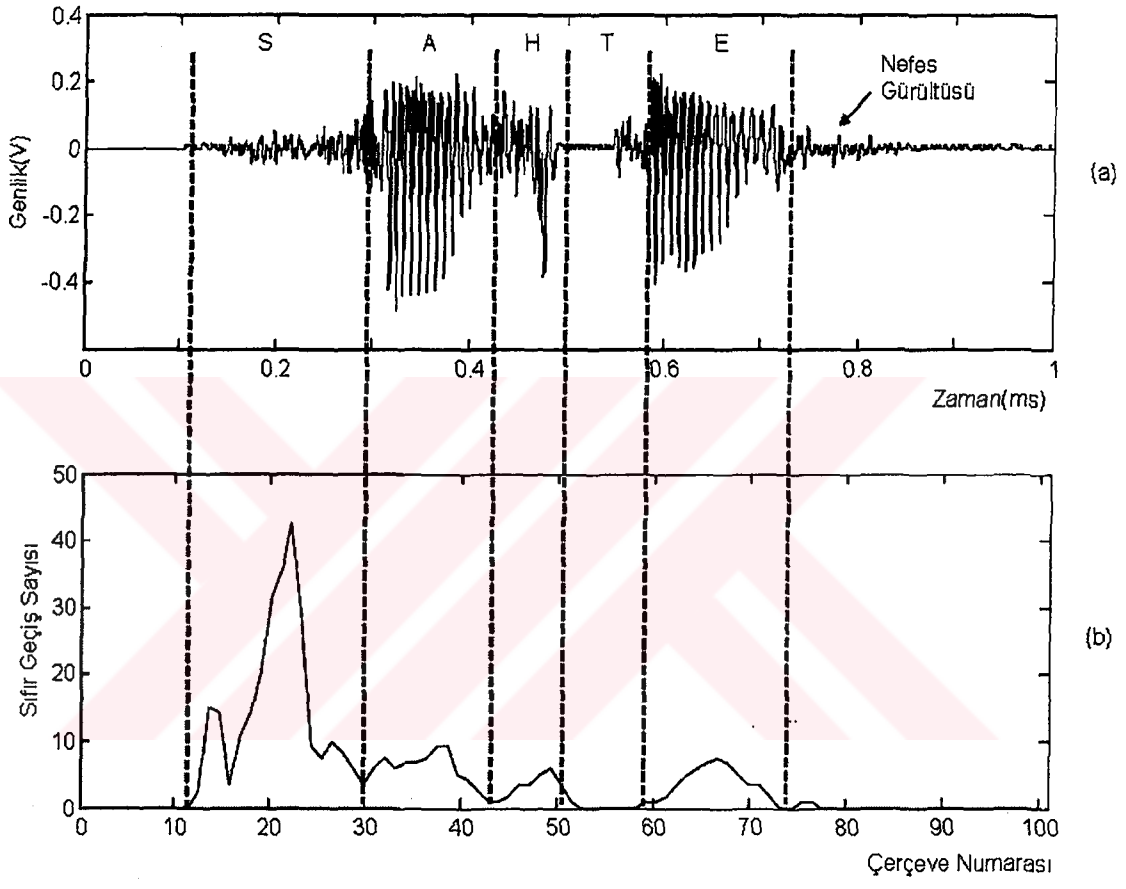
Kelimenin başlangıç ve bitim noktalarını saptamada kullanılan diğer yöntem sıfır geçiş sayısıdır. Sıfır geçişlerinin sayısı, işaret sıklığının basit bir ölçüsüdür. Ayrık zamanlı işaretlerde art arda gelen iki örneğin işaretlerine bakılarak sıfır geçiş sayısı hesaplanır. Pencereleme sonucunda kısa zamanlı parçalara dönüştürülen ses işareti için m. pencerenin sıfır geçiş sayısı Eşitlik (2.4) kullanılarak hesaplanır.

$$Z_m = \sum_{n=0}^{N-1} \left| \mathrm{sgn}(x_n) - \mathrm{sgn}(x_{n+1}) \right| ,$$

$$\mathrm{sgn}(x_n) = \begin{cases} 1 & x_n \geq 0 \\ -1 & x_n < 0 \end{cases}, \qquad (2.4)$$

Burada N bir penceredeki örnek sayısını ifade etmektedir.

Konuşma işaretinin insan kulağı tarafından ideal olarak anlaşılabilmesi için 300 Hz ile 3,2 KHz arası yeterlidir. 1984 yılında Bell laboratuarlarında yapılan çalışmalar telefon hatlarında o güne dek en ideal olarak 3,3 KHz kabul edilen üst bant taşıma sınırının, aslında 3,2 KHz olduğunu göstermiştir (Rodman, 2003). Alt sınır ise 280 veya 300 Hz'dir.



Şekil 2.8: 'SAHTE' kelimesi için (a) zaman-genlik (b) çerçeve-sıfır geçiş sayısı şekilleri

Ötümlü sesler konuşma tanımada ve kelimenin anlaşılabilirliği konusunda ötümsüz seslere oranla çok daha fazla belirleyicidirler. Ötümsüz sesler yüksek sıfır geçiş sayısına, yüksek sıklığa veya yüksek frekans bileşenlerine, sahiptirler fakat konuşma iletiminde veya konuşma tanıma sistemlerinde bu yüksek frekans bileşenleri göz ardı edilebilirler. Sürtünmeli ötümsüz sesler yüksek sıklığa veya diğer bir söyleyişle yüksek frekans bileşenlerine sahip seslerdir.

Periyodik olmayan ve düşük enerji seviyelerine sahip sürtünmeli ötümsüz sesler gürültü karakteristiği göstermektedir. Bu durum kelime sınırlarını belirlemeyi zorlaştırmaktadırlar. Ötümsüz sesler (özellikle sürtünmeli ötümsüz sesler) ile başlayan veya biten kelimelerin sınırlarını belirlemede sıfır geçiş sayısı önemli bir göstergedir.

Şekil 2.7 ve Şekil 2.8'de örnek olarak verilen 'sahte' kelimesinden de görülebileceği gibi sürtünmeli ötümsüz sesler ile başlayan kelimelerin başlangıcını veya sonunu enerji seviyelerine bakarak belirlemek zor iken sıfır geçiş sayısına bakarak belirlemek çok daha kolaydır. Şekil 2.7 (b)'de gösterilen kelime sonlarında meydana gelen nefes gürültüsü genellikle düşük sıklığa sahiptir. Bu durumda bitiş noktası saptamada enerji seviyelerine bakılması yanlışlıklara neden olabilirken, bitiş noktası sıfır geçiş sayısına göre saptandığında genellikle daha doğru bir saptama yapılır.

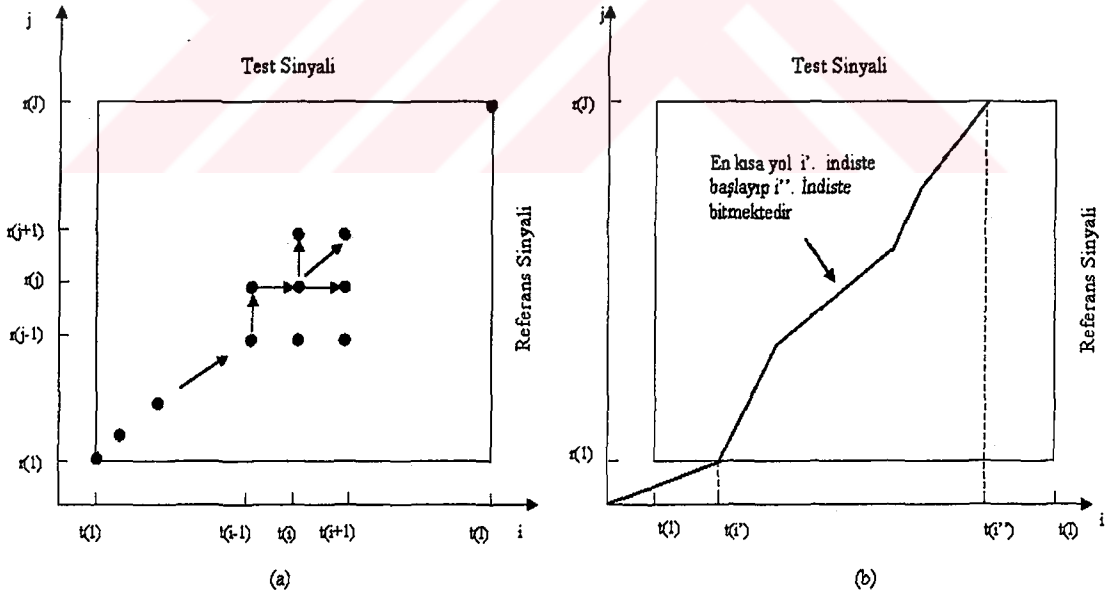
Başlangıç-bitiş noktalarını saptamak için işaret enerji seviyeleri veya sıfır geçiş sayısı özelliklerinden sadece birini kullanmak yerine bu iki özelliği aynı algoritmada kullanmak daha doğrudur.

2.7.2.3. Yüksek dereceli spektral saptama

Yüksek gürültülü ortamlarda kelimelerin başlangıç–bitiş noktalarını belirlemek için işaret enerji sevilerini ve sıfır geçiş hızını kullanmak yeterli değildir. Özellikle sıfır geçiş sayısı çeşitli gürültü tipleri için hassastır (Bing-Fei and Kun-Ching, 2005). Yüksek dereceli spektral saptama başlangıç-bitiş noktalarını konuşma ve gürültü sınıflamada kullanılan bir yöntemdir. Konuşma işaretinin frekans boyutundaki özellikleri incelenerek sınırlar tespit edilmeye çalışılır. Bu yöntem ağır hesaplama yüküne sahiptir ve yöntem için eşiğin önceden belirlenmesi zordur. verimi iyi olmasına rağmen hız faktörünün çok önemli olduğu gerçek zamanlı uygulamalarda tercih edilmemektedir.

2.7.3. Dinamik zaman eğreltimi

Belirli bir kelimenin telaffuzu, kişiden kişiye göre değişebilmektedir. Hatta aynı kişi tarafından farklı zamanlarda telaffuz edilen aynı kelime bile farklılık gösterebilmektedir. Aynı kelimenin seslendirilmesi, bir seslendirmede uzun, bir seslendirmede ise daha kısa zamanda gerçekleştirilebilir. Aynı zamanda, ses işaretinde kimi sesler daha uzun, kimileri ise daha kısa yer almaktadır (Yaldır ve Baygın, 2006). Dinamik zaman eğreltimi yöntemi farklı konuşma hızlarıyla uğraşan konuşma tanıma uygulamalarında kullanılır. Bu yöntem genellikle iki ses işaretinin en uygun bir biçimde aynı boyuta getirilmesini sağlar. Ses işaretleri zaman içerisinde daraltılarak veya genişletilerek aynı boyuta getirilirler. Dinamik zaman eğreltimi basit bir algoritmaya ve en az donanıma gereksinim duyan uygulamalarda başarıyla uygulanmaktadır (Deller et al, 1993). Ses işareti zamanla genliği ve sıklığı değişen bir işarettir. Dinamik zaman eğreltimi sayısallaştırılmış iki ses işaretinin örneklerinin birbirlerine yakınlığı kontrol edilerek en uygun eşleştirmeyi yapar.



Şekil 2.9: Dinamik zaman eğreltimi için (a) maliyeti hesaplanmış noktalar arasında olası geçişler (b) maliyeti en düşük yol

Şekil 2.9'da gösterilen referans işareti J adet, test işareti ise I adet örnekten oluşmaktadır. Bu iki işaretin aynı boyuta getirilmesi işleminde öncelikle her noktanın

birbirlerine benzerliği hesaplanır. Bu işlem maliyet belirleme olarak adlandırılır. Her bir noktanın maliyeti belirlendikten sonra her iki işaretinde ilk noktası olan (1,1) başlangıç noktasından son noktaya doğru en düşük maliyetli yol saptanır. Noktalar arasında geçiş belli kurallar çerçevesinde gerçekleştirilir. Bir noktadan diğer bir noktaya geçiş için üç olasılık vardır. Örneğin D(i-1,j-1) noktasından x-y düzleminde çapraz, D(i,j-1) noktasından x-y düzleminde x yönünde D(i-1,j) noktasından x-y düzleminde y yönünde hareketle D(i,j) noktasına ulaşılabilir. Başlangıç noktasından bitiş noktasına en kısa yolun maliyeti aşağıdaki şekilde hesaplanır.

$$D(i,j) = \min[D(i-1,j-1), D(i,j-1), D(i-1,j)] + d(i,j) \tag{2.5}$$

Burada d(i,j), (i,j) noktasına geline dek oluşan maliyettir.

Referans ve test işaretleri arasındaki minimum yol belirlendikten sonra test işareti referans işaretinden uzun ise daraltılarak, kısa ise genişletilerek her iki işaret aynı boyuta getirilir.

2.7.4. Özellik çıkartım yöntemleri

Özellik çıkartım, sayısal konuşma işaretin parçalarının ayırt edici özelliklerinin tanımlanması ve önemli özniteliklerinin çıkarılıp özellik vektörünün elde edilmesi işlemidir. Ses işareti için özellik çıkartım yöntemleri parametrik ve parametrik olmayan yöntemler olarak iki grupta incelenir. Parametrik olmayan yöntemler akustik-fonetik yaklaşım olarak da bilinir. Bu yaklaşım da sesin frekans, genlik gibi fiziksel özellikleri incelenir ve sesler sınıflandırılır. Parametrik olmayan yaklaşımda kullanılan özellik çıkartım yöntemlerine örnek olarak;

1. Süzgeç bankası
2. Güç spektral analiz verilebilir.
3. Doğrusal öngörülü kodlama

Parametrik yaklaşım, Şekil 2.3'de gösterilen Kaynak-Süzgeç modelini temel alan ve ses işaretinin üretim mekanizmasını tahmin edilmesine yönelik bir yaklaşımdır.

Amaç, glottal darbe üreticiden veya rasgele gürültü üreticiden alınan işaretin son şeklini aldığı ses yolu süzgecinin parametrelerini elde etmektir. Bu parametreler tamamıyla oluşan sesi modellemektedir. Doğrusal Öngörülü (Linear Prediction) parametrik yaklaşımda model oluşturma yöntemlerinin bir alt kümesidir (Furui, 1989).

2.7.4.1. Doğrusal öngörülü kodlama

Doğrusal Öngörülü Kodlama (DÖK-Linear Prediction Coding) konuşma ve konuşmacı tanıma sistemlerinde yaygın olarak kullanılan parametrik yöntemdir. Doğrusal öngörü yönteminin dayandığı temel olan bir sesli ifade örneğinin geçmiş sesli ifade örneklerine bakılarak yaklaşık olarak elde edilebileceği fikrini DÖK en iyi şekilde gerçekleştirir. DÖK incelemesi yapılarak sayısal ses işaretinin küçük parçaları için özellik vektörleri oluşturulur. DÖK yönteminin konuşma ve konuşmacı tanıma sistemlerinde geniş çaplı kullanılmasının nedenleri aşağıdaki gibi sıralanabilir (Rabiner and Juang, 1993):

1. DÖK ses işaretinin iyi bir modelini sağlar. Bu durum konuşma işaretinin durağan kabul edildiği 5-20 ms'lik ötümlü ses kısımları için doğrudur. Konuşmanın ötümsüz kısımları için DÖK ötümlü bölgeler için olandan daha az etkilidir fakat konuşma işaretinin anlaşılabilirliğini daha fazla ötümlü sesler belirlediği için bu yöntem konuşma tanıma amaçları için kullanışlı bir model sağlar.

2. Konuşma işaretinin analizinde makul bir kaynak-ses yolu ayrımı sağlar. Sonuç olarak ses yolu karakteristiklerinin küçük betimlenmelerine olanak sağlar. Bu karakteristikler konuşmanın bileşenlerini oluşturur.

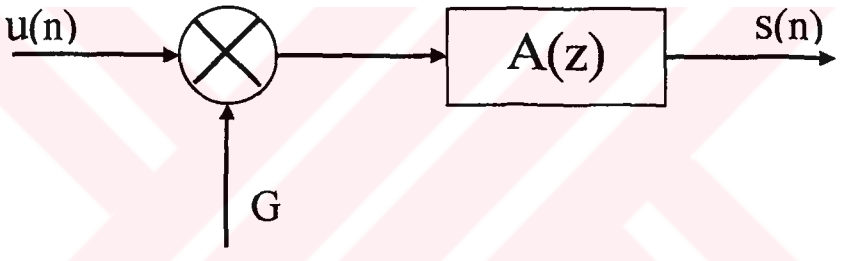
3. DÖK analitik olarak çözümlenebilir bir modeldir. DÖK modeli matematiksel olarak basit ve kesindir. Ayrıca yazılımda ve donanımda uygulaması kolaydır. Örneğin DÖK hesaplamalar; akustik-fonetik süzgeç bankası modelinin gerektirdiği hesaplamalardan oldukça azdır.

4. DÖK modeli konuşma tanıma uygulamaları için iyi sonuç verirler. Çalışmalar DÖK konuşma tanıma verimlerinin süzgeç bankasını temel alan uygulamalara göre daha iyi sonuç verdiğini göstermiştir.

DÖK modeli, n. ayrık zamanda verilmiş s(n) konuşma örneği, yaklaşık olarak geçmiş p konuşma örneğinin doğrusal bir birleşiminden aşağıdaki şekilde elde edilir:

$$s(n) \cong a_1 s(n-1) + a_2 s(n-2) + a_3 s(n-3) + ... + a_p s(n-p) \qquad (2.6)$$

$a_1, a_2, a_3, ..., a_p$ katsayıları ilgili konuşma penceresi için sabit kabul edilir ve Doğrusal Öngörücü (Linear Predictor) katsayılar olarak adlandırılırlar. Şekil 2.10'da konuşma işaretinin Doğrusal Öngörü Modeli verilmiştir.



Şekil 2.10: Konuşma işaretinin Doğrusal Öngörü Modeli

Konuşma işareti s(n)'i üretmek için u(n) uyarım kaynağı, G kazancı ile çarpılır ve tüm-kutup sisteme, $H(z) = \frac{1}{A(z)}$, giriş olarak verilir. Eşitlik (2.6)'ya uyarım terimini, Gu(n), ekleyerek eşitliği aşağıdaki şekle dönüştürebiliriz.

$$s(n) = \sum_{k=1}^{p} a_k s(n-k) + Gu(n) \qquad (2.7)$$

Eşitliği z boyutunda yazarsak aşağıdaki eşitliği elde ederiz.

$$S(z) = \sum_{k=1}^{p} a_k z^{-k} S(z) + GU(z) \qquad (2.8)$$

Eşitlik (2.8)'deki sitemin transfer fonksiyonu aşağıdaki gibidir.

$$H(z) = \frac{S(z)}{GU(z)} = \frac{1}{1 - \sum_{k=1}^{p} a_k z^{-k}} = \frac{1}{A(z)} \tag{2.9}$$

Doğrusal Öngörü analizinde temel problem doğrudan konuşma işaretinden öngörücü (predictor) katsayılar kümesini, $\{a_1, a_2, a_3, ..., a_p\}$, hesaplamaktır. Konuşma işaretinin spektral karakteristikleri zamanla değiştiğinden n. zaman için verilen öngörücü katsayılar, n. nokta civarındaki konuşma işaretinin kısa parçalarından tahmin edilir. Böylelikle n. zamanda tahmin edilen konuşma işaretinin o andaki gerçek konuşma işaretinden farkı minimize edilerek öngörücü katsayılar hesaplanır. Geçmiş konuşma örneklerinden tahmin edilen konuşma işaretini ŝ(n) kabul edersek Eşitlik (2.6) aşağıda verildiği şekilde yeniden ifade edilebilir:

$$\hat{s}(n) = \sum_{k=1}^{p} a_k s(n-k) \tag{2.10}$$

Hata işareti, e(n), aşağıdaki şekilde hesaplanacaktır.

$$e(n) = s(n) - \hat{s}(n) = s(n) - \sum_{k=1}^{p} a_k z^{-k} \tag{2.11}$$

m noktadan oluşan pencerelenmiş tahmini işaret için oluşan toplam hata, hataların karelerinin toplamı alınarak hesaplanır.

$$E = \sum_{m} e_n^2(m) = \sum_{m} \left[s_n(m) - \sum_{k=1}^{p} a_k s_n(m-k) \right]^2 \tag{2.12}$$

En ideal a katsayılarını üretmek için hata işaretinin minimize edilmesi gerekmektedir. Hata işareti a_k'e göre türevi alınıp sıfıra eşitlediğimizde minimum hata değerine ulaşırız.

$$\frac{\partial E}{\partial a_k} = \frac{\sum_m \left[s_n(m) - \sum_{k=1}^{p} a_k s_n(m-k) \right]^2}{\partial a_k} = 0 \qquad (2.13)$$

Eşitlik (2.13) düzenlendiğinde aşağıdaki eşitliği elde ederiz.

$$\sum_m s_n(m-i)s_n(m) = \sum_{k=1}^{p} a_k \sum_m s(m-i)s_n(m-k) \qquad (2.14)$$

Elde edilen bu ifadeden a_k katsayıları hesaplanabilir.

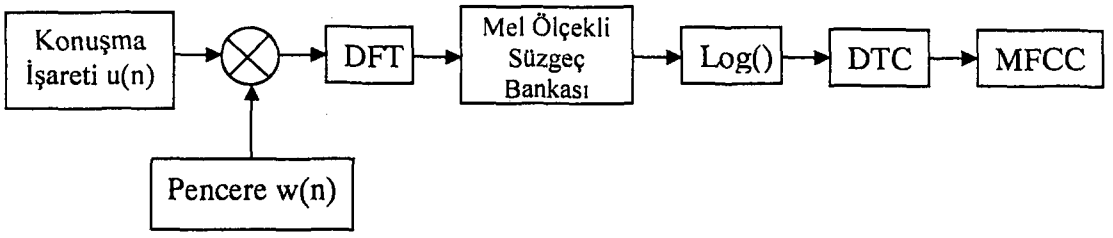
DÖK modelinden türetilen kepstral (cepstral) katsayılar bir diğer özellik çıkartım yöntemi olup, bu alanda uygulanan en baskın yöntemdir. Kepstral katsayıları, işaretin güç spektrumunun logaritmasının ters Fourier dönüşümü şeklinde ifade edilebilir. Bu katsayılar, konuşma işaretinin ses genlik yanıtını ifade eder (Fatihoğlu, 2003). Kepstral katsayılar, DÖK modelindeki a_k parametre takımındaki katsayıların yinelemeli ilişkisinden aşağıdaki gibi elde edilebilir (Farell and Mammone,1994):

$$c_1 = a_1$$
$$c_n = \sum_{i=1}^{n-1} (1 - \frac{i}{n}) a_i c_{n-i} + a_n \qquad (2.15)$$
$$1 < n \leq p$$

Eşitlik (2.15)'te p, a ve c parametre takımı katsayılarının derecesini göstermektedir.

2.7.4.2. Mel-Frekans ölçekli kepstral katsayıların elde edilmesi

1970'lerin başında Oppenheim, Shaffer ve Rabiner konuşma işleme uygulamaları içinde kepstrum tekniklerini tanımladılar ve kullandılar. İnsan kulağının davranışını taklit eden mel-frekans ölçeği temel alan mel-frekans ölçekli kepstral katsayılar (MFCC) pek çok uygulamada başarıyla kullanılmıştır (Koç, 2002).



Şekil 2.11: MFCC elde etme süreci

Şekil 2.11'de MFCC elde etme süreci verilmiştir. Katsayıları elde etme sürecinde aşağıdaki işlemler gerçekleştirilir:

- Konuşma işareti birbiri içine geçmiş tekrarlanan (overlap) çerçeveler şekilde 5-20 ms'lik küçük parçalara ayrılır ve pencerelenir

- Her bir çerçevenin güç spektrumunu bulmak için Ayrık Zamanlı Fourier Dönüşümü (DFT) yapılır.

- Güç spektrumu, denklemi Eşitlik (2.16)'da verilmiş olan 1 KHz altında doğrusal ölçekli ve 1 KHz yukarısında logaritmik ölçekli mel-frekans süzgeç bankasından geçirilir.

$$mel(f) = 2595 * \log 10(1 + \frac{f}{700}) \qquad (2.16)$$

- Süzgeç çıkışındaki elde edilen değerlerin logaritmasının ayrık kosinüs dönüşümü (DCT) alınarak mel-frekans ölçekli kepstral katsayılar (MFCC) hesaplanmış olur.

2.7.5. Euclidean uzaklığı

Kısa zamanlı sayısal konuşma işaretleri için bulunan özellik vektörlerinin veri tabanında bulunan seslerden en fazla hangisine benzediğinin belirlenmesi işleminde kullanılan en basit ve en hızlı yöntem Euclidean uzaklığının hesaplanmasıdır. Yapılan işlem, çok boyutlu uzayda test ve referans işaretlerinin özellik vektörlerinin

birbirlerine uzaklıklarını belirlemektir. Test ve referans vektörlerinin birbirlerinden uzaklıkları d(t,r) ifadesi ile gösterilir ve iki vektör için olası iki durum vardır.

