

T.C
SELÇUK ÜNİVERSİTESİ
FEN BİLİMLERİ ENSTİTÜSÜ

KONUŞMACI TANIMA SİSTEMLERİNDE
DALGACIK DÖNÜŞÜMÜ
Onur AYGÜN
YÜKSEK LİSANS TEZİ
BİLGİSAYAR MÜHENDİSLİĞİ ANABİLİM DALI

Konya, 2006

ÖZET

Yüksek Lisans Tezi

KONUŞMACI TANIMA SİSTEMLERİNDE DALGACIK DÖNÜŞÜMÜ

Onur AYGÜN

Selçuk Üniversitesi Fen Bilimleri Enstitüsü

Bilgisayar Mühendisliği Anabilim Dalı

Danışman: Prof. Dr. Ferruh YILDIZ

2006, 79 Sayfa

Konuşma ve konuşmacı tanıma günümüz bilgi teknolojilerinde özel ve ayrı bir yere sahiptir.. Konuşma tanıma ile metinlerin elektronik ortama aktarılması diğer yöntemlerden çok daha hızlı yapılmaktadır. Ayrıca konuşma tanıma ile cihazların sesle yönetilmesi de sağlanabilmektedir. Konuşmacı tanıma ise biyometrik teknolojinin bir parçası olarak karşımıza çıkmakta ve insanların seslerinden tanınmasını sağlamaktadır. Bu iki tür tanımada frekans analizi hayati bir öneme sahiptir ve hemen hemen tüm adımlarda spektral düzlemde işlemler yapılır. Dalgacık dönüşümü ise son zamanların yükselen değeri olarak ortaya çıkmakta ve spektral analizlerde fourier analizinin yerine geçmeye aday durumdadır.

Bu çalışmada hem konuşma hem de konuşmacı tanımada geçilmesi gereken aşamalarından bahsedilmiştir. Her bir aşamada kullanılan genel yöntemler ve analizler hakkında bilgi verilmiştir. Frekans analizini ve dalgacık dönüşümü genel olarak ele alınmıştır. En son olarak ta dalgacık dönüşümünün konuşma ve konuşmacı tanıma işlemlerinde kullanım şekillerine yer verilmiştir.

Anahtar Kelimeler: Konuşmacı Doğrulama, Frekans Analizi, Öznitelik Çıkarımı, Dalgacık Dönüşümü

ABSTRACT

Master Thesis

WAVELET TRANSFORM IN SPEAKER RECOGNITION SYSTEMS

Onur AYGÜN

Selçuk Universty

Graduate School of Natural and Applied Sciences

Department of Computer Engineering

Supervisor: Prof. Dr. Ferruh YILDIZ

2006, 79 Page

Speech and speaker recognition is a special and important issue in nowadays information technology. With speech recognition, transferring text into digital platform can be made very fast. In addition to speech recognition provide to control devices by the voice commands. Speaker recognition is a branch of the biometric technology and provides to recognize people from their voice.

In each of these recognition, frequency analysis has a a important role and all process steps includes procedures over spectral domain. Wavelet transformation is rising over the information technology. This transformation technique candidate for supersedence of fourier transform on the spectral analysis.

This thesis is concerned with outline of speech and speaker recognition technology. General methods and analyses is described at every stage of recognition process. Frequency analysis and wavelet transform discussed briefly. At the end, utilization of wavelet transform on the speech and speaker recognition.is mentioned.

Key Words : Speaker Verification, Frequency Analysis, Feature Extraction, Wavelet Transform

İÇİNDEKİLER

ÖZET	i
ABSTRACT	ii
İÇİNDEKİLER	iii
ŞEKİL LİSTESİ	v
1. GİRİŞ	1
2. KONUŞMA VE KONUŞMACI TANIMA.....	2
2.1. BİYOMETRİK TEKNOLOJİ	2
2.2. SES VE KONUŞMA.....	3
2.2.1. Ses Nedir?.....	3
2.2.2. Konuşma Nedir?	4
3. KONUŞMACI TANIMA.....	7
3.1. KONUŞMACI TANIMA TEKNİKLERİ	7
3.1.1. Konuşmacı Tanıma Teknolojilerinin Sınıflandırılması.....	7
3.1.2. Konuşmacı Tanıma Sistemlerinin Genel Yapıları.....	9
3.1.3. Konuşma Sinyallerinin Temel Özellikleri	11
3.1.4. Konuşma Sinyallerinin Sayısallaştırılması	11
4. SİNYAL MODELLEMEDE TEMEL YAKLAŞIMLAR.....	15
4.1. SPEKTRAL DÖNÜŞÜM	15
4.2. SPEKTRAL ANALİZ	17
4.2.1. Temel Frekans (Perde) (Pitch)	17
4.2.2. Güç	17
4.2.3. Spektral Analiz.....	20
4.2.3.1. Sayısal Filtre (Süzgeç) Bankası.....	21
4.2.3.2. Fourier Dönüşüm Filtre Bankaları	24
4.2.3.3. Ters-Spektral (Kepstral) Katsayılar	24
4.2.3.4. Doğrusal Ön Kestirim Katsayıları.....	27
(Linear Prediction Coefficients).....	27
4.2.3.5. PLP (Perceptual Linear Prediction).....	29
4.2.3.6. RASTA (RelAtive SpekTrA).....	31
4.3. PARAMETRE DÖNÜŞÜMÜ.....	32
4.3.1. Fark Alma (Differentiation).....	32
4.3.2. Birleştirme (concatenation).....	32
4.4. NORMALİZASYON/ UYARLAMA TEKNİKLERİ	36
4.4.1. Parametre Domeni Normalizasyonu.....	36
4.4.2. Uzaklık/Benzerlik Domenindeki Normalizasyon.....	37
4.4.3. Model Bölgesi Adaptasyonu	38
4.5. KONUŞMACI MODELLEME.....	38
4.5.1. Dinamik Time Warping Temelli Konuşmacı Modelleme	39
4.5.2. Gauss Karışım Modelleri (Gaussian Mixture Model)	41
4.5.3. Saklı Markov Model (HMM) Tabanlı Konuşmacı Modelleri.....	42
4.5.4. Uzun dönemli istatistik tabanlı yöntemler.	44
4.5.5. Vektör Nicemleme (Vector Quantization).....	44
5. DALGACIK DÖNÜŞÜMÜ (WAVELET TRANSFORM).....	45
5.1. FREKANS ANALİZİ.....	45
5.1.1. Frekans Bölgesi	45
5.1.2. Spektrum Örnekleri.....	47
5.2. FOURIER DÖNÜŞÜMÜ	48
5.3. FOURIER DÖNÜŞÜM AİLESİ	51
5.4. AYRIK FOURIER DÖNÜŞÜMÜ (DFT).....	52
5.4.1. Kutupsal Gösterim.....	53

5.4.2. Pencerelemiş Fourier Dönüşümleri.....	55
(Windowed Fourier Transforms).....	55
5.4.3. Hızlı Fourier Dönüşümleri (Fast Fourier Transforms).....	55
5.5. DALGACIK (WAVELET) DÖNÜŞÜMÜ.....	56
5.6. DALGACIK ANALİZİ.....	57
5.6.1. Sürekli Dalgacık Dönüşümü.....	59
5.7. ANA DALGACIKLARIN ÖZELLİKLERİ.....	63
5.7.1. Örnek Ana Dalgacık Şekilleri ve formülleri.....	64
6.KONUŞMA/KONUŞMACI TANIMADA WAVELET.....	66
7. DENEYSEL ÇALIŞMA.....	69
SONUÇ.....	76
KAYNAKLAR.....	78

ŞEKİL LİSTESİ

Şekil 2. 1. Biyometrik Sistemlerin Genel Çalışma Prensibi	3
Şekil 2. 2. Ses Yolu (Ses Üretim Organları)	4
Şekil 2. 3. Konuşma Süreci	5
Şekil 3. 1. Konuşmacı Tanıma Alt Dalları	9
Şekil 3. 3. Konuşmacı Doğrulama Sistemlerinin Genel Çalışma Prensibi	10
Şekil 3. 4. Konuşma Sinyalinin Sayısallaştırma Adımları	11
Şekil 3. 5. Analog Sinyalin Sayısallaştırma Adımları	12
Şekil 3. 6. Analog/Sayısal Çevrim	12
Şekil 3. 7. Konuşma Sinyaline Uygulanan Ön İşlemler	13
Şekil 4. 1. Preemphasis Filtresinin Frekans Cevabı	16
Şekil 4. 3. Çerçeve Bazlı Analiz	20
Şekil 4. 4. Spektral Analiz Yöntemleri	21
Şekil 4. 5. Bark Ölçeği Dönüşümü	22
Şekil 4. 6. Mel - Frekans Dönüşümü	22
Şekil 4. 7. Mel Ölçek Filtre Bankası	23
Şekil 4. 8. Sesli ve Sessiz Konuşma Sinyalleri ve Cepstrumları	26
Şekil 4. 9. Perceptual Linear Prediction Yöntemi	30
Şekil 4. 10. DTW Temelli bir sistemin genel yapısı	40
Şekil 4. 11. Markov Modeli ve geçiş örneği	43
Şekil 5. 1. İki Frekansın Birleşimi	45
Şekil 5. 2. Zaman-Frekans-Genlik Düzlemi	45
Şekil 5. 3. Sürekli Bir Sinyal ve Frekans Bileşenleri	46
Şekil 5. 4. Bazı Sinyallerin Sinüsoidal Bileşenleri	48
Şekil 5. 5. Fourier Dönüşümünde Dörtgen ve Polar Gösterim	55
Şekil 5. 6. Sinüs İşareti ve Ana Dalgacık	57
Şekil 5. 7. Dalgacık Analizinin Aşamaları	58
Şekil 5. 9. Dalgacık Dönüşümünden Hesaplanan Katsayıların Zamanla Değişen İki Boyutlu Gösterimi	61
Şekil 5. 10. Dalgacık Dönüşümünden Hesaplanan Katsayıların Zamanla Değişen Üç Boyutlu Gösterimi	61
Şekil 5. 11. Meksika Şapkası	64
Şekil 5. 12. Mayer Dalgacığı	64
Şekil 5. 13. Morlett Dalgacığı	65
Şekil 5. 14. Gauss Dalgacığı	65
Şekil 7. 1 Kullanıcı Veritabanı Ekranı	70
Şekil 7. 2 Kullanıcılara ait Örnek Toplama Ekranı	70
Şekil 7. 3 Aynı kişiye ait iki telaffuz örneği	71
Şekil 7. 4 Yukarıdaki iki telaffuzun önışlemeden sonraki halleri	72
Şekil 7. 5 İki telaffuzun spektrogram görüntüleri	73
Şekil 7. 6 Kullanılan Mel-Filtre Bankları	73
Şekil 7. 7 İki telaffuza ait filtre çıkışları	74

1. GİRİŞ

İnsanların ve hayvanların en önemli yeteneklerinden birisi dostlarını diğerlerinden ayırt edebilme ve tanıyabilme özelliğidir. En eski zamanlardan beri bu yetenek hayatta kalmayı sağlamıştır. Bahsi geçen “*diğerlerini*” yanlış tanımak iyi niyetin suistimal edilmesine , elinizdeki malların veya servetin çalınmasına veya değerli bir bilginin yanlış ellere geçmesine neden olabileceği gibi sizin yada sevdiklerinizin hayatının tehlikeye girmesine neden olabilir. Terside de yani dostunuzu da “*diğeri*” zannetmek sosyal ve bireysel ilişkilerinizin bozulmasına neden olabilir.

İnsanlarda başkalarını tanıma ve tanımlama yeteneği bebeklikten başlayarak doğal olarak gelişir. Genel olarak yakınlarımızı yüzünden ve sesinden tanırız. Ancak bunlar tanıma yöntemlerinden sadece ikisidir. Günümüzde kimlik doğrulama işlemleri hayatımızın her alanında kullanılmakta ve önemli bir yer tutmaktadır. Bilgi teknolojileri sayesinde bu kimlik tanıma ve doğrulama işlemleri yüz yüze veya kişisel irtibat sağlanarak yapılma gereği ortadan kalkmaktadır. Geleneksel bağlamda kimlik doğrulama işlemleri için insanlara bir PIN (Kişisel Tanımlama Numarası) ve şifre verilmekte. Ve doğrulama işlemleri bunlar üzerinden yürütülmektedir (Yun 2006). Ancak gelin görün ki bu şifre mekanizması karşımızdakinin gerçekten o olup olmadığını belirleyemezler. İşlemleri doğru kişi tarafından yapıldığına emin olmak için bu işlemlerin yüz yüze yapılması gerekmektedir. Herkesin karşısına bir insan koymak ve kullanıcıların bu insanlarla temasa geçmesini sağlamak hem zaman kaybına yol açmakta hem de maliyetleri artırmaktadır. Çözüm olarak ta tanıma işleminin makinelere devredilmesi gereği ortaya çıkmaktadır. Bu noktada karşımıza biyometri kavramı ortaya çıkar.

2. KONUŞMA ve KONUŞMACI TANIMA

2.1. Biyometrik Teknoloji

Biyometrik, bireyin ölçülebilir fiziksel ve davranışsal karakteristiklerini tanıyarak kimlik saptamak üzere geliştirilmiş otomatik sistemler için kullanılan bir terimdir. Artan güvenlik ihtiyacının şifrelerle karşılanamayacağı görüldüğü için biyometrik sistemlerin geliştirilmesi kaçınılmaz olmuştur.

İnsan bedeninin bir parçasının ölçülmesi ile veri elde edilen Fiziksel Biyometrikler şunlardır:

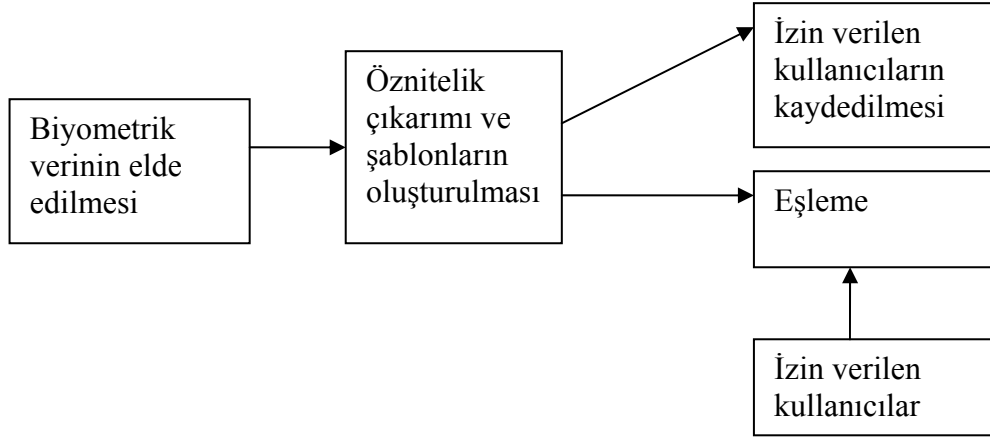
- Yüz Tanıma
- İris Tanıma
- Parmak İzi Tanıma
- Retina Tarama
- El İzi Tanıma
- Damar Tanıma
- DNA analizi
- Yüz Sıcaklık Eğrileri (Termogram)

Bir davranışın ölçülmesi ile veri elde edilen Davranışsal Biyometrikler;

- Ses Tanıma
- İmza Tanıma

Biyometrik Sistemler, bireyin belli biyolojik karakteristiklerini sadece o kişiye özel tek ve benzersiz bir koda dönüştürür. Bu kod elektronik ortama kaydedilir ve aktif Kimlik Saptanmasında kayıtlar ile ilgili kişi anında karşılaştırılır ve sonuca varılır (Markowitz 2006) .

Genel olarak biyometrik sistemler aşağıdaki gibi çalışır (Markowitz 2006).



Şekil 2. 1. Biyometrik Sistemlerin Genel Çalışma Prensibi

2.2. Ses ve Konuşma

2.2.1. Ses Nedir?

Ses, insan kulağını etkileyerek işitme duyusu oluşturan hava molekülleri titreşimleri, ya da bunların neden olduğu ufak hava basınç değişimleri gibi, ya da bu fiziksel olayın neden olduğu işitsel izlenim olarak tanımlanır. Sesin en temel fiziksel özellikleri ise yükseklik, şiddet ve tınıdır.

Sesin yüksekliği frekansı ile gösterilir. Frekansı yüksek seslere ince (tiz), frekansı düşük seslere kalın (pes) sesler denir. Bazı sesler o kadar yüksektir ki bunlar kulakları incitebilir.

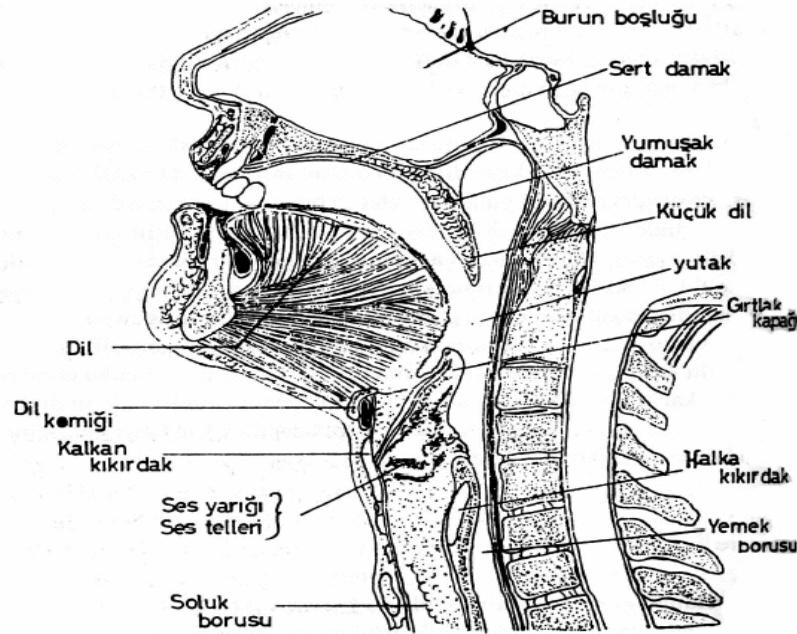
Sesin şiddeti, ses dalgalarının genliği ile belirlenir. Ses dalgalarının genliği arttıkça, sesin yüksekliği artar. Ses şiddeti desibel (dB) cinsinden ölçülür. Desibel ölçeği logaritmiktir. Böylece duyulabilen ses seviyesini 1 kabul edip, acı verici olarak kabul edilen ve bunun 10 trilyon katına kadar olan sesleri 1 dB ile 130 dB arasında ifade etmek mümkün olmuştur. Normal konuşmadan çıkan sesin gürültü seviyesi 60-70 dB civarında iken, jet uçaklarının çıkardıkları gürültü 120 dB'e, Satürn roketinin gürültüsü ise 200 dB'e ulaşır. Burada hemen belirtmelidir ki, gürültü seviyesindeki 10 dB'lik bir artış bizim gürültüyü iki kat daha fazla algılamamıza neden olur.

Sesin tınısı ise onu tanıtan bir özelliğidir. Örneğin, bir çocuk sesinin tınısı genelde bir yetişkinin sesinin tınısından daha büyüktür. Müzik ölçeğinde her bir nota farklı bir tınıya sahiptir. Sesin tınısı da yine frekansa bağlı bir özelliktir. Yüksek tınlı seslerin frekansı da büyüktür. Örneğin, bir piyano ile keman aynı notayı çaldıklarında kulağımızda uyandırdıkları duygular farklıdır. Bunu keman ile piyanonun verdiği seslerin tınları farklıdır şeklinde ifade ederiz. Sesin tınısı, ses kaynağının yapısını belirleyen bir özelliktir.

2.2.2. Konuşma Nedir?

Konuşma insanlar arası iletişimin en önemli parçasıdır. Konuşmanın tanımı şöyle yapılabilir: Beyinde oluşan bir iletinin konuşma örgenlerinden yararlanılarak dinleyen kişiye ses titreşimleriyle iletilmesidir. Görüldüğü gibi konuşmanın farklı özellikleri vardır. Bu özellikler: zihinsel, fizyolojik ve fizikseldir.

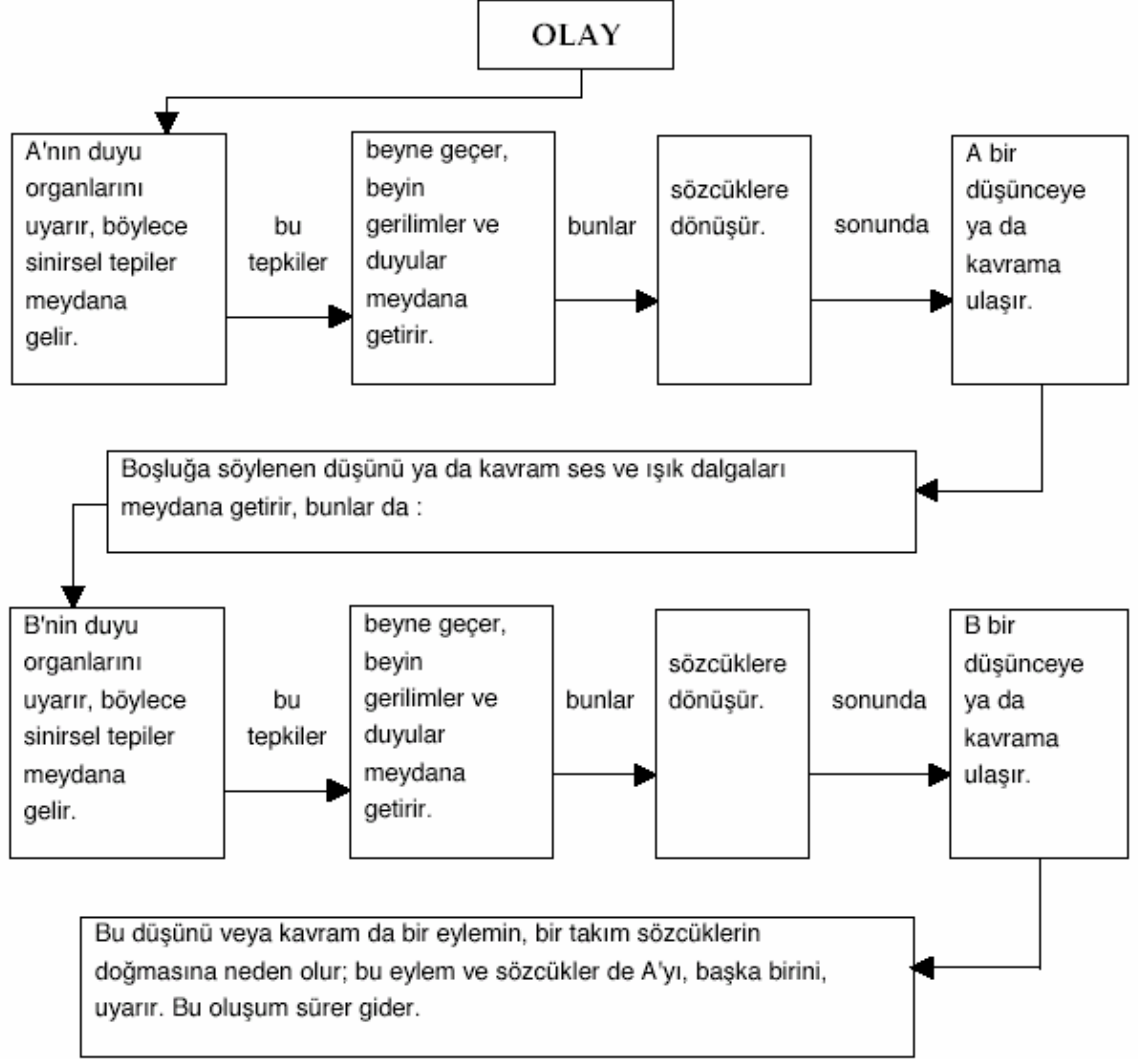
Zihinsel özellik konuşmanın beyinde oluşturulması, fizyolojik özellik beyinde oluşan bu durumun sese dönüştürülmesi için konuşma örgenlerinin hazırlanması, fiziksel özellik ise sesin duyulabilir olmasını sağlayan ses titreşimleridir (Çelik 2002).



Şekil 2. 2. Ses Yolu (Ses Üretim Organları)

Tanımda dikkat edilmesi gereken bir nokta da konuşmayı dinleyecek bir dinleyicinin olması gerektiğidir. Konuşma iletişimdir. İletişim için bir ileti gerekir. İletişimin sağlanması bu iletinin gönderilmesi demektir. Bunun için de konuşma alıcı

ve verici arasında gerçekleşir. Alıcı dinleyen, verici ise konuşandır. Konuşma etken, dinlemek ise edilgen bir eylemdir. İletişimde etken durumdaki konuşmanın oluşması bir sürecin sonunda gerçekleşir. Bu süreci şöyle bir şema ile gösterilmektedir (Aydın 2005):



Şekil 2. 3. Konuşma Süreci

Konuşma sesinin oluşması için akciğerlerdeki havanın dışarı çıkarken gırtlakın sağ ve sol yanında bulunan ikisi gerçek ikisi yalancı olan dört ses teline çarpması gerekmektedir.

İnsan konuşurken, soluk şu örgenlere çarpılarak dışarı çıkar: *Akciğer, soluk borusu, gırtlak, ses telleri, küçük dil, ağız boşluğu, geniz, burun boşluğu, damak, dişetleri, diş, dil, dudaklar.*

Sesin oluşmasında birinci derecede rol oynayan ses telleri, önde kalkan kıkırdağın içiyle halka kıkırdağın iç kenarları arasına yerleşmişlerdir. Arkada, üçgen piramit biçimindeki ibriksi kıkırdakların iç yüzeyine bağlıdır.Boğumlama konuşma örgenlerinin akciğerden gelen soluğa biçim vermesidir. Bu, sesi anlaşılır kılar. Boğumlama konuşmanın temel ögesidir. İnsanların başlangıçta, boğumlamayı öğrenmeden önce, hayvanlar gibi sesler çıkardıkları varsayılır. Süreç içinde boğumlamayı öğrendikleri zaman konuşmaya başlamışlardır. Diyaframın, göğüs kaslarının, kaburgaların yardımıyla akciğerden gelen basınçlı hava, ses tellerindeki titreşimle ses yarığında, yani gırtlak içinde sesi oluşturur (ancak kimi seslerin oluşumunda ses telleri hareketsizdir). Bu durumda titreşimin üretimi gırtlakta gerçekleşir denilebilir. Tınlama ise ses yarığında yukarıda yutak, ağız ve burun boşluklarında sağlanır (Aksan, 1995). Bu örgenlerle birlikte daha önce belirtilen soluğun dışarı çıkarken çarptığı örgenler çeşitli kapanma, engelleme ve hareketlerle sesin değişik biçimlerde oluşmasını sağlar. İşte bu süreç boğumlama olarak adlandırılmaktadır.

