

**YILDIZ TEKNİK ÜNİVERSİTESİ
FEN BİLİMLERİ ENSTİTÜSÜ**

**ENSTRÜMAN SESLERİNİN TANINMASI VE
SINIFLANDIRILMASI**

Elektronik ve Haberleşme Yüksek Mühendisi Bülent BOLAT

**FBE Elektronik ve Haberleşme Anabilim Dalı Haberleşme Programında
Hazırlanan**

DOKTORA TEZİ

Tez Savunma Tarihi	: 20.02.2006
Tez Danışmanı	: Yrd. Doç. Dr. Ünal KÜÇÜK (YTÜ)
Jüri Üyeleri	: Prof. Metin YÜCEL (YTÜ) Doç. Dr. Tülay YILDIRIM (YTÜ) Prof. Dr. Ahmet Hamdi KAYRAN (İTÜ) Prof. Dr. Ümit AYGÖLÜ (İTÜ)

İSTANBUL, 2006

İÇİNDEKİLER

	Sayfa
SİMGE LİSTESİ	iv
KISALTMA LİSTESİ	v
ŞEKİL LİSTESİ	vi
ÇİZELGE LİSTESİ	vii
ÖNSÖZ	ix
ÖZET	x
ABSTRACT	xi
1. GİRİŞ	1
1.1 Tezin Amaç ve Önemi	2
1.2 Kullanılan Veri Kümesi	3
1.3 Tezin Düzeni	4
2. GEÇMİŞ ÇALIŞMALAR	5
2.1 Ses Bölütleme ve Sınıflandırma	7
2.2 İçeriğe Bağlı Ses Erişimi	8
2.3 Video İndeksleme İçin Ses Analizi	8
2.4 Konuşma ve Konuşmacı Tanıma	9
2.5 Müzikal Enstrüman Tanıma	9
2.5.1 İnsanlar Üzerinde Yapılan Deneyler	9
2.5.2 Otomatik Sistemlerin Başarımı	11
3. SES TANIMADA KULLANILAN ÖZNİTELİKLER	13
3.1 Fiziksel Öznitelikler	13
3.1.1 Enerji	13
3.1.2 Sıfır Geçiş Sayısı	14
3.1.3 Atak Süresi	14
3.1.4 Genlik Modülasyonu	14
3.1.5 Vibrato	14
3.2 Spektral Öznitelikler	14
3.2.1 Spektral Ağırlık Merkezi	15
3.2.2 Harmonik Bozulma	15
3.2.3 Spektral Düzensizlik	15
3.2.4 Spektral Akı	16
3.3 Doğrusal Öngörü Katsayıları	16
3.4 Mel Frekans Cepstral Katsayılar	18
4. OLASILIKSAL SINIR AĞI ve AKTİF ÖĞRENME	20

4.1	Olasılıksal Sinir Ağları	21
4.2	Aktif Öğrenme	23
4.3	PNN İle Aktif Öğrenme.....	25
5.	UYGULAMA	28
5.1	Özniteliklerin Hesaplanması.....	29
5.2	Pasif Öğrenen PNN İle Enstrüman Tanıma.....	30
5.3	Aktif Öğrenen PNN İle Enstrüman Tanıma	38
6.	SONUÇLAR.....	44
KAYNAKLAR.....		49
ÖZGEÇMİŞ.....		54

SİMGE LİSTESİ

HD	Harmonik Bozulma
IRR	Spektral Düzensizlik
SF	Spektral Akı
$R(n)$	Öznitelik Fonksiyonu
$c(n)$	Karmaşık Cepstrum
c_{mel}	Mel Cepstrum
P	Karşılaşma Sıklığı
L	Maliyet Fonksiyonu
$F(.)$	Olasılık Yoğunluk Fonksiyonu
σ	Bant Genişliği

KISALTMA LİSTESİ

CASA	Computational Auditory Scene Analysis
GMM	Gaussian Mixture Models
HMM	Hidden Markov Models
LP	Lienar Preediction
LPC	Linear Prediction Coefficients
MFCC	Mel Frequency Cepstral Coefficients
MLP	Multi Layer Perceptron
PNN	Probabilistic Neural Networks
QBC	Query By Comittee
RBF	Radial Basis Function Networks
RUB	Repeat Until Bored
SOM	Self Organizing Maps
SVM	Support Vector Machines
YSA	Yapay Sinir Ağları
LVQ	Learning Vector Quantisation
DCT	Discrete Cosine Transform
k-NN	k- Nearest Neighborhood Network
RMS	Root Mean Square
RMS E	Root Mean Square Energy
PCA	Principal Component Analysis
LDA	Linear Discriminant Analysis
ZCR	Zero Cross Rate
SC	Spectral Centroid
DFT	Discrete Fourier Transform
FFT	Fast Fourier Transform
F ₀	Ana Harmonik Frekansı

ŞEKİL LİSTESİ

	Sayfa
Şekil 1.1 İnsan ve makine yeteneklerinin karşılaştırılması (Martin, 1999).....	2
Şekil 2.1 Ses tanıma işlemi.....	7
Şekil 3.1. MFCC'nin FFT yardımıyla hesaplanması.....	19
Şekil 4.1. İki sınıflı bir PNN.....	22
Şekil 4.2. Pasif Öğrenme	23
Şekil 4.3. Aktif öğrenme. Eğitim kümesi öğrenen tarafından seçilmiştir.....	23
Şekil 4.4. PNN ile aktif öğrenme akış şeması	27
Şekil 5.1. Hiyerarşik sınıflandırma yapısı	29
Şekil 5.2 Uygulanan enstrüman tanıma sisteminin blok şeması	29

ÇİZELGE LİSTESİ

	Sayfa
Çizelge 2.1 İnsan deneklerle yapılan tanıma deneylerinin sonuçları	10
Çizelge 5.1. 20. dereceden LP katsayıları ile aile içi doğru sınıflandırma yüzdeleri	31
Çizelge 5.2. 15. dereceden LP katsayıları ile aile içi doğru sınıflandırma yüzdeleri	31
Çizelge 5.3. 10. dereceden LP katsayıları ile aile içi doğru sınıflandırma yüzdeleri	31
Çizelge 5.4. 5. dereceden LP katsayıları ile aile içi doğru sınıflandırma yüzdeleri	31
Çizelge 5.5. 15. derece LP katsayıları kullanıldığında Yaylılar ailesi için aile içi sınıflandırma test kümesi sonuçları	31
Çizelge 5.6. 20. derece LP katsayıları kullanıldığında Pirinç Nefesliler ailesi için aile içi sınıflandırma test kümesi sonuçları	32
Çizelge 5.7. 15. derece LP katsayıları kullanıldığında Pizzicato ailesi için aile içi sınıflandırma test kümesi sonuçları	32
Çizelge 5.8. 10. derece LP katsayıları kullanıldığında Klarnet ailesi için aile içi sınıflandırma test kümesi sonuçları	32
Çizelge 5.9. 10. derece LP katsayıları kullanıldığında Saksofon ailesi için aile içi sınıflandırma test kümesi sonuçları	33
Çizelge 5.10. 10. derece LP katsayıları kullanıldığında Çift Kamışlılar ailesi için aile içi sınıflandırma test kümesi sonuçları	33
Çizelge 5.11. 10. derece LP katsayıları kullanıldığında Flüt ailesi için aile içi sınıflandırma test kümesi sonuçları	33
Çizelge 5.12. LP derecesinin enstrüman ailelerinin tanınmasına etkisi	34
Çizelge 5.13. 4. dereceden MFCC katsayıları ile aile içi doğru sınıflandırma yüzdeleri	34
Çizelge 5.14. 6. dereceden MFCC katsayıları ile aile içi doğru sınıflandırma yüzdeleri	34
Çizelge 5.15. 8. dereceden MFCC katsayıları ile aile içi doğru sınıflandırma yüzdeleri	34
Çizelge 5.16. 10. dereceden MFCC katsayıları ile aile içi doğru sınıflandırma yüzdeleri	34
Çizelge 5.17. 12. dereceden MFCC katsayıları ile aile içi doğru sınıflandırma yüzdeleri	34
Çizelge 5.18. 14. dereceden MFCC katsayıları ile aile içi doğru sınıflandırma yüzdeleri	35
Çizelge 5.19. 16. dereceden MFCC katsayıları ile aile içi doğru sınıflandırma yüzdeleri	35
Çizelge 5.20. 8. derece MFCC katsayıları kullanıldığında Yaylılar ailesi için aile içi sınıflandırma test kümesi sonuçları	35
Çizelge 5.21. 6. derece MFCC katsayıları kullanıldığında Pizzicato ailesi için aile içi sınıflandırma test kümesi sonuçları	35
Çizelge 5.22. 6. derece MFCC katsayıları kullanıldığında Pirinç Nefesliler ailesi için aile içi sınıflandırma test kümesi sonuçları	36
Çizelge 5.23. 6. derece MFCC katsayıları kullanıldığında Klarnet ailesi için aile içi sınıflandırma test kümesi sonuçları	36
Çizelge 5.24. 6. derece MFCC katsayıları kullanıldığında Çift Kamışlılar ailesi için aile içi sınıflandırma test kümesi sonuçları	36
Çizelge 5.25. 10. derece MFCC katsayıları kullanıldığında Saksofon ailesi için aile içi sınıflandırma test kümesi sonuçları	37
Çizelge 5.26. 14. derece MFCC katsayıları kullanıldığında Flüt ailesi için aile içi sınıflandırma test kümesi sonuçları	37
Çizelge 5.27. Farklı derecelerden MFCC katsayıları ile aile tanıma yüzdeleri	37
Çizelge 5.28. Farklı derecelerden LPC'ler için hiyerarşik tanıma yüzdeleri	38
Çizelge 5.29. Farklı derecelerden MFCC'ler için hiyerarşik tanıma yüzdeleri	38
Çizelge 5.30. Aktif Öğrenen PNN kullanıldığında 10. dereceden LP katsayıları ile aile içi doğru sınıflandırma yüzdeleri	38
Çizelge 5.31. Aktif Öğrenen PNN ile 10. derece LP katsayıları kullanıldığında Yaylılar ailesi için aile içi sınıflandırma test kümesi sonuçları	39

Çizelge 5.32. Aktif Öğrenen PNN ile 10. derece LP katsayıları kullanıldığında Pizzicato ailesi için aile içi sınıflandırma test kümesi sonuçları.....	39
Çizelge 5.33. Aktif Öğrenen PNN ile 10. derece LP katsayıları kullanıldığında Pirinç Nefesliler ailesi için aile içi sınıflandırma test kümesi sonuçları	39
Çizelge 5.34. Aktif Öğrenen PNN ile 10. derece LP katsayıları kullanıldığında Klarnet ailesi için aile içi sınıflandırma test kümesi sonuçları.....	40
Çizelge 5.35. Aktif Öğrenen PNN ile 10. derece LP katsayıları kullanıldığında Flüt ailesi için aile içi sınıflandırma test kümesi sonuçları.....	40
Çizelge 5.36. Aktif Öğrenen PNN ile 10. derece LP katsayıları kullanıldığında Saksofon ailesi için aile içi sınıflandırma test kümesi sonuçları.....	40
Çizelge 5.37. 6. dereceden MFCC katsayıları kullanıldığında aktif öğrenen PNN ile aile içi doğru sınıflandırma yüzdeleri	41
Çizelge 5.38. Aktif Öğrenen PNN ile 6. derece MFCC katsayıları kullanıldığında Pirinç Nefesliler ailesi için aile içi sınıflandırma test kümesi sonuçları	41
Çizelge 5.39. Aktif Öğrenen PNN ile 6. derece MFCC katsayıları kullanıldığında Saksofon ailesi için aile içi sınıflandırma test kümesi sonuçları	41
Çizelge 5.40. Aktif Öğrenen PNN ile 6. derece MFCC katsayıları kullanıldığında Pizzicato ailesi için aile içi sınıflandırma test kümesi sonuçları	42
Çizelge 5.41. Aktif Öğrenen PNN ile 6. derece MFCC katsayıları kullanıldığında Klarnet ailesi için aile içi sınıflandırma test kümesi sonuçları	42
Çizelge 5.42. Aktif Öğrenen PNN ile 6. derece MFCC katsayıları kullanıldığında Flüt ailesi için aile içi sınıflandırma test kümesi sonuçları.....	42
Çizelge 5.43. Aktif Öğrenen PNN ile 6. derece MFCC katsayıları kullanıldığında Çift Kamışlılar ailesi için aile içi sınıflandırma test kümesi sonuçları	43
Çizelge 5.44. Aktif Öğrenen PNN ile aile tanıma sonuçları	43
Çizelge 6.1. Aktif Öğrenen PNN ile enstrüman tanıma test sonuçları	46
Çizelge 6.1. (Devam) Aktif Öğrenen PNN ile enstrüman tanıma test sonuçları.....	47

ÖNSÖZ

Yüksek lisans tezimden sonra doktora tezimi de yönetmeyi kabul ettiği için ve çalışmam boyunca verdiği destek için Dr. Ünal Küçük'e teşekkür ederim. Tezimi bitirebilmeme olanak sağlayan, her aşamada bana destek olan ve tezde kullandığım aktif öğrenme algoritmasını birlikte tasarladığımız Dr. Tülay Yıldırım'a da teşekkür borçluyum. Dr. Cumhuriyet Erkut, MUMS CD'lerini temin etmesinin yanı sıra, (tıpkı yüksek lisans tezimde olduğu gibi) tez konumu belirlememde de etkin rol oynadı. Sevgili Cumhuriyet'a olan borcum ödenemez.

Teşekkür ederken, ailemi de unutmayacağım. Sevgili eşim Gonca ve tezime başladıktan sonra ailemize katılan, yaşantıma neşe ve umut katan oğlum Oğuz, bu çalışmada sizlerin de payı büyük. Hem mecazi, hem de gerçek anlamda ilk öğretmenim olan annem ve orta okul fen bilgisi öğretmenim olan babam, belki de fark etmeden bilimi sevmemde baş rolü oynamışlardır. Araştırmanın önemini onlardan öğrendim. Onlara layık bir evlat olabilmek için gösterdiğim çabanın boşa gitmediğini ümit ederim. Kardeşim ve meslekdaşım Suna'yı da unutmamam gerekir. Son olarak oda arkadaşım Yük. Müh. Ümit Bakırcı'ya teşekkür etmek istiyorum. Her türlü aksiliklerime sebatla katlandı.

Tezime doğrudan katkısı olmasa da, tıkanıp kaldığım noktalarda gerçeklikten kaçabilmeme yardım edenler Douglas Noel Adams, Frank Patrick Herbert ve John Ronald Reuel Tolkien oldu. Düş gücümü geliştirebilmem onların sayesinde. Bu üçlünün arasına Dr. Isaac Asimov ve Dr. George Gamow da eklenebilir.

En son olarak, yaşamı daha eğlenceli hale getiren Gölge Mavi arkadaş grubuna da teşekkür etmeliyim. Ne yaptığımdan tam emin olamasalar bile, iyi şeyler yaptığıma inandıkları anda cesaretimi arttırıp takdir ettikleri için.

ÖZET

Müzikal seslerin içerik analizi yapısal kodlama, ses veri tabanı sorgulama sistemleri, otomatik notaya dönüştürme, müzik eğitimi gibi çok çeşitli ve geniş bir uygulama alanına sahiptir. Bu uygulamaların önemli alt görevlerinden biri otomatik enstrüman tanıma işlemidir.

Bu çalışmada; müzikal enstrüman seslerinin aktif öğrenen bir PNN ile tanınması üzerinde durulmuştur. Öznitelik olarak farklı derecelerden LPC ve MFCC katsayıları kullanılmıştır. Mevcut aktif öğrenme algoritmaları PNN için uygun değildir. Bu nedenle, yeni bir aktif öğrenme algoritması geliştirilmiştir.

Uygulama esnasında pasif PNN'ler kullanılarak en yüksek başarıyı veren öznitelikler tespit edildikten sonra, aktif öğrenme algoritması yalnız bu özniteliklerle birlikte uygulanmıştır. Yapılan deneyler sonucunda aktif öğrenen PNN'in başarımının daha yüksek olduğu gösterilmiştir.

Anahtar kelimeler: İşlemsel işitsel durum analizi, olasılıksal sinir ağları, aktif öğrenme, enstrüman tanıma.

ABSTRACT

Musical Audio Content Analysis has a wide application area such as structural encoding, sound database search systems, automatic note transcription, musical education etc. The most important part of these applications is the automatic instrument recognition task.

In this work an active learning PNN was used to recognise instrumental sounds. LPC and MFCC coefficients with different orders were used as features. Existing active learning algorithms are not suitable for PNN. Hence, a new active learning algorithm for PNN was developed.

The best feature sets were found by using passive PNNs and these sets were used with active learning PNNs. By realizing some experiments, it was shown that the entire performance was improved by using the active learning algorithm

Keywords: Computational auditory scene analysis, Probabilistic Neural Networks, active learning, instrument recognition.

1. GİRİŞ

Müzikal seslerin içerik analizi; yapısal kodlama, ses veri tabanı sorgulama sistemleri, otomatik notaya dönüştürme, müzik eğitimi gibi çok çeşitli ve geniş bir uygulama alanına sahiptir. Bu tür uygulamaların önemli alt görevlerinden biri otomatik enstrüman tanıma işlemidir. Enstrüman tanıma işleminin bir başka önemi de, bu alandaki çalışmalardan elde edilen sonuçların konuşma tanıma, konuşmacı tanıma gibi farklı alanlarda da uygulanabilir oluşudur.

Enstrüman tanıma sistemlerinin gelecek on yıl içerisinde önemli bir ticari meta haline gelmesi beklenmektedir. MPEG-7 hareketli görüntü ve ses sıkıştırma standardı, ses katmanı içerisinde bir enstrüman tanımlama bilgisi içermektedir. Bu bilginin insan gücü ile elde edilerek katman içerisine yerleştirilmesi zor ve pahalı bir yöntemdir. Bunun dışında, günümüzde ses veri tabanları ve ses müzeleri kurulmaktadır. Bir ses veri tabanı içerisinde yapılacak her türlü sorgu için MPEG-7 standardındakine benzer bir İşlemsel İşitsel Durum Analiz (Computational Auditory Scene Analysis - CASA) bilgisi gerekmektedir. Bu bilginin insan gücü ile elde edilmesi de yine zor ve pahalı bir işlemdir. Bir örnek vermek gerekirse, İngiliz Ulusal Müzesi, bir milyondan fazla ses CD'sinden oluşan bir koleksiyona sahiptir. Müze, halihazırda koleksiyon içinde yalnızca albüm, şarkı ya da sanatçı adıyla arama hizmeti verebilmekte, fakat örneğin "saksofon ve kontrbas ile caz stili çalınan Bach parçaları" gibi aramalar yapılamamaktadır. Bu çalışmada önerilen yöntem, bu tür sorunlara da alternatif bir çözüm olarak değerlendirilebilir.

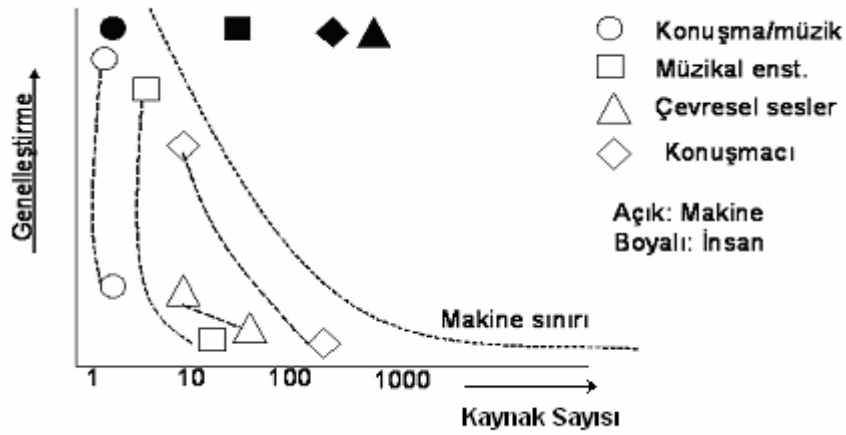
19. yüzyılda Helmholtz ile başlayan müzikal sinyallerin tanımlanması problemi, halen kesin bir sonuca ulaşamamıştır. İnsan beyninin ses sinyallerini ne şekilde işlediği halen çözülememiş problemlerden biridir. Var olan işlemsel işitsel durum analizi çalışmaları kabul edilebilir seviyede başarılı olmalarına rağmen halen insan becerilerinin uzağındadır.

