

T.C.  
İSTANBUL AYDIN ÜNİVERSİTESİ  
FEN BİLİMLERİ ENSTİTÜSÜ



SES VE GÖRÜNTÜ İŞLEME EŞLEŞTİRMESİ VE İŞİTME ENGELLİLER  
İÇİN BİR KONUŞMA TERAPİ UYGULAMASI

YÜKSEK LİSANS TEZİ

Alperen KAÇAR

Bilgisayar Mühendisliği Ana Bilim Dalı  
Bilgisayar Mühendisliği Programı

Temmuz , 2019



T.C.  
İSTANBUL AYDIN ÜNİVERSİTESİ  
FEN BİLİMLERİ ENSTİTÜSÜ



SES VE GÖRÜNTÜ İŞLEME EŞLEŞTİRMESİ VE İŞİTME ENGELLİLER  
İÇİN BİR KONUŞMA TERAPİ UYGULAMASI

YÜKSEK LİSANS TEZİ  
Alperen KAÇAR  
Y1113.010028

Bilgisayar Mühendisliği Ana Bilim Dalı  
Bilgisayar Mühendisliği Programı

Tez Danışmanı: Prof. Dr. Ali GÜNEŞ

Temmuz / 2019







T.C.  
İSTANBUL AYDIN ÜNİVERSİTESİ  
FEN BİLİMLER ENSTİTÜSÜ MÜDÜRLÜĞÜ

**Yüksek Lisans Tez Onay Belgesi**

Enstitümüz Bilgisayar Mühendisliği Ana Bilim Dalı Bilgisayar Mühendisliği Tezli Yüksek Lisans Programı Y1113.010028 numaralı öğrencisi **Alperen KAÇAR** ' ın "**SES VE GÖRÜNTÜ İŞLEME EŞLEŞTİRİLMESİ VE İŞİTME ENGELLİLER İÇİN BİR KONUŞMA TERAPİ UYGULAMASI**" adlı tez çalışması Enstitümüz Yönetim Kurulunun 12/06/2019 tarih ve 2019/12 sayılı kararıyla oluşturulan jüri tarafından ..... ile Tezli Yüksek Lisans tezi olarak ..... edilmiştir.

Öğretim Üyesi Adı Soyadı

İmzası

Tez Savunma Tarihi : 09/07/2019

1) Tez Danışmanı: Prof. Dr. Ali GÜNEŞ

2) Jüri Üyesi : Dr. Öğr. Üyesi Ahmet GÜRHANLI

3) Jüri Üyesi : Doç. Dr. Metin ZONTUL

*[Handwritten signatures of Prof. Dr. Ali GÜNEŞ, Dr. Öğr. Üyesi Ahmet GÜRHANLI, and Doç. Dr. Metin ZONTUL]*

Not: Öğrencinin Tez savunmasında **Başarılı** olması halinde bu form **imzalanacaktır**. Aksi halde geçersizdir.



## YEMİN METNİ

Yüksek Lisans tezi olarak sunduğum “ Ses ve görüntü işleme eşleştirmesi ve işitme engelliler için bir konuşma terapi uygulaması ” adlı çalışmanın, tezin proje safhasından sonuçlanmasına kadarki bütün süreçlerde bilimsel ahlak ve geleneklere aykırı düşecek bir yardıma başvurulmaksızın yazıldığını ve yararlandığım eserlerin Bibliyografya’da gösterilenlerden oluştuğunu, bunlara atıf yapılarak yararlanılmış olduğunu belirtir ve onurumla beyan ederim. (.../.../20..)

Aday / İmza







*Eşime ve Oğluma,*



## ÖNSÖZ

Bu çalışmada ses ve görüntü işleme yöntemleri kullanılarak işitme engelli bireyler için konuşma terapi yazılım uygulaması gerçekleştirilmiştir.

Öncelikle tez konusunu seçerken isteklerimi göz önünde bulundurarak bana yardımcı olan, çalışmamın planlanması, araştırılması, yürütülmesinde ve oluşumunda ilgi ve desteğini esirgemeyen tez danışmanım Prof. Dr. Ali GÜNEŞ'e, bilgi ve tecrübelerinden yararlandığım Doç. Dr. Metin ZONTUL'a teşekkürlerimi sunarım.

**Temmuz 2019**

**Alperen KAÇAR**





## İÇİNDEKİLER

### Sayfa

<b>ÖNSÖZ</b> .....	<b>ix</b>
<b>İÇİNDEKİLER</b> .....	<b>xi</b>
<b>KISALTMALAR</b> .....	<b>xv</b>
<b>ÇİZELGE LİSTESİ</b> .....	<b>xvii</b>
<b>ŞEKİL LİSTESİ</b> .....	<b>xix</b>
<b>ÖZET</b> .....	<b>xxiii</b>
<b>ABSTRACT</b> .....	<b>xxv</b>
<b>1. GİRİŞ</b> .....	<b>1</b>
1.1 Problemin tanımı .....	1
1.2 Literatür taraması .....	2
1.3 Amaç .....	6
<b>2. GENEL BİLGİLER</b> .....	<b>9</b>
2.1 Konuşma Üretim Süreci .....	9
2.2 Konuşma Algılama Süreci .....	12
2.3 İşitme Engellilerde Dil ve Konuşma .....	15
2.4 İşitme Kaybı Olan Çocuklarda Dil ve Konuşma Gelişimi .....	16
2.5 Konuşma Tanıma .....	17
2.5.1 Konuşma tanıma sistemlerinin sınıflandırılması .....	19
2.5.1.1 Ayrık sözcük tanıma sistemleri.....	19
2.6 Konuşma Terapisi .....	22
2.7 İnsan Dudak Okuma Becerileri .....	22
<b>3. SAYISAL SİNYAL İŞLEME</b> .....	<b>25</b>
3.1 Giriş.....	25
3.1.1 Konuşma üretimi.....	25
3.1.2 Konuşma tanıma .....	26
<b>4. GÖRÜNTÜ İŞLEME</b> .....	<b>29</b>
4.1 Giriş.....	29
4.2 Sayısal Görüntü İşleme .....	29
<b>5. SES İŞLEME</b> .....	<b>33</b>
5.1 Giriş.....	33
5.2 Sıkıştırma .....	34
5.3 Konuşma Sentezi ve Tanıma.....	34
5.4 Doğrusal Olmayan Ses İşleme .....	36
5.5 Niceleme.....	37
5.6 Özellik Vektörlerinin Çıkarımı .....	40
5.6.1 Çerçeveleme.....	41
5.6.2 Pencereleme .....	42
5.6.3 FFT dönüşümleri.....	42
5.6.4 Mel frekansı saptırması.....	43
5.7 Ses İşleme Uygulamaları.....	43
5.7.1 Ses sınıflandırması.....	43
5.7.2 Ses parmak izi.....	44

5.7.3 Ses bölümlenme .....	44
5.7.4 Ses kaynağı ayrımı .....	44
5.7.5 Müzik önerisi .....	45
<b>6. KONUŞMA TERAPİSİ .....</b>	<b>46</b>
6.1 Giriş .....	46
6.2 Dil Edinimi .....	48
6.2.1 Nöropsikolojik açıdan dil edinimi .....	48
6.2.2 Dil sisteminin işleyişinde dilbilimsel bakış açısı .....	49
6.3 Konuşma Eylemi .....	51
6.4 İşitme Engellilerin Dil ve Konuşma Becerilerinin Geliştirilmesi .....	52
6.4.1 Artikülasyon eğitimi .....	53
6.4.2 Fonemlerin karıştırılması .....	53
6.5 Dilin Öğretilmesi .....	53
6.6 Konuşmanın Öğretilmesi .....	55
6.7 Konuşma Üretim Gelişiminde İşitsel Geri Bildirim .....	56
<b>7. KONUŞMA TANIMA.....</b>	<b>58</b>
7.1 Giriş .....	58
7.2 Görüntü Ön İşleme .....	59
7.2.1 Yüz lokasyon tespiti .....	60
7.2.1.1 Bilgi tabanlı yöntemler .....	61
7.2.1.2 Özellik tabanlı yöntemler .....	62
7.2.1.3 Şablon eşleme yöntemleri .....	64
7.2.1.4 Görünüm tabanlı yöntemler .....	65
7.2.2 Dudak lokasyon tespiti .....	65
7.2.2.1 Model tabanlı dudak algılama yöntemleri .....	66
7.2.2.2 Görüntü tabanlı dudak algılama yöntemleri .....	71
7.3 Görüntü Özellik Çıkarımı .....	73
7.3.1 Görünüm Özelliklerine Dayalı Yaklaşımlar .....	73
7.3.2 Geometrik özelliklere dayalı yaklaşımlar .....	73
7.3.3 Görüntü dönüşümü temelli yaklaşımlar .....	74
7.4 Ses Ön İşleme .....	74
7.4.1 Filtreleme .....	80
7.4.2 Çerçeveleme .....	82
7.4.3 Pencereleme .....	84
7.4.4 Hızlı Fourier Dönüşümü (FFT) .....	86
7.4.5 Mel frekansı kepsral katsayıları .....	87
<b>8. KONUŞMA TERAPİSİ UYGULAMA TASARIMI.....</b>	<b>91</b>
8.1 Sistem Mimarisi .....	92
8.2 Konuşma Eğitimi .....	93
8.3 Görüntü Ön Uç .....	94
8.3.1 Ön işlem adımı .....	95
8.3.1.1 Yüz tespiti .....	96
8.3.1.2 Ağız tespiti .....	98
8.3.1.3 Ağız bölgesinin yeniden boyutlandırılması .....	100
8.3.1.4 Rgb-gray dönüşümü .....	100
8.3.2 Görsel özellik çıkarımı .....	101
8.4 Ses Ön Uç .....	104
8.4.1 Ses Kaydı .....	107
8.4.2 Ses özellik çıkarımı .....	107
8.4.3 Ses kayıtlarının grafiksel gösterimi .....	109

8.4.4 Sistemin eğitilmesi.....	110
8.4.5 Ses Tanımlaması.....	112
8.4.6 Seslerin karşılaştırılması.....	113
8.5 Konuşma Seslerinin Karşılaştırılması.....	118
8.5.1 HMM Sınıflandırıcısı.....	119
8.5.2 Dinamik zaman atlama (DTW).....	122
<b>9. SONUÇLAR VE ÖNERİLER.....</b>	<b>129</b>
<b>KAYNAKLAR .....</b>	<b>135</b>
<b>EKLER.....</b>	<b>147</b>
<b>ÖZGEÇMİŞ.....</b>	<b>159</b>





## **KISALTMALAR**

<b>AAM</b>	: Active Appearance Models
<b>ADC</b>	: Analog Digital Converter
<b>ASM</b>	: Active Shape Models
<b>ASR</b>	: Automatic Speech Recognition
<b>AVSR</b>	: Audio Visual Speech Recognition
<b>CMYK</b>	: Cyan, Magenta, Yellow and Black
<b>DCA</b>	: Digital Analog Converter
<b>DCT</b>	: Discrete Cosine Transform
<b>DTW</b>	: Dynamic Time Warping
<b>DWT</b>	: Discrete Wavelet Transform
<b>FFT</b>	: Fast Fourier Transform
<b>HCI</b>	: Human Computer Interface
<b>HMM</b>	: Hidden Markov Model
<b>HSV</b>	: Hue, Saturation and Value
<b>KNN</b>	: K-Nearest Neighborhood
<b>LDA</b>	: Linear Discriminant Analysis
<b>LPC</b>	: Linear Prediction Coding
<b>LPCC</b>	: Linear Prediction Cepstral Coefficient
<b>MFCC</b>	: Mel Frequency Cepstral Coefficient
<b>RGB</b>	: Red, Green and Blue
<b>STAR</b>	: Speech Training Aid Research
<b>SVM</b>	: Support Vector Machine



## ÇİZELGE LİSTESİ

### Sayfa

<b>Çizelge 8.1</b> : HMM ile öğrenci ve eğitimci konuşma uzaklık değerleri.....	120
<b>Çizelge 8.2</b> : “Abi” kelimesi’nin eğitimciler için DTW uzaklık değerleri .....	125
<b>Çizelge 8.3</b> : “Bak” kelimesi’nin eğitimciler için DTW uzaklık değerleri.....	126
<b>Çizelge 8.4</b> : “Beyaz” kelimesi’nin eğitimciler için DTW uzaklık değerleri .....	126
<b>Çizelge 8.5</b> : Kelimelerin öğrenci ve eğitimci konuşma uzaklık değerleri.....	127







## ŞEKİL LİSTESİ

### Sayfa

Şekil 2.1: Konuşma işlemi.....	9
Şekil 2.2 : Ses telleri ve gırtlak.....	11
Şekil 2.3 : Ünsüzlerin artikülasyon şeklinde sınıflandırılması.....	12
Şekil 2.4 : Kulağın yapısı.....	13
Şekil 2.5 : İnsan kulağının fonksiyonel diyagramı (Smith, 1999).....	14
Şekil 2.6 : Ses yoğunluğu (Smith, 1999).....	15
Şekil 2.7 : Ses teknolojisi alanları.....	19
Şekil 2.8 : Sözcük tabanlı sesli ifade tanıma.....	20
Şekil 2.9 : Fonem tabanlı sesli ifade tanıma.....	21
Şekil 2.10 : Makine için akustik ve dil modellerinin eğitimi.....	21
Şekil 2.11 : Makine ile konuşma tanıma işlemi.....	21
Şekil 3.1 : Analog ve Sayısal sinyal dönüşümü.....	25
Şekil 3.2 : Konuşma tanıma sistemi.....	26
Şekil 4.1 : Bir görüntünün sayısallaştırılması.....	30
Şekil 4.2 : Siyah ve beyaz bir görüntünün kodlanmış görüntüsü.....	31
Şekil 5.1 : İnsan konuşma modeli.....	35
Şekil 5.2 : konuşma ve gürültü spektrumu. (Smith, 1999).....	36
Şekil 5.3 : Analog dijital dönüşümü ve nicelemesi.....	37
Şekil 5.4 : Analog Sinyal.....	38
Şekil 5.5 : Dijital Sinyal.....	38
Şekil 5.6 : Örnekleme.....	39
Şekil 5.7 : (a) Bir niceleyicinin doğrusal olmayan karakteristik eğrisi. (b) Niceleme hatası( $e$ ) ve olasılık yoğunluk fonksiyonu (PDF) $PE(e)$ .....	39
Şekil 5.8 : Özellik vektörü çıkarımı.....	41
Şekil 5.9 : 20-30 milisaniye arasındaki bir ses sinyalinin detaylı gösterimi.....	41
Şekil 5.10 : Hamming penceresi.....	42
Şekil 5.11 : Hamming penceresi uygulama örneği.....	42
Şekil 6.1 : İşitme kaybı seviyeleri.....	47
Şekil 7.1 : Görsel konuşma tanıma sistemi.....	60
Şekil 7.2 : (a) $n = 1$ , orijinal görüntü. (b) $n = 4$ . (c) $n = 8$ . (d) $n = 16$ . Orijinal ve karşılık gelen düşük çözünürlüklü görüntüler.....	61
Şekil 7.3 : Bilgi tabanlı yukarıdan aşağıya yöntemlerinde kullanılan tipik bir yüz (Yang ve Huang, 1994).....	62
Şekil 7.4 : Yatay ve dikey profiller (Yang ve diğ., 2002).....	63
Şekil 7.5 : 14x16 piksel oranı şablon yöntemi.....	65
Şekil 7.6 : Snakes yaklaşımı kullanarak dudak lokasyon bölümlenmesi.....	68
Şekil 7.7 : Snakes yaklaşımı ile örnek bir dudak bölümlenmesi (Kass ve diğ., 1988).....	68
Şekil 7.8 : Dudak tespiti için geleneksel ASM sonucu (Jang, 2007).....	69
Şekil 7.9 : Şekil ve gri seviye görünüm modeli.....	70
Şekil 7.10 : AAM araştırması örneği.....	70

Şekil 7.11 : Ağızın deforme edilebilir şablonlar tarafından çıkarılması (a) ağız kapalı deforme edilebilir şablon, (b) ağızın açık deforme edilebilir şablon (Zhang, 1997)...	71
Şekil 7.12 : Konuşma dalgabıçımı .....	75
Şekil 7.13 : Konuşma Tanıma Süreci .....	76
Şekil 7.14 : Merhaba kelimesinin aynı kişi tarafından farklı söylenişi .....	77
Şekil 7.15 : Merhaba kelimesinin dalga formu kesiti .....	78
Şekil 7.16 : Ses dalgasının tek boyutlu yükseklik gösterimi .....	78
Şekil 7.17 : Ses dalgasının bir kesitinin örneklenmiş gösterimi .....	78
Şekil 7.18 : Merhaba Ses dalgasının ilk 4 örnekleme değerleri.....	78
Şekil 7.19 : İlk 20 milisaniyelik değerler.....	79
Şekil 7.20 : Değerlerin gösterimi.....	79
Şekil 7.21 : Enerji değerleri .....	80
Şekil 7.22 : Ses frekans değerleri.....	80
Şekil 7.23 : Merhaba Ses dosyasının tam spektrogramı .....	80
Şekil 7.24 : Ön vurgulama Matlab kodu.....	81
Şekil 7.25 : Ön vurgulama işlemi etkisi.....	82
Şekil 7.26 : Örtüşen çerçevelerin görünümleri .....	82
Şekil 7.27 : Altta yatan çerçevelerin görünümleri .....	82
Şekil 7.28 : Hamming penceresi Matlab kodu.....	84
Şekil 7.29 : Hamming penceresi görseli .....	85
Şekil 7.30 : Bir sinyalin penceresi .....	85
Şekil 7.31 : FFT uygulanmış ses dalgası .....	87
Şekil 7.32 : Konuşmacı tanımlaması için kullanılan yaklaşım şeması .....	88
Şekil 7.33 : MFCC hesaplama adımları.....	88
Şekil 7.34 : Mel filterbank görseli. ....	89
Şekil 8.1 : Sistem mimarisi .....	93
Şekil 8.2 : Görsel konuşma eğitim kaydı.....	93
Şekil 8.3 : Konuşma eğitim uygulama görüntüsü.....	94
Şekil 8.4 : Görsel Ön uç işleme adımları .....	94
Şekil 8.5 : Görsel özellik çıkarım işlemi.....	95
Şekil 8.6 : Haar cascade sınıflandırıcı eğitimi .....	96
Şekil 8.7 : Viola ve Jones algoritması temel prensibi.....	97
Şekil 8.8 : Sistem yüz tespit görseli .....	98
Şekil 8.9 : Viola-Jones algoritması ile ağız tespiti sonrası .....	98
Şekil 8.10 : Yüz oran bilgileri.....	99
Şekil 8.11 : Dudak tespiti.....	99
Şekil 8.12 : Konuşmada yakalanan dudak görüntüleri .....	100
Şekil 8.13 : Dudak bölgesinin gri formata dönüşüm sonucu.....	100
Şekil 8.14 : Görünüm ve şekil tabanlı görsel özellik çıkarım işlemi .....	101
Şekil 8.15 : 2D-DCT Görsel özelliklerin çıkarılma işlemi .....	102
Şekil 8.16 : Görüntünü DCT metoduna gönderimi.....	103
Şekil 8.17 : Mavi kelimesinin seçili çerçevelerdeki DCT çıkarım sonucu.....	104
Şekil 8.18 : Ses sinyali dalga formunun kararlı yapısı .....	104
Şekil 8.19 : Ses özelliklerinin çıkarımı .....	105
Şekil 8.20 : Sistemin görsel gösterimi .....	105
Şekil 8.21 : Konuşma kaydı .....	107
Şekil 8.22 : Ses bilgisi özellik çıkarım modeli .....	108
Şekil 8.23 : İki ses sinyalinin MFCC değerlerinin grafiksel gösterimi .....	109
Şekil 8.24 : Sesin grafiksel gösterimi .....	109
Şekil 8.25 : Sesin grafiksel gösterimi .....	110

Şekil 8.26 : Gerçek zamanlı konuşmanın ses dalgası ve spektrum gösterimi .....	110
Şekil 8.27 : Sisteme tanıtılacak örnek kelimeler .....	110
Şekil 8.28 : Kelimeleri sisteme tanıtmak .....	111
Şekil 8.29 : Özellik vektörleri.....	111
Şekil 8.30 : Sistemin test dosyası .....	112
Şekil 8.31 : Matlab bağlantı kodları .....	112
Şekil 8.32 : Matlab path toolbox ekleme .....	113
Şekil 8.33 : Ses bilgilerinin sisteme yüklenmesi .....	113
Şekil 8.34 : Sinyal gürültü temizleme metodu.....	114
Şekil 8.35 : Ses sinyalinin temizlenmiş görüntüsü .....	114
Şekil 8.36 : Ses bilgilerinin grafiksel gösterimi.....	114
Şekil 8.37 : Ses sinyalinin sayısal benzerlik gösterimi.....	115
Şekil 8.38 : İki sinyal arası zaman serisi hesaplama.....	115
Şekil 8.39 : DTW metodu ile iki sinyalin karşılaştırılması .....	116
Şekil 8.40 : İki farklı sesin benzerlik değeri .....	116
Şekil 8.41 : Aynı kişi tarafından söylenen aynı kelimenin benzerlik değeri.....	117
Şekil 8.42 : Aynı kişi tarafından söylenen aynı kelimenin benzerlik grafiği .....	117
Şekil 8.43 : Aynı kişi tarafından söylenen aynı kelimenin MFCC değer grafiği ....	118
Şekil 8.44 : İki farklı sesin uzaklığı .....	118
Şekil 8.45 : HMM Akustik modelin eğitim diyagramı.....	120
Şekil 8.46 : DTW ile farklı kişilerin aynı kelimeyi söyleyişi .....	122
Şekil 8.47 : DTW ile aynı kişinin aynı kelimeyi söyleyiş gösterimi .....	123
Şekil 8.48 : DTW ortak bir zaman ekseninde genişletilmesi.....	123
Şekil 8.49 : DTW sistemlerindeki temel izole kelimeler.....	124
Şekil 8.50 : DTW sistemlerindeki temel izole kelime tanıma .....	124
Şekil 8.51 : İki nokta arasındaki uzaklığın iki boyutlu uzayda gösterimi .....	125
Şekil 9.1 Araba kelimesinin eğitici ve öğrenci ses bilgisinin karşılaştırılması.....	129
Şekil 9.2 Kal kelimesinin eğitici ve öğrenci ses bilgisinin karşılaştırılması.....	129
Şekil 9.3 Baba kelimesinin eğitici ve öğrenci ses bilgisinin karşılaştırılması .....	129
Şekil 9.4 Beş kelimesinin eğitici ve öğrenci ses bilgisinin karşılaştırılması .....	130
Şekil B.1 : Araba kelimesinin eğitici ve öğrenci ses bilgisinin karşılaştırılması ....	151
Şekil B.2 (devam) : Kal kelimesinin eğitici ve öğrenci ses bilgisinin karşılaştırılması .....	151
Şekil B.3 (devam): Baba kelimesinin eğitici ve öğrenci ses bilgisinin karşılaştırılması .....	151
Şekil B.4 (devam) : Araba kelimesinin eğitici ve öğrenci ses bilgisinin zamansal eşitlemesi.....	152
Şekil B.5 (devam) : Beyaz kelimesinin eğitici ve öğrenci ses bilgisinin zamansal eşitlemesi.....	152
Şekil B.6 (devam) : İki kelimesinin eğitici ve öğrenci ses bilgisinin zamansal eşitlemesi.....	153
Şekil B.7 (devam): Kal kelimesinin eğitici ve öğrenci ses bilgisinin zamansal eşitlemesi.....	153
Şekil B.8 (devam) : Eğitici ve öğrenci ses bilgisi karşılaştırma .....	154
Şekil B.9 (devam): Altı Kelimesinin eğitici ve öğrenci ses spektrogramları.....	154
Şekil B.10 (devam) : Baba Kelimesinin eğitici ve öğrenci ses spektrogramları.....	155
Şekil B.11 (devam) : Kal Kelimesinin eğitici ve öğrenci ses spektrogramları .....	155
Şekil B.12 (devam) : Kırmızı Kelimesinin eğitici ve öğrenci ses spektrogramları	156
Şekil B.13 (devam) : Mor sesinin diğer sesler ile DTW mesafesi.....	156
Şekil B.14 (devam) : Abi sesinin diğer sesler ile DTW mesafesi.....	157

<b>Şekil B.15 (devam) :</b> Mavi sesinin diğer sesler ile DTW mesafesi .....	157
<b>Şekil B.16 (devam) :</b> Bir sesinin diğer sesler ile DTW mesafesi.....	158
<b>Şekil B.17 (devam) :</b> Araba sesinin diğer sesler ile DTW mesafesi .....	158



## SES VE GÖRÜNTÜ İŞLEME EŞLEŞTİRMESİ VE İŞİTME ENGELLİLER İÇİN BİR KONUŞMA TERAPİ UYGULAMASI

### ÖZET

Kendi kendine öğrenebilen, karar verebilen akıllı sistemler çağı yaşanmaktadır. Teknolojik gelişmelerin, pratik ve teorik çalışmalar ile iş dünyasının ve toplumun her kesimini dönüştürmeye yönelik olarak birçok farklı alanda kullanıldığı görülmektedir. Artık bilgisayarlar onlara yapmasını söylediğimiz şeyden çok daha fazlasını yapma becerisi ve kapasitesine sahiptirler.

Birçok alanda bilgisayar sistemleri hızları ve doğruluk dereceleri göz önüne alındığında hızlı çözümler geliştirmekte ve son derece başarılı olmaktadır. Eğitim, sağlık, savunma ve güvenlik alanlarında insan gibi düşünebilen, karar verebilen akıllı sistemler, insanların ihtiyaçlarına yönelik çözümler üretilmesini hedeflemektedir.

Bu çalışmanın amacı, işitme engelli bireylerin konuşma ve dil kazanımı gelişimine destek sağlamaktır. Bu bağlamda ses ve görüntü işleme yöntemleri kullanılarak bireyin konuşma sırasındaki ses bilgisi ve görünür dudak hareketlerine ait özelliklerini çıkaran ve görsel geri bildirim sağlayan bir yazılım gerçekleştirilmesi hedeflenmektedir. Gelecekte bu teknolojilerin, işitme engelli bireylerin dil ve konuşma yeteneklerinin geliştirilmesinde önemli bir işlem adımı olarak görülmektedir.

**Anahtar Kelimeler:** *Ses ve görüntü işleme, konuşma tanuma, bilgisayar destekli konuşma terapisi*



## **AUDIO AND IMAGE PROCESSING MATCHING AND A SPEECH THERAPY APPLICATION FOR HEARING-IMPAIRED PEOPLE**

### **ABSTRACT**

We are living in the era of smart systems, which can learn and make decisions on its own. It is seen that technological developments are used in many different fields in order to transform all segments of the business world and the society with practical and theoretical studies. Computers now have the ability and capacity to do more than what we tell them to do.

In many areas, computer systems develop fast solutions and are extremely successful considering their levels of speed and accuracy. Smart systems that can think and make decisions like humans in the fields of education, health, defense and security aim to produce solutions for the needs of people.

The purpose of this study is to support the development of speech and language acquisition of hearing-impaired individuals. In this context, it is aimed to realize a software which provides visual feedback and extracts the characteristics of the individual's audio information and visible lip movements during the speech by using audio and image processing methods. In the future, these technologies are seen as an important operational step in improving the language and speaking skills of hearing-impaired individuals.

**Keywords:** *Audio and image processing, speech recognition, computer assisted speech therapy*





# 1. GİRİŞ

## 1.1 Problemin tanımı

İnsanlar sosyal varlıklardır. Konuşarak duygu ve düşüncelerini başka insanlara ileterek etkileşim ve paylaşım içerisindedirler. İşitme engelli bireyler konuşma kabiliyetlerine sahip olsalar bile geri dönütlerdeki eksiklikler onların konuşma ve dil gelişimini etkileyerek konuşma bozukluğuna sebep olmaktadır. Bu olumsuz etki ileriki yaşlarda akademik başarılarına, sosyal çevreden kopuşa, iş bulma ve sürdürme üzerinde olumsuz etkiye sebep olmaktadır.

Türkiye’de her yıl yaklaşık olarak 1000 ile 2000 arasında çocuk işitme engeliyle doğmakta veya konuşma ve dil öğrenmeye başlamadan önce işitme bozukluğu yaşamaktadır. Ağır işitme engelli çocuklar işitme cihazlarından minimum düzeyde faydalanmaktadır. 18-48 aylık çocuklar tipik olarak anormal sesler çıkarabilmektedir. Ancak yaşlarına kıyasla neredeyse kelime dağarcığına sahip değildirler. Bu sorun işitme cihazı ile çözülmeye çalışılsa da işitme cihazı, daha büyük çocuklar ve yetişkinlerin konuşma kalitelerini artırma ve daha akıcı bir etkileşim için yetersiz kalmaktadır (Bernstein ve diğ., 1988).

İşitme engelli bireyler üzerinde gerçekleştirilen araştırmalar ve raporlar göstermektedir ki engelli bireyler sosyal çevre ve toplumla tam olarak etkin katılım sağlayamamaktadır. Ancak engelli bireylerin, toplumsal yaşama tam ve etkin katılımlarını sağlayacak ve sosyal çevrelerini geliştirecek ulusal ve uluslararası düzeyde somut adımlar atmaları beklenmektedir. Yayınlanan raporlarla birlikte ulusal ve küresel düzeylerde eyleme geçilmesi tavsiye edilmektedir.

İletişim sisteminde dil ve konuşma birbiriyle bütünleşik kavramlardır. İletişimde var olan dil, işaretler ve sözcükler gibi temel birimlerden oluşan bir dizgedir (Baykoç, 1987). Konuşma ve dil gelişiminde, çocuğun sesleri taklit ederek konuşma hatalarını düzeltmesi önemli faktörler arasındadır. İşitme engelli bireyler için işitsel geri bildirim eksikliği, konuşma engeline neden olmaktadır. Bu nedenle işitme engelliler

uygun konuşma üretim mekanizmasına sahip olmalarına rağmen konuşma gelişimini gösteremezler.

İşitme engeliyle doğan çocuklar, başkalarının konuşmalarını taklit etmek ve karşılaştırmak için sınırlı bir akustik konuşma hedefine sahiptir. İşitme engellilerin farklı duyu organları konuşmayı öğrenen çocukların duyu organı gibi işitsel geri bildirim yerine geçmesi gerektiği belirtilmektedir (Engwall ve diğ., 2006). İşitme engeliyle dünyaya gelen çocuklar, konuşma taklitlerini öğrenmek ve konuşma hareketlerini geliştirmek için konuşmanın fonetik özelliklerin görsel ipuçlarına güvenmek durumundadır. İşitme engelli bireyler, dudak hareketlerini taklit edebilselerde, doğru ses üretimi ve tonlamada yeteri düzeyde gelişim gösterememektedirler.

İşitme engelli bireyler nadir olarak kendiliğinden konuşma gerçekleştirirler. Konuşmaların anlaşılabilirliğini geliştirmek için konuşma terapistleri eşliğiyle konuşma eğitimi almalıdırlar. Kelimelerin telaffuzu konuşmacının yaşına, lehçesine, cinsiyetine, ruh haline, sağlık ve vb. koşullara bağlı olarak büyük ölçüde değişir. Çocukların dil gelişiminde önemli bir süreç adımı konuşma seslerinin fonemlere göre kategorik bir algı oluşturmaktır. Bir dildeki telaffuz eğitimindeki en büyük zorluk öğrenciye bilinmeyen ayrımları öğretmektir.

Çocuklarda tespit edilen işitme kaybının derecesi ve işitme kaybının tipi ile erken tanı ve uygun yöntemler belirlenerek kişiye özel yapılacak eğitim ve rehabilitasyon hizmetleri, işitme kaybının neden olduğu dil üzerindeki olumsuz etkilerden çocuğun daha az etkilenmesini sağlayacaktır. Bu nedenle işitme engelli bireylerin konuşma sesleri ve dudak hareketlerinin görsel geri bildirimlerle desteklenecek bir sistemin kurulması onların dil ve konuşma gelişimi ile kendi ihtiyaçlarının karşılanmasına destek olmak, çocuğun sosyal hayatta ve yaşamında daha olumlu etkiler sağlayacaktır.

## **1.2 Literatür taraması**

Makineler tarafından konuşma tanıma çalışmaları kronolojik olarak bakıldığında zaman 1936 yılında AT&T Bell Labs tarafından üretilen Voder isimli elektronik konuşma sentezleyicisi olduğu görülmektedir. 1950' yıllardan başlayarak 1952 Bell Laboratuvarlarında izole rakam tanıma amaçlı kişiye bağımlı sistem geliştirilmiştir.

Örüntü tanıma kavramlarının konuşma tanıma sistemleri için kullanılması ilk olarak Velichko ve Zagoruyko (1970) tarafından önerilmiştir. 5000 kelimenin kullanıldığı deneylerde elde edilen bir konuşmacı için tanınmanın ortalama güvenilirliği 0,95. Tanıma için hesaplanan süre ise 2-4 saniyedir. 1971 yılında DARPA tarafından sürekli konuşma tanımayı anlayacak bilgisayar sistemi geliştirme çabaları, 1975’li yıllarda IBM araştırma merkezinde de konuşma tanıma üzerine çalışmalar başlatılmış ve 80’li yıllarda kısa ofis yazışmaları için ilk dikte sistemi çalışması yapılmıştır (Gürel ve Arslan, 2016). 1978 yılında Texas Instruments tarafından “Speak and Spell” için bir konuşma çipi tasarlanmıştır. 1984 yılında SpeechWorks şirketi kurularak telefon üzerinden otomatik konuşma tanıma sistemleri geliştirilmiştir. 1995 yılında Dragon Systems tarafından kelime tabanlı dikte yazılımı piyasaya sürülmüştür.

1990 yılından itibaren DRA konuşma araştırma birimi, küçük çocuklar için konuşma ve dil gelişimine yönelik konuşma tanıma teknoloji uygulamalarını araştırmıştır. STAR adıyla bilinen ilk proje, öğretmenlerin küçük çocuklara çeşitli iletişim ve dil becerilerinin geliştirilmesinde yardımcı olması için bir “konuşma eğitimi yardımı” aracına gereksinim duyduklarına dair farkındalığı ile desteklenmiştir. Amaç, işitme engelli çocuk tarafından konuşulan kelimenin “iyi” ya da “kötü” telaffuzlarını ayırt edebilen bilgisayar destekli bir sistem geliştirmektir (Russell ve diğ. , 1996). Aynı konuşma tanıma teknolojisi daha sonraki yıllarda animasyonlu “Konuşan Kitaplar” serisine dahil edilmiştir.

Bilgisayar destekli konuşma ve dil gelişimi uygulamaları çok büyük bir potansiyele sahip olmakla birlikte işitme engelli bireyin daha fazla eğitim yardımı almasını desteklemektedir.

Adams ve diğ. (1989), işitme yada konuşma engelliler için konuşma sinyali, ses perdesi, ses yoğunluğu, ses spektrum gibi ses özelliklerini açıklayan IBM Speech Viewer konuşma sistemini tanıtmışlardır. Bu sistem ile konuşma ifadelerini gerçek zamanlı görsel geri bildirimler gerçekleştirilmiştir.

Javkin ve diğ. (1994), İşitme engelli çocukların konuşma eğitim sistemleri için ses parametrelerini modelleyen, konuşmadan metne dönüştürmek için akustik bilgileri kullanan yeni bir yöntem benimsemiştir. Daha önceki yapmış oldukları görsel artikülator ve akustik analizlerden türetilmiş bir dizi konuşma parametreleri

sunmuşlardır. Geliştirdikleri yeni modelde ise öğretmenin konuşması sisteme girilerek çocuğun izleyebileceği bir eğitim modeli kullanmışlardır. Sonraki süreçte çocuklara, konuşmalarının geribildirimleri sağlanmıştır.

Vicsi ve diğ.(2000), görsel-işitsel telaffuzu öğretme ve eğitim metodu üzerine bir yazılım gerçekleştirmeyi amaçladılar. Yöntem, konuşma ve işitme bozukluğu olan çocukların konuşma üretimlerini daha iyi kontrol etmelerine yardımcı olmak için tasarlanmıştır. İngilizce, İsveççe, Slovence ve Macarca, bireysel ses hazırlık aşamasından dört dilde cümleler halinde sesleri uygulamak için hazırlanmıştır. Sistem, genel bir dilden bağımsız olarak ölçme ve veritabanı işlemlerini gerçekleştirir. Veritabanı ile tüm katılımcı diller ve farklı ses grupları için modüller oluşturmayı mümkün kılmıştır.

Bunnell ve diğ. (2000), konuşma eğitimi, değerlendirme ve düzeltme sistemi, konuşma ve dil patolojilerine, eklemleme sorunları olan çocukların tedavisinde yardımcı olmak için etkileşimli bir video oyunu şeklinde tasarlanmıştır. Seçilen kelimelerin anlamalarını öğretmeyi amaçlamaktadırlar. Kelimeyi öğretebilmek için çocuklara, hecelerden başlayarak ve sözcük / ifadelere doğru ilerleyen, birbirini takip eden aşamalı konuşma görev dizisi verilmiştir. Kelimelerin söyleyişini değerlendirmek için gizli markov model (HMM) kullanılmıştır.

Hsiao ve diğ. (2001), işitme engelli çocukların konuşma eğitimi ve öğrenmesi için bilgisayar destekli bir yazılım tasarladılar. Bu yazılımda, işitme engelli çocukların konuşma becerilerini eğitmek ve geliştirmek için görsel bir asistan yöntemi kullanılmıştır. Yazılım ile animasyonlu görüntüleme sistemi oluşturulmuştur. Sistem, artikulatörler arasındaki ilişkileri dinamik bir şekilde göstermektedir. Konuşma sırasında burun, ağız, dil ve dudakların pozisyonu ve değişimi farklı profillerden gösterilerek işitme engelli çocukların doğru konuşma tekniğini yakalayabilmesi amaçlanmıştır.

Engwall ve diğ. (2006), yapmış oldukları çalışmada ARTUR adlı bilgisayar destekli konuşma eğitimi yardımcısının insan-bilgisayar arayüzünü değerlendirmeyi amaçladılar. Çalışmaların ana özelliği konuşmadaki eklem bölgelerinin nasıl iyileştirileceği konusunda önerilerde bulunabilmesidir. İki kullanıcı grubu 9-14 yaşları arasındaki bilgisayar destekli yoğun konuşma eğitimi tecrübesine sahip ve 6 yaşından küçük, önceden bilgisayar destekli konuşma eğitimi tecrübesi olmayan üç

çocuk ve terapistlerden oluşmaktadır. Bütün çocuklar, genel olarak dil bozukluklarına sahiptir. Yapmış oldukları çalışmalarda, arayüzlerin küçük çocuklar için daha önceden eğitim veya talimat olmadan da kullanılabilir olduğunu ancak daha fazla motivasyon faktörünün olması gerektiği belirtilmiştir.

Werda ve diğ. (2006), insan konuşma kalitesine ilişkin deneysel karşılaştırmaları yapmışlardır. Konuşma uyaranları, konuşmalarının kalitesini bilerek değiştiren iki normal konuşmacıdan ve Hintçe konuşma eğitimi yardımı İSTRA üzerinde konuşma terapisi alan işitme engelli bir çocuktan elde edilmektedir. İşitme engelli konuşmacının söylediği bir kelimeyi daha önceden elde edilmiş bir şablonla ne kadar iyi eşleştirdiğini göstererek puan üreten bir bilgisayar sisteminden oluşmaktadır. Verilerin korelasyonel analizi, konuşma kalitesini değerlendirmede insanların bilgisayardan biraz daha iyi olduğunu ancak bilgisayarların çok daha güvenilir olduğunu göstermiştir. Bilgisayar destekli konuşma terapilerinin insan yargılamasının alternatifi olabileceği belirtilmektedir.

Huang ve diğ. (2009), işitme engelli çocuklara yönelik olarak konuşma rehabilitasyonu için yeni bir dil eğitim sistemi ele alınmıştır. Sistem, Nios işlemci ile donatılmıştır. Çocuk seslerini tek bir fonem ile tanıma temeline dayanmaktadır. Sesler arasındaki benzerliği yakalamak için LPCC kullanılmıştır.

Hu ve diğ. (2013), işitme engelli çocukların dil becerilerini geliştirmeyi amaçladılar. Geleneksel yaklaşım olarak öğretmenlerin hareketlerinin kullanılması, verimsiz bir yol olarak görülmüştür. STAS adını verdikleri konuşma eğitim yardımı sistemi önerilmiştir. Çin'de işitme engelliler okulunda deneysel bir çalışma yürütülmüştür. Katılımcılar iki deneyimli okul öğretmeni ile yirmi öğrenciden oluşmaktadır. Çalışmalarının deneysel sonuçlarında, çocukların konuşmalarının doğruluğunu ve akıcılığını özellikle STAS kullanarak iyileştirdiği belirtilmiştir.

Wankhede (2014), işitme engellilerin konuşma eğitimi yardımları için görsel geri bildirim sağlayan öğrenme, konuşma ve iletişim kurma fırsatını yaratmayı amaçlamıştır. Farklı yaşlardaki konuşan çocuklardan ses yolu şeklini elde ederek LPC analizi kullanarak bir uygulama gerçekleştirilmiştir.

Lijuan ve diğ. (2015), çalışmalarında işitme engelli çocuklar için konuşma eğitimlerine yönelik olarak bir solunum eğitim sistemini araştırmışlardır. Çocukların konuşma eğitimi sürecinde ortaya çıkan nefes ile ilgili sorunların çözümüne yönelik

olarak bir eğitim sistemi kullanılmıştır. Sistem etkileşimli olarak yarı saydam bir solunum sistemi modeline ve çoklu sensörlere dayalı solunum eğitim ölçüm sisteminden oluşmaktadır. Geliştirmiş oldukları sistemde nefes bilgileri hakkında veri toplayarak ölçüm işlemi gerçekleştirilmiştir. İşitme engelli çocuklar, telaffuz sürecini taklit etmek için solunum eğitim modelini kullanabileceği belirtilmiştir. Sistem, konuşmacıların nefesleri hakkında veri toplayabilir, ölçebilir ve işitme engelli çocukların solunum eğitiminde kullanılabilir. Rehabilitasyon eğitimlerinden sonra net, doğru ve yüksek sesler çıkaran çocuklar için bir temel hazırladığı belirtilmiştir.

Dünya üzerinde İngilizce, Çince, Japonca, Hintçe dilinde farklı ülkelerde farklı uygulamalar ile birlikte işitme engelli bireyler için geliştirilen konuşma terapi uygulamaları ve çalışmaları bulunmasına karşın ne yazık ki ülkemizde Türkçe için yapılan çalışmalar diğerleri ile karşılaştırıldığında oldukça düşüktür. Konuşma tanıma teknoloji çalışmaları yapılmasına karşın, konuşma terapi, konuşma gelişimi üzerine Yalçın (2008) tarafından “Konuşma teknolojisi yardımıyla ilköğretim birinci sınıf öğrencilerine ilkokuma yazma öğretimi için bir yazılım geliştirme” adlı doktora teziyle konuşma tanıma teknolojisini ilköğretim alanına uygulanmasından daha başka uygulamaların da gerçekleştirilmesi gerekmektedir.

### **1.3 Amaç**

Geçtiğimiz yıllar boyunca işitme engelli bireylere yönelik olarak konuşma eğitimi yardımlarının bir dizi kapsamlı incelemesi ortaya çıkmıştır. Konuşma terapi yöntemlerinde gelişimsel düşünceler ve eğitim yardımcılarının en etkin kullanımı olarak çocukluk yıllarında ve özellikle de okul öncesi dil edinimi döneminde yapılması gerektiği belirtilmesine karşın bilgisayar destekli konuşma sistemlerinin dilbilimsel ve bilişsel özelliklere bakılmayarak tasarlandığı görülmektedir.

Klasik konuşma terapisi eğitimlerinde, tonlama, ses çalışmaları, nefes eğitimleri, konuşma egzersizleri görsel geri bildirimle verme imkanına sahip değildir. Konuşma terapi eğitimlerinde, çocuğa çıkarmış olduğu sesleri ve yapmış olduğu dudak hareketlerini görsel geri bildirimlerle desteklemek durumundadır. Bu nedenle konuşma terapi eğitimlerinde temel ses frekans düzeyleri, dudak ve ses yolu hareketlerini görsel geri bildirimler şeklinde verilmelidir.

Konuşma gelişiminde doğru geri bildirim ve tekrarlanan eğitimler, konuşmanın otomatikleşmesini sağlamak için gereklidir. Ancak bunu gerçekleştirecek olan terapistin zaman kısıtlaması, nedeniyle zordur. Bilgisayar destekli konuşma terapi uygulamaları öğrenci ve öğretmen için kolaylık sağlayabilecek bir sistem kurulması için gereklidir. Özellikle öğretmenin öğrencinin sürekli olarak dudak hareketlerini ve ses çıkışını kontrol etmesi anlamında zorluk yaşamasına neden olmaktadır. Konuşma sırasında söylenen kelimenin doğru yada yanlış olduğunu öğretmek ve açıklamak için çocuğa anında ve anlamlı görsel geribildirim sunmalıdır.

Diğer bir sınırlılık olarak belirtilen çocuğun konuşma terapisini yeteri düzeyde alamadığı bu nedenle konuşma terapisi potansiyel olarak etkili bir şekilde konuşma eğitimi yardımı içerse bile zaman yetersizliğinden kaynaklı mevcut faydanın sınırlı olduğu belirtilir (Bernstein ve diğ. , 1988). Bir öğrencinin terapi sırasında ilerlemesini devam ettirmesi için eğitimi sürekli olması gerekmektedir. Kısa seanslarla haftada yalnızca bir ya da iki sefer gerçekleştirilen bireysel konuşma terapisi gelişim adına yeterli görülmemektedir.

Konuşma eğitimlerinde karşılaşılan sorunlardan bazıları da teknik ve yeterli ekipmanın olmayışı. Geçmişten bu yana geliştirilen cihazların dikkatli ve çok fazla ayar gerektiği için kullanımının oldukça zor olduğu belirtilmektedir. Ancak teknolojik gelişmelerle birlikte kullanımı kolay, hızlı ve yararlı sistemler geliştirilmelidir.

Konuşma terapi sistemi geliştirilirken çocuğun gelişimsel düşünceler, eğitim yardımlarının en etkili kullanımının çocuk yıllarında ve özellikle de okul öncesi dil edinimi döneminde yapılması gerektiğine işaret edilse de, çoğu yardımcı programın dilbilimsel ve bilişsel özelliklere bakılmaksızın tasarlandığı görülmektedir.

Başka bir problem, konuşma eğitimi dışındaki yardımların sınırlı erişebildiği olmuştur. Türkiye'de konuşma terapistlerinin eksikliği, işitme engelli bireylerin yeteri kadar konuşma eğitimi alamamalarına neden olmaktadır. Konuşma terapi uygulaması, işitme engeliyle birlikte diğer konuşma bozukluğu yaşayan down sendromlu çocuklar, Dizatri ve Apraksi gibi motor konuşma bozukluğu yaşayan bireyler yada farklı konuşma bozukluğuna sahip bireylerin, konuşma ve dil gelişimine katkıda bulunabilir.

Dil ve konuşma eğitimi, süreklilik isteyen bir eğitimidir. İşitme engellilerin bilgisayar destekli konuşma terapi uygulaması, her yerden erişilebilir sürekli ve raporlanabilir nesnel bilgiler içereceği için gelişim tablosu rahat bir şekilde incelenebilir. İşitme engelli bireyin gelişim sistemi, terapistle birlikte öğrenme verimliliğini arttırabilir.

İşitme engellilerin eğitimleri, yaşamda karşılaştıkları engelleri aşabilmeleri ve hayatlarını bağımsız olarak devam ettirebilmeleri açısından önem arz etmektedir. Özellikle diğer bireylerle iletişim kurmada sorunlarla karşı karşıya kalan işitme engelli bireylerin iletişim ihtiyaçlarını gidermek gayesiyle oluşturulan yöntemler, onların hayata tutunmasını ve daha fazla bağlanmasını kolaylaştıracak seviyededir.

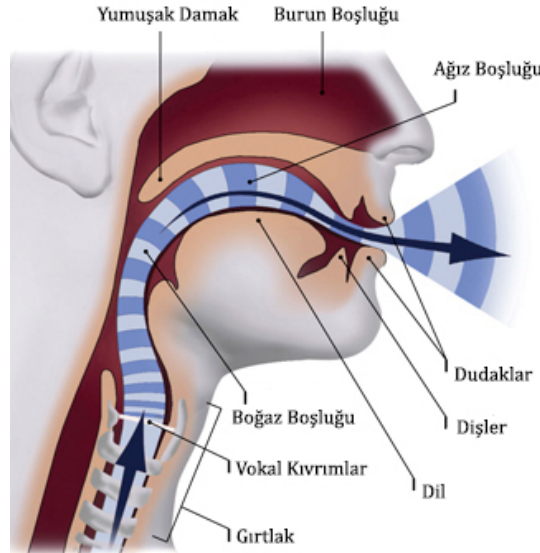
Bu çalışmanın amacı, konuşma eğitim sistemlerinde olduğu gibi görsel geri bildirim sağlayan bir yazılım ile işitme engelli bireylerin konuşma işlemlerini doğru şekilde öğrenmesini sağlamak ve öğrenme, iletişim kurma becerilerinin gelişimine katkı sunmaktır. Çalışma, eğitici, öğrenci ve veli arasında çok işlevli bir asistan sistemi olarak nitelendirilebilir. Böylece işitme engelli bireylerin gelişimi için aile ve eğitim kurumlarına yardımcı bir sistem sağlanmış olacaktır.



## 2. GENEL BİLGİLER

### 2.1 Konuşma Üretim Süreci

Dünyanın bütün sesleri arasında, insan için en önemlisi konuşma sesleridir (Neyman ve Bogomilsky, 2001). İnsanın düşüncelerini sözel olarak kompleks bir şekilde gerçekleşmesi konuşma olarak nitelendirilir Gerçeker ve diğ. (2000). Kendilerini ifade etmek için insanlar, kelimelerden oluşan bir çok farklı konuşma sesini üretme yeteneğine sahiptir. Bu yetenek, Şekil 2.1'de gösterildiği gibi beyin, akciğerler, gırtlak, ses telleri ve hareketli artikülasyon topluluğu gibi vücudun bir çok kısmı arasındaki karmaşık etkileşim sonucunda gerçekleşir. Akustik bakış açısından ise konuşma farklı sürelerdeki duraklamalar tarafından kesilen çeşitli seslerin bir akışıdır. Gerçeker ve diğ. (2000) ' a göre insan sesi bedenimizin türlü perdelerde çeşitli tınılarda kişiden kişiye değişen ve kendine özgü bir şekli olduğunu belirtir. Konuşma seslerinin özellikleri yükseklik, güç, tını ve süre gibi akustik farklılıklara göre belirlenir (Neyman ve Bogomilsky, 2001) .



Şekil 2.1: Konuşma işlemi

Vokal kıvrımların birbirine yakın olması sonucu havanın sağladığı basınçtan dolayı bir ton çıkarmak için titreşim işleminin gerçekleşmesi fonlama olarak tanımlanır. Hareketli artikülasyon konumlarıyla yutak, burun, ağız gibi oyuk bölgelerin boyutu

ve şekli tonun frekanslarını yükseltir. Bu şekilde dilin her fonetik işlemi için benzersiz bir ses üretir.

Konuşma sesi, üç farklı fonksiyonel birim tarafından desteklenen karmaşık hareketler sonucunda oluşan bir hava dalgasıdır (Honda, 2008). Konuşma için akciğer hava basıncı kısa bir solunumdan sonraki solunum sisteminin işlevlerinden kaynaklanmaktadır. Nefes veriş sırasında akciğerlerden gelen hava ses tellerini titreştirerek ham sesi oluşturmaktadır. Ses tellerinin saniyedeki titreşimleri erkeklerde 100-150, kadınlarda 200-250, bebeklerde 400 keredir. Ses tonu ise ham sesin boğaz, burun ve ağız boşluklarında şekillenmesiyle meydana gelir. Kişiyeye özgü ses, ağız içinde dil, diş, dudak, damak, yumuşak damak da çeşitli pozisyonlara girmesiyle konuşma sesine dönüşür.

Araştırmacılar konuşma üretiminin farklı birimlerin karşılıklı olarak etkileşimde olduğunu ve etkinleşme seviyesine göre karar verilerek konuşmanın gerçekleştiğini belirtir. Diğer bir yaklaşımda ise konuşmanın, konuşma birimleri arasında ayrı ayrı gerçekleştiğini belirtir (Altıparmak, 2017).

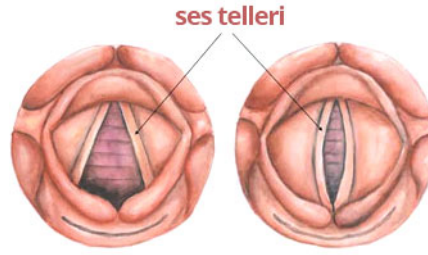
Konuşma ile düşüncelerin organize ve sembolik bir şekilde sözel olarak ifade edilmesi insanın fikirlerinin tanımlanmasını sağlayan bir araç olarak nitelendirilir. Konuşma işlemi sırasında dolaylı yada dolaysız olarak vücudumuzun büyük bir bölümü bu işleme katılmaktadır. Konuşma fizyolojik ve psikolojik yönleri bulunan kompleks ve genel bir beceri gerektiren fiili davranıştır Gerçeker ve diğ. (2000).

Konuşma üretimi için insan sesi farklı aşamalardan oluşmaktadır;

Respirasyon aşaması, solunum aşaması olarak da bilinir. Soluk alma ve soluk verme fazlarından oluşur. Soluk alma sırasında diyafram düştüğünde meydana gelir. Akciğerlerin atmosferik olarak hava basıncına kıyasla akciğerlerde negatif bir basınç meydana gelir. Ağız ve burun boşluklarına, trakeden aşağıya ve akciğerlere dolar. Soluk verme sırasında ise solunumdan sonra akciğerlerde pozitif basınçla sonuçlanır. Hava yolu açıksa dış ve iç basıncı eşitlemek için ciğerlerden dışarı doğru bir hava çıkışı gerçekleşir.

Fonasyon aşaması, ses tellerinin yerleştiği gırtlakta meydana gelen titreşim aşamasıdır. Akciğerlerden gelen hava ses telleri aracılığıyla zorlandığında ses telleri titreterek fonlamayı oluşturur. Şekil 2.2'deki gırtlakta üretilen ses, ses tellerinin gerginliğine bağlı olarak değişir. Ses tellerinin uzama ya da gerilimi daha yüksek

frekansların oluşmasına neden olurken kısılma ve gevşemesi daha yavaş titreşimlerle daha düşük frekanslara neden olur.



**Şekil 2.2 : Ses telleri ve gırtlak**

Rezonans aşaması, oluşan ham sesin şekle girerek insana özgü ses tonunun oluşum aşamasıdır. Seslerimizin her birinin rezonans frekansı, bir çello veya bas rezonansının bir gitardan farklı olması gibidir.

Artikülasyon aşaması ise konuşma seslerinin üretim aşamasıdır.

Konuşma üretim aşamasında;

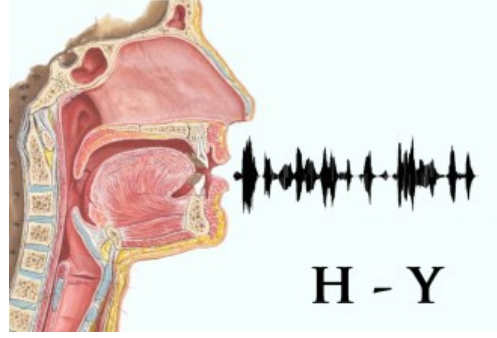
- dil,
- dudaklar,
- dişler,
- damak,
- yumuşak damak / küçük dil yer alır.

Bir fonem (ses birimi) belirli bir konuşma sesi için teknik terimdir. Fonemler sesli ve sessiz harflerden oluşur.

- Sesli Harfler (a, e, ı, i, o, ö, u, ü) .
- Sessiz harfler ise (b, c, ç, d, f, g, ğ, h, j, k, l, m, n, p, r, s, ş, t, y, z)' dir. Ancak sesli harflerin kelimelere bağlı olarak farklı ses çıkışına sahiptir.

