

TOBB EKONOMİ VE TEKNOLOJİ ÜNİVERSİTESİ
FEN BİLİMLERİ ENSTİTÜSÜ

İNSAN SESİNİN AYIRT EDİCİ KAPASİTESİNİN İRDELENMESİ

YÜKSEK LİSANS TEZİ

Sinan Erkam TANDOĞAN

Elektrik ve Elektronik Mühendisliği Anabilim Dalı

Tez Danışmanı: Prof. Dr. Bülent TAVLI

AĞUSTOS 2018

Fen Bilimleri Enstitüsü Onayı

.....
Prof. Dr. Osman EROĞUL
Müdür

Bu tezin Yüksek Lisans derecesinin tüm gereksinimlerini sağladığını onaylarım.

.....
Doç. Dr. Tolga GİRİCİ
Anabilimdalı Başkanı

TOBB ETÜ, Fen Bilimleri Enstitüsü'nün 151211049 numaralı Yüksek Lisans Öğrencisi **Sinan Erkam TANDOĞAN**'ın ilgili yönetmeliklerin belirlediği gerekli tüm şartları yerine getirdikten sonra hazırladığı "**İNSAN SESİNİN AYIRT EDİCİ KAPASİTESİNİN İRDELENMESİ**" başlıklı tezi **22.03.2018** tarihinde aşağıda imzaları olan jüri tarafından kabul edilmiştir.

Tez Danışmanı : **Prof. Dr. Bülent TAVLI**
TOBB Ekonomi ve Teknoloji Üniversitesi

Eş Danışman : **Prof. Dr. Hüsrev Taha SENCAR**
TOBB Ekonomi ve Teknoloji Üniversitesi

Jüri Üyeleri : **Prof. Dr. İsmail AVCIBAŞ(Başkan)**
Üsküdar Üniversitesi

Doç. Dr. İmam Şamil YETİK
TOBB Ekonomi ve Teknoloji Üniversitesi

Dr. Öğr. Üyesi Hüseyin Uğur YILDIZ
TED Üniversitesi

TEZ BİLDİRİMİ

Tez içindeki bütün bilgilerin etik davranış ve akademik kurallar çerçevesinde elde edilerek sunulduğunu, alıntı yapılan kaynaklara eksiksiz atıf yapıldığını, referansların tam olarak belirtildiğini ve ayrıca bu tezin TOBB ETÜ Fen Bilimleri Enstitüsü tez yazım kurallarına uygun olarak hazırlandığını bildiririm.

Sinan Erkam TANDOĞAN

ÖZET

Yüksek Lisans Tezi

İNSAN SESİNİN AYIRT EDİCİ KAPASİTESİNİN İRDELENMESİ

Sinan Erkam TANDOĞAN

TOBB Ekonomi ve Teknoloji Üniversitesi
Fen Bilimleri Enstitüsü
Elektrik ve Elektronik Mühendisliği Anabilim Dalı

Danışman: Prof. Dr. Bülent TAVLI

Tarih: Ağustos 2018

Biyometrik tabanlı kimlik doğrulama sistemleri yaygın olarak parolalar yerine kullanılmaya başlamıştır. Bir mikrofon kullanılarak kolayca elde edilebileceği için ses biyometrisi tüm biyometriler arasında daha popülerdir. Ses biyometrisinin kullanımı gün geçtikçe artmasına rağmen konuşmacı doğrulama sistemlerinin kapasitesi ile ilgili çalışmalar sınırlıdır. Hatta bu alandaki çalışma sonuçları birbirleri ile çelişerek bu konudaki problemleri çözmek yerine konuşmacı sistemlerine olan güvenin azalmasına sebep olmaktadır.

Bu nedenlerden ötürü, bu tezde, ses tabanlı kimlik doğrulama sistemlerinin diğer bir deyişle konuşmacı doğrulama sistemlerinin kapasiteleri entropi açısından araştırılmıştır. Bu konu üç temel başlık altında incelenmiştir. İlk olarak biyometrik tabanlı sistemler için şimdiye kadar önerilen yöntemler detaylı bir şekilde incelenmiş ve bu yöntemlerin ses tabanlı kimlik doğrulama sistemlerine uygun olup olmadığı da araştırılmıştır. İkinci olarak konuşmacı doğrulama sistemlerinde kullanılan en gelişmiş yöntemlerden bahsedilmiştir. Konuşmalardan çıkartılan özellikler, bu özellikleri temsil etmek için kullanılan modeller ve bu modellerde kullanılan ses

tabanlı kimlik doğrulama yöntemleri ayrı ayrı incelenmiştir. Son olarak kullanılan veri kümelerinin kişi ve süre gibi kısıtlarından dolayı açık kaynaklar kullanılarak 20000'den fazla kişiden oluşan veri kümesi oluşturulmuştur.

Kapasiteyi ölçmek için en gelişmiş konuşmacı doğrulama sistemi ile uyumlu yeni bir yaklaşım önerilmiş ve bu yaklaşımın matematiksel alt yapısı detaylı bir şekilde açıklanmıştır. Bu yaklaşım farklı durumlarda farklı veri kümeleri kullanılarak incelenmiştir. Son olarak kapasite tahmini ile ilgili yeni araştırma konularından bahsedilmiştir.

Anahtar Kelimeler: Konuşmacı doğrulama, İ-vektör, Entropi, Karşılıklı bilgi ölçütü, Biyometrik bilgi.



ABSTRACT

Master of Science

EXAMINATION OF DISTINCTIVE CAPACITY OF HUMAN VOICE

Sinan Erkam TANDOĞAN

TOBB University of Economics and Technology
Institute of Natural and Applied Sciences
Electrical and Electronics Engineering Science Programme

Supervisor: Prof. Dr. Bülent TAVLI

Date: August 2018

Biometric-based authentication systems have been begun to be widely used instead of passwords. Because voice can be captured easily by using a microphone, voice is more popular between all biometric modalities. Although the use of voice biometrics is increasing day by day, the studies about capacity of speaker verification systems are limited. Moreover, the results of these studies conflict with each other and which in turn raise doubts reliability of speaker verification systems instead of answering questions.

Because of these reasons, in this thesis, the capacity of voice-based authentication systems, in other words, speaker verification systems, is investigated in terms of entropy. The subject has been examined under three main headings. Firstly, proposed approaches up to now for measuring capacity of biometric systems are examined in detail and whether these approaches are suitable for voice-based authentication systems or not was also investigated. Secondly, state-of-the-art methods used in speaker verification systems are overviewed. The features extracted from the speeches, the models used for representation of the features, and voice-based authentication methods for these models are examined separately. Thirdly, because the dataset used in speaker verification systems contains limited number of speakers

and speeches, by using open sources a new dataset containing more than 20000 speakers is created.

A new approach suitable with state-of-the-art speaker verification system is proposed for measuring capacity and the mathematical background of this approach is explained in detail. This approach is examined in different cases by using different datasets. Finally, new research topics on capacity estimation are mentioned.

Keywords: Speaker verification, I-vector, Entropy, Mutual entropy, Biometric information.



TEŐEKKÜR

Yüksek lisans süresi boyunca beni yönlendiren, değerli vaktini ve emeğini esirgemeyen Prof. Dr. Bülent TAVLI ve Doç. Dr. Hüsrev Taha SENCAR ile çalışma imkânına sahip oldum. Öncelikle tez danışmanım Prof. Dr. Bülent TAVLI'ya ve eş danışmanım Doç. Dr. Hüsrev Taha SENCAR'a sonsuz teşekkürlerimi sunarım.

Yüksek lisans eğitimi boyunca desteğini eksik etmeyen başta ailem ve Yusuf Abdulaziz YILMAZ olmak üzere, araştırmamda yardımlarını eksik etmeyen laboratuvar arkadaşlarıma, alfabetik sıraya göre, Ekrem Talha SELAMET, Enes ALTINIŐIK ve Hamdi Alperen ÇETİN'e teşekkür etmek istiyorum.

Son olarak TOBB Ekonomi ve Teknoloji Üniversitesi Elektrik-Elektronik Mühendisliđi Bölümü öğretim üyelerine, yüksek lisans süresi boyunca burs sağladığı için TOBB Ekonomi ve Teknoloji Üniversitesi'ne ve BİDEB 2228-A kapsamı ile kısmi burs veren TÜBİTAK'a teşekkür ediyorum.

İÇİNDEKİLER

Sayfa

TEZ BİLDİRİMİ	iii
ÖZET	iv
ABSTRACT	vi
TEŞEKKÜR	viii
İÇİNDEKİLER	ix
ŞEKİL LİSTESİ	x
ÇİZELGE LİSTESİ	xi
KISALTMALAR	xii
SEMBOL LİSTESİ	xiii
1. GİRİŞ	1
1.1 Tezin Amacı	1
1.2 Tez Organizasyonu	2
2. LİTERATÜR ÖZETİ	3
2.1 Biyometrik Tabanlı Sistemlerde Kapasite	3
2.1.1 Hamming mesafesinin istatistiği	3
2.1.2 Biyometrik veriler arasında bağıl entropi	5
2.1.3 Biyometrik tabanlı kimlik doğrulama sistem sonuçları arasında bağıl entropi	6
2.2 Ses Tabanlı Kimlik Doğrulama.....	7
3. VERİ KÜMESİ VE KARŞILIKLI BİLGİ ÖLÇÜTÜ	13
3.1 Veri Kümeleri	13
3.2 Karşılıklı Bilgi Ölçütü Yöntemi	22
3.2.1 Nicemleme	24
4. KAPASİTE SONUÇLARI	27
4.1 Karşılıklı Bilgi Ölçütü Yöntemi.....	27
4.2 Seste Karşılık Bilgi Ölçütü	28
4.3 Farklı Kapasite Ölçüm Yöntemleri	31
4.4 Çeşitliliğin Arttırılması	31
5. SONUÇ ve GELECEK ÇALIŞMALAR	35
KAYNAKLAR	37
ÖZGEÇMİŞ	39

ŞEKİL LİSTESİ

Sayfa

Şekil 2.1	: Kişilerin kimliklerini oluşturma diyagramı	7
Şekil 2.2	: Kimlik doğrulama diyagramı.....	7
Şekil 2.3	: MFCC çıkartılması	8
Şekil 2.4	: İki kişinin i-vektörlerinin ilk 30 boyutu.....	11
Şekil 2.5	: İki kişinin i-vektör modelinin ilk 30 boyutu.....	11
Şekil 3.1	: TED veri kümesinde videolardan ses dosyalarının elde edilmesi	14
Şekil 3.2	: TED veri kümesindeki seslerin bir kısmının ekran görüntüsü	14
Şekil 3.3	: TED veri kümesindeki altyazıların bir kısmının ekran görüntüsü.....	15
Şekil 3.4	: TED veri kümesinde hizalama aşaması çıktısı örneği	15
Şekil 3.5	: TEDx Talks veri kümesinde videolardan ses dosyalarının elde edilmesi.....	16
Şekil 3.6	: TEDx Talks veri kümesindeki seslerin bir kısmının ekran görüntüsü....	17
Şekil 3.7	: TEDx Talks veri kümesindeki altyazıların bir kısmının ekran görüntüsü.....	18
Şekil 3.8	: TEDx Talks veri kümesinde hizalama aşaması çıktısı örneği	18
Şekil 3.9	: Filmlerden ses dosyalarının elde edilmesi.....	19
Şekil 3.10	: Bir film için hizalama aşaması çıktısı örneği.....	20
Şekil 3.11	: Bir film için kelimelerin birleştirilmesi aşaması çıktısı örneği	21
Şekil 3.12	: Karşılıklı bilgi ölçütünün Venn şeması	22
Şekil 3.13	: Nicemlenmemiş i-vektörlerin ilk 20 boyutu.....	25
Şekil 3.14	: İ-vektörlerin 2 seviye ile nicemlenmesi.....	25
Şekil 3.15	: İ-vektörlerin 32 seviye ile nicemlenmesi.....	26

ÇİZELGE LİSTESİ

Sayfa

Çizelge 4.1 : Nicemleme seviyesinin hata oranına etkisi.....	28
Çizelge 4.2 : TED veri kümesinde nicemleme seviyesi ile karşılıklı bilgi ölçütü ilişkisi	29
Çizelge 4.3 : TED veri kümesinde kişi başı örnek sayısının karşılıklı bilgi ölçütüne etkisi	30
Çizelge 4.4 : TEDx Talks veri kümesinde nicemleme seviyesi ile karşılıklı bilgi ölçütü ilişkisi	30
Çizelge 4.5 : TEDx Talks veri kümesinde kişi başı örnek sayısının karşılıklı bilgi ölçütüne etkisi.....	31
Çizelge 4.6 : TED veri kümesinde diğer kapasite ölçüm yöntemleri	31
Çizelge 4.7 : Ses efektleri ve parametreler.....	32
Çizelge 4.8 : TED veri kümesindeki seslerin efekt ile değiştirilmesi sonucu karşılıklı bilgi ölçütü.....	32
Çizelge 4.9 : Filmlerden elde edilen ses dosyalarında karşılıklı ölçütü.....	33

KISALTMALAR

AKD	: Ayrık Kosinüs Dönüşümü
JFA	: Ortak Faktör Analizi (Joint Factor Analysis)
G-PLDA	: Gauss Olasılıksal Doğrusal Ayırıcı Analizi (Gaussian Probabilistic Linear Discriminant Analysis)
GMM	: Gauss Karışım Model (Gauss Mixture Model)
HD	: Hamming Mesafesi (Hamming Distance)
i-vektör	: Kimlik Vektörü (Identity Vector (i-vector))
MFCC	: Mel Frekansı Kepstral Katsayıları (Mel Frequency Cepstral Coefficients)
UBM	: Evrensel Arkaplan Modeli (Universal Background Model)

SEMBOL LİSTESİ

Bu çalışmada kullanılmış olan simgeler açıklamaları ile birlikte aşağıda sunulmuştur.