Herhangi bir sözcüğün söylenmesi için genel olarak o sözcüğü oluşturan sesbirimlerin her biri için belirli ve kimi zaman birbirinden çok değişik hareketlerin yapılması gerekir.

3. KONUŞMACI TANIMA

En temel seviyede konuşma, iki tür bilgi içerir.konuşmanın içeriği hakkında bilgi ve konuşmacının kimliği hakkında bilgi. Bu bilgileri kullanan uygulamalar ikiye ayrılabilir, “*Konuşma Tanıma*” ve “*Konuşmacı Tanıma*”.

Bilgisayar temelli konuşma tanıma programları konuşulan metnin içeriğini çıkarır ve kullanırlar. Bu sayede bilgisayarda giriş için kullanılan klavye ve fare gibi çevre birimlerinin yerini almaya çalışırlar. Bu güne kadar yapılan çalışmalarla bilgisayarla konuşarak iletişim kurmakta kaçınılmaz olacaktır.

Konuşmacı Tanıma (Speaker Recognition) konuşmaya ait ses dalgalarından elde edilen bilgiye dayanılarak konuşanın kim olduğunu otomatik olarak bulma işlemidir. Bu teknik, erişim kontrolünün sesle yapıldığı sistemlere erişmeye çalışan kişilerin, kimliklerinin belirlenebilmesini mümkün kılmaktadır (Furui 1996). Bu sistemler içerisinde sesli arama, telefon üzerinden bankacılık işlemleri, telefonla alışveriş, veritabanı erişim servisleri, bilgi ve rezervasyon servisleri, sesli posta, gizlilik gerektiren bilgilere erişimde güvenlik kontrolleri, bilgisayarlara uzaktan erişim, sayılabilir. Konuşmacı tanıma teknolojisi yeni eklenen servislerle günlük hayatımızda kolaylıklar sağlamaktadır. Konuşmacı tanıma teknolojilerinin kullanıldığı en önemli alanlardan biride adli uygulamalardır.

3.1. Konuşmacı Tanıma Teknikleri

3.1.1. Konuşmacı Tanıma Teknolojilerinin Sınıflandırılması

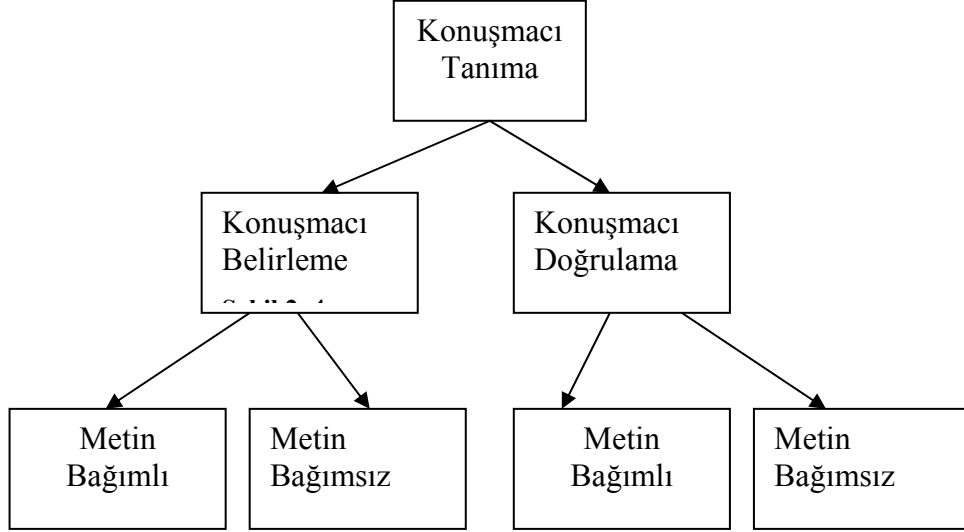
Konuşmacı Tanıma, Konuşmacı Belirleme ve Konuşmacı Doğrulama olarak ikiye ayrılabilir. Konuşmacı belirleme, elde edilen telaffuz verileriyle, konuşan kimsenin kayıtlı konuşmacılardan hangisi olduğunun belirlenmesi işlemidir. Konuşmacı doğrulama işlemi ise iddia eden kişinin gerçekten o olup olmadığına karar verilmesidir. Belirleme ve Doğrulama işlemlerinin temel farklılık karar verme alternatiflerinin sayısıdır. Belirleme işleminde bu alternatiflerin sayısı kayıtlı popülasyon sayısı kadardır. Doğrulama da ise karar verme alternatif sayısı sadece “kabul” ya da “red” olarak ikidir. Konuşmacı belirlemenin performansı

popülasyonun artmasına baęlı olarak düşer, tanıma da ise popülasyona baęlı deęildir ve sabit bir yaklaşım gösterir.

Konuşmacı belirlemede, belirlenecek kullanıcının var olmaması durumuna göre “açık küme” kavramı da mevcuttur. Bu durumda bir başka karar alternatifi ortaya çıkar : “ Bilinen hiçbir modelle eşleşmedi” Hem belirleme de hem de doğrulamada, karar vermede kullanılacak eşleşmenin ne kadar doğru olduğunu belirlemek için bir başka bir eşik testi uygulanır. Buna göre kararın kabul edileceğine yada yeni bir denemenin daha yapılacağına karar verilebilir.

Konuşmacı tanıma metotları ayrıca metin baęımlı veya metin baęımsız metotlar olarak ikiye ayrılırlar. Bu metotların ilkinde konuşmacının hem eğitim hem de tanıma basamaklarında aynı kelimeleri veya cümleleri telaffuz etmesi gerekmektedir. Diğerinde ise belirli bir metnin konuşulması gerekmemektedir. Metin baęımlı sistemler genellikle şablon eşleme tekniklerini kullanır. Bu teknikte konuşma örneęi zaman ekseninde her bir referans modelle karşılaştırılır. Ve konuşmanın başından sonuna kadar benzerlik miktarı karşılaştırılır. Metin baęımlı tanıma sistemleri göreceli olarak çok daha basittir. Bu yöntem sesin bileşenleri olan fonem ve hecelere birebir uygulanabilir. Bu yöntemle, “Metin Baęımsız” yöntemlere göre çok daha büyük başarımlar elde edilir.

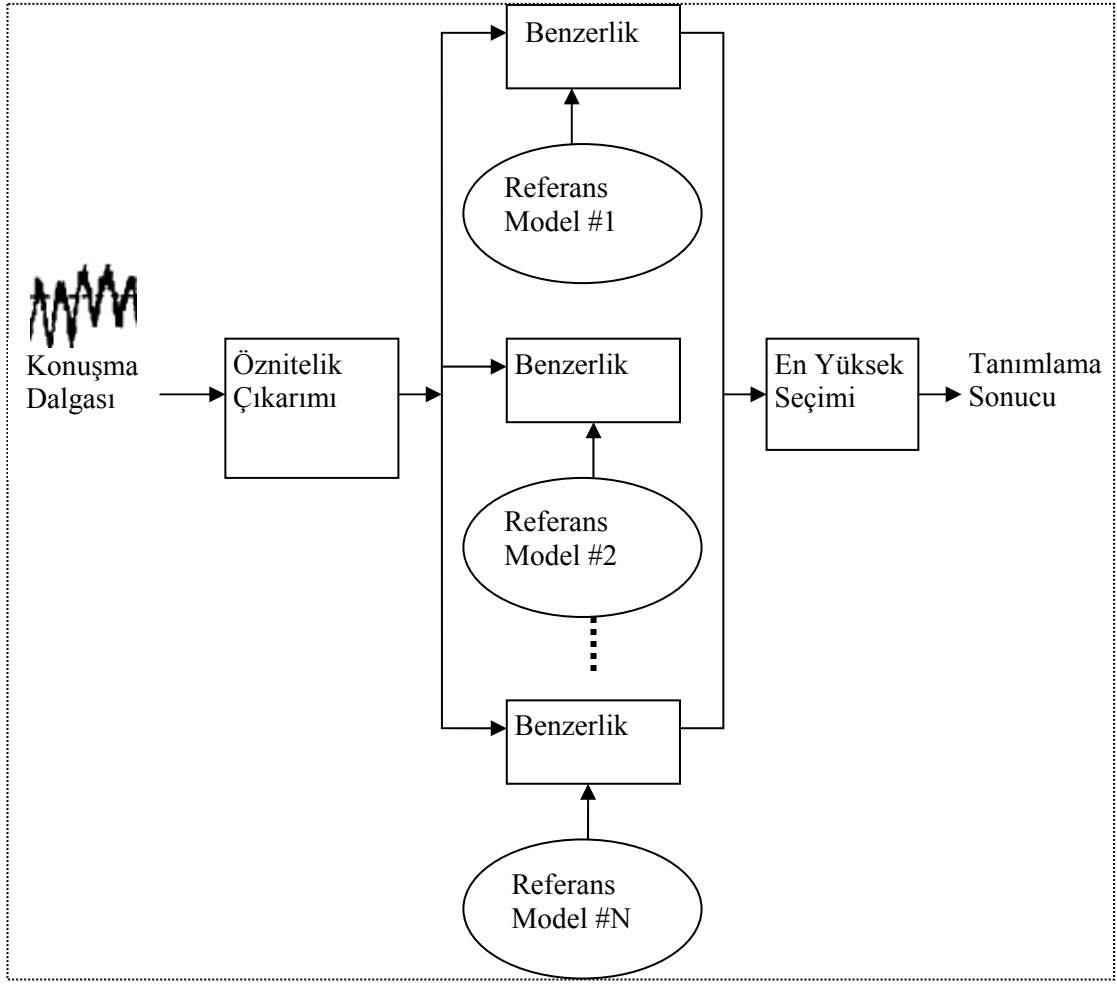
Ancak adli veya gizlilik gerektiren uygulamalarda önceden belirlenmiş kelimeler kullanılamaz. Ek olarak insanoęlu konuşmanın içerięinden baęımsız olarak konuşmacıyı tanıır. Bu yüzden metin baęımsız sistemler son zamanlarda daha popüler olmuşlardır. Metin baęımsız sistemlerin bir başka avantajı da kelimeleri tekrarlamaya gerek kalmadan sırasal olarak istenen seviyeye ulaşana kadar analiz eder (Furui 1996).



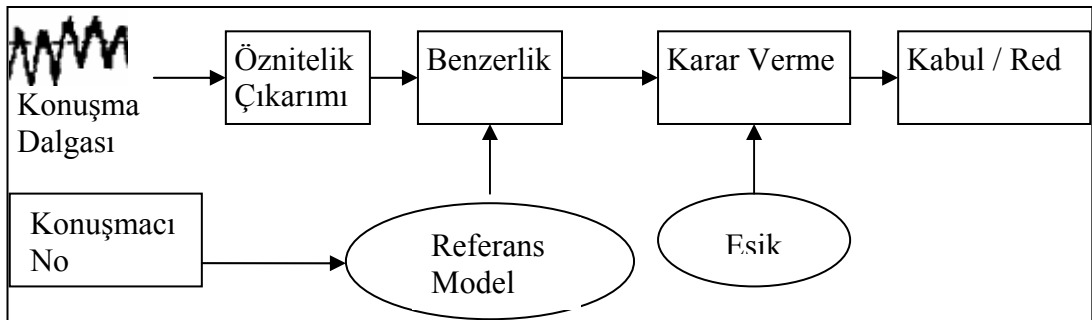
Şekil 3. 1. Kuşmacı Tanıma Alt Dalları

3.1.2. Kuşmacı Tanıma Sistemlerinin Genel Yapıları

Aşağıdaki Şekil konuşmacı tanıma sistemlerinin genel yapısını gösterir. Kuşmacı belirlemede, bilinmeyen konuşmacıdan alınan telaffuz analiz edilir ve bilinen modellerle karşılaştırılır. Sistem tarafından yapılan analiz sonucunda giriş en uygun eşleştirme yapılır. Kullanıcı doğrulama da ise bilinmeyen bir kullanıcı tarafından bir kimlik beyan edilir. Bu bilinmeyen kullanıcıya ait giriş iddia ettiği kimlikle karşılaştırılır. Eğer karşılaştırma belli bir eşik değerinin üstünyse kimlik doğrulanır. Yüksek bir eşik değeri sahtekarların sistem tarafından kabul edilmesini güçleştirir ama gerçek beyanların reddedilme riski de artar. Düşük eşik değerinin getireceği sakıncalarda ortadadır. Uygun seviye de bir eşik değer tanımlamak için doğru ve yanlış beyanların değerlendirmesinin yapılması gerekmektedir. Bu amaçla ROC (Receiver Operating Characteristics) eğrisi kullanılabilir (Markowitz 2006).



Şekil 3. 2. Kuşmacı Belirleme Sistemlerinin Genel Çalışma Prensibi



Şekil 3. 3. Kuşmacı Doğrulama Sistemlerinin Genel Çalışma Prensibi

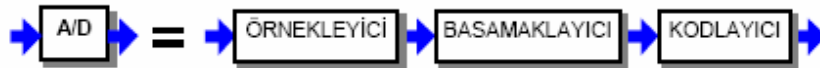
3.1.3. Konuşma Sinyallerinin Temel Özellikleri

Sesli ifadeyi oluşturan ses dalgasının, sıklık(frequency) ve genlik (amplitude) özellikleri vardır. Genlik sesin şiddetini, taşıdığı enerjiyi belirler. Sıklık ise sesin tizlik ve peslik özelliklerini belirler. Sesli ifadenin, farklı sıklıkta çok sayıda sinüsel sinyalin üst üste binmiş biçimi olduğu söylenebilir. Konuşmacı tanıma sistemleri de genelde sesli ifade sinyallerinin analizinden, ayrıştırılmasından yararlanmaktadır. Sesli ifadenin içerdiği seslerdeki sıklık ve genlik değerleri her bir fonem için farklılık göstermektedir.

Tanımanın temelinde bu ayırıcı özellikleri bulmak ve sınıflandırmak yatmaktadır. İnsanın duyabileceği ses için bazı alt ve üst sınırlar vardır. İnsan kulağı kabaca 20Hz ile 20KHz sıklıklar arasındaki seslere duyarlıdır. Bu aralığın dışında kalan sesler algılanamaz. Genlik için de bir sınır söz konusudur. İnsan kulağı en az 20db şiddetindeki sesleri duyabilir. 120db değerinden daha fazla şiddetteki sesler ise insan kulağında hasara yol açmaktadır. Ses dalgası periyodik bir dalgadır. Periyodik bir dalgayı bir dizi sinüsel dalganın birleşimi şeklinde ifade etmek mümkündür. Sesli ifade için de bu durum geçerlidir. Periyodik bir sinyali oluşturan sinüsel dalgalar harmonik olarak adlandırılır. Bu dalgalardan periyodik sinyalin sıklığıyla aynı sıklığa sahip olanına temel harmonik veya temel sıklık (*Fundamental Frequency*) denir(Mengusoglu 2002). Sesli ifade tanıma sistemlerinin çoğu sesli ifadenin harmoniklerinden yararlanmaktadır. Temel harmonik ise sesli ifadedeki sözcüklerin sınırlarını bulmak için sıkça kullanılan bir yöntemdir.

3.1.4. Konuşma Sinyallerinin Sayısallaştırılması

Konuşmacı tanımanın ilk aşaması, sesli ifadenin elde edilmesidir. Bu amaçla genellikle mikrofon veya telefon kullanılır. Bu aşamada elde edilen konuşma sinyali analogdur. Sürekli zamandaki analog sinyalin öncelikle ayrık zamana çevrilmesi gerekir.

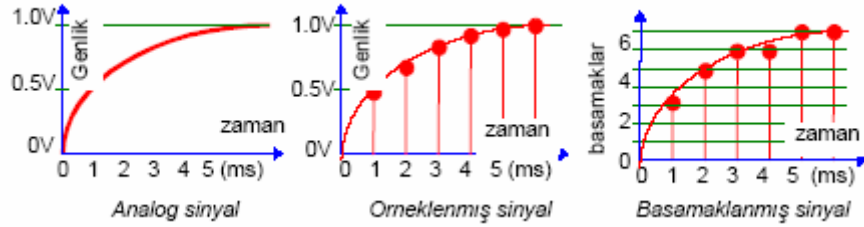


Şekil 3. 4. Konuşma Sinyalinin Sayısallaştırma Adımları

Ayrık zamana dönüştürme işlemi *örnekleme* ile yapılmaktadır. Belli zaman aralıklarında sinyalin değeri ölçülür ve elde edilen bu değere *örnek* (sample) adı verilir. Eğer analog sinyal yüksek frekans bileşenlerine sahipse bilgi kaybına

uğramamak için örnekleme işlemini daha yüksek oranlarda yapmamız gerekir. Genel olarak analog sinyalden bilgi kaybına uğramamak istiyorsak en yüksek frekansın iki katı kadar bir örnekleme oranına ihtiyacımız vardır. Bu Nyquist oranı olarak bilinir. Aksi halde *spektrum örtüşmesi (aliasing)* yüzünden bozulmalar meydana gelir ve orijinal sinyal tekrar elde edilemez.

Alınan örneklerin genlikleri herhangi bir değerde olabilir. Buna karşılık işaretin sayısal çevrilebilmesi için kullanılacak seviye sayısının sınırlı olması gerekir. Bu sayı, her bir örnek için kullanılacak kod uzunluğu ya da bit sayısı tarafından belirlenir. Örnek olarak 8-bit'lik bir kodlama yapılacaksa 256 seviye, 3-bit'lik bir kodlama yapılacaksa sadece 8 seviye kullanılabilir. Seviye veya basamak sayısının artması alıcı tarafta sayısal/analog dönüştürücü çıkışında elde edilecek sinyalin kalitesini belirler. Daha iyi kalite için daha çok bit ve daha çok basamak kullanmak gerekir. Örnek olarak 0-1V arası değişen bir sinyali 3-bitlik bir kodlama ile sayısallaştırmak istiyorsak basamak sayısı 8, aralık sayısı ise $8 - 1 = 7$ dir. 1 volt 7 aralığa bölünürse iki basamak arası 0,143V olur. Basamak sayısı belli olduktan sonra her basamağa karşı düşen bir kod oluşturulur. Bu, genelde, basamak numarasının ikili sayı sistemindeki karşılığıdır.

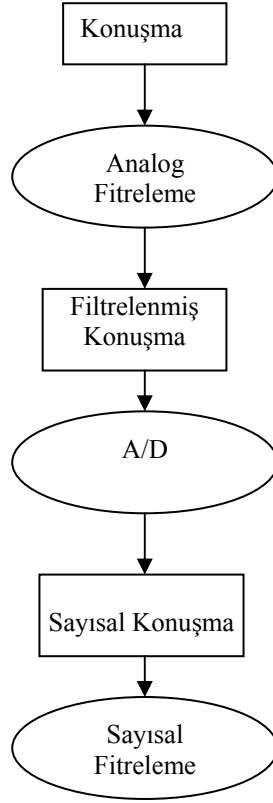


Şekil 3. 5. Analog Sinyalin Sayısallaştırma Adımları

Gerilim (V)	0	0,143	0,286	0,429	0,571	0,714	0,857	1,00
Basamak	0	1	2	3	4	5	6	7
Kod	000	001	010	011	100	101	110	111
İletilen Sinyal	[Sinyal dalgası]							

Şekil 3. 6. Analog/Sayısal Çevrim

Analog sinyalin sayısallaştırılmasından önce ve sonra filtreleme işlemleri yapılarak sinyal üzerindeki gürültünün giderilmesi, anlamlı bilgilerin ön plana çıkarılması da gerekebilir. Aşağıdaki şekil konuşma sesinin elde edilme aşamalarını gösterir



Şekil 3. 7. Konuşma Sinyaline Uygulanan Ön İşlemler

Analog filtreler genellikle mikrofonun içerisine yerleştirilmişlerdir ve gürültü temizleme amacıyla kullanılır. Sayısal Filtreler ise elde edilmiş sayısal sinyalin öznitelik vektörlerinin daha iyi çıkarılabilmesi için sinyalin genişletilmesi amacıyla kullanılır. Pek çok öznitelik vektörü çıkarma yöntemi zaten bir filtreleme yöntemini de beraberinde getirmektedir.

Analogdan sayısala çevrimden sonraki aşama ise elde edilmiş sinyalin kodlanmasıdır. Genel olarak, **Dalgaformu kodlama (PCM)**, **Kaynak kodlama** ve **Hibrid kodlama** olarak üç çeşit kodlama biçimi vardır. Dalgaformu kodlama, çevrilmiş sinyalin eski haline mümkün olduğu kadar yakın dönüştürülebilmesini amaçlar. Ve sinyalin örnekleme basamaklandırılması temeline dayanır(Mengusoglu 1999).

PCM (Pulse Code Modulation) sesin doğrusal niceleme ile dijital formata dönüştürüldüğü en basit yöntemdir. Temel olarak 8 kHz'de ses sinyalini örnekleyerek nicemler. Çıkış akışı yaklaşık 64 kBit/saniyedir. Bu sebeple bu çeşit kodlama gerçek zamanlı sistemlerde, yüksek bant genişliği isteyeceği, hafıza ve

kaynak sıkıntısı yaratacağı için pek uygun değildir ancak basitliği nedeniyle de tercih edilmektedir.

DPCM (Differential Pulse Code Modulation), PCM'e göre daha etkili bir yöntemdir çünkü ses sinyali içindeki gereksiz kısımları, daha sonra önceki ve sonrakinden örneklenebilecek şekilde atar. Böylece sıkıştırıcının tek yaptığı birbiri ardı sıra gelen örneklerdeki farkı belirtmektir. Çözme işlemi sırasında bu sinyaller yeniden oluşturulur.

ADPCM (Adaptive Differential Pulse Code Modulation) 32 kBit/saniye gibi oranlarda çok yüksek ses kalitesi sağlayabilir. 16, 24,32 ve 40 kBit/saniyelik bit akış oranlarında çalışacak şekilde standart hale gelmiştir. ADPCM algoritma olarak PCM'den farklıdır çünkü örneklenmiş ses sinyalinin nicelenmesinin yerine önkestirilen ve nicelenen sinyal arasındaki farkı niceler. İyi bir önkestirimde gerçek sinyal ile tahmini sinyal arasındaki fark çok küçük olacaktır ve bu da daha düşük bit akış hızı anlamına gelecektir. Arkasında çalışan niceleyici tek tip değildir ve farklı sinyal modellerinde kullanılmak üzere eniyilenebilir. Sinyalin yeniden üretilmesi nicelenmiş farkın tahmini sinyale eklenmesiyle bulunur. Bu sayede orijinal sese çok yakın bir sinyal elde edilmiş olur.

Kaynak kodlama ise model tabanlıdır ve genellikle insan sesinin modellenmesi için geliştirilmiştir. Kodlayıcılar içinde belki de en yaygın olanı ve en çok kullanılanı Doğrusal Önkestirimci Kodlayıcı (Linear Predictive Coder) olmuştur. Bu yaklaşımda sesin örnekleme zamanındaki değerinin, geçmiş değerlerin lineer bir fonksiyonu olarak kestirilebileceği varsayılır.

4. SİNYAL MODELLEMEDE TEMEL YAKLAŞIMLAR

Bir analog sinyalin modellenmesi konuşma/konuşmacı tanıma işleminin ilk aşamasıdır. Sinyal modelleme konuşmaya ait örneklerin bir olasılık uzayındaki olayları gösterecek olan gözlem vektörlerine dönüştürülme işlemini ifade eder.

Sinyal Modelleme dört temel işleme ayrılabilir. Bunlar Spektral (Tayf) Çevrim, Spektral Analiz, Parametrik Dönüşüm ve İstatistiksel Modellemedir.

Sayısal Sinyal Modelleme sistemlerini etkileyen üç temel unsur bulunmaktadır. Bunlardan birincisi **algısal-anlamlı** parametreler olarak adlandırılan ve analog sinyal içerisinde insan duyu sistemi tarafından da kullanılan göze çarpan parametrelerin ortaya çıkarılmasıdır. İkincisi elde edilen bu parametrelerin iletim kanalına, konuşmacıya ve çeviricilerdeki farklılıklara karşı sağlamlığıdır. Buna da **dayanıklılık** veya **değişmezlik**(invariance) problemi denir. Son olarak ta, yeni yapılan çalışmalarla birlikte, bu parametrelerin spektral dinamikleri veya zamanla spektrumdaki değişiklikleri ortaya çıkarmaları da istenmektedir. Bunlara da **zamansal ilişki** (temporal correlation) denir (Joseph 1993).

4.1. Spektral Dönüşüm

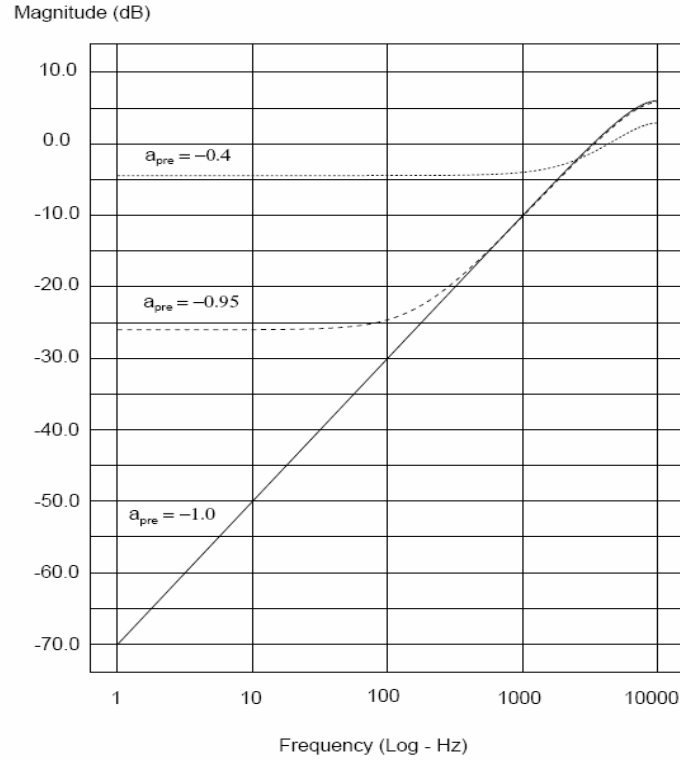
Spektral dönüşüm iki temel işlemi gerektirir. Birincisi **A/D dönüşümü** – sinyalin ses basınç dalgasından sayısal bir sinyale dönüştürülmesi ve **sayısal filtreleme** – sinyal içerisindeki önemli frekans bileşenlerinin ortaya çıkarılması. Bu konu bir önceki kısımda anlatılmıştı. A/D dönüşümlerinde kullanılan mikrofonlar genellikle hattaki frekans kirliliği (50/60 Hz) , yüksek ve düşük frekans bilgisinin kaybolması veya lineer olmayan bozulmalar gibi kenar etkilerine maruz kalırlar. Sayısallaştırma işleminin ana amacı da konuşma sinyalindeki örneklenmiş verinin mümkün olduğunca yüksek SNR'a (Sample to Noise Ratio)(Örnek-Gürültü Oranı) ulaşmalarını sağlamaktır.