Ünsüzlerin artikülasyon şeklinde sınıflandırılması;

- Her iki dudakta meydana gelen : (p,b,m,)
- Diş-dudak ünsüzleri: alt dudak ve üst diş teması (f,v)
- Diş: dil üst dişlerle teması ile (t)
- Diş yuvası: dilin ucu alveolar sırtıyla temas eder (t,d,s,z,n,l)
- Damak: Dil damağa yaklaşması (j,r)
- Dil arkası teması ile (k,g)



**Şekil 2.3 :** Ünsüzlerin artikülasyon şeklinde sınıflandırılması

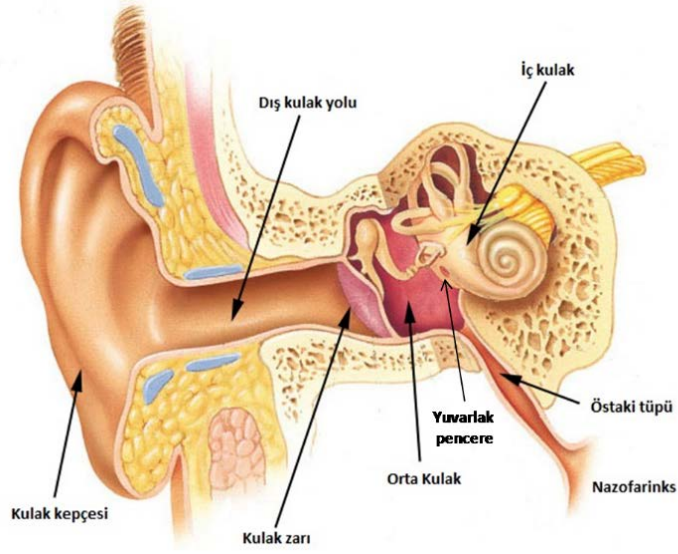
Konuşma sesleri ünlüler ve ünsüzler olarak iki gruba ayrılır. Ünlüler, tonlu sesler iken ünsüzler çoğunlukla gürültülü seslerdir. Ünsüz harflerin y,s,o (200-800) ve ünlülerde ise e,a,i (1500-4200) arasında frekans değerlerine sahiptir. Orta frekanslarda ise bu değer (1000-1400) arasında bulunmaktadır. Sesli harfler arasındaki fark her sesli harfin karakteristiği ile belirlenir.

Unutulmamalıdır ki konuşma bir çok organın ortak çalışması sonucu oluşan bir beceridir. Konuşma eğitiminde, konuşma organlarının konuşmaya hazır hale getirilmesi çalışmaları yapılmalıdır.

## **2.2 Konuşma Algılama Süreci**

Çevremizde sesler havada dolaşan görünmez titreşimlerden meydana gelir. Herhangi bir nedenden dolayı meydana gelen ses, titreşimlerle hava içerisinde tüm yönlere ses dalgaları eşliğinde gönderilir. Ses dalgaları benzersiz bir yapıya sahiptir. Bu sebeple sesler birbirinden farklıdır, kimi ses dalgaları ince, kalın ya da yüksek, yumuşak olabilir. İnsan kulağı etraftaki sesleri yakaladığı zaman beynimiz bu sesleri anlayabileceği mesajlara dönüştürür. Bu ses dalgasının netliği kulağımızın ne kadar iyi çalıştığına da bağlıdır. Kulağımıza gelen sesler insan beyni tarafından şaşırtıcı bir sinir ağında birleştirilerek yorumlanır.

Ses dalgaları havadan sıvıya geçmeye çalıştığı zaman sesin bir kısmı ara yüzden iletilirken enerjinin geri kalanı yansıtılır. Bu havanın düşük bir mekanik empedansa sıvının ise yüksek bir mekanik empedansa sahip olmasından kaynaklanmaktadır (Smith, 1999).



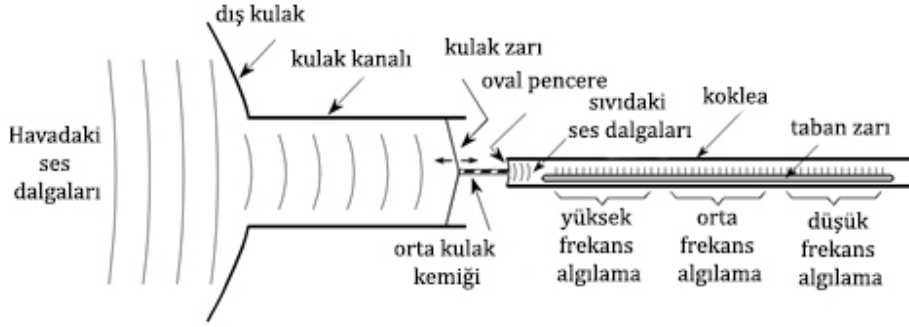
**Şekil 2.4 : Kulağın yapısı**

Dış Kulak, Şekil 2.4’de gösterildiği gibi başın iki yanında bulunan deri ve kıkırdaktan oluşan yaklaşık 3 cm boyunda ve 0,5 cm çapında kulak kanalından oluşmaktadır. Bu yapı çevresel sesleri, kafatası içerisinde bulunan orta ve iç kulak organlarına yönlendirir. Seslerin yoğunlaştırarak toplanmasını ve ses kaynağının belirlenmesini sağlayarak gelen seslerin orta kulağa yönlendirilmesi işlemini gerçekleştirir.

Orta Kulak, Şekil 2.4’de gösterildiği gibi iç kulağın sıvısına giren ses enerjisinin payını arttıran bir empedans eşleştirme ağıdır (Smith, 1999). Empedans dönüşümlerinin çoğu kulak zarı (havadan gelen ses) ve oval pencere arasındaki alan farkından kaynaklanır. Kulak zarı ve oval pencere arasındaki fark ses dalgasının basıncını yaklaşık olarak 15 kat arttırmaktadır. Gelen ses dalgalarını, iç kulak sıvısına mekanik basınç dalgalarını aktarır. Çekiç kemiği, örs kemiği ve üzengi kemiğinden oluşan kulak kemikçikleri bu titreşimi iç kulağa iletir.

İç Kulak, Şekil 2.4’de gösterildiği gibi, koklea içinde yer alan baziler membrandır, koklea siniri oluşturan yaklaşık 12.000 duyuşal hücre için destekleyici yapıdır. Baziler membran oval pencerenin yanında sert ve karşı uca doğru daha esnek hale gelip bir frekans spektrum analizörü olarak hareket etmesine izin verir. Yüksek frekanslı bir sinyale maruz kaldığında, baziler membran sert olduğu yerde rezonans ederek oval pencereye yakın sinir hücrelerinin uyarılmasıyla sonuçlanır. Aynı şekilde, düşük frekanslı sesler, baziler membranın uzak ucundaki sinir hücrelerini uyarır. Bu koklear sinirdeki özel liflerin özel frekanslara cevap vermesini sağlar. Bu

organizasyon yer prensibi olarak adlandırılır ve işitme yol boyunca beyne doğru korunur. Mikrofon gibi görev yapan parçalardan oluşan bu yapı titreşim sonucu oluşan elektrik sinyallerini dönüştürür. Elektrik sinyalleri işitme siniri ile beyne aktarılır.



**Şekil 2.5** : İnsan kulağının fonksiyonel diyagramı (Smith, 1999) .

1. Sesler kulak kanalına girerek dış ortamdan gelen ses dalgaları kulak kanalı içerisinde geçerek kulak zarına çarpar.
2. Kulak zarı ve işitme kemikleri titreşir. Bu ses dalgaları kulak zarını ve orta kulakta bulunan üç kemiği (kemikçikleri) titreştirir.
3. Sıvı orta kulak içinde akar oluşan titreşimler koklea olarak bilinen spiral şekilli iç kulaktaki sıvı içinde hareket eder ve kokleadaki tüylü hücreleri hareket ettirir. Tüylü hücreler hareketi algılar ve bunu işitme siniri için kimyasal sinyallere dönüştürür.
4. İşitme sinirleri beyinle iletişim kurar. Bunun ardından, işitme siniri aldığı bilgileri elektrik darbeleri ile beyne gönderir ve bu darbeler beyinde ses olarak algılanır.

İnsan kulağı tarafından algılanan en zayıf ses 0 dB olarak bilinir. Normal konuşma yaklaşık olarak 60 dB iken işitme güçlüğü çeken insanlarda algılanan ses düzeyi yaklaşık olarak 140 dB civarında gerçekleşir. İnsanların sinyal olarak 1 dB ile meydana gelen değişiklik ile ses şiddetindeki değişikliği tespit edebilir. En hafif fısıltıyla en gürültülü gök gürültüsüne kadar algılanabilecek yalnızca 120 kadar yüksek ses seviyesi bulunmaktadır (Smith, 1999).

	Watts / cm <sup>2</sup>	Decibel	Örnek Ses
Daha Yüksek Sesl	10 <sup>-3</sup>	140 dB	Ağrı
	10 <sup>-4</sup>	130 dB	
	10 <sup>-5</sup>	120 dB	Rahatsızlık
	10 <sup>-6</sup>	110 dB	
	10 <sup>-7</sup>	100 dB	Rock Konseri Gürültüsü
Daha Yumuşak Ses	10 <sup>-8</sup>	90 dB	
	10 <sup>-9</sup>	80 dB	Endüstri Gürültüsü
	10 <sup>-10</sup>	70 dB	
	10 <sup>-11</sup>	60 dB	Normal Konuşma
	10 <sup>-12</sup>	50 dB	
	10 <sup>-13</sup>	40 dB	
	10 <sup>-14</sup>	30 dB	Zayıf Ses
	10 <sup>-15</sup>	20 dB	
	10 <sup>-16</sup>	0 dB	
	10 <sup>-17</sup>	-10 dB	
10 <sup>-18</sup>	-20 dB		

**Şekil 2.6 :** Ses yoğunluğu (Smith, 1999)

Şekil 2.6'da ses şiddeti ile algılanan ses yüksekliği arasındaki ilişkiyi göstermektedir. Ses yoğunluğu desibel (ses gücü yüksekliği) olarak ifade edilmektedir. Ses yoğunluğu, birim alan başına güç (watt / cm gibi) veya daha yaygın desibel adı verilen logaritmik bir ölçekte ifade edilir.

### 2.3 İşitme Engellilerde Dil ve Konuşma

Çocuklarda anadilin ve konuşma becerilerinin gelişiminde işitme önemli bir işlem adıdır. Çocuklar çevrelerindeki konuşmaları dinleyerek taklit ederek öğrenirler. Konuşma döneminde yaşadıkları işitme sorunları konuşma ve artikülasyon gelişimini etkileyerek iletişim engeline yol açar. Bu iletişim engeli, sosyal uyumsuzluk ve eğitim performansı gibi sorunlara neden olur.

İnsanoğlunun, dil ve konuşma biçimi onun toplumun geri kalanıyla iletişim kurmasını sağlamaktadır. Dil edinimi, insanoğlunda genetik olarak doğuştan var olmaktadır (Chomsky, 2001). Dil edinimi ve konuşma eğitimi, etkili bir iletişim kurabilmek için gereklidir. Bu etmenler herhangi bir şekilde zarar görürse yaşamlarında etkin bir şekilde iletişim kurmasını zorlaştıracak veya engelleyecektir. Bunun sonucunda ise birey akademik becerilerinin gelişiminde farklı sorunlarla karşılaşacaktır (Tüfekçioğlu, 1992). İşitme engellilerin dil ve konuşma becerilerini

geliştirme teknikleri gelişimsel bir sıra içinde önce dilin sonra konuşmanın öğretilmesi olarak ele alındığı görülmektedir. (Türk ve Arslan, 2004)

Çocuk doğumdan sonraki dönemde homurtu ve esneme gibi sesler çıkarmaktadır. Bu sesler ileriki yaşlarda çevresinden gördüğü sesleri taklit ederek gelişir ve anlamlı sesler çıkarmaya başlar. Erken çocukluk döneminin dilin kazanılmasında önemli olduğu, herhangi bir nedenle dil gelişiminde geri kalınması durumunda ileriki yaşlarda bir çok sorunla karşılaşacağını dil bilimciler tarafından vurgulanmaktadır (Kol, 2011) .

Akçamete (1993)' e göre iletişim, dil ve konuşma kavramları birbiriyle bağlantılı kavramlar olarak belirtirken, iletişimin dili kullanarak ya da kullanmaksızın insanlar arasındaki duygu, düşünce ve yaşantıların ifade edilmesi olarak nitelendirir.

Okul öncesi dönemde çocuklar dil gelişimini büyük oranda tamamlamaktadır. Çocuk, 5 yaşına kadar olan yaşamında yaklaşık 2000 kelime bilgisine sahip olarak yetişkin gibi anadilini konuşabilir. Doğumdan itibaren çocuğun çıkardığı sesler ileriki zamanlarda anlamlı hale gelmeye başlar. Ancak işitme engeli yaşayan çocuklarda bu sesler ileriki yıllarda geri dönütlerdeki yetersizliklerden dolayı kesilmektedir.

Çocukların dil ve konuşma kazanımlarında çevre ve geri dönütler de önemlidir. İşiten çocuklar çevre uyarıcılarını alabildikleri için kısa zamanda kendi dillerinin gramer yapısını öğrenirler. Aynı zamanda anne, baba ve diğer uyarıcı faktörleri ile birlikte çocuk kısa sürede etkili dönütler ile gelişim gösterir. İşitme engeli ile doğan çocuklarda ise işitmeyle birlikte yaşanan geri dönüt yetersizliği dil ve konuşma gelişimlerinde sorunlara neden olur.

## **2.4 İşitme Kaybı Olan Çocuklarda Dil ve Konuşma Gelişimi**

Fonolojik Gelişim evresinde söz öncesi ve söz düzeyi evleri bulunmaktadır. Söz öncesi dönemde 6-7 aya kadar olan gelişim evresinde işitme engelli bebeklerin vokal gelişimi diğer bebeklerden farklı değildir. Fonasyon, gıgıldama ve genişletme evrelerinde gelişim özellikleri benzerdir. Mırıldanma evresine gelindiğinde, işitsel geri dönütler olmadığı takdirde, işitme engelli bebeklerin babıldamaları azalır.



Söz düzeyi döneminde işitme kaybının türü, derecesi, tanılanma yaşı faktörler göz önünde bulundurulduğunda ilk sözcük evresinde yaşlılarına göre gecikme gözleneceği belirtilmektedir.

Morfolojik gelişim ve sentaks Gelişiminde, kısa ve basit cümleler kullanmayı tercih ederken ilgi tümceciği, edilgen çatı gibi karmaşık cümleleri anlamada ve yazmada güçlük yaşarlar. Özne, nesne, eylem gibi bazı cümle yapılarını sıklıkla kullanırken, edat, yardımcı fiil ve bağlaç gibi yapıları daha az kullanırlar. Cümlede sözcük sonlarını duymada güçlük yaşamalarından dolayı eksik kullanım ya da hatalı anlama sorunu yaşayabilirler.

Semantik gelişiminde sınırlı bir sözcük dağarcığına sahiptirler. Soyut sözcükleri anlama yetenekleri somut sözcüklere göre daha zordur. Mecazı anlam, özdeyiş, benzetmeler ya da çoklu anlama sahip olan sözcükleri anlamada sorun yaşarken uygun bir kullanım şeklini yapamazlar.

Pragmatik yapıda ise söyleşi kurma ve yürütme stratejilerinde, iletişim çeşitliliğinde sınırlılık gösterirken, söyleşi konusu değiştirme ya da bitirmede güçlük yaşarlar.

Konuşma seslerinin ediniminde ilk 6 aylık süreç boyunca çevresinde konuşulan dilde pek çok sesi ayırt edebilirken, ilk yılın sonuna doğru dilin seslerini ayırt ederek o dile ait fonetik kategorileri yeniden düzenler. İşitsel engelinden dolayı konuşma seslerini duymayı ve bir sesi diğerinden ayırt etmeyi engellemesinden dolayı konuşma anlaşılabilirlikleri düşük olmasına neden olur.

## **2.5 Konuşma Tanıma**

İnsan gibi düşünen, karar veren, konuşan sistemlerin gerçekleştirilme çabası geçmişten bugüne kadar bilim adamları tarafından araştırma çalışmaları arasındadır. Konuşma tanıma sistemi, bilgisayarın önceden kaydedilen ses ve metin eğitim setleriyle dile özgü olan akustik ve dilbilgisinin modellenmesi ve bu modelleri kullanarak kaydedilen konuşma içeriğinin metne dönüştürülmesidir (Gürel ve Arslan, 2016).

Konuşma tanıma sistemlerinin geliştirilmesi ve anlaşılması dil bilimi, bilgisayar bilimi, matematik, istatistik, elektrik mühendisliği gibi farklı disiplinler arası bir çalışmayı gerektirmektedir.

İnsanlar tarafından konuşma tanıma kulaktan başlayan sinyal seviyesi yükseltilerek ve frekans incelemesi yapılarak beyne gönderilen ve beyin tarafından dil ve kavram ilişkisi ile anlamlı hale getirilen bir sistem olarak belirtilir (Gürel ve Arslan, 2016). Makineler tarafından baktığımız da ses sinyalleri mikrofon vasıtasıyla elektrik sinyalleri dönüştürülerek ses kartı vasıtasıyla dijital hale çevrilen yanı sayısallaştırarak programlar vasıtasıyla anlamlandırılan bir sistem olarak karşımıza çıkar.

Konuşma tanıma sistemleri birçok aygıtta etkili arabirim olarak yaygın kullanılmaktadır. Kişisel bilgisayarlar, robotlar, cep telefonları ve araç otomasyonları, dikte kontrol sistemleri ile bilgisayarlara komut verme işlemlerinde, sesli aramalar ile telefon üzerinden konuşma örnek olarak verilebilir. Tıp alanında, işitme engellilere yönelik konuşma terapi uygulamaları, çocukların dil gelişimi için dil tanıma sistemleri Russell ve arkadaşları (1996), güvenlik sistemleri gibi farklı alanlarda kullanılmaktadır. İlerleyen yıllarda konuşma tanıma sistemleri hayatımızda daha etkin rol alarak hızlı ve kolay çözümler sunması beklenmektedir.

Konuşma tanıma sistemleri gürültüsüz ortamlarda sadece seslendirme ile tanımada doğruluk oranı %95 üzerinde iken gürültülü ortamlarda konuşma tanıma doğruluk oranı önemli derecede azalmaktadır (Shin ve diğ., 2011). Bunun yanında telaffuz farklılıkları nedeniyle kelimelerin doğru bir şekilde elde edilememesi konuşma tanıma sistemlerinde doğruluk oranının neden olur.

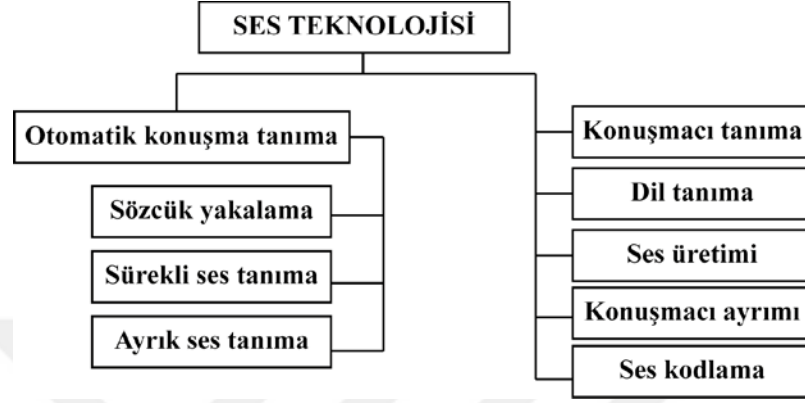
Gelişen teknolojiyle birlikte günümüzde insanlar işlemlerini hızlı ve etkin yollarla gerçekleştirmek istemektedir.

Ses ve konuşma tanıma sistemleri farklı alt problemlere ayrılır;

- konuşmacı belirleme,
- konuşmacı tanıma,
- konuşmacıdan bağımsız tanıma sistemleri,
- konuşmacıya bağımlı tanıma sistemleri,
- ayırık sözcük tanıma,
- anahtar sözcük yakalama,
- sürekli konuşma tanıma sistemleri.

Günümüzde konuşma tanıma ile birçok farklı uygulamalar gerçekleştirilmiş bulunmaktadır. Eğitim alanında ilköğretim öğrencileri için konuşma tanıma

teknolojisi yardımıyla ilk okuma yazma Yalçın (2008), işitme engelliler için konuşma terapi sistemleri, konuşmadan yazıya çevirme, güvenlik sistemleri girişi gibi farklı alanlarda uygulamalar geliştirilmektedir. Konuşma tanıma ile tıp, eğitim ve birçok alanda yeni uygulamaların geliştirilmesi insan-bilgisayar etkileşimini daha farklı bir noktaya taşıyacaktır.



Şekil 2.7 : Ses teknolojisi alanları

### 2.5.1 Konuşma tanıma sistemlerinin sınıflandırılması

Konuşma tanıma sistemleri Şekil 2.7’de gösterildiği gibi ihtiyaçlara göre farklı şekilde sınıflandırılmıştır.

- Ayrık sözcük tanıma sistemleri.
- Sözcük yakalama sistemleri.
- Sürekli konuşma tanıma sistemleri.

#### 2.5.1.1 Ayrık sözcük tanıma sistemleri

İzole edilmiş kelimelerin tanınması diğer sistemlere göre daha kolay bir uygulamadır. Tanıma sözcük temellidir herhangi bir şekilde konuşma sürekliliği bulunmamaktadır. Ayrık sözcük tanıma sözcüklerin birbirinden bağımsız olarak telaffuz edilmesi tanımayı kolaylaştırmaktadır (Yalçın, 2008).

- En yalın konuşma tanıma şeklidir.
- Sözcükler arası duraklar belirgindir ve tanıma kolaylaşmaktadır.
- Sözcükler birbirinden bağımsız olarak ele alınır.
- Seslendirme sorunu yoktur.

Sözcük tanıma sistemlerinin kendi aralarında ayrık sözcük yada sözcük yakalama, sürekli konuşma tanıma sistemlerinin olmasının yanında konuşmacıya bağımlılığına göre ;

- Konuşmacıya bağımlı sesli ifade tanıma sistemleri,
- Konuşmacıdan bağımsız sesli ifade tanıma sistemleri.

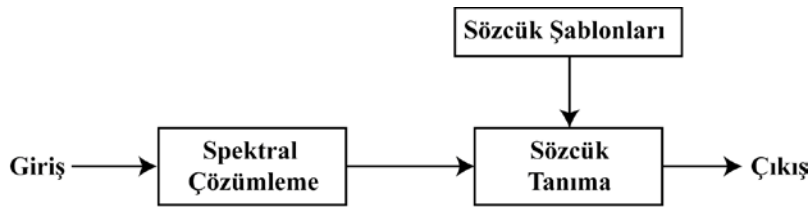
Şeklinde ikiye ayrılmaktadır.

Konuşmacıya bağımlı sistemlerde kişi için deneme kayıtları oluşturulmalıdır. Yeni bir kişi eklendiği zaman kayıtlar tekrar güncellenerek eklenmelidir. Konuşmacıdan bağımsız sistemlerde ise herhangi bir kişi tarafından seslendirilen ifade tanınır. Konuşmacıdan bağımsız sistemlerin oluşturulması konuşmacıya bağımlı sistemlerden daha zor ve zahmetlidir.

Diğer bir sınıflandırma şekli ise tanıma için seçilecek kelime sayılarının büyüklüğü ile ilgilidir (Yalçın, 2008) .

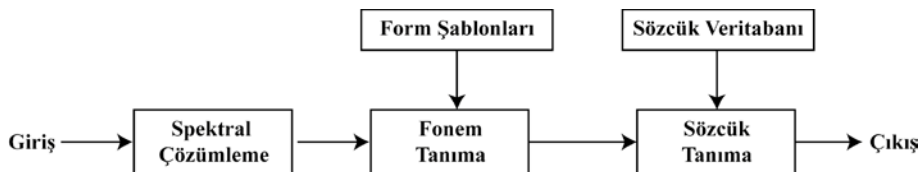
- Sözcük tabanlı tanıma sistemleri,
- Fonem tabanlı sesli ifade tanıma sistemleri.

Sözcük tabanlı tanıma sistemlerinin doğruluk derecesi fonemler arasındaki geçişleri olumsuz etkisinin olmamasından dolayı daha yüksektir (Yalçın, 2008). Ancak sürekli sesli ifade tanıma sistemlerinde sözcükler arası geçişler sorun yaratmaktadır. Ayrıca dilde bulunan çok sayıdaki sözcük sistemin işleyici zamanı ve bellek gereksinimini arttıracaktır.

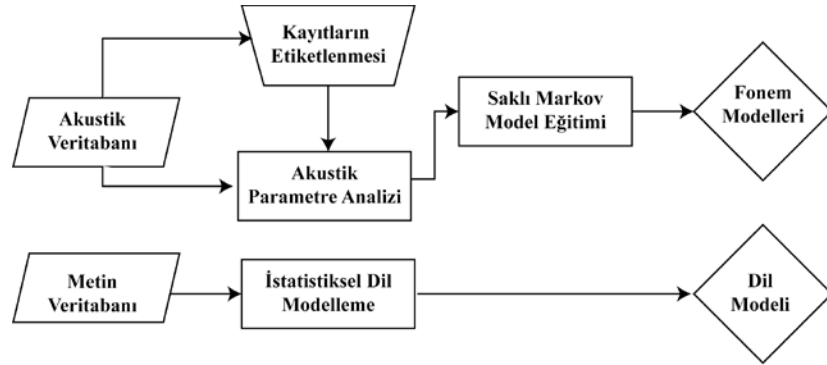


**Şekil 2.8 :** Sözcük tabanlı sesli ifade tanıma

Fonem tabanlı tanıma sistemlerinde ise doğruluk derecesinin daha düşük olmasının yanında az olan fonem sayısından dolayı hızlı sonuç gerçekleşecektir.



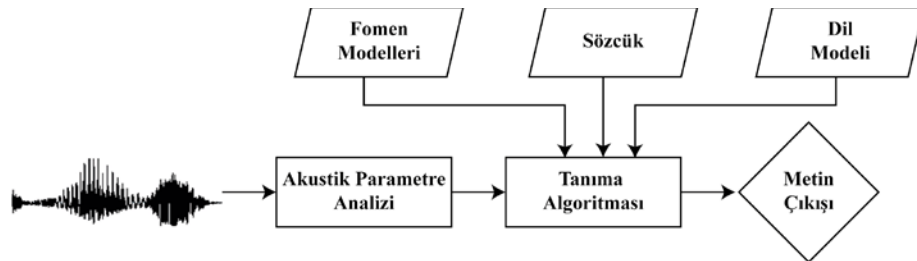
**Şekil 2.9 :** Fonem tabanlı sesli ifade tanıma



**Şekil 2.10 :** Makine için akustik ve dil modellerinin eğitimi

Model eğitim için Şekil 2.10’da yapılan işlemlerin özeti (Gürel ve Arslan, 2016). Model eğitimi iki amaca yönelik olarak yapılmaktadır. Ses birimlerinin akustik incelemesi ile ortalama istatistiklerinin çıkarılması ile farklı seslerin spektrumlarındaki farklılıkların matematiksel olarak modellenmesi. Örnek olarak /a/ sesini /e/ sesinden ayırt edebilmemiz için bu seslerin birçok kişiden örnek alınarak frekans incelemesi yapılması ve her ses için ortalama spektrum ’un elde edilmesi gerekir. Ancak bu bilgilerin elde edilmesi herhangi bir dil için söylenecek cümlelerin sayısının fazlalığından dolayı yüksek konuşma tanıma performansı vermez. Dille ilgili istatistikler sözcüklerin sıklık oranları veya birbirini takip etme olasılıkları akustik olasılığa eklenerek yüksek konuşma tanıma gerçekleştirilmektedir (Gürel ve Arslan, 2016).

Şekil 2.11’de eğitim modelinden sonraki aşama konuşma tanıma işleminin makineler tarafından gerçekleştirme adımlarını gösterir. Mikrofon vasıtasıyla alınan akustik sinyaller elektriksel sinyallere dönüştürülerek ses kartı vasıtasıyla sayısal olarak dijital ortama aktarılır.



**Şekil 2.11 :** Makine ile konuşma tanıma işlemi

Sözcük tanıma sistemlerindeki amaç tüm sözcüğün tanınması işlemidir. Birinci şekilde sesli ifadeden sözcük tanımaya geçilirken ikinci yaklaşımda fonem tanıma işlemi gerçekleştirildikten sonra sözcük veritabanı kullanılarak fonemler sözcüklere dönüştürülmektedir (Yalçın, 2008).

## **2.6 Konuşma Terapisi**

Konuşma sırasında söylenen kelimelerin anlaşılma, harflerin güçlükçe çıkarılması, yanlış harf üretimi gibi problemler konuşma bozuklukları olarak nitelendirilir. Aynı zamanda konuşma sırasında harf, hece yutmaları gibi sorunlarda meydana gelmektedir. Çocukların kelimeleri telaffuz etmedeki konuşma engeli, konuşma terapisi yöntemleriyle dil gelişimi ve konuşma becerilerinin gelişimine yardımcı olabilir. Konuşma terapisi, çocuğun konuşması dil gelişimi, anlama ve ifade etme becerilerini geliştirmeye odaklanan bir eğitimidir.

Konuşma terapistleri çocuğun kelime ve cümleleri oluşturmak için ağzın koordine edilmesini (ekleme, akıcılık, ses seviyesi düzenlemeleri), dil anlama, ifade etme gibi sorunların çözümü için uygun yöntem ve tekniklerin geliştirilmesi için çalışır.

Konuşma terapisi, down sendromlu, işitme engelliler, beyin hasarı geçirmiş ve iletişim yeteneklerini etkileyen enfeksiyon hastalığı yaşamış, bilişsel veya gelişimsel gecikmeler yaşayan bireylerin ihtiyaç duyduğu bir eğitimidir.

## **2.7 İnsan Dudak Okuma Becerileri**

Konuşma, doğası gereği görsel ve işitsel yöntemlerden gelen bir süreçtir. İnsanların görsel ve işitsel olarak gelen uyarıları ayrı ayrı sunulduğu zaman algılanan sesin hatalı olabileceği üzerinde yaptığı çalışmayla bilinen McGurk etkisi olarak da adlandırılan yöntemde görüntü ve ses, konuşmanın tamamlayıcı ikilisidir (McGurk ve MacDonald, 1976).

Dudak okuma görsel konuşma işleme olarak da bilinen, konuşma sırasında dudak hareketlerinin düzenini temel alarak sözel kelimelerin tanınması anlamına gelir (Rathee, 2016).

İnsan konuşması doğada iki modludur. Sesli konuşma ile konuşmacı tarafından üretilen akustik dalga biçimini ifade ederken, görsel konuşma konuşmacının dudak, dil ve diğer yüz kaslarının hareketini ifade eder (Chen, 2001).

Dudak okuma sadece konuşmayı anlamak için dudak hareketleri gibi görsel konuşma bilgisinin kullanılması değil aynı zamanda sesli konuşma tanımayı geliştirmek için özellikle gürültülü ortamlarda yararlı olan görsel bilginin kullanılmasını ifade eder. Dudak okuma görsel işitsel konuşma işlemenin analiz özelliğini temsil etmektedir. Son yıllarda görsel konuşma tanıma sistemleri bilgisayar topluluğunun dikkatini çekmiştir (Chen, 2001).

Dudak okuma becerisine sahip bir kişi, cümlenin anlamını konuşmacının (dil, dudaklar, dişler gibi) konuşmacı artikülasyonlarının hareketine, görsel ve mimik hareketlerine birlikte bakarak anlayabilmektedir. Bu artikülasyonların hareketleri, akustik konuşma sinyalinin içeriği hakkında bilgi verir. İnsanlar görsel bilgileri, sözdizimsel, anlamsal bilgileri birleştirerek konuşmacının yüz hareketlerini gözlemleyerek anlamayı öğrenebilirler. Dudak okuma işitme engelli bireyle için karşı tarafın konuşmasını anlamak için yaygın olarak kullanılmaktadır.

Dudak okumasına ek olarak, insanların sözlü iletişime yardımcı olmak için görüşlerini kullanabilecekleri başka birçok yol vardır. İlk olarak, görsel sinyal dikkatin odaklanmasına yardımcı olmak için kullanılabilir. Konuşmanın arka plan gürültüsünden ayrılmasına yardımcı olmak için hangi konuşmacının konuştuğunu da gözlemleyebilirsiniz. İkincisi, görünür konuşma, dinleyicinin akustik konuşmayı anlamada güçlük çekmesi durumunda faydalı olan ek bir bilgi kaynağı sağlar. Arka plan gürültüsü, konuşmacıların varlığı gibi çeşitli zorluklar ortaya çıkabilir. Bu durumlar için, işitme engelli olmayan insanlar bile, bir ölçüde dudak okuma tekniklerini kullanırlar (Summerfield, 1992) .

Dudak okuması, gürültülü ortamlarda bile bozulmadığı için faydalı olabilmektedir. Yabancı bir dilden gelen cümleler veya aksanlı konuşmacı tarafından konuşulan kelimeler anlaşılması zordur, bu tip durumlarda görsel bilgiler eklenerek daha kolay tanıma işlemi gerçekleştirilir (Reisberg ve diğ., 1987) .

17. yüzyıldan beridir konuşmacının yüz hareketlerinde konuşma hakkında aktarılan yararlı bilgiler olduğu belgelenmiştir. İşitme engelli bireyler dudak okuma tekniklerini başarılı bir şekilde kullanabilmekte ve birçoğu akıcı konuşmaları anlayabilmektedir (Matthews ve diğ. , 2002). Bunun yanında normal işiten bireyler bile konuşmacının yüzünü görebilmenin bile özellikle gürültülü koşullar altında anlaşılabilirliği önemli ölçüde geliştirdiğini belirtmektedir (Erber, 1975).

İşitme engelli kişiler genellikle işitme cihazı aracılığıyla dinlerken konuşmacının dudak hareketlerini ve yüzünü izleyerek konuşmayı algırlar. Normal işiten bireylerde özellikle gürültülü ortamlarda başkasıyla iletişim kurduğu zaman görsel ipuçlarına güvenme eğilimindedirler. Normal işiten ve işitme engelli bireyler üzerinde yapılan çalışmalarda görsel-ışitsel modellerin birleştirilmesiyle tanıma performansı sadece ses ve görme yoluyla algılamaktan daha üstün olduğu ve böylece tanıma olasılığının da artırıldığı belirtilmiştir (Erber, 1975).

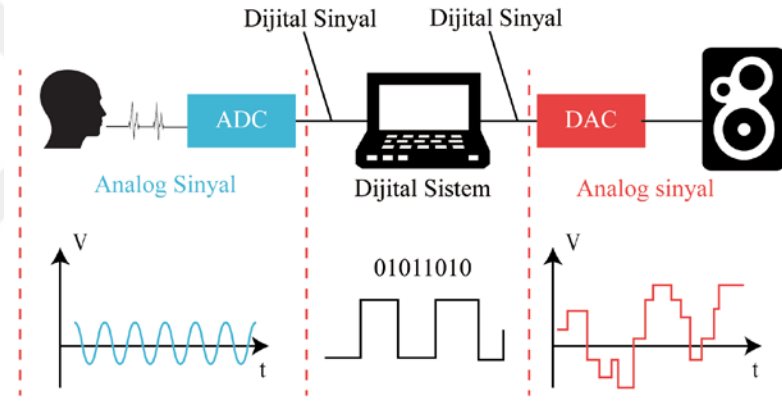




### 3. SAYISAL SİNYAL İŞLEME

#### 3.1 Giriş

Sayısal sinyal işleme, bilim ve mühendislik alanlarında kullanılan ve çeşitli teknolojilerin gelişmesinde ve güçlenmesinde rol oynayan önemli bir işlem adıdır. İletişim, sağlık, robotik, uzay teknolojileri gibi alan çalışmalarında matematik, istatistik ve kendi algoritmaları ile birlikte sayısal sinyal işleme yöntemleri kullanılmaktadır. Örnek olarak Şekil 3.1’de konuşma sinyalinin bilgisayarın anlayacağı bir şekilde sayısallaştırılması gösterilmektedir.



Şekil 3.1 : Analog ve Sayısal sinyal dönüşümü

Sinyal işleme ile ses sinyalleri, elektromanyetik dalgalar, resim gibi öğeler algılanıp dönüştürülerek haberleşme sistemleri, kontrol sistemleri gibi birçok alanda uygulanabilir hale getirilme işlemi gerçekleştirilir.

#### 3.1.1 Konuşma üretimi

Konuşma üretimi ve tanıma sistemleri, insan-makine arasında etkileşim için kullanılan bir yöntemdir. El ve göz kullanımını yerine ağız ve kulakların kullanımını düşündüğümüz zaman eller ve gözlerin başka bir şey yapması gerektiğinde konuşma tanıma sistemleri çok kullanışlıdır. Örnek olarak araba kullanırken, ameliyat yaparken ya da güvenlik sistemlerinde konuşma tanıma kolaylık sağlayan teknoloji

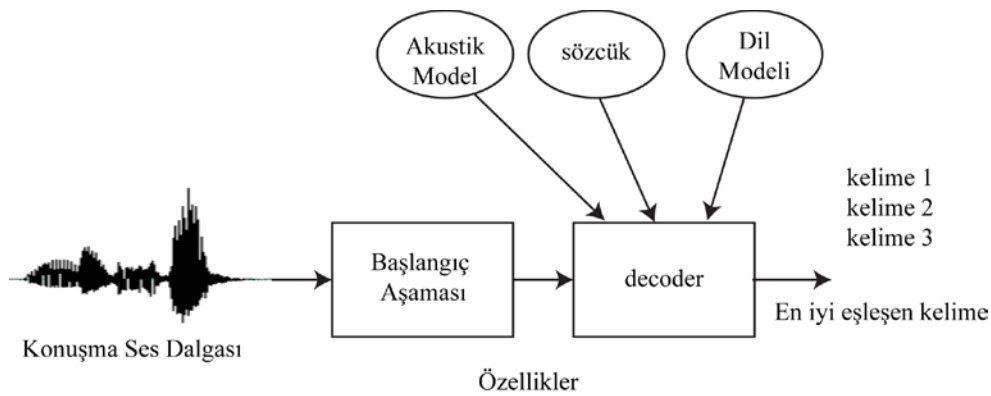
olarak insan kullanımına sunulabilir. Bilgisayar tarafından üretilen konuşma yaklaşımı dijital kayıt ve ses simülasyonları gibi yaklaşımlardır.

Dijital kayıt işlemlerinde insan konuşması genellikle sıkıştırılmış bir biçimde sayısallaştırılarak saklanır. Oynatma sırasında sayısallaştırılan sinyaller tekrar analog sinyallere dönüştürülür. Vokal ses simülatörleri, insan konuşmasının yarattığı fiziksel mekanizmaları taklit ederek konuşma işlemini gerçekleştirirler.

### 3.1.2 Konuşma tanıma

İnsan konuşma tanınması, konuşma oluşturmadan daha zorlu bir süreçtir. Konuşma tanıma, insan beyninin iyi yaptığı şeylerin klasik bir örneğidir. Normal konuşma sırasında başkalarının seslerini kolayca tanıyıp anlayabiliyoruz. Ancak, konuşmacının cinsiyeti, sözlü davranışları ve kelimelerin söyleniş biçimi gibi farklı nedenlerden dolayı insanlar aynı şeyi söylerken bile çok fazla konuşma çeşitliliği vardır. Sayısal bilgisayarlar çok miktarda veri depolayabilir, geri çağırabilir, yüksek hızlarda matematiksel hesaplamalar yapabilir ve bu işlemleri tekrar tekrar hiç yorulmadan yüksek verimli bir şekilde gerçekleştirirler. Bunun yanında bilgisayarlar duygusal verilerle karşılaştığında çok düşük performansa sahiptir. Sonuç olarak bilgisayar sistemleri konuşma tanımada insan kadar doğru konuşma tanıma yeteneğine sahip değildir. Bilgisayar sistemlerine sesin anlaşılması için gerekli işlem adımlarını öğretmek önemli bir sorundur.

Sayısal sinyal işleme ses tanıma işlemi için iki adım gerçekleştirir; özellik çıkarımı ve özellik eşleştirmesi. Şekil 3.2’de gösterildiği gibi gelen ses sinyalindeki her kelime analiz edilerek kelime işleme gerçekleştirilir daha sonra en yakın eşleşmeyi tanımlamak için sözcük örnekleri ile karşılaştırılır.



Şekil 3.2 : Konuşma tanıma sistemi

Günümüz şartlarında kullanıcıların konuşmalarını metne çevrilmesini sağlayan otomatik konuşma tanıma sistemleri geliştirilmiştir. Çok sayıda ücretli ve ücretsiz hizmetler bulunmaktadır.





## 4. GÖRÜNTÜ İŞLEME

### 4.1 Giriş

Görüntü işleme, bir görüntüyü sayısal forma dönüştürerek üzerinde yararlı bilgiler çıkarmak ve gelişmiş bir görüntü elde etmek adına gerçekleştirilen yöntemlerdir. Görüntüler farklı özelliklere sahip birçok sinyal barındırır. Çoğu, sinyal zaman içindeki bir parametrenin ölçüsü iken, görüntüler mesafe üzerindeki bir parametrenin ölçüsüdür (Smith, 1999). Girişin bir video karesi veya fotoğrafın olduğu, çıktının bu görüntü ile ilişkili görüntü veya özellik / özellikler olabildiği bir sinyal işleme türüdür. Genellikle görüntü işleme sistemi, görüntülere önceden ayarlanmış sinyal işleme yöntemlerini uygularken iki boyutlu sinyaller olarak işlemden geçirmeyi içerir.

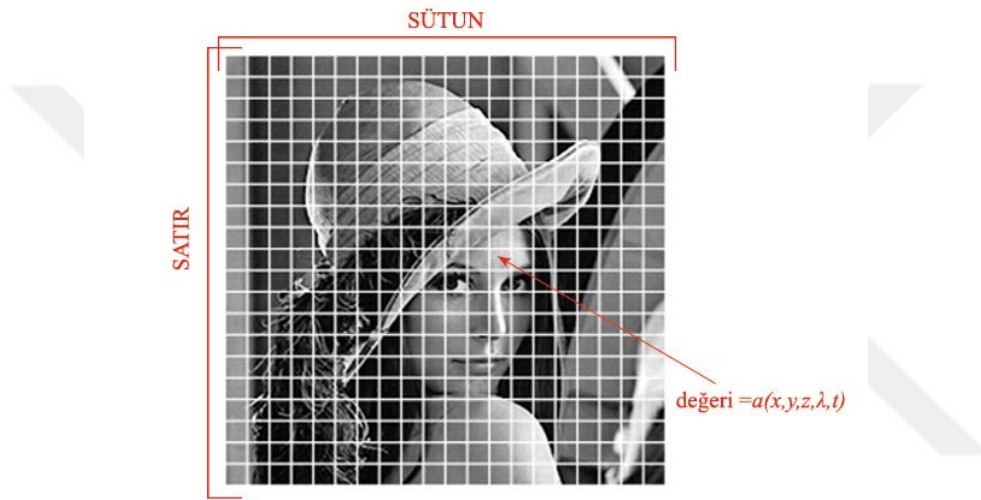
Günümüzde, görüntü işleme hızla büyüyen teknolojiler arasındadır. Mühendislik ve bilgisayar bilimleri disiplinlerinde de temel araştırma alanını oluşturur. Görüntü işleme tıp alanında, 1895'te Wilhelm Conrad Röntgen, x-ışınlarının önemli miktarda maddeden geçebileceğini keşfetti. Tıbbi röntgen sistemleri dünyaya sadece birkaç yılda yayıldı. Uzay teknolojileri alanında ise, kötü bir resimden en iyi şekilde yararlanmanız gerekir. İnsansız uydulardan ve uzay araştırma araçlarından alınan görüntüler için bu olası bir durumdur.

Sayısal sinyal işleme teknikleri, olumsuz koşullar altında çekilen görüntülerin kalitesini çeşitli şekillerde artırabilir: parlaklık ve kontrast ayarı, kenar algılama, gürültü azaltma, odak ayarı, hareket bulanıklığı azaltma, gibi görüntü üzerinde iyileştirme işlemleri gerçekleştirilebilir.

### 4.2 Sayısal Görüntü İşleme

Görüntüler 2 boyutlu olarak bir  $f(x, y)$  fonksiyonu şeklinde gösterilir. Bir  $x, y$  değeri görüntünün düzlem koordinatları olarak nitelendirilir. Bir  $f$  fonksiyonu ise herhangi  $(x, y)$  koordinatındaki yeşinlik veya gri seviyesidir. Sayısal görüntülerin işlenmesi sayısal yeşinlik işleme alanıdır. 2 boyutlu  $f(x, y)$  şeklinde gösterilen bir resim  $N$

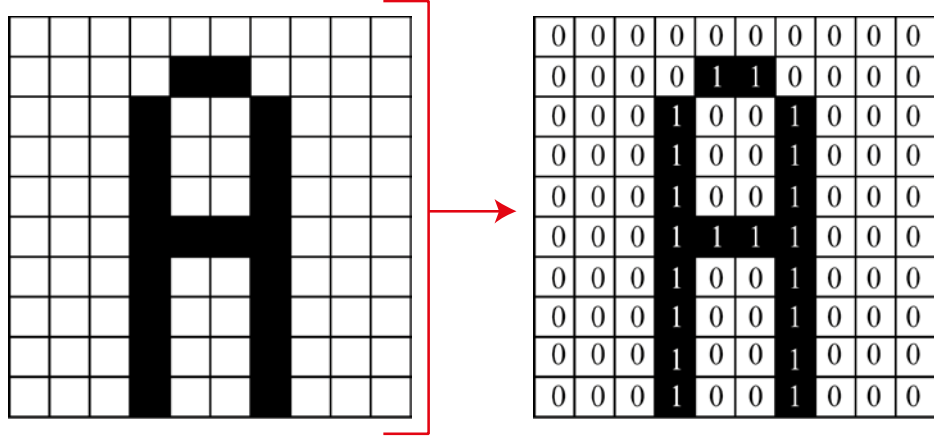
*satır* ve *M sütuna* ayrılır. Bu *N* ve *M* değerleri satır ve sütun olan matris formu analog bir resmin sayısallaştırılmış gösterim formudur. Genel olarak bir  $(x, y)$  yüzeyine çarpan bir sinyal olarak düşünebilir. Şekil 3.1’de, 2 boyutlu *N* satır ve *M* sütunundan oluşmuş bir görüntünün sayısallaştırılması gösterilmektedir. Bir satır ve bir sütunun kesişim bölgesi piksel olarak adlandırılır. Piksel sayısal görüntüleri göstermede yaygın olarak kullanılır. Koordinatlara atanan değerler  $(x, y) \{x = 0, 1, 2, 3, \dots, N - 1\}$  ve  $\{y = 0, 1, 2, 3, \dots, M - 1\}$  bu değerlerin bulunduğu koordinatlar  $a[x, y]$  şeklinde gösterilir. Bu kesimlerde bulunan piksel değerleri derinlik (*z*), renk ( $\lambda$ ) gibi birçok değişken işlevine sahiptir.



**Şekil 4.1** : Bir görüntünün sayısallaştırılması

Şekil 4.1’de gösterilen görüntü  $N = 20$  satır ve  $M = 20$  sütuna bölünmüştür. Koordinatlardaki piksel  $(x = 13, y = 10)$  noktasındaki 100 tam sayı değerine sahip parlaklık göstergesidir. Her piksele atanan değer, en yakın tam sayı değerine yuvarlanan pikseldeki ortalama parlaklıktır.

Görüntünün en küçük birimi piksel olarak tanımlandığını söylemiştik. 800x600 bir resimde 800 sütun ve 600 satırlı bir matris mantığı vardır. Matrisin her bir elemanı 1 pikseli oluşturur. İkili görüntülerde ya da binary image olarak belirtilen görseller siyah ve beyaz piksellerden oluşur. 1 piksel 1 bit yer kaplar. Şekil 4.2’ de siyah ve beyaz pikseller 1 ve 0 ile kodlanmıştır.



**Şekil 4.2 :** Siyah ve beyaz bir görüntünün kodlanmış görüntüsü

Görüntü işleme uygulamaları, görüntü analizi ve bilgisayarlı görme gibi alanlarda yaygın kullanıldığı belirtilirken tam olarak sınırlarının belirlenmesi konusunda herhangi bir görüş birliği bulunmamaktadır. Ancak ortak bir görüş olarak görsel girdilerin kullanımıyla öğrenilme ve sonuç çıkarılma işlemini gerçekleştiren insan görmesine benzetim çalışmalarını oluşturur.





## 5. SES İŞLEME

### 5.1 Giriş

İki ana insan duyusu görme ve duymadır. Buna bağlı olarak, sayısal sinyal işleme'nin çoğu görüntü ve ses işleme ile ilgilidir. Sinyal işleme yöntemleri ses ve görüntü işleme alanlarında devrim niteliğinde değişiklikler yaptı.

Konuşma üretimi özellikle konuşma oluşturma ve tanıma, insan ve makine arasında kurulacak iletişim yönteminin önemli bir kısmını oluşturur. Bilgisayarlar tarafından üretilen konuşma sistemleri dijital kayıt ve ses simülasyonlarını kapsamaktadır. Dijital kayıtlarla insan sesinin sıkıştırılarak sayısal olarak saklama işlemi gerçekleştirilir. Bu sıkıştırılan ve sayısallaştırılan veriler tekrar analog sinyale dönüştürülür. Konuşma simülatörleri insan konuşmalarını taklit etmeye çalışan karmaşık sistemlerdir.

Konuşma tanıma sistemleri konuşma simülasyon sistemlerinden daha zorlu bir sürece sahiptir. Konuşma tanıma insan beyninin gerçekleştirdiği iyi işlerden sadece biridir. Ancak konuşma tanıma bilgisayar sistemleri adına zorlu bir süreçtir. Dijital bilgisayarlar ile büyük miktarda veriler saklanabilir, bir kaynaktan başka bir kaynağa iletebilir, yüksek hızlarda matematiksel hesaplamalar yapılabilir. Bu işlemleri çok hızlı tekrarlar ile verimli bir şekilde yapabilmesine rağmen duyuşal verilerle karşılaştığı zaman ne yazık ki düşük performans göstermektedir.

Sesi bilgisayarla işleme, sesin bilgisayarın anlayabileceği sayılara dönüştürmek demektir. Sayısal sinyal işleme yöntemleri işlem hızı ve bellek kapasitesi yüksek bilgisayarların gelişmesiyle birlikte gelişme göstermiştir. Hemen her alanda sinyal işleme yöntemleri uygulanmakta olup insanın yaşamında önemli bir parça olan sesin işlenmesi de bu alanlardan sadece biridir. Birçok insana ait ses sinyallerini alıp ve bu sinyaller üzerinden işlemler yapılarak konuşma tanıma sistemleri gerçekleştirilir. Bu ses sinyallerinin işlenmesi ile dikte kontrol sistemleri, komut kontrol sistemleri, haber yazım sistemleri gibi farklı sistemler gerçekleştirilir. Dikte sistemleri ile bilgisayar ya da cep telefonu gibi ortamlarda daha hızlı yazım işlemi

sağlanmaktadır. Klavye kullanmadan sadece yazılmasını istediğiniz kelimeyi seslendirerek istenilen kelimeleri bilgisayar ortamına ses işleme yolu ile aktarılır.

Ses sinyal işlemenin genel amacı, seslerin geliştirilmesi, kullanıcı ve alıcılar için akustik koşulların iyileştirilmesini sağlamaktır. Seslerdeki çeşitli parazit sinyallerinin temizlenmesi, çeşitli sinyal kesitlerinde mevcut bulunan eksik ya da karışmış sinyali uygun şekilde değiştirerek sinyal üzerinde analiz işlemlerinin gerçekleştirilmesini sağlar. Örnek olarak herhangi bir müzik parçasındaki melodi, uyum ve ritminde çeşitli sinyal parametreleri kullanılarak elde edilir. Sinyaldeki periyot temel frekans için kullanılabilir. Herhangi iki müzik parçasındaki benzerlik oranı, müzik kategorilerine göre sınıflama, tür sınıflandırması, sanatçı kimliği, alet tanıma, müzik notası gibi çeşitli bilgi çıkarsama işlemi yapılabilir. Yine aynı şekilde ses sinyal işleme ile konuşulan kelimeyle ilgili özellik çıkarımı ve uygulamaları ilgi çekici olabilir. İnsan konuşma diksiyon ile konuşmacının duygusal durumu arasındaki ilişkinin bulunması, konuşmadan kelime tanıma, insan tanıma gibi sistemlerin gerçekleştirilmesi sağlanabilir.

## **5.2 Sıkıştırma**

Telekomünikasyonda veri hızı önemlidir, çünkü doğrudan sinyal iletme maliyeti ile orantılıdır. Bit tasarrufu, para tasarrufu ile aynıdır. Sıkıştırma, niceleme seviyelerindeki ayarlamalar ile ses sinyalinin veri hızını azaltmak yaygın olarak kullanılan bir tekniktir.

Tolere edilebilecek en yüksek ses 120 dB'dir. Bununla birlikte, kulak, yaklaşık 1 dB'den (genlikte %12) daha yakın sesler arasında ayırım yapamaz. Diğer bir deyişle, bir milyon genlik aralığında logaritmik aralıklarla tespit edilip, algılanabilen yaklaşık 120 farklı ses seviyesi vardır (Smith, 1999). Bu işlem sesi sayısallaştırmak için önemlidir. Niceleme seviyeleri eşit aralıklarla yerleştirilmişse, telefon kalitesinde konuşma elde etmek için 12 bit kullanılmalıdır. Ancak, niceleme seviyeleri eşit değilse, insanın işitme özelliklerini eşleştiren yalnızca 8 bit gereklidir. Bu oldukça sezgiseldir; eğer sinyal küçükse, seviyelerin birbirine çok yakın olması gerekir; eğer sinyal büyükse, daha büyük bir aralık kullanılabilir.

## **5.3 Konuşma Sentezi ve Tanıma**

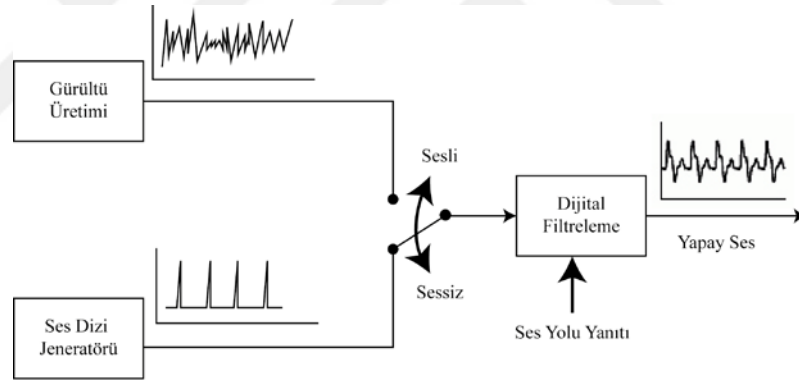
Konuşma tanıma, konuşmacının işitsel sinyallerinin analizi ile söylediği sözleri tanımlama işlemidir. Konuşma tanıma sistemleriyle insan bilgisayar etkileşimi ve

iletişimi gerçekleştirebilir. Konuşma tanıma süreci sesin kayıt edilmesiyle başlar sesin işlenmesi, özelliklerinin çıkarılıp karşılaştırma işlemleri yapılarak konuşmanın tanınması ile son bulur.

Bilgisayar destekli, konuşma tanıma ve üretimi zorlu bir süreçtir; düşük başarılı yöntemlerle farklı yaklaşımlar denenmiştir. Sayısal sinyal işleme, üzerinde araştırmaların gerçekleştirileceği aktif bir çalışma alanıdır.

Konuşma sentezi ve tanıma için tüm teknikler, Şekil 4.1’ de gösterilen insan konuşma üretimi modeline dayanmaktadır (Smith, 1999). Ses, ciğerlerden, ses tellerinden, ağızdan ve / veya burundan dışarı çıktığında meydana gelir. Ses telleri, boyun bölgesinde gırtlığımızın içerisinde gerilmiş iki ince doku kanalıdır. Değişen kas gerginliğine cevap olarak, ses telleri 50 ila 1000 Hz arasındaki frekanslarda titrer ve boğaza periyodik hava püskürmeleri sağlar.

Şekil 5.1’de gösterildiği gibi sesler, ses dizi jeneratörü tarafından temsil edilir, perde (yani, dalga formunun temel frekansı) ayarlanabilir bir parametredir.



**Şekil 5.1 :** İnsan konuşma modeli

Buna karşılık, sürtünme sesleri ses tellerinin titreşiminden rastgele ses olarak çıkar. Bu, hava akışı neredeyse dil, dudaklar ve / veya dişler tarafından engellendiğinde ve daralmanın yakınında hava türbülansına neden olduğunda meydana gelir. Şekil 5.1’de gösterilen modelde, sürtünmeler bir gürültü üretici ile temsil edilmektedir.

Şekil 5.1’deki gösterimde olduğu gibi ses kaynaklarının her ikisi de, dil, dudak, ağız, boğaz ve burun kanallarından oluşan akustik oyuklar tarafından değiştirilir. Bu yapılar yoluyla ses yayılımı doğrusal bir işlem olduğundan, doğrusal bir filtre olarak gösterilebilir. Çoğu durumda, modelde öz yinelenmeli bir filtre kullanılır ve öz nitelik katsayıları filtrenin özelliklerini belirtir. Akustik oyuklar birkaç santimetre boyutlara sahip olduğundan, frekans tepkisi temel olarak kilohertz aralığında bir rezonans

serisidir. Ses işleme teriminde, bu rezonans piklerine format frekansları denir. Dilin ve dudakların göreceli pozisyonunu değiştirerek, format frekansları hem frekans hem de genlikte değiştirilebilir.

