YILDIZ TEKNİK ÜNİVERSİTESİ
FEN BİLİMLERİ ENSTİTÜSÜ

GİZLİ MARKOV MODELİ İLE GENİŞ SÖZLÜKLÜ SÜREKLİ KONUŞMA TANIMA

Erkan USLU

FBE Bilgisayar Mühendisliği Anabilim Dalında Hazırlanan

YÜKSEK LİSANS TEZİ

Tez Danışmanı: Yrd. Doç. Dr. M. Elif KARSLIGİL

İSTANBUL, 2007

YILDIZ TEKNİK ÜNİVERSİTESİ
FEN BİLİMLERİ ENSTİTÜSÜ

GİZLİ MARKOV MODELİ İLE GENİŞ SÖZLÜKLÜ SÜREKLİ KONUŞMA TANIMA

Erkan USLU

FBE Bilgisayar Mühendisliği Anabilim Dalında Hazırlanan

YÜKSEK LİSANS TEZİ

Tez Danışmanı: Yrd. Doç. Dr. M. Elif KARSLIGİL
Jüri: Yrd. Doç. Dr. Banu DİRİ
Jüri: Prof. Dr. Fikret GÜRGEN

İSTANBUL, 2007

İÇİNDEKİLER

Sayfa

SİMGE LİSTESİ iv

KISALTMA LİSTESİ vi

ŞEKİL LİSTESİ vii

ÇİZELGE LİSTESİ ix

ÖNSÖZ x

ÖZET xi

ABSTRACT xii

1. GİRİŞ 1

1.1 Akustik 2
1.2 İnsanda Konuşma, Duyma 2
1.3 Konuşma Tanıma 4
1.4 Önceki Çalışmalar 5

2. KONUŞMA TANIMADA KULLANILAN YÖNTEMLERE GENEL BAKIŞ 10

2.1 Özellik Çıkarım Yöntemleri 10
2.1.1 Sıfır Geçiş Sayısı 11
2.1.2 Enerji Seviyesi, Tepe Değeri ve Tepe Değer Geçiş Sayısı 11
2.1.3 Doğrusal Öngörülü Kodlama İle Özellik Çıkarımı 11
2.2 Özellik Sınıflandırma Yöntemleri 13
2.2.1 Dinamik Zaman Bükmesi 13
2.2.2 Yapay Sinir Ağları 15

3. FONETİK 17

3.1 Parçalı Sesbirimleri 17
3.1.1 Türkçe'de Parçalı Sesbirimleri 17
3.1.2 İngilizce'de Parçalı Sesbirimleri 18
3.2 Parçalarüstü Sesbirimleri 19

4. MEL FREKANS KEPSTRAL KATSAYI ÖZELLİKLERİ KULLANILARAK GİZLİ MARKOV MODELİ İLE GENİŞ SÖZLÜKLÜ SÜREKLİ KONUŞMA TANIMA 20

4.1 Önişlemler 21
4.1.1 Ses Verisinin Sayısallaştırılması 21

4.1.2 Ön-Vurgulama 22
4.1.3 Pencereleme 23
4.2 Mel Frekans Sepstrum Yöntemi ile Özellik Çıkarma 25
4.2.1 Spektral Analiz 26
4.2.2 Filtre Bankası Uygulama 29
4.2.3 Log Enerji 30
4.2.4 Ayrık Kosinüs Dönüşümü 31
4.2.5 Özellik Vektörünün Oluşturulması 31
4.2.6 K-Ortalama Vektör Kuantalama Yöntemi ile Özellik Azaltılması 31
4.3 Akustik Model 32
4.3.1 Markov Zincirleri 32
4.3.2 Gizli Markov Modeli 35
4.3.2.1 İleri/Geri Algoritması 36
4.3.2.2 Viterbi Algoritması 38
4.3.2.3 Baum-Welch Algoritması 39
4.3.3 Gizli Markov Modeli'nin Geniş Sözlüklü Sürekli Konuşma Tanımada Kullanımı 45
4.4 Dil Modeli 46

5. UYGULAMA 47

5.1 Ön İşlemler 47
5.2 Özellik Çıkarımı 48
5.3 Model Oluşturma ve Eğitim 51
5.4 Test ve Tanıma 55
5.4.1 İki Rakamlı Ayrık Konuşma Deney Kurgusu 56
5.4.2 Dokuz Rakamlı Ayrık Konuşma Deney Kurgusu 57
5.4.3 Kısıtlı Sayıda Tekrarlı Sürekli Konuşma Deney Kurgusu 58

6. SONUÇLAR VE ÖNERİLER 61

KAYNAKLAR 64

ÖZGEÇMİŞ 68

SİMGE LİSTESİ

$p(\vec{r},t)$	Basıncın zamana ve yere bağlı fonksiyonu
$u(\vec{r},t)$	Ortamdaki parçacıkların hızının zamana ve yere bağlı fonksiyonu
$\phi_n(i,k)$	Kovaryans fonksiyonu
$r_n(x)$	Otokorelasyon fonksiyonu
$w(n)$	Pencereleme fonksiyonu
$X(w)$	Fourier dönüşümü sonucu veya ayrık zamanlı Fourier dönüşümü sonucu
$x(t)$	Zaman boyutunda sinyal
$F(s)$	Laplace dönüşümü sonucu veya karmaşık değişkenli fonksiyon
$f(t)$	Ters Laplace dönüşümü sonucu veya zaman değişkenli fonksiyon
$x[n]$	Zaman boyutunda ayrık sinyal
$X[k]$	Ayrık Fourier dönüşümü sonucu
m	Mel frekans skalasından bir değişken
f	Frekans skalasından bir değişken
$P(A\|B)$	Koşullu olasılık
λ	Gizli Markov Modeli
S	Gizli Markov Modeli durum kümesi
A	Gizli Markov Modeli durumlar arası geçiş matrisi
B	Gizli Markov Modeli gözlem sembolleri gerçekleşme olasılık matrisi
π	Gizli Markov Modeli başlangıç durumu olma olasılık vektörü
N	Gizli Markov Modeli durum sayısı
M	Gizli Markov Modeli gözlem sembolleri sayısı
a_{ij}	i durumundan j durumuna geçiş olasılığı
$b_j(k)$	j durumunda k gözlem sembolünün gerçekleşme olasılığı
π_i	i durumun başlangıç durumu olma olasılığı
Q	Durum dizisi
O	Gözlem dizisi
q_t	t anında bulunulan durum
v_k	Gözlem kümesinin k'nıncı elemanı
o_t	t anındaki gözlem sembolü
α	İleri değişkeni
β	Geri değişkeni

$\hat{\alpha}$	Normalize edilmiş Forward değişkeni
$\hat{\beta}$	Normalize edilmiş Backward değişkeni
$\xi_t(i,j)$	t anında i'den j'ye geçiş olasılığı
$\gamma_t(i)$	t anında i durumunda bulunma olasılığı
$\delta_t(i)$	i durumunda sonlanan t adet gözlem için en yüksek gerçekleşme olasılığı
$\psi_t(j)$	En olası durum dizisi
$\bar{\lambda}$	Tahmin edilen Gizli Markov Modeli
$\bar{a}_{ij}$	Tahmin edilen geçiş olasılığı
$\bar{b}_j(k)$	Tahmin edilen gözlem sembolü gerçekleşme olasılığı
$\bar{\pi}_i$	Tahmin edilen başlangıç durumu olma olasılığı
$N(\mu,\Sigma)$	Ortalama vektörü μ, kovaryans matrisi Σ olan normal dağılım
σ	Standart sapma
c_{jk}	Katışım model katsayısı
μ_{jk}	Ortalama veya ortalama vektörü
Σ_{jk}	Kovaryans matrisi
$\bar{c}_{jk}$	Tahmin edilen katışım model katsayısı
$\bar{\mu}_{jk}$	Tahmin edilen ortalama veya ortalama vektörü
$\bar{\Sigma}_{jk}$	Tahmin edilen kovaryans matrisi

KISALTMA LİSTESİ

ANN	Artificial Neural Network (Yapay Sinir Ağları)
CMU	Carnegie Mellon University
dB	desibel
DCT	Discrete Cosine Transform (Ayrık Kosinüs Dönüşümü)
DFT	Discrete Fourier Transform (Ayrık Fourier Dönüşümü)
DTFT	Discrete Time Fourier Transform (Ayrık Zamanlı Fourier Dönüşümü)
DTW	Dynamic Time Warping (Dinamik Zaman Bükmesi)
FFT	Fast Fourier Transform (Hızlı Fourier Dönüşümü)
GMM	Gizli Markov Modeli
HMM	Hidden Markov Model (Gizli Markov Modeli)
Hz	Hertz
kHz	kilo Hertz
LPC	Linear Predictive Coding (Doğrusal Öngörülü Kodlama)
MFCC	Mel Frequency Cepstral Coefficients
MFKK	Mel Frekans Kepstral Katsayıları
ms	milisaniye

ŞEKİL LİSTESİ

Sayfa

Şekil 1.1 Konuşma ve duyma sürecine ilişkin şematik yapı 3
Şekil 1.2 Ses oluşumuna ilişkin anatomik yapı [2] 3
Şekil 1.3 Duymaya ilişkin anatomik yapı [1] 4
Şekil 1.4 Kratzenstein'in beş farklı sesli harf için kullandığı rezonans tüpleri [3] 6
Şekil 1.5 Wheatstone'un konuşma makinesi [3] 6
Şekil 2.1 Ses Tanıma sistemine ilişkin genel blok yapı (Cole, 1997) 10
Şekil 2.2 El yazısı tanıma için Dinamik Zaman Bükmesi yönteminin görsel olarak gerçekleştirilmesi (Niels, 2004) 14
Şekil 2.3 Çok katmanlı perseptron için genel gösterim 15
Şekil 4.1 Geniş Sözlüklü Konuşma Tanıma sistemi genel yapısına ilişkin blok diyagram 21
Şekil 4.2 Çözünürlüğün örneklemeye etkisi 22
Şekil 4.3 N=32 için dikdörtgen pencere fonksiyonu 23
Şekil 4.4 N=32, $\sigma = 0.4$ için Gauss pencere fonksiyonu 24
Şekil 4.5 N=32 için Hamming pencere fonksiyonu 24
Şekil 4.6 N=32 için Hann pencere fonksiyonu 24
Şekil 4.7 N=32 için Bartlett pencere fonksiyonu 25
Şekil 4.8 Özellik çıkarım adımları blok şeması 26
Şekil 4.9 Hızlı Fourier Dönüşümü adımları (Cooley, Tukey, 1965) 29
Şekil 4.10 Filtre bankası örneği 30
Şekil 4.11 a. Ergodik, b. Soldan sağa, c. Soldan sağa zorlanmış Markov model çeşitleri. 34
Şekil 4.12 İleri algoritması için yineleme adımı şematik gösterimi. (Rabiner, 1989) 37
Şekil 4.13 Geri algoritması için yineleme adımı şematik gösterimi. (Rabiner, 1989) 38
Şekil 4.14 a. Hatalar karesi, b. Hataların mutlak değeri, c. Düzgün maliyet fonksiyonu (Barkat, 2005) 40
Şekil 4.15 $\gamma_t(i)$ değişkeni için şematik gösterim. (Rabiner, 1989) 41
Şekil 4.16 $\xi_t(i, j)$ değişkeni için şematik gösterim. (Rabiner, 1989) 42
Şekil 4.17 Tek boyutlu iki katışımlı normal dağılım örneği 44
Şekil 4.18 GMM eğitim ve test aşamaları blok yapısı 46
Şekil 5.1 a. Ön-vurgulama ve normalizasyon öncesi, b. sonrası ses verisi 47
Şekil 5.2 a. Bir pencere genişliğinde ses verisi, b. Uygun boyutlu Hamming pencere fonksiyonu, c. Ses verisinin pencere fonksiyonundan geçirilmiş hali, d. İlgili

pencere için hesaplanmış MFKK katsayıları 49
Şekil 5.3 a. Ses verisi, b. Ses verisine ilişkin hesaplanmış MFKK katsayıları..... 50
Şekil 5.4 İki boyutlu bir uzayda Öklid uzaklığı..... 50
Şekil 5.5 Fon veya fonem için temel GMM model parametreleri ilk durum ataması..... 51
Şekil 5.6 Bir eğitim modeli için kelime bazında ve fon bazında çözümleme verisi 52
Şekil 5.7 İlk koşulları verilmiş genel GMM için a. A matrisi, b. B matrisi 53
Şekil 5.8 Sıfırdan dokuza kadar tekrarlı rakamlar için dil modeli..... 55
Şekil 5.9 İki rakam ayrık telaffuzu için test modeli 56
Şekil 5.10 Birden dokuza kadar olan rakamların ayrık telaffuzu için test modeli 57
Şekil 5.11 İki rakamlı sürekli konuşma tanıma test modeli 59
Şekil 5.12 Cümle test modeli..... 60

ÇİZELGE LİSTESİ

Sayfa

Çizelge 3.1 İngilizce'de kullanılan sesli fonemlerin sınıflandırılması [4] 18
Çizelge 3.2 İngilizce'de kullanılan yarım sesli fonemler [4] 18
Çizelge 3.3 İngilizce'de kullanılan sessiz fonemlerin sınıflandırılması [4] 18
Çizelge 3.4 Türkçe için hece yapısı ve ilgili yapılara ilişkin örnekler 19
Çizelge 5.1 Gözlem sembolü gerçekleşme olasılığı 54
Çizelge 5.2 20 adet test örneğinin iki rakamlı ayrık rakam tanıma modeli ile sınıflandırma sonuçları 56
Çizelge 5.3 İki rakamlı test modeli için k-katlı çapraz geçerlilik sınaması sonuçları 57
Çizelge 5.4 45 adet test örneğinin dokuz rakamlı ayrık rakam tanıma modeli ile sınıflandırma sonuçları 58
Çizelge 5.5 İki rakamlı süreli konuşma tanıma sonuçları 59
Çizelge 5.6 Sürekli konuşma cümle test model sonuçları 60

ÖNSÖZ

Danışmanım Sayın Yrd. Doç. Dr. M. Elif Karslıgil'e tez çalışmam boyunca beni yönlendirdiği, bilgisini, değerli görüşlerini ve deneyimlerini bana aktardığı için teşekkür ederim.

Yıldız Teknik Üniversitesi, Donanım Anabilimdalı öğretim üyeleri başta olmak üzere YTÜ Bilgisayar Mühendisliği bölümündeki bütün arkadaşlarıma bana desteklerinden ve katkılarından dolayı teşekkür ederim.

Lisansüstü eğitim sürecinde beraber ilerlediğimiz Selim Nasır ve İlktan Ar'a bana olan destek ve katkılarından dolayı teşekkür ederim.

Sevgili anneme, sevgili babama ve sevgili kardeşime, bana göstermiş oldukları sabır ve fedakarlıklardan dolayı çok teşekkür ederim.

Erkan Uslu
Mayıs, 2007

ÖZET

Bilgisayar bilimlerinde ses ile ilgili çalışmalar genel olarak üç ana başlık altında toplanabilir: bunlar konuşma tanıma, konuşmacı tanıma-doğrulama ve konuşma sentezlemedir. Ana başlıklardan konuşma tanıma ile makine-bilgisayar tarafından insan konuşmasının anlaşılması veya bundan bilgi çıkarımı hedeflenmektedir.

Bu tez çalışmasında çok boyutlu bir tanım uzayına sahip olan konuşma tanıma probleminin geniş sözlüklü sürekli konuşma tanıma gereksinimlerini karşılayacak şekilde çözümüne yönelik bir uygulama gerçekleştirilmiştir. Mel Frekans Kepstral Katsayı (MFKK) özellik çıkarımı ve Gizli Markov Modeli (GMM) ile özellik sınıflandırılması bu çalışmanın temel adımlarını oluşturmaktadır.

MFKK özellikleri konuşmacı bağımsız olarak ses verisine ilişkin özelliklerin ortaya konmasında etkili bir yöntemdir. GMM de ardışıl özelliklerin geliş sırası dikkate alınarak sınıflandırılmasını sağlayan bir yöntemdir.

MFKK özellik çıkarım adımlarının ses verisine uygulanması ile reel değerlikli özellik vektörleri elde edilir. MFKK katsayılarına K-ortalama yönteminin uygulanması ile tek boyutlu ayrık değerlikli bir özellik uzayına geçilir.

Herbir fon için dört durumlu, soldan sağa, ayrık çıkış olasılıklarına sahip GMM temel modelleri uygun ilk durumları verilerek oluşturulur. Eğitimde kullanılan ses verisine ilişkin çözümleme doğrultusunda temel GMM'ler bir araya getirilerek sesler arası geçiş sayılarına göre olasılıklandırılırlar. Tüm eğitim seti üzerinde Baum-Welch algoritması çoklu gözlem durumu dikkate alınarak uygulanır ve tüm temel GMM'ler için model parametreleri güncellenir. Kullanılan yaklaşım ile GMM modelinin eğitim aşamasında ses üzerinde etiketleme, bölümleme, kelime başlangıcı ve bitişi işaretleme gereği olmadan ses verisine ilişkin istatistiki yapı elde edilebilmektedir.

Sistem başarısı iki farklı kelimeyi ayrık olarak tanıma, çok sayıda kelimeyi ayrık olarak tanıma, kısıtlı sayıda tekrarlı kelimeleri sürekli konuşma yapısında tanıma, sürekli konuşma yapısında cümle tanıma deneysel kurguları üzerinde incelenmiştir.

Anahtar Kelimeler: Geniş sözlüklü sürekli konuşma tanıma, Gizli Markov Modeli, Mel Frekans Kepstral Katsayıları.

ABSTRACT

In computer sciences researches on speech can be viewed in three major fields, which are speech recognition, speaker recognition-verification and speech synthesis. The aim of speech recognition, as a major field, is information extraction from speech data by machine or computer.

With this thesis an application for large vocabulary continuous speech recognition requirements, which is a portion of multi dimensional speech recognition problem space, is built. On this purpose mel frequency cepstral coefficients (MFCC) feature extraction is used and feature vectors are classified by hidden Markov models (HMM).

MFCC is an effective feature extraction method for speaker independent recognition and HMM can handle sequential feature vectors.

Real valued MFCC feature vectors are firstly clustered into discrete observations with K-Means algorithm. Discrete observations are then used with left to right, discrete probability emitting HMM models.

Statistical nature of speech data can be figured out by the HMM model with no need to labeling, segmentation or signing word start and endings.

As HMM training is made on phonemes, test speech data can be decoded with flexibly built test HMM model.

System is tested on several models such as, two word discrete recognition, nine words discrete recognition, repetition constraint continuous recognition.

Keywords: Large vocabulary continuous speech recognition, hidden Markov model, mel frequency cepstral coefficients.

1. GİRİŞ

Teknolojinin gelişiminde canlıların sahip oldukları özelliklerin insan yapımı araç ve gereçlere aktarılması isteği önemli yer tutmaktadır. Canlıların bünyesinde var olan özelliklerin bir kısmı bir işin yapılmasına ilişkin en verimli yöntem olduğu için, bir kısmı ise insan yapımı daha hızlı, daha verimli ve daha düşük hata paylarıyla çalışan sistemler geliştirilmiş olmasına rağmen canlı açısından anlaşılabilir tek özellik olması açısından teknolojik olarak kendine yer bulabilmektedir. Çok daha hızlı haberleşme metotlarının varolmasına rağmen konuşma ve duyma da insan açısından anlaşılabilir en önemli iletişim aracı olması bakımından bahsedilen ikinci tip özellikler arasında sınıflandırılabilir ve insan-bilgisayar arayüzü olarak aranan bir özellik konumundadır.

Ses ile ilgili ilk çalışmalar öncelikle sesin fiziki yapısını anlamaya yöneliktir. Sonraları biyolojik olarak insanda ses oluşumu ve duymanın fizyolojisi anlaşılmaya çalışılmıştır.

Ses oluşumunun yapısının ortaya konması ile yapay konuşma kaynakları modellenmiş ve gerçekleştirilmiştir. Gerçekleştirilen bu sistemler teknolojik gelişimi izleyerek sırayla mekanik, elektro-mekanik ve elektronik sistemler olarak ortaya çıkmıştır.

Günümüzde konuşma sentezleme olarak kendine yer bulan sistemler kişisel bilgisayarlarda da birtakım kısıtlar çerçevesinde gerçekleştirilebilen uygulamalar olarak kullanılabilmektedir.

Sinyal işleme ve bilgisayar bilimlerindeki gelişmelere paralel olarak konuşma tanımaya yönelik yöntem ve metotlar ortaya konmuştur. Konuşma tanımada matematiksel alt yapının gelişimi ile zaman boyutunda şablon karşılaştırma yöntemlerinden, stokastik hibrit sistemlere uzanan bir gelişim ortaya çıkmıştır. Günümüzde konuşma tanıma; otomatik çeviri, otomobil içi konuşma tanıma, komut tanıma, insan-bilgisayar arayüzü, ev otomasyonu, robotik uygulamaları, bilgisayar temelli dil öğreniminde telaffuz geliştirme, otomatik veri girişi ve benzeri konularda uygulama alanı bulmaktadır.

Konuşma tanıma yöntemlerinin gelişimi ile biyometrik ölçütlerden biri olarak da konuşmacı tanıma ve doğrulama sistemleri geliştirilmiştir. Konuşmacı tanıma ve doğrulama sistemleri güvenlik uygulamalarında geniş bir kullanım alanı bulmaktadır.

1.1 Akustik

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

Sesin bir ortamda yayılması ile ortama ilişkin basınç, yoğunluk, ortamdaki parçacıkların hızı ve sıcaklık değerlerinde uzamsal ve zamansal değişiklikler gerçekleşir. Bu değişimlerin sağlaması gereken süreklilik ve sınır koşulları göz önüne alınarak, $p(\vec{r},t)$ basıncın zaman ve yere göre değişimini, $u(\vec{r},t)$ parçacık hızının zamana ve yere göre değişimini göstermek üzere yoğunluğu ρ olan bir ortamda düzlemsel bir ses dalgası denklem (1.1) ile verilen koşulu sağlaması gerektiği söylenebilir. (Juang, Rabiner, 2005)

$$-\frac{\partial p}{\partial x} = \rho \frac{\partial u}{\partial t} \tag{1.1}$$

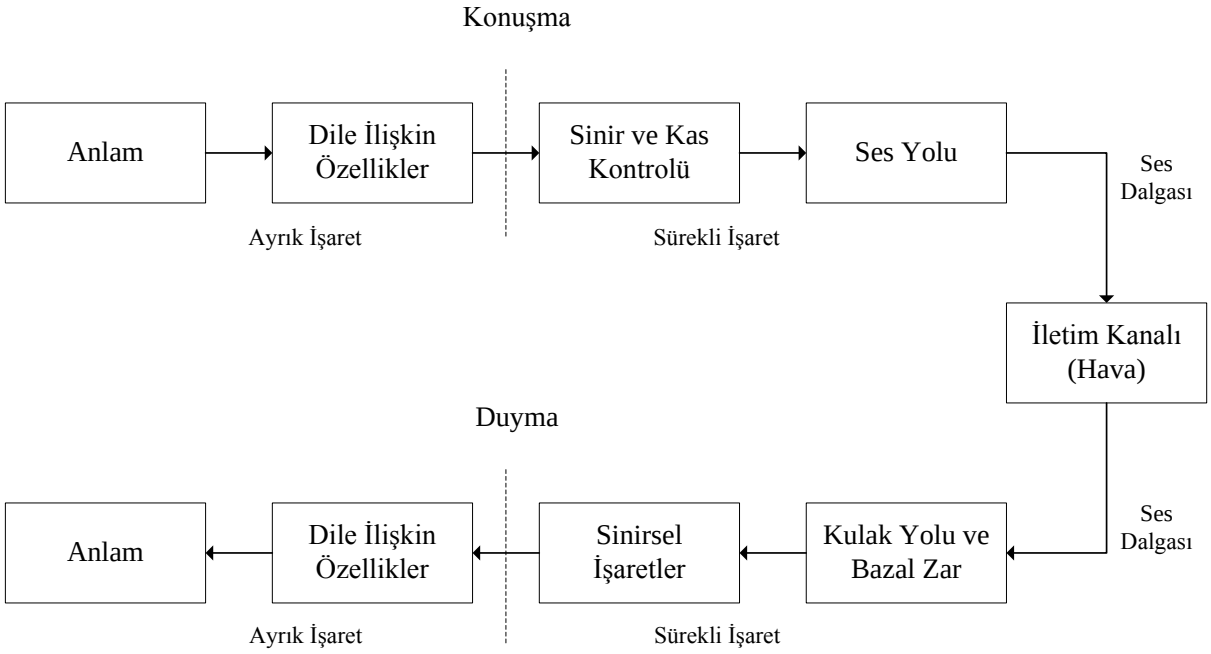
Ses dalgasının yayıldığı ortamın düzgün olmayan bir tüp olarak modellenmesi ile süreklilik ve sınır koşulları yardımıyla insanda ses yoluna ilişkin rezonans frekansları, formant değerleri, anti-rezonans değerleri gibi fiziki özellikler hesaplanabilmektedir. Hesaplanan bu değerler ses tanımada akustik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

1.2 İnsanda Konuşma, Duyma

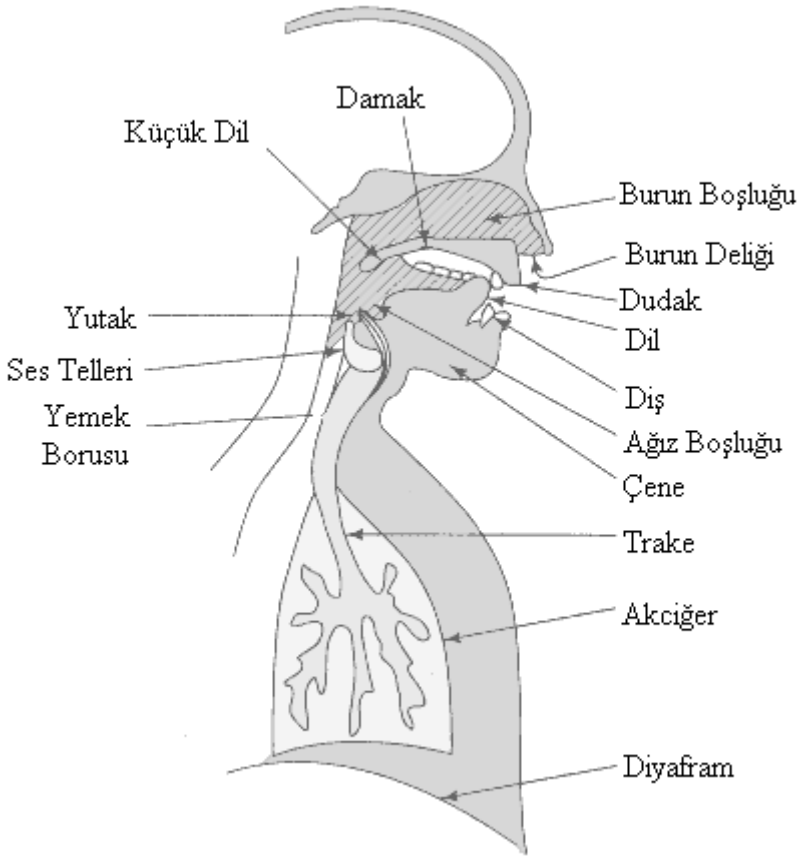
Konuşma tanıma ve konuşma sentezinde insan konuşması ve duyması model olarak alındığı için insanda konuşmanın ve duymanın anlaşılması önemli ve gereklidir.

