

172181

İSTANBUL TEKNİK ÜNİVERSİTESİ ★ FEN BİLİMLERİ ENSTİTÜSÜ

**DAYANIKLI SES HASH'LEME İLE
İÇERİK TANILAMA**

YÜKSEK LİSAN TEZİ
Müh. Ozan GÜRSOY
(504021504)

Tezin Enstitüye Verildiği Tarih : 19 Aralık 2005

Tezin Savunulduğu Tarih : 2 Şubat 2006

Tez Danışmanı : Prof.Dr. Bilge GÜNSEL
Diğer Jüri Üyeleri Yrd.Doç.Dr. Şule Gündüz ÖĞÜDÜCÜ
Doç.Dr. M. Ertuğrul ÇELEBİ

B. Günel
Ş. Öğüdücü
M. E. Çelebi

ŞUBAT 2006

ÖNSÖZ

Lisans ve yüksek lisans eğitimlerimi tamamladığım İTÜ Bilgisayar Mühendisliği Bölümünde bana emeği geçen tüm hocalarıma teşekkür ederim.

Bana olan güven ve desteklerini, bana rağmen, hiçbir zaman esirgemeyen arkadaşlarıma teşekkür ederim.

Yükseklisans eğitimim sırasında bir süre Kanada'daki laboratuvarında çalışma olanağı bulduğum hocam Prof. Dr. Yusuf ALTINTAŞ'a, bana okul ve hayat yolunda yaptığı katkılardan dolayı çok teşekkür ederim.

Bu tez, Tübitak'ın 104E151 numaralı Sayısal Ses Damgalama ve Ses Parmakizi Kullanarak Güvenli Ses Tanılama projesi kapsamında hazırlanmıştır.

Bu çalışmanın başından sonuna beni destekleyen, bilgi ve tecrübeleriyle bana yol gösteren, değerli hocam Prof. Dr. Bilge GÜNSEL'e çok teşekkür ederim.

Tüm zorluklarda yanımda olan aileme minnettarım.

Şubat, 2006

Ozan GÜRSOY

İÇİNDEKİLER

KISALTMALAR	v
TABLO LİSTESİ	vi
ŞEKİL LİSTESİ	vii
SEMBOL LİSTESİ	viii
ÖZET	ix
SUMMARY	xi
1. GİRİŞ	1
1.1. Ses Parmakizleri	2
1.1.1. Ses parmakizi tanımı	2
1.1.2. Ses parmakizi özellikleri	3
1.2. Ses Parmakizi Kullanım Alanları	4
1.2.1. Yayın takibi	4
1.2.2. Bağlantılı ses sistemleri	4
1.2.3. P2P ağlarda filtreleme	4
1.2.4. Müzik arşivleri	5
1.3. Literatürdeki Ses Hash'leme Algoritmaları	6
1.3.1. Ritm analizine dayalı parmakizi çıkarma (Beat-ID)	6
1.3.2. Bozulma ayırtetme analizine dayalı parmakizi çıkarma (DDA)	6
2. SES HASHLEME İLE PARMAKİZİ ÇIKARIMI VE İÇERİK TANILAMA	8
2.1. Ses Parmakizlerinin Oluşturulması	11
2.2. Veri Tabanı Yapısı Ve Arama Algoritması	15
2.3. Hata Analizi	18
2.3.1. Uygulama karmaşıklığı	19
3. UYGULAMA ARAYÜZÜ VE DENEYSEL SONUÇLAR	20
3.1. Uygulama Arayüzü	20
3.2. Deneysel Sonuçlar	24
3.2.1. FM radyo yayını bozucu etkisine dayanıklılık	25
3.2.2. Mp3 sıkıştırmaya dayanıklılık	25
3.2.3. Ses örnekleme hızı dönüşümlerine dayanıklılık	28
3.2.4. Hızlandırma saldırısına dayanıklılık	29
3.2.5. Ses parmakizi uzunluğu (granularity)	31
3.2.6. Gürültü ve diğer bozulmalar	32
4. SONUÇLAR VE PLANLANAN ÇALIŞMALAR	33

KAYNAKLAR	34
ÖZGEÇMİŞ	35



KISALTMALAR

DDA	: Distortion Discriminant Analysis
M2S	: Music2Share [5]
MFCC	: Mel Frequency Cepstral Coefficients [12]
BHO	: Bit Hata Oranı



TABLO LİSTESİ

	<u>Sayfa No</u>
Tablo 3.1: 44100Hz’lik radyo kaydının veri tabanıyla karşılaştırılması	26
Tablo 3.2: 128Kbit Mp3 sıkıştırmanın veri tabanıyla karşılaştırılması	26
Tablo 3.3: 96Kbit Mp3 sıkıştırmanın veri tabanıyla karşılaştırılması	27
Tablo 3.4: 32Kbit Mp3 sıkıştırmanın veri tabanıyla karşılaştırılması	27
Tablo 3.5: Örnekleme miktarı 22050Hz’e düşürülen ses dosyası	28
Tablo 3.6: Örnekleme miktarı 11025Hz’e düşürülen ses dosyası	29
Tablo 3.7: Örnekleme miktarı 8000Hz’e düşürülen ses dosyası	29
Tablo 3.8: %1 Hızlandırma etkisi	30
Tablo 3.9: %2 Hızlandırma etkisi	30
Tablo 3.10: %3 Hızlandırma etkisi	31
Tablo 3.11: Ses parmakizi uzunlukları	31
Tablo 3.12: Tüm Bozulmalar	32

ŞEKİL LİSTESİ

	<u>Sayfa No</u>
Şekil 1.1: Music2Share Genel Mimarisi [5]	5
Şekil 2.1: Ses HASH'leme ve Veri Tabanı ile Radyo Yayını Takibi	9
Şekil 2.2: Ses Parmakizi çıkarım blok çizelgesi	12
Şekil 2.3: Hanning Pencereleme Fonksiyonu	12
Şekil 2.4: Bant Enerjileri	14
Şekil 2.5: Ses Parmakizi	15
Şekil 2.6: Veritabanında Arama	16
Şekil 2.7: 128Kbps Mp3 Parmakizi Hash Değerlerindeki Hatalı Bit Sayısı	17
Şekil 3.1: Ses Parmakizi Oluşturma	20
Şekil 3.2: Ses Parmakizinin Veritabanında Aranması	21
Şekil 3.3: Veritabanına Yeni Kayıt Girişi	22
Şekil 3.4: Ses Parmakizlerinin Karşılaştırılması	23
Şekil 3.5: Bit Hata Oranlarının Bulunması	24
Şekil 3.6: Karşılaştırma Yöntemi	25

SEMBOL LİSTESİ

P_f	: Yanlış Tanıma Oranı
P_n	: Tanıyamama Oranı
T	: Eşik Değeri
$F(n,m)$	: n. Çerçeveye Ait Ses Alt-parmakızinin m. Biti
$E(n,m)$	: Ses Verisinin n. Çerçevesindeki m. Enerji Bant Değeri



DAYANIKLI SES HASH'LEME İLE İÇERİK TANILAMA

ÖZET

Müzik, teknolojik gelişmelerin de etkisiyle, günlük yaşantımızda daha fazla yer almaya başlamıştır. Bu sektörün gelişmesi, gün geçtikçe daha fazla ve çeşitte müzik üretilmesini ve daha çok insanın üretilen müziğe ilgi duymasını sağlamıştır. P2P ağlar, ev bilgisayarı kullanıcılarının amatör, fakat geniş içerikli dijital müzik arşivleri oluşturmalarını sağlamıştır. Televizyonlarda yayınlanan müzik videoları, gördükleri talep sonucunda, müzik yayını yapan kanalların artmasını sağlamıştır. Televizyonun etkisine rağmen, özellikle metropol insanının yollarda harcadığı zaman düşünülürse, radyolar etkilerini yitirmemiş, tersine yayın yapan kanal sayısı artmıştır. Müziğin günümüz insanının yaşamında aldığı rolü farkedenden mobil sektör, bu alanda kullanıcılarının ilgisini çekecek yeni servislerle kullanımlarını arttırmayı hedeflemişlerdir.

Bu yaygınlaşma, müzik içeriklerinin etiketlenmesi, depolanması ve tanınması konularında bilimsel ve ticari ilginin artmasıyla sonuçlanmıştır.

Bu tez çalışmasında, Phillips Research tarafından tanımlanan “Robust Audio Fingerprinting” yöntemine [1] benzer bir yöntem ile, bilinmeyen bir ses verisini tanılama sistemi gerçekleştirilmiştir. Ses tanılama sistemi, bilinmeyen bir ses verisinin ayırt edici algısal özelliklerinin bulunup, daha önceden oluşturulmuş bir veri tabanı içinde, model eşleme yöntemine dayalı karşılaştırılması sonucunda, ses verisini tanımlamaya yarayan bilgilerin elde edilmesidir.

Günümüzde ses tanılama sistemleri, radyo yayını izleme, mobil uygulamalarla şarkı tanıma, P2P ağlarda yasal olmayan şarkıların filtrelenmesi, telif hakları ve müzik arşivlerinin oluşturulup düzenlenmesi gibi değişik alanlarda kullanılmaktadır.

Ses hash'lemesi ile, şarkı ve benzeri ses kayıtlarını benzersiz olarak tanılamaya yarayan ses parmakizlerinin kodlanması sağlanır. Ses tanılama sistemi iki ana bölümden oluşur: ses parmakizleri çıkarma algoritması ve çıkarılan ses parmakizlerinin, çok sayıda şarkıya ait parmakizleri bulunan bir veri tabanında en etkin şekilde aranması.

Ses parmakizleri, ses verisinin her 11.6 milisaniyesi için oluşturulan 32 bitlik ses alt-parmakizlerinden (hash değerlerinden) oluşur. Bu, sesin 11.6 milisaniyelik bir çözünürlükle incelendiği anlamına gelir. Alt-parmakizleri zaman ve frekans uzayı boyunca hesaplanan enerji farklarından elde edilir. 256 hash değerinin oluşturduğu dizi bir ses parmakizine karşı düşer. Bu da ses verisinde yaklaşık 3 saniyelik bir bölümdür ve ses verisinin tanınması için yeterli bilgiyi içerir.

Ses parmakizi arama algoritmasının temel hedefi, bir parmakizi bloğunu mümkün olan en etkin şekilde veri tabanındaki parmakizleriyle karşılaştırmaktır. Bire bir eşleşmeden çok en benzeyen ses parmakizlerinin bulunması amaçlandığı için karşılaştırma basit bir işlem değildir. Aranacak ses parmakizleri genelde işlenmiş ya da bozulmuş ses verilerinden çıkarıldığı için bire bir eşleme kullanılamamaktadır.

Tanımlanan sistem ses verisindeki bozulmalara karşı oldukça dayanıklıdır. Ayrıca ses verisini sadece 3 saniyelik bir bölümle tanımlama yeteneğine sahiptir. Parmakizlerinin veri tabanında aranmasını hızlandırmak için karşılaştırma sayısını azaltan indeks tablosu hazırlanmıştır.

Bundan sonraki araştırmalar, diğer ses parmakizi çıkarma algoritmalarının denenmesine ve bu tezde uygulanan arama algoritmasının daha da iyileştirilmesine odaklanacaktır.



ROBUST AUDIO HASHING FOR CONTENT IDENTIFICATION

SUMMARY

Music started to take part in our daily lives more and more, thanks to technological progress, which provides more and variety music produced as time goes on and more people interest in this music. P2P networks, provide home pc users compose amateur but with large contents digital music archives. The number of music channels increased, because of the attention for the music videos broadcasted on television. Especially if metropol people who spare too much time on traffic is thought, in spite of the bad effects of TV, radio didn't lose its popularity. On the contrary the number of radio channels increased. The mobile sector which has comprehended importance of the music for the people's daily lives, aimed to increase the utilization with new services which can draw attention of the users.

This widespread came to a conclusion with an increasing scientific and commercial demand for labelling, storage, and identification of the contents of songs.

In this thesis studying, by the method which is called "Robust Audio Fingerprinting" by Phillips Research [1], an unknown audio data's identification system is described. Audio identification system uses the distinctive perceptual features of an unknown audio data to retrieve the identifying info of the audio data from an already prepared database of audio fingerprints by comparing based on model pairing.

Today audio identification system is used in radio broadcast monitoring, identification songs with mobile utilization, filtering the illegal songs at P2P networks, at copyrights and usage of music database.

Audio hashing provides coding audio fingerprint, which is used for identification uniquely a segment of music. An audio fingerprint system basically consists of two parts: a fingerprint extraction algorithm and an algorithm to efficiently search such a fingerprint in a fingerprint database containing the fingerprints of many recordings.

Audio fingerprinting extraction is based on extracting bit sequences of 32 bits, so called sub-fingerprints (hash values), for every 11.6 ms of audio. This means audio is examined under a solubility with 11.6 ms. These sub-fingerprints are obtained from energy differences along both the time and the frequency axis. A sequence of 256 sub-fingers equals a fingerprint block, which corresponds to 3 seconds of audio, contains enough information to enable reliable identification.

The basic aim of the audio fingerprint search algorithm is to compare a fingerprint block with the fingerprints in the database in an efficiently way as possible. It is not an insignificant task, because the matter is not to find an exact copy but the most similar one. This is due to the fact that a fingerprint block that has to be identified is generally taken from an audio excerpt that has been processed or degraded.

The described audio fingerprint extraction algorithm is highly robust against audio signal degradation. Besides it needs only 3 seconds of audio to identify a song. The fingerprint search algorithm is highly efficient and fast.

Future research will focus on investigating other fingerprint extraction and optimizing the current fingerprint search algorithm.



1. GİRİŞ

Günümüzde yaygınlaşan müzik tüketimi pek çok değişik yoldan müziğe erişimi kolaylaştırmıştır. Bu sayede oluşan geniş içerikli müzik arşivlerinde, insanların istedikleri müziğe hızlı ve kolayca ulaşması, içeriğin güvenilirliği ve kalitesi, sağlanması gereken önemli özellikler olarak karşımıza çıkmaktadır.

Ses içeriği tanılama, bilinmeyen bir ses işaretinin ayırt edici özelliklerinin bulunup, daha önceden oluşturulmuş bir veri tabanı içinde, model eşleme yöntemine dayalı karşılaştırılması sonucunda, ses işaretini tanımlamaya yarayan bilgilerin elde edilmesidir. Ses verisini özetlemede öznitelik çıkarımı [1] ya da hash fonksiyonları [1] kullanılabilir.

