YILDIZ TEKNİK ÜNİVERSİTESİ
FEN BİLİMLERİ ENSTİTÜSÜ

SAKLI MARKOV MODEL TABANLI MÜZİK PARÇASI TANIMA SİSTEMİ

Bilgisayar Müh. Güngör Tumak

FBE Bilgisayar Mühendisliği Anabilim Dalında Hazırlanan

YÜKSEK LİSANS TEZİ

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

İSTANBUL, 2009

YILDIZ TEKNİK ÜNİVERSİTESİ
FEN BİLİMLERİ ENSTİTÜSÜ

SAKLI MARKOV MODEL TABANLI MÜZİK PARÇASI TANIMA SİSTEMİ

Bilgisayar Müh. Güngör Tumak

FBE Bilgisayar Mühendisliği Anabilim Dalında Hazırlanan

YÜKSEK LİSANS TEZİ

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

İSTANBUL, 2009

İÇİNDEKİLER

Sayfa

SİMGE LİSTESİ

w_i	i.sınıf
μ	Ortalama
$P(w_i)$	Sınıfın önsel olasılığı
$P(w_i\|x)$	Sınıfın şartlı olasılık yoğunluk işlevi
$P(x\|w_i)$	Sınıfta x'in olasılık yoğunluk işlevi
C_j	j. Kod vektörü
$p(\vec{r},t)$	Basıncın yere ve zamana bağlı fonksiyonu
$u(\vec{r},t)$	Ortamdaki parçacıkların hızının zamana ve yere bağlı fonksiyonu
$\phi_n(i,k)$	Kovaryans fonksiyonu
$r_n(x)$	Otokorelasyon fonksiyonu
$w(n)$	Pencereleme fonksiyonu
$X(w)$	Fourier dönüşümü sonucu veya ayrık zamanlı Fourier dönüşümü sonucu
$x(t)$	Zaman boyutunda sinyal
$F(s)$	Laplace dönüşümü sonucu veya karmaşık değişkenli fonksiyon
$f(t)$	Ters laplace dönüşümü sonucu veya zaman değişkenli fonksiyon
$x[n]$	Zaman boyutunda ayrık sinyal
$X[k]$	Ayrık fourier dönüşümü sonucu
m	Mel frekans skalasında bir değişken
f	Frekans skalasında bir değişken
$P(A\|B)$	Koşullu olasılık
λ	Saklı markov modeli
S	Saklı markov modeli durum kümesi
A	Saklı markov modeli durumlar arası geçiş matrisi
B	Saklı markov modeli gözlem sembolleri gerçekleşme olasılık matrisi
π	Saklı markov modeli başlangıç durumu olma olasılık vektörü
N	Saklı markov modeli durum sayısı

M	Saklı markov modeli gözlem sembolleri sayısı
a_{ij}	i durumundan j durumuna geçiş olasılığı
$b_j(k)$	j durumunda k gözlem sembolünün gerçekleşme olasılığı
π_i	i durumunun başlangıç durumu olma olasılığı
Q	Durum dizisi
O	Gözlem dizisi
q_t	t anında bulunan durum
v_k	Gözlem kümesinin k'nıncı elemanı
o_t	t anındaki gözlem sembolü
α	İleri değişken
β	Geri değişkeni

KISALTMA LİSTESİ

ANN	Artificial Neural Network (Yapay Sinir Ağları)
A/D	Analogtan Dijitale Çevirici
CMU	Carnegie Mellon University
coHMM	Competitive Hidden Markov Model
dB	desibel
D/A	Dijitalden Analoğa Çevirici
DCT	Discrete Cosine Transform (Ayrık Kosinüs Dönüşümü)
DFT	Discrete Fourier Transform (Ayrık Fourier Dönüşümü)
DTFT	Discrete Time Fourier Transform (Ayrık Zamanlı Fourier Dönüşümü)
DTW	Dynamic Time Warping (Dinamik Zaman Bükmesi)
FFT	Fast Fourier Transform (Hızlı Fourier Dönüşümü)
FRR	False Rejection Rate
GMM	Gauss Mixture Model(Gauss karışım modeli)
HMM	Hidden Markov Model (Saklı Markov Modeli)
Hz	Hertz
kHz	kilo Hertz
LPC	Linear Predictive Coding (Doğrusal Öngörülü Kodlama)
MFCC	Mel Frequency Cepstral Coefficients(Mel Frekansı Kepstral Katsayıları)
MIR	Music information retrival(Müzik Bilgi Erişimi)
Ms	Milisaniye
Sn	Saniye
VN	Vektör Nicemleme

ŞEKİL LİSTESİ

ÇİZELGE LİSTESİ

ÖNSÖZ

Yıldız Teknik Üniversitesi'nde Bilgisayar Mühendisliği lisans ve yüksek lisans eğitimim sırasında aldığım sınıflandırma ve istatistik konularını içeren derslerin ve müzik merakımın sonucu olarak tez çalışmam olarak karar verdiğim "Saklı Markov Model Tabanlı Müzik Parmak izi Sistemi" konusunda başarılı bir çalışma yapmayı hedefledim.

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

Ayrıca tez konusunda ilerlememde bana yardımcı olan arkadaşlarıma ve aileme katkılarından dolayı teşekkür ederim.

Güngör Tumak

Mart, 2009

ÖZET

SAKLI MARKOV MODEL TABANLI MÜZİK PARÇASI TANIMA SİSTEMİ

Güngör TUMAK
Bilgisayar Mühendisliği, Yüksek Lisans Tezi

İçerik tabanlı müzik tanıma sistemleri müzik parçalarını müzik parmak izi olarak isimlendirilen imzalar şeklinde saklarlar. Müzik parmak izi, müzik kaydını özetleyen içerik tabanlı kompakt bir imzadır. Bu yöntem, müzik parçalarının herhangi bir tanım bilgisine ihtiyaç duymadan, formattan bağımsız olarak tanınmasına imkan sağlar. Bu tez çalışmasında bir müzik parçasını kısa bir bölümünden tanıyabilen içerik tabanlı bir Müzik Bilgi Erişimi Sistemi tasarlanmış ve gerçekleştirilmiştir.

Sistemde, müzik parçalarından Mel Frekansı Kepstral Katsayıları (Mel Frequency Cepstrum Coefficents - MFCC) özellikleri ile tanımlayıcı akustik bilgiler çıkarılmış ve Saklı Markov Modeli (Hidden Markov Model – HMM) ile modellenmiştir.

MFCC özellikleri müzik verisine ilişkin özelliklerin ortaya konmasında etkili bir yöntemdir. HMM de ardışıl özelliklerin geliş sırası dikkate alınarak sınıflandırılmasını sağlayan bir yöntemdir. MFCC adımlarının müzik verisine uygulanmasıyla elde edilen 12 uzunluklu özellik vektörlerine delta özellikleri eklenerek 36 uzunluklu olarak kullanılmıştır. Özellik vektörleri gerekli normalizasyon işlemlerinin ardından modellemeye uygun hale getirilmiştir.

Konuşma tanımada her HMM bir fonemi modeller. Fakat müzikte fonem kavramı yoktur. Çalışmada, konuşma tanımadaki fonemlere benzer yapıda müziği ifade eden akustik müzik birimlerinin eğitmensiz olarak HMM ile modellemesi yapılmıştır. Bu akustik müzik birimlerinin her biri ayrı bir müzik olayını ifade ettiği için AudioGen olarak ifade edilir. AudioGen'lerin eğitimi için eğitim verisinde bulunan müzik parçalarından rastgele alınmış 10-15 sn'lik bölümler birleştirilip toplu bir eğitim seti hazırlanmıştır. Oluşturulan eğitim setinden kümeleme ve HMM eğitim algoritmaları ile istenilen sayıda AudioGen üreten tasarım çalışması yapılmıştır. Her bir AudioGen 3 olaylı ergonomik HMM modelinden oluşmaktadır. Çalışmada 32 adet AudioGen kullanılmıştır.

Müzik parçalarının parmak izlerinin üretilebilmesi için AudioGen'ler kullanılır. Müzik parçası küçük parçalara bölünür ve her parçaya kendisini oluşturmuş olma olasılığı en yüksek olan AudioGen atanır. Müzik parçasından dakikada 800 AudioGen içerecek şekilde parmak izleri oluşturulur. Parmak izleri AudioGen'ler dizilimi şeklinde olduğu için AudioDNA olarak adlandırılır.

Sistem, eğitim seti içerisinde bulunan müzik parçalarının, farklı kaynaklardan çalınan 10 saniyelik bölümlerinden tanıma yapılması deneysel kurgusu üzerine tasarlanmış ve test edilmiştir. Bu müzik parçası bölümleri için kısa AudioDNA'ler oluşturulur ve veritabanındaki müzik parçası parmak izleri ile karşılaştırılır. Tasarlanan müzik parmak izi yapısının canlı genlerine benzerliğinden ötürü karşılaştırma yöntemi olarak biyoinformatikte kullanılan Smith-Waterman gen hizalama algoritması kullanılmıştır. Dış kaynaktan gelen müzik verisinde ortam gürültüsünün etkisini azaltmak için bilgisayarın mikrofon portu ile müzik çalar arasından ses kablosu kullanılmıştır. Farklı deneysel kurgularla yapılan testlerde %60 - %87,5 arasında tanıma başarısı elde edilmiştir.

Anahtar Kelimeler: Müzik parmak izi, müzik bilgi erişimi, müzik parçası tanıma, saklı markov modeli, mel frekans kepstral katsayıları.

JÜRİ:

1. Yrd. Doç. Dr. M. Elif KARSLIGİL (Danışman)
2. Yrd. Doç. Dr. Songül ALBAYRAK
3. Yrd. Doç. Dr. Soner ÖZGÜNEL

Kabul tarihi: 16.03.2009
Sayfa Sayısı: 61

ABSTRACT

HIDDEN MARKOV MODEL BASED SONG IDENTIFICATION SYSTEM

Güngör TUMAK
Computer Enginering, M.S. Thesis

Content based music information systems store songs in a compact fingerprints. At the core of the presented system is a fingerprint extraction system. An audio fingerprint is a content-based compact signature that summarizes an audio recording. Audio Fingerprinting technologies allow the monitoring of audio independently of its format and without the need of meta-data or watermark embedding. With this thesis a content based music information retrival(MIR) system is designed and implemented.

On this purpose mel frequency cepstral coefficients (MFCC) feature extraction is used and feature vectors are classified by hidden markov models (HMM). MFCC is an effective feature extraction method for music information retrival and HMM can handle sequential feature vectors.

Real valued MFCC feature vectors with 12 dimensions are firstly exracted from music signal. Delta features added vectors are used with 36 dimensions. After normalization step, feature vectors are used in modelling.

In speech, the target for each HMM is a phoneme (or other phonetic related characteristic), but in music there are not such "phonemes". A way to define some properties for the units that can suit the music identification problem and to model with unsupervised HMM is presented. HMMs are named AudioGens. Training set is prepared from random selected 10-15 second of each song. Each AudioGen is 3 state ergonomic HMMs. 32 AudioGens are used for the system.

AudioGens are used to create fingerprints from songs. Fingerprints are named AudioDNA because they are sequence of AudioGens. Fingerprints include 800 AudioGens in a minute of song.

System is tested on 10 seconds music parts playing in a different source. Short AudioDNA fingerprints are extracted from these music parts and matched with AudioDNAs in the music fingerprint database. Music fingerprint scheme is similar with living being genes and Smith-Waterman gene sequence alignment algorithm is used for matching process. Direct cable connection between music player and computer microphone port is used to reduce environment noise. Recognition success ratio of system is between %60 and %87,5 with different test variations.

Keywords: Music fingerprint, audio fingerprint, music information retrival, hidden markov model, mel frequency cepstral coefficients.

JURY:

1. Assist. Prof. M. Elif KARSLIGİL (Supervisor)
2. Assist. Prof. Dr. Songül ALBAYRAK
3. Assist. Prof. Dr. Soner ÖZGÜNEL

Date: 16.03.2009
Page: 61

1. GİRİŞ

Arabada radyo dinlediğimizi ve çalan bir parçanın dikkatimizi çektiğini düşünelim. Bu şarkının kim tarafından söylendiğini nasıl öğrenebiliriz? Radyoyu arayabiliriz fakat bu çok zor bir yol. Peki cep telefonumuzda bir iki tuşa basarak çalan şarkının kısa bir bölümünü dinlettiğimizi ve mesaj olarak şarkının ve şarkıcının adının geldiğini düşünelim. Bu tip bir sistemin gerçeklenebilmesi için içerik tabanlı bir müzik tanıma uygulamasına ihtiyaç duyulmaktadır. İçerik tabanlı müzik tanıma sistemleri müzik parçalarını müzik parmak izi adında kompakt imzalar şeklinde saklarlar. Tez kapsamında yapılan çalışmanın temelinde müzik parçalarından içerik tabanlı parmak izleri çıkarılması bulunmaktadır.

Parmak izinin kullanımı yüzyıldan fazla öncesine dayanmaktadır. 1893 yılında Sir Francis Galton her insanın parmak izinin birbirinden farklı olduğunu kanıtlamıştır. Yaklaşık 10 yıl sonra ise İskoçya'da insanların parmak izlerini tanımak için Sir Edward Henry tarafından geliştirilen sistem kullanılmaya başlanmıştır. Parmak izi her insan için tekil olan bir özet ya da imza olarak görülebilir. Önemli bir nokta ise bu imzadan insanın diğer özelliklerinin ortaya çıkarılamayışıdır. Örnek olarak bir parmak izi insanın saç ya da göz rengi konusunda bilgi içermez.

Parmak izi konusunda büyüyen bilimsel ve endüstriyel ilgi ile çeşitli multimedya nesnelerinin parmak izleri konusunda çalışmalar yapılmıştır (Cano ve Battle, 2002; Cheng vd., 2001; Allamanche vd., 2001; Neuschmied vd., 2001; Fragoulis vd., 2001). Bu büyüyen endüstri de çeşitli firmaların ortaya çıkmasını sağlamıştır. Bunlarda bazıları Auditude, Relatable, Audible Magic, Shazam, Moodlogic, Yacast firmalarıdır[1][2][3][4][5][6].

Müzik parmak izinin en iyi bilinen özelliği tanınmayan müzik parçasını formattan bağımsız olarak uygun şarkı ve şarkıcı bilgisi ile tanımlayabilmesidir. Müzik parmak izi ya da içerik tabanlı müzik tanıma(Content Based Audio Identification - CBID) yöntemleri müzik parçasının algısal parçalarından oluşan ses parmak izini çıkarır ve bir veritabanında saklar (Cano ve Battle, 2002). Bilinmeyen bir müzik parçası geldiğinde ise bu müzik parçasının parmaki izi hesaplanır ve veritabanındaki kayıtlar ile karşılaştırılır. Çeşitli parmak izi karşılaştırma yöntemleri kullanılarak parazitlenmiş ya da kirlenmiş ses kayıtlarının da tanınması sağlanabilir.

Müzik parçası tanıma işlemi en basit yaklaşım ile sayısallaştırılmış dalga şeklinin doğrudan karşılaştırılması olarak düşünülebilir. Verimli şekilde düşünülen bir yaklaşımda ise müzik

dosyalarının kompakt gösterimi için MD5(Message Digest 5) ya da CRC(Cylic Redundancy Checking) gibi bir erişim yöntemi kullanılabilir. Bu şekilde bir yöntemde karşılaştırma için tüm dosya değil dosyaların kompakt gösterimleri kullanılabilir. Fakat bu tip özetleme yöntemleri dosyadaki bir bitin değişmesinde bile tüm hash değerinin değişmesine neden olur. Bu tip yöntemlerle yapılan karşılaştırmalar sesteki en ufak parazitlerde bile başarısız sonuçlar verecektir ki aslında sadece dosya bazlı işlemler yaptıkları için içerik tabanlı tanıma olarak kabul edilemezler (Cano ve Battle, 2002).

İdael bir müzik parmak izi yönteminin bazı gereksinimleri vardır. Bir parçayı sıkıştırma derecesinden, bozulmadan ya da iletim kanalındaki parazitten mümkün olduğu kadar bağımsız olarak tanımalıdır. Uygulamaya bağlı olarak beş on saniyelik ses parçasından tüm müzik parçasını tanımalıdır. Bu da çıkarılan ve veri tabanında tutulan parmak izlerinin senkronizasyonunu zorlaştıran, yer değiştirme ile alakalı metodlara ihtiyaç duyar. Ayrıca idael bir parmakizi sistemi bozulmanın diğer kaynakları olan pitching(şarkının hızlı ya da yavaş çalınması), eşitleme, arka plan gürültüsü, D/A - A/D çevrimi, konuşma ya da ses kodlayıcıları(GSM ya da MP3 gibi) gibi özelliklerle de ilgilenmelidir. Ayrıca sistem hesaplama konusunda da efektif olmalıdır. Hesaplamanın karmaşıklığı da parmak izinin boyutu, arama algoritmasının ve parmak izi çıkarma algoritmasının karmaşıklığına bağlıdır.

Müzik parmak izi konusundaki tasarım prensipleri çeşitli araştırma alanlarında yinelenmektedir. Karmaşık multimedya nesnelerini tanımlamak için kullanılan kompakt imzalar hızlı indeksleme ve erişim için bilgi erişimi konusuna hizmet etmektedir. Bir multimedya nesnesini indekslemek için boyutunu azaltmak gerekmektedir (Baeza-Yates vd., 1999), (Kimura vd., 2001). Şifrelenmiş hash değerleri gibi içerik tabanlı sayısal imzalar da içerik koruyucu özellikleriyle hash değerlerinin evrimleşmiş halleri olarak görülebilirler. Ayrıca örnek karşılaştırma noktasından bakıldığında ses nesnelerinin niteliklerinin onun ana karakteristik özelliklerinden çıkarılması sınıflandırma sisteminin temelini oluşturmaktadır (Hatisma vd., 2001; Mihak vd, 2001).

1.1 Müzik Parmak İzi Kavramları

Bu bölümde bir müzik parçasının özetini oluşturan müzik parmak izinin özelliklerinden bahsedilecektir.

1.1.1 Müzik Parmak İzi Tanımı

Müzik parmak izi bir müzik parçasının kısa bir özeti olarak görülebilir. Bu yüzden parmak izi

fonksiyonu F çok miktarda bitlerden oluşan X ses objesini, belirli sayıda bitler içeren parmak izine eşlemelidir.

Daha önce bahsedildiği gibi hash fonksiyonları bu özelliği sağlasa da ikili dosyadaki çok küçük değişiklikler bile tüm hash değerini değiştirecektir. Bir müzik parçasının orjinal CD versiyonu ile MP3 versiyonunun ikili dosyasının tamamen birbirinden farklı olduğu düşünülürse bu tip kriptografik hash fonksiyonlarının müzik parmak izi sistemlerinde kullanılması mümkün değildir (Cano ve Battle, 2002).

Bir diğer geçerli soru ise algısal olarak benzer ses nesneleri için matematiksel olarak benzer parmak izleri üreten fonksiyonların tasarlanıp tasarlanamayacağıdır. Soru geçerli olsa da bu tip bir modelin oluşturulması temelde mümkün değildir. Daha kesin söylemek gerekirse, algısal benzerlik geçişli değildir. X ve Y benzer ve Y ve Z benzer ise X ve Z nesnelerinin benzer olması gerekmez. Fakat matemaktiksel olarak geliştirilebilecek bir algısal benzerlik modellemesi bu ilişkiyi sağlayacaktır.

Sonuç olarak algısal benzer ses nesnelerinden benzer parmak izleri üretilebilmelidir. Matematiksel olarak ifade edecek olursak parmak izi fonksiyonu F ve eşik değeri T olarak ifade edersek; eğer ses nesneleri benzer ise $| F(X) - F(Y) | <= T$ ve eğer farklı ise $|F(X) - F(Y)| > T$ olarak gösterilir.

1.1.2 Müzik Parçası Tanıma Sistemlerinin Parametreleri

Müzik parmak izini uygun şekilde tanımlamak için bir müzik parçası tanıma sisteminin temel parametrelerini şu şekilde gösterebiliriz.

- Sağlamlık (Robustness) : Bir ses parçası çeşitli sinyal bozulmaları ardından tanınabilir mi? Yüksek derecede sağlamlığı sağlayabilmek parmak izini oluşturan algısal özelliklerin bozulmalardan etkilenmemesi gerekmektedir. Yanlış negatif oranı sağlamlığı göstermek için kullanılır. Yanlış negatif algısal olarak benzer ses parçalarının parmak izlerinin birbirlerinden oldukça farklı olduğu zaman oluşur.