Bir müzikal ses, dört temel algısal özelliğe sahiptir: perde, güç, süre ve tını. Bu algısal özellikler, dinleyicinin dinlemekte olduğu sesi tanımlamasına olanak sağlar. Perde, güç ve sürenin fiziksel anlamları kesin bir dille tanımlanabildiği halde, tını için böyle bir tanım yapılamamıştır. Müzikal sesler için perde, sinyalin ana harmoniğinin frekansına denk düşmektedir. Güç, sinyalin ne kadar kuvvetli olduğunun göstergesidir. Üçüncü özellik, sesin varlığının algılanabildiği süredir. Tını, ANSI tarafından aynı perde ve eşit güçte olan farklı kaynaklardan gelen seslerin birbirinden ayırt edilebilmesini sağlayan özellik olarak tanımlanmaktadır. Görüldüğü üzere bu tanım tınının ne işe yaradığını anlatmakta, ancak

gerçekte ne olduğunu ve ne şekilde ölçülebileceğini söylememektedir. Geçmişte yapılmış olan fizyo-akustik çalışmalarda, benzer seslerin birbirlerinden ayrılmasını sağlayacak çeşitli akustik öznitelikler tanımlanmıştır. Bu öznitelikler probleme tam bir çözüm sunamasa da önemli ölçüde yararlı olmuşlardır.

Ses sinyallerinin bilgisayarlar yardımıyla tanınması çalışmalarında iki temel işlem basamağı tanımlanmıştır. İlk basamak tanıma işleminde kullanılacak özniteliklerin kestirimi, ikinci basamak ise bu öznitelikleri kullanarak tanıma işlemini gerçekleştirecek bir yapının tasarımıdır. Literatürde yüzlerce çeşit öznitelik grupları tanımlanmıştır. Tanıyıcı yapılar olarak genellikle Gizli Markov Modelleri (Hidden Markov Model - HMM), Gauss Karışım Modelleri (Gaussian Mixture Model - GMM) ve Yapay Sinir Ağları kullanılmaktadır.

Yapay ses tanıma sistemleri halen gerçek dünyadan gelen verileri değerlendirmede yetersizdir. Bu tür sistemler gerçek dünyadan gelen gürültülü ve karmaşık sesleri değil, stüdyo ortamında hazırlanmış, izole ve/veya monofonik sesleri kullanırlar. Brown (1999), 4 nefesli enstrümanı insan yeteneklerine yakın bir başarımla sınıflandırabilmiştir. Martin'in (1999) ve Eronen'in (2001) sistemleri çok daha fazla sayıda enstrümanı sınıflandırabilmektedir, ancak her iki sistemin başarımı da insan yeteneklerinin çok altındadır. Genel olarak, sınıflandırılması istenen enstrüman sayısı arttırıldıkça yapay sistemlerin başarımı insan başarımından daha hızlı azalmaktadır (Şekil 1.1).



Şekil 1.1 İnsan ve makine yeteneklerinin karşılaştırılması (Martin, 1999).

1.1 Tezin Amaç ve Önemi

Bu çalışmada enstrümantal seslerin çeşitli öznitelik grupları kullanılarak yapay sinir ağları

aracılığı ile tanınması amaçlanmıştır. Çalışmada kullanılan öznitelik grupları literatürde en yüksek başarıyı veren öznitelikler arasından seçilmiştir. Geçmişte yapılmış olan çalışmalarda daha çok öznitelik gruplarının üzerinde durulmuştur ancak tanıma işlemini gerçekleştiren yapıların fazlaca bir öneminin olmadığı düşünülmüştür. Bu yaklaşım bir noktaya kadar doğru olmakla birlikte, tanıyıcı yapıların da önemli sayılması gerektiği göz ardı edilmemelidir. Gerçekleştirilen çalışmayı literatürden farklı kılan en önemli özellik, başarımlar açısından tanıyıcı yapıların da en az tanımlayıcı öznitelikler kadar önem taşıdığı düşüncesidir. Bu çalışmada, insan öğrenme yöntemlerine geleneksel makina öğrenmesinden daha fazla benzeyen aktif öğrenme yöntemleri kullanılarak daha yüksek başarımlar elde edilebileceği kanıtlanmaya çalışılacaktır. Çalışma boyunca ilk aşamada geleneksel Olasılıksal Sinir Ağları (PNN) kullanılacak, daha sonra yeni bir aktif öğrenme algoritması ile tanıma işlemi gerçekleştirilecektir. Elde edilen sonuçlar hem kendi aralarında, hem de geçmiş çalışmalarla karşılaştırılarak, önerilen öğrenme yönteminin başarımlarını arttırdığı savı doğrulanmaya çalışılacaktır. Tanıyıcı sistemin başarımlarını sisteme uygulanan test kümesinin doğru tanıma oranı, başka bir değişle doğruluk yüzdesi kullanılarak ölçülecektir.

1.2 Kullanılan Veri Kümesi

Müzikal seslerin içerik analizi çalışmalarında birkaç farklı veri tabanı kullanılmaktadır. Iowa Üniversitesi tarafından hazırlanan veri kümesi internet üzerinden ücretsiz olarak temin edilebilir ancak bu veri kümesi diğerlerine göre oldukça küçüktür. l'Institut de Recherche et Coordination Acoustique/Musique (IRCAM) veri kümesi yine internet üzerinden erişime açık, ancak ücretlidir. Üçüncü ve en büyük veri kümesi McGill Üniversitesi tarafından hazırlanan McGill University Master CD Samples (MUMS)'dir. MUMS veri kümesi neredeyse tüm klasik batı müziği enstrümanlarını, vurmaları ve bazı caz parçalarını kapsamaktadır. Veri kümesi toplam 11 CD'den oluşmaktadır. Bu çalışmada kullanılan ses örnekleri MUMS CD-1, 2 ve 3'ten elde edilmiştir.

MUMS ses veri kümesindeki sesler ön iyileştirme yapılmadan 44.1 KHz. örnekleme hızında örneklenmiş ve Sony PCM 3202 Dash kaydedici ile kaydedilmiştir. B&K 4300 ve Sennheiser MKH40 mikrofonlar ile alınan sesler bir Sony/MCI MXP 3630 karıştırıcıdan geçirilmiştir. Bazı sesler, karıştırıcı üzerindeki 4 kanallı Wien tipi bir dengeleyici ile bir miktar iyileştirilmiştir. Tüm enstrümanlar profesyonel müzisyenler tarafından çalınmıştır. Her bir enstrüman tek bir icracı tarafından çalınmıştır.

1.3 Tezin Düzeni

Sunulan bu çalışmada batı müziği enstrümanlarının aktif öğrenen Olasılıksal Sinir Ağları ile tanınmasına çalışılmıştır. Öznitelik olarak doğrusal öngörü katsayıları ve mel frekans cepstral katsayıları kullanılmıştır.

Çalışmanın ikinci bölümünde geçmişte yapılmış olan çalışmalar özetlenmiştir. Ses tanıma işlemlerinin temel sorun ve çözüm önerilerine bu bölümde değinilmiştir. Üçüncü bölümde literatürde en sık kullanılan özniteliklere yer verilmiştir. Dördüncü bölümde aktif öğrenme ve aktif öğrenen bir Olasılıksal Sinir Ağı yapısı üzerinde durulmuştur. Beşinci bölümde yapılan uygulamalarda elde edilen sonuçlar özetlenmiş, altıncı bölümde ise bu sonuçlar irdelenmiştir.

2. GEÇMİŞ ÇALIŞMALAR

Akustik dalgaların analizi ile ilgili çalışmalar, Galileo'ya kadar gitmektedir. Galileo'nun bir telin gerginliği ile telin titreşirken ürettiği sesin temel frekansı arasındaki ilişkiyi tanımladığı 1636 yılında, Fransız matematikçi, filozof ve rahip Martin Mersenne de perde periyodunun saniyedeki titreşim sayısı ile ilişkili olduğunu göstermişti. Akustik dalgaların yapısına yönelik bir sonraki önemli çalışma ise, 19. yüzyılda, Helmholtz tarafından yapılmıştır. Helmholtz, akustik süzgeçler kullanarak her hangi bir karmaşık ses sinyalinin bir dizi basit sinyalin toplamından ibaret olduğunu ispatlamış, bu ispatı kullanarak ilk toplamsal ses sentezcisi sayılabilecek mekanik bir enstrüman geliştirmiştir. Bu enstrüman, her biri bir tokmak ile uyarılan ve tekdüze bir ses üreten yaylardan oluşmaktaydı. Helmholtz bu düzenek yardımıyla aynı anda birden fazla yayı titreştirerek basit seslerin toplamından oluşan karmaşık sesler üretmiştir (Pierce, 1983). Bu basit çalışmalar ile başlayan akustik sinyal işleme uygulamaları, günümüzde ayrı bir bilim dalı oluşturabilecek kadar genişlemiştir. Bu genişleme ile birlikte akustik sinyal işleme kendi içinde alt gruplara bölünmüştür.

Ses sinyallerinin tanınması ve sınıflandırılması uygulamalarına genel olarak İşlemsel İşitsel Durum Analizi (CASA) adı verilir. CASA işlemleri arasında konuşma tanıma, konuşmacı tanıma, konuşma/müzik ayrımı, çevresel seslerin tanınması, tıbbi tanı sistemleri, müzikal seslerin tanınması gibi çok çeşitli ve farklı uygulama alanları sayılabilir. Bu alanlar arasında konuşma ve konuşmacı tanıma uygulama açısından en büyük alanı kaplarken, işitsel içerik analizi ve ses bölütleme uygulamaları da özellikle sinema ve televizyon endüstrisinin gelişimine paralel olarak ivmelenen bir hızla ticari uygulama alanları bulabilmektedir (Zhang ve Kuo, 2001).

İnsan duyma sistemi, kendisine ulaşan akustik dalgayı tanıma, tanımlama ve sınıflandırma yeteneklerine sahiptir. Bir insan, duymakta olduğu sesin bir köpeğe ait olduğunu, köpeğin havlayışının neşeli ya da sinirli oluşunu, köpeğin büyük mü, küçük mü olduğunu ayırt edebilir. Hatta kişi bir köpek besliyor ise, köpeğin kendine ait olup olmadığını bile söyleyebilir. İnsan duyma sisteminin böylesi karmaşık görevleri yerine getirebilmesi, sistemin anlaşılabilmesini güçleştirmektedir.

Ses tanıma işlemini zorlaştıran ana neden, aynı kaynak tarafından farklı zamanlarda üretilen seslerin farklı oluşudur. Eğer her kaynak her seferinde aynı dalgayı üretseydi, daha önceden saklanmış olan dalga şekilleri ile alınan dalga şekli karşılaştırılarak tanıma işlemi gerçekleştirilebilirdi. Ancak, gerçek dünyada farklı zamanlarda üretilen ses dalgaları

farklıdır. Bunun nedeni, sesi üreten fiziksel mekanizmaların farklı zamanlarda yoğunlukla farklı oluşundan kaynaklanır. Ayrıca, ses kaynağının dinleyiciye göre konumu ile ortamın akustik parametrelerinin farklı oluşu da önemli etkenlerdir (Eronen, 2001).

Dinleyici kendisine gelen sesi tanıyabilmek için kaynağın zaman ve mekandan bağımsız olan karakteristik özelliklerinden yararlanmak zorundadır. Bu özelliklere akustik sabitler adı verilir (Handel, 1995, McAdams, 1993 ve Martin, 1999). Ses üreten mekanizmaların sesin üretilme şekline bağlı bazı değişmez özellikleri vardır. Bu özellikler arasında uyarım şekli ve süresi, kaynağın yapıldığı malzeme, titreşen parçacıkların boy ve şekilleri sayılabilir. Bu değişmezlerce şekillendirilen öznel özellikler bulunabilirse, bulunan öznel özelliklerden yararlanarak akustik değişmezleri tanımlamak ve bu değişmezleri kullanarak tanıma işlemini gerçekleştirmek mümkün olabilir (Martin, 1999).

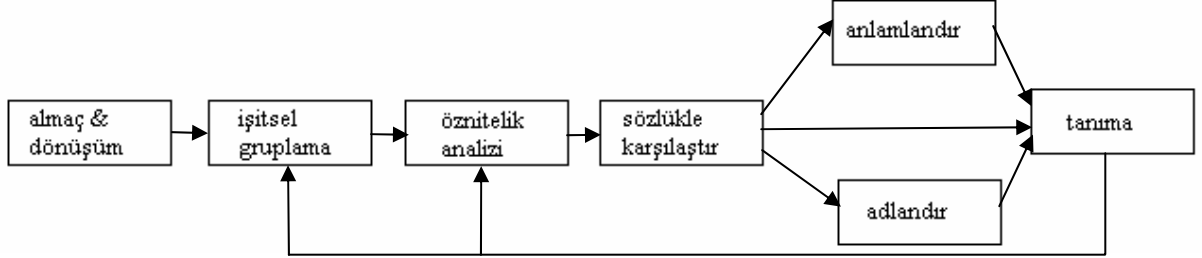
Ne yazık ki, akustik özelliklerin zamanla yavaş bir tempoyla ve sürekli olarak değişmesi, bu değişmezlerin elde edilmesini güçleştirmektedir. Akustik ortamın doğrusal olması da bir diğer zorlaştırıcı etkidir. Aynı anda birden fazla kaynak varsa, dinleyiciye ulaşan dalga kaynakların ürettiği dalgaların doğrusal bir bileşimi olacaktır (Handel, 1995).

Tanıma işlemi bir öğrenme ve deneyim süreci gerektirmektedir. Örneğin, insanlar farklı ortamların ses dalgalarını nasıl değiştirdiğini öğrenirler. Akustik özellikler ile kaynağın fiziksel özellikleri arasındaki ilişki kurmak için bir öğrenme sürecinden geçerler. Tanıma işlemini, uzun dönem belleklerinde oluşturdukları bir ses sözlüğü yardımıyla yaparlar (McAdams, 1993).

McAdams (1993), ses tanıma işleminin Şekil 2.1'deki gibi çok katmanlı bir yapı ile gerçekleştirildiğini öne sürmüştür. Kulağa ulaşan akustik dalga iç kulaktaki koklea organının yardımıyla beyin yorumlayabileceği elektriksel sinyallere dönüştürülür. Bu dönüşüm sırasında ilk frekans analizi ve dinamik sıkıştırma işlemi de gerçekleştirilir. İkinci adımda birden fazla kaynaktan gelen seslerin birbirlerinden ayrıştırılarak gruplanması gerçekleştirilir (Bregman, 1990). Bundan sonraki aşamada her bir ses için ayrı ayrı öznel özellik kestirimi yapılır. Bu noktada başlangıçtaki akustik dalga, kaynakları temsil eden bir grup akustik değişmeze dönüştürülmüş durumdadır. İşitsel sözlük ile karşılaştırma aşamasında girişteki işitsel veri, bellekte saklanmakta olan kaynak ve olaylara ilişkin bilgi ile karşılaştırılarak en yakın olduğu grupla özdeşleştirilir. Son aşamada ise, anlamlandırma, tanımlama, tanıma gibi işlemler gerçekleştirilir (Eronen, 2001).

Şekil 2.1'deki geri beslemeler bazı işitsel algılama olaylarını açıklayabilmek için konmuştur.

Örneğin, gürültü seviyesinin çok yüksek olduğu yerlerde bile, kişinin kendi adının söylendiğini algılayabilmesi bunlara bir örnektir. Ana dilin anlaşılabilmesi için gereken sinyal-gürültü oranının yabancı bir dili anlayabilmek için gerekenden daha düşük olması da bu olaylar arasında sayılabilir (Eronen, 2001, Bregman, 1990).



Şekil 2.1 Ses tanıma işlemi

CASA uygulamaları bir önceki bölümde bahsedildiği gibi çok farklı alanlara hitap edebilmektedir. Bu alanların amaç ve kapsamı her ne kadar birbirinden farklı da olsa, kullanılan araç ve yöntemler genellikle ortaktır. Bu nedenle, bu çalışmanın amacı olan müzikal seslerin sınıflandırılması işlemine bir alt yapı oluşturabilmek açısından diğer CASA uygulamalarına da kısaca göz atmak faydalı olacaktır. Bu bölümde, geçmişte yapılmış olan çeşitli CASA uygulamaları üzerinde kısaca durularak kullanılan araç ve yöntemler incelenmiştir.

2.1 Ses Bölütleme ve Sınıflandırma

Ses bölütleme ve sınıflandırma uygulamalarında en temel problem konuşma ve müzik sinyallerini ayırt etmedir. Bu iki sinyal türü, ses sinyalleri arasında en önemli iki grubu oluşturur (Zhang ve Kuo, 2001). Saunders (1996), öznitelik olarak yalnızca enerji ve ortalama sıfır geçiş sayısını kullanmış ve basit bir eşikten geçirme işlemi ile bir uygulama gerçekleştirmiştir. Bu uygulamada %90'dan fazla bir başarımla elde edilmiştir. Berenzweig ve Ellis (2001) sonsal olasılık öznitelikleri (posterior probability features) ve cepstral katsayılar kullanarak gerçekledikleri çalışmada sınıflandırıcı olarak Gizli Markov Model (Hidden Markov Model – HMM) kullanmışlardır. Bir müzik kaydı içinden solistin şarkı söylediği bölgeleri tespit etmeye çalıştıkları bu uygulamada sonsal olasılık öznitelikleri ile %70'e yakın başarımla elde ederken cepstral katsayılar ile %74 başarımla elde etmişlerdir. Bolat ve Küçük (2004), bir radyal tabanlı fonksiyon ağı (Radial Basis Function Networks - RBF) sıfır geçiş sayısı, ortalama kareköksel enerji ve izgesel ağırlık merkezinden oluşan bir öznitelik grubu ile eğitmiş ve %91.1 başarımla elde etmişlerdir. Aynı öznitelik grubunu kullanan Bolat ve diğerleri

(2004), konuşma/müzik ayrımını aktif öğrenen bir PNN ile %97 başarımla gerçekleştirmişlerdir. Her iki çalışmada da aynı veri kümesi kullanılmıştır. Toplam güç, alt bant güçleri, parlaklık, bant genişliği, perde periyodu gibi algısal öznitelikler ve mel frekans cepstral katsayılarını (Mel Frequency Cepstral Coefficients - MFCC) kullanan Guo ve diğerleri (2001) algısal öznitelikler için %89 başarımla elde ederken, MFCC ile %92 başarımla elde etmişlerdir.

Ses bölütleme, konuşma ve müzik dışındaki kategorileri de katarak daha geniş bir probleme dönüştürülebilir. Ses sinyalinin konuşma, sessizlik, gülme ve konuşma dışı sesler olarak ayrıldığı bir uygulamada (Kimber ve Wilcox, 1996) sınıflandırıcı olarak HMM, öznitelik olarak da cepstral katsayılar kullanılmıştır. Bu uygulamada amaç, bir toplantıdaki tartışmanın parçacıklara bölünmesidir. Lu ve diğerleri (2001) 5 sınıfa (sessizlik, müzik, arka plan sesleri, saf konuşma ve karışık konuşma) ayırdıkları sesleri MFCC ve algısal/fiziksel öznitelikler yardımıyla sınıflamışlardır. Sınıflayıcı olarak Destek Vektör Makinesi kullanılan bu çalışmada %95 başarıma ulaşılmıştır.