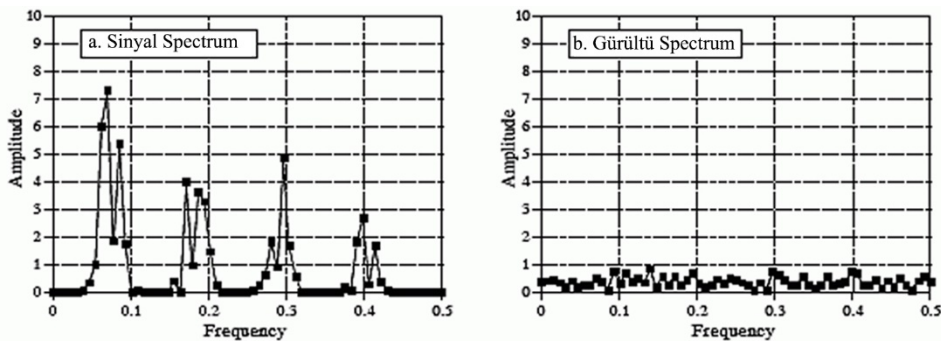
#### 5.4 Doğrusal Olmayan Ses İşleme

Dijital sinyal işleme, zamanla değişmeyen sistemlere dayanmaktadır. Doğrusallık ve zamanla değişmezlik özellikle giriş çıkış sinyallerinin belirli bir genlik aralığına bağlı olduğu sistemler için geçerlidir. Ancak ses işleme sistemlerinde analog seslerin doğrusal olmayan özellikleri bulunur. Dijital filtreleme, ses sinyallerini birçok yönden iyileştirebilir. Örneğin Wiener filtreleme, çoğunlukla sinyal olan frekansları, esas olarak gürültü frekanslarından ayırmak için kullanılır.

İlk doğrusal olmayan teknik, konuşma sinyallerinde geniş bant gürültüsünü azaltmak için kullanılır. Bu ses türü şunları içerir: manyetik bant hissi, analog devrelerde elektronik gürültü, mikrofondan kaynaklanan gürültüler, kalabalık tezahüratlar, vb. Sinyallerde bulunan gürültüler, gerçek frekansları kaplayabilir. Bu gürültüleri filtrelemek için doğrusal filtreleme kullanılır.

Sinyal ve gürültünün mevcut olduğu bir durumda frekans genliğine bakılarak gürültü ve sinyal kısmen ayrılabilir. Genlik büyük olduğu durumlarda muhtemelen bir sinyalin varlığından bahsedilir ve bu nedenle tutulmalıdır. Genlik küçükse, çoğu zaman gürültü olarak nitelendirilir bu nedenle sıfıra ayarlanmalıdır.

Bu tekniği görmenin bir başka yolu, zamana göre değişen bir Wiener filtresidir. Şekil 5.2'de gösterildiği gibi Wiener filtresinin frekans yanıtı çoğunlukla sinyal olan frekansları geçerken gürültü olan frekansları reddeder.



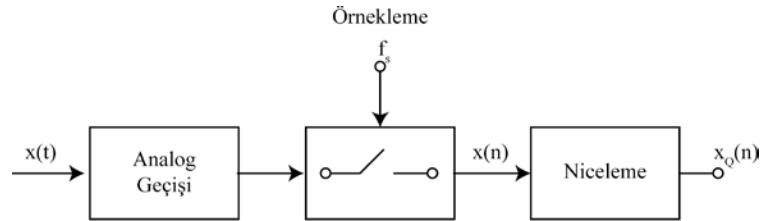
Şekil 5.2 : konuşma ve gürültü spektrumu. (Smith, 1999)

Bu, filtrenin frekans tepkisi belirlenebilmesi için önceden sinyal ve gürültü spektrumları hakkında bir bilgi gerektirir. Bu doğrusal olmayan teknik, Wiener

filtresinin frekans yanıtının, o bölümün spektrumuna dayanarak her bölüm için yeniden hesaplanması dışında, aynı fikri kullanır. Başka bir deyişle, filtrenin frekans tepkisi, sinyalin özelliklerinin belirlediği şekilde bölümden bölüme değişir.

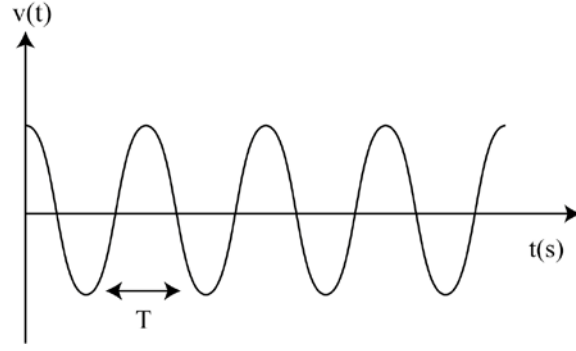
### 5.5 Niceleme

Analog sinyallerin sayısallaştırılması, anlık olarak analog sinyalin karşılık gelen sayısal bir değere eşitlenmesi işlemini içerir. Örnekleme yöntemi ile analog sinyal üzerinde birkaç nokta seçilir ve bu noktalar değeri yakın bir sayı ile birleştirilir. Örnekleme hızı hertz olarak saniye başına örnekleme sayısıdır. Daha yüksek bir örnekleme hızı daha iyi ses kalitesini gösterir, bunun yanında depolama miktarının artması anlamına da gelir. Yaygın olarak kullanılan örnekleme oranları 8 kHz, 16 kHz, 44.1 kHz'dir. Bit çözünürlüğü ise her bir ses sinyalinin örnekleme noktasını temsil etmek için kullanılan bit sayısıdır. 8-bit yaklaşık olarak 0~255 değer aralığını temsil ederken 16-bit 32768~32767 değer aralığını temsil etmektedir. Başka bir deyişle her örnekleme noktası 8 veya 16-bitlik bir tamsayı ile temsil edilir. Şekil 5.3' de gösterildiği gibi sürekli bir ses sinyalinin  $x(t)$  Analog Dijital dönüşümü için temel işlem yapısı,  $x_Q(n)$  in niceleme dizisini veren  $x(n)$ 'in örnekleme ve nicelendirilmesidir. Örneklenen bir sinyalin sürekli genlikle sayısallaştırılması işlemi, niceleme olarak adlandırılır.



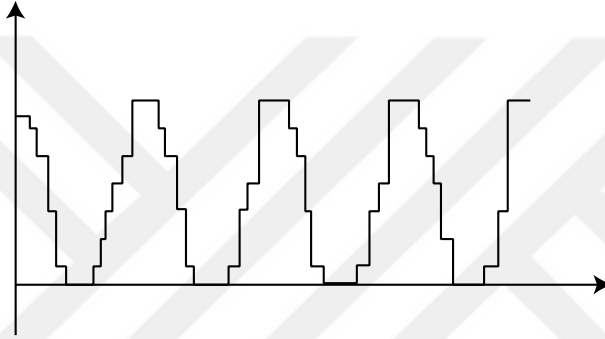
**Şekil 5.3 :** Analog dijital dönüşümü ve nicelemesi

Şekil 5.3'de gösterilen Analog dijital dönüştürücüler, analog sinyali bir dizi dijital değer oluşturma görevini gerçekleştirir.



**Şekil 5.4 : Analog Sinyal**

Şekil 5.4’de bir analog sinyalin görüntüsü gösterilmektedir. Bir analog sinyalin, dijital sinyal olarak gösterimi şekil 5.5’dedir.



**Şekil 5.5 : Dijital Sinyal**

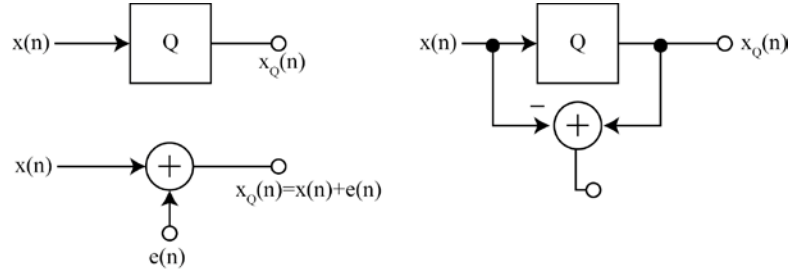
Niceleme fiziksel miktarlar sayısal olarak gösterildiği zaman gerçekleşmektedir (Widrow, 1961). Ölçüm değerleri en yakın birim sayılarına karşılık gelen tamsayılardır. Yuvarlama hataları (+-1/2) birim arasında değerlere sahiptir. Fiziksel miktar ne kadar küçükse, aynı zamanda fiziksel miktarları temsil etmek için gereken sayısal değerler o kadar büyük olacaktır ve bu sayıların depolanması zorlu, masraflı bir süreç olacaktır. Niceleme işlemlerinde doğruluk ve ekonomi arasında bir denge kurulması gerekmektedir.

Widrow,(1961) ‘ e göre orijinal bir  $x(n)$  sinyale rastgele  $(n)$  sinyali eklenmesi olarak modelleme;

$$x_q(n) = x(n) + e(n) \quad (5.1)$$

Bu katkı modeli, sayısallaştırılmış çıktı ile hata sinyaline göre girdi arasındaki farka dayanmaktadır.

$$e(n) = x_Q(n) - x(n). \quad (5.2)$$



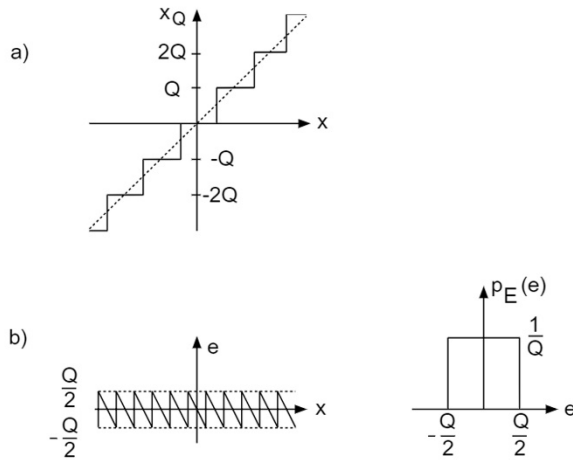
**Şekil 5.6 : Örnekleme**

Çıktı  $x_Q(n)$  'nin bu doğrusal modeli ancak girdi genliği geniş bir dinamik aralığa sahip olduğunda ve  $e(n)$  niceleme hatası  $x(n)$  sinyali ile ilişkilendirilmediğinde geçerlidir. Ardışık niceleme hatalarının istatistiksel bağımsızlığı nedeniyle, hata sinyalinin kendi kendine korelasyonu  $r_{EE}(m) = \sigma_E^2 \cdot \delta(m)$  ile verilir ve bir güç yoğunluğu spektrumu  $S_{EE}(e^{j\Omega}) = \sigma_E^2$  elde edilir.

Doğrusal olmayan niceleme işlemi,  $Q$ 'nun niceleme aşamasını gösterdiği Şekil 5.7(a)'da gösterildiği gibi doğrusal olmayan bir karakteristik eğri ile tanımlanır. Niceleme işlemi çıkışı ve girişi arasındaki fark, Şekil 5.7(b)'de gösterilen niceleme hatasını  $e(n) = x_Q(n) - x(n)$  verir.

Ölçme hatasının tek biçimli olasılık yoğunluk fonksiyonu;

$$P_E(e) = \frac{1}{Q} \text{rect}\left(\frac{e}{Q}\right) \quad (5.3)$$



**Şekil 5.7 :** (a) Bir niceleyicinin doğrusal olmayan karakteristik eğrisi. (b) Niceleme hatası( $e$ ) ve olasılık yoğunluk fonksiyonu (PDF)  $P_E(e)$ .

Bir niceleme hatası ve olasılık yoğunluk fonksiyonu (PDF)  $P_E(e)$  ile rastgele bir değişkenin  $E$  momenti,  $E^m$ 'in beklenen değeri olarak tanımlanır:

$$E\{E^m\} = \int_{-\infty}^{\infty} e^m p_E(e) de \quad (5.4)$$

Sayıların genlik örnekleme için niceleme teoremi Widrow (1961) tarafından verilmiştir. Zaman eksenini sayısallaştırma benzerliği, Shannon'a bağlı örnekleme teoremidir (Shannon, 1948).

## 5.6 Özellik Vektörlerinin Çıkarımı

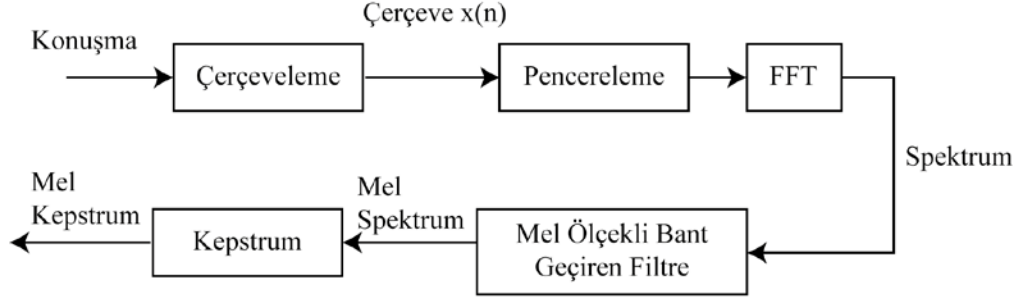
Ses sinyallerinin analizinde genellikle kısa süreli analiz yöntemleri benimsenir çünkü çoğu ses sinyali kısa süre içinde kararlıdır. Örnek olarak 20 ms. aralıklarda çerçeveselenen ses sinyalinde akustik özellikler yakalanır.

Genel olarak ses sinyallerindeki her çerçevede en belirgin akustik özellikler (Jang R., 2019);

- Ses: bu özellik ile sinyallerin genliği ile ilişkili olan ses sinyalinin yüksekliğini gösterir. Genel olarak ses sinyallerinde enerji ve yoğunluk olarak da adlandırılır.
- Ses perdesi: bu özellik, temel frekans tarafından temsil edilecek ses sinyallerinin titreşim hızını veya eş zamanlı olarak ses sinyallerinin temel periyodunun karşılığını temsil eder.
- Tını: temel ses sinyalleri periyodu içinde dalga formu ile karakterize edilen ses sinyallerinin Türkçedeki bir sesli harf gibi anlamlı içeriğini temsil eder.

Örneklenen ses bilgileri, çerçevelere bölünerek daha etkin işlem gerçekleştirilir. Ses sinyalindeki akustik özelliklerin çıkarılmasıyla öznitelikler elde edilir. MFCC, konuşma tanıma alanında yaygın olarak kullanılan öznitelik çıkarma yöntemlerinden biridir.





**Şekil 5.8 : Özellik vektörü çıkarımı**

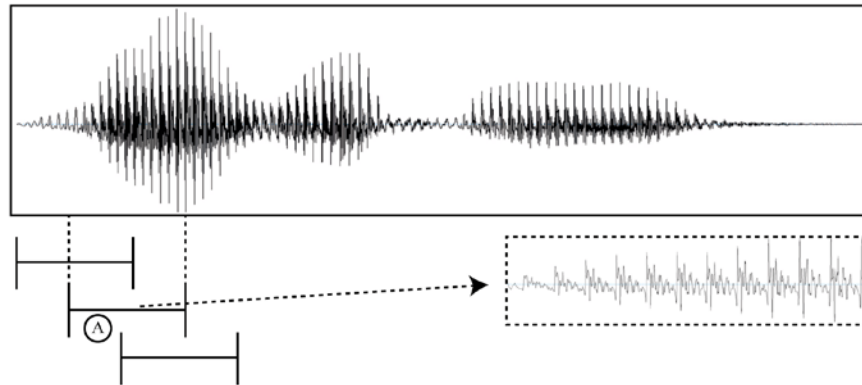
Öznitelik çıkarma aşağıdaki aşamalardan oluşur;

1. Çerçeve Bloklama (Frame Blocking)
2. Pencereleme (Windowing)
3. Fourier dönüşümü (FFT)
4. Mel-Frekansı saptırması (Mel-Frequency Warping)
5. Kepstrum (Cepstrum)

Öznitelik çıkarma işlemi Şekil 5.8’ de gösterildiği gibi ses bilgisinden konuşma tanımlayan akustik özellik verilerini özetleyerek tanımlama işlemini verimli hale getirir.

### 5.6.1 Çerçeveleme

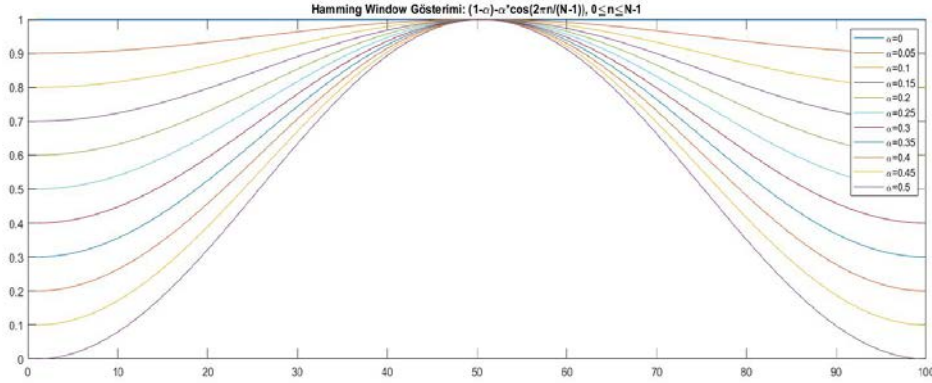
Ses sinyalleri kısa zaman aralıklarında kararlı kaldıkları için bu zaman aralıklarında işlenir. Ses sinyalleri, şekil 5.9’da gösterildiği gibi genellikle 20-30 ms. arasındaki değişen uzunluklarda çerçevelere bölünür.



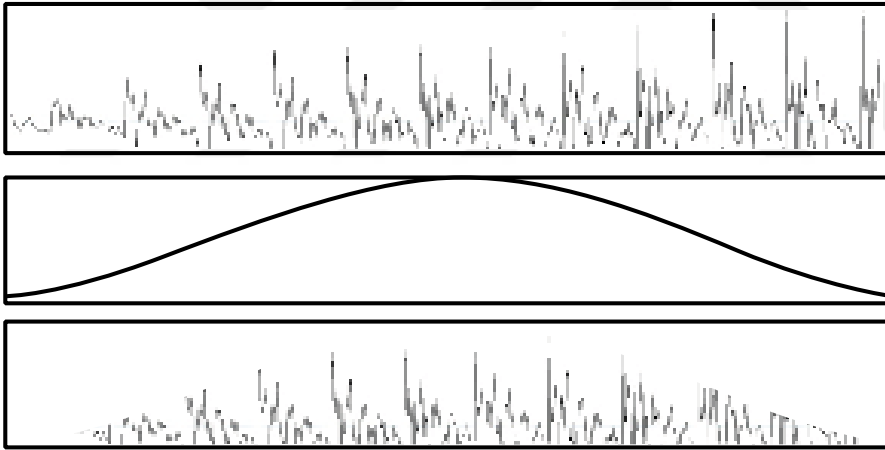
**Şekil 5.9 : 20-30 milisaniye arasındaki bir ses sinyalinin detaylı gösterimi**

## 5.6.2 Pencereleme

İkinci aşama olarak bütün çerçevelerin pencereleme işleminin gerçekleştirilmesi. Şekil 5.10' da gösterildiği gibi pencerelemenin amacı çerçeve başında ve sonundaki süreksizlikleri Şekil 5.11'de gösterildiği şekilde ortadan kaldırmaktır. En çok kullanılan pencereleme fonksiyonlarından birisi Hamming fonksiyonudur.



Şekil 5.10 : Hamming penceresi



Şekil 5.11 : Hamming penceresi uygulama örneği

## 5.6.3 FFT dönüşümleri

Bir sonraki aşama her çerçevede Hızlı Fourier Dönüşüm (FFT) işlemini gerçekleştirmektir. Bu dönüşüm ile Ayrık Fourier (DFT) dönüşümünün hızlı bir şekli olarak tanım kümesini zaman uzayından frekans uzayına geçirir. Hızlı Fourier Dönüşümü (FFT), ses ve akustik ölçüm biliminde önemli bir ölçüm yöntemidir. Sinyal hakkında frekas bilgisi sağlamak için spektral bileşenlere dönüştürür. FFT'ler, arıza analizi, kalite kontrolü ve makine veya sistemlerin durum izlemesi için kullanılır.

#### **5.6.4 Mel frekansı saptırması**

Herhangi bir konuşma tanıma sisteminde ilk adım, özellik çıkarımıdır. Dile ait içeriği tanımlamak ve arka plan gürültüsü gibi ses sinyali bilgisinde istenmeyen ses özellikleri atarak iyi olan ses sinyalini tanımlamaktır.

Konuşma, insan tarafından üretilen seslerin, dil, dişler gibi ses yolu şekli ile filtrelenmesidir. Bu filtreleme ile hangi sesin çıkacağı belirlenir. Üretilen sesin doğru bir gösteriminin belirlenmesi sesin doğru çıkarımını sağlar.

İnsan kulağı frekansları doğrusal olmayan bir şekilde algılamaktadır. İnsan kulağı 1kHz 'e kadar olan frekansları doğrusal olarak algılamakta daha yüksek değerleri logaritmik algılar. Mel-ölçüsü insan kulağının frekans yanıtını karakterize eden bir ölçüdür. MFCC ses yolunun kısa süreli güç spektrumunu temsil eder. Mel Frekans Keprstral Katsayıları (MFCC), otomatik konuşma ve konuşmacı tanımda yaygın olarak kullanılan bir yöntemdir.

#### **5.7 Ses İşleme Uygulamaları**

Ses bilgisi, içerisinde binlerce bilgi barındırmaktadır. Ses sinyalleri konuşmacının dil, yaş ve cinsiyetine bağlı olarak değişir. Ses işleme uygulamaları ile ilgili olarak konuşmacı tanıma modülleri, ses kayıtlarındaki bozukluk ve gürültüleri gidermek için iyileştirme işlemleri, ses kayıtlarındaki güvenilirliği tespit etmek için montaj analiz işlemleri yada ortamdaki konuşmacıların birbirinden ayrıştırılması için konuşmacı ayrıştırma işlemleri ve ses bilgilerinden konuşmacı doğrulama işlemleri gibi farklı uygulama alanlarına sahiptir.

##### **5.7.1 Ses sınıflandırması**

Ses sınıflandırması, ses işleme alanında temel bir sorundur. Görev esas olarak seslerden özellikler çıkarmak ve sonra sesin hangi sınıfa ait olduğunu belirlemek. Ses sınıflandırması ile ilgili olarak çalgı aleti tanıma, sanatçı tanıma gibi faydalı uygulamalar bulunmaktadır.

Müzik sınıflandırılması ile doğrudan ilgili olan görevler aşağıda listelenmiştir (Fu ve diğ. , 2011).

- Tür Sınıflandırması.
- Ruh hali Sınıflandırması.
- Sanatçı Kimliği.

- Alet Tanıma.
- Müzik Notu.

Bir ses sınıflandırma görevini çözmek için yaygın bir yaklaşım, kullanışlı özellikleri çıkarmak için ses girişlerini önceden işlemek ve daha sonra buna bir sınıflandırma algoritması uygulamaktır. Örnek olarak 5 saniyelik bir ses bilgisinden köpek havlaması ya da bir matkap sesi olup olmadığının belirlenmesi. Bunun gibi işlemlerle başa çıkmak için ses özelliğinin çıkarılması ve daha sonra uygun bir sınıfa yerleştirilmesi gerekir.

### **5.7.2 Ses parmak izi**

Ses parmak izi uygulamasının amacı, sesin dijital “özetini” belirlemektir. Bu, bir ses örneğinden gelen sesi tanımlamak için yapılır. Ses parmak izi, diğer ses kliplerine göre verimli bir şekilde eşleştirilebilen kompakt bir gösterim elde etme yeteneği sağlar. Yaygın bir kullanım durumu, örnek sorgulamaya göre müzik tanıma özelliğidir: bir kullanıcı restoran, alışveriş merkezi veya arabadaki bir şarkıyı dinler ve şarkı hakkında daha fazla bilgi edinmek ister. Shazam ve SoundHound, cep telefonlarındaki popüler müzik tanıma uygulamalarının örnekleridir.

### **5.7.3 Ses bölümleme**

Ses bölümlendirmesi ile bir ses sinyalindeki anlamsal olarak işimize yarayacak geçici bölümleri otomatik olarak ortaya çıkarmaktır. Bu süreçler vasıtasıyla, anlamsal olarak benzer işitsel sahneler bir araya getirilir.

Bölümleme, kelimenin tam anlamıyla belirli bir nesneyi, tanımlanmış bir özellik setine dayanarak parçalara (veya segmentlere) bölmek anlamına gelir. Bölümleme, özellikle ses veri analizi için, önemli bir ön işleme adımıdır. Bunun nedeni, gürültülü ve uzun bir ses sinyalini, daha sonraki işlemler için kullanılan kısa homojen bölümlere (kullanışlı kısa ses dizileri) ayrıştırmamızdır. Örnek bir kalp sesi bölümlendirme, yani kalbe özgü sesleri tanımlamaktır.

### **5.7.4 Ses kaynağı ayrımı**

Ses Kaynağı Ayrımı, bir sinyal karışımından bir veya daha fazla kaynak sinyalinin izole edilmesinden oluşur. Bunun en yaygın uygulamalarından biri, eşzamanlı çeviri (örneğin, karaoke) için yapılan şarkı sözlerini tanımlamaktır.

### 5.7.5 Müzik önerisi

İnternet sayesinde artık istediğimiz zaman dinleyebileceğimiz milyonlarca şarkı var. Müzik öneri sistemleri, dinleyicilere yeni müzikler önererek otomatik olarak bu aşırı bilgi yüklemesine yardımcı olur. Spotify ve Deezer gibi içerik sağlayıcılar oldukça gelişmiş müzik öneri motorları geliştirmiştir. Bu modeller, özelleştirilmiş öneri listeleri oluşturmak için diğer birçok özellik arasında dinleme geçmişini kullanır.

Bir regresyon / derin öğrenme modeli eğiterek dinleme tercihlerini kişiselleştirme zorluğuyla baş edebiliriz. Bu, işbirlikçi bir filtreleme modelinden elde edilen şarkıların gizli temsillerini tahmin etmek için kullanılabilir. Bu şekilde, herhangi bir kullanım verisi olmasa bile, bir şarkının ortak filtreleme alanındaki temsilini tahmin edebiliriz.

## 6. KONUŞMA TERAPİSİ

### 6.1 Giriş

Yaklaşık 70 yıl boyunca araştırmacılar konuşma başarısını geliştirmek için teknolojiyi kullanma konusunda başarılı sonuçlar elde edemediler (Bernstein ve diğ., 1988). Gelişen teknoloji ve sinyal işlemedeki yeniliklerle birlikte konuşma eğitim uygulamaları için itici güç sağlayan gelişmeler yaşandı.

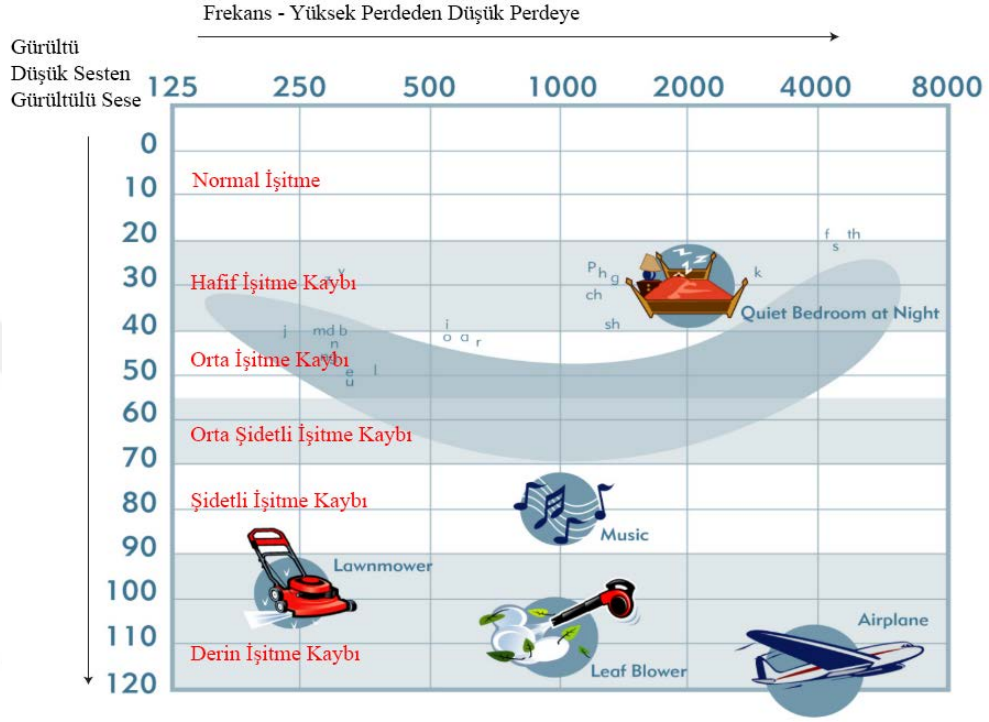
Türkiye’de her yıl yaklaşık olarak 1000 ila 2000 arasında çocuk işitme engeliyle doğar veya konuşma ve dil öğrenmeye başlamadan önce işitme bozukluğu yaşarlar. Ağır işitme engelli çocuklar işitme cihazlarından minimum düzeyde fayda sağlamaktadır. 18-48 aylık çocuklar tipik olarak anormal seslere ve neredeyse hiçbir kelime dağarcığına sahip değildirler. Daha büyük çocuklar ve yetişkinler konuşmaların kalitesi işitme cihazı ile akıcı bir etkileşim gerçekleştirmek için yeterli değildir (Bernstein ve diğ., 1988).

1990 yılından itibaren DRA konuşma araştırma birimi, küçük çocuklar için konuşma ve dil gelişimine yönelik konuşma tanıma teknoloji uygulamalarını araştırmıştır. STAR (Speech Training Aid Research) adıyla bilinen ilk proje, öğretmenlerin küçük çocuklara çeşitli iletişim ve dil becerilerinin geliştirilmesinde yardımcı olması için bir “konuşma eğitimi yardımı” aracına gereksinim duyduklarına dair farkındalığı ile desteklendi. Amaç çocuğun metinsel, görsel ya da sözlü bir ifadeye cevaben işitme engelli çocuk tarafından konuşulan kelimenin “iyi” ya da “kötü” telaffuzlarını ayırt edebilen bilgisayar tabanlı bir sistem geliştirmektir (Russell ve diğ. , 1996). Aynı konuşma tanıma teknolojisi daha sonraki yıllarda animasyonlu “Konuşan Kitaplar” serisine dahil edilmiştir. Bilgisayar destekli konuşma ve dil gelişimi uygulamaları çok büyük bir potansiyele sahip olmakla birlikte işitme engelli bireyin daha fazla eğitim yardımı almasını desteklemektedir.

İnsanlar frekans tabanlı duyarlar ve her frekansta duyulan ses duyarlılığı farklıdır. Normal işiten insanlar 20 Hz ile 20000 Hz arasında duyarken, yaş ilerledikçe bu frekans aralığı daralır. İşitme testlerinde kişinin frekans bazlı duyduğu sesler tespit

edilir. Ses şiddeti desibel olarak ifade edilirken kişinin duyabildiği aralık 0-120 db'dir.

125 Hz'den başlayan sesler yükseldikçe incelmektedir. 0 db başlayan sesler ise yükseldikçe ses şiddeti artar. İnsanlar yaşlandıkça duyu organı olan kulaklarda yaşlanmaktadır.



Şekil 6.1 : İşitme kaybı seviyeleri

İşitme engel türlerinden Şekil 6.1'de gösterildiği gibi;

- Hafif Şiddetli Kayıpla 20-30 db.
- Orta dereceli kayıplar 40-60 db.
- Orta ileri kayıplar 60-70 db.
- İleri derece kayıplar 70-90 db.
- Çok ileri derece kayıplar ise 91db'den yukarı olan işitme kayıplarıdır.

İşitme engeli, bireylerde mevcut işitme sisteminde herhangi bir nedenden oluşan sorun sebebiyle akustik bileşenlerin normalden az ya da hiç algılayamaması sonucunda meydana gelen bir engel durumu olarak tanımlanır. İşitme engelli bireyler Dünya ve Türkiye üzerinde önemli bir sayıdadır. İşitme engelli bireyler de var olan iletişim sorunundan dolayı işiten dünyanın bir parçası oldukları için eğitilmesi planlanmalıdır (Neumeyer ve diğ. , 2000).

Genel olarak dil iletişim kurmada bir araç, konuşma ise bu aracı iletme yolu olarak belirtilir (Akçamete, 1993) . İşitme engeliyle birlikte iletişim becerilerini, işitme kaybının başlama yaşı, işitme kaybı derecesi, işitme kaybı tipi gibi etmenlerde belirlemektedir.

İşitme engelliyle birlikte kişinin sadece dil ve konuşma gelişimi değil aynı zamanda bilişsel, motor, sosyal ve duygusal gelişim yönünden de olumsuz etkilenmektedir. İşitme engelli bireyler yeteri kadar işitsel uyarıcıya sahip olamadıkları için öğrendiklerini pekiştirip davranış haline getiremezler (Uğurtay ve diğ., 1999). İşitme engelli çocukların açık ve anlaşılır bir konuşmaya sahip olamaması düşüncelerini ifade etmede ve başkalarının düşüncelerini anlamada engel oluşturmaktadır. İşitme engelli çocukların bilişsel becerilerdeki gecikmeleri eğitim ve yaşantı eksikliğinden kaynaklanmaktadır. Yeteri kadar eğitim verilmezse öğrendiklerini pekiştirip davranış haline getiremezler.

Konuşma terapisindeki amaç konuşma bozukluklarının terapi yoluyla tedavisidir. Konuşma için bireylerin doğru seslendirmesi, vurgu ve ritim kontrolünün öğretilmesiyle konuşmada meydana gelecek hataların giderilmesidir (Türk ve Arslan, 2004). İşitme engelli bireylerde tipik olarak konuşma üretimi anlaşılabilir düzeydedir. Gelişen teknolojiyle birlikte görüntü ve ses işleme teknolojileri konuşma eğitimi yardımları konusunda itici güç sağladı. İşitme engelli bireylerin konuşma eğitimi konusunda bilgisayar destekli uygulamalarla yardım sağlamaya çalışılmaktadır. Bilgisayar destekli konuşma tanıma yöntemleriyle birlikte konuşma terapisindeki teşhis ve tedavi aşamalarında kullanılmaya başlanmıştır. Bilgisayar destekli konuşma terapi yöntemleriyle birlikte kullanıcı ve bilgisayar arasında birebir ve etkin bir etkileşim kurulması gerçekleşmektedir.

## **6.2 Dil Edinimi**

Dil insanın ayrıcalık belgesi olduğunu fizyolojik, biyolojik ve sosyal bir sürecin sonunda ortaya çıktığını belirtmektedir (Ergenç, 2018).

### **6.2.1 Nöropsikolojik açıdan dil edinimi**

Bütün çağdaş bilimsel veriler, beyin, bilişsel süreçler ve dil konusundaki tüm verilerin alanlar arası bir yaklaşımla ele alınması gerektiğini ortaya koymuştur (Ergenç, 2018).



Nöropsikolojik açıdan dil insan hayatında sosyal gelişmelerle yaratılan kodlar sistemi sesbilim, biçimbilim, sözdizim, anlambilim ve kullanımbilim düzlemlerinden oluşmaktadır. Bütün bu düzlemlerin aktif bir şekilde kullanımı ve beynin katılımıyla ortaya dil kavramı çıkmaktadır.

Psikofizyolojik açıdan bakıldığında zaman ise beyindeki aktif oluşumlar sonucunda dilin kazanılması ve kullanılması ifade edilmektedir. Eğer sinir sisteminde bu fonksiyon için yüksek derecede özelleşmiş ve aynı zamanda esneklik kazanmış fizyolojik bir ön hazırlık yoksa dilin, ne kazanılması ne de kullanılması olasıdır (Ergenç, 2018).

Doğumdan itibaren, doğal ve sosyal çevre vasıtasıyla bilinçlenen ve biçimlenen beyin, eğitim yoluyla sistematik bir şekilde yeniden biçimlenir. Konuşma ve adlandırma süreci genel bir eğitim evresi şeklinde tanımlanırken; okuma, yazma, matematiksel işlemlerin gerçekleştirilmesi belirli bir yaş olgunluğuna göre gerçekleşmektedir.

Ergenç (2018)' e göre dil edinimi en basit düzeyde, dil fonksiyonlarının ortaya konma aşamaları şeklinde konuşma, anlama, okuduğunu anlama, sesli okuma, yazma ve adlandırma şeklinde gerçekleşmektedir.

### **6.2.2 Dil sisteminin işleyişinde dilbilimsel bakış açısı**

Chomsky (2001)'e göre insanlar doğuştan, dil öğrenebilmek için özel bir mekanizmaya sahiptir. Öğrenme becerisi konuşulan dilin içselleştirilerek, dilin kurallarını anlayıp öğrenmesini ve sonrada bu kurallara uygun olarak konuşmasını sağlar. Chomsky, her ifadeyi dilbilimsel sistemde kavramların anlamsal yönü ve konuşulan sözcükler ile düşünsel olarak seslerin kavramlarını anlayarak gerçekleştirir (Chomsky, 2001).

Dil ediniminin ilk evreleri tekrar ve taklit ön planda yer almaktadır. Çocuğun doğumundan itibaren yakın çevresi ile ilişkilerinde iletişimsel paylaşım kurallarını öğrenerek büyümektedir. Çevresindeki nesnelere hakkında iletişim nasıl gerçekleştiği kavramını anlamaya çalışmaktadır. İlk sözcükleri kullansa dahi henüz tam olarak karşılıklarını oturtamamaktadır.

Dil ediniminin ikinci safhasında sözcüklerin temsil boyutunu kavrayarak dilin anlamsal şekilde geliştirmeye çalıştığı bilişsel süreçlerdir. Çocuk bilişsel süreçte soyut kavramlar hakkında kurgulamayı öğrenip somuttan sembolik düzleme geçmektedir.

Dil ses, biçim, sözdizimi, anlam ve kullanım gibi birbirinden bağımsız farklı sistemlerin bağlantılı olarak işlemesiyle ortaya çıkan büyük bir sistemdir. Dil 'bir sistemler sistemidir' diyerek 20. yy başında, dilbilimin diğer bilim alanları arasında yer almasına neden olan Ferdinand de Saussure (1976); dilin soyut ve somut kullanım yönüne sahip olduğunu belirterek dilin soyut yönünü yani işleyişini ortaya koymak olduğunu söylemiştir. Chomsky ve Halle (1968), göre ise her insanın, bir dili edinebilme ve kullanabilme yetisiyle doğduğunu ve insan dilinin zihnin karmaşık prensipleri olduğunu belirlemiştir. İnsan yeni bilgileri dile getirerek sonsuz sayıdaki tümceyi anlayıp farklı biçimde kullanabilme ve üretebilme yetisine sahiptir.

İnsan beyni ürettiğimiz sözcüklerin anlamının yanında ses tonunda bulunan duygu durumunda anlamlı bir şekilde çözebilir. Yaşamımızda kullandığımız bu özelliklerin tümü 10 milyar kadar nörona sahip ve 1 milyar tümcenin hesaplanmasını sağlayan nöronlarda gerçekleştirilir (Jackendoff, 1994).

Araştırmalara göre insan diline ait bilgilerin yanı sıra yaşama ait tüm imgelerinde o imgeyi oluşturan özelliklerin, nöronlar aracılığıyla saklandığını göstermektedir. Her dil kullanıcısının beyinde depolanmış olarak bulunan ve dil kurallarını içeren bir zihinsel dilbilgisinin (ya da evrensel dilbilgisinin) varlığı, modern dilbilimin temel kuramsal yapısını oluşturur. Jackendoff (1994)' a göre kalıtım varsayımına göre dil gelişim süreci ancak, destekleyici beyin yapıları tamamlandığında ortaya çıkabilir. Yeni doğan bir bebeğin beyin gelişimi tam olarak sağlamadığı için ve dil ediniminde çevreden gelen bilgilerin alınması sürecinin yeterli olmadığı bu sürece çevreden gelen bilgilerin anlamlandırılmasıyla birlikte yeni yollardan kullanılmasını gerektiğini belirtmektedir.

Modern dilbilim çalışmaları dilin işleyişini sesbilimsel, sözdizimsel ve anlambilimsel olarak ortaya koymaktadır (Chomsky, 2001). Dildeki bu bileşenler çocuk tarafından düşünce ile somuta dökülmüş ses arasında zamanla içselleştirilen bir dizi kural içermektedir. .

Dil temel olarak sözcüklere, sözcüklerde düzenli seslemlere ve konuşma seslerine bölünebilmektedir. Sesbilgisel yapısı ile konuşurken kişinin zihninde ses yolu düzeneklerinin dizgilerinin belirlenmesi ve beyin vasıtasıyla bu dizgeyi ses yolu kaslarına talimatlara dönüştürerek nasıl davranmaları gerektiğini belirtmek

zorundadır. Konuşma ve duyma sırasında da bu tür komut talimatları beyin tarafından verilmektedir.

Sesbilimsel yapı, konuşma seslerini bir araya getirip sözcük üretmemizin ilk aşamasını oluştururken, sözdizimsel yapı, zihinsel dilbilgisi kuralları ile sözcüklerin bir araya gelmesi ile öbek ve tümcelerin ortaya çıkmasını sağlamaktadır. Anlamli birimlerin oluşturulma süreci sözdizimsel kurallar bütününün işleyişi, edinim sürecindeki çocuğun zihinsel dilbilgisinin kullanmasını sağlamaktadır (Ergenç, 2018). Beyin, dilin anlatım özelliğini, sözcükleri, sesbirimleri ile anlamlarını ve ayırıcı özelliklerini ve sözcükleri yerleştireceği örüntüleri depolayarak gerçekleştirir.

Anlambilimsel açıdan sözcüklerarası ve tümcelerarası ilişkiler kurulabilmektedir. İnsanlar nesne ve olaylarla ilgili olarak yaşananları, dünya bilgilerini, deneyimler ve kültür öğelerini dilsel yapılara başvurarak kendi bilişsel düşüncelerini anlatabilmek için kullanırlar. Anlambilimsel bileşeni oluşturan bilgi, sözcük anlamı, tümce ya da önerme anlamı ve dolaylı anlamla açıklanabilmektedir (Ergenç, 2018). Dilin dördüncü bileşeni olarak kullanımbilimsel bileşenden söz etmek gerekirse dil diğer üç bileşenin çalışmasıyla oluşturduğu iletişimsel veriler, somuta dökülüp alıcı/dinleyici/okuyucu tarafından algılanma sürecine girdiğinde kullanımbilimsel bileşen de devreye girer.

Konuşma eylemi sırasında herhangi bir tümcenin üretiminde sözcükler belirli bir sözdizimsel sırayla ifade edilirken anlamsal ve sesbilimsel özellikler de bilinmekte ve devreye girmektedir. İnsanlar konuşma sırasında aktarmak istediği anlamı bir dizi sesi bir araya getirerek sözcükleri oluşturup tümceleri ifade etmektedir (Ergenç, 2018).

Dil edinimi süreci beynin tümevarım yöntemiyle hareket etmemektedir. Doğuştan gelen bilgilerle ve gelen uyaranlar doğrultusunda süreci tamamlamaktadır. Dil ediniminin normal sürecinde meydana gelen olumsuz durumlar dil için özelleşmiş bölümde olduğu, genel olarak işlemlerin sağlıklı bir şekilde işlediği belirlenmiştir.

### **6.3 Konuşma Eylemi**

Kişinin beyninde gerçekleşen olgular ve soyut kavramlar dil, söz ya da yazıyla somutlaşarak bireyin çevreyle etkileşim kurmasına sağlar. Kişinin ruhsal ve toplumsal kişiliğinin belirginleşmesinde ve düşüncelerin gelişiminde önemli bir

işleve sahip olan dil, farklı görüş, fikir, yönlendirme, etkileme belirli amaçlar için deneyimlerinin aktarmasını sağlayan yöntemlerin başında gelmektedir.

Toplumsal yaşamda en sık kullanılan iletişim yöntemlerinden ses dayalı bildirimler kişinin konuşma eylemiyle düşüncelerinin açıklayıp aktarmasını sağlarken diğer yandan dil dizgesinin biçimlerini kullandığı için dille ilişkilidir. Etkin bir iletişim sisteminin gerçekleşebilmesi için verici (konuşmacı/yazar), alıcı (konuşucu/okuyucu), bildirim yolu ya da kanal, bir kod, bir mesaj ve bir gönderi gerekir.

Konuşma eylemi, konuşucunun amacı ve sezdirimleriyle dinleyicinin çıkarımlarının bir toplamı olarak yalınlaştırılabilir. Konuşucunun amacı ve sezdirimleri, dinleyicinin çıkarımlarının belirleyicisi olarak düşünüldüğünde, betimleme açısından da bir başlangıç noktası oluşturacaktır.

#### **6.4 İşitme Engellilerin Dil ve Konuşma Becerilerinin Geliştirilmesi**

İnsanoğlunun karmaşık bir yapıya sahip olan dil ve konuşma biçimi onun toplumun geri kalanıyla iletişim kurmasını sağlamaktadır. Chomsky'ye göre dil edinimi insanoğlunda genetik olarak doğuştan var olduğunu belirtmektedir. Dil edinimi ve konuşma eğitimi, etkili bir iletişim kurabilmek için gereklidir. Bu etmenler herhangi bir şekilde zarar görürse yaşamlarında etkin bir şekilde iletişim kurmasını zorlaştıracak veya engelleyecektir. Bunun sonucunda kişinin yaşamının sonraki yıllarında akademik becerilerinin gelişiminde sorunlarla karşı karşıya kalacağını belirtmektedir (Tüfekçioğlu, 1992). İşitme engellilerin dil ve konuşma becerilerini geliştirme teknikleri gelişimsel bir sıra içinde ve önce dilin sonra konuşmanın öğretilmesi olarak ele alındığı görülmektedir (Neumeyer ve diğ., 2000). Konuşma ve dil bozuklukları dünya sağlık örgütünün istatistiklerine göre insan nüfusunun en az %3.5 'nin iletişim becerilerini olumsuz bir şekilde etkilediğini belirtmektedir.

Konuşma terapisinde amaç, doğru seslendirmenin, vurgu ve ritim kontrolünün öğretilmesi, konuşmadaki hata ve bozuklukların giderilmesidir. Terapist hastaya çeşitli testler uygulayarak konuşma sorunlarını tespit etmek, uygun tedavi yöntemleriyle sorunları gidermeye çalışmaktadır.

Konuşma terapisinde teşhis ve tedavilerin aşamalarında otomatik konuşma tanıma yöntemlerinin kullanımı yaygınlaşmaktadır. Artikülasyon eğitimi için tasarlanan yazılım araçlarında konuşma tanıma kullanıcısıyla bilgisayar arasındaki birebir

etkileşim kurulmasını sağlamaktadır. Telaffuzun belirlenmesi ve doğru telaffuz için fonem tanıma ve ses perde eğrilerinin analizi için farklı teknikler uygulanmaktadır.

#### **6.4.1 Artikülasyon eğitimi**

Konuşma seslerindeki yaşanan sorunlardan kaynaklanmaktadır. Öncelikle yapılması gereken hangi seslerde sorun yaşandığının tespitidir. Bunun için artikülasyon testlerine başvurulması gerekir. Sorunlar tespit edildikten sonra bu sorunun niteliği tespit edilmelidir.

- Ses Değiştirilmesi: Bir sesin yerine başka bir sesin söylenmesi : “/K/ yerine /T/, /L/ yerine /Y/ demek gibi ” Örnek Katı yerine Tatı Salon Yerine Sayon.
- Ses Atlama (Düşürme): Kişi bazı sesleri çıkarma ile ilgili sorun yaşar ve ilgili sesi atlar/ yutar: Örnek: Tara yerine ta.a hayır yerine ayır.
- Sesin eklenmesi: Kelimeye yeni bir ses ekleme aşağı-aşşağı, eşek-eşsek der.
- Sesin bozulması: sesleri tam doğru olmamakla birlikte gerçeğine yakındır. Daha çok yöresel konuşma biçimlerinde rastlanır. Testi, desti, kardeş gardeş

İşitme engelli bireylerin bir dili öğrenmesinde psikolojik faktörde önemlidir. İlgi ile bireyi harekete geçirmeyi ve hedefleri takip etmeye iten psikolojik bir güçtür. Anlama ve tekrarlamada dilin öğrenilmesini etkiler.

Bir metni okurken veya bir cümlenin sözlerini tanıyan sözlü bir sunum duymak, kavrayışın başlangıcıdır. Kelimenin anlamı ile birlikte, cümlenin anlamının zihinsel bir yapısını inşa etmek için kelimeler arasındaki sözdizimsel veya dilbilgisel ilişkiler tanınmalıdır.

#### **6.4.2 Fonemlerin karıştırılması**

Sık karşılaşılan sorunlardan biride fonemlerin birbiriyle karıştırılmasıdır. Terapist hastaya uygulamalı olarak ilgili fonemlerin seslendirilişini göstermelidir. Daha sonra hasta ilgili fonem çiftlerinin geçtiği sözcük ve cümleleri okumalıdır. Terapist hastanın kaydettiği ilerlemeyi izlemekte ve gerektiğinde müdahale etmektedir.

#### **6.5 Dilin Öğretilmesi**

İletişim kurabilmek için gerekli olan dil, toplum arasında kullanılan sınırsız olarak bilgilerin aktarılmasını sağlayan en temel ve en etkili iletişim aracıdır. Dil, duygusal ve sosyal iletişimin en önemli birimlerinden biridir. Sosyal ve duygusal iletişim birimlerinden olan dilin iki temel bileşeni bulunmaktadır. Bunlardan birincisi alıcı

dildir. Alıcı dil ile sesleri algılama, soyut ve somut kelimeleri anlama, söylenenleri yapma ve eleştirel dinleyebilme ve yargılama yetenekleri şeklindedir. Diğer dil bileşeni ise ifade edici dildir. İfade edici dil zihinsel kavramın nefes alma, ses çıkarma, artikülasyon mekanizmaları kullanılarak ses şeklinde ifade edilmiş biçimidir.

İnsanların birbiriyle sözel olarak iletişim kurabilmesi için alıcı dil ve ifade edici dil etkin bir şekilde gelişmiş olması gerekmektedir. Çünkü bireylerin kendilerini başkalarına ifade edebilmesi, alıcı dilin daha önceden gelişmiş olmasına bağlıdır.

Dil kazanımı aşağıdaki aşamaları izlemektedir.

- Seslemler
- Tek sözcüklü tümceler
- İki sözcüklü tümceler
- Üç veya daha fazla sözcüklü tümceler
- Karmaşık tümceler

İşitme engelli bireyin işitme sisteminde kişinin akustik uyaranları normalden az algılaması veya hiç algılamamasıyla özellikle bireyin iki yaşına kadar olan süreçte dilbilgisel olarak isimleri algılama, ayırt etme ve anlamasını kısıtlayacak hale gelmişse ana dilini öğrenme sürecinde büyük bir risk ile karşı karşıya kalır. İleriki yaşlarda da bu sebeple meydana gelecek akademik yetersizlik ve iletişim sorunu çocuğun birçok problemle karşı karşıya kalmasına neden olur.

Çocuklarda tespit edilen işitme kaybının derecesi ve işitme kaybının tipi ile erken tanı ve uygun yöntemler belirlenerek kişiye özel yapılacak eğitim ve rehabilitasyon hizmetleri, işitme kaybının neden olduğu dil üzerindeki olumsuz etkilerden çocuğun daha az etkilenmesini sağlayacaktır. Kendi ihtiyaçlarının karşılanmasına destek olmak çocuğun sosyal hayatta ve yaşamında daha olumlu etkiler sağlayacaktır.

Yeterli dil becerilerinin kazandırılması iki temel yaklaşımdan söz edilmektedir. Gramatik yöntem olarak adlandırılan doğru dil yapılarıyla öğrencilerin dile ait becerilerin kazandırıldığı “Analitik yöntemdir”. Diğer yöntem ise doğal yoldan normal çocuklarında gelişim aşamaları izlenerek dile ait becerilerin geliştirilmesini amaçlayan “doğal yöntemdir” Doğal yöntemin temelinde işitme engelli çocuğa doğal yaşantılar aracılığıyla dili öğretmeye dayalıdır. İşitme engelli çocuklar birlikte oyunlar, resimler, öyküler, karşılıklı konuşma ile dili öğrenebileceği

dođal ortamın sađlanması amaçlanmaktadır. Çocuk böylece işitme cihazıyla birlikte işitsel ve yüz ifadelerine dayanan görsel uyaranlar aracılığıyla dili öğrenmektedir (Akçamete, 1993).

## 6.6 Konuşmanın Öğretilmesi

Konuşma, konuşma organlarının düzenli devinimleri sonucu üretilen anlamlı ses dizgeleri olarak tanımlanır (Girgin, 2006). Konuşma insana özgü iletişim biçimlerinden en çok kullanılan yöntemdir. İşitme engelli kişiler ünlü ve ünsüz harf üretiminde farklı zorluklar yaşamaktadır. Ünlülerin üretimi ünsüzlerden daha zordur. Ünsüz üretiminde de mevcut /b-/p/ veya /d-/t/ gibi ünsüzleri karıştırdıkları ya da bazı ünsüzleri atladıkları belirtilir (Akçamete, 1993).

İşitme engellilerin konuşma yetersizliği üzerine yapılan çalışmalarda işitme engellilerin %15,4'nün konuşmaları anlaşılır düzeydeyken %29,4 'ünün anlaşılır, %12,8'inin konuşamaz düzeyde olduğu belirlenmiştir (Akçamete, 1993).

Nickelson, işitme engellilerin konuşma öğretiminde ve algılamalarında işitme cihazlarına, görsel araçlara ve dokunma sistemlerine yer veren çok duyulu bir programın önemini vurgulamaktadır (Akçamete, 1993).

Konuşma öğretiminde görsel konuşma ve element yöntemi, babıldama yöntemi, akustik yöntem, ortak merkezli yöntem, dokunsal, görsel, işitsel yöntem adı altında değişik yöntemlerde geliştirilmiştir.

Ortak Merkezli yöntemde işiten çocukların basitleştirilmiş ses sistemleri basitten karmaşığa doğru bir sistemle öğretilmiştir. Bu sistemde bazı seslerin artikülasyon ve ayırımında anlamda, dudaktan okumada, konuşmadaki yetersizlikler ile ilgili sorunlar yaşanmıştır. Pat, mat, bat gibi kelimelerin ilk harflerin telaffuzunda zorluklar yaşanmıştır. Bu nedenle ilk olarak söylenişi yakın olan daha sonra ilişkili olan sesler tanıtılmıştır.

Telaffuz eğitimi için nefes alıştırmaları ile seslenim, artikülasyon ve taklit çalışmalarının yer aldığı hazırlık aşaması;

1. Ses üretiminde sistematik çalışmalar.
2. Eksikliklerin düzeltildiği telaffuz becerisi.

Dokunsal, görsel, işitsel yöntemde ise konuşma eğitimi çok duyulu dođal bir yaklaşım olarak tanımlanmaktadır. İşitme engelli çocuklar zamanlarının büyük bir

kısmını konuşmayı kullanmaları ve kelimeyi bütün olarak öğrenmeleri için cesaretlendirmektedir. Sözel iletişim toplumun en temel iletişim aracı olduğu için işitme engelli bireylerin toplumla kaynaştırılması ve çevrenin olası potansiyellerinden faydalanmalarını sağlanmalıdır. Artikülasyon konuşma sürecinin temeli değil doğal sonucu olarak görülmektedir.

İşitsel Global yöntemde ise işitsel görevleri geliştirmeye dayalı bu sistem erken tanı ile etkili araç ve gereçlerle aileyi içine alan geniş çaplı bir uyum programı olarak görülmektedir.

Genel olarak uzman görüşleri sözel iletişim sürecinde görsel işaretlerin, jest ve mimiklerin beden hareketlerinin işitmenin yanı sıra dikkate alınması gereken parçalar olduğunu belirterek tek duyulu bir sistem yerine çok duyulu bir yaklaşım içerdiği görüşünde birleşmektedirler (Akçamete, 1993) .

#### **6.7 Konuşma Üretim Gelişiminde İşitsel Geri Bildirim**

İşitme cihazı olan veya hafif işitme kaybı yaşayan çocuklar, konuşması için işitsel geri bildirim alırlar. Bu geri bildirim hem konuşma hem de dil ediniminde önemlidir. İşitme engelli çocukların herhangi bir farklı engeli olmaması şartı ile konuşma ve dilin anlaşılmasındaki zorluğu yeterli işitsel geri bildirim eksikliğinden kaynaklı olabilmektedir. İşitme engelli bazı kişilerin konuşmayı başarmasının terapist ile uzun süreli bireysel konuşma eğitiminin bir sonucu olduğu gerçeğiyle desteklenir (Bernstein ve diğ. , 1988).





