





**İSTANBUL TEKNİK ÜNİVERSİTESİ ★ FEN BİLİMLERİ ENSTİTÜSÜ**

**SES PARMAKIZI KULLANILARAK REKLAM TANIMA**

**YÜKSEK LİSANS TEZİ**

**Hüseyin ÇABUK**

**Bilgisayar Mühendisliği Anabilim Dalı**

**Bilgisayar Mühendisliği Programı**

**MAYIS 2015**



**SES PARMAKIZI KULLANILARAK REKLAM TANIMA**

**YÜKSEK LİSANS TEZİ**

**Hüseyin ÇABUK  
(504031548)**

**Bilgisayar Mühendisliği Anabilim Dalı**

**Bilgisayar Mühendisliği Programı**

**Tez Danışmanı: Yrd. Doç. Dr. Gökhan İnce**

**MAYIS 2015**



İTÜ, Fen Bilimleri Enstitüsü'nün 504031548 numaralı Yüksek Lisans Öğrencisi **Hüseyin ÇABUK**, ilgili yönetmeliklerin belirlediği gerekli tüm şartları yerine getirdikten sonra hazırladığı “**SES PARMAKIZI KULLANILARAK REKLAM TANIMA**” başlıklı tezini aşağıdaki imzaları olan jüri önünde başarı ile sunmuştur.

**Tez Danışmanı :**      **Yrd. Doç. Dr. Gökhan İnce**      .....  
İstanbul Teknik Üniversitesi

**Jüri Üyeleri :**      **Doç.Dr. Mustafa Ersel Kamaşak**      .....  
İstanbul Teknik Üniversitesi

**Yrd. Doç. Dr. Cenk Demiroğlu**      .....  
Özyeğin Üniversitesi

....      .....  
....

**Teslim Tarihi :**      **04 Mayıs 2015**  
**Savunma Tarihi :**      **26 Mayıs 2015**





*Eşime ve aileme,*



## ÖNSÖZ

Tez çalışmam süresince yanında çalışmaktan onur duyduğum, bilgisini, desteğini, sabrını ve hoşgörüsünü esirgemeyen değerli hocam Yrd. Doç. Dr. Gökhan İnce'ye teşekkürü bir borç bilirim. Kendisinin danışmanlığı ile çalışma fırsatı bulmuş olmam benim için büyük bir şanstı.

Bu zorlu süreçte bana daima destek olan ve sabır gösteren eşime, ayrıca bana emek verip bugünlere getiren aileme de teşekkürlerimi sunarım.

Mayıs 2015

Hüseyin ÇABUK  
(Bilgisayar Mühendisi)



## İÇİNDEKİLER

### Sayfa

ÖNSÖZ .....	vii
İÇİNDEKİLER .....	ix
KISALTMALAR.....	xi
ÇİZELGE LİSTESİ.....	xiii
ŞEKİL LİSTESİ.....	xv
ÖZET .....	xvii
SUMMARY .....	xix
<b>1. GİRİŞ.....</b>	<b>1</b>
1.1 Tezin Amacı.....	1
1.2 Literatür Araştırması .....	2
1.2.1 Ses Parmakizi Çalışmaları.....	3
1.2.2 Reklam Tanıma Çalışmaları .....	6
1.2.3 Tezin Organizasyonu .....	7
<b>2. İÇERİK TABANLI SES TANIMA SİSTEMLERİ.....</b>	<b>9</b>
2.1 Genel Yapı .....	9
2.1.1 Ön işleme.....	10
2.1.2 Öznitelik Çıkarma .....	10
2.1.3 Parmakizi Modelleme.....	11
2.1.4 Veritabanında Arama ve Mesafe Hesabı .....	11
2.1.5 Hipotez Testi .....	12
2.2 Ses Parmakizi Çıkarma İçin Kullanılan Ses Öznitelikleri .....	12
2.2.1 Yüksek Düzey Öznitelikler (High-Level Features) .....	12
2.2.2 Alçak Düzey Öznitelikler (Low-Level Features) .....	13
2.3 Ses parmakizi ile tanıma sistemlerinin performans parametreleri .....	14
2.3.1 Gürbüzlük (Robustness) .....	14
2.3.2 Güvenilirlik (Reliability) .....	14
2.3.3 Parmakizi boyutu (Fingerprint size).....	14
2.3.4 Parçalılık (Granularity).....	15
2.3.5 Arama süresi ve ölçeklenebilirlik (Search speed and scalability) .....	15
2.4 Ses parmakizi uygulama alanları.....	15
2.4.1 Ses içeriği gözlemlemek.....	15
2.4.1.1 Dağıtımçı (Distributor) tarafında gözlemlemek.....	15
2.4.1.2 İletim kanalları tarafında gözlemlemek .....	16
2.4.1.3 Kullanıcı tarafında gözlemlemek.....	16
2.4.2 Katma değerli servisler.....	16

2.4.3 Bütünlük kontrolü sistemleri .....	16
2.4.4 Tekrarlı kayıtların bulunması.....	17
<b>3. ÖNERİLEN YÖNTEM.....</b>	<b>19</b>
3.1 Temel prensipler .....	19
3.2 Takımyıldızı Haritaları (Constallation map).....	20
3.3 Kombinasyonel Karmalama (Combinatorial Hashing) .....	21
3.4 Arama ve Skorlama .....	23
3.4.1 Arama .....	24
3.4.2 Alternatif Karma Değerleriyle Arama.....	25
3.4.3 Histogram Normalizasyonu.....	25
3.4.4 Skorlama.....	26
<b>4. DENEYLER VE SONUÇLAR.....</b>	<b>29</b>
4.1 Yazılım ve Donanım Özellikleri .....	29
4.2 Deney Şartları .....	29
4.3 Değerlendirme Kriterleri .....	30
4.4 Ön Deneyler (Laboratuar Deneyleri ve Sonuçları).....	31
4.4.1 Yapısal İyileştirmelerin Etkisi .....	31
4.4.1.1 Alternatif Karma Değerlerinin Aramada Kullanılması .....	32
4.4.1.2 En Güçlü Zirvelerin Kullanılması.....	33
4.4.1.3 Histogram Normalizasyonu .....	34
4.4.1.4 Güvenilirlik Kontrolü.....	34
4.4.2 Sorgu süresinin etkisi.....	35
4.4.3 Sistem parametrelerin etkisi .....	36
4.5 Deney Sonuçları .....	36
4.5.1 Şarkı Kümesi Deney Sonuçları.....	36
4.5.2 Reklam Kümesi Deney Sonuçları.....	37
4.6 Tartışma .....	37
<b>5. SONUÇ VE ÖNERİLER .....</b>	<b>41</b>
<b>KAYNAKLAR.....</b>	<b>45</b>
<b>ÖZGEÇMİŞ .....</b>	<b>49</b>

## **KISALTMALAR**

<b>SNR</b>	: Signal To Noise Ratio
<b>PCM</b>	: Pulse-code modulation
<b>TP</b>	: True Positive
<b>TN</b>	: True Negative
<b>FP</b>	: False Positive
<b>FN</b>	: False Negative
<b>Hz</b>	: Hertz
<b>FFT</b>	: Fast Fourier Transform
<b>ms</b>	: Milisaniye
<b>sn</b>	: Saniye
<b>GHz</b>	: GigaHertz
<b>ADT</b>	: Audio Degradation Toolbox
<b>dB</b>	: Decibel
<b>MFCC</b>	: Mel Frequency Cepstral Coefficient
<b>BFCC</b>	: Bark-Frequency Cepstral Coefficients
<b>AdaBoost</b>	: Adaptive Boosting
<b>DDA</b>	: Distortion Discriminant Analysis
<b>OPCA</b>	: Oriented Principle Component Analysis
<b>CD</b>	: Compact Disc
<b>DVD</b>	: Digital Versatile Disc
<b>TV</b>	: Television
<b>AVM</b>	: Alış Veriş Merkezi





## ÇİZELGE LİSTESİ

	<u>Sayfa</u>
<b>Çizelge 4.1:</b> Deney Kümesi Çeşitleri .....	30
<b>Çizelge 4.2:</b> Alternatif karma değerleriyle aramanın başarıma etkisi .....	32
<b>Çizelge 4.3:</b> Alternatif karma değerleriyle aramanın arama süresine etkisi.....	32
<b>Çizelge 4.4:</b> En güçlü zirvelerin kullanılmasının etkisi.....	33
<b>Çizelge 4.5:</b> Histogram normalizasyonunun etkisi .....	34
<b>Çizelge 4.6:</b> Güvenilirlik kontrolünün etkisi .....	35
<b>Çizelge 4.7:</b> Sorgu süresinin etkisi .....	36
<b>Çizelge 4.8:</b> $\alpha$ , $\beta$ ve $\sigma$ parametrelerinin belirlenmesi için yapılan deney sonuçları	37
<b>Çizelge 4.9:</b> Şarkılar veritabanındaki anma, kesinlik, sonuç oluşturma süresi ve karma sayısı sonuçları.....	38
<b>Çizelge 4.10</b> Reklamlar veritabanındaki anma, kesinlik, sonuç oluşturma süresi ve karma sayısı sonuçları. ....	39



## ŞEKİL LİSTESİ

	<u>Sayfa</u>
Şekil 2.1 : İçerik-bazlı Ses Tanıma Sistemi .....	9
Şekil 3.1 : Spektrogram.....	20
Şekil 3.2 : Zirve Noktaları .....	20
Şekil 3.3 : Kombinasyonel karmalama .....	21
Şekil 3.4 : Karma üretimi.....	21
Şekil 3.5 : Veri yapısı .....	22
Şekil 3.6 : $\beta$ ve $\sigma$ parametrelerine göre başarımlar oranları.....	23
Şekil 3.7 : Hash sayısı oranı.....	23
Şekil 3.8 : Histogram .....	25
Şekil 3.9 : $\alpha$ ve $\beta$ parametrelerine göre başarımlar oranları.....	27



## SES PARMAKIZI KULLANILARAK REKLAM TANIMA

### ÖZET

Akıllı telefonların kullanımı her geçen gün daha da fazla artmaktadır. Artık insanlar günlük hayatlarında birçok faaliyetle eş zamanlı olarak akıllı telefon kullanmaktadırlar. Akıllı telefonun eş zamanlı olarak en çok kullanıldığı faaliyetlerden biri de televizyon izlemektir. Akıllı telefon kullanıcılarının %84'ü televizyon izlerken eşzamanlı olarak akıllı telefonlarını kullanmaktadırlar.

Televizyon izlerken akıllı telefon kullanımına yönlendiren sebeplerden biri de televizyonda izlenen konu ile ilgili daha detaylı bilgiye ulaşmak için telefonda arama yapmaktır. Metinsel olarak arama yapmak zahmetli ve uzun bir işlemdir. Özellikle reklam sektörünün, tanıttıkları ürünle ilgili kullanıcıların daha kolay ve hızlı şekilde bilgiye ulaşması için başka çözümlere ihtiyacı vardır.

Çalışmamızda, televizyonda oynayan reklamın, akıllı telefonda bir uygulamaya dinletilip tanınmasını sağlayacak bir algoritmanın geliştirilmesi amaçlanmıştır. Reklam tanınması için ses parmakizi (audio fingerprinting) yöntemlerinin kullanılmasına karar verilmiştir. Geliştirilecek yöntemin, ses tanıma sistemlerinin sahip olması gereken gürbüzlük, güvenilirlik, veri boyutu küçüklüğü, parçalılık, arama süresi kısıtlılığı ve hesaplama maliyeti küçüklüğü özelliklerine sahip olması hedeflenmiştir. Farklı reklamlarda, şarkılardan farklı olarak, aynı müzik veya konuşma bölümleri geçebilmektedir, bu nedenle güvenilirlik kontrolü aşamasında, müzik tanıma sistemlerinden farklı bir yaklaşım geliştirilmesi amaçlanmıştır.

Ses parmakizi olarak, spektrogramdaki zirve noktalarının aralarındaki zaman ve frekans farklarından yararlanılarak üretilen karmalar kullanılmıştır. Gürültü ve sinyal bozulmaları durumlarında bile spektrogramdaki zirve noktaların en azından bir kısmının korunması beklenmektedir.

İlgili yöntemin prototip geliştirmesi yapıldıktan sonra deneylerde bazı zayıf yönleri tespit edilmiştir. Genelde, spektrogramdaki zirve noktaların zaman veya frekans yönünde küçük kaymaları nedeniyle ortaya çıkan bu problemlere çözümler geliştirilmiştir. Yöntemin aynı başarı oranlarına daha az veri kullanarak ulaşabilmesi için de katkılar sunulmuştur. Güvenilirlik kontrolü aşamasında da iki eşik değeri parametrelili bir çözüm üretilmiştir.

Yapısal katkılar tamamlandıktan sonra, yapılan birçok deney ile yöntemin en iyi başarı oranlarını verdiği sistem parametreleri belirlenmiştir.

Deneylerde şarkılar ve reklamlardan oluşan 2 deney kümesinin, çeşitli seviyelerde beyaz gürültülü, pembe gürültülü, kahverengi gürültülü, kırılma uygulanmış, bar ortamı etkisi uygulanmış, canlı kayıt etkisi uygulanmış, akıllı telefon kayıt etkisi uygulanmış, akıllı telefon çalma etkisi uygulanmış alt-deney kümeleriyle, ayrıca bir

de İstanbul'daki bir alış veriş merkezinde akıllı telefon ile kaydedilmiş versiyonları kullanılmıştır.

Deney sonuçları, anma, kesinlik, kullanılan veri boyutu ve tanıma süresi açılarından değerlendirilmiş, baz alınan yöntemle karşılaştırılmıştır. Bahsedilen bütün kriterlerde baz alınan yöntemle göre daha iyi sonuçlar elde edildiği tespit edilmiştir.

# COMMERCIAL IDENTIFICATION USING AUDIO FINGERPRINTING

## SUMMARY

Smart phone usage is rapidly increasing day by day. Nowadays, in their daily lives, people often use their smart phones simultaneously with another activity. One of the most often conducted activities that goes with smart phone usage is watching television. 84% of smart phone owners, use their smart phones while watching television.

One of the main factors that leads people to use their smart phones while watching television is to search for more detailed information about the topics they are watching on TV. Text-based searching is inconvenient and also time consuming. Especially advertising industry needs some other resolution to let their customers find information about their products more quickly and easily.

In this thesis, it is aimed to develop an algorithm that can be used in a smart phone application to identify a TV commercial by listening through microphone. It is decided to use audio fingerprinting techniques for commercial identification. The algorithm is aimed to have the properties that most of the audio fingerprinting systems should have, such as robustness, reliability, granularity, fingerprint size, search speed and scalability. Some commercials contain same part of music or speech, so another type of reliability check method is needed to be developed, apart from the ones used in song identification systems. Audio fingerprinting literature is reviewed and one algorithm is chosen to be used as base, to satisfy the said requirements.

According to the algorithm, temporal and frequencial distances between the peaks in audio spectrogram were used to generate the fingerprint. A time-frequency point is a candidate peak if it has a higher energy content than all its neighbors in a region centered around the point. At least some of the peaks are supposed to survive in presence of noise or signal distortions.

Base algorithm was developed as a prototype and some weaknesses were identified in the initial tests. It was seen that the problems are often caused by small shifts of peaks in time or frequency directions. Especially in situations with high level of noise, it was observed that, those shifts could be much more and cause success rates to decrease excessively. Some solutions were presented for the mentioned problems.

Firstly, in the database search step, hashes, that were generated using the distances between the peaks, were searched in the fingerprint database with some alternative values to increase the success rates. Although increase in success rates were observed in the tests, database search duration was also increased as expected. As search duration was still under real time, it was evaluated as applicable.

In the initial test, it was observed that using all the peaks in search step was decreasing the success rates and also increasing the search duration and computation cost, because most of the peaks couldn't survive in high levels of noise. Therefore another contribution was made to use only the strongest  $n$  peaks in search step. Value of  $n$  was decided to be calculated for every second of the query using the equation 3.1. This contribution helped achieve the same success rates using smaller data sizes.

It was observed that shifts of peaks caused some problems also in scoring step. According to this, right answer had smaller score than it should have, so it took longer time to exceed the threshold. To overcome this problem, another contribution, that is called histogram normalization, was presented. This contribution led higher success rates.

Another weakness of the method was identified in the tests with commercial test sets. Since more than one commercial could have same music or speech parts, querying with this parts resulted in false positives with the original hypothesis testing method. Hypothesis testing step was also improved with a method that uses two threshold parameters, which are called matching rate and power rate. Test results showed that false positives rate decreased excessively after this contribution.

After the structural contributions were completed, a number of tests were employed to find the optimal values for the 3 system parameters,  $\alpha$  (alpha),  $\beta$  (beta) and  $\sigma$  (sigma), to achieve the best success rates.

Two sets, consisting of songs and commercials, were used as test sets. A number of degradations including white noise, pink noise, brown noise, clipping, bar environment effect, smart phone recording effect, smart phone playback effect and live recording effect were applied to test sets. Furthermore, another test set was generated by recording with a smart phone in a real shopping mall in İstanbul.