1. $d(r,t) \geq 0,$ (2.17)

2. $d(r,t) = 0,$ sadece r = t durumunda (2.18)

Test ve referans vektörleri arasındaki mesafe ne kadar küçük ise iki telaffuz birbirine o kadar yakındır. Euclidean uzaklığı aşağıdaki şekilde hesaplanır.

$$d(r,t) = \sqrt{\sum_{k=1}^{N} |r_k - t_k|^2} = \sqrt{[r-t]^T [r-t]} \quad (2.19)$$

BÖLÜM 3. SES İŞARETİNDE GÜRÜLTÜNÜN BASTIRILMASI

3.1. Gürültü Nedir?

Sistemin kendisinden veya dış nedenlerden oluşan ve işaretin orijinal yapısında bozulmalara neden olan, istenmeyen işaret bileşenlerine gürültü denir. İnsan kulağının işitme sınırı olan 20 Hz – 20 KHz arasında bir frekansa ve belli bir genliğe sahip rahatsız edici, istenmeyen işaretlere ses gürültüsü denir. Ses gürültüsü göreceli bir kavram olup uçak sesi gürültü olarak kabul edilirken, bazen uyku sırasında sivrisineğin çıkardığı ses de gürültü olabilmektedir. Konuşma işareti için gürültü anlaşılabilirliği etkilediği zaman bir problem olarak değerlendirilmektedir. Ses işaretindeki gürültü bileşenlerinin nedenlerinden bazıları aşağıda sıralanmıştır;

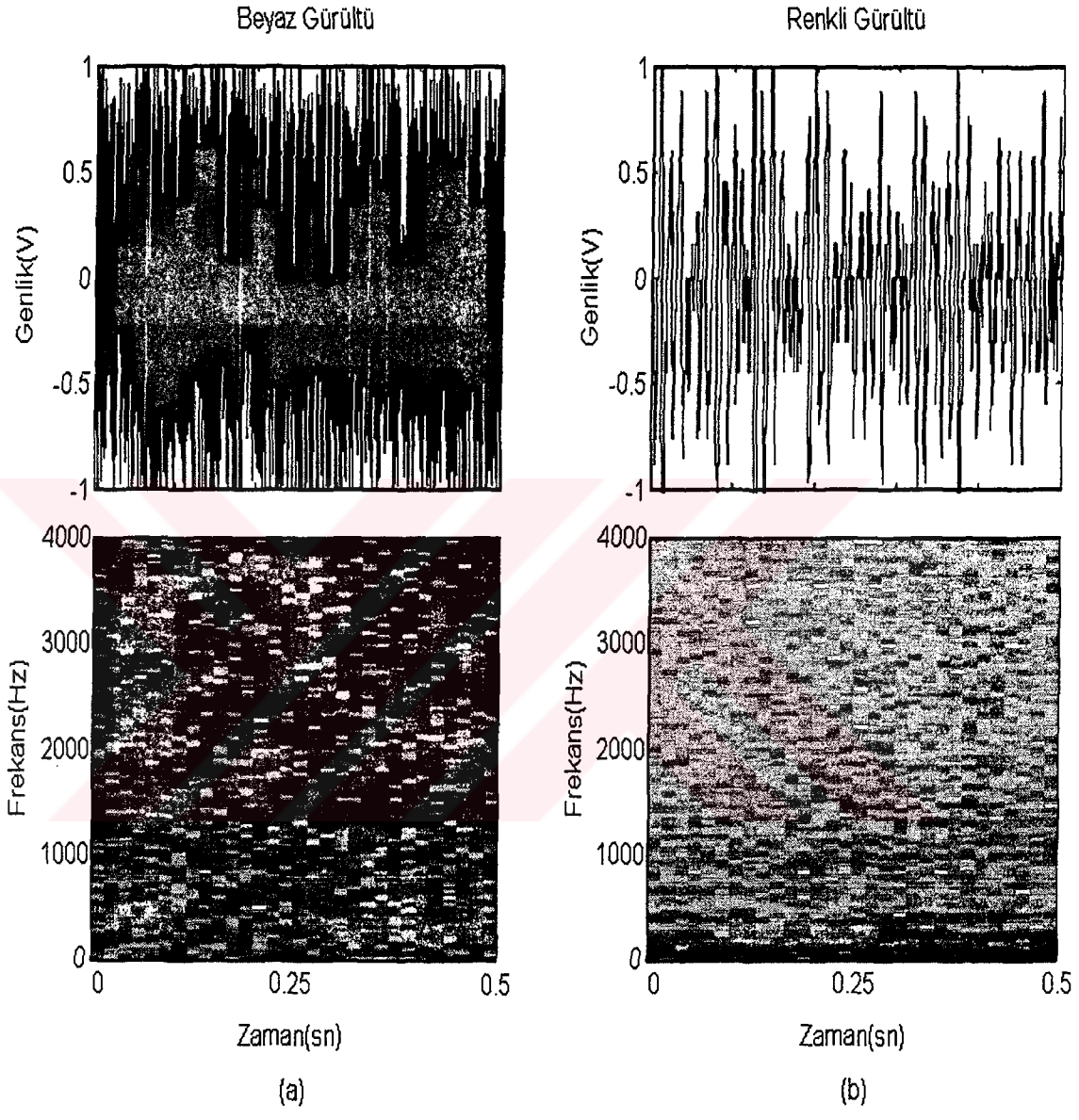
- Kayıtta kullanılan mikrofonun kalitesi
- Kullanılan kayıt cihazının kalitesi
- İletimde meydana gelen bozulmalar
- Ortam gürültüsü

Gürültü işaretleri, frekans spektrumlarına göre ikiye ayrılırlar:

Beyaz Gürültü: işaret gücü tüm frekans bölgesine eşit oranda dağılmış gürültü işaretleridir. Şekil 3.1 (a)'da 0,5 saniyelik beyaz gürültü işaretinin frekans spektrumu verilmiştir. Şekilde de görüldüğü gibi beyaz gürültü işaretinin güç spektrumu, her frekansta yaklaşık aynı değere sahiptir. Doğada mevcut bulunan bütün renklerin karışımı nasıl beyaz ışığı meydana getirirse, bütün frekans aralıklarına sahip sürekli spektrumlu sesler de "Beyaz Gürültü"yü meydana getirir.

Renkli Gürültü: Böyle seslerde birkaç frekans yoğun olarak yer alır. Şekil 3.1 (b)'de BMW 750i marka otomobil içerisinden kaydedilmiş yarım saniyelik renkli gürültü

işaretinin frekans spektrumu verilmiştir. Şekilde de görüldüğü gibi kayıt yapılan araç içindeki gürültü 0–500 Hz arsında frekans bileşenlerine sahiptir. Ortamda bulunan ses gürültülerinin tümü renkli gürültülerdir.



Şekil 3.1: 0,5 sn. uzunluğunda (a) beyaz gürültü işareti ve frekans spektrumu (b)renkli gürültü işareti ve frekans spektrumu

3.2. Uyarlamalı Gürültü Bastırma

Ortamdan veya sistemin kendisinden kaynaklı gürültü tarafından bozulmuş işareti tahmin etmenin genel yöntemi, mevcut gürültüyü bastırmayı amaçlayan ve çıkışta işareti yaklaşık değişmeden elde edebilecek bir süzgeçten geçirmektir (Windrow et al, 1975). Temel olarak süzgeçler sabit süzgeçler ve uyarlamalı süzgeçler olmak üzere iki gruba ayrılır. Sabit süzgeçlerin tasarımı işaret ve gürültü hakkında ön bilgiye ihtiyaç duyar. Konuşma işareti gibi frekansı ve genliği değişken işaretlerde gürültü bastırma amacıyla kullanılmaları mümkün değildir. Diğer taraftan uyarlamalı süzgeçler işaretteki değişimlere göre parametrelerini uyarlayabilme yeteneğine sahiptirler ve tasarımları sırasında işaret ile ilgili ön bilgiye ihtiyaç duymazlar. Uyarlamalı süzgeçler, depolanmış veya gerçek zamanlı işlenen verileri kullanan çeşitli uygulamalar içinde kullanılabilir. Uyarlamalı süzgeç kullanım alanları, ortamdaki değişikliklere otomatik olarak uyum sağlayabilen gürültü bastırma süzgeçleri, bilinmeyen bir sistemin bir sayısal süzgeç olarak modellenerek tanımlanması olabilir (Vinger and Torresen, 2003). İşaretin karakteristiği ile ilgili bilgiye sahip olunamayan ortamlarda ki gerçek zamanlı uygulamalarda uyarlamalı süzgeçleme yeterli bir süzgeçleme verimini mümkün kılar. Uyarlamalı süzgeçleme işlemi için literatürde çok sayıda yinelemeli algoritma çeşitleri geliştirmiştir. Bu algoritmalardan hangisinin tercih edileceği aşağıdaki ölçütlerin biri veya bir kaçı göz önüne alınarak yapılır (Haykin, 2002):

- Yakınsama hızı: algoritma için ihtiyaç duyulan yineleme sayısını tanımlar.

- İz sürebilme: uyarlamalı süzgeç algoritması durağan olmayan bir ortamda işletildiğinde, algoritmanın ortamdaki istatistiksel değişimleri izlemesi istenir.

- Sağlamlık: uyarlamalı süzgeçler için sağlamlık, küçük bozulmalarda sonuçta büyük tahmin hatalarına neden olmamasıdır.

- Hesaplama Gereksinimleri: algoritmanın bir yineleme işlemi içindeki çarpma, bölme, toplama, çıkarma gibi işlemlerin sayısını, verinin ve programın depolanması için hafıza gereksinimlerini kapsar.

- Yapı: algoritma içinde akan bilginin yapısıdır.

Gerçek zamanlı uygulamalarda gürültü bastırmadaki başarı ile birlikte en önemli ölçüt algoritmanın çalışma hızıdır.

Doğrusal bir süzgecinin darbe yanıtı süzgeç hafızası olarak tanımlanır. Bu temele dayanarak doğrusal süzgeçler FIR (Finite-Duration Impulse Respond - Sonlu Darbe Yanıtı) ve IIR (Infinite-Duration Impluse Respond) olarak iki gruba ayılmıştır.

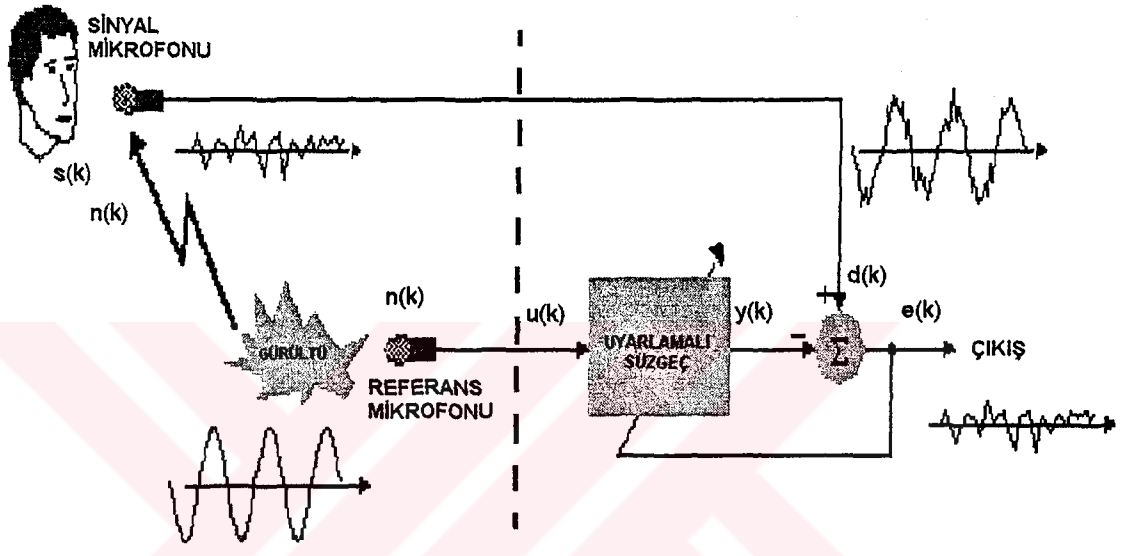
Gürültü bastırma işleminde en önemli kavram, ses işareti gücünün gürültü gücüne oranının logaritması alınarak hesaplanan işaret-gürültü oranı, SNR (Signal Noise Rate), dır. İşaret gürültü oranı aşağıdaki eşitlik kullanılarak hesaplanabilir.

$$SNR = 10\log_{10}\frac{SesSinyalininGücü}{GürültüSinyalininGücü} \tag{3.1}$$

3.3. Konuşma İşaretinde Gürültünün Uyarlamalı Süzgeç Kullanılarak Bastırılması

Konuşma ve konuşmacı tanıma sistemlerinin verimini etkileyen en büyük etkenlerden biri şüphesiz gürültüdür. Gürültü işaretinin şiddeti arttıkça anlaşılabilirlik ve sistemin tanıma verimi düşmektedir. İnsan ile karşılaştırıldığında arka plan gürültüsünün var olduğu ortamlarda modern konuşma cihazlarının verimi önemli derecede düşer (Hussain and Campbell, 1998). Tez çalışmamızda amaçlanan araç içerisindeki sistemlerin ses ile kontrolü işlemini gerçekleyebilmenin temel koşullarından biri araç içerisinde ki gürültüyü etkisiz hale getirmek yani bastırmaktır. Bu amaçla çeşitli çalışmalar yapılmıştır. Öncelikle araç içindeki gürültünün uyarlamalı süzgeçleme yoluyla bastırılmasından söz edilecektir. Şekil 3.2'de gürültü bastırma işleminin yapısı verilmiştir.

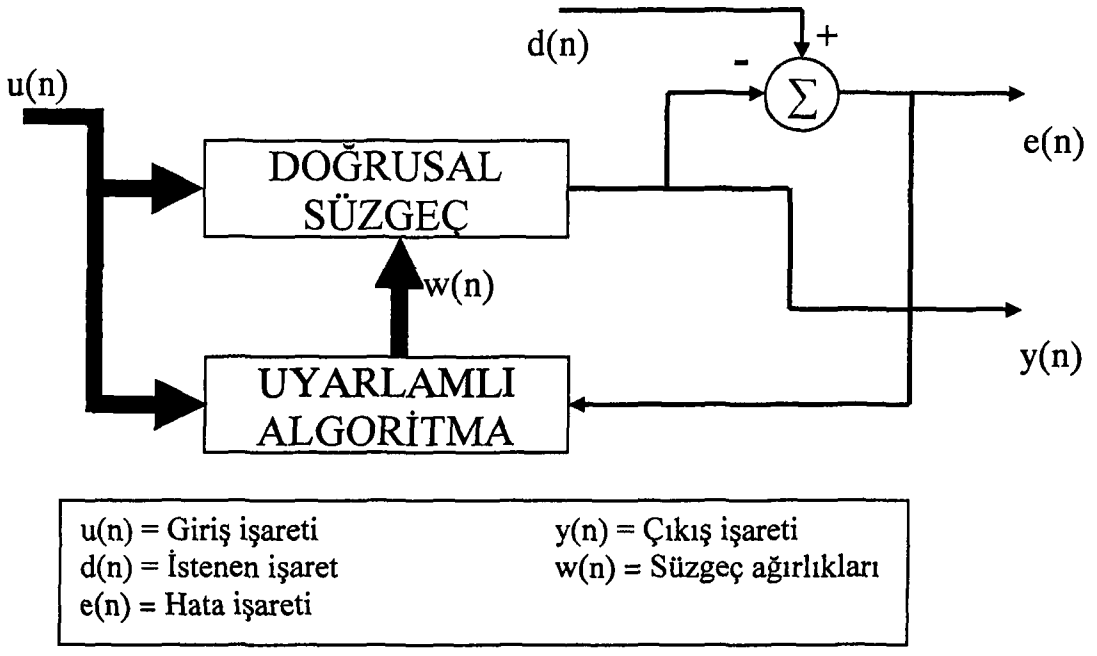
Araç içerisindeki gürültüyü bastırabilmek için sistemde konuşmacının ağzına yakın bir yerden konuşma işareti almak üzere bir ve sadece ortamdaki gürültüyü yani referans işaretini almak üzere konuşmacıdan uzakta bir veya birkaç mikrofon birlikte kullanılır. Referans işareti için kullanılan mikrofon sayısı ne kadar fazla olursa gürültünün karakteristiği o kadar daha iyi belirlenir. Bu durum sistem veriminin daha da iyi olmasını sağlar.



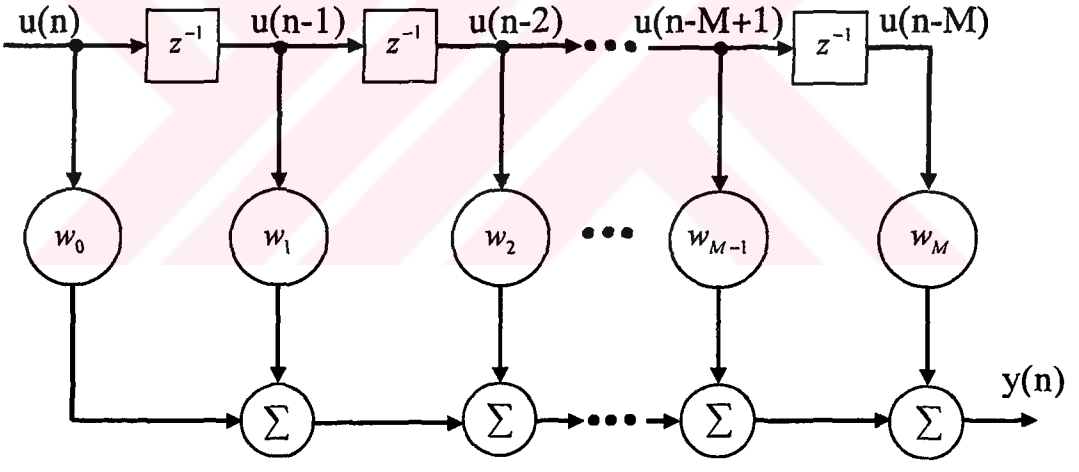
Şekil 3.2: Gürültü bastırma işleminin yapısı

Uyarlamalı süzgeç yapısı ve yapıya ait kavramlar Şekil 3.3'te gösterilmektedir. Gürültülü ortamda kaydedilen konuşma işareti, u(n), doğrusal süzgeçten geçirilerek gürültü bastırılması amaçlanmaktadır.

Tez çalışmamızda doğrusal süzgeç olarak Şekil 3.4'te gösterilen FIR (Sonlu Darbe Yanıtlı) süzgeç kullanılmıştır. M. dereceden bir FIR süzgeç giriş işaretinin kendinden önceki M örneğini süzgeç ağırlıklarıyla, w(n), çarpar ve çarpım sonuçları toplanarak çıkış değerini, y(n), elde eder. Çıkış işareti, y(n), ile istenen işaret, d(n), arasındaki fark hata işaretini, e(n), verir.



Şekil 3.3: Uyarlamalı süzgeç yapısı ve yapıya ait kavramlar



Şekil 3.4: M. derece FIR süzgeç iç yapısı

Konuşma işareti frekansı ve genliği zaman içerinde çok hızlı değişebilen dinamik bir yapıya sahiptir. Gürültülü konuşma işaretinin süzgeçlenerek yalın konuşma işaretinin elde edilebilmesi için süzgecin ağırlıklarının, w(n), konuşma işaretine aynı hızda uyum göstermesi gereklidir. Bu yapının sağlanabilmesi için Şekil 3.3'te gösterildiği gibi bir uyarlamalı algoritma ile süzgeç ağırlıkları yenilenir.

Ağırlıkların yenilenmesi işleminde belirleyici olan hata işaretinin büyüklüğüdür. Eğer hata işareti kabul edilebilir sınırın altında ise ağırlıklar değiştirilmeyecek aksi durumda ağırlıklar hatayı küçültecek yönde yenilenecektir. Uyarlamalı süzgeçlerde kullanılan bazı algoritmalar aşağıda verilmiştir:

- LMS (En küçük ortalama kareler)
- NLMS (Normalleştirilmiş en küçük ortalama kareler)
- RLS (Özyinelemeli en küçük kareler)
- Genetik algoritmalar

3.3.1. En küçük ortalama kareler algoritması (Least Mean Squares - LMS)

Uyarlamalı süzgeç ağırlıklarının, w(n), yenilenmesinde kullanılan en temel algoritma LMS algoritmasıdır. M. dereceden bir FIR süzgeç için süzgeç ağırlıkları başlangıç değeri olarak sıfır alınır.

$$w(0) = 0,$$
$$w = \{w_1, w_2, w_3, ..., w_m\} \tag{3.2}$$

Süzgeç çıkışı, y(n), giriş işaretinin, u(n), önceki M örneği ile fitre ağırlıklarının, w(n), çarpılması ile elde edilir.

$$y(n) = w(n)u^T(n), \tag{3.3}$$

$u(n)$ = n anında Mx1 boyutunda giriş vektörü

$$= [u(n), u(n-1), u(n-2), ..., u(n-M+1)]$$

Çıkış işareti ile istenen işaret arasındaki fark hata işaretinin, e(n), değerini verir.

$$e(n) = d(n) - y(n) = d(n) - w(n)u^T(n), \tag{3.4}$$

Hata işareti yeni ağırlıların bulunmasında kullanılacaktır.

$$w(n+1) = w(n) + \mu.e(n)u^*(n), \tag{3.5}$$

Eşitlik (3.5)'deki μ adım boyu olarak adlandırılır ve aşağıdaki belirtilen aralıkta seçilmelidir.

$$0 < \mu < \frac{2}{M.S_{\max}}, \tag{3.6}$$

Burada $S_{\max}$, M boyutlu giriş işaretinin, u(n), güç spektral yoğunluğunun maksimum değeridir.

Adım boyu, μ, değerinin seçimi algoritmanın verimi için çok önemlidir çünkü ağırlıların yenilenmesinde seçilecek büyük adımlar ideal ağırlıkların aşılmasına ve dolayısıyla ters yönde hataya, küçük adımlar ise en ideal ağırlıkların elde edilmesinde gecikmeye neden olacaktır. Adım boyu, μ, değerinin seçimi genellikle deneysel olarak yapılır.

3.3.2. Normalleştirilmiş en küçük ortalama kareler algoritması (Normalized Least Mean Squares - NLMS)

LMS uyarlamalı süzgeçlerinin tasarımında ve uygulamalarında en önemli zorluklardan biri adım boyu, μ, değerinin seçimidir. Durağan işaretler için LMS algoritmasının yakınsaması Eşitlik (3.6) değerine bağlıdır. Genelde giriş işaretinin güç spektral yoğunluğu, $S_{\max}$, bilinmemektedir, sadece sınırları tahmin edilmektedir (Hayes, 1996). NLMS algoritmasında adım boyu, μ, değeri Eşitlik (3.7)'de verildiği şekilde giriş işaretindeki değişimin büyüklüğü referans alınarak her yinelemede tekrar hesaplanır. Böylelikle adım boyu bir sabit olmaktan çıkarılır ve uyarlamalı hale getirilmiş olur.

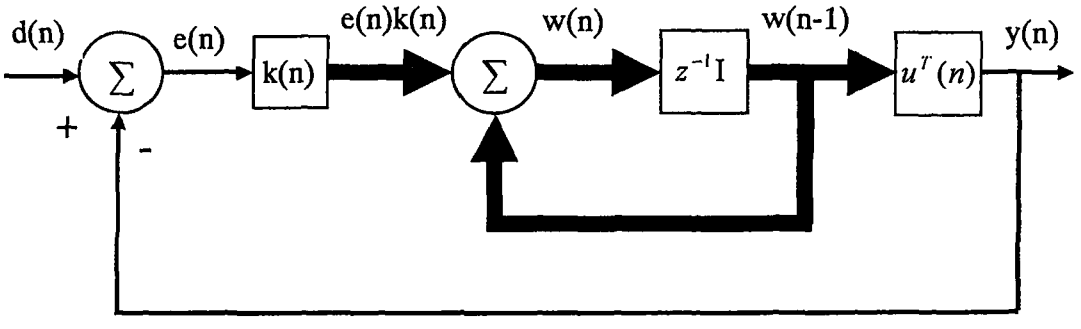
$$\mu(n) = \frac{\mu_0}{u(n)u^T(n)}, \quad \mu_0 = \text{uyarlama sabiti} \tag{3.7}$$

Uyarlama sabiti, μ_0, deneysel olarak belirlenebileceği gibi giriş ve hata işaretleri hakkında ön bilgi mevcut ise hesaplanarak elde edilebilir (Haykin, 2002). NLMS algoritması adım boyu, $\mu(n)$, değerinin zamanla değişen bir parametre olarak kullanması dışında LMS algoritmasıyla tamamen aynıdır. Dolayısıyla algoritmanın diğer parametreleri Eşitlikler (3.2), (3.3), (3.4) ve (3.5) kullanılarak hesaplanır.

NLMS algoritmasının en önemli özelliği ilişkili (correlated) veya ilişkisiz (uncorelated) giriş işareti için standart LMS algoritmasından daha hızlı bir yakınsama göstermesidir (Nagumo et al, 1994).

3.3.3. Özyinelemeli en küçük kareler algoritması (Recursive Least Mean Squares -RLS)

Uyarlamalı süzgeçlerde kullanılan diğer bir algoritma da RLS algoritmasıdır. RLS algoritmasının önemli bir özelliği yakınsa oranının standart LMS algoritmasından daha hızlı olmasıdır. Verimdeki bu iyileşme hesaplama karmaşıklığında bir artışı da beraberinde getirmektedir (Haykin, 2002). Şekil 3.5'de RLS algoritmasının işaret akış diyagramı verilmiştir.



Şekil 3.5: RLS algoritması işaret akış diyagramı

Başlangıç ağırlık değerleri için Eşitlikler (3.2) ve (3.8) kullanılır.

$$P(0) = \delta^{-1} I, \tag{3.8}$$

Burada,

I = birim matris

$$\delta = \begin{cases} \text{Küçük Pozitif Sabit, Büyük SNR için} \\ \text{Büyük Pozitif Sabit, Küçük SNR için} \end{cases}$$

Ağırlıkların yenilenmesi sürecinde kullanılan eşitlikler aşağıda verilmiştir.

$$\pi(n) = P(n-1)u(n), \tag{3.9}$$

$$k(n) = \frac{\pi(n)}{\lambda + u^T(n)\pi(n)}, \tag{3.10}$$

$$w(n) = w(n-1) + k(n)e(n), \tag{3.11}$$

Eşitlik (3.10)'da λ unutma faktörüdür.

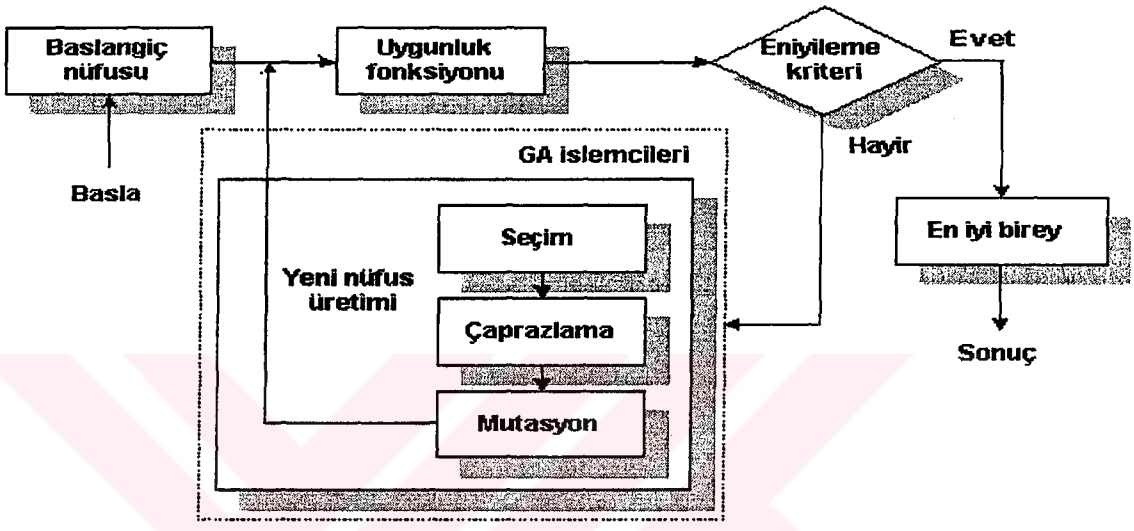
Şekil 3.5'de ve yukarıda verilen eşitliklerde görüldüğü gibi süzgeç ağırlıklarının yenilenmesi sürecinde işlem yükü oldukça fazladır. Eşitlik (3.10)'da ki unutma faktörü, λ, durağan olmayan ortamda süzgeçleme işlemi sırasında gözlemlenebilir verinin istatistiksel değişimlerini takip etmeyi mümkün kılar. Unutma faktörü 1 değerinden küçük pozitif sabit bir sayıdır. RLS süzgecinin çıkışı, y(n), eşitlik (3.3) ve hata işareti, e(n), eşitlik (3.4) kullanılarak hesaplanabilir.

