

**DAR BANTLI KONUŐMA KALİTESİNİN İZGEL ZARF  
GENİŐLETİLEREK İYİLEŐTİRİLMESİ**

**QUALITY ENHANCEMENT OF NARROWBAND SPEECH BY  
SPECTRAL ENVELOPE EXTENSION**

**GÜRHAN BULU**

Hacettepe Üniversitesi  
Lisansüstü Eğitim-Öğretim ve Sınav Yönetmeliğinin  
ELEKTRİK ve ELEKTRONİK Mühendisliği Anabilim Dalı İçin Öngördüğü  
DOKTORA TEZİ  
olarak hazırlanmıştır.

2008

Fen Bilimleri Enstitüsü Müdürlüğü'ne,

Bu çalışma jürimiz tarafından ELEKTRİK ve ELEKTRONİK Mühendisliği ANABİLİM DALI'nda DOKTORA TEZİ olarak kabul edilmiştir.

Başkan :  
Prof.Dr. A. Salim Kayhan

Üye (Danışman) :  
Yrd.Doç.Dr. A. Semih Bingöl

Üye :  
Doç.Dr. H. Gökhan İlk

Üye :  
Yrd.Doç.Dr. Mücahit Üner

Üye :  
Yrd.Doç.Dr. Atila Yılmaz

ONAY

Bu tez ....../....../2008 tarihinde Enstitü Yönetim Kurulunca kabul edilmiştir.

....../....../2008

Prof.Dr. Erdem Yazgan  
FEN BİLİMLERİ ENSTİTÜSÜ MÜDÜRÜ

# DAR BANTLI KONUŞMA KALİTESİNİN İZGEL ZARF GENİŞLETİLEREK İYİLEŞTİRİLMESİ

GÜRHAN BULU

ÖZ

İnsan sesinin 7–8 kHz'e kadar izgel bileşenlerinin olduğu bilinmektedir. Öte yandan, telefon sistemlerinde insan sesi genellikle 3.4 kHz'lik bir alçak geçirgen süzgeçten geçirilip 8 kHz'te örneklenir. Kodlama ve benzeri işlemlerden kaynaklanan hiçbir kayıp olmasa bile yüksek frekans bölgesinin süzülmesinden kaynaklanan bir kalite kaybı söz konusudur. Bu kaybın düzeyi anlaşılabilirliği pek etkilememekle beraber konuşma kalitesinde hissedilir bir bozulmaya yol açmaktadır. Bu çalışmada izgel zarfın düşük frekanslı bölgelerinden faydalanılarak, süzülen yüksek frekanslı bölgeler elde edilmeye çalışılmış ve izgel zarfın genişletilmesi olarak da nitelenebilecek bu işlem için başvuru çizelgesi ve yapay sinir ağları yöntemleri kullanılmıştır.

Daha sonra genişletilen bu izgel zarfın yüksek frekanslı bölgelerinin altı, değişik yöntemler kullanılarak doldurularak konuşma kalitesi artırılmaya çalışılmıştır. Bu yöntemlerin en basiti izgel zarfın altının gürültüyle doldurulmasıdır. Bir diğer yöntem ise izgel zarfın altının sinüslerle doldurulmasıdır. Sonuncu ve en başarılı yöntemde ise kaynak-süzgeç modeliyle uyarım sinyalinin de genişletilerek konuşma kalitesinin artırılmasıdır. Bu geliştirilen yöntemlerin telefon kalitesindeki konuşmalar için kullanılacağı öngörülmekle birlikte, daha düşük bant genişliğine sahip konuşmalar (örneğin 2 kHz) için de kullanılabilecekleri gösterilmiştir.

**Anahtar Kelimeler:** İzgel zarf genişletme, izgel zarf dışdeğerleme, dar bantlı konuşma, konuşma kalitesi iyileştirme, vektör nicemleme, başvuru çizelgesi, sinir ağları.

Danışman: Yrd.Doç.Dr. A. Semih Bingöl, Hacettepe Üniversitesi, Elektrik ve Elektronik Mühendisliği Bölümü

# QUALITY ENHANCEMENT OF NARROWBAND SPEECH BY SPECTRAL ENVELOPE EXTENSION

GÜRHAN BULU

## ABSTRACT

It is known that human speech has components up to 7-8 kHz. However in telephony, speech is usually lowpass filtered at 3.4 kHz and sampled at 8 kHz. Although this filtering operation has little effect on intelligibility, it causes a perceptible degradation in speech quality. This paper deals with the reconstruction of the lost high frequency content of telephone speech using the low frequency spectral information. Reconstruction methods based on a lookup table and artificial neural networks are used to obtain the high frequency region. This extended high frequency region is used for enhancement of speech quality in various methods. One of them is filling the high region with random noise, the other one is filling the high region with sinusoids, the last and the most effective method is to fill the high region by using source excitation model. These methods are developed to enhance the telephone quality speech, however they may be used for speech signals which have lower bandwidths (2 kHz).

**Keywords:** Spectral envelope extension, spectral envelope extrapolation, narrow-band speech, speech quality enhancement, vector quantization, look-up table, neural networks.

Advisor: Yrd.Doç.Dr. A. Semih Bingöl, Hacettepe University, Department of Electrical and Electronics Engineering

## TEŐEKKÜR

Tez alıőması sırasında karőılaőılan sorunların özümünde yardımcı olan ve alıőmanın sađlıklı bir őekilde yürütölmesi için yol gösteren tez danıőmanım Yrd. Do. Dr. A. Semih Bingöl'e,

Bilgi ve deneyimlerini paylaşan deđerli Tez İzleme Komitesi ve Tez Savunma Sınavı jüri üyelerine,

Tez yazımı sırasında formatın hazırlanması konusunda yardımlarından dolayı Dr. ađatay Yavuzıılmaz'a,

Her türlü maddi ve manevi desteklerini esirgemeyen anneme, babama ve kardeőime,

alıőma sırasında gösterdikleri anlayıőtan ötürü sevgili eőim Ayően'e ve kızım Elif Servi'ye,

itenlikle teőekkür ederim.

## İÇİNDEKİLER DİZİNİ

	<u>Sayfa</u>
ÖZ .....	i
ABSTRACT .....	ii
TEŞEKKÜR .....	iii
İÇİNDEKİLER DİZİNİ .....	iv
ŞEKİLLER DİZİNİ .....	vi
ÇİZELGELER DİZİNİ .....	viii
SİMGELER VE KISALTMALAR DİZİNİ .....	ix
1. GİRİŞ .....	1
2. KONUŞMANIN OLUŞUMU VE MODELLENMESİ .....	5
2.1. Konuşmanın Oluşumu .....	5
2.2. Konuşmanın Kaynak-Süzgeç Modellenmesi .....	6
2.3. Konuşmanın Sinüs Modellenmesi .....	8
2.4. Perde Frekansının Bulunması .....	10
2.5. İzgel Zarfın Modellenmesi .....	11
2.6. ÇİF Gösterimi ve Özellikleri .....	13
3. TABLO YÖNTEMİ KULLANARAK İZGEL ZARFIN GENİŞLETİLMESİ .....	19
3.1. Vektör Nicemleme .....	20
3.2. Vektör Nicemleyici Kullanarak Başvuru Çizelgesi Oluşturulması .....	21
3.3. İzgel Zarfın Genişletilmesi .....	25
4. YAPAY SİNİR AĞLARIYLA İZGEL ZARFIN GENİŞLETİLMESİ .....	27
4.1. Yapay Sinir Ağları .....	27
4.2. Çok Katmanlı Sinir Ağları .....	30
4.3. Geri Yayılım Algoritması .....	31
4.3.1. Çıkış Katmanındaki Türevin Hesaplanması .....	35
4.3.2. Gizli Katmanındaki Türevin Hesaplanması .....	36
4.4. Telefon Kalitesindeki Konuşmanın İzgel Zarfının Genişletilmesi .....	37
4.5. 4. kHz İle Örneklenmiş Konuşmanın İzgel Zarfının Genişletilmesi .....	41

5. KONUŞMA KALİTESİNİN ARTIRILMASI .....	43
5.1. İzgel Zarfın Gürültüyle Doldurulması .....	43
5.2. İzgel Zarfın Sinüslerle Doldurulması .....	43
5.3. Kaynak-Süzgeç Modeliyle Kalitenin Artırılması .....	45
5.3.1. Uyarım Sinyalinin Genişletilmesi .....	45
5.3.1.1. Uyarım Sinyalinin Harmonik Sinüslerle Genişletilmesi	47
5.3.1.2. Uyarım Sinyalinin Sinüslerle Genişletilmesi .....	48
5.4. Geniş Bantlı Konuşma İçin Kazancın Bulunması .....	49
5.5. Öznel Testler .....	51
5.5.1. Ortalama Kanı Puan Testleri .....	52
5.5.2. Karşılaştırmalı Testler .....	55
6. SONUÇ .....	56
KAYNAKLAR DİZİNİ .....	58
EKLER DİZİNİ .....	62
ÖZGEÇMİŞ .....	71

## ŞEKİLLER DİZİNİ

	<u>Sayfa</u>
Şekil 2.1. Ses sistemi .....	5
Şekil 2.2. Ses oluşum modeli .....	6
Şekil 2.3. Konuşma ve gerçel kepstrumu .....	12
Şekil 2.4. ÇİF'lerin İzgel zarfta dağılımları .....	14
Şekil 2.5. Çizgi izgel zarfların değişimiyle izgel güç yoğunluğunun değişimi ...	15
Şekil 2.6. 8 kHz ile örneklenmiş sesli konuşma parçası ve özellikleri .....	17
Şekil 2.7. 8 kHz ile örneklenmiş sessiz konuşma parçası ve özellikleri .....	18
Şekil 3.1. Başvuru çizelgesi .....	19
Şekil 3.2. 7 kHz ile bant sınırlı orijinal ve 3.4 kHz ile bant sınırlı izgel zarftan tablo yöntemiyle 7 kHz'e genişletilen izgel zarflar .....	26
Şekil 4.1. Biyolojik nöron .....	28
Şekil 4.2. Nöron modeli .....	29
Şekil 4.3. Çok katmanlı YSA .....	30
Şekil 4.4. Çok katmanlı sinir ağları notasyonu .....	32
Şekil 4.5. Çok Katmanlı YSA .....	38
Şekil 4.6. Konuşma parametrelerinin sinir ağları kullanarak genişletilmesi ....	39
Şekil 4.7. İzgel bozulunun alt sınır eğrisi .....	39
Şekil 4.8. 7 kHz'e bant sınırlı orijinal ve 3.4 kHz'e bant sınırlı izgel zarftan YSA kullanılarak genişletilen izgel zarflar .....	40
Şekil 4.9. 3.4 kHz'e bant sınırlı orijinal ve 1.8 kHz'e bant sınırlı izgel zarftan YSA kullanılarak genişletilen izgel zarflar .....	42
Şekil 5.1. Dar bantlı konuşmanın gürültüyle genişletilmesi .....	44
Şekil 5.2. Dar bantlı konuşmanın periyodik sinüslerle genişletilmesi .....	44
Şekil 5.3. Dar bantlı konuşmanın kaynak-süzgeç modeliyle genişletilmesi ....	45
Şekil 5.4. 16 kHz ile örneklenmiş konuşma parçaları .....	46
Şekil 5.5. 16 kHz ile örneklenmiş konuşma ve izgegrafiği .....	47
Şekil 5.6. Dar bantlı uyarım sinyalinin sinüslerle genişletilmesi .....	49



Şekil 5.7. 16 <i>kHz</i> ile örneklenmiş konuşma için orijinal ve genişletilen uyarım sinyalinin yüksek frekanslı bölgeleri . . . . .	50
Şekil 5.8. 16 <i>kHz</i> ile örneklenmiş konuşma için orijinal ve genişletilmiş uyarım sinyallerinin izgegrafikleri . . . . .	51
Şekil 5.9. 16 <i>kHz</i> ile örneklenmiş konuşma için orijinal ve kestirilen kazançlar .	52

## ÇİZELGELER DİZİNİ

	<u>Sayfa</u>
Çizelge 5.1. Ortalama kanı puanlama cetveli .....	52
Çizelge 5.2. Ortalama kanı puanlaması için kullanılan konuşma parçalarının elde edilış yöntemleri ve örnekleme frekansları .....	53
Çizelge 5.3. Test konuşmaları için OKP sonuçları .....	54
Çizelge 5.4. 15 dinleyici için için OKP sonuçları .....	54

## SİMGELER VE KISALTMALAR DİZİNİ

TK	: Tüm Kutup
ÇİF	: Çizgi İzgel Frekans
BÇ	: Başvuru Çizelgesi
SA	: Sinir Ağları
LBG	: Linde-Buzo-Gray
$H(z)$	: Konuşma yolu izgel zarf süzgeci
$A(z)$	: Ters süzgeç
$P(z), Q(z)$	: ÇİF polinomları
$s(t), s[n]$	: Konuşma sinyali
$e(t), u(t)$	: Uyarım sinyali
$F_s$	: Örnekleme frekansı
$\omega_\ell^k$	: $k$ . çerçevdeki $\ell$ . sinüsün frekansı
$A_\ell^k$	: $k$ . çerçevdeki $\ell$ . sinüsün genliği
$\theta_\ell^k$	: $k$ . çerçevdeki $\ell$ . sinüsün evresi
$c(n)$	: Bir sinyalin gerçel (real) kepstrumu
$X(e^{j\omega})$	: $x[n]$ sinyalinin AZFD'si (Ayrık-Zaman Fourier Dönüşümü)
$T_0, f_0$	: Perde periodu ve perde frekansı
$\hat{\mathbf{A}}$	: Vektör nicemleyici için yeniden oluşum alfabesi (çizelgesi)
$\hat{\mathbf{x}} = q(\mathbf{x})$	: $q$ nicemleyicisi için yeniden oluşum vektörü
$\varepsilon(n)$	: Sinir ağlarında çıkıştaki karesel hata
$\varepsilon_{avg}$	: Sinir ağlarında çıkıştaki ortalama karesel hata
$m_I, m_L$ ve $m_0$	: $I$ ., çıkış ve giriş katmanlarındaki nöron sayıları
$w_{l, j, i}$	: $l$ . katmandaki $j$ . nörona, $(l - 1)$ . katmandaki $i$ . nörondan gelen ağırlık
$D^k$	: $k$ . çerçevdeki izgel bozulum
$D_M$	: İki izge arasındaki kayma çıkarıldıktan sonraki izgel bozulum
$G_c$	: İki izge arasındaki kayma
$a_i$	: Doğrusal öngörülü kodlama parametreleri
$c_i$	: Kepstral katsayılar

$y_i$  : Ağırlıklandırılmış kestral katsayılar  
 $h(y)$  : Entropi  
 $I(x; y)$  : Karşılıklı bilgi miktarı

## 1. GİRİŞ

Konuşma insanlar tarafından kullanılan vazgeçilmez bir iletişim aracıdır. Günümüzde, konuşma kodlama ve iletiminde pek çok yöntem kullanılmaktadır. Bunlardan en temel ve basit olan ve tüm telefon sistemlerinde de kullanılan yöntemde, konuşma yaklaşık 3.4 kHz'lik bir alçak geçirgen süzgeçten (örtüşme önler süzgeç) geçirildikten sonra 8 kHz'de örneklenir. En temel kodlama yönteminde her bir örnek 8 bit ile nicemlenerek 64 kbit/s veri hızı elde edilmiş olur. Bu parametrelerle elde edilen konuşma orta-yüksek kalitede olmakta ve "telefon kalitesi" adı ile anılmaktadır. Günümüzde bant genişliğinden tasarruf etmek için konuşma daha düşük bit hızlarında iletilmektedir. Daha düşük bit hızları sağlamak için son 30 yılda basitten karmaşığa onlarca konuşma kodlayıcı geliştirilmiştir. 64 kbit/s veri hızındaki orijinal konuşmaya yakın kalite günümüzde 2.4 kbit/s civarında hızlarla elde edilebilmektedir. Ancak "telefon kalitesi" konuşma kodlayıcılarında 3.4 kHz bant genişliği ve 8 kHz örnekleme hızı değişmeden kalmıştır.

Konuşma, ancak 7 kHz'lik bir alçak geçirgen süzgeçten geçirilerek 16 kHz'le örneklenip 13 – 16 bit ile nicemlendiğinde yüksek kaliteli sayılabilir. (ETSI EG 201 377-1). Dolayısıyla 3.4 kHz'e bant sınırlanmış (orta-yüksek kalitede) konuşmada 3.4 – 7 kHz arasında kalan kısım yok farzedilmektedir. Yok sayılan bu kısım konuşmanın anlaşılabilirliğini pek fazla etkilememekte fakat kalitede bir takım kayıplara yol açmaktadır. Literatürde bu kaybın konuşmanın ne kadarına karşılık geldiği pek araştırılmamakla beraber bazı kaynaklarda %10-20 arasında bir kayıptan söz edilmektedir (O'Shaughnessy, 1987).

Literatürde konuşma kalitesini artırmaya yönelik birçok çalışma bulunmakla birlikte bunların hemen hemen tamamı gürültü azaltma, mikrofon karakteristiği düzeltme vb. konulara yöneliktir. Alçak geçirgen süzgeçten dolayı meydana gelen kalite kaybına odaklanan pek fazla çalışma yoktur. Olan çalışmalar ise başlıca iki ana grup altında toplanabilir.

Birinci gruptaki çalışmalarda geniş bantlı konuşma ya frekans ekseninde sıkıştırılır yada yüksek frekans bileşenlerine ait bir takım yan bilgiler (side information) dar bantlı konuşma ile beraber gönderilir. Bu gruba örnek çalışmalardan bir tanesi yüksek bant genişliğine sahip olan konuşmanın frekans ekseninde sıkıştırılıp daha düşük bant genişliğinde (3.4 kHz) konuşmanın elde edilmesi ve elde edilen sıkıştırıl-

mış konuşmanın klasik ses iletim ve kodlama teknikleriyle kodlanmasıdır (Patrick *et al.*, 1983). Bir diğer çalışmada ise yüksek frekansa ait çok kısıtlı bilginin (150 b/s) dar bantlı konuşmayla beraber gönderilerek konuşma kalitesinin artırılmasıdır (Epps and Holmes, 2001). Bu gruna örnek verilebilecek son çalışmada ise yüksek frekans bileşenlerine ait yan bilgilerin sıkıştırılıp gizli damgalama (watermarking) teknikleriyle dar bantlı konuşma içine gömülerek karşı tarafa iletilmesi ve uygun alıcılarda bu bilgiler kullanılarak yüksek kaliteli konuşmanın elde edilmesidir (Geiser and Vary, 2007).