Sinyal dönüşümü tamamlandıktan sonra son adım olarak **Önvurgulama** (preemphasis) FIR fitresi kullanılır (Mengusoglu 2002):

$$H_{pre}(z) = \sum_{k=0}^{N_{pre}} a_{pre}(k) z^{-k} \quad (4.1)$$

$$H_{pre}(z) = 1 + a_{pre} z^{-1} \quad (4.2)$$

tipik olarak a_{pre} için $[-1.0, -0.4]$ aralığında bir değer seçilir.



Şekil 4. 1. Preemphasis Filtresinin Frekans Cevabı

Bu filtreyi kullanmanın iki avantajı vardır. Birincisi konuşma sinyalinde sesli kısımlar doğası gereği ses üretim sistemlerine göre her onlukta yaklaşık olarak 20 dB kadar negatif spektral eğilime sahiptirler. Önvurgulama filtreler, analizin etkinliğini artırmak için bu eğilimi doğal seviyeye çekerler. Bir başka şekilde de söylemek gerekirse insanın duyması spektrumun 1 kHz bölgesi üzerinde daha duyarlıdır. Önvurgulama filtresi ile bu bölge yükseltilir.

4.2. Spektral Analiz

Konuşma/Konuşmacı tanıma sistemlerinde kullanılan spektral ölçümleri iki kısma ayırabiliriz. Birincisi **Güç** : sinyalin kaba spektral ölçümleri , ikincisi ise Spektral Genlik : spektrumdaki belirli frekans aralıkları üzerindeki güç ölçümü. Konuşma/Konuşmacı Tanıma sistemlerinin parametre kümesinde bu iki ölçümde bulunmaktadır. Son zamanlarda temel frekans bileşeni de bu konularda çalışanların dikkatini çekmekte ve üzerinde durulmaktadır.

4.2.1. Temel Frekans (Perde) (Pitch)

Konuşmada sesli harflerin söylenmesi sırasında ses tellerinin titreşimine ait frekans temel frekans olarak adlandırılmaktadır. Temel Frekansın (f_0) konuşma sinyali içerinden hesaplanması oldukça güçtür. Bugün bu işlemin yapılması için dört ana sınıf algoritma bulunmaktadır.

- Gold-Rabiner Algoritması (40)
- LPC10e Algoritması(42)
- Gürbüz Perde Bulma(44)
- Ters-Spektrum Perde Tespiti (45)

Temel frekans bileşeni insan duyma sistemiyle paralellik kurmak amacıyla genellikle lineer bir ölçek yerine logaritmik bir ölçekte değerlendirilir (Joseph 1993).

4.2.2. Güç

Günümüzde konuşmacı tanıma sistemlerinde güç ölçülerinin kullanımı standart olmuştur.

$$P(n) = \frac{1}{N_s} \sum_{m=0}^{N_s-1} (w(m)s(n - \frac{N_s}{2} + m))^2, \quad (4.3)$$

N_s gücü hesaplanacak örnek sayısını, $s(n)$ sinyali, $w(m)$ ağırlık fonksiyonunu, ve n pencerenin merkezini belirleyen örnek indisini belirtir. Pek çok tanıma sistemi gücü direk olarak kullanmak yerine insan duyma sistemini taklit ederek.

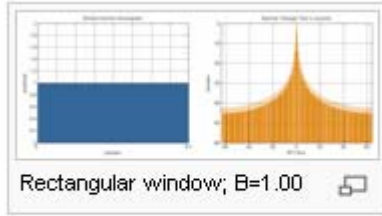
$$p_{dB}(n) = \log_{10}(P(n) * 10) \quad (4.4)$$

formülünü kullanır.

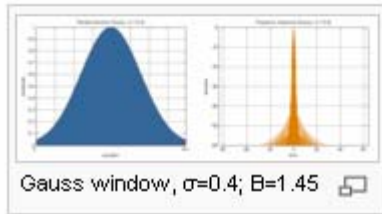
Ağırlık fonksiyonu pencere fonksiyonu olarak ta nitelendirilir. Pencere fonksiyonunda amaç eldeki sinyali belli aralıklarla sınırlamaktır. Örneğin aralık içerisinde sabit bir değer alan, aralık dışında sıfır olan fonksiyon **dikdörtgen pencere** fonksiyonudur. Bir fonksiyon veya sinyal bu pencere ile çarpıldığında aralık dışındaki değerleri sıfır olur.

Sesli ifade verilerinin önışleme aşamasında kullanılan belli başlı pencereleme yöntemleri aşağıda yer almaktadır.

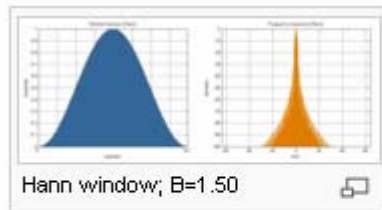
- Dikdörtgen Pencere
- Barlett Penceresi
- Hanning Penceresi
- Hamming Penceresi
- Blackman Penceresi
- Kaiser Penceresi.



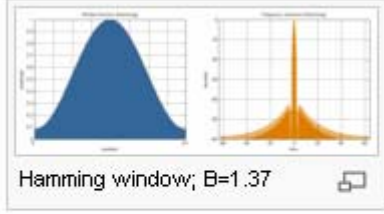
$$w(n) = 1$$



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



$$w(n) = 0.5 \left(1 - \cos \left(\frac{2\pi n}{N-1} \right) \right)$$



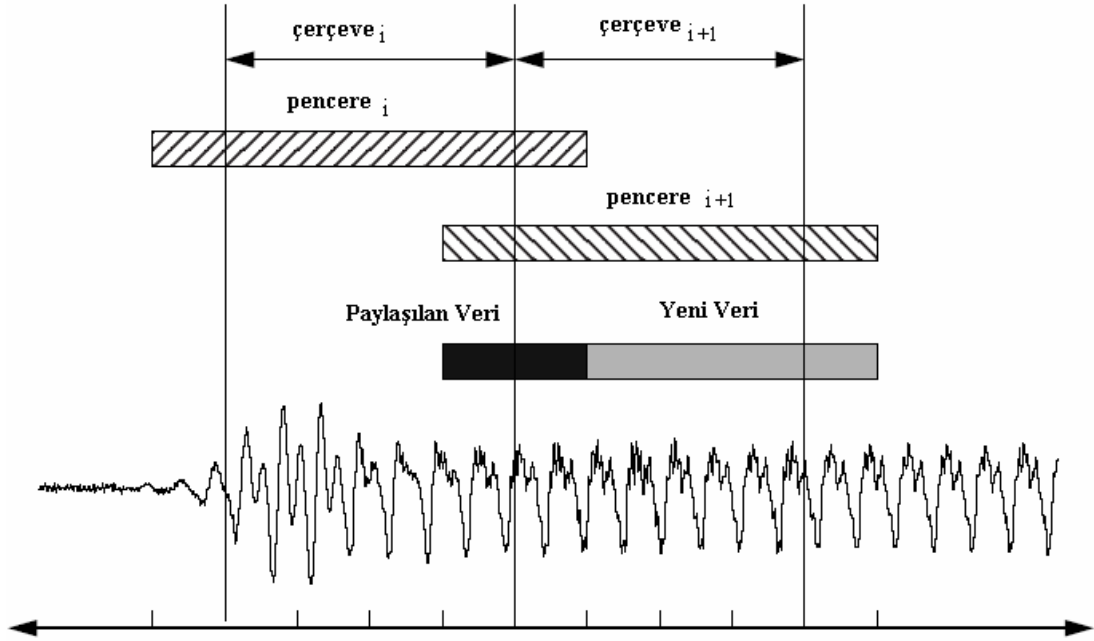
$$w(n) = 0.53836 - 0.46164 \cos\left(\frac{2\pi n}{N-1}\right)$$

Şekil 4. 2. Pencereleme Fonksiyonları

Pencerenin amacı pencerenin merkezinde kalan örnekleri ağırlandırarak ön plana çıkarmaktır. Böylece yavaş değişim gösteren parametrelerin elde edilmesini sağlayan bir fonksiyon elde edilir.

Güç konuşma/konuşmacı tanıma sistemlerindeki diğer parametreler gibi çerçeve temelinde hesaplanır. **Çerçeve süresi** (frame duration) T_f parametreler kümesinin geçerli olduğu zaman uzunluğu olarak tanımlanır. Çerçeve periyodu ise benzer bir şekilde arka arkaya parametre hesaplamaları arasında geçen süredir. Çerçeve oranı (Frame Rate) ise saniyede hesaplanan çerçevelerin sayısıdır(Hz). Çerçeve süresi 20 ms ile 10 ms arasında seçilir.

Aynı şekilde önemli olan bir kavramda gücün hesaplandığı aralıktır. Toplamın hesaplanacağı örnek sayısı N_s **pencere süresi** olarak bilinir. Pencere süresi ile çerçeve süresi birlikte güç hesabındaki sinyalin değişimini izlerler. Genel uygulamalarda aralarında bir oran vardır. Ve eğer çerçeve oranı 20 ms ise Pencere süresi 30 ms dir. Eğer 20 ms lik bir pencere süresi kullanılıyorsa 10 ms lik çerçeve süresi kullanılır. Genel olarak anlatmak gerekirse kısa çerçeve süreleri spektrumdaki hızlı değişimleri yakalar.



Şekil 4. 3. Çerçeve Bazlı Analiz

Bazı kaynaklarda bu tür analize örtüşmeli (overlapping analysis) adı verilir. Çünkü her çerçeve değişiminde sinyal verisinin sadece bir kısmı değişir. Örtüşme oranı aşağıdaki gibi hesaplanır

$$\text{Örtüşme \%} = \frac{(T_w - T_f)}{T_w} \times 100\% \quad (4.5)$$

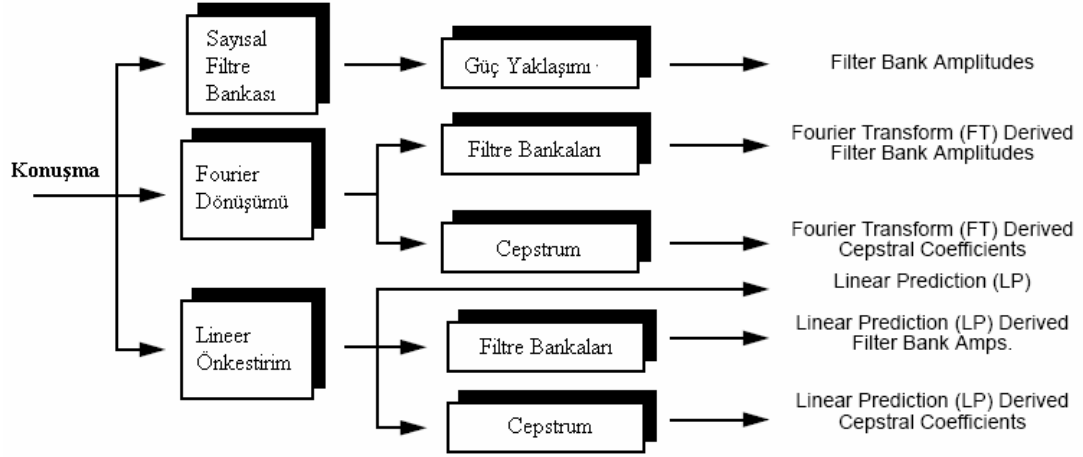
Burada T_w pencere süresi, T_f ise çerçeve süresidir.

Eğer $T_w < T_f$ ise örtüşme yoktur ve oranı sıfırdır.

Örtüşme oranını artırmanın bir amacı durağan olmayan kanal gürültülerini ve pencere yerleşiminden kaynaklanan gürültüleri azaltmaktır. Öte yandan fazlaca yumuşatılmış hesaplar sinyaldeki gerçek değişimlerin gözlenmesini engelleyebilir.

4.2.3. Spektral Analiz

Konuşma/Konuşmacı tanıma sistemlerinde kullanılan 6 ana spektral analiz yöntemi vardır.



Şekil 4. 4. Spektral Analiz Yöntemleri

4.2.3.1. Sayısal Filtre (Süzgeç) Bankası

Sayısal filtre bankaları konuşma tanıma sistemlerinde kullanılan en temel kavramlardan birisidir. İnsan duyma sistemine uyum sağlayan dahili basamakların kaba bir modeli olarak söz edilebilir. Filtre bankalarının temsiline iki ana güdüsü vardır. Birincisi kulaktaki zarın yerleşiminin saf tonlara karşı olan tepkisi tonun frekansı ile logaritmik olarak orantılıdır. İkincisi ise deneyler sonucunda, insanların, belirli bir bant genişliğindeki seslere ait karmaşık frekansları tek başına algılayamadığını göstermiştir. Bu sese ait bileşenlerden birisi bu bant genişliğinin dışına çıkarsa ayırt edilebilir. Bu bant genişliğine **kritik bant genişliği** denir ve sese ait merkez frekansın %10 ile %20 civarındadır.

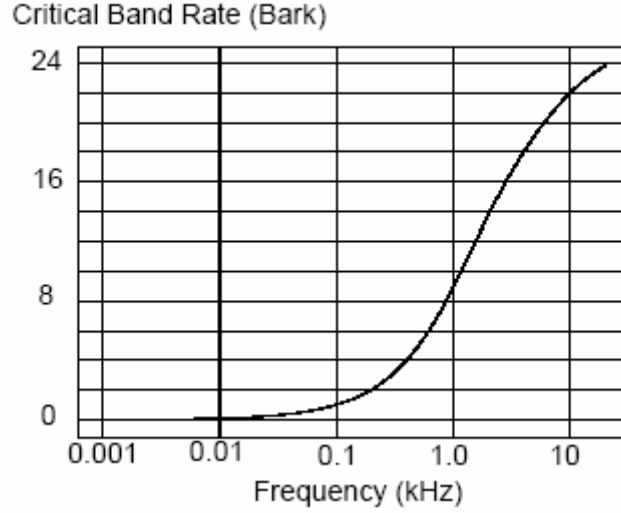
Akustik frekans f 'i algısal frekans ölçeğine aşağıdaki gibi eşleyebiliriz.

$$Bark = 13 \operatorname{atan}\left(\frac{0.76f}{1000}\right) + 3.5 \operatorname{atan}\left(\frac{f^2}{(7500)^2}\right) \quad (4.6)$$

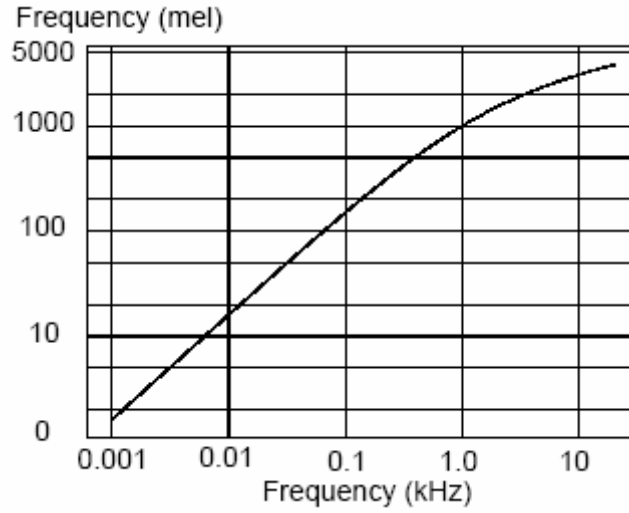
Bu algısal frekans ölçeğinin birimlerine kritik bant oranı veya *Bark* denir.

Konuşma tanıma uygulamalarında bu tip eşlemelerden en çok kullanılanı Mel-ölçeğidir.

$$\text{Mel Frekans} = 2595 \log_{10}(1 + f / 700.0) \quad (4.7)$$



Şekil 4. 5. Bark Ölçeği Dönüşümü

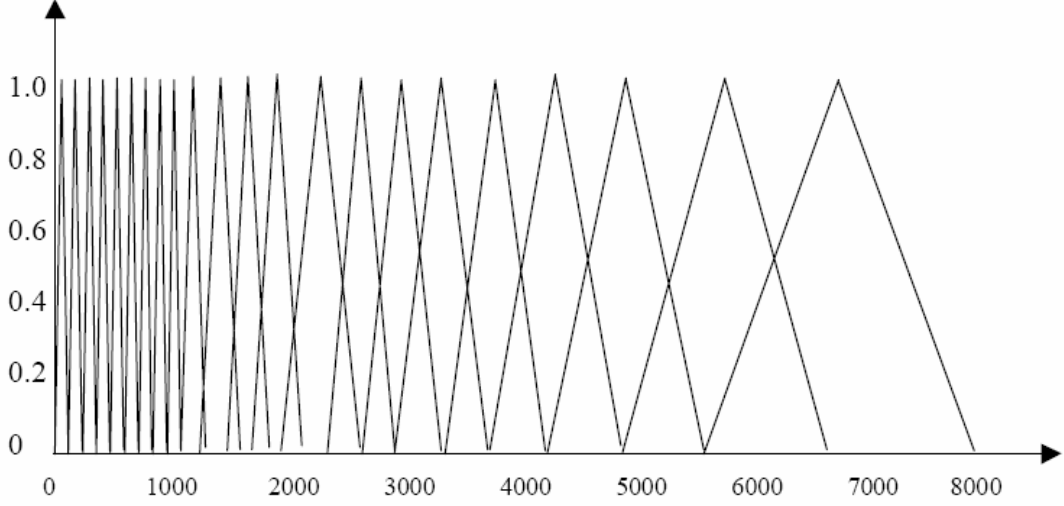


Şekil 4. 6. Mel - Frekans Dönüşümü

Örneğin 256 değer için alınan güç spektrumunda 256 boyutlu bir özellik vektörü kullanılmaz. Bunun yerine, bu değerler belirli sayıda kanallara ayrılır. Bu kanalların her biri filtre bankasındaki birer bant geçişli filtre olarak düşünülür. Bu şekilde filtrelenmiş değerler o andaki çerçevenin spektrumu olarak alınır. Burada bahsedilen bant geçişli filtrelerin bant genişliklerinin eşit olması zorunlu değildir. Değişken uzunluklu bant genişliklerine sahip filtrelerin kullanılması mümkündür. Ardışık olmayan, yani üst üste bindirilmiş filtre kullanımı da mümkündür.

Filtre seçimi yapılırken amaç sesli ifadenin özelliklerinin en iyi biçimde bir vektör ile ifade edilmesidir. Bunun için günümüzde sık kullanılan ve iyi sonuç verdiği gözlenen *Mel-ölçek* filtre bankası iyi bir çözümdür. (Şekil 14). Bu filtre

bankasında üçgen şeklinde filtreler vardır. 0-1000 Hz arasında 100 Hz uzunluğunda ardışık, 1000-8000 Hz arasında ise logaritmik olarak genişliği artan filtreler içermektedir. Bu durumda 8 kHz sıklığındaki bir ses için toplam 22 tane filtre kullanılmakta ve her bir çerçeve için 22 elemanlı bir özellik vektörü kullanılmaktadır.



Şekil 4. 7. Mel Ölçek Filtre Bankası

Yukarıda belirtilen *mel* ve *bark* ölçeği formülleriyle filtre oluşturmak için kritik bant genişliklerinin belirlenmesi gerekir. Kritik bant genişlikleri aşağıdaki formülle belirlenir.

$$BW = 25 + 75[1 + 1,4(f / 1000)^2]^{0,69} \quad (4.8)$$

Mel ve *Bark* ölçeklerinin her ikisi de frekans aralığının algısal (*perceptual*) olarak anlamlı ve doğrusal olan bir aralığa dönüştürülmesi için kullanılır. Buradaki tekniklerin birleştirilmesiyle elde edilen sinyal çözümü tekniği **kritik bant filtre bankası** (*critical band filter bank*) olarak adlandırılır. Bu filtreler doğrusal fazlı FIR bant geçişli filtrelerdir ve yukarıda belirtilen *mel* ve *bark* ölçeklerinde doğrusal sıralanmışlardır.

Filtre çıkışları daha önce bahsedilen güç hesapları kullanılarak işlenir. Filtre bankası çözümlemesinin ürettiği veriler her bir veri çerçevesi için bir dizi güç değeridir.

4.2.3.2. Fourier Dönüşüm Filtre Bankaları

Spektrum oluşturmanın bir diğer yöntemi de Fourier dönüşümüne dayalı filtre bankasının kullanılmasıdır. Bu yöntem kritik bant genişliğine dayalı filtre bankasına göre gerçekleştirimi daha kolay olan bir yöntemdir. Yapılan işlem sesli ifade sinyali üzerinde bir fourier dönüşümü yapıp, sinyalden belirli değerler için örnekler almaktır. Bir sinyalin kesikli fourier dönüşümü

$$s(f) = \sum_{n=0}^{N_s-1} s(n) e^{-j\left(\frac{2\pi f}{f_s}\right)n} \quad (4.9)$$

Bu formülde f hertz cinsinden frekans, f_s örnek olarak alınan frekans, N_s örneklerin alındığı pencerenin uzunluğudur. Bu denklem kullanılarak sinyalin belirli frekanslardaki spektrumu hesaplanabilir. Bu basit olarak fourier dönüşümünün kullanılmasıdır. Bunun dışında, daha iyi bir spektral yoğunluk elde etmek için her bir spektral değer elde edilirken komşu değerlerin ağırlıklı bir toplamı alınmaktadır. Bu ağırlıklı toplam alınırken kullanılan komşuların belirlenmesi için mel ölçeği kullanılmaktadır.

$$S_{ort}(f) = \frac{1}{N_{os}} \sum_{n=0}^{N_{os}} w_{FB}(n) S(f + \delta f(f, n)) \quad (4.10)$$

Burada N_{os} ortalama değeri elde edilecek örnek sayısını, $w_{FB}(n)$ ağırlık fonksiyonunu ve $\delta f(f, n)$ de f frekansının komşuluğunda ortalaması hesaplanacak frekansları tanımlayan fonksiyonu veya fonksiyonları gösterir.

DFT dışında fourier dönüşümü için FFT (fast fourier transformation) yöntemi de kullanılmaktadır. Bu sadece DFT yönteminin hesap yükü bakımından etkinleştirilmiş bir biçimdir. FFT çok hızlıdır; $N \log N$ toplama ve $N \log N/2$ çarpma işlemi gerektirir. DFT ise N^2 karmaşıklığındadır.

4.2.3.3. Ters-Spektral (Kepstral) Katsayılar

1970 lerin başından itibaren benzeryapılı (homomorfik) sinyal işleme teknikleri konuşma tanıma alanında kullanılmaya başlanmıştır. Benzer yapılı sistemler doğrusal sistemlerin bir sınıfı olarak kabul edilirler. Doğrusal sistemler homomorfik sistemlerin özel bir durumudur.

$$D[[x_1(n)]^\alpha \cdot [x_2(n)]^\beta] = \alpha D[x_1(n)] + \beta D[x_2(n)] \quad (4.11)$$

Yukarıdaki formül içerisinde çarpma, üs alma ve toplama bağlı bir süperpozisyon ilkesidir.

Benzeryapılı sistemler uyarım sinyalinin, ses yolunun şeklinden ayırabilecek bir yol sunmalarından dolayı konuşma işleme için kullanışlı bulunmaktadır.

Ayırma işlemi dekonvolüsyon olarak anılır ve aşağıdaki gibi ifade edilir.

$$s(n) = g(n) \otimes v(n)$$

Burada $s(n)$ sesli ifadeyi, $v(n)$ gırtlığı, yani, sesin izlediği yolu, $g(n)$ ise asıl ses sinyalini, yani ses telleri tarafından üretilen ve değişime uğramamış ses sinyalini temsil eder.

Frekans bölgesindeki (domen) gösterim ise

$$S(f) = G(f).V(f) \quad (4.12)$$

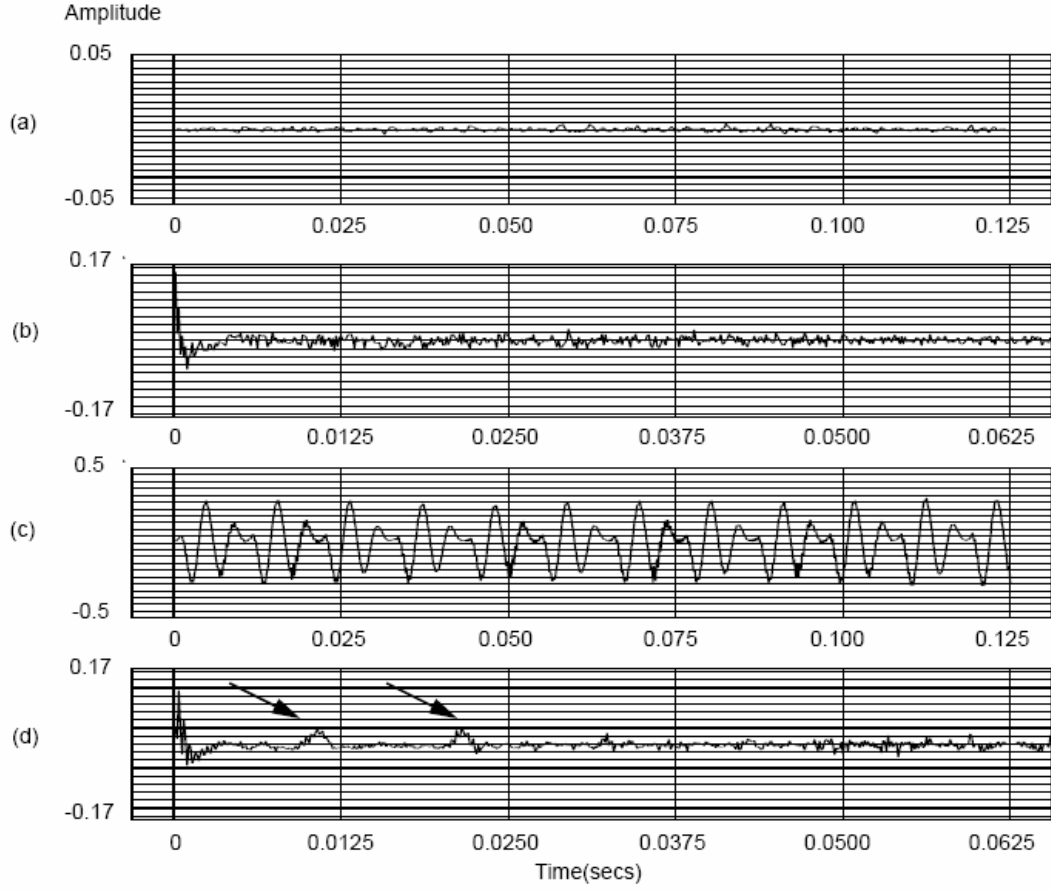
Eğer iki tarafında logaritmasını alacak olursak eşitlik

$$\log(S(f)) = \log(G(f).V(f)) = \log(G(f)) + \log(V(f)) \quad (4.13)$$

Böylece logaritmik düzlemde uyarımı ve ses yolunun şeklini alışılmış sinyal işleme teknikleri ile ortaya çıkarabiliriz. Ters-Spektrum (Cepstrum) kavramı ilk defa 1963 (Bogart ve ark.) yılında ortaya atılmıştır.

$$c(n) = \frac{1}{N_s} \sum_{k=0}^{N_s-n} \log_{10} |S_{ort(k)}| e^{j \frac{2\pi}{N_s} kn}, 0 \leq n \leq N_s - 1 \quad (4.14)$$

Burada $c(n)$, n . Ters-Spektrum olarak adlandırılır. $S(k)$, ters-spektrum değerinin ait olduğu frekans aralığı için alınan fourier dönüşümünü belirtir. N_s o andaki çerçevenin boyunu gösterir. Burada dikkat edileceği gibi $c(0)$ doğrudan o andaki DFT spektrum değerini gösterir. Sesli ifadenin gürültüden ayırt edilmesi için önce spektrumun logaritması alınır ve ters fourier dönüşümü yapılır. Bu şekilde belirlenen ters-spektrum değerleri fourier dönüşümünden türetilmiş ters-spektral katsayılar (fourier transform derived cepstral coefficients) olarak adlandırılır.