3.3.4. Süzgeç ağırlıklarının yenilenmesinde genetik algoritmaların kullanılması

Genetik algoritmalar (GA) doğada gözlemlenen evrimsel sürece benzer bir şekilde çalışan arama ve eniyileme yöntemidir. Karmaşık çok boyutlu arama uzayında "en iyinin hayatta kalma" ilkesine göre bütünsel en iyi çözümü arar (Söke, 2003).

GA arama uzayından bir grup aday çözüm alınarak bir nüfus oluşturur. Sonra doğal ayıklama, mutasyon ve çaprazlama gibi genetik işlemleri uygulayarak daha iyi ve daha uyumlu kromozomları bularak, gelecek kuşaklara iletir.

Temel bir GA Şekil 3.6'da görüldüğü gibi dört adımdan oluşur. İlk adım başlangıç nüfusunu oluşturacak kromozomların rasgele olarak seçilmesidir. Nüfusun büyüklüğü eniyileme süresince sabit kalır ve nüfus büyüklüğü araştırmacı tarafından belirlenir. İkinci adım problemin çözümü için oluşturulan nüfus içindeki kromozomların sürekliliğine ya da elenmesine karar veren uygunluk fonksiyonudur ve kromozomların uygunluk değerlerinin hesaplanması bu adımda gerçekleştirilir.



Şekil 3.6: Genetik Algoritma akış diyagramı

Genetik algoritmalarda üçüncü adımda bireylerin uygunluk değerlerine bakılarak herhangi bir bireyin problem için çözüm olup olmayacağına karar verilir. Başlangıç nüfusunun uygunluk değerleri için eniyileme ölçütü karşılanıyorsa problemin çözümü olan birey elde edilmiş demektir. Eğer eniyileme ölçütü karşılanmıyorsa genetik algoritma işlemcileri ile yeni bireylerin üretimi yapılır.

Dördüncü adımda yeni nüfus üretimi için GA işlemleri; seçim, çaprazlama, mutasyon gerçekleştirilir. Seçim işlemi, uygunluk değerleri yüksek bireylerin nesiller aracılığıyla daha çok üretilmelerine imkan vermek için yapılan bir genetik işlemdir. Çaprazlama İşlemi, olası çözümü temsil eden bireyler arasındaki bilgi değişimini gerçekleştirerek daha iyi çözümlerin üretilmesini sağlar. Çaprazlama işlemi ise biyolojide iki kromozomun çaprazlanmasına benzer şekilde, sayı dizileriyle gösterilen iki bireyin belirli parçalarının yerleri karşılıklı değiştirilerek iki yeni birey elde edilir. Mutasyon işlemi, bir bireyde herhangi bir genin değerinin değiştirilmesi

ya da genlerin yerlerinin değiştirilmesiyle olur. Daha önceden belirlenmiş mutasyon oranı, herhangi bir bireyin mutasyona uğrayıp uğramayacağını belirler (Söke, 2003).

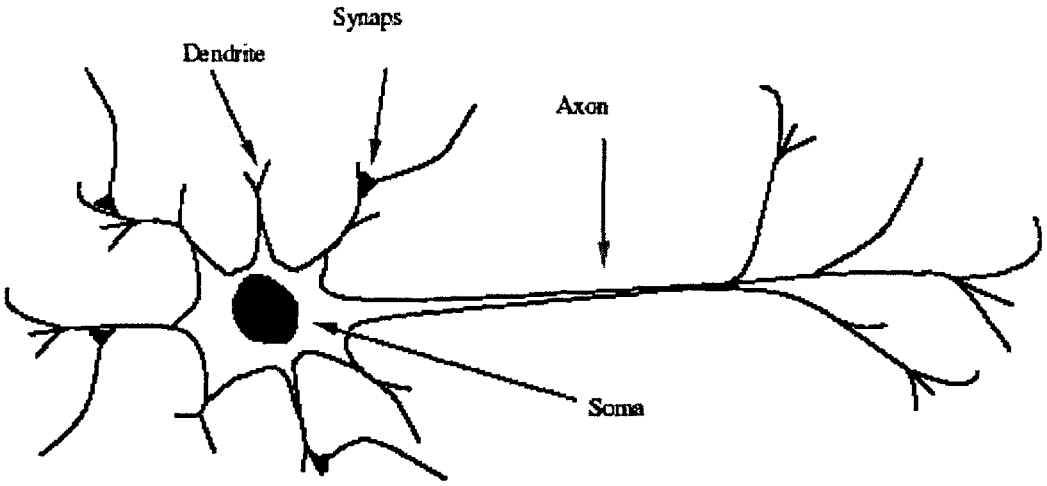
GA işlemleri sırasında kromozomların çaprazlama olasılığı genellikle yüksek bir değer seçilir (0.6–1.0). Eğer bu ebeveynlerin eşleştirilmelerine izin verilmişse bir çaprazlama işlemi gerçekleşir. Mutasyon olma olasılığı genellikle düşük bir değerdir (0.01–0.1). Böylelikle iyi kromozomların bozulmaması sağlanır. İkinci adımdan itibaren aynı işlemler belirlenen sayıda kuşaklar boyunca tekrarlanarak eniyileme gerçekleştirilir (Fatihoğlu, 2003).

GA için araştırma alanları çok geniştir ve genellikle küresel en iyi değere yakın bir noktada yakınsama sağlarlar (Mastorakis et al, 2003). Genetik Algoritmalar, hem sayısal hem de analog süzgeçlerin tasarımında başarılı bir şekilde uygulanmıştır (Arslan and Horrocks, 1996).

3.4. Gürültü Bastırma İşleminde Yapay Sinir Ağlarının Kullanılması

YSA (Yapay Sinir Ağları) insan beyninin çalışması ve biyolojik sinir sistemi modellenerek geliştirilmiş bir veri yorumla sistemidir. YSA deneysel bilgiyi saklama, kullanıma hazır hale getirme yeteneğine sahip basit işleme birimlerinden oluşan, paralel ve dağıtık düzende çalışabilen bir işlemcidir.

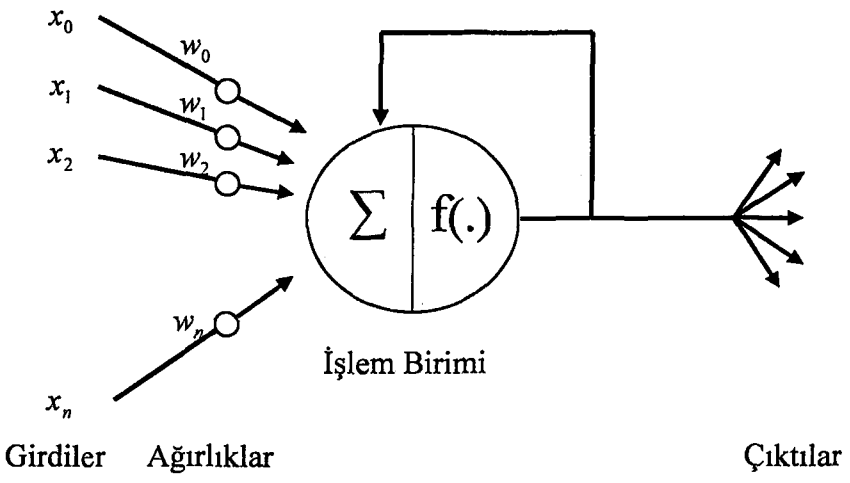
Şekil 3.7'de biyolojik bir beyin sinir hücresi ve bileşenleri gösterilmiştir. Biyolojik sinir ağlarında giriş işaretlerini alan, yorumlayan ve uygun çıktıyı ileten temel işlemci işlem birimi olarak adlandırılır. Bir işlem birimi; gövde, gövdeye giren işaret alıcıları (dendrit) ve gövdeden çıkan işaret iletici (akson) olmak üzere üç kısımdan oluşur (Sağıroğlu et al, 2003). Aksonların bitimi ince yollarla ayrılabilir ve bu yollar diğer hücreler için dendritleri oluşturur. Akson-dendrit bağlantı elemanı sinaps olarak adlandırılır.



Şekil 3.7: Biyolojik bir beyin sinir hücresi ve bileşenleri

YSA, biyolojik sinir sisteminin çalışma şeklinin benzetimini yapan programlardır. YSA öğrenme, ezberleme ve veriler arasındaki ilişkiyi ortaya çıkarma kapasitesine sahiptirler.

Biyolojik sistemlerde öğrenme, işlem birimleri arasındaki sinaptik bağlantıların ayarlanması ile olur. YSA'da öğrenme, biyolojik sisteme benzer şekilde eğitim olarak adlandırılan girdi/çıktı verilerinin işlenmesiyle, yani eğitm algoritmasının bu verileri kullanarak bağlantı (sinaps) ağırlıklarını bir yakınsama sağlanana kadar, tekrar tekrar ayarlamasıyla gerçekleşir (Yurtoğlu, 2005).



Şekil 3.8: Yapay sinir hücresinin genel yapısı

Yapay sinir ağlarının temel işlem elemanı olan yapay işlem birimleri, doğal işlem birimlerinin dört temel fonksiyonunun benzetimini yaparlar. Bir yapay işlem biriminin temel yapısı Şekil 3.8'de verilmiştir. Öncelikle yapay işlem birimi giriş değerleri, x(i), ağırlıklarla, w(i), çarpılır. Daha sonra yapay işlem birimi içerisinde tüm sonuçlar toplanarak bir transfer fonksiyonuna gönderilir ve sonuç üretilir. Bu sonuç son olarak bir çıktıya dönüştürülür.

Tablo 3.1'de biyolojik sinir sistemi ile yapay sinir ağını yapısal olarak karşılaştırılması verilmiştir.

Tablo 3.1: Biyolojik sinir sistemi bileşenlerinin yapay sinir ağında karşılıkları

Biyolojik Sinir Ağı	Yapay Sinir Ağı
Nöron (sinir hücresi)	İşlem birimi
Dendrit	Toplama fonksiyonu
Hücre gövdesi	Transfer fonksiyonu
Akson	Yapay sinir hücresi çıkışı
Sinaps	Ağırlıklar

YSA'larda kullanılan bazı transfer fonksiyonları aşağıda verilmiştir.

Doğrusal $$f(s) = s \tag{3.12}$$

Sigmoid $$f(s) = \frac{1}{1+e^{-s}} \tag{3.13}$$

Hiperbolik tanjant $$f(s) = \frac{1-e^{-2s}}{1+e^{-2s}} \tag{3.14}$$

Doğrusal ve doğrusal olmayan aktivasyon fonksiyonlarının YSA'larında kullanılması karmaşık ve çok farklı problemlere uygulanmasını mümkün kılmıştır. Yapay sinir ağlarının özellikleri ve avantajları aşağıda verilmiştir (Kaya ve diğ., 2005):

1. Doğrusal Olmama: Bütün sonuçların her bir girdiyle etkileşim içinde olduğu anlamına gelir. Doğrusal bir sistemde girdilerin biri değiştirildiğinde, çıktı bu değişime orantılı olarak değişir ve bu etki sadece değiştirilen girdinin değeri ile bağlantılıdır.

2. Hata Toleransı: YSA paralel dağılmış parametreli bir sistem olduğundan her bir işlem birimi yalıtılmış bir ada gibidir. Bu nedenle işlem birimleri zarar görse bile bu sistemin veriminin düşmesine neden olabilir ancak sistem hiçbir zaman durmaz. YSA sistemlerinin hata toleranslı olmasının sebebi, bilginin sisteme dağıtılıyor olmasıdır. Bu özellik sistemin durmasının önemli bir zarara neden olacağı uygulamalarda önem kazanır.

3. Eğitim: YSA ağırlık katsayılarını değiştirerek, somut problemin çözümü için kendini adapte etme özelliğine sahiptir. Bu durum çözülen problem alanı hakkında çok da bilgiye sahip olmadan ağın parametrelerini hesaplamaya imkân verir.

4. Öğrenme: YSA'nın bir problemi öğrenmesi için problemin giriş verilerine karşılık gelen çıkış verilerinin veya sadece giriş verilerinin sağlanması gerekir. Giriş ve çıkış verileri ile tanımlanan öğrenme kümesi yeteri sayıda örnek içermelidir. Öğrenme süreci sistem girişleri ve sistem çıkışı arasındaki ilişkiyi veya giriş verileri arasındaki ilişkilendirmeyi tanımlayan ağırlık verilerinin elde edilme süreci olarak ta tanımlanabilir.

5. Genelleme: YSA eğitilen verileri öğrettikten sonra eğitimde kullanılmayan veriler için bir yanıt oluşturur.

6. Bellek: YSA'nın önemli bir özelliği, bilgiyi saklama şeklidir. Biyolojik sistemlerde veri dağınık yapıda saklanır. YSA'da bellek, birçok yerel bellekler oluşturularak dağıtılır. Bağlantı ağırlıkları YSA bellek biçimleridir. Ağırlıkların değerleri, ağın o anki bilgi durumunu temsil eder. Örneğin; bir giriş/istenen çıkış çiftinin belirtilen bilgi parçası ağın içinde birçok bellek biçimine dağıtılmıştır. Bellek birimleri ile diğer saklı bilgiler, bu bilgiyi paylaşırlar.

YSA'lar çok çeşitli şekillerde sınıflandırılabilirler. Temel olarak YSA'lar ağ yapılarına ve öğrenme yaklaşımlarına göre olmak üzere iki sınıfa ayrılabilirler.

Ağ yapılarına göre YSA' lar iki gruba ayrılırlar;

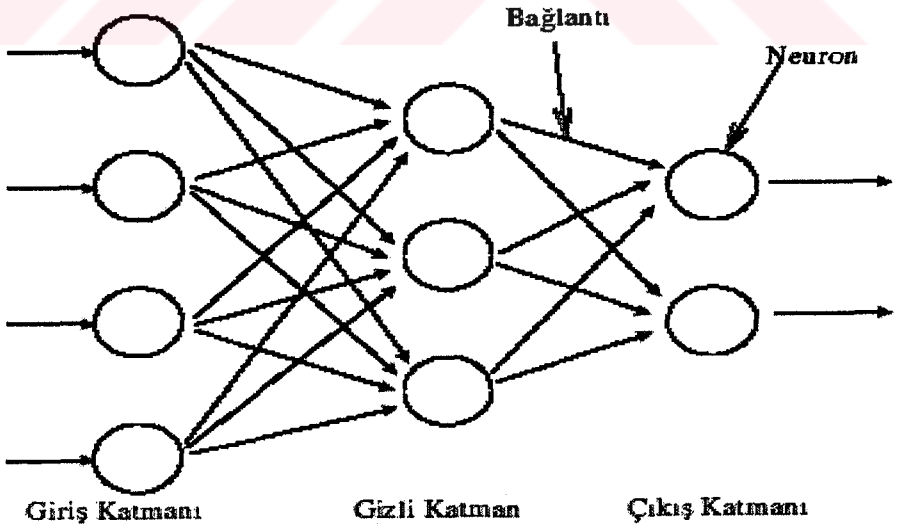
- Tek katmanlı ağlar
- Çok katmanlı ağlar

YSA'lar öğrenme yaklaşımlarına göre üç gruba ayrılır;

- Eğiticili Öğrenme
- Eğiticisiz Öğrenme
- Takviyeli Öğrenme

3.4.1. Çok katmanlı ileri beslemeli ağlar

Tek katmalı ağlardaki giriş ve çıkış katmanından başka, bir ya da daha fazla sayıda gizli katman içeren ağlara çok katmanlı ağ denir. Dış dünya tarafından doğrudan müdahale edilmediği için gizli katman adı verilen katmanda bulunan işlem birimlerine de gizli işlem birimleri adı verilir (Aydın, 2005). Tez çalışmamızda da kullanılan çok katmalı ileri beslemeli bir ağ modeli Şekil 3.9'da gösterilmektedir.



Şekil 3.9: Çok katmanlı ileri beslemeli ağ modeli

Çok katmanlı ağlar(ÇKA) tek katmanlı ağlara göre daha karmaşık problemlere çözüm getirebilmektedir. Ancak bu ağların eğitilmesi katman ve işlem birimi sayısı arttığından daha uzundur.

3.4.2. YSA öğrenme kuralları

Yapay Sinir Ağlarında çok sayıda öğrenme kuralı kullanılmaktadır. Bu öğrenme kurallarının büyük çoğunluğu en eski ve en çok bilinen öğrenme kuralı olan Hebb kuralının birer sürümüdürler. Bunların yanında, farklı öğrenme kurallarına ait transfer fonksiyonları da vardır. Kullanılmakta olan bazı önemli öğrenme kuralları şunlardır (Yurtoğlu, 2005):

- Hebb kuralı
- Hopfield kuralı
- Eğimli Değişim (Gradient Descent) kuralı
- Kohonen öğrenme kuralı
- Delta kuralı

Tez çalışmamızda kullanılan Delta öğrenme kuralı ilk olarak Widrow ve Hoff tarafından geliştirilmiş daha çok mühendislik kökenli bir algoritmadır (Yurtoğlu, 2005). Bu kural, işlem biriminin çıkışı ile istenilen çıkış değerleri arasındaki farkı azaltan, giriş bağlantılarını güçlendiren ve sürekli olarak değiştiren bir düşünceye dayanmaktadır. Delta kuralı, ortalama karesel hatayı (MSE), bağlantı ağırlık değerlerinin değiştirilmesi ile düşürme prensibine dayanır. Bu nedenle de bu algoritma en küçük kareler (LMS) kuralı olarak da bilinmektedir. Hata aynı anda bir katmandan bir önceki katmanlara geriye yayılarak azaltılır. Ağın hatalarının düşürülmesi işlemi, çıkış katmanından giriş katmanına ulaşıncaya kadar devam eder. Delta öğrenme algoritması hataların karesinin en küçük olduğu noktayı bulurken dereceli azaltma yöntemini (gradient descent) kullanmaktadır. Bu yöntemde, hatanın kareleri, koordinatlarını ağırlıkların oluşturduğu uzayda bir çanak oluşturmaktadır. Delta algoritması mevcut ağırlık vektörünü, bulunduğu konumdan hatanın en küçük olduğu çanağın dibine doğru ilerletir.

YSA, işaret işlemenin pek çok alanına uygulama imkanı vardır. Birbirinden bağımsız kaynaklardan farklı işaretlerin eşzamanlı süzgeçlenmesi veya karıştırılması durumunda, orijinal kaynak işaretini çıkaracak gözlem süreci YSA tarafından gerçekleştirilebilir (Amari and Cickocki, 1998). YSA'nın uyarlamalı süzgeç olarak kullanımı yeni bir kavram değildir ve uyarlamalı sinir ağı süzgeçleri çeşitli yinelemeli algoritmaların içinde kullanılmıştır (Ngia et al, 1998).

FIR sinir ağları, doğrusal sonlu darbe yanıtlı süzgeçler tarafından yenilenen düzenli skaler snapslere sahip ileri beslemeli YSA'lardır. Bu YSA'lara, geçmiş giriş değerlerine bağlı dinamik haritalama gerçekleştirebilme kapasitesiyle zaman serilerinin tahmini, doğrusal olmayan sistem tanımlaması ve işaret işleme uygulamalarını mümkün kılar. En bilenen şekli, sonlu darbe yanıtlı süzgeçlerin katsayıları ile ağırlıkların yer değiştirildiği FIR sinir ağlarıdır (Cholewo and Zurada, 1998). Bu ağlarda kullanılan en yaygın ağ mimarisi Çok Katmanlı Ağdır(ÇKA). ÇKA'da ki gizli katmandaki işlem birimleri delta öğrenme kuralıyla minimize edilerek eğitilir (Back and Tsoi, 1991).

3.5. Gürültü Bastırma İşleminde Uyarlamalı Sinir-Bulanık Çıkartım Sisteminin (Adaptive Neuro-Fuzzy Inference System - ANFIS) Kullanılması

Bulanık mantık günlük yaşamda her gün kullandığımız davranışlarımızı tanımladığımız ve yorumladığımız yapıya ulaşmamızı sağlayan matematiksel bir disiplindir. Bulanık mantık, ilk kez 1965 yılında Profesör Lotfi ZADEH tarafından ortaya konulmuştur (Hellmann, 2001). Klasik kontrol yöntemleri ile modellenmesi çok zor olan ya da modellendiği durumda sistemin tanım aralığı civarındaki değerler için doğru tanımlamanın yapılamadığı durumlarda kullanılabilecek bir kontrol metodudur. Temelini, doğru ve yanlış değerlerin belirlendiği Bulanık Küme Kuramı (Fuzzy Set Theory) oluşturur (Steimann, 1997). Burada yine klasik mantık da olduğu gibi (1) ve (0) değerleri vardır. Ancak bulanık mantık yalnızca bu değerlerle yetinmeyip bunların ara değerlerini de kullanarak; örneğin uzaklıkla ilgili bir problemde mesafenin yalnızca yakın ya da uzak olduğunu belirtmekle kalmayıp ne kadar yakın ya da ne kadar uzak olduğunu da belirtir (Jantzen, 1998).

Dilsel ifadeleri kullanan Bulanık Mantık; biraz sıcak, ılık, uzun, çok uzun, soğuk gibi günlük hayatımızda kullandığımız kelimeler yardımıyla insan mantığına en yakın doğrulukta elektronik denetimi sağlayabilir.

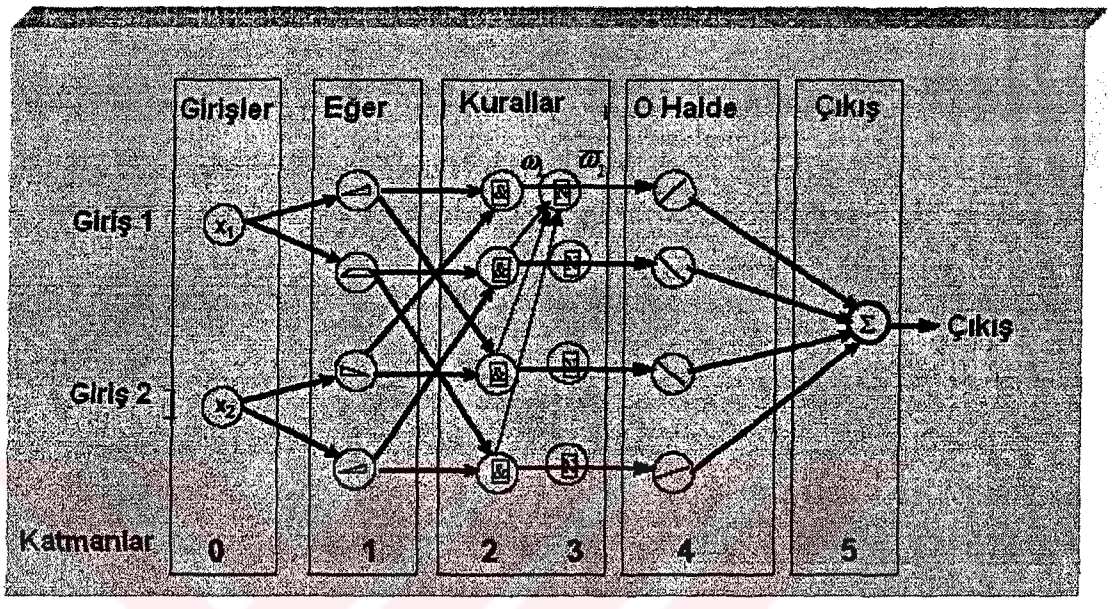
Temel olarak bulanık mantık sistemleri, insan tecrübelerinden türetilmiş kurallar olarak adlandırılan "Eğer ... ise, o halde ..." (If-Then) değerlendirmesi yaparlar (Ishibuchi et al, 1995). Tüm kurallar paralel olarak değerlendirilir ve kural sayısı önemsizdir. Kurallar da değişkenlerden ve değişkenleri tanımlayan amaçlardan söz edilir. Bulanık mantık sistemleri aşağıdaki aşamalar göz önüne alınarak oluşturulur;

- Kural Tabanı (kurallar) oluşturma
- Doğruluk Değerlerini (bazı değer ve durumlar için) Belirleme
- Kuralların öncelik derecelerini (kuvvetlerini) belirleme
- Her durum için Bulanık Mantık çıkışlarını belirleme

ANFIS bulanık mantık sistemlerinin ve yapay sinir ağlarının en iyi özellikleri birleştirilerek 1992 yılında Roger Jang tarafından geliştirilmiştir (Helie et al, 2004). Bu yapıda eğiticili öğrenme kuralına sahip ileri beslemeli yapay sinir ağlarının tüm türlerinden oluşturulan bir küme olarak uyarlamalı uyarlamalı tanımlanmıştır. Uyarlamalı ağ, işlem birimlerini ve bu birimlerin bağlantılarını içeren bir ağ yapısıdır. İşlem birimlerinin hepsi veya bir kısmı, çıkışları kendi parametrelerine bağlı olan ve öğrenme kurallarının öngörülmüş hata değerini minimize etmek için bu parametrelerin nasıl değiştirileceği açıkça belirtilmiş uyarlamalı işlem birimleridir (Jang, 1993). Uyarlamalı ağların temel öğrenme kuralı dereceli azalma (gradient descent) ve zincir kuralını esas alır ve daha sonra ise melez (hybrid) öğrenme kuralı kullanılarak çalışmalar gerçekleştirilmiştir. Şekil 3.10'da bir ANFIS ağının yapısı verilmiştir. ANFIS içindeki işlem birimleri farklı yapılara sahiptiler (Bonissone, 2002):

- Değer işlem birimleri: YSA'daki transfer fonksiyonlarına benzer yapıdaki genelleştirilmiş Bell fonksiyonları tarafından tanımlanan üyelik fonksiyonları
- Kural işlem birimleri: Genellikle çarpım işlemi
- Normalleştirme işlem birimleri: Toplama ve aritmetik bölme işlemi

- Fonksiyon işlem birimleri: Doğrusal yansıma ve normalleştirilmiş ağırlıklarla çarpma .
- Çıkış işlem birimleri: Cebirsel toplama



Şekil 3.10: ANFIS ağının yapısı

Şekil 3.10'daki ANFIS ağı yapısındaki her katmanın işlevi aşağıda açıklanmıştır (Bonissone, 2002):

- K0: durum değişkenleri, ANFIS giriş katmanı içindeki düğümlerdir.
- K1: her bir durum değişkeni terim kümesi, üyelik değerlerini hesaplayan ANFIS değerler katmanı içindeki düğümlerdir.
- K2: bulanık mantık yapısındaki her bir kural, kural eşleştirme faktörünü, w_i, hesaplamak için yumuşak minimumlaştırma veya çarpma işlemi yapan ANFIS kurallar katmanı içindeki düğümlerdir.
- K3: her bir kural eşleştirme faktörü, w_i, normalleştirme katmanı içerisinde $\overline{w_i}$ ile ölçeklendirilir.
- K4: her bir $\overline{w_i}$ ağırlığı fonksiyon katmanı içinde kural çıkışı üretir.
- K5: her bir kural çıkışı birleştirilerek çıkış katmanına gönderilir.