Test results were examined in terms of recall, precision, required data size and search time, and then compared with the base algorithm. These result values also refer to the required performance parameters of any audio fingerprinting system, which are robustness, reliability, fingerprint size, granularity and search time. It has been observed that the developed system is more advantageous than the base algorithm. In another words, search time and storage need were decreased and recall and precision were increased with the contributions.

For now, developed system is running as a standalone windows application, written in C++ programming language. When a song or commercial, that were introduced to the system beforehand, is played, testing is started by clicking the button on the user interface and then application starts listening from the microphone and tries to find a match in the database.

For now, the fingerprints of songs and commercials, that are introduced to the system beforehand, are stored in a text file. Before testing step, another button is used to start reading this fingerprints from text file into system memory. Since memory is limited, the number of song and commercial fingerprints is limited too. In tests, a maximum of 500 songs and commercials were used. As the system to be usable in real life applications, it is needed to make some developments to use a real database as fingerprint storage. The database structure needs to be indexable to prevent search



times from increasing as the fingerprint count goes up, otherwise the system will be unusable.

As the system to be usable in real life applications, a client-server architecture must be developed, with a smart phone application as client, and fingerprint database as server. Smart phone application should listen from the microphone, generate the fingerprint, and send the fingerprint to server instead of all the song data. Server should search the fingerprint in database, make scoring and hypothesis testing and response the client appropriately.

Audio fingerprint generation speed with the smart phone application, should be observed and if it is above the real time, some optimizations should be done to speed up the process.

Searching and scoring processes in server side, are convenient for parallel processing. If the server is developed in parallel processing architecture, searching and scoring times can be reduced excessively.



## 1. GİRİŞ

Google tarafından yapılan bir çalışmaya göre [1], günümüzde medya erişiminin çok büyük bir kısmı (%90), akıllı telefonlar, tabletler, dizüstü bilgisayarlar ve televizyonlar gibi ekranlı cihazlar aracılığıyla; kalan kısım da (%10) gazete, dergi ve radyo aracılığıyla sağlanmaktadır. Ekranlı cihazlar içinde en büyük pay ise %38 ile akıllı telefonlardadır.

Kullanıcıların cihaz kullanımı arttıkça çoklu cihaz kullanımı da artmaktadır. Cihaz kullanıcılarının %98'si aynı gün içinde birden fazla cihaz kullanmaktadır. Eş zamanlı cihaz kullanımı da çok yaygındır. Televizyon izleyicilerinin büyük kısmı aynı anda ikinci bir cihaz daha kullanmaktadır [1] [2]. Kullanıcılar sıklıkla televizyonda izledikleri konuyla ilgili ikinci bir ekrandan arama yaparlar. En yaygın eş zamanlı kullanılan cihaz akıllı telefonlardır. [1]

Reklam sektörü, kullanıcıların bu eğilimini göz önünde bulundurup uygun stratejiler geliştirmelidir. Televizyon reklamları kısa sürelidir ve detaylı bilgi içermez. Bu yüzden izleyicinin dikkatini çektikten sonra, daha detaylı bilgi ve tanıtım fırsatı için, ikinci ekrana yönlendirmek bir avantaj sağlayabilir. Yalnız, bunun için kullanıcıdan ikinci bir ekrandan arama yapmasını istemek yeterince ikna edici değildir. İzleyiciyi sıkmadan, kolayca ve hızlı bir şekilde dikkatini ikinci bir ekrana aktarmak gerekmektedir.

Bunun yerine, kullanıcının ilgili reklam çıktığında akıllı telefonundaki bir uygulamayı başlatması, bu uygulamanın da telefonun mikrofonundan reklamı dinleyip, hangi reklam olduğunu anlaması ve sonrasında istenen kurguya göre akışın devam ettirilmesi daha iyi bir çözüm olacaktır.

### 1.1 Tezin Amacı

Bu çalışmada, reklam tanınması için kullanılabilecek bir yöntem tasarlanması, prototip bir uygulama geliştirilmesi ve başarı oranlarının incelenmesi amaçlanmıştır. Reklam

tanınması için ses parmakizi (audio fingerprinting) yöntemlerinin kullanılmasına karar verilmiştir. Geliştirilecek sistemin, ses tanıma sistemlerinden beklenen gürbüzlük, güvenilirlik, veri boyutu küçüklüğü, parçalılık, arama süresi ve hesaplama maliyeti performans parametrelerini en iyi şekilde sağlaması hedeflenmiştir. Çalışmamızda reklam tanınması amaçlandığı için, şarkılardan farklı olarak, birden çok reklamda aynı müzik veya konuşma bölümleri geçebilmektedir, bu nedenle güvenilirlik kontrolü ile ilgili, müzik tanıma sistemlerinden farklı bir yaklaşım geliştirilmesi amaçlanmıştır.

Öncelikle ses parmakizi çalışmaları incelenmiş, uygun yöntem belirlenmiştir. Belirlenen yöntemin prototip geliştirmesi yapılmış ve üretilen çok çeşitli deney kümeleriyle tekrar tekrar testler yapıp, yöntemin anma (recall), kesinlik (precision), tanıma süresi ve veri boyutu açısından sonuçları incelenmiş ve eksikleri belirlenip bu eksikleri giderecek şekilde yöntem katkıları sağlanmıştır. Son aşamada tekrar birçok test yapılarak en iyi sonuç üretecek yöntem parametre değerleri tespit edilmiştir.

## **1.2 Literatür Araştırması**

Haitisma ve arkadaşları tarafından yapılan çalışmada [3] ses parmakizi şu şekilde tanımlanmıştır: Ses parmakizi, bir ses dosyasının kısa bir özetidir. Bu nedenle, bir ses parmakizi fonksiyonu  $F$ , büyük bir ses nesnesini, sınırlı boyuttaki bir parmakizine dönüştürür. Bu noktada kriptografideki karma (hash) fonksiyonlarıyla benzerlik kurabiliriz. Kriptografik karma fonksiyonu büyük bir nesneyi, küçük bir karma değerine dönüştürür. Bir kriptografik karma fonksiyonu iki büyük objenin eşit olup olmadığını anlamak için, sadece karma değerlerinin eşit olup olmadığını bakmamızı sağlar. Karma değerlerinin matematiksel olarak eşit olması, çok küçük bir hata ihtimali ile, orjinal objelerin de eşit olduğu anlamına gelir. Kriptografik karma fonksiyonları ile, bir nesnenin, verilen bir nesne kümesi içinde olup olmadığını anlamak kolaylaşır. Bütün veriyi saklamak ve karşılaştırmak yerine sadece karma değerleri saklanıp karşılaştırılarak o veri objesinin kümede olup olmadığı anlaşılabilir.

Bu açıklamaya göre, kriptografik karma fonksiyonlarının, ses parmakizi fonksiyonu olarak kullanılabilmesi anlaşılabilir. Fakat bu doğru değildir, çünkü parmakizi fonksiyonu, karma fonksiyonlarının sağladığı matematiksel eşitlik değil, algısal eşitlik sağlamalıdır. Mesela bir şarkı farklı sıkıştırma teknikleriyle saklanabilir ve her

biri, karma fonksiyonundan farklı karma deęerleri üretir. Ama aynı şarkıyı farklı sıkıştırma teknikleriyle kaydedilmiş farklı dosyalardan çalıp dinlediğimizde bile aynı şarkı olduğunu anlarız. Parmakizi yöntemleri, dinlediğimizde algıladığımız şekilde bir karşılaştırma yapabilmemizi sağlamalıdır.

Çalışmamız televizyon reklamlarının ses parmakizi yöntemleriyle tanınması olduğu için, literatür araştırması iki bölümde ele alınmıştır, ilk bölümde ses parmakizi ile müzik tanıma çalışmalarından, daha sonra da reklam tanıma çalışmalarından bahsedilmiştir.

### **1.2.1 Ses Parmakizi Çalışmaları**

Çalışmamızda reklam tanınması amaçlandığı için, sadece müzik için kullanılabilir yüksek düzey ses öznitelikleri (Melodi, ses reng, ritim, harmoni, perde) tercih edilmemiştir. Çünkü reklamlarda müzik içeren bölümler kadar konuşma veya sessiz bölümler de olabilmektedir. Bu nedenle yüksek düzey ses öznitelikleri kullanan çalışmaların detaylarına girilmemiştir. Bu çalışmalara örnek olarak [4] ve [5] çalışmaları gösterilebilir.

Aşağıda alçak düzey ses öznitelikleri kullanan çalışmalara biraz daha yakından bakılmıştır :

En çok kullanılan sistemlerden biri [3] örtüşen pencereler (overlapping windows) kullanmaktadır. Spektral gösterim için 300-2000 Hz arasındaki 33 Bark-frequency cepstral coefficients (BFCC) bantı kullanmaktadır ve her 11,6 ms'de, 370 ms'lik bir dilim ses verisi almaktadır. Bu büyük örtüşme alt-parmakizlerinin zaman içinde hızla değil, yavaş yavaş değişmesini, böylece parmakizinin küçük kaymalara karşı değişmez olmasını sağlamaktadır. Her spektral dilim için, sonraki ve önceki frekans bandına ve frekans dilimlerine olan farkına göre alt-parmakizi denilen, 32 bit'lik bir dizi üretilmektedir. Bu alt-parmakizleri ses sinyalindeki küçük değişimlere karşı hassas değildir, çünkü gerçek fark deęerleri değil, sadece aradaki farka göre azalan veya artan olduğunu gösteren deęerler tutulmaktadır. Bu parmakizi yönteminde karşılaştırma kolay olmaktadır, çünkü basit bir Hamming Uzaklığı (hamming distance) kullanılabilir.

Bu çalışmada ortalama 5 dk'lık şarkıların bulunduğu 10.000 şarkılık bir database için yaklaşık 250.000.000 alt-parmakizi üretilmektedir. Arama aşamasında bütün bu veritabanının incelenmesi mümkün olamayacağı için, bunun yerine Haitzma ve arkadaşları, olası sinyal bozulmalarında bile, her sorgu şarkısının alt-parmakizlerinden en az bir tanesinin veritabanıyla doğru bir şekilde ve tamamen eşleşeceğini kabul etmişlerdir. Bu da onlara tamamen eşleşen alt-parmakizlerini bulmak için HashTable kullanabilme imkanı vermektedir. Bir tam eşleşme bulunduğu zaman, eşleşen şarkının zamana göre sıralı sonraki alt-parmakizleri incelenerek eşleşmenin doğru olup olmadığından emin olunmaktadır. Bu da bütün sorgu uzunluğunda basit Hamming Uzaklığı (Hamming Distance) hesabının yapılabilmesini sağlamaktadır. Yapılan aramanın tamamı, bulunan tam eşleşmelerle ilgili yapılmaktadır. Eğer tek bir tam eşleşmenin bile olamayacağı, çok büyük oranda sinyal bozulmaları bekleniyorsa, yöntemde bir değişiklik yapıp sorgudan, küçük Hamming Uzaklığı mesafedeki eşleşmeler de aranmaktadır. Bu çalışma Philips Research tarafından yapılmıştır ve 2005 yılında Gracenote [6] ve Philips arasında yapılan anlaşmayla, Gracenote'un MusicID sisteminin çalışma algoritması olarak kullanılmıştır. [7]

Bir diğer çalışmada [8] ses tanınması için "bilgisayarla görme" (computer vision) tekniklerinin kullanılabileceği gösterilmiştir. İlk yöntemdeki [3] parametrelerin aynısı kullanılarak ses dalgası sinyali, zaman-frekans olarak iki boyutlu gösterime dönüştürülmektedir ve aynı şekilde 33 Bark-frequency cepstral coefficients (BFCC) bantı kullanılmaktadır. [3] çalışmasından farkı ise, 32 bitlik alt parmak izinin üretilmesi için kullanılan filtrelerin, resim işleme çalışmalarında sıkça kullanılan eğitime yöntemleriyle seçilmiş olmasıdır. Şarkıların orjinalleri ve belirli bozulmalara uğratılmış halleri kullanılarak oluşturulan eğitim setiyle, AdaBoost eğitim yöntemleri kullanılarak, 25000 aday filtre arasından, en ayırd edici ve bozulmalara karşı dayanıklı 32 filtre belirlenmiş ve 32 bitlik alt-parmakizleri bu filtreler kullanılarak oluşturulmuştur. Böylece ilk yönteme göre daha ayırd edici ve dayanıklı parmakizleri elde edildiği ve daha iyi sonuçlar alındığı belirtilmiştir.. MusicBrainz online şarkı veritabanı [9], şarkı tanınması için bu yöntemden esinlenilerek geliştirilmiş ChromaPrint algoritmasını kullanmaktadır. [10]

Bir diğerk çalıřmada da [11], [8] çalıřmasında olduđu gibi resim iřleme yöntemlerinden yararlanılmıřtır. Önce spektrogram oluřturulup, sonra spektrogram resimlerine Haar dalgacıklar (Haar wavelet) uygulanmaktadır. Bu iřlem sonucunda, spektrogram resmindeki piksel sayısı kadar dalgacık katsayısı (co-efficient) ortaya çıkmaktadır. Yazarlar, sadece en güçlü t dalgacıđı kullanmanın yeterli olacađına karar vermiřtirler. Bunun yanında, en güçlü t dalgacıđının da güçlerinin bilinmesine gerek olmadıđı, sadece artı/eksi olarak iřaretinin bilinmesinin yeterli olacađı sonucuna varılmıř ve bu şekilde iřaretlerin artı veya eksi oluřuna göre parmakizi oluřturulmuřtur.

Bir diğerk çalıřmada [12] spektrogram oluřturulmuř ve spektrogramdan yararlanılarak notaların bařlangıç noktaları (onset) bulunmuřtur. Bu noktaların arasındaki zaman farkları kullanılarak parmakizi üretilmiřtir. Spektrogram 8 banda bölünmüř, her bant için ayrı ayrı nota bařlangıçları ve parmakizleri hesaplanmıřtır. Genel olarak saniyede 1 bařlangıç noktası olacak şekilde parametreler ayarlanmıřtır. Her bařlangıç noktasının, kendisinden sonra gelen 6 bařlangıç noktasıyla arasındaki süre farkı kullanılmıřtır. Sonuç olarak saniyede "8 bant x 1 bařlangıç noktası x her bařlangıç noktası için 6 mesafe" olacak şekilde toplam 48 karma üretilmiřtir. The Echo Nest [13] online müzik servisleri platformunun řarkı tanıma algoritması [12] çalıřmasındaki yöntemi kullanmaktadır.

Microsoft Research'de yapılan bir çalıřmada [14] algısal olarak ađırlařtırılmıř log spektrogramı kullanılmıřtır. Bu spektrogramdan gürültüye karřı toleranslı parmakizleri üretmek için Distortion Discriminant Analysis (DDA) yöntemi kullanılmıřtır. Bu çalıřmadaki parmakizleri önceki çalıřmalara göre daha karmařık ama aynı zamanda her parmakizi daha uzun süreli ses parçalarını özetlemektedir. DDA, Linear Discriminant Analysis'in bir varyantı olan Oriented Principle Component Analysis (OPCA)'e dayanmaktadır. OPCA, ses parçalarının bozulmuř hallerinin de, eğitim ařamasında mevcut olduđunu kabul etmektedir. OPCA, Signal To Noise Ratio (SNR)'yı maksimize edecek şekilde parametreler üretmektedir. Son durumda sistem, 110.000 girdiyi, 64 çıktıya dönüřtürmektedir. Bu 64 çıktı, aramada kullanılacak olan alt parmakizleridir.

Bir bařka çalıřmada [15] yine spektrogram üretilmiřtir ve bu defa spektrogramdaki zirve noktaları (peak) bulunup parmakizi üretiminde bu noktalar kullanılmıřtır. Sesin

bozulması durumunda zirve noktalarının korunması ihtimali yüksek olduğu için bu çalışmada zirve noktaları tercih edilmiştir. Bulunan zirve noktalarının frekans değerleri ve zaman farkları kullanılarak parmakizi oluşturulmuştur. En çok bilinen şarkı tanıma servislerinden biri olan Shazam [16] bu çalışmadaki algoritmayı kullanmaktadır.

Biz de çalışmamızda bu yöntemi uyguladığımız için, 3. bölümde çalışmanın detayları anlatılacaktır.

### **1.2.2 Reklam Tanıma Çalışmaları**

Televizyon reklamlarıyla ilgili çalışmalar iki grupta toplanabilir, reklam algılama (Commercial detection) ve reklam tanıma (Commercial identification). Reklam algılama, televizyondaki yayın akışı içindeki reklamların başlangıç ve bitişlerini bulmayı amaçlamaktadır. Reklam tanıma ise yayın akışındaki reklamların, bilinen reklam veritabanı içindeki hangi reklam olduğunu amaçlamaktadır. Bizim çalışmamız reklam tanıma içerisinde yer almaktadır. Reklam tanıma çalışmalarını da kullandıkları özniteliklere göre üçe ayırabiliriz : Görsel öznitelik kullananlar, ses öznitelikleri kullananlar ve ikisini birlikte kullananlar.