Öznitelik çıkarımı kullanılan ses tanılama yöntemlerinin performansı, kullanılan özniteliklere ve sayısına bağlı olarak değişir. Nitelikler genellikle ondalık değerlerle ifade edilir. Giriş verisindeki bozulmalara karşı esnektir.

Hash fonksiyonları giriş verisini benzersiz bir şekilde tanılayacak ikili sistemde sayılar üretir. Genel kullanım alanı şifrelemedir. Büyük miktardaki veriyi özetleyen değerler üretir. Ayırtıcılığı çok fazladır ancak giriş verisindeki bozulmalara karşı kırılgandır. Bu nedenle de video ve ses gibi değişik kalitelere kaydedilebilen içerikler için uygun değildir, modifiye edilmeleri gerekir.

Ses parmakizi çıkarma, ses tanılama sistemlerinin büyük boyutlu ses verilerini benzersiz bir şekilde ve çok daha küçük boyutlarda tanımlamak için kullandıkları bir yöntemdir. Ses parmakizleri ses verisiyle ilgili tanımlayıcı bilgilere ulaşmamızı sağlayan indekslerdir. Ses parmakizlerinin kullanılması; depolama kaynak ihtiyacının azaltılması, karşılaştırmalarda önemsiz algısal özelliklerin ortadan kalkması, aramaları kolaylaştıran bir veri yapısına geçilmesi gibi avantajlar sağlar.

Sıradan kullanıcılar bile orijinal ses verileri üzerinde değişik oynamalar yapıp bu verileri tekrar kaydedebilmektedirler. Ayrıca ses tanılama sistemleri ses verisini radyo, P2P ağlar gibi içeriğin kalitesinin ve güvenilirliğinin azaldığı kanallardan

alabildikleri için ses parmakizi çıkarımlarının ses verisindeki bozulmalara karşı oldukça dayanıklı olmaları gerekir.

Günümüzde ses tanılama sistemleri, radyo yayını izleme, mobil uygulamalarla şarkı tanıma, P2P ağlarda yasal olmayan şarkıların filtrelenmesi, telif hakları ve müzik veri tabanları uygulamalarında kullanılmaktadır.

Tez kapsamında, 1. Bölümde, ses parmakizi tanımı, özellikleri ve uygulama alanları ayrıntılı olarak incelendikten sonra, literatürdeki ses parmakizi uygulamaları hakkında bilgi verilmekte. Gerçeklenen ses hash'leme sistemi, ses parmakizi çıkarımı ve veri tabanı yapısıyla arama algoritması 2. Bölümde açıklanmakta. 3. bölümde uygulama ve gerçekleştirilen sistemle elde edilen sonuçlar, 4. ve son bölümde çıkarılan sonuçlar ve ileriye yönelik çalışmalar özetlenmekte.

1.1 Ses Parmakizleri

1.1.1 Ses parmakizi tanımı

Ses parmakizleri bir çeşit dayanıklı hash'leme algoritmasıdır. Orijinal verinin algılanabilen ses yapısını benzersiz bir şekilde belirterek ve çok daha az kaynak kullanarak büyük miktardaki veriyi ifade eder [2].

Şifrelemede yaygın olarak kullanılan hash'leme fonksiyonları, genelde oldukça büyük olan X verisini, daha küçük miktardaki bir hash değeriyle eşler. Bu fonksiyonlar X ve Y gibi iki büyük veriyi, $H(X)$ ve $H(Y)$ hash değerlerini kullanarak karşılaştırmaya yarar. Farklı veriler için aynı hash değerinin üretilmesi hash'leme fonksiyonunun hata oranını verir. Eğer n bitlik hash değerleri kullanılırsa, bu hata oranı 2^{-n} ile gösterilir.

Her ne kadar hash'leme fonksiyonları ses parmakizi çıkarımı için uygun gibi görünse de, uygulanan sistemde amaç bire bir matematiksel eşleşme bulmak değil, algısal benzerlikleri ortaya çıkarmaktır. Ses verisi, farklı kalitelerde kaydedildiğinde çok değişik işaretler oluşturduğu halde algısal olarak benzerliğini korur ve insan algı sistemi tarafından aynı veri olarak kabul edilir. Ses parmakizi sistemi de insan algısal sistemini taklit etmeye çalışmaktadır.

Tasarlanan ses hash'leme algoritması, algısal benzerliği bulunan ses verileri için benzer ses parmakizleri üretmelidir. Ses parmakizi fonksiyonu, F , benzer X ve Y verileri için

$$\|F(X) - F(Y)\| \leq T \quad (1.1)$$

birbirine benzemeyen X ve Y için de

$$\|F(X) - F(Y)\| > T \quad (1.2)$$

üretecek şekilde bir eşik değerine, T, sahip olmalıdır.

1.1.2 Ses parmakizi özellikleri

Bir ses parmakizi sisteminin sağlaması gereken özellikler kullanıldığı uygulamaya göre değişiklik gösterse de temelde her uygulama için gerekli nitelikler vardır. Bu özelliklerin öncelik ve etkinlikleri uygulamadan uygulamaya değişebilir. Ses parmakizi sistemlerini değerlendirme ölçütleri aşağıda belirtilmiştir:

- **Dayanıklılık:** Ses işaretindeki bozulmalara rağmen ses verisinin doğru bir şekilde tanılanmasıdır. Hatalı olarak tanıyamama oranı bu özelliği belirtmede kullanılır. Daha dayanıklı ses parmakizi oluşturmak için parmakizleri işaret değişimlerine karşı sabit, algısal özelliklerden çıkarılmalıdır.
- **Güvenilirlik:** Bir giriş verisi için farklı zamanlarda farklı sonuçlar üretmemesidir. Aynı zamanda bulunan sonucun doğruluğu da güvenilirliği tanımlamada kullanılır. Giriş verisinin hatalı tanılanma oranı ile ters orantılıdır.
- **Parmakizi Boyu:** Hızlı arama ve bellek kaynak gereksinimlerini belirler. Kısaldığı zaman ise ayrıştırıcılığın azalmasına ve hatalı tanıma oranının artmasına neden olur.
- **Çözünürlük:** Ses parmakizi çıkarmak için kullanılması gereken ses verisinin boyunu belirler. Literatürde 5 saniyenin altında olması kabul edilmiştir. Kısa olması, ses verisinin tamamına ihtiyaç duymadan tanılamayı sağlar.
- **Çok Yönlülük:** Farklı ses formatlarında verileri tanılayabilme yeteneğidir. Ses veri tipinin kısıtlanamayacağı uygulamalar için gerekli bir özelliktir.
- **Karmaşıklık:** Parmakizi çıkarma algoritmasının hesaplama maliyeti, aramalarda parmakizlerinin karşılaştırılma maliyeti ses parmakizi sisteminin karmaşıklığını belirleyen etkenlerdir.

- **Ölçeklenirlik:** Geniş veri tabanları ile çalışıldığında ses verisinin tanılama süresi ve aynı anda karşılanabilecek tanılama istekleri ses parmakizi sisteminin ölçeklenebilirliğini gösterir.

1.2 Ses Parmakizi Kullanım Alanları

1.2.1 Yayın takibi

Ses parmakizi sistemlerinin en yaygın kullanıldığı alandır. Radyo yayınlarının zamana göre kayıtlarının tutulması, izlenirlik ölçümleri, reklâm ve telif haklarının takibi gibi olanaklar sağlar. Yayın takibi günümüzde birçok yerde hala insanlar tarafından yapılmaktadır. Parmakizi uygulamalarıyla bu işlem otomatize edilebilir.

Yayın takibi, parmakizi çıkarımının yapıldığı bir kaç alt uygulama ile parmakizi veri tabanı sunucusu olarak hizmet veren bir merkezi uygulama ilişkilendirilerek gerçekleştirilebilir. Merkezi uygulama yayın takibi yapan uygulamalardan gelen parmakizlerini veri tabanında karşılaştırarak her kanala ait program listesinin döktümünü çıkarır.

1.2.2 Bağlantılı ses sistemleri

Bağlantılı ses sistemleri; müziğin, destekleyici ek bilgilerle bağlantılı olarak kullanıldığı uygulamalar için kullanılan genel bir terimdir [3]. Örnek olarak, radyoda dinlediğiniz bir parçanın adını ve kime ait olduğunu öğrenmek istediğinizde, özel bir mobil servis numarasını arayıp çalan parçayı telefonunuza dinleterek, parça hakkında bilgilerin bulunduğu bir kısa mesaj alabilirsiniz.

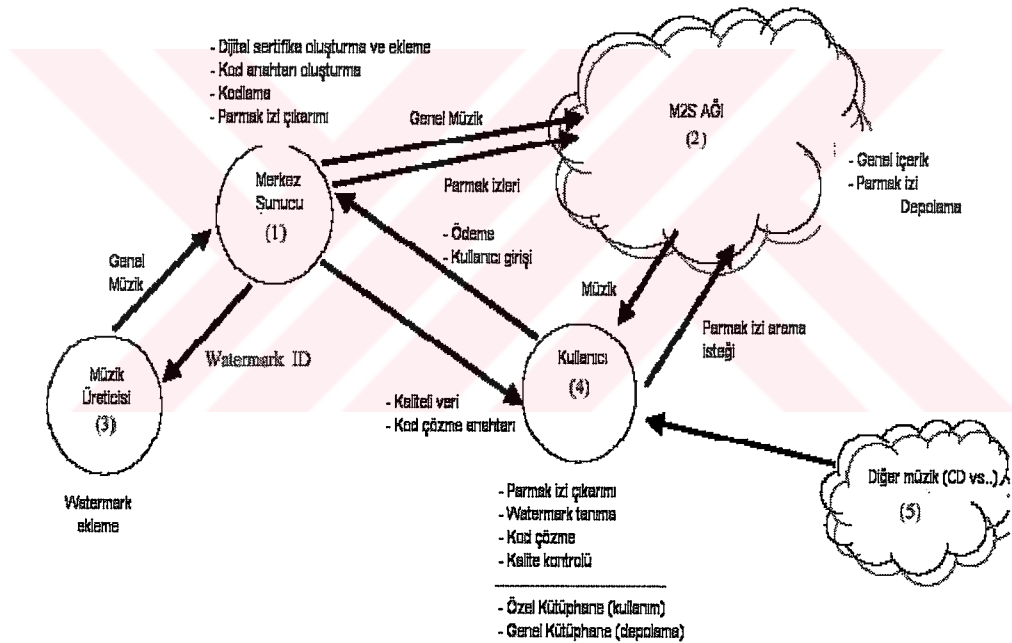
Diğer bir örnek de, bilgisayarınızda dinlediğiniz bir müziğin ses parmakizini internet üzerinden yollayarak sizi parçayla ilgili bilgi alabileceğiniz ya da albümü satın alabileceğiniz internet sitelerine yönlendiren bir uygulama olabilir.

1.2.3 P2p ağlarda filtreleme

Bu tür uygulamalar için en yaygın bilinen örnek Napster'dır [4]. 1999'da, Napster'ın istemci uygulamasını indiren kullanıcıların, geniş müzik arşivlerini ücretsiz olarak paylaştıkları ve indirdikleri bir sistem yaygın bir şekilde kullanılmaya başlandı. Daha sonra telif hakları yasalarıyla korunan içeriklerin paylaşılması engellendi. Başta dosya ismiyle filtreleme yöntemi uygulayan Napster, kullanıcıların şarkı isimlerini

bozarak bu engeli aşmaları sonucu, Relatable [5] tarafından geliştirilen ses parmakizi sistemini faaliyete geçirdi.

Yasal müzik paylaşım sistemlerinde yalnızca filtreleme yerine daha etkin ve yararlı düzenlemeler yapılabilir. Music2Share [6] ile, hem kullanıcı memnuniyetini arttıran hem de müzik dağıtımında ve paylaşımında yasal kontrollerin yapılabildiği bir mimari önerilmektedir. Kullanıcılar satın aldıkları içerikle ilgili hiçbir kısıtlamayla karşılaşmazlar. Ayrıca sahip oldukları içeriklerin düzenlenmesi ve kalitelerinin artırılması sağlanır. M2S mimarisi bir P2P ağ ve bir merkezi kontrol sunucusundan oluşur. İlki müzik dosyalarının saklanması, aktarılması, tanınması ve kalitelerinin kontrolünden sorumluyken; ikinci kısım tüm ağı kontrol edip güvenilirlik denetimi ve ödeme işlemlerini takip eder [6]. Önerilen sistemin genel yapısı Şekil 1.1'de görülebilir.



Şekil 1.1: Music2Share Genel Mimarisi [5]

1.2.4 Müzik arşivleri

Ev bilgisayar kullanıcıları, yaygınlaşan müzik paylaşımı ve teknolojik gelişimler sonucunda binlerce şarkı içeren arşivlere sahip oldular. İnternette ve paylaşım ağlarından indirilerek, CD'lerden bilgisayarlara aktararak oluşturulan bu arşivler çoğunlukla doğru düzenlenmemiş olurlar. Şarkı bilgileri eksik, yanlış ya da

tutarsızdır. Ses parmakizi sistemlerinin bağılı olduğu parmakizi veri tabanlarında şarkı bilgilerinin doğru tutulduğu kabul edilirse, ev kullanıcılarının arşivlerini kolaylıkla düzenleyebilmeleri sağlanır. Bu amaçla ID3Man [7] uygulaması, Auditude [8] parmakizi teknolojisiyle eksik ya da yanlış etiketlenmiş mp3 dosyalarının düzeltilmesini sağlamaktadır.

1.3 Literatürdeki Ses Hash'leme Algoritmaları

1.3.1 Ritm analizine dayalı parmakizi çıkarma (Beat-ID)

Beat-ID [9], sesin ritim bileşenine dayalı bir parmakizi çıkarım algoritmasıdır. İnsanlar tarafından da çok iyi bir şekilde ayırt edilebilen ritim, sesteki oldukça dayanıklı bir özellik olarak görülmektedir. Ritim özelliği ses parmakizi boylarının küçülmesini de sağlamaktadır.

Önerilen algoritmada ses verisinin boyundan bağımsız olarak 32 baytlık parmakizleri kullanılmaktadır. Oluşturulan parmakizleri iki bölümden oluşur: ortalama ritim periyodu ve bu periyot boyunca zamandaki ortalama enerji dağılımı.