- Güvenirlik: Bir şarkı ne kadar sıklıkla yanlış tanındı? Bu oran da yanlış negatif tarafından oluşturulur.

- Parmak izi boyutu: Parmak izi için ne kadar depolama alanı gerekiyor? Hızlı arama için parmak izleri genelde RAM hafızada bulunur. Parmak izi boyutları genelde şarkı ya da saniye başına boyut olarak tanımlanır. Bu boyut parmak izlerinin bulunacağı veritabanı sunucusunun hafıza boyutunu belirler.

- Öğe boyutu (Granularity): Bir müzik parçasını tanımak için kaç saniyelik örnek gerektiği uygulamaya göre değişmektedir. Bazı uygulamalarda tanıma için şarkının tamamı gerekirken bazılarında ise sadece küçük bir parçası yeterli olmaktadır.

- Arama hızı ve ölçeklenirliği: Bir parmak izini parmak izi veritabanında bulmak ne kadar sürmektedir? Bir veritabanında binlerce şarkı olabilir. Ticari uygulamalarda bu konu temel parametrelerden biridir.

1.2 Uygulama Alanları

Bu bölümde müzik parçası tanıma sistemlerinin kullanım alanları açıklanmıştır.

1.2.1 Yayın İzleme

Yayın izleme müzik parçası tanıma sistemlerinin en bilinen kullanım alanıdır (Cheng vd., 2001), (Allamanche vd., 2001). Televizyon, radyo ya da web yayını için otomatik çalma listesi oluşturabilir veya program, reklam doğrulaması yapabilir. Şu anda yayın izleme halen manual bir süreç olarak uygulanmaktadır.

1.2.2 Bağlı Ses Sistemi(Connected Audio)

Bağlı ses sistemi, müziğin ek ve destekleyici bilgiye bir şekilde bağlı olduğu müşteri uygulamalarının genel adıdır. Özet bölümünde verilen cep telefonu ile şarkı tanıma örneği bunun örneklerinden biridir. Bu iş şu anda bazı şirketler tarafından yapılmaktadır[4].

Diğer bir örnek ise radyoda çalan bir şarkının bir butona basılarak bilgisayardan satın alınması olarak gösterilebilir.

1.2.3 Dosya Paylaşımı için Filtreleme Teknolojisi

Filtreleme içerik paylaşımında aktif bir arıcılık rolü oynar. Bu konudaki ilk örnek dosya paylaşımı sitesi Haziran 1999 yılında başlayan napster'dır[7]. Bu sistemi kullanarak kullanıcılar geniş bir müzik arşivini bilgisayarlarına indirebilmekteydi. Daha sonra kanuni nedenlerle telif hakkı bulunan şarkıların paylaşımı yasaklandı. Bu nedenle napster 2001 Mart ayında napster dosya isimlerini taban alan bir filtreleme sistemini kullanıma aldı. Bu sistem fazla efektif değildi çünkü kullanıcılar dosya isimleri değiştirmeye başladılar. Bunun üzerine naspter Relatable[16] tarafından hazırlanan bir ses parmak izi sistemi kullanmaya başladı.

Müşteri tarafında ise bu teknoloji ilk bakışta negatif bir etki yaratsa da çeşitli potensiyel faydaları da bulunmaktadır. İlk olarak bu teknoloji ile parmak izi veritabanında bulunan tanım bilgileri kullanılarak müzik parçaları organize edilebilir. İkincisi ise parmak izi yöntemi parçanın gerçekten istenen parça olduğunu garanti eder.

1.2.4 Otomatik Müzik Kütüphanesi Organizasyonu

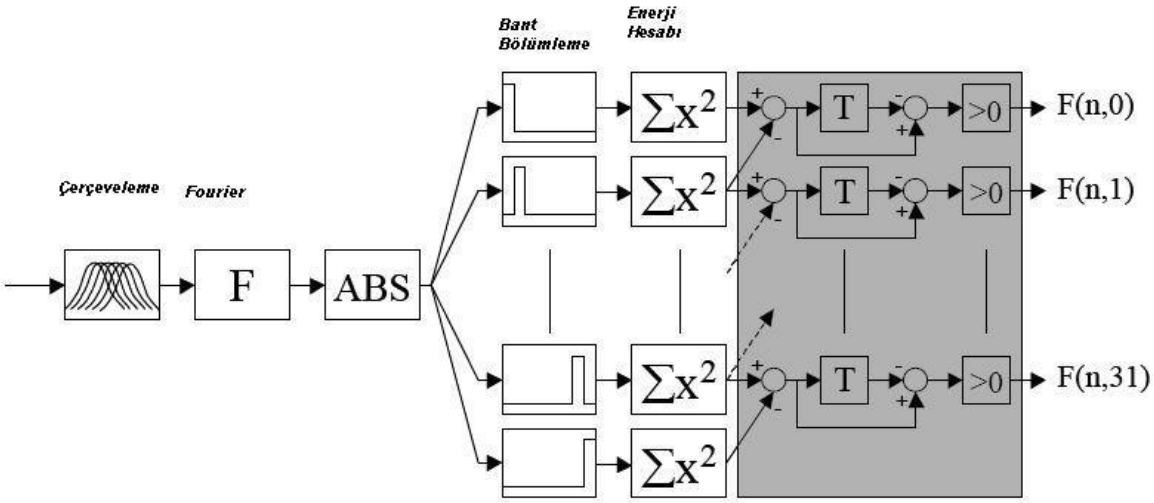
Günümüzde her bilgisayar kullanıcısının binlerce şarkılık müzik arşivleri bulunmaktadır. Bu müzik parçaları ise genelde mp3 gibi sıkışmış formatta saklanmaktadır. Bu şarkılar CD formatından diske kopyalama ya da internetten indirme gibi çeşitli şekillerde geldiği için fazla organize edilmiş şekilde tutulmamaktadır. Bu müzik parçalarının otomatik organizasyonunu yani şarkıcı, albüm gibi tanım bilgilerinin bulunması ve bu bilgilere göre saklanmasını sağlayan uygulamalar geliştirilmiştir. Bu uygulamalar müzik parçalarının parmak izini çıkartır ve internet üzerinden müzik parmak izi veritabanlarından uygun tanım bilgileri ile karşılaştırırlar. Bu tip uygulamalara örnek olarak ID3Man[8] isimli Auditude[3] uygulamaları gösterilebilir.

1.3 Müzik Parçası Tanıma Konusundaki Geçmiş Çalışmalar

Müzik parçası tanıma konusunda bugüne kadar çeşitli çalışmalar yapılmıştır. Bu bölümde uygulamamıza referans olan belli başlı çalışmalardan bahsedilmiştir.

Jaap Haitsma ve Ton Kalter'in 2002 yılında Philips Research bünyesinde yaptığı çalışmada 3 saniyelik müzik parçası bölümünden tanıma yapılması amaçlanmıştır. Müzik parçasından frekans spekturumu kullanılarak özellikler çıkarılmıştır. Elde edilen frekanslardan insan duyma sistemine(Human Auditory System - HAS) uygun olanlar seçilmiş ve frekans değişimlerine göre hazırlanan tasarım ile parmak izi çıkarımı yapılmıştır. Her 11.8 ms için 32 bit alt parmak izleri oluşturulmuştur. 3 saniyelik parmak izi için 256 adet alt parmak izi kullanılmıştır. Arama ve karşılaştırma için kontrol tablosu tabanlı bir yöntem geliştirilmiştir. Genel yapısı şekilde görülebilir (Haitsma ve Kalter, 2002).

Pedra Cano, Eloi Battle, Harald Mayer ve Helmut Neuschmied tarafından yapılan çalışmada AudioDNA adı verilen müzik parmak izi çıkarma çalışması yapılmıştır. Müzik parmak izi Yapılan çalışmanın amacı radyodan çalan müzik parçalarının tanımlanmasıdır. Müzik parçasından MFCC özellikleri çıkartılmış ve HMM ile özellik vektörleri modellenmiştir. Konuşma tanıma temel alınarak tasarlanmıştır. Arama ve karşılaştırma için string karşılaştırma yöntemleri kullanılmıştır (Cano, Battle vd, 2002).



Şekil 1.1 Philips Research Parmak izi Çıkarma Yöntemi (Haitsma ve Kalter, 2002)

E. Battle, J. Masip ve E. Guaus 2004 yılında "AMADEUS" adı verdikleri çalışmayı yayınladılar. Bu çalışmada (Cano ve Battle, 2002)'de temelleri attılan yöntemleri geliştirerek kullanılmıştır. Müzik parçasından çıkarılan özelliklere MFCC yanında sıfır geçiş sayısı (Zero Crossing Rate), Spectral Cendroid(Kepstral Merkez), Spectral Flatness(Kepstral Düzlük) gibi yeni özellikler eklenmiştir. Arama ve karşılaştırma bölümü için HMM tabanlı yöntemler kullanılmıştır (Battle, Masip ve Guaus, 2004).

2. SES VERİSİ ÖZELLİKLERİ

Bu bölümde müzik tanıma sistemlerinde kullanılan yöntemlerin açıklanmasında kolaylık sağlanması için analog ve sayısal ses verisi hakkında çeşitli kavramlardan bahsedilmiştir.

2.1.1 Ses Dalgası Özellikleri

Ses dalgaları 20 Hz.-20 Khz. arasında mekanik titreşim yapan cisimlerin (katı, sıvı, gaz) insan kulağı ile teması olan bir ortamda oluşturdukları dalgalara denir. Dalgalar genel olarak, mekanik ve elektromanyetik dalgalar olmak üzere iki ana gruba ayrılır. Elektromanyetik dalgalar, yayılmak için bir ortama ihtiyaç duymazlar ve boşlukta da yayılabirler. Mekanik dalgalar ise, enerjilerini aktarabilmek için ortam taneciklerine ihtiyaç duyarlar. Bu yüzden boşlukta (örneğin uzayda) yayılamazlar. Ses dalgaları da mekanik dalgalar olduklarından yayılmak için maddesel bir ortama ihtiyaç duyarlar.

Ses, nesnelerin titreşiminden meydana gelen ve uygun bir ortam içerisinde (hava, su vb.) bir yerden başka bir yere, sıkışma (compressions) ve genleşmeler (rarefactions) şeklinde ilerleyen bir dalgadır. Dolayısıyla ses, bir basınç dalgasıdır.

Doğada tek frekansı olan sese nadiren rastlanmaktadır. Ses dalgasının zamana göre değişimi, saf sinüs biçiminde ise böyle bir ses dalgasının tek bir frekansı vardır. Saf sinüsün dışındaki, birden fazla dalgaya kompleks ses dalgası denir. Kompleks ses dalgaları, temel frekans ve temel frekans harmoniklerinden oluşan sinüs dalgaları biçiminde düşünülebilir.

2.1.1.1 Frekans (Sıklık)

Bir dalganın frekansı, dalganın hava veya başka bir ortam içinden geçerken ortamdaki partiküllerin ne sıklıkta titreştiğine bağlıdır. Frekans ileri geri titreşimlerin zamana bağlı olarak ölçülmesi ile hesaplanır. Saniyedeki titreşim sayısı özel olarak Hertz birimi ile ifade edilir (1Hertz=1döngü/saniye).

Yüksek frekans değerleri için Hertz'in bin katı olan ‘kilohertz’ (kHz) birimi kullanılır. İnsan kulağının duyabildiği sesler 20 ile 20000 Hz (20kHz) arasında frekansa sahip olabilir. Eğer bir frekans 20 Hz'in altında ise bu tür titreşimlere ‘ses altı’ titreşimler, frekans 20 kHz' in üzerinde ise bunlara da ‘ses üstü’ titreşimler denilmektedir.

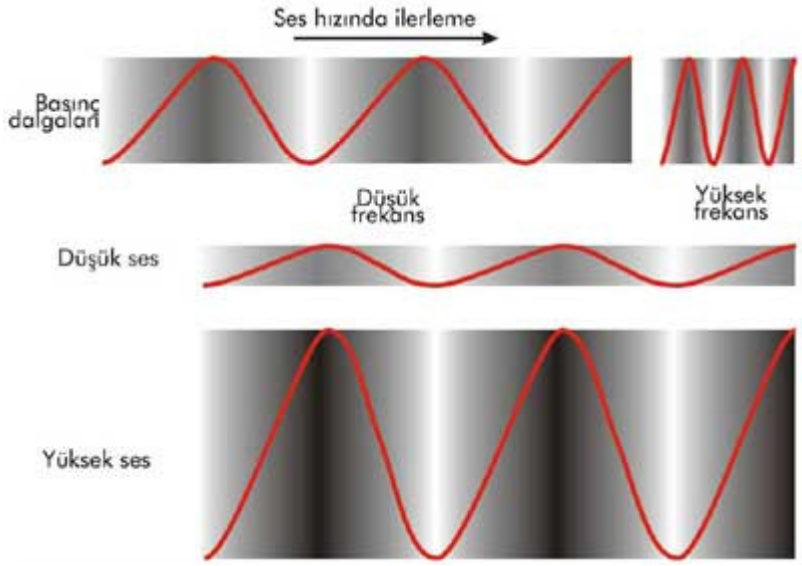
2.1.1.2 Genlik (Amplitüd)

Genlik, ses dalgalarının dikey büyüklüğünün bir ölçüsüdür. Ses dalgalarını oluşturan sıkışma

ve genleşmeler arasındaki fark, dalgaların genliğini belirler. Ses dalgaları havada veya başka bir ortamda titreşen objeler tarafından üretilir. Örneğin titreştirilen bir gitar teli, yaptığı periyodik salınım hareketi ile, hava moleküllerinin belli bir frekansta sıkışmasını ve genleşmesini sağlar. Bu şekilde teldeki enerji havaya iletilmiş olur. Enerjinin miktarı, teldeki titreşim genliğine bağlıdır. Eğer tele fazla enerji yüklenirse, tel daha büyük bir genlikle titreşir. Teldeki titreşim genliği ne kadar fazla ise ortam tanecikleri (örneğin hava molekülleri) tarafından taşınan enerji de o kadar fazladır. Enerji ne kadar fazla ise sesin şiddeti de o kadar büyük olacaktır. Bu ifadeler, titreşen tüm cisimler için geçerlidir.

2.1.1.3 Dalga Boyu

Bir dalganın ardışık iki tepe veya iki çukur noktası arasındaki mesafe bize dalga boyunu verir. Dalga boyu l (lambda) ile gösterilir.



Şekil 2.1 Sesin grafiksel gösterimi

Bir basınç dalgası olan sesin grafiksel gösterimi. Grafiklerde koyu renkli bölgeler sıkışmaları, açık renkli bölgeler ise genleşmeleri simgelemektedir. Eğriler ise bu sıkışma ve genleşmelerin iki boyutlu grafiksel temsilleridir. Dikkat edilirse, sıkışma miktarı arttıkça (yüksek seste olduğu gibi) sesin şiddeti de artmaktadır.

2.1.1.4 Ton

Müzikte, diatonik (doğal major) gamda bir 'tam aralık' olarak tanımlanan ton, belli bir frekansta ve perdede üretilen saf ses anlamında kullanılır. Örneğin bir ses çatalı (diyapozon)

titreştirildiğinde ortaya çıkan 440 Hz frekansındaki ‘Do (C)’ notası, saf bir tondur. Saf tonlar doğal ortamda fazla karşılaşılmayan ve genellikle müzik aletleri veya ses üreteçleri aracılığıyla üretilen seslerdir. Yüksek frekanslı (yüksek perdeden) sesler tiz, düşük frekanslı (düşük perdeden) sesler pes (bas) olarak algılanır.

2.1.1.5 Tını

Sesin ‘rengini’ ifade eden bir terimdir. Aynı oktavda, aynı notayı (tonu) aynı yoğunlukta ve aynı uzunlukta çalan bir kemanla bir flüt arasındakı temel fark ‘tını farkı’dır. Enstrümanları oluşturan bileşenlerin doğal frekanslarındaki farklılıklar, sonuçta oluşan sesin farklı bir tınıda olmasını sağlar. Bu sayede, farklı müzik aletlerinden çıkan özdeş notaları kolaylıkla ayırdedebiliriz. Tını, sesin harmonik (doğuşkan) yapısına bağlı olarak değişir.

2.1.1.6 Sesin Şiddeti ve Desibel Ölçeği

Şiddet, ses dalgalarının taşıdıkları enerjiye bağlı olarak birim alana uyguladıkları kuvvettir. Birimi genellikle ‘metrekare başına Watt’ (W/m2) olarak ifade edilir. Sesin şiddeti, ses kaynağına olan uzaklığın karesi ile ters orantılıdır.

2.1.1.7 Desibel (dB)

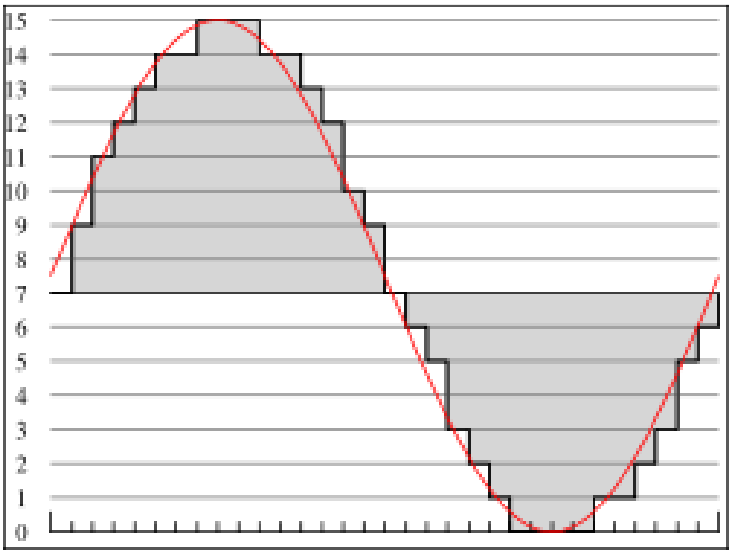
İnsan kulağı çok düşük ve çok yüksek şiddette sesleri duyabilme yeteneğine sahiptir. İnsan kulağının algılayabileceği en düşük ses şiddeti, ‘eşik şiddet’ olarak bilinir. Kulağa zarar vermeden işitilebilen en yüksek sesin şiddeti ise, eşik şiddetinin yaklaşık 1 milyon katı kadardır. İnsan kulağının şiddet algı aralığı bu kadar geniş olduğundan, şiddet ölçümü için kullanılan ölçek de 10'un katları, yani logaritmik olarak düzenlenmiştir. Biz buna ‘desibel ölçeği’ adını vermekteyiz. Sıfır desibel mutlak sessizliği değil; işitilemeyecek kadar düşük ses şiddetini (ortalama 1.10-12 W/m2) gösterir.

Desibel, bir oranı veya göreceli bir değeri gösterir ve ‘*bel*’ biriminin 10 katıdır. Alexander Graham Bell' in anısına bel adı verilen birim, iki farklı büyüklüğün oranının logaritması olarak tanımlanmaktadır. Yani ‘1 bel’, birbirlerine oranları 10 olan iki büyüklüğü göstermektedir (örneğin 200/20). Bu oranın çok büyük olmasından dolayı "Desibel" adı verilen ve oranların logaritmasının 10 katı olarak tanımlanan birim daha yaygın olarak kullanılmaktadır. Bu sayılardan biri bilinen bir sayı olarak alındığından “Desibel”, söz konusu bir büyüklüğün (Pi) referans büyüklüğe (Pref) oranının logaritmasının 10 katıdır (dB=10.log [Pi/Pref]).

dBA ise insan kulağının en çok hassas olduğu orta ve yüksek frekansların özellikle vurgulandığı bir ses değerlendirmesi birimidir. Gürültü azaltması veya kontrolünde çok kullanılan dBA birimi, ses yüksekliğinin sübjektif değerlendirmesi ile ilişkili bir kavramdır. Eşik şiddetindeki ses 'sıfır' desibeldir ve 1.10-12 W/m2 değerine eşdeğerdir. 10 kat daha şiddetli ses 1.10-11 W/m2; yani 10 dB iken, 100 kat daha şiddetli ses 20 dB'dir.