İkinci gruptaki çalışmalarda ise dar bantlı konuşma iletim kanalında herhangi bir değişikliğe uğramadan ve hiçbir ek bilgiye sahip olmadan iletilir. Bu çalışmanın da dahil olduğu bu başlık altındaki çalışmalarda konuşma kalitesi artırma işlemleri alıcı tarafında ve konuşmanın özellikleri kullanılarak yapılır. Esas olarak bu çalışmalar geniş bantlı konuşmanın izgel zarfının elde edilmesi ve elde edilen bu zarf kullanılarak konuşmanın kayıp bölgelerinin sentezlenmesi üzerine kurulmuştur. Bu tür çalışmalara ilk örnekler izgel zarf genişletmede tablo yöntemini kullanan çalışmalardır (Carl and Heute, 1994), (Yoshida and Abe, 1994). Bu çalışmalarda izgel zarfın altı kaynak-süzgeç yöntemiyle doldurulmuştur. Daha sonraki çalışmalarda ise yüksek frekanslı bölgenin izgel zarfının altı çok bantlı uyarım (multiband excitation) (Chan and Hui, 1996) ve sinüs yöntemiyle doldurulmuştur (Epps and Holmes, 1998). Tablo yöntemiyle izgel zarf genişletmedeki bir sonraki aşama ise sesli ve sessiz konuşma bölgeleri için iki farklı tablonun kullanıldığı çalışmadır (Epps and Holmes, 1999). Daha sonra ardarda gelen konuşma parçalarının birbirine bağlı olduğundan yola çıkılarak Saklı Markov Modeli (Hidden Markov Model) tablo yöntemiyle birlikte kullanılmaya başlanmıştır (Jax and Vary, 2000), (Jax, 2004), (Yao and Chan, 2005), (Rohde and Vedstesen, 2007). Bu yöntemlerin dışında ayrıca konuşma izgel zarfının konuşmanın özellikleri (perde, formant vb.) kullanılarak genişletildiği (Gustafsson *et al.*, 2001), (Gustafsson *et al.*, 2006); geniş bantlı konuşma parametrelerinin dar bantlı parametrelerden bir dönüşüm matrisi kullanılarak elde edildiği (Chennoukh *et al.*, 2001) ve uyarım sinyalinin harmonikliğini parametrik olarak modelleyip konuşmanın bant genişliğini artıran çalışmalar da literatürde mevcuttur (Vaseghi *et al.*, 2006). Ancak bu çalışmaların çoğunda özellikle tablo yöntemini kullanan çalışmalarda, kullanılan yöntemlerin detayları hakkında yeterli bilgi mevcut değildir. Dolayısıyla bu çalışmada kullanılan yöntemlerin büyük bir bölümü sıfırdan geliştirilmiştir.

Bu çalışmada izgel zarfın örtüşme önler süzgeç tarafından atılan yüksek frekans bölgesinin yine aynı süzgeç tarafından geçirilen düşük frekans bölgesinden yararlanılarak elde edilmesi amaçlanmaktadır. Konuşmanın izgel zarfını oluşturan ses yolunun kabaca 250000-500000 ( $2^{18} - 2^{19}$ ) kadar farklı konfigürasyona sahip olduğu literatürde belirtilmektedir (Lahouti *et al.*, 2006). İzgel zarfın 3.4 kHz üstü bölgesi de bu sonlu sayıda ses yolu konfigürasyonu ile üretildiğinden bu bölgenin izgel zarfını düşük frekans bölgesinin izgel zarfından elde etmek mümkündür. Hatta yüksek frekans bölgesinin anlaşılabilirliğe katkısı çok daha az olduğundan bu bölge için daha az sayıda farklı konfigürasyon kullanılabilir. İzgel zarf dışdeğerleme yada genişletme adı verilen bu yöntemle kalite artırma işlemi başlıca iki ana çalışma alanı altında toplanabilir. Birinci alan yüksek frekanslı izgel zarfın düşük frekanslı izgel zarftan nasıl elde edileceği; ikincisi ise elde edilen yüksek frekans zarfının içinin nasıl ve hangi yöntemlerle doldurulacağıdır.

Bu çalışmada konuşmanın izgel zarfının genişletilmesi için iki farklı yöntem kullanılmıştır. Bu yöntemler sırasıyla yüksek frekanslı izgel zarfın

- Başvuru Çizelgesi (Look-Up Table)
- Sinir Ağları (Neural Networks)

kullanılarak elde edilmesidir. Genişletilen bu izgel zarfın altında kalan bölgenin doldurulması için ise üç yöntem kullanılmıştır. Bu yöntemler sırasıyla izgel zarfın yüksek frekans bölgesinin

- Gürültüyle
- Sinüslerle
- Kaynak-süzgeç yöntemiyle

doldurulması olarak adlandırılabilirler.

Bu tez çalışması altı bölümden oluşmaktadır. Bölüm 2'de konuşmanın oluşumu ve modellenmesi ile ilgili kısa bilgi verilmiştir. Ayrıca izgel zarfın genişletilmesinde kullanılacak konuşma parametrelerinin elde edilmesi ve bu parametrelerin işlem yapmaya uygun gösterim biçimleri tartışılmıştır. Bölüm 3'te izgel zarfın tablo yöntemi

(look-up table) kullanılarak genişletilmesi üzerinde durulmuştur. Bu yöntemde girişte dar bant ve çıkışta geniş bant parametreleri için iki farklı tablo kullanılmaktadır. Ancak bu tablolar eş zamanlı ve birbirine bağlı biçimde LGB (Linde-Buzo-Gray) vektör nicemleme (vector quantization) algoritması kullanılarak oluşturulmuştur (Linde *et al.*, 1980), (Paliwal and Atal, 1993). Bölüm 4'te izgel zarf genişletme işlemi, bu çalışmadaki biçimiyle daha önce kullanılmayan yapay sinir ağları kullanılarak gerçekleştirilmeye çalışılmıştır. Bu çalışmada kullanılan sinir ağları çok karmaşık bir yapıya sahip olmayan çok katmanlı sinir ağlarıdır ve bu ağlar geriye yayılım algoritması kullanılarak eğitilmişlerdir. Bölüm 5'te ise tablo yöntemi ve sinir ağları kullanılarak genişletilen zarfların altı yukarıda bahsedilen üç yöntemle doldurularak konuşma kalitesi artırılmaya çalışılmıştır. Bu bölümde ayrıca genişletilen izgel zarfa ait kazanç parametresinin hesaplanma yöntemi ve kaynak süzgeç modelinde kullanılan uyarım sinyalinin genişletilmesi üzerinde de durulmuştur. Bölüm 6'da ise elde edilen sonuçlar değerlendirilmiş ve ileride yapılabilecek çalışmalara baz olabilecek öneriler yapılmıştır.

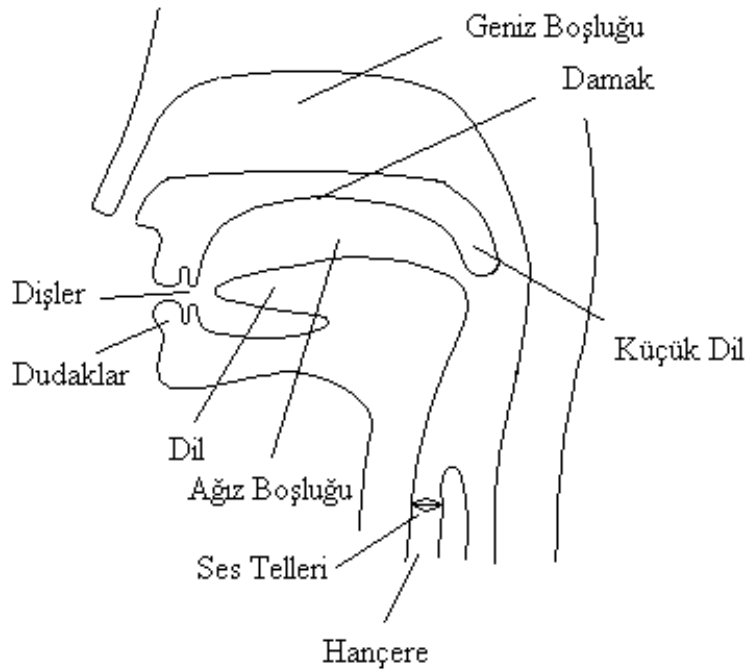


## 2. KONUŞMANIN OLUŞUMU VE MODELLENMESİ

İzgel zarf genişleterek konuşma kalitesini artırmak için öncelikle konuşma sinyali hakkında temel bazı bilgilere sahip olmamız gerekmektedir. Çalışmanın bu bölümünde konuşmanın oluşumu, modellenmesi ve özellikleri hakkında bazı kısa bilgilere yer verilecektir.

### 2.1 Konuşmanın Oluşumu

İnsanlar konuşmak için akciğerlerinde biriken havayı kullanırlar. Konuşmanın olmadığı durumda, akciğerlerde biriken bu hava hiçbir engelle karşılaşmadan dışarı verilir. İnsanlar konuşacakları zaman ise bu havayı Şekil 2.1'de görülen ses sistemlerinde çeşitli rezonans, türbülans, ya da patlamalar oluşturarak dışarı verirler (Rabiner and Schafer, 1978), (Özen, 2002).



Şekil 2.1. Ses sistemi

Akciğerlerden sonra hava, herhangi bir değişikliğe uğramadan hançerenin sonundaki ses tellerine kadar gelir. Bundan sonra iki farklı durum söz konusudur. Birinci durumda, ses tellerinin titreşimi ile gelen hava akımına belli bir şekil kazandırılıp Şekil 2.6'daki gibi az çok periyodik dalga şekline sahip olan sesli (sedalı) seslerin (a, e, ı, i, o, ö, u, ü, b, c, d, g, j, l, m, n, r, v, y, z) üretilmesi sağlanır. İkinci durumda ise gelen hava akımı ses telleri titreşmediği için Şekil 2.7'deki gibi periyodik bir uyarıma dönüş-

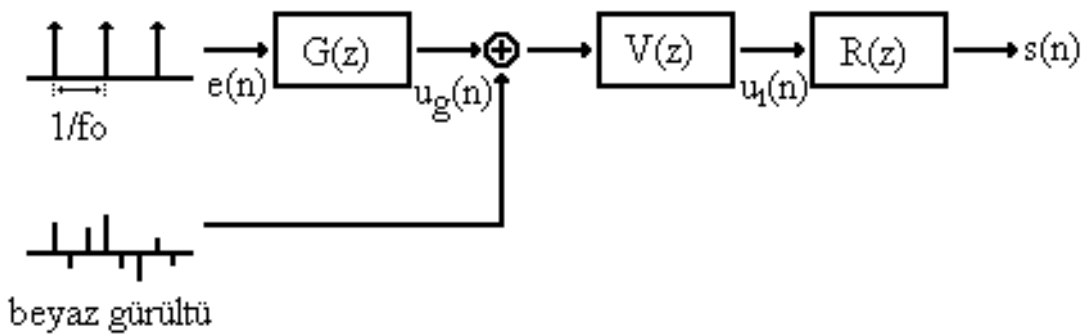
mez ve böylece sessiz seslerin (sedasız) ( $p, ç, t, k, f, h, s, ş$ ) oluşumu sağlanır. Ses tellerinin saniyede yapmış olduğu titreşim miktarı perde frekansı,  $f_0$  olarak adlandırılır. Ses tellerini geçip gırtlaktan yukarı doğru ilerleyen hava akımı, arka damağın sonunda asılı bulunan küçük dil tarafından ağız veya geniz boşluklarına yönlendirilir. Türkçe'de bulunan  $m$  ve  $n$  sesleri geniz yolunu kullanırlar. Türkçe'de bulunan diğer sesler ise ağız boşluğunda oluşurlar. Ağız boşluğundaki seslerin oluşumunda dil, damak, dişler ve dudaklar rol oynar.

## 2.2 Konuşmanın Kaynak-Süzgeç Modellenmesi

Şekil 2.2'de de görülen bu modelde (Rabiner and Schafer, 1978) konuşmanın akciğerlerden gelen havanın ses üretim yolunu modelleyen bir süzgeci uyarılmasıyla oluştuğu kabul edilir. Genelde, sesli ve sessiz sesler için sırasıyla dürtü dizisi ve beyaz gürültü olmak üzere iki farklı uyarım dizisi kullanılır. Sesli ve sessiz konuşma parçalarına örnekler Şekil 2.6a ve Şekil 2.6a'da ve bunlara ait uyarım sinyalleri ise sırasıyla Şekil 2.6a ve Şekil 2.6a'de görülebilir. Burada ses tellerinin etkisini  $G(z)$ , ağız ve geniz boşluğunun etkisini  $V(z)$  ve dudakların etkisini  $R(z)$  üzerinde toplayacak biçimde

$$H(z) = G(z)V(z)R(z) \quad (2.1)$$

şekilde tanımlanabilir.



Şekil 2.2. Ses oluşum modeli

Modelde bulunan  $G(z)$  transfer fonksiyonu, ses tellerinin oluşturduğu hava akımını modellemektedir. Periyodik dürtü dizisi, ses tellerinin perde frekansında,  $f_0$ , titreşmesini sağlamaktadır.  $G(z)$ 'nin çıkışında elde edilen  $u_g(n)$  glotal uyarım (glottal excitation) dalga biçimi olarak bilinir. Perde frekansı, konuşmanın yapısına göre metin

içinde deęişiklik gösterir ve bu deęişiklik konuşmanın anlamını ve doğallığını belirler. Sessiz seslerde ses telleri titreşmedięi için  $G(z) = 1$  olarak alınabilir ve bu durumda konuşmayı oluşturan süzgeçin

$$H(z) = V(z)R(z) \quad (2.2)$$

akcięerlerden gelen havanın ses yolunda takıntıya uğrayarak oluşturduęu beyaz gürültüyle sürüldüęü düşünülür. Beyaz gürültünün gücü konuşmadan konuşmaya farklılıklar gösterir ve model hatasıyla orantılıdır (Rabiner and Schafer, 1978).

Ses tellerinden gelen hava akımı (sesli konuşma için dürtü dizisi, sessiz konuşma için beyaz gürültü), ağız ve geniz boşluklarında şekillenir. Ağız ve geniz boşluklarının bu etkisi  $V(z)$  ile modellenir. Buralardaki ses yolu organlarının aldığı biçime göre, hava akımının bazı frekans bölgeleri yükseltilirken, bazı frekans bölgeleri bastırılır. Yükseltelen bu frekans noktaları, formant frekansları olarak adlandırılırlar.  $V(z)$  karmaşık eşlenik kutuplardan oluşan bir süzgeçtir. Ağız ve geniz boşluęundan sonra konuşma dudaklardan çıkarak oluşumunu tamamlar. Dudakların konuşmaya etkisi yüksek geçirgen bir filtre olarak modellenir ve bir tane sıfır içerir

$$R(z) = 1 - z^{-1}. \quad (2.3)$$

Sonuçta konuşmanın

$$S(z) = E(z)H(z) \quad (2.4)$$

biçiminde,  $E(z)$  uyarım sinyalinin ses üretim yolunu modelleyen  $H(z)$  süzgeçinden geçirilmesiyle oluştuęu varsayılır. Önceden de bahsedildięi gibi sesli ve sessiz konuşma için

$$H_v(z) = G(z)V(z)R(z) \quad (2.5)$$

$$H_u(z) = V(z)R(z)$$

olacak biçimde iki farklı biçimde modellenebilir. Konuşma sinyalinde sesli bölgeler sessizlere oranla çoğunluktadır ve sesli konuşma parçaları tüm kutup (all-pole)

süzgeçle modellenir, bu yüzden  $H(z)$  TK (Tüm Kutup) süzgeç olarak seçilir

$$H(z) = \frac{g}{A(z)} = \frac{g}{1 + a_1 z^{-1} + \dots + a_p z^{-p}} \quad (2.6)$$

ve süzgeç katsayılarını hesaplamak için çeşitli hızlı yöntemler mevcuttur (Rabiner and Schafer, 1978), (Kay, 1988), (Moon and Stirling, 2000). Burada  $g$  filtre kazancı ve  $p$  de TK süzgecin derecesidir. Süzgeç derecesi genelde

$$p = 2B + (2 \text{ ya da } 4) \quad (2.7)$$

eşitliği kullanılarak bulunabilir. Burada  $B$  konuşmanın kHz cinsinden bant genişliğidir. Örnek olarak Şekil 2.6'daki gibi 3.4 kHz'e bant sınırlı 8 kHz ile örneklenmiş sesli bir konuşma parçasında 3–4 tane formant frekansı vardır. Her bir formantın 2. dereceden tüm kutup bir filtreye modellenebileceği düşünülürse formantlar için 6–8 tane parametre gerekmektedir. Ayrıca 2 – 4 tane parametrenin de genel izgel dengeyi modelleyeceği düşünülürse, 8 kHz ile örneklenmiş konuşma için 10-12. dereceden bir TK süzgecin, 16 kHz ile örneklenmiş bir konuşma için ise 16-20. dereceden bir TK süzgecin yeterli olacağı görülür. Bunun yanında Şekil 2.7'deki gibi sessiz konuşma parçaları için ise formantlar tanımlı değildir ancak sessiz konuşma parçaları konuşmada az oldukları için ve 10-12. dereceden TK modellendiklerinde pek fazla hata oluşmadığı için, sessiz konuşma parçaları da TK modellenirler (Rabiner and Schafer, 1978).