Simgeler	Açıklama
$EtkinAlan_i$	i 'nci kişinin 2048 bitlik iris verisinin etkin bitleri
$EtkinAlan_j$	j 'nci kişinin 2048 bitlik iris verisinin etkin bitleri
U_k	Kişi kümesindeki k 'ninci kişi
U_x	Kişi kümesindeki rastgele bir kişi
a_j	j 'nci boyuttaki biyometrik veri
i_{model}	Kişinin i -vektör modeli
$iris_i$	i 'nci kişinin 2048 bitlik iris verisi
$iris_j$	j 'nci kişinin 2048 bitlik iris verisi
i_{test}	Doğrulaması gerçekleştirilecek test i -vektörü
u_{UBM}	Kanal ve konuşmacı bağımsız GMM vektörü
σ^2	HD değerlerinin varyansı
z	Konuşmacıya bağlı özdeğer vektörü
Σ	GMM'deki gauss dağılımlarının kovaryans matrisi
A	k -boyutlu biyometrik veri
B	Özkanal matrisi
D	Köşegen artık matrisi
$D(p(x) q(x))$	$p(x)$ ile $q(x)$ arasındaki bağıl entropi
$H(U_x)$	Bir kişinin entropisini
$H(U_x A)$	Biyometrik veri bilindiğinde olabilecek kişilerin entropisi
$H(A)$	Biyometrik verinin entropisi
$H(A U_x)$	Kişi bilindiğinde biyometrik verinin entropisi
$H(X)$	X rastgele değişkeninin entropisi
$I(U_x; A)$	Karşılıklı bilgi ölçütü
K	HD değerlerinin dağılımının binom derecesi
M	Bir ses parçasından çıkartılan GMM vektörü
$P(U_n)$	Her kişinin olasılığı
$P(x)$	X rastgele değişkeninin olasılık kütle fonksiyonu
T	Değişken matrisi
U	Kişi kümesi
V	Özsas matrisi
i	i -vektör
m	Kanal ve konuşmacı bağımsız GMM vektörü
p	HD değerlerinin beklenen değeri
$p(x)$	X biyometrisinin kişi içindeki dağılımı

$q(x)$	X biyometrisinin tüm kişiler içindeki dağılımı
w	GMM'deki gauss dağılımlarının ağırlık vektörü
x	Kanala bağlı özdeğerler vektörü
y	Konuşmacıya bağlı özdeğer vektörü
θ	Eşik değeri
λ	Gauss dağılım kümesi
μ	GMM'deki gauss dağılımlarının beklenen değer vektörü



1. GİRİŞ

Günümüzde kullanılan birçok sistemde, kullanıcıların sisteme giriş yetkisinin olup olmadığı kimlik doğrulama yöntemleri ile kontrol edilmektedir. Bu kimlik doğrulama işleminde en yaygın kullanılan yöntem parolalardır. Parolaların seçilebileceği örnek uzayının büyüklüğü artıkça seçilen parolanın tahmin edilebilirliği azalarak sistemin güvenliği artar. Sistemin güvenliğini örnek uzayının büyüklüğü haricinde parolaların örnek uzayda nasıl bir dağılıma sahip oldukları da önemli bir şekilde etkiler. Örneğin 4-rakamlık parolaların örnek uzayı 13.3 bitlik bir büyüklüğe sahip olması gerekirken kullanıcıların parola seçme alışkanlıkları göz önüne alındığında bu değer 8.4 bite düşmektedir [1]. Yine aynı çalışmada 6-rakamlık parolaların örnek uzayının büyüklüğü 19.9 bitten 13.2 bite düşmektedir. Sonuç olarak parolaları daha uzun seçmek örnek uzayın büyüklüğünü arttırsa da kullanıcıların parola seçme alışkanlıkları dikkate alındığında örnek uzayın büyüklüğü beklenilenden daha küçük olmaktadır. Örnek uzayın etkin büyüklüğünün azalmaması için kullanıcılara rastgele parolalar vermek ise kullanıcıların parolalarını hatırlamaları konusunda problem yaşamalarına neden olmaktadır. Bu sebeple parolalar ile kimlik doğrulama yerine biyometrik tabanlı kimlik doğrulama uygulamaları kullanılmaya başlanmış olup gün geçtikçe daha yaygın bir şekilde kullanılmaktadır. Biyometrik tabanlı kimlik doğrulama sistemlerinin getirdiği kolaylığa rağmen sağladığı güvenliği belirlemek için bu sistemlerin örnek uzayının büyüklüğü ve bu biyometrilerin kullanıcıya özgü olup olmadığı detaylı bir şekilde incelenmelidir.

1.1 Tezin Amacı

Diğer biyometriklere göre kullanımı daha kolay olan ses tabanlı kimlik doğrulama sistemlerinin kullanımı gün geçtikçe artmaktadır. Bu sistemlerin güvenlik kapasiteleri örnek uzayın etkin büyüklüğü ile direkt bağlantılıdır. Bu çalışmada insan sesinin ayırt edici kapasitesinin detaylı bir şekilde irdelenmesi amaçlanmıştır. Bu

alıřma kapsamında insan sesinin rnek uzayının etkin byklė llmeye alıřılmıřtır.

1.2 Tez Organizasyonu

Bu tez alıřmasının ikinci blmnde literatrde biyometrik tabanlı kimlik doėrulama sistemlerinde rnek uzayın etkin byklėn lmek iin nerilen yntemler ve ses tabanlı kimlik doėrulama sistemlerinde sesin nasıl iřlendiėi detaylı bir řekilde anlatılmıřtır. nc blmde, ses tabanlı kimlik doėrulama sistemlerinde kullanılan veri kmeleri, tez kapsamında oluřturulan veri kmeleri, kapasitenin hesaplanması iin nerilen karřılıklı bilgi lt yntemi ve bu yntemin matematiksel modeli detaylı bir řekilde anlatılmıřtır. Drdnc blmde, karřılıklı bilgi ltn hesaplayabilmek iin kiřilerin temsilinde kullanılan zniteliklerin nicemlenmesinin hata oranına etkisi ve farklı durumlardaki karřılıklı bilgi lt deėerleri arařtırılmıřtır. Beřinci blmde ise bu alıřma ile elde edilen sonular ve gelecek alıřmalardan bahsedilmiřtir.

2. LİTERATÜR ÖZETİ

2.1 Biyometrik Tabanlı Sistemlerde Kapasite

Literatürde biyometrik tabanlı kimlik doğrulama sistemlerinin örnek uzayını hesaplamaya çalışan temelde 3 farklı yaklaşım vardır. Bu kısımda bu yaklaşımların çalışma prensipleri ve çalışma koşulları incelenmiştir.

2.1.1 Hamming mesafesinin istatistiği

İris tabanlı kimlik doğrulama sistemlerinin kapasitesini ölçmek için iris verisinin istatistiksel özellikleri kullanılmıştır [2], [3]. Bu yöntemde, kişilerin iris resimlerinden çıkartılan 2048 bitlik veriler arasındaki farklılık miktarının dağılımı incelenmiştir. 2048 bitlik veriler arasındaki farklılık miktarını ölçerken iki kişinin 2048 bitlik iris verileri arasında Hamming mesafesi (HD-Hamming Distance) hesaplanmıştır. Kişinin iris resimlerinde her zaman kişinin irisinin tamamı bulunmamaktadır. Bu nedenle, bazı iris resimlerinde 2048 bit yerine daha az veri çıkarılmaktadır. 2048 bit yerine daha az bitlik verinin çıkarılmasından dolayı olabilecek karışıklıkları engellemek için elde edilen farklılık miktarı etkin iris verisinin büyüklüğü ile normalize edilmektedir. Eğer iris resminin kullanılamaz kısımları varsa etkisiz eleman kullanılarak her durumda 2048 bitlik verinin elde edilmesi sağlanmıştır. İki iris verisi arasındaki HD ifadesi Eşitlik (2.1)'deki gibi hesaplanmıştır.

$$HD = \frac{\text{Toplam}(XOR(iris_i, iris_j))}{\text{Toplam}(AND(EtkinAlan_i, EtkinAlan_j))} \quad (2.1)$$

Eşitlik (2.1)'de $iris_i$, $iris_j$ ifadeleri sırayla i 'nci ve j 'nci kişilerin 2048 bitlik iris verisini, $EtkinAlan_i$ ve $EtkinAlan_j$ ifadeleri ise i 'nci ve j 'nci kişilerin 2048 bitlik iris verilerinin hangilerinin resimden çıkarılabildiğini içeren vektörlerdir. İki iris

verisi arasındaki farklılığı bulmak için mantık kapılarından XOR kapısı kullanılmakta ve toplanarak kaç tane bitin birbirinden farklı olduğu bulunmaktadır. Etkin alanların kesişim yerlerini bulmak için mantık kapılarından AND kapısı kullanılmakta ve toplanarak kaç bitin etkin değer aldığı bulunmaktadır. İki ifadenin oranı alınarak HD ifadesi normalize edilmiştir.

Kişilerin iris resimlerinden çıkarılan 2048 bitlik veriler kullanılarak hesaplanan HD'lerin dağılımının varyans ve beklenen değerini sırayla Eşitlik (2.2)'deki σ^2 ve p yerine kullanılarak bu dağılımın kaçınıcı dereceden binom dağılımı ile aynı özelliklere sahip olduğu hesaplanır.

$$K = p(1 - p)/\sigma^2 \quad (2.2)$$

Hesaplanan K değeri 2048 bitlik veri içindeki bağımsız eleman sayısına denk gelmektedir. Diğer bir ifade ile bu yöntemle 2048 bitlik veri içindeki bağımlı elemanlar elenerek, bağımsız eleman sayısı bulunmaktadır. Bu yöntem kullanılarak iris tabanlı kimlik doğrulama sisteminin kapasitesi 249 bit ve 218 bit olarak ölçülmüştür [2], [3].

Bu yöntem ile hesaplanan değer sistemin kapasitesi olarak kullanılabilmesi için 2048 bitlik verideki her bir bit eşit olasılıklı Bernoulli dağılımına sahip olmalıdır. Ayrıca, 2048 bitlik iris verisindeki her bit eşit öneme sahip olmalıdır. Bir biyometriden çıkartılan biyometrik verinin ilk elemanı ile son elemanı kişinin kimliği hakkında eşit bilgi veriyorsa bu yöntem o biyometrik veri için kullanılabilir. Ama eğer biyometrik verinin elemanları arasında bir önem sırası varsa bu yöntem kullanılamaz. Çünkü HD hesaplanırken biyometrik verinin her elemanının HD'ye etkisi birbirine eşittir.

Normalde biyometrik verilerin kişinin farklı örneklerinde değişmemesi istense bile çoğu biyometrikte kişinin farklı örneklerinden elde edilen biyometrik veri değişiklik göstermektedir. Bu yöntemin eksik özelliklerinden biri ise biyometrik verinin kişi içinde değişmesi göz önüne alınmamıştır. Kişi içi değişim olmadığı kabul edilmiştir.

Yöntemin diğer önemli eksik özelliklerinden biri ise verilerin ikili kodlama ile kodlanması gerekliliğidir. İkili kodlama yerine kullanılacak diğer kodlama

yöntemlerinde iki veri arasındaki farklılık miktarı XOR kapısı kullanılarak hesaplanamaz.

2.1.2 Biyometrik veriler arasında bağıl entropi

İkinci temel yaklaşımda, herhangi bir kişinin biyometrik verisinin dağılımı ile biyometrik verinin tüm kişilerdeki dağılımı arasındaki farklılık ölçülmektedir. Bu farklılık ölçülürken bağıl entropi ifadesi hesaplanarak kullanılmaktadır. Bağıl entropi ifadesi Eşitlik (2.3)'teki gibi hesaplanmaktadır.

$$D(p(x)||q(x)) = \int p(x) \log_2 \frac{p(x)}{q(x)} \quad (2.3)$$

Eşitlik (2.3)'teki $p(x)$, $q(x)$ ve $D(p(x)||q(x))$ ifadeleri sırayla biyometrik verinin kişi içindeki dağılımını, biyometrik verinin tüm kişiler içinde dağılımını ve biyometrik verinin kişi içi dağılımı ile tüm kişilerdeki dağılımı arasındaki bağıl entropiyi temsil etmektedir. Bu yöntem kullanılarak yapılan yüz ve iris tabanlı kimlik doğrulama sistemlerinin kapasitesi sırayla 45 bit ve 278 bit olarak ölçülmüştür [5], [4].

Bu yöntemin doğru bir şekilde uygulanabilmesi için öncelikle biyometrik verinin dağılımı doğru bir şekilde tahmin edilmelidir. Özellikle bir kişinin biyometrik verisinin beklenen değerden uzak değerlerdeki dağılımlarının doğru olarak tahmin edilebilmesi için kişiden çok fazla sayıda örnek alınması gerekmektedir. Kişinin biyometrik verisinin dağılımının kuyruk kısmında yapılabilecek küçük hatalar, bağıl entropi formülünü orantılı yapısından dolayı bağıl entropi değerini çok etkileyebilir. Biyometrik verilerin hem tüm kişilerde hem de kişinin farklı örneklerinde Gauss dağılımına sahip olduğu kabul edilmektedir ve biyometrik veriler kullanılarak bu Gauss dağılımlarının varyans ve beklenen değeri hesaplanır. Bazı biyometrik verilerin dağılımı tek bir Gauss dağılımı yerine farklı dağılımlarla daha iyi modellenebilmektedir.

Shannon tarafından yapılan entropi tanımı [6] ayrık değişkenler için geçerlidir ve Eşitlik (2.4)'teki denklemlerle hesaplanmaktadır.

$$H(X) = - \sum_x P(x) \log_2 P(x) \quad (2.4)$$

Eşitlik (2.4)'teki $H(X)$, X değişkeninin entropisini ve $P(x)$, X değişkenin olasılık kütle fonksiyonunu temsil etmektedir. Sürekli değişkenler için tanımlanan entropi formülü Eşitlik (2.4)'ün integral formudur ve X değişkeninin olasılık yoğunluk fonksiyonuna bağlı olarak negatif değerler alabilmektedir. Bu nedenle, sürekli değişkenler için hesaplanan entropi, değişkenin bilgi miktarı yerine dağılımı hakkında kısmi bilgi vermektedir. Bu yöntemde ise bağıl entropi sürekli değişkenler üzerinden hesaplanmaktadır.

2.1.3 Biyometrik tabanlı kimlik doğrulama sistem sonuçları arasında bağıl entropi

Son kapasite ölçüm yönteminde biyometrik sistemin entropisi tanımlanarak hesaplanmıştır [7]. Bir biyometrik sistemin entropisi, herhangi bir kişinin biyometrik verisinin sistemde kullanılmasıyla kişinin kimliğindeki belirsizlikteki ortalama düşüğe eşittir. Biyometrik sistemin entropisi asimptotik olarak bağıl entropi ifadesi ile hesaplanabilmektedir. Biyometrik sistemin kapasitesi hesaplanırken biyometrik sistemin entropi ifadesi kullanıldığı için bağıl entropi ifadesi direkt olarak biyometrik verilerin dağılımları arasında hesaplanmamaktadır. Kimlik ile biyometrik verinin eşleşmesi durumunda sistemin çıktısının dağılımı ile kimlik ile biyometrik verinin eşleşmemesi durumunda sistemin çıktısının dağılımı arasındaki bağıl entropi değeri kullanılarak biyometrik tabanlı kimlik doğrulama sistemlerinin kapasitesi ölçülmeye çalışılmıştır [7], [8]. Bu yöntem kullanılarak parmak izi tabanlı kimlik doğrulama sisteminin kapasitesi 12.62 bit olarak ölçülmüştür [7].

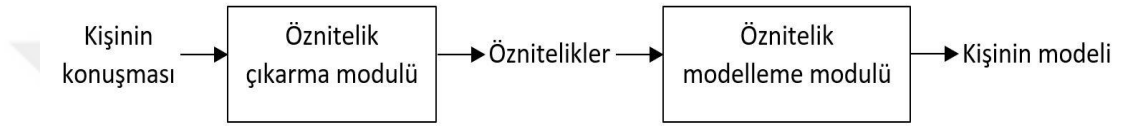
Bu yöntemde biyometrik tabanlı sistemde kullanılan doğrulama yöntemi direkt olarak hesaplanan kapasiteye etki etmektedir. Normalde daha fazla bilgi içeren biyometrik veriler sadece karşılaştırma sisteminden dolayı çok daha düşük entropi sonucu verebilir.