2.1.2 Sayısal Ses Sinyali

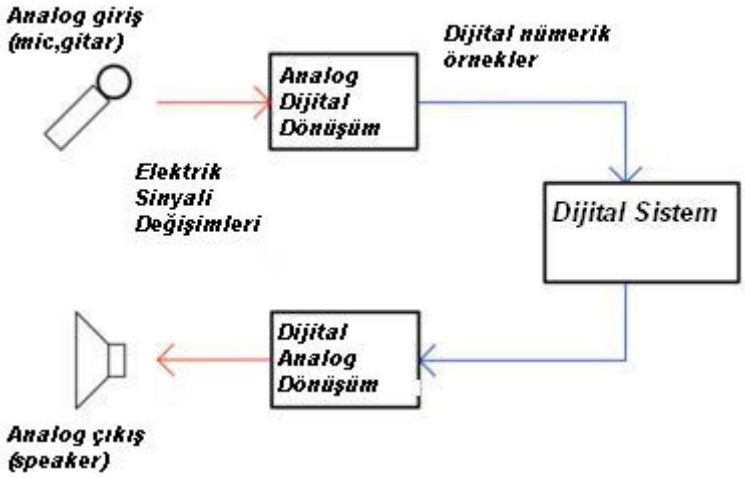
Sayısal ses, analog ses kaydının digital olarak ifade edilmesidir. Analog bir sinyal sayısal sinyale belirli bir örnekleme frekansı(sampling rate) ve bit çözünürlüğünde(bit resolution) çevrilir. Sayısal ses birden fazla kanal içerebilir. Bunlar stereo için iki kanal, surround ses için ikiden fazla ve mono için tek kanaldır. Genel olarak daha yüksek örnekleme frekansı ve bit çözünürlüğü daha kaliteli ses anlamına gelmektedir. Aynı zamanda örnekleme frekansının ve bit çözünürlüğünün artması sayısal veri miktarını arttırmaktadır.



Şekil 2.2 Analog sinyalin örnekleme ve 4 bit çözünürlüğü ile sayısal sinyale dönüştürülmesi

2.1.2.1 Dönüşüm İşlemi

Bir sayısal sinyal, analog-sayısal dönüştürücü (analog-to-digital converter - ADC) ile analog sinyalin sayısal sinyale dönüşmesinden oluşur. ADC belirli bir örnekleme frakansında ve bit çözünürlüğünde çalışır. Örneğin bir müzik CD'si 44.1 Khz (saniyede 44.100 örnek) ve örnek başına 16 bit çözünürlüğe sahiptir.



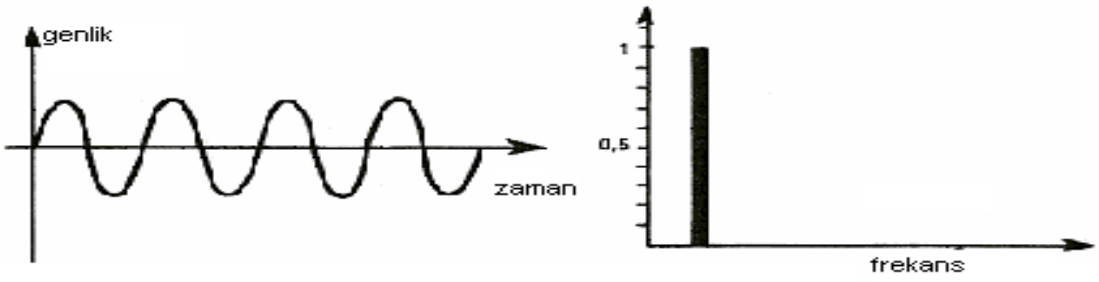
Şekil 2.3 Analog – Sayısal dönüştürücü örneği

Sinyali, belirsizliğe yer vermeyecek şekilde tanımlayabilmek için sinyaldeki en büyük frekans, örnekleme frekansının yarısından büyük olmamalıdır. Aksi takdirde bir sinyal olduğundan farklı görünebilir, buna aliasing denir. Örneklenen sinyalin bu davranışı Nyquist Teoremi ve Shannon Teoremi ile açıklanır. Nyquist frekansı bir sinyaldeki en yüksek frekans değerinin yarısıdır.

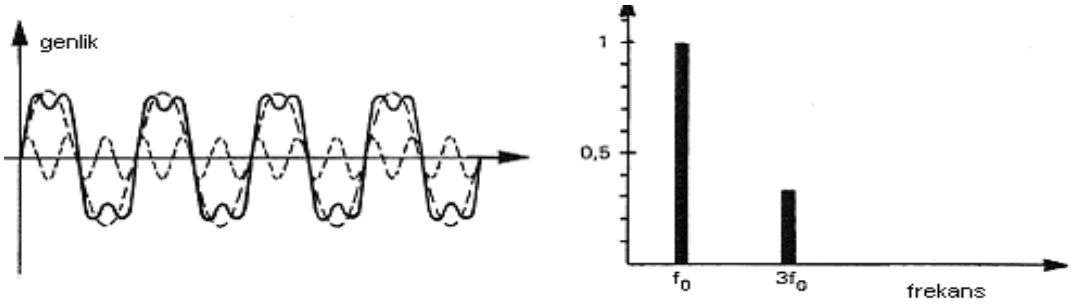
2.1.3 Zaman ve Frekans Domain

Ses sinyali, zaman domain ve frekans domain olarak ifade edilebilir. Zaman domaini ses sinyalinin gösterimidir. Zamana bağlı değişen sinyal genliğini gösterir. Frekans domain ise ses frekansını ve sinyal genliğini gösterir.

Zaman domainden frekans domaine geçiş FFT (fast foruier transform) ile yapılır. Her ses kendisini oluşturan basit sinüs sinyalleri ile ifade edilebilir.



Şekil 2.4 Sinüs sinyali ve frekans domain'de gösterimi



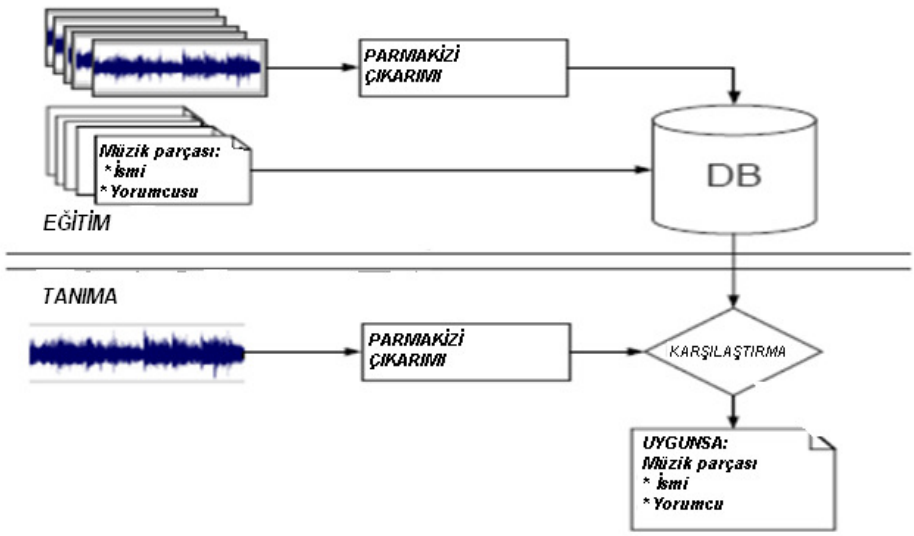
Şekil 2.5 Ses sinyalinin frekans domain'de gösterimi

3. MÜZİK PARÇASI TANIMA SİSTEMLERİNE GENEL BAKIŞ

Konuya farklı tarzlarda çözümler sunulabilmesine rağmen müzik parçası tanıma işlemi genel anlamda aynı yaklaşımları paylaşır. Şekil 3.1'de görüldüğü gibi iki temel işlem vardır. Bunlar parmak izi çıkarılma ve karşılaştırma algoritmasıdır. Parmak izi çıkarma işleminin aşağıdaki özelliklere sahip olması gerekmektedir.

- Veritabanında bulunan çok sayıda parmak izinden ayırt edici özelliği bulunmalıdır.
- Sesteki bozulmalara karşı değişmez olmalıdır.
- Az yer tutmalıdır.
- Hesaplama karmaşıklığı düşük olmalıdır.

Pratikte bu şartların tümünü aynı anda sağlayan özellikleri bulmak oldukça zordur. Uygulamanın türüne göre gerekli şartları sağlayan özellikler kullanılmalıdır.



Şekil 3.1 Müzik parçası tanıma sistemi genel yapısı (Cano ve Batlle, 2002)

Parmak izi çıkarma işlemi, bir ön yüz ve bir parmak izi modelleme bloğundan oluşur. Ön yüz sinyalden çeşitli ölçüleri hesaplar. Parmak izi modelleme bloğu ise son parmak izi gösterimini hazırlar. Bu bir vektör, bir vektörler parçası, bir kodlama kitabı, HMM(Hidden Markov Model – Saklı Markov Modeli) ses sınıflarına ait bir sıralama ya da müziksel şekilde anlamlı yüksek seviye özelliklerden oluşabilir.

Arama algoritması, bir kayıttan oluşturulan parmak izini kullanarak veri tabanında en uygun

eşlemeyi bulmak için arama işlemini yapar. Karşılaştırma yöntemlerinden biri uzaklık hesaplamaktır. Fakat karşılaştırma sayısı arttığında ve uzaklık hesaplaması zorlaştığında arama hesabını hızlandırmak için bazı metodlara ihtiyaç duyulur. İyi bir arama metodunun aşağıdaki özellikleri sağlaması gerekmektedir.

- Hız : Sıralı tarama ve uzaklık hesaplama geniş veritabanları için çok yavaş olabilir.
- Doğruluk : Hatalı Red(False Rejection Rate - FRR) oranı düşük olmalıdır.
- Hafızanın Efektif Kullanılması : Az hafızaya ihtiyaç duymalıdır.
- Kolayca güncellenebilirlik : Nesneleri kolayca giriş, silme ve güncelleme özelliğine sahip olmalıdır.

3.1 Ön Yüz (Front-End)

Ön yüz ses sinyalini ilgili özelliklerin sırası haline getirerek parmak izi modelleme bölümünü besler. Ön yüzün dizaynında çeşitli etkenler rol oynar.

- Boyut azatma.
- Algısal anlamlı parametreler(insan duyma sistemine benzer)
- Değişmezlik ve sağlamlık(kanal bozulmalarına, arka plan sesine karşı)

3.1.1 Ön işleme (Preprocessing)

Ses sinyali sayısallaştırılır ve genel bir formata sokulur. Genelde sağ ve sol kanalları ortalayan mono basit bir formata çevirilir (16 bit PCM). Örnekleme oranı 5 ile 44.1 Khz arasında değişen kesin bir formata getirilir. Telefon ile tanıma gibi durumlarda ses kanallarını simule etmek için ön işleme uygulanabilir.

3.1.2 Çerçeveleme ve örtüştürme (Framing-Overlapping)

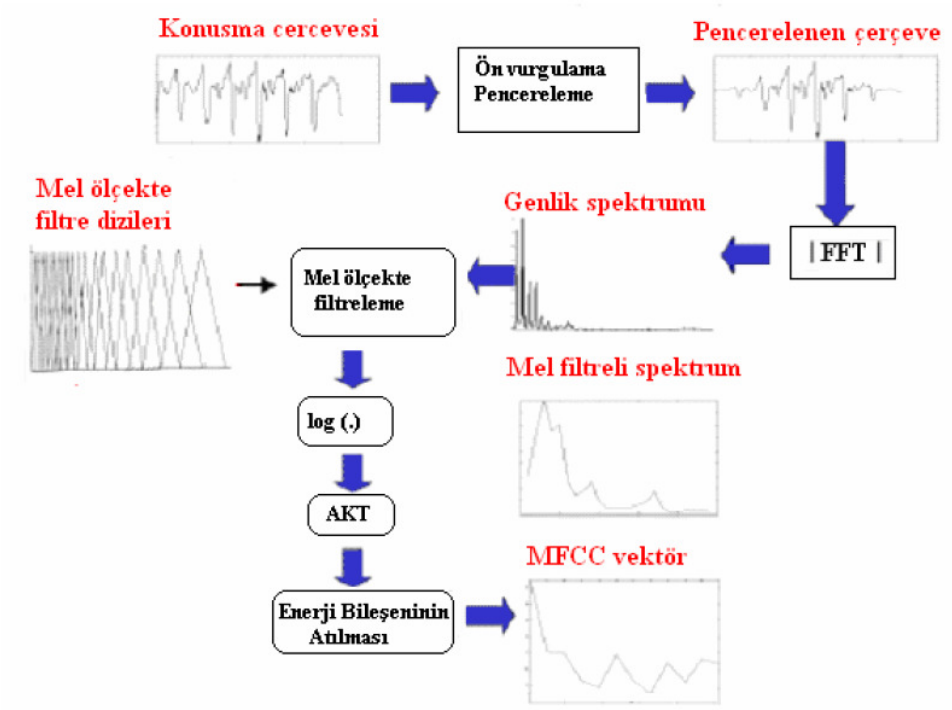
Karakteristiklerin ölçümünde önemli bir varsayım sinyalin milisaniyelik aralıklarla durağan olduğunu kabul etmektir. Bu yüzden sinyal akustik olayların temelini oluşturacak boyutlarda ses çerçevelerine (frame) ayrılır. Saniyede hesaplanan çerçeve sayısı çerçeve oranı olarak adlandırılır. Her çerçevenin başında ve sonunda bulunan devamsızlıkları en aza düşürmek için incelen bir pencere fonksiyonu kullanılır. Kaydırmada bilgi kaybı olmaması için örtüşme (overlapping) uygulanır.

3.1.3 Lineer Dönüşümler (Lineer Transforms)

Lineer dönüşüm fikrinin arkasında bir ölçü setini yeni bir özellik setine dönüştürmek yatar. Eğer dönüşüm doğru seçilirse gereksiz bilgi miktarı azalacaktır. Bu konuda en çok kullanılan dönüşüm Hızlı Fourier Dönüşümüdür (Fast Fourier Transform - FFT). Bunun dışında farklı dönüşümlerde düşünülebilir. Bunlar Ayrık Kosinüs Dönüşümü (Discrete Cosine Transforms - DCT), Modüle Edilmiş Kompleks Dönüşüm (Modulated Complex Transform - MCLT) gibi olabilir.

3.1.4 Özellik Çıkarımı (Feature Extraction)

Frekans gösterimi hazırlanıp, ek dönüşümler uygulanarak son akustik vektörler elde edilir. Bu aşamada geniş yelpazede algoritmalar bulunabilir. Amaç boyutu azaltmak ve aynı zamanda bozulmalara karşı değişmezliği arttırmaktır. İnsan duyu sistemine uyum sağlamak için algısal anlamlı parametreler çıkarılmalıdır. Bu konuda birçok sistem kritik spektrum band analizleri yapar. Genelde konuşma analizinde kullanılan ama müzik tanımaya da uygun olan MFCC (Mel Frequency Cepstrum Coefficents) özelliklerinin çıkarılması şekil 3.2'de görülebilir(Logan, 2000)



Şekil 3.2 MFCC özelliklerinin çıkarılması

3.2 Modelleme

Parmak izi modeli bloğuna genelde çerçevelere (frame) ayrılmış şekilde hesaplanmış özellik vektörleri sırası giriş olarak verilir. Kayıtlarda ve veritabanında çerçeve-zaman ilişkisinde fazlalıkları ortadan kaldırmak parmak izi boyutunu düşürmede faydalıdır. Seçilen model uzaklık ölçüsünü ve hızlı erişim için indeksleme algoritmasını belirler.

Kısa bir şekilde parmak izi oluşturmanın bir yolu bir şarkının vektör zincirini tek bir vektörde özetlemektir. Etaantrum ve Musicbrainz yaptıkları model mp3 dosyalarını tanım bilgilerine eşleme amaçlıdır. Hesaplama olarak oldukça efektif olan bu modeller oldukça kompakt parmak izleri oluşturur. Düşük karmaşıklık için geliştirildikleri için sağlamlık konusunda fazla başarılı değildir.

Ayrıca parmak izi özelliklerin sıralaması olarak modellenebilir. Bu tip parmak izleri ikili vektör dizileri şeklinde bulunabilir (Haitsma ve Kalter, 2002).

Bir diğer yaklaşımdaki çalışmalarda model ayrıca global fazlalıkları da ortadan kaldırır (Cano ve Battle, 2002), (Battle, Masip ve Guaus 2004). Bu mantık büyük ölçüde konuşma tanımadan esinlenmiştir. Konuşmada ses sınıfı alfabesi kullanılması konuşma datasını fazla bilgi kaybı yaşamadan ve fazlalık veriyi ortadan kaldırarak yazıya dönüştürülmesini sağlamaktadır. Benzer şekilde müzik bilgisini de ses sınıflarının sonlu bir alfabesinden oluşan cümleler olarak görebiliriz. Örneğin algısal olarak benzer bir davul sesi bir çok pop şarkısında geçmektedir. Bu yaklaşım parmak izinin, ses parçalarının gösterimi olan bir grup ses sınıfının indeksleri şeklinde oluşmasını sağlar. Ses sınıfları eğitmensiz olarak HMM ile modellenir.

3.3 Uzaklık Hesabı ve Arama Yöntemleri

Bu bölümde kullanılan modelleme tekniğine bağlı olarak değişen uzaklık hesabı ve arama yöntemlerinden bahsedilmiştir.

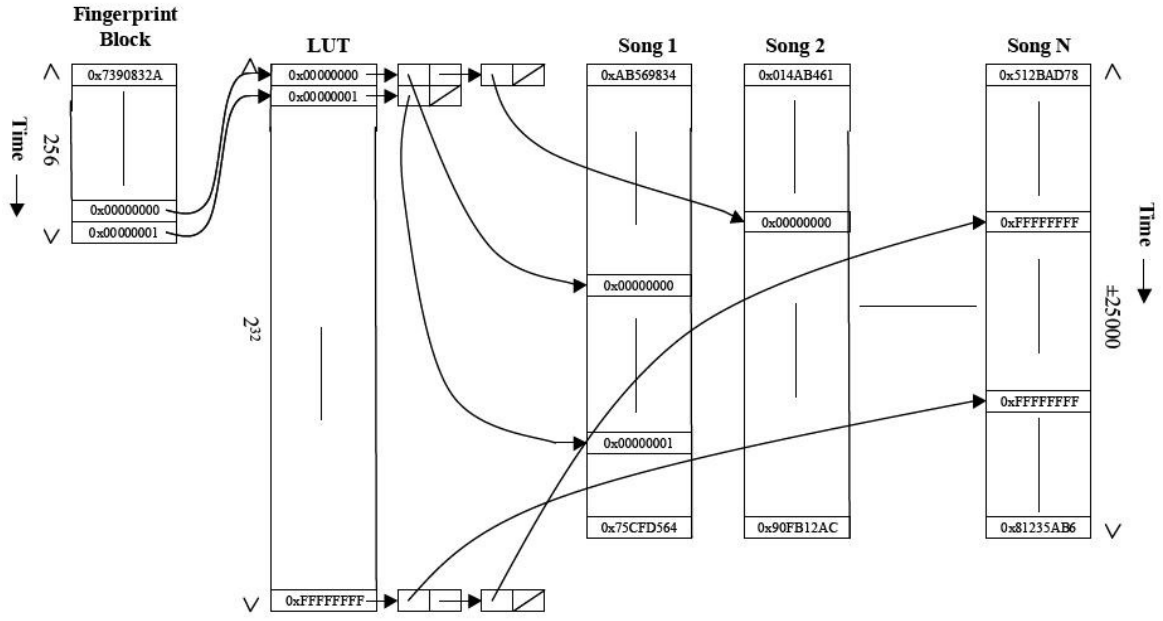
3.3.1 Uzaklık Hesabı

Benzerlik değerlendirme için uzaklık ölçüsü seçilen model ile oldukça ilgilidir. Vektör dizilerini karşılaştırırken genelde iki vektör dizisi arasındaki ilişki hesaplanır. Euclidean uzaklık hesabı ile vektörler arası uzaklık bulunabilir ve knn (k nearest neigbourhood) en yakın komşu algoritmasıyla sınıflandırma yapılabilir. Ayrıca çeşitli durumlarda Manhattan uzaklığı ya da Hamming uzaklığı da kullanılabilir.

Bazı sistemlerde parmak izi yanlızca referanslardan oluşmaktadır. Bu modeller HMM ya da kod kitabı şeklinde kompakt şekilde gerçekleştirilmiştir. Kod kitabı şeklindeki modeller vektör kuantalama ile oluşturulur. Bu durumlarda uzaklık hesapları bilinmeyen ses kaydından alınan özellikler ile depoda bulunan parmak izleri arasında yapılır. HMM sınıfları kullanılan yöntemde, parmak izleri arasında özellik dizisi sorgusu için viterbi algoritması kullanmaktadır. Veritabanındaki en benzer pasaj eşleme olarak seçilir (Cano ve Battle, 2002).