Ritim bulma algoritması üç adımda tanımlanmaktadır: ritim periyodunun alınması, ritim periyodunun ortalama ve gerçek ataklarının kestirilmesi. Tüm bu hesaplamalarda “beklenti maksimizasyonu” (Expectation Maximization) kullanılır.

Ses işareti kimi ön işlemlerden geçirildikten sonra, ses parmakizinin ilk iki baytında ortalama ritim periyodu tutulur. Geri kalan 30 bayt, ortalama enerji dağılımının 30 alt-bandının saklanması için kullanılır. Her bir parmakizi 10–20 saniyelik ses verisine karşı gelir. Bu değer, ses parmakizi çözünürlüğü için MPEG21 grubunca belirlenmiş olan 5 saniye sınırının oldukça üstündedir. Her ses verisi için birkaç parmakizi hesaplanır. Genelde ses verisinin başından ve ortasından alınan iki parmakizi yeterli olmaktadır.

1.3.2 Bozulma ayırdetme analizine dayalı ses parmakizi çıkarma (DDA)

DDA (Distortion Discriminant Analysis) [10], ses parmakizi algoritmaları içinde en karmaşıklarıdır. Hangi özelliklerin sesi en dayanıklı olarak tanımlayabileceğini, yapay sinir ağları ve öğrenme algoritmaları kullanarak bulmayı hedefler.

Önerilen sistem, öncelikle ses öznitelik çıkarma algoritmasının sağlaması gereken koşulları belirtir:

- Giriş verisinin çok boyutluluğu azaltılmalıdır.
- Bulunan ses öz nitelikleri giriş verisindeki bozulmalara karşı dayanıklı olmalıdır.
- Ses öz nitelikleri giriş verisi hakkında bilgi verebilmelidir, farklı giriş verileri farklı öz niteliklere karşı düşmelidir.
- Öz nitelik çıkarım işlemi verimli olmalıdır.

DDA öz nitelikleri bir lineer, çok katmanlı yapay sinir ağı ile hesaplanır. Yapay sinir ağında her bir katman, yönlü ana bileşen analizi (oriented principal component analysis) ile boyutsal indirgeme gerçekleştirir.

0.4 saniyelik giriş verileri için öz nitelik vektörleri hesaplanır. Sistem öğrenme girişinde, bozulmamış veriden hesaplanmış vektör kümesi ile aynı verinin N değişik bozulmuş halinden hesaplanan vektör kümelerini alır. Giriş vektörleri arasındaki maksimum sapma hesaplanarak daha az hatalı tanıma ve tanımama oranları elde edilir. Ancak bu hesaplamalar fazladan ön-işlem maliyeti getirir. Öğrenme sonucunda elde edilen ayırt edici öz vektörler diğer ses dosyalarının modellenmesinde kullanılır.

2. SES HASH'LEME İLE PARMAKİZİ ÇIKARIMI VE İÇERİK TANILAMA

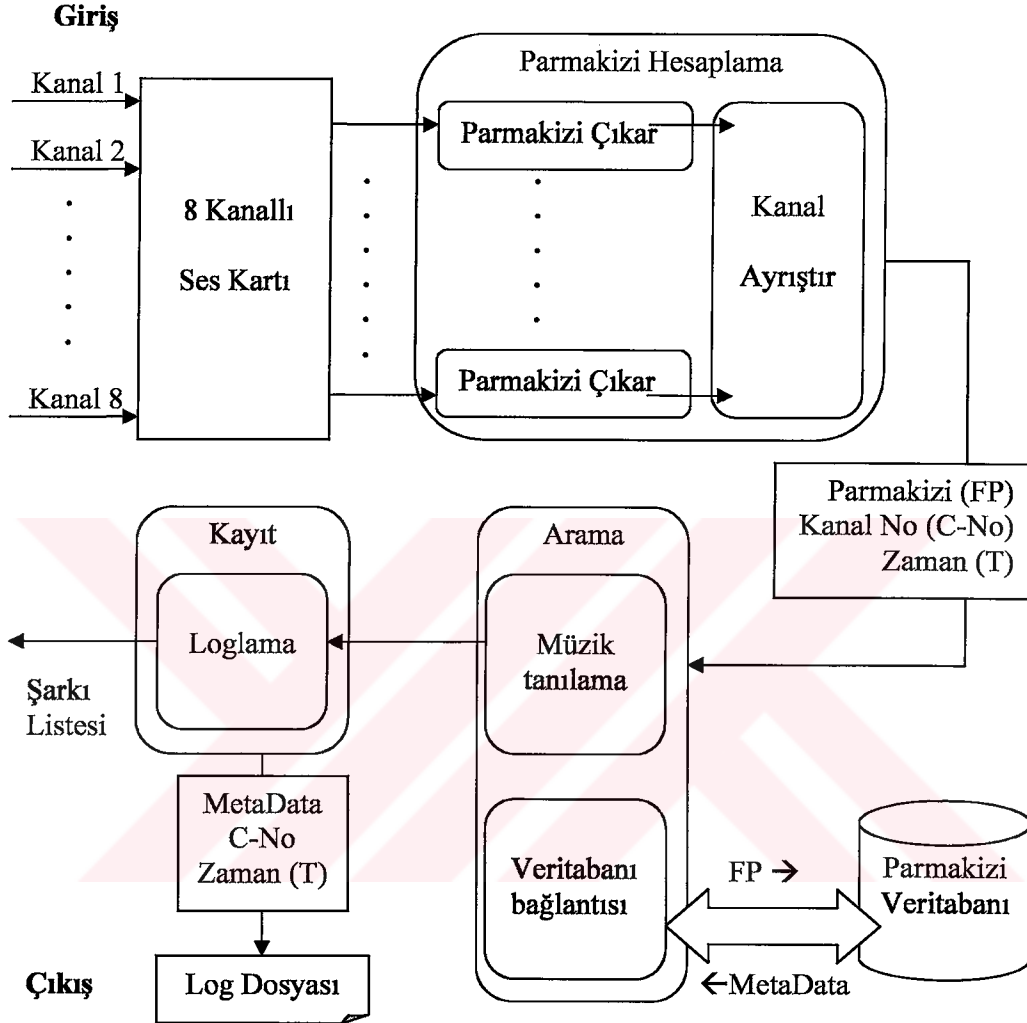
Bu çalışmada, Phillips Research tarafından tanımlanan “Robust Audio Fingerprinting”[1] yönteminin bir benzeri oluşturularak, bilinmeyen bir ses verisini tanılama sistemi gerçekleştirilmiştir. Ses tanılama sistemi, bilinmeyen bir ses verisinin ayırt edici algısal özelliklerinin bulunup, daha önceden oluşturulmuş bir veri tabanı içinde, model eşleme yöntemine dayalı karşılaştırılması sonucunda, ses verisini tanımlamaya yarayan bilgilerin elde edilmesidir.

Uygulama alt yapısı iki ana bölümden oluşur: Ses verisinin okunup ses parmakizlerinin çıkarılması ve ses verilerinden veritabanı oluşturup tanılama için gelen ses parmakizlerinin ait oldukları ses verisi ile eşleştirilmesi.

Bölüm 2.1’de ayrıntılı olarak tanıtılan ses parmakizleri, sesteki algısal öznitelikleri modellemek için sesin anlamsal (semantik) olmayan özelliklerini kullanır, çünkü semantik özellikler hesaplanması daha zor, zamanla ve kişiden kişiye değişkenlik gösteren, her tür müziğe kolaylıkla uygulanamayan özelliklerdir. Ses parmakizleri hash’leme yöntemiyle oluşturulduklarından ses verisini benzersiz olarak tanırlar ve anlamsal bilgi içermezler. Aynı ses verisinin farklı bölümlerinden çıkarılan iki ses parmakizi de birbirine benzemez. Kullanılan hash’leme fonksiyonunun başarımı yanlış tanılama oranını sıfıra indirir.

Uygulamanın etkinliği önemli ölçüde kullanılan veritabanına bağlıdır. Kısıtlı sayıda ses verisi içeren bir veritabanı ile dar kapsamda başarımlar yüksek olsa da kullanıcı isteklerini karşılayamayan yetersiz bir sistem oluşturulmuş olur. Veritabanı, hash’leme sistemi ile baştan sona parmakizi çıkarılmış ses verilerinin, kullanıcılar tarafından kullanılması muhtemel anlamsal bilgilerle birlikte veritabanına girilmesi ile oluşturulur. Gelen ses verisindeki bozulmalara karşı dayanıklı bir sistem oluşturmak için veritabanında yapılacak aramalarda bire bir ses parmakizi eşlemesi yerine, belli bir eşik değerinin altında benzeyen ses parmakizleri bulunmaya çalışılmaktadır. Bu da arama işleminin basit bir parmakizi karşılaştırmasından daha

karmaşık bir hale sokar. Geniş veritabanları arama algoritmalarını hızlandıracak yöntemlere ve yüksek bellek kaynaklarına ihtiyaç duyar.



Şekil 2.1: Ses HASH'leme ve Veri Tabanı ile Radyo Yayını Takibi

Bu tez kapsamında hazırlanan uygulamanın, 104E151 numaralı Tübitak projesi kapsamında, radyo yayınlarını izleyerek gerçek zamanda, yüksek başarımla, birden çok kanalın yayın akışının takibi ve loglanması için kullanılması planlanmaktadır. Şekil 2.1'de genel yapısı tanıtılan sistem, ses verisi girişini radyo yayınlarının yakalandığı bir ses kartından soket bağlantısıyla alıp, ses parmakizlerini çıkardıktan sonra bulunan parmakizlerini veritabanında eşleyerek kanal bazında yayın akışlarının kaydını tutma işlemlerini gerçekleyecektir.

Kartın bulunduğu sistem sunucu olarak çalışarak gelen istemcilere ilgili radyo kanalının gerçek zamanda yayın verisini kurulan soket bağlantısı üzerinden aktaracak. Sistem limitlerine göre belirlenecek istemci sayısına ulaşılan kadar sunucu yeni istekleri kabul edecek.

İstemciler soket üzerinden aldıkları yayın verisini ses hash'leme algoritması ile ses parmakizlerine dönüştürerek kanal ve zaman bilgisi ile arama modülüne aktaracak.

Arama modülü ses parmakizi veritabanına bağlanarak her bir kanal için yayınlanan ses verisini tanılayarak çalan parçanın bilgilerini çekecek. Bu bilgiler kanal ve zaman bilgileri ile birlikte kayıt modülünde loglanacak ve ilgili istemci kullanıcıya sunulacak.

Parmakizi hesaplama modülü, gelen ses verilerini 3'er saniyelik bölümler halinde ses parmakizlerine dönüştürür. Ses parmakizleri 0.37 saniyelik ses verisinden hesaplanan ses alt-parmakizlerinden oluşur. Ses verisi frekans uzayında insan duyu sisteminin duyarlı olduğu frekans aralığında logaritmik olarak sıralanan bantlara ayrılır. Bant enerjileri arasındaki farkın işareti kullanılarak ses alt-parmakizlerinin her bir biti hesaplanır. Ses verisi üzerinden yüksek üst üste binme oranı ve çerçeveleme yöntemi kullanılır. Sonuçta 256 ses alt-parmakizinden oluşan dizi bir ses verisini oluşturur ve bu bilgi ses verisini veritabanında tanılamak için yeterlidir.

Bir ses verisinin bu sistem aracılığı ile tanılanabilmesi için sistemin kullandığı(bağlandığı) ses parmakizi veritabanında, ses verisinin ait olduğu parçanın daha önceden kaydedilmiş olması gerekir. Aranılan ses parmakizinin veritabanındaki bir parmakizi ile eşleşebilmesi için her iki parmakizinin tüm bitleri belli bir hata oranı altında eşleşmelidir. Parmakizini oluşturan bit sayısı ile doğru eşleşmeyen bit sayısı arasındaki oran "Bit Hata Oranı" olarak tanımlanır.

Bit Hata Oranı sistem için belirlenen eşik değerinin altında olduğun ses verisinin ait olduğu parça ses veritabanında bulunur ve veritabanında kayıtlı bilgiler çekilir.

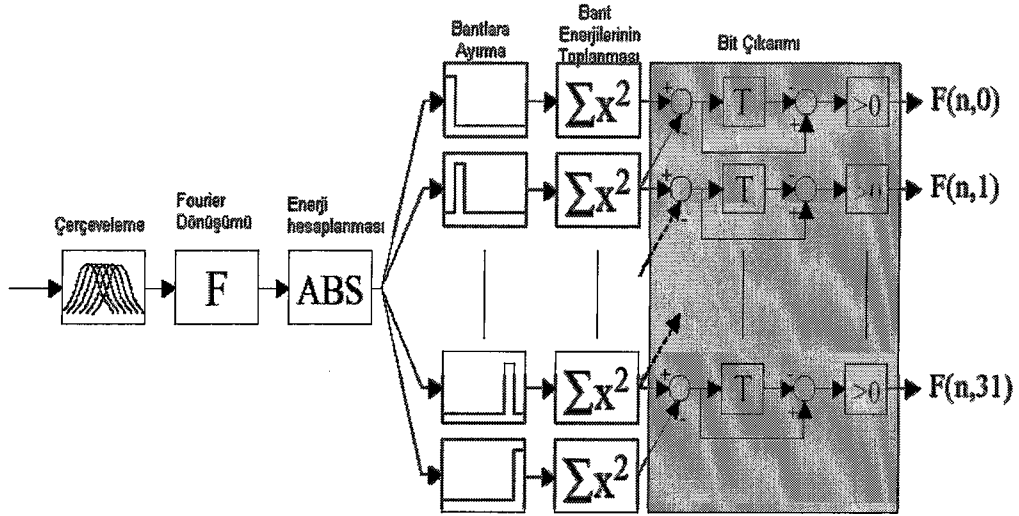
Ses veritabanı arama hızının artırılması için bir eşleme tablosu bulundurulur. Bu tabloda, veritabanına kayıtlı parçalarda geçen ses alt-parmakizlerinin parçalara göre indeksleri tutulur. Yeni bir arama isteği geldiğinde ses parmakizindeki her alt-parmakizi önce eşleme tablosunda aranarak veritabanındaki hangi parçanın kaçırıncı indeksinde geçtiğı bulunur. Bu sayede veritabanında olası tüm yerlerde parmakizi karşılaştırması yapmak yerine ses alt-parmakizlerinin geçtiğı bölümler kullanılır.