## 7. KONUŞMA TANIMA

### 7.1 Giriş

Akıllı telefon, robotlar, otonom araçlar ve birçok akıllı makineler için insan-makine arayüzlerinin yaşamımızda daha aktif ve etkin hale gelerek yaygınlaşması beklenmektedir. Konuşma tanınmanın etkin bir şekilde kullanılması insan-makine etkileşimini güçlü bir şekilde elde edilmesini sağlayacaktır. Görsel işitsel konuşma tanıma özellikle gürültülü ortamlarda sesin bozulmasıyla daha güvenilir konuşma tanıma için umut verici bir çözüm olarak düşünülür. Görsel işitsel konuşma tanınmanın temel fikri bozulan sesin tamamlanması için görsel bilgilerden yani dudak hareketlerinden bilgileri sağlamak. Görsel işitsel konuşma tanıma sistemlerinde özellikle görsel girdiler çok önemlidir çünkü tanıma performansını önemli ölçüde etkiler.

Bilgisayarlara konuşmanın klavyeyle yazmaya tercih edileceği bazı durumlar olabilmektedir. İnsan genel olarak yazabildiğinden daha hızlı konuşur ve üstelik farklı işlerle uğraşırken bile konuşabilirler. Konuşma tanıma sistemi mikrofon ya da benzer bir elektronik donanım sayesinde elde edilmiş akustik konuşma bilgisinin, o konuşmanın sözcük ya da anlamlı bir bileşene dönüştürme işlemi olarak tanımlanır (Nilsson, 2018).

Konuşma, doğası gereği görsel ve işitsel yöntemlerden gelen bir süreçtir. İnsanların görsel ve işitsel olarak gelen uyarınları ayrı ayrı sunulduğu zaman algılanan sesin hatalı olabileceği üzerinde yaptığı çalışmayla bilinen McGurk etkisi olarak da adlandırılan yöntemde görüntü ve ses, konuşmanın tamamlayıcı ikilisidir (McGurk ve MacDonald, 1976).

İnsandan insana iletişimde ana bilgi kanalları konuşma ve görsel ipuçlarıdır ve bu iki kanal iletişimi etkin kılmak için genellikle birbirini tamamlar. Bu nedenle, kullanıcı dostu bir bilgisayar sonunda kullanıcıyı dinleyebilmeli, görebilmeli ve kullanıcıya görsel-işitsel geri bildirim sunabilmelidir. Bilgisayarın kullanıcının ne söylediğini anlamasını sağlamak için, akustik konuşma sinyalinden metinsel ve anlamsal bilgiler çıkarmak için konuşma tanıma ve doğal dil işleme gereklidir. Kullanıcının sesli

talimatlarını etkili bir şekilde anlamak için, gürültülü ortamlarda bile, bilgisayarın kullanıcıdan gelen görsel ipuçlarına güvenmesi gerekebilir.

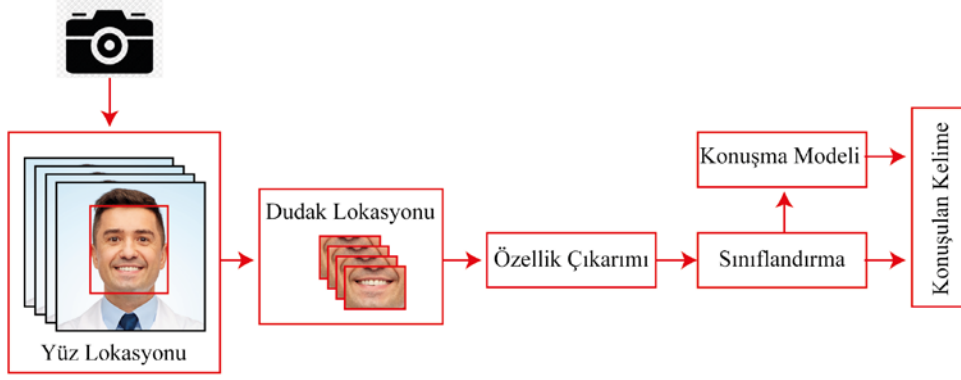
İnsan konuşması görsel ve işitsel olarak iki modludur. Sesli konuşma sinyali konuşmacı tarafından üretilen akustik dalga biçimini ifade ederken, görsel konuşma sinyali konuşmacıların dudak, dil ve diğer yüz kaslarının hareketlerini belirtir (Chen, 2001). Konuşma, sesin titreşimi ve burun boşluğu, dil, dişler ve dudaklar dahil olmak üzere artikülasyon organlarından oluşan ses yolunun yapılandırılması ile gerçekleşir. Bu artikülasyonların bazıları görünür olduğundan, akustik ve görünür konuşma arasında doğal bir ilişki vardır. McGurk etkisi, bir kişiye çelişkili sesli ve görsel uyarılarla sunulduğunda, kişinin her iki şekilde de bulunmayan bir sesi algılayabileceğini göstererek iki taraflılığı oluşturur. Örneğin, bir kişi / ba / sesini “duyduğunda”, ancak / ga / diyen bir konuşmacının dudak hareketini “izlerken”, kişi / ga / veya / ba / 'yi algılayabilir. Bunun yerine, algılanan şey / da / 'ya yakın bir şeydir. McGurk etkisi, bir kişi tarafından algılanan konuşmanın sadece akustik ipuçlarına değil, dudak hareketleri gibi görsel ipuçlarına da bağlı olduğunu göstermektedir. McGurk etkisinin bazı görsel-işitsel kombinasyonlar için diğerlerinden daha güçlü olduğu bildirilmiştir (Chen, 2001) .

## 7.2 Görüntü Ön İşleme

Tipik bir görsel konuşma tanıma sistemi Şekil 7.1, üç aşamadan oluşmaktadır (Hassanat, 2009).

- İnsan yüzlerinin tespiti
- Lokalize işlemi / Dudak lokasyonu
- Dudak okuması

Görsel konuşma tanıma sistemlerinin doğruluğu hassas bir dudak lokasyonu ve bunun yanında sağlam özellik çıkarma sisteminin gerçekleştirilmesine bağlıdır.



**Şekil 7.1 :** Görsel konuşma tanıma sistemi

Görüntü işlemede ön işlem adımı giriş görüntüsünden daha iyi sonuçlar elde etmede yardımcı olduğu için önemli rol oynamaktadır. Giriş görüntüsünden gürültü çıkarma işlemi, netlik ayarı, ışıklandırma gibi ayarlar en iyi şekilde sonuçların elde edilmesinde yardımcı olacaktır.

### 7.2.1 Yüz lokasyon tespiti

Yüz tespiti, yüzle ilgili birçok uygulamada (yüz tanıma, dudak okuma, yaş, cinsiyet ve ırk tanıma, duygu tahmini) önemli bir ön işlem adımındır. Yüz tespitinin amacı herhangi bir görüntüde yüz olup olmadığının belirlenerek varsa yüz bölgesinin görüntü üzerindeki konumunu ve yüzün boyut bilgilerini geri döndürmektir. Yüz tanıma, dudak okuma, duygu tahmini gibi uygulamaların doğruluk oranı, yüz tespit işlem adımının güvenilirliğine bağlıdır. Genel olarak yüz algılama ve tespitinde güvenilirliği etkileyen faktörlerin başında pozlama, yönlendirme ve aydınlatma koşulları ile birlikte gelen parametrik farklılıklar vardır. Bunu çözmek için farklı algoritma ve yöntemler kullanılır. En başarılı yöntemlerden bazıları, tüm olası konumlar, ölçekler ve yönlendirmeler için görüntü üzerinde  $20 \times 20$  (veya daha fazla) piksel gözlem penceresi kullanmak. Bu yöntemler destek vektör makineleri, sinir ağı veya özellik çıktılarının histogramlarına dayanan maksimum olabilirlik yaklaşımı kullanılmaktadır.

Yüz tespitiyle ilgili bazı zorluklar bulunmaktadır. Pozlama sorununda, yüz görüntüleri kamera-yüz pozisyonuna bağlı olarak değişiklik göstermektedir. Kişinin farklı profillerdeki görüntüsünün sonucunda göz, burun gibi yüz özellikleri kısmen yada tamamen gözükemeyebilir. Yapısal Bileşenlerin varlığı ya da yokluğu kişiye bağlı olarak sakal, bıyık, gözlük gibi yüz özelliklerinde bulunması ya da

bulunmaması bu bileşenler arasındaki farklılık gösteren şekil, renk ve boyut gibi büyük çeşitlilikler söz konusu olabilmektedir. Yüz ifadesi zorluğunda, yüz görünümü, bir kişinin yüz ifadesinden doğrudan etkilenen özelliklerdendir. Kapanma bir görüntüdeki yüz ifadelerinin önünde başka nesnelere ya da kişiler tarafından kapanması sonucunda oluşabilecek zorluklardandır. Yine görüntünün yönü, yüz görüntülerinin farklı eksen ve yerlerde olması ile ilgili zorluklara sebep olur. Görüntüleme koşulları, görüntüler oluşturulduğu zaman, aydınlatma ya da kamera özelliklerine bağlı faktörler yüzün görünümünü etkilemektedir.

Yang ve diğ. (2002) yaptıkları çalışmada fotoğraflarda yüz tanıma yöntemlerini hareketsiz görüntülerde dört kategoride sınıflandırmıştır.

### 7.2.1.1 Bilgi tabanlı yöntemler

Bilgiye dayalı yöntemlerde insan yüzü hakkında genel bilgiler gereklidir. Bu yöntem ile birlikte insan yüzü bilgisinden türetilen kurallar ile yüz algılama yöntemleri geliştirilir. Örnek olarak yüzün simetrik olan göz, burun ve ağız gibi özelliklerin görüntülerde yakalanarak yüz bölgesinin tespit edilme işleminin gerçekleştirilmesi. Yüz özellikleri arasındaki ilişkiler göreceli mesafeler, konumları ile temsil edilebilir. Giriş görüntüsündeki yüz özellikleri çıkartılarak yüz'e ait kodlanmış kurallara göre tanımlama işlemi gerçekleştirilir.

Bu yaklaşımdaki sorun insan bilgisinin iyi tanımlanmış kurallara dönüştürmedeki zorluklardır. Ayrıntılı kurallar bilgisinde, olası kurallara uymayan görüntülerdeki yüzleri tespit etmede başarısız olabilir, yanlış kural bilgileri verilebilir.

Yang ve Huang (1994) yüz tespiti için bilgi tabanlı yöntem kullanmıştır. Kullandıkları yöntemde, olası yüzler için giriş görüntüsü taranarak bir dizi kural uygulamışlardır. Yüksek seviyedeki kurallarda bir yüzün nasıl görüldüğünün tanımlarıdır, düşük seviyelerdeki kurallar ise yüz özelliklerinin ayrıntılarına dayanır.



**Şekil 7.2 :** (a)  $n = 1$ , orijinal görüntü. (b)  $n = 4$ . (c)  $n = 8$ . (d)  $n = 16$ . Orijinal ve karşılık gelen düşük çözünürlüklü görüntüler.

Şekil 7.2’de gösterildiği gibi her kare hücre, her bir pikselin yoğunluğunun, o hücrede bulunan piksellerin ortalama yoğunluğuyla değiştirildiği  $n \times n$  piksellerinden oluşur (Yang ve Huang, 1994).



**Şekil 7.3 :** Bilgi tabanlı yukarıdan aşağıya yöntemlerinde kullanılan tipik bir yüz (Yang ve Huang, 1994) .

Ortalama ve alt örnekleme ile çoklu çözülme hiyerarşisi oluşturulur ve bir örnek Şekil 7.2’de gösterilmektedir. Yüz adaylarını en düşük çözünürlükte konumlandırmak için kullanılan kodlanmış kuralların örnekleri şunları içerir. Yüzün orta kısmı Şekil 7.3’deki koyu gölgeli kısımlar. Temel olarak homojen bir yoğunluğa sahip dört hücreye sahiptir, yüzün üst yuvarlak kısmı Şekil 7.3’deki hafif gölgeli kısımlar temel olarak homojen bir yoğunluğa sahiptir ve merkezin ortalama gri değerleri arasındaki fark bölüm ve üst yuvarlak bölüm önemlidir.

En düşük çözünürlükteki (Seviye 1) görüntü yüz adayları için aranır ve bunlar daha ince çözünürlüklerde daha da işlenir. 2. Seviyede, 2. Seviyeden alınan yüz adaylarına lokal histogram eşitlemesi ve ardından kenar tespiti yapılır. Geri kalan aday bölgeler daha sonra Seviye 3’te gözler ve ağız gibi yüz özelliklerine cevap veren başka bir setle incelenir. Her ne kadar yüksek bir algılama oranına yol açmasa da, çoklu yüz hiyerarşisi ve aramaları yönlendirmek için kurallar kullanma düşünceleri sonraki yüz tanıma çalışmalarında kullanılmıştır (Kotopoulos ve Pitas, 1997).

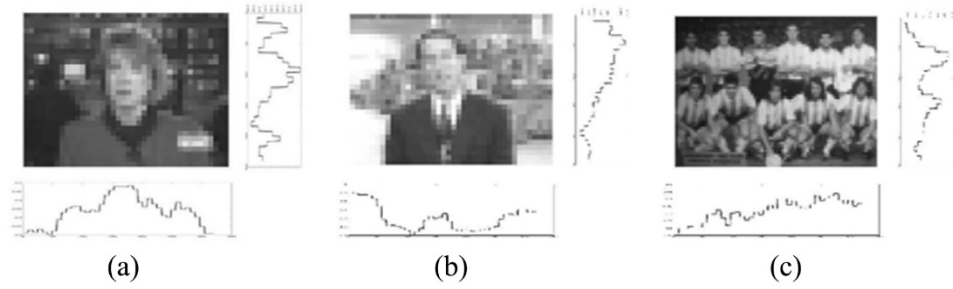
### **7.2.1.2 Özellik tabanlı yöntemler**

Bilgi tabanlı yöntemlerin yerine araştırmacılar yüz tespiti için yüzlerin değişmez özelliklerini bulmaya çalışmışlardır. Özellik tabanlı yöntemlerde, insanların farklı poz ve aydınlatma koşullarında dahi yüzleri ve nesnelere zahmetsizce tespit edebildikleri değişkenler üzerinde değişmeyen özelliklerin olması gerektiği bilgisine dayanmaktadır. Özellik tabanlı yöntem, yüzün yapısal özelliklerini çıkararak yüzleri

bulmaktadır. Yüz özelliklerini algılamak için sayısız yöntem önermişlerdir. Kaşlar, gözler, burun, ağız ve saç çizgisi gibi yüz özellikleri yayın olarak kenar tespit işlemi kullanılarak elde edilmektedir. Çıkarılan özelliklere dayanarak ilişkilerin tanımlanması ve bir yüz varlığını doğrulamak için istatistiksel model oluşturulur. Özellik tabanlı yöntemlerin dezavantajı görüntünün aydınlatma, gürültü gibi nedenlerden dolayı bozulabilmesidir. Bu yöntem gerçekleştirilirken özellik sınırları, yüz için zayıflatılabilirken, gölgeler çok sayıda güçlü kenara neden olabilir.

### Yüz hatları

Yüz hatlarını karışık arka plandan çıkarmak adına yerleştirme yöntemi önermiştir (Sirohey, 1993). Kenar haritalama yöntemiyle kenarların çıkartılması ve gruplanması için Canny (1986) kenar detektörü kullanılır.



**Şekil 7.4 :** Yatay ve dikey profiller (Yang ve diğ., 2002)

Yatay ve dikey profillerde tepe noktalar aranarak tek bir yüzü tespit etmek mümkündür. Ancak Şekil 7.4 (b) ve 7.4 (c) 'deki karmaşık arka plandaki yüzleri tespit etmekte zorlanır.

Baş bölgesi ve arka plan arasında fark karmaşık arka plan görüntüsüne sahip 48 görselde %80 lik bir doğruluk elde eder. Farklı algoritmalarda yüz modelinde gözler, elmacık kemikleri ve burnu temsil eden iki koyu renkli ve açık renkli bölgeler bulunur. Model, yüzlerin, kaşların ve dudakların ana hatlarını temsil etmek için çizgiler kullanır. Uzaysal ilişkiyi korumak için üçgen konfigürasyonu kullanılır. Bunun için Laplacian algoritması kullanılabilir. Daha sonra görüntü belirli üçgen oluşumları bulmak için taranarak çizgiler tespit edilirse yüz algılanması gerçekleştirilir.

## **Doku**

İnsan yüzü, farklı nesnelere ayırmak için kullanılacak ayrı bir dokuya sahiptir. Yüz dokularının tanımlanması ile bir yüzün varlığı tespit edilebilir. Üç tür özellik göz önünde bulundurulur: cilt, saç ve diğerleri.

## **Ten rengi**

İnsan ten rengi yüz tespiti, el takibi gibi birçok uygulamada etkili bir özellik olarak kullanılmıştır. Renk bilgileri yüz bölgelerini ve belirli yüz özelliklerini tanımlamak için etkili bir arabirimdir. Bununla birlikte, bu gibi ten rengi modelleri, ışık kaynağının spektrumunun önemli ölçüde değiştiği yerlerde etkili değildir. Başka bir deyişle, arka plandaki ve ön plan aydınlatmasındaki değişiklikler nedeniyle renk görünümü genellikle dengesizdir.

## **Çoklu özellikler**

Son zamanlarda, yüzleri bulmak veya tespit etmek için çeşitli yüz özelliklerini birleştiren çok sayıda yöntem önerilmiştir. Birçoğu yüz adaylarını bulmak için ten rengi, ebadı ve şekli gibi global özellikleri kullanır ve daha sonra kaş, burun ve saç gibi yerel, ayrıntılı özellikleri kullanarak bu adayları doğrular.

### **7.2.1.3 Şablon eşleme yöntemleri**

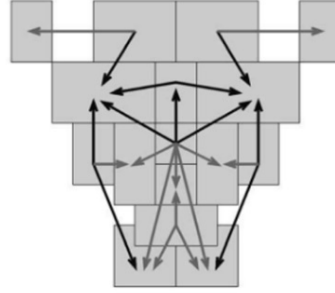
Şablon eşleme standart bir yüz deseni manuel olarak önceden tanımlanır veya bir fonksiyonla parametrelendirilir. Bir giriş görüntüsü verildiğinde, standart desenlerle olan korelasyon değerleri yüz konturu, gözler, burun ve ağız için bağımsız olarak hesaplanır. Bir yüzün varlığı, korelasyon değerlerine göre belirlenir. Örneğin insan yüzü gözler, yüz çevresi, burun ve ağız şekline bölünebilir. Ayrıca bir yüz modeli yalnızca kenar algılama yöntemi kullanılarak kenarlardan oluşturulabilir. Bu yaklaşımın uygulanması basit olma avantajına sahiptir. Bununla birlikte, yüz tanıma için yetersiz olduğu kanıtlanmıştır. Çünkü ölçek, poz ve şekildeki farklılıkları etkili bir şekilde ele alamaz.

#### **7.2.1.3.1 Önceden tanımlı şablonlar**

Bir yüzü modellemek için Şekil 7.5'deki gibi gözler, burun, ağız ve yüz çevresi için birçok alt plaka kullanılmıştır. Her alt öge, çizgi segmentleri cinsinden tanımlanır. Giriş görüntüsündeki çizgiler, en büyük eğim değişimine dayanarak ayıklanır ve sonra alt nesnelere göre eşleştirilir. Alt görüntüler ve kontur şablonları arasındaki



korelasyon, yüzlerin aday konumlarını tespit etmek için ilk önce hesaplanır. Daha sonra diğer alt noktalarla eşleştirme aday pozisyonlarında gerçekleştirilir. Başka bir deyişle, birinci aşama ilgi odağını veya ilgi alanını belirler ve ikinci aşama bir yüzün varlığını belirlemek için detayları inceler. Dikkat ve alt konulara odaklanma fikri, daha sonra yüz tanıma çalışmaları ile benimsenir.



**Şekil 7.5 : 14x16 piksel oranı şablon yöntemi**

#### **7.2.1.4 Görünüm tabanlı yöntemler**

Şablonların uzmanlar tarafından önceden tanımlandığı şablon eşleştirme yöntemlerinin aksine, görünüm tabanlı yöntemlerde “şablonlar” resimlerdeki örneklerden öğrenilir. Yüz modellerini bulmak için bir dizi eğitim verisi gerekmektedir. Genel olarak, görünüme dayalı yöntemler, yüz ve yüz dışı görüntülerin ilgili özelliklerini bulmak için istatistiksel analiz ve makine öğrenmesinden elde edilen tekniklere dayanır. Öğrenilen özellikler, sonuç olarak yüz tespiti için kullanılan dağıtım modelleri veya ayırt edici işlevler biçimindedir. Görünüm tabanlı yöntemler yüz tanıma için özellik çıkarımında da kullanılır.

#### **7.2.2 Dudak lokasyon tespiti**

Dudak ve ağız bölgesi, insan konuşma üretim sisteminin görsel parçalarıdır. Bu parçalar görsel konuşma bilgisini saklar, bu nedenle herhangi bir görsel konuşma tanıma (AVSR) sisteminin görsel bilgiyi yakalamak için ağız ve dudak bölgelerini tespit edip yerelleştirmesi zorunludur. Dudak okuma yaklaşımı için dudakların görünmesi gerekmektedir. İnsan yüzü bilgisinde dudaklar yüzün alt orta kısmında yer almaktadır.

Gelişen teknolojiyle birlikte son on yılda yüz algılama, yüz tanıma, görsel konuşma tanıma ve yüz ifadesinden duygu tanıma gibi insan yüzlerini otomatik işlenmesi ve analizi ile ilgili uygulamaların sayısı giderek artmıştır. Bu tür sistemler ileriki

yıllarda daha fazla artarak gündelik yaşantımızın önemli bir parçası haline gelecektir. Tıpkı Heartrow havaalanında yüz tanıma sistemiyle güvenlik gereksiniminin karşılanması gibi. Bu tip uygulamalar dudak ve ağız bölgesiyle özel olarak ilgilenmektedir. Bu tür uygulamalar için, sağlam ve gerçek zamanlı dudak algılama / lokalizasyon yöntemi, güvenilirlik ve başarısına katkıda bulunan önemli bir faktördür. Dudak bölgesi yüzün en fazla deforme olabilen kısmı olduğu için bu bölgenin saptanması, takibi, aydınlatma koşullarındaki değişiklik, poz, kafa rotasyonu, yüz ifadelerindeki değişiklikler analiz ve işlem sırasında performansı olumsuz etkileyen faktörler arasında önemli bir problem olarak yer alır.

Literatürde yer alan şimdiye kadarki dudak tespit yöntemleri kendi güçlü ve zayıf yönlerine sahiptir. Genel olarak doğruluk süreleri arttığı zaman işlem sürelerinin artmasına neden olur.

Dijital görüntülerde dudak algılama / lokalizasyon için birçok teknik literatürde bildirilmiştir ve iki ana çözüm türü olarak kategorize edilebilir (Hassanat, 2009);

- Model bazlı dudak algılama yöntemleri. Bu tür modeller, Snakes, Aktif Şekil Modelleri (ASM), Aktif Görünüm Modelleri (AAM) ve deforme edilebilir şablonlar olarak adlandırılan spline tabanlı deformasyonların kullanımını içerir.
- Görüntü tabanlı dudak algılama yöntemleri. Bunlar uzaysal bilgi, piksel rengi ve yoğunluğu, çizgiler, köşeler, kenarlar ve hareket kullanımını içerir.

#### **7.2.2.1 Model tabanlı dudak algılama yöntemleri**

Model tabanlı yöntemler, fotoğrafta beklenenden önceki bir modeli kullanır ve tipik olarak modelin yeni bir görüntüyle en iyi uyuşmasını bulmaya çalışır. Modeli eşleştirdikten sonra, hedefin gerçekten var olup olmadığını ölçmek veya test yapmak mümkündür (Cootes, 2000) .Yeni bir görüntü, modelin görüntü verisiyle en iyi şekilde uyuşabileceği eşleşmeyi bularak yorumlanabilir.

Böyle bir yöntemin avantajları

- Yaygın olarak uygulanabilir. Aynı algoritma, yalnızca farklı eğitim örnekleri sunarak, birçok farklı soruna uygulanabilir.
- Eğitim örneklerinin ek açıklamasında sistemde uzman bilgisi yakalanabilir.

- Modeller izin verilebilir varyasyonun kompakt bir gösterimini sağlar, ancak eğitim setinde görülenlerden farklı bir keyfi varyasyona izin vermeyecek kadar özeldir.
- Sistemin, eğitim setinden öğrendiklerinden başka, modellenmekte olan nesnelerin doğası hakkında birkaç önceden varsayımda bulunması gerekir

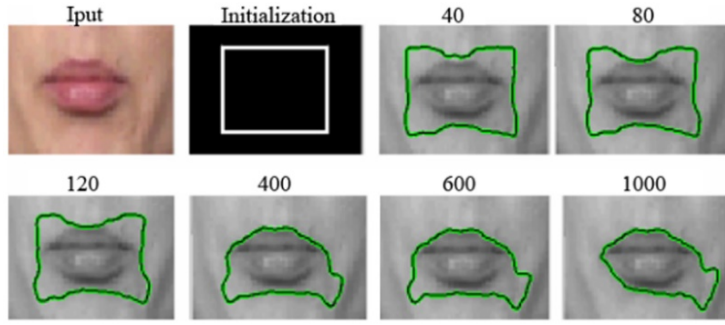
Bu yaklaşım, eğitim yüz görüntülerini olsun ya da olmasın dudak modellerini veya girilen bir görüntüdeki dudakları aramak için tanımlanmış modeli kullanmaya bağlıdır. Belirlenen kriterlere göre modele en uygun olan tespit edilen dudakların yeri olarak beyan edilir. Bu yöntemler Aktif Konturlar (Yılanlar), Aktif Şekil Modelleri, Aktif Görünüm Modelleri ve deforme şablonlar içerir.

### Snakes

Snakes veya aktif kontur modelleri ilk kez Kass, Witkin ve Terzopoulos tarafından tanıtılmıştır (Kass ve diğ., 1988). Snakes, şekil tespiti için genel bir yöntem olarak önerilmiştir. Optimizasyon teknikleri, spline'lara dayanmaktadır. Fikir, lokal minimuma uyması için spline enerjilerini yinelemeli olarak en aza indirmektir. Snakes yöntemi ile algılanan şekil arasındaki en uygun uyum, 6.1 denklemindeki enerjileri en aza indirerek bulunur:

$$E_{snake} = \int_0^1 (E_{int}S(i) + E_{img}S(i) + E_{con}S(i))di \quad (7.1)$$

Snakes, dış sınırlama kuvvetleri tarafından yönlendirilen onu çizgiler ve kenarlar gibi özelliklere doğru çeken görüntü kuvvetlerinden etkilenen, enerjiyi en aza indiren bir splinedir. Snakes yöntemi, aktif kontur modellerdir ve yakındaki kenarlara kilitlenir ve onları tam olarak konumlandırır. Ölçek boşluğunun devamı, bir özelliği çevreleyen yakalama bölgesini büyütme için kullanılabilir. Snakes yöntemi, kenarların, çizgilerin ve sübjektif konturların tespiti, hareket takibi için kullanılabilir.



**Şekil 7.6 :** Snakes yaklaşımı kullanarak dudak lokasyon bölümlemesi.

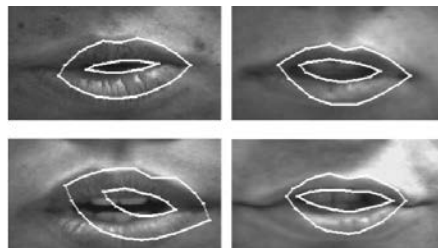
Her görüntü üzerindeki sayı, Şekil 7.6’da ayarlanan Snakes parametreleri için gereken yineleme sayısını göstermektedir (Hassanat, 2009).



**Şekil 7.7 :** Snakes yaklaşımı ile örnek bir dudak bölümlemesi (Kass ve diğ., 1988)

### **Aktif şekil modelleri**

Bu model, bir uzmanı tarafından açıklanan bir dizi görüntüden eğitilmiştir. Eğitim seti üzerinde şekil ve görünümdeki farklılıkları analiz ederek, değişimi taklit edebilecek bir model oluşturulur. Yeni bir görüntüyü yorumlamak için görüntüye bir model örneği ile en iyi eşleşen parametreleri bulmalıyız. Bunu verimli bir şekilde yapabilen bir algoritma tanımlarız. Şekil 7.8’deki gibi modeli görüntüye sığdırdıktan sonra, parametreler veya model noktası konumları, sınıflandırma veya ölçüm yapmak için daha sonraki işlemlere girdi olarak kullanılabilir (Cootes, 2000).



### Şekil 7.8 : Dudak tespiti için geleneksel ASM sonucu (Jang, 2007)

Bir dudak yapısını bulmak için önce onun modelini oluşturmalıyız. İstatistiksel bir görünüm modeli oluşturmak için, tipik örneklerin açıklanmalı görüntülerini gerektirir. Öncelikle, hedefin şeklini tanımlayan ve her eğitim görüntüsünde güvenilir bir şekilde bulunabilecek uygun bir simge setine karar vermeliyiz. Aktif şekil modelleri (ASM), özellikle dudak ve ağız bölgesi olmak üzere yüz özellikleri tespiti için başarıyla kullanılabilir.

#### **Aktif görünüm modelleri**

Aktif Görünüm Modelleri (AAM), ASM ile aynıdır, ancak bir dönüm noktası boyunca kenar profilini kullanmak yerine, AAM şekil bilgisiyle birlikte tüm görüntünün gri ölçek bilgilerini içerecek şekilde genişletilir (Cootes ve diğ., 1998). Bir modelin geçerli örnekleri olmak üzere çözümleri kısıtlayarak güçlü sonuçlar elde edebilirler. Ek olarak, bir görüntüyü bir dizi model parametresi olarak açıklayabilme yeteneği, sahne yorumlaması için doğal bir temel oluşturur. Bu faydaları gerçekleştirmek için, nesne görünüm modeli mümkün olduğu kadar eksiksiz olmalıdır. Hedef nesnenin herhangi bir görüntüsüne çok yakın bir yaklaşım sentezleyebilmelidir.

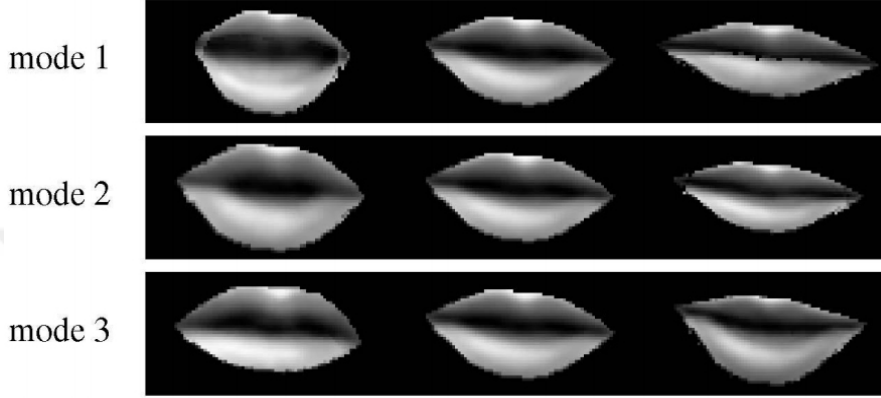
Bir AAM, ilgilenilen nesnenin neredeyse geçerli olan herhangi bir örneğe genelleyecek şekli ve gri seviye görünümünün istatistiksel bir modelini içerir. Bir eğitim aşaması sırasında, model parametresi yer değiştirmeleri ile bir eğitim görüntüsü ile sentezlenmiş bir model örneği arasında oluşan artık hatalar arasındaki ilişkiyi öğreniriz. Bir resimle eşleştirmek için mevcut kalıntıları ölçeriz ve mevcut parametrelerdeki değişiklikleri tahmin etmek için modeli kullanırız, bu da daha iyi bir uyum sağlar. İyi bir genel eşleşme, birkaç başlangıçta, zayıf başlangıç tahminlerinden bile elde edilir.

ASM ve AAM arasında üç temel fark vardır (Cootes ve diğ., 1998):

- ASM, imge noktalarının etrafındaki görüntü dokusunun modellerini kullanır, oysa AAM tüm bölgenin görüntü dokusunun bir modelini kullanır.
- ASM sınır etrafında arama yapar, oysa AAM sadece görüntüyü geçerli konum altında gösterir.

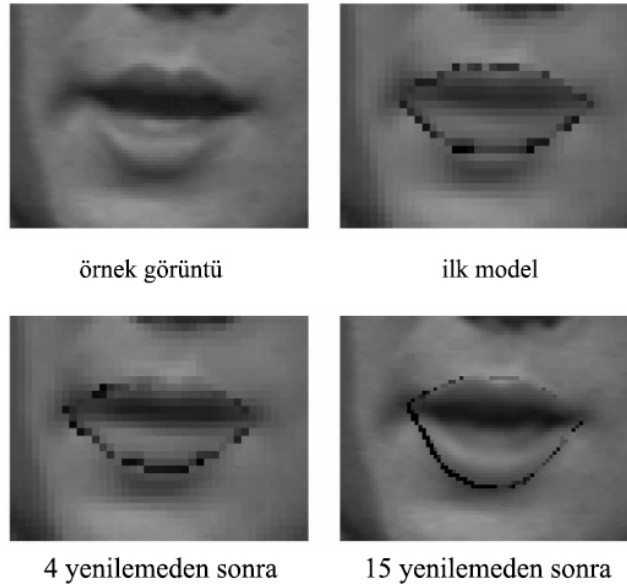
- ASM, model ile görüntü noktaları arasındaki mesafeyi en aza indirmeye çalışırken, AAM sentezlenmiş model görüntüsü ve hedef imaj arasındaki farkı en aza indirmeyi amaçlamaktadır.

Şekil 7.9'da, belirli bir veri tabanında eğitilmiş birleşik görünüm modelinin  $\pm 2$  standart sapma ortalamasında ilk üç varyasyon modu gösterilmiştir (Matthews ve diğ., 2002).



**Şekil 7.9 :** Şekil ve gri seviye görünüm modeli ,

Matthews ve diğ. (2002), Şekil 7.10' da AAM yöntemini kullanarak 15 yinelemeden sonra dudak sapmatama sonuçlarını göstermiştir.

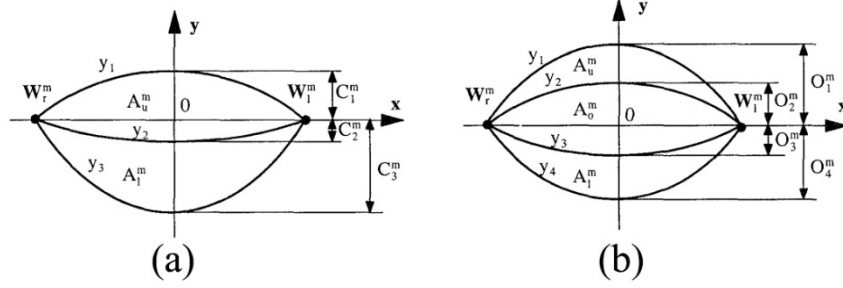


**Şekil 7.10 :** AAM araştırması örneği

Snakes yönteminden farklı olarak AAM yöntemi ile az sayıda yineleme işlemi gerçekleştirilir, çünkü modeller şekil hakkında daha önceden bazı bilgilere sahiptir.

## Deforme edilebilir şablonlar

Deforme edilebilir şablon, belirli bir nesnenin hareketlerini izlemek için kullanılan parametrel bir matematiksel modeldir (Yuille ve diğ., 1989). Verilen nesneye uyması için kendini adapte edebildiği için aktif modeldir. Deforme olabilen dudak şablonu, şeklini ilgili konturlar boyunca bir dizi integralin değerine göre şekil 7.11'deki gibi ayarlar.



**Şekil 7.11 :** Ağzın deforme edilebilir şablonlar tarafından çıkarılması (a) ağzı kapalı deforme edilebilir şablon, (b) ağzın açık deforme edilebilir şablon (Zhang, 1997).

Deforme olabilen bir şablona dayanarak dudak saptama ve / veya takibi ile uğraşan diğer çalışmalar arasında Zhang (1997) bulunmaktadır. Bu model ile, bir ağzın açık yada kapalı olup olmadığını belirleyen algoritmaların geliştirilmesini kolaylaştırır.

### 7.2.2.2 Görüntü tabanlı dudak algılama yöntemleri

Dudaklar ve ağız bölgesi, insan konuşma sisteminin ana görsel kısmını oluşturmaktadır. Bu bölümler görsel konuşma bilgisinin çıkarıldığı kısımlardır. Dudak bölgesinin segmentasyonu dudak hatlarına odaklanarak dudak bölgesinin izlenmesi için farklı yöntemler önermiştir. Görüntü tabanlı yöntemlerde renk  $v$  kenarlar gibi bölge işaretleme yardımıyla bir dizi karakteristik nokta tanımlanabilir. Bu tür yöntemler, noktaların başlangıç pozisyonuna karşı hassas ve belirsizdir.

Dudak ve cilt etrafındaki yüzün rengi arasında bir fark olduğu için renk bilgilerini kullanarak dudakları tespit etmek son zamanlarda araştırmacıların ilgisini çekmektedir. Dudak tespiti işleminde renk farkı tekniğinin kullanılması zamandan ve kaynak kullanımından tasarruf sağlayarak düşük bellek kullanımı ile umut verici yöntemlere izin vermektedir. RGB, HSV, YCbCr gibi renk sistemlerini kullanarak çeşitli araştırmacıların çalışma konuları arasına girmiştir. Bazı araştırmacılar dudak kenarları ve dudak hareketlerinden karma yaklaşım kullanmışlardır (Eveno ve diğ., 2002).

Dudak özellik çıkarımının ana özelliklerinden birisi sistemin doğruluğu ve güvenilirliğidir. Yapılan araştırmalar renk bilgisinin güçlü bir araç olduğunu göstermektedir. Gri seviyeli yaklaşıma ek olarak renkli görüntü analizinde dudakların yerinin bulunması, sakal, diş ve dili belirlemek kolaylaşmıştır. Renk temelli tekniklerde piksel ve komşuluk düzeyinde çalışarak renk özelliklerine göre pikselleri ayırt etmeye ve renklendirmeye çalışır. Dudak izleme algoritmaları genel olarak RGB görüntüsüne uygun olarak renk dönüşüm işlemini gerçekleştirir. Dudak rengi ile dudakların etrafındaki yüz bölgesinin rengi arasındaki fark renk bilgisini kullanarak dudakların tespit edilmesi zaman alıcı olmadığından daha az kaynak kullanımı ile daha hızlı işlemler gerçekleştirilerek ortaya çıkan renk bilgisini kullanarak dudak algılama için gelecek vaat eden farklı yöntemler sunar. Gerçekleştirilen sistemlerde RGB deki kırmızı ve yeşil renkler HSV renk sisteminin renk tonu bilgisi ve YCbCr renk sistemindeki kırmızı ve mavi bileşeni içerir. Görüntü tabanlı yöntemler hesaplama açısından hızlı ancak aydınlatmadan olumsuz yönde etkilenirler. Görüntü verileri düşük seviyeden incelenerek nesnelere tanımlamak amacıyla bölge ve kenar gibi yerel yapılar arar.

RGB renk uzayında, cilt ve dudak pikselleri farklı bileşenlere sahiptir. Her ikisi içinde kırmızı yaygındır. Ayrıca ten rengi karışımında ve dudaklarda maviden daha fazla yeşil renk pikselleri vardır. Deri, dudaklardan daha sarı görünür çünkü kırmızı ve yeşil arasındaki fark cilt bölgesinde daha fazladır (Eveno ve diğ., 2002).

Görüntüdeki ten rengi birçok faktörden etkilenir (Sneha, 2017) ;

- Aydınlatma: Işık kaynağı dağılımında ve iç mekân, dış mekân, açık tonlar, gölgeler ve beyaz olmayan ışıklar gibi aydınlatma seviyesindeki bir değişiklik, görüntüdeki cildin renginde bir değişime neden olur, yani renk sıkıntısı sorunu. Aydınlatma varyasyonu performansı ciddi şekilde düşürür.
- Kamera özellikleri: Aynı aydınlatma altında bile, aynı kişinin ten rengi dağılımı, kamera sensörü özelliklerine bağlı olarak bir kameradan diğerine değişmektedir.
- Etnik: Ten rengi, kişiden kişiye ve belirli etnik gruplara, bölgelere bağlı olarak da değişmektedir. Asya, Afrika, Kafkas ve İspanyol gruplarına ait insanların ten rengi birbirinden farklıdır.
- Bireysel özellikler: Yaş, vücut bölgeleri ve cinsiyet gibi bireysel özellikler ten rengini de etkiler.



- Diğer faktörler: Makyaj, saç modeli ve gözlük gibi görünümleri, arka plan renkleri, gölgeler ve hareket gibi farklı faktörler de ten rengi görünümünü etkiler.

### **7.3 Görüntü Özellik Çıkarımı**

Görsel işitsel konuşma tanıma sistemlerinin çoğu, görsel ve işitsel sinyallerin birbirini tamamlaması prensibinden dolayı özellik çıkarımı ve görsel konuşma özelliği tanıma şeklinde iki ana adımdan oluşmaktadır. Özellik çıkarma işlemi için mevcut yaklaşımlar

#### **7.3.1 Görünüm Özelliklerine Dayalı Yaklaşımlar**

Görünüme dayalı yöntemler, ağız bölgesinin gri ya da renkli görüntülerine uygulanarak piksel değerlerini göz önünde bulundurur. Normal olarak özellik çıkarılacak ilgili bölge bir tür boyut azaltma yaklaşımları kullanılır.

Eigenlips, PCA kullanarak ağız bölgesinin çıkarım işlemini gerçekleştiren yöntemdir. Bu yaklaşım Turk ve Pentland (1991) yöntemlerinden esinlenmiştir. Görsel ve işitsel kaynakların tamamlayıcılığından faydalanmayı amaçlayan Eigenlips tabanlı sistem dikey dudak hareketlerinin ve ağız yapısının özellik çıkarım işleminin bazı filtreleme teknikleri kullanarak dudak okuma yöntemi işlemini gerçekleştirmektedir.

Hazen ve arkadaşları (2004), segment bazlı modelleme stratejisi kullanarak konuşmacıdan bağımsız görsel işitsel konuşma tanıma (AVSR) sistemi geliştirmişlerdir. Bu sistemde, konuşmacıya ait dudak bölgesinin görsel ölçümlerinden Gizli Markov Modeli (HMM) kullanarak toplar. Başka yöntemlerde de SVM-HMM sistemi ile dudakların manuel algılanmasını sağlayan sistemlerde geliştirilmiştir.

#### **7.3.2 Geometrik özelliklere dayalı yaklaşımlar**

Geometrik özelliklere dayalı yaklaşımlar Petajan tarafından 1984 yılında konuşma tanıma sistemlerinin performansını arttırmak için sistem tasarımına sahip ilk VSR çalışmasını içermektedir (Petajan, 1984). Bu yaklaşıma göre özellik ağız boyutu, genişliği, alanı ve çevresi gibi geometrik özellikleri kullanmaya dayanıyordu. Bu özellikler, basit bir eşikleme tekniği kullanılarak yüz görüntülerinden elde edilen ağız ikili görüntülerinden elde edilir. Petajan, konuşma tanıma sisteminin performansını

artıran tanıma için Dinamik Zaman Atlama (DTW) yöntemini kullandı (Petajan, 1984). Geometrik özelliklere dayalı olarak başka bir çalışmada, Werda ve arkadaşları (2006) tarafından dudak lokasyonu, dudak takibi, görsel özellik çıkarma ve konuşma tanıma işlemlerinin gerçekleştirildiği otomatik dudak özellik çıkarma prototipine sahip (ALİFE) çalışmasıdır. Deneyleri, doğal koşullar altında birden fazla konuşmacı (kadın ve erkek) tarafından söylenen %72.73 Fransızca ünlülerin doğruluğunu vermiştir

### **7.3.3 Görüntü dönüşümü temelli yaklaşımlar**

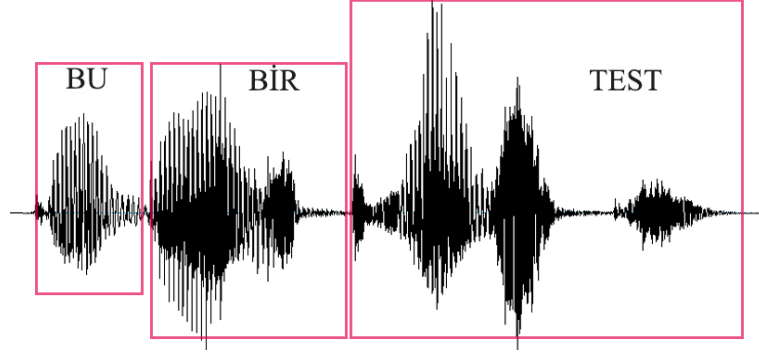
Sridharan ve Lucey (2008), görsel-işitsel konuşma tanıma sistemlerinde başın pozisyonuna bakılmaksızın konuşma tanıma sistemi geliştirmişlerdir. Görsel özellikleri çıkarmadan önce yüz tespit ve tahmini için (Viola ve Jones, 2001) yöntemini kullandılar. Görsel özellik çıkarma için LDA ile DCT görüntü boyutlarını düşürerek HMM sınıflandırma işlemini kullanmışlardır.

DCT katsayılarından en ayırt edici özellik vektörlerini çıkarmak için DCT'yi ağız bölgesinden özellik çıkarımı için kullanmışlardır (Jun ve Hua, 2009). LDA kullanılarak boyut azaltma işlemini gerçekleştirmişlerdir. Ayrıca, kelimeleri tanımak için HMM kullanmışlardır.

## **7.4 Ses Ön İşleme**

İşitme duyumuz, ses üreten nesnelere yerleri ve özellikleri hakkında zengin bilgiler sağlar. Herhangi bir kuşun sesi, yada enstrüman eşliğinde çalan bir müziğin şarkı sözlerine eşlik edebiliriz. İşitme sistemimiz, kulaklarımıza ulaşan karmaşık sesleri işleyebilir ve ölçülen duyuşal girdilerin analizi ve gruplandırmasıyla üst düzey bilgiler elde edebilir. Akustik sinyalden kaynakların ayrıştırılması ve tanımlanması sürecine işitsel analiz olarak yorumlanabilir (Rao P. , 2007).

Konuşma, bilgi alışverişinin en önemli, etkili ve uygun şeklidir. Konuşma ile insanlar arası iletişimin en doğal şekli sağlanır. Otomatik Konuşma Tanıma sisteminde, ortam gürültüsünün istenmediği ve elde edilecek özelliklerin en iyi şekilde temsil edileceği koşullarda ön işlem aşaması önemlidir. Konuşma sinyalinin ön işlenmesinde, gürültülü sesin önlenmesinde, konuşmanın ön vurgulanmasında ve boyutsallığın azaltılmasında sistemin yıllarca daha verimli olmasını sağlayan bir tekniktir.



**Şekil 7.12** : Konuşma dalgabıçımı

Konuşma tanıma sistemlerinde ilk olarak Şekil 7.12’de gösterilen konuşma dalgabıçımı bilgisayarın anlayabileceği şekle dönüştürülmelidir. Bu sebeple konuşma sinyali sayısal hale getirilerek frekans ya da ses perdesi gibi parametreler belirlenir. Bu parametrelerin zamanla değişim şekli dalgabıçımını, ses içeren birimlere ayrıştırılarak kullanılır. Dalgabıçımından metne dönüştürme işlemi zorlu ve karmaşık bir süreçtir. Telaffuz edilen sözcüklerin başlangıç ve bitişi karmaşık örüntüler içinde üst üste biner ve insandan insana kelimelerin telaffuzu farklılık gösterir. Örnek olarak masa kelimesi [mm a ss ha] yada [m aaaa ss haa] şeklinde farklı telaffuz edilebilir.

Sesler, gerçek dünyadaki zamana göre değişen sinyallerdir ve anlamlarının tümü bu zaman değişkenliği ile ilgilidir. Bu nedenle, anlama, karşılaştırma, değiştirme ve yeniden sentez işlemlerini kolaylaştırmak için zamanla değişen seslerin ayırt edici özelliklerinden faydalanmak için ses analizi işlemleri gerçekleştirilir (Rocchesso, 2003).

Ön işleme, konuşma algılayıcının ilk kısmıdır ve özellik çıkarma işleminden önce hazırlık seviyesi olarak yorumlanabilir. Yararlı konuşma bilgisi, uygun, optimum konuşma tanımayı elde etmek için ön işleme bölümünde uygun bir biçimde çıkarılmalı ve temsil edilmelidir. Ham sesin ön düzenlemesi ve işlenmesi önemli bir standart işlem adıdır. Tipik olarak temiz bir ses dosyası için çeşitli istenmeyen öksürük, hıçkırık, rahatsız edici tepe noktaları gibi bölümlerin kaldırılarak sesin normal hale getirilmesi gerekmektedir.

Konuşma tanıma sistemlerinin geliştirilmesinde ön işleme adımı, sinyal ayırt etme ve özellik vektörleri oluşturmak için konuşma tanımanın ilk aşaması olarak kabul edilmektedir. Ön işleme adımı ile özellik çıkarma analiz işlemleri için daha kabul

edilebilir olmasını sağlar. Özellik çıkarma işlemi ayrıca sınıflandırma için kullanılan modellere uygun bir forma sinyali dönüştürmeyi içermektedir.

Bir konuşma tanıma sisteminde aranan birkaç özellik;

- Kelimeler arasında yüksek ayırım gücü.
- Düşük kullanıcı değişikliği.
- Konuşma sinyalindeki gürültüden kaynaklanan bozulmaların azlığı.

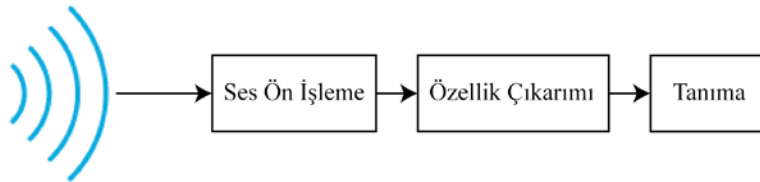
Amaç, konuşma sinyalinde akustik korelasyona sahip bir ifadenin bir dizi özelliğini bulmak, yani sinyal dalga biçiminin işlenmesiyle bir şekilde hesaplanabilen veya tahmin edilebilen parametrelere sahip özellikleri bulmaktır. Sinyal modelinde konuşma sinyalinin ön işlenmesinden sonraki adım, özellik çıkarmadır. Özellik çıkarmayı, konuşma sinyalinin parametrelendirilmesidir. Konuşma sinyalinin algısal olarak anlamlı bir gösteriminin üretilmesi amaçlanır. Özellik çıkarmayı tipik olarak sinyali dijital forma dönüştürerek enerji veya frekans gibi sinyalin önemli karakterlerini ölçümler (Singh ve diğ., 2012).

Özellik Çıkarmayı hedefleri (Singh ve diğ., 2012) ;

- Konuşma sinyalini akustik olarak tanımlanabilir çeşitli bileşenlere ayırmak.
- Hesaplamaları mümkün kılmak için düşük değişim oranlı bir dizi özellik elde etmek.

Bu teknolojiler ile birlikte bilgisayarların insanın en doğal iletişim biçimi olan konuşmayı tanımları amaçlanmaktadır.

Özellik çıkarmayı üç temel işleme bölünebilir: spektral analiz, parametrik dönüşüm ve istatistiksel modellemedir.



**Şekil 7.13 : Konuşma Tanıma Süreci**

Konuşma işleme bir dizi ilgi alanı kapsar;

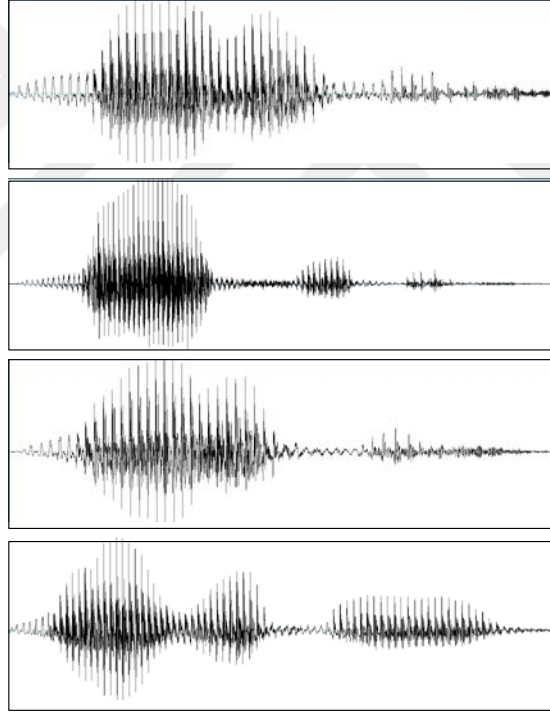
- Konuşma tanıma: konuşma sinyali bilgisinden dile ait içerik çıkarma,
- Konuşmacı tanıma: konuşmacıların kimlik bilgisinin sesleriyle tanıma,

- Konuşma Sentezi: Telekomünikasyon için konuşma sinyallerinin sıkıştırılması işlemi,
- Konuşma geliştirme: konuşma sinyallerinin anlaşılabilirliğini ve algılanma kalitesinin artırılması veya iyileştirilmesi.

Konuşma işleme uygulama alanları:

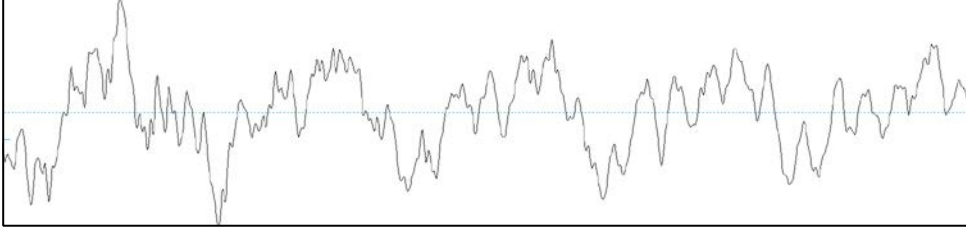
- İnsan-Bilgisayar arayüzleri (konuşma girişi, duygusal durum tanıma)
- Telekomünikasyon (konuşma çeviri, konuşma geliştirme)
- Yardımcı Teknolojileri (işitme engelliler, görme engelliler )
- Güvenlik (biyometri, adli tıp, savunma sanayii)

Bir konuşma sırasında konuşmanın şekli ve hızı kişiden kişiye değişmektedir. Bir kişi “merhaba” kelimesini Şekil 7.14’de gösterildiği gibi farklı hızlarda ve sürelerde söyler. Konuşma bilgilerinde çok farklı ve uzun ses bilgileri üretilebilmektedir.



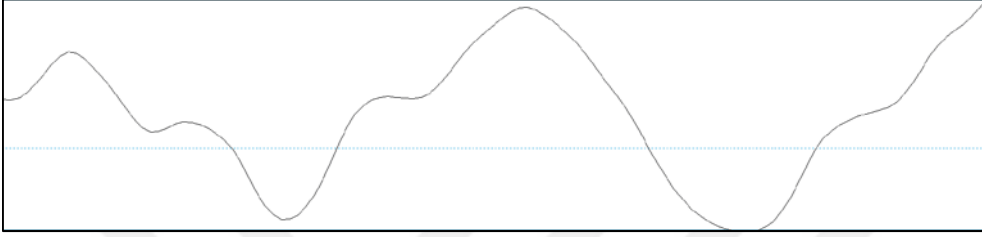
**Şekil 7.14 :** Merhaba kelimesinin aynı kişi tarafından farklı söylenişi

Ses dalgaları bilgisayarın anlayacağı bir şekilde bitlere çevrilerek sayısallaştırılması gerekmektedir. Konuşma tanıma sistemlerindeki ilk adım ses dalgalarını bilgisayara beslememiz gerekir. Fakat ses, dalgalar şeklinde iletilir. “Merhaba” diyen bir ses dalgası Şekil 7.15’de gösterildiği gibidir.

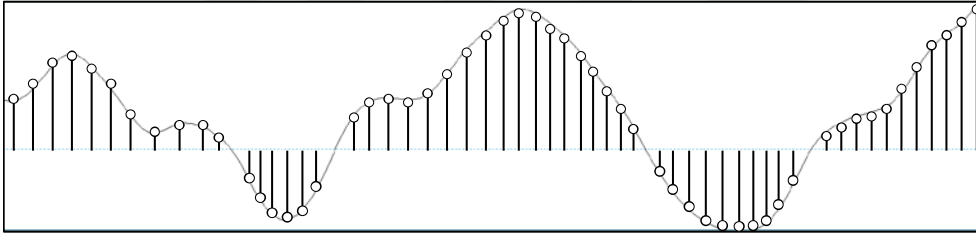


**Şekil 7.15** : Merhaba kelimesinin dalga formu kesiti

Ses dalgaları Şekil 7.15’ de gösterildiği gibi zamanın her anında dalganın yüksekliğine bağlı olarak tek bir değere sahiptir.