Bu yöntemle ilgili diğer bir konu ise karşılaştırma sonuçlarının dağılımlarının modellenmesinde az sayıda örnek kullanılması ve bu dağılımların Gauss dağılım

kullanarak modellenmesidir. Önceki yönteme benzer şekilde dağılımların kuyruk kısımlarında yapılabilecek küçük hatalar kapasitenin yanlış ölçülmesine neden olabilir.

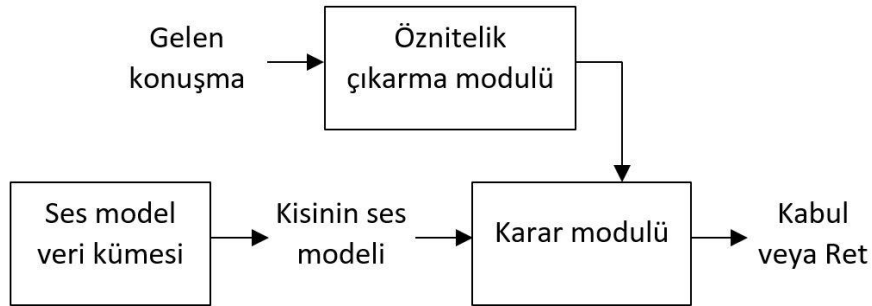
2.2 Ses Tabanlı Kimlik Doğrulama

Ses tabanlı kimlik doğrulama sistemleri iki temel basamaktan oluşmaktadır [9]. Sistemin ilk basamağında, kişilerden alınan konuşmalardan öznitelikler çıkartılır ve bu öznitelikler modellenerek kişilerin kimlikleri oluşturulur. Kimlik oluşturma diyagramı Şekil 2.1’de verilmiştir.



Şekil 2.1 : Kişilerin kimliklerini oluşturma diyagramı.

İkinci basamakta ise kişinin yeni konuşmasından çıkartılan öznitelikler kişinin kimliği ile karşılaştırılarak kişinin kimlik doğrulaması geliştirilir ve genel olarak akış diyagramı Şekil 2.2’de verilmiştir. Kimlik doğrulaması gerçekleştirilecek kişinin ses örneğinden öznitelikleri çıkarılır ve kişinin bulunduğu veri kümesindeki daha önceden oluşturulan kişinin modeli ile karşılaştırılarak karar verilir.



Şekil 2.2 : Kimlik doğrulama diyagramı.

Ses tabanlı kimlik doğrulamalarında en yaygın kullanılan öznitelik Mel Frekanslı Cepstral Katsayıları (MFCC – Mel Frequency Cepstral Coefficients)’dır. Ses sinyalinden MFCC çıkartılırken izlenen basamaklar Şekil 2.3’te verilmiştir.



Şekil 2.3 : MFCC çıkartılması.

Ses sinyali öncelikle sabit akustik karakter gösteren parçalara bölünür ve Hamming gibi pencereleme yöntemleri ile parçalara ayırmanın spektral etkisi azaltılır. Her bir parçanın frekans alanındaki değerleri hesaplanır ve mutlak değerleri alınır. Daha sonra Mel ölçekte dizilmiş üçgen süzgeç dizileri kullanılarak insan kulağının hassas olduğu frekans bileşenleri yükseltilir. Logaritması alınarak frekans alanında çarpım haline bulunan sesin izlediği yol ile boğaz yapısının etkisi birbirinden ayrılır. Son olarak Ayrık Kosinüs Dönüşümü (AKD) ile sesin MFCC değerleri elde edilir.

MFCC değerlerinin dağılımı kişiden kişiye farklılık göstermektedir. Bu nedenle, MFCC değerlerinin dağılımını temsil edilecek dağılımlar kullanılarak kişiler modellenmeye çalışılmıştır. MFCC değerlerinin dağılımı standart dağılımlardan çok farklılık göstermektedir. Ses sinyalinin MFCC değerleri hesaplandıktan sonra bu MFCC değerlerinin dağılımı Gauss Karışım Modelleri (GMM-Gauss Mixture Model) kullanarak modellenmiştir [10]. Burada kişiler Gauss dağılım kümeleri, $\lambda = \{w, \mu, \Sigma\}$ ile temsil edilmektedir. Burada w , karışımın ağırlık vektörü, μ , Gauss dağılımlarının beklenen değer vektörü ve Σ , Gauss dağılımının kovaryans matrisidir. Farklı beklenen değerlere sahip Gauss dağılımların farklı ağırlıklarla toplanması ile elde edilen GMM birçok standart olmayan dağılımın modellenmesinde etkili bir şekilde kullanılabilir. Kişinin ses örneğinden her 25-milisaniyelik parçadan 19 adet MFCC değeri hesaplanır. 25-milisaniyelik pencereleme 10 milisaniyelik kayma ile tüm ses dosyasına uygulanır. Çıkarılan MFCC'ler ve beklenti maksimizasyonu algoritması kullanılarak kişinin Gauss dağılım kümesindeki karışımın ağırlık vektörü, beklenen değer vektörü ve kovaryans matrisi tahmin edilerek kişinin GMM parametreleri elde edilir. Kişinin kimlik doğrulaması, gelen sesten çıkarılan MFCC'lerin, kişi için daha önceden oluşturulan konuşma modelindeki GMM parametreleri ile aynı dağılıma sahip olma olasılığı hesaplanarak belli bir değerden büyük olup olmamasına göre karar verilerek gerçekleştirilir.

Daha sonraki çalışmalarda MFCC'leri kimlik doğrulamasında direkt kullanmak yerine bir vektör kullanılmasının daha iyi sonuç verdiği gösterilmiştir [11]. Her ses parçasından elde edilen MFCC değerleri için GMM hesaplanır ve bu GMM'nin beklenen değerleri kullanılarak oluşturulan vektör kullanılmaktadır. Kişinin kimlik doğrulaması gerçekleştirilirken bu oluşturulan vektörler bazı makine öğrenme yöntemleri ile işlenir.

Bu ses vektörleri kullanılarak gerçekleştirilen kimlik doğrulama yaklaşımlarında, kişinin konuşmasından çıkartılan GMM vektörünün içerdiği kanal ve konuşmacı bileşenlerinin ayrı ayrı modellenmesi (bu işlem Ortak Faktör Analizi (JFA - Joint Factor Analysis) olarak adlandırılır) önerilmiştir [12]. Gelen bir sesten çıkartılan GMM vektörü, M , bu yaklaşımda Eşitlik (2.5)'teki gibi modellenir.

$$M = m + Vy + Bx + Dz \quad (2.5)$$

Burada m kanal ve konuşmacı bağımsız vektörü, B özkanal matrisi, V özses matrisi, D köşegen artık matrisi, x kanalın etkisini içeren vektörü, diğer bir deyişle kanala bağlı özdeğerleri, y ve z konuşmacının etkisini içeren vektörleri temsil etmektedir. Konuşmacı etkisini içeren vektörlerin doğrulaması için daha önceden kişinin bilinen sesleri kullanılarak elde edilen kişinin modeli ile gelen ses parçasından elde edilen konuşmacı etkisini içeren vektörlerin aynı dağılıma sahip olup olmadığı hipotez testi kullanılarak karar verilir.

Ses tabanlı kimlik doğrulama sistemlerinde en son teknoloji olarak kimlik vektörü (i -vektör - identity vector) yöntemi kullanılmaktadır [13]. Bu yöntemin JFA yöntemi ile en temel farkı, kanal bilgisinin konuşmacı bilgisi içerdiği için, konuşmacı bileşenler ile kanal bileşenlerini ayrı ayrı modellenmesinin yerine bu bileşenlerin beraber modellenmesidir.

$$M = u_{UBM} + Ti \quad (2.6)$$

Sesten elde edilen GMM vektörü M , u_{UBM} konuşmacı ve kanal bağımsız bileşen, Universal Background Model (UBM) , T değişken matrisi ve i i -vektör ile Eşitlik (2.6)'daki gibi modellenir.

Sesler kullanılarak elde edilen GMM vektörlerinde daima sabit bir bileşen bulunmaktadır. Bunun sebebi her sesin içinde konuşmacıdan ve kanaldan bağımsız bileşenlerin bulunmasıdır. Bu sabit bileşen, M ifadesinin Eşitlik (2.6)'daki gibi modellenirken, u_{UBM} ifadesi ile belirtilmektedir. Yani her ses sinyali içinde doğal bir ofset GMM vektörü bulunmaktadır.

Gelen ses parçasından elde edilen GMM vektörünün boyutu çok büyük olmasına rağmen bu vektörden elde edilen i -vektör genel olarak 200 boyutludur. Bu çok boyutlu GMM vektörü ile sesten çıkartılan 200 boyutlu i -vektörler arasındaki doğrusal geçişi sağlayan özetleme fonksiyonu (Hashing Function) T matrisi ile temsil edilmektedir. T matrisi tüm kişilerin i -vektörlerinin normal dağılıma sahip olmasını sağlamaktadır. T ve u_{UBM} sistemin eğitim aşamasında, eğitim verileri kullanılarak hesaplanır.

i -vektör yöntemi kullanılırken öncelikle bir grup insanın verisi, sistemin eğitim aşamasında kullanılır. Bu gruptaki kişilerin seslerinden çıkartılan M vektörlerinin ortalaması alınarak öncelikle u_{UBM} ifadesi hesaplanır. Sonrasında ise kişilerin ses dosyalarından çıkartılan M ile hesaplanan u_{UBM} ifadesi kullanılarak eğitim veri kümesindeki kişilerin i -vektörlerinin dağılımının normal dağılıma sahip olacak şekilde dağılması için gerekli T matrisi tahmin edilir. Yinelemeli şekilde beklenti maksimizasyonu algoritması ile minimum sapma algoritması kullanılarak T matrisinin son hali hesaplanır. T matrisi son halini aldığı anda eğitim veri kümesindeki kişilerin i -vektörlerinin dağılımı normal dağılıma sahip olacaktır.

Eğitim aşamasından sonra sistemde kimlik doğrulaması yapılacak her kişi için o kişinin ses dosyaları kullanılarak M vektörleri hesaplanır ve kişinin i -vektörleri hesaplanır. Şekil 2.4'te iki kişinin ses dosyalarından çıkarılmış ilk kişinin 10 tane, ikinci kişinin 5 tane 200 boyutlu i -vektörlerin ilk 30 boyutu verilmiştir. İlk 10 sütun bir kişiye diğer kalan 5 sütun diğer kişiye aittir. Kişinin farklı ses parçalarından çıkartılan i -vektörlerin ortalaması alınarak kişinin i -vektör modeli oluşturulur. Şekil 2.5'te Şekil 2.4'te i -vektörleri verilen iki kişinin i -vektör modellerinin ilk 30 boyutu verilmiştir.

$$\frac{\langle i_{test}, i_{model} \rangle}{\|i_{test}\| \times \|i_{model}\|} \stackrel{\geq}{\leq} \theta \quad (2.5)$$

Doğrulaması gerçekleştirilecek kişinin sesinden elde edilen test i-vektörü i_{test} , kişinin daha önceden oluşturulmuş i-vektör modeli i_{model} ve eşik değeri θ ile temsil edilmiştir.

İ-vektörlerin karşılaştırılmasında en sık kullanılan yöntem Gauss Olasılıksal Doğrusal Ayırıcı Analizi (G-PLDA - Gaussian Probabilistic Linear Discriminant Analysis) yöntemidir [14]. G-PLDA yaklaşımında bir kişinin i-vektörlerin belirli bir beklenen değer etrafında Gauss dağılımına sahip oldukları kabul edilir ve bu dağılımın parametreleri kişinin i-vektör modeli oluşturulurken kişinin farklı ses dosyalarından çıkartılan i-vektörler kullanılarak tahmin edilir. Kişinin modeli ile gelen sesten çıkartılan i-vektörün aynı dağılımdan olması veya farklı dağılımlardan olması durumları hipotez testi yöntemi kullanılarak karar verilir.

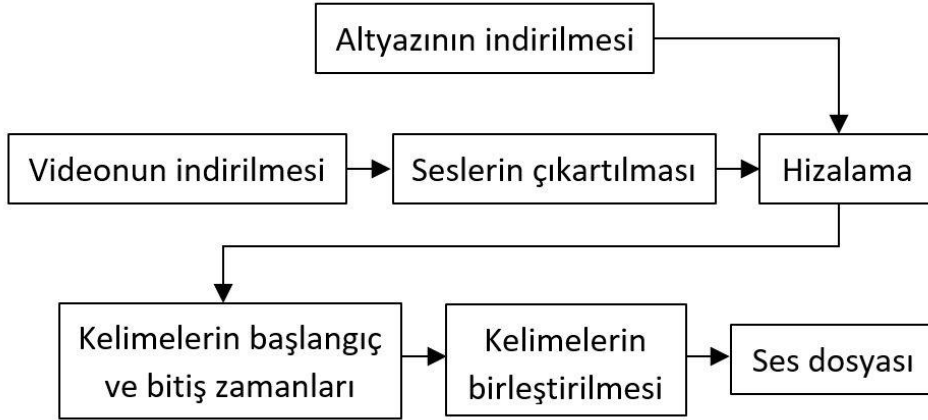
3. VERİ KÜMESİ VE KARŞILIKLI BİLGİ ÖLÇÜTÜ

3.1 Veri Kümeleri

Ses işleme alanında birçok veri kümesi olmasına rağmen bunların sadece birkaç tanesi ses tabanlı kimlik doğrulama alanında kullanılabilir. Bu veri kümelerinden en çok kullanılanları NIST SRE 2004-2010, TIMIT ve YOHO'dur. Bu veri kümelerinden NIST SRE 2004-2010'da 976, TIMIT'te 630 ve YOHO'da 138 kişinin ses dosyaları bulunmaktadır. Bu veri kümelerindeki sesler genellikle telefon konuşmalarından veya sessiz odalarda kaliteli mikrofonlar kullanılarak konuşmacılara rastgele cümleler söylenilerek kaydedilmesi ile elde edilmiştir.

Ses tabanlı kimlik doğrulama sistemlerinin kapasitesi ölçülürken, veri kümesinde bulunan kişi sayısı ve bu kişilerden alınan ses kayıtlarının uzunluğu kilit rol oynamaktadır. Standart veri kümelerinde, kişi sayısının az olması, ses kaydının özel durumlarda gerçekleştirilmiş olması ve bir kişiden alınan ses kaydının uzunluğunun az olması gibi dezavantajlar bulunmaktadır. Bu nedenlerden dolayı, bu çalışmada kullanılmak üzere, farklı açık kaynaklar kullanarak 3 adet veri kümesi oluşturulmuştur. Bu 3 veri kümesinden ilk ikisi sırayla Youtube'daki TED [URL-1] ve TEDx Talks [URL-2] kanalları ve sonuncusu ise çeşitli filmler kullanılarak oluşturulmuştur.