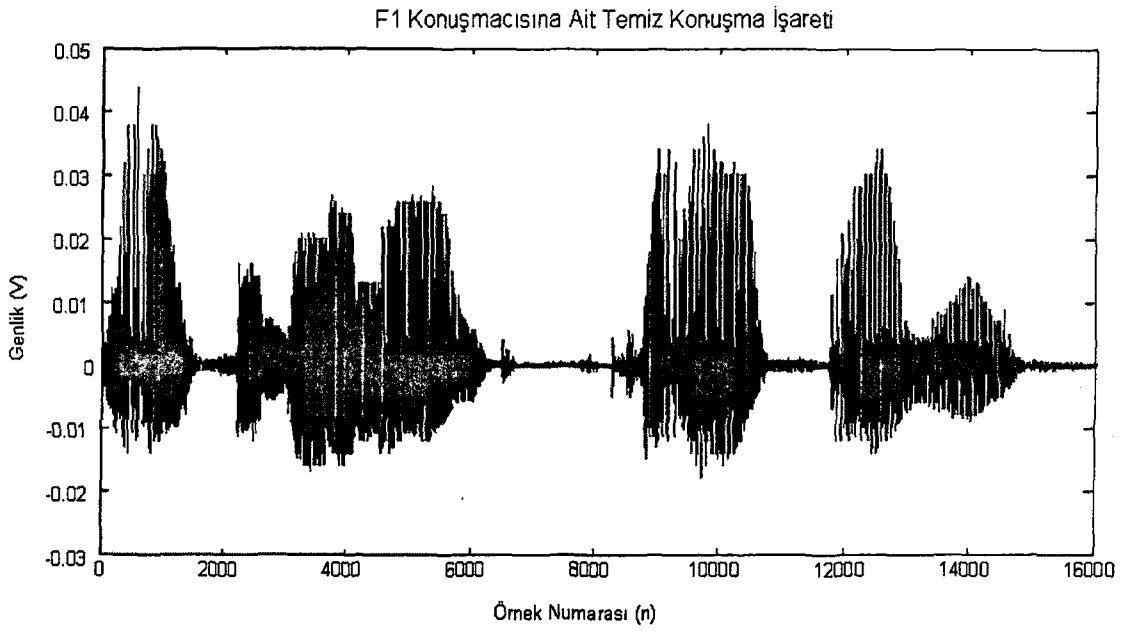
3.6. Gürültü Bastırma Çalışmaları

Gürültü bastırma işlemi için yapılan deneysel çalışmalar için öncelikle veritabanı oluşturulmuştur. Konuşma verileri Center for Spoken Language Understanding (CSLU) konuşmacı tanıma sürüm 1.1 veritabanından seçilmiştir. Oluşturduğumuz konuşma veritabanı verileri, 5 bayan ve 5 erkek konuşmacının Ad-Soyad ve yaşadıkları şehri telaffuz ettikleri yaklaşık olarak 1,5–5 sn. arasında süren cümlelerdir. Her bir konuşmacının verisi 8 KHz örnekleme frekansına sahip, 16 bit, mono kanal ve "wav" uzantılı dosyalar şeklindedir. Bayan konuşmacılar; F1, ..., F5 ve erkek konuşmacılar; M1, ..., M5 şeklinde etiketlenmiştir.

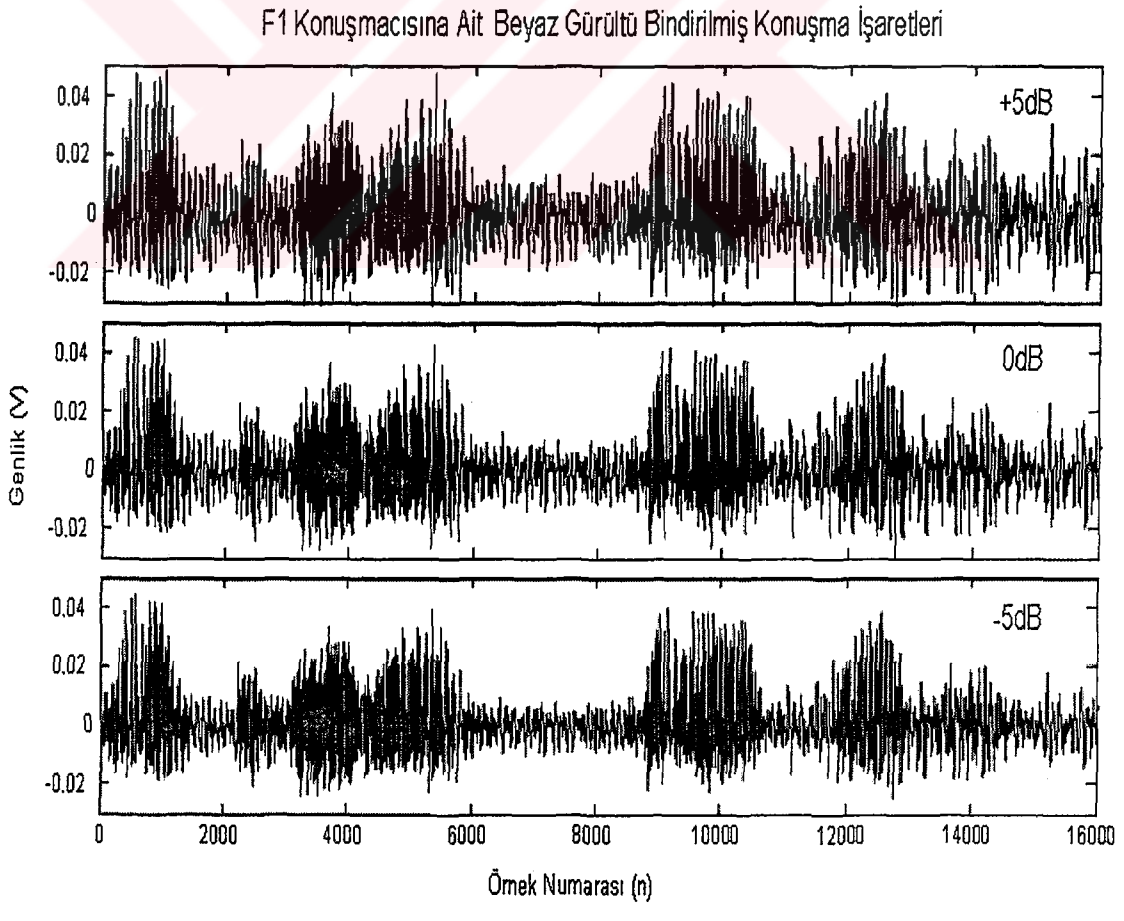
Çalışmamızda Gaussian beyaz gürültü ve BMW 750i marka otomobil içerisinde kaydedilmiş ses gürültüsü de renkli gürültü olarak kullanılmıştır. Konuşmacı işaretlerine her gürültü türü için, üç farklı SNR; +5dB, 0dB, -5dB oranlarında gürültü bindirilerek çalışmalar gerçekleştirilmiştir. Şekil 3.11. (a)'da birinci bayan (F1) konuşmacıya ait temiz konuşma işareti, (b)'de üzerine +5dB, 0dB ve -5dB beyaz gürültü bindirilmiş F1'e ait konuşma işaretleri , (c)'de ise üzerine +5dB, 0dB ve -5dB renkli gürültü bindirilmiş F1'e ait konuşma işaretleri örnek olarak verilmiştir.

Çalışmalar P4 2.0 GHz işlemci ve 512 MByte RAM hafızaya sahip bilgisayar üzerinde ve MATLAB 6.5 sürüm (R13) programında gerçekleştirilmiştir. Çalışmalarımızda gürültü bastırma işlemi için NLMS, RLS gibi uyarlamalı süzgeç algoritmalarının yanı sıra GA, ileri beslemeli çok katmanlı YSA ve ANFIS gibi yapay zeka ile hesaplama (soft computing) algoritmaları kullanılmıştır.

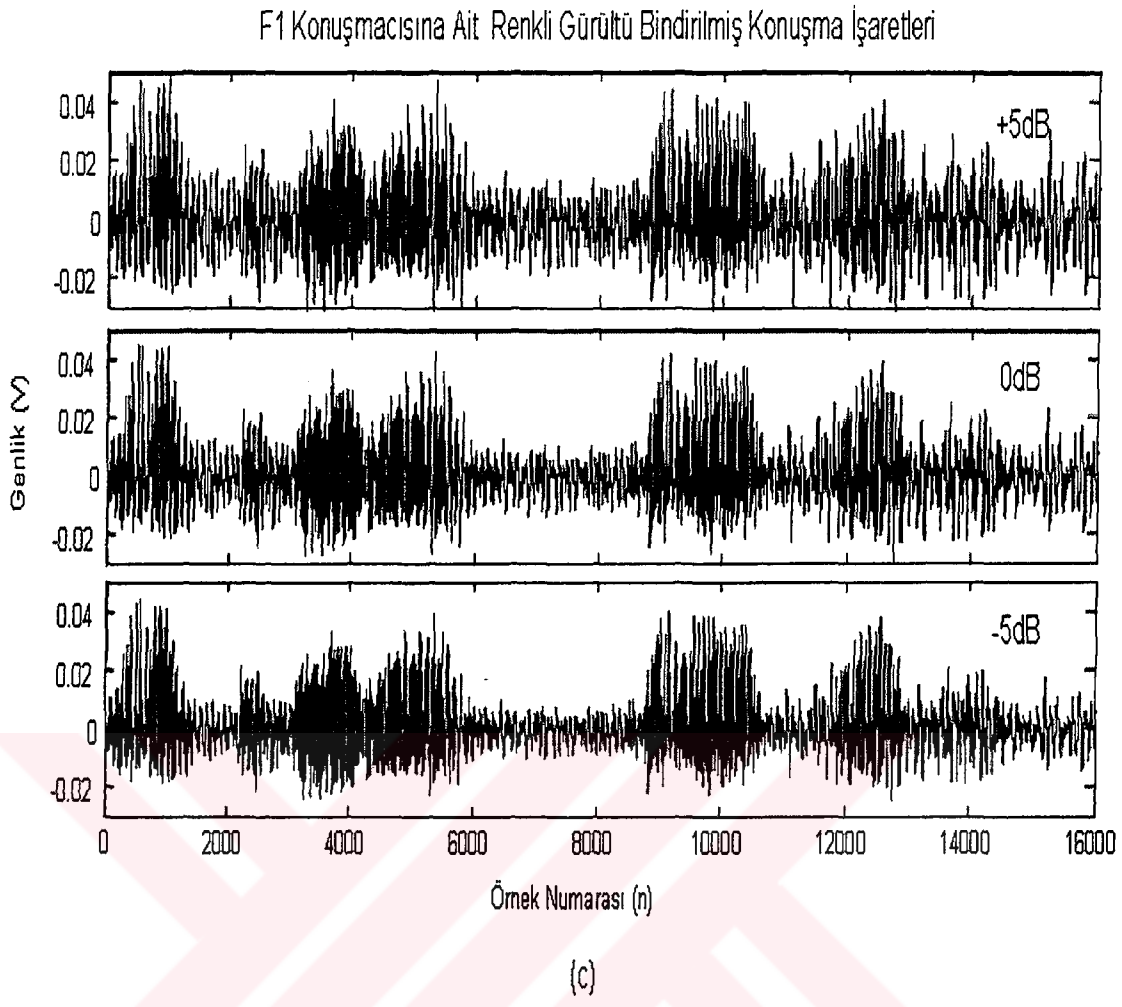
Yapılan çalışmalar esnasında gözlenebilirliği arttırmak amacıyla Şekil 3.12'de gösterilen grafik kullanıcı arayüzü (Graphical User Interface - GUI) hazırlanmıştır. Hazırlanan GUI penceresi, bilgisayar üzerinde işlemler devam ederken kullanıcıya algoritma özellikleri, gürültü çeşidi, gürültü şiddeti, konuşmacı bilgisi, hata değerleri gibi işlem bilgilerinin yanı sıra gürültülü işaret ve süzgeçlenmiş işaret şekillerini göstermektedir.


F1 Konuşmacısına Ait Temiz Konuşma İşareti
Genlik (V)
0.05
0.04
0.03
0.02
0.01
0
-0.01
-0.02
-0.03
0
2000
4000
6000
8000
10000
12000
14000
16000
Örnek Numarası (n)

(a)

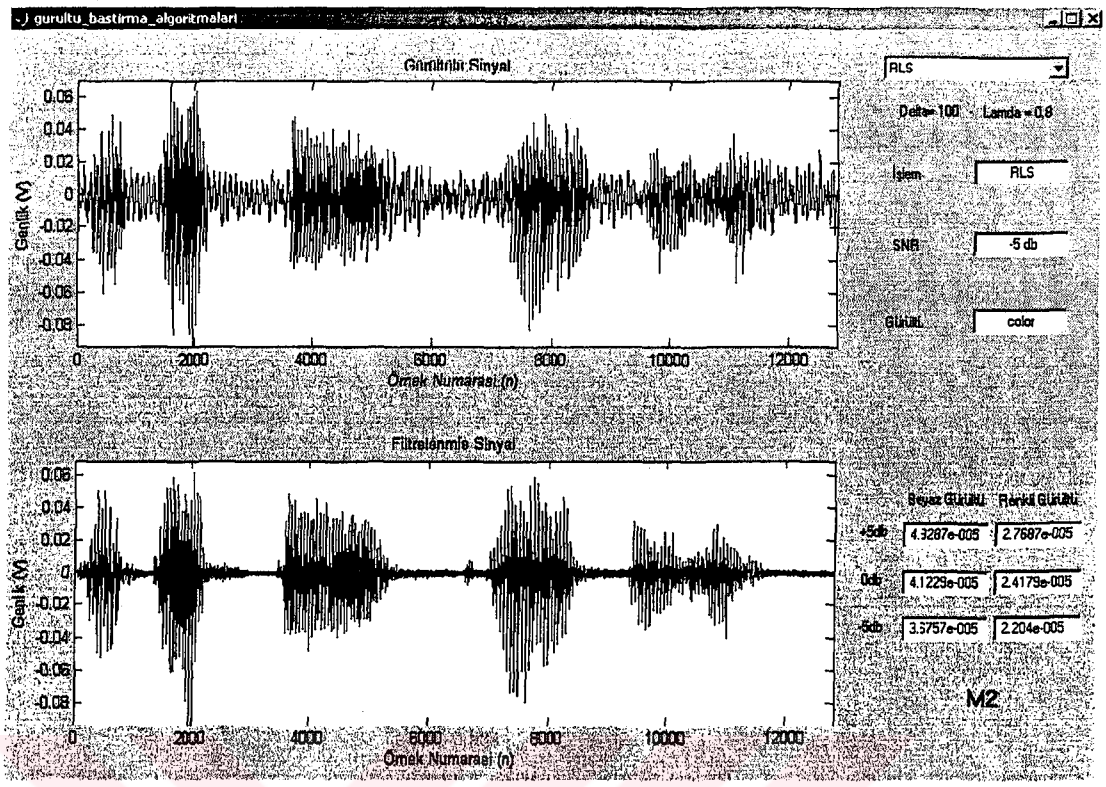

F1 Konuşmacısına Ait Beyaz Gürültü Bindirilmiş Konuşma İşaretleri
+5dB
0dB
-5dB
Genlik (V)
0.04
0.02
0
-0.02
0
2000
4000
6000
8000
10000
12000
14000
16000
Örnek Numarası (n)

(b)



Şekil 3.11: F1'e ait; (a) temiz konuşma işareti, (b) +5dB, 0dB, -5dB beyaz gürültü bindirilmiş konuşma işaretleri, (c) +5dB, 0dB, -5dB renkli gürültü bindirilmiş konuşma işaretleri

Algoritmaların gürültülü ses işaretlerini, ne kadar hata ile temizleme verimleri ve gerçek zamanlı uygulamalarda kullanılabilirliği belirlemek amacıyla işlem süreleri karşılaştırılmıştır. Alt bölümlerde kullanılan yöntemler ve elde edilen sonuçlar açıklanmıştır.



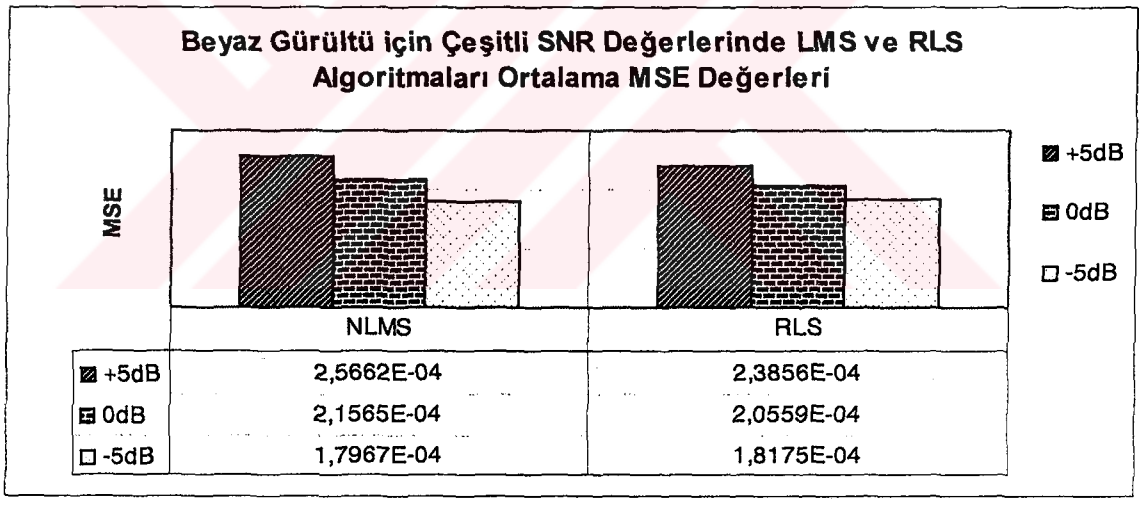
Şekil 3.12: Gürültü bastırma işlemi için hazırlanan GUI

3.6.1. NLMS uygulaması

Çalışmamızda ilk olarak 10. dereceden bir FIR süzgeç için ağırlıkların yenilenmesi işleminde NLMS algoritması kullanılmıştır. Düz bir frekans spektrumuna sahip işaretler için iyi sonuç veren LMS algoritmasının verimi ses işareti gibi frekans spektrumu geniş bir bölgeye dağılmış, genliği ve frekansı çok hızlı değişen dinamik işaretler için hata işaretini minimize etmekten çok uzaktır ve genellikle hatayı sonsuza götürür. NLMS algoritmasında adım boyu işaret değişimlerine bağlı olarak her yinelemede ayarlanır ve algoritmanın dinamik bir yapıya sahip işaretlerini süzgeçlemesini mümkün kılar. NLMS algoritmasında uyarlama sabiti, μ_0, çeşitli denemeler sonunda 0.01 seçilmiştir.

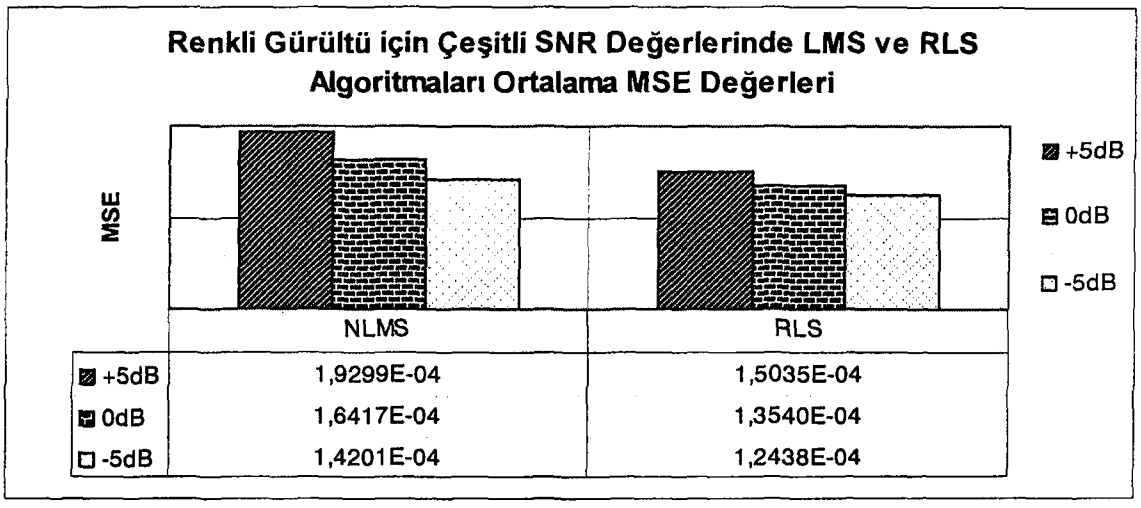
3.6.2. RLS uygulaması

Çalışmamızda 10. dereceden bir FIR süzgeç için ağırlıkların yenilenmesi işleminde RLS algoritması kullanılmıştır. RLS algoritmasında, δ sabiti uygulanan SNR değerleri de göz önüne alınarak çeşitli denemeler sonucunda 100 ve unutma sabiti , λ, ise çeşitli denemeler sonucunda 0.8 olarak alınmıştır. Oluşturulmuş veritabanı üzerinde yapılan çalışmalar sonucunda her bir SNR değeri (+5dB, 0dB, -5dB) için on konuşmacının MSE değerleri hesaplanmıştır. Elde edilen MSE değerlerinin ortalaması alınarak, beyaz gürültü için NLMS ve RLS algoritmalarının ortalama MSE değerleri karşılaştırılan grafik Şekil 3.13'te, renkli gürültü için NLMS ve RLS algoritmalarının ortalama MSE değerleri karşılaştırılan grafik Şekil 3.14'de verilmiştir. Verilen iki grafikte ortalama MSE değerleri incelendiğinde RLS algoritmasının iki gürültü türü içinde daha iyi sonuç verdiği görülmektedir.

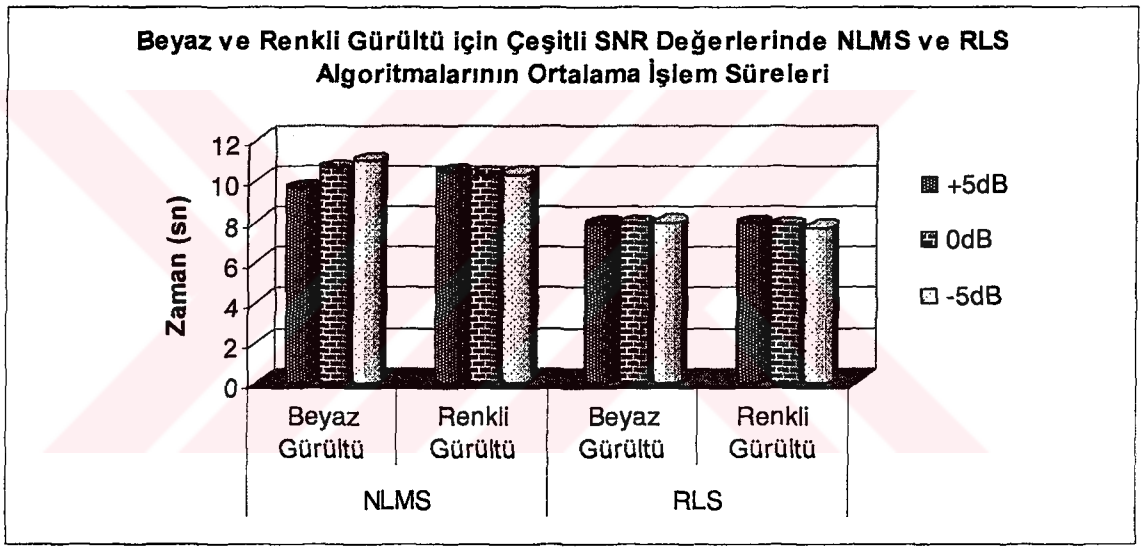


Şekil 3.13: Beyaz gürültü için NLMS ve RLS algoritmalarının karşılaştırılması

Her iki gürültü çeşidi ve her bir SNR değeri için iki algoritmanın ortalama işlem sürelerinin karşılaştırıldığı grafik Şekil 3.15'de verilmiştir. Grafikte görüldüğü gibi RLS algoritması tüm çalışmalarda NLMS algoritmasına göre daha hızlı bir yakınsama göstermektedir.



Şekil 3.14: Renkli gürültü için NLMS ve RLS algoritmalarının karşılaştırılması



Şekil 3.15: NLMS ve RLS algoritmalarının ortalama işlem sürelerinin karşılaştırılması

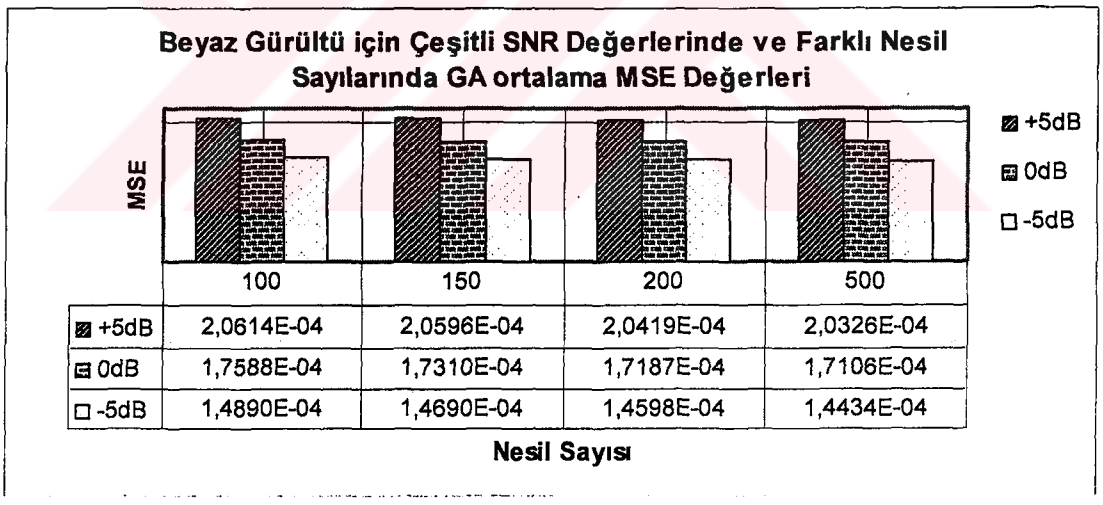
3.6.3. GA uygulaması

Gürültünün bastırılması işlemi için çalışmalarımızda kullanılan GA başlangıç parametreleri ve işlemleri Tablo 3.2'de gösterilmiştir. GA parametreleri sabit tutulurken, dört farklı nesil sayısı için denemeler yapılmış ve en iyi verime sahip nesil sayısı tespit edilmeye çalışılmıştır.