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

Ses oluşumu mekanik olarak akciğer, soluk borusu, ses telleri, gırtlak, çene, dil, diş, küçük dil, damak, burun boşluğu ve dudak gibi organların etkisiyle sağlanmaktadır (Şekil 1.2). Bu yapı bir sistem yapısı ile modellenerek konuşmaya ilişkin akustik model elde edilmiştir (Rabiner, Juang, 1993).



Şekil 1.1 Konuşma ve duyma sürecine ilişkin şematik yapı



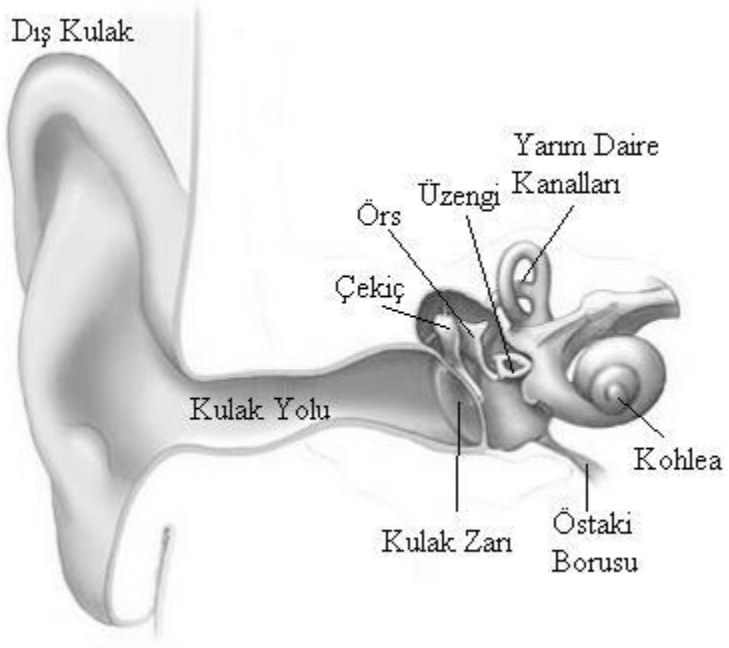
Şekil 1.2 Ses oluşumuna ilişkin anatomik yapı **[2]**

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

$$P_r(z) = S(z)\ T(z)\ R(z) \tag{1.2}$$

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

Duyma ise ses üretiminde varolan transfer fonksiyonlarının tersleri ile oluşturulmuş bir sistem olarak düşünülebilir. Hava yolu iletilen ses dalgaları dış kulak yardımıyla kulak yoluna iletilir. Kulak zarında oluşan titreşimler çekiç, örs ve üzengi kemikleri ile duyma organı kohleaya ulaştırılır. Kohleada bulunan bazilar zar üzerindeki kılcal yapılar ve lif yoğunluğunun kohlea boyunca giderek artan yapı göstermesi, kohleanın bir ucunda düşük frekansa, diğer ucunda yüksek frekansa duyarlılığı sağlar. Bazilar zar üzerindeki kılcal korti reseptör hücreleri ile sinirsel işarete dönüştürülen veri duyma siniri ile beyne iletilir.



Şekil 1.3 Duymaya ilişkin anatomik yapı [**1**]

1.3 Konuşma Tanıma

Konuşma tanıma günlük hayatta pek çok ortam ve uygulamada kendine yer bulmuş bir konudur. Genel olarak konuşma tanıma ile insan konuşmasının bilgisayar tarafından

anlaşılması ve buradan bilgi çıkarımı hedeflenmektedir.

Konuşma tanıma problemi birçok farklı boyut ile tanımlanabilmektedir. Konuşma tanıma üzerine yapılan çalışmalar artık günümüzde bahsedilen çok boyutlu problem uzayının belirli bölgelerinde özelleşmektedir. Problem uzayı aşağıda bahsedilen boyutlar ve uç durumlar ile tanımlanabilmektedir:(Cole, 1997)

Konuşma modu: ayrık kelime tanıma ↔ sürekli konuşma tanıma

Tanıma birimi: sözcük tabanlı ↔ fonem tabanlı

Konuşma stili: okuma ↔ spontane konuşma

Konuşmacıya bağlılık: konuşmacı bağımsız tanıma ↔ konuşmacı bağımlı tanıma

Sözlük genişliği: 20 kelime ↔ 20000 kelime

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

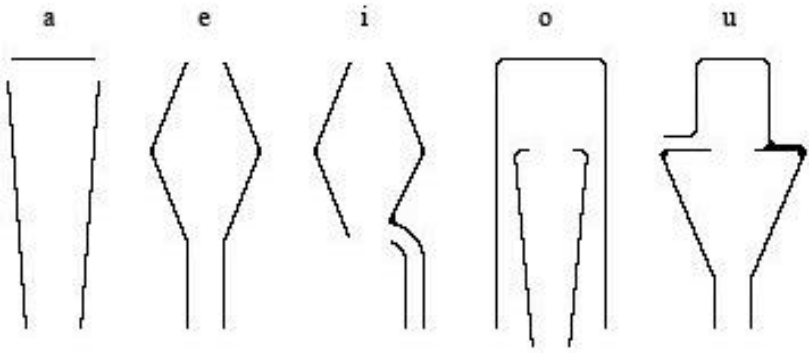
Sinyal için gürültü oranı: 10 dB ↔ 25 dB

Kullanılan alıcı: gürültü önleyen mikrofon ↔ cep telefonu

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

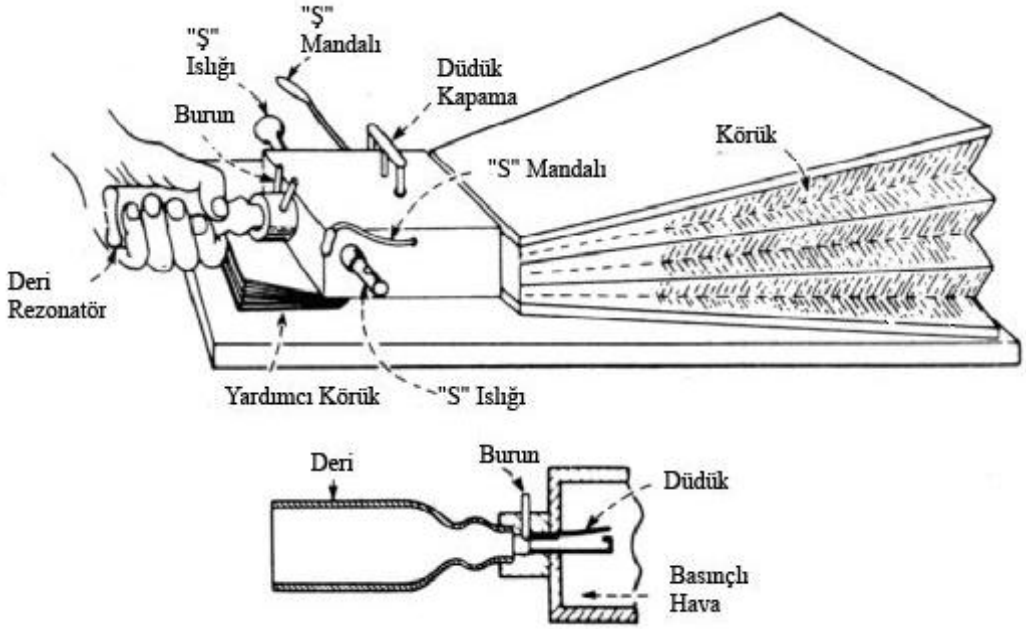
1.4 Önceki Çalışmalar

Ses ile ilgili bilimsel çalışmalar tarihsel sırayla ses sentezleme, ses sinyali işleme ve konuşma tanıma üzerine yoğunlaşmıştır. Ses sentezleme için sırasıyla mekanik, elektro-mekanik ve elektronik sistemler kullanılmıştır. Ses sinyali işleme tarihsel sırasıyla zaman boyutunda, frekans boyutunda ve kepstral özelliklerin elde edilmesi ile ilgilenmiştir. Konuşma tanıma konusunda ise tarihsel olarak şablon karşılaştırma, örüntü tanıma ve istatistiki modellerle ilgilenilmiştir.



Şekil 1.4 Kratzenstein'in beş farklı sesli harf için kullandığı rezonans tüpleri **[3]**

1770'li yıllarda Christian Kratzenstein (Juang, Rabiner, 2005) rezonans tüpleri ile sesli harf sentezlemeyi başarmıştır. 1791'de Wolfgang von Kempelen (Juang, Rabiner, 2005) mekanik kontrollü akustik-mekanik konuşma makinesini üretmiştir. 1800'lerin ortalarında Charles Wheatstone (Juang, Rabiner, 2005) deri rezonatörlü konuşma makinesini yapmıştır. Charles Wheatstone yaptığı bu aletle istediği sesli harfi el kontrolü ile deri rezonatörden sağlayabilmekteydi. 1939'da Homer Dudley elektrik mekanik konuşma makinesini yapmıştır.



Şekil 1.5 Wheatstone'un konuşma makinesi **[3]**

Konuşma tanımaya ilişkin ilk çalışmalar akustik ve fonetik özelliklere dayanmaktaydı. Akustik-fonetik özelliklerden biri de sesin oluşmasında ses tellerinin ses yayılımına olanak sağlayan titreşim frekansları olan formantlardır.

1952'de formant frekanslarından hareketle tek konuşmacı için ayrık rakam tanıma sistemi

Davis, Biddulph ve Balashek tarafından Bell Laboratuvarları'nda gerçekleştirilmiştir, bu sistemde rakamlar, sesli harflerinin rezonans frekanslarından faydalanılarak tanınmaktaydı. 1956 yılında Olson ve Belar on hece için tek konuşmacılı tanıma sistemini geliştirmiştir. 1959'da MIT'den Forgie ve Forgie konuşmacı bağımsız olarak b_<sesli harf>_t formatındaki farklı sesli harflerle oluşturulmuş on farklı kelimenin tanınmasına ilişkin bir sistem gerçekleştirmiştir. 1960'lı yıllarda Japon bilim adamları (Suzuki ve Nakata, 1961) sesli harf tanımayı, Sakai ve Doshita (1962) sürekli konuşma tanıma için öncü bir örnek olan ses tanımayı gerçekleştirmiştir. Ayrıca NEC Laboratuarı (Nagata, Kato, Chiba, 1963) rakam tanımaya yönelik çalışmalar gerçekleştirmiştir.

1959'da Fry ve Denes dört sesli dokuz sessiz harf ile İngilizce için olası ses sıralarını da hesaba katarak konuşma tanımada istatistiki yapı kullanımının ilk örneğini oluşturmuştur. 1964'te RCA Laboratuarları'ndan Tom Martin konuşmada sözler için bitiş noktası belirleme ile ilgili çalışmalarda bulunmuştur. Ayrıca Tom Martin (1964) ve Vintsyuk (1968) eş zamanlı olarak yaptıkları çalışmalar ile dinamik programlama yardımıyla ses sinyalinin zaman yayılımından kaynaklanan farklılıkları gidererek ses sinyallerindeki benzerlikleri ortaya çıkarmışlardır. 1978'de Sakoe ve Chiba konuşma tanıma için dinamik zaman bükümü ile örüntü tanıma gerçekleştirmiştir.

1972'de H. Fletcher ses sinyali için spektral özellikler ile ses karakteristikleri arasındaki yakın ilişkiyi ortaya koymuştur. Bu çalışmadan sonra ses tanımada spektral özelliklerin kullanımı yaygınlaşmaya başlamıştır.

1970'lerde Doğrusal Öngörülü Kodlama (Linear Predictive Coding) yöntemi ortaya konmuştur. Atal B. (Atal, Hanauer, 1971) ve Itakura (Itakura, Saito, 1970) ses sinyali için dalga yapısından hareketle ses yolu için transfer fonksiyonu parametrelerinin tahminini gerçekleştiren LPC yönteminin temellerini ortaya koymuştur. Itakura (1975) ile Rabiner ve Levinson (1979) LPC yönteminin örüntü tanıma ile birlikte kullanımı için çalışmalar gerçekleştirmiştir.

Konuşma tanıma ile ilgili olarak ilk ticari ürün Tom Martin'in kurduğu Threshold Technology Inc. tarafından gerçekleştirilen VIP-100 sistemidir. VIP-100 yüksek başarı ve geniş bir pazar bulamamış olmasına rağmen 1970'lerde Amerika Savunma Bakanlığı'nın desteğinde konuşma tanıma biriminin kurulmasında etkili olmuştur.

Benzer çalışmalar ve bunların ortaya koyduğu yeni fikirler açısından incelendiğinde Lowerre'in (1990) kullanımını önerdiği sonlu durum makineleri ve çizge arama (graph

search), CMU tarafından Hearsay(II)'de kullanılan asenkron paralel süreçler ve IBM'in ses kontrollü daktilosu Tangora'da kullandığı dil modeli gibi yöntemlerin sonraki çalışmalara öncülük etmiş olduğu görülmektedir.

1980'lerde konuşma tanımada şablon karşılaştırma ve örüntü tanıma temelli yöntemlerden istatistik modelli tanıma sistemlerine geçiş görülmektedir. 1972'de Leonard E. Baum Markov zincirlerinin bir türevi olarak iki aşamalı olasılık dağılımları birbirine bağlayan Gizli Markov Modeli'ni ortaya koymuştur. Aynı çalışmasında Leonard E. Baum Gizli Markov Modeli'ne ilişkin en uygun parametrelerin elde edilme yönteminden bahsetmiştir. Gizli Markov Modeli temelli sistemlerin yeniden gündeme gelmesi ve konuşma tanımada kullanılması ise1980'lerde J. D. Ferguson (1980) ile S. E. Levinson, L. R. Rabiner ve M. M. Sondhi'nin (1983) çalışmaları ile olmuştur.

Gizli Markov Modeli'nin ilk önceleri ayrık değerlikli değişkenler için çalıştığı ortaya konmuştur. Daha sonraları Gizli Markov Modeli için içbükey olasılık dağılımlarının kullanılabilir olduğu gösterilmiştir. Liporace (1982) ise Gauss olasılık dağılımı gibi simetrik eliptik dağılımların da Gizli Markov Modeli ile kullanılabilir olduğunu göstermiştir. 1985 ve 1986 yıllarında Bell Laboratuarları'ndan yapılan çalışmalar ile (Juang, Levinson, Sondhi, 1986) (Juang, 1985) Gizli Markov Modeli için olasılık katışım modellerinin kullanılabilirliği gösterilmiştir. Böylelikle Gizli Markov Modeli'nin kullanım alanı konuşmacı tanımlanması ve konuşmacıdan bağımsız sürekli konuşma tanıma görevlerini yapabilecek düzeyde genişletilmiş oldu.

Konuşma tanımada kullanılan yöntemlerden bir diğeri yapay sinir ağları olmuştur. İlk olarak McCullough ve Pitts tarafından 1943'te ortaya konan yapay sinir ağları, 1980'lerde hata geri yayılım (error back-propagation) metodu bulunana kadar geniş bir kullanım alanı bulamamıştır. Hata geri yayılım metodunun ortaya konması ile yapay sinir ağlarının fonem tanıma, ayrık rakam tanıma gibi işlerde başarılı sonuç verdiği gösterilmiştir (Lippmann, 1990).

1990'larda minimum sınıflandırma hatası kavramının benimsenmesi ile ayırıcı eğitim (discriminative training) yöntemi ve Destek Vektör Makinesi (Support Vector Machine) gibi çekirdek tabanlı (kernel based) metodlar konuşma tanımada kullanılmaya başlandı (Juang, Lee, Wu Chou, 1997) (Bahl, Brown, deSouza, Mercer, 1986).

1996'da Hu ve arkadaşları dil için temel birim olarak gördükleri hecelerden ve hece benzeri ses birimlerinden hareketle İngilizce ay isimleri için 29 heceden oluşan tanıma sistemleri ile

%84'lük başarı oranı yakalamışlardır. Aynı yıl Boulard bezer bir çalışmayı Almanca için gerçekleştirmiştir.

Hauenstein Gizli Markov Modeli ve yapay sinir ağlarının birlikte kullanıldığı bir sistem tasarlayarak bu yapının sadece bir yöntemin kullanıldığı diğer tasarımlarından daha iyi sonuç verdiğini göstermiştir (1996)

Wu ve arkadaşları tarafından yürütülen bir çalışma ile hece ve fon veya fonem bazında hibrit sitemlerinin tanıma başarısını arttırdığı gösterilmiştir.(1997)

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

Günümüzde ticari veya araştırma amaçlı geliştirilen konuşma tanıma sistemleri için elde edilen hata oranları: rakam tanıma için %0.3, 1000 kelimelik sözlüklü okuma sesi için %3, 20,000 kelimelik sözlüklü okuma sesi için %6, 10,000 kelimelik sözlüklü karşılıklı konuşma için %20, telefon üzerinden 10,000 kelimelik sözlüklü konuşma için %30 seviyelerinde gerçekleşmektedir.

Bu çalışmanın amacı fonem temelli bir konuşma tanıma sistemi tasarlamaktır. Bu amaçla mel frekans kepstral katsayı özellikleri üzerinden Gizli Markov Modeli ile sınıflandırma gerçekleştirilmektedir.

Tez çalışmasında 2. bölümde genel olarak konuşma tanıma sistemlerinde Gizli Markov Modeli dışındaki sınıflayıcılar ile mel frekans kepstral katsayı özellikleri dışında kullanılan özellik çıkarım yöntemlerinden bahsedilmiştir.

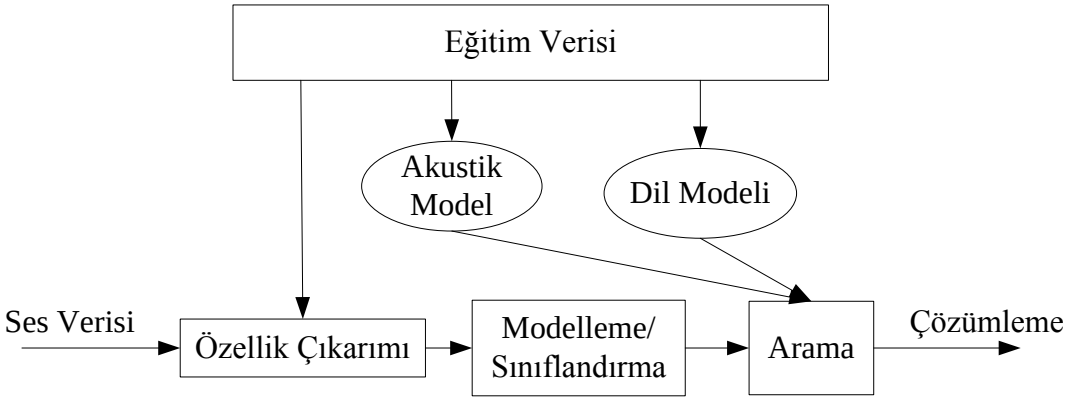
3. bölümde Türkçe ve İngilizce için konuşma tanımada kullanılan parçalı sesbirimleri ile genel olarak parçalarüstü sesbirimlerinden bahsedilmiştir.

4. bölümde Gizli Markov Modeli kullanarak mel frekans kepstral katsayı özellikleri üzerinden konuşma tanımanın gerçekleştirilmesine ilişkin teorik yapı; önişlemler, özellik çıkarımı, akustik model ve dil modeli başlıkları altında anlatılmıştır.

5. bölümde uygulamanın gerçekleştirilmesinden bahsedilerek 6. bölümünde elde edilen sonuçlar değerlendirilmiştir.

2. KONUŞMA TANIMADA KULLANILAN YÖNTEMLERE GENEL BAKIŞ

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



Şekil 2.1 Ses Tanıma sistemine ilişkin genel blok yapı (Cole, 1997)

Özellik çıkarım yöntemleri olarak Fourier analizi, kepstral katsayıların hesaplanması, kepstral fark katsayılarının hesaplanması, sinyal enerji seviyesi hesaplanması, sıfır geçiş sayısı hesaplanması, tepe değer hesaplanması, tepe değer gerçekleşme sayısı belirlenmesi, otokorelasyon katsayılarının hesaplanması, Doğrusal Öngörülü Kodlama (Linear Predictive Coding) gibi metotlardan bahsedilebilir.

Özellik vektörlerinin sınıflandırılmasında Gizli Markov Modeli (Hidden Markov Model), Dinamik Zaman Bükmesi (Dynamic Time Warping) ve Yapay Sinir Ağları (Artificial Neural Networks) yöntemlerinin kullanımı söz konusu olmaktadır.

Konuşma tanımada kullanılan yöntemlerden bu çalışmanın konusu olan Mel Frekans Kepstral Katsayıları (MFKK) özelliklerinin Gizli Markov Modeli (GMM) ile sınıflandırılması detaylı olarak dördüncü bölümde anlatılacaktır.

2.1 Özellik Çıkarım Yöntemleri

Ses verisine ilişkin özellikler zaman, frekans veya sepstrum boyutunda tanımlanabilmektedir. Zaman boyutunda hesaplanabilen özellikler sıfır geçiş sayısı, enerji seviyesi, tepe değeri, tepe değer geçiş sayısı gibi özelliklerdir. Ses verisinin frekans yapısından hareketle filtre bankası uygulama ve frekans spektrumu hesaplama yöntemleri uygulanabilir. Sepstrum boyutu

spektrumun spektrumu olarak tanımlanabilir ve ses verisine ilişkin doğrusal ya da logaritmik frekans ölçeğine göre hesaplanabilir. İstatiski bir metot olan doğru uydurma yöntemini temel alan Doğrusal Öngörülü Kodlama ve bu metot ile hesaplanan özelliklerin sepstrum değerleri de ses verisine ilişkin özellikler olarak kullanılabilmektedir (Zhou, L., Fang, D. 1988).

2.1.1 Sıfır Geçiş Sayısı

Ses verisine ilişkin iki örnek arasında işaret değişimi sıfır geçişi olarak tanımlanmaktadır (Zhou, L., Fang, D. 1988). Sıfır geçiş sayısı konuşmacı bağımlı konuşma tanıma uygulamalarında ayırt edici özellik olarak ortaya çıkmaktadır. Çocukların ve bayanların sesinin yüksek frekanslı olması ise sıfır geçiş sayısı özelliğinin konuşmacı bağımsız tanıma sistemlerinde ayırt ediciliğini olumsuz yönde etkilemektedir.

2.1.2 Enerji Seviyesi, Tepe Değeri ve Tepe Değer Geçiş Sayısı

Ses verisine ilişkin enerji seviyesi ayırt edici bir özellik olmasa da diğer özellik çıkarım yöntemlerine önişlem olarak kelime başlangıç ve bitişlerini belirlemekte kullanılır (Zhou, L., Fang, D. 1988).

Tepe değeri ve tepe değer geçiş sayısı sıfır geçiş sayısı gibi konuşmacıya bağlı olarak telaffuzlar arasında ayırt edici özellik ortaya koyabilmektedir (Zhou, L., Fang, D. 1988).

2.1.3 Doğrusal Öngörülü Kodlama İle Özellik Çıkarımı

Doğrusal Öngörülü Kodlama (Linear Predictive Coding – LPC) yöntemi ses dalgasının durağan kabul edilebileceği kısa bir aralıkta n'inci örneğin kendinden önce gelen P adet ses örneğinin lineer bir birleşimi olarak yazılabileceği varsayımından hareketle ortaya konmuştur (Atal, 1971). Bu varsayımın altında, konuşma sırasında ses oluşumuna ilişkin organların ve ses yolunun kısa süreler boyunca akustik model çerçevesinde durağan bir yapıda kaldığı gerçeği yatmaktadır. Böylelikle kısa süreler boyunca ses yolunun akustik yapısı parametrik olarak modellenmiş olacaktır. Bu varsayım denklem (2.1) ile formülize edilmiştir.

$$\bar{s}(n) = \sum_{k=1}^{P} a_k s(n-k) \tag{2.1}$$

n anı için ses yolunun akustik modelinin değişmediği kısa aralığın, ses sinyalinin denklem (2.2) ile verilen pencereleme fonksiyonundan geçirilmesiyle elde edilmiş olan $s_n(m)$ ile ifade edildiği durumda, gerçek değer ile tahmin değeri arasındaki farktan dolayı denklem (2.3) ile

verilen tahmin hatası oluşur.

$$s_n(m) = \begin{cases} s(n+m)w(m) & 0 \le m \le N-1 \\ 0 & diger \end{cases} \tag{2.2}$$

$$e_n(m) = s_n(m) - \bar{s}_n(m) \tag{2.3}$$

Tahmin katsayılarına ilişkin en iyi kestirim denklem (2.4) ile verilen hata kareleri toplamının a_k katsayılarına göre minimize edilmesi ile (2.5) elde edilebilir.

$$\begin{aligned} E_n &= \\ &= [e_n(m)]^2 \\ &= \sum_{m=0}^{N-1} \left[s_n(m) - \sum_{k=1}^{P} a_k s_n(m-k) \right]^2 \end{aligned} \tag{2.4}$$

$$\frac{\partial E_n}{\partial a_k} = 0 \qquad k = 1,..., P \tag{2.5}$$

Denklem (2.5) ile verilen koşul tüm a_k katsayıları için uygulandığında çözümün denklem (2.6)'i sağlayan değerler olduğu görülür (Rabiner, Juang, 1993). $\phi_n(i,k)$ terimi kovaryans fonksiyonu olarak isimlendirilir.

$$\sum_{m=0}^{N-1} s_n(m-i)s_n(m) = \sum_{k=1}^{P} a_k \underbrace{\sum_{m=0}^{N-1} s_n(m-i)s_n(m-k)}_{\phi_n(i,k)} \tag{2.6}$$

Denklem (2.6) ile verilen koşulu sağlayan a_k katsayılarının bulunması için iki yöntem kullanılır. Bu yöntemlerden ilki otokorelasyon katsayıları yöntemi diğeri ise kovaryans matrisi yöntemidir.