Şekil 4. 8. Sesli ve Sessiz Konuşma Sinyalleri ve Cepstrumları

Yukarıdaki şekilde (a) ve (c) sessiz ve sesli konuşma dalgalarını gösterir. Sırasıyla (b) ve (d) de karşılık gelen ters-spektralar vardır.

Burada hesaplanan ters-spektrum değeri sesli ifade tanımada kullanılan önemli bilgileri elde etmede etkilidir. Fourier dönüşümünde kullanılan frekans mel ölçeğine örneklenirse elde edilen ters-spektrum değerleri mel-ter-spektrum değerleri olarak adlandırılır. Genelde ters-spektrum değerlerinin ilk 20 tanesi özellik vektörü olarak kullanılmaktadır. Yani sesli ifade hakkında önemli bilgiler içeren ters-spektrum değerleri ilk 20 tanesidir. ters-spektrum değerleri doğrusal olmayan işlemlerle hesaplandığından dolayı gürültüye duyarlı oldukları söylenebilir. Bu sebeple gürültülü ortamlarda sesli ifade tanıma uygulamalarında tercih edilmezler.

Mel-ter-spektrum değerlerini hesaplamak için kullanılan bir diğer yöntemde fourier dönüşümü yerine mel ölçeğindeki filtre dizisi de kullanılabilir.

4.2.3.4. Doğrusal Ön Kestirim Katsayıları

(Linear Prediction Coefficients)

Sesli ifade çözümü tekniklerinden en güçlüsü doğrusal tahmin (Linear Prediction) yöntemidir. Bu yöntem sesli ifade parametrelerinin tahmininde kullanılan en etkin yöntemlerden biridir. Bu parametreler, temel frekans, biçimlendirici (formant) frekansları, spektrum, gırtlığı modelleyen fonksiyonlar ve sesli ifadenin sıkıştırılmasında kullanılan kodlama teknikleridir. Bu metodun önemi ürettiği doğru sonuçlara ve bu sonuçlara ulaşırken görece olarak daha hızlı çalışmasıdır.

LP çözümlemenin dayandığı temel fikir, bir sesli ifade örneğinin geçmiş sesli ifade örneklerine dayanarak yaklaşık olarak elde edilebileceğidir. Su andaki örnek sesli ifade ile doğrusal olarak tahmin edilen sesli ifade arasındaki farkların karelerinin toplamı en aza indirilmeye çalışılarak, sesli ifadenin tahminini sağlayacak bir dizi birim katsayı bulunabilir. Bu katsayılara tahmin edici katsayılar (Linear Predictor Coefficients) denir ve tahmin edilene sesli ifadenin doğrusal olarak birleştirilmesi sırasında kullanılan ağırlık katsayıları olarak da tanımlanabilirler. LP yöntemi sesli ifadenin doğrusal, zamana bağlı değişen bir sistem olarak modellenmesine dayanır.

Verilen bir $s(n)$ sinyalinde, sinyali daha önceki örneklerin lineer bileşimi olarak modelleyebiliriz.

$$s(n) = \sum_{i=1}^{N_{LP}} a_{LP}(i)s(n-i) + e(n) \quad (4.15)$$

olarak tanımlandığında N_{LP} modeldeki katsayıların sayısını, $\{a_{LP}\}$ ler lineer ön kestirim katsayılarını ve $e(n)$ de modeldeki hatayı (tahmin edilen ile gerçek değer arasındaki fark) temsil eder. Buradaki hata miktarı modelimizin kalitesini ortaya koyar.

Sesli ifade çözümü bağlamında LP ses dalgasının formüle edilmesi olarak düşünülebilir. Temel olarak, all-pole türünden doğrusal filtrelerle sesli ifadenin kaynak filtreleme modeli oluşturulur. Bir sonraki sesli ifade örneğinin doğrusal olarak tahmini geçmiş örneklerin ağırlıklı toplamı aşağıdaki formül ile yapılabilir.

$$\hat{s}_n = \sum_{i=1}^p a_i s_{n-i} \quad (4.16)$$

Kayıpsız bir borunun transfer fonksiyonu all-pole modeli kullanılarak tanımlanabilmektedir. Sesli ifade bağlamında gırtlak, kayıpsız bir boru olarak kabul edilirse, ses telleri tarafından oluşturulan seslerin bu boru içinden geçişi de all-pole modeli ile modellenebilir. Ancak, Gırtlak silindirler biçiminde değildir, Gırtlak kayıpsız değildir, Gırtlak içinde odacıklar bulunan bir yapıya sahiptir. Bazı sessiz harfler dudaklara yakın bir bölgede ve ses tellerinin katkısı olmadan oluşmaktadır. Bütün bu olumsuz yönlerine rağmen yeterince LP katsayısı kullanılarak sesli ifadedeki seslerin yaklaşık bir değerini elde edecek yaklaşımlar mümkündür.

LP parametrelerinin tahmin edilmesi

N değerden oluşan bir sesli ifade örneği verilmiş olsun. Amaç en uygun sonucu üretecek olan a_i katsayılarını tahmin etmek için hesaplamalar yapmaktır. En uygun sonucu elde etme testi için farkların karesini en aza indirme yöntemi kullanılır. Herhangi bir anda asıl sesli ifade ile tahmin edilen arasındaki hata formülüyle hesaplanabilir.

$$e_n = s_n - \hat{s}_n = s_n - \sum_{i=1}^p a_i s_{n-i} \quad (4.17)$$

Bu durumda farkların kareleri toplamı aşağıdaki formül ile hesaplanır.

$$E = \sum_{n=0}^{N-1} e_n^2 = \sum_{n=0}^{N-1} \left(s_n - \sum_{k=1}^p a_k s_{n-k} \right)^2 \quad (4.18)$$

Burada E'nin en küçük değerini aldığı an türevinin sıfır olduğu andır. Yani yukarıdaki formülün ak için türevini alıp sıfıra eşitlersek gerçek sesli ifade ile tahmin edilen arasında en az hata olduğu durumu belirlemiş oluruz. Buna göre;

$$\frac{\partial E}{\partial a_j} = 0 = -\sum_{n=0}^{N-1} \left(2 \left(s_n - \sum_{k=1}^p a_k s_{n-k} \right) s_{n-j} \right) = -2 \sum_{n=0}^{N-1} s_n s_{n-j} + 2 \sum_{n=0}^{N-1} \sum_{k=1}^p a_k a_{k-j} s_{n-j} \quad (4.19)$$

eşitliğinde yeniden bir düzenleme yapılarak (Es. 3.13) formülü elde edilir.

$$\sum_{n=0}^{N-1} s_n s_{n-j} = \sum_{n=0}^{N-1} \sum_{k=1}^p a_k a_{k-j} s_{n-j} \quad (4.20)$$

Bu eşitlik sesli ifade örneği $s_{-p} \dots s_{-1}$ noktaları için LP katsayılarını a_k bulmayı sağlar. Bu eşitliğin çözümü için üç yol vardır.(Robinson, 1995)

- Otokorelasyon(autocorrelation) yöntemi
- Eşdeğişiklik (Kovaryans) yöntemi.
- Kafes(lattice) yöntemi.

Sesli ifade tanımada genellikle otokorelasyon yöntemi kullanılır. Bunun sebebi kullanılan etkin hesaplama yöntemi ve ürettiği daha durağan sonuçlardır. Bu yöntemlerle amaçlanan LP katsayılarının tahmini için yukarıda belirtilen doğrusal denklemin çözümünü yapmaktır. Bu denklemin çözümü sonucunda elde edilen LP katsayıları (a_k) sesli ifade tanımada kullanılabilir.

LPC Ters-Spektrum değerleri, LP katsayılarının(parametrelerinin) doğrudan kullanılmasıyla elde edilir. Bunun için aşağıdaki özyineli yaklaşım kullanılır.

$$c_k = a_k + \frac{1}{k} \sum_{i=1}^{k-1} i c_i a_{k-i} \quad (4.21)$$

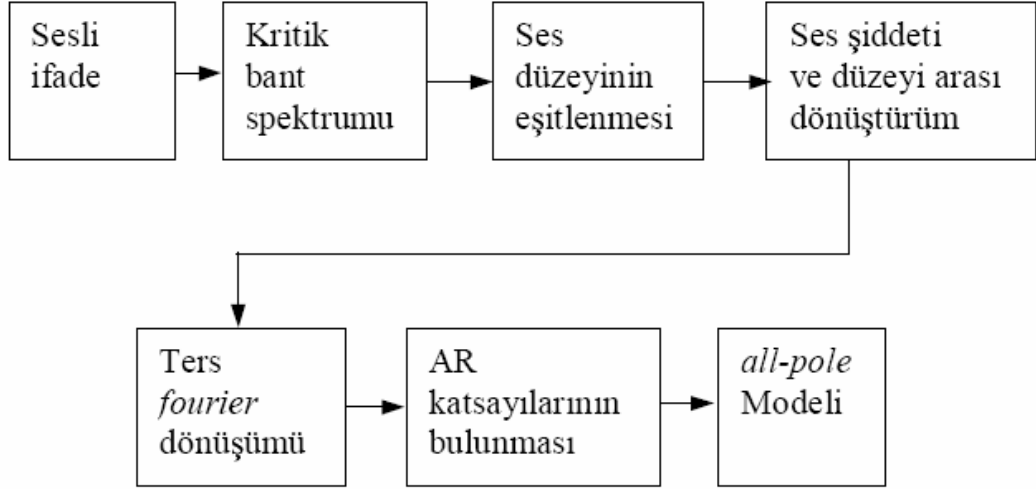
Burada c_k Ters-Spektrum değerlerinden k indisine sahip olanını temsil etmektedir a_{k-1} . ise ilgili LP katsayısını belirtir.

4.2.3.5.PLP (Perceptual Linear Prediction)

PLP yöntemi, DFT (ayrık fourier dönüşümü) ve LP tekniklerinin birleştirilmesi ile sesli ifade parametrelerinin hesaplanmasıdır. Bu yöntem insan kulağının duyma sistemini LP yönteminden daha iyi modellemeye yöneliktir.

LP tekniğinde sesli ifade modellenirken tüm frekanslardaki sesler eşdeğer tutulmaktadır. Bu durum insan kulağıyla uyumlu değildir. 800 Hz değerinden daha düşük frekanslarda duyma miktarı frekansla birlikte düşer. insan kulağı daha çok duyma frekans aralığının ortasındaki frekanslara duyarlıdır. Bu sorunu çözmek için birçok çalışma yapılmıştır. Bu çalışmalardan biri de bulunan LP katsayılarının mel ölçeğine uyarlanması olmuştur. Bir başka yaklaşım da LP tekniği uygulamadan önce

sesli ifadenin güç spektrumunun alınmasıdır.(Hermansky, 1990). PLP yöntemi de bu yaklaşımı kullanmaktadır.



Şekil 4. 9. Perceptual Linear Prediction Yöntemi

Kritik bant spektrumu

Spektrum hesaplanırken ilk olarak, uygun bir çerçeve boyutu belirlenip bu çerçeve üzerinde DFT uygulanmaktadır. Elde edilen sonuçlar bark ölçeğinde filtrelenerek kritik bant çözümlenmesi yapılmaktadır.

Ses düzeyinin eşitlenmesi

Spektrum hesaplandıktan sonra, kulaktaki algılamanın tüm ses düzeyleri için aynı olmadığı gerçeğine dayanarak ses düzeyinin kulağın algılama durumuna göre eşitlenmesi için elde edilen spektrum, es-düzeylilik eğrisi adı verilen bir eğri kullanılarak ön iyileştirme yapılır.

Ses şiddeti-düzeyi arası dönüşüm

Ses şiddeti ve algılanan duyma düzeyi arasındaki doğrusal olmayan bir ilişkinin varlığına dayanılarak bu ilişkinin modellenmesi yapılır. Şiddet ve algılama düzeyi arasındaki bu ilişki $L = I^{0.33}$ formülüyle modellenmiştir. Yani algılanan sesin düzeyi sesin şiddetinin küp köküne eşittir. Bu iyileştirme de önceki aşamada yapılan düzey eşitleme işleminde olduğu gibi, daha sonra yapılacak olan LP işleminde daha az parametrenin yeterli modelleme için yeterli olmasına olanak sağlamaktadır.

Ters *fourier* dönüşümü (*Inverse Fourier Transformation*)

Yukarıdaki iyileştirmelerden sonra ters *fourier* dönüşümü (IDFT) yapılarak üzerinde iyileştirmeler yapılmış sesli ifadenin tekrar elde edilmesi sağlanır.

***Autoregressive(AR)* modelleme**

Burada son aşamada elde edilmiş olan, üzerinde kulağın algılamasına dayalı iyileştirmelerin yapıldığı sesli ifade sinyali için bir LP modeli oluşturulur. Yapılan deneyler oluşturulan bu modelin daha az parametre ile doğrudan oluşturulan LP modelinden daha iyi sonuç verdiğini göstermektedir. (Hermansky, 1990).

4.2.3.6. RASTA (RelAtive SpekTrA)

Özellik vektörü oluşturma da kullanılan RASTA yönteminde, sesli ifade içindeki çevresel etkilerin, yani gürültünün, modellenmesine dayalı bir sesli ifade modelleme yöntemi kullanılır. RASTA ismi, görelî spektrum (RelAtive SpecTrA) ifadesinin kısaltılmasıyla elde edilmiştir. Yukarıda belirtilen PLP yöntemi üzerine gürültü modelleme tekniği eklenerek elde edilen bir yöntemdir.

RASTA yönteminin dayandığı temel, insan kulağının sesli ifadeyi algılamasının daha önceki seslerden önemli derecede etkilendiğidir. Yani sesli ifadenin algılanması daha önce duyulan seslere bağlıdır. Daha değişik bir ifadeyle algılama şu andaki ses ile önceki ses arasındaki spektral farka bağlıdır. Bu durumda insan kulağı yavaş değişen seslere daha az duyarlıdır denebilir.

Yapılan sesli ifade çözümlemesinin yavaş değişen seslere daha az duyarlı yapılması insan kulağının bu özelliğinin de modellenmesini sağlar. Bunu yapmak için daha önce belirtilen PLP yönteminde kullanılan filtreleme yönteminde değişiklikler yapılmıştır. Kullanılan filtreler spektral sıfır değeri keskinleştirilmiş, yani sıfır frekans düzeyine aniden inen filtrelerle değiştirilmiştir. Böylece frekanslardaki yavaş değişimlerin etkisi azaltılmıştır.

Buraya kadar konuşma tanıma sistemlerinde kullanılan sinyal ölçüm yöntemlerini anlattık. Bundan sonrada elde ettiğimiz bu parametrelerinin nasıl yumuşatılıp birleştirilebileceğini ve sinyal parametrelerine dönüştürüleceğini inceleyeceğiz.

4.3. Parametre Dönüşümü

Sinyal parametreleri iki temel işlemle sinyal ölçümlerinden elde edilir. Fark Alma ve birleştirme. Bu işlemlerin ardından elimize sinyalin ham hesaplarından elde edilen parametre vektörleri geçer.

4.3.1. Fark Alma (Differentiation)

Fark alma işlemi (delta), özellik vektörlerinin herhangi bir biçimde farklarının alınmasıdır. Fark vektörleri sesli ifadenin kısa süreli değişimlerinin açığa çıkmasını sağlayan özellik vektörleridir. Fark alma işlemi sonucu elde edilen vektörler zaman boyutunda özellik vektörlerinin benzerliğini açığa çıkarır.

$$\dot{s}(n) \equiv \frac{d}{dt}s(n) \approx s(n) - s(n-1) , \quad (4.22)$$

$$\dot{s}(n) \equiv \frac{d}{dt}s(n) \approx s(n+1) - s(n) , \quad (4.23)$$

$$\dot{s}(n) \equiv \frac{d}{dt}s(n) \approx \sum_{m=-N_d}^{N_d} m s(n+m) . \quad (4.24)$$

Eş.4.22 geri fark almayı , Eş.4.23 ise ileri fark işlemini gösterir. Eş.4.24 de regresyon analizi olarak tanımlayabileceğimiz ideal bir fark alma denklemini ifade eder.

Bu diferansiyel denklemin sonucunda bir delta parametresi elde edilir. Aynı şekilde ikinci seviyeden bir türev aldığımızda ise delta-delta parametresini elde ederiz.

4.3.2. Birleştirme (concatenation)

Birleştirme (concatenation) işlemi birden fazla özellik vektörünün yan yana eklenmesiyle gerçekleştirilir. Birleştirilen özellik vektörleri farklı yöntemlerle de

elde edilmiş olabilir. Ayrıca sadece bir çerçeve üzerinden hesaplanan özellik vektörleri birleştirilebileceği gibi, farklı zamanlarda hesaplanan özellik vektörlerini de birleştirmek mümkündür. Örneğin enerji değerini bir ters-spektrum vektörünün başına eklemek veya ardı ardına iki çerçeve için hesaplanmış iki ters-spektrum vektörünü yan yana eklemek mümkündür.

Herhangi bir sinyal için elde ettiğimiz sinyal ölçüm matrisini aşağıdaki gibi ifade edelim

$$\underline{X} = \begin{bmatrix} x(0, 0) & x(0, 1) & \dots & x(0, N_x - 1) \\ x(1, 0) & x(1, 1) & \dots & x(1, N_x - 1) \\ \dots & \dots & \dots & \dots \\ x(N_f - 1, 0) & x(N_f - 1, 1) & \dots & x(N_f - 1, N_x - 1) \end{bmatrix}$$

Burada $x(n,m)$ n. Çerçeve'deki m. sinyal ölçümüne, N_f sinyaldeki toplam çerçeve sayısını ve N_x te her çerçeve için yapılan ölçüm sayısına karşılık gelir.

Yukarıdaki X matrisi sinyalin tüm zamandaki ölçüm değerlerini tutar. Pratik sistemlerde sinyaller gerçek zamanda çerçeve çerçeve ölçülür. Tüm değerlerin bir matriste toplanması büyük bir gecikmeye sebebiyet verir. Ancak matris şeklindeki bu gösterim sinyal modelinin bütün olarak görülmesini sağlar. Sinyal Ölçüm matrisinde genellikle güç ve ters-spektral katsayılar kümesinin harmanlanmış ölçüleri kullanılır.

Daha sonra parametre yumuşatma işlemine yardımcı iki matris tanımlayacağız. Birincisi gecikme matrisi, $\underline{\tau}$ ile gösterilen bu matris şimdiki zamandan olan gecikmeyi gösterecek :

$$\underline{\tau} = \begin{bmatrix} \tau(0, 0) & \tau(0, 1) & \dots & \tau(0, N_{\tau_0} - 1) \\ \tau(1, 0) & \tau(1, 1) & \dots & \tau(1, N_{\tau_1} - 1) \\ \dots & \dots & \dots & \dots \\ \tau(N_p - 1, 0) & \tau(N_p - 1, 1) & \dots & \tau(N_p - 1, N_{\tau_p} - 1) \end{bmatrix}$$

$$= \left[\underline{\tau}_0 \ \underline{\tau}_1 \ \dots \ \underline{\tau}_{N_p-1} \right]^T .$$

burada $\underline{\tau}_i$ i. Gecikme vektörünü ve N_p de toplam sinyal parametresini gösterir.

Gecikme vektörü her sinyal parametresi parçası için kaç tane ölçüm yapıldığına bağlı olarak değişik boyutlarda olabilir. Aslında $\underline{\tau}$ vektörlerin bir vektörüdür. Bu formüldeki her bir vektörün uzunluğu N_{τ_i} ile gösterilir.

Daha sonra ölçülere uygulanacak filtrelerin ağırlıklarını tutan bir ağırlık matrisi \underline{W} tanımlayacağız. Bu ağırlık matrisi gecikme matrisi ile birebir eşlenen bir matristir.

$$\begin{aligned} \underline{W} &= \begin{bmatrix} w(0, 0) & \dots & w(0, N_{\tau_0} - 1) \\ \dots & \dots & \dots \\ w(N_p - 1, 0) & \dots & w(N_p - 1, N_{\tau_p} - 1) \end{bmatrix} \\ &= \left[\overline{w}_0 \ \overline{w}_1 \ \dots \ \overline{w}_{N_p-1} \right]^\dagger . \end{aligned}$$

\overline{w}_i , $\underline{\tau}$ içerisinde karşılık gelen vektörüyle aynı boyutta i. katsayı vektörüdür.

Aynı zamanda X matrisindeki sütunlara karşılık gelen W deki her bir satır için bir indis vektörü I tanımlanır.

$$I = \left[I_0 \ I_1 \ \dots \ I_{N_x-1} \right]$$

Ölçüm vektörünü filtrelemek için aşağıdaki konvolüsyon benzeri işlem tanımlanır

$$V = X \star W$$

yıldız işlemi aşağıdaki algoritma ile hesaplanır.

```
For 0 ≤ n ≤ Nf - 1 {  
  For 0 ≤ i ≤ Nx - 1 {  
    
$$V[n, i] = \sum_{j=0}^{N_f-1} W(i, j) X[n + \tau(i, j), I_i]$$
  
  }  
}
```

Bu işleme birleştirme işlemi denir. Her bir çerçeve için tüm sinyal parametrelerinin dahil edildiği tek bir parametre vektörü elde etmemizi sağlar.

4.4. Normalizasyon/ Uyarlama Teknikleri

Otomatik konuşmacı tanıma sistemlerinin performansını etkileyen en önemli faktör her denemede (oturumlar arası veya zamanda) değişen sinyal karakteristiğidir. Değişimler kullanıcının kendisinden kaynaklanabileceği gibi kayıt ve iletim ortamlarından yada gürültüden kaynaklanabilir. Konuşmacılar her deneme de kesinlikle aynı telaffuzu tekrar edemezler. Araştırmalar göstermiştir ki aynı oturumda kaydedilen konuşmalar, farklı zamanlarda ve oturumlarda yapılanlardan daha tutarlı olmaktadır.

Konuşmacı tanıma sistemleri için bu değişimleri uyarlamak önemlidir. Uzun bir periyotta yüksek bir tanıma oranı elde etmek için her konuşmacıya ait referans modelin ve doğrulama eşliğinin uyarlanması çok önemlidir. Bu Değişimler arasındaki farkı gidermek için üç çeşit normalleştirme ve uyarlama tekniği kullanılır. Bu teknikler, parametre bölgesinde (sinyaller ve öznitelikler) , uzaklık/benzerlik bölgesinde ve model bölgesindedir.

4.4.1. Parametre Domeni Normalizasyonu

Parametre bölgesindeki en tipik normalleştirme tekniği, spektral eşitlemedir. Bu eşitleme aynı zamanda kör eşitleme, kör dekonvalisyon veya ters-spektral ortalama normalleştirmesidir. Bu yöntemle lineer kanal efektlerinde ve uzun dönem spektral değişimlerde azalma sağlanır. Bu metot özellikle uzun telaffuzları kullanan metin-bağımlı konuşmacı tanıma uygulamalarında etkilidir. Bu metotta tüm telaffuz boyunca spektral katsayıların ortalanması sağlanır ve bu ortalama değerler her çerçevenin ters-spektral katsayılarından çıkarılır. Log spektral bölgesindeki ek değişimler bu yöntem sayesinde telafi edilir. Ancak kaçınılmaz olarak metin bağımlı ve konuşmacıya özel özniteliklerden bazıları kaybolur. Bu yüzden konuşmacı tanıma uygulamalarındaki kısa telaffuzlarda kullanımı uygun değildir.

Gish farklı telefon hatlarından iletilen konuşmalar üzerinde uygulanacak basit ön filtreleme teknikleriyle metin bağımsız-konuşmacı uygulamalarının performanslarının önemli derecede arttığını göstermiştir. Gish ve arkadaşları çok değişkenli Gauss olasılıklı yoğunluk fonksiyonları kullanılarak iletim kanallarının istatistiksel olarak modellenebileceğini göstermişlerdir. Tabi bu yöntemin kullanılabilmesi , kanallara ait yeterli sayıda eğitim örneğinin elde edilmesiyle mümkündür. Ters-spektral katsayıların zamandaki türevlerinin (delta-cepstral

coefficient) eğitim ve test aşaması arasındaki lineer kanal hatalarına karşı dayanıklı olduğunu göstermiştir.

4.4.2. Uzaklık/Benzerlik Domenindeki Normalizasyon

Higgins ve arkadaşları uzaklık(benzerlik) oranı kullanarak uzaklık değerleri için bir normalleştirme yöntemi geliştirmişlerdir. Benzerlik oranı sisteme kendisini tanıtan bir kişinin gerçek olması ile sahtekar olmasına dayalı gözlemlere göre elde edilen bir orandır. Olasılık oranının matematiksel ifadesi

$$\log L(X) = \log p(X | S = S_c) - \log p(X | S \neq S_c) \quad (4.25)$$

Burada S ve S_c Giriş konuşmacısını ve doğru kişi olmasını gösterir. Genellikle $\log L$ nin pozitif değerde olması geçerli bir talebi, negatif olması ise bir sahtekarı gösterir. Eşitliğin sağ tarafını normalleştirme terimi olarak ifade edeceğiz.

Eğer biz referans konuşmacıların tüm konuşmacıları temsil etini varsayarsak, gerçek konuşmacı S haricindeki tüm konuşmacılar için X noktasındaki yoğunluk en yakın kullanıcıya referans eden yoğunluk tarafından baskılanır. Böylece aşağıdaki karar kriterini elde ederiz.

$$\log L(X) = \log p(X | S = S_c) - \max_{S \in Ref, S \neq S_c} \log p(X | S) \quad (4.26)$$

Bu karar kriteri iki sebepten dolayı gerçekçi değildir. Birincisi en yakın referans konuşmacıyı seçmek için, tüm konuşmacılara ait koşullu olasılıkların tümü hesaplanmalıdır ki bu da büyük hesaplama yükü getirir. İkincisi maksimum koşullu olasılıklar referans kümesinde en yakın konuşmacının ne kadar yakın olduğunu bağlı olarak konuşmacıdan konuşmacıya değişir.