### 2.3 Konuşmanın Sinüs Modellenmesi

Konuşma üretim modeline göre, konuşma dalga biçimi  $s(t)$ , glotal uyarım dalga biçimi olan  $e(t)$ 'nin, doğrusal zamanla değişen bir süzgeçten geçirilmesiyle oluşur. Eğer zamanla değişen süzgecin dürtü yanıtı (impulse response)  $h(\tau; t)$  ise, konuşma dalga biçimi,

$$s(t) = \int_0^t h(t - \tau; t) e(\tau) d\tau \iff S(w) = H(w)E(w) \quad (2.8)$$

şeklinde yazılabilir (McAulay and Quatieri, 1986). Sinüs modelinde buradaki uyarım sinyalinin rastgele genlik, frekans ve evrelerden oluşan sinüslerin toplamı biçiminde

ifade edilebileceği önerilmektedir. Bu durumda uyarım dalga biçimi,

$$e(t) = Re \sum_{\ell=1}^{L(t)} a_{\ell}(t) \exp \left\{ j \left[ \int_0^t \omega_{\ell}(\sigma) d\sigma + \phi_{\ell} \right] \right\} \quad (2.9)$$

şeklinde ifade edilebilir. Burada,  $a_{\ell}(t)$  ve  $\omega_{\ell}(t)$   $\ell$ 'inci sinüs bileşenlerinin sırasıyla genlik ve anlık frekansını,  $\phi_{\ell}$  ise sinüslerin genellikle aynı evrede olmamalarından kaynaklanan evre kaymasını tanımlar. Bunun yanında ses üretim aktarım işlevi de

$$H(\omega; t) = M(\omega; t) \exp[j\Phi(\omega; t)] \quad (2.10)$$

şeklinde ifade edilebilir. Eş. 2.9'daki uyarım parametrelerinin de ses üretim dürtü yanıtı boyunca sabit olduğu kabul edilerek, Eş. 2.9 ve Eş. 2.10, Eş. 2.8'de yerine konulduğunda konuşma dalga biçimi,

$$s(t) = \sum_{\ell=1}^{L(t)} a_{\ell}(t) M[\omega_{\ell}(t); t] \exp \left\{ j \left[ \int_0^t \omega_{\ell}(\sigma) d\sigma + [\Phi(\omega_{\ell}(t); t)] + \phi_{\ell} \right] \right\} \quad (2.11)$$

şeklinde elde edilir (Burada  $Re$  (gerçek kısım) simgelem karışıklığını önlemek için geçici olarak ihmal edilmiştir). Glotal ve ses üretim yolu, genlik ve evrelerinin etkileri birleştirilerek Eş. 2.11 kısaca

$$s(t) = \sum_{\ell=1}^{L(t)} A_{\ell}(t) \exp[j\psi_{\ell}(t)] \quad (2.12)$$

biçiminde ifade edilebilir. Burada

$$A_{\ell}(t) = a_{\ell}(t) M[\omega_{\ell}(t); t] \quad (2.13)$$

ve

$$\psi_{\ell}(t) = \int_0^t \omega_{\ell}(\sigma) d\sigma + \Phi[\omega_{\ell}(t); t] + \phi_{\ell} \quad (2.14)$$

sırasıyla  $\ell$ 'inci frekans izi ( $\omega_\ell(t)$ ) için genlik ve evreyi belirtir. Ancak konuşma (bu çalışma boyunca olduğu gibi) önce uygun bir frekansta örneklenip çerçeve çerçeve (frame) işlendiği için kullanılan süzgeç bir çerçeve boyunca sabit kabul edilebilir. Ayrıca konuşma kompleks üstel yapıdan kurtarılıp daha basit ve anlaşılır bir biçimde ifade edilmek istenirse Eş. 2.12'teki konuşma ayrık zamanlı (discrete-time) bir sinyal olarak

$$s[n] = \sum_{\ell=1}^{L^k} A_\ell^k \cos(\omega_\ell^k n + \theta_\ell^k) \quad (2.15)$$

biçiminde ifade edilebilir. Burada  $n = 0, 1, 2, \dots, S - 1$  ve  $S$  ise her bir çerçevenin uzunluğudur. Ayrıca  $L^k$  k'inci çerçevedeki sinüslerin sayısını,  $A_\ell^k$ ,  $\omega_\ell^k$  ve  $\theta_\ell^k$  ise sırasıyla  $\ell$ 'inci sinüsün genliğini, anlık frekansını ve evresini belirtir. Bu parametreler her bir çerçeve için Kısa Zaman Fourier Dönüşümü (KZFD) (STFT) ile bulunur ve çerçeveden çerçeveye eşleştirme algoritmalarıyla konuşma yeniden ve hemen hemen hiç kayıpsız sentezlenebilir. Ayrıca Eş. 2.15'teki frekanslar sesli konuşma parçaları için perde frekansının tam katı olarak seçilebilir. Evreler önceki evrelerden ve genlikler ise izgel zarftan perde frekansının tam katlarında örnekler alınarak elde edilerek konuşma

$$s[n] = \sum_{\ell=1}^{L^k} A_\ell^k \cos(\ell \omega_0^k n + \theta_\ell^k) \quad (2.16)$$

biçiminde ifade edilebilir (Quatieri, 2002). Sessiz konuşma parçaları için perde frekansı tanımlı olmamakla beraber, konuşmanın sessiz bölgelerinde perdenin 100 Hz alınması konuşmada pek bir kayba yol açmaz (McAulay and Quatieri, 1986), (Quatieri, 2002).

## 2.4 Perde Frekansının Bulunması

Bu bölümde, çalışma sırasında kullanacağımız perde frekansının bulunma yöntemi hakkında kısaca bilgi verilecektir. Perde frekansını bulmak için literatürde pek çok yöntem bulunmaktadır. Bu çalışmada perde frekansı kepstrum yöntemiyle hesaplanacaktır (Noll, 1967). Bir  $x(n)$  sinyalinin gerçel (real) kepstrumu,

$$c(n) = \frac{1}{2\pi} \int_{-\pi}^{\pi} \log |X(e^{j\omega})| e^{j\omega n} d\omega \quad (2.17)$$

biçiminde tanımlanır. Burada  $X(e^{j\omega})$ ,  $x(n)$  sinyalinin Ayrık-Zaman Fourier Dönüşümüdür (AZFD) ve

$$X(e^{j\omega}) = \sum_{n=-\infty}^{+\infty} x[n] e^{-j\omega n} \quad (2.18)$$

biçiminde tanımlanır. Ancak sayısal ortamda gerçel kepstrum ve AZFD Eş. 2.17 ve Eş. 2.18'de  $\omega = (2\pi/N)k$  konularak

$$c[n] = \frac{1}{N} \sum_{k=0}^{N-1} \log |X[k]| e^{j(2\pi/N)kn}$$

$$X[k] = \sum_{n=0}^{N-1} x[n] e^{-j(2\pi/N)kn} \quad (2.19)$$

biçiminde hesaplanabilir (Oppenheim and Schaffer, 1989).

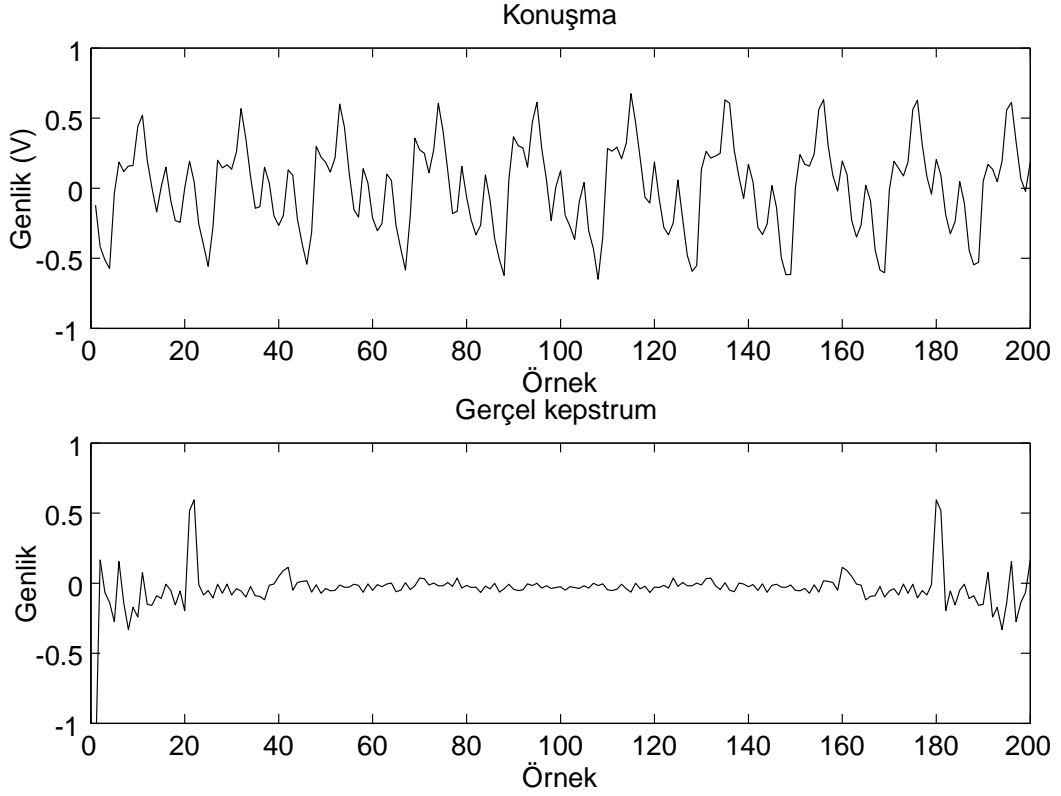
Sesli konuşma parçalarının kepstrumlarına bakıldığında perde periyodunda bir tepeye rastlanır, ancak sessiz seslerde böyle bir şey söz konusu değildir. Kepstrumun bu özelliğinden faydalanılarak perde frekansı bulunabilir. Bunun için önce perde periyodunun olabileceği aralıktaki (2.5 – 20 ms) en büyük tepe bulunur ve bu tepenin değeri bir eşik değeri ile karşılaştırılır. Eşik değeri perde periyodunun araştırıldığı bölgenin ortalama değerinin 4 – 5 katı olarak alınabilir. Eğer bulunan tepe değeri eşikten büyük ise tepenin yeri perde periyodu,  $T_0$  olarak alınır ve perde frekansı da

$$f_0 = \frac{1}{T_0} \quad (2.20)$$

kullanılarak hesaplanabilir. Örneğin, Şekil 2.3a'da 8 kHz ile örneklenmiş sesli bir konuşmanın 25 ms'lik (200 örnek) bir parçası, Şekil 2.3b'de ise bu konuşma parçasının gerçel kepstrumu görülmektedir. Bu örnekte perde periodun yaklaşık olarak 2.625 ms (21 örnek) olduğu görülebilir ve bu durumda perde frekansı 380 Hz olarak hesaplanır.

## 2.5 İzgel Zarfın Modellenmesi

Konuşmanın klasik olarak glotal uyarım dalga biçiminin ses üretim yolunun (vocal tract) rezonans karakteristiklerini modelleyen zamanla değişen doğrusal bir süzgeç-



Şekil 2.3. Konuşma ve gerçek kepstrumu

ten geçirildikten sonra oluşturduğundan bir önceki bölümde bahsedilmişti. Bu TK süzgeç,  $H(z)$ , konuşmanın izgel zarfını da modeller ve  $p$ 'inci dereceden analiz süzgeci,  $A(z)$  (ters süzgeç),

$$A(z) = \sum_{i=0}^p a_i z^{-i} \quad (2.21)$$

şeklinde tanımlanır. Bu katsayılar  $a_0, a_1, \dots, a_p$  doğrusal öngörümlü kodlama (DÖK, LPC) yöntemiyle bulunabilir. Doğrusal öngörüm orijinal örnek ile öngörülen örnek arasındaki hatanın belirli kriterlere göre küçültülmesine elde edilir. Buradaki hata,

$$e(n) = s(n) - \hat{s}(n) = s(n) + \sum_{i=1}^p a_i s(n-i) \quad (2.22)$$

biçiminde ifade edilir. Toplam karesel hata da

$$\alpha = \sum e^2(n) \quad (2.23)$$

biçimindedir ve bu hata çeşitli yöntemlerle küçültülüp konuşmanın izgel zarfını modelleyen bu katsayılar bulunabilir. Daha ayrıntılı bilgi (Makhoul, 1975) ve (Rabiner



and Schafer, 1978)'de mevcuttur.

Buradaki süzgeç 8 kHz'de örneklenmiş bir konuşma için genellikle 10. dereceden tüm kutup (TK) bir süzgeçtir ve bu tüm kutup süzgeç konuşmanın izgel zarfını modeller. Burada 10. dereceden TK parametreleri bilinen konuşmanın 3.4 – 7 kHz'lik bölgesinin izgel zarfı bulunmaya çalışılacaktır. 16 kHz'te örneklenmiş bir konuşmanın izgel zarfı da 16-20. dereceden TK parametreleriyle ifade edilebilir. Dolayısıyla izgel genişletmedeki amaç 10. derecedeki TK parametrelerinden 20. dereceden TK parametrelerini elde etmektir.

Ancak TK katsayıları işlem yapmaya (nicemlenmeye, kodlamaya) pek elverişli değildir, çünkü bu katsayılar nicemlenirken meydana gelebilecek çok küçük nicemleme hataları bile izgel zarfı büyük ölçüde etkileyebilir, hatta tüm kutup süzgeç kararsız bile olabilir. Bu yüzden bu katsayıların işlem yapmaya (kodlama vb.) daha elverişli olan diğer katsayılar cinsinden ifade edilmeleri gerekmektedir (örneğin logaritmik alan oranları (log-area ratios) (Viswanathan and Makhoul, 1975) , arcsinüs yansıma katsayıları (arcsine reflection coefficients) (Gray and Markel, 1976) ve çizgi izgel frekansları (line spectral frequencies) (Itakura, 1975)). Biz burada TK parametrelerini vektör nicemlenmeye de en uygun gösterim biçimi olan ÇİF'ler (çizgi izgel frekanslar) cinsinden göstereceğiz.

## 2.6 ÇİF Gösterimi ve Özellikleri

Bu bölümde ÇİF tanımlanacak ve bazı özellikleri açıklanacaktır. Konuşma yolu izgel zarfı, TK süzgeç katsayıları cinsinden ifade edildiğinde TK süzgeç,

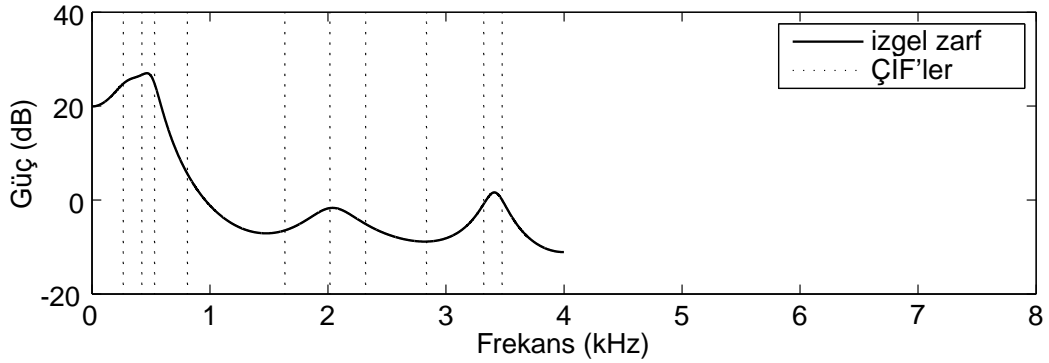
$$H(z) = \frac{1}{A(z)} = \frac{1}{1 + a_1 z^{-1} + \dots + a_p z^{-p}} \quad (2.24)$$

biçiminde yazılır ve  $A(z)$  ters süzgeç olarak tanımlanır. Burada  $p$  TK süzgecin derecesi ve  $a_i$ 'ler de katsayılarıdır. ÇİF'ların tanımlanabilmesi için önce ters süzgeç polinomundan faydalanarak,

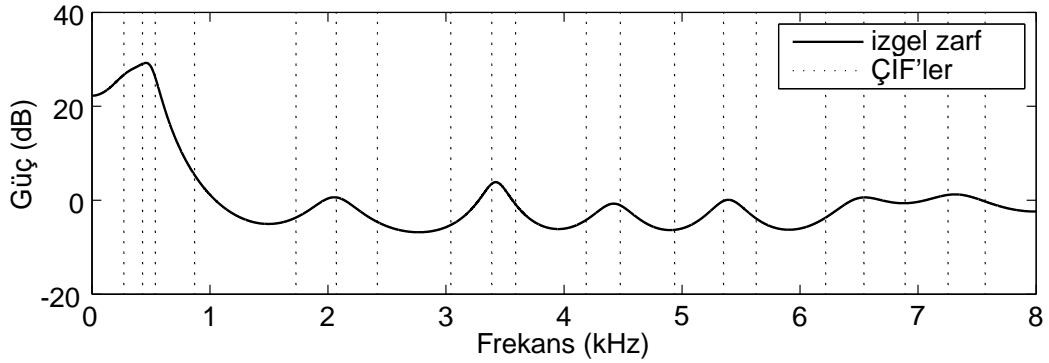
$$P(z) = A(z) + z^{-(p+1)} A(z^{-1}) \quad (2.25)$$

$$Q(z) = A(z) - z^{-(p+1)} A(z^{-1})$$

polinomlarının tanımlanması gerekir.  $P(z)$  ve  $Q(z)$  polinomlarının kökleri ÇİF olarak tanımlanırlar ve bu kökler şu özelliklere sahiptirler (Paliwal and Atal, 1993):



a. 8 kHz ile örneklenmiş konuşma için izgel zarf ve ÇİF'ler



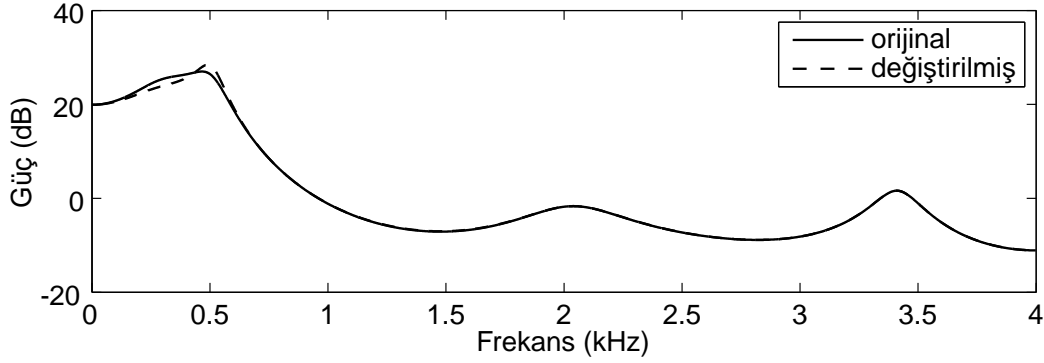
b. 16 kHz ile örneklenmiş konuşma için izgel zarf ve ÇİF'ler

Şekil 2.4. ÇİF'lerin İzgel zarfta dağılımları

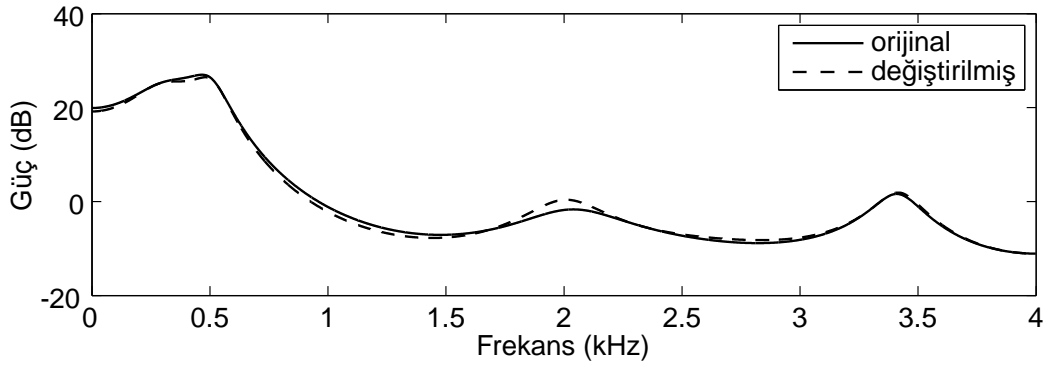
- $P(z)$  ve  $Q(z)$ 'in tüm kökleri birim çember üzerindedir.
- $P(z)$  ve  $Q(z)$ 'in kökleri birbiri içine geçmişlerdir ve bu kökler artan bir sırada bulunurlar.
- TK parametrelerinden ÇİF'in elde edilmesi tersinirdir.
- ÇİF'in duyarlılığı bölgeseldir.