$\phi_n(i,k)$ terimi için değişken dönüşümü yapılırsa kovaryans fonksiyonu otokorelasyon katsayıları cinsinden denklem (2.7) ile verilen şekilde yazılabilir.

$$\phi_n(i,k) = r_n(i-k) = \sum_{m=0}^{N-1-(i-k)} s_n(m)s_n(m-i-k) \qquad \begin{matrix} 1 \le i \le P \\ 0 \le k \le P \end{matrix} \tag{2.7}$$

Otokorelasyon katsayıları için $r(x) = r(-x)$ simetriklik koşulu da dikkate alınarak otokorelasyon katsayıları yöntemi denklem (2.8) ile ifade edilebilir.

$$\begin{bmatrix} r_n(0) & r_n(1) & r_n(2) & \cdots & r_n(P-1) \\ r_n(1) & r_n(0) & r_n(1) & \cdots & r_n(P-2) \\ r_n(2) & r_n(1) & r_n(0) & \cdots & r_n(P-3) \\ \vdots & \vdots & \vdots & & \vdots \\ r_n(P-1) & r_n(P-2) & r_n(P-3) & \cdots & r_n(0) \end{bmatrix} = \begin{bmatrix} a_1 \\ a_2 \\ a_3 \\ \vdots \\ a_P \end{bmatrix} \begin{bmatrix} r_n(1) \\ r_n(2) \\ r_n(3) \\ \vdots \\ r_n(P) \end{bmatrix} \tag{2.8}$$

$m = m' + i$ değişken dönüşümü yapılarak $\phi_n(i,k)$ terimi m' cinsinden yazılırsa kovaryans matrisi yöntemi denklem (2.9) ile verilen şekliyle elde edilir.

$$\begin{bmatrix} \phi_n(1,1) & \phi_n(1,2) & \phi_n(1,3) & \cdots & \phi_n(1,P) \\ \phi_n(2,1) & \phi_n(2,2) & \phi_n(2,3) & \cdots & \phi_n(2,P) \\ \phi_n(3,1) & \phi_n(3,2) & \phi_n(3,3) & \cdots & \phi_n(3,P) \\ \vdots & \vdots & \vdots & & \vdots \\ \phi_n(P,1) & \phi_n(P,2) & \phi_n(P,3) & \cdots & \phi_n(P,P) \end{bmatrix} = \begin{bmatrix} a_1 \\ a_2 \\ a_3 \\ \vdots \\ a_P \end{bmatrix} \begin{bmatrix} \phi_n(1,0) \\ \phi_n(2,0) \\ \phi_n(3,0) \\ \vdots \\ \phi_n(P,0) \end{bmatrix} \tag{2.9}$$

Gerek kovaryans matrisi yöntemi gerekse otokorelasyon katsayıları yöntemi ile $\{a_1,\ldots,a_p\}$ LPC katsayıları ilgili deklem sistemlerinin çözülmesiyle elde edilir.

2.2 Özellik Sınıflandırma Yöntemleri

Ses verisi doğası gereği ardışıl bir yapıdadır. Bu sebeple hesaplanan özellik vektörleri için elde edildikleri sıra da göz önüne alınarak uygun sınıflayıcılar kullanılmalıdır.

Ses verisinin ardışıl yapısına göre Dinamik Zaman Bükmesi, Yapay Sinir Ağları, Gizli Markov Modeli, şablon karşılaştırma gibi yöntemler uygun şekilde uyarlanarak kullanılabilir.

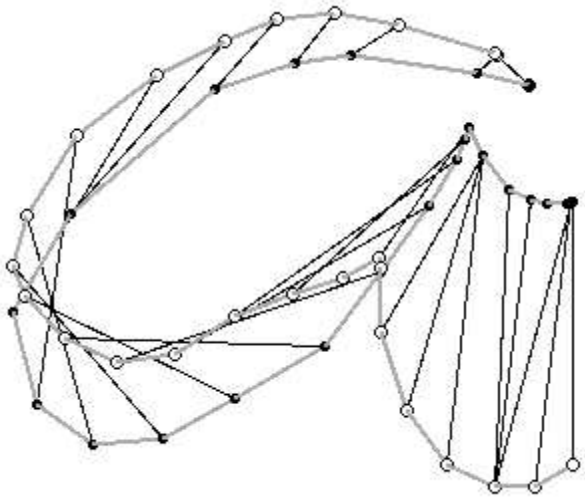
Özellik vektörlerinin sınıflandırılması ile ilgili olarak Dinamik Zaman Bükmesi ve Yapay Sinir Ağları yöntemlerinden bu bölümde bahsedilecektir.

2.2.1 Dinamik Zaman Bükmesi

1983 yılında Kruskal ve Liberman iki eğri arasındaki mesafenin belirlenebilmesi için yeni bir yöntem önerdiler. Bu yöntem ile eğriler arasında sahip olunan ortak örüntüyü yakalamak eğriler zamansal olarak farklı yerleşime sahip olsalar da mümkün olmaktadır. Şekil 2.2 ile Dinamik Zaman Bükmesi yönteminin el yazısı tanımada görsel olarak iki ayrı eğri üzerinde nasıl gerçekleştirildiği gösterilmiştir.

Ses tanımada Dinamik Zaman Bükmesi (dynamic time warping - DTW) yönteminin, aynı ses dizisinin bir kişi tarafından dahi farklı zamanlarda farklı zaman aralıklarına yayılarak telaffuz

edilebileceğinden hareketle kullanılabilir olduğu düşünülmüştür. Bu yöntemde test örneklerinin elde bulunan şablonlar ile karşılaştırılmasında uygun noktaların şablondaki karşılıklarına denk getirilmesi için ses sinyalinin zaman aralıklarında genişletilip daraltılması söz konusu olmaktadır.



Şekil 2.2 El yazısı tanıma için Dinamik Zaman Bükmesi yönteminin görsel olarak gerçekleştirilmesi (Niels, 2004)

N_1 ve N_2 iki ayrı seri için örnek sayısı olmak üzere ilk seriden i ve ikinci seriden j örnekleri denklem (2.10) ile verilen koşulu sağlarsa bu iki nokta doğrusal örtüşür olarak kabul edilirler.

$$\frac{i-1}{N_1} N_2 \leq j \leq \frac{i}{N_1} N_2 \tag{2.10}$$

Doğrusal örtüşüm koşulları dışında süreklilik, sınır ve monotonluk koşullarının da uygulanması örtüşen noktaların bulunmasında dinamik zaman bükmesi yöntemini ortaya çıkartmıştır.

Süreklilik koşulu Vuori (2002) tarafından formülize edilmiştir ve denklem (2.11) ile verilmiştir. Süreklilik koşulu ile iki seri arasında kaç adet örtüşüme izin verileceği c parametresi ile belirlenir.

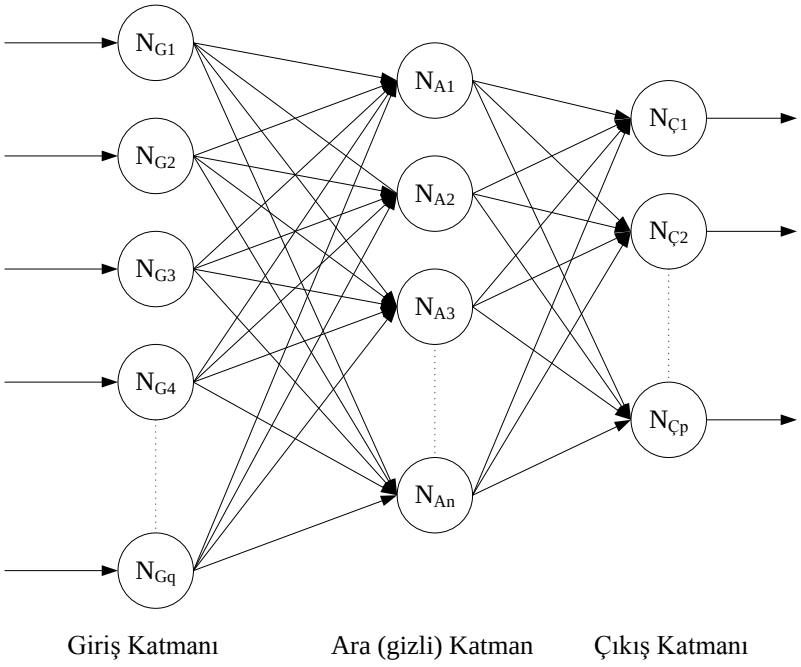
$$\frac{N_2}{N_1} i - cN_2 \leq j \leq \frac{N_2}{N_1} i + cN_2 \tag{2.11}$$

Sınır koşulları ile iki seri için ilk ve son noktaların örtüşümü zorlanır. Monotonluk koşulu ile örtüşümler ileri yönlü olarak aranır.

2.2.2 Yapay Sinir Ağları

Yapay Sinir Ağları (Artificial Neural Networks - ANN) lineer ayrıştırılamayan örneklerin sınıflandırmasında oldukça başarılı olan bir yöntemdir. Genel olarak bir giriş, bir çıkış ve bir ya da daha fazla sayıda ara katmadan oluşan, her katmanda çok sayıda karar düğümü bulunduran bir yapıdadır. Şekil 2.3 ile tipik bir çok katmanlı perseptron yapısı sembolize edilmiştir (Rainer, 1989).

Sinir ağları canlılarda var olan sinirsel iletim yapısını taklit etmektedir. Canlılarda sinirsel iletim herbir çift sinir hücresi arasında kimyasal, sinir hücresi içerisinde elektriksel olarak gerçekleşmektedir. Sinir hücreleri arasındaki iletişimin kimi iletimi bloke edici yönde kimisi ise iletimi sağlayacak yönde gerçekleşmektedir. Kendinden önceki hücrelerden aldığı kimyasal uyarıların toplamını değerlendiren sinir hücresi ilgili uyarının iletilmeye değer olup olmadığına karar vererek kendinden sonra gelen sinir hücresi için iletimi engelleyici ya da sağlayıcı kimyasal salınımı gerçekleştirir. Benzer yapı yapay sinir ağlarında nöron birimleri arasındaki bağlar için atanan ağırlıklar, nöron çıkış değerleri ve nöronların sahip olduğu karar fonksiyonları ile sağlanmaktadır.



Şekil 2.3 Çok katmanlı perseptron için genel gösterim

Eğitim aşaması sonucu yapay sinir hücreleri arası bağlantı ağırlık katsayıları eldeki eğitim örneklerini sağlayacak şekilde güncellenir. Bir test örneğinin sınıflandırılması ise test örneğine ilişkin özelliklerin giriş katmanına uygulanarak, güncellenmiş ağırlık katsayıları

üzerinden yapılan hesaplama sonucu çıkış katmanından sınıf bilgisinin alınması ile gerçekleştirilir.

Konuşma tanımada Yapay Sinir Ağları kullanımı bir özellik vektörünün hangi sesbirimine karşılık geldiğini belirlemek için kullanılır. Komut verme tarzı konuşma tanıma uygulamalarında ise tüm bir ses verisi Yapay Sinir Ağı'na verilerek komut tek ve ayrık olarak çözümlenebilir.

3. FONETİK

Konuşmada anlam farkı yaratan en küçük ses birimi fon olarak adlandırılmaktadır. Fonlar tüm diller için ortak bir yapıda değildir. Her dil için kendine özgü sesbirimleri mevcuttur. Kimi dillerde sesbirimleri doğrudan harflere karşılık gelmekteyken, kimi dillerde sesler harflerden farklı olarak simgesel telaffuz birimleri olarak fonemlere karşılık gelmektedir.

Konuşma tanımada sesbirimleri parçalı ses birimleri ve parçalarüstü sesbirimleri olarak iki anabaşlık altında incelenir (Mengüşoğlu, 1999).

3.1 Parçalı Sesbirimleri

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

3.1.1 Türkçe'de Parçalı Sesbirimleri

Türkçe Altay dilleri içerisinde Oğuz grubuna dahildir. Türkçe'de kullanılan sesbirimleri ünlü-ünsüz olma durumuna, ses tellerinin titreşimin, seslerin çıkış yeri ya da sesi çıkaran organ veya seslerin çıkış biçimine göre sınıflandırılmaktadır (Mengüşoğlu, 1999).

Türkçe'de ünlü sesler {a, e, ı, i, o, ö, u, ü}, ünsüz sesler ise {b, c, ç, d, f, g, h, j, k, l, m, n, p, r, s, s, t, v, y, z} olarak sınıflandırılmaktadır.

Ünsüz sesler, ses tellerinde titreşime sebep olup olmamalarına göre ötümlü sesler {b, d, g, v, z, j, c, l, r, m, n, y} ve ötümsüz sesler {p, t, k, f, s, ş, ç, h} olarak sınıflandırılmaktadır.

Ünlü sesler çıkış yerlerine göre dil önü ünlü sesleri {i, ü, e, ö} ve dil arkası ünlü sesleri {ı, u, a, o} olarak ayrıştırılırken ünsüz sesler çıkış yeri ve yardımcı organa göre çift dudak {p, b, m}, alt dudak-üst dişler {f, v}, dil ucu-diş arkası {t, d}, dil ucu- diş eti {s, z, n, r, l, ç, c}, dil önü-sert damak {ş, j, y}, dil arkası-damak {k, g}, ses teli-gırtlak {h} sesleri olmak üzere sınıflandırılmaktadır.

Seslerin çıkış biçimlerine göre sınıflandırılmasında ses organlarının aldıkları durumlar önemlidir. Ünlü sesler çıkış biçimlerine göre dil-damak arası açıklığa göre dar {i, ü, ı, u}, orta {e, ö, ü}, geniş {a};dudakların biçimine göre düz {i, ı, e, a}, yuvarlak {u, ü, o, ö} olarak

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

3.1.2 İngilizce'de Parçalı Sesbirimleri

İngilizce Hint-Avrupa dil ailesinden Cermen dilleri grubuna dahildir. İngilizce'de sesler ile yazı alfabesi birbirinden farklılık göstermektedir. Uluslararası Fonetik Alfabesi standardına göre İngilizce'de 50 parçalı sesbirimi bulunmakla birlikte İngilizce alfabe 26 harften oluşmaktadır. İngilizce parçalı ses birimleri çizelge 3.1, 3.2 ve 3.3 ile verilmiştir.

Çizelge 3.1 İngilizce'de kullanılan sesli fonemlerin sınıflandırılması **[4]**

	Tek Sesli Ünlüler					Çift Sesli Ünlüler		
	Kısa		Uzun			Kapalı Son		Orta Açık Son
	Ön	Arka	Ön	Orta	Arka			
Kapalı	ɪ	ʊ	iː		uː			ɪə ʊə
Orta Açık	ɛ	ʌ		ɜː	ɔː	eɪ ɔɪ	əʊ	ɛə
Açık	æ	ɒ		ɑː		aɪ	aʊ	

Çizelge 3.2 İngilizce'de kullanılan yarım sesli fonemler **[4]**

ɪ	ə	l̩	n̩	m̩

Çizelge 3.3 İngilizce'de kullanılan sessiz fonemlerin sınıflandırılması **[4]**

	Bilabial	Labio dental	Labio velar	Dental	Alveolar	Post alveolar	Palatal	Velar	Glottal
Stop	p b				t d			k g	
Affricate						tʃ dʒ			
Nasal	m				n			ŋ	
Fricative		f v		θ ð	s z	ʃ ʒ		x	h
Approximant			ʍ w		ɹ		j		
Lateral approximant					l				

İngilizce'de sesbirimleri genel olarak ünlü ve ünsüz sesler olarak ikiye ayrılır. Ünlü sesler tek sesli, çift sesli ve yarım sesli olarak sınıflandırılmaktadır. Ünsüz harfler ise çıkış yerlerine ve

çıkış biçimlerine göre sınıflandırılmaktadır.

3.2 Parçalarüstü Sesbirimleri

Parçalarüstü sesbirimleri heceler veya sözcüklerdir. Ağzın tek bir hareketiyle çıkartılabilen bir ünlü ya da bir ünlü ile bir veya birkaç ünsüz sesin birleşmesiyle oluşturulan ses modeline hece denir. Hecelerin birleştirilmesinden sözcükler oluşur.

Türkçe için altı çeşit hece yapısından söz edilebilir. Türkçe için altı çeşit hece yapısı ve ilgili yapıya ilişkin bir örnek çizelge 3.4 ile verilmiştir (Mengüşoğlu, 1999).

Çizelge 3.4 Türkçe için hece yapısı ve ilgili yapılara ilişkin örnekler

Hece Yapısı	**Örnek**
Ünlü	a
Ünlü ünsüz	ak
Ünlü ünsüz ünsüz	ilk
Ünsüz ünlü	ba
Ünsüz ünlü ünsüz	tel
Ünsüz ünlü ünsüz ünsüz	sert

Konuşma tanımaya ilişkin dar sözlük ile çalışılan uygulamalarda parçalı sesbirimleri yerine hece veya sözcük boyutunda sesbirimleri modellenerek kullanılmaktadır.

4. MEL FREKANS KEPSTRAL KATSAYI ÖZELLİKLERİ KULLANILARAK GİZLİ MARKOV MODELİ İLE GENİŞ SÖZLÜKLÜ SÜREKLİ KONUŞMA TANIMA

Konuşma tanımanın temel amacı verilen bir özellik dizisinden (X), en olası kelime çözümlemesinin (W^*) elde edilmesidir (Cole, 1997). Bu durum denklem (4.1) ile verilmiştir.

$$W^* = \arg\max_{W \in \varpi} P(W|X) \tag{4.1}$$

Koşullu olasılıkların Bayes kuralı çerçevesinde yeniden yazılması (4.2) ile aranan kelime çözümlemesinin, özellik dizisi gerçekleşme olasılığı ve kelimenin gerçekleşme olasılığına bağlı olarak yazılabileceğini gösterir (4.3).

$$P(W|X) = \frac{P(X|W)P(W)}{P(X)} \tag{4.2}$$

$$W^* = \arg\max_{W \in \varpi} P(W|X) = \arg\max_{W \in \varpi} \frac{P(X|W)P(W)}{P(X)} = \arg\max_{W \in \varpi} P(X|W)P(W) \tag{4.3}$$

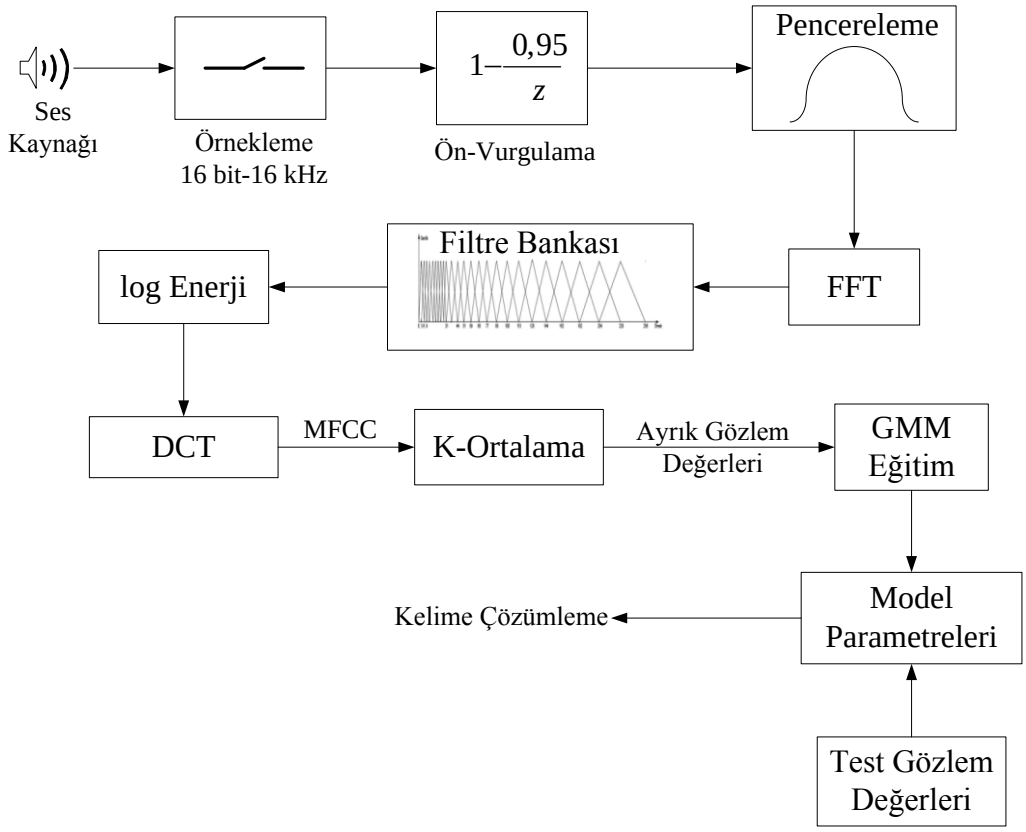
Konuşma tanımada özellik dizisi gerçekleşme olasılığı akustik model ile kelime gerçekleşme olasılığı dil modeli ile ifade edilmektedir.

Özellik dizisi gerçekleşme olasılığının, önceden gerçekleşmiş durumlardan hareketle ortaya konmasında Mel Frekans Kepstral Katsayıları ile ses verisine ilişkin ayırt edici özellik değerleri çıkartılmakta, Gizli Markov Modeli ile de ardışıl özellik değerleri çift dereceli stokastik bir model ile sınıflandırılmaktadır.

Kelimenin tüm ϖ sözlüğü içerisindeki gerçekleşme olasılığı P(W) da dil modeli ile ortaya konmaktadır.

Şekil 4.1 ile Geniş Sözlüklü Konuşma Tanıma sistemine ilişkin yapı blok diyagram olarak verilmiştir. Temel olarak blok diyagram sırasıyla önişlemler, özellik çıkarımı, ayrık gözlem dizisi elde edilmesi, Gizli Markov Modeli parametrelerinin elde edilmesi ve hesaplanan yeni parametreler üzerinden tanıma işleminin gerçekleştirilmesi işlerini özetlemektedir.

Alt başlıklarda gerekli önişlemler, özellik çıkarım adımları, akustik model ve dil modeline ilişkin temel yapı anlatılmıştır.



Şekil 4.1 Geniş Sözlüklü Konuşma Tanıma sistemi genel yapısına ilişkin blok diyagram

4.1 Önişlemler

Konuşma tanımada önişlemler olarak nitelendirilen bölüm analog ses sinyalinin sayısallaştırılması, yüksek frekans bileşenlerinin kuvvetlendirilmesi ve seçilen örnekleme frekansı doğrultusunda pencereleme için uygun boyutun, kaydırma boyutunun ve pencereleme fonksiyonun belirlenmesini içerir.

Ses sinyali yaklaşık olarak 10-30 ms'lik düzeylerde durağan kabul edilmektedir. Bu sebeple seçilen örnekleme frekansı doğrultusunda sonraki adımların gereksinimleri de düşünülerek 10-30 ms'yi sağlayacak boyutta pencere boyutu seçilmelidir.

4.1.1 Ses Verisinin Sayısallaştırılması

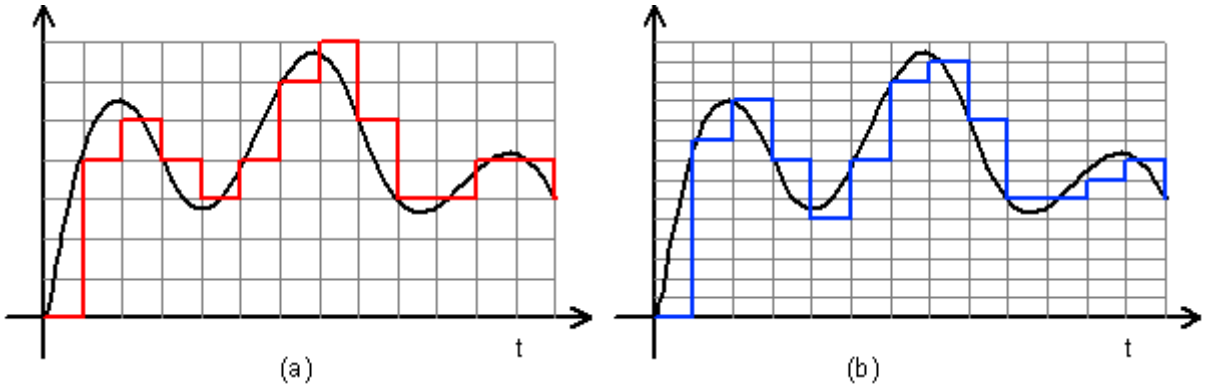
Analog ses sinyali analog-sayısal dönüştürücüler aracılığı ile sayısal sinyale dönüştürülür. Konuşma tanıma probleminin çeşitli boyutlarından hangisi ile ilgileniliyorsa ona ilişkin analog-sayısal dönüştürücü kullanılmalıdır. Çevre şartlarından, fon gürültüsünün etkisinden uzak bir konuşma tanıma çalışması için gürültü önleyen özel mikrofonlarla ses kaydı yapılmalıdır. Gürültülü ortamlarda konuşma tanıma uygulaması yapılacak ise tek kanallı

genel amaçlı mikrofonlar kullanılabilir.

Ses verisinin sayısallaştırılmasında bir başka boyut örnekleme frekansının seçimidir. Örnekleme teorisinin gereği bir sinyal en yüksek frekansının en az iki katı ile örneklenirse sinyal genel karakteristiğini kaybetmemiş olur ve sinyal örnekler üzerinden yeniden elde edilebilir.