3.3.2 Arama Yöntemleri

Parmak izi karşılaştırılması için uzaklık ölçümünün yanında, bilinmeyen ses kaydının milyonlarca kayıtlık bir veritabanında yapılacak olan karşılaştırmasının ne kadar efektif olacağı da kullanılabilirlik için önemli bir kriterdir. Arama metodu parmak izi gösterimine bağlıdır. Vektör gösteriminde geçerli vektör uzayı arama yöntemleri kullanılır. Genel amaç sorgulama yaparken uzaklık ölçümlerinin sayısını azaltmak için bir veri yapısı geliştirmek ve indekslemektir. Çoğu indeksleme algoritmasında yakınlık araması için eşitlik sınıfları oluşturulur, diğerleri göz ardı edilerek bu sınıflar arasından arama yapılır. Basit uzaklık kullanılmasının ardında bulunan fikir hızlıca birçok sonucu elemek ve indeks kullanarak kaba-kuvvet (brute force) karşılaştırmalarının üstesinden gelmektir. Şekil 3.3 de mümkün olan parmak izi varyasyonlarının bir kontrol (look-up) tablosu kullanılarak şarkılardaki pozisyonları ile indexlenmesi gösterilmiştir.



Şekil 3.3 Kontrol tablosuna dayalı arama metodu (Haitsma ve Kalter, 2002)

4. MÜZİK PARÇASI TANIMA SİSTEMİ TASARIMI

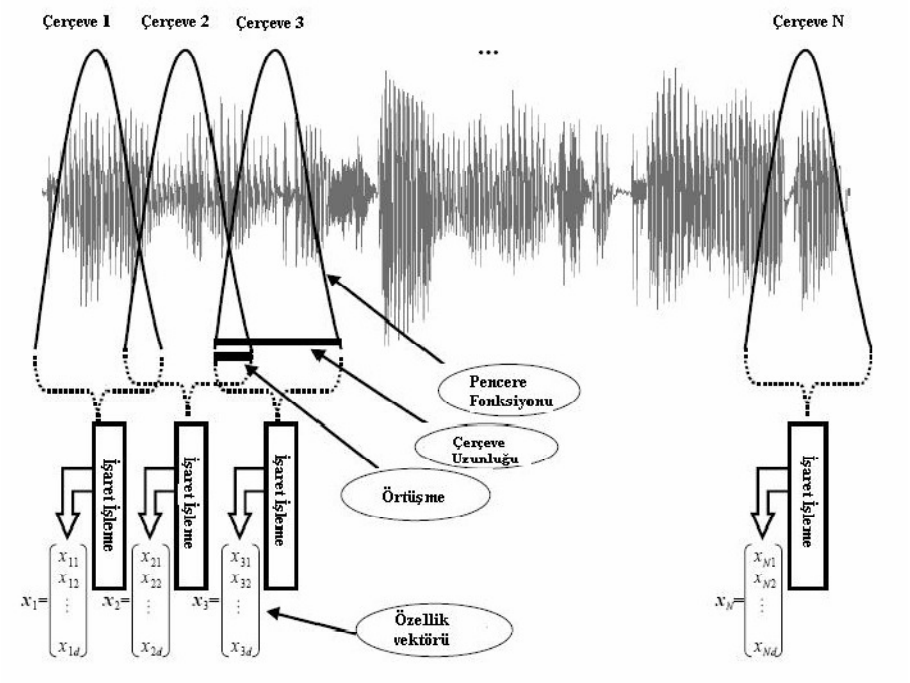
Bu bölümde, bu çalışmada gerçekleştirilen müzik parmak izi tabanlı müzik parçası tanıma sistemi tasarım ve gerçekleme detayları anlatılmıştır. Bu amaçla bu çalışma kapsamında, özellikle sinyalin frekans ve akustik özelliklerine bakılarak her şarkı için parmak izleri oluşturan bir sistem geliştirilmiştir. Sistemin genel amacı bir müzik parçasının yorumcusunu ve müzik parçasının adını tanımlamasıdır. Bunun için sistem müzik dosyasını alır, içeriğini analiz eder, parmak izini çıkarır ve veritabanındaki kayıtlarla karşılaştırır.

Diğer yöntemlerin arasında içerik tabanlı tanıma sistemlerinin diğer dosya karşılaştırma, yardımcı veri ya da damga bilgisi tutma gibi fazladan tanım verisi içeren yöntemlerden daha esnek ve güçlü olduğu bilinmektedir (Cano vd., 2002). İçerik tabanlı tanıma sistemleri müzik sinyalinin akustik özelliklerini temel alır. Bunun için farklı özellik çıkarma, modelleme ve karşılaştırma yöntemleri kullanılan farklı sistemler geliştirilebilir.

Yapılan çalışmada sinyal üzerinde yapılan ön işlemler ardından özellik çıkarmak için MFCC ve modelleme için HMM kullanılmıştır.

4.1 Ön Yüz

Ses işareti, ses yolunun yapısı itibariyle sıkça değiştiği için işaret kısa bölümler halinde işlenmelidir. İşaret yeterince küçük parçalar halinde işlendiği zaman daha kararlı sessel özellikler gösterecektir (Deller vd., 1993). Böylece işaretin kısa dönemli bir bölümünden özellikler çıkarılmış olacaktır. Yapılan bu işleme kısa-dönem analizi denilmektedir. Şekil 4.1 de kısa dönem analizinin adımları gösterilmektedir. Kısa dönem analizinde işaret belirli kısımları örtüşen çerçevelere ayrılır. Örtüşmenin sebebi bilgi kaybını engellemektir. Her bir çerçeve uzunluğu önceden tanımlanmış bir pencere fonksiyonu ile çarpılır. Pencere fonksiyonu ile çarpılan çerçevelere pencerelenmiş çerçeveler de denilmektedir. Müzik tanıma sistemlerinde kullanılan çeşitli pencere fonksiyonu vardır ancak bunlardan en yaygın olanı Hamming Penceresidir (Batlle, Masip ve Guaus, 2001; Logan, 2002).



Şekil 4.1 Kısa dönem analizi (Hanilci, 2007)

Bir çerçeveden elde edilen özellikler kümesine özellik vektörü adı verilir. Kısa dönem analizinden sonra özellik vektörlerinin elde edilmesinde değişik yöntemler kullanılmaktadır. Bu çalışmada müzik parçasını karakterize eden özellikler olarak MFCC kullanılmaktadır.

Müzik parçası tanımada ön işlemler olarak nitelendirilen bölüm analog ses sinyalinin sayısallaştırılması, yüksek frekans bileşenlerinin kuvvetlendirilmesi ve seçilen örnekleme frekansı doğrultusunda çerçeveleme için uygun boyutun, kaydırma boyutunun ve pencereleme fonksiyonunun belirlenmesini içerir.

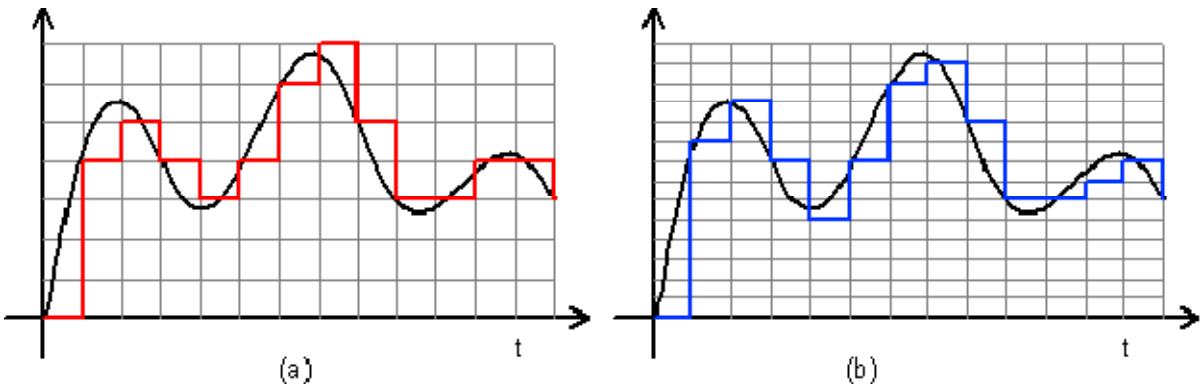
Ses sinyali yaklaşık olarak 10-30 ms'lik düzeylerde durağan kabul edilmektedir (Karpov 2003). Bu sebeple seçilen örnekleme frekansı doğrultusunda sonraki adımların gereksinimleri de düşünülerek 10-30 ms'yi sağlayacak boyutta çerçeve boyutu seçilmelidir.

4.1.1 Ön İşleme

Analog ses sinyali analog-sayısal dönüştürücüler aracılığı ile sayısal sinyale dönüştürülür. Dış ortamlardan alınan ses verisi için bu işlemi ses kayıt imkanı sağlayan cihazlar yapabilmektedir. Müzik tanıma probleminin çeşitli boyutlarından hangisi ile ilgileniliyorsa ona ilişkin analog-sayısal dönüştürücü kullanılmalıdır. Çevre şartlarının ve fon gürültüsünün etkisini azaltmak için özel gürültü engelleyici mikrofonlar kullanılabilir ya da müzik verisi kablo üzerinden alınabilir. Ses verisinin sayısallaştırılmasında bir başka boyut örnekleme

frekansının seçimidir. Örnekleme teorisinin gereği bir sinyal en yüksek frekansının en az iki katı ile örneklenirse sinyal genel karakteristiğini kaybetmemiş olur ve sinyal örnekler üzerinden yeniden elde edilebilir. İnsan kulağı 20 Hz ile 20 kHz frekans aralığındaki ses sinyalini duyabilir. Bu bakımdan analog ses sinyali 20 kHz'lik alçak geçiren filtreden geçirildikten sonra 40 kHz ile örneklenirse insan duyması simüle edilmiş olur. (Rabiner ve Juang, 1993).

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



Şekil 4.2 Genlik - Zaman görünümünde çözünürlüğün örneklemeye etkisi

4.1.2 Çerçeveleme

Ses sinyali, parametrelerin sabit kaldığı kabul edilen çerçeve olarak adlandırılan küçük parçalara ayrılmalıdır. Çünkü tüm işaret boyunca FFT hesaplanırsa, farklı müzik birimlerine ait spektral bilgilerin tutulmasında kayıplar oluşur. Bu nedenle tüm işaretin FFT'sini almak yerine çerçevenin FFT'si hesaplanır. Çerçeve uzunluğu 10-30 msn arasında degişir. Bu aralıkta ses oldukça sabit akustik karakteristik gösterir (Karpov, 2003). Her bir çerçeveye örtüşme uygulanır. Çerçevelerin örtüşme oranı, çerçeve uzunluğunun % 30'u ile % 75' i arasında alınır (Kinnunen, 2003). Örtüşme uygulanması ile çerçeve sonundaki işaretlerin önemlerini kaybetmemesi saglanır.

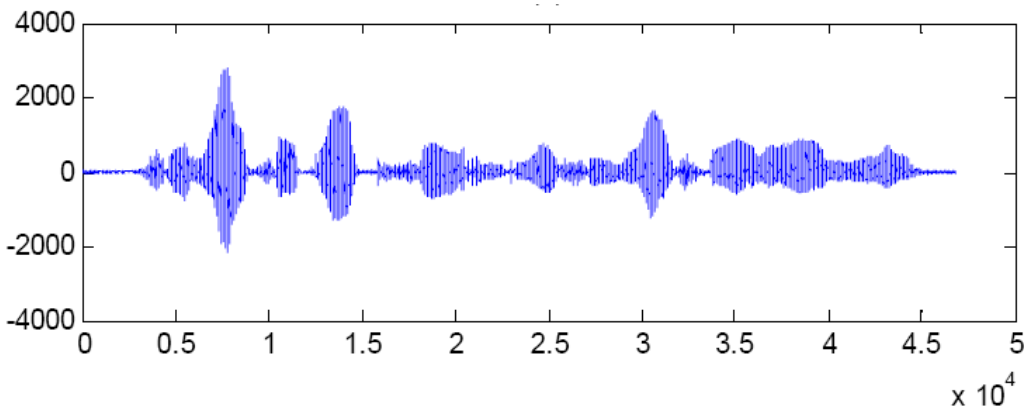
Ses işareti N adet örnekten olusan çerçevelere ve komşu M örnekten oluşan çerçevelere bölünür ($M < N$). İlk çerçeve N örnekten oluşurken ikinci çerçeve ilk çerçeveden M örnek sonra başlar ve ilk çerçevenin N-M çerçeve kadar üzerine biner (Rabiner ve Juang 1993).

4.1.3 Ön vurgulama (Pre-emphasis)

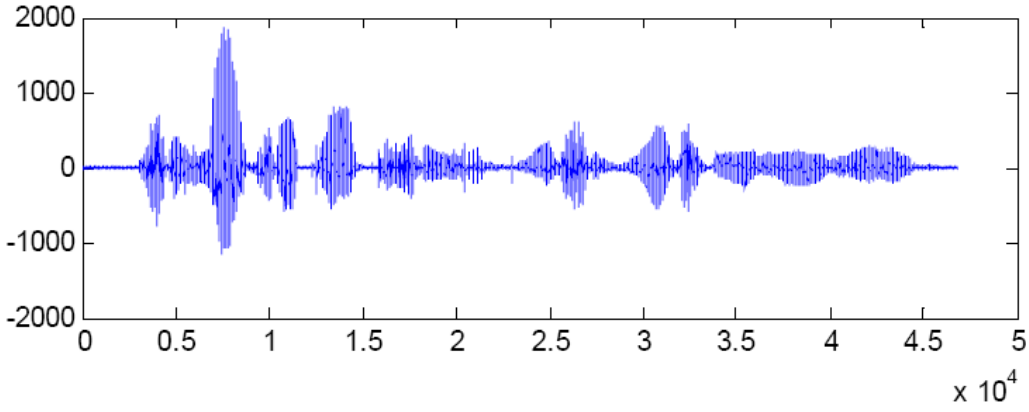
Ön vurgulama işleminde giriş işareti birinci dereceden bir FIR (Finite Impulse Response) süzgeç girişine uygulanır. Birinci dereceden süzgecin transfer fonksiyonu,

$$H(z) = 1 - 0.95z^{-1} \quad (4.1)$$

şeklindedir. Ön vurgulama işleminin amacı sinyalin yüksek frekans bileşenlerini daha baskın hale getirmektir. Şekil 4.3 orijinal ses işaretini ve şekil 4.4 ön vurgulama işlemi yapıldıktan sonra süzgeç çıkışında elde edilen işareti göstermektedir.



Şekil 4.3 Orjinal ses sinyali



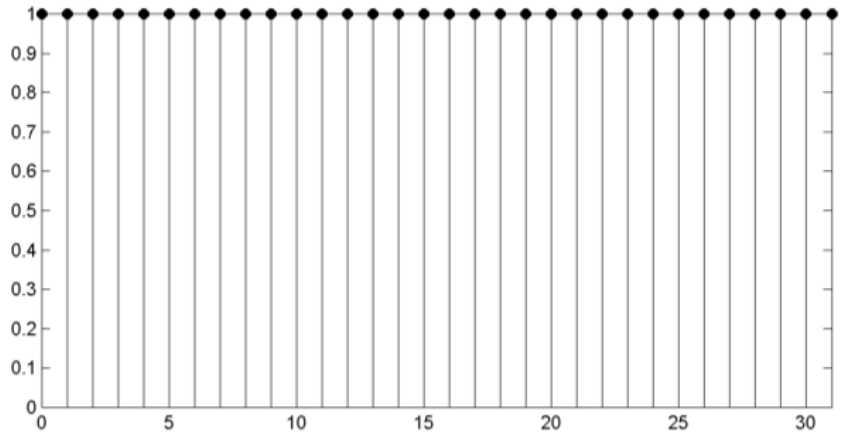
Şekil 4.4 Ön vurgulama ardından elde edilen sinyal

4.1.4 Pencereleme

Pencerelemenin amacı çerçeveleme işlemi sonucunda olusan spektral etkilerin azaltılmasıdır. Pencereleme ile çerçevelerde süreksizliğin önüne geçilir (Rabiner ve Juang, 1993). Bu sayede sesin orta bölgeleri güçlendirilirken kenar bölgeleri zayıflatılır. Pencereleme işlemlerinde

çeşitli fonksiyonlar uygulanabilir. Bunlara örnek olarak dikdörtgen, Gauss, Hamming, Hann, Bartlett, üçgen, Bartlett-Hann, Blackman, Kaiser, Nuttal, Blackman-Harris ve Blackman-Nuttal fonksiyonları verilebilir.

Dikdörtgen pencere fonksiyonu denklem (4.2) ile belirtilmiş olup zaman boyutundaki yapısı şekil 4.5 ile verilmiştir.

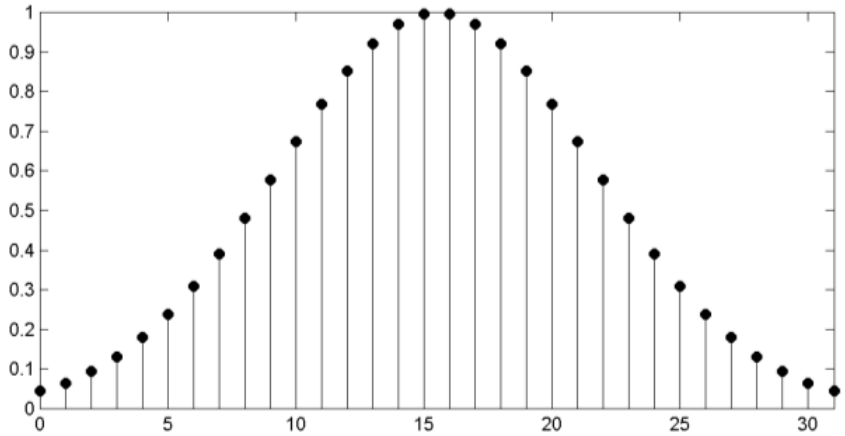


Şekil 4.5 N=32 için dikdörtgen pencere fonksiyonu

$$w(n) = 1 \tag{4.2}$$

σ standart sapma olmak üzere Gauss pencere fonksiyonu denklem (4.3) ile belirtilmiş olup zaman boyutundaki yapısı (şekil 4.6) ile verilmiştir.

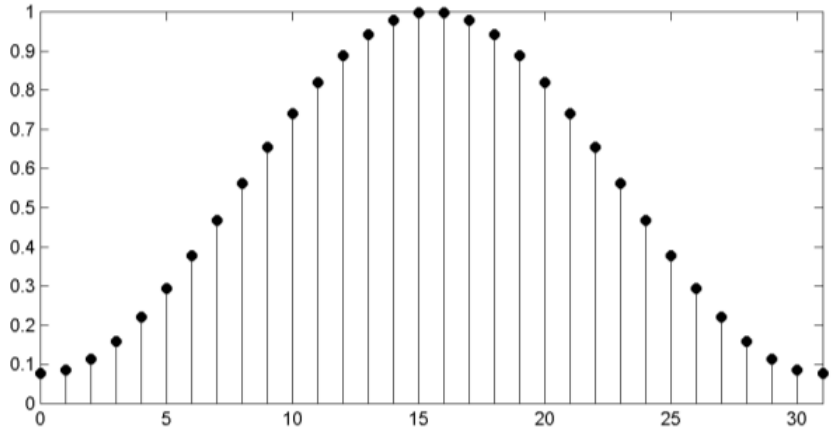
$$w(n) = e^{-\frac{1}{2}\left(\frac{n-\frac{N-1}{2}}{\sigma\frac{N-1}{2}}\right)^2} \tag{4.3}$$



Şekil 4.6 N=32, σ = 0.4 için Gauss pencere fonksiyonu

Hamming pencere fonksiyonu denklem (4.4) ile verilmiş olup zaman boyutundaki yapısı (şekil 4.7) ile verilmiştir.

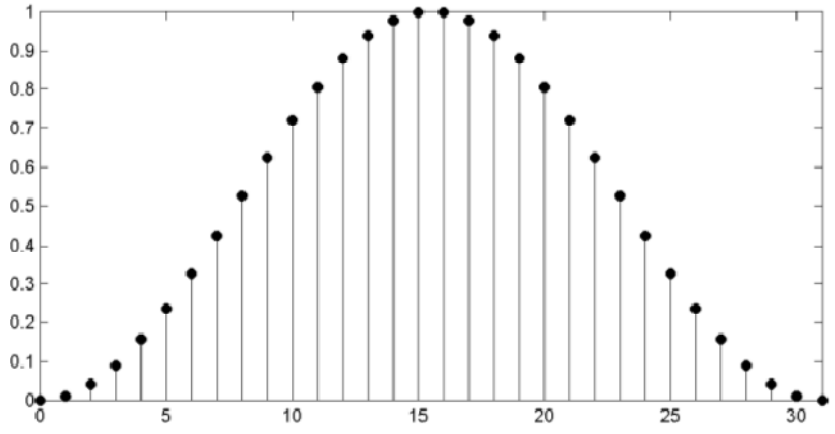
$$w(n) = 0{,}53836 - 0{,}46164 \cos\left(\frac{2\pi n}{N-1}\right) \tag{4.4}$$



Şekil 4.7 N=32 için Hamming pencere fonksiyonu

Hann pencere fonksiyonu denklem (4.5) ile verilmiş olup zaman boyutundaki yapısı şekil 4.8 ile verilmiştir.