Görsel öznitelik kullanılarak üretilen parmakizi boyutları büyük ve hesaplama olarak da zor olduğu için gerçek zamanlı çalışacak sistemlerde uygulanabilirliği azdır. Biz de çalışmamızda ses tanıma tekniklerinden yararlandık.

Ses parmakizi yöntemlerini kullanan bir çalışmada [17] spektral merkez öznitelikleri kullanılarak parmakizi üretilmiştir. Başka bir çalışma [18] enerji paketlerini (energy envelope) kullanarak ses dalgasını segmentlere ayırmaktadır ve bu segmentler arasındaki mesafeleri kullanarak parmakizi üretmektedir. Bir diğer çalışmada [19] spektrogramdaki 16 alt bant için merkezler hesaplanıp kullanılmıştır. [20] çalışmasında hem ses hem de görsel parmakizi kullanılmıştır. Ses parmakizi olarak spektrogramdaki alt bantların, komşu zamanlardaki değişimleri kullanılmıştır. Video parmakizi olarak da Harris tanımlayıcıları (descriptor) ve bitişik video çerçevelerindeki değişimlerin bir kombinasyonu kullanılmıştır. Bir diğer çalışmada [21] yine görsel ve ses parmakizleri beraber kullanılmıştır. Her ikisi için ayrı ayrı parmakizi üretilip ayrı ayrı



arama yapılmaktadır ve ikisinin birden sonuçları beraber değerlendirilip tek sonuca varılmaktadır. Ses parmakizi olarak [3] çalışmasındaki yöntem kullanılmıştır.

### **1.2.3 Tezin Organizasyonu**

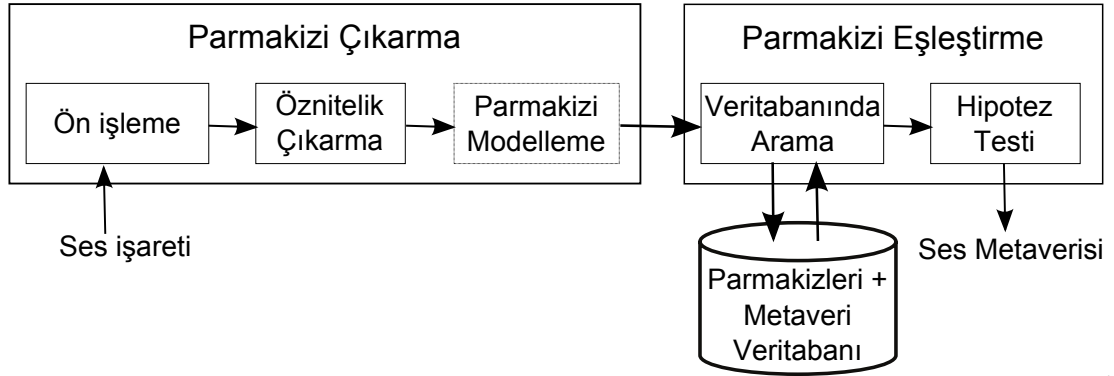
Bölüm 2’de içerik tabanlı ses tanıma sistemlerinin genel yapısı anlatılmış, bu sistemlerde genelde izlenen yöntemlerden, kullanılan örnek ses özniteliklerinden ve ses tanıma uygulama alanlarından bahsedilmiştir. Bölüm 3’de önerilen yöntemin temel prensiplerine, detaylı işleyişine ve sunulan katkılara yer verilmiştir. Bölüm 4’de deney sonuçları verilmiş ve tartışılmıştır. Bölüm 5’de ise tezle ilgili sonuçlar, öneriler ve ileride yapılabilecek çalışmalardan bahsedilmiştir.



## 2. İÇERİK TABANLI SES TANIMA SİSTEMLERİ

### 2.1 Genel Yapı

İçerik tabanlı ses tanıma sistemlerinde kullanılabilecek bir çok yöntem olsa da, genelde yöntemlerin bazı ortak yanları vardır. İçerik tabanlı ses tanıma sistemlerinin genel akışı Şekil 2.1’de gösterildiği şekilde, bir tane "Parmakizi Çıkarma" ve bir tane de "Parmakizi Eşleştirme" algoritması modüllerinden oluşmaktadır.



Şekil 2.1: İçerik-bazlı Ses Tanıma Sistemi

Parmakizi Çıkarma modülü, ses kaydının algısal karakteristiklerini, sağlam ve kısa bir forma (parmakizi) dönüştürür. Bu modül, kendi içinde üç alt-modülden oluşur : Ön işleme, öznitelik çıkarma ve parmakizi modelleme modülleri. Ön işleme modülü sinyale, analogdan sayısala dönüştürme, tek kanala düşürme, örnekleme oranını değiştirme gibi bazı işlemler uygulayarak, sinyalin işlenmeye hazır biçime dönüştürülmesini sağlar. Öznitelik çıkarma modülü, sinyalle ilgili önceden belirlenmiş, ayırt edici olan bazı değerleri ölçümler. Parmakizi modelleme bölümü de bu ölçüm değerlerinden son parmakizi formunu ortaya çıkarır.

Bir parmakizi verildiğinde, parmakizi eşleştirme modülü, bu parmakizini, veritabanında tanımlı olan diğer parmakizleriyle karşılaştırır ve en iyi eşleşmeyi bulur. Bu nedenle, bir şekilde parmakizlerinin mesafelerini (distance) hesaplayan bir algoritma

gereklidir. Veritabanındaki ses kaydı sayısı fazla ve mesafe hesabı da zor bir iş olduğu için, bu hesabı hızlı yapabilecek yöntemlere ihtiyacımız vardır.

Son modül olan hipotez testi modülü de, bir güvenilirlik ölçüsü hesaplar ve sistemin bulduğu tanıma sonucunun ne kadar güvenilir olduğuna karar verir.

### **2.1.1 Ön işleme**

Ön işleme kısmında; gerekliyse öncelikle ses dijital hale dönüştürülür ve ortak, belirlenmiş bir formata getirilir. Mesela 16 bit PCM, tek kanal ve sabit bir örnekleme frekansına dönüştürülür.

Bir sinyal birkaç milisaniye süresince sabit, değişmez kabul edilebilir. Bu nedenle sinyal çerçeve (frame) denilen bu parçalara bölünür. Bir saniyede kullanılan çerçeve sayısına çerçeve frekansı (frame rate) denir. Çerçevelerin başındaki ve sonundaki devamsızlıkları minimize etmek için bazı pencere fonksiyonları kullanılır. Çerçeveler zaman olarak bir miktar üst üste gelecek şekilde ayarlanır ki, sesteki kaymalara karşı dayanıklı olsun.

Daha sonra sinyal, zaman domeninden, frekans domenine dönüştürülür. Frekans domeninde birçok işlem daha kolay yapılabilmektedir. En sık kullanılan dönüştürme yöntemi Fast Fourier Transform yöntemidir.

### **2.1.2 Öznitelik Çıkarma**

Frekans domenine dönüştürülmüş sinyalin önceden belirlenmiş, ayırt edici olan birtakım değerlerinin ölçümleri yapılır. Bu aşamada çok çeşitli yöntemler kullanılabilir. Buradaki amaç boyutluluğu azaltmak ve sesteki bozulmalara karşı dayanıklılığı arttırmaktır. Algısal olarak daha anlamlı öznitelikler çıkarabilmek için insan duyma sistemiyle ilgili bilgilerin göz önünde bulundurulması çokça rastlanır bir durumdur.

Ölçülen bu öznitelik değerleri hataya açık değerlerdir. Bu değerler direk parmakizi olarak kullanılırsa çok kırılğan bir parmakizi elde edilir. Bunun yerine bu değerlerin zamana göre değişimi veya birbirlerine oranları kullanılarak daha sağlam bir

parmakizi elde edilebilir. Bu mutlak değerli öznitelikler, daha küçük başka değerlere dönüştürülerek hafıza ve işlem gücü olarak kazanç elde edilebilir.

### **2.1.3 Parmakizi Modelleme**

Parmakizi modelleme modülü genellikle çerçeve çerçeve hesaplanmış bir dizi öznitelik vektörünü girdi olarak alır. Burada seçilen model, arama modülündeki mesafe hesabını ve indeksleme yapısını belirler. Girdi olarak gelen öznitelik vektörlerindeki gereksiz kısımlar göz ardı edilerek parmakizi boyutu küçültülebilir. Bir örnekteki veya bütün veritabanındaki çok tekrarlı vektörler gereksiz kabul edilebilir. Genel olarak, girdi olan birden çok öznitelik vektörünü tek bir vektörde özetleyerek parmakizi üretilebilir. Bunun için ortalama ve sapmalar kullanılabilir.

### **2.1.4 Veritabanında Arama ve Mesafe Hesabı**

Bir tanıma sisteminin kullanılabilirliğini belirleyen faktörlerden biri de bilinmeyen bir ses parçasının bilinen milyonlarca ses parçasıyla etkin bir şekilde karşılaştırmasını yapabilmesidir. Karşılaştırma yöntemi parmakizi modeline bağlıdır. Genel yaklaşım, bir sorgu verildiğinde yapılacak olan mesafe hesabı sayısını azaltmak için bir dizin (index) yapısı oluşturmaktır. Birçok dizinleme yöntemi, benzer sınıfları gruplar, bazı sınıfları gözardı eder ve geri kalan sınıflar için kaba kuvvet arama yapar. Bazı yöntemler, basit bir mesafe hesabıyla çoğu adayı hızlıca eler, pahalı mesafe hesabı kullanan yöntemler de dizin kullanarak kaba kuvvet arama yapmaktan kaçınırlar. Haitsma ve arkadaşları [22] parmakizi parçalarını dizinde (index) tutar ve bir aranan ses parçasının en az bir tane parmakizi parçasının tam olarak bulunabileceğini kabul ederek, bu dizin içinde arama yaparak hızlıca olası ses parçası aday sayısını küçük bir gruba indirir. Kalan az sayıdaki aday arasında da kaba kuvvet arama yaparlar.

Veritabanında arama sonucunda, bazı ses parçalarıyla belirli miktarlarda eşleşme elde edilmektedir. Ses parçalarıyla eşleşme miktarını ifade etmek için mesafe hesabı yöntemleri kullanılır. Mesafesi en kısa olan ses parçası en güçlü adaydır. Mesafe metrikleri büyük oranda, seçilen parmakizi modeline bağlıdır. Mesela farklı boyutlardaki vektör serilerini karşılaştırırken Öklid Uzaklığı (Euclidean distance) yöntemi kullanılabilir. Öznitelik vektörlerinin nicemlendiği (quantized) durumlarda

Manhattan Uzaklığı (Manhattan distance) veya nicemlemenin ikili (binary) olduğu durumlarda Hamming Uzaklığı (Hamming distance) yöntemleri sıklıkla kullanılır.

### **2.1.5 Hipotez Testi**

Bu son adım, aranan sorgunun veritabanında olup olmadığına karar verir. Aranan parmakizinin veritabanındaki parmakizleriyle karşılaştırılması sırasında, mesafe hesaplama sonucu üretilen skorlar elde edilir. Doğru bir eşleşme olduğuna karar vermek için skorun belirli bir eşik değerini (threshold) aşması gerekir. Eşik belirlemek kolay değildir, çünkü göz önünde bulundurulması gereken birçok faktör vardır:

- Kullanılan parmakizi modeli
- Sorgunun ayırdedici bilgi miktarı
- Veritabanındaki parmakizlerinin benzerliği
- Veritabanı boyutu

Veritabanı boyutu büyüdükçe yanlış tanıma ihtimali artar.

## **2.2 Ses Parmakizi Çıkarma İçin Kullanılan Ses Öznitelikleri**

Ses analizi çalışmalarında kullanılan ses öznitelikleri temelde yüksek düzey öznitelikler (high-level features) ve alçak düzey öznitelikler (low-level features) olarak 2 gruba ayrılabilir. [23]

### **2.2.1 Yüksek Düzey Öznitelikler (High-Level Features)**

Yüksek düzey öznitelikler, bir insanın, sesi (müziği) dinlerken edindiği bilgi çeşitlerini içerir. Melodi, ses rengi (timbre), ritim, harmoni, perde gibi öznitelikler bu grupta sayılabilir. Bu tip özniteliklerin elde edilmesi zordur ve genelde müzik tipindeki seslerde kullanıma uygundur. Çalışmamız reklam tanıma konusunda olduğu için, bu tip öznitelikler kullanılmamıştır. Bu nedenle yüksek düzey özniteliklerin detaylarına girilmemiştir.

### **2.2.2 Alçak Düzey Öznitelikler (Low-Level Features)**

Alçak düzey ses öznitelikleri, ses sinyallerinin birtakım ölçümleridir. Birçok alçak düzey ses öznitelikleri, sesin kısa zamanlı spektrumundan (Short Time Spectrum) elde edilir. Az sayıda da olsa bazı ses öznitelikleri de zaman alanındaki ses dalgasının ölçümlerinden elde edilir. Aşağıda, kullanılabilir bazı öznitelikler verilmiştir. Bunların yanısıra, yapılacak çalışmanın ihtiyaçlarına göre, tamamen yeni öznitelikler de türetilip kullanılabilir.

Kısa zamanlı enerji (Short-Term Energy), zaman domeninde (time domain) çıkarılan bir özniteliktir. Zamana göre, ses sinyalindeki enerji değişimi, parmakizi olarak kullanılabilir. Ama tek başına yeterli miktarda ayırd edici değildir.

Kısa zamanlı sıfır geçiş oranı da (Short-Term Zero Crossing Rate) zaman domeninde (time domain) çıkarılan bir özniteliktir. Ses dalgasının, zamanda ilerledikçe sıfır noktasından geçişlerinin sayısı kullanılarak parmakizi üretilebilir. Ama tek başına bu öznitelik de yeterince ayırd edici değildir.

Müzik olayları başlama noktaları (Onset) ile şekillenir, bir nota ani bir vuruşla başlar, sürer ve gittikçe gücünü kaybeder. Başlama noktaları bulma işlemi, notaların başlangıç zamanlarının işaretlenmesidir. Başlama noktalarının aralarındaki mesafeler öznitelik olarak kullanılabilir.

İnsan tarafından üretilen sesler dil, diş gibi ses üretme bölgelerinin şekilleri ile filtrelenirler. Bu şekiller, hangi sesin çıkacağını belirler. Bu şekli tam olarak belirleyebilirsek, bize, üretilen sesle ilgili kesin bir bilgi verecektir. Ses bölgesinin şekli, kendisini kısa zamanlı güç spektrumu (Short Time power spectrum) paketinde gösterir. Mel/Log-Frequency Cepstral Katsayıları (Coefficients) bu paketi tam olarak ifade etmektedir [24]. MFCC öznitelikleri otomatik konuşma ve konuşmacı tanıma sistemlerinde sıkça kullanılırlar.

Bir ses sinyalinin, orta-üst spektral bantlardaki genlik geçişlerinin ölçülerek, detaylı spektral-zaman yapısının gösterimine spektral akış (spectral flux) denir. Ortaya çıkan öznitelik, satırlarında frekans bantları, sütunlarında da geçiş frekansları olan ve her bir spektral banttaki güç değişimini ifade eden bir matristir.

Perde-sınıfı profili (Pitch-Class Profile, Chromagram), spektrumun 12 frekans bandına bölünmesiyle elde edilir. Batı müziğinde eşit aralıklı 12 perde sınıfı vardır, bu nedenle

Perde-sınıfı profilinde de 12 bant vardır. Müzikte tam 1 oktav aralıklı iki nota aynı algılandığı için gerçek frekans değeri kullanılmadan, bütün oktavlardaki notalar aynı perde sınıfına denk gelecek şekilde spektrum bölünür. Bazen daha detaylı perde bilgisi kullanabilmek için, spektrum 12'nin katları şeklinde bantlara da bölünebilir.

Parlaklık, ses sinyalindeki yüksek frekanslı içeriğin ölçümüdür. Parlaklığın hesaplanması için, kısa zamanlı Fourier güç spektrumunun frekans merkezi bulunur.

Spektrumun, birtakım alt frekans bantlarına bölünmesi ve her alt bantın frekans olarak merkezinin bulunması ile spektral alt bant merkezleri (Spectral Subband Centroids) elde edilir. Genellikle, spektrumun bölüneceği alt bant sayısına, insan kulağının duyma frekans aralıklarına göre karar verilir.

### **2.3 Ses parmakizi ile tanıma sistemlerinin performans parametreleri**

Yine aynı çalışmada [3] ses parmakizi sistemlerinden beklenen performans parametreleri aşağıdaki şekilde sıralanmıştır.