2.2 İçeriğe Bağlı Ses Erişimi

İçeriğe bağlı ses erişiminde önemli bir teknik, mırıldanarak sorgulamadır (Zhang ve Kuo, 2001). Günümüzde ticari uygulamalarının da bulunduğu bu teknikte kullanıcı bir ses veri tabanı içinde aramakta olduğu bir eseri, bir kısmını mırıldanarak arayabilmektedir. (Foote, 1997)'de önerilen sistemde öznitelik olarak MFCC kullanılmış ve ağaç tabanlı bir sınıflandırıcı kurulmuş, ancak önerilen yöntem müzik eserlerini çevresel seslerden ayırmakta fazla başarılı olmamıştır. Melodi konturunun perde periyodunun bağlı değişimi ile temsil edilebileceğini öneren bir çalışmada (Ghias ve diğerleri, 1995) 183 şarkılık bir veri tabanının %90'ı doğru olarak sınıflandırılmıştır. İçeriğe bağlı çevrim içi ses sınıflandırma ve bölütleme için önerilen bir başka yöntemde (Zhang ve Kuo, 1998) fiziksel öznitelikler kullanılarak %90 başarımla elde edilmiştir. Bu çalışmada sınıflandırıcı olarak HMM kullanılmıştır. Li ve diğerleri (2001) MFCC, LPC ve çeşitli fiziksel öznitelikler kullandıkları bir çalışmada, en yüksek başarımla veren öznitelik grubunun MFCC olduğunu göstermişlerdir.

2.3 Video İndeksleme İçin Ses Analizi

(Liu ve diğerleri, 1997) ve (Liu ve diğerleri, 1998)'de, 5 farklı video grubunu tanıyabilmek için ses analizi kullanılmıştır. Öznitelik olarak sessizlik oranı, konuşma oranı ve alt bant enerji oranı kullanılan bu uygulamada tanıyıcı olarak da Çok Katmanlı Ağ (Multi Layer Perceptron -

MLP) ve HMM kullanılmıştır. Her iki çalışmanın sonunda da HMM'nin başarımının daha yüksek olduğu bildirilmiştir.

2.4 Konuşma ve Konuşmacı Tanıma

CASA uygulamaları arasında hem akademik, hem de ticari olarak en fazla araştırma konuşma ve konuşmacı tanıma üzerine yapılmıştır. Konuşma ve konuşmacı tanıma literatürü, bir cilde sığdırılamayacak kadar geniş olduğundan, burada yalnızca bazı önemli çalışmalara çok kısa olarak yer verilmiştir.

Konuşma tanıma uygulamalarında en sık kullanılan tanıyıcı sistemler HMM (Gravier ve diğerleri, 1999, Gülmezoğlu ve diğerleri, 1999, Russel ve Jackson, 2005, Ito ve diğerleri, 2005, Mashao, 1996, Sarıkaya, 2001) ve LVQ (Learning Vector Quantisation) (Kondo ve diğerleri, 1994, Sarıkaya, 2001)' dur. Öznitelik olarak ise MFCC (Gravier ve diğerleri, 1999, Hyun ve Lee, 1999, Sarıkaya, 2001, Lee ve diğerleri, 2003, Choi ve diğerleri, 2002, Hariharan ve diğerleri, 2001), cepstrum (Gülmezoğlu ve diğerleri, 1999, Sarıkaya, 2001), süzgeç bankası çıkışı (Gravier ve diğerleri, 1999), LPC (Gülmezoğlu ve diğerleri, 1999, Mashao, 1996, Lee ve diğerleri, 2003), spektrum (Kondo ve diğerleri, 1994, Wang, 2003, Sarıkaya, 2001), DCT (Discrete Cosine Transform) (Hariharan ve diğerleri, 2001) ve sıfır geçiş sayısı (Gülmezoğlu ve diğerleri, 1999,) kullanılmıştır.

Konuşmacı tanıma uygulamalarında da, benzer şekilde HMM (Matsui ve Furui, 1994, Reynolds ve Rose, 1995, 1996, Reynolds ve Carlson, 1995) en çok kullanılan tanıyıcı yapıdır. LVQ da yine sıklıkla kullanılmaktadır (Matsui ve Furui, 1994, Soong ve diğerleri, 1982). En sık kullanılan öznitelikler ise cepstrum (Matsui ve Furui, 1994, Higgins ve diğerleri, 1991), MFCC (Seddik ve diğerleri, 2004, Reynolds ve Rose, 1995, 1996, Reynolds ve Carlson, 1995) ve LPC (Soong ve diğerleri, 1982, Tishby, 1991, Higgins ve diğerleri, 1991)'dir.

2.5 Müzikal Enstrüman Tanıma

2.5.1 İnsanlar Üzerinde Yapılan Deneyler

İnsan dinleyiciler ile yapılan deneyler ne yazık ki fazla yaygın değildir. Mevcut olan deneyler ise deneylerin düzenleniş şartları açısından tatmin edici değildir. Çoğu çalışma kısıtlı sayıda enstrüman ve dar bir nota aralığında yapılmıştır.

Brown ve diğerleri (2001) bir grup dinleyici ile yaptığı çalışmanın sonucunda 4 enstrümanın ortalama %85 başarımla doğru sınıflandırıldığını bulmuştur. Bu çalışmada kullanılan

enstrüman sesleri obua, klarnet, saksofon ve flüte aittir. Sesler izole notalar şeklinde kaydedilmiştir. Dinleyici grubu, 15 müzisyenden oluşmaktadır. Dinletilen seslerin enstrümanlara göre dağılımı dinleyicilerden gizlenmiştir. Bu deneyde flüt %93 ile en yüksek tanınma oranına erişirken, klarnetin tanınma oranı %71’de kalmıştır. Bu çalışmada toplam 60 farklı ses kaydı kullanılmıştır. Brown’un bir başka çalışmasında (Brown, 1999) 2 enstrüman (obua ve saksofon) kullanılmış ve %89 başarımla elde edilmiştir.

Bu alandaki en geniş araştırma Martin (1999) tarafından yapılmıştır. Bu çalışmada 14 enstrüman kullanılmış ve %46 başarıma erişilmiştir. Martin dinleyici olarak 14 müzisyen kullanmıştır. Deney grubuna 14 farklı enstrümanın sesleri sırayla dinletilerek, dinletilen sesin 27 ayrı enstrümandan hangisine ait olabileceği sorulmuştur. Bazı enstrümanlar farklı çalma tekniklerinde çalınmış, bu teknikler de farklı enstrümanlar gibi nitelendirilmiştir. Tek tek enstrümanların tanınma başarımları %46 olarak bulunurken, enstrümanların ait oldukları aileler ise %92 başarımla doğru tespit edilmiştir. Bu çalışmada kullanılan sesler, MUMS veri kümesinden alınmıştır. Çalışmanın ikinci aşamasında ise 10 saniyelik solo pasajlar kullanılmıştır. Uzun süreli solo pasajların kullanılmasıyla birlikte enstrüman tanıma başarımları %67’ye yükselirken, aile tanıma başarımları %97 olmuştur.

Çizelge 2.1 İnsan deneklerle yapılan tanıma deneylerinin sonuçları

	Doğruluk Yüzdesi	Enstrüman Sayısı
Eagleson ve Eagleson(1947)	56	9
Saldanha ve Corso (1964)	41	10
Berger (1964)	59	10
Clark ve Miller (1964)	90	3
Strong ve Clark (1967)	85	8
Campbell ve Heller (1978)	72	6
Kendall (1978)	84	3
Brown (1999)	89	2
Martin (1999)	46/67	27
Brown ve diğerleri (2001)	85	4

Campbell ve Heller (1978) 6 farklı enstrümanla yaptığı bir deneyde %72 başarıma erişirken,

Berger'in (1964) eski sayılabilecek bir çalışmasında 10 enstrüman için %59 başarımlar rapor edilmiştir. Bu alanda yapılmış olan önemli çalışmaların bir özeti Çizelge 2.1'de verilmiştir.

2.5.2 Otomatik Sistemlerin Başarımı

Enstrüman seslerinin otomatik tanınması üzerine geniş bir literatür mevcuttur. Bu çalışmalarda çoğunlukla tek bir kaynaktan alınan izole notalar kullanılmıştır. Nispeten eski çalışmalarda enstrüman sayıları sınırlı tutulmuş ve genellikle enstrümanların üretebileceği tüm notalar değil, bir ya da birkaç oktavlık kısımları kullanılmıştır. Yeni sayılabilecek çalışmalarda ise, izole notalar yerine ticari CD'lerden elde edilen solo pasajlar kullanılmıştır. Polifonik tanıma yönelik birkaç çalışma da mevcut olmakla birlikte, bu alandaki araştırmalar henüz yeterli olgunluğa ulaşabilmiş değildir.

Otomatik tanıma sistemleri içinde en fazla enstrümanla yapılan çalışma Eronen'e (Eronen, 2001) aittir. Eronen'in bu çalışmasında farklı derecelerden LP, MFCC ve delta MFCC katsayıları ile F_0 , atak süresi, spektral ağırlık merkezi, 4–8 Hz AM modülasyon gücü, 10–40 Hz AM modülasyon gücü gibi öznitelikler kullanılmıştır. 30 farklı enstrüman önce hiyerarşik bir yapı yardımıyla 6 aileye bölünmüş, daha sonra aile içi tanıma işlemi gerçekleştirilmiştir. Hiyerarşinin her düğümünde GMM ya da k-NN (k- Nearest Neighborhood Networks) tipi tanıyıcılar kullanılmış, en alt seviyede ise sadece k-NN kullanılmıştır. Eronen, en yüksek test başarımlarını veren öznitelik grubunun %32 ile MFCC olduğunu göstermiştir. Hesaplanan tüm özniteliklerin kullanılmasıyla başarımlar %37 olarak bulunmuştur. Aile tanımada ise en yüksek başarımlar %94,7 olarak elde edilmiştir. Bu çalışmadaki toplam başarımlar %80,6'dır. Eronen (2003) öznitelik olarak MFCC katsayıları kullandığı bir başka çalışmasında 27 enstrümanı HMM yardımıyla %68 başarımla sınıflandırmıştır.

Fujinaga ve MacMillan (2000) 23 enstrümanla yaptıkları çalışmada %68 doğru tanıma oranına ulaşmıştır. Fraser ve Fujinaga (1999) ise 23 enstrümanla %64 başarımlar elde etmiştir. Bu iki çalışmada öznitelikler spektrumdan hesaplanmış ve en iyi sonucu veren öznitelikler bir genetik algoritma ile bulunmuştur. Kaminskyj ve Materka (1995) gitar, piyano, marimba ve akordiyonla yaptığı çalışmada %98 doğru sınıflama oranına erişirken, Kostek ve Czyzewski (2001) obua, trompet, keman ve çello ile %93 başarımlar elde etmiştir. Kaminskyj ve Materka öznitelik olarak kareköksel ortalama enerjiyi (RMS) kullanmıştır. Bu çalışmalarda sınıflandırıcı olarak k-NN kullanılmıştır. Kostek ve Czyzewski ise öznitelikleri spektral zarf ve atak bölgesinden çıkarmış, sınıflandırıcı olarak da 2 katmanlı ileri beslemeli ağ kullanmıştır. Kostek (2005) bir başka çalışmasında 12 enstrümanı ayırık dalgacık

dönüşümünden elde ettiği öznitelikler yardımıyla sınıflandırmıştır. MLP kullanılarak gerçekleştirilen bu çalışmada, toplam başarımlar yerine enstrümanların doğru tanıma yüzdeleri verilmiştir. Bu sonuçlara göre en yüksek başarımla tanınan enstrüman trompet (%90,48), en düşük başarımla tanınan enstrüman ise keman (%56,19) olmuştur.

Kitahara ve diğerleri (2003) 19 enstrümandan oluşan bir veri kümesini %79,73 başarımla sınıflandırmışlardır. Enstrümanlar piyano, gitar, telli, piriç nefesli, saksofon, çift kamışlı, klarnet ve kamışsızlar olmak üzere 8 alt gruba bölünmüş ve hiyerarşik bir sınıflandırma gerçekleştirilmiştir. Tanıyıcı olarak 3-NN kullanılan bu çalışmada özgün öznitelik vektörünün uzunluğu 120'dir. Ancak çalışmada PCA (Principal Component Analysis) ve LDA (Linear Discriminant Analysis) kullanılarak öznitelik vektörünün boyutu azaltılmıştır. Bahsedilen başarımla 10-parça çapraz doğrulama ile erişilmiştir.

Martin ve Kim (1998), MUMS veri kümesini kullanarak gerçekleştirdikleri çalışmada 14 enstrümanı k-NN kullanarak %72 başarımla sınıflandırmışlardır. Bu çalışmada hiyerarşik bir tanıma mekanizması kullanılmıştır. Martin (1999) bir başka çalışmasında 14 enstrümanı 27 olası sınıfa ayıran bir sistem geliştirmiştir. Bu çalışmada önce enstrüman ailelerini bulan, daha sonra enstrümanı tespit eden bir Bayes sınıflandırıcı kullanılmıştır. Öznitelikler, weft gösteriminin (Ellis, 1997) geliştirilmiş bir türevi olan log-lag correologram dönüşümünden elde edilmiştir. Sınıflandırıcı, her aşamada bir önceki aşamadan elde edilen sonuca bağlı olarak farklı öznitelik gruplarını kullanmaktadır. Hesaplanan toplam öznitelik sayısı yüzden fazladır. Sistemin toplam başarımları %39 olmuştur.

3. SES TANIMADA KULLANILAN ÖZNİTELİKLER

Öznitelik kestirimi çoğunlukla bir sınıflandırıcı sistemin ilk aşamasını oluşturur. Ses tanıma sistemlerinde kullanılan öznitelikler çeşitli şekillerde gruplandırılabilir. En temel gruplandırma, yapısal ve algısal öznitelikler olarak yapılabilir (Gerhard, 2000). Algısal öznitelikler insan duyma sisteminden esinlenilerek hesaplanan özniteliklerdir. Perde periyodu, tını ve ritim gibi öznitelikler bu gruptadırlar. Yapısal öznitelikler ise, ses sinyalinin matematiksel ve istatistiksel özelliklerine karşılık düşerler.

Yapısal öznitelikler kendi içinde alt gruplara bölünebilir. Bu alt gruplar spektral, cepstral, LPC ve fiziksel öznitelikler olarak sayılabilir. Spektral öznitelikler sinyalin spektrumundan, cepstral öznitelikler ise cepstrumdan hesaplanır. Fiziksel öznitelikler sinyalin fiziksel özelliklerine karşılık düşen özniteliklerdir ve her biri farklı bir yöntemle hesaplanır. En çok kullanılan fiziksel öznitelikler arasında sıfır geçiş sayısı, alçak frekans AM modülasyon miktarı, atak süresi, enerji gibi öznitelikler sayılabilir.

Bu bölümde, en yaygın olarak kullanılan öznitelikler üzerinde durulacaktır.

3.1 Fiziksel Öznitelikler

Fiziksel öznitelikler sinyalin fiziksel özellikleri ile ilişkili olduğundan hesaplanması nispeten daha kolay özniteliklerdir. Algısal modellere ihtiyaç duymaması nedeniyle fiziksel öznitelikler özellikle işlemsel karmaşıklığın düşük olmasının istendiği durumlarda yaygın olarak kullanılmaktadır.

3.1.1 Enerji

Enerji, sinyalin herhangi bir zamanda ne kadar yoğun olduğunun bir göstergesidir (Gerhard, 2000). Enerji, sinyalin hangi periyotlarla zayıflayıp kuvvetlendiğinin gözlenmesi için kullanılabilir. Enerji, genellikle kısa süreli olarak ölçülür. Sinyal dikdörtgen pencereler yardımıyla kısa süreli parçacıklara bölünür. Parçacıklardan hesaplanan enerjiler, sinyalin enerjisinin zamanla değişimini gösterir. Sinyalin enerjisinin karekökü olarak tanımlanan ortalama kareköksel enerji (Root Mean Square Energy – RMS E) de öznitelik olarak kullanılmaktadır (Bolat ve Küçük, 2004).

Enerjiye ilişkin bir başka öznitelik de alt bant enerji dağılımıdır. Alt bant enerji dağılımı sinyalin iki ya da daha fazla alt banda bölünerek her bir alt bandın taşıdığı enerji miktarının ölçülmesi ile elde edilir (Gerhard, 2000).

3.1.2 Sıfır Geçiş Sayısı

Sıfır geçiş sayısı (Zero Crossing Rate - ZCR), genliğin işaretinin değişme sayısıdır. İlk olarak, ZCR ile sinyalin temel frekansı arasında basit bir ilişki kurulabilir. Sinyal tek bir bileşenden oluşuyorsa, bir periyot boyunca iki kez işaret değiştirecektir. Böylece ZCR, sinyalin temel frekansının iki katına eşit olur. Ancak, sinyal izgesel olarak zengin ise bir periyottaki ZCR değeri yükselecektir (Gerhard 2000). ZCR, ölçümü çok kolay olduğu için yaygın olarak kullanılan bir özneliktir. Literatürde sıfır geçiş sayısından yararlanılarak hesaplanabilen çeşitli öznelikler tanımlanmıştır (Kedem, 1986, Gerhard 2000).

3.1.3 Atak Süresi

Atak süresi, sinyalin başladığı an ile en yüksek değerine ulaştığı an arasında geçen süre olarak ölçülür. Literatürde çeşitli atak süresi ölçüm teknikleri tanımlanmıştır (Eronen, 2001, Martin, 1999, Hainsworth ve Macleod, 2003).

3.1.4 Genlik Modülasyonu

AM modülasyonu öznelikleri RMS enerji zarfı yardımıyla elde edilebilir. 4 – 8 Hertz arası AM modülasyonu tremolo, 10–40 Hertz arası AM modülasyonu ise sesin pürüzlülüğü ile ilişkilidir. AM modülasyonu hesaplanırken RMS enerji zarfı önce bir Hamming penceresi ile çarpılır. Pencereyle çarpılmış sinyalin FFT'si alınarak spektrumun 4–8 ya da 10–40 Hz. bantları arasındaki maksimum genliğe sahip frekans bileşeni, bu banttaki AM modülasyon frekansını gösterir (Eronen, 2001).

3.1.5 Vibrato

Vibrato, çalınan enstrümanın titreşen kısmının ya da uyarıcı sinyalin icracı tarafından periyodik olarak değiştirilmesi ile elde edilen doğal bir efekttir. Bu periyodik değişimler üretilen sesin ana harmonik frekansının periyodik olarak ileri geri değişmesine neden olur. Bazı enstrümanlar vibratolu çalmaya izin verirken, bazıları vermezler. Ayrıca, bazı enstrümanların standart çalınma şekilleri kendiliğinden vibratoya neden olur. Bu nedenle, vibrato enstrüman tanıma için kullanışlı bir özneliktir (Martin, 1999). Vibrato ölçümü için çeşitli yöntemler literatürde tanımlanmıştır (Martin, 1999, Rossignol ve diğerleri, 1999, Herrera ve Bonada, 1998).

3.2 Spektral Öznelikler

Bir sinyalin spektrumu, sinyalin içindeki farklı frekanslardaki bileşenlerin dağılımını gösterir.

İnsan kulağının doğrusal olmayan bir spektral analiz gerçekleştirdiği bilinmektedir (Moore, 1995). Buradan yola çıkarak, spektral özniteliklerin makine tanınmasında da kullanılabileceği söylenebilir. Spektral öznitelikler çok geniş bir aile oluşturmaktadır. Burada, spektral öznitelikler arasında en fazla kullanılanlara yer verilecektir.