2.1 Ses Parmakizlerinin Oluřturulması

Çoėu ses parmakizi çıkarım algoritması benzer işlemleri uygular. Ses işareti önce çerçevelere (frame) ayrılır. Her çerçeve için öznitelik hesaplaması yapılır. Hesaplanan bu özniteliklerin belli bir dereceye kadar işaret bozulmalarına karşı dayanıklı olması beklenir. Aynı zamanda insan parmakizi gibi ses verisini benzersiz bir şekilde tanılması ses parmakizinin temel hedefidir. Bu amaçla daha önceden Fourier katsayıları[1][11], Mel Frequency Cepstral Coefficients (MFCC)[12], spektrum düzlüėü (spectral flatness)[13], spektrum keskinliėi (spectral sharpness) [13], doğrusal öngörölü kodlama katsayıları (Linear Predictive Coding Coefficients) [13] öznitelik olarak kullanılmıştır.

Ses hash'lemesi ile ses verisinin belli bir bölümünü benzersiz olarak tanılamaya yarayan ses parmakizlerinin kodlanması sağlanır. Tanılama amacıyla ses verisinin tamamının ses parmakizi bir veritabanında tutulur. Çerçeveleme ile alt birimlere ayrılan ses verisi, frekans uzayında işlenerek tek başına bir bilgi içermeyen ses alt-parmakizlerine kodlanır. Zaman uzayında ardışıl ses alt-parmakizleri, ses verisini veritabanında eşlemeye yarayacak ses parmakizini oluşturur. Ses verisini tanılamak için gerek ve yeter bilgi içeren alt-parmakizi bloėuna ses parmakizi denir.

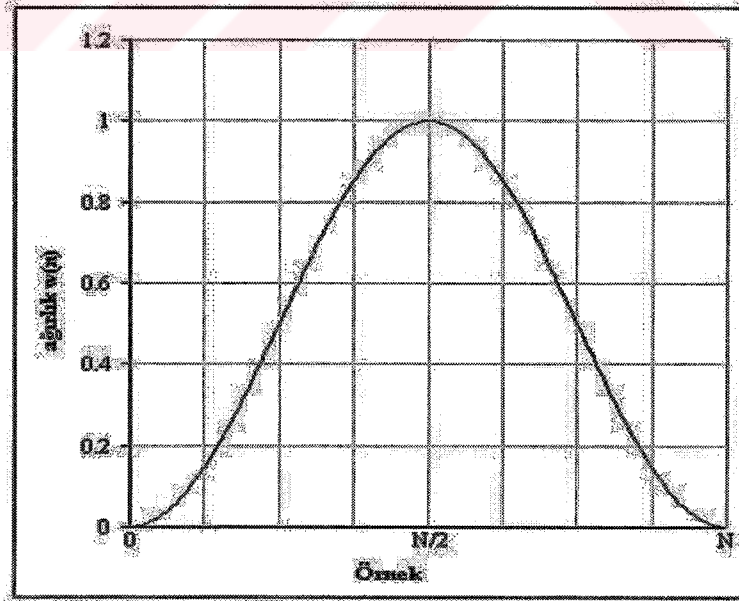
Sistemde ses verisi 0.37 saniyelik çerçevelere ayrılarak kullanılır. Çerçeveleme için Hanning pencereleme fonksiyonu kullanılır. Sistemin senkronizasyon bozukluklarına karşı dayanıklılıėını arttırmak için pencereleme fonksiyonu 31/32'lik üstü üste binme oranı ile kullanılır. Bu sayede ses versinin herhangi bir yerinden başlayarak oluşturulan ses parmakizlerinin de sistem tarafından başarıyla tanılanması sağlanır. Üst üste binme etkisiyle ses verisinin her 11.61 milisaniyesi için bir kodlama gerçekenmiş olur. Ses verisinin incelenme çözünürlüėü 11.61 milisaniyeye kadar düşürölmüş olur. Bu sayede en kötü durumda bile, tanılama için gelen ses parmakizinin zaman uzayında başlama ve bitiş sınırları ile veri tabanına giriş yapılırken kullanılan ses parmakizi çerçevelerinin başlama ve bitiş zamanları arasında 5.8 milisaniye gibi oldukça küçük ve sistemin doğru çalışmasını etkileyemeyecek bir kayma olur. Yüksek üst üste binme oranı çerçeve faz farkını çok düşük bir değerde tutmuş olur. Üst üste binen çerçevelerin kulanımı art arda gelen hash değeriğerlerinin zaman içinde çok yavaş değışim göstererek benzer olmalarını da sağlar.



Şekil 2.2: Ses Parmakizi çıkarım blok çizelgesi

Şekil 2.2’de genel yapısı tanıtılan ses parmakizi hesaplama modülü, önce giriş verisine Hanning çerçeveleme fonksiyonunu uygular (Şekil 2.3). Hanning pencereleme fonksiyonunun ağırlık değerleri Denklem 2.1’deki formüle göre hesaplanır.

$$w(n) = 0.5 - 0.5\cos(2\pi n/(N-1)) , (N-1) \geq n \geq 0 \quad (2.1)$$



Şekil 2.3: Hanning Pencereleme Fonksiyonu

Çerçevelemeden geçirilen giriş verileri Fourier Dönüşümü (Denklem 2.2) ile frekans uzayına geçirilir. Bunun nedeni en önemli algısal ses öz nitelikleri frekans uzayında yer almasıdır. Daha önceden de belirtildiği gibi uygulamada kullanılan ses hash'leme fonksiyonu sesin algısal öz niteliklerini modellemeye çalışmaktadır. Ancak Fourier Dönüşümü, insan duyu sisteminin tersine çerçeve kaymaları ile oluşan faz değişimlerine oldukça duyarlıdır. Bu olumsuz etkiden kurtulmak için ses verisi frekans uzayına geçirilirken enerjiler hesaplanır.

$$X(f) = \sum_{n=0}^N x'(n) e^{-2\pi i f n / N} \quad (2.2)$$

Tüm bu işlemlerden sonra hash'leme fonksiyonunun kullandığı ses öz nitelikleri hesaplanır. Hesaplanan enerji değerleri frekans uzayında üst üste binmeyen 33 banta ayrılır. Frekans enerji bantları insan duyu sisteminin en duyarlı olduğu 300Hz-2000Hz arasında logaritmik olarak genişleyecek şekilde seçilir. Logaritmik sıralanma insan duyu sisteminin modellenmesi amacıyla seçilmiştir, çünkü insan duyusu logaritmik genişleyen bantlarda çalışır. Logaritmik genişleyen bantlar Denklem 2.3'e göre hesaplanır.

$$f(i) = f(0)(f(n-1) / f(0))^{i/(n-1)}, \quad 0 \leq i < n \quad (2.3)$$

Her bant aralığına düşen enerji değerleri toplanarak frekans bant enerjileri hesaplanmış olur (Denklem 2.4).

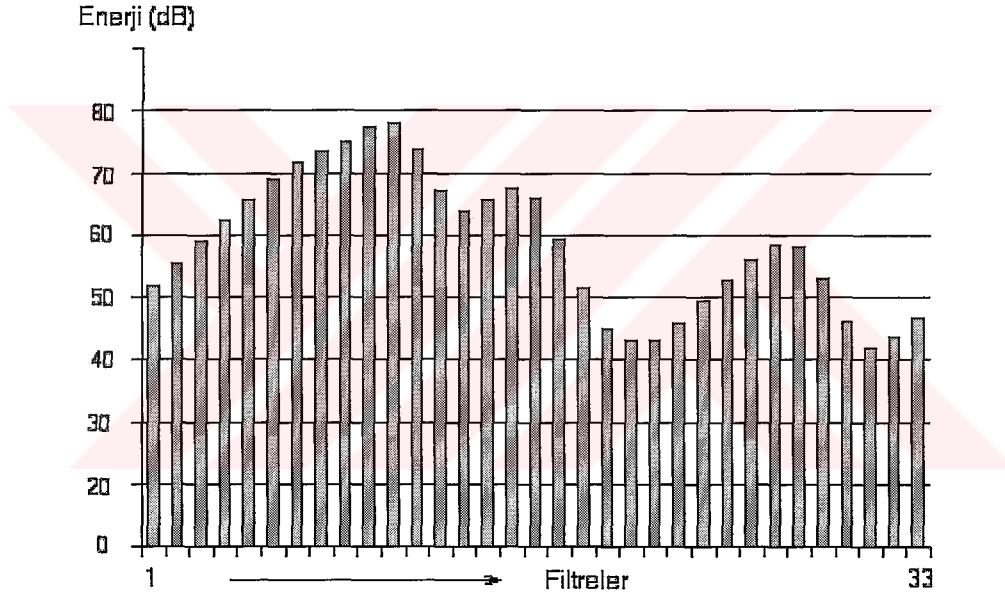
$$E(i) = \sum_{f=f_i}^{f=f_{i+1}} |X(f)|^2, \quad 300\text{Hz} \leq f_0, f_1, \dots, \leq 2000\text{Hz} \quad (2.4)$$

Frekans uzayında ardışıl bant enerjileri hash'leme fonksiyonunun giriş verisini oluşturur. Ardışıl bant enerjileri arasındaki fark hesaplanır. Sonucun işareti ses çerçevesine karşı gelecek alt-parmakızının (hash değerinin) bir bitini belirler. 33 enerji bandından hesaplanan 32 enerji bant farkı birer bit değerine karşı gelir. Sonuçta 32 bitlik bir ses alt-parmakızı kodlanmış olur. Ses hash'lemesi için kullanılan bu öz niteliğin pek çok bozulmaya karşı yüksek dayanıklılıkta olduğu görülmüştür. Bant enerjileri kullanılarak oluşturulan hash değerinin bir bitinin hesaplanması Denklem 2.5'te gösterilmiştir.

$$F(n,m) = \begin{cases} 1 & \text{if } E(n,m) - E(n,m+1) - (E(n-1,m) - E(n-1,m+1)) > 0 \\ 0 & \text{if } E(n,m) - E(n,m+1) - (E(n-1,m) - E(n-1,m+1)) \leq 0 \end{cases} \quad (2.5)$$

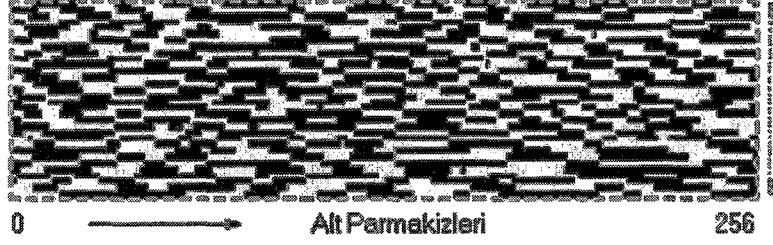
n. çerçevenin m. enerji bandı E (n,m) ile, n. çerçeve için hesaplanan hash değerinin m. biti F (n,m) ile gösterilmiştir.

Şekil 2.4'te ses verisinin 0.37 saniyelik bir çerçevesindeki enerji farklarının frekans bantlarına bağlı değişimi görülmektedir. Art arda gelen iki enerji bandı arasındaki farklar oldukça büyüktür. Yalnızca 21. ve 22. bantlar birbirine çok yakındır, bunların dışındaki farklar en az 2–3 dB civarındadır.



Şekil 2.4: Bant Enerjileri.

Ses verilerini tanılamaya yarayan bilgi, art arda gelen 256 ses alt-parmakizinden oluşan ses parmakizinde bulunur. Bir ses parmakizi yaklaşık 3 saniyelik ses verisine karşı gelir ve toplamda 8192 bitten oluşur. Üst üste binme nedeniyle ortalama 5 dakikalık bir şarkıda yaklaşık 25000 ses parmakizi bulunur.



Şekil 2.5: Ses Parmakizi

Hesaplanan 256 hash değeri yanyana geldiğinde (Şekil 2.5) ses verisi için parmakizi niteliğinde bir görüntü oluşturur ve bu parmakizi ses verisini diğerlerinden ayırabilmemizi sağlar.

2.2 Veritabanı Yapısı Ve Arama Algoritması

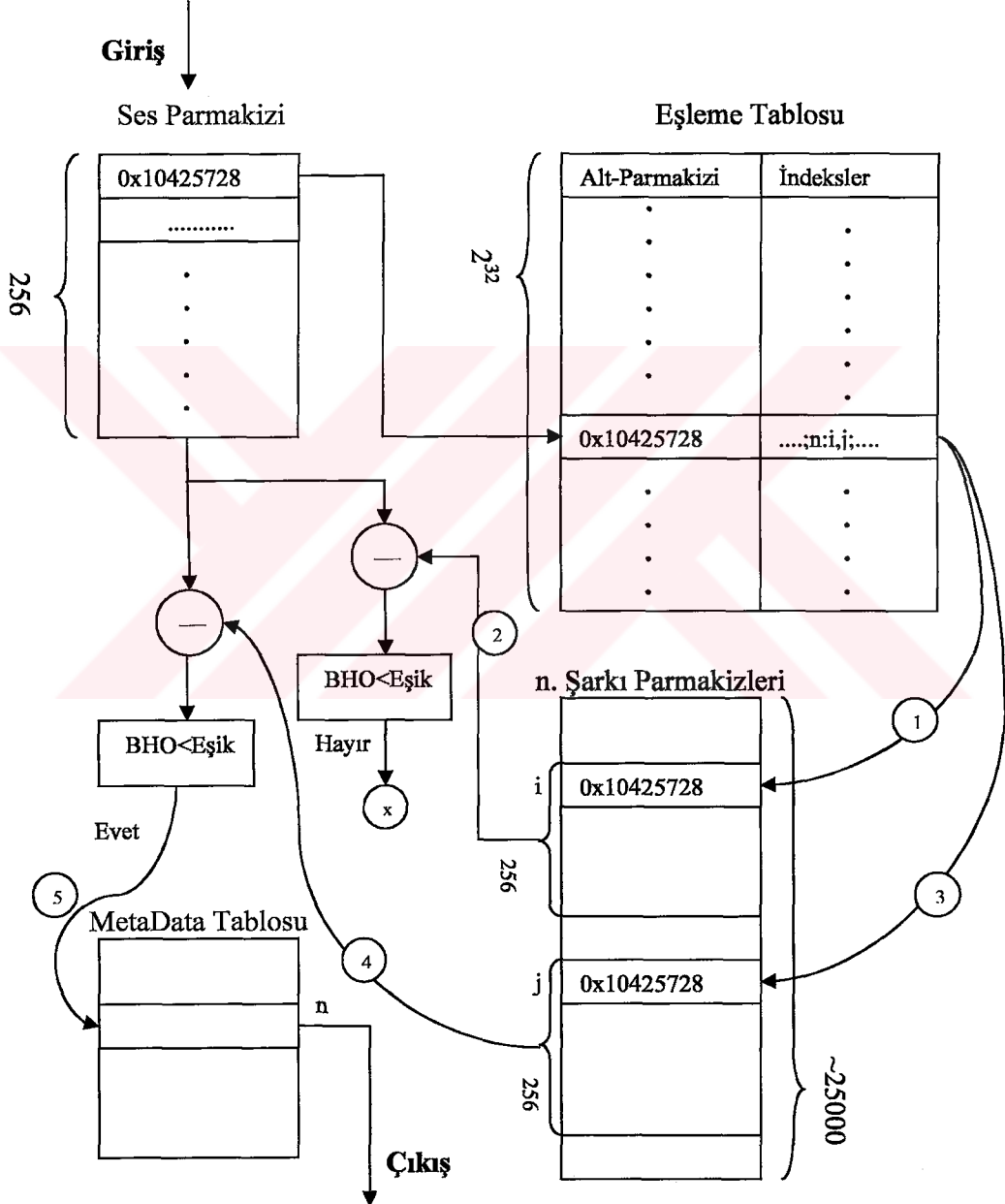
Ses parmakizlerinin veri tabanında aranması basit bir işlem değildir, çünkü bire bir eşleşen parmak izi yerine en çok benzeyen (Bit Hata Oranı hata oranının altında) parmak iz(ler)i aranmaktadır. Aranacak bir parmakizini veri tabanında olası tüm yerlerle karşılaştırmak ve BER eşik değerinin altında kalan parçaları bulmak, tek bir parmakizi araması için onlarca dakika ayırmak anlamına gelir.