#### **2.3.1 Gürbüzlük (Robustness)**

Gürbüzlük, bir ses nesnesinin, yüksek sinyal kayıplarında - bozulmalarında bile tanınabilmesini ifade eder. Bir ses parmakizi sisteminin gürbüz olabilmesi için, bir dereceye kadar sinyal bozulmalarına karşı değişmeyen özellikler kullanması lazım. Bir ses parmakizi sisteminin gürbüzlüğü, test sonuçlarındaki "yanlış negatif" (FN) değeriyle ifade edilir. Yani tanınması gereken ama tanınmayan test nesnelere oranı, sistemin ne kadar gürbüz olduğunu ifade eder. "yanlış negatif" (FN) değeri ne kadar küçükse sistemin gürbüzlüğü o kadar yüksek kabul edilir.

#### **2.3.2 Güvenilirlik (Reliability)**

Güvenilirlik, bir ses nesnesinin tanıma işlemi sonucunda, yanlış olarak tanınmaması gerektiğini ifade eder. Test sonuçlarındaki "yanlış pozitif" (FP) değeri sistemin güvenilirliğini ifade eder. "yanlış pozitif" değeri ne kadar küçükse sistemin o kadar çok güvenilir olduğu kabul edilir.

#### **2.3.3 Parmakizi boyutu (Fingerprint size)**



Belirli süredeki bir ses verisini ifade etmek için kullanılması gereken parmakizinin ne kadar saklama alanına ihtiyaç duyduğunu ifade eder. Genellikle arama işleminin hızlı yapılabilmesi için parmakizleri hafızada tutulur, parmakizi boyutu da ne kadar şarkının hafızada tutulabileceğini belirler.

### **2.3.4 Parçalılık (Granularity)**

Bir ses verisinin tanınabilmesi için gerekli olan minimum ses verisi miktarını ifade eder. Bu miktar ne kadar küçükse, sistemin parçalılığı o kadar yüksektir ve iyidir.

### **2.3.5 Arama süresi ve ölçeklenebilirlik (Search speed and scalability)**

Bir ses verisinin, veritabanında aranma - bulunma süresidir. Bu sürenin olabildiğince kısa olması gerekmektedir. Özellikle gerçek zamanlı uygulamalarda, arama süresinin kısa olması önemlidir. Veritabanındaki ses parmakizi sayısının çok artması durumunda da arama süresinin fazla artmaması beklenir. Bu da ölçeklenebilirliği ifade etmektedir.

Bu 5 parametre birbiriyle sıkı ilişkilidir. Mesela kısa sürede şarkı tanınabilmesi için, daha büyük boyutlu parmakizi kullanılmalıdır ki aynı güvenilirlik elde edilebilsin. Bir başka örnek olarak, genellikle daha sağlam parmakizi kullanıldığında arama süresi de kısaldır. Gerçekleştirilecek uygulamanın ihtiyaçlarına göre bu parametrelerden hangilerinin daha kritik olduğuna karar verilebilir ve ona uygun şekilde bir sistem tasarlanabilir.

## **2.4 Ses parmakizi uygulama alanları**

Ses parmakizi kullanılarak, bilinmeyen bir ses parçasının, bilinen ses verilerinden biri olup olmadığının bulunması üzerine çeşitli uygulamalar yapılmıştır. Pedro Cano ve arkadaşları olası uygulamaları aşağıdaki şekilde gruplandırmışlardır [25] :

### **2.4.1 Ses içeriği gözlemlemek**

Telif hakları kontrolü veya çalma listeleri oluşturmak için, ses içeriği yayınlarının takip edilmesidir.

#### **2.4.1.1 Dağıtıcı (Distributor) tarafında gözlemlemek**

İçerik dağıtımıcılar, içeriği son kullanıcılara yayınlamak için hakları olup olmadığını bilmek isteyebilirler. Ses parmakizi teknikleri, TV ve Radyo kanallarının arşivindeki, tanımsız içeriklerin tanımlanması için kullanılabilir.

#### **2.4.1.2 İletim kanalları tarafında gözlemlemek**

Bir çok ülkede radyo istasyonları yayınladıkları müzik parçaları için telif hakkı öderler. Bu nedenle hak sahipleri radyo kanallarını gözlemleyerek telif haklarının uygun şekilde ödenip ödenmediğini takip ederler. Telif hakkı konusu olmasa bile, hak sahipleri istatistiki bilgi çıkarmak için bu şekilde gözlem yaparlar. Reklam sahipleri de, kanalları gözlemleyerek, verdikleri reklamların düzgün bir şekilde yayınlandığından emin olmak isterler.

Torrent veya napster gibi paylaşım sistemlerinde de telif hakkı ihlalleri yapılmaktadır. Bu sistemlere ses parmakizi tanıma özelliği eklenerek telif hakkı olan içerikler kontrol edilebilir.

#### **2.4.1.3 Kullanıcı tarafında gözlemlemek**

Kullanıcı bir şarkıyı çalarken de ses parmakizi yöntemiyle tanınır, telif hakları kapsamında kullanılma kontrolü yapılabilir. Bu sistem için, destekleyen cihazlar (CD - DVD çalarlar, bilgisayarlar) gerekmektedir. Aynı zamanda bu cihazların, bilinen şarkı veritabanına erişim için internet'e bağlantılarının da olması gerekmektedir.

#### **2.4.2 Katma değerli servisler**

Dinleyiciler, dinledikleri bir ses içeriği ile ilgili birtakım bilgilere erişmek isterler. Bu bilgiler arasında, şarkının adı, bestecisi, yorumcusu, yayınlanma tarihi, sanatçı biyografisi, albüm kapağı, albümdeki diğer şarkılar sayılabilir. Sıradan dinleyiciler bu tip bilgiler öğrenmek isterken, müzisyen bir dinleyici şarkıda kullanılan enstrümanları öğrenmek isteyebilir. Bir müzik mühendisi, şarkının kayıt süreciyle ilgili bilgilere ulaşmak isteyebilir. Kısacası kullanıcı profiline göre değişebilecek şekilde çok çeşitli bilgilere ulaşmak istenebilir. Bu bilgilere ulaşmak için öncelikle şarkının ses parmakizi yöntemleriyle tanınması gerekmektedir.

#### **2.4.3 Bütünlük kontrolü sistemleri**

Bazı uygulamalarda, bir ses verisi kullanılmadan önce bütünlüğünden ve bozulmamış olduğundan emin olmak istenir. Mesela sıkıştırılmış bir ses dosyasının, kullanılmadan önce algısal olarak ne kadar bozulduğunun kontrol edilmesi istenebilir. Veya reklam veren bir kurum, reklamının bütün ve bozulmadan yayınlandığını kontrol etmek isteyebilir.

#### **2.4.4 Tekrarlı kayıtların bulunması**

Günümüzdeki çok büyük multimedya arşivleri, milyarlarca kayıt içermektedir. Bu kadar büyük sayıdaki kaydın içindeki tekrarlı kayıtların bulunup teke indirilmesi önemli bir işlemdir. Bu kadar büyük sayıdaki kayıt için bu işlemin manual olarak yapılabilmesi mümkün olmadığından, ses parmakizi teknikleri kullanılır. [26]



### 3. ÖNERİLEN YÖNTEM

Çalışmamızda, televizyon veya radyo yayınları üzerinde reklam tanınması yapan bir sistem geliştirilmesi amaçlanmıştır. Sistemin hem televizyon hem de radyo yayınlarındaki reklamları tanıyabilmesi amaçlandığı için görsel parmakizi teknikleri değil, ses parmakizi teknikleri kullanılmasına karar verilmiştir. Ses parmakizi ile ilgili yapılan çalışmaların bazıları sadece şarkı tanınması ile ilgilendiği için ses parmakizini oluşturacak öznitelikler olarak, yüksek düzey öznitelikler (Melodi, ses reng, ritim, harmoni, perde) kullanılmıştır [27]. Bizim çalışmamızda, reklamlarda müzikli bölümlerin yanında konuşma ağırlıklı bölümlerin de olduğunu düşünerek, yüksek düzey öznitelikler yerine alçak düzey özniteliklerin kullanılmasına karar verilmiştir. Literatürdeki alçak düzey ses öznitelikleri kullanan çalışmalar incelendiğinde, [15] çalışmasını baz alıp, bu algoritmanın performansını iyileştirmeye ve gürültüye karşı gürbüzleştirmeye yarayan bir sistemin geliştirilmesine karar verilmiştir.

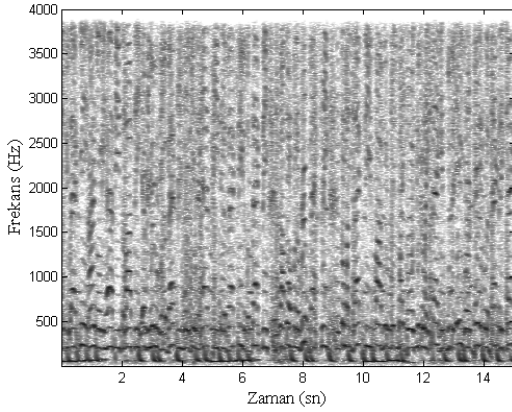
#### 3.1 Temel prensipler

Parmakizi olarak kullanılacak özniteliklerin zamansal olarak yerel, başlangıç noktasından bağımsız, sağlam ve yeterli miktarda bağımlı olmaları gerekmektedir. Zamansal olarak yerel olması, parmakizini oluşturan karmaların (hash) herbirini oluşturan özniteliklerin zamansal olarak komşu ses örneklerinden alınmış olmasıdır. Böylece uzak noktadaki olaylar birbirini etkilemeyecektir, bu sayede bir ses verisinin tanınması için tamamına ihtiyaç duyulmayıp kısa bir bölümü yeterli olacaktır. Başlangıç noktasından bağımsız olması, sorgunun şarkı başından itibaren dinletilmesinin zorunlu olmaması, herhangi bir noktadan başladığımızda da tanıma yapılabilmesidir. Sağlamlık gürültü ve sinyal bozulmalarına karşı dayanıklı olmasıdır. Yeterli miktarda bağımlılık ise, karmaları oluşturan öznitelikler arasında yeterli miktarda bağımlılık olmasıdır. Fazla bağımlılık olursa çok kırılgan ve tekrar

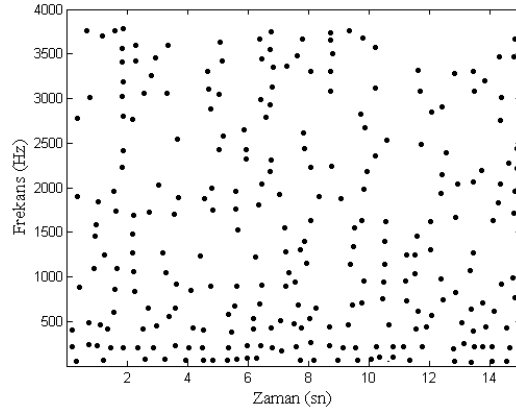
üretilemeyen karmalar, az bağımlılık olursa da yanlış tanıma sonuçları ortaya çıkabilmektedir.

### 3.2 Takımyıldızı Haritaları (Constallation map)

Wang ve arkadaşları, yüksek oranda gürültü ve sinyal bozulmasına karşı dayanıklı olacağını düşündükleri için spektrogram zirvelerini parmakizi özneliği olarak kullanmaktadırlar. Spektrogramdaki bir zaman-frekans noktası, kendi etrafındaki bir bölgedeki bütün komşularından daha yüksek bir enerji düzeyine sahipse zirve noktasıdır. Sonuç olarak Şekil 3.1'deki karışık spektrogram, Şekil 3.2'de görülen seyrek bir koordinatlar kümesine düşürülmüş olur. Gürültü sebebiyle bazı yeni zirve noktaları ortaya çıksa bile, diğer zirveleri çok etkilemeyecektir, çünkü zirveler yerel olarak birbirinden bağımsızdır. Benzer şekilde bazı zirveler silinse bile büyük bölümü korunacaktır. Bu koordinat gösterimi, literatürde takımyıldızı haritası (Constallation map) olarak isimlendirilmiştir.



Şekil 3.1: Spektrogram



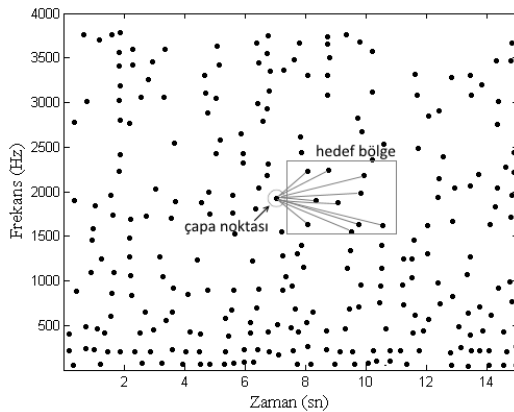
Şekil 3.2: Zirve Noktaları

Çalışmamızda, tek kanallı, 8000 Hz örnekleme oranındaki ses kayıtlarını kullandık. Ses dalga sinyalini, her birinde 2048 örnek sayısı olan ve her adımda 128 örnek ilerleyecek şekilde pencerelere (frame) böldük. Küçük pencereler, yüksek frekanslı olayları zaman ekseninde daha detaylı incelemeye, büyük penreceler ise uzun süreli olayları frekans ekseninde daha detaylı incelemeye olanak sağlar [28]. Genelde çalışmalarda 2048 örnek uzunluğunda pencere boyutu kullanıldığı için ve bizim denemelerimizde de iyi sonuç verdiği için biz de çalışmamızda bu şekilde kullandık. Pencerelerin örtüşmesi (overlap), parmakizi için kullanılacak özneliliklerin zaman

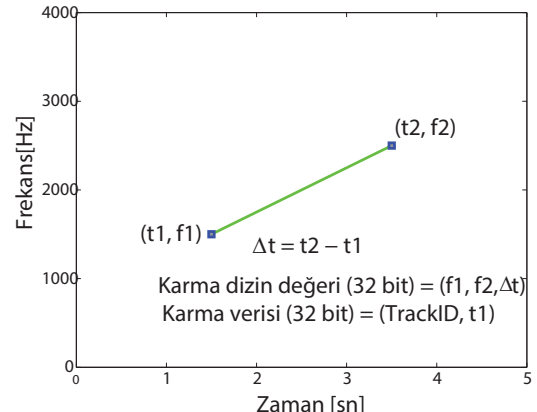
içinde yavaş değişmesini sağlar. Böylece ses kaydının başlangıç noktası pencerenin başı olmasa da öznitelikler çok değişmez [3]. Denemelerimiz sonucunda hem saklama alanı boyutu gereksinimi, hem de hesaplama maliyeti olarak en optimum örtüşme oranının 15/16 olduğuna karar verdik. Her pencereye Hanning Window fonksiyonu uygulayarak, başlangıç ve bitişindeki devamsızlıkların azalmasını sağladık. Daha sonra pencerelere Fast Fourier Transform (FFT) uygulayarak sonuçta ses kaydının spektrogramını elde ettik. Spektrogramdaki noktaların zirve olarak kabul edilmesi için 200 milisaniye ve 500 Hz komşuluğundaki en büyük enerjiye sahip olması şartını uyguladık.

### 3.3 Kombinasyonel Karmalama (Combinatorial Hashing)

Doğrudan zirve noktalarını kullanarak doğru ses kaydını ve başlangıç noktasını bulmak yavaş olacaktır, çünkü yeterince bağımlılık içermemektedirler. Bunun yerine, parmakizi karmaları, takımyıldızı haritasındaki zirve noktalarının ikili kombinasyonlar şeklinde eşleştirilmesiyle elde edilmektedirler. Çapa noktaları (anchor point) seçilir ve her çapa noktasının bir hedef bölgesi vardır. Her çapa noktası, hedef bölgesi içinde kalan zirve noktalarıyla sırayla eşleşir. Her eşleşme çifti, iki tane frekans değeri ( $f_1$  ve  $f_2$ ) ve bir de bu noktalar arasındaki zaman farkından ( $\Delta t$ ) oluşan bir karma değeri üretir. Şekil 3.3 ve Şekil 3.4'da bu işlemin detayları görülebilir. Bu karma değerleri gürültü ve ses sıkıştırma kodlayıcılarının varlığında bile tekrar üretilebilir.



Şekil 3.3: Kombinasyonel karmalama



Şekil 3.4: Karma üretimi

Daha sonra bu karmalar 32 bit olarak saklanır. Bu karma üretimi, tanınmak istenen bütün ses kayıtları için tekrarlanır ve karma değerlerine göre bir dizin (index)

oluşturulur. Ayrıca karma değerleriyle ilişkili olarak bu karma değerinin hangi ses kaydında geçtiği ve çapa noktasının, o ses kaydının başlangıcından itibaren ne kadar zaman farkı olduğu da tutulur, ama bu değerler karma dizinine dahil edilmez. Böylece 32 bit karma değeri ve 32 bit ses kaydı tanımlayıcısı (TrackID) ve çapa noktasının başlangıç mesafesi olacak şekilde toplam 64 bitlik veri yapıları elde edilir. Veri yapısı detayları Şekil 3.5’da gösterilmiştir.