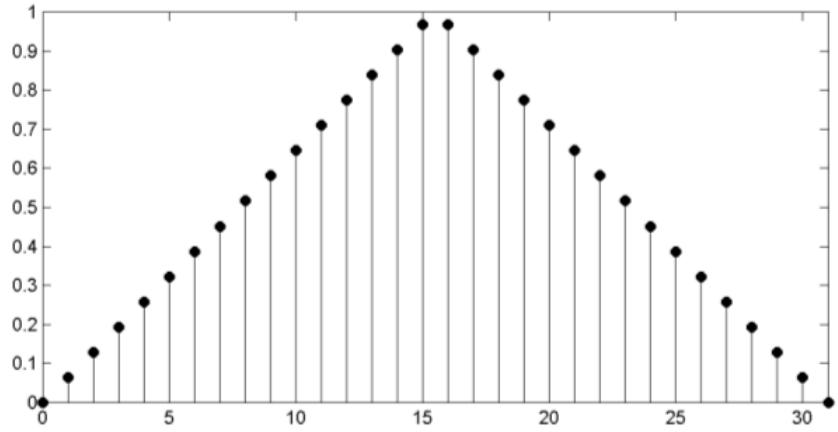
$$w(n) = 0{,}5\left(1 - \cos\left(\frac{2\pi n}{N-1}\right)\right) \tag{4.5}$$



Şekil 4.8 N=32 için Hann pencere fonksiyonu

Bartlett pencere fonksiyonu denklem (4.6) ile verilmiş olup zaman boyutundaki yapısı (şekil 4.9) ile verilmiştir. Bartlett pencere fonksiyonu başlangıç ve bitiş noktalarında sıfır değerini alan üçgen pencere fonksiyonudur. Bartlett pencereleme işlemi özellik çıkarmada kepstral katsayılar hesaplanırken filtre bankası şeklinde uygulanır.

$$w(n) = \frac{2}{N-1} \cdot \left(\frac{N-1}{2} - \left| n - \frac{N-1}{2} \right| \right) \tag{4.6}$$



Şekil 4.9 N = 32 için Barlett pencere fonksiyonu

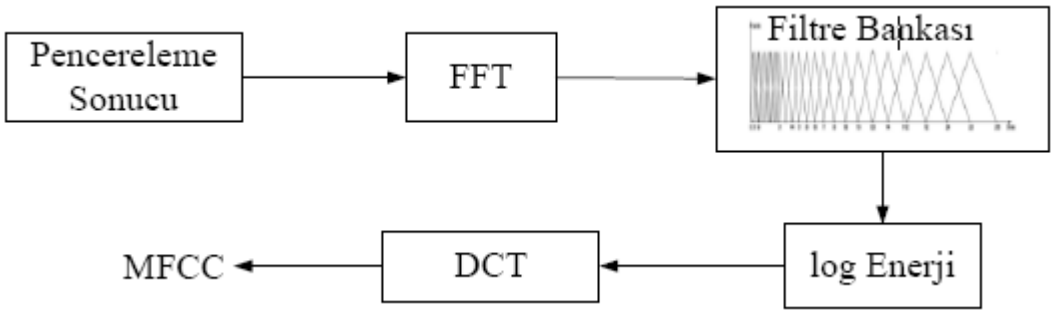
Tüm bu formülasyonlarda $0 < n < N$ koşulu geçerlidir ve N pencere boyutunu göstermektedir.

Müzik tanıma açısından pencerelemenin önemi, kısa zamanlı ayrık Fourier dönüşümü yapıldığında, pencereleme sonucu sinyalin başlangıç ve bitiş değerlerini birbirine yaklaştırması ile pencere uçlarında oluşabilecek süreksizlikleri önlemesindedir. Bu sayede elde edilecek kısa zamanlı frekans spektrumu süreksizliklerden kaynaklanacak gereksiz spektrum verisinden ayıklanmış olacaktır. Bahsedilen duruma uygun düşecek şekilde müzik tanımada en çok kullanılan pencereleme fonksiyonları Hamming ve Hann pencere fonksiyonlarıdır. Hamming ve Hann pencere fonksiyonları uygulanarak pencere uçlarındaki süreksizlikler giderilebilmekte fakat uç noktaların özellikleri kaybedilmektedir. Bu sebeple pencereleme birbiri ardınca değil pencere boyutundan daha kısa boyutta kaydırılarak uygulanır. Kaydırma genel olarak pencere boyutunun yarısı kadar uygulanır.

4.2 MFCC ile Özellik Çıkarma

Ses verisi için çeşitli özelliklerden bahsedilebilir. Bir kısım özellikler genlik, sıfır geçiş sayısı, frekans spektrumu, ses perdesi (pitch) ve sinyalin taşıdığı enerji gibi ses dalgası ile ilişkilendirilebilecek özelliklerdir. Bunun yanında MFCC, otokorelasyon katsayıları, LPC (Linear Predictive Coding - Doğrusal Öngörülü Kodlama) katsayıları gibi bir takım sinyal işleme teknikleri ya da istatistiksel işlemlerle elde edilebilecek özelliklerden de bahsedilebilir.

Müzik tanıma açısından başarılı sonuçlar vermiş olan mel frekans kepstral katsayılarının elde edilmesi işlemi ses verisinin belirli frekans bantlarına düşen güç değerlerinin bir ortalamasının alınarak bantlar arası lineer bağımsız katsayıların hesaplanması ile gerçekleştirilmektedir (Logan, 2002). Bu amaçla zaman boyutundaki ses verisi için frekans spektrumu ayrık Fourier dönüşümü ile bulunur. İnsan duyması simüle edilecek şekilde 1-2 kHz civarına kadar lineer olarak eşit daha sonrası için logaritmik olarak eşit aralıklı örtüşümlü filtre bankası uygulanarak belirli frekans bantlarına düşen ortalama güç yaklaşık olarak bulunur. Ortalama güç değerleri logaritmik olarak yeniden ifade edilerek frekans spektrumu normalize edilir.



Şekil 4.10 Özellik çıkarım adımları blok şeması (Hanilci, 2007)

Ayrık kosinüs dönüşümü ile de hesaplanan özellik değerleri birbirlerinden lineer bağımsız bir şekilde dönüştürülürler. Uygulamalarda mel frekans kepstral katsayılarının yanında birinci dereceden ve ikinci dereceden fark katsayıları da hesaplanarak özellik vektörü olarak kullanılmaktadır.

4.2.1 Spektral Analiz

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

Genel olarak bir sinyalin zaman boyutundan frekans boyutuna geçirilmesi dönüşüm olarak adlandırılmaktadır. Temel olarak zaman boyutunda sürekli işaretler için frekans dönüşümü yine sürekli bir işaret üretmektedir, benzer şekilde tanım gereği zaman boyutunda ayrık işaretler için de frekans dönüşümü sürekli bir sinyal üretmektedir.

Zaman boyutunda sürekli sinyallerin frekans dönüşümleri için Fourier dönüşümü ve Laplace dönüşümü kullanılır. Tanım olarak Fourier dönüşüm çifti $\{X(\omega), x(t)\}$ denklem (4.7) ve (4.8) ile, Laplace dönüşüm çifti $\{F(s), f(t)\}$ ise denklem (4.9), (4.10) ile verilmiştir.

$$X(\omega) = \int_{-\infty}^{\infty} x(t) e^{-j\omega t} dt \tag{4.7}$$

$$x(t) = \frac{1}{2\pi} \int_{-\infty}^{\infty} X(\omega) e^{j\omega t} dt \tag{4.8}$$

$$F(s) = \int_{0^-}^{\infty} e^{-st} f(t) dt \qquad s = \sigma + j\omega \tag{4.9}$$

$$f(t) = \frac{1}{2\pi j} \int_{\gamma - j\infty}^{\gamma + j\infty} e^{st} F(s) ds \qquad \begin{matrix} s = \sigma + j\omega \\ \gamma \in \Re \end{matrix} \tag{4.10}$$

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

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

Zaman boyutunda ayrık sinyallerin frekans boyutuna geçirilmesinde ayrık zamanlı Fourier

dönüşümü ve z dönüşümü kullanılır. z dönüşümü Laplace dönüşümünün ayrık zamanlı sinyaller için uygulanan benzeridir. Ayrık zamanlı Fourier dönüşüm çifti denklem (4.11), (4.12) ile verilmiştir.

$$X(\omega) = \sum_{n=-\infty}^{\infty} x[n] e^{-j\omega n} \tag{4.11}$$

$$x[n] = \frac{1}{2\pi} \int_{-\pi}^{\pi} X(\omega) e^{j\omega n} d\omega \tag{4.12}$$

Daha önce bahsedildiği gibi zaman boyutunda ayrık işaretin frekans boyutuna geçirilmesi sonucu sürekli bir fonksiyon üretilmektedir. Bilgisayar ortamında bu verilerin işlenebilmesi için tıpkı analog ses verisinin örneklenmesi gibi, ayrık zamanlı Fourier dönüşümündeki ω frekans bileşeni de bir periyot boyunca $\frac{1}{N}$ aralıklarla örneklenerek kullanılır. Bu şekildeki uygulama ayrık Fourier dönüşümü olarak adlandırılır ve denklem (4.13), (4.14) ile tanımlanmaktadır.

$$X[k] = \sum_{n=0}^{N-1} x[n]\, e^{-j\frac{2\pi k}{N} n} \qquad 0 \le k \le N-1 \tag{4.13}$$

$$x[n] = \frac{1}{N} \sum_{k=0}^{N-1} X[k] e^{j\frac{2\pi k}{N} n} \qquad 0 \le n \le N-1 \tag{4.14}$$

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

Ayrık Fourier dönüşümü hesaplama işlemi denklem (4.12) ile verildiği şekliyle N^2 çarpma ve toplama gerektirmektedir. 1965 yılındaki çalışmaları ile Cooley ve Tukey ayrık Fourier dönüşümü için karmaşıklığı r^p örnekli bir veri dizisi için $rN \log_r N$ mertebesine indirgeyecek hızlı Fourier dönüşümünü ortaya koymuşlardır. Denklem (4.14), (4.15) ile verilen tanımlar ve N'in

$N = r_1 r_2$ şeklinde çarpanlarına ayrılabildiği varsayımıyla ayrık Fourier dönüşümü denklem (4.16) ile verilen şeklide yeniden yazılabilir.

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

$$\begin{aligned} k &= k_1 r_1 + k_0 \qquad k_0 = 0,1,..r_1 - 1 \qquad k_1 = 0,1,..r_2 - 1 \\ n &= n_1 r_2 + n_0 \qquad n_0 = 0,1,..r_2 - 1 \qquad n_1 = 0,1,..r_1 - 1 \end{aligned} \tag{4.15}$$

$$X[k_0, k_1] = \sum_{n_0=0}^{r_2-1} \sum_{n_1=0}^{r_1-1} x[n_0, n_1] W^{k n_1 r_2} W^{k n_0} \tag{4.16}$$

Ayrıca denklem (4.16) ile verilen terim için yapılabilecek indirgemeye yer verilmiştir.

$$W^{k n_1 r_2} = W^{(k_0 + k_1 r_1) n_1 r_2} = W^{k_0 n_1 r_2} W^{k_1 n_1 r_1 r_2} = W^{k_0 n_1 r_2} W^{k_1 n_1 N} = W^{k_0 n_1 r_2} \underbrace{e^{-2\pi j k_1 n_1}}_{1} = W^{k_0 n_1 r_2} \tag{4.17}$$

$$X[k_0, k_1] = \sum_{n_0=0}^{r_2-1} \underbrace{\sum_{n_1=0}^{r_1-1} x[n_0, n_1] W^{k_0 n_1 r_2}}_{A[n_0, k_0]} W^{k n_0} \tag{4.18}$$

Denklem (4.18) ile verilen ayrık Fourier dönüşümüne ilişkin $A[n_0, k_0]$ teriminin her bir elemanını hesaplamak r_1 işlem gerektirmektedir. A için $n_0 k_0 = N$ eleman hesaplanabileceği düşünülürse, A teriminin hesaplanması $N r_1$ karmaşıklıktadır. $X[k_0, k_1]$ teriminin verilen bir A terimi üzerinden hesaplanmasında ise herbir eleman için r_2 işlem gerekmektedir. X $k_0 k_1 = N$ için eleman hesaplanması gerektiği düşünülürse, X teriminin hesaplanabilmesi için $N r_2$ karmaşıklık gerekmektedir. Toplamda ise X teriminin hesaplanması için N^2 karmaşıklık yerine $N(r_1 + r_2)$ mertebesinde bir karmaşıklık yeterli olabilmektedir. N'in çok sayıda çarpanı olan bir bileşik sayı olması durumunda ise örneğin $N = r^p$ şeklinde yazılabiliyorsa hızlı Fourier

dönüşümü ile $rN\log_r N$ karmaşıklık mertebesine indirgenebilir. Genel olarak $r = 2$ seçilmesi durumunda ise (Cooley ve Tukey, 1965) ikili bilgisayar sistemlerinin adresleme ve çarpma işlemlerinde kolaylık sağlayacaklarını söylemektedir. (Şekil 4.11) ile hızlı Fourier dönüşümünün gerçekleştirilmesine ilişkin adımlar verilmiştir.

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

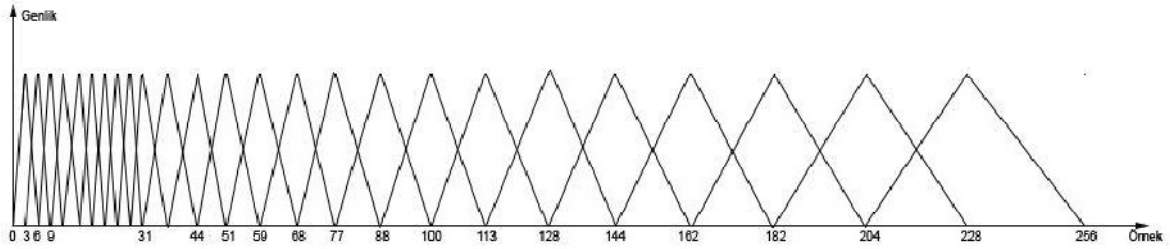
Şekil 4.11 Hızlı fourier dönüşümü adımları (Cooley ve Tukey, 1965)

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

4.2.2 Filtre Bankası Uygulama

Frekans boyutunda hesaplanan genlik değerleri kullanılan pencereleme fonksiyonu boyutunda elde edilmektedir. Özellik vektörünün bu boyutu ile kullanılması yerine belirli bantlara düşen ortalama genlik değerleri alınarak hem özellik vektörü boyutu düşürülmüş olmakta hem de belirli bir uygunlaştırma, filtreleme sağlanmaktadır.

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



Şekil 4.12 Filtre bankası örneği

Şekil 4.12 ile 16 kHz ile örneklenmiş bir sinyal için uygulanabilecek filtre bankasına ilişkin bir örnek verilmektedir. Bu örnekte sinyal 16 kHz ile örneklendiği için en yüksek frekansı 8 kHz olarak kabul edilir. Kullanılan pencereleme fonksiyonu boyutu 512 olduğu durumda, FFT'den elde edilecek 512 örneklik simetrik genlik katsayısından kullanılacak olan ilk 256 örneğin hangi filtrede hesaba katılacağı yatay eksendeki örnek sınırlarıyla verilmiştir. Burada 256'ıncı örnek 8kHz frekansına denk gelmektedir.

Logaritmik eşit aralıkların oluşturulmasında yine insan duymasından hareketle ortaya konmuş olan mel frekans ölçeği kullanımı söz konusudur. Frekans skalası ile mel frekans skalası arasındaki dönüşüm denklem (4.19) ve (4.20) ile verilmektedir.

$$\mathrm{m} = 2595 \cdot \log(1 + \frac{f}{700}) \tag{4.19}$$

$$f = 700 \cdot (10^{\frac{\mathrm{m}}{2595}} - 1) \tag{4.20}$$

4.2.3 Log Enerji

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

4.2.4 Ayrık Kosinüs Dönüşümü

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

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

Kullanılan K adet filtre sonucu elde edilen değerler X_k (k=1, ..., K) ile gösteriliyor ise log enerji ve ayrık kosinüs dönüşümü sonucu birlikte denklem (4.21) ile verilebilir (Young,).

$$MFCC(i) = \sum_{k=1}^{K} \log(X_k) \cos(i(k - \frac{1}{2}) \frac{\pi}{K}) \tag{4.21}$$

4.2.5 Özellik Vektörünün Oluşturulması

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

4.2.6 Özellik Vektörü Normalizasyonu

Özellik vektörleri üzerinde eğitim ve test ortam farklılıkları ve bazı gürültü etkilerini azaltmak amacıyla normalizasyon işlemleri yapılır. Bu teknikleri CMS(Cepstral Mean Substraction – Kepstral Ortalama Çıkarma) ve MN(Maximum Normalizing – Maksimum Normalizasyonu) vektör normalizasyon teknikleridir.

MN normalizasyonunda özellik vektörü kümesinin her satırı o satırın mutlak maksimum değerine bölünür.

CMS normalizasyonunda özellik vektörü kümesinin her satırında o satırın ortalama değeri çıkartılır.

4.3 Modelleme

Müzik verisinden elde edilmiş özelliklerin ses birimleri boyutunda sonlu durum makinesi şeklinde istatistiki olarak modellenmesi akustik model olarak adlandırılır. Ses birimleri boyutunda modelleme HMM ile yapılır. Bu bölümde HMM ve HMM'in temelini oluşturan Markov zincirleri ve Markov modeli ile ilgili bilgi verildikten sonra gerekli parametrelerin hesaplanmasının nasıl yapıldığı anlatılacaktır.

4.3.1 Markov Modelleri

Sınıflama yöntemleri içerik bağımlı ve içerik bağımsız olmak üzere iki gruba ayrılır. İçerik bağımsız sınıflamada bir deney sonucunda ortaya çıkan özelliklerin bağımsız olduğu kabul edilir. Bu varsayım sınıflar arasında ilişki olmadığı anlamına gelir. Ses tanıma, müzik tanıma, konuşmacı tanıma gibi uygulamalarda ardışıl özellik vektörleri birbirinden bağımsız değildir. Bu nedenle sınıflama bütün özellik vektörleri aynı anda kullanılarak yapılmalıdır. Markov Modelleri içerik bağımlı sınıflandırıcılar grubuna girmektedir (Theodoridis ve Koutroumbas 2003). Bu türden sınıflandırıcıların temel başlangıç noktası Bayes sınıflandırıcılardır. Diğer bir deyişle bir x özellik vektörü, $\boldsymbol{P(w^i|x) > P(w^j)} \quad \boldsymbol{\forall j \neq i}$ ise bir gözlem dizisine karşılık gelen muhtemel bir sınıf dizisi olsun. Bu türden sınıf dizilerinin toplam sayısı M^K kadardır. Sınıflandırma yapmaktaki amacımız "hangi Ω_i sınıf dizisi, X gözlem dizisi ile uyumludur?" sorusuna cevap vermektir. Bu yaklaşım, x_1 dizisinin w_{i_1} sınıfına, x_2 dizisinin w_{i_2} sınfına… atanmasına karşılık gelir. Bayes kuralına göre bir X gözlemi Ω_i sınıfına aşağıdaki şartlarda atanır. şartını sağlıyorsa w_i sınıfına atanır. K adet gözlemden oluşan bir dizi (özellik vektörü) $X = x_1, x_2, \ldots, x_k$ ve bu gözlem vektörlerinin alacağı sınıflar w_i, $(i = 1,2, \ldots . M)$ olsun;

$$P(\Omega_i|X) > P(\Omega_j|X) \qquad \forall\, j \neq i \tag{4.22}$$

Bu ifade aşağıdaki ifadeye denktir.

$$P(\Omega_i)p(\Omega_i|X) > P(\Omega_j)p(\Omega_j|X) \quad \forall\, i \neq j \tag{4.23}$$

Bayes sınıflandırıcısı ile bir X gözleminin ait olduğu sınıfa karar verebilmek için M_K adet sınıf dizisi üzerinden olasılık hesabı yapmak ve bunlardan maksimum olasılık veren sınıf dizisini elde etmek gerekmektedir. Birçok uygulama için bu tür bir yaklaşım hesaplama karmaşası yüzünden oldukça zordur. Bunun yerine sınıflar arası ilişkiyi kullanan modeller kullanılabilir.