3.2.1 Spektral Ağırlık Merkezi

Spektral ağırlık merkezi (Spectral Centroid – SC) basit fakat kullanışlı bir özniteliktir. Araştırmalar, SC'nin sesin algılanan parlaklığı ile ilişkili olduğunu göstermiştir. İnsan dinleyiciler tarafından daha parlak olarak nitelendirilen seslerin SC'leri diğerlerine göre daha yüksek bulunmuştur (Eronen, 2001). SC, spektrumun frekansa göre birinci dereceden momenti olarak tanımlanır. SC, Ayırık Fourier Dönüşümü (DFT) yardımıyla hesaplanabileceği gibi (Serra ve Bonada, 1998), bir süzgeç bankasının çıkışının ağırlıklı ortalaması alınarak (Eronen, 2001) da hesaplanabilir. Eşitlik 3.1 SC'nin N noktalı DFT ile hesaplanışını göstermektedir:

$$SC = \sum_k \frac{k}{N} f_k \frac{|X(k)|}{\sum_k |X(k)|} \quad (3.1)$$

3.2.2 Harmonik Bozulma

Müzikal sinyaller periyodik sesler olarak tanımlanırlar ancak bu tanım tam doğru değildir. Diğer harmonikler, ana harmoniğin tam katları değil, ancak tam katlarına oldukça yakındırlar. Sinyalin mükemmel harmoniklikten ne kadar uzak olduğunun bir ölçütü, harmonik bozulmadır (Serra ve Bonada, 1998). Harmonik bozulma şu şekilde hesaplanabilir:

$$HD = \sum_{i=1}^I |f_i - (F_0 \cdot i)| \frac{a_i}{\sum_{i=1}^I a_i} \quad (3.2)$$

Burada a_i , i . bileşenin genliği, F_0 ana harmoniğin frekansı, f_i ise i . bileşenin frekansıdır.

3.2.3 Spektral Düzensizlik

Spektral düzensizlik (Spectral Irregularity – IRR) bir spektral zarftan elde edilen harmonik genliklerin zaman ortalamalarının standart sapması olarak tanımlanmıştır (Eronen, 2001, Fujinaga ve MacMillan, 2000). A_k k . bileşenin genliğinin zaman ortalaması, r bileşen sayısı olmak üzere IRR şu şekilde hesaplanır:

$$IRR = 20 \log_{10} \left(\sum_{k=2}^{r-1} \left| A_k - \frac{A_{k-1} + A_k + A_{k+1}}{3} \right| \right) \quad (3.3)$$

3.2.4 Spektral Akı

Spektral akı (Spectral flux -SF) delta genlik spektrumunun Öklid normudur ve şu şekilde hesaplanır (Huang ve Hansen, 2004):

$$SF = \|S_i - S_{i-1}\|_2 = \frac{1}{N} \left(\sum_{k=0}^{N-1} (S_i(k) - S_{i-1}(k))^2 \right)^{1/2} \quad (3.4)$$

Burada S_i , i . çerçevenin genlik spektrumudur.

3.3 Doğrusal Öngörü Katsayıları

Doğrusal öngörü (LP) analizi tüm-kutup sinyal modelleme yöntemlerinden biridir. Bu yöntemde spektrum bir tüm-kutup süzgeç ile modellenir. Bu kutuplar, spektral tepelerde odaklanmıştır. İnsan kulağının süzgecin sıfırlarına neredeyse duyarsız olduğu bilinmektedir (Eronen, 2001). Bu nedenle, tüm-kutup bir süzgeç, ses işleme uygulamaları için yeterlidir.

İleri yönlü doğrusal öngöründe amaç, doğrusal ayırık zamanlı bir sistemin bir sonraki çıkış örneği $\hat{y}(n)$ 'i daha önceki p adet çıkışın doğrusal bir birleşimi ile elde etmektir:

$$\hat{y}(n) = \sum_{i=1}^p a_i y(n-i) \quad (3.5)$$

Buradaki a_i katsayıları LP katsayıları olarak adlandırılır. LP katsayıları tarafından tanımlanan tüm-kutup LP süzgecinin transfer fonksiyonu aşağıdaki gibidir.

$$H(z) = \frac{1}{A(z)} = \frac{1}{1 - \sum_{i=1}^p a_i z^{-i}} \quad (3.6)$$

LP'de amaçlanan, kısa süreli ortalama karesel hata e 'yi en az yapacak öngörü katsayılarını bulmaktır.

$$e = E \left\{ \left| y(n) - \sum_{i=1}^p a_i y(n-i) \right|^2 \right\} \approx \sum_{n=-\infty}^{\infty} \left| y(n) - \sum_{i=1}^p a_i y(n-i) \right|^2 \quad (3.7)$$

Burada $E\{\cdot\}$, beklendiği değer operatörüdür. Ortalama karesel hatayı en az yapan çeşitli algoritmalar mevcuttur. Burada, Rabiner (1993) tarafından geliştirilen özilişki yöntemine yer

verilecektir.

(3.7) denklemini minimize edebilmek için e 'nin a_i 'lere göre kısmi türevlerinin sıfıra eşitlenmesi gerekmektedir:

$$\frac{\partial e}{\partial a_i} = 0, i = 1, 2, \dots, p \quad (3.8)$$

Böylelikle normal denklemler elde edilir:

$$\sum_n y(n)y(n-k) - \sum_{i=1}^p a_i \sum_n y(n-i)y(n-k) = 0, k = 0, 1, \dots, p-1 \quad (3.9)$$

$y(n)$ durağan ve reel bir dizi ise,

$$R(k) = \sum_{m=-\infty}^{\infty} y(m)y(m+k) \quad (3.10)$$

özilişki fonksiyonu olmak üzere, (3.11) eşitliği şu şekilde düzenlenebilir:

$$\begin{bmatrix} R(0) & R(1) & \dots & R(p-1) \\ R(1) & R(0) & \dots & R(p-2) \\ \dots & \dots & \dots & \dots \\ R(p-1) & R(p-2) & \dots & R(0) \end{bmatrix} \begin{bmatrix} a_1 \\ a_2 \\ \dots \\ a_p \end{bmatrix} = \begin{bmatrix} R(1) \\ R(2) \\ \dots \\ R(p) \end{bmatrix} \Leftrightarrow \mathbf{R}\mathbf{a}=\mathbf{r} \quad (3.11)$$

Bu denklem takımı, Yule-Walker eşitliği olarak bilinir. $\mathbf{R}$ matrisinin çeşitli özelliklerinden yararlanarak (simetrik Toeplitz bir matris olması gibi) bu denklem takımı etkin bir şekilde çözülebilir. En etkin yöntemlerden biri, Durbin'in yinelemesidir. Durbin'in yinelemesi şu şekilde işler:

$$E^{(0)}=R(0) \quad (3.12)$$

$$k_i = \frac{R(i) - \sum_{j=1}^{i-1} a_j^{(i-1)} R(i-j)}{E^{(i-1)}}, \quad 1 \leq i \leq p \quad (3.13)$$

$$a_i^{(i)} = k_i \quad (3.14)$$

$$a_j^{(i)} = a_j^{(i-1)} - k_i a_{i-j}^{(i-1)}, \quad 1 \leq j \leq i-1 \quad (3.15)$$

$$E^{(i)} = (1 - k_i^2) E^{(i-1)} \quad (3.16)$$

Bu denklemlerin $i=1,2,...p$ için yinelemeli bir şekilde çözülmesi ile p . adımda elde edilen $a_j = a_j^{(p)}$ katsayıları, aranan LP katsayılarıdır. Aynı şekilde, k_j 'ler de yansıtma katsayılarıdır.

3.4 Mel Frekans Cepstral Katsayılar

Formantlar konuşma ve konuşmacı tanıma uygulamaları için oldukça kullanışlıdır. Formant kestiriminde önemli bir yöntem de cepstral katsayılarıdır. Herhangi bir $y(n)$ sinyalinin cepstrumu şu şekilde hesaplanır:

$$c(n) = F^{-1} \{ \log |F \{ y(n) \}| \} \quad (3.17)$$

Burada $F\{\cdot\}$, ayrık Fourier dönüşümüdür. Yukarıdaki denklem peş peşe iki DFT gerektirdiğinden pratikte uygulanabilir değildir. Cepstral katsayılar, LP katsayılarından ya da mel skalalı bir süzgeç bankasından daha az işlem yaparak elde edilebilir. DFT'nin doğrusal frekans çözünürlüğü sunmasına karşın insan algı sisteminin doğrusal olmayan frekans çözünürlüğüne sahip olması da (3.17) denkleminin pratik anlamda kullanışsız olmasına sebep olmaktadır (Eronen, 2001).

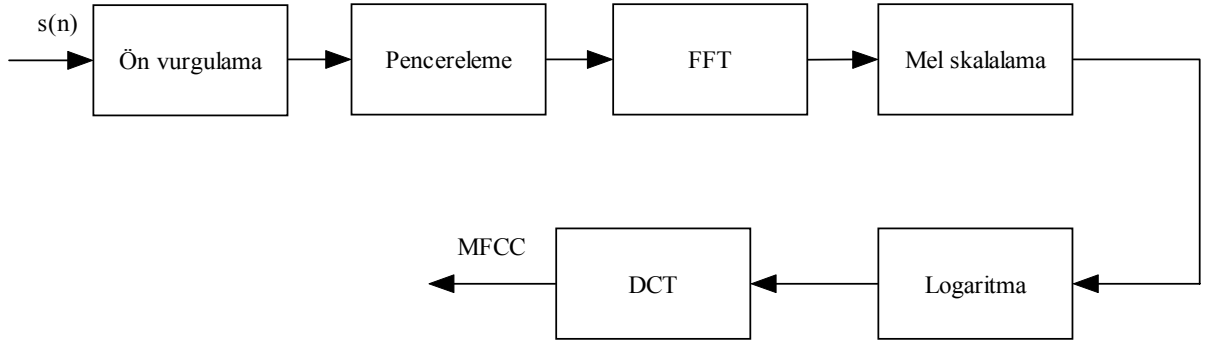
Mel birimi, insan duyma sisteminden esinlenerek geliştirilmiş doğrusal olmayan bir frekans ölçüm birimidir. Mel skalası ise doğrusal frekans eksenini (Hz) ile algısal frekans eksenini (mel) arasında bir birim dönüşümüdür (Sarıkaya, 2001, Eronen, 2001). Bu dönüşüm aşağıdaki eşitlik uyarınca gerçekleştirilir:

$$f_{mel} = 2595 \log_{10} \left(1 + \frac{f_{linear}}{700} \right) \quad (3.18)$$

Mel-frekans cepstral dönüşüm, Davis (1980) tarafından tanımlandıktan kısa bir süre sonra otomatik ses tanıma sistemleri için vazgeçilmez bir öznelite kestirim yöntemi olmuştur. MFCC hesaplamak için literatürde çeşitli yöntemler vardır. FFT temelli yöntem (Eronen, 2001) ve sabit-Q dönüşümüne dayalı yöntem (Brown, 1999) bunların arasında en sık kullanılan iki yöntemdir. Şekil 3.1 FFT temelli yöntemle ilişkin blok şemayı göstermektedir. Giriş sinyali öncelikle bir pencere fonksiyonu ile çarpılır. Pencerelemiş sinyalin genlik spektrumu, FFT kullanılarak elde edilir. Elde edilen spektrum, insan duyma sisteminden elde edilmiş bir süzgeç bankasına uygulanır. Süzgeç bankası, 27 logaritmik sıralanmış süzgeci (peş peşe 2 süzgecin merkez frekanslarının oranı 1.0711703'tür) takip eden 13 doğrusal sıralanmış süzgeçten (peş peşe 2 süzgecin merkez frekansları arasında 133.33 Hz. vardır) oluşmaktadır. Genlik spektrumu ile bu üçgensel süzgeçler çarpılarak her kanaldaki değerler toplanır.

Böylelikle her kanal için bir spektral genlik elde edilmiş olur. Süzgeç bankası çıkışlarının logaritması alınır. Logaritmik süzgeç bankası çıkışları m_j 'lerin Ayırık Kosinüs Dönüşümü (DCT), mel-cepstrumu verir:

$$c_{mel}(i) = \sum_{j=1}^N m_j \cos\left(\frac{\pi i}{N}(j-0.5)\right) \quad (3.19)$$



Şekil 3.1. MFCC'nin FFT yardımıyla hesaplanması

4. OLASILIKSAL SINIR AĞI ve AKTİF ÖĞRENME

Makine öğrenmesinde öğrenme işlemi, giriş uzayı X 'in küçük bir alt kümesi olan x ile, çıkış uzayı Y 'nin bir alt kümesi olan y arasında bir haritalama işlevi olarak tanımlanabilir. Öğrenen yapı (s), $X \rightarrow Y: s_w(x)=y$ olarak tanımlanan bir işlevi yerine getirir. Burada w alt indisi, öğrenene ait, eğitime süreci boyunca güncellenen uyarlanabilir bir parametre kümesidir (Hasenjager ve Ritter, 2000, Sugiyama ve Ogawa, 2000). Eğitime sürecinin amacı, kabul edilebilir herhangi bir E hata işlevinin w parametrelerinin uyarlanması ile en aza indirgenmesi olarak tarif edilebilir. Hata terimi, öğrenen ve öğretmenin aynı veri üzerinde varmış oldukları kararlar arasındaki fark olarak da nitelendirilebilir. W parametrelerinin uyarlanması öğrenenin verdiği kararların öğretmenin kararlarına yeteri kadar yaklaşmasına kadar sürdürülür. Öğretmenin asıl amacı eğitime verisinin tamamen öğrenilmesi (ezberleme) değil, öğretmenin gerçekleştirdiği işlevin kabul edilebilir bir modelini çıkarabilmektir (Hasenjager ve Ritter, 2000). Öğrenen yalnız eğitime verisini doğru işlemekle yetinmemeli, olay uzayından alınan ve daha önce hiç karşılaşmamış olduğu yeni örnekler hakkında da doğru kararlar verebilmelidir. Bu yeteneğe genelleştirme adı verilir.

Geleneksel yapay sinir ağı eğitime algoritmalarında öğrenen yapı (öğrenen), ortamı gözlemleyerek öğrenir. Öğrenen, kendisine verilen eğitime verisini kullanarak ortam hakkında genelleme yapabilmelidir. Eğitime verisi bilinmeyen bir kaynak tarafından üretilen giriş-çıkış çiftleridir. Kaynağın olasılık yoğunluk fonksiyonu da bilinmemektedir. Öğrenenin genelleştirebilme başarımı, öğrenenin yapısı, öğrenme yönergesi ve eğitime verisi gibi çeşitli faktörlere bağlıdır (Hasenjager ve Ritter, 2000). Geçmiş yıllarda yapılan araştırmalarda öğrenme başarımını ve genelleştirme yetisini arttırabilmek amacıyla daha çok öğrenme yönergesinin optimizasyonu üzerine yoğunlaşmıştır.

Genel olarak, eğitime verisi olay uzayından rasgele seçilen örneklerden oluşur. Eğitime setinin büyümesi ile birlikte, öğrenenin giriş uzayının daha büyük bir kısmı hakkındaki bilgisi artarken, bir yandan da bazı örnekler fazlalık bilgiye dönüşür. Bu nedenle, öğrenme ilerlerken örnek başına ortalama bilgi miktarı azalır (RayChaudhuri ve Hamey, 1995, Hasenjager ve Ritter, 2000, Takizawa ve diğerleri, 2000).

Aktif öğrenmede, öğrenen yalnızca pasif bir gözlemci değildir. Öğrenen, genelleştirme başarımını arttırabilecek yeni eğitime örneklerini seçebilme yetisine sahiptir. Benzer bir şekilde, büyüyen eğitime verisi ile birlikte fazlalık bilgiye dönüşmüş olan eski örnekleri de eğitime kümesinden çıkarabilmektedir (RayChaudhuri ve Hamey, 1995, Hasenjager ve Ritter,

2000, Takizawa ve diğerleri, 2000, Thrun, 1995, Leisch ve Hornik, 1998). Bu iki yeni yetiyi kullanan aktif öğrenen, tüm olay uzayını daha iyi temsil edebilecek bir eğitim kümesini kendi başına oluşturabilir. Aktif öğrenen tarafından seçilen böyle bir eğitim kümesi ile öğrenenin genelleştirme başarımı arttırılabilir.

4.1 Olasılıksal Sinir Ağları

İlk olarak Specht (1988, 1990a ve 1990b) tarafından önerilen Olasılıksal Sinir Ağları (PNN), Bayes-Parzen kestiriciler olarak da bilinir. K_1 veya K_2 sınıflarından birine ait olan, p boyutlu bir x örüntü vektörü olsun. K_1 ve K_2 sınıflarının olasılık yoğunluk fonksiyonları $F_1(x)$ ve $F_2(x)$ olsun. Bayes Karar Kuralı'na göre,

$$\frac{F_1(x)}{F_2(x)} > \frac{L_1 P_2}{L_2 P_1} \quad (4.1)$$

ise, x K_1 sınıfına aittir. Aynı şekilde eğer

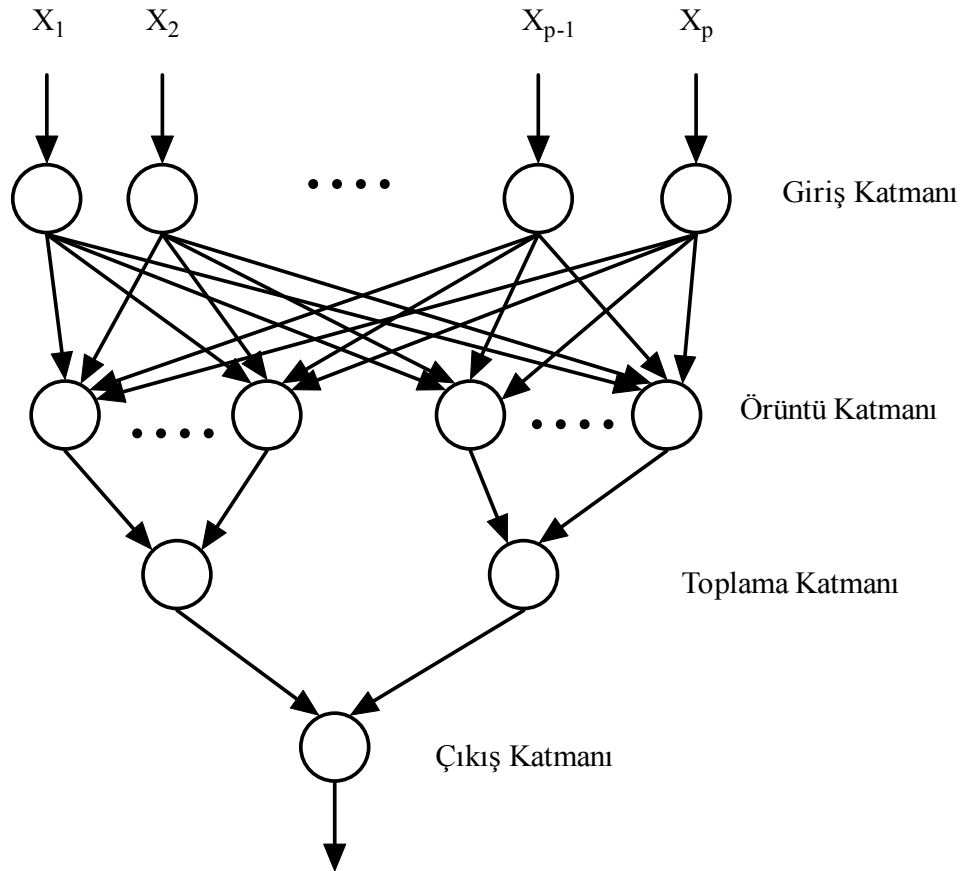
$$\frac{F_1(x)}{F_2(x)} < \frac{L_1 P_2}{L_2 P_1} \quad (4.2)$$

ise, x K_2 sınıfına aittir. Burada L_1 , x vektörü K_1 sınıfına ait ise x 'in hatalı sınıflandırılmasına ilişkin bir maliyet parametresi, L_2 ise x vektörü K_2 sınıfına ait ise x 'in hatalı sınıflandırılmasına ilişkin maliyet parametresidir. P_1 , K_1 sınıfından bir örnek ile karşılaşma sıklığı, P_2 , K_2 sınıfından bir örnekle karşılaşma sıklığıdır. Çoğu uygulamada, maliyet parametreleri ile karşılaşma sıklıkları eşit kabul edilir. Bu durumda (4.1) ve (4.2) eşitlikleri uyarınca, sınıfların olasılık yoğunluk fonksiyonlarının bulunması problemin çözümü için yeterlidir (Goh, 2002, Specht, 1988, 1990a ve 1990b). Bu yoğunluk fonksiyonları eğitim kümesinde yer alan örnekler kullanılarak hesaplanabilir. PNN'lerde yoğunluk fonksiyonları Parzen penceresi (Parzen, 1962) adı verilen parametrik olmayan bir kestirim yöntemi ile hesaplanır. Hesaplanan olasılık yoğunluk fonksiyonları ve karşılaşma sıklıklarını kullanan PNN, verilen bir giriş vektörünün en yüksek olasılıkla ait olabileceği sınıfı belirleyebilir. K_i sınıfının j . eğitim vektörü x_j ile gösterilsin. Bu durumda K_i sınıfının olasılık yoğunluk fonksiyonunun Parzen kestirimi şu şekildedir:

$$F_i(x) = \frac{1}{(2\pi)^{p/2} \sigma^m n} \sum_{j=1}^n \exp \left[-\frac{(\mathbf{x} - \mathbf{x}_j)^T (\mathbf{x} - \mathbf{x}_j)}{2\sigma^2} \right] \quad (4.3)$$

Burada n eğitme kümesindeki K_i sınıfına ait eleman sayısı, p giriş vektörünün boyutu, j örüntü numarası ve σ ise bir yumuşatma terimidir (Parzen, 1962, Specht, 1988, 1990a ve 1990b, Goh, 2002). Bu yumuşatma terimine bant genişliği, ya da genişlik adı verilir.