Tablo 3.2: GA başlangıç parametreleri ve işlemleri

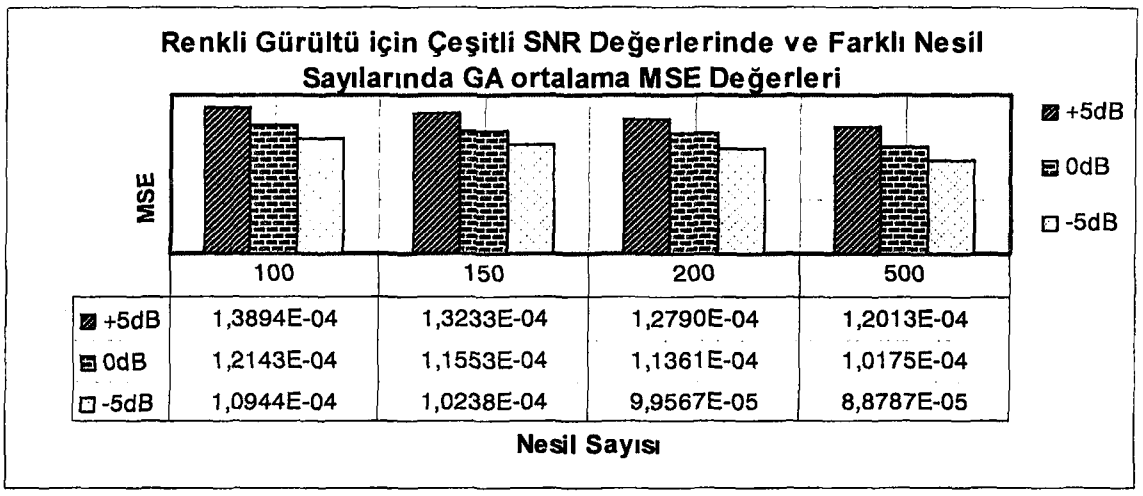
Nesil sayısı	N=100, 150, 200, 500
Nüfus büyüklüğü	100
Kromozom uzunluğu	10
Kromozom tipi	Reel
Mutasyon olasılığı	$p_m = 0.1$
Çaprazlama olasılığı	$p_c = 0.8$
Seçim tipi	Rulet tekeri
Mutasyon tipi	Kromozomdan kromozoma
Çaprazlama tipi	Tek noktada
Seçicilik	Basit

Çalışmalar sonucunda her bir SNR değeri (+5dB, 0dB, -5dB) için on konuşmacının MSE değerlerinin ortalaması hesaplanmıştır. Beyaz gürültü için100, 150, 200 ve 500 nesilde ortalama MSE değerleri karşılaştırılan grafik Şekil 3.16'de, renkli gürültü için ortalama MSE değerleri karşılaştırılan grafik ise Şekil 3.17'de verilmiştir.



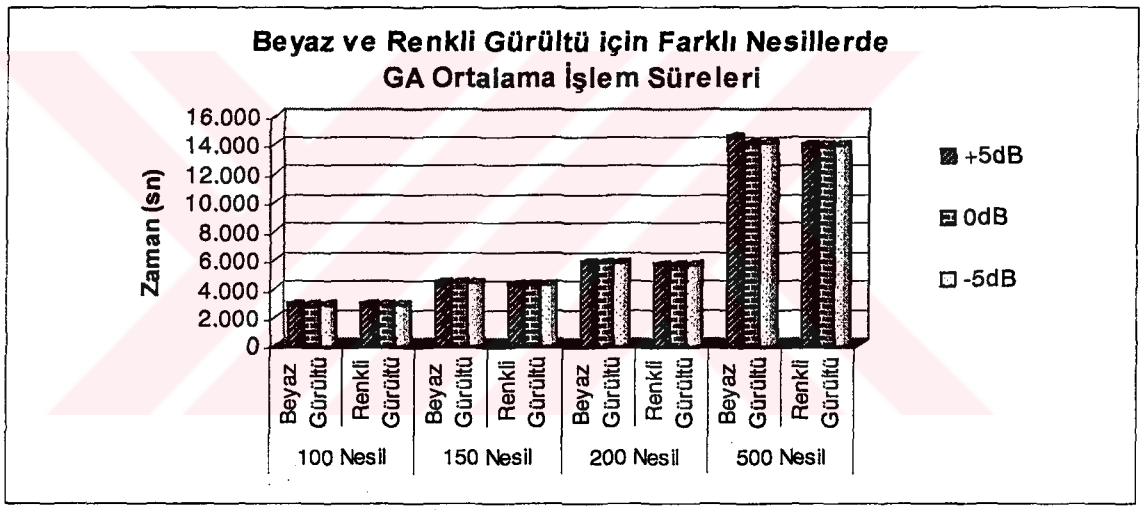
Şekil 3.16: Beyaz gürültü için farklı nesil sayılarında GA verimlerinin karşılaştırılması

Verilen iki grafikte ortalama MSE değerleri incelendiğinde GA beklenildiği gibi en yüksek nesil sayısı olan 500 nesilde en düşük hata değerleri ile gürültüyü bastırmıştır. Fakat işlemlerin gerçekleşme süreleri incelendiğinde nesil sayısı arttıkça işlem sürelerinin arttığı Şekil 3.18'de verilen GA işlem süreleri karşılaştırma grafiğinde görülmektedir.



	100	150	200	500
+5dB	1,3894E-04	1,3233E-04	1,2790E-04	1,2013E-04
0dB	1,2143E-04	1,1553E-04	1,1361E-04	1,0175E-04
-5dB	1,0944E-04	1,0238E-04	9,9567E-05	8,8787E-05

Şekil 3.17: Renkli gürültü için farklı nesil sayılarında GA verimlerinin karşılaştırılması

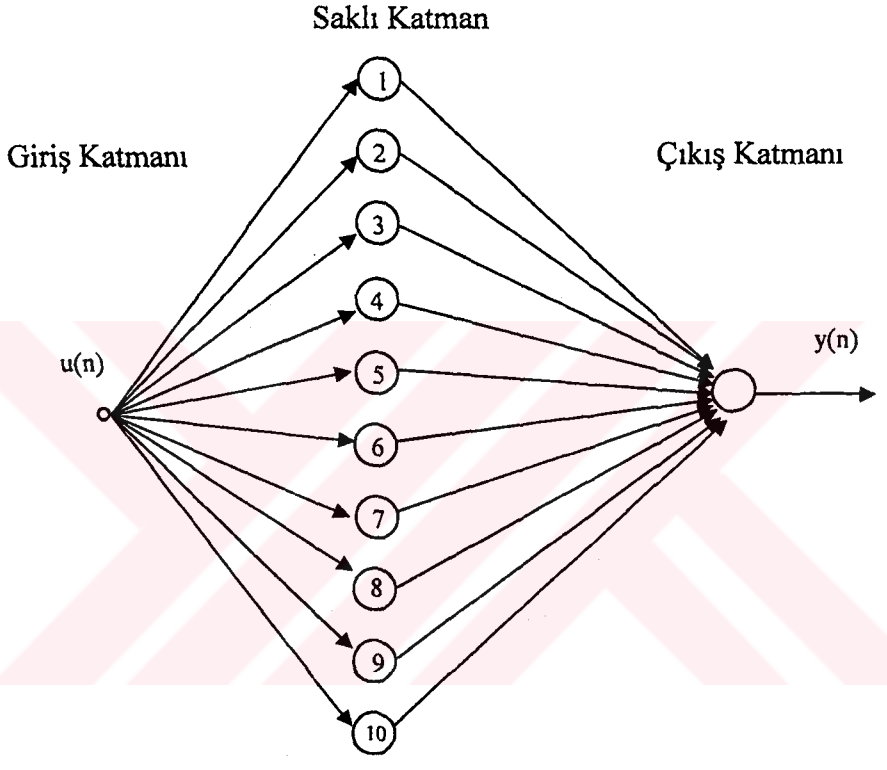


Şekil 3.18: Farklı nesil sayıları için GA ortalama işlem sürelerinin karşılaştırılması

Tez çalışmamızın temel amacı olan gerçek zamanlı ses işleme uygulaması göz önüne alındığında algoritmaların verimleri kadar işlem süreleri de önem taşımaktadır. Bu durum göz önüne alınarak tüm algoritmaların birlikte karşılaştırıldığı çalışmada GA için 100 nesil kullanılması kararlaştırılmıştır.

3.6.4. YSA uygulaması

Gürültü bindirilmiş veritabanı konuşma verileri, yapısı Şekil 3.19'da gösterilen tek gizli katmanda hiperbolik tanjant transfer fonksiyonuna sahip 10 işlem birimi ve çıkış katmanında doğrusal tek işlem birimi olan ileri beslemeli ÇKA kullanılarak eğitilmiştir.

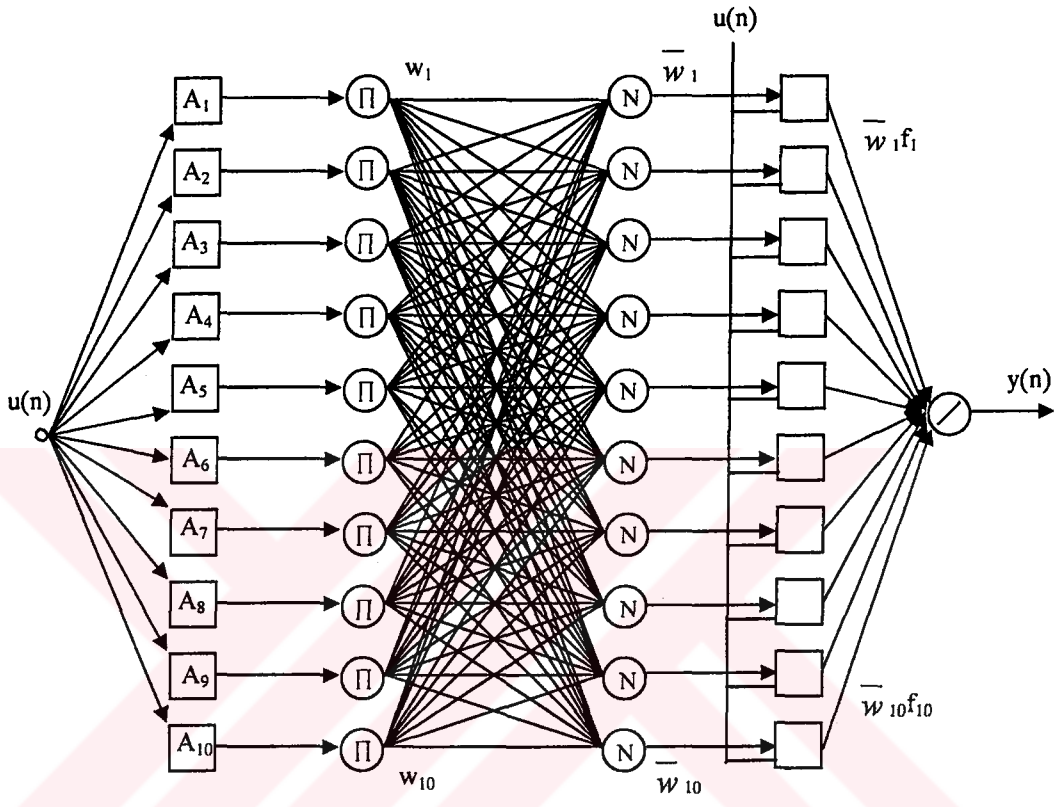


Şekil 3.19: Bir girişli, gizli katmanda on işlem birimi ve bir çıkışa sahip ÇKA yapısı

ÇKA eğitiminde Levenberg-Marquard geriye yansıtma algoritması kullanılmıştır (Hagan and Menjah, 1994). Eğitim işleminin başlangıcında ağın tüm ağırlıkları Nguyen-Widrow algoritmasına göre belirlenmiştir (Nguyen and Widrow, 1990). Çalışmalarda ağın eğitimi 500, 750, 1000 ve 2000 kez tekrarlanarak gerçekleştirilmiştir. MSE değerleri ve süreler dikkate alındığında en iyi verimin 1000 kez tekrarlamada elde edildiği görülmüştür ve tüm algoritmaların birlikte karşılaştırıldığı çalışmada YSA için 1000 kez tekrarlanan sonuçların kullanılması kararlaştırılmıştır.

3.6.5. ANFIS uygulaması

Gürültü bindirilmiş veritabanı konuşma verileri, yapısı Şekil 3.20'da gösterilen bir ANFIS kullanılarak gürültüden temizlenmeye çalışılmıştır.



Şekil 3.20: Bir girişli, on kurallı 'Sugeno Bulanık' modeli eşdeğer ANFIS yapısı

Tablo 3.3'de kullanılan ANFIS yapısının özellikleri verilmiştir (Jang, 1993; Jang and Sun, 1995; Jang and Gulley, 1995; Jantzen, 1998).

Çalışmalarda ağın eğitimi, eğitim verisi için 100, 250, 500, 750 ve 1000 kez tekrarlanarak gerçekleştirilmiştir. MSE değerleri ve süreler dikkate alındığında en iyi verimin 500 kez tekrarlamada elde edildiği görülmüştür ve tüm algoritmaların birlikte karşılaştırıldığı çalışmada ANFIS için 500 kez tekrarlanan sonuçların kullanılması kararlaştırılmıştır.

Tablo 3.3: ANFIS yapısının özellikleri

Giriş sayısı	1
Çıkış sayısı	1
Kural sayısı	10
Bulanık Mantık çıkartım sistemi	Sugeno
andMethod	prod
orMethod	max
defuzzMethod	wtaver
impMethod	prod
aggMethod	max

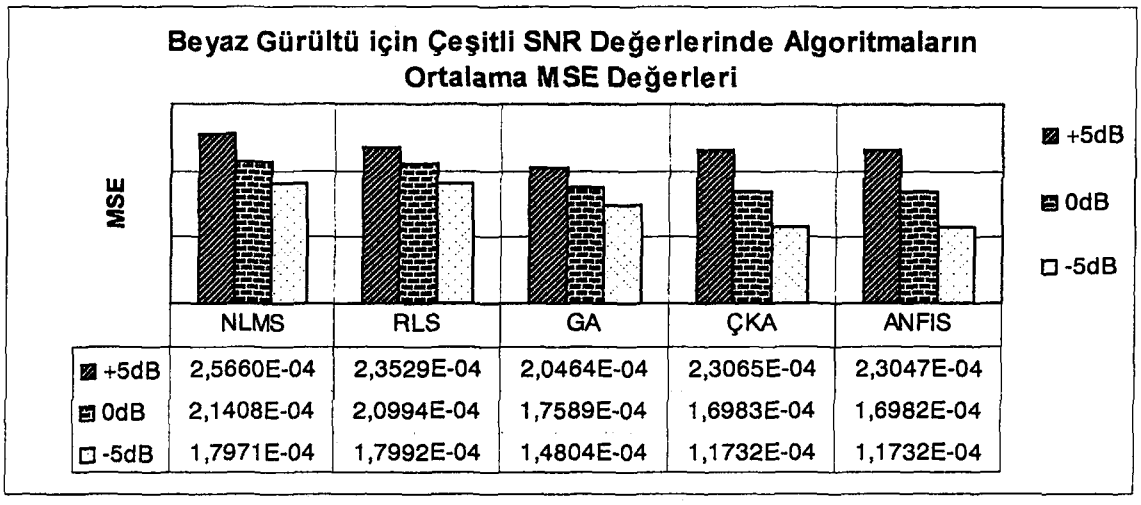
ANFIS, YSA ve bulanık mantık özelliklerini kullanılarak geliştirilmiş bir yapı olduğu için verim açısından YSA çalışmasından daha iyi sonuçlar vermesi umulmaktadır. Bu durum tüm çalışmalar sonucu elde edilen Tablo 3.4'de açık bir şekilde görülmektedir. Tablo 3.5'de kullanılan algoritmalar işletim süreleri açısından incelendiğinde 500 kez tekrarlamalı olarak gerçekleştirilen ANFIS ve 1000 kez tekrarlamalı olarak gerçekleştirilen YSA çalışmalarının yaklaşık eşit sürelerde tamamlandığı görülmektedir. ANFIS yapısının YSA yapısına göre karmaşık olması nedeniyle işlem süreleri açısından beklenen yavaş çalışma verimi Tablo 3.5'de açıkça görülmektedir. Sonuç olarak 500 kez tekrarlamalı ANFIS yapısı çok küçük bir farkla da olsa iki kat fazla tekrarlamalı YSA yapısının gerçekleştirdiği verimden daha iyi sonuç vermiştir. Fakat süre açısından YSA daha iyi sonuç vermektedir.

3.6.6. Algoritmaların karşılaştırılması

Aynı işaretler üzerinde algoritmaların verimlerinin incelenmesi amacıyla yapılan çalışmanın MSE sonuçları Tablo 3.4'te ve işlem süreleri sonuçları Tablo 3.5'te verilmiştir.

Oluşturulan veritabanı üzerinde tüm algoritmaların aynı işaretler için verimlerinin karşılaştırılması amacıyla yapılan çalışmalar sonucunda her bir SNR değeri (+5dB, 0dB, -5dB) için on konuşmacının MSE değerleri hesaplanmıştır. MSE değerlerinin ortalaması alınarak, beyaz gürültü için NLMS, RLS, GA, ÇKA ve ANFIS algoritmalarının ortalama MSE değerlerinin karşılaştırıldığı grafik Şekil 3.21'de,

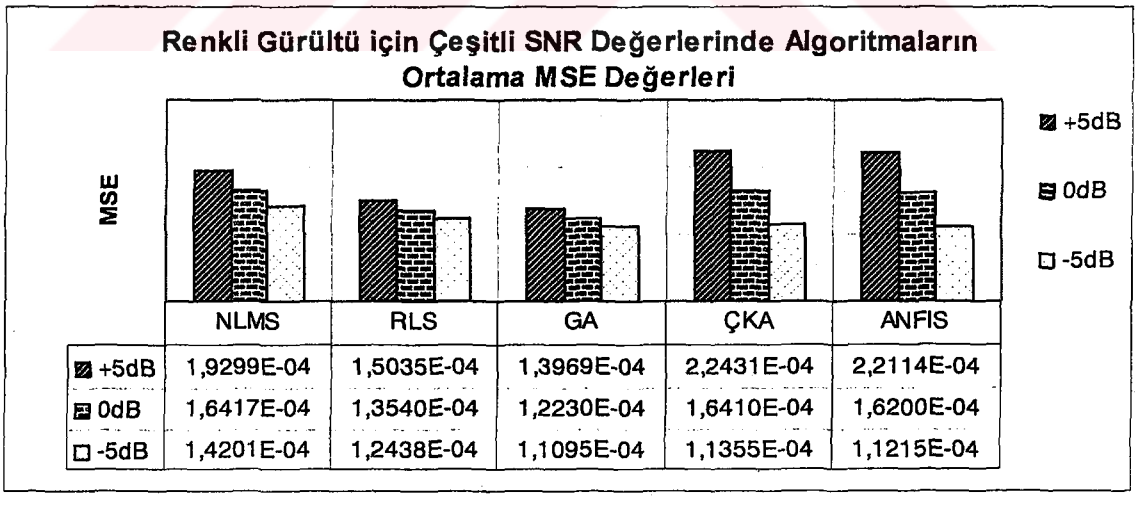
renkli gürültü için NLMS, RLS, GA, ÇKA ve ANFIS algoritmalarının ortalama MSE değerlerinin karşılaştırıldığı grafik Şekil 3.22'de verilmiştir.



	NLMS	RLS	GA	ÇKA	ANFIS
+5dB	2,5660E-04	2,3529E-04	2,0464E-04	2,3065E-04	2,3047E-04
0dB	2,1408E-04	2,0994E-04	1,7589E-04	1,6983E-04	1,6982E-04
-5dB	1,7971E-04	1,7992E-04	1,4804E-04	1,1732E-04	1,1732E-04

Şekil 3.21: Beyaz gürültü için NLMS, RLS, GA, ÇKA ve ANFIS algoritmalarının karşılaştırılması

Şekil 3.21 incelendiğinde beyaz gürültü +5dB SNR için en düşük hata değerlerinin GA tarafından elde edildiği görülmektedir. Beyaz gürültü 0dB ve -5dB SNR için en düşük hata değerleri ise ANFIS ve çok katmanlı YSA tarafından elde edilmiştir.

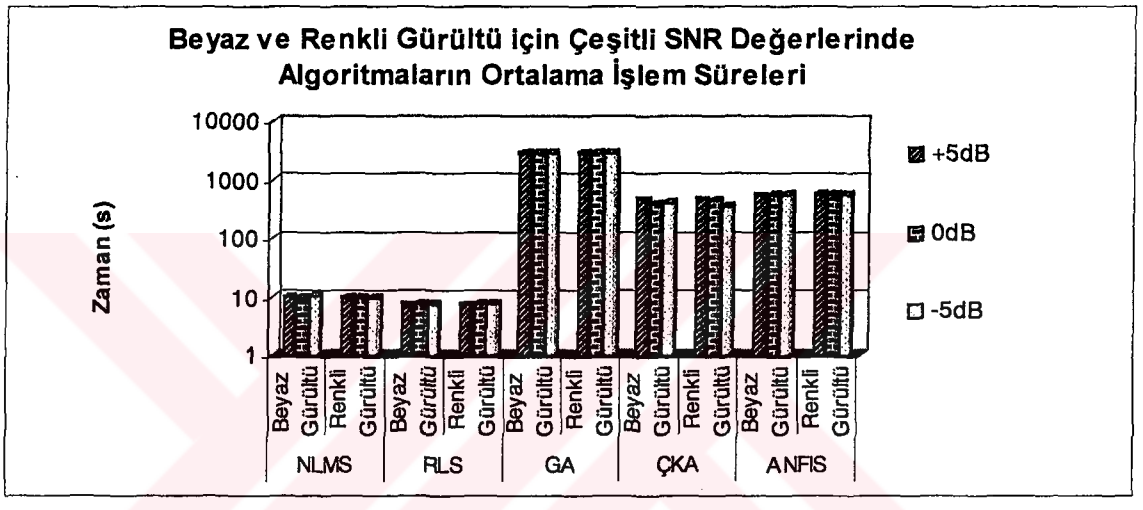


	NLMS	RLS	GA	ÇKA	ANFIS
+5dB	1,9299E-04	1,5035E-04	1,3969E-04	2,2431E-04	2,2114E-04
0dB	1,6417E-04	1,3540E-04	1,2230E-04	1,6410E-04	1,6200E-04
-5dB	1,4201E-04	1,2438E-04	1,1095E-04	1,1355E-04	1,1215E-04

Şekil 3.22: Renkli gürültü için NLMS, RLS, GA, ÇKA ve ANFIS algoritmalarının karşılaştırılması

Şekil 3.22 incelendiğinde ise renkli gürültü için tüm SNR değerlerinde en düşük ortalama hataya sahip algoritmanın GA olduğu görülmektedir.

Şekil 3.23'de beyaz ve renkli gürültü için NLMS, RLS, GA, çok katmanlı YSA ve ANFIS algoritmalarının çeşitli SNR değerleri için ortalama işlem sürelerinin karşılaştırıldığı grafik verilmiştir. Algoritmaların gerçekleşme süreleri arasında çok büyük farklar olduğu için Şekil 3.23'de zaman ekseni logaritmik artan bir eksen olarak verilmiştir.



Şekil 3.23: NLMS, RLS, GA, ÇKA ve ANFIS algoritmalarının ortalama işlem sürelerinin karşılaştırılması

İşlem süreleri açısından incelendiğinde her iki gürültü türü içinde en hızlı algoritmanın RLS olduğu Şekil 3.23'de görülmektedir. Tez çalışmamızda amaçlanan araç içi gürültüsü gibi renkli gürültülerin bastırılması işlemini en küçük hata değerleriyle gerçekleştiren GA, işlem süreleri acısından incelendiğinde en yavaş sonuçlanan algoritmadır ve gerçek zamanlı uygulamalarda kullanılması mümkün değildir. Gerçek zamanlı uygulamalar için işlem süreleri ve hata değerleri göz önüne alındığında en uygun gürültü bastırma algoritmasının RLS olduğu görülmüştür.