Veritabanına giriş yapmak için bir ses verisinin baştan sona tüm ses alt-parmakizlerinin çıkarılması gerekir. Tüm hash değerleri ses verisine ait özel bir tabloda tutulur. Hash değerleriyle birlikte arama sonucunda ses verisi hakkında verilebilecek bilgiler de veritabanına girilir. Tüm ses verileri için MetaData olarak adlandırılan bilgiler tek bir tabloda tutulur ve tekrarlanmayan benzersiz bir “ID” değeri ile arama tablosuna bağlanır. Arama tablosu ses verilerinin özel tablolarında doğru yerlere daha kolay ve hızlı erişme imkanı sağlar. Bu tabloda, veritabanındaki ses verilerine ait her bir hash değerinin hangi şarkının kaçınıcı indeksinde geçtiği tutulur. Böylece arama için gelen bir ses parmakizindeki her bir alt-parmakizinin veritabanındaki hangi ses verisinin kaçınıcı indeksinde yer aldığı kolaylıkla bulunmuş olur.

10,000 şarkılık bir veri tabanında, her şarkının ortalama 5 dakika uzunluğunda olduğu kabul edilirse, 250 milyon hash değeri ve 250 milyon ses parmak izi bulunur. Her aramada 250 milyon karşılaştırma yapmak yerine, uygulanan sistem için çok daha etkin bir yöntem kullanılmıştır (Şekil 2.6). Gelen ses parmakizinde geçen alt-parmakizlerinin veritabanında bulunduğu yerlerden ses parmakizleri çekerek

karşılaştırma yapmak, tüm veritabanını taramaktan 800,000 kat daha hızlı bir yöntem sunmaktadır[3].

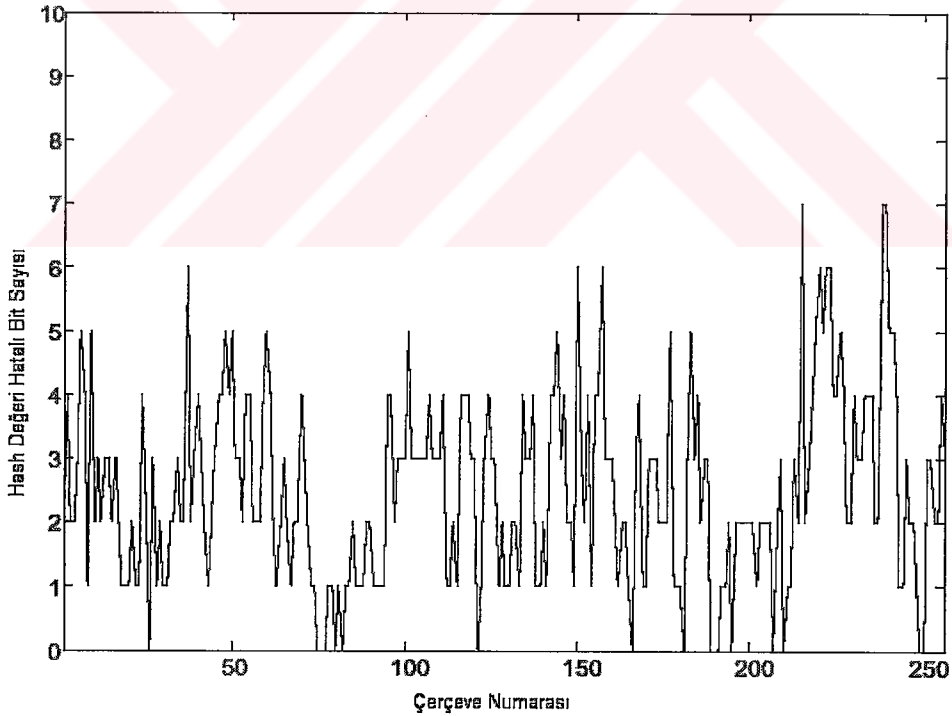
32 bitlik hash değerleri 2^{32} farklı değer alır. Pratikte 2^{32} değer için bilgi tutmak bellek kaynakları açısından çok uygulanabilir değildir. Ancak her hash değeri için bir kayıt tutmak yerine yalnızca şarkılarda geçen hash değerlerinin tutulduğu bir arama tablosu uygulamada kullanılır.



Şekil 2.6: Veritabanında Arama

Arama için gelen ses parmakizinden sırayla bir alt-parmakizi alınarak arama tablosundan bu alt-parmakizinin geçtiği ses verilerinin tabloları ve indeksleri çekilir. Daha sonra bu indekslerden yer alan ses parmakizleri ile arama için gelen ses parmakizi karşılaştırılarak aralarında benzeşmeyen bit sayısı bulunur. Hatalı bit sayısının ses parmakizinde yer alan tüm bit sayısına oranı, karşılaştırılan iki ses parmakizinin benzerliğini tanımlar. Bit Hata Oranı, sistemin kabul ettiği eşik değerinin altındaysa bu iki parmak izi eş olarak kabul edilir ve arama için gelen ses parmakizine karşı gelen ses verini tanımlanmış olur. Uygulamada eşik değeri olarak 0.25 gibi yüksek bir değer kullanılmıştır.

Uygulanan arama sistemi, gelen ses parmakizindeki hash değerlerinden en az bir tanesinin hiçbir bozulmadan etkilenmeden ve hiçbir biti değişmeden kaldığı kabulüne dayanmaktadır [1]. Şekil 2.7’de, 128Kbit’lik bir Mp3 dosyasından oluşturulan parmakizine ait hash değerlerindeki hatalı bit sayısı gösterilmektedir. 256 hash değerinden 17’sinde hiç hatalı bit bulunmamaktadır. Bu da başta yapılan kabulün doğruluğunu gösterir.



Şekil 2.7: 128Kbps Mp3 Parmakizi Hash Değerlerindeki Hatalı Bit Sayısı

10,000 şarkılık veritabanında 250 milyon hash değeri tutulduğuna göre; bir hash değerinin şarkılarda bulunma oranı:

$$0.058 = 250 \times 10^6 / 2^{32} \quad (2.6)$$

olacaktır. Ses parmakizindeki tüm hash değerleri için veri tabanında karşılaştırma yapıldığında toplamda sadece $(256 \times 0.058=)$ 15 karşılaştırma işlemi yapılmış olur. Gerçekte hash değerleri ideal rastlantısalılıkta dağılmadığı için bir hash değerinin şarkılarda bulunma oranı pratikte, hesaplanandan yaklaşık 20 kat daha yüksektir. Sonuçta 300 karşılaştırma ile parmakizi arama işlemi tamamlanır.

Eğer yüksek bozulmalarda da ses parmakizinin tanınması isteniyorsa, gelen alt-parmakizinin hiç bozulmaya uğramadığı kabulü yerine bir bitinde hata olabileceği kabul edilir. Bu durumda her bir alt-parmakizi yerine bir bitinin değişimiyle elde edilecek toplam 33 alt-parmakizi için arama yapılır. Sonuçta arama sayısı 33 kat artar ancak yine de kabul edilir sınırlar içinde kalır[3].

2.3 Hata Analizi

Tasarlanan sistem giriş olarak en az 3 saniyelik bir ses verisi ya da buna karşı gelen ses parmakizi alır. Amaç, bu veriyi doğru bir şekilde tanılamaktır. Bunun için daha önceden hazırlanmış bir ses veritabanı kullanılır. Veritabanı ses verilerinin orjinallerinden çıkarılan ses parmakizlerini bulundurmaz. Ancak arama için gelen ses parmakizi bir radyo yayınından, telefondan ya da mp3 sıkıştırma ile oluşturulmuş bir dosyadan alınmış olabilir. Tüm bu ve benzer bozulmalara karşı gelen ses parmakizinin başarıyla tanınması istenmektedir.

Sistemde iki tür hata oluşabilir: gelen ses parmakizinin veritabanında karşılığı olduğu halde tanınmaması ve bir ses parmakizinin yanlış tanınarak ait olmadığı bir ses verisiyle eşleştirilmesi.

İki farklı 3 saniyelik ses verisinden çıkarılan ses parmak izleri arasındaki Hamming uzaklığı belirli bir eşik değerinin (T) altındaysa, bu iki ses verisi eş olarak adlandırılır. Eşik değeri, T , yanlış tanıma (P_f) oranı ile doğru orantılıdır. Benzer ses verilerini tanıyamama oranı, P_n , ise T azaldıkça artar.

Hash çıkarma işlemi rastlantısal (bağımsız ve özdeş dağılımlı) bitler üretir. Hatalı bit sayısı binominal dağılım gösterir (Denklem 2.7a)

$$P_f(\alpha) = \frac{1}{2} \operatorname{erfc} \left(\frac{(1-2\alpha)}{3\sqrt{2}} \sqrt{n} \right) \quad (2.7a)$$

Ses parmakızinde bulunan bit sayısı n'dir(256x32=8192). Ses alt-parmakızinde yer alan her bir bitin 0 ya da 1 değeri alma olasılığı ise p'dir(0.5). Burada toplam bit sayısı yeterince büyük olduğu için binominal dağılım yerine Gauss dağılımı kullanılabilir. Buna göre, "Bit Hata Oranı" (BER)(=α), ortalama değeri μ=nxp ve standart sapması σ = √(np(1-p)) olan bir normal Gauss dağılımı ile yanlış tanıma oranı hesaplanabilir (Denklem 2.7b)

$$P_f(\alpha) = \frac{1}{\sqrt{2\pi}} \int_{(1-2\alpha)\sqrt{n}}^{\infty} e^{-x^2/2} dx = \frac{1}{2} \operatorname{erfc} \left(\frac{(1-2\alpha)}{\sqrt{2}} \sqrt{n} \right) \quad (2.7b)$$

Bu modelleme ile, örneğin, bit hata oranı olarak 0.35 gibi yüksek bir değer alındığında bile (8192 bitten 2876'sının hatalı olması), hatalı tanıma oranı 10⁻²⁰ 'li değerler alır. Ses hash'lemesinin ayrıştırıcılığı çok yüksektir ve yanlış tanılamayı neredeyse ortadan kaldırır.

2.3.1 Uygulama karmaşıklığı

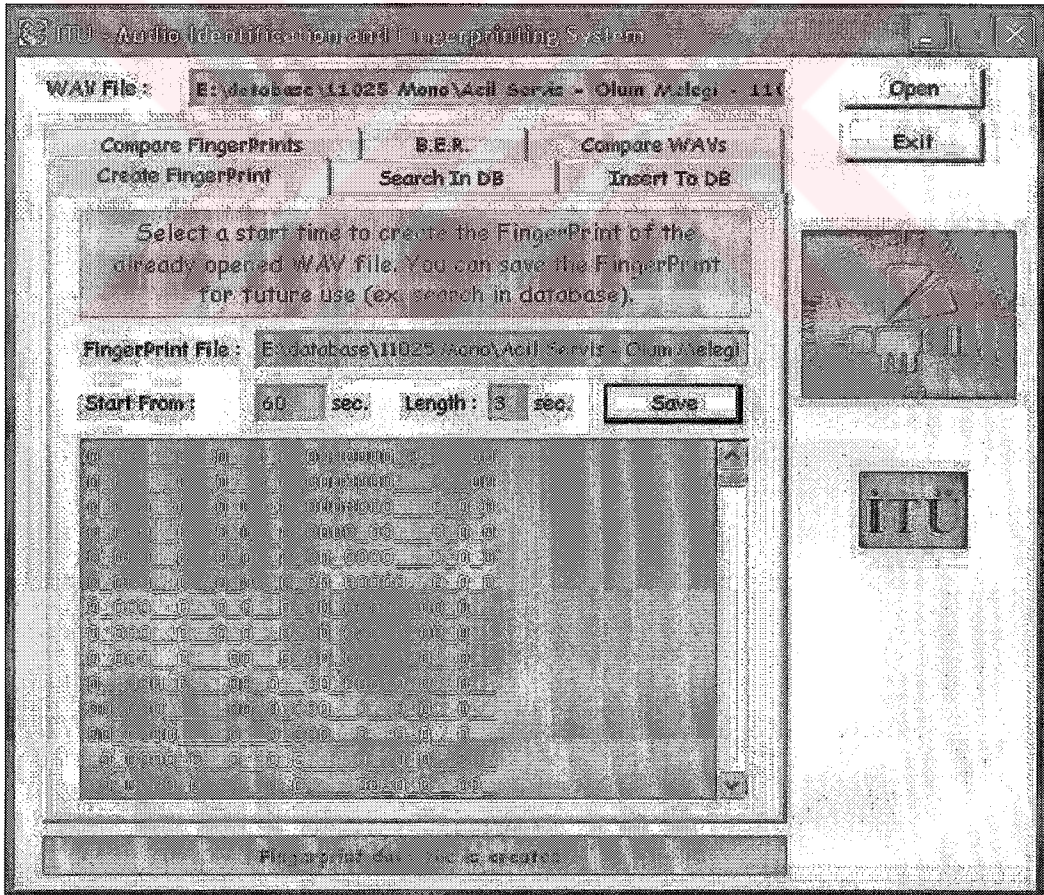
Yazılım Ms Visual Studio .Net ortamında C# dili kullanılarak kodlanmıştır. Ses veritabanı için Ms Sql Server 2000 kullanılmıştır. Veritabanı işlemlerinin tamamı Sp çağrılarını ile gerçekleştirilerek uygulama arayüzünün kod kısmında bu çağrılar dışında herhangi bir veritabanı işlemi yapılması engellenmiştir.