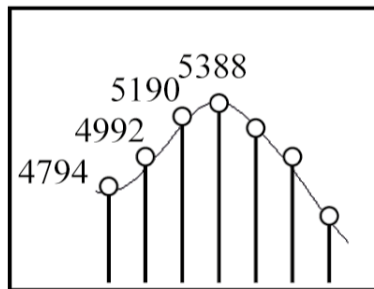


**Şekil 7.16** : Ses dalgasının tek boyutlu yükseklik gösterimi



**Şekil 7.17** : Ses dalgasının bir kesitinin örneklenmiş gösterimi

Ses dalgası örnekleme ile saniyede binlerce defa okunarak o andaki ses dalgasının yüksekliğini temsil eden bir sayı ile kaydedilir. Temel olarak Şekil 7.17’de sıkıştırılmamış bir “Merhaba” ses dosyasının kesitinin örneklemesidir. Bir CD kalitesinde sesi 44.1 kHz ‘de (saniyede 44.100) örneklenir. Fakat konuşma tanıma sistemlerinde 16 kHz (saniyede 16.000) örnekleme oranı, insan konuşmasının frekans aralığını kapsamak için yeterli bir değerdir.



**Şekil 7.18** : Merhaba Ses dalgasının ilk 4 örnekleme değerleri

Her sayı değeri ikinci dalganın 1/16000 'inde ses dalgasının genliğini temsil etmektedir. Örnekleme orijinal bir ses dalgasının tahmini değerleri olarak Şekil 7.18'deki gibi düşünebiliriz, çünkü birebir tüm sinyal değerlerinin örnekleme işlemi gerçekleştirilmemektedir. Okumalar arasındaki boşluklarda veri kaybının olması muhtemeldir.

Her ses dalgasının genliğini, 1/16000 'inci ikinci aralıklarla temsil eden bir sayı dizisine sahibiz. Ancak bu örnekleri doğrudan işleyerek konuşma tanıma zor bir süreçtir. Bunun yerine ses verisinde bazı ön işlemleri yaparak sorunu daha kolay hale getirebiliriz. Örneklenmiş sesler 20 milisaniye uzunluğunda parçalar halinde grupta işlemi yapabiliriz. Şekil 7.19'da gösterildiği gibi ilk 20 milisaniye için toplam 320 sayısal değeri elde ederiz.

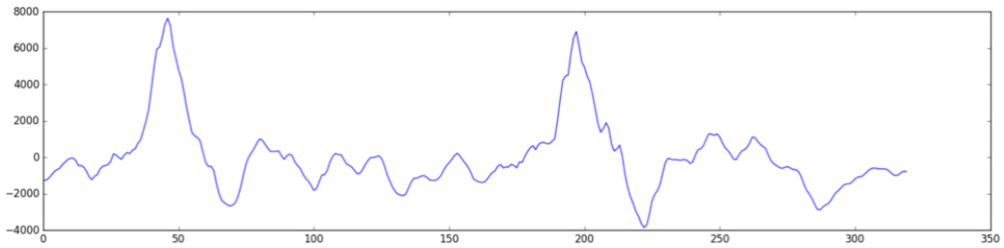
```

-1274, -1252, -1160, -986, -792, -692, -614, -429, -286, -134, -57, -41, -169, -456, -450, -541, -761, -1067, -1231, -1047, -952, -645, -489, -448,
-397, -212, 193, 114, -17, -110, 128, 261, 198, 390, 461, 772, 948, 1451, 1974, 2624, 3793, 4968, 5939, 6057, 6581, 7302, 7640, 7223, 6119, 5461,
4820, 4353, 3611, 2740, 2004, 1349, 1178, 1085, 901, 301, -262, -499, -488, -707, -1406, -1997, -2377, -2494, -2605, -2675, -2627, -2500, -2148, -
1648, -970, -364, 13, 260, 494, 788, 1011, 938, 717, 507, 323, 324, 325, 350, 103, -113, 64, 176, 93, -249, -461, -606, -909, -1159, -1307, -1544, -
1815, -1725, -1341, -971, -959, -723, -261, 51, 210, 142, 152, -92, -345, -439, -529, -710, -907, -887, -693, -403, -180, -14, -12, 29, 89, -47, -
398, -896, -1262, -1610, -1862, -2021, -2077, -2105, -2023, -1697, -1360, -1150, -1148, -1091, -1013, -1018, -1126, -1255, -1270, -1266, -1174, -10
83, -707, -468, -300, -116, 92, 224, 72, -150, -336, -541, -820, -1178, -1289, -1345, -1385, -1365, -1223, -1004, -839, -734, -481, -396, -580, -52
7, -531, -376, -458, -581, -254, -277, 50, 331, 531, 641, 416, 697, 810, 812, 759, 739, 888, 1008, 1977, 3145, 4219, 4454, 4521, 5691, 6563, 6909,
6117, 5244, 4951, 4462, 4124, 3435, 2671, 1847, 1370, 1591, 1900, 1586, 713, 341, 462, 673, 60, -938, -1664, -2185, -2527, -2967, -3253, -3636, -38
59, -3723, -3134, -2380, -2032, -1831, -1457, -804, -241, -51, -113, -136, -122, -158, -147, -114, -181, -338, -266, 131, 418, 471, 651, 994, 1295,
1267, 1197, 1291, 1110, 793, 514, 370, 174, -90, -139, 104, 334, 407, 524, 771, 1106, 1087, 878, 703, 591, 471, 91, -199, -357, -454, -561, -605,
-552, -512, -575, -669, -672, -763, -1022, -1435, -1791, -1999, -2242, -2563, -2853, -2893, -2740, -2625, -2556, -2385, -2138, -1936, -1803, -1649,
-1495, -1460, -1446, -1345, -1177, -1088, -1072, -1003, -856, -719, -621, -585, -613, -634, -638, -636, -683, -819, -946, -1012, -964, -836, -762,
7881

```

Şekil 7.19 : İlk 20 milisaniyelik değerler

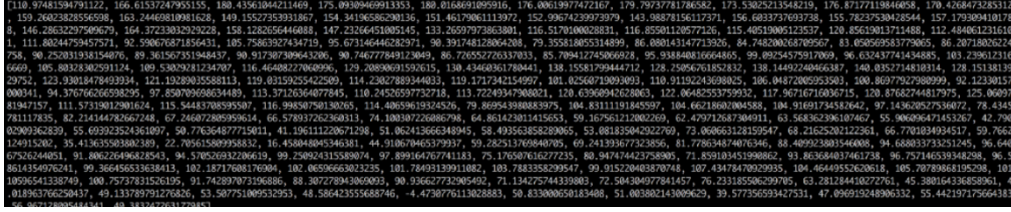
Bu değerlerin basit bir grafik olarak çizilmesi Şekil 7.20'de gösterildiği gibi bize 20 milisaniyelik süreye göre orijinal ses dalgasının kaba bir gösterimini verir.



Şekil 7.20 : Değerlerin gösterimi

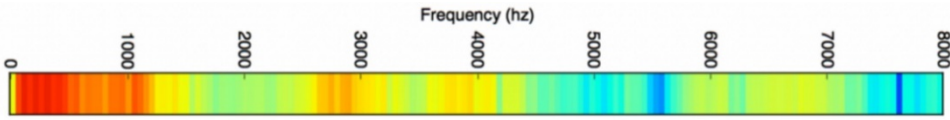
Kısa ses kayıtları bile karmaşık ses frekanslarını barındırır. Düşük sesler, orta sesler, yüksek sesler insan konuşmasının karmaşık ses yapısını oluşturmaktadır. Bu ses değerlerinin ne kadar enerjiye sahip olduğunu Fourier dönüşümleri ile gerçekleştirmekteyiz. Bu dönüşümler karmaşık ses dalgalarını onu oluşturan basit ses dalgalarına ayırır. Böylece ses dalgalarının ne kadar enerji değerine sahip olduğunu bulmamızı sağlar. Sonuç olarak her frekans aralığı düşük perdeden yüksek perdeye kadar bir aralığa sahiptir. Şekil 7.21'de 20 milisaniyelik ses bölümünün her 50 Hz bandındaki enerji değerlerini gösterir.





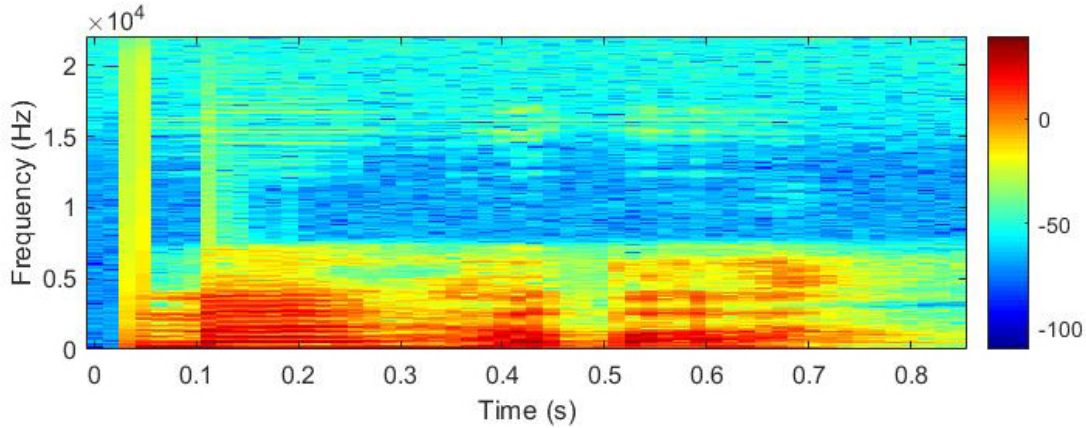
Şekil 7.21 : Enerji değerleri

Sayısal değerler ses bilgisi hakkında anlaşılır bir karşılık ifade edemez. Bu değerlerin grafiksel gösterimi, ses yapısının anlaşılmasını kolaylaştırmaktadır.



Şekil 7.22 : Ses frekans değerleri

Ses frekans değerleri Şekil 7.22'de gösterildiği gibidir. Bu işlemler 20 milisaniyelik ses parçasında tekrarlanırsa Şekil 7.23 deki gibi bir ses spektrogramı ile sonuçlanır.



Şekil 7.23 : Merhaba Ses dosyasının tam spektrogramı

Bir spektrogram, bir sinyalin frekans spektrumunun görsel bir gösterimidir.

#### 7.4.1 Filtreleme

Filtreleme işlemi iletişim sistemlerinde kanal eşitleme, gürültü azaltma, ses işleme, video işleme, biyomedikal sinyal işleme verilerin analizi gibi uygulamalarda yaygın olarak kullanılmaktadır. Bir konuşma belli başlı duraklamalar ile ayrılmış çeşitli bileşenlere sahiptir. Her bileşenin başlangıç ve bitiş noktaları vardır. Başlangıç ve bitiş noktasını belirleyen başlı enerji kıvrımlarına sahiptir. Başlangıç noktasını takip eden yüksek bir kenar ile bitiş noktasından önceki bir alçak kenar özelliği vardır.

Bir konuşma sinyali  $s(n)$  filtreleme işlemine gönderilir;



$$s_2(n) = s(n) - a*s(n-1) \quad (7.2)$$

Buradaki  $s_2(n)$  çıkış sinylidir ve  $a$ 'nın değeri genellikle 0.9 ile 4.0 arasındadır. Filtrenin z-dönüşümü

$$H(z) = 1 - a*z^{-1} \quad (7.3)$$

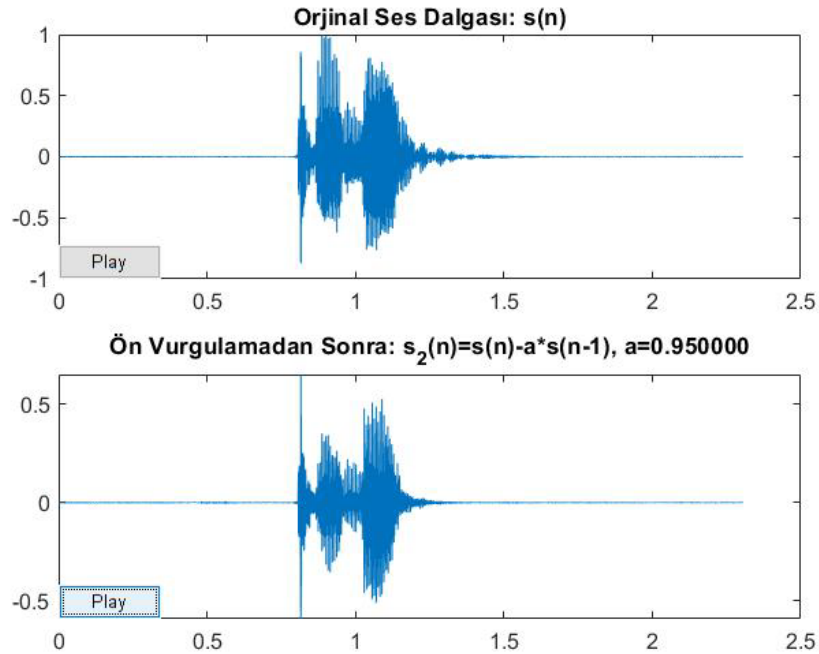
Şekil 7.24'de gösterildiği gibi ön vurgulamanın amacı, insanların ses üretim mekanizmaları sırasındaki yüksek frekanslı kısmı telafi ederek bu frekansların öneminin artırılmasını sağlamaktır.

```

onVurgulama.m
1 - waveFile='kalem.wav';
2 - [y,fs]=audioread(waveFile);
3 - a=0.95;
4 - y2 = filter([1, -a], 1, y);
5 - time=(1:length(y))/fs;
6 - [y1,fs1]=audioread(waveFile);
7 - audiowrite('kalem.wav',y1,fs1);
8 - subplot(2,1,1);
9 - plot(time, y);
10 - title('Orjinal Ses Dalgası: s(n)');
11 - subplot(2,1,2);
12 - plot(time, y2);
13 - title(sprintf('Ön Vurgulamadan Sonra: s_2(n)=s(n)-a*s(n-1), a=%f', a));
14
15 - subplot(2,1,1);
16 - set(gca, 'unit', 'pixel');
17 - axisPos=get(gca, 'position');
18 - uicontrol('string', 'Play', 'position', [axisPos(1:2), 60, 20], 'callback', 'sound(y, fs)');
19 - subplot(2,1,2);
20 - set(gca, 'unit', 'pixel');
21 - axisPos=get(gca, 'position');
22 - uicontrol('string', 'Play', 'position', [axisPos(1:2), 60, 20], 'callback', 'sound(y2, fs)');

```

Şekil 7.24 : Ön vurgulama Matlab kodu





gücüne ihtiyaç duyulmasına rağmen çerçevelerin parametre değerlerinde daha yumuşak bir değişikliğe neden olabilir.

Bir sonraki adımda, çerçeveler içindeki konuşma sinyaline belirli ağırlıkların atandığı bir pencereleme işlemi gerçekleştirilir. Giriş konuşma sinyali, çerçeve boyutunun  $1/3 \sim 1/2$  'si isteğe bağlı olarak 20~30 ms'lik karelere bölünmektedir. FFT kullanımını kolaylaştırmak için genellikle çerçeve boyutları iki kuvvet değerine sahiptir.

Örnekleme hızı 16 kHz ve çerçeve boyutunun 320 örnekleme noktasına sahip olduğunu varsayarsak;

Kare süresi,

$$\frac{320}{16000} = 0.02 \text{ sn (20ms)}. \quad (7.4)$$

Ayrıca, örtüşme 160 nokta ile temsil ediliyorsa kare hızı,

$$\frac{16000}{320 - 160} = 100 \text{ (saniyedeki kare sayısı)} \quad (7.5)$$

Çerçeveleme, blok şeklinde işlemeyi kolaylaştırmak için sürekli konuşma örnekleri akışını sabit uzunlukta bileşenlere ayırma işlemidir. Aynı şekilde, konuşma yarı durağan bir sinyal olarak düşünülebilir.

Pencereleme işlemi;

- İşleme için en kullanışlı olan çerçevenin ortasındaki bilgilere vurgu yapmak,
- Yüksek frekanslı seslerin etkisini azaltarak spektral sızıntıyı önlemek.

Sonuç olarak, konuşma sinyali zaman içinde (yarı-durağan) yavaşça değişmektedir, yani sinyal kısa bir süre boyunca (5-100 msn) incelendiğinde, sinyal oldukça sabittir. Bu nedenle, konuşma sinyalleri çoğu zaman konuşma işlemede kısa süreli spektral analiz olarak da adlandırılan kısa zaman bileşenlerinde analiz edilir.

Bu basitçe, sinyalin tipik olarak 20-30 msn karelere bölünmesi veya engellenmesi anlamına gelir. Bu yönde, bitişik çerçeveler normal olarak %30-50 ile birbirleriyle örtüşürler. Bu, pencereden dolayı konuşma sinyalinin hayati bir bilgisini kaybetmemek için yapılır. Çerçeve çok kısaysa, güvenilir bir spektral tahmin elde

etmek için yeterli örneğe sahip olamayız, eğer daha uzunsu sinyal çerçeve boyunca çok fazla değişir.

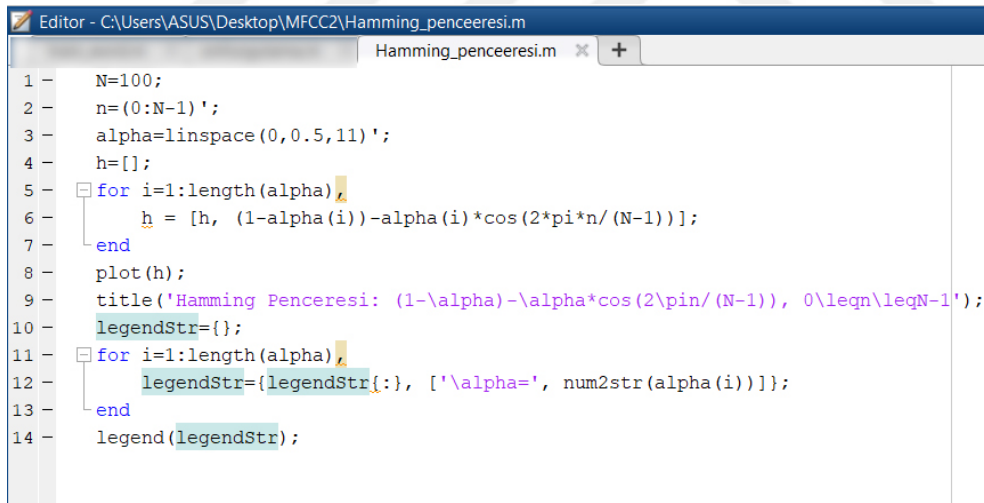
### 7.4.3 Pencereleme

İkinci aşama olarak bütün çerçevelerin pencereleme işleminin gerçekleştirilmesidir. Pencereleme yöntemiyle çerçevenin başında ve sonunda bulunan süreksizlik gibi bilgilerin ortadan kaldırılmasıdır. Hamming pencereleme fonksiyonu en çok kullanılan pencereleme fonksiyonlarından birisidir. Hamming pencerelemesi ile her çerçevenin, çerçevedeki ilk ve son noktaların sürekliliğini sağlamak için bir sonraki pencerede bir hamming penceresi ile çarpılması gerekir.

Bir çerçevedeki sinyal  $s(n), n = 0, \dots, N - 1$  ile gösterilirse, Hamming pencerelemesinden sonraki sinyal  $s(n) * w(n)$ , burada  $w(n)$  tanımlanmış Hamming penceresi tarafından ;

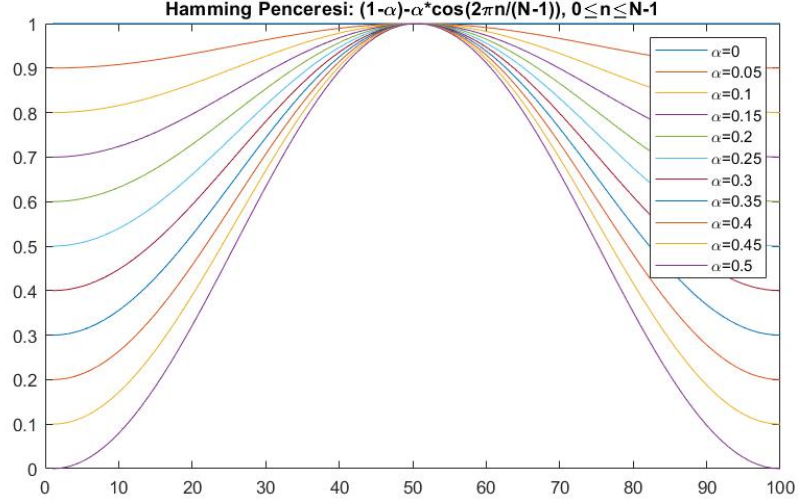
$$w(n, \alpha) = (1 - \alpha) - \alpha \cos(2\pi n / (N - 1)), \quad 0 \leq n \leq N - 1 \quad (7.6)$$

sonraki Hamming pencereleri için farklı değerlere karşılık gelir.



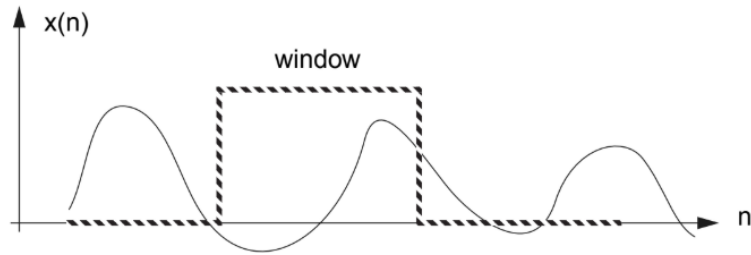
```
Editor - C:\Users\ASUS\Desktop\MFCC2\Hamming_penceeresi.m
Hamming_penceeresi.m x +
1 - N=100;
2 - n=(0:N-1)';
3 - alpha=linspace(0,0.5,11)';
4 - h=[];
5 - for i=1:length(alpha)
6 -     h = [h, (1-alpha(i))-alpha(i)*cos(2*pi*n/(N-1))];
7 - end
8 - plot(h);
9 - title('Hamming Penceresi: (1-\alpha)-\alpha*cos(2\pin/(N-1)), 0\leqn\leqN-1');
10 - legendStr={};
11 - for i=1:length(alpha)
12 -     legendStr={legendStr{:}, ['\alpha=', num2str(alpha(i))]};
13 - end
14 - legend(legendStr);
```

Şekil 7.28 : Hamming penceresi Matlab kodu



**Şekil 7.29 :** Hamming penceresi görseli

Bu aşamada, sinyal bölümlere göre çerçevenmiştir. Her bir çerçeve, N'nin çerçevenin uzunluğu olduğu  $N(n)$  uzunluğunda bir pencere fonksiyonu ile çarpılmaktadır. Pencereleme, sinyalin önceden tanımlanmış özelliklerini vurgulamak için bir konuşma sinyali bölümünün dalga biçimini verilen bir şekle sahip bir zaman penceresi ile çarpma işlemidir. Konuşma sinyalinin süreksizliğini her karenin başında ve sonunda azaltmak için, sinyal sıfıra veya sıfıra yakın bir noktaya konulmalı ve bu nedenle uyumsuzluğu en aza indirilmelidir. Mel Frekanslı Kepstrum Katsayıları (MFCC) ve ardışık kareler arasındaki spektral tahminler arasındaki korelasyonu arttırmak için sinyalin her bir karesinin pencerelenmesi ile elde edilebilir. Şekil 7.30'daki gibi pencereleme işlemi, sinyalin yumuşatılmasında yardımcı olduğu için sinyalin kesilmesinden kaynaklanan problemleri önlemek için her zaman bir konuşma sinyaline uygulanır.



**Şekil 7.30 :** Bir sinyalin penceresi

Yaygın pencere işlevlerinin türleri aşağıda verilmiştir (İbrahim, Odiketa, & Ibiyemi, 2017):

Dikdörtgen pencereleme :

$$w(n) = 1, 0 \leq n \leq M - 1 \quad (7.7)$$

Üçgen pencereleme :

$$w(n) = 1 - \left[ 1 - \frac{2n}{M-1} \right], 0 \leq n \leq M - 1 \quad (7.8)$$

Hanning pencerelemesi :

$$w(n) = 0.5 - 0.5 \cos \left( \frac{2n\pi}{M-1} \right), 0 \leq n \leq M - 1 \quad (7.9)$$

Hamming pencerelemesi :

$$w(n) = 0.54 - 0.46 \cos \left( \frac{2n\pi}{M-1} \right), 0 \leq n \leq M - 1 \quad (7.10)$$

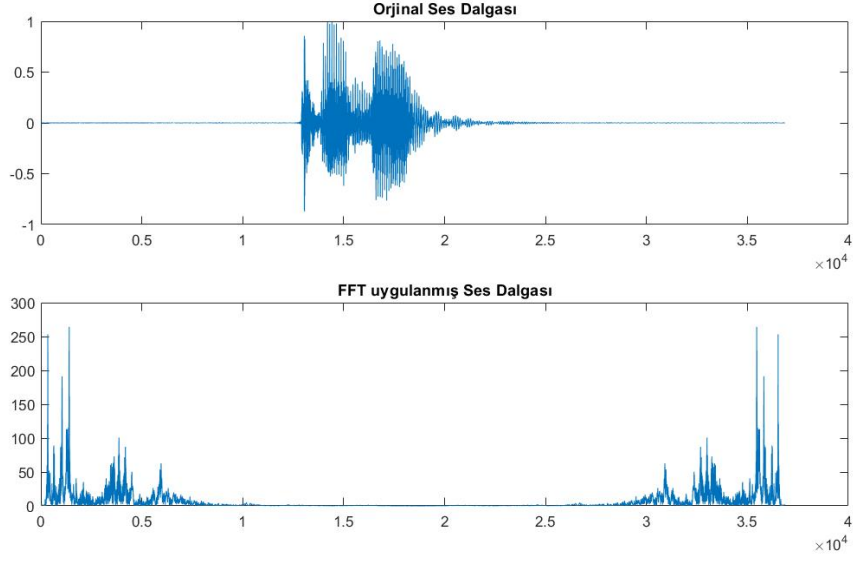
Barlett pencerelemesi :

$$w(n) = 0.42 - 0.5 \cos \left( \frac{2n\pi}{M-1} \right) + 0.08 \cos \left( \frac{4n\pi}{M-1} \right), 0 \leq n \leq M - 1 \quad (7.11)$$

#### 7.4.4 Hızlı Fourier Dönüşümü (FFT)

İnsan konuşması ses tellerinin titreşimi yoluyla zaman içerisinde değişen bir sinyal aralığında üretildiği için özellikleri spektral aktivitenin parametrelenmesi şekliyle gösterilebilir. Konuşma sinyali zaman aralığı şeklinde gösterilmesinden ziyade, genliği, sıklığı ve fazı olarak gösterilir.

Pencere uzunluğu ile ilgili bir denge vardır: pencere uzunluğu küçükse, bir yandan kısa sürede meydana gelen olayı tanıma (örneğin patlayıcı maddeler) daha kolay olurken, bir yandan frekans spektrumu tahmini zayıf olacaktır.



**Şekil 7.31 : FFT uygulanmış ses dalgası**

Bir sinyalin sahip olduğu frekansların ölçülebilmesi için Şekil 7.31’de gösterildiği gibi Hızlı Fourier dönüşümleri (FFT) sinyal işlemede sıkça kullanılır. Dönüşüm işlemi ile bir sinyal farklı parametrelerle ifade edilir. FFT, sistemimizde bulunan sinyallerin hangi türden olduğunu görmemize yardımcı olur. Karmaşık sinyalleri ayrı sinüs dalgalarına böler. Herhangi bir sinyal, farklı sinüs dalgalarının (Fourier Dönüşümü) toplamı olarak düşünülebilir. Hızlı Fourier dönüşümleri (FFT), spektral analiz, konuşma sinyallerindeki farklı zaman aralıklarının frekanslar üzerindeki farklı enerji dağılımına karşılık geldiğini göstermek ve ses özellik çıkarımında her karenin frekans bilgilerini elde etmek için kullanılır.

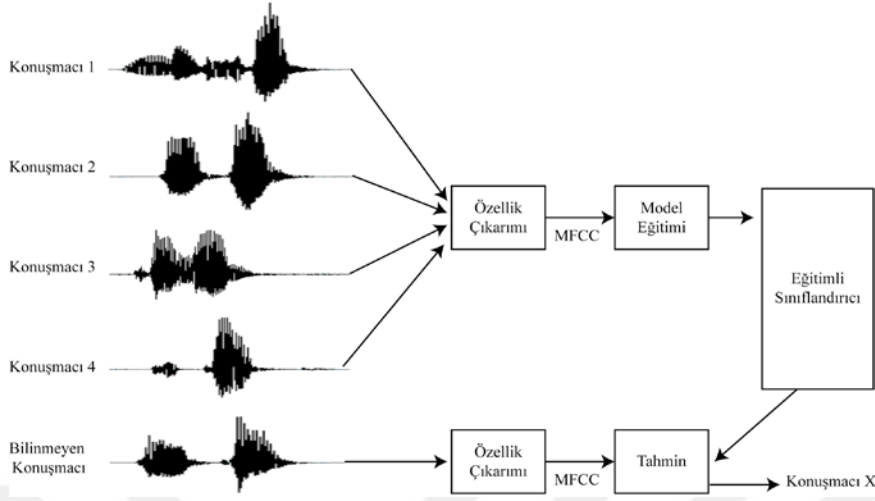
Bir çerçeve üzerinde FFT işlemi gerçekleştirdiğimiz zaman çerçeve içindeki sinyalin periyodik ve sürekli olduğunu varsayıyoruz. İlk ve son noktalarda süreklilik etkisini arttırmak için her çerçeve Hamming penceresi ile çarpılır.

#### **7.4.5 Mel frekans kepsral katsayıları**

Genel olarak bir konuşma insan tarafından üretilen seslerin dil, dişler ya da ses yolu tarafından filtrelenmesidir. Bu şekilde ses üretim işlemi gerçekleşir. Ses yolunun şekli güç spektrumu şeklinde gözükmektedir. Mel frekans kepsral katsayıları (MFCC), otomatik konuşma ve konuşmacı tanımda yaygın olarak kullanılan bir özelliktir. MFCC, frekanslara göre insan algı hassasiyetini dikkate alır.

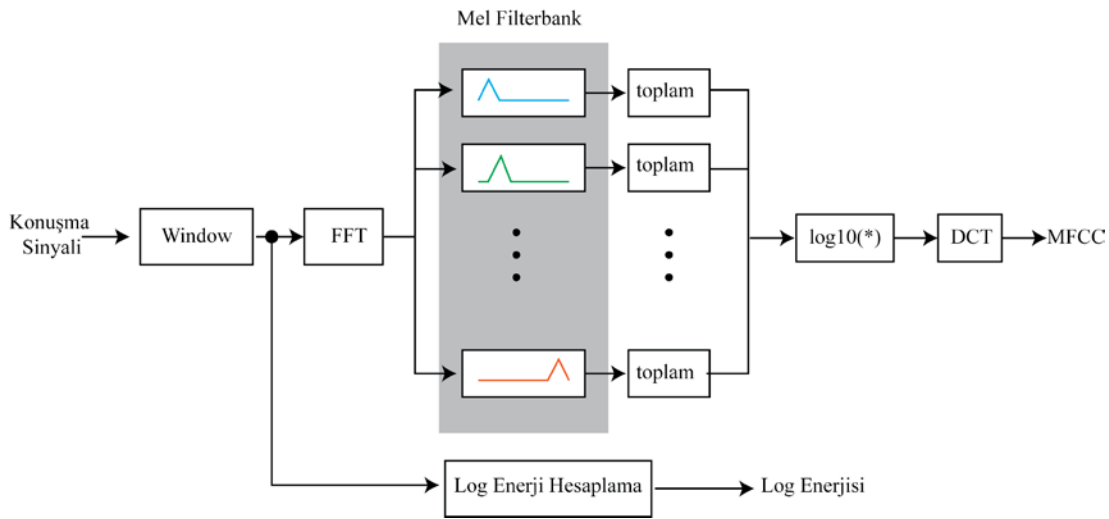
Konuşma tanımayla birlikte konuşmacı tanıma sistemlerinde de konuşma özellik çıkarım işlemleri sıklıkla kullanılır. Kaydedilen konuşmalardan elde edilen

özelliklere dayanarak insan seslerini tanımlamak için Şekil 7.23’de gösterildiği gibi bir makine öğrenme yaklaşımı uygulanır.



Şekil 7.32 : Konuşmacı tanımlaması için kullanılan yaklaşım şeması.

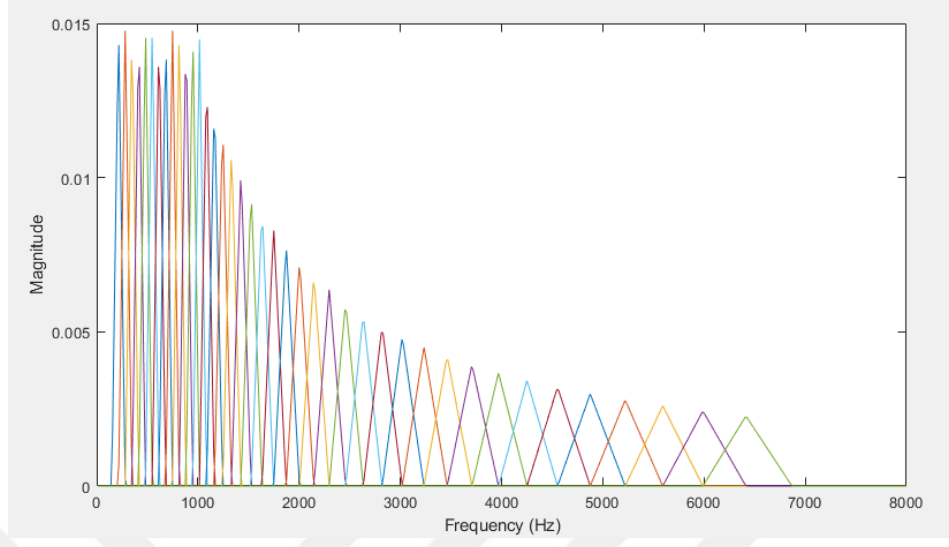
Mel frekansı, saf bir tonun algılanan frekansını veya adını gerçek ölçülen frekansıyla ilişkilendirir. İnsanlar düşük frekanslarda küçük değişiklikleri ayırt etmede yüksek frekanslarda olduklarından daha iyidir. Bu ölçeği kullanmak, özelliklerin insanların duyduklarıyla daha yakından eşleşmesini sağlar. MFCC, tanıma görevlerinde kullanılmak üzere konuşma sinyallerinden elde edilen özelliklerdir. Konuşmanın kaynak filtre modelindeki ses yolunu temsil ettiği bilinmektedir. MFCC, koklea anlayışına dayanarak Şekil 7.33’deki gibi ses yolu hakkındaki bilgileri az sayıda katsayıya sıkıştırılmaktadır.



Şekil 7.33 : MFCC hesaplama adımları.



Mel filtresi Şekil 7.34’de gösterildiği gibi ilk 10 üçgen filtreyi doğrusal olarak, kalan filtreleri logaritmik olarak yerleştirir. Tek tek bantlar enerji için bile ağırlıklıdır.



Şekil 7.34 : Mel filterbank görseli.



## 8. KONUŞMA TERAPİSİ UYGULAMA TASARIMI

Dudak algılama ve segmentasyonu, dudak okuma, görsel ve işitsel konuşma tanıma, yüz ifade analizi gibi alanlardaki potansiyel uygulamaları nedeniyle araştırma alanında yoğun ilgi görmüştür. Görsel-işitsel konuşma tanıma sistemleri robotik, kişisel bilgisayarlar ve cep telefonlarında yaygın olarak kullanılmaktadır. Sadece işitsel olarak konuşma tanıma sistemleri gürültülü ortamlarda performansı düşmektedir.

Dudak okuma sistemlerinde dudak segmentasyonu sonraki adımda işlenecek temel bilgileri sağladığı için en önemli ve kritik ön işlem adımına sahiptir. Araştırmacılar tarafından gerçekleştirilen birden farklı dudak segmentasyonu yöntemi bulunmaktadır.

Kişiler arası çeşitli duygu, düşünce ve bilgilerin görsel ve akustik yöntemlerle karşı tarafa iletilmesi sözlü iletişim ile gerçekleşir. Geleneksel olarak bilgi işleme sistemleri tek bir model üzerine odaklanmıştır. Konuşma algı ve üretim gereği bimodal bir süreçtir (Chen ve Rao, 1998). İnsan konuşması, vokal kordun titreşimi ve burun boşluğu, dil, dişler ve dudaklar dahil olmak üzere artikülatör organlardan oluşan vokal yolun konfigürasyonu ile üretilir. Bu artikülatör organları kullanarak, yüz ifadeleri oluşturan kaslarla birlikte bir konuşma eylemi gerçekleştirilir. Artikülatörlerin bazıları görülebilir olduğundan, akustik ve görünür konuşma arasında doğal bir ilişki vardır. İnsan konuşmasının bimodal doğası McGurk etkisinde en uygun şekilde gösterilebilir (McGurk ve MacDonald, 1976). McGurk etkisi, insanların çatışan akustik ve görsel uyaranlarla sunulduğunda, algılanan sesin her iki yöntemde de bulunmasının zor olduğunu göstermektedir. Görsel-işitsel konuşma işleminin amacı, tanıma, kodlama, geliştirme ve konuşmacı tanıma için konuşmayı en iyi şekilde işlemek amacıyla, konuşmanın akustik ve görsel özelliklerinden faydalanmaktır.

Görsel işitsel konuşma dudak okuma bilgileri ve akustik konuşmayı yakalamak için çok sayıda görüntü işleme ve ses işleme yöntemlerini kullanan çok disiplinli bir

çalışma alanıdır. Konuşma tanımada, ilgili konuşma ve görüntü bilgilerinin eş zamanlı analizini ele alır.

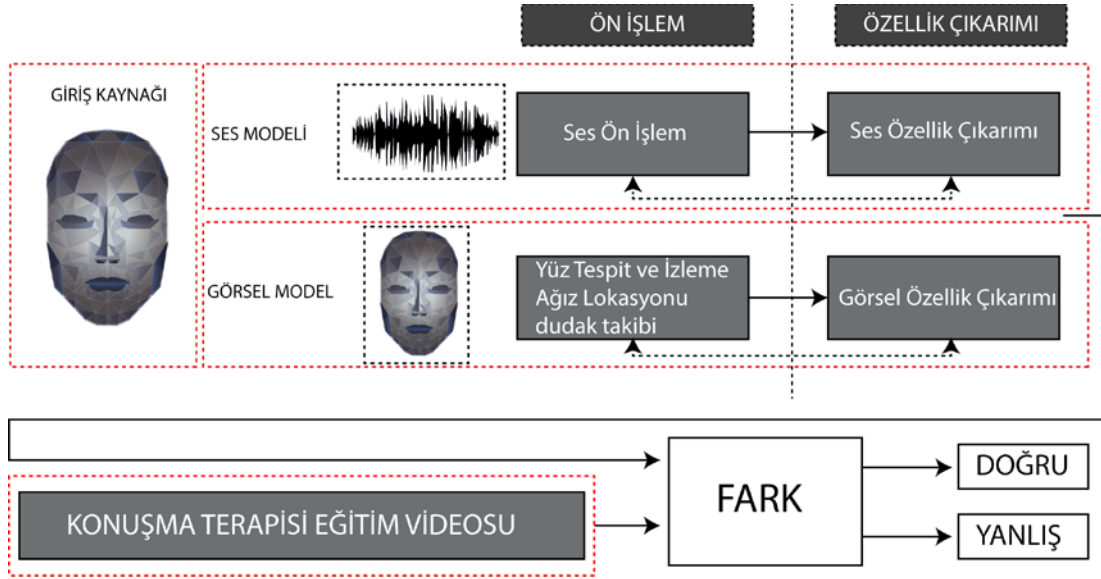
Akustik tabanlı konuşma tanıma sisteminin doğru kelime oranı sessiz ortamlarda %95 'ten fazladır, ancak gürültülü ortamlarda bu oran zayıflar. Bunun en büyük nedeni etkin bir gürültü ayırım işlemini gerçekleştirememesidir. Dudak okuma sisteminde doğru kelime oranı görsel ipuçlarının belirsizliğinden dolayı tek başına yeterli olamamaktadır. Bunun sebebi ise farklı fonemlerin aynı görüntülere karşılık gelebilmesi. Bununla birlikte görsel dudak okuma sistemleri gürültülü ortamlarda sesten etkilenmediği için güçlü bir şekilde çalışmaktadır. İşitme engellilerin, konuşma tanıma sisteminin doğru kelime telaffuzu için dudak hareketleri bilgilerini kullanarak akustik konuşma tanıma sistemiyle görsel konuşma tanıma sisteminin birleştirilmesiyle görsel işitsel konuşma tanıma sistemi oluşturmayı amaçlamaktayız.

Bozukluk insanların konuşma tanıma, konuşma artikulatorlerinin motor kontrolünden yoksun olması nedeniyle zor bir görevdir. Düzensiz konuşmanın sağlığını geliştirmek için çoklu yöntem konuşma tanıma kullanılabilir. Çalışmamızda, konuşma ve görsel bileşenlere dayalı işitme engelliler için konuşma eğitimiyle birlikte konuşma tanıma sistemi tanıtılmaktadır. Mel-Frekanslı Kestirme Katsayıları (MFCC), akustik konuşma sinyalini temsil eden özellikler olarak kullanılır. Görsel taraf için DCT, konuşmacının ağız bölgesinden çıkarılır.

Konuşulan kelimeler arasındaki konuşmacının doğruluk seviyelerini kontrol edilerek konuşmasının geliştirilmesi sağlanabilir. Ses sinyal analiz işlemlerinde spektral analiz işlemlerini uygulayarak sesin biçim bilgisi zaman ve frekans alanındaki standart özelliklerini elde ederek ses ve görüntü işleme çalışmaları gerçekleştirilir.

## **8.1 Sistem Mimarisi**

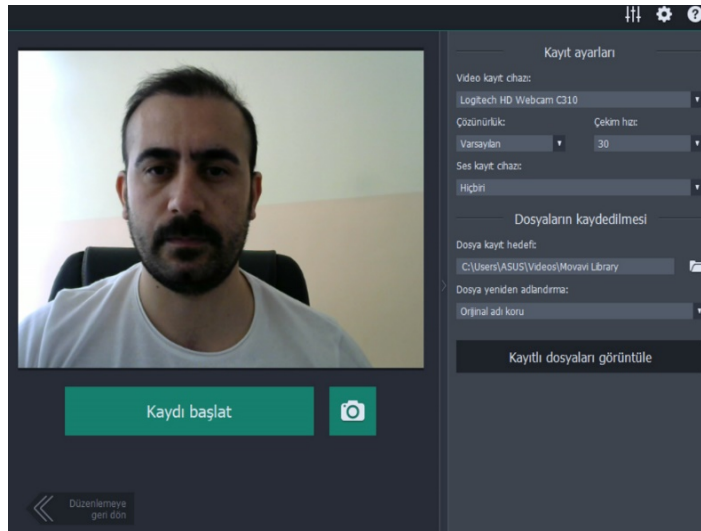
İşitme engelliler için konuşma terapi uygulamamızla birlikte işitsel konuşma tanıma sistemi Şekil 8.1'deki gibi tasarlanmıştır. Kullanıcı konuşulan kelimeye özgü eğitim videosunu izledikten sonra kendisinden beklenen kelimenin telaffuz şeklini program girişine vermektedir. İlgili veriler ses ve görsel olarak bölünür. İki çalışma parçacığı, ses ön ucu ve görüntü ön ucu vardır.



**Şekil 8.1 : Sistem mimarisi**

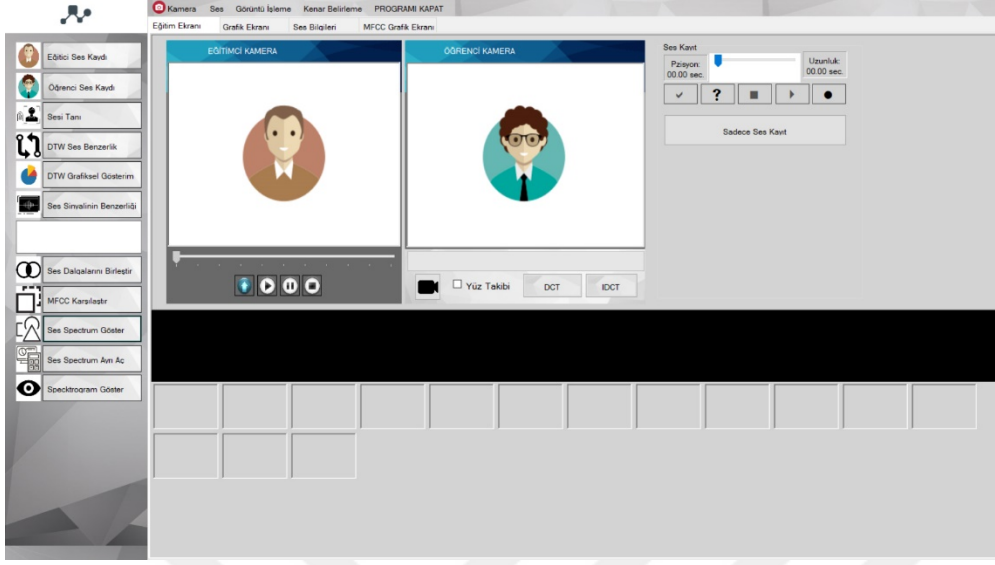
## 8.2 Konuşma Eğitimi

Konuşma eğitim setimizde 5 kişiden belirlediğimiz kelimelerin öğretimi için veri topladık. Elde ettiğimiz konuşma eğitim verileri kelimelerin söylenişinde gündelik konuşma da olduğu gibi söylenmesi istenmiştir. Kayıt işlemi Movavi Video Editor programıyla varsayılan çözünürlükte 30 fps çekim hızı ile Logitech HD Webcam C310 kamerası ile gerçekleştirilmiştir. Eğitim sırasında işaret diliyle de sembolik olarak desteklenmiştir. Şekil 8.2’de terapi eğitim için görüntülerin elde edilmesi işlemi vardır.



**Şekil 8.2 : Görsel konuşma eğitim kaydı**

Konuşma eğitim sistemi çalıştırıldıktan sonraki gelen kısım Şekil 8.3’de gösterildiği gibidir.



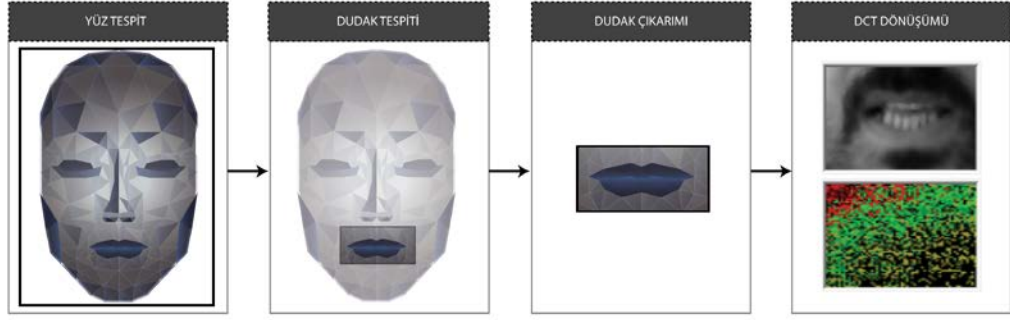
Şekil 8.3 : Konuşma eğitim uygulama görüntüsü

### 8.3 Görüntü Ön Uç

Görüntü işleme, gerçek bilgiyi, yani görüntüden yararlı bilgileri çıkarmak için herhangi bir görüntüyü işleyebilen bir tekniktir. Giriş videosunun görüntüleri üzerinde ön işleme Şekil 8.4 ve görsel özellik çıkarımı Şekil 8.5’de gösterilmiş ve alt bölümlerde açıklanmıştır.



Şekil 8.4 : Görsel Ön uç işleme adımları



**Şekil 8.5 :** Görsel özellik çıkarım işlemi

Sistemin görüntü ön ucu, ön işlem ve özellik çıkarımı olmak üzere iki aşamada gerçekleşmektedir.

Ön işlem, basamağı giriş görüntüsünün temizlenmesi, yüz lokasyon tespiti, ağız lokasyon tespiti ve dudak hareket takibi işlemlerini içermektedir. Bu aşama, görüntü işleme sonucunda elde edilecek sonuçların sağlığı açısından oldukça büyük bir önem taşımaktadır. Bir video, hızlı ardışık olarak gösterilen görüntü dizisidir. Kamera aracılığı ile elde edilen görüntüler öncelikle gürültü çıkarma işlemi, netlik ayarı, ışıklandırma gibi ayarlar ile giriş görüntüsü optimize edilir.

Görüntü temizleme işleminden sonra, dudak bölgesinin çıkarılması için yüz algılama işlemi gerçekleştirilir. Dudak görüntüleri, sürekli hareket eden çok küçük bir parçalar olduğundan dudakların aranması gereken alanı sınırlamak için, önce yüz algılamanın yapılması gerekmektedir. Yüz algılandıktan sonra daha fazla işlem için belirlenen görüntüden kırpma yapılır.

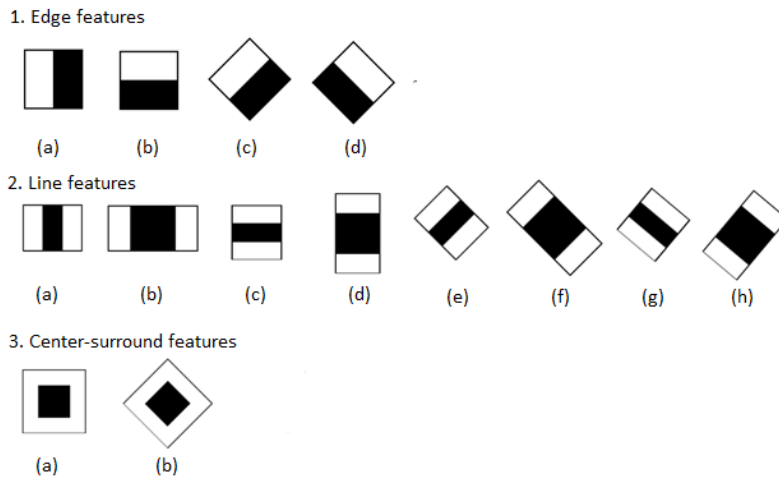
### 8.3.1 Ön işlem adımı

Görüntü işlemede, ön işlem adımı istenmeyen bozulmaları önlemek veya bazı görüntü özelliklerini geliştirmek için en düşük soyutlama seviyesindeki görüntü verilerinin iyileştirilmesi işlemidir. Giriş görüntüsünden daha iyi sonuçlar elde etmede yardımcı olduğu için önemli bir işlem adımıdır. Elde edilen görüntülerden gürültüyü çıkarma, doğruluk ve netliğin daha iyi olması için gereklidir. Işıklıdırma koşulları, görüntüdeki gürültü gibi olumsuzlukları ortadan kaldırarak en iyi sonuç göstermede yardımcı olacaktır. Görsel özellikler ağız bölgesinden çıkarılır, bir sonraki görsel özellikler çıkarımı için ağız bölgesinin hazırlanmasına ihtiyaç vardır. Uygulanan ön işlem süreçleri kısaca açıklanmıştır.

### 8.3.1.1 Yüz tespiti

Yüz tespiti, yüzle ilgili birçok uygulama (yüz tanıma, dudak okuma, yaş, cinsiyet ve ırk tanıma, duygu tahmini) için önemli bir ön işlem adımudur. Bu uygulamaların doğruluk oranı, yüz tespit işlem adımının güvenilirliğine bağlıdır. Yüz tespitinin amacı herhangi bir görüntüde yüz olup olmadığının belirlenerek varsa yüzün konum ve boyut bilgilerini geri döndürmektir. Genel olarak yüz algılama ve tespitinde güvenilirliği etkileyen faktörlerin başında pozlama, yönlendirme ve aydınlatma koşulları ile birlikte gelen parametrik farklılıklar vardır. Bunu çözmek için farklı algoritma ve yöntemler kullanılır. En başarılı yöntemlerden bazıları, tüm olası konumlar, ölçekler ve yönlendirmeler için görüntü üzerinde  $20 \times 20$  (veya daha fazla) piksel gözlem penceresi kullanmaktır. Bu yöntemlerden bazıları, destek vektör makineleri, yapay sinir ağlarıdır. Bazı araştırmacılar ten rengini renkli görüntülerdeki yüzü tespit etmek için kullanır.

Haar Cascade sınıflandırıcısında göre bulunması istenen nesnelere ilk önce tanıtılması gerekmektedir. Daha sonra ona benzer şekillerin bulunduğu resim ve video çerçeveleri taranarak nesne bulunmaya çalışılır. Sınıflandırıcı eğitiminde, Şekil 8.6'daki gibi çerçevelerle taranan resim, beyaz bölgedeki ve siyah bölgedeki piksel değerlerinin toplamının karanlık aydınlık değerleri kontrol edilerek belirli bir hedef değere oluşturulur.



Şekil 8.6 : Haar cascade sınıflandırıcı eğitimi

Performansını artırmak için zayıf sınıflandırıcıların cascade kullanan ikili bir sınıflandırıcı olan AdaBoost'u temel alır. Boosting algoritması birçok zayıf



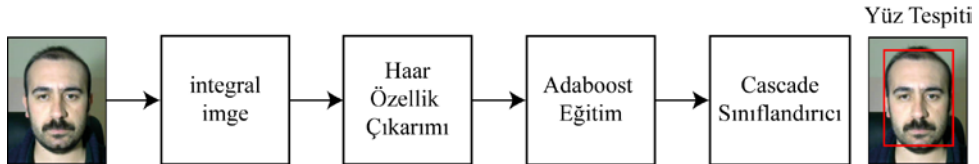
öğreniciyi (weak learner) bir araya getirerek daha güçlü bir öğrenici (strong learner) elde etmeye çalışır.

Ağız bölgesinin tespit işlemi için videodan her bir çerçevede yüz bölgesinin tespit edilmesi gereklidir. Viola ve Jones'un "Hızlı Nesne Tespiti" isimli makalesinde önerdiği yaklaşım etkili bir nesne algılama yöntemidir (Viola ve Jones, 2001). Yüz tespit işlemi için bu yöntem kullanılmış ve sistem, birçok olumlu ve olumsuz görüntü ile eğitilmiştir. Yüz bölgesinin doğru bir şekilde belirlenmesi ile ağız bölgesinin belirlenme işlemi kolaylaşır.

Viola ve Jones (2001) algoritmasında Şekil 8.7.'de gösterildiği gibi, 4 temel aşama bulunmaktadır.

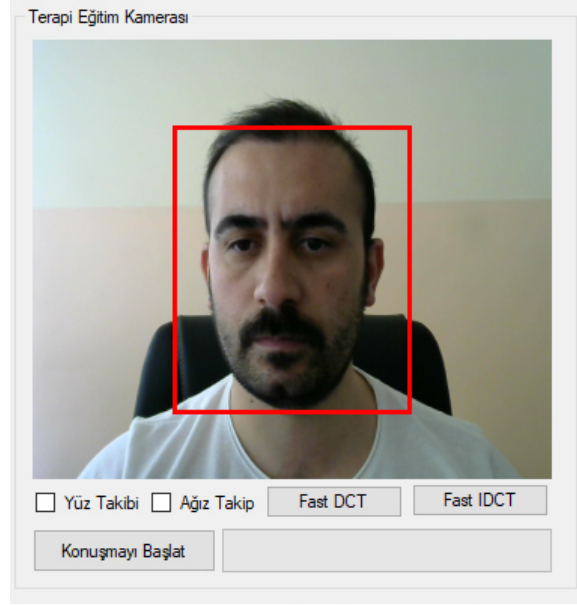
1. İntegral imge.
2. Haar özellik çıkarımı.
3. AdaBoost eğitimi.
4. Cascade sınıflandırıcı.

Bu yöntem, her bir pikselin incelenmesine gerek duyulmadan bir görüntünün analiz edilmesine olanak verir.



**Şekil 8.7 :** Viola ve Jones algoritması temel prensibi

Viola ve Jones algoritmasını kullandıktan sonra bir video görüntüsündeki yüz tespit işlemi Şekil 8.8'de gösterilmiştir.



**Şekil 8.8 :** Sistem yüz tespit görseli

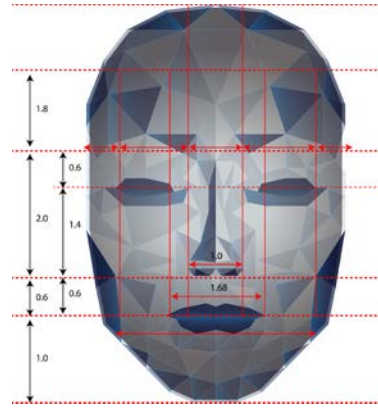
### 8.3.1.2 Ağız tespiti

Ağız bölgesini tespit etmek için, öncelikle yüz bölgesi tespit edilmelidir. Viola-Jones algoritması ile yüz bölgesi algılandıktan sonra ki işlem, yüz bölgesini oran bilgilerinden faydalanarak ağız bölgesinin çıkarılması işlemidir. Viola-jones algoritması yüz bölgesinin tespitinde başarılı sonuçlar versede Şekil 8.9’da görüldüğü gibi ağız bölgesi tespitinde iyi sonuçlar verememektedir.