Eş.(4.25) de gösterilen normalleştirme eşitliğini hesaplamak için yandaş konuşmacılar denilen bir konuşmacı kümesi seçilir. Higgins ve arkadaşları iddia eden konuşmacının yakınındaki popülasyonu temsil eden kullanıcıların seçilmesini önermişlerdir.

$$\log L(X) = \log p(X | S = S_c) - \log \sum_{S \in Yandaş, S \neq S_c} p(X | S) \quad (4.27)$$

DeneySEL sonuçlar göstermiştir ki bu normalleştirme metodu konuşmacıların ayrıştırılabilirliğini artırır ve sadece talep eden konuşmacıya ait modelin kullanıldığı puanlamaya göre konuşmacı bağımlı veya metin bağımlı eşiklemeye olan ihtiyacı

azaltır. Bir başka deneysel çalışmaya göre sayısı birden beşe kadar değişen yandaş konuşmacı seçmek doğrulama performansını artırır.

4.4.3. Model Bölgesi Adaptasyonu

Bir sonraki konuda anlatılacak olan SMM (Saklı Markov Modeli) hem metin-bağımlı hem de metin bağımsız sistemlerde sıkça kullanılmaktadır. Ancak SMM arka plandaki gürültüye karşı hassastır. Öyle ki gürültülü seslerde tanıma oranında büyük düşüşler gözlenmiştir. Bu problemin üstesinden gelmek için paralel model kombinasyonu tekniği kullanılır (PMC).

Bu metotta konuşmayla gürültünün birleştirildiği SMM kullanılır. Gürültülü konuşmaya ait SMM oluşumunda, gözlenen olasılıklar (ortalamalar ve katsayılar), lineer ters-spektral bölgede gözlenen olasılıklarla birleştirilerek hesaplanır. İçinde ses olmayan sadece gürültü bulunan bir SMM 'i eğitildikten sonra temel bir gürültü modeli olarak konuşma SMM lerine uyarlanır.

Bu yöntem yalnızca durağan gürültülere değil, aynı zamanda konuşma içerisindeki başka insanların sesinde olduğu gibi zamanla değişen gürültülere de uygulanabilir. Bu metodun etkinliği gürültü eklenmiş konuşmalara ait deneysel çalışmalarda gözlenmiştir. Sözü edilen yöntem son yıllardaki çalışmalarla diğer başka türlü gürültülerle ve çarpımsal bozulmalarla baş edecek şekilde geliştirilmektedir.

4.5. Konuşmacı Modelleme

Konuşmacı Tanıma Sistemlerinin ilk aşaması bilinen tüm konuşmacılara ait modellerin oluşturulması ve eğitilmesidir. Bu eğitim aşamasında evren modeli ve yardımcı modellerde oluşturulur.

Konuşmacı Modelleme iki grup halinde sınıflandırılabilir.

- Şablon Modeller
- Olasılıklı (İstatistiksel) Modeller

Şablon Modeller her konuşmacı için şablon öznitelik vektörü oluşturulması esasına dayanır ve metin bağımlı sistemlerde kullanılır. Şablonlar birkaç oturumda elde edilmiş şifre kelimelere ait öznitelik vektörlerinin normalize edilmesiyle elde edilir.

DTW (Dynamic Time Warping) tekniđi řablon modeldeki ölçülerin benzerliđini kullanan bir tanıma çeşididir.

İstatistiksel Modellerde, olasılıklı süreçler altında hesaplanan parametreler tarafından modellenmiş parametrik rastgele süreçler altında toplanır. Olasılıklı süreçler konuşmacı tanımda daha başarılı sonuçlar vermektedir. Eğer tanıma işlemi metin bağımlı ise, yani konuşmacının sesindeki metin içeriđi sistemde bulunuyorsa HMM modelleme kullanılır. Eğer sistem metin bağımlı değil ise konuşmacı modellerini oluşturmak için GMM (Gaussian Mixture Model) adı verilen tek durumlu HMM'ler kullanılır.

4.5.1. Dinamik Time Warping Temelli Konuşmacı Modelleme

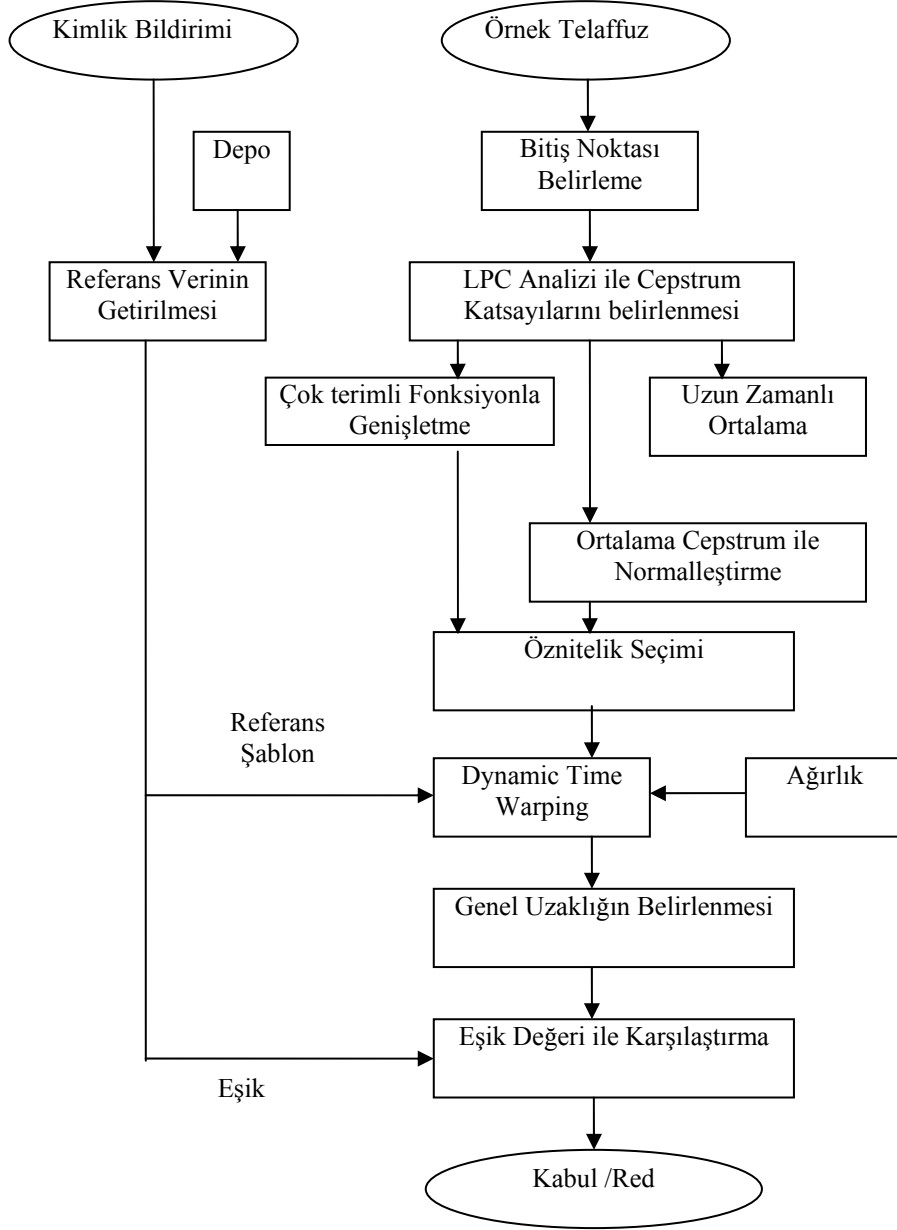
Dinamik Time Warping řablon eşleştirme problemlerinde kullanılır. DTW, Dinamik Programlama denilen ve "bir optimizasyon probleminde alt problemlere ait çözümlerin saklanması ve sonradan tekrar hesaplanmamasını sađlayan bir algoritma" olarak tanımlanan yöntemi temel alır. İçerisinde bir geri izleme adımı barındıran bu yöntem en iyi çözümlerin bulunmasını sađlar.

Bu yaklaşımda her telaffuz bir öznitelik vektörü olarak belirlenir. Genellikle kısa dönemli spektral öznitelik vektörleri ve farklı denemelerle alınan aynı telaffuzdaki deđişimlerin normalleştirilmesi sonucunda elde edilen öznitelik vektörleri DTW yardımıyla karşılaştırılır.

DTW referans řablonla, test řablonları arasındaki minimum uzaklıđı bulmaya çalışır. Minimum uzaklık ařađdaki formülle hesaplanır.

$$D(i, j) = d(i, j) + \min \left\{ \begin{array}{l} D(i, j) \\ D(i, j - 1) \\ D(i - 1, j) \end{array} \right\}, \quad i = 1, n; \quad j = 1, N. \quad (4.28)$$

burada i ve j karşılaştırılacak özellik vektörlerini d(i,j) öklid uzaklıđını D(,) de bulunan noktadan olan genel uzaklıđı temsil eder. Bařlangıç noktasında D(0,0)=d(0,0) dır.



Şekil 4. 10. DTW Temelli bir sistemin genel yapısı

Şekil tipik bir DTW tabanlı sistemin yapısını gösterir. İlk konuşmadaki 10 ms lik kısa bölümlere ait 10 LPC ters-spektral katsayı çıkarılır. Önceki bölümlerde bahsedilen spektral eşitleme tekniği, iletim bozulmalarını ve konuşmacılar arası farklılığı telafi etmek için her ters-spektral katsayıya uygulanır. Daha sonra normalleştirilmiş ters-spektral katsayılar her 10 ms lik dilimlerde çıkarılır. Parametre kümesinin zaman fonksiyonu referans şablonla aralarındaki uzaklığı belirlemek için zamanda karşılaştırma yapmada kullanılır. Daha sonra tüm uzaklıklar doğrulama kararını vermek için bir eşikle karşılaştırılır.

4.5.2. Gauss Karışım Modelleri (Gaussian Mixture Model)

Gauss Karışım modelleme(GMM) istatistiksel desen eşleştirme problemlerinde kullanılan modellerin oluşturulmasında kullanılan istatistiksel bir yöntemdir. Konuşma da parametrik rastgele işlemler olarak tanımlanabileceğinden, konuşmacı tanıma da, istatistiksel desen karşılaştırma tekniklerinin kullanılması da uygundur. Öznitelik çıkarım aşamasından sonra elde edilen öznitelik vektörleri konuşma sinyalindeki metin içeriği ile ilgili bilgilerin yanı sıra, ses yolunun şekli ve boğumlama gibi konuşmacıyla ilişkili bilgileri de içerir.

Elde edilen öznitelik vektörlerinin çok boyutlu Gauss olasılık yoğunluk fonksiyonuna (pdf) göre elde edildikleri varsayılır.

D boyutlu öznitelik vektörü x için gauss pdf'i aşağıdaki gibi tanımlanır.

$$b_i(x) = \frac{1}{\sqrt{(2\pi)^D |\Sigma_i|}} e^{-\frac{1}{2}(x-\mu_i)^T(\Sigma_i)^{-1}(x-\mu_i)} \quad (4.29)$$

Burada μ_i ve Σ_i Gauss karışımının i . bileşenin ortalaması ve kovaryans matrisidir. $()^T$ matris transpozunu gösterir.

Eşitlikteki $b_i(x)$ i . Gauss bileşeninin olasılık yoğunluğudur.

Verilen Konuşmacı modeli $S_j, P(x|S_j)$ ye ait öznitelik vektörü x 'in benzerlik ihtimali M adet Gauss yoğunluğunun ağırlıklı toplamı şeklinde bulunabilir.

$$P\langle x|S_j \rangle = \sum_{i=1}^M p_i b_i(x) \quad (4.30)$$

Burada p_i i . bileşenin karışım ağırlığıdır.

Gauss Karışım modelinin parametreleri aşağıdaki gibi tanımlanır.

$$S_j = \{p_i, \mu_i, \Sigma_i\}, i = 1, \dots, M. \quad (4.31)$$

ve karışım ağırlıklarının toplamı aşağıdaki kriteri sağlamalıdır.

$$\sum_{i=1}^M p_i = 1 \quad (4.32)$$

Eğitim esnasında Model Parametreleri Maximum Likelihood (ML) hesaplamasından esinlenerek ortaya konulan Expectation Maximization (EM) algoritmasından elde edilir. Eğitim safhasının sonunda Konuşmacı tanıma amacıyla Maximum Likelihood sınıflandırıcılar tarafından kullanılacak olan S_j konuşmacı modellerini elde ederiz.

GMM teknikleri kullanan konuşmacı tanıma sistemleri gerçek yaşamda kullanılabilecek kadar iyi sonuç vermektedirler. Gauss karışım modellerinin performansı başlangıç durumuna bağlı değildir, eğitim verisinden konuşmacıya özel karakteristikleri öğrenebilirler. Ayrıca bozuk telaffuzların bulunduğu durumlarda bile başarılı sonuçlar vermektedir.

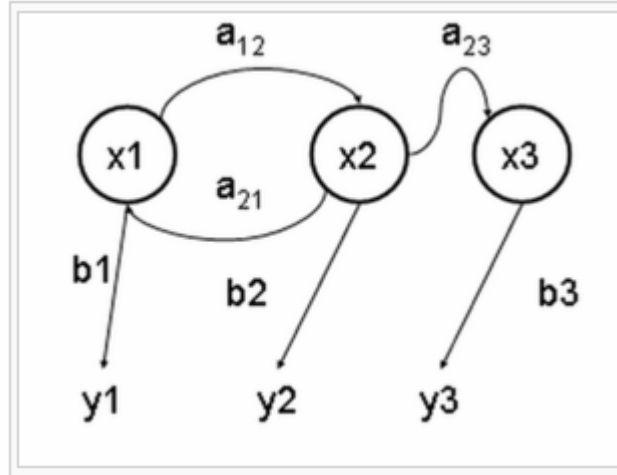
GMM ler ile ilgili asıl kısıtlama eşleşmeyen durumlarda kullanıldığında performanstaki azalmadır. Bu durumda daha önce bahsedilen adaptasyon teknikleri kullanılarak konuşmacı modellerinin sağlamlaştırılması sağlanabilir.

GMM bazlı konuşmacı tanıma genellikle metin bağımsızdır. Metin bağımlı sistemlerde kullanılması isteniyorsa HMM gibi başka yöntemlerle birleştirilebilir.

4.5.3. Saklı Markov Model (HMM) Tabanlı Konuşmacı Modelleri

Saklı Markov Modelleme (HMM) tekniği, sistem metin-bağımlı ise kullanılır. Her konuşmacının söylediği şifre kelimeler veya HMM'ler tarafından modellenir.

HMM'de konuşma ayrık durağan işlemler olarak farzedilir ve her telaffuz ayrık durağan durumlar ve onların arasındaki anlık geçiş olarak modellenir. Tüm cümlenin modellenebilmesi için bu durumlar bir zincir olarak elde edilecek şekilde organize edilir.



Şekil 4. 11. Markov Modeli ve geçiş örneği

şekilde bir markov modeli görülmektedir. Markov modelinde

- $x \rightarrow$ Gizli durumlar
- $y \rightarrow$ Gözlenebilir çıkışlar
- $a \rightarrow$ Geçiş olasılıkları
- $b \rightarrow$ Çıkış olasılıklarını ifade eder

Her durum bir diğerine geçişlerle bağlıdır. Geçişler, geçiş olasılıklarına (a_{ij}) bağlı olarak durum değiştirmeye imkan verir. Durumlara iliştilen emisyon olasılıkları (b_j) bir özellik vektörünün, referansın belirli bir zaman aralığıyla olan spektral benzerliğini gösterir. Sistem girdisine göre oluşturulan özellik vektörleri dizisine bağlı olarak, model üzerinde birinci durumdan başlayan farklı yollar izlenebilir. Bazı durumların tekrarı veya atlanması kullanıcının konuşma hızındaki değişimlere sistemin adaptasyonunu sağlar. Bir kelimenin tanınabilmesi için referans olarak alınan durumdan itibaren izlenen yolun en son duruma, kabul edilebilir bir olasılıkla ulaşması gereklidir (Yapanel, 1997).

HMM'ler ile çözülebilecek üç temel problem vardır.

- Model parametreleri verildiği zaman çıkış sırasının olasılığını hesaplamak (forward algorithm)
- Model parametreleri verildiğinde istenen bir çıkışı sağlayacak olan gizli durumların sırasının hesaplanması (Viterbi algorithm)

- Verilen bir çıkış sırasına göre durum geçişlerinin ve çıkış olasılıklarının hesaplanması (Baum-Welch algorithm)

HMM'ye dayalı konuşmacı tanıma sistemi işlem basamaklarının değişik adımlarında bu üç algoritmayı kullanarak bir sonuç üretir. Modelleme aşamasında forward-backward algoritmaları kullanılarak konuşmacıya ait model parametreleri hesaplanır. Karar verme aşamasında ise Viterbi algoritmasından yararlanılarak oluşturulan model ile var olan model arasındaki benzerlik ölçülür.

4.5.4. Uzun dönemli istatistik tabanlı yöntemler.

Telaffuz serilerine ait spektral özniteliklerin ortalaması ve varyansı gibi uzun dönemli örnek istatistikleri metin bağımsız öznitelik olarak kullanılmaktadır. Ancak kanal etkilerine karşı aşırı hassastırlar

4.5.5. Vektör Nicemleme (Vector Quantization)

Konuşmacıya ait kısa dönemli eğitim öznitelik vektörleri, konuşmacının esas karakteristiğini temsil etmede kullanılabilir. Ancak eğitim vektörleri uzun olduğu zaman bu temsil çok fazla hafıza ve hesaplama karmaşıklığına yol açacağı için pratik kullanıma uygun olmayabilir. Bu yüzden eğitim verisini sıkıştırarak etkili bir şekilde kullanım amacıyla VQ teknikleri kullanılır.

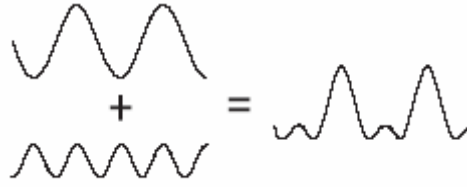
Bu metotta konuşmacıya özel özniteliklerin etkin kısımlarını temsil eden az sayıda öznitelik temsilcilerinin bulunduğu VQ kod kitabı kullanılır. Konuşmacıya özel kod kitabı her konuşmacıya ait eğitim özniteliklerinin kümelenmesiyle elde edilir. Tanıma aşamasında giriş telaffuzu, tüm referans konuşmacıların bulunduğu kod kitabı kullanılarak vektör niceleme işlemine tabi tutulur. ve tüm telaffuz üzerindeki VQ bozulması kullanılarak karar verme işlemi yapılır.

5. DALGACIK DÖNÜŞÜMÜ (WAVELET TRANSFORM)

5.1. Frekans Analizi

5.1.1. Frekans Bölgesi

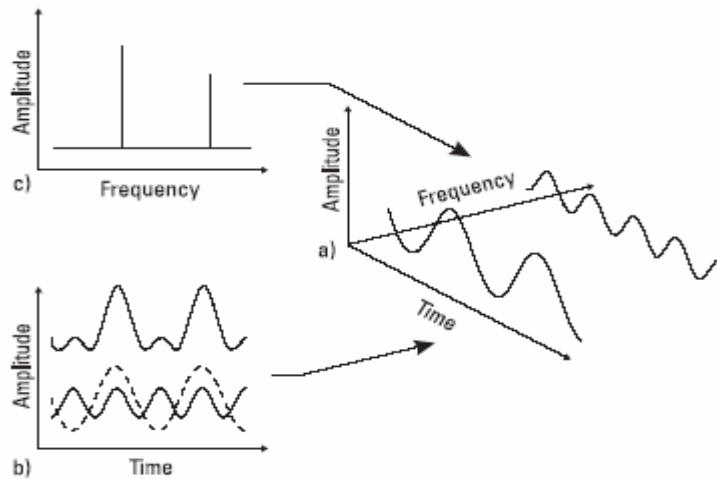
100 yıldan uzun zaman önce Jean Baptiste Fourier her hangi bir sinyalin, sinüs dalgalarının birbirine eklenmesi ile elde edilebileceğini göstermişti. Aşağıdaki şekil bir dalga formunun iki sinüs dalgasının birleşimi şeklinde gösterimini göstermektedir.



Şekil 5. 1. İki Frekansın Birleşimi

Elimizdeki bir sinyal, sinüs dalgalarının genlik, frekans ve faz değerlerinin doğru şekilde seçilerek birleştirilmesiyle elde edilebilir.

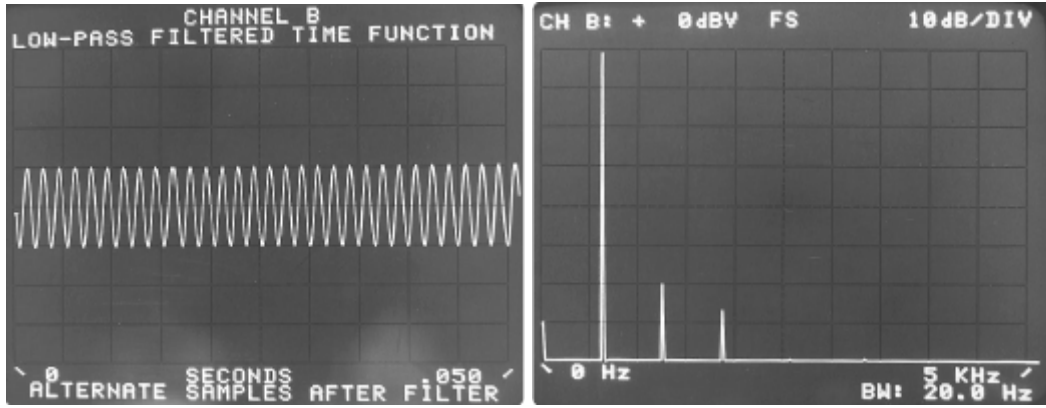
Yukarıda yaptığımız işlemin tersine herhangi bir sinyali birden fazla sinüs dalgasına ayırabiliriz. Gerçek hayatta karşılaştığımız bir sinyal sadece belirli sinüs dalgalarının tek bir birleşimi sonucu elde edilir.



Şekil 5. 2. Zaman-Frekans-Genlik Düzlemi

Yukarıda şekil x te gösterilen işaretin bileşenleri olan sinüs dalgalarının üç boyutlu gösterimi verilmiştir. Buradaki iki boyut bildiğimiz zaman-genlik bölgesi üçüncü boyut ise frekans bölgesidir. Şekil 11 de frekans ve genlik eksenleri görülmektedir. Bu gösterime frekans bölgesi. Girişten elde ettiğimiz her sinüs dalgası burada dik çizgilerle gösterilmektedir. Çizginin yüksekliği genliği, pozisyonu ise frekansını göstermektedir. Bu şekilde giriş sinyalimizi, frekans bölgesinde tekil olarak karakterize etmiş oluruz. Sinyalimizin frekans bölgesindeki gösterimine *sinyal spektrumu*, spektrumdaki her bir çizgi ise sinyalimizin *bileşeni* denir.

Peki, niçin frekans analizine ihtiyaç duyarız. Buna basit örnekler vererek açıklayalım. Gürültülü bir iş makinesindeki seslerden mukavemet hatalarını bulmak istersek geniş sinyallerin arasındaki küçük sinyalleri ayırtmamız gerekir. Bu şekilde normalde fark edilemeyecek bileşenler ayırarak hata önleme rutinleri gerçekleştirilebilir.

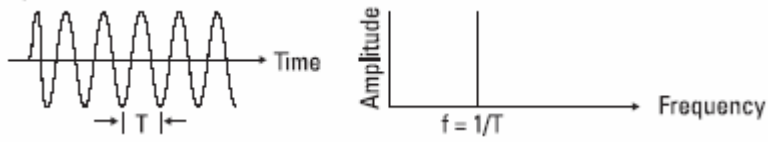


Şekil 5. 3. Sürekli Bir Sinyal ve Frekans Bileşenleri

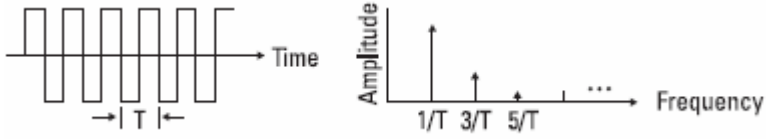
İlk bakışta frekans analizi garip ve bilinmedik gelebilir. Ancak frekans analizi, günlük hayatın önemli bir parçasıdır. Kulağımız ve beynimiz mükemmel bir frekans analizörüne örnektir. Beynimiz kulağa gelen ses spektrumunu çok küçük dar bantlara ayırır ve her banttaki gücü belirler. Bu sayede gürültülü arkaplan seslerinden küçük sesleri ayırtabilir. Ek bir örnek olarak Bir doktor kalbimizi ve solunumumuzu dinleyerek alışılmadık sesleri bulmaya çalışır.

Frekans bölgesinin hayatımıza o kadar da uzak olmadığını gördük. Görmediğimiz ve alışık olmadığımız frekans bölgesinin grafiksel gösterimidir. Bazı bilindik sinyallerin frekans spektrumundaki gösterimleri bize bu konuda yardımcı olacaktır.