Oluşturulan ilk veri kümesinde Youtube'daki TED kanalındaki videolar kullanılmıştır. Öncelikle veri kümesini oluşturmaya başladığımız tarihte kanalda bulunan 2000 civarı video ve bu videoların altyazıları indirilmiştir. Bu videoların bazıları indirme sırasında oluşan hatalar, bazıları altyazılarının mevcut olmaması ve bazıları da videonun konuşma içermemesinden dolayı veri kümesine dâhil edilmemiştir. TED kanalı kullanılarak oluşturulan veri kümesi 1914 videodan çıkartılan ses dosyalarından oluşturulmuştur. TED kanalından alınan videolardan veri kümesindeki ses dosyalarının elde edilmesi arasında yapılan işlemler Şekil 3.1'de genel hatları ile verilmiştir.



Şekil 3.1 : TED veri kümesinde videolardan ses dosyalarının elde edilmesi.

Şekil 3.1’de gösterildiği üzere, ilk olarak video ve videonun altyazısı indirilmiştir. İndirilen videoların sesleri, dönüştürücü programlar kullanılarak genellikle ‘.mp3’ formatında elde edilmiştir. İşleme alınan ilgili video seslerinin bir kısmı Şekil 3.2’de, altyazılarının bir kısmı da Şekil 3.3’te verilmiştir.

- | | | |
|--------------------------------------|-----------------------------------|-------------------------------------|
| FrankGehry_1990-320k.mp3 | MichaelMoschen_2002-320k.mp3 | BillStrickland_2002-320k.mp3 |
| DanGilbert_2005G-320k.mp3 | RichardLedgett_2014-320k.mp3 | EvelynGlennie_2003-320k.mp3 |
| LennartGreen_2005-320k.mp3 | RichardBranson_2007-320k.mp3 | StevenPetranek_2002-320k.mp3 |
| BillGates_2010-320k.mp3 | RichardDawkins_2002-320k.mp3 | PhilipRosedale_2008P-320k.mp3 |
| TimBrown_2008P-320k.mp3 | AlGore_2008-320k.mp3 | JaneGoodall_2002-320k.mp3 |
| AmoryLovins_2012S-320k.mp3 | DavidPogue_2008P-320k.mp3 | RobertFischell_2005-320k.mp3 |
| LarryBrilliant_2006-320k.mp3 | JehaneNoujaim_2006-320k.mp3 | SarahJones_2015-320k.mp3 |
| HisHolinessTheKarmapa_2009I-320k.mp3 | InwinRedlener_2008-320k.mp3 | DavidHolt_2004-320k.mp3 |
| RickSmolan_2007P-320k.mp3 | ReginaDugan_2012-320k.mp3 | DanDennett_2006-320k.mp3 |
| DavidRockwell_2002-320k.mp3 | DaveEggers_2008-320k.mp3 | VincentMoon_2014G-320k.mp3 |
| Sheenalyengar_2010G-320k.mp3 | BillClinton_2007-320k.mp3 | ThomasBarnett_2005-320k.mp3 |
| JaneGoodall_2007G-320k.mp3 | BryanStevenson_2012-320k.mp3 | VilayanurRamachandran_2007-320k.mp3 |
| HelenFisher_2006-320k.mp3 | AndrewSolomon_2013P-320k.mp3 | DonaldHoffman_2015-320k.mp3 |
| CKWilliams_2001-320k.mp3 | Sting_2014-320k.mp3 | BrianGreene_2012-320k.mp3 |
| PhilZimbardo_2008-320k.mp3 | MichaelMerzenich_2004-320k.mp3 | BruceAylward_2011-320k.mp3 |
| SamanthaPower_2008-320k.mp3 | NuclearDebate_2010-320k.mp3 | RayKurzweil_2005-320k.mp3 |
| AubreydeGrey_2005G-320k.mp3 | AmyTan_2008-320k.mp3 | StevenPinker_2003-320k.mp3 |
| JuanEnriquez_2003-320k.mp3 | SherwinNuland_2001-320k.mp3 | AllanSavory_2013-320k.mp3 |
| GaryHaugen_2015-320k.mp3 | WadeDavis_2003-320k.mp3 | FrankGehry_2002-320k.mp3 |
| StevenStrogatz_2004-320k.mp3 | JamieOliver_2010-320k.mp3 | PaulaScher_2008P-320k.mp3 |
| MaryRobinson_2015W-320k.mp3 | RicardoSemler_2014G-320k.mp3 | JamesNachtwey_2007-320k.mp3 |
| PaulSerenio_2005-320k.mp3 | AntonioDonatoNobre_2010X-320k.mp3 | EstherPerel_2015-320k.mp3 |
| KarenArmstrong_2008-320k.mp3 | MaeJemison_2002-320k.mp3 | BarrySchuler_2008P-320k.mp3 |
| DavidPogue_2006-320k.mp3 | GarrettLisi_2008-320k.mp3 | AnoteTong_2015Z-320k.mp3 |
| ShekharKapur_2009I-320k.mp3 | MartineRothblatt_2015-320k.mp3 | DeborahRhodes_2010W-320k.mp3 |
| RickWarren_2006-320k.mp3 | SethBerkley_2010-320k.mp3 | BruceSchneier_2010X-320k.mp3 |
| SarahJones_2009-320k.mp3 | BobMankoff_2013S-320k.mp3 | AmyCuddy_2012G-320k.mp3 |

Şekil 3.2 : TED veri kümesindeki seslerin bir kısmının ekran görüntüsü.

DanGilbert_2005G-320k.txt	EdwardBurtynsky_2005-320k.txt	KaryMullis_2002-320k.txt
EdwardSnowden_2014-320k.txt	BillStrickland_2002-320k.txt	PeterHirschberg_2007P-320k.txt
AndrewSolomon_2013X-320k.txt	PhilipRosedale_2008P-320k.txt	RichardBranson_2007-320k.txt
StevenPetranek_2002-320k.txt	DaveEggers_2008-320k.txt	JeffHawkins_2003-320k.txt
InvinRedlener_2008-320k.txt	BryanStevenson_2012-320k.txt	DavidPogue_2008P-320k.txt
LaurieSantos_2010G-320k.txt	JaneGoodall_2002-320k.txt	NormanFoster_2007P-320k.txt
RayKurzweil_2005-320k.txt	MichaelMerzenich_2004-320k.txt	TonyRobbins_2006-320k.txt
AmoryLovins_2012S-320k.txt	DanielKraft_2011X-320k.txt	PaulSerenio_2005-320k.txt
PeterDonnelly_2005G-320k.txt	DanielWolpert_2011G-320k.txt	ThomasBarnett_2005-320k.txt
MaeJemison_2002-320k.txt	NuclearDebate_2010-320k.txt	CharlieRose_2014-320k.txt
EthanZuckerman_2010G-320k.txt	BruceFeiler_2010P-320k.txt	DavidAgus_2009P-320k.txt
EdBoyden_2011-320k.txt	StevenPinker_2003-320k.txt	DanGilbert_2004-320k.txt
AlaindeBotton_2011G-320k.txt	JimmyWales_2005G-320k.txt	StuartBrown_2008P-320k.txt
LarryBrilliant_2006-320k.txt	JuanEnriquez_2003-320k.txt	NicholasNegroponte_1984-320k.txt
StewartBrand_2004-320k.txt	DorisKearnsGoodwin_2008-320k.txt	GarrettLisi_2008-320k.txt
HelenFisher_2006-320k.txt	SamHarris_2010-320k.txt	ThomasPiketty_2014S-320k.txt
BarrySchwartz_2010S-320k.txt	GlennGreenwald_2014G-320k.txt	ClayShirky_2005G-320k.txt
TempleGrandin_2010-320k.txt	DavidSRose_2007U-320k.txt	DeanOrnish_2004-320k.txt
StevenLevitt_2005G-320k.txt	JonathanHaidt_2012S-320k.txt	MelindaGates_2012X-320k.txt
CarneRoss_2009P-320k.txt	CameronSinclair_2006-320k.txt	SarahJones_2015-320k.txt
RishiManchanda_2014S-320k.txt	JeremyGilley_2011G-320k.txt	JacksonKatz_2012X-320k.txt
RussellFoster_2013G-320k.txt	AmyCuddy_2012G-320k.txt	RicardoSemler_2014G-320k.txt
Bono_2005-320k.txt	MorganSpurlock_2011-320k.txt	EveEnsler_2004-320k.txt
ChrisBangle_2002-320k.txt	CalebChung_2007P-320k.txt	StuartFirestein_2013-320k.txt
JeremyHoward_2014X-320k.txt	PeterWard_2008-320k.txt	BillClinton_2007-320k.txt
PaulSnelgrove_2011G-320k.txt	AnnaDeaverSmith_2005-320k.txt	DanAriely_2012X-320k.txt
AmyTan_2008-320k.txt	AnnCooper_2007P-320k.txt	NatalieJeremijenko_2009P-320k.txt

Şekil 3.3 : TED veri kümesindeki altyazıların bir kısmının ekran görüntüsü.

Hizalama aşamasında, CMU Sphinx [15] programını kullanarak altyazıda geçen kelimelerin, ses dosyasında tam olarak ne zaman geçtiği tespit edilmiştir. Altyazıdaki kelimeler açıklama olup olmadığı CMU Sphinx programına verilmeden önce belirlenerek açıklama olan kelimeler temizlenmiştir. Hizalama işleminden elde edilen örnek sonuçlar Şekil 3.4’te verilmiştir.

```

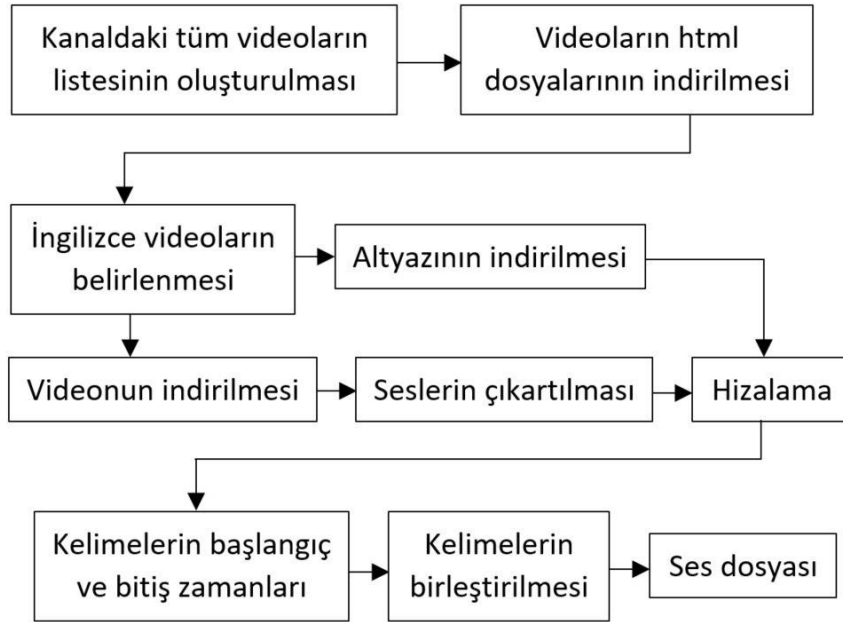
parents, 1.000, [10020:10720], -7.0104504E7, 0.0
and, 1.000, [11200:11360], -7925385.5, 0.0
one, 1.000, [11370:11610], -3.3325646E7, 0.0
teenage, 1.000, [11620:12080], -7.118608E7, 0.0
girl, 1.000, [12090:12410], -1.4291976E8, 0.0
one, 1.000, [13430:13790], -7725871.0, 0.0
evening, 1.000, [13800:14090], -6.4678124E7, 0.0
there, 1.000, [14100:14230], -1.09454976E8, 0.0
was, 1.000, [14240:14400], -1.30749464E8, 0.0
driving, 1.000, [14410:14820], -1.56438E8, 0.0
home, 1.000, [14830:15210], -2.1883384E8, 0.0
all, 1.000, [15690:16370], 0.0, 0.0
of, 1.000, [16380:16450], -3.8000988E7, 0.0
a, 1.000, [16460:16490], -5.0451216E7, 0.0
sudden, 1.000, [16500:16780], -5.7295256E7, 0.0
they, 1.000, [16790:16890], -1.01720968E8, 0.0
got, 1.000, [16920:17280], -4870082.0, 0.0
a, 1.000, [17350:17380], -1.89743728E8, 0.0
terrible, 1.000, [17390:18090], -1.96688144E8, 0.0
car, 1.000, [18100:18390], -3.03593824E8, 0.0
accident, 1.000, [18400:18850], -3.49918656E8, 0.0
it, 1.000, [19960:20040], -5954218.0, 0.0
was, 1.000, [20050:20180], -2.1141714E7, 0.0
so, 1.000, [20190:20480], -4.3479204E7, 0.0

```

Şekil 3.4 : TED veri kümesinde hizalama aşaması çıktısı örneği.

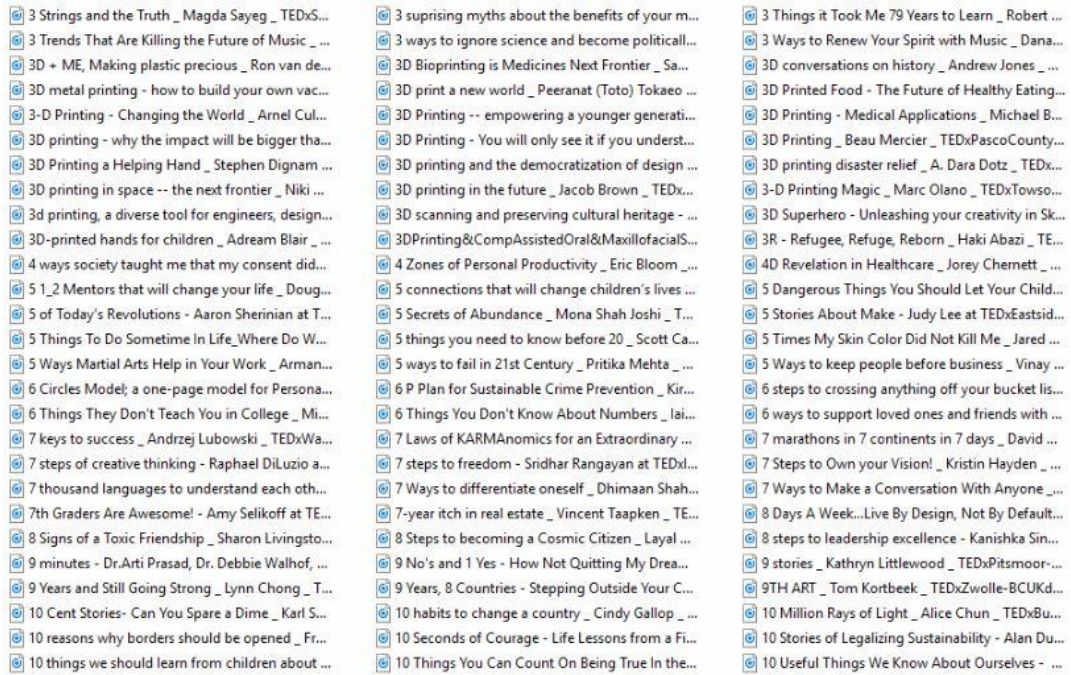
CMU Sphinx program çıktısının değişkenleri virgüllerle ayrılmış olarak elde edilmiştir. Değişkenler sıra ile konuşulan kelimeyi, kelimenin cümlede içinde bir açıklama gibi geçip geçmediğini (altyazıdaki kelimeler temizlenip verildiği için sonuçlarda daima 1 değerini alıyor), kelimenin başladığı ve bittiği süreyi (milisaniye cinsinde), kelimenin ses dosyasındaki formu ile programdaki ses formunun benzerliğini belli eden değeri, kelimenin cümlede kaçınıcı sırada geçtiği (kelimeler ayrı ayrı programa verildiği için daima 0 değerini alıyor) belirtmektedir. Başlangıç ve bitiş zamanları bulunan ardışık kelimeler hem altyazıda aynı cümle içinde ardışık olarak geçiyorsa hem de birinin bitişi ile diğerinin başlangıcı arasında geçen süre 1 saniyeden az ise bu kelimeler birleştirilerek kelime grupları elde edilmiştir. Bu kelime gruplarının süreleri en az 5 saniye olacak şekilde birleştirilerek ses dosyaları elde edilmiş ve TED veri kümesi oluşturulmuştur.