Yukarıdaki özelliklere uygun olarak 8 kHz ve 16 kHz'te örneklenmiş aynı konuşma parçalarına ait izgel güç yoğunlukları ve ÇİF'ler Şekil 2.4'te görülebilir. Şekil dikkatli incelendiğinde iki ya da üç ÇİF'in bir formantın özelliklerini belirlediği görülebilir. ÇİF'lerin bölgesel duyarlılığı ise Şekil 2.5'te görülebilir. Şekil 2.5.a'da ikinci ÇİF

424'ten 445'e ve Şekil 2.5.b'de ise beşinci ÇİF 1636'dan 1719'a değiştiği zamanki izgel zarflar gösterilmektedir. Görüldüğü gibi bir ÇİF değerindeki değişiklik o bölgedeki izgel güç yoğunluğunu etkilemekte, ama daha uzak bölgeler neredeyse bu değişimden hiç etkilenmemektedir.



a. İkinci çizgi izgel frekans 424'den 445'e değiştiği zaman

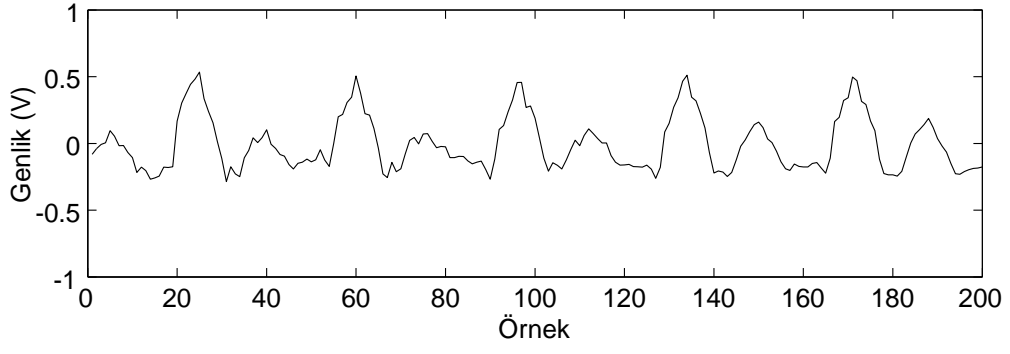


b. Beşinci çizgi izgel frekans 1636'dan 1719'e değiştiği zaman

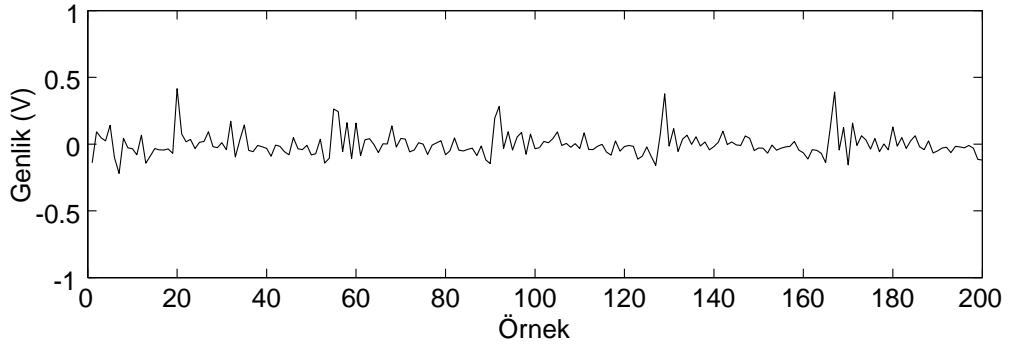
Şekil 2.5. Çizgi izgel zarfların değişimiyle izgel güç yoğunluğunun değişimi

ÇİF'lerin tüm bu kullanışlı özelliklerine rağmen bu çalışma sırasında olumsuz bir özelliği ile karşılaşmıştır. Örneğin Şekil 2.4'te 16 kHz ile örneklenmiş konuşmaya ait ÇİF'lerin 4 kHz'in altında kalanlarının, 8 kHz ile örneklenmiş olan aynı konuşmaya ait olanlarla tamamen aynı olmadıkları görülebilir. Hatta başka konuşma örneklerinde 16 kHz ile örneklenmiş konuşmaya ait ÇİF'lerden 4 kHz'in altında kalanların sayısının 10'dan fazla yada az olduğu durumlarla karşılaşmıştır. Bu durum bu bölümden sonra anlatılacak olan izgel zarf dışdeğerlemede birtakım problemler çıkarmıştır. Bu problemlerden en önemlisi tablo yönteminde oluşturulacak yüksek frekanslı bölgeye ait ÇİF'lerin olmaları gereken bölgenin  $[F_s/2F_s]$  ( $F_s$  : örnekleme frekansı) dışına taşmalarıdır. Bu durumda sınır bölgelerindeki ÇİF'ler birbirleri içine

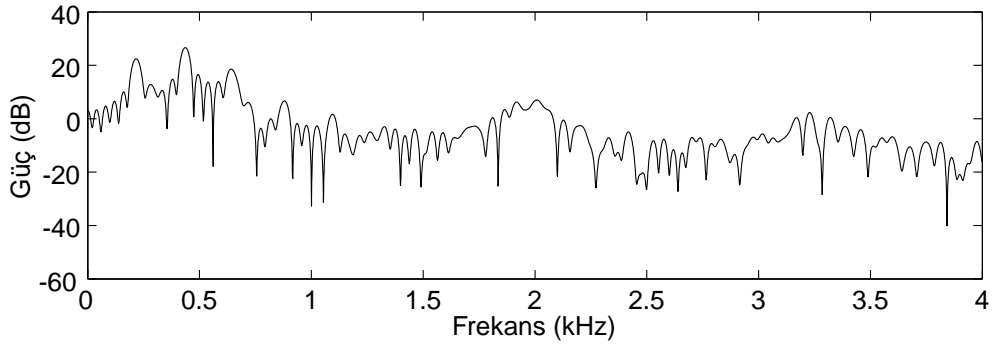
geçebilmekte ve geri dönüşümde elde edilecek DÖK katsayıları tamamen değişebilmekte ve hatta bazen DÖK katsayılarını elde etmek mümkün olmamaktadır. Benzer problemlerle sinir ağlarıyla izgel zarf genişletmede de karşılaşılmaktadır.



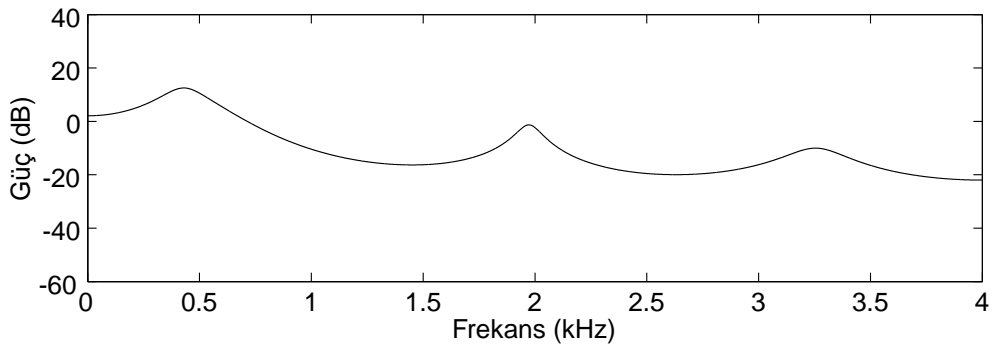
a. Sesli konuşma parçası



b. Sesli konuşma için uyarım sinyali

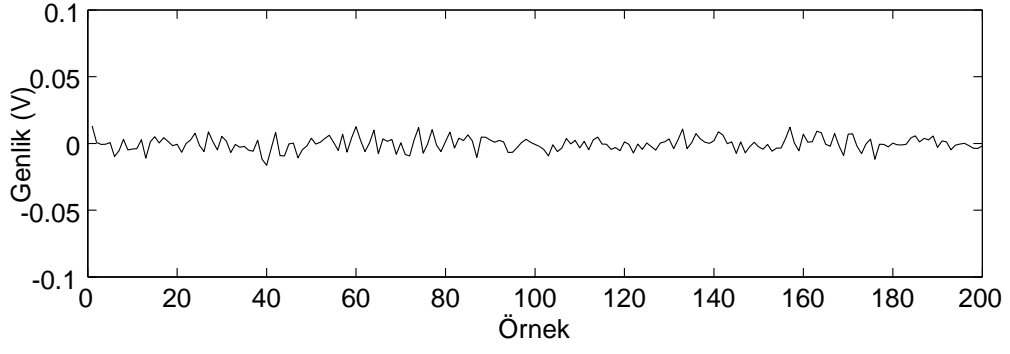


c. Sesli konuşmanın izgel güç yoğunluğu

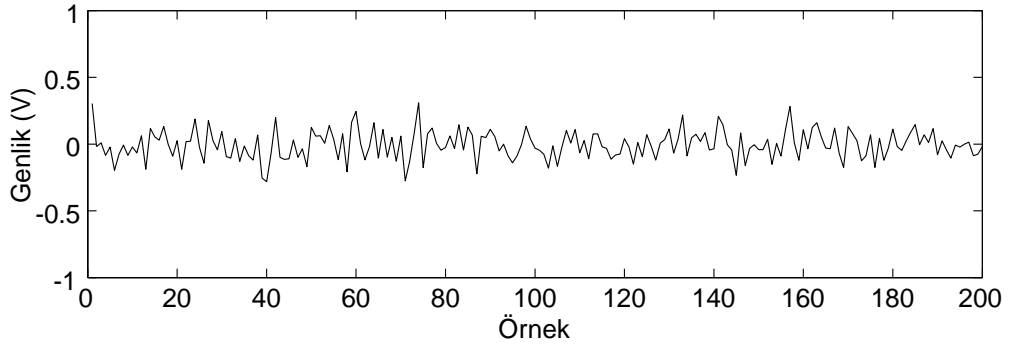


d. Sesli konuşmanın izgel zarfı

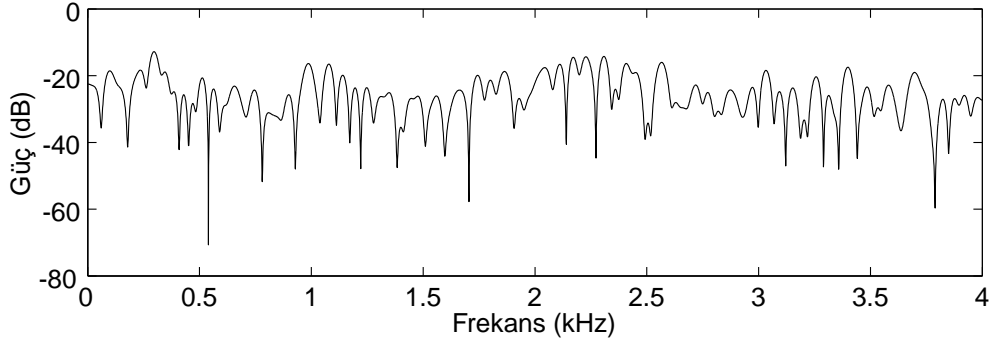
Şekil 2.6. 8 kHz ile örneklenmiş sesli konuşma parçası ve özellikleri



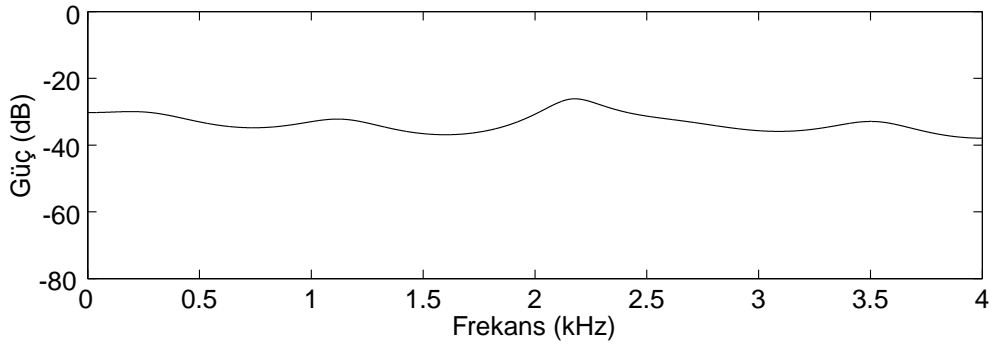
a. Sessiz konuşma parçası



b. Sessiz konuşma için uyarım sinyali



c. Sessiz konuşmanın izgel güç yoğunluğu

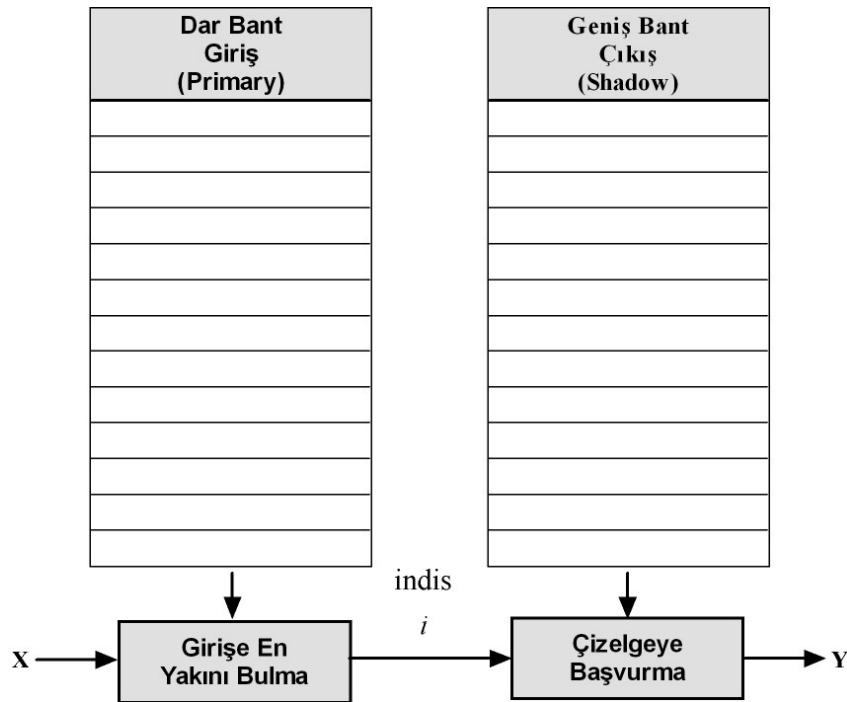


d. Sessiz konuşmanın izgel zarfı

Şekil 2.7. 8 kHz ile örneklenmiş sessiz konuşma parçası ve özellikleri

### 3. TABLO YÖNTEMİ KULLANILARAK İZGEL ZARFIN GENİŞLETİLMESİ

Bu bölümde dar bantlı konuşmaya ait parametrelerden geniş bantlı konuşma parametreleri tablo yöntemi kullanılarak elde edilmeye çalışılacaktır. Bu yöntemin basit bir işleyiş şeması Şekil 3.1'de görülebilir. Bu şema dikkatli incelendiğinde iki tane tablonun birbirine paralel bir şekilde çalıştıkları gözlenebilir. Sistemin girişinde dar bantlı konuşmaya ait TK parametrelerinden elde edilen ÇİF'ler, çıkışında ise geniş bantlı konuşmaya ait ÇİF'ler yer almaktadır. Çıkış çizelgesi genel olarak iki yöntem kullanılarak tasarlanabilir. Birinci yöntemde tüm geniş bant parametreleri ikinci yöntemde ise sadece yüksek frekansa ait parametreler kullanılır. Sistem girişine gelen dar bantlı her bir vektör için başvuru çizelgesinde bu vektöre en yakın kod vektörü ve ona ait indis bulunur. Daha sonra bu indise karşılık gelen çıkış setindeki vektör konuşmanın o parçası için geniş bantlı parametre vektörü (ÇİF) olarak alınır.



Bunun için öncelikle bu işlemi yapacak başvuru ve çıkış çizelgelerinin oluşturulması gerekmektedir. Bu çizelgeleri oluşturmak için kullanılabilen en iyi yöntem bir vektör nicemleyici kullanmaktır. Bu bölümde önce vektör nicemleyici tasarımı basamakları kısaca anlatılacak ardından bu vektör nicemleyici tasarım basamakları

değiştirilerek (modify), kullanılacak başvuru ve çıkış çizelgelerinin tasarımları yapılacaktır (Bulu ve Bingöl, 2007). Daha sonra bu yöntem kullanılarak konuşma parametreleri genişletilerek, geniş bantlı konuşma parametreleri elde edilecektir. Son olarak, elde edilen bu parametrelerden ve orijinal konuşma parametrelerinden oluşturulan izgel zarflar arasındaki farklar hesaplanarak kullanılan yöntemin başarımı irdelenecektir.

### 3.1 Vektör Nicemleme

$N$  düzeyli  $k$  boyutlu bir vektör nicemleyici  $q$  ile gösterildiğinde ve giriş vektörü de

$$\mathbf{x} = \{x_0, x_1, \dots, x_{k-1}\} \quad (3.1)$$

olduğunda, yeniden oluşum vektörü

$$\hat{\mathbf{x}} = q(\mathbf{x}) \quad (3.2)$$

şeklinde yazılır ve bu yeniden oluşum alfabesi (codebook),

$$\hat{\mathbf{A}} = \{\mathbf{y}_i; i = 1, \dots, N\} \quad (3.3)$$

içinden seçilir.

LBG (Linde-Buzo-Gray) Vektör Nicemleme Algoritması: Burada  $N$  düzeyli  $k$  boyutlu bir vektör nicemleyicinin tasarlanması basamakları anlatılacaktır. Bu algoritmanın dört tane basamağı vardır (Linde *et al.*, 1980), (Paliwal and Atal, 1993):

- **Sıfırıncı basamak (Başlangıç durumu):** Bu basamakta ilk önce bozulum eşiği (distortion threshold)  $\varepsilon \geq 0$  seçilir. Daha sonra bir eğitim dizisi (training sequence)  $\{\mathbf{x}_j; j = 0, \dots, M - 1\}$  için gerekli vektörler bulunur ve başlangıç için  $N$  düzeyli bir kod çizelgesi  $\hat{\mathbf{A}}_0$ , eğitim dizisinden rastgele seçilir. Ayrıca yineleme sayısı (iteration number) sıfıra eşitlenir ( $m = 0$ ) ve bir önceki bozulum da sonsuz olarak alınır ( $D_{-1} = \infty$ ).
- **Birinci basamak:** Eğitim kümesindeki  $\{\mathbf{x}_j; j = 0, \dots, M - 1\}$  her bir vektör kod çizelgesi  $\{\hat{\mathbf{A}}_m = \mathbf{y}_i; i = 1, \dots, N\}$  ile en küçük bozulumu sağlayacak biçimde



yeniden oluşum vektörlerinin komşuluklarına

$$\rho(\hat{\mathbf{A}}_m) = \mathbf{S}_i; \quad i = 1, \dots, N \quad (3.4)$$

alınırlar. Buradaki en küçük bozulum ve her bir komşuluktaki  $\mathbf{x}_j \in \mathbf{S}_i$

$$d(\mathbf{x}_j, \mathbf{y}_i) \leq d(\mathbf{x}_j, \mathbf{y}_k) \quad k = 1, \dots, N \quad (3.5)$$

olacak şekilde seçilir. daha sonra ortalama bozulum ise

$$D_m = \{\hat{\mathbf{A}}_m, \rho(\hat{\mathbf{A}}_m)\} = \frac{1}{M} \sum_{j=0}^{M-1} \min_{\mathbf{y} \in \hat{\mathbf{A}}_m} d(\mathbf{x}_j, \mathbf{y}) \quad (3.6)$$

kullanılarak hesaplanır.