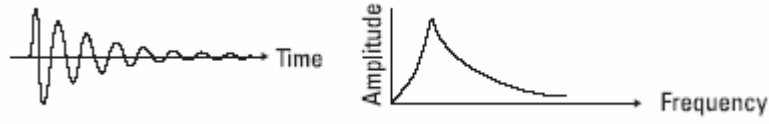
5.1.2. Spektrum Örnekleri



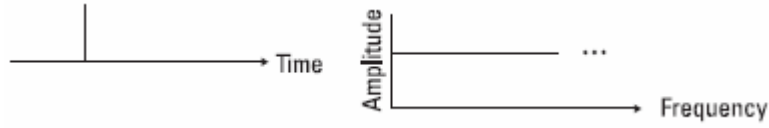
Sinüs Dalgası



Kare Dalga



Sönümlü Dalga



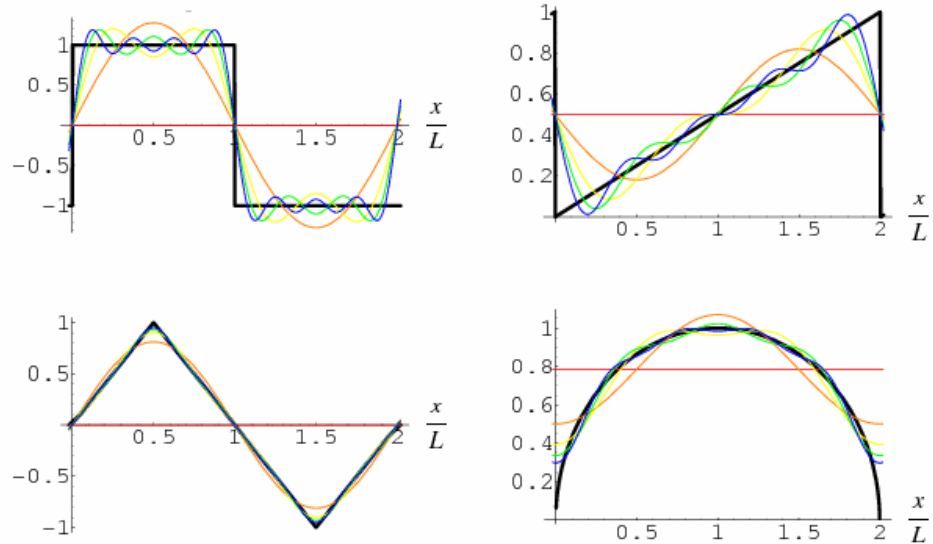
Dürtü (impulse)

5.2. Fourier Dönüşümü

Fourier Analizi, sinyallerin, sinüsoidal bileşenlere ayrıştırılması temeline dayanan matematiksel teknik ailesidir. Fourier Analizi ismini Fransız matematikçi ve fizikçi Jean Baptise Joseph Fourier (1768-1830) den almıştır. Fourier ısının yayılımıyla ilgilenmiş ve 1807 yılında sıcaklığın dağılımında sinüsoidlerin kullanımını gösteren bir yayın çıkarmıştır. Bu yayında, herhangi bir sürekli sinyalin düzgün seçilmiş sinüsoidlerin toplamı biçiminde gösterilebileceğini iddia etmiştir. Bu iddia ortalığı çok karıştırmıştır. Bir yanda Fouriere destek veren Laplace bir yanda da Fourier'in tezine şiddetle karşı çıkan Lagrange, Laplace ve diğer bilim adamlarının yayını onaylamasına rağmen, yayını veto etmiştir. Lagrange böyle bir yaklaşımın devamsızlık bölgeleri olan veya kare dalga gibi köşeleri olan sinyalleri temsil edemeyeceği savunmuştur. Bu yayın yaklaşık 15 sene sonra Lagrange öldükten sonra ancak kabul edilip yayınlanabilmiştir.

Aslında Lagrange görüşünde haklı idi. Sinüsoidlerin toplamı asla köşeli bir sinyali ifade edemezdi. Ancak çok yaklaşabilirdiniz. O kadar yakın ki iki nokta arasındaki enerji sıfır olmaktaydı.

Aşağıdaki şekilde dört sinyalin yaklaşık sinüsoidal bileşenleri gösterilmektedir.



Şekil 5. 4. Bazı Sinyallerin Sinüsoidal Bileşenleri

Bu konu matematikte fourier serileri altında incelenmektedir. Ve Fourier serileri tanım itibariyle periyodik olan bir $f(x)$ fonksiyonun sonsuz sayıdaki sinüs ve kosinüslerin toplamının açılımıdır. Fourier serileri sinüs ve kosinüs fonksiyonlarının ortogonal ilişkilerini kullanır. İleride de dalgacık konusunda bahsi geçeceği için Ortogonallikten biraz bahsetmekte fayda var.

$f(x)$, $g(x)$ iki fonksiyon, $w(x)$ ağırlık fonksiyonu ve $a \leq x \leq b$ olmak üzere

$$\langle f(x) | g(x) \rangle \equiv \int_a^b f(x)g(x)w(x)d(x) = 0 \quad (5.1)$$

ise bu iki fonksiyon birbirine diktir (ortogonal) denir.

Ek olarak

$$\int_a^b [f(x)]^2 w(x)d(x) = 1 \quad (5.2)$$

$$\int_a^b [g(x)]^2 w(x)d(x) = 1 \quad (5.3)$$

ise bu iki fonksiyon ortonormaldir denir.

Şimdi fourier serilerini kullanarak bir fonksiyonun nasıl ifade edileceğine bakalım

$$f(x) = \frac{1}{2}a_0 + \sum_{n=1}^{\infty} a_n \cos(nx) + \sum_{n=1}^{\infty} b_n \sin(nx), \quad (5.4)$$

Burada

$$a_0 = \frac{1}{\pi} \int_{-\pi}^{\pi} f(x)dx \quad (5.5)$$

$$a_n = \frac{1}{\pi} \int_{-\pi}^{\pi} f(x) \cos(nx)dx \quad (5.6)$$

$$b_n = \frac{1}{\pi} \int_{-\pi}^{\pi} f(x) \sin(nx)dx \quad (5.7)$$

formülleriyle katsayılar bulunur.

Lagrange'ın itiraz ettiği nokta olan süreksizlik noktaları ve keskin dönüşler için bir düzeltme getirilmiştir. Buna göre $f(x)$ fonksiyonu $\overline{f}(x)$ fonksiyonuna dönüştürülür.

$$\overline{f} \equiv \begin{cases} \frac{1}{2} \left[\lim_{x \rightarrow x_0^-} f(x) + \lim_{x \rightarrow x_0^+} f(x) \right] & \text{for } -\pi < x_0 < \pi \\ \frac{1}{2} \left[\lim_{x \rightarrow -\pi^+} f(x) + \lim_{x \rightarrow \pi^-} f(x) \right] & \text{for } x_0 = -\pi, \pi \end{cases} \quad (5.8)$$

Bu şekilde sürekli noktalarda $f(x)$ fonksiyonuna eşit olan bir fonksiyon elde edilmiş olur.

Şimdi fourier serilerinin sinyal işlemedeki şekline bir bakalım:

Sinyal işlemede fonksiyon değişkeni zamandır ve t ile gösterilir. Her periyodik sinyal doğal olarak $f(t) = f\left(\frac{2\pi t}{T}\right)$ eşitliğini sağlar buradaki T sinyalin periyodudur. Bu yeni eşitliğe dayalı olarak Fourier serilerinin formülünü yeniden yazarsak

$$f(t) = \frac{1}{2} a_0 + \sum_{n=1}^{\infty} a_n \cos\left(\frac{2\pi n t}{T}\right) + \sum_{n=1}^{\infty} b_n \sin\left(\frac{2\pi n t}{T}\right) \quad (5.9)$$

olur.

Buradaki a_n, b_n katsayılarında aşağıdaki gibi değişir.

$$\begin{aligned} a_n &= \frac{2}{T} \int_{-T/2}^{T/2} f(t) \cos\left(\frac{2\pi n t}{T}\right) dt \\ b_n &= \frac{2}{T} \int_{-T/2}^{T/2} f(t) \sin\left(\frac{2\pi n t}{T}\right) dt \end{aligned} \quad (5.10)$$

İşimizi biraz daha basitleştirmek ve kullanılabilirliğini artırmak üzere

$e^{inx} = \cos(nx) + i \sin(nx)$ euler eşitliğini kullanarak

$$f(t) = \sum_{n=-\infty}^{\infty} c_n e^{-i \frac{2\pi n t}{T}} \quad (5.11)$$

Buradaki karmaşık katsayı c_n aşağıdaki gibi hesaplanır.

$$c_n = \frac{1}{T} \int_{-T/2}^{T/2} f(t) e^{i \frac{2\pi n t}{T}} dt \quad (5.12)$$

Bu eşitlikler göz önünde bulundurularak asıl konumuz olan frekans dönüşümünü sağlamak üzere fourier dönüşümünün nasıl yapıldığını görelim.

$$\omega = \frac{2\pi n}{T} \text{ olsun , } T \rightarrow \infty \text{ iken } d\omega = \frac{2\pi}{T} dn \text{ olur. ve } F(\omega) \text{ sürekli}$$

fonksiyonunu da c_n ile değiştirdiğimiz zaman aşağıdaki fonksiyonel simetriyi elde ederiz.

$$\begin{aligned} f(t) &= \frac{1}{2\pi} \int_{-\infty}^{+\infty} F(\omega) e^{-i\omega t} d\omega \\ F(\omega) &= \int_{-\infty}^{+\infty} f(t) e^{i\omega t} dt \end{aligned} \quad (5.13)$$

Zamana bağlı $f(t)$ fonksiyonundan $F(\omega)$ dönüşüm fonksiyonunu elde ettik. Yukarıda açıkça görülebileceği gibi ω değişkeninin birimi frekanstır. Bu yüzden elde ettiğimiz dönüşüm fonksiyonuyla zaman bölgesinden frekans bölgesine geçiş yapmış olduk.

5.3. Fourier Dönüşüm Ailesi

Fourier Dönüşümü işlenen sinyalin türüne göre dört kategoriye ayrılabilir. Bir sinyal *sürekli* veya *ayrık* olabilir. Aynı zamanda *periyodik* veya *aperiyodik* olabilir. Bu iki özelliğin birleşimi aşağıda gösterildiği gibi dört ayrı kategori meydana getirir.

Aperiyodik-Sürekli Sinyaller

Bu gruptaki sinyallere, bozulan üsteller ve Gauss eğrisi örnek gösterilebilir. Bu sinyaller hem pozitif hem de negatif sonsuzluğa bir periyodik özellik göstermeden uzayabilir. Bu tip için kullanılan dönüşüme sadece **Fourier Dönüşümü** denir.

Periyodik-Sürekli Sinyaller

Sinüs dalgaları, kare dalga gibi kendisini düzgün bir şablonda eksi sonsuz artı sonsuz aralığında sürekli tekrar eden sinyaller bu gruba girer ve bu sinyallere uygulanan dönüşüme **Fourier Serileri** adı verilir.

Aperiyodik-Ayrık Sinyaller

Bu sinyaller ayrık zamanlarda eksi sonsuzdan artı sonsuza kadar tanımlıdır ve kendilerini hiçbir şekilde tekrar etmezler. Bu tür sinyaller için uygulanan dönüşüme **Ayrık Zamanlı Fourier dönüşümü** denir.

Periyodik-Ayrık İşaretler

Bu türde sinyaller ayrık zamanlı ve periyodik tarzda eksi sonsuzdan artı sonsuza kadar kendini tekrarlayan sinyallerdir. Bu tür sinyallere uygulanan dönüşüme bazı kaynaklarda Ayrık Fourier Serileri dense de genellikle *Ayrık Fourier Dönüşümü* olarak adlandırılır.

Yukarıda saydığımız dört dönüşüm biçiminin birbirlerinden farkları ilk başta önemsiz gibi görünebilir. Ancak uygulamaya dönüş biçimlerinde büyük farklar vardır. Ve ilerideki kaynak araştırmalarında sözle de olsa farklarının bilinmesi faydalı olacaktır. Yukarıda ilginç olan bir nokta da tüm dönüşümlerin sonsuz uzunluktaki sinyaller üzerinde yapılmasıdır. Ancak pratikte elimizde sonsuz uzunlukta bir sinyalin bulunması ve onun işlenmesi mümkün değildir. Elimizdeki çoğu sinyal ayrık olup sınırlı sayıda örneğe sahiptir. Bu ikilemi çözümenin iki yolu vardır. Birincisi sinyalimizi sonsuz olarak düşünmektir. Bunun için sinyalimiz eksi sonsuza ve artı sonsuza uzatılır ve örneklem dışında kalan kısımları sıfır değerli düşünülür. Böyle yaptığımız zaman elimizde aperiodyk ve ayrık bir sinyal bulunur. Ve bu sinyal türü DTFT ile işlenebilir. Ancak Aperiodyk bir sinyal sonsuz sayıda sinüsoidal bileşene ayrılır ve bilgisayar tarafından hesaplanamaz. Bu soruna ikinci bir çözüm de elimizdeki sinyalin hem sol tarafında hem de sağ tarafında kendini tekrar ettiğini varsaymamız olacaktır. Bu sayede elimizde ayrık ve periyodik bir sinyal olacaktır. Bu sinyalin dönüşümü de yukarıda bahsettiğimiz DFT ile gerçekleştirilir.

5.4. Ayrık Fourier Dönüşümü (DFT)

Ayrık Fourier Dönüşümü aşağıdaki eşitlikteki gibi ifade edilir.

$$X_m = \sum_{k=0}^{N-1} x_k e^{\frac{-j2\pi mk}{N}} \quad (5.14)$$

Aşağıdaki farklar haricinde yukarıda bahsettiğimiz Sürekli Fourier Dönüşümüne benzer.

- Zamanın sürekli fonksiyonu olan sinyalimiz $f(t)$ zaman dizisi x_k ya dönüşmüştür.
- Sürekli zaman olan t , zaman indeksi veya örnek numarası olan k ya dönüşmüştür. (k 0,1,2... N-1 değerlerini alır).

- Sinyalin uzunluğu N adet örneğe çevrilmiştir
- İntegral işlemi Toplam işlemiyle yer değiştirmiştir.
- Frekans fonksiyonu olan $F(\omega)$ frekans dizisi olan X_k halini almıştır.
- Sürekli f frekans fonksiyonu 0,1,2,...N-1 değerlerini alan m frekans indeksi olmuştur.

Frekans domeninden zaman domenine geri dönüşü sağlayan Ters Ayırık Fourier Dönüşümü de aşağıdaki gibi ifade edilir.

$$x_k = \frac{1}{N} \sum_{m=0}^{N-1} X_m e^{\frac{j2\pi mk}{N}} \quad (5.15)$$

Ayrık Fourier dönüşümünde N adet örnek için Gerçel ve Sanal olmak üzere iki çeşit veri elde edilir. her birisinin sayısı $(N/2)+1$ dir. Bu iki kısım elde edilen Sinüs ve Kosinüs lerin genlik bilgilerini içerir ve bunlara frekans bileşenleri adı verilir. Gerçel kısımlar (**Re X[]**) kosinüs bileşenlerin genliklerini, Sanal kısımlar (**Im X []**) ise Sinüs bileşenlerin genliklerini gösterir. Frekans bölgesi de zaman bölgesiyle aynı sinyali ifade eder ancak farklı bir biçimde gösterir. Bu şekilde fourier dönüşümü kullanarak zaman bölgesinden frekans bölgesine geçiş *ayrıştırma, analiz veya ileri fourier dönüşümü olarak adlandırılır*. Frekans bölgesinden, zaman bölgesine geçiş ise *sentez veya Ters Ayırık Fourier Dönüşümü* olarak adlandırılır.

Yukarıdaki formülden de hesaplanabilmesine rağmen daha anlaşılır olduğu için gerçel ve sanal kısımlar aşağıdaki gibi ayrı ayrı hesaplanabilir.

$$\text{Re } X[k] = \sum_{i=0}^{N-1} x[i] * \text{Cos}(2\pi ki / N) \quad (5.16)$$

$$\text{Im } X[k] = -\sum_{i=0}^{N-1} x[i] * \text{Sin}(2\pi ki / N) \quad (5.17)$$

5.4.1. Kutupsal Gösterim

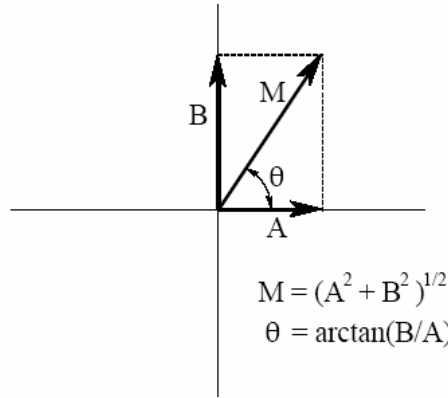
Yukarıda da gösterildiği gibi Frekans bölgesi sinüs ve kosinüslerin genliklerini tutar. Bu tür gösterime *dörtgen* (rectangular) gösterim denir. Frekans bölgesinin bir başka gösterimi de *kutupsal* gösterim (polar notation) dir. Sanal ve Gerçek kısımlar *büyüklük* (magnitude) ve *faz* (phase) $\text{Mag } X[]$ ve $\text{Phase } x[]$ olarak gösterilir. Mag ve Phase değerleri karşılık gelen sanal ve gerçel kısımlar kullanılarak

hesaplanırlar. Bu yüzden birebirdirler. Yani $\text{Mag } X[0]$ ve $\text{Phase } X[0]$, $\text{Re } X[0]$ ve $\text{Im } X[0]$ kullanılarak hesaplanır.

Polar dönüşümde:

$A \cdot \text{Cos}(x) + B \cdot \text{Sin}(x) = M \cdot \text{Cos}(x + \theta)$ eşitliğinden faydalanılır. Bu eşitliğe göre aynı frekanstaki farklı genliğe sahip bir sinüs ile kosinüs dalgasının toplamı farklı genlik ve faz farkına sahip bir başka Kosinüs işaretine karşılık gelmektedir.

Yeni elde ettiğimiz M ve θ değerleri hiçbir bilgi kaybına neden olmaz. A ve B değerleri M ve θ kullanılarak tekrar elde edilebilir.



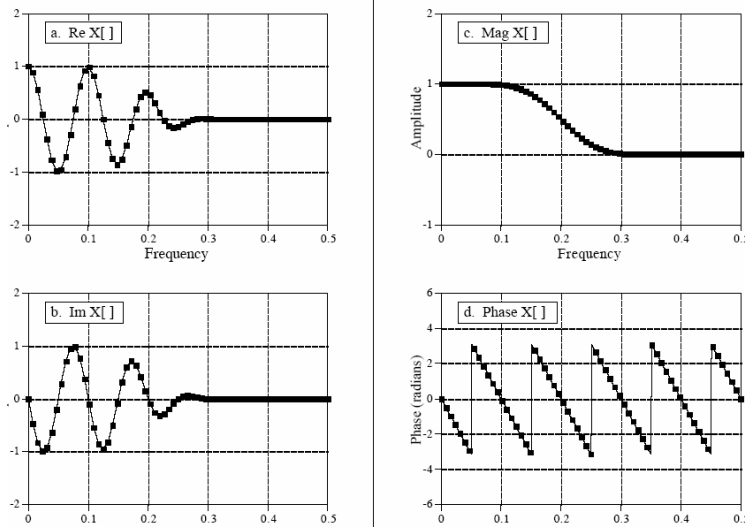
$$\text{Mag}X[k] = \sqrt{\text{Re } X[k]^2 + \text{Im } X[k]^2} \quad \text{Re } X[k] = \text{Mag}X[k] \cos(\text{Phase}X[k])$$

$$\text{Phase}X[k] = \arctan\left(\frac{\text{Im } X[k]}{\text{Re } X[k]}\right) \quad \text{Im } X[k] = \text{Mag}[k] \sin(\text{Phase}X[k])$$

Polar ve dörtgen gösterim kesin olarak aynı bilgiyi ifade etse de uygulamalarda birisinin diğerine göre daha kolay kullanıldığı alanlar bulunmaktadır. Dörtgen gösterim genelde hesaplamalar ve bilgisayar programları için idealdir. Buna karşın grafik gösterimlerin çoğu kutupsal formdadır.

Dörtgen Gösterim

Kutupsal Gösterim



Şekil 5. 5. Fourier Dönüşümünde Dörtgen ve Polar Gösterim

5.4.2. Pencerelemiş Fourier Dönüşümleri

(Windowed Fourier Transforms)

Eğer $f(t)$ periyodik olmayan bir sinyalse, periyodik olan sinüs ve kosinüslerin toplamı sinyelimizi doğru olarak göstermezler. Elimizdeki sinyali periyodik hale getirmek için yapay olarak uzatabiliriz. Fakat sonlanma noktalarında ek süreklilikler gerektirir. Pencerelemiş Fourier dönüşümleri bu sorun için bir çözümdür ve periyodik olmayan sinyallerin daha iyi bir şekilde temsil edilmesini sağlar. WFT, sinyaller hakkında hem zaman bölgesinde hem de frekans bölgesinde eşzamanlı bilgi vermek üzere kullanılabilirler.

WFT ile giriş sinyali $f(t)$ bölgelere ayrılır ve her bölge kendi frekans içerikleri için ayrı olarak analiz edilir. Eğer sinyal keskin geçişlere sahip ise giriş verilerini sonlanma noktalarında sıfır olacak şekilde pencereleriz. Bu pencereleme sonlanma noktalarını ortadaki noktalardan daha az vurgulayacak bir ağırlık fonksiyonuyla gerçekleştirilir. Pencerenin etkisi, sinyali zamanda yerelleştirmesidir.

5.4.3. Hızlı Fourier Dönüşümleri (Fast Fourier Transforms)

Ayrıık fourier dönüşümlerinde n adet örneğe sahip bir sinyalin dönüşüm işlemlerinin hesaplanması için $n \times n$ bir matrisin uygulanması gerekir. $n \times n$ bir matrisin bir vektör ile çarpılması n^2 lik aritmetik işlem maliyeti demektir. Ancak verimiz içerisindeki örnekler düzenli bir şekilde dağıtılmış ise $n \times n$ matrisimiz yerine çok daha küçük birkaç matris ile $n \log n$ lik aritmetik işlemle aynı sonuçlar elde edilir. Bu yüzden bu işleme Hızlı Fourier Dönüşümü (FFT) adı verilir.

5.5. Dalgacık (Wavelet) Dönüşümü

Dalgacıklar, veriyi farklı frekans bileşenlerine ayıran ve sonra her bir bileşeni, bileşenin ölçeğiyle eşleştirilmiş bir çözünürlükle çalışan matematiksel fonksiyonlardır. İşaretin (sinyalin) süreksizliklere ve keskin, sivri uçlara sahip olduğu fiziksel durumları incelemede zayıf kalan geleneksel Fourier yöntemleri üzerine avantajlara sahiptir. Dalgacıklar matematik, kuantum fiziği, elektrik mühendisliği ve sismik jeoloji alanlarından bağımsız olarak geliştirildiler. Son on yıl içerisinde farklı disiplinlerin bir araya gelmesi ve bu alanlardaki bilgi değiş tokuşu ile resim sıkıştırma, türbülans, insanın görme gücü, radar ve depremi önceden tahmin etme gibi birçok yeni dalgacık uygulamalarına yol açtı.

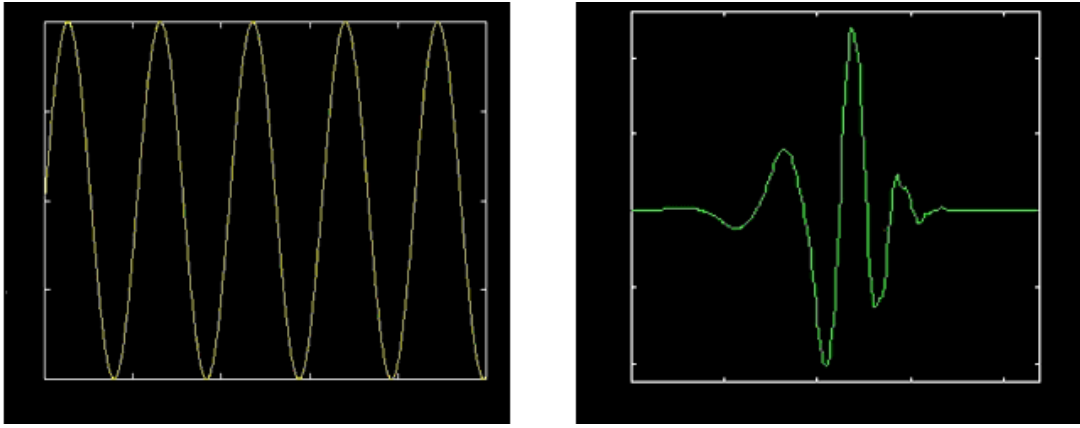
Dalgacık algoritmaları veriyi farklı ölçek ve çözünürlükte incelerler. Eğer bir sinyale geniş bir pencereden bakarsak büyük özelliklerinin fark ederiz. Eğer küçük bir pencereden bakarsak ta küçük özelliklerini fark edebiliriz. Açıkça konuşmak gerekirse dalgacık analizinde sonuç hem ormanı hem de ağaçları görmektir. Bu dalgacıkları ilginç ve kullanışlı kılar. Onlarca yıldır bilim adamları, değişken sinyalleri analiz edebilmek için Fourier analizinin temelini oluşturan sinüs ve kosinüs fonksiyonlarından daha kullanışlı fonksiyonlar bulmaya çalıştılar. Keskin dönüşlere ve süreksizliklere yaklaşmada geline son nokta dalgacık analizidir.

Wavelet analizinde öncelik, çözümleyici veya ana dalgacık (mother wavelet) denilen prototip fonksiyonu seçmektir. Ana dalgacığın sıkıştırılmış yüksek frekanslı versiyonuyla zaman analizi yapılır. Aynı şekilde ana dalgacığın genişletilmiş düşük frekanslı versiyonuyla da bu analiz tekrar edilir. Çünkü orijinal sinyal veya fonksiyon bir yayılmış bir dalgacığın (wavelet fonksiyonlarının lineer birleşimlerinin katsayıları kullanılarak) katsayıları cinsinden ifade edilebilir. Ve eğer verinize iyi uyarlanmış bir dalgacık seçerseniz veya belli bir eşik değerinin altındaki katsayıları atarsanız veriniz çok yaklaşık ancak kısmen temsil edilmiş olur. Bu kısmi kodlama dalgacıkları veri sıkıştırma alanında harika bir araç yapmaktadır.

Dalgacıkların kullanıldığı diğer uygulamaları alanların içerisinde astronomi, akustik, nükleer enerji, alt bant kodlama, sinyal ve görüntü işleme, nöropsikoloji, müzik, MR görüntüleme, ses ayırma, optik, fraktaller, türbülans, deprem tahmini, insan görmesi ve kısmi diferansiyel denklemlerin çözümü gibi soyut matematik uygulamaları sayılabilir.

5.6. Dalgacık Analizi

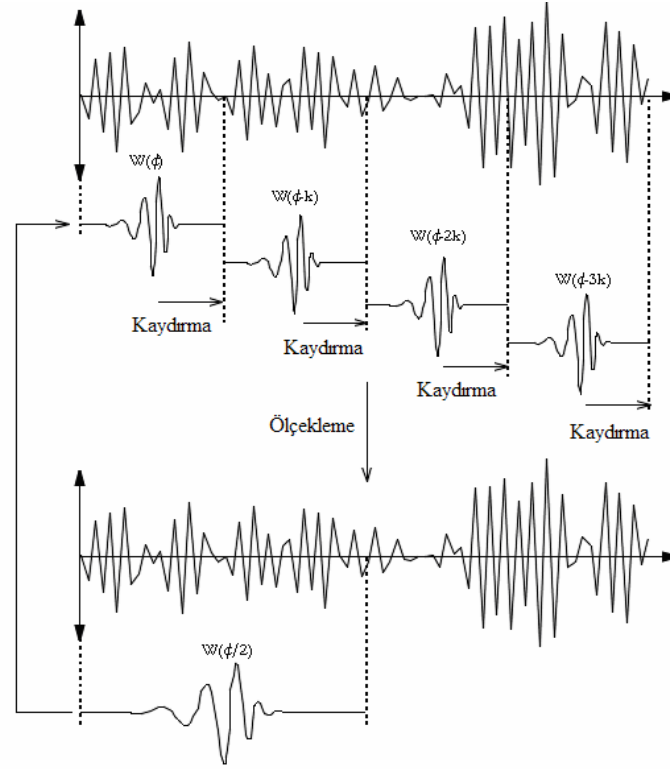
Dalgacık analizi sinyallerin işlenebilmesi için alt bileşenlerine ayırması bakımından Fourier analizine benzer. Fourier Dönüşümü bir sinyali farklı frekanslardaki sinüs ve kosinüs bileşenlerine böler. Wavelet dönüşümü ise sinyali ana dalgacığın ölçeklenmiş ve kaydırılmış alt dalgacıklarına böler. Dubechies 5 ve sinüs sinyalinin verildiği aşağıdaki şekilde ne kadar büyük farklar olduğu görülmektedir. Karşılaştırdığımız zaman sinüs dalgası yumuşak ve sonsuz uzunluktadır. Dalgacık ise düzensiz bir yapıda öz olarak sunulmuştur.