Tablo 3.4: NLMS, RLS, GA, ÇKA ve ANFIS algoritmalarının MSE sonuçları

Konuşmacı Etiketleri	Gürültü Oranları (SNR)	NLMS		RLS		GA		ÇKA		ANFIS	
		Beyaz Gürültü	Renkli Gürültü	Beyaz Gürültü	Renkli Gürültü	Beyaz Gürültü	Renkli Gürültü	Beyaz Gürültü	Renkli Gürültü	Beyaz Gürültü	Renkli Gürültü
F1 (2,01 s.)	+5 dB	2,2340E-05	2,1122E-05	2,1294E-05	1,3983E-05	1,8731E-05	1,5606E-05	2,1844E-05	2,1548E-05	2,1840E-05	2,1051E-05
	0 dB	1,9107E-05	1,6694E-05	1,7362E-05	1,1680E-05	1,5021E-05	1,3071E-05	1,6337E-05	1,6043E-05	1,6338E-05	1,5698E-05
	-5 dB	1,4854E-05	1,3115E-05	1,4251E-05	9,7325E-06	1,2156E-05	1,0849E-05	1,1432E-05	1,1317E-05	1,1434E-05	1,1074E-05
F2 (1,89 s.)	+5 dB	1,1344E-03	8,1153E-04	1,0026E-03	5,5833E-04	8,8881E-04	5,5965E-04	1,1299E-03	1,0800E-03	1,1291E-03	1,0760E-03
	0 dB	9,2445E-04	6,5630E-04	8,6961E-04	4,6895E-04	7,5459E-04	4,5540E-04	8,4053E-04	7,9570E-04	8,4050E-04	7,9524E-04
	-5 dB	7,3129E-04	5,2183E-04	6,8586E-04	3,9505E-04	5,9997E-04	4,0409E-04	5,8336E-04	5,6027E-04	5,8336E-04	5,5874E-04
F3 (1,31 s.)	+5 dB	2,5680E-05	2,2516E-05	2,2887E-05	1,5773E-05	1,9111E-05	1,4242E-05	1,8773E-05	1,7612E-05	1,8772E-05	1,7403E-05
	0 dB	2,0911E-05	1,8666E-05	1,9720E-05	1,4188E-05	1,6084E-05	1,3257E-05	1,3779E-05	1,3131E-05	1,3768E-05	1,3022E-05
	-5 dB	1,7615E-05	1,5483E-05	1,7345E-05	1,2745E-05	1,4329E-05	1,1568E-05	9,5341E-06	9,2881E-06	9,5470E-06	9,1888E-06
F4 (1,36 s.)	+5 dB	2,6063E-05	1,9604E-05	2,3718E-05	1,4652E-05	2,1536E-05	1,5228E-05	2,1815E-05	2,1013E-05	2,1783E-05	2,0447E-05
	0 dB	2,2583E-05	1,6563E-05	2,0855E-05	1,3080E-05	1,8204E-05	1,3297E-05	1,5551E-05	1,5413E-05	1,5540E-05	1,4992E-05
	-5 dB	1,8853E-05	1,4629E-05	1,8542E-05	1,2031E-05	1,5736E-05	1,2680E-05	1,0671E-05	1,0658E-05	1,0660E-05	1,0376E-05
F5 (4,09 s.)	+5 dB	4,7504E-05	3,7172E-05	4,6354E-05	2,8618E-05	3,8509E-05	2,8909E-05	2,9171E-05	2,9550E-05	2,9130E-05	2,8015E-05
	0 dB	4,1766E-05	3,4950E-05	4,0924E-05	2,7010E-05	3,2912E-05	2,7469E-05	2,0366E-05	2,0624E-05	2,0359E-05	1,9572E-05
	-5 dB	3,8797E-05	3,3073E-05	3,8004E-05	2,5861E-05	3,1348E-05	2,7234E-05	1,3830E-05	1,3935E-05	1,3846E-05	1,3221E-05
M1 (1,92 s.)	+5 dB	3,5499E-04	2,8576E-04	3,4231E-04	2,4686E-04	2,9053E-04	2,1347E-04	3,2164E-04	3,3153E-04	3,2087E-04	3,2926E-04
	0 dB	3,0965E-04	2,5225E-04	3,1083E-04	2,2752E-04	2,4782E-04	1,8990E-04	2,3561E-04	2,4084E-04	2,3519E-04	2,3908E-04
	-5 dB	2,7513E-04	2,2769E-04	2,8744E-04	2,1440E-04	2,2047E-04	1,7093E-04	1,6378E-04	1,6772E-04	1,6374E-04	1,6644E-04
M2 (1,66 s.)	+5 dB	5,4010E-05	3,7554E-05	4,8480E-05	2,8767E-05	4,3067E-05	3,0208E-05	5,3513E-05	5,5558E-05	5,3622E-05	5,2707E-05
	0 dB	4,5003E-05	3,0901E-05	4,2282E-05	2,5121E-05	3,7933E-05	2,6516E-05	3,9708E-05	3,8949E-05	3,9728E-05	3,6951E-05
	-5 dB	3,7269E-05	2,6968E-05	3,5495E-05	2,2899E-05	3,0171E-05	2,3838E-05	2,6660E-05	2,5652E-05	2,6639E-05	2,4484E-05
M3 (2,03 s.)	+5 dB	1,4717E-05	1,1723E-05	1,4022E-05	9,0261E-06	1,2883E-05	8,4638E-06	1,7567E-05	1,7266E-05	1,7543E-05	1,7077E-05
	0 dB	1,3608E-05	9,9033E-06	1,1952E-05	7,9375E-06	1,0581E-05	7,5563E-06	1,2438E-05	1,2192E-05	1,2439E-05	1,2139E-05
	-5 dB	1,1593E-05	8,5863E-06	1,0390E-05	7,0525E-06	9,1107E-06	6,9621E-06	8,1437E-06	7,9363E-06	8,1455E-06	7,8818E-06
M4 (1,62 s.)	+5 dB	1,3677E-04	1,0356E-04	1,3018E-04	9,6123E-05	1,0805E-04	7,3271E-05	9,0828E-05	8,9874E-05	9,0729E-05	8,6353E-05
	0 dB	1,1357E-04	9,4498E-05	1,1700E-04	9,3007E-05	9,5425E-05	6,6780E-05	6,5925E-05	6,6454E-05	6,5796E-05	6,3733E-05
	-5 dB	9,8940E-05	8,8025E-05	1,0661E-04	9,1284E-05	8,2265E-05	7,1194E-05	4,4066E-05	4,5172E-05	4,4031E-05	4,3508E-05
M5 (1,36 s.)	+5 dB	7,4955E-04	5,7935E-04	7,0109E-04	4,9140E-04	6,0516E-04	4,3783E-04	6,0141E-04	5,7914E-04	6,0133E-04	5,6310E-04
	0 dB	6,3018E-04	5,1102E-04	6,4888E-04	4,6552E-04	5,3036E-04	4,0971E-04	4,3801E-04	4,2161E-04	4,3850E-04	4,0953E-04
	-5 dB	5,5279E-04	4,7072E-04	5,8528E-04	4,5272E-04	4,6483E-04	3,7011E-04	3,0171E-04	2,8357E-04	3,0180E-04	2,7654E-04

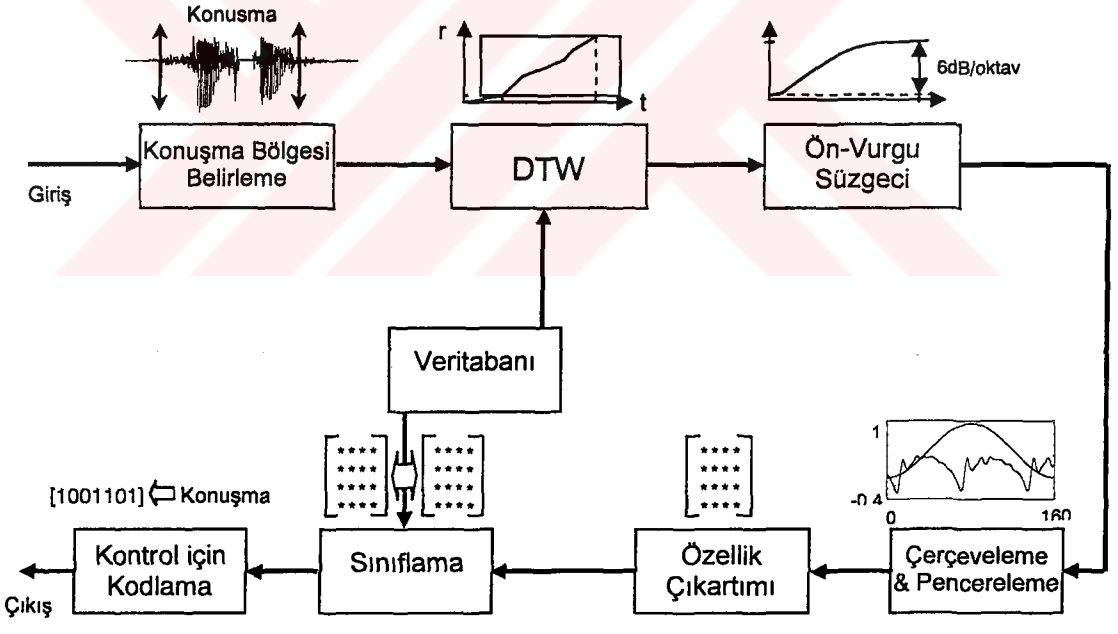
Tablo 3.5: NLMS, RLS, GA, ÇKA ve ANFIS algoritmalarının işlem süreleri (saniye)

Konuşmacı Etiketleri	Gürültü Oranları (SNR)	NLMS		RLS		GA		ÇKA		ANFIS	
		Beyaz Gürültü	Renkli Gürültü	Beyaz Gürültü	Renkli Gürültü	Beyaz Gürültü	Renkli Gürültü	Beyaz Gürültü	Renkli Gürültü	Beyaz Gürültü	Renkli Gürültü
F1 (2,01 s.)	+5 dB	8,09	7,81	6,80	5,48	3073,70	2978,20	483,44	485,66	560,91	568,20
	0 dB	7,86	6,69	6,56	5,91	3070,50	2979,00	523,59	492,47	571,31	565,67
	-5 dB	7,95	7,92	5,52	6,91	3072,30	2979,00	480,28	176,19	568,75	564,89
F2 (1,89 s.)	+5 dB	7,03	7,91	6,28	6,27	2886,00	2885,90	457,91	516,81	533,61	530,39
	0 dB	7,20	7,36	5,45	6,53	2886,00	2884,50	459,16	464,78	532,19	529,75
	-5 dB	7,19	5,63	5,33	6,88	2885,70	2884,60	460,09	464,49	531,56	530,58
F3 (1,31 s.)	+5 dB	3,47	3,44	2,75	2,56	1954,60	1953,90	319,67	318,22	369,27	375,20
	0 dB	2,38	3,03	2,66	2,99	1953,50	1954,60	319,72	319,31	369,95	372,44
	-5 dB	3,41	3,02	3,17	1,95	1952,60	1953,00	314,75	318,39	369,70	370,70
F4 (1,36 s.)	+5 dB	3,20	4,36	2,81	2,56	2047,70	2046,40	330,61	339,13	382,64	385,78
	0 dB	3,16	2,61	2,73	2,24	2047,30	2047,50	329,59	324,91	389,03	383,78
	-5 dB	3,75	3,58	3,36	2,91	2045,60	2047,60	65,73	327,39	384,63	382,17
F5 (4,09 s.)	+5 dB	46,70	42,25	31,89	33,89	6169,20	5980,80	935,34	914,52	1153,20	1150,90
	0 dB	45,67	48,06	36,69	33,36	6170,10	5980,70	166,05	915,78	1162,10	1249,10
	-5 dB	51,69	46,77	36,03	32,97	6163,70	5980,30	953,45	325,94	1150,80	1156,70
M1 (1,92 s.)	+5 dB	7,25	7,36	5,69	6,08	2886,10	2910,10	449,25	467,83	541,86	543,63
	0 dB	8,80	8,41	6,24	6,25	2886,10	2891,30	469,89	452,97	539,97	540,19
	-5 dB	8,55	8,50	6,42	6,25	2886,00	2889,50	113,22	453,69	547,53	541,41
M2 (1,66 s.)	+5 dB	6,16	5,03	4,36	3,45	2512,60	2419,20	414,05	383,36	469,22	472,20
	0 dB	5,59	5,19	4,03	3,55	2513,50	2420,60	396,94	385,09	479,28	467,53
	-5 dB	4,94	4,77	3,64	3,72	2513,50	2420,70	396,63	388,84	469,50	481,97
M3 (2,03 s.)	+5 dB	7,63	7,56	5,84	6,42	3072,60	3072,40	487,33	478,94	583,17	581,94
	0 dB	7,92	7,59	5,88	6,45	3073,30	3072,40	489,45	481,17	580,97	573,51
	-5 dB	7,83	7,61	6,42	6,50	3072,40	3073,10	501,97	313,97	575,91	579,09
M4 (1,62 s.)	+5 dB	5,58	5,31	4,64	3,48	2421,20	2428,30	384,08	377,73	459,42	461,23
	0 dB	5,38	5,41	4,11	4,31	2428,70	2430,30	384,44	379,11	462,11	463,99
	-5 dB	5,61	4,92	4,72	3,69	2429,80	2428,00	383,88	378,13	457,01	462,26
M5 (1,36 s.)	+5 dB	3,38	3,19	2,28	2,89	2051,70	2047,40	346,49	322,03	384,92	387,97
	0 dB	3,31	3,50	3,06	2,58	2062,70	2048,80	348,48	317,84	384,94	386,22
	-5 dB	2,95	3,45	3,14	2,66	2046,50	2053,60	332,56	324,48	385,78	390,97

BÖLÜM 4. ARAÇ İÇİ CİHAZLARIN SESLE KONTROLÜ

4.1. Sesle Kontrol İşleminin Yapısı

Çalışmamızda araç içindeki cihazların ve bileşenlerin sesle kontrolünü sağlamak amaçlanmaktadır. Konuşma tanıma yöntemi olarak ayrık kelime tanıma yöntemi kullanılmıştır. Bu yöntemin başarılı bir şekilde çalışmasının temel koşulu konuşma bölgesinin çok iyi şekilde belirlenmesi ve dolayısıyla DTW işleminin en iyi biçimde yapılmasıdır (Rabiner et al, 1978; Myers et al, 1980; Furui, 1986). Hazırlanan sistemin blok diyagramı Şekil 4.1'de verilmiştir.



Şekil 4.1: Konuşma tanıma işleminin blokları

Bir mikrofon aracılığıyla alınan konuşma işareti, Dijital-Analog Çevirici (Digital Analog Converter-DAC) ile ayrık zamanlı işaret haline getirilmiş ve 8 KHz örnekleme frekansı ile örneklenerek Şekil 4.1'de konuşma tanıma sisteminin girişine

uygulanmıştır. Sistemin blokları sırasıyla aşağıda açıklanan işlemleri gerçekleştirirler:

- Konuşma Bölgesini Belirleme: Girişten gelen konuşma işaretinin sınırlarını, işaretin enerji seviyesine ve sıfır geçiş sayısına bakarak belirler. Bu sınırlar dışında kalan sessiz bölge atılarak bloğun çıkışında konuşma işareti yalın olarak elde edilmiş olur.

- Dinamik Zaman Eğreltimi: Sessiz bölgeden ayrıştırılış konuşma işareti veri tabanındaki kelimeler ile iki işaretin zaman düzleminde boyutlarının eşitlenmesi işlemi olan dinamik zaman eğreltimi işlemi uygulanır.

- Ön-Vurgu Süzgeci: Konuşma üretme sisteminde sürtünme, ısıl kayıplar ve dudaklardaki saçılma kaybı gibi nedenlerden bir miktar enerji tüketilir. Yüksek frekanslarda meydana gelen bu kayıpları karşılamak, işareti spektral olarak düzleştirmek ve daha sonra işaret işleme içindeki etkilere karşı hassasiyetini azaltmak için işaretin birinci dereceden bir FIR süzgeçten geçirilmesi işlemidir. Ön-Vurgu süzgeci olarak genellikle birinci dereceden bir süzgeç sistemi kullanılır (Rabiner and Juang, 1993).

$$H(z) = 1 - az^{-1}, \qquad 0.9 \le a \le 1 \tag{4.1}$$

Bu durumda Ön-Vurgu süzgecinin çıkışı, giriş işaretine bağlıdır ve aşağıdaki eşitlikle elde edilir.

$$\hat{s}(n) = s(n) - as(n-1), \qquad 0.9 \le a \le 1 \tag{4.2}$$

a değeri sıklıkla 0.95 civarında seçilir. Çalışmamızda a değeri 0.95 olarak kullanılmıştır.

- Çerçeveleme ve Pencereleme: Ön-Vurgu süzgecinden geçirilen ve 20 ms 'lik parçalara bölünerek çerçevelenen konuşma işareti Hanning pencere ile çarpılarak pencerelenir. 20 ms 'lik parçalara bölünerek pencerelenen konuşma işareti, pencere geçişlerinde meydana gelebilecek bilgi kayıplarını önlemek amacıyla her pencere bir diğerinin içine geçecek şekilde 10 ms 'lik aralıklarla başlatılır.

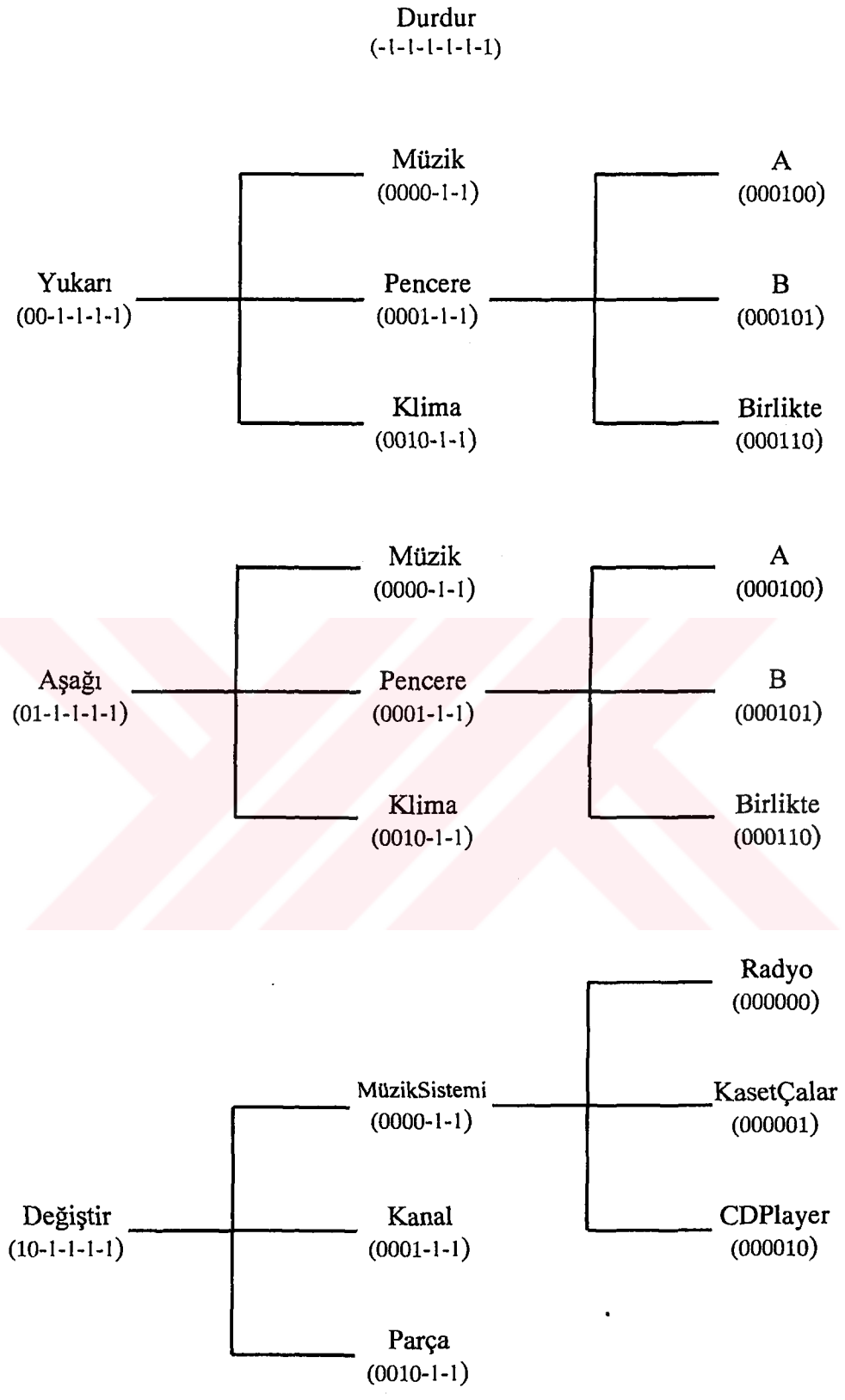
- Özellik Çıkartımı: Sistemde kullanılacak özellik çıkartımı yöntemini belirleme aşamasında kepstral sayıların elde edilmesi için üç farklı yöntem kullanılmıştır. Bu yöntemlerden ikisi DÖK analizi yapan; Özilinti Analizi (Autocorrelation Analysis) ve Ortak Değişinti Metodu (Covariance Method), üçüncüsü ise Mel Frekans Ölçekli Kepstral Katsayılar (MFCC) Yöntemidir. Bu yöntemler içerisinden tez çalışmamızda amaçlanan gerçek zamanlı uygulama için en uygun yöntem belirlenmeye çalışılmıştır.

- Sınıflandırma: Konuşma işareti özellik matrisi, veritabanında tutulan muhtemel aynı konuşma işareti özellik matrisinin benzerliği ile Euclidean uzaklığı hesaplanarak bulunur.

- Kontrol için Kodlama: Giriş konuşma işaretine en yakın veritabanı konuşma işareti belirlendikten sonra araç içi cihazların veya bileşenlerinin kontrolü için kullanılacak kodlar üretilir.

4.2. Veritabanının Oluşturulması

Veritabanının oluşturulmasına, çalışmada kullanılacak kelimelerin belirlenmesiyle başlanır. Konuşma tanıma işleminde araç içi cihazların ve bileşenlerin kontrolü amacına uygun kelimeler seçilmiştir. Çalışmamızda araç bileşenlerinden müzik sisteminin, klimanın ve pencerenin sesle kontrolü amaçlanmaktadır. Komut listesi Şekil 4.2'de verilen ağaç yapısı şeklindedir. Sistem, ses komutlarına göre içeriği değiştirilecek bir kontrol kod kümesine sahiptir. Bu kod kümesinin içeriği başlangıç değeri olarak altı adet -1'den oluşmaktadır. Şekil 4.2'de görüldüğü gibi ağaç yapısı üç bloktan oluşmaktadır ve her blok kontrol kod kümesinin iki değerini değiştirmektedir. Sistem başlangıç aşamasında kullanıcıdan ağaç yapısının ilk bloğunu oluşturan "yukarı", "aşağı", "değiştir" kelimelerinden birini beklemektedir. Eğer bu aşamada söylenen telaffuzun özellik matrisi bu kelimelerden herhangi birinin özellik matrisine belirlenen eşik değerinden daha yakınsa, kontrol kod kümesinin içeriği Şekil 4.2'de kelimelerin altında verilen ve kelimeye karşılık gelen kodlar olarak değiştirilecektir. İlk telaffuz beklenen kelimelerden biri ise ağaç yapısı ilgili kelimenin dallarında gösterilen kelimelerden birini bekleyecektir.



Şekil 4.2: Komut listesinin ağaç yapısı

Örneğin "yukarı" kelimesinden sonra kontrol kod kümesinin içeriği '00-1-1-1-1' şeklinde değiştirilecek ve ikinci telaffuz olarak "müzik", "pencere", "klima" kelimelerinden biri beklenecektir. Eğer ikinci telaffuz bu kelimelerden herhangi biri ile eşleştirilirse kontrol kod içeriğinin üç ve dördüncü değerleri kelimeye karşılık gelen kodlar olarak değiştirilecektir. İkinci telaffuz bu kelimelerden hiç biri ile eşleştirilemez ise kontrol kod kümesinin içeriği başlangıç değeri olan '-1-1-1-1-1-1' şekline döner. Şekil 4.2'de görüldüğü gibi üçüncü dal bloğuna sahip iki komut vardır. Bu komutlar birden fazla birleşene sahip "pencere" ve "müziksistemi" komutlarıdır. Üçüncü blok için de ilk iki bloktaki çalışma yapısı geçerlidir.

Sistem ağaç yapısının her aşamasında söylenen telaffuz ile "durdur" kelimesinin telaffuzunu kontrol etmektedir. Eğer telaffuz edilen kelime "durdur" kelimesi ile eşleştirilirse kontrol kod kümesi içeriği '-1-1-1-1-1-1' olarak değiştirilir ve sistem başlangıç durumuna getirilmiş olur. Böylelikle herhangi bir aşamada sistemin yapabileceği yanlış iptal edilebilir veya telaffuz edilecek yanlış bir komuttan geri dönüş sağlanabilir.

Tablo 4.1: Sistemde kullanılan komutların listesi ve komutların işlevleri

Komut Listesi	Kontrol Kod Kümesi	Komuta Karşılık Gelen Görev
Durdur	-1-1-1-1-1-1	Sistemi başlangıç durumuna getirir
Yukarı → müzik	0000-1-1	Müzik sesini arttırır
Yukarı → pencere → A	000100	Sol pencereyi kapatır
Yukarı → pencere → B	000101	Sağ pencereyi kapatır
Yukarı → pencere → birlikte	000110	İki pencereyi birlikte kapatır
Yukarı → klima	0010-1-1	Klimanın derecesini arttırır
Aşağı → müzik	0100-1-1	Müzik sesini azaltır
Aşağı → pencere → A	010100	Sol pencereyi açar
Aşağı → pencere → B	010101	Sağ pencereyi açar
Aşağı → pencere → birlikte	010110	İki pencereyi birlikte açar
Aşağı → klima	0110-1-1	Klimanın derecesini azaltır
Değiştir → müziksistemi → radyo	100000	Müzik sisteminde radyoyu çalar
Değiştir → müziksistemi → kasetcalar	100001	Müzik sisteminde kaseti çalar
Değiştir → müziksistemi → cdplayer	100010	Müzik sisteminde CD'yi çalar
Değiştir → kanal	1001-1-1	Radyo kanalı değiştirilir
Değiştir → parça	1010-1-1	Çalan parça değiştirilir (kaset veya CD)

Komutlar, yapacağı işler ve bu işlere karşılık gelen kontrol kod kümesi içerikleri Tablo 4.1'de verilmiştir. Pencere kontrolünü sağlayan komutlarda "sol" ve "sağ" kelimelerinin telaffuzları dolayısıyla özellik matrisleri birbirine çok yakın olduğu

için "sol" kelimesinin yerine "A", "sağ" kelimesinin yerine de "B" telaffuz edilmesi uyun görülmüştür. Sisteme Tablo 4.1'in son satırındaki komutlar; sırasıyla "değiştir" ve "parça" telaffuz edildiğinde, sistemde o anda kasetçalar veya CD oynatıcıdan hangisi aktif ise çalmakta olan parçayı değiştirerek sonraki parçaya geçilir.

Sistemde kullanılacak kelimelerin belirlendikten sonra veritabanı oluşturma çalışmalarına geçilmiştir. Veritabanının oluşturulması için 10 erkek ve 10 bayandan beşer kez belirlenen kelimeleri telaffuz edilmesi istenmiştir. Kelime telaffuzlarından üçü eğitim ve ikisi test amaçlı kullanılmak üzere veritabanına kaydedilmiştir. Kişilerin eğitim amaçlı kullanılmak üzere kaydedilen ilk üç telaffuzları her bir kelimenin kişiden bağımsız ortak özellik matrislerini çıkarmada kullanılmıştır. Kişilerin test amaçlı kaydedilen telaffuzları ise sistemin verimini ölçmede kullanılmıştır.

4.3. Özellik Çıkartım Yönteminin Belirlenmesi

DÖK analizi konuşma kodlama, konuşma işaretinden konuşma veya konuşmacı belirleyici parametrelerin çıkartılması gibi birçok alanda kullanılır. Ancak çok iyi bir tanıma verimi sağlamaz. DÖK analizi sonucu elde edilen değerler kepstral katsayılara dönüştürülerek tanıma verimleri arttırılır ve konuşma-konuşmacı tanıma sistemlerinde kullanılabilir (İnal et al, 2000). MFCC yönteminin insan kulağının algılama özelliğini temel alarak çalışması, MFCC 'nin konuşma tanıma sistemi için çok etkili bir yöntem olmasını sağlar. Yapılan deneysel çalışmalar konuşma ve konuşmacı tanıma sistemlerinde MFCC 'nin çok başarılı olduğunu göstermiştir (Koç, 2002).