İnsan kulağı 20 Hz ile 20 kHz frekans aralığındaki ses sinyalini duyabilir. Bu bakımdan analog ses sinyali 20 kHz'lik alçak geçiren filtreden geçirildikten sonra 40 kHz ile örneklenirse insan duyması simüle edilmiş olur. Fakat konuşma tanıma için 16 kHz'lik örnekleme frekansı yeterli olabilmektedir (Rabiner, Juang, 1993).

Ses verisinin sayısallaştırılmasında bir başka boyut da veri çözünürlüğüdür yani kuantalama ile elde edilecek verinin kaç bitlik olarak saklanacağı problemidir. Konuşma tanımada genel olarak 16 bitlik çözünürlük kullanılır. Şekil 4.2 ile çözünürlüğün arttırılması ile kuantalama hatasının azaltılabildiği gösterilmiştir.



Şekil 4.2 Çözünürlüğün örneklemeye etkisi

4.1.2 Ön-Vurgulama

Ses sinyalinde yüksek frekanslı bileşenler genel olarak düşük genliğe sahiptir. Bu yüksek frekanslı bileşenler olduğu gibi kullanılırlarsa taşıdıkları özellikler tam olarak elde edilemeyecektir. Ses sinyalindeki yüksek frekanslı bileşenlerin özelliklerini ortaya çıkarabilmek için denklem (4.4) ile belirtilen transfer fonksiyonuna sahip ön-vurgulama işlemi uygulanır (Juang, Rabiner, 1993).

$$H(z) = 1 - \frac{a}{z} \qquad 0 \le a \le 1 \tag{4.4}$$

Denklem (4.4) ile belirtilen transfer fonksiyonun zaman boyutunda uygulanması denklem (4.5) ile elde edilebilir.

$$x'(n) = x(n) - ax(n-1) \tag{4.5}$$

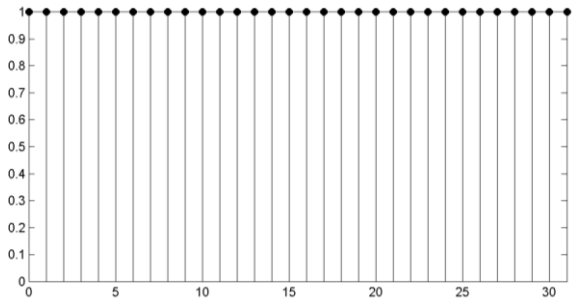
Genel olarak a için 0,95 gibi bir değer kullanılır.

4.1.3 Pencereleme

Pencereleme işlemi ile ses verisinin kısa zamanlı frekans spektrumunun elde edilmesi sağlanabilir. Pencereleme işleminde kullanılan fonksiyon ses sinyalinin durağan özellik gösterdiği varsayılan süre boyutunda tanımlanarak ses sinyali üzerinde gezdirilmek suretiyle ses sinyali ile pencereleme fonksiyonu belirli aralıklarla çarpılır.

Pencereleme işlemlerinde çeşitli fonksiyonlar uygulanabilir: bunlara örnek olarak dikdörtgen, Gauss, Hamming, Hann, Bartlett, üçgen, Bartlett-Hann, Blackman, Kaiser, Nuttal, Blackman-Harris ve Blackman-Nuttal fonksiyonları verilebilir.[**6**]

Dikdörtgen pencere fonksiyonu denklem (4.6) ile belirtilmiş olup zaman boyutundaki yapısı şekil (4.3) ile verilmiştir.[**6**]

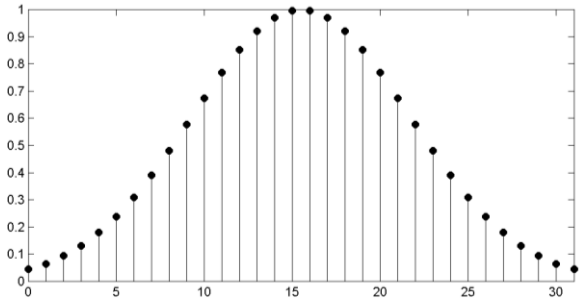


Şekil 4.3 N=32 için dikdörtgen pencere fonksiyonu

$$w(n) = 1 \tag{4.6}$$

σ standart sapma olmak üzere Gauss pencere fonksiyonu denklem (4.7) ile belirtilmiş olup zaman boyutundaki yapısı şekil 4.4 ile verilmiştir. [**6**]

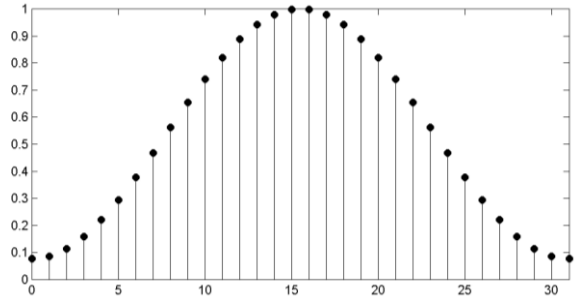
$$w(n) = e^{-\frac{1}{2}\left(\frac{n-\frac{N-1}{2}}{\sigma\frac{N-1}{2}}\right)^2} \tag{4.7}$$



Şekil 4.4 N=32, $\sigma = 0.4$ için Gauss pencere fonksiyonu

Hamming pencere fonksiyonu denklem (4.8) ile verilmiş olup zaman boyutundaki yapısı şekil 4.5 ile verilmiştir. [**6**]

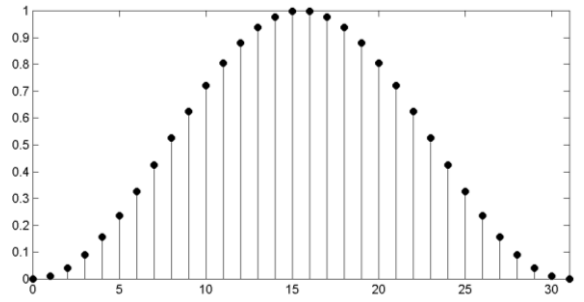
$$w(n) = 0{,}53836 - 0{,}46164 \cos\left(\frac{2\pi n}{N-1}\right) \tag{4.8}$$



Şekil 4.5 N=32 için Hamming pencere fonksiyonu

Hann pencere fonksiyonu denklem (4.9) ile verilmiş olup zaman boyutundaki yapısı şekil 4.6 ile verilmiştir. [**6**]

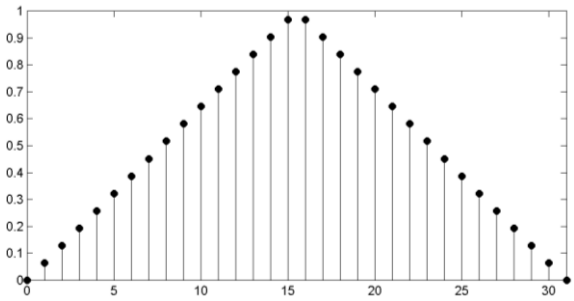
$$w(n) = 0{,}5\left(1 - \cos\left(\frac{2\pi n}{N-1}\right)\right) \tag{4.9}$$



Şekil 4.6 N=32 için Hann pencere fonksiyonu

Bartlett pencere fonksiyonu denklem (4.10) ile verilmiş olup zaman boyutundaki yapısı şekil 4.7 ile verilmiştir. Bartlett pencere fonksiyonu başlangıç ve bitiş noktalarında sıfır değerini alan üçgen pencere fonksiyonudur. Bartlett pencereleme işlemi özellik çıkarmada kepstral katsayılar hesaplanırken filtre bankası şeklinde uygulanır.

$$w(n) = \frac{2}{N-1} \cdot \left(\frac{N-1}{2} - \left| n - \frac{N-1}{2} \right| \right) \tag{4.10}$$



Şekil 4.7 N=32 için Bartlett pencere fonksiyonu

Tüm bu formülasyonlarda $0 \le n \le N-1$ koşulu geçerlidir ve N pencere boyutunu göstermektedir.

Konuşma tanıma açısından pencerelemenin önemi kısa zamanlı ayrık Fourier dönüşümü yapılacağında, pencereleme sonucu sinyalin başlangıç ve bitiş değerlerini birbirine yaklaştırması ile pencere uçlarında oluşabilecek süreksizlikleri önlemesindedir. Bu sayede elde edilecek kısa zamanlı frekans spektrumu süreksizliklerden kaynaklanacak gereksiz spektrum verisinden ayıklanmış olacaktır.

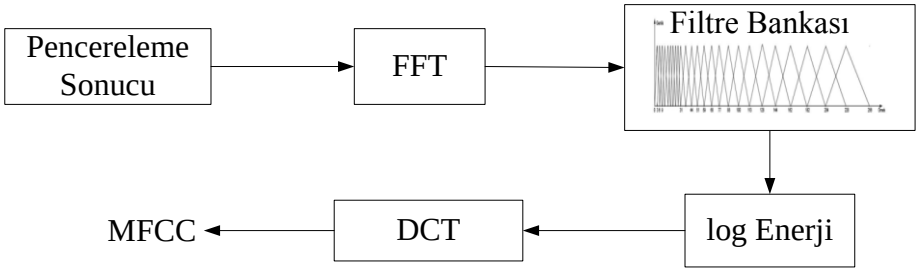
Bahsedilen duruma uygun düşecek şekilde konuşma tanımada en çok kullanılan pencereleme fonksiyonları Hamming ve Hann pencere fonksiyonlarıdır. Hamming ve Hann pencere fonksiyonları uygulanarak pencere uçlarındaki süreksizlikler giderilebilmekte fakat uç noktaların özellikleri kaybedilmektedir. Bu sebeple pencereleme birbiri ardınca değil pencere boyutundan daha kısa boyutta kaydırılarak uygulanır. Kaydırma genel olarak pencere boyutunun yarısı kadar uygulanır.

4.2 Mel Frekans Sepstrum Yöntemi ile Özellik Çıkarma

Ses verisi için çeşitli özelliklerden bahsedilebilir. Bir kısım özellikler genlik, sıfır geçiş sayısı,

frekans spektrumu, formant* değerleri, ses perdesi (pitch) ve sinyalin taşıdığı enerji gibi ses dalgası ile ilişkilendirilebilecek özelliklerdir. Bunun yanında mel frekans sesptrum katsayıları, otokorelasyon katsayıları, Doğrusal Öngörülü Kodlama katsayıları gibi bir takım sinyal işleme teknikleri ya da istatistiksel işlemlerle elde edilebilecek özelliklerden de bahsedilebilir.

Sürekli konuşma tanıma açısından başarılı sonuçlar vermiş olan Mel Frekans Kepstral Katsayıları'nın elde edilmesi işlemi ses verisinin belirli frekans bantlarına düşen güç değerlerinin bir ortalamasının alınarak bantlar arası lineer bağımsız katsayıların hesaplanması ile gerçekleştirilmektedir. Bu amaçla zaman boyutundaki ses verisi için frekans spektrumu ayrık Fourier dönüşümü ile bulunur. İnsan duyması simüle edilecek şekilde 1-2 kHz civarına kadar lineer olarak eşit daha sonrası için logaritmik olarak eşit aralıklı örtüşümlü filtre bankası uygulanarak belirli frekans bantlarına düşen ortalama güç yaklaşık olarak bulunur. Ortalama güç değerleri logaritmik olarak yeniden ifade edilerek frekans spektrumu normalize edilir.



Şekil 4.8 Özellik çıkarım adımları blok şeması

Ayrık kosinüs dönüşümü ile de hesaplanan özellik değerleri birbirlerinden lineer bağımsız bir şekilde dönüştürülür. Uygulamalarda Mel Frekans Kepstral Katsayıları'nın yanında birinci dereceden ve ikinci dereceden fark katsayıları da hesaplanarak özellik vektörü olarak kullanılmaktadır.

4.2.1 Spektral Analiz

Önişlemler ile belirli boyutlarda pencereler halinde düzenlenmiş zaman boyutundaki ses verisi spektral analiz ile frekans özellikleri ortaya konacak şekilde frekans boyutuna geçirilir.

Genel olarak bir sinyalin zaman boyutundan frekans boyutuna geçirilmesi dönüşüm olarak adlandırılmaktadır. Temel olarak zaman boyutunda sürekli işaretler için frekans dönüşümü

* Akustik teoride ses yolunun şekil yapısı itibariyle sahip olduğu rezonans frekans değerleri.

yine sürekli bir işaret üretmektedir, benzer şekilde tanım gereği zaman boyutunda ayrık işaretler için de frekans dönüşümü sürekli bir sinyal üretmektedir.

Zaman boyutunda sürekli sinyallerin frekans dönüşümleri için Fourier dönüşümü ve Laplace dönüşümü kullanılır. Tanım olarak Fourier dönüşüm çifti $\{X(\omega), x(t)\}$ denklem (4.11), (4.12) ile, Laplace dönüşüm çifti $\{F(s), f(t)\}$ ise denklem (4.13), (4.14) ile verilmiştir. [7]

$$X(\omega) = \int_{-\infty}^{\infty} x(t) e^{-j\omega t} dt \tag{4.11}$$

$$x(t) = \frac{1}{2\pi} \int_{-\infty}^{\infty} X(\omega) e^{j\omega t} dt \tag{4.12}$$

$$F(s) = \int_{0^-}^{\infty} e^{-st} f(t) dt \qquad s = \sigma + j\omega \tag{4.13}$$

$$f(t) = \frac{1}{2\pi j} \int_{\gamma - j\infty}^{\gamma + j\infty} e^{st} F(s) ds \qquad \begin{matrix} s = \sigma + j\omega \\ \gamma \in \Re \end{matrix} \tag{4.14}$$

Bu iki dönüşüm çifti arasındaki fark Fourier dönüşümünün sürekli rejime ilişkin inceleme imkanı vermesi, Laplace dönüşümünün ise hem geçici rejim hem de sürekli rejime ilişkin inceleme imkanı veren daha genel bir dönüşüm olmasıdır. Ayrıca Fourier dönüşümünün uygulanabilmesi için sinyalin sonlu sayıda sonlu ekstremum noktalara sahip olması gerekir.

Sese ilişkin sinyal işleme için gerek ses oluşumunda ya da yayılımında oluşacak sistem geçici hal cevabı ile ilgilenilmediği gerekse de ses sinyalinin sonsuz süreksizlikleri olmayacağı için Fourier dönüşümü uygun düşmektedir.

Zaman boyutunda ayrık sinyallerin frekans boyutuna geçirilmesinde ayrık zamanlı Fourier dönüşümü ve z dönüşümü kullanılır. z dönüşümü Laplace dönüşümünün ayrık zamanlı sinyaller için uygulanan benzeridir. Ayrık zamanlı Fourier dönüşüm çifti denklem (4.15), (4.16) ile verilmiştir. [7]

$$X(\omega) = \sum_{n=-\infty}^{\infty} x[n] e^{-j\omega n} \tag{4.15}$$

$$x[n] = \frac{1}{2\pi} \int_{-\pi}^{\pi} X(\omega) e^{j\omega n} d\omega \tag{4.16}$$

Daha önce bahsedildiği gibi zaman boyutunda ayrık işaretin frekans boyutuna geçirilmesi sonucu sürekli bir fonksiyon üretilmektedir. Bilgisayar ortamında bu verilerin işlenebilmesi

için tıpkı analog ses verisinin örneklenmesi gibi, ayrık zamanlı Fourier dönüşümündeki ω frekans bileşeni de bir periyot boyunca $1/N$ aralıklarla örneklenerek kullanılır. Bu şekildeki uygulama ayrık Fourier dönüşümü olarak adlandırılır ve denklem (4.17), (4.18) ile tanımlanmaktadır. [7]

$$X[k] = \sum_{n=0}^{N-1} x[n]\, e^{-j\frac{2\pi k}{N}n} \qquad 0 \le k \le N-1 \tag{4.17}$$

$$x[n] = \frac{1}{N}\sum_{k=0}^{N-1} X[k] e^{j\frac{2\pi k}{N}n} \qquad 0 \le n \le N-1 \tag{4.18}$$

Ayrık Fourier dönüşümü genel olarak reel ya da karmaşık katsayılı giriş sinyallerine uygulanabilmektedir. Ses sinyalleri için ise reel katsayılı giriş sinyalleri söz konusudur.

Ayrık Fourier dönüşümü hesaplama işlemi denklem (4.17) ile verildiği şekliyle N^2 çarpma ve toplama gerektirmektedir. 1965 yılındaki çalışmaları ile Cooley ve Tukey ayrık Fourier dönüşümü için karmaşıklığı r^p örnekli bir veri dizisi için $rN \log_r N$ mertebesine indirgeyecek hızlı Fourier dönüşümünü ortaya koymuşlardır. Denklem (4.19), (4.20) ile verilen tanımlar ve N'in $N = r_1 r_2$ şeklinde çarpanlarına ayrılabildiği varsayımıyla ayrık Fourier dönüşümü denklem (4.21) ile verilen şeklide yeniden yazılabilir.

$$W = e^{-\frac{2\pi j}{N}} \tag{4.19}$$

$$\begin{aligned} k &= k_1 r_1 + k_0 \qquad k_0 = 0,1,..r_1 - 1 \qquad k_1 = 0,1,..r_2 - 1 \\ n &= n_1 r_2 + n_0 \qquad n_0 = 0,1,..r_2 - 1 \qquad n_1 = 0,1,..r_1 - 1 \end{aligned} \tag{4.20}$$

$$X[k_0, k_1] = \sum_{n_0=0}^{r_2-1}\sum_{n_1=0}^{r_1-1} x[n_0, n_1] W^{kn_1r_2} W^{kn_0} \tag{4.21}$$

Ayrıca denklem (4.22) ile verilen terim için yapılabilecek indirgemeye yer verilmiştir.

$$W^{kn_1r_2} = W^{(k_0+k_1r_1)n_1r_2} = W^{k_0n_1r_2} W^{k_1n_1r_1r_2} = W^{k_0n_1r_2} W^{k_1n_1N} = W^{k_0n_1r_2} \underbrace{e^{-2\pi j k_1 n_1}}_{1} = W^{k_0n_1r_2} \tag{4.22}$$

$$X[k_0, k_1] = \sum_{n_0=0}^{r_2-1} \underbrace{\sum_{n_1=0}^{r_1-1} x[n_0, n_1] W^{k_0n_1r_2}}_{A[n_0, k_0]} W^{kn_0} \tag{4.23}$$

Denklem (4.23) ile verilen ayrık Fourier dönüşümüne ilişkin $A[n_0, k_0]$ teriminin her bir elemanını hesaplamak r_1 işlem gerektirmektedir. A için $n_0 k_0 = N$ eleman hesaplanabileceği düşünülürse, A teriminin hesaplanması Nr_1 karmaşıklıktadır. $X[k_0, k_1]$ teriminin verilen bir A terimi üzerinden hesaplanmasında ise herbir eleman için r_2 işlem gerekmektedir. X için $k_0 k_1 = N$ eleman hesaplanması gerektiği düşünülürse, X teriminin hesaplanabilmesi için Nr_2 karmaşıklık gerekmektedir. Toplamda ise X teriminin hesaplanması için N^2 karmaşıklık yerine $N(r_1 + r_2)$ mertebesinde bir karmaşıklık yeterli olabilmektedir. N'in çok sayıda çarpanı olan bir bileşik sayı olması durumunda ise örneğin $N = r^p$ şeklinde yazılabiliyorsa hızlı Fourier dönüşümü ile karmaşıklık $rN \log_r N$ mertebesine indirgenebilir. Genel olarak $r = 2$ seçilmesi durumunda ise Cooley ve Tukey (1965) ikili bilgisayar sistemlerinin adresleme ve çarpma işlemlerinde kolaylık sağlayacaklarını söylemektedir. Şekil 4.9 ile hızlı Fourier dönüşümünün gerçekleştirilmesine ilişkin adımlar verilmiştir.

```
N = 2^m
For k = 1, ...,m
        For j = 1, ..., 2^(k−1)
                t = (j − 1) * N
                For n = 0, ...,N/2 − 1
                        wfactor = (e^(i2π/N n))^(2k−1)
                        temp = wfactor *(f_(n+t) − f_(n+t+N/2))
                        f_(n+t) = (f_(n+t) + f_(n+t+N/2))
                        f_(n+t+N/2) = temp
                EndFor
        EndFor
EndFor
```

Şekil 4.9 Hızlı Fourier Dönüşümü adımları (Cooley, Tukey, 1965)

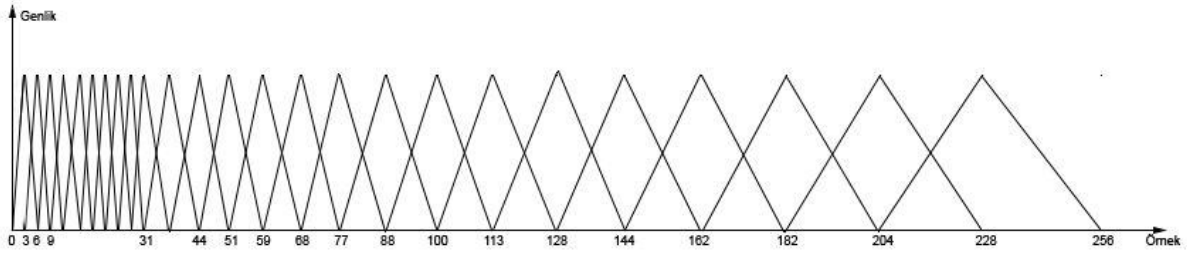
Bahsedilen hızlı Fourier dönüşümü uygulanarak zaman boyutundaki ses verisine ilişkin frekans özellikleri elde edilmiş olmaktadır. Frekans özellikleri karmaşık katsayılı bir dizi olarak elde edilir. Bu karmaşık katsayı dizisinden genlikler elde edilerek özellikler olarak genlik değerleri tutulur. Sonuçta kullanılan pencere boyutunda simetrik bir genlik değerleri vektörü elde edilir.

4.2.2 Filtre Bankası Uygulama

Frekans boyutunda hesaplanan genlik değerleri kullanılan pencereleme fonksiyonu boyutunda elde edilmektedir. Özellik vektörünün bu boyutu ile kullanılması yerine belirli bantlara düşen

ortalama genlik değerleri alınarak hem özellik vektörü boyutu düşürülmüş olmakta hem de belirli bir uygunlaştırma, filtreleme sağlanmaktadır.

Filtre bankası uygulamasında belirli bir frekans değerine kadar lineer eşit aralıklı, daha sonrası için ise logaritmik eşit aralıklı kısmen örtüşümlü üçgen filtre bankası uygulanmaktadır. Böylelikle insan kulağının sahip olduğu yapı simüle edilmiş olmaktadır.



Şekil 4.10 Filtre bankası örneği

Şekil 4.10 ile 16 kHz ile örneklenmiş bir sinyal için uygulanabilecek filtre bankasına ilişkin bir örnek verilmektedir. Bu örnekte sinyal 16 kHz ile örneklendiği için en yüksek frekansı 8 kHz olarak kabul edilir. Kullanılan pencereleme fonksiyonu boyutu 512 olduğu durumda, FFT'den elde edilecek 512 örneklik simetrik genlik katsayısından kullanılacak olan ilk 256 örneğin hangi filtrede hesaba katılacağı yatay eksendeki örnek sınırlarıyla verilmiştir. Burada 256'ıncı örnek 8kHz frekansına denk gelmektedir.

Logaritmik eşit aralıkların oluşturulmasında yine insan duymasından hareketle ortaya konmuş olan mel frekans ölçeği kullanımı söz konusudur. Frekans skalası ile mel frekans skalası arasındaki dönüşüm denklem (4.24), (4.25) ile verilmektedir.

$$\mathrm{m} = 2595 \cdot \log(1 + \frac{f}{700}) \tag{4.24}$$

$$f = 700 \cdot (10^{\frac{\mathrm{m}}{2595}} - 1) \tag{4.25}$$

4.2.3 Log Enerji

Çoğunluğu mel frekans skalasında eşit aralıklarda oluşturulan pencerelerden elde edilen filtre bankasının uygulanması sonucu hesaplanan özellikler istatistiki olarak sola çarpık olarak adlandırılan yapıda elde edilir. Özelliklerin istatistiki yapısını normal dağılıma yaklaştırmak için özellik değerleri logaritmaları alınarak kullanılır.

4.2.4 Ayrık Kosinüs Dönüşümü

Mel Frekans Kepstral Katsayıları'nın hesaplanmasında son aşama ayrık kosinüs dönüşümünün uygulanmasıdır. Ayrık kosinüs dönüşümü uygulanması ile özellik vektöründe hem boyut azaltılmış olmakta hem de özellik değerlerinin kosinüs fonksiyonlarının lineer bağımsız bir bileşeni olarak yazılması ile aralarında korelasyon bulunmayan özellikler ortaya çıkarılmış olmaktadır.

Kullanılan filtre bankasının 25 adet filtreden oluştuğu durum için ayrık kosinüs dönüşümü sonucu boyutu 25 olan simetrik bir özellik vektörü elde edilir. 25 adet özellikten ilk 13 tanesi ilgili ses sinyali parçasına ilişkin Mel Frekans Kepstral Katsayıları olarak alınır. İlk MFKK katsayısı işaretin DC bileşenine ilişkin bilgi içermektedir ve ayırt edici bir özellik taşımadığı için özellik vektörüne katılmaz. Sonuçta 25 adet filtrenin kullanılması ile 12 boyutlu bir özellik vektörü elde edilir.

Kullanılan K adet filtre sonucu elde edilen değerler X_k (k=1, ..., K) ile gösteriliyor ise log enerji ve ayrık kosinüs dönüşümü sonucu birlikte denklem (4.26) ile verilebilir (Young).

$$MFCC(i) = \sum_{k=1}^{K} \log(X_k) \cos(i(k - \frac{1}{2}) \frac{\pi}{K}) \tag{4.26}$$

4.2.5 Özellik Vektörünün Oluşturulması

Hesaplanan MFKK katsayıları yanında, ardışık MFKK katsayıları arasındaki birinci ve ikinci dereceden farklar ve ses sinyaline ilişkin enerji değerini de içeren bir özellik vektörü oluşturulur. Ardışık MFKK katsayıları arasındaki birinci ve ikinci derece farklardan hesaplanan özellikler delta özellikleri olarak adlandırılmaktadır.