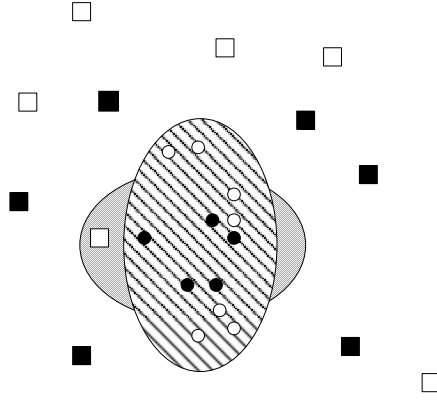
Şekil 4.1. iki sınıflı durum için basit bir PNN yapısını göstermektedir. İlk katman giriş katmanıdır ve p adet giriş değişkenini ($x_1, x_2, \dots, x_p$) temsil etmektedir. Giriş nöronları, girişteki p değişkeni ikinci katmandaki tüm nöronlara dağıtır. Örüntü katmanı olarak adlandırılan ikinci katman ile giriş katmanı tam bağlantılıdır. Bu katmandaki her bir nöron, eğitme kümesindeki bir örneğe karşılık gelmektedir. (4.3) denklemindeki üstellerin toplanması işlemi toplama katmanı içinde gerçekleştirilir. Her bir sınıfın toplama katmanı diğerlerinden bağımsızdır. Toplama katmanının ağırlıkları 1 olarak sabittir. Toplama katmanının çıkışı, örüntü katmanından gelen değerlerin basit toplamından ibarettir. Böylelikle toplama katmanının çıkışında, o anda değerlendirilen giriş vektörünün o katmanın ait olduğu sınıftan gelme olasılığı elde edilmiş olur. Çıkış katmanında ise Bayes Karar Kuralı işletilerek en yüksek olasılık değerine sahip olan sınıfı gösteren ikili bir değer üretilir (Specht, 1988, 1990a ve 1990b, Goh, 2002).



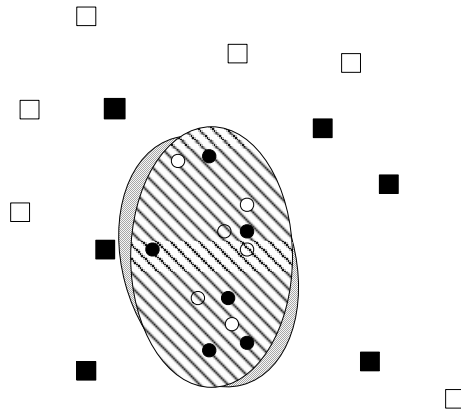
Şekil 4.1. İki sınıflı bir PNN

4.2 Aktif Öğrenme

Şekil 4.1 ve 4.2 ikili bir sınıflama problemini göstermektedir. Sınıf 1 daireler, sınıf 2 ise dikdörtgenler ile gösterilmiştir. Boyalı daire ve dikdörtgenler eğitime kümesine, beyaz daire ve dikdörtgenler ise test kümesine aittir. Soldan sağa taralı alan öğretmenin karar sınırını, sağdan sola taralı alan ise öğrenenin yaklaşımını göstermektedir. Öğrenen ve öğretmenin farklı kararlar verdiği alan, genelleme hatasına karşılık düşmektedir. Örnekte görüldüğü gibi öğrenen tüm eğitime verisini doğru sınıflandırmıştır, fakat öğretmen ve öğrenenin farklı kararlar verdiği bölgeler mevcuttur (Şekil 4.2). Eğitime kümesine yeni örnekler eklendiğinde genelleştirme hatası azalabilmektedir (Şekil 4.3). Bu yeni örneklerin seçimi rasgele yapılabileceği gibi, belli kurallara bağlı olarak da yapılabilir.



Şekil 4.2. Pasif Öğrenme



Şekil 4.3. Aktif öğrenme. Eğitime kümesi öğrenen tarafından seçilmiştir.

Etkin bir aktif öğrenme algoritması, hem genelleştirme hatasını hem de eğitime verisini azaltabilmelidir (RayChaudhuri ve Hamey, 1996, Saar-Tsechansky ve Provost, 2001). Sınıflandırma problemlerinde hedeflenen ortalama karesel hata gibi bir hata terimini en aza indirmek değil, sınıflar arasındaki sınırları (karar sınırı) doğru belirleyebilmektir (Takizawa ve diğerleri, 2000).

Aktif öğrenme stratejileri aktif örnekleme ve aktif seçim olarak iki gruba ayrılabilir. Aktif örneklemede yeni eğitime verileri eski verilerin çeşitli işlemlerden geçirilmesi ile elde edilir. Bu tür algoritmalarda olay uzayından yeni veriler elde etmeye gerek yoktur. Aktif seçim algoritmalarında ise yeni eğitime verileri olay uzayından ve daha önce eğitime verisi içerisinde olmayan örnekler arasından seçilir.

Literatürde çeşitli aktif seçim yaklaşımları mevcuttur. Bu yaklaşımların büyük kısmı iki alt grupta toplanabilir. İlk gruptaki algoritmalar küçük bir eğitime kümesi ile başlar ve kurallı bir şekilde yeni eğitime verilerini eğitime kümesine eklerler. Bu tür algoritmalara artımlı algoritmalar denir. Diğer grupta yer alan algoritmalar ise büyük bir eğitime kümesi ile başlar ve bu kümenin içindeki fazlalık bilgi niteliğindeki örnekleri eğitime kümesinden çıkarırlar. Bu tür algoritmalar ise eksiltmeli algoritmalar olarak adlandırılırlar.

Plutowski ve Halbert (1993) artımlı bir algoritma tasarlamışlardır. Bu algoritmada ortalama karesel hatada maksimum azalmayı sağlayacağına karar verilen yeni bir eğitime verisi eğitime kümesine eklenir ve yapay sinir ağı bu kümeyle yeniden eğitilir. Bir başka artımlı algoritmada (Leisch ve diğerleri, 1998), ağ ilk olarak küçük bir eğitime kümesi ile eğitilir. Eğitime kümesinde yer almayan ve en yüksek hataya sahip örnek x_n tespit edilerek eğitime kümesine eklenir.

Komite ile Sorgulama (Query by committee - QBC) algoritmasında ise birden fazla ağ aynı eğitime kümesi kullanılarak eğitilir. Daha sonra bu ağlara test verisi uygulanır. Oluşturulan komitenin üyeleri MLP türü ağlardır. Hatanın geriye yayılımı algoritmasında başlangıç ağırlıkları raslantısal olarak belirlendiği için, aynı eğitime kümesi ve eğitime parametrelerine sahip olan ağlar birbirlerinden farklı olabilmektedir. Bu şekilde oluşturulan komitenin her bir üyesi aynı örnek üzerinde farklı kararlara varabilirler. Bu duruma anlaşmazlık adı verilir. Üzerinde en fazla anlaşmazlık oluşan örnek, ağ tarafından öğrenilmesi en zor olan örnektir. Öğrenilmesi zor olan bu örnek eğitime verisine eklenerek tüm komite üyeleri yeni oluşturulmuş eğitime kümesi ile tekrar eğitilir ve tüm işlem yenilenir.

RayChaudhuri ve Hamey (1995) QBC algoritmasına benzer yeni bir algoritma önermişlerdir.

Tong ve Koller (2001a) Bayescil ağlarda parametre kestirimine yönelik Kullback-Leibler diverjansına dayalı bir aktif öğrenme yöntemi önermişlerdir. Yöntem, Kullback-Leibler diverjansı temelli bir risk faktörünü kullanmaktadır. Bu yöntemde en düşük risk faktörüne sahip örnek tespit edilerek eğitime verisine eklenmektedir.

Destek Vektör makinası (Support Vector Machine – SVM) doğal işleyişi içinde eğitime kümesini küçülten bir yapıdır (Cortes ve Vapnik, 1995). SVM, sınıflar arasındaki mesafeyi en büyük yapacak olan hiper düzlemleri bulmaya dayalı bir yöntemdir. Destek vektörleri denen küçük bir eğitime kümesi bu hiper düzlemleri temsil etmek için yeterlidir. Bulunan destek vektörleri en yüksek bilgiyi taşıyan eğitime verileridir. Tong ve Koller SVM’yi temel alan bir aktif öğrenme yöntemi geliştirmişlerdir.

Bir başka kullanışlı algoritma da Sıkılana Kadar Tekrarla (Repeat Until Bored – RUB) algoritmasıdır (Munro, 1991). Bu algoritmada baştan belirlenen bir hata eşliğinden daha fazla hataya sebep olan eğitime örneği eğitime kümesi içinde tekrarlanır. Tekrarlama işlemi örneğin hatası eşik değerinin altına düşene kadar sürdürülür. Bu işlem, küçük çocuklarda gözlemlenen en temel doğal öğrenme güdüsünün bir taklididir.

4.3 PNN İle Aktif Öğrenme

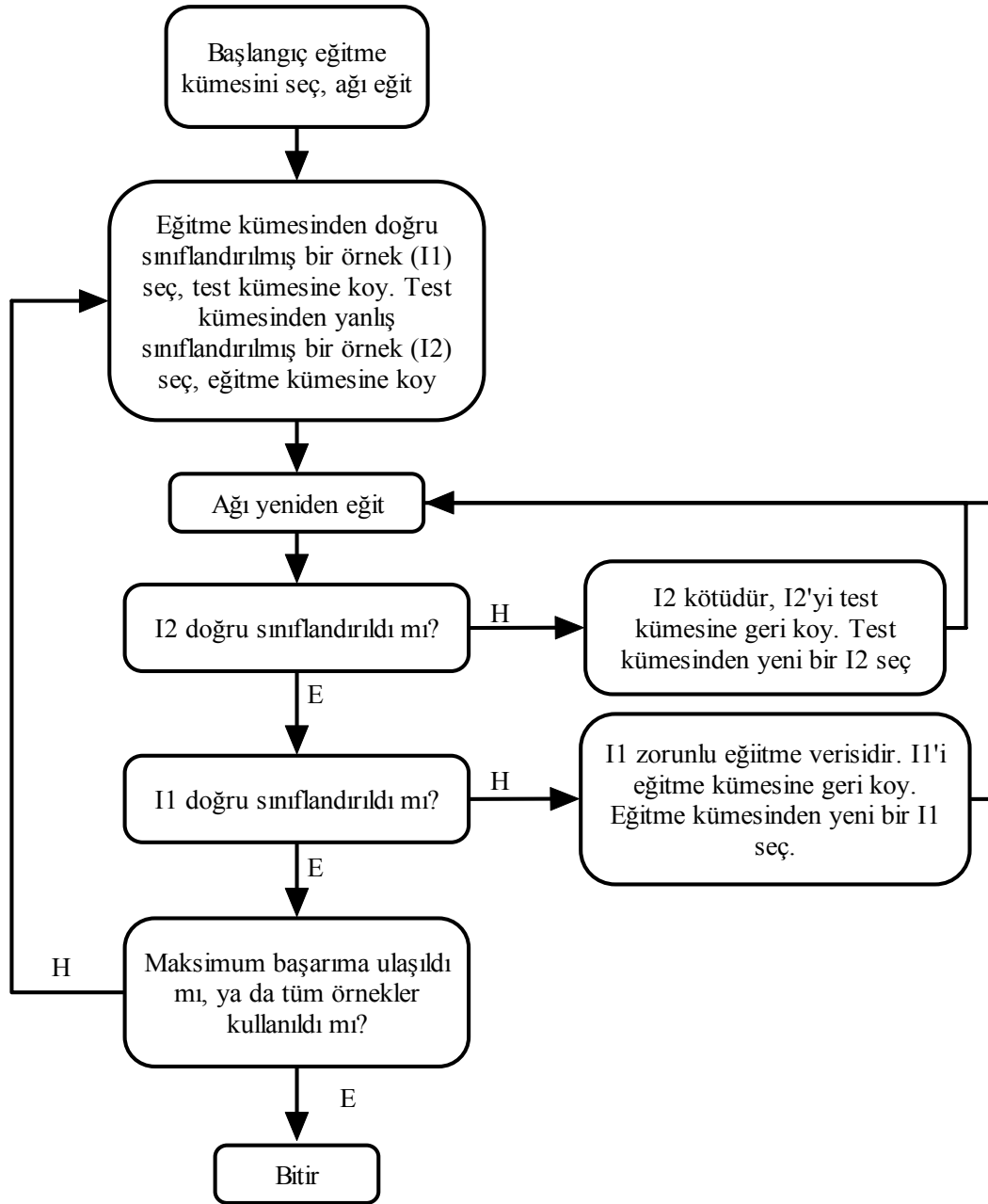
Bilinen aktif öğrenme algoritmalarının bir hata ölçütüne (örneğin ortalama karesel hata, en yüksek mutlak hata, toplam karesel hata vs.) ya da eğitime sürecinin rasgele oluşuna (MLP’nin başlangıç şartlarının rasgele oluşu gibi) ihtiyaç duyması, bu algoritmaların PNN ile birlikte kullanılmasına engel oluşturur. MLP, RBF gibi ağların aksine PNN’in çıkışı ikili kodlanmış sınıf numarası olduğundan çıkışta sürekli bir hata sinyali oluşturulamaz. İkili kodlanmış bir çıkışla yapılan hatanın miktarı ölçülemez. Bu yüzden hata ölçütüne dayalı aktif öğrenme algoritmaları PNN için uygun değildir. Diğer taraftan PNN’in eğitime fazı tek bir adımda tamamlanır ve MLP’nin aksine eğitime sürecinde rasgelelik yoktur. Aynı eğitime kümesi ile eğitilen PNN her seferinde aynı sonucu verir. Bu nedenle, QBC türü algoritmalar da PNN ile birlikte kullanılamaz. Bu nedenlerden dolayı, PNN için yeni bir aktif öğrenme algoritması geliştirilmelidir. PNN ile birlikte kullanılabilecek bir algoritma Bolat ve Yıldırım (Bolat ve Yıldırım, 2003a, 2003b, 2004, 2005) tarafından geliştirilmiştir. Bu algoritma, aktif seçimli bir algoritmadır.

PNN öğrenmesinin ilk adımı uygun bir bant genişliği değerinin belirlenmesidir. Bant genişliği deneme yanılma yoluyla hızlı bir şekilde bulunabilir. En uygun bant genişliği belirlendikten

sonra veri takas algoritması ile daha iyi bir eğitime kümesi belirlenebilir.

Algoritma rasgele seçilen bir eğitime kümesi ile birlikte başlar. İlk eğitime işleminden sonra test kümesi ağı uygulanır. Eğitime kümesinden rasgele seçilen doğru sınıflandırılmış bir örnek (I1) test kümesine konur. Test kümesinden rasgele seçilen hatalı sınıflandırılmış bir örnek (I2) ise eğitime kümesine eklenir. Ağ yeni oluşturulmuş eğitime kümesi ile yeniden eğitilir. İkinci eğitmeden sonra I2 yeni kümesi içinde yanlış sınıflandırılmış ise bu örnek ağ tarafından doğru olarak sınıflandırılmamaktadır. I2 “kötü” olarak adlandırılır ve özgün kümesine geri konarak bir başka hatalı sınıflandırılmış test örneği (I3) eğitime kümesine eklenerek işlem tekrarlanır. Kötü olarak işaretlenmiş olan örnekler algoritma içinde bir daha ele alınmazlar. I2 doğru sınıflandırılmış ise, I1’in durumuna bakılır. Eğer I1 test kümesi içindeyken hatalı sınıflandırılmış ise, eğitime için zorunlu bir örnektir. I1 eğitime kümesine geri konur ve yerine bir başka doğru sınıflandırılmış örnek test kümesine atılarak eğitime tekrarlanır. Eğer yer değiştirme işlemi sonrasında test başarımı artmış ise, bu basamak başarı ile tamamlanmıştır. Algoritma yeni bir I1-I2 çifti için tekrarlanır. Tekrarlama işlemi başarı en yüksek olana kadar, ya da işlenebilir veri kalmayana kadar tekrarlanır (Bolat ve Yıldırım, 2003a, 2003b, 2004, 2005, Bolat ve diğerleri, 2004) (Şekil 4.4).

Algoritmanın son basamağı en yüksek başarıyı veren maliyet fonksiyonlarının bulunmasıdır. Genellikle maliyet fonksiyonları her sınıf için eşit kabul edilir. Ancak, veri kümesindeki sınıflardan biri ya da bir kaç tüm veri kümesinin çok ufak bir bölümünü oluşturuyorsa, maliyet fonksiyonlarının düzgün seçilmesi ağın başarımını arttırmaktadır. Bu çalışmada kullanılan MATLAB Neural Network Aracı’nda yer alan PNN yapısı içerisinde maliyet fonksiyonlarına erişim engellenmiştir. Yine de maliyet fonksiyonu dolaylı olarak değiştirilebilir. Örnek sayısı az olan sınıfların eğitime kümesinde birden fazla kez yer alması, o sınıfın rastlanma sıklığını arttırmaktadır. (4.1) denkleminde de görülebileceği gibi, P_i rastlanma sıklığının arttırılması ile L_i maliyet fonksiyonunun azaltılması eşdeğerdir. Bu işleme yineleme adı verilir. Yineleme, daha iyi bir test başarımı elde edilene kadar tekrarlanır. Kabul edilebilir miktarda yineleme işleminden sonra başarı halen artmıyorsa, yineleme işlemine son verilir (Bolat ve Yıldırım, 2003a, 2003b, 2004, 2005).