Gerçek zamanlı uygulamalarda kullanılan sistemin verimi kadar işlemlerin gerçekleşme süreleri de önemlidir. İşlem süreleri açısından yöntemler, oluşturulan veritabanından 3 bayan ve 3 erkek toplam 6 konuşmacının "pencere" telaffuzları alınarak her yöntemin aynı telaffuz için özellik matrislerini üretme süreleri incelenerek karşılaştırılmıştır. Özilinti Analizi-kepstral katsayı dönüşümü, Ortak Değişinti Metodu-kepstral katsayı dönüşümü ve MFCC Yönteminin aynı telaffuzlar

için işlem süreleri Tablo 4.2'de verilmiştir. Tablo 4.2 aynı telaffuz için yöntemlerin işlem süreleri incelendiğinde MFCC yönteminin en hızlı yöntem olduğu görülmektedir.

Tablo 4.2: Yöntemlerin işlem süreleri (saniye)

Konuşmacı Etiketleri	Özilinti (Autocorrelation)	Ortak Değişinti (Covariance)	MFCC
F1 (0,61 s.)	1,062	0,060	0,040
F2 (0,65 s.)	1,101	0,070	0,060
F3 (0,69 s.)	1,231	0,070	0,040
M1 (0,59 s.)	1,112	0,070	0,060
M2 (0,58 s.)	0,981	0,061	0,060
M3 (0,56 s.)	1,011	0,060	0,060
Ortalama	1,083	0,065	0,053

Karşılaştırılan yöntemler içerisinde konuşma ve konuşmacı tanıma verimi açısından en iyi yöntem olduğu bilinen MFCC yöntemi, işlem süreleri acısından karşılaştırdığında en hızlı yöntemdir. Tüm bu karşılaştırmalar sonucunda, tez çalışmamızda amaçlanan gerçek zamanlı konuşma tanıma sisteminde özellik çıkartım yöntemi olarak MFCC yönteminin kullanılması kararlaştırılmıştır.

4.4. Konuşma Tanıma Sisteminin Verimi

Tez çalışmamızda gerçekleştirilen araç içindeki cihazların sesle kontrolüne yönelik konuşma tanıma sisteminin temiz konuşma işareti için veriminin ölçümünde aşağıdaki işlemler gerçekleştirilmiştir:

- 10 erkek konuşmacıdan alınan her bir komut telaffuzunun ilk üç kaydı komuta ait özellik matrisi oluşturmada kullanılmıştır. Her bir komut için on kişiden alınan üçer telaffuz olmak üzere toplam otuz telaffuzun ortalama uzunluğu bulunmuştur ve ortalamaya en yakın uzunluktaki telaffuz referans olarak kabul edilmiştir. Daha sonra her bir komut için kayıtlı bulunan otuz telaffuz DTW işlemi ile eşit boyuta getirilmiştir. MFCC yöntemi kullanılarak özellik matrisleri çıkarılmıştır ve elde edilen eşit boyutlu matrislerin ortalaması alınarak komuta ait kişiden bağımsız ortalama özellik matrisi elde edilmiştir. Her bir komut için bu işlem

tekrar edilerek erkeklerin telaffuzları için ortalama özellik matrisler klasörleri oluşturulmuştur.

- Yukarıda erkek konuşmacılar için gerçekleştirilen işlemler 10 bayan konuşmacı içinde tekrarlanarak bayanların telaffuzları için ortalama özellik matrisler klasörleri oluşturulmuştur.

- 10 erkek – 10 bayan olmak üzere toplam yirmi kişiye ait konuşma verileri yukarıda anlatılan yöntem tekrarlanarak kişiden bağımsız özellik matrisleri oluşturulmuştur.

Tasarlanan sistemde komutların Tablo 4.1'de verilen sırada söylenmesi gerekmektedir. Konuşmacılardan alınan son iki kayıt sistemin veriminin test edilmesi amacıyla kullanılmıştır. Komut telaffuz kayıtları belirtilen sırayla sisteme uygulanmıştır ve işlemler sonucunda kontrol kümesi içeriği kontrol edilerek sistemin verimi değerlendirilmiştir. Bu amaçla yapılan verim testlerinin sonuçları Tablo 4.3'te verilmiştir.

Tablo 4.3. incelendiğinde en iyi sonuçların bayan konuşmacılar için bayan konuşmaların eğitim telaffuzlarından elde edilen özellik matrisleri kullanıldığında ve erkek konuşmacılar için ise erkek konuşmaların eğitim telaffuzlarından elde edilen özellik matrisleri kullanıldığında elde edilmiştir. Aksine, en kötü sonuçlar bayan konuşmacılar için erkek konuşmaların eğitim telaffuzlarından elde edilen özellik matrisleri kullanıldığında ve erkek konuşmacılar için ise bayan konuşmaların eğitim telaffuzlarından elde edilen özellik matrisleri kullanıldığında elde edilmiştir. Burada sistem verimi Tablo 4.1'deki her hangi bir komut bütünü için kontrol kod kümesi doğru bir şekilde sistem tarafından belirlenmiş ise; sistem ilgili komutu doğru sınıflandırmış olur. Aslında her bir komut içindeki kelimelerden biri yanlış sınıflandırıldığında sistem doğrudan bu komutu başarısız kabul etmektedir. Bu durum sistem veriminde bir düşüşe sebep olmaktadır. Halbuki sistem her bir telaffuzu çok daha yüksek verimle sınıflandırmaktadır.

Tablo 4.3: Sistem verimi ölçüm sonuçları (%)

Konuşmacı Etiketleri	Erkek Konuşmacılar Ortalama Özellik Vektörleri	Bayan Konuşmacılar Ortalama Özellik Vektörleri	Tüm Konuşmacılar için Ortak Ortalama Özellik Vektörleri
M1	100,00	16,67	83,33
M2	96,67	53,33	100,00
M3	100,00	33,33	53,33
M4	100,00	73,30	100,00
M5	100,00	33,33	70,00
M6	100,00	90,00	100,00
M7	73,33	46,67	60,00
M8	100,00	60,00	100,00
M9	93,33	100,00	86,67
M10	93,33	66,67	100,00
F1	86,67	100,00	86,67
F2	53,33	93,33	86,67
F3	46,67	100,00	100,00
F4	50,00	100,00	73,33
F5	46,67	100,00	93,33
F6	83,33	100,00	100,00
F7	56,67	100,00	100,00
F8	53,33	100,00	100,00
F9	30,00	93,33	73,33
F10	46,67	100,00	100,00
Erkek Konuşmacıların Ortalamaları	95,67	57,33	85,33
Bayan Konuşmacıların Ortalamaları	55,33	98,67	91,33

Sistemin kişiden bağımsız çalışması amaçlanmaktadır. Ancak bayanların ve erkeklerin ses yollarının fiziksel farklıları gerçekleştirilen sistemin kişiden bağımsızlığını kısıtlamakta ve verimini düşürmektedir. Tüm konuşmacılar için elde edilen özellik matrisi kullanılarak yapılan çalışmalarda ise bayan sesleri %91.33, erkek sesleri ise %85.33 başarımla sınıflandırılmıştır. Bu iki değerin ortalaması alındığında sistemin bayan ve erkek ayrımı olmaksızın kişiden bağımsız başarı verimi %88.33 olarak bulunmuştur.

Elde edilen sonuçlar bayan ve erkeklerin ses yollarının fiziksel farklılıklarının neden olduğu akustik-parametrik farklılıkları ve bunların ne kadar belirleyici olduğunu açıkça göstermektedir. Bayan özellik matrisleri ile karşılaştırılan erkek sesleri ve erkek özellik matrisleriyle karşılaştırılan bayan sesleri %58 'in altında bir başarı

göstermektedir. Bu özellik kullanılarak konuşmacının bayan mı olduğu yoksa erkek mi olduğu konusunda yorum yapılabilmektedir. Bu bilgi kullanılarak gerçekleştirilecek bir sistem; eğitim amaçlı alınacak komut telaffuzlarıyla öncelikle kullanıcının cinsiyetini belirleyebilir. Kullanıcının cinsiyetine göre sistem, kullanacağı özellik matrislerini erkek veya bayan telaffuzlardan elde edilmesini otomatik olarak belirleyebilir. Böylelikle sistem her zaman %95'in üzerinde verimle çalışabilir.

4.5. Konuşma Tanıma Sisteminin Süzgeçlenmiş Gürültülü Konuşma İşaretleri İçin Veriminin Belirlenmesi

Konuşma tanıma sistemleri gürültülü konuşma işaretlerini sınıflamada başarılı oldukları sevide güvenilirdirler. Gürültünün seviyesi arttıkça konuşma tanımadaki verim düşmektedir. Bölüm 3'te gerçekleştirmiş olduğumuz çalışmada araç içersindeki gürültünün bastırılması işleminde gerçek zamanlı bir uygulama için en iyi verimin, süzgeç ağırlıkları RLS algoritması tarafından yenilenen uyarlamalı bir FIR süzgeç tarafından sağlandığı belirlenmiştir. Oluşturduğumuz konuşma tanıma sisteminin, uyarlamalı süzgeç tarafından gürültüden arındırılan işaretler için veriminin belirlenmesi aşamasında aşağıda açıklanan çalışmalar yapılmıştır.

Sistem veriminin Bölüm 4.4'te anlatıldığı şekilde bayan konuşmacıların telaffuzları ile oluşturulan bayan konuşmacılara ait özellik matrisleri kullanılarak; öncelikle bayan konuşmacılardan test amacıyla alınan konuşma işaretlerine sırasıyla +5dB, 0dB, -5dB SNR oranlarında BMW 750i marka otomobil içerisinde kaydedilmiş renkli gürültü bindirilmiştir. Daha sonra uyarlamalı süzgeç kullanılarak işaretler gürültüden temizlenmeye çalışılmıştır. Uyarlamalı süzgeç çıkışında elde edilen işaretler, Tablo 4.1'de belirtilen komut sıralarıyla sisteme uygulanmıştır ve işlemler sonucunda kontrol kümesi içeriği kontrol edilerek sistemin verimi değerlendirilmiştir. Çalışmalar sonunda elde edilen, gürültüden temizlenmiş konuşma işaretleri için sistem verimine ilişkin sonuçlar Tablo 4.4'de verilmiştir.

Tablo 4.4: Sistemin gürültüden temizlenmiş konuşma işaretleri için verimi (%)

SNR	Sistem Verimi
+5dB	61,00
0dB	63,33
-5dB	66,67

Üzerine gürültü bindirilmiş konuşma işareti, gürültünün şiddetine göre belirleyici karakteristik özelliklerini kaybetmektedir. Uyarlamalı süzgeç tarafından süzgeçlenmiş gürültülü konuşma işareti asıl özelliklerine tamamen dönemez. Sadece işaretin üzerinde barındırdığı gürültü karakteristikleri kısmen bastırılır. Ayrık konuşma tanıma sistemleri gürültülü konuşma işaretlerini sınıflandırma da, iki temel problem ile karşılaşırlar. Birinci problem konuşma bölgesinin saptanmasıdır. Üzerine gürültü bindirilmiş konuşma işaretinin sınırlarını belirlemek oldukça zordur. Oluşturduğumuz sistem konuşma bölgesi saptama aşamasında işaretin enerjisine ve sıfır geçiş sayısına bakmaktadır. Uyarlamalı süzgeç çıkışında elde edilen işaretin sınırları her zaman doğru bir şekilde saptanamamaktadır. Buda sistemin verimini düşürmektedir. İkinci problem ise konuşma işaretinin ayırt edici özelliklerinin gürültü işareti tarafından çok fazla değiştirilmesidir. Konuşma işaretinin sınırları doğru olarak belirlense de, karakteristiği çok fazla değiştiği için sistem konuşmayı doğru bir şekilde sınıflandıramamaktadır. Temiz bayan konuşma işaretleri için sistemin verimi %98.67 iken gürültü tarafından etkilenmiş aynı konuşma işaretleri için SNR oranına göre %61–%67 değerleri arasındadır. Tablo 4.4. 'te de görüldüğü gibi gürültü işaretinin gücü arttıkça sistemin verimi düşmektedir. Araç içi gürültünün SNR oranı genellikle -15dB kabul edildiği göz önünde bulundurulduğunda, çalışmamızda yüksek SNR oranlarına sahip gürültü işaretleri uygulanmıştır. Temiz bayan konuşmacıların işaretlerine SNR oranı -15dB düzeyinde aynı karakteristikte renkli gürültü eklediğimizde ise sistem verimi %75'in üzerindedir. Buradan da anlaşılacağı üzere gürültü oranının sistem verimine etkisi azımsanmayacak derecede önemlidir.

4.6. Gerçek Zamanlı Konuşma Tanıma Sistemi

Şekil 4.1'de gösterilen sistemin gerçek zamanlı uygulaması Matlab 6,5 sürümünde Data Acquisition Toolbox kullanılarak, Intel(R) Pentium(R) Mobile işlemcide, 1,6 GHz işlemci hızına ve 512 MByte RAM hafızaya sahip bir bilgisayar aracılığı ile gerçekleştirilmiştir. Tablo 4.4'de kullanılan Data Acquisition Toolbox özellikleri verilmiştir (Data Acquisition Toolbox User's Guide, 2002).

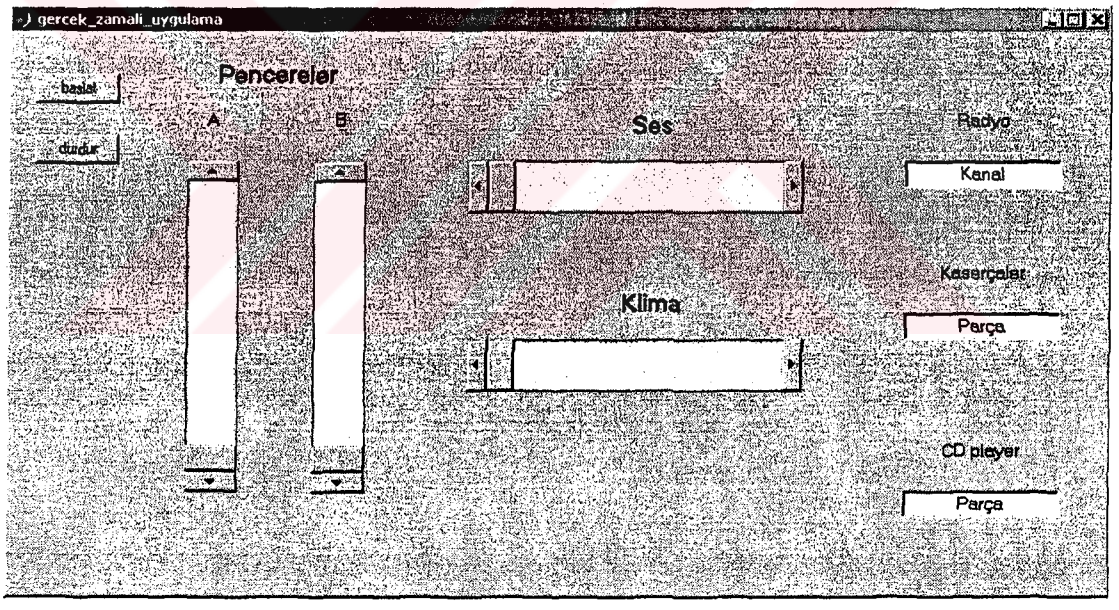
Tablo 4.5: Data Acquisition Toolbox özellikleri

Analog giriş	Winsound
Kanal sayısı	1
Örnekleme oranı	8000
Tetikleme tekrar sayısı	0
Tetikleme tipi	Manual
Zamanlayıcı fonksiyon	Veri_al
Tetiklemede alınan örnek sayısı	4000
Zamanlayıcı fonksiyon periyodu	0,5 s.
Tampon konfigürasyonu	[4000 20]

Data Acquisition toolbox, Matlab yorumlayıcısının dış ortamdan veri okumasına veya dış ortama veri göndermesine imkan sağlayan M-dosya fonksiyonlardan ve MEX-dosya dinamik bağlantı kütüphanelerinden (DLLs) oluşan araç kutusudur. Tez çalışmamızda bu araç kutusu, ses kartından veri okuma amacıyla kullanılmıştır. Analog giriş "winsound" seçildiğinde Matlab bir saniyede örnekleme oranı kadar veriyi ses kartından okur. Kullanıcı tarafından yazılan "Zamanlayıcı fonksiyon", "zamanlayıcı fonksiyon periyodu" kadar sıklıklarla çalıştırılır. Oluşturduğumuz veri edinim yapısıyla ses kartından her 0,5 s.'de bir 4000 veri, veri_al fonksiyonu aracılığı ile okunmaktadır. 0,5 s. sıklıklarla gelen veriler yine veri_al fonksiyonunda bir değişkene birbiri ardına atanarak konuşma bölgesi saptama algoritmasına uygulanmaktadır. "veri_al" fonksiyonunun ve "veri_al" içerisinde kullanılan diğer fonksiyonların kaynak kodu Ek-A'da verilmiştir. Herhangi bir anda konuşma bölgesi saptama algoritması konuşma bölgesini belirlediği anda konuşma işaretini Şekil 4.1'de gösterilen bir sonraki bloğa gönderir. Sistem; DTW işleminden itibaren konuşma işaretini işleyerek kod kontrol kümesinin içeriğini belirlerken diğer taraftan veri_al fonksiyonu ses kartından 0,5 s.'de bir veri okumaya devam edecektir.

Sistemin oluşturulan yapıya göre konuşmayı işlemesi ve veri tabanında ağaç yapısının ilgili dalına göre gerekli telaffuzlarla kontrol ederek kod kontrol kümesinin içeriğini belirlemesi 0,5 s.'nin altında bir sürede gerçekleşir. Bu durum dış ortamdan alınan konuşma verilerinde herhangi bir kayıp yaşanmamasını sağlar. Eğer komut ağaç yapısının bir dalında dörtten fazla komut konulursa mevcut sistemin işlemleri tamamlaması 0,5 s.'den fazla sürmektedir ve bu durumda dış ortamdan alınan veride kayıp olmasına neden olmaktadır. Daha fazla komut sayısına sahip uygulamaları daha hızlı donanımsal özelliklere sahip bir bilgisayar üzerinde gerçekleştirmek mümkündür. Aslında hedeflediğimiz sistem, mikrodenetleyici tabanlı bir donanımda gerçeklenirse daha fazla komut işlenmesi mümkün olacaktır.

Gerçek zamanlı uygulamanın çalışmasını görsel olarak izleyebilmek amacıyla Şekil 4.3'de verilmiş olan GUI penceresi hazırlanmıştır.



Şekil 4.3: Gerçek zamanlı uygulama için hazırlanan GUI penceresi

BÖLÜM 5. SONUÇLAR VE ÖNERİLER

Konuşma tanıma ile ilgili çalışmalar 50 yılı aşkın süredir devam etmektedir. Bu alanda bugüne dek birçok başarılı yöntem geliştirilmiş olmasına rağmen gelinen nokta başlangıçta hedeflenen düzeyin gerisindedir. Bunun sebebi ses işaretinin dinamik bir yapıya sahip olması, özelliklerinin kişiden kişiye farklılıklar göstermesi, konuşma tanıma sürecinde insan doğasında geçirilen evrelerle ilgili bilgilerin yetersiz olmasıdır. Konuşma tanıma işlemi çok zor bir görev olmasına rağmen, bugün geldiğimiz noktada konuşma tanıma sistemleri pek çok uygulamanın bazı aşamalarında kullanılabilmektedir.

Konuşma tanıma sistemlerinin araç içerisinde kullanımı 1980 'lerde yüksek hızlı işlemcilerin üretilmeye başlamasıyla üzerinde çalışılmaya başlanmış bir konudur. O günlerden günümüzde dek bu konuda birçok ticari ve akademik çalışmalar yapılmış olmasına ve hatta büyük otomobil üreticilerinin konuşma tanıma sistemlerini ürettikleri araçlarda kullanacaklarını duyurmalarına rağmen gelinen noktada hala araç içerisinde konuşma tanıma sistemleri ticari olarak kullanılamamıştır. Bunun sebebi şimdiye dek geliştirilmiş olan sistemlerin güvenilirliğinin ve veriminin istenen düzeyde olmamasıdır. Bu durum konuşma tanıma işleminin zorluğunu açıkça göstermektedir.

Konuşma tanıma sistemlerinin araç içerisinde kullanımını zorlaştıran diğer bir etken ise araç içerisindeki gürültünün (motor, çalışan cihaz ve diğer insanların konuşmaları vb.) neden olduğu etkidir. Gerçek zamanlı konuşma tanıma sisteminin araç içerisinde kullanılması işleminde gürültünün bastırılması işlemi başlı başına bir konu olarak karşımıza çıkmaktadır. Gürültü bastırma işlemiyle ilgili tezimizde gerçekleştirilen çalışmalar sonucunda karşılaştırdığımız yöntemlerden araç içerisindeki renkli gürültüyü en iyi bastırma verimine GA yapısının sahip olduğu saptanmıştır. Fakat GA yapısı gürültü bastırma işlemindeki iyi veriminin yanında gerçek zamanlı bir uygulamada kullanılamayacak kadar yavaştır. Tez çalışmamızda amaçlanan araç

içindeki cihazların sesle kontrolüne yönelik konuşma tanıma sisteminin gerçek zamanlı uygulaması için gürültü bastırma işleminde kullanılabilecek yöntemin belirlenmesinde, yöntemin gürültü bastırma verimi kadar çalışma hızı da önemlidir. Bu doğrultuda gerçekleştirilen çalışmalar sonucunda karşılaştırılan yöntemlerden gerçek zamanlı bir uygulamada kullanılabilecek en iyi yöntemin süzgeç katsayıları RLS algoritması kullanılarak yenilenen uyarlamalı bir süzgeç olduğu saptanmıştır.

Tez çalışmamızda amaçladığımız araç içindeki cihazların sesle kontrolüne yönelik konuşma tanıma sisteminin gerçek zamanlı uygulaması işleminde oluşturduğumuz sistemde kullanılan yöntemlerin seçiminde konuşma tanıma işlemini gerçekleştirme verimleri kadar işlem hızları da dikkate alınmıştır. Bu amaçla çok hızlı bir yöntem olduğu bilinen DTW algoritması referans ve test işaretlerinin boyutlarını eşitleme işleminde kullanılmıştır. Ayrıca özellik çıkartım yöntemlerinden; Özilintili (Autocorrelation) Analiz, Ortak Değişinti (covariance) Yöntemi ve Mel-Frekans Ölçekli Kepstral Katsayılar (MFCC) yöntemi işlem süreleri açısından incelenmiştir. Yöntemlerden konuşma tanıma verimi en iyi olduğu bilinen MFCC yönteminin aynı zamanda en hızlı yöntem olduğu görülmüştür ve oluşturulan gerçek zamanlı konuşma tanıma sisteminde özellik çıkartım yöntemi olarak MFCC yöntemi kullanılmıştır.

Oluşturulan sistemin verimi incelendiğinde erkek konuşmacıların sesleri kullanılarak eğitilen sistemin erkek konuşmacının komutlarını tanıma başarısı %95.67, bayan konuşmacıların sesleri kullanılarak eğitilen sistemin bayan konuşmacıların komutlarını tanıma başarısı %98.67 olarak saptanmıştır. İki grubun konuşmacılarının sesleri kullanılarak eğitilen sistemin bayan veya erkek ayırt etmeksizin konuşmacının komutlarını tanıma başarısı %88.33 olarak saptanmıştır. Bu durum gerçekleştirilecek bir sistemde konuşmacıları bayan veya erkek olarak sınıflamanın verimi pozitif yönde arttıracağını göstermektedir. İleri çalışmalarda, erkek yada bayan sesleri ile eğitilmiş iki ayrı veritabanına sahip bir sistemin, sürücünün bayan veya erkek olmasına göre kullanacağı veri tabanını otomatik olarak belirleyen ve böylelikle daha iyi verim gösteren bir sistem gerçekleştirilebilir.

Ayrıca en iyi verimi sağlayan bayan konuşmacıların verisine çeşitli SNR oranlarında renkli gürültü eklenerek sistem verimi saptanmıştır. İdeal araç ortamı göz önüne alınarak çalışmamızda gerçekleştirilen gürültülü konuşma işareti ile sistem başarımı %75'in üzerindedir. Gerçek bir uygulamada bu değerin iyileştirilmesi araç içine konumlandırılacak referans ve konuşmacı mikrofon dizilerinin en uygun biçimde yerleştirilmesiyle mümkündür.

Tez çalışmamızda gerçekleştirilen gerçek zamanlı konuşma tanıma sistemi uygulamasında ayrık kelime tanıma yöntemi kullanılmıştır ve bu yüzden komut listesi ağaç yapısı şeklinde verilmiştir. Ağaç yapısının her dalında sistem en fazla dört kelime için karşılaştırma yapmaktadır ve bu kelimelerden giriş işaretine en yakın olan saptanmaktadır. Sistemde her ağaç dalının üç kelime ile sınırlandırılmasının nedeni sistemin gerçekleştirildiği işlemcinin hızının daha fazla kelime kullanımına izin vermemesi, daha fazla kelime kullanımının gecikmelere ve dolayısıyla veri kaybına neden olmasıdır. Kullanılacak daha hızlı bir işlemci veya donanım ile komut sayısını dolayısıyla sistemin gerçekleştireceği işlem sayısını arttırmak mümkün olacaktır. Bunun yanında mikrodenetleyicili bir konuşma tanıma sisteminde bu algoritmaların koşturulması sonucunda amaçlanan sitemin bir araç içinde bilgisayardan bağımsız bir şekilde kullanılabileceğinin mümkün olduğu söylenebilir.

Yaygın olmamakla birlikte özürlü insanlar için tasarlanan araçlarda konuşma işareti ile araç kontrolü gerçekleştirilmektedir. Bu tür uygulamalarda sürücünün özrüne göre çeşitli değişiklikler yapılmakta, dolayısıyla konuşma tanıma sistemi konuşmacıya bağımlı bir şekilde tasarlanmaktadır. Bu çalışmamızda ise konuşmacıdan bağımsız konforlu ve güvenli bir sürüş sağlayacak özelliklerde araç içi konuşma tanıma sisteminin benzetimi yapılmıştır. Otomobil üreticilerinin maliyet ve güvenilirlik endişeleri giderilirse, önümüzdeki yıllarda bu tür sistemlerin ticari amaçlı yaygınlaşması mümkün olacaktır.

KAYNAKLAR

1. AMARI, S., CICKOCKI, A., 1998. Adaptive Blind Signal Processing—Neural Network Approaches, Proceedings of the IEEE, vol. 86, no. 10, pp. 2026.

2. ARSLAN, T., HORROCKS, D.H., 1996. Hierarchical Frequency Grid Management for Arbitrary Responce FWL IIR Filter Design Using a Genetic Algorithm, IEEE International Symposium on Circuits and Systems, pp. 104-106.