10 bit $f_1$	10 bit $f_2$	12 bit $\Delta t$	32 bit TrackID + $t_1$
--------------	--------------	-------------------	------------------------

**Şekil 3.5:** Veri yapısı

Çapa noktası sayısı ve hedef bölge boyutu belirlenmesi önemlidir. Seçilen çapa noktası sayısı ve hedef bölgesi boyutu, oluşturulacak karma sayısını belirleyecek ve dolayısıyla da saklama alanı ihtiyacı ve hesaplama maliyetini belirleyecektir.

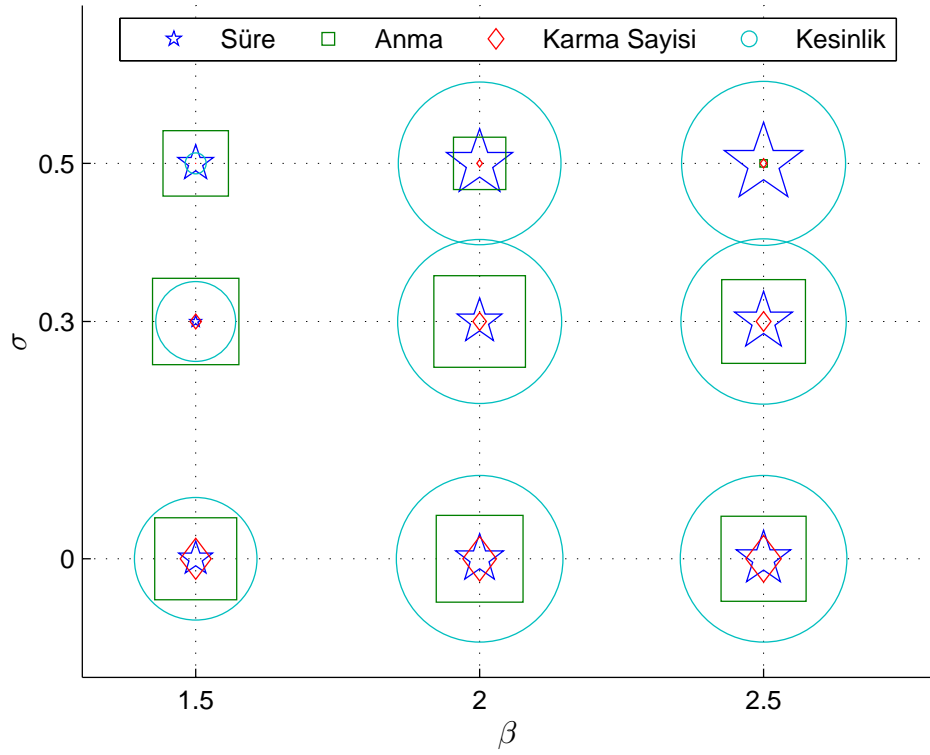
Çalışmamızda, hedef bölge boyutu olarak çapa noktasından sonraki 700 milisaniye ve çapa noktasının altı ve üstündeki 2500 Hz lik frekans alanı kullanılmıştır.

Çalışmamızda, bütün zirve noktalarının kullanılması yerine, hesaplama maliyeti ve veri boyutunu düşürmek için sorgulama sırasında sadece en güçlü zirvelerin *zaman içerisinde azalan bir şekilde* seçilerek kullanılmasına karar verilmiştir. Sorgulanan ses parçası içinde ilerlendikçe Denklem (3.1) uyarınca, her saniye azalan şekilde en güçlü bağlı  $n$  değeri ile çarpılarak bulunan zirve sayısı kadar karma değeri kullanılmaktadır.

$$n = \frac{1}{e^{\sigma \cdot t}} \quad (t = 1, 2, 3, \dots [sn]) \quad (3.1)$$

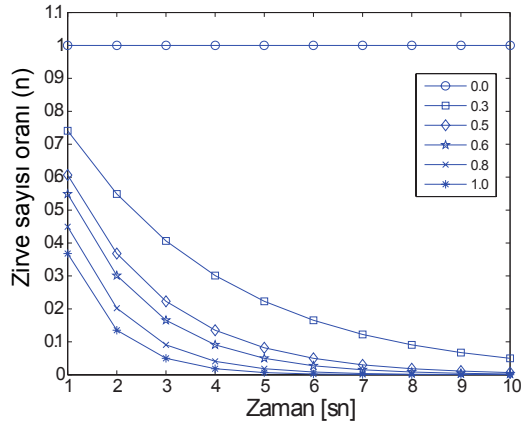
*Karma azaltma katsayısı*,  $\sigma$ 'nın en uygun değerini tespit etmek amacıyla,  $\alpha$ 'nın sabit tutulup  $\beta$  ve  $\sigma$ 'nın farklı değerleri için deneyler yapılmıştır ve buluşsal olarak  $\sigma$  için en uygun değer 0,3 olduğu sonucuna varılmıştır. Şekil 3.6’de, bahsedilen deneylerin sonuçları gösterilmiştir. Şekilde başarı oranları büyüklükleriyle orantılı olacak şekilde gösterilmiştir. Buna göre, anma ve kesinliğin büyük, süre ve karma sayısının ise küçük gösterildiği durum tercih edilmiştir.





Şekil 3.6:  $\beta$  ve  $\sigma$  parametrelerine göre başarımları

Şekil 3.7’de  $\sigma$ ’nın farklı değerleri alması durumunda karma sayısının zamana göre bağlı olarak değişimi gösterilmiştir.



Şekil 3.7: Hash sayısı oranı

### 3.4 Arama ve Skorlama

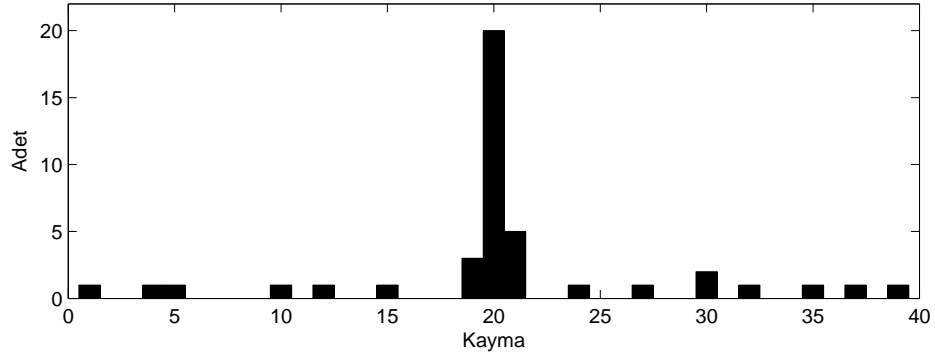
Bir ses tanıma sisteminin performansını belirleyen en önemli aşamalarından biri de, verilen bir sorgunun, veritabanında varolan çok sayıda kayıt arasından hızlı ve

doğru şekilde aranabilmesidir. Buradaki kritik nokta, arama süresinin, veritabanındaki kayıt sayısının artışıdan olabildiğince az etkilenmesidir. Bu amaç doğrultusunda, kullanılan parmakizinin dizinlemeye (index) uygun olması önemlidir.

Arama sonucunda veritabanında bulunan alt-parmakizlerine göre, birden fazla aday elde edilir. Bulunan adayların sorguya benzerlik oranları skorlarıdır. En yüksek skorlu adayın skoru önceden belirlenen bir eşik değerinde yüksek ise, sorgunun bulunduğu karar verilir. Skorlama aşamasında da en önemli noktalardan biri eşik değerinin doğru belirlenmesidir.

### 3.4.1 Arama

Arama yapılacak ses kaydı parçacığı için yukarıdaki karma üretimi işlemi uygulanır. Çıkarılan karma listesi veritabanında aranır ve bulunanlar için, sorgu ve bulunan karmaya karşılık gelen başlangıçtan zaman mesafesi ikilileri saklanır. Bu ikililer de bulunan karmanın ses kaydı tanımlayıcısına (TrackId) göre gruplanır. Bütün karmalar için arama işlemi tamamlandıktan sonra, her TrackId için bulunan karmalar incelenir. Her TrackId için bulunan zaman mesafe ikilileri, aranan ve veritabanında bulunan ses kayıtları arasındaki ilişkiyi gösteren bir dağılım diyagramı ortaya çıkarır. Eğer aranan ses kaydı ile veritabanındaki kayıt aynı kayıt ise, örtüşen özniteliklerin, her seferinde kayıt başlangıcından itibaren benzer mesafelerde oluşması beklenir, mesela bir kayıttaki bir karma sırasının aynı zaman aralıklarıyla diğer ses kaydında da olması gibi. Doğru ses kaydı için, örtüşen her öznitelik noktası için  $X = Y + kayma(offset)$  şeklinde,  $X$ 'in veritabanındaki ses kaydının başlangıcından olan mesafe,  $Y$ 'nin sorgulanan ses kaydının başlangıcından olan mesafe,  $kayma$ 'nın ise sabit bir zaman farkını ifade edeceği bir ilişki olmalıdır. Bu fomülden yola çıkarak, her eşleşen nokta için  $kayma$  değeri hesaplanıp,  $kayma$  değerlerinin dağılımını gösteren bir histogram oluşturulur. Oluşturulan histogramda güçlü bir  $kayma$  değeri varsa doğru ses kaydının bulunduğu anlaşılır. Yoksa bir sonraki TrackID için aynı şekilde histogram oluşturulur ve güçlü bir  $kayma$  değeri var mı incelenir. Şekil 3.8'deki histogramda 20  $kayma$  değeri güçlü şekilde öne çıkıyor. Bu da, sorgulanan ses kaydının, veritabanındaki ses kaydının 20. saniyesinden başlanarak arama yapıldığını ifade ediyor.



**Şekil 3.8:** Histogram

Çalışmamızda, sorgulanan ses parçası için, başlangıçtan itibaren, bir sonuç bulunana kadar her seferinde 1 saniye daha uzatılarak ilgili kısım için parmakizi çıkarılır ve veritabanındakiler arasında aranır. İterasyon aralığının deneysel olarak 1 sn olarak seçilmesinin nedeni, 1 sn'den kısa aralıklar için zirve noktalarının kararsız davranması ve uzun aralıklarda da tanıma süresinin artmasıdır.

### 3.4.2 Alternatif Karma Değerleriyle Arama

Ön çalışmada, sinyal bozulmaları veya gürültü durumlarında zirve noktalarının zaman ve frekans ekseninde kayma gösterebildiği gözlenmiştir. Bu nedenle sorgudaki karmalar veritabanında aranırken alternatif yakın değerlerinin de aranmasına karar verilmiştir. Karma değerini oluşturan  $f_1$  ve  $f_2$  frekans değerlerinin  $\pm 1$  frekans bandı, zirvelerin arasındaki  $\Delta t$  zaman farkının da yine  $\pm 1$  çerçeve adımı olarak farklı değerleri alacağı şekilde toplam  $3^3 = 27$  farklı alternatifle arama yapılmaktadır. Bu şekilde başarı oranının önemli derecede arttığı tespit edilmiştir.

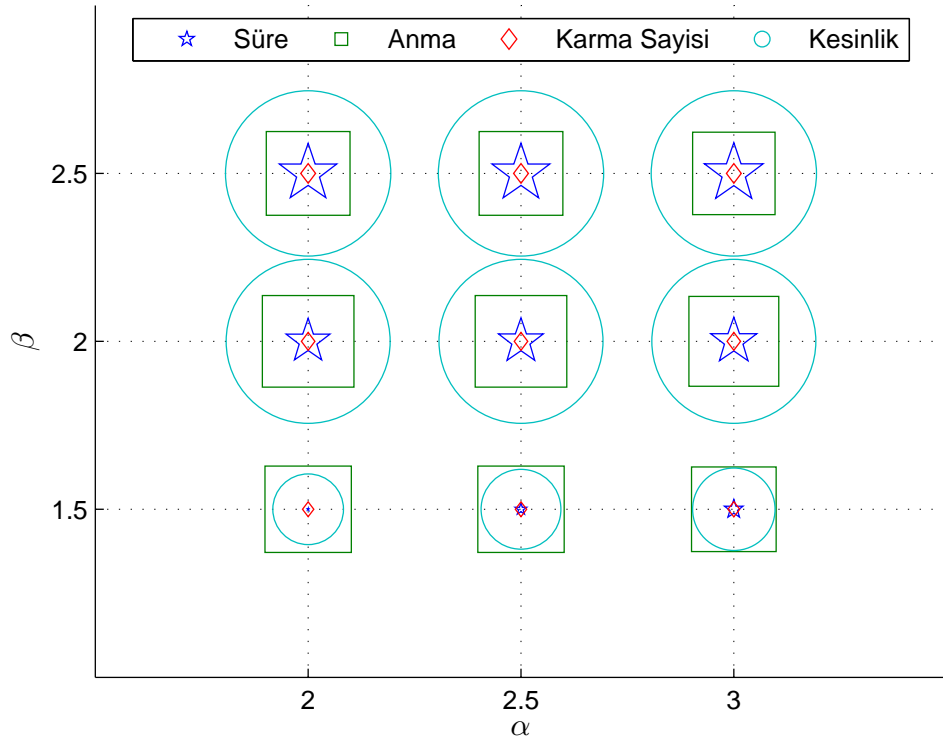
### 3.4.3 Histogram Normalizasyonu

Çalışmamızda, oluşturulan histogramlar incelenmiş ve gerçek mesafe değerinin skorunun, çevresindeki birkaç pozisyona yayılabildiği görülmüştür. Mesela Şekil 3.8'da görülen 20. kayma doğru değerken, 21. kaymanın da diğer kaymalardan yüksek değerde olmasının sebebinin aslında 20. kaymadaki bazı karmalarda oluşan bozulmalardan kaynaklandığı tespit edilmiştir. Şekil 3.8'da bu durum sorun yaratmamaktadır, çünkü bozulma miktarı küçüktür. Ama gürültü miktarı arttıkça bu bozulma miktarının artabildiği ve en yüksek kaymanın skorunun eşik değerini

aşmamasına sebep olduğu, bu nedenle de başarımlar oranlarını düşürdüğü görülmüştür. Bunun önüne geçmek için histogram üzerinde skor değerlendirilmesi yapılmadan önce bir normalizasyon yapılmasına karar verilmiştir. Buna göre histogramdaki komşu değerler, kendi aralarındaki en büyük değer ile birleştirilmektedir. Bu şekilde, histogramdaki zirve daha net bir şekilde ortaya çıkmaktadır ve güvenilirlik kontrolünü daha hızlı aşabilmektedir. Histogram normalizasyon işlemi de başarımlar oranının yükselmesine sebep olmuştur.

#### 3.4.4 Skorlama

Çalışmamızda, skor kontrolüyle ilgili de [15] çalışmasına bir eklenti yapılmıştır. Çalışmamız reklam tanınması amacıyla yapıldığı için ve reklamlarda tekrarlı bölümlerin sıkça görülmesinden dolayı, sadece histogram skorunun önceden belirlenmiş bir eşik değerini aşması yeterli görülmemiştir. Çünkü özellikle aynı markanın farklı reklamlarında aynı müziklerin geçiyor olması sık rastlanan bir durumdur. Bu nedenle sorgulanan reklam parçacığı ortak olan müzik bölümünden ise, bu müziği içeren reklamların hepsiyle aynı skor oranında eşleşecektir. Bu nedenle güven kriteri olarak iki kriter belirlenmiştir ve ikisinin de sağlanması gerekliliği kabul edilmiştir. Buna göre, birinci parametre olan *eşleşme oranı* ( $\alpha$ ), sorgudan üretilen karmaların ne kadarının veritabanında bulunduğunu ifade eder. İkinci parametre ise, üretilen histogramdaki en güçlü değer, ikinci en güçlü değere oranı olarak nitelenen, *güç oranıdır* ( $\beta$ ). Izgara arama (grid search) optimizasyonu ile  $\alpha = [0,015 - 0,035]$  ve  $\beta = [1,5 - 3,5]$  aralığındaki değerleri için yapılan testler sonucunda,  $\alpha = 0,025$  ve  $\beta = 2,0$  değerlerinin en iyi sonuç verdiği gözlemlenmiştir. Şekil 3.9'de  $\sigma$ 'nın sabit olarak 0,3 değeri kullanılarak,  $\alpha$  ve  $\beta$ 'nin farklı değerleriyle yapılan deneylerde alınan başarımlar oranları gösterilmiştir. Buna göre, anma ve kesinliğin büyük, süre ve karma sayısının ise küçük gösterildiği durum tercih edilmiştir.



Şekil 3.9:  $\alpha$  ve  $\beta$  parametrelerine göre başarımları



## 4. DENEYLER VE SONUÇLAR

### 4.1 Yazılım ve Donanım Özellikleri

Anlatılan yöntem C++ programlama dilinde bir konsol uygulaması olarak gerçekleştirildi ve Windows işletim sisteminde çalışacak şekilde derlendi. Test uygulaması iki görevi gerçekleştirecek şekilde geliştirildi. İlk görev, verilen bir klasör altındaki ses kayıtlarının parmakizinin çıkarılıp bir metin dosyasına yazılması, ikinci görev ise önceden yazılmış bu parmakizi verisinin hafızaya alınarak, verilen bir klasör altındaki ses kayıtlarının, hafızadaki parmakizleri arasında aranması ve sonuçların bir metin dosyasına yazılmasıdır.