- **İkinci Basamak:** Eğer

$$\frac{D_{m-1} - D_m}{D_m} \leq \varepsilon \quad (3.7)$$

ise algoritmadan çıkılır ve  $\hat{\mathbf{A}}_m$  yeniden oluşum alfabetesi olarak alınır.

- **Üçüncü basamak:** Eğer ikinci basamaktaki koşul sağlanmazsa, birinci basamakta elde edilen komşulukların ağırlık merkezleri (centroid) hesaplanarak yeni kod çizelgesi bulunur ve  $m$  bir artırılarak birinci basamağa dönülür.

### 3.2 Vektör Nicemleyici Kullanarak Başvuru Çizelgesi Oluşturulması

Bu çalışmada düşük bant genişliğine sahip konuşmanın  $p$  boyutlu, yüksek bant genişliğine sahip konuşmanın ise  $q$  ( $p < q$ ) boyutlu ÇİF'le ifade edildiğini düşünelim. Aslında ilk bakışta ÇİF'in özellikleri düşünüldüğünde bir tane  $q$  boyutlu,  $N$  düzeyli bir vektör nicemleyicinin tasarlanmasının yeterli olabileceği düşünülebilir. Ancak yüksek bant genişliğine sahip konuşmanın  $q$  tane ÇİF'inden ilk  $p$  tanesinin düşük bant genişliğine sahip konuşmanın diğerleriyle örtüşmesinin garantisi olmadığı gibi bunlar bu dar bantın dışında da olabilirler. Bütün bu durumlar göz önüne alındığında  $p$  ve  $q$  boyutlu,  $N$  düzeyli iki tane vektör nicemleyicinin birbiriyle paralel olarak tasarlanması gerekliliği ortaya çıkmaktadır.

Başvuru çizelgesi için giriş ve çıkış eğitim setlerindeki her bir vektörü sırasıyla,

$$\mathbf{x} = \{x_0, x_1, \dots, x_{p-1}\} \quad (3.8)$$

$$\mathbf{v} = \{v_0, v_1, \dots, v_{q-1}\}$$

biçiminde tanımlanabilir. Giriş ve çıkış vektörleri için alfabeler oluşturulurken eğitim setindeki her bir vektör  $P$  ve  $Q$  nicemleyicilerinden geçerek başvuru çizelgesindeki ve çıkıştaki vektörleri

$$\hat{\mathbf{x}} = P(\mathbf{x}) \quad (3.9)$$

$$\hat{\mathbf{v}} = Q(\mathbf{v})$$

şeklinde oluştururlar.  $p$  boyutlu her bir vektör sırasıyla başvuru çizelgesi olan

$$\hat{\mathbf{A}} = \{\mathbf{y}_i; i = 1, \dots, N\} \quad (3.10)$$

ve  $q$  boyutta bunlara karşı gelen

$$\hat{\mathbf{B}} = \{\mathbf{z}_i; i = 1, \dots, N\} \quad (3.11)$$

setinden seçilir. Burada LBG vektör nicemleme algoritması, izgel dışdeğerlemede kullanılacak olan başvuru çizelgesini oluşturmak için biraz değiştirilerek kullanılacaktır.

- **Sıfırıncı basamak:** Bu basamakta ilk önce bozulum eşiği  $\varepsilon \geq 0$  seçilir. Daha sonra eğitim dizileri  $\{\mathbf{x}_j; j = 0, \dots, M - 1\}$ ,  $\{\mathbf{v}_j; j = 0, \dots, M - 1\}$  sırasıyla düşük ve yüksek bant genişliğine sahip konuşmaların aynı bölgelerinden alınır. Başlangıç için  $N$  düzeyli kod çizelgeleri  $\hat{\mathbf{A}}_0$ ,  $\hat{\mathbf{B}}_0$  eğitim dizilerinin aynı konuşma parçalarından senkronize bir biçimde seçilir. Ayrıca yineleme sayısı sıfıra eşitlenir ( $m = 0$ ) ve bir önceki bozulum da sonsuz olarak alınır ( $D_{-1} = \infty$ ).
- **Birinci basamak:** Eğitim kümelerindeki  $\{\mathbf{x}_j; j = 0, \dots, M - 1\}$   $\{\mathbf{v}_j; j = 0, \dots, M - 1\}$  her bir vektör kod çizelgesi sadece  $\mathbf{x}_j$ ,  $\{\hat{\mathbf{A}}_m = \mathbf{y}_i; i = 1, \dots, N\}$  ile en küçük bozulumu sağlayacak biçimde ve  $\mathbf{v}_j$ 'lerde bunlara paralel biçimde,  $\{\hat{\mathbf{B}}_m =$

$\mathbf{z}_i; i = 1, \dots, N$  yeniden oluřum vektörlerinin komřuluklarına

$$\rho(\hat{\mathbf{A}}_m) = \mathbf{S}_i; \quad i = 1, \dots, N \quad (3.12)$$

$$\rho(\hat{\mathbf{B}}_m) = \mathbf{T}_i; \quad i = 1, \dots, N$$

alınırlar. Buradaki en küçük bozulum ve her bir komřuluktaki  $\mathbf{x}_j \in \mathbf{S}_i$

$$d(\mathbf{x}_j, \mathbf{y}_i) \leq d(\mathbf{x}_j, \mathbf{y}_k) \quad k = 1, \dots, N \quad (3.13)$$

olacak řeklinde seçilir. daha sonra ortalama bozulum ise

$$D_m = \{\hat{\mathbf{A}}_m, \rho(\hat{\mathbf{A}}_m)\} = \frac{1}{M} \sum_{j=0}^{M-1} \min_{\mathbf{y} \in \hat{\mathbf{A}}_m} d(\mathbf{x}_j, \mathbf{y}) \quad (3.14)$$

kullanılarak hesaplanır.

- **İkinci Basamak:** Eğer

$$\frac{D_{m-1} - D_m}{D_m} \leq \varepsilon \quad (3.15)$$

ise algoritmadan çıkılır ve  $\hat{\mathbf{A}}_m$  başvuru çizelgesi  $\hat{\mathbf{B}}_m$  ise bunların haritalanacağı çıkış vektörleri olarak alınır.

- **Üçüncü basamak:** Burada birinci basamakta elde edilen komřulukların ağırlık merkezleri hesaplanır. Daha sonra  $m$  bir artırılarak birinci basamağa dönülür.

Ancak yukarıda anlatılan LBG vektör nicemleyici tasarım algoritmasından türetilen başvuru çizelgesi oluřturma yöntemi teorik olarak çalışmasına rağmen başvuru çizelgesi boyutu büyüdükçe pratik olarak uygulanması oldukça güçtür, özellikle eğitim zamanları çok uzamaktadır. Zaten günümüzde kullanılan vektör nicemleyicilerde de bu güçlükten dolayı parçalı (split) vektör nicemleme yöntemleri kullanılmaktadır. Ayrıca yukarıda anlatılan yöntemle çizelge tasarlanırken referans olarak giriş alınmış ve tüm bozulum hesaplamaları giriş üzerinden yapılmıştır. Bu işlem çıkış referans olarak alınarak da yapılabilir ve önce girişin haritalanacağı çıkış tablosu tasarlanıp sonra girişteki başvuru çizelgesi tasarlanabilir. Bu işlemin basamakları ise şöyle özetlenebilir.

Önce kestirilecek izgel zarfın parametreleri olarak kabul edilen

$$\hat{\mathbf{v}} = \{v_p, v_{p+1}, \dots, v_{q-1}\} \quad (3.16)$$

için  $N$  düzeyli  $(q-p)$  boyutlu bir vektör nicemleyici önceden anlatılan klasik vektör nicemleme yöntemi kullanılarak tasarlanır. Bu nicemleyicide yeniden oluşum vektörü,  $\hat{\mathbf{v}} = \hat{Q}(\mathbf{v})$  şeklinde yazılır ve bu kod çizelgesi yeniden oluşum alfabeti

$$\hat{\mathbf{B}} = \{\hat{\mathbf{z}}_i; i = 1, \dots, N\} \quad (3.17)$$

içinden seçilir. Bu yeniden oluşum alfabeti başvuru çizelgesinin çıkış vektörleridir. Daha sonra ise başvuru çizelgesi bir önce anlatılan yöntemdekine benzer şekilde şöyle tasarlanabilir.

- **Sıfırıncı basamak:** Bu basamakta ilk önce bozulum eşiği  $\varepsilon \geq 0$  seçilir. daha sonra bir eğitim dizileri  $\{\mathbf{x}_j; j = 0, \dots, M-1\}$ ,  $\{\hat{\mathbf{v}}_j; j = 0, \dots, M-1\}$  sırasıyla düşük ve yüksek bant genişliğine sahip konuşmaların aynı bölgelerinden alınır. Başlangıç için  $N$  düzeyli kod çizelgesi  $\hat{\mathbf{A}}_0$ , eğitim dizisinde rastgele seçilir. Ayrıca yinleme sayısı sıfıra eşitlenir ( $m = 0$ ) ve bir önceki bozulum da sonsuz olarak alınır ( $D_{-1} = \infty$ ).
- **Birinci basamak:** Eğitim kümelerindeki  $\{\mathbf{x}_j; j = 0, \dots, M-1\}$   $\{\hat{\mathbf{v}}_j; j = 0, \dots, M-1\}$  her bir vektör kod çizelgesi sadece  $\hat{\mathbf{v}}_j$ ,  $\{\hat{\mathbf{B}}_m = \hat{\mathbf{z}}_i; i = 1, \dots, N\}$  ile en küçük bozulumu sağlayacak biçimde yeniden oluşum vektörlerinin komşuluklarına

$$\rho(\hat{\mathbf{A}}_m) = \mathbf{S}_i; \quad i = 1, \dots, N \quad (3.18)$$

alınırlar. Buradaki en küçük bozulum ve her bir komşuluktaki  $\mathbf{x}_j \in \mathbf{S}_i$

$$d(\mathbf{x}_j, \mathbf{y}_i) \leq d(\mathbf{x}_j, \mathbf{y}_k) \quad k = 1, \dots, N \quad (3.19)$$

olacak şekilde seçilir. daha sonra ortalama bozulum ise

$$D_m = \{\hat{\mathbf{A}}_m, \rho(\hat{\mathbf{A}}_m)\} = \frac{1}{M} \sum_{j=0}^{M-1} \min_{\mathbf{y} \in \hat{\mathbf{A}}_m} d(\mathbf{x}_j, \mathbf{y}) \quad (3.20)$$

kullanılarak hesaplanır.

- **İkinci Basamak:** Eğer

$$\frac{D_{m-1} - D_m}{D_m} \leq \varepsilon \quad (3.21)$$

ise algoritmadan çıkılır ve  $\hat{\mathbf{A}}_m$  başvuru çizelgesi olarak alınır.

- **Üçüncü basamak:** Burada birinci basamakta elde edilen komşulukların ağırlık merkezleri hesaplanır. Daha sonra  $m$  bir artırılarak birinci basamağa dönülür.

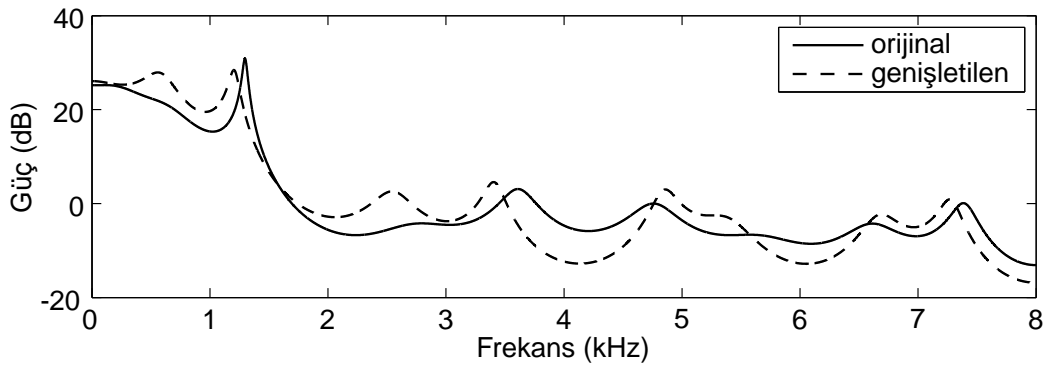
### 3.3 İzgel Zarfın Genişletilmesi

Bu çalışmada esas olarak 8 kHz ile örneklenmiş 3.4 kHz ile bant sınırlı (telefon kalitesi) konuşma parçalarının bant genişliklerinin 7 kHz'e çıkarılarak kalitelerinin artırılması hedeflenmiştir. Telefon kalitesindeki konuşma parçaları genel olarak 10. dereceden TK modellendiği için bu çalışmada da bu genel uygulamaya uyulmuştur. Yüksek kaliteli konuşma parçaları ise 20. dereceden TK modellenmiştir.

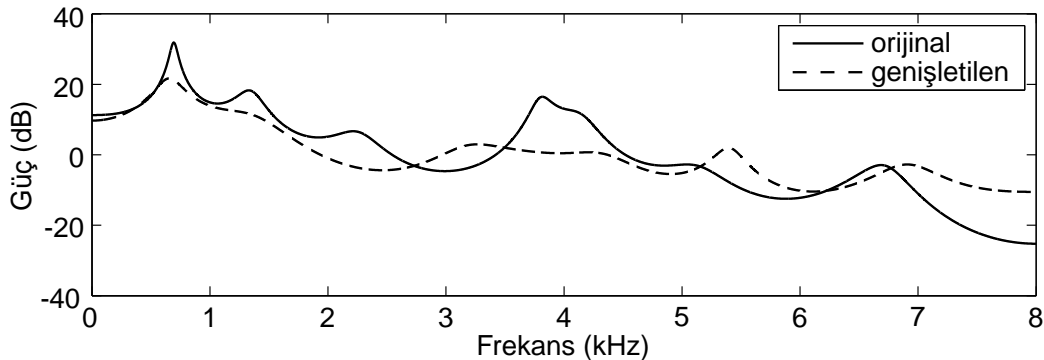
Dolayısıyla tablo yöntemindeki kullanılacak giriş çizelgesi 10. dereceden TK parametrelerinden elde edilmiş ÇİF'ler kullanılarak eğitilir. Çıkış çizelgesinin eğitiminde ise 20. dereceden TK parametrelerinden elde edilen 20 tane ÇİF kullanılır. Burada bir önceki bölümde anlatıldığı gibi iki tane eğitim çıkış tablosu tasarım yöntemi mevcuttur. Birinci yöntemde 20 tane ÇİF'in tümü çıkış tablosunda yer alır. Ancak tüm parametreler çıkış tablosunda yer almasına rağmen tablo eğitimi sırasında hesaplanan bozulum ölçütünde sadece son 10 parametre kullanılmaktadır. İkinci yöntemde ise sadece genişletilecek olan son 10 tane ÇİF çıkış tablosunda yer alır. Geriye kalan ilk 10 parametre dar bantlı konuşmadan elde edilenlerle eşit olarak kabul edilir. Çıkışların son 10 tane ÇİF olarak seçilmesi zaman zaman problemlere yol açmaktadır. Bu problemlerin en önemlisi çıkış tablosunda ilk ÇİF'lerden birinin  $\pi/2$ 'den küçük olması olasılığıdır. Böyle bir durumda en küçük değer  $\pi/2$ 'den küçük olmayacak biçimde düzeltmeler yapılmalıdır.

Çalışma sırasında kullanılan tablo yapısı Şekil 3.1'de görülmektedir. Bu yapı önce eğitilerek giriş ve çıkış tabloları elde edilmiş daha sonra geniş bant konuşma parametrelerinin elde edilmesi için kullanılmıştır. Bu yapının eğitimi için toplam 40 dakikalık bir konuşma veri tabanı kullanılmıştır. Bu yöntemle genişletilen, eğitim setinde olmayan konuşma parçaları için EK.1'de anlatılan ortalama izgel bozulum

ölçütü yaklaşık olarak 3.9 dB olarak bulunmuştur. Bulunan bu değerin Ek.2'de anlatılan izgel bozulumun bilgi teorisi bakış açısıyla bulunan ve Şekil 4.7'de görülen alt sınır eğrisiyle karşılaştırılması gerekmektedir (Jax, 2004). Bu karşılaştırmanın yapılabilmesi için ise bu eğride karşılaşılan karşılıklı bilgi miktarının (mutual information) bilinmesi gerekmektedir ve bu bilgi miktarı 3.4 kHz'e bant sınırlı olan telefon kalitesindeki konuşma parametreleri ile 7 kHz'e bant sınırlı olan konuşma parametreleri için karşılıklı olarak 2.3 – 2.4 bittir (Jax, 2004). Bu durumda izgel zarflar arasında oluşan bozulumun yaklaşık 3 dB'den küçük olamayacağı sonucuna varılır. Tablo yöntemiyle genişletilen izgel zarflar bu sınırdan bir miktar büyük olmakla birlikte literatürde yapılan benzer çalışmalarla uyumluluk göstermektedir (Epps and Holmes, 1998), (Jax and Vary, 2000), (Jax, 2004). Tablo yöntemiyle genişletilen izgel zarf örnekleri Şekil 3.2'de görülmektedir. Bu örneklerden Şekil 3.2.a'da izgel bozulum 3.8 dB, Şekil 3.2.b'de ise izgel bozulum 5.3 dB'dir.



a. İzgel Bozulum 3.8 dB



b. İzgel Bozulum 5.3 dB

Şekil 3.2. 7 kHz ile bant sınırlı orijinal ve 3.4 kHz ile bant sınırlı izgel zarftan tablo yöntemiyle 7 kHz'e genişletilen izgel zarflar

## 4. YAPAY SİNİR AĞLARIYLA İZGEL ZARFIN GENİŞLETİLMESİ

Çalışmanın bu bölümünde geniş bantlı konuşma parametreleri yapay sinir ağları (YSA) kullanılarak elde edilmeye çalışılmış ve bu işlem için çok katmanlı yapıya sahip olan bir sinir ağı seçilmiştir. Sinir ağının girişi dar bantlı konuşma parametreleri çıkışı ise genişletilen zarfa ait parametrelerdir. Bu yapı basit ve anlaşılabilir olmakla birlikte, aynı zamanda benzer işlemi yapan yöntemlerle karşılaştırılabilir sonuçlar verdiği için seçilmiştir. Burada önce YSA'lar ve kullanılan yöntem hakkında kısa bilgi verilecek daha sonra yöntemin uygulanış biçimleri ve sonuçlara yer verilecektir.