Şekil 5. 6. Sinüs İşareti ve Ana Dalgacık

Dalgacıkların düzensiz yapısı, süreksiz ve keskin geçişlere sahip sinyallerin analizine izin verirken öz (compact) yapıları sayesinde de sinyallerin niteliklerinin zamana bağlı olarak belirlenmesini de sağlar.

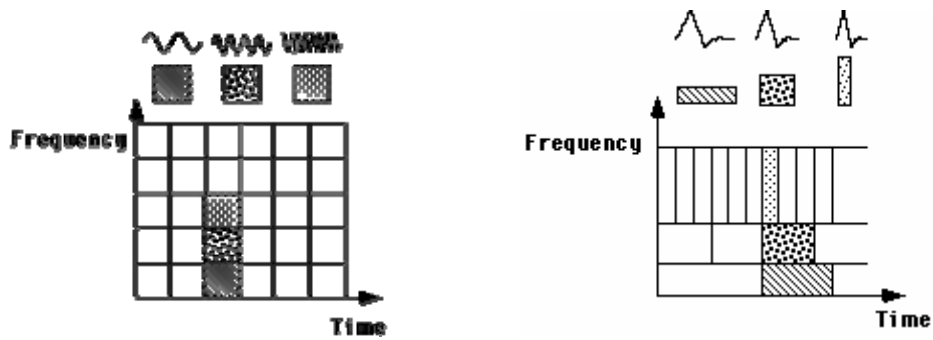
Genellikle durağan olmayan türdeki sinyallerin işlenmesinde, sinyalin zaman ve frekans bölgeleri arasında bir ilişki elde etmek çok faydalı olmaktadır. Fourier dönüşümü frekans bölgesi hakkında bilgiler verirken, zamandaki yerel bilgiler bu işlem sırasında kaybolmaktadır. Bu problemin nedeni frekansların, zaman içerisindeki konumlarıyla, frekans bölgesindeki özelliklerini birleştirebilme yeteneği olmamasıdır. Bunun neticesinde frekans bölgesindeki en ufak değişiklik tüm zaman boyunca değişikliğe neden olmaktadır. Fourier Dönüşümünün tersine, dalgacık dönüşümü ana dalgacığın kaydırılmasıyla zaman bölgesinde, ana dalgacığın ölçeklendirilmesiyle de frekans bölgesinde yerinin belirlenmesine izin verir. Ana dalgacığa uygulanan kaydırma ve ölçekleme işlemleri, sinyalin yerel bölgeleri ile dalgacık arasındaki karşılıklı ilişkiyi temsil edecek olan katsayıları belirlemede kullanılır.



Şekil 5. 7. Dalgacık Analizinin Aşamaları

Dalgacık analizlerinde sinyal bölme problemini çözmek için tamamen ölçeklenebilir olarak ayarlanmış pencereleri kullanır. Pencere sinyal boyunca kaydırılarak her pozisyonu için spektrumu yeniden hesaplanır. Ve bu işlem biraz daha uzun veya kısa pencerelerle her döngü için tekrarlanır. En sonunda sonuç zaman-frekans temsillerinin toplamı olarak bulunur.

Bu temsillerin toplamı yüzünden çok çözünürlüklü analiz olarak ta adlandırılır. Dalgacıklar söz konusu olunca genellikle zaman-frekans gösterimi olarak değil zaman-ölçek gösterimi kullanılır.



Şekil 5. 8. Fourier ve Dalgacık Analizlerinin Zaman Frekans Ölçeğinde Gösterimi

5.6.1. Sürekli Dalgacık Dönüşümü

Wavelet Dönüşümü Temel dalgacık fonksiyonu olan $\Psi(t)$ nin a ölçek parametresini ve b zaman parametresi kullanılarak, zamana bağlı $f(t)$ fonksiyonuna uygulanmasıdır. Ve aşağıdaki eşitlik kullanılarak dalgacık dönüşümü yapılır.

$$W(a,b) = \frac{1}{\sqrt{a}} \int_{-\infty}^{\infty} f(t) \cdot \overline{\Psi\left(\frac{t-b}{a}\right)} dt \quad (5.18)$$

Eş. (5.18) $W(a,b)$ Wavelet dönüşümünün katsayılarıdır. $\overline{\Psi(t)}, \Psi(t)$ nin karmaşık eşleniğidir.

Tersi olarak dalgacık katsayıları kullanılarak bir sinyal ters dalgacık dönüşümüyle tekrar oluşturulabilir

$$f(t) = \frac{1}{C_{\Psi}} \int_{-\infty}^{\infty} W(a,b) \cdot \frac{1}{\sqrt{a}} \cdot \Psi\left(\frac{t-b}{a}\right) \cdot \frac{1}{a^2} da db \quad (5.19)$$

Eş.(5.19) da C_{Ψ} kullanılan dalgacığa ait bir sabittir.

Eş.(5.19) u basitleştirmek amacıyla dalgacık katsayıları ölçek ve zamanı ayırık zamanda diyadik değişkenler m ve n olarak aldığımız zaman Ayırık Dalgacık Dönüşümü (DWT) nü elde ederiz.

$$a = a_0^m \quad \text{ve} \quad na_0^m b_0 \quad m, n \in Z$$

$$DWT(m,n) = a_0^{\frac{m}{2}} \int f(t) \cdot \Psi(a_0^{-m} t - nb_0) dt \quad (5.20)$$

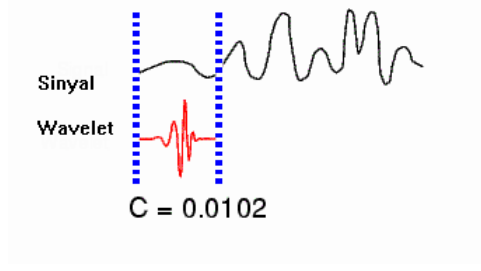
Eşitliğini elde ederiz eşitlikteki a_0, b_0 ölçek ve zaman eksenlerinde yapacağımız örnekleme oranlarını belirlemede kullanılan sabitlerdir.

Şimdi yukarıda anlattığımız dönüşüm işlemini basamak basamak uygulanmasına bakalım.

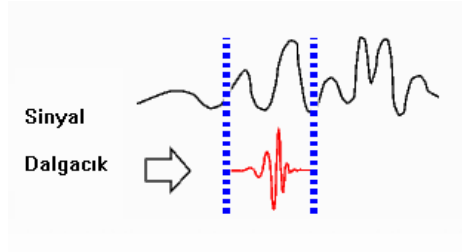
Adım 1. Kullanılacak dalgacık seçilir. Ve sinyalinin ilk bölümüyle karşılaştırılır.

Adım 2. Sinyalinin o bölgesi ve Dalgacık kullanılarak C sabit sayısı hesaplanır. C sayısı ne kadar yüksekse dalgacığımızla sinyaliniz o kadar benzerdir.

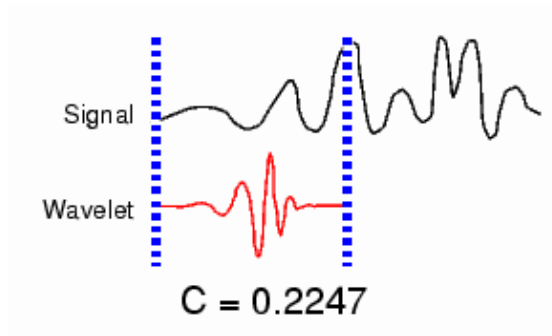
Ancak sinyalin o bölgesinin enerjisiyle dalgacığın enerjisi eşitse C bir korelasyon katsayısı olarak kullanılabilir.



Adım 3 Dalgacığı sağa doğru kaydır ve tüm sinyali kapsayana kadar adım 1 ve 2 yi tekrar et



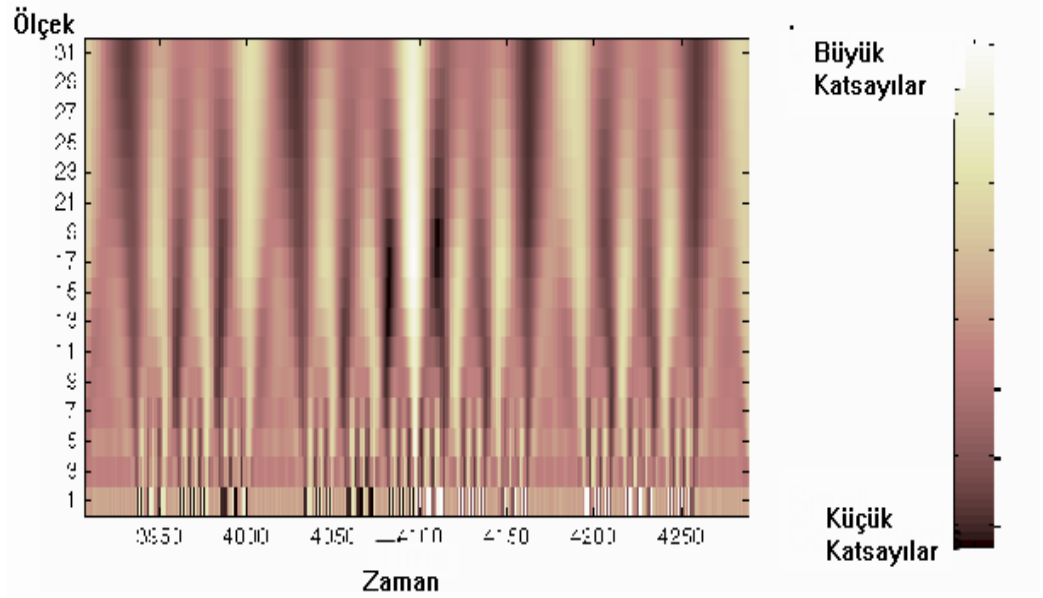
Adım 4 Dalgacığı ölçeklendir (genişlet) ve Adım1 den 3' e kadar tekrar et



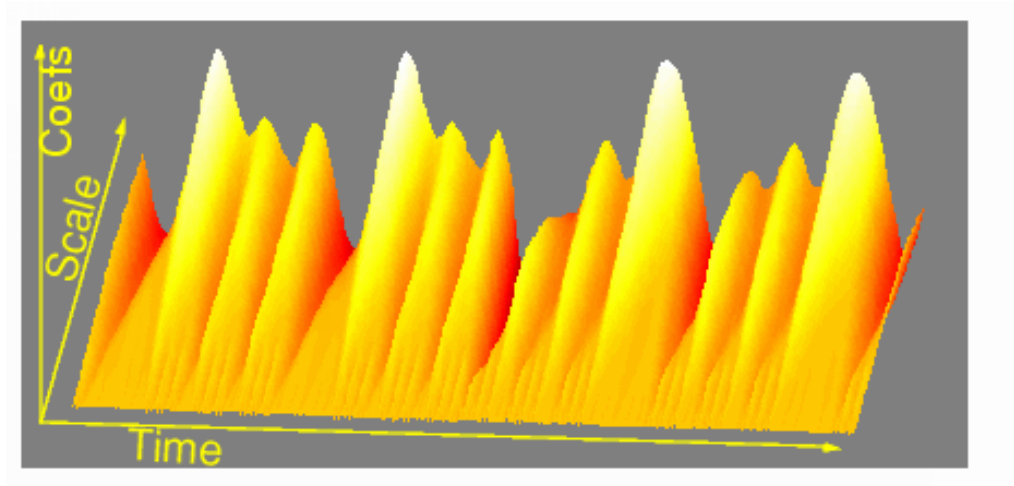
Adım 5. Tüm ölçekler için adım1 den 4 kadar tekrar et

Yukarıdaki adımları tamamladığımızda sinyalin farklı bölgelerinde farklı ölçeklerde elde edilmiş katsayıları elde edeceğiz. Bu Katsayılar sinyalinizmin dalgacıklar üzerine uygulanmış regresyon sonuçlarıdır.

Peki bu katsayıları görsel olarak nasıl anlaşılabilir hale getirebiliriz. Şekil 5.9. ve Şekil 5.10 bu katsayıların grafiksel gösterim biçimlerinden ikisidir.



Şekil 5. 9. Dalgacık Dönüşümünden Hesaplanan Katsayıların Zamanla Değişen İki Boyutlu Gösterimi

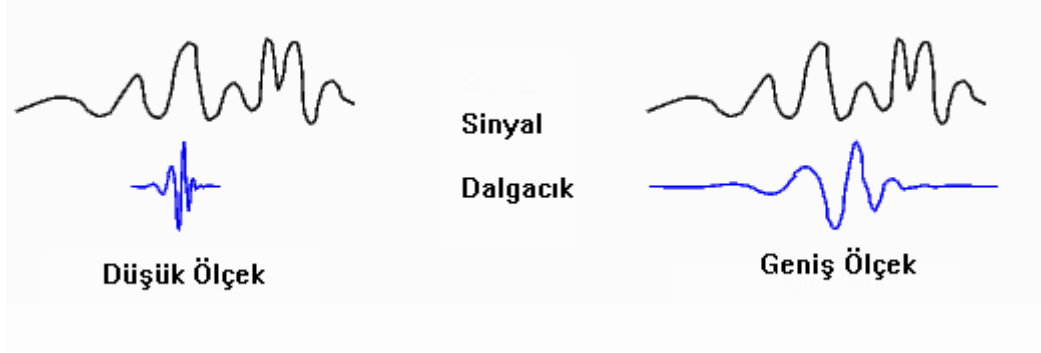


Şekil 5. 10. Dalgacık Dönüşümünden Hesaplanan Katsayıların Zamanla Değişen Üç Boyutlu Gösterimi

Wavelet dönüşümünde elimizde katsayılar ve ölçek bulunur. Peki bunu nasıl yorumlayacağız ?

Genişletilmiş dalgacık daha uzun sinyal bölgelerinde çalışır. Böylece dalgacık katsayıları daha kaba sinyal özelliklerini ortaya koyar.

Sıkıştırılmış dalgacıklar ise dar bantlarda çalışarak hızlı değişen küçük özellikleri meydana çıkarır.



Düşük Ölçek $a \Rightarrow$ Şıkışık Wavelet \Rightarrow Hızlı Değişen Detaylar \Rightarrow Yüksek Frekans(ω)

Yüksek Ölçek $a \Rightarrow$ Geniş Wavelet \Rightarrow Yavaş Değişen, Kaba Özellikler \Rightarrow Düşük Frekans(ω)

5.7. Ana Dalgacıkların Özellikleri

Dalgacık dönüşümünün baz fonksiyonu olan dalgacıklar ve onlara yardımcı ölçeklendirme fonksiyonlarının önemi aşıkardır. Dalgacık dönüşümü amaç değil, bir araçtır. Yapacağımız işe göre en uygun dalgacık fonksiyonunu bulmak ise bize düşmektedir. Ancak yeni bir dalgacık geliştirmek kolay değildir. Yeni bir dalgacığın geliştirilmesinde derin matematiksel altyapıya ihtiyaç vardır. Kullanılacağı yere göre bugüne kadar bazı dalgacık aileleri geliştirilmiştir. Kullanılacakları yere göre dalgacıklarda bazı özellikler öne çıkmaktadır. Önemli kriterlerden bazıları ise şunlardır.

Dalgacık desteği (wavelet support): Bu özellik zaman t veya frekans ω sıfıra indiğinde sıfıra yakınsama hızıdır. Zaman ve frekansın yerini saptamak için çok önemlidir.

Simetri : (symmetry) Özellikle görüntü işlemede defaz dan kaçınmak için önemlidir.

Kaybolma anlarının sayısı (vanishing moment) : Sıkıştırma amaçlı olarak kullanılacağı zaman önemlidir.

Düzenlilik (Regularity): Yeniden düzenlenen bir görüntünün yumuşaklığı veya lineer olmayan regresyonlarda yaklaşık fonksiyonların uygunluğu gibi daha ince özelliklerin çıkarılması için önemlidir.

Ölçeklendirme fonksiyonu (ϕ) na sahip olmak

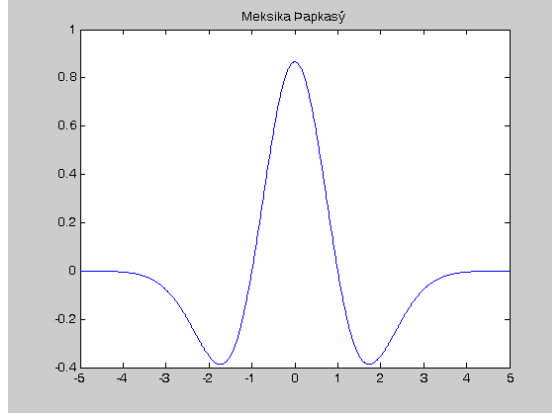
Ortogonallik ve biortogonallik : Sonuç analizlerinde kolaylık sağlar

Bunların yanında daha az önemli ancak kullanışlı diğer özellikler

- Açıkça bir şekilde anlaşılacak ifade olmaları
- Çizelgeleme kolaylığı
- Kullanım Kolaylığına

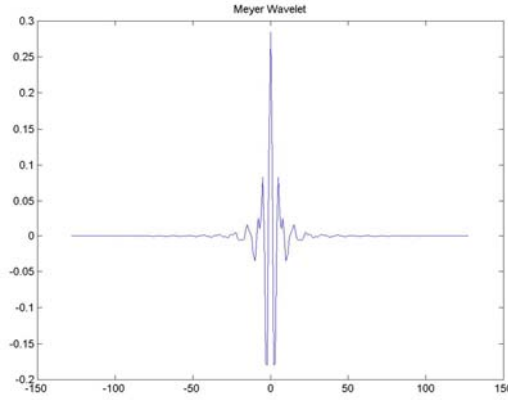
dikkat etmek gerekir.

5.7.1. Örnek Ana Dalgacık Şekilleri ve formülleri



Şekil 5. 11.Meksika Şapkası

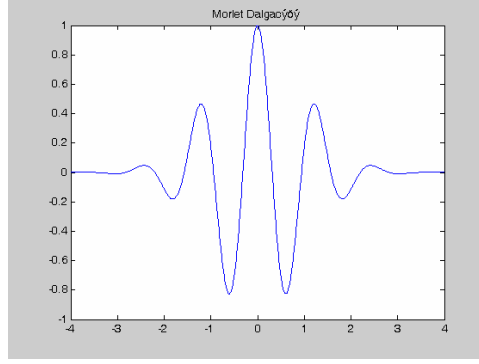
$$\psi(x) = \left(\frac{2}{\sqrt{3}}\pi^{-1/4}\right)(1-x^2)e^{-x^2/2}$$



Şekil 5. 12. Mayer Dalgacı

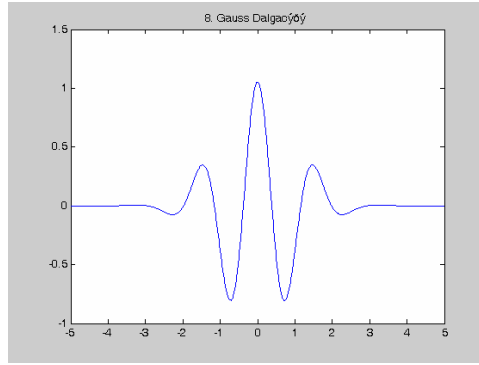
$$\hat{\psi}(\omega) = (2\pi)^{-1/2} e^{i\omega/2} \sin\left(\frac{\pi}{2} \nu \left(\frac{3}{2\pi} |\omega| - 1\right)\right) \quad \text{eğer} \quad \frac{2\pi}{3} \leq |\omega| \leq \frac{4\pi}{3}$$

$$\hat{\psi}(\omega) = (2\pi)^{-1/2} e^{i\omega/2} \cos\left(\frac{\pi}{2} \nu \left(\frac{3}{4\pi} |\omega| - 1\right)\right) \quad \text{eğer} \quad \frac{4\pi}{3} \leq |\omega| \leq \frac{8\pi}{3}$$



Şekil 5. 13. Morlett Dalgacı

$$\psi(x) = e^{-x^2/2} \cos(5x)$$



Şekil 5. 14. Gauss Dalgacı

$$F = C_p e^{-x^2}$$

6.KONUŞMA/KONUŞMACI TANIMADA WAVELET

Dalgacık dönüşümü yaklaşık 30 senelik gelişiminin ardından teknolojinin alt disiplinlerinde de kullanılmaya başlanmıştır. Şu bir gerçektir ki yeni çıkan bir yöntem öncelikle kendisini ispat etmek zorundadır. Matematik fizik gibi temel bilimlerdeki çalışmalar belli bir doyuma ulaştıktan sonra bu yöntem uygulamalardaki yerini almaya başlar.

Dalgacık dönüşümü ve onların uygulamaları özellikle son on yıl içerisinde konuşmacı tanımayla ilgili disiplinlerde de kullanılmaya başlanmıştır. Dalgacık dönüşümünün konuşma ve konuşmacı tanıma sistemlerindeki belli başlı uygulama alanları aşağıdaki gibi sıralanabilir.

- Elde edilen konuşma dalgalarının iyileştirilmesi
- Konuşma iletimin kanallarının dayanıklı hale getirilmesi
- Konuşma dalgalarının kodlanması
- Daha etkin öznitelik vektörü çıkarmak amacıyla DFT'nin yerini alması
- Konuşma ve Konuşmacı Modellemede diğer yöntemlerle birlikte kullanılması

(Avcı ve Akpolat 2006) nin sundukları çalışmada Dalgacık paket ayrıştırma yöntemi kullanılarak gerçek zamanda öznitelik vektörlerinin elde edilmesini sağlamışlardır.ve Adaptif Fuzzy sisteme iletilmesinde kullanılmıştır. Bu çalışmada gürültülü verilere karşı %92 lik bir başarıyı sağlamıştır.

Ganbari ve Ark.,2005 yaptıkları çalışmada VAD adını verdikleri bir sisteme uyguladıkları dalgacık eşikleme algoritması sayesinde Gauss, pink ve çoklu konuşmacı gürültülerine karşı etkili bir sistem geliştirmişlerdir. Benzer şekilde (Lu ve ark.2004) parçalı gürültülü seslerin analizinde kullanılacak olan bir kazanç faktörü elde etmişlerdir. Bunu yaparken dalgacık tabanlı bir iyileştirme yöntemi geliştirmişlerdir.

Durağan olmayan geniş bant gürültülerinin konuşma sinyalinden ayıklanmaları çok zordur. (Lou ve Ark. 2002) de geliştirilen dalgacık dönüşümünü kullanan bir yöntemle ses sinyalleri öncelikle dalgacık uzayında bileşenlerine ayrılmıştır. Daha sonra Karhunen-Loeve dönüşümünü temel alan vektör uzayında iyileştirilmiş ve tekrar birleştirilmiştir. Gerçekleştirilen çalışma sonunda Durağan olmayan geniş bant gürültülerine karşı kullanılan diğer yöntemlere göre başarı elde edilmiştir.

(Erçelebi 2002) tarafından yapılan çalışmada ise ses tanıma sistemlerinin önemli bir konusu olan temel frekansın (pitch) belirlenmesi için ikinci nesil dalgacık dönüşümleri kullanılmış ve gürültülü ortamlarda bile başarılı olabilen bir yöntem geliştirilmiştir. Geliştirilen bu yöntem hem düşük temel frekanslı hem de yüksek temel frekanslı konuşmacıların seslerinde başarılı olmuştur.

Obaidat ve Arkadaşları tarafından yapılan çalışmada ise diydik dalgacık dönüşümleri kullanılarak pitch periyodun belirlenmesi sağlanmıştır ve temel frekansın bulunmasında % 100'e yakın bir başarı elde edilmiştir.

(Namba M ve Ishida Y 1998) tarafından yapılan çalışmada parametre normalizasyonun en temel ve en çok kullanılan yöntemlerinde birisi olan Kör dekonvolüsyon (blind deconvolution) yöntemine, ortogonal dalgacık yöntemiyle bir yaklaşım sergilenmiş eigen vektörlerinin hesaplanmasında daha hızlı sonuçlar elde edilmiştir.

(Singh ve Ark.1997) tarafından yapılan ve Gizli markov modellerinin kullanıldığı çalışmada Ayrık Dalgacık dönüşümünü kullanan bir iyileştirme yöntemiyle hem bölütleme de hem de tanıma performansında artış sağlamışlardır.

(Lung 2004) tarafından gerçekleştirilmiş bir çalışma Adaptif fuzzy C-ortalamarı ve dalgacık dönüşümü kümelemesi aşamalarından geçirilen ses sinyalinden daha kaliteli öznelik vektör elde edilmeye çalışılmıştır. Elde edilen bu öznelik vektörleri konuşmacı tanıma da kullanılmış ve %95 lik bir doğru tanıma yüzdesi elde edilmiştir.

Yine (Lung 2006) tarafından kaleme alınan makalede dalgacık paketleri kullanılarak elde edilen öznitelikler fuzzy-neural bir ağına verilerek eğitilmiş sonuçta yüksek oranda konuşma tanıma başarısı sağlanmıştır.

7. DENEYSEL ÇALIŞMA

Tezimizin son aşamasında, bundan önceki bölümlerde değinilen konulara göre; içerisinde dalgacık dönüşümünün de kullanılacağı bir konuşmacı tanıma uygulaması geliştirilmiştir. Konuşmacı tanımının kullanım amacını doğru yansıtması için uygulamamızda bir güvenlik kontrol mekanizması oluşturulmuştur. Programda kullanıcılar sisteme tanıtıldıktan sonra her kullanıcıya bir ID (tekil tanımlayıcı) tanımlanır. Kullanıcılar sisteme kendilerine verilen bu ID ve ses örneğini kullanarak giriş yaparlar. Uygulamamız Metin-Bağımlı bir sistem üzerine inşa edilmiştir. Bu yüzden kullanıcılar hem eğitim hem de test aşamalarında kendi isimlerini telaffuz edeceklerdir.

Uygulama aşağıda belirtilen aşamalardan oluşur.