Uygulamada birincil karmaşıklığı ve işlem yükünü yüksek örnekleme değerleri (44100Hz) ile gelen ses verilerinin Fourier Dönüşümlerinin hesaplanması oluşturmaktadır. Ses verisi 2'nin kuvveti olacak şekilde pencerelemeden geçirilerek Hızlı Fourier Dönüşümünün uygulanabilmesi sağlanmıştır. Bu dönüşüm için Numerical Recipes'de [14] tanıtılan algoritma C# dilinde gerçekleştirilmiştir. Veritabanında arama karmaşıklığı; bir arama tablosu oluşturularak, hash değerlerini indeksli olarak onluk düzende sayılar halinde tutarak ve ses verilerine ait tablolara sıralı indeks değerleri ile erişerek ortadan kaldırılmaya çalışılmıştır.

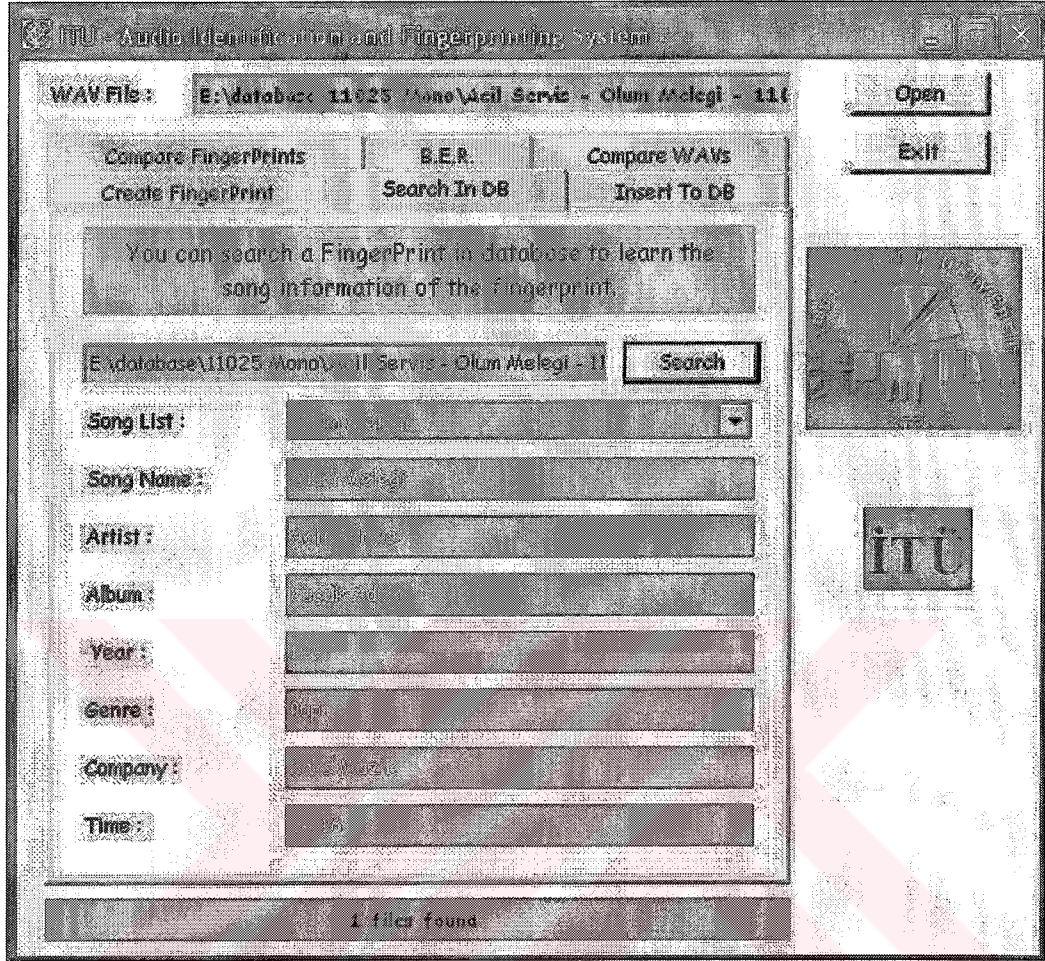
3. UYGULAMA ARAYÜZÜ VE DENEYSEL SONUÇLAR

3.1 Uygulama Arayüzü

Uygulama arayüzü kullanıcıya değişik işlemler yapma olanağı sunmaktadır. Sistem çalışırken ilk yapılması gereken ses klbinden bir ses parmakizi oluşturmaktır (Şekil 3.1). Bu amaçla bir ses dosyası açıldıktan sonra, istenen ses parmakizinin uzunluğu (saniye cinsinden) ve ses dosyasındaki başlangıç anı belirtilerek bir ses parmakizi oluşturulabilir. Ses parmakizi hem arayüzden kullanıcıya sunulur hem de *.fpf uzantılı bir dosya olarak kaydedilebilir. Daha sonra bu dosya uygulamanın diğer bölümlerinde kullanılabilir.

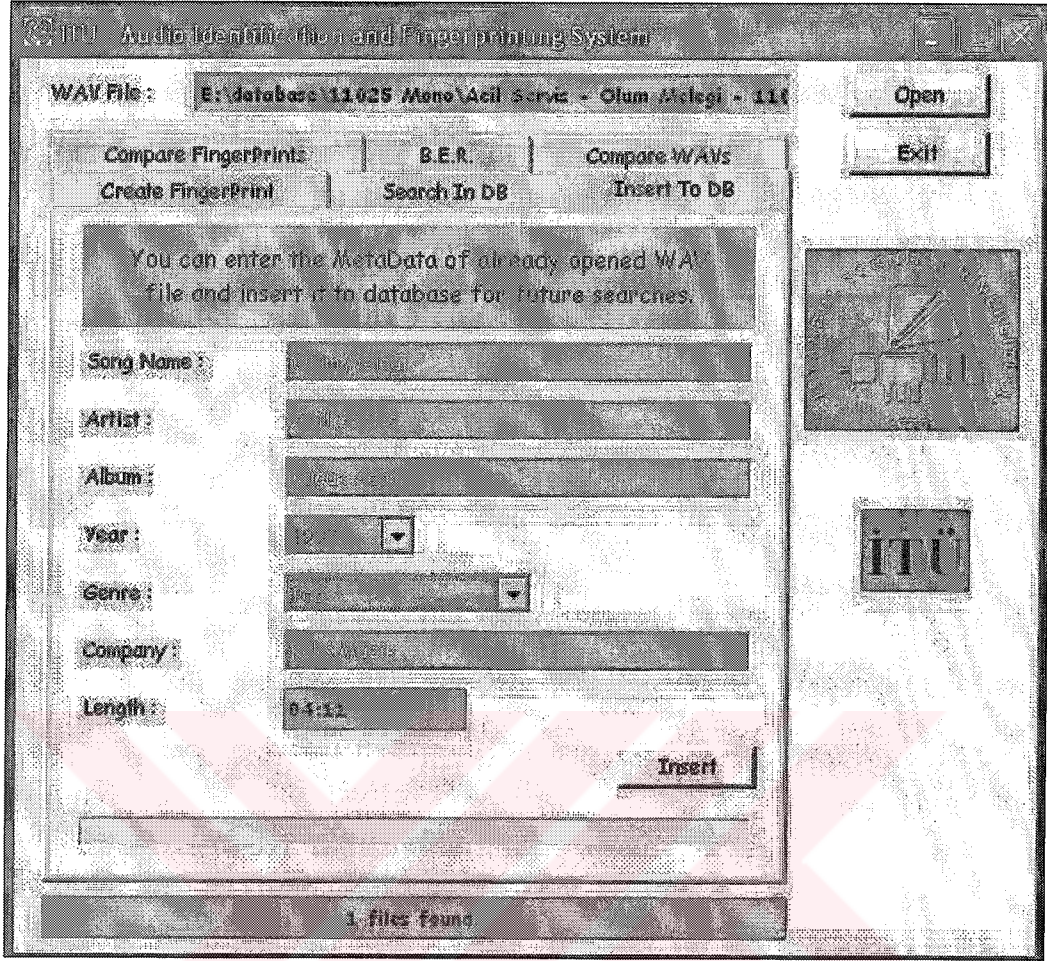


Şekil 3.1: Ses Parmakizi Oluşturma



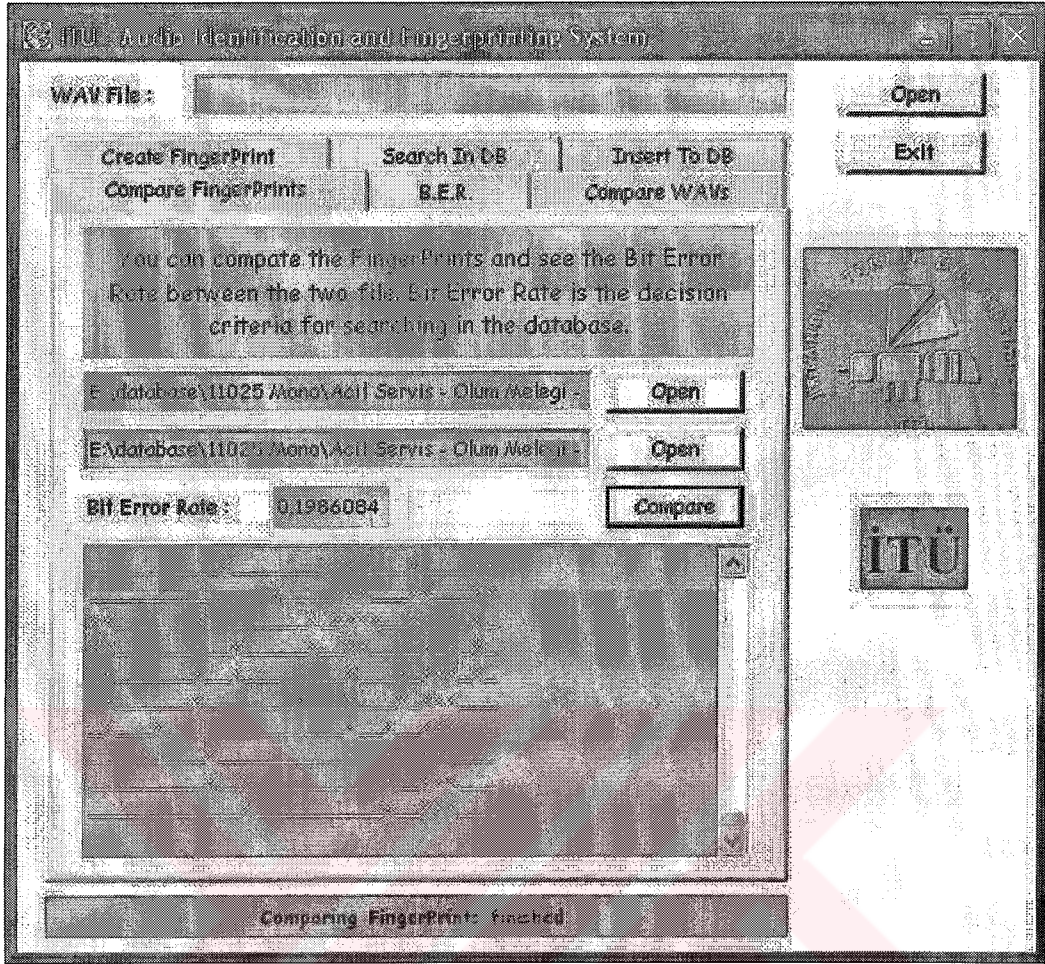
Şekil 3.2: Ses Parmakizinin Veritabanında Aranması

Bir önceki adımda oluşturulan ses parmakizleri ses veritabanında aranmak üzere arayüze giriş parametresi olarak verilirse, sistem ses parmakizinin ait olduğu parçayı arar ve bulduğunda veritabanında parçayla birlikte kaydedilmiş bilgileri çekerek kullanıcıya sunar (Şekil 3.2). Burada parça ismi, albüm, grup ismi, tür, yayınlanma yılı gibi pek çok “meta-data” olarak adlandırılan bilgiye ulaşılabilir. İstenirse bu bilgiler kullanılarak arama-getirme olanaklıdır.



Şekil 3.3: Veritabanına Yeni Kayıt Girişi

Geliştirilen sistem arayüzü, yeni bir ses klibi alındığında bununla ilgili veritabanı kaydı yapılmasına olanak tanır (Şekil 3.3). Uygulama içinde açılan wav dosyası ile ilgili gerekli bilgiler kullanıcı tarafından girildikten sonra veritabanına eklenir. Ancak sistem tekrarları önlemek için ses verisini veritabanına eklemeden önce bir ses parmakizi çıkarıp veritabanında arama yapar. Eğer eşleşen bir kayıt bulunursa parçanın veritabanına girişine izin verilmez. Böylece kayıt tekrarı engellenmiş olur.

