

T.C.

GEBZE YÜKSEK TEKNOLOJİ ENSTİTÜSÜ
MÜHENDİSLİK VE FEN BİLİMLERİ ENSTİTÜSÜ

SES İŞARETLERİNİN ANALİZİNDE
KISA ZAMANLI FOURIER VE DALGACIK
DÖNÜŞÜMLERİNİN UYGULAMALARI

12 72 84

Birol KESKİN

YÜKSEK LİSANS TEZİ

ELEKTRONİK MÜHENDİSLİĞİ

ANABİLİM DALI

T.C. YÜKSEKÖĞRETİM KURULU
DOKÜMANTASYON MERKEZİ

TEZ DANIŞMANI

Yrd. Doç. Dr. Şeref Naci ENGİN

12 72 84

GEBZE

2002

...Binal...KESKİN.....'in tez çalışması, G.Y.T.E. Mühendislik ve Fen Bilimleri Enstitüsü Yönetim Kurulu'nun 20.08.2002 tarih ve 2002/26 sayılı kararıyla oluşturulan jüri tarafından Elektronik Müh. Anabilim Dalında YÜKSEK LİSANS tezi olarak kabul edilmiştir.

JÜRİ

ÜYE
(Tez Danışmanı)

: Yrd. Doç. Dr. Ş. Naci ENGİN



ÜYE

: Doç. Dr. E. Murat ESİN



ÜYE

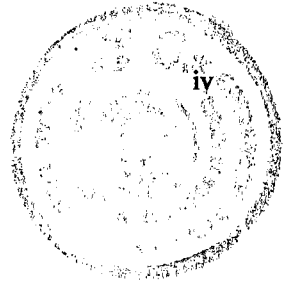
: Yrd. Doç. Dr. Oğuz KUCUR



ONAY

G.Y.T.E Mühendislik ve Fen Bilimleri Enstitüsü Yönetim Kurulu'nun 28.11.2002 tarih ve 2002/40 sayılı kararı.

İmza / Mühür

ÖZET

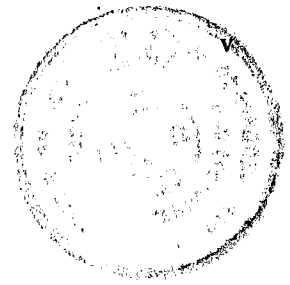
Bu çalışmada frekans, zaman-frekans ve zaman-ölçek temelli işaret analiz yöntemleri incelenmiş ve konuşma işaretlerinin tanınmasında kullanılmıştır.

Bilindiği gibi, en temel dönüşüm yöntemlerinden biri olan Fourier dönüşümü uygulandığı işaretin sadece frekans bilgisini vermekte, zaman bilgisine yer vermemektedir. Ancak konuşma işaretleri durağan işaretler değildir, frekans içerikleri zamanla değişir. Bu nedenle, frekansın yanı sıra zaman bilgisine de yer veren Kısa Zamanlı Fourier Dönüşümleri (KZFD) konuşma işaretlerini incelemek için daha uygun bir yöntemdir. Bu yöntemde, küçük bir aralıkta işaretin durağan olduğu kabul edilir ve işaret küçük pencerelere bölünerek incelenir. Ne var ki, KZFD’nde analiz boyunca pencere uzunluğu sabit olduğu için arzu edilen zaman-frekans çözünürlükleri elde edilememektedir.

KZFD’nün bu dezavantajından kurtulmak amacıyla Dalgacık Dönüşümüne başvurulmuştur. Dalgacık dönüşümünde pencere uzunluğu analiz boyunca kendiliğinden değiştiği için istenilen zaman-frekans çözünürlükleri elde edilebilir. Bu nedenle dalgacık dönüşümü, frekans içeriği zamanla oldukça değişken bir yapıya sahip konuşma işaretlerini incelemek için daha uygun bir yöntemdir.

Kısa zamanlı Fourier ve dalgacık dönüşümlerinden yararlanılarak 0 ile 9 arasındaki rakamların ses kayıtlarından elde edilen konuşma işaretlerinin ayırt edici özellikleri çıkartılmaya çalışılmıştır. Elde edilen özellikler ses tanıma algoritmalarına verilerek konuşmacı-bağımlı ve konuşmacı-bağımsız rakam tanıma deneyleri yapılmıştır.

Bu çalışma, KZFD ile çok başarılı ses tanıma sonuçları alındığını göstermiştir. Bununla birlikte dalgacık dönüşümü konuşma işaretlerinin analizi için etkili bir yöntem olduğunu kanıtlamış ve ses tanıma deneylerinde önemli katkılar sağlamıştır.



SUMMARY

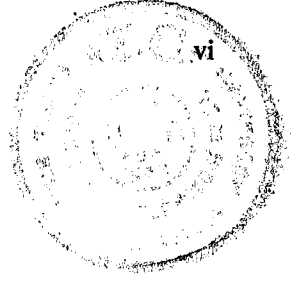
In this work speech signals are studied with frequency, time-frequency and time-scale based analysis methods and some speech recognition experiments are carried out by using these methods.

The Fourier transform, which is the basic transformation method, gives only the frequency information of the signal. However, speech signals are non-stationary signals and their frequency properties change over the time. Therefore STFT (Short-Time Fourier Transform), which also provides time information, is a more suitable technique in the analysis of the speech signals. But in STFT the length of the window is the same over the time and it cannot produce desired time-frequency resolutions.

In order to correct this deficiency of STFT, Wavelet Transform is used. The size of the window during the analysis changes and hence it could yield better time-frequency resolutions. This advantage makes the wavelet transform a more suitable technique for analyzing the speech signals, which present quite non-stationary characteristics.

By using the short-time Fourier and wavelet transforms feature extraction is done for the speech signals of sound records of digits from 0 to 9. The features obtained from these methods are fed into the speech recognition algorithms for digit recognition.

This work shows that very successful speech recognition results have been obtained by means of STFT. Besides this, the wavelet transform has proven itself as an effective method for the analysis of speech signals and provided significant contributions in the speech recognition experiments.

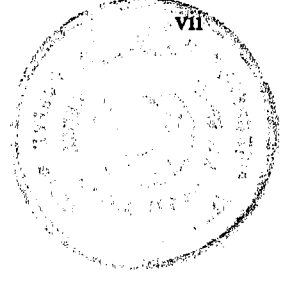


TEŞEKKÜR

Tez çalışmamda bana yol gösteren, gerekli ilgi ve desteği esirgemeyen tez danışmanım Yıldız Teknik Üniversitesi, Elektrik-Elektronik Fakültesi Elektrik Mühendisliği Bölümü tam zamanlı ve GYTE Mühendislik Fakültesi Elektronik Mühendisliği Bölümü kısmi zamanlı öğretim üyesi hocam Sayın Yrd. Doç. Dr. Şeref Naci ENGİN'e, çalışmalarımda bana yardımcı olan tüm mesai arkadaşlarıma, özellikle Gürhan Günce GÜRSEL'e, çalışmalarım için gerekli izni sağlayan kurumum ASELSAN A.Ş.'ye ve çalışmalarımı gerçekleştirdiğim GYTE'nün tüm akademik ve idari personeline katkılarından dolayı teşekkürü borç bilirim.

Tez çalışmam süresince bana destek olan sevgili arkadaşım Müjgan ESİM'e, bana göstermiş olduğu hoşgörü, anlayış ve destek için sonsuz teşekkür ederim.

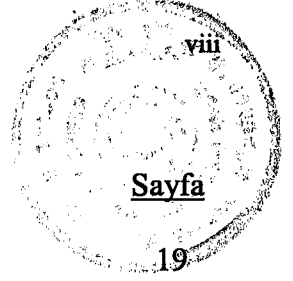
Ayrıca haklarını hiçbir zaman ödeyemeyeceğim, benden maddi ve manevi desteklerini esirgemeyen sevgili aileme teşekkür eder, bu çalışmayı onlara armağan ederim.



İÇİNDEKİLER DİZİNİ

Sayfa

ÖZET	iv
SUMMARY	v
TEŞEKKÜR	vi
İÇİNDEKİLER DİZİNİ	vii
SİMGELER VE KISALTMALAR DİZİNİ	ix
ŞEKİLLER DİZİNİ	xi
ÇİZELGELER DİZİNİ	xiv
1 GİRİŞ	1
1.1 Konu ve Önemi	1
1.2 Tezin Amacı ve İçeriği	2
1.3 Ses ve Konuşma İşaretleri	4
1.4 Sesin Fiziksel Özellikleri	4
1.4.1 Ses Şiddeti	4
1.4.2 Ses Desibel Seviyeleri	5
1.4.3 Ses Yüksekliği	6
1.5 Sayısal İşaret İşleme	7
2 SES TANIMA	10
2.1 Giriş	10
2.2 Ses Tanımanın Tarihçesi	11
2.3 Ses Tanıma Mimarisi	12
2.4 Örüntü Tanıma	14
2.5 Dinamik Zaman Sıkıştırması	16
3 FOURIER ANALİZİ	19
3.1 Giriş	19



3.2	Baz Fonksiyonları	19
3.3	Fourier Serileri	21
3.4	Fourier Dönüşümü	22
3.5	Ayrık Fourier Dönüşümü	23
3.6	Hızlı Fourier Dönüşümü	24
3.7	Kısa Zamanlı Fourier Dönüşümü (KZFD)	25
4	DALGACIK ANALİZİ	31
4.1	Giriş	31
4.2	Dalgacıklar	31
4.3	Sürekli Dalgacık Dönüşümü	40
4.3.1	Giriş	40
4.3.2	Ölçekleme (Dilation, Scaling) ve Öteleme (Translation)	42
4.3.3	Sürekli Dalgacık Katsayılarının Hesaplanması	45
4.3.4	Konuşma İşaretlerinin Sürekli Dalgacık Dönüşümü ile İncelenmesi	47
4.4	Ayrık Dalgacık Dönüşümü	52
4.4.1	Giriş	52
4.4.2	Ölçek ve Dalgacık Fonksiyonları	53
4.4.3	Alt-Bantlara Ayırma ve Ayrık Dalgacık Analizi	57
4.4.4	Ayrık Dalgacık Sentezi	61
4.4.5	Ayrık Dalgacık Katsayılarının Hesaplanması	63
4.4.6	Konuşma İşaretlerinin Ayrık Dalgacık Dönüşümü ile İncelenmesi	66
5	DENEY SONUÇLARI	72
5.1	Konuşma İşaretlerinin Elde Edilmesi	72
5.2	Örüntü Matrisinin Elde Edilmesi	72
5.3	Konuşmacı-Bağımlı Ses Tanıma Deney Sonuçları	73
5.4	Konuşmacı-Bağımsız Ses Tanıma Deney Sonuçları	79
6	SONUÇLAR	81
	KAYNAKLAR	83
	ÖZGEÇMİŞ	85



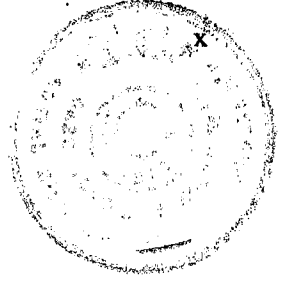
SİMGELER VE KISALTMALAR DİZİNİ

KZFD	Kısa Zamanlı Fourier Dönüşümü
Hz	Hertz
kHz	Kilohertz
W	Watt
m	Metre
N	Newton
dB	Decibel
s	saniye
ARPA	Advanced Research Projects Agency (İleri Araştırma Projeleri Ajansı)
ms	Milisaniye
AFD	Ayrık Fourier Dönüşüm
HFD	Hızlı Fourier Dönüşümü

I	Ses şiddeti (W/m^2)
p	Ses basıncı (N/m^2)
c	Ses hızı (m/s)
ρ_0	Ortam yoğunluğu (Ns/m^3)
L	Ses enerji seviyesi (dB)
S	Ses yüksekliği (Sone)
$u(i,j)$	İki özellik vektörü arasındaki uzaklık
U_n	İki örüntü arasındaki toplam uzaklık
$c_{T,R}$	Test ve referans örüntüleri arasındaki minimum uzaklık
$\overline{f(t)}$	$f(t)$ fonksiyonunun kompleks eşleniği

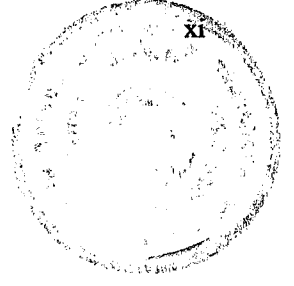
$$\langle f(t), g(t) \rangle \quad \text{iç çarpım} \quad = \int_{-\infty}^{\infty} f(t) \overline{g(t)} dt$$

$$\hat{f}(t) \quad f(t) \text{ fonksiyonunun ikili (dual)}i$$



$$\|f(t)\| \quad f(t) \text{ fonksiyonunun normu} \quad = \sqrt{\int_{-\infty}^{\infty} |f(t)|^2 dx}$$

$F(\omega)$	$f(t)$ fonksiyonunun Fourier dönüşümü
$w(t)$	Pencere fonksiyonu
$G(t, \omega)$	Kısa zamanlı Fourier dönüşümü
Δt	Zaman penceresi
$\Delta \omega$	Frekans penceresi
$\Psi(t)$	Ana dalgacık fonksiyonu
b	Öteleme
a	Ölçek
$\phi(t)$	Ölçek fonksiyonu
$\psi(t)$	Dalgacık fonksiyonu
$c(k)$	Ayrık dalgacık dönüşümü yaklaşıklık katsayıları
$d(k)$	Ayrık dalgacık dönüşümü ayrıntı katsayıları
$h_0(n)$	Ayrık dalgacık dönüşümü alçak-geçiren analiz filtresi
$h_1(n)$	Ayrık dalgacık dönüşümü yüksek-geçiren analiz filtresi
$g_0(n)$	Ayrık dalgacık dönüşümü alçak-geçiren sentez filtresi
$g_1(n)$	Ayrık dalgacık dönüşümü yüksek-geçiren sentez filtresi

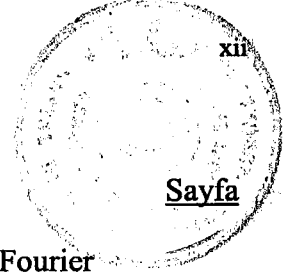


ŞEKİLLER DİZİNİ

Sekil

Sayfa

1.1 “Dokuz” kelimesinin aynı kişiye ait iki farklı kayıt örneğinin (a) işlenmemiş halleri, (b) Fourier dönüşümleri.	8
2.1 Ses tanıma mimarisi.	13
2.2 Örüntü tanıma blok diyagramı.	14
2.3 Konuşma işaretlerinin pencerelenmesi ve örüntü matrisinin elde edilmesi.	15
2.4 Dinamik zaman sıkıştırması grafiği.	17
3.1 50 Hz ve 200 Hz frekans bileşenlerine sahip bir işaretin (a) işlenmemiş hali ve (b) Fourier dönüşümü.	23
3.2 “Yedi” kelimesinin iki farklı örneğinin (a) işlenmemiş halleri, (b) Fourier dönüşümleri ve (c) KZFD’leri.	27
3.3 KZFD’nde zaman-frekans penceresi.	28
3.4 Matematiksel olarak üretilmiş bir işaretin (a) işlenmemiş hali, (b) Fourier dönüşümü, ve (c) 1.6 ms, (d) 3 ms, (e) 6 ms pencere uzunlukları ile KZFD grafikleri.	30
4.1 (a) Sinüs dalgası ve (b) Daubechies dalgacığı.	32
4.2 (a) Kare dalgacığının ölçek fonksiyonu, (b) dalgacık fonksiyonu.	35
4.3 (a) db4 (Daubechies) dalgacığının ölçek fonksiyonu, (b) dalgacık fonk.	35
4.4 (a) bior5.5 (biortogonal) dalgacığının ölçek fonksiyonu, (b) dalgacık fonksiyonu.	36
4.5 (a) coif3 (coiflet) dalgacığının ölçek fonksiyonu, (b) dalgacık fonksiyonu.	36
4.6 (a) sym8 (simettrili) dalgacığının ölçek fonksiyonu, (b) dalgacık fonk.	37
4.7 Morlet dalgacık fonksiyonu.	37
4.8 Meksika şapkası dalgacık fonksiyonu.	38
4.9 (a) Meyer dalgacığının ölçek fonksiyonu, (b) dalgacık fonksiyonu.	38
4.10 Dalgacık dönüşümde zaman-frekans penceresi.	40
4.11 (a) İşlenmemiş işaret, (b) Fourier dönüşümü, (c) KZFD, (d) dalgacık dönüşümü için zaman-frekans çözünürlükleri.	41
4.12 (a) $a=16$, (b) $a=4$, (c) $a=1$ ölçek değerleri için Daubechies dalgacıkları.	42



Sekil

- 4.13 (a) $a=8.5$ ölçek değeri için Morlet dalgacık fonksiyonu ve (b) Fourier dönüşümü. 43
- 4.14 (a) 500 Hz ve 3 kHz frekans bileşenlerine sahip bir işaretin işlenmemiş hali, (b) Morlet dalgacı ve (c) Meksika şapkası dalgacı ile dalgacık dönüşümleri. 44
- 4.15 (a) $b=0.4$, (b) $b=0.8$, (c) $b=1.2$ öteleme değerleri ile db12 (Daubechies) dalgacı. 44
- 4.16 (a) “Altı” kelimesinin işlenmemiş hali, (b) sürekli dalgacık dönüşümü. 46
- 4.17 (a) “Altı” kelimesinin işlenmemiş hali, (b) Fourier dönüşümü, (c) KZFD ve (d) sürekli dalgacık dönüşümü. 48
- 4.18 (a) “Altı” kelimesinin işlenmemiş hali ve (b) 4 ms, (c) 8 ms, (d) 16 ms, (e) 32 ms pencere uzunlukları ile KZFD’leri. 49
- 4.19 (a) “Altı” kelimesinin işlenmemiş hali, (b) 300-400 ms aralığının işlenmemiş hali, (c) Fourier dönüşümü, (d) sürekli dalgacık dönüşümü ve (e) 1.5 ms, (f) 3 ms, (g) 6 ms pencere uzunlukları ile KZFD’leri. 51
- 4.20 Ayırık dalgacık dönüşümünde ölçek ve zaman eksenlerinin örneklenmesi. 53
- 4.21 (a) Morlet dalgacı ve (b) Fourier dönüşümü. 53
- 4.22 Ayırık dalgacık dönüşümünde ilk 4 seviye için dalgacık fonksiyonu ile frekans bandının kaplanması. 54
- 4.23 (a) db13 (Daubechies) dalgacığının ölçek fonksiyonu ve (b) Fourier dönüşümü. 55
- 4.24 Ayırık dalgacık dönüşümünde ilk 3 seviye için ölçek ve dalgacık fonksiyonları ile frekans bandının kaplanması. 55
- 4.25 (a) db13 (Daubechies) dalgacığının alçak geçiren ve (b) yüksek geçiren analiz filtreleri. 58
- 4.26 Aşağı örnekleme (Örnekleme frekansının yarıya düşürülmesi). 59
- 4.27 İki seviyeli ayırık dalgacık dönüşümü. 60
- 4.28 3 seviyeli ayırık dalgacık dönüşümünü ağacı ve frekans bantları. 60
- 4.29 (a) db13 (Daubechies) dalgacığının alçak geçiren ve (b) yüksek geçiren sentez filtreleri. 61
- 4.30 İki seviyeli ters ayırık dönüşümü ağacı. 62

Şekil

4.31 Ölçek fonksiyonu, dalgacık fonksiyonu, alçak-geçiren analiz filtresi ve yüksek-geçiren analiz filtresi.	63
4.32 İşaretin alçak-geçiren filtreden geçirilmesi.	63
4.33 Alçak-geçiren filtreden geçirilen işaretin aşağı-örneklenmesi.	63
4.34 İşaretin yüksek-geçiren filtreden geçirilmesi.	64
4.35 Yüksek-geçiren filtreden geçirilen işaretin aşağı örneklenmesi.	64
4.36 c_1 katsayılarının alçak-geçiren filtreden geçirilmesi.	64
4.37 Alçak-geçiren filtreden geçirilen c_1 katsayılarının aşağı-örneklenmesi.	65
4.38 c_1 katsayılarının yüksek-geçiren filtreden geçirilmesi.	65
4.39 Yüksek-geçiren filtreden geçirilen c_1 katsayılarının aşağı-örneklenmesi.	65
4.40 (a) "Dokuz" kelimesinin işlenmemiş hali ve (b) Fourier dönüşümü.	66
4.41 "Dokuz" kelimesinin işlenmemiş hali, ve 7. seviyeye kadar yaklaşıklık ve ayrıntı katsayıları.	67
4.42 "Dokuz" kelimesinin ayrık dalgacık katsayılarının elde edilmesi.	69
4.43 "Dokuz" kelimesine ait ayrık dalgacık analizi ağacı.	70
4.44 "Dokuz" kelimesinin işlenmemiş hali ve 7. seviyeye kadar ayrık dalgacık analizi ($f \approx c_7 + d_7 + d_6 + d_5 + d_4 + d_3 + d_2 + d_1$).	71

ÇİZELGELER DİZİNİ

<u>Çizelge</u>	<u>Sayfa</u>
1.1 Bazı ses kaynaklarının seviyeleri.	5
4.1 Morlet dalgacığın için ölçek-frekans değerleri karşılıkları.	47
4.2 11.025 kHz örnekleme frekansı kullanıldığında alt-bantlara ayırma ile elde edilen frekans bantları.	66
5.1 Birinci grup veri üzerinde KZFD ile konuşmacı-bağımlı ses tanıma sonuçları.	74
5.2 Birinci grup veri üzerinde 1. seviye ayrık dalgacık dönüşümü ile konuşmacı-bağımlı ses tanıma sonuçları.	76
5.3 Birinci grup veri üzerinde 2. seviye ayrık dalgacık dönüşümü ile konuşmacı-bağımlı ses tanıma sonuçları.	77
5.4 Birinci grup veri üzerinde 3. seviye ayrık dalgacık dönüşümü ile konuşmacı-bağımlı ses tanıma sonuçları.	77
5.5 Birinci grup veri üzerinde değişik seviyeler için ayrık dalgacık dönüşümü ile konuşmacı-bağımlı ses tanıma sonuçları.	78
5.6 İkinci grup veri üzerinde KZFD ile konuşmacı-bağımsız ses tanıma sonuçları (eğitme birinci grup veri ile).	79
5.7 İkinci grup veri üzerinde ayrık dalgacık dönüşümü ile konuşmacı-bağımsız ses tanıma sonuçları (eğitme birinci grup veri ile).	79
5.8 İkinci grup veri üzerinde KZFD ile konuşmacı-bağımsız ses tanıma sonuçları (eğitme ikinci grup veri ile).	80
5.9 İkinci grup veri üzerinde ayrık dalgacık dönüşümü ile konuşmacı-bağımsız ses tanıma sonuçları (eğitme ikinci grup veri ile).	80

1 GİRİŞ

1.1 Konu ve Önemi

Teknolojinin ilerlemesi ile birlikte sinyal analizi mühendislik çalışmalarında önemli bir yer tutmuştur. Analiz edilebilecek çok çeşitli sinyaller vardır. Sismik dalgalar, makine titreşimleri, tıbbi görüntüler ve sinyaller, müzik işaretleri ve konuşma işaretleri bunlardan birkaçıdır. Genellikle işaretler zamanın bir fonksiyonu olarak gösterilir. İşaretin bu haline *işlenmemiş* (ham) işaret denilmektedir. Fakat işlenmemiş işaret çoğu zaman yeterli bilgi vermez. Bu nedenle işaretler üzerinde çeşitli dönüşümler uygulanır.

Fourier analizi dönüşümler arasında en temel olanıdır, işaretleri değişik frekanslı sinüs bileşenlerinin toplamı şeklinde gösterir ve işaretin frekans spektrumunu (tayfını) verir. Fakat dönüşüm sonucunda zaman bilgisi kaybedilmektedir. Bu nedenle durağan olmayan işaretler için Fourier dönüşümü uygun bir yöntem değildir. Örneğin anlık değişimleri, başlangıç ve bitiş zamanlarını, kısa süreli tepeleri tespit edemeyecektir. Fourier dönüşümünün bu eksikliğini gidermek için araştırmacılarca değişik yöntemler geliştirilmiştir. KZFD (*Kısa Zamanlı Fourier Dönüşümü*) ve *dalgacık dönüşümü* (wavelet transform) bu yöntemlerdendir ve hem zaman hem de frekans bilgisine yer vermektedir.

KZFD’nde işaret pencerelere bölünerek Fourier dönüşümleri elde edilir. Seçilen bu pencere kaydırılarak işaretin parça parça tüm zaman eksenı boyunca Fourier dönüşümü alınır. Pencerenin konumu zaman bilgisini, hesaplanan Fourier katsayıları ise frekans bilgisini verir. Fakat KZFD’nde analiz boyunca pencere büyüklüğü sabittir, istenilen zaman-frekans çözünürlüğü elde edilememektedir. Zaman çözünürlüğünü arttırmak için pencere genişliği azaltıldığında frekans çözünürlüğü kötüleşir. Tersı durumda pencere genişliği arttırıldığında frekans çözünürlüğü iyileşirken, zaman çözünürlüğü kötüleşir.

Dalgacık dönüşümünde pencere genişliği analiz sırasında kendiliğinden değişerek istenilen zaman-frekans çözünürlüğünü verir. Dalgacık dönüşümü işaretleri değişik genişlikteki ve ölçekteki dalgacık fonksiyonlarının toplamı şeklinde

gösterir. Küçük ölçekli dalgacıklar yüksek frekanslı bileşenleri (hızlı değişimleri) dar pencereler ile, büyük ölçekli dalgacıklar ise alçak frekanslı bileşenleri (yavaş değişimleri) geniş pencereler ile analiz ederek istenilen zaman-frekans çözünürlüğünü verir.

Dalgacık dönüşümü bu avantajı nedeniyle sinyal analizinde kullanılmak için çok uygun bir araçtır. Dalgacık dönüşümü ile değişken yapıdaki konuşma işaretleri analiz edilebilir ve ayırt edici özellikleri çıkartılabilir.

1.2 Tezin Amacı ve İçeriği

Tez çalışmasında Fourier, kısa zamanlı Fourier ve dalgacık dönüşümleri incelenmiş ve birbirleriyle karşılaştırılmıştır. 0-9 arasındaki rakamlardan oluşan konuşma işaretleri bu dönüşümlerle analiz edilerek ayırt edici özellikleri çıkartılmaya çalışılmıştır. Son olarak da elde edilen özellikler ses tanıma algoritmalarına verilerek en uygun yöntemin hangisi olduğu araştırılmıştır.

Bölüm 1’de ses ve konuşma işaretlerinin frekans içeriği, enerji seviyeleri, şiddet ve yükseklikleri gibi genel özellikleri anlatılmıştır. Sayısal işaret işleme yöntemleri ana hatları ile açıklanmış, tez çalışmasında kullanılan sinyal işleme yöntemleri ve bu yöntemlerin nerelerde, nasıl kullanıldığı verilmiştir.

Bölüm 2’de ses tanıma hakkında bilgi verilmiştir. Ses tanımanın tarihçesi, ses tanıma mimarisi, hangi bileşenlerden oluştuğu ve kullanılan yöntemler anlatılmıştır. Çalışmada ses tanıma yöntemi olarak kullanılan örüntü tanıma (pattern recognition) yöntemi anlatılmıştır.

Bölüm 3’te analiz yöntemlerine giriş yapılmış, ve baz fonksiyonları, ortogonal ve biortogonal baz kümeleri hakkında bilgi verilmiştir. Fourier analizi ve KZFD anlatılmıştır. Fourier dönüşümünün neden yetersiz olduğu ve KZFD’nün Fourier dönüşümüne göre avantajları sunulmuştur.

Bölüm 4’te dalgacık analizi anlatılmıştır. Dalgacık fonksiyonları ve bir fonksiyonun dalgacık olabilmesi için sağlaması gereken koşullar verilmiştir. Sürekli dalgacık dönüşümü ve ayrık dalgacık dönüşümü anlatılmış, dalgacık dönüşümünün diğerlerine göre üstünlükleri konuşma işaretlerini analiz ederek sunulmuştur. Ayrık

dalgacık dönüşümünün temelini oluşturan alt-bantlara ayırma (subband coding) yöntemi anlatılmıştır.

Bölüm 5'te yapılan ses tanıma deneylerinin sonuçları verilmiştir. İlk olarak deneylerde kullanılan işaretlerin nasıl toplandığı anlatılmıştır. Daha sonra KZFD ve dalgacık dönüşümleri ile bu işaretlerin nasıl işlendiği ve özellik vektörlerinin nasıl oluşturularak örüntü tanıma yöntemine verildiği açıklanmıştır. Elde edilen ses tanıma deney sonuçları karşılaştırmalı olarak yorumlanmıştır.

Bölüm 6'ta bu tez çalışmasından alınan genel sonuçlar sunulmuş ve buna göre gelecekte yapılacak çalışmalar için önerilerde bulunulmuştur.



1.3 Ses ve Konuşma İşaretleri

İnsan kulağının duyabileceği her hangi bir titreşime *ses* denir. İnsan kulağı yaklaşık 20 Hz - 20 kHz frekans aralığındaki sesleri işitebilmektedir ve bu frekans aralığı *işitme frekans* (audio) aralığı olarak adlandırılır [Truax,1999].

Konuşma ise dilbilgisi içeren ses işaretlerinden oluşur. Konuşma işaretlerinin frekans aralığı 10 kHz'e kadar çıkabilir. Bu çalışmada kullanılan konuşma işaretleri incelendiğinde, ses işleme ile ilgili kaynakların da belirttiği gibi 4 kHz'in üstündeki frekans bileşenlerinin büyüklüklerinin çok düşük kaldığı görülmüştür. [Fallside and Woods,1985], [Rabiner and Schafer,1978]. Buna göre en yüksek frekans bileşeninin 4 kHz olduğu kabul edilirse Nyquist Teoremi'ne göre örnekleme frekansı en az 8 kHz olmalıdır [Denbigh,1998]. Bu örnekleme oranı telefon kalitesidir. Daha yüksek ses kalitesi gerektiren sayısal ses kayıtları ise 44.1 kHz (CD kalitesi) ya da 44.8 kHz (Sayısal Teyp Kayıt kalitesi) örnekleme frekansı ile yapılmaktadır [Truax,1999]. Bu çalışmada kullanılan konuşma işaretlerin elde edilmesinde seçilen örnekleme frekansı ve kodlama gibi sayısal işaret işleme yöntemleri Bölüm 1.5'te sunulmuştur.

1.4 Sesin Fiziksel Özellikleri

1.4.1 Ses Şiddeti

Ses şiddeti, bir saniyede birim alandan sesin yaydığı enerji miktarıdır ve aşağıdaki gibi tanımlanır [Gold and Morgan,1999]:

$$I = \frac{\overline{p^2}}{\rho_0 c} \quad (1.1)$$

p ses basıncı, c ses hızı ve ρ_0 ortam yoğunluğudur. $\rho_0 c$ çarpımı karakteristik empedansı vermektedir. Görüldüğü gibi ses şiddeti, basıncın karesi ile doğru orantılıdır.

1.4.2 Ses Desibel Seviyeleri

Ses seviyeleri, referans bir seviyeye göre ölçülen ses enerji değerleridir ve desibel cinsinden verilir.

1 kHz frekans için duyma eşiği 10^{-12} W/m² olarak tanımlanır [Gold and Morgan,1999]. Bu duyma eşiği referans alınarak ses enerji seviyesi aşağıdaki gibi verilir:

$$L = 10 \log \left| \frac{I_1}{I_2} \right| \quad (1.2)$$

Denklem (1.2) ile verilen ses enerji seviyesi ifadesine duyma eşiği değeri yerleştirilirse, dB cinsinden duyma eşiği 0 dB olacaktır.

Ses enerji yoğunluğunun basıncın karesi ile orantılı olduğu hatırlanırsa ses seviyesi basınç cinsinden de tanımlanabilir:

$$L = 20 \log \left| \frac{p_1}{p_2} \right| \quad (1.3)$$

Referans ses basıncı $p_2=2 \times 10^{-5}$ N/m² olarak tanımlanırsa ses basınç seviyesi bulunabilir. Bazı ses kaynaklarının ses seviyeleri Çizelge 1.1'de verilmiştir [Truax,1999]:

Duyma eşiği	0 dB	Motorsiklet (9.1 m)	88 dB
Yaprak hışırtısı	20 dB	Mikser (0.9 m)	90 dB
Fısıltı (0.9 m)	30 dB	Yeraltı metrosu içi	94 dB
Ev gürültüsü	40 dB	Dizel kamyon (9.1 m)	100 dB
Cadde gürültüsü	50 dB	Çim biçme mak. (0.9 m)	107 dB
Ort. konuşma	60 dB	Perçin mak. (0.9 m)	115 dB
Araba içi	70 dB	Elektrikli testere (0.9 m)	117 dB
Şarkı söyleme (0.9 m)	75 dB	Rock müziği (1.8 m)	120 dB
Araba gürültüsü (7.6 m)	80 dB	Jet uçağı (30.4 m)	130 dB

Çizelge 1.1 Bazı ses kaynaklarının seviyeleri.

1.4.3 Ses Yüksekliği

İnsanlarda duyular (duyma, koklama, görme, vb.) uyarının şiddeti arttıkça logaritmik olarak artar. Birçok deney bu kuralı doğrulamış ve desibel ölçeğinin kullanılmasını gerektirmiştir.

Duyum insandan insana değişebileceği için ses yüksekliği de ancak öznel olarak ölçülebilir. Bu amaçla *sone* (son) birimi tanımlanmıştır: 40 dB seviyeli 1000 Hz frekanslı bir ton işaretinin ses yüksekliği 1 sone'dur. Sesin yüksekliği ile şiddeti arasındaki ilişki sone birimi cinsinden deneysel olarak aşağıdaki gibi verilmiştir [Gold and Morgan,1999]:

$$S \propto I^{0.3} \quad (1.4)$$

Yukarıdaki denklem frekansa bağlı olarak değişecektir. Frekansa göre ses yükseklik değerleri için ses yükseklik eğrileri verilmektedir [Gold and Morgan,1999].

1.5 Sayısal İşaret İşleme

Mühendislikte kullanılan işaretler analog işaretlerdir. Bilgisayarların geliştirilmesiyle sağladıkları işlem gücü ve hızı, bu makinelerin mühendislik çalışmalarında yaygın şekilde kullanılmasına neden olmuştur. Bilgisayarların sadece sayısal verilerle çalışması kullanılan analog işaretlerin de sayısala dönüştürülmesini zorunlu kılmıştır. Sayısal işaret işleme yöntemleri de bu noktada devreye girmektedir.

Konuşma işaretlerinin analizinde, sayısal işaret işleme yöntemlerinin kullanılabilmesi için önce analog işaretler sayısala çevrilmelidir. Sayısal işaret elde edildikten sonra istenilen işaret işleme yöntemi uygulanabilir. Konuşma işaretlerine uygulanan belli başlı işaret işleme yöntemleri şunlardır:

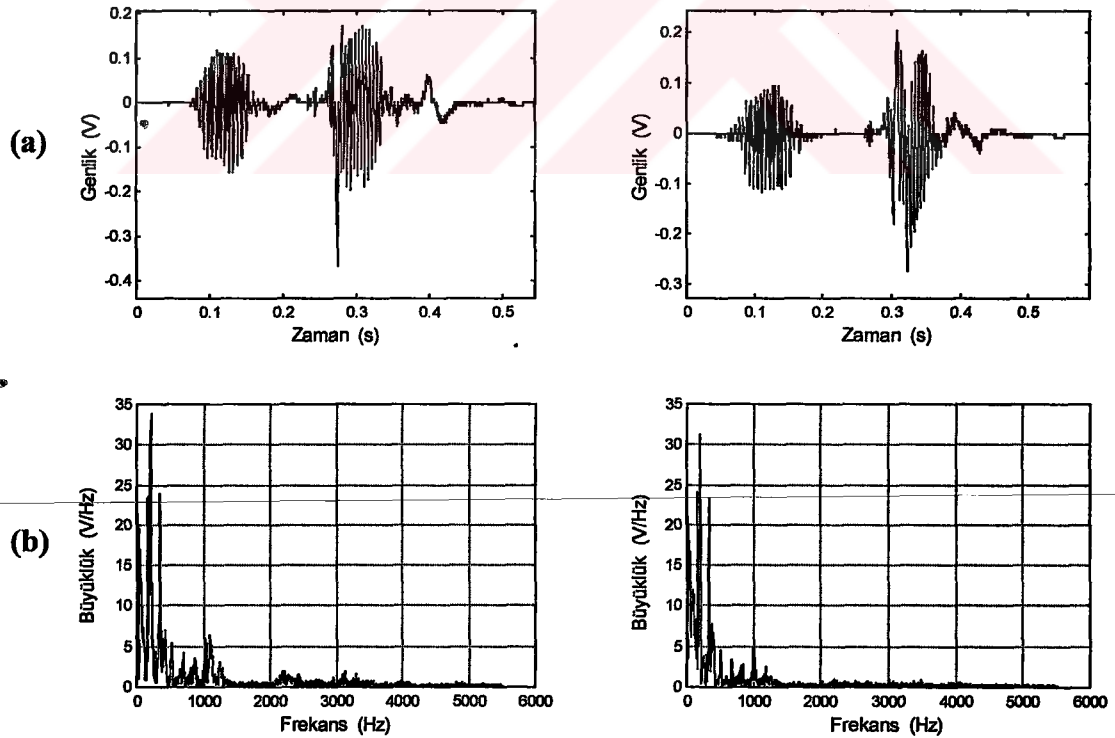
Örnekleme : Analog işareti sayısala dönüştürmede yapılan ilk işlemidir. Analog bir işareti sayısala çevirmek için işaret her T saniyede bir örneklenirse, elde edilen işaret $1/T$ örnekleme frekanslı olacaktır. Teorik olarak bu örnekleme hızı ile temsil edilebilecek maksimum frekans bileşeni $1/(2T)$ olacaktır. Telefon işaretleri 8 kHz ile örneklenmektedir [Rabiner and Schafer,1978]. Bu çalışmada incelenen işaretleri temsil edecek veri sayısını düşük tutmak için ses kayıt programının sunduğu frekans seçeneklerinden telefon işaretlerinin örnekleme frekansına en yakın olan 11025 Hz kullanılmıştır.

Kodlama : Örnekleme sonucunda, analog işaretin eşit zaman aralıkları ile alınmış genlik değerleri elde edilir. Kodlama işleminde, bu genlik değerleri 0 ve 1 bitlerinden oluşan sayısal kodlara dönüştürülmektedir. Böylece analog işaret ikili bitlerden oluşan sürekli bir dizi halinde gösterilmiş olur. Bu çalışmada kullanılan konuşma işaretleri kayıt edilirken Darbe Kod Modülasyonu (Pulse Code Modulation) kullanılmıştır. Darbe Kod Modülasyonu, örnekleme sonucunda elde edilen işareti logaritmik bir ölçeğe göre kuantize (quantize) eden bir kodlama yöntemidir [Rabiner and Schafer,1978].

Ses analizi ve sentezi : Ses analizi, ses işaretlerini kendisini oluşturan küçük (zaman ve frekans) bileşenlerine ayırma işlemidir. Bu işlemin tam tersi, yani küçük bileşenlerden ses işaretini tekrar elde etme işlemi ise ses sentezidir. Örnekleme ve

kodlamadan sonra, yani işaret sayısal hale getirildikten sonra elde edilen bilgi işaretin zaman-genlik bilgisidir ve işaretin yüksekliği hakkında bilgi verir. Zaman tanım bölgesi (zaman domeni) ses işaretlerinin incelenmesi için yeterli olmamaktadır. Bu nedenle, ses işaretlerin daha farklı özelliklerini de görebilmek için çeşitli analiz yöntemleri kullanılmaktadır. Ses analizinde sayısal işaret işleme yöntemlerinden en çok Fourier dönüşümü temelli spektrum, kepstrum (cepstrum) ve doğrusal kestirimli kodlama (Linear Predictive Coding) analizleri kullanılmaktadır [Gold and Morgan,1999]. Ayrıca dalgacık dönüşümü, Wigner-Ville ve Choi-Williams gibi çeşitli zaman-frekans dağılımları kullanılabilir [Quin and Chen,1996].

Şekil 1.1’de “dokuz” rakamının aynı kişiden aynı kayıt ortamında alınmış iki farklı örneği ve Fourier dönüşümleri verilmiştir. Görüldüğü gibi işlenmemiş hallerinde farklılıklar olmasına rağmen Fourier dönüşümleri oldukça benzerdir, aynı frekans tepeleri gözükmektedir. Bu nedenle zaman tanım bölgesinde konuşma



Şekil 1.1 “Dokuz” kelimesinin aynı kişiye ait iki farklı kayıt örneğinin (a) işlenmemiş halleri, (b) Fourier dönüşümleri.

işaretlerini karşılaştırmak yanıltıcı olabilir. Bu yüzden konuşma işaretleri ile çalışırken bu işaretlerin frekans içeriği ile ilgilenilmektedir ve kişiden, kayıt ortamından, bağımsız özellikleri çıkartılmaya çalışılmaktadır.

Ses analizinde işaretler dönüşüm yöntemlerinden birisiyle kendisini oluşturan bileşenlerine ayrılır. Fourier dönüşümü işaretleri sinüs ve kosinüslerin toplamı, dalgacık dönüşümü ise dalgacıkların toplamı şeklinde ifade eder. Ses işaretleri zamanla çok değişken işaretler oldukları için, Fourier dönüşümü bu işaretleri analiz etmek için yeterince elverişli bir yöntem değildir. Ses işaretlerinin anî değişimleri, frekans bileşenlerinin konumu gibi özelliklerini tespit etmek için kısa zamanlı Fourier dönüşümünü kullanmak daha uygundur. Bu çalışmada, frekans bilgisi yanında zaman bilgisine de yer veren KZFD ve dalgacık dönüşümleri ile çalışılmıştır.

Ses sentezinde ise analiz sonucunda elde edilen bileşenlerden ters dönüşümler kullanılarak gerçek işaret elde edilir. Gerek Fourier dönüşümü gerekse dalgacık dönüşümünün ters dönüşümlerini uygulamak mümkündür. Örneğin ters Fourier dönüşümü filtreleme, ters dalgacık dönüşümü ise gürültüden arındırma ve veri sıkıştırma amaçlı olarak kullanılmaktadır [Misiti et al.,2000].

Ses tanıma : Ses tanıma her hangi bir ses işaretini anlamlı dil bileşenlerine çevirme işlemidir. Bu anlamlı ses bileşenleri bir cümle, bir kelime hatta bir harf bile olabilir. Bu çalışmada 0-9 arasındaki rakamlardan oluşan konuşma işaretleri ile ses tanıma yapılmıştır. Bir sonraki bölümde ses tanıma hakkında bilgi verilecek ve bu çalışmada kullanılan ses örüntü tanıma yöntemi anlatılacaktır.

2 SES TANIMA

2.1 Giriş

Daha önce ses tanıma, “her hangi bir ses işaretini anlamlı dil bileşenlerine çevirme işlemi” olarak tanımlanmıştı. Ses tanıma işlemi, tek kelime tanıma, sürekli ses tanıma ve konuşmacı tanıma şeklinde sınıflandırılabilir.

Tek kelime tanıma işleminde dil modeline gerek yoktur. Tanınan kelime eğitme sonucu elde edilen veri tabanı ile karşılaştırılır ve en uygun sınıfa atanmaya çalışılır.

Sürekli ses tanımada, konuşma işaretleri cümleler halinde tanınmaya çalışılır. Cümle içindeki kelimeler ayrıştırılır ve kelimeler tek tek tanınır. Fakat sürekli ses tanımada sürecin içine dil modelleri de girmektedir. Çünkü ayrı ayrı kelimelerin oluşturduğu diziye toplu bir anlam kazandırılmak zorundadır. Bu nedenle çok iyi bir dil veri tabanı hazırlanmış olmalıdır.

Konuşmacı tanımada verilen ses ya da konuşma işaretlerinin hangi konuşmacıya ait olduğu araştırılır.

Ses tanıma pratiğe uygulamada zor bir işlemdir ve araştırmacılarca üzerinde çok çalışılan bir konudur [Long and Datta,1996], [Delfs and Jondral,1998], [Cassidy,2001]. Bu konuda dikkat edilmesi gereken belli başlı sorunlar şunlardır:

- Cümleler sürekli, bazen cümle içindeki kelimeler arasında boşluk bırakılmamaktadır. Bu durumda kelimeleri ayırt etmek zorlaşmaktadır.
- Konuşma işaretlerinin telaffuzu, kişiden kişiye, yöreden yöreye, kişinin o anki ruhsal duruma, yaşına, cinsiyetine göre değişebilmektedir. İşaretlerin içeriği bu etkenlerden herhangi birinin değişmesi ile farklı olacaktır.
- Kayıt edilecek ses işaretleri ortamın akustiğine, mikrofona ve ortam gürültüsüne göre değişecektir.

Ses tanımada kullanılacak algoritmalar bu etkenleri de göz önünde bulundurarak seçilmelidir. Seçilen algoritma işareti işlemek için çok fazla zaman

gerektirmemelidir, ses kayıt edildikten sonra tanıma işlemi için çok fazla zaman ayrılmamalıdır.

2.2 Ses Tanımanın Tarihçesi

Ses tanımanın tarihi 1920'li yıllara kadar uzanmaktadır. Bu tarihlerde geliştirilen Radio Rex isimli bir oyuncak ilk ses tanıma cihazıdır. Ancak bu oyuncak "Rex" kelimesi dışında, yeterli düzeyde 500 Hz enerjiye sahip birçok kelimeye de cevap vermiştir [Rabiner and Juang,1993].

1930 ve 1940'lı yıllarda da çeşitli çalışmalar yapılmış olmasına rağmen dikkate değer çalışmalar 1950'lilerde gerçekleştirilmiştir. 1952'de Bell Laboratuvarları'nda rakam tanıyan bir sistem geliştirilmiştir. Bu sistemde işaretlerin iki banttaki enerjilerine bakılmaktaydı ve %98'e varan başarı oranı elde edilmiştir.

1959'da Londra Koleji'nden (College of London) Denes akustik bilginin yanında dilbilgisi olasılıklarını da kullanmıştır. Denes'e göre belirli bir dil bileşenin söylenilme olasılığı önceki dil bileşenine de bağlıdır.

1963'te Widrow yapay sinir ağlarını kullanarak rakam tanıma yapmıştır.

1964'te Martin yapay sinir ağlarını hece (phoneme) tanıma için kullanmıştır.

1960'lı yıllarda işaretlerin spektrumunu incelemek için üç yeni yöntem geliştirilmiştir: Hızlı Fourier dönüşümü, kepsral (cepstral) analizi ve doğrusal kestirimli kodlama (Lineer predictive coding). Bunların yanında ses tanıma kullanılmak üzere yeni örüntü tanıma yöntemleri de geliştirilmiştir: dinamik zaman sıkıştırması ve gizli Markov modelleri.

1970'lerde ARPA (Advanced Research Projects Agency, İleri Araştırma Projeleri Ajansı) 15 milyon dolarlık bir proje başlatmıştır. Projenin amacı 1000 kelimelik veri tabanında birkaç konuşmacı, sürekli konuşma ve dilbilgisi kurallarını kapsayacak bir ses tanıma sistemi geliştirmektir. Sadece Bruce Lowerre adında bir üniversite öğrencisi istenilen başarıyı sağlayan bir sistem geliştirebilmiştir.

1984'te ikinci ARPA projesi başlatılmıştır. Bu projede 1000 kelimelik veri tabanından oluşturulan cümleler tanınmaya çalışılmıştır. İkinci ARPA projesi ses tanıma ile uğraşan bir çok yeni laboratuvarın, araştırma merkezinin açılmasını ve gelişmesini sağlamıştır.

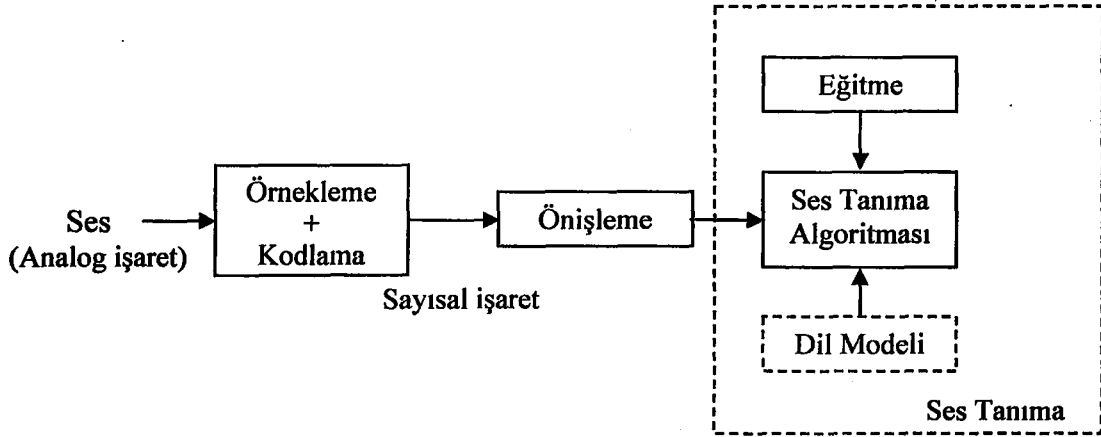
1986'da ses tanıma sistemlerinin performanslarının, ortak bir veri tabanı ile karşılaştırılabilmesi amacıyla TIMIT veri tabanı geliştirilmiştir. 630 konuşmacının söylediği 10 cümlelerin ses kayıtları yapılmıştır. Ses kayıtları Texas Instruments firmasında, kayıtların dil bileşenlerine ayrılması ise Massachusetts Teknoloji Enstitüsü'nde (Massachusetts Institute of Technology) yapılmıştır. Bu nedenle de bu veri tabanına her iki kuruluşun isminin baş harflerden oluşan TIMIT veri tabanı denilmiştir.

1990'lı yıllardan bugüne ses tanıma çalışmaları hızla artmış ve günlük hayata girmeye başlamıştır. Telefon uygulamalarında, fiziksel engelliler için geliştirilen sistemlerde, konuşmadan-yazıya ve yazıdan-konuşmaya çeviri, yabancı dilde çeviri vb. uygulamalara entegre edilmektedir.

Gelişim hızlı olsa da ses tanımada hala çözülmeye çalışılan konular vardır. Özellikle konuşmadan-yazıya çeviri, konuşmacı tanıma, konuşmacı-bağımsız ses tanıma, gürültüye dayanıklılık gibi sorunlar üzerinde çalışılmaktadır [Janer et al.,1996], [Sarikaya et al.,1998], [Hau and Zhu,2000].

2.3 Ses Tanıma Mimarisi

Ses tanıma mimarisi ana hatları ile Şekil 2.1'de verilmiştir. İlk önce ses işaretleri mikrofona gibi bir dönüştürücü ile elektriksel işarete dönüştürülmektedir. Bu elektriksel işaret örneklendikten sonra kodlanarak bit dizisi halinde kayıt edilmektedir.



Şekil 2.1 Ses tanıma mimarisi.

Ön-işleme, *özellik çıkartma* olarak da adlandırılabilir. Bu süreçte ses işaretlerinin ayırt edici özellikleri elde edilir. Ses işaretlerini ön-işlemeye sokmanın iki temel nedeni vardır. Birincisi aynı ses işaretinden, kaç kez (ve de değişik kişiler tarafından) söylenirse söylensin değişmeyen özellikler elde etmektir. Böylece konuşma işaretlerinin kişiden ve kayıt ortamından kaynaklanan, gürültü ve değişimlerden bağımsız bir karşılaştırması yapılabilir. İkincisi ham işarete veri sayısı çok fazladır, bu tanıma işlem süresini uzatır. İşaretlerin ayırt edici özellikleri çıkartılarak sadece bu özellikleri karşılaştırılabilir.

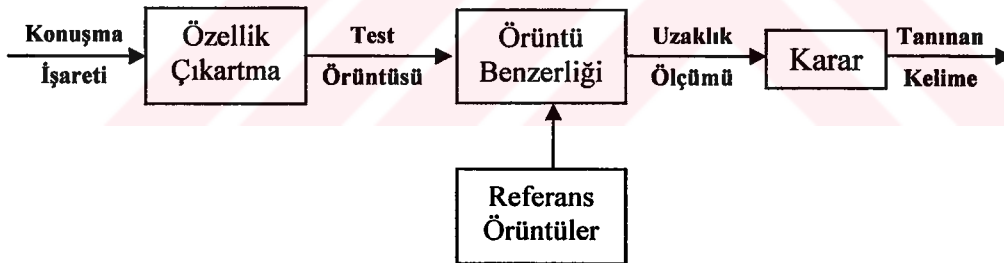
Ön-işleme, işaretin durağan olduğu kabul edilebilecek kadar küçük analiz pencereleri ile yapılır. Özellik çıkartma bu pencerelere çeşitli dönüşümler uygulanarak yapılır. Bu şekilde tüm işaret pencerelenerek analiz edilir ve her pencereye ait özellik vektörleri oluşturulur. İşareti temsil eden bu özellik vektörleri dizisine *örüntü* (pattern) denir. Bu şekilde işaretlerin örüntüleri elde edilerek yapılan ses tanıma işlemine de *örüntü tanıma* (pattern recognition) denir.

Ses tanıma süreci, eğitme, algoritma ve dil modelinden oluşmaktadır. Eğitme sürecinde önceden kayıt edilen ses işaretlerinin özellik vektörleri hesaplanarak örüntüleri oluşturulur. Eğitme sürecinde elde edilen bu örüntüler ses tanıma işlemine sokulacak işaretin örüntüsü ile karşılaştırılır. Bu karşılaştırma ses tanıma algoritması tarafından yapılır. İki örüntü arasındaki *uzaklık* (benzerlik ölçüsü) hesaplanır ve hangi sınıf için uzaklık en küçükse işaret o sınıfa atanır. Bir sonraki bölümde

anlatılacak olan örüntü tanıma, şablon eşleme (template matching), dinamik zaman sıkıştırması, yapay sinir ağları ve gizli Markov modelleri gibi yöntemleri içermektedir [Rabiner and Juang,1993]. Dil modeli çok iyi bir sözlüğe ve dilbilgisine sahip olmalıdır. Ayır ayrı tanınan kelimelere toplu halde bir anlam verilmeye, başka bir deyişle cümle oluşturulmaya çalışılır.

2.4 Örüntü Tanıma

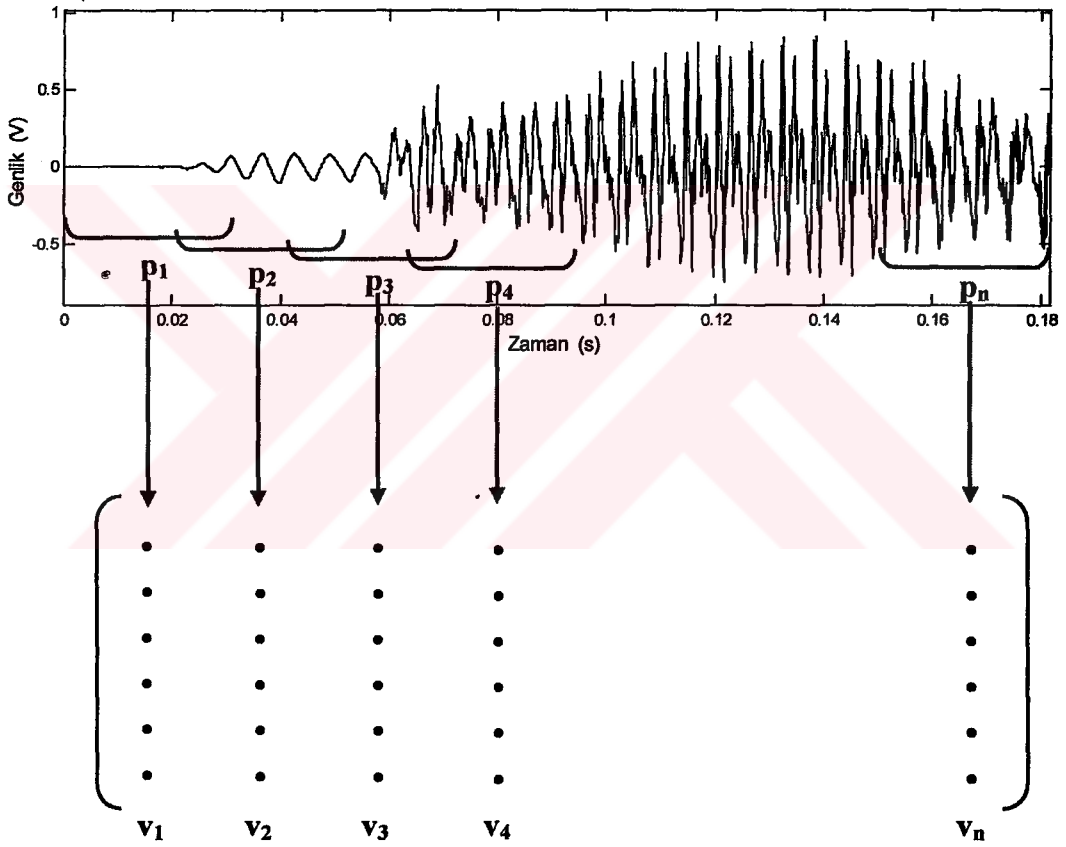
Şekil 2.2’de blok diyagramı verilen örüntü tanımada, konuşma işaretlerinin özellik vektörlerinden oluşan örüntüler elde edilir ve eğitime sonucu elde edilen örüntülerle karşılaştırılır. Tek kelime tanıma uygulamalarında eğitime ile her sınıf için ayrı ayrı referans örüntüler oluşturulur ve kaydedilir. Test kelimesinin özellik vektör örüntüsü her sınıfa ait kelimelerin referans örüntüleriyle karşılaştırılarak uygun sınıfa atanır.



Şekil 2.2 Örüntü tanıma blok diyagramı.

Konuşma işaretleri zamanla çok değişken işaretler oldukları için durağan kabul edilebilecek kadar dar pencereler halinde analiz edilirler. Ses tanımada genellikle 20-30 ms (bu çalışmada toplam kayıt süresinin 1/20 ve 1/30’u kadar) pencere uzunluğu kullanılmaktadır [Gold and Morgan,1999], [Kim et al.,2000]. Şekil 2.3’te gösterildiği gibi pencere geçişlerinde bilgi kaybetmemek için pencere merkezleri kaydırılarak pencereleme yapılır. Her pencerenin ayrı ayrı Fourier, doğrusal kestirimli kodlama, kepstrum ve dalgacık dönüşüm gibi dönüşümleri alınarak özellik vektörleri elde edilir. Bu özellik vektörleri kullanılan analiz yöntemi sonucu elde edilen katsayılardan oluşacaktır [Kaisheng and Zhigang,1998], [Tüfekçi

and Gowdy,2000]. Şekil 2.3'te gösterilen örüntü tanıma matrisinin sütunları özellik vektörlerinden oluşmaktadır ve pencere sayısı kadar sütun vardır. Satır sayısını ise kullanılan dönüşüm sonucunda elde edilen katsayılar belirleyecektir. Bu çalışmada, ses tanıma deneylerinde kullanılan konuşma işaretlerinin özellik vektörleri KZFD ya da ayrık dalgacık dönüşümü katsayılarından elde edilmiştir ve 30 ms pencere uzunluğunun yanında 20 ve 45 ms pencere uzunlukları da denenmiştir.



Şekil 2.3 Konuşma işaretlerinin pencerelenmesi ve örüntü matrisinin elde edilmesi.

Örüntü tanıma için kullanılan başlıca yöntemler şunlardır: Şablon eşleme, dinamik zaman sıkıştırması, gizli Markov modelleri ve yapay sinir ağları. Bu yöntemler hakkında kısaca bilgi verilecektir.

Şablon eşlemede, örüntüler vektör vektör karşılaştırılır. Karşılaştırma işlemi özellik vektörleri arasındaki uzaklığın hesaplanması ile yapılır. Örneğin referans

örüntünün 1. vektörü (matrisin 1. sütunu) test örüntüsünün 1. vektörü ile karşılaştırılır. Aynı şekilde 2. vektörler (örüntü matrislerinin 2. sütunları) birbirleri ile karşılaştırılır. Bu işlem son vektöre gelinceye kadar devam ettirilir. Örüntüler arasındaki toplam uzaklık, vektör uzaklıklarının toplamı ya da ortalaması olacaktır. Test örüntüsü hangi sınıf için minimum uzaklık veriyorsa o sınıfa atanacaktır. Uzaklık hesaplanması Euclid uzaklığı, Manhattan uzaklığı, Gauss sınıflandırması gibi yöntemlerle yapılır [Cassidy,2001].

Dinamik zaman sıkıştırması, farklı uzunluktaki örüntülerin karşılaştırılabilmesi için geliştirilmiştir. Şablon eşlemede iki örüntünün vektör sayısı (başka bir deyişle sütun ya da pencere sayısı) aynı olmak zorundadır. Fakat gerçekte iki örüntü çoğu kez aynı uzunlukta olmaz (aynı sayıda özellik vektörüne sahip olmaz). Aynı kelime için yapılmış kayıtlar tek kişi tarafından seslendirilmiş olsa dahi farklı uzunluklarda olabilir. Değişik uzunluktaki iki örüntüyü karşılaştırmak için zaman normalizasyonu yapılmalıdır. Zaman normalizasyonu dinamik zaman sıkıştırması ile gerçekleştirilir. Bu çalışmada gerçekleştirilen ses tanıma deneyleri bir sonraki bölümde anlatılacak olan dinamik zaman sıkıştırması yöntemi ile yapılmıştır.

Gizli Markov modelleri, olasılık temelli bir örüntü tanıma yöntemidir. Bu yöntemde özellik vektörlerinin olasılık dağılımları hesaplanır ve örüntüler olasılık dağılımlarından oluşan diziler halinde modellenir. Eğitim sürecinde, tüm sınıflara ait model parametreleri hesaplanır. Tanıma sürecinde ise, test örüntüsü bu model parametreleri kullanılarak en uygun sınıfa atanır [Gold and Morgan,1999].

Yapay sinir ağları, örüntüleri nöron denilen birimlerden oluşan bir ağ şeklinde modeller. Nöronlar birbirlerine transfer fonksiyonları aracılığıyla bağlıdır. Eğitim sırasında, her sınıfa ait transfer fonksiyonları elde edilir. Bu transfer fonksiyonları kullanılarak test örüntüsünün her sınıf için uzaklıkları hesaplanır ve uzaklığın minimum olduğu sınıfa atanır [Rabiner and Juang,1993].

2.5 Dinamik Zaman Sıkıştırması

Dinamik zaman sıkıştırması farklı uzunluktaki iki örüntüyü karşılaştırmayı sağlayan (başka bir deyişle iki örüntü arasındaki uzaklığı hesaplayan) bir yöntemdir [Rabiner and Juang,1993].

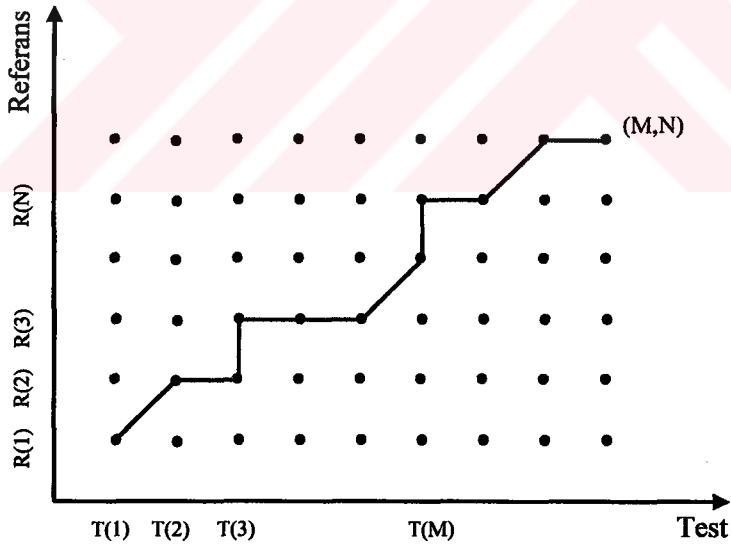
Karşılaştırılacak test ve referans örüntüleri aşağıdaki gibi tanımlansın:

Test : $T(i)$ $i=1, \dots, M$ (test örüntüsünün uzunluğu)

Referans : $R(j)$ $j=1, \dots, N$ (referans örüntüsünün uzunluğu)

Burada $T(i)$ ve $R(i)$ özellik vektörleridir. İki vektör arasındaki uzaklık u ile tanımlanırsa, (i,j) noktasına karşılık gelen uzaklık $u(i,j)$ olacaktır. Şekil 2.4'teki grafikte görüldüğü gibi uzaklık hesaplama $u(1,1)$ ile başlamakta ve bir yol boyunca ilerleyerek $u(M,N)$ ile bitmektedir. Buna göre izlenen herhangi bir yol için toplam uzaklık aşağıdaki gibi olacaktır:

$$U_n = \sum_{k=1}^M \sum_{l=1}^N u(k,l) \quad (2.1)$$



Şekil 2.4 Dinamik zaman sıkıştırması grafiği.

Denklem (2.1)'de verilen U_n toplam uzaklıkları, izlenebilecek tüm yollar için hesaplanır. İki örüntü arasındaki en yakın uzaklık ise U_n uzaklıklarının minimumu olacaktır:

$$c_{T,R} = \min U_n = \min \sum_{k=1}^M \sum_{l=1}^N u(k,l) \quad (2.2)$$

Test ve referans örüntüleri arasındaki $c_{T,R}$ minimum uzaklığını bulmak için olası tüm U_n uzaklıklarının bulunması gereklidir. Ancak örüntülerin uzunlukları göz önüne alındığında tüm U_n uzaklıklarını hesaplamak çok zaman alacaktır. Bu nedenle tüm yollar yerine sadece bazı yollar için toplam uzaklığın hesaplandığı ve bunlar arasından minimum uzaklığın alındığı yöntemler geliştirilmiştir [Sakoe and Chiba,1978], [Itakura, 1975]. Bu çalışmada hem Sakoe&Chiba'nın hem de Itakura'nın önerdiği dinamik zaman sıkıştırma yöntemleri denenmiştir.

3 FOURIER ANALİZİ

3.1 Giriş

Fourier analizi, bir işareti zaman tanım bölgesinden frekans tanım bölgesine dönüştüren bir yöntemdir ve işaretin tüm frekans spektrumunu verir. Fourier analizi hem Fourier serilerini hem de Fourier dönüşümünü kapsar. Fourier serileri periyodik fonksiyonlar için kullanılır. Fourier serilerindeki periyot sonsuza yaklaştırarak Fourier dönüşümünü elde edilebilir [Goswami and Chan,1999]. Tüm analiz yöntemleri gibi Fourier dönüşümü de bir baz kümesi kullanmaktadır. Fourier analizine geçmeden önce baz fonksiyonları hakkında bilgi verilecektir.

3.2 Baz Fonksiyonları

Herhangi bir $f(t)$ fonksiyonu aşağıdaki gibi bir seri açılımı şeklinde gösterilebilir:

$$f(t) = \sum_j \alpha_k g_k(t) \quad (3.1)$$

Bu açılımda α_k katsayıları açılım katsayıları, $g_k(t)$ fonksiyonları ise açılım fonksiyonlarıdır.

$f(t)$ fonksiyonu için bu açılım tek ise bu açılım kümesi *baz* denir. Bu durumda baz fonksiyonları $g_k(t)$ birbirlerinden lineer bağımsızdırlar. Matematiksel dönüşümlerin (Fourier, KZFD, Dalgacık...) analiz kısmında bu açılım katsayıları bulunmakta, sentez kısmında ise bu katsayılar ve baz fonksiyonlarının oluşturduğu lineer kombinasyon hesaplanmaktadır.

Eğer bu baz kümesi *ortogonal* ise katsayıları hesaplamak çok kolay olacaktır. Ortogonalite koşulu aşağıdaki gibi ifade edilir:

$$\langle g_k(t), g_l(t) \rangle = \int g_k(t) \overline{g_l(t)} dt = 0 \quad k \neq l \quad (3.2)$$

Vektör uzayı düşünülürse, ortogonal iki vektör birbirine 90° dik, dolayısıyla da vektörel çarpımları sıfır olacaktır. Baz fonksiyonları $\|g_l(t)\|=1$ norm şartını sağlamak için normalize edilir ve aşağıdaki *ortonormallik* koşulunu sağlarlar:

$$\langle g_k(t), g_l(t) \rangle = \int g_k(t) \overline{g_l(t)} dt = \delta(k-l)^1 \quad (3.3)$$

Denklem (3.1)'in her iki tarafı da $g_l(t)$ ile çarpılırsa indisleri farklı olan baz fonksiyonlarının çarpımları ortogonalite nedeniyle sıfır olacak, ve sadece Denklem (3.4)'ün sağ tarafındaki $\alpha_k g_k(t) \overline{g_l(t)}$ terimi kalacaktır. Ortonormal oldukları için de kalan iki terimin çarpımı bire ($\delta_{l,k} = 1$) eşit olacak ve α_k katsayısı elde edilecektir:

$$\begin{aligned} \langle f(t), g_l(t) \rangle &= \int f(t) \overline{g_l(t)} dt \\ &= \int \sum_j \alpha_j g_j(t) \overline{g_l(t)} dt \\ &= \sum_j \alpha_j \delta_{j,l} \\ &= \alpha_l \end{aligned} \quad (3.4)$$

Fourier dönüşümde ortogonal baz olarak $\sin(k\omega_0 t)$ ve $\cos(k\omega_0 t)$ fonksiyonları kullanılmaktadır. Dalgacık dönüşümde ise baz olarak dalgacık fonksiyonları kullanılmaktadır.

Her zaman ortogonal bir baz kümesi oluşturmak mümkün olmayabilir. Dalgacık dönüşümde ortogonal baz kümelerinin yanında *biortogonal* baz kümeleri de kullanılmaktadır. Biortogonal baz kümeleri, birbirinin ikili (duali) iki ayrı baz kümesinden oluşmaktadır ve birbirlerine ortogonaldirler. Bu baz kümelerini kullanarak yukarıda anlatılan açılımın benzeri elde edilebilir [Goswami and Chan, 1999]:

$$\langle g_k(t), \hat{g}_l(t) \rangle = \int g_k(t) \overline{\hat{g}_l(t)} dt = \delta_{k,l} \quad (3.5)$$

$$\begin{aligned} \alpha_k &= \langle f(t), \hat{g}_k(t) \rangle \\ &= \int f(t) \overline{\hat{g}_k(t)} dt \end{aligned} \quad (3.6)$$

¹ Delta fonksiyonu $\delta(k-l)=1 \quad k=l$
 $\delta(k-l)=0 \quad k \neq l$

$$\begin{aligned}\hat{\alpha}_k &= \langle f(t), g_k(t) \rangle \\ &= \int f(t) \overline{g_k(t)} dt\end{aligned}\quad (3.7)$$

3.3 Fourier Serileri

Periyodu T olmak üzere $f(t)$ gibi bir fonksiyonun Fourier serisi açılımı aşağıdaki gibi verilir:

$$f(t) = \sum_{k=-\infty}^{\infty} \alpha_k e^{\frac{j2\pi kt}{T}} \quad (3.8)$$

$\{e_k\} = \{e^{j2\pi kt/T}\}$ fonksiyon kümesi ortogonal bir baz oluşturmaktadır. Bu ortogonalite özelliğinden faydalanarak α_k Fourier katsayıları şu şekilde hesaplanabilir:

$$\alpha_k = \frac{1}{T} \int_0^T f(t) e^{-\frac{j2\pi kt}{T}} dt \quad (3.9)$$

Sinüs ve kosinüs fonksiyonları kullanılarak Fourier serisi açılımı değişik bir şekilde gösterilebilir. $e^{j\omega} = \cos\omega + j\sin\omega$ Euler formülü kullanılarak açılım aşağıdaki gibi tekrar düzenlenebilir:

$$f(t) = a_0 + \sum_{k=1}^{\infty} a_k \cos \frac{2\pi kt}{T} + b_k \sin \frac{2\pi kt}{T} \quad (3.10)$$

Bu seri açılımında gözüken a_0 , a_k ve b_k Fourier katsayıları ise aşağıdaki gibi elde edilir:

$$a_0 = \frac{1}{T} \int_{-T/2}^{T/2} f(t) dt \quad (3.11)$$

$$a_k = \frac{2}{T} \int_{-T/2}^{T/2} f(t) \sin \frac{2\pi kt}{T} dt \quad (3.12)$$

$$b_k = \frac{2}{T} \int_{-T/2}^{T/2} f(t) \cos \frac{2\pi kt}{T} dt \quad (3.13)$$

3.4 Fourier Dönüşümü

Fourier serisi açılımındaki katsayıların yerine açık ifadesi yazılır

$$f(t) = \sum_{k=-\infty}^{\infty} \left[\frac{1}{T} \int_{-T/2}^{T/2} f(t') e^{-\frac{j2\pi kt}{T}} dt' \right] e^{\frac{j2\pi kt}{T}} \quad (3.14)$$

ve Fourier serisinde iki frekans bileşeni arasındaki uzaklık $\Delta\omega = 2\pi/T$ olarak tanımlanarak ve formülde yerine konularak

$$f(t) = \frac{1}{2\pi} \sum_{k=-\infty}^{\infty} \Delta\omega \left[\frac{1}{T} \int_{-T/2}^{T/2} f(t') e^{-jk\Delta\omega} dt' \right] e^{-jk\Delta\omega} \quad (3.15)$$

T periyodu sonsuza götürüldüğünde $\Delta\omega$ $d\omega$ 'ya, $k\Delta\omega$ ise ω 'ya yaklaşacaktır ve seri toplamı entegrale dönüşecektir.

$$f(t) = \frac{1}{2\pi} \int_{-\infty}^{\infty} \left[\int_{-\infty}^{\infty} f(t') e^{-j\omega t'} dt' \right] e^{j\omega t} d\omega \quad (3.16)$$

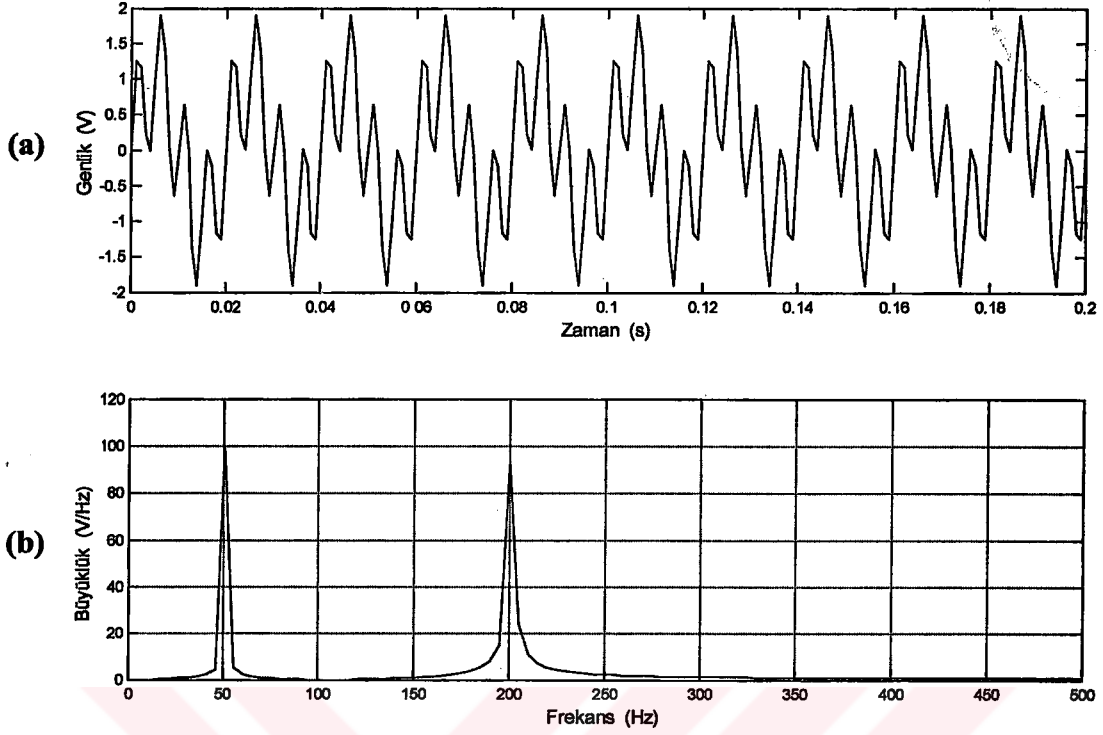
Köşeli parantez içindeki entegral ifadesi $F(\omega)$ ile gösterilirse

$$F(\omega) = \int_{-\infty}^{\infty} f(t) e^{-j\omega t} dt \quad (3.17)$$

$$f(t) = \frac{1}{2\pi} \int_{-\infty}^{\infty} F(\omega) e^{j\omega t} d\omega \quad (3.18)$$

Fourier dönüşümü çifti elde edilmiş olunur. Yukarıdaki ilk denklem $f(t)$ fonksiyonunun Fourier dönüşümü, ikincisi ise ters Fourier dönüşümdür.

Şekil 3.1'de matematiksel olarak üretilmiş $f(t) = A_1 \sin(2\pi 50t) + A_2 \sin(2\pi 100t)$ fonksiyonu ve Fourier dönüşümü verilmiştir. Görüldüğü gibi işaretin içerdiği 50 ve 100 Hz frekans bileşenleri dönüşümde açıkça ortaya çıkmaktadır. Zaman gösteriminden bu frekans değerlerini hesaplayarak ya da ölçerek elde etmek oldukça zordur. İşaret oluşturulurken örnekleme frekansı olarak 1000 Hz seçildiği için dönüşümde frekans spektrumu 500 Hz değerine (örnekleme frekansının yarısı) kadar verilmiştir.



Şekil 3.1 50 Hz ve 200 Hz frekans bileşenlerine sahip bir işaretin (a) işlenmemiş hali ve (b) Fourier dönüşümü.

3.5 Ayırık Fourier Dönüşümü

Gerçek hayatta hesaplamalar bilgisayarlarda yapılmaktadır ve bu nedenle incelenen işaretler de örneklenerek sayısal olarak çevrilmektedir. Kullanacak matematiksel dönüşümlerin de ayırık elemanlardan oluşan bu işaretlere uygulanabilir olması gereklidir. Bu amaçla ayırık Fourier dönüşümü geliştirilmiştir.

AFD (Ayırık Fourier Dönüşümü) bir işaretin sonlu sayıdaki örneklerinden Fourier dönüşümünün elde edilmesini sağlar.

T periyotlu bir $f(t)$ işaretini aşağıdaki gibi Fourier serisine açabileceği daha önce gösterilmişti:

$$f(t) = a_0 + \sum_{k=1}^{\infty} a_k \cos \frac{2\pi kt}{T} + b_k \sin \frac{2\pi kt}{T} \quad (3.19)$$

$$a_k = \frac{2}{T} \int_0^T f(t) \sin \frac{2\pi kt}{T} dt \quad (3.20)$$

$$b_k = \frac{2}{T} \int_0^T f(t) \cos \frac{2\pi kt}{T} dt \quad (3.21)$$

$f(t)$ fonksiyonu periyodik olduğu için entegral sınırlarının $[-T/2, T/2]$ yerine $[0, T]$ olması herhangi bir değişiklik yapmayacaktır.

Fourier serisi F_k ile tanımlanır ve kompleks gösterim kullanılırsa dönüşüm sınırları değiştirilmiş olur:

$$\begin{aligned} F_k &= a_k - b_k \\ e^{-j\frac{2\pi kt}{T}} &= \cos\left(\frac{2\pi kt}{T}\right) - j \sin\left(\frac{2\pi kt}{T}\right) \\ F_k &= \frac{1}{T} \int_0^T f(t) e^{-j\frac{2\pi kt}{T}} dt \quad k \geq 0 \end{aligned} \quad (3.22)$$

Eğer $f(t)$ fonksiyonu N adet örneğe sahip $\{f_r\}$ ayrık serisinden oluşuyorsa örnekleme aralığı $\Delta = T/N$ ve $t = r\Delta$ olacaktır ve ayrık Fourier dönüşümü aşağıdaki gibi ifade edilebilecektir:

$$F(k) = \frac{1}{N} \sum_{n=0}^{N-1} f(n) e^{-j\frac{2\pi kn}{N}} \quad (3.23)$$

Ters Fourier dönüşümü ise aşağıdaki gibi ifade edilir:

$$f(n) = \sum_{k=0}^{N-1} F(k) e^{j\frac{2\pi kn}{N}} \quad (3.24)$$

3.6 Hızlı Fourier Dönüşümü

Denklem (3.23) ile tanımlanan AFD N^2 çarpma ve $N(N-1)$ toplama işlemi gerektirmektedir. N büyük değerler aldıkça özellikle N^2 adet çarpma işlemi önemli bir problem olmaktadır. İşlem sayısını azaltmak ve dönüşümü hızlandırmak için

HFD (Hızlı Fourier Dönüşümü) geliştirilmiştir. HFD'nü kullanarak AFD katsayıları yaklaşık $N \log_2 N$ işlem ile hesaplanabilir [Goswami and Chan,1999].

Çeşitli HFD algoritmaları vardır. Bunlardan en çok kullanılanı Danilson ve Lanczos'un geliştirdiği *radix-2 HFD* algoritmasıdır. MATLAB alet kutusunda da bu algoritma kullanılmaktadır ve bu çalışmada verilen Fourier dönüşümleri bu yöntem ile hesaplanmıştır [Brian and Breiner,1995].

Danilson ve Lanczos'una göre bir işaretin N adet veriden oluşan AFD katsayıları, $N/2$ adet veriden oluşan iki ayrı AFD katsayıları takımının toplamı şeklinde gösterilebilir. Bu algoritma son iki AFD katsayısına ulaşana kadar sürdürülebilir. Eğer N değeri 2'nin kuvveti değilse, sıfır ekleyerek 2'nin kuvveti haline getirilebilir.

3.7 Kısa Zamanlı Fourier Dönüşümü (KZFD)

Fourier dönüşümünü incelendiğinde zaman bilgisini içermediği görülür. Sonuçta sadece büyüklük-frekans bilgisi elde edilmektedir. Eğer incelenen işaretin frekans içeriği zamanla değişmiyorsa (başka bir deyişle durağan ise) Fourier dönüşümü analiz için yeterli olacaktır. Fakat işaretin frekans içeriği zamanla değişiyorsa (durağan değilse) Fourier dönüşümü bununla ilgi bir bilgi veremeyecektir.

Fourier dönüşümünün bu eksikliğini gidermek amacıyla araştırmacılarca KZFD geliştirilmiştir. KZFD hem zaman hem de frekans bilgisine yer vermektedir. KZFD'nde işaret pencerelere bölünerek Fourier dönüşümleri elde edilir. Seçilen bu pencere kaydırılarak işaretin parça parça tüm zaman eksenini boyunca Fourier dönüşümü alınır.

KZFD'nün ifade edilebilmesi için önce pencere fonksiyonu tanımlanmalıdır. $w(t)$ pencere fonksiyonu olmak üzere $f(t)$ fonksiyonunun KZFD aşağıdaki gibi tanımlanmaktadır [Goswami and Chan,1999]:

$$G(t, \omega) = \int_{-\infty}^{\infty} f(t)w(t-b)e^{-j\omega t} dt \quad (3.25)$$

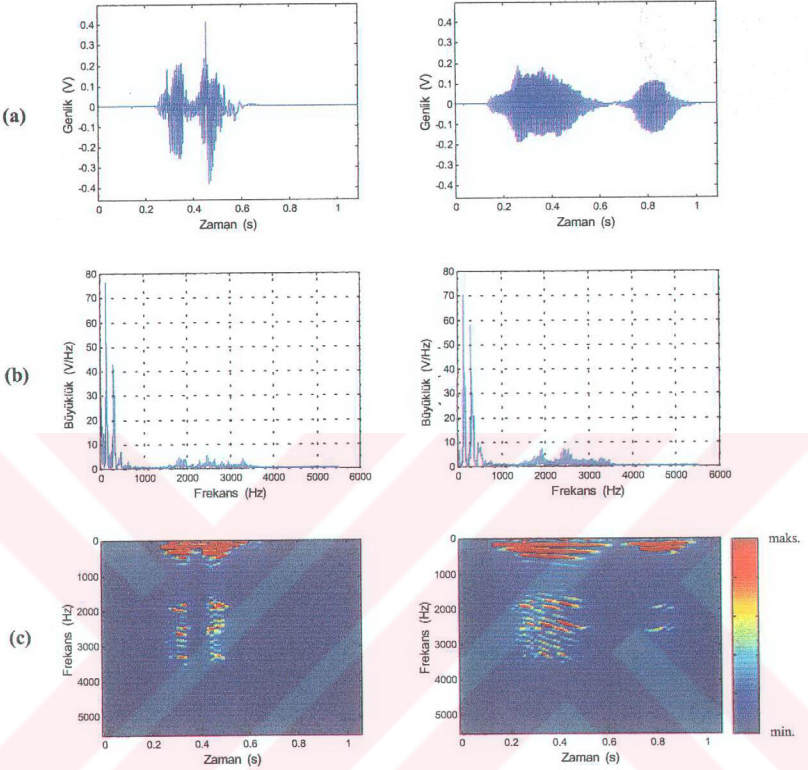
Burada, b parametresi Fourier dönüşümü katsayılarının hesaplandığı pencerenin büyüklüğünü belirler ve zaman bilgisini verir. Bu çalışmada pencere fonksiyonu olarak Hanning pencere fonksiyonu kullanılmıştır [Denbigh,1998].

$f(t)$ fonksiyonunu ters KZFD ile tekrar elde etmek için Denklem (3.25)'in her iki tarafı da $\overline{w(t-b)}$ ile çarpılır ve b 'ye göre entegrali alınır:

$$f(t) = \frac{1}{2\pi\|w(t)\|} \int_{-\infty}^{\infty} e^{-j\omega t} d\omega \int_{-\infty}^{\infty} G(t, \omega) \overline{w(t-b)} dt \quad (3.26)$$

Şekil 3.2'de "yedi" rakamının iki farklı örneğinin işlenmemiş halleri, Fourier dönüşümleri ve KZFD'leri verilmiştir. Grafiği verilen KZFD'leri üç boyutlu bilgi taşımaktadır; x-ekseni zaman ve y-ekseni frekans bilgisini, grafikteki renkler ise büyüklük bilgisini vermektedir. Grafiğin sağ tarafında renk palet çubuğu verilmiştir. Kırmızı renk en büyük değeri, mavi renk ise en küçük değeri temsil etmektedir. Grafikte kırmızı renkte olan kısımlar orada frekans bileşenlerinin büyük olduğunu gösterir. Tezdeki tüm KZFD ve dalgacık dönüşümü grafikleri bu şekilde incelenmelidir.

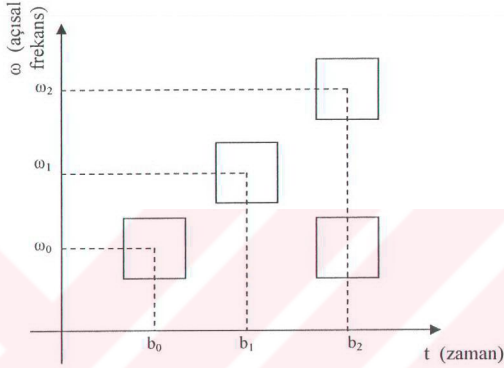
Soldaki örnek daha hızlı (süresi yaklaşık 0.3 s), sağdaki ise daha yavaş (süresi yaklaşık 0.8 s) söylenmiştir. Beklenildiği gibi her iki işaretin Fourier dönüşümleri benzerdir. Ancak Fourier dönüşümü işaretlerin zaman içeriği hakkında herhangi bir bilgi vermediğinden dolayı işaretleri ayırt etmek mümkün değildir. Zaman bilgisini de içeren KZFD'leri incelendiğinde kelimelerin farklı sürelerde söylendiği anlaşılabilmektedir. Kelimeleri oluşturan heceler ve bu hecelerin içerdiği frekans bileşenleri ayırt edilebilmektedir. Örneğin her iki kelime de yaklaşık 2000 Hz, 2500 Hz ve 3400 Hz frekans bileşenlerine sahiptir. Fourier dönüşümünde de bu frekans tepeleri gözükmemektedir, fakat konumları anlaşılamamaktadır.



Şekil 3.2 “Yedi” kelimesinin iki farklı örneğinin (a) işlenmemiş halleri, (b) Fourier dönüşümleri ve (c) KZFD’leri.

KZFD’nde işarete çok dar bir pencereden bakılmakta ve işaretin sadece bu pencere boyunca durağan olduğu kabul edilerek Fourier dönüşümü alınmaktadır. Daha sonra bu pencere kaydırılarak tüm zaman eksenı boyunca Fourier katsayıları hesaplanmaktadır. Şekil 3.2 (c)’deki KZFD’leri hesaplanırken 23 ms pencere uzunluğu kullanılmış ve her 7.6 ms’de bir pencere kaydırılarak Fourier dönüşümü alınmıştır.

Şekil 3.3'te KZFD'nde pencere büyüklükleri verilmiştir. Görüldüğü gibi pencere büyüklüğü analiz boyunca sabittir. Fourier dönüşümünde pencere büyüklüğü sonsuzdur ve dönüşüm tüm işaret süresince uygulanır. Bu nedenle frekans çözünürlüğü çok iyidir fakat zaman çözünürlüğü yoktur. KZFD ise hem frekans hem de zaman çözünürlüğü vermektedir.



Şekil 3.3 KZFD'nde zaman-frekans penceresi.

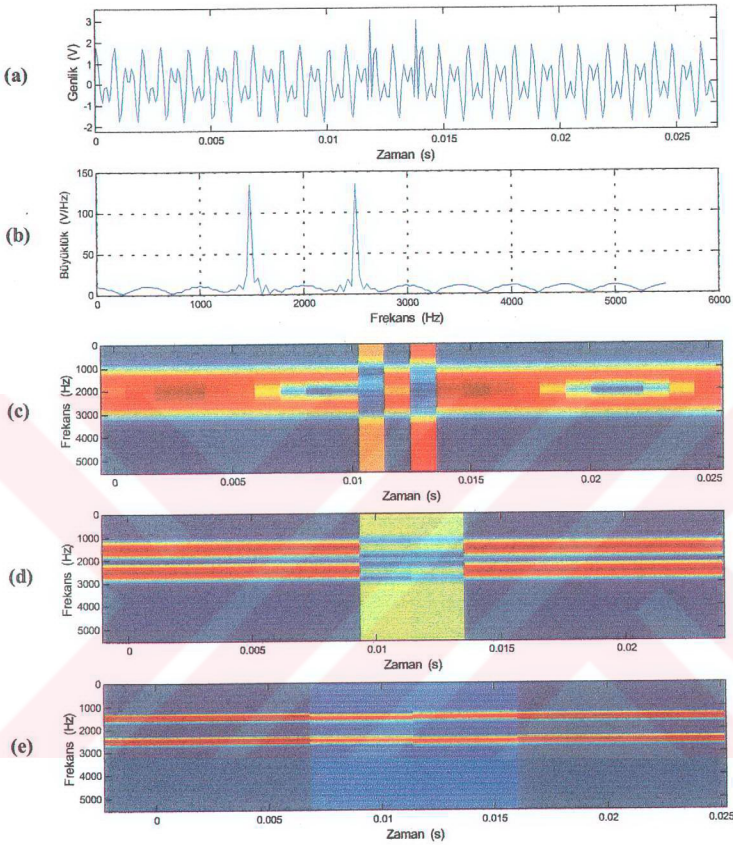
KZFD'nde pencere dar olduğunda zaman çözünürlüğümüz iyi, frekans çözünürlüğü kötü olacaktır; tersi durumda pencere geniş olduğunda ise frekans çözünürlüğü iyi, zaman çözünürlüğü kötü olacaktır. Pencere geniş tutulduğunda düşük frekanslı bileşenler, dar tutulduğunda ise yüksek frekanslı bileşenler başka bir deyişle ayrıntılar görülebilir. Görüldüğü gibi her iki tür bileşen de aynı anda istenilen çözünürlükte görülemeyeceği için seçim yapılması gerekmektedir. Bu nedenle pencere büyüklüğünün sabit olması KZFD'nün bir dezavantajıdır.

KZFD'nün bu dezavantajı *Heisenberg belirsizlik ilkesi* ile açıklanabilir [Goswami and Chan,1999]. Δt zaman penceresi ve $\Delta \omega$ frekans penceresi olmak üzere, $\Delta t \Delta \omega$ zaman-frekans penceresi (zaman-frekans çözünürlüğü) Denklem (3.27) ile verilen belirsizlik ilkesi ile sınırlıdır:

$$\Delta t \Delta \omega \geq \frac{1}{2} \quad (3.27)$$

Şekil 3.4'te matematiksel olarak üretilmiş $f(t)=A_1\sin(2\pi1500t)+A_2\sin(2\pi2500t)$ fonksiyonun değişik pencere uzunlukları ile alınmış KZFD'leri verilmiştir. İşarete, kayıt süresinin 12. ve 14. ms'lerinde başlayan ve 0.8 ms süren iki ani darbe yerleştirilmiştir. 1.6 ms uzunluğunda pencere ile KZFD alındığında iki darbeye rahatlıkla ayırt edilebilmekte, fakat frekans çözünürlüğümüz kötüleştiği için iki frekans bileşenini ayırt etmek zorlaşmaktadır. Pencere uzunluğu arttırıldığında darbeleri ayırt etmek zorlaşmaktadır, çünkü seçilen pencere uzunluğu darbeleri analiz edemeyecek kadar büyük olmaktadır. 3 ms ve 6 ms pencere uzunlukları ile darbeler ayırt edilememektedir. Fakat frekans çözünürlüğü iyileştiği için frekans bileşenleri kolayca ayırt edilebilmektedir.

Görüldüğü gibi KZFD'nde pencere büyüklüğünün analiz boyunca sabit olması istenilen zaman-frekans çözünürlüğünün elde edilememesine sebep olmaktadır. KZFD'nün sabit uzunluktaki pencerelerle analiz yapmasından kaynaklanan bu dezavantaj, araştırmacıları kısa süreli ani değişimleri kısa pencerelerle, uzun süreli değişimleri daha geniş pencerelerle analiz edebilen (başka bir deyişle düşük frekans bileşenleri için geniş zaman aralığı, alçak frekans bileşenleri için dar zaman aralığı sağlayan) yöntemler araştırmaya yöneltmiştir. Bir sonraki bölümde bu özelliği sağlayan ve yaygın olarak kullanılan yöntemlerden birisi olan dalgacık dönüşümü incelenmiştir.



Şekil 3.2 Matematiksel olarak üretilmiş bir işaretin (a) işlenmemiş hali, (b) Fourier dönüşümü, ve (c) 1.6 ms, (d) 3 ms, (e) 6 ms pencere uzunlukları ile KZFD grafikleri.

4 DALGACIK ANALİZİ

4.1 Giriş

Fourier dönüşümü temelli KZFD'nün sabit uzunluklu pencereler kullanması ve belirsizlik ilkesi kısıtlaması nedeniyle arzu edilen eş zamanlı zaman-frekans çözünürlüğü sağlayamaması gibi temel eksikleri araştırmacıları pencere uzunluğu değişken, zamanda yerelleşebilen ve frekansta seviye seviye çözünürlük sağlayabilen yeni dönüşümler araştırmaya yöneltmiştir. *Dalgacık dönüşümü* bunların en popülerlerindendir.

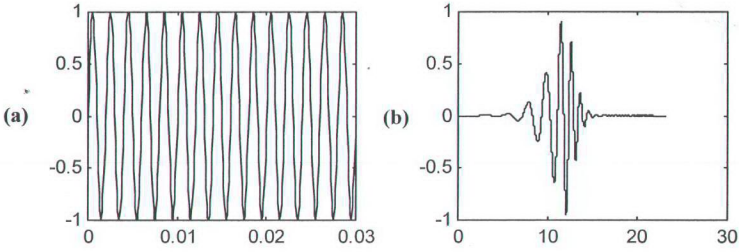
Dalgacıkların temeli 1900'lü yıllara kadar uzanır. *Dalgacık* kavramı ilk defa 1909 yılında Alfred Haar tarafından kullanılmıştır. *Dalgacık teorisi* ise ilk kez Jean Morlet tarafından ortaya konmuştur [Misiti et al.,2000].

Dalgacık analizi yöntemi ilk olarak Yves Meyer ve arkadaşları tarafından geliştirilmiştir. Kullanılan algoritmanın temeli Stephane Mallat'ın 1988'deki çalışmalarına dayanmaktadır. Dalgacık analizi 1990'lardan itibaren büyük ilgi toplamaya başlamış ve çeşitli alanlara uygulanmıştır. Bu yıllarda özellikle Ingrid Daubechies, Ronald Coifman, and Victor Wickerhauser dalgacık analizi ile ilgili çalışmaların öncüsü olmuştur [Daubechies,1996].

Dalgacık analizi işaretlerdeki süreksizliklerin ve saf frekans bileşenlerinin tespit edilmesi, ses ve titreşim gibi işaretlerin analizi, gürültüden arındırma, görüntü sıkıştırma gibi uygulamalarda kullanılmaktadır [Strang and Gilbert,1996], [Daubechies,1998], [Torrence and Compo,1998].

4.2 Dalgacıklar

Dalga, bir doğrultu boyunca salınım yapan bir fonksiyon olarak tanımlanabilir [Burrus et al.,1998]. Fourier dönüşümlerinde kullanılan sinüsler dalgalara bir örnektir. Dalgacık dönüşümde kullanılan dalgalar ise sınırlı bir uzunluğa sahiptirler, ve sonsuza kadar salınım yapmazlar. Şekil 4.1'de sinüs ve dalgacık fonksiyonları birlikte verilmiştir.



Şekil 4.1 (a) Sinüs dalgası ve (b) Daubechies dalgacığı.

Önceki bölümlerde anlatıldığı gibi bir işaret baz fonksiyonlarının lineer kombinasyonu şeklinde gösterilebilir. Fourier analizinde baz fonksiyonu olarak sinüsler kullanılmaktadır. Dalgacık analizinde ise bu amaçla dalgacıklar kullanılmaktadır. Dalgacık analizinin avantajı baz fonksiyonuna *ölçekleme* (dilation, scaling) ve *öteleme* (translation) şansı vermesidir. Dalgacık dönüşümünde işaretler analiz edilirken analiz penceresinin uzunluğu ve konumu kendiliğinden değişmektedir. KZFD’nde ise analiz penceresi kaydırılabildiği halde uzunluğu analiz boyunca sabit kalmaktaydı.

$\Psi_{b,a}(t)$ (ana dalgacık) baz fonksiyonu olmak üzere $f(t)$ fonksiyonunun sürekli dalgacık dönüşümü aşağıdaki gibi tanımlanır [Goswami and Chan,1999]:

$$W_{\psi} f(b, a) = \int_{-\infty}^{\infty} f(t) \overline{\Psi_{b,a}(t)} dt \quad (4.1)$$

Denklem (4.1) ile tanımlanmış olan dalgacık dönüşümünde dalgacık fonksiyonunun ifadesi açık değildir. Bunun nedeni analiz için kullanılacak dalgacık fonksiyonlarının çeşitli olmasıdır. Dalgacık dönüşümü için ters dönüşüm ise şu şekilde tanımlanır:

$$f(t) = \frac{1}{C_{\psi}} \int_{-\infty}^{\infty} db \int_{-\infty}^{\infty} \frac{1}{a^2} [W_{\psi} f(b, a)] \Psi_{b,a}(t) da \quad (4.2)$$

Yukarıdaki denklemdeki C_ψ sabiti seçilen dalgacığa bağlıdır ve kabul edilebilirlik koşulunu verir. Bir fonksiyonun dalgacık fonksiyonu olabilmesi için aşağıda verilen *kabul edilebilirlik* koşulunu (admissibility condition) sağlaması gereklidir [Goswami and Chan,1999]:

$$C_\psi = \int_{-\infty}^{\infty} \frac{|\Psi(\omega)|^2}{|\omega|} d\omega < \infty \quad (4.3)$$

Kabul edilebilirlik koşulu, ters dalgacık dönüşümü ile bir işaretin kayıpsız olarak tekrar elde edilebilmesini sağlar. Kabul edilebilirlik koşulunun sağlanması için dalgacık fonksiyonun Fourier dönüşümü $\Psi(\omega)$, $\omega=0$ noktasında sıfır olmalıdır:

$$\Psi(0) = \int_{-\infty}^{\infty} \Psi(t) dt = 0 \quad (4.4)$$

Denklem (4.3) ve (4.4) ile verilen koşullara göre dalgacık fonksiyonu *bant-geçiren* özelliği gösterecektir. Çünkü sıfır noktasında frekans bileşeni yoktur ve Fourier dönüşümü sonsuza gitmemektedir. Bunun yanı sıra Denklem (4.4) göz önüne alındığında dalgacık fonksiyonunun zaman tanım bölgesindeki ortalama değeri sıfır olacaktır. Ortalama değeri sıfır olan bir fonksiyon salınım yapacak, yani dalga özelliği gösterecektir. [Jawerth and Sweldens,1994]

Dalgacık fonksiyonunu oluşturabilmek için fonksiyonun *sıfır moment sayısı* yüksek olmalıdır. Sıfır moment sayısı kavramını açıklamak için Denklem (4.1) ile verilen dalgacık dönüşümünü (kolaylık olması için $b=0$ seçilerek) aşağıdaki gibi $t=0$ noktasında Taylor serisine açalım:

$$\gamma(a,0) = \frac{1}{\sqrt{a}} \left[\sum_{p=0}^n f^{(p)}(0) \int \frac{t}{p!} \Psi\left(\frac{t}{a}\right) dt + O(n+1) \right] \quad (4.5)$$

Seri açılımında $f^{(p)}$, $f(t)$ fonksiyonunun p . türevini ve $O(n+1)$ ise açılımın geri kalanını ifade etmektedir. Bu açılım için $\Psi(t)$ fonksiyonunun momentleri aşağıdaki gibi tanımlanır:

$$M_p = \int t^p \Psi(t) dt \quad (4.6)$$

Bu tanıma göre Denklem (4.5) ile verilen açılım tekrar düzenlensin:

$$\gamma(a,0) = \frac{1}{\sqrt{a}} \left[\sum_{p=0}^n f(0)M_0 a + \frac{f^{(1)}(0)}{1!} M_1 a^2 + \frac{f^{(2)}(0)}{2!} M_1 a^3 + \dots \right. \\ \left. \dots\dots + \frac{f^{(n)}(0)}{n!} M_1 a^{n+1} + O(a^{n+2}) \right] \quad (4.7)$$

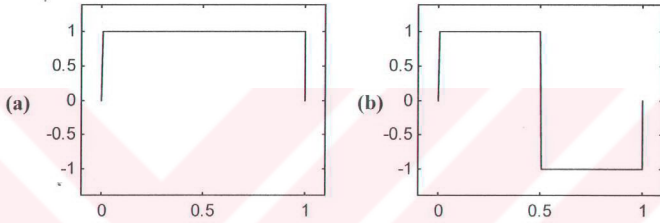
Kabul edilebilirlik koşulu nedeniyle 0. moment M_0 sıfırdır ve bu nedenle Denklem (4.7)'nin sağ tarafındaki ilk terim sıfır olacaktır. M_1 'den sonraki momentler den ne kadar çoğu sıfır olursa, açılım o oranda dalgacık katsayılarına yaklaşacaktır. Eğer bir dalgacığın sıfır moment sayısı N ise dalgacığın yaklaşma derecesi de N olacaktır. Dalgacık fonksiyonlarının özellikleri verilirken sıfır moment sayıları da verilir.

Yukarıda anlatılan koşulları sağlayan çeşitli dalgacık fonksiyonları vardır [Misiti et al.,2000]. Bunlardan bazılarının açık ifadeleri vardır ve analitik olarak hesaplanabilirler. Açık ifadeleri olmayanlar ise yineleme (iterasyon) ya da spektrum yöntemleri ile elde edilebilir [Strang and Gilbert,1996], [Sweldens AND Schroder,1996].

İşaret analizinde, incelenen işarete en uygun dalgacık bulunarak bu dalgacıkla çalışılabilir. Bu çalışmada MATLAB dalgacık alet kutusundaki dalgacıklar ile konuşma işaretlerinin dalgacık analizi yapılmıştır, bu denemeler sonucunda ve literatürde de belirtildiği gibi sürekli dalgacık dönüşümü için Morlet dalgacığı, ayrık dalgacık dönüşümü için ise Daubechies dalgacıklarının uygun olduğu görülmüştür. Bu nedenle bu çalışmada Morlet ve Daubechies dalgacıklar kullanılmıştır [Long and Datta,1996], [Torrence and Compo,1998], [Hau and Zhu,2000]. Her dalgacığın kendine has özellikleri vardır. MATLAB dalgacık alet kutusunda yer alan dalgacıklar ve özellikleri şunlardır:

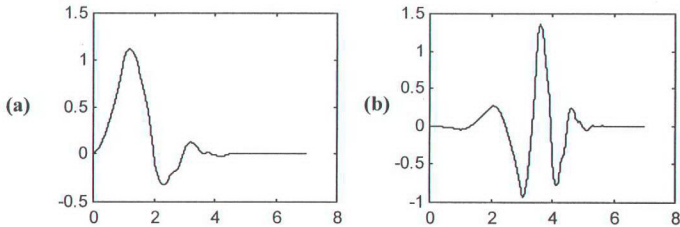
Kare (Haar) Dalgacığı : En eski ve basit dalgacıktır. Simetriktrir ve analiz ortogonaldır. Hem sürekli hem de ayrık dalgacık dönüşümü yapılabilir. Sıfır moment sayısı 1'dir. Şekil 4.2'de $\varphi(t)$ (ölçek) fonksiyonu ve $\psi(t)$ (dalgacık) fonksiyonu verilen kare dalgacığın açık ifadesi aşağıdaki gibidir:

Ölçek fonksiyonu Şekil 4.2 (a)	$\begin{cases} \varphi(t) = 1 & t \in [0,1] \\ \varphi(t) = 0 & t \notin [0,1] \end{cases}$
Dalgacık fonksiyonu Şekil 4.2 (b)	
	$\begin{cases} \psi(t) = 1 & t \in [0,0.5[\\ \psi(t) = -1 & t \in [0.5,1[\\ \psi(t) = 0 & t \notin [0,1[\end{cases}$



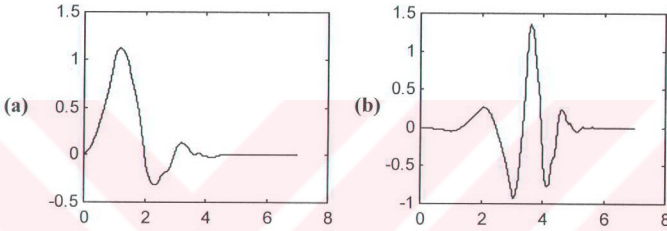
Şekil 4.2 (a) Kare dalgacığının ölçek fonksiyonu, (b) dalgacık fonksiyonu.

Daubechies dalgacıkları : Açık ifadeleri yoktur. Simetrik değildir, analiz ortogonaldır. N derece olmak üzere Daubechies dalgacığının uzunluğu $2N-1$ kadardır. 1. derece Daubechies dalgacığı (db1) kare dalgacığı vermektedir. Hem sürekli hem de ayrık dalgacık dönüşümü yapılabilir. Sıfır moment sayısı N 'dir. Şekil 4.3'te 'db4' (Daubechies 4. derece) dalgacığının ölçek ve dalgacık fonksiyonu verilmiştir. Görüldüğü gibi uzunlukları 7 birimdir.



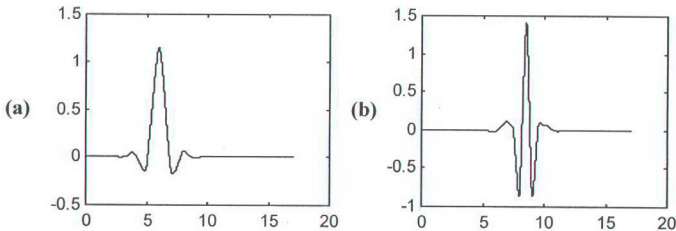
Şekil 4.3 (a) db4 (Daubechies) dalgacığının ölçek fonksiyonu, (b) dalgacık fonk.

Biortogonal dalgacıklar : Açık ifadeleri yoktur. Simetrikler ve analiz biortogondur. Biortogonal dalgacıklar ile sentez tam olarak mümkündür, fakat ortogonal dalgacıklar ile bu mümkün değildir. N_s sentez derecesi, N_a analiz derecesi olmak üzere biortogonal dalgacığının uzunluğu $2N+1$ kadardır. Hem sürekli hem de ayrık dalgacık dönüşümü yapılabilir. Sıfır moment sayısı ölçek fonksiyonu için N_a-1 'dir. Şekil 4.4'te 'bior5.5' (biortogonal 5. derece) dalgacığının ölçek ve dalgacık fonksiyonu verilmiştir.



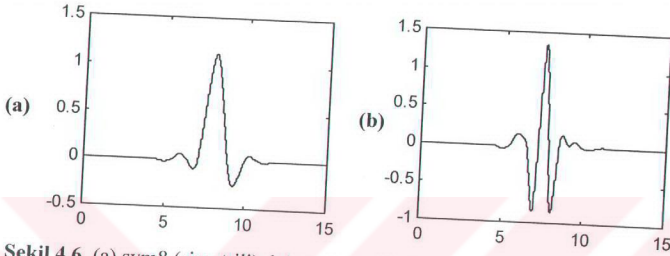
Şekil 4.4 (a) bior5.5 (biortogonal) dalgacığının ölçek fonksiyonu, (b) dalgacık fonksiyonu.

Coiflet dalgacıkları : Açık ifadeleri yoktur. Simetriye yakındır ve analiz ortogondur. N derece olmak üzere Coiflet dalgacığının uzunluğu $6N-1$ kadardır. Hem sürekli hem de ayrık dalgacık dönüşümü yapılabilir. Sıfır moment sayısı ölçek fonksiyonu için $2N$, dalgacık fonksiyonu için $2N-1$ 'dir. Şekil 4.5'te 'coif3' (Coiflet 3. derece) dalgacığının ölçek ve dalgacık fonksiyonu verilmiştir.



Şekil 4.5 (a) coif3 (coiflet) dalgacığının ölçek fonksiyonu, (b) dalgacık fonksiyonu.

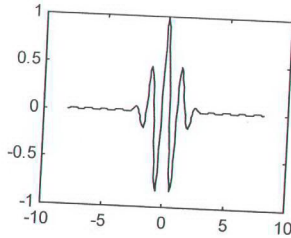
Simetrlili (Symlet) dalgacıkları : Açık ifadeleri yoktur. Simetriye yakındırlar ve analiz ortogonaldır. N derece olmak üzere simetrlili dalgacığın uzunluğu $2N-1$ kadardır. Hem sürekli hem de ayrık dalgacık dönüşümü yapılabilir. Sıfır moment sayısı ölçek fonksiyonu için N 'dir. Şekil 4.6'da 'sym4' (simetrlili 4. derece) dalgacığının ölçek ve dalgacık fonksiyonu verilmiştir.



Şekil 4.6 (a) sym8 (simetrlili) dalgacığının ölçek fonksiyonu, (b) dalgacık fonk.

Morlet dalgacığı : Simetriktir, analiz ortogonal ya da biortogonal değildir. Ölçek fonksiyonu olmadığı için ayrık dalgacık dönüşümü yapılamaz. Şekil 4.7'de dalgacık fonksiyonu verilen Morlet dalgacık fonksiyonun açık ifadesi aşağıdaki gibidir:

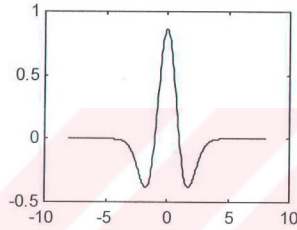
$$\psi(x) = e^{\frac{-x^2}{2}} \cos(5x) \quad (4.8)$$



Şekil 4.7 Morlet dalgacık fonksiyonu.

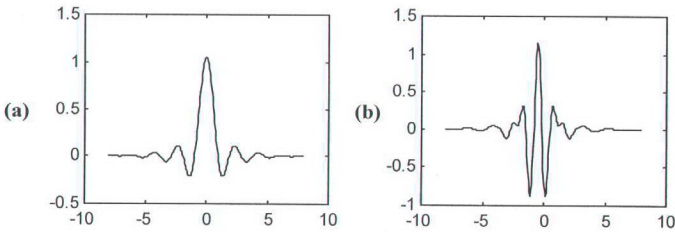
Meksika şapkası (Mexican Hat) dalgacı : Simetriktir, analiz ortogonal ya da biortogonal değildir. Ölçek fonksiyonu olmadığı için ayrık dalgacık dönüşümü yapılamaz. Şekil 4.8'de dalgacık fonksiyonu verilen Meksika şapkası dalgacık fonksiyonunun açık ifadesi aşağıdaki gibidir:

$$\psi(x) = \frac{2}{\sqrt{3}\sqrt[4]{\pi}} e^{\frac{-x^2}{x}} (1-x^2) \quad (4.9)$$



Şekil 4.8 Meksika şapkası dalgacık fonksiyonu.

Meyer dalgacı : Açık ifadesi yoktur. Simetriktir, analiz ortogonaldır. Hem sürekli hem de ayrık dalgacık dönüşümü yapılabilir. Şekil 4.9'da Meyer dalgacığının ölçek ve dalgacık fonksiyonu verilmiştir.



Şekil 4.9 (a) Meyer dalgacığının ölçek fonksiyonu, (b) dalgacık fonksiyonu.

Dalgacık analize yapılan giriş ve kullandığı dalgacık çeşitleri özellikleri ile kısaca sunulduktan sonra izleyen bölümlerde sürekli dalgacık ve ayrık dalgacık dönüşümleri anlatılacak, tanıtılan dalgacıkların bu dönüşümlerde nasıl kullanılabilecekleri incelenecektir.



4.3 Sürekli Dalgacık Dönüşümü

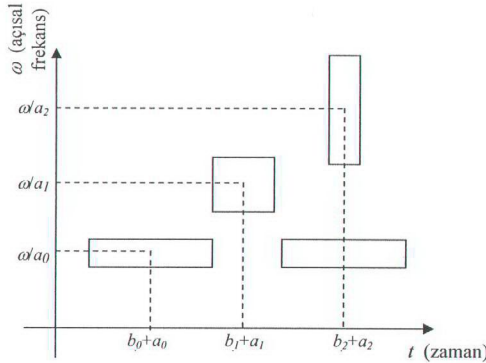
4.3.1 Giriş

Sürekli dalgacık dönüşümünde analiz fonksiyonu olarak önceden matematiksel olarak tanımlanmış dalgacıklar kullanılmaktadır. Dalgacıkların büyüklüğü (ölçeği) ve konumu değiştirilerek dalgacık katsayıları hesaplanmaktadır. Denklem (4.1) ile verilen sürekli dalgacık dönüşümde kullanılan baz fonksiyonu, $\Psi_{b,a}(t)$ ana dalgacık fonksiyonu aşağıdaki gibi ifade edilir:

$$\Psi_{b,a}(t) = \frac{1}{\sqrt{a}} \psi\left(\frac{t-b}{a}\right) \quad a > 0 \quad (4.10)$$

Bu fonksiyon analiz sırasında kullanılan pencerenin uzunluğunu ve konumunu belirler. Ana dalgacık fonksiyonundaki parametrelerden b öteleme, a ise ölçek parametresidir. Ölçek, dalgacığın büyüklüğünü, öteleme ise dalgacığın zaman eksenindeki yerini belirler. Normalizasyon faktörü $a^{-1/2}$, $\|\Psi_{b,a}\| = \|\Psi\|$ norm eşitliğini sağlamak için konulmuştur.

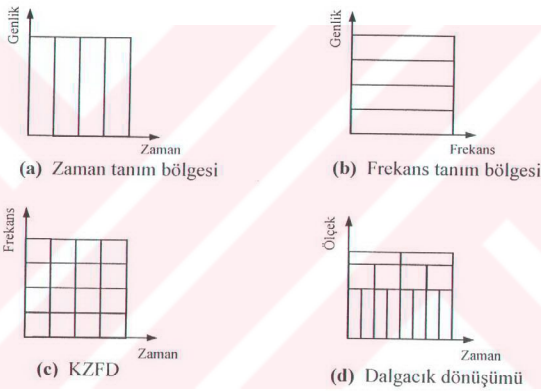
KZFD anlatılırken pencere uzunluğunun analiz boyunca sabit olduğu söylenmişti. Şekil 4.10'da görüldüğü gibi dalgacık dönüşümünde pencere uzunluğu değişkendir.



Şekil 4.10 Dalgacık dönüşümde zaman-frekans penceresi.

Dalgacık dönüşümünün KZFD'ne göre en önemli üstünlüğü, pencere uzunluğu ve yerinin analiz sırasında kendiliğinden değişken olmasıdır. Bu şekilde dalgacık dönüşümü ile çoklu-çözünürlük analizi yapılabilir

Şekil 4.11'de değişik dönüşümler için çözünürlükler verilmiştir. İşlenmemiş işarette frekans çözünürlüğü sıfırdır ve frekans bilgisi yoktur. Fourier dönüşümünde ise zaman çözünürlüğü sıfırdır, zaman bilgisi yoktur. KZFD hem zaman hem de frekans çözünürlüğü verir, fakat her ikisi için de çözünürlükler sabittir. Dalgacık dönüşümünde ise çoklu-çözünürlük söz konusudur. Alçak frekanslar için geniş zaman aralığı, yüksek frekanslar için dar zaman aralığı sağlamaktadır.

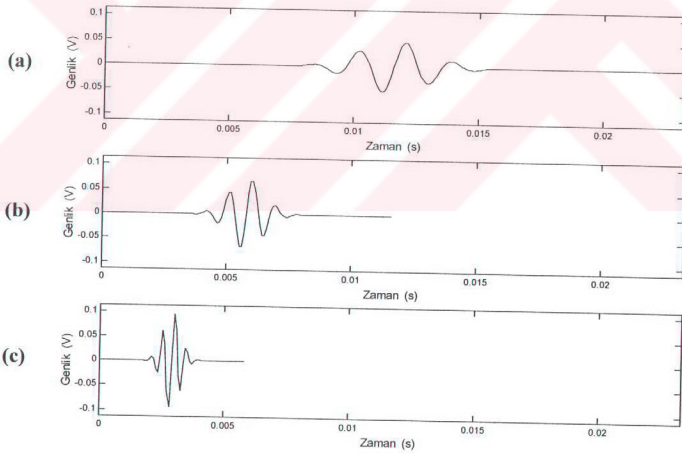


Şekil 4.11 (a) İşlenmemiş işaret, (b) Fourier dönüşümü, (c) KZFD, (d) dalgacık dönüşümü için zaman-frekans çözünürlükleri.

4.3.2 Ölçekleme (Dilation, Scaling) ve Öteleme (Translation)

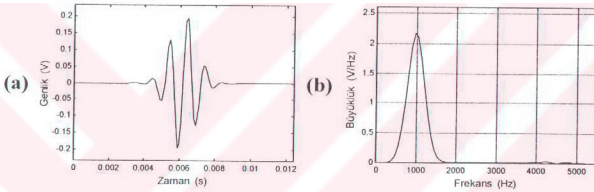
Dalgacık dönüşümünün baz fonksiyonu olan ana dalgacık fonksiyonu Denklem (4.10) ile verilmişti. Bu denklemde a ölçek, b ise öteleme parametresidir.

Ölçek parametresini daha iyi anlayabilmek için, dalgacık dönüşümü en temel dönüşüm yöntemi olan Fourier dönüşümü ile karşılaştırılabilir. Fourier dönüşümünün baz fonksiyonu olan sinüs fonksiyonundaki ω açısal frekansı, fonksiyonun genişliğini (periyodunu) belirler ve değeri arttıkça sinüs fonksiyonu daralır. Dalgacık fonksiyonunda ise ω 'ya (tersi olarak) karşı gelen a ölçek parametresi fonksiyonun (dalgacığın) hem genişliğini hem de genliğini (yani büyüklüğünü) belirler. Şekil 4.12'de ölçek parametresi ile dalgacık fonksiyonunun büyüklüğü arasındaki ilişki Daubechies dalgacıkları ile gösterilmiştir. Görüldüğü gibi ölçek küçüldükçe, dalgacığın uzunluğu azalmakta, genliği ise artmaktadır.



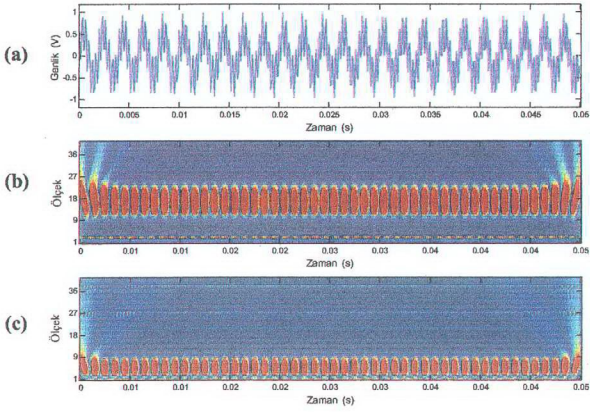
Şekil 4.12 (a) $a=16$, (b) $a=4$, (c) $a=1$ ölçek değerleri için Daubechies dalgacıkları.

Belirtildiği gibi ölçek parametresi frekans ile ters orantılıdır. İkisi arasındaki ilişkiyi bulmak için dalgacık fonksiyonun frekans spektrumu incelenebilir. Şekil 4.13'te Morlet dalgacı 8.5 ölçek değeri ile gösterilmiştir. Frekans spektrumundan görüldüğü gibi bandının orta noktası 1 kHz'dir. Dolayısıyla 1 kHz frekans bileşenleri dalgacık dönüşümünde 8.5 ölçek değerine karşılık gelecektir. Bu ilişki ters orantılı olarak değişmektedir, örneğin 500 Hz frekansına 17 ölçeği karşılık gelecektir. Bu değerler 11025 Hz örnekleme frekansı için hesaplanmıştır. Örnekleme frekansı değiştiğinde frekans-ölçek ilişkisi de değişecektir. Morlet dalgacığına ilişkin ölçek-frekans karşılıkları Bölüm 4.3.4'te Çizelge 4.1'de verilmiştir.



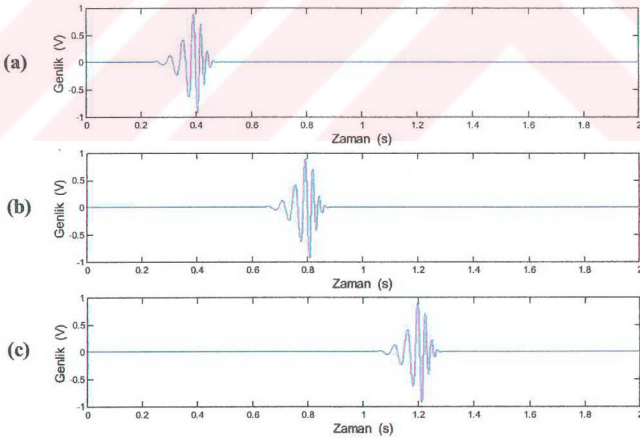
Şekil 4.13 (a) $a=8.5$ ölçek değeri için Morlet dalgacık fonksiyonu ve (b) Fourier dönüşümü.

Ölçek-frekans ilişkisi dalgacıktan dalgacığa farklılık gösterebilir. Yukarıda verilmiş olan ilişki Morlet dalgacı için geçerlidir. Ölçek-frekans ilişkisi aynı örnekleme frekansında çalışılsa bile başka dalgacıklar için değişebilir. Örneğin Şekil 4.14'te matematiksel olarak üretilmiş $f(t)=A_1\sin(2\pi 500t)+A_2\sin(2\pi 3000t)$ fonksiyonu ve dalgacık dönüşümleri verilmiştir. Görüldüğü gibi iki frekans (500 Hz ve 3 kHz) bileşeni, Morlet ve Meksika şapkası dalgacıkları ile yapılmış iki ayrı dalgacık dönüşümünde farklı ölçek değerlerinde belirmiştir.



Şekil 4.14 (a) 500 Hz ve 3 kHz frekans bileşenlerine sahip bir işaretin işlenmemiş hali, (b) Morlet dalgacığı ve (c) Meksika şapkası dalgacığı ile dalgacık dönüşümleri.

Öteleme parametresi, dalgacığın zaman eksenindeki yerini belirlemektedir. $f(t)$ gibi bir fonksiyonun b kadar ötelenmesi $f(t-b)$ ile ifade edilir. Şekil 4.15'te Daubechies dalgacığının ötelenmesi gösterilmiştir.



Şekil 4.15 (a) $b=0.4$, (b) $b=0.8$, (c) $b=1.2$ öteleme değerleri ile db12 (Daubechies) dalgacığı.

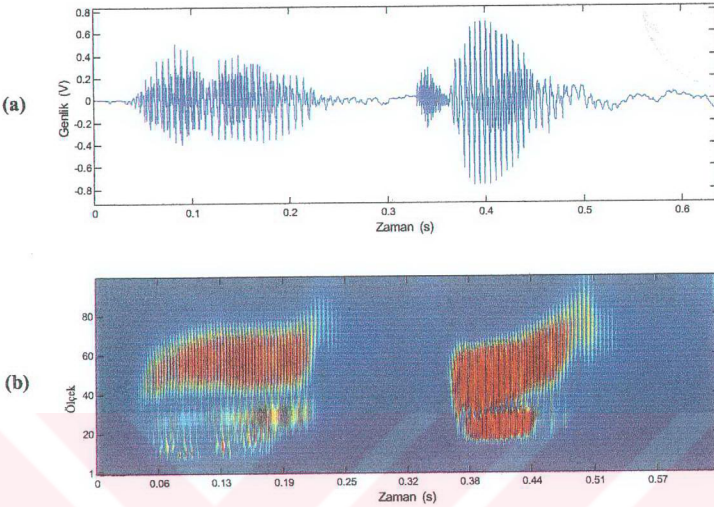
Öteleme ve ölçekleme anlatılırken görüldüğü gibi Şekil 4.12’de b öteleme, Şekil 4.15’te ise a ölçek parametresi sabit tutulmuştur. Gerçekte bu parametreler Bölüm 4.3.3’te anlatıldığı gibi belli bir sırayla eşgüdümlü olarak değişir.

4.3.3 Sürekli Dalgacık Katsayılarının Hesaplanması

Sürekli dalgacık dönüşümünü oluşturan katsayılar hesaplanırken sırasıyla aşağıdaki adımlar izlenir:

- Bir dalgacık seçilir ve seçilen pencere için sürekli dalgacık katsayıları hesaplanır.
- Dalgacık, işaret boyunca kaydırılır ve karşı gelen pencere için yeni katsayılar hesaplanır.
- İşaretin sonuna gelinceye kadar tüm pencereler için sürekli dalgacık katsayıları hesaplanır.
- Dalgacığın ölçeği (pencere büyüklüğü) değiştirilir ve ilk üç işlem tekrarlanır.
- Diğer ölçekler için de aynı yolla katsayılar hesaplanır.

Şekil 4.16’da “altı” kelimesinin ses kaydından oluşan işaretin Morlet dalgacığı ile elde edilmiş *sürekli dalgacık dönüşümü* görülmektedir. İşaret toplam 7000 elemanlı bir diziden oluşmaktadır. Sürekli dalgacık dönüşümü katsayıları 1-100 arasındaki ölçek değerleri için hesaplanmıştır ve bunun sonucunda 100×7000 boyutunda bir matris elde edilmiştir. Matrisin her satırı bir ölçek değerine karşılık gelmektedir. Sütun sayısı ise zaman veya yer (lokasyon) bilgisi ile ilgilidir.



Şekil 4.16 (a) “Altı” kelimesinin işlenmemiş hali, (b) sürekli dalgacık dönüşümü.

Sürekli dalgacık dönüşümünde her ölçek değeri için bütün zaman eksenini boyunca katsayılar hesaplanmaktadır. Sonuçta oluşan matris oldukça büyük boyutlardadır. Bu hem işlem süresini uzatmakta, hem de çok fazla sayıda veri saklanması nedeniyle olmaktadır. Bu sebeple ayrık dalgacık dönüşümü geliştirilmiştir. Bölüm 4.4’te anlatılacak olan dalgacık dönüşümünde hem ölçek hem zaman eksenini uygun şekilde örneklenerek sadece belirli noktalarda dalgacık katsayıları hesaplanmaktadır. Bu nedenle elde edilen bu dalgacık dönüşümü ayrık dalgacık dönüşümü olarak adlandırılır.

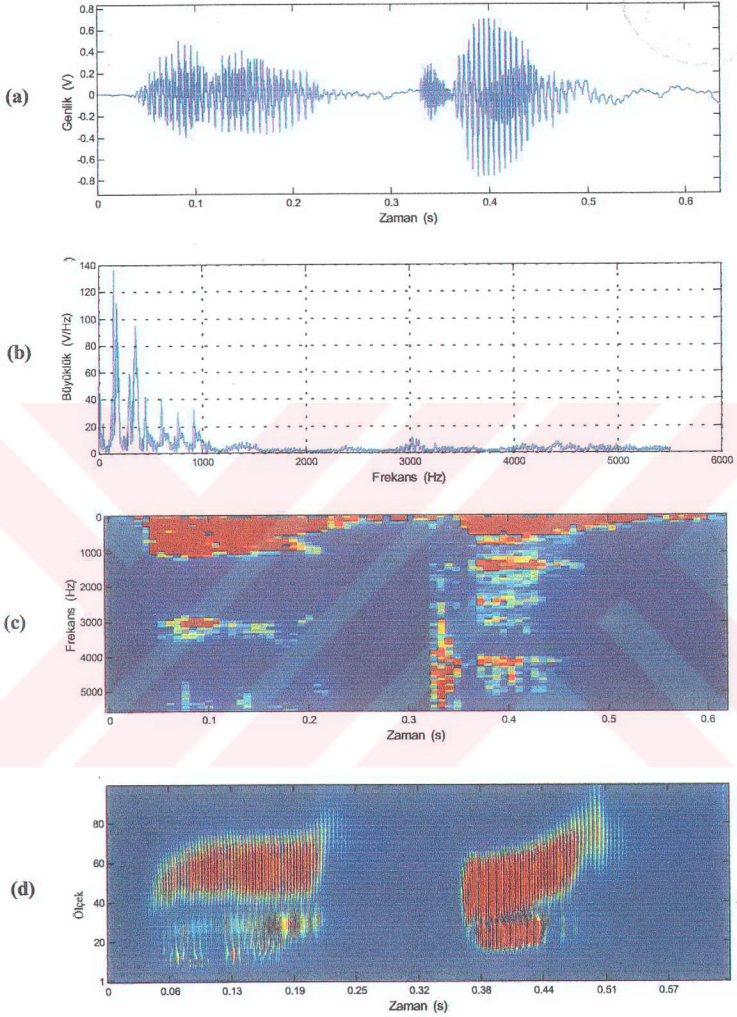
4.3.4 Konuşma İşaretlerinin Sürekli Dalgacık Dönüşümü ile İncelenmesi

Sürekli dalgacık dönüşümü zaman-ölçek analizi yapmaktadır. Küçük ölçekler yüksek frekanslı, büyük ölçekler ise alçak frekanslı bileşenleri temsil etmektedir. Sürekli dalgacık dönüşümünde belirli örnekleme frekansları için hangi ölçek değerinin hangi frekans değerine karşılık geldiği Çizelge 4.1’de verilmiştir ve bu değerler Morlet dalgacığı için geçerlidir. Çizelgedeki değerler frekans içerikleri bilinen matematiksel olarak üretilmiş çeşitli işaretlere Fourier ve dalgacık dönüşümleri uygulanarak elde edilmiştir.

Frekans	Örnekleme Frekansı	Ölçek
2000 Hz	11025 Hz	4,25
2000 Hz	22050 Hz	8,5
1000 Hz	11025 Hz	8,5
1000 Hz	22050 Hz	17
500 Hz	11025 Hz	17
500 Hz	22050 Hz	34
250 Hz	11025 Hz	34
250 Hz	22050 Hz	68
125 Hz	11025 Hz	68
125 Hz	22050 Hz	136

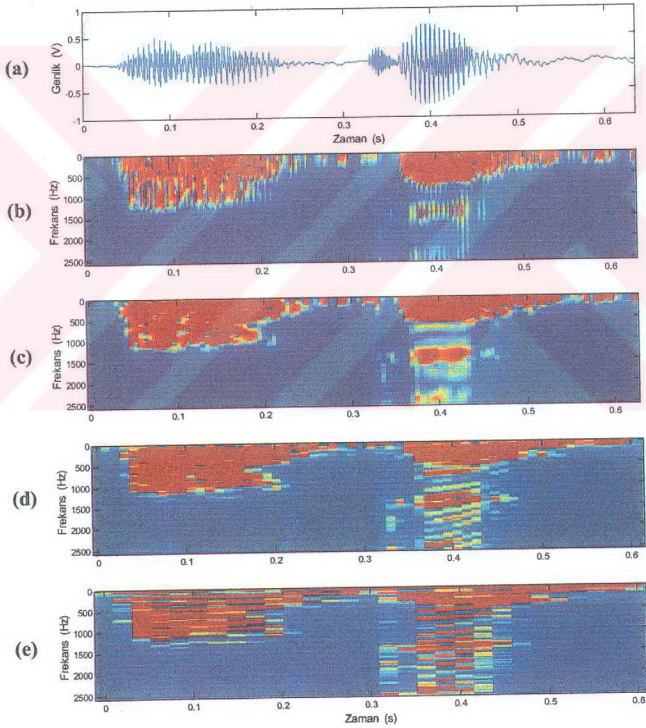
Çizelge 4.1 Morlet dalgacığı için ölçek-frekans değerleri karşılıkları.

Konuşma işaretleri incelenirken tüm frekans spektrumunu kaplamak ve gereksiz hesaplamadan kaçınmak amacıyla, yukarıdaki çizelgede verilen ölçek-frekans değerleri dikkate alınarak 1-100 arası ölçek değerleri seçilmiştir. Şekil 4.17’de “altı” kelimesinin Fourier dönüşümü, KZFD ve sürekli dalgacık dönüşümü verilmiştir. İşaretin Fourier dönüşümünde görüldüğü gibi 1000 Hz’in altında iki baskın frekans bileşeni vardır. Dalgacık dönüşümünde bu iki bileşen rahatlıkla ayırt edilebilirken, KZFD’nde bu ayırım o kadar açık değildir. Ayrıca dalgacık dönüşümüyle elde edilen grafik daha fazla bilgi taşımaktadır. İşaretin işlenmemiş halinde görülen tepeler KZFD’nde görülemezken, dalgacık dönüşümde görülebilmektedir ve frekans çözünürlüğü daha iyidir.



Şekil 4.17 (a) “Altı” kelimesinin işlememiş hali, (b) Fourier dönüşümü, (c) KZFD ve (d) sürekli dalgacık dönüşümü.

KZFD grafiğini daha iyi inceleyebilmek için frekans eksenini daha dar bir aralıkta çizdirilebilir. Şekil 4.18’de verilmiş olan KZFD grafikleri 0-2500 Hz frekans aralığı için çizdirilmiştir. Şekilde “altı” kelimesine ait ses kayıt örneğinin değişik pencere büyüklükleri için KZFD katsayılarının grafikleri verilmiştir. Grafiklerden görüldüğü gibi pencere büyüklüğü küçüldükçe zaman çözünürlüğü iyileşmekte fakat frekans çözünürlüğü kötüleşmektedir. Şekil 4.18 (b)’de görüldüğü gibi 4 ms pencere uzunluğu ile elde edilmiş ilk grafikte frekans bileşenleri ayırt edilemiyor. Şekil 4.18 (d) ve (e)’de gösterildiği gibi pencere büyüklüğü arttırıldığında ise tam tersi zaman çözünürlüğü azalırken frekans çözünürlüğü iyileşmektedir ve 0.3 – 0.5 s arasında oluşan iki ayrı frekans bileşeni ayırt edilebilmektedir.

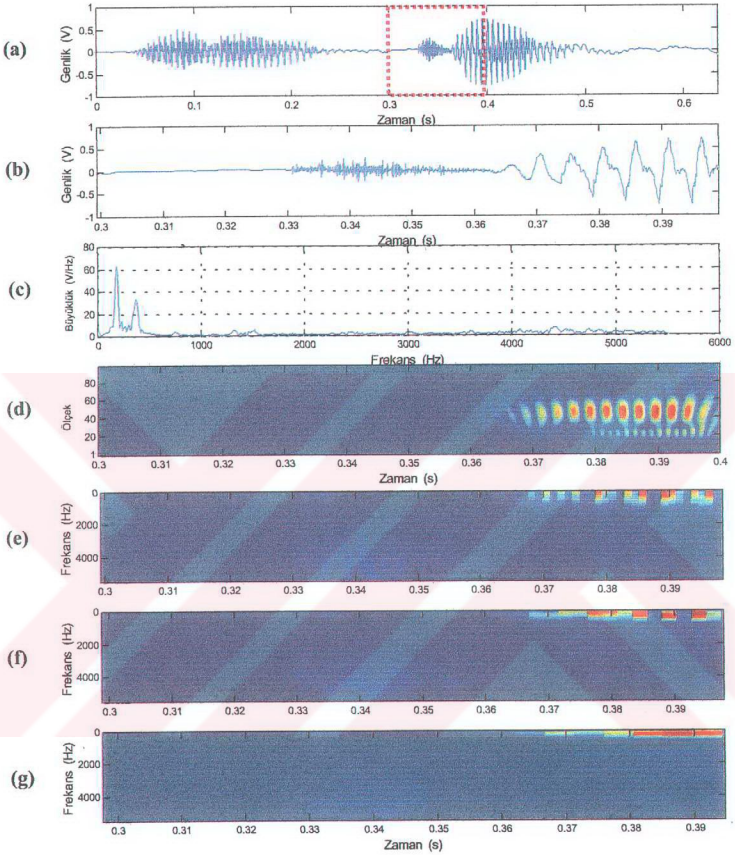


Şekil 4.18 (a) “Altı” kelimesinin işlenmemiş hali ve (b) 4 ms, (c) 8 ms, (d) 16 ms, (e) 32 ms pencere uzunlukları ile KZFD’leri.

Yukarıdaki örnekte anlatıldığı gibi, KZFD’nde pencere uzunluğunun analiz boyunca sabit olması nedeniyle analiz yaparken pencere uzunluğu önceden uygun şekilde belirlenmek zorundadır. Ancak dalgacık dönüşümde böyle bir zorunluluk yoktur. Daha önce de belirtildiği gibi pencere büyüklüğü analiz boyunca kendiliğinden değişmektedir.

Daha önce Şekil 4.17 incelenirken “altı” kelimesinin işlenmemiş halinde gözüken öbeklerinin dalgacık dönüşümde rahatça görülebildiği halde KZFD’nde hemen hemen hiç görülemediği belirtilmişti. Bunun nedeni KZFD’ndeki pencere uzunluğunun bu tepeler için fazla büyük olmasıdır. Pencere uzunluğunu daha fazla küçültmek frekans çözünürlüğü kötüleşeceği için mümkün değildir. Bu sebeple dalgacık dönüşümü yerel ayrıntıları incelemek için daha uygun bir yöntemdir. Bu etki işaretin bir bölümüne yakından bakılarak daha iyi görülebilir.

Şekil 4.19’da aynı işaretin sadece 300-400 ms aralığı incelenmektedir. KZFD değişik pencere uzunlukları için hesaplanmıştır. Dalgacık dönüşümünde frekans tepeleri belirgindir, ayrıca Fourier dönüşümünde gözüken iki frekans bileşeni de ayırt edilebilmektedir. KZFD’nde ise her iki özellik de görülememektedir. Sadece pencere genişliğinin en küçük olduğu ilk grafikte tepeler görülebilmektedir. Fakat bunlar da dalgacık dönüşümünde olduğu gibi keskin değildir. Frekans bileşenleri ise hiçbirinde ayırt edilememektedir.



Şekil 4.19 (a) “Altı” kelimesinin işlenmemiş hali, (b) 300-400 ms aralığının işlenmemiş hali, (c) Fourier dönüşümü, (d) sürekli dalgacık dönüşümü ve (e) 1.5 ms, (f) 3 ms, (g) 6 ms pencere uzunlukları ile KZFD’leri.

4.4 Ayırık Dalgacık Dönüşümü

4.4.1 Giriş

Sürekli dalgacık dönüşümünde tüm ölçek değerleri için dalgacık katsayıları bütün zaman eksenini boyunca hesaplanmaktadır. Hesaplama işlem sayısını azaltmak amacıyla araştırmacılarla ayırık dalgacık dönüşümü geliştirilmiştir. Ayırık dalgacık dönüşümünde ölçek değerleri *dyadic* (2'nin kuvvetleri şeklinde) seçilerek yalnızca belirli noktalarda dalgacık katsayıları hesaplanmakta ve böylece işlem sayısı azaltılmaktadır.

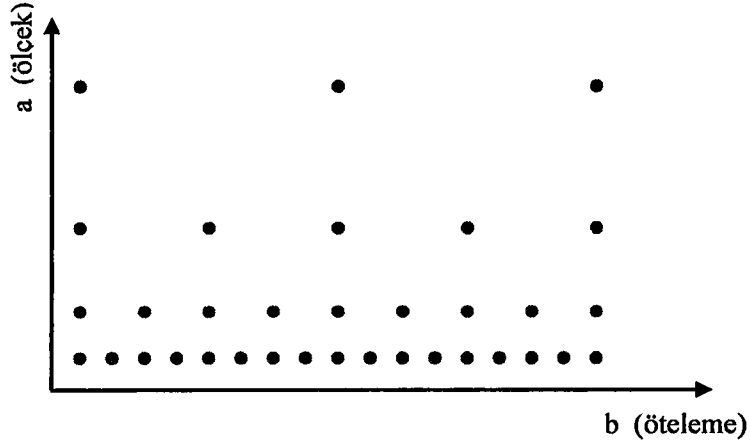
Denklem (4.10) ile verilmiş olan ana dalgacık denkleminde a ölçek katsayısı yerine 2^j ve b öteleme katsayısı yerine $k2^j$ konulur ve denklem tekrar düzenlenirse ayırık dönüşüm için dalgacık fonksiyonu bulunabilir:

$$\begin{aligned}\Psi_{b,a}(t) &= \frac{1}{\sqrt{a}} \psi\left(\frac{t-b}{a}\right) \\ &= \frac{1}{\sqrt{2^{-j}}} \psi\left(\frac{t-2^{-j}}{2^{-j}}\right) \\ \psi_{j,k}(t) &= 2^{j/2} \psi(2^j t - k) \quad j, k \in \mathbb{Z}\end{aligned}\quad (4.11)$$

Bu dalgacık fonksiyonu kullanılarak, $f(t)$ işaretinin Denklem (3.1) ile verilen seri açılımı aşağıdaki gibi ifade edilebilir:

$$f(t) = \sum_{j,k} \alpha_{j,k} 2^{j/2} \psi(2^j t - k) \quad (4.12)$$

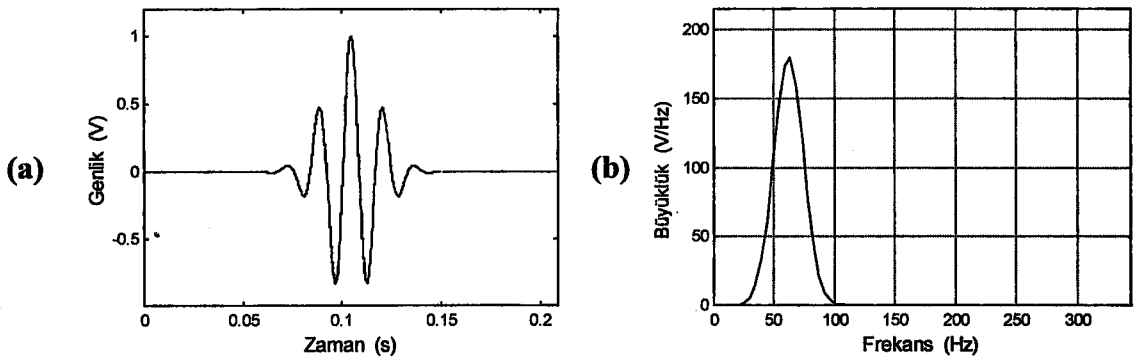
j indisi dalgacığın ölçeğini (büyüklüğünü) belirlemektedir, j arttıkça ölçek azalmaktadır. Ayırık dalgacık dönüşümü seviye seviye yapılmaktadır. Dönüşüm seviyesi ilerledikçe dalgacığın ölçeği 2 kat artmaktadır. Böylece ölçek eksenini 2'nin kuvvetleri şeklinde örneklenmiş olur. Görüldüğü gibi dönüşüm seviyesi, j indisi ile ters orantılıdır. Zaman eksenini de aynı şekilde örneklenmektedir. Seviye her arttırıldığında dalgacık fonksiyonu iki kat genişlemekte ve hesaplanacak katsayı miktarı yarıya düşmektedir. Şekil 4.20'de ayırık dalgacık dönüşümde ölçek ve zaman eksenlerinin örneklenmesi grafik halinde gösterilmiştir.



Şekil 4.20 Ayırık dalgacık dönüşümünde ölçek ve zaman eksenlerinin örneklenmesi.

4.4.2 Ölçek ve Dalgacık Fonksiyonları

Dalgacık fonksiyonları anlatılırken bu fonksiyonların band geçiren karakteristik gösterdiği söylenmişti. Şekil 4.21’de Morlet dalgacığı ve frekans spektrumu verilmiştir. Dalgacık fonksiyonunun bu bant-geçiren özelliği kullanılarak analiz edilen işaretin tüm frekans spektrumu değişik ölçekteki dalgacıklarla kaplanabilir.

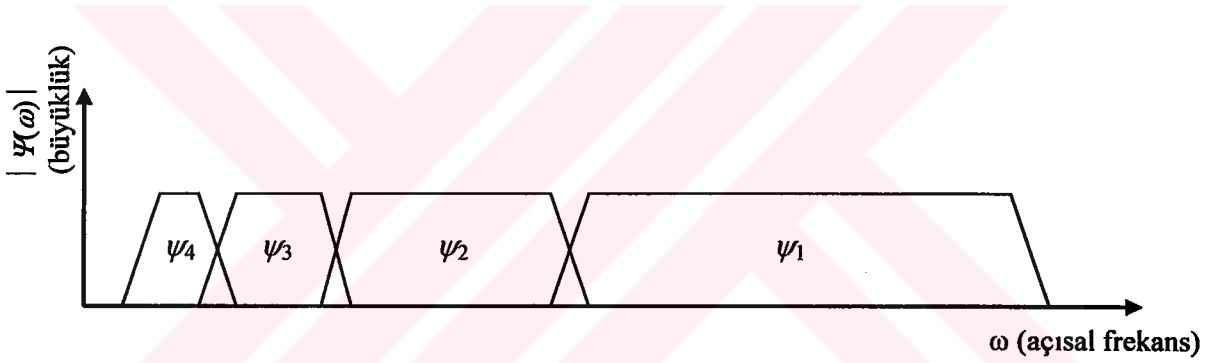


Şekil 4.21 (a) Morlet dalgacığı ve (b) Fourier dönüşümü.

Fourier dönüşümünün aşağıda Denklem (4.13) ile verilen özelliğine göre, bir işareti zaman tanım bölgesinde genişletme frekans tanım bölgesinde aynı oranda daraltmaya karşılık gelmektedir [Goswami and Chan,1999].

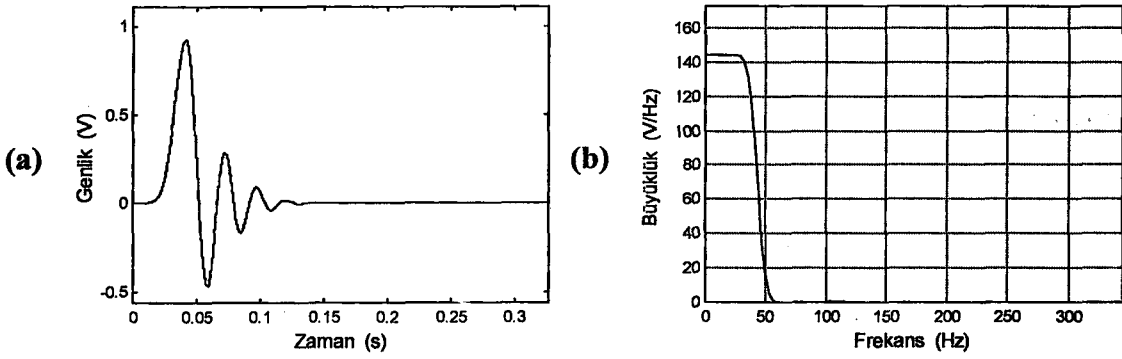
$$F\{f(at)\} = \frac{1}{|a|} F\left(\frac{\omega}{a}\right) \quad (4.13)$$

Daha önce değinildiği gibi ayrık dalgacık dönüşümünde seviye her arttırıldığında dalgacık 2 kat genişlemekte ve Denklem (4.13)'e göre frekans bandı yarıya düşmektedir. Bu şekilde dönüşüm seviyesi arttırılarak (j indisi azaltılarak), Denklem (4.12)'de tanımlanmış olan açılımdaki gibi bir işaretin tüm frekans spektrumu bantlar halinde kaplanabilir. Şekil 4.22'de görüldüğü gibi ayrık dalgacık seviyesi arttıkça dalgacık fonksiyonun bant genişliği yarıya düşmektedir.



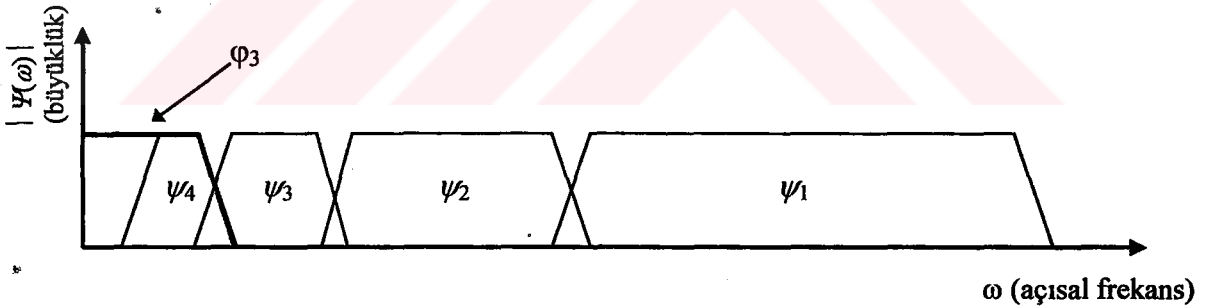
Şekil 4.22 Ayrık dalgacık dönüşümünde ilk 4 seviye için dalgacık fonksiyonu ile frekans bandının kaplanması.

Ancak bu yöntemde tüm spektrumu kaplamak için sonsuz sayıda dalgacık gerekmektedir (Başka bir deyişle analizi sonsuza kadar götürmek gereklidir). Çünkü her seferinde dalgacık fonksiyonunun bantı yarıya düşmektedir ve sıfıra ulaşmak mümkün olmayacaktır. Bu sorunu gidermek ve analizde kullanılacak dalgacık fonksiyonu sayısını düşük tutmak için alçak-geçiren frekans karakteristiği olan ölçek fonksiyonu geliştirilmiştir. Şekil 4.23'te ölçek fonksiyonu Fourier dönüşümü ile verilmiştir.



Şekil 4.23 (a) db13 (Daubechies) dalgacığının ölçek fonksiyonu ve (b) Fourier dönüşümü.

Ölçek fonksiyonunun alçak-geçiren özelliği kullanılarak işaretin tüm frekans bandı kullanılan dalgacık fonksiyonu sayısını sonsuza götürmeye gerek kalmadan kaplanabilir. Belirli bir seviyede analiz kesilerek geri kalan frekans bandı ölçek fonksiyonu tarafından kaplanabilir. Şekil 4.24'te görüldüğü gibi 3. seviyeye kadar analiz dalgacık fonksiyonu tarafından götürülür ve, geri kalan kısım ise ölçek fonksiyonu tarafından tamamlanır ($f \approx \varphi_4 + \psi_3 + \psi_2 + \psi_1$)¹.



Şekil 4.24 Ayırık dalgacık dönüşümünde ilk 3 seviye için ölçek ve dalgacık fonksiyonları ile frekans bandının kaplanması.

Yukarıda değinilen sorunlar nedeniyle ayırık dalgacık dönüşümünde $\psi_{j,k}(t)$ dalgacık fonksiyonun yanında, ikinci bir baz fonksiyonu olarak $\varphi_k(t)$ ölçek fonksiyonu da tanımlanır. Bu iki baz fonksiyonu kullanılarak $f(t)$ işaretinin ayırık dalgacık açılımı şu şekilde ifade edilebilir [Burrus et al,1998]:

¹ φ ve ψ fonksiyonların frekans bandlarının toplamı, f fonksiyonunun frekans spektrumunu verir.

$$f(t) = \sum_{k=-\infty}^{\infty} c(k) \varphi_k(t) + \sum_{j=0}^{\infty} \sum_{k=-\infty}^{\infty} d(j,k) \psi_{j,k}(t) \quad (4.14)$$

Denklem daha genel bir biçimde aşağıdaki gibi yazılabilir:

$$f(t) = \sum_k c_{j_0}(k) 2^{j_0/2} \varphi(2^{j_0} t - k) + \sum_k \sum_{j=j_0}^{\infty} d_j(k) 2^{j/2} \psi_{j,k}(2^j t - k) \quad (4.15)$$

Yukarıdaki açılımdaki katsayılar ayrık dalgacık dönüşümü katsayılarıdır. Burada $d(j,k)$ *ayrıntı* (detail) katsayıları, $c(k)$ ise *yaklaşıklık* (approximation) katsayıları olarak adlandırılmaktadır. j indisi ölçeği (frekans bilgisini), k indisi ise konumu (zaman bilgisini) vermektedir. Açılımda yaklaşıklık katsayıları sadece j_0 ölçeği (son seviyeye ait ölçek değeri) için hesaplanmakta, ayrıntı katsayıları ise j_0 'dan sonraki tüm ölçekler (son seviyeye kadarki tüm ölçek değerleri) için hesaplanmaktadır.

$\varphi_k(t)$ ölçek ve $\psi_{j,k}(t)$ dalgacık fonksiyonları aşağıda verilen ortogonallik koşulunu sağlıyorsa dönüşüm katsayılarını hesaplamak kolaylaşacaktır. Bu fonksiyonlar için ortogonallik koşulu aşağıdaki gibi verilir:

$$\langle \varphi_k(t), \psi_l(t) \rangle = \int \varphi_k(t) \overline{\psi_l(t)} dt = \delta_{k,l} \quad (4.16)$$

Denklem (4.14)'ün her iki tarafı ayrı ayrı $\varphi_k(t)$ ve $\psi_{j,k}(t)$ ile çarpılarak ve dalgacık baz kümelerinin ortogonallik özelliği kullanılarak açılım katsayıları aşağıdaki gibi hesaplanabilir:

$$c_j(k) = \langle f(t), \varphi_k(t) \rangle = \int f(t) \overline{\varphi_k(t)} dt \quad (4.17)$$

$$d_{j,k}(k) = \langle f(t), \psi_{j,k}(t) \rangle = \int f(t) \overline{\psi_{j,k}(t)} dt \quad (4.18)$$

4.4.3 Alt-Bantlara Ayırma ve Ayırık Dalgacık Analizi

Denklem (4.15) ile verilmiş olan ayırık dalgacık dönüşümünde amaç $d_j(k)$ ve $c(k)$ katsayılarını bulmaktır. Bu katsayıları bulmak için alçak-geçiren ve yüksek-geçiren filtrelerin kullanıldığı *alt-bantlara ayırma* metodu kullanılmaktadır. $c(k)$ yaklaşıklık katsayıları *alçak-geçiren* filtreler ile elde edilmektedir. Bir işaretin ana hatları alçak frekans bileşenlerinde gizlidir. İşaret alçak-geçiren filtreden geçirildiğinde hala gerçek işarete benzemektedir. $c(k)$ katsayılarına yaklaşıklık katsayıları olarak adlandırılmasının sebebi budur. Bunun tersi olarak $d_j(k)$ ayrıntı katsayıları yüksek frekans bileşenlerini (başka bir deyişle hızlı değişimleri, ayrıntıları) göstermektedir ve *yüksek-geçiren* filtreler ile elde edilmektedir.

Sürekli dalgacık dönüşümde farklı ölçek çözünürlükleri elde etmek için dalgacığın ölçeği değiştirilmektedir. Ayırık dalgacık dönüşümünde ise farklı ölçek çözünürlükleri sağlamak için alçak-geçiren ve yüksek-geçiren filtreler kullanılmaktadır. Değişik kesim frekanslı filtreler ile değişik ölçek çözünürlükleri elde edilmektedir ve bu filtreler ile frekans spektrumu alt frekans bantlarına bölünerek analiz edilen işaret bu alt bantlar cinsinden gösterilmektedir.

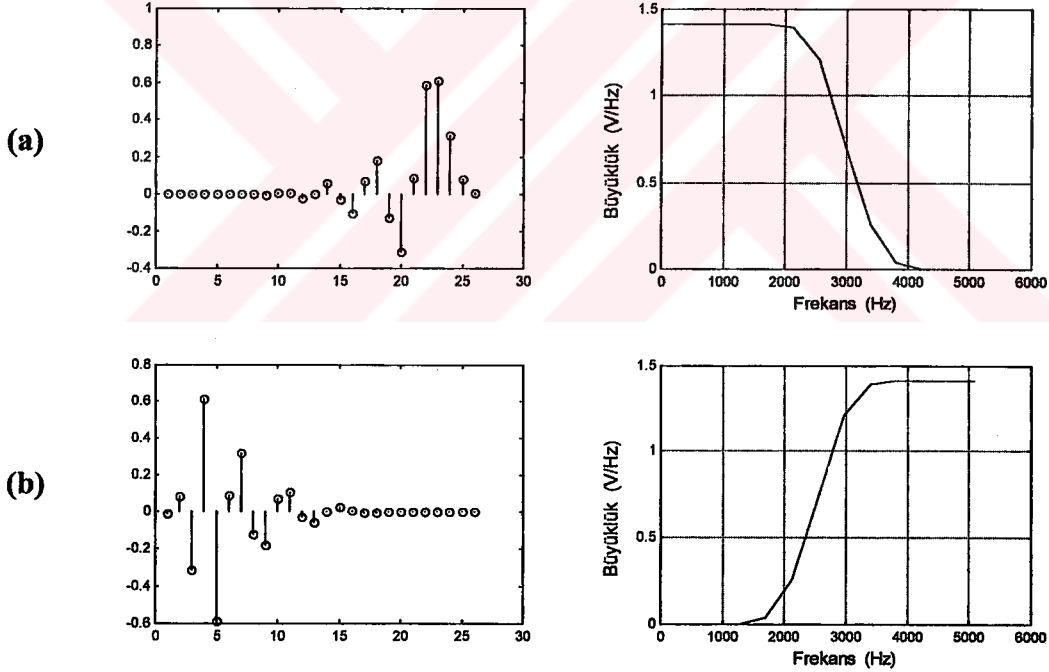
Daha önce Şekil 4.24'te gösterildiği gibi spektrumun en alt bandı ölçek fonksiyonu tarafından kaplanmaktadır. φ_3 3. seviye ölçek fonksiyonu 3. seviyeden sonraki tüm dalgacık fonksiyonlarını kapsamaktadır. Eğer φ_3 ölçek fonksiyonuna aynı ölçek değerindeki ψ_3 dalgacık fonksiyonu ilave edilirse bir üst seviyedeki φ_4 ölçek fonksiyonu elde edilir ($\varphi_3 \approx \varphi_4 + \psi_4$). Görüldüğü gibi herhangi bir ölçek fonksiyonunu bir sonraki ölçek fonksiyonu cinsinden elde edilebilir. Ölçek fonksiyonlarının arasındaki bu ilişki, *yinelemeli* (iteratif) olarak aşağıdaki verilebilir [Daubechies,1998]:

$$\varphi_k(t) = \sum_n h_0(n) \sqrt{2} \varphi_k(2t - n) \quad (4.19)$$

Aynı şekilde dalgacık fonksiyonu da ölçek fonksiyonları cinsinden yinelemeli (iteratif) olarak gösterilebilir:

$$\psi_{j,k}(t) = \sum_n h_1(n) \sqrt{2} \varphi_k(2t - n) \quad (4.20)$$

Denklem (4.16) *çoklu-çözünürlük analiz* (multi-resolution) denklemi ya da *ölçekleme* (dilation) denklemi olarak adlandırılmaktadır. Buradaki $h_0(n)$ katsayıları ölçek filtresidir. İkinci denklem ise *dalgacık denklemi*, $h_1(n)$ katsayıları dalgacık filtresidir. $h_0(n)$ filtresi alçak geçiren, $h_1(n)$ ise yüksek geçiren filtredir. Bu filtreler ayrık dalgacık dönüşümünün temelini oluşturmaktadır. Daha önce değinildiği gibi analiz sırasında işaret seviye seviye bu filtrelerden geçirilerek alt-bantlara ayrılmaktadır ve bu bantlara ait katsayılar elde edilmektedir. Şekil 4.25'te Daubechies dalgacığının 1. seviye analiz filtreleri Fourier dönüşümleriyle birlikte verilmiştir. Görüldüğü gibi filtrelerin kesim frekansları yaklaşık 2.7 kHz'dir (11025 Hz ile örneklenmiş bir işaretin frekans spektrumunun orta noktası).



Şekil 4.25 (a) db13 (Daubechies) dalgacığının alçak geçiren ve (b) yüksek geçiren analiz filtreleri.

Ayrık dalgacık dönüşümü katsayıları Denklem (4.19) ve (4.20)'de tanımlanmış olan filtreler cinsinden aşağıdaki gibi gösterebilir [Daubechies,1998]:

$$c_j(k) = \sum_m h_0(m-2k)c_{j+1}(m) \quad (4.21)$$

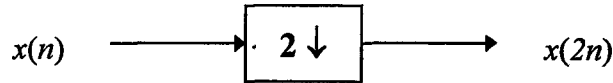
$$d_j(k) = \sum_m h_1(m-2k)c_{j+1}(m) \quad (4.22)$$

Yukarıdaki denklemlerde görüldüğü gibi $c_j(k)$ katsayıları bilindikten sonra diğer seviyelere ait $c_j(k)$ ve $d_j(k)$ katsayıları $h_0(n)$ ve $h_1(n)$ filtreleri kullanılarak hesaplanabilir.

Dalgacık katsayılarını hesaplamak için verilen Denklem (4.21) ve (4.22) sayısal işaret işlemede çok yaygın olarak kullanılan filtreleme denklemine benzemektedir [Denbigh,1998]:

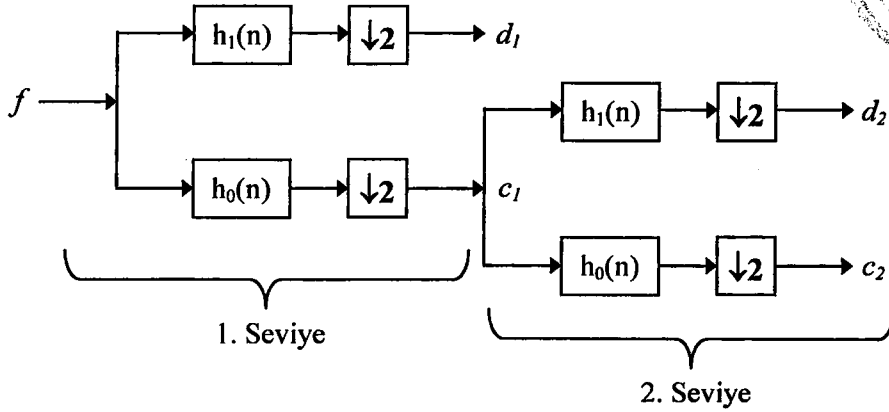
$$y(n) = x(2n) = \sum_{k=0}^{N-1} h(k)x(n-k) \quad (4.23)$$

$x(n)$ giriş dizisi, $h(n)$ filtre katsayıları olmak üzere filtrelenmiş çıkış dizisi $y(n)$ olacaktır. $c_j(k)$ ve $d_j(k)$ katsayılarının denklemleri *aşağı-örnekleme* (down-sampling) denklemidir. Şekil 4.26'da gösterildiği gibi aşağı-örnekleme işleminde $x(n)$ giriş dizisi için $y(n)=x(2n)$ gibi bir çıkış dizisi oluşmakta, eleman sayısı yarıya düşmektedir. Bu işlemin frekans tanım bölgesindeki karşılığı frekans spektrumunun alt yarısını seçip almaktır. Eleman sayısının yarıya düşmesi örnekleme frekansını da yarıya düşürecektir, dolayısıyla, işaretin sahip olduğu en yüksek frekansın yarı değerine kadarki tüm frekanslar kalacaktır.



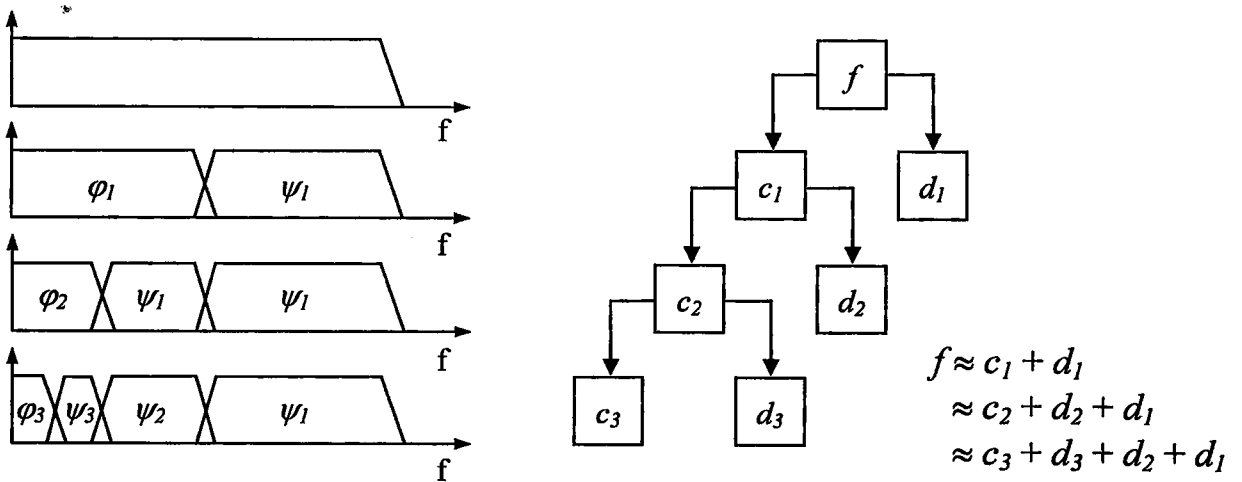
Şekil 4.26 Aşağı örnekleme (Örnekleme frekansının yarıya düşürülmesi).

Ayrık dalgacık dönüşümünde de bir çeşit aşağı-örnekleme yapılmaktadır. İşaret alçak ve yüksek geçiren filtrelerden geçirilerek örneklenmekte ve frekans bantlarına ayrılmaktadır. Bu süreç Şekil 4.27'de gösterilmiştir.



Şekil 4.27 İki seviyeli ayrık dalgacık dönüşümü.

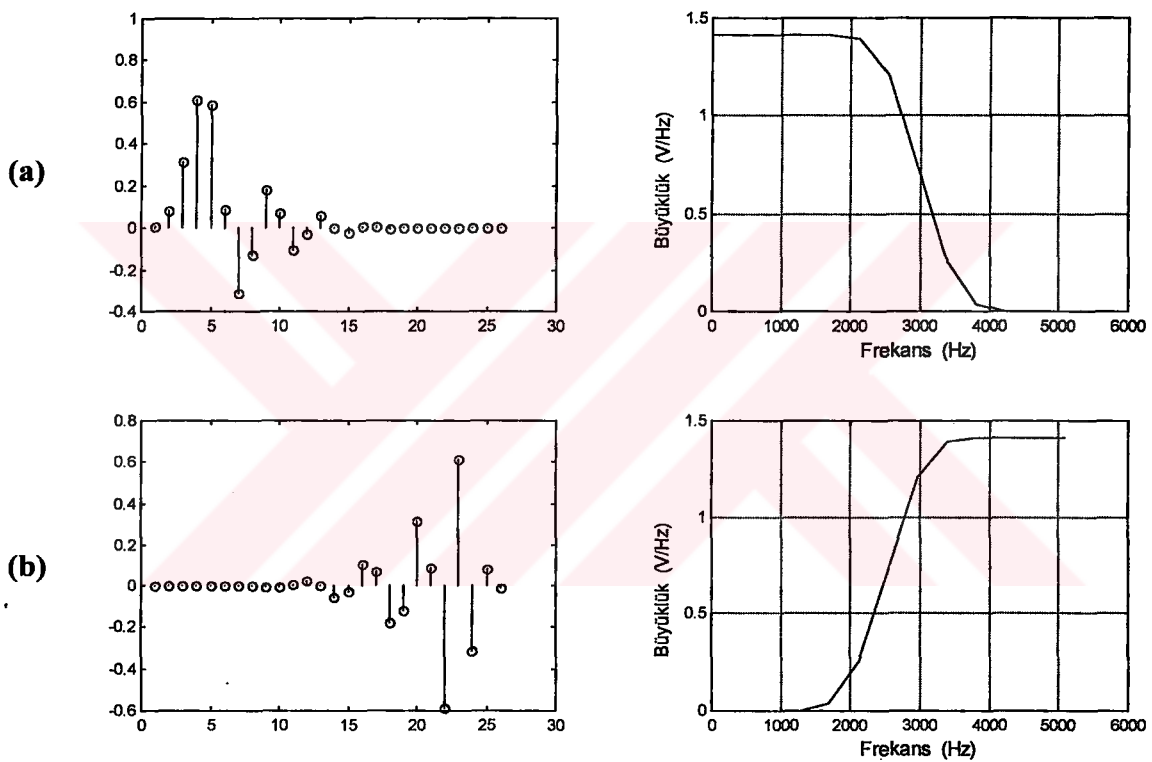
Ayrık dalgacık dönüşümünde analiz edilecek işaret $h_0(n)$ ve $h_1(n)$ filtrelerinden geçirilerek 1. seviye dalgacık katsayıları c_1 ve d_1 elde edilir. Daha sonra c_1 (1. seviye yaklaşıklık katsayıları) tekrar filtrelendir ve 2. seviye katsayıları c_2 ve d_2 elde edilir. Her seviyede aynı şekilde c_k yaklaşıklık katsayıları filtrelenerek bir sonraki seviyeye ait dalgacık katsayıları elde edilir. Ayrık dalgacık dönüşümüne ilişkin dönüşüm ağacı Şekil 4.28’te gösterilmiştir. Dönüşümde seviye ilerledikçe aşağı örnekleme nedeniyle veri sayısı ve kaplanan frekans bandı yarıya düşmektedir.



Şekil 4.28 3 seviyeli ayrık dalgacık dönüşümünü ağacı ve frekans bantları.

4.4.4 Ayırık Dalgacık Sentezi

Ayrık dalgacık dönüşümünün sentez kısmı da aynı şekilde alçak ve yüksek geçiren filtreler ile yapılmaktadır. Sentez filtrelerinin tek farkı, analiz filtrelerinin zaman eksenine göre tersi (time reversal) olmalarıdır. Şekil 4.29'da Daubechies sentez filtreleri verilmiştir. Görüldüğü frekans spektrumları Şekil 4.25'te verilmiş olan analiz filtrelemlerinkiler ile aynıdır.

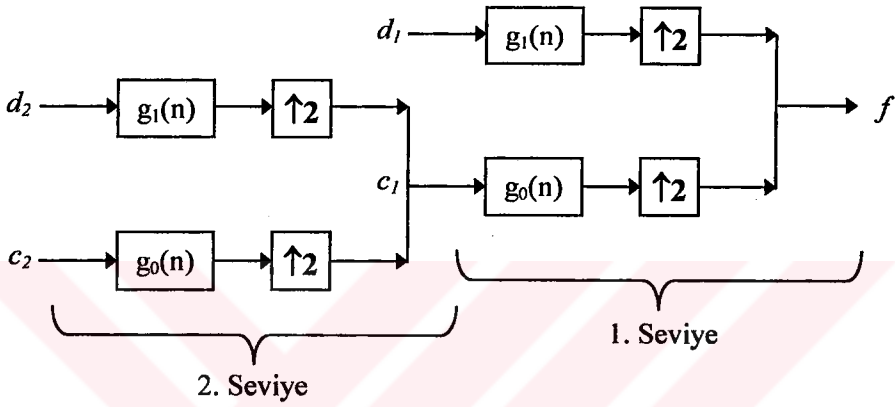


Şekil 4.29 (a) db13 (Daubechies) dalgacığının alçak geçiren ve (b) yüksek geçiren sentez filtreleri.

İşareti ayırık dalgacık dönüşümü katsayılarından sentezlemek için yaklaşıklık katsayıları alçak geçiren filtreden ve ayrıştırma katsayıları yüksek geçiren filtreden geçirilerek bir önceki seviyenin yaklaşıklık katsayıları elde edilir. Aynı işlem sonraki seviyeler için de tekrarlanır ve işarete ulaşılır. Bu sürece ait denklem aşağıda verilmiştir:

$$c_{j+1}(k) = \sum_m g_0(k-2m)c_j(m) + \sum_m g_1(k-2m)d_j(m) \quad (4.24)$$

Sentez filtreleri *yukarı-örnekleme* (up-sampling) yapmaktadır, yani eleman sayısını iki katına çıkarmaktadır. Ters ayrık dönüşüm ağacı Şekil 4.30'de verilmiştir. Analizde olduğu gibi sentezde de süreç yaklaşıklık katsayıları üzerinden yapılmaktadır. 2. seviye katsayıları c_2 ve d_2 filtrelenerek 1. seviye yaklaşıklık katsayıları c_1 elde edilir. Aynı şekilde 1. seviye katsayıları c_1 ve d_1 filtrelenerek işaret elde edilir.

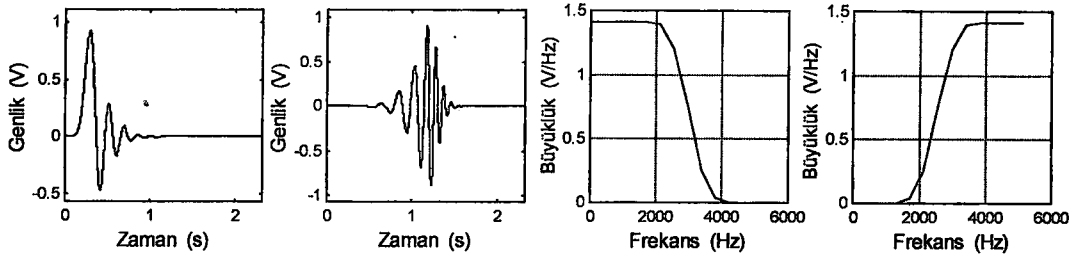


Şekil 4.30 İki seviyeli ters ayrık dönüşüm ağacı.

4.4.5 Ayırık Dalgacık Katsayılarının Hesaplanması

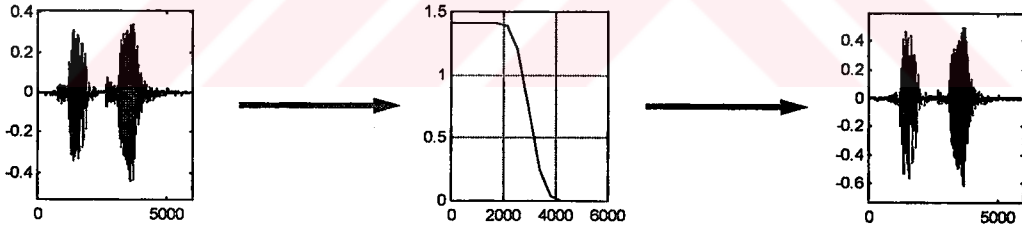
Ayrık dalgacık dönüşümünü oluşturan katsayılar hesaplanırken sırasıyla aşağıdaki adımlar izlenir:

- Bir dalgacık seçilir. Bu dalgacığa ait ölçek ve dalgacık fonksiyonu, analiz filtreleri elde edilir. Bkz. Şekil 4.31.



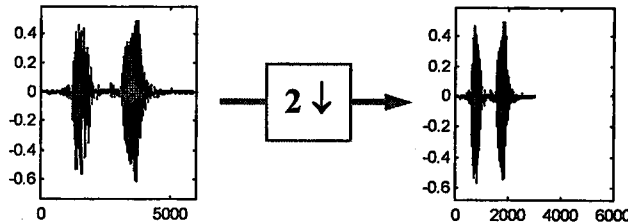
Şekil 4.31 Ölçek fonksiyonu, dalgacık fonksiyonu, alçak-geçiren analiz filtresi ve yüksek-geçiren analiz filtresi.

- İşaret alçak-geçiren filtreden geçirilir ve frekans spektrumunun alt yarısı (0-2756 Hz) alınır. Bkz. Şekil 4.32.



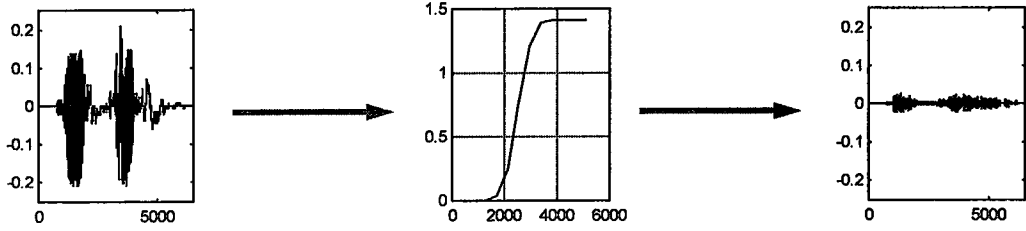
Şekil 4.32 İşaretin alçak-geçiren filtreden geçirilmesi.

- Alçak-geçiren filtreden geçirilen işaret aşağı örneklenir (eleman sayısı yarıya indirilir), ve c_1 1. seviye yaklaşıklık katsayıları elde edilir. Eleman sayısını yarıya düşürmek için işareti oluşturan dizinin çift indisli elemanları atılır. Bkz Şekil 4.33.



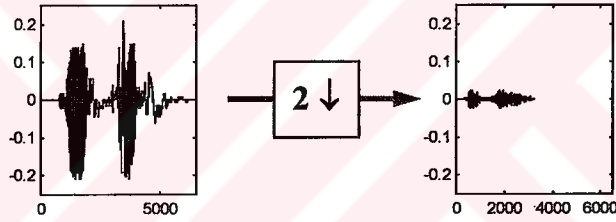
Şekil 4.33 Alçak-geçiren filtreden geçirilen işaretin aşağı-örneklenmesi.

- İşaret yüksek-geçiren filtreden geçirilerek frekans bandının üst yarısı (2756-5512 Hz) alınır. Bkz. Şekil 4.34.



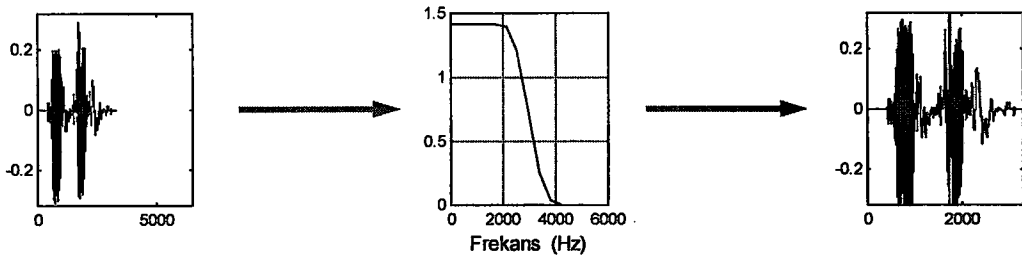
Şekil 4.34 İşaretin yüksek-geçiren filtreden geçirilmesi.

- Yüksek-geçiren filtreden geçirilen işaret aşağı-örneklenir ve d_1 1. seviye ayrıştırma katsayıları elde edilir. Bkz. Şekil 4.35.



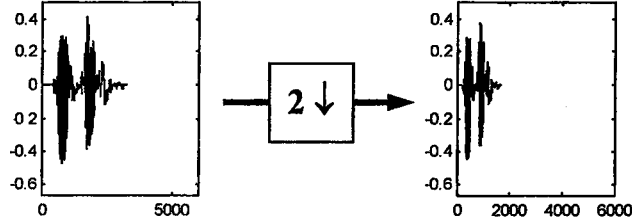
Şekil 4.35 Yüksek-geçiren filtreden geçirilen işaretin aşağı örneklenmesi.

- c_1 1. seviye yaklaşıklık katsayıları alçak-geçiren filtreden geçirilerek frekans bandının tekrar alt yarısı alınır (0-1378 Hz). Bkz. Şekil 4.36.



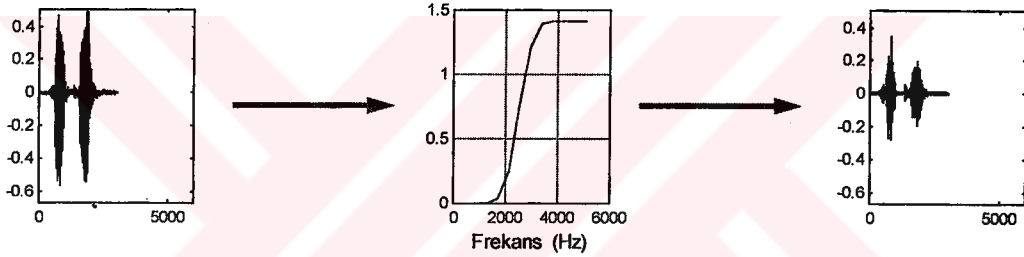
Şekil 4.36 c_1 katsayılarının alçak-geçiren filtreden geçirilmesi.

- Alçak-geçiren filtreden geçirilen c_1 1. seviye yaklaşıklık katsayıları aşağı-örneklenerek c_2 2. seviye yaklaşıklık katsayıları elde edilir. Bkz. Şekil 4.37.



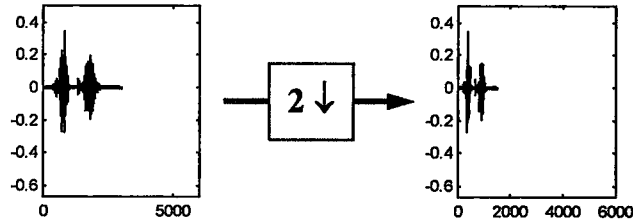
Şekil 4.37 Alçak-geçiren filtreden geçirilen c_1 katsayılarının aşağı-örneklenmesi.

- c_1 1. seviye yaklaşıklık katsayıları yüksek-geçiren filtreden geçirilir ve frekans bandının tekrar üst yarısı alınır (1378-2756 Hz). Bkz. Şekil 4.38.



Şekil 4.38 c_1 katsayılarının yüksek-geçiren filtreden geçirilmesi.

- Yüksek-geçiren filtreden geçirilen c_1 1. seviye yaklaşıklık katsayıları aşağı-örneklenerek d_2 2. seviye ayrıntı katsayıları elde edilir. Bkz. Şekil 4.39.

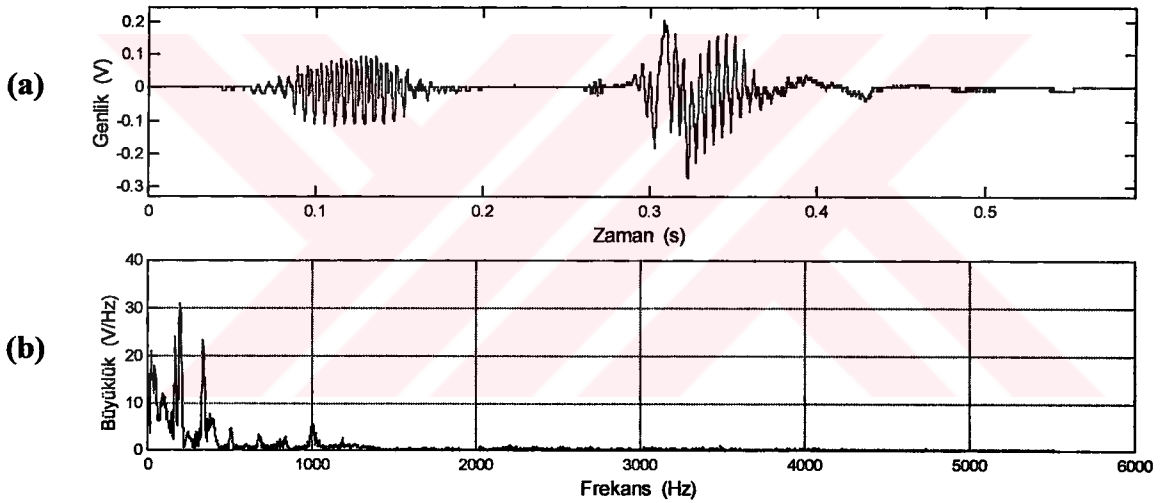


Şekil 4.39 Yüksek-geçiren filtreden geçirilen c_1 katsayılarının aşağı-örneklenmesi.

- Diğer seviyeler içinde sırasıyla tüm adımlar uygulanır ve tüm seviyeler için ayrık dalgacık katsayıları hesaplanır.

4.4.6 Konuşma İşaretlerinin Ayırık Dalgacık Dönüşümü ile İncelenmesi

Bu bölümde "dokuz" kelimesinin ses kayıt örneği ayırık dalgacık dönüşümü ile incelenecektir. Şekil 4.40'da görüldüğü gibi işaret toplam 6500 elemanlı bir diziden oluşmaktadır ve 11025 Hz ile örneklenmiştir. Ayırık dalgacık dönüşümünde her seviyede frekans bandının yarıya düştüğü göz önünde bulundurularak analiz 7. seviyeye kadar ilerletilmiştir. Çizelge 4.2'de 1-7 arası seviyeler için yaklaşıklık ve ayrıştırma katsayılarına ait frekans bantları verilmiştir. En düşük kesim frekansı 43 Hz'dir, bu değer analiz için yeterli olacaktır.

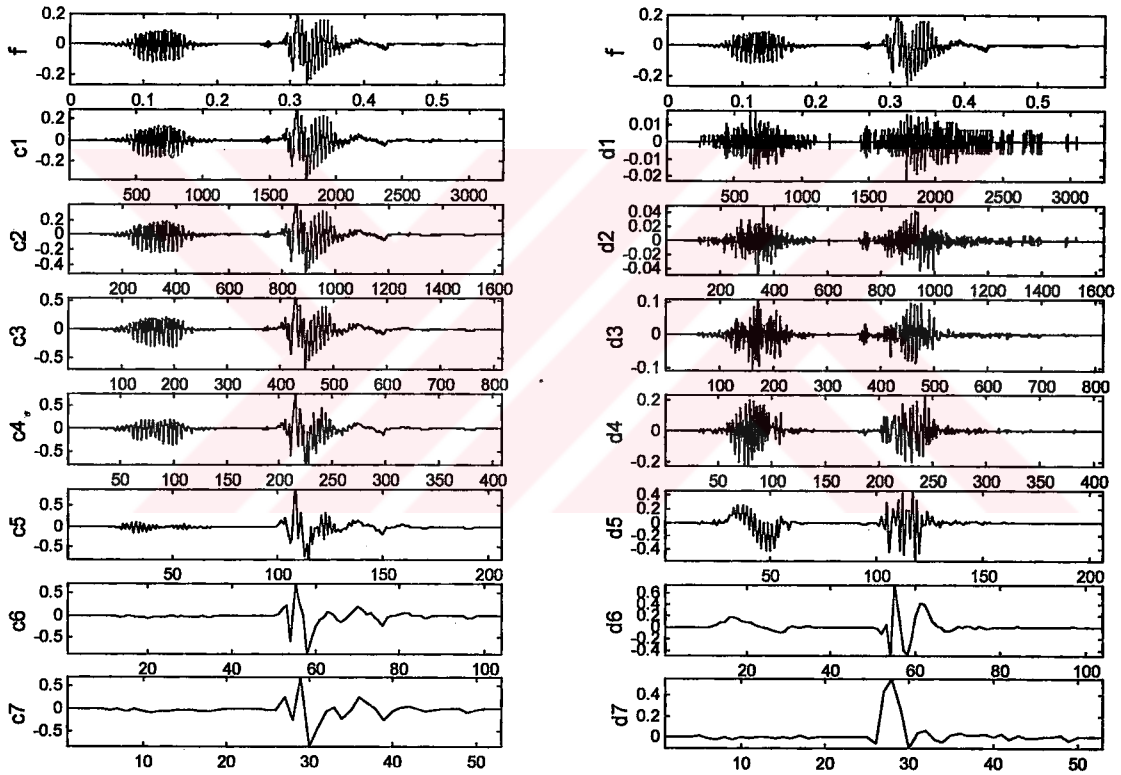


Şekil 4.40 (a) "Dokuz" kelimesinin işlenmemiş hali ve (b) Fourier dönüşümü.

Seviye	d(k)	c(k)
1	2756-5512 Hz	0-2756 Hz
2	1378-2756 Hz	0-1378 Hz
3	689-1378 Hz	0-689 Hz
4	344-689 Hz	0-344 Hz
5	172-344 Hz	0-172 Hz
6	86-172 Hz	0-86 Hz
7	43-86 Hz	0-43 Hz

Çizelge 4.2 11.025 kHz örnekleme frekansı kullanıldığında alt-bantlara ayırma ile elde edilen frekans bantları.

İşaretin 7. seviyeye kadarki ayrık dalgacık dönüşümü katsayıları Şekil 4.41’de verilmiştir. En üstteki grafikte işlenmemiş işaret, sol taraftaki kolonda ayrıştırma katsayıları ve sağ taraftaki kolonda yaklaşıklık katsayıları görülmektedir. Seviye arttıkça eleman sayısı yarıya düşmektedir. 7. seviye katsayıları yaklaşık 50 elemandan oluşmaktadır. Eleman sayısının azalacağı düşünüldüğünde dönüşümü daha da ilerletmek doğru olmayacaktır. Örneğin eleman sayısı 8. seviyede 25’e, 9. seviyede 12’ye düşecektir ve işareti temsil etmekten uzaklaşacaktır.



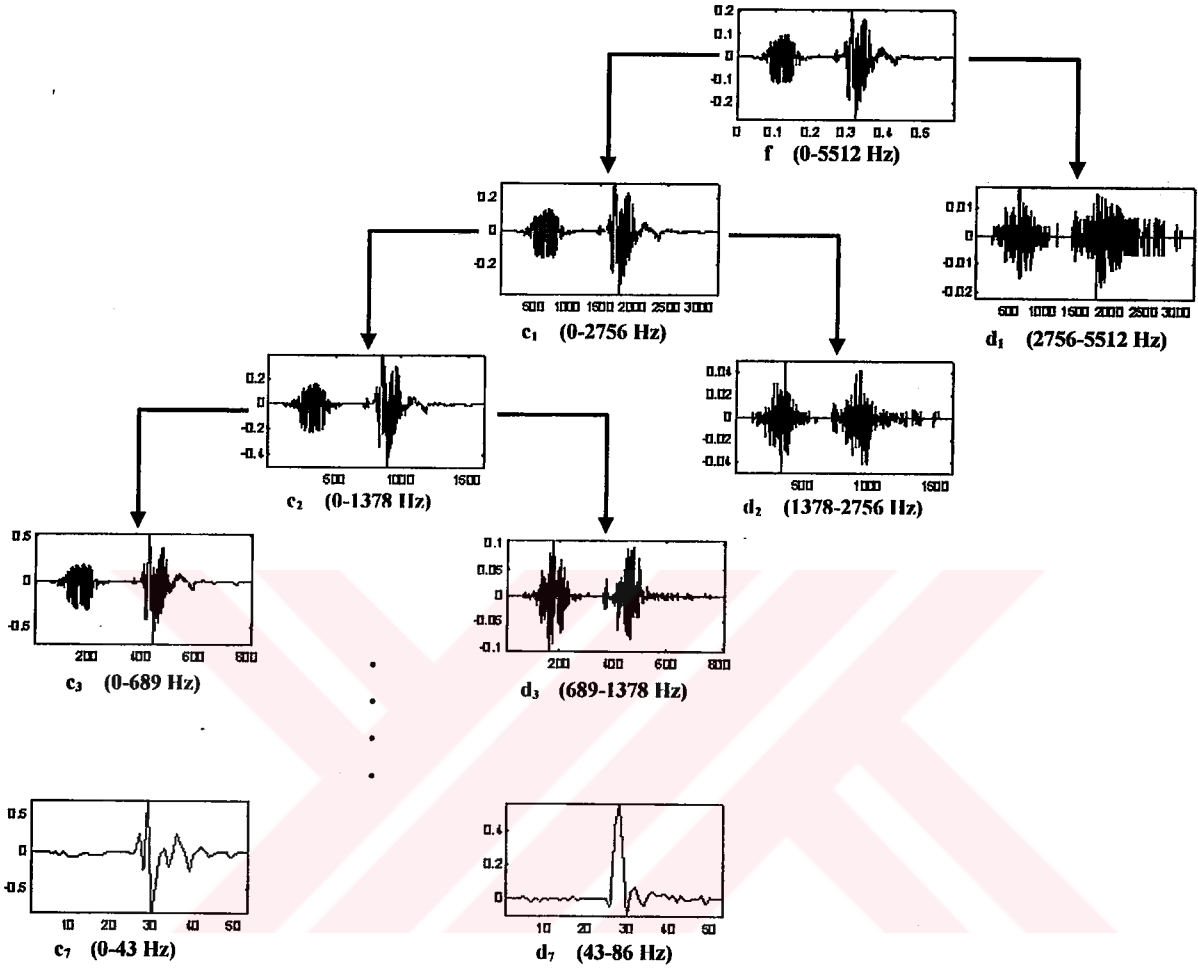
Şekil 4.41 “Dokuz” kelimesinin işlenmemiş hali, ve 7. seviyeye kadar yaklaşıklık ve ayrıştırma katsayıları.

İşaretin ana hatları alçak frekans bileşenlerinde gizlidir, bu nedenle yaklaşıklık katsayılarının grafikleri işarete benzemektedir. İncelenen “dokuz” kelimesinin ses örneğinin ayrık dalgacık dönüşümü katsayıları elde edildikten sonra bilgisayar ortamında ‘wav’ dosyası olarak kayıt edilmiştir. Aşağı-örnekleme nedeniyle örnekleme frekansı yarı değerine düştüğü için ‘wav’ dosyası kayıt edilirken

örnekleme frekansları da değiştirilmiştir. Aksi taktirde katsayılar seslendirilmesi hatalı olacaktır.

4. seviyeye kadarki yaklaşıklık katsayıları dinlendiğine “dokuz” kelimesi anlaşılabilir. 3. ve 4. seviyeler de ses çok boğuk gelmektedir. Ayrıntı katsayıları dinlendiğinde ise beklenildiği gibi kelime anlaşılabilir. Yaklaşıklık katsayıları dinlendiğinde ses kayıt ortamı nedeniyle gelen gürültüler süzülmediğinden ses boğuk olmasına rağmen daha gürültüsüzdür.

Ayrık dalgacık dönüşümü anlatılırken analizin yaklaşıklık katsayıları üzerinden gittiği, her seviyede yaklaşıklık katsayıları filtrelenerek bir sonraki seviyenin yaklaşıklık ve ayrıntı katsayıları bulunduğu gösterilmiştir. İncelenen “dokuz” rakamı için bu süreç Şekil 4.42’de verilmiştir. Ayrık dalgacık dönüşümünde işaret alt-geçiren ve üst-geçiren filtreler ile süzülerek alt bantlara ayrılmaktadır. Filtreleme sonucunda aşağı-örnekleme işlemi nedeniyle eleman sayısı her seviyede yarıya düşmektedir.



Şekil 4.42 “Dokuz” kelimesinin ayrık dalgacık katsayılarının elde edilmesi.

“Dokuz” kelimesinin ses kayıt örneği ayrık dalgacık dönüşümü ile Şekil 4.42’de gösterildiği gibi frekans bantlarına ayrılmıştır, ve 7. seviyeye kadarki yaklaşıklık ve ayrıntı katsayıları hesaplanmıştır. Seviyelere ait frekans bantları Çizelge 4.2 ile verilmişti. Konuşma işaretlerinin analizinde bu katsayılar kullanılırken en fazla bilgi taşıyan ve en az yer kaplayan kombinasyonlar kullanılır. Sadece tek bir seviye katsayıları kullanılabileceği gibi, birkaç seviyeye ait katsayılarda kullanılabilir. Bu çalışmada yapılan ses tanıma deneylerinde ayrık dalgacık katsayıların nasıl kullanıldığı Bölüm 5’te sunulmuştur.

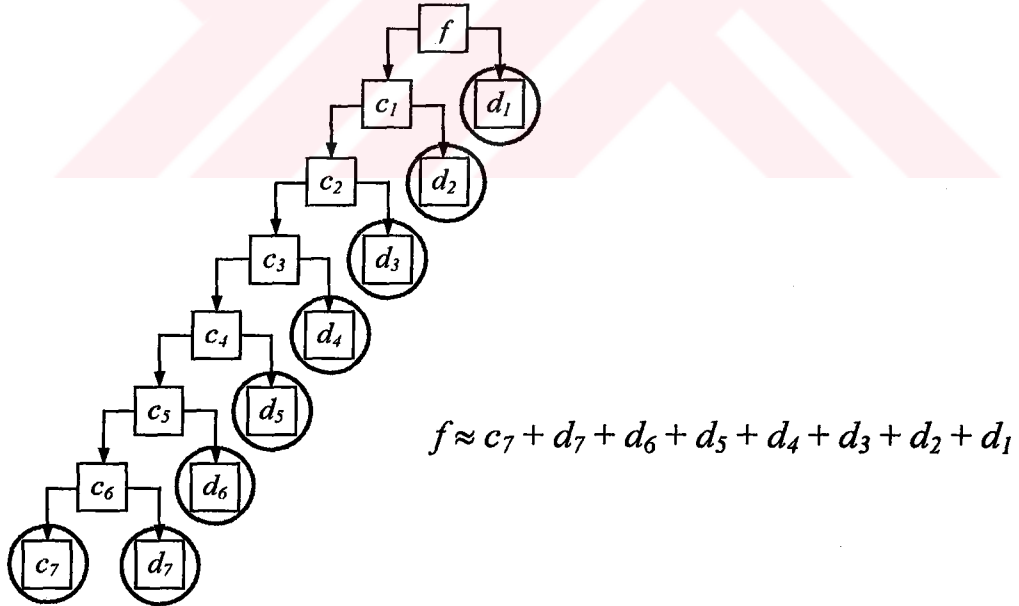
Analiz sonucunda elde edilen katsayılardan Şekil 4.43'teki analiz ağacında gösterildiği gibi sadece belirli bir kısmı seçilirse, işaret tüm frekans bantlarını kaplayacak biçimde aşağıdaki gibi temsil edilebilir:

$$f \approx c_7 + d_7 + d_6 + d_5 + d_4 + d_3 + d_2 + d_1$$

Frekans bantları dikkate alınırse yukarıdaki gösterime göre kullanılan frekans bantları Hz cinsinden aşağıdaki gibi olacaktır:

$$f \approx (0-43) + (43-86) + (86-172) + (172-344) + (344-689) + (689-1378) + (1378-2756) + (2756-5512)$$

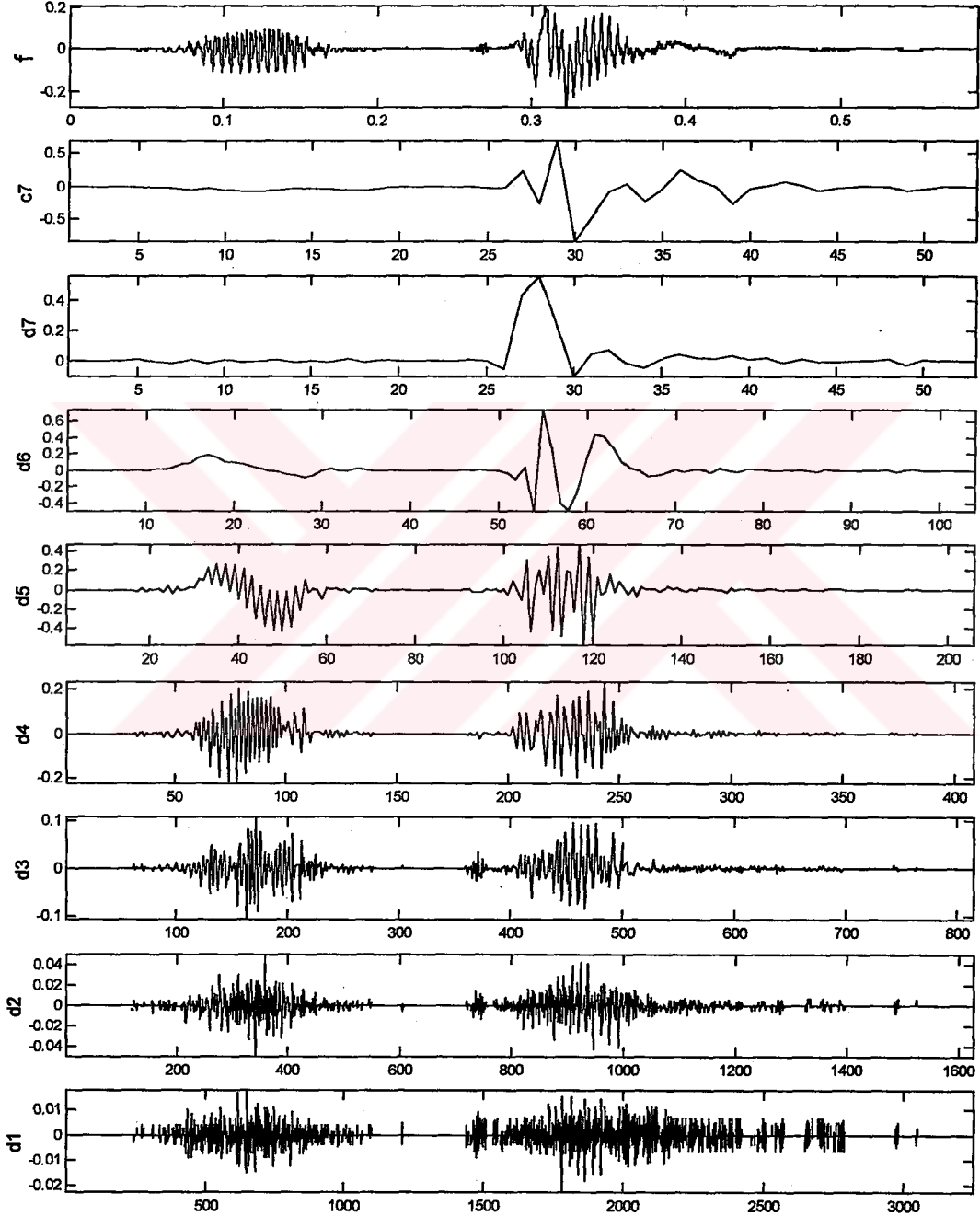
Bu şekilde işaretin tüm frekans bandı kaplanmış olmakta ve işaret bütün olarak temsil edilebilmektedir. Bölüm 4.4.2'de $f(t)$ işaretin ayrık dalgacık seri açılımı Denklem (4.15) ile verilmişti. Bu açılımda yaklaşıklık katsayıları sadece son seviye için, ayrıntı katsayıları ise son seviyeye kadarki tüm seviyeler için hesaplanmaktaydı. Bu nedenle Şekil 4.43'te sunulan gösterim ayrık dalgacık seri açılımı ile de tutarlıdır.



Şekil 4.43 "Dokuz" kelimesine ait ayrık dalgacık analizi ağacı.

Kullanılan katsayıların grafikleri ($c_7, d_7, d_6, d_5, d_4, d_3, d_2, d_1$) Şekil 4.44'te toplu olarak verilmiştir. Görüldüğü gibi c_7 7. seviye yaklaşıklık katsayısı, işaretin eğilimi (kabaca zarfını) vermektedir. Ayrıntı katsayıları ise c_7 eğrisi üzerinde salınım yapıyormuş gibi düşünülebilir. d_1-d_7 ayrıntı katsayılar ise işaretin her seviyedeki

ayrıntılarını verir. Her iki katsayılar dizisi konum bilgisi de taşıyarak işaretin tüm frekans alanını yansıtmış olur.



Şekil 4.44 "Dokuz" kelimesinin işlenmemiş hali ve 7. seviyeye kadar ayırık dalgacık analizi ($f \approx c_7 + d_7 + d_6 + d_5 + d_4 + d_3 + d_2 + d_1$).

5 DENEY SONUÇLARI

5.1 Konuşma İşaretlerinin Elde Edilmesi

Çalışmada kullanılan konuşma işaretleri bilgisayar ortamında Microsoft Ses Kaydedicisi programı kullanılarak kayıt edilmiştir. Kayıt yapılırken konuşmacının ağzı ile mikrofon arasındaki uzaklığın 1-2 cm'yi geçmemesine dikkat edilmiştir. Örnekleme frekansı olarak 11025 Hz seçilmiştir. İşaretler 'wav' dosyası formatında saklanmış ve okutulmuştur. İşaretler kayıt edildikten sonra bilgisayar ortamında başındaki ve sonundaki bilgi içermeyen fazlalıklar atılmış, ortalama 0.6-0.7 s uzunluklara getirilmiştir.

İki grup veri üzerinde çalışılmıştır. Birinci grup veri, tek kişi tarafından seslendirilmiş her rakamdan 10'ar adet toplam, 100 örnekten oluşmaktadır. İkinci grup veri ise değişik konuşmacılardan elde edilmiştir. 5 konuşmacı (2 bayan, 3 erkek) her rakamdan 2'ser adet seslendirmiştir. Birinci grup veri *konusmacı-bağımlı*, ikinci grup veri ise *konusmacı-bağımsız* ses tanıma deneyleri için kullanılmıştır.

Elde edilen bu konuşma işaretlerinin KZFD ve ayırık dalgacık dönüşümü kullanılarak örüntüleri oluşturulmuş, ve Sakoe&Chiba ve Itakura dinamik zaman sıkıştırması yöntemleri ile ses tanıma deneyleri yapılmıştır [Theodoridis and Koutroumbas,1998].

Tüm denemeler Pentium 4, 1.6 GHz işlemcili bir bilgisayar üzerinde MATLAB'te yazılan programlar ile yapılmıştır. Sonuçlara ait süreler bilgisayar konfigürasyonuna göre değişeceği için sadece karşılaştırma amacıyla verilmiştir.

5.2 Örüntü Matrisinin Elde Edilmesi

Örüntü matrisinin elde edilmesi için işaretler ilk önce pencerelenmiştir. Pencere fonksiyonu olarak Hanning pencere fonksiyonu kullanılmıştır. Üç değişik pencere uzunluğu denenmiştir: 45 ms (445 elemanlı), 30 ms (334 elemanlı) ve 20 ms (223 elemanlı). Pencere geçişlerinde bilgi kaybetmemek için pencereler uzunluklarının 1/3'ü oranında kaydırılmıştır. Örneğin 600 ms uzunluğundaki bir

işaret 30 ms'lik pencerelere ayrılırsa toplam pencere sayısı kaydırma nedeniyle 30 olacaktır.

KZFD ile örüntü matrisi oluşturulurken işaret ilk önce pencerelenmiş ve daha sonra her pencerenin Fourier dönüşümü alınmıştır. En uzun pencere 445 elemanlı olduğu için HFD uzunluğu 512 seçilmiştir (sadece 20 ms, yani 223 elemanlı, pencere uzunluğu için 256 uzunluklu HFD de denenmiştir). Her pencere bir özellik vektörüne karşılık gelecektir ve uzunluğu 512 olacaktır. Örüntü matrisinin sütun sayısını ise pencere sayısı belirleyecektir. Elde edilen matrisin boyutu [*HFD uzunluğu* $\times$ *Pencere sayısı*] olacaktır.

Ayrık dalgacık dönüşümü ile örüntü matrisi oluşturulurken işaret pencerelenmiş ve her pencerenin değişik seviyelerde ayrık dalgacık dönüşümü alınmıştır. Özellik vektörleri bazen tek bir seviyeye ait dalgacık katsayılarından bazen de birkaç seviyeye ait dalgacık katsayılarından oluşturulmuştur. Özellik vektörlerini oluşturmak için ise iki yöntem izlenmiştir. Birinci yöntemde sadece tek bir seviyeye ait katsayılar alınmıştır. Bu yöntem için örüntü matrisin boyutu [*Katsayı sayısı* $\times$ *Pencere sayısı*] olacaktır. İkinci yöntemde ise farklı seviyelere ait yaklaşıklık ve ayrıntı katsayıları birlikte alınmıştır. Özellik vektörü ise her seviye katsayılarının enerjilerinden oluşturulmuştur. Enerji hesaplamak için kullanılan formül aşağıda Denklem (5.1)'de verilmiştir:

$$E = \frac{1}{N} \sum_{n=1}^N e_n^2 \quad (5.1)$$

Burada N ayrık dalgacık dönüşüm katsayılarının sayısıdır. Sonuçta elde edilen örüntü matrisinin boyutu ise [*Dönüşüm seviyesi* $\times$ *Pencere sayısı*] olacaktır.

5.3 Konuşmacı-Bağımlı Ses Tanıma Deney Sonuçları

Konuşmacı-bağımlı ses tanıma deneyleri birinci grup veri ile yapılmıştır. İşaretlerin %30'u eğitime için geri kalan %70'i ise test için ayrılmıştır. Eğitime için her rakam grubundan 3'er, toplam 30 örnek; test için 7'şer, toplam 70 örnek kullanılmıştır. Test için ayrılan işaretler eğitime sırasında kullanılmamıştır,

dolayısıyla ses tanıma algoritması bu işaretlerle ilk kez test aşamasında karşılaşmaktadır.

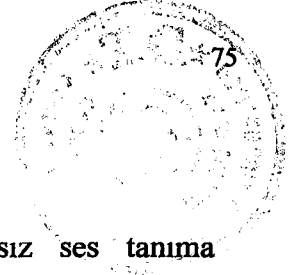
KZFD ile konuşmacı-bağımlı ses tanıma sonuçları Çizelge 5.1’de verilmiştir. Her üç pencere uzunluğu (45 ms, 30 ms, 20 ms) da denenmiştir. Tanıma oranı maksimum %97’dir. Pencere uzunluğunu küçültmek başarıyı arttırmıştır. Bu etki ilerde verilecek ayırık dalgacık sonuçlarında daha açık görülmektedir.

Çizelgede verilen eğitime zamanları 30 örneğin örüntü matrislerinin elde edilmesi için geçen toplam zamandır. Test zamanlarında ise 70 örneğin örüntü matrislerinin elde edilmesi ve referans örüntüler ile karşılaştırılması için geçen süredir. Test zamanlarında esas belirleyici olan ses tanıma algoritmasında harcanan zamandır. Algoritmaya ne kadar fazla veri girerse zaman o oranda artmaktadır. Görüldüğü gibi pencere uzunluğu arttırıldığında test zamanları oldukça artmaktadır. Bunun nedeni algoritmaya verilen matrisin boyutunun artmasıdır (pencere sayısının artması sebebiyle).

Çizelgedeki sonuçlar hem Itakura hem de Sakoe&Chiba dinamik zaman sıkıştırması yöntemi için verilmiştir. Görüldüğü gibi Itakura yöntemi daha iyi sonuç vermiştir. Test zamanları yine Itakura’nın lehinedir, ancak bu üstünlük fazla değildir.

	<i>Tanıma Oranı</i>	<i>Tanıma Yüzdesi</i>	<i>Pencere Uzunluğu</i>	<i>HFD Uzunluğu</i>	<i>Eğitime zamanı</i>	<i>Test Zamanı</i>
ITAKURA	67/70	94.29%	45.0 ms	512	0.43 s	72.17 s
	68/70	97.14%	30.0 ms	512	0.50 s	127.79 s
	68/70	97.14%	20.0 ms	512	0.65 s	296.43 s
	68/70	97.14%	20.0 ms	256	0.43 s	280.45 s
SAKOE CHIBA	59/70	84.29%	45.0 ms	512	0.44 s	75.28 s
	61/70	87.14%	30.0 ms	512	0.48 s	121.23 s
	65/70	92.86%	20.0 ms	512	0.63 s	318.43 s
	67/70	95.71%	20.0 ms	256	0.44 s	289.17 s

Çizelge 5.1 Birinci grup veri üzerinde KZFD ile konuşmacı-bağımlı ses tanıma sonuçları.



Ayrık dalgacık dönüşümü ile yapılan konuşmacı-bağımsız ses tanıma deneylerinde kullanılan katsayıların seçimine göre iki yöntem izlenmiştir. İlk yöntemde tek bir seviye katsayıları alınmıştır. İkincisinde ise birkaç seviyeye ait katsayılar birlikte kullanılmıştır.

Birinci yöntemle ait sonuçlar Çizelge 5.2, 5.3 ve 5.4'te verilmiştir. Bu çizelgelerdeki tüm sonuçlar yaklaşıklık katsayıları ile elde edilmiştir. Ayrıntı katsayıları ile yapılan birkaç denemede sifıra yakın tanıma sonuçları alındığı için tek seviye denemelerinde bu katsayılarla daha fazla deney yapılmamıştır. Bunun nedeni işaretlerle ilgili esas bilginin alçak frekanslarda yer almasıdır. Dönüşüm seviyesi arttıkça tanıma oranı artmıştır. Çünkü frekans çözünürlüğü artmaktadır. 1. seviye yaklaşıklık katsayıları 0-2756 Hz, 2. seviye ve 3. seviye katsayıları ise 0-1378 Hz ve 0-689 Hz frekans bantlarına sahiptir. Bu dönüşüm seviyelerinin kullanılmalarının sebebi konuşma işaretlerinin önemli frekans bileşenlerinin bu bantlarda olmasıdır. 0-344 Hz bandına sahip 4. seviye katsayıları ile yapılan birkaç denemede ise tanıma oranı çok düşük kalmıştır. Bunun nedeni konuşma işaretlerinin bu alçak frekans bandında fazla bilgi taşımasıdır.

Ayrık dalgacık dönüşümleri db2, db6 ve db12 (Daubechies) dalgacıkları için alınmıştır. Dalgacık derecesinin değişmesi başarı oranını pek fazla etkilememiştir. En iyi başarı oranını %78 ile db6 dalgacığı verdiği için bundan sonraki denemelerde bu dalgacık kullanılmıştır.

KZFD ile yapılan deneylerde olduğu gibi burada da Itakura yöntemi daha iyi sonuçlar vermiştir. Çizelge 5.2, 5.3 ve 5.4'te görüldüğü gibi pencere uzunluğunun düşürülmesi ile başarı oranı artmaktadır.

Ayrık dalgacık dönüşümünü Fourier dönüşümünden daha çok hesap gerektirdiği için gerekli eğitime zamanı daha fazladır. Ancak bu dönüşümle ses tanıma algoritmasına daha az veri girildiği için test zamanları daha düşüktür. Görüldüğü gibi seviye arttıkça eğitime zamanları uzamaktadır. Bütün örüntü tanıma yöntemleri için eğitime bir kez yapıldığından tanıma sürelerinin kısa olması daha önemlidir. Çünkü seviye yükseldikçe ayrık dalgacık dönüşümü hesabındaki işlem sayısı artmaktadır.

İkinci yöntemde, ayrık dalgacık dönüşümünde frekans spektrumunun alt-bantlara bölünmesinden yararlanmak düşüncesiyle birkaç seviyeye ait katsayılar birlikte kullanılmıştır. Birinci yöntemde en iyi tanıma sonuçları db6 dalgacığı ve 20 ms pencere uzunluğu ile elde edildiğinden, ikinci yöntemde sadece bu değerler kullanılmıştır.

Çizelge 5.5'te görüldüğü gibi bu yöntemle tanıma oranı %80'lere çıkmıştır. Çünkü bu yöntemde birden fazla seviyeye ait katsayılar birlikte kullanılarak işaretin frekans spektrumu farklı alt-bantlarla temsil edilmiş ve böylece pencere içindeki frekans çözünürlüğü daha iyileştirilmiştir. Örneğin 5. seviye yaklaşıklık ve 1-5. seviye ayrıntı katsayılarının kullanıldığı (c_5, d_{1-5}) yönteminde işaret. $f \approx c_5 + d_5 + d_4 + d_3 + d_2 + d_1$ şeklinde 6 katsayı dizisi ile temsil edilmektedir. Bu 6 katsayı dizisinin Denklem (5.1) ile verilen bant enerjileri ayrı ayrı hesaplanmıştır.

	Tanıma Oranı	Tanıma Yüzdesi	Pencere Uzunluğu	Dönüşüm Seviyesi	Dalgacık	Eğitme zamanı	Test Zamanı
ITAKURA	27/70	38,57%	45,0 ms	1	db2	4,86 s	73,81 s
	29/70	41,43%	30,0 ms	1	db2	6,25 s	112,11 s
	45/70	64,29%	20,0 ms	1	db2	8,13 s	252,28 s
	31/70	44,29%	45,0 ms	1	db6	5,28 s	64,31 s
	27/70	38,57%	30,0 ms	1	db6	6,09 s	110,65 s
	45/70	64,29%	20,0 ms	1	db6	8,81 s	241,85 s
	26/70	37,14%	45,0 ms	1	db12	15,25 s	62,85 s
	25/70	35,71%	30,0 ms	1	db12	20,53 s	105,41 s
	45/70	64,29%	20,0 ms	1	db12	29,84 s	241,80 s
SAKOE & CHIBA	27/70	38,57%	45,0 ms	1	db2	4,61 s	67,95 s
	13/70	18,57%	30,0 ms	1	db2	5,81 s	114,70 s
	25/70	35,71%	20,0 ms	1	db2	8,39 s	246,40 s
	15/70	21,43%	45,0 ms	1	db6	4,83 s	67,29 s
	12/70	17,14%	30,0 ms	1	db6	6,63 s	118,48 s
	25/70	35,71%	20,0 ms	1	db6	9,11 s	278,92 s
	14/70	20,00%	45,0 ms	1	db12	15,23 s	68,13 s
	12/70	17,14%	30,0 ms	1	db12	19,89 s	122,07 s
	25/70	35,71%	20,0 ms	1	db12	30,58 s	278,30 s

Çizelge 5.2 Birinci grup veri üzerinde 1. seviye ayrık dalgacık dönüşümü ile konuşmacı-bağımlı ses tanıma sonuçları.

	Tanım Oranı	Tanım Yüzdesi	Pencere Uzunluğu	Dönüşüm Seviyesi	Dalgacık	Eğitme zamanı	Test Zamanı
ITAKURA	32/70	45,71%	45,0 ms	2	db2	4,92 s	57,79 s
	35/70	50,00%	30,0 ms	2	db2	6,20 s	96,58 s
	49/70	70,00%	20,0 ms	2	db2	9,13 s	228,59 s
	31/70	44,29%	45,0 ms	2	db6	5,25 s	58,64 s
	34/70	48,57%	30,0 ms	2	db6	6,73 s	100,76 s
	46/70	65,71%	20,0 ms	2	db6	10,56 s	233,11 s
	31/70	44,29%	45,0 ms	2	db12	15,77 s	59,71 s
	33/70	47,14%	30,0 ms	2	db12	20,48 s	109,42 s
	47/70	67,14%	20,0 ms	2	db12	30,50 s	241,12 s
SAKOE & CHIBA	16/70	22,86%	45,0 ms	2	db2	4,89 s	64,99 s
	14/70	20,00%	30,0 ms	2	db2	6,39 s	106,42 s
	24/70	34,29%	20,0 ms	2	db2	9,25 s	259,72 s
	18/70	25,71%	45,0 ms	2	db6	5,44 s	61,74 s
	13/70	18,57%	30,0 ms	2	db6	6,78 s	115,54 s
	26/70	37,14%	20,0 ms	2	db6	10,42 s	259,78 s
	19/70	27,14%	45,0 ms	2	db12	15,66 s	63,35 s
	14/70	20,00%	30,0 ms	2	db12	20,48 s	11,73 s
	27/70	38,57%	20,0 ms	2	db12	31,59 s	254,96 s

Çizelge 5.3 Birinci grup veri üzerinde 2. seviye ayrık dalgacık dönüşümü ile konuşmacı-bağımlı ses tanıma sonuçları.

	Tanım Oranı	Tanım Yüzdesi	Pencere Uzunluğu	Dönüşüm Seviyesi	Dalgacık	Eğitme zamanı	Test Zamanı
ITAKURA	31/70	44,29%	45,0 ms	3	db2	6,28 s	56,25 s
	33/70	47,14%	30,0 ms	3	db2	6,65 s	105,00 s
	48/70	68,57%	20,0 ms	3	db2	9,89 s	224,22 s
	33/70	47,14%	45,0 ms	3	db6	5,64 s	56,01 s
	32/70	45,71%	30,0 ms	3	db6	7,23 s	106,38 s
	55/70	78,57%	20,0 ms	3	db6	10,78 s	241,54 s
	31/70	44,29%	45,0 ms	3	db12	18,28 s	70,86 s
	34/70	48,57%	30,0 ms	3	db12	21,48 s	113,86 s
	51/70	72,86%	20,0 ms	3	db12	34,30 s	285,86 s
SAKOE & CHIBA	17/70	24,29%	45,0 ms	3	db2	5,62 s	59,46 s
	15/70	21,43%	30,0 ms	3	db2	6,86 s	106,16 s
	26/70	37,14%	20,0 ms	3	db2	10,06 s	244,48 s
	18/70	25,71%	45,0 ms	3	db6	5,61 s	71,29 s
	15/70	21,43%	30,0 ms	3	db6	7,52 s	106,23 s
	25/70	35,71%	20,0 ms	3	db6	10,74 s	237,26 s
	16/70	22,86%	45,0 ms	3	db12	16,25 s	60,56 s
	13/70	18,57%	30,0 ms	3	db12	20,94 s	102,80 s
	24/70	34,29%	20,0 ms	3	db12	32,80 s	250,82 s

Çizelge 5.4 Birinci grup veri üzerinde 3. seviye ayrık dalgacık dönüşümü ile konuşmacı-bağımlı ses tanıma sonuçları.

	<i>Tanırna Oranı</i>	<i>Tanırna Yüzdesi</i>	<i>Pencere Uzunluđu</i>	<i>Dönüşüm Seviyesi</i>	<i>Dalgacık</i>	<i>Eğitme zamanı</i>	<i>Test Zamanı</i>
ITAKURA	57/70	81.43%	20.0 ms	c5, d1-5	db6	38.83 s	216.58 s
	56/70	80.00%	20.0 ms	c6, d1-6	db6	51.94 s	220.37 s
	57/70	81.43%	20.0 ms	c7, d1-7	db6	63.31 s	212.55 s
	56/70	80.00%	20.0 ms	c8, d1-8	db6	72.60 s	209.82 s
	55/70	78.57%	20.0 ms	d1-5	db6	9.59 s	201.12 s
	56/70	80.00%	20.0 ms	d1-6	db6	10.47 s	209.54 s
	58/70	82.86%	20.0 ms	d1-7	db6	11.83 s	209.65 s
	56/70	80.00%	20.0 ms	d1-8	db6	12.77 s	223.45 s
SAKOE & CHIBA	41/70	58.57%	20.0 ms	c5, d1-5	db6	39.48 s	228.58 s
	37/70	52.86%	20.0 ms	c6, d1-6	db6	54.14 s	277.27 s
	36/70	51.43%	20.0 ms	c7, d1-7	db6	71.05 s	233.59 s
	36/70	51.43%	20.0 ms	c8, d1-8	db6	72.33 s	241.70 s
	30/70	42.86%	20.0 ms	d1-5	db6	9.69 s	227.51 s
	34/70	48.57%	20.0 ms	d1-6	db6	10.56 s	230.79 s
	41/70	58.57%	20.0 ms	d1-7	db6	11.42 s	230.51 s
	37/70	52.86%	20.0 ms	d1-8	db6	12.80 s	231.34 s

Çizelge 5.5 Birinci grup veri üzerinde değışik seviyeler için ayrıık dalgacık dönüşümü ile konuşmacı-bağımlı ses tanıma sonuçları.

5.4 Konuşmacı-Bağımsız Ses Tanıma Deney Sonuçları

Konuşmacı-bağımsız ses tanıma deneyleri ikinci grup veri üzerinde yapılmıştır. Eğitime için ise hem birinci grup (tek bir kişi tarafından seslendirilmiş), hem de ikinci grup veri (5 kişi tarafından seslendirilmiş) kullanılmıştır. Konuşmacı-bağımlı ses tanıma deneylerinde en iyi tanıma sonuçları db6 dalgacı, 20 ms pencere uzunluğu, d_{1-7} ve d_{1-8} ($f \approx d_8 + d_7 + d_6 + d_5 + d_4 + d_3 + d_2 + d_1$) yöntemi ile elde edildiğinden, konuşmacı-bağımsız ses tanıma denemelerinin hepsinde bu değerler kullanılmıştır.

İlk denemede eğitime için birinci grup veriler, tanıma için ikinci grup veriler kullanılmıştır. Çizelge 5.6'da KZFD ile, Çizelge 5.7'de ise ayrık dalgacık dönüşümü ile alınan sonuçlar verilmiştir. Görüldüğü gibi ayrık dalgacık dönüşümü daha iyi sonuç vermiştir. Bunun yanı sıra ayrık dalgacık dönüşümü test zamanları daha düşüktür.

	Tanım Oranı	Tanım Yüzdesi	Pencere Uzunluğu	HFD Uzunluğu	Eğitime zamanı	Test Zamanı
ITAKURA	20/100	20.00%	20.0 ms	512	0.44 s	516.60 s
SAKOE & CHIBA	15/100	15.00%	20.0 ms	512	0.42 s	563.68 s

Çizelge 5.6 İkinci grup veri üzerinde KZFD ile konuşmacı-bağımsız ses tanıma sonuçları (eğitime birinci grup veri ile).

	Tanım Oranı	Tanım Yüzdesi	Pencere Uzunluğu	Dönüşüm Seviyesi	Dalgacık	Eğitime zamanı	Test Zamanı
ITAKURA	30/100	30.00%	20.0 ms	d1-7	db6	11.58 s	424.39 s
	28/100	28.00%	20.0 ms	d1-8	db6	12.78 s	421.50 s
SAKOE CHIBA	17/100	17.00%	20.0 ms	d1-7	db6	11.62 s	488.04 s
	12/100	12.00%	20.0 ms	d1-8	db6	12.78 s	483.97 s

Çizelge 5.7 İkinci grup veri üzerinde ayrık dalgacık dönüşümü ile konuşmacı-bağımsız ses tanıma sonuçları (eğitime birinci grup veri ile).

İkinci denemede ise eğitime aşamasında ikinci grup veriler kullanılmıştır. Ancak ses tanıma algoritmasının eğitime sırasında karşılaştığı işaretle testte bir daha karşılaşmamasına dikkat edilmiştir. Örneğin bir kişinin ses kayıtları test edilirken eğitime için geri kalan 4 kişinin ses kayıtları kullanılmıştır. Bu denemede de ayrık dalgacık dönüşümü daha iyi sonuç vermiştir. Bunun yanı sıra ilk denemeye göre tanıma oranı da artmıştır. Bunun nedeni bu seslerin aynı ortamda kayıt edilmesi ve/veya eğitime için daha fazla işaret kullanılması olabilir.

Test zamanlarının birinci grup veri ile yapılan denemelere göre daha uzun olmasının nedeni eğitime için kullanılan işaret sayısının fazla olmasıdır. İlk denemede eğitime için 30 işaret kullanılırken iken burada 80 işaret kullanılmaktadır.

Önceki deneylerde olduğu gibi burada da Itakura yöntemi daha başarılı olmuştur.

	Tanıma Oranı	Tanıma Yüzdesi	Pencere Uzunluğu	HFD Uzunluğu	Eğitime zamanı	Test Zamanı
ITAKURA	34/100	34,00%	20,0 ms	512	1,44 s	1513,12 s
SAKOE & CHIBA	11/100	11,00%	20,0 ms	512	1,63 s	2174,89 s

Çizelge 5.8 İkinci grup veri üzerinde KZFD ile konuşmacı-bağımsız ses tanıma sonuçları (eğitime ikinci grup veri ile).

	Tanıma Oranı	Tanıma Yüzdesi	Pencere Uzunluğu	Dönüşüm Seviyesi	Dalgacık	Eğitime zamanı	Test Zamanı
SAKOE	42/100	43,00%	20,0 ms	d1-7	db6	46,50 s	1589,59 s
CHIBA	30/100	30,00%	20,0 ms	d1-8	db6	50,25 s	1656,27 s
ITAKURA	16/100	16,00%	20,0 ms	d1-7	db6	52,09 s	1814,85 s
	12/100	12,00%	20,0 ms	d1-8	db6	55,13 s	1900,14 s

Çizelge 5.9 İkinci grup veri üzerinde ayrık dalgacık dönüşümü ile konuşmacı-bağımsız ses tanıma sonuçları (eğitime ikinci grup veri ile).

6 SONUÇLAR

Bu tez çalışmasında Fourier, Kısa Zamanlı Fourier ve Dalgacık dönüşümleri incelenmiştir. Zaman-frekans analizi yapan KZFD ve zaman-ölçek analizi yapan dalgacık dönüşümleri kullanılarak konuşma işaretlerinin ayırt edici özelliklerini çıkartmaya yönelik yöntemler geliştirilmiş ve bu yöntemlerle elde edilen örüntüler ses tanıma deneylerinde kullanılmıştır.

Bu çalışmalar sonucu alınan genel sonuçlar aşağıda sıralanmıştır:

- Çoklu-çözünürlük analizi yapan dalgacık dönüşümü, değişken yapıdaki ses işaretlerini incelemek için KZFD'ne göre daha uygun bir yöntemdir.
- Dalgacık dönüşümü yöntemi ile işaretlerin yerel değişimlerini yakalamak mümkündür.
- Konuşmacı-bağımlı ses tanıma deneylerinde KZFD temelli özellik vektörleri ile oluşturulmuş örüntüler daha başarılı olmuştur. Konuşmacı-bağımsız ses tanıma deneylerinde ise dalgacık dönüşümü daha iyi performans sergilemiştir.
- Alt-bantlara ayırma yöntemini kullanan ayrık dalgacık dönüşümü ile işaretlerin bütün frekans bantlarından ayrı ayrı bilgi edinmek mümkündür. Bu şekilde oluşturulan özellik vektörleri daha küçük boyutludur, bu da ses tanıma süresini kısaltmaktadır.
- Ayrık dalgacık dönüşümü özellik vektörü oluşturmada daha esnektir, çeşitli özellik vektörü oluşturma yöntemleri geliştirilmiştir.
- Kullanılan örüntü tanıma yöntemlerinden Itakura yöntemi, Sakoe&Chiba yöntemine göre her denemede daha başarılı ses tanıma sonuçları vermiştir.



Bu sonuçlar ışığında gelecek çalışmalar için aşağıdaki öneriler sunulmuştur:

- Gelecekteki ses tanıma çalışmalarında gizli Markov modelleri ve yapay sinir ağları gibi yöntemler kullanılabilir. Bu yöntemlerle konuşma işaretleri daha iyi modellenebilmektedir. Dinamik zaman sıkıştırması yöntemi, eğitme sonucu elde edilmiş referans örüntüye her tanıma işleminde ihtiyaç duymaktadır. Bu nedenle tanıma süresi uzamaktadır. Gizli Markov modelleri ve yapay sinir ağları ise her seferinde referans örüntüye ihtiyaç duymadıklarından tanıma süreleri oldukça kısadır.
- Eğitim örnek sayısı artırılarak ve ses işaretleri kayıt ortamından kaynaklanan gürültülerden arındırılarak özellikle konuşmacı-bağımsız ses tanımada daha iyi sonuçlar alınabilir.
- Dalgacık dönüşümü başka işaretlerin analizlerinde de kullanılabilir. Burada geliştirilen özellik çıkartma yöntemleri EKG, EEG gibi biyolojik işaretlerin, titreşim sinyallerinin ve borsa değişim eğilimlerinin analizlerinde kullanılabilir.

KAYNAKLAR

- [1] B. Truax, Handbook for Acoustic Ecology, World Soundscape Project, Simon Fraser University, and ARC Publications, 1999.
- [2] F. Fallside and William A. Woods, Computer Speech Processing, Prentice Hall, 1985
- [3] L. Rabiner and R. W. Schafer, Digital Processing of Speech Signals, Prentice Hall, 1978
- [4] P. Denbigh, System Analysis and Signal Processing; With Emphasis on the Use Of MATLAB, Addison-Wesley Publishing Group, 1998.
- [5] B. Gold and N. Morgan, Speech and Audio Signal Processing, Processing and Perception of Speech and Music, John Wiley & Sons, INC., 1999.
- [6] S. Quin and D. Chen, Joint Time-Frequency Analysis, Prentice Hall, 1996.
- [7] M. Misiti, Y. Misiti, G. Oppenheim and J. M. Poggi, Wavelet Toolbox User's Guide, MathWorks Inc., 2000.
- [8] C. J. Long and S. Datta, "Wavelet Based Feature Extraction for Phoneme Recognition", Proceedings of ICSLP'96, 1: 264-267, 1996.
- [9] C. Delfs and F. Jondral, "Classification of Transient Time-Varying Signals Using DFT and Wavelet Packet Based Methods", IEEE Conference on Acoustics, *Speech and Signal Processing*, 3:1569, 1998.
- [10] S. Cassidy. *Lecture Notes on Speech Recognition*. Speech Hearing and Language Research Centre, Macquarie University, 2001
- [11] L. Janer, J. Martí, C. Nadeu and E. Lleida-Solano, "Wavelet Transforms For Non-uniform Speech Recognition Systems", *Proceedings of ICSLP'96* Volume 4, pp. 2348-2351, 1996.
- [12] R. Sarikaya, B. L. Pellom, and J. H. Hansen, "Wavelet Packet Transform Features with Application to Speaker Identification". *NORSIG'98*, pp. 81-84, 1998.
- [13] Y. Hau and X. Zhu, "A New Feature in Speech Recognition Based on Wavelet Transform", *IEEE*, 2000 (0-7803-5747-7).
- [14] L. Rabiner and B. H. Juang, *Fundamentals of Speech Recognition*, Prentice Hall, 1993.
- [15] K. Kim, D. H. Youn and C. Lee, "Evaluation of Wavelet Filters for Speech Recognition", *IEEE*, 2000 (0-7803-6583-6).
- [16] Y. Kaisheng, C. Zhigang, "A Robust Speech Feature – Perceptive Scalogram Based on Wavelet Analysis". *Proceedings of ICSP'98*, pp. 662-665, 1998.
- [17] Z. Tufekci, J. N. Gowdy "Feature Extraction Using Discrete Wavelet Transform for Speech Recognition" *Proceedings of Southeastcon 2000*, pp. 116-123, 2000

- [18] H. Sakoe and S. Chiba, "Dynamic Programming Algorithm Optimization for Spoken Word Recognition", *IEEE Transactions on Acoustic Speech, and Signal Processing*, ASSP-26, pp. 43-49, 1978.
- [19] F. Itakura, "Minimum Prediction Residual Principle Applied to Speech Recognition", *IEEE Transactions on Acoustic Speech, and Signal Processing*, ASSP-23, pp. 57-72, 1975
- [20] J. C. Goswami and A. K. Chan, *Fundamental of Wavelets, Theory, Algorithms, and Applications*, John Wiley & Sons, INC 1999.
- [21] A. Brian and M. BREINER, *MATLAB for Engineers*, Addison-Wesley Publishing Group, 1995.
- [22] I. Daubechies, "Where do wavelets come from? A personal point of view", *The Proceedings of the IEEE, Special Issue on Wavelets* Volume 84, No 4, pp. 510-513, 1996.
- [23] G. Strang and T. Gilbert, *Wavelets and Filter Banks*, Wellesley-Cambridge Press, 1996.
- [24] I. Daubechies, "Recent results in wavelet applications", *Journal of Electronic Imaging*, Volume 7, No 4, pp. 719-724, 1998.
- [25] C. Torrence and G. P. Compo, "A Practical Guide to Wavelet Analysis", *Bulletin of the American Meteorological Society*, Volume 79, No 1, pp. 61-78, 1998.
- [26] C. S. Burrus, R. A. Gopinath and H. Guo, *Introduction to Wavelets and Wavelet Transforms*. Prentice Hall, 1998.
- [27] B. Jawerth and W. Sweldens, "An overview of wavelet based multiresolution analyses", *SIAM Rev.*, Volume 36, No 3, pp. 377-412, 1994.
- [28] W. Sweldens and P. Schröder, "Building your own wavelets at home", *Wavelets in Computer Graphics SIGGRAPH'96* Course notes, pp. 15-87, 1996.
- [29] S. Theodoridis and K. Koutroumbas, *Pattern Recognition*, Academic Press, 1998.

ÖZGEÇMİŞ



Biröl KESKİN, 1976 yılında İzmit'te doğdu. İlk öğrenimini İzmit 7.Boru İlkokulu'nda, orta ve lise öğrenimini Kocaeli Anadolu Lisesi'nde tamamlamıştır. 1995 yılında girdiği İstanbul Teknik Üniversitesi, Elektrik-Elektronik Fakültesi, Elektronik ve Haberleşme Mühendisliği Bölümü'nden 1999 yılında mezun olduktan sonra aynı yıl Gebze Yüksek Teknoloji Enstitüsü, Elektronik Mühendisliği Bölümü'nde yüksek lisans eğitimine başlamıştır. Halen ASELSAN A.Ş., Haberleşme Cihazları Grubu, Askeri Haberleşme Sistemleri Müdürlüğü'nde elektronik mühendisi olarak çalışmaktadır.