4.2.6 K-Ortalama Vektör Kuantalama Yöntemi ile Özellik Azaltılması

Hesaplanan özellik değerleri çok boyutlu reel değerlikli özellik vektörleri olarak elde edilmektedir. Kullanılacak Gizli Markov Modeli gözlem sembolleri olasılık dağılımı yapısına göre (B değerlerinin sürekli ya da ayrık olması durumu) veya gözlem sembolleri için ayrık olasılık dağılımı kullanılacak ise vektör kuantalama ile reel değerlikli özellik vektörleri belirli sayıda ayrık durum ile ifade edilebilir.

Vektör kuantalama için K-ortalama algoritması kullanılabilir. K-ortalama algoritması ile verilen özellik vektörleri istenen sayıda merkez vektörü ile temsil edilebilmektedir. Bu şekildeki bir kuantalama ile herbir pencereden elde edilen özellik vektörü bir sınıf olarak

kullanılabilir. Sonuçta (özellik vektörü boyutu)x(pencere sayısı) adet özellik değeri, (pencere sayısı) adet özellik sayısına indirgenebilir.

K-ortalama algoritmasının adımları şu şekildedir:

Eğitim setinden k adet rassal örnek başlangıç sınıf merkezleri olarak atanır

Yakınsama durumuna kadar

Eğitim setindeki her örnek en yakın merkeze sahip sınıfa atanır

Herbir sınıf için yeni merkez değerleri hesaplanır

4.3 Akustik Model

Ses verisinden elde edilmiş özelliklerin gerek sesbirimleri boyutunda gerekse kelime boyutunda sonlu durum makinesi şeklinde istatistiki olarak modellenmesi akustik model olarak adlandırılır.

Sesbirimleri boyutunda modelleme Gizli Markov Modeli ile yapılırken, kelime boyutunda modelleme Markov modeli ile gerçekleştirilmektedir. Kelime boyutunda modelleme kelimelerin hangi kural ile dizileceğini irdelediğinden dilbilgisi ve dilin genel yapısı ile yakından ilgilidir ve bu sebeple dil modeli olarak adlandırılmaktadır.

Bu bölümde Gizli Markov Modeli ve Gizli Markov Modeli'nin temelini oluşturan Markov zincirleri ve Markov modeli ile ilgili bilgi verildikten sonra gerekli parametrelerin hesaplanmasının nasıl yapıldığı anlatılacaktır. Ayrıca Gizli Markov Modelinde kullanılacak değişkenlerin sürekli ya da ayrık olması koşulunda parametre tahminin nasıl yapılacağından da bahsedilecektir. Daha sonra dil modelinin genel yapısı hakkında bilgi verilecektir.

4.3.1 Markov Zincirleri

Markov süreçleri, bir dizi durum ve herhangi bir an için bu durumlardan birinden diğerine geçişin o anda bulunulan durumun ile birlikte önceki bulunulan durumlara da bağlı olarak tanımlandığı stokastik süreçlerdir.

Markov süreçlerinin bir alt başlığı olan Markov zincirleri ise ayrık zamanlı, durumlar arası geçişin sadece bulunulan duruma bağlı olarak tanımlandığı süreçlerdir. Buna göre Markov zincirleri model λ ile tanımlanabilir. (Rabiner , 1989)

$$\lambda = \{S, A, \pi\} \tag{4.27}$$

λ modelinde

S : N adet ayrık elemandan oluşan durum kümesi

A : NxN'lik durumlar arası geçiş matrisi

π : N boyutlu başlangıç durum vektörü

olarak tanımlamaktadır. Model, Q={q_1, q_2, q_3, ..., q_T } T adet durum gözlem sırası olmak üzere denklem (4.28), (4.29) ile verilen koşulları sağlar. (Rabiner , 1989)

$$P(q_t = S_j \mid q_{t-1} = S_i, q_{t-2} = S_k, \cdots) = P(q_t = S_j \mid q_{t-1} = S_i) \tag{4.28}$$

$$P(q_t = S_i \mid q_{t-1} = S_j) = P(q_{t+l} = S_i \mid q_{t+l-1} = S_j) \tag{4.29}$$

Denklem (4.28) birinci dereceden Markov zinciri koşulunu belirtir ve sonraki durumun sadece bulunulan duruma bağlı olarak tanımlanması anlamındadır (Rabiner , 1989). Denklem (4.29) ise durumlar arası geçiş parametrelerinin zamanla değişmediğini gösterir (Rabiner , 1989).

Durum geçiş matrisi A için denklem (4.30-4.33) koşulları geçerlidir (Rabiner , 1989).

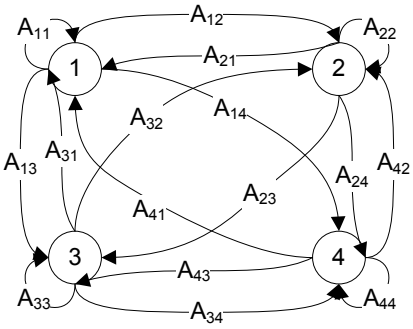
$$a_{ij} = P(q_t = S_j \mid q_{t-1} = S_i) \qquad 1 \le i, j \le N \tag{4.30}$$

$$A = [a_{ij}] \tag{4.31}$$

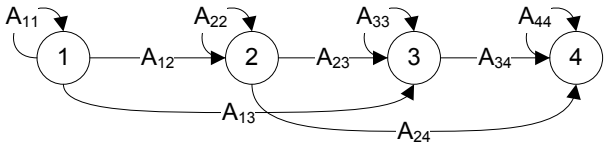
$$a_{ij} \ge 0, \qquad \forall i, j \tag{4.32}$$

$$\sum_{j=1}^{N} a_{ij} = 1, \qquad \forall i \tag{4.33}$$

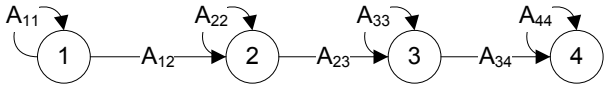
Geçiş matrisinin tanımladığı geçişler doğrultusunda oluşan gizli Markov modelleri farklı isimler ile adlandırılmaktadır. Geçiş matrisinde tanımlı geçişlere göre Markov model çeşitleri şekil 4.11 ile verilmiştir.



(a)



(b)



(c)

Şekil 4.11 a. Ergodik, b. Soldan sağa, c. Soldan sağa zorlanmış Markov model çeşitleri.

Başlangıç durum vektörü π için denklem (4.34-4.36) koşulları geçerlidir.

$$\pi_i = P(q_1 = S_i), \qquad 1 \le i \le N \tag{4.34}$$

$$\pi_i \ge 0, \qquad \forall i \tag{4.35}$$

$$\sum_{i=1}^{N} \pi_i = 1 \tag{4.36}$$

Yukarıda yapılan tanımlamalar ve sınırlandırmalar çerçevesinde, koşullu olasılık kuralı denklem (4.37)'den hareketle, verilen herhangi bir Q={q_1, q_2, q_3, ..., q_T } gözlem sırası için bu gözlem sırasının gerçekleşme olasılığı denklem (4.38) ile verilmiştir.

$$P(\mathrm{A},\mathrm{B}) = P(A \mid B)P(B) \tag{4.37}$$

Denklem (4.37) A ve B olaylarının birlikte gerçekleşme olasılığının B olayının ve B olayı varken A olayının gerçekleşme olasılıklarına bağlı olduğunu ifade etmektedir.

$$
\begin{aligned}
P(Q) = & \\
& = P(q_1, q_2, q_3, .. q_T) \\
& = P(q_T \mid q_1, q_2, q_3, .. q_{T-1}) P(q_1, q_2, q_3, .. q_{T-1}) \\
& = P(q_T \mid q_{T-1}) P(q_1, q_2, q_3, .. q_{T-1}) \\
& = P(q_T \mid q_{T-1}) P(q_{T-1} \mid q_{T-2}) P(q_1, q_2, q_3, .. q_{T-2}) \\
& \vdots \\
& = P(q_T \mid q_{T-1}) P(q_{T-1} \mid q_{T-2}) \cdots P(q_2 \mid q_1) P(q_1) \\
& = \pi_{\mathrm{i}=q_1} \prod_{t=2}^{T} a_{q_{t-1} q_t}
\end{aligned}
\tag{4.38}
$$

Denklem (4.38) ile Q gözlem sırası, koşullu olasılık kuralı doğrultusunda, tek tek parçalara ayrılmaktadır. Her adımda koşullu olasılıklar, birinci dereceden Markov zinciri özelliği uygulanarak durumlar arası geçiş olasılıklarına indirgenebilmektedir. Son olarak kalan $P(q_1)$ ise başlangıç durumu olasılığı belirtmektedir.

4.3.2 Gizli Markov Modeli

Markov zincirlerinde her bir gözlemin bir duruma karşılık geldiği varsayımı söz konusudur. Markov zincirlerinden farklı olarak Gizli Markov Modeli'nde durumların doğrudan gözlenebilir olmadığı varsayılır; gözlemlerin gerçek durumların bir olasılık fonksiyonu olarak gerçekleştiği varsayılır. Buna göre Gizli Markov Modeli'nde gözlemler, gözlemlerden durumlara geçiş, gizli durumlar, durumlar arası geçişler söz konusu olmaktadır.

Gizli Markov Modeli $\lambda=\{N, M, A, B, \pi\}$ ile tanımlanabilir. Model parametrelerine ilişkin tanımlar aşağıda verilmiştir (Rabiner , 1989).

N: modeldeki durum sayısıdır. Durum kümesi $S=\{S_1, S_2, ..., S_N\}$ ile gösterilmekte ve t anında bulunulan durum q_t ile ifade edilmektedir.

M: ayrık gözlem sembolleri sayısıdır. Gözlem sembolleri modelin fiziki çıktılarına karşılık gelmektedir. Gözlem sembolleri kümesi $V=\{v_1, v_2, ..., v_M\}$ ile gösterilmekte ve t anında gözlenen sembol O_t ile ifade edilmektedir.

A: Durumlar arası geçiş olasılık matrisidir. Sözel olarak; t anında i. durumda bulunulup t+1 anında j. durumda bulunma olasılığını ifade eder. $A=[a_{ij}]$ denklem (4.39) ile tanımlanmıştır.

$$a_{ij} = P(q_{t+1} = S_j \mid q_t = S_i) \qquad 1 \le i, j \le N \tag{4.39}$$

B: Gözlem sembolleri olasılık dağılımıdır. Sözel olarak t anında j. durumda bulunulup t. gözlemde k sembolünün üretilme olasılığıdır. B=[$b_j(k)$] denklem (4.40) ile tanımlanmıştır.

$$b_j(k) = P(O_t = v_k \mid q_t = S_j) \qquad \begin{matrix} 1 \le j \le N \\ 1 \le k \le M \end{matrix} \tag{4.40}$$

Doğrudan gözlenemeyen durumların birer olasılık dağılımı olarak meydana getirdiği gözlem sembollerinin gerçekleşme olasılıkları $b_j(k)$ denklem (4.41), (4.42) ile verilen koşulları sağlar.

$$b_j(k) \ge 0, \qquad \forall j,k \tag{4.41}$$

$$\sum_{k=1}^{M} b_j(k) = 1, \qquad \forall k \tag{4.42}$$

π: Başlangıç durumu olasılık vektörüdür. Sözel olarak i. durumun ilk durum olma olasılığıdır. π=[π_i] denklem (4.43) ile tanımlanmıştır.

$$\pi_i = P(q_1 = S_i) \qquad 1 \le i \le N \tag{4.43}$$

Yapılan tanımlamalar çerçevesinde Gizli Markov Modeli için üç problem tanımlanabilir (Rabiner , 1989). Bunlar:

1. Verilen $O=O_1O_2...O_T$ gözlem sırası ve λ={A, B, π} modeli için P(O|λ) gözlem sırasının λ modeli ile gerçekleşme olasılığın hesaplanması.
2. Verilen $O=O_1O_2...O_T$ gözlem sırası ve λ={A, B, π} modeli için en anlamlı $Q=q_1q_2...q_T$ durum sırasının hesaplanması.
3. Verilen $O=O_1O_2...O_T$ gözlem sırası ve λ={A, B, π} modeli için P(O|λ) olasılığını maksimize edecek şekilde model parametrelerinin tahmin edilmesi.

Bahsedilen her bir probleme ilişkin çözümler sırasıyla İleri/Geri (Forward/Backward) algoritması, Viterbi algoritması ve Baum-Welch algoritması ile en etkin şekilde gerçekleştirilebilmektedir.

4.3.2.1 İleri/Geri Algoritması

Gözlem sembolleri sırası gerçekleşme olasılığı ile ilgili olarak tanımlanmış olan problem 1'in çözümü için tek başına İleri ya da tek başına Geri algoritması yeterli olabilmektedir.

İleri algoritması için $\alpha_t(i)$ değişkeni t anına kadarki kısmi gözlem sırası sonucunda S_i durumunda bulunulma olasılığı olarak tanımlanır ve denklem (4.44) ile ifade edilir.

$$\alpha_t(i) = P(O_1, O_2, ..., O_t, q_t = S_i \mid \lambda) \tag{4.44}$$

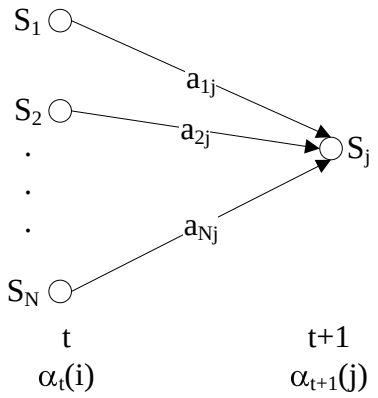
İleri algoritması aşağıda bahsedilen adımlarla gerçekleştirilir (Rabiner , 1989):

Başlangıç: $\alpha_1(i) = \pi_i b_i(O_1)$ (4.45)

Çıkarım: $\alpha_{t+1}(j) = \left[\sum_{i=1}^{N} \alpha_t(i) a_{ij} \right] b_j(O_{t+1})$ (4.46)

Bitiş: $P(O \mid \lambda) = \sum_{i=1}^{N} \alpha_T(i)$ (4.47)

Sözel olarak; İleri algoritması ile ilk adımda i durumunun başlangıç durumu olması ve bu durumun gözlenen O_1 sembolünü üretmesi ihtimali $\alpha_1(.)$ değeri olarak atanır. İkinci adımda şekil 4.12 ile özetlenebilecek, tüm durumlardan j durumuna geçiş ve bu yeni durumda O_{t+1} gözlem sembolünün üretilme olasılığı hesaplanır.



Şekil 4.12 İleri algoritması için yineleme adımı şematik gösterimi. (Rabiner, 1989)

Üçüncü adım ile de gözlem sırasının gerçekleşme olasılığı $\alpha_T(.)$ tüm durumlar üzerinden toplanarak hesaplanması ile bulunur.

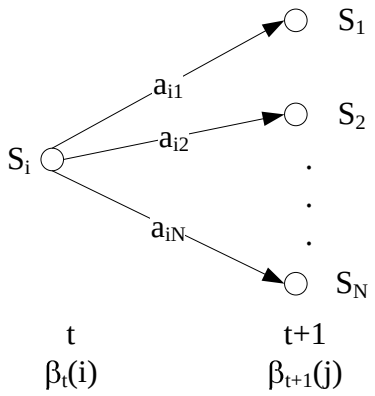
Geri algoritması için $\beta_t(i)$ değişkeni t anında S_i durumunda bulunulup t+1 anından itibaren sonrası için gözlem sırasının gerçekleşme olasılığı olarak tanımlanır ve denklem (4.48) ile ifade edilir.

$$\beta_t(i) = P(O_{t+1}, O_{t+2}, ..., O_T \mid q_t = S_i, \lambda) \tag{4.48}$$

Geri algoritması aşağıda bahsedilen adımlarla gerçekleştirilir (Rabiner , 1989):

Başlangıç: $\beta_T(i) = 1 \qquad 1 \le i \le N$ (4.49)

Çıkarım: $\beta_t(i) = \sum_{j=1}^{N} a_{ij} b_j(O_{t+1}) \beta_{t+1}(j) \qquad \begin{matrix} 1 \le i \le N \\ 1 \le t \le T-1 \end{matrix}$ (4.50)



Şekil 4.13 Geri algoritması için yineleme adımı şematik gösterimi. (Rabiner, 1989)

4.3.2.2 Viterbi Algoritması

Çözümleme problemi olarak da bilinen Gizli Markov Modeli'ne ilişkin problem 2, gözlenen sembollerin altında gerçekleşen optimum durum dizisini ortaya çıkarmayı amaçlamaktadır. Optimumluk tanımı ya durumların tek tek her biri için ya da tüm bir gözlem dizisinin bütünü için yapılabilir. Tek tek durumlar üzerinden optimizasyon gerek mantık dışı sonuçlar doğurabileceği gerekse karmaşık olduğu için tercih edilmez.

Viterbi algoritması ile verilen bir O={O_1, O_2, ..., O_T} gözlem sembolleri dizisi için en uygun durum sırası Q={q_1, q_2, ..., q_T} tek olarak elde edilebilir.

$$\delta_t(i) = \max_{q_1, q_2, \ldots, q_{t-1}} P(q_1, q_2, \ldots, q_t = i, O_1, O_2, \ldots, O_t \mid \lambda) \tag{4.51}$$

Denklem (4.51) ile O_1'den O_t'ye kadar olan gözlem sembollerini sağlayan ve t anında S_i durumunda bulunulan durum dizisi için maksimum olasılık tanımlanır.

$$\delta_{t+1}(j) = (\max_i \delta_t(i) a_{ij}) b_j(O_{t+1}) \tag{4.52}$$

Denklem (4.52) ile j durumuna hangi i. durumdan geçişin en yüksek olasılıkla gerçekleşebileceği hesaplanarak j durumunda O_{t+1} gözlem sembolü üretilme olasılığı δ_{t+1} olarak atanır.

Durumlar arası en olası geçişler $\psi_t(j)$ ile tutularak, son adımdan itibaren en olası yol geriye doğru takip edilerek istenen durum dizisi elde edilir.

Viterbi algoritması şu adımlardan oluşur (Rabiner , 1989):

Başlangıç:

$$\delta_1(i) = \pi_i b_i(O_1) \qquad 1 \le i \le N \tag{4.53}$$

$$\psi_1(i) = 0 \tag{4.54}$$

Yineleme:

$$\delta_t(j) = \max_i(\delta_{t-1}(i)a_{ij})b_j(O_t) \qquad \begin{matrix} 1 \le i \le N \\ 2 \le t \le T \\ 1 \le j \le N \end{matrix} \tag{4.55}$$

$$\psi_t(j) = \arg\max_i(\delta_{t-1}(i)a_{ij}) \qquad \begin{matrix} 1 \le i \le N \\ 2 \le t \le T \\ 1 \le j \le N \end{matrix} \tag{4.56}$$

Bitiş:

$$P^* = \max_i(\delta_T(i)) \qquad 1 \le i \le N \tag{4.57}$$

$$q_T^* = \arg\max_i(\delta_T(i)) \qquad 1 \le i \le N \tag{4.58}$$

Yol Takibi:

$$q_t^* = \psi_{t+1}(q_{t+1}^*) \qquad t = T-1, T-2, ..., 1 \tag{4.59}$$

4.3.2.3 Baum-Welch Algoritması

Gizli Markov Modeli için eğitim aşaması model parametrelerinin tahmini olarak gerçekleştirilir.

Parametre tahmini genel olarak rassal değişkenler için yapılıyorsa Bayes tahmin edici ile, rassal olmayan değişkenler için yapılıyor ise En Büyük Olabilirlik (Maximum Likelihood) tahmin edici ile gerçekleştirilmektedir (Barkat, 2005).

Bayes tahmin edici için tahmin ile gerçek değer arasındaki farktan dolayı bir $C(\theta, \hat{\theta})$ maliyet

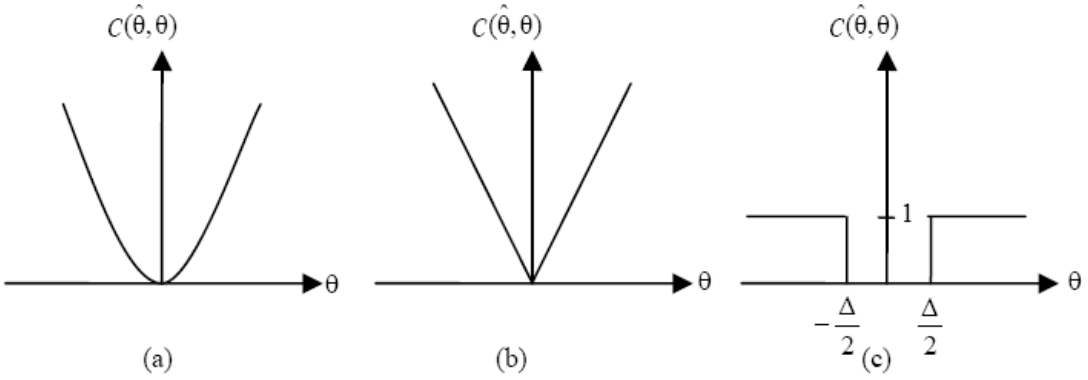
fonksiyonu ile maliyet fonksiyonun beklenen değeri olarak $\Re = E[C(\theta, \hat{\theta})]$ risk fonksiyonu tanımlanır. $\Re$ risk fonksiyonu minimize edilerek en iyi $\hat{\theta}$ tahmini yapılmış olur.

Maliyet fonksiyonu olarak denklem (4.60-4.62) ile verilen fonksiyonlar kullanılabilir.

Hataların karesi: $C(\theta, \hat{\theta}) = (\theta - \hat{\theta})^2$ (4.60)

Hataların mutlak değeri: $C(\theta, \hat{\theta}) = \left|\theta - \hat{\theta}\right|$ (4.61)

Düzgün maliyet fonksiyonu: $C(\theta, \hat{\theta}) = \begin{cases} 1, & \left|\theta - \hat{\theta}\right| \geq \dfrac{\Delta}{2} \\ 0, & \left|\theta - \hat{\theta}\right| < \dfrac{\Delta}{2} \end{cases}$ (4.62)



Şekil 4.14 a. Hatalar karesi, b. Hataların mutlak değeri, c. Düzgün maliyet fonksiyonu (Barkat, 2005)

Y rassal değişkenine ait Y_1, Y_2 ..., Y_K örnekleri için y_1, y_2, ...,y_K değerleri elde ediliyorsa ve $f(\theta \mid y)$ koşullu olasılık yoğunluk fonksiyonu olmak üzere, verilen maliyet fonksiyonları $\Re$ risk fonksiyonunda yerine konularak;

Hatalar karesi için: $\hat{\theta} = E[\theta \mid \mathrm{y}]$, ortalama (4.63)

Hataların mutlak değeri için: $\hat{\theta} = medyan(\theta)$ (4.64)

Düzgün maliyet fonksiyonu için: $\hat{\theta} = \mathrm{mod}(\theta)$ (4.65)

tahmin edicileri hesaplanabilir. Düzgün maliyet fonksiyonu ile hesaplanan tahmin ediciye En Büyük Sonsal (Maximum A Posteriori - MAP) tahmin edici denir.

En Büyük Olabilirlik tahmin edici verilen Y değişkeni için ve P(Y|Θ) olasılık modeli için $\hat{\theta}$

tahmin değerini denklem (4.66) ile bulur.

$$\hat{\theta} = \arg\max_{\Theta}(P(Y \mid \Theta)) = \arg\max_{\Theta}(\prod_{i=1}^{K} P(y_i \mid \Theta)) \tag{4.66}$$

Beklenti En Büyükleme, (Expectation Maximization - EM) algoritması sonuca En Büyük Olabilirlik ile ulaşan algoritmalardan biridir. EM algoritması şu adımlardan oluşur (Barkat, 2005):

Başlangıç: Θ için ilk değer atanır

Döngü: P(Y|X, Θ) hesaplanır

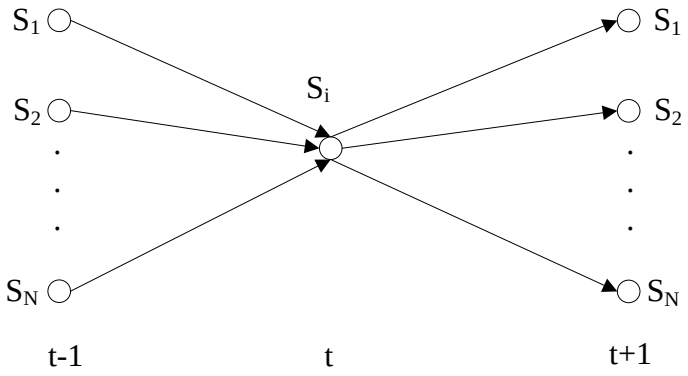
$\sum_{Y} P(Y \mid X, \Theta) \log P(X, Y \mid \hat{\Theta})$, $\hat{\Theta}$ değerine göre maksimize edilir

Baum-Welch algoritması En Büyük Olabilirlik ile parametre tahmini yapmaktadır ve Beklenti En Büyükleme algoritmasının adımlarının izlemektedir.