Markov modelleri en çok kullanılan içerik bağımlı sınıflandırıcılardan biridir. Bir $W_{i_{1,}}, W_{i_{1,}}, \ldots, W_{i_k}$ sınıf dizisi için Markov modeli aşağıdaki varsayımı yapmaktadır.

$$P(w_{i_k}|w_{i_{k-1}}, w_{i_{k-2}}, \ldots . . w_{i_1}) = P(w_{i_k}|w_{i_{k-1}}) \tag{4.24}$$

Bu varsayımın anlamı sınıflar arası bağımlılık sadece birbirini takip eden iki sınıf ile sınırlıdır.

Bu tip bir model birinci dereceden Markov model olarak adlandırılır. (4.24) denklemini olasılığın zincir kuralı ile yeniden yazarsak aşağıdaki ifade elde edilir.

$$P(\Omega_i) \equiv P(w_{i_1}, w_{i_2}, ..., w_{i_N})$$

$$= P(w_{i_N} \left| w_{i_{N-1}}, ..., w_{i_1}) P(w_{i_{N-1}} \right| w_{i_{N-2}}, ..., w_{i_1}) ... P(w_{i_1})$$

Buradan

$$P(\Omega_i) = P(w_{i_1}) \prod_{k=2}^{N} P(w_{i_k} \left| w_{i_{k-1}}) \right. \tag{4.25}$$

elde edilir. Gözlemlerin istatistiksel olarak bağımsız olduğu bir sınıf dizisi için

$$P(X \left| \Omega_i) P(\Omega_i) = P(w_{i_1}) p(x_1 \right| w_{i_1}) \prod \prod P(w_{i_k} \left| w_{i_{k-1}}) p(x_k \right| w_{i_k}) \tag{4.26}$$

Şeklinde ifade edilebilir.

4.3.2 Saklı Markov Modeli

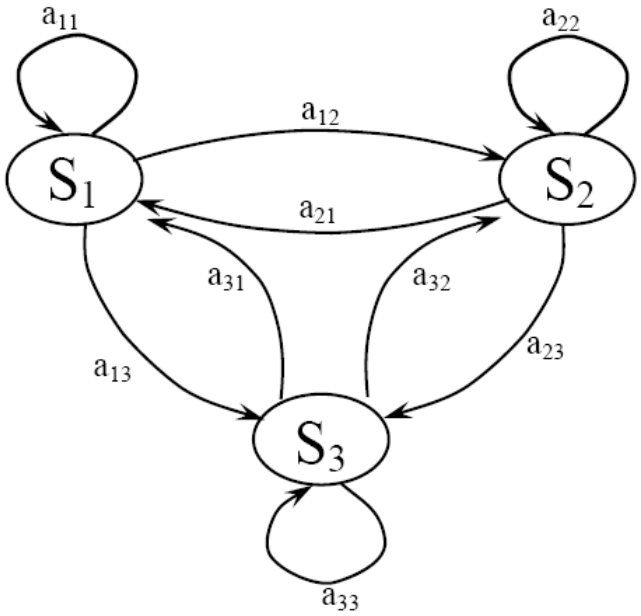
Saklı Markov Modelleri (HMM) bir örüntüye ait çerçevelerin spektral özelliklerini temsil etmede sıklıkla kullanılan ve en çok bilinen yöntemlerden biridir. HMM'nin (veya diğer istatistiksel yöntemler) temelinde yatan en önemli varsayım, ses sinyalinin parametrik bir rastsal süreç olarak iyi bir şekilde karakterize edilebilmesi ve bu rastsal süreç parametrelerinin doğru bir şekilde hesaplanabilmesidir (Rabiner ve Juang, 1993).

HMM'nin temel kuramı Baum ve arkadaşları tarafından 1960'ların sonlarında belirtilmiştir. Konuşma tanıma uygulamalarında ise 1970'lerde Baker, Jelinek ve arkadaşları tarafından kullanılmıştır. HMM, konuşma tanıma uygulamalarında en gelişmiş tekniktir (Juang vd., 1985; Rabiner vd., 1988). HMM ayrıca müzik parçası tanıma uygulamalarında da kullanılmaktadır.

HMM sınıflar arası ilişkiyi tanımlamada kullanılan bir modelleme tekniğidir. HMM stokastik bir modeldir ve istatistiksel özellikleri zamanla değişen dizilerin modellenmesinde sıkça kullanılmaktadır. HMM'de durumlar doğrudan gözlenemez. Bir durumda hangi

sembolün/gözlemin üretileceği başka bir olasılık ifadesi ile belirlenir. Bu modelde bir takım en iyileme tekniklerinin gözlem dizilerine uygulanması sonucu en yüksek olasılıklı durum dizisi elde edilir.

HMM basitçe gözlem dizisi $(x_1, x_2, \ldots, x_k)$ üreten ve sonlu sayıda durumdan oluşan bir rastsal süreçtir. Bu nedenle HMM belirli sayıda durumdan ve gözlem dizisinden oluşmaktadır. Gözlem dizisi bir durumdan diğer bir duruma geçişin bir sonucudur. Şekil 4.13'de 3 durumlu bir HMM gösterilmektedir.



Şekil 4.13 3 durumlu markov modeli

Bu şekildeki bir Markov model ergodik model olarak ifade edilmektedir. Ergodik modelin anlamı her durumdan her duruma geçişin olmasıdır. Şekilde gösterilen modelde a_{ij} 'ler durumlar arası geçiş olasılıklarını belirtmektedir. Matematiksel olarak durumlar arası geçiş olasılıkları şu şekilde ifade edilmektedir.

$$a_{ij} = P[q_t = S_j \,|\, q_{t-1} = S_i] \qquad 1 \le i, j \le N \tag{4.27}$$

Durum geçiş olasılıkları aşağıdaki şartı sağlamalıdır.

$$a_{ij} \ge 0$$

$$\sum_{j=1}^{N} a_{ij} = 1$$

Denklem (4.27) 'de N durum sayısını göstermektedir. Durum geçiş olasılıkları durum geçiş matrisi ile ifade edilmektedir. Şekil 4.11 deki 3 durumlu Markov modeli için durum geçiş matrisi şu şekildedir:

$$A = \begin{bmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{bmatrix}.$$

Bir λ HMM modeli aşağıdaki parametrelerden oluşmaktadır:

Modelde kullanılan durum sayısı N.

Gözlem sembol sayısı M.

Durumlar arası geçiş olasılık matrisi A.

Gözlem sembol olasılık matrisi B.

Başlangıç durum olasılıkları π.

Tüm bu parametreler yardımıyla bir HMM modeli şu şekilde belirtilmektedir:

$$\lambda = (A, B, \pi) \tag{4.28}$$

Bu sayede modelin tüm parametreleri belirtilmiş olur. Durum geçiş olasılık matrisi A olan ve başlangıç durum olasılıkları verilen bir sistemde her hangi bir durum dizisinin bu model tarafından üretilme olasılığı aşağıdaki ifadeden elde edilir:

$$P(q|A, \pi) = \pi_{q0} a_{q0q1} a_{q1q2} \ldots a_{qT-1qT} \quad . \tag{4.29}$$

O gözleminin (özellik vektörü) üretilme olasılığı şu şekilde ifade edilir (Rabiner ve Juang 1993).

$$b_i(O_t) = P(O_t | q_t = S_i) \tag{4.30}$$

Gözlem dizisinin üretilme olasılığı

$$P(O|q,B) = b_{q1}(O_1)b_{q2}(O_2)...b_{qT}(O_T) \tag{4.31}$$

şeklinde olacaktır. Gözlem dizisi O ve durum dizisi q'nun sistem tarafından üretilme olasılığı

$$P(O,q|\pi,A,B) = \pi_{q0}\prod_{t=1}^{T} a_{qt-1qt}b_{qt}(O_t) \tag{4.32}$$

olacaktır. O gözlem dizisinin verilen model tarafından üretilme olasılığı ise

$$P(O|\lambda) = \sum_{q} P(O,q|\lambda) \tag{4.33}$$

olur. HMM'de gözlem sembolleri sürekli veya ayrık olabilir. Sürekli olması durumunda gözlem sembol olasılıkları sürekli bir olasılık dağılım işlevi ile ifade edilir. Bu durumda model Sürekli Saklı Markov Model adını alır.

4.3.2.1 Saklı Markov Modelde Üç Temel Problem ve Çözümü

HMM'nin geliştirilmesinde çözülmesi gereken üç temel problem bulunmaktadır (Rabiner, 1989).

Bunlar :

- Verilen bir O gözlem dizisi ve λ modeli için, bu gözlem dizisinin model tarafından üretilme olasılığı $P(O|\lambda)$ etkin bir şekilde nasıl hesaplanır?
- Verilen bir O gözlem dizisi ve λ modeli için en yüksek olasılığı verecek q durum dizisi nasıl bulunur?
- $P(O|\lambda)$ olasılığını maksimum yapacak model parametreleri nasıl ayarlanır?

Bu üç problem sırasıyla değerlendirme (evaluation), model yapısını öğrenme (learn structure) problemi ve tahmin (estimation) problemi olarak adlandırılmaktadır.

4.3.2.2 Değerlendirme Problemi

Olasılık hesabının nasıl yapılacağı ile ilgilenmektedir. Denklem kullanılarak $P(O|\lambda)$'yı hesaplamak mümkündür ancak bunun için gereken işlem sayısı fazladır. Bu durum sayısı ve

gözlem sembol sayısının yüksek olduğu sistemler için oldukça karmaşık bir hal almaktadır. Bu nedenle olasılık hesabındaki işlem sayısının azaltmak için ileri-geri işlemi kullanılmaktadır (Rabiner ve Juang, 1993). İleri-geri işlemi sayesinde olasılık hesabı etkin bir şekilde yapılabilmektedir.

4.3.2.3 Tahmin Problemi

Verilen bir gözlem dizisi için tahmin probleminin çözümü ile bu diziyi en yüksek olasılıkla üretecek model parametreleri bulunmuş olur. Müzik tanıma sisteminde bu işlem eğitim aşamasına karşılık gelmektedir. Model parametrelerinin elde edilmesinde kullanılan gözlem dizisi eğitim dizisi olarak adlandırılır ve müzik tanımada bu dizi özellik vektörüdür.

HMM 'de tahmin probleminin çözümü için Baum-Welch yöntemi kullanılmaktadır (Rabiner, 1989). Problemin çözümünde en büyük olabilirlik (Maximum Likelihood) yöntemi kullanılır.

4.3.2.4 Model Yapısını Öğrenme Problemi

O gözlem dizisini üreten en yüksek olasılıklı durum dizisinin ne olduğu ile ilgilenir. Bu olasılığını en büyük yapan dizi Viterbi algoritması ile elde edilir.

4.3.3 Saklı Markov Modelin Müzik Tanımada Kullanımı

Bu çalışmanın ana modelleme bölümü HMM tabanlıdır. HMM zaman içerisinde değişen bir süreci istatistiksel olarak modellemek için çok güçlü bir araçtır. Arkasındaki fikir son derece basittir. Sadece çıktısına ulaşılabilen, bilinmeyen bir kaynaktan gelen olasılıklı bir süreç düşündüldüğünde HMM bu süreç için oldukça uygundur. Fakat müziği segmentlere ayırma konusunda HMM eğitim süreçleri yetersiz kalmaktadır. Çünkü segmentlere ayırmak için herhangi bir ön bilgi bulunmamaktadır. HMM eğitimi için etiketlenmiş bir gözlem serisine ihtiyaç vardır. Örneğin, konuşma tanımada eğitim seti fonem şeklinde etiketlenmiş konuşma verilerinden oluşur. Bu fonem örnekleri kullanılarak herbir fonem için HMM'ler oluşturulur ve tanıma aşamasında bu HMM'ler kullanılır. Fakat müzik tanımada karşılan problem daha farklı ve bu nedenle çok daha zordur. Ana amaç müzik parçalarındaki benzer segmentleri herhangi bir ön bilgi olmadan belirleyebilmek ve sonuç olarak her segment için HMM'ler eğitebilmektir. Bu tam olarak eğitmensiz(unsupervised) HMM eğitimine örnektir. Bu tür HMM'ler Rekabete Dayalı Saklı Markov Model(Competitive Hidden Markov Model - CoHMM) olarak adlandırılabilir (Batlle ve Cano, 2000). Normal HMM'den sadece eğitim aşamasında ayrılır.

Çalışmada müzik parçası akustik olayların dizilimi olarak ifade edilmektedir. Böylece benzer bir akustik dizi geldiğinde de müzik parçası tanınacaktır. Biyolojik terminolojide bu akustik olaylara AudioGen adı verilmiştir(Neuschmied, Mayer ve Batlle, 2001). Müzik parçası parmak izi ise AudioGen'lerin dizilimi şeklinde AudioDNA olarak adlandırılmıştır.

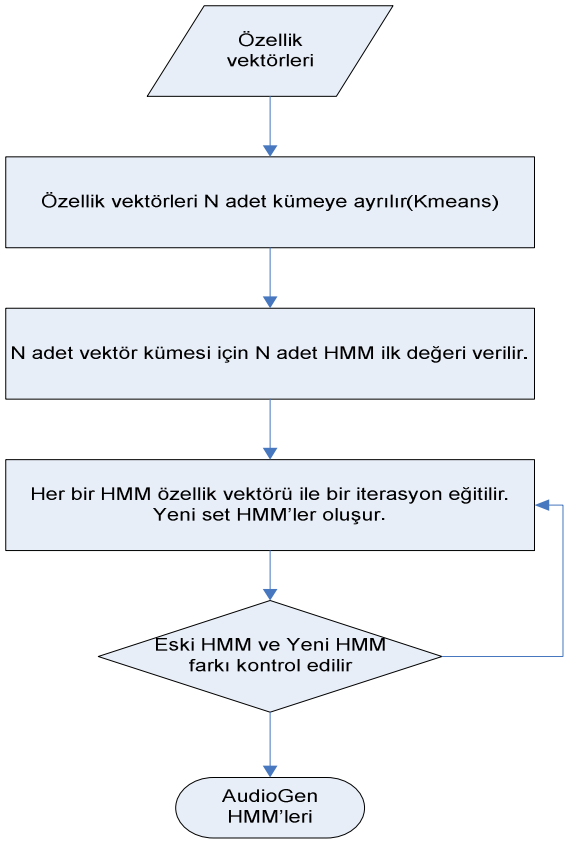
Müzik parçasında konuşmada olduğu gibi fonemler yoktur. Notaların kullanımının ise çeşitli dezavantajları vardır. Çünkü notalar üstüste çalınabilir ve üzerlerinde şarkı sözleri gelmektedir. Bu nedenle AudioGen'ler yeterli miktarda akustik olayları içerdiği beklenen bir müzik parçası topluluğu içerisinden eğitmensiz olarak eğitilen HMM modelleri şeklinde oluşturulur. Her bir HMM'nin belirli bir akustik olayı gösteren özellik vektörlerini ürettiği varsayılır. Aslında AudioGen HMM'lerinin müzik için herhangi bir fiziksel anlamları bulunması beklenmemektedir.

Sistem tasarımı gereği iki çeşit eğitim aşaması içermektedir. Birincisi AudioGen'lerin üretilmesi, ikincisi üretilen AudioGen'lerden herbir müzik parçası için parmak izlerinin üretilmesidir. AudioGen'lerin eğitimi bir kere yapılırken, parmak izi üretimi her şarkı için tekrarlanmaktadır.

4.3.3.1 AudioGen

AudioGen eğitimi için eğitim setinde bulunan müzik parçalarından rastsal olarak alınmış belirli zaman süresindeki bölümlerinden oluşan AudioGen eğitim verisi oluşturulur. Çalışmaya göre oluşturulması istenen AudioGen sayısı belirlenir. Ardından eğitimsiz olarak herbiri bir HMM olan AudiGen'ler oluşurulur.

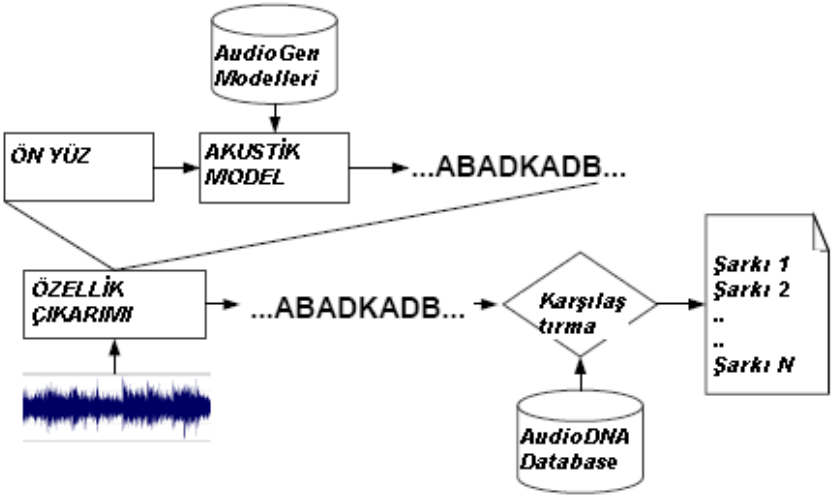
Eğitmensiz HMM eğitimlerinin yapılması için kesinleşmiş bir yöntem bulunmamaktadır. Literatürde bazı çalışmalarda bu tür bir eğitimin CoHMM saklı markov modelleriyle yapılabileceği belirtilmiştir(Batlle ve Cano, 2000). Fakat bu HMM türünün eğitimi için ortaya konmuş açık kesin bir yöntem bulunmamaktadır. Bu nedenle bu çalışmada girişi eğitim seti ve çıkışı istenilen sayıda akusik olayı ifade eden AudioGen HMM'leri olan bir tasarım yapılmıştır. Bu tasarıma göre AudioGen'ler oluşturulmuştur. Özeti Şekil 4.14'dedir.



Şekil 4.14 AudioGen Üretimi

4.3.3.2 AudioDNA

AudioGen'ler müzik parçalarının parmak izlerinin oluşturulması için kullanılır. Müzik parçasının kısa süreli bölümlerinin özellik vektörleri çıkarılır ve bu gözlem verilerini oluşturma olasılığı en yüksek olan AudioGen'lere atanır. Oluşturulan müzik parmak izine AudioGen'ler dizisi şeklinde oluşacağından AudioDNA adı verilir. Üretim süreci Şekil 4.15'te verilmiştir.



Şekil 4.15 AudioDNA Üretim ve Kullanımı

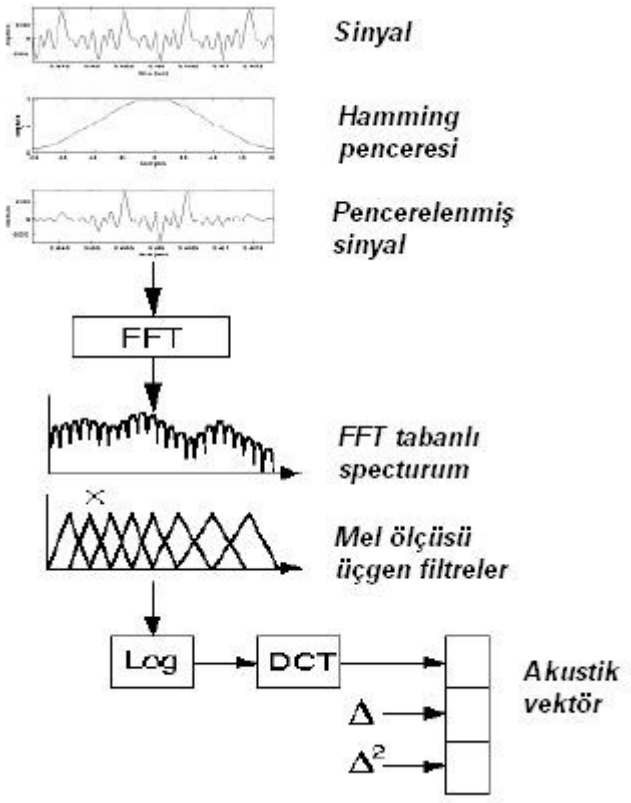
5. UYGULAMA

Uygulama aşamasının, ses sinyalinin sayısallaştırılmasından başlayarak ön işlemler, özellik çıkarma, HMM model oluşturma, HMM eğitim, müzik parçası modeli oluşturma ve HMM çözümleme süreçlerinin gerçekleştirilme adımları ve kullanılan değişkenler bu başlık altında anlatılacaktır. Ayrıca farklı uygulama modelleri üzerinden elde edilen deneysel sonuçlara ve tanıma başarı oranlarına da yer verilecektir.