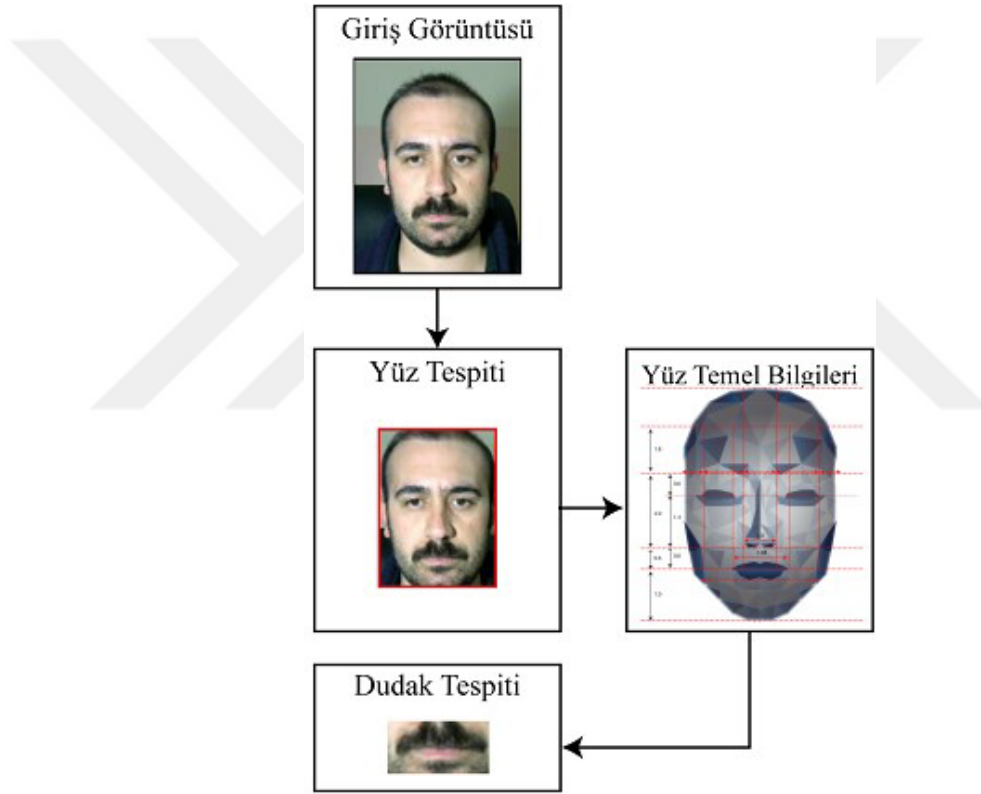
**Şekil 8.9 :** Viola-Jones algoritması ile ağız tespiti sonrası

Ağız bölgesi konuşma bilgilerinin görsel olarak çıkarıldığı bölge olduğu için daha da önemli bir işlem adıdır. Doğru görsel bilgilerin çıkarılması doğru bölgenin tespit edilmesine bağlıdır. Şekil 8.10’ da gösterildiği gibi yüz oran bilgileri dudak çıkarımı için matematiksel ve mantıksal bilgiler sunar.



**Şekil 8.10 : Yüz oran bilgileri**

Dudak bölgesinin çıkarımı, bu mantıksal bilgilere dayanılarak gerçekleştirilir.

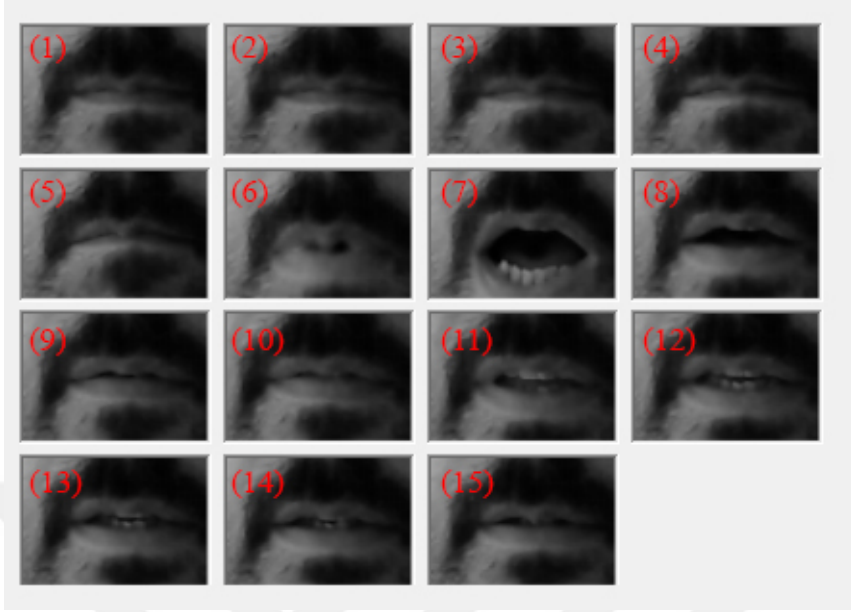


**Şekil 8.11 : Dudak tespiti**

Çalışmamızda dudak bölgesi tespiti için izlenen adımlar şunlardır:

1. Viola-jones algoritması ile yüz tespit işlemi gerçekleştirilir.
2. Yüz tespit işleminden sonra belirlenen alan kare içerisine alınır.
3. Kare içerisindeki alanın koordinat bilgileri elde edilir.
4. Koordinat bilgilerine göre yüz oran bilgisinden faydalanılarak ağız bölgesi çıkarma işlemi gerçekleştirilir.

Şekil 8.12’ de gösterildiği gibi ağız tespit işleminden sonra elde edilen görüntüler özellik çıkarım işlemi için hazır hale gelirler.



Şekil 8.12 : Konuşmada yakalanan dudak görüntüleri

### 8.3.1.3 Ağız bölgesinin yeniden boyutlandırılması

Ağız bölgesi, dikdörtgen n'nin bir tamsayı olduğu  $2^n$  şeklinde olacak şekilde yeniden boyutlandırılır. Bu işlem, DCT özelliklerinin hesaplanmasının giriş görüntüsündeki dudak konumundan etkilenmemesi için yapılır.

### 8.3.1.4 Rgb-gray dönüşümü

Giriş dudak bölgesi Şekil 8.13’de gösterildiği gibi RGB biçiminden 0 (Siyah) ile 255 (beyaz) aralığında gri formata dönüştürülür.



Şekil 8.13 : Dudak bölgesinin gri formata dönüşüm sonucu

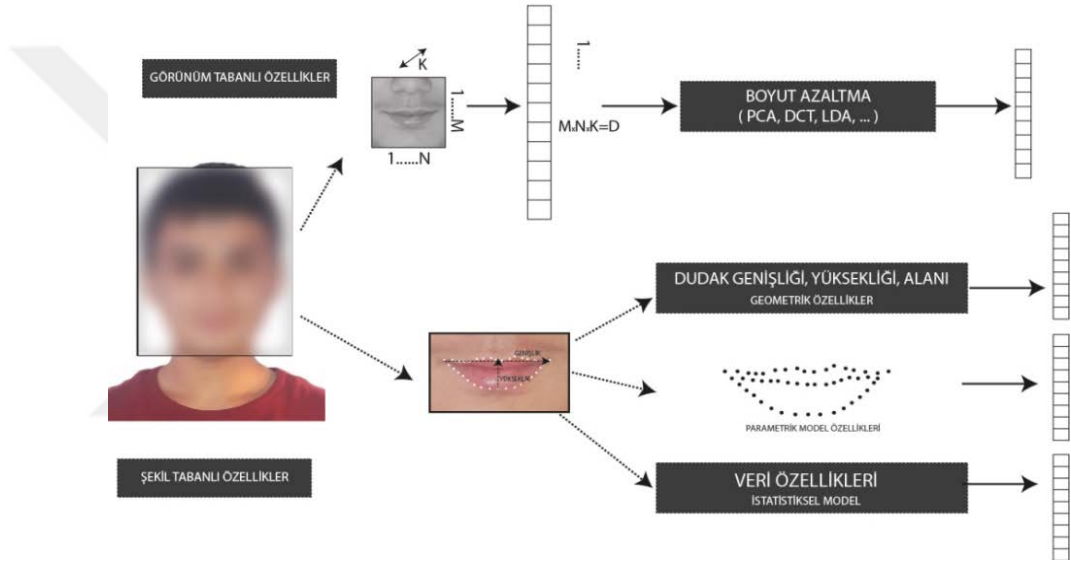
Viola-Jones algoritması kullanılarak yüz ve ağız tespitinin etkinliğini test etmek için, farklı kullanıcılardan 20 rastgele örnek görüntüsü seçildi ve yüz ve ağız algılama yüzdeleri kaydedildi. Yüz ve ağız bölgesi tespiti algılama oranı % 95’ e ulaşır.

### 8.3.2 Görsel özellik çıkarımı

Ağız bölgesi bulunduğunda; Kendi güçlü ve zayıf yönleri olan dudak kontur özelliği elde etmek için birçok algoritma kullanılabilir.

Özellik çıkarma için mevcut teknikler üç kategoriye ayrılabilir:

1. Bazı şablonlar tarafından tanımlanan bir yer işareti noktaları dizisi olarak gösterilebilen görsel şekil.
2. Özelliklerin doğrudan piksellerden elde edildiği görünüm tabanlı görsel özellikler.
3. Hem (1) hem de (2) 'nin kombinasyonu olan hibrit yaklaşımlar.



**Şekil 8.14 :** Görünüm ve şekil tabanlı görsel özellik çıkarma işlemi

Görünüm tabanlı görsel özellikler: Görünüme dayalı görsel özellikler, Ayrık Kosinüs Dönüşümü (DCT), Ayrık Dalgacık Dönüşümü (DWT) ve Temel Bileşen Analizi (PCA) olarak görüntü dönüştürme temelli özellik çıkarma yöntemlerini içerir. Bu algoritmalar genellikle boyut indirmek için uygulanarak gereksiz verileri ortadan kaldırır. DCT' ler, bilim ve mühendislik alanında, ses ve görüntülerin kayıplı sıkıştırılmasından (küçük yüksek frekanslı bileşenlerin atılabildiği yerlerde) kısmi diferansiyel denklemlerin sayısal çözümü için spektral yöntemlere kadar çok sayıda uygulama için önemlidir.

Görsel özelliklerin çıkarılması DCT ile gerçekleştirilir. DCT, Ayrık Fourier dönüşümü (DFT) ile benzerdir. Görüntüyü uzaysal alandan frekans alanına dönüştürme işlemidir.

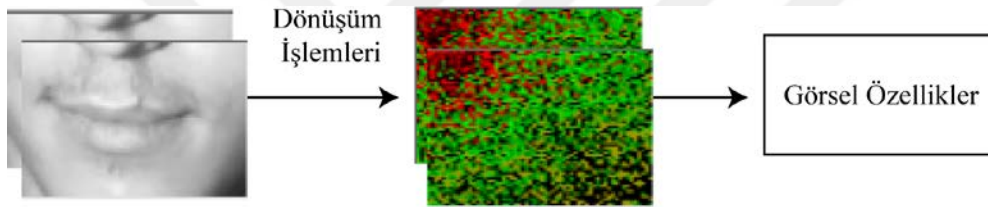
$M * N$  boyutunda  $F(u, v)$  fonksiyonunun 2D DCT'si aşağıdaki denklem ile tanımlanır:

$$F(u, v) = a_u a_v \sum_{x=0}^{M-1} \sum_{y=0}^{N-1} f(x, y) \cos \frac{(2x+1)u\pi}{2N} \cdot \cos \frac{(2y+1)v\pi}{2N} \quad (8.1)$$

$F(x, y)$ ,  $x$  satırındaki  $y$  sütunundaki pikselin yoğunluğudur.

$$a_u, a_v = \begin{cases} \sqrt{\frac{1}{2}}, & u, v = 0 \\ 1, & u, v = 1, \dots, N \end{cases} \quad (8.2)$$

Şekil 8.15'da gösterildiği gibi DCT, giriş ağız bölgesi görüntüsünün boyutuna sahip bir özellik matrisi üreterek dudak özellikleri olarak kaydedilir.



**Şekil 8.15 :** 2D-DCT Görsel özelliklerin çıkarılma işlemi

Görünüm tabanlı, şekil ya da model tabanlı özellik çıkarımları da bulunmaktadır. Model tabanlı özellikler kullanıcının dudaklarının genişlik ve yüksekliği ile ilgilidir. Bu tip özellik çıkarımlarında bilgi kaybı söz konusudur, çünkü tüm bölge için değil dudaklar hakkındaki bazı bilgilere dayanır (Potamianos ve diğ., 2003). Görünüm tabanlı özellik çıkarım yöntemlerinde, tüm ağız bölgesi piksellerinin konuşma tanıma konusunda bilgilendirici olduğunu varsayar.

Konuşma yakalama işlemi ile elde edilen çerçevelerdeki görüntüler DCT işlemini gerçekleştirmek için Şekil 8.16'da gösterildiği gibi metoda gönderilir.

```

private void btnFastDct_Click(object sender, EventArgs e)
{
    ortalamaDegerler = new List<double>();

    for (int i = 0; i < totaldeckCount; i++)
    {
        bmp = imgList[i].ToBitmap();
        ImgDCT = new FastDCT2D(bmp, WindowSize);
        ImgDCT.FastDCT();
        pbName[i].Image = (Image)ImgDCT.DCTMap;
        Bitmap p = ImgDCT.DCTMap;
        MemoryStream ms = new MemoryStream();
        p.Save(ms, System.Drawing.Imaging.ImageFormat.Jpeg);
        byte[] imageconvert = ms.ToArray();
    }
}

```

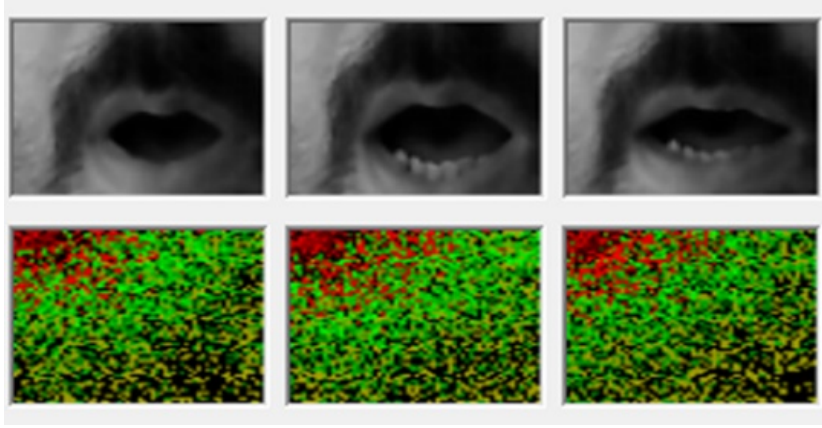
**Şekil 8.16** : Görüntünü DCT metoduna gönderimi

Görünüme dayalı özelliklerin örnekleri DCT, Ayrık Dalgacık Dönüşümü (DWT) ve Doğrusal Diskriminant Analizi (LDA). Görsel-işitsel konuşma tanımda DCT' nin iyi performansından dolayı genel olarak kullanılmaktadır (Estellers ve Thiran, 2012). Bu çalışmada da DCT uygulanmış ve konuşmacının ağız bölgesinden çıkarılmıştır.

Daha sonra, son görsel özellik vektörünü hesaplamak için üç seçim yöntemi uygulanır (Salama ve diğ., 2014).

- DCT matrisinden toplam maksimum değerleri hesaplar ve 1. sınıf 2. türevlerini ekler.
- DCT matrisinin sol üst köşe bölgesini alır. Bu bölgenin boyutunun 3 olarak ayarlandığını, bu nedenle bölgenin 3 \* 3 boyutunda olacağını ve son özellik vektörünün 9 öge olmaktadır.
- İkinci yöntemle aynıdır ancak sol üst köşe bölgesinin tamamı dışında yalnızca maksimum özelliklerde çalışır.



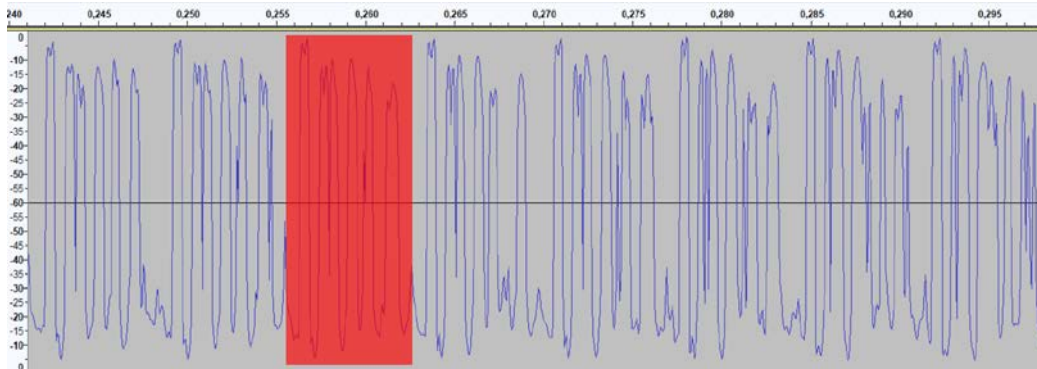


**Şekil 8.17 :** Mavi kelimesinin seçili çerçevelerdeki DCT çıkarım sonucu

#### 8.4 Ses Ön Uç

Sesler, gerçek dünyadaki zamana göre değişen sinyallerdir. Anlamalarının tümü bu zaman değişkenliği ile ilgilidir. Bu nedenle, anlama, karşılaştırma, değiştirme ve yeniden sentez işlemlerini kolaylaştırmak için zamanla değişen seslerin ayırt edici özelliklerinden faydalanmak için ses analizi işlemleri gerçekleştirilir (Rocchesso, 2003). İlgili özelliklerden faydalanmak için yapılan ses işleme işlemi ön işlem ve özellik çıkarımı olmak üzere iki aşamadan meydana gelmektedir.

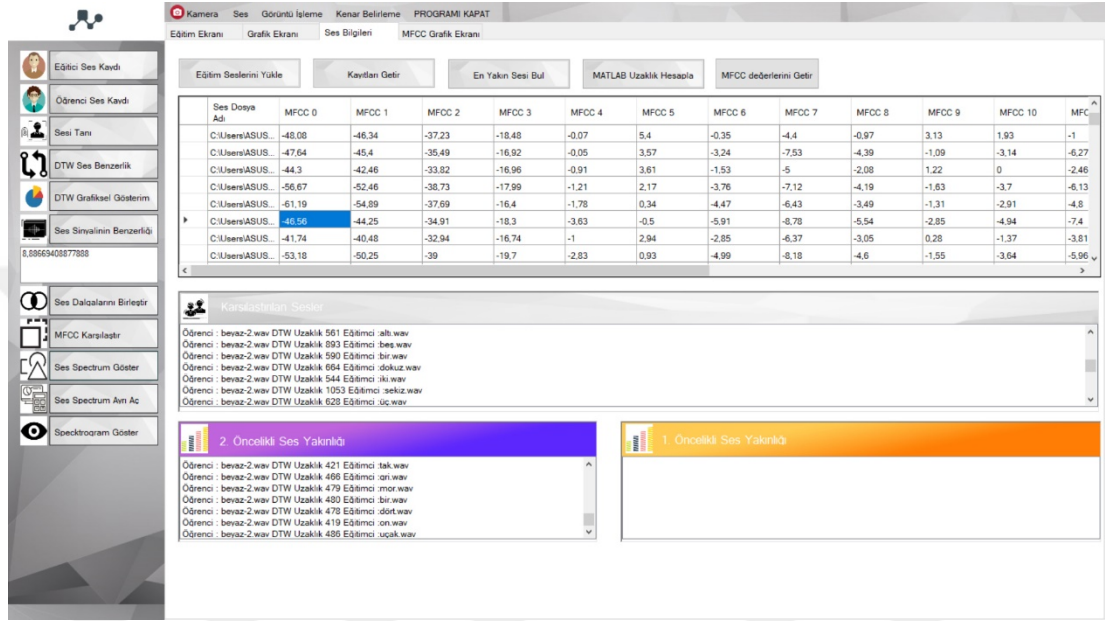
Ön işleme, konuşma algılayıcının ilk kısmıdır ve özellik çıkarma işleminden önce hazırlık seviyesi olarak yorumlanabilir. Optimum konuşma tanımayı elde etmek için yararlı konuşma bilgisi, ön işleme bölümünde uygun bir biçimde çıkarılmalı ve temsil edilmelidir. Ham sesin ön düzenlenmesi ve işlenmesi önemli bir işlem adıdır. Tipik olarak temiz bir ses dosyası için çeşitli istenmeyen öksürük, hıçkırık, rahatsız edici tepe noktaları gibi bölümlerin kaldırılarak sesin normal hale getirilmesi sağlanır.



**Şekil 8.18 :** Ses sinyali dalga formunun kararlı yapısı

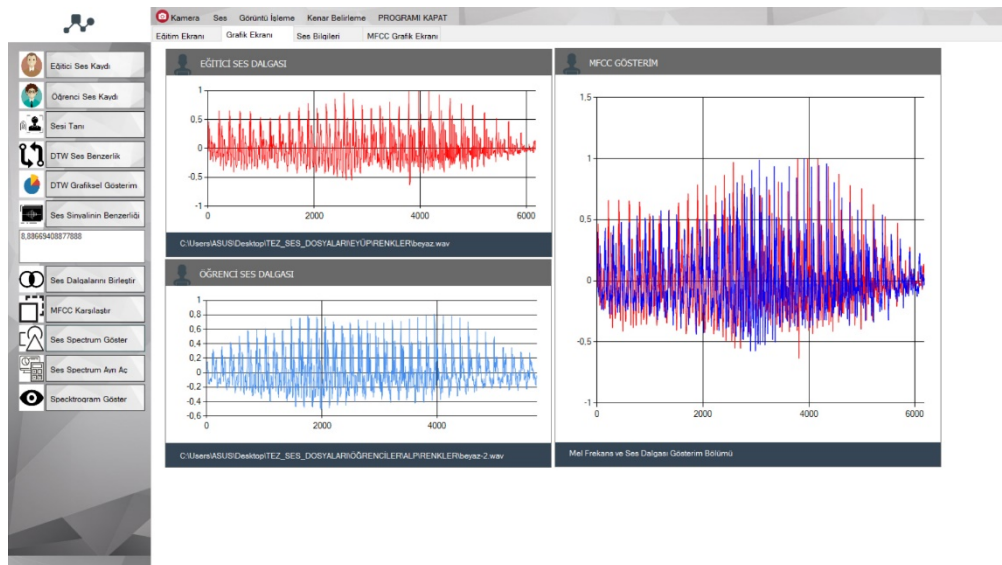


Akustik özellikleri elde etmeden önce konuşma bilgisinde yapılması gereken işlemler vardır. Ses sinyalleri belli zaman aralıklarında Şekil 8.18’de gösterildiği gibi kararlı yapıdadır. Bu kararlı yapıdaki bilgilerden gerekli özellikler çıkarmak için konuşma sinyali çerçeveleme ve segmentasyon işlemi ile küçük parçalara bölünür. Şekil 8.19’ da gösterildiği gibi eğitimcilerden elde ettiğimiz ses özelliklerinin öğrencinin konuşma sesi özellikleri ile karşılaştırma işlemi gösterilmektedir.



Şekil 8.19 : Ses özelliklerinin çıkarımı

Çıkarılan özellikler konuşma tanıma ve grafiksel olarak yorumlama için Şekil 8.20’de gösterildiği gibi sistemde kullanılır.



Şekil 8.20 : Sistemin görsel gösterimi

Ses sinyalindeki pencereleme işlemini gerçekleştirdiğimiz zaman her karenin temel analiz işlemi ile en belirgin akustik özellikler ses yoğunluğu (kuvveti), ses perdesi ya da basamağı, ses sinyalinin anlamlı içeriği araştırılır. Bu içerikler ses sinyallerinin dalga formuyla ilişkili olabilmektedir. İnsan seslerinin özellikleri fiziksel büyüklüklerle ilişkilendirilir.

Çalışmamızda ses özelliklerinin çıkarılması için temel yaklaşım;

- Ses sinyalleri bir dizi çerçeve bloklarına dönüştürülür. Her karenin zaman süresi yaklaşık 20 ~30 ms'dir. Bu kare yakalama süresinin büyük yada küçük olması ses sinyallerinin zamanla değişen özelliklerinin yakalanması adına önemlidir. Çerçeve sürelerinin küçük olması geçerli bir akustik özellik çıkarımı için yeterli olmaz. Çerçeve boyutları 2'nin katları (256, 512, 1024) şeklinde alınır böylece FFT dönüşümleri için uygun hale getirilir.
- Birbirine komşu çerçeveler arasındaki farkı azaltmak için birbiri üzerine bindirilmesi gerekmektedir. Genellikle örtüşme orijinal çerçevenin 1/2 den 2/3'ü kadardır. Örtüşme oranı arttıkça daha fazla hesaplama işlemi gerekir.
- Çerçevelerdeki ses sinyallerinin sabit olduğu bilgisine dayanılarak sıfır geçişleri, ses seviyesi, aralık, MFCC gibi akustik özellikleri çıkarabiliriz.

Ses özellikleri çıkarımında sıklıkla kullanılan temel terminolojiler;

- Çerçeve boyutu, her çerçevenin içindeki örnekleme noktalarıdır.
- Çerçeve örtüşmesi, ardışık çerçeveler arasındaki örtüşme örnekleme noktalarıdır.
- Çerçeve adımı, eksi bindirmenin çerçeve boyutuna eşitidir.
- Kare hızı, saniyedeki kare sayısıdır. Kare adımına bölünen örnek frekansına eşitidir.

Örnek olarak elimizde 16000 frekans değerine sahip ve 25 ms kare süresi olan ve 15 ms bir örtüşme süresine sahip bir ses sinyalin olduğunu düşünürsek;

Çerçeve boyutu;

$$f_s * \frac{25}{1000} = 400 \text{ (örnek nokta)} \quad (8.3)$$

Çerçeve örtüşmesi;

$$f_s * \frac{15}{1000} = 240 \text{ (örnek nokta)} \quad (8.4)$$

Çerçeve adımı;

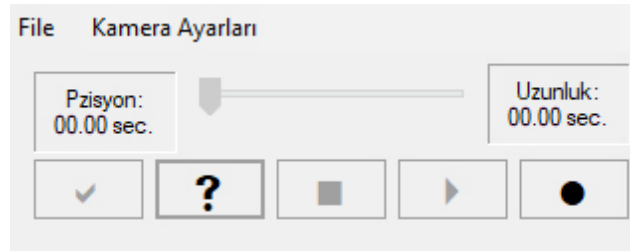
$$400 - 240 = 160 \text{ (örnek nokta)} \quad (8.5)$$

Çerçeve hızı;

$$\frac{f_s}{160} = 100 \text{ çerçeve/sn} \quad (8.6)$$

#### 8.4.1 Ses Kaydı

İşitme engellilerin telaffuzlarının eğitici eşliğinde karşılaştırılarak benzerlik oranına göre tekrar edilmesi ve bunların analizi için kaydedilmesi gerçekleştirilmiştir. İşitme engelli bireylerin ses bilgi girişi Şekil 9.21'deki tasarım ekranından sağlanmaktadır. Aynı şekilde eğitimcide aynı ekrandan ses kayıt işlemini gerçekleştirmektedir. Ses kayıt sırasında dışarıdan gelen sesleri ve parazitleri en aza indirmek için snopy marka harici mikrofon kullanılmıştır. Çünkü ses özelliklerinin çıkarımında olabildiğince temiz bir ses kaydı yapılması gerekmektedir. Aksi takdirde özellik çıkarımında istemediğimiz gürültüler ses bilgisine eklenebilir.



Şekil 8.21 : Konuşma kaydı

Eğitim ve test için söylenen kelimeleri 22.050 Hz hızında ve 16-bit olarak kaydetme işlemini gerçekleştirdik.

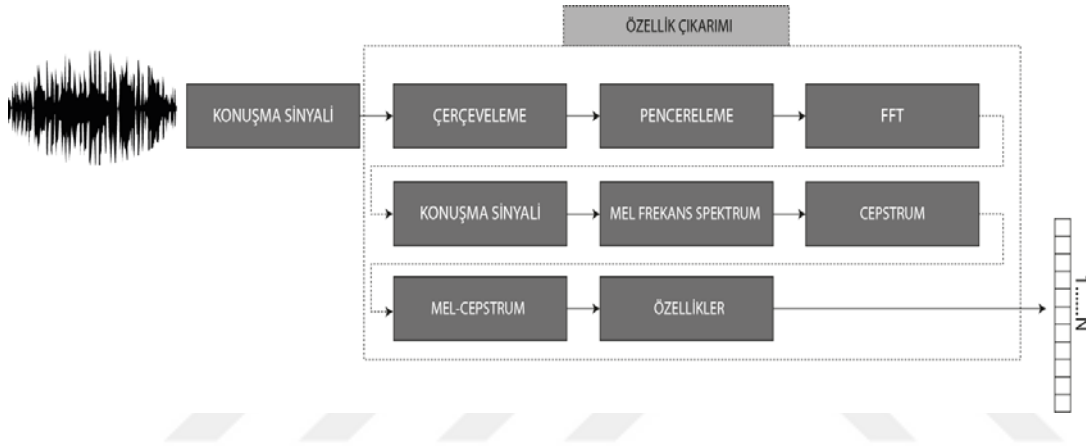
#### 8.4.2 Ses özellik çıkarımı

Herhangi bir konuşma tanıma sisteminde ilk adım özellik çıkarım işlemidir. Mel frekanslı kestral katsayılar (MFCC) zaman alanındaki sesin frekans alanına dönüştürülme işlemini gerçekleştirerek konuşma sinyalinde bulunan bilgilere ulaşılabilmesini gerçekleştirir. MFCC, frekanslara göre insan algı hassasiyetini

dikkate alır ve bu nedenle konuşma tanıma için kullanılan en iyi tekniklerden birisidir.

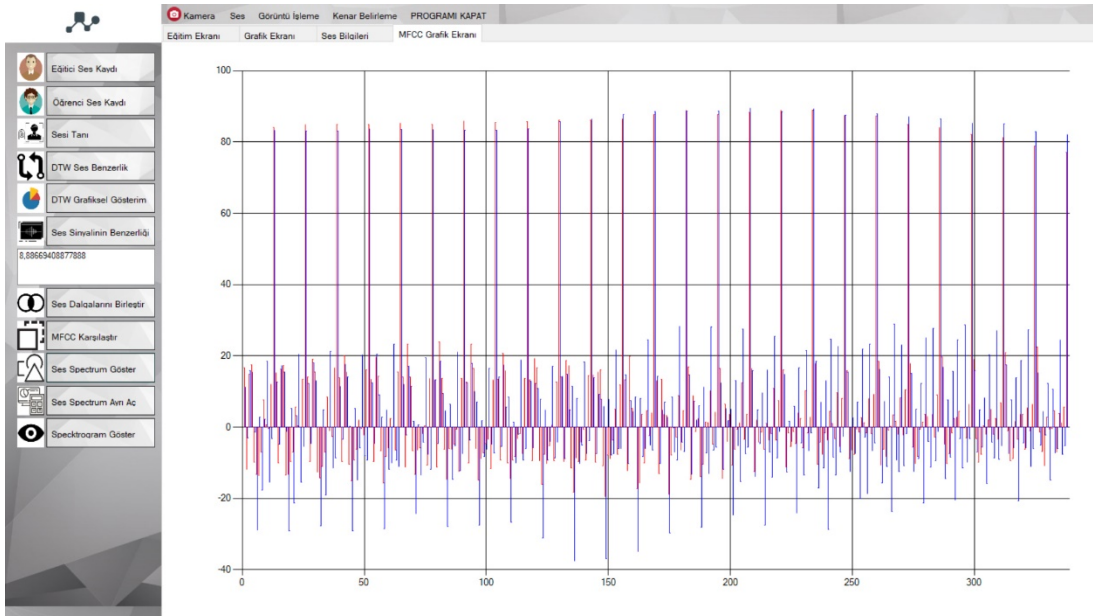
Özellik çıkarımı işleminin temel amacı konuşma sinyalindeki her sözcük için benzersiz, ayırt edici, sağlam, verimli özellik bilgilerinin çıkarımıdır. Konuşma şekli her konuşmacı için konuşma sinyalinin hızına ve duygusal durumuna bağlı olarak değişebilmektedir.

Önerilen yaklaşımımızda, akustik özellikleri, temel ve türetilmiş özellikler de dahil MFCC özelliklerini kullanarak çıkartıyoruz. Şekil 8.22, MFCC'nin hesaplanmasında ana adımları gösterir.



Şekil 8.22 : Ses bilgisi özellik çıkarım modeli

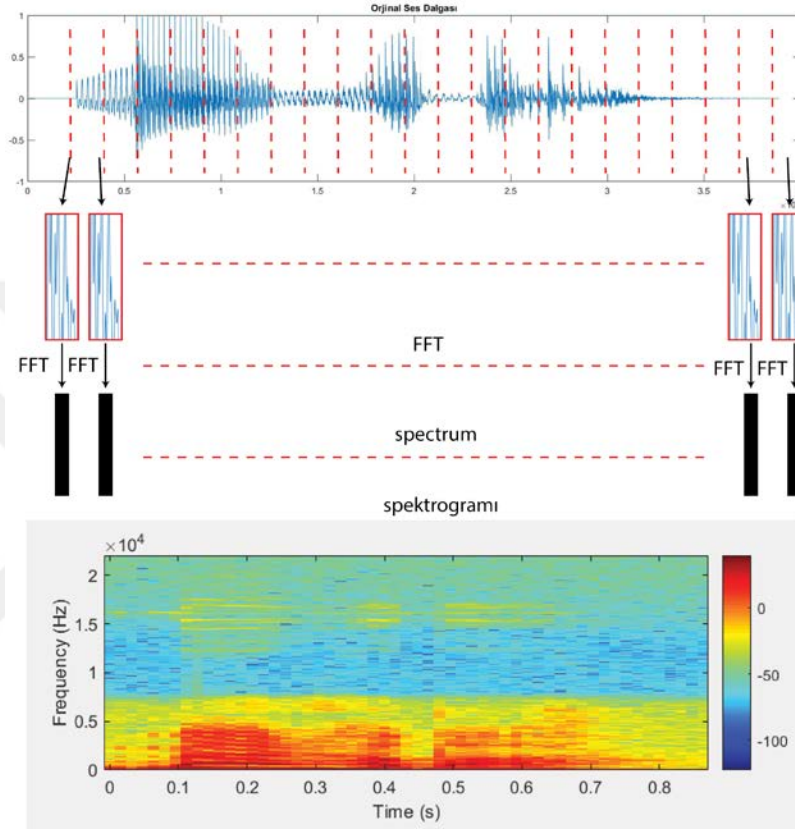
Kaydedilen kelimelerin HMM modelini oluşturmada ve ses işlem adımları için MFCC özellik çıkarma yöntemi olarak belirledik. Önceden kaydedilmiş bir ses dosyasının MFCC değerleri için grafiksel gösterimi Şekil 8.23'de gösterilmiştir.



Şekil 8.23 : İki ses sinyalinin MFCC değerlerinin grafiksel gösterimi

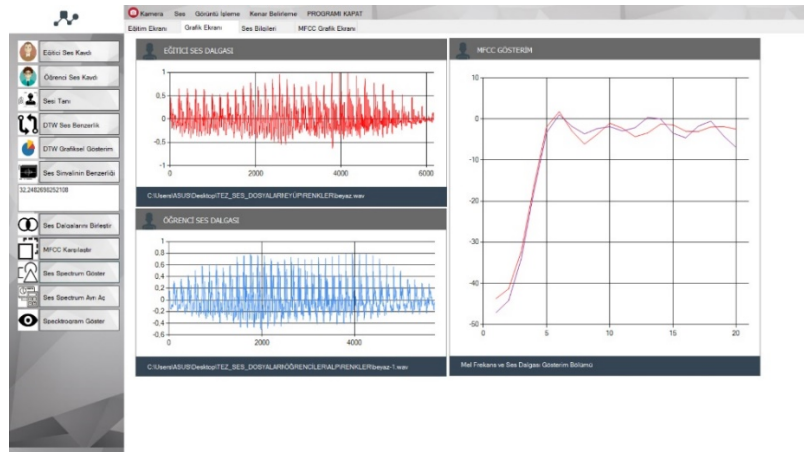
### 8.4.3 Ses kayıtlarının grafiksel gösterimi

Eğitim sırasında ya da önceden kaydedilen sesleri, ses dalgası ve ses spektrumu olarak Şekil 8.24'deki gibi göstermek kişinin görsel olarak sesinin neye benzediğini görmesi açısından faydalı olmaktadır.



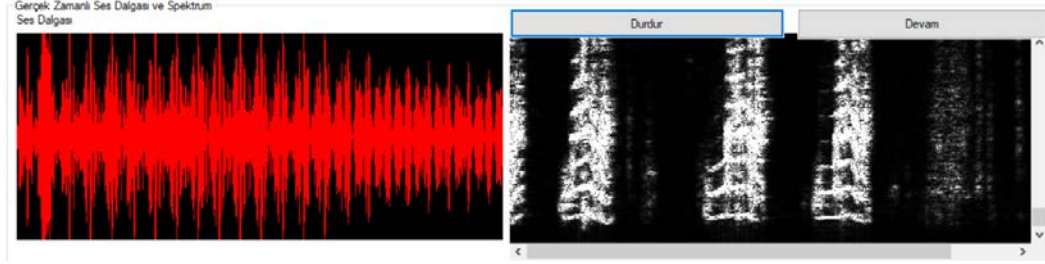
Şekil 8.24 : Sesin grafiksel gösterimi

Çalışmamız esnasında gerçek zamanlı olarak öğrenci ve eğitmenin ses bilgilerini Şekil 8.25'deki gibi görsel gösterim işlemini gerçekleştirdik.



## Şekil 8.25 : Sesin grafiksel gösterimi

Kişinin konuşma sırasındaki sesinin grafiksel gösterimi, özellikle işitme engelli bireylerde ses kalitesini ve söyleyiş şeklini geliştirip değiştirebilmesi adına önemli bir özelliktir. Gerçek zamanlı, kişinin konuşmasının ses dalgası ve spektrum gösterimi şekil 8.26'daki gibidir.



Şekil 8.26 : Gerçek zamanlı konuşmanın ses dalgası ve spektrum gösterimi

### 8.4.4 Sistemin eğitilmesi

Konuşma tanıma sisteminde izole edilmiş kelimelerin sisteme eğitilmesi gerekmektedir. İki matlab dosyamız, train\_word ve train\_all\_words bulunmaktadır. Train\_word dosyası ile parametre olarak verdiğimiz wav dosyasını işleyerek HMM model parametreleri üretiriz. Şekil 8.27'de Train\_all\_words dosyası ile train\_word fonksiyonunu kullanarak sisteme tanıtılacak tüm sözcükleri işleminden geçirmektedir.

```
Editor - C:\Users\ASUS\Desktop\konuşmatanım\train_all_words.m
train_word.m  train_all_words.m  dosya_okuma.m  +
1 function train_all_words
2     warning('off','MATLAB:dispatcher:InexactCaseMatch');
3     numStates = 7;
4
5     % EĞİTİM KELİMELERİ - AİLE
6     train_word('ABİ', numStates);
7     train_word('AMCA', numStates);
8     train_word('ANNE', numStates);
9     train_word('ANNEANNE', numStates);
10    train_word('BABAANNE', numStates);
11    train_word('BABA', numStates);
12    train_word('DEDE', numStates);
13    train_word('HALA', numStates);
14    train_word('TEYZE', numStates);
15
16    % EĞİTİM KELİMELERİ - GÜNLÜK KELİMELER
17    train_word('BAK', numStates);
18    train_word('DÖN', numStates);
19    train_word('GEL', numStates);
20    train_word('GİT', numStates);
21    train_word('KAL', numStates);
22    train_word('SELAM', numStates);
23    train_word('TAK', numStates);
24
25    % EĞİTİM KELİMELERİ - RENKLER
26    train_word('BEYAZ', numStates);
27    train_word('GRİ', numStates);
28    train_word('KIRMIZI', numStates);
29    train_word('MAVİ', numStates);
30    train_word('MOR', numStates);
31    train_word('PEMBE', numStates);
32    train_word('SİYAH', numStates);
33
Command Window
>> train_all_words
Eğitim Kelimesi: ABİ
Eğitim Kelimesi: AMCA
Eğitim Kelimesi: ANNE
Eğitim Kelimesi: ANNEANNE
Eğitim Kelimesi: BABAANNE
Eğitim Kelimesi: BABA
Eğitim Kelimesi: DEDE
Eğitim Kelimesi: HALA
Eğitim Kelimesi: TEYZE
Eğitim Kelimesi: BAK
Eğitim Kelimesi: DÖN
Eğitim Kelimesi: GEL
Eğitim Kelimesi: GİT
Eğitim Kelimesi: KAL
Eğitim Kelimesi: SELAM
Eğitim Kelimesi: TAK
Eğitim Kelimesi: BEYAZ
Eğitim Kelimesi: GRİ
Eğitim Kelimesi: KIRMIZI
Eğitim Kelimesi: MAVİ
Eğitim Kelimesi: MOR
Eğitim Kelimesi: PEMBE
Eğitim Kelimesi: SİYAH
Eğitim Kelimesi: TURUNCU
Eğitim Kelimesi: BİR
Eğitim Kelimesi: İKİ
Eğitim Kelimesi: ÜÇ
Eğitim Kelimesi: DÖRT
Eğitim Kelimesi: BEŞ
Eğitim Kelimesi: ALTI
Eğitim Kelimesi: YEDİ
Eğitim Kelimesi: SEKİZ
Eğitim Kelimesi: DOKUZ
```

Şekil 8.27 : Sisteme tanıtılacak örnek kelimeler



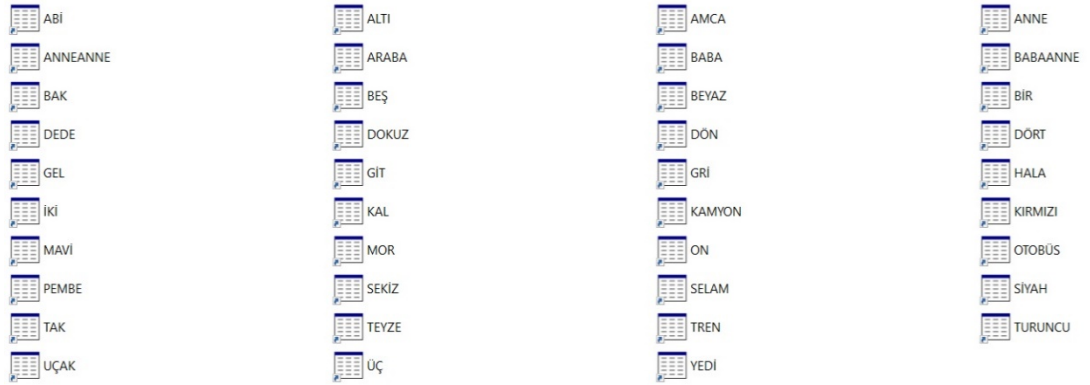
```

Editor - C:\Users\ASUS\Desktop\konuşmatanımı\train_word.m*
train_word.m* x
1 function train_word( wordName, numStates )
2 disp(['Training word: ' wordName]);
3 wav_file_name = ['training_words/' wordName '.wav'];
4 [y, fs] = wavread(wav_file_name);
5 observationVector = melcepst(y, fs, '', 22, floor(3*log(fs)), 128, 32);
6 DIAG_COV = 1;
7 QUIET = 1;
8 N = numStates;
9 A = sparse(0.85*diag(ones(1,N))+0.15*diag(ones(1,N-1),1));
10 A(N,N) = 1;
11 NIT = 10;
12 p = size(observationVector, 2);
13 X = [];
14 st = [];
15 st = [st; size(X,1)+1];
16 X = [X; observationVector];
17 T = size(X,1);
18 [my_mu,Sigma] = hmm_init(X, st, N, DIAG_COV,QUIET);
19 Sigma = ones(N,1)*mean(Sigma);
20 logl = zeros(1, NIT);
21 for n = 1:NIT
22 [tmp, logl(n), gamma] = hmm_mest(X, st, A, my_mu, Sigma, QUIET);
23 [my_mu, Sigma] = mix_par(X, gamma, DIAG_COV, QUIET);
24 Sigma = ones(N,1)*(sum(sum(gamma).*ones(1,p)).*Sigma)/T;
25 end
26 my_sigma = Sigma(1,:);
27 eval(['save hmm_model_files/' wordName ' my_mu my_sigma']);
28 end

```

**Şekil 8.28 : Kelimeleri sisteme tanıtmak**

Train\_word fonksiyonu ile bir kelimeyi sisteme tanıtarak HMM model parametrelerini oluşturmak istediğimiz wav dosyasından okutarak voicebox toolbox'ında yer alan MELCEPST fonksiyonu ile şekil 8.29'daki gibi kelimelerin özellik vektörlerini oluşturduk.



**Şekil 8.29 : Özellik vektörleri**

Eğitim modelleri oluşturduktan sonra sisteme yeni bir kelime vererek daha önceden eğitim esnasında kullandığımız kelimeler arasında en benzerini bulmasını istedik.

```

1 function test_word( wordName, numStates )
2 %for i=1:160
3 %   trainedWords{i}=num2str(i);
4 %end
5 trainedWords = {'BİR'; 'İKİ'; 'ÜÇ'; 'DÖRT'; 'BEŞ'; 'ALTI'; 'YEDİ'; 'SEKİZ'; 'DOKUZ'; 'O
6 'BEYAZ'; 'SİYAH'; 'MAVİ'; 'MOR'; 'KIRMIZI'; 'PEMBE'; 'GRİ'; 'TURUNCU'
7 'ANNE'; 'BABA'; 'DEDE'; 'ABİ'; 'AMCA'; 'TEYZE'; 'HALA'; 'ANNEANNE'; 'B
8 'SELAM'; 'GEL'; 'GİT'; 'KAL'; 'DÖN'; 'BAK'; 'TAK';
9 'ARABA'; 'TREN'; 'KAMYON'; 'OTOBÜS'; 'UÇAK'
10 };
11 wav_file_name = ['testing_words/' wordName '.wav'];
12 [y, fs] = wavread(wav_file_name);
13 observationVector = melcepst(y, fs, '', 22, floor(3*log(fs)), 128, 32);
14 N = numStates;
15 A = sparse(0.85*diag(ones(1,N))+0.15*diag(ones(1,N-1),1));
16 A(N,N) = 1;
17 numIter = 10;
18 numberOfFiles = length(trainedWords);
19 for w=1:numberOfFiles
20   fileName = trainedWords{w};
21   modelFileName = ['hmm_model_files/' fileName '.mat'];
22   load(modelFileName);
23   Sigma = ones(N,1) * my_sigma;
24   score(w) = hmm_vit(observationVector, A, [1 zeros(1,N-1)], my_mu, Sigma, 1
25   disp("Dosya İsmi:" + fileName + " Skor : " + score(w));
26 end % for w=3:num...
27 result = find(score >= max(score));
28 disp(['Tanınan Kelime : ' trainedWords{result}]);
29 end

```

```

>> test_word('ABİ',7)
Dosya İsmi:BİR Skor :-60324.5771
Dosya İsmi:İKİ Skor :-47862.2892
Dosya İsmi:ÜÇ Skor :-58096.0206
Dosya İsmi:DÖRT Skor :-45371.489
Dosya İsmi:BEŞ Skor :-52889.6082
Dosya İsmi:ALTI Skor :-38170.3779
Dosya İsmi:YEDİ Skor :-49176.2411
Dosya İsmi:SEKİZ Skor :-41685.5429
Dosya İsmi:DOKUZ Skor :-44990.533
Dosya İsmi:ON Skor :-52163.3685
Dosya İsmi:BEYAZ Skor :-43217.8527
Dosya İsmi:SİYAH Skor :-38852.0028
Dosya İsmi:MAVİ Skor :-35855.4965
<missing>
Dosya İsmi:KIRMIZI Skor :-43743.7572
Dosya İsmi:PEMBE Skor :-42690.4528
Dosya İsmi:GRİ Skor :-42825.1832
Dosya İsmi:TURUNCU Skor :-40416.6821
Dosya İsmi:ANNE Skor :-44366.6623
Dosya İsmi:BABA Skor :-45343.888
Dosya İsmi:DEDE Skor :-45923.1703
Dosya İsmi:ABİ Skor :-21494.023
Dosya İsmi:AMCA Skor :-36440.0963
Dosya İsmi:TEYZE Skor :-42365.7031
Dosya İsmi:HALA Skor :-37302.5249
Dosya İsmi:ANNEANNE Skor :-44031.5331
Dosya İsmi:BABAANNE Skor :-87912.7315
Dosya İsmi:SELAM Skor :-43240.8493
Dosya İsmi:GEL Skor :-44553.6843
Dosya İsmi:GİT Skor :-43649.7347
Dosya İsmi:KAL Skor :-45175.2085
Dosya İsmi:DÖN Skor :-43544.8861

```

Şekil 8.30 : Sistemin test dosyası

Matlab programında konuşma tanıma sistemi için kullandığımız fonksiyonlar Visual Studio ortamında kendi sistemimizle bağlantıyı şekil 8.31’de göstermiş olduğumuz yöntemle gerçekleştirdik.

```

1 reference
private void btnSesTani_Click(object sender, EventArgs e)
{
    MLab.MLab matlab = new MLab.MLab();
    string bilgi = matlab.Execute(@"cd C:\Users\ASUS\Desktop\konuşmatanıma");
    MessageBox.Show(bilgi);
    MessageBox.Show(matlab.Execute(@"train_all_words"));
    string kelime = path;
    MessageBox.Show(matlab.Execute(@"test_word('" + kelime + "', 5)"));
}

```

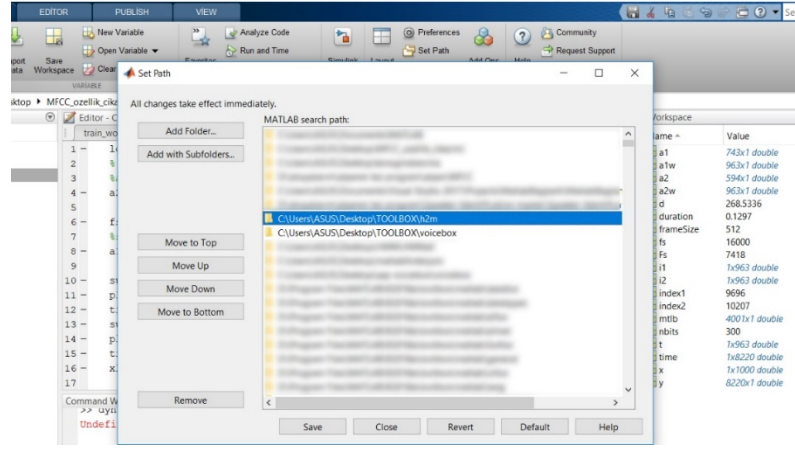
Şekil 8.31 : Matlab bağlantı kodları

### 8.4.5 Ses Tanımlaması

Konuşulan kelimenin analiz işleminin gerçekleştirilmesi ve tanınması farklı işlem adımlarına sahiptir. Konuşma tanıma da ses özelliklerinin çıkarımıyla birlikte örüntü tanıma ve sınıflandırma gibi işlem basamakları bulunmaktadır. Konuşma tanıma için Matlab R2018a programı kullanılmıştır. Matlab teknik ve mühendislik işlemleri için yüksek performanslı bir dildir. Matlab ile birlikte konuşma işlemede kullanılan birçok yararlı metoda sahip, özellikle MFCC hesaplamalar için MELCEPST kütüphanesine sahip Voicebox ile birlikte ve HMM toolbox kullanarak konuşma



tanıma işlemini gerçekleştirdik. Şekil 8.32’de Matlaba toolbox ekleme işlemi gösterilmektedir.



Şekil 8.32 : Matlab path toolbox ekleme

#### 8.4.6 Seslerin karşılaştırılması

Eğitici ile birlikte ses çalışmasının yapılmasından sonra gerçekleştirilecek ses benzerlik özelliklerinin gösterimi için ilk olarak eğitici ve öğrenci ses kayıt bilgilerinin yüklenmesi gerekmektedir. Önceden kaydedilmiş wav formatındaki ses dosyaları Şekil 8.33’deki butonlarla sisteme yüklenerek gerekli analiz işlemleri gerçekleştirilir.



Şekil 8.33 : Ses bilgilerinin sisteme yüklenmesi

Ses dosyalarının yüklenmesinden sonra ses bilgisi bilgisayar tarafından işlem yapılacak şekilde sayısal formata dönüştürülür. Çeşitli aygıtlar ile yapılan ses kayıtlarında sıklıkla gürültü bilgileri de yer almaktadır. Konuşma sinyali kalitesi ve anlaşılabilirliği gürültü ile karıştığı zaman özellik çıkarım işlemi orijinal bilgiye gereksiz gürültü bilgilerinin eklenmesine neden olmaktadır. Bu sebeple sinyaldeki orijinal bilgiyi çıkarmak için konuşma sinyalinde istenmeye gürültüyü azaltmak yada kaldırmak önemli bir işlem adımıdır. Yüklenen ses dosyaları şekil 8.34’deki metoda gönderilerek gürültü temizleme işlemi gerçekleştirilir.

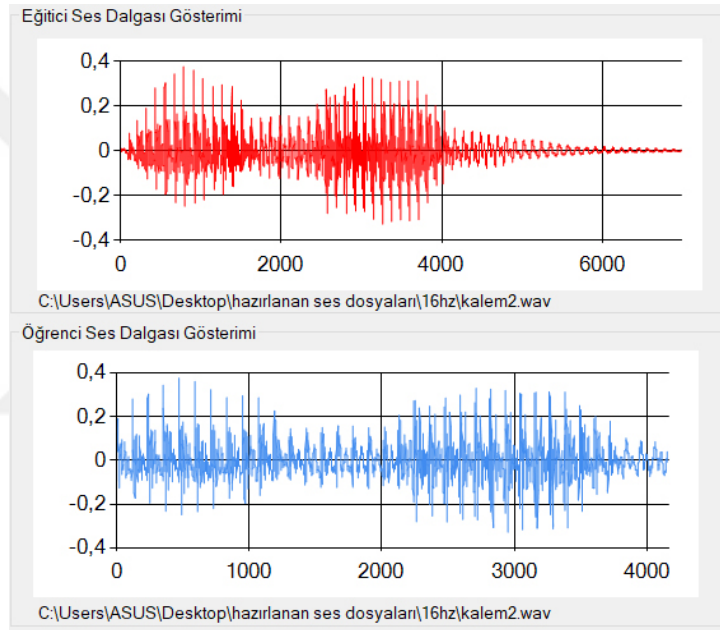
```

1 reference
public static double[] RemoveSilence(double[] pSignal, int pSamplingRate, double pSignalLengthInMilliseconds, double pFrameLengthInMilliseconds)
{
    SignalFrame[] originalFrames = DivideSignalToFrames(pSignal, pSamplingRate, pSignalLengthInMilliseconds, pFrameLengthInMilliseconds);
    SignalFrame[] filteredFrames = RemoveSilentSegments(originalFrames);
    int signalLength = 0;
    foreach (SignalFrame frame in filteredFrames)
    {
        signalLength += frame.Data.Length;
    }
    double[] filteredSignal = new double[signalLength];
    int index = 0;
    foreach (SignalFrame frame in filteredFrames)
    {
        frame.Data.CopyTo(filteredSignal, index);
        index += frame.Data.Length;
    }
    return filteredSignal;
}

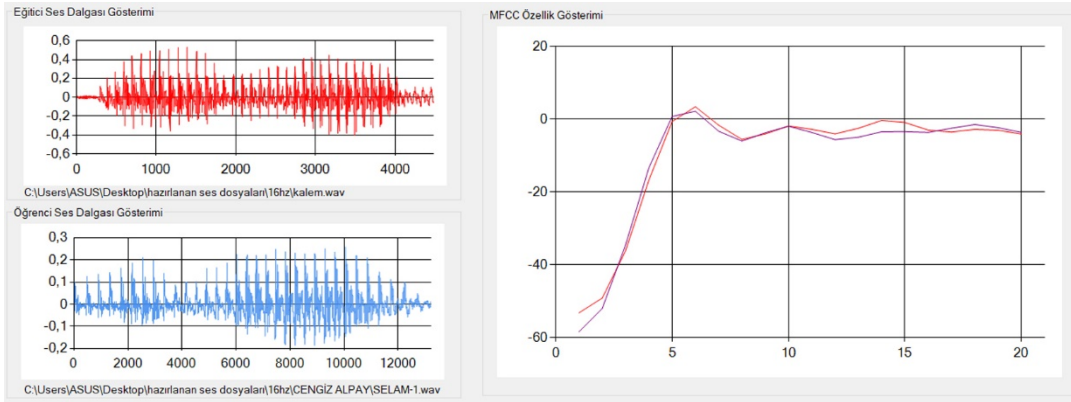
```

**Şekil 8.34 : Sinyal gürültü temizleme metodu**

Şekil 8.35 ve Şekil 8.36’da sayısal olarak elde edilen ses sinyallerinden mevcut gürültülerin temizlendikten sonraki grafiği gösterilmiştir.