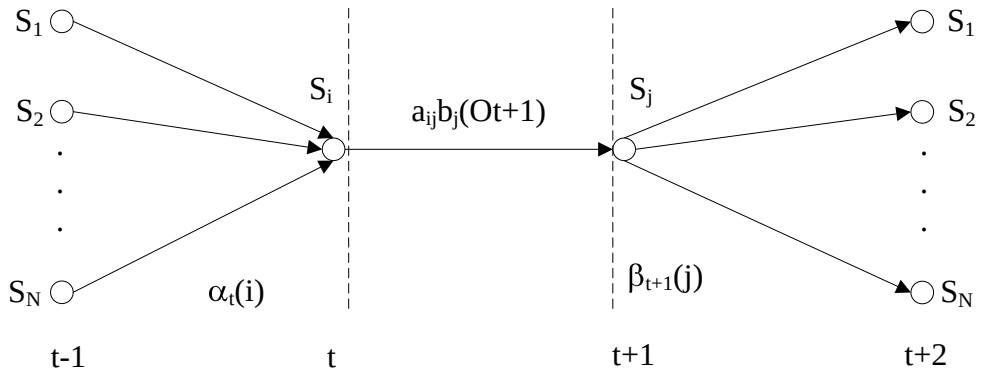
Baum-Welch algoritması için $\xi_t(i, j)$ ve $\gamma_t(i)$ değişkenleri tanımlanır. $\xi_t(i, j)$ verilen bir gözlem sırası ve başlangıç λ modeli için t anında i durumunda bulunulup t+1 anında j durumuna geçiş olarak tanımlanır ve denklem (4.67) ile ifade edilir. $\gamma_t(i)$ ise verilen model ve gözlem sırası için t anında i durumunda bulunulma olasılığı olarak tanımlanır ve denklem (4.68) ile ifade edilir.

$$\xi_t(i, j) = P(q_t = S_i, q_{t+1} = S_j \mid O, \lambda) \tag{4.67}$$

$$\gamma_t(i) = P(q_t = S_i \mid O, \lambda) \tag{4.68}$$



Şekil 4.15 $\gamma_t(i)$ değişkeni için şematik gösterim. (Rabiner, 1989)



Şekil 4.16 $\xi_t(i,j)$ değişkeni için şematik gösterim. (Rabiner, 1989)

$$\xi_t(i,j) = \frac{\alpha_t(i)a_{ij}b_j(O_{t+1})\beta_{t+1}(j)}{\sum_{i=1}^{N}\sum_{j=1}^{N}\alpha_t(i)a_{ij}b_j(O_{t+1})\beta_{t+1}(j)} \tag{4.69}$$

$$\gamma_t(i) = \sum_{j=1}^{N}\xi_t(i,j) = \frac{\alpha_t(i)\beta_t(i)}{\sum_{i=1}^{N}\alpha_t(i)\beta_t(i)} \tag{4.70}$$

$\gamma_t(i)$ değişkenini zaman indeksi üzerinden toplayarak S_i durumundan geçişlerin sayısı için bir ortalama değer hesaplanabilir. Benzer şekilde $\xi_t(i,j)$ değişkenini zaman indeksi üzerinden toplayarak S_i, S_j geçişleri sayısı için ortalama bir değer hesaplanabilir. $\bar{\lambda} = \{\bar{A}, \bar{B}, \bar{\pi}\}$ tahminleri için:

$\bar{a}_{ij}$: ($S_i \rightarrow S_j$ geçişleri sayısı beklenen değeri)/(S_i geçişleri sayısı beklenen değeri)

$\bar{\pi}_i$: İlk durum olarak S_i'nin görülmesi

$\bar{b}_j(k)$: (S_j durumunda v_k gözlem sembolünün üretilmesi)/(S_j durumundan geçişler)

$$\bar{a}_{ij} = \frac{\sum\xi_t(i,j)}{\sum\gamma_t(i)} \tag{4.71}$$

$$\bar{b}_j(k) = \frac{\sum_{t=1,O_t=v_k}^{T}\gamma_t(j)}{\sum_{t=1}^{T}\gamma_t(j)} \tag{4.72}$$

$$\bar{\pi}_i = \gamma_1(i) \tag{4.73}$$

İlk λ modeli ile başlayıp $\overline{\lambda}$ tahmini modeli hesaplanırsa ya $\lambda = \overline{\lambda}$ kritik değerine ya da $P(O \mid \overline{\lambda}) > P(O \mid \lambda)$ koşulunu sağlayacak şekilde daha iyi bir tahmine ulaşıldığı Baum ve arkadaşları tarafından (1972) çalışmaları ile gösterilmiştir.

Uygulamada model parametre tahmini ile ilgili karşılaşılan problemlerden birisi olan uzun gözlem dizileri ile birlikte α ve β değişkenlerinin art arda çarpımlar sonucu hesap duyarlılığının altında çok küçük değerler alması durumudur. Bu problemin önüne geçilebilmesi için her adımda α ve β değişkenleri denklem (4.74), (4.75) ile verilen ölçeklemeye tabi tutulmaktadır. α için kullanılan ölçekleme katsayıları β için de aynen uygulanabilir. Parametre tahmininde α ve β değişkenleri yerine $\hat{\alpha}$ ve $\hat{\beta}$ ölçeklenmiş değişkenleri kullanılabilir.

$$\hat{\alpha}_t(i) = \frac{\alpha_t(i)}{\sum_{i=1}^{N} \alpha_t(i)} \tag{4.74}$$

$$\hat{\beta}_t(i) = \frac{\beta_t(i)}{\sum_{i=1}^{N} \alpha_t(i)} \tag{4.75}$$

Model parametreleri için yukarıda bahsedilen tahmin ediciler tek bir gözlem dizisi için hesaplanan değerlerdir. Çoğu zaman için tek bir gözlem dizisi konuşma tanıma gibi işlerde tüm geçiş ve gözlem sembolleri gerçekleşme olasılıklarının tamamını eğitmeye yeterli olmamaktadır. Birden çok gözlem dizisi için parametre tahmini, gözlemlerin birbirinden bağımsız olduğu varsayımı ile herbiri T_m (m=1, ..., M) uzunlukta toplam M adet farklı gözlem dizisi için denklem (4.76-4.78) ile verilmektedir (Xiaolin, 2000).

$$\overline{a}_{ij} = \frac{\sum_{m=1}^{M} \sum_{t=1}^{T_m - 1} \xi_t^{(m)}(i, j)}{\sum_{m=1}^{M} \sum_{t=1}^{T_m - 1} \gamma_t^{(m)}(i)} \tag{4.76}$$

$$\overline{b}_j(k) = \frac{\sum_{m=1}^{M} \sum_{t=1, O_t^{(k)} = v_k}^{T_m} \gamma_t^{(m)}(j)}{\sum_{m=1}^{M} \sum_{t=1}^{T_m} \gamma_t^{(m)}(j)} \tag{4.77}$$

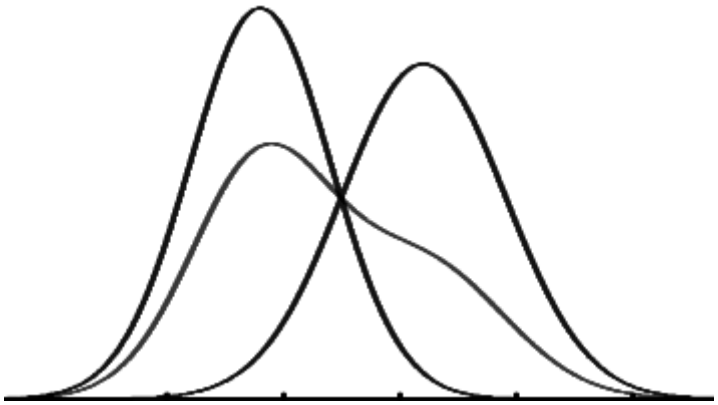
$$\bar{\pi}_i = \frac{1}{M}\sum_{m=1}^{M}\gamma_1^{(m)}(i) \tag{4.78}$$

Gözlem sembolleri gerçekleşme olasılığı $b_j(o_t)$ değişkenlerinin sürekli olasılık dağılımı fonksiyonu olarak modellenmesi durumu söz konusu olabilmektedir. $b_j(o_t)$'nin sürekli olduğu durumların modellenmesinde yaygın olarak katışımlı normal dağılım (Gaussian mixture) fonksiyonunun kullanımı söz konusu olmaktadır.

$$b_j(o_t) = \sum_{k=1}^{M} c_{jk} N\left(o_t, \mu_{jk}, \Sigma_{jk}\right) \tag{4.79}$$

Denklem (4.79) ile verilen tanımda c_{jk} ile j'ninci durum için kullanılan k'nıncı katışım modeline ait katsayı, $N\left(o_t, \mu_{jk}, \Sigma_{jk}\right)$ ile de o_t gözlem sembolü için μ_{jk} ortalama vektörü, Σ_{jk} kovaryans matrisi olan normal dağılımdan hesaplanacak olan olasılık değeri ifade edilmektedir.

Şekil 4.17 ile tek boyutlu iki katışımlı normal dağılıma ilişkin örnek verilmiştir.



Şekil 4.17 Tek boyutlu iki katışımlı normal dağılım örneği

Gözlem değerlerinin gerçekleşme olasılıklarının, bir sürekli olasılık dağılımı olarak modellenmesi ile tahmini gerçekleştirilecek parametrelere c, μ ve Σ değerleri de eklenir. Denklem (4.80) ile ayrık olasılık dağılımlı gözlem değerleri için hesaplanan $\gamma_t(i)$ değerinin genelleştirilmiş hali olan $\gamma_t(j,k)$ değeri verilmiştir. $\gamma_t(j,k)$ değerinden hareketle sürekli olasılık dağılımına sahip GMM modelleri için tahmini yapılacak diğer parametrelerin güncellenecek değerleri denklem (4.81-4.83) ile verilmiştir. c, μ ve Σ için verilen tahmin değerleri tek gözlem dizisine ilişkin olarak verilmiştir (Rabiner, 1989)

$$\gamma_t(j,k) = \left[\frac{\alpha_t(j)\beta_t(j)}{\sum_{j=1}^{N} \alpha_t(j)\beta_t(j)} \right] \left[\frac{c_{jk} N\left(o_t, \mu_{jk}, \Sigma_{jk}\right)}{\sum_{m=1}^{M} c_{jm} N\left(o_t, \mu_{jm}, \Sigma_{jm}\right)} \right] \tag{4.80}$$

$$\overline{c}_{jk} = \frac{\sum_{t=1}^{T} \gamma_t(j,k)}{\sum_{t=1}^{T} \sum_{k=1}^{M} \gamma_t(j,k)} \tag{4.81}$$

$$\overline{\mu}_{jk} = \frac{\sum_{t=1}^{T} \gamma_t(j,k) \cdot o_t}{\sum_{t=1}^{T} \gamma_t(j,k)} \tag{4.82}$$

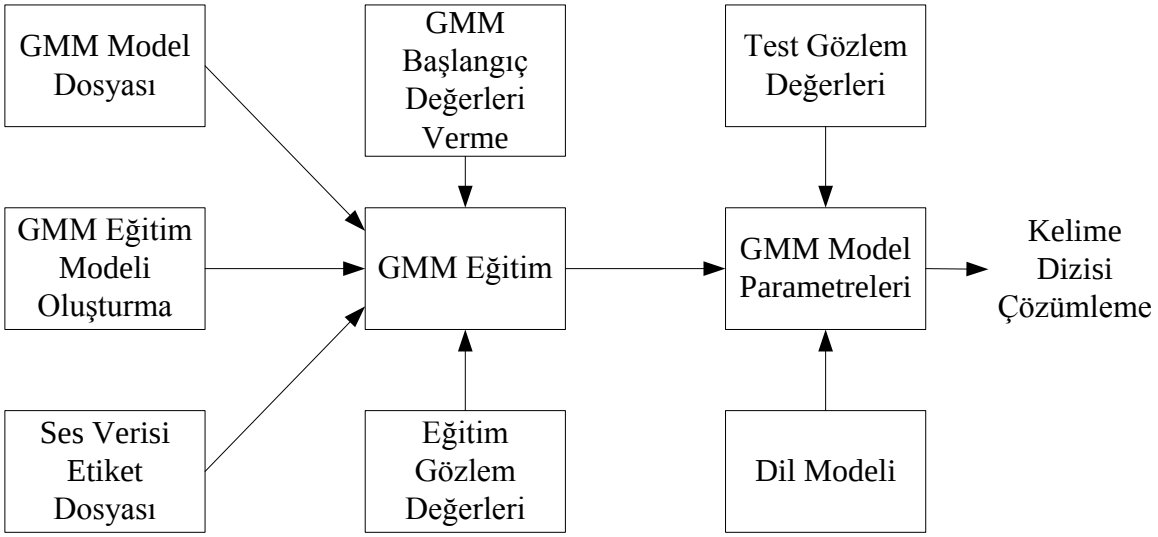
$$\overline{\Sigma}_{jk} = \frac{\sum_{t=1}^{T} \gamma_t(j,k)(o_t - \mu_{jk})(o_t - \mu_{jk})^t}{\sum_{t=1}^{T} \gamma_t(j,k)} \tag{4.83}$$

4.3.3 Gizli Markov Modeli'nin Geniş Sözlüklü Sürekli Konuşma Tanımada Kullanımı

Geniş sözlüklü sürekli konuşma tanımada parçalı sesbirimlerinin kullanımı gerek tanıma uzayını esnek bir şekilde genişletebildiği, gerekse eğitim örnek sayısı kısıtlarını kolayca aşabildiği için uygun düşmektedir. (Juang, Rabiner, 1989). Parçalı sesbirimleri yani fon/fonemlerin konuşma tanımada kullanımı için herbir fon/fonem için benzer temel GMM tanımı yapılır. Genel olarak temel GMM'ler dört durumlu soldan sağa modeller olarak tanımlanır.

Eğitim aşamasında herbir eğitim örneğine karşılık bir GMM oluşturulmaktadır. Herbir eğitim örneği için ilgili fon çözümlemesinde geçen fona ilişkin temel GMM'ler birbirleri arası geçişler sesler arası geçiş sayısına göre olasılıklandırılarak bir araya getirilir. Tüm eğitim örnekleri için oluşturulan GMM'ler üzerinde Baum-Welch algoritması çoklu gözlem durumu dikkate alınarak uygulanır. Sonuçta herbir fona ilişkin temel GMM parametreleri güncellenmiş olarak elde edilmektedir.

Şekil 4.18 ile GMM için eğitim ve test aşamaları blok şema halinde verilmiştir.



Şekil 4.18 GMM eğitim ve test aşamaları blok yapısı

Test ve tanıma aşamasında güncel değerli temel GMM'ler kurgulanan test modeli çerçevesinde biraraya getirilerek tanıma arama uzayı oluşturulur. Verilen bir gözlem dizisini en yüksek olasılıkla sağlayan yol üzerindeki fonlar tanıma sonucu olarak elde edilir.

4.4 Dil Modeli

Dil modeli ile amaçlanan söz dizisinde k. kelime w_k'nın gerçekleşme olasılığının kendinden önceki $W_1^{k-1} = w_1, w_2, ..., w_{k-1}$ kelime dizisine bağlı olarak ortaya konmasıdır. Kelimenin gerçekleşme olasılığının kendinden önceki n-1 kelimeye bağlı olarak ortaya konması N-gram dil model, olarak adlandırılır ve denklem (4.84) ile ifade edilir.

$$P(w_k \mid W_1^{k-1}) = P(w_k \mid W_{k-n+1}^{k-1}) \tag{4.84}$$

N-gram için olasılık değerleri doğrudan eğitim setinden frekans sayımları ile hesaplanabilmektedir. N-gramların oluşturulmasında büyük sözlükle çalışılıyor olması bazı N-gramların eğitim setinde gözlenememesi problemini doğurabilir. Bu durumda gözlenen N-gram olasılıklarına olduğundan küçük olasılıklar atanarak kalan artık olasılıklar gözlenemeyen N-gramlara paylaştırılabilir. Bazı durumlarda ise eğitim seti anlamlı N-gram oluşturmaya yeterli olmamaktadır, bu durumda ise daha düşük kelime sayılı dil modellerinden ağırlıklandırarak istenen N-gramlar elde edilmiş olur.

Test aşamasında oluşturulan arama uzayları tüm kelimeler arası geçişleri belirlediği için test modelleri üzerinden bir N-gram dil modeli yapısı oluşturulduğu söylenebilir.

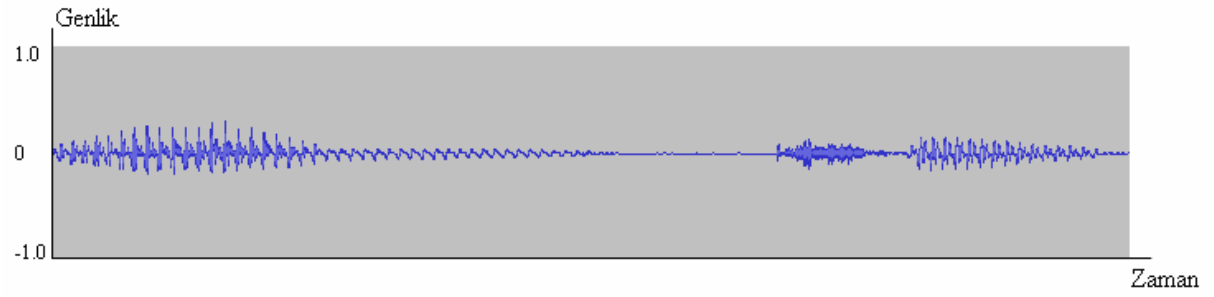
5. UYGULAMA

Uygulama aşamasının, ses sinyalinin sayısallaştırılmasından başlayarak önişlemler, özellik çıkarma, GMM model oluşturma, GMM eğitim, dilbilgisi ya da dil modeli oluşturma ve GMM çözümleme süreçlerinin gerçekleştirilme adımları bu başlık altında anlatılacaktır. Ayrıca farklı uygulama modelleri üzerinden elde edilen deneysel sonuçlara ve tanıma başarı oranlarına da yer verilecektir.

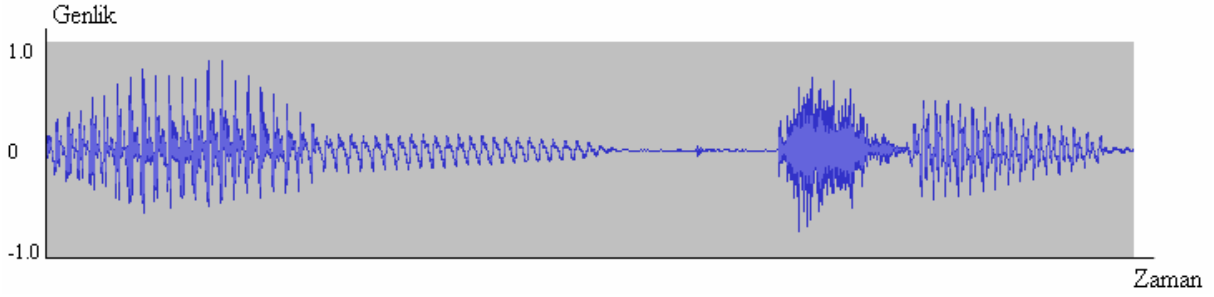
Eğitim ve test aşaması için kullanılan İngilizce ses örnekleri CMU tarafından 1991 yılında kaydedilmiş olan AN4 veritabanından alınmıştır [**5**]. Bu veritabanında kişilerden isim, adres, telefon numarası ve doğum günü gibi kişisel verileri sesli söylemeleri istenmiş ve bu şekilde 16 kHz'te 16 bit çözünürlükle ses verisi sayısallaştırılarak kaydedilmiştir. Türkçe konuşma tanıma deney kurgusunda kullanılan ses verileri 16 kHz'te 16 bit çözünürlükte sayısallaştırılmış bir konuşmacıya ait ses örnekleridir.

5.1 Ön İşlemler

Ses verisi üzerinde işlem yapılırken farklı ortamlarda, farklı kişilerden kayıtlar yapıldığı göz önüne alınarak ses verisi -1 ile 1 arasına normalize edilerek kullanılmıştır. Böylelikle tüm ses örnekleri için en yüksek genlik değerleri eşitlenmiş olmaktadır.



(a)



(b)

Şekil 5.1 a. Ön-vurgulama ve normalizasyon öncesi, b. sonrası ses verisi

Yüksek frekanslı özelliklerin ortaya çıkarılması için (4.5) eşitliğindeki ön-vurgulama işlemi a=0.95 alınarak uygulanmıştır. Ön-vurgulama ve normalizasyon öncesi ve sonrası ses verisindeki değişikliğe ilişkin bir örnek yatay eksen zamanı düşey eksen genlik değerlerini göstermek üzere şekil 5.1 ile verilmiştir.

Sese ilişkin kısa süreli durağan frekans özelliklerinin ortaya çıkarılması için 256 örneklik denklem (4.8) ile ifade edilen Hamming pencere fonksiyonları 128'er örnek kaydırılarak uygulanmıştır. 16 kHz'lik sinyal dikkate alındığında 256 örnek 16 ms'lik ses sinyaline karşılık gelmektedir.

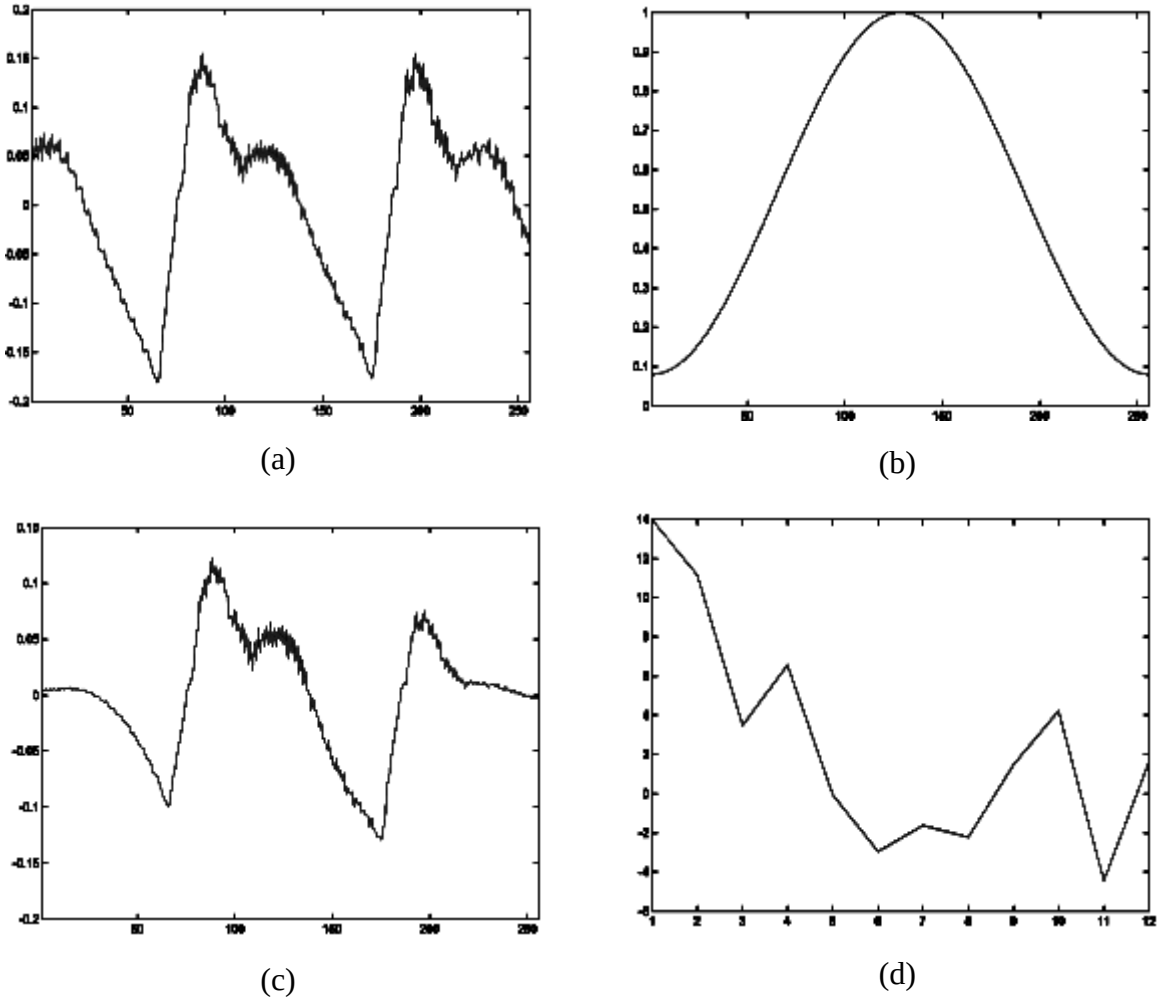
5.2 Özellik Çıkarımı

Pencereleme fonksiyonu sonucu elde edilen 256 örneklik özellik vektörüne FFT uygulanarak frekans boyutunda simetrik 256 örneklik yeni özellik vektörü elde edilir. Simetrik yapı dikkate alınarak özellik vektörünün gereksiz ikinci yarısı özellik vektöründen çıkarılır. Böylelikle frekans boyutunda 128 örneklik özellik vektörü elde edilir. Frekans boyutunda elde edilen özellikler karmaşık değerlikli olmakla birlikte sahip oldukları genlik özellikleri hesaplanarak özellik değerleri olarak kullanılır.

İnsan kulağındaki bazal zara benzetim sağlamak için 0-1000 Hz arası 10 adet lineer eşit aralıklı, 1-8 kHz arası mel ölçeğinde 15 adet logaritmik eşit aralıklı üçgen pencerelerle filtre bankası oluşturularak 128 örneklik özellik vektörüne uygulanır. Sonuçta boyutu 25 olan bir özellik vektörü elde edilir.

Özellik değerlerinin logaritmaları alınarak özellikler istatistik olarak normal dağılıma yaklaştırıldıktan sonra ayrık kosinüs dönüşümü uygulanarak 25 adet simetrik yapıda MFKK katsayıları elde edilir. Simetrik yapı dikkate alınarak son 12 MFKK katsayısı ile DC özellikleri içerdiği için ilk MFKK katsayısı özellik vektöründen çıkarılarak 12 boyutlu MFKK özellik vektörü elde edilir. Şekil 5.2 ile bir pencere genişliğindeki ses verisi, uygulanacak pencereleme fonksiyonu, pencereleme sonucu ve pencereye ilişkin hesaplanan MFKK katsayıları verilmiştir.

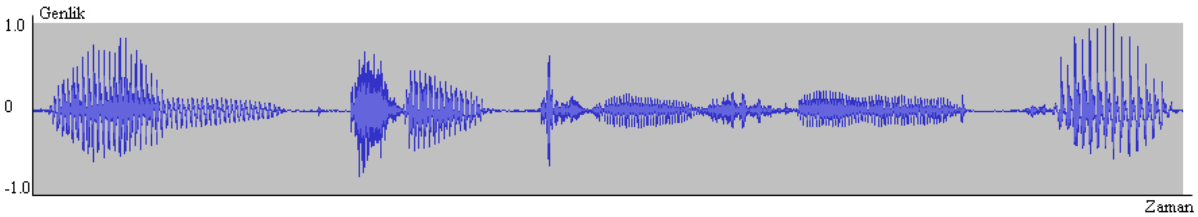
Şekil 5.3 ile ise yaklaşık 2 sn'lik bir ses verisinden elde edilen MFKK katsayıları; pencerelere, MFKK katsayı sırasına ve genlik değerine göre grafiksel olarak gösterilmiştir.