Daha fazla kişiden oluşan ikinci veri kümesi için Youtube'daki TEDx Talks kanalındaki videolar kullanılmıştır. Videolardan ses dosyalarının elde edilmesi TED veri kümesinin elde edilmesine benzer ama birebir aynı değildir. TEDx Talks kanalından alınan videolardan veri kümesindeki ses dosyalarının elde edilmesi arasındaki işlemler Şekil 3.5'te verilmiştir.



Şekil 3.5 : TEDx Talks veri kümesinde videolardan ses dosyalarının elde edilmesi.

Youtube'daki bir kanalın yüklediği tüm videoların listesine erişmek özellikle kanalda çok fazla video varsa zor olmaktadır. Bu sebeple bir kod yazarak otomatik olarak kanala yüklü videoların bulunduğu sayfalar değiştirilerek tüm videoların listesi oluşturulmuştur. Listede 90000'den fazla video bulunmaktadır. TED kanalının aksine TEDx Talks kanalında birçok dilde videolar bulunmaktadır. Tüm videolar arasında İngilizce olan videoları belirlemek için listedeki 90000'den fazla videonun html dosyaları indirilerek, html dosyası içinde videoların İngilizce olduğunu belirtecek kelimelerin olup olmadığı kontrol edilmiştir. İngilizce olan video sayısı 43000 seviyesine inmiştir. Bu videolar hepsi kullanım için uygun değildir. Çünkü bu videoların bazılarının ne yüklenmiş bir altyazısı bulunmaktadır ne de Youtube tarafından otomatik oluşturulan altyazısı bulunmaktadır. Bunun yanında video veya altyazı indirilirken yaşanan sorunlar, video içeriğinde konuşmanın olmaması veya anlaşılacak kadar gürültülü olması gibi nedenlerden dolayı 20741 videodan ses dosyası oluşturulabilmiştir. İngilizce olan videolar belirlendikten sonra bu videolar içinde altyazısı bulunan videolar altyazıları ile birlikte indirilmiştir. Elde edilen videoların ses içerikleri dönüştürücü programlar kullanılarak genellikle '.m4a' formatında elde edilmiştir. İşleme alınan ilgili video seslerinin bir kısmı Şekil 3.6'da, altyazılarının bir kısmı da Şekil 3.7'de verilmiştir.



Şekil 3.6 : TEDx Talks veri kümesindeki seslerin bir kısmının ekran görüntüsü.

- 3 design principles to help us overcome everyd...
- 3 keys to prepare us for an A.I. revolutionized f...
- 3 secrets to practicing self designed learning _ ...
- 3 steps to save mankind - Yanki Margalit at TE...
- 3 Trends That Are Killing the Future of Music _ ...
- 3D + ME, Making plastic precious _ Ron van de...
- 3D metal printing - how to build your own vac...
- 3-D Printing - Changing the World _ Arnel Cul...
- 3D Printing - You will only see it if you underst...
- 3D printing and the democratization of design ...
- 3-D Printing Magic _ Marc Olano _ TEDxTowso...
- 3D Superhero - Unleashing your creativity in Sk...
- 3R - Refugee, Refuge, Reborn _ Haki Abazi _ TE...
- 4D Revelation in Healthcare _ Jorey Chernett _ ...
- 5 Jhooth Maaf _ Anshulika Dubey _ TEDxUSICT...
- 5 things Africa taught me _ Martina Resselmann ...
- 5 to thrive in the 9-5 _ Nicole van Hattem _ TE...
- 5,000 small things are the next big thing - Amy...
- 6 steps to knowing how to know _ Natalie Fole...
- 7 Excellent Tools to Bring Out Your Creativity _ ...
- 7 seconds...PTG _ Doug Graff _ TEDxBaldwinHi...
- 7 things to do before you graduate college _ Li...
- 7th Graders Are Awesome! - Amy Selikoff at TE...
- 8 Signs of a Toxic Friendship _ Sharon Livingsto...
- 9 No's and 1 Yes - How Not Quitting My Drea...
- 9TH ART _ Tom Kortbeek _ TEDxZwolle-BCUKd...
- 10 Million Rays of Light _ Alice Chun _ TEDxBu...
- 3 Easy Steps to Change the World _ Richard Yi...
- 3 magical approaches for achieving a perfect li...
- 3 simple rules to turn food waste into good tas...
- 3 Strings and the Truth _ Magda Sayeg _ TEDxS...
- 3 ways to ignore science and become politicall...
- 3D Bioprinting is Medicines Next Frontier _ Sa...
- 3D print a new world _ Peernat (Toto) Tokaeo ...
- 3D Printing - Medical Applications _ Michael B...
- 3D Printing _ Beau Mercier _ TEDxPascoCounty...
- 3D printing disaster relief _ A. Dara Dotz _ TEDx...
- 3D printing on the moon _ Sergei Dobrianski _ ...
- 3D Vision in insects _ Jenny Read _ TEDxOldha...
- 4 'C's for Your Career _ Elizabeth Handover _ TE...
- 4-Day School Week _ Jariel Vazquez _ TEDxPasc...
- 5 Keys to Success For Social Entrepreneurs - LI...
- 5 Things I Learned in Free Fall _ David Hodge _ ...
- 5 transformations to achieve universal literacy ...
- 6 Centimetres of Separation _ Christopher Hog...
- 6 Things They Don't Teach You in College _ Mi...
- 7 keys to success _ Andrzej Lubowski _ TEDxWa...
- 7 steps of creative thinking - Raphael DiLuzio a...
- 7 thousand languages to understand each oth...
- 7-year itch in real estate _ Vincent Taapken _ TE...
- 8 steps to leadership excellence - Kanishka Sin...
- 9 stories _ Kathryn Littlewood _ TEDxPitsmoor...
- 10 and 90 - How to Think and Act Positively _ L...
- 10 Minutes to Happiness - Ja-Nae Duane at TE...
- 3 Essential Skills To Thrive In The 21st Century ...
- 3 Months, 22 Days and My Letters to Ubin _ Inc...
- 3 simple steps to chase your dreams _ Saurabh ...
- 3 surprising myths about the benefits of your m...
- 3 Ways to Renew Your Spirit with Music _ Dana...
- 3D conversations on history _ Andrew Jones _ ...
- 3D Printed Food - The Future of Healthy Eating...
- 3D Printing - The Future is Here _ Mark Lengsf...
- 3D Printing _ Hyeongyeol (David) Park _ TEDxY...
- 3D printing human tissue - where engineering ...
- 3D Printing Opportunities _ Stephen Elford _ TE...
- 3D, let's create! - Martijn Joris at TEDxLeuven-f...
- 4 lessons I learned while battling cancer _ Sara ...
- 5 1_2 Mentors that will change your life _ Doug...
- 5 of Today's Revolutions - Aaron Sherinian at T...
- 5 Things To Do Sometime In Life_Where Do W...
- 5 Ways Martial Arts Help in Your Work _ Arman...
- 6 P Plan for Sustainable Crime Prevention _ Kir...
- 6 Things You Don't Know About Numbers _ lai...
- 7 Laws of KARMAnomics for an Extraordinary ...
- 7 steps to freedom - Sridhar Rangayan at TEDxI...
- 7 Ways to differentiate oneself _ Dhimaan Shah...
- 8 Days A Week...Live By Design, Not By Default...
- 8 things personal branding does for you _ Rob...
- 9 unexpected rules for an awesome life _ Karste...
- 10 billion people for dinner _ Nina Fedoroff _ T...
- 10 questions I asked the ocean - Nick Freedma...

Şekil 3.7 : TEDx Talks veri kümesindeki altyazıların bir kısmının ekran görüntüsü.

Hizalama aşamasında TED veri kümesine benzer şekilde, CMU Sphinx programını kullanarak altyazıda geçen kelimelerin, ses dosyasında tam olarak ne zaman geçtiği tespit edilmiştir. İşlenen ses ve altyazı dosyalarının birkaç bilgisayarda eş zamanlı çalıştırıldığı için bulunan kelimelerin önüne bulunduğu tarih ve zaman yazılarak oluşabilecek hatalar engellenmiştir. Altyazıdaki kelimeler CMU Sphinx programına verilmeden önce ayıklanmıştır. Hizalama işleminden elde edilen örnek sonuçlar Şekil 3.8’de verilmiştir.

```

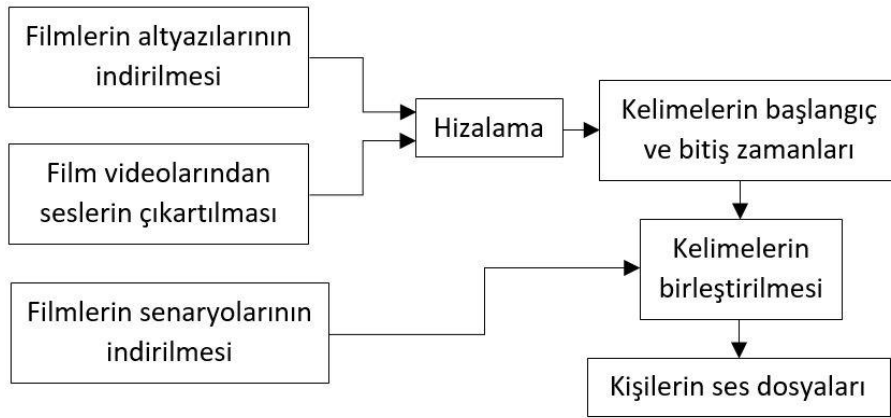
Wed Sep 20 08:16:47 MSK 2017 you, 1.000, [10870:10920], -4.82278976E8, 0.0
Wed Sep 20 08:16:47 MSK 2017 guys, 1.000, [10930:11180], -4.9138448E8, 0.0
Wed Sep 20 08:16:47 MSK 2017 one, 1.000, [11190:11370], -5.30769888E8, 0.0
Wed Sep 20 08:16:47 MSK 2017 amazing, 1.000, [11380:11780], -5.60057664E8, 0.0
Wed Sep 20 08:16:47 MSK 2017 job, 1.000, [11790:12020], -6.210768E8, 0.0
Wed Sep 20 08:16:47 MSK 2017 you've, 1.000, [12030:12110], -6.57961984E8, 0.0
Wed Sep 20 08:16:47 MSK 2017 done, 1.000, [12150:12310], -4883560.0, 0.0
Wed Sep 20 08:16:47 MSK 2017 here, 1.000, [12310:12450], -3.1279186E7, 0.0
Wed Sep 20 08:16:47 MSK 2017 and, 1.000, [12450:12570], -5.210168E7, 0.0
Wed Sep 20 08:16:47 MSK 2017 i, 1.000, [13680:13940], -1.20785E7, 0.0
Wed Sep 20 08:16:47 MSK 2017 am, 1.000, [13940:14110], -5.1711296E7, 0.0
Wed Sep 20 08:16:47 MSK 2017 here, 1.000, [14110:14230], -7.9255872E7, 0.0
Wed Sep 20 08:16:47 MSK 2017 theme, 1.000, [14490:14770], -1.3958072E8, 0.0
Wed Sep 20 08:16:47 MSK 2017 acceleration, 1.000, [14770:15350], -1.8307032E8, 0.0
Wed Sep 20 08:16:47 MSK 2017 and, 1.000, [15350:15860], -2.78031296E8, 0.0
Wed Sep 20 08:16:47 MSK 2017 forget, 1.000, [16450:16730], -4.41292128E8, 0.0
Wed Sep 20 08:16:47 MSK 2017 into, 1.000, [16740:16960], -4.8844128E8, 0.0
Wed Sep 20 08:16:47 MSK 2017 us, 1.000, [16970:17120], -5.24467232E8, 0.0
Wed Sep 20 08:16:47 MSK 2017 got, 1.000, [17730:17940], -6.41074304E8, 0.0

```

Şekil 3.8 : TEDx Talks veri kümesinde hizalama aşaması çıktısı örneği.

CMU Sphinx program çıktısında kelimenin bulunduğu tarih ve zamanın devamındaki değişkenler virgüllerle ayrılmış olarak elde edilmiştir. TED veri kümesi ile aynı şekilde, değişkenler sıra ile konuşulan kelimeyi, kelimenin cümlede içinde bir açıklama gibi geçip geçmediğini (altyazıdaki kelimeler temizlenip verildiği için sonuçlarda daima 1 değerini alıyor), kelimenin başladığı ve bittiği süreyi (milisaniye cinsinde), kelimenin ses dosyasındaki formu ile programdaki ses formunun benzerliği belli eden değeri, kelimenin cümlede kaçınıcı sırada geçtiği (kelimeler ayrı ayrı programa verildiği için daima 0 değerini alıyor) belirtmektedir. Başlangıç ve bitiş zamanları bulunan ardışık kelimeler hem altyazıda aynı cümle içinde ardışık olarak geçiyorsa hem de birinin bitişi ile diğerinin başlangıcı arasında geçen süre 1 saniyeden az ise bu kelimeler birleştirilerek kelime grupları elde edilmiştir. Bu kelime gruplarının süreleri en az 5 saniye olacak şekilde birleştirilerek ses dosyaları elde edilmiş ve TEDx Talks veri kümesi oluşturulmuştur.

Oluşturulan iki veri kümesindeki sesler hep sunum yapan birilerinin konuşmasından oluşmuştur. Kişinin farklı ruh hallerindeki konuşmalarından oluşan bir veri kümesi oluşturmak için sonuncu veri kümesini filmlerde geçen konuşmalar kullanılmıştır. Bu veri kümesinin diğer iki veri kümesinden temel farkı konuşulan kelimelerin kim tarafından söylendiği direkt olarak bilinmemesidir. Bu problemi çözmek için filmlerin senaryoları kullanılmıştır. Film videolarından veri kümesindeki ses dosyalarının elde edilmesi arasındaki işlemler Şekil 3.9'da genel hatları ile verilmiştir.