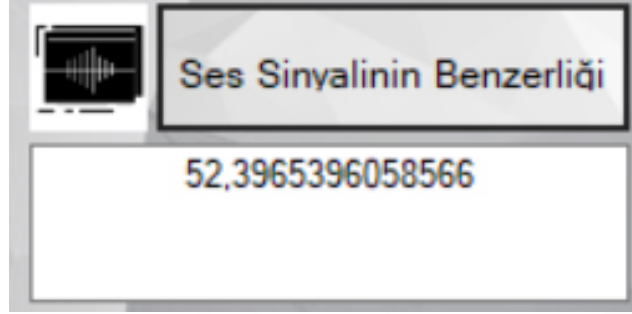


**Şekil 8.35 : Ses sinyalinin temizlenmiş görüntüsü**



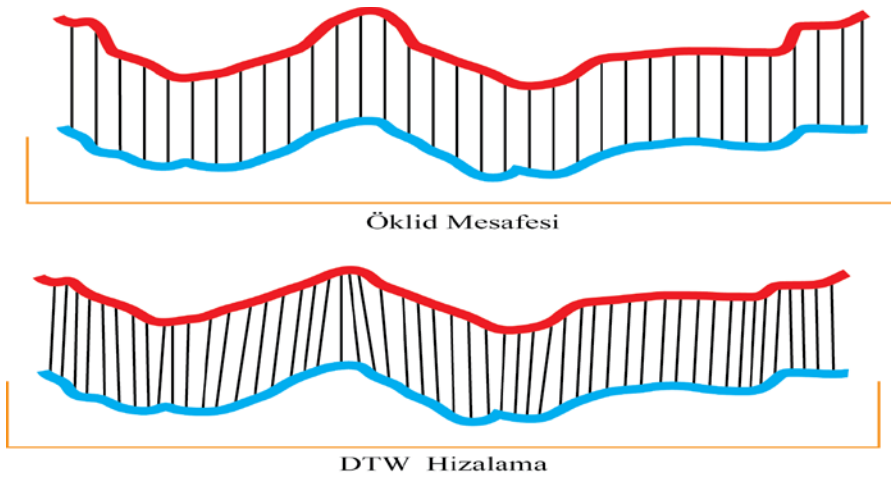
**Şekil 8.36 : Ses bilgilerinin grafiksel gösterimi**

Eđitici ve đrencinin ses bilgilerinin Őekil 8.36'daki grafiksel gsterimindeki benzerlik sadece insan tarafından yorumlanabilir bir zelliktir. Bu benzerlik oranının sayısal bir karŐılıđının bulunması gerekmektedir. Sayısal karŐılıđı elde etmek iin Őekil 8.37'da gsterildiđi gibi iki ses sinyalinin benzerlik iŐlemine sayısal bir deđer Őeklinde gerekleŐtirmeliyiz.



**Őekil 8.37 :** Ses sinyalinin sayısal benzerlik gsterimi

Bir zaman dizisi olarak konuŐulan kelimeler olayların gerekleŐtiđi zaman sırasına gre temsil edilen bir olgudur. evremizde birok Őey zaman serisine dayanarak gerekleŐmektedir. İzole edilmiŐ kelimelerin konuŐma tanıma yaklaŐımlarında her kelimenin bir prototipi ile karŐılaŐtırarak en yakın eŐleŐmeyi ya da o Őablona benzerlik oranını bulmayı hedefler. Őablon olarak nceden konuŐulan kelimelerin bir zellik vektrleri dizisi olarak dŐunebiliriz. đrencinin eđitilmesi istenen kelimeyi konuŐtuđu zaman bunu Őablonu ile karŐılaŐtırılması, her bir zellik vektrlerinin ikili olarak karŐılaŐtırılmasıyla sađlanabilir. İki konuŐma dizisi arasındaki toplam mesafe, zellik vektrleri arasındaki ayrı ayrı mesafelerin toplamı veya ortalaması olacaktır.



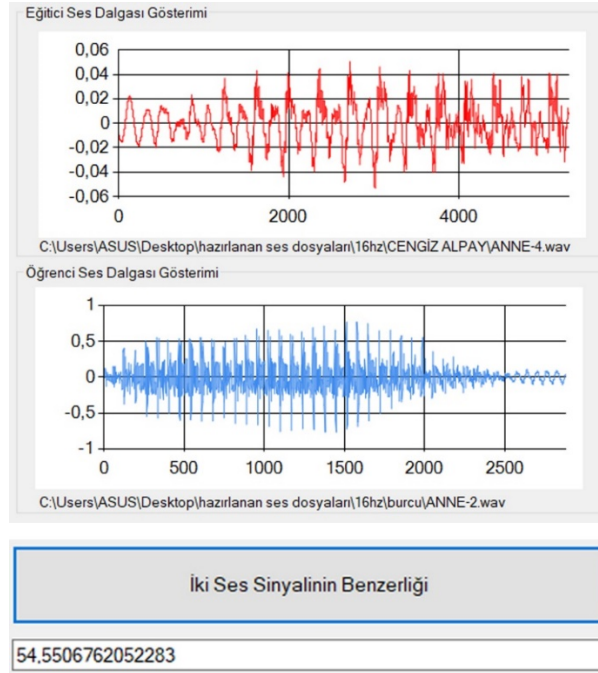
**Őekil 8.38 :** İki sinyal arası zaman serisi hesaplama

Şekil 8.38'deki gibi kusursuz iki zaman serisi arasındaki benzerlik bilgisini Öklid mesafesi hesaplanabilirdi. Ancak bu iki zaman serisinin senkronize olması ve tam olarak aynı hız ve zamanda hareket etmesi durumundan benzerlik için iyi bir ölçüm işlemidir. Dizilerin senkronize olmadığı durumlarda Öklid mesafeleri büyüterek iki seri arasındaki benzerlik oranında düşmesi ile sonuçlanır. Çalışmamızda iki ses sinyalinin benzerlik değeri için dinamik zamana atlama (DTW) algoritmasını kullandık.

```
private void button4_Click(object sender, EventArgs e)
{
    DTWHesapla sdt = new DTWHesapla(mfc1donen, mfc2donen);
    sdt.computeDTW();
    ArrayList liste = sdt.getDistanceList();
    textBox1.Text = sdt.getSum().ToString();
}
```

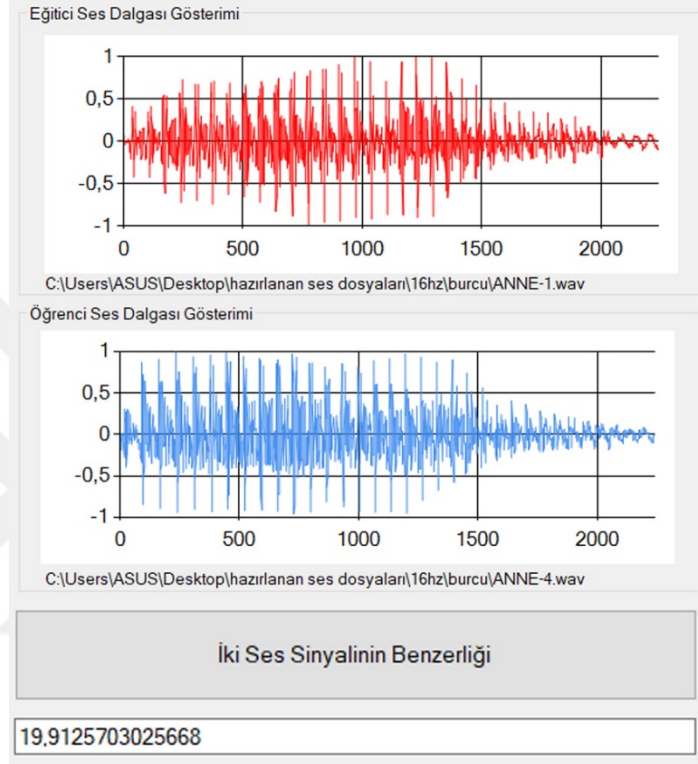
**Şekil 8.39 :** DTW metodu ile iki sinyalin karşılaştırılması

İki ses sinyalinin şekil 8.39'da gösterildiği gibi DTW metoduna gönderilerek karşılaştırma sonucunda dönecek olan bilgi benzerlik bilgisini verir. Bu dönen bilgi şekil 8.40'da gösterildiği gibi 0 değerine yakın olması iki ses sinyalin birbirine benzer olduğunu gösterir.

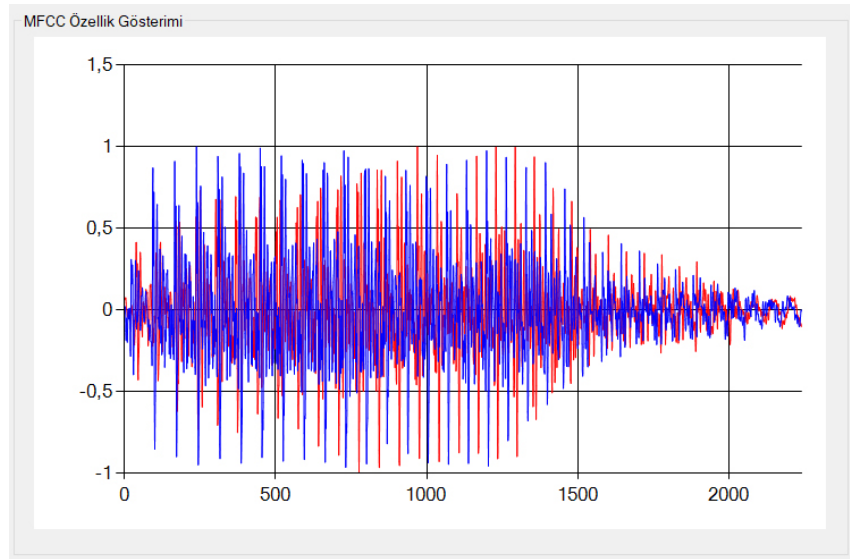


**Şekil 8.40 :** İki farklı sesin benzerlik değeri

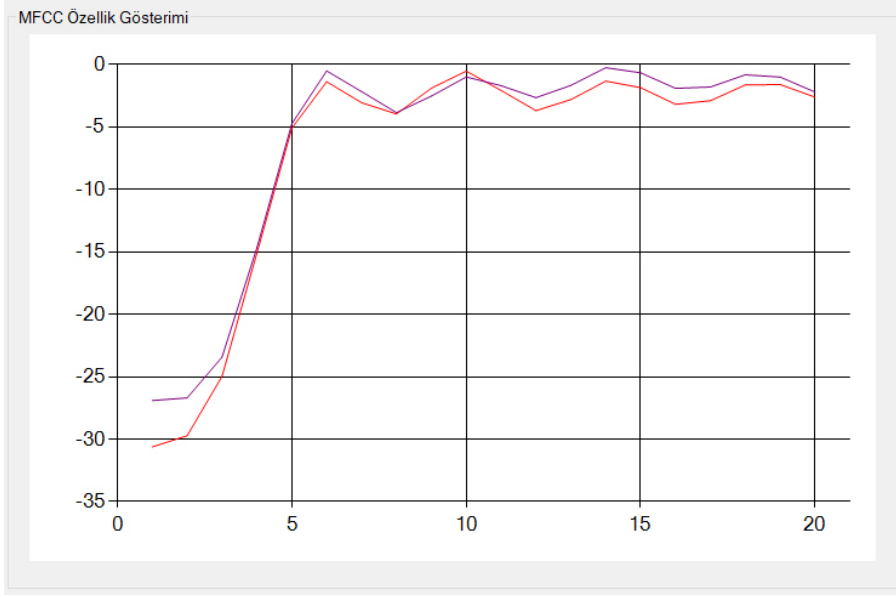
Aynı kişi tarafından söylenen aynı kelimelerin söylenişi sırasında bile Şekil 8.41’de gösterildiği gibi farklı ses dalga değerleri oluşur. İki ses sinyalinin benzerlik değerlerinin 0 yakın olması kelimenin söylenişinin doğru olduğu anlamını gelir. Eğitimci ile birlikte gerçekleştirilecek çalışmalarda işitme engelli öğrenci farklı eğitimciler tarafından söylenen kelimeleri eğitim videosundan ve grafiksel olarak görerek ona yakın bir ses çıkışı sağlayabilir.



Şekil 8.41 : Aynı kişi tarafından söylenen aynı kelimenin benzerlik değeri



Şekil 8.42 : Aynı kişi tarafından söylenen aynı kelimenin benzerlik grafiği



**Şekil 8.43 :** Aynı kişi tarafından söylenen aynı kelimenin MFCC değer grafiği

Ses bilgisinin eğitici ile olan benzerlik ölçüsü ile karşılaştırılabileceği gibi daha önceki kayıtlardan elde edilen ses bilgileri ile karşılaştırılarak diğer ses bilgilerine olan benzerlik değerleri Şekil 8.44 gösterildiği gibi ölçülür.

Ses Dosyası Adı	MFCC 0	MFCC 1	MFCC 2	MFCC 3	MFCC 4	MFCC 5	MFCC 6	MFCC 7	MFCC 8	MFCC 9	MFCC 10	MFCC 11
C:\Users\ASUS\...	-52.56	-49.78	-38.9	-19.96	-3.24	0.68	-4.95	-7.94	-4.25	-1.1	-3.14	-5.52
C:\Users\ASUS\...	-40.94	-39.88	-32.67	-16.6	-0.9	2.77	-3.36	-6.86	-3.16	0.31	-1.73	-4.59
C:\Users\ASUS\...	-46.14	-42.7	-31.53	-14.76	-1.48	0.83	-4.02	-6.24	-3.41	-1.53	-3.71	-5.74
C:\Users\ASUS\...	-45.95	-44.91	-37.25	-19.56	-1.56	3.77	-2.11	-6.3	-3.29	-0.08	-2.1	-4.91
C:\Users\ASUS\...	-40.42	-41.99	-39.49	-24.61	-5.13	3.79	0.55	-3.1	-1.16	0.19	-3.65	-7.41
C:\Users\ASUS\...	-49.26	-47.44	-38.09	-19.06	-0.6	4.55	-1.86	-6.77	-4.28	-0.85	-1.99	-4.09
C:\Users\ASUS\...	-46.34	-37.23	-18.48	-0.07	5.4	-0.35	-4.4	-0.97	3.13	1.93	-1	-1
C:\Users\ASUS\...	-47.64	-45.4	-35.49	-16.92	-0.05	3.57	-3.24	-7.53	-4.39	-1.09	-3.14	-6.27

**Şekil 8.44 :** İki farklı sesin uzaklığı

## 8.5 Konuşma Seslerinin Karşılaştırılması

Son adımda, giriş sözcükleri HMM ve K-NN sınıflandırıcı tarafından tanımlanır. Tanımlanacak kelimeler hazırlanmış olduğumuz veritabanını kullanarak karşılaştırma işlemi gerçekleştirilir. Gizli Markov Modelleri (HMM) normal ve konuşma bozukluğu yaşayan insanlar için konuşma tanıma uygulamaların da bir sınıflayıcı



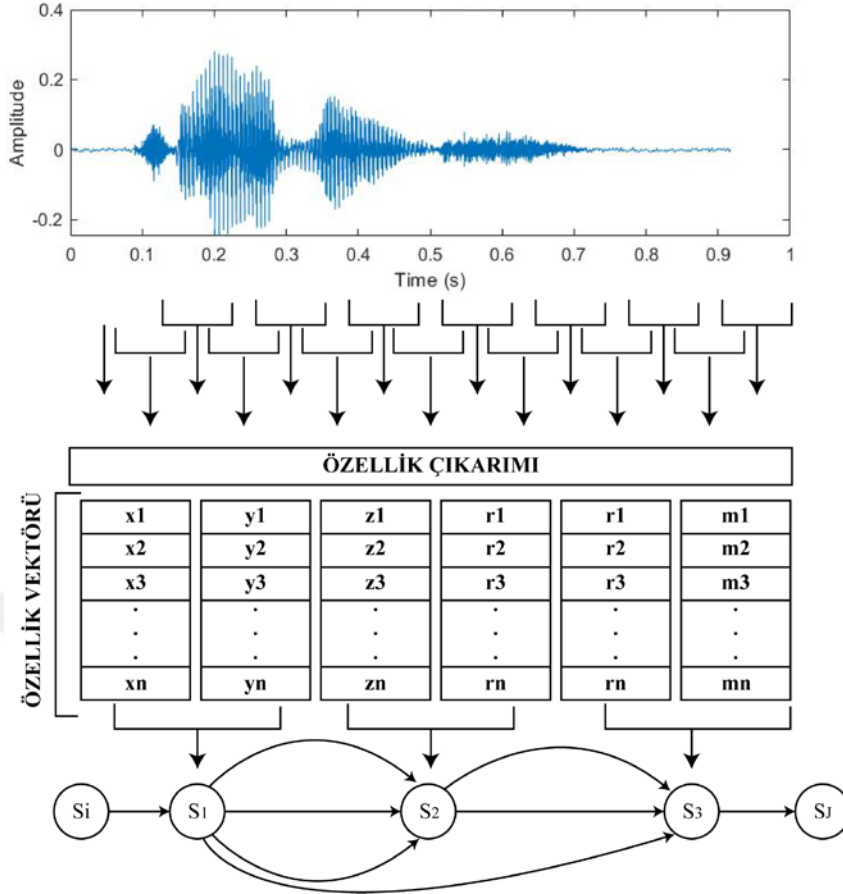
olarak kullanılır. HMM’de izole edilmiş kelimeleri modellemek için her kelimenin durumunu önceden tanımlanır. K-NN sınıflandırıcı ise parametrik olmayan örüntü sınıflandırma yöntemidir. HMM ve K-NN sınıflandırıcıları kullanarak sesli ve görsel özelliklerin birleşik vektörüne uygulanır.

### **8.5.1 HMM Sınıflandırıcısı**

Gizli Markov modeli (HMM), çok çeşitli zaman serisi verilerini modellemek için kullanılan popüler bir istatistiksel yöntemdir. Doğal dil işleme (NLP) yöntemlerinde konuşmanın etiketlenmesi ve kelime öbeklerine ayrıştırma başarılı sonuçlar verir (Arfeen ve Aggarwal, 2016). HMM’ler sinyal işlemede özellikle konuşma tanıma sistemlerinde yaygın olarak kullanılır.

Konuşma tanıma sistemleri birbirine bağımlı uygulamalara örnektir. Örnek olarak bir sözcükte arka arkaya gelen harfler bağımlı bir yapıya sahiptir; Türkçede “b” harfinden sonra “a” harfinin gelme olasılığı yüksekken “f” harfinin gelme olasılığı çok düşüktür. Konuşma tanıma, sesin temel sesbirimciklerinden oluştuğu ve belirli bir biçimde sıralandığı varsayılır. Sözcükler o dile ait sözdizimsel ve anlambilimsel kurallara bağlı olarak belirli bir biçimde sıralanabildiği ve bu sıralamaların cümlelere karşılık geldiği söylenebilir (Alpaydın, 2018).

HMM, bir zaman diliminde durum değiştiren sonlu bir durum makinesidir. HMM tabanlı konuşma tanıma sistemlerinde, her bir kelimeye karşılık gelen vektörler dizisinin bir Markov zinciri tarafından üretildiği varsayılır. Şekil 8.45’de HMM akustik modelin eğitim diyagramı verilmiştir.



**Şekil 8.45 :** HMM Akustik modelin eğitim diyagramı.

HMM’de izole kelimeleri modellemek için, her kelimenin durum numarasını önceden tanımlamalıyız. Sınıflandırma işleminde en iyi durum geçiş şeklini araştırır ve Viterbi algoritmasını kullanarak olasılık değerini belirleriz. Her bir HMM’ nin olasılık değerini karşılaştırarak HMM’nin en büyük olasılıklı değeriyle belirleriz.

Sistemimizde kullandığımız HMM ile konuşma tanıma modelini işitme engelli öğrencilerin konuşmaları üzerinde gerçekleştirdik. İşitme engelli öğrencinin Abi, Bak, Beyaz gibi söylemiş olduğu kelimeleri diğer kelimelerle olan HMM uzaklık değerleri Çizelge 8.1’de gösterilmektedir.

**Çizelge 8.1 :** HMM ile öğrenci ve eğitimci konuşma uzaklık değerleri

KELİME	ABİ	BAK	BEYAZ
BİR	-24733.177	-18225.2312	-17195.3418
İKİ	-20806.5403	-13803.702	-12875.5111
ÜÇ	-22489.3205	-14558.0416	-12721.7663
DÖRT	-21237.2322	-11424.4435	-12661.8109
BEŞ	-22727.8317	-14505.2292	-13668.7915
ALTI	-22176.7696	-9614.722	-11889.2332
YEDİ	-22564.9758	-15648.2083	-14271.8969



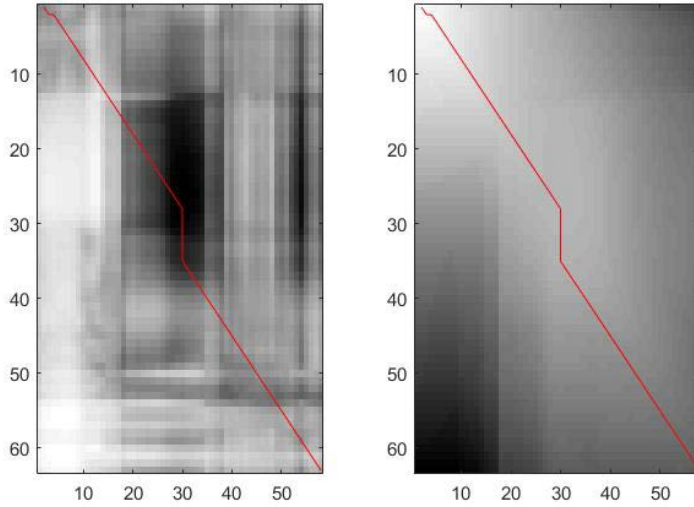
SEKİZ	-20534.4728	-13896.9983	-13934.913
DOKUZ	-28084.4104	-10865.0184	-15960.2735
ON	-27065.9424	-11977.205	-15841.5259
BEYAZ	-22304.8377	-11129.2684	-11281.0969
SİYAH	-19613.3852	-8942.4953	-12632.2699
MAVİ	-15405.4994	-10129.8667	-11503.0207
MOR	-16320.5036	-14271.8969	-21891.8691
KIRMIZI	-20190.7974	-11538.9703	-11786.0773
PEMBE	-21217.0709	-10320.8107	-11942.1191
GRİ	-18377.6444	-12637.2963	-13371.708
TURUNCU	-17580.2401	-9807.2508	-10514.9832
ANNE	-21088.2739	-10615.355	-13133.9053
BABA	-31066.6009	-10962.1437	-16320.5036
DEDE	-21479.8385	-13435.7713	-12689.7563
ABİ	-19829.966	-10750.2228	-16879.8424
AMCA	-17101.8713	-9449.4167	-12080.4356
TEYZE	-21891.8691	-12765.8692	-12657.0537
HALA	-28687.8451	-11488.5566	-17877.5624
ANNEANNE	-21163.9315	-10985.0956	-12351.15
BABAANNE	-21437.3281	-10289.0122	-13615.3899
SELAM	-20630.8123	-10212.6016	-11471.3165
GEL	-23291.1687	-13030.6117	-12821.3888
GİT	-23208.2054	-17691.7511	-17657.9288
KAL	-27167.4986	-10382.4301	-15205.3389
DÖN	-21008.6113	-11091.7082	-12780.0796
BAK	-21626.8137	-9388.0675	-13290.1098
TAK	-23799.3772	-9024.3653	-13390.303
ARABA	-23330.2139	-9355.7674	-13567.6145
TREN	-19585.0025	-13592.1841	-12622.4957
KAMYON	-19580.6428	-9997.3055	-14521.2724
OTOBÜS	-17697.5532	-12029.405	-12822.2859
UÇAK	-23125.2829	-9307.6774	-12275.0518

HMM, konuşma tanıma işlemi ile Abi kelimesinin en yakın eşleştiği kelime Mavi, Bak kelimesinin Siyah, Beyaz kelimesinin Turuncu olarak tanıma işlemi gerçekleştirmiştir.

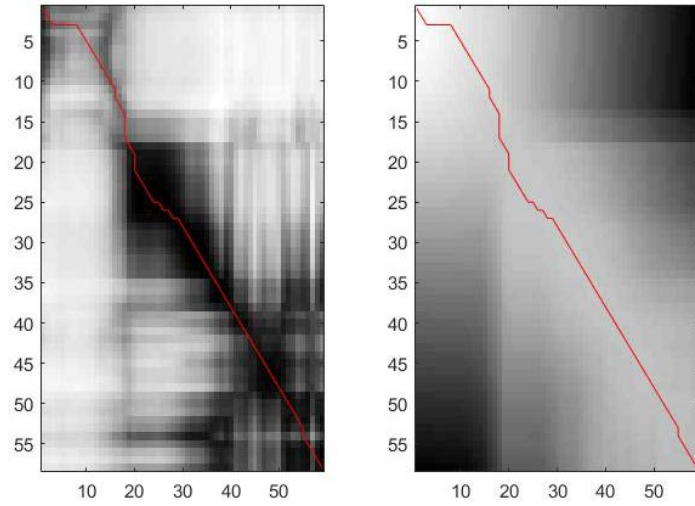
### 8.5.2 Dinamik zaman atlama (DTW)

Bir zaman serisi, zaman içinde art arda yapılan bir gözlemler topluluğudur. DTW, zamanlamada değişkenlik gösterebilecek iki zaman serisi arasındaki benzerliği ölçmek için kullanılan bir algoritmadır. Bir zaman serisi değerleri tek başına bir anlam ifade etmezken bunun grafiksel gösterimi bize farklı uygulama alanları için bilgi verebilir. Gerçek zamanlı olarak kalp atış değerleri, konuşma desenleri, DNA yapıları, el yazıları hatta şekiller bile bir dizi zaman serisine dönüştürülebilir.

Konuşma işlemi, istenen bilginin bir konuşma sinyalinden elde edilmesi ile gerçekleşir. Akustik desenlerin eşleştirilmesine dayanan konuşma tanıma sistemleri, zaman içerisinde değişim gösteren ölçülerin karşılaştırılması ilkesine dayanır. Aynı kelime farklı kişiler tarafından söylendiğinde oluşan konuşma sinyallerinde gözlemlenen farklılık Şekil 8.46'de görülebilir. Bununla beraber aynı kişi aynı kelimeyi aynı sürelerde söylese dahi konuşma sinyallerinde yine farklılıklar meydana gelir. Bu farklılıklar Şekil 8.47'de verilen görsellerde gözlemlenebilir.

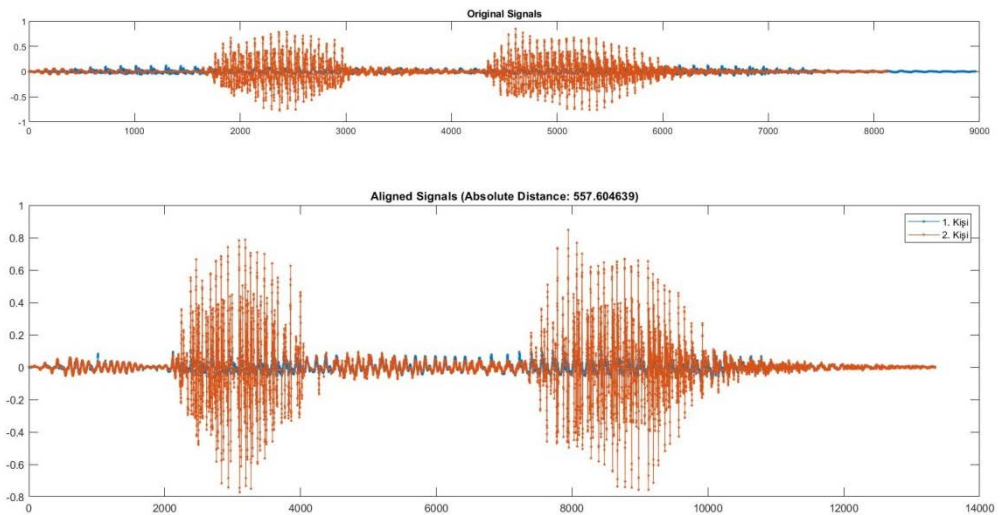


Şekil 8.46 : DTW ile farklı kişilerin aynı kelimeyi söyleyişi



**Şekil 8.47 :** DTW ile aynı kişinin aynı kelimeyi söyleyiş gösterimi

İki konuşma deseni arasında global bir mesafe elde etmek için bir zaman hizalaması yapılmalıdır (Kaleka, 2010). Dinamik zaman atlama yöntemi ile iki konuşma desenine karşılık gelen özellikler ortak bir zaman ekseninde aynı yerde görüntülenebilir. Böylece sinyaller arasındaki benzerlikler vurgulanır. Dinamik zaman atlama (DTW), konuşmanın spektral dizi karşılaştırması ile ilgili sorunların çözümünde uygulanabilir. Şekil 8.48’de gösterildiği gibi Dinamik zaman atlama yöntemi, sesli ifadeleri sıkıştırma yada genişletme yoluyla ses desenlerinin referans karşılaştırmalarını yapar.



**Şekil 8.48 :** DTW ortak bir zaman ekseninde genişletilmesi

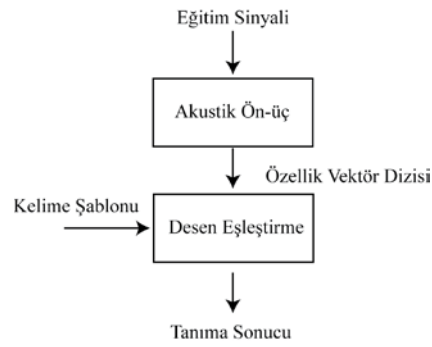
İzole edilmiş kelime tanıma sistemlerinde, sözcüklerdeki her kelimenin akustik deseni veya şablonu, bir zaman özelliği olarak depolanır. Tanıma işlemi, tanınacak olan kelimenin akustik desenlerinin karşılaştırılması ve daha sonra tanınmış kelime

ile en iyi eşleşen kelimenin seçilmesi işlemidir. Bu, iki farklı desenin zaman uyumunun karşılaştırılması anlamına gelir.

Çalışmamız sırasında örüntü eşleştirme işlemini gelen konuşma bilgilerini bellekteki var olan bilgilerle eşleştirmeye çalıştık. DTW, izole sözcükler için konuşma tanıma sistemlerinde kullanılan desen eşleştirme tekniğidir. Temel olarak gelen kelimeyi bellekte kayıtlı referans kelimesiyle eşleştirmeye çalışır. Gizli Markov modelleri konuşma tanımda sıklıkla kullanılmasına rağmen, örnek kelimeler, şablonlar arasındaki zamanlama farklılıklarını barındırmak için Dinamik Zaman Atlama (DTW) adı verilen dinamik programlama teknikleri kullanılmaktadır. Genel olarak örnek bir zaman çerçevesindeki şablonlarda maksimum eşleşmeyi bulmaya çalışmak yatar.



**Şekil 8.49 :** DTW sistemlerindeki temel izole kelimeler



**Şekil 8.50 :** DTW sistemlerindeki temel izole kelime tanıma

Konuşmacıdan bağımsız konuşma tanıma sistemlerinde belirli bir konuşmacıyı tanımak için sistemin eğitilmesi gerekmez. DTW'nin avantajları arasında verimli donanım kullanımı ile birlikte eğitim verisinin basit ve hızlı oluşturulmasının gerçekleştirilmesidir. Çalışmamızda sesler arasındaki benzerliği puanlayarak aynı kelimenin iki farklı ifadesini senkronize ederek eşleştirme işlemini gerçekleştirdik.

Elde edilen iki konuşma sinyali arasındaki özellik vektörleri arasındaki mesafe ölçüsü hesaplanır. Bu mesafe Öklid bağıntısı kullanılarak gerçekleştirilir.

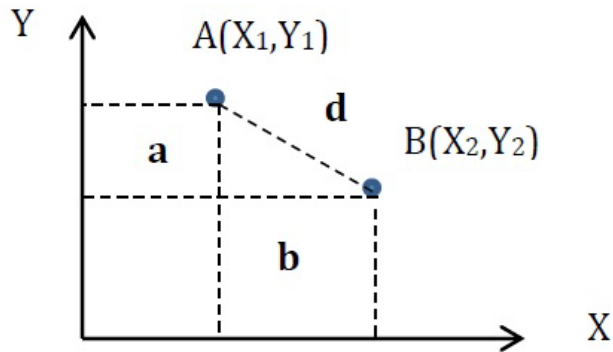
A ve B noktaları arasındaki uzaklığı belirleyen Öklid bağıntısı şu şekilde hesaplanır;

$$d(A, B) = \sqrt{(X_1 - X_2)^2 + (Y_1 - Y_2)^2} \quad (8.7)$$

Bu bağıntı genelleştirilecek olursa, aşağıdaki gibi bir bağıntıya ulaşılır;

$$d(i, j) = \sqrt{\sum_{k=1}^p (X_{ik} - X_{jk})^2} \quad (8.8)$$

Konuşma işlemede aynı kelime farklı sürelerde söylenebilir. Farklı kelimelerin söyleniş şeklinin aynı zamanda gerçekleştirilmesi zordur. İki konuşma deseni arasında global bir mesafe elde etmek için zaman hizalaması yapılmalıdır. “Zaman - Zaman” matrisi, hizalamayı görselleştirmek için kullanılır. Söylenen kelimeler arasındaki uzaklığın iki boyutlu gösterimi Şekil 8.51’de gösterilmiştir.



**Şekil 8.51 :** İki nokta arasındaki uzaklığın iki boyutlu uzayda gösterimi

Eğitimcilerin söylemiş olduğu kelimelerin DTW uzaklık mesafesi değerleri Çizelge 8.2, Çizelge 8.3, Çizelge 8.4’ de verilmiştir.

**Çizelge 8.2 :** “Abi” kelimesi’nin eğitimciler için DTW uzaklık değerleri

	<b>E1</b>	<b>E2</b>	<b>E3</b>	<b>E4</b>	<b>E5</b>
<b>E1</b>	0	53,77	40,26	11,95	21,83
<b>E2</b>	53,77	0	28,36	47,71	49
<b>E3</b>	40,26	28,36	0	35,69	31,80
<b>E4</b>	11,95	47,71	35,69	0	24,40
<b>E5</b>	21,83	49	31,80	24,40	0

**Çizelge 8.3 :** “Bak” kelimesi’nin eğitimciler için DTW uzaklık değerleri

	E1	E2	E3	E4	E5
E1	0	30,20	118,59	59,78	56,74
E2	30,20	0	30,20	54,748	43,01
E3	118,59	95,80	0	86,521	60,32
E4	59,78	54,748	86,521	0	55,635
E5	56,74	43,01	60,32	55,635	0

**Çizelge 8.4 :** “Beyaz” kelimesi’nin eğitimciler için DTW uzaklık değerleri

	E1	E2	E3	E4	E5
E1	0	24,74	26,77	40,53	31,37
E2	24,74	0	14,23	36,97	29,74
E3	26,77	14,23	0	37,02	23,74
E4	40,53	36,97	37,02	0	44,85
E5	31,37	29,74	23,74	44,85	0

Çizelge 8.5’de işitme engelli bir öğrencinin, eğitimciler arasındaki DTW uzaklık değerleri gösterilmiştir.

**Çizelge 8.5 : Kelimelerin öğrenci ve eğitimci konuşma uzaklık değerleri**

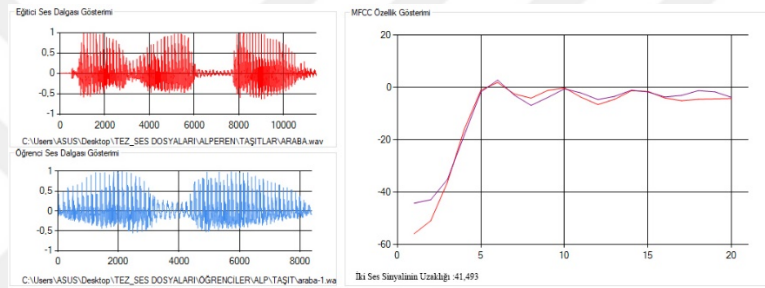
KELİMELER	EĞİTİCİ 1 ÖĞRENCİ	EĞİTİCİ 2 ÖĞRENCİ	EĞİTİCİ 3 ÖĞRENCİ	EĞİTİCİ 4 ÖĞRENCİ	EĞİTİCİ 5 ÖĞRENCİ
Bir	34,97	39,16	57,38	42,47	35,57
İki	72,43	49,83	59,28	65,85	60,19
Üç	123,90	84,76	83,09	131,21	76,36
Dört	110,45	68,77	75,00	74,03	89,10
Beş	64,61	81,34	77,36	103,73	39,14
Altı	38,03	25,40	51,86	35,81	18,66
Yedi	49,86	40,35	30,17	80,58	41,55
Sekiz	89,04	45,04	60,34	73,93	26,34
Dokuz	61,92	20,08	55,74	45,73	53,98
On	30,98	38,04	36,39	83,97	46,75
Beyaz	25,25	47,01	32,25	33,97	30,62
Siyah	45,62	26,21	72,52	40,13	31,04
Mavi	24,12	35,49	44,31	22,18	22,91
Mor	75,21	34,77	92,76	37,90	54,91
Kırmızı	83,91	71,70	52,40	60,44	29,61
Pembe	60,71	49,85	43,52	67,17	58,44
Gri	55,52	47,66	29,48	18,77	20,18
Turuncu	69,77	46,83	75,73	34,45	49,34
Anne	76,15	57,24	82,58	60,39	46,13
Baba	27,15	19,28	32,46	67,63	54,47
Dede	13,89	53,93	26,20	54,94	27,35
Abi	42,53	52,22	45,43	40,96	53,62
Amca	29,62	27,98	14,62	44,90	55,90
Teyze	70,78	51,50	66,32	71,60	39,67
Hala	91,06	42,20	55,89	87,51	91,75
Anneanne	71,15	46,12	65,93	41,87	15,75
Babanne	37,95	39,19	46,85	44,90	45,41
Selam	29,02	18,26	61,20	14,67	35,01
Gel	43,05	30,63	22,67	57,09	50,86
Git	48,99	43,04	42,16	37,85	32,24
Kal	91,24	45,81	76,88	23,10	57,14
Dön	45,13	26,02	71,06	15,31	27,69
Bak	40,94	69,77	38,58	74,31	88,16
Tak	42,13	77,12	21,56	83,64	47,73
Araba	16,83	28,93	40,83	41,49	33,62
Tren	39,86	29,32	44,48	46,09	39,52
Kamyon	79,15	70,68	83,78	64,82	38,05
Otobüs	88,96	16,35	29,58	51,03	47,13
Uçak	142,14	102,65	90,67	60,64	95,24



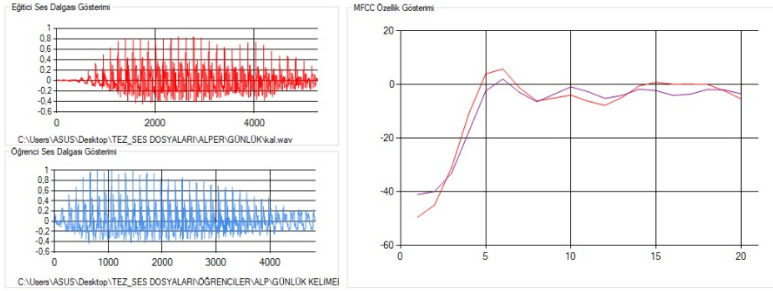


## 9. SONUÇLAR VE ÖNERİLER

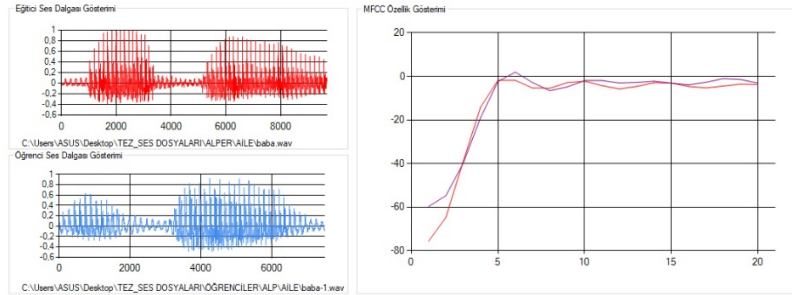
Çalışmada öncelikle işitme engeli olmayan ve normal konuşma yeteneğine sahip eğitimcilerin ses bilgileri kendi aralarında kıyaslanmıştır. Daha sonra, eğitimcilerle işitme engelli bireylerin konuşmaları karşılaştırılmıştır. Bu kıyaslamalara ait detaylı grafiklere ekler bölümünde verilmiştir. Buna göre işitme engelli bireyler arasında az da olsa kelime çıkarma yeteneğine sahip olan bireylerin ses bilgilerinin eğitimcilerin konuşmaları ile benzerlik gösterdiği Şekil 9.1, Şekil 9.2, Şekil 9.3’de görülmektedir.



Şekil 9.1 Araba kelimesinin eğitimci ve öğrenci ses bilgisinin karşılaştırılması

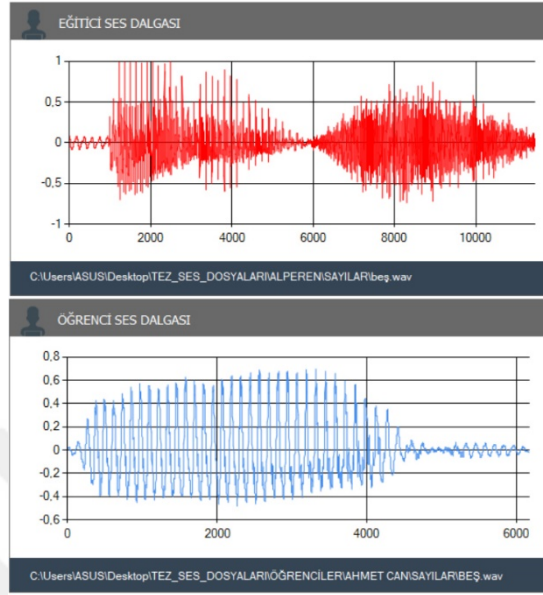


Şekil 9.2 Kal kelimesinin eğitimci ve öğrenci ses bilgisinin karşılaştırılması



Şekil 9.3 Baba kelimesinin eğitimci ve öğrenci ses bilgisinin karşılaştırılması

Bunun yanında doğuştan itibaren herhangi bir koklear implant cihazı kullanmamış ve konuşma yeteneğine sahip olmayan işitme engellilerin konuşmalarında Şekil 9.4’de gösterildiği gibi büyük farklılıklar ve eksiklikler bulunduğu gözlenmiştir.



**Şekil 9.4** Beş kelimesinin eğitici ve öğrenci ses bilgisinin karşılaştırılması

Türkiye’de mevcut işitme engelli bireylerin eğitiminde ciddi zorluklar yaşanmakta ve eksiklikler bulunmaktadır. Klasik konuşma terapi eğitimlerinde özellikle eğitmenin öğrencinin sürekli olarak dudak hareketlerini ve ses çıkışını kontrol etmesi oldukça güçtür ve bu durum eğitimin kalitesini etkilemektedir. Konuşma sırasında söylenen kelimenin doğru yada yanlış olduğunu öğretmek ve açıklamak için çocuğa anında ve anlamlı görsel geribildirim sunulabilmelidir. Bu alandaki bu yetersizlikler dikkate alınarak bu çalışmada, işitme engelli bireylerin konuşma terapi eğitimleri için kullanılacak türkçe bilgisayar destekli bir konuşma terapisi yazılımı gerçekleştirilmiştir. Sistem işitme engelli bireylerin konuşma sinyallerini görselleştirme ve işitme engelli bireyin ses bilgilerini görerek tekrar etmelerini sağlamak için tasarlanmıştır. İşitme engelli bireyler öncelikle belirlenmiş kelimeler için kaydedilen örnek eğitim videosunu izleyecek ve daha sonra ona yakın bir konuşma gerçekleştirecektir. Yazılımımız ses işleme ve görüntü işleme olmak üzere iki sistem parçasından oluşmaktadır. Her bir parça ön işlem ve özellik çıkarımı olmak üzere iki basamaktan meydana gelmektedir. Görüntü işleme kısmında yüz lokasyon tespiti, ağız lokasyon tespiti ve dudak hareketlerinin görsel özelliklerinin gösterim işlemleri gerçekleştirilmektedir. Ses işleme kısmında ise ses sinyallerinin temizlenmesi ve ses bilgisi özelliklerinin elde edilmesi işlemleri yapılmaktadır.

Konuşma tanıma Gizli Markov modelleri (HMM) sıklıkla kullanılır ancak örnek kelimeler şablonlar arasındaki zamanlama farklılıklarını barındırdığı için Dinamik Zaman Atlama (DTW) modeli gibi dinamik programlama teknikleri daha başarılı sonuçlar vermektedir. Bu nedenle bu çalışmada HMM ve DTW teknikleri bir arada kullanılmıştır. HMM ve DTW grafikleri için detaylı karşılaştırma bilgileri ekler bölümünde verilmiştir.

Bu yazılımın bir çok işitme engelli bireyin konuşma terapi sürecine önemli katkılar sunacağı düşünülmektedir. Çünkü yazılım, görüntülerden ve ses bilgilerinden elde edilen verileri görselleştirerek hem konuşma terapistine hem de işitme engelli bireye görsel geribildirimler sunmaktadır. Grafikler ile geri dönüt işlemlerinin gerçekleştirilmesi, konuşmayı öğrenme sürecinde işitme engelli bireyi uzun süreli ve etkili bir eğitim için motive edebilir. Özel eğitim uzmanına yardımcı olarak eğitim öğretim sürecinin verimliliğini de arttırabilir. Ayrıca özel eğitime ihtiyaç duyan bireyler için geliştirilen ve geliştirilecek olan sistemler, bu alanda yetişmiş insan gücü eksikliği sorununu da çözüm getirebilir. Bu sistemlerin istikrarlı, taşınabilir ve daha az maliyetli olması da avantajları arasında sayılabilir. Bu gerekçelere dayanarak çalışmamızın bu boşluğu dolduracak nitelikte olduğu söylenebilir.

İşitme engelli çocukların iletişim becerisi kazandırılması için işitme kalitesinin en yoğun ve uygun bir şekilde kullanılması kişiye özel eğitim ortamlarında uygun yaşantılarla desteklenerek geliştirileceğine inanmaktayız. Ses ve görüntü işleme yöntemleri ile gerçekleştirdiğimiz konuşma terapi uygulamamız bir ekran vasıtasıyla çocuğun konuşma üretimini ve buna bağlı olarak terapistin doğru konuşma modelini karşılaştırması etkili bir konuşma materyali olarak kullanılabilir.

Bilgisayar destekli konuşma terapi sistemleri geliştirilirken genel olarak bazı sorulara cevap vermesi ve değerlendirme yapılması gerekmektedir. Bu eğitim sistemi uzun soluklu olup erken yaşlardan itibaren aile ve çevre tarafından desteklenip kabul görmesi gerekmektedir.

Belli başlı sorular ;

1. Bilgisayar destekli yapılan çalışma sonucunda konuşmanın ne derece iyileştirilip, geliştirildiği.
2. Konuşma terapistlerinin bilgisayar destekli konuşma terapisi uygulamasından nasıl faydalanıp yararlandığı.

3. Yardımın amaçlanan algısal, bilişsel ve dikkat çekici özelliklerini dikkate alıp almadığı (Yardımanın farklı yaşlarda veya gelişim seviyelerinde kullanım kolaylığının belirlenmesi gerektiği).
4. İşitme engelli bireyin konuşma yeteneğinin olup olmadığı ve istenen konuşma özelliklerini geliştirmesine yardım edip etmediği gibi sorunlara cevap aranmalıdır.

Çocukların bilgisayarla yaptıkları konuşma etkinlikleri terapistlerle konuşmaktan daha az korkutucu gelmiştir. Onlar için bu çalışma heyecan verici olmaktadır. Bilgisayar kullanımı çocuğun motivasyonu için önemli bir faktör olarak görünmektedir. Teknolojinin çekiciliğinin yanında bilgisayarın tarafsız olarak görülmesi kelimeleri tekrar tekrar denemek için özverili olmasını sağlamaktadır.

Konuşma terapi etkinlikleri sırasında küçük yaş gurupları arasındaki çocukların herhangi bir şekilde kelime bilgisine sahip olmadığını fark ettik. Bu sebeple küçük yaştaki çocuklar için kelime telaffuzundan ziyade seslendirme, ses süresi kontrolleri, nefes kontrolleri, tekrarlanan hece üretimi, odaklanma yoğunluğu gibi eğitim içeriği genişletilebilir. Fizyolojik ve akustik önlemler, fonlamanın hızı ve kalitesi ve nefes akışı üzerindeki kontrol gibi temel özelliklerin altında yatan bazı faaliyetlerin tahminlerini elde etmek için kullanılabilir.

Çeşitli konuşma tanıma sistemlerinde sesle çalışan basit bilgisayar oyunları bulunmaktadır. Buradaki amaç çocuğa ilgi çekici olarak çocuğa geri bildirim sağlamaktır. Ses yoğunluğunun zamansal çizelgesinin gösterilmesi çocukta herhangi bir bilgi sağlamada yeterli olmayacaktır. Geliştirmeyi amaçladığımız sistemdeki görsel bilgiler çocuk için yeteri kadar geri dönüt sağlamamaktadır. Ses kontrollerini grafiksel olarak oyunlarla gösterimi daha ilgi çekebileceğini düşünmekteyiz. Farklı oyunlarla seslendirme sonrası farklı görsel geri bildirim veren sistemlerin tasarlanarak geri bildirimler olmadan da konuşma yeteneğinin kontrol edilmesi sağlanabilir.

İşitme engelli birey doğru bir kelime üretimi gerçekleştirdiği zaman, bu üretimin sürdürülebilir olması gerekmektedir. Bu çalışma, öğretim ve ev uygulamaları için bir eğitim asistanı sistemi şeklinde sunulabilir. Bu eğitimlerin doğru becerileri geliştirebilmesi için önemli yatırımlar ve destekler sunulmalıdır. Bu konuşma terapisindeki en önemli unsurlardan birisidir. Hedef konuşma üretimi çeşitli

seviyelerde tekrarlanmalı ve uygulanmalıdır. Bu doğru bir konuşma üretiminin istikrarlı bir sistemle kurulmasıyla oluşturulur. Eğitim ortamından bağımsız olarak evde yada farklı bir yerde konuşma terapi uygulaması sürdürülebilir. Bu sebeple eğitim verimliliği ve sürekliliği arttırılabilir.

Türkiye’de mevcut işitme engelli kişiler için eğitim öğretim gücünde ciddi eksiklikler bulunmaktadır. Özel eğitime ihtiyaç duyan bireyler için geliştirilecek olan sistemler eğitim öğretim verimliliğini arttırarak insan gücü eksikliği sorununa çözüm önerisi getirebilir. Etkileşimli sistemler ile birlikte istikrarlı, taşınabilir ve daha ucuz maliyetli sistemler kurulabilir.

Ses işleme ve görüntü işleme ile birlikte gerçekleştirilecek dudak okuma ve ses işleme teknikleri konuşma tanıma sistemlerinin daha etkin olarak kullanılmasını sağlayabilir. Önerdiğimiz konuşma terapi uygulaması ile birlikte daha kapsamlı deneyler ve işitme engellilerin dahil edildiği arayüz çalışmaları yapılacaktır. Gerçekleştirilecek çalışmalarla birlikte daha büyük bir konuşma veritabanı toplanarak makine öğrenmesi ve yapay zeka teknolojileri ile birlikte daha etkin bir sistem çalışması yapılacaktır. Bu insan-makine etkileşim seviyesini arttırmayı sağlayarak teknolojinin özellikle işitme engellilerin faydasına olacak gelişmeler sağlayabilir.



## KAYNAKLAR

- Abdulghafour, M. (2003).** Image Segmentation Using Fuzzy Logic And Genetic Algorithms. *Journal of WSCG*, 11(1).
- Adams, F. R., Crepy, H., Jameson, D., & Thatcher, J. (1989).** IBM products for persons with disabilities. *1989 IEEE Global Telecommunications Conference and Exhibition 'Communications Technology for the 1990s and Beyond'*, (s. 980-984). Dallas,USA.
- Akçamete, G. (1993).** İşitme Engellilerde Dil ve Konuşma. *Özel Eğitim Dergisi*, 1(3), 2-9.
- Akmeşe, P. P., & Kayhan, N. (2016).** İşitme Engelliler Öğretmenleri ile Odyoloji ve Konuşma Bozuklukları Uzmanlarının İşitme Kayıplı Çocukların Eğitimi Hakkındaki Görüşleri. *Mehmet Akif Ersoy Üniversitesi Eğitim Fakültesi Dergisi*(40), 88-112.
- Alpaydın, E. (2018).** *Yapay Öğrenme* (4. b.). İstanbul, Sarıyer: Boğaziçi Üniversitesi Yayınevi.
- Altıparmak, A. (2017).** Psikodilbilimsel Konuşma Üretimi Modellerine İlişkin Bir İnceleme. *Söylem Filoloji Dergisi*, 2(2), 314-347.
- Arfeen, Z., & Aggarwal, J. (2016).** Speech Recognition based on Hidden Markov Model Toolkit (HTK) with BODO Language. *International Journal of Advanced Research in Computer Science and Software Engineering*, 6(9), 383-388.
- Baykoç, D. N. (1987).** 12-30 aylık Türk çocuklarında dilin kazanılması. *Çocuk Gelişimi ve Eğitimi Dergisi*(2), 36-38.
- Bernstein, L. E., Goldstein, M. H., & Mahshie, J. J. (1988).** Speech training aids for hearing-impaired individuals: 1. Overview and aims. *Journal of Rehabilitation Research and Development*, 25(4), 53-62.
- Bovik, A. (2000).** *Handbook of Image and Video Processing*. San Diego: Academi Press.

- Butler, K. A., & K. Jacob, R. J.** (1997). Human-Computer Interaction: Introduction and Overview. *Human Factors in Computing Systems, CHI '97: Looking to the Future, Extended Abstracts*, (s. 105-106). Atlanta, USA.
- Canny, J.** (1986). A Computational Approach to Edge Detection. *IEEE Trans Pattern Analysis and Machine Intelligence*, 8(6), 679-698.
- Card, S. K., Moran, T. P., & Newell, A.** (2008). *The Psychology of Human-Computer Interaction*. New Jersey: CRC Press.
- Cengiz, D. U., Kolcu, D., & Ercan, M. K.** (2016). İşitme Engelli Çocuklarda Dil Kazanımı ve Konuşma Eğitimi. *International Congress on Woman and Child Health and Training*. Kocaeli.
- Chen, T.** (2001). Audiovisual Speech Processing Lip Reading and Lip Synchronization. *IEEE Signal Processing Magazine*, 9-21.
- Chen, T., & Rao, R.** (1998). Audio-Visual Integration in Multimodal Communication. *Proceeding of the IEEE*, 86, 837-852.
- Chomsky, N.** (2001). *Dil ve Zihin*. (A. Kocaman, Çev.) Ankara: Ayraç Yayınevi.
- Cootes, T.** (2000). An Introduction to Active Shape Models.
- Cootes, T. F., Edwards, G. J., & Taylor, C. J.** (1998). Active appearance models. *European Conference on Computer Vision*, (s. 484-498).
- Çağiltay, K.** (2018). *Teoriden Pratiğe İnsan-Bilgisayar Etkileşimi ve Kullanılabilirlik Mühendisliği* (2. b.). Ankara: Seçkin Yayınları.
- Duda, R. O., Hart, P. E., & Stork, D. G.** (2001). *Pattern Classification* (2. b.). A Wiley-Interscience Publication.
- Engwall, O., Bälter, O., Öster, A.-M., & Kjellström, H.** (2006). Designing the user interface of the computer-based speech training system ARTUR based on early user tests. *Behaviour & Information Technology*, 25(4), 353-365.
- Erber, N. P.** (1975). Auditory-Visual Perception of Speech. *Journal of Speech and Hearing Disorders*, 40(4), 481-492.
- Estellers, V., & Thiran, J.-P.** (2012). Multi-pose lipreading and audio-visual speech recognition. *EURASIP Journal on Advances in Signal Processing*.