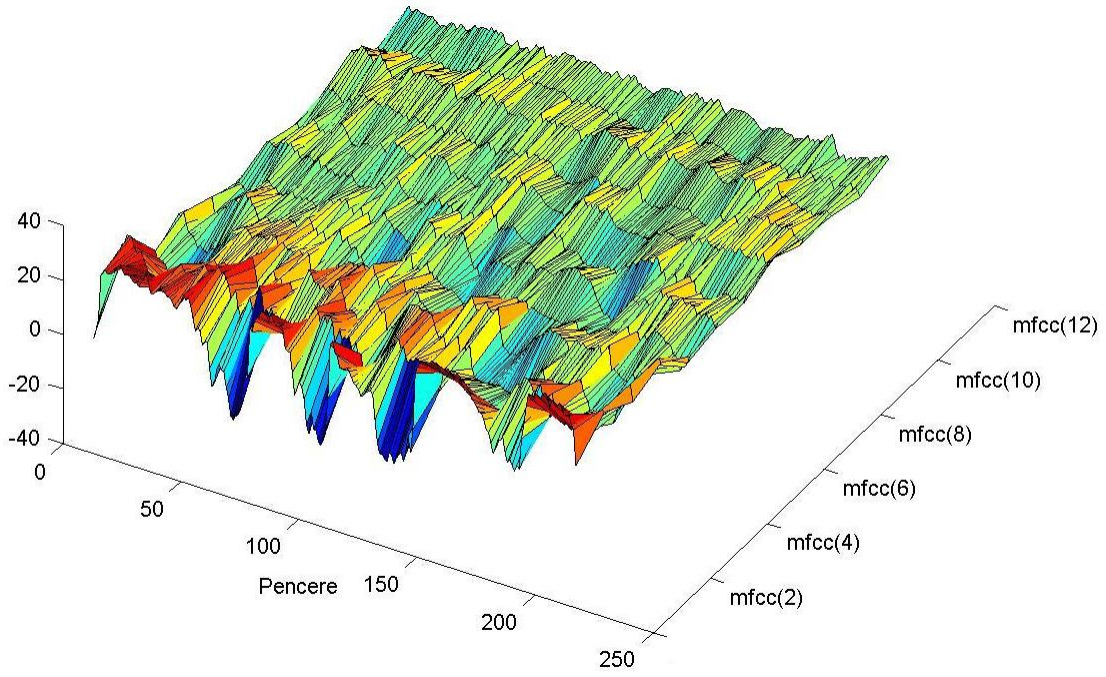
(a) (b)

(c) (d)

Şekil 5.2 a. Bir pencere genişliğinde ses verisi, b. Uygun boyutlu Hamming pencere fonksiyonu, c. Ses verisinin pencere fonksiyonundan geçirilmiş hali, d. İlgili pencere için hesaplanmış MFKK katsayıları



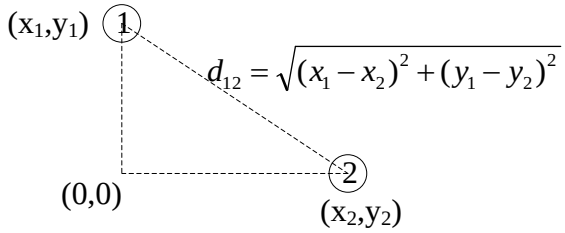
(a)



(b)

Şekil 5.3 a. Ses verisi, b. Ses verisine ilişkin hesaplanmış MFKK katsayıları

Hesaplanan MFKK katsayıları 12 boyutlu bir vektör olarak elde edilir ve bunlardan herbiri reel değerlikli özelliklerdir. Uygulamada bu reel değerli 12 boyutlu vektörler K-ortalama algoritması ile daha önceden belirlenmiş merkezlerden kendisine en yakın olana atanarak, ayrık bir gözlem değeri alır.

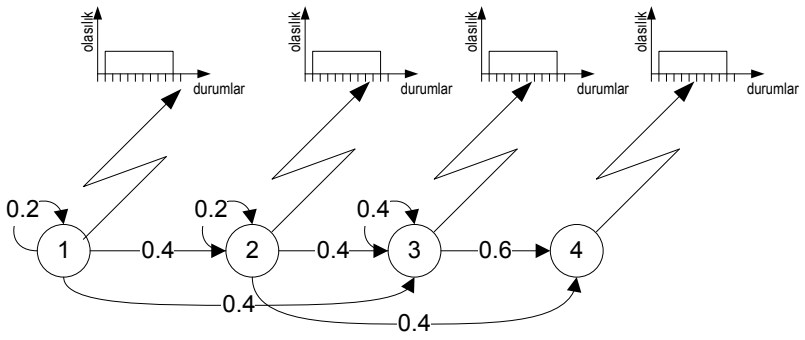


Şekil 5.4 İki boyutlu bir uzayda Öklid uzaklığı

Merkez vektörlerinin belirlenmesi K-ortalama eğitim aşaması ile gerçekleştirilir. GMM modelinin eğitiminde kullanılacak olan ses sayısının yarısı kadar öbek oluşturacak şekilde K-ortalama algoritması için k parametresi belirlenerek, Öklid uzaklığı (şekil 5.4) temel alınarak güdümsüz eğitim gerçekleştirilir. Sonuçta ses (fon veya fonem) sayısının yarısı kadar merkez vektör bulunmuş olur. Bu noktadan itibaren akustik modelin oluşturulması ve GMM modelinin eğitiminden bahsedilecektir.

5.3 Model Oluşturma ve Eğitim

GMM modeli fon veya fonemler baz alınarak oluşturulmaktadır. Her bir fon veya fonem başlangıç, gövde, çıkış ve bağlantı olmak üzere 4 durumlu GMM modeline sahiptir. Bu temel GMM modeli şekil 4.11 b ile verilen soldan sağa modelidir. Fon veya fonemler için oluşturulan temel GMM modelleri A parametresine ilişkin ilk koşulları atanmış halleriyle şekil 5.5 ile verilmiştir. B parametresine ilişkin olarak ise ilk değer ataması herbir durumdan tüm gözlem değerlerine eşit olasılık verecek şekilde olmaktadır. Şekil 5.5 ile verilen ilk değerler yapılan denemeler sonucu en başarılı tanıma oranını sağlayan değerlerdir.



Şekil 5.5 Fon veya fonem için temel GMM model parametreleri ilk durum ataması

Fon veya fonem bazında tüm temel GMM modelleri oluşturulup ilk değerleri atandıktan sonra eğitim seti metinsel kodlaması doğrultusunda bu modellerin uygun şekilde birleştirilmesi gerekmektedir.

Eğitim seti sürekli konuşma içeren ses dosyalarından oluşmaktadır. Herbir ses dosyasında birbirlerinden farklı rakam dizileri, farklı kişilerce söylenmiş halleriyle kayıtlı bulunmaktadır. Ayrıca ses dosyalarına ilişkin cümle bazında çözümlemeler de metin olarak eğitim setine dahildir.

GMM model eğitiminde tek bir eğitim seti tüm fon veya fonemlere ilişkin model parametrelerinin eğitimini sağlayamayacağı için çok sayıda gözlem örneği ile eğitim yapılmalıdır. Bu amaçla da tüm gözlem değerlerini birden gerçekleyecek model parametreleri hesabı bir seferde yapılmalıdır.

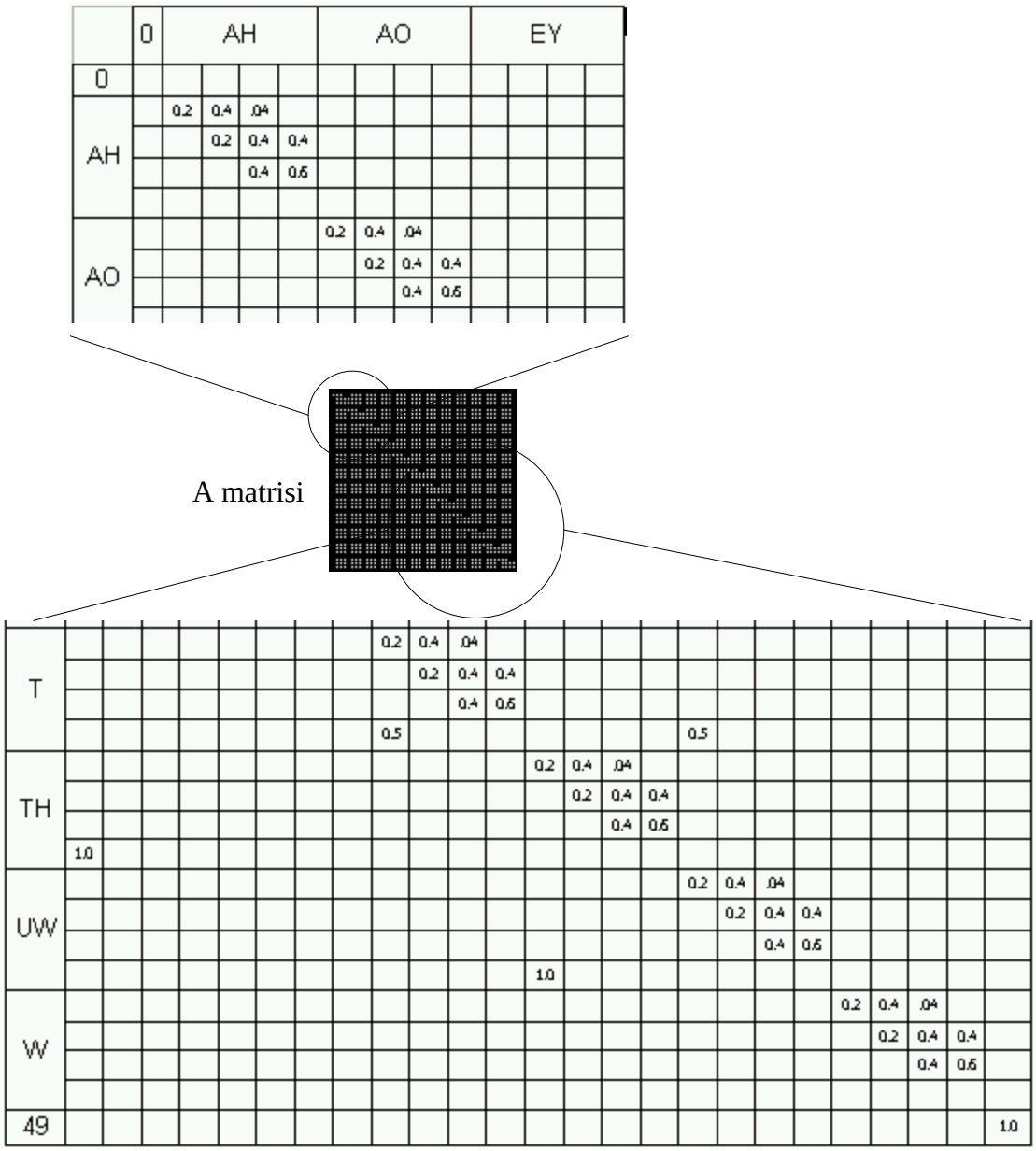
Yukarıda bahsedilen durumu gerçeklemek amacıyla herbir eğitim örneğine ilişkin genel GMM modeli eğitim örneğinde geçen herbir fon veya fonemi içerecek şekilde, birbirini takip eden fon veya fonemler arası geçişler uygun şekilde olasılıklandırılarak oluşturulur. Ayrıca eğitim örneğinin tümü için yapay bir başlangıç ve bir çıkış durumu eklenip bu özel durumlar

için özel gözlem değerleri tanımlanarak tüm modelin takip etmesi gerek durum dizisi koşullandırılmaktadır.

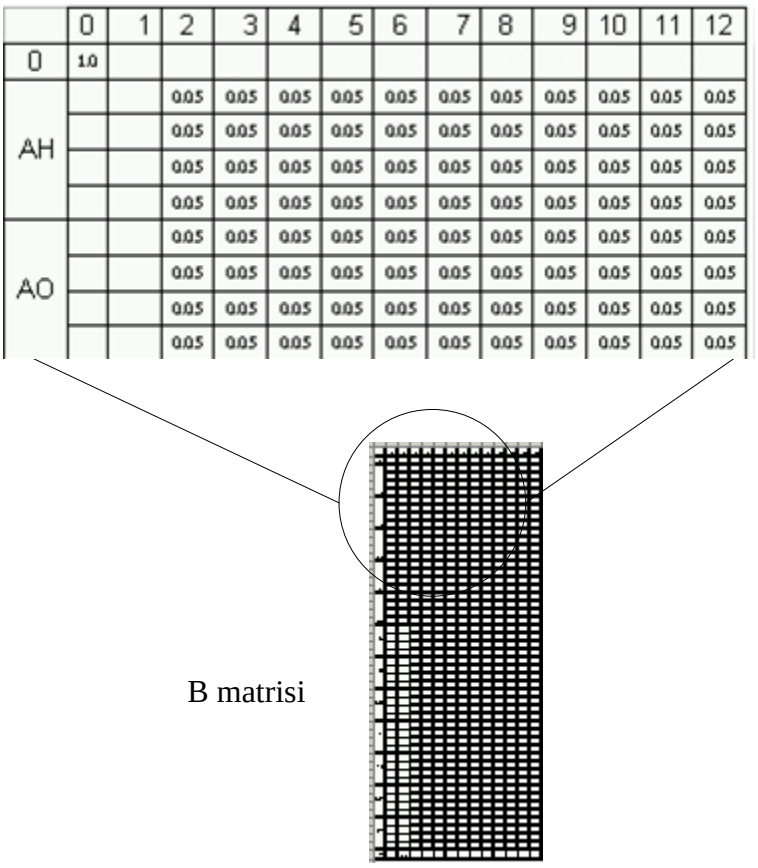
<S> ONE FOUR EIGHT TWO THREE </S>

SIL W AH N F AO R EY T T UW TH R IY SIL

Şekil 5.6 Bir eğitim modeli için kelime bazında ve fon bazında çözümleme verisi



(a)



(b)

Şekil 5.7 İlk koşulları verilmiş genel GMM için a. A matrisi, b. B matrisi

Bir eğitim örneğinden oluşturulacak genel GMM modeli için hangi fon veya fonem temel modellerinin bir araya getirileceği eğitim örneğine ilişkin kelime bazında çözümlenmiş metinden hareketle, bir telaffuz sözlüğünden kelimelere karşılık gelen fon-fonem çözümleme ile belirlenir. Şekil 5.6 ile kelime ve fon bazında çözümleme örneği verilmiştir.

Şekil 5.7 ile bir eğitim örneğine ilişkin ilk koşulları verilmiş genel GMM modelinde durumlar arası geçiş matrisi A'nın ve gözlem sembolleri gerçekleşme olasılığı matrisi B'nin nasıl olasılıklandırıldığı gösterilmiştir. A matrisinde diyagonal olasılıklar ses içi diğer geçişler sesler arası geçişi göstermektedir.

Temel GMM'lerin birleştirilmesinden oluşturulan cümle için GMM modeli Baum-Welch algoritması ile eğitilirken iki yöntem izlenebilir. Bunlardan ilki hem A hem de B parametrelerinin birlikte güncellenmesi yöntemidir. İlk yöntem ile karşılaşılan bir problem ise genel olarak eğitim örneğinde kelimelere ilişkin başlangıç ve bitiş noktaları işaretlenmeden kullanılacağı için, gözlem sembolü olarak en çok karşılaşılan değerlerin gerçekleşme olasılıklarına, tüm durumlarda, diğer gözlem sembollerine göre daha yüksek değer

atanmasıdır. Uygulanabilecek ikinci yöntem ise A parametresinin atanan başlangıç değerleri sabit tutularak sadece B parametrelerinin güncellenmesidir.

Bahsedilen yöntemlere ilişkin elde edilen bir eğitim sonucu çizelge 5.1 ile verilmiştir. Bu örnekte giriş ve çıkış durumlarının da katılmasıyla oluşturulmuş durum sayısı N=30 ve gözlem sembol sayısı M=23 değerlerine sahip GMM modeli için, satırlar durum numaralarını göstermek üzere, eğitim örneğinde en çok gözlenen 6. sembolün, eğitime ilişkin uygulanan farklı yöntemler sonucu elde edilen B olasılıkları verilmiştir.

Çizelge 5.1 Gözlem sembolü gerçekleşme olasılığı

	A sabit	A değişken
Durum	6. gözlem sembolü	6. gözlem sembolü
0	0,00	0,00
1	0,00	0,23
2	0,00	0,23
3	0,00	0,23
4	0,00	0,23
5	0,00	0,23
6	0,00	0,23
7	0,00	0,23
8	0,00	0,25
9	0,51	0,25
10	0,00	0,26
11	0,00	0,25
12	0,00	0,26
13	0,46	0,26
14	0,28	0,26
15	1,00	0,26
16	0,32	0,13
17	0,00	0,13
18	0,00	0,13
19	0,00	0,13
20	0,00	0,26
21	0,30	0,26
22	0,00	0,26
23	0,92	0,26
24	0,50	0,23
25	0,00	0,23
26	0,00	0,23
27	0,00	0,23
28	0,11	0,23
29	0,00	0,00

Her iki eğitim yaklaşımı ile de gözlem dizisi için hesaplanacak olan gerçekleşme olasılığı aynı olsa da A'nın sabit tutularak B parametresinin güncellenmesi, belirli durumların belirli gözlem sembollerine bağlanmasını sağladığı ve durumlardan ilgisiz gözlem sembollerine çıkış olasılıklarını sıfırladığı için tercih edilmiştir.

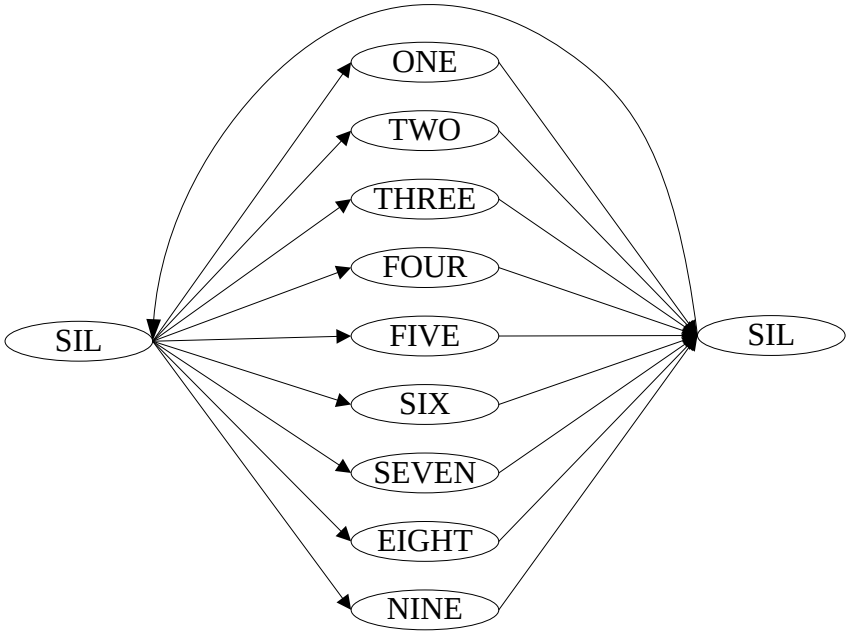
Çok sayıda eğitim örneğinden elde edilen gözlem dizilerinin, A matrisinin sabit tutulup, B matrisinin Baum-Welch ile eğitim sonucu güncellenmesi ile tüm fon veya fonemlerde durumlardan gözlem sonucu gerçekleşme olasılıkları hesaplanmış olmaktadır.

Tüm fon veya fonemler için elde edilen eğitim sonuçları test aşamasında kullanılmak üzere uygun görülen formatta veri olarak saklanmaktadır.

5.4 Test ve Tanıma

Test aşaması için, fon veya fonemlere ait temel GMM'ler bir dil modeli çerçevesinde bir araya uygun sırayla getirilerek eldeki test gözlem dizisi için en olası yol bulunmaktadır.

Test aşamasına ilişkin örnek bir dil modeli şekil 5.8 ile verilmiştir. Bu dil modeli sıfırdan dokuza kadar olan rakamların daha önceden kısıtlanmamış sayıda tekrarlandığı durumu tanımaya yöneliktir.



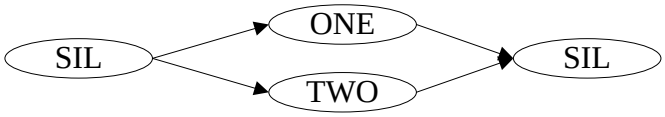
Şekil 5.8 Sıfırdan dokuza kadar tekrarlı rakamlar için dil modeli

Verilen dil modeline göre herbir yol için fon veya fonem çözümlemesi yapıldıktan sonra

uygun temel GMM modelleri sırayla test GMM modeli oluşturmak için bir araya getirilir. Fon veya fonemler arası geçişler uygun şekilde olasılıklandırıldıktan sonra Viterbi algoritması ile test gözlem sembolleri dizisini en iyi sağlayan yol, yani kelime dizisi ortaya konabilir.

Test amaçlı olarak çeşitli deney kurguları üzerinden sistemin başarısı irdelenmeye çalışılmıştır. Bu amaçla iki adet rakamın ayrık olarak söylendiği durum, birden dokuza kadar rakamların ayrık söylediği durum iki kelimeli sürekli rakam telaffuzu ve yine birden dokuza kadar olan rakamların tekrarlı söylendiği durum deney kurguları olarak denenmiştir. Her deney kurgusunda 30 adet sürekli konuşma yapısında rakam telaffuzu içeren eğitim seti ile eğitilmiş GMM modeli kullanılmıştır.

5.4.1 İki Rakamlı Ayrık Konuşma Deney Kurgusu



Şekil 5.9 İki rakam ayrık telaffuzu için test modeli

Şekil 5.9 ile iki rakamlı ayrık telaffuz deney kurgusu için kullanılan test GMM modeli verilmiştir. Bu test modelinde ses verisinin sessiz bir kısım ile başladığı daha sonra “one” veya “two” kelimelerinden birisinin telaffuz edileceği son olarak da sessiz bir kısmın daha geleceği koşullandırılmaktadır. 10 adet “one” ve 10 adet “two” telaffuzu ile yapılan test sonucunda elde edilen sınıflandırma sonuçları çizelge 5.2 ile verilmiştir. 20 adet test örneği ile iki rakamlı ayrık tanıma modeli ile sınıflandırılması sonucu tanıma başarısı %75 olarak elde edilmiştir. Test örneklerinin sürekli konuşma yapısındaki ses veritabanından ilgili telaffuzlar kesilerek oluşturulmuş olduğu gözönüne alınırsa, ayrık konuşma tanımaya uygun olarak hazırlanmış bir ses veri tabanı ile daha yüksek bir başarı elde edilmesi beklenmektedir.

Çizelge 5.2 20 adet test örneğinin iki rakamlı ayrık rakam tanıma modeli ile sınıflandırma sonuçları

		Sınıflandırma Sonucu	
		“one”	“two”
Gerçek Sınıf	“one”	**9**	1
	“two”	4	**6**

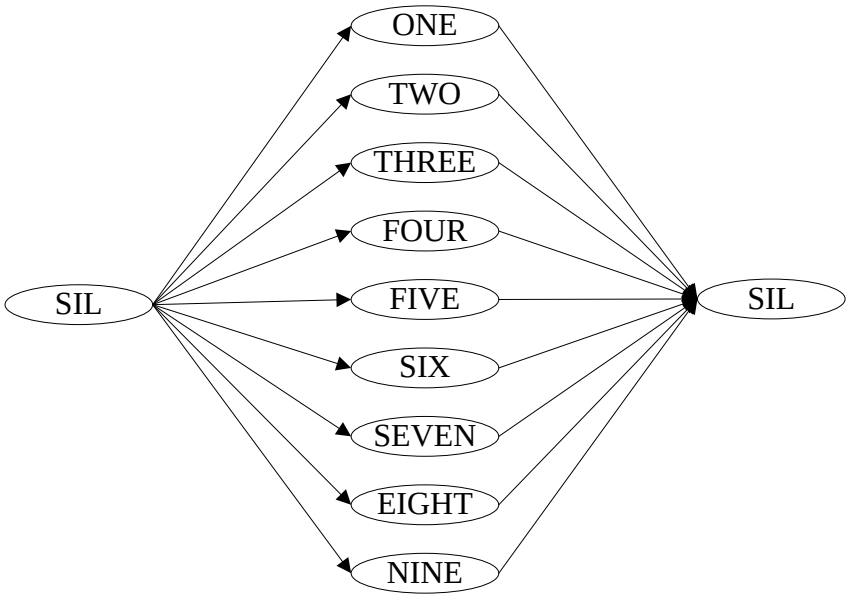
Şekil 5.9 ile verilen test modelinin başarısı k=5 için k katlı çapraz geçerlilik sınaması ile test edildiğinde ise çizelge 5.3 ile verilen sonuç elde edilmiştir. Ortalama tanıma başarısı %85 olarak gerçekleşmiştir.

Çizelge 5.3 İki rakamlı test modeli için k-katlı çapraz geçerlilik sınaması sonuçları

	Doğru Sınıflandırılan		**Hatalı Sınıflandırılan**		**Toplam Doğru**
	"one"	*"two"*	*"one"*	*"two"*	
Sınama 1	10	6	0	4	16
Sınama 2	10	8	0	2	18
Sınama 3	10	7	0	3	17
Sınama 4	9	8	1	2	17
Sınama 5	9	8	1	2	17
Genel Toplam	100				85

5.4.2 Dokuz Rakamlı Ayrık Konuşma Deney Kurgusu

Şekil 5.10 ile birden dokuza kadar olan rakamların ayrık söylendiği durum için bir test modeli verilmiştir. Bu modelde başta sessiz bir kısım sonra birden dokuza kadar herhangi rakam ve son olarak yine sessiz bir kısım daha koşullanmıştır.