Şekil 3.9 : Filmlerden ses dosyalarının elde edilmesi.

Film videolarında kelimenin kim tarafından söylendiğini belirlenmesi için senaryolar çok önemlidir. Yaklaşık 1000 adet filmde güvenilir senaryosu olan 249 film belirlenmiştir. Filmlerin videolarından ses kısımları dönüştürücü programlar kullanılarak çıkarılmıştır. Filmlerin altyazıları indirilmiş ve çıkartılan ses dosyaları ile birlikte CMU Sphinx programı çalıştırılmış ve kelimelerin başlangıç ve bitiş zamanları elde edilmiştir. Diğer iki veri kümesinde videolardan çıkartılan sesler boyut olarak büyük olmadığı için çıkartılan ses tek parça şekilde verilmiştir. Ama filmlerden çıkartılan seslerin boyutu büyük olduğu için bilgisayarda CMU Sphinx programı çalışırken hafıza hatasına neden olmuştur. Bu problemi çözmek için hem altyazıyı hem de filmde çıkartılan sesi daha küçük parçalara ayırarak CMU Sphinx programına girdi olarak verilmiştir. Şekil 3.10, bir filmin CMU Sphinx çıktısını içermektedir.

```
is, 1.000, [117960:118140], -4793310.0, 0.0
anything, 1.000, [118150:118420], -3.260054E7, 0.0
i, 1.000, [118430:118480], -7.6638216E7, 0.0
can, 1.000, [118490:118580], -8.6541648E7, 0.0
do, 1.000, [118590:118690], -1.02163728E8, 0.0
for, 1.000, [118700:118850], -1.19508656E8, 0.0
you, 1.000, [118860:118940], -1.45006096E8, 0.0
monsieur, 1.000, [118950:119230], -1.58666544E8, 0.0
no, 1.000, [120810:121070], -3.1428004E7, 0.0
yes, 1.000, [131750:132010], -6269635.0, 0.0
monsieur, 1.000, [132020:132320], -4.7647496E7, 0.0
looking, 1.000, [132550:133550], 0.0, 0.0
around, 1.000, [133670:134000], -5.0313524E7, 0.0
comrades, 1.000, [152220:152630], -4814008.0, 0.0
why, 1.000, [152630:152790], -6.8524488E7, 0.0
should, 1.000, [152790:152990], -9.4780832E7, 0.0
we, 1.000, [152990:153080], -1.26806368E8, 0.0
lie, 1.000, [153080:153180], -1.40607888E8, 0.0
to, 1.000, [153270:153430], -1.74971664E8, 0.0
each, 1.000, [153440:153570], -2.02055536E8, 0.0
other, 1.000, [153580:153790], -2.24568432E8, 0.0
it's, 1.000, [154050:154200], -5940801.5, 0.0
wonderful, 1.000, [154210:154680], -3.1521988E7, 0.0
let's, 1.000, [154200:155100], -3.0639776E7, 0.0
be, 1.000, [155110:155360], -1.75211376E8, 0.0
honest, 1.000, [155370:155790], -2.1816136E8, 0.0
have, 1.000, [155800:155980], -2.84087072E8, 0.0
we, 1.000, [155990:156130], -3.13430656E8, 0.0
anything, 1.000, [156140:156540], -3.38354432E8, 0.0
like, 1.000, [156550:156760], -4.04926816E8, 0.0
it, 1.000, [156770:156850], -4.39872224E8, 0.0
in, 1.000, [156860:156980], -4.5473008E8, 0.0
```

Şekil 3.10 : Bir film için hizalama aşaması çıktısı örneği.

CMU Sphinx programının çıktısında bulunan ardışık kelimeler hem altyazıda aynı cümle içinde ardışık olarak geçiyorsa hem de birinin bitişi ile diğerinin başlangıcı arasında geçen süre 1 saniyeden az ise bu kelimeler birleştirilerek kelime grupları elde edilmiştir. Bu kelime grupları senaryoda geçmiş ve kim tarafından söylendiği belirlenebilmiş ise söyleyen kişinin ses dosyasına eklenmiştir. Aksi takdirde bu kelime grubu kullanılmamıştır. Şekil 3.9'daki kelimelerin birleştirilmesi aşamasının çıktısı Şekil 3.11'de verilmiştir.

423100	425620	leon
426100	426340	swana
426340	427040	swana
427040	430460	
431220	432020	swana
432720	433890	leon
434040	434560	swana
434750	436580	leon
437480	438290	leon
438890	439260	leon
439270	440880	leon
441020	444220	leon
444400	445540	swana
446910	448380	leon
449130	449870	leon
450860	452070	swana
452120	453840	leon
454460	456430	swana
457090	457760	leon
459090	460190	leon
460200	461120	leon
461300	461580	
461930	463550	leon
465970	468970	leon
475750	478080	maid
478860	480480	swana
480510	481730	swana
481730	481760	

Şekil 3.11 : Bir film için kelimelerin birleştirilmesi aşaması çıktısı örneği.

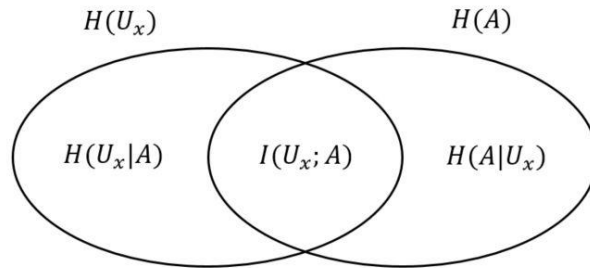
Şekil 3.11'de bir satırdaki ilk sayı kelime grubunun başladığı zamanını, ikinci sayı bitiş zamanını ve diğer kelimeler ise senaryoda kelimelerin kimin söylediğini belirtmektedir. Eğer bir satır sadece iki sayıdan oluşuyorsa bu kelime grubunun senaryoda kim tarafından söylendiği belirlenememiştir. Hem filmlerde arka plan gürültüsünün çok olması hem de filmlerin tam senaryolarının bulunamamasından dolayı diğer iki veri kümesine göre elde edilen yeni film veri kümesi hem daha az kişinin ses dosyasını hem de kişi başı daha az ses dosyası içermektedir. 249 filmde 1595 kişinin ses verisi elde edilmiştir. En az 10 ses dosyası olan 556, 20 ses dosyası olan 286 kişi bulunmaktadır.

3.2 Karşılıklı Bilgi Ölçütü Yöntemi

Bağıl entropi yöntemi kullanılarak ses tabanlı sistemlerin kapasitesinin ölçüldüğü literatürde iki farklı çalışma bulunmaktadır. İlk çalışmada konuşmacı doğrulama sisteminde kişiler GMM modeli kullanılarak modellenmiştir [16]. Ses tabanlı sistemin kapasitesini ölçerken rastgele bir kişinin GMM'lerinin dağılımı ile tüm toplumun GMM'lerinin dağılımları arasındaki bağıl entropi ifadesi hesaplanmıştır. Sonuç olarak bu çalışmada [16] 138 kişilik YOHO veri kümesi kullanıldığında kapasite 14 bit olarak ölçülmüştür. Diğer çalışmada ise, konuşmacı doğrulama sisteminde kişiler i-vektör kullanılarak modellenmiştir [17]. Rastgele bir kişinin i-vektörünün dağılımı ile tüm kişilerin i-vektörlerinin dağılımı arasında bağıl entropi ifadesi hesaplanmıştır. Sonuç olarak bu çalışmada [17] 976 kişilik NIST SRE 2004-2010 veri kümesi kullanıldığında kapasite 120 bit olarak ölçülmüştür.

Biyometrik tabanlı sistemlerde kapasite ölçmek için öncelikle biyometrik bilgi tanımı doğru bir şekilde yapılmalıdır. Önceki çalışmalardaki [7], [5] tanımlara benzer bir şekilde, bir kişinin kimliğinin belirsizliğindeki, kişinin biyometrik verisinden dolayı, ortalama düşüşe biyometrik bilgi [18] denir. Diğer bir deyişle, kişiler kümesinin entropisi ile biyometrik veriden elde edilen bilginin ait olabileceği kişiler kümesinin entropisi arasındaki farka biyometrik bilgi denir.

Biyometrik bilgi, kişinin kimliği ile biyometrik veri arasındaki karşılıklı bilgi ölçütü ile ölçülebilir. $U = \{U_1, U_2, \dots, U_N\}$, N kişiden oluşan bir kümeyi, U_x ve U_n , sırayla kümeden rastgele bir kişi ve n 'inci kişiyi ve $A = [a_1, a_2, \dots, a_k]$, k -boyutlu biyometrik veriyi temsil etmektedir. Karşılıklı bilgi ölçütü ile entropiler arasındaki bağıntı Venn şeması formatında Şekil 3.12'de verilmiştir. Karşılıklı bilgi ölçütü Eşitlik (3.1) ile hesaplanabilir.



Şekil 3.12 : Karşılıklı bilgi ölçütünün Venn şeması.

$$I(U_x; A) = H(U_x) - H(U_x|A) = H(A) - H(A|U_x) \quad (3.1)$$

Hem Şekil 3.12 hem de Eşitlik (3.1)'de $H(U_x)$, $H(U_x|A)$, $H(A)$ ve $H(A|U_x)$ sırayla bir kişinin entropisini, biyometrik verisi bilindiğinde bir kişinin entropisini, biyometrik verinin entropisini ve kişi bilindiğinde biyometrik verinin entropisini temsil etmektedir. Karşılıklı bilgi ölçütünü hesaplariken Eşitlik (3.1)'in $H(A) - H(A|U_x)$ kısmının hesaplanması $H(U_x) - H(U_x|A)$ kısmının hesaplanmasına göre daha kolay olacağı için bundan sonra karşılıklı bilgi ölçütü hesaplanırken $H(A) - H(A|U_x)$ değeri hesaplanmış ve kullanılmıştır. $H(A|U_x)$ ifadesi, veri kümesindeki tüm kişiler için ayrı ayrı $H(A|U_x = U_n)$ ifadesi hesaplanarak Eşitlik (3.2)'deki gibi birleştirilmesi ile hesaplanmıştır.

$$H(A|U_x) = \sum_{n=1}^N P(U_n) H(A|U_x = U_n) \quad (3.2)$$

Veri kümesindeki her kişinin olasılığı $P(U_n)$, birbirine eşit olarak hesaplandığında karşılıklı bilgi ölçütü Eşitlik (3.3)'teki hali alır.

$$I(U_x; A) = H(A) - \frac{1}{N} \sum_{n=1}^N H(A|U_x = U_n) \quad (3.3)$$

Biyometrik verinin entropisi hesaplanırken k-boyut birbirinden bağımsız ise entropi Eşitlik (3.4)'teki gibi hesaplanabilir.

$$H(A) = H(a_1) + \dots + H(a_k) \quad (3.4)$$

Bu çalışmada kullanılan ses tabanlı kimlik doğrulama sisteminde kullanılan i-vektörlerin birbirinden bağımsız olmasından dolayı Eşitlik (3.3)'teki $H(A)$ ifadesi yerine Eşitlik (3.4)'teki ifade yazılarak karşılıklı bilgi ölçütü Eşitlik (3.5)'teki veya Eşitlik (3.6)'daki gibi hesaplanabilir.

$$I(U_x; A) = \sum_{j=1}^k H(a_j) - \frac{1}{N} \sum_{n=1}^N \sum_{j=1}^k H(a_j|U_x = U_n) \quad (3.5)$$

$$I(U_x; A) = \sum_{j=1}^k \left(H(a_j) - \frac{1}{N} \sum_{n=1}^N H(a_j | U_x = U_n) \right) \quad (3.6)$$

3.2.1 Nicemleme

Ses biyometrisinden elde edilen i-vektörler sürekli değer aldıkları için onlar için hesaplanacak entropi değeri i-vektörlerin dağılımı hakkında bilgi verirken, bilgi miktarı olarak direkt kullanılamaz. Bu nedenle, i-vektör değerleri nicemlenerek ayrık değişken haline getirilebilir. İ-vektörlerin ayrık değerler alması, hem i-vektörlerin entropisi hesaplanırken kullanılan olasılık kütle fonksiyonlarını daha doğru olarak tahmin edilmesine hem de hesaplanan i-vektör entropisinin bilgi miktarı olarak kullanılabilmesine olanak sağlar. Nicemleme yapılırken iki temel yaklaşım bulunmaktadır. Bunlar düzgün nicemleme (uniform quantization) ve düzgün olmayan nicemlemedir (non-uniform quantization). Bu iki yaklaşım arasındaki temel fark nicemleme sınırları arasındaki farkın sabit olup olmamasıdır. Düzgün nicemlemede nicemlenen değişkenin dağılımı daha çok korunurken düzgün olmayan nicemlemede farklı özellikler korunur. Düzgün olmayan nicemleme yöntemlerinden biri olan, Llodys tarafından önerilmiş [19] ve literatüre Llodys yöntemi olarak geçmiş nicemleme yönteminde nicemleme hatasının karesinin ortalamasını (Mean Square Distortion) minimize edilir. Nicemleme aralıkları hesaplanırken eğitim verileri kullanılarak yinelemeli bir şekilde nicemleme sınırları belirlenir.

Bu tez çalışmasında i-vektörler nicemlenirken MATLAB'ta bulunan Llodys nicemleme yöntemi kullanılmıştır. Şekil 3.13'te eğitim veri kümesinde bulunan i-vektörlerin örnek olarak 10 tanesinin ilk 20 boyutu verilmiştir. İ-vektörler 2 seviyeli nicemlendiğinde elde edilen sonuçlar Şekil 3.14'te verilmiştir. İ-vektörlerin boyutları nicemlenirken her boyut için nicemleme sınırları birbirinden bağımsız olarak hesaplanarak elde edilmiştir. Bu sebeple farklı boyutlardaki değerler, diğer bir deyişle farklı satırlardaki değerler, genel olarak birbirinden bağımsızdır.