Şekil 4.4. PNN ile aktif öğrenme akış şeması

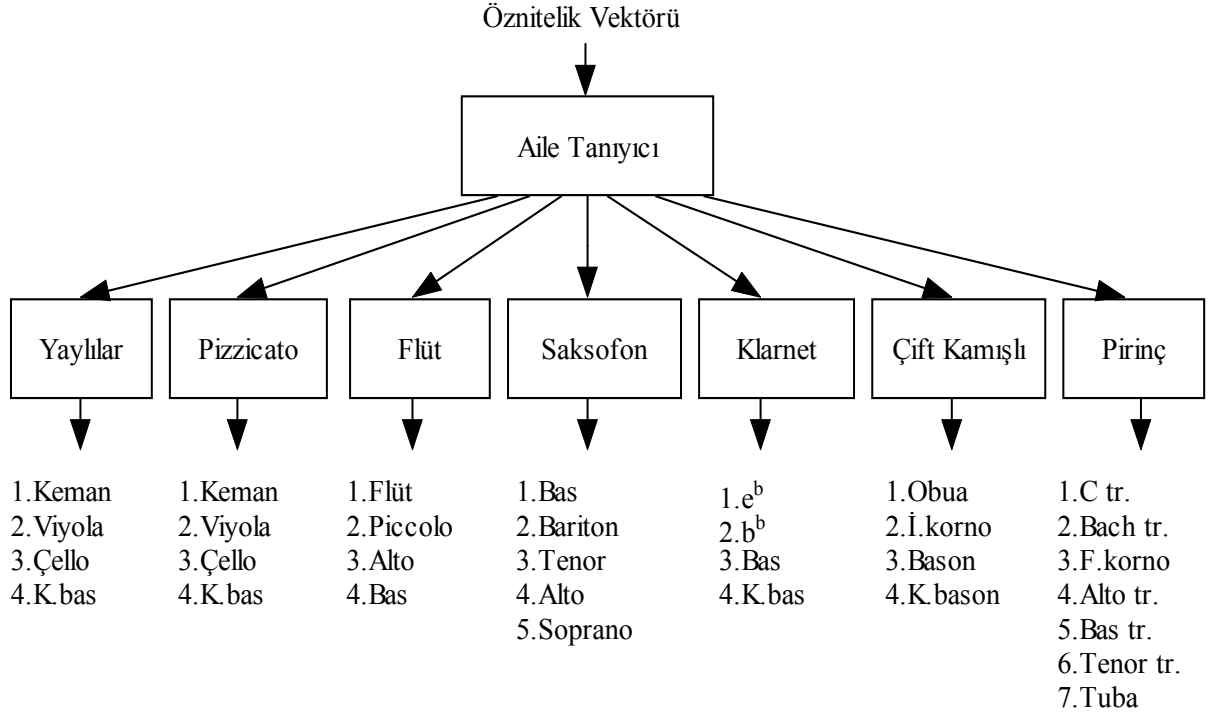
5. UYGULAMA

Uygulamada kullanılan ses örnekleri McGill Üniversitesi'nce hazırlanmış olan McGill University Master CD Samples (MUMS) veri kümesinden alınmıştır. Ses örnekleri 44100 Hz örnekleme frekansında ve 16 bit çözünürlükle kaydedilmiştir. CD üzerindeki her bir iz, bir enstrümanın ürettiği tüm notaları barındırmaktadır. Ses örneklerini bu haliyle işlemek zor olduğundan ilk olarak her bir iz .WAV formatına dönüştürülmüş ve GNU/Linux işletim sistemi altında TAON yazılımı yardımıyla parçalanmıştır. Böylelikle her bir nota ayrı bir .WAV dosyada saklanmıştır. Veri kümesinde 28 farklı enstrümana ait toplam 974 nota mevcuttur. Yaylılar ailesi arşeli ve pizzicato olmak üzere iki ayrı teknikle çalınmıştır. Farklı çalma teknikleri farklı enstrümanlar olarak değerlendirildiğinde toplam enstrüman sayısı 32 olmaktadır.

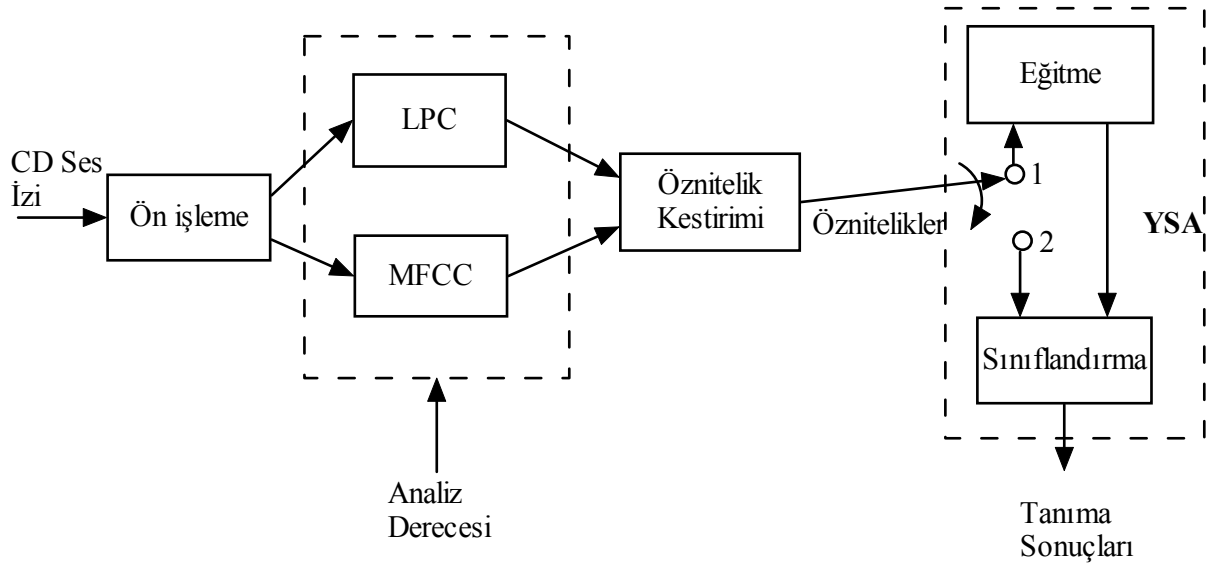
Öznitelik kestirimi yapılırken her bir nota 20 ms uzunluğunda çerçevelere bölünmüş ve her çerçeve için öznitelikler hesaplanarak bir vektöre atanmıştır. Tüm çerçeveler için bu işlem yapıldıktan sonra, bulunan öznitelik vektörlerinin ortalaması alınmış ve böylece her bir nota bir öznitelik vektörü ile temsil edilmiştir.

Tanıma işlemi, tek bir adımda gerçekleştirilebilir. Ancak 32 sınıflı bir veri kümesinin sınıflandırılması yapay sinir ağları için zor bir görevdir. Bu zorluğu aşmanın bir yolu, sınıflandırmayı aşama aşama gerçekleştirmektir. Bu yöntem hiyerarşik sınıflandırma adı verilir. Eronen (2001) ve Martin (1998, 1999) çeşitli sınıflandırma hiyerarşileri önermişlerdir. Hiyerarşik sınıflandırmada her düğüm bir yapay sinir ağına karşılık düşmektedir. PNN diğer ağlara göre çok daha fazla bellek tükettiğinden mümkün olduğu kadar az sayıda yapay sinir ağına ihtiyaç duyacak bir hiyerarşi kullanmak gereklidir. Kurulabilecek en basit hiyerarşik model iki aşamalıdır. İlk aşamada işlenen sesin hangi enstrüman ailesine ait olduğuna karar verilir. İkinci aşamada ise, sesin hangi enstrümana ait olduğu tespit edilir (Şekil 5.1).

Her bir ses örneği için öznitelik olarak kullanılmak üzere farklı derecelerden LPC ve MFCC katsayıları hesaplanmıştır. Hesaplanan öznitelikler kullanılarak bir dizi deneme yapılmıştır. Denemelerin ilk aşamasında pasif bir PNN kullanılmış ve en yüksek başarıımı veren öznitelik grubu belirlenmiştir. Daha sonra, en yüksek başarıımı veren özniteliklerle birlikte aktif öğrenen bir PNN kullanılarak aynı işlemler tekrarlanmış ve aktif öğrenmenin başarıma katkısı araştırılmıştır. Yapılan tüm işlemler, MATLAB 6.5 ortamında gerçekleştirilmiştir (Şekil 5.2).



Şekil 5.1. Hiyerarşik sınıflandırma yapısı



Şekil 5.2 Uygulanan enstrüman tanıma sisteminin blok şeması

5.1 Özniteliklerin Hesaplanması

Yapılan uygulamada öznitelik olarak LP ve MFCC katsayıları kullanılmıştır. Farklı dereceler için hesaplanan katsayıların başarıma etkisini gözleyebilmek amacıyla LP katsayıları 5, 10, 15 ve 20. dereceler için hesaplanırken, MFCC katsayıları 16. dereceden hesaplanmıştır. MFCC katsayılarının yalnızca bir kez hesaplanmış olmasının nedeni, n. dereceden MFCC

katsayılarının $n-1$. dereceden katsayıları içermesidir.

Simülasyon ve hesaplamalar MATLAB 6.5 kullanılarak gerçekleştirilmiştir. LPC katsayıları MATLAB Signal Processing Toolbox içerisinde yer alan betikle hesaplanmış, MFCC katsayıları ise Slaney (1998) tarafından yazılmış olan ücretsiz bir betik yardımıyla hesaplanmıştır.

5.2 Pasif Öğrenen PNN İle Enstrüman Tanıma

Uygulamada kullanılan 32 enstrüman yaylılar (keman, viyola, çello, kontrbas), pizzicato yaylılar (keman, viyola, çello, kontrbas), flütler (flüt, pikolo, alto flüt, bas flüt), klarnetler (si bemol, mi bemol, bas, kontra bas), saksofonlar (bas, bariton, alto, tenor, soprano), piriç nefesliler (do trompet, Bach trompet, Fransız kornosu, alto trompet, tenor trompet, bas trompet, tuba) ve çift kamışlılar (obua, basson, kontra basson, İngiliz kornosu) olmak üzere 7 aileye bölünmüştür. Uygulamanın bu aşamasında aile içi tanıma işlemini gerçekleştiren 7 ayrı yapay sinir ağı tasarlanmıştır. Veri kümesinin yaklaşık %70'i eğitime için ayrılırken kalan kısmı test için kullanılmıştır. Tüm PNN'ler için en uygun bant genişlikleri deneme yanılma yoluyla bulunmuştur. Farklı derecelerden LP katsayıları kullanılarak elde edilen ağlar için eğitime ve test doğrulukları Çizelge 5.1 – 5.4 arasında verilmiştir. Çizelgelerde her aile için elde en yüksek test başarımı koyu renkle belirtilmiştir. En yüksek başarımın elde edildiği durumlar için sınıflandırma sonuçları ise Çizelge 5.5 – 5.11'de gösterilmiştir. Bir sonraki aşama, enstrümanların ailelerine göre sınıflandırılmasıdır. Aile tanıma işleminde kullanılan eğitime kümesi, aile içi tanımada kullanılan eğitime kümeleri birleştirilerek elde edilmiş ve böylelikle en son aşamadaki hiyerarşik sınıflama deneyinde her hangi bir düğümde yer alan yapay sinir ağının eğitime kümesindeki bir verinin bir başka düğümdeki ağın test kümesinde yer alması engellenmiştir. Geri kalan veri, test için ayrılmıştır. Farklı LP dereceleri ile yapılan uygulamalarda elde edilen sonuçlar Çizelge 5.12'de verilmiştir. Öznitelik olarak MFCC kullanıldığında elde edilen sonuçlar Çizelge 5.13–5.19'da verilmiştir. En yüksek başarımın elde edildiği durumlar için test sonuçları Çizelge 5.20–5.26 arasında yer almaktadır. Farklı derecelerden MFCC'ler kullanılarak yapılan aile tanıma deneylerine ilişkin sonuçlar Çizelge 5.27'de sunulmuştur. Farklı derecelerden LPC ve MFCC kullanılarak yapılan hiyerarşik sınıflama denemelerinden elde edilen sonuçlar ise Çizelge 5.28 ve 5.29'da gösterilmiştir.

Çizelge 5.1. 20. dereceden LP katsayıları ile aile içi doğru sınıflandırma yüzdeleri

	Yaylı	Pizzicato	Klarnet	Çift Kamışlı	Saksofon	Flüt	Pirinç Nefesli
Eğitme	%100	%100	%91,67	%100	%100	%91,86	%100
Test	%76,92	%65,96	%74,29	%61,54	%52,94	%56,76	%52,38

Çizelge 5.2. 15. dereceden LP katsayıları ile aile içi doğru sınıflandırma yüzdeleri

	Yaylı	Pizzicato	Klarnet	Çift Kamışlı	Saksofon	Flüt	Pirinç Nefesli
Eğitme	%100	%100	%95,24	%100	%97,37	%100	%100
Test	%76,92	%63,83	%71,43	%74,36	%52,94	%56,76	%47,62

Çizelge 5.3. 10. dereceden LP katsayıları ile aile içi doğru sınıflandırma yüzdeleri

	Yaylı	Pizzicato	Klarnet	Çift Kamışlı	Saksofon	Flüt	Pirinç Nefesli
Eğitme	%100	%100	%94,05	%100	%94,73	%94,19	%100
Test	%71,54	%65,96	%82,86	%74,36	%76,47	%64,86	%46,03

Çizelge 5.4. 5. dereceden LP katsayıları ile aile içi doğru sınıflandırma yüzdeleri

	Yaylı	Pizzicato	Klarnet	Çift Kamışlı	Saksofon	Flüt	Pirinç Nefesli
Eğitme	%98,41	%96,3	%100	%97,7	%89,47	%97,67	%100
Test	%55,77	%53,19	%77,14	%48,72	%52,94	%62,16	%36,5

Çizelge 5.5. 15. derece LP katsayıları kullanıldığında Yaylılar ailesi için aile içi sınıflandırma test kümesi sonuçları

		Ağın Kararı			
		Keman	Viyola	Çello	Kontra bas
Ağa uygulanan	Keman (13)	13			
	Viyola (12)		12		
	Çello (14)	1	6	7	
	Kontra bas (13)	1	4		8

Çizelge 5.6. 20. derece LP katsayıları kullanıldığında Piriñ Nefesliler ailesi için aile içi sınıflandırma test kümesi sonuçları

		Ağın Kararı						
		C Trompet	Bach Trompet	Fransız Kornosu	Alto Trompet	Tenor Trompet	Bas Trompet	Tuba
Ağa uygulanan	C Trompet (10)	10						
	Bach Tr. (10)	10	0					
	Fr. Korno (11)			7				4
	Alto Tr. (4)			2	1	1		
	Tenor Tr. (11)	1		5		5		
	Bas Tr. (7)				1	6	0	
	Tuba (10)							10

Çizelge 5.7. 15. derece LP katsayıları kullanıldığında Pizzicato ailesi için aile içi sınıflandırma test kümesi sonuçları

		Ağın Kararı			
		Keman	Viyola	Çello	Kontra bas
Ağa uygulanan	Keman (12)	11			1
	Viyola (10)	2	8		
	Çello (12)	2	1	9	
	Kontra bas (13)	1		9	3

Çizelge 5.8. 10. derece LP katsayıları kullanıldığında Klarnet ailesi için aile içi sınıflandırma test kümesi sonuçları

		Ağın Kararı			
		b ^b Klarnet	e ^b Klarnet	Bas Klarnet	Kontrbas Klarnet
Ağa uygulanan	b ^b Klarnet (11)	10	1		
	e ^b Klarnet (10)	2	7		1
	Bas Klarnet (7)			6	1
	Kontrbas Klarnet (7)			1	6

Çizelge 5.9. 10. derece LP katsayıları kullanıldığında Saksofon ailesi için aile içi sınıflandırma test kümesi sonuçları

		Ağın Kararı				
		Bas Saks.	Bariton Saks.	Tenor Saks.	Alto Saks.	Soprano Saks.
Ağa uygulanan	Bas Saks. (1)	1				
	Bariton Saks. (4)		4			
	Tenor Saks. (4)		2	2		
	Alto Saks. (4)			1	3	
	Soprano Saks. (4)			1		3

Çizelge 5.10. 10. derece LP katsayıları kullanıldığında Çift Kamışlılar ailesi için aile içi sınıflandırma test kümesi sonuçları

		Ağın Kararı			
		Obua	İngiliz Kornosu	Bason	Kontra bason
Ağa uygulanan	Obua(10)	7	3		
	İ. Kornosu (9)	4	5		
	Bason (10)			10	
	K. Bason (10)			3	7

Çizelge 5.11. 10. derece LP katsayıları kullanıldığında Flüt ailesi için aile içi sınıflandırma test kümesi sonuçları

		Ağın Kararı			
		Flüt	Piccolo	Alto Flüt	Bas Flüt
Ağa uygulanan	Flüt (11)	4	3	4	
	Piccolo (9)		8	1	
	Alto Flüt (9)	2		7	
	Bas Flüt (8)	1		2	5

Çizelge 5.12. LP derecesinin enstrüman ailelerinin tanınmasına etkisi

	LPC 5	LPC 10	LPC 15	LPC 20
Eğitme	%100	%99,85	%99,71	%99,85
Test	%51,56	%54,33	%61,59	%58,13

Çizelge 5.13. 4. dereceden MFCC katsayıları ile aile içi doğru sınıflandırma yüzdeleri

	Yaylı	Pizzicato	Flüt	Klarnet	Saksofon	Pirinç Nf.	Çift Kmş.
Eğitme	%97,62	%89,82	%100	%88,1	%100	%100	%100
Test	%59,62	%65,96	%45,95	%80	%70,59	%63,49	%74,36

Çizelge 5.14. 6. dereceden MFCC katsayıları ile aile içi doğru sınıflandırma yüzdeleri

	Yaylı	Pizzicato	Flüt	Klarnet	Saksf.	Pirinç Nf.	Çift Kmş.
Eğitme	%98,41	%98,15	%100	%84,52	%100	%100	%100
Test	%65,39	%68,09	%43,24	%88,57	%76,47	%61,91	%74,36

Çizelge 5.15. 8. dereceden MFCC katsayıları ile aile içi doğru sınıflandırma yüzdeleri

	Yaylı	Pizzicato	Flüt	Klarnet	Saksofon	Pirinç Nefesli	Çift Kamışlı
Eğitme	%100	%96,3	%100	%84,52	%95,75	%100	%100
Test	%65,39	%61,7	%51,35	%74,29	%76,47	%61,91	%71,8

Çizelge 5.16. 10. dereceden MFCC katsayıları ile aile içi doğru sınıflandırma yüzdeleri

	Yaylı	Pizzicato	Flüt	Klarnet	Saksofon	Pirinç Nefesli	Çift Kamışlı
Eğitme	%100	%100	%100	%86,91	%100	%100	%100
Test	%63,46	%53,19	%48,65	%71,43	%88,24	%60,32	%74,36

Çizelge 5.17. 12. dereceden MFCC katsayıları ile aile içi doğru sınıflandırma yüzdeleri

	Yaylı	Pizzicato	Flüt	Klarnet	Saksofon	Pirinç Nefesli	Çift Kamışlı
Eğitme	%100	%100	%100	%86,91	%100	%100	%100
Test	%63,46	%51,06	%48,65	%68,57	%82,35	%60,32	%71,8

Çizelge 5.18. 14. dereceden MFCC katsayıları ile aile içi doğru sınıflandırma yüzdeleri

	Yaylı	Pizzicato	Flüt	Klarnet	Saksofon	Pirinç Nefesli	Çift Kamışlı
Eğitme	%100	%100	%100	%84,52	%100	%100	%94,25
Test	%55,77	%57,45	%54,05	%71,43	%76,47	%58,73	%71,8