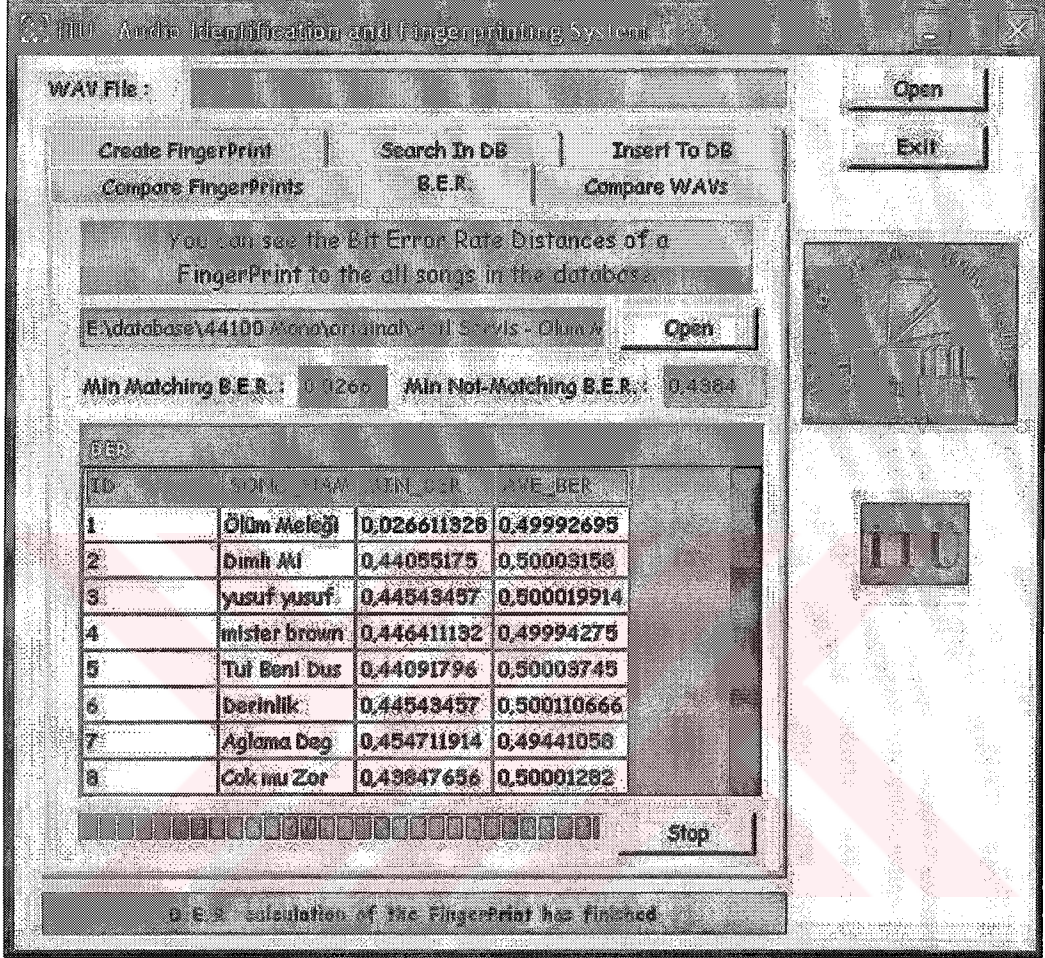


Şekil 3.4: Ses Parmakizlerinin Karşılaştırılması

Sistem klip tanılama işlemi için ses parmakizlerini kullanır ve bunlar arasındaki ilişkiyi Bit Hata Oranı (BHO) olarak raporlar. Bu amaçla, kullanıcının elinde bulunan iki ses parmakizini karşılaştırıp aralarındaki bit hata oranını öğrenebileceği bir bölüm de sistem içerisinde yer almaktadır (Şekil3.4). Kullanıcı eşit uzunlukta ses kliplerine ait parmakizlerini karşılaştıracak ve bir tanılama yapmak isteyecektir. Bu nedenle arayüzde ses parmakizi uzunluğu kontrolü bulunmaktadır.

Bir radyo istasyonundan dinlenen parçaya ilişkin parmakizi bulunduktan sonra, sistem, veritabanındaki tüm ses verilerinin olası tüm ses parmakizleri ile karşılaştırılması sonucu, ses parmakizinin ait olduğu parçaya ve veritabanındaki diğer parçalara olan bit hata oranı uzaklıklarını hesaplar ve en yakın olunan ses klipi ile eşleme yapar. Şekil 3.5’te görüldüğü gibi, geliştirilen arayüz, her bir ses kliline ait parmakizinin kendisine olan en küçük uzaklığını (MIN_BER) ve veri tabanındaki

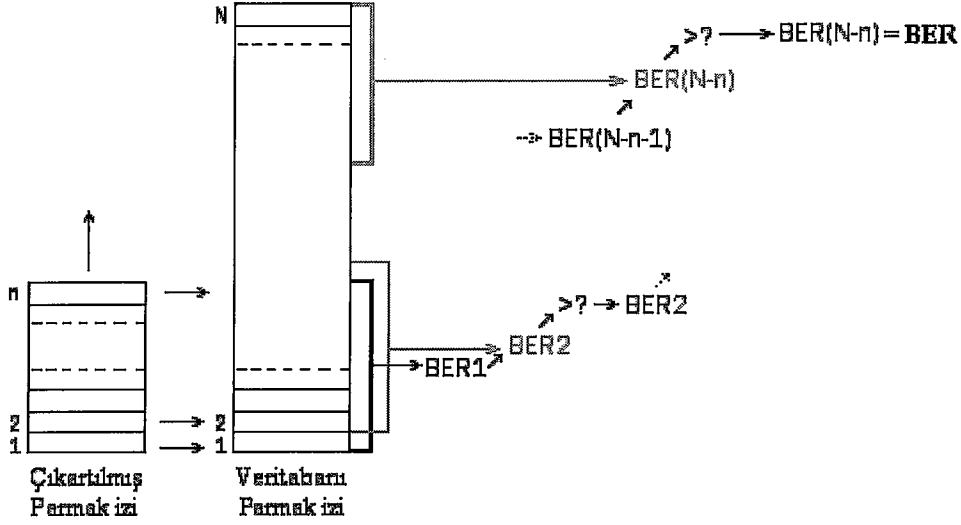
tüm parçalardan ortalama uzaklığını (AVE_BER) raporlar. Böylelikle geliştirilen ses hash'lemesinin başarımı test edilebilmektedir.



Şekil 3.5: Bit Hata Oranlarının Bulunması

3.2 Deneysel Sonuçlar

Sistemin başarımının irdelenmesi amacıyla bir test arayüzü geliştirilmiştir. Testlerde amaç, gelen parmakizinin veri tabanındaki tüm kayıtlara ne kadar yakın ya da uzak olduğunun araştırılması ve raporlanmasıdır. Parmakizine en yakın kaydın bulunması yerine, parmakizinin veri tabanındaki her bir şarkıya en benzer bölümlerinin bulunması ve bu benzerliğin hatalı bit oranı (BER) cinsinden değerlendirilmesi amaçlanmıştır. Bu amaçla kullanılan karşılaştırma yöntemi Şekil 3.6'da sunulmuştur.



Şekil 3.6: Karşılaştırma Yöntemi

3.2.1 Fm radyo yayını bozucu etkisine dayanıklılık

Geliştirilen sistemin bir yayın izleme ve raporlama sistemi olarak kullanılması hedeflendiğinden, İTÜ Radyo vericisi kullanılarak yapılan FM radyo yayını ile kaydedilen 44100Hz örnekleme hızlı Mono ses dosyalarından oluşturulan parmakizleri 44100Hz Mono şarkıların hash değerlerinin tutulduğu veri tabanında aranmıştır. Sonuçlar Tablo 3.1’de sunulmaktadır. Sonuçlardan görüldüğü gibi, parmakizinin ait olduğu şarkıya olan en küçük uzaklığı $BER = 0,102294922$ dir ve kabul edilir sınırlarda kalmıştır, parmakizi başarıyla tanınmıştır. Parmak izinin veri tabanındaki diğer şarkılara uzaklığı 0,4 den fazladır ve tüm ses klipleri için benzer sonuçlar alınmıştır ki bu da sistemin tutarlılığının bir göstergesidir.

3.2.2 Mp3 sıkıştırmaya dayanıklılık

Günümüzde ses dosyaları birçok uygulamada sıkıştırılmış olarak kullanılmaktadır. Bu nedenle, ses parmakizi sisteminin sıkıştırmalara karşı dayanıklılığını test etmek için değişik oranlarda (128Kbit, 96Kbit, 32Kbit) mp3 formatıyla sıkıştırılmış ses verilerinden çıkarılan parmak izleri 44100Hz ile örneklenmiş ses verilerine ait parmakizlerinin saklandığı veri tabanıyla karşılaştırılmıştır (bkz Tablo 3.2, Tablo 3.3, Tablo 3.4). Sonuçta 32Kbit sıkıştırmada bile parmakizinin, ait olduğu şarkıya olan

uzaklığı (hatalı bit oranı) sistemin çalışması sırasında kullanılan eşik değeri olan 0.25'in altında çıkmıştır. Diğer şarkılarla uzaklıklar yine 0.45'ler mertebesinde. Bu da sistemin ayrıştırıcılığını göstermektedir.

Tablo 3.1: 44100Hz'lik radyo kaydının veri tabanı ile karşılaştırılması

Ölüm Meleği 44100 mono (50. sn) radyo		
Şarkı Adı	Min. BHO	Ort. BHO
Ölüm Meleği	0,102294922	0,499948681
Dımlı Mi	0,441650391	0,50006089
Yusuf Yusuf	0,441162109	0,500039295
Mister Brown	0,441772461	0,5000683
Tut Beni Dusmeden	0,440673828	0,500172802
Derinlik	0,44543457	0,500152089
Aglama Degmez Hayat	0,453369141	0,496889394
Cok mu Zor	0,434326172	0,500026873
Bu Sabahların Bir Anlamı Olmalı	0,44909668	0,500025085
Bu Sabahların Bir Anlamı Olmalı(radyo mix)	0,44519043	0,50009797
Iz Birakanlar Unutulmaz	0,443359375	0,499927093

Tablo 3.2: 128Kbit Mp3 sıkıştırmanın veri tabanı ile karşılaştırılması

Ölüm Meleği 44100 mono (50. sn) 128Kbit		
Şarkı Adı	Min. BER	Ort. BHO
Ölüm Meleği	0,0859375	0,499950377
Dımlı Mi	0,442138672	0,50006242
Yusuf Yusuf	0,446411133	0,500019637
Mister Brown	0,439941406	0,499939654
Tut Beni Dusmeden	0,442138672	0,500101913
Derinlik	0,446166992	0,500063102
Aglama Degmez Hayat	0,454345703	0,498252174
Cok mu Zor	0,44128418	0,49998222
Bu Sabahların Bir Anlamı Olmalı	0,445800781	0,499943413
Bu Sabahların Bir Anlamı Olmalı(radyo mix)	0,445678711	0,500063218
Iz Birakanlar Unutulmaz	0,440673828	0,499872403

Tablo 3.3: 96Kbit Mp3 sıkıştırmanın veri tabanıyla karşılaştırılması

Ölüm Meleği 44100 mono (50. sn) 96Kbit		
Şarkı Adı	Min. BER	Ort. BHO
Ölüm Meleği	0,12121582	0,499936252
Dımlı Mi	0,44519043	0,499989647
Yusuf Yusuf	0,443725586	0,500049148
Mister Brown	0,440063477	0,499910658
Tut Beni Dusmeden	0,442138672	0,499916013
Derinlik	0,444580078	0,499998584
Aglama Degmez Hayat	0,445678711	0,494726214
Cok mu Zor	0,436035156	0,500004426
Bu Sabahların Bir Anlamı Olmalı	0,446899414	0,499874993
Bu Sabahların Bir Anlamı Olmalı(radyo mix)	0,44543457	0,500047649
Iz Birakanlar Unutulmaz	0,448730469	0,499848647

Tablo 3.4: 32Kbit Mp3 sıkıştırmanın veri tabanıyla karşılaştırılması

Ölüm Meleği 44100 mono (50. sn) 32Kbit wav		
Şarkı Adı	Min. BER	Ort. BHO
Ölüm Meleği	0,222045898	0,499928157
Dımlı Mi	0,430786133	0,49996773
Yusuf Yusuf	0,451171875	0,500023404
Mister Brown	0,44543457	0,499925558
Tut Beni Dusmeden	0,442626953	0,499960565
Derinlik	0,449951172	0,500089594
Aglama Degmez Hayat	0,446044922	0,495762118
Cok mu Zor	0,447875977	0,499996541
Bu Sabahların Bir Anlamı Olmalı	0,439697266	0,499875271
Bu Sabahların Bir Anlamı Olmalı(radyo mix)	0,443725586	0,500116343
Iz Birakanlar Unutulmaz	0,435913086	0,499864836

3.2.3 Ses örnekleme hızı dönüşümlerine dayanıklılık

Vericilerden yapılan yayın kalitesi yüksek olsa da alıcılar çok farklılık gösterebilmektedir. Bazı uygulamalarda ise iletim ortamından ve yayın türünden kaynaklanan band genişliği sınırlamaları olmaktadır. Bu nedenle ses tanılama sisteminin farklı örnekleme hızlarında çalışabilir olması önemlidir.

Tablo 3.5: Örnekleme hızı 22050Hz'e düşürülen ses dosyası

22050 örnekleme ölüm meleği (50. sn)		
Şarkı Adı	Min. BER	Ort. BHO
Ölüm Meleği	0,102172852	0,499943
Dımlı Mi	0,444824219	0,500028
Yusuf Yusuf	0,449951172	0,500057
Mister Brown	0,446777344	0,499929
Tut Beni Dusmeden	0,439575195	0,500005
Derinlik	0,447631836	0,500067
Aglama Degmez Hayat	0,449829102	0,4962
Cok mu Zor	0,436645508	0,500013
Bu Sabahların Bir Anlamı Olmalı	0,434082031	0,499914
Bu Sabahların Bir Anlamı Olmalı(radyo mix)	0,447753906	0,500076
Iz Bırakanlar Unutulmaz	0,43762207	0,499891

Farklı örnekleme hızlarında sistem başarımını gözlemlemek amacıyla, veri tabanını oluşturmada kullanılan 44100Hz Mono wav dosyaları değişik örnekleme hızlarında örneklendikten sonra tekrar 44100Hz'e çıkarılmıştır. Tablo 3.5 ten görüldüğü gibi, örnekleme hızı dönüşümleri sonucunda diğer parçalara uzaklıklar değişmemektedir. Örnekleme hızı 8000Hz gibi çok düşük bir değerde iken bile, eşik seviyesinin altında kalan bit hata oranıyla ses verileri veri tabanındaki karşılıklarıyla eşlenebilmektedir.