Şekil 5.10 Birden dokuza kadar olan rakamların ayrık telaffuzu için test modeli

Verilen test modeli çerçevesinde birden dokuza kadar olan her bir rakam için 5 adet test örneği ile sistem sınıflandırması sonucu çizelge 5.4 ile verilmiştir. Toplam 45 adet test örneğinin sınıflandırılması sonucu sistem başarısı %60 olarak gerçekleşmiştir.

Çizelge 5.4 45 adet test örneğinin dokuz rakamlı ayrık rakam tanıma modeli ile sınıflandırma sonuçları

		Sınıflandırma Sonucu								
		“one”	“two”	“three”	“four”	“five”	“six”	“seven”	“eight”	“nine”
Gerçek Sınıf	“one”	**4**			1					
	“two”	1		3				1		
	“three”			**4**	1					
	“four”	1			**2**			2		
	“five”					**4**	1			
	“six”			1			**4**			
	“seven”	1						**3**		1
	“eight”	1		1					**3**	
	“nine”				1	1				**3**

Şekil 5.10 ile verilen model k=5 için k katlı çapraz geçerlilik sınaması ile değerlendirildiğinde başarı %55olarak gerçekleşmiştir. 45 örnek için her bir sınamada doğru sınıflandırılan kelime sayıları sırayla 22, 25, 20, 30, 24 olarak gerçekleşmiştir.

Türkçe ayrık rakam tanıma amaçlı deneysel bir uygulama ise uygun telaffuz sözlüğü oluşturularak gerçekleştirilmiştir. Bu deneysel uygulamada şekil 5.10’deki yapıya benzer Türkçe rakamlar için bir test modeli oluşturulmuştur. Sistem her rakamdan 2 adet olmak üzere toplam 18 adet örnekle test edilmiştir ve doğru sınıflandırma oranı %77 olarak gerçekleşmiştir.

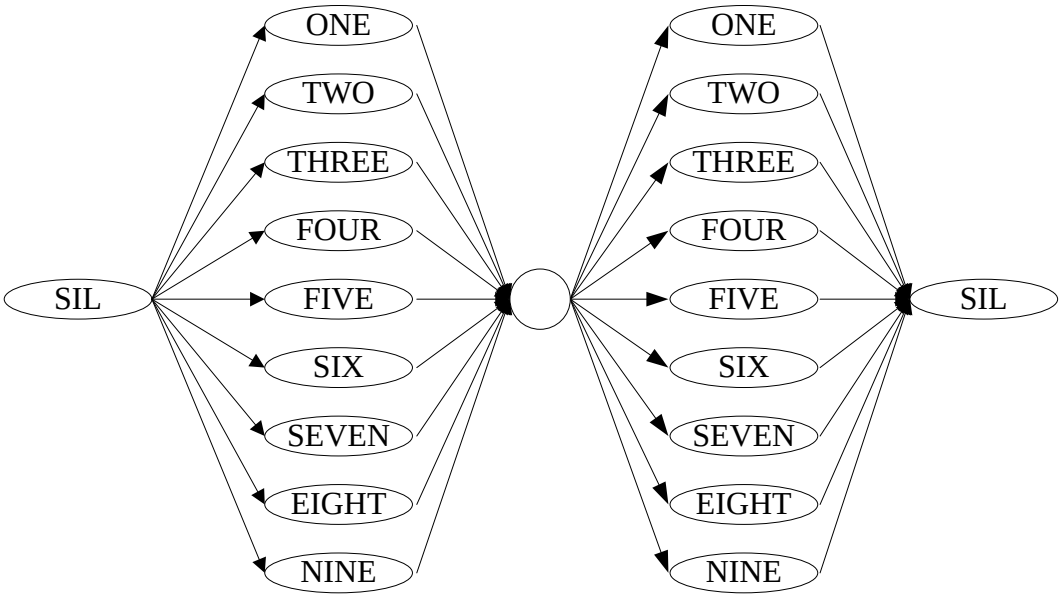
5.4.3 Kısıtlı Sayıda Tekrarlı Sürekli Konuşma Deney Kurgusu

Ayrık konuşma tanımada başarı ölçütü olarak doğru olarak çözümlenen kelime oranı verilebilirken, sürekli konuşma tanıma için iki ayrı hata ölçütü verilebilir; bunlar cümle ve kelime boyutunda tanıma hatası olarak isimlendirilir ve denklemler (5.1-5.2) ile tanımlanır (Jurafsky, D., Martin, J. H., 1999).

$$\text{kelime hata oranı}=100\times\frac{\text{ekleme+silme+hatalı}}{\text{doğru çözümleme toplam kelime sayısı}} \tag{5.1}$$

$$\text{cümle hata oranı}=100\times\frac{\text{en az bir hatalı çözümleme içeren cümle sayısı}}{\text{toplam cümle sayısı}} \tag{5.2}$$

Şekil 5.11 ile verilen kısıtlı iki rakamlı sürekli konuşma tanımaya yönelik bir test modelidir. Bu test modeli ile 31 adet ikili rakam telaffuzu içeren sürekli konuşma yapısındaki cümle test edilmiştir. Test sonuçları çizelge 5.5 ile verilmiştir. Test sonucunda cümle hata oranı %61, ortalama kelime hata oranı ise %40 olarak gerçekleşmiştir.

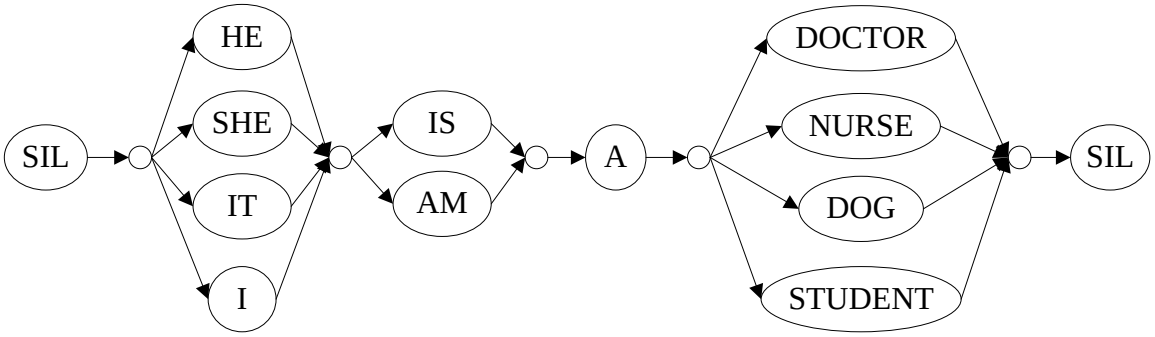


Şekil 5.11 İki rakamlı sürekli konuşma tanıma test modeli

Çizelge 5.5 İki rakamlı süreli konuşma tanıma sonuçları

Bir kelime hatalı		İki kelime de hatalı	İki kelime de doğru
İlk kelime hatalı	İkinci kelime hatalı		
4	9	6	12
13			

Şekil 5.12 ile örnek bir sürekli konuşma cümle test modeline yer verilmiştir. Bu deneysel uygulamada GMM parametrelerinin eğitimi 34 adet sürekli konuşma yapısında cümle ile gerçekleştirilmiştir.



Şekil 5.12 Cümle test modeli.

Çizelge 5.6 ile test örneklerinin gerçek çözümlemeleri ile elde edilen çözümlemelere yer verilmiştir.

Çizelge 5.6 Sürekli konuşma cümle test model sonuçları

Test 1	SIL	HE	IS	A	DOCTOR	SIL
Sonuç 1	SIL	HE	IS	A	DOCTOR	SIL
Test 2	SIL	SHE	IS	A	NURSE	SIL
Sonuç 2	SIL	SHE	IS	A	***STUDENT***	SIL
Test 3	SIL	IT	IS	A	DOG	SIL
Sonuç 3	SIL	IT	IS	A	DOG	SIL
Test 4	SIL	I	AM	A	STUDENT	SIL
Sonuç 4	SIL	I	AM	A	STUDENT	SIL
Test 5	SIL	SHE	IS	A	DOCTOR	SIL
Sonuç 5	SIL	SHE	IS	A	DOCTOR	SIL

6. SONUÇLAR VE ÖNERİLER

Tez çalışmasında çok boyutlu bir tanım uzayına sahip olan konuşma tanıma probleminin geniş sözlüklü sürekli konuşma tanıma gereksinimlerini karşılayacak şekilde çözümüne yönelik bir uygulama gerçekleştirilmiştir. Mel Frekans Kepstral Katsayı (MFKK) özellik çıkarımı ve Gizli Markov Modeli (GMM) ile özellik sınıflandırılması bu çalışmanın temel adımlarını oluşturmaktadır.

MFKK özellikleri konuşmacı bağımsız olarak ses verisine ilişkin özelliklerin ortaya konmasında etkili bir yöntemdir. GMM de ardışıl özelliklerin geliş sırası dikkate alınarak sınıflandırılmasını sağlayan bir yöntemdir.

MFKK özellik çıkarım adımlarının ses verisine uygulanması ile reel değerlikli özellik vektörleri elde edilir. MFKK katsayılarına K-ortalama yönteminin uygulanması ile tek boyutlu ayrık değerlikli bir özellik uzayına geçilir.

Sistem eğitim aşamasında girdi olarak eğitim ses verisi ile ses verisine ilişkin kelime boyutunda çözümleme verisinin kabul eder, çıktı olarak fonem temelli GMM'ler için model parametrelerini üretir. Herbir fon için dört durumlu, soldan sağa, ayrık çıkış olasılıklarına sahip GMM temel modelleri uygun ilk durumları verilerek oluşturulur. Eğitimde kullanılan ses verisine ilişkin çözümleme doğrultusunda temel GMM'ler bir araya getirilerek sesler arası geçiş sayılarına göre olasılıklandırılırlar. Tüm eğitim seti üzerinde Baum-Welch algoritması çoklu gözlem durumu dikkate alınarak uygulanır ve tüm temel GMM'ler için model parametreleri güncellenir. Kullanılan bu yaklaşım ile GMM modelinin eğitim aşamasında ses üzerinde etiketleme, bölümleme, kelime başlangıcı ve bitişi işaretleme gereği olmadan ses verisine ilişkin istatistiki yapı elde edilebilmektedir.

Test aşamasında hesaplanan GMM parametreleri ile uygun şekilde oluşturulmuş test modeli üzerinden test ses verisine ilişkin kelime boyutunda çözümleme elde edilmektedir. Sistem başarısı iki farklı kelimeyi ayrık olarak tanıma, çok sayıda kelimeyi ayrık olarak tanıma, kısıtlı sayıda tekrarlı kelimeleri sürekli konuşma yapısında tanıma, sürekli konuşma yapısında cümle tanıma deneysel kurguları üzerinde incelenmiştir. Deneysel kurgunun karmaşıklığına göre sistem başarısı %60-%77 düzeylerinde gerçekleşmiştir.

Sistem, yararlanılan ses veritabanı İngilizce olduğu için İngilizce için bir konuşma tanıma sistemi olarak tasarlanmıştır. Fakat sistem aynı zamanda fon-fonem temelli olduğu için sistem üzerinde hiçbir değişiklik yapmadan Türkçe bir konuşma veritabanı ve Türkçe telaffuz

sözlüğü ile Türkçe için bir konuşma tanıma sistemi olarak kullanılabilir.

Tez çalışması için kullanılan GMM statik matris yapısındadır ve ara durumların dinamik olarak eklenip çıkartılmasına olanak tanımamaktadır. Bu sebeple benzer çalışmaların daha kolay yürütülebilmesi için GMM için daha kullanışlı bir veri yapısının tasarlanıp kullanılması önerilmektedir.

Her fon veya fonem için, temel GMM geçiş olasılık parametresi A değerlerinin özel olarak ayrı ayrı önceden deneysel olarak belirlenmesinin sınıflandırma başarını arttıracağı düşünülmektedir. Tez çalışmasında tüm fon-fonemler için temel GMM A geçiş değerleri şekil 5.5 verilen ilk değerleri ile kullanılmıştır.

Test edilecek örneklerin test modelinde aranan kelimelerden oluşması, sistemin hatalı kabul oranı açısından doğru olarak değerlendirilememesi sonucunu doğurmaktadır. Bu sebeple test örneği için hesaplanan gerçekleşme olasılığının belirli bir kesim değerinin altında kalması durumunda sözlük dışı kelime olarak sınıflandırılmasını sağlayacak bir düzenleme önerilmektedir.

MFKK özellik çıkarımı duymaya ilişkin biyolojik yapıyı modellemektedir. LPC özellikleri ile de konuşma oluşumuna ilişkin biyolojik yapı modellenmektedir. İleriki çalışmalar için MFKK ile LPC katsayıları üzerinden hesaplanan kepstral katsayıların, birlikte, özellik değerleri üzerinde karşılıklı bir çeşit doğrulama gerçekleştirilecek şekilde kullanılmasıyla, özelliklerin daha doğru olarak (sınıflandırmaya olumlu katkı sağlayacak şekilde) elde edilebileceği düşünülmektedir.

Sistemin geneli düşünüldüğünde başarıya etki edecek parametreler olarak kullanılan ayrık fon sayısı, özellik çıkarımında seçilecek pencereleme boyutu, GMM için ayrık gözlem sembol sayısı ve bunu belirleyen K-ortalama için merkez vektör sayısı, Baum-Welch algoritması için sonlandırma koşulu gibi değerler ön plana çıkmaktadır. Fon sayısı arttıkça daha geniş bir arama uzayı kullanılacağı için tanıma başarısı göreceli olarak azalmaktadır. GMM eğitiminde ayrık gözlem sembolleri kullanıldığından sistemi aşırı eğitim durumundan korumak için ayrık gözlem sembol sayısı optimum düzeyde tutulmuştur. Ayrık gözlem sayısının yüksek veya düşük tutulması tanıma başarısını olumsuz yönde etkilemektedir. Baum-Welch algoritmasında sonlandırma koşulu olarak tüm gözlem dizileri için toplam gerçekleşme olasılığının belirli bir değerin altına inmesi durumu kullanılmıştır. Bu sebeple sonlandırma koşul değerinin çok düşük seçilmesi yine sadece verilen eğitim örneklerinde aşırı uygunluğa sebep olmaktadır.

Sistem başarısı hesaplanırken İngilizce deneysel uygulamalar için çeşitli hızlara sahip, bay-bayan konuşmaları kullanılmıştır, Türkçe deneysel uygulamalarda ise tek konuşmacıya ait ortalama hızlı konuşma ses kayıtları kullanılmıştır.

Sonuç olarak bu tez çalışması ile gereksinimler ve kısıtlar göz önüne alınarak sürekli konuşma tanıma için tüm adımları kapsayacak şekilde bir sistem tasarımı gerçekleştirilmiştir.

KAYNAKLAR

Atal, B. S., Hanauer, S. L., Speech Analysis and Synthesis by Linear Prediction of the Speech Wave, J. Acoust. Soc. Am. Vol. 50, No. 2, pp. 637-655, Aug. 1971.

Bahl L. R., Brown P. F., deSouza P. V., Mercer L. R., Maximum Mutual Information Estimation of Hidden Markov Model Parameters for Speech Recognition, Proc. ICASSP 86, Tokyo, Japan, pp. 49-52, April 1986.

Barkat, M., (2005) Signal detection and estimation, Artech House Publishing.

Baum L. E., An Inequality and Associated Maximization Technique in Statistical Estimation for Probabilistic Functions of Markov Processes, Inequalities, Vol. 3, pp. 1-8, 1972.

Boulard, D. S. (1996), A new ASR approach based on independent processing and recombination of frequency bands. In Proceedings of the International Conference on Spoken Language Processing (ICSLP 1996) (Vol 1. pp. 426-429) ACM Digital Library.

Cole R., Mariani J., Uszkoreit H., Zaenen A., Zue V. (1997) Survey of the state of the art in human language technology, Cambridge University Pres

Cooley, J.W., Tukey, J.W., An Algorithm for the Machine Computation of Complex Fourier Series, Mathematics of Computation, Vol. 19,pp. 297-301, April 1965

Davis K. H., Biddulph R., Balashek S., Automatic Recognition of Spoken Digits, J. Acoust. Soc. Am., Vol 24, No. 6, pp. 627-642, 1952.

Dudley, H. and Tarnoczy, (1950), T. H., The Speaking Machine of Wolfgang von Kempelen, J. Acoust. Soc. Am., Vol. 22, pp. 151.

Dudley, H., Riesz, R. R., and Watkins, S. A., (1939), A Synthetic Speaker, J. Franklin Institute, Vol. 227, pp. 739-764.

Ferguson J. D., Hidden Markov Analysis: An Introduction, in Hidden Markov Models for Speech, Institute for Defense Analyses, Princeton, NJ 1980.

Fisher R. A., On an Absolute Criterion for Fitting Frequency Curves, Gonville and Caius College, Cambridge, Statistical Science, Vol. 12, No. 1. (Feb., 1997), pp. 39-41.

Flanagan J. L., Speech Analysis, Synthesis and Perception, Second Edition, Springer-Verlag, 1972.

Forgie J. W., Forgie C. D., Results Obtained from a Vowel Recognition Computer Program, J. Acoust. Soc. Am., Vol. 31, No. 11, pp. 1480-1489, 1959.

Fry, D. B. and Denes, P., (1959), The Design and Operation of the Mechanical Speech Recognizer at University College London, J. British Inst. Radio Engr., Vol. 19, No. 4, pp. 211-229.

Ganapathiraju, A., Hamaker, J., Picone, J., & Doddington, G. (2001). Syllable-based large vocabulary continuous speech recognition. IEEE Transactions on Speech and Audio Processing, 9(4), 358-366.

Hauenstein, A. (1996), The syllable re-revisited (Tech. Rep. No. tr-96-035). Munich, Germany: Siemens AG, Corporate Research and Development.

Hu, Z., Schalkwyk, J., Barnard, E., & Cole, R. (1996). Speech recognition using syllable-like

units. In Proceedings of the International Conference on Spoken Language Processing, Philadelphia, PA, (vol. 2, pp. 1117-1120). Washington DC: IEEE Computer Society.

Itakura F., Minimum Prediction Residual Principle Applied to Speech Recognition, IEEE Trans. Acoustics, Speech and Signal Proc., Vol. ASSP-23, pp. 57-72, Feb. 1975.

Itakura, F., Saito, S.A Statistical Method for Estimation of Speech Spectral Density and Formant Frequencies, Electronics and Communications in Japan, Vol. 53A, pp. 36-43, 1970.

Jelinek F., Bahl L. R., Mercer R. L., Design of a Linguistic Statistical Decoder for the Recognition of Continuous Speech, IEEE Trans. On Information Theory, Vol. IT-21, pp. 250-256, 1975.

Juang, B. H. and Rabiner L. R., (2005), Automatic Speech Recognition-A Brief History of the Technology, Elsevier Encyclopedia of Language and Linguistics, Second Edition.

Juang, B. H., (1985), Maximum Likelihood Estimation for Mixture Multivariate Stochastic Observations of Markov Chains, AT&T Tech. J., Vol. 64, No. 6, pp. 1235-1249, July-Aug.

Juang, B. H., Levinson, S. E. and Sondhi, M. M., (1986), Maximum Likelihood Estimation for Multivariate Mixture Observations of Markov Chains, IEEE Trans. Information Theory, Vol. It-32, No. 2, pp. 307-309, March.

Juang, B. H., Wu Chou, (2003), Pattern Recognition in Speech And Language Processing Crc Press.

Juang, B.H., Lee C.H. and Wu Chou, Minimum classification error rate methods for speech recognition, IEEE Trans. Speech & Audio Processing, T-SA, vo.5, No.3, pp.257-265, May 1997.

Jurafsky, D., Martin, J. H., (1999) Speech and Language Processing, Prentice Hall.

Kapadia S., (1998), Discriminative Training of Hidden Markov Models, Ph.D. Thesis, University of Cambridge Downing College.

Kruskal J. B., Liberman M., The symmetric time-warping problem: from continuous to discrete. In David Sanko and Joseph B. Kruskal, editors, Time Warps, String Edits, and Macromolecules: The Theory and Practice of Sequence Comparisons. Addison-Wesley, Reading, Massachusetts, 1983.

Lee K. F., Large-vocabulary speaker-independent continuous speech recognition: The Sphinx system, Ph.D. Thesis, Carnegie Mellon University, 1988.

Levinson S. E., Rabiner L. R., Sondhi M. M., An Introduction to the Application of the Theory of Probabilistic Functions of a Markov Process to Automatic Speech Recognition, Bell Syst. Tech. J., Vol. 62, No. 4, pp. 1035-1074, April 1983.

Liporace L. A., Maximum Likelihood Estimation for Multivariate Observations of Markov Sources, IEEE Trans. On Information Theory, Vol. IT-28, No. 5, pp. 729-734, 1982.

Lippmann R. P., Review of Neural Networks for Speech Recognition, Readings in Speech Recognition, A. Waibel and K. F. Lee, Editors, Morgan Kaufmann Publishers, pp. 374-392, 1990.

Lowerre, B., (1990), The HARPY Speech Understanding System, Trends in Speech Recognition, W. Lea, Editor, Speech Science Publications, 1986, reprinted in Readings in

Speech Recognition, A. Waibel and K. F. Lee, Editors, pp. 576-586, Morgan Kaufmann Publishers.

Martin T. B., Nelson A. L., Zadell H. J., Speech Recognition by Feature Abstraction Techniques, Tech. Report AL-TDR-64-176, Air Force Avionics Lab, 1964.

McCullough W. S., Pitts W. H., A Logical Calculus of Ideas Immanent in Nervous Activity, Bull. Math Biophysics, Vol. 5, pp. 115-133, 1943.

Mengüşoğlu, E., (1999), Bir Türkçe Sesli İfade Tanıma Sisteminin Kural Tabanlı Tasarımı ve Gerçekleştirimi, Yüksek Mühendislik Tezi, Hacettepe Üniversitesi.

Nagata K., Kato Y., and Chiba S., Spoken Digit Recognizer for Japanese Language, NEC Res. Develop., No. 6, 1963.

Niels, R., (2004), Dynamic Time Warping: An Intuitive Way of Handwriting Recognition? Master's thesis, Radboud University Nijmegen.

Olson, H. F. and Belar, H., (1956), Phonetic Typewriter, J. Acoust. Soc. Am., Vol. 28, No. 6, pp. 1072-1081.

Rabiner L. R., (1989) A Tutorial on Hidden Markov Models and Selected Applicaitons in Speech Recognition, Proceedings of The IEEE, Vol. 77, No: 2, February, pp 257-286.

Rabiner L. R., Levinson S. E., Rosenberg A. E., Wilpon J. G., Speaker Independent Recognition of Isolated Words Using Clustering Techniques, IEEE Trans. Acoustics, Speech and Signal Proc., Vol. Assp-27, pp. 336-349, Aug. 1979.

Rabiner, L., Juang, B. H., (1993), Fundamentals of Speech Recognition, PTR Prentice-Hall.

Sakai J., Doshita S., The Phonetic Typewriter, Information Processing 1962, Proc. IFIP Congress, Munich, 1962.

Sakoe H., Chiba S., Dynamic Programming Algorithm Quantization for Spoken Word Recognition, IEEE Trans. Acoustics, Speech and Signal Proc., Vol. ASSP-26, No. 1, pp. 43-49, Feb. 1978.

Steven B. Davis, Paul Mermelstein, Comparison of Parametric Representations for Monosyllabic Word Recognition in Continuously Spoken Sentences IEEE Transactıons On Acoustics, Speech, And Signal Processing, Vol. Assp-28, No: 4, pp 357-366, August 1980.

Suzuki J., Nakata K., Recognition of Japanese Vowels—Preliminary to the Recognition of Speech, J. Radio Res. Lab, Vol. 37, No. 8, pp. 193-212, 1961.

Vapnik V. N., Statistical Learning Theory, John Wiley and Sons, 1998.

Vintsyuk T. K., Speech Discrimination by Dynamic Programming, Kibernetika, Vol. 4, No. 2, pp. 81-88, Jan.-Feb. 1968.

Vuori, V., Clustering writing styles with a self-organizing map. In Proceedings of The 8th International Workshop on Frontiers in Handwriting Recognition, pages 345-350, Niagara-on-the-Lake, 2002.

Wu, S., Shire, M., Greenberg, S., & Morgan, N. (1997), Integrating syllable boundary information into automatic speech recognition. In Proceedings of the IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), Munich, Germany (vol. 1, pp. 987-990). Washington, DC: IEEE Computer Society.

Xiaolin Li, Marc Parizeau, Réjean Plamondon, Training Hidden Markov Models With Multiple Observations - A Combinatorial Method, IEEE Transactıons on Pattern Analysis and Machine Intelligence, Vol. 22, No: 4, pp 371-377, April 2000.

Young S., et. al., the HTKBook, http://htk.eng.cam.ac.uk/

Zhou, L., Fang, D. (1988) The Research on Speech Feature Representation Methods and Distance Measure Methods. IEEE.

İNTERNET KAYNAKLARI

[1] http://ocw.mit.edu/OcwWeb/Health-Sciences-and-Technology/HST-721Fall-2005/CourseHome/index.htm

[2] http://www.owlnet.rice.edu/~elec301/Projects01/speech_syn/speech.htm

[3] http://www.ims.uni-stuttgart.de/~moehler/ISCA-SynSIG/historical

[4] http://www.arts.gla.ac.uk/IPA/IPA_chart_(C)2005.pdf

[5] http://www.speech.cs.cmu.edu/databases/an4/

[6] http://en.wikipedia.org/wiki/Hamming_window

[7] www.cambridge.org/us/features/chau/webnotes/chap2laplace.pdf

ÖZGEÇMİŞ

Doğum tarihi	01.03.1982	
Doğum yeri	Ordu	
Lise	1996-1999	Özel Fatih Erkek Fen Lisesi
Lisans	1999-2003	İstanbul Teknik Üniversitesi İşletme Fakültesi İşletme Mühendisliği
Lisans	2003-2006	Yıldız Teknik Üniversitesi Elektrik-Elektronik Fakültesi Elektronik ve Haberleşme Mühendisliği
Yüksek Lisans	2003-2007	Yıldız Teknik Üniversitesi Fen Bilimleri Enstitüsü Bilgisayar Mühendisliği

Çalıştığı kurum(lar)

	2006-Devam	YTÜ Bilgisayar Mühendisliği, Donanım Anabilim Dalı, Araştırma Görevlisi