Çizelge 5.19. 16. dereceden MFCC katsayıları ile aile içi doğru sınıflandırma yüzdeleri

	Yaylı	Pizzicato	Flüt	Klarnet	Saksofon	Pirinç Nefesli	Çift Kamışlı
Eğitme	%97,62	%100	%100	%100	%100	%100	%100
Test	%50	48,94	%40,54	%54,29	%82,35	%57,14	%71,8

Çizelge 5.20. 8. derece MFCC katsayıları kullanıldığında Yaylılar ailesi için aile içi sınıflandırma test kümesi sonuçları

		Ağın Kararı			
		Keman	Viyola	Çello	Kontra bas
Ağa uygulanan	Keman (13)	13			
	Viyola (12)		12		
	Çello (14)	8	4	2	
	Kontra bas (13)	3	2	1	7

Çizelge 5.21. 6. derece MFCC katsayıları kullanıldığında Pizzicato ailesi için aile içi sınıflandırma test kümesi sonuçları

		Ağın Kararı			
		Keman	Viyola	Çello	Kontra bas
Ağa uygulanan	Keman (12)	12			
	Viyola (10)	4	6		
	Çello (12)		2	10	
	Kontra bas (13)			9	4

Çizelge 5.22. 6. derece MFCC katsayıları kullanıldığında Pirinç Nefesliler ailesi için aile içi sınıflandırma test kümesi sonuçları

		Ağın Kararı						
		C Trompet	Bach Trompet	Fransız Kornosu	Alto Trompet	Tenor Trompet	Bas Trompet	Tuba
Ağa uygulanan	C Trompet (10)	3	7					
	Bach Tr. (10)		10					
	Fr. Korno (11)			6		5		
	Alto Tr. (4)	1			3			
	Tenor Tr. (11)				7	4		
	Bas Tr. (7)			2		2	3	
	Tuba (10)							10

Çizelge 5.23. 6. derece MFCC katsayıları kullanıldığında Klarnet ailesi için aile içi sınıflandırma test kümesi sonuçları

		Ağın Kararı			
		b ^b Klarnet	e ^b Klarnet	Bas Klarnet	Kontrbas Klarnet
Ağa uygulanan	b ^b Klarnet (11)	9	2		
	e ^b Klarnet (10)		10		
	Bas Klarnet (7)			7	
	Kontrbas Klarnet (7)			2	5

Çizelge 5.24. 6. derece MFCC katsayıları kullanıldığında Çift Kamışlılar ailesi için aile içi sınıflandırma test kümesi sonuçları

		Ağın Kararı			
		obua	İngiliz Kornosu	Bason	Kontra bason
Ağa uygulanan	Obua(10)	10			
	İ. Kornosu (9)	7	2		
	Bason (10)		3	7	
	K. Bason (10)				10

Çizelge 5.25. 10. derece MFCC katsayıları kullanıldığında Saksofon ailesi için aile içi sınıflandırma test kümesi sonuçları

		Ağın Kararı				
		Bas Saks.	Bariton Saks.	Tenor Saks.	Alto Saks.	Soprano Saks.
Ağa uygulanan	Bas Saks. (1)	1				
	Bariton Saks. (4)		3	1		
	Tenor Saks. (4)			4		
	Alto Saks. (4)				3	1
	Soprano Saks. (4)					4

Çizelge 5.26. 14. derece MFCC katsayıları kullanıldığında Flüt ailesi için aile içi sınıflandırma test kümesi sonuçları

		Ağın Kararı			
		Flüt	Piccolo	Alto Flüt	Bas Flüt
Ağa uygulanan	Flüt (11)	3	8		
	Piccolo (9)		9		
	Alto Flüt (9)	5	1	3	
	Bas Flüt (8)	1		2	5

Çizelge 5.27. Farklı derecelerden MFCC katsayıları ile aile tanıma yüzdeleri

	MF 4	MF 6	MF 8	MF 10	MF 12	MF 14	MF 16
Eğitme	95,76%	89,33%	92,10%	100%	100%	100%	96,64%
Test	56,55%	63,10%	62,76%	61,38%	58,97%	56,20%	54,82%
Toplam	84,08%	81,52%	83,37%	88,50%	87,82%	86,96%	84,19%

Çizelge 5.28. Farklı derecelerden LPC'ler için hiyerarşik tanıma yüzdeleri

	LPC 5	LPC 10	LPC 15	LPC 20
Eğitme	%97,95	%97,81	%97,08	%93,28
Test	%30,34	%37,24	%36,55	%36,21

Çizelge 5.29. Farklı derecelerden MFCC'ler için hiyerarşik tanıma yüzdeleri

	MFCC4	MFCC6	MFCC8	MFCC10	MFCC12	MFCC14	MFCC16
Eğitme	%92,25	%87,28	%89,62	%98,39	%98,39	%97,36	%94,44
Test	%34,83	%40,69	%40,35	%38,28	%35,17	%33,79	%28,37

5.3 Aktif Öğrenen PNN İle Enstrüman Tanıma

Çalışmada aktif öğrenme kullanılmasının asıl nedeni, başarıyı daha yüksek bir yapay sinir ağı yapısı geliştirmektir. Bu nedenle aktif öğrenme denemelerinde yalnızca pasif öğrenen PNN ile yapılan uygulamalar sonucunda en yüksek başarıyı veren öz nitelikler ele alınmıştır. Tanıma işleminde kullanılan asıl tanıyıcı mekanizma hiyerarşik sınıflayıcıdır. Aile ve aile içi tanıma için kullanılan yapay sinir ağları ise yalnızca ara işlem basamaklarını oluşturmaktadır. Bu nedenle, aktif öğrenmede yalnızca 10. derece LPC ve 6. derece MFCC'ler kullanılmıştır. Aktif öğrenme, aile içi sınıflandırıcılara uygulanmış ve her aile için elde edilen sonuç eğitme kümeleri birleştirilerek aile sınıflandırmada kullanılan eğitme kümesi elde edilmiştir. Çizelge 5.30, aktif öğrenen PNN ve 10. derece LPC ile elde edilen aile içi sınıflandırma doğruluklarını göstermektedir. Çizelge 5.31–5.36'da aile içi sınıflandırma test sonuçları yer almaktadır. 6. derece MFCC ile yapılan aile içi doğru sınıflandırma yüzdeleri Çizelge 5.37'de verilmiştir. Aile içi test sonuçları ise Çizelge 5.38–5.43'te yer almaktadır. 10. derece LPC ve 6. derece MFCC katsayıları için aile tanıma sonuçları Çizelge 5.44'te gösterilmiştir.

Çizelge 5.30. Aktif Öğrenen PNN kullanıldığında 10. dereceden LP katsayıları ile aile içi doğru sınıflandırma yüzdeleri

	Yaylı	Pizzicato	Klarnet	Çift Kamışlı	Saksofon	Flüt	Pirinç Nefesli
Eğitme	%100	%100	%93,65	%100	%95,74	%94,19	%100
Test	%96,15	%91,49	%97,14	%100	%94,12	%78,38	%68,25

Çizelge 5.31. Aktif Öğrenen PNN ile 10. derece LP katsayıları kullanıldığında Yaylılar ailesi için aile içi sınıflandırma test kümesi sonuçları

		Ağın Kararı			
		Keman	Viyola	Çello	Kontra bas
Ağa uygulanan	Keman (13)	13			
	Viyola (12)		12		
	Çello (14)		1	13	
	Kontra bas (13)			1	12

Çizelge 5.32. Aktif Öğrenen PNN ile 10. derece LP katsayıları kullanıldığında Pizzicato ailesi için aile içi sınıflandırma test kümesi sonuçları

		Ağın Kararı			
		Keman	Viyola	Çello	Kontra bas
Ağa uygulanan	Keman (12)	11		1	
	Viyola (10)		10		
	Çello (12)	1		11	
	Kontra bas (13)	1		1	11

Çizelge 5.33. Aktif Öğrenen PNN ile 10. derece LP katsayıları kullanıldığında Piring Nefesliler ailesi için aile içi sınıflandırma test kümesi sonuçları

		Ağın Kararı						
		C Trompet	Bach Trompet	Fransız Kornosu	Alto Trompet	Tenor Trompet	Bas Trompet	Tuba
Ağa uygulanan	C Trompet (10)	10						
	Bach Tr. (10)	10	0					
	Fr. Korno (11)			11				
	Alto Tr. (4)	1		1	2			
	Tenor Tr. (11)			3		8		
	Bas Tr. (7)	4				1	2	
	Tuba (10)							10

Çizelge 5.34. Aktif Öğrenen PNN ile 10. derece LP katsayıları kullanıldığında Klarnet ailesi için aile içi sınıflandırma test kümesi sonuçları

		Ağın Kararı			
		b ^b Klarnet	e ^b Klarnet	Bas Klarnet	Kontrbas Klarnet
Ağa uygulanan	b ^b Klarnet (11)	11			
	e ^b Klarnet (10)	1	9		
	Bas Klarnet (7)			7	
	Kontrbas Klarnet (7)				7

Çizelge 5.35. Aktif Öğrenen PNN ile 10. derece LP katsayıları kullanıldığında Flüt ailesi için aile içi sınıflandırma test kümesi sonuçları

		Ağın Kararı			
		Flüt	Piccolo	Alto Flüt	Bas Flüt
Ağa uygulanan	Flüt (11)	8		3	
	Piccolo (9)		9		
	Alto Flüt (9)	2		7	
	Bas Flüt (8)	1		1	6

Çizelge 5.36. Aktif Öğrenen PNN ile 10. derece LP katsayıları kullanıldığında Saksofon ailesi için aile içi sınıflandırma test kümesi sonuçları

		Ağın Kararı				
		Bas Saks.	Bariton Saks.	Tenor Saks.	Alto Saks.	Soprano Saks.
Ağa uygulanan	Bas Saks. (1)	1				
	Bariton Saks. (4)		4			
	Tenor Saks. (4)		1	3		
	Alto Saks. (4)				4	
	Soprano Saks. (4)					4

Çizelge 5.37. 6. dereceden MFCC katsayıları kullanıldığında aktif öğrenen PNN ile aile içi doğru sınıflandırma yüzdeleri

	Yaylı	Pizzicato	Flüt	Klarnet	Saksofon	Pirinç Nefesli	Çift Kamışlı
Eğitme	%98,41	%98,15	%100	%84,52	%100	%100	%100
Test	%100	%89,36	%100	%91,43	%88,24	%85,71	%94,87

Çizelge 5.38. Aktif Öğrenen PNN ile 6. derece MFCC katsayıları kullanıldığında Pirinç Nefesliler ailesi için aile içi sınıflandırma test kümesi sonuçları

		Ağın Kararı						
		C Trompet	Bach Trompet	Fransız Kornosu	Alto Trompet	Tenor Trompet	Bas Trompet	Tuba
Ağa uygulanan	C Trompet (10)	7	3					
	Bach Tr. (10)		10					
	Fr. Korno (11)			10		1		
	Alto Tr. (4)				4			
	Tenor Tr. (11)			3		7	1	
	Bas Tr. (7)			1			6	
	Tuba (10)							10

Çizelge 5.39. Aktif Öğrenen PNN ile 6. derece MFCC katsayıları kullanıldığında Saksofon ailesi için aile içi sınıflandırma test kümesi sonuçları

		Ağın Kararı				
		Bas Saks.	Bariton Saks.	Tenor Saks.	Alto Saks.	Soprano Saks.
Ağa uygulanan	Bas Saks. (1)	1				
	Bariton Saks. (4)		4			
	Tenor Saks. (4)			4		
	Alto Saks. (4)				2	2
	Soprano Saks. (4)					4

Çizelge 5.40. Aktif Öğrenen PNN ile 6. derece MFCC katsayıları kullanıldığında Pizzicato ailesi için aile içi sınıflandırma test kümesi sonuçları

		Ağın Kararı			
		Keman	Viyola	Çello	Kontra bas
Ağa uygulanan	Keman (12)	12			
	Viyola (10)	1	9		
	Çello (12)			10	2
	Kontra bas (13)			2	10

Çizelge 5.41. Aktif Öğrenen PNN ile 6. derece MFCC katsayıları kullanıldığında Klarnet ailesi için aile içi sınıflandırma test kümesi sonuçları

		Ağın Kararı			
		b ^b Klarnet	e ^b Klarnet	Bas Klarnet	Kontrbas Klarnet
Ağa uygulanan	b ^b Klarnet (11)	10	1		
	e ^b Klarnet (10)	1	9		
	Bas Klarnet (7)			7	
	Kontrbas Klarnet (7)			1	6

Çizelge 5.42. Aktif Öğrenen PNN ile 6. derece MFCC katsayıları kullanıldığında Flüt ailesi için aile içi sınıflandırma test kümesi sonuçları

		Ağın Kararı			
		Flüt	Piccolo	Alto Flüt	Bas Flüt
Ağa uygulanan	Flüt (11)	8		3	
	Piccolo (9)		9		
	Alto Flüt (9)	2		7	
	Bas Flüt (8)	1		1	6

Çizelge 5.43. Aktif Öğrenen PNN ile 6. derece MFCC katsayıları kullanıldığında Çift Kamışlılar ailesi için aile içi sınıflandırma test kümesi sonuçları

		Ağın Kararı			
		Obua	İngiliz Kornosu	Bason	Kontra bason
Ağa uygulanan	Obua(10)	10			
	İ. Kornosu (9)	2	7		
	Bason (10)			10	
	K. Bason (10)				10

Çizelge 5.44. Aktif Öğrenen PNN ile aile tanıma sonuçları

	LPC Eğitim	LPC Test	MFCC Eğitim	MFCC Test
Yaylı	%100	%92,23	%94,44	%88,46
Pizzicato	%100	%63,83	%100	%89,36
Flüt	%100	%54,05	%89,54	%97,3
Klarnet	%100	%65,71	%86,9	%40
Saksofon	%100	%52,94	%34,04	%17,65
Pirinç Nefesli	%99,31	%34,92	%97,95	%73,02
Çift Kamışlı	%100	%64,1	%94,25	%41,03

6. SONUÇLAR

LPC'ler ile yapılan aile içi tanıma denemelerinde, en yüksek başarımlar klarnet ailesinde elde edilirken, en düşük başarıma sahip enstrüman ailesi pirinç nefesliler olmuştur. LP derecesinin aile içi tanıma başarımına etkisine bakıldığında, LP derecesinin artmasıyla birlikte başarımın önce artmaya, bir noktada sonra ise düşmeye başladığı görülmektedir. Bu durum, Eronen tarafından da gözlemlenmiştir. Başarımdaki düşüşün iki önemli nedeni vardır. LPC derecesinin artması, öznitelik vektörünün taşıdığı gereksiz bilgi miktarını arttırmaktadır. Artan gereksiz bilgi miktarı ile birlikte ağırlı hata yapma olasılığı artmaktadır. Bunun dışında, öznitelik vektörünün boyutu istatistiksel ağların zayıf noktasıdır. Öznitelik vektörünün çok büyük olduğu durumlarda istatistiksel ağların başarımı düşmektedir. Genel ortalamaya bakıldığında aile içi tanıma denemelerinde 10. derece LPC katsayıları en yüksek başarımları vermektedir.

Aile tanıma denemesinde en yüksek başarımlar 15. derece LPC katsayıları ile elde edilmiştir. Birleştirilmiş denemede ise, en yüksek başarımlar 10. derece LPC katsayıları ile %37,24 olarak elde edilmiştir. Toplam başarımlar ise %79,74 olarak bulunmuştur. Buradan, aile içi başarımın genel başarıma etkisinin aile tanıma başarımından daha fazla olduğu söylenebilir, ancak bu önermenin doğruluğunun denetlenmesi gerekmektedir.

MFCC katsayıları ile yapılan aile içi tanıma denemelerinde LPC ile yapılan denemelerden biraz daha iyi sonuçlar elde edilmiştir. En yüksek ortalama başarımlar 6. derece katsayılarla elde edilmiştir. Aile tanıma denemelerinde ise en yüksek başarımlar yine 6. dereceden katsayılarla elde edilmiştir. MFCC katsayıları ile enstrüman tanıma başarımı, %40,69 olarak bulunmuştur. Bulunan bu değer, geçmişte yapılmış çalışmalar ile yarışabilecek düzeydedir. Ancak, bu çalışmada kullanılan hiyerarşik ağın önceki çalışmalara göre daha basit olduğu da unutulmamalıdır. Eronen'in yaklaşık olarak aynı veri kümesi ve MFCC katsayıları kullanarak yaptığı çalışmada (Eronen, 2001) %32 başarımlar elde ettiği göz önüne alındığında, olasılıksal sinir ağının bu problem için daha uygun bir ağ yapısı olduğu söylenebilir.

Aktif öğrenen PNN ile yapılan denemelerde, 10. derece LPC katsayıları kullanıldığında aile içi doğru sınıflandırma yüzdeleri yükselmiştir. Örneğin, çift kamışlılar ailesi tamamen doğru sınıflandırılabilmiştir. Aktif öğrenme algoritmasının kullanılması ile birlikte, 10. derece LPC katsayıları için hiyerarşik sınıflandırma test başarımı %54,14'e yükselirken, toplam başarımlar %85,01 olmuştur.

Aktif öğrenen PNN'e öznitelik olarak 6. derece MFCC katsayıları uygulandığında, aile içi

tanıma deneyi sonucunda flüt ailesi tamamen doğru sınıflandırılırken, yaylı ailesi de test aşamasında tamamen doğru sınıflandırılmıştır. Hiyerarşik sınıflandırma sonucunda %88,3 eğitme başarımına erişilirken, test başarımı ise %65,17 olarak bulunmuştur. Toplam başarım ise %81,42 olarak gerçekleşmiştir.

Aktif öğrenme ile elde edilen sonuçlar, önceki çalışmalarla rekabet edebilecek durumdadır. Test başarımı, geçmiş çalışmalardan daha yüksektir. En yüksek başarımın elde edildiği iki çalışma Eronen (2001) ve Martin'e (1998) aittir. Bu iki çalışmadaki test başarımları sırasıyla %39 ve %37'dir. Her iki çalışmada da daha az sayıda enstrüman kullanılmıştır. Hiyerarşik sınıflandırıcılar daha karmaşıktır. Bunlara ek olarak, Martin'in çalışmasında yüzün üzerinde öznitelik kullanılmıştır. Gerçekleştirilen çalışmada daha basit bir hiyerarşi ile, daha basit öznitelikler yardımıyla daha fazla enstrüman daha yüksek bir başarımla sınıflandırılmıştır. Enstrüman tanıma işlemine ait test kümesi sonuçları Çizelge 6.1'de verilmiştir. Çizelgede sütunlar ağın verdiği yanıtı, satırlar ise ağa uygulanan girişleri göstermektedir. Koyu yazılmış olan sayılar satırda gösterilen enstrümanın ağ tarafından kaç kez doğru sınıflandırıldığını, açık yazılmış sayılar ise bu enstrümanın kaç kez yanlış sınıflandırıldığını göstermektedir. Çift çizgi ile kare içine alınmış olan sayı grupları, aynı aile içindeki enstrümanları göstermektedir.

Elde edilen sonuçlar Martin'in insan deneklerle yaptığı deneyle (Martin, 1998) karşılaştırıldığında (enstrüman tanıma %45,9; aile tanıma %91,7), gerçekleştirilen sistemin enstrüman tanımada daha başarılı olmasına (6. derece MFCC ile toplam %81,42) rağmen aile tanımada daha başarısız olduğu (6. derece MFCC ile toplam %84,3) görülecektir. Bu durum daha karmaşık aile tanıma sistemleri kullanılarak giderilebilir. Örneğin Martin (Martin, 1998) ya da Eronen'in (Eronen, 2001) kullandığı hiyerarşiler insan başarımından daha yüksek aile tanıma yüzdelerine erişmiştir. Ancak daha karmaşık hiyerarşilerin kullanılması ile birlikte sistemin içerdiği YSA sayısı, buna bağlı olarak da toplam bellek gereksinimi hızla artmaktadır. PNN, yapısı gereği diğer YSA modellerine göre çok fazla bellek tüketmektedir. Basit bir örnek vermek gerekirse, bir MLP eğitme verisinin boyundan bağımsız olarak yalnızca içerdiği ağırlıklar kadar belleğe ihtiyaç duyarken, PNN eğitme kümesindeki tüm verileri saklamak zorundadır. Aşırı bellek tüketimi tasarlanan sistemin bir yetersizliği olarak değerlendirilebilir. Hiyerarşinin bazı düğümlerinde bellek gereksinimi daha düşük ağlar kullanmak bu sorunun aşılmasında bir yöntem olarak önerilebilir.

Çizelge 6.1. Aktif Öğrenen PNN ile enstrüman tanıma test sonuçları

		Ağın Kararı															
		1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16
		keman	viyola	çello	kontra bas	keman	viyola	çello	kontra bas	flüt	piccolo	alto flüt	bas flüt	bb klarnet	eb klarnet	bas klarnet	kontrbas k
Ağa Uygulanan	1	11									2						
	2		10								2						
	3			12													
	4				13												
	5					12											
	6					1	9										
	7				1			8	1								
	8							2	9								
	9									11							
	10										9						
	11											9					
	12												7				
	13													1			
	14	2								1		1			0		
	15															6	
	16															1	6
	17				1												
	18																4
	19		3														
	20	2															
	21	1								1							
	22																
	23	1								4							
	24					2											
	25					1											
	26																
	27																
	28				1	5	2										
	29																
	30	1								2							
	31									1							
	32												2				

Çizelge 6.1. (Devam) Aktif Öğrenen PNN ile enstrüman tanıma test sonuçları

		Ağın Kararı															
		17	18	19	20	21	22	23	24	25	26	27	28	29	30	31	32
		bas sax	bariton sax	tenor sax	alto sax	soprano sax	c trompet	Bach tr.	Fr. Korno	alto tr.	tenor tr.	bas tr.	tuba	obua	E. Korno	bason	k. bason
Ağa Uygulanan	1																
	2																
	3			2													
	4																
	5																
	6																
	7												2				
	8												2				
	9																
	10																
	11																
	12																1
	13						4	4							2		
	14					1	2	3									
	15			1													
	16																
	17	0															
	18		0														
	19			1													
	20				2												
	21					0	1	1									
	22						6	3						1			
	23							5									
	24								8		1						
	25									3							
	26							3			7	1					
	27							1				6					
	28												2				
	29						6	2						2			
	30					2	1	2							1		
	31								4							5	
	32																8

Çalışmada sunulan sistemin en önemli zayıflığı, yalnızca monofonik ses kayıtlarını işleyebilmesidir. Her ne kadar ticari master kayıtlar her bir enstrüman için farklı kanallardan kayıt yapıyor olsa da, bu kayıtlar CD haline getirilip piyasaya sürüldüğünde tüm kanallar birleştirilerek stereo haline getirilmektedir. Dolayısı ile gerçekleştirilen çalışma gerçek

dünyadan gelen problemler için bir çözüm üretememektedir. Polifonik seslerin tanınması üzerine yapılacak olan çalışmaların gerçek dünyadan gelen verilerin değerlendirilmesinde daha gerçekçi sonuçlar vereceği mutlaktır. Ancak bu tür çalışmaların gerçekleştirilebilmesi için, öncelikle insan duyma sisteminin karmaşık bir arka plan arasından tek bir kaynağa ait sesleri nasıl ayırttığına dair bir teori kurmak gereklidir. Ayrıca, polifonik seslerin tanınmasında insan başarımının ne olduğu da halen araştırılmış değildir.

Son iki paragrafta bahsedilen konular dışında geleceğe yönelik çalışmalar olarak önerilebilecek bir başka konu da, başarımı daha yüksek öznelilikler keşfetmektir. Daha önce hiç kullanılmamış öznelilikler aramak yanında, mevcut özneliliklerin ses sinyalinin tümü yerine daha küçük parçaları (örneğin sadece atak bölgesinde) için hesaplanması da ileride yapılması planlanan çalışmalar arasında sayılabilir. Ayrıca, farklı YSA modellerinin başarıma etkisi de değerlendirilmelidir.

KAYNAKLAR

- Berenzweig, A. L. Ve Ellis, D. P. W., (2001) "Locating Singing Voice Segments within Music Signals", IEEE Workshop on Applications of Signal Processing to Acoustics and Audio, October 2001, NewYork, 119-122.
- Berger, K. W., (1964), "Some Factors in the Recognition of Timbre", Journal of the Audio Engineering Society, 30, 396-406.
- Bolat, B. ve Yıldırım, T., (2003a) "A Data Selection Method For Probabilistic Neural Networks", TAINN2003, International 12th Turkish Symposium on Artificial Intelligence and Neural Networks, 2003, Çanakkale.
- Bolat, B. ve Yıldırım, T., (2003b), "Performance Increasing Methods for Probabilistic Neural Networks", Pakistan Journal of Information Technologies, 2 (3), 250-255.
- Bolat, B. ve Küçük, Ü., (2004) "İstatistiksel Sinir Ağları ile Konuşma/Müzik Sınıflandırma", IEEE 12. Sinyal İşleme ve İletişim Uygulamaları Kurultayı, 28-20 Nisan 2004, Kuşadası, 227-229.
- Bolat, B. ve Yıldırım, T., (2004), "A Data Exchange Method for Probabilistic Neural Networks", Journal of Electrical and Electronics Engineering, 4 (2), 1137-1140.
- Bolat, B., Küçük, Ü. ve Yıldırım, T., (2004) "Aktif Öğrenen PNN ile Konuşma/Müzik Sınıflandırma", Akıllı Sistemlerde Yenilikler ve Uygulamaları Sempozyumu, 23-25 Haziran 2004, İstanbul, 187-189.
- Bolat, B. ve Yıldırım, T., (2005), "Active Learning for Probabilistic Neural Networks", Lecture Notes in Computer Science, 3610 (1), 110-118.
- Bregman, A., (1990), "Auditory Scene Analysis", MIT Pres, Massachussets.
- Brown, J. C., (1999), "Computer Identification of Musical Instruments Using Pattern Recognition with Cepstral Coefficients as Features", Journal of the Acoustical Society of America, 105, 1933 – 1941.
- Brown, J. C., Houix, O. ve McAdams, S., (2001), "Feature Dependence in the Automatic Identification Of Musical Woodwind Instruments", Journal of the Acoustical Society of America, 109, 1064 - 1072.
- Campbell, W. C. ve Heller, J. J., (1978) "The Contribution of the Legato Transient to Instrument Identification", Proceedings of Research Symposium on Psychology and Acoustics of Music, 1978, 30-44.
- Choi, E., Hyun, D. ve Lee, C., (2002), "Optimizing Feature Extraction for English Word Recognition", IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP'02), May 2002, Florida, 1813-1816.
- Clark, M. ve Milner, P., (1964), "Dependence of Timbre on the Tonal Loudness Produced by Musical Instruments", Journal of the Audio Engineering Society, 12, 28-31.
- Eagleson, H. W. ve Eagleson, O. W., (1947), "Identification of Musical Instruments When Heard Directly and Over a Public-Address System", Journal of the Acoustical Society of America, 19, 338-342.
- Ellis, D., (1997) "The Weft: A Representation for Periodic Sounds", Proceedings of the

International Conference on Acoustics, Speech and Signal Processing, April 1997, Munich, 1307-1310.

Eronen A., (2001) "Automatic Musical Instrument Recognition", Yüksek Lisans Tezi, Tampere University of Technology.

Eronen, A., (2003) "Musical Instrument Recognition Using ICA-Based Transform of Features and Discriminatively Trained HMMs", 7th International Symposium on Signal Processing and Its Applications, July 2003, 2: 133-136.

Foote, J., (1997) "Content-Based Retrieval of Music and Audio", Proceedings of SPIE'97, 1997, Dallas.

Fraser, A. ve Fujinaga, I., (1999) "Toward Realtime Recognition of Acoustic Musical Instruments", Proceedings of International Computer Music Conference (ICMC99), 1999, Beijing, 175-177.

Fujinaga, I. ve MacMillan, K., (2000) "Realtime Recognition of Orchestral Instruments", Proceedings of International Computer Music Conference (ICMC00), 2000, Berlin, 141-143.

Gerhard, D., (2000), "Audio Signal Classification", Doktora Tezi, Simon Frasier University, Kanada.

Ghias, A., Logan, J. ve Chamberlin, D., (1995) "Query by Humming- Musical Information Retrieval in an Audio Database", Proceedings of ACM Multimedia Conference, 1995, Anaheim, 231-235.

Goh, T. C., (2002) "Probabilistic Neural Network for Evaluating Seismic Liquefaction Potential", Proceedings of IEEE International Symposium on Intelligent Systems, 2002, Varna, 16-20.

Gravier, G., Sigelle, M. ve Chollet, G., (1999), "Markov Random Field Modelling for Speech Recognition", Australian Journal of Intelligent Information Processing Systems, 5(4), 245-252.

Guo, G., Zhang, H. ve Li, S. Z., (2001) "Boosting for Content-Based Audio Classification and Retrieval: an Evaluation", Proceedings of IEEE International Conference on Multimedia and Expo, August 2001, Tokyo.

Gülmezoğlu, M. B., Dzhaferov, V., Keskin, M. ve Barkana, A., (1999) "A Novel Approach to Isolated Word Recognition", IEEE Transactions on Speech and Audio Processing, 7(6):620-628.

Hainsworth, S. ve Macleod, M., (2003) "Onset Detection in Musical Audio Signals", Proceedings of the International Computer Music Conference, September 2003, Singapore.

Handel, S., (1995), "Timbre Perception and Auditory Object Identification", B. C. J. Moore (ed.), Hearing, Academic Pres, New York.

Hariharan, R., Kiss, I. ve Viikki, S., (2001), "Noise Robust Speech Parametrization Using Multiresolutional Feature Extraction", IEEE Transactions on Speech and Audio Processing, 9(8):856-865.

Hayes, M. H., (1996) Statistical Digital Signal Processing and Modelling, John Wiley & Sons, New York.

- Haykin, S., (1991) Adaptive Filter Theory, Prentice Hall, New York.
- Herrera, P. ve Bonada, J., (1998) "Vibrato Extraction and Parametrization in the Spectral Modeling Synthesis Framework", 1st COST-G6 Workshop on Digital Audio Effects, November 1998, Barcelona.
- Higgins, A., Bahler, L. ve Porter, J., (1991), "Speaker Verification Using Randomized Phrase Prompting", Digital Signal Processing, 1(2):89-106.
- Huang, R. ve Hansen, J. H. L., (2004) "High-Level Feature Weighted GMM Network for Audio Stream Classification", International Conference on Spoken Language Processing, October 2004, Jeju Island, 1-4.
- Hyun, D. ve Lee, C., (1999) "Optimization of Mel-Cepstrum for Speech Recognition", IEEE International Conference on Systems, Man, and Cybernetics, 12-15 October 1999, 500 – 503.
- Ito, T., Takeda, K. ve Itakura, F., (2005), "Analysis and Recognition of Whispered Speech", Speech Communication, 45:139-152.
- Kaminskyj, I. ve Materka, A., (1995) "Automatic Source Identification of Monophonic Musical Instrument Sounds", IEEE International Conference on Neural Networks, 1995, USA, 189-194.
- Kedem, B., (1986), "Spectral Analysis and Discrimination by Zero-Crossings", Proceedings of the IEEE, 74(11), 1477-1493.
- Kendall, R. A., (1986), "The Role of Acoustic Signal Partitions in Listener Categorization of Musical Phrases", Music Perception, 4(2), 185-214.
- Kimber, D. ve Wilcox, L., (1996) "Acoustic Segmentation for Audio Content Analysis", Proceedings of Interface Conference, July 1996, Sydney.
- Kitahara, T., Goto, M. ve Okuno, H. G., (2003) "Musical Instrument Identification Based on F0-Dependent Multivariate Normal Distribution", International Conference Multimedia and Expo, July 2003, 3: 409-412.
- Kondo, K., Kamata, H. ve Ishida, Y., (1994) "Speaker-independent Spoken Digits Recognition Using LVQ", IEEE World Congress on Computational Intelligence, 27 June-2 July 1994, 4448 – 4451.
- Kostek, B., (2005), "Musical Instrument Classification and Duet Analysis Employing Music Information Retrieval Techniques", Proceedings of the IEEE, 92(4), 712-729.
- Kostek, B. ve Czyzewski, A., (2001) "Automatic Recognition of Musical Instrument Sounds – Further Developments", Proceedings of 110th Audio Engineering Society Convention, 2001, Amsterdam.
- Lee, C., Hyun, D., Choi, E., Go, J. ve Lee, C., (2003), "Optimizing Feature Extraction for Speech Recognition", IEEE Transactions on Speech and Audio Processing, 11(1):80-87).
- Li, D., Sethi, I. K., Dimitrova, N. ve McGee, T., (2001), "Classification of General Audio Data for Content-Based Retrieval", Pattern Recognition Letters, 22:533-544.
- Liu, Z., Huang, Y., Wang, Y. ve diğerleri, (1997) "Audio Feature Extraction and Analysis for Scene Classification", Proceedings of IEEE 1st Multimedia Workshop, 1997.
- Liu, Z., Huang, Y., ve Wang, Y., (1998) "Classification of TV Programs Based on Audio

Information Using Hidden Markov Model”, Proceedings of IEEE Second Workshop on Multimedia Signal Processing, December 1998, California, 27-32.

Lu, L., Li, S. Z ve Zhang, H. J., (2001) “Content-Based Audio Segmentation Using Support Vector Machines”, Proceedings of IEEE International Conference on Multimedia and Expo, August 2001, Tokyo, 956-959.

Martin, K. D., (1998) “Musical Instrument Recognition : A Pattern Recognition Approach”, Presented at 136th Meeting of the Acoustical Society of America, 1998, Norfolk.

Martin, K. D., (1999) “Sound-Source Recognition: A Theory and Computational Model”, Doktora Tezi, Massachusetts Institute of Technology, Massacushets.

Mashao, D. J., (1996) “Experiments on a Parametric Nonlinear Spectral Warping for an HMM-based Speech Recognizer”, IEEE International Conference on Acoustics, Speech and Signal Processing, May 1996, 17-20.

Matsui, T. ve Furui, S., (1994), “Comparison of Text-Independent Speaker Recognition Methods Using VQ-Distortion and Discrete/Contunious HMM’s”, IEEE Transactions on Speech and Audio Processing, 2(3):456-459.

McAdams, S., (1993), “Recognition of Auditory Sound Sources and Events. Thinking in Sound: The Cognitive Psychology of Human Audition”, Oxford University Pres, Oxford.

Moore, B. C. M., (1995), “Hearing”, Academic Press, Toronto.

Parzen, E., (1962), “On Estimation of a Probability Density Function and Mode”, Annals of Mathematical Statistics, 33, 1065 – 1076.

Pierce, J. R., (1983), “The Science of Musical Sound”, Scientific American Library, New York.

Rabiner, L. ve Juang, B., (1993) Fundamentals of Speech Recognition, Prentice Hall, New York.

Reynolds, D., (1996), “MIT Lincoln Laboratory Site Presentation”, Speaker Recognition Workshop, March 1996, MD.

Reynolds, D. ve Carlson, B., (1995) “Text Dependent Speaker Verification Using Decoupled and Integrated Speaker and Speech Recognizers”, Proceedings of EUROSPEECH, 1995, Madrid, 647-650.

Reynolds, R. ve Rose, R., (1995), “Robust Text Independent Speaker Identification Using Gaussian Mixture Speaker Model”, IEEE Transactions on Speech and Audio Processing, 3(1):72-83.

Rossignol, S., Depalle, P., Soumagne, J., Rodet, X. Ve Collette, J.-L., (1999) “Vibrato: Detection, Estimation, Extraction, Modification”, 2nd COST-G6 Workshop on Digital Audio Effects, December 1999, Trondheim.

Russel, M. J. ve Jackson, P. J. B., (2005), “A Multiple Linear/Linear segmental HMM With a Formant-Based Intermediate Layer”, Computer Speech and Language, 19:205-225.

Saldanha, E. L., Corso, J. F., (1964), “Timbre Cues And The Identification of Musical Instruments”, Journal of the Acoustical Society of America, 36, 2021-2026.

Sarıkaya, R., (2001), “Robust and Efficient Techniques for Speech Recognition in Noise”,

Doktora Tezi, Duke Üniversitesi.

Saunders, J., (1996) "Real-time Discrimination of Broadcast Speech/Music", Proceedins of ICASSP'96, May 1996, Atalanta, 993-996.

Schmid, C. E., (1997) "Acoustic Pattern Recognition of Musical Instruments", Doktora Tezi, University of Washington.

Seddik, H., Rahmouni, A. ve Sayadi, M., (2004) "Text Independent Speaker Recognition Using the Mel Frequency Cepstral Coefficients and a Neural Network Classifier", First International Symposium on Control, Communications and Signal Processing, 631-634.

Serra, X. ve Bonada J., (1998) "Sound Transformations Based on the SMS High Level Attributes", Proceedings of COST G6 Conference on Digital Audio Effects, 1998, Barcelona.

Slaney, M., (1998), "Auditory Toolbox", Teknik Rapor, Rapor No 1998-010, Interval Research Corp.

Specht, D. F., (1988) "Probabilistic Neural Networks for Classification, Mapping or Associative Memory", IEEE International Conference on Neural Networks, 1988, 525-532.

Specht, D.F., (1990a), "Probabilistic Neural Networks", Neural Networks, 3, 109-118.

Specht, D.F., (1990b), "Probabilistic Neural Networks and the Polynomial Adaline as Complementary Techniques for Classification", IEEE Transactions on Neural Networks, 1(1):111-121.

Strong, W. ve Clark, M., (1967), "Perturbations of Synthetic Orchestral Wind Instrument Tones", Journal of the Acoustical Society of America, 41, 277-285.

Tishby, N. Z., (1991), "On the Application of Mixture AR Hidden Markov Models to Text Independent Speaker Recognition", IEEE Transactions on Acoustics, Speech and Signal Processing, 39(3):563-570.

Wang, Y., (2003), "Representing Signals Using Only Timing and Feature Extraction for Automatic Speech Recognition", Doktora Tezi, Rhode Island Üniversitesi.

Zhang, T. ve Jay Kuo, C. C., (1998) "Content-Based Classification and Retrieval of Audio", 43rd Annual Meeting-Conference on advanced Signal Processing Algorithms, Architectures, and Implementations VII, San Diego, July 1998, SPIE 3461:432-443.

ÖZGEÇMİŞ

Doğum tarihi	20.11.1973	
Doğum yeri	Kahraman Maraş	
Lise	1986–1989	İstanbul Bahçelievler Lisesi
Lisans	1989–1996	Yıldız Teknik Üniversitesi Elektrik-Elektronik Fak. Elektronik ve Haberleşme Mühendisliği Bölümü
Yüksek Lisans	1996–1998	Yıldız Teknik Üniversitesi Fen Bilimleri Enstitüsü Elk. ve Hab. Müh. ABD., Haberleşme Programı
Doktora	2002–2006	Yıldız Teknik Üniversitesi Fen Bilimleri Enstitüsü Elk. ve Hab. Müh. ABD., Haberleşme Programı

Çalıştığı kurumlar

1996–1997	Only Bilgisayar
1997-Devam ediyor	YTÜ Elektrik-Elektronik Fak. Araştırma Görevlisi