Eğitim ve test aşaması için kullanılan müzik parçası örnekleri CD kalitesinde kaydedilmiş müzik parçalarıdır. Bu parçalar 22.050 kHz örnekleme frekansı ve 16 bit çözünürlüğünde PCM wave formatına dönüştürülmüştür.

5.1 Ön Yüz

Ön yüz bölümünde müzik parçalarını özellik çıkarma bölümüne hazırlamak için bu çalışmada bahsedilen yöntemler kullanılmıştır. Bu yöntemler ve kullanılan parametreler açıklanmıştır.



Şekil 5.1 Uygulama Ön Yüz (Batlle, Masip ve Guaus, 2004)

5.1.1 Ön İşleme

Ön işleme aşamasında eğitim seti için CD kalitesinde bulunan müzik parçaları, wave formatına çevrilmiştir. Wave formatındaki müzik parçalarının örnekleme frekansı 22,050 Khz, örnek boyutu 16 bit olarak alınmıştır.

5.1.2 Çerçeveleme

Sinyale kısa dönem analizi yapabilmek için çerçeveleme işlemi yapılmıştır. Her bir çerçeve 512 örneklik boyutta alınmıştır ve 256 örneklik kaydırmalar yapılarak çerçeveleme yapılmıştır. 512 örneklik pencere boyutu 22,050 Khz ile örneklenmiş müzik kaydında yaklaşık 23,21 ms'lik zaman dilimine denk gelmektedir. Bu da durağan kabul edilen zaman dilimi için uygundur (Kinnunen, 2003). Ayrıca 256 örneklik kaydırmalar sayesinde örtüşme oranı %50 seviyesindedir.

5.1.3 Ön vurgulama

Sinyale ön vurgulama yapılarak yüksek frekanslı bileşenler baskın hale getirilmiştir.

$$\boldsymbol{H(z) = 1 - 0.95z^{-1}} \qquad (5.1)$$

Ön vurgulama için 0.95 önvurgulama oranı kullanılmıştır.

5.1.4 Pencelereleme

Pencelereleme için hamming window kullanılmıştır.

$$w(n) = 0{,}53836 - 0{,}46164 \cos\left(\frac{2\pi n}{N-1}\right) \qquad (5.2)$$

N frame uzunluğu nedeniyle 512 değerindedir. Pencereleme ardından çerçevelerde süreksizliğin önüne geçilir (Rabiner ve Juang, 1993).

5.2 Özellik Çıkarma

Özellik çıkarma bölümünde üzerinde ön işlemler yapılmış müzik verisinden özellik vektörleri çıkartılmıştır.

5.2.1 MFCC özelliklerinin elde edilmesi

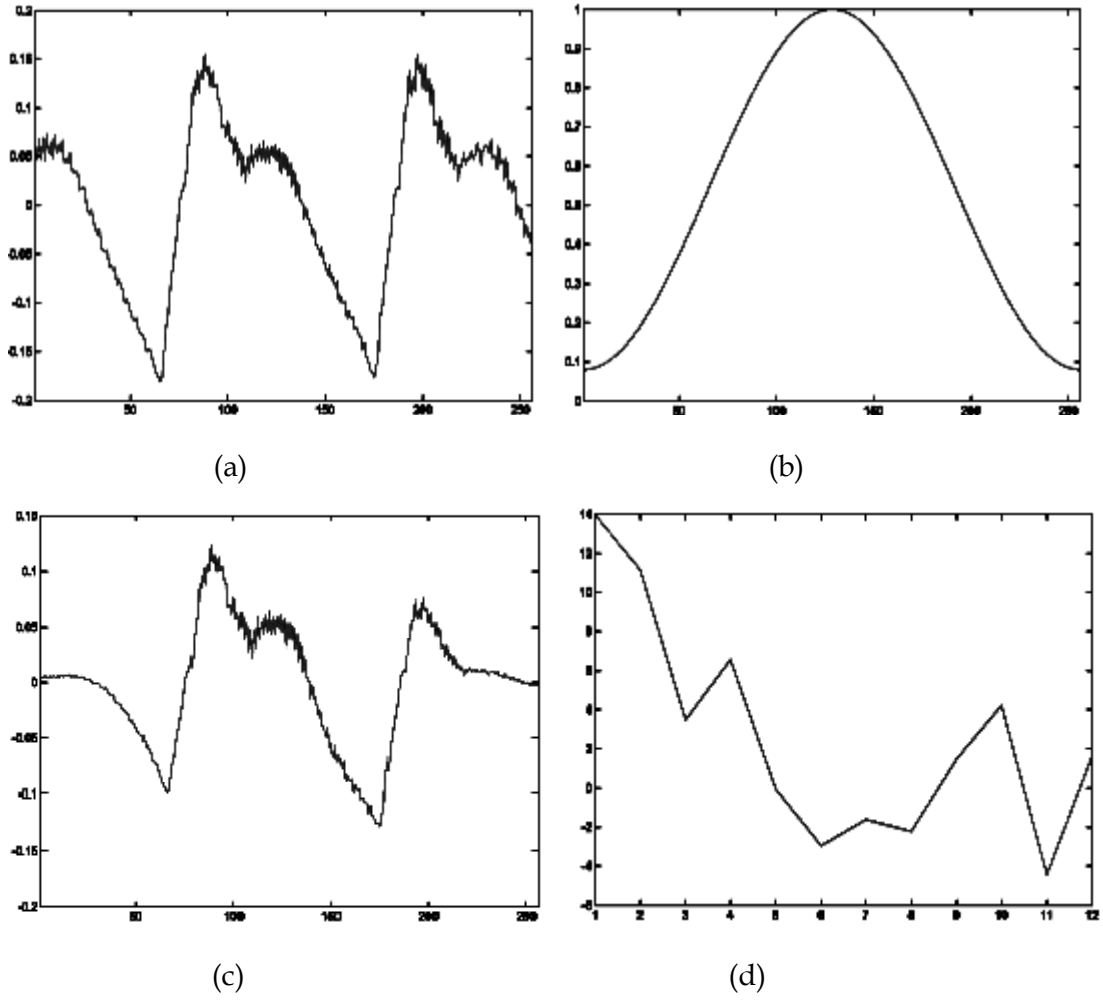
Pencereleme fonksiyonu sonucu elde edilen her bir 512 örneklik ses çerçevesine FFT uygulanarak frekans boyutunda simetrik 512 örneklik özellik vektörü elde edilir. Simetrik

yapı dikkate alınarak özellik vektörünün gereksiz ikinci yarısı özellik vektöründen çıkarılır. Böylelikle frekans boyutunda 256 örneklik özellik vektörü elde edilir. Frekans boyutunda elde edilen özellikler karmaşık değerlikli olmakla birlikte sahip oldukları genlik özellikleri hesaplanarak özellik değerleri olarak kullanılır.

İnsan kulağındaki bazal zara benzetim sağlamak için 23 adet logaritmik eşit aralıklı üçgen pencerelerle filtre bankası oluşturularak 256 örneklik özellik vektörüne uygulanır. Sonuçta boyutu 25 olan bir özellik vektörü elde edilir.

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

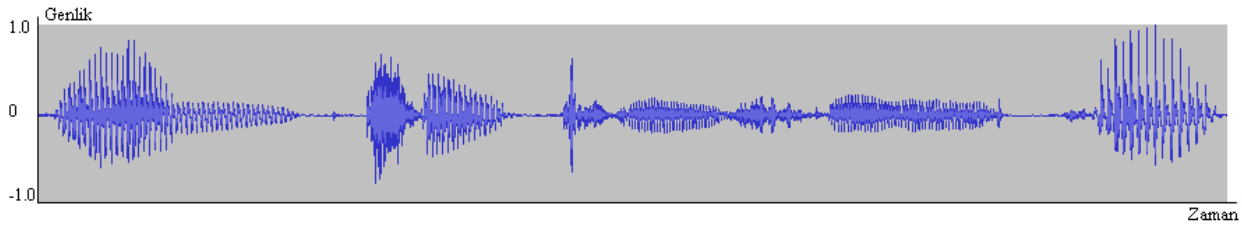
Şekil 5.3 ile ise bir ses verisinden elde edilen MFCC katsayıları, pencerelere, MFCC katsayı sırasına ve genlik değerine göre grafiksel olarak gösterilmiştir.



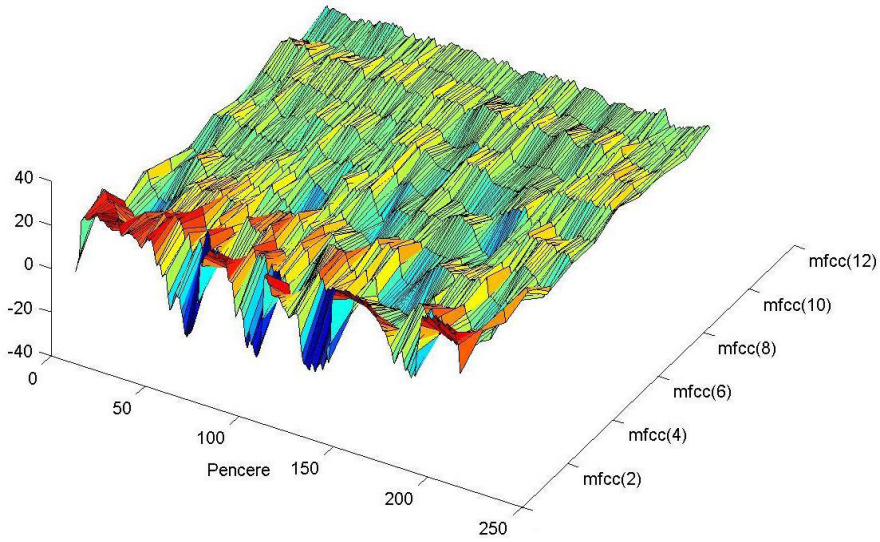
(a) (b)

(c) (d)

Şekil 5.2 (a) Bir pence genişliğinde ses verisi (b) Uygun boyutlu Hamming pencere fonksiyonu (c) Ses verisinin pencere fonksiyonundan geçirilmiş hali (d) İlgili pencere için hesaplanmış MFCC katsayıları



(a)



(b)

Şekil 5.3 (a) Ses verisi (b) Ses verisine ilişkin hesaplanmış MFCC katsayıları

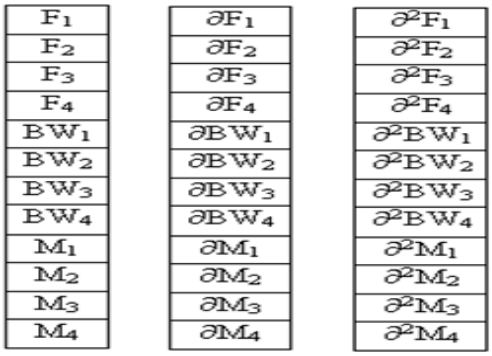
Her bir çerçeveden bir adet 12 boyutlu MFCC vektöre elde edilmiş olur. Müzik parçasının süresine göre elde edilen vektör sayısı belirlenir. Bu vektörler müzik parçasının özellik vektörleri olarak isimlendirilir.

5.2.2 Delta ve Delta-Delta(Acceleration) Özellikleri

Elde edilen 12 boyutlu özellik vektörlerine delta ve delta-delta özellikleri eklenir.

n. vektörün $(n-1)$. vektörden çıkarılmasıyla delta vektörü

n. delta vektöründen $(n-1)$. delta vektörünün çıkarılmasıyla delta-delta vektörleri elde edilir.



Şekil 5.4 Özellik Vektörleri(MFCC, delta MFCC, delta-delta MFCC)

Sonuçta 12 boyutlu vektörümüze delta ve delta-delta özellikleri eklenerek 36 boyutlu özellik vektörleri elde edilir.

5.2.3 Özellik Vektörü Normalizyonu

36 boyutlu özellik vektörleri üzerinde eğitim ve test ortam farklılıkları ve bazı gürültü etkilerini azaltmak amacıyla normalizasyon işlemleri yapılmıştır.

Kullanılan normalizasyon işlemleri MN ve CMS teknikleridir. İki normalizasyon tekniği kullanıldığında önce MN daha sonra CMS kullanılmıştır.

Bu normalizasyon teknikleri genelde konuşma tanıma uygulamalarında kullanılır. Müzik parçası tanıma sırasında eğitim ve test setine uygulanması uygunsuzluk ortaya çıkarmaktadır. Bu nedenle uygulamaya göre özel teknikler geliştirilerek kullanılmıştır. Özellikle eğitim setinde tam bir müzik parçasının çeşitli yerlerinde birbirinden oldukça farklı özellikler görülebilmektedir. Bu nedenle bu müzik parçasının özellik vektörlerinden alınan mean ve maximum değerleri tüm parçada değerlendirmek başarısız sonuçlara neden olmaktadır. Bunun önüne geçebilmek için eğitim aşamasında müzik parçası 5-10 sn'lik parçalara bölünerek her bir parça için normalizasyon işlemleri ayrı ayrı değerlendirilmiştir.

5.3 Model Oluşturma ve Eğitim

Bu bölümde AudioGen ve AudioDNA oluşturma detaylarından bahsedilecektir. HMM eğitiminde ve testlerinde açık kaynaklı java jahmm kütüphanesinden faydalanılmıştır[13].

5.3.1 AudioGen Eğitimi

AudioGen eğitimi için eğitim setinden her bir müzik parçasından 15 saniyelik bölümler

rastgele olarak alınmıştır. Her birinden çıkartılan özellik vektörleri birleştirilerek AudioGen eğitim seti oluşturulmuştur.

Model eğitimi başlamadan önce AudioGen sayısı belirlenmesi gerekmektedir. Bu çalışmada bu sayı 32 olarak belirlenmiştir. Bu sayı müzik parçalarının 32 ayrı akustik olay içerdiği varsayılarak belirlenmiştir. Her akustik olay için bir HMM üretilmiştir. Her HMM ergonomik 3 olaylı HMM olarak belirlenmiştir.

İki HMM arasındaki eşik değeri 0.01 olarak alınmıştır. Uzaklık yöntemi olarak Kullback-Leibler uzaklık yöntemi kullanılmıştır[12].

Oluşturulan AudioGen'lerin her birine bir harf karşılığı verilmiştir. Örneğin 0. HMM "A", 1. HMM "B" ile isimlendirilmiştir.

5.3.2 Müzik parmak izi (AudioDNA) Eğitimi

AudioGen'ler oluşturulduktan sonra sıra müzik parçalarının parmak izlerinin oluşturulmasına gelir. Bu bölümde müzik parçasının uzunluğuna göre dakikada 800 AudioGen içerek şekilde müzik parçası AudioGen'lerle karşılaştırılır. Her 0.075 saniyelik bölüm için özellik vektörleri çıkarılır ve en uygun olan AudioGen ile etiketlenir. Bu bölümde her bir SMM modelin ilgili gözlemi oluşturabilme olasılıkları hesaplanmaktadır. Böylece her bir müzik parçası için yaklaşık 3 KB boyutunda müzik parmak izleri çıkartılmış olur. Örneğin Şebnem Ferah yorumcusunun Okyanuslar adlı müzik parçasından kısa bir bölüm "CCHACJCCBCCCCCCJJJDJDCDDADDDEAA" şeklinde özetlenmiştir.

5.4 Karşılaştırma

Eğitim ardından sistem her bir şarkı için parmak izleri oluşturmuştur. Test için eğitim setinde bulunan şarkılardan oluşan ve farklı ortamlarda kaydedilen 10 sn'lik müzik parçaları kullanılmıştır. Her bir test müzik parçasının 10 sn'lik bölümleri için müzik parmak izleri çıkartılmış ve veritabanında bulunan parmak izleri ile karşılaştırılmıştır.

İki parmak izinin karşılaştırılması temelde string arama işlemidir. Daha detaylı şekilde bakılırsa müzik parçası parmak izlerinin canlı genlerine benzer şekilde sıralandığını görürüz. Bu da karşılaştırma yöntemimizin aynı zamanda bir biyoinformatik problemi olduğunu ortaya koyar.

Biyoinformatikte iki gen sekansının karşılaştırılması ve benzer pasajların bulunması için çeşitli yöntemler bulunmaktadır. Bunlar yerel ve genel dizi hizalama algoritmaları olarak

ikiye ayrılır. Genel hizalamada Needleman-Wunsch, yerel hizalamada ise Smith-Waterman hizalama yöntemleri kullanılır.

Müzik parmak izlerinin karşılaştırılmasına uygun olan local hizalama yöntemi Smith-Waterman yöntemidir. Bu yöntemde tutan ve tutmayan genlere puanlar verilerek test müzik bölümünün her bir şarkıya yakınlığı için bir skor elde edilir. Test verisi, en yüksek skorlu müzik parçasına atanır.

5.5 Test ve Tanıma

Eğitim ve test için öncelikle 6 ayrı yorumcunun 78 müzik parçası kullanılmıştır. İkinci aşamada ise 2 yorumcunun daha müzik parçaları eklenerek, set 103 müzik parçasına çıkartılmış ve eğitim ve test aşamaları tekrarlanmıştır. Müzik parçaları, eğitim için, orjinal CD formatından 22050 Hz örnekleme frekansında 16 bit çözünürlükte wave PCM formatına çevrilerek kaydedilmiştir.

Eğitim için öncelikte müzik parçalarından rastgele 10 – 15 sn'lik bölümler alınmış ve 32 adet AudioGen üretilmiştir. Daha sonra her bir şarkı için dakikada 800 adet AudioGen içeren AudioDNA'lar üretilmiştir.

AMD Opteron 2.59 Ghz işlemci ve 2 GB RAM kullanan PC'de AudioGen üretimi ortalama 6 saat, AudoDNA'ların üretimi ise 1 saat sürmektedir. Uygulama Java programlama dili ile geliştrilmiştir.

Testler müzik parçalarının farklı bölümlerinin dış bir kaynaktan sisteme kaydedilmesi ile yapılmıştır. Bu nedenle testler müzik parçasının farklı bölümleri ile tanıma yapıldığında farklı sonuçlar verecektir. Test için müzik parçasının çaldığı kaynağın formatı önemli değildir. Bir mp3 playerdan, radyodan ya da bilgisayardan çalan müzik parçasının istenilen örnekleme frekansı ve bit çözünürlüğünde kaydedilmesi ya da uygulamaya dinletilmesi yeterlidir.

Test sırasında müzik verisinin dış kaynaklardan alınmasında iki yöntem uygulanmıştır. Bunlar dış kaynaktan çalan müzik parçasının mikrofon portundan ses kablosu aracılığı ile, ikincisi ise mikrofon ile dinleme yapılarak sisteme alınmasındır. Kullanılan kablo ile bir mp3 çaların ses çıkışı bilgisayarın mikrofon portuna bağlanmaktır. Mikrofon ise son derece basit bir kulaklık mikrofon takımının mikrofonudur. Müzik çalarda çalan müzik, mikrofon aracılığı ise sisteme giriş olarak verilmiştir.

Kablo ile yapılan testlerde ortam gürültüsünün minimum düzeye indirilmesi sağlanmıştır.

Böylece sistem modelinin başarısı test edilmiş olur. Mikrofonla yapılan kayıtlarla yapılan testlerde ise ortam gürültüsünün sistem üzerindeki etkisi nedeniyle sistem başarısı düşmüştür.

Testler müzik parçalarının farklı 10 sn'lik bölümleri kullanılarak yapılmıştır. Sistem, yapısı gereği şarkının 10 sn'lik herhangi bir bölümünden müzik parçasını tanımalıdır.

Eğitim ve test aşamasında ilk sette kullanılan müzik parçalarının yorumculara göre sayıları Çizelge 5.1'de gösterilmiştir.

Çizelge 5.1 Test ve Eğitim Müzik Parçalarının Dağılımı

Yorumcu	**Şarkı Sayısı**
Vega	25
Yaşar	10
Şebnem Ferah	10
Kurban	11
Mirkelam	12
Duman	10
Toplam	**78**

Müzik parçalarının yorumculara göre gruplanması sadece mantıksal bir gruplamadır. Sistem için her müzik parçası birbirinden farklıdır. Ayrıca testlerde aynı ve farklı yorumcular arasında testler yapılarak yorumcu değişimlerinin müzik parçası tanımaya etkisi gözlenmiştir. Yapılan testler her bir ayrı müzik parçasının tanınması üzerine yapılmıştır.

5.5.1 Kablo Bağlantısı ile Yapılan Testler

Testler yorumcuların aralarında ve tüm eğitim setinde olmak üzere iki şekilde yapılmıştır. Yorumcuların arasında yapılan testler aynı yorumcunun bir ya da iki albümünde bulunan müzik parçaları arasında yapılmıştır. Tüm eğitim setinde ise yorumcu ya da albüm ayrımı olmadan tüm eğitim setinde tanıma çalışması yapılmıştır. Test sonuçları Çizelge 5.2, Çizelge 5.3 ve Çizelge 5.4 üzerinden görülebilir.

Çizelge 5.2 Yorumcuların kendi şarkıları arasında yapılan testler

	Test Sayısı	**Başarılı Tanıma**	**Başarı Oranı(%)**
Vega	20	17	85
Yaşar	20	18	90
Şebnem Ferah	20	19	95
Kurban	20	16	80
Duman	20	18	90
Mirkelam	20	17	85
Toplam	**120**	**105**	**87,5**

Tek yorumcu ve kablo üzerinden kayıt yapılan testlerde Çizelge 5.2'de görüldüğü gibi yüksek başarı elde edilmektedir. Bu aşamada testler her yorumcunun kendi şarkıları içinde yapıldığı için hatalı tanımalarda aynı yorumcunun farklı müzik parçaları görülmektedir.

Çizelge 5.3 Tüm şarkılar arasında yapılan testler

	Test Sayısı	**Başarılı Tanıma**	**Başarı Oranı(%)**
Vega	20	14	70
Yaşar	20	15	75
Şebnem Ferah	20	16	80
Kurban	20	13	65
Duman	20	15	75
Mirkelam	20	14	70
Toplam	**120**	**87**	**72,5**

Tüm müzik parçaları ve kablo üzerinden kayıt yapılan testlerde Çizelge 5.3 'te görüldüğü

başarı yüksektir fakat bir miktar düşüş göze çarpmaktadır. Bu düşüşün eğitim setinin artması ve farklı yorumcuların müzik parçalarında benzer pasajlar bulunmasının neden olduğu düşünülmektedir.

Çizelge 5.4 Tüm şarkılar arasında yapılan testlerde hataların dağılımı

	Vega	**Yaşar**	**Şebnem Ferah**	**Kurban**	**Duman**	**Mirkelam**
Vega	1	1	2	1	0	1
Yaşar	2	0	0	0	1	2
Şebnem Ferah	2	1	1	0	0	0
Kurban	1	0	2	2	2	0
Duman	0	1	2	2	0	0
Mirkelam	1	3	0	1	0	1

Çizelge 5.4'te yapılan hatalı tanımaların yorumcu bazında dağılımları görülmektedir.

5.5.2 Mikrofon Kaydı ile Yapılan Testler

Testler yorumcuların aralarında ve tüm eğitim setinde olmak üzere iki şekilde yapılmıştır. Yorumcuların arasında yapılan testler aynı yorumcunun bir ya da iki albümünde bulunan müzik parçaları arasında yapılmıştır. Tüm eğitim setinde ise yorumcu ya da albüm ayrımı olmadan tüm eğitim setinde tanıma çalışması yapılmıştır. Test sonuçları Çizelge 5.5, Çizelge 5.6, Çizelge 5.7 üzerinde görülebilir.

Çizelge 5.5 Yorumcuların kendi şarkıları arasında yapılan testler

	Test Sayısı	**Başarılı Tanıma**	**Başarı Oranı(%)**
Vega	20	13	65
Yaşar	20	14	70

Şebnem Ferah	20	16	80
Kurban	20	12	60
Duman	20	14	70
Mirkelam	20	16	80
Toplam	**120**	**85**	**70,8**

Bu aşamada testler her yorumcunun kendi şarkıları içinde yapıldığı için hatalı tanımalar aynı yorumcunun farklı müzik parçaları olarak oluşur.

Çizelge 5.6'da görüldüğü gibi mikrofonla yapılan testlerde başarı bir miktar düşmektedir. Ancak başarı yeterince yüksek seviyededir. Başarının düşmesinin nedeni ortam gürültüsünün test müzik kaydına eklenmesidir. Ön yüzde filtreleme ve özellik vektörü normalizasyonu yapılarak gürültünün etkisi azaltılmaya çalışılmıştır. Çalışmanın asıl amacı tanıma modelinin oluşturulması olduğu için mikrofonla yapılan testlerde başarının düşmesi beklenen bir sonuçtur.

Çizelge 5.6 Tüm şarkılar arasında yapılan testler

	Test Sayısı	**Başarılı Tanıma**	**Başarı Oranı(%)**
Vega	20	12	60
Yaşar	20	13	65
Şebnem Ferah	20	14	70
Kurban	20	11	55
Duman	20	10	50
Mirkelam	20	12	60
Toplam	**120**	**72**	**60**

Çizelge 5.7'de yapılan hatalı tanımaların yorumcu bazında dağılımları görülmektedir.

Çizelge 5.7 Tüm şarkılar arasında yapılan testlerde hataların dağılımı

	Vega	**Yaşar**	**Şebnem Ferah**	**Kurban**	**Duman**	**Mirkelam**
Vega	2	0	3	2	0	1
Yaşar	2	1	0	0	1	3
Şebnem Ferah	2	0	0	2	1	1
Kurban	1	0	1	3	3	1
Duman	0	1	2	4	1	2
Mirkelam	3	3	1	1	0	0

Sistemde elde edilen en düşük başarı oranına mikrofonla tüm eğitim seti içinde yapılan testlerde ulaşılmıştır. Bu sonuçta en büyük etkenin ortam gürültüsünün etkisi olduğu görülmektedir. Ayrıca eğitim setinin artması ve şarkılar arasındaki benzer pasajların ortaya çıkması da bu sonucun nedenlerindedir. Başarının artırılması için daha kaliteli bir mikrofon kullanılabilir veya sistem ön yüzünde gürültü temizleme üzerine daha detaylı bir çalışma yapılabilir.

5.5.3 Farklı Müzik Türleri ile Yapılan Testler

Eğitim ve test aşamasında ikinci sette kullanılan müzik parçalarının yorumculara göre sayıları Çizelge 5.8'de gösterilmiştir. İkinci sette, birinci grup müzik parçalarına yeni yorumcuların müzik parçaları eklenmiştir. Bu testlerde sisteme eklenen farklı türlerdeki müzik parçalarının etkisini görmek amacıyla yapılmıştır. İlk sette bulunan müzik parçaları rock erkek ve bayan yorumcu, pop erkek yorumcu müzik parçalarından oluşmuştur. İkinci sette pop bayan ve halk müziği erkek yorumcuları eklenmiştir. Testler ortam gürültüsünün etkisi azaltılma amacıyla müzik çaların ses çıkışının sistemin mikrofon portuna bağlanması ile yapılmıştır. AudioGen ve AudioDNA eğitimi yeni set ile tekrar yapılmıştır.

Çizelge 5.8 İkinci set test ve eğitim müzik parçalarının dağılımı

Yorumcu	**Şarkı Sayısı**
Vega	25
Yaşar	9
Şebnem Ferah	10
Kurban	9
Mirkelam	12
Duman	16
Sezen Aksu	10
Aşık Veysel	12
Toplam	**103**

Çizelge 5.9 Test sonuçları

	Test Sayısı	**Başarılı Tanıma**	**Başarı Oranı(%)**
Vega	24	18	75
Yaşar	28	24	85,71
Şebnem Ferah	23	18	78,26
Kurban	24	14	58,33
Duman	20	13	65
Mirkelam	30	26	86,66
Sezen Aksu	30	28	93,33
Aşık Veysel	18	17	94,44
Toplam	**197**	**158**	**80,20**

Eklenen yeni türdeki müzik parçalarının genel tanıma başarısı oranında önemli bir değişiklik oluşturmadığı görülmüştür. Yorumcu bazında bakıldığında ise yeni eklenen yorumcularda yüksek başarı görülürken önceki testlerde bulunan bazı yorumcularda tanıma başarısının düştüğü görülmüştür. Çizelge 5.9 ve Çizelge 5.10'da görüldüğü gibi "Kurban" yorumcusunun başarısı düşmüştür. Bunun sebebinin diğer müzik parçalarına bakıldığında daha farklı bir müzik türü icra etmeleri olduğu düşünülmektedir. Yeni müzik parçalarının eklenmesi ve AudioGen'lerin bu şekilde üretilmesiyle bu fark daha çok ortaya çıkmış daha az AudioGen ile tanımlanır olmuş ve ayırt edicilikleri düşmüştür.

Çizelge 5.10 Hatalı tanımaların yorumcu bazında dağılımı

	Vega	**Yaşar**	**Şebnem Ferah**	**Kurban**	**Duman**	**Mirkelam**	**Sezen Aksu**	**Aşık Veysel**
Vega	2	0	2	0	0	1	1	0
Yaşar	1	0	1	2	0	0	0	0
Şebnem Ferah	2	0	1	1	0	0	0	0
Kurban	7	0	0	3	0	0	0	0
Duman	2	0	1	3	1	0	0	0
Mirkelam	3	0	0	0	0	1	0	0
Sezen Aksu	0	0	0	0	1	0	1	0
Aşık Veysel	0	0	0	0	1	0	0	0

6. SONUÇLAR ve ÖNERİLER

Tez çalışmasında çok boyutlu bir tanım uzayına sahip müzik parçası tanıma probleminin çözümüne yönelik bir uygulama geliştirilmiştir. MFCC özellik çıkarımı ve HMM ile özellik sınıflandırılması bu çalışmanın temel adımlarını oluşturmaktadır.

MFCC özellikleri format bağımsız olarak ses verisine ilişkin özelliklerin ortaya konmasında etkili bir yöntemdir. HMM de ardışıl özelliklerin geliş sırası dikkate alınarak sınıflandırılmasını sağlayan bir yöntemdir.

MFCC özellik çıkarım adımlarının ses verisine uygulanması ile reel değerlikli özellik vektörleri elde edilir. MFCC katsayıları üzerinde model eğitimi uygulanarak HMM modeline geçilir.

Eğitim aşamasında hesaplanan HMM parametreleri kullanılarak uygun şekilde oluşturulmuş AudioDNA müzik parmak izleri üzerinden test müzik parçalarının tanınması yapılmaktadır. Deneysel kurguların değişimine göre sistem başarısı %60 -%87,5 düzeylerinde gerçekleşmiştir.

AudioGEN adedi ve AudioDNA oluşturulurken kullanılan özellik vektörü sayısı değiştirilerek sistemde farklı başarı oranları elde edilebilir. Aynı şekilde oluşturulan AudioGen sayısı da değiştirilebilir. Uzun zaman alacak deneysel kurgularla bu özelliklerin optimum noktaları bulunabilir.

Sistemin başarılı şekilde çalışabilmesi için farklı müzik türlerinden dengeli sayıda müzik parçaları ile AudioGen üretimi yapılmalıdır. Böylece farklı türlerdeki ses olaylarını ifade edebilecek yeterli sayıda AudioGen'ler üretilebilecektir. Yapılan testlerde, genel setten farklı müzik türünde az sayıda müzik parçasının bulunduğu eğitim setlerinde bu müzik türündeki müzik parçalarında başarının düştüğü görülmüştür. Bunun nedeni müzik türünün ses olayları için az sayıda AudioGen üretilmesi ve bu yüzden ayırt edilebilirliklerini kaybetmeleridir.

Yapılan testlerde ortam gürültüsünün etkisinin ortadan kaldırılmasının tanıma başarısını oldukça arttırdığı gözlenmiştir. Mikrofon ile yapılan testlerde mikrofonun kalitesini arttırmak başarıyı yükseltecektir.

Ayrıca yapılan deneysel testler sonucunda sistemin daha uzun sureli dinlemelerde daha başarılı olduğu görülmüştür. Bu da gözlem sırası sayısının artmasının HMM model ile tanıma başarısını arttırdığını göstermektedir.

MFCC özellik çıkarımı duymaya ilişkin biyolojik yapıyı modellemektedir. LPC özellikleri ile de konuşma oluşumuna ilişkin biyolojik yapı modellenmektedir. İleriki çalışmalar için MFCC ile LPC katsayıları üzerinden hesaplanan kepstral katsayılarının birlikte, özellik değerleri üzerinde

karşılıklı bir çeşit doğrulama gerçekleştirilecek şekilde kullanılmasıyla, özelliklerin daha doğru olarak (sınıflandırmaya olumlu katkı sağlayacak şekilde) elde edilebileceği düşünülmektedir.

Sistem genelinde göze çarpan bir nokta yapılan hatalı tanımalarda bayan yorumcuların genelde bayan yorumculara, erkek yorumcuların ise erkek yorumculara benzetilmesidir. Bunun nedeninin kullanılan özellik vektörleri olduğu düşünülmektedir. MFCC özellikleri konuşma tanıma çalışmalarında öncelikli kullanılan bir yöntemdir. Bu özelliğinden dolayı benzer insan seslerine yakınsama oluşturduğu düşünülmektedir. Ayrıca benzer şekilde sert sesli parçaların sert parçalara ve daha yavaş parçaların yavaş parçalara benzetildiği görülmüştür. Bu gözlem de benzer enstruman seslerinin birbirlerine benzerliklerinden dolayı benzer özellik vektörlerini oluşturduğunu düşündürmektedir.

Müzik tanıma farklı şartlarda uygulanabilmelidir. Müzik parçalarına ait ses örnekleri, farklı mikrofon telefon veya cep telefonu ahizesi, farklı ses iletim ortamları ve farklı akustik çevrelerden toplanmalı ve sistemin bu örneklerde başarılı tanıma yapması sağlanmalıdır. Gerçek dünyada oluşabilecek çesitli gürültülere karşı sistemin dayanıklı hale getirilmesi gerekir. Bu şartlarda yeni veya var olan gürültü azaltma, normalizasyon, filtreleme gibi teknikler kullanılarak gerçek dünyada müzik tanıma başarımı arttırılabilir.

Testlerde bazı müzik parçalarının tanıma oranı yükselirken iken bazı müzik parçalarının tanıma oranının düştüğü görülmüştür. Bu tanıma oranını örnek bazında değiştiren özelliklerin bulunup değerlendirmesi sistem başarısını yükseltecektir.

Sistem kapalı sistem olarak geliştirildi. Sistemde bulunmayan bir müzik parçası ile sistem test edilmeye çalışıldığında en yüksek skoru veren benzer müzik parçasına benzetilecektir. Karşılaştırma skorları üzerinde çalışma yapılarak, eşik değerleri elde edilebilir. Böyle bir çalışma ile sistem açık sisteme dönüştürülebilir.

AudioDNA parmak izi müziğin süresine göre üretildiği için içerisinden müzik parçasının küçük parçalar halinde parmak izlerine ulaşılabilir. Bu sayede yapılan 10 sn'lik tanımalarda müzik parçasının hangi 10 sn'lik bölümünün tanındığı bilgisine de ulaşılabilir.

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

KAYNAKLAR

Allamanche E., Herre J., Hellmuth O., Bernhard Fröbach B. ve Cremer M. (2001), "AudioID: Towards Content-Based Identification of Audio Material", 100th AES Convention, Amsterdam, The Netherlands, May 2001.

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

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

Batlle, E., Cano, P. (2000). "Automatic Segmentation for Music Classification using Competitive Hidden Markov Models".

Batlle, E., Masip, J. Ve Guaus, E. (2004). "Amadeus A Scalable HMM-based Audio Information Retrieval System".

Batlle E., Masip J. ve Guaus E. (2001). "Automatic Song Identificaiton in Noisy Broadcast Audio"

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

Burges J., Platt J., Jana S. (2003), "Distortion Discriminant Analysis for Audio Fingerprinting".

Burges J., Platt J., Jana S. (2002), "Extracting Noise Robust Features From Audio Data".

Cano P., Battle E., Haitsma J. ve Kalter T. (2002), "A review of Algorithms for Audio Fingeprints", Universiat Pompeu Fabra, Philips Research.

Cano, P., Batlle, E., Mayer, H., Neuschmied, H. (2002). "Robust Sound Modeling for Song Detection in Broadcast Audio".

Cheng Y. (2001), "Music Database Retrieval Based on Spectral Similarity", International Symposium on Music Information Retrieval (ISMIR) 2001, Bloomington, USA, October 2001.

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

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

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

Fragoulis D., Rousopoulos G., Panagopoulos T., Alexiou C. ve Papaodysseus C. (2001), "On the Automated Recognition of Seriously Distorted Musical Recordings", IEEE Transactions on Signal Processing, vol.49, no.4, p.898-908, April 2001.

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

Neuschmied H., Mayer H., and Batlle E., (2001). "Identification of Audio Titles on the Internet," in International Conference on Web Delivering of Music.

Haitsma J. ve Kalter T. (2002), "A Highly Robust Audio Fingerprinting System", Philips Research

Haitsma J., Kalter T. ve Oostveen J. (2001). "Robust AudioHashing for Content Identification", in Proc. of the Content-Based Multimedia Indexing, Firenze, Italy, Sept. 2001.

Hanilçi, C (2007)," Konuşmacı Tanıma Yöntemlerinin Karşılaştırmalı Analizi".

Juang, B. H. and Rabiner L. R. (1990). "The Segmental K-Means Algorithm for Estimating Parameters of Hidden Markov Models". IEEE Transactions on Acoustic, Speech and Signal Processing 38 (9), p. 1639-164

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

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

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

Kinnunen T. (2003). "Spectral Features for Automatic Speaker Recognition." PhD Thesis, University of Joensuu, Finland.

Kinnunen T., Karpov E. and Franti, P. (2006). "Real-Time Speaker Identification and Verification". IEEE Transactions on Audio, Speech and Language Processing 14 (1), p. 77-288

Linde, Y., Buzo, A. and Gray, R. M. 1980. "An Algorithm for Vector Quantization". IEEE Transactions on Communications 28 (1), p. 84-95

Logan B. (2002), MFCC for Music Modeling

Mahedero J., Tarasov V., Batlle E. (2004), "Industrial Audio Fingerprinting Distributed System with Corba and Web Services".

Mihak M. ve Venkatesan R. (2001), "A perceptual audio hashing algorithm: a tool for robust audio identification and information hiding," in 4th Workshop on Information Hiding, 2001.

Neuschmied H., Mayer H. ve Battle E. (2001), "Identification of Audio Titles on the Internet", Proceedings of International Conference on Web Delivering of Music 2001, Florence, Italy, November 2001.

Papaodysseus C., Roussopoulos G., Fragoulis D., Panagopoulos T. ve Alexiou C. (2001), "A new approach to the automatic recognition of musical recordings," J. Audio Eng. Soc., vol. 49, no. 1/2, pp. 23–35, 2001.

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

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

Rabiner, L. R. and Juang, B. H., 1993. "Fundamentals of Speech Recognition". Prentice Hall, New Jersey

Sukittanon S. ve Atlas L. (2002), "Modulation Frequency Features For Audio Fingerprinting".

İNTERNET KAYNAKLARI

[1] Auditude website <http://www.auditude.com>

[2] Relatable website <http://www.relatable.com>

[3] Audible Magic website <http://www.audiblemagic.com>

[4] Shazam website <http://www.shazamentertainment.com>

[5] Moodlogic website <http://www.moodlogic.com>

[6] Yacast website <http://www.yacast.com>

[7] Napster website <http://www.napster.com>

[8] ID3Man website <http://www.id3man.com>

[9] Winamp website <http://www.winamp.com>

[10] Winamp website <http://www.winamp.com>

[11]Musicbrainz website <http://www.musicbrainz.org>

[12] Kullback-Leibler distance <http://www.csse.monash.edu.au/~lloyd/tildeMML/KL/>

[13] A HMM implementation in Java
<http://www.run.montefiore.ulg.ac.be/~francois/software/jahmm/>

ÖZGEÇMİŞ

Doğum tarihi	02.04.1982	
Doğum yeri	İstanbul	
Lise	1996-2000	YDA Otakçılar Lisesi
Lisans	2000-2004	Yıldız Üniversitesi Mühendislik Fak. Bilgisayar Mühendisliği Bölümü
Yüksek Lisans	2005-2008	Yıldız Teknik Üniversitesi Fen Bilimleri Enstitüsü Bilgisayara Mühendisliği Anabilim Dalı
Çalıştığı kurum(lar)		
	2006-....	Saftaş Bilgi ve Bilgisayar Teknolojileri A.Ş.