- Eveno, N., Caplier, A., & Coulon, P. Y.** (2002). A parametric model for realistic lip segmentation. *7th International Conference on Control, Automation, Robotics and Vision*, (s. 1426-1431). Singapore.
- Foley, J. D., Dam, A. v., Feiner, S. K., & Hughes, J. F.** (2014). *Computer graphics: principles and practice* (Third Edition b.). New Jersey: Pearson Education.
- Fu, Z., Lu, G., Ting, K. M., & Zhang, D.** (2011). A Survey of Audio-Based Music Classification and Annotation. *IEEE Transactions on Multimedia*, 13(2), 303-319.
- Gao, W., Zhang, X., Yang, L., & Liu, H.** (2010). An improved Sobel edge detection. *3rd International Conference on Computer Science and Information Technology*, (s. 67-71). Chengdu, China.
- Gerçeker, M., Yorulmaz, İ., & Ural, A.** (2000). Ses ve Konuşma. *K.B.B. ve Baş Boyun Cerrahisi Dergisi*, 71-78.
- Girgin, M. C.** (2006). İşitme Engelli Çocukların Konuşma Edinimi Eğitiminde Dinleme Becerilerinin Önemi. *Ankara Üniversitesi Eğitim Bilimleri Fakültesi Özel Eğitim Dergisi*, 7(1), 15-28.
- Gonzalez, R. C., & Woods, R. E.** (2018). *Sayısal Görüntü İşleme*. (Z. Telatar, H. Tora, F. Arı, & A. Kalaycıoğlu, Çev.) Ankara: Palme Yayınevi.
- Gürel, A., & Arslan, L. M.** (2016). Konuşma Tanıma için İnsan-makine Karşılaştırması. *Dilbilim Araştırmaları Dergisi*, 77-90.
- Gürsakal, N.** (2017). *Makine Öğrenmesi ve Derin Öğrenme* (1 b.). Bursa: Dora Yayınları.
- Hassanat, A.** (2009). *Visual Words for Automatic Lip-Reading*. University of Buckingham. UK: University of Buckingham.
- Hazen , T. J.** (2006). Visual model structures and synchrony constraints for audio-visual speech recognition. *IEEE Transactions on Audio, Speech, And Language Processing*, 14(3).
- Hazen , T. J., Saenko, K., La , C.-H., & Glass, J.** (2004). A segment-based audio-visual speech recognizer: Data collection, development and initial

- experiments. *In Proceedings of the International Conference on Multimodal Interfaces*, (s. 235-242). Pennsylvania.
- Honda, K.** (2008). Physiological Processes of Speech Production. J. Benesty, M. M. Sondhi, & Y. A. Huang içinde, *Springer Handbook of Speech Processing* (s. 7-26). Berlin, Heidelberg.
- Hsiao, M.-L., Li, P. T., Lin, P. Y., Tang, S.-T., Lee, T.-C., & Young, S.-T.** (2001). A computer based software for hearing impaired children's speech training and learning between teacher and parents in Taiwan. *2001 Conference Proceedings of the 23rd Annual International Conference*, (s. 1457-1459). Istanbul, Turkey.
- Hu, Y., Wang, T., & Huang, R.** (2013). Development of Speech Training Aid System for Hearing-impaired Children. *IEEE 13th International Conference on Advanced Learning Technologies*, (s. 212-214).
- Huang, N., Wu, H., & Song, Y.** (2009). A Nios II Based English Speech Training System for Hearing-Impaired Children. *2009 International Conference on Computer Engineering and Technology*, (s. 452-456). Singapore, Singapore.
- Hudgins, C. V., & Numbers, F. C.** (1942). An investigation of the intelligibility of the speech of the deaf. *Genetic Psychology Monographs*(25), 289-392.
- Ibrahim, Y. A., Odiketa, J. C., & Ibiyemi, T. S.** (2017). Preprocessing Technique in Automatic Speech Recognition for Human Computer Interaction : An Overview. *Annals. Computer Science Series*, 15(1), 186-191.
- Jang, K.-S.** (2007). Lip Contour Extraction based on Active Shape Model and Snakes. *International Journal of Computer Science and Network Security*, 7(10), 148-153.
- Javkin, H., Keate, E., Antonanzas-Barroso, N., Yamada, Y., & Youdelman, K.** (1994). Automatic model parameter generation for the speech training of deaf children. *Proceedings of ICASSP '94. IEEE International Conference on Acoustics, Speech and Signal Processing*, (s. 229-232). Adelaide, SA, Australia.

- Ji, X., Zhang, C., Wang, J., & Kai, W.** (2009). Fast 2-D 8×8 discrete cosine transform algorithm for image coding. *Science in China Series F Information Sciences*, 52(2), 215-225.
- Jun, H., & Hua, Z.** (2009). Research on Visual Speech Feature Extraction. *International Conference on Computer Engineering and Technology*. Singapore.
- Kaleka, J. S.** (2010). Isolated Word Recognition using Dynamic Time Warping. *Proceedings of the International Conference on Circuits, Systems, Signals*, (s. 293-295).
- Kass, M., Witkin, A., & Terzopoulos, D.** (1988). Snakes: Active contour models. *International Journal of Computer Vision*, 1(4), 321-331.
- Kol, S.** (2011). Erken Çocuklukta Bilişsel Gelişim ve Dil Gelişimi. *Sakarya Üniversitesi Eğitim Fakültesi Dergisi*, 1-21.
- Kotropoulos, C., & Pitas, I.** (1997). Rule-Based Face Detection in Frontal Views. *Proc. Int'l Conf. Acoustics, Speech and Signal Processing*, 4, 2537-2540.
- Kulkarni, A. D.** (2000). *Computer Vision and Fuzzy-Neural Systems*. New Jersey: Prentice Hall PTR.
- Kumar, G., & Bhatia, P. K.** (2014). A Detailed Review of Feature Extraction in Image Processing Systems. *2014 Fourth International Conference on Advanced Computing & Communication Technologies*, (s. 5-12). Rohtak, India.
- Lavagetto, F.** (1995). Converting speech into lip movements: a multimedia telephone for hard of hearing people. *IEEE Transactions on Rehabilitation Engineering*, 3(1), 90-102.
- Lijuan, S., Jian, Z., Lirong, W., Zhiyong, A., Biao, S., & Hao, L.** (2015). Design of Measurement System for Respiratory Training Device Based on Speech Training for Hearing-Impaired Children. *2015 IEEE Fifth International Conference on Big Data and Cloud Computing*, (s. 203-206). Dalian, China.
- Ling, D.** (1976). *Speech and the hearing-impaired child : theory and practice*. Washington : Alexander Graham Bell Association for the Deaf.

- Lucey, S.** (2002). *Audio-visual Speech Processing*. Queensland University of Technology.
- Matthews, I., Cootes, T. F., Bangham, J. A., Cox, S., & Harvey, R.** (2002). Extraction of visual features for lipreading. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 24(2), 198-213.
- McGurk, H., & MacDonald, J.** (1976). Hearing lips and seeing voices. *Nature*(264), 746-748.
- Neumeyer, L., Franco, H., Digalakis, V., & Weintraub, M.** (2000). Automatic scoring of pronunciation quality. *Speech Communication*, 30(2-3), 83-93.
- Neyman, L., & Bogomilsky, M. R.** (2001). *Anatomi, işitme ve konuşma organlarının fizyolojisi ve patolojisi*.
- Nickerson, R. S., & Stevens, K. N.** (1972). Teaching speech to the deaf: can a computer help. *72 Proceedings of the ACM annual conference*, 240-251.
- Nilsson, N. J.** (2018). *Yapay Zeka Geçmişi ve Geleceği*. (M. Doğan, Çev.) İstanbul: Boğaziçi Üniversitesi Yayınevi.
- Patel, N., Dangarwala, K., Shah, A., & Mistry, M.** (2014). A Study of Digital Image Filtering Techniques in Spatial Image Processing. *International Conference on Convergence of Technology*, (s. 1-7).
- Petajan, E. D.** (1984). Automatic lipreading to enhance speech recognition. *in Proc, Global Telecomm*, (s. 265-272). Atlanta.
- Potamianos, G., Neti, C., Gravier, G., Garg, A., & Senior, A. W.** (2003). Recent Advances in the Automatic Recognition of Audio-Visual Speech. *Proceedings of the IEEE*, 91(9), 1-16.
- Ranjan, R., & Dubey, R. K.** (2016). Isolated Word Recognition using HMM for Maithili dialect. *2016 International Conference on Signal Processing and Communication (ICSC)*, (s. 323-327). Noida, India.
- Rao, K.** (1989). Overview Of Image Processing. *Readings in Image Processing*.
- Rao, P.** (2007). Audio Signal Processing. B. Prasad, & S. R. Mahadeva içinde, *Speech, Audio, Image and Biomedical Signal Processing using Neural Networks*. Springer.

- Rathee, N.** (2016). A novel approach for lip reading based on neural network. *International Conference on Computational Techniques in Information and Communication Technologies (ICCTICT)*, (s. 1-6). New Delhi, India.
- Ravi, P., & Ashokkumar, A.** (2017). Analysis of Various Image Processing Techniques. *International Journal of Advanced Networking & Applications (IJANA)*, 8(5).
- Reisberg, D., McLean, J., & Goldfield, A.** (1987). Easy to hear, but hard to understand: A lipreading advantage with intact auditory stimuli. *Hearing by eye: The psychology of lip-reading*, 97-113.
- Riella, R., Linarth, A., Lippmann, L., & Nohama, P.** (2001). Computerized System to Aid Deaf Children in Speech Learning. *IEEE EMBS International Conference*.
- Rocchesso, D.** (2003). *Introduction to Sound Processing*. California: Creative Commons.
- Russell, M., Brown, C., Skilling, A., Series, R., Wallace, J., Bonham, B., & Barker, P.** (1996). Applications of automatic speech recognition to speech and language development in young children. *Proceeding of Fourth International Conference on Spoken Language Processing. ICSLP '96*. Philadelphia, USA.
- Salama, E. S., El-Khoribi, R. A., & Shoman, M. E.** (2014). Audio-Visual Speech Recognition for People with Speech Disorders. *International Journal of Computer Applications*, 96(2), 51-56.
- Scanlon, P., & Potamianos, G.** (2005). Exploiting lower face symmetry in appearance-based automatic speechreading.
- Schafer, R. W., & Rabiner, L. R.** (1975). Digital representations of speech signals. *Proceedings of the IEEE*, 63(4), 662-677.
- Scharcanski, J., & Venetsanopoulos, A. N.** (1997). Edge detection of color images using directional operators. *IEEE Transactions on Circuits and Systems for Video Technology*, 7(2), 397-401.
- Shannon, C. E.** (1948). A Mathematical Theory of Communication. *The Bell System Technical Journal*, 27, 379-423, 623-656.

- Shin, J., Lee, J., & Kim, D.** (2011). Real-time lip reading system for isolated Korean word recognition. *Pattern Recognition*, 44(3), 559-571.
- Singh, B., Kaur, R., Devgun, N., & Kaur, R.** (2012). The process of Feature Extraction in Automatic Speech Recognition System for Computer Machine Interaction with Humans: A Review. *International Journal of Advanced Research in Computer Science and Software Engineering*, 2(2).
- Sirohey, S.** (1993). *Human Face Segmentation and Identification*. Univ. of Maryland: Technical Report CS-TR-3176.
- Smith, S. W.** (1999). *Digital Signal Processing*. San Diego, California: California Technical.
- Sneha, K. S.** (2017). Image Based Techniques for Lip Segmentation –A Review. *International Journal of Innovative Research in Science, Engineering and Technology*, 6(8), 16809-16811.
- Sridharan, S., & Lucey, P.** (2008). A Visual Front-End for a Continuous Pose-Invariant Lip-reading System. *2nd International Conference on Signal Processing and Communication Systems*. Australia.
- Summerfield, Q.** (1992). Lipreading and audiovisual speech perception. *Trans.R. Soc. Lond*, 335(1273), 71-78.
- Torfi, A., Iranmanesh, S. M., Nasrabadi, N., & Dawson, J.** (2017). 3D Convolutional Neural Networks for Cross Audio-Visual Matching Recognition. *IEEE Journals & Magazines*, 5, 22081 - 22091.
- Turk, M., & Pentland, A.** (1991). Face Recognition Using Eigenfaces. *Proceedings of IEEE Computer Vision and Pattern Recognition*, (s. 586-591).
- Türk, O., & Arslan, L. M.** (2004). Konuşma Terapisine Yönelik Konuşma Tanıma Yöntemleri. *Signal Processing and Communications Applications Conference*.
- Uğurtay, Y., N. G., & Erturan, N.** (1999). 6-10 yaş arası çok ileri derecede işitme engelli çocukların algı, bellek ve küçük kas motor gelişimlerinin incelenmesi. *M.Ü. Atatürk Eğitim Fakültesi Eğitim Bilimleri*(11), 379-404.

- Velichko, V. M., & Zagoruyko, N. G.** (1970). Automatic recognition of 200 words. *International Journal of Man-Machine Studies*, 2(3), 223-234.
- Vicsi, K., Roach, P., Öster, A., Kacic, Z., Barczikay, P., Tantos, A., . . . Sfakianaki, A.** (2000). A Multimedia, Multilingual Teaching and Training System for Children with Speech Disorders. *International Journal of Speech Technology*, 3(3), 289-300.
- Viola, P., & Jones, M.** (2001). Rapid Object Detection using a Boosted Cascade of Simple Features. *Proceedings of the International Conference on Computer Vision and Pattern Recognition*, 511-518.
- Vygotsky, L. S.** (1998). *Düşünce ve Dil*. (S. Koray, Çev.) İstanbul: Toplumsal Dönüşüm Yayınları.
- Wankhede, S. N.** (2014). Designing visual Speech Training Aids for Hearing Impaired Children. *International Journal of Advanced Research in Electrical, Electronics and Instrumentation Engineering*, 3(4), 8726-8733.
- Watson, C. S., Reed, D. J., Kewley-Port, D., & Maki, D.** (1989). The Indiana Speech Training Aid (ISTRA). *Comparisons between Human and Computer-Based Evaluation of Speech Quality*, 32(2), 245-251.
- Werda, S., Tmar, M., & Hamadou, A. B.** (2006). ALiFE: Automatic Lip Feature Extraction: A New Approach for Speech Recognition Application. *2nd International Conference on Information & Communication Technologies*. Damascus, Syria.
- Widrow, B.** (1961). Statistical analysis of amplitude-quantized sampled-data systems. *Transactions of the American Institute of Electrical Engineers, Part II: Applications and Industry*, 79(6), 555-568.
- Witt, S. M., & Young, S. J.** (2000). Phone-level pronunciation scoring and assessment for interactive language learning. *Speech Communication*, 30(2-3), 95-108.
- Yalçın, N.** (2008). Konuşma tanıma teorisi ve teknikleri. *Kastamonu Eğitim Dergisi*, 16(1), 249-266.
- Yang, G., & Huang, T. S.** (1994). Human Face Detection in Complex Background. *Pattern Recognition*, 27(1), 53-63.

**Yang, M.-H., Kriegman, D., & Ahuja, N.** (2002). Detecting Faces in Images: A Survey. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 24(1), 34-58.

**Young, I. T., Gerbrands, J. J., & van Vlie, L. J.** (2004). *Fundamentals Of Image Processing*. Delft University of Technology.

**Yuille, A. L., Cohen, D. S., & Hallinan, P. W.** (1989). Feature extraction from faces using deformable templates. *IEEE Computer Society Conference on Computer Vision and Pattern Recognition*.

**Zhang, L.** (1997). Estimation of the mouth features using deformable templates. *Proceedings of International Conference on Image Processing*, (s. 328-331).

#### **İnternet Kaynakları :**

**Acartürk, C., & Çağıltay , K.** (2019, ocak 12). *İnsan Bilgisayar Etkileşimi ve ODTÜ'de Yürütülen Çalışmalar*. Akademik Bilişim Konferansları: <https://ab.org.tr/ab06/bildiri/59.pdf> adresinden alındı

**Brynjolfsson, E., & McAfee, A.** (2019, Şubat 3). *The Business of Artificial intelligence*. MIT Management Sloan School: <http://mitsloanexperts.mit.edu/the-business-of-artificial-intelligence-erik-brynjolfsson-and-andrew-mcafee/> adresinden alındı

**Dil ve Konuşma Bozuklukları :Ailem Özel Eğitim Kurumları.** (2019, Ocak 20). Ailem Özel Eğitim Kurumları: <http://www.ailemozelegitim.com/bilgi/dil-ve-konusma-bozukluklari/isitme-kaybi-olan-cocuklarda-dil-ve-konusma-gelisimi> adresinden alındı

**Ergenç, İ.** (2018, ocak 2). *Dilin Beyindeki Gerçekleşimi ve Konuşma Eylemi*. Ankara Üniversitesi Açık Ders Malzemeleri: <https://acikders.ankara.edu.tr> adresinden alındı

**Human-Computer Interaction (HCI).** (2019, Ocak 27). Interaction Design Foundation: <https://www.interaction-design.org/literature/topics/human-computer-interaction> adresinden alındı

**Jang, R.** (2019, Ocak 20). *Audio Signal Processing and Recognition*. <http://mirlab.org>: <http://mirlab.org/jang/books/audioSignalProcessing/> adresinden alındı



**McCollum, P.** (2018, Ekim 3). *An Introduction to Back-Propagation Neural Networks*. Encoder The Newsletter of the Seattle Robotics Society: <http://www.seattlerobotics.org/encoder/199811/neural.html> adresinden alındı

**Yelken, K.** (2019, 02 21). *Türk Kulak Burun Boğaz ve Baş Boyun Cerrahisi Vakfı*. <http://www.tkbbv.org.tr/>: <http://www.tkbbv.org.tr/menu/111/ses-nasil-olusur-doc-dr-kursat-yelken> adresinden alındı





## **EKLER**

**EK A:** Yazılıma dahil edilen kelimeler

**EK B:** Eğitim Grafik Örnekleri





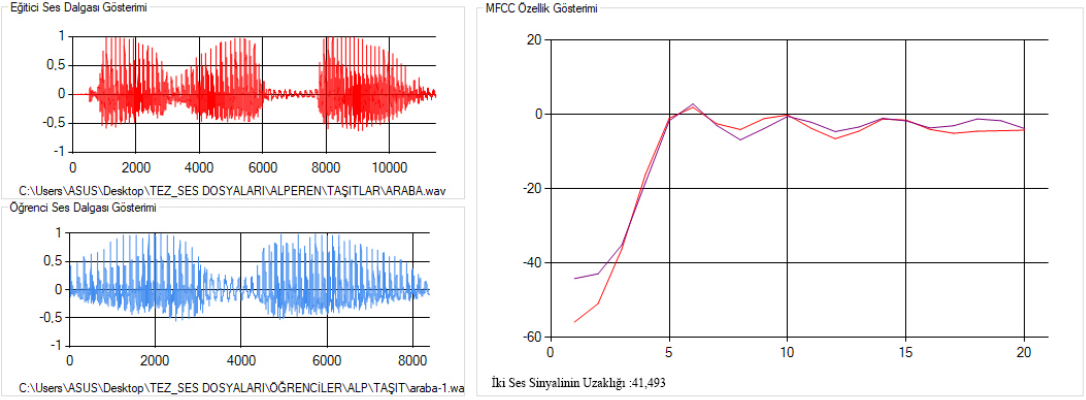
**EK A: Yazılıma dahil edilen kelimeler**

**Çizelge A : Yazılıma dâhil edilen kelimeler**

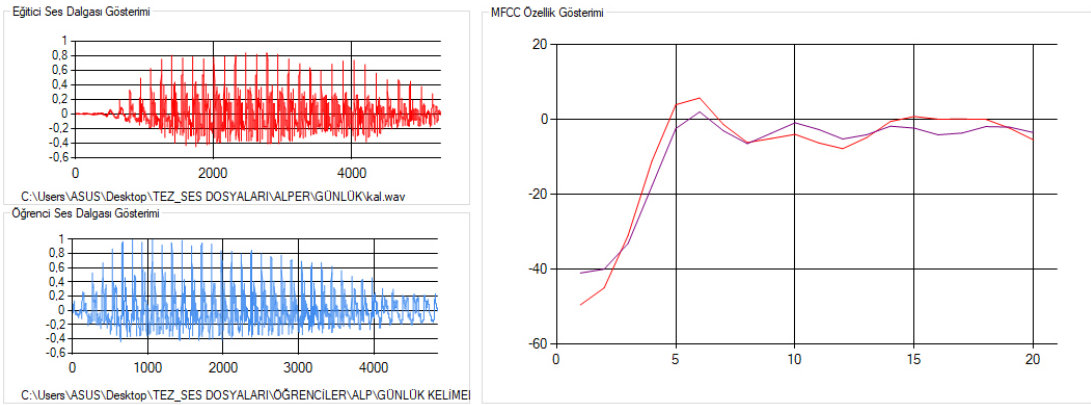
Sayılar	Renkler	Aile	Günlük Kelimeler	Taşıt
Bir	Beyaz	Anne	Selam	Araba
İki	Siyah	Baba	Gel	Tren
Üç	Mavi	Dede	Git	Kamyon
Dört	Mor	Abi	Kal	Otobüs
Beş	Kırmızı	Amca	Dön	Uçak
Altı	Pembe	Teyze	Bak	
Yedi	Gri	Hala	Tak	
Sekiz	Turuncu	Anneanne		
Dokuz		Babaanne		
On				



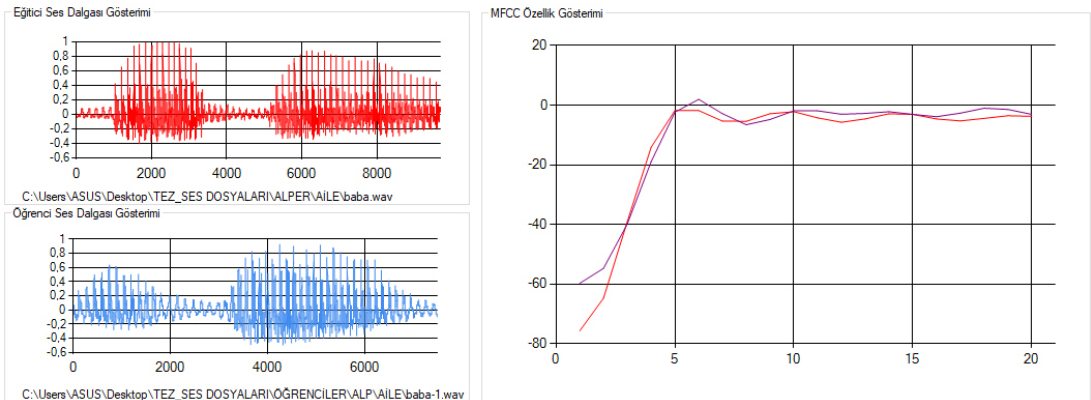
## EK B: Eğitim Grafik Örnekleri



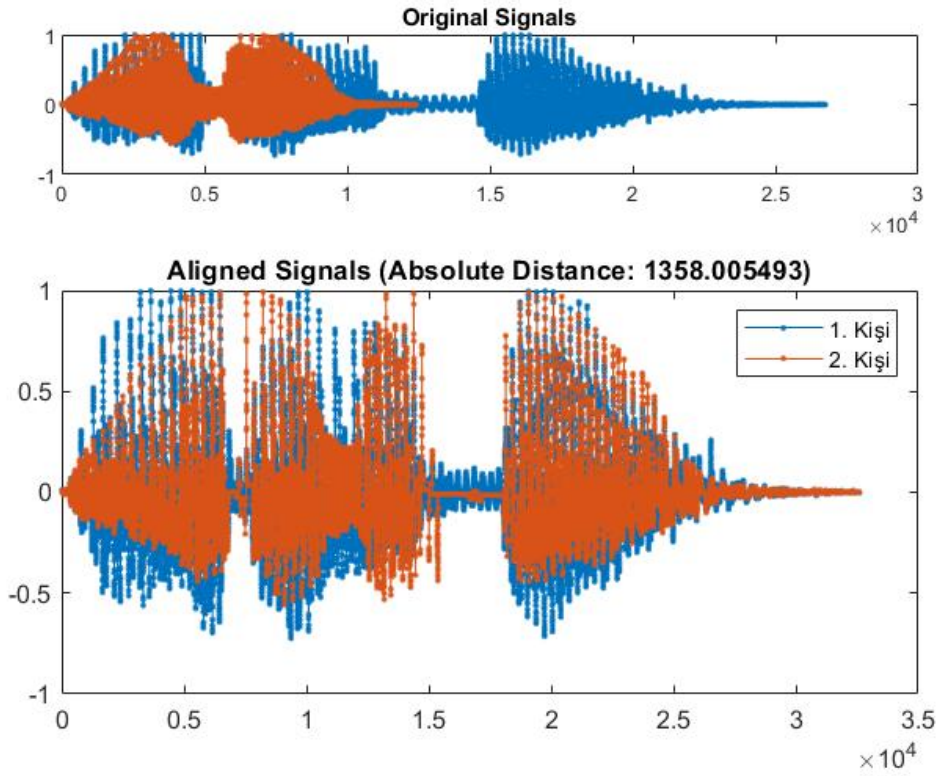
Şekil B.1 : Araba kelimesinin eğitici ve öğrenci ses bilgisinin karşılaştırılması



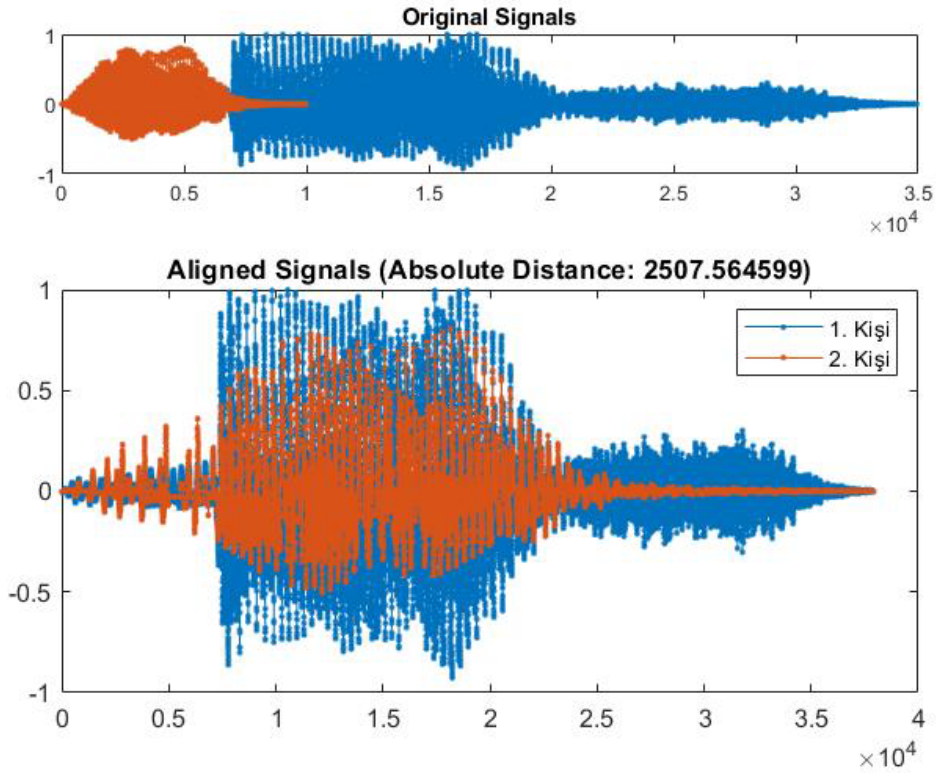
Şekil B.2 (devam) : Kal kelimesinin eğitici ve öğrenci ses bilgisinin karşılaştırılması



Şekil B.3 (devam): Baba kelimesinin eğitici ve öğrenci ses bilgisinin karşılaştırılması

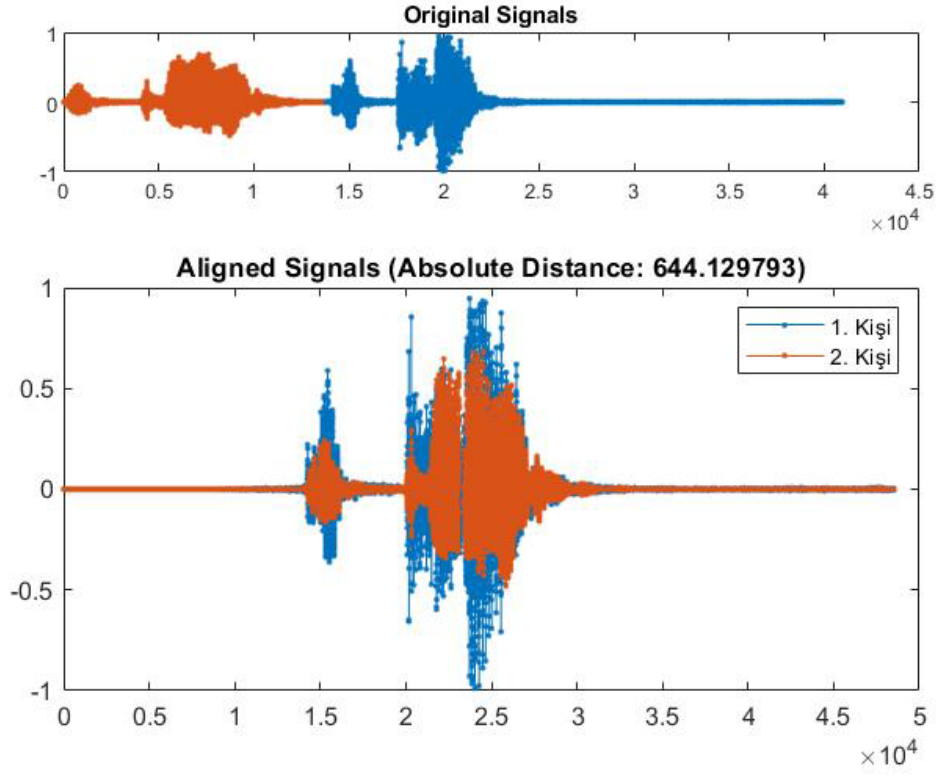


**Şekil B.4 (devam)** : Araba kelimesinin eğitici ve öğrenci ses bilgisinin zamansal eşitlemesi

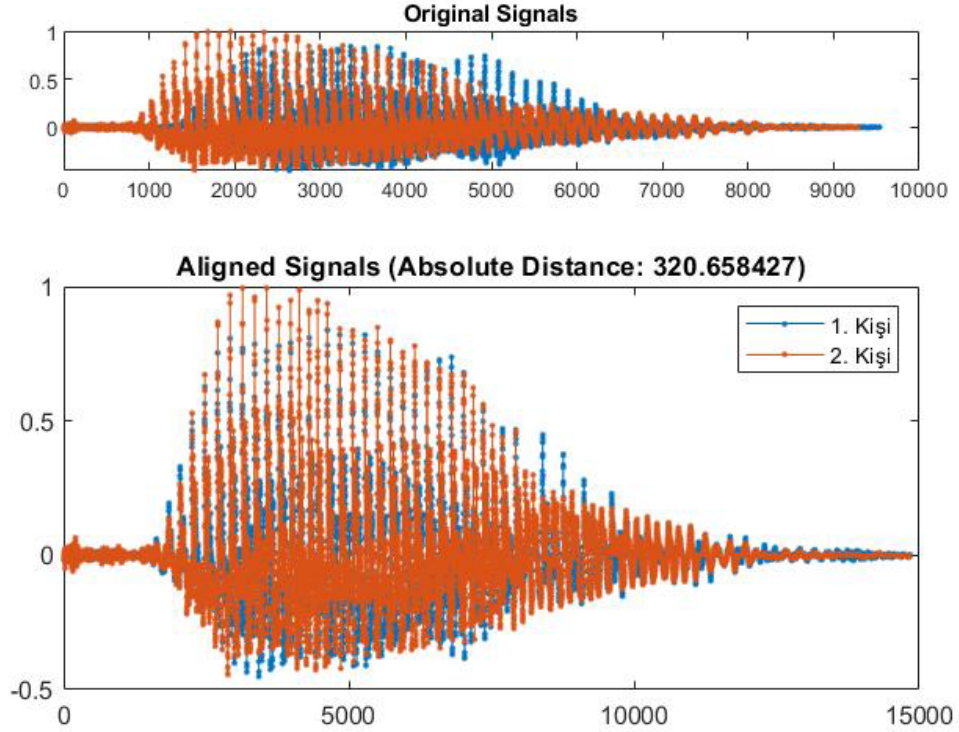


**Şekil B.5 (devam)** : Beyaz kelimesinin eğitici ve öğrenci ses bilgisinin zamansal eşitlemesi

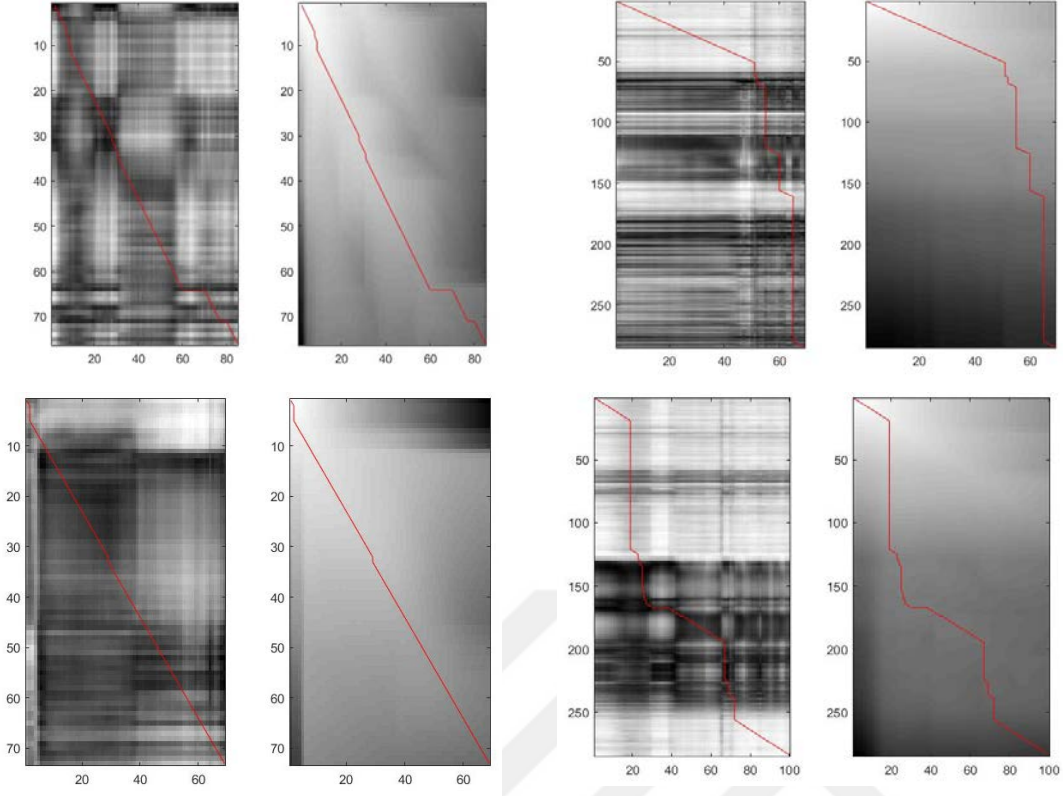




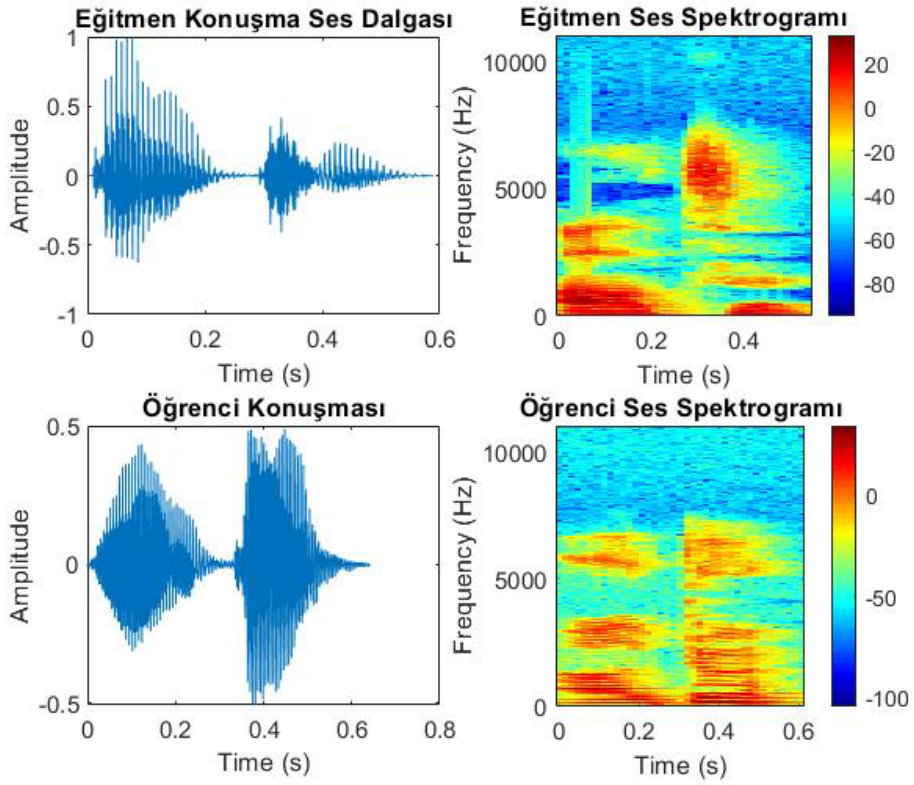
**Şekil B.6 (devam)** : İki kelimesinin eğitici ve öğrenci ses bilgisinin zamansal eşitlemesi



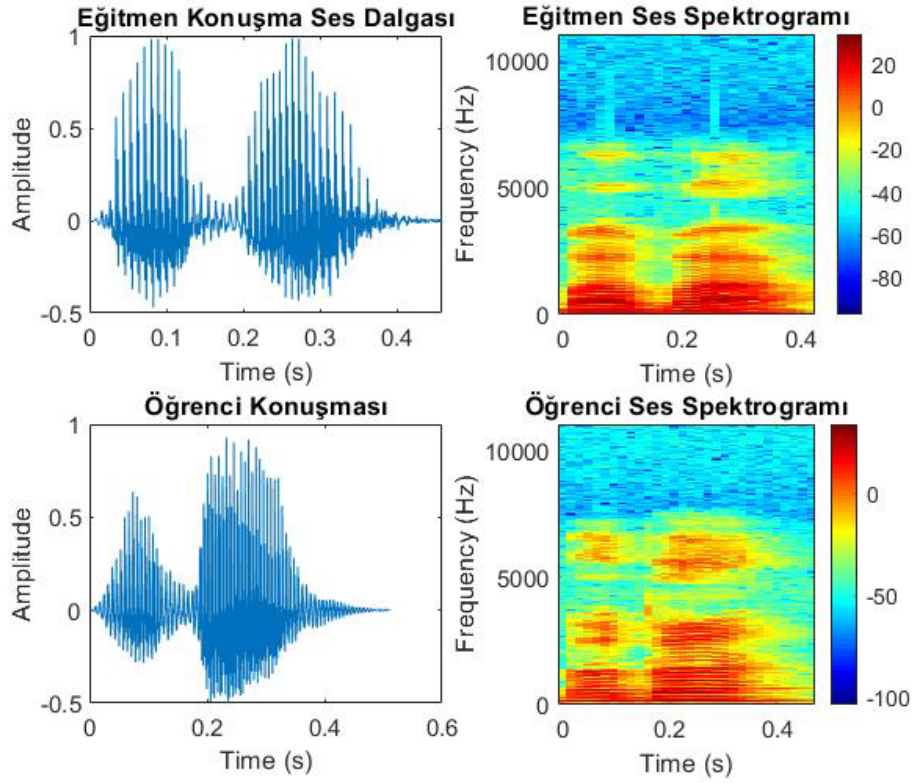
**Şekil B.7 (devam)**: Kal kelimesinin eğitici ve öğrenci ses bilgisinin zamansal eşitlemesi



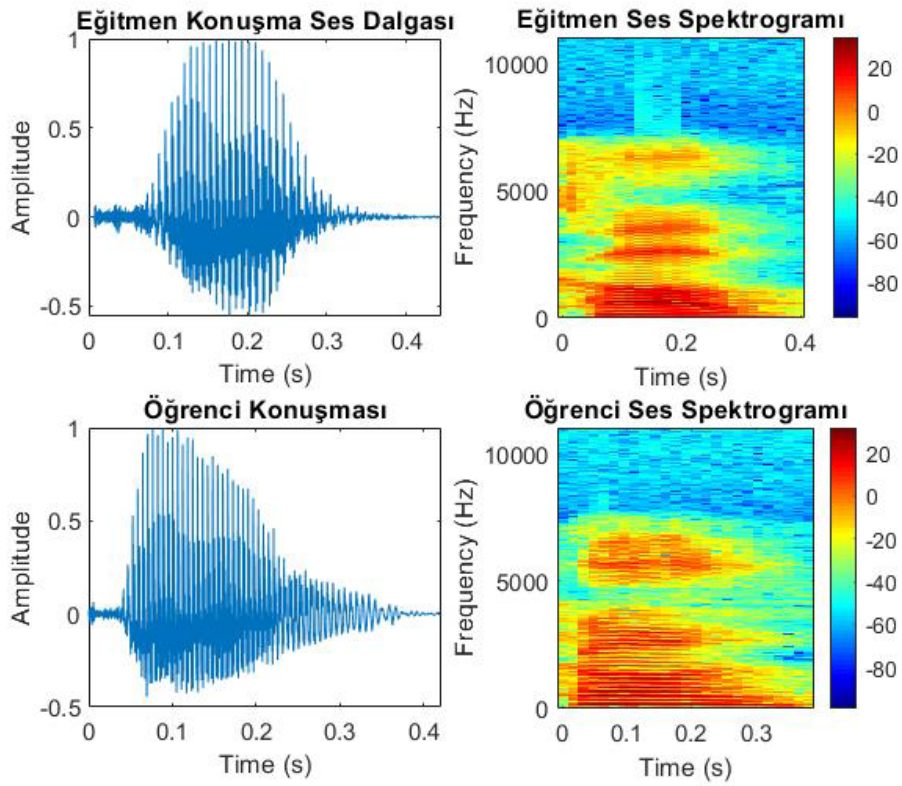
Şekil B.8 (devam) : Eğitici ve öğrenci ses bilgisi karşılaştırma



Şekil B.9 (devam): Altı Kelimesinin eğitici ve öğrenci ses spektrogramları

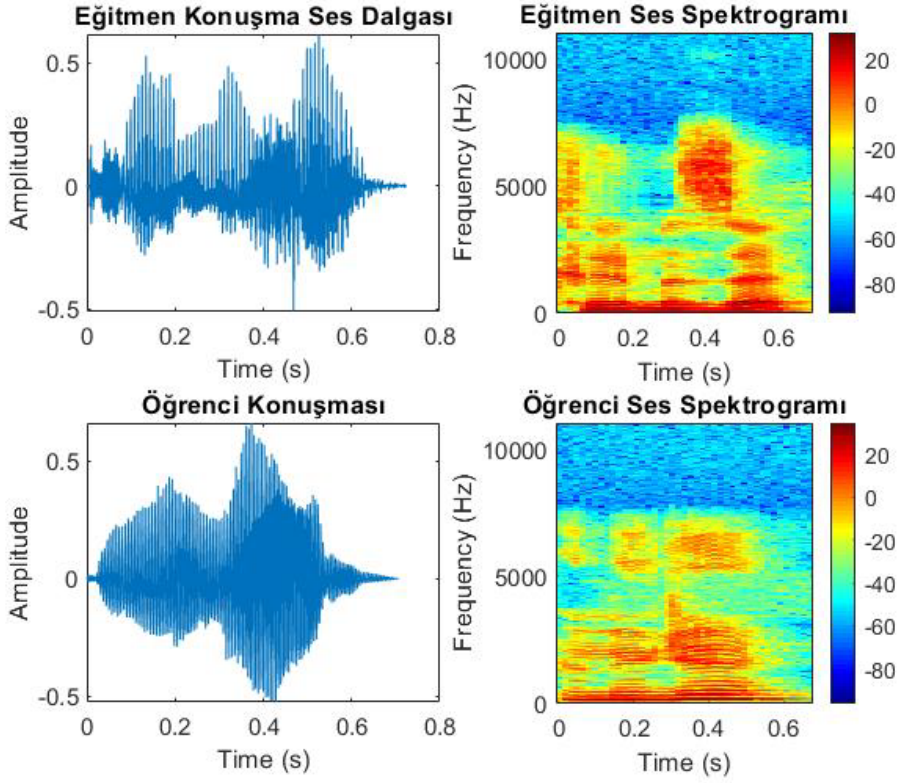


Şekil B.10 (devam) : Baba Kelimesinin eğitici ve öğrenci ses spektrogramları

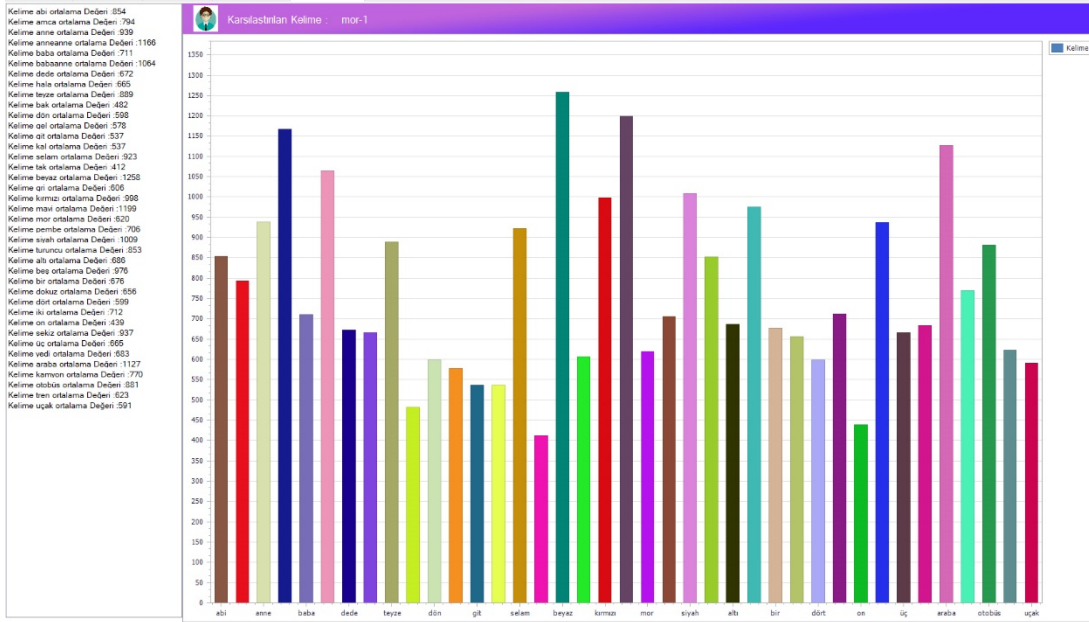


Şekil B.11 (devam) : Kal Kelimesinin eğitici ve öğrenci ses spektrogramları

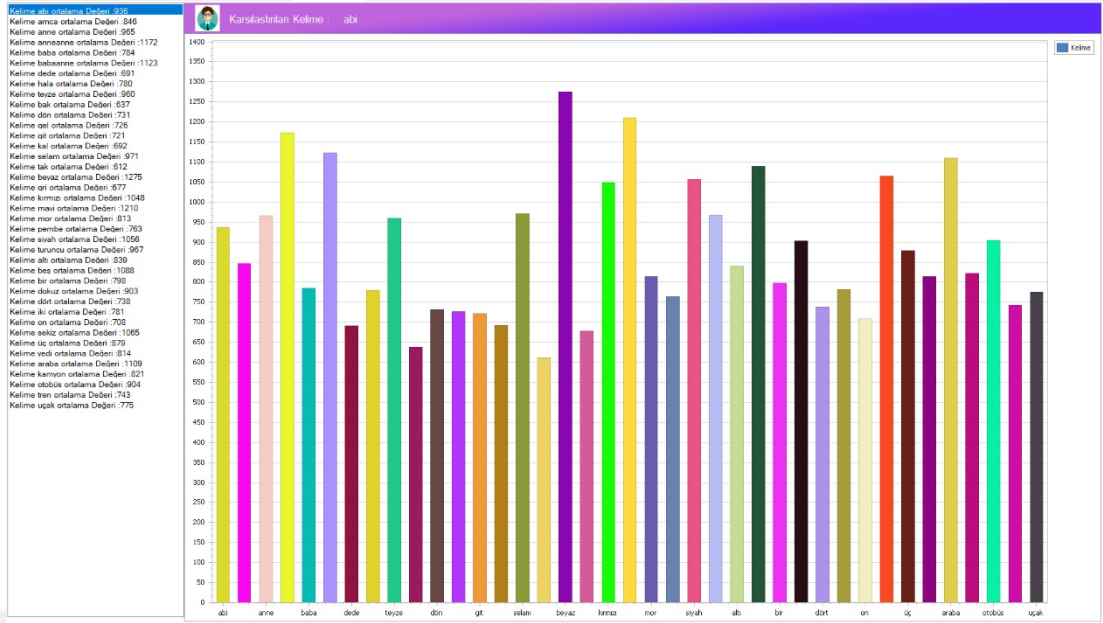




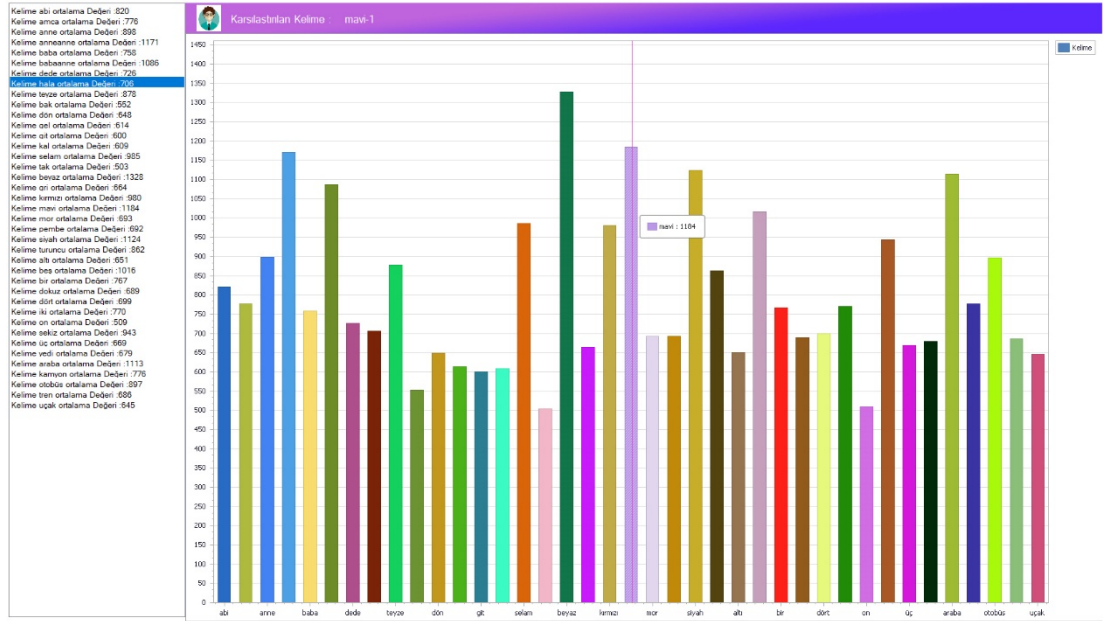
Şekil B.12 (devam) : Kırmızı Kelimesinin eğitici ve öğrenci ses spektrogramları



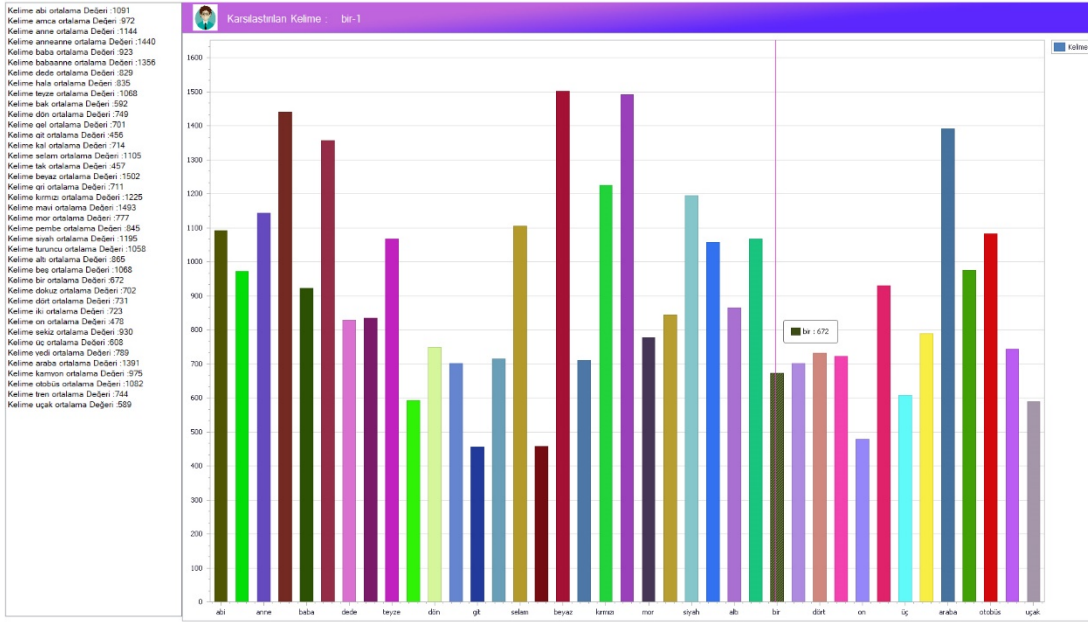
Şekil B.13 (devam) : Mor sesinin diğer sesler ile DTW mesafesi



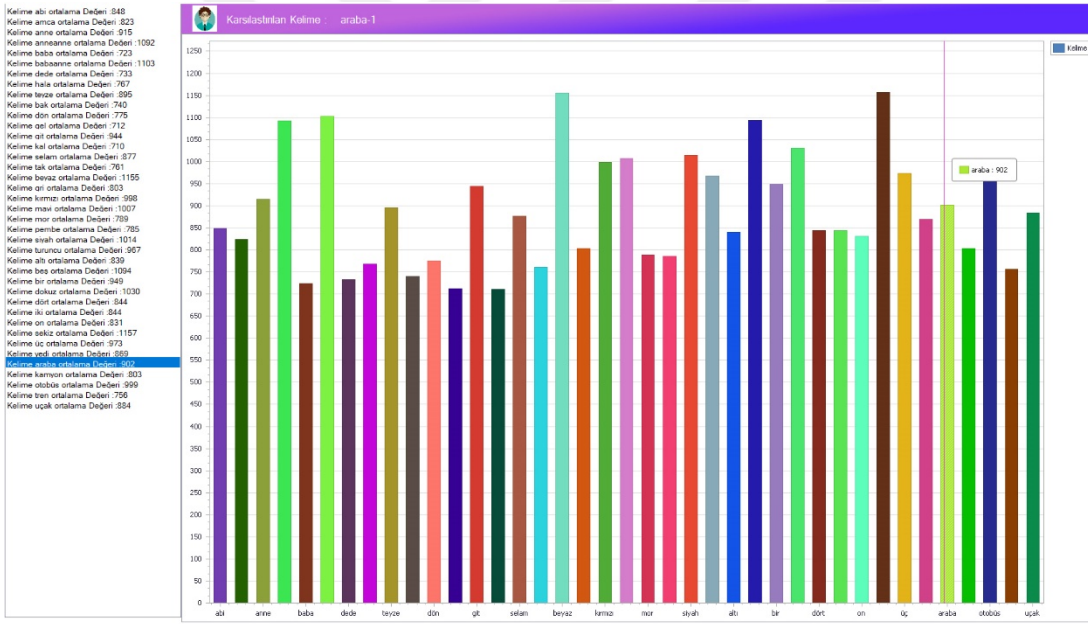
Şekil B.14 (devam) : Abi sesinin diğer sesler ile DTW mesafesi



Şekil B.15 (devam) : Mavi sesinin diğer sesler ile DTW mesafesi



Şekil B.16 (devam) : Bir sesinin diğer sesler ile DTW mesafesi



Şekil B.17 (devam) : Araba sesinin diğer sesler ile DTW mesafesi

## ÖZGEÇMİŞ



**Ad-Soyad** : Alperen KAÇAR  
**Doğum Tarihi ve Yeri** : 16/11/1984 - İzmir  
**E-posta** : nerepla@hotmail.com

### ÖĞRENİM DURUMU

**Lisans** : 10.07.2007, Fırat Üniversitesi, Teknik Eğitim Fakültesi, Bilgisayar Öğretmenliği.  
22.01.2018, Fırat Üniversitesi, Mühendislik Fakültesi Ana Bilim Dalı, Bilgisayar Mühendisliği Programı.

**Yüksek Lisans** : 2019, İstanbul Aydın Üniversitesi, Mühendislik Fakültesi Ana Bilim Dalı, Bilgisayar Mühendisliği Programı