Ses dosyalarının okunması için "libsndfile" [29] kütüphanesi, Fast Fourier Transform işlemi için de FFTW [30] kütüphanesi kullanılmıştır.

Testler, Intel Core i7-4702MQ 2.2 GHz işlemcili, 64 bit Windows işletim sistemli bir dizüstü bilgisayarda yapılmıştır. Test kümesinde yer alan akıllı telefon kayıtları için Samsung Galaxy S4 cep telefonu üzerinde çalıştırılan "Smart Voice Recorder" uygulaması kullanılmıştır.

### 4.2 Deney Şartları

Deneyleerde iki ayrı veritabanı üzerinde çalışılmıştır. Birincisi şarkılardan, ikincisi ise reklamlardan oluşmaktadır. İki kümeyle de testler yapıp, yöntemin hem şarkılar, hem de reklamlar için performansı incelenmiştir. Veritabanlarındaki ses dosyaları hem eğitim hem de test için kullanılmıştır.

Şarkılar veritabanındaki "Marsyas GTZAN Genre Collection" [31] şarkılarından 500 tanesi kullanılmıştır. Reklam veritabanı olarak Turkcell İletişim Hizmetleri'nin Youtube kanalındaki 144 reklamı [32] kullanılmıştır. Kullanılan reklam sorgu

parçalarının %15'i, reklamların ortak olan müzik veya konuşma kısımlarından seçilmiştir, böylece sistemin kesinlik sonuçları (yanlış alarm) incelenmiştir.

Test kümelerindeki ses dosyaları çeşitli gürültü ve sinyal bozulmalarına maruz bırakılmıştır. Bu işlem için Audio Degradation Toolbox (ADT) [33] aracından yararlanılmıştır. Ayrıca belirtilen deney kümeleri, İstanbul'daki gürültülü bir alışveriş merkezinde akıllı telefon kaydı ile gerçek ortam testlerine de tabi tutulmuştur. Deney şartlarının detayları Çizelge 4.1 altında tanımlanmıştır.

**Çizelge 4.1:** Deney Kümesi Çeşitleri

<b>Kısaltma</b>	<b>Gürültü Tipi</b>
<i>Temiz</i>	Gürültü eklenmemiş
<i>BG 5dB</i>	ADT ile 5 dB SNR oranında Beyaz Gürültü (White Noise) eklenmiş
<i>BG 0dB</i>	ADT ile 0 dB SNR oranında Beyaz Gürültü (White Noise) eklenmiş
<i>BG -5dB</i>	ADT ile 5 dB SNR oranında Beyaz Gürültü (White Noise) eklenmiş
<i>KG 5dB</i>	ADT ile 5 dB SNR oranında Kahverengi Gürültü (Brown Noise) eklenmiş
<i>KG 0dB</i>	ADT ile 0 dB SNR oranında Kahverengi Gürültü (Brown Noise) eklenmiş
<i>KG -5dB</i>	ADT ile -5 dB SNR oranında Kahverengi Gürültü (Brown Noise) eklenmiş
<i>PG 5dB</i>	ADT ile 5 dB SNR oranında Pembe Gürültü (Pink Noise) eklenmiş
<i>PG 0dB</i>	ADT ile 0 dB SNR oranında Pembe Gürültü (Pink Noise) eklenmiş
<i>PG -5dB</i>	ADT ile -5 dB SNR oranında Pembe Gürültü (Pink Noise) eklenmiş
<i>Kırpma%10</i>	ADT ile %10 oranında kırpma (clipping) uygulanmış
<i>Kırpma%20</i>	ADT ile %20 oranında kırpma (clipping) uygulanmış
<i>Kırpma%50</i>	ADT ile %50 oranında kırpma (clipping) uygulanmış
<i>Canlı Kayıt</i>	ADT ile canlı kayıt etkisi uygulanmış
<i>ATÇ</i>	ADT ile Akıllı Telefon Çalması (SmartPhone Playback) etkisi uygulanmış
<i>ATK</i>	ADT ile Akıllı telefon Kaydı (SmartPhone Recording) etkisi uygulanmış
<i>Bar 5dB</i>	ADT ile 5 dB SNR oranında bar ortamı sesi eklenmiş
<i>Bar 0dB</i>	ADT ile 0 dB SNR oranında bar ortamı sesi eklenmiş
<i>Bar -5dB</i>	ADT ile -5 dB SNR oranında bar ortamı sesi eklenmiş
<i>AVM</i>	Gürültülü bir alışveriş merkezinde yapılan gerçek kayıt

### 4.3 Değerlendirme Kriterleri

Gerçeklenen sistem, anma (recall), kesinlik (precision), sonuç dönme süresi ve üretilen karma sayısına göre değerlendirilmiştir. Anma, tanınması gereken ses parçalarının ne kadarının doğru tanındığını gösterir. Anma hesabı Denklem (4.1)'de verilmiştir. Kesinlik, sistemin tanıdığını iddia ettiği ses parçalarının ne kadarının tanınması gerekenlerden oluştuğudur. Kesinlik hesabı Denklem (4.2)'e göre yapılmaktadır.



Sonuç dönme süresi, tanıma işleminin ne kadar sürede olduğunu ifade etmektedir. Üretilen karma sayısı ise, tanıma işlemi için kullanılan toplam karma sayısını ifade etmektedir. Anma ve kesinlik değerlerinin yüksek, sonuç dönme süresi ve karma sayısının ise küçük olması beklenmektedir.

Ses parmakizi sistemlerinden beklenen gürbüzlük performans parametresini sonuçlardaki anma değeri, güvenilirlik performans parametresini sonuçlardaki kesinlik değeri, parmakizi boyutu performans parametresini sonuçlardaki karma sayısı, parçalılık ve arama süresi performans parametrelerini de sonuçlardaki sonuç dönme süresi ifade etmektedir.

$$Anma = \frac{TruePositives}{TruePositives + FalseNegatives} \quad (4.1)$$

$$Kesinlik = \frac{TruePositives}{TruePositives + FalsePositives} \quad (4.2)$$

#### 4.4 Ön Deneyler (Laboratuar Deneyleri ve Sonuçları)

Yöntem geliştirilirken, her adım sonunda deneyler yapıp eksikleri tespit edilip gerekli iyileştirme çalışmaları yapılmıştır. Sistem performansını etkileyen faktörler ve buna yönelik çalışmalar bu bölümde incelenmiştir. İlk olarak bölüm 4.4.1’de yapısal (algoritmik) iyileştirmelerin etkisi ele alınmıştır. Bölüm 4.4.2’de sorgu süreleri bakımından varolan yöntem ve önerilen yöntem karşılaştırılmıştır. Bölüm 4.4.3’de ise  $\alpha$ ,  $\beta$  ve  $\sigma$  sistem parametrelerinin en iyi sonuç veren değerlerinin belirlenmesi için yapılan deneylerden bahsedilmiştir. Ön deneyler ve beraberinde yapılan iyileştirmeler sonrasında sistemden beklenen başarımların değerlerine ulaşılmıştır.

##### 4.4.1 Yapısal İyileştirmelerin Etkisi

Bu bölümde [15] çalışmasına yapılan yapısal katkıların deney sonuçlarına etkisi ele alınacaktır.

Yapısal iyileştirmelerin etkisi incelenirken, sonuçların daha kolay anlaşılabilmesi için birçok parametre sabitlenmiş, sadece iyileştirme yönteminin etkisi ortaya çıkarılmıştır. Buna göre; 4.4.1.1, 4.4.1.2 ve 4.4.1.3 bölümlerinde test kümesi olarak 5 sn

uzunluğunda, "ADT-W-5" gürültü tipindeki, şarkı veri kümesindeki 500 şarkı, 4.4.1.4 bölümünde ise reklam deney kümesindeki 24 reklam sorgu olarak kullanılmıştır.

#### 4.4.1.1 Alternatif Karma Değerlerinin Aramada Kullanılması

Baz algoritmayı teşkil eden [15] çalışmasındaki yöntem incelendiğinde bazı karma değerlerinin ufak değişimlere uğradığı için veritabanında bulunamadığı anlaşılmıştır. Bu değişimlerin sebebinin, spektrogramdaki zirve noktalarının yerlerinin zaman ve frekans doğrultusunda birkaç pozisyon sapması olduğu anlaşılmıştır. Bu nedenle zirve noktalarından üretilen karmalar veritabanında aranırken bulunamazsa,  $f_1$ ,  $f_2$  ve  $\Delta t$  zaman farkı değerleri  $\pm 1$  pozisyon olarak değiştirilip farklı karma değerleriyle veritabanında arama yapılmaktadır. Bu şekilde, her bir parametre için 3 farklı değer olabileceği için toplamda her karma değeri için  $3^3 = 27$  farklı değer için sorgulama yapılmaktadır. Benzer şekilde  $f_1$  ve  $f_2$  için  $\pm 1$  frekans kutusu (frequency bin) olarak değiştirilirken  $\Delta t$   $\pm 3$  örnek olacak şekilde, toplam  $3 \times 3 \times 7 = 63$  alternatifli arama şeklinde de deneme yapılmıştır. Bu iki denemenin başarımlarına etkisi sonucu Çizelge 4.2'de verilmiştir.

**Çizelge 4.2:** Alternatif karma değerleriyle aramanın başarıma etkisi

	Temel Arama [15]	27 Alternatifli Arama	63 Alternatifli Arama
<i>Kesinlik</i>	1.0	1.0	1.0
<i>Anma</i>	0.77	0.83	0.82

Alternatif değerlerle sorgulamanın, başarıma etkisi pozitif olmasına karşın, beklenen şekilde arama süresini uzattığı görülmüştür. 1 sn'lik bir karma dizisinin, var olan yöntem [15] ve alternatif sorgulamalı olarak veritabanında aranma süreleri Çizelge 4.3'de verilmiştir.

**Çizelge 4.3:** Alternatif karma değerleriyle aramanın arama süresine etkisi

	Var Olan Yöntem [15]	27 Alternatifli Arama	63 Alternatifli Arama
<i>Süre [sn]</i>	0.01	0.18	0.40

Yapılan deneyler sonucunda, 27 alternatifli arama yönteminin kullanılmasına karar verilmiştir, çünkü başarı oranı olarak hem var olan yöntemden [15], hem de 63 alternatifli arama yönteminden daha iyi sonuç vermiştir. Bununla birlikte arama süresindeki artış, gerçek zamanın hala altındadır ve kabul edilebilir olduğuna karar verilmiştir. Ayrıca diğer yapısal iyileştirmelerle birlikte arama süresinde düşüş sağlamıştır.

#### 4.4.1.2 En Güçlü Zirvelerin Kullanılması

Wang ve arkadaşları çalışmalarında, saniyede üretilen zirve sayısı ve her çapa noktası için kaç tane karma üretildiğiyle ilgili net bir bilgi vermemişlerdir. Bu değerlerin, uygulamanın ihtiyaçlarına göre belirlenebileceğini yazmışlardır.

Var olan yöntemin geliştirilmesi sonucunda, seçtiğimiz zirve komşuluk alanı boyutu ve çapa noktası hedef bölge boyutu değerlerine göre, ortalama olarak saniyede yaklaşık 25 zirve noktası ve bu zirvelerin eşleştirilmesi sonucunda da saniyede ortalama 525 karma üretildiği gözlenmiştir.

Saniyedeki karma sayısı, saklama yeri ihtiyacını ve arama süresini belirlemektedir. Hem arama süresini kısaltmak için, hem de saklama yeri ihtiyacını azaltmak için, zirve noktaları bulunduktan sonra, karma üretimi aşamasında bütün zirve noktalarını kullanmak yerine, her hedef bölge için sadece **en güçlü** 6 tane zirve noktasının kullanılması düşünülmüştür. Bu şekilde başarı oranının çok düşmeyeceği ama arama süresi ve saklama yeri ihtiyacının büyük ölçüde azalacağı öngörülmüştür. Yapılan deneyler de bu doğrultuda sonuçlar vermiştir. Çizelge 4.4’de görülebileceği gibi, başarı oranı düşmezken arama süresi önemli derecede düşmüştür.

**Çizelge 4.4:** En güçlü zirvelerin kullanılmasının etkisi

	Var Olan Yöntem [15]	En güçlü 6 zirve ile
<i>Kesinlik</i>	1.0	1.0
<i>Anma</i>	0.77	0.77
<i>Saklama alanı</i> <i>(saniyedeki karma sayısı)</i>	525	68
<i>Arama süresi [sn]</i> <i>(saniyedeki karmayı arama süresi)</i>	0.01	0.003

Çalışmanın ilerleyen aşamalarında, sabit olarak en güçlü 6 zirvenin kullanılması yerine, sorgu içinde ilerledikçe, her saniye azalan miktarda zirvenin kullanılmasına karar verilmiştir. Her saniye, Denklem (3.1) formülüne göre kullanılacak zirve sayısı belirlenmektedir.

#### 4.4.1.3 Histogram Normalizasyonu

Veritabanında bulunan karmalarla oluşturulan histogramlar incelendiğinde, bazı değerlerin birkaç komşu değere dağılabildiği görülmüştür. Bunun sebebinin de karmaları oluşturan  $f_1$  değerlerindeki ufak kaymalar olduğu tespit edilmiştir. Aslında bu ufak kaymaların hepsi de doğru değer olduğu için, histogram oluşturulduktan sonra, histogramdaki komşu değerlerin, kendi aralarındaki en büyük değer ile birleştirilmesine karar verilmiştir. Bu şekilde, histogramdaki zirve daha net bir şekilde ortaya çıkmaktadır ve güvenilirlik kontrolünü daha hızlı aşabilmektedir. Çizelge 4.5’de histogram normalizasyonu işleminin anma değerine katkısı görülebilmektedir.

**Çizelge 4.5:** Histogram normalizasyonunun etkisi

	<b>Var Olan Yöntem [15]</b>	<b>Histogram normalizasyonu ile</b>
<i>Kesinlik</i>	1.0	1.0
<i>Anma</i>	0.77	0.86

#### 4.4.1.4 Güvenilirlik Kontrolü

Wang ve arkadaşları [15] çalışmasında, güvenilirlik kontrolü olarak, histogramdaki maksimum değer, önceden belirlenmiş bir eşik değerini geçmesi şartını kullanmışlardır. Eşik değerinin belirlenmesi için de, uygulamaya göre değişebilecek, kabul edilebilir bir yanlış pozitif (FP) değerinin ve maksimum başarı oranının belirlenip, bu kriterleri sağlayacak şekilde bir eşik değeri seçilmesi gerektiğini belirtmişlerdir. Bu yöntemle göre, özellikle reklam tanınması denemelerinde, reklamlarda aynı müziklerin geçtiği durumlar olabildiği için, birden fazla reklamın eşik değerini geçebildiği, veya bir tanesinin eşik değerini geçip başka bir reklamın da eşik değerine çok yakın skor aldığı görülmüştür. Bu tip durumlarda en büyük skorlu sonucun "doğru" olarak kabul edilmesinin yanlış sonuçlar (FP) doğurabildiği tespit edilmiştir. Bu durumu

engellemek için, iki parametrelilik bir güvenilirlik kontrolü yapılmasına karar verilmiştir. Birinci parametre ( $\alpha$ ), orjinal yöntemdekine benzer olarak, sorgudan çıkarılan karma sayısının ne kadarının aday ile eşleştiğini gösterir. Ön çalışmalarda, bu şart için eşik olarak 3 değeri seçilmiştir, yani sorgudaki karmaların %3'lük kısmının veritabanındaki aday ile eşleşmesi gerekliliğidir. İkinci parametre ise ( $\beta$ ), en güçlü adayın, ikinci en güçlü adayın kaç katı olduğunun kontrol edilmesidir, ön çalışmalarımızda bu oranın 1,5'den büyük olması şartı aranmıştır.

Bu durumun net bir şekilde görülebilmesi için reklamlar deney kümesindeki, başları aynı müzikle başlayan 24 reklam ile test yapılmıştır. Sonuçlar Çizelge 4.6'de gösterilmiştir. Görüldüğü gibi, sadece eşleşme sayısına bakıldığında yanlış tanıma (FP) oranı bir hayli yüksek olabilmektedir, önerdiğimiz yöntemin kullanılması durumunda ise, tanıma oranı düşmekte, ama yanlış tanıma olmamaktadır. Tanımamak, yanlış tanımaya göre tercih edilir bir durumdur, çünkü tanımadığımız sürece dinlemeye devam edip ilerleyen saniyelerde tanıma ihtimalimiz vardır, ama yanlış tanıma durumu düzeltilebilir değildir.