6.6448e-05	4.8322e-04	-7.2329e-04	4.8322e-04	-1.6081e-04	-6.5436e-04	3.9213e-04	4.8322e-04	-4.9959e-05	-7.9159e-04
-8.2277e-04	5.6671e-04	4.8211e-04	-0.0014	-8.2277e-04	1.0035e-06	-0.0015	-0.0014	-6.9967e-04	4.0255e-04
-8.9081e-05	-5.5646e-04	0.0011	8.3319e-04	8.9998e-05	2.7020e-04	7.3271e-04	9.4026e-04	8.9998e-05	4.5130e-04
6.0551e-04	4.8949e-04	-6.6146e-05	2.6277e-04	-7.3569e-04	4.8949e-04	-6.6146e-05	0.0016	8.6001e-04	6.0551e-04
-6.4462e-04	9.0090e-04	3.8419e-04	1.9523e-05	6.8393e-04	-2.5332e-04	-3.4813e-04	1.9523e-05	6.8393e-04	0.0010
-1.2354e-05	-6.4088e-04	-5.2455e-04	-6.4088e-04	3.8060e-04	2.8313e-04	-6.4088e-04	-5.2455e-04	-4.1715e-04	-0.0010
3.4226e-04	5.5142e-04	4.4409e-04	2.3911e-04	1.3672e-04	-7.4742e-04	3.4326e-05	2.3911e-04	0.0011	-6.5092e-05
3.6448e-05	5.6945e-04	3.6448e-05	8.0491e-04	8.0491e-04	5.6945e-04	0.0012	9.3562e-04	5.6945e-04	-7.0078e-05
5.9494e-05	-5.8019e-04	-8.3263e-04	-8.3263e-04	-9.6993e-04	0.0020	-7.0462e-04	-5.8019e-04	7.3863e-04	-8.3263e-04
-3.7223e-04	-0.0011	-0.0016	-4.8141e-04	-6.3850e-05	-8.2196e-04	1.4750e-04	4.0674e-05	4.9081e-04	-4.8141e-04
-1.4964e-04	6.2677e-04	-5.1290e-04	-5.1290e-04	1.8002e-04	-5.1290e-04	-1.4964e-04	-2.6726e-04	-0.0012	5.1210e-04
-0.0020	4.8298e-04	2.5087e-04	-2.1724e-04	8.7596e-04	0.0015	-8.3163e-04	3.6469e-04	8.7596e-04	-3.3384e-04
-2.3774e-04	1.1735e-04	-0.0011	4.0782e-07	8.6804e-04	-9.0418e-04	8.6804e-04	2.3164e-04	-0.0011	0.0012
2.6588e-04	-2.1605e-04	-9.0072e-04	6.2637e-04	2.9243e-05	-9.0875e-05	8.8427e-04	-2.1605e-04	0.0018	0.0016
-4.8212e-05	-5.9418e-04	-0.0012	-3.7107e-04	-0.0018	6.4189e-05	-4.8110e-04	0.0010	6.4189e-05	-4.8212e-05
0.0015	9.9340e-04	-3.3704e-04	-0.0011	-3.3704e-04	-1.0724e-04	1.2797e-04	-7.0853e-04	-3.3704e-04	6.0497e-04
-4.2073e-04	0.0013	3.4475e-04	6.0127e-04	2.2015e-04	0.0013	9.0973e-05	9.0973e-05	9.0973e-05	9.0973e-05
-0.0016	-4.6665e-04	-0.0014	0.0011	7.4479e-04	8.7187e-04	7.4479e-04	4.9856e-04	-0.0018	-8.7535e-04
-1.7905e-04	-5.3377e-05	6.0154e-04	-5.3377e-05	0.0011	7.4664e-04	-0.0013	-1.7905e-04	-8.1895e-04	-5.5756e-04
4.9569e-04	7.4793e-04	0.0015	-3.6321e-04	3.7343e-04	-7.5242e-04	-7.5242e-04	-1.1541e-04	-6.1694e-04	2.4996e-04

Şekil 3.15 : İ-vektörlerin 32 seviye ile nicemlenmesi.

4. KAPASİTE SONUÇLARI

Elde edilen ses dosyalarından i -vektörler elde edilirken literatürde en sık kullanılan parametreler kullanılmıştır. MFCC değerleri hesaplanırken 25-milisaniyelik Hamming pencereler 10-milisaniye kayma (15-milisaniye çakışma) ile uygulanmıştır. Her 25-milisaniyelik pencereden 19-boyutlu MFCC değerleri hesaplanmış ve pencerenin enerjisi de hesaplanarak 20-boyutlu vektör elde edilmiştir. Ardışık pencerelerden gelen 20-boyutlu vektörler kullanılarak her boyuttaki değerlerin birinci ve ikinci türevi hesaplanarak toplamda bir pencereden 60-boyutlu vektör elde edilmiştir. Her ses dosyasının uzunluğu 5 saniyedir ve oluşturulan veri kümelerindeki çoğu kişinin birden fazla ses dosyası bulunmaktadır. Ses dosyasından çıkartılan MFCC'ler 512-dereceli GMM kullanılarak modellenmiştir. Eğitim veri kümeleri kullanılarak i -vektör sistemindeki u_{UBM} ve T elemanları hesaplanmıştır. Ses dosyalarından 200-boyutlu i -vektörler elde edilmiştir. Kimlik doğrulama sisteminde kullanılan kişilerin birden çok ses dosyası kullanılarak birden çok i -vektörü çıkarılmış ve aynı kişinin farklı ses dosyalarından çıkartılarak elde edilen i -vektörlerin ortalaması alınarak kişilerin i -vektör modelleri elde edilmiştir. Kimlik doğrulama sisteminde kullanılan her kişinin en az bir ses parçasından çıkartılan i -vektör kimlik doğrulamasında kullanılacağı için kişinin modeli hesaplanırken kullanılmamıştır. Tüm bu i -vektör çıkarma işlemleri ve i -vektör doğrulama işlemleri MSR araç kutusu [20] kullanılarak gerçekleştirilmiştir.

4.1 Karşılıklı Bilgi Ölçütü Yöntemi

Youtube'daki TED kanalından elde edilen 1914 kişiden oluşan veri kümesinin 993 kişilik kısmı eğitim veri kümesi için kullanılıp kalan 921 kişilik kısmı kullanılarak kimlik doğrulama sisteminin hata oranı hesaplanmıştır. Hata oranı olarak kimlik doğrulama sistemlerinde hatalı olarak doğrulama yapma oranının hatalı olarak doğrulama yapmama oranına eşit olduğu değer kullanılmaktadır. Llodys nicemleme kullanılarak i -vektörler farklı seviyelerde nicemlenmiştir. Aynı işlemler TEDx Talks

kanalından elde edilen 20741 kişiden oluşan veri kümesinde de tekrarlanmıştır. Verinin 5000 kişilik kısmı eğitim veri kümesi için kullanılıp kalan 15741 kişilik kısmı kullanılarak kimlik doğrulama sisteminin hata oranı hesaplanmıştır. İki veri kümesinin sonuçları Çizelge 4.1’de verilmiştir.

Çizelge 4.1 : Nicemleme seviyesinin hata oranına etkisi.

Nicemleme Seviyesi	Nicemleme YOK	2	4	8	16	32
TED Veri Kümesi Hata Oranı (%)	2.89	4.77	2.38	2.06	2.17	2.17
TEDxTalks Veri Kümesi Hata Oranı (%)	2.39	4.96	2.56	2.36	2.38	2.45

Nicemleme seviyesi artıka hata oranı nicemleme olmayan duruma yakınsayacaktır. Bazı seviyelerde hata oranı nicemleme olmayan durumun daha altına düşmüştür. Nicemleme seviyesi küçük değerlerde iken kişiler arasında i-vektörler farklılık göstermezken yüksek seviyelerde bir kişinin farklı ses dosyalarından çıkartılan i-vektörleri farklılık göstermeye başlamış ve orta bir seviyede kişinin farklı i-vektörleri az farklılık gösterip kişiler arasında i-vektörleri farklılık göstermiştir. Nicemleme olmayan seviyenin altındaki hata oranlarının görülmesinin diğere bir nedeni ise nicemleme değeri olarak nicemleme aralığında bir değeri vermek yerine nicemleme seviyesi verilmesidir. Bu şekilde, i-vektörler için karşılaştırmada kullanılan G-PLDA yaklaşımında daha fazla ayırt edebilme gücü elde edilmiştir. Her iki veri kümesinde de en az hata oranı 8 seviyeli nicemleme durumunda elde edilmiştir.

4.2 Seste Karşılık Bilgi Ölçütü

Kendi oluşturduğumuz 1914 kişiden oluşan TED veri kümesini kullanarak ses tabanlı kimlik doğrulama sistemlerinin kapasitesini ölçmek için Eşitlik (3.1)’deki karşılıklı bilgi ölçütü hesaplanmıştır. Sistemi eğitmek için kullanılan 993 kişinin ses dosyalarından çıkartılan i-vektörler kullanılarak nicemleme aralıkları tahmin edilmiştir. Belirlenen nicemleme aralıkları kullanılarak 921 kişinin her ses dosyasından çıkartılan i-vektörler ayrı ayrı nicemlenmiştir. 921 kişinin tüm i-vektörleri kullanılarak her bir nicemleme seviyesinde değeri alma olasılıkları tahmin edilmiş ve $H(A)$ ifadesi Eşitlik (3.4)’teki gibi her boyut ayrı ayrı hesaplanıp toplanması ile elde edilmiştir. Sonra her kişinin için ayrı ayrı $H(A|U_x = U_n)$ ifadesi

hesaplanıp ortalaması alınarak rastgele bir kişi için $H(A|U_x)$ ifadesi bulunmuştur. Elde edilen sonuçlar Çizelge 4.2’de verilmiştir. Hata oranının en az olduğu 8 seviyeli nicemleme durumu dikkate alındığında $I(U_x; A)$, 62.33 bit olarak hesaplanmıştır.

Çizelge 4.2 : TED veri kümesinde nicemleme seviyesi ile karşılıklı bilgi ölçütü ilişkisi.

Nicemleme Seviyesi	$H(A)$	$H(A U_x)$	$I(U_x; A)$
2	199.59	171.21	28.38
4	319.26	272.32	46.94
8	483.31	420.97	62.33
16	615.88	571.47	80.41
32	780.39	679.90	100.49

Yine TED veri kümesi kullanılarak, $H(A|U_x = U_n)$ ifadesi hesaplanırken bir kişinin i-vektörlerinin hangi olasılıklarla hangi nicemleme seviyesinde değer aldıkları tahmin edilirken kullanılan örnek sayısının ölçülen kapasiteye etkisini göstermek için her kişiden ortalama 5, 10, 20, 40 ve 71 örnek kullanılmıştır. $I(U_x; A)$ ifadesi hesaplandığında elde edilen sonuçlar Çizelge 4.3’te verilmiştir. Daha fazla örnek kullanılarak i-vektörlerin dağılımları daha doğru tahmin edilerek daha doğru sonuçlar elde edilmiştir. Özellikle yüksek nicemleme seviyelerinde örnek sayısındaki artış $I(U_x; A)$ ifadesine daha çok etki etmektedir. Bu duruma bir örnek vermek gerekirse, 6 yüzlü bir hileli zarın her yüzünün gelme olasılıkları hesaplanırken en az 6 gözlem gerçekleştirmek gereklidir. Gözlem sayısını arttırarak olasılıklar daha doğru bir şekilde tahmin edilebilir. Benzer şekilde, bir kişinin i-vektörleri tüm nicemleme seviyelerinde eşit sayıda i-vektörü olacak şekilde dağılıyorsa kişinin en az nicemleme seviyesi kadar örneği alınarak bu dağılım doğru olarak tahmin edilebilir. Eğer kişinin i-vektörleri eş dağılıma sahip değilse dağılımın daha doğru tahmin edilmesi için nicemleme seviyesinden daha fazla örneği kullanılarak dağılım tahmin edilmelidir. Küçük nicemleme seviyelerinde sonucun belirli bir değere doğru yakınsadığı görülmektedir. Yüksek seviyeli nicemleme durumlarında da benzer bir yakınsamanın görülmesi için daha fazla örnek kullanılarak tahmin gerçekleştirilmelidir. Kullanılan TED veri kümesi kişi başı ortalama 71 örnek içermektedir.

Çizelge 4.3 : TED veri kümesinde kişi başı örnek sayısının karşılıklı bilgi ölçütüne etkisi.

Nicemleme Seviyesi	5 Örnek	10 Örnek	20 Örnek	40 Örnek	71 Örnek
2	59.94	42.21	34.14	30.29	28.38
4	117.21	81.75	62.21	52.00	46.94
8	190.34	124.83	88.97	71.10	62.33
16	292.45	190.32	127.65	95.76	80.41

Diğer oluşturduğumuz 20741 kişiden oluşan TEDx Talks veri kümesini kullanarak $I(U_x; A)$ ifadesi hesaplandığında elde edilen sonuçlar Çizelge 4.4'te verilmiştir. Sistemi eğitmek için kullanılan 5000 kişinin ses dosyalarından çıkartılan i-vektörler kullanılarak nicemleme aralıkları tahmin edilmiştir. Belirlenen nicemleme aralıkları kullanılarak 15741 kişinin her ses dosyasından çıkartılan i-vektörler ayrı ayrı nicemlenmiştir. Diğer aşamalar Çizelge 4.1 oluşturulmasına benzer şekilde gerçekleştirilmiştir. Hata oranının en az olduğu 8 seviyeli nicemleme durumu dikkate alındığında $I(U_x; A)$, 56.80 bit olarak hesaplanmıştır. Hem daha fazla kişinin kullanılması hem de kişilerin i-vektörlerinin dağılımlarını daha fazla örneğin kullanılması ile tahmin edilmesinden dolayı Çizelge 4.4'te verilen sonuçlar Çizelge 4.2'ye göre daha doğru değerler içermektedir.

Çizelge 4.4 : TEDxTalks veri kümesinde nicemleme seviyesi ile karşılıklı bilgi ölçütü ilişkisi.

Nicemleme Seviyesi	$H(A)$	$H(A U_x)$	$I(U_x; A)$
2	199.19	171.50	27.63
4	368.86	324.03	44.82
8	544.51	487.70	56.80
16	727.29	657.55	69.74
32	898.00	808.99	89.01

Çizelge 4.3'te TED veri kümesini kullanarak elde edilen örnek sayısının etkisi, TEDx Talks kümesinde tekrardan hesaplanmıştır. Her kişiden 10, 20, 40, 80 ve 100 örnek alınarak kişilerin i-vektörlerinin dağılımını tahmin edip $I(U_x; A)$ ifadesi hesaplandığında elde edilen sonuçlar Çizelge 4.5'te verilmiştir. Sonuç olarak, hem kişi sayısının çok olduğunda hem de dağılım tahmininin daha çok örnek kullanılarak gerçekleştirildiğinde ölçülen kapasite düşmektedir. Daha fazla örnek kullanılması

sonuçları iyileştirse de diğer yandan, TEDx Talks veri kümesinde de kişilerin toplam ses dosyası sayısı sınırlıdır.

Çizelge 4.5 : TEDxTalks veri kümesinde kişi başı örnek sayısının karşılıklı bilgi ölçütüne etkisi.

Nicemleme Seviyesi	10 Örnek	20 Örnek	40 Örnek	80 Örnek	100 Örnek
2	42.90	34.52	30.22	27.87	27.32
4	87.78	64.18	51.90	45.49	44.09
8	142.67	96.42	71.63	58.41	55.56
16	225.51	144.74	97.97	73.02	67.74
32	332.24	216.51	139.37	95.69	85.79

4.3 Farklı Kapasite Ölçüm Yöntemleri

Biyometrik kapasiteyi ölçmek için önerilen Hamming mesafesinin (HD) istatistiği ve bağıl entropi yöntemlerini TED veri kümesindeki ses dosyaları üzerinde denendiğinde elde edilen sonuçlar ile $I(U_x; A)$ ifadesinin farklı nicemleme seviyelerindeki sonuçları Çizelge 4.6'da verilmiştir. HD istatistiğini hesaplarken öncelikle kişilerin i-vektör değerlerini nicemleyerek her ses parçasından çıkarılan 200-boyutlu i-vektörlerden 200 bitlik veriler elde edilmiştir. Kişilerin 200 bitlik verileri arasında HD hesaplanarak, bu sonuçların varyans ve beklenen değeri bulunmuştur. Varyans ve beklenen değer kullanılarak kapasite hesaplanmıştır. Diğer bir kapasite ölçme yönteminde, kişilerin i-vektör dağılımları ile toplumun i-vektör dağılımı arasındaki bağıl entropi hesaplanarak kişiler üzerinden ortalaması alınmıştır.

