

**KARADENİZ TEKNİK ÜNİVERSİTESİ  
FEN BİLİMLERİ ENSTİTÜSÜ**

**BİLGİSAYAR MÜHENDİSLİĞİ ANABİLİM DALI**

**SESBİRİM TABANLI KONUŞULAN DİLİN TANINMASI**

**YÜKSEK LİSANS TEZİ**

**Kamil Öncü Şen**

**AĞUSTOS 2007  
TRABZON**

**KARADENİZ TEKNİK ÜNİVERSİTESİ  
FEN BİLİMLERİ ENSTİTÜSÜ**

**BİLGİSAYAR MÜHENDİSLİĞİ ANABİLİM DALI**

**SESİRİM TABANLI KONUŞULAN DİLİN TANINMASI**

**KAMİL ÖNCÜ ŐEN**

**Karadeniz Teknik Üniversitesi Fen Bilimleri Enstitüsünce  
“Yüksek Lisans (Bilgisayar Mühendisi) ”  
Unvanı Verilmesi İçin Kabul Edilen Tezdir.**

**Tezin Enstitüye Verildiđi Tarih : 17/07/2007**

**Tezin Savunma Tarihi : 01/08/2007**

**Tez Danışmanı : Yrd. Doç. Dr. Cemal KÖSE**

**Jüri Üyesi : Yrd. Doç. Dr. Hüseyin PEHLİVAN**

**Jüri Üyesi : Yrd. Doç. Dr. Zafer KÜÇÜK**

**Enstitü Müdürü : Prof. Dr. EMİN ZEKİ BAŐKENT**

**Trabzon 2007**

## ÖNSÖZ

Tez konumu seçmemde bana destek olan, uzun tez sürecinde zorlandığım ve bunaldığım zamanlarda, gerek önerdiği yöntemler ve verdiği destek, gerekse de gösterdiği yakın ilgisinden ötürü sayın danışman hocam Yrd. Doç. Dr. Cemal KÖSE'ye en içten dileklerle teşekkür ederim.

Her şeyden önce beni bu günlere getiren, her türlü fedakârlığı yaparak benim yetişmemi sağlayan, gereken her şeyi yapıp beni hiçbir zaman yalnız bırakmayarak her zaman yanımda olan ve sıcacık sevgilerini her zaman hissettiren çok sevgili Anneme, Babama, Kardeşlerime Dedeme, Babaanneme ve tüm aileme layık bir evlat olabilmek temennisi ile kendilerine saygı ve sevgilerimi sunar her şey için çok teşekkür ederim.

Bu çalışma bir noktaya geldiyse, bana kader ortaklığı yapan, en zor zamanlarımda bana desteklerini esirgemeyen, çalışma arkadaşlarım ve dostlarım Salih ARAS'a, Mehmet Emin TENKECİ'ye, Murat AYKUT'a, Yasin KAYA'ya ve Özkan BİNGÖL'e tüm kalbimle teşekkür ederim.

Çok değerli dostum Uğur ŞEVİK'e her zaman benim yanımda olup bana destek olarak hiçbir zaman beni yalnız bırakmadığı için ona minnet ve teşekkürlerimi sunarım.

Bana yurt dışında ki eğitimde danışmanlığı yapan ve çalışmalarımı benden desteğini esirgemeyen Vienna University of Technology öğretim elemanı sayın Yrd.Doç.Dr. Andreas RAUBER'e teşekkürlerimi sunarım.

Aynı evi paylaşıp bana katlanmak zorunda kalan ve zor zamanlarımda benden hiçbir zaman desteğini esirgemeyen dostum Mehmet ALKANAT'a her şey için çok teşekkür ederim.

Her zaman bana sıcak dostluklarını hissettiren beni her konuda destekleyen ve yanımda olan Hakan KARATAŞ'a ve Yağmur GÖK'e teşekkür ederim.

Viyana'da yaşadığım süre zarfında bana zor günlerimde dostluklarını ve desteklerini esirgemeyen başta Özgür ATAGAN olmak üzere tüm Meidling yurdundaki dostlarıma ve Türk Erasmus grubu öğrencilerine en içten teşekkürlerimi sunarım.

Bu çalışmanın hazırlanması esnasında bana desteği olan adını sayamadığım nice insana ve Karadeniz Teknik Üniversitesi'ne teşekkür ederim.

Kamil Öncü ŞEN  
Trabzon, 2007

## İÇİNDEKİLER

ÖNSÖZ .....	III
İÇİNDEKİLER.....	IV
ÖZET .....	VI
ŞEKİLLER DİZİNİ .....	VIII
TABLolar DİZİNİ.....	X
1. GENEL BİLGİLER.....	1
1.1. Giriş .....	1
1.2. Ses ve Fizyolojisi.....	2
1.2.1. Konuşma .....	4
1.2.2. Ses Yolunun İşlevi ve Akustik Analiz.....	5
1.2.3. Ses Sisteminde Uyarım.....	6
1.2.4. İşitme Sistemi .....	6
1.3. Sesbirim Tanıma.....	7
1.3.1. Genel Bilgiler.....	7
1.3.1.1. Konuşmacı Bağımlı ve Bağımsız Sistemler .....	8
1.3.1.2. Ayrışık Kelime Tanıma .....	8
1.3.1.3. Sürekli Ses Tanıma .....	8
1.3.1.4. Sesbirim Tabanlı Ses Tanıma .....	9
1.3.1.5. Kelime Tabanlı Ses Tanıma.....	9
1.3.2. Ses Tanıma Kullanıcıları ve Kullanım Alanları .....	10
1.3.3. Sesbirim Tanımanın Avantajları ve Sınırlamaları .....	10
1.4. Ses Tanıma Süreci .....	12
1.4.1. İşlenmemiş Konuşma.....	13
1.4.2. Sesin Sayısal Kodlanması.....	13
1.5. Sinyal İşleme Teknikleri (Sinyal analizi) .....	14
1.5.1. Çok Çözünürlülük Analizi.....	16
1.5.2. Fourier Dönüşümü .....	17
1.5.3. Dalgacık (Wavelet) Dönüşümü .....	21
1.5.4. Saç Dalgacık Dönüşümü (Haar Wavelet Transform ).....	24
1.6. Sesbirim ile Ses Sinyalinin Eşlenmesi.....	26
1.6.1. Çapraz Korelasyon.....	26
1.6.2. Dinamik Zamanlı Çarpıtma .....	28

1.7.	Konuşma Kısmının Kayıt İçinde Tespit Edilmesi .....	36
1.7.1.	Sıfır geçiş oranı (Zero Crossing Rate) .....	37
1.7.2.	Enerji Seviyelerinin Tespiti .....	37
1.7.3.	Konuşmanın Başlangıç ve Bitiş Yerinin Tespiti .....	38
2.	YAPILAN ÇALIŞMALAR .....	41
3.	SONUÇLAR .....	50
3.1.1	Hızlı Fourier Dönüşümü ve Çapraz Korelasyon ile üretilen tablolar .....	51
3.1.2	Hızlı Fourier Dönüşümü ve Dinamik Zaman Çarpıtma ile üretilen tablolar .....	53
3.2.1	Saç Dalgacık Dönüşümü ve Çapraz Korelasyon ile üretilen tablolar .....	54
3.2.2	Saç Dalgacık Dönüşümü ve Dinamik Zaman Çarpıtma ile üretilen tablolar .....	56
3.3.1	Zaman Boyutu ve Çapraz Korelasyon ile üretilen tablolar .....	57
3.3.2	Zaman Boyutu ve Dinamik Zaman Çarpıtma ile üretilen tablolar .....	59
3.4	Sesbirimler ile Türkçe ve İngilizce Dili ile İlgili Tablolar .....	61
4.	ÖNERİLER .....	66
5.	KAYNAKLAR .....	67

## ÖZET

İlkel çağlardan günümüze kadar geçen zaman zarfı içerisinde, insanlar arası iletişim aracı hiç şüphe yoktur ki ses'dir. İnsanlar duygu, düşünce ve hareketlerini ifade etmede gerçekleştirdikleri seslerle insanlara ifade etmektedir. Benzer bir ilişki insanla bilgisayar arasında da kurulabileceği düşünülmüş ve bu aşamada çalışmalar başlamıştır.

Ses tanıma sistemlerinin gelişimine bakıldığında zaman yurt dışında kayda değer çalışmalara rastlamak mümkündür. Ancak yurdumuzda aynı düzeyde olduğu pek söylenemez. Bu çalışmada bu açığı belirli bir düzeyde kapatmaya yönelik olmuştur. Amaç Türkçede sesbirim tabanlı bir tanıma sistemi gerçekleştirmektir.

Bu amaç doğrultusunda bir ses editörü yapılarak önce sesin alınması ve incelenerek sesbirimlerin kaydedilen kelimelerin içerisinde alınarak sesbirim veri tabanı oluşturulması gerçekleştirilmiştir. Bu veri tabanı oluşturulurken on farklı kadından ve on farklı erkekten ses kayıt örnekleri alınarak bu kayıtlar hem sesbirim veri tabanı oluşturmada hem de sesbirimlerin kayıtlar içerisinde aranmasında kullanılmıştır. Aynı zamanda radyodan alınan kayıtlardan sesbirim araştırması yapabilmek amaçlı 5 farklı kayıt alınmıştır.

Çalışmada Kelime belirleme yöntemi, Hızlı Fourier dönüşüm yöntemi ve Saç Dalgacık Dönüşümü yöntemi kullanılarak Türkçedeki kelimelerin ve sesbirimlerin spektral analizi gerçekleştirilmiştir. Çapraz Korelasyon yöntemi ve Dinamik Zaman Çarpıtma yöntemi kullanılarak kelimelerin içerisinde sesbirimlerin aranması yapabilecek bir sistem gerçekleştirilmiştir. Bu sistem sayesinde kelimelerin içinde belirlenen sesbirimin varlığı araştırılır ve eğer sesbirim bu kelimenin içerisinde var ise bulunduğu yer veya yerler belirlenir. Bu şekilde sistemin denetimi de yapılabilmektedir ve sesbirimler yazıya dönüştürülebilirler. Türkçe kayıtlarda İngilizce sesbirimler ve İngilizce kayıtlarda Türkçe sesbirimler aranarak dillerin birbirlerinin sesbirimlerine olan duyarlılığı test edilerek konuşulan dil belirlenebilirliği araştırılmıştır.

**Anahtar Kelimeler:** Ses Analizi, Ses Tanıma, Sesbirim Tanıma, Kelime Belirleme, Hızlı Fourier Dönüşümü, Çapraz Korelasyon, Dinamik Zaman Çarpıtma

## SUMMARY

### **Speaker Independent Phoneme Based Spoken Language Recognition System**

Speech is the most important means for the communication between the human beings since primitive ages. Mankind transmits their feeling, thoughts and desires to the others by using speech. People thought that there could be make a communication between a human and a computer like communication between human being and started to work to carry out this imagination.

When you look the improvement on this topic, there are many considerable researches made in developed countries, especially in western countries. However, we could not say the same for the researches made for Turkish. In this study, our most important aim is aid to cover this gap. This study's aim is making a phoneme based speech recognition and language identification system.

Through the aim, a sound editor is firstly implemented to make sound records, investigate these records, extract the phonemes in these speech records and create our own phoneme database. To build the database, ten speech records (male and female) for each phoneme are taken from ten different persons. In addition to this, five Turkish and five English speech samples are also taken from the radio for the investigation the phonemes if they are exist or not in these records. In this study, we have used Word Detection, Fast Fourier Transform and Wavelet Transform for the spectral analysis of words and phonemes. Then, these spectral analysis results are examined by employing Cross Correlation technique and Dynamic Time Warping technique. Hence, the system could search the phonemes in the words and if the phoneme exists in the word, then it finds where it is and how many times the phone exist in this word. Finally, the obtained results are discussed in various respects.

**Keywords:** Speech Analysis, Speech Recognition, Phoneme Recognition, Fast Fourier Transform, Haar Wavelet Transform, Word Detection

## ŞEKİLLER DİZİNİ

	<u>Sayfa No</u>
Şekil 1. Periyodik olmayan ve periyodik olan iki ses parçacığı verilmiştir .....	2
Şekil 2. Ses yolunun yapısı .....	5
Şekil 3. Standart bir sesbirim sistemi yapısı .....	12
Şekil 4. Ses sinyallerinin özellik vektörlerine dönüştürülmesi.....	15
Şekil 5. Çok çözünürlülük temeline dayanan bir işaret .....	16
Şekil 6. 50 Hz lik işaret.....	17
Şekil 7. Şekil 6'daki 50 Hz'lik işaretin Fourier dönüşümü sonuçları.....	18
Şekil 8. $x(t)$ fonksiyonunun frekans grafiği .....	19
Şekil 9. a-b) Şekil Fourier dönüşümü sonucu.....	20
Şekil 10. Değişken frekanslı bir işaret .....	20
Şekil 11. Şekil 10.'daki değişken frekanslı işaretin fft dönüşümü .....	21
Şekil 12. Dalgacık dönüşümü seviyelerinde elde edilen frekans aralıkları .....	23
Şekil 13. Filtre analizinin blok diyagramı .....	23
Şekil 14. Haar Wavelet'inin görünümü .....	25
Şekil 15. X ve Y gibi iki sinyal sinyalin koordinat sistemine yerleştirilerek temsili olarak çapraz korelasyon yapılmasının grafiksel olarak gösterilmesi.....	26
Şekil 16. Yerel yolun olası yol haritası.....	29
Şekil 17. Yerel sınırlandırmanın matristeki oluşumu .....	31
Şekil 18. Genel sınırlandırmanın yol haritasını sınırlamasının koordinat sisteminde gösterimi .....	31
Şekil 19. Çapalı başlangıç, serbest bitiş sınırlandırmasının koordinat sisteminde gösterimi .....	32
Şekil 20. Serbest başlangıç, serbest bitiş yapısının sınırlandırmasının koordinat sisteminde gösterimi.....	33
Şekil 21. İki vektörün DTW ile hesaplanmasını grafiksel olarak anlatımı.....	34
Şekil 22. İki vektörün DTW ile hesaplanmasını grafiksel olarak anlatımı.....	35
Şekil 23. İki farklı vektörün dtw aracılığı ile hesaplanmasındaki eşeleme yapılmasının grafiksel anlatımı.....	36
Şekil 24. Bir ses parçasının aşamalarının grafiksel olarak anlatımı .....	39
Şekil 25. Başlangıç ve bitiş noktalarının aranmasının grafiksel anlatımı.....	40



Şekil 26.	Ses kayıt işlemini gerçekleyen programın kullanıcı ara yüz görüntüsü .....	42
Şekil 27.	Ses editörü olarak yapılmış programın kullanıcı ara yüz görüntüsü .	42
Şekil 28.	Gerçek zamanlı FFT ve FDCT yapan programın ekran görüntüsü ...	44
Şekil 29.	Karşılaştırılan iki sesbirimin ekran görüntüsü.....	45
Şekil 30.	FFT uygulanmış verilerin grafik ortamda sunulmasının ekran görüntüsü .....	46
Şekil 31.	Seçili seslere saç dalgacık dönüşümü uygulanmasının grafiksel gösterimi .....	46
Şekil 32.	çevrelerin farklı zamanda dalgacık dönüşümü değerleri grafiksel gösterimi .....	47
Şekil 33.	Zamanda alınan farklı pencerelere uygulanmış FFT sonuç grafiği.	48
Şekil 34.	Karşılaştırma işleminde eşlemenin grafiksel ifade edilmesi .....	49

## TABLolar DİZİNİ

	<u>Sayfa No</u>
Tablo 1. Temel ünlü sesbirimler .....	4
Tablo 2. Türkçede yer alan temel ünsüzler .....	4
Tablo 3. Ötümlü - sürekli olan sessiz sesbirimler birbiri ile olan ilişkisi .....	51
Tablo 4. Ötümsüz - sürekli olan sessiz sesbirimler .....	52
Tablo 5. Ötümsüz - süreksiz olan sessiz sesbirimler .....	52
Tablo 6. Ünlülerin birbiri içerisinde benzerliği .....	52
Tablo 7. Ötümlü - süreksiz olan sessiz sesbirimler .....	53
Tablo 8. Ötümsüz - sürekli olan sessiz sesbirimler .....	53
Tablo 9. Ötümsüz - sürekli olan sessiz sesbirimler .....	54
Tablo 10. Ötümlü - sürekli olan sessiz sesbirimler birbiri ile olan ilişkisi .....	54
Tablo 11. Ötümlü - süreksiz olan sessiz sesbirimler .....	55
Tablo 12. Ötümlü - süreksiz olan sessiz sesbirimler .....	55
Tablo 13. Ötümsüz - sürekli olan sessiz sesbirimler .....	55
Tablo 14. Ötümlü - süreksiz olan sessiz sesbirimler .....	56
Tablo 15. Ötümlü - süreksiz olan sessiz sesbirimler .....	56
Tablo 16. Ötümsüz - sürekli olan sessiz sesbirimler .....	56
Tablo 17. Ötümsüz - sürekli olan sessiz sesbirimler .....	57
Tablo 18. Ötümlü - süreksiz olan sessiz sesbirimler .....	57
Tablo 19. Ötümsüz - sürekli olan sessiz sesbirimler .....	58
Tablo 20. Ötümsüz - süreksiz olan sessiz sesbirimler .....	58
Tablo 21. Ünlülerin birbiri içerisinde benzerliği .....	58
Tablo 22. Ötümlü - sürekli olan sessiz sesbirimler birbiri ile olan ilişkisi .....	59
Tablo 23. Ötümlü - süreksiz olan sessiz sesbirimler .....	60
Tablo 24. Ötümsüz - sürekli olan sessiz sesbirimler .....	60
Tablo 25. Ötümsüz - süreksiz olan sessiz sesbirimler .....	60
Tablo 26. Sesbirimlerin seçili konuşma içinde tanınmasının doğruluk yüzdesi..	61
Tablo 27. Sesbirimlerin seçili konuşma içinde tanınmasının doğruluk yüzdesi..	61
Tablo 28. Sesbirimlerin seçili konuşma içinde tanınmasının doğruluk yüzdesi..	62
Tablo 29. Sesbirimlerin seçili konuşma içinde tanınmasının doğruluk yüzdesi..	62

Tablo 30. Radyo kayıtlarında aranan sesbirimlerin eşlenme oranları .....	63
Tablo 31. Türkçe fonemin Türkçe metin ve İngilizce metin içerisinde araştırılması.....	63
Tablo 32. İngilizce fonemin İngilizce metin ve Türkçe metin içerisinde araştırması.....	64
Tablo 33. Türkçe sesbirimlerin İngilizce konuşmanın içinde aranması .....	65

# 1. GENEL BİLGİLER

## 1.1. Giriş

Günümüze kadar yapılmış olan sesbirim tanıma sistemleri geliştirmeye yönelik çalışmalar henüz istenen düzeyde sonuçlar üretmekten uzaktır. Bu eksiklik bütün diller için geçerli olduğu gibi, Türkçe içinde geçerlidir.

Çalışmada amaç bir yönüyle sesin analizine yönelik matematiksel analizleri içerirken, diğer yönüyle bu analizler neticesinde ortaya çıkacak modellerin yazılımsal olarak konuşmanın sesbirimlerle eşleşmesinin gerçekleşmesi sağlayan bir program geliştirmektir.

Amaçlanan sistemlerin gerçekleştirilmesi ile hem ses analizine teorik bir katkı, hem de sesbirim tanıma alanında Türkçe ile ilgili bir eksik kısmen giderilmesi olacak bir çalışma geliştirmek. Çalışma, farklı yöntemleri bir arada kullanma ve farklı sistemler oluşturularak yöntemlerin araştırmasına olanak sağlamaktadır.

Bu çalışmada tek bir konuşmacıya bağımlı olmayan, konuşmacının bir metin okumasıyla sesbirim tabanlı, konuşma içerisinde belirlenen sesbirimin eşlenme yöntemleriyle eşlenerek konuşmada varlığının ve varsa konuşmadaki adedinin belirlenmesi amaçlanmıştır. Farklı konuşmacılara ait konuşmalarda da sesbirim taraması yapılacaktır.

Çalışmanın ilk bölümünde sesin önemli özelliklerinden, işitme ile konuşmanın nasıl gerçekleştiği, sesin analiz edilmesi ve ses işleme tekniklerinden ve sesbirimlerin seslerle eşlenmesi işleminin aşamalarından bahsedilmiştir. Bu aşamalar, analog ses sinyalinin sayısallaştırılmasını kapsayan, sayısal ses verisi üzerinde yapılan sinyal analizini içeren sinyal işleme teknikleri ve farklı iki ses sinyali üzerinde eşlenmenin yapıldığı ses sinyallerinin modellenmesi olarak sıralanabilir. Ayrıca sesbirim tanıma sistemi kullanıcıları ve kullanım alanları ile sesbirim tanıma sistemlerinin getirmiş olduğu avantajlar ve dezavantajlarda irdelenmiştir. İkinci bölümde, çalışmada kullanılmış programlar ve geliştirilmiş sistemler anlatılmıştır. Üçüncü bölümde ise yöntemlerden elde edilmiş sonuçlar aktarılmıştır. Dördüncü bölümde ise yapılan çalışmanın daha etkin olabilmesini sağlayacak öneriler sunulmuştur.

## 1.2. Ses ve Fizyolojisi

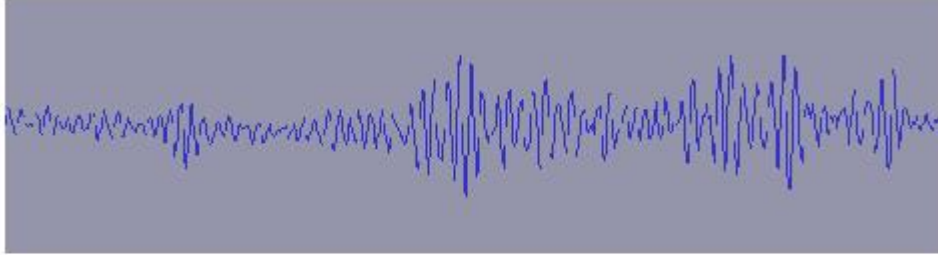
Ses, insan kulağını etkileyerek işitme duyusu oluşturan hava molekülleri titreşimleri veya bunların neden olduğu hava basınç değişimleri gibi işitsel izlenim olarak tanımlanabilir.

Ses fizyolojisi ile ilgili bazı önemli kavramları şöyle açıklanabilir:

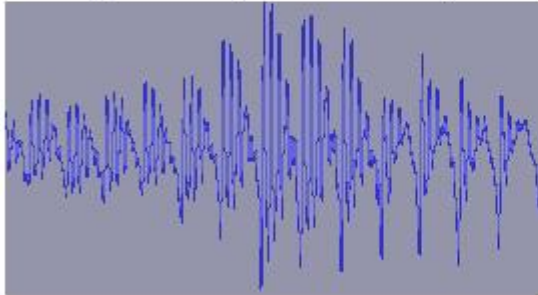
Ses Dalgası : Ses , sıvı, katı ve gaz ortamlarında 20 Hz ile 20 KHz arasındaki insan kulağının algılayabileceği basınç değişiklikleri olarak tanımlanmaktadır. Bu frekans aralığındaki mekanik dalgalar işitme duyusunu uyardıklarından, insan için duyma duyusunun oluşmasını sağlayan ses dalgalarını oluşturur.

İnsan kulağına bir ses dalgası geldiğinde kulak ses dalgasındaki basınç değişikliklerini sinirlerdeki itkilere çevirir ve bunlar beyinde duyulan sesler olarak yorumlanır.

En basit ses dalgasının sadece bir frekansı ve sabit bir genliği vardır. Şekil 1’de basit iki ses parçacığı kullanılarak periyodik ve periyodik olmayan iki ses örnek grafiği ile verilmeye çalışılmıştır.



Periyodik olmayan bir ses olan "ç" sessiz harfi



Periyodik bir ses olan "a" sesli harfi

Şekil 1. Periyodik olmayan ve periyodik olan iki ses parçacığı verilmiştir

Frekans : İki tepe arasındaki uzaklık dalga boyudur. Bir saniyede gözlenen dalga tepesi sayısına frekans denir. Frekans sesin tizliğini belirler. Saniyedeki çevrim CPS (Cycle Per Second-) veya Hertz (Hz) ile gösterilir. Düşük frekanslar bas sesler, yüksek frekanslar ise yüksek seviyeli seslerdir.

Genlik: Sesin diğer bir karakteristik özelliği genliğidir. Sesler yumuşak veya yüksek olurlar. Yumuşaklık veya yükseklik, havayı sıkıştırmak için kullanılan güce bağlı olan bir tür küçük veya büyük basınca karşılık gelir. Ses gücü veya seviyesi için desibel (dB) birimi kullanılmaktadır. Kulağın algılama özelliği ile ilgili yapılan çalışmalarda ses gücünün artması ile hissedilen ses artışının doğrusal olmadığı ve logaritmik bir ses şiddeti ile duyma olduğu anlaşılmıştır. Bu nedenle algılanan ses logaritmik bir büyüklüktür. Haberleşme sistem ve cihazlarının yapısı ve ölçü birimleri de bu sebepten logaritmik olarak düzenlenmiştir.

Gürültü : Periyodik olmayan titreşimlerdir. Kulağın, teknik duyumu bakımından sınırları zorlayan ve psikolojik rahatsızlık doğuran sesler gürültü olarak tanımlanır.

Sesbirim (Fonem) : Ses en küçük yapısal birimlerden en büyük birim cümleye kadar, dili oluşturan bütün yapıların bünyesindeki en küçük ve en basit dil unsurudur. Kaba bir deyişle Türkçede sesler ünlü ve ünsüz sesbirim olarak ikiye ayrılabilir. Ünlüler, oluşumları sırasında ses geçidinde tıkanıklığa uğramadan hiçbir engelle karşılaşmayan sesbirimlerdir. (Yazıcı ve Köse, 1993)

Türkçede ünlüler ağız ünlüleridir ve 8 ünlü sesbirim vardır. Bunların her biri çalışmada incelenmiştir. Türkçe ünlüler dar ve geniş olmak üzere ikiye ayrılırlar. Dar ünlüler ı, i, u, ü ve geniş ünlüler ise a,e,o ve ö'dür. Bu ünlülerden yedi tanesi yer bakımından veya yapı bakımından değişebilmelerine rağmen genellikle kısa sesbirimlerdir. "I" ünlüsü ise çok daha uzun bir sesbirimdir. Sözcüklerin birinci hecelerinde sekiz ünlü sesbiriminin herhangi biri bulunabildiği halde, yapısal uyum nedeni ile sözcüklerin ikinci heceden bulunabilecek ünlü sayısı dörde inmektedir. Eklerde ise sadece iki ünlü birim ("-yor" hariç) bulunabilmektedir.

Ünlüler dilin ve ses yolunun aldığı biçimlere göre üç ölçüte göre değerlendirilebilir. Birinci ölçüt, söyleniş özelliğine ve dilin aldığı biçime göre sesin kalın veya ince, ikinci ölçüt alt çenenin durumuna göre sesin dar veya geniş son olarak üçüncü ölçütte ünlüler durakların durumuna göre düz veya yuvarlak olarak değerlendirilebilir. Bu sınıflandırma Türkçe'de ses uyumunun temel ölçütlerini oluşturur. Yine yapısal olarak Türkçede iki ünlü ardı ardına gelemez (bazı istisnalar hariç). Aynı

şekilde Türkçede sözcüklerin başında ardı ardına iki ünsüz gelemez, fakat cümle içerisindeki kelimelerde gelebilir.

Aşağıdaki Tablo1 ve Tablo2’de İstanbul Türkçesi temel alınarak verilmiş olan temel ünlü ve ünsüz sesbirimler verilmiştir.

Tablo 1. Temel ünlü sesbirimler

AĞIZ AÇIKLIĞINA GÖRE		DİLİN DURUMUNA GÖRE			
		ÖN ÜNLÜLER		ART ÜNLÜLER	
		Kapalı Dar	Açık Geniş	Kapalı Dar	Açık geniş
Dudakları Durumuna Göre	DÜZ	i	e	ı	a
	YUVARLAK	ü	ö	u	o

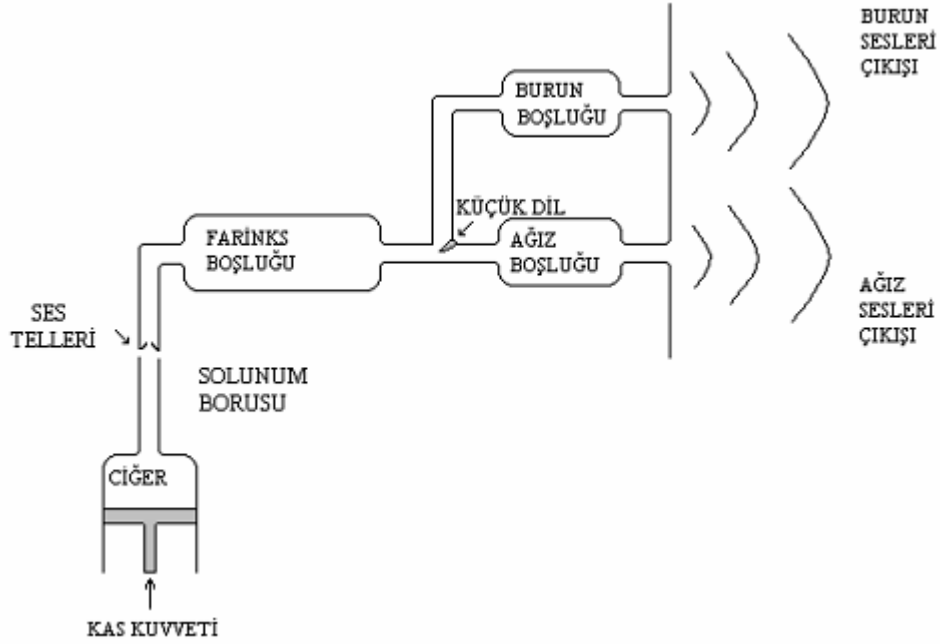
Tablo 2. Türkçede yer alan temel ünsüzler

			ÇIKIŞ YERLERİNE GÖRE						
			Çift Dudak	Diş Dudak	Dişeti	Diş eti-Damak	Damak	Yumuşak Damak	Gırtlak
Ç	Kapanma Ünsüzleri	Ötümsüz	p		t		k (önde)	k (artta)	
		Ötümlü	b		d		g (önde)	g (artta)	
k	Kapanma Dar. Üns.	Ötümsüz				ç			
		Ötümlü				c			
ş	Darama Ünsüzleri	Ötümsüz		f	s	ş			
		Ötümlü		v	z	j	y (yarı ünlü)	g	h
B	Burun Ünsüzleri	Ötümsüz							
		Ötümlü	m		n				
ç	Yan Ünsüzler	Ötümsüz							
		Ötümlü			ı	ı			
m	Çarpmalı Ünsüzler	Ötümsüz							
		Ötümlü			r				

### 1.2.1. Konuşma

Ses dalgası, ses üretim sistemini meydana getiren anatomik yapıların istemli hareketleri sonucunda oluşan, akustik bir basınç dalgasıdır. Bu sistemin ana bölümleri ciğerler, nefes borusu, gırtlak, boğaz, ağız boşluğu ve burun boşluğudur. Teknik terimsel olarak boğaz ve ağız boşluğu ‘ses yolu’ olarak tanımlanır. Ses yolu, gırtlak çıkışından

başlayıp, dudaklarda sona erer. Burun yolu, damaktan başlar burun deliklerinde sona erer. Ses üretimi için kritik olan anatomik yapılar, ses telleri, damak, dil, dişler ve dudaklardır. Ağız iyice açıldığı zaman ağız boşluğunun arka tarafında duran damağın yumuşak uzantısına 'küçük dil' denir. Ses yolunu oluşturan bu anatomik yapılar, farklı pozisyonlar alarak değişik sesleri oluştururlar. Ses yolunun yapısı Şekil2'de görülmektedir.



Şekil 2. Ses yolunun yapısı

### 1.2.2. Ses Yolunun İşlevi ve Akustik Analiz

İnsan sesinin dalga şeklini incelendiğinde, fiziksel sistem zamana bağlı olarak değiştiği için, dalga şeklinin de zamanla değişime uğradığı görülür. Konuşma sesleri, kısa süreler boyunca benzer akustik özellikler gösteren ses parçalarına ayrılabilir.

Ses sinyallerinin zamana bağlı dalga şekillerine bakılarak, sinyal periyotları, yoğunlukları, süreleri ve her bir ses parçasının sınırları tespit edilebilir. Ancak, ardı ardına gelen sesler birbirlerini etkilerse bu sınırları belirlemek mümkün olmayabilir. İnsanların ses üretme ve sesi algılama sistemlerindeki organların yapılarından



kaynaklanan bazı sınırlamalardan dolayı, ses dalga şeklinde deęişime sebep olan yapıların pozisyonlarının kısa zaman aralıklarında deęişmedięi farz edilebilir.

### 1.2.3. Ses Sisteminde Uyarım

Sesin başlıca özellięi, uyarım şeklidir. İki temel uyarım şekli vardır. Bunlar sıra ile *ötümlü* ve *ötümsüz* uyarımlardır. Bunlardan başka, ötümlü uyarım, ötümsüz uyarım ve sessizlięin birleşimi sonucunda oluşan dört tip uyarım daha vardır; *karışık*, *patlamalı*, *fısıltı* ve *sessiz uyarım* (Kuş, 1998).

Ötümlü sesler, havanın nefes borusundan veya ses telleri arasından geçmeye zorlanmasıyla gırtlakta, ötümsüz sesler ise hava akımının sıkıştırılmış ses yolundan geçmeye zorlanması ile üretilir. Ardı ardına gelen seslerden ötümlü-ötümsüz olarak üretilen seslere ‘karışık sesler’ denir. Patlamalı sesler ses yolunun son kısmı kapatılarak içeride basınç oluşması ile meydana gelen ses şeklidir. Basınçlı hava, aniden bırakılarak bu sesler üretilir.

### 1.2.4. İşitme Sistemi

İşitme sistemi işitme olgusundan sorumludur. Bu sistem akustik ses dalgalarını alır ve onları beyin tarafından yorumlanan sinir kodlarına dönüştürür. İnsanın dış dünya ile ilişki kurmasını sağlayan beş duyu organından biri de kulaktır. Kulak ses titreşimlerini alır ve sinir uyarıları yardımıyla beyne ulaşmasını sağlar. Gerçek ses algılaması beyinde olur. Kulak ve beyin arasındaki ilişki, seslerin algılanmasını, işlemlerden geçmesini ve seslerin taklit edilebilecek şekilde öğrenilmesini sağlar. İşitme sistemi dört ana bölümden oluşur: *dış kulak*, *orta kulak*, *iç kulak* ve *beyine giden sinir yolları*.

İşitme duyusunun gerçekleşme gereksinimleri;

- a) Sesin olması,
- b) Sesin kulaęa ulaşması,
- c) O sesin insan kulaęının alabileceęi frekans ve şiddet sınırları içinde olması,
- d) Sesin kulaktaki dış, orta ve iç bölümleri aşması,
- e) Sesin işitme merkezine ulaşması ve merkezce algılanması, gerekmektedir.

### 1.3. Sesbirim Tanıma

#### 1.3.1. Genel Bilgiler

Sesbirim tanıma ses tanıma gibi bir akustik sinyali bir karakter setine dönüştüren bir işlemdir. Son yıllarda geniş bir uygulama alanı oluşmuştur. Tanıma işlemi bazı farklı teknolojilere ve uygulamalara gereksinim duyar. 1950'li yıllardan günümüze kadar yapılan araştırmalarda en yüksek doğruluk yüzdesini artırmak hedeflenen amaçların başında gelir. Bu amaca ulaşmak için gerçekleştirilen sesbirim tanıma sistemleri sinyal işleme, akustik, model tanıma, dilbilim, bilgi teorisi, haberleşme ve psikoloji ve bilgisayar bilimi gibi birçok olgunun birleşiminden meydana gelir. Ses ve sesbirim tanıma alanında yapılan çalışmalarda bu olgular tamamen bilinçli bir şekilde uygulanmalıdır. Sesbirim tanıma ve ses tanıma bu disiplinlerle bağlantılıdır. Bu disiplinlerden ve sesbirim tanımadaki uygulamalarının ne şekilde olduğu hakkında kısaca bilgi verilirse:

**Sinyal İşleme:** Bir ses sinyalinden gerekli bilgiyi en etkili ve en uygun şekilde elde edilme işlemidir.

**Fizik :** İnsan sesinin akustik ve fiziksel mekanizması ile ilgilenen işlemdir.

**Model Tanıma:** Veriyi ilk örnek alan modellerde, gruplamak ve bir çift modelin özelliklerini temel alarak eşleyen işlemdir.

**Haberleşme ve Bilgi Teorisi :** Ses modellerini belirleyen metotlardır.

**Dilbilim :** Dil içinde oluşan ifadelerin sessel, morfolojik, sentaktik, semantik ve faydacı düzeylerdeki yapısını ve işlevlerini ve bu düzeyler arasındaki ilişkileri inceler.

**Fizyoloji :** Ses tanıma çalışmalarında insan sinir sisteminin mekanizmasını anlamak için kullanılır ve anlaşılan yapı bilgisayar alanında kullanılır.

**Bilgisayar Bilimi :** Yazılım ve donanımda çeşitli metotların uygulanması için gerekli ve etkili algoritmaların oluşturulması ile ilgili çalışır.

**Psikoloji :** Amaçlarda, insanlar tarafından kullanılan teknolojiyi anlamak üzere çalışır.

Ses tanıma işleminde kullanılan tekniklerin etkinlik sınırlarının değerlendirilmesinde kullanılan bazı koşullar vardır ve bu koşullar aşağıdaki gibi sıralanabilir:

### **1.3.1.1. Konuşmacı Bağımlı ve Bağımsız Sistemler**

Konuşmacı bağımlı sistemler (Speaker Dependent Systems) için belirli bir kullanıcı tarafından daha önce tanımlanmış bir kelime ya da cümle ele alınır. Bu çeşit sistemler olanakların ve zamanın sistemin bir konuşmacıya bağımlı olarak eğitilmesi için yeterli olduğu uygulamalar için kullanılırlar.

Konuşmacı bağımsız sistemler (Speaker Independent Systems) konuşmacılardan alınan çok miktarda ses örnekleriyle bir ön öğrenmeden gerçekleşir. Kullanıcı böyle bir sistemi hemen kullanmaya başlayabilir. Bu sistemin dezavantajı, bir dil için bütün konuşmacı varyasyonlarını modellemenin imkânsız olmasıdır. Konuşmacı bağımsız sistemler özel konuşmacı eğitimine gerek duymazlar.

Konuşmacı bağımlı sistemlerin performansı konuşmacı bağımsız sistemlere göre daha yüksektir. Bunun yanında bir sistemin kullanım alanını arttırmak için amaç konuşmacıdan bağımsız bir sistem olmasıdır. Tahmin edilebileceği gibi, bunu gerçekleştirmek konuşmacıya bağımlı bir sistem geliştirmekten daha zordur.

### **1.3.1.2. Ayrışık Kelime Tanıma**

Ayrışık kelime tanıma (Isolated Word Recognition), kısa aralıklarla seslendirilen kelimelerin tanınmasıdır. Doğal bir konuşmada, bütün kelimeler arasında duraklama olmaz. Duraklamalar bazı kelime grupları arasında vardır. Bu avantaj sayesinde sistem kelimeler arası sınırları bulmakla uğraşmaz. Kelimeler analiz edilir ve daha önceden hazırlanmış modellerle karşılaştırılırlar.

### **1.3.1.3. Sürekli Ses Tanıma**

Sürekli ses tanıma (Continuous Speech Recognition) sistem ister gerçek zamanlı olsun ister kayıt olsun ara verilmeden seslendirilen kelimelerin tanınması amaçlanır. Bu sistemler bir tanıma işleminin asıl hedefidir. Bir kelimenin ne zaman ya da nasıl sonlandırıldığı sorun değildir. Kelimeler gerçek zamanlı olarak tanınırlar ve sonrasında bir tetiklemeye öncülük ederler. Bu tanıma modeli için başlıca sorunlar, konuşmadaki değişkenlikler, telaffuzlar ve gerçek zamanlı işlem olarak sıralanabilir.

#### **1.3.1.4. Sesbirim Tabanlı Ses Tanıma**

Fonların sesbirimlere dönüştürülmesi ve tanıma işleminin buna dayandırılmasına sesbirim tabanlı ses tanıma (Phoneme Based Speech Recognition) denir. Fon dildeki anlam ayırıcı en küçük öğedir. Konuşma organlarının, anlamlı kelimeler ve cümleler oluşturmak için düzenli olarak çalışması sonucu oluşan olgulardır. Fonların simgesel olarak ifade edilmesi sonucu oluşan simgeler sesbirim olarak adlandırılır. Ters bir tanımlama yapılacak olursa sesbirimlerin seslendirilişiyle fonlar oluşur. Tanıma sırasında seçilen birimler ikili sesbirim (diphone), üçlü sesbirim (triphone), hece veya kelimenin tamamı olabilir (Mengüşoğlu, 1999). Türkçe’de her bir fon alfabetik bir simge ile ifade edilebildiğinden dolayı alfabetik simgeler aynı zamanda sesbirim olarak da adlandırılabilir. Sesbirim tabanlı bir sistemde sözcük sınırı yoktur. Bu tür sistemlerin karşı karşıya bulunduğu önemli bir konu da sesbirimlerin ardı ardına sıralanması sırasında ortaya çıkan geçişlerdir. Bu geçişler sırasında insanın gırtlak yapısının bir sonucu olarak bir sesbirimin seslendirilmesi, bir sonraki sesbirim başladığında hala bitmemiş olur. Bu sebeple kesin bir sınır konulamaz. Bu durum sesli ifadenin kesimlenmesinde bir dezavantaj oluşturmaktadır. Bir başka deyişle, sesli ifadenin sesbirimlere ayrılması sırasında sesbirim sınırlarının belirlenmesinde zorluklar görülür. Bu çalışmada, sesbirim tabanlı ses tanıma ve bahsedilen sorun sesbirimlerinin editör yardımıyla parçalanarak sesbirimlerinin doğal biçimde seçilip işlemlerin gerçekleşmesi için en yüksek doğruluk sağlanmaya çalışıldı.

#### **1.3.1.5. Kelime Tabanlı Ses Tanıma**

Kelime Tabanlı Ses Tanıma (Word Based Speech Recognition) için öngörülen en küçük birimin kelime olan sistemlerdir. Doğruluk yüzdesi sesbirim tabanlı ses tanıma sistemlerine göre daha yüksektir. Çünkü sesbirimler arası geçişlerin olumsuz etkisi burada gözlenmez. Bu tür sistemlerde kelime sayısı sınırlı tutulmak zorundadır.

### 1.3.2. Ses Tanıma Kullanıcıları ve Kullanım Alanları

Sesbirim tanıma yazılımı konuşulan kelimeleri yazılı metne çevirmekte bilgisayar kullanıcılarına yardımcı olmaktadır. Yazılım birçok kullanıcı tarafından farklı ihtiyaç ve amaçlarla kullanılır. 3 tip kullanıcı vardır (Fransson vd., 2001):

1) Elleri kullanma problemi olan kullanıcılar : Yazı yazmak için ellerini kullanmakta zorlanan insanlardır, buna rağmen doğru bir şekilde konuşabilirler. Bunlar bazı problemleri olan kullanıcılarıdır: sinirsel problemi olan veya bir kaza geçirmiş zayıf durumda olanlar veya iyi yazamayanlar vb. örnek verilebilir.

2) Profesyoneller : Yazı yazmak için zamanı olmayanlar veya iyi yazamayanlardır. Bu grup bazı çalışan insanları, tıbbi ve kanuni alanlarda çalışanları veya yazılmış yayınları takip etme ihtiyacı olanları içerir. Yazılım, zaman ve para kazandırır ve kullanıcılara rapor alabilmelerini sağlar. Tipik kullanıcıları doktorlar, avukatlar, psikologlar, satış sorumluları ve diğerleridir.

3) Öğrenme zorluğu olan kullanıcılar : Doğru olarak yazmalarını önleyen öğrenme yetersizliği olan insanlardır.

### 1.3.3. Sesbirim Tanımanın Avantajları ve Sınırlamaları

Sesbirim tanıma metodunun avantajları ve dezavantajları bulunur. Hala çözümü olmayan problemler teknikte bazı sınırlamalara neden olur. İlk olarak, teknik kontrollü ve sessiz ortamda iyi çalışır. Yüksek ses seviyeleri tekniğin faydalı olan avantajlarını bulmayı zorlaştırır.

Eski basit ses tanıyıcılarda içerik düşüncesi yoktur. Bir çıktıdaki sonuçlar sadece bir kelime topluluğudur ve kullanıcı için anlamı vardır.

Tanıyıcı daha fazla kullanılabilir işlem gücü gerektirdiği için uygulamanın kalını bundan olumsuz etkilenir. Bir tanıyıcı için diğer bir problem, araştırmalar en normal sesin kısa kelimeler içerdiğini gösterirken uzun kelimeleri ayırmanın daha kolay olmasıdır.

Sesbirim tanıma konusunda hâlihazırda uygulamadaki sıkıntı, her insanın konuşma tarzının ve ağzının farklı olması, bunun sebebi bir standardın olmamasıdır. Tanınabilir bir lehçeye rağmen, bir sesbirim tanıma aracı herkes için çalışmayı garanti

etmez. Bazı insanlar, diğerleri en sessiz ortamlarda bile tanıma yapamıyorken, en gürültülü ortamlarda bile bu aracı kullanabilirler.

Bazı konuşma tanıma araçları frekans ölçüsünün tamamını kullanamayan donanıma sahiptir. Bu çok kapsamlı olmayan donanım özellikle yüksek frekanslı girişi kapsamaz ve çıktı sonuçları güvenilir değildir. Erkek sesinin kadın sesine tercih edilme nedeni budur.

Sekizinci ve son sınırlama tanıyıcının kapasite seviyesi ile ilgilidir. Mükemmel şartları olan bir iyi-fonksiyonlu tanıyıcı bile sürekli olarak çeşitli hatalar yapar. Tanıyıcı, hatalı sesbirim duyabilir, söylenen sesbirimi atlayabilir ve sesbirimleri yanlış anlayabilir. Kelimenin doğruluk oranı %95'dir. Ama unutmamalıdır ki bir doğruluk oranının %95 olması istatistiksel bakış açısıyla 8–10 kelimeli cümlelerin yarı zamandan daha fazlasında en az bir hata ile tanınmasıdır. Ses tanıyıcılar için tüm bu kısıtlamaları maddeler halinde özetleyecek olursak:

1. Sessiz, kontrollü ortamda en iyi çalışır.
2. Eski basit ses tanıyıcılar dilsizdir, kelimeleri duygusuz söylerler.
3. Sesteş kelimelerin kullanılması karmaşıklığı artırır.
4. Yüksek miktarda işlem gücü ister.
5. Kısa kelimelerdeki sesbirimlerin ayırt edilme zorlukları vardır.
6. Herkes için çalışma garantisi yoktur.
7. Bazı tanıyıcıların diğerlerine göre daha fazla zahmetli işleme tarzı vardır.
8. En iyi durumda sesbirim doğruluğu %95'e ulaşır.

Bütün bunlardan başka, sesbirim tanıma bir iş yeri için çok büyük potansiyellere sahiptir ve yetersizlikleriyle insanlar için eğitimsel uyum süreci vardır. Çoğu durumda bir sesbirim tanıma aracı çalışmazsa, bu kullanıcının davranışına ve bilgisine bağlıdır. Bu gibi yetersiz bilgiyi önlemek için iki ölçüm vardır. Kullanıcılar sistemi ses ile çalıştırmak için iyi hazırlanmış olmalıdır ve güncel teknolojileri bilmelidirler.

#### 1.4. Ses Tanıma Süreci

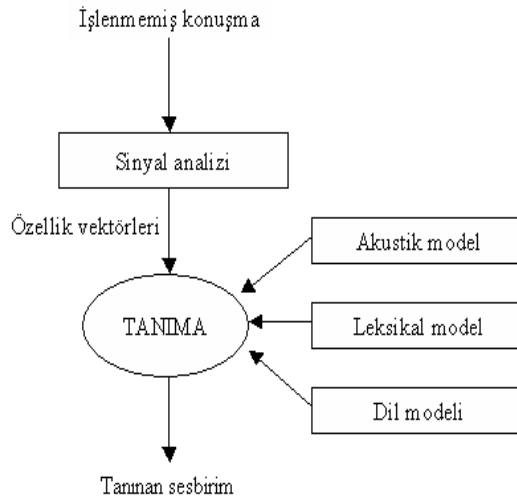
Sesbirim tanıma işlemi söylenmiş bir kelimedeki sesbirimlere karşılık gelen özellik vektörleri dizisinin haritalanması olarak düşünülebilir. Konuşmadaki özellik parametreleri belirlendikten sonra bu parametreler için istatistiksel bir model bulunur. Buna akustik modelleme denir (Yaşaroğlu, 2003).

Konuşmacıların söylediği konuşma en iyi sinyallerinin akustik modelleri veritabanını oluşturur. Bütün bu veri tabanında arama ve verilen bir sinyal için sözlükteki eşleşmeyi seçme tanıma işlemini gerçekleştirir.

Bir tanıma sürecinin çözümünde kullanılan denklem (1)'deki gibidir.

$$W = \arg \max P(W | A) = \frac{P(W)P(A | W)}{P(A)} \quad (1)$$

Formülde amaçlanan akustik sinyal A'nın verdiği en uygun W sesbirimini bulmaktır. P (W) sesbiriminin olasılığını belirler. P (W|A), W sesbiriminin verdiği akustik sinyalin durum olasılığıdır. P (A) kaydedilen sinyalin olasılığıdır. Son amaç, P (W|A) ile P (W) ürünlerinin maksimum değerini bulmaktır.



Şekil 3. Standart bir sesbirim sistemi yapısı

Sesbirim tanıma sürecinde konuşmacının söyledikleri sisteme verildiğinde öncelikle kaydedilir. Böylelikle, elimizde işlenmemiş konuşma verisi olur. Sonraki aşama ses sinyalinin sayısallaştırılmasıdır. Sayısallaştırılan ses sinyali üzerinde sinyal işleme teknikleri (sinyal analizi) uygulanarak sesin ayırt edici özellikleri ortaya çıkartılır. Sonra sinyal işleme sonucunda oluşan özellik vektörleri bazı teknikler ile modellenir. Modellemenin sonucunda sesbirim söylenişine karşılık gelen özellik vektörleri dizisinden bir sesbirim modeli elde edilir ve bu model veritabanından seçilen şablonla karşılaştırılır. Bütün bu veritabanında arama ve verilen sinyal için en iyi eşleşmeyi seçme ile sesbirim tanıma işlemi gerçekleştirilir. Bütün bu aşamalar sonraki bölümlerde ayrıntılı olarak açıklanacak.

#### **1.4.1. İşlenmemiş Konuşma**

Ses girişini alma işlemi farklı teknolojiler ve uygulamalar gerektirir. Konuşma girdi cihazı genellikle bir mikrofon veya bir telefondur. Konuşma çoğunlukla yüksek bir frekansta örneklenir, (örneğin bir mikrofonda 16 KHz veya telefonda 8 KHz olarak). Böylelikle, zaman boyutunda genlik değeri elde edilmiş olunur.

#### **1.4.2. Sesin Sayısal Kodlanması**

Ses analogtur ve işlenebilmesi için öncelikle analog formdan sayısal forma dönüştürülmesi gerekir. Bunu yerine getirmek için geliştirilmiş olan farklı kodlama metotları vardır. Kodlama metotlarının çoğu kodlayıcıya girilen ve kod çözücüde görülen ses arasındaki fark edilir gecikmeden kaçınmak için yüklenmesi gereken, sayısal hatlar üzerindeki gerçek zamanlı ses iletimi için geliştirilmiştir. Bu gerekliliğin sayısal kodlamada mesajı saklamak için kullanımı uygulanamaz, bu nedenle bu uygulama için değişken fazlalığını kullanmak gibi daha büyük bir özgürlük vardır.

Burada ses kodlamada ağırlıklı durulabilecek ses iletiminin iki özelliği vardır. İlk olanı insan işitsel sisteminin sınırlı kapasitesidir. İşitsel sınırlar konuşmacının sesin yeniden üretimindeki çeşitli eksikliklerden etkilenmemesini sağlar. Ses kodlama sistemi tasarlanırken bu aynı zamanda avantajlı da olabilir. Şöyle ki, konuşma mekanizması fizyolojisi meydana gelebilecek sinyal tipleri üzerinde güçlü kısıtlamalar koyar ve bu



gerçek bir ses hattından alınan insan ses üretimindeki bazı durumları modellemede kullanılabilir (Holmes, 2001).

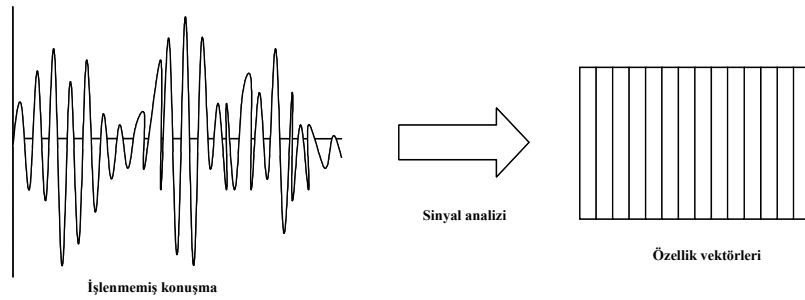
Kodlama metotları üç genel sınıfa bölünebilir:

1. Basit dalga formu kodlayıcıları, veri oranı 16 kbits/s üzerinde işlem yapar:
  - Darbe Kodu Modülasyonu (Pulse Code Modulation-PCM)
  - Uyarlanabilir Diferansiyel Darbe Kodu Modülasyonu (Adaptive Differential Pulse Code Modulation-ADPCM)
  - Delta Modülasyonu (Delta Modulation-DM)
2. Analiz/Sentez sistemleri şunlardır:
  - Kanal Ses Kodlayıcılar (Channel Vocoders)
  - Sinüsoid Kodlayıcılar (Sinusoidal Coders)
  - LPC Ses Kodlayıcılar (LPC Vocoders)
  - Biçimlendirici Ses Kodlayıcılar (Formant Vocoders)
  - Etkin Parametre Kodlama (Efficient Parameter Coding)
  - Parçasal/Fonetik Yapıdaki Ses Kodlayıcılar (Vocoders based on segmental/phonetic structure)
3. Orta düzey sistemler, yukarıdaki iki kategorinin bazı özelliklerine sahiptir ve 4-32 kbits/s bölgesindeki geniş bir alanı kapsar:
  - Alt-Bant Kodlama (Sub-band Coding)
  - Kalanı basit kodlama ile doğrusal tahmin (Linear prediction with simple coding of the residual)
  - Uyarlanabilir Kestirimci Kodlama (Adaptive predictive coding)
  - Çoklu-sinyal LPC (Multipulse LPC)
  - Kod-uyarımlı Lineer Tahmin (Code-excited linear prediction-CELP)

### **1.5. Sinyal İşleme Teknikleri (Sinyal analizi)**

Ses sinyaline, tanımada gerekli olan faydalı özelliklerini ortaya çıkarmak için sinyal işleme teknikleri uygulanır. Amaç sinyaldeki sesbirim bilgisini kötü kayıt

ortamından ötürü gürültü, konuşmacı farklılıkları veya konuşmacının duygusal durumu gibi faktörlerin tesiri olmayacak şekilde saklamaktır. Sesbirim tanıma sistemleri genelde ses sinyallerinin analizinden faydalanmaktadırlar. Ses dalgasının içerdiği frekans ve genlik değerleri her bir sesbirim için farklılık göstermektedir. Ses sinyalinin analizi sonucunda bu frekans ve genlik bilgilerini içeren özellik vektörleri oluşur. Bir özellik vektörü genelde her bir kısa zaman aralığındaki (10 ms) bir ses sinyali penceresinden (20~30 ms)'den hesaplanır. Söylenen kelime bu özellik vektörlerinin bir dizisi olarak gösterilir. Sonraki aşamada bu özellik vektörleri tanıyıcıya giriş olarak verilir.



Şekil 4. Ses sinyallerinin özellik vektörlerine dönüştürülmesi

Bazı sinyal işleme teknikleri herhangi bir önemli veri kaybı olmadan ayırıcı özellikleri bulduğu ve sıkışmış veriyi açabildiği için daha kullanışlıdır (Tebelskis, 1995; Albers, 1998, Kosko, 1992, Sydral, 1995, Jackson, 1986). En popülerleri arasında:

Hızlı Fourier Dönüşümü (Fast Fourier Transform-FFT) gözle görülebilecek biçimde yorumlanabilen zamandaki ayırık frekans bileşenlerini verir. Frekanslar düşük düzeyde lineer olan ama yüksek düzeyde logaritmik olan *Mel* scale kullanılarak ve insan kulağının fiziksel karakteristiklerine uygun olarak sınıflandırılırlar. Fourier dönüşümü ile genlik-zaman boyutu, frekans-genlik boyutuna dönüştürülür.

Wavelet transform (Dalgacık dönüşümleri), frekans bilgilerinin yanı sıra konumsal bilgileri de verebilmektedir. Bir işaretin Wavelet (dalgacık) dönüşümü alındıktan sonra tekrar başlangıç işaretini veride herhangi bir kayıp ya da bozulma olmadan geri üretmek mümkündür.

Discrete cosine transform (ayırık kosinüs dönüşümü) fourier dönüşümüyle alakalıdır ve ayırık fourier dönüşümü ile çok benzerdir fakat burada sadece gerçekteki sayılar

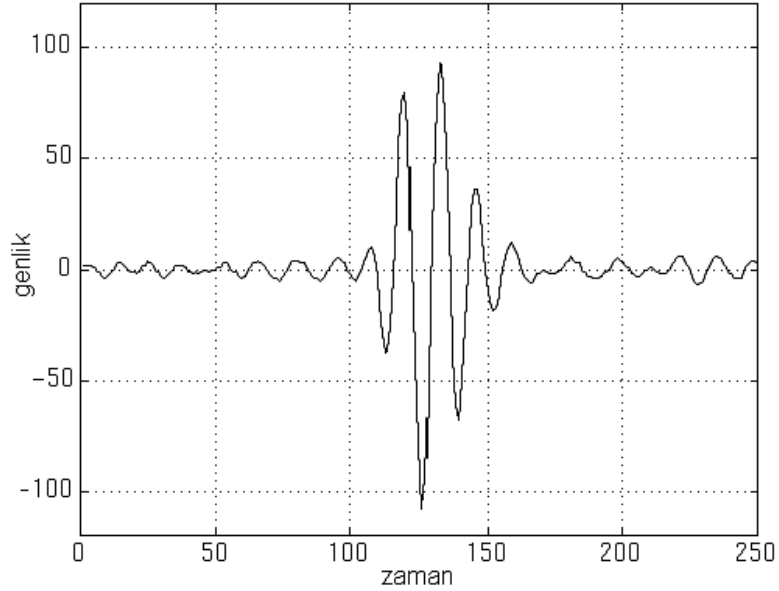
kullanılır. DCT sinyallerin sinüzoid toplamlarının frekans veya genlik değerleri farkları şeklinde açıklar.

Kepstrum (Cepstral Analysis) güçlü sinyal spektrumu logaritmasında Fourier dönüşümün tersi bir işlem yapar.

### 1.5.1. Çok Çözünürlülük Analizi

Fourier tanımlayıcısı ses algılamada kullanılan güçlü bir sistemdir. İşaretteki herhangi bir kayma Fourier tanımlayıcısının vereceği sonuca etki etmez. Çünkü Fourier dönüşümü işarete genel olarak bakmaktadır. Yani Fourier dönüşümü, sadece işarete bulunan frekans bileşenlerini vermektedir. O frekansın hangi zamanda meydana geldiği konusunda herhangi bir bilgi vermemektedir.

Çok çözünürlülük analizi; yüksek frekanslarda iyi zaman ve zayıf frekans çözünürlüğü, düşük frekanslarda iyi frekans ve zayıf zaman çözünürlüğü vermesi için tasarlanmıştır. Pratik uygulamalardaki işaretler de genellikle böyle işaretlerdir.



Şekil 5. Çok çözünürlülük temeline dayanan bir işaret

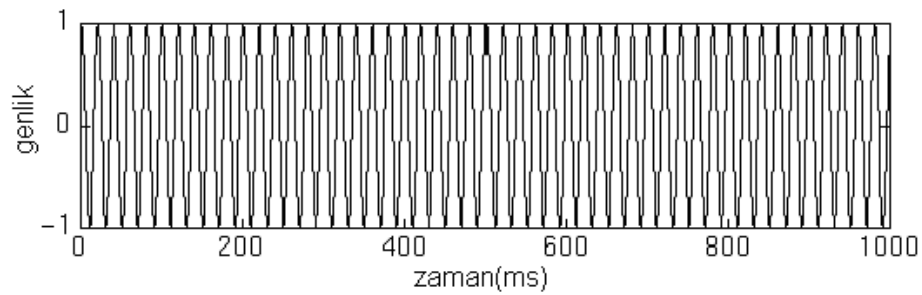
Şekil 5.'teki işaret çok çözünürlülük amacına uygun bir işarettir (<http://users.rowan.edu/~polikar/WAVELETS/WTtutorial.html>, Wavelet Tutorial). Tüm işaret genel olarak nispeten düşük frekans bileşenine sahip olduğu halde işaretin ortaya yakın kısımlarında biraz daha yüksek frekans bileşeninin hâkim olduğu görülmektedir.

### 1.5.2. Fourier Dönüşümü

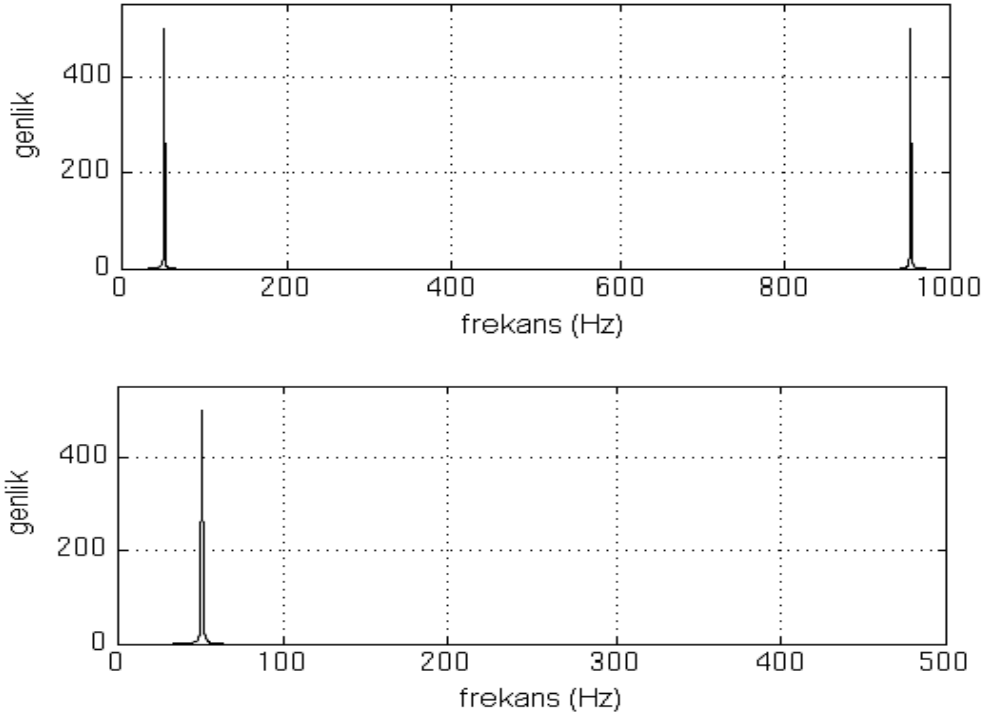
Pratikteki pek çok işaret zaman boyutunda tanımlanır. İşaretin grafiği göz önünde alındığında eksenlerden biri zaman (bağımsız değişken), diğeri ise genliktir (bağımlı değişken). İşaret işleme ile ilgili bir çok uygulamada zaman-genlik gösterimi ile sonuca gitmek zordur. Çoğu durumda en önemli bilgi frekans bileşeninde yatmaktadır.

İşaretteki değişim çok hızlı ise o işaretin frekansı yüksek, değişim yavaş ise o işaretin frekansı düşüktür denir. İşarete herhangi bir değişim görülüyorsa o işarete sıfır frekanslı işaret ya da frekansı olmayan işaret denir.

Bir işaretin frekans bileşenleri *Fourier Dönüşümü* kullanılarak hesaplanır. Zaman alanındaki bir işaretin Fourier dönüşümü alındığı takdirde o işaretin frekans-genlik bilgileri elde edilir. Fourier dönüşümü uygulanması sonunda elde edilen bilgi eksenlerinden biri frekans diğeri ise genliktir. Bu bilgiler, işarete her frekanstan ne kadar bulunduğunu göstermektedir.



Şekil 6. 50 Hz lik işaret



Şekil 7. Şekil 6'daki 50 Hz'lik işaretin Fourier dönüşümü sonuçları

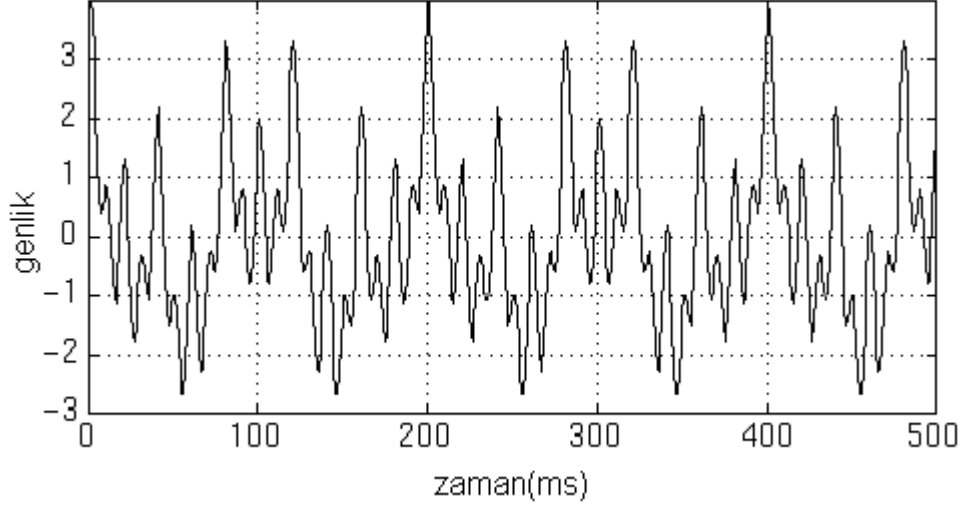
Şekil 7.a'daki grafikte iki adet tepe değeri görülmektedir. Bunun nedeni, gerçek değerli işaretlerin frekanslarının simetrik olmasıdır

(<http://users.rowan.edu/~polikar/WAVELETS/WTtutorial.html>, Wavelet Tutorial).

Fourier dönüşümünün mühendislerce ve matematikçilerce kullanılan türleri vardır. Bunlardan bazıları; Hilbert dönüşümü, kısa süreli Fourier dönüşümü, Wigner dağılımları, Radon dönüşümü, Wavelet dönüşümüdür. Tüm dönüşüm türlerinin kendine özgü uygulama alanlarının olmasının yanında onlara ait üstünlükleri ve sakıncaları da bulunmaktadır.

Dalgacık dönüşümü Fourier dönüşümünden türetildiğinden dolayı öncelikle Fourier dönüşümünün iyi algılanması gerekmektedir. Fourier dönüşümü geri dönüşümlü bir dönüşümdür. Yani işlenmiş işareten işlenmemiş işarete dönüşüm ve tersi dönüşüm mümkündür. Fakat belirli bir anda bunlardan sadece birinin bilgisini almak mümkün olabilmektedir. Bir başka deyişle zaman boyutundaki grafiğinde frekans bilgisi alınamazken, zaman bilgisine de Fourier dönüşümü uygulanmış işareten ulaşmak mümkün değildir.

Sabit frekanslı işaretlerde Fourier dönüşümünün zaman hakkında bilgi verememesinin önemi yoktur. Frekans zamanla değişmediğinden dolayı frekans o işaret boyunca sürekli vardır.



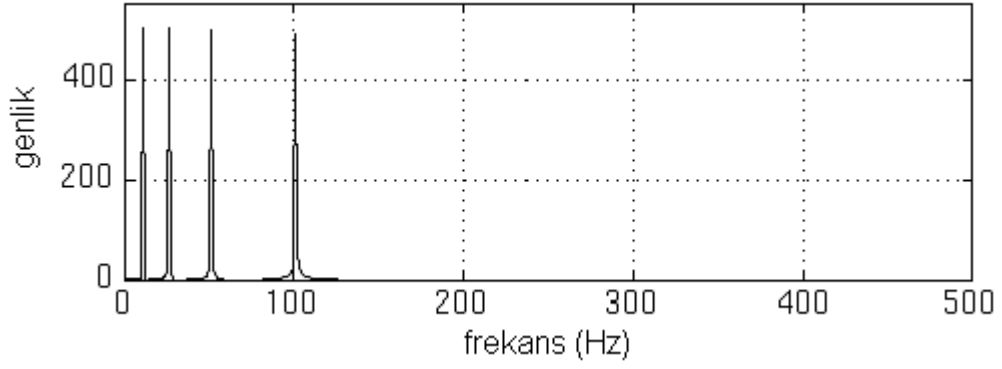
Şekil 8.  $x(t)$  fonksiyonunun frekans grafiği

Şekil 8  $x(t) = \cos(2\pi 10t) + \cos(2\pi 25t) + \cos(2\pi 50t) + \cos(2\pi 100t)$  fonksiyonlu  $x(t)$ 'nin 10, 25, 50, 100 Hz frekanslarındaki görüntüsüdür. Zaman içinde frekansın değişiklik gösterdiği işaretlere ise değişken frekanslı işaretler denir. Şekil 8.'deki görüntü, değişken frekanslı işaretlere bir örnektir.

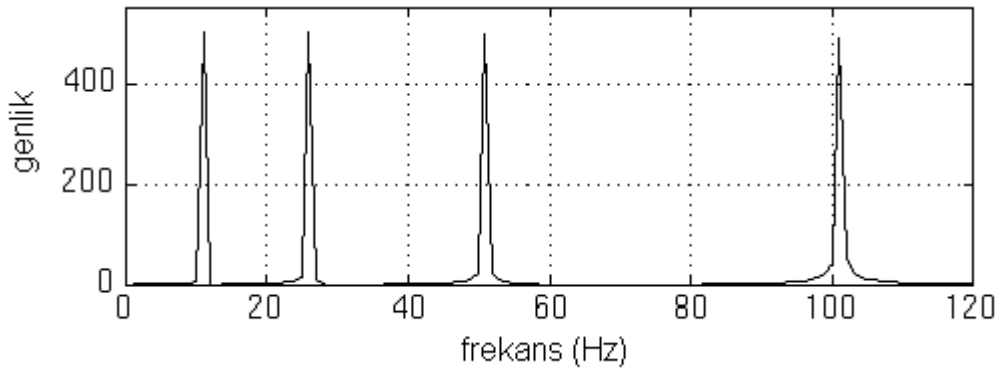
Şekil 9. ile Şekil 11. karşılaştırıldığında aralarında benzerlik görülmektedir. Ancak bunlardan bir tanesi sabit frekanslı işaret diğeri ise değişken frekanslı işarettir.

Fourier dönüşümünün hangi frekans bilgisinin hangi anda var olduğunu gösterememesi, değişken frekanslı işaretler için uygun bir teknik olmadığı göstermektedir.

Dalgacık dönüşümü zaman-frekans bilgisini verebilmektedir. Bu bilgiyi kısa zamanlı Fourier dönüşümü ve Wigner dağılımları gibi dönüşümler de verebilmektedir.



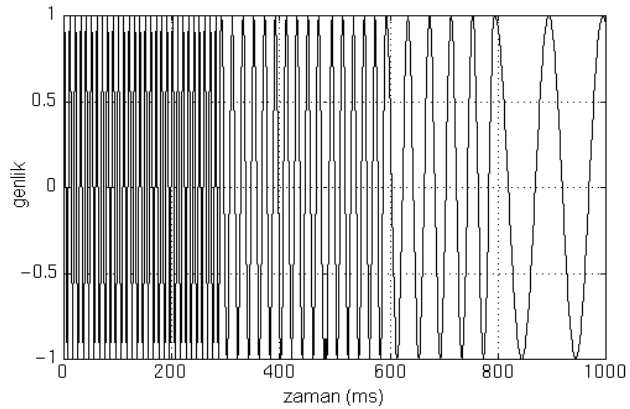
(a)



(b)

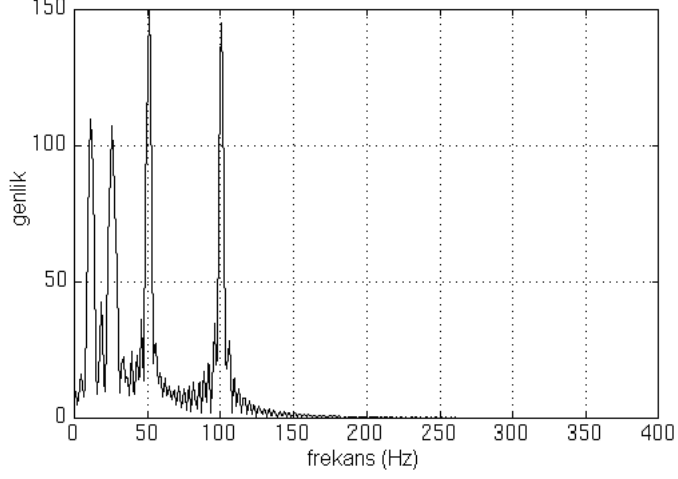
Şekil 9. a-b) Şekil Fourier dönüşümü sonucu

Dalgacık dönüşümü, kısa zamanlı Fourier dönüşümünde karşılaşılan bazı çözünürlük temelli problemlerin aşılması için kısa zamanlı Fourier dönüşümüne alternatif olmak üzere geliştirilmiştir.



Şekil 10. Değişken frekanslı bir işaret

Bu işaretin Fourier dönüşümü Şekil 11’de gösterilmiştir.



Şekil 11. Şekil 10.’daki değişken frekanslı işaretin fft dönüşümü

Zaman alanındaki işaret yüksek geçirgenlikli ve alçak geçirgenlikli filtrelerden geçirilerek işaretin yüksek frekanslı bölgeleri ve alçak frekanslı bölgeleri ayrı ayrı elde edilmiş olur. Bir sonraki adım için genellikle bir önceki adımda elde edilen alçak frekanslı bölge alınır ve filtreler tekrar uygulanır. Bu işlem önceden belirlenmiş bir seviyeye gelinceye kadar tekrarlanır. Bu işleme *ayrışım* (decomposition) denir. Böylelikle ayrışım işlemi sonunda frekans grupları elde edilir. Ancak kısa zamanlı Fourier dönüşümü ile belirli bir zamanda hangi frekansın olduğu bilinemez. Bunun yerine o frekansın hangi zaman aralığında olduğu belirlenebilir. Zaman aralığı kavramı da çözünürlük problemini ortaya çıkarmaktadır. Yüksek frekanslı işaretler zaman ekseninde daha çok örneklenebildiklerinden dolayı zaman ekseninde daha hatasızdır. Buna karşın, alçak frekanslı işaretler ise belirli bir aralık boyunca aynı frekans bilgisini taşıdıklarından dolayı frekans ekseninde daha hatasızdır.

### 1.5.3. Dalgacık (Wavelet) Dönüşümü

Dalgacık dönüşümü kısa zamanlı Fourier dönüşümündeki çözünürlük problemini ortadan kaldırmak üzere alternatif olarak tasarlanmıştır. Wavelet dönüşümü, kısa zamanlı Fourier dönüşümü ile aynı biçimde uygulanır. İşaret fonksiyon ile çarpılır ve sonuçlar



toplanır ([http://en.wikipedia.org/wiki/Haar\\_wavelet](http://en.wikipedia.org/wiki/Haar_wavelet) , Haar Wavelet). 1976'da Croiser, Esteban ve Galand ayrık zaman işaretlerini bileşenlerine ayırmaya yönelik bir teknik geliştirdi (Gonzalez,2002). Aynı yıl Crochiere, Weber ve Flanagan ses işaretlerini analiz etti ve bu analize *alt bant kodlama* adı verildi (Gonzalez,2002), 1983'te Burt *piramitsel kodlama* adı altında alt bant kodlamasına çok benzeyen yeni bir teknik geliştirdi (Gonzalez,2002). 1989'da Vetterli ve Le Gall alt bant kodlama yöntemini, piramitsel kodlamadaki bazı fazlalıkları atarak geliştirdi (Gonzalez,2002).

İlk Dalgacık dönüşümü Macar matematikçi Alfred Haar tarafından keşfedilmiştir. Haar Wavelet (Saç Dalgacık) dönüşümü en basit Wavelet dönüşümüdür (Graps,1995) Bu dönüşüm temel olarak girdi değerlerinin ikişer ikişer ortalamasını alır. Bunun yanında ortalamadan oluşan farkı da ek bilgi olarak tutar. Bu işlem özyinelemeli olarak devam eder. İşlem sonunda  $2^n - 1$  tane fark ve bir tane toplam ortalama değer oluşur ([http://en.wikipedia.org/wiki/Haar\\_wavelet](http://en.wikipedia.org/wiki/Haar_wavelet) , Haar Wavelet)

Dalgacık dönüşümü, frekans bilgilerinin yanı sıra konumsal bilgileri de verebilmektedir. Bir işaretin Dalgacık dönüşümü alındıktan sonra tekrar başlangıç işaretini veride herhangi bir kayıp ya da bozulma olmadan geri üretmek mümkündür.

Dalgacık analizindeki temel fikir, bir  $f$  işaretini  $\psi_i$  fonksiyonuna bağlı olarak bileşenlerine ayırmaktır (Lu, C.S,1997)

$$f = \sum_i a_i \psi_i \quad (2)$$

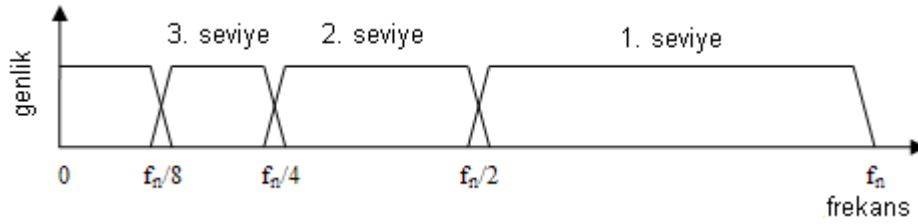
Buradaki  $\psi_i$  fonksiyonunun uygun Dalgacık ailesinden seçilmesi çok önemlidir. İşaretle,  $\psi_i$  fonksiyonu birbirine uygun olmalıdır.

Bir  $x[n]$  işaretinin ayrık Dalgacık dönüşümü bu işaret üzerinden alçak geçirgenlikli  $g[n]$  ve yüksek geçirgenlikli  $h[n]$  filtresinin geçirilmesiyle hesaplanır (Lu, C.S,1997).  $g[n]$  ve  $h[n]$  filtreleri ana Dalgacık adı verilen bir fonksiyondan türetilmişlerdir ( Busch, C. ,1997).

$$y_{alçak}[n] = \sum_{k=-\infty}^{\infty} x[k]g[2n-k] \quad (3)$$

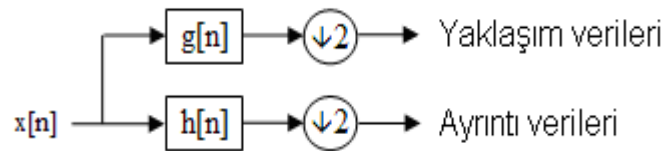
$$y_{yüksek}[n] = \sum_{k=-\infty}^{\infty} x[k]h[2n-k] \quad (4)$$

Bu iki işlemin ardından girdi verisi yüksek ve alçak frekanslı bileşenlerine ayrılmış olur. Örnek olarak, girdi verisi 1000 Hz.lik bir işaretten oluşuyorsa, Dalgacık dönüşümü sonucunda 0–500 Hz ve 500–1000 Hz. frekans aralıklarından ibaret iki işaret elde edilir.



Şekil 12. Dalgacık dönüşümü seviyelerinde elde edilen frekans aralıkları

Filtreleme işleminin sonucunda yüksek geçirgenlikli filtreden elde edilen verilere ayrıntı verileri, alçak geçirgenlikli filtreden elde edilen verilere ise yaklaşım verileri denmektedir (<http://en.wikipedia.org/wiki/Wavelet>, Wavelet).



Şekil 13. Filtre analizinin blok diyagramı

Ters dönüşüm yapılarak ana veriyi tekrar kayıpsız olarak elde etmek mümkündür ( Beng T. Beng T. TAN, Minyue Fu, Andrew Spray ve Phillip Dermody). Bu dönüşüme *ters Dalgacık dönüşümü* denir.

$N$  adet veri için, hızlı Fourier dönüşümü  $O(n \log n)$  işlem karmaşıklığına sahiptir. Buna karşın ayrık Dalgacık dönüşümü  $O(n)$  işlem karmaşıklığına sahiptir. Buradan da görüldüğü gibi, Dalgacık dönüşümü Fourier dönüşümü ile kıyaslandığında daha hızlı sonuç üretebilen bir yöntemdir.

#### 1.5.4. Saç Dalgacık Dönüşümü (Haar Wavelet Transform )

Dalgacık dönüşümü sürekli ve ayrık Dalgacık dönüşümü olmak üzere iki çeşittir. Sürekli Dalgacık dönüşümü Fourier dönüşümü gibi integraller ve analitik eşitlikler yardımıyla hesaplanmaktadır. Günümüzde bu hesaplamalar bilgisayarlarla yapıldığından dolayı, bu dönüşümlerin ayrık çeşitleri de türetilmiştir.

Ana Dalgacık olarak kullanılabilen pek çok temel fonksiyon mevcuttur. Dönüşümde kullanılacak olan bütün fonksiyonlar ana Dalgacık tarafından üretildiği için, Dalgacık dönüşümünün karakteristiğini ana Dalgacık belirleyecektir. Bu yüzden uygulamalarda uygun Dalgacık türünün seçilmesi büyük önem taşımaktadır. Uygulamalarda Dalgacık türü seçilirken Dalgacığın şekline ve analiz edebilme yeteneğine bakılır (Favero ve King, 1993). Çalışmada kullanılan dalgacık dönüşümü aşağıda verilmiştir.

Saç dalgacık ilk bilinen Dalgacık'tır. 1909 yılında Alfred Haar tarafından bulunmuştur. Bu dönüşüm çeşidi aynı zamanda Daubechies D-2 dönüşümü olarak da bilinir. Mevcut Dalgacık türleri içerisinde en basit olandır. Bunun yanında tekrarlayan (recursive) algoritma yapısına sahip olduğu için bu algoritma gerçek zamanlı uygulamalarda diğer dalgacık dönüşümlerine göre daha çok tercih edilirler.

Bu yapı aşağıdaki örnekle daha iyi anlaşılır;

- Veri dizisi ikili şekilde gruplanır ve bu ikililerin ortalamaları alınır
- Alınan ortalamalarla yeni bir dizi oluşturulur
- Oluşturulmuş yeni dizi orijinal dizinin yarısını oluşturur

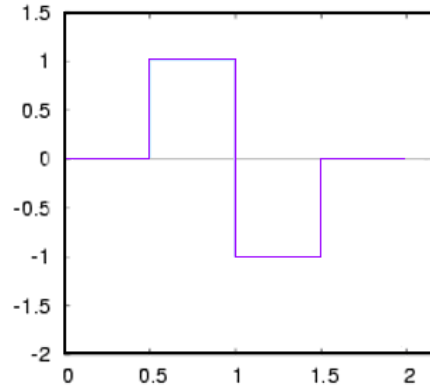
- Diğer yarısı ikili grupların ilk elemanından ortalama değerleri çıkartılarak hesaplanarak dizinin diğer yarısı doldurulur
- Aynı işlemler yeni oluşturulmuş diziye de uygulanır
- İlk adım:

$$\begin{array}{c}
 [ 3 \ 5 \ 4 \ 8 \ 6 \ 2 \ 4 \ 5 \ 7 \ 9 ] \\
 \text{ortalama} \quad \swarrow \quad \searrow \quad \text{fark} \\
 [ 4 \ 6 \ 4 \ 5 \ 8 \ -1 \ -2 \ 2 \ -1 \ -1 ]
 \end{array}$$

- İkinci adım:

$$\begin{array}{c}
 [ 4 \ 6 \ 4 \ 5 \ 8 \ -1 \ -2 \ 2 \ -1 \ -1 ] \\
 \text{ortalama} \quad \swarrow \quad \searrow \quad \text{fark} \\
 [ 5 \ 5 \ 4 \ 0 \ -1 \ -1 \ -1 \ 4 \ -2 \ 0 ]
 \end{array}$$

Şeklinde hesaplanır.



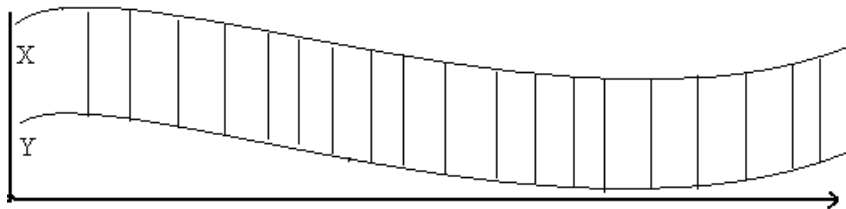
Şekil 14. Haar Wavelet'inin görünümü

## 1.6. Sesbirim ile Ses Sinyalinin Eşlenmesi

Belirli bir sesbirimin belirli bir ses parçasının içerisinde aranması yapılırken kullanılan yöntemlerden biride eşleme yöntemidir. Çalışmada eşleme işlemi kullanırken iki algoritma kullanıldı, bu algoritmalar sırası ile Çapraz Korelasyon (Cross Correlation) ve Dinamik Zamanlı Çarpıtma (Dinamik Time Warping) yöntemlerinden yararlanılmıştır. İlk olarak çapraz korelasyon yöntemi kullanılmıştır. Bu yöntem hızlı bir algoritmaya sahip olmasına karşın karşılaştırılma yapılacak iki sinyalde gürültü veya kayma söz konusu ise eşleme doğruluk oranı istenilen veya beklenen düzeyde gerçekleşmemiştir. (Şen,2007). Bu yöntemle sonucu temel alarak bilinen iki sinyalin karşılaştırılmasında kayma veya gürültünü olduğu anlaşılabilir fakat bu kayma veya gürültünün nerede olduğu belirlenemez. Bahsi geçen bu eksiklikler göz önüne alınarak çalışmaya eşleme işlemlerinde daha başarılı olduğu bilinen dinamik zamanlı çarpıtma yöntemi ilave edilmiştir. Dinamik zamanlı çarpıtma yöntemi karşılaştırılacak iki sinyalde kayma veya gürültü varlığı olsa bile belirli bir düzeyin üzerinde başarılı bir eşleme yapmaktadır. Bu yöntemleri daha detaylı bilgisi aşağıda verilmiştir.

### 1.6.1. Çapraz Korelasyon

Çalışmada eşleme işlemi için kullanılan yöntemlerden biri çapraz korelasyon (Cross correlation) yöntemidir. Bu yöntem temel olarak iki sinyalin ne kadar ilişkili olduğunu tespit etmekte kullanılır. Bu yöntem aşağıdaki şekil deki gibi eşleme yapar;



Şekil 15. X ve Y gibi iki sinyal sinyalin koordinat sistemine yerleştirilerek temsili olarak çapraz korelasyon yapılmasının grafiksel olarak gösterilmesi

Örnekte görüldüğü üzere çapraz korelasyon yöntemi X sinyalinin her bir elemanını Y sinyalinin her bir elemanının eş x-ekseni noktasında karşılaştırmaya sokarak bir benzerlik değeri üretir. Bu üretim safhasında aşağıdaki denkliği kullanır; ( Guang Gong)

$$\gamma(\tau) = \frac{E\{[x(t) - \mu_x] - [y(t) - \mu_y]\}}{\sigma_x \sigma_y} \quad (5)$$

Burada  $\gamma(\tau)$  devamlı çapraz korelasyonu temsil eder.  $\mu_x$  ve  $\mu_y$  X ve Y sinyallerinin ortalamalarını ifade eder.  $E\{\cdot\}$  beklenen değeri ifade eder.  $\sigma_x$  ve  $\sigma_y$  eğerleri ise standart sapmayı ifade etmektedir.

Ayrık çapraz korelasyon  $\gamma(\tau)$  aşağıdaki denklemle yapılandırır:

$$\gamma_k = \sum_{t=1}^N \frac{(x_t - \bar{x})(y_{t-k} - \bar{y})}{N * \sigma_{x,k} \sigma_{y,k}} \quad (6)$$

Burada  $k=[0,1,\dots,(N-1)]$ .  $\bar{x}$  ve  $\bar{y}$  iki serinin ortalamalarını gösterir ve bu değerler aşağıdaki eşitliklerle hesaplanır:

$$\bar{x} = \frac{1}{N} \sum_{t=1}^N x_t \quad (7)$$

$$\bar{y} = \frac{1}{N} \sum_{t=1}^N y_t \quad (8)$$

Standart sapmalar  $\sigma_{x,k}$  ve  $\sigma_{y,k}$  hesaplanmaları ise aşağıdaki eşitliklerde verilmiştir:

$$\sigma_{x,k} = \left[ \frac{1}{N} \sum_{t=1}^{N^*} (x_t - \bar{x})^2 \right]^{1/2} \quad (9)$$

$$\sigma_{y,k} = \left[ \frac{1}{N} \sum_{t=1}^{N^*} (y_t - \bar{y})^2 \right]^{1/2} \quad (10)$$

Bu eşitlikte  $\gamma(\tau)$  ifadesi karşılaştırmadaki korelasyon katsayısı olarak değerlendirilir ve bize iki sinyal arasındaki benzerliği 0 ile 1 değerleri arasında verir. Denkliklerden ve şekilden anlaşılacağı üzere burada kullanılan sinyallerin eleman sayıları eşit olmalıdır. Eğer eşit değilse eksik olan sinyal eşitlik değerini bozmayacak şekilde doldurulmalıdır. Bu yapının dezavantajları tanımlamalardan ve şekilden de anlaşılacağı üzere kayma var ve bilinmiyorsa veya sinyallersen biri gürültü içeriyorsa o zaman eşleme sonucu istenilen değeri üretemeyecektir. Eğer bu iki sinyalin ne olduğunu biliyorsa o zaman bu işlem neticesinde sinyallerde kayma veya gürültü olup olmadığı bu yöneltme belirlenebilir. Fakat bu çalışmada ses verilerinde ki kayma varlığı hakkında herhangi bilgi yoktur veya ses verilerininin gürültü içerip içermediği ve eğer içeriyorsa sesin hangi kısmında içerdiği bilinmediğinden bu yöntem gürültülü koşullarda veya kayma olan verilerin karşılaştırılmasında beklenen düzeyde başarı sağlayamadığı görüldü. Bunun neticesinde çalışmaya DTW (dinamik zamanlı çarpıtma ) yöntemi eklendi.

### 1.6.2 Dinamik Zamanlı Çarpıtma

N boyutlu uzayda  $\mathbf{x} = [x_1, x_2, \dots, x_n]$  ve  $\mathbf{y} = [y_1, y_2, \dots, y_n]$  gibi iki nokta arasındaki fark ölçüm modeli en popüler yöntemlerden biri Euclidean'dır ve formülüzasyonu aşağıda ki gibidir (E. A. Yfantis ve T. Lazarakis A);

$$\text{Fark}(\mathbf{x}, \mathbf{y}) = |\mathbf{x} - \mathbf{y}| = \left[ (x_1 - y_1)^2 + (x_2 - y_2)^2 + \dots + (x_n - y_n)^2 \right]^{1/2} \quad (11)$$

Fakat bu yöntem x'in ve y'nin eleman sayıları eşit olduğunda kullanılabilir aksi durumda veya durumlarda bu yöntem kullanılamaz. Farklı boyutlardaki x'in ve y'nin

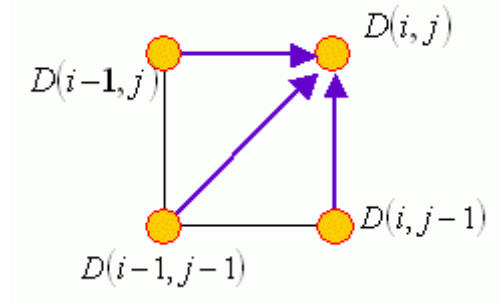
farklarının ölçülmesi için daha esnek bir yöntem ihtiyacı vardır. Bu esnekliği bize sunan yöntemlerin başında Dinamik Zamanlı Çarpıtma (DTW – Dynamic Time Warping) yöntemi gelmektedir. Bu yöntemle iki sinyal arasındaki fark ortaya çıkartarak araştırmacıya sinyallerin benzerlikleri hakkında yorum yapma olanağı sunar. Bu yöntemle iki sinyalin birbirine benzerliğini veya farkını bulunabilir. Yöntem çapraz korelasyondan farklı olarak yapısında sezgisel bir yapı barındırarak daha ayrıntılı bir eşleme sonucu sinyaller arasındaki farkı bulur. Bu yapıda benzerlik değil fark elde edilir. Aslında bu iki kavram birbirinin tümleyeni olduğundan ve birinin bilinmesi ötekini de bilinmesine olanak verdiğinden sonuçta iki sinyalin benzerliğine bu yöntem yardımı ile de erişilir.

Bu yapının çalışma mantığı şu şekilde verilebilir;

İki adet  $t$  ve  $r$  gibi iki vektör verilmiş olsun ve bu vektörlerin uzunlukları sıra ile  $m$  ve  $n$  olsun. DTW'nin amacı bu iki vektör arasındaki minimum yol haritasını  $\{(p_1, q_1), (p_2, q_2), \dots, (p_k, q_k)\}$  bulmaktır öyle ki bu yol haritasındaki minimum yol  $\sum_{i=1}^k |p_i - r(q_i)|$  minimum olacak şekilde olmalıdır ve şu kısıtlamaları izlemelidir:

Sınır koşulu:  $(p_1, q_1) = (1, 1)$ ,  $(p_k, q_k) = (m, n)$ . Bu başlangıç *çapası*'na ve *bitiş çapası*'na tipik bir örnektir.

Yerel sınır: Yoldaki herhangi bir düğüm  $(i, j)$ ,  $(i-1, j)$ ,  $(i, j-1)$ ,  $(i-1, j-1)$  ile sınırlandırılmıştır. Bu yerel sınırlar yolun haritalanmasında ilk ve ikinci değerlendirilmesinde monoton olarak azalmayı garanti eder. Bunun haricinde,  $t$ 'den gelen herhangi bir eleman için en az ilişkili bir eleman bulunabilir tabi olarak bunun akside geçerlidir.



Şekil 16. Yerel yolun olası yol haritası



DTW için en uygun değer olacak yol haritası çizmek için aşağıdaki yol izlenir;

Optimum değer fonksiyonu:  $D(i,j)$ 'nin  $t(i,j)$  ve  $r(i,j)$  arasındaki DTW farklarının  $(1,1)$ 'den  $(i,j)$ 'ye kadar olan kısmın belirlenmesi

Tekrarlama:  $D(i,j) = iyt(i) - r(j)iy + \min\{D(i-1,j), D(i-1,j-1), D(i,j-1)\}$ ,

$D(1,1) = iyt(1) - r(1)iy$  olan başlangıç koşulu ile bulunması.

Son sonuç:  $D(m,n)$

Uygulama esnasında  $m \times n$  boyutlu ve  $D(1,1)$  başlangıç koşulu ile oluşturulmuş  $D$  matrisi gerekmektedir. Bu matris üretildikten sonra tekrarlamalı formülle bütün matris elemanlarına değerler sütundan sütuna ve satırdan satıra olacak şekilde sırayla üretilir. Sonuç  $O(m \cdot n)$  hesaplama karmaşıklığı ile  $D(m,n)$ 'e erişilebilir.

Eğer optimum yolun yanında minimum farkı da öğrenmek istiyorsak her bir düğüme gelen optimum giriş yelpazesi tutulur. Verilerin işlenmesinden sonra hızlı bir şekilde geri iz arama yapılması kaidesiyle iki vektör arasında ki optimum yol haritası bulunabilir. DTW için geri yön algoritması şu şekildedir

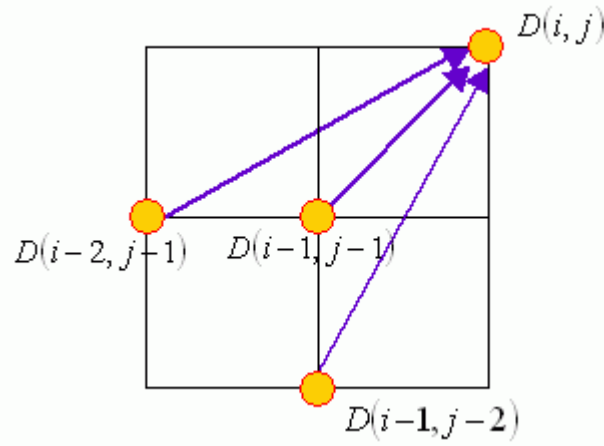
Optimum değer fonksiyonu:  $D(i,j)$ 'yi  $t(i:m)$  ve  $r(j:n)$  arasındaki DTW farkı  $(i,j)$ 'den  $(m,n)$ 'e olan yol haritası kullanılarak hesapla.

Tekrarlama:  $D(i,j) = iyt(i) - r(j)iy + \min\{D(i+1,j), D(i+1,j+1), D(i,j+1)\}$ , bu hesabı yaparken  $D(m,n) = iyt(m) - r(n)iy$  başlangıç koşulu kullanılır.

Son sonuç:  $D(1,1)$

Geri yön algoritmasının sonucu ileri yön algoritmasıyla aynı çıkabilir.

Yaygın olarak kullanılan bir başka sırlandırmada  $27 \times 45 \times 63$  şeklindedir ve aşağıda verilmiştir:

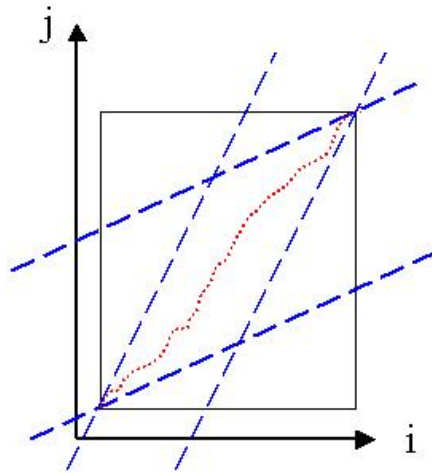


Şekil 17. Yerel sınırlandırmanın matristeki oluşumu

Bu yerel yol sınırlandırması aşağıdaki sonuçları oluşturabilir:

- Bazı giriş verileri yol haritasında atlanabilir. Bu giriş verilerinin atlanması eğer ki giriş verileri ardı ardına gelecek şekilde oluşmuşsa avantajlı bir durumdur.
- Optimum değer fonksiyonu toplam farkın üzerine kurulu olduğundan ötürü bazen yol haritası yerel yolda  $45\phi X$  derecelik açı yerine  $27\phi X$  veya  $63\phi X$  açı değerlerini alarak toplam farkı en küçük değeri almasını dener.

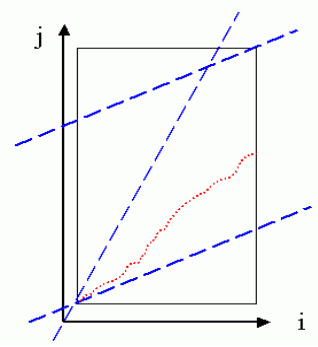
Eğer gerek duyulan yol “çapalı başlangıç, çapalı bitiş” ise o zaman yerel yol sınırlandırmaları genel yol sınırlarını eyleme geçirir, aşağıdaki şekilde görüldüğü gibi;



Şekil 18. Genel sınırlandırmanın yol haritasını sınırlamasının koordinat sisteminde gösterimi

Şekil 18.'de olurlu bölge paralel kenarla gösterilmiştir. Matrisi doldururken hesaplamaların kaydedilmesi paralel kenarın dışındaki elemanların atlanmasıyla yapılabilir. Buna ek olarak mümkün olan yol ancak ve ancak giriş vektörlerinin uzunluklarının oranları sonucu 0,5 ile 2 arasında ise var olur.

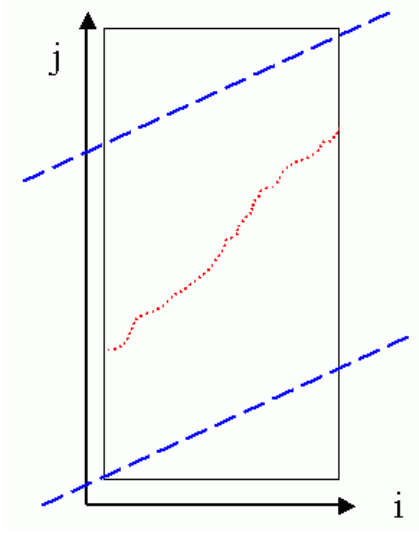
Eğer ki gerek duyulan yol haritası “çapalı başlangıç, serbest bitiş” şeklinde ise o zaman genel yol sınırlandırmaları Şekil 19.'da verilmiştir:



Şekil 19. Çapalı başlangıç, serbest bitiş sınırlandırmasının koordinat sisteminde gösterimi

Olurlu bölge yamuk içerisinde şekilde gösterilmiştir. Bu “çapalı başlangıç, serbest bitiş” yöntemi “çapalı başlangıç, çapalı bitiş” yöntemi ile karşılaştırıldığında görülüyor ki buradaki hesaplama ve kaydetme daha azdır.

Eğer ki yol haritası “serbest başlangıç, serbest bitiş” gerektiriyor ise genel yol sınırlandırması aşağıdaki şekilde verilmiştir.



Şekil 20. Serbest başlangıç, serbest bitiş yapısının sınırlandırmasının koordinat sisteminde gösterimi

Olurlu bölge yukarıdaki şekilde paralelkenar olarak yukarıdaki şekilde gösterilmiştir. Bu “serbest başlangıç, serbest bitiş” yöntemi “çapalı başlangıç, serbest bitiş” yöntemi ile karşılaştırıldığında görülüyor ki bu yöntemde hesaplama ve kaydetme daha azdır. Bu çalışmada kullanılan DTW yönteminde kullandığımız genel yol sınırlandırması da “serbest başlangıç, serbest bitiş” yöntemidir.

Not olarak vermek gerekirse genel yol sınırları  $27\phi X$ ,  $45\phi X$ ,  $63\phi X$  açılı yerel yol sınırları olarak gerçekleştirilebilir. Bununla birlikte bazen  $0\phi X$ ,  $45\phi X$ ,  $90\phi X$  açılı yerel yol sınırları kullanılmasıyla yol haritasına makul bir şekil verilmek amacıyla genel yol sınırlarının eklenebilir. Özet olarak genel yol sınırlarının kullanılması iki amaca yardımcı eder:

- Hesaplama yükünün azaltılması
- Yol haritasını makul bir yol ile sınırlandırmak

Aşağıda bu yapıya uygun iki örnek verilerek iki DTW'nin nasıl gerçekleştirildiği anlatılmaya çalışılacak:

- Birinci tip dtw  $27\phi X$ - $45\phi X$ - $63\phi X$  yerel yol sınırları ile gerçekleştirilen
- İkinci tip dtw  $0\phi X$ - $45\phi X$ - $90\phi X$  yerel yol sınırları ile gerçekleştirilen

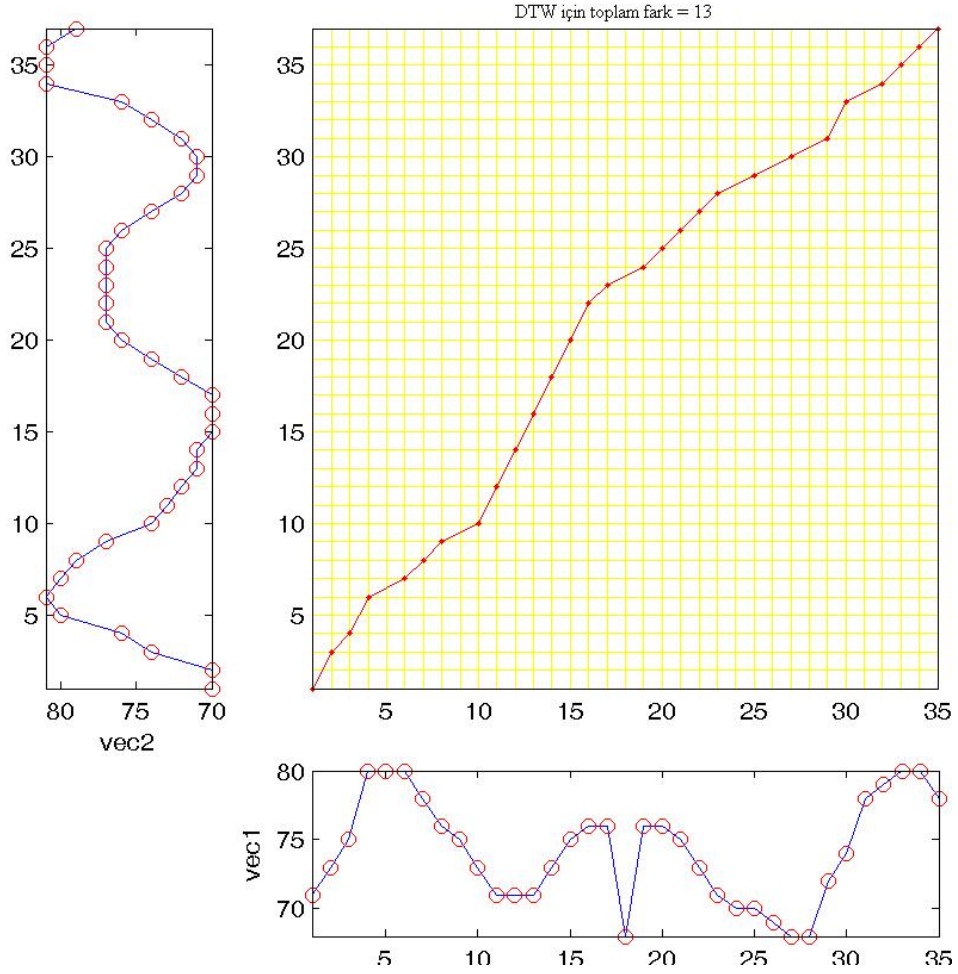
İlk örnek birinci tip dtw ile gerçekleştirilen:

Giriş verileri:

vek1=[71 73 75 80 80 80 78 76 75 73 71 71 71 73 75 76 76 68 76 76 75 73 71 70 70 69  
68 68 72 74 78 79 80 80 78];

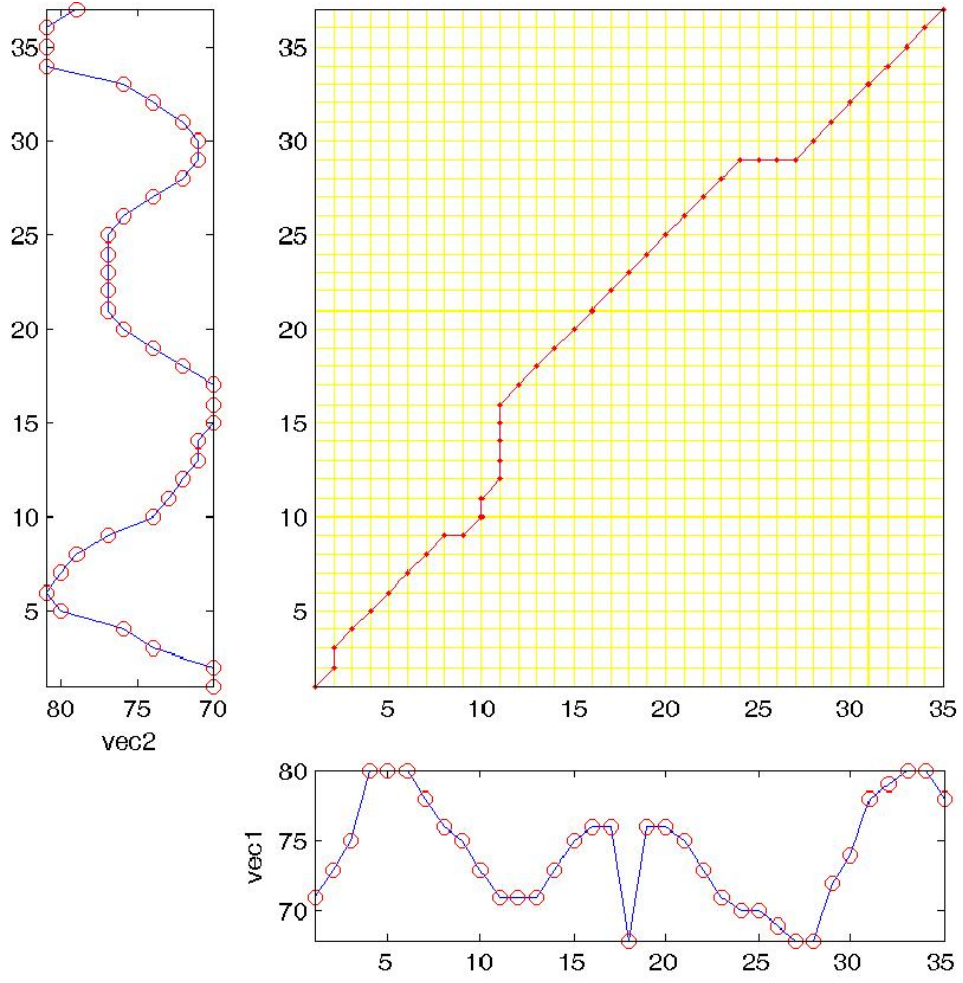
vek2=[69 69 73 75 79 80 79 78 76 73 72 71 70 70 69 69 69 71 73 75 76 76 76 76 76 75  
73 71 70 70 71 73 75 80 80 80 78];

Sonuç grafiksel olarak;



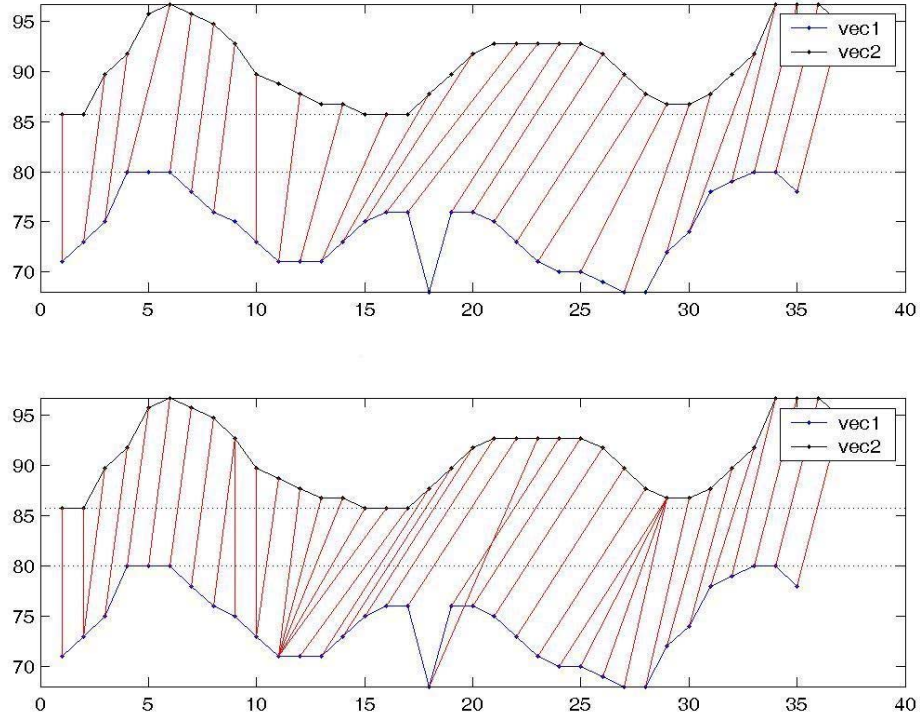
Şekil 21. İki vektörün DTW ile hesaplanmasını grafiksel olarak anlatımı

Aynı giriş verileri kullanılarak ikinci yöntemin kullanıldığı DTW Şekil 22.'de verilmiştir.



Şekil 22. İki vektörün DTW ile hesaplanmasını grafiksel olarak anlatımı

Aşağıda Şekil 23.'de DTW yönteminin çalışma mantığını daha iyi anlaşılması eşleşmenin nasıl yapıldığının grafiksel olarak anlatımı verilmiştir.



Şekil 23. İki farklı vektörün dtw aracılığı ile hesaplanmasındaki eşleme yapılmasının grafiksel anlatımı

### 1.7. Konuşma Kısımının Kayıt İçinde Tespit Edilmesi

Yapılan çalışmada kelime belirleme algoritmasının tüm aşamaları uygulanmamıştır. Bu yüzden kelime belirleme (word detection) demek yerine konuşma kısmının kayıt içinde tespit edilmesi olarak ifade edilmiştir. Çalışmada, kelime belirleme algoritmasının üç aşaması kullanılmıştır. Bu safhalar sırası ile başlangıç-noktası belirleme (Start-point Detection), bitiş-noktası belirlenmesi (End-point Detection) ve sıfır geçiş oranı (Zero Crossing Rate) olarak söylenebilir.

Konuşma kısmının kayıt içinde tespit edilmesi sisteme olan katkısı iki şekilde açıklanabilir. Birincisi, fazla işlem yapılması engellenerek sistem hızlandırılmış olur ikinci olarak ise sadece konuşma kısımlarının karşılaştırmada kullanılması karşılaştırma işleminin doğruluk oranını olumlu şekilde artırması olarak açıklanabilir. Konuşmanın belirlenmesinde öncelikle eşik değeri (Threshold) ve enerji düzeyleri belirlenmiştir. Bu değerler kullanılarak arka plan gürültüsü elimine edilmiştir. Zero Crossing Rate (Sıfır geçiş oranı) ve Time feature Energy (zaman enerji özelliği) (G.S. Ying, C.D. Mitchell,

*L.H. Jamieson*)değerlerinin belirlenmesi gerçekleştirilmelidir. Bu aşamanın ardından bulunan değerlerin yardımıyla alınan pencerede konuşmanın başlangıç ve bitiş değerleri belirlenmesi yapılmalıdır. Sırası ile bahsi geçen yapıların formülüzasyonunu aşağıda verilmiştir:

### 1.7.1. Sıfır geçiş oranı (Zero Crossing Rate)

$$ZCR = 0,5 \sum_{i=1}^N (s(i+1) - s(i)) \quad (12)$$

Burada N pencerenin eleman sayısıdır (bu projede sabit bir değer yoktur. Çerçeveler sesbirimlerin kendisi olarak seçilmiştir) ve s (i) ses işaretidir. Bu ana formül yardımıyla  $Z_N$  (pencere ZCR değeri),  $Z_f$  (ileri yön ZCR),  $Z_b$  (geri yön ZCR) değerlerinin hesaplanması aşağıdaki gibidir (J. G. Wilpon, L. R. Rabiner, and T. Martin);

$Z_f = \{Z_1 + Z_2 / 2\}$  eğer ki  $0,5 < Z_1 / Z_2 \leq 2$  aksi takdirde  $Z_f = \min (Z_1 / Z_2)$  olarak değerlendirilir.

$Z_b = \{Z_k + Z_{k-1} / 2\}$  eğer ki  $0,5 < Z_k / Z_{k-1} \leq 2$  aksi takdirde  $Z_b = \min (Z_k / Z_{k-1})$  olarak değerlendirilir

$Z_N = \{Z_f + Z_b / 2\}$  eğer ki  $0,5 < Z_f / Z_b \leq 2$  aksi takdirde  $Z_N = \min (Z_f / Z_b)$  olarak değerlendirilir.

### 1.7.2. Enerji Seviyelerinin Tespiti

Enerji seviyesi hesaplanmasında aşağıdaki formül.13 kullanılır,

$$E_k = \sum_{n=((k-1)F/2)+1}^{(k+1)F/2} (S_n)^2 \quad (13)$$



$$k= 1,2\dots K$$

F= Her pencerenin elemen sayısı

K= Toplam pencere sayısı

Bu formül ana denklemdir ve  $E_f$ 'nin (ileri yön enerji) ve  $E_b$ 'nin (geri yön enerji) hesaplanmasında da kullanılacaktır.  $E_f$  ve  $E_b$  nin hesaplanması (12) yardımıyla aşağıdaki gibidir;

$$E_f = \{E_1 + E_2 / 2\} \text{ eğer ki } 0,5 < E_1/E_2 \leq 2 \text{ aksi takdirde } E_f = \min(E_1, E_2) \text{ değerini alır.}$$

$E_b = \{E_k + E_{k-1} / 2\}$  eğer ki  $0,5 < E_k/E_{k-1} \leq 2$  aksi takdirde  $E_b = \min(E_k, E_{k-1})$  değerini alır.

$E_N = \{E_f + E_b / 2\}$  eğer ki  $0,5 < E_f/E_b \leq 2$  aksi takdirde  $E_N = \min(E_f, E_b)$  değerini alır ve bu tüm pencerenin enerjisini temsil eder.

### 1.7.3. Konuşmanın Başlangıç ve Bitiş Yerinin Tespiti

Belirlenmiş ZCR ve enerji seviyeleri yardımıyla konuşmanın başlangıç ve bitiş yerinin tespiti  $P_{f3}$ ,  $P_{b3}$ ,  $P_{f2}$ ,  $P_{b2}$  değerleri ile hesaplanır. Bu değerlerin elde edilme yöntemleri aşağıda verilmiştir;

$P_{f3} = \operatorname{argmin}\{E_k > T_E, k=1,2,3\dots K\}$  ve  $T_E = C_E * E_N$  olarak hesaplanır. ( $C_E$  denemeler sonucunda elde edilen katsıyı)

$$P_{b3} = \operatorname{argmax}\{E_k > T_E, k=K, K-1, \dots, 1\} \text{ geriye doğru olarak hesaplanır.}$$

$P_{f3} - P_{b3} < t_{\min}$  olduğu takdirde bu aralıkta konuşma yok denilebilir ve pencere işleme sokulmayabilir.

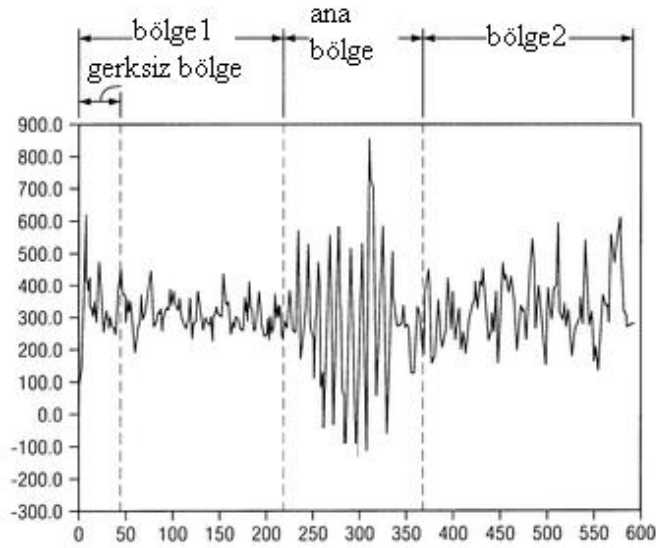
Enerji katsayılarıyla belirlenen aralık daha ayrıntılı ve kesin olarak ZCR katsayılarıyla hesaplanması aşağıdaki şekilde gerçekleşir;

$$P_{f2} = \operatorname{argmin}\{Z_k > T_{Zf}, k=P_{f3}, P_{f3-1}, \dots, 1\} \text{ ve } T_{Zf} = C_{Zf} * Z_N$$

hesaplanır. ( $C_{Zf}$  denemeler sonucunda elde edilir)

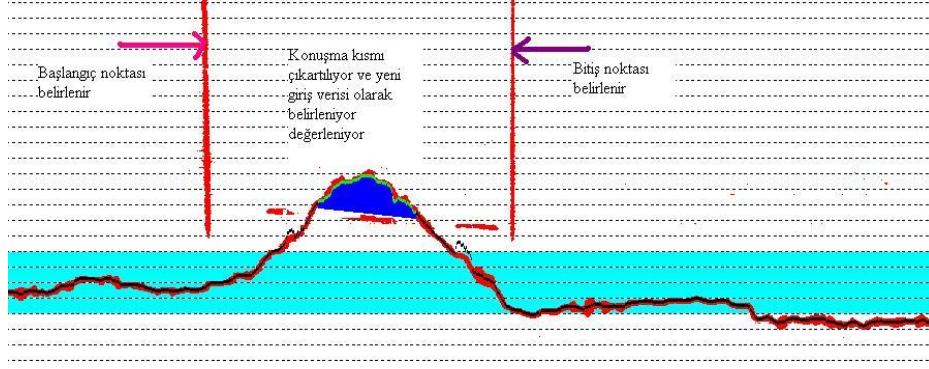
$P_{b2} = \operatorname{argmax}\{Z_k > T_{Zb}, k=P_{b3}, P_{b3+1}, \dots, K\}$  bulunur ve  $T_{Zb} = C_{Zb} * Z_N$  hesaplanır. ( $C_{Zb}$  denemeler sonucunda bulunur)

Sonuç olarak, elde edilen konuşma aralığı  $P_{f2}$  ve  $P_{b2}$  olarak belirlenmiş olur. Bu değerler aynı zamanda konuşma kısmının elde edilmesini sağlar. Bu sınırlar yeni verinin elemanlarının başlangıç ve bitiş değerlerini vermiş olur ki bu konuşma temel olarak gürültüden veya boşluktan arındırılmış olur. Bu ise ses tanımayı daha etkin kılar. Formülüzasyonda kullanılan  $C_E$ ,  $C_{Zf}$  ve  $C_{Zb}$  katsayıları bulunurken sistemde önce konuşma olmadan ses kaydı yapıldı ve sonra kayıtlardan biri rasgele seçilerek onun maksimum değeri bulundu ve bu iki örneğinin değerlerinin ortalamaları alınarak belirlendi. Burada amaç bu değişkene bağlı eşik değeri belirleme sistem üzerinde minimum düzeyde olumsuz etkisi yapsın. Şekil 24.'de bu işlemin gerçekleşmesini sağlayan bir görüntü üzerinde yön bilgileri verilmiştir:



Şekil 24. Bir ses parçasının aşamalarının grafiksel olarak anlatımı

Bir ses parçasığı alınarak sesin başlangıç ve bitiş yerleri Şekil 25.'de verilmiştir.



Şekil 25. Başlangıç ve bitiş noktalarının aranmasının grafiksel anlatımı

## 2. YAPILAN ÇALIŞMALAR

Bu bölümde yapılan çalışma adım adım anlatılacaktır. Proje başlangıcında üzerinde çalışılacak olgu ses olduğundan, öncelikle bir ses editörü yapılmıştır. İlk editör ses kayıt işlemi için gerçekleştirilmiştir. İkinci editör ise, kaydedilen seslerin incelenebilmesi için yapılmıştır. İkinci editör vasıtasıyla sesin istenilen kısmı üzerinde işlemler yapılabilir hale getirilmiştir. Seçilen ses parçası büyütülüp ayrıntılı biçimde incelenebilmektedir. Böylece, sesbirimlere daha yakından bir inceleme gerçekleştirilmiştir.

Sesbirimler, ses editörleri vasıtasıyla yapılan kayıtlarda inceleme yapılarak bulunmuştur. Sesbirimler yapısal olarak incelendikten sonra kayıtlardan dikkatli bir şekilde alınarak sesbirim kütüphanesi oluşturulmuştur. Kütüphanenin içeriği, 10 farklı bayan, 10 farklı erkek ve 10 tane farklı zamanda alınmış radyodan kayıtlardan oluşmuştur.

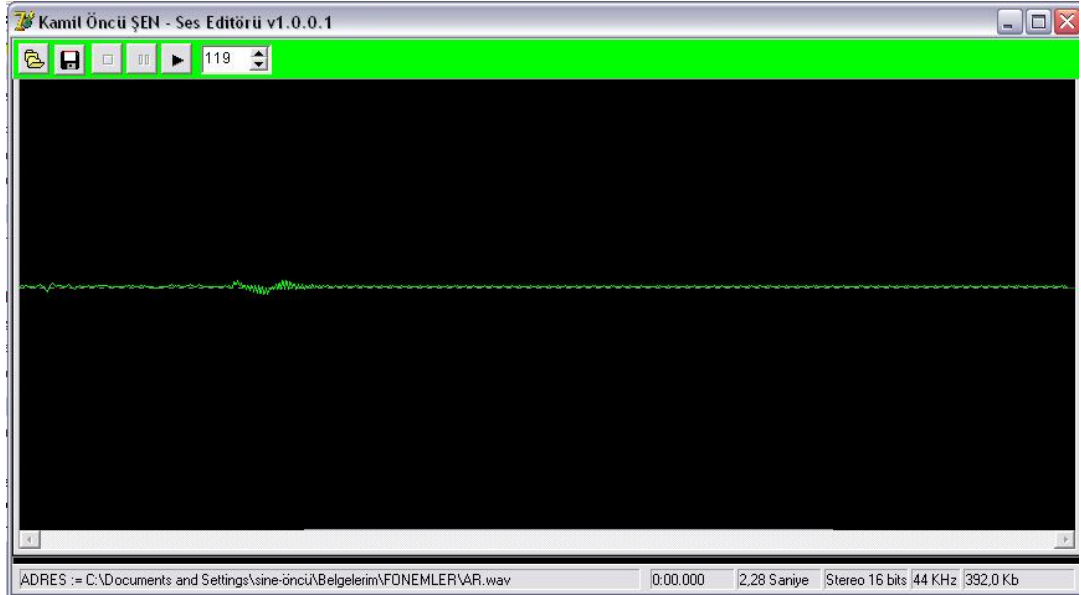
Çalışmada sesbirimler, öncelikle yalın kayıt alınmış ve karşılaştırmalarda bu şekilde kullanılmıştır. Fakat yalın alınan kayıtlardan sadece yapısal analiz yapılabileceği anlaşılmıştır. Çünkü aynı konuşmacıdan alınan kayıtların içerisindeki sesbirim arama işleminde istenilen başarıya ulaşamadığı gözlenmiştir. Bu başarısızlık, sesbirimlerini oluştururken kelimelerden yapılmış ses kayıtlarından sesbirimleri elde etmeye yönlendirmiş ve bu yöntemle sesbirim kütüphanesi oluşturulmuştur. Bu işlem yapılırken önce sesi harflerle başlayan kelimeler seçilmiştir. Böylece, sesli harfler kolaylıkla kelimedenden ayrılarak yalın bir halde sesli sesbirimler olarak kütüphanesi oluşturulmuştur.

Sessiz harfler ise benzer bir yol izlenerek belirlenmiştir. Kayıt edilecek sessiz harfle başlayan bir kelime seçilerek sessiz harf sesbirimi elde edilmiştir. Daha önce belirlenmiş ve elde edilmiş sesli harflerin karakteristiği bilindiğinden, sessiz sesbirim çıkartılma işlemi daha kolay bir şekilde gerçekleştirilmiştir.

Şekil 26. ve 27.'da ses kaydetme ve ses editörünün ekran görüntülerini göstermektedir.



Şekil 26. Ses kayıt işlemini gerçekleyen programın kullanıcı ara yüz görüntüsü



Şekil 27. Ses editörü olarak yapılmış programın kullanıcı ara yüz görüntüsü

Çalışmada ses birimler elde edildikten sonra karşılaştırma işleminde kullanılacak olan kayıtların elde edilmesine gelinmiştir. Bu kayıtlar ilk olarak belirli metinler seçilmesi esasına dayandırılmıştır fakat sonra bu sistemin yavaş olmasından ötürü kelimelerde sesbirim aramak olarak değiştirilmiştir. Bir başka deyişle arama yapılacak örneklem boyutu küçültülmüştür. Temel olarak aynı işlemler gerçekleşeceğinden dolayı da amaçtan da bir sapma olmamıştır sadece daha küçük ölçekli örnekler kullanılmıştır. Sistemde çok fazla grafik işlemi olması ve kullanılan algoritmaların yavaş olması bu kararı vermede etkili olmuştur. Şekil 26. ve 27.'de

sesbirimleri kelimedden ayırmakta kullanılmak üzere tasarlanmış programın ekran görüntüleri verilmiştir.

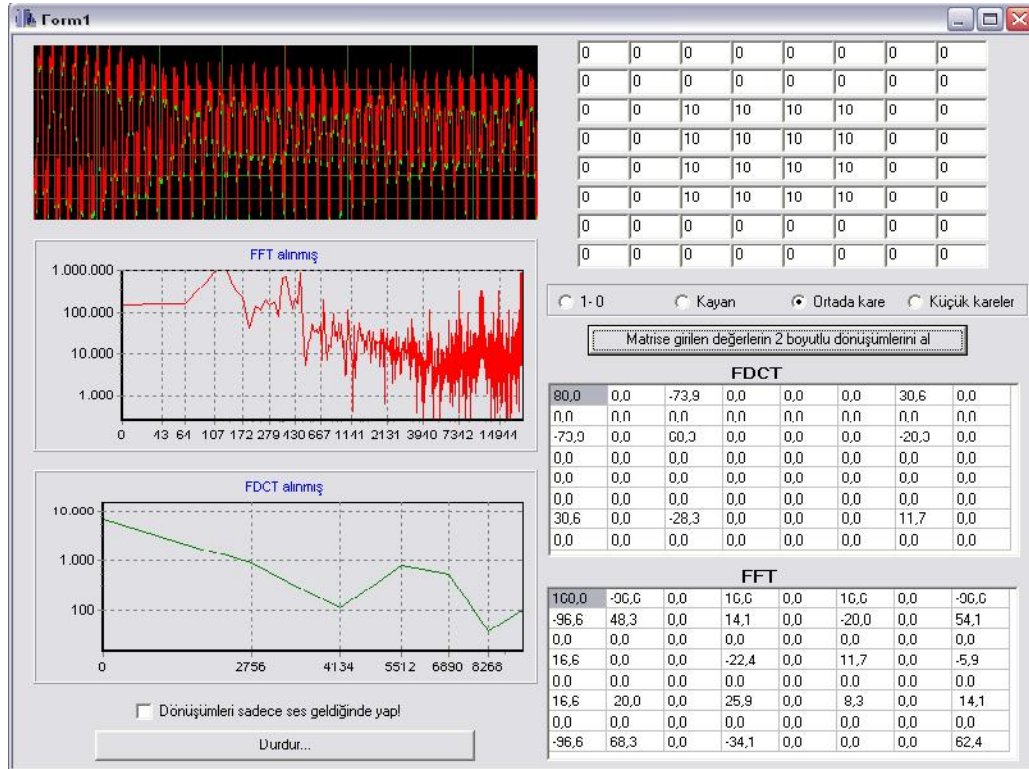
Çalışmada sadece ses editörü ile sesler ayrıştırılıp aynı zamanda göz organı yardımıyla gürültüden ayrıştırılma yapılarak kullanılan algoritmaları başarısı arttırılmaya çalışılmıştır. Ana programda kelime belirleme algoritması yöntemi ile kelimelerdeki var olan gürültü ve boşluk kısımlarının elenmesini gerçekleyerek benzerlik bulma algoritması ve fark bulma algoritmalarının eşleşme başarıları arttırıldığı gözlemiştir.

İki farklı program sayesinde önce ses verileri kaydedilip sonra bu ses verilerinden istenilen veriler elde edilmiş ve gözle basitçe gürültüden ayrıştırılmıştır. Bu işlemden sonra, kelime belirleme işlemi ana program tarafından başlatılmıştır. Bu çalışmada, verilerden aranacak olan sesbirim verisini temel alan bir çerçeveleme yöntemi gerçekleştirilmiştir. Bu çerçeveleme sınırları belli olmayan bir yapıya sahip ve seçilen sesbirimin uzunluğuna bağlıdır. Aranacak sesbirim bir bütün olarak saklanır ve sesbirimin aranacağı ses verisi içerisinde seçili sesbirim boyutunda çerçeveler alınır.

Sesbirim çerçevesi arama yapılacak ses verisi içerisinde bir bit kaydırma yöntemi ile kaydırma işlemi yapılarak arama yapılır. Bu arama yavaş görülebilir fakat çalışmada hedef farklı yöntemlerin kullanımını yaparak kullanıcıya bilgi vermek olduğundan hız önemsenmemiştir. Amaç kullanıcıya kendi belirleyebileceği bir sistem hazırlamasına olanak sağlayarak sistemi doğruluk oranına bağlı olarak kendi çalışmasına yön verebilmesini sağlamaktır. Seçilen sesbirimler çerçeve boyutlarını belirlediğinden ötürü dönüşüm aşamasında da hızlandırmaya yönelik yapılar kullanılamamaktadır çünkü işlenecek veri esnektir bu esneklikte sabit bir sistem kurarak dönüşüm algoritmalarını hızlandırmaya yönelik bir çalışma gerçekleştirilememiştir.

Çalışmaya ek olarak gerçek zamanlı çalışmalarda kullanılması ve hız performansının veri boyutu ile nasıl etkilendiğine örnek olabilmesi için bir gerçek zamanlı dönüşüm programı tasarlanmış ve uygulanmıştır. Bu çalışmada hedeflenen, mikrofondan alınan verilere iki dönüşüm uygulanması şeklindedir. Ayrıca programda birde kullanılan dönüşüm fonksiyonlarının doğru sonuç üretmediğini kontrol amaçlı bir kontrolör görevi yapacak bir sistem monte edilmiştir. Bu sistem metin kutuları yardımıyla kullanıcıya istediği 8x8'lik bir matris verisi giriş verisi olarak vermesine olanak sağlar ve bu veriler dönüşüm fonksiyonları tarafından dönüşüme tabi tutulur. Sonuçlar, bir ızgara metin kutusu yardımıyla ekrana basılır. Eğer, kullanıcı dönüşüm

için verdiği veri matrisinin dönüşümden elde edilmiş verilerine de sahip ise o zaman hazırlanmış sistemin ürettiği sonuçlarla karşılaştırarak bu dönüşüm sisteminde kullanılan fonksiyonların doğru çalışıp çalışmadığından emin olabilmesi sağlanmıştır. Bu programın ekran görüntüsü Şekil 28.'de verilmiştir.



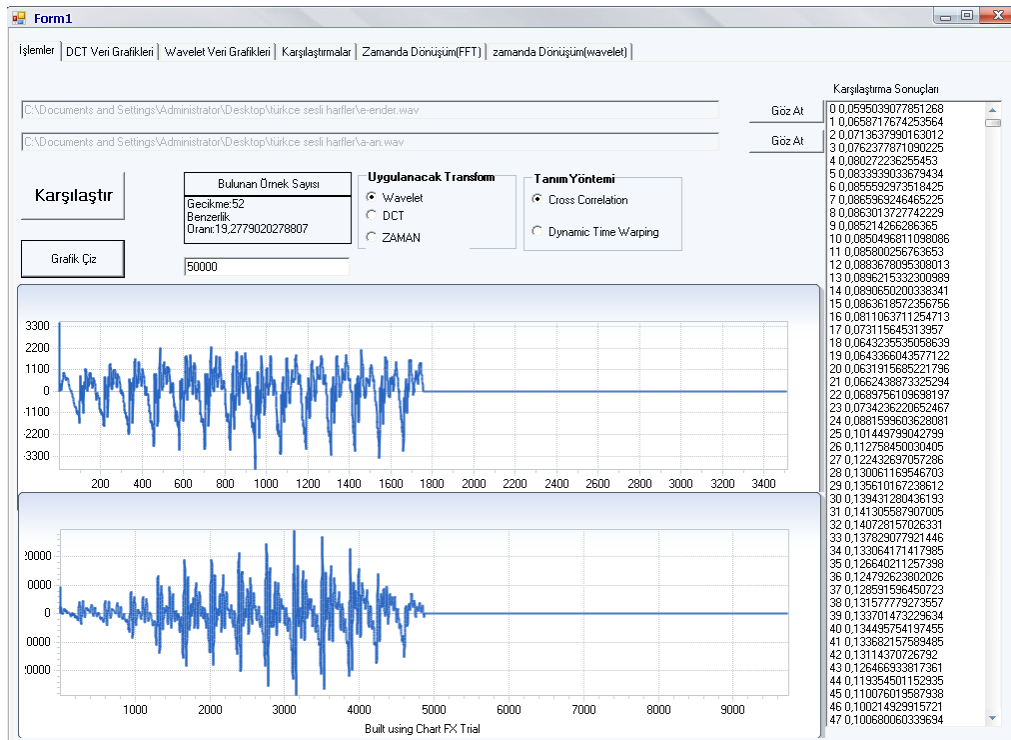
Şekil 28. Gerçek zamanlı FFT ve FDCT yapan programın ekran görüntüsü

Sistem karşılaştırma işleminde kullanacağı sesbirim verisini aldı ve sesbirimin karşılaştırılacağı ses kaydını da belirledikten sonra ilk önce kelime belirleme algoritması yardımıyla ses verisinde eğer varsa gürültü ve boşluk olan kısımlar optimize edildikten sonra çerçeveler belirlenmiştir. Bu belirleme sesbirimin boyutsal uzunluğu doğrultusunda gerçekleştirilmiştir. Belirlenen çerçeveler seçili dönüşüm yöntemine tabi tutularak zaman boyutundan frekans boyutuna dönüştürülmüştür. Bu ses verisi üzerinde işlem yaparken kolaylık sağlayacak ve doğruluk oranını olumlu yönde etkileyecektir. Eğer, kullanıcı zaman boyutunda çalışma yapmak isterse kurulan sistem buna olanak sağlamaktadır. Ses dönüşümü yapılacak ise o zaman iki yöntem kullanılır bunlar Saç Dalgacık Dönüşümü ve Fourier Dönüşümü'dür. Bu yapılar seçilmesindeki en etkin sebep bu yöntemler kendi türlerinin önde gelenleri olmalarındandır. Bu

dönüşümler kendi sınıflarındaki ana dönüşümlerdir. Diğer dönüşümler bunların türevleridir. Böylece dönüşüm olgusu için kıyaslama yapabilme olanağı sunulmuş ve sesin işlenmesi kısmı iki safhalı olarak bitirilmiştir. Sesin frekans boyutuna dönüştürülmesiyle ses verileri belli frekans karakteristiği özelliği üzerinde işlem yapabilme şansına yakalanmıştır.

Çalışmada bu safhaların kullanıcı tarafından rahatça gözlem yapabilmesine olanak sağlamak amacıyla yapılan dönüşümler oluşturulan pencerede iki boyutlu frekans-genlik bilgileriyle grafiksel ortama aktarılarak gözlemsel nitelik artırılmıştır.

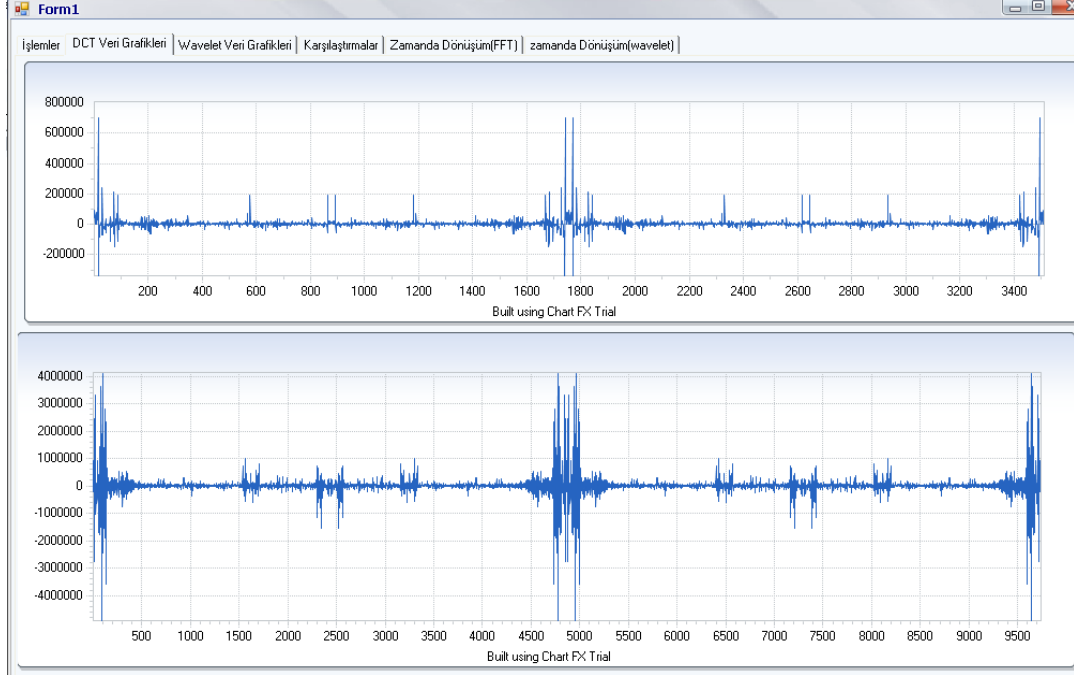
Gözlemci ham veriler için gerçekleşmiş zaman-genlik bilgilerini içeren grafiklerle ve iki boyutlu frekans-genlik dönüşüm sonuçlarını barındıran grafiklerle gözlemsel kararlarına dayanarak kendi çalışmasına yön verebilme olanağı sunulmuştur. Şekil 29.'da bu aşamaların çalışma çerçevesinde gerçekleşmiş programın "a" ünlüsü ile "e" ünlüsünün saç dalgacık dönüşümü uygulanmış ve çapraz korelasyon uygulanması sonucu aralarındaki benzerliğin belirlendiği ekran görüntüleri verilmiştir.



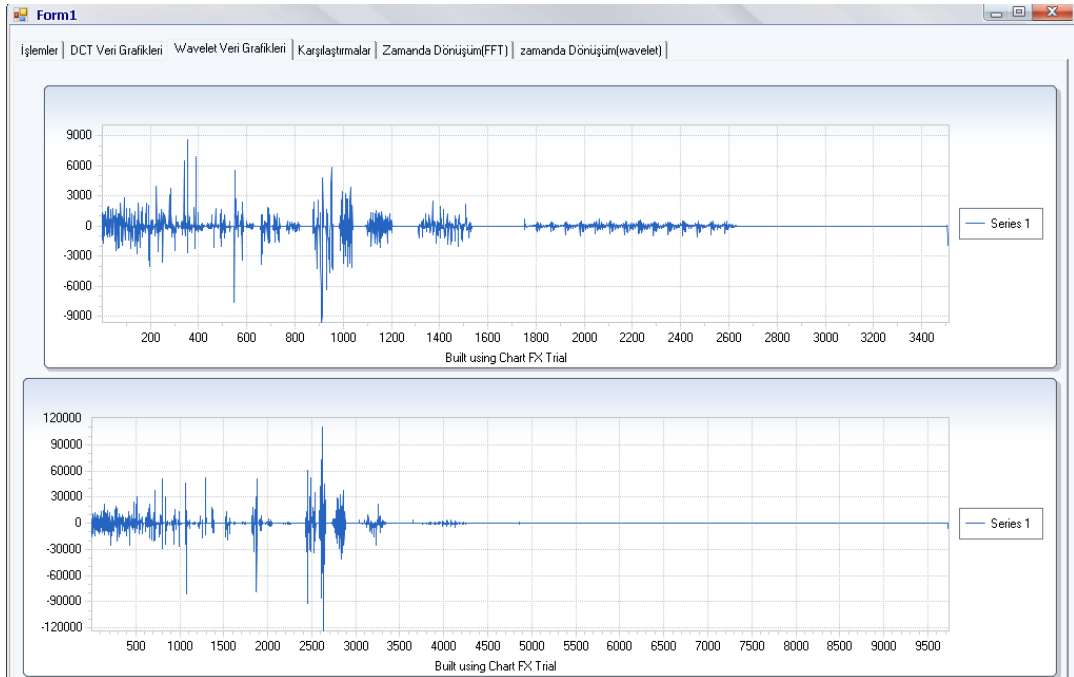
Şekil 29. Karşılaştırılan iki sesbirimin ekran görüntüsü



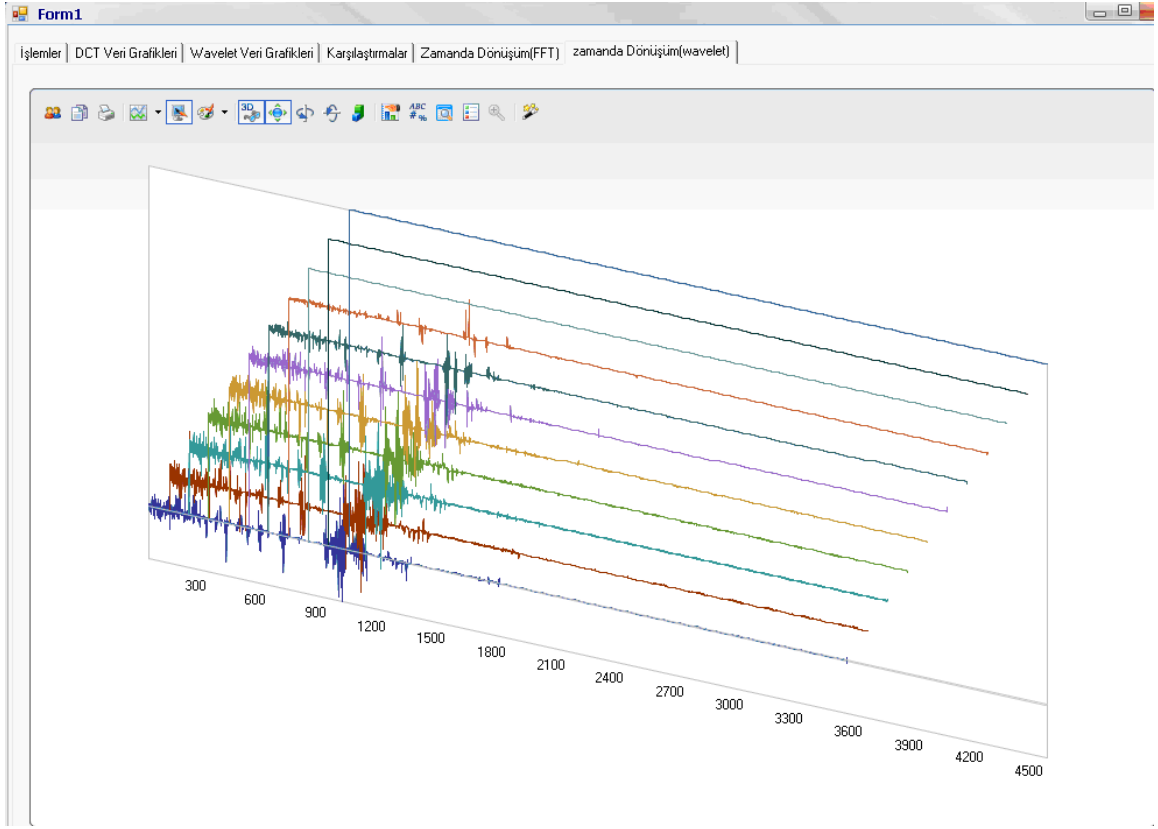
Programın ilk sekmesinde seçilen seslerin zaman-genlik (ham halinin) grafiksel gösterimi ve sonuçlar kısmı vardır. Şekil 30. ve Şekil31.'de ise bu seçilmiş ses verilerinin dönüşümden sonra oluşan formunun ekran görüntüleri verilmiştir.



Şekil 30. FFT uygulanmış verilerin grafik ortamında sunulmasının ekran görüntüsü

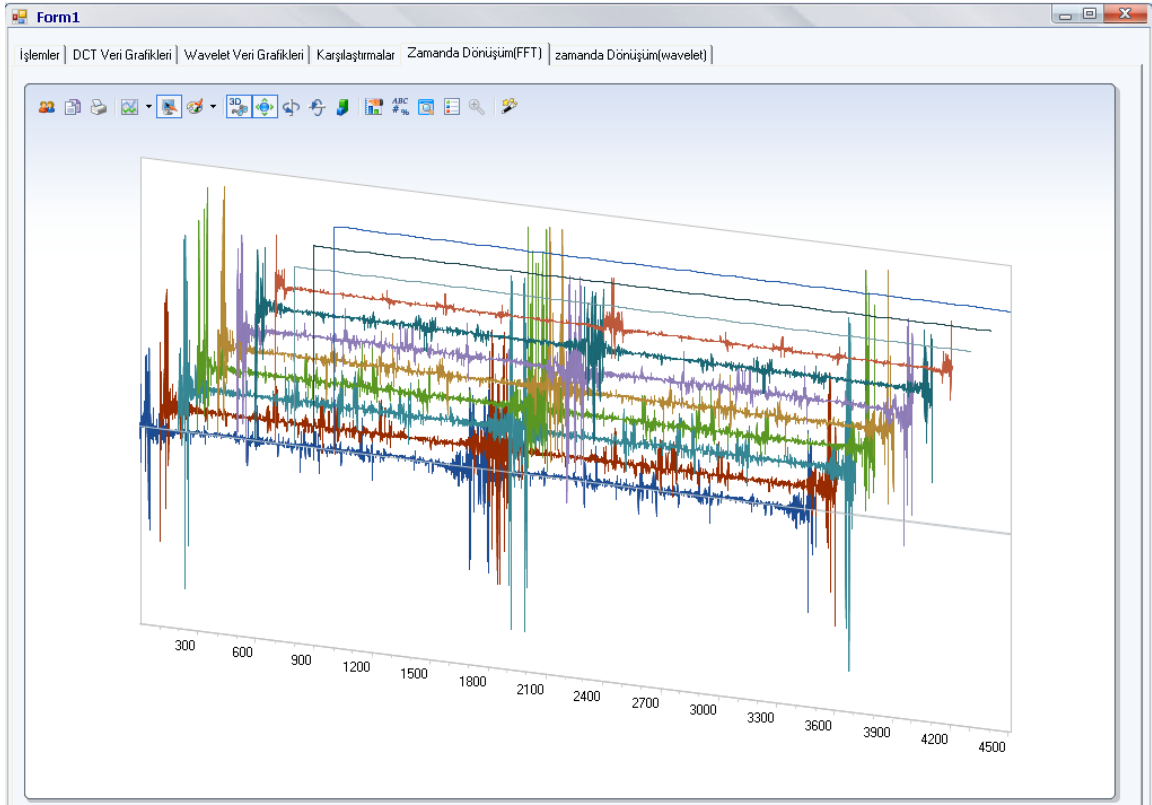


Şekil 31. Seçili seslere saç dalgacık dönüşümü uygulanmasının grafiksel gösterimi



Şekil 32. çevrelerin farklı zamanda dalgacık dönüşümü değerleri grafiksel gösterimi

Şekil 32.'de gözlemciye verilmek istenen bilgi şu şekilde belirtilebilir. Daha öncede söylenildiği gibi çalışmada belirlenen çerçeve sesbirim boyutundadır yani tamamen sesbirime bağlıdır. Belirlenmiş seste seçilmiş sesbirimin aranması işlemi belirlenmiş sestten sesbirim çerçeve boyutunda çerçeveler alınarak işleme sokulması şeklindedir. Şekil 32.'de sestten alınan bu çerçevelere uygulanan dalgacık dönüşümleri periyodik olarak alınmıştır. Bu periyot seçili sesin boyutundan sesbirimin boyutunun çıkartılması ve bu sayının 10'a bölünmesiyle kaç tane örnek alınacağı belirlendi ve bu sayıya göre mod alınarak grafikselleştirilecek örnekler belirlenerek grafiği gerçekleştirir. Şekil 33.'de aynı örnekler için FFT dönüşümü ile gerçekleştirilmiş halinin grafiksel olarak ekran görüntüsü verilmiştir.



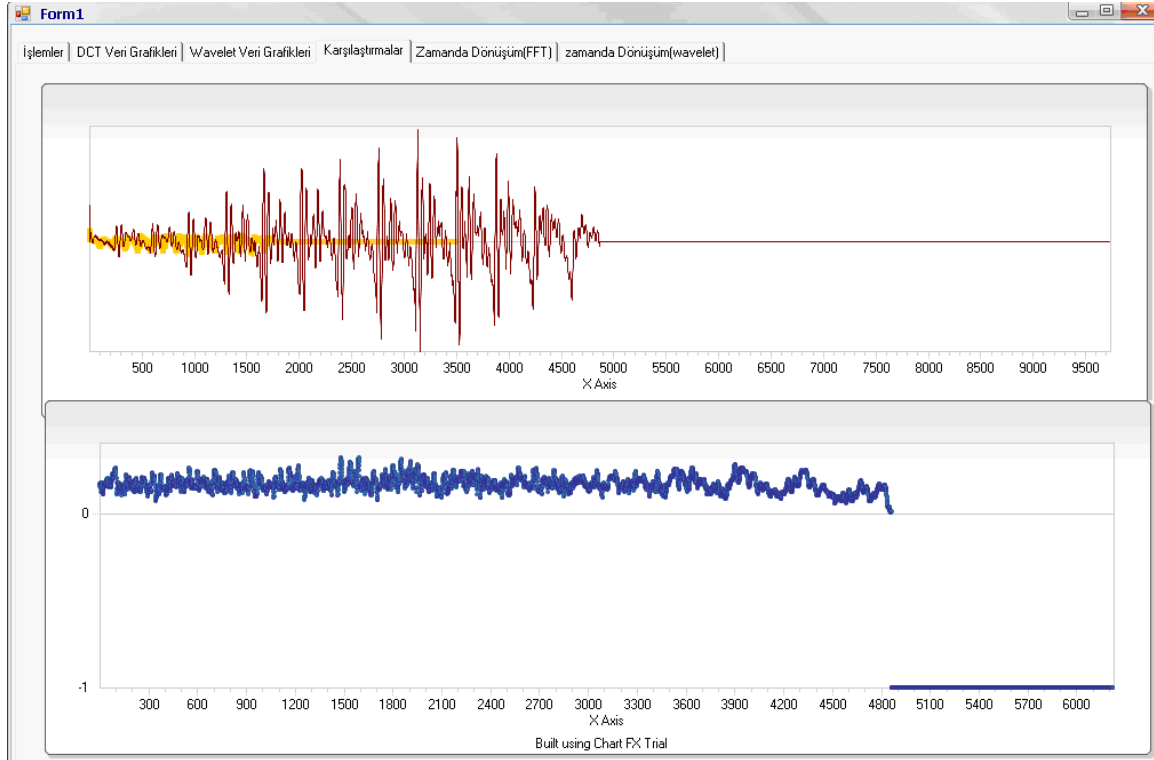
Şekil 33. Zamanda alınan farklı pencerelere uygulanmış FFT sonuç grafiği

Şekil 33. de verilen grafikteki fft sonuç grafikleri farklı pencerelerden alınmış olup çerçeveler arası farkın daha net olarak görülmesi amacıyla grafikselleştirilmiştir. Bu sonuçlar aşağıda ki şekilde grafikselleştirilmiştir.

- |                                    |                                       |
|------------------------------------|---------------------------------------|
| 1. Mavi fft sonucu 105.çerçeve     | 6. Mor fft sonucu 525. çerçeve        |
| 2. Bordo fft sonucu 210.çerçeve    | 7. Yeşil fft sonucu 630. çerçeve      |
| 3. Türkuaz fft sonucu 315. çerçeve | 8. Kahverengi fft sonucu 635.çerçeve  |
| 4. Yeşil fft sonucu 420.çerçeve    | 9. Açık mavi fft sonucu 740. çerçeve  |
| 5. Sarı fft sonucu 420. çerçeve    | 10. Koyu mavi fft sonucu 845. çerçeve |

Çalışmada ses verileri işlenmesinin aşamaları bitiminin ardından ses verilerinin karşılaştırma işlemi yapılmaktadır. Karşılaştırma işlemleri için iki farklı yöntem bulunmaktadır. Bunlardan biri çapraz korelasyon yöntemi diğeri ise dinamik zaman

çarpıtma yöntemidir. Burada temelde bu iki yöntem karşılaştırma yapmaktadır fakat felsefeleri farklıdır. Çapraz korelasyon verilen ses verilerinin arasındaki benzerlik oranını bulmaya çalışırken, dinamik zaman çarpıtma verilen ses verileri arasındaki farkı ortaya çıkartmaya çalışır. Fakat sonuç olarak bu yöntemler sayesinde iki farklı ses verisinin karşılaşma sonuçları gözlemciye sunulmaktadır. Bu sonuçlar, hem yazı olarak hem de grafiksel olarak sunulmaktadır. Aşağıda ses işleme için FFT ve karşılaştırma işlemi için çapraz korelasyon kullanılmış “a” sesli harfi ile ”e” sesli harflerin karşılaştırılması sonucu alınan sonucun grafiksel olarak gösterimi Şekil 34.’de verilmiştir. Bu grafiklerden ilki, ses verisinin zaman-genlik boyutundaki grafiği üzerine eşlemenin eşik değerini aşan pencere kısımlarına sesbirimin zaman-genlik değerleri grafiksel olarak ses grafiğinin üstüne basılmıştır.



Şekil 34. Karşılaştırma işleminde eşlemenin grafiksel ifade edilmesi

Şekil 34.’teki ikinci grafik ise noktasal olarak tüm çerçevelerdeki eşlenmelerin sonuçlarının verildiği grafiksel sunumdur. Gözlemci için eşlenmeler hem yazısal olarak hem de grafiksel olarak verilmiştir. Böylece, gözlemciye sağlıklı karar verebilmesi için mümkün olan tüm sonuçlar ve aşamalar sunulmuştur.

### 3. SONUÇLAR

Yapılmış olan çalışmada alınan sonuçların değerlendirilmesi tablolaştırılarak sunulmaya çalışılacaktır. Yapılmış olan çalışma için iki farklı kayıt alınmıştır. Bu kayıtlar sırasıyla kişilerden ve radyodan alınan kayıtlardır. Bu kayıtlardan kişilerden alınan kayıtlardan sesbirim veri tabanı oluşturulmuştur. Burada alınan kayıtlarda konuşmacıdan belirli kelimeleri söylemesi istenmiş ve kayıtlar kulağa sabitlenmiş bir mikrofon aracılığı ile konuşmacılardan kayıtlar alınmıştır. Kayıt alınırken kulağa sabitlenen mikrofon kullanılmasının en önemli nedeni her konuşmacıdan sabit mesafede ses kaydının alınabilmesi içindir.

Alınan kayıtlardan sesbirimler ses editörü vasıtası ile çıkartılarak bir sesbirim veri tabanı oluşturulmuştur. Bu veri tabanından iki şekilde yararlanılmıştır. İlk olarak sesbirimlerinin kendi aralarında ki benzerliklerin ve farklarının belirlenmesi, ikinci olarak belirli ses ve radyo kayıtlarının içinde belirlenmiş ses birimlerinin aranması şeklindedir. Ses birimlerin birbirlerine olan benzerliği aranırken iki farklı yol kullanılmıştır. İlk olarak aynı konuşmacıdan alınmış sesbirimler içerisindeki benzerlik aranmış ve tablolar farklı yöntemler kullanılarak oluşturulmuştur. Kullanılan yöntemler şu şekilde sıralanabilir. Sesin işlenmesi safhasında FFT (hızlı fourier dönüşümü) ve saç dalgacık dönüşümü kullanılmıştır. Aynı zamanda sesin eşlenmesi işlemi safhasında verilecek veriler hiçbir dönüşüm tarafından işlenmeden verilmiştir. Seslerin eşlenmesi safhasında da çapraz korelasyon ve dinamik zaman çarpıtma yöntemleri kullanılmıştır.

Anlaşılabacağı üzere 6 farklı şekilde sesbirimler üzerinde işlem yapılabilmektedir. Bu yöntemler sırası ile şöyledir:

Saç dalgacık dönüşümü --çapraz korelasyon

Saç dalgacık dönüşümü -- dinamik zaman çarpıtma

Hızlı fourier dönüşümü -- çapraz korelasyon

Hızlı fourier dönüşümü -- dinamik zaman çarpıtma

Zamanda ses verisi -- çapraz korelasyon

Zamanda ses verisi -- dinamik zaman çarpıtma

Çalışma kullanılan bu yöntemler ile oluşturulmuş örnekler aşağıda sınıflanarak verilmiştir. İlk sınıf olarak FFT (hızlı fourier dönüşüm) ve eşleme yöntemleri, ikinci sınıf olarak dalgacık dönüşümü ve eşleme yöntemleri ve son sınıf olarak zaman boyutundan alınan veriler ve eşleme yöntemlerinin oluşturduğu sistemlerden alınan sonuçlar tablolaraştırılarak verilmiştir. Bu sistemlerde olarak sessiz ve sesli sesbirimlerin arasındaki ilişki irdelenecektir. Bu inceleme yapılırken sessiz harfler kendi aralarında ki sınıflar taban alınarak tablolama yapılmıştır. Bu sınıflar ötümlü- sürekli olan [m, v, j, l, n, r, z, ğ, y], ötümlü-süreksiz [b,c,d,g], ötümsüz – sürekli [f,s,h], ötümsüz süreksiz [p,ç,t,k] irdelenmiştir.

### 3.1.1 Hızlı Fourier Dönüşümü ve Çapraz Korelasyon ile üretilen tablolar

Tablo 3. Ötümlü - sürekli olan sessiz sesbirimler birbiri ile olan ilişkisi

	<b>M</b>	<b>V</b>	<b>J</b>	<b>L</b>	<b>N</b>	<b>R</b>	<b>Z</b>	<b>Ğ</b>	<b>Y</b>
<b>M</b>	83	74	23	83	87	77	60	47	60
<b>V</b>	74	79	17	71	76	73	81	69	19
<b>J</b>	23	17	68	11	73	5	16	49	16
<b>L</b>	83	71	11	72	70	69	45	65	77
<b>N</b>	87	76	73	70	65	73	16	67	54
<b>R</b>	77	73	5	69	73	67	70	74	11
<b>Z</b>	57	81	16	45	16	70	56	81	17
<b>Ğ</b>	47	69	49	65	67	74	81	78	48
<b>Y</b>	60	19	16	77	54	11	17	48	72

Tablo 3.'de verile sonuçlar Hızlı fourier dönüşümü – çapraz korelasyon yöntemlerinin, farklı bayan konuşmacılardan alınan sesbirimlere uygulanması sonucunda oluşan benzerlik yüzdeleri alınmıştır. Alınan ses örnekleri farklı bayan konuşmacılara ait olup her birine söylenen farklı kelimelerden alınan sesbirimlerin birbirleri arasındaki ilişki irdelenmiş ve yapılacak olan sesbirim tanıma tabanlı konuşulan dilin belirlenmesi işleminde hangi sesbirimin hangi sesbirimle benzerliği var ve onun araştırması yapılarak hatalar için bir öngörü sağlanmış olunur.

Tablo 4. Ötümsüz - sürekli olan sessiz sesbirimler

	F	S	H
F	100	32	72
S	32	100	74
H	72	74	100

Tablo 4.'da görülen sonuç tablosu bir erkek konuşmacıdan alınan ses kayıtlarından oluşturulmuş ötümsüz-sürekli sessiz sesbirimleri arasında ki benzerlik yüzdeleri tabloda verilmiştir. Bu benzerlik hesaplanırken hızlı fourier dönüşümü yöntemi ile çapraz korelasyon yöntemi uygulanmıştır.

Tablo 5. Ötümsüz - süreksiz olan sessiz sesbirimler

	P	Ç	T	K
P	67	35	43	47
Ç	35	76	29	16
T	43	29	68	17
K	47	16	17	61

Tablo 5.'de görülen sonuç tablosu iki farklı erkek konuşmacıdan alınan ses kayıtlarından oluşturulmuş ötümsüz-süreksiz sessiz sesbirimleri arasında ki benzerlik yüzdeleri tabloda verilmiştir. Bu benzerlik hesaplanırken zaman boyutundan alınan verilere hızlı fourier dönüştürme yöntemi ile çapraz korelasyon uygulanması ile elde edilen sonuçların tablosudur.

Tablo 6. Ünlülerin birbiri içerisinde benzerliği

	A	E	İ	I	O	Ö	U	Ü
A	70	36	25	27	42	11	12	29
E	36	62	25	32	46	17	15	14
İ	25	25	74	33	46	13	14	16
I	27	32	33	80	24	27	24	17
O	42	46	46	24	76	11	43	47
Ö	11	17	13	27	24	71	14	17
U	12	15	14	24	43	14	76	42
Ü	29	14	16	17	47	37	42	44

Tablo 6.'de görülen sonuç tablosu iki farklı erkek konuşmacıdan alınan ses kayıtlarından oluşturulmuş sesli sesbirimleri arasında ki benzerlik yüzdeleri tabloda verilmiştir. Bu benzerlik hızlı fourier dönüşümü ile çapraz korelasyon ses verilerine uygulanması ile elde edilen sonuçların tablosudur. Buradaki tablodan çıkarılacak sonuçlarla sesli harfler arasındaki benzerlik çıkartılabilir ve işlemlerinde sonunda elde edilen eşleşme yüzdesinin değerlendirmesi buna göre yapılabilir.

### 3.1.2 Hızlı Fourier Dönüşümü ve Dinamik Zaman Çarpıtma ile üretilen tablolar

Tablo 7. Ötümlü - süreksiz olan sessiz sesbirimler

	B	C	D	G
B	78	3	67	21
C	3	76	11	7
D	67	11	71	21
G	21	7	21	69

Tablo 7.'da görülen sonuç tablosu farklı bayan konuşmacılardan alınan ses kayıtlarından oluşturulmuş sesbirimleri arasında ki ilişki hızlı fourier dönüşümü ile dinamik zamanlı çarpıtma yöntemini uygulanması sonucu üretilen benzerlik yüzdeleri sonuç tablosudur.

Tablo 8. Ötümsüz - sürekli olan sessiz sesbirimler

	F	S	H
F	100	41	10
S	41	100	32
H	10	32	100

Tablo 8.'da görülen sonuç tablosu bir bayan konuşmacıdan alınan ses kayıtlarından oluşturulmuş ötümsüz-sürekli sessiz sesbirimleri arasında ki benzerlik yüzdeleri tabloda verilmiştir. Bu benzerlik hesaplanırken hızlı fourier dönüşümü yöntemi ile dinamik zaman çarpıtma yöntemi uygulanmıştır.



Tablo 9. Ötümsüz - sürekli olan sessiz sesbirimler

	F	S	H
F	100	57	44
S	57	100	43
H	44	43	100

Tablo 9.'de görülen sonuç tablosu bir erkek konuşmacıdan alınan ses kayıtlarından oluşturulmuş ötümsüz-sürekli sessiz sesbirimleri arasında ki ilişki diğer bir deyişle benzerlik yüzdeleri saç dalgacık dönüşümü ile tabloda verilmiştir.

### 3.2.1 Saç Dalgacık Dönüşümü ve Çapraz Korelasyon ile üretilen tablolar

Tablo 10. Ötümlü - sürekli olan sessiz sesbirimler birbiri ile olan ilişkisi

	<b>M</b>	<b>V</b>	<b>J</b>	<b>L</b>	<b>N</b>	<b>R</b>	<b>Z</b>	<b>Ğ</b>	<b>Y</b>
<b>M</b>	100	45	18	72	79	62	67	70	46
<b>V</b>	45	100	15	54	66	76	63	70	65
<b>J</b>	18	15	100	15	20	41	14	35	13
<b>L</b>	72	54	15	100	76	64	65	71	41
<b>N</b>	79	66	20	76	100	81	83	83	53
<b>R</b>	62	76	41	64	81	100	85	74	88
<b>Z</b>	67	63	14	45	80	85	100	81	40
<b>Ğ</b>	70	70	35	71	83	74	81	100	51
<b>Y</b>	46	65	13	41	53	88	40	51	100

Tablo 10.'de verilen sonuçlar saç dalgacık dönüşümü ile çapraz korelasyon sonucu üretilmiş benzerlik yüzdeleri sonuçlarının tablolaştırılmış şeklidir. Burada kullanılan sesbirim örnekleri bir erkek kullanıcıdan alınan örnek seslerden oluşturulmuş sesbirimlerdir. Bu sesbirimlerin kendi içlerindeki benzerliği bulunarak yapılacak eşleşme işleminde bulunan sonuçların doğruluğu ve eğer şüphe duyulacak bir durum söz konusu ise hangi ses ile karıştırılmış olacağının tahmin edilmesinde kullanılması amaçlı olarak üretilmiştir.

Tablo 11. Ötümlü - süreksiz olan sessiz sesbirimler

	B	C	D
B	100	10	43
C	10	100	25
D	43	25	100

Tablo 11.'de verilen sonuç tablosu bayan bir konuşmacıdan alınan ses verilerine saç dalgacık dönüşümü ve çapraz korelasyon uygulanması sonucu üretilen benzerlik yüzdeleri sonuç tablosudur.

Tablo 12. Ötümlü - süreksiz olan sessiz sesbirimler

	B	C	D	G
B	100	8	94	78
C	8	100	40	14
D	94	40	100	79
G	78	14	79	100

Tablo 12.'de görülen sonuç tablosu bir erkek konuşmacıdan alınan ses kayıtlarından oluşturulmuş sesbirimler arasında ki ilişki hızlı fourier dönüşümü ile çapraz korelasyon yöntemini uygulanması sonucu üretilen sonuç tablosudur. Görüldüğü üzere burada "b" ile "d" arasındaki benzerlik önemsenecek kadar fazladır. Buradan anlaşıyor ki bu tablolar dikkate alınarak sesbirim aramaların sonuçları hakkında yorum yapılmalıdır.

Tablo 13. Ötümsüz - sürekli olan sessiz sesbirimler

	F	S	H
F	100	32	78
S	74	100	74
H	78	32	100

Tablo 13.'de görülen sonuç tablosu bir bayan konuşmacıdan alınan ses kayıtlarından oluşturulmuş ötümsüz-sürekli sessiz sesbirimleri arasında ki benzerlik yüzdeleri tabloda verilmiştir. Bu benzerlik hesaplanırken saç dalgacık dönüşüm yöntemi ile çapraz korelasyon yöntemi uygulanmıştır.

### 3.2.2 Saç Dalgacık Dönüşümü ve Dinamik Zaman Çarpıtma ile üretilen tablolar

Tablo 14. Ötümlü - süreksiz olan sessiz sesbirimler

	B	C	D	G
B	100	32	17	18
C	32	100	11	30
D	17	11	100	18
G	18	30	18	100

Tablo 14.'de görülen sonuç tablosu bir erkek konuşmacıdan alınan ses kayıtlarından oluşturulmuş sesbirimler arasında ki ilişki saç dalgacık dönüşümü - dinamik zaman çarpıtma yöntemleri uygulanması ile elde edinilen benzerlik yüzdeleri sonuç tablosudur.

Tablo 15. Ötümlü - süreksiz olan sessiz sesbirimler

	B	C	D	G
B	64	37	24	18
C	37	73	33	11
D	24	33	77	23
G	18	11	23	68

Tablo 15.'de görülen sonuç tablosu farklı iki bayan konuşmacıdan alınan ses kayıtlarından oluşturulmuş sesbirimler arasında ki ilişki sesbirimlerine saç dalgacık dönüşümü ile dinamik zaman çarpıtma uygulanarak elde edilen benzerlik yüzdelerinin sonuçların tablosu verilmiştir.

Tablo 16. Ötümsüz - sürekli olan sessiz sesbirimler

	F	S	H
F	100	35	17
S	35	100	30
H	17	30	100

Tablo 16.'de görülen sonuç tablosu bir erkek konuşmacıdan alınan ses kayıtlarından oluşturulmuş ötümsüz-sürekli sessiz sesbirimleri arasında ki ilişki diğer bir deyişle benzerlik yüzdeleri tabloda verilmiştir.

Tablo 17. Ötümsüz - sürekli olan sessiz sesbirimler

	F	S	H
F	100	35	17
S	35	100	30
H	17	30	100

Tablo 17.'da görülen sonuç tablosu bir bayan konuşmacıdan alınan ses kayıtlarından oluşturulmuş ötümsüz-sürekli sessiz sesbirimleri arasında ki benzerlik yüzdeleri tabloda verilmiştir.

### 3.3.1 Zaman Boyutu ve Çapraz Korelasyon ile üretilen tablolar

Tablo 18. Ötümlü - süreksiz olan sessiz sesbirimler

	B	C	D	G
B	100	14	78	34
C	14	100	11	7
D	78	11	100	70
G	34	7	70	100

Tablo 18.'da görülen sonuç tablosu bir erkek konuşmacılardan alınan ses kayıtlarından oluşturulmuş sesbirimler arasında ki ilişki zaman boyutundan alınan verilerle çapraz korelasyon yöntemi uygulanmış ötümlü-süreksiz sessizlerinin benzerlik yüzde oranlarını göstermektedir.

Tablo 19. Ötümsüz - sürekli olan sessiz sesbirimler

	F	S	H
F	100	59	46
S	59	100	26
H	46	26	100

Tablo 19.'da görülen sonuç tablosu bir bayan konuşmacıdan alınan ses kayıtlarından oluşturulmuş ötümsüz-sürekli sessiz sesbirimleri arasında ki benzerlik yüzdeleri tabloda verilmiştir. Bu benzerlik hesaplanırken zaman boyutundan alınan verilere çapraz korelasyon yöntemi uygulanması ile elde edilen sonuçların tablosudur.

Tablo 20. Ötümsüz - süreksiz olan sessiz sesbirimler

	P	Ç	T	K
P	50	42	30	22
Ç	42	53	16	34
T	30	14	33	18
K	22	34	18	32

Tablo 20.'de görülen sonuç tablosu iki farklı erkek konuşmacıdan alınan ses kayıtlarından oluşturulmuş ötümsüz-süreksiz sessiz sesbirimleri arasında ki benzerlik yüzdeleri tabloda verilmiştir. Bu benzerlik hesaplanırken zaman boyutundan alınan verilere zaman boyutundan ses verilerine çapraz korelasyon uygulanması ile elde edilen sonuçların tablosudur.

Tablo 21. Ünlülerin birbiri içerisinde benzerliği

	A	E	İ	I	O	Ö	U	Ü
A	29	15	7	19	22	23	27	11
E	15	34	13	26	36	17	27	18
İ	7	13	35	25	45	6	33	31
I	19	26	25	40	53	16	12	62
O	22	36	46	53	24	9	46	18
Ö	23	14	7	16	9	8	15	9
U	27	27	33	13	46	15	26	53
Ü	11	8	31	60	18	9	53	47

Tablo 21.'de görülen sonuç tablosu iki farklı erkek konuşmacıdan alınan ses kayıtlarından oluşturulmuş sesli sesbirimleri arasında ki benzerlik yüzdeleri tabloda

verilmiştir. Bu benzerlik zaman boyutundan alınan veriler ile çapraz korelasyon yöntemi ses verilerine uygulanması ile elde edilen sonuçların tablosudur. Görüldüğü üzere eşleşme oranları çok başarısızdır. Bunun sebebi ses verilerinin zamanda başlama noktalarının farklı olması ve ya gürültüden ötürüdür. Buradaki bu denli başarısız sonuç alınması dönüşüm yapılmasının ne denli doğru olduğunun göstergesidir.

### 3.3.2 Zaman Boyutu ve Dinamik Zaman Çarpıtma ile üretilen tablolar

Tablo 22. Ötümlü - sürekli olan sessiz sesbirimler birbiri ile olan ilişkisi

	<b>M</b>	<b>V</b>	<b>J</b>	<b>L</b>	<b>N</b>	<b>R</b>	<b>Z</b>	<b>Ğ</b>	<b>Y</b>
<b>M</b>	100	46	17	43	39	11	10	29	37
<b>V</b>	46	100	16	54	46	77	3	18	61
<b>J</b>	17	16	100	15	1	33	35	20	45
<b>L</b>	43	54	15	100	43	32	31	26	65
<b>N</b>	39	46	19	56	100	26	24	26	24
<b>R</b>	11	56	41	55	26	100	35	29	33
<b>Z</b>	10	43	13	57	24	35	100	28	43
<b>Ğ</b>	29	51	56	62	29	29	28	100	31
<b>Y</b>	37	45	12	26	22	53	43	31	100

Tablo 22.'de Zaman boyutu – dinamik zaman çarpıtma yönteminin bir erkek konuşmacıdan alınan sesbirimlere uygulanmasının benzerlik yüzdelerini gösteren sonuç tablosu verilmiştir. Burada zaman boyutundan kastedilen alınan ses örnekleri herhangi bir dönüşüme tabi tutulmadan ses eşleme safhasına verilen verilerdir. Zaman genlik boyutundan kasıt veriler mikrofondan alınarak “\*.wav” dosyasına çevrilmiş haldeki ham halleri kastedilmek isteniyor.

Burada zaman boyutundan alınan veriler üzerinde yapılan eşleşmeler gözle fark edilebilecek derecede düşük yani başarısız eşleşme olduğu görülebilmektedir. Buna en önemli faktör olarak farklı zamanlarda konuşma kısımlarının gerçekleşmesi gösterilebilir. Eğer ki konuşmalar eş zamanlı başlanabilse o zaman sonucun daha iyi olacağını aynı sesbirimlerde olan eşleşme yüzdesindeki başarıdan kaynaklı öngörülebilir. Burada başarılı sonuçlar alınmasının en önemli sebeplerinden biride kelime belirleme yapısından kaynaklıdır. Kelime belirleme ile zamandaki kaymalar belli oranda kapatabilmektedir. Bu sayede başarıda bir artım sağlanmıştır.

Tablo 23. Ötümlü - süreksiz olan sessiz sesbirimler

	B	C	D	G
B	100	29	26	27
C	29	100	13	12
D	26	13	100	15
G	27	12	15	100

Tablo 23.'de görülen sonuç tablosu bir bayan konuşmacıdan alınan ses kayıtlarından oluşturulmuş sesbirimler arasında ki ilişki zaman boyutundan dönüşüme uğramamış ses verilerine dinamik zaman çarpıtma uygulanması sonucu elde edilmiş benzerlik yüzdeleri gözükmetedir.

Tablo 24. Ötümsüz - sürekli olan sessiz sesbirimler

	F	S	H
F	100	41	17
S	41	100	36
H	17	36	100

Tablo 24.'de görülen sonuç tablosu bir bayan konuşmacıdan alınan ses kayıtlarından oluşturulmuş ötümsüz-sürekli sessiz sesbirimleri arasında ki benzerlik yüzdeleri tabloda verilmiştir. Bu benzerlik hesaplanırken zaman boyutundan alınan verilere dinamik zaman çarpıtma yöntemi uygulanması ile elde edilen sonuçların tablosudur.

Tablo 25. Ötümsüz - süreksiz olan sessiz sesbirimler

	P	Ç	T	K
P	100	53	37	34
Ç	53	100	32	34
T	37	32	100	74
K	34	34	74	100

Tablo 25.'de görülen sonuç tablosu iki farklı erkek konuşmacıdan alınan ses kayıtlarından oluşturulmuş ötümsüz-süreksiz sessiz sesbirimleri arasında ki benzerlik yüzdeleri tabloda verilmiştir. Bu benzerlik hesaplanırken zaman boyutundan alınan

verilere zaman boyutundan ses verilerine dinamik zaman çarpıtma uygulanması ile elde edilen sonuçların tablosudur.

### 3.4 Sesbirimler ile Türkçe ve İngilizce Dili ile İlgili Tablolar

Tablo 26. Sesbirimlerin seçili konuşma içinde tanınmasının doğruluk yüzdesi

Tanınan sesbirim	Yöntem	Cinsiyet	Adet	Bulunma sayısı	Bulunan sayısı	Doğruluk oranı (%)
F	Hfd-çk	E	10	2	1	%57
F	Hfd-çk	B	10	2	1	%51
H	Hfd-çk	E	10	2	1	%56
H	Hfd-çk	B	10	2	0	%48
J	Hfd-çk	E	10	1	1	%59
J	Hfd-çk	B	10	1	1	%53
P	Hfd-çk	E	10	2	1	%59
P	Hfd-çk	B	10	2	1	%62
Ş	Hfd-çk	E	10	2	2	%74
Ş	Hfd-çk	B	10	2	2	%66
Ü	Hfd-çk	E	10	2	2	%77
Ü	Hfd-çk	B	10	2	2	%72

Tablo 27. Sesbirimlerin seçili konuşma içinde tanınmasının doğruluk yüzdesi

Tanınan sesbirim	Yöntem	Cinsiyet	Adet	Bulunma sayısı	Bulunan sayısı	Doğruluk oranı (%)
B	Hfd-dzç	E	10	2	1	%67
B	Hfd-dzç	B	10	2	1	%53
D	Hfd-dzç	B	10	2	1	%54
D	Hfd-dzç	E	10	2	1	%58
L	Hfd-dzç	B	10	2	2	%54
L	Hfd-dzç	E	10	2	2	%61
N	Hfd-dzç	E	10	2	2	%74
N	Hfd-dzç	B	10	2	2	%76
R	Hdf-dzç	E	10	3	2	%63
R	Hdf-dzç	B	10	3	2	%66
V	Hfd-dzç	B	10	2	2	%57
V	Hfd-dzç	E	10	2	2	%58

**Sdd:** Saç dalgacık dönüşümü

**Hfd:** Hızlı fourier dönüşümü

**ÇK:** Çapraz korelasyon

**Dzç:** Dinamik zaman çarpıtma



Tablo 28. Sesbirimlerin seçili konuşma içinde tanınmasının doğruluk yüzdesi

Tanınan sesbirim	Yöntem	Cinsiyet	Adet	Bulunma sayısı	Bulunan sayısı	Doğruluk oranı (%)
A	sdd-çk	E	10	3	3	%77
A	sdd-çk	B	10	3	3	%73
E	Sdd-çk	E	10	3	3	%79
E	Sdd-çk	B	10	3	3	%78
M	Sdd-çk	E	10	2	2	%73
M	Sdd-çk	B	10	2	2	%74
Ö	Sdd-çk	E	10	3	3	%77
Ö	Sdd-çk	B	10	3	3	%71
S	Sdd-çk	E	10	2	2	%74
S	Sdd-çk	B	10	2	2	%72
U	Sdd-çk	E	10	3	3	%81
U	Sdd-çk	B	10	3	3	%79

Tablo 29. Sesbirimlerin seçili konuşma içinde tanınmasının doğruluk yüzdesi

Tanınan sesbirim	Yöntem	Cinsiyet	Adet	Bulunma sayısı	Bulunan sayısı	Doğruluk oranı (%)
Ç	Sdd-dzç	E	10	2	2	%74
Ç	Sdd-dzç	B	10	2	2	%69
I	Sdd-dzç	E	10	3	3	%79
I	Sdd-dzç	B	10	3	3	%81
K	Sdd-dzç	E	10	2	1	%57
K	Sdd-dzç	B	10	2	1	%53
O	Sdd-dzç	E	10	3	3	%78
O	Sdd-dzç	B	10	3	3	%76
Q	Sdd-dzç	E	10	1	1	%55
Q	Sdd-dzç	B	10	1	0	%41

**Sdd:** Saç dalgacık dönüşümü

**Hfd:** Hızlı fourier dönüşümü

**ÇK:** Çapraz korelasyon

**Dzç:** Dinamik zaman çarpıtma

Yukarıdaki tablo 26, 27, 28 ve 29 da farklı yöntemler kullanılarak erkek ve bayan'dan alınan Türkçe sesbirimlerin Türkçe konuşma içerisinde aranmasının sonuçları karşılaştırılmalı olarak verilmiştir. Bu araştırmanın sonuçlarını olarak kullanarak kurulan sistemlerin performanslarını ve tanıma işlemlerinin başarı oranları değerlendirilmesi yapılabilir.

Tablo 30. Radyo kayıtlarında aranan sesbirimlerin eşlenme oranları

Taninan harf	Yöntem	Doğruluk oranı (%)	Bulunma adedi	Tanınma adedi
A	Hfd-çk	%47	3	1
E	Hfd-dzç	%43	3	0
İ	Sdd-çk	%37	2	0
Ö	Sdd-dzç	%52	2	2
Ç	Hfd-dzç	%44	1	0
Th (i)	Hfd-çk	%39	1	0
Th (e)	Hfd-çk	%43	1	0
-tion	Sdd-çk	%61	1	1
Eu	Sdd-çk	%56	1	1

**Sdd:** Saç dalgacık dönüşümü

**Hfd:** Hızlı fourier dönüşümü

**ÇK:** Çapraz korelasyon

**Dzç:** Dinamik zaman çarpıtma

Tablo 30.'de İngilizce sesbirimler (th (i),th (e),-tion,eu) ve Türkçe sesbirimler (a,e,i,ö,ç) seçilerek radyodan yapılan kayıtlar içerisinde aranarak eşlenme yüzdeleri tablo27'de verilmiştir. Bu örnekle anlaşılacağı üzere hazırlanan sistem sadece Türkçe için değil sesbirim bilgisine sahip olunan her hangi bir dil içinde kullanılabilir. Seçilen sesbirimler İngilizcede en çok kullanılan ve İngiliz dilini bilen insanların çoğunluğu bu sesbirimleri doğru telaffuz yani şive etkisi minimumdur ve açık telaffuz edebildiği için tercih edilmiştir.

Tablo 31. Türkçe fonemin Türkçe metin ve İngilizce metin içerisinde araştırılması

Taninan harf	Yöntem	Kayıt adet	Türkçe metin Doğruluk oranı	Bulunma adedi/Tanınma adedi	İngilizce metin Doğruluk oranı	Bulunma adedi/Tanınma adedi
A	Hfd-çk	10	%77	3/3	%18	3/0
E	Hfd-çk	10	%79	3/3	%8	3/0
İ	Hfd-çk	10	%79	3/3	%15	3/0
Ö	Hfd-çk	10	%77	3/3	%3	3/0

**Sdd:** Saç dalgacık dönüşümü

**Hfd:** Hızlı fourier dönüşümü

**ÇK:** Çapraz korelasyon

**Dzç:** Dinamik zaman çarpıtma

Tablo 31. Türkçe sesbirimlerin Türkçe konuşma ve İngilizce konuşma içerisinde aranmasının sonuç tablosudur. Bu tablodan anlaşılacağı üzere sistem denenmek istenmiştir. Bu deneme işlemi Türkçe sesbirimlerin İngilizce konuşmanın içerisinde varlığının bulunmasına yöneliktir. Bu deneme ile elde edilen sonuç şu şekilde açıklanabilir; dil belirlenmesi o dilin sesbirimlerinin konuşmada aranması ile elde edilebilirliği yönünde görüş elde edebiliriz. Farklı bir dilin sesbirimleri farklı dillerde çok küçük eşleşme değerlerine sahip oluyor ki bu tarafımızca belirlenen %45'lik eşik değerini aşamamaktadır. Bu aşamama sonucunda bir eşelemeye sonuç üretilememektedir. Bu üretimsizlikten dilin sesbirimleri kendisine özgüdür olduğudur. Bu özellik sayesinde konuşulan dil belirlenebilir.

Tablo 32. İngilizce fonemin İngilizce metin ve Türkçe metin içerisinde araştırması

Tanınan harf	Yöntem	Kayıt adedi	Türkçe metin Doğruluk oranı	Bulunma adedi/Tanınma adedi	İngilizce metin Doğruluk oranı	Bulunma adedi/Tanınma adedi
Th(i)	Hfd-çk	5	%39	1/0	%8	1/0
Th(e)	Hfd-çk	5	%43	1/0	%13	1/0
-tion	Hfd-çk	5	%61	1/1	%14	1/0
Eu	Hfd-çk	5	%56	1/1	%16	1/0

**Sdd:** Saç dalgacık dönüşümü

**Hfd:** Hızlı fourier dönüşümü

**ÇK:** Çapraz korelasyon

**Dzç:** Dinamik zaman çarpıtma

Tablo 32. İngilizce sesbirimlerin Türkçe konuşmanın içinde aranmasının sonuç tablosudur. Bu tablodan anlaşılacağı üzere sistem İngilizce sesbirimlerin Türkçe konuşmanın içerisinde varlığının bulunmasına yönelik bir deneme yapılmıştır. Bu deneme ile elde edilen sonuç şu şekilde açıklanabilir; dil belirlenmesi o dilin sesbirimlerinin konuşmada aranması ile elde edilebilirliği yönünde görüş elde edebiliriz. Farklı bir dilin sesbirimleri farklı dillerde çok küçük eşleşme değerlerine sahip oluyor ki bu tarafımızca belirlenen %45'lik eşik değerini aşamamaktadır. Bu aşamama sonucunda bir eşelemeye sonuç üretilememektedir. Bu üretimsizlikten anlayacağımız dilin sesbirimleri kendisine özgüdür.

Tablo 33. Türkçe sesbirimlerin İngilizce konuşmanın içinde aranması

Taninan harf	Yöntem	Doğruluk oranı (%)
D (i)	Hfd-çk	%11
D (ı)	Hfd-çk	%17
Ş	Sdd-çk	%22
O	Sdd-çk	%26

**Sdd:** Saç dalgacık dönüşümü

**Hfd:** Hızlı fourier dönüşümü

**ÇK:** Çapraz korelasyon

**Dzç:** Dinamik zaman çarpıtma

Tablo 33. Türkçe sesbirimlerinden İngiliz dili sesbirimleri benzerleri olarak düşünülen sesbirimler seçilerek konuşmanın içinde aranması sonuçları verilmiştir. Ses olarak birbirine benzeyen bu sesbirimlerin eşlenmesi sonucundan şu çıkartılabilir ki diller benzer sesbirimlerine sahip olabilir fakat bu sesbirimlerin dilin özelliğine bağlı olarak telaffuzu sonucunda farklılaşırlar ve dillerde kendi özelliklerini barındıran sesbirimleri kullanılarak çalışmadaki program yardımıyla belirleme yapılabilir.

#### 4. ÖNERİLER

Yapılan çalışma kullanılan dilin ne olduğunun sesbirime dayalı tespitini kullanılacak olduğundan buradaki tanınacak dillerin sesbirimleri çok net olarak uzmanca araştırılarak en uygun sesbirimler bulunmalıdır. Bu çalışmada konuşmacı bağımsız bir tanıma sistemi oluşturulduğundan ses kayıtlarının netliği ve telaffuz farkının minimum olmasına dikkat edilmesi gereklidir. Sistemsiz olarak yapılan çalışmaya bir akıllı sistem eklentisi olarak eklenmesi, sistemin çok farklı yöntemlerle denenebilmesine olanak sağlayacaktır. Yapılanlara ek olarak ya bir yapay sinir ağı veya saklı markov modeli eklenmesi önerilebilir.

Sistemin mevcut pencereleme sistemi sesin işlenmesi ve eşlenmesi safhalarındaki veri yükü bakımından çok zorlamakta ve yavaşlatmaktadır. Pencere boyutları küçültülerek pencereler içerisine akıllı sistem bir sistem monte edilerek sistemin büyük boyutlu pencerelerden ötürü yavaşlatılması engellenebilir.

Sistemde pencere gezdirmeye işlemi bir birimlik kaydırma ilkesi doğrultusundadır. Bu yapı sistemi oldukça yavaşlatmasına rağmen sonuç üretimine olumsuz bir etki yapmamaktadır. Göze alınabilecek hata payları doğrultusunda farklı pencere kaydırma yöntemleri uygulanarak sistemin mevcut hızı arttırılabilir.

Bu öneriler doğrultusunda sistem daha verimli ve hızlı hale getirilebilir.

## 5. KAYNAKLAR

1. Beng T. TAN, Minyue Fu, Andrew Spray and Phillip Dermody ,The use of transforms in phoneme recognition, 1996
2. Busch, C., Wavelet Based Texture Segmentation of Multi-Modal Tomographic Images, Comput. & Graphics, 21/3/ (1997) 347–358
3. Chin-Teng Lin, Senior Member, IEEE, Jiann-Yow Lin, and Gin-Der Wu, A Robust Word Boundary Detection Algorithm for Variable Noise-Level Environment, IEEE, 2002
4. Eamonn J. Keogh and Michael J. Pazzani, Scaling up Dynamic Time Warping to Massive Datasets, 3rd European Conference on Principles and Practice of Knowledge Discovery in Databases, 1999
5. E. A. Yfantis and T. Lazarakis A. , On Time Aligment and Metric Algorithms for Speech Recognition, Angelopoulos Image Processing Laboratory Computer Science Department University of Nevada, 1999
6. Guang G., Correlation Among Signal Set Correlation Among Signal Sets, Department of Electrical & Computer Engineering University of Waterloo Canada, 2004
7. Graps, A., An Introduction to Wavelets, IEEE Computational Science & Engineering, 2, 2(1995) 50-61
8. Gonzalez R.C. and Woods R.E., Digital Image Processing, Prentice Hall, Upper Saddle River, New Jersey, 2002
9. <http://users.rowan.edu/~polikar/WAVELETS/WTtutorial.html>, Wavelet Tutorial, 12/06/2007
10. [http://en.wikipedia.org/wiki/Haar\\_wavelet](http://en.wikipedia.org/wiki/Haar_wavelet) , Haar Wavelet, 15/04/2007
11. <http://en.wikipedia.org/wiki/Wavelet>, Wavelet, 14/07/2007
12. J. G. Wilpon, L. R. Rabiner, and T. Martin An improved word detection algorithm for telephone quality speech, The Journal of the Acoustical Society of America , 1983
13. HOLMES J., Speech Synthesis and Recognition, Taylor and Francis, 2001
14. Lu, C.S., Chung, P.C., ve Chen, C.F., Unsupervised Texture Segmentation via Wavelet Transform, Pattern Recognition, 30/5 (1997) 729-742

15. Mengüšođlu, E., “Bir Tükçe Sesli İfade Tanıma Sisteminin Kural Tabanlı Tasarımı ve Gerçekleřtirimini”, Master Tezi, Hacettepe Üniversitesi, 1999
16. R. Favero and R. King. Wavelet parameterization for speech recognition. Preprint, 1993.
17. Robinson, T., ‘Speech Anaylsis Tutorial’, <http://svrwww.eng.cam.ac.uk/~ajr/speechAnalysis/>, 1998
18. ŐEN K.Ö., RAUBER A., KÖSE C, ARAS S., Konuřulan Dilin Tanınması, UMES, 2007
19. TEBELSKIS J., Speech Recognition Using Neural Networks, PhD. Thesis, Carnegie Mellon University, Pittsburgh Pennsylvania, 1995
20. Uytterhoeven, G., Wavelets:Software and Applications, Doktora Tezi, Haverlee, 1999.
21. Yařarođlu Y., Multi-Model Video Summarization Using Hidden Markov Models for Content-Based Multimedia Indexing, MSc. Thesis, Middle East Technical University, 2003
22. Yazıcı R. ve Köse C., Kısıtlamasız Türkçe Söz Sentezi, 1. Sinyal iřleme ve uygulamaları kurultayı, Bođaziçi Üniversitesi, 1993
23. Ying G.S., C.D. Mitchell and L.H. Jamieson, Endpoint detection of isolated utterance based on a modified teager energy measurement, CiteSeer, 1993

## ÖZGEÇMİŞ

Kamil Öncü ŞEN, 1978 Ankara doğumludur. İlköğrenimini 1984–1989 yılları arasında Gazi Osman Paşa İlkokulu’nda, orta ve lise öğrenimini 1989–1996 yılları arasında Özel Samanyolu Erkek Koleji’nde tamamladı. 1999 yılında Karadeniz Teknik Üniversitesi, Fen Edebiyat Fakültesi, İstatistik ve Bilgisayar Bilimleri Bölümü’nde Lisans Programı’na başladı ve 2003 yılında bu bölümden 3.’lük derecesi ile mezun oldu. 2003 yılında K.T.Ü. Fen Bilimleri Enstitüsü Bilgisayar Mühendisliği Anabilim Dalı’nda yüksek lisansa başladı. 2005-2006 yılları arasında Bilgisayar mühendisliği bölümünde asistan olarak görev yaptı. 2006–2007 eğitim öğretim yılını Viyana Teknik Üniversitesinde tamamladı.