3. AYDIN, Ö., 2005. Yapay Sinir Ağlarını Kullanarak Bir Ses Tanıma Sistemi Geliştirilmesi, Yüksek Lisans Tezi, Trakya Üniversitesi Fen Bilimleri Enstitüsü, Edirne.

4. BACK, A.D., TSOI, A.C., 1991. FIR and IIR Synapse, a New Neural Network Architecture for Time Series Modeling, Neural Computation, Vol 3, No 3, pp. 375-385.

5. BAYGIN, M.K., YALDIR, A.K., 2006. Linear Predictive Coding ve Dynamic Time Warping Teknikleri Kullanarak Ses Tanıma Sistemi Gerçekleştirilmesi, Bilgi Teknolojileri Kongresi, Pamukkale Üniversitesi,Türkiye.

6. BING-FEI, W, KUN-CHING, W., 2005. Robust Endpoint Detection Algorithm Based on the Adaptive Band-Partitioning Spectral Entropy in Adverse Environments, IEEE Transaction on Speech and audio processing, vol. 13, No. 5.

7. BONISSONE, P.P., 2002. Adaptive Neural Fuzzy Inference Systems (ANFIS): Analysis and Applications, GE CRD Schenectady, NY USA.

8. CAVEDON, L., WENG, F., MISHRA, R., BRATT, H., RAGHUNATHAN, B., CHENG, H., SCHMIDT, H., MIRKOVIC, D., BEI, B., PON-BARRY, H., SCHEIDECK, T., LATHROP, B., CHEN, J., PETERS, S., SHRIBERG, L., BERGMANN, C., 2004. Developing a Conversational In-Car Dialog System, The 12th International Congress on Intelligent Transportation Systems, San Francisco CA, USA.

9. CHOLEWO, T.J., ZURADA, J.M., 1998. Second Order Derivatives in Feedforward FIR Neural Networks, Proc. of the International Joint Conference on Neural Networks, Anchorage, Alaska, vol. 2, pp. 1074-77.

10. DAVIS, C.D., 2001. Automatic Speech Recognition and Access: 20 years, 20 Months, or Tomorrow?, Issue of Hearing Loss: The Journal of Self Help for Hard of Hearing People 22(4), p. 11-14.

11. DELLER, J.R, PROAKIS, J.G, HANSEN, J.H.L, 1993. Discrete-Time Processing of Speech Signals, ISBN 0-02-328301-7, Macmilan Publish Company.

12. FATİHOĞLU, Y.S., 2003. Metinden Bağımsız Kapalı-Set Konuşmacı Tanıma, Yüksek Lisans Tezi, Kocaeli Üniversitesi Fen Bilimleri Enstitüsü, Kocaeli.

13. FARELL K. R. AND MAMMONE R.J., 1994. Speaker Recognition Using Neural Networks and Conventional Classifiers, IEEE Trans. On Speech and Audio Processing, Vol.2, No.1, part II.

14. FORLINES, C., NIELSEN, B.S., RAJ, B., WITTENBURG, K., WOLF, P., 2005. A Comparison Between Spoken Queries and Menu-based Interfaces for in car Digital Music Selection, Mitsubishi Electric Research Laboratories Inc., Cambridge, Massachusetts.

15. FURUI, S., 1989. Digital Speech Processing, Synthesis, and Recognition, ISBN 0-08-2477965-7, Marcel Dekker Inc.

16. FURUI, S., 1986. Speaker-independent isolated word recognition using dynamic features of speech spectrum, IEEE Transaction on Acustic, Speech, and Signal Processing, Vol. Assp-34, No. 1.

17. FURUI, S., 2000. Speech Recognition Technology in the Ubiquitous/Wearable Computing Environment, Acoustics, Speech, and Signal Processing, ICASSP '00. Proceedings. 2000 IEEE International Conference on.

18. HAGAN, M.T.,MENJAH, MB., 1994. Training Feedforward Networks with the Marquardt Algorithm, IEEE Transaction Neural Networks, 5(6), pp. 989-993.

19. HANSEN, J.H.L., ANGKITITRAKUL, P., PLUCIENKOWSKI, J., GALLANT, S, YAPANEL, U, PELLOM, B., WARD, W., COLE, R., 2001. CU-Move : Analysis & Corpus Development for Interactive in-Vehicle Speech Systems, Eurospeech-2001, pp. 2023-2026, Aalborg, Denmark.

20. HAYES, M.H., 1996. Statistical Digital Signal Processing and Modelling, ISBN 0-0471 59431-8, Jonh Wiley & Sons, Inc.

21. HAYKIN, S., 2002. Adaptive Filter Theory, ISBN 0-13-090126-1, Prentice Hall Inc.

22. HEISTERKAMP, P., 2001. Linguatronic Product-Level Speech System for Mercedes-Benz Cars, DaimlerChrysler AG Research and Technology Speech Understanding, Germany.

23. HELIE, S., CHARTIER, S., PROULX, R., 2004, Applying Fuzzy Logic to Neural Modeling, Proceedings of the Sixth International Conference on Cognitive Modelling (pp. 352-353), Mahwah, NJ.

24. HELLMANN, M., 2001. Fuzzy Logic Introduction, Laboratoire Antennes Radar Telecom, France.

25. HEUVEL, H., BOUDY, J., COMEYNE, R., EULER, S., MORENO, A., RICHARD, G., 1999. The speechDat-Car Multilingual Speech Databases for in-Car Applications: Some First Validation Results, Proceedings EUROSPEECH'99, Budapest, Hungary, vol.5, pp. 2279-2282.

26. HUSSAIN, A., CAMPBELL, D.J., 1998. Binaural Sub-Band Adaptive Speech Enhancement Using Artificial Neural Networks, Speech Communication, Vol.25, Elsevier Science B.V., Amsterdam, The Netherlands, pp.177-186.

27. İNAL, M., BÜTÜN, E., ERKAN, K., YILDIRIM, M., ÇEKEN, C., 2000. Comparison of Linear Predictive Analysis Methods for ANN-Based Speaker Identification, Neurel-2000, 5. Seminar on Neural Network Application in Electrical Engineering, Faculty of Elec. Eng. Univ. of Belgrade, YU, pp.109-112.

28. ISHIBUCHI, H., NOZAKI, K., YAMAMOTO, N., TANAKA, H., 1995. Selecting Fuzzy If-Then Rules for Classification Problems Using Genetic Algorithms, IEEE Transaction on Fuzzy Systems, vol. 3, No. 3.

29. JANG, J.R., 1993. ANFIS: Adaptive-Network-Based Fuzzy Inference System, IEEE Transaction on System, Man, and Cybernetics, vol. 23, No. 3.

30. JANG, R., SUN, C.T., 1995. Neuro-Fuzzy Modeling and Control, Proceedings of The IEEE, Vol. 83, No. 3.

31. JANG, R., GULLEY, N., 1995. Fuzzy Logic Toolbox User's Guide, The Mathworks Inc.

32. JANTZEN, J., 1998. Tutorial On Fuzzy Logic, Technical University of Denmark, Department of Automation, Tech. report no 98-E 868, Bldg 326, DK-2800 Lyngby, DENMARK.

33. JANTZEN, J., 1998. Neurofuzzy Modelling, Technical University of Denmark, Department of Automation, Bldg 326, DK-2800 Lyngby, DENMARK, (1998)

34. KAYA, İ., OKTAY, S., ENGİN, O., 2005. Kalite Kontrol Problemlerinin çözümünde Yapay Sinir Ağlarının Kullanımı, Erciyes Üniversitesi Fen Bilimleri Enstitüsü Dergisi, 21 (1-2), pp. 92-107.

35. KOÇ, A., 2002. Acoustic Feature Analysis for Robust Speech Recognition, Yüksek Lisans Tezi, Boğaziçi Üniversitesi, İstanbul.

36. LAMEL, L., RABINER, L.R., ROSENBERG, A., WILPON, J., 1981. An Important Endpoint Dedector for Isolated Word Recognition, IEEE Transaction on Acustic, Speech, and Signal Processing,vol. ASSP-29, No. 4.

37. LI, M., AND OKAMURA, A.M., 2003. Recognition of Operator Motions for Real Time Assistance using Virtual Fixtures, Proceedings of the11th Symposium on Haptic Interfaces for Virtual Environment and Teleoperator Systems, (HAPTICS'03).

38. MASTORAKIS, N.E.,GONOS, I.F.,SWAMY, M.N.S., 2003. Design of Two-Dimensional Recursive Filters Using Genetic Algorithms, IEEE Transactions on Curcuits and Systems—I: Fundamental Theory and Applications, vol 50, no. 5.

39. MENG, T.H., 1994. Normalized Data Nonlinearities for LMS Adaptation, IEEE Transaction on Acustic, Speech, and Signal Processing, vol. 42,pp.1352-1365.

40. MYERS, C.S., RABINER, L.R., ROSENBERG, A.E., 1980. An Investigation of the Use of Dynamic Tinie Warping for Word Spotting and Connected Speech Recognition, CH1559—4/80/0000—0437$OO.75©. 1980 IEEE.

41. MYERS, C.S., RABINER, L.R., ROSENBERG, A.E., 1980. Performance tradeoffs in dynamic time warping algorithms for isolated word recognition, IEEE Transaction on Acustic, Speech, and Signal Processing, Vol. Assp-28, No.6.

42. NAGUMO, J.I., NODA, A., 1967. A Learning Method for System Identification, IEEE Transaction Automatic Control, vol AC-12,pp. 282-287.

43. NGIA, L.S.H., SJOBERG, J., VIBERG, M., 1998. Adaptive Neural Nets Filter Using a Recursive Levenberg-Marquardt Search Direction, Proc. Asilomar Conf. Signals, Systems, Computers, pp. 697-701.

44. NGUYEN, N.D., WIDROW, B., 1990. Improving the Learning Speed of 2-Layer Neural Networks bu Choosing Initial Values of Adaptive Weights, International Joint Conference of Neural Networks, vol.3, pp. 21-26.

45. RABINER, L.R., SAMBUR, M.R., 1975. An Algorithm for Determining the Endpoints of Isolated Utterances, Bell Syst. Tech. Journal, vol. 54, pp. 297-315.

46. RABINER, L.R., JUANG, B.H., 1993. Fundamentals of Speech Recognition, ISBN 0-013-015157-2, PTR Prentice-Hall Inc.

47. RABINER, L.R., ROSENBERG, A.E., LEVINSON, S.E., 1978. Considerations in Dynamic Time Warping Algorithms for Discrete Word Recognition, IEEE Transaction on Acustic, Speech, and Signal Processing, Vol. Assp-26, No. 6.

48. RODMAN, J., 2003. The Effect of Bandwith on Speech Intelligibility, Polycom, The United States.

49. RUSSELL, M., BROWN, C., SKILLING, A., SERIES, R., WALLACE, J., BONHAM, B., BARKER, P., 1996. Applications of Automatic Speech Recognition of Speech and Language Development in Yough Children, British Crown Copyright / DERA.

50. SAĞIROĞLU, Ş., BEŞDÖK, E., ERLER, M., 2003. Mühendislikte Yapay Zeka Uygulamaları, Yapay Sinir Ağları, ISBN 9759594854, Ufuk Yayınları.

51. SÖKE, A., 2003. Genetik Algoritma ve Benzetilmiş Tavlama ile İki Boyutlu Giyotinsiz Kesme Problemlerine Olasılıksal Yaklaşım, Yüksek Lisans Tezi, Kocaeli Üniversitesi Fen Bilimleri Enstitüsü.

52. STEIMANN, F., 1997. Fuzzy set theory in medicine, Artificial Intelligence in Medicine, 11 (1997) 1 - 7, Elsevier Science.

53. TÜRK, O., ŞAYLI, Ö., DUTAĞACI, H., ARSLAN, L., 1999. Alt-bant İşlemeye Dayalı Bir Ses Sınıflandırma Sistemi, Boğaziçi Üniversitesi, Türkiye.

54. VAN THONG, J-M, GODDEAU, D., LITVINOVA, A., LOGAN, B., MORENO, P., AND SWAIN, M., 2000. SpeechBot: a Speech Recognition based Audio Indexing System for the Web, International Conference on Computer-Assisted Information Retrieval, Recherche d'Informations Assistee par Ordinateur (RIAO), Paris, pp 106-115.

55. VINGER, K.A.,TORRESEN, J., 2003. Implementing Evolution of FIR-Filters Efficiently in an FPGA, In proc. of 2003 NASA/DoD Conference on Evolvable Hardware, Chicago, Illinois, USA.

56. WALD, M., 2004. Using Automatic Speech Recognition to Enhance Education for All Students: Turning a Vision into Reality, 34th ASEE/IEEE Frontiers in Education Conference,Savannah, GA.

57. WANG, H., YANG, C., WANG, J., WU, C., CHIEN, J., 2005. TAICAR - The Collection and Annotation of an in-Car Speech Database Created in Taiwan, Computational Linguistic and Chinese Language Processing, vol. 10, no. 2, pp. 237-250.

58. WATIKA, T., 2004. Speech Based Interfaces in Vehicles, R&D Review of Toyota CRDL, Vol. 39 No.1.

59. WENG, F., CAVEDON, L., MISHRA, R., BRATT, H., RAGHUNATHAN, B., CHENG, H., SCHMIDT, H., MIRKOVIC, D., BEI B., PON-BARRY, H., SCHEIDECK, T., LATHROP, B., CHEN, J., PETERS, S., SHRIBERG, L., BERGMANN, C., 2005. A Flexible Conversational Dialog System for MP3 Player, Proceedings of HLT/EMNLP 2005 Demonstration Abstracts, pages 24–25, Vancouver.

60. WESTPHAL, M., AND WAIBEL, A., 1999. Towards Spontaneous Speech Recognition For On-Board Car Navigation And Information Systems, Proceedings of the Eurospeech 99, pp. 1955-1958.

61. WINDROW, B., GLOVER, J.R., KAUNITZ, J., WILLIAMS, C.S., HEARN, R.H., ZEIDLER, J.R, DONG, E., GOODLIN, R.C, 1975. Adaptive Noise Cancelling: Principles and Aplications, Proceeding of The IEEE, vol. 63, No. 12.

62. YAPANEL, U., ZHANG, X., HANSEN, J.H.L., 2002. High Performance Digit Recognition in Real Car Environments, ICSLP-2002:Inter. Conf. on Spoken Language Processing, vol. 2, pp. 793-796, Denver, CO USA.

63. YURTOĞLU, H., 2005. Yapay Sinir Ağları Metodolojisi ile Öngörü Modellenmesi, Uzmanlık Tezi, DPT, Yayın no. 2683.

64. "Data Acquisition Toolbox User's Guide", The Mathworks Inc. 2002.

EK A: PROGRAM KODLARI

% Gerçek zamanlı uygulamada 0.5 s'de bir çalıştırılan "veri_al" fonksiyonu

```
function veri_al(obj,event,handles)
 global bilgi;
 x = peekdata(obj, obj.SamplesPerTrigger);
 obj.veri=[obj.veri x'];
 if length(obj.veri)>100000 obj.veri(end-16000:end)=[];end
 sinir=konusma_bolgesi(obj.veri);
 if sinir(2)~=0
    bilgi=dynamictw2(obj.veri(sinir(1):sinir(2))',bilgi)
    obj.veri=[];
 end
```

% Konusma bölgesini saptamada kullanılan "konusma_bolgesi" fonksiyonu

```
function [sinirlar]=konusma_bolgesi(x,fs,l,step)
%işaretin ilk 100 ms'lik kısmının sessiz bölge olduğu kabul edilir
 if nargin<2 fs=8000; end
 if nargin<3 l=floor(fs*20e-3); end
 if nargin<4 step=floor(l/2);
 if length(x)==0
    sinirlar=[0 0];
    return;
 end
   e_x=20*log10(frame(x,'energy',l,step)); %x pencere ayrılır
   e_x(find(e_x<0))=0;   % ve her pencerenin enerjisi bulunur
   zcr=zcr_func(x);  %zcr_func ile sıfır geçiş sayısı bulunur
   e_x(length(zcr)+1:end)=[]; ort=mean(e_x(1:10));maxi=max(e_x);
   esik_zcr=mean(zcr(1:10));
   if ort<1              %enerji icin esik belirlenmesi
      esik=maxi*ort;
   else if maxi/ort<ort
         esik=ort;
      else
         esik=maxi/ort;
      end
   end
   while esik>maxi/3 esik=esik/2; end
   %işaretin enerjisine göre başlangıç noktasının belirlenmesi
   i=1; bas=[];
   while i<length(e_x)-2
      j=0;
      while (esik <e_x(i+j)) & (i+j<length(e_x))
         j=j+1;
      end
```

```
    if j>10
       bas=[bas;i j];
       i=i+j;
     end
     i=i+1;
  end
  %şaretin sıfır geçiş sayısına göre başlangıç noktasının belirlenmesi
  i=1;bas_zcr=[];
  while i<length(zcr)
    j=0;
    while (esik_zcr <zcr(i+j)) & (i+j<length(zcr))
      j=j+1;
    end
    if j>5
      bas_zcr=[bas_zcr;i j];
       i=i+j;
     end
     i=i+1;
  end
  if length(find(bas))==0 | length(find(bas_zcr))==0
    disp('konuşma saptanmadı'); sinirlar=[0 0]; return;
  end
  if size(bas,1)==1
    sinyal_e=bas(1,:);
  end
  %işaretin enerjisine göre bitiş noktasının belirlenmesi
  while size(bas,1)>1
    if bas(1,2)<20 & bas(2,1)-bas(1,1)-bas(1,2)>5
      bas(1,:)=[];sinyal_e=bas(1,:);
    else if bas(2,1)-bas(1,1)-bas(1,2)<=5
          bas(1,2)=bas(2,2)+bas(2,1)-bas(1,1);
          sinyal_e=bas(1,:);bas(2,:)=[];
        else
          sinyal_e=bas(1,:); bas(2,:)=[];
        end
    end
  end
  bas=bas(1,1)+bas(1,2);
  %işaretin sıfır geçiş sayısına göre batiş noktasının belirlenmesi
  if size(bas_zcr,1)==1
    sinyal_zcr=bas_zcr(1,:);
  end
  while size(bas_zcr,1)>1
    if bas_zcr(1,2)<15 & bas_zcr(2,1)-bas_zcr(1,1)-bas_zcr(1,2)>5
      bas_zcr(1,:)=[];sinyal_zcr=bas_zcr(1,:);
    else if bas_zcr(2,1)-bas_zcr(1,1)-bas_zcr(1,2)<=5 | (bas_zcr(2,2)>6 & bas_zcr(2,1)-
bas_zcr(1,1)-bas_zcr(1,2)<20)
        bas_zcr(1,2)=bas_zcr(2,2)+bas_zcr(2,1)-bas_zcr(1,1);
          sinyal_zcr=bas_zcr(1,:); bas_zcr(2,:)=[];
        else
          sinyal_zcr=bas_zcr(1,:); bas_zcr(2,:)=[];
        end
    end
```

```
    end
    bas_zcr=bas_zcr(1,1)+bas_zcr(1,2);
    % iki algoritmanın ortak başlangıç noktası
    if sinyal_e(1,1)-sinyal_zcr(1,1)>0 & mean(e_x(sinyal_zcr(1,1):sinyal_e(1,1)-1))>esik/4
       baslangic=sinyal_zcr(1,1)*step;
    else
       baslangic=sinyal_e(1,1)*step;
    end
    if bas-bas_zcr==0
       bitis=bas*step;
    end
    % iki algoritmanın ortak bitiş noktası
    if bas-bas_zcr>0 & mean(e_x(bas_zcr:bas))>esik*3
       bitis=bas*step;
    else if bas-bas_zcr<0 & mean(e_x(bas:bas_zcr))>esik*.25
          bitis=bas_zcr*step;
       else
          bitis=bas*step;
       end
    end
    %sinyalin son noktası ile bitiş arası 50 ms'den daha büyük ise bitiş;
    if (length(x)-bitis)>step*10
       sinirlar=[baslangic bitis+step];
    else
       sinirlar=[baslangic 0];
    end
  end
```

%Dinamik zaman Eğreltimi işlemi ve Kontrol Kümesinin İçeriğinin
%belirlenmesinde kullanılan "dynamictw2" fonksiyonu

```
function bilgi=dynamictw2(t,bilgi,Fs)
 if nargin<3 Fs=8000; end
 ison=3;ii=0;
 %veritabınından ağaç yapısına uygun verilerin okunması
 while ii<=ison
   if ii==0
      r=readwav('durdur'); disp('durdur');load durdur;
   elseif bilgi(1)==-1
      switch ii
         case 1, r=readwav('yukari');disp('yukarı');load yukari;
         case 2, r=readwav('asagi'); disp('aşağı');load asagi;
         case 3,  r=readwav('degistir');disp('degistir');load degistir;
      end
   elseif bilgi(1)==0 & (bilgi(2)==1 | bilgi(2)==0)
      if bilgi(3)==0 & bilgi(4)==1
         switch ii
            case 1, r=readwav('A');disp('A');load A;
            case 2, r=readwav('B');disp('B');load B;
            case 3, r=readwav('birlikte');disp('birlikte');load birlikte;
         end
      else
        switch ii
            case 1, r=readwav('muzik'); disp('müzik');load muzik;
```

```
            case 2, r=readwav('pencere'); disp('pencere');load pencere;
            case 3, r=readwav('klima');disp('klima');load klima;
          end
        end
      elseif bilgi(1)==1 & bilgi(2)==0
        if bilgi(3)==0 & bilgi(4)==0
          switch ii
            case 1, r=readwav('radio');disp('radio');load radio;
            case 2, r=readwav('kasetcalar');disp('kasetçalar');load teyp;
            case 3, r=readwav('cdplayer');disp('CDplayer');load CDplayer;
          end
        else
          switch ii
            case 1,  r=readwav('muziksistemi');disp('müziksistemi');load muziksistemi;
            case 2,  r=readwav('kanal');disp('kanal');load kanal;
            case 3,  r=readwav('parca');disp('parça');load parca;
          end
        end
      end
      %dinamik zaman eğreltimi
      D1 = specgram(r,512,Fs,512,384);
      D2 = specgram(t,512,Fs,512,384);
      SM = simmx(abs(D1),abs(D2));
      [p,q,C] = dp(1-SM);
      D2i1 = zeros(1, size(D1,2));
      for i = 1:length(D2i1); D2i1(i) = q(min(find(p >= i))); end
      D2x = pvsample(D2, D2i1-1, 128);
      d2x = istft(D2x, 512, 512, 128);
      d2x = resize(d2x', length(d1),1);
      melc_r=melc;%melc_r=melcepst(r,8000,'N',12,26,160,80);
      %boyutu düzenlenmiş sinyali MFCC katsayılarının çıkartılması
      melc_ty=melcepst(d2x,8000,'N',12,26,160,80);
      %özellik matrislerinin yakınlıklarının hesaplanması
      distmelc=mean(disteusq(melc_r,melc_ty))
      %telaffuz durdur ise;
      if ii==0 & distmelc<700
        ii=ison+1
        bilgi=[-1 -1 -1 -1 -1 -1 ];enkucuk=[];disp('işlemler durduruldu');
      elseif ii>0
        enkucuk(ii)=distmelc;ii=ii+1;
      else
        ii=ii+1;
      end
    end
    %komutun saptanması ve komut kod kümesinin içeriğinin belirlenmesi
    if (length(enkucuk)>0)
     [minn,yer]=min(enkucuk);
     if minn<900
       minn
      if bilgi(1)==-1
        switch yer
          case 1,bilgi=[0 0 -1 -1 -1 -1];    %yukarı
          case 2,bilgi=[0 1 -1 -1 -1 -1 ];   %aşağı
          case 3,bilgi=[1 0 -1 -1 -1 -1 ];   %degistir
```

```
      end
    elseif bilgi(1)==0 & (bilgi(2)==1 | bilgi(2)==0)
      if bilgi(3)==0 & bilgi(4)==1
        switch yer
          case 1,if bilgi(5)==-1 bilgi(5)=0,bilgi(6)=0; end  %sol pencere(A)
          case 2,if bilgi(5)==-1 bilgi(5)=0,bilgi(6)=1; end  %sağ pencere(B)
          case 3,if bilgi(5)==-1 bilgi(5)=1,bilgi(6)=0; end  %birlikte
        end
      else
        switch yer
          case 1,if bilgi(3)==-1  bilgi(3)=0;bilgi(4)=0; end  %ses (muzik)
          case 2,if bilgi(3)==-1  bilgi(3)=0;bilgi(4)=1; end  %pencere
          case 3,if bilgi(3)==-1  bilgi(3)=1;bilgi(4)=0; end  %klima
        end
      end
    elseif bilgi(1)==1 & bilgi(2)==0
      if bilgi(3)==0 & bilgi(4)==0
        switch yer
          case 1,if bilgi(5)==-1 bilgi(5)=0,bilgi(6)=0; end  %radyo
          case 2,if bilgi(5)==-1 bilgi(5)=0,bilgi(6)=1; end  %kasetçalar
          case 3,if bilgi(5)==-1 bilgi(5)=1,bilgi(6)=0; end  %CDplayer
        end
      else
        switch yer
          case 1,if bilgi(3)==-1  bilgi(3)=0;bilgi(4)=0; end  %müziksistemi
          case 2,if bilgi(3)==-1  bilgi(3)=0;bilgi(4)=1; end  %kanal
          case 3,if bilgi(3)==-1  bilgi(3)=1;bilgi(4)=0; end  %parça
        end
      end
  end
  elseif   minn>1900
   bilgi=[-1 -1 -1 -1 -1 -1];
  end
end
```

ÖZGEÇMİŞ

1980 yılında Bursa'da doğdu. İlk ve orta öğrenimini Bursa'da tamamladı. Lise öğrenimini Bursa Hürriyet Endüstri Meslek Lisesinde tamamladı. 1997 yılında Kocaeli Üniversitesi Teknik Eğitim Fakültesi Elektronik ve Bilgisayar Eğitimi Bölümü, Elektronik Öğretmenliği programında yüksek öğrenimine başladı. 2002 yılında Elektronik öğretmeni olarak bu bölümden mezun oldu. Aynı yıl Sakarya Üniversitesi Fen Bilimleri Enstitüsü Elektronik ve Bilgisayar Eğitimi Bölümünde yüksek lisans eğitimine başladı. 2003 yılında eğitimine askerlik nedeniyle ara verdi. 2004 yılında Kocaeli Üniversitesi Fen Bilimleri Enstitüsü Elektronik ve Bilgisayar Eğitimi Bölümünde yüksek lisans eğitimine tekrar başladı. 2004 yılı Mayıs ayında halen çalışmakta olduğu Kocaeli Üniversitesi Teknik Eğitim Fakültesi Elektronik ve Bilgisayar Eğitimi Bölümünde araştırma görevlisi olarak göreve başladı.