Çizelge 4.6 : TED veri kümesinde diğer kapasite ölçüm yöntemleri.

	Hamming Mesafesi ile	Bağıl Entropi ile	$I(U_x; A)$ 4 seviye nicemleme	$I(U_x; A)$ 8 seviye nicemleme
Kapasite	195.08	109.34	46.94	62.33

4.4 Çeşitliliğin Arttırılması

Ses tabanlı kimlik doğrulama sistemlerinde kullanılan veri kümelerinde, kişilerden alınan ses parçalarında kişilerin genel olarak konuşma tarzları aynıdır. Gerçekte ise bir kişi gün içinde birçok farklı ruh halinde birçok farklı tarzda konuşmaktadır. Kendi oluşturduğumuz TED ve TEDx Talks veri kümelerinde de konuşmalar genel

olarak sunum tarzında olduğu için kişinin farklı tarz konuşmalarını içermemektedir. Bu durum hem ölçülen hata oranını azaltır hem de hesaplanan kapasiteyi yükseltir. Daha doğru bir kapasite tahmini yapmak için veri kümesindeki kişilerin olabildiğince farklı tarzda konuşmalarını içermelidir.

Bu duruma çözüm olarak öncelikle TED veri kümesini kullanarak kişilerin ses dosyalarına farklı ses efektlerini [21] uygulayarak kişilerin farklı tarzdaki konuşmaları elde edilmeye çalışılmıştır. Uygulanan ses efektleri ve parametreler ile bu efektlerin kaç ses dosyasında kullanıldığı Çizelge 4.7’de verilmiştir.

Çizelge 4.7 : Ses efektleri ve parametreler.

Ses Efektleri ve Parametresi	Örnek Sayısı
Efeksiz	15
Ses seviyesini yarıya düşürmek	7
Ses seviyesini 2 katına çıkarmak	7
Yükseklik değişimi (Pitch shift) (0.9 ve 1.1)	14
Yürüme gürültüsü eklemek (0 dB)	7
Restoran gürültüsü eklemek (0 dB)	7
Trafik gürültüsü eklemek (0 dB)	7
Yankı (Echo)	7

TED veri kümesindeki eğitim kısmında kullanılan veriler içinde efektler kullanılmıştır. Efektli sesler kullanarak $I(U_x; A)$ hesaplandığında elde edilen sonuçlar Çizelge 4.8’de verilmiştir.

Çizelge 4.8 : TED veri kümesindeki seslerin efekt ile değiştirilmesi sonucu karşılıklı bilgi ölçütü.

Nicemleme Seviyesi	$H(A)$	$H(A U_x)$	$I(U_x; A)$
2	193.61	177.85	15.76
4	309.65	280.92	28.72
8	472.58	431.15	41.42
16	639.83	581.08	58.75
32	766.53	687.71	78.81

Elde edilen sonuçlar Çizelge 4.2'deki sonuçlara göre beklenildiği gibi daha düşük çıkmıştır. Kişilerin i-vektörlerinin entropisi, $H(A|U_x)$, yükselirken nicemleme sınırlarının tekrardan etkili sesler kullanılarak belirlenmesinden dolayı tüm i-vektörlerin entropi değeri, $H(A)$, düşmüştür. Etkili seslerin kullanılması sonucu 8 seviyeli nicemleme için 62.33 bit olan $I(U_x; A)$ ifadesi aynı seviye nicemleme için 41.42 bit olarak hesaplanmıştır.

Farklı tarzdaki konuşmaların ses üzerindeki etkileri efektlerle tam olarak sağlanamadığı için efekt kullanmak yerine direkt filmlerde geçen konuşmalar kullanarak kişilerin farklı ruh hallerindeki konuşmalarının elde edilmesine çalışılmıştır. Bu amaçla, 1000 film incelenerek internet üzerinde senaryosu bulunan 249 film belirlenmiştir. Filmlerden çıkarılan ses dosyaları ile filmlerin altyazıları kullanılarak, oluşturduğumuz diğer veri kümelerine benzer şekilde, konuşulan kelimelerin başlama ve bitiş zamanları CMU Sphinx programı kullanılarak tespit edilmiştir. Başlama ve bitiş zamanı bilinen kelimelerin filmdeki hangi karakter tarafından söylendiğinin belirlenmesi için filmin senaryosu kullanılmıştır. Daha doğru bir karakter kelime eşleştirmesi yapabilmek için altyazıda kelimenin geçtiği cümledeki diğer kelimelerde senaryoda aranmıştır. Bu şekilde filmdeki karakterlerin konuştuğu zaman aralıkları belirlenerek kişilerin ses dosyaları oluşturulmuştur. Her kişiden 10, 20 ve 40 parça ses dosyası kullanılarak $I(U_x; A)$ hesaplandığında elde edilen sonuçlar Çizelge 4.9'da verilmiştir.

Çizelge 4.9 : Filmlerden elde edilen ses dosyalarında karışıklık bilgi ölçütü.

Nicemleme Seviyesi	10 Örnek	20 Örnek	40 Örnek
2	23.58	14.87	10.44
4	60.27	34.35	21.32
8	114.49	64.32	37.63
16	200.49	114.17	63.95
32	310.96	189.47	107.42

Bu hesaplamaları yaparken en büyük sorun fazla parçası olan kişi sayısının sınırlı olmasıdır. 249 filmde 1595 kişinin ses verisine ulaşılmasına rağmen 10 parça ses dosyası olan 556, 20 parçası olan 286, 40 parçası olan 132 kişi bulunmaktadır. Daha fazla örnek kullanılarak dağılımların tahmininin daha doğru yapılabilmesine rağmen kişi sayısındaki azalma ölçüm sonucunu olumsuz etkilemektedir. Elde edilen

37.63 bitlik kapasite tam olarak doğru olmasa da önceki sonuçlarla karşılaştırıldığında efekt kullanılarak elde edilen sonuçlara yakındır. Önceki tablolardaki sonuçlar göz önüne alındığında, örnek sayısının artması ile $I(U_x; A)$ değerindeki düşüşü ve kişi sayısının artması ile $I(U_x; A)$ değerindeki düşüş hesaba katıldığında $I(U_x; A)$ değerinin 37 bitten az olacağı tahmini yapılabilir.



5. SONUÇ ve GELECEK ÇALIŞMALAR

Bu tez çalışmasında gün geçtikçe kullanımı artan ses tabanlı kimlik doğrulama sistemlerinin kapasitesini ölçmek için öncelikle açık kaynaklar kullanılarak büyük veri kümeleri elde edilmiştir. Geliştirilen kapasite ölçme yöntemi ile ses tabanlı kimlik doğrulama sistemlerinin kapasitesi tahmin edilmiştir. Örnek sayısının etkisi, kişi sayısının etkisi ve kişilerin farklı ruh halleri ile konuşmalarının kapasiteye etkisi detaylı bir şekilde incelenmiştir. Literatürde önerilen diğer biyometrik tabanlı sistemlerin kapasite ölçüm yöntemleri incelenmiştir. Bu çalışmada biyometrik tabanlı sistemlerin güvenliklerine hata oranından farklı olarak entropi alanından bakılarak parolalarla daha rahat kıyaslanabilir bir bakış açısı geliştirilmiştir.

Bu çalışma, cevapladığı soruların yanında araştırılması gereken yeni problemlere öncülük etmektedir. Geliştirilen yöntemle birçok biyometrik tabanlı kimlik doğrulama sisteminin kapasite hesaplanarak karşılaştırma yapılabilir. İris, parmak izi ve yüz gibi farklı biyometrik tabanlı sistemlerin kapasiteleri ölçülerek bu biyometriklerden hangisinin hangi güvenlik seviyesinde kullanılmasının daha doğru olacağı belirlenmelidir. Bu konunun araştırılması gelecek çalışmalar arasındadır.

Bu tez çalışmanın sonucunda yeni oluşan bir diğer problem ise tahmin edilen kapasitelere ulaşabilecek sistemlerin geliştirilmesidir. Bu kapasitelere ulaşan sistemlerin geliştirilmesi biyometrik verilerin güvenlik alanında vazgeçilmez olması sonucunu doğurabilir.



KAYNAKLAR

- [1] Wang, D., Gu, Q., Huang, X., & Wang, P., (2017, April). Understanding human-chosen pins: Characteristics, distribution and security. In *Proceedings of the 2017 ACM on Asia Conference on Computer and Communications Security* (pp. 372-385). ACM.
- [2] Daugman, J., (2003). The importance of being random: statistical principles of iris recognition. *Pattern recognition*, 36(2), 279-291.
- [3] Daugman, J., (2006). Probing the uniqueness and randomness of IrisCodes: Results from 200 billion iris pair comparisons. *Proceedings of the IEEE*, 94(11), 1927-1935.
- [4] Youmaran, R., & Adler, A., (2012). Measuring biometric sample quality in terms of biometric feature information in iris images. *Journal of Electrical and Computer Engineering*, 2012, 22.
- [5] Adler, A., Youmaran, R., & Loyka, S., (2006, May). Towards a measure of biometric information. In *Electrical and Computer Engineering, 2006. CCECE'06. Canadian Conference on* (pp. 210-213). IEEE.
- [6] Shannon, C. E., (2001). A mathematical theory of communication. *ACM SIGMOBILE mobile computing and communications review*, 5(1), 3-55.
- [7] Takahashi, K., & Murakami, T., A metric of information gained through biometric systems. In *Proc. 20th Int'l Conf. Pattern Recognition (ICPR'10)* (pp. 1184-1187).
- [8] Sutcu, Y., Tabassi, E., Sencar, H. T., & Memon, N., (2013, November). What is biometric information and how to measure it?. In *Technologies for Homeland Security (HST), 2013 IEEE International Conference on* (pp. 67-72). IEEE.
- [9] Bimbot, F., Bonastre, J. F., Fredouille, C., Gravier, G., Magrin-Chagnolleau, I., Meignier, S., ... & Reynolds, D. A., (2004). A tutorial on text-independent speaker verification. *EURASIP Journal on Advances in Signal Processing*, 2004(4), 101962.
- [10] Reynolds, D. A., & Rose, R. C., (1995). Robust text-independent speaker identification using Gaussian mixture speaker models. *IEEE transactions on speech and audio processing*, 3(1), 72-83.
- [11] Campbell, W. M., Sturim, D. E., & Reynolds, D. A., (2006). Support vector machines using GMM supervectors for speaker verification. *IEEE Signal Process. Lett.*, 13(5), 308-311.

- [12] **Kenny, P., Boulianne, G., Ouellet, P., & Dumouchel, P.,** (2007). Joint factor analysis versus eigenchannels in speaker recognition. *IEEE Transactions on Audio, Speech, and Language Processing*, 15(4), 1435-1447.
- [13] **Dehak, N., Kenny, P. J., Dehak, R., Dumouchel, P., & Ouellet, P.,** (2011). Front-end factor analysis for speaker verification. *IEEE Transactions on Audio, Speech, and Language Processing*, 19(4), 788-798.
- [14] **Garcia-Romero, D., & Espy-Wilson, C. Y.,** (2011). Analysis of i-vector length normalization in speaker recognition systems. In *Twelfth Annual Conference of the International Speech Communication Association*.
- [15] **Lamere, P., Kwok, P., Gouvea, E., Raj, B., Singh, R., Walker, W., ... & Wolf, P.,** (2003, April). The CMU SPHINX-4 speech recognition system. In *IEEE Intl. Conf. on Acoustics, Speech and Signal Processing (ICASSP 2003), Hong Kong* (Vol. 1, pp. 2-5).
- [16] **Yang, C.,** (2014). *Security in Voice Authentication* (Doctoral dissertation, Worcester Polytechnic Institute).
- [17] **Nautsch, A., Rathgeb, C., Saeidi, R., & Busch, C.,** (2015, April). Entropy analysis of i-vector feature spaces in duration-sensitive speaker recognition. In *Acoustics, Speech and Signal Processing (ICASSP), 2015 IEEE International Conference on* (pp. 4674-4678). IEEE.
- [18] **Tandogan, S. E., Sencar, H. T., & Tavli, B.,** (2017, December). Towards measuring uniqueness of human voice. In *Information Forensics and Security (WIFS), 2017 IEEE Workshop on* (pp. 1-6). IEEE.
- [19] **Lloyd, S.,** (1982). Least squares quantization in PCM. *IEEE Transactions on Information Theory*, 28(2), 129-137.
- [20] **Sadjadi, S. O., Slaney, M., & Heck, L.,** (2013). MSR identity toolbox v1. 0: A MATLAB toolbox for speaker-recognition research. *Speech and Language Processing Technical Committee Newsletter*, 1(4), 1-32.
- [21] **Uzun, E., & Sencar, H. T.,** (2014). A preliminary examination technique for audio evidence to distinguish speech from non-speech using objective speech quality measures. *Speech Communication*, 61, 1-16.

[URL-1] <https://www.youtube.com/user/TEDtalksDirector> alındığı tarih: 11.05.2016

[URL-2] <https://www.youtube.com/user/TEDxTalks> alındığı tarih: 10.04.2017

ÖZGEÇMİŞ

Ad-Soyad : Sinan Erkam TANDOĞAN
Uyruğu : T.C.
Doğum Tarihi ve Yeri : 09.11.1993 - KONAK
E-posta : sinantandogan@etu.edu.tr

ÖĞRENİM DURUMU:

- **Lisans** : 2015, TOBB Ekonomi ve Teknoloji Üniversitesi,
Mühendislik Fakültesi, Elektrik-Elektronik Mühendisliği

MESLEKİ DENEYİM VE ÖDÜLLER:

Yıl	Yer	Görev
2015-2018	TOBB Ekonomi ve Teknoloji Üniversitesi	Burslu Y. Lisans Öğrencisi

YABANCI DİL: İngilizce

TEZDEN TÜRETİLEN YAYINLAR, SUNUMLAR VE PATENTLER:

- **Tandogan, S. E.**, Sencar, H. T., & Tavli, B., (2017, December). Towards measuring uniqueness of human voice. In *Information Forensics and Security (WIFS), 2017 IEEE Workshop on* (pp. 1-6). IEEE.

DİĞER YAYINLAR, SUNUMLAR VE PATENTLER:

- Yilmaz, Y. A., **Tandogan, S. E.**, Hayran, Z., Giden, I. H., Turduev, M., Kurt, H., 2017. Theoretical and experimental investigations of efficient light coupling with spatially varied all dielectric striped waveguides, *Journal of Applied Physics*, 122(3), 033101.