**Çizelge 4.6:** Güvenilirlik kontrolünün etkisi

	Var Olan Yöntem [15]	Güvenilirlik kontrolü geliştirmesi
<i>Kesinlik</i>	0.59	1.0
<i>Anma</i>	0.87	0.29

#### 4.4.2 Sorgu süresinin etkisi

Sorgu ses parçasının süresinin artması, üretilen ve dolayısıyla aramada kullanılan karma sayısının da artması anlamına geldiği için, başarı oranlarını iyi yönde etkilemesi beklenmektedir. Tezin amaçlarından biri olan parçalılık, kısa süreli sorgularda da başarılı sonuçlar üretmeyi gerektirmektedir. Bu beklentileri deneyle incelemek amacıyla, farklı sorgu süreleriyle deneyler yapıp, var olan yöntem [15] ile geliştirilen yöntemin anma ve kesinlik sonuçları karşılaştırılmıştır. Sonuçlar Çizelge 4.7'de verilmiştir. Beklendiği şekilde, sorgu süresinin uzamasıyla, başarı oranları da artmaktadır. İki yöntemin sonuçları karşılaştırıldığında da özellikle kısa süreli sorgularda "Önerilen Yöntemin" daha iyi sonuç verdiği görülmüştür.

**Çizelge 4.7:** Sorgu süresinin etkisi

Süre	Kriter	Var Olan Yöntem [15]	Önerilen Yöntem
1 sn	Kesinlik	1.0	0.87
	Anma	<b>0.06</b>	<b>0.36</b>
2 sn	Kesinlik	1.0	0.98
	Anma	<b>0.30</b>	<b>0.73</b>
3 sn	Kesinlik	1.0	0.99
	Anma	<b>0.54</b>	<b>0.84</b>
5 sn	Kesinlik	1.0	1.0
	Anma	0.77	0.84
10 sn	Kesinlik	0.99	1.0
	Anma	0.95	0.97

#### 4.4.3 Sistem parametrelerin etkisi

Uygulanan yapısal iyileştirmelerin birleştirilmesi sonucu algoritma son haline getirilmiştir. Algoritmanın son halini alması sonrasında,  $\alpha$ ,  $\beta$  ve  $\sigma$  parametrelerinin en iyi sonuç veren değerlerinin belirlenmesi için tekrar deneyler yapılmıştır.

Bu deneylerde şarkı deney kümesindeki şarkılar kullanılmıştır ve kullanılan gürültü tipleri şunlardır : "BG 5dB", "BG 0dB", "BG -5dB", "Bar 0dB", "Bar -5dB", "Canlı Kayıt". Gürültü tiplerinin açıklamaları Çizelge 4.1'de verilmiştir.

$\alpha$ ,  $\beta$  ve  $\sigma$  parametrelerinin değişik değerleri ve belirtilen deney kümeleri kullanılarak yapılan birçok deneyin sonuçları Çizelge 4.8'de verilmiştir. Sonuç süresi, anma, karma sayısı ve kesinlik değerlerinin en iyi olduğu parametreler olarak  $\alpha=0.025$ ,  $\beta=0.20$  ve  $\sigma=0.3$  değerlerinin kullanılmasına karar verilmiştir ve bu değerler Bölüm 4.5'de anlatılan deneylerde kullanılmıştır.

#### 4.5 Deney Sonuçları

##### 4.5.1 Şarkı Kümesi Deney Sonuçları

Çizelge 4.9'de, Wang'ın [15] çalışması ile önerilen yöntemin, şarkı deney kümelerinin farklı koşullardaki gürültü ve sinyal bozulması durumlarındaki anma, kesinlik, sonuç dönme süresi ve karma sayısı sonuçları birlikte verilmiştir.

**Çizelge 4.8:**  $\alpha$ ,  $\beta$  ve  $\sigma$  parametrelerinin belirlenmesi için yapılan deney sonuçları

$\alpha$	$\beta$	$\sigma$	Süre (sn)	Anma	Karma Sayısı	Kesinlik
0.02	1.5	0	2.03	0.936	713	0.971
		0.3	1.879	0.936	237	0.950
		0.5	2.228	0.876	147	0.940
	2	0	2.345	0.955	830	1.0
		0.3	2.419	0.955	274	0.998
		0.5	3.09	0.844	170	0.997
	2.5	0	2.551	0.95	909	1.0
		0.3	2.8	0.928	298	0.999
		0.5	3.604	0.792	180	0.998
0.025	1.5	0	2.194	0.922	794	0.972
		0.3	1.916	0.937	238	0.953
		0.5	2.234	0.875	147	0.940
	2	0	2.512	0.939	914	1.0
		<b>0.3</b>	<b>2.439</b>	<b>0.954</b>	<b>275</b>	<b>0.998</b>
		0.5	3.09	0.844	169	0.997
	2.5	0	2.709	0.933	989	1.0
		0.3	2.813	0.928	298	0.999
		0.5	3.604	0.792	180	0.998
0.03	1.5	0	2.409	0.909	895	0.979
		0.3	1.978	0.931	241	0.954
		0.5	2.255	0.874	148	0.941
	2	0	2.68	0.92	995	1.0
		0.3	2.472	0.949	276	0.998
		0.5	3.094	0.844	169	0.997
	2.5	0	2.872	0.913	1070	1.0
		0.3	2.835	0.924	298	0.999
		0.5	3.604	0.792	180	0.998

#### 4.5.2 Reklam Kümesi Deney Sonuçları

Çizelge 4.10'de, Wang'ın [15] çalışması ile önerilen yöntemin, reklam deney kümelerinin farklı koşullardaki gürültü ve sinyal bozulması durumlarındaki anma, kesinlik, sonuç dönme süresi ve karma sayısı sonuçları birlikte verilmiştir.

#### 4.6 Tartışma

Sorgudan üretilen karmaların alternatifleriyle birlikte veritabanından sorgulanması ve histogram normalizasyonu işlemleriyle, anma değerlerinin yükseltilmesi hedeflenmiş ve deney sonuçlarıyla tespit edilmiştir. Şarkı deney kümesinde ortalama anma değeri baz algoritma [15] ile 0,88 iken, iyileştirmelerimiz sayesinde 0,96'ya yükselmiştir.

**Çizelge 4.9:** Şarkılar veritabanındaki anma, kesinlik, sonuç oluşturma süresi ve karma sayısı sonuçları.

Deneyler	<i>Şarkılar Deney Kümesi</i>							
	<i>Var olan yöntem [15]</i>				<i>Önerilen yöntem</i>			
	Anma	<b>Kesinlik</b>	<u>Süre</u>	<i>Hash</i>	Anma	<b>Kesinlik</b>	<u>Süre</u>	<i>Hash</i>
<i>Temiz</i>	1.0	<b>1.0</b>	<u>1014</u>	280	1.0	<b>1.0</b>	<u>1038</u>	<i>106</i>
<i>BG 5dB</i>	0.95	<b>0.998</b>	<u>1711</u>	472	1.0	<b>1.0</b>	<u>1438</u>	220
<i>BG 0dB</i>	0.86	<b>1.0</b>	<u>2364</u>	653	0.99	<b>1.0</b>	<u>1967</u>	<i>315</i>
<i>BG -5dB</i>	0.57	<b>0.998</b>	<u>3604</u>	995	0.93	<b>0.998</b>	<u>3186</u>	483
<i>KG 5dB</i>	1.0	<b>1.0</b>	<u>1016</u>	280	1.0	<b>1.0</b>	<u>1067</u>	<i>114</i>
<i>KG 0dB</i>	1.0	<b>1.0</b>	<u>1023</u>	282	1.0	<b>1.0</b>	<u>1094</u>	<i>118</i>
<i>KG -5dB</i>	1.0	<b>0.998</b>	<u>1059</u>	292	1.0	<b>1.0</b>	<u>1096</u>	<i>119</i>
<i>PG 5dB</i>	0.98	<b>1.0</b>	<u>1358</u>	375	1.0	<b>1.0</b>	<u>1262</u>	<i>169</i>
<i>PG 0dB</i>	0.87	<b>0.991</b>	<u>2059</u>	568	0.97	<b>1.0</b>	<u>1902</u>	241
<i>PG -5dB</i>	0.62	<b>0.989</b>	<u>3137</u>	866	0.81	<b>0.99</b>	<u>3411</u>	368
<i>Bar 5dB</i>	0.98	<b>0.998</b>	<u>2761</u>	762	0.99	<b>1.0</b>	<u>1382</u>	<i>149</i>
<i>Bar 0dB</i>	0.92	<b>1.0</b>	<u>3016</u>	832	0.99	<b>0.998</b>	<u>1922</u>	<i>191</i>
<i>Bar -5dB</i>	0.68	<b>0.998</b>	<u>3629</u>	1002	0.85	<b>0.995</b>	<u>3877</u>	257
<i>Kırpma%10</i>	1.0	<b>1.0</b>	<u>1057</u>	292	1.0	<b>1.0</b>	<u>1086</u>	<i>116</i>
<i>Kırpma%20</i>	1.0	<b>0.998</b>	<u>1141</u>	315	1.0	<b>1.0</b>	<u>1121</u>	<i>120</i>
<i>Kırpma%50</i>	0.96	<b>0.998</b>	<u>1667</u>	460	0.99	<b>0.998</b>	<u>1327</u>	<i>148</i>
<i>Canlı Kayıt</i>	0.54	<b>0.993</b>	<u>3551</u>	980	0.9	<b>0.995</b>	<u>2961</u>	224
<i>ATÇ</i>	1.0	<b>0.998</b>	<u>1068</u>	295	1.0	<b>1.0</b>	<u>1125</u>	<i>110</i>
<i>ATK</i>	0.9	<b>1.0</b>	<u>2269</u>	626	0.99	<b>1.0</b>	<u>1607</u>	<i>172</i>
<i>AVM</i>	0.84	<b>1.0</b>	<u>2287</u>	631	0.9	<b>0.989</b>	<u>3028</u>	202
<i>Ortalama</i>	0.88	<b>0.998</b>	<u>2094</u>	578	0.96	<b>0.998</b>	<u>1887</u>	202

Reklam deney kümesinde ise ortalama anma değeri 0,69'dan 0,75'e yükselmiştir. Reklam kümesi deneylerinde kullanılan sorguların yaklaşık %15'i bilinçli olarak, birden çok reklamda ortak olan müzik/konuşma bölümlerinden seçildiği için anma değerinin şarkılar kümesine göre düşük kalması normaldir. Özellikle gürültü seviyesi arttıkça (SNR<0 dB) baz yöntemin anma değerlerinin hızla düştüğü, ama önerilen yöntemin nispeten çok etkilenmediği görülmüştür.

Güvenilirlik kontrolü aşamasında  $\alpha$  ve  $\beta$  parametrelerinin dahil edilmesi ile, kesinlik değerlerinin yükseltilmesi sağlanmıştır. Şarkı veritabanında ortalama kesinlik değeri iki çalışmada da 0,998 olarak değişmezken, reklam deney kümesindeki %15'lik ortak müzik/konuşma bölümleri nedeniyle, [15] çalışması 0,896 ortalama kesinlik değeri verirken, önerilen algoritma sonucunda 0,994 ortalama kesinlik değerine ulaşmıştır.



**Çizelge 4.10:** Reklamlar veritabanındaki anma, kesinlik, sonuç oluşturma süresi ve karma sayısı sonuçları.

Deneyler	<i>Reklamlar Deney Kümesi</i>							
	<i>Var olan yöntem [15]</i>				<i>Önerilen yöntem</i>			
	Anma	<b>Kesinlik</b>	<u>Süre</u>	<i>Hash</i>	Anma	<b>Kesinlik</b>	<u>Süre</u>	<i>Hash</i>
<i>Temiz</i>	0.96	<b>0.958</b>	<u>1000</u>	<i>256</i>	0.82	<b>1.0</b>	<u>2707</u>	<i>160</i>
<i>BG 5dB</i>	0.78	<b>0.846</b>	<u>1888</u>	<i>521</i>	0.8	<b>0.992</b>	<u>3401</u>	<i>328</i>
<i>BG 0dB</i>	0.66	<b>0.895</b>	<u>2727</u>	<i>753</i>	0.78	<b>1.0</b>	<u>3917</u>	<i>430</i>
<i>BG -5dB</i>	0.32	<b>0.958</b>	<u>3986</u>	<i>1100</i>	0.75	<b>1.0</b>	<u>4981</u>	<i>562</i>
<i>KG 5dB</i>	0.92	<b>0.916</b>	<u>1077</u>	<i>297</i>	0.8	<b>1.0</b>	<u>2949</u>	<i>180</i>
<i>KG 0dB</i>	0.9	<b>0.895</b>	<u>1091</u>	<i>301</i>	0.79	<b>1.0</b>	<u>3006</u>	<i>179</i>
<i>KG -5dB</i>	0.9	<b>0.902</b>	<u>1126</u>	<i>311</i>	0.79	<b>1.0</b>	<u>3057</u>	<i>188</i>
<i>PG 5dB</i>	0.83	<b>0.881</b>	<u>1650</u>	<i>455</i>	0.78	<b>1.0</b>	<u>3459</u>	<i>277</i>
<i>PG 0dB</i>	0.73	<b>0.881</b>	<u>2280</u>	<i>629</i>	0.72	<b>1.0</b>	<u>3975</u>	<i>329</i>
<i>PG -5dB</i>	0.48	<b>0.902</b>	<u>3245</u>	<i>896</i>	0.61	<b>1.0</b>	<u>5115</u>	<i>422</i>
<i>Bar 5dB</i>	0.8	<b>0.874</b>	<u>2084</u>	<i>575</i>	0.82	<b>1.0</b>	<u>3624</u>	<i>209</i>
<i>Bar 0dB</i>	0.71	<b>0.916</b>	<u>3042</u>	<i>840</i>	0.76	<b>0.967</b>	<u>4503</u>	<i>251</i>
<i>Bar -5dB</i>	0.37	<b>0.958</b>	<u>4476</u>	<i>1235</i>	0.57	<b>0.938</b>	<u>6325</u>	<i>296</i>
<i>Kırpma%10</i>	0.88	<b>0.881</b>	<u>1007</u>	<i>278</i>	0.8	<b>1.0</b>	<u>2828</u>	<i>169</i>
<i>Kırpma%20</i>	0.87	<b>0.874</b>	<u>1091</u>	<i>301</i>	0.81	<b>1.0</b>	<u>2885</u>	<i>181</i>
<i>Kırpma%50</i>	0.83	<b>0.853</b>	<u>1762</u>	<i>486</i>	0.8	<b>1.0</b>	<u>3185</u>	<i>234</i>
<i>Canlı Kayıt</i>	0.24	<b>0.923</b>	<u>4084</u>	<i>1127</i>	0.64	<b>1.0</b>	<u>5045</u>	<i>258</i>
<i>ATÇ</i>	0.83	<b>0.839</b>	<u>1231</u>	<i>340</i>	0.79	<b>1.0</b>	<u>3153</u>	<i>179</i>
<i>ATK</i>	0.54	<b>0.839</b>	<u>2874</u>	<i>793</i>	0.78	<b>1.0</b>	<u>3732</u>	<i>239</i>
<i>AVM</i>	0.55	<b>0.993</b>	<u>3594</u>	<i>992</i>	0.75	<b>0.983</b>	<u>4089</u>	<i>213</i>
<i>Ortalama</i>	0.69	<b>0.896</b>	<u>2332</u>	<i>644</i>	0.75	<b>0.994</b>	<u>3854</u>	<i>270</i>

Zamana bağlı olarak, kullanılan karma sayısının Denklem 3.1'e göre azaltılması ile, kullanılan toplam karma sayısının azaltılması sağlanmıştır. Şarkı deney kümesinde ortalama karma sayısı [15] çalışması ile 578 iken, son durumda 202'ye düşmüştür. Reklam deney kümesinde ise ortalama karma sayısı değeri 644'den 270'e düşmüştür.

Yapılan katkılar aynı zamanda sonuç üretme süresini de düşürmüştür. Şarkı deney kümesinde ortalama sorgu dönüş süresi baz algoritma ile 2,09 sn iken, son durumda 1,88 sn'ye düşmüştür. Reklam deney kümesinde ise ortalama sonuç üretme süresi 2,33 sn'den 3,85 sn'ye yükselmiştir. Reklam deney kümesindeki sürenin uzamasının sebebi, bahsedilen %15'lik sorgu için kesin sonuç üretilemediği için uzun süre (10 sn) boyunca dinlenmeye devam edilmesidir. [15] çalışması ise bu ortak sorgu kısımlarının olduğu kümeyi, içinde geçen reklamlardan herhangi birisi olarak tanıdığı için, kısa sürede sonuç dönüyor gibi görünse de, kesinlik olarak kötü sonuç üretmektedir.



## 5. SONUÇ VE ÖNERİLER

Tezin amacı olan, reklam tanınması için kullanılabilecek bir yöntem tasarlanması bu çalışmada gerçekleştirilmiştir. Reklam tanıma işleminin, ses parmakizi teknikleri kullanılarak gerçekleştirilmesine karar verilmiş, bu doğrultuda, ses parmakizi çalışmaları incelenmiştir. Reklam tanınması için uygun olduğu düşünülen bir yöntem belirlenip prototip geliştirmesi yapılmıştır. Geliştirilen yöntem, spektrogramdaki zirve noktaların aralarındaki zamansal ve frekanssal mesafeler kullanılarak üretilen parmakizine dayanmaktadır. Büyük oranda gürültü ve sinyal bozulmaları durumlarında bile spektrogramdaki zirve noktaların en azından bir kısmının korunduğu tespit edilmiştir.

Prototip sistem geliştirmesi sonrasında farklı gürültü ve sinyal bozulmasına sahip deney kümeleriyle testler yapılmış ve yöntemin zayıf noktaları tespit edilmiştir. Bahsedilen zayıf noktaların sebebinin, spektrogramdaki zirve noktaların bir kısmının korunsa bile, zaman veya frekans yönünde ufak mesafelerle yer değiştirebilmesinden kaynaklandığı tespit edilmiştir. Özellikle yüksek gürültülü durumlarda ( $SNR \leq -5$  dB) zirve noktalarındaki yer değiştirmelerin arttığı ve sistemin başarımlarını ciddi şekilde düşürdüğü tespit edilmiştir. Bahsedilen problemlere çözüm önerileri sunulmuştur.

İlk olarak, zirvelerin aralarındaki mesafeler kullanılarak üretilen karma değerleri, zirvelerin yer değiştirmesi nedeniyle değişime uğradığı için, veritabanında bulunamaması problemini çözmek için, alternatif karma değerleriyle veritabanında arama işlemi yapılmıştır. Yapılan bu katkı ile, başarımlarında ciddi iyileşmeler tespit edilmiştir. Yalnız bu işlemin dezavantajı olarak veritabanında arama süresinin de beklendiği gibi uzadığı görülmüştür. Arama süresinin, gerçek zamanın (real time) altında kalması gerektiği için, yapılan deneylerle başarımlarının arttığı, ama arama süresinin makul değerlerde kaldığı şekilde, 27 alternatifli arama yöntemi uygulanmasına karar verilmiştir.

Ön deney sonuçları incelendiğinde, bütün zirvelerin aramada kullanılmasının, hem başarımlar oranını düşürdüğü hem de arama süresini arttırdığı tespit edilmiştir. Bunun sebebinin de, zirvelerin güçsüz olanlarının, gürültü ve sinyal bozulması durumlarında kolaylıkla kaybolabilmesi olduğu görülmüştür. Bu nedenle arama işleminde bütün zirvelerin kullanılması yerine, sadece en güçlü  $n$  zirvenin kullanılmasına karar verilmiştir.  $n$  sayısının da arama işlemi süresince her saniye azalacak şekilde, Denklem 3.1'e göre belirlenmesine karar verilmiştir. Yapılan deneyler sonucunda, yapılan bu katkı ile, hem başarımlar oranlarının arttığı, hem de daha az zirve kullandığımız için arama süresinin ciddi oranda azaldığı tespit edilmiştir. Böylece, alternatifli arama geliştirmesi nedeniyle artmış olan arama süresinin daha da fazlası geri kazanılmıştır.

Zirve noktalarının gürültü ve sinyal bozulmaları durumlarında kaymaları sonucunda ortaya çıkan bir başka problemin de, yöntemin skorumla aşamasında ortaya çıktığı tespit edilmiştir. Buna göre, doğru sonucun skorunun, olması gerekenden düşük kaldığı ve dolayısıyla güvenilirlik kriterini aşmasının uzun sürdüğü görülmüştür. Bu problemi gidermek için de skorumla aşamasında histogram normalizasyon işlemi uygulanmıştır. Bu katkı sonucunda yapılan deneylerde, başarımlar oranının arttığı ve bulma süresinin de kısaldığı tespit edilmiştir.

Reklam deney kümeleriyle yapılan deneylerde, yöntemin bir başka zayıf noktası tespit edilmiştir. Buna göre, aynı müzik veya konuşmaların geçtiği farklı reklamların tanınmak istenmesi durumunda, sorgu birden fazla reklama aynı skor ile benzeşebilmektedir. Varolan yöntem bu tip durumlarda birden fazla aday var mı diye bakmadan, herhangi bir adayın skoru, önceden tanımlanmış bir değeri geçerse, sorgunun bulunduğu sonucuna varıyordu. Bu da yanlış pozitif (FP) sonuçların artmasına sebep oluyordu. Bunun yerine, iki eşik değeri parametresi kullanan bir güvenilirlik kontrolü geliştirilmiştir. Buna göre hem adayın skorunun önceden belirlenmiş bir değeri aşması, hem de yeterince güçlü bir şekilde tek aday olması şartı aranmıştır. Bu geliştirme sonucunda yapılan deneylerde, kesinlik başarımlar oranının, özellikle reklam deney kümesinde ciddi şekilde arttığı tespit edilmiştir.

Bahsedilen yapısal iyileştirmeler tamamlandıktan sonra birçok deney yapılmış ve başarımlar oranlarının en yüksek olduğu sistem parametreleri tespit edilmiştir.

Son olarak, birçok gürültü ve sinyal bozulmaları içeren şarkı ve reklam deney kümeleriyle deneyler yapıp sistemin anma, güvenilirlik, kullanılan veri boyutu ve sorgunun bulunma süresi sonuçları incelenmiştir. Bahsedilen sonuç değerleri, aynı zamanda ses parmakizi sistemlerinden beklenen gürbüzlük, güvenilirlik, parmakizi boyutu, parçalılık ve arama süresi performans parametrelerini de ifade etmektedir. Bütün sonuç değerlerinde, baz alınan yönteme göre ciddi şekilde iyileştirmeler elde edildiği görülmüştür. Buna göre karma sayısının ve ortalama ses eşleme süresinin azaldığı, hassasiyet ve anma değerlerinin arttığı teyit edilmiştir.

Çalışma şimdilik bir windows uygulaması olarak çalışmaktadır. Önceden sisteme tanıtılmış şarkı veya reklamlardan biri çalınırken, uygulama ekranındaki bir butona basıldığında, sistem mikrofonundan dinleyerek tanıma işlemi yapılmaktadır.

Şu aşamada, sisteme tanıtılan şarkı veya reklamların ses parmakizleri bir metin dosyasında saklanmaktadır, tanıma işleminin başlamasından önce, başka bir buton ile bu metin dosyasından okunup hafızaya yüklenmektedir. Hafızada tutulabilecek veri boyutu sınırlı olduğu için, sisteme tanıtılabilecek şarkı veya reklam sayısı da sınırlı kalmaktadır. Yapılan testlerde en fazla 500 şarkı veya reklam kullanılmıştır. Sistemin daha kullanılabilir olması için gerçek bir veritabanı alt yapısı kullanacak şekilde geliştirilmesi gerekmektedir. Kullanılacak veritabanının dizin (index) yapısına sahip olması zorunludur, aksi halde arama süreleri, veritabanına yazılacak içerik sayısı arttıkça artacaktır ve sistem kullanılamaz hale gelecektir.

Sistemin gerçek anlamda kullanılabilmesi için, akıllı telefon uygulaması geliştirilmesi gerekmektedir. Akıllı telefon uygulaması, mikrofonundan dinleyip, sorgunun parmakizini üretip, sunucuya bütün ses verisi yerine sadece parmakizi verisini göndermeli, sunucu ise bu parmakizini veritabanında arayıp skorlama ve güvenilirlik kontrolü sonrası sonuca karar vermeli ve uygulamaya olumlu veya olumsuz cevap dönmelidir. Sunucunun arama işlemini yapıp cevap dönmesi süresince akıllı telefon uygulaması mikrofondan dinlemeye devam etmeli, sunucunun cevabının olumsuz olması durumunda, daha uzun süreli ses verisi içeren parmakizini sunucuya göndermelidir.

Akıllı telefon uygulaması ile, ses parmakizi üretim hızı incelenmelidir, eğer gerçek zamanlı çalışmayı engelleyecek kadar yavaşlık tespit edilirse uygun şekilde çözümler üretilmelidir. Sistem parametreleri değiştirilerek, zirve hesabı hızlandırılabilir veya kullanılacak zirve sayısı değiştirilerek hesaplama maliyeti azaltılabilir.

Sunucu tarafındaki arama ve skarlama işlemi, paralel olarak çalışmaya çok uygun işlemlerdir. Sunucu tarafı geliştirmesi paralel çalışma mimarisine uygun şekilde yapılırsa, sorgulanan parmakizi içerisindeki çok sayıda karmanın veritabanında aranma süresi ciddi şekilde azaltılabilir.

## KAYNAKLAR

- [1] **The New Multi-screen World - Understanding Cross-platform Consumer Behavior**, [https://think.withgoogle.com/databoard/media/pdfs/the-new-multi-screen-world-study\\_research-studies.pdf](https://think.withgoogle.com/databoard/media/pdfs/the-new-multi-screen-world-study_research-studies.pdf), alındığı tarih: 11.04.2015.
- [2] **Cross-Screen Engagement : Multi-screen pathways reveal new opportunities for marketers to reach and engage consumers**, [http://www.ux-lady.com/resources/Cross\\_ScreenWhitepaper-microsoft.pdf](http://www.ux-lady.com/resources/Cross_ScreenWhitepaper-microsoft.pdf), alındığı tarih: 11.04.2015.
- [3] **Haitsma, J. ve Kalker, T.**, 2002. A Highly Robust Audio Fingerprinting System, Proceedings of ISMIR.
- [4] **Grosche, P., Serrà, J., Müller, M. ve Arcos, J.L.**, 2012. Structure-Based Audio Fingerprinting for Music Retrieval, Proceedings of the 13th International Society for Music Information Retrieval Conference.
- [5] **Kirovski, D. ve Attias, H.**, 2002. Beat-ID: identifying music via beat analysis, Multimedia Signal Processing, 2002 IEEE Workshop.
- [6] **Gracenote**, <http://www.gracenote.com/>, alındığı tarih: 12.04.2015.
- [7] **Gracenote News**, [http://www.gracenote.com/company\\_info/press/2005/2005083000/](http://www.gracenote.com/company_info/press/2005/2005083000/), alındığı tarih: 12.04.2015.
- [8] **Ke, Y., Hoiem, D. ve Sukthankar, R.**, 2005. Computer Vision for Music Identification, Proceedings of the 2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'05) - Volume 1 - Volume 01, Washington, DC, USA.
- [9] **MusicBrainz**, <https://musicbrainz.org/>, alındığı tarih: 12.04.2015.
- [10] **Chromaprint**, <https://oxygen.sk/2011/01/how-does-chromaprint-work/>, alındığı tarih: 12.04.2015.
- [11] **Baluja, S. ve Covell, M.**, 2008. Waveprint: Efficient Wavelet-based Audio Fingerprinting, *Pattern Recogn.*, **41(11)**, 3467–3480, <http://dx.doi.org/10.1016/j.patcog.2008.05.006>.
- [12] **Ellis, D.P., Whitman, B. ve Porter, A.**, 2011. Echoprint: An open music identification service, *ISMIR 2011 Miami: 12th International Society for Music Information Retrieval Conference, October 24-28*.

- [13] **The Echonest** , <http://the.echonest.com/>, alındığı tarih: 12.04.2015.
- [14] **Burges, C., Platt, J. ve Jana, S.**, 2003. Distortion Discriminant Analysis for Audio Fingerprinting, *IEEE Transactions on Speech and Audio Processing*, **11(3)**, 165–174, <http://research.microsoft.com/apps/pubs/default.aspx?id=67120>.
- [15] **Wang, A.**, 2003. An Industrial Strength Audio Search Algorithm, International Symposium/Conference on Music Information Retrieval.
- [16] **Shazam**, <http://www.shazam.com/>, alındığı tarih: 12.04.2015.
- [17] **Lee, S. ve Seo, J.S.**, 2007. A TV Commercial Monitoring System Using Audio Fingerprinting, Proceedings of the 6th International Conference on Entertainment Computing, Shanghai, China.
- [18] **Zhao, D., Wang, X., Qian, Y., Liu, Q. ve Lin, S.**, 2008. Fast commercial detection based on audio retrieval, Multimedia and Expo, 2008 IEEE International Conference on.
- [19] **Song, Y.**, 2012. AN AUTOMATIC COMMERCIAL SEARCH APPLICATION FOR TV BROADCASTING USING AUDIO FINGERPRINTING, *Doktora Tezi*, Texas University.
- [20] **quan Ouyang, J., Nie, H., Zhang, M., li, Z. ve Li, Y.**, 2011. Fusing audio-visual fingerprint to detect {TV} commercial advertisement, *Computers and Electrical Engineering*, **37(6)**, 991 – 1008, <http://www.sciencedirect.com/science/article/pii/S0045790611001194>.
- [21] **Saracoglu, A., Esen, E., Ates, T.K., Acar, B.O., Zubari, U., Ozan, E.C., Ozalp, E., Alatan, A.A. ve Ciloglu, T.**, 2009. Content Based Copy Detection with Coarse Audio-Visual Fingerprints., CBMI.
- [22] **Haitsma, J., Kalker, T. ve Oostveen, J.**, 2001. Robust audio hashing for content identification, *International Workshop on Content-Based Multimedia Indexing*, **4**, 117–124.
- [23] **Casey, M., Veltkamp, R., Goto, M., Leman, M., Rhodes, C. ve Slaney, M.**, 2008. Content-Based Music Information Retrieval: Current Directions and Future Challenges, *Proceedings of the IEEE*, **96(4)**, 668–696.
- [24] **Mel Frequency Cepstral Coefficient (MFCC) tutorial**, <http://practicalcryptography.com>, alındığı tarih: 12.04.2015.
- [25] **Cano, P., Batlle, E., Gómez, E., Gomes, R.D.C.T. ve Bonnet, M.**, 2005. Audio Fingerprinting Concepts and Applications, Springer-Verlag.
- [26] **Kekre, H.B., Bhandari, N., Nair, N., Padmanabhan, P. ve Bhandari, S.** A Review of Audio Fingerprinting and Comparison of Algorithms, *International Journal of Computer Applications*.



- [27] **Grosche, P., Serrà, J., Müller, M. ve Arcos, J.L.**, 2012. Structure-based audio fingerprinting for music retrieval, Int. Soc. for Music Information Retrieval Conf. (ISMIR), Porto, Portugal.
- [28] **Ra's, Z. ve A.Wieczorkowska, A.**, 2010. Advances in Music Information Retrieval, Springer, Berlin, 4. sürüm.
- [29] **Libsndfile**, <http://www.mega-nerd.com/libsndfile/>, alındığı tarih: 14.04.2015.
- [30] **FFTW**, <http://www.fftw.org/>, alındığı tarih: 14.04.2015.
- [31] **Marsyas GTZAN Genre Collection**, [http://marsyas.info/download/data\\_sets/](http://marsyas.info/download/data_sets/), alındığı tarih: 14.04.2015.
- [32] **Turkcell İletişim Hizmetleri Reklamları**, [https://www.youtube.com/playlist?list=PLhhDHGY-cvwdmunHFicQAQwcGCIVXyx\\_U](https://www.youtube.com/playlist?list=PLhhDHGY-cvwdmunHFicQAQwcGCIVXyx_U), alındığı tarih: 14.04.2015.
- [33] **Mauch, M. ve Ewert, S.**, 2013. The Audio Degradation Toolbox and its Application to Robustness Evaluation, Proceedings of the 14th International Society for Music Information Retrieval Conference (ISMIR 2013).



## ÖZGEÇMİŞ



**Ad Soyad: Hüseyin Çabuk**

**Doğum Yeri ve Tarihi: İstanbul, 08.06.1980**

**Adres: İstinye Boğaziçi Siteleri C Bloklar 3.Blok Daire:7 İstinye/İstanbul**

**E-Posta :husmanc@gmail.com**

**Lisans: Yıldız Teknik Üniversitesi - Bilgisayar Mühendisliği**

**Mesleki Deneyim ve Ödüller: 2003-2005 : AEC Mühendislik ve Bilişim Hizmetleri (Yazılım Geliştirici), 2005-2015 : Artı Teknoloji (Yazılım Geliştirici, Yazılım Ekibi Lideri)**

**Yayın ve Patent Listesi: Çabuk, H., Yüksel, Ç., Mocan, Z., Diri, B., Amasyalı, M. F., 2003 : Metin Analizi Ve Sorgulama (MAvS). 11. Sinyal İşleme ve İletişim Uygulamaları Kurultayı, Haziran 18-20, 2003 Koç Üniversitesi, İstanbul, Türkiye.**

### TEZDEN TÜRETİLEN YAYINLAR/SUNUMLAR

- **Çabuk H., İnce G., 2015: Ses Parmakizi Kullanılarak Reklam Tanıma. IEEE 23. Sinyal İşleme ve İletişim Uygulamaları Kurultayı, Mayıs 16-19, 2015 Malatya, Türkiye.**