### 4.1 Yapay Sinir Ağları

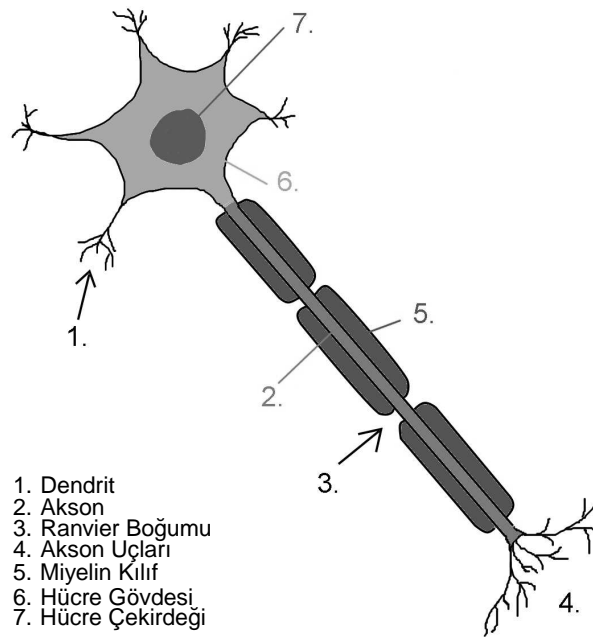
Yapay sinir ağları, insan beyninin özelliklerinden biri olan öğrenme yolu ile yeni bilgiler türetebilme, yeni ilişkiler kurabilme (associations) ve keşfedebilme gibi yetenekleri herhangi bir yardım almadan otomatik olarak gerçekleştirmek amacı ile geliştirilen bilgisayar sistemleridir. YSA, olayların örneklerine bakmakta, onlardan ilgili olay hakkında genellemeler yapmakta, bilgiler toplamakta ve daha sonra hiç görmediği örnekler ile karşılaşınca öğrendiği bilgileri kullanarak o örnekler hakkında karar verebilmektedir.

1990'lı yıllardan beri bilgisayarların öğrenmesini amaçlayan YSA teknolojisinde oldukça hızlı bir gelişme görülmüştür. YSA'lar sahip oldukları özelliklerden dolayı, tanımlanan bir işi herhangi bir yardım almadan otomatik olarak gerçekleştirmek amacı ile geliştirilen bilgisayar sistemleri olduklarından yeni gelişmelere neden olmuşlardır. Bunun yanında nasıl çalıştığı bilinmeyen insan beyni hakkında yapılan araştırmalara da önemli katkılar sağlamışlardır ve sağlamaktadırlar. Başka bir bakış açısıyla YSA teknolojisi insanlığın doğayı araştırma ve taklit etme çabalarının en son ürünlerinden bir tanesi olarak da yorumlanabilir.

YSA'lar, basit biyolojik sinir sisteminin çalışma şeklinin benzetimini yapmak (simulate) için tasarlanan programlardır. Benzetimi yapılan sinir sistemi Şekil 4.2'dekine benzer sinir hücrelerini (nöronlar) içerir ve bu sinir hücreleri çeşitli şekillerde birbirlerine bağlanarak ağı oluştururlar. Bu ağlar öğrenme, hafızaya alma ve veriler arasındaki ilişkiyi ortaya çıkarma kapasitesine sahiptirler. Diğer bir ifadeyle, YSA'lar, normalde bir insanın düşünme ve gözlemlemeye yönelik doğal yeteneklerini gerektiren problemlere çözüm üretmektedir. Bir insanın, düşünme ve gözleme yetenekle-

rini gerektiren problemlere yönelik çözümler üretebilmesinin temel sebebi ise insan beyninin ve dolayısıyla insanın sahip olduğu yaşayarak veya deneyerek öğrenme yeteneğidir.

İnsan beyninde öğrenme, nöronlar arasındaki sinaptik (synaptic) bağlantıların ayarlanması ile olur. Yani, insanlar doğumlarından itibaren bir yaşayarak öğrenme süreci içerisine girerler. Bu süreç içinde beyin sürekli bir gelişme göstermektedir. Yaşayıp tecrübe ettikçe sinaptik bağlantılar ayarlanır ve hatta yeni bağlantılar oluşur. Bu sayede öğrenme gerçekleşir. Bu durum YSA'lar için de geçerlidir. Öğrenme, eğitime yoluyla örnekler kullanarak olur; başka bir deyişle, gerçekleşme girdi ve çıktı verilerinin işlenmesiyle, yani eğitime algoritmasının bu verileri kullanarak bağlantı ağırlıklarını (weights of the synapses) bir yakınsama sağlanana kadar, tekrar tekrar ayarlamasıyla olur.

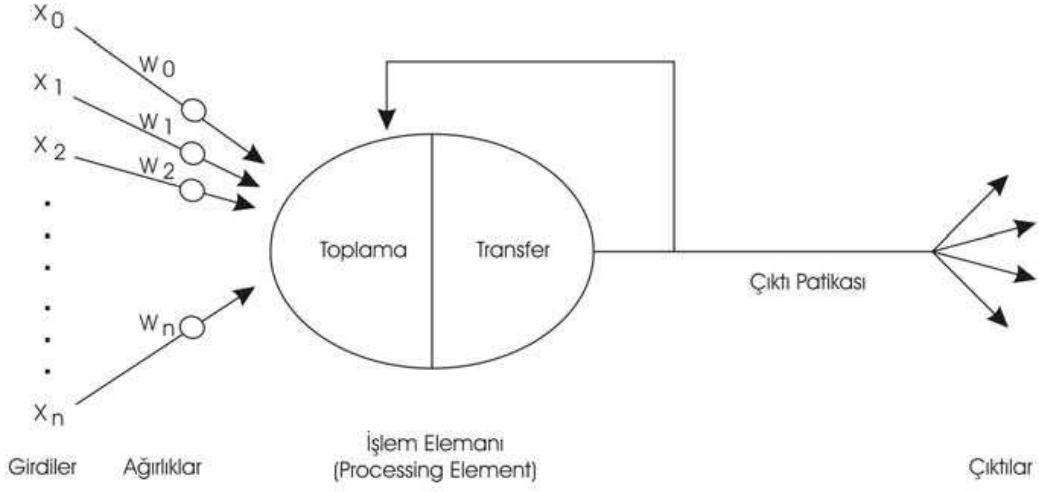


Şekil 4.1. Biyolojik nöron

YSA'lar, ağırlıklandırılmış şekilde birbirlerine bağlanmış bir çok işlem elemanlarından (nöronlar, Şekil 4.2) oluşan matematiksel sistemlerdir. Bir işlem elemanı, aslında aktarım işlevi (transfer fonksiyonu) olarak da anılan bir denklemdir. Bu işlem elemanı, diğer nöronlardan sinyalleri alır; bunları birleştirir, dönüştürür ve sayısal bir sonuç ortaya çıkarır. Genelde, işlem elemanları kabaca gerçek nöronlara karşılık



gelirler ve bir ağ içinde birbirlerine bağlanırlar; bu yapı da sinir ağlarını oluşturmaktadır.



Şekil 4.2. Nöron modeli

Sinirsel (neural) hesaplamanın merkezinde dağıtılmış, uyarlanır ve doğrusal olmayan işlem kavramları vardır. YSA'lar, geleneksel işlemcilerden farklı şekilde işlem yapmaktadırlar. Geleneksel işlemcilerde, tek bir merkezi işlem elemanı her hareketi sırasıyla gerçekleştirir. YSA'lar ise herbiri büyük bir problemin bir parçası ile ilgilenen, çok sayıda basit işlem elemanlarından oluşmaktadır. En basit şekilde, bir işlem elemanı, bir girdiyi bir ağırlık kümesi ile ağırlıklandırır, doğrusal olmayan bir şekilde dönüşümünü sağlar ve bir çıktı değeri oluşturur. İlk bakışta, işlem elemanlarının çalışma şekli yanıltıcı şekilde basittir. Sinirsel hesaplamanın gücü, toplam işlem yükünü paylaşan işlem elemanlarının birbirleri arasındaki yoğun bağlantı yapısından gelmektedir.

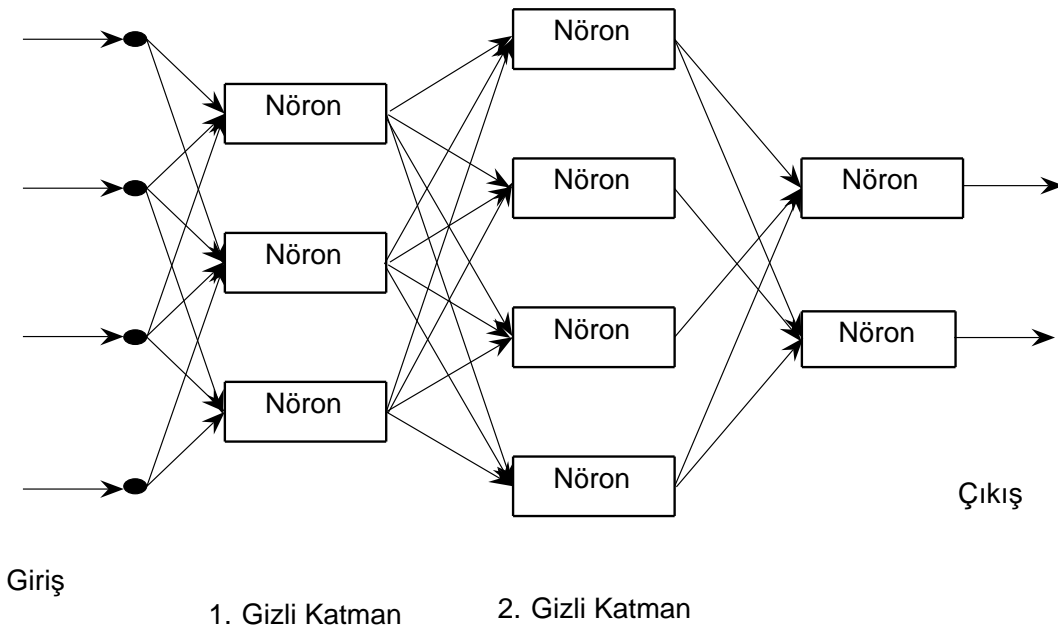
Çoğu YSA'da, benzer karakteristiğe sahip nöronlar tabakalar halinde yapılandırılırlar ve transfer fonksiyonları eş zamanlı olarak çalıştırılırlar. Hemen hemen tüm ağlar, veri alan nöronlara ve çıktı üreten nöronlara sahiptirler.

YSA'nın ana ögesi olan matematiksel fonksiyon, ağın mimarisi tarafından şekillenir. Daha açık bir şekilde ifade etmek gerekirse, fonksiyonun temel yapısını ağırlıkların büyüklüğü ve işlem elemanlarının işlem şekli belirler. YSA'ların davranışları, yani girişteki veriyi çıkıştaki veriyle nasıl ilişkilendirdikleri, öncelikli olarak nöronların transfer fonksiyonlarından, nasıl birbirlerine bağlandıklarından ve bu bağlantıların

ağırlıklarından etkilenir.

## 4.2 Çok Katmanlı Sinir Ağları

Bu çalışmada YSA, düşük bant genişliğine sahip konuşma parçalarına ait parametrelerden yüksek bant genişliğine sahip konuşma parametrelerini elde etmek için kullanılacaktır. Daha önce benzer bir çalışmaya rastlanılmadığı için nasıl bir sinir ağı seçileceği konusu aslında tam olarak bilinmemekle birlikte, bu çalışmada Şekil 4.3'deki gibi çok katmanlı bir YSA kullanılmıştır (Moon and Stirling, 2000). Kullanılan bu çok katmanlı yapı geriye yayılım algoritması kullanılarak eğitilmiştir. Daha sonra eğitilen bu yapı dar bantlı konuşma parametrelerinden geniş bantlı konuşma parametrelerinin elde edilmesi için kullanılmıştır.



Şekil 4.3. Çok katmanlı YSA

Çok katmanlı sinir ağları yapıları itibariyle çok basit fakat oldukça etkili ve kullanışlıdır. Bu ağlar birbirine ağırlıklandırılarak bağlanmış birçok nörondan oluşurlar. Genel olarak üç ana katmanları vardır:

- Giriş katmanı:
- Gizli katman(lar):

- Çıkış Katmanı:

Bu çalışmada kullanılan sinir ağırları öğreticiyle öğrenme yöntemi kullanılarak eğitilmiştir. Bu yöntemde geriye yayılım algoritması kullanılarak, hesaplanan çıkış ile istenen değer arasındaki hata mümkün olduğunca küçültülmeye çalışılır.

$N$  tane giriş ve çıkıştan oluşan bir eğitim setinde giriş verileri,  $x(n)$ ,  $n = 1, 2, \dots, N$  ve istenen çıkış verileri de  $d(n)$ ,  $n = 1, 2, \dots, N$  olsun. Bu durumda sinir ağırlarının o anda hesaplanan çıkış verilerinin  $y(n)$ ,  $n = 1, 2, \dots, N$  olduğu durumda çıkıştaki karesel hata,

$$\varepsilon(n) = \frac{1}{2} \| \mathbf{d}(n) - \mathbf{y}(n) \|^2 \quad (4.1)$$

şeklinde ve bu hatanın ortalaması da

$$\varepsilon_{avg} = \frac{1}{N} \sum_{n=1}^N \varepsilon(n) \quad (4.2)$$

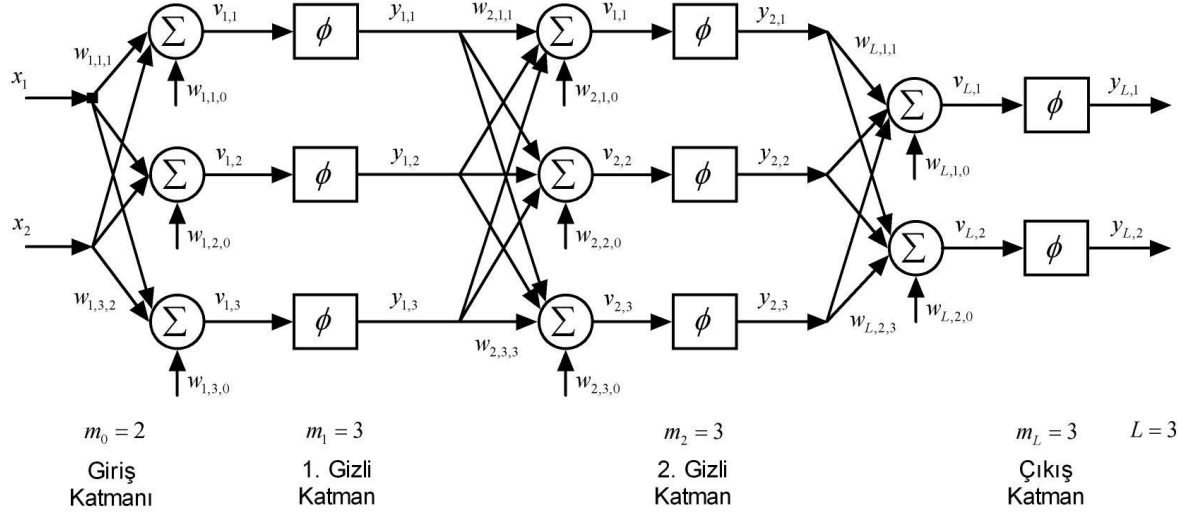
kullanılarak hesaplanır.

YSA, bu ortalama karesel hatayı en küçültecek biçimde nöronlar arasındaki bağlantıları değiştirerek eğitilirler. Eğitilen bu sinir ağırları daha sonra tasarlanan sistem için kullanılabilirler.

### 4.3 Geri Yayılım Algoritması

Geri yayılım algoritması bayır inişi (steepest descent - gradient descent) temelli bir eğitim algoritmasıdır. Bu algoritmayı anlatmak için öncelikle Şekil 4.4'deki simge ve gösteriliş biçimlerinin açıklanması gerekmektedir (Moon and Stirling, 2000).

- $n$ , index numarası giriş ve çıkış verilerinin yerini göstermektedir.  $x(n)$  ve  $d(n)$  sırasıyla  $n$ 'inci giriş ve çıkışı göstermektedir.
- $k$  (üstsimge olarak) yineleme sayısını göstermektedir.
- $l$  katman sayısını göstermektedir,  $l = 0$  giriş katmanını ve  $l = L$  ise çıkış katmanını belirtmektedir.



Şekil 4.4. Çok katmanlı sinir ağı notasyonu

- $i$ , girişi (nöron girişlerini) belirtmek için,  $j$ , ise nöron numarasını belirtmek için kullanılmaktadır.
- $m_l$   $l$ 'inci katmandaki nöron sayısını belirlemek için kullanılır. Burada  $m_L$  çıkış katmanındaki  $m_0$  ise giriş katmanındaki nöron sayısını belirtir.
- $w_{l,j,i}$   $l$ 'inci katmandaki  $j$ 'inci nörona,  $(l-1)$ 'inci (bir önceki) katmandaki  $i$ 'inci nörondan gelen ağırlığı belirtir. Eğitim sırasında  $n$ 'inci girişe göre ağırlıklardaki değişim  $\Delta w_{l,j,i}(n)$  biçiminde gösterilebilir ve bu durumda  $k$ 'inci yinelemedeki ağırlık  $w_{l,j,i}^{[k]}$  olur.  $l$ 'inci katmandaki ağırlıklar  $m_l \times (1 + m_{l-1})$  boyutunda bir matris ile ifade edilebilir. Bu matristeki kolon sayısının bir fazla olması her bir nörona sabit değer olarak 1'in giriş olarak kabul edilmesindedir.
- $y_{l,j}(n)$ ,  $x(n)$  girişi için  $l$  katmanındaki  $j$ 'inci nöronun çıkışıdır.  $l$ 'nin belirtilmediği durumda çıkış katmanı kastedilmektedir.  $y_{l,0} = 1$  gizli katmanlardaki sabiti belirtmek için kullanılır. Bu çıkışlar aynı zamanda vektörel olarak da gösterilebilirler. Örneğin çıkış katmanındaki nöronlar için çıkış vektörü,

$$\mathbf{y}(n) = \begin{bmatrix} y_1(n) \\ y_2(n) \\ \vdots \\ y_{m_L}(n) \end{bmatrix} \quad (4.3)$$

şeklinde ve  $l$  katmanındaki nöronlar için çıkış vektörü ise

$$\mathbf{y}_l(n) = \begin{bmatrix} 1 \\ y_{l,1}(n) \\ y_{l,2}(n) \\ \vdots \\ y_{l,m_l}(n) \end{bmatrix} \quad (4.4)$$

biçiminde gösterilebilir.

- $l$  katmanındaki  $j$ 'inci nöronun  $n$  girişine göre faaliyeti işlevi, girişlerin ve ağırlıkların doğrusal kombinasyonu olarak,

$$v_{l,j}(n) = \sum_{i=0}^{m_{l-1}} w_{l,j,i} y_{l-1,i}(n) \quad (4.5)$$

biçiminde tanımlanabilir.

- Herbir nöronun çıkışı faaliyetin doğrusal olmayan bir işlevdir ve

$$y_{l,j} = \phi_{l,j}(v_{l,j}(n)) \quad (4.6)$$

biçiminde tanımlanabilir. Bu doğrusal olmayan işlev aslında herbir nöron için farklı olabilir ancak genelde hepsi için tek bir işlev kullanılır ve bu durumda  $\phi_{l,j}(\cdot) = \phi(\cdot)$  olur.

- Çıkış katmanı için istenen değere,  $d_{L,j}(d_j)$ , eğitim kümesinden doğrudan ulaşılabilir ve tüm çıkış değerleri

$$\mathbf{d}(n) = \begin{bmatrix} d_1(n) \\ d_2(n) \\ \vdots \\ d_{m_L}(n) \end{bmatrix} \quad (4.7)$$

biçiminde vektörel olarak gösterilebilir. Gizli katmanlar için istek değerleri ise geri yayılım algoritmasıyla bulunabilir.

- $l$  katmanındaki bir nöron için istenen değer ile nöron çıkış değeri arasındaki hata

$$e_{l,j}(n) = d_{l,j}(n) - y_{l,j}(n) \quad (4.8)$$

biçiminde tanımlanır. Bu durumda Eş. 4.1'deki hata

$$\varepsilon(n) = \varepsilon_L(n) = \frac{1}{2} \sum_{j=1}^{m_L} e_{L,j}^2(n) = \frac{1}{2} \sum_{j=1}^{m_L} [d_{L,j}(n) - y_{L,j}(n)]^2 \quad (4.9)$$

biçiminde ve ortalama hata ise

$$\varepsilon_{av} = \frac{1}{N} \sum_{n=1}^N \varepsilon_L(n) \quad (4.10)$$

şeklinde ifade edilebilir.

Eğitim algoritmasındaki temel amaç ortalama hatayı en küçültmektir. Ortalama hata ise ağdaki tüm ağırlıkların bir işlevidir ve dolayısıyla bu hatanın küçültülmesi için tüm ağırlıkların her bir yinelemede (iteration) güncellenmesi gerekmektedir. Bu güncelleme işlemi ise  $n$ 'inci giriş için bayır inişi yöntemini kullanarak

$$w_{l,j,i}^{[k+1]} = w_{l,j,i}^{[k]} - \mu \frac{\partial \varepsilon(n)}{\partial w_{l,j,i}^{[k]}} \quad (4.11)$$

biçiminde yapılır ve burada  $\mu$  öğrenme parametresi olarak tanımlanır. Bu güncelleme işlemi her bir giriş kümesi için  $n = 1, 2, 3, \dots, N$  ağırlıklar bir sonuca yakınsayınca kadar ve muhtemelen birden fazla kere devam eder.

Bu algorithmada öğrenme parametresinin uygun şekilde seçilmesi kritik noktalardan bir tanesidir. Eğer  $\mu$  olması gerekenden büyük bir değer seçilirse algorithmadaki adımlar çok büyük olur ve ağırlıklar bir yere yakınsayamaz; tam tersi durumda ise adımlar çok küçük olur ve bölgesel en uç noktaları geçemezler ve dolayısıyla ağırlıklar yine olması gereken yere yakınsayamazlar. Diğer bir kritik nokta ise kısmi türevlerin

$$\frac{\partial \varepsilon(n)}{\partial w_{l,j,i}^{[k]}} \quad (4.12)$$

alınmasıdır. Çıkış katmanı ile saklı katmanlardaki türevlerin alınmasında birtakım farklılıklar vardır, çünkü saklı katmanlarda olamayan istenen değer,  $d_j(n)$  çıkış katmanında mevcuttur.

### 4.3.1 Çıkış Katmanındaki Türevin Hesaplanması

Çıkış katmanındaki ( $l = L$ ) türev hesaplanırken Eş. 4.12, zincir kuralı uygulanarak kısmi türevler cinsinden,

$$\frac{\partial \varepsilon(n)}{\partial w_{L,j,i}^{[k]}} = \frac{\partial \varepsilon(n)}{\partial e_{L,j}(n)} \frac{\partial e_{L,j}(n)}{\partial y_{L,j}(n)} \frac{\partial y_{L,j}(n)}{\partial v_{L,j}(n)} \frac{\partial v_{L,j}(n)}{\partial w_{L,j,i}^{[k]}} \quad (4.13)$$

biçiminde ifade edilebilir. Ardından bütün kısmi türevler sırayla

- Birinci kısmi türev Eş. 4.1 ve Eş. 4.10 kullanılarak

$$\frac{\partial \varepsilon(n)}{\partial e_{L,j}(n)} = e_{L,j}(n) \quad (4.14)$$

- İkinci kısmi türev Eş. 4.8 kullanılarak

$$\frac{\partial e_{L,j}(n)}{\partial y_{L,j}(n)} = -1 \quad (4.15)$$

- Üçüncü kısmi türev Eş. 4.6'den faydalanarak

$$\frac{\partial y_{L,j}(n)}{\partial v_{L,j}(n)} = \phi'_{L,j}(v_{L,j}(n)) \quad (4.16)$$

şeklinde ifade edilir. Buradaki  $'$ , sembolü argümana göre türevi belirtmektedir, ancak bu türev faaliyet işlevi,  $\phi_{l,j}(\cdot)$  seçilene kadar alınmamıştır.

- Son olarak dördüncü türev ise Eş. 4.5 yardımıyla

$$\frac{\partial v_{L,j}(n)}{\partial w_{L,j,i}^{[k]}} = y_{L-1,i}(n) \quad (4.17)$$

biçiminde bulunur.

Tüm bu kısmi türevler kullanılarak

$$\frac{\partial \varepsilon(n)}{\partial w_{L,j,i}^{[k]}} = -e_{L,j}(n) \phi'_{L,j}(v_{L,j}(n)) y_{L-1,i}(n) \quad (4.18)$$

elde edilir. Eş. 4.13'deki ilk üç terim birleştirilerek

$$\frac{\partial \varepsilon(n)}{\partial w_{L,j,i}^{[k]}} = \delta_{l,i}(n) \frac{\partial v_{L,j}(n)}{\partial w_{L,j,i}^{[k]}} \quad (4.19)$$

biçiminde yazılabilir. Burada  $\delta_{l, i}$  bölgesel gradyan (local gradient) olarak adlandırılır ve

$$\delta_{l, i}(n) = \frac{\partial \varepsilon(n)}{\partial v_{l, j}(n)} \quad (4.20)$$

biçiminde tanımlanır. Eş. 4.18 bir önceki bölgesel gradyan kullanılarak

$$\frac{\partial \varepsilon(n)}{\partial w_{L, j, i}^{[k]}} = \delta_{l, i}(n) y_{L-1, i}(n) \quad (4.21)$$

biçiminde tekrar yazılabilir ve bu eşitlikte bulunan çıkış katmanındaki bölgesel gradyan ise

$$\delta_{l, i}(n) = -e_{L, j}(n) \phi'_{L, j}(v_{L, j}(n)) \quad (4.22)$$

kullanılarak hesaplanabilir.

### 4.3.2 Gizli Katmanındaki Türevin Hesaplanması

Hatanın türevi bölgesel gradyan kullanılarak

$$\frac{\partial \varepsilon(n)}{\partial w_{L, j, i}^{[k]}} = \delta_{l, i}(n) \frac{\partial v_{L, j}(n)}{\partial w_{L, j, i}^{[k]}} = \delta_{l, i}(n) y_{L-1, i}(n) \quad (4.23)$$

biçiminde yazılabilir ve buradaki bölgesel gradyan aşağıdaki gibi hesaplanır.

$$\begin{aligned} \delta_{l, i}(n) &= \frac{\partial \varepsilon(n)}{\partial v_{l, j}(n)} \\ &= \sum_{p=1}^{m_{l+1}} \frac{\partial \varepsilon(n)}{\partial v_{l+1, p}(n)} \frac{\partial v_{l+1, p}(n)}{\partial v_{l, j}(n)} \\ &= \sum_{p=1}^{m_{l+1}} \delta_{l+1, i}(n) \frac{\partial v_{l+1, p}(n)}{\partial v_{l, j}(n)} \end{aligned} \quad (4.24)$$

Burada

$$v_{l+1, p}(n) = \sum_{i=0}^{m_l} w_{l+1, p, i} \phi_{l, i}(v_{l, j}(n)) \quad (4.25)$$



şeklinde tanımlanabilir. Eş. 4.24'deki kısmi türev de

$$\frac{\partial v_{l+1, p}(n)}{\partial v_{l, j}(n)} = w_{l+1, p, i} \phi'_{l, i}(v_{l, j}(n)) \quad (4.26)$$

kullanılarak bulunur ve bu kısmi türev Eş. 4.24'de yerine konulursa gizli katmanlardaki bölgesel gradyan da aşağıdaki gibi hesaplanabilir:

$$\delta_{l, i}(n) = \phi'_{l, i}(v_{l, j}(n)) \sum_{i=0}^{m_l} \delta_{l+1, p}(n) w_{l+1, p, i}(n) \quad (4.27)$$

Yukarıdaki eşitliklerde geçen  $\phi_{l, j}(\cdot)$ , sinir ağları için gerekli olan doğrusal olmayan bir işlevdir. Genelde, herhangi bir monotonik artan işlev kullanılabilir. Ancak genelde bu işlev için

$$\phi(v) = \frac{1}{1 + e^{-v}} \quad (4.28)$$

biçiminde tanımlanan sigmoid işlev kullanılır. Çünkü bu işlevin türevi çok kolay hesaplanabilir.  $y = \phi(v)$  olduğunda türev aşağıdaki gibi bulunur:

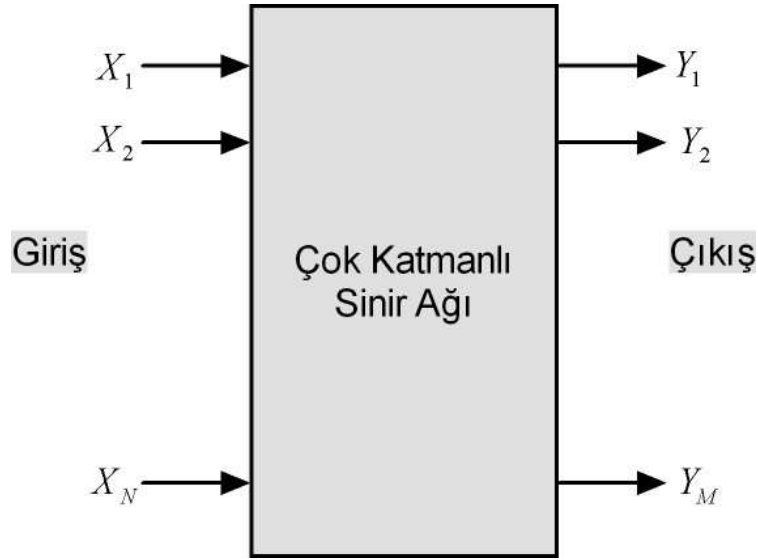
$$\phi'(v) = y(1 - y) \quad (4.29)$$

#### 4.4 Telefon Kalitesindeki Konuşmanın İlgel Zarfının Genişletilmesi

Bu çalışmada esas olarak 8 kHz ile örneklenmiş 3.4 kHz ile bant sınırlı (telefon kalitesi) konuşma parçalarının bant genişliklerini 7 kHz'e çıkarılarak kalitelerinin artırılması hedeflenmiştir. Telefon kalitesindeki konuşma parçaları genel olarak 10. dereceden TK modellendiği için bu çalışmada da bu genel uygulamaya uyulmuştur. Yüksek kaliteli konuşma parçaları ise 20. dereceden TK modellenmiştir. Dolayısıyla Şekil 4.5'te görülen ve sinir ağlarının giriş ve çıkışlarını gösteren basit yapının girişleri 10. dereceden TK parametrelerinden elde edilmiş ÇİF'ler, çıkışları ise 20. dereceden TK parametrelerinden elde edilen 20 tane ÇİF'in son on tanesidir.

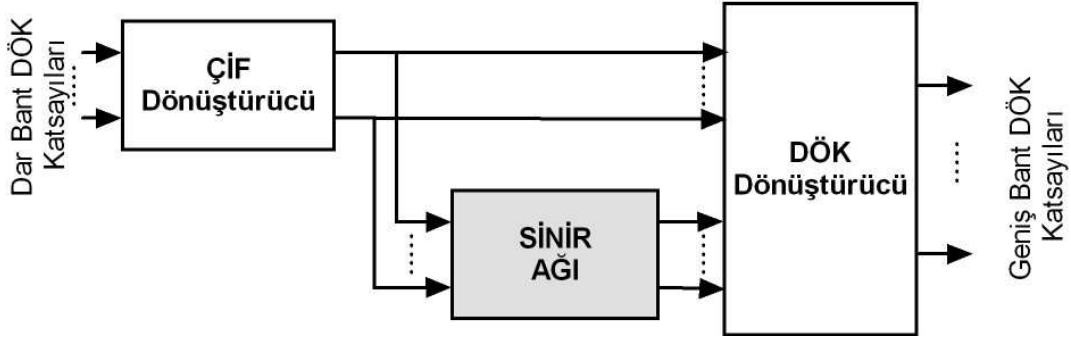
Burada son on tane ÇİF'in çıkış olarak kullanılmasındaki sebep ilk 10 tanesinin dar bantlı konuşmadan elde edilenlerle eşit olarak kabul edilmelerindedir. Çıkışların son 10 tane ÇİF olarak seçilmesi zaman zaman problemlere yol açmaktadır. Bu

problemlerin en önemlisi sinir ağından elde edilen en küçük ÇİF'in dar bantlı konuşmaya ait olan sonuncu parametreden küçük olması durumudur. Bu durumda elde edilen çıkışların ilk 2 tanesi atılır ve geniş bantlı konuşma 18. dereceden TK modellenmiş olarak kabul edilir. Eğer sinir ağının çıkışı 20 tane geniş bant parametrelerini kestirecek biçimde seçilirse böyle bir problemle karşılaşılmaz.



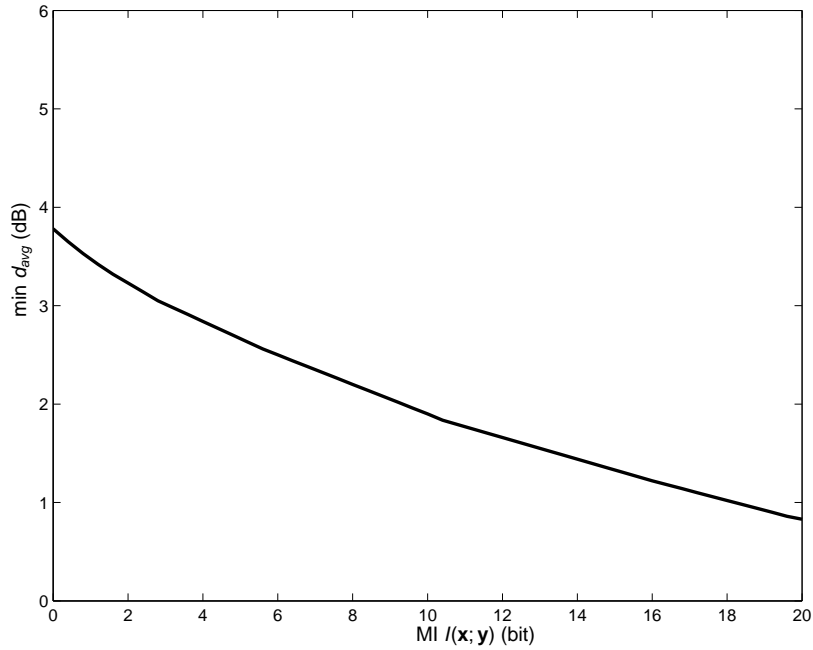
Şekil 4.5. Çok Katmanlı YSA

Şekil 4.5'te görülen basit yapı sadece Şekil 4.3'teki gibi çok katmanlı sinir ağının giriş ve çıkışlarını göstermektedir. Burada kullanılan çok katmanlı sinir ağında 10 nörondan oluşan giriş ve yine 10 nörondan oluşan çıkış katmanının yanında bir tane de 24-30 tane nörondan oluşan gizli katman bulunmaktadır. Çalışma sırasında bu yapı önce eğitilmiş daha sonra geniş bant konuşma parametrelerinin elde edilmesi için kullanılmıştır. Bu ağın eğitimi için toplam 40 dakikalık bir konuşma veri tabanı kullanılmıştır. Bu veri tabanındaki konuşmalar 25 ms.'lik %50 örtüşen pencerelere ayrılarak 96000 tane eğitim vektörü elde edilmiştir. Bu eğitim setinin kullanıldığı eğitimin sonunda kullanılan eğitim parametrelerinden bağımsız olarak ( $\mu = 0.8$ ,  $\mu = 0.5$ ,  $\mu = 0.2$ ) ortalama karesel hata yaklaşık olarak 0.0035 bulunmuştur. Öğrenme parametresinin değişimi sadece en küçük hataya ulaşma süresini değiştirmiştir. Bu veri tabanı ile eğitilen sinir ağı, Şekil 4.6'teki gibi bir yapıda, dar bantlı konuşma parametrelerinin genişletilmesinde kullanılmıştır. Bu yöntemle genişletilen ve eğitim setinde olmayan konuşma parçaları için EK.1'de anlatılan ortalama izgel bozulum yaklaşık olarak 3.8 dB olarak bulunmuştur.



Şekil 4.6. Konuşma parametrelerinin sinir ağları kullanarak genişletilmesi

Ek.2'de anlatılan izgel bozulumun bilgi teorisi bakış açısıyla bulunan alt sınır eğrisi Şekil 4.7'de görülebilir (Jax, 2004). Bu eğri karşılıklı bilgi miktarına (mutual information) göre izgel bozulumun üst sınırını göstermektedir. 3.4 kHz'e bant sınırlı olan telefon kalitesindeki konuşma parametreleri ile 7 kHz'e bant sınırlı olan konuşma parametreleri karşılıklı olarak 2.3 – 2.4 bit bilgi içerirler (Jax, 2004). Bu durumda izgel zarflar arasında oluşan bozulumun yaklaşık 3 dB'den küçük olamayacağı sonucuna varılır.

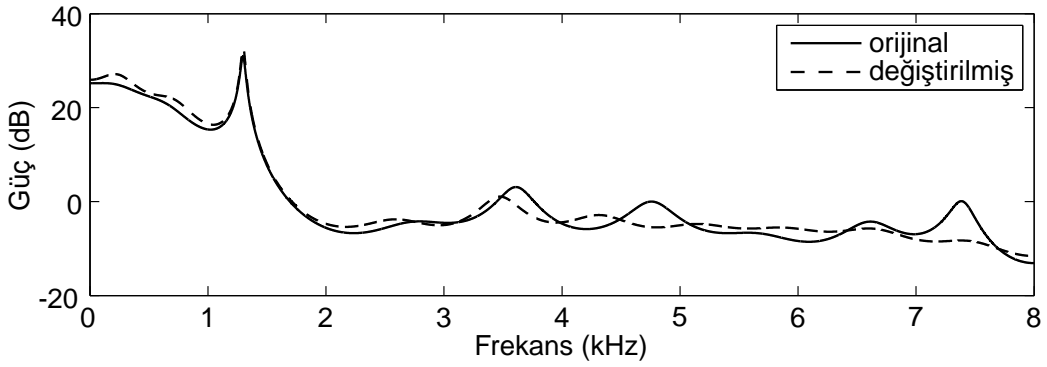


Şekil 4.7. İzgel bozulumun alt sınır eğrisi

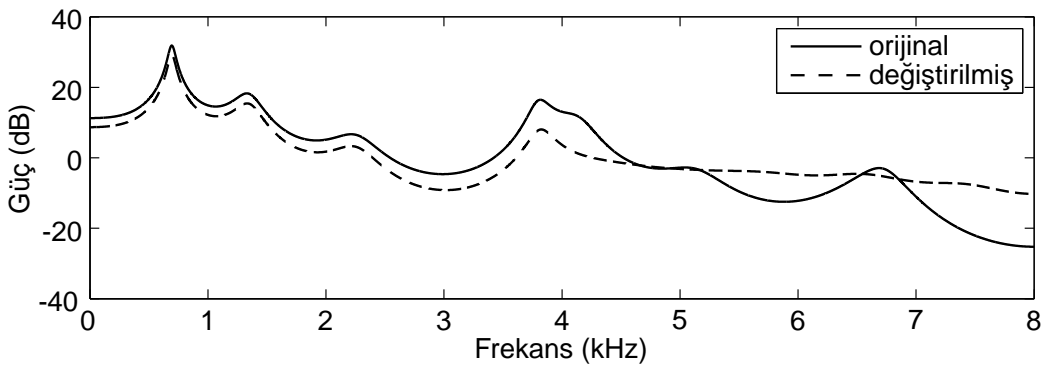
Bu koşullar altında konuşmalar için buluna 3.8 dB'lik izgel bozulum teorik sınırdan büyük olmakla birlikte olabilirlik sınırları dahilinde oldukları düşünülmektedir. Eğitim

setindeki konuşma parçaları için ortalama izgel bozulum yaklaşık olarak 3 dB bulunmuştur. Bu da teorik sınırın hemen üst sınırına karşılık gelmektedir. Şekil 4.8'de 3.4 kHz'e bant sınırlı konuşmadan YSA kullanarak genişletilen ve 7 kHz'e bant sınırlı olan orjinal konuşmaya ait izgel zarflar görülmektedir. Şekil 4.8.a'da izgel bozulum 1.7 dB Şekil 4.8.b'de ise izgel bozulum 5 dB'dir.

Ayrıca Bölüm 2'de de bahsedildiği gibi konuşma, sesli ve sessiz olmak üzere farklı özelliklere sahip iki ana türe ayrılabilir. Bundan yola çıkarak izgel zarfların ve model parametrelerinin de iki türlü olduğu düşünülebilir. Bu sebeplerden dolayı bir yerine, sesli ve sessiz konuşma parametreleri için ayrı ayrı olmak üzere aynı özelliklerde iki tane sinir ağı kullanılarak konuşma parametreleri genişletilmeye çalışılmıştır. Bu modelde kullanılan sinir ağları tek olarak kullanılan sinir ağıyla aynı özelliklere sahiptirler. Bu yöntemle aynı eğitim setiyle eğitilen sinir ağlarında ortalama izgel bozulum 0.1 dB iyileşerek 3.7 dB olarak bulunmuştur.



a. İzgel Bozulum 1.7 dB



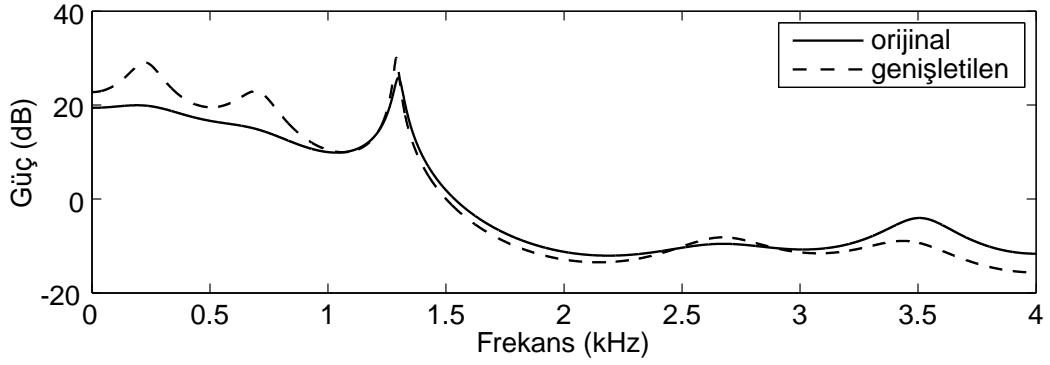
b. İzgel Bozulum 5 dB

Şekil 4.8. 7 kHz'e bant sınırlı orijinal ve 3.4 kHz'e bant sınırlı izgel zarftan YSA kullanılarak genişletilen izgel zarflar

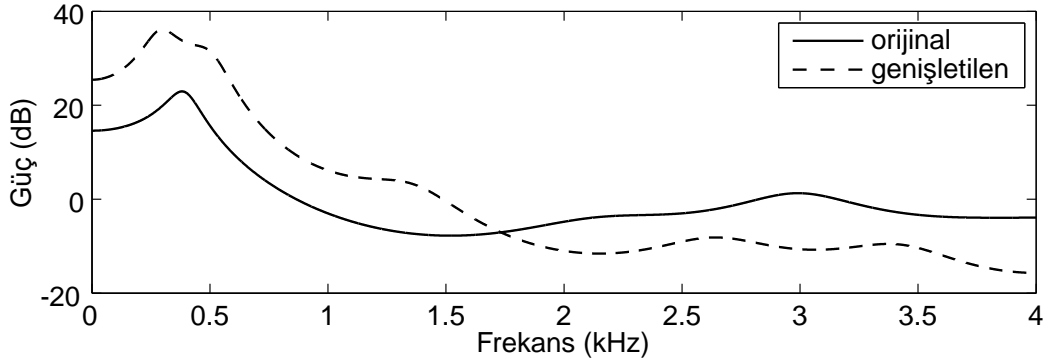
#### 4.5 4. kHz İle Örneklenmiş Konuşmanın İzgel Zarfının Genişletilmesi

İzgel zarf genişletme algoritmaları genel olarak telefon kalitesindeki konuşmalar için tasarlanmakla birlikte daha dar bantlı konuşma parçaları için de kullanılabilirler. Burada örnek olarak 4 kHz ile örneklenmiş 1.8 kHz'e bant limitli olan konuşmaları 3.5 kHz bant genişliğine çıkarmak için izgel zarfları genişletilmeye çalışılacaktır. Bunun için öncelikle 1.8 kHz'e bant genişliği olan konuşma 6. dereceden TK modellenip parametreleri bulunmuş ve bunlar Şekil 4.5'de görülen sinir ağının girişleri olarak alınmışlardır. Sinir ağının çıkışı ise telefon kalitesine sahip konuşmaya ait 10. dereceden parametrelerin son 4 tanesidir. Bu durumda giriş katmanı 6 nöron, çıkış katmanı 4 nöron ve gizli katman ise 12-16 nöron olacak biçimde seçilmiştir. Bu sinir ağı yapısı da bir önceki bölüme benzer bir şekilde ve aynı eğitim setiyle eğitilmiş ve ortalama bozulum yaklaşık olarak 3.9dB bulunmuştur.

Çalışmanın bu ve bundan önceki bölümünde dar bantlı izgel zarflar genişletilerek geniş bantlı konuşmaya ait izgel zarflar elde edilmeye çalışılmıştır. Bundan sonraki bölümde ise özellikle sinir ağları kullanılarak genişletilen izgel zarflar kullanılarak konuşmanın bant genişliğini artırarak kalitesi yükseltilmeye çalışılacaktır.



a. İzgel Bozulum 2.7 dB



b. İzgel Bozulum 8 dB

Şekil 4.9. 3.4 kHz'e bant sınırlı orijinal ve 1.8 kHz'e bant sınırlı izgel zarftan YSA kullanılarak genişletilen izgel zarflar

## 5. KONUŞMA KALİTESİNİN ARTIRILMASI

Çalışmanın bundan önceki bölümlerinde dar bantlı konuşmanın izgel zarfından geniş bantlı konuşmaya ait izgel zarf elde edilmeye çalışılmıştır. Bu bölümünde genişletilen zarflar kullanılarak daha yüksek frekans bileşenlerine sahip olan daha kaliteli konuşmalar elde edilmeye çalışılacaktır. Bu işlem için çalışmada başlıca üç yöntem kullanılmıştır:

- Genişletilen izgel zarfın gürültüyle doldurulması.
- Genişletilen izgel zarfın sinüslerle doldurulması.
- Kaynak Süzgeç Modeli kullanılarak kalitenin artırılması.

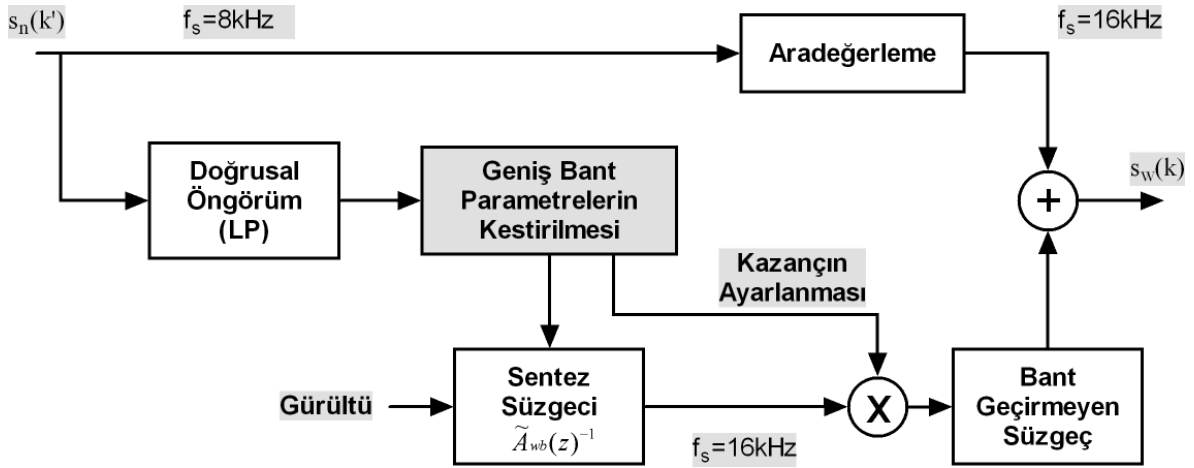
Bu bölümde yapılan tüm uygulamalarda izgel zarf genişletme işlemi için sinir ağları yöntemi kullanılmıştır. Bu yöntemin kullanılmasındaki sebep diğer izgel zarf genişletme yöntemlerine göre daha özgün olmasıdır. Ayrıca bölümün sonlarında, kaynak-süzgeç modeli kullanılarak konuşma kalitesinin artırılması sırasında gerekli olan, geniş bantlı konuşma için uyarım sinyalinin ve kazancın bulunma yöntemleri hakkında bilgi verilmiştir.

### 5.1 İzgel Zarfın Gürültüyle Doldurulması

Bu yöntem temel olarak konuşmanın yüksek frekanslı bölgelerinin harmonik bileşenlerden daha çok gürültü bileşenlerine sahip olduğu tezine dayanır ve bu tez konuşmanın sessiz kısımları için tamamen doğru olmakla beraber sesli kısımlar için tamamen doğru değildir. Konuşmanın sesli kısımlarının yaklaşık olarak 5.5 kHz'e kadar harmonik bileşenleri vardır ve ancak bu frekanstan sonra gürültü bileşenleri baskın olmaya başlarlar. Bu yöntemle elde edilen yüksek bant genişliğine sahip konuşmanın kalitesi dar bantlı konuşmadan biraz daha fazladır. Bu yöntemin genel işleyiş şeması Şekil 5.1'de görülebilir.

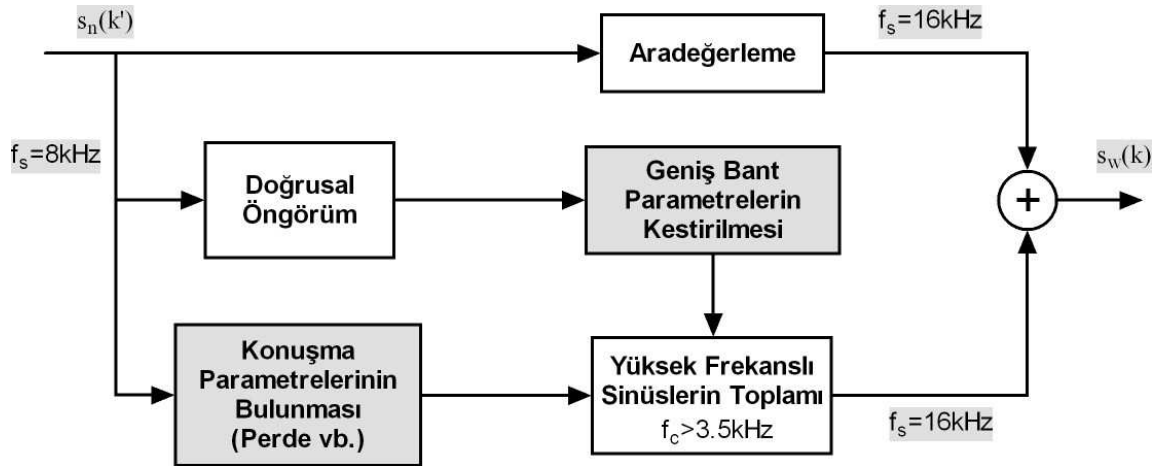
### 5.2 İzgel Zarfın Sinüslerle Doldurulması

Bu yöntem konuşmanın sinüslerin toplamı olarak ifade edilebileceği temeline dayanır ve genişletilen izgel zarfın altı sadece sinüslerle doldurulur. Sesli konuşma parçaları için perde frekansının katlarında sinüsler, sessiz konuşma parçaları için ise



Şekil 5.1. Dar bantlı konuşmanın gürültüyle genişletilmesi

rastgele evreye sahip 100 Hz'in katlarında sinüsler düşük bant genişliğindeki konuşmaya eklenir. Her iki durumda da sinüslerin genlikleri genişletilen izgel zarf örneklenerek elde edilir. Bu yöntemle elde edilen konuşmanın kalitesi gürültüyle doldurulan izgel zarfa oranla daha fazladır. Sinüs modelinde perde frekansının tam katlarında sinüsler kullanılacağı gibi dar bantlı konuşmadan elde edilen sinüs frekansları da yüksek frekans bandına kaydırılarak kullanılabilir. Bu durumda da sinüslerin genlikleri genişletilen izgel zarftan elde edilir.

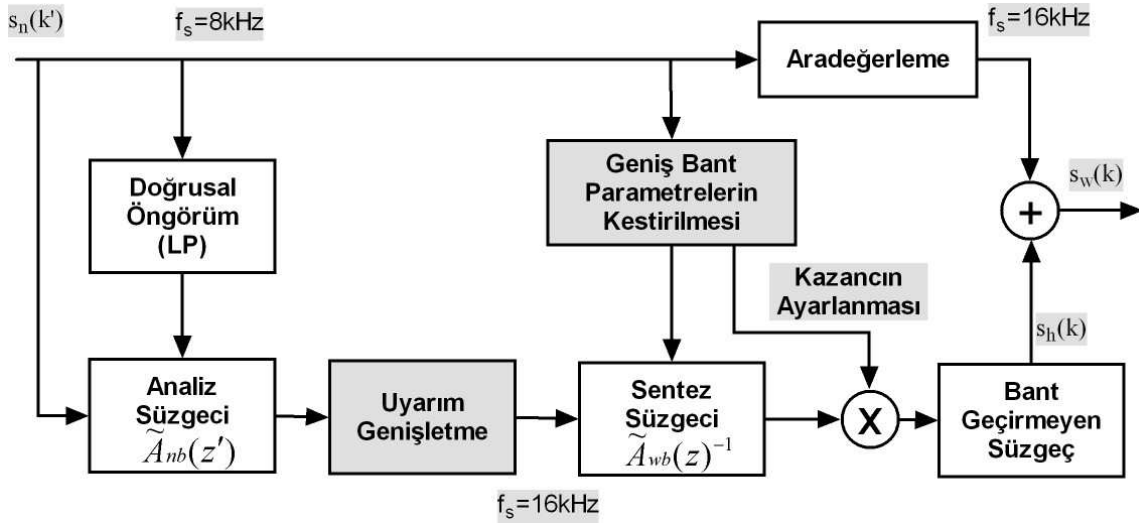


Şekil 5.2. Dar bantlı konuşmanın periyodik sinüslerle genişletilmesi



### 5.3 Kaynak-Süzgeç Modeliyle Kalitenin Artırılması

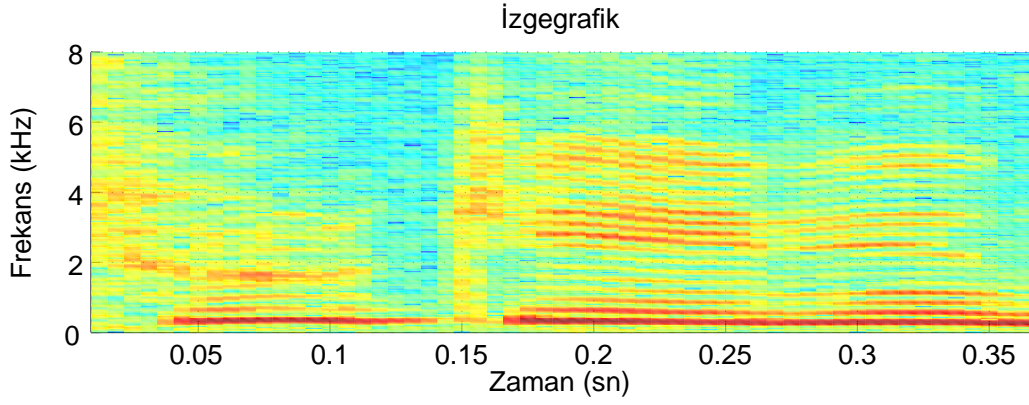
Bu yöntemde konuşmanın Kaynak-Süzgeç modeliyle oluşturduğu varsayılmıştır. Şekil 5.3'te genel işleyiş seması görülen bu yöntemde önce düşük bant genişliğine sahip konuşmaya ait TK parametreleri bulunmuş ve konuşma ters süzgeçten geçirilip uyarım sinyali de ayrıca elde edilmiştir. Daha sonra zarf ve aynı zamanda uyarım sinyali Bölüm 5.3.1'deki yöntemler kullanılarak genişletilmiştir. Ardından uyarım sinyali ve genişletilen TK parametreleri kullanılarak geniş bantlı konuşma sinyali elde edilmiştir. Ancak zaten düşük bant genişliğine sahip orijinal konuşmaya sahip olduğu için bu yüksek bant genişliğine sahip konuşmanın sadece yüksek frekanslı bölgeleri bir yüksek geçirgen bir süzgeçle süzülüp elde edilmiştir. Bundan sonra sadece yüksek frekans bileