

T.C.
FIRAT ÜNİVERSİTESİ
FEN BİLİMLERİ ENSTİTÜSÜ

AKILLI YÖNTEMLER İLE KONUŞMACI TANIMA

Eda RENÇBER

Yüksek Lisans Tezi

Elektronik ve Bilgisayar Eğitimi Anabilim Dalı

Danışman: Yrd. Doç. Dr. Davut HANBAY

AĞUSTOS-2011

T.C.
FIRAT ÜNİVERSİTESİ
FEN BİLİMLERİ ENSTİTÜSÜ

AKILLI YÖNTEMLER İLE KONUŞMACI TANIMA

YÜKSEK LİSANS TEZİ

Eda RENÇBER

(07131103)

Tezin Enstitüye Verildiği Tarih: 09 Ağustos 2011
Tezin Savunulduğu Tarih: 25 Ağustos 2011

Tez Danışmanı: Yrd. Doç. Dr. Davut HANBAY (F.Ü)
Diğer Jüri Üyeleri: Doç. Dr. Abdulkadir ŞENGÜR (F.Ü)
Doç. Dr. Sami EKİCİ (F.Ü)

AĞUSTOS-2011

ÖNSÖZ

Bu tez çalışması, Fırat Üniversitesi Fen Bilimleri Enstitüsü Elektronik - Bilgisayar Eğitimi Anabilim Dalı Yüksek Lisans Programı'nda hazırlanmıştır.

Bu çalışmada farklı kişilere ait ses işaretlerinin yapay sinir ağları ile uyarlamalı sinirsel bulanık ağ gibi akıllı metotları kullanıp sınıflandırma yapılarak sistem başarısı tespit edilmeye çalışılmıştır. Bu uygulamada amaç, ses işaretinin ön işlem aşamasının başarıyla yapılabilmesi ve ön işlem sonucundan sonra bu işaretin ayırt edici özelliklerinin çıkartılması ve uygun bir düzende yapay sinir ağı ile uyarlamalı sinirsel bulanık ağa verilmesidir. Çalışmanın son aşaması ise değişik tekniklerle özelliği çıkartılmış kişilere ait ses sinyallerinin, tanınma başarılarının kıyaslanması olmuştur.

Bu tez çalışmasının gerçekleşmesinde kıymetli zamanını ayırarak yardım ve desteğini esirgemeyen ve uyarıları ile yol gösteren tez danışmanım Sayın Yrd. Doç. Dr. Davut HANBAY'a, tez çalışmam boyunca yardımlarını esirgemeyen Sayın Doç. Dr. İbrahim TÜRKOĞLU'na, Sayın Doç. Dr. Engin AVCI'ya ve Sayın Doç. Dr. Abdulkadir ŞENGÜR'e, bilgilerini sürekli olarak benimle paylaşan arkadaşım Öznur ERKUŞ'a, manevi desteğiyle beni hiçbir zaman yalnız bırakmayan çok değerli anneme, babama ve eşime en içten teşekkürlerimi ve şükranlarımı sunarım.

Eda RENÇBER
ELAZIĞ - 2011

İÇİNDEKİLER

Sayfa No

ÖNSÖZ	II
İÇİNDEKİLER	III
ÖZET	V
SUMMARY	VI
ŞEKİLLER LİSTESİ	VII
TABLolar LİSTESİ	VIII
SEMBOLLER LİSTESİ	IX
KISALTMALAR	X
1. GİRİŞ	1
1.1. Sesin Fizyolojisi ve Konuşmanın Oluşum Mekanizması.....	2
1.2. Konuşmacı Tanıma.....	4
1.3. Konuşmacı Tanıma ile ilgili Yapılan Çalışmalar.....	5
1.4. Çalışmanın Amacı.....	7
2. KONUŞMACI TANIMA SİSTEMLERİ.....	8
2.1. Konuşmadan Öznitelik Çıkarma Metotları.....	8
2.1.1. MFKK (Mel-Frekansı Kepstrum Katsayıları)	8
2.1.1.1. Çerçeveleme (Frame Blocking).....	9
2.1.1.2. Pencereleme (Windowing).....	9
2.1.1.3. Hızlı Fourier Dönüşümü (HFD).....	9
2.1.1.4. Mel-Frekansına Çevirme.....	10
2.1.1.5. Kepstrum.....	11
2.1.2. Dalgacık (Wavelet) Dönüşümü ile Entropi Hesaplama.....	11
2.1.2.1 Dalgacık Dönüşümü.....	13
2.1.2.2. Sürekli Dalgacık Dönüşümü.....	15
2.1.2.3. Ayırık Dalgacık Dönüşümü.....	16
2.1.2.4. Dalgacık Paket Analizi.....	17
2.1.2.5. Entropi Hesaplama Teknikleri.....	18
2.2. Konuşmacı Tanımadaki Kullanılan Sınıflandırma Yöntemleri.....	20

2.2.1. Gaussian Karma Model ile Konuşmacı Tanıma.....	21
2.2.2. Vektör Nicemleme Tekniğine Dayalı Konuşmacı Tanıma.....	21
2.2.3. Saklı Markov Modelleri ile Konuşmacı Tanıma.....	22
2.2.4. Dinamik Zaman Eğiirme ile Konuşmacı Tanıma.....	23
3. AKILLI YÖNTEMLER.....	25
3.1. Yapay Sinir Ağları (YSA)	25
3.1.1. Biyolojik Sinir Ağları ile Yapay Sinir Ağlarının Karşılaştırılması.....	25
3.1.2. Yapay Sinir Ağlarının Yapıları ve İşlem Elemanları.....	29
3.2. Uyarlamalı Sinirsel Bulanık Ağ (USBA) Mimarisi.....	33
3.2.1. USBA İçin Geri Yayılımlı Öğrenme Algoritması.....	35
4. AKILLI YÖNTEMLER İLE KONUŞMACI TANIMA UYGULAMALARI	41
4.1. MFKK Kullanarak YSA ve USBA Akıllı Yöntemler ile Konuşmacı Tanıma...	41
4.1.1. MFKK ve YSA ile Konuşmacı Tanıma.....	41
4.1.2. MFKK ve USBA ile Konuşmacı Tanıma.....	44
4.2. DD Kullanarak YSA ve USBA Akıllı Yöntemler ile Konuşmacı Tanıma.....	46
5. SONUÇLAR VE ÖNERİLER.....	50
KAYNAKLAR	52
ÖZGEÇMİŞ	

ÖZET

Bu çalışmada YOHO veri tabanından alınan cümleler üzerinde Yapay Sinir Ağları (YSA) ve Uyarlamalı Sinirsel Bulanık Ağ (USBA) kullanılarak metin bağımlı konuşmacı tanıma uygulamaları geliştirilmiştir. Konuşmacı tanıma sisteminin öznitelik çıkarım aşamasında Mel Frekansı Kepstrum Katsayıları (MFKK) ve Dalgacık Dönüşümlü (DD) entropi değerleri kullanılmıştır. Elde edilen özellik vektörleri, YSA ve USBA tabanlı modellere giriş olarak uygulanmış ve modeller eğitilmiştir. Test aşamasında ise verilen konuşma örneğinden konuşmacının kim olduğu tespit edilmeye çalışılmıştır. Aynı zamanda MFKK öznitelik vektörleri ve DD sonucu elde edilen entropi değerleri ile YSA ve USBA gibi akıllı metotların çeşitli parametrelerinin değişmesinin konuşmacı tanıma üzerindeki etkileri incelenerek karşılaştırılmıştır.

Anahtar Kelimeler: Konuşmacı Tanıma, Yapay Sinir Ağları, Uyarlamalı Sinirsel Bulanık Ağ, MFKK, Dalgacık Dönüşümü, Entropi.

SUMMARY

Speaker Recognition with Intelligent Methods

In this study, text depended speaker recognizer system; based on Neural Network (NN) and Adaptive Network Fuzzy Inference System (ANFIS) are designed by using YOHO database. In feature extracting stage of speaker recognizer, MFCC and Wavelet Transform (WT) entropy values were used. The obtained feature vectors are used as inputs to the NN and ANFIS models. In test stage, recognition of the speakers is performed from the given speech samples. At the same time, MFCC feature vectors and wavelet transforms entropy values are used for examining the NN and ANFIS model parameters effects on the systems performance and finally the results are compared.

Key Words: Speaker Recognition, Neural Network, Adaptive Network Based Fuzzy Inference Systems, MFCC, Wavelet Transform, Entropy.

ŞEKİLLER LİSTESİ

Sayfa No

Şekil 1.1. Konuşmacı tanıma sınıflandırma.....	2
Şekil 1.2. Ses sistemi.....	3
Şekil 1.3. Konuşmacı tanıma karar verme süreci.....	4
Şekil 2.1. Öznitelik çıkarma aşamaları.....	8
Şekil 2.2. Mel yayılımlı filtre bankası.....	10
Şekil 2.3. Entropi katsayıları çıkarma.....	11
Şekil 2.4. Pencerelemiş Fourier dönüşümü için zaman-frekans diyagramı.....	13
Şekil 2.5. Dalgacık dönüşümü için zaman-frekans diyagramı.....	14
Şekil 2.6. Dalgacık dönüşümü için iki boyutlu zaman-frekans diyagramı	14
Şekil 2.7. Yaygın olarak kullanılan dalgacık örnekleri.....	15
Şekil 2.8. ADD ayrışım ağacı.....	17
Şekil 2.9. DPA ağaç yapısı.....	18
Şekil 2.10. Düşük entropi kavramı.....	19
Şekil 2.11. Yüksek entropi kavramı.....	19
Şekil 2.12. Basit bir soldan sağa SMM.....	22
Şekil 3.1. Yapay sinir ağlarının yapısı.....	25
Şekil 3.2. Biyolojik nöronun şematik yapısı.....	26
Şekil 3.3. Bir yapay sinir ağı hücresi modeli.....	27
Şekil 3.4. Yapay sinir ağları için kullanılan eşik fonksiyonları.....	28
Şekil 3.5. 2 girişli 4 kurallı bir USBA sınıflandırıcı yapısı.....	33
Şekil 4.1. YSA ile konuşmacı tanıma algoritması.....	43
Şekil 4.2. USBA ile konuşmacı tanıma algoritması.....	45
Şekil 4.3. Dikdörtgen pencere ile pencerelemiş sinyal	47
Şekil 4.4. DD tabanlı YSA veya USBA ile konuşmacı tanıma algoritması.....	48

TABLÖLAR LİSTESİ

Sayfa No

Tablo 4.1. MFKK ile konuşmacı tanıma eğitim ve test cümleleri.....	42
Tablo 4.2. Çok katmanlı yapay sinir ağının yapısı ve eğitim parametreleri.....	44
Tablo 4.2. USBA modelinin yapısı ve eğitim parametreleri.....	45
Tablo 4.4. DD ile konuşmacı tanıma eğitim ve test cümleleri.....	46
Tablo 4.5. Uyarlamalı sinirsel bulanık ağ sisteminin parametreleri ve başarı yüzdesi.....	49

SEMBOLLER LİSTESİ

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

KISALTMALAR

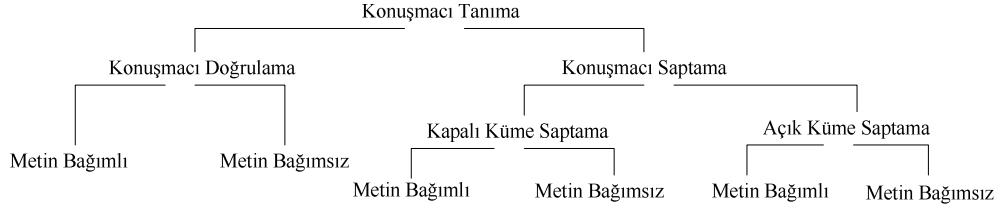
ADD	: Ayrık Dalgacık Dönüşümü
AFD	: Ayrık Fourier Dönüşümü
DD	: Dalgacık Dönüşümü
DDA	: Doğrusal Diskriminant Analiz
DPA	: Dalgacık Paket Ayırıştırma
DZE	: Dinamik Zaman Eğirme
FD	: Frekans Dönüşümü
GKM	: Gaussian Karma Model
HFD	: Hızlı Fourier Dönüşümü
KZFD	: Kısa Zamanlı Fourier Dönüşümü
MFKK	: Mel Frekansı Kepstrum Katsayıları
SDD	: Sürekli Dalgacık Dönüşümü
SMM	: Saklı Markov Model
USBA	: Uyarlamalı Sinirsel Bulanık Ağ
VN	: Vector Nicemleme
YSA	: Yapay Sinir Ağları
ZFG	: Zaman Frekans Gösterimi

1. GİRİŞ

Doğadaki canlılar arasında kullanılan temel iletişim araçlarından biri sestir. Sesin insanlar arasında iletişim amacıyla kullanılmasıyla konuşma ortaya çıkmıştır. Konuşma işareti, parmak izi ve retina gibi bireylere özgü özellikler taşır. Bu nedenle de birçok bilimsel çalışmaya konu olmuştur. Bu çalışmalarda temel araç insan sesi olduğundan veri olarak ses işareti kullanılır. Alınan ses işaretinin kullanılabilmesi için sayısallaştırılması ve kullanılacak amaca uygun özelliklerinin çıkartılması gerekir. Böylece konuşmacı tanıma sistemlerinde, bu ses işareti ile konuşmacı modellenerek, tanıma işlemlerinde kullanılacak karakteristik şablonlar elde edilir. Tanıma aşamasında ise önceden elde edilen eğitim şablonlarıyla test için alınan konuşmadan elde edilen şablonlar karşılaştırılır. Şablonlar arasında benzerlik olduğuna karar verildiğinde konuşmacı kabul edilir aksi durumda reddedilir [1].

Konuşmacı tanıma iki ana bölüme ayrılır; konuşmacı doğrulama (speaker verification) ve konuşmacı saptama (speaker identification) [2]. Konuşmacı doğrulama, bilinmeyen bir ses örneğinin, iddia edilen kişiye ait olup olmadığının belirlenmesidir. Konuşmacı doğrulama; iletişim hizmetlerinde, bankacılık hizmetlerinde, özel kayıtlara vb. güvenli erişimler sağlamak amacıyla birçok uygulamada kullanılabilir. Konuşmacı saptama ise bilinmeyen bir ses örneğinin, belli konuşmacıların ses kayıtlarından oluşan bir veritabanı içerisinde hangi kişiye ait olduğunun bulunmasıdır.

Konuşmacı tanıma, metin bağımlı ya da metin bağımsız olabilir. Metin bağımlı sistemlerde konuşulan metin sistem tarafından önceden bilinmektedir. Metin bağımsız sistemlerde ise metin, herhangi bir sözdizimi olabilir [3]. Diğer taraftan konuşmacı tanıma, açık küme ya da kapalı küme olabilir. Kapalı kümede bilinmeyen ses örneği, veritabanındaki konuşmacılardan birisine aittir. Açık kümede ise ses örneği veritabanındaki konuşmacılardan hiç birisine ait olmayabilir [4]. Dolayısı ile açık küme konuşmacı tanıma sistemlerinde, ret sonucunu da veren fazladan bir olasılık daha bulunur. Şekil1.1’de konuşmacı tanıma genel sınıflandırma şeması verilmiştir [5].



Şekil 1.1. Konuşmacı tanıma sınıflandırma

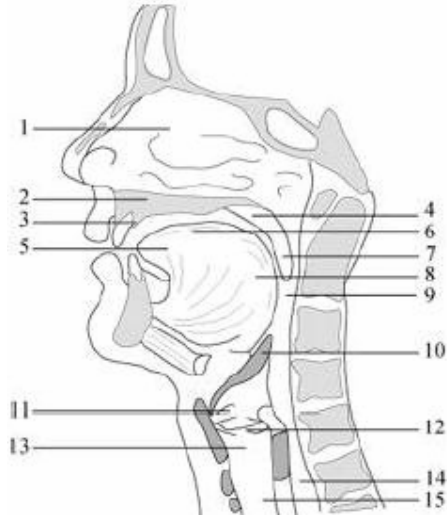
İnsanların bazı durumlarda farklı ton ve vurguda konuşmaları, hastalık ve duygusal koşullar sonucu çıkarılan seslerin farklı olması, ortamdaki gürültü, daha da kötüsü bir başkasının kişinin sesini taklit etmesi konuşma ve konuşmacı tanıma sistemlerinin karmaşıklığını ve geliştirilmeye açık konular olduğunu göstermektedir.

Konuşmacı tanıma sistemleri genellikle sesli arama, telefon üzerinden bankacılık işlemleri, telefonla alışveriş, veritabanı erişim servisleri, bilgi ve rezervasyon servisleri, sesli posta, gizlilik gerektiren bilgilere erişimde, güvenlik kontrolleri, adli uygulamalar ve bilgisayarlara uzaktan erişim gibi alanlarda kullanılır.

1.1.1.1. Sesin Fizyolojisi ve Konuşmanın Oluşum Mekanizması

Ses, gırtlak, akciğerler, kas ve iskelet sistemi ile psiko-nörolojik sistemlerin birbirleriyle koordineli olarak çalışması sonucunda meydana gelir.

Ses dalgaları; birinden diğerine ulaşabilen titreşimlerdir. Akustik bir sinyal olan sesin oluşumu için maddesel ortamda titreşim yapabilen bir ses kaynağına ve ses dalgalarının yayılabileceği iletici bir ortama ihtiyaç vardır. Her madde, komşu partikül üzerine kendinde bulunan hareketi nakletmek için etki yapar [6]. Şekil 1.2’de sesin oluşumunda etkisi olan tüm organlar ayrıntılı olarak verilmiştir.



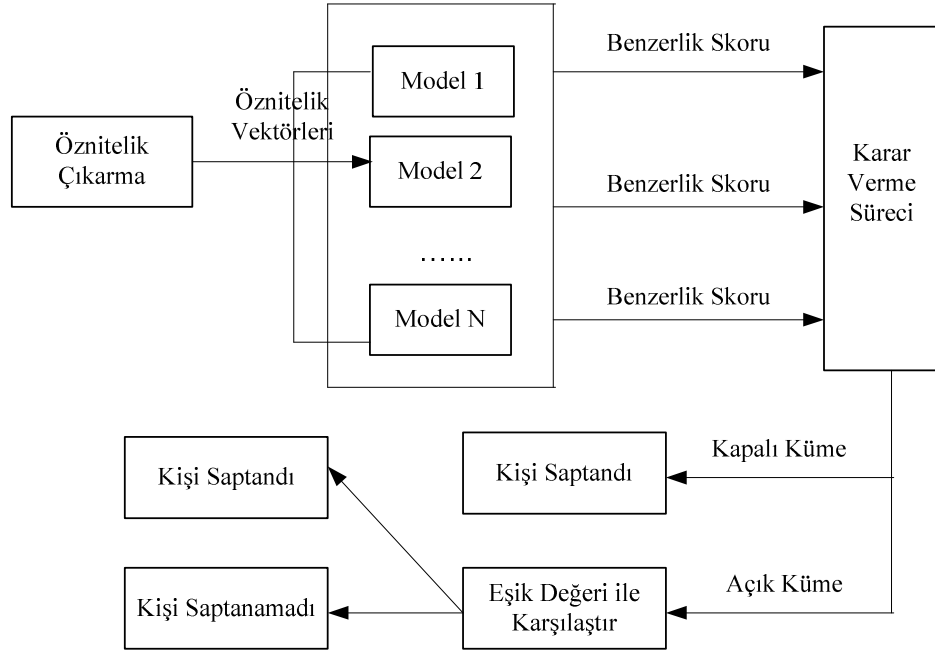
Şekil 1.2. Ses sistemi

1. Burun boşluğu
2. Sert damak
3. Alveoral krep
4. Yumuşak damak
5. Dil ucu
6. Dil sırtı
7. Küçük dil
8. Dil kökü
9. Yutak
10. Epiglot kıkırdak
11. Sahte ses telleri
12. Ses telleri
13. Gırtlak
14. Yemek borusu
15. Soluk Borusu.

Akciğerlerdeki hava karınsal kasların yardımı ile birlikte ses tellerinin arasından geçerek vızıltı sesine benzer ham bir ses oluşturur. Bu ses nefes borusu ağzı üzerinde ses telleri boyunca, geniz, ağız boşluğu, dil, burun ve sinüslerde rezonansa uğrayarak konuşmayı oluşturur.

1.2. Konuşmacı Tanıma

Konuşmacı tanıma, ses dalgalarının içerdiği bireysel bilgiden faydalanarak konuşanın kim olduğunun otomatik tanındığı bir işlemdir. Bu teknik, erişim kontrolünün sesle yapıldığı sistemlere erişmeye çalışan kişilerin, kimliklerinin belirlenebilmesini mümkün kılmaktadır [1]. Tanımlama işlemlerinde Şekil 1.3'te gösterilen basamaklar uygulanır [5].



Şekil 1.3. Konuşmacı tanıma karar verme süreci

Şekil 1.3'deki şemada her bir blok, sistemin bir parçasını temsil eder. Konuşmacı tanıma öznitelik vektörleri çıkartma ve öznitelik eşleştirme olarak iki bölümden oluşur. Öncelikle ses işaretinin girişte ön işlem yapılması sonucu boyutu azaltılır. Ardından kişileri temsil eden sayısal vektörler oluşturulur. Daha sonra bu özellik vektörleri önceden belirlenen modeli eğitmek için kullanılır. Sistemin en sonunda karar mekanizması yer alır. Karar mekanizmasının girişinde bulunan test vektörü ile eğitilmiş model kullanılarak test örneğindeki sesin hangi konuşmacıya ait olduğu belirlenmiş olur.

1.3. Konuşmacı Tanıma ile ilgili Yapılan Çalışmalar

1976 yılında Marvin R. Sambur OLP tekniğini kullanarak %94; 1983'te Li ve Wrench örüntü eşleme ile %96; F. Soong, A. Rosenberg, L. Rabiner ve B. Juang 1985 yılında VN ile %98; T. Matsui ve S. Furui 1991'de VN yöntemi ile %99; Douglas A.Reynolds ve Richard C. Rose 1995'te SMM ile %80.8; 1999 yılında Zhong-Xuan Yuan, Bo-Ling Xu ve Chong-Zhi Yu MFKK ile %99.1; Q. Jin ve A. Waibel 2000'de DDA ile %97; M. Selvi, U. Yapanel ve F.Gürgen 2002'de WMFCC ile %98.9; Jiuqing Deng ve Qixiu Hu 2003'te örüntü eşleme ile %98.5; L. Zhang, B. Zheng ve Z. Yang 2005'te GA yöntemi ile %93.85 başarı elde etmişlerdir [5].

Uzunçarşılı 2005 yılında yaptığı çalışmada MFKK öznitelik vektörlerini kullanarak VN tekniklerine dayalı, kapalı küme, metinden bağımsız bir konuşmacı tanıma sistemi tasarlayarak konuşmacı sayısına bağlı olarak farklı oranlarda başarı elde etmiştir [6].

Başaran 2007'de MFKK kullanarak YSA ve VN gibi farklı algoritmalar ile başarıyı etkileyen parametreler üzerinde çalışmıştır. Daha sonra bu iki yöntemin başarıya olan etkilerini incelemiştir [7].

Hanilçi 2007 yılında daha önce yapılan çalışmalardan VN ve SMM yöntemlerinin kullanılması sonucu elde edilen başarıları incelemiştir. Kendi yapmış olduğu çalışmasında ise kişi sayısını arttırıp VN ile SMM yöntemlerdeki parametreleri de değiştirerek kendi sistemini oluşturmuştur. Daha önce yapılan çalışma ile sonuçları karşılaştırmış ve daha fazla başarı sağlamıştır [2].

Dalgacık dönüşümü özellikle son on yıl içerisinde konuşmacı tanımayla ilgili çalışmalarda da kullanılmaya başlanmıştır. Bu çalışmalar;

1997'de Singh ve arkadaşları tarafından yapılan ve SMM'nin kullanıldığı çalışmada ADD ile, bir iyileştirme yöntemi kullanarak hem bölütlemeye hem de tanıma performansında artış sağlamışlardır [8].

1998 yılında M. Namba ve Y. Ishida tarafından yapılan çalışmada parametre normalizasyonun en temel ve en çok kullanılan yöntemlerinden birisi olan Kör dekonvolüsyon (blind deconvolution) yöntemine, ortogonal dalgacık yöntemiyle bir yaklaşım uygulanmış öz değer vektörlerinin hesaplanmasında daha hızlı sonuçlar elde edilmiştir [8].

Durağan olmayan geniş bant gürültülerinin konuşma sinyalinden ayıklanmaları oldukça zordur. Lou ve arkadaşları 2002'de dalgacık dönüşümünü kullanan bir yöntemle ses

sinyalleri öncelikle dalgacık uzayında bileşenlerine ayırmıştır. Daha sonra Karhunen-Loeve dönüşümünü temel alan vektör uzayında iyileştirmiş ve tekrar birleştirmiştir. Gerçekleştirilen çalışmanın sonunda durağan olmayan geniş bant gürültülerine karşı kullanılan diğer yöntemlere göre daha yüksek başarı elde edilmiştir [8].

2002’de Erçelebi tarafından yapılan çalışmada ise ses tanıma sistemlerinin önemli konularından biri olan temel frekansın belirlenmesi için DD kullanılmış ve gürültülü ortamlarda bile başarılı olan yeni bir yöntem geliştirilmiştir. Geliştirilen bu yöntem hem düşük temel frekanslı hem de yüksek temel frekanslı konuşmacıların seslerinde başarı sağlamıştır [8].

2004’te Lung tarafından gerçekleştirilmiş bir çalışmada ise adaptif bulanık C ve DD kümelemesi aşamalarından geçirilen ses sinyalinden daha kaliteli öznitelik vektörler elde edilmeye çalışılmıştır. Elde edilen bu öznitelik vektörleri konuşmacı tanımada kullanılarak %95’lik bir doğru tanıma yüzdesi elde edilmiştir [8].

Benzer şekilde 2004’te Lu ve arkadaşları parçalı gürültülü seslerin analizinde kullanılacak olan bir kazanç faktörü elde ederek dalgacık tabanlı bir iyileştirme yöntemi geliştirmişlerdir [8].

2005’te Ganbari ve arkadaşları yaptıkları çalışmada VAD adını verdikleri bir sisteme uyguladıkları dalgacık eşikleme algoritması ile gauss, pink ve çoklu konuşmacı gürültülerine karşın etkili bir sistem ortaya çıkarmışlardır [8].

Yine 2006’da Lung tarafından kaleme alınan makalede dalgacık paketleri kullanılarak elde edilen öznitelik vektörleri sinirsel bulanık bir ağa verilerek eğitilmiş, sonuçta da yüksek oranda konuşma tanıma başarısı sağlamışlardır [8].

Avcı ve Akpolat’ın 2006’da sundukları çalışmada, DPA yöntemini kullanarak gerçek zamanda öznitelik vektörlerinin elde edilmesini sağlamışlardır. Yapılan çalışmada gürültülü verilere karşı %92’lik bir başarı sağlanmıştır [8].

Obaidat ve arkadaşları tarafından yapılan çalışmada ise diydik dalgacık dönüşümleri kullanılarak pitch (perde) periyodun belirlenmesi sağlanmıştır. Temel frekansın bulunmasında % 100’e yakın bir başarı elde edilmiştir [8].

Aygün 2006’da metin bağımlı olarak gerçekleştirilen sistemde MFKK ile elde edilen öznitelik katsayıları dalgacık dönüşümü işlemine tabii tutularak 4. seviye katsayılar elde etmiştir. Elde edilen bu yaklaşım ve detay katsayılar birleştirilerek ikinci bir öznitelik vektörü olarak değerlendirmeye alınmıştır. MFKK ve dalgacık katsayıları DZE algoritması kullanılarak birbirleriyle karşılaştırılmış ve her birinin birbirine olan uzaklığı

hesaplanmıştır. Daha sonraki test aşamasında bu uzaklıklar kullanılmıştır. Genel itibariyle gerçekleştirilen sistem %70'lik bir başarı elde etmiştir [8].

1.4. Çalışmanın Amacı

Güvenlik ve şifreleme gibi kişiye bağımlı sistemlerde işlemlerin doğru kişi tarafından yapılması en önemli konudur. Bu nedenle işlemlerin doğru kişi tarafından yapıldığından emin olmak için yüz yüze yapılması gerekmektedir. Herkesin karşısına bir insan koymak ve kullanıcıların bu insanlarla temasa geçmesini sağlamak hem zaman kaybına yol açmakta hem de maliyetleri artırmaktadır. Bu soruna çözüm olarak, tanıma işleminin makinelerle gerçekleştirilmesi önerilmiştir. Bu çalışmalar sayesinde makine yardımı ile etkin konuşmacı tanıma sistemlerinin geliştirilmesi amaçlanmaktadır. Konuşmacı tanıma sistemleri sayesinde kullanıcıların anahtar ya da akıllı kart taşımaya gerek duymadan seslerini bir çeşit şifre gibi kullanabilmeleri sağlanmaktadır.

Yapılan bu çalışmada YOHO veritabanı kullanılarak kapalı küme, metin bağımlı konuşmacı tanıma (saptama) sistemi geliştirilecektir. İlk önce MFKK ve DD ile öznitelik vektörleri çıkarımı yapılacaktır. Daha sonra YSA ve USBA gibi akıllı metotlar kullanılarak, akıllı konuşmacı tanıma sistemleri geliştirilecektir. Elde edilecek olan sistemlerin sonuçları analiz edilerek en uygun olanı belirlenecektir.

2.KONUŞMACI TANIMA SİSTEMLERİ

2.1. Konuşmadan Öznitelik Çıkarma Metotları

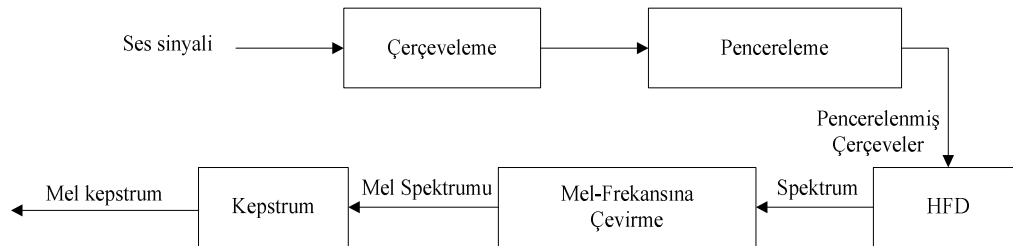
2.1.1. MFKK (Mel-Frekansı Kepstrum Katsayıları)

Ses sinyalindeki akustik özelliklerin belirlenmesine öznitelik çıkarma denir. MFKK, konuşma tanıma alanında en çok kullanılan öznitelik çıkarma yöntemlerinden birisidir [9, 10]. Öznitelik vektörlerini çıkarmanın amacı, ses örneğindeki, konuşmacıyı tanımlayan akustik özelliklerden ödün vermeden yüklü miktardaki veriyi özetlemektir [11]. Böylece tanıma işlemi kolaylıkla gerçekleşir ve daha verimli olur. MFKK öznitelik çıkarma işlemi konuşmacı tanıma sisteminin hem kaydolma hem de saptama aşamasında kullanılır.

Konuşmacı saptama sisteminin ilk aşaması olan kaydolma aşamasında, konuşmacılardan alınan ses örnekleri çeşitli sinyal işleme yöntemleriyle, konuşmacıya ait olan öznitelik vektörlerini elde etmek maksadıyla işlenirler. İşlem sonunda her konuşmacı için bir model oluşturularak konuşmacı veritabanına kaydedilir. Bu veritabanı daha sonra saptama aşamasında kullanılmak üzere saklanır.

Saptama aşamasında ise, verilen bir bilinmeyen ses örneği işlenip öznitelik vektörleri çıkarılır. Daha sonra bu vektörler bir karşılaştırma algoritması yardımıyla konuşmacı veri tabanında kayıtlı olan diğer modellerle karşılaştırılır. Daha sonra her model için bir benzerlik değeri bulunur. En benzer model, ses örneğinin sahibi olarak belirlenir.

Öznitelik vektörleri Şekil 2.1'deki aşamalardan oluşur [12, 13].



Şekil 2.1. Öznitelik çıkarma aşamaları

2.1.1.1. Çerçeveleme (Frame Blocking)

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

2.1.1.2. Pencereleme (Windowing)

İkinci aşamada ise çerçevelenen sinyal pencerelenir. Pencerelemenin amacı çerçevenin başında ve sonunda bulunan süreksizlikleri ortadan kaldırmaktır. Bu aşamada en çok kullanılan pencereleme yöntemi olarak Hamming fonksiyonu kullanılır. Hamming fonksiyonu Denklem 2.1. ile tanımlanır [14]. N , pencere süresini gösterir.

$$p(t) = \begin{cases} 0,54 + 0,46 \cos 2\pi t / N & -N/2 \leq t \leq N/2 \\ 0, & |t| > N/2 \end{cases} \quad (2.1)$$

2.1.1.3. Hızlı Fourier Dönüşümü (HFD)

Konuşma işaretiyle çalışmanın en temel tekniklerinden biri güç spektrumudur. Güç spektrumu, konuşma sinyalinin zaman içindeki frekans bilgilerini gösterir. Konuşma sinyalinin güç spektrumunun hesaplanmasındaki ilk işlem Ayrık Fourier Dönüşümünün yapılmasıdır. HFD, AFD hesaplamak için kullanılan hızlı bir algoritmadır. N örnekli her çerçevenin zaman domeninden frekans domenine çevrilmesi için HFD uygulanır. HFD, N örneklik bir sette alt indis $\{X_n\}$ Denklem 2.2. de olduğu gibi tanımlanabilir:

$$X_n = \sum_{k=0}^{N-1} x_k e^{-2\pi jkn / N}, \quad n=0,1,2, \dots, N-1 \quad (2.2)$$

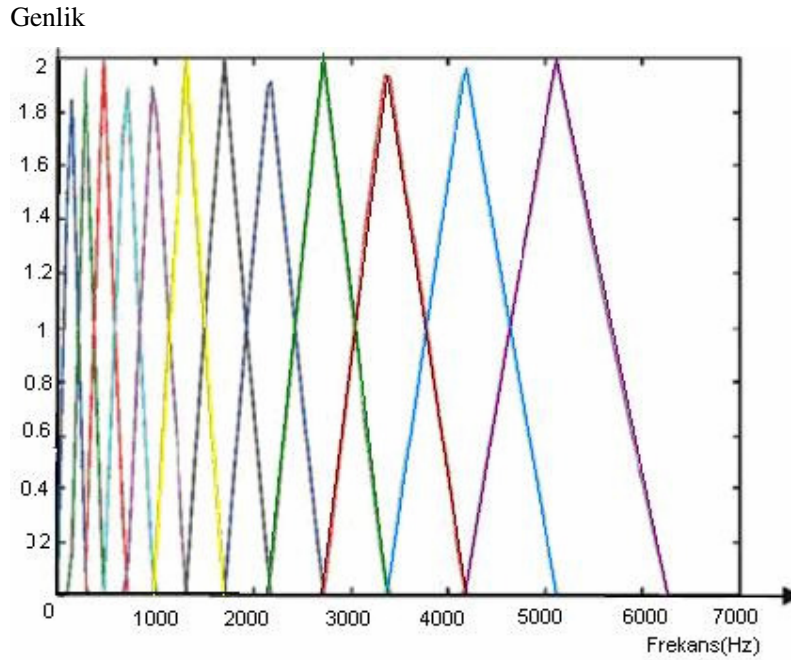
2.1.1.4. Mel-Frekansına Çevirme

Normal bir insan kulağı frekansları doğrusal olmayan bir şekilde algılar. Araştırmalara göre ölçüler, 1 kHz e kadar doğrusal, daha yüksek değerlerde ise logaritmiktir [11, 15]. İnsan kulağının frekans cevabını karakterize eden bu ölçüye melodi ölçüsü denir. Bir frekansı mel-frekansına çevirmek için Denklem 2.3. kullanılır [15]:

$$mel(f) = 2595 * \log (1 + f / 700) \quad (2.3)$$

f , Hertz biriminden frekansı, $mel(f)$ ise mel frekansını göstermektedir.

Konuşmacı tanımanın bu aşamasında bant geçiren süzgeç (band-pass filter) kullanılır. Bu filtre bankı üçgen bant geçiren özelliktedir. Bu filtrenin yayılım özelliğinden dolayı band genişliği sabit bir mel frekans aralığına sahiptir. Bu filtre bankı, sinyalin HFD ile frekans domenine çevrilmiş haline uygulanır. Bu mel dönüştürücü filtre bankındaki her bir filtre, frekans domeninde bir histogram kutusu (kutuların birbiri üzerine binmesi) oluşturur.



Şekil 2.2. Mel yayımlımlı filtre bankası

2.1.1.5. Kepstrum

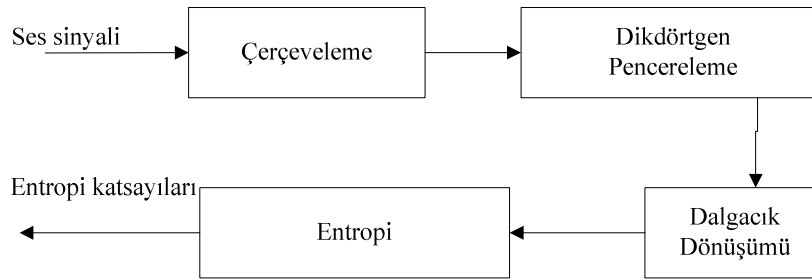
Öznitelik çıkarmanın son aşamasında logaritması alınmış mel spektrumunun frekans domeninden tekrar zaman domenine çevrilmesi gerekmektedir. Bu işlemin sonucunda Mel-Frekansı Kepstrum Katsayıları (MFKK) elde edilir. Her çerçeve için elde edilen vektörlere de öznitelik vektörleri denir. MFKK Denklem 2.4. deki gibi hesaplanır:

$$MFKK_i = \sum_{k=1}^{20} X_k \cos \left[i \left(\frac{k-1}{2} \right) \frac{\pi}{20} \right], \quad i=1, 2, \dots, M \quad (2.4)$$

M , kepstrum katsayılarının sayısını; X_k , k . filtrenin log-enerji çıktısını göstermektedir.

2.1.2. Dalgacık (Wavelet) Dönüşümü ile Entropi Hesaplama

DD ile entropi katsayıları hesaplama işlemi sistemin hem kaydolma hem de saptama aşamasında kullanılır. Entropi katsayılarının elde edilme aşamaları Şekil 2.3'te verilmiştir.



Şekil 2.3. Entropi katsayıları çıkarma

Konuşma ve konuşmacı tanıma sistemlerinde kullanılan spektral ölçümler iki bölüme ayrılır. Birincisi güç, sinyalin sahip olduğu kaba spektral ölçümleri, ikincisi ise spektral genlik, spektrumdaki belirli frekans aralıkları üzerindeki güç ölçümüdür. Konuşma ve konuşmacı tanıma sistemlerinin parametre kümesinde bu iki ölçüm yer alır. Konuşmacı tanıma sistemlerinde güç ölçülerinin kullanımı Denklem 2.5. deki gibidir.

$$P(n) = \frac{1}{N_s} \sum_{m=0}^{N_s-1} (w(m)s(n - \frac{N_s}{2} + m)), \quad (2.5)$$

N gücü hesaplanacak örnek sayısını, $w(m)$ ağırlık fonksiyonunu, $s(n)$ sinyali ve n pencerenin merkezini belirleyen örnek indisini gösterir. Pek çok tanıma sistemi gücü direkt olarak kullanmak yerine insan duyma sistemini taklit ettiği için aşağıdaki Denklem 2.6. kullanır.

$$P_{db}(n) = \log_{10}(P(n) * 10) \quad (2.6)$$

Güç, konuşma ve konuşmacı tanıma sistemlerinde bulunan diğer parametreler gibi çerçeve temelinde hesaplanır. Çerçeve süresi T_f , parametreler kümesinin geçerli olduğu zaman uzunluğu olarak tanımlanır. Çerçeve periyodu ise benzer şekilde arka arkaya yapılan parametre hesaplamaları arasında geçen süredir. Çerçeve oranı ise saniyede hesaplanan çerçevelerin sayısıdır. Aynı zamanda önemli olan bir kavram da gücün hesaplandığı aralıktır. Toplamın hesaplanacağı örnek sayısı olan N_s pencere süresi olarak bilinir. Pencere süresi ile çerçeve süresi birlikte güç hesabındaki sinyalin değişimini izlerler.

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

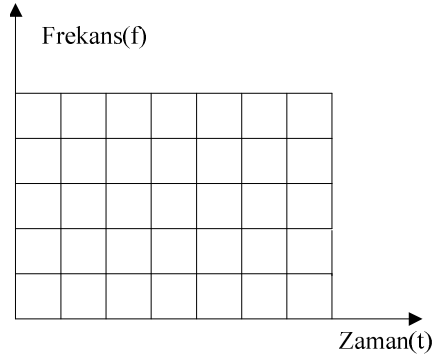
- Dikdörtgen Pencere
- Barlett Pencere
- Hanning Pencere
- Hamming Pencere
- Blackman Pencere
- Kaiser Pencere.

2.1.2.1. Dalgacık Dönüşümü

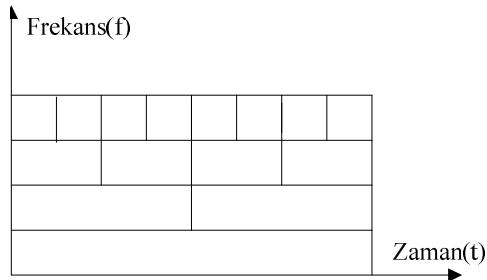
DD ilk defa jeofizik alanında kullanılmıştır [16]. Daha sonra yapılan çalışmalarla bu teori daha sağlamlaştırılarak ayrık zamanlı işaretlere uyarlaması yapılmıştır. Bugün dalgacıklar ve dalgacık paketleri işaret işlemenin farklı alanlarında kullanılmaktadır.

Dalgacık dönüşümü durağan olmayan sinyallerin analizini sağlar. Zaman-frekans çözünürlüğü optimum olarak elde edilir. Analizi yapılacak fonksiyon istenildiği gibi seçilebilir. Dalgacık dönüşümü ve dalgacık analizleri fourier dönüşümünün genelleştirilmiş halidir [17].

Dalgacık dönüşümünde, sinyal sürecinde kaydırılan ölçeklenebilir modüle edilmiş pencereler kullanılır ve her yeni konumda onun spektral davranışı incelenir. Farklı çözünürlüğe sahip sinyallerin frekans-zaman sunumları gerçekleştirilmiş olur. Aslında pencerelenmiş fourier analizinde Şekil 2.4'te görüldüğü gibi zaman frekans seviyeleri için sabit pencereler kullanılır.

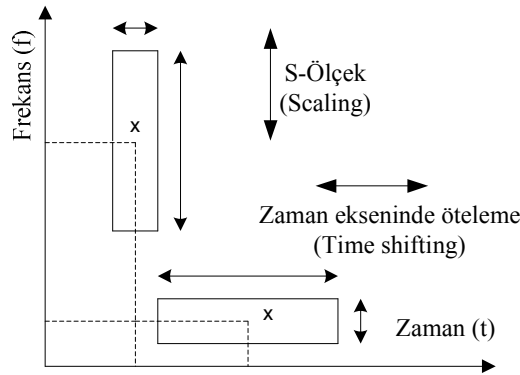


Şekil 2.4. Pencerelenmiş fourier dönüşümü için zaman-frekans diyagramı



Şekil 2.5. Dalgacık dönüşümü için zaman-frekans diyagramı

Fourier dönüşümünde eşit aralıklarla işlem yapılırken dalgacık dönüşümünde düşük frekanslarda geniş zaman aralığında yüksek frekanslarda ise küçük zaman aralığında işlem yapılmaktadır. Yani dönüşümünde boyutları değişik pencereler zaman-frekans domeninde kullanılmaktadır (Şekil 2.5).



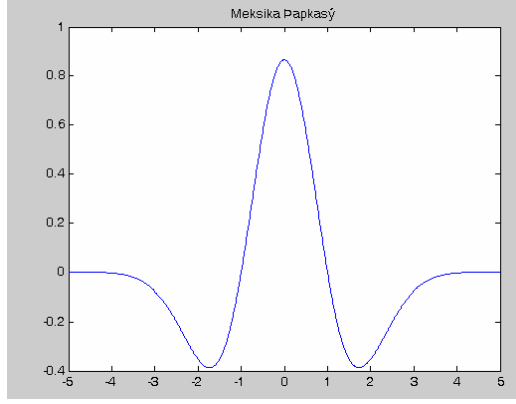
Şekil 2.6. Dalgacık dönüşümü için iki boyutlu zaman-frekans diyagramı

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

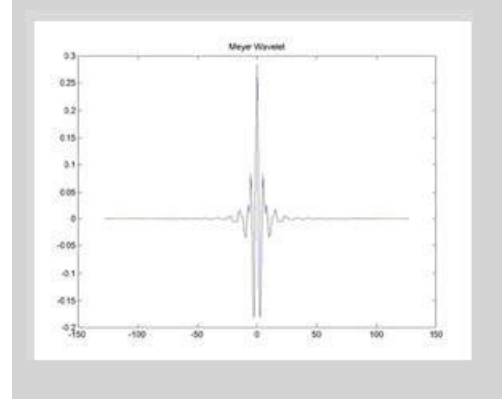
Dalgacık analiz metotlarının geleneksel metotlara göre üstünlükleri aşağıdaki gibi sıralanabilir:

- Spektrumun düşük frekans bölümlerinde, DD oldukça hızlıdır. Bu özelliklerinden dolayı, DD'nin durağan olmayan işaretlerin daha esnek zaman-frekans gösterimlerinin elde edilmesinde etkili bir araç olduğu görülmektedir [16, 18].
- Frekans spektrumundaki farklı bölgeler için, daha basit farklı frekans çözünürlükleri seçebilir.
- Analiz için spektrumdaki birkaç frekans bandı kullanılacaksa, tüm spektrumu hesaplamaya gerek yoktur.

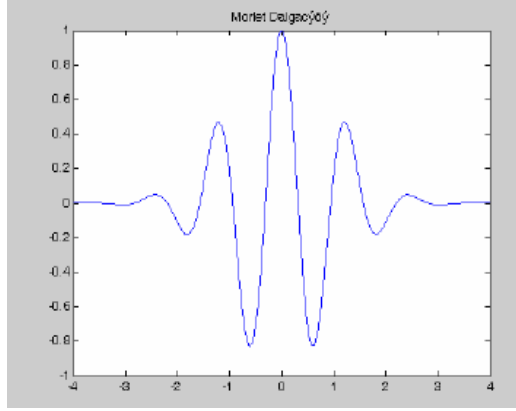
Şekil 2.7'de yaygın olarak kullanılan dalgacıkların bazıları verilmiştir [19].



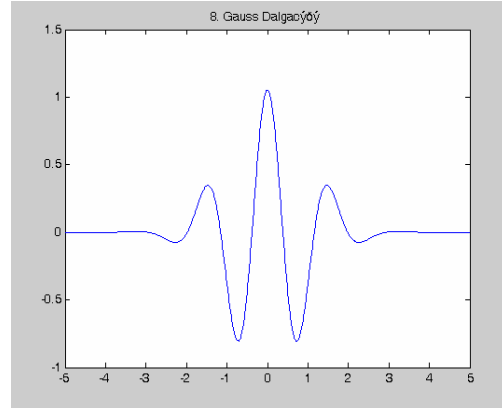
(a) Meksika Şapkası



(b) Meyer Dalgacı



(c) Morlet Dalgacı



(d) Gauss Dalgacı

Şekil 2.7. Yaygın olarak kullanılan dalgacık örnekleri

2.1.2.2. Sürekli Dalgacık Dönüşümü

Sürekli DD (SDD) sıkıştırılmalar, genişlemeler ve dönüşümler ile tüm zaman ve ölçeklerdeki işaretler sayesinde ana dalgacık arasında uygunluk sağlar.

$$SDD(a,b) = - \int_{-\infty}^{\infty} x(t) \frac{1}{\sqrt{a}} \psi\left(\frac{t-a}{b}\right) dt \quad (2.7)$$

Burada $x(t)$ işaret, $\psi(t)$ dalgacıktır. b ölçekleme etmeni olup, farklı frekans seviyelerinde ayrışım filtreleri tanımlar. a ise öteleme etmeni olup her seviye için ayrışım filtrelerini ölçeklendirir. Dönüşüm sonucu ise zamanın bir fonksiyonu olarak işaret ve

dalgacık arasında karşılıklı nasıl bir ilişki (korelasyon) olduğunu gösterir. İşaret ve dalgacık iyi eşleşirse işaret ve dalgacık arasındaki korelasyon da yüksek olur. Uygulamalara bağlı olarak istenilen bir dalgacık türü seçilir. SDD'deki ölçekleme kavramı frekans dönüşümündeki (FD) frekansın tersine benzer. Dalgacık yüksek oranda sıkışık olduğunda işaretin sahip olduğu yüksek frekans detaylarını elde eder. Dalgacık tamamen geniş olursa dalgacığın uzunluğu, işaretin uzunluğu ile daha uygun benzerlik gösterir. Böylece işaretin düşük frekans eğilimleri ortaya çıkar.

2.1.2.3. Ayrık Dalgacık Dönüşümü

Sürekli dalgacık dönüşümü'nün hesaplama yükü fazla olduğu için dönüşümü sağlamak için kullandığı bilgi miktarı oldukça büyüktür. Daha etkili bir yol olan Ayrık Dalgacık Dönüşümü (ADD), ana dalgacığın sadece belirli genişlemelerinde işarete bakar.

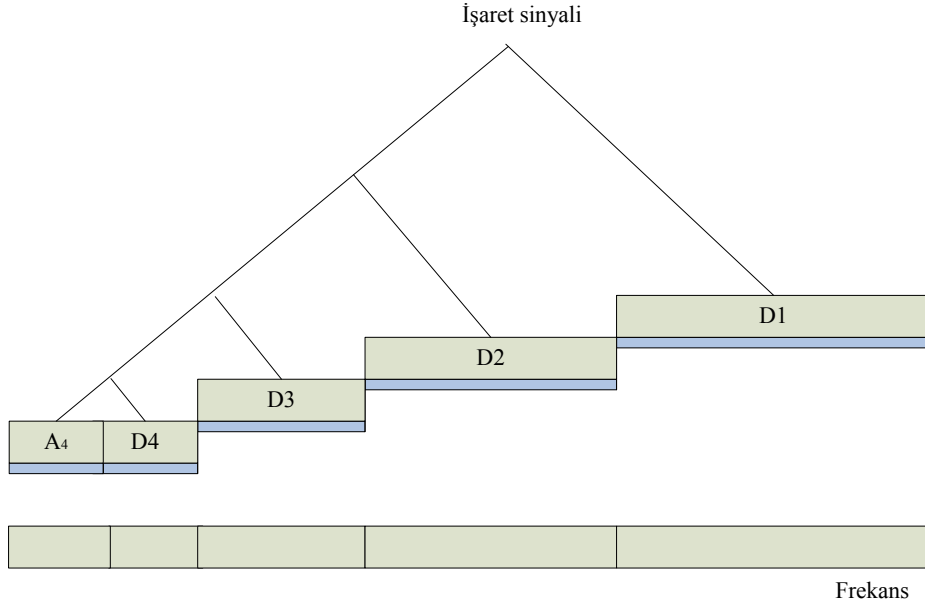
Tek bir filtre belirli bir frekans cevabına sahiptir. İşaret sinyali filtreden geçtikten sonra işaretin içindeki bilginin bir bölümü kaybolur. Bu sebeple tek bir filtre, filtrelenmiş işaretten tekrar asıl işareti elde etmek için kullanılamaz. Çünkü kaybolan bilgi tekrar kazanılmaz. Eğer iki filtre kullanılırsa, biri düşük frekans bilgisini, diğeri yüksek frekans bilgisini tutar. Yani işaret içindeki bilginin tümünü içermiş olur. Böylece bu iki filtrenin çıkışı asıl işaret sinyali tekrar elde etmek için birleştirilebilir. Filtre kümesi veya filtre bankası, spektrum bilgisini ayırmak için kullanılarak işareti gittikçe daha ince frekans bantlarına ayırtırmayı sağlar. Filtrelerin çıkışından elde edilen veri örneklenmelidir. Çünkü işaret, filtre bankasının her bir seviyesinden geçirildiği zaman verinin miktarı iki katına çıkmaktadır.

İdeal olan, işareti daha etkili temsil etmektir. Eğer işaret sinyali m tane zaman ile ayrıştırılırsa asıl işaretten $2m$ zaman daha fazla veriye sahip olması yararlı değildir. Filtre kümesinin, kusursuz yeniden yapılandırma filtrelerine benzer davranabilmesi için özel karakteristiklere sahip olması gerekir. Çünkü gerçek filtreler kusursuz kesim frekanslarına sahip değildirler. Böylece bilginin tümünü tekrar elde etmek için bitişik filtreler arasında bazı binişmeler vardır. Bu görevi yerine getirmek için geliştirilen filtreler kuvadratur filtreler olarak adlandırılır [16, 20]. ADD, bu tür filtrelerden elde edilir. Şekil 2.8'de görüldüğü gibi işaretin ADD'si, ayrışımının her bir seviyesinde yüksek frekans bileşenleri çıkarılarak, asıl işaretin gittikçe kaba bir yaklaşımının elde edilmesini sağlayacak çok

çözünürlüklü bir ayrıştırıcıdır. Asıl işaret sinyali, en yüksek seviyedeki yaklaşık işaret ile daha düşük seviyelerdeki detay işaretleri birleştirilerek yeniden oluşturulabilir.

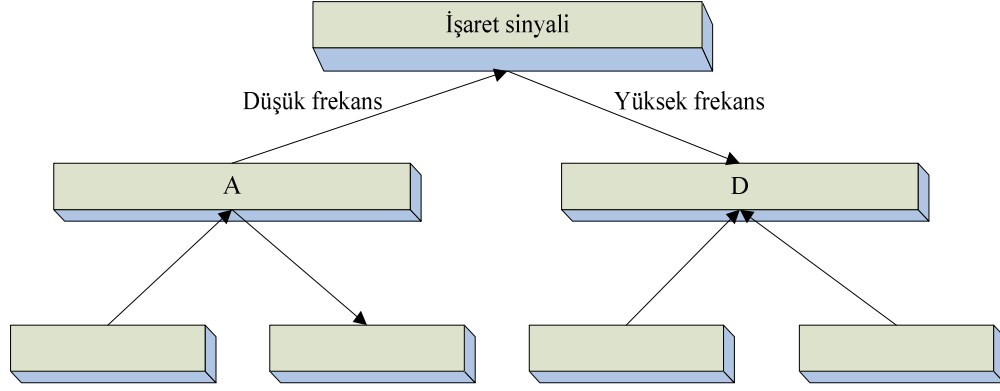
2.1.2.4.Dalgacık Paket Analizi

DPA, ADD'ye benzer olup işaret üzerinde mümkün olabilecek birçok ayrışmaları üretir. Filtre bankası her bir zaman için düşük frekans bileşenlerini tam olarak ayrıştırmak yerine, yüksek frekans bileşenlerini daha iyi ayrıştırmaya olanak sağlar. Böylece tüm zaman-frekans düzlemi Şekil 2.9'da görüldüğü gibi mümkün olabilecek tüm alt bölümlere, farklı zaman-frekans pencereleri ile bölünür [21].



Şekil 2.8. ADD ayrışım ağacı

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



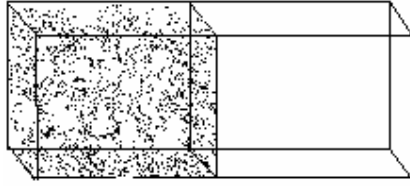
Şekil 2.9. DPA ağaç yapısı

2.1.2.5. Entropi Hesaplama Teknikleri

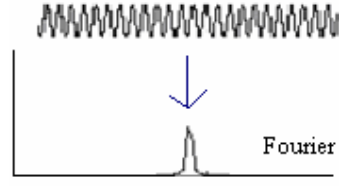
Bir sistemin düzensizliğinin ölçüsü olarak, entropi kavramı tanımlanabilir. Özellikle, termodinamik fizikte bir sistemin düzenliliğini ölçmek için çok fazla kullanılan bir terimdir. Bu tanımlamada düzensizlik kelimesi ile anlatılmak istenen ifade, sistemin toplam enerjisinin mevcut olan tanecikler arasındaki dağılımıdır. Entropi kavramını, haberleşme sahasında ilk olarak Shannon kullanmıştır [16, 22]. Entropi hesaplama tekniği, durağan olmayan bir işaretin düzensizlik derecesini ölçmek için kullanılan uygun bir araçtır [23]. Ayrıca entropi kavramı, bir olayın içerdiği bilginin ortalama miktarını ölçmek amacı ile de kullanılmaktadır [24].

Şekil 2.10 ve Şekil 2.11’de entropi kavramının anlaşılabilmesine yönelik olarak, termodinamik ve işaret işleme alanlarındaki kullanımı karşılaştırmalı olarak gösterilmiştir. Şekil 2.10’dan görüleceği gibi gaz molekülleri belirli bir alana sıkıştırılmıştır. Buradaki gaz molekülleri daha düzenli bir yapıda oldukları için düşük entropi değerine sahiptirler. Şekil 2.11’de ise gaz molekülleri sıkıştırıldığı bölgeden serbest bırakılarak tüm kapalı alana yayılmışlardır. Burada ise gaz moleküllerinin düzenli yapısı bozulduğu için gaz molekülleri yüksek entropi değerine sahiptirler. Termodinamikte kullanılan bu kavram, iletişim teorisinde bir işarettaki düzensizliği veya belirsizliği ölçmek için yararlanabilecek iyi bir araçtır. İşaret işleme açısından Şekil 2.10 ve Şekil 2.11 incelendiğinde, işaret durağan ise frekans bölgesinde dar bir spektrum ile düşük bir entropi değeri elde edilir (Şekil 2.10). İşaret durağan değilse, frekans spektrumu geniş bir banda yayılarak yüksek bir entropi değeri oluşturur. (Şekil 2.11) [16, 20].

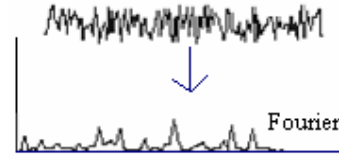
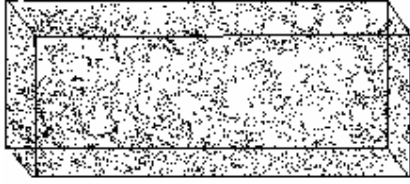
Termodinamik



İşaret işleme



Şekil 2.10. Düşük entropi kavramı



Şekil 2.11. Yüksek entropi kavramı

Entropi ile ilgili ilk çalışmalar, termodinamik alanında 19. yüzyıl da yapılmıştır. 1940'lı yıllarda haberleşme kanallarının doluluğunu tahmin etmek ve haberleşme işaretlerinin içerdiği bilgi miktarını ölçmek için entropi kavramı Shannon tarafından kullanılmıştır [16, 22]. Son yıllarda ise entropi kavramı işaret işleme alanında da yaygın olarak kullanılır hale gelmiştir. Yapılan çalışmalardan bazıları; entegre devre teknolojisinde karbon üzerindeki elektron enerji kaybını bulmak [25], JPEG görüntü kodlamada oluşabilecek gürültüyü tespit etmek [26], EEG işaretlerinin spektral entropilerini kullanarak anestezi derinliğini tahmin etmek [27], biyomedikal NMR işaretlerinin dalgacık dönüşüm uzayından istatistiksel bilgilerini elde ederek işaretin gürültüsüz şeklini kestirmek [23] ve MEG biyomedikal görüntülerin yeniden yapılandırılmasında Norm entropi metodu kullanarak yüksek kalitede MEG görüntüleri elde etmek [18] sayılabilir.

İşaret işlemede yaygın olarak kullanılan entropi hesaplama çeşitleri; Shannon, Eşik (Threshold), Logaritmik Enerji, Norm ve Sure metotlarıdır [16, 28]. Bunun dışında farklı entropi hesaplama teknikleri de bulunmaktadır. Aşağıdaki entropi yöntemlerinde; s işareti, s_i ise işaretin i . katsayısını göstermektedir. İşaret işlemede yaygın kullanılan entropi türleri ve denklemleri;

$$1. \text{ Shannon Entropi: } E(s) = -\sum_i s_i^2 \cdot \log_2(s_i^2) \quad \text{ve } \log(0)=0 \text{ kabul edilerek} \quad (2.8)$$

$$2. \text{ Norm Entropi: } E(s) = \sum_i |s_i|^p \quad \text{ve } 1 \leq p < 2 \quad (2.9)$$

$$3. \text{ Logaritmik Enerji: } E(s) = \sum_i \log_2(s_i^2) \quad \text{ve } \log(0)=0 \text{ kabul edilerek} \quad (2.10)$$

$$E(s) = \sum_i E(s_i) \quad \varepsilon \text{ pozitif bir eşik değeridir}$$

$$4. \text{ Eşik Entropi: } |s_i| > \varepsilon \Rightarrow E(s_i) = 1 \quad \text{ve} \quad |s_i| \leq \varepsilon \Rightarrow E(s_i) = 0 \quad (2.11)$$

$$5. \text{ Sure Entropi: } |s_i| \leq \varepsilon \Rightarrow E(s_i) = \sum_i \min(s_i^2, \varepsilon^2) \quad \varepsilon \text{ pozitif bir eşik değeridir} \quad (2.12)$$

Belirtilen entropi hesaplama tekniklerinin, işaret işlemedeki kesin kullanım alanları belirli olmayıp, uygulamalara göre başarımları değişebilmektedir. Fakat durağan olmayan işaretler için zamana bağımlı entropi hesaplaması daha kullanışlı olacaktır. Bu sebeple işaret katsayıları s_i 'lerin zamana bağımlı olması gerekir [23].

2.2. Konuşmacı Tanımda Kullanılan Sınıflandırma Yöntemleri

Konuşmacı tanıma bir örüntü sınıflandırma işlemidir. Test kelime veya cümlesine ait özellik vektörlerini kullanarak bu test cümlesini hangi konuşmacının söylediğini bulma görevi sınıflandırıcınınır. Bu görevi yerine getirmek için her konuşmacıya ait eğitim verileri ile akustik modeller oluşturulur. Sınıflama aşamasında test kelime veya cümlesine ait özellik vektörlerinin eğitim kümesindeki konuşmacılara ait şablonlarla olan benzerliğine bakılır. Elde edilen benzerlik ölçütü ile konuşmacı tanıma sistemi test cümlesinin kim tarafından söylendiği belirlenir.

Konuşmacı tanıma uygulamalarında birçok sınıflandırıcı teknikleri kullanılmaktadır. Bu bölümde kullanılan sınıflandırıcılar hakkında genel bilgilere yer verilecektir.

2.2.1. Gaussian Karma Model ile Konuşmacı Tanıma

Gaussian karma modelin en önemli özelliği keyfi biçimli dağılım yoğunluklarının düzgün kestirimlerinin yapılabilmesidir. Verilen bir konuşmacının eğitim verisini kullanarak konuşmacı eğitim öznitelik vektörlerinin dağılımına en iyi karşılık gelen Gaussian Karma Model parametrelerinin, yani λ 'nın tahminidir. GKM ile konuşmacı tanıma sistemi için $S = \{1, 2, \dots, S\}$ şeklindeki S adet konuşmacı $\lambda_1, \lambda_2, \dots, \lambda_S$ Gaussian karma modelleriyle gösterilir. Burada amaç verilen gözlem dizisi için maksimum sonsal olasılığı veren konuşmacı modelini bulmaktır.

Bu parametrelerin tahmini için birçok yöntem vardır. En yaygın olarak kullanılan yöntem, en çok olabilirlik tahmin yöntemidir. En çok olabilirlik tahmin yönteminin amacı verilen eğitim verisi için GKM olabilirlik fonksiyonunu maksimize eden model parametrelerini bulmasıdır. T adet eğitim vektöründen oluşan $X = \{X_1, \dots, X_T\}$ dizisi için GKM olasılık fonksiyonu, denklem 2.13. ile hesaplanır.

$$p(X/\lambda) = \prod_{t=1}^T p(\tilde{x}_t/\lambda) \quad (2.13)$$

λ başlangıç modeli ile başlayan beklenti maksimizasyon algoritmasının temeli $p(X|\bar{\lambda}) \geq p(X|\lambda)$ olacak şekilde yeni bir $\bar{\lambda}$ modelinin kestirimine dayanır. Bir sonraki iterasyon için bulunan yeni model başlangıç modeli kabul edilerek bir yakınsama seviyesine ulaşıncaya kadar işlem tekrar edilir. i . akustik sınıf için sonsal olasılık ise denklem 2.14. ile gösterilir:

$$p(i|\tilde{x}_t, \lambda) = \frac{p_i b_i(\tilde{x}_t)}{\sum_{k=1}^M p_k b_k(\tilde{x}_t)} \quad (2.14)$$

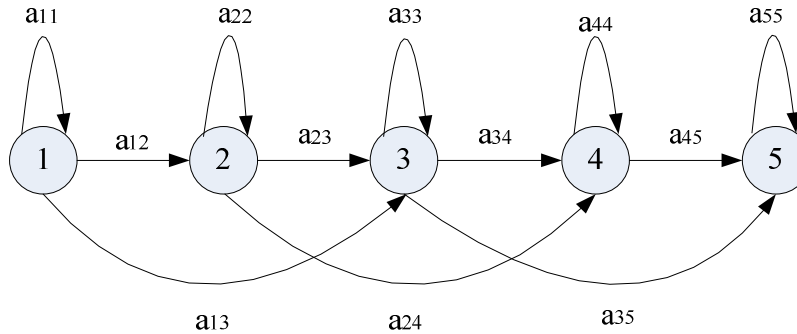
2.2.2. Vektör Nicemleme Tekniğine Dayalı Konuşmacı Tanıma

Vektör nicemleme tekniğine dayalı konuşmacı tanıma sistemlerinde eğitim verileri kullanılarak her bir konuşmacının öznitelik vektörleri elde edilir. Öznitelik vektörleri kişiden kişiye değişen özelliklerdir. Kişilerin gırtlak yapısından dolayı konuşmacıları ayırt

etmeye yarayan niteliklerdir. Her konuşmacının öznitelik vektörleri çıkarıldıktan sonra, bu vektörler kümelendirme algoritmalarından herhangi biri kullanılarak eğitilir ve her bir konuşmacı için optimum referans model olan kod kitabı tasarlanır. Konuşmacıyı tanıma aşamasında ise konuşmacının test verileri kullanılır. Bu test verilerinin de öznitelik vektörleri çıkarıldıktan sonra referans model olarak hazırlanan kod kitabı vektörleriyle en yakın uzaklığı veren kod vektörleri eşleştirilerek sonuca varılır.

2.2.3. Saklı Markov Modelleri ile Konuşmacı Tanıma

Dizilerin modellenmesinde kullanılan diğer bir istatistiksel model ise Saklı Markov Modelleridir (SMM) [29]. SMM sonlu sayıda durumdan oluşan ve her durumun özellik vektörüne ait olasılık yoğunluk fonksiyonunu içeren bir süreçtir. SMM’de durumlar birbirlerine bir durum geçiş işlevi aracılığı ile bağlanmışlardır. Durum geçiş olasılıkları, a_{ij} bir durumdan diğer bir duruma geçiş olasılıklarını belirtmektedir.



Şekil 2.12. Basit bir soldan sağa SMM

Şekil 2.12’de bir markov modeli görülmektedir. Markov modelinde;

rakamlar(1, 2, 3, 4, 5)→Gizli durumlar

a→Geçiş olasılıklarını ifade eder.

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

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

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

2.2.4. Dinamik Zaman Eğirme ile Konuşmacı Tanıma

Dinamik Zaman Eğirme(DZE) konuşmacı tanıma uygulamalarının ilk dönemlerinde kullanılmaya başlanan bir sınıflandırma tekniğidir.

Konuşmanın dalga şeklinde; konuşma hızı, ses yüksekliği veya stres gibi nedenlerden dolayı konuşan kişi aynı olsa bile farklılıklar görülür. Bazı örüntüler yüksek genlik ve uzun süreye sahipken bazıları ise daha düşük genlikli ve kısa süreli olabilir. Konuşma hızındaki değişikliklerden dolayı bir konuşmacının farklı zamanlarda söylediği aynı cümleler arasında zamanlama açısından farklılıklar ortaya çıkmaktadır. Zamanlamadaki bu problem test cümlesi ile eğitim cümlesi arasındaki önemli benzerlikleri bulmak için DZE algoritması kullanılır.

DZE, test cümlesi ile eğitim şablonunu karşılaştırırken muhtemel yollardan optimum olanı bulmaya çalışır. Verilen bir referans (eğitim) şablonu R ve test cümlesi T için; $R(n)$ ve $T(n)$ sırasıyla eğitim ve test cümlelerindeki çerçeve sayıları olsun. DZE, T 'nin zaman eksenini olan n 'yi R 'nin zaman eksenine eşleştiren bir $m = w(n)$ fonksiyonu bulmaya çalışır.

$$D = \min_{w(n)} \left[\sum d(T(n), R(w(n))) \right] \quad (2.15)$$

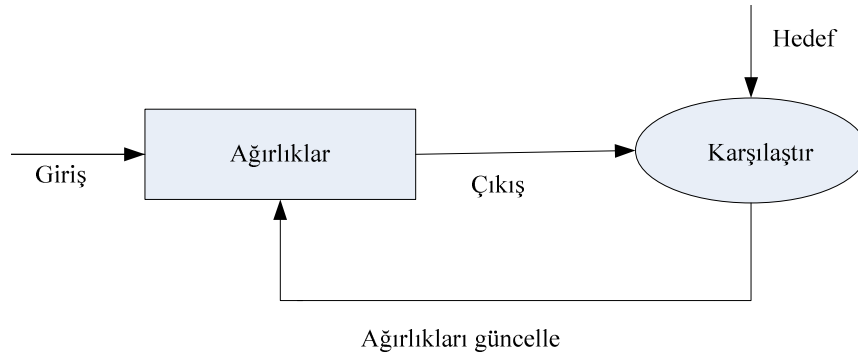
Denklemden d , T cümlesine ait n . çerçeve ile R cümlesine ait $w(n)$. çerçeve arasındaki bir uzaklık ölçütüdür. D ise en iyi yolu veya en iyi eşleşmeyi temsil eden uzaklık ölçütüdür. Verilen bir test cümlesine ait özellik vektörü dizisi için, DZE bütün referans şablonlar arasından en iyi eşleşme uzaklıklarını bulur. Sistem de bu uzaklıkları saklar ve test cümlesinin en küçük uzaklık veren şablona ait olduğu kararı verir. DZE genellikle metine bağımlı konuşmacı tanıma uygulamalarında kullanılmaktadır [30].

3. AKILLI YÖNTEMLER

3.1. Yapay Sinir Ağları (YSA)

YSA insan beyninin öğrenme yoluyla yeni bilgiler üretebilme, oluşturabilme ve keşfedebilme gibi yetenekleri herhangi bir yardım almadan otomatik olarak gerçekleştirebilen bilgisayar sistemleridir. YSA insan beyninden esinlenerek öğrenme sürecinin matematiksel olarak modellenmesi sonucunda ortaya çıkmıştır. Bu sistemlerin başlıca belirgin özellikleri uyarlanabilen, öğrenebilen, tamamıyla paralel ve paralel dağıtılmış bir hafızaya sahip olmalarıdır [31].

Sinir ağları genellikle eğitilerek belirli bir girişe karşılık olarak verilen, istenilen çıkışı elde etmek için kullanılırlar. Bu durum aşağıdaki Şekil 3.1’de verilmiştir. Ağın çıkışı arzu edilen hedefe ulaşıncaya kadar çıkış ile hedef karşılaştırılarak ağın eğitimi (ağırlık değerlerinin ayarlanması) gerçekleştirilir. En basit tanımıyla yapay sinir ağları, bilinen en iyi eğri uydurma tekniklerinden birisidir [32].

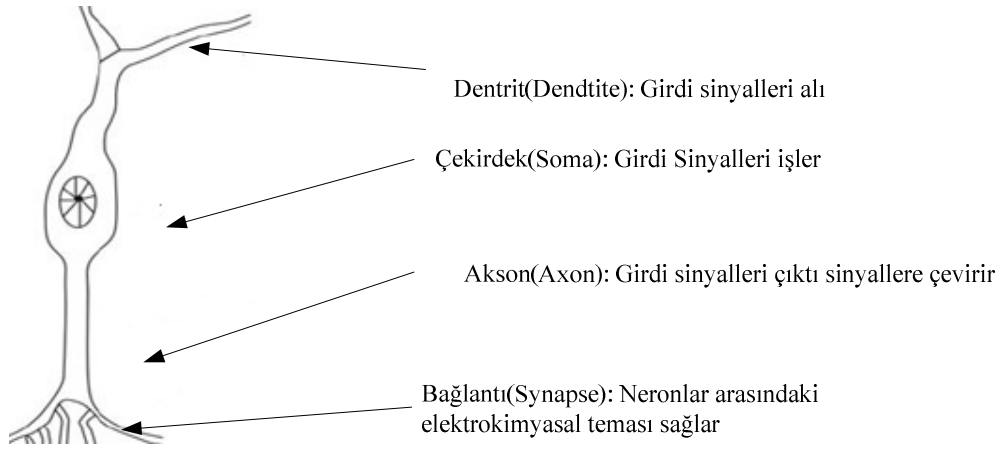


Şekil 3.1. Yapay sinir ağlarının yapısı

3.1.1. Biyolojik Sinir Ağları ile Yapay Sinir Ağlarının Karşılaştırılması

Biyolojik sinir ağının temel elemanı olan sinir hücresi, sinir sistemi içerisindeki görev ve fonksiyonlarına göre değişik şekil ve büyüklükte olabilir (Şekil 3.2). Sinir sisteminin en basit yapısı olan bir nöronun bir ucunda “dendrit” adı verilen ve hücreye, diğer hücrelerden veya dış dünyadan gelen bilgileri toplayan bağlantı elemanı bulunurken diğer ucunda ise

tek bir life benzeyen, hücreden diğer hücelere aynı zamanda da dış dünyaya bilgi taşıyan “akson” adı verilen bağlantı elemanı vardır. Akson diğer hücelerle birleşme esnasında dağınık dallara ayrılır. Bu iki uçtaki bağlantı noktalarının, elektrofizyolojik olarak hücelerde bulunan bilgileri işlemede önemli bir yeri vardır. Hücelerin birbiri ile elektrik işaretleri vasıtasıyla irtibat kurduğu gözlemlenmiştir. İşaretler, bir hücrenin aksonundan, diğer hücrenin dentritine gönderilir. Bir akson birden fazla dentrit ile bağlantı kurabilir. Bu bağlantıların yapıldığı yer “synaps” olarak adlandırılır.

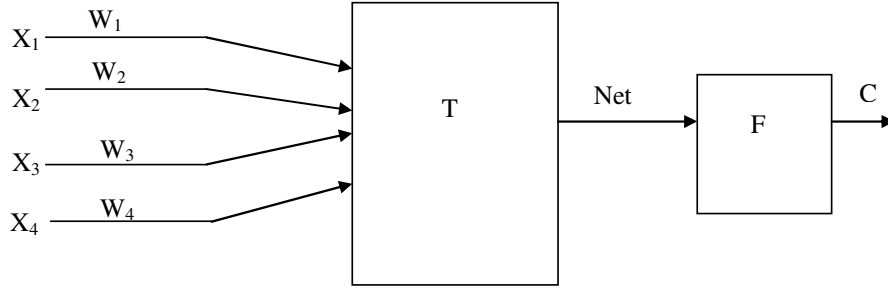


Şekil 3.2. Biyolojik nöronun şematik yapısı

Hüceler elektrik sinyalini hücre duvarındaki voltajı değiştirerek üretebilirler. Bu ise hücrenin içinde ve dışında dağılmış olarak bulunan iyonlar vasıtasıyla gerçekleşir. Bu iyonlar potasyum, sodyum, kalsiyum ve klordur. Nöronun içinde potasyum yoğunluğu daha çok bulunurken nöronun dışında da sodyum yoğunluğu daha çok bulunur.

Bir hücre, diğer hücreye elektrik enerjisini bu kimyasal iyonlar aracılığıyla transfer eder. Bazı iyonlar elektrik ve magnetik kutuplaşmalara sebep olurken, bazıları kutuplaşmadan kurtulup hücre zarını açarak iyonların hücreye geçişini kolaylaştırır. İnsan sinir ağı sistemi, bu şekilde çalışan milyonlarca hücrenin bir araya gelmesinden meydana gelir. Beyinde korteks denilen bölgede her nöronun bir karşılığı bulunur. Bir nöronun çıkışı kendisine bağlı olan bütün nöronlara iletilir. Fakat korteks, işin yapılabilmesi için hangi nöron harekete geçirilecekse sadece ona komut gönderir [16, 33].

Bir yapay sinir ağının yapısını belirleyen birçok faktör vardır. Yapay sinir hüceleri veya işlem elemanları, sinir ağının yapısal modelleri ile ağın sahip olduğu kural ve stratejisi bunların başında gelir [33].



Şekil 3.3. Bir yapay sinir ağı hücresi modeli

X_i : Giriş W_i : Ağırlık T : Toplam Fonksiyonu
 F : Eşik Fonksiyonu $\mathcal{C}$: Çıkış

Şekil 3.3'te verilmiş olan bir yapay sinir ağı hücresinin matematiksel ifadesi Denklem 3.1. deki gibidir.

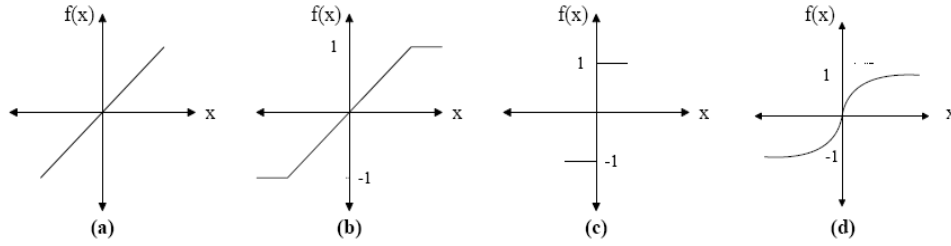
$$Net = \sum_{i=1}^4 w_i \cdot x_i = w_1 \cdot x_1 + w_2 \cdot x_2 + w_3 \cdot x_3 + w_4 \cdot x_4 \quad (3.1)$$

$$\mathcal{C} = F(Net)$$

Tıpkı biyolojik sinir hücresi gibi, işlem elemanına birden fazla giriş gelir ama sadece bir çıkış gider. Girişler, dentritler gibi diğer yapay hücrelerden bağlantılar aracılığıyla işlem elemanına bilgi gelmesini sağlarlar. Bazı durumlarda bir işlem elemanı kendisine de bilgi geri gönderebilir (geri besleme) [34]. Bu bilgiler elemanlar arasında bulunan bağlantı hatları üzerinde depolanır. Her bağlantının sahip olduğu bir ağırlığı vardır. Bu ağırlık bir işlem elemanının diğer üzerindeki etkisini gösterir. Ağırlık büyüdükçe etki de büyür. Ağırlığın sıfır olması hiç bir etkinin olmaması, negatif olması ise etkinin ters yönde olması anlamına gelir. Bu ağırlıklar sabit olabilecekleri gibi değişken de olabilirler.

YSA bilgi işleme sistemlerinin yeni neslini temsil eder. Genellikle YSA; model seçimi ve sınıflandırılması, işlev tahmini, en uygun değeri bulma ve veri sınıflandırılması gibi işlerde başarılı sonuçlar elde etmiştir. Geleneksel bilgisayarlar ise özellikle model seçme işinde verimsizken sadece algoritmaya dayalı hesaplama işlemleri ile kesin aritmetik işlemlerde hızlıdırlar [35].

Eşik fonksiyonları, işlem elemanlarının sınırsız sayıdaki girişini önceden belirlenmiş sınırdaki çıkış olarak belirlerler. En çok kullanılan dört tane eşik (aktivasyon) fonksiyonu vardır. Bunlar, (a) lineer, (b) rampa, (c) basamak ve (d) sigmoid fonksiyonudur. Bu fonksiyonlar Şekil 3.4'te gösterilmiştir.



Şekil 3.4. Yapay sinir ağları için kullanılan eşik fonksiyonları

Şekil 3.4.(a) , lineer fonksiyonunu gösterir ve formülü (Denklem 3.2);

$$f(x) = \begin{cases} +\tau, & x > 0 \text{ ise} \\ -\delta, & \text{diğer durumlarda} \end{cases} \quad (3.2)$$

Şekil 3.4.(b)'deki lineer fonksiyon $[-\tau, +\tau]$ sınırları arasında kısıtlandığında rampa eşik fonksiyonu olur ve formülü (Denklem 3.3);

$$f(x) = \begin{cases} +\tau, & x \geq t \text{ ise} \\ x, & |x| < t \\ -\tau, & x \leq -t \end{cases} \quad (3.3)$$

$+\tau$ ($-\tau$) işlem elemanının maksimumu (minimumu) çoğu zaman doyma seviyesi olarak adlandırılan çıkış değeridir. Eğer eşik fonksiyonu bir giriş işaretine bağlı ise yaydığı $+\tau$ giriş toplamı pozitif, bağlı değilse eşik basamak fonksiyonu $[-\delta]$ olarak adlandırılır.

Şekil 3.4.(c)'de gösterilen basamak fonksiyonun formülü (Denklem 3.4);

$$f(x)=\begin{cases} +\tau, & x>0 \\ -\tau, & x<0 \end{cases} \quad (3.4)$$

Son ve en önemli eşik fonksiyonu (bu çalışmada kullanılan) sigmoid fonksiyonudur. Şekil 3.4.(d) de gösterilen S biçimindeki sigmoid fonksiyonu; seviyeli, lineer olmayan çıkış veren, sınırlı, monoton artan fonksiyondur. Denklem(3.5);

$$f(x) = \frac{1}{1+e^{-x}} \quad (3.5)$$

Toplama fonksiyonu, bir işlem elemanına gelen net girişi hesaplayan bir fonksiyondur. Net giriş genellikle gelen bilgilerin, ilgili bağlantılarının ağırlıkları ile çarpılarak toplanması sonucunda hesaplanır. Bu nedenle adına, toplama fonksiyonu denmiştir. Eşik fonksiyonu da, toplama fonksiyonu tarafından hesaplanan net girişi alarak, işlem elemanının çıkışını belirleyen fonksiyondur. Genellikle türevi alınabilen bir fonksiyon olması tercih edilir (Şekil 3.4).

Toplama ve çıkış fonksiyonları, ilgili probleme bağlı olarak farklı şekillerde olabilirler. İşlem elemanının çıkış ünitesi ise çıkış fonksiyonunun ürettiği dürtüyü diğer işlem elemanlarına veya dış dünyaya aktarma görevini yapar. İşlem elemanları ağıın topolojik yapısının etkisiyle tamamen birbirinden bağımsız ve paralel olarak çalışabilirler.

3.1.2. Yapay Sinir Ağlarının Yapıları ve İşlem Elemanları

YSA gösterilen örnekten öğrenerek kendi kurallarını oluşturan bir sistemdir. Yapay sinir ağının sahip olduğu bilgi, işlem elemanları arasındaki bağlantı hatları üzerinde saklanarak ağırlıklar aracılığıyla gösterilir. Ağ, eldeki mevcut örneklerden genellemeler yaparak, olaylar hakkında girişler ve çıkışlar arasındaki ilişkiyi öğrenir. Bu genelleme ile yeni oluşan ve ortaya çıkan daha önce hiç görülmemiş olaylar hakkında karar verir. Yani ağa, bir örnek olay verildiğinde, giriş katmanından alınıp ara katmanlarda işlenerek, ağın o olay hakkında ürettiği sonuç, çıkış katmanına gönderilir. Bu bilgiyi işleme, ağın sahip olduğu tecrübeye bağlı olarak, bilginin ara katmanlarda ilişkilendirilmesi ile gerçekleştirilir. Bu ilişkilendirme modelden modele farklılık gösterir. Örneğin, ara katmanlardaki işlem elemanları sahip oldukları bağlantılar sayesinde kendi kararlarını

üretir ve çıkış katmanındaki işlem elemanlarına iletirler. Çıkış katmanındaki işlem elemanları da, yine ilgili ağırlıkları kullanarak ağın en son kararını meydana getirir. Bu ağırlıklar, ilgili olayın belirli özelliklerini hafızada tutan elemanlar gibi düşünülebilirler. Bilgi işleme ise (örüntü tanıma gibi), bir olay gösterildiğinde hafızadan ilgili özellikleri çağırmak ve bunlar ile ilgili girişleri birlikte analiz ederek karar vermek şeklinde yorumlanabilir [36].

Yapay sinir ağlarında bilgi, ağdaki bağlantıların sahip olduğu ağırlıklarda depolanır. Bir ağda öğrenme kısaca, istenen bir işlevi yerine getirecek şekilde ağırlıkların belirlenmesi sürecidir. Yapay sinir ağlarında öğrenme, sinirler arasındaki ağırlıkların değiştirilmesi ile gerçekleşmektedir [37]. Yapay sinir ağının öğrenme işleminin gerçekleşebilmesi için, sahip olduğu bütün ağırlıkların, ilgili problemde öğrenilmesi istenen özellikleri genelleştirecek şekilde doğru değerlere sahip olması gerektirir. Bu doğruluk ne kadar artarsa ağın sınıflandırma kabiliyeti dolayısı ile öğrenme işlemi o kadar iyi olacaktır. Doğru ağırlık değerleri, bir öğrenme kuralına göre belirlenir. Çoğunlukla bağlantılara başlangıç değerleri olarak rastgele ağırlıklar atanarak bu ağırlıklar eldeki örnekler incelendikçe bir kurala göre değiştirilerek doğru ağırlık değerleri bulunmaya çalışılır. Öğrenme yöntemleri temelde üç grupta toplanır [33, 36, 38].

Eğitici Öğrenme: Dışarıdan bir eğiticinin sinir ağına müdahalesi söz konusu olan bir öğrenmedir. Eğitme işlemi, sinir ağına giriş ve çıkış bilgileri sunmaktan oluşur [37]. Yapay sinir ağına giriş / çıkış ikilisinden oluşan örnekler sunulur. Bu ikili, ağın öğrenmesi gereken özellikleri temsil ederek daha çok eğitime kümesi olarak tanımlanır. Ağ giriş bilgisini alır ve o anki bağlantı ağırlıklarının temsil ettiği bilgi ile bir çıkış oluşturur. Bu çıkış, olması gereken çıkış ile karşılaştırılarak aradaki hata tekrar ağa gönderilerek ağırlıklar bu hatayı azaltacak şekilde değiştirilirler.

Takviyeli Öğrenme: Bu yöntemde de yine eğiticiye ihtiyaç vardır. Bu yöntemin eğitici öğrenme yönteminden tek farkı, eğiticinin ağın üretmesi gereken sonuç yerine, onun ürettiği sonucun sadece doğru ya da yanlış olduğunu göstermesidir. Bu işlem ağa bir takviye işaretinin gönderilmesi sonucu gerçekleştirilir. Bu yöntem, ilgili örnek için beklenen çıkışın oluşturulmadığı durumlarda faydalı olmaktadır.

Eğitici Öğrenme: Bu yöntemde herhangi bir eğiticiye ihtiyaç yoktur. Bundan dolayı çoğu zaman bu yönteme kendi kendine organize olma da denir. Ağ, kendine gösterilen örnekleri alarak belli bir şarta göre sınıflandırma yapar. Bu şart önceden bilinmeyebilir. Ağ, kendi öğrenme şartlarını yine kendisi oluşturur.

Öğrenme işlemi, başlangıçta rastgele seçilmiş ağırlıkların, belirli bir kurala göre değiştirilmesine dayandığı daha önce belirtilmişti. Bu öğrenme kurallarının önemli olanlarından bazıları aşağıda verilmiştir. Bu algoritmalarda, $w_i = [w_{i1} \ w_{i2} \ ... \ w_{in}]$ vektörü ağdaki i . hücreye gelen n tane girişin ağırlık katsayılarını belirtirken, $x = [x_1 \ x_2 \ ... \ x_n]$ vektörü ağa gelen n tane girişi, y_i ve d_i 'de sırasıyla i . hücrenin hesaplanan ve istenen çıkışını temsil etmektedirler.

Hebb Öğrenme Algoritması: Hebb öğrenme algoritmasında amaç, hücrenin çıkış üretmesine sebep olan bağlantıların ağırlık katsayılarını büyüterek etkilerini daha da arttırmaktır. Buna göre bu değişim aşağıda Denklem 3.6. ve Denklem 3.7. deki gibi tanımlanmıştır.

$$w_i(t+1) = w_i(t) + \varepsilon \cdot f(w_i(t) \cdot x(t)) \cdot x(t) \quad (3.6)$$

veya

$$w_i(t+1) = w_i(t) + \varepsilon \cdot y_i(t) \cdot x(t) \quad (3.7)$$

Burada ε 'a öğrenme katsayısı veya öğrenme oranı denir. Dışarıdan istenilen bir çıkış uygulanmadığı için eğiticişiz bir öğrenme algoritmasıdır.

Yarışmacı Öğrenme Algoritması: Bu algoritmada esas olan, giriş işaretine en yakın olacak bağlantı ağırlık katsayılarını bulmaktır. Bu, w 'lara ilişkin nöron, kazanan nöron olarak isimlendirilir. Sadece kazanan nöron bir çıkış üretmekte dolayısıyla bu nörona ilişkin w 'lar değişime uğramaktadır. Eğiticişiz öğrenmenin tipik bir örneği olan bu algoritmanın sonucunda ağa gelen girişler sınıflandırılmaktadır. Denklem 3.15. incelersek;

$$w(t+1) = w(t) + \varepsilon \cdot (x(t) - w_m(t)) \quad (3.15)$$

Burada w_m kazanan hücrenin bağlantı ağırlık katsayısı olup Denklem 3.16. daki gibi tanımlanmıştır.

$$w_m \cdot x = \max(w_i \cdot x) \quad i = 1, 2, \dots, p \quad (3.16)$$

Algılayıcı (Perceptron) Öğrenme Algoritması: Rosenblatt tarafından önerilen algılayıcı öğrenme algoritmasında w değişimi, nöron çıkışı ile istenilen çıkışın farkı ile orantılıdır. Dolayısıyla eğitici bir öğrenme algoritmasıdır (Denklem 3.8).

$$w_i(t+1) = w_i(t) + \varepsilon [d_i(t) - \text{sgn}(w_i(t).x(t))]x(t) \quad (3.8)$$

Ağın ürettiği çıkış ile istenen çıkış birbirinden farklı ise, hata azaltılacak yönde hareket edilir. Hata sıfır olduğu zaman w değişimi durmuş olur.

Delta Kuralı (Eğim düşme) Öğrenme Algoritması: “Delta kuralı” öğrenme algoritması, eğim düşme (gradient descent) öğrenme algoritması, sadece türetilebilir fonksiyonlara sahip olan ağlara uygulanır. Bu algoritma eğitici bir öğrenme algoritmasıdır. w değişimi, hatanın eğiminin tersi yönde olur. Böylece, hata fonksiyonunun minimum olduğu yere doğru hareket edilmiş olur (Denklem 3.9).

$$w_i(t+1) = w_i(t) - \varepsilon \cdot \frac{\partial E(t)}{\partial w(t)} \quad (3.9)$$

Burada $E(t)$, hata fonksiyonunu temsil etmektedir ve Denklem 3.10. gibi tanımlanır.

$$E(t) = \frac{1}{2} (d_i(t) - y_i(t))^2 \quad (3.10)$$

$$E(t) = \frac{1}{2} [d_i(t) - f(w_i(t).x(t))]^2 \quad (3.11)$$

$$\frac{\partial E(t)}{\partial w(t)} = -[d_i(t) - f(w_i(t).x(t))] \frac{\partial f(w_i(t).x(t))}{\partial (w_i(t).x(t))} . x(t) \quad (3.12)$$

Bilinen hata geriye yayılım algoritmasının esası da, bu öğrenme kuralına dayanır.

Widrow-Hoff Öğrenme Algoritması: Widrow-Hoff öğrenme algoritması, eğitici bir öğrenme algoritması olup w değişimi ağın çıkışından bağımsızdır (Denklem 3.13.)

$$w_i(t+1) = w_i(t) + \varepsilon (d_i(t) - w_i(t).x(t)).x(t) \quad (3.13)$$

İlintili Öğrenme Algoritması: Hebb öğrenme algoritmasının eğitici uyarlanmış halidir. $f(w_i(t).x_i(t))$ yerine $d_i(t)$ istenen çıkış alındığında Denklem 3.14 elde edilir.

$$w_i(t+1) = w_i(t) + \varepsilon.d_i(t).x(t) \quad (3.14)$$

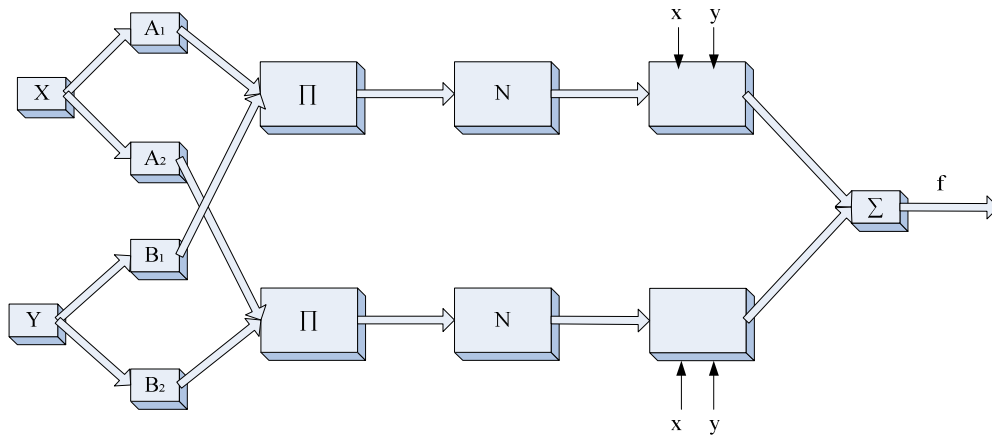
Outstar Öğrenme Algoritması: Eğitici öğrenme algoritmalarından olan bu öğrenme algoritmasında amaç, w 'ları istenen çıkışa benzetebilmektir (Denklem 3.17).

$$w(t+1) = w(t) + \varepsilon.(d_i(t) - w(t)) \quad (3.17)$$

3.2. Uyarlamalı Sinirsel Bulanık Ağ (USBA) Mimarisi

Sinirsel bulanık sistemler, yapay sinir ağlarının paralel hesaplayabilme ve öğrenme kabiliyeti ile bulanık mantığın uzman bilgisini kullanarak sonuçlar çıkarabilme özelliklerinin birleşiminden oluşur. Böylece sinirsel bulanık sistemler sayesinde YSA daha anlaşılır bir hale gelir [39].

USBA, ele alınan problem için oluşturulan yapıya göre olası tüm kuralları atayabilmekte ya da kuralların veriler yardımıyla uzman tarafından atanmasına imkan vermektedir. USBA'nın kural oluşturabilmesi veya kural oluşturulmasına olanak sağlaması uzman görüşlerinden faydalanmış olduğu anlamına gelmektedir. USBA 5 katmandan meydana gelir. Bu sistem Şekil 3.5'te gösterilmiştir. USBA yapısındaki her katmana ait düğüm işlevleri ve katmanların işleyişi sırasıyla şöyledir [40].



Şekil 3.5. 2 girişli 4 kurallı bir USBA sınıflandırıcı yapısı

Bu USBA yapısı yukarıda da görüldüğü gibi 5 katmandan meydana gelmiştir. Burada x ve y giriş, f ise çıkış olarak alınırsa temel kural yapısı Denklem 3.8. deki gibi yazılabilir:

$$\text{Eğer } x \text{ } A_I \text{ ve } y \text{ } B_I \text{ ise } f_1 = p_1x + q_1y + r_1 \quad (3.18)$$

Burada p_i , q_i ve r_i lineer çıkış parametrelerini göstermektedir. İki girişli ve bir çıkışlı olan bir USBA' nın temel yapısı Şekil 3.5'te görülmektedir. Bu yapıda, 5 katman ve 4 adet eğer-ise kuralı kullanılmıştır.

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

$$O_{1,i} = \mu_{Ai}(x), \quad i=1,2 \quad O_{1,i} = \mu_{Bi-2}(y) \quad i=3,4 \text{ için} \quad (3.19)$$

Burada x ve y girişleri temsil eder. Bu katmanın sahip olduğu çıkış, kuralların varsayım (eğer) kısımlarının üyelik fonksiyonlarına olan üyelik dereceleridir. Kullanılan üyelik fonksiyonu Denklem 3.20. de verilmiştir. Buradan da görüldüğü üzere üyelik fonksiyonu olarak Gauss üyelik fonksiyonu kullanılmıştır.

$$\mu_{Ai}(x), \mu_{Bi-2}(y) = \exp(-(x_i - c_i)/(a_i))^2 \quad (3.20)$$

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

$$O_{2,1} = w_i = \mu_{Ai}(x) * \mu_{Bi}(y), \quad i=1,2 \quad (3.21)$$

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

$$O_{3,i} = \bar{w}_i = w_i / (w_1 + w_2), \quad i=1,2 \quad (3.22)$$

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

$$O_{4,i} = \bar{w}_i * f_i = w_i * (p_i x + q_i y + r_i \quad (3.23)$$

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

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

$$O_{5,i}=\text{toplam çıkış} = \sum_i \bar{w}_i f_i = \frac{\sum_i w_i f_i}{\sum_i w_i} \quad (3.24)$$

3.2.1. USBA İçin Geri Yayılımlı Öğrenme Algoritması

Geri yayılımlı öğrenme algoritmasında ağın çıkışından elde edilen hata değeri giriş katmanına kadar geriye doğru yansıtılarak gerekli ağırlık değişkenlerinin ayarlanması gerçekleştirilir [41]. Burada amaç e hata ölçütünü öğrenme işlemi sonunda bütün giriş örnekleri için sıfıra götürmektir. Ağın çıkışında bulunan hata değeri, istenen değer d ile gerçek çıkış f arasındaki farktır ve Denklem 3.25. deki şekilde ifade edilir.

$$e = d - f \quad (3.25)$$

Bu hata değerine en küçük kareler yöntemi uygulanarak hata ölçütü olan E değeri

Denklem 3.26. da verildiği gibi hesaplanır.

$$E = \frac{1}{2} e^2 \quad (3.26)$$

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

5. Katman: Bu katmanda herhangi bir ağırlık değeri ayarlaması yapılmaz, sadece çıkışta bulunan hatanın Denklem 3.27. deki gibi katman çıkışına yansımaları hesaplanır.

$$\delta^5 = -\frac{\partial E}{\partial f} = -\left[\frac{\partial E}{\partial e} \frac{\partial e}{\partial f} \right] = e \quad (3.27)$$

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

$$\Delta p_i = -\frac{\partial E}{\partial p_i} = \left[-\frac{\partial E}{\partial f} \right] \left[\frac{\partial f}{\partial p_i} \right] \quad (3.28)$$

$$f = \sum_i \bar{\mu}_i (p_i x + q_i y + r_i) \quad i = 1, 2 \quad (3.29)$$

$$\frac{\partial f}{\partial p_i} = \bar{\mu}_i x \quad (3.30)$$

buradan,

$$\Delta p_i = \delta^5 \bar{\mu}_i x \quad (3.31)$$

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

$$\Delta q_i = \delta^5 \bar{\mu}_i y \quad (3.32)$$

$$\Delta r_i = \delta^5 \bar{\mu}_i \quad (3.33)$$

olur.

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

$$n_i(k+1) = n_i(k) + \eta_n \Delta n_i(k+1) \quad (3.34)$$

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

$$\delta_i^3 = -\frac{\partial E}{\partial \bar{\mu}_i} = \left[-\frac{\partial E}{\partial f} \right] \left[\frac{\partial f}{\partial \bar{\mu}_i} \right] \quad i = 1, 2 \quad (3.35)$$

$$f = \sum_i \bar{\mu}_i f_i \quad (3.36)$$

$$\frac{\partial f}{\partial \bar{\mu}_i} = f_i \quad (3.37)$$

$$\delta_i^3 = \delta^4 f_i \quad (3.38)$$

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

$$\delta_i^2 = -\frac{\partial E}{\partial \mu_i} = \left[-\frac{\partial E}{\partial f} \frac{\partial f}{\partial \bar{\mu}_i} \right] \left[\frac{\partial \bar{\mu}_i}{\partial \mu_i} \right] \quad i = 1, 2 \quad (3.39)$$

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

$$\frac{\partial \bar{\mu}_i}{\partial \mu_i} = \left[\frac{a - \mu_i}{a^2} \right] \quad (3.40)$$

olarak bulunur. Bu ifade Denklem 3.41. de yerine konularak aşağıdaki gibi son durum elde edilir.

$$\delta_i^2 = \delta_i^3 \left[\frac{a - \mu_i}{a^2} \right] \quad (3.41)$$

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

$$\delta_i^1 = -\frac{\partial E}{\partial \mu_{Ai}} = -\left[\frac{\partial E}{\partial f} \frac{\partial f}{\partial \bar{\mu}_i} \frac{\partial \bar{\mu}_i}{\partial \mu_i} \right] \left[\frac{\partial \mu_i}{\partial \mu_{Ai}} \right] \quad i = 1, 2 \quad (3.42)$$

$$\frac{\partial \mu_i}{\partial \mu_{Ai}} = \mu_{Bi} \quad (3.43)$$

$$\delta_i^1 = \delta_i^2 \mu_{Bi} \quad (3.44)$$

$$\delta_{i+2}^1 = -\frac{\partial E}{\partial \mu_{Bi}} = -\left[\frac{\partial E}{\partial f} \frac{\partial f}{\partial \bar{\mu}_i} \frac{\partial \bar{\mu}_i}{\partial \mu_i} \right] \left[\frac{\partial \mu_i}{\partial \mu_{Bi}} \right] \quad (3.45)$$

$$\frac{\partial \mu_i}{\partial \mu_{Bi}} = \mu_{Ai} \quad (3.46)$$

$$\delta_{i+2}^1 = \delta_i^2 \mu_{Ai} \quad (3.47)$$

$$\mu_{Ai} = \exp \left[-\left(\frac{x - m_i}{\sigma_i} \right)^2 \right] \quad v_i = -\left[\frac{x - m_i}{\sigma_i} \right]^2 \quad \mu_{Ai} = \exp[v_i] \quad (3.48)$$

Çan eğrisi şekilli üyelik işlevlerinin orta noktalarındaki değişim aşağıdaki gibi

hesaplanır.

$$\Delta m_i = -\frac{\partial E}{\partial m_i} = -\left[\frac{\partial E}{\partial f} \frac{\partial f}{\partial \bar{\mu}_i} \frac{\partial \bar{\mu}_i}{\partial \mu_i} \frac{\partial \mu_i}{\partial \mu_{Ai}} \right] \left[\frac{\partial \mu_{Ai}}{\partial v_i} \frac{\partial v_i}{\partial m_i} \right] \quad i = 1, 2 \quad (3.49)$$

$$\frac{\partial \mu_{Ai}}{\partial v_i} = \exp[v_i] = \mu_{Ai} \quad (3.50)$$

$$\frac{\partial v_i}{\partial m_i} = 2 \left[\frac{x - m_i}{\sigma_i^2} \right] \quad (3.51)$$

$$\Delta m_i = 2 \left[\frac{x - m_i}{\sigma_i^2} \right] \delta_i^1 \mu_{Ai} \quad \delta_i^1 = \delta_i^2 \mu_{Bi} \quad (3.52)$$

$$\Delta m_i = 2 \left[\frac{x - m_i}{\sigma_i^2} \right] \delta_i^2 \mu_{Ai} \mu_{Bi} \quad \mu_i = \mu_{Ai} \mu_{Bi} \quad (3.53)$$

$$\Delta m_i = 2 \left[\frac{x - m_i}{\sigma_i^2} \right] \delta_i^2 \mu_i \quad (3.54)$$

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

$$\Delta m_{i+2} = 2 \left[\frac{y - m_{i+2}}{\sigma_{i+2}^2} \right] \delta_i^2 \mu_i \quad (3.55)$$

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

$$\Delta \sigma_i = -\frac{\partial E}{\partial \sigma_i} = -\left[\frac{\partial E}{\partial f} \frac{\partial f}{\partial \bar{\mu}_i} \frac{\partial \bar{\mu}_i}{\partial \mu_i} \frac{\partial \mu_i}{\partial \mu_{Ai}} \right] \left[\frac{\partial \mu_{Ai}}{\partial v_i} \frac{\partial v_i}{\partial \sigma_i} \right] \quad i = 1, 2 \quad (3.56)$$

$$\frac{\partial v_i}{\partial \sigma_i} = 2 \left[\frac{(x - m_i)^2}{\sigma_i^3} \right] \quad (3.57)$$

$$\Delta\sigma_i = 2 \left[\frac{(x - m_i)^2}{\sigma_i^3} \right] \delta_i^2 \mu_i \quad (3.58)$$

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

$$\Delta\sigma_{i+2} = 2 \left[\frac{(y - m_{i+2})^2}{\sigma_{i+2}^3} \right] \delta_i^2 \mu_i \quad (3.59)$$

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

$$m_i(k+1) = m_i(k) + \eta_m \Delta m_i(k+1) \quad i = 1, 2, 3, 4 \quad (3.60)$$

$$\sigma_i(k+1) = \sigma_i(k) + \eta_m \Delta \sigma_i(k+1) \quad i = 1, 2, 3, 4 \quad (3.61)$$

Burada toplam 8 adet değişken ayarlanır.

4. AKILLI YÖNTEMLER İLE KONUŞMACI TANIMA UYGULAMALARI

Bu bölümde yapılan çalışmalar detaylı bir şekilde anlatılmıştır. Veri tabanı olarak uluslararası YOHO veri tabanından alınan ses örnekleri kullanılmıştır. YOHO veritabanında kaydolma ve saptama aşamaları için ayrı kümeler vardır ve her bir kümede 69 kişiye ait çok sayıda cümleler yer almaktadır. 20 konuşmacıdan alınan cümleler hem eğitim hem de test aşamasında kullanılmıştır. Belirlenen yöntemlerin tasarlanıp test edilmesi için de Matlab programı kullanılmıştır.

Yapılan çalışmalar iki aşamadan oluşmaktadır. Birinci aşama eğitim aşaması, ikinci aşama ise test aşamasıdır. Eğitim aşamasında, kaydolma (enrollment) ve saptama (identification) işlemleri gerçekleştirilecektir. Kaydolma aşamasında konuşmacıların tüm ses örneklerinin öznitelik özellikleri çıkarılarak bir veritabanında saklanacaktır. Kaydolma aşamasının amacı her konuşmacı için bir referans modeli oluşturmaktır. Daha sonra bu modeller saptama aşamasında kullanılacaktır.

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

Kaydolma aşamasında elde ettiğimiz MFKK öznitelik vektörleri hem YSA hem de USBA ile eğitilerek test edilmiş ve başarı sonuçları karşılaştırılmıştır. Daha sonra yeni bir yaklaşımla konuşma sinyallerine dalgacık dönüşümü uygulanıp entropi katsayıları elde edilmiştir. Bu katsayılar yine YSA ve USBA gibi akıllı metotlar ile eğitilip test edilmiş ve çok fazla başarı sağlamadığı görülmüştür.

4.1. MFKK Kullanarak YSA ve USBA Akıllı Yöntemler ile Konuşmacı Tanıma

4.1.1. MFKK ve YSA ile Konuşmacı Tanıma

Öznitelik vektörlerinin bulunması sürecinde YOHO veri tabanında bulunan 10 bayan, 10 erkek konuşmacının söylediği 5 cümle kullanılmıştır. Kullanılan cümleler Tablo 4.1’de gösterilmiştir.

Tablo 4.1. MFKK ile konuşmacı tanıma eğitim ve test cümleleri

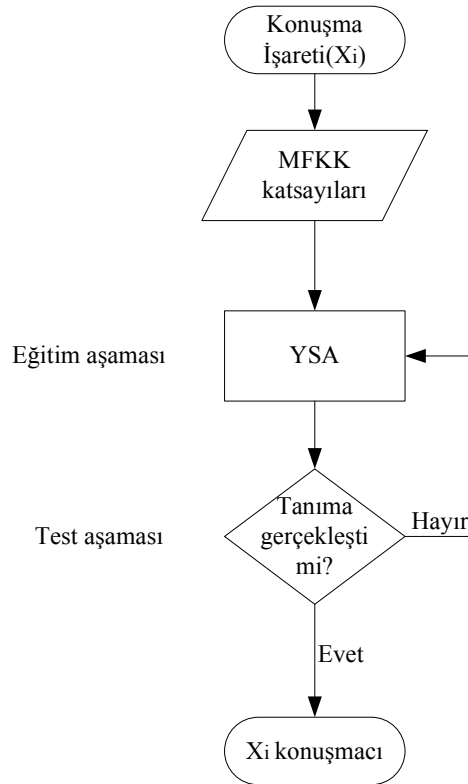
	1. cümle	2. cümle	3. cümle	4. cümle	5. cümle
Bayan1	23_71_84	27_56_74	32_97_74	43_95_31	45_61_89
Bayan2	25_54_63	47_64_34	56_35_79	61_86_95	69_47_21
Bayan3	25_43_81	29_32_91	32_96_81	41_29_52	41_63_96
Bayan4	21_87_76	24_59_76	37_91_75	37_92_63	42_64_63
Bayan5	21_43_65	25_96_57	26_67_72	37_84_96	41_34_93
Bayan6	23_31_67	45_93_37	54_43_42	67_81_29	72_96_35
Bayan7	34_43_29	43_67_71	45_56_84	67_86_34	73_84_62
Bayan8	21_35_96	23_72_82	47_54_69	65_29_52	71_86_57
Bayan9	23_95_54	24_39_76	35_57_64	71_86_57	45_61_57
Bayan10	27_94_61	42_35_89	53_76_21	57_36_91	73_26_52
Erkek1	24_83_96	26_34_92	47_39_49	49_75_57	61_75_52
Erkek2	27_31_95	35_97_64	51_25_52	61_82_56	73_42_53
Erkek3	24_73_61	26_41_73	45_95_46	52_62_37	59_36_87
Erkek4	54_85_32	57_25_84	62_84_29	59_36_87	84_31_46
Erkek5	27_73_36	29_93_62	35_52_54	37_21_53	46_71_84
Erkek6	29_63_85	51_97_74	61_57_92	65_39_53	74_32_46
Erkek7	25_83_74	42_29_93	43_21_85	53_84_41	54_76_39
Erkek8	25_76_83	29_31_53	35_96_84	43_59_67	56_41_75
Erkek9	47_89_96	49_32_69	56_39_71	63_71_54	87_45_64
Erkek10	27_96_75	47_84_59	56_95_71	62_32_83	69_54_32

Bunun için önce konuşmacılara ait konuşma sinyalleri 30 ms'lik çerçevelere ayrılmıştır. Pencereleme işleminde Hamming pencere kullanılmıştır. Pencerelemiş konuşma sinyal çerçevelerinin HFD'si alınarak konuşma sinyali frekans bölgesine taşınmıştır. Bir sonraki adımda HFD spektrumundan mel spektrumuna geçilmiştir. Bunun için HFD spektrumunun genlik değerlerinin mutlak kare değerleri hesaplanarak güç spektrumu oluşturulmuştur.

Mutlak genlik değerlerinin kareleri hesaplanan HFD noktaları, band geçiren üçgen dalgalardan oluşan mel-filtre bankasından geçirilmiştir. Kullanılan mel filtre bankası, 13

adet doğrusal olarak yerleştirilen filtreyi takip eden 27 adet logaritmik filtreden meydana gelmiştir.

Mel spektrumun elde edilmesinden sonra mel kepstrum katsayıları elde edilmiştir. Bunun için mel filtre bankası çıkışının logaritması alınarak Ayrık Kosinüs Dönüşümü ile zaman bölgesine geçiş yapılır. Oluşturulan bu katsayılar mel-ölçeklendirilmiş HFD temelli kepstrum katsayılarıdır. AKD kullanmanın sebeplerinden biri THFD (Ters Hızlı Fourier Dönüşümü) uygulamasının AKD'ye göre kompleks aritmetik işlemler gerektirmesidir. Bütün bu aşamalardan sonra YSA için 13 adet USBA için de 4 adet MFKK katsayısı elde edilmiştir. Şekil 4.1'de MFKK tabanlı YSA akıllı metoduyla konuşmacı saptama aşamaları adım adım gösterilmiştir.



Şekil 4.1. YSA ile konuşmacı tanıma algoritması

10 erkek ve 10 kadın konuşmacının öznitelik vektörleri kullanılarak eğitim dizi oluşturulmuştur. İlk aşama eğitim safhasıdır. Çalışmamız metine bağımlı olduğu için özellik çıkarım aşamasında elde edilen 20 x 5 x 13 adet özellik vektörü yapay sinir ağları

ile akıllı bir sınıflandırma yapılarak eğitilmiştir. Kullanılan çok katmanlı algılayıcı yapılı yapay sinir ağının eğitim parametreleri ve yapısı Tablo 4.2’de gösterilmektedir.

Tablo 4.2. Çok katmanlı yapay sinir ağının yapısı ve eğitim parametreleri

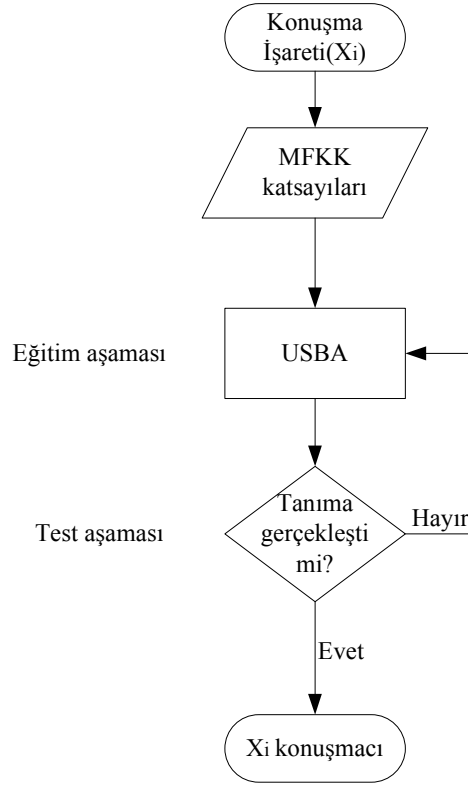
Katman sayısı	3
Katman hücre sayıları	Giriş: 13, Gizli Katman: 30, Çıkış: 1
Aktivasyon Fonksiyonları	Giriş: sigmoid, ara: sigmoid, çıkış:lineer
Öğrenme algoritması	Geriye yayılım algoritması
Toplam karesel hata	0.001

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

Eğitim aşamasından sonra test aşamasına geçilmiştir. Sistem gürültü eklenmeden test edildiğinde YSA için doğru konuşmacı tanıma oranı %100 olmuştur. Konuşmacıların eğitimde kullandıkları cümlelere enerji ortalaması sıfır olan -1, -5, -10, 1, 5, 10 dB’lik beyaz gürültü eklenerek eğitilen sistem test edildiğinde herhangi bir başarımlı elde edilememiştir.

4.1.2. MFKK ve USBA ile Konuşmacı Tanıma

Burada, özellik çıkarım safhasında elde edilen $20 \times 5 \times 4$ özellik vektörü kullanılarak uyarlamalı sinirsel bulanık ağ ile akıllı bir tanıma yapılmıştır. Öncelikle çerçevelenip pencerelenen sinyalin hızlı fourier dönüşümü (HFD) alınarak mel ölçekli filtreden geçirilip mel frekansı kepsrum katsayıları elde edilmiştir. Şekil 4.2’de MFKK tabanlı USBA akıllı metoduyla konuşmacı saptama aşamaları adım adım gösterilmiştir.



Şekil 4.2. USB A ile konuşmacı tanıma algoritması

Kullanılan çok katmanlı algılayıcı yapıli uyarlamalı sinirsel bulanık ağ sisteminin eğitim parametreleri ve yapısı Tablo 4.3’de gösterilmektedir.

Tablo 4.3. USB A modelinin yapısı ve eğitim parametreleri

Katman sayısı	5
Katman hücre sayıları	Giriş: 4, Çıkış: 1
Aktivasyon fonksiyonları	Gbellmf
Eğitim parametreleri Öğrenme kuralı	Melez Öğrenme Algoritması (Doğrusal olmayan giriş parametreleri için Geri yayınım ve doğrusal çıkış parametreleri için En küçük kareler yöntemlerinin birleşimleri)

MFKK katsayıları eğitildikten sonra test aşamasına geçilmiştir. Gürültü eklenmeden test edildiğinde USBA için doğru tanıma oranı %100 olmuştur. Konuşmacıların eğitimde kullandıkları cümlelere enerji ortalaması sıfır olan -1, -5, -10, 1, 5, 10 dB’lik beyaz gürültü eklenerek eğitilen sistem test edildiğinde iyi bir başarı elde edilememiştir.

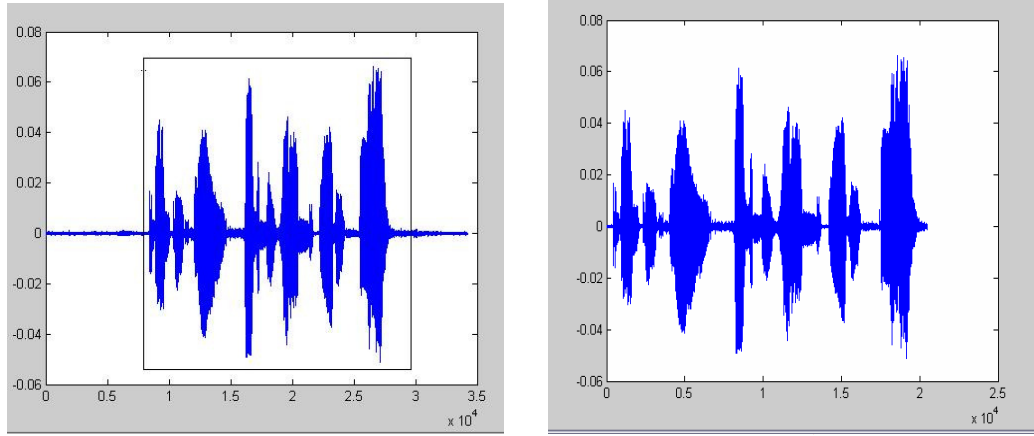
4.2. DD Kullanarak YSA ve USBA Akıllı Yöntemler ile Konuşmacı Tanıma

Öznitelik vektörlerinin bulunması sürecinde YOHO veri tabanından 10 bayan, 10 erkek konuşmacının söylediği 3 cümle kullanılmıştır. Kullanılan cümleler Tablo 4.4’de verilmiştir.

Tablo 4.4. DD ile konuşmacı tanıma eğitim ve test cümleleri

	1. cümle	2. cümle	3. cümle
Bayan1	23_71_84	27_56_74	32_97_74
Bayan2	25_54_63	47_64_34	56_35_79
Bayan3	25_43_81	29_32_91	32_96_81
Bayan4	21_87_76	24_59_76	37_91_75
Bayan5	21_43_65	25_96_57	26_67_72
Bayan6	23_31_67	45_93_37	54_43_42
Bayan7	34_43_29	43_67_71	45_56_84
Bayan8	21_35_96	23_72_82	47_54_69
Bayan9	23_95_54	24_39_76	35_57_64
Bayan10	27_94_61	42_35_89	53_76_21
Erkek1	24_83_96	26_34_92	47_39_49
Erkek2	27_31_95	35_97_64	51_25_52
Erkek3	24_73_61	26_41_73	45_95_46
Erkek4	54_85_32	57_25_84	62_84_29
Erkek5	27_73_36	29_93_62	35_52_54
Erkek6	29_63_85	51_97_74	61_57_92
Erkek7	25_83_74	42_29_93	43_21_85
Erkek8	25_76_83	29_31_53	35_96_84
Erkek9	47_89_96	49_32_69	56_39_71
Erkek10	27_96_75	47_84_59	56_95_71

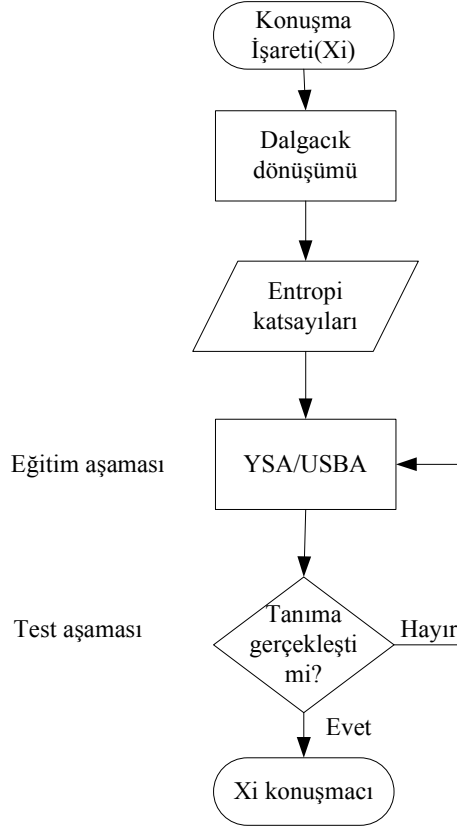
DD entropi katsayılarını elde etmek için öncelikle konuşmacılara ait konuşma sinyalleri çerçevelenmiştir. Çünkü işaret sinyali oldukça fazla bilgi içermektedir. Dolayısıyla işlemlerin daha kolay yapılabilmesi için ses işareti belirli sayıda örnek içeren çerçevelere ayrılmıştır. Ardından dikdörtgen pencere kullanılarak pencerelenmiştir. Bu sayede sinyal bu pencere ile çarpıldığında aralık dışındaki değerleri sıfır olmuştur. Şekil 4.3'te olduğu gibi;



Şekil 4.3. Dikdörtgen pencere ile pencerelenmiş sinyal

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

Yapılan çalışmada, “sym1” dalgacık türü kullanılarak entropi katsayıları hesaplanmıştır. Entropi tipi olarak ta Shannon kullanılmış ve sonuçta 4 tane entropi katsayısı elde edilmiştir. Şekil 4.4'te DD tabanlı YSA ve USBA akıllı metoduyla konuşmacı tanıma adım adım gösterilmiştir.



Şekil 4.4. DD tabanlı YSA veya USB ile konuşmacı tanıma algoritması

Özellik çıkarım safhasında elde edilen $20 \times 3 \times 4$ özellik vektörü YSA ile eğitilmiştir. Test aşamasında ise 1, 5, 10, -1, -5, -10 dB değerinde farklı gürültü değerleri eklenerek test edilmiş fakat bir başarı elde edilememiştir.

Aynı katsayılar USB ile akıllı bir sınıflandırma yapılarak eğitilmiştir. Eğitim aşamasında elde ettiğimiz katsayılar öncelikle üyelik fonksiyon sayısı 2 ve 3 olan farklı fonksiyonlar ile tek tek eğitilmiştir. Test aşamasında ise 1, 5, 10, -1, -5, -10 dB değerinde farklı gürültü değerleri kullanılarak test edilmiş fakat herhangi bir başarı sağlanamamıştır. Gürültü eklenmeden sistem çalıştırıldığında ise farklı başarı değerleri elde edilmiştir. Kullanılan uyarlamalı sinirsel bulanık ağ sisteminin eğitim parametreleri ve başarı yüzdesi Tablo 4.5.'te gösterilmektedir.

Tablo 4.5. Uyarlamalı sinirsel bulanık ağı sisteminin parametreleri ve başarı yüzdesi

Fonksiyon / Üyelik Fonksiyon Sayısı	Saptanan Kişi Sayısı	Saptanmayan Kişi Sayısı	Saptama Başarı Yüzdesi
gbellmf mf=2	30	30	% 50
gbellmf mf=3	45	15	% 75
gaussmf mf=2	30	30	% 50
gaussmf mf=3	40	20	% 66.66
trimf mf=2	11	49	% 18.33
trimf mf=3	24	36	% 40
pimf mf=2	24	36	% 40
pimf mf=3	34	26	% 56.66
trapmf mf=2	26	34	% 43.33
trapmf mf=3	38	22	% 63.33
psigmf mf=2	27	33	% 45
psigmf mf=3	30	30	% 50

5. SONUÇLAR VE ÖNERİLER

Bu tezde konuşmacı tanıma sistemlerinin genel yapısı incelenmiş, metin bağımlı ve metin bağımsız sistemler için kullanılmakta olan yöntemler özelliklerine göre sınıflandırılmış ve her biri ayrı ayrı incelenmiştir.

Gerek daha kolay gerçekleştirilebileceği gerekse de daha fazla başarımlı sağlayacağı düşüncesiyle metin bağımlı bir sistem tasarımı tercih edilmiştir. Metin bağımlı konuşmacı saptama uygulamasının gerçekleşmesi sırasında her bir işlem basamağı incelenmiştir.

Öznitelik vektörlerinin belirlenmesinde ise MFKK yöntemi ile birlikte, karşılaştırma yapmak amacıyla DD uygulanmıştır. DD kullanmadaki amaç sinyalin önileme aşamasındaki gürültüleri temizlemektir. Daha sonra DD ile elde edilen veri dizilerinin entropi değerleri hesaplanmıştır. Modelleme aşamasında ise YSA ve USBA gibi akıllı metotlar kullanılarak karşılaştırma yapılmıştır.

Akıllı modelleme çalışmalarında yapılan ölçümler ve hesaplamalar sonucunda elde edilen veriler herhangi bir ön işleme tabi tutulmadan kullanıldığı için sistemin akıllı modelini oluştururken çok sayıda örnek sayısına ihtiyaç duyulmaktadır. Çok sayıda örnek kullanıldığında ise hesaplama süresinin ve hafıza gereksinimin artması ve eğitilen sistemin sistem davranışını ezberlemesi gibi bazı sorunlar ortaya çıkmıştır.

MFKK katsayıları kullanılarak yapılan YSA ve USBA tabanlı modelleme çalışmalarında, gürültünün tanımayı oldukça etkilediği görülmüştür. Sonuçta istenilen test başarımlarını gösterememiştir.

DD ile elde edilen entropi katsayıları kullanılarak yapılan YSA ve USBA tabanlı modelleme çalışmalarında ise gürültünün YSA ve USBA akıllı metotları kullanıldığında tanımayı gerçekleştirilemediği görülmüştür. Gürültü eklenmeden sistem test edildiğinde ise MFKK katsayıları ile yapılan tanıma başarımlarının DD entropi ile yapılan tanıma başarımlarına göre çok daha yüksek olduğu gözlemlenmiştir. Konuşmacı saptama başarısını etkileyen faktörlerin entropi türü, dalgacık tipi ve dalgacığın seviyesi gibi parametreler olduğu tespit edilmiştir. Aynı zamanda konuşmacı sayısı ile konuşulan kelime sayısının artışı tanıma başarısını etkileyerek sistem performansını düşürmüştür. Uluslararası konuşmacı veritabanı olan YOHO sisteminden 20 kişilik konuşma kümesi ile yapılan testler göstermiştir ki önerilen dalgacık dönüşümü gerek tanıma gerekse güvenilirlik açısından önceki algoritmalara göre daha iyi sonuçlar üretememiştir.

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

KAYNAKLAR

- [1] **Furui, S.**, 1996. An Overview of Speaker Recognition Technology, Automatic Speech and Speaker Recognition, Kluwer Academic, Boston, pp. 31-56.
- [2] **Hanilçi, C.**, 2007. Konuşmacı Tanıma Yöntemlerinin Karşılaştırmalı Analizi, *Yüksek Lisans Tezi*, Uludağ Üniversitesi, Bursa.
- [3] **Wang, L., et al.**, 2002. Capture Interspeaker Information with a Neural Network for Speaker Identification, IEEE Neural Networks, vol. 13, no. 2, pp. 436-445.
- [4] **Erzin, E. ve ark.**, 2005. Multimodal Speaker Identification Using an Adaptive Classifier Cascade Based on Modality Reliability, IEEE Multimedia, vol. 7, no. 5, pp. 840-852.
- [5] **Mut, O.**, 2005. Konuşmacı Tanıma, Yüksek Lisans Tezi, Gebze İleri Teknoloji Enstitüsü Mühendislik ve Fen bilimleri Enstitüsü, Kocaeli.
- [6] **Uzunçarşılı, M.**, 2005. Vektör Nicemleme Tekniklerine Dayalı Konuşmacı Tanıma Algoritmalarının İncelenmesi, *Yüksek Lisans Tezi*, Ankara Üniversitesi Fen Bilimleri Enstitüsü, Ankara.
- [7] **Başaran, S.**, 2007. Yapay Sinir Ağları Kullanarak Konuşmacı Tanıma, *Yüksek Lisans Tezi*, Uludağ Üniversitesi Fen Bilimleri Enstitüsü, Bursa.
- [8] **Aygün, O.**, 2006. Konuşmacı Tanıma Sistemlerinde Dalgacık Dönüşümü, *Yüksek Lisans Tezi*, Selçuk Üniversitesi Fen Bilimleri Enstitüsü, Konya.
- [9] **Davis, S., Mermelstein, P.**, 1980. Comparison of Parametric Representations Formonosyllabic Word Recognition in Continuously Spoken Sentences, IEEE Acoustics, Speech, and Signal Processing, vol. 28, no. 4, pp. 357-366.
- [10] **Ningping, F., Justinian, P. R.**, 2003. Enhanced VQ-Based Algorithms for Speech Independent Speaker Identification, Audio and Video based Biometric Person Authentication, pp. 470-477.
- [11] **Gish, H. and Schmit, M.**, 1994. Text-Independent Speaker Identification, IEEE Signal Processing Magazine, vol. 11, pp. 18-32.
- [12] **Kinnunen, T., et al.**, 2005. Real-Time Speaker Identification and Verification, IEEE Transactions on Speech and Audio Processing no. 99, pp. 1-12.
- [13] **Molau, S., et al.**, 2001. Computing Mel-frequency cepstral coefficients on the power spectrum, IEEE Int. Conf. Acoustics, Speech, and Signal Processing, vol. 1, pp. 73-76.

- [14] **Atal, B. S.**, 1976. Automatic Recognition of Speakers from Their Voices, IEEE, vol. 64, no. 4, pp. 460-475.
- [15] **Thomas, F.**, 2002. Quatieri, Discrete-Time Speech Signal Processing: Principles and Practice, Prentice Hall, MIT Lincoln Laboratory.
- [16] **Türkoğlu, İ.**, 2002. Durağan Olmayan İşaretler İçin Zaman-Frekans Entropilerine Dayalı Akıllı Örüntü Tanıma, *Doktora Tezi*, Fırat Üniversitesi Fen Bilimleri Enstitüsü, Elazığ, s. 112.
- [17] **Keeton, P. I. J. and Schlindwein, F. S.**, 1997. Application of Wavelets in Doppler Ultrasound, Sensor Review, MCB University Press, vol. 17, no. 1, pp. 38-45.
- [18] **Hoang, M., Aaron, R. ve Shiffman, C. A.**, 1997. Maximum Entropy Method for Magnetoencephalography, IEEE Trans. on Biomedical Engineering, vol. 44, no. 1, pp. 98-102.
- [19] **Akay, M.**, 1994. Investigating the Effects of Vasodilator Drugs on the Turbulent Sound Caused by Femoral Artery Stenosis Using Short-Term Fourier and Wavelet Transform Methods, IEEE Trans. on Biomedical Engineering, vol. 41, pp. 921-928.
- [20] **Strang, G. and Nguyen, T.**, 1996. Wavelets and Filter Banks, Wellesley-Cambridge Press, Cambridge, MA, p. 390.
- [21] **Kil, D. H. and Shin, F. B.**, 1996. Pattern Recognition and Prediction with Applications to Signal Characterization. AIP Press, USA, p. 417.
- [22] **Shannon, C. E.**, 1948. A Mathematical Theory of Communication, Bell System Technology Journal, vol. 27, pp. 379-423.
- [23] **Tonga, S., Bezerianos, A., Paula, J., Zhub, Y. and Thakora, N.**, 2002. Nonextensive Entropy Measure of EEG Following Brain in Jury from Cardiac Arrest, Elsevier Physica A, vol. 305, pp. 619- 628.
- [24] **Principe, J. C., Euliano, N. R. and Lefebvre, W. C.**, 2000. Neural and Adaptive Systems. John Wiley & Sons, New York, p. 656.
- [25] **Overwijk, M. H. F. and Reefman, D.**, 2000. Maximum-Entropy Deconvolution Applied to Electron Energy-Loss Spectroscopy, Pergamon Micron, vol. 31, pp. 325-331.
- [26] **Li, X.**, 2000. Edge Directed Statistical Inference with Applications to Image Processing, *Doktora Tezi*, Princeton Üniversitesi, p. 131.
- [27] **Zhang, X.S. and Roy, R.J.**, 2001. Derived Fuzzy Knowledge Model for Estimating the Depth of Anesthesia, IEEE Trans. on Biomedical Engineering, vol. 48, no. 3, pp. 312-323.

- [28] **Coifman, R. R. And Wickerhauser, M. V.,** 1992. Entropy-Based Algorithms for Best Basis Selection, IEEE Trans. on Information Theory, vol. 38, no. 2, pp. 713-718.
- [29] **Rabiner, L. R.,** 1989. A Tutorial on Hidden Markov Models and Selected Applications in Speech Recognition. Proc IEEE, 77, 257-286
- [30] **Campbell, J. P.,** 1997. Speaker Recognition: A Tutorial. Proc. IEEE 85 (9), 1437-1462.
- [31] **Özbay, Y.,** 1999. EKG Aritmilerini Hızlı Tanıma, *Doktora Tezi*, Selçuk Üniversitesi Fen Bilimleri Enstitüsü, Konya.
- [32] **Demuth, H., Beale, M., and Hagan, M.,** 2006. Neural Network Toolbox, for use with MATLAB, User Guide Version 5, Mathworks.
- [33] **Türkoğlu, İ.,** 1996. Yapay Sinir Ağları ile Nesne Tanıma, *Yüksek Lisans Tezi*, Fırat Üniversitesi Fen Bilimleri Enstitüsü, Elazığ.
- [34] **Avcı, E., Akpolat, Z. H.,** 2006. Speech Recognition Using A Wavelet Packet Adaptive Network Based Fuzzy Inference System, Expert Systems with Applications, 31(3), 495-503.
- [35] **Civalek, Ö.,** 1998. Plak ve Kabukların Nöro-Fuzzy Tekniği ile Lineer ve Non-Lineer Statik-Dinamik Analizi, *Yüksek Lisans Tezi*, Fırat Üniversitesi, Elazığ.
- [36] **Kelesoglu, O., ve Fırat, A.,** 2006. Tuğla Duvardaki ve Tesisattaki Isı Kaybının Yapay Sinir Ağları ile Belirlenmesi, Fırat Üniversitesi Fen ve Mühendislik Bilimleri Dergisi, 18(1), 133-141.
- [37] **Elmas, Ç.,** 2007. Yapay Zeka Uygulamaları, Seçkin Yayınları, Ankara, s. 88.
- [38] **Feixue, W., Wenxian, Y., Guirong, G.,** 1996. Recurrent Neural Network for High-Resolution Radar Ship Target Recognition, Radar, Proceedings., CIE International Conference of , pp.200 – 203.
- [39] **Jang, S.R.,** 1993. ANFIS: Adaptive Network-Based Fuzzy Inference Systems, IEEE, Trans. on Man. and Cybernetics, vol. 23, pp. 665-685.
- [40] **Hanbay, D.,** 2007. Yapay Sinir Ağı Tabanlı Akıllı Yöntemlerle Karmaşık Sistemlerin Modellenmesi, *Doktora Tezi*, Fırat Üniversitesi Fen Bilimleri Enstitüsü, Elazığ.
- [41] **Elmas, Ç.,** 2003. Bulanık Mantık Denetleyiciler, Seçkin Yayınları, Ankara, s. 230.

ÖZGEÇMİŞ

Eda RENÇBER

Fırat Üniversitesi
Teknik Eğitim Fakültesi
Elektronik ve Bilgisayar Eğitimi Bölümü
Elazığ

Tel: 424 – 2389079
E.posta: edalim_23@hotmail.com

- 1982** Elazığ ‘da doğdu.
- 1996 – 2000** Elazığ Balakgazi Yabancı Dil Ağırlıklı Lisesini tamamladı.
- 2001 – 2005** Fırat Üniversitesi Teknik Eğitim Fakültesi Elektronik – Bilgisayar Eğitimi Bilgisayar Öğretmenliği Bölümünden mezun oldu.
- 2005 –** T.C. M.E.B. da öğretmen olarak çalışmaktadır.
- 2009 –** Fırat Üniversitesi, Fen Bilimleri Enstitüsü, Elektronik – Bilgisayar Eğitimi Anabilim Dalında yüksek lisans tez çalışmasına başladı.