Tablo 3.6: Örneklemme hızı 11025Hz'e düşürülen ses dosyası

11025 örneklemme ölüm meleği (50. sn)		
Şarkı Adı	Min. BER	Ort. BHO
Ölüm Meleği	0,168212891	0,499963
Dımlı Mi	0,438476563	0,500022
Yusuf Yusuf	0,446411133	0,50006
Mister Brown	0,43737793	0,499951
Tut Beni Dusmeden	0,438964844	0,500019
Derinlik	0,44909668	0,500072
Aglama Degmez Hayat	0,443237305	0,496505
Cok mu Zor	0,442749023	0,499983
Bu Sabahların Bir Anlamı Olmalı	0,445556641	0,499938
Bu Sabahların Bir Anlamı Olmalı(radyo mix)	0,440063477	0,500072
Iz Bırakanlar Unutulmaz	0,443359375	0,499918

Tablo 3.7: Örneklemme hızı 8000Hz'e düşürülen ses dosyası

8000 örneklemme ölüm meleği (50. sn)		
Şarkı Adı	Min. BER	Ort. BHO
Ölüm Meleği	0,183349609	0,499941
Dımlı Mi	0,434692383	0,500007
Yusuf Yusuf	0,434692383	0,500027
Mister Brown	0,440429688	0,49999
Tut Beni Dusmeden	0,443115234	0,499984
Derinlik	0,445800781	0,500142
Aglama Degmez Hayat	0,448730469	0,501612
Cok mu Zor	0,44152832	0,500017
Bu Sabahların Bir Anlamı Olmalı	0,444335938	0,499948
Bu Sabahların Bir Anlamı Olmalı(radyo mix)	0,442382813	0,500133
Iz Bırakanlar Unutulmaz	0,437011719	0,499891

3.2.4 Hızlandırma saldırısına dayanıklılık

Ses dosyalarının orijinalden daha hızlı çalınması yayın akışına uyma ve programı kısaltma amacıyla çok sık uygulanmaktadır. Bu nedenle hızlandırma saldırısına karşı başarıımı ölçmek amacıyla, ses klipleri üzerinde hızlandırma saldırısı uygulandıktan

sonra çıkarılan parmakizlerinin veri tabanında aranması gerçekleştirildi. Yapılan testlerde kullanılan eşik değerinin altında sonuçlar elde edildi ve ses klipleri doğrulukla tanındı (bkz. Tablo 3.8, Tablo 3.9, Tablo 3.10).

Tablo 3.8: %1 Hızlandırma etkisi

%1 Hızlandırma ölüm meleği (50. sn)		
Şarkı Adı	Min. BER	Ort. BHO
Ölüm Meleği	0,102416992	0,499988
Dımlı Mi	0,434814453	0,500163
Yusuf Yusuf	0,440185547	0,500031
Mister Brown	0,44934082	0,500016
Tut Beni Dusmeden	0,448608398	0,500107
Derinlik	0,447143555	0,499948
Aglama Degmez Hayat	0,450317383	0,505808
Cok mu Zor	0,441894531	0,500126
Bu Sabahların Bir Anlamı Olmalı	0,444458008	0,500135
Bu Sabahların Bir Anlamı Olmalı(radyo mix)	0,447753906	0,500025
Iz Birakanlar Unutulmaz	0,437255859	0,499931

Tablo 3.9: %2 Hızlandırma etkisi

%2 Hızlandırma ölüm meleği (50. sn)		
Şarkı Adı	Min. BER	Ort. BHO
Ölüm Meleği	0,121337891	0,499935
Dımlı Mi	0,447021484	0,499994
Yusuf Yusuf	0,448364258	0,500086
Mister Brown	0,443725586	0,500006
Tut Beni Dusmeden	0,430419922	0,500153
Derinlik	0,446166992	0,50011
Aglama Degmez Hayat	0,455566406	0,501015
Cok mu Zor	0,443359375	0,500045
Bu Sabahların Bir Anlamı Olmalı	0,438232422	0,500051
Bu Sabahların Bir Anlamı Olmalı(radyo mix)	0,435668945	0,500109
Iz Birakanlar Unutulmaz	0,442382813	0,499986

Tablo 3.10: %3 Hızlandırma etkisi

%3 Hızlandırma ölüm meleği (50. sn)		
Sarkı Adı	Min. BER	Ort. BHO
Ölüm Meleği	0.154785156	0.499887
Dımlı Mi	0.441162109	0.499964
Yusuf Yusuf	0.444458008	0.500059
Mister Brown	0.441040039	0.500047
Tut Beni Dusmeden	0.444091797	0.500087
Derinlik	0.446899414	0.50007
Ağlama Degmez Hayat	0.447265625	0.499008
Bu Sabahların Bir Anlamı Olmalı	0.446289063	0.500034
Bu Sabahların Bir Anlamı Olmalı(radyo mix)	0.449829102	0.500086
Iz Bırakanlar Unutulmaz	0.439941406	0.499844

3.2.5 Ses parmakizi uzunluğu (granularity)

Günümüz uygulamalarında bir ses klibinin 5 saniyeden daha kısa bir süre içerisinde tanımlanması gerekmektedir. Bu nedenle sistemin gerek duyduğu en küçük parmakizi uzunluğunun belirlenmesi önemli bir başarıım parametresidir.

Tablo 3.11: Ses Parmakizi Uzunlukları

	11025Hz Wav Dosyaları							
	1sn parmakizi		2sn parmakizi		3sn parmakizi		4sn parmakizi	
	Eşleşen	Diğerleri	Eşleşen	Diğerleri	Eşleşen	Diğerleri	Eşleşen	Diğerleri
Synch.	0,02867	0,39816	0,02796	0,43165	0,0274	0,44628	0,0271	0,4518
fast96	0,12242	0,40551	0,16429	0,43238	0,21118	0,44433	0,26787	0,4516
mp3_64	0,10735	0,41102	0,10489	0,42635	0,10729	0,44482	0,12151	0,4539
mp3_32	0,19448	0,40514	0,18658	0,43475	0,18212	0,44775	0,18878	0,4492

Geliştirilen sistemde ses parmakizi uzunluğu olarak 3 saniye kullanılmaktadır. Ancak, başarıımı gözlemleyebilmek amacıyla, değişik uzunluklarda ses parmakizleri ile de testler yapılmıştır (Tablo 3.11). Küçük veritabanlarında 3 saniyeden kısa ses parmakizleri de iyi çalışmaktadır. Ancak burada kullanılan veritabanının özellikleri dikkate alınmalıdır. Testlerin yapıldığı 68 parçalı veritabanında toplam 1,660,279 değişik ses alt-parmakizi (hash değeri) bulunmaktadır ve bunların sadece 7680 tanesi birden fazla parçada yer almaktadır. Bu hem ses hash'lemesinin ayıricılığını hem de kısa ses parmakizlerinin de başarıyla çalışabilmesini açıklamaktadır. Ancak alınan işaret bozulmuş ise veya arama veritabanı çok geniş ise, yüksek başarımla çalışabilmek için 3 saniyelik ses parmakizlerinin kullanılmasının uygun olacağı sonucuna varılmıştır. Tablo 3.11 de farklı uzunluktaki parmak izleri ile çeşitli ataklara karşı başarıım raporlanmaktadır.

3.2.6 Gürültü ve diğer bozulmalar

İletim ortamından kaynaklanan bozulmalara karşı tanılama sisteminin başarımını gözlemlemek açısından ses dosyalarına farklı işaret-gürültü oranlarında Gauss gürültüsü eklenmiştir. Tablo 3.12 de gürültü ve diğer bozulmalar için, 68 parçalık genel bir veritabanında 44100Hz'lik parçalardan alınan ses parmakizleri ile yapılan testlerin sonuçları bulunmaktadır. Tablodan da görüldüğü gibi, sistemin gürültü atağına karşı 10dB'ye kadar başarımın yüksektir.

Tablo 3.12: Tüm Bozulmalar

	44100Hz 68 parça içinde arama	
	Eşleşen	Diğerleri
Synchronize	0,026611328	0,442903889
Radio	0,102294922	0,442825545
Fast %1	0,08581543	0,441994738
Fast %2	0,121337891	0,443025959
Fast %3	0,154785156	0,441909107
Fast %4	0,210327148	0,44267068
mp3 320Kbit	0,043701172	0,442411964
mp3 192Kbit	0,061889648	0,442377347
mp3 128Kbit	0,0859375	0,44205304
mp3 96Kbit	0,12121582	0,442894779
mp3 64Kbit	0,180053711	0,442284427
mp3 32Kbit	0,222045898	0,443924178
Resample to11025	0,168212891	0,442492129
Resample to22050	0,102172852	0,442721694
Resample to 8000	0,183349609	0,442391922
Snr 5DB	0,294067383	0,441305722
Snr 10DB	0,231933594	0,442322688
Snr 20DB	0,122070313	0,441486416
Snr 25DB	0,094482422	0,442492129
Snr 30DB	0,072265625	0,441542896

Genel olarak tüm testlerde ses parmakizlerinin ait olmadıkları parçalara Bit Hata Oranı uzaklıkları %44 seviyesinde çıkmıştır. Bu, 0.25 eşik değeri kullanan sistemin yanlış tanılama olasılığının düşüklüğünü göstermektedir.

Örnekleme hızının düşürülmesinde 8KHz'e kadar rahatlıkla tanılama başarılmıştır. Hızlandırma etkisinde %4'e kadar tanılama başarılmaktadır. Mp3 sıkıştırmada 128Kb için oldukça iyi bir bit hata oranı elde edilmiştir. Sıkıştırma etkisinde 32Kb gibi düşük bir değerde sistem eşik değerine yakın ama yine de altında bir sonuç üretmiştir. 10dB'lik gürültü atağı sistemin tanılama sınırında yer alır. Bunun üzerindeki (20dB-30dB) ataklarda başarım sağlanmıştır. Radyo yayını da başarıyla tanılanabilmiştir.

4. SONUÇLAR VE PLANLANAN ÇALIŞMALAR

Bu tez çalışmasında, yüksek performanslı bir ses hash'leme sistemi gerçekleştirilmiştir. Ses parmakizi çıkarımı her 11.6 milisaniye için 32 bitlik değerler oluşturularak gerçekleştirilmiştir. Hash değerleri frekans bantları boyunca enerji farkları incelenerek oluşturulmuştur. Ses parmakizi 256 ardışık hash değerine karşı gelmektedir. Veri tabanında arama, alt-parmakizlerine karşı gelen şarkı indekslerinde karşılaştırma işlemi uygulanarak yapılmaktadır. Ses parmakizi çıkarımı ve parmakizlerinin veri tabanında aranması yazılımla gerçekleştirilmiştir. Yapılan testler, gerçekleştirilen uygulamanın zamanda sıkıştırma, radyo yayını, mp3 sıkıştırma, faz kaymaları ve örnekleme sayısının azaltılması gibi etkilere karşı hayli dayanıklı olduğunu göstermiştir.

Farklı parmakizi çıkarım algoritmalarının geliştirilmesi, birden fazla kanaldan alınan yayını dinleyerek gerçek zamanlı program logları tutan bir sistemin gerçekleştirilmesi ilerde yapılması planlanan çalışmalardır.

KAYNAKLAR

- [1] **Haitsma, J.A., Kalker, T. and Oostveen, J.**, 2001. Robust Audio Hashing for Content Identification, Content Based Multimedia Indexing, Brescia, Italy, September, pp. 19-21.
- [2] **Cano, P., Batlle, E., Gomez, E., Gomes, L. de C.T. and Bonnet, M.**, 2002. Audio Fingerprinting: Concepts And Applications, 1st International Conference on Fuzzy Systems and Knowledge Discovery, Singapore, November.
- [3] **Haitsma, J.A.**, 2002. Audio Fingerprinting - A New Technology To Identify Music, M. Sc. Thesis, Technische Universiteit Eindhoven, Eindhoven.
- [4] **Napster**, <http://www.napster.com>
- [5] **Relatable**, <http://www.relatable.com>
- [6] **Kalker, T., Epema, D.H.J., Hartel, P.H., Lagendijk, R.L., Steen, M.v.**, 2004. Music2Share - Copyright-Compliant Music Sharing in P2P Systems, Proceedings of the IEEE, **92**, pp. 961-970.
- [7] **ID3Man**, <http://www.id3man.com>
- [8] **Auditude**, <http://www.auditude.com>
- [9] **Kirovski, D., Attias, H.**, 2002. Beat-ID: Identifying Music via Beat Analysis, IEEE International Workshop on Multimedia and Signal Processing, pp. 190-193.
- [10] **Burges, C.J.C., Platt, J.C., Jana, S.**, 2003. Distortion Discriminant Analysis for Audio Fingerprinting, IEEE Trans. Speech and Audio Processing, **11**, pp. 165-174.
- [11] **Fragoulis, D., Rousopoulos G., Panagopoulos T., Alexiou C. and Papaodysseus C.**, 2001. On the Automated Recognition of Seriously Distorted Musical Recordings, IEEE Transactions on Signal Processing, **49**, pp. 898-908.
- [12] **Logan B.**, 2000. Mel Frequency Cepstral Coefficients for Music Modeling, Proceeding of the International Symposium on Music Information Retrieval (ISMIR), Plymouth, USA, October.
- [13] **Allamanche, E., Herre, J., Hellmuth, O., Fröbach, B.B. and Cremer, M.**, 2001. AudoID: Towards Content-Based Identification of Audio Material, 100th AES Convention, Amsterdam, Netherland, May.
- [14] **Press, W.H., Teukolsky, S.A., Vetterling, W.T., Flannery, B.P.**, 2002. Numerical Recipes in C, Cambridge University Press, New York.

ÖZGEÇMİŞ

5 Eylül 1980’de Denizli’de doğdu. Orta öğretimi Denizli Anadolu Lisesi’nde tamamladıktan sonra 1998’de İstanbul Teknik Üniversitesi Bilgisayar Mühendisliği bölümünde lisans eğitime başladı. 2002’den beri İTÜ Fen Bilimleri Enstitüsü Bilgisayar Mühendisliği Bölümünde yüksek lisans eğitimini sürdürmektedir.

Kasım 2001, Mart 2004 tarihleri arasında TURKCELL Bilgi Teknolojileri Bölümünde yarı-zamanlı olarak çalışmıştır.

Mart-Haziran 2004’te Kanada’da bulunan University of British Columbia üniversitesinde Manufacturing Automation Laboratuvarında ziyaretçi çalışan olarak bulunmuştur.

Ocak 2005’ten bu yana DONE İletişim ve Bilgi Sistemleri kurumunda yazılım geliştirme konusunda çalışmaktadır.

