

154661

İSTANBUL TEKNİK ÜNİVERSİTESİ ★ FEN BİLİMLERİ ENSTİTÜSÜ

**ÖTÜMLÜ KONUŞMA SESİ KALINTI SİNYALLERİNİN
1/F ÖZELLİĞİNİN İNCELENMESİ**

**YÜKSEK LİSANS TEZİ
Müh. Süleyman BAYKUT
504021318**

**Tezin Enstitüye Verildiği Tarih : 25 Haziran 2004
Tezin Savunulduğu Tarih : 30 Haziran 2004**

Tez Danışmanı : Prof.Dr. Tayfun AKGÜL

Tayfun Akgül

Diğer Jüri Üyeleri : Prof.Dr. Serhat ŞEKER (İ.T.Ü.)

Şeker

Doç.Dr. M. Ertuğrul ÇELEBİ (İ.T.Ü.)

Mehmet E Çelebi

HAZİRAN 2004

ÖNSÖZ

Bu tezin gerçekleşmesinde büyük pay sahibi olan, değerli bilgi ve tavsiyelerini hiçbir zaman esirgemeyen ve bana benden daha fazla güvenen Sn. Prof. Dr. Tayfun AKGÜL'e şükranlarımı sunuyorum.

Maddi-manevi her şeyimi borçlu olduğum hayatımın en değerli varlıkları anne ve babama, çalışmam boyunca beni yalnız ve moralsiz bırakmayan arkadaşlarım Murat, Volkan ve Uğur'a, kıymetli fikir ve yardımlarından dolayı Tolga Esat'a teşekkür ediyorum.

Haziran 2004

Süleyman BAYKUT



İÇİNDEKİLER

KISALTMALAR	v
TABLO LİSTESİ	vi
ŞEKİL LİSTESİ	vii
SEMBOL LİSTESİ	ix
ÖZET	xi
SUMMARY	xii
1. GİRİŞ	1
2. KONUŞMA SESİNİN ÜRETİLMESİ VE MODELLENMESİ	3
2.1. Konuşma Sesi Üretim Sistemi	3
2.1.1. Konuşma Seslerinin Uyarım Sinyallerine Göre Sınıflandırılması	5
2.1.2. Konuşma Seslerinin Ayırt Edici Fiziksel Nitelikleri	9
2.2. Konuşma Sesinin Modellenmesi	9
2.2.1. Sıfır-Kutup Ses Üretim Modeli	10
2.2.1.1. Levinson-Durbin Yineleme Algoritması	15
3. $1/f$ SÜREÇLER	18
3.1. $1/f$ Süreçler ve Öz-benzeşimlilik	18
3.2. $1/f$ Süreçlerin Modellenmesi	19
3.3. Ses Kalıntı Sinyalinin $1/f$ Özelliği	23
4. YÖNTEM	25
4.1. Ötümlü/Ötümsüz/Sessizlik Ayrımı	26
4.1.1. Kısa Zamanlı Enerji	26
4.1.2. Kısa Zamanlı Ortalama Genlik	27
4.1.3. Ortalama Sıfır Seviyesi Geçiş Sayısı	27
4.1.4. Birim Aralıklı Normalize Özilişki Katsayısı	29
4.1.5. Perde Periyodu Kestirimi	30
4.1.5.1. Gerçel Kepstrum Yöntemi	30
4.1.5.2. Kısa Zamanlı Özilişki Yöntemi	32
4.1.5.3. Kısa Zamanlı Genlik Farkı Yöntemi	34
4.2. Konuşma Sesi Kalıntı Sinyalinin Elde Edilmesi	35
4.2.1. Bölmeleme ve Hanning Pencereleme	35
4.2.2. Ön-vurgulama (Beyazlatma)	37
4.2.3. Doğrusal Öngörü Analizi ve Tersinir Süzgeçleme	38
4.3. Dalgacık Tabanlı Analiz	40
4.3.1. Dalgacık Dönüşümü	40

4.3.2. $1/f$ Süreçlerin Dalgacık Tabanlı Analizi ve Öz-Benzeşim Parametresi Kestirimi	48
4.3.3. Ses Kalıntı Sinyalinin Öz-benzeşim Parametresinin Kestirimi	48
4.3.4. Ön-vurgulamanın γ Kestirimine Etkisi	52
5. DENEYLER	55
5.1. Öz-benzeşim Parametresi İle Konuşmacı Cinsiyeti Belirleme	57
5.2. Ses Kalite Artırımı	61
5.3. Öz-benzeşim Parametresinin Ses İletiminde Kullanılması	64
6. SONUÇLAR VE ÖNERİLEN ÇALIŞMALAR	67
6.1. Sonuçlar	67
6.2. Önerilen Çalışmalar	68
KAYNAKLAR	69
ÖZGEÇMİŞ	71



KISALTMALAR

ADD	: Ayrık Dalgacık Dönüşümü
AGS	: Alçak Geçiren Süzgeç
APDD	: Ayrık Parametrelili Sürekli Zamanlı Dalgacık Dönüşümü
AR	: Öz-bağımlı (Autoregressive)
ÇÇA	: Çoklu Çözünürlük Analizi
dB	: Desibel
DD	: Dalgacık Dönüşümü
fBm	: Kesirli Brownian Devinimleri (Fractional Brownian Motions)
FD	: Fourier Dönüşümü
fGn	: Kesirli Gauss Gürültüsü (Fractional Gaussian Noise)
FIR	: Sonlu Uzunluklu Darbe Cevabı (Finite-Length Impuls Response)
KZFD	: Kısa-Zaman Fourier Dönüşümü
Ö/Ö	: Ötümlü/Ötümsüz
SDD	: Sürekli zamanlı Dalgacık Dönüşümü
TDD	: Ters Dalgacık Dönüşümü
YGS	: Yüksek Geçiren Süzge.

TABLO LİSTESİ

	<u>Sayfa No</u>
Tablo 4.1. Dalgacık katsayı ölçekleri ve ilişkili frekans bölgeleri	49
Tablo 5.1. Dil kamburunun ağız içindeki konumuna göre İngilizce ünlü sesler	55
Tablo 5.2. Ünlü sesler ve bu seslere ait γ değerleri	57
Tablo 5.3. Örnek dizilerden elde edilen ortalama değer vektörleri ve kovaryans matrisleri	59

ŞEKİL LİSTESİ

	<u>Sayfa No</u>
Şekil 2.1 : Ses üretim sistemi	3
Şekil 2.2 : Ses Yolu	4
Şekil 2.3.a : Ötümlü ses parçacığı zaman-genlik değişimi	6
Şekil 2.3.b : Ötümsüz ses parçacığı zaman-genlik değişimi	6
Şekil 2.4 : Ötümlü bir ses parçasının güç izgesi	6
Şekil 2.5 : Ötümsüz ses parçasının güç izgesi	7
Şekil 2.6 : Fonem grupları	8
Şekil 2.7 : Ünlü seslerin üretiminde dilin ağız içerisindeki konumu	8
Şekil 2.8 : Konuşma sesi üretimi blok çizeneği	9
Şekil 2.9 : Ses sentez modeli	10
Şekil 2.10 : Doğrusal öngörü hata süzgeci	12
Şekil 2.11 : Analiz/sentez blok çizeneği	13
Şekil 3.1 : Farklı H değerlerine sahip fBm süreçler	22
Şekil 3.2 : Farklı H değerlerine sahip fGn süreçler	23
Şekil 4.1 : Ses kalıntı sinyali çıkarımı blok çizeneği.....	25
Şekil 4.2.a : “Head” sözcüğünün zaman-genlik değişimi.....	28
Şekil 4.2.b : Her bir 512 örneklilik bloğun enerjisi (dB).....	28
Şekil 4.2.c : Her bir 512 örneklilik bloğun ortalama genlik değeri.....	28
Şekil 4.2.d : Herbir 512 örneklilik bloğun ortalama sıfır geçiş sayısı.....	28
Şekil 4.3.a : “Head” sözcüğünün zaman-genlik değişimi.....	29
Şekil 4.3.b : Her bir 256 örneklilik bloğun birim aralıklı normalize ilişkisi	29
Şekil 4.3.c : Her bir 256 örneklilik bloğun ilk öngörü katsayısı (α_1)	29
Şekil 4.4.a : 16 kHz ile örneklenmiş /EH/ sesinin 32 ms’lik bölümü	32
Şekil 4.4.b : Hanning pencerelenmiş ses parçası	32
Şekil 4.4.c : Sinyalin FD’nün mutlak değeri	32
Şekil 4.4.d : Sinyalin FD’nün mutlak değerinin 10 tabanında logaritması	32
Şekil 4.4.e : İlk 30 değeri sıfırlanmış gerçel kepstrum	32
Şekil 4.5.a : 32ms uzunluğunda ötümlü ses parçası.....	34
Şekil 4.5.b : %55 oranında merkez kırpma yapılmış ses sinyali	34
Şekil 4.5.c : Sinyalin kısa zamanlı öz-ilişki işlevi	34
Şekil 4.5.d : Merkez kırpılmış sinyalin öz-ilişki işlevi	34
Şekil 4.6.a : Ötümlü ses parçası	35
Şekil 4.6.b : %40’ı oranında merkez kırpma yapılmış ses sinyali	35
Şekil 4.6.c : Sinyalin kısa zamanlı genlik farkı işlevi	35
Şekil 4.6.d : Merkez kırpılmış sinyalin genlik farkı işlevi	35
Şekil 4.7.a : Hanning pencere	36
Şekil 4.7.b : Dikdörtgen pencere	36
Şekil 4.7.c : Pencerelerin güç izgeleri	36
Şekil 4.8 : Kayan Hanning Pencere	37
Şekil 4.9.a : 512 Örneklilik ses sinyali	38

Şekil 4.9.b	: Sinyalin önvurgulanmış hali	38
Şekil 4.9.c	: Sinyalin güç izgesi	38
Şekil 4.9.d	: Ön-vurgulama süzgecinin güç izgesi	38
Şekil 4.9.e	: Ön-vurgulanmış sinyalin güç izgesi	38
Şekil 4.10	: Ses parçasının güç izgesi ve doğrusal öngörü zarfı	39
Şekil 4.11.a	: Konuşma sesi sinyalinin güç izgesi	39
Şekil 4.11.b	: Tersinir süzgecin güç izgesi	39
Şekil 4.11.c	: Ses kalıntı sinyalinin güç izgesi	39
Şekil 4.12.a	: 512 örnekleli ötümlü ses parçası	40
Şekil 4.12.b	: Hanning pencerelenmiş ses parçası	40
Şekil 4.12.c	: Ön-vurgulanmış ses parçası	40
Şekil 4.12.d	: Ses kalıntı sinyali	40
Şekil 4.13.a	: Durağan sinyal	41
Şekil 4.13.b	: Zaman ilerledikçe frekansı değişen, durağan olmayan sinyal.....	41
Şekil 4.13.c	: Durağan sinyalin güç izgesi	41
Şekil 4.13.d	: Durağan olmayan sinyalin güç izgesi	41
Şekil 4.14	: Zaman-frekans çözünürlükleri	42
Şekil 4.15	: Dalgacık dönüşümünde ölçekleme ve öteleme işlemleri	44
Şekil 4.16	: Ayrık dalgacık dönüşümü algoritması	47
Şekil 4.17	: Kalıntı sinyalinin dalgacık tabanlı analizi	49
Şekil 4.18	: 6.-9. Ölçeklerin dalgacık katsayıları	50
Şekil 4.19	: 6.-9. Ölçeklerin dalgacık katsayılarının dağılımları	51
Şekil 4.20	: Ölçeklerdeki dalgacık katsayılarının normalize öz-ilişki dizileri .	51
Şekil 4.21	: Dalgacık katsayılarının değişinti gelişimi	52
Şekil 4.22	: Ses üretiminin doğrusal modeli	53
Şekil 4.23.a	: Hanning pencerelenmiş 512 örnekleli ses parçacığı	54
Şekil 4.23.b	: Ses kalıntı sinyali	54
Şekil 4.23.c	: Dalgacık katsayılarının değişinti gelişimleri	54
Şekil 4.23.d	: Önvurgulanmış, pencerelenmiş 512 örnekleli ses parçacığı	54
Şekil 4.23.e	: Ön-vurgulanmış ses parçacığının kalıntı sinyali	54
Şekil 5.1	: Ünlü seslerin zaman-genlik değişimleri	56
Şekil 5.2	: γ kestirimi blok çizeneği	56
Şekil 5.3	: Olasılık yoğunluk işlevleri	58
Şekil 5.4	: Ortak olasılık yoğunluk işlevleri	60
Şekil 5.5	: Dar bantlı sinyalden geniş bantlı sinyale geçiş blok çizeneği	62
Şekil 5.6.a	: (a) 8kHz ile örneklenmiş 32ms'lik ötümlü ses parçacığı	62
Şekil 5.6.b	: (b) Ses kalıntı sinyali	62
Şekil 5.7	: Değişinti gelişimine ve fazladan bir ölçek ekleme	63
Şekil 5.8.a	: Sentezlenen geniş bantlı ses sinyali	63
Şekil 5.8.b	: Üretilen kalıntı sinyali	63
Şekil 5.9	: Ses sentezi blok çizeneği	65
Şekil 5.10.a	: 512 örnekleli ötümlü /UW/ sesi	65
Şekil 5.10.b	: Ses kalıntı sinyali	65
Şekil 5.10.c	: Sentezlenen ses sinyali	65
Şekil 5.10.d	: Sentezlenen kalıntı sinyali	65

SEMBOL LİSTESİ

P	: Perde periyodu
v_k	: Ses yolu aktarım işlevinin sıfırları
ζ_k	: Ses yolu aktarım işlevinin kutupları
b_k	: Ses yolu aktarım işlevinin pay polinomunun katsayıları
a_k	: Sistem parametreleri
q	: Ses yolu aktarım işlevinin pay polinomu derecesi
p	: Sistem derecesi (Öngörü derecesi)
G	: Kazanç sabiti
E_n	: Karesel ortalama öngörüm hatası
Γ	: Öz-ilişki matrisi
α	: Öngörü katsayıları vektörü
ρ	: Özilişki vektörü
k_i	: Kısmi ilişki katsayıları” (partial correlation coefficients, PARCOR)
H	: Hurst parametresi
γ	: Öz-benzeşim parametresi
σ_x^2	: Sinyalin değışintisi
D	: Fraktal boyutu
t	: Sürekli zaman değışkeni
f	: Frekans değışkeni
f_s	: Örnekleme frekansı
ω	: Açısal Frekans
σ, c, a	: Sabit, pozitif, gerçel sayı
t_0	: Zaman artımı
$\varepsilon, \Delta, \sigma_A^2, \sigma_U^2$	: Küçük, sabit, pozitif gerçel sayı
$\mathcal{N}$	: Gauss dağılımı (Normal dağılım)
τ	: Aralık (lag) değeri
E_s	: Kısa zamanlı enerji
N	: Pozitif tamsayı
M_s	: Kısa Zamanlı Ortalama Genlik
Z_s	: Ortalama sıfır seviyesi geçiş sayısı
C_1	: Birim aralıklı normalize özilişki katsayısı
C^-	: Alt kırpma sınırı
C^+	: Üst kırpma sınırı
π	: Pi sayısı
b	: Sabit önvurgulama süzgeç katsayısı ($0.9 \leq b \leq 1$)
$*$	: Eşlenik ifadesi
t'	: Öteleme parametresi
c_ψ	: Uygunluk sabiti
m	: Ölçekleme indisi
n	: Öteleme indisi
a_0	: Ölçekleme artımları
τ_0	: Öteleme artımları

$\mathbf{a}_n^m$	: Yaklaşık dalgacık katsayıları
$\mathbf{x}_n^m$	: Detaylı dalgacık katsayıları
T	: Örnekleme aralığı
λ	: γ dizisi
σ_λ^2	: λ dizisinin değışintisi
μ_λ	: λ dizisinin ortalaması
$\mathbf{M}_i$	: Ortalama değer vektörü
$\mathbf{W}_i$	: Kovaryans vektörü



ÖTÜMLÜ KONUŞMA SESİ KALINTI SİNYALLERİNİN 1/F ÖZELLİĞİNİN İNCELENMESİ

ÖZET

Bu çalışmada, İngilizce'deki ötümlü seslerin kalıntı sinyallerinin $1/f$ özelliği dalgacık tabanlı analiz ile incelenmektedir. Ele alınan ses üretim modelinde, ses yolu doğrusal bir sistem olarak modellenmiştir. Bu sistemi uyaran sinyal $1/f$ gürültüsüdür ve ses kalıntı sinyali olarak adlandırılmaktadır. Ünlü sesler en belirgin ötümlü sesbirimi grubunu oluştururlar. Bu sebeple incelemede erkek ve kadın konuşmacılar tarafından üretilen /Y/, /IH/, /EI/, /EH/, /AE/, /ER/, /AH/, /AW/, /OA/, /OO/, /UW/ ve /UH/ ünlü sesleri ele alınmıştır. Kalıntı sinyalini elde etme aşamasında, ilkin, sesin güç izgesi beyazlatılmakta ve beyazlatılmış izge üzerine doğrusal öngörü analizi uygulanarak ses yolu parametreleri elde edilmektedir. Daha sonra tersinir süzgeçleme ile süzgecin uyarıcısı olan kalıntı sinyaline ulaşılır. Kalıntı sinyaline dalgacık dönüşümü (wavelet transform) uygulanarak dalgacık katsayıları elde edilmektedir. Kalıntı sinyalinin $1/f$ davranışı gösterdiği varsayımıyla, bu sinyalin güç izgesinin "güç-yasası" ilişkisine sahip olduğu söylenebilir. Güç yasası ilişkisi dalgacık katsayılarının değişinti gelişimlerinde de görülmektedir. Bu çalışmada dalgacık katsayılarının değişinti gelişimleri kullanılarak öz-benzeşim parametreleri kestirilmiş, bu parametrelerin kadın ve erkeklere ait farklı seslerdeki davranışları incelenerek cinsiyet ayrımında etkin bir kıstas olup olamayacağı araştırılmıştır. Konuşma sesinin doğallığının ana etmeni, uyarıcı kalıntı sinyalinin rasgele öz-benzeşimli yapısında gizlidir. Ses kalıntı sinyalinin öz-benzeşimli yapısının bir parametre ile modellenebileceği gösterilmiş ve bu parametrenin kullanılabileceği alanlar incelenmiştir. Öz-benzeşimli yapı araştırılarak, gerçeğe daha yakın yapay ses üretiminde kullanılabilecek uyarıcı sinyalin ne şekilde olması gerektiği hakkında bilgi edinilmeye çalışılmıştır.

ANALYSIS OF $1/f$ CHARACTERISTICS OF VOICED SPEECH RESIDUAL SIGNAL

SUMMARY

In this thesis, the $1/f$ nature of voiced speech residual signal is examined by wavelet based analysis method. Speech signals are generally classified as voiced or unvoiced. Voiced speech signals are considered to be generated by the vocal cords vibration signal exciting the vocal tract. In our model, the vocal tract is considered as a linear system. The excitation signal is the $1/f$ noise which is called the *speech residual signal*. Since the vowels are the largest and the most evident voiced phoneme group, we study some of these vowels, i.e., /IY/, /IH/, /EI/, /EH/, /AE/, /ER/, /AH/, /AW/, /OA/, /OO/, /UW/ and /UH/ which are generated by several men and women. To extract the speech residual, first, we force to whiten the power spectrum of the speech signal by using a preemphasize filter and then perform the linear predictive analysis on the whitened speech to obtain the vocal tract parameters. The speech residual signal is obtained by the inverse filter. A wavelet decomposition technique is applied to the residual signal to obtain the wavelet coefficients. The power-law relationship is observed in the progression of the variances of these coefficients along scales. The self-similarity parameters (the slope of the progression) are then estimated. We investigate and compare the behavior of the self-similarity parameters for the speeches of 40 men and 40 women. A speaker gender identification method is explored based on the self-similarity parameter. The random fractalness of the speech residual signals is considered to be a key factor in their naturalness. In this thesis, the random fractalness of the speech residual signal is explored in order to learn how to design more realistic source signals for speech synthesis.

1. GİRİŞ

Çevre ile iletişimi sağlayan konuşma, insan yaşamı için vazgeçilmez bir özelliktir. Gelişen teknolojiyle modern ses cihazlarının yaygınlaşması, konuşma sesini daha verimli bir şekilde inceleme olanağı sunmuştur. Bu sayede, konuşma sesi sentezi, yüksek kalitede ses iletimi ve ses tanıma gibi pek çok yeni uygulama alanları geliştirilmiştir.

Ses üretim sistemi, birbirinden çok farklı tınlar, tonlamalar ve ses bükümleri meydana getirebilen, mükemmel bir tasarım ve işleyişe sahiptir. Ses üretimi, “uyarıcı sinyal” ve “tınlatıcı” olmak üzere iki bölüme ayrılabilir. Uyarıcı sinyal, ötümlü seslerde, ses tellerinin yarı periyodik titreşimiyle oluşurken, ötümsüz seslerde gürültü benzeri bir sinyal olur. Bu sinyal “ses kalıntı sinyali” olarak adlandırılmaktadır. Ses üretim sistemindeki ses yolu (vocal tract) ise tınlatıcı görevi görmektedir. Çok karmaşık bir yapısı olan konuşma sesinin iletilmesinde ve yapay olarak üretilmesinde karşılaşılan önemli problemlerden biri “doğallık” problemidir. Ses yolunun tınlatma etkisi, çeşitli yöntemlerle gerçeğe yakın bir şekilde modellenmesine rağmen, üretilen sesin makinemsi bir ses olduğu gözlenir. Çoğu uygulamada sentezlenen ses gayet anlaşılır olmuş fakat mekaniklikten ve vızıltıdan kurtulmak mümkün olmamıştır. Bunun sebebi uyarıcı sinyal olarak basit bir darbe dizisinin kullanılmasıdır. Doğala yakın ses üretimini gerçekleştirebilmek için doğallık olgusunun modellenmesi gerekliliği ortaya çıkmaktadır. Yapılan araştırmalar süreçlerin doğallıklarının, rasgele fraktallı (öz-benzeşimli) yapılarında saklı olduğunu göstermiştir. Dolayısıyla süreçlerin fraktallı yapılarının ölçülmesi ve parametrik hale getirilmesi, doğallığın ölçülmesi ve modellenmesi ile aynı anlama gelecektir. Rasgele fraktallı yapı ses sinyallerinde de gözlenmektedir. Gerçekte, sesin doğallığının ana etmeni, uyarıcı sinyalin rasgele fraktallı yapısında gizlidir. Bu yüzden ses üretiminde uyarıcı sinyalin fraktallı yapısı anlaşılabilir ve modellenebilirse, öngörümüze göre, doğala çok daha yakın ses üretimi mümkün olabilecektir. Ötümlü seslerin kalıntı sinyalleri $1/f$ özellik gösteren öz-benzeşimli süreçlerdir. Bu süreçler, öz-benzeşim parametresi γ ile ifade edilebilirler. Dolayısıyla γ değerinin kestirimi ile sinyalin fraktallı yapısı

modellenebilmektedir. Bu fraktallı yapının modellenmesinde, öz-benzeşimli sinyalleri modellemekte önemli bir paya sahip “Kesirli Brownian Devinimleri (Fractional Brownian Motions, fBm)” ve fBm’lerin türevleri olan “Kesirli Gauss Gürültüleri (Fractional Gaussian Noises, fGn)” kullanılmaktadır. Bu çalışmada, ses kalıntı sinyali fBm türü bir sinyal olarak modellenmiştir.

İlk olarak, ötümlü bir ses sinyalinden ses kalıntı sinyalinin çıkarım aşamaları anlatılmaktadır. Daha sonra, ötümlü ses kalıntı sinyallerinin öz-benzeşimli yapıları dalgacık tabanlı analiz ile incelenmiştir. Kestirilen öz-benzeşim parametrelerinin kadın ve erkeklere ait farklı seslerdeki davranışları gözlenerek, bu parametrenin cinsiyet ayrımında bir kıstas olup olamayacağı araştırılmıştır. Ayrıca kalıntı sinyalinin öz-benzeşimli yapısı γ parametresi ile modellenerek, ses iletiminde ve gerçeğe daha yakın yapay ses üretiminde kullanılabilecek uyarıcı sinyalin ne şekilde olması gerektiği incelenmiştir.

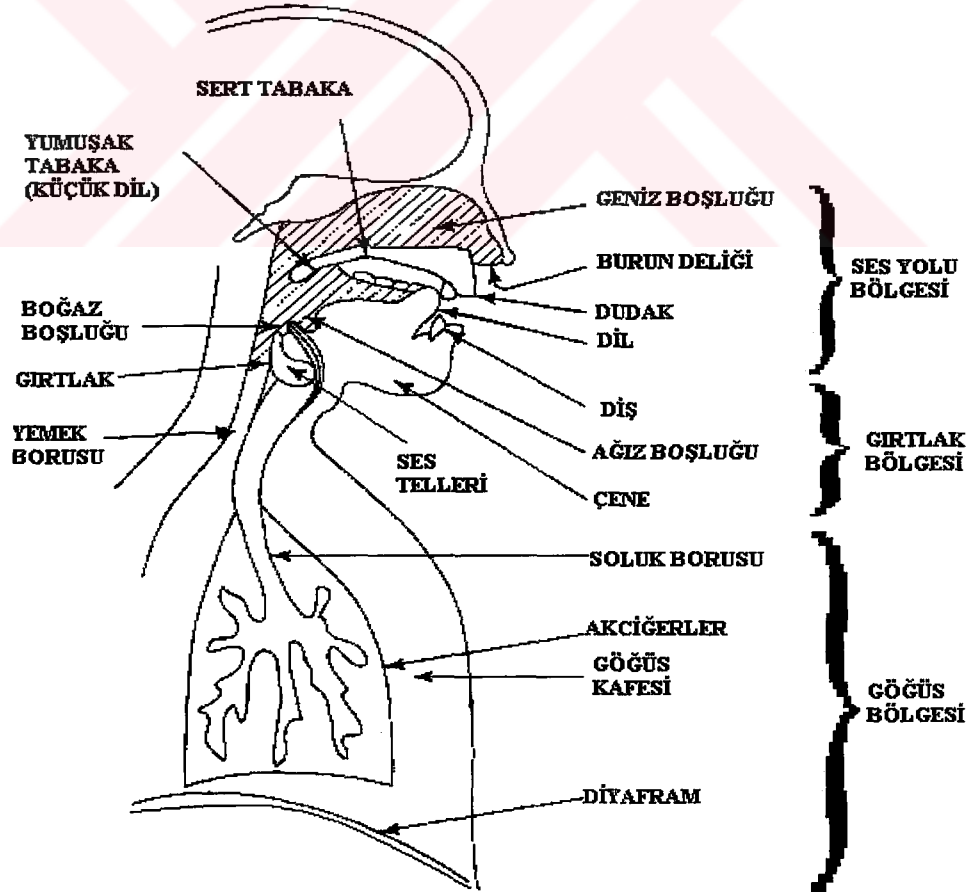
Bu tez 6 bölümden oluşmaktadır. 1. Bölüm giriş bölümüdür. Bölüm 2’de ses üretim sistemi, ses sinyalinin yapısı ve sıfır-kutup ses üretim modeli anlatılmaktadır. Çoğu uygulamada, tüm-kutup ses üretim modeli sıfır-kutup modeline tercih edilmektedir. Tüm-kutup modelin sistem parametrelerinin doğrusal öngörü analizi ile kestirimi yine Bölüm 2’de anlatılmaktadır. Bölüm 3’te fraktallı yapı, $1/f$ süreç ve öz-benzeşimli süreç kavramları açıklanmaktadır. 4. Bölümde, sesin analizinde ve sentezinde büyük önem taşıyan ötümlü/ötümsüz ayrımı, ses kalıntı sinyalinin elde ediliş aşamaları, dalgacık dönüşümü ve son olarak dalgacık tabanlı analiz ile öz-benzeşim parametrelerinin kestirim işlemleri anlatılmaktadır. Bölüm 5’te, farklı kişilerce söylenen sözcüklere 4. Bölümünde anlatılan işlemler uygulanmıştır. Varılan sonuçlar ve önerilen çalışmalar Bölüm 6’da tartışılmıştır.

2. KONUŞMA SESİNİN ÜRETİLMESİ VE MODELLENMESİ

2.1 Konuşma Sesi Üretim Sistemi

Konuşma sesi, ses üretim sistemini oluşturan organların istemli hareketleri sonucu meydana gelen akustik bir basınç dalgasıdır [1,2]. Konuşma sesi üretim sistemi genel olarak üç temel bölümde incelenir: göğüs bölgesi, gırtlak bölgesi ve ses yolu bölgesi [1]. Bu bölümler Şekil 2.1’de gösterilmişlerdir.

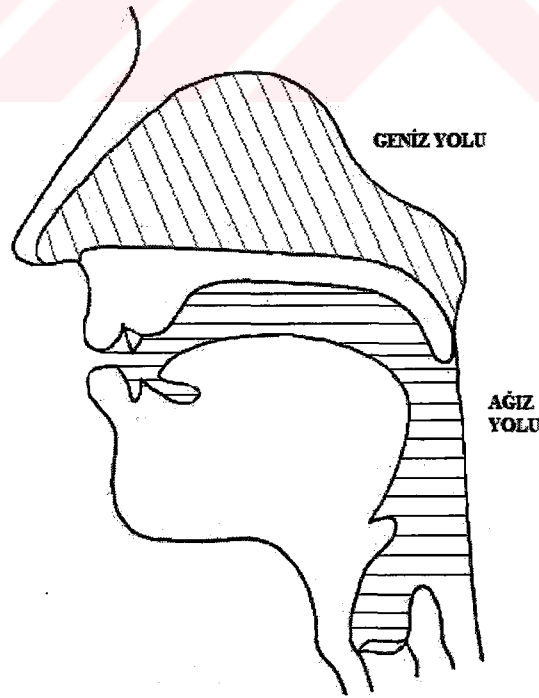
Göğüs bölgesi organları diyafram, göğüs kafesi ve akciğerlerdir. Bu organlar, ses üretiminde, gırtlak bölgesindeki titreşimi oluşturmak için gerekli hava akımını sağlamakla görevlidirler.



Şekil 2.1 Ses üretim sistemi, [2]

Gırtlak bölgesi organları göğüs bölgesinden gelen hava akımını modüle ederek ses yolu için kaynak sinyalinin üretimini sağlarlar. Soluk borusu ve ses telleri bu bölgede bulunmaktadır. Soluk borusu yaklaşık 13 cm uzunluğunda 1.5-2 cm çapındadır. Akciğerlerden gırtlığa kadar uzanarak akciğerlerden gelen hava akımının ses tellerine ulaşmasını sağlar [1,2]. Ses telleri iki adettir ve nefes borusunun üst kısmında gırtlak bölgesindeki kalkansı kıkırdaklar arasında bulunur. Erkeklerin ses telleri kadınlarınkiler göre daha kalın ve uzundur. Ses tellerinin uzunluğu erkeklerde ortalama 20-25 mm kadınlarda ise 16-20 mm uzunluğundadır [1].

Gırtlak ile başlayıp dudaklarda sona eren bölüm “ağız yolu” (oral track) olarak bilinmektedir. Üst damağın arka kısmından burun deliklerine kadar olan kısım ise “geniz yolu” (nasal tract) adını alır. Ses üretiminde birlikte görev alan ağız yolu ve geniz yolunun ortak adı “ses yolu”dur. Geniz yolu ve ağız yolu Şekil 2.2’de gösterilmiştir. Ağız yolunun uzunluğu yetişkin erkeklerde 16-18 cm civarında, kadınlarda ise 13-14 cm civarındadır. Ses oluşumu sırasında, üretilecek sesin özeliğine göre, ses yolunun kesit alanı 0-25 cm² arasında değişir [2]. Geniz yolu da ses çeşitliliği sağlamada önemli bir bölüm olup, yetişkin erkeklerde yaklaşık 12 cm civarındadır. Kadınlarda geniz yolu erkeklerinkine göre daha kısadır [2].



Şekil 2.2 Ses yolu

Ses üretimi akustik bir süzgeçleme işlemi şeklinde düşünülebilir. Buradaki süzgeç ses yolunu modelleyen süzgeçtir. Ses üretim sisteminde bulunan ses telleri, küçük dil, damaklar, burun, dil, dişler ve dudaklar sesin niteliğini belirleyen önemli unsurlar olup değişik konumlara girerek farklı seslerin oluşmasını sağlarlar [1-3].

Sesin oluşabilmesi için ses tellerinin akciğerlerden gelen hava ile uyarılıp titreşmesi gerekmektedir ancak sesin oluşumu için bu yeterli değildir. Ses telleri tarafından üretilen bu titreşim bir tınlaticıdan (resonator) geçerse anlamlı hale gelir. Bu düzenek müzik aletlerinde de görülmektedir . Örneğin keman tellerinde oluşturulan titreşim, tellerin arka kısmındaki ahşap kasa olmasa, kötü bir gıcırtıdan başka bir şey duyulmaz. Kemandaki ahenkli sesi duyabilmemizi kasa bölümü mümkün kılar. Konuşma sesi oluşumunda da vücudumuzdaki tınlaticılar, kemandaki ahşap kasanın görevini üstlenir. Gırtlak, ağız boşluğu ve geniz boşluğu ana tınlaticılardır. Göğüs boşluğu, sırt, kafatası hatta karın bölgesi ve vücudun diğer organları ses üretiminde yardımcı tınlaticılar olarak görev yaparlar [1].

2.1.1 Konuşma Seslerinin Uyarım Sinyallerine Göre Sınıflandırılması

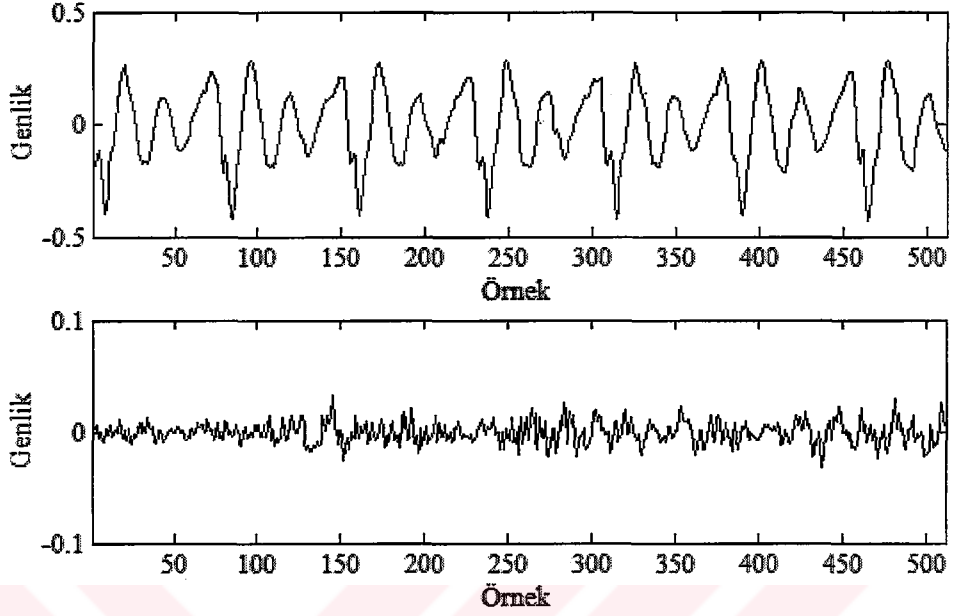
Konuşma seslerinin oluşumunda genel olarak üç uyarım sinyali mevcuttur: yarı-periyodik uyarım sinyali, gürültü uyarım sinyali ve darbe uyarım sinyali. Üç uyarımın dışında, bunların kombinasyonu olan karışık, fısıltı ve sessiz diye adlandırılan uyarım sinyalleri de bulunmaktadır [1-3].

Uyarım sinyali şekli, konuşma sesinin en önemli özelliklerinden biridir. Ses tellerinin titreşimiyle oluşan sözde periyodik (quasi-periodic) sinyalin ses yolunu uyarması sonucunda oluşan seslere “ötümlü (voiced) sesler” denir. Ötümlü ses üretimi durumunda ses telleri belli bir periyotta açılıp kapanarak yarı periyodik uyarım sinyalini üretirler.

Hava akımının daraltılmış ses yolundan geçmeye zorlanmasıyla bir gürültü sinyali oluşur. Bu gürültü uyarım sinyali kullanılarak üretilen seslere de “ötümsüz (unvoiced) sesler” adı verilir. Ötümsüz seslerin üretiminde ses tellerinde bir titreşim olmamaktadır.

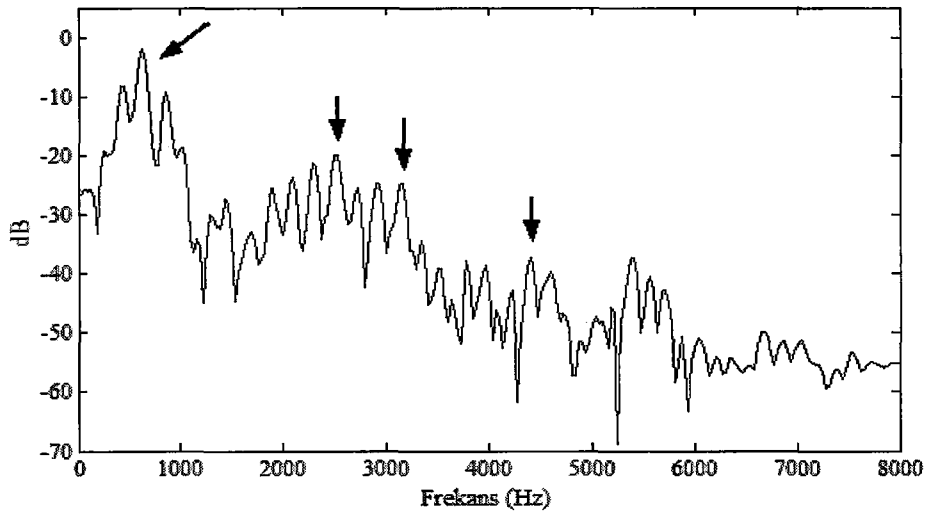
Ötümlü sesler yarı periyodik bir yapı sergilerken, ötümsüz sesler gürültü benzeri bir yapıya sahiptirler. Örneğin “HEAD” sözcüğündeki /EH/ ünlü sesi ötümlü bir ses, /H/ sesi ötümsüz bir sestir. Şekil 2.3.a’da 16 kHz ile örneklenmiş 512 örneklik /EH/

ötümlü sesinin zaman-genlik değişimi, (b)'de ise 512 örneklilik ötümsüz /H/ sesinin zaman genlik değişimi gösterilmektedir.

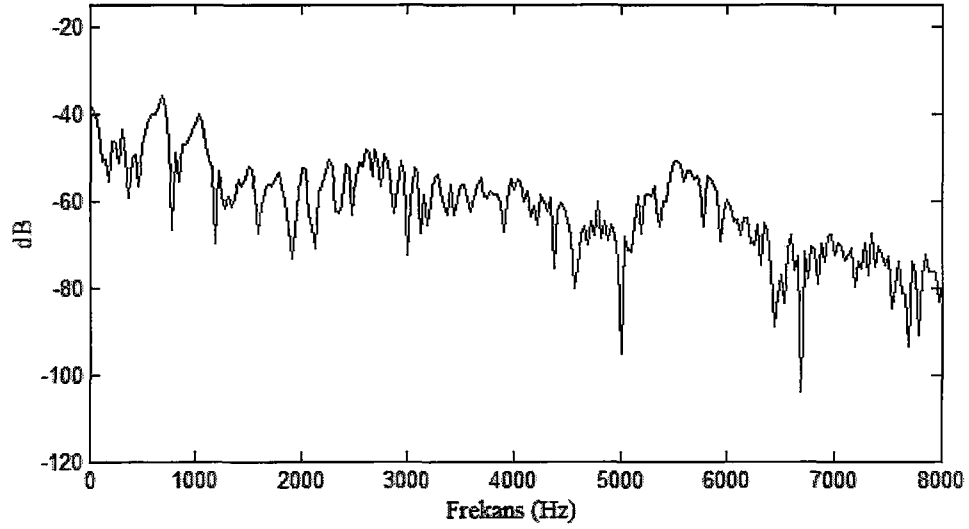


Şekil 2.3 (a) Ötümlü ses parçasığı zaman-genlik değişimi, (b) Ötümsüz ses parçasığı zaman-genlik değişimi

Ötümlü ve ötümsüz seslerin güç izgeleri de farklıdır. Şekil 2.4'te ötümlü bir ses parçasının güç izgesi gösterilmektedir. Şekil 2.5'te de ötümsüz ses kalıntı sinyalinin güç izgesi gösterilmektedir. Ötümlü seslerin güç izgeleri incelendiğinde, ses yolunun tınlama frekanslarına karşı gelen frekanslarda yüksek genlikli tepcikler görülür. Bu frekans değerlerine “formant” adı verilir.



Şekil 2.4 Ötümlü bir ses parçasının güç izgesi ve formantlar



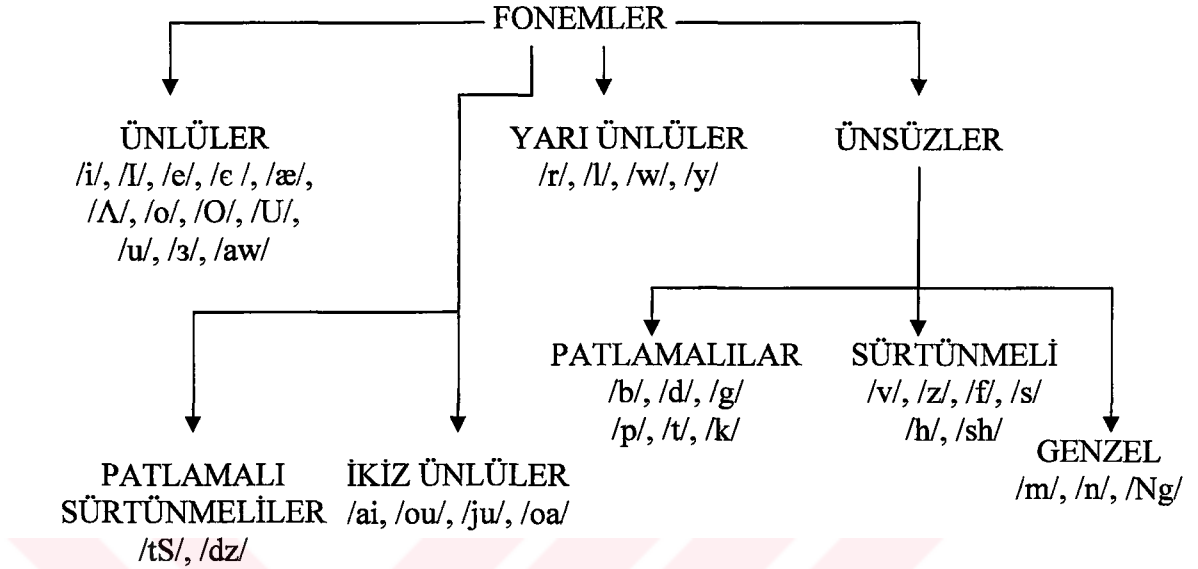
Şekil 2.5 Ötümsüz ses parçasının güç izgesi

Ses yolu işlevindeki her bir değişken (dilin yeri, dudakların şekli vs.) değiştikçe oluşan sese ait formant değerleri de değişir. Bir seste sayısız formant mevcuttur. Buna karşılık, ses telleri tarafından üretilen akustik enerji 5000 Hz'in altındaki frekanslarda yoğunlaştığı için ses ilk 3-4 formant ile tanımlanmaktadır [2,3].

- 1. Formant: Frekans değeri en düşük fakat şiddeti en yüksek formanttır. Ünlü seslerde 300 Hz ile 1000 Hz arasında değerler alır. Dil, ağız tavanına ne kadar yakınsa yani ağız boşluğu ne kadar küçükse formant'ın değeri o kadar düşüktür. Ağız boşluğunun hacmi arttıkça bu formantın değeri de yükselir [2,3].
- 2. Formant: Ses yolunun uzunluğuyla ilgilidir. 850 Hz ile 2500 Hz arasında değerler alır. Ses yolunun uzunluğu arttıkça 2. Formantın değeri azalır. Örneğin “boot” sözcüğündeki /U/ sesi oluşumunda dudaklar yuvarlatılıp öne doğru uzatılır bu da ses yolunun boyunu uzatmış olur. Bu sesin 2. formant değeri 850 Hz gibi küçük bir değerdir [2,3].
- 3. Formant: 1. ve 2. Formant değerlerinden daha yüksek değerler alır. Sesin kalitesini belirlemede önemlidir.

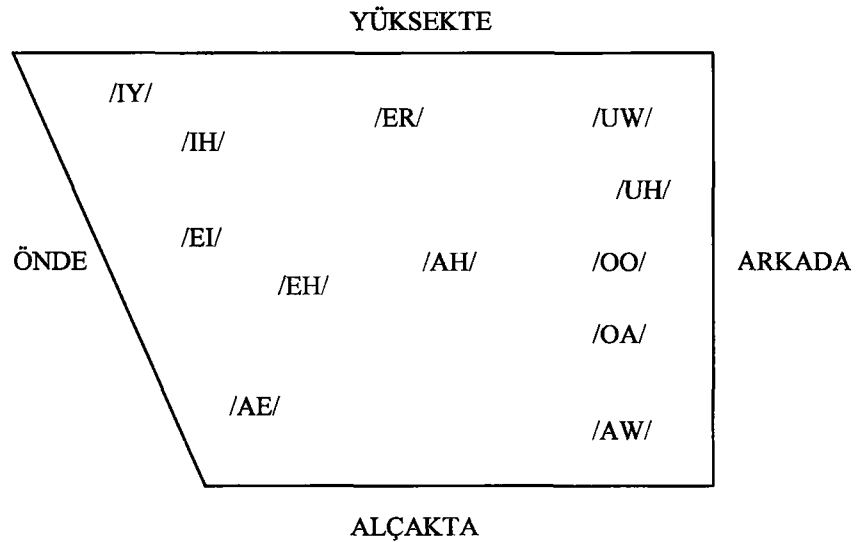
Dilin ayırt edici en temel birimine fonem (phoneme) adı verilir. İngilizce’de yaklaşık 40 adet fonem mevcuttur [3]. Bu fonemlerin sınıflandırması Şekil 2.6’da gösterilmektedir. Fonemler ses tellerinin hareketleri ve ses yolunun şekil özelliklerine göre birbirlerinden farklılık gösterirler. Örneğin ses tellerinin titreşimi,

dilin ağız içerisindeki yeri ve şekli, küçük dilin durumu, dudakların durumu gibi etmenler farklı fonemlerin oluşmasını sağlarlar.



Şekil 2.6 Fonem grupları

Ünlü sesler en belirgin ve en geniş ötümlü ses grubunu oluştururlar. Dilin kambur kısmının ağız içerisinde ileride, ortada ve geride olması durumlarına göre 3 ayrı gruba ayrılırlar. [1-3]. Bu seslerin üretiminde ses telleri belli bir frekansta sözde-periyodik salınım yapmaktadırlar. Ünlü seslerin üretiminde, dilin ağız içindeki konumu ve üretilen ünlü sesler Şekil 2.7’de gösterilmiştir.



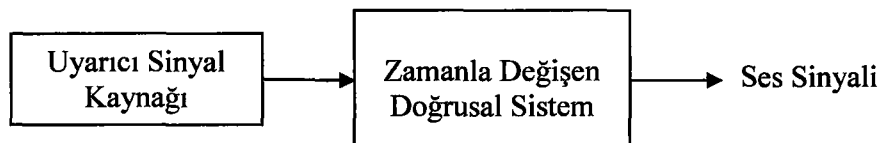
Şekil 2.7 Ünlü seslerin üretiminde dilin ağız içerisindeki konumu

2.1.2 Konuşma Seslerinin Ayırt Edici Fiziksel Nitelikleri

- *Temel Frekans:* Ses tellerinin 1 saniyede oluşturduğu bir tam açılıp kapanma periyodu sayısına “sesin temel frekansı” denir. Temel frekans “perde frekansı” olarak da bilinir. Ses tellerinin uzunluğu, kalınlığı, kütlesi, gerginliği, elastikiyet kıvamı ve nefes borusunda oluşan basınç temel frekansı belirleyen etmenlerdir. Ses tellerinin uzunluğu, kalınlığı ve kütlesi arttıkça, üretilen sesin temel frekansı düşük olur. Ses tellerinin gerginliğinin ve ses telleri üzerinde oluşan hava basıncının artması ise temel frekansı yükseltir. Temel frekans arttıkça üretilen ses inceler (tizleşir). Erkek ve kadın sesleri arasındaki en belirgin ayrım, ses tellerinin yapısından ve şeklinden kaynaklanan perde frekansı farklılığıdır. Erkeklerde perde frekansı 110-150 Hz arasında değişirken kadınlarda 200-300 Hz arasında değişir [1-3].
- *Sesin Şiddeti:* Sesin şiddeti, ses dalgasının yayılma doğrultusuna dik 1cm^2 'lik bir alanda 1 sn'de verdiği ses enerjisi olup genellikle dB olarak ölçülür. Sesin şiddeti ses tellerinde üretilen titreşimin genliğiyle doğru orantılıdır. Sesin şiddetini etkileyen esas etken ciğerler tarafından oluşturulan basınçtır. İnsan sesi yaklaşık 130 dB'ye kadar yükselebilir [1-3].
- *Sesin Genişliği:* İnsan sesinin erişebileceği en düşük temel frekanslı ses ile en yüksek temel frekanslı ses arasındaki farka “ses genişliği” adı verilir.
- *Tını:* Ses tellerindeki titreşim sinyali vücudumuzdaki tınlatıcılar tarafından modüle edilerek farklı tınılarda sesler oluşturmamızı sağlarlar. Sesin tınısı kişiden kişiye farklılıklar göstererek sesin özgünlüğünü sağlar.

2.2 Konuşma Sesinin Modellenmesi

Konuşma sesi üretim sisteminin matematiksel olarak bir modele oturtulması, ses üretiminin daha iyi anlaşılabilmesi ve ses sinyaline ait özelliklere ulaşılmasını kolaylaştırır [1-3]. Şekil 2.8'de konuşma sesi üretimi basit bir blok çizeneği ile gösterilmiştir.

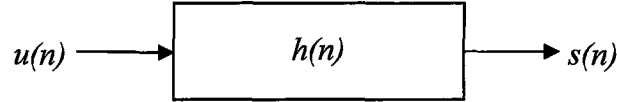


Şekil 2.8 Konuşma sesi üretimi blok çizeneği

Konuşma seslerinin üretiminde, kaynak sinyal “uyarıcı sinyal” bloğu ile modellenir. Konuşma sesi, istatistiksel özellikleri kısa zaman aralıklarında değişen, durağan olmayan bir sinyaldir. Ses yolunun parametreleri sürekli değişmektedir. Bu yüzden ses yolu, “zamanla değişen doğrusal sistem” bloğu ile modellenmektedir.

2.2.1 Sıfır-Kutup Ses Üretim Modeli

Sıfır-kutup modelinde, ses sinyali, uyarıcı sinyal ile ses yolu birim darbe cevabının evrişimi ile oluşur. Şekil 2.9’da ses üretim modeli gösterilmektedir.



Şekil 2.9 Ses sentez modeli

$s(n)$ ses sinyali, $h(n)$ ses yolunun birim darbe cevabı, $u(n)$ uyarıcı sinyal sinyal olmak üzere, aşağıdaki evrişim yazılabilir:

$$s(n) = u(n) * h(n) \quad (2.1)$$

z-bölgesine geçilirse:

$$S(z) = U(z)H(z) \quad (2.2)$$

elde edilir. Ses yolunu modelleyen aktarım işlevi, $H(z)$, sıfır ve kutupları bulunan bir aktarım işlevi gibi düşünülmektedir [2]:

$$H(z) = G \frac{\sum_{k=0}^q b_k z^{-k}}{1 - \sum_{k=1}^p a_k z^{-k}} = G \frac{\prod_{k=1}^q (1 - v_k z^{-1})}{\prod_{k=1}^p (1 - \zeta_k z^{-1})} \quad (2.3)$$

Burada, v_k ’lar aktarım işlevinin sıfırlarını, b_k ’lar pay polinomunun katsayılarını ($k=0,...,q$), ζ_k ’lar aktarım işlevinin kutuplarını, a_k ’lar ise payda polinomunun katsayılarını ($k=1,...,p$) temsil etmektedirler. G kazanç sabitidir. q pay polinomunun derecesi, p ise payda polinomunun derecesidir. Aktarım işlevindeki kutuplar ağız yolu tınlatma etkisini modellerken, sıfırlar ise geniz yolu etkisini ve ses yolundaki bazı enerji soğurucu bölgeleri modellemektedir. Ses yolunun şekli, dolayısıyla tınlatma etkisi ve ses tellerinde oluşan titreşim sinyalinin niteliği zamanla değişmektedir. Buna rağmen konuşma sesi sinyalleri 20-40 ms gibi küçük zaman dilimlerinde durağan kabul edilmektedirler [2,3]. Kısa bir süre içerisinde ses yolunun

şeklinin ve uyarıcı sinyalin niteliğinin değişmediği varsayılmaktadır. Bu yüzden ses sinyali çok küçük parçacıklara ayrılarak işlenir.

İnsan kulağı sıfırlardan çok kutuplara daha duyarlıdır [2]. Bu yüzden $H(z)$, genellikle tüm-kutup bir işlev olarak seçilir. Sonuç olarak (2.3)'te belirtilen aktarım işlevi aşağıdaki gibi, sadece kutuplardan oluşan tüm-kutup bir aktarım işlevi halini alır:

$$H(z) = G \frac{1}{1 - \sum_{k=1}^p a_k z^{-k}} \quad (2.4)$$

Verilen bir ses sinyalinden a_k sistem parametrelerinin kestiriminde “Doğrusal Öngörü Analizi (Linear Predictive Analysis)” etkin bir yöntemdir.

Doğrusal öngörü analizindeki temel düşünce, her bir konuşma sesi sinyali örneğinin kendisinden önce gelen örneklerin doğrusal bir birleşimi şeklinde ifade edilebileceğidir. Bağntı (2.2) ve (2.4) birleştirilirse:

$$H(z) = \frac{S(z)}{U(z)} = \frac{G}{1 - \sum_{k=1}^p a_k z^{-k}} \quad (2.5)$$

olduğu görülür. (2.5) bağıntısında çapraz çarpım yapılırsa ve tersinir z-dönüşümü ile zaman ortamına geçilirse;

$$S(z) \left[1 - \sum_{k=1}^p a_k z^{-k} \right] = S(z) - \sum_{k=1}^p a_k S(z) z^{-k} = GU(z) \quad (2.6)$$

$$s(n) = \sum_{k=1}^p a_k s(n-k) + Gu(n) \quad (2.7)$$

elde edilir. Bu bağıntı öz-bağlanımlı (Autoregressive, AR) model olarak adlandırılır [4,5]. a_k 'ların kestirimi doğrusal öngörü analizi olarak bilinir [1-4].

İlk aşamada ses üretim sisteminin uyarıcısı $u(n)$ bilinmediğinden dolayı $u(n)=0$ kabul edilir. $s(n)$ ses sinyali örneği, geçmiş örneklerinin doğrusal kombinasyonundan yaklaşık olarak kestirilmeye çalışılır [1-5]. Ses sinyalinin kestirilen yaklaşık değeri $\hat{s}(n)$ aşağıdaki bağıntıda verilmektedir:

$$\hat{s}(n) = \sum_{k=1}^p \alpha_k s(n-k) \quad (2.8)$$

Bu bağıntı “ p ’inci dereceden doğrusal öngörücü” olarak adlandırılır [1-5]. α_k ’lar öngörü katsayılarıdır ($k=1,...,p$). $\hat{s}(n)$, $s(n)$ ’in ağırlıklandırılmış p tane geçmiş örneğinden elde edilir. p ’inci dereceden doğrusal öngörücünün sistem işlevi, $D(z)$, sonlu uzunluklu darbe cevaplı (Finite-Length Impuls Response, FIR) bir süzgeçtir:

$$D(z) = \sum_{k=1}^p \alpha_k z^{-k} \quad (2.9)$$

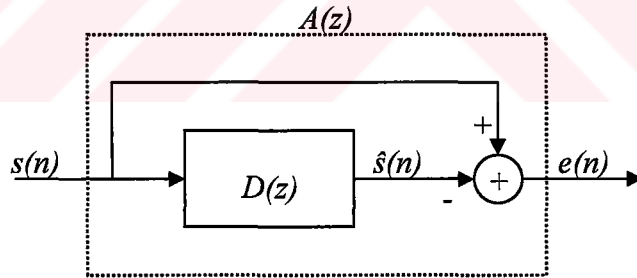
p adet geçmiş örnekten $s(n)$ sinyalinin öngörümü sırasında bir öngörüm hatası yapılır. Bu hata, $e(n)$, aşağıdaki gibi ifade edilmektedir [1-5]:

$$\begin{aligned} e(n) &= s(n) - \hat{s}(n) \\ &= s(n) - \sum_{k=1}^p \alpha_k s(n-k) \end{aligned} \quad (2.10)$$

z -dönüşümü yardımıyla öngörü hata süzgeci $A(z)$ aşağıdaki gibi ifade edilir:

$$A(z) = 1 - \sum_{k=1}^p \alpha_k z^{-k} = 1 - D(z) \quad (2.11)$$

Şekil 2.10’da öngörü hata süzgeci, doğrusal öngörü süzgeci ile birlikte gösterilmiştir.

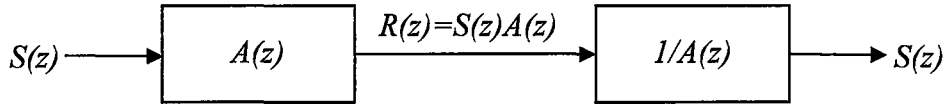


Şekil 2.10 Doğrusal öngörü hata süzgeci

Tüm-kutup bir sistemin çıkışı olan $s(n)$ ’in (2.7)’deki gibi ifade edildiği gösterilmişti. Eğer α_k öngörü parametreleri a_k sistem parametrelerine eşit olursa öngörücü hatası aşağıdaki şekilde olur:

$$\begin{aligned} e(n) &= s(n) - \sum_{k=1}^p \alpha_k s(n-k) \\ &= \sum_{k=1}^p a_k s(n-k) + Gu(n) - \sum_{k=1}^p \alpha_k s(n-k) \\ &= Gu(n) \end{aligned} \quad (2.12)$$

Görüldüğü gibi giriş sinyalinin sistem kazancı ile çarpımı, $Gu(n)$, öngörüm hatasına eşittir. Uyarıcı sinyal, orijinal sinyal ile öngörülen sinyal arasındaki fark olduğu için “kalıntı sinyali” adını alır. İlerleyen bölümlerde uyarıcı sinyal olarak kalıntı sinyalinden bahsedilecektir ve ses kalıntı sinyali “ $r(n)$ ” ile ifade edilecektir. $r(n)$, $s(n)$ ’in $A(z)$ ’ten geçirilmesiyle elde edilebilmektedir. Bu yüzden $A(z)$ süzgecine “tersinir süzgeç” adı verilir. Benzer bir şekilde, $r(n)$ ’in $1/A(z)$ süzgecinden geçirilmesiyle de $s(n)$ elde edilir. Aşağıdaki şekilde analiz/sentez modeli gösterilmektedir:



Şekil 2.11 Analiz/sentez blok çizeneği

Öngörü hata süzgeci ile (2.4) bağıntısında verilen tüm-kutup ses yolu aktarım işlevi arasındaki ilişki:

$$A(z) = \frac{1}{H(z)} \quad (2.13)$$

şeklindedir.

Bilinmeyen sistem parametreleri (2.10)’daki öngörü hatasının karesel ortalamasını en aza indirecek şekilde kestirilirler [1-3]. Öngörü hatalarının karesel ortalaması aşağıdaki şekilde ifade edilir:

$$E_n = E \left\{ e(n)^2 \right\} \quad (2.14)$$

(2.10)’daki eşitlik yukarıdaki denklemde yerine konulursa:

$$E_n = E \left\{ \left(s(n) - \sum_{k=1}^p \alpha_k s(n-k) \right)^2 \right\} \quad (2.15)$$

elde edilir. E_n ’in, α_k ’lara göre türevi alınır ve sıfıra eşitlenirse, E_n ’i en küçük yapan α_k değerlerine ulaşılır [1]-[5].

$$\frac{\partial E_n}{\partial \alpha_i} = 0, \quad 1 \leq i \leq p$$

$$2E \left\{ \left(s(n) - \sum_{k=1}^p \alpha_k s(n-k) \right) \left(-\frac{\partial}{\partial \alpha_i} \sum_{k=1}^p \alpha_k s(n-k) \right) \right\} = 0 \quad (2.16)$$

$$-\frac{\partial}{\partial \alpha_i} \sum_{k=1}^p \alpha_k s(n-k) = -s(n-i) \quad (2.17)$$

(2.17) eşitliği (2.16)'da yerine konulursa:

$$0 = 2E \left\{ \left(s(n) - \sum_{k=1}^p \alpha_k s(n-k) \right) (-s(n-i)) \right\} \quad (2.18)$$

$$E \{ s(n)s(n-i) \} = \sum_{k=1}^p \alpha_k E \{ s(n-k)s(n-i) \}, \quad 1 \leq i \leq p \quad (2.19)$$

elde edilir. Yukarıdaki denklem öz-ilişki işlevi ρ cinsinden ifade edilebilir:

$$\rho(i-k) = E \{ s(n-k)s(n-i) \}, \quad 1 \leq i \leq p, \quad 1 \leq k \leq p \quad (2.20)$$

$$\rho(i-0) = \sum_{k=1}^p \alpha_k \rho(i-k), \quad i = 1, 2, \dots, p \quad (2.21)$$

Bu denklem “normal denklemler” adını alır [1-5]. Öz-ilişki işlevinin bazı özellikleri şunlardır [1,4,5].

- $\rho(\tau)$ çift bir işlevdir ($\rho(\tau) = \rho(-\tau)$)
- $\rho(0) \geq \rho(\tau)$, (tüm τ değerleri için).
- $\rho(0)$, gerekirci (deterministic) sinyallerde sinyalinin enerjisine, rasgele veya periyodik sinyallerde sinyalin gücüne eşittir.

Normal denklemlerin çözümü için türetilen, en iyi α_k değerleri için en küçük E_n aşağıdaki şekilde hesaplanmaktadır:

$$E_n = \rho(0) - \sum_{k=1}^p \alpha_k \rho(k) \quad (2.22)$$

(2.21)'deki normal denklemler matris formunda ($\Gamma \mathbf{a} = \mathbf{p}$) gösterilebilirler:

$$\begin{bmatrix} \rho(0) & \rho(1) & \dots & \rho(p-1) \\ \rho(1) & \rho(0) & \dots & \rho(p-2) \\ \rho(2) & \rho(1) & \dots & \rho(p-3) \\ \vdots & \vdots & \ddots & \vdots \\ \rho(p-1) & \rho(p-2) & \dots & \rho(0) \end{bmatrix} \begin{bmatrix} \alpha_1 \\ \alpha_2 \\ \alpha_3 \\ \vdots \\ \alpha_p \end{bmatrix} = \begin{bmatrix} \rho(1) \\ \rho(2) \\ \rho(3) \\ \vdots \\ \rho(p) \end{bmatrix} \quad (2.23)$$

Γ öz-ilişki matrisi Toeplitz bir matristir. Ana köşegen üzerindeki elemanlar birbirine eşittir ayrıca ana köşegene paralel köşegenler üzerindeki elemanlar da birbirine eşittir. Γ matrisinin Toeplitz özellikte olması normal denklemlerin çözümünde büyük kolaylık sağlamaktadır çünkü Γ 'nın sütunları doğrusal olarak birbirinden bağımsızdır dolayısıyla Γ tersi alınabilir bir matristir [1],4,5]. Öngörü katsayıları, α_k 'lar aşağıdaki denklemle elde edilirler:

$$\alpha = \Gamma^{-1} \rho \quad (2.24)$$

Yukarıdaki matris tersi alma yöntemi p^3 tane işlem gerektirdiği için, yüksek dereceli öngörücülerde tercih edilmez. Onun yerine, p^2 tane işlem gerektiren yinelemeli Levinson-Durbin yöntemi kullanılır.

2.2.1.1 Levinson-Durbin Yineleme Algoritması

Bu yöntemdeki temel mantık, $(i+1)$ dereceli çözümü i dereceli çözümden elde etmektir. Yineleme, p 'inci dereceye kadar devam eder. Yinelemeli yöntem öz-ilişki matrisinin, köşegenleri boyunca aynı elemanlara sahip olmasından yararlanır. Algoritmanın her adımında öngörücü derecesine bağlı hata hesaplanabildiğinden, bazı durumlarda öngörücü derecesi sabit tutulmayarak, yineleme, hatanın en küçük olduğu adımda durdurulur. Algoritmanın adımları şu şekilde sıralanmaktadır:

- 0'ıncı dereceden öngörücü hatası ve öngörü katsayısı belirlenir (İlk öngörücü katsayısı sıfır sayılır):

$$E^0 = \rho(0), \quad \alpha_0^0 = 0 \quad (2.25)$$

- i 'inci kutup modeli için bir ağırlık faktörü belirlenir:

$$k_i = \frac{\left(\rho(i) - \sum_{j=1}^{i-1} \alpha_j^{i-1} \rho(i-j) \right)}{E^{i-1}} \quad (2.26)$$

- Elde edilen ağırlık faktörleri kullanılarak i . kutup modeli için öngörü katsayıları güncellenir:

$$\begin{aligned} \alpha_i^i &= k_i \\ \alpha_j^i &= \alpha_j^{i-1} - k_i \alpha_{i-j}^{i-1}, \quad 1 \leq j \leq i-1 \end{aligned} \quad (2.27)$$

Burada en yüksek mertebeli katsayı k_i 'ye eşitlenmiştir. k_i değerleri “kısmi ilişki katsayıları” (partial correlation coefficients, PARCOR) olarak bilinmektedirler.

- $(i-1)$ 'inci derece öngörü hatası ve k_i kullanılarak, i 'inci derece için öngörücü hatası güncellenir:

$$E^i = (1 - k_i^2) E^{i-1} \quad (2.28)$$

- Adımlar $i=1,2,\dots,p$ için tekrarlanır. p 'inci adımda p 'inci dereceden öngörü katsayısına ulaşılır (“*” işareti optimum öngörücüyü ifade etmektedir):

$$\alpha_j^* = \alpha_j^p, \quad 1 \leq j \leq p \quad (2.29)$$

Levinson-Durbin algoritmasıyla çözülen öz-ilişki yönteminin özellikleri şu şekilde sıralanmaktadır:

i-) Kısmi ilişki katsayılarının genlikleri 1'den küçüktür ($|k_i| < 1$) [5]. (2.28) bağıntısındaki karesel ortalama öngörü hatası daima 1'den büyüktür. Çünkü hiç hata yapılmama durumu mümkün değildir. O halde:

$$E^i = (1 - k_i^2) E^{i-1} > 0 \Rightarrow (1 - k_i^2) = \frac{E^i}{E^{i-1}} > 0 \quad (2.30)$$

olur. Yukarıdaki eşitlik düzenlenirse $|k_i| < 1$ olduğu görülür:

$$1 - k_i^2 > 0 \Rightarrow k_i^2 < 1 \Rightarrow |k_i| < 1 \quad (2.31)$$

ii-) Öngörücü polinomunun kökleri daima birim dairenin içindedir. Yani $H(z)$ aktarım işlevinin kutupları birim dairenin içindedir ($H(z)$ minimum fazlı bir işlevdir) [2].

iii-) Öngörü katsayıları dizisi, $[\alpha_1 \alpha_2 \alpha_3, \dots, \alpha_p, G]$, ile öz-ilişki değerleri dizisi, $[\rho_n(0), \rho_n(1), \dots, \rho_n(p)]$ birbirlerinden türetilebilirler. Aynı ilişki kısmi ilişki katsayıları ile öngörü katsayıları arasında da vardır.

- $s(n)$ sinyalinin değerleri kullanılarak, öz-ilişki yöntemiyle kestirilmiş, p . dereceden bir aktarım işlevi $H(z)$ 'nin birim darbe cevabı $h(n)$ olsun. $h(n)$ 'in öz-ilişki işlevi $\rho_h(\tau)$ ile ifade edilir. Birim darbe cevabı $h(n)$ 'in enerjisinin $s(n)$ 'in enerjisine eşit olduğu varsayılırsa, $|\tau| < p$ için $\rho_h(\tau) = \rho(\tau)$ eşitliği yazılabilir. Bu özellik öz-ilişki eşleme özelliği olarak bilinmektedir [1,5].

Ses kalıntı sinyali $1/f$ özellik gösteren rasgele öz-benzeşimli yapıya sahip bir sinyaldir. Bu yapının incelenip modellenmesinden önce “öz-benzeşimli süreç” ve “ $1/f$ süreç” kavramlarının anlaşılması gerekmektedir. Bu kavramlar bir sonraki bölümde açıklanmaktadır.

3. 1/f SÜREÇLER

Bu bölümde 1/f süreçler ve bu süreçlerin matematiksel modelleri incelenmektedir.

3.1 1/f Süreçler ve Öz-benzeşimlilik

“1/f özellik gösteren süreçler” diğer adıyla “güç yasası ilişkisi süreçleri” ölçekleme ile değişmeyen (scale-invariant) öz-benzeşimli süreçler olup, doğal veya suni birçok olguda karşımıza çıkmaktadır. Örneğin süper iletkenlerdeki ısı ve elektromanyetik titreşimler, kalp atış hızı değişkenliği, trafik akışı değişkenliği, borsa indeksi [6], konuşma sesi kalıntı sinyali [7,8], sualtı gürültüsü [9], lazer ışığı yayılımı [10] ve DNA çiftlerinin dizilimi [11,12] gibi birbirinden farklı birçok sinyallerde bu özellik gözlenmektedir.

Eğer bir $x(t)$ rasgele süreci istatistiksel olarak öz-benzeşimli ise onun zaman ortamındaki istatistiksel özellikleri öteleme ile değişirken, ölçekleme (daraltma-genişletme) ile değişmez. Öz-benzeşimli süreçler aşağıdaki ölçekleme yapısına sahiptirler [6]:

$$x(t) \stackrel{p}{=} c^{-H} x(ct) \quad (3.1)$$

Burada “ p ” istatistiksel davranışı simgelemektedir. c ise herhangi bir sabit pozitif gerçel sayıdır. H Hurst parametresi olup $0 < H < 1$ aralığında değerler almaktadır. Geniş anlamda (wide sense) öz-benzeşimli süreçler için (3.1) deki eşitlik 2. dereceden istatistikler olan, ortalama ve kovaryans için de genişletilebilir [6]:

$$\begin{aligned} \mu_x(t) &\stackrel{\Delta}{=} E[x(t)] \\ &= c^{-H} \mu_x(ct) \end{aligned} \quad (3.2)$$

$$\begin{aligned} \Phi_x(t_1, t_2) &\stackrel{\Delta}{=} E[x(t_1)x(t_2)] \\ &= c^{-2H} \Phi_x(ct_1, ct_2) \end{aligned} \quad (3.3)$$

Burada, $\mu_x(t)$, $x(t)$ sürecinin ortalama işlevi, $\Phi_x(t)$ ise $x(t)$ sürecinin kovaryans işlevidir. Sinyal istatistiksel olarak öz-benzeşimli ise, H kestirilebilir. H parametresi kestirime, konuşma sesi işlemeden bilgisayar ağlarına kadar birçok alanda karşılaşılmaktadır [6-13].

$1/f$ süreçlerin güç izge yoğunlukları, birkaç onkat (decade) boyunca $|\omega|^{-\gamma}$ ile orantılıdır. Genel olarak ifade edilecek olursa; güç izgesi aşağıdaki şekilde olan rasgele süreçler “ $1/f$ gürültü” olarak adlandırılırlar [6]:

$$S_x(\omega) \approx \frac{\sigma_x^2}{|\omega|^\gamma} \quad (3.4)$$

Burada $S_x(\omega)$, $x(t)$ sürecinin güç izgesi, ω açısal frekans, γ sabit gerçel bir sayı olan izgesel üstel (öz-benzeşim parametresi) ve σ_x^2 sinyalin değışintisidir. (3.4)’teki eşitlik “Güç Yasası İlişkisi” olarak bilinmektedir [6,13]. Bu alanlarda, gözlenen bir örnekler topluluğundan γ parametresi kestirimi, öngörü, kontrol ve teşhis (tanı) gibi birçok farklı amaç için kullanılmaktadır.

Genel olarak, güç yasası birçok onkat için genişletilebilir. Veri uzunluğu, alçak frekans bölgelerindeki izgesel bilgilere ulaşmayı sınırlı hale getirirken, veri çözünürlüğü, yüksek frekans bölgelerindeki izgesel içeriğe ulaşmayı sınırlandırır [6]. Öz-benzeşimliliğin doğal sonucu olarak, $1/f$ süreçler fraktallı yapıya sahiptirler ve etkin bir fraktal boyutu tanımı yapılabilir. Fraktal boyutu “ D ” ile ifade edilir ve sinyalin pürüzlülüğünün bir ölçütüdür. Herhangi bir işlevin fraktal boyutu, $1 \leq D \leq 2$ aralığında değerler almaktadır [6]. D ’nin büyük değerlerine sahip işaretlerin görünüşünde fark edilir bir biçimde pürüzlülük mevcuttur. $1/f$ süreçlerde D ile H arasında aşağıdaki ilişki mevcuttur [6]:

$$D = 2 - H \quad (3.5)$$

H ’nin değerindeki bir artma D ’nin değerindeki bir azalmaya karşılık gelmektedir.

3.2 $1/f$ Süreçlerin Modellenmesi

“Kesirli Brownian Devinimleri (Fractional Brownian Motions, fBm)” sıfır ortalamalı, Gauss dağılımlı, durağan olmayan, sürekli, rasgele süreçlerdir [14]. fBm, öz-benzeşimli süreçlerin modellenmesinde yaygın olarak kullanılan etkin bir modeldir [6,14-16]. Güç yasası ilişkisindeki öz-benzeşim parametresi, $1 < \gamma < 3$

aralığında değerler alır. γ 'nın yaklaşık "2" olduğu özel durum süreçleri ise "Klasik Brownian Devinimi" olarak bilinir [14]. fBm'ler, sınırlı zaman aralığında sınırlı güce sahip, durağan olmayan rasgele süreçlerdir. Bu süreçlerde düşük frekans bölgelerindeki güç daha fazladır ve aşağıdaki özellikleri sağlarlar [6,14-16]. fBm süreç $B_H(t)$ ile gösterilecek olursa ($0 \leq t < \infty$, $0 < H < 1$):

Özellik I. $B_H(t)$ 'nin artımları durağandır.

Özellik II. $B_H(0) = 0$, $E\{B_H(t)\} = 0$, $t \geq 0$

Özellik III. $B_H(t + t_0) - B_H(t) \stackrel{d}{=} \mathcal{N}(0, \sigma|t_0|^H)$

Burada, " $\stackrel{d}{=}$ " sonlu boyutta eşitliği ifade eder. σ sabit bir gerçel sayı, t_0 zaman artımı, $\mathcal{N}(0, \sigma|t_0|^H)$ ise "0" ortalamalı, " $\sigma|t_0|^H$ " değişintili Gauss (normal) dağılımını ifade etmektedir. fBm'lerin kovaryans işlevlerinin özellikleri ise ($t_1, t_2 > 0$) :

Özellik IV. $E\{B_H(t_1)B_H(t_2)\} = \frac{\sigma^2}{2} (|t_1|^{2H} + |t_2|^{2H} - |t_1 - t_2|^{2H})$

Özellik V. $E\{B_H^2(t)\} = \sigma^2 |t|^{2H}$

Özellik VI. $B_H(ct) = c^{*H} B_H(t)$, $c > 0$

Özellik VII. $E\{B_H(ct_1)B_H(ct_2)\} = E\{(c^H B_H(t_1))(c^H B_H(t_2))\}$

şeklinde sıralanabilir[6,14-16]. Burada $E\{.\}$ beklenen değer ifadesidir. σ ve c sabit pozitif gerçel sayılardır. $B_H(t)$ 'nin Gauss dağılımlı olması ve yukarıdaki özellikleri sağlaması onun istatistiksel olarak öz-benzeşimli olduğunun bir göstergesidir [15]. *Özellik II.* ve *Özellik III.* rasgele yürüyüş (random walk) süreçlerinin de bir ifadesidir. Dolayısıyla fBm'leri rasgele yürüyüş süreçleri olarak adlandırmak mümkündür. Bu süreçlerin üretilmesinde "0" noktasından başlanır ve her adımda bir önceki değere Gauss dağılımlı rasgele bir değer eklenerek bir sonraki adıma geçilir.

$1/f$ süreçlerin bir diğer modeli "Kesirli Gauss Gürültüsüdür (Fractional Gaussian Noise, fGn)". fGn süreci $G_{H,\Delta}(t)$ ($0 \leq t < \infty$, $0 < H < 1$, Δ küçük, sabit, pozitif gerçel değer) ile ifade edilecek olursa, fBm'lerin türevleri olduğu ve aşağıdaki özellikleri sağladığı gösterilebilir [6, 14-16]:

Özellik I. $G_{H,\Delta}(t) = \frac{1}{\Delta} (B_H(t + \Delta) - B_H(t))$

$$\text{Özellik II. } G_{H,\Delta}(t) \stackrel{d}{=} \mathcal{N}(0, \sigma|\Delta|^{H-1})$$

Özellik II.'de görüldüğü gibi fGn'ler sıfır ortalamalı, durağan Gauss süreçleridir. Yeterli büyüklükteki aralık (lag, τ) değerleri için öz-ilişki işlevleri aşağıdaki gibi ifade edilir [6]:

$$\begin{aligned} \text{Özellik III. } E\{G_{H,\Delta}(t+\tau)G_{H,\Delta}(t)\} &= \frac{\sigma^2}{2} [(\tau+1)^{2H} - 2\tau^{2H} + (\tau-1)^{2H}] \\ &\underset{\tau \rightarrow \infty}{\approx} \sigma^2 H(2H-1)\tau^{2H-2} \end{aligned}$$

fBm'ler durağan olmayan süreçler olmalarına rağmen artımları olan fGn'ler durağandır. fGn'ler, güç yasası ilişkisindeki öz-benzeşim parametresinin, $-1 < \gamma < 1$ aralığında değerler aldığı, yüksek-frekans bölgelerindeki gücün yüksek olduğu süreçlerdir [6]. γ 'nın sıfıra eşit olduğu özel durum, durağan beyaz Gauss gürültüsüdür.

Özellik III.'ten görüleceği gibi, $0 < H < 1/2$ olduğu durumda fBm, negatif ilişkili artımlara sahip bir süreçtir. Belirli eğilimde devam eden bir hareket yoktur. Süreçteki bir yükselmeden sonra muhtemelen bir alçalma gelir. Bu süreçler “Kısa menzilli bağımlı (Short Range Dependent) süreçler” olarak bilinmektedirler [6]. $1/2 < H < 1$ olduğu durumda, fBm süreçlerde, belirli bir eğilimde devam eden sürekli bir hareket vardır. Süreçteki bir yükselmeden sonra muhtemelen bir diğer yükselme gelir. fBm'lerin öz-ilişki işlevleri incelendiğinde $H > 1/2$ için çok yavaş bir düşüşe sahip oldukları görülür. H değeri büyüdükçe süreç daha az düzensizlik gösterir. H 'nin, $1/2 < H < 1$ aralığında olduğu süreçler “Uzun menzilli bağımlı (Long range dependent)” süreçler olarak bilinmektedirler [6].

γ parametresi ile H parametresi arasındaki ilişki fBm için:

$$\gamma = 2H + 1 \quad (3.6)$$

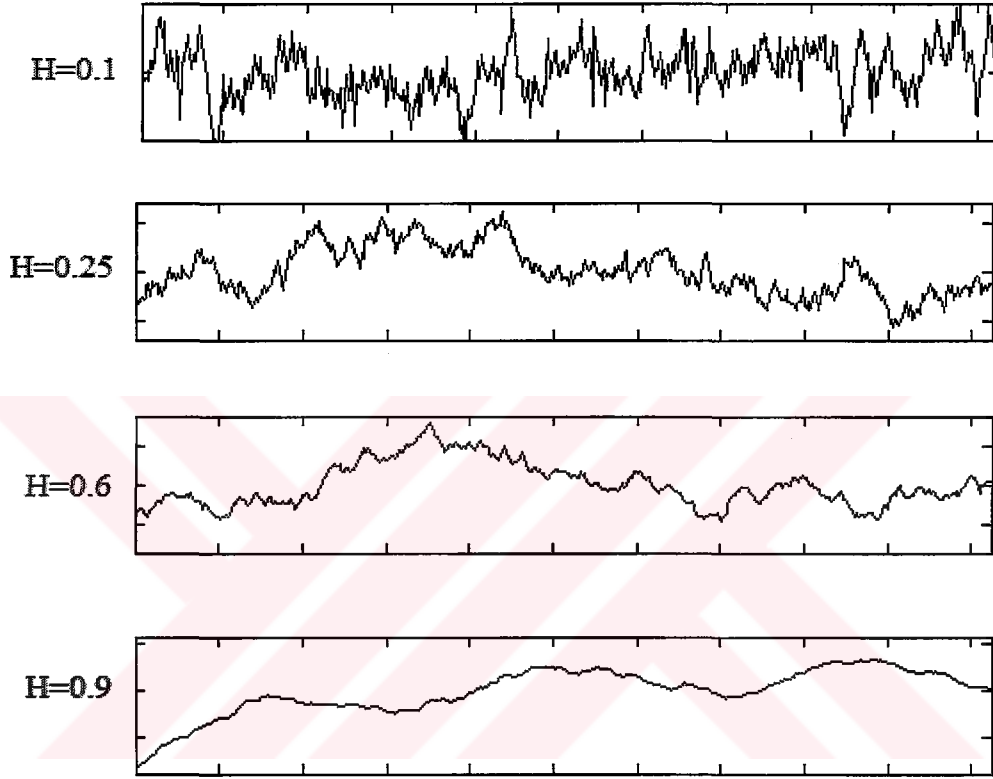
fGn için:

$$\gamma = 2H - 1 \quad (3.7)$$

olur [6, 14].

Daha önce de belirtildiği gibi, $1/f$ süreçlerde fraktal boyutu D ile H arasında $D=2-H$ ilişkisi mevcuttur ($0 < H < 1$). H değerindeki bir artma D değerindeki bir azalmaya

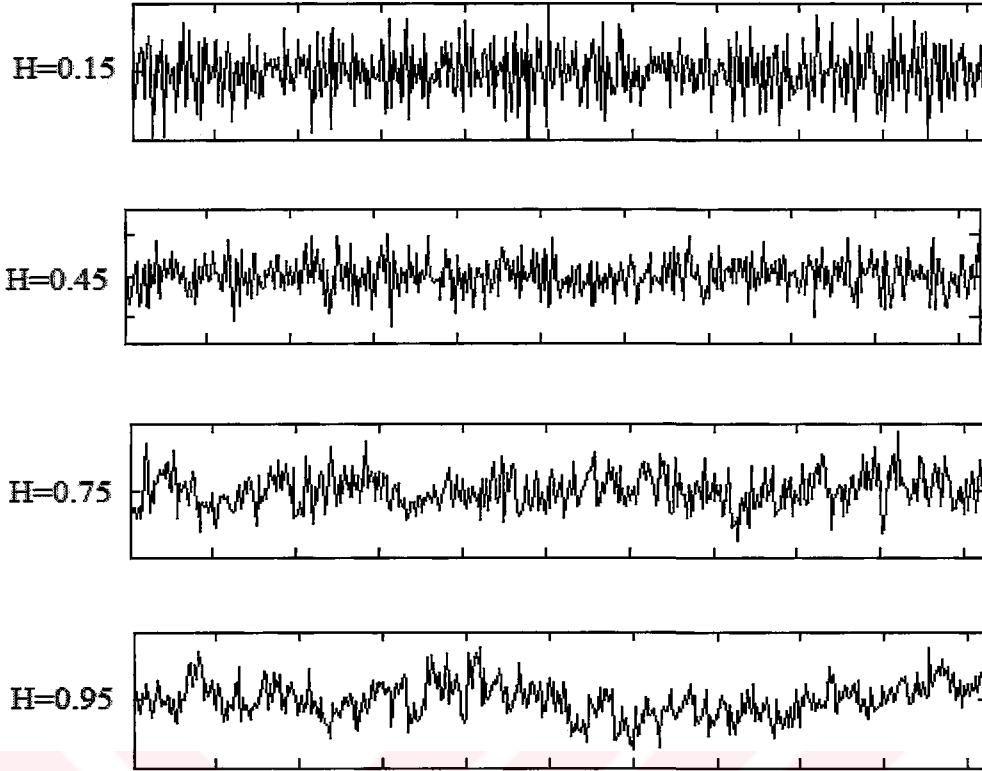
karşılık gelmektedir. Bunun gayet doğal olduğu açıktır. H artarsa γ da artar. γ 'nın artması, gücün yüksek frekans bölgelerinden alçak frekans bölgelerine doğru artması anlamına gelir ve örnek salınımlarının (sample fluctuation) görünürde artan bir yumuşaklık göstermesini sağlar [6]. Şekil 3.1'de farklı H değerlerine sahip dört fBm sürecin zaman-genlik değişimi verilmiştir. Şekil 3.2'de ise farklı H değerlerine sahip dört fGn sürecin zaman-genlik değişimi gösterilmektedir.



Şekil 3.1 Farklı H değerlerine sahip fBm süreçler

$1/f$ tipi süreçlerin daha genel olarak gözlenen bir başka şekli mevcuttur ki bu süreçler yaklaşık olarak $1/f$ izgesine sahip olan “yaklaşık $1/f$ süreçler”dir. Bu süreçlerde güç yasası ilişkisi sadece belli frekans bölgelerinde geçerliliğini korur. Yaklaşık $1/f$ süreçler analizi kolay süreçler olup $1/f$ süreçlerin çoğu özelliğine sahiptirler. σ_A^2 ve σ_U^2 birbirine yakın sabit değerler olmak üzere, bu süreçlerin güç izgeleri aşağıdaki gibi sınırlandırılarak ifade edilir [6]:

$$\frac{\sigma_A^2}{|\omega|^\gamma} \leq S_x(\omega) \leq \frac{\sigma_U^2}{|\omega|^\gamma} \quad (3.8)$$



Şekil 3.2 Farklı H değerlerine sahip fGn süreçler

3.3 Ses Kalıntı Sinyalinin $1/f$ Özelliği

$1/f$ süreçlerin güç izgeleri, $|\omega|^{-\gamma}$, izgesel düzlemde -3γ dB/oktav'lık bir düşüşe eşdeğerdir. Aşağıda görüldüğü gibi her bir oktavda 3γ dB'lik düşüş meydana gelir:

$$\begin{aligned}
 -10\gamma \log_{10} 2 &\approx -3\gamma \text{ dB / oktav} \\
 -10\gamma \log_{10} 4 &\approx -6\gamma \text{ dB / oktav} \\
 -10\gamma \log_{10} 8 &\approx -9\gamma \text{ dB / oktav} \\
 -10\gamma \log_{10} 16 &\approx -12\gamma \text{ dB / oktav} \\
 \vdots &\quad \quad \quad \vdots
 \end{aligned} \tag{3.9}$$

Konuşma sesi kalıntı sinyali öz-benzeşimli bir sinyal olup, güç izgesi güç yasası ilişkisini sağlamaktadır. Ses kalıntı sinyalinin $1/f$ özelliği ilk olarak Koser tarafından incelenmiştir [7]. Konuşma sesi kalıntı sinyalleri, frekans bölgesinde yaklaşık -6 dB/oktav düşüş ile tanımlanırlar (-12 dB/oktav boğaz bölgesi titreşimini, 6 dB/oktav dudak yayılımını tanımlar) [2,17]. -6 dB/oktav'lık azalma izge ortamında $1/\omega^2$ yapısına eşdeğerdir. Bölüm 3.2'de belirtildiği gibi, yaklaşık olarak $1/\omega^2$ yapısı gösteren ($\gamma \approx 2$) sinyaller klasik fBm olarak nitelendiğinden [14] incelediğimiz konuşma sesi kalıntı sinyalleri de fBm tipi sinyaldir. Konuşma sesinin doğallığı

kalıntı sinyalinin $1/f$ yapısında saklıdır [17]. Bu yapı tek bir parametre (γ) ile modellenebilir ve kodlama ve ses kalitesi artırımında kullanılabilir [8]. Kalıntı sinyalinin $1/f$ yapısının konuşma sesi kodlamasında ve ses kalitesi artırımında nasıl kullanılabileceği Bölüm 5'te anlatılmaktadır.

Bir sonraki bölüm olan “Yöntem” bölümünde ses kalıntı sinyali elde ediliş aşamaları anlatılmakta ve ses kalıntı sinyalinin $1/f$ yapısı dalgacık tabanlı analiz ile incelenmektedir.

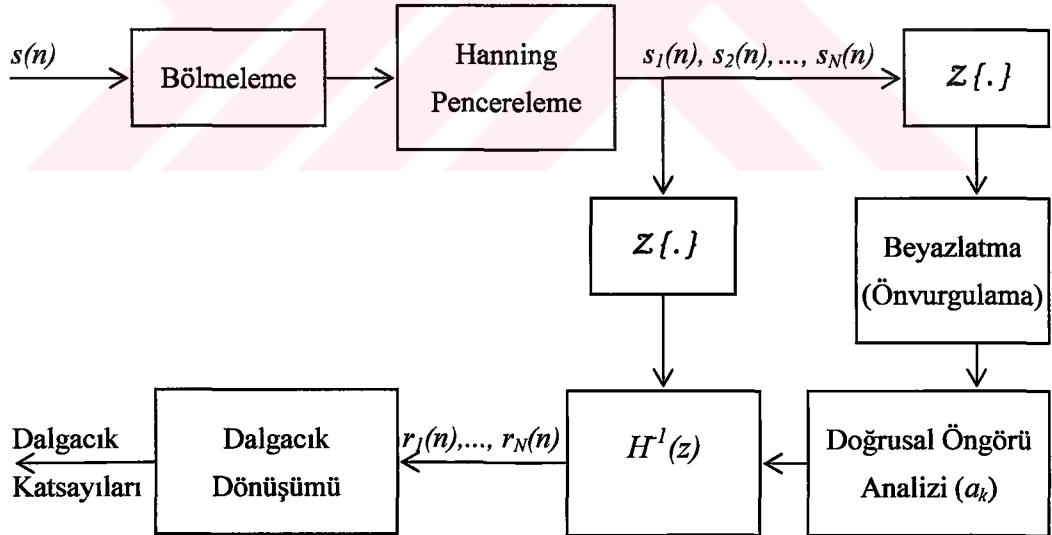


4. YÖNTEM

Bu bölümünde, ses kalıntı sinyalinin elde edilme aşamaları ve kalıntı sinyalinin öz-benzeşim parametresinin kestirim işlemleri anlatılmaktadır.

Ötümlü seslerin üretimi sırasında ses yolunu uyaran ses tellerindeki sözde periyodik titreşim sinyali “Ses Kalıntı Sinyali” olarak adlandırılmaktadır. Bölüm 2’de belirtildiği gibi ünlü sesler en belirgin ötümlü sesbirimi grubunu oluştururlar. Bu sebeple incelemede ünlü sesler ele alınmış ve kalıntı sinyalinin $1/f$ özelliği incelenmiştir.

İlk olarak, elimizdeki sözcüklerin ötümlü (dolayısıyla ünlü) kısımları ayrılmıştır. Ötümlü kısımların kalıntı sinyalleri elde edilmiş ve bu sinyaller üzerine dalgacık tabanlı öz-benzeşim parametresi kestirim yöntemi uygulanarak öz-benzeşim parametreleri kestirilmiştir. Yöntemin blok çizeneği Şekil 4.1’de gösterilmektedir.



Şekil 4.1 Ses kalıntı sinyali çıkarımı blok çizeneği

Kalıntı sinyalini elde etmede ilkin, konuşma sesi sinyali durağan kabul edilebileceği küçük parçacıklara ayrılır. Bu parçacıklar Hanning pencereden geçirilerek pencerelenmiş ses parçacıkları elde edilir. Daha sonra, konuşma sesi sinyallerinin güç izgeleri beyazlatılarak, beyazlatılmış izge üzerine doğrusal öngörü analizi uygulanır. Doğrusal öngörü analizi ile ses yolu parametreleri elde edilir ve bu

parametreler yardımıyla tersinir süzgeç oluşturur. Hanning pencerelemiş ses parçacıkları tersinir süzgeçten geçirilerek süzgecin uyarıcısı olan kalıntı sinyaline ulaşılır. Son olarak kalıntı sinyalinin dalgacık tabanlı analiz ile öz-benzeşim parametresi kestirilir.

4.1 Ötümlü/Ötümsüz/Sessizlik Ayrımı

Çalışmamızda ötümlü seslerin kalıntı sinyallerinin elde edilmesi ve $1/f$ özelliğinin incelenmesi söz konusu olduğundan, konuşmaların ötümlü kısımlarını ayrıştırarak, sinyalden soyutlamak gerekmektedir. Ötümlü/Ötümsüz ayrımı için birçok kıstas mevcuttur. Bu kıstaslardan bazıları şunlardır [1-3]:

- Kısa Zamanlı Enerji
- Kısa Zamanlı Ortalama Genlik
- Ortalama Sıfır Seviyesi Geçiş Sayısı
- Birim Aralıklı Normalize Özilişki katsayısı
- Perde Periyodu

Bu kıstaslar aşağıda anlatılmaktadır.

4.1.1 Kısa Zamanlı Enerji

$s_n(n)$ sinyali, $[0, N-1]$ aralığında “1” değeri alan diğer yerlerde “0” olan bir dikdörtgen pencere ile çarpılmış $s(n)$ sinyalinin N uzunluğunda bir parçası olmak üzere, $s_n(n)$ ’in kısa zaman enerjisi (dB cinsinden) aşağıdaki gibi ifade edilir [2]:

$$E_s = 10 \log(\epsilon + \sum_{n=0}^{N-1} |s_n(n)|^2) \quad (4.1)$$

Burada “ ϵ ”, logaritma ifadesinin içerisinde sıfır olmasını engellemek üzere kullanılan küçük, pozitif, sabit bir gerçel sayıdır. Konuşma sesi sinyallerinin genlikleri zamana bağlı olarak değişmektedir. Ses sinyalinin kısa zaman enerjisi, bu değişimi gözlemlemede iyi bir yöntemdir. Ötümlü seslerin kısa zaman enerjileri sessizliğe göre çok yüksektir. Ötümsüz seslerin kısa zaman enerjileri ise ötümlü seslerin enerjilerinden düşük, sessizliğin enerjisinden büyüktür [1-3].

4.1.2 Kısa Zamanlı Ortalama Genlik

Ötümlü/Ötümsüz ayrımında kısa zaman enerjisi etkin bir yöntem olmasına karşın, yüksek genlikli ses sinyallerine karşı duyarlı olması olumsuz bir özelliktir. Bu sorunu ortadan kaldırmak için “Kısa zamanlı ortalama genlik” kullanılabilir. N boyunda bir $s_n(n)$ sinyalinin kısa zamanlı ortalama genliği, M_s , aşağıdaki gibi ifade edilir [2]:

$$M_s = \frac{1}{N} \sum_{n=0}^{N-1} |s_n(n)| \quad (4.2)$$

Burada, sinyalin mutlak değer karelerinin ortalamaları yerine mutlak değer ortalamaları alınmaktadır.

4.1.3 Ortalama Sıfır Seviyesi Geçiş Sayısı

Ayrık sinyallerde ardı ardına gelen örnekler farklı işaretlere sahiplerse “Sıfır Geçiş” meydana gelir. Ortalama sıfır seviyesi geçiş sayısı konuşma sesi analizinde yararlı bir özelliktir. $s_n(n)$ sinyalinin ortalama sıfır seviye geçiş sayısı, Z_s , aşağıdaki gibi hesaplanır [2]:

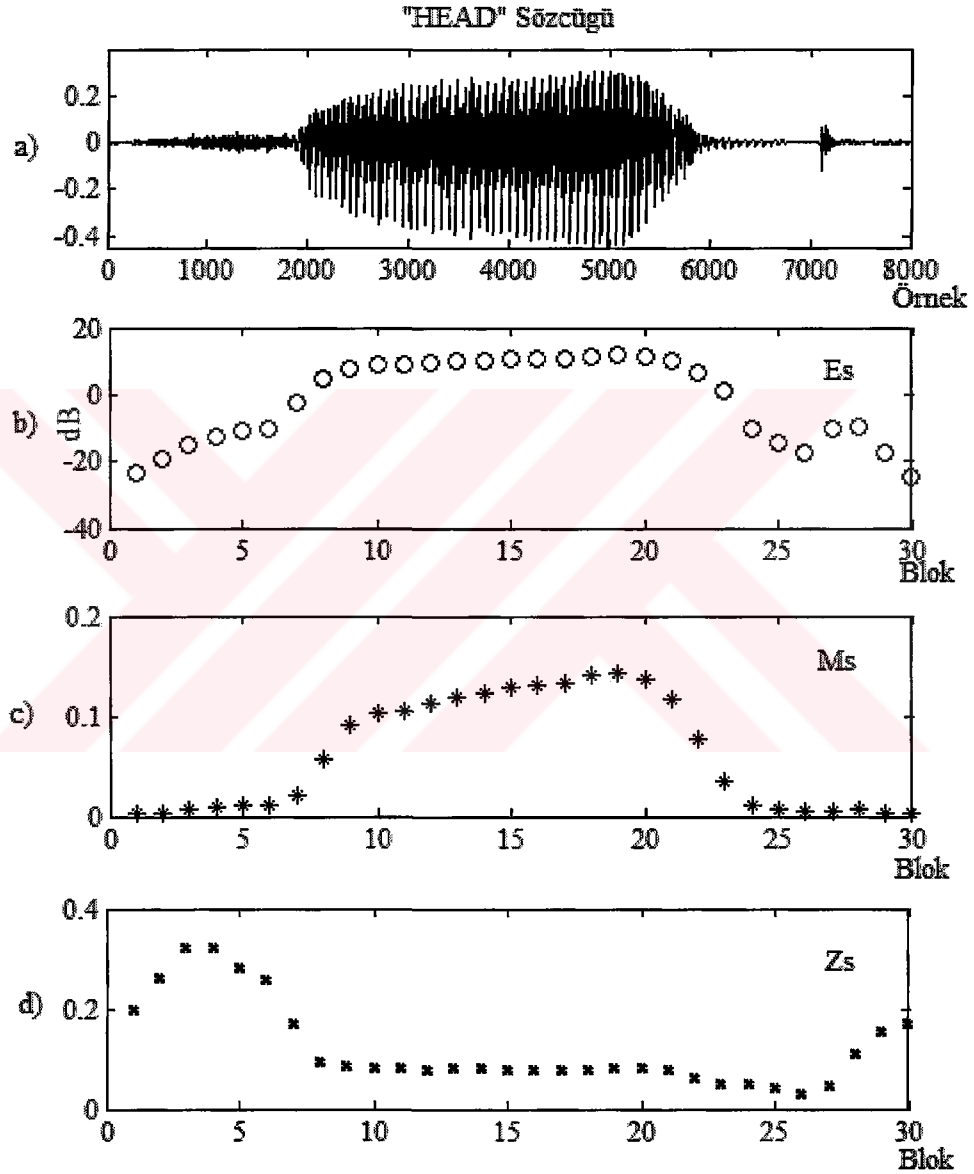
$$Z_s = \frac{1}{N} \sum_{n=0}^{N-1} \frac{|sgn\{s_n(n+1)\} - sgn\{s_n(n)\}|}{2} \quad (4.3)$$

$$sgn\{s_n(n)\} = \begin{cases} +1, & s_n(n) \geq 0 \\ -1, & s_n(n) < 0 \end{cases}$$

Ötümsüz seslerin güç izgeleri incelendiğinde, yüksek frekans bölgelerindeki genliklerin daha büyük olduğu görülür. Ötümsüz ses sinyallerinin yüksek frekanslı bileşenleri daha fazladır. Yüksek frekans bileşenleri fazla olan sinyallerde sıfır seviyesi geçiş sayısı da yüksek olur. Dolayısıyla ötümsüz seslerin kısa zaman ortalama sıfır geçiş sayısı ötümlü seslere göre daha yüksektir. Teorik olarak konuşma sinyalinin “sessizlik” kısmında sıfır geçiş sayısı sıfır olur ancak ses sinyalindeki gürültü nedeniyle Z_s hiçbir zaman sıfır olmaz, ötümlü kısmın sayısına yakın bir değerde olur.

Şekil 4.2.a’da 16kHz ile örneklenmiş “HEAD” sözcüğünün zaman-genlik değişimi gösterilmiştir. Bu sözcük birbiri üzerine 256 örnek (16 ms) ile örtüşen 512 örnekli (32 ms) 30 parçaya ayrılmıştır. Şekil 4.2.b’de herbir parçanın kısa zamanlı enerjisi gösterilmektedir. Görüldüğü gibi ötümsüz /H/ sesine karşılık gelen kısımlardaki

enerjiler sessiz blokların enerjilerinden daha yüksek iken, /EH/ sesinin bulunduğu ötümlü ses bloklarının enerjilerinden daha düşüktür. Ötümlü-patlamalı bir ses olan /D/ sesinin bulunduğu blokların enerjileri, ötümlü kısımdan daha düşüktür ancak sessiz bölümlerden ve ötümsüz bölümlerden daha yüksektir. Şekil 4.2.c’de her bir ses parçasının kısa zamanlı ortalama genlik değerleri bulunmaktadır. Şekil 4.2.d’de ise her bir konuşma sesi parçasığının ortalama sıfır geçiş sayısı gösterilmektedir.

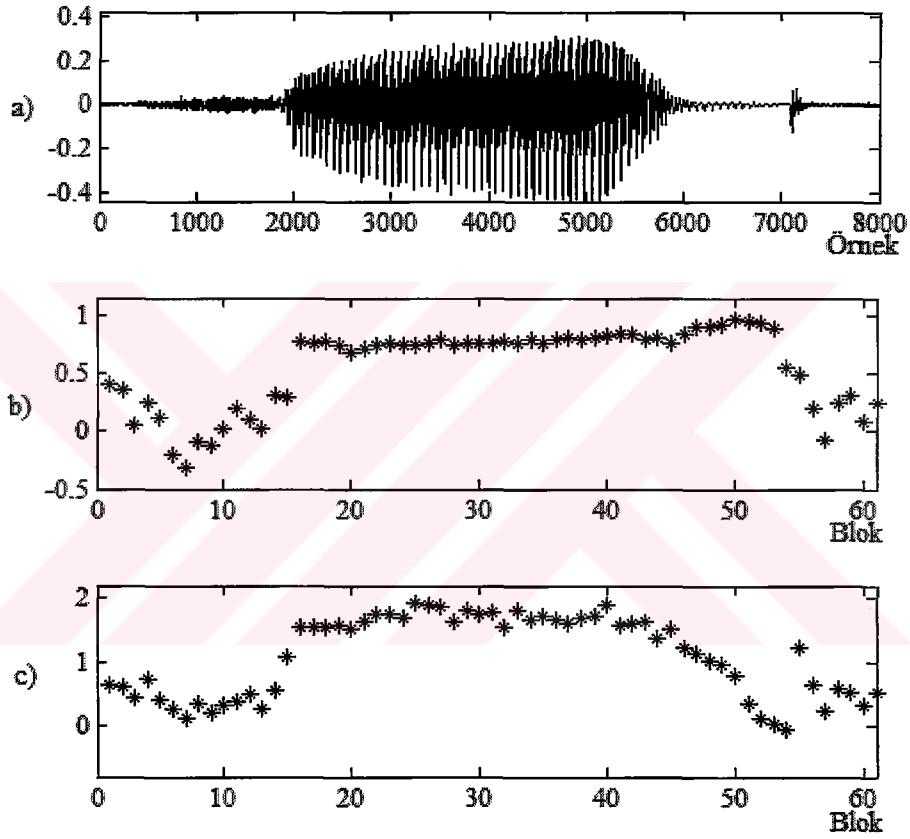


Şekil 4.2 (a) "Head" sözcüğünün zaman-genlik değişimi, (b) Her bir 512 örneklilik bloğun enerjisi (dB), (c) Her bir 512 örneklilik bloğun ortalama genlik değeri, (d) Herbir 512 örneklilik bloğun ortalama sıfır geçiş sayısı

4.1.4 Birim Aralıklı Normalize Özilişki katsayısı

Birim aralıklı normalize özilişki katsayısı, C_l , ses sinyalindeki birbirine komşu örneklerin ilişkisini gösterir. $[0, N-1]$ aralığında tanımlı, kısa zamanlı $s_n(n)$ konuşma sesi sinyali için, C_l , aşağıdaki şekilde ifade edilmektedir [2]:

$$C_l = \frac{\sum_{n=l}^N s_n(n) s_n(n-1)}{\sqrt{\sum_{n=l}^N s_n^2(n) \sum_{n=0}^{N-1} s_n^2(n)}} \quad (4.4)$$



Şekil 4.3 (a) “Head” sözcüğünün zaman-genlik değişimi, (b) Her bir 256 örneklik bloğun birim aralıklı normalize ilişkisi katsayısı (C_l), (c) Her bir 256 örneklik bloğun ilk öngörü katsayısı (α_l)

Ötümlü seslerde enerji alçak frekans bölgelerinde yoğunlaşmaktadır. Yani çok hızlı (yüksek frekanslı) değişimler görülmemektedir. Dolayısıyla, ötümlü bölgelerdeki ses örnekleri komşu örneklerle sıkı bir ilişkiye sahiptirler. Ötümlü ses parçası için C_l yaklaşık “1” olurken, ötümsüzlerde bu değer “1”den çok daha küçüktür. Ötümlü/ötümsüz ayrımında, ilk doğrusal öngörü katsayısı α_l de kullanılabilir. Ötümlü seslerin güç izgeleri ile ötümsüz seslerin güç izgeleri birbirinden çok farklıdır, doğal olarak doğrusal öngörü katsayıları da birbirinden farklı olacaktır.

Şekil 4.3.a'da "HEAD" sözcüğünün genlik-zaman değişimi verilmiştir. Şekil 4.3.b'de birbiri üzerine 128 örnek (8 ms) ile örtüşen 256 örnekli (16 ms) konuşma sesi bloklarında hesaplanan C_l 'in değişimi gösterilmiştir. Aynı bloklarda hesaplanan α_l 'in değişimi ise Şekil 4.3.c'de gösterilmiştir. α_l hesaplanırken, ses blokları Hanning pencerelenmiş ve öngörücü mertebesi 12 seçilmiştir. Ayrıca yüksek frekans bölgelerindeki düşük genlikli tepeciklerin doğrusal öngörüsünün daha iyi yapılabilmesi için sabit ön-vurgulama süzgeci kullanılmıştır. (Ön-vurgulama Bölüm 4.2.2'te anlatılmaktadır.)

4.1.5 Perde Periyodu Kestirimi

Ötümlü seslerin üretimi sırasında ses tellerinde periyodik bir titreşim mevcuttur. Bu titreşimin ses yolunu uyarması sonucunda oluşan ötümlü sesler periyodik bir yapıya sahip olurlar. Ötümsüz seslerin üretiminde ise uyarıcı sinyal gürültü olduğu için, bu sesler, gürültü benzeri, dolayısıyla periyodik olmayan bir yapıda olurlar. Bu fark, ötümlü/ötümsüz sesler arasındaki en önemli farklardan biridir [1-3]. Konuşma sesi parçalarında perde periyodu kestirimi yapılarak periyodik bir sinyal olup olmadığı kontrol edilebilir ve ötümlü/ötümsüz kararı verilebilir.

Perde periyodu kestirimi için bir çok yöntem önerilmiştir [1-3]. Bu bölümde, gerçel kepstrum yöntemi, kısa zamanlı özilişki yöntemi ve kısa zamanlı genlik farkı yöntemi özetlenmektedir.

4.1.5.1 Gerçel Kepstrum Yöntemi

Kepstrum yöntemi, perde periyodu kestiriminde yaygın olarak kullanılan bir yöntemdir [1-3]. Yapılan işlem özetle, ses sinyalinden, perde periyodu bilgisini barındıran uyarıcı sinyali ayırmak ve perde periyodunu kestirmektir. Bölüm 2'de belirtildiği gibi, konuşma sesi sinyali, uyarıcı sinyal $r(n)$ ile ses yolu darbe cevabı $h(n)$ 'in evrişimi şeklindedir:

$$s(n) = r(n) * h(n) \quad (4.5)$$

Yukarıdaki evrişim frekans alanında çarpıma denk düşmektedir. Frekans alanında çarpılmış sinyalleri, doğrusal bir işlemle birbirinden ayırmak mümkün değildir. Ses yolu etkisini ve uyarıcı sinyali birbirinden ayırmak için, bir doğrusal işlem oluşturan kepstrum kullanılmaktadır. Kepstrum, bir sinyalin Fourier dönüşümünün logaritmasının ters Fourier dönüşümüdür. $s_n(n)$, $w(n)$ pencere işlevi ile pencerelenmiş

$s(n)$ konuşma sesi sinyalinin N uzunluğunda bir parçası olsun. “Kısa Zamanlı Gerçek Kepstrum”, $c_s(n)$, aşağıdaki gibi ifade edilir [1-3]:

$$c_s(n) = F^{-1} \{ \log |F \{ s_n(n) \}| \} = F^{-1} \{ C_s(\omega) \} \quad (4.6)$$

Burada $F\{\cdot\}$, kısa zamanlı ayrık Fourier dönüşümünü simgelemektedir. Pencerenin boyu, perde periyodunun en az 2 katı seçilmelidir. Pencere işlevi olarak Hanning pencere seçilmiştir. Pencerelemiş $s(n)$ sinyali, $r(n)$ 'nin pencerelemiş şekli olan $r_w(n, m)$ ile $h(n)$ 'in evrişimidir [2]:

$$\begin{aligned} s(n)w(n-m) &= (r(n) * h(n))w(n-m) \\ &= r_w(n, m) * h(n) \\ r_w(n, m) &= r(n)w(n-m) \\ S(\omega) &= R_w(\omega)H(\omega) \end{aligned} \quad (4.7)$$

Yukarıdaki eşitliğin her iki tarafının logaritması alınır ve ters Fourier dönüşümü uygulanırsa:

$$C_s(\omega) = \log(S(\omega)) = \log(R_w(\omega)) + \log(H(\omega)) \quad (4.8)$$

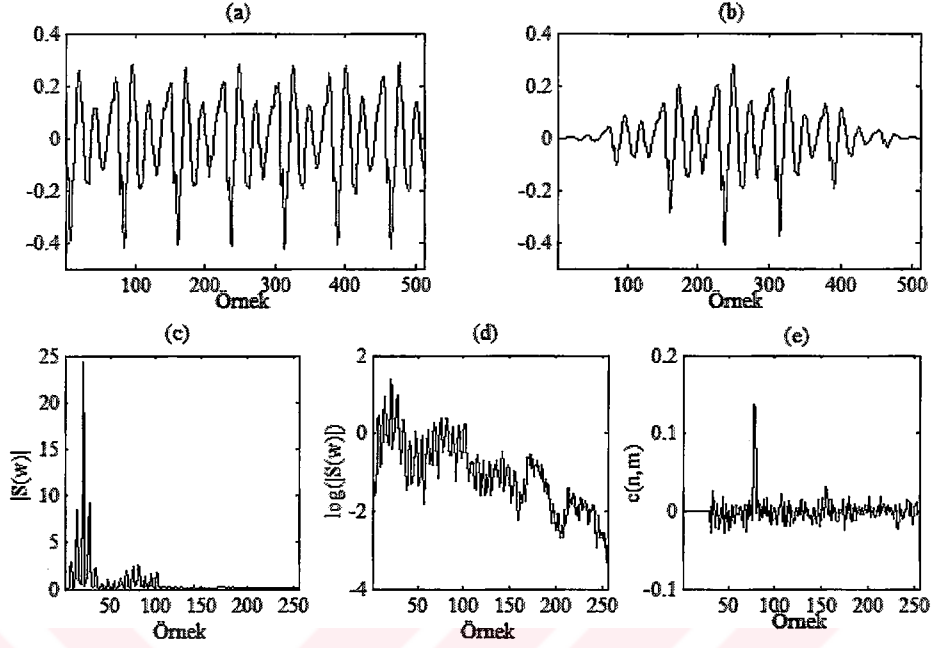
$$\begin{aligned} c_s(n) &= F^{-1} \{ C_s(\omega) \} = F^{-1} \{ C_R(\omega) \} + F^{-1} \{ C_H(\omega) \} \\ &= c_r(n) + c_h(n) \end{aligned} \quad (4.9)$$

elde edilir. Görüldüğü gibi $c_r(n)$ ile $c_h(n)$ toplam halindedir ve birbirlerinden ayrılabilirler [2]:

$$c_s(n) = \begin{cases} c_r(0) + c_h(0) & n = 0 \\ c_h(n) & 0 < n < P \\ c_r(n) & n \geq P \end{cases} \quad (4.10)$$

Burada P perde periyodudur. $c_r(n)$, P periyotlu bir darbe dizisi gibidir. Ancak n arttıkça darbelerin genlikleri düşer çünkü ses sinyali tam anlamıyla periyodik bir sinyal değildir. $c_r(n)$ 'de P periyotlu tepecikler görüldüğünden, tepe yerlerine bakarak perde frekansı kestirilir. Şekil 4.4.a'da 16 kHz ile örneklenmiş, 512 örnekli (32ms) bir ötümlü konuşma sesi parçası (/EH/) bulunmaktadır. Şekil 4.4.b, (a)'daki sinyalin Hanning pencerelemiş halidir. (c) ve (d)'de sırasıyla, (b)'deki sinyalin kısa süreli Fourier dönüşümü ve elde edilen Fourier dönüşümünün 10 tabanına göre logaritması

gösterilmektedir (Şekil 4.4'te negatif frekans bölgeleri dikkate alınmamıştır). (e) şıkında ise elde edilen kepstrum sinyali verilmektedir.



Şekil 4.4 (a) 16 kHz ile örneklenmiş /EH/ sesinin 32 ms'lik bölümü, (b) Hanning pencerelenmiş 32 ms'lik ses parçası, (c) (b)'deki sinyalin FD'nün mutlak değeri, (d) (b)'deki sinyalin FD'nün mutlak değerinin 10 tabanında logaritması, (e) İlk 30 değeri sıfırlanmış gerçel kepstrum.

(4.10)'da belirtildiği gibi, kepstrum $c_s(n)$ 'in ilk P tane değeri $c_h(n)$, daha sonraki değerleri $c_r(n)$ 'dir. $c_s(n)$, $n=0$ noktasında en büyük değerini alır. Bilgisayar analizinde, kepstrumun ilk 30 değeri değerlendirme dışı bırakılarak daha sonra gelen değerlerin en büyük genlikli olanı bulunmuş ve bulunduğu yer perde periyodu olarak belirlenmiştir. İlk 30 tane değeri (1,87 ms) değerlendirme dışı bırakmak herhangi bir sorun oluşturmaz, çünkü erkek seslerinin perde frekansları 110-150 Hz (perde periyodu 6,6-9 ms), kadın seslerinin perde frekansları 200-300 Hz (perde periyodu 3,33-5 ms) arasında değişmektedir. Şekil 4.4.e'de görüldüğü gibi kepstrumun 77, 154, 231 (P , $2P$, $3P$) numaralı örneklerinde genlikleri giderek azalan tepecikler vardır. Buradan perde periyodunun 77 örnek yani 4,8ms ($77/16000$) olduğu, perde frekansının ise 208 Hz ($1/4,8 \cdot 10^{-3}$) olduğu bilgisine ulaşılır.

4.1.5.2 Kısa Zamanlı Öz-İlişki Yöntemi

Öz-ilişki işlevi, aynı süreç içerisinde bulunan, birbirinden τ kadar uzak iki rasgele değişkenin arasındaki doğrusal ilişkinin derecesinin bir göstergesidir. Bir sinyalin kısa zamanlı öz-ilişki işlevi, $\rho_n(\tau)$, aşağıdaki gibidir[1-4]:

$$\begin{aligned}\rho_n(\tau) &= \sum_{m=0}^{N-\tau-1} s_n(m) s_n(m+\tau) \\ \rho_n(\tau) &= s_n(\tau) * s_n(-\tau)\end{aligned}\quad (4.11)$$

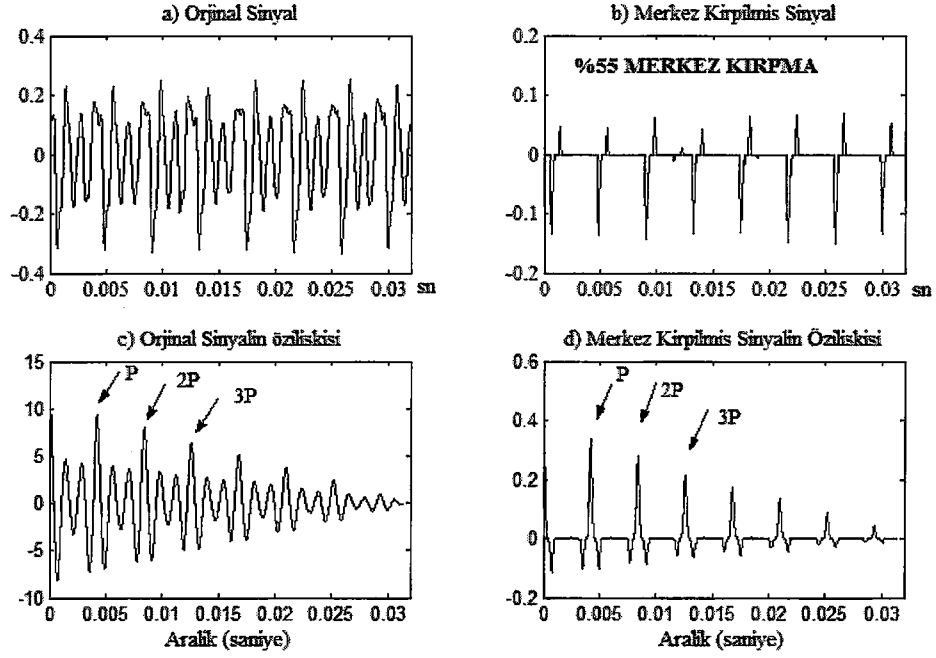
Bu eşitlikte, τ aralık (lag) değeridir. $s_n(n)$, ayrık zamanlı konuşma sinyali $s(n)$ 'in N boyunda bir pencere ile pencerelenmiş halidir:

P ile periyodik olan bir sinyalin öz-ilişki işlevi hesaplanırsa P 'nin katlarında ($\tau=0, \pm P, \pm 2P, \pm 3P..$) yüksek genlikli tepeliklerin oluştuğu görülür. Öz-ilişki işlevi de P ile periyodik olur. τ değeri perde periyoduna eriştiği anda, orijinal sinyal, $s_n(n)$, ile kaydırılan sinyal, $s_n(n+\tau)$ arasındaki fark minimum olacaktır, dolayısıyla bu iki sinyalin çarpımları büyük bir değer alacaktır. Periyodik yapıda olmayan ötümsüz seslerin öz-ilişkilerinde bu tepelere rastlanmaz. Kısa zamanlı öz-ilişki işlevi hesaplanarak perde periyodu kestirimi yapılabilir [1-3].

Bu yöntemin bazı olumsuz yanları da mevcuttur. Yöntem, gürültüye karşı kısmen dayanıklı olsa da, örnekleme hızından etkilenir. Çünkü perde periyodu, doğrudan örneklerdeki kayma ile hesaplanır. Düşük örnekleme hızı, düşük çözünürlük demektir. Bu durumda düşük örnekleme hızlarında yöntemin etkinliği azalır. Bir diğer sorun, ötümlü sesler tam olarak periyodik olmadıklarından dolayı, $\rho_n(\tau)$, P 'nin tam katlarında maksimum değerleri almayabilir. Bazı durumlarda formant değerleri baskın olabilir ve öz-ilişki, temel periyot (perde periyodu) yerine formantları yansıtabilir [1,2]. Bu problemleri ortadan kaldırmak için, “Merkez Kırpma (Center Clipping) Yöntemi” geliştirilmiştir [2]. Bu yöntemde, kısa zamanlı ses sinyali bir ön işleminden geçirilerek öz-ilişki işlevi hesaplanır. Ön işlemin amacı, sinyaldeki periyodikliği daha belirgin hale getirmektir:

$$C[s_n(n)] = \begin{cases} s_n(n) - C^+, & s_n(n) > C^+ \\ 0, & C^- \leq s_n(n) \leq C^+ \\ s_n(n) - C^-, & s_n(n) < C^- \end{cases} \quad (4.12)$$

Burada C^- ve C^+ alt ve üst kırpma sınırlarıdır. Sınır değeri, perde periyodu kestirimi için zararlı olan tepeleri kırpacak kadar büyük seçilirken, periyodikliğin tamamen kaybolmayacağı kadar da küçük seçilir (Genelde sinyalin en yüksek değerinin %30-70'i arasında seçilirler [2]). Kırpılmış sinyal, belirli bir aralıkla dizilmiş darbe dizisine benzer. Böylelikle, öz-ilişki işlevindeki tepeler daha belirgin hale gelirler.



Şekil 4.5 (a) 32ms uzunluğunda ötümlü ses parçası, (b) (a)'daki sinyalin en büyük değerinin %55'i oranında merkez kırpma yapılmış ses sinyali, (c) (a)'daki sinyalin kısa zamanlı öz-ilişki işlevi, (d) Merkez kırpılmış sinyalin öz-ilişki işlevi

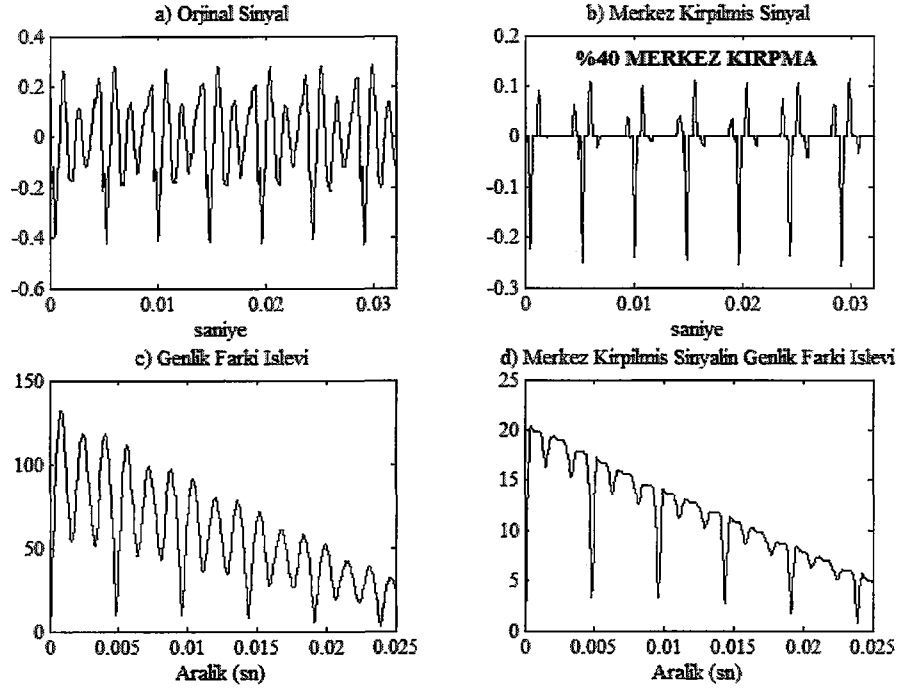
Şekil 4.5.a'da 512 örnek (32 ms) uzunluğunda dikdörtgen pencere ile pencerelenmiş ötümlü /EH/ sesi parçası bulunmaktadır. Bu sinyalin kısa zamanlı öz-ilişki işlevi (c)'de gösterilmiştir. (b) ve (d)'de sırasıyla, maksimum değerinin %55'i oranında merkez kırpılmış ses sinyali ve bu sinyalin öz-ilişki işlevi bulunmaktadır.

4.1.5.3 Kısa Zamanlı Genlik Farkı Yöntemi

“Kısa Zamanlı Genlik Farkı” aşağıdaki gibi ifade edilmektedir [1,2]:

$$\Delta M_n(\tau) = \sum_{m=0}^{N-\tau-1} |s_n(m) - s_n(m+\tau)| \quad (4.13)$$

τ aralık (lag) değeri, perde periyodu P ye eriştiği anda, orijinal sinyal $s_n(n)$ ile bu sinyalin zamanda kaydırılmış şekli $s_n(n-\tau)$ arasındaki fark en aza inecektir. Dolayısıyla $\Delta M_n(\tau)$, P 'nin tam katları olan τ noktalarında minimum değerlerde olacaktır. (Eğer sinyal tam olarak periyodik ise, $\Delta M_n(\tau)$, bu noktalarda sıfır değerini alır.) Ortalama genlik farkı işlevi öz-ilişki işlevine benzer bir işlevdir. Öz-ilişki yönteminde mevcut olan sorunlar burada da görülmektedir. Ötümlü sesler tam olarak periyodik olmadıklarından dolayı, $\Delta M_n(\tau)$, P 'nin tam katları olan τ noktalarında sıfır değerini almazlar. Baskın formant değerleri de perde periyodu kestirimini kötü yönde etkileyebilir. Bu durumda, merkez kırpma yöntemi kullanılabilir.



Şekil 4.6 (a) 32 ms'lik ötümlü ses parçası, (b) (a)'daki sinyalin en büyük değerinin %40'ı oranında merkez kırılma yapılmış ses sinyali, (c) (a)'daki sinyalin kısa zamanlı genlik farkı işlevi, (d) (b)'deki sinyalin genlik farkı işlevi

Şekil 4.6.a'da 32 ms uzunluğunda dikdörtgen pencere ile pencerelenmiş, ötümlü /EH/ sesi parçası bulunmaktadır. Bu sinyalin kısa zamanlı genlik farkı işlevi (c)'de gösterilmiştir. Genlik farkı işlevi hesaplanırken en büyük aralık değeri $\tau_{max}=25$ ms seçilmiştir. (b) ve (d)'de sırasıyla, maksimum değerinin %40'ı oranında merkez kırılmış ses sinyali ve bu sinyalin kısa zamanlı genlik farkı işlevi bulunmaktadır.

4.2 Konuşma Sesi Kalıntı Sinyalinin Elde Edilmesi

Elimizdeki konuşma sesi sinyallerinin ötümlü kısımları ayrıldıktan sonra, bu sinyallerin üretiminde uyarıcı görevi gören kalıntı sinyalleri elde edilmiştir. Bu bölümde ses kalıntı sinyalinin elde edilme aşamaları anlatılmaktadır.

4.2.1 Bölmeleme ve Hanning Pencereleme

Konuşma sesi sinyali durağan olmayan bir sinyaldir. İstatistiksel özellikleri çok kısa zaman aralıklarında hızlı bir şekilde değişir. Fakat 20-40 ms gibi zaman aralıklarında durağan kabul edilebilirler [2,3]. Bu yüzden konuşma sesi sinyali küçük bölümlere ayrılarak incelenir. Bölmeleme işlemi, belli bir aralıkta sıfırdan farklı, bu aralığın dışında sıfır değerlerini alan pencere işlevleri ile gerçekleştirilir. Sinyal ile pencere

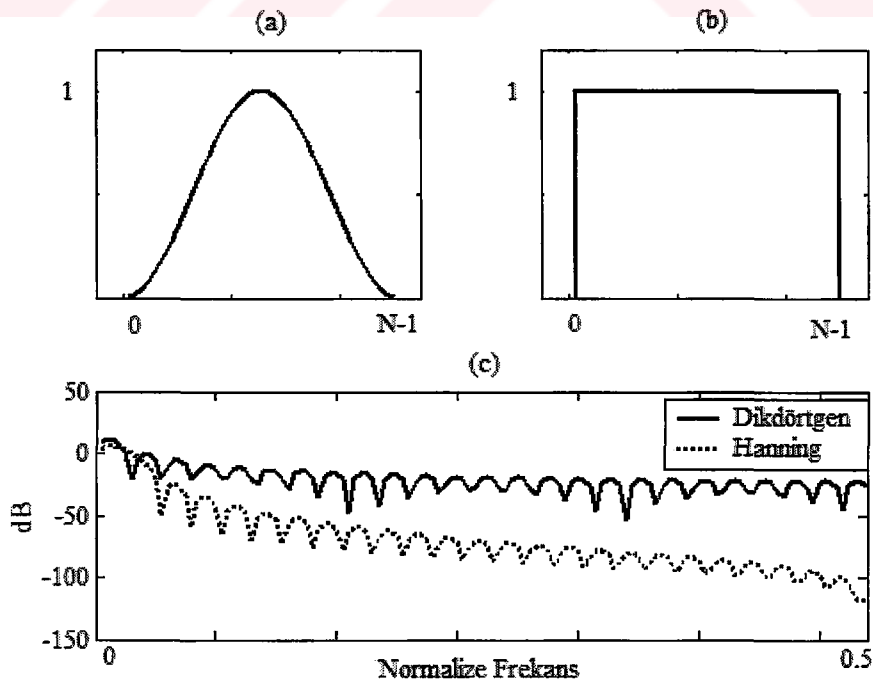
çarpılarak küçük ses parçacıkları elde edilir. Çeşitli pencere işlevleri mevcuttur. Dikdörtgen, Hamming, Hanning, Blackman en yaygın olarak kullanılan pencere işlevleridir [2,3]. Deneylerimizde Hanning penceresi tercih edilmiştir. N boyundaki Hanning Pencere işlevi aşağıdaki şekilde ifade edilir:

$$w(n) = \frac{1}{2} - \frac{1}{2} \cos\left(\frac{2\pi n}{N-1}\right), \quad 0 < n < N-1 \quad (4.14)$$

Dikdörtgen pencere ise:

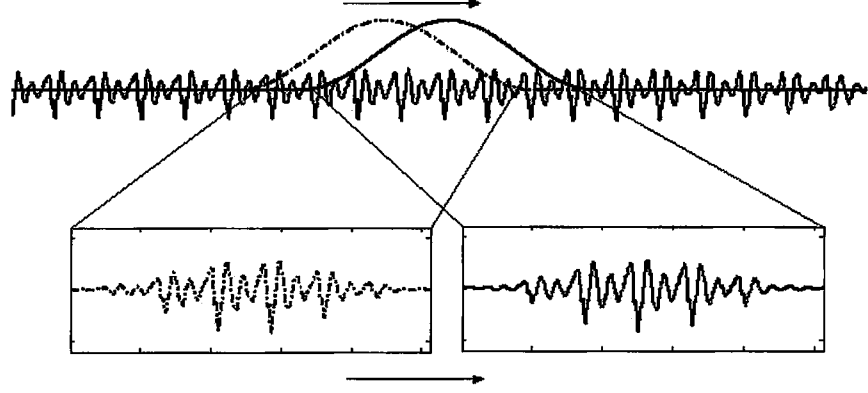
$$w(n) = \begin{cases} 1, & 0 \leq n \leq N-1 \\ 0, & \text{diğer} \end{cases} \quad (4.15)$$

şeklindedir. Dikdörtgen pencere, $[0, N-1]$ aralığındaki tüm sinyal değerlerini aynen bıraktığı ve bu aralığın dışında kalan değerleri sıfırladığı için, sinyal başlangıcı ve sonunda keskin değişimlere sebebiyet verir. Uçlara doğru yumuşak bir düşüş gösteren pencereler dikdörtgen pencereye tercih edilirler. Şekil 4.7'de Hanning pencere ile dikdörtgen pencere işlevleri ve bu işlevlerin güç izgeleri gösterilmektedir. Şekil 4.7.c'de görüldüğü gibi, Hanning pencerenin ana lobu, dikdörtgen pencereninkine göre daha geniştir. Bu istenmeyen bir durumdur. Ancak, yan loblarının genliklerinin dikdörtgen pencereninkilere göre daha düşük olması sebebiyle Hanning pencere tercih edilir.



Şekil 4.7 (a) Hanning pencere, (b) Dikdörtgen pencere, (c) Pencerelerin güç izgeleri

Deneylerimizde konuşma sesi sinyali, birbiri üzerine 25 ms ile örtüşen 32 ms uzunluğundaki Hanning pencere işlevi ile çarpılarak 32 ms'lik küçük ses parçacıkları elde edilmiştir. Şekil 4.8'de kayan pencere yöntemi gösterilmektedir. Kayan pencere yöntemi ile sinyaldeki hızlı değişimler yakalanabilmekte ve sinyalin özellikleri daha iyi kestirilebilmektedir.



Şekil 4.8 Kayan Hanning pencereler

4.2.2 Ön-vurgulama (Beyazlatma)

Daha önce de belirtildiği gibi, ötümlü seslerin güç izgeleri incelendiğinde, frekans arttıkça genliği azalan bir davranışa sahip oldukları gözlemlenmektedir. Yüksek frekans bölgelerindeki formantların genlikleri düşüktür. Bu bölgedeki zayıf genlikli formantların doğrusal öngörüsü için, doğrusal öngörü analizine bir ön işlem olarak, konuşma sesi sinyalinin güç izgesinin beyazlatılması gerekmektedir. Ön-vurgulama süzgeci işte bu beyazlatma görevini yerine getirmektedir. Eğer bir sinyalin güç izgesinin dinamik erimi (dynamic range) yüksekse, yani en düşük değer ile en yüksek değer arasındaki fark büyükse, doğrusal öngörü analizinde kullanılan öz-ilişki matrisi kötü koşullu olur [2]. Ön-vurgulama işlemi ile, yüksek frekans bölgelerindeki tepeliklerin genlikleri arttırılarak dinamik erim azaltılır.

Ötümlü seslerin kalıntı sinyalleri, frekans bölgesinde yaklaşık -6 dB/oktav'lık bir düşüş gösterir (-12 dB/oktav boğaz bölgesi titreşimini, +6 dB/oktav dudak yayılımını tanımlar [2,17]). Bu düşüşü dengelemek için yaklaşık +6 dB/oktav'lık, birinci dereceden sonlu darbe cevaplı (FIR) süzgeç olan sabit ön-vurgulama süzgeci kullanılır:

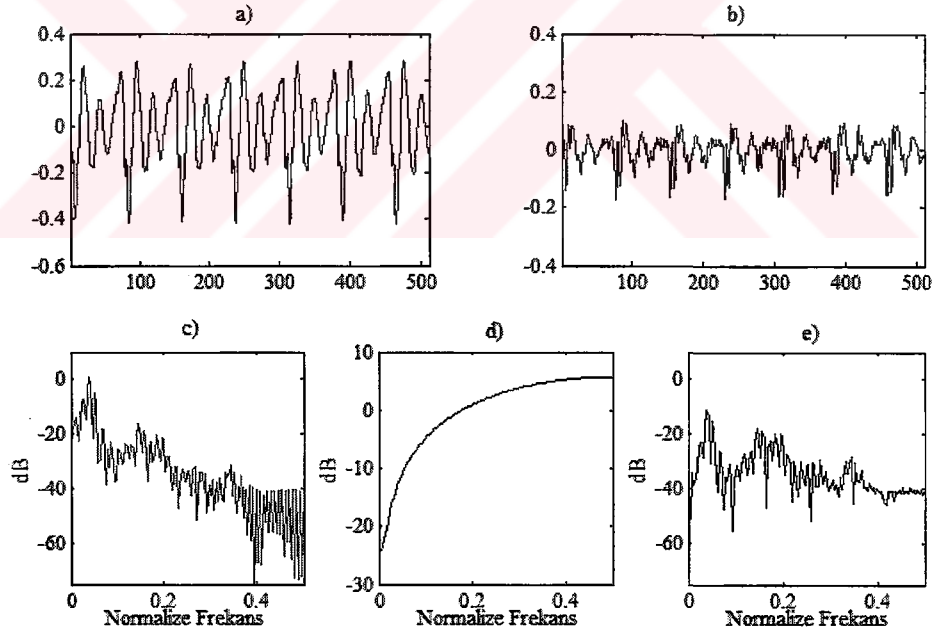
$$B(z) = 1 - bz^{-1} \quad (4.16)$$

Buradaki b katsayısı genelde $0.9 \leq b \leq 1$ aralığında seçilir [2] (Deneylerimizde b katsayısı 15/16 seçilmiştir). Şekil 4.9 (a) ve (b)'de, sırasıyla, 512 noktalık ötümlü ses sinyali ve bu sinyalin ön-vurgulanmış şekli gösterilmektedir. Şekil 4.9 (c)'de ses sinyalinin güç izgesi, (d)'de ön-vurgulama süzgecinin güç izgesi, (e)'de ise ön-vurgulanmış sinyalin güç izgesi bulunmaktadır. (d) ve (e) şıkları karşılaştırılırsa, ön-vurgulamanın, alçak frekans bölgelerindeki tepeciklerin genliklerini azalttığı, yüksek frekans bölgelerindeki tepeciklerin genliklerini artırdığı görülür.

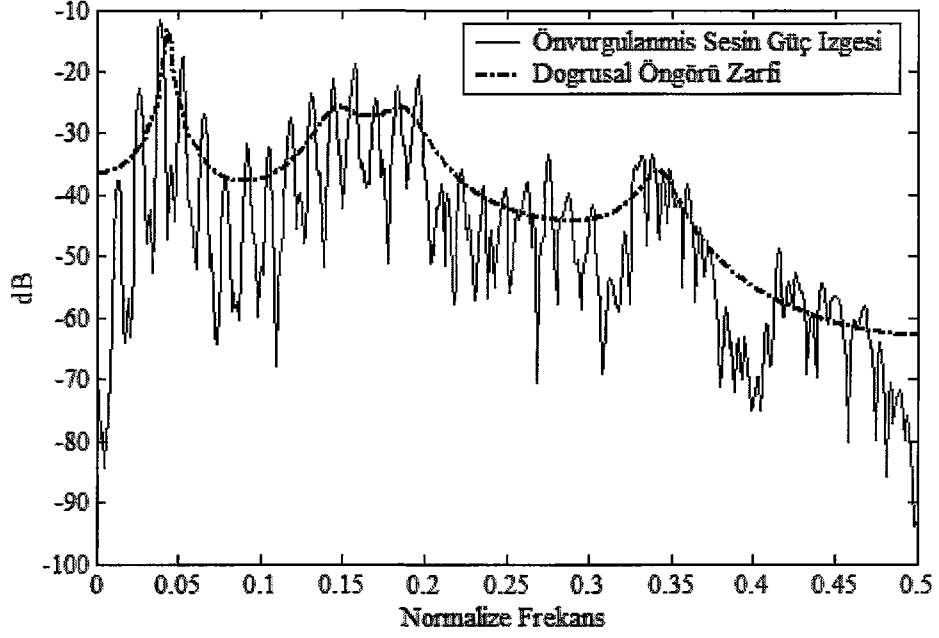
4.2.3 Doğrusal Öngörü Analizi ve Tersinir Süzgeçleme

Bölmeleme ve ön-vurgulama işlemlerinden geçirilen ses sinyaline doğrusal öngörü analizi uygulanarak, sistem parametreleri a_k katsayıları, yardımıyla Şekil 2.10'da gösterilen tersinir süzgeç ($H^{-1}(z)=A(z)$) oluşturulur. Ses sinyali bu tersinir süzgeçten geçirilerek, ses kalıntı sinyali, $r(n)$, elde edilir. Tersinir süzgeçleme z -bölgesinde aşağıdaki şekilde ifade edilir:

$$R(z) = H^{-1}(z)S(z) \quad (4.17)$$

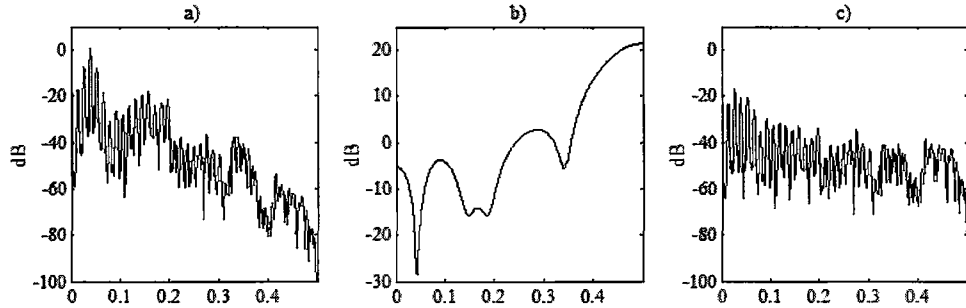


Şekil 4.9 (a) 512 Örneklilik ses sinyali, (b) (a)'daki sinyalin ön-vurgulanmış hali, (c) (a)'daki sinyalin güç izgesi, (d) Ön-vurgulama süzgecinin güç izgesi, (e) Ön-vurgulanmış sinyalin güç izgesi

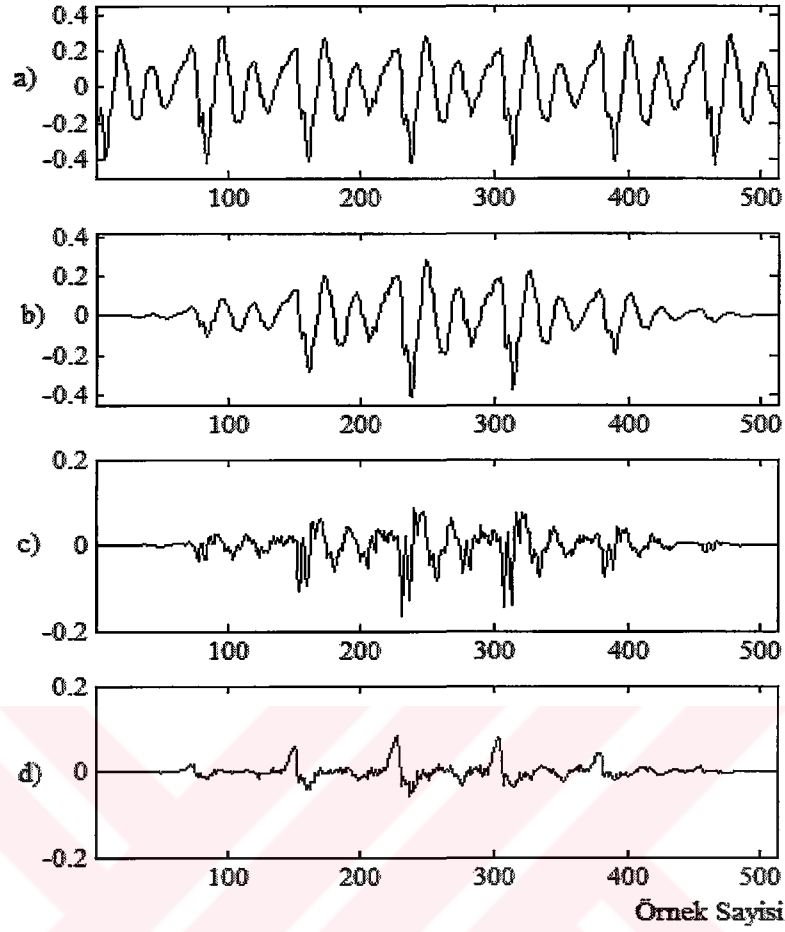


Şekil 4.10 Önvurgulanmış ve Hanning pencerelenmiş ses parçasının güç izgesi ve doğrusal öngörü zarfı

Deneylerimizde öngörücü derecesi $p=12$ seçilmiştir. Şekil 4.10'de Hanning pencerelenen ve ön-vurgulanan ötümlü ses parçasının güç izgesi ve doğrusal öngörü zarfı gösterilmektedir. Sonuç olarak $R(z)$, $S(z)$ 'in $H^l(z)$ 'ten geçirilmesiyle elde edilebilmektedir. Şekil 4.11 (a)'da Hanning pencerelenmiş ses sinyalinin güç izgesi, (b)'de doğrusal öngörü analizi yardımıyla elde edilen tersinir süzgecin güç izgesi, (c)'de ise kalıntı sinyalinin güç izgesi gösterilmektedir.



Şekil 4.11 (a) Konuşma sesi sinyalinin güç izgesi, (b) Tersinir süzgecin güç izgesi, (c) Ses kalıntı sinyalinin güç izgesi



Şekil 4.12 (a) 512 örnekli ötümlü ses parçası, (b) Hanning pencerenilmiş ses parçası (c) Ön-vurgulanmış ses parçası, (d) Ses kalıntı sinyali

Şekil 4.12 (a)'da 16 kHz ile örneklenmiş, 512 örnekli ses parçasığı bulunmaktadır. (b) ve (d)'de sırasıyla, bu sinyalin, Hanning pencerenilmiş ve ön-vurgulama işleminden geçirilmiş halleri gösterilmektedir. Doğrusal öngörü ve ardından tersinir süzgeçleme yardımıyla (d)'deki ötümlü ses kalıntı sinyali elde edilmiştir.

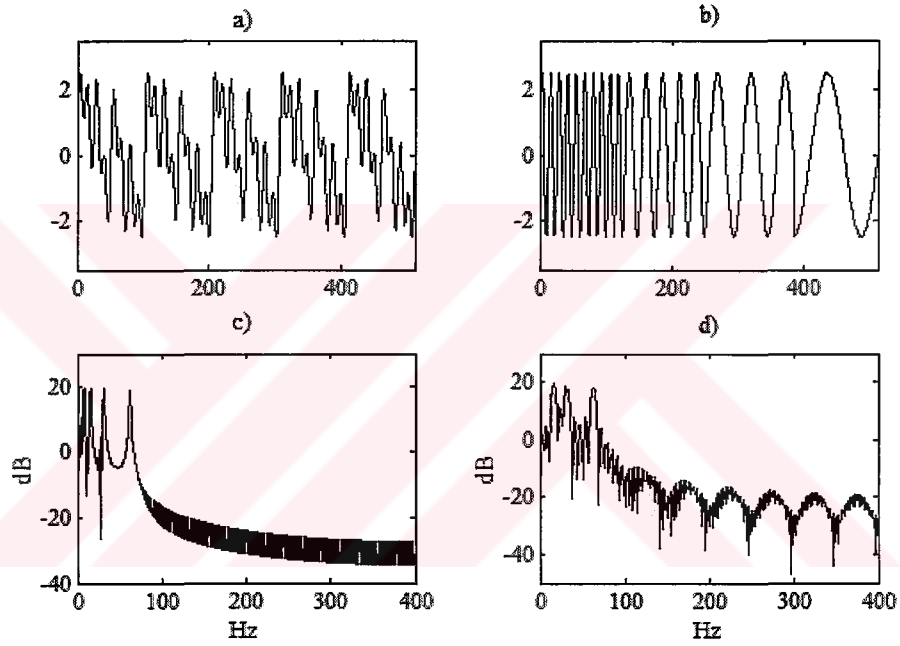
4.3 Dalgacık Tabanlı Analiz

Bölüm 4.2'de anlatılan işlemlerle elde edilen ses kalıntı sinyallerinin $1/f$ özelliğinin incelenmesinde ve öz-benzeşim parametrelerinin kestiriminde dalgacık tabanlı analiz etkin bir yöntemdir [6]. Bu bölümde dalgacık dönüşümü ve $1/f$ süreçlerin dalgacık tabanlı analizi anlatılmaktadır.

4.3.1 Dalgacık Dönüşümü

Bir sinyalin ayırt edici özellikleri genellikle sinyalin içerdiği frekans bilgilerinde saklıdır. Frekans ortamına geçiş için “Fourier Dönüşümü (FD)” etkin bir yöntemdir. Sinyale FD uygulanarak o sinyalin barındırdığı frekans değerleri elde edilebilir.

Durağan olmayan sinyallerde aynı anda hem frekans hem de zaman bilgisine sahip olmak gerekir. Eğer sinyal, ani frekans değişimlerine sahip, durağan olmayan bir sinyalse, sinyalin barındırdığı frekans değerleri FD ile elde edilir ama bu frekanstaki bileşenlerin hangi zaman diliminde ortaya çıktığı bilinemez [18,19]. Örneğin Şekil 4.13'teki sinyalleri karşılaştırabiliriz.



Şekil 4.13 (a) Farklı frekanslardaki 4 sinüs sinyalinin toplamı olan durağan sinyal, (b) Zaman ilerledikçe frekansı değişen, durağan olmayan sinyal, (c) (a)'daki sinyalin güç izgesi, (d) (b)'deki sinyalin güç izgesi

Şekil 4.13.a'da $f_1=5$, $f_2=10$, $f_3=20$ ve $f_4=40$ Hz frekans değerlerine sahip 4 sinüs sinyalin toplamı olan durağan bir sinyal gösterilmektedir. Bu frekans bileşenleri sinyalin tanımlı olduğu tüm zaman bölgesinde eşit miktarda mevcuttur:

$$s(n) = \sin(2\pi f_1 n/N) + \sin(2\pi f_2 n/N) + \sin(2\pi f_3 n/N) + \sin(2\pi f_4 n/N), n=[1,2,...,N]$$

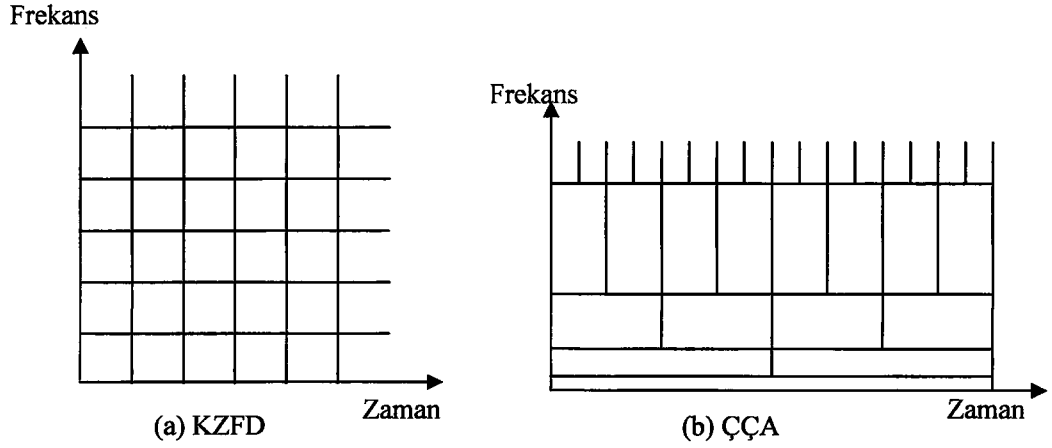
Şekil 4.13.b'deki sinyal de 5, 10, 20 ve 40Hz frekans değerlerine sahip 4 sinüs sinyalin birleşiminden oluşmuştur. Frekanslar farklı zaman aralıklarında ortaya çıktığından sinyal durağan değildir. Şekil 4.13 (c) ve (d)'de sinyallerin güç izgeleri verilmiştir. Güç izgelerindeki yüksek genlikli 4 tepecik 4 farklı frekans bileşenini

göstermektedir. Görüldüğü gibi her iki durumda da FD, sinyallerin ana frekans bileşenlerini ayırmıştır ancak bu frekanslara sahip bileşenlerin hangi zaman aralığında bulunduğu hakkında bir bilgi vermemektedir. Bu sorunu ortadan kaldırmak için “Kısa-Zaman Fourier Dönüşümü, (KZFD)” geliştirilmiştir. Bu yöntemde sinyal, sonlu uzunlukta bir pencere işlevi ile çarpılarak, durağan kabul edilebileceği kısa zaman aralıklarında işlenmektedir. Böylelikle, hangi frekans bileşenlerinin hangi zaman aralıklarında ortaya çıktığı belirlenebilir.

KZFD, aşağıdaki şekilde tanımlanır:

$$F_{KZ}(t', f) = \frac{1}{2\pi} \int_t [x(t)w^*(t-t')] \exp(-2j\pi ft) dt \quad (4.18)$$

Burada $x(t)$, dönüşüme uğrayan sinyal, $w(t)$, pencere işlevidir (“*” eşlenik ifadeyi temsil etmektedir). t' öteleme parametresi, f ise frekans değişkenidir. Dönüşüm sonucunda, sinyalin, 2-boyutlu zaman-frekans gösterimi elde edilir. KZFD’de pencere genişliği önemli bir özelliktir. Pencere boyunun küçük seçilmesi halinde frekans çözünürlüğü azalırken, büyük seçilmesi halinde ise zaman bilgisi kaybolur. Pencere boyu tüm zaman aralıkları için sabit tutulduğundan, sabit zaman-frekans çözünürlüğü sağlanır. Pencere boyuyla alakalı “İyi frekans çözünürlüğü mü? - İyi zaman çözünürlüğü mü?” ikilemini ortadan kaldırmak için “Çoklu-Çözünürlük Analizi (ÇÇA)” geliştirilmiştir. ÇÇA’inde zaman çözünürlüğü ve frekans çözünürlüğü değişkendir [18,19]. Şekil 4.14.b’deki dikdörtgenler “değişken çözünürlük” kavramını açıklamak için yaygın olarak kullanılmaktadırlar. Buradaki her bir kutucuk, ÇÇA’nin zaman-frekans bölgesindeki bir değerine karşılık gelmektedir. Kutucuklar belli bir alana sahiplerdir. Bu da o bölge içerisindeki tek bir noktanın hangi frekans ve zamana ait olduğunun bilinmemesi anlamına gelir. Zaman-frekans alanındaki aynı kutucuk içerisinde bulunan tüm noktalar tek bir dalgacık dönüşümü değeriyle ifade edilirler. Kutucukların yükseklikleri ve genişlikleri değişmesine rağmen alanları sabit kalmaktadır. Her bir kutucuk zaman-frekans bölgesinde eşit alanları tanımlar fakat farklı zaman ve frekans çözünürlükleri sağlar. Alçak frekans bölgelerindeki kutucukların boylarının kısa olması frekans çözünürlüğünün yüksek olması anlamına gelirken, enlerinin geniş olması da zaman çözünürlüğünün düşük olması anlamına gelmektedir.



Şekil 4.14 Zaman-frekans çözünürlükleri

ÇÇA’inde yüksek frekans bölgelerine doğru gidildikçe kutucukların boyları artarken enleri daralır. Dolayısıyla azalan bir frekans çözünürlüğü ve artan bir zaman çözünürlüğü elde edilir. Çoğu sinyalde yüksek frekans bileşenleri kısa zaman aralıklarında, alçak frekans bileşenleri de uzun zaman aralıklarında bulunduğu için, ÇÇA’nin “farklı frekanslarda farklı çözünürlük” yaklaşımı gayet faydalı sonuçlar verir. Zaman-frekans alanındaki kutucuklar KZFD için Şekil 4.14.a’da gösterilen sabit boyutlu kareler halini alır. Zaman ve frekans çözünürlükleri analiz boyunca değişmemektedir.

Durağan, öteleme ile değişmeyen sinyaller için Fourier dönüşümü nasıl iyi sonuçlar veriyorsa, durağan olmayan, ölçekleme ile değişmeyen sinyaller için de dalgacık dönüşümü o derece iyi sonuçlar verir. Dalgacık dönüşümünde sinyal bir dalgacık tabanı ile çarpılarak pencerelenir. Ancak pencerelenmiş parçanın Fourier dönüşümü alınmaz. Ayrıca, ÇÇA’de olduğu gibi, pencerenin genişliği her bir izgesel bileşen için değiştirilir. ÇÇA, dalgacık tabanının ölçeklenip kaydırılmasıyla gerçekleştirilir. Bu, dalgacık dönüşümünün önemli bir özelliğidir.

Sonlu genlikli ve sonlu uzunluklu bir $x(t)$ sinyalinin “Sürekli zamanlı Dalgacık Dönüşümü (SDD)” aşağıdaki gibi ifade edilir [18,19]:

$$SDD_x(\tau, a) = \frac{1}{\sqrt{a}} \int x(t) \psi^* \left(\frac{t-\tau}{a} \right) dt \quad (4.19)$$

Burada $\psi(t)$ ana dalgacık (dalgacık tabanı) işlevidir. τ öteleme, a ölçekleme parametresidir. Ana dalgacığın ötelenmiş ve ölçeklenmiş hali olan $(1/\sqrt{a})\psi((t-\tau)/a)$, yardımcı işlevler olup, $(1/\sqrt{a})$ ifadesi enerji normalizasyonu sağlamaktadır. SDD’den

$x(t)$ sinyalini tekrar elde etmek için, yani SDD'nin terse dönülebilir olması için $\psi(t)$ sonlu uzunlukta bant geçiren bir sinyal olmalı ve aşağıdaki uygunluk koşulunu sağlamalıdır [6,18,19]:

$$\int_{-\infty}^{+\infty} \frac{|\Psi(\omega)|^2}{|\omega|} d\omega < \infty \quad (4.20)$$

$\Psi(\omega)$, ana dalgacığın Fourier dönüşümüdür:

$$\Psi(\omega) = \int_{-\infty}^{+\infty} \psi(t) \exp(-j\omega t) dt \quad (4.21)$$

$$|\Psi(\omega)| < \infty, \forall \omega \quad \Psi(0) = 0 \quad \int \psi(t) dt = 0 \quad (4.22)$$

(4.20) ve (4.22) koşulları sağlandıktan sonra $x(t)$ sinyali tekrar şu şekilde elde edilir [18,19]:

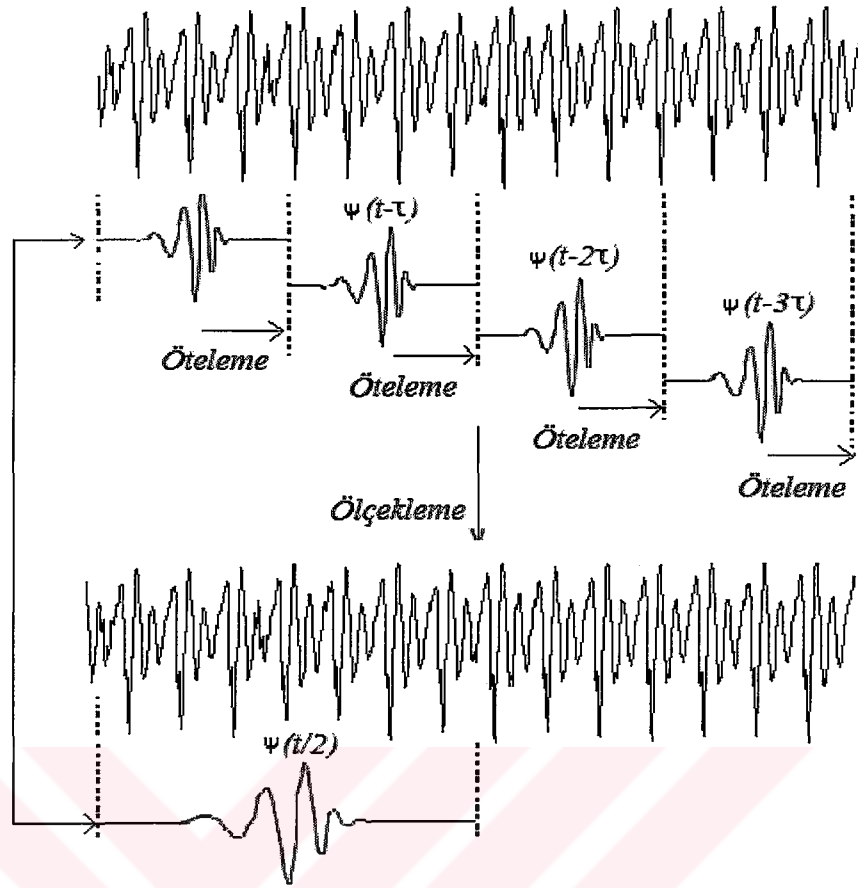
$$x(t) = \frac{1}{c_\psi} \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} SDD_x(a, \tau) \frac{1}{\sqrt{a}} \psi\left(\frac{t-\tau}{a}\right) \frac{1}{a^2} da dt \quad (4.23)$$

c_ψ , (4.20)'te belirtilen uygunluk sabitidir ve kullanılan dalgacık tabanına bağlıdır.

$$c_\psi = \int_{-\infty}^{+\infty} \frac{|\Psi(\omega)|^2}{|\omega|} d\omega < \infty \quad (4.24)$$

En iyi geri dönüşüm, $\psi(t)$ 'nin sınırlı enerjili ve ideal bant geçiren olduğu durumda sağlanır.

SDD, dalgacık tabanının ölçeklenip ötelenerek sinyalle çarpılması ve tüm zaman aralığında tümlev alınması ile gerçekleştirilir. Şekil 4.15'te öteleme ve ölçekleme işlemleri gösterilmektedir.



Şekil 4.15 Dalgacık dönüşümünde ölçekleme ve öteleme işlemleri

Belli koşullar altında, $x(t)$ sinyalini, “Ayrık Parametrelili Sürekli Dalgacık Dönüşümü (APDD)” ile ifade etmek mümkündür. SDD’de a ve τ parametreleri sürekli değerler alır ancak bu, uygulamalarda gerekmeyen bir durumdur ve hesaplama karmaşıklığı yaratır. a ve τ ’nın belli sayıdaki ayrık değerleri ile $x(t)$ sinyali karakterize edilebilir [6,18,19]. Ölçekleme parametresi, a , logaritmik olarak örneklenirken, öteleme parametresi, τ , ölçekleme parametresine göre örneklenir:

$$a = a_0^{-m}, \quad \tau = n\tau_0 a_0^{-m} \quad (4.25)$$

m ölçekleme ve n öteleme indisleridir ($-\infty < m < \infty$, $-\infty < n < \infty$, m, n : tamsayı). a_0 ve τ_0 uygun ölçekleme ve öteleme artımlarıdır ($a_0 > 1$, $\tau_0 > 0$). Ayrıklaştırma işlemi sonucunda, ana dalgacığın ötelenmiş ve ölçeklenmiş hali olan birimlik dalgacık tabanları oluşur:

$$\psi_n^m(t) = a_0^{m/2} \psi(a_0^m t - n\tau_0) \quad (4.26)$$

$$\psi_0^0(t) = \psi(t) \quad (4.27)$$

$a_0=2$ ve $\tau_0=1$ seçilirse yaygın olarak kullanılan ikinci birimlik dalgacık tabanları oluşur:

$$\psi_n^m(t) = 2^{m/2} \psi(2^m t - n) \quad (4.28)$$

Bir sinyalin ayrık parametrelili birimlik dalgacık dönüşümü (APDD) aşağıdaki şekilde ifade edilir:

$$APDD(m, n) = \int x(t) \psi_n^m(t) dt \quad (4.29)$$

$x(t)$ sinyali, ötelenmiş ve ölçeklenmiş dalgacık tabanlarının ağırlıklandırılmış toplamları şeklinde oluşturulabilir:

$$x(t) = \sum_m \sum_n x_n^m \psi_n^m(t) \quad (4.30)$$

Burada x_n^m 'ler, $x(t)$ 'nin ayrık parametrelili dalgacık dönüşümüdür ve “dalgacık katsayıları” olarak bilinirler:

$$x_n^m = \int x(t) \psi_n^m(t) dt \quad (4.31)$$

(4.31) bağıntısı sinyalin ve ikinci birimlik dalgacığın ayrık zamanlı olması durumunda aşağıdaki gibi ifade edilir ($t=kT$ de örnekleme yapılır. T Örnekleme aralığı “1” seçilmiştir):

$$x_n^m = 2^{m/2} \sum_k x(k) \psi(2^m k - n) \quad (4.32)$$

Bu eşitlik ayrık dalgacık dönüşümünün (ADD) ifadesidir [6,18,19]. ADD’de sinyal farklı frekans bantlarında farklı çözünürlüklerle analiz edilir. Bunu yaparken sinyalin “yaklaşık (a_n^m)” ve “detaylı (x_n^m)” dalgacık katsayıları elde edilir. ADD’de birimlik dalgacık tabanları, farklı kesim frekanslarına sahip alçak ve yüksek geçiren süzgeçler ile ifade edilirler. Sinyalin yüksek frekans bölgelerini analiz etmek için sinyal yüksek geçiren süzgeçler (YGS) dizisinden geçirilir (x_n^m ’ler bulunur). Aynı şekilde alçak frekans bölgelerini analiz etmek için de sinyal alçak geçiren süzgeçler (AGS) dizisinden geçirilir (a_n^m ’ler bulunur). Zaman çözünürlüğü süzgeçleme ile değiştirilirken, frekans çözünürlüğü değişimi (ölçek değişimi) alt ve üst örnekleme ile sağlanır [18,19]. Alt örneklemede sinyaldeki örneklerden bazıları atılır. Örneğin 2 ile alt örnekleme yapılırken sinyalin her iki örneğinden biri atılarak sinyal boyu

yarısına düşürülür. 2 ile üst örnekleme de ise her iki örneğin arasına bir yeni örnek katılarak sinyal boyu 2 katına çıkarılır. Bu işlem sıfır ekleyerek veya enterpolasyon ile sağlanır.

Sinyal ilk olarak yarım bant alçak geçiren süzgeçten geçirilir (Süzgecin kesim frekansı en yüksek frekansın yarısı kadardır). Bu işlem, en yüksek frekansın yarısından yukarı olan tüm frekans bileşenlerini yok eder. Ayrık sinyaller için frekans “raydan” ile ölçülür [3]. Örnekleme frekansı $f_s=2\pi$ raydandır. Sonuçta, sinyalin barındırabileceği en yüksek frekans π raydan olacaktır. AGS’den geçen sinyal, 2 ile alt örnekleme ile örnek sayısı yarıya düşürülür. En büyük frekans değeri de π ’den $\pi/2$ ye düştüğü için, örnek atma ile herhangi bir bilgi kaybı yaşanmaz. Şimdi geriye kalan sinyalde örnek sayısı yarıya inmiştir ve en yüksek frekans bileşeni $\pi/2$ ’dir. Sinyalin ölçeği 2 katına çıktığından frekans çözünürlüğü de 2 katına çıkar. Ölçek AGS ile değişmez, alt örnekleme ile değişir. AGS frekans bölgesindeki bilginin yarısını yok edeceğinden zaman çözünürlüğü yarıya iner. Alt örneklemenin zaman çözünürlüğüne bir etkisi yoktur çünkü atılan örnekler AGS’den sonra gereksiz hale gelmişlerdir. Aynı işlemler bir YGS ile de tekrarlanırlar. Bu işlemler matematiksel olarak aşağıdaki gibi ifade edilirler (AGS’nin darbe cevabı $h(n)$, YGS’nin darbe cevabı $g(n)$ ’dir):

$$a_n^m = \sum_{k=-\infty}^{\infty} h(k-2n)a_k^{m+1} \quad (4.33)$$

$$x_n^m = \sum_{k=-\infty}^{\infty} g(k-2n)a_k^{m+1} \quad (4.34)$$

a_n^{M+1} $x(n)$ ’e eşittir (M mevcut ölçek sayısı, $M=\log_2 N$). $h(n)$ ve $g(n)$ arasında aşağıdaki gibi bir ilişki mevcuttur (L süzgeç boyu):

$$g(L-1-n) = (-1)^n h(n) \quad (4.35)$$

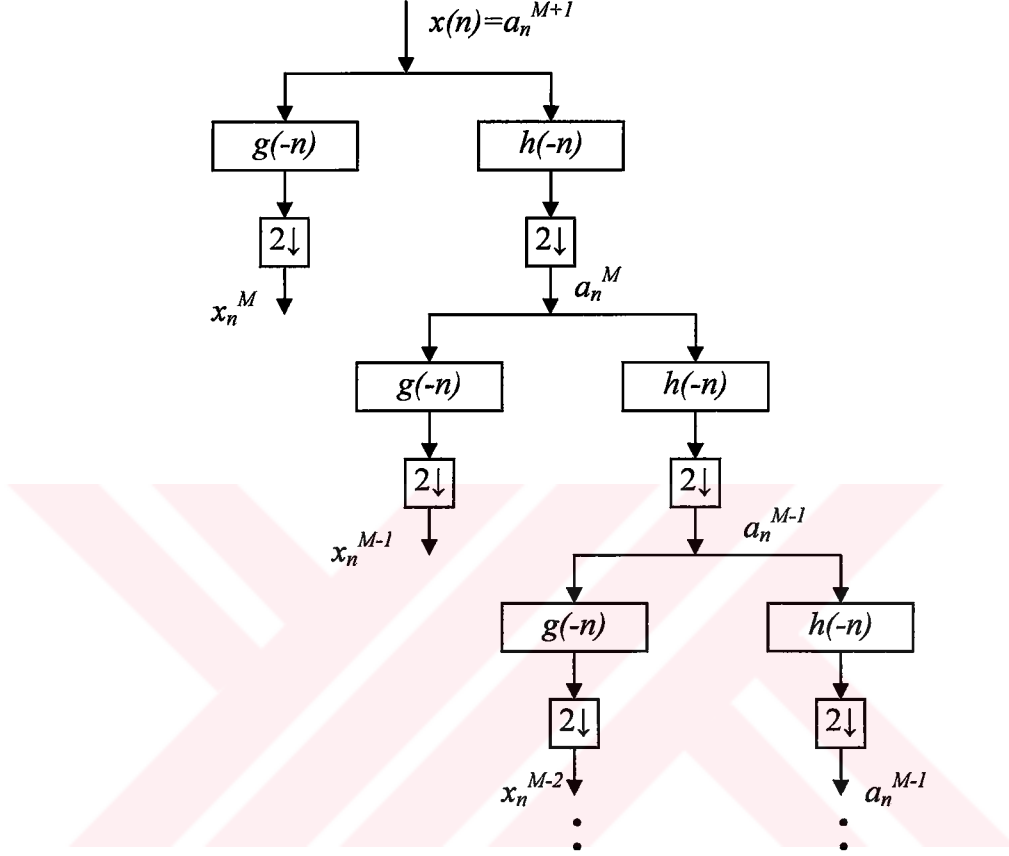
Şekil 4.16’da ayrık dalgacık dönüşümü ve katsayıların elde edilişi gösterilmektedir.

Dalgacık katsayılarından sinyali tekrar üretme algoritmayı tersine çevirmekle mümkün olur:

$$a_n^{m+1} = \sum_k \{h(n-2k)a_k^m + g(n-2k)x_k^m\} \quad (4.36)$$

Eğer kullanılan süzgeçler ideal yarım bant süzgeç değilse geri dönüşüm tam anlamıyla gerçekleştirilemez.

Dalgacık dönüşümü ve dalgacık katsayıları öz-benzeşimli sinyallerin analizinde önemli bir yere sahiptirler.



Şekil 4.16 Ayırık dalgacık dönüşümü algoritması

4.3.2 $1/f$ Süreçlerin Dalgacık Tabanlı Analizi ve Öz-benzeşim parametresi Kestirimi

Dalgacık dönüşümü durağanlık gerektirmedüğinden dolayı, $1/f$ süreçlerin incelenmesinde etkin bir araçtır.

$x(t)$, güç yasası ilişkisini sağlayan $1/f$ süreç olsun. Bu sürecin dalgacık dönüşümüyle elde edilen dalgacık katsayıları x_n^m 'ler aynı ölçek boyunca ve ölçekler arasında ilişkisiz, sıfır ortalamalı, normal dağılımlı rasgele değişkenlerdir. x_n^m 'lerin değışintileri [6]:

$$Var x_n^m = \sigma^2 2^{-\gamma m} \quad (4.37)$$

şeklindedir.

(4.37) eşitliği yaklaşık $1/f$ süreçler için:

$$\sigma_A^2 2^{-\gamma m} \leq \text{Var} x_n^m \leq \sigma_U^2 2^{-\gamma m} \quad (4.38)$$

halini alır. σ_A^2 ve σ_U^2 birbirine yakın sabit değerlerdir.

x_n^m 'lerin bir diğer ve en önemli özelliği; x_n^m 'lerin değişimleri logaritmik düzlemde bir doğru üzerinde sıralanmaktadır. Güç yasası ilişkisi dalgacık katsayılarının değişimi gelişimlerinde de görülür [6]. (4.37)'deki eşitliğin her iki tarafının 2 tabanında logaritması alınırsa karşımıza bir doğru denklemi çıkmaktadır:

$$\log_2(\text{var } x_n^m) = \log_2(\sigma^2 2^{-\gamma m}) = 2 \log_2(\sigma^2) - \gamma m \quad (4.39)$$

(4.39) denklemi $f(m)=c-\gamma m$ şeklinde olup bu denklemin eğiminin ters işaretlisi öz benzeşim parametresi γ 'nın kestirimini verir.

4.3.3 Ses Kalıntı Sinyalinin Öz-benzeşim Parametresinin Kestirimi

Doğal ve suni birçok olguda rastlandığı gibi, ötümlü seslerin kalıntı sinyallerinde de $1/|\omega|^\gamma$ şeklinde bir güç izgesi gözlenir [7,8]. Öz-benzeşim parametresinin kestiriminde dalgacık dönüşümlü analiz etkin bir yoldur.

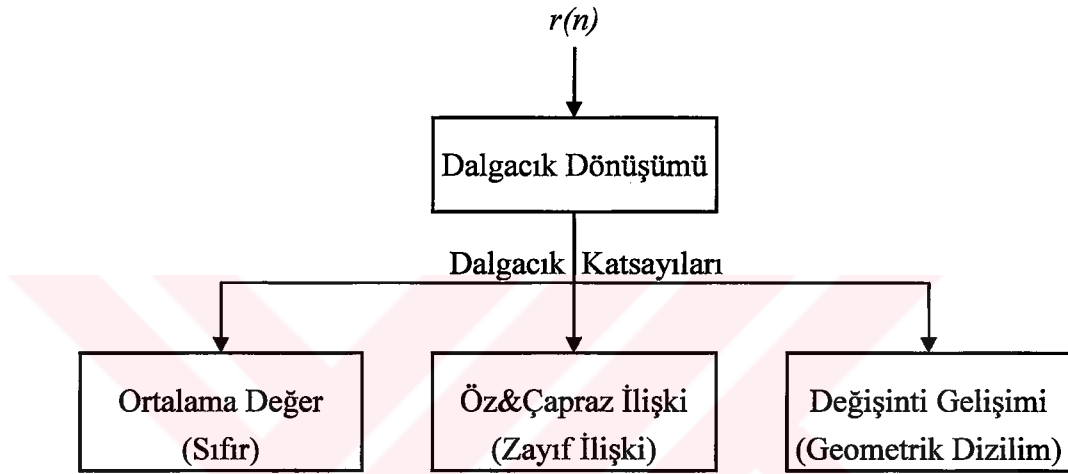
Ses kalıntı sinyaline dalgacık dönüşümü uygulanarak dalgacık katsayıları, x_n^m 'ler, elde edilir. Dalgacık katsayılarının elde edilmesi aşamasında, ikici, birimdik dalgacık tabanı olarak 5. dereceden Daubechies dalgacığı kullanılmıştır.

Tablo 4.1 Dalgacık katsayı ölçekleri ve ilişkili frekans bölgeleri

Ölçek	Katsayı sayısı	İlişkili frekans Bölgesi (Hz)
1	2	15.625-31.25
2	3-4	31.25-62.50
3	5-8	62.50-125
4	9-16	125-250
5	17-32	250-500
6	33-64	500-1000
7	65-128	1000-2000
8	129-256	2000-4000
9	257-512	4000-8000

İncelememizde, kalıntı sinyali 512 örnekten oluştuğu için, dalgacık analizi 9 ikici ölçeklidir ($2^9=512$). Ses sinyali 16kHz ile örneklendiğinden, en yüksek dereceli ölçek olan 9. ölçek 4kHz ile 8kHz arasındaki frekans bölgesine karşılık gelmektedir. Tablo 4.1’de ölçeklere ait dalgacık katsayılarına karşılık düşen frekans değerleri verilmiştir[8].

Kalıntı sinyalinin $1/f$ özeliğinde olduğunu göstermek için, dalgacık katsayıları 3 aşamalı bir işleme tabi tutulurlar. Bu işlemler Şekil 4.17’deki blok çiziminde gösterilmektedir.



Şekil 4.17 Kalıntı sinyalinin dalgacık tabanlı analizi, [8].

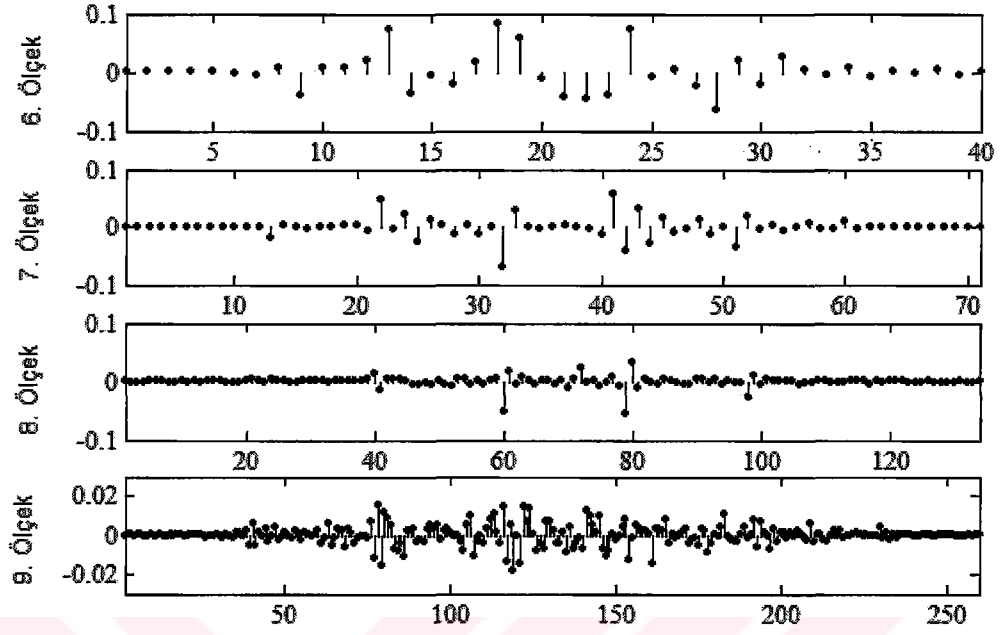
1. Elde edilen, her bir ölçekteki ayırık dalgacık katsayıları yaklaşık sıfır ortalamalı olmalı ve Gauss dağılımına sahip olmalıdır:

$$E \{x_n^m\} \cong 0 \quad (4.40)$$

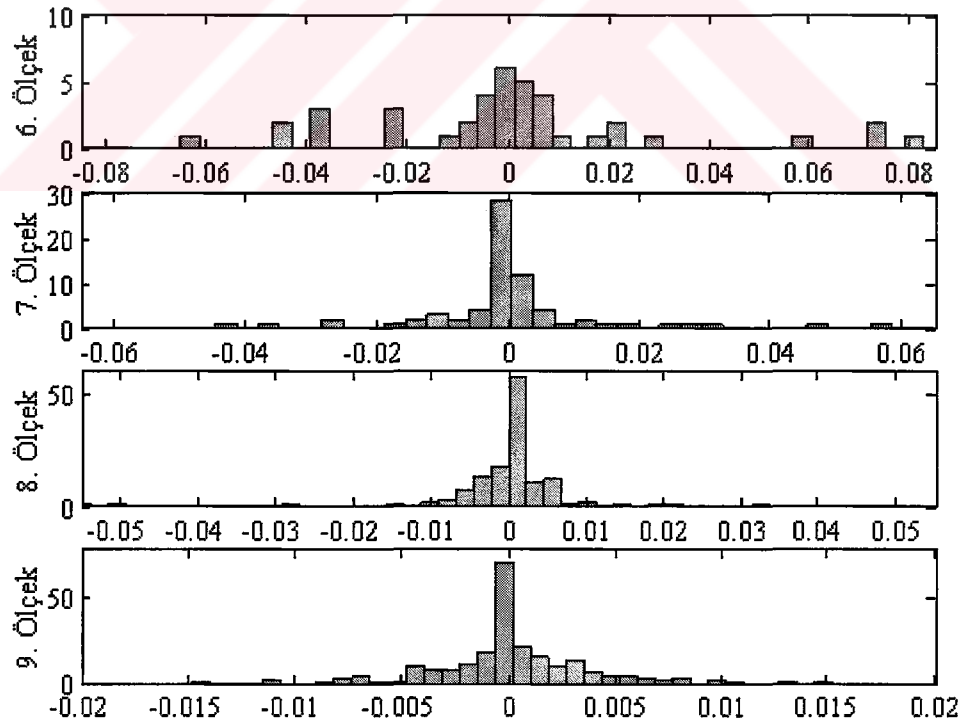
Şekil 4.18’de bir ses kalıntı sinyalinin dalgacık dönüşümü ile elde edilmiş, 6.-9. ölçeklerdeki dalgacık katsayıları gösterilmektedir. Her bir ölçekteki katsayıların ortalamaları sıfır olup yaklaşık Gauss dağılımına sahiptirler. Katsayıların dağılımları Şekil 4.19’da verilmiştir.

2. Her bir ölçekteki dalgacık katsayılarının öz-ilişkileri ve katsayıların ölçekler arası çapraz ilişkileri hesaplanmalıdır. Katsayılar, aynı ölçek içerisinde ve farklı ölçekler arasında, tamamen ilişkisiz yada zayıf ilişkili olmalıdırlar. Şekil 4.20’de her bir ölçekteki dalgacık katsayılarına ait normalize öz-ilişki dizileri gösterilmektedir.

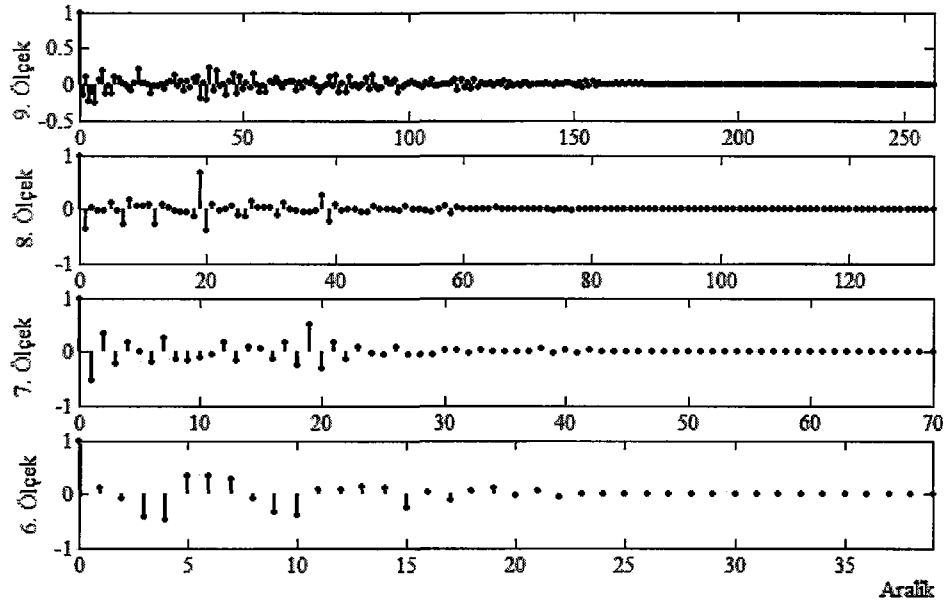
$\tau=0$ anında “1” değerini aldıkları ve diğer τ değerleri için sıfıra yakın değerlerde oldukları görülmektedir.



Şekil 4.18 6-9. Ölçeklerin dalgacık katsayıları

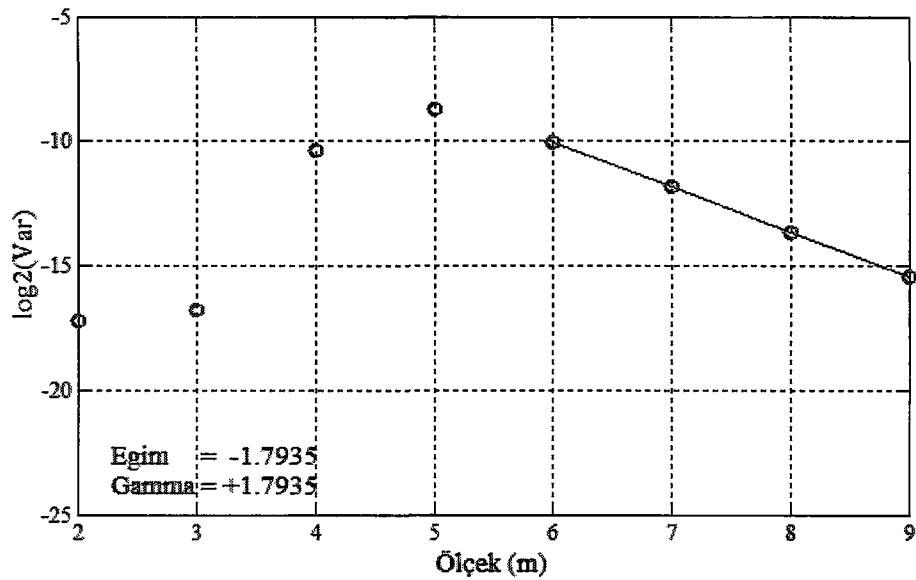


Şekil 4.19 6-9. Ölçeklerin dalgacık katsayılarının dağılımları



Şekil 4.20 Ölçeklerdeki dalgacık katsayılarının normalize öz-ilişki dizileri

3. Son ve en önemli koşul; yüksek frekans bölgelerine (500Hz'den yüksek bölgeler [8]) karşılık gelen ölçeklerdeki dalgacık katsayılarının değışintilerinin geometrik bir gelişim sergilemesi gerekmektedir. Dalgacık katsayılarının değışintileri (4.37) bağıntısıyla hesaplanarak 2 tabanında logaritmaları alınır, (4.39)'deki doğru denkleminde ulaşılır. Doğru denklemin eğiminin ters işaretlisi öz benzeşim parametresi γ 'nın kestirimini verir. Şekil 4.21'de, Şekil 4.18'de verilen dalgacık katsayılarının değışinti gelişimleri gösterilmektedir. Görüldüğü gibi 6-9 ölçekler arası değerler yaklaşık bir doğru üzerinde sıralanmaktadır. Bu değerler doğrusal uydurma yolu ile bir doğruya oturtularak, eğimin ters işaretlisi olan γ kestirilir.

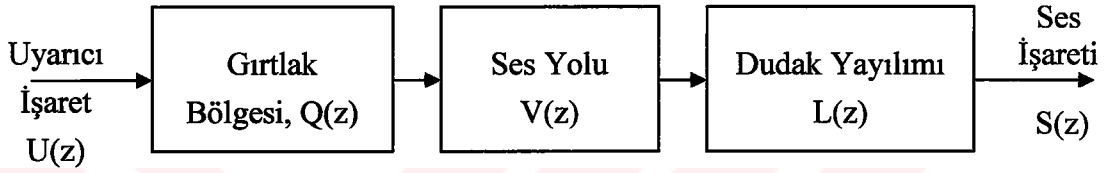


Şekil 4.21 Dalgacık katsayılarının değışinti gelişimi

4.3.4 Ön-vurgulamanın γ Kestirimine Etkisi

Doğrusal Öngöründen önce yapılan “izge beyazlatma” işlemi, diğer adıyla “ön-vurgulama”, γ parametresinin doğru kestiriminde büyük öneme sahiptir. Ön-vurgulama kalıntı sinyalinin $1/f$ yapısını muhafaza eder. Bu özellik, şu şekilde açıklanabilir:

Ses üretim sisteminin her bir bölgesini ayrı bir süzgeç ile modellemek mümkündür. Bölüm 2’de anlatılan ses yolu aktarım işlevi, $H(z)$, gırtlak bölgesini tanımlayan süzgeç $Q(z)$, ses yolu etkisini tanımlayan süzgeç $V(z)$ ve dudak yayılımını modelleyen süzgeç $L(z)$ ’in seri bağlanması sonucunda oluşan bir süzgeç gibi düşünülmektedir [3]. Modelin blok çizeneği Şekil 4.22’de gösterilmiştir.

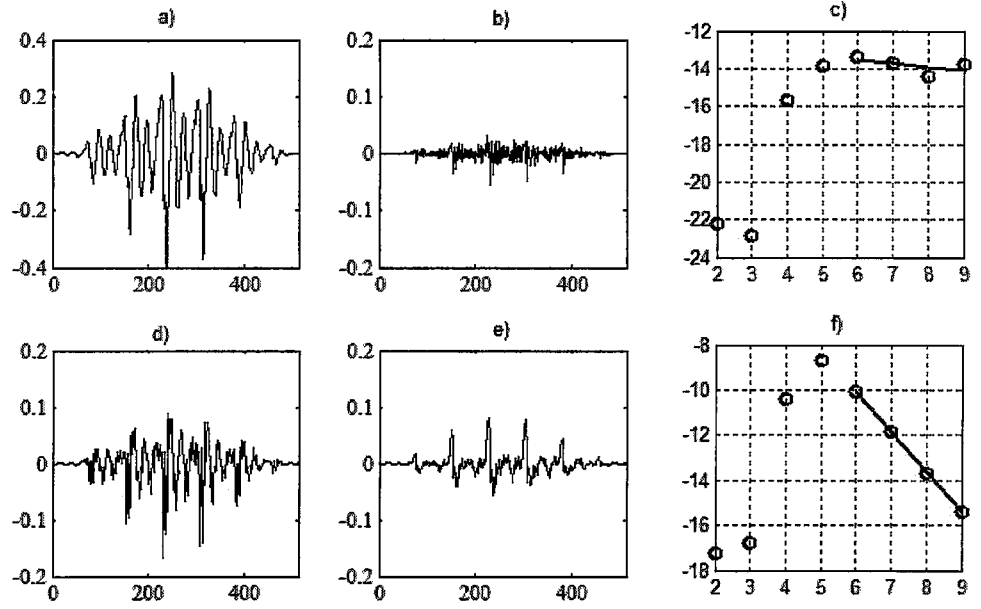


Şekil 4.22: Ses üretiminin doğrusal modeli

$Q(z)$, $z \approx 1$ ’de 2 gerçel kutbu olan minimum fazlı bir süzgeç ile modellenmektedir. $L(z)$ ise $z \approx 1$ ’de sıfırı olan 1. dereceden tüm-sıfır bir süzgeç ile modellenir. Şekil 4.22’teki süzgeçlerin seri bağlanması, z -alanında aşağıdaki şekilde gösterilir:

$$S(z) = U(z)Q(z)V(z)L(z) \quad (4.41)$$

Dudak yayılımı süzgeci, gırtlak bölgesi süzgecinin kutuplarından birinin izgesel etkisini yok eder. Sonuçta geriye ses yolu süzgeci $V(z)$ ile $z \approx 1$ ’de bir kutbu olan $Q(z)$ ifadesi kalır. Doğrusal öngöründen önce ses sinyali, $z \approx 1$ ’de sıfırı olan bir süzgeç ile ön-vurgulanırsa $Q(z)$ ’in etkisi tamamen giderilmiş olacaktır. Geriye sadece ses yolu süzgeci kalacaktır. Ön-vurgulanmış sinyalden, doğrusal öngörü yardımıyla $V(z)$ kestirilir ve tersinir süzgeçleme ile $R(z) = U(z)Q(z)$ elde edilir. Böylelikle $R(z)$ ’in $1/f$ özelliği korunmuş olur. Diğer yandan, $Q(z)$ hiç yokmuş gibi düşünülüp, ön-vurgulama yapılmaksızın doğrusal öngörü yapılsa, tersinir süzgeçleme ile $U(z)$ ’e yakın bir ifade elde edilecektir. Dolayısıyla $1/f$ ’lik ise ortadan kaybolacaktır.



Şekil 4.23 (a) Hanning pencerelenmiş 512 örnekli ses parçacığı, (b) (a)'daki ses parçacığının kalıntı sinyali, (c) Dalgacık katsayılarının değışinti gelişimleri, (d) Önvurgulanmış, Hanning pencerelenmiş 512 örnekli ses parçacığı, (e) Önvurgulanmış ses parçacığının kalıntı sinyali, (g) Dalgacık katsayılarının değışinti gelişimleri

Şekil 4.23 (a)'da 16kHz ile örneklenmiş 512 noktalı, Hanning pencerelenmiş bir ötümlü ses parçası, (b)'de bu sesin kalıntı sinyali, (c)'de dalgacık katsayılarının değışinti gelişimi verilmektedir. (d), (e) ve (f)'de ise sırasıyla, (a)'daki sinyalin önvurgulanmış hali, önvurgulanmış sesin kalıntı sinyali ve dalgacık katsayılarının değışinti gelişimleri gösterilmektedir. Şekil 4.23.c'de görüldüğü gibi, önvurgulama yapılmadan elde edilen kalıntı sinyalinin γ değeri sıfıra yakındır ($\gamma \approx 0.19$ olarak kestirilmiştir. $-1 < \gamma < 1$ aralığında olan süreçler fGn süreçleridir). $H \approx 0.5$ olup beyaz gürültü özelliği göstermektedir. Bilindiği gibi, ötümsüz seslerin uyarıcıları gürültü özelliği gösterir, fakat incelenen ses ötümlü bir sestir. Dolayısıyla kestirim yanlıştır. Önvurgulama yapıldıktan sonra elde edilen kalıntı sinyalinin γ değeri yaklaşık 1.7935 olarak kestirilmiştir.

5. DENEYLER

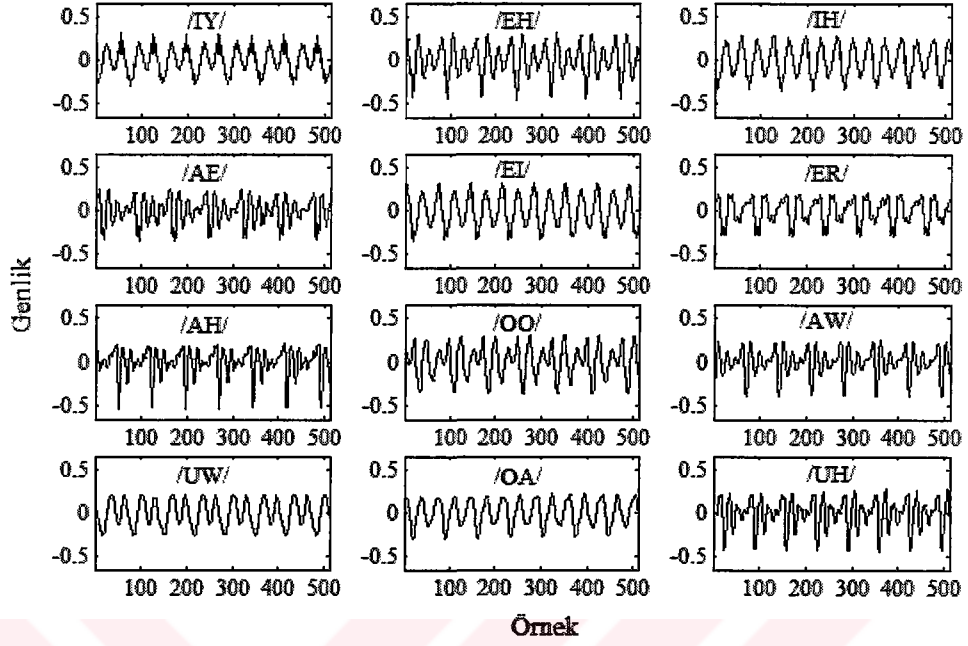
Ses bankamızda 40 farklı yetişkin erkek ve 40 farklı yetişkin kadın tarafından üretilmiş, İngilizce'deki 12 ünlü sesi barındıran, 16kHz ile örneklenmiş, sözcükler mevcuttur. Bu sözcükler Tablo 5.1'de gösterilmiştir.

Tablo 5.1: Dil kamburunun ağız içindeki konumuna göre İngilizce ünlü sesler

Dilin Durumu	Sözcükler	Sesler
ÖNDE	He <u>e</u> d	/IY/
	H <u>i</u> t	/IH/
	Hay <u>e</u> d	/EI/
	He <u>a</u> d	/EH/
	Ha <u>a</u> d	/AE/
ORTADA	He <u>a</u> rd	/ER/
	Ho <u>o</u> d	/AH/
ARKADA	Ha <u>r</u> d	/AW/
	Ha <u>w</u> ed	/OA/
	Who' <u>o</u> d	/OO/
	Ho <u>o</u> t	/UW/
	H <u>u</u> t	/UH/

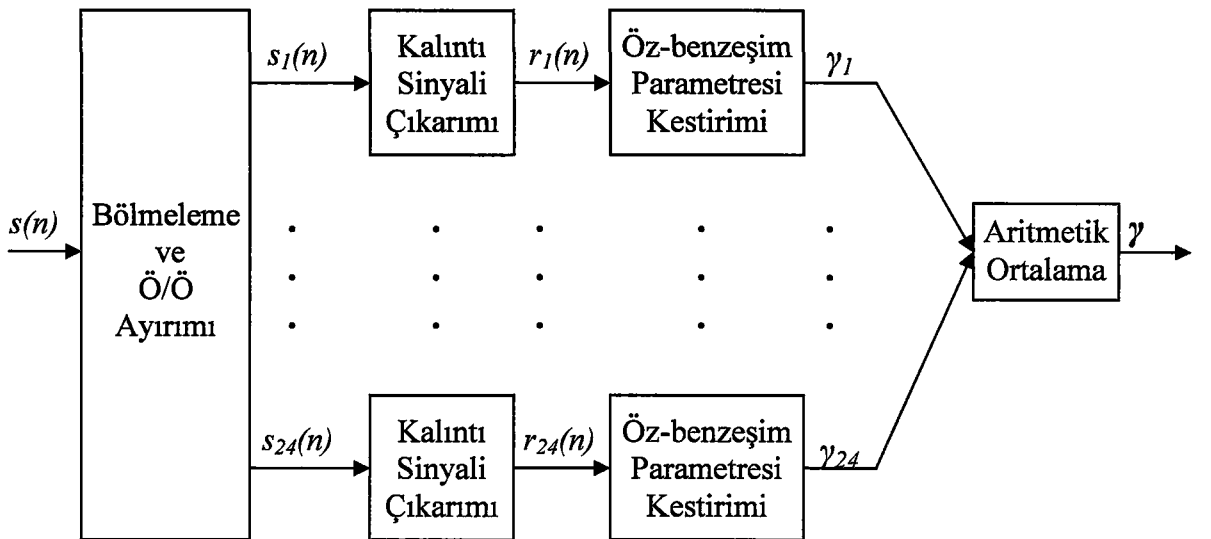
Tablo 5.1'deki sözcükler, birbiri üzerine 400 örnek (25ms) ile örtüşecek şekilde kayarak ilerleyen, 512 örneklilik Hanning pencere ile çarpılarak, 512 örnek (32ms) uzunluğuna sahip ses parçacıkları elde edilmiştir. Bu ses parçacıkları, Bölüm 4.1'de anlatılan kıstaslara göre ötümlü/ötümsüz olarak sınıflandırılmışlardır. Tablo 5.1'deki sözcüklerin ötümlü kısımları /IY/, /IH/, /EI/, /EH/, /AE/, /ER/, /AH/, /AW/, /OA/,

/OO/, /UW/ ve /UH/ ünlü sesleridir. Şekil 5.1’de bu seslerin 512 örneklilik parçalarının zaman-genlik değişimleri gösterilmektedir.



Şekil 5.1 Ünlü seslerin zaman-genlik değişimleri

Ötümlü olarak sınıflandırılan bloklardan ardı ardına gelen 24 tanesine, Bölüm 4.2’de anlatılan işlemler uygulanarak kalıntı sinyali elde edilmiş ve dalgacık tabanlı analiz yöntemi ile kalıntı sinyallerinin öz-benzeşim parametreleri, γ ’lar, kestirilmiştir. Kestirilen 24 γ parametresinin ortalamaları alınarak, incelenen sese ait γ değeri bulunmuş olur. γ parametresi kestirimi blok çizeneği Şekil 5.2’de gösterilmektedir.



Şekil 5.2 γ kestirimi blok çizeneği [Ö/Ö: Ötümlü/Ötümsüz ayırımı]

Şekil 5.2'deki blok çizenekte anlatılan işlemler, elimizdeki tüm sözcüklere uygulanarak, Tablo5.1'deki her bir ünlü sesin γ parametresi kestirilmiştir. Farklı kişiler tarafından söylenen aynı sese ait 40 γ 'nın tekrar ortalaması alınarak, Tablo 2'deki değerler elde edilir.

Tablo 5.2 Ünlü sesler ve bu seslere ait γ değerleri

	/IY/	/IH/	/EI/	/EH/	/AE/	/ER/
γ (Erkek)	1.1563	1.3655	1.3040	1.5257	1.5664	1.6464
γ (Kadın)	1.3840	1.8108	1.7485	1.8365	1.8688	1.6504
	/AH/	/AW/	/OA/	/OO/	/UH/	/UW/
γ (Erkek)	1.7103	1.8296	1.7159	1.5497	1.6599	1.5423
γ (Kadın)	1.8248	1.7914	1.6029	1.7070	1.8033	1.4385

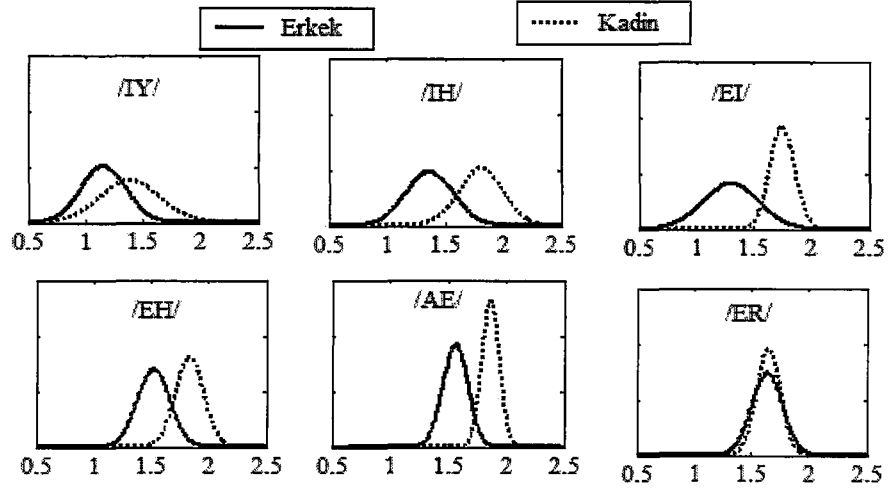
5.1 Öz-benzeşim Parametresi İle Konuşmacı Cinsiyeti Belirleme

γ parametrelerinin sestten cinsiyet belirlemede ne derece etkin olabileceğini görmek için klasik “En az hata olasılığı karar verme yöntemi (Minimum probability-of-error decision)” kullanılmıştır. İlk olarak incelenen sese ait γ değerlerini içeren bir dizi oluşturulur ($\lambda = [\gamma_1 \ \gamma_2 \ \gamma_3 \ \dots \ \gamma_{40}]$). Daha sonra γ 'lar, bu dizinin ortalama değeri ve değişintisine sahip bir Gauss olasılık dağılım işlevine oturtulurlar. 1-boyutlu Gauss dağılımı aşağıdaki şekilde ifade edilir:

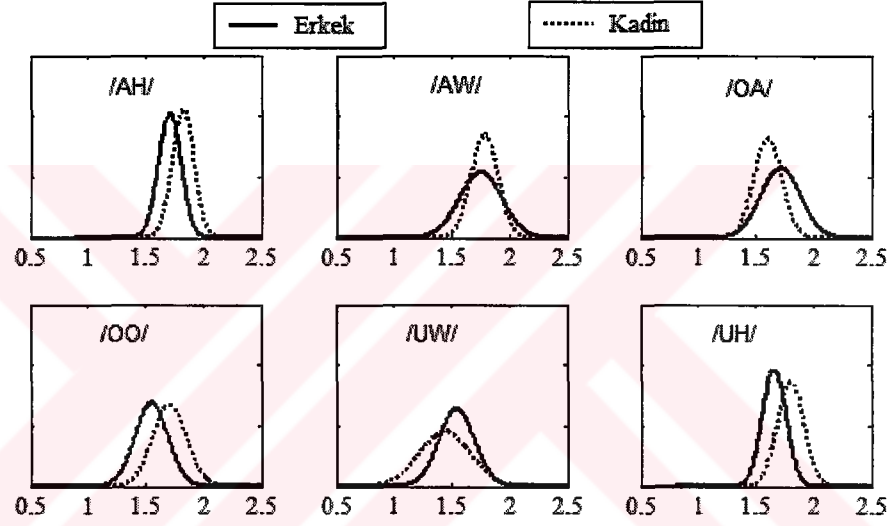
$$g(\gamma) = \frac{1}{\sqrt{2\pi\sigma_\lambda^2}} \exp\left(-\frac{1}{2}\left(\frac{\gamma - \mu_\lambda}{\sigma_\lambda}\right)^2\right) \quad (5.1)$$

Burada, σ_λ^2 , λ dizisinin değışintisi, μ_λ , dizinin ortalama değeridir. Elde edilen olasılık dağılım işlevleri, Şekil 5.3.a ve Şekil 5.3.b'de gösterilmektedirler.

Ötümlü seslerden bazılarının (/IH/, /EI/, /EH/, /AE/) kalıntılarından elde edilen γ değeri kadn ve erkek seslerinde daha fazla farklılık gösterirken bazı seslerde (/ER/, /AW/, /UW/) ise γ değeri birbirinden ayıramayacak şekilde yakın çıkmışlardır.



Şekil 5.3.a Olasılık yoğunluk işlevleri



Şekil 5.3.b Olasılık yoğunluk işlevleri

İki farklı sese ait öz-benzeşim parametresinin kombinasyonu ile sınıflandırma yapmak şu şekilde gerçekleştirilebilir:

Yapılan sınıflandırmada, 20 ayrı kadın ve 20 ayrı erkek tarafından söylenen “HEAD” ve “HIT” sözcüklerinin ötümlü kısımları olan, sırasıyla /EH/ ve /IH/ sesleri deneme verileri olarak kullanılmıştır. İlk olarak, bu ötümlü seslerden elde edilen yirmişer değere sahip örnek γ dizileri aşağıdaki gibi oluşturulur:

$$\gamma_{mE} = [\gamma_{m1,E} \ \gamma_{m2,E} \ \gamma_{m3,E}, \dots, \gamma_{m20,E}], \gamma_{mI} = [\gamma_{m1,I} \ \gamma_{m2,I} \ \gamma_{m3,I}, \dots, \gamma_{m20,I}]$$

$$\gamma_{wE} = [\gamma_{w1,E} \ \gamma_{w2,E} \ \gamma_{w3,E}, \dots, \gamma_{w20,E}], \gamma_{wI} = [\gamma_{w1,I} \ \gamma_{w2,I} \ \gamma_{w3,I}, \dots, \gamma_{w20,I}].$$

($\gamma_{mk,E}$: k . erkek, /EH/ sesine ait değer; $\gamma_{mk,I}$: k . erkek, /IH/ sesine ait değer; $\gamma_{wk,E}$: k . kadın, /EH/ sesine ait değer; $\gamma_{wk,I}$: k . kadın, /IH/ sesine ait değer; $1 \leq k \leq 20$).

Örnek γ dizileri yardımıyla, M_i ortalama değer vektörü ve W_i kovaryans vektörü elde edilerek 2-boyutlu ortak Gauss dağılımı işlevi tanımlanır:

$$g_i(X) = (2\pi)^{-1} |W_i|^{-1/2} \exp\left(-\frac{1}{2} (X - M_i)^H W_i^{-1} (X - M_i)\right) \quad (5.2)$$

$$X = [\gamma_E \ \gamma_I]^T$$

i , erkek için “ m ”, kadın için “ w ” seçilmiştir:

$$M_i = \frac{1}{N_i} \sum_{k=1}^{N_i} [\gamma_{ik,E} \ \gamma_{ik,I}] \quad W_i = \frac{1}{N_i} \sum_{k=1}^{N_i} \left(\begin{bmatrix} \gamma_{ik,E} \\ \gamma_{ik,I} \end{bmatrix} [\gamma_{ik,E} \ \gamma_{ik,I}] - M_i M_i^T \right) \quad (5.3)$$

M_m vektörünün ilk elemanı γ_{mE} 'nin ortalama değeri, ikinci elemanı ise γ_{mI} 'nin ortalama değeridir ($N_i=20$). Aynı şekilde, M_w vektörünün ilk elemanı γ_{wE} 'nin ortalama değeri, ikinci elemanı ise γ_{wI} 'nin ortalama değeridir. Örnek γ dizilerinden yararlanılarak üretilen ortalama değer vektörleri ve kovaryans matrisleri Tablo 5.3'de gösterilmektedir.

Tablo 5.3 Örnek dizilerden elde edilen ortalama değer vektörleri ve kovaryans matrisleri

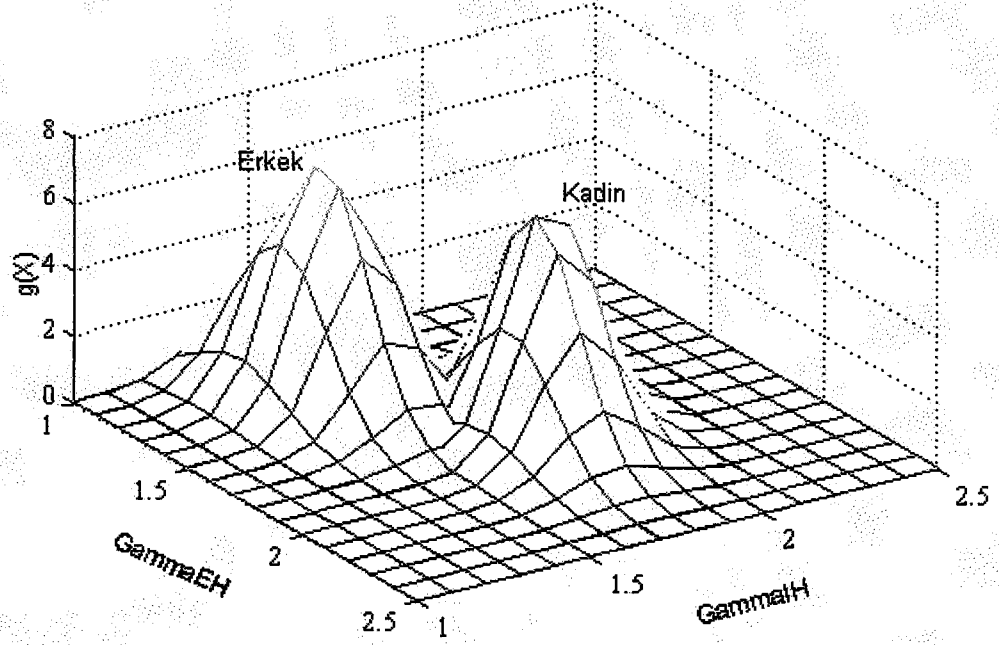
		γ_{EH}	γ_{IH}
Erkek	Ortalama, M_m	$[1.5003 \ 1.3386]^T$	
	Kovaryans, W_m	$\begin{bmatrix} 0.0137 & 0.0041 \\ 0.0041 & 0.0384 \end{bmatrix}$	
Kadın	Ortalama, M_w	$[1.8390 \ 1.7931]^T$	
	Kovaryans, W_w	$\begin{bmatrix} 0.0149 & -0.006 \\ -0.006 & 0.0372 \end{bmatrix}$	

Ortak olasılık yoğunluk işlevleri elde edildikten sonra, aşağıdaki hipotez yazılabilir:

Hipotez-m : Erkek sesi, *Hipotez-w* : Kadın sesi

- Eğer $p_i g_i(X) \geq p_j g_j(X)$ ise *Hipotez-i*'yi seç.

$g_i(X)$: Ortak olasılık yoğunluk işlevi, $X=[\gamma_E \ \gamma_I]^T$, p_i : Önsel olasılık, $i=m, w$ (*Sınıf-m*: Erkek sesi, *Sınıf-w*: Kadın sesi). $g_m(X)$ ve $g_w(X)$ 2-boyutlu ortak olasılık yoğunluk işlevleri Şekil 5.4'te gösterilmiştir.



Şekil 5.4 Ortak olasılık yoğunluk işlevleri

p_i 'ler eşit varsayılarak, hipotez aşağıdaki gibi basitleştirilebilir:

$$d_i = (X - M_i)^H W_i^{-1} (X - M_i) \quad (5.4)$$

$p_i g_i$ 'nin en büyük olması için d_i 'nin en küçük olması gerekir . Sonuçta d_i 'yi en küçük yapan i sınıfı seçilir. Örneğin bir kişiye ait “head” ve “hit” sözcüklerinde yapılan parametre kestirimi sonucunda X vektörü, $X=[1.56 \ 1.53]^T$ şeklinde bulunmuş olsun. X 'e bakarak bu kişinin cinsiyeti, d_m ve d_w hesaplanarak saptanır (M_i ve W_i değerleri Tablo 5.3'te verilmektedir) :

$$\begin{aligned} d_m &= ([1.56 \ 1.53] - [1.5003 \ 1.3386]) \begin{bmatrix} 0.0137 & 0.0041 \\ 0.0041 & 0.0384 \end{bmatrix}^{-1} \left(\begin{bmatrix} 1.56 \\ 1.53 \end{bmatrix} - \begin{bmatrix} 1.5003 \\ 1.3386 \end{bmatrix} \right) \\ &= 1.06 \\ d_w &= ([1.56 \ 1.53] - [1.8390 \ 1.7931]) \begin{bmatrix} 0.0149 & -0.006 \\ -0.006 & 0.0372 \end{bmatrix}^{-1} \left(\begin{bmatrix} 1.56 \\ 1.53 \end{bmatrix} - \begin{bmatrix} 1.8390 \\ 1.7931 \end{bmatrix} \right) \\ &= 9.29 \end{aligned}$$

Görüldüğü gibi, $d_m < d_w$ olduğu için, bu sesler bir erkeğe aittir.

Yapılan incelemelerde, büyük γ değerine sahip seslerin daha düz ve yumuşak, küçük γ değerine sahip seslerin ise daha sert ve pürüzlü olduğu gözlenmiştir. Bu ilişki

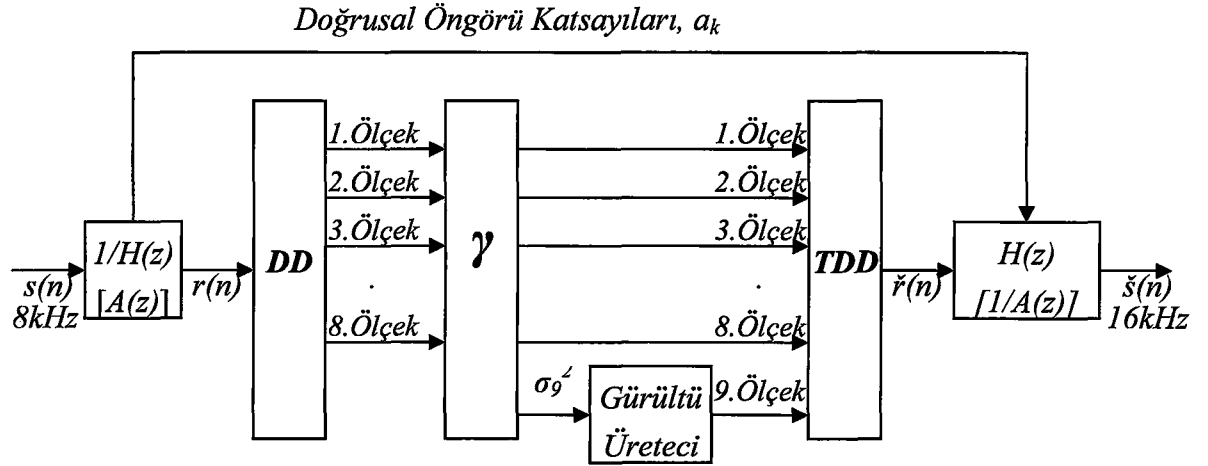
fraktal boyutu ile açıklanabilir. Bölüm 3'te bahsedildiği gibi, fraktal boyutu D ile H arasında $D=2-H$ bağıntısı bulunmaktadır. Bir sinyalin γ değeri yüksek ise H değeri de yüksek olur dolayısıyla D küçük olur. Fraktal boyutu küçük olan sinyal daha az dalgalanma gösterir ve yumuşak bir biçime sahiptir. Yapılan incelemelerde gözlenen bir diğer durum ise, elde edilen γ değerlerinin, genellikle, kadın seslerinde erkek seslerindekiyle göre daha yüksek çıktığıdır.

5.2 Ses Kalite Artırımı

Öz-benzeşim parametresi ve dalgacık tabanlı analiz, dalgacık katsayılarının bir ölçek genişletilmesi suretiyle ses kalitesi iyileştirmesinde kullanılabilir [8]. Başka bir deyişle, ses işaretine ekstra yüksek frekans bileşenleri eklenerek dar bantlı ses sinyali geniş bantlı ses sinyaline dönüştürülebilir. Ses kalitesi artırım işlemleri aşağıdaki örnekle açıklanmaktadır:

$s(n)$, 8kHz ile örneklenmiş 256 örnek (32ms) uzunluğunda ötümlü bir ses parçasığı olsun. Yöntem bölümünde anlatılan işlemler ile ses kalıntı sinyaline ulaşılır. Kalıntı sinyali de 256 örnekli bir sinyaldir. Bu sinyale dalgacık dönüşümü uygulanarak dalgacık katsayıları elde edilir. Toplam 8 ölçek mevcuttur ($2^8=256$). Her bir ölçekteki dalgacık katsayılarının değışintilerinin logaritmaları bir doğru üzerinde dizilmektedir. Bu doğrudan yararlanarak mevcut olmayan 9. ölçek katsayılarının değışintisi, σ_9^2 , kestirilebilir. Dalgacık katsayılarının sıfır ortalamalı, Gauss dağılımlı ve zayıf ilintili oldukları göz önünde bulundurularak, 9. ölçek katsayıları yerine, sıfır ortalamalı, σ_9^2 değışintili beyaz gürültü konulabilir. Sonuçta ölçek sayısı dokuza çıkacaktır. Ters dalgacık dönüşümü yardımıyla kalıntı sinyali tekrar elde edilir. 9 ölçek mevcut olduğundan, yeni kalıntı sinyali $2^9=512$ örnekli olacaktır. Kalıntı sinyali elde etme işlemleri tersine çevrilerek ses sinyali üretilir. Yeni ses sinyalinde 32 ms'lik parçacıkta 512 örnek olacaktır. Dolayısıyla $512/0.032=16$ kHz ile örneklenmiş geniş bantlı sinyale ulaşılmış olur.

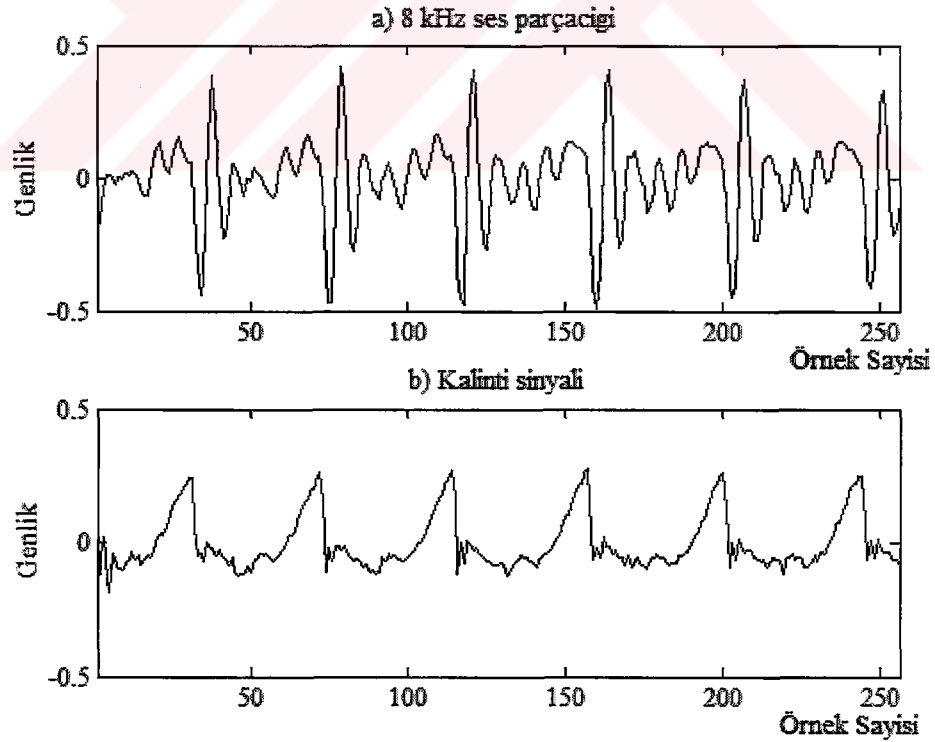
Dar bantlı sinyali geniş bantlı sinyale dönüştürme aşamaları Şekil 5.5'te gösterilmektedir. Burada $s(n)$ 8kHz ile örneklenmiş ses parçasığı, $r(n)$, ses kalıntı sinyalidir. $\check{r}(n)$ üretilen kalıntı sinyali ve $\check{s}(n)$ ise sentezlenen, 16kHz örnekleme frekansına sahip geniş bantlı ses parçasığıdır.



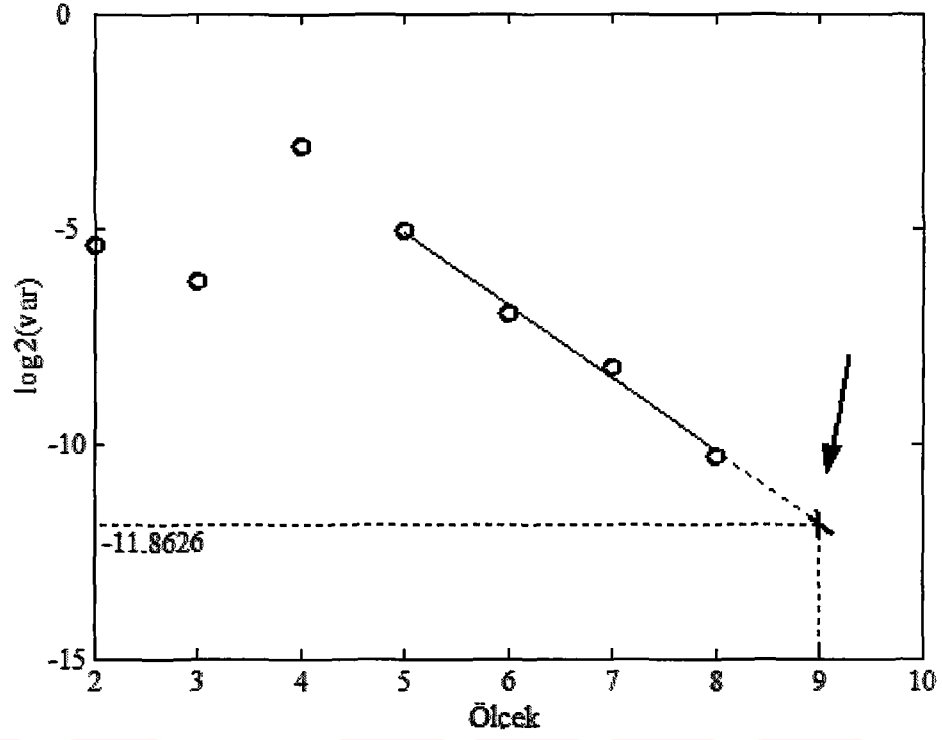
Şekil 5.5 Dar bantlı sinyalden geniş bantlı sinyale geçiş blok çizeneği (**DD**: Dalgacık dönüşümü, **TDD**: Ters dalgacık dönüşümü, γ : Öz-benzeşim Parametresi Kestirimi)

Şekil 5.6.a'da 8 kHz ile örneklenmiş, 256 örnek uzunluğunda (32ms) bir ötümlü ses parçasığı gösterilmektedir. (b)'de ise bu ses parçasığının kalıntı sinyali bulunmaktadır.

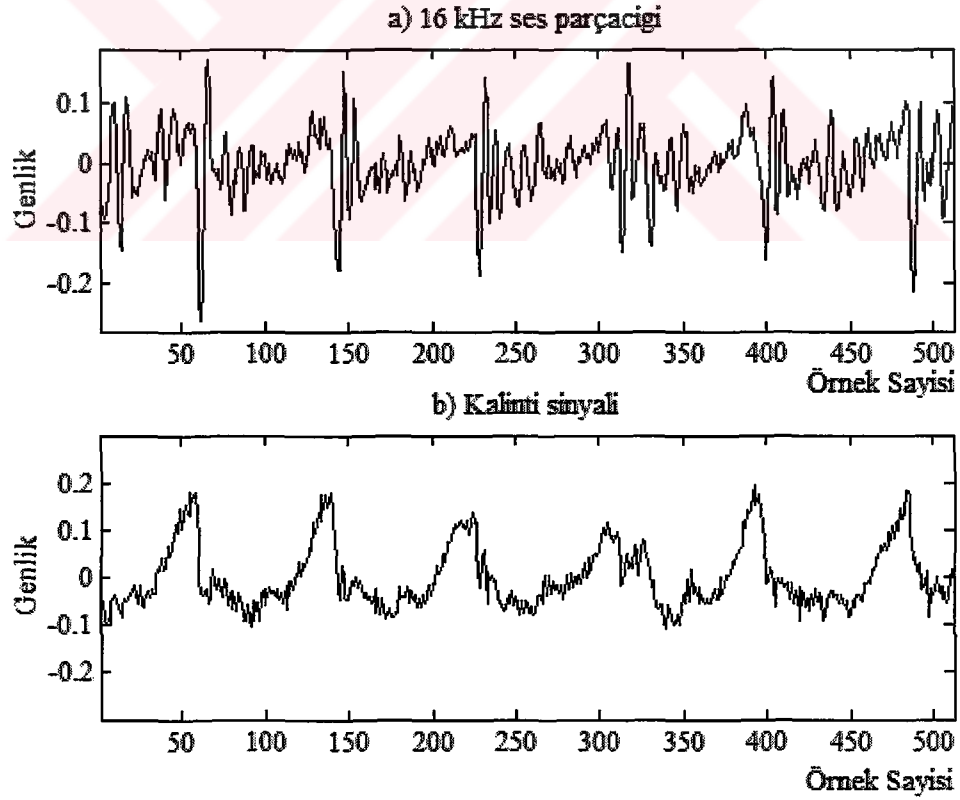
Şekil 5.7'de, mevcut olmayan 9.ölçek katsayılarının değışintisinin yerinin kestirimi gösterilmektedir. Yukarıda anlatılan işlemlerden sonra elde edilen geniş bantlı ses sinyali ve kalıntı sinyali Şekil 5.8 (a) ve (b)'de verilmektedir.



Şekil 5.6 (a) 8kHz ile örneklenmiş 32ms'lik ötümlü ses parçasığı, (b) Ses kalıntı sinyali



Şekil 5.7 Dalgacık katsayılarının değışinti gelişimleri ve fazladan bir ölçek ekleme



Şekil 5.8 (a) Sentezlenen geniş bantlı ses sinyali, (b) Üretilen kalıntı sinyali

5.3 Öz-benzeşim Parametresinin Ses İletiminde Kullanılması

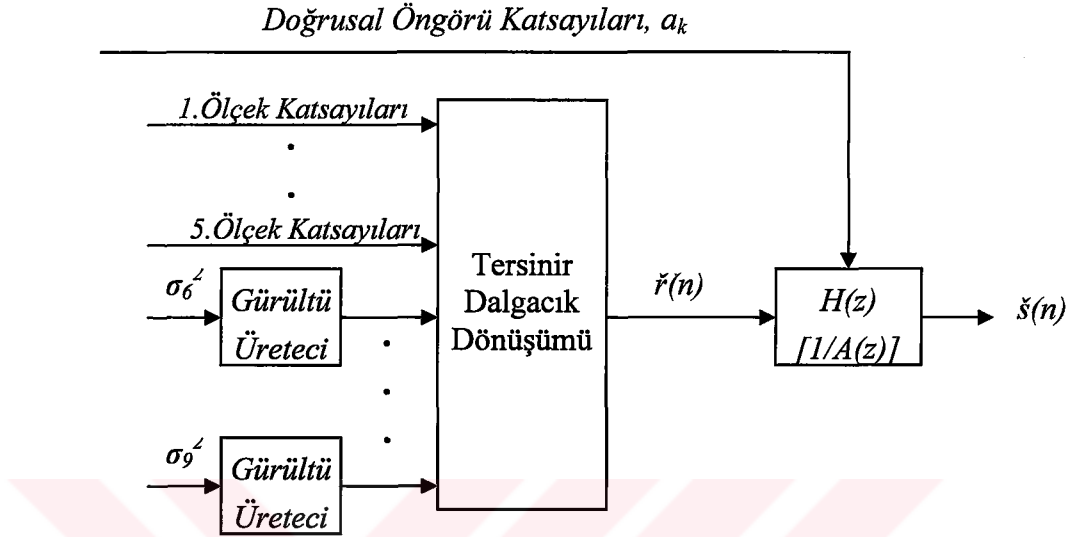
Sesin doğallığının ana etmeni, uyarıcı sinyalin rasgele fraktallı yapısında gizlidir [17]. Bu fraktallı yapı, Bölüm 4'te anlatıldığı şekilde öz-benzeşim parametresi elde edilerek modellenenir. Konuşma sesi sinyalinin doğallığının bir ölçütü olan öz-benzeşim parametresi ses iletimi uygulamalarında sesin doğala yakın olarak sentezlenmesinde kullanılabilir [17,20].

Düşük bit hızlarında ses iletimi gerçekleştiren “Darbe-Uyarımlı Doğrusal Öngörüm Kodlaması” yönteminde, verici bölümünde doğrusal öngörü analizi yapılarak a_k öngörü katsayıları ve perde periyodu elde edilir. Alıcı tarafa a_k katsayıları, ötümlü/ötümsüz kararı ve perde periyodu bilgisi ulaştırılır. Alıcıda, ötümsüz seslerin sentezi sırasında uyarıcı sinyal olarak beyaz Gauss gürültüsü kullanılır. Ötümlü seslerin sentezinde ise uyarıcı sinyal olarak darbe dizisi kullanılır. Darbe dizisinin periyodu sesin perde periyoduna eşit seçilir. Uyarıcı sinyaller, a_k parametrelerinden oluşturulan sistemden geçirilerek ses üretimi gerçekleştirilir. Kullanılan uyarıcı sinyal, kalıntı sinyalinin gayet basitleştirilmiş halidir ve rasgele fraktallı yapıya sahip değildir. Bu yüzden, üretilen ses sinyali doğaldan uzak, mekanik ve vızıltılı bir ses olur. Ancak uyarıcı sinyal olarak darbe dizisi yerine, γ parametresi önceden verici tarafında belirlenerek alıcıya gönderilmiş, bir rasgele fraktallı sinyal kullanılırsa, doğala daha yakın ses üretimi mümkün olacaktır [20]. Bir diğer öngörüm kodlaması olan “Kod-uyarımlı Doğrusal Öngörüm (Code-Excited Linear Prediction, CELP)” kodlamasında ses sinyallerinin olabildiğince parametrik bir şekilde ifade edilmesi istenir. CELP’te uyarıcı sinyalin özellikleri bir vektöre atanır ve bu vektör iletilir. Sesin doğallığı da bir parametre ile nitelenip CELP’te kullanılan kod vektörünün bir elemanı yapılırsa, vericide üretilecek sesin doğala yakınlığı sağlanabilir [20].

Son olarak, “Kalıntı Uyarımlı Doğrusal Öngörüm (Residual Excited Linear Prediction, RELP) kodlamasında, tersinir süzgeçleme yardımıyla elde edilen kalıntı sinyalinin tamamı yada, çoğu uygulamada olduğu gibi, küçük bir kısmı, karşı tarafa iletilerek ses sentezinde uyarıcı sinyal olarak kullanılır [2]. Ses kalıntı sinyali perde periyodu bilgisini içerisinde barındırır. Dolayısıyla alıcı tarafa perde periyodu bilgisi ayrı olarak gönderilmez. Kalıntı sinyalinde ses bilgisinin çoğu 1000 Hz’in altındaki frekans bölgelerindedir. Bu yüzden, ses kalıntı sinyali 1kHz kesim frekansına sahip alçak geçiren süzgeçten geçirilerek iletilir. Böylelikle bit iletim verimliliği sağlanır.

Kalıntı sinyalinin iletilen küçük kısmında yüksek frekans bileşenleri bulunmamaktadır. Bu bileşenler çeşitli yöntemlerle tekrar üretilmeye çalışılır [2].

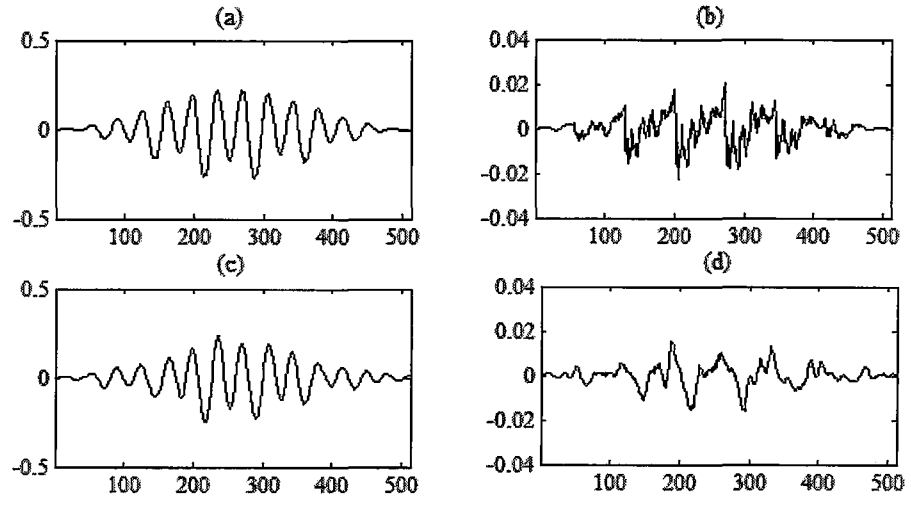
Dalgacık tabanlı analiz de RELP kodlayıcılarda kullanılabilir. Kalıntı sinyalinin kendisi yerine dalgacık katsayıları ve öz-benzeşim parametresinin iletilmesi verimliliği artıracaktır.



Şekil 5.9 Ses sentezi blok çizeneği

İlk olarak ses kalıntı sinyalinin dalgacık katsayıları ve γ parametresi elde edilir. Küçük ölçeklerdeki dalgacık katsayıları ve γ değeri gönderilerek alıcı tarafta uyarıcı sinyali tekrar üretmek mümkün olabilir. Ses sentez modeli Şekil 5.9'da gösterilmiştir. 1.ölçek ile 5. ölçek katsayıları aynen gönderilir. 6.-9. ölçek katsayılarının değışintileri kestirilir. Bu değışintilere sahip sıfır ortalamalı Gauss gürültüleri ölçeklerin katsayıları olarak kullanılırlar. Tersinir dalgacık dönüşümü ile ses kalıntı sinyali elde edilir. Son olarak bu sinyal, öngörü parametreleri yardımıyla oluşturulan $H(z)$ süzgecinden geçirilerek ses sinyali sentezlenir.

Şekil 5.10 (a) ve (b)'de 16 kHz ile örneklenmiş, 512 örnekli /UW/ ötümlü sesi ve bu sesin kalıntı sinyali gösterilmektedir. (c) ve (d)'de ise yukarıda anlatıldığı şekilde sentezlenen ses sinyali ve kalıntı sinyali verilmiştir.



Şekil 5.10 (a) 512 örneklilik /UW/ ötümlü sesi, (b) (a)'daki sinyalin kalıntısı, (c) Sentezlenen ses sinyali, (d) Sentezlenen kalıntı sinyali

6. SONUÇLAR ve ÖNERİLEN ÇALIŞMALAR

6.1 Sonuçlar

Bu çalışmada ses kalıntı sinyalinin rasgele fraktallı yapısı araştırılmış ve $1/f$ süreç özelliği dalgacık tabanlı analiz ile sınanmıştır. İlk olarak kalıntı sinyalinin çıkarım aşamaları anlatılmıştır. Ses kalıntı sinyaline dalgacık dönüşümü uygulanarak dalgacık katsayıları elde edilmiştir. Her bir ölçekteki dalgacık katsayılarının sıfır ortalamalı, yaklaşık Gauss dağılımlı ve zayıf ilintili oldukları, değişimlerinin logaritmalarının bir doğru üzerinde sıralandığı gözlenmiştir.

Dalgacık katsayılarının değişim gelişimi yardımıyla elde edilen öz-benzeşim parametresi, ses kalıntı sinyalinin fraktallı yapısını modellemekte kullanılabilir. Bu parametrenin çeşitli uygulamalarda bir öznitelik olarak kullanılabileceği önerilmiştir. Örneğin düşük bantlı ses iletiminde, ses kalitesi artırımında, doğala yakın ses sentezinde kullanılabilir. Sesin fraktallı yapısının, dolayısıyla doğallığının bir parametreyle ifade edilmesi sesin modellenmesinde önemli bir gelişmedir.

Dalgacık tabanlı analizin RELP kodlayıcılarda kullanılabileceği bir örnek gösterilmiştir. Ses kalıntı sinyalinin tamamının iletimi yerine dalgacık katsayıları ve öz-benzeşim parametresinin iletilmesi verimliliği artıracaktır. Daha az sayıda olan küçük ölçek katsayıları aynen gönderilir. Fazla sayıdaki yüksek ölçek katsayılarının değişimleri gönderilip, bu değişimlere sahip sıfır ortalamalı Gauss gürültüleri ölçeklerin katsayıları olarak kullanılırlar. Elde edilen katsayılar yardımıyla uyarıcı sinyal ve ardından ses sinyali sentezlenir. Böylelikle kodlama verimliliği sağlanmış olur.

Öz-benzeşim parametresinin, kadın ve erkek seslerinde farklı olup olmadığı araştırılmıştır. /IH/, /EI/, /EH/, /AE/ ötümlü seslerinin kalıntılarından elde edilen γ değerleri kadın ve erkek seslerinde daha fazla farklılık gösterirken bazı seslerde (/ER/, /AW/, /UW/) ise γ değerleri birbirinden ayrılmayacak şekilde yakın çıkmışlardır. Öz-benzeşim parametresi ile sesin yumuşaklığı/pürüzlülüğü arasında bir bağlantı kurulmuştur. Büyük γ değerlerine sahip seslerin, küçük γ değerlerine

sahip seslerden daha yumuşak ve pürüzsüz olduğu ve genel olarak, kadın seslerine ait γ değerlerinin erkek seslerine ait γ değerlerinden daha yüksek çıktığı gözlenmiştir.

6.2 Önerilen Çalışmalar

- Öz-benzeşim parametresi kestirim yöntemi, hasta ve sağlıklı insanlar tarafından üretilen seslere uygulanarak, parametrelerin farklılık gösterip göstermediği incelenebilir. Ses tellerindeki rahatsızlığın γ parametresine etkisi araştırılabilir.
- γ parametresi, bilinen kadın/erkek sesi ayırımı teknikleriyle kombine edilerek daha etkin bir sınıflandırma tekniği geliştirilebilir.
- Ses kalıntı sinyalinin fraktallı $1/f$ yapısı, ses sinyalinin doğallığının bir ölçütüdür. Ses sentezinde bu göz önüne alınarak ses üretimi yapılırsa mekaniklikten uzak, doğala daha yakın ses üretimi mümkün olabilir. Perde periyodu bilgisini içinde barındıran öz benzeşimli kalıntı sinyalleri üretilerek ses sentezi gerçekleştirilebilir.
- Yazılı metinden konuşma sentezleme uygulamalarında, uyarıcı kalıntı sinyalin fraktallı yapısı göz önüne alınarak doğala yakın ses üretimi gerçekleştirilebilir.
- Ses sentezi sadece 512 örneklik küçük ses parçacıklarında sınanmıştır. Bu işlemler çok daha uzun konuşma sinyallerine uygulanarak kalite artırımının başarımı ve kodlama verimliliği ölçülebilir.

KAYNAKLAR

- [1] **Quateri, T., F.**, 2002. Discrete-Time Speech Signal Processing Principles and Practice, Prentice Hall Inc.
- [2] **Deller, J. R., Hansen, J. H. I., Proakis, J. G.**, 2000. Discrete-Time Processing of Speech Signals (Second edition), IEEE Press, New York, NY.
- [3] **Gold, B., Morgan, N.**, 2000. Speech And Audio Signal Processing, Processing and Perception of Speech and Music, John Wiley&Sons Inc.,New York, NY.
- [4] **Manolakis, D.G., Ingle V.K., Kogon S.M.**, 2000. Statistical and Adaptive Signal Processing, McGraw Hill, Boston.
- [5] **Kaiser, J. F., Mitra, S. K.**, 1993. Handbook for digital signal processing, John Wiley&Sons Inc., New York, NY.
- [6] **Wornell, G. W.**, 1995. Signal Processing With Fractals:A Wavelet-Based Approach, Prentice Hall, NJ.
- [7] **Koser, R. W., Onaral, B.**, 1996. 1/f Nature of Speech: Implications for Enhancement and Coding, *MS Thesis*, Drexel University, PA.
- [8] **Lee, Y.**, 1998. Speech Quality Enhancement by Exploring 1/f Nature of Speech Residual, *MS Thesis*, Drexel University, PA.
- [9] **Özkurt, T. E., Akgül, T.**, 2003. Analysis of 1/f Nature of Sea-Wave Noise Data, *3rd WSEAS International Conference on Signal Processing, Computational Geometry& Artificial Vision*, Greece, November 14-18.
- [10] **Musha, T., Gabor, B., Shoji, M.**, 1990. 1/f photon-number fluctuations in quartz observed by light scattering, *Physics Review Letters*, Vol. 64, No. 20, 2394.
- [11] **Voss, R. F.**, 1992. Evolution of long-range fractal correlations and 1/f noise in DNA base sequences, *Physics Review Letters*, Vol. 68, No. 25, 3805-3808.
- [12] **Li, W., Marr, T. G., Kaneko, K.**, 1994. Understanding long-range correlations in DNA sequences, *Physica D*, Vol. 75, 392-416.
- [13] **Ninnes, B.**, 1998. Estimation of 1/f Noise, *IEEE Transactions on Information Theory*, Vol. 44, No. 1, 32-46.
- [14] **Mandelbrot, B. B., Van Ness, J. W.**, 1968. Fractional Brownian Motions, Fractional Noise and Applications, *SIAM Review*, Vol. 10, No. 4, 422-436.
- [15] **Gonçalves, P. Abry, P. ,** 1995. Wavelets, Spectrum Analysis and 1/f Processes, *Lecture Notes in statistics (Edtd. by Antoniadis)*, 103, 15-29.

- [16] Dieker, T., 2002. Simulation of fractional Brownian motion, *MS Thesis*, Vrije University, Amsterdam.
- [17] Aoki, N., Ifukube, T., 2001. Enhancing the Naturalness of Synthesized Speech by Using the Random Fractalness of Vowel Source Signals, *Electronics and Communications in Japan*, Vol. 84, No. 1, 11-20.
- [18] Chan, Y.T., 1995, Wavelet Basics, Kluwer Academic Publishers, Norwell, Massachusetts.
- [19] Mallat, S., 1998. A wavelet Tour os Signal Processing (Second edition), Academic Pres, CA.
- [20] Maragos, P., Young, K. L., 1990. Fractal Excitation Signals for CELP Speech Coders, *International Conference on Acoustics, Speech, and Signal Processing*, Vol. 2, 669-672.



ÖZGEÇMİŞ

Süleyman BAYKUT, 1980 yılında Tekirdağ'da doğdu. İlk, orta ve lise eğitimini aynı şehirde tamamladı. 1998 yılında İstanbul Üniversitesi Elektrik-Elektronik Mühendisliği bölümünde lisans eğitimine başladı. 2002 yılında İstanbul Üniversitesi'nden mezun oldu ve aynı yıl İstanbul Teknik Üniversitesi Telekomünikasyon Mühendisliği Bölümü'nde yüksek lisans eğitimine başladı. Haziran 2004'te İstanbul Teknik Üniversitesi Elektrik-Elektronik Fakültesi Elektronik ve Haberleşme Bölümü, Telekomünikasyon Anabilim Dalı'nda araştırma görevlisi olarak göreve başladı. Halen aynı bölümde araştırma görevlisi olarak çalışmaktadır.