Kullanıcı Veritabanı Oluşturma

Bu aşamada kullanıcıya ait bilgiler kaydedilir. Ve her bir kullanıcıya bir ID verilir. Yine bu bölümde tanımlanan kullanıcılar için ses örnekleri mikrofon aracılığı ile toplanır ve kaydedilir. Yapılan literatür çalışmalarında konuşmacı doğrulama sistemlerinin her bir konuşmacıdan 3-10 telaffuz örneği aldığı görülmüştür. Ancak bu çalışmaların pek çoğu sadece üç telaffuz örneği alınmasının hatalı red oranlarının yüksek olmasına sebep olduğunu söylemektedir. 10 telaffuz örneği alınması ise hem yüksek hesaplama maliyeti getirmekte hem de beş örneğe göre çok da önemli bir başarı yakalamamaktadır. Bu yüzden her kullanıcıdan beş adet örnek alınmasına karar verilmiştir.

Örnekler mikrofon aracılığı ile 8 kHz ve tek kanallı (mono) olarak toplanmıştır. 8 kHz seçilmesinin nedeni hem hesaplama yükünün az olması (Her ses örneği byte olarak depolanır.) hem de telefon kalitesi olarak bilinmesidir. Eğer başarılı sonuçlar alınacak olursa sistem otomatik olarak telefon aracılığı ile yapılacak olan doğrulama mekanizmalarında da kolaylıkla kullanılabilir.

Kullanıcı Kayıt Sistemi

ID: 1001
 Adı: ONUR
 Soyadı: AYGÜN
 Baba Adı: SAİT AHMET
 Doğum Tarihi: 23.09.1976
 Doğum Yeri: ALTINDAĞ
 Çalıştığı Bölüm: BİLGİ İŞLEM
 Yetki: 1

Kişiyeye Ait Ses Örnekleri

SıraNo	Tarih	Dosya Adı	SIR.
2	27.12.2006	1001_2.WAV	
3	27.12.2006	1001_3.WAV	
4	27.12.2006	1001_4.WAV	

Resim Yükle

Katsayıları Göster

Sınırlama **PreEmphasis** **Katsayılar** **DTW** **Örnek Ekle** **Örneği sil**

Şekil 7. 1 Kullanıcı Veritabanı Ekranı

1920

Kayıt **Çal** **Dur** **Kaydet**

Süre 0.00

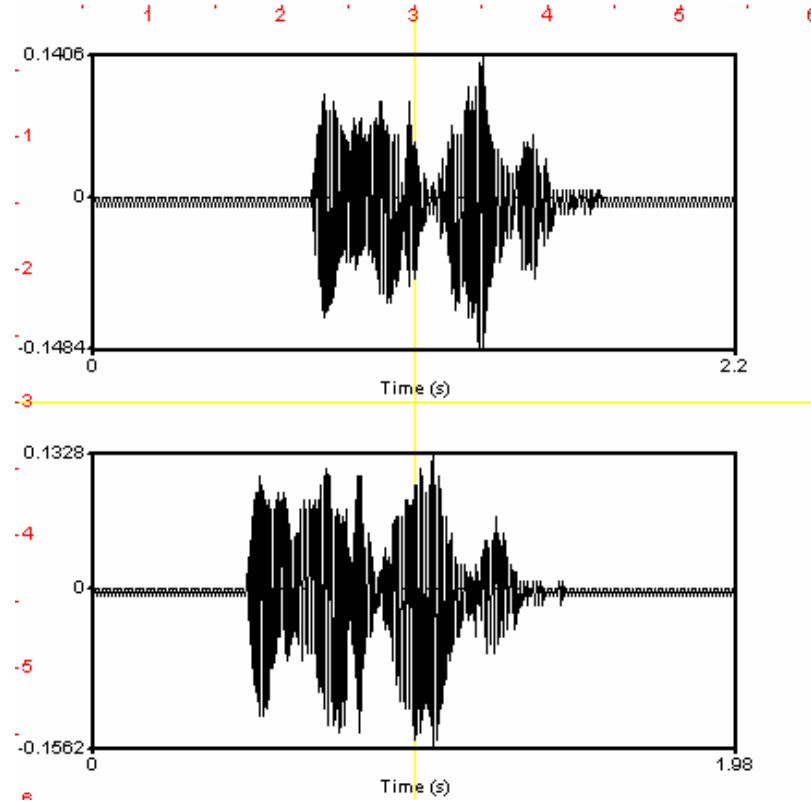
Kapat

Şekil 7. 2 Kullanıcılara ait Örnek Toplama Ekranı

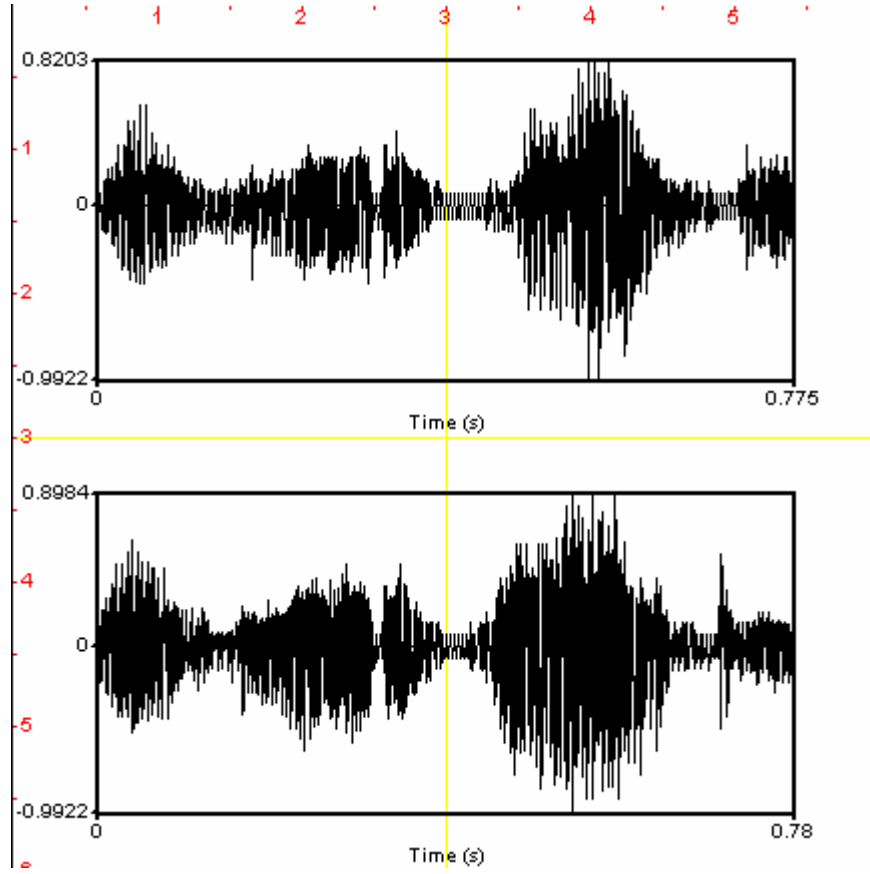
Ön İşleme

Alınan örnekler direk veritabanına kaydedilmeyip gürültü temizleme ve bölütleme işlemleri yapılmaktadır. Bu şekilde doğrulama işlemlerinin başarısı artırılmaktadır. Gürültü temizleme aşamasında kayıt araçlarından ve ortamdan kaynaklanan gürültülerin giderilerek spektral düzlemdeki sapmaların önüne geçilmesi hedeflenmektedir. Bu aşama da dalgacık dönüşümü kullanılmaktadır. Dubechies dalgacığına ait katsayılar kullanılarak önce sinyal dönüşüme uğratılmakta sonra elde edilen detay katsayılar düzenlenerek ters dalgacık dönüşümü uygulanmaktadır.

Bölütleme aşamasında kullanıcıdan alınan telaffuz örneklerinin başındaki ve sonundaki sessiz kısımlar çıkarılmaktadır. Bu sayede hem hesaplama yükü azaltılmakta hem de karşılaştırma aşamasında birbirine uyumlu parçalar kullanılmaktadır. Bunun için 10 mslik pencerelenmiş sinyal kısa dönemli enerji değişimlerine ve sıfırı geçme sayılarına göre incelenmekte ve belli bir eşik değerinin üzerine çıkılana kadar ses sinyalinden çıkarılmaktadır. Bu eşik değerleri enerji için $TE=0.035$ ve sıfırı geçme sayıları $Tf=0.060$ olarak belirlenmiştir.



Şekil 7. 3 Aynı kişiye ait iki telaffuz örneği



Şekil 7. 4 Yukarıdaki iki telaffuzun önışlemeden sonraki halleri

Öznitelik Çıkarımı

Bu aşamada MFCC (Mel Frequency Cepstral Coefficient) lar elde edilerek öznitelik vektörü çıkarılmaktadır. MFCC nin aşamaları şunlardır.

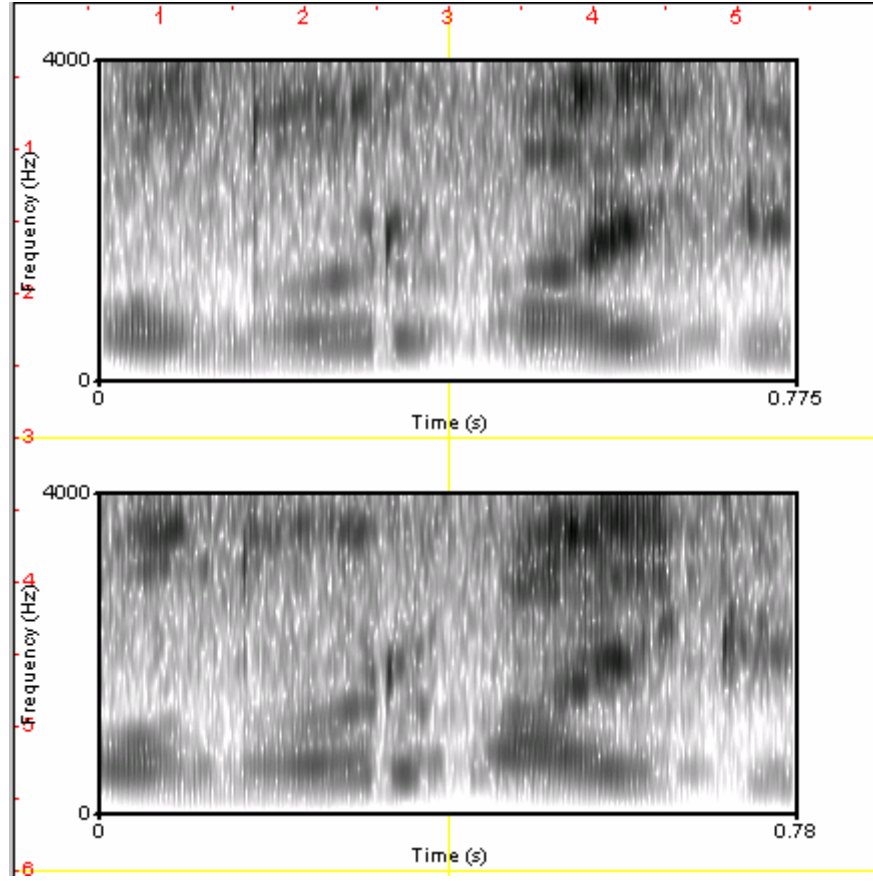
Öncelikle ses sinyaline aşağıdaki şekilde önvurgulama filtresi uygulanmaktadır.

$$H(z) = 1 - \alpha \cdot z^{-1}$$

Daha sonra sinyale kısa dönemli fourier dönüşümü uygulanır ve güç spektrogramı elde edilir.

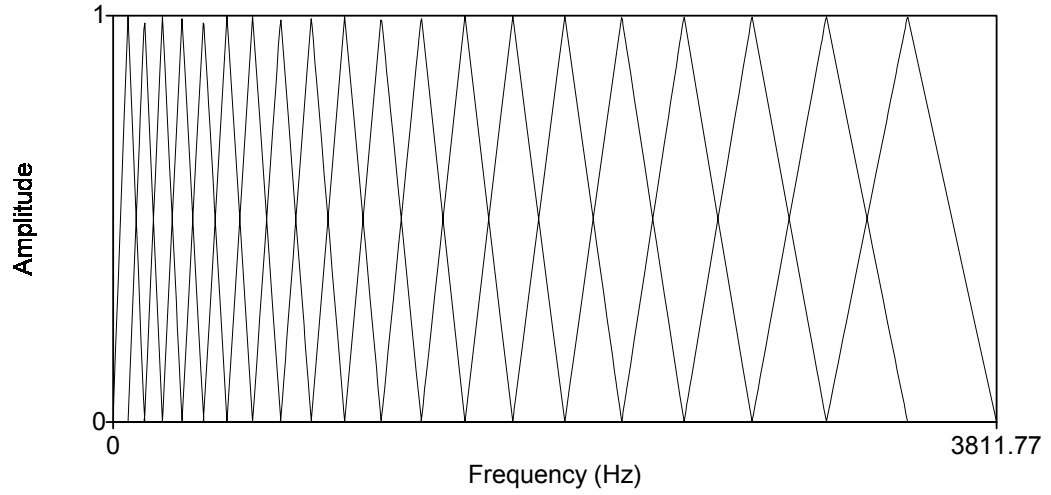
$$P(\omega) = |STFT(\omega(n) * s(n))|^2$$

Burada $w(n)$ konuşma çerçevesini $s(n)$ ise Pencere fonksiyonunu gösterir. Pencere fonksiyonu Hamming penceresi olarak belirlenmiştir.

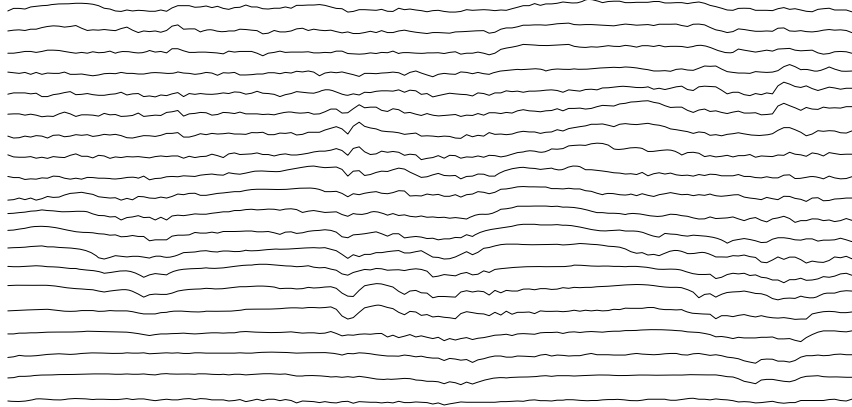
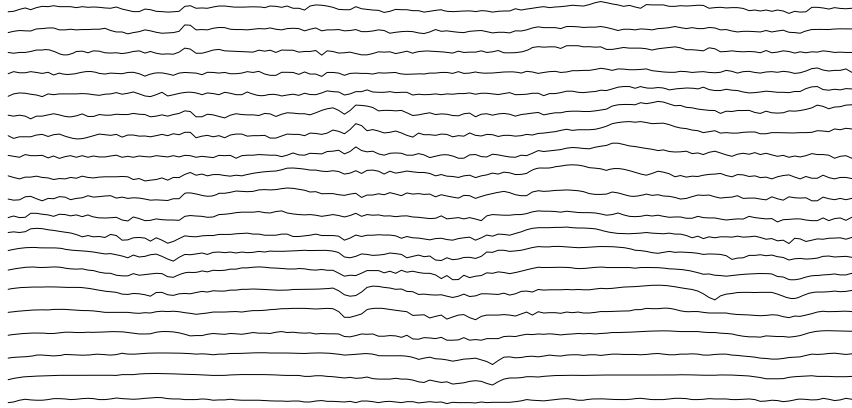


Şekil 7.5 İki telaffuzun spektrogram görüntüleri

Elde edilen güç spektrogramı $P(w)$, Mel-ölçek üçgen filtreleri ile eşuzaklıklı frekans alt-bantlarına dönüştürülür.



Şekil 7.6 Kullanılan Mel-Filtre Bankları



Şekil 7. 7 İki telaffuza ait filtre çıkışları

Bu frekans alt-bantların logaritmik enerjilerine, ayrık kosinüs dönüşümü uygulanarak MFCC katsayıları elde edilir.

$$c_i = \sum_{j=1}^F E_j \cdot \cos\left(\frac{\pi \cdot i}{L}(j-0.5)\right), \quad 0 \leq i \leq M$$

Buradaki M elde edilmek istenen katsayıların sayısıdır. Uygulamamızda 12 MFCC katsayısı elde edilmiştir. Aşağıda ilk beş çerçeve için 12 şer tane katsayı örnek olarak gösterilmiştir

1. -00.7991 00.3172 -00.4058 00.1039 -00.1600 00.0347 -00.3803 -00.1938 -00.2293 -00.0820 -00.0136 00.2009
2. -00.5568 00.5296 -00.5107 -00.0418 -00.2778 00.0967 -00.3319 00.1214 -00.2093 00.0439 00.0079 00.0824
3. -01.0044 00.0757 -00.1393 -00.1944 -00.2096 00.1042 -00.1498 00.2965 -00.0946 00.1022 00.0495 00.0977
4. -00.9859 -00.2312 -00.4586 00.3231 -00.3009 -00.1433 -00.2003 00.1977 -00.1794 00.1753 -00.0515 00.0237
5. -00.5666 00.2900 -00.4292 00.5126 -00.2452 -00.1386 -00.1820 00.0782 -00.0536 -00.0700 00.0562 -00.1375

MFCC katsayıları elde edildikten sonra önışleme ve gürültü temizleme işleme tabi tutulan ses sinyaliniz dalgacık dönüşümü işleme tabii tutulur. Dalgacık dönüşümünde Daubechies 4 dalgacığı seçilmiş ve 4 .seviye katsayılar elde edilmiştir. Elde edilen bu yaklaşım ve katsayılar birleştirilerek ikinci bir öznelik vektörü olarak değerlendirmeye alınır.

Değerlendirme

Bu aşamada her bir çerçeve için elde edilmiş olan katsayılar veritabanına kaydedilmiştir. Her kişi için beşer adet telaffuz alınmıştır. Alınan bu telaffuzlara ait MFCC ve Dalgacık katsayıları DTW algoritması kullanılarak birbirleriyle karşılaştırılmış ve her birinin birbirine olan uzaklığı hesaplanmıştır. Daha sonra ki test aşamasında bu uzaklıklar kullanılmaktadır.

Test aşamasında kullanıcıdan tekrar bir telaffuz alınır ve sırasıyla yukarıdaki işlemlerden geçirilir. Bu işlemler sonunda elde edilen katsayılar kullanılarak veritabanındaki katsayılarla karşılaştırılır. Eğer örneğimiz eğitim sırasında belirlenen uzaklıklardan en büyük değerde olanına %20 oranında yakın ise kişinin doğru kişi olduğu sonucuna varılır.

SONUÇ

Bu tezde konuşmacı tanıma sistemlerinin genel yapısı incelenmiş, metin bağımlı ve metin bağımsız sistemler için kullanılmakta olan yöntemler özelliklerine göre sınıflandırılmış ve her biri ayrı ayrı incelenmiştir. Çalışmadaki ilk amaç geliştirilecek uygulamanın konuşmacı tanıma kavramı içerisinde bulunan konuşmacı belirleme mi yoksa konuşmacı doğrulama mı olacağına karar vermektir. Sonuç olarak daha faydalı olacağı düşüncesiyle uygulamanın konuşmacı doğrulama sistemi olarak tasarlanması gerektiği düşünülmüştür. Daha sonraki aşama da ise metin bağımlı ve metin bağımsız sistem arasında bir tercih yapılması gerekmiştir. Gerek daha kolay gerçekleştirilebileceği gerekse de daha fazla başarı sağlayacağı görüldüğünden metin bağımlı bir tasarım tercih edilmiştir.

Metin bağımlı konuşmacı doğrulama uygulamasının gerçekleşmesi sırasında her bir işlem basamağı incelenmiş ve öznitelik vektörlerinin belirlenmesi aşamasında Mel-Frekans Ters-Spektrum Katsayılarının (MFCC) kullanılmasının diğer yöntemlere göre daha avantajlı olacağı görülmüştür. Bununla birlikte MFCC yöntemi ile karşılaştırma amacıyla dalgacık katsayılarının kullanıldığı bir model daha geliştirilmiştir. Modelleme ve karşılaştırma aşamasında ise DTW tabanlı bir teknik kullanılmıştır..

Sinyalin ön işleme aşamasında, gürültü temizleme için tezin araştırma konusu olan dalgacık dönüşümü uygulanmıştır. Dalgacık dönüşümü sonucu elde edilen detay katsayılar ve yaklaşım katsayıları literatürde dalgacık eşikleme yöntemi adı verilen bir yöntemle beyaz gürültüden arındırılmıştır.

Aşağıdaki sonuç tablosu

Bir kişi için alınan ses dalgalarına ait MFCC DTW uzaklık sonuçları Tablo 1'de, Dalgacık katsayılarına ait DTW uzaklık sonuçları Tablo 2'de gösterilmektedir.

Tablo 1

Kişi A	Örnek 1	Örnek 2	Örnek 3	Örnek 4	Örnek 5
Örnek 1	0	0.3297	0.8039	0.0348	0.9314
Örnek 2	0.3297	0	0.0172	0.2148	0.0170
Örnek 3	0.8039	0.0172	0	0.4257	0.0383
Örnek 4	0.0348	0.2148	0.4257	0	0.3257
Örnek 5	0.9314	0.0170	0.0383	0.3257	0

Tablo 2

Kişi A	Örnek 1	Örnek 2	Örnek 3	Örnek 4	Örnek 5
Örnek 1	0	0.8317	1.979	0.0648	1.335
Örnek 2	0.8317	0	0.7033	1.0421	0.8931
Örnek 3	1.9794	0.7033	0	1.6933	1.1719
Örnek 4	0.0648	1.0421	1.6933	0	1.1925
Örnek 5	0.9314	0.8931	1.1719	1.1925	0

Yapılan testler sonucunda öznitelik vektörü olarak MFCC katsayılarının seçilmesinin saf halde kullanılan dalgacık katsayılarından daha başarılı olduğu gözlenmiştir.

Genel itibariyle gerçekleştirilen sistem %70 lik bir başarı ile çalışmıştır. Yani hem doğru kişilerin reddedilmesinde hem de yanlış kişilerin kabul edilmesinde yüksek bir oran vardır. Bugünkü teknolojilerin ortalama %95 oranında bir başarı elde ettiği göz önüne alınırsa gerçekleştirilen sistemin başarısız olduğu ortaya çıkmaktadır. Bu yüzden diğer sistemlerle ve tekniklerle bir karşılaştırma yapılmamıştır.

Bundan sonraki çalışmalarda, bu tezde yalın olarak kullanılan dalgacık katsayıları ile MFCC ve LPC benzeri yöntemler geliştirilmeye çalışılacaktır. Bu işlemler yapılırken , bilinen tüm dalgacıklar geliştirilen bir karşılaştırma yöntemiyle incelenerek ses sinyalleri için uygun olan bir dalgacık belirlenecektir. Bunun nedeni her bir dalgacığın şeklinin diğerinden farklı olması ve bu şekil farklılıklarının incelenen sinyallere göre farklı sonuçlar doğurmasıdır. Aynı zamanda bu tezde şablon eşleme modeli olarak kullanılan DTW yöntemindeki eksiklikler incelenecek ve model üzerinde bazı iyileştirmeler yapılacaktır.

KAYNAKLAR

Rabiner, L.R., Juang, B.H., Lee C.H., An Overview of Automatic Speech Recognition, Automatic Speech and Speaker Recognition, Advanced Topics :1-30. Kluwer Academic Publishers, 1996

Furui, S., An Overview of Speaker Recognition Technology, Automatic Speech and Speaker Recognition: 31-56, Boston: Kluwer Academic, 1996

Markowitz, J. The kinds of errors speaker verification and other biometric systems make, VoiceId Quarterly Vol:6 Issue 4, 2006

Yun, Y.W. "123 of Biometric Technology", Synthesis Journal 2002, <http://www.itsc.org.sg/synthesis/2002/biometric.pdf>

Çelik Z.Ö., Sözlü Anlatım (Konuşma), "Sözlü ve Yazılı Anlatım", Ünite 3 Anadolu Üniversitesi Yayınları, 2004

Mengüşoğlu, E., 1999 "Bir Türkçe Sesli İfade Tanıma Sisteminin Kural Tabanlı Tasarımı ve Gerçekleştirimi", Master Tezi, Hacettepe Üniversitesi

Joseph W. P. Signal modeling techniques in speech recognition. Proceedings of the IEEE, Vol. 81, No. 9, pages 1215--1247, 1993.

Mengüşoğlu, E., Confidence Measures for Speech/Speaker Recognition and Applications on Turkish LVCSR, Doktora Tezi, Faculte Polytechnique de Mons, 2002

Gersho A., "On the Structure of Vector Quantizers," *IEEE Transactions on Information Theory*, vol. 28, no. 2, pp. 157-166, 1982.

Rabiner, L. R., A tutorial on Hidden Markov Models and Selected Applications in Speech Recognition. Proceedings of IEEE, vol. 77, no. 2, pp. 257-286., 1989

Aydın Ö., Yapay Sinir Ağlarını Kullanarak Bir Ses Tanıma Sisteminin Gerçekleştirilmesi, Yüksek Lisans Tezi, Trakya Üniversitesi, 2005

Ertürk S Sayısal İşaret İşleme, Birsen Yayınevi, İstanbul, ISBN 975-511-309-6, 1996

Graps A. ,An Introduction to Wavelet Analysis, IEEE Computational Science and Engineering, 1995, Vol. 2, Num. 2

Louis, P. Maaß, Reider A, "Wavelets, Theory and Applications", John Wiley & Sons, Chichester, A. K., 1997

Daubechies I., Ten Lectures on Wavelets, SIAM, Philadelphia, 1992

Nyquist, H., "Certain Topics in Telegraph Transmission Theory", 1928

Lu C.T., Wang H.C., Enhancement of single channel speech based on masking property and wavelet transform, *Speech Communication*, 41 409–427, 2003

Erçelebi E. Second generation wavelet transform-based pitch period estimation and voiced/unvoiced decision for speech signals, *Applied Acoustics* 64 25–41, 2003

Obaidat M.S., Lee C., Estimation of pitch period of speech signal using a new dyadic wavelet algorithm, *Information Sciences* 119 21-39, 1999

Namba M., Ishida Y., Wavelet transform domain blind deconvolution, *Signal Processing* 68 119-124, 1998

Sing R., Davis K. Hidden Markov model-based speech recognition with intermediate wavelet transform domains, *Computer Speech and Language* 11, 253–273, 1997

Lung S.Y. Wavelet feature selection based neural networks with application to the text independent speaker identification, *Pattern Recognition* 39 1518 – 1521, 2006

Lardies J., Identification of a dynamical model for an acoustic enclosure using the wavelet transform, *Applied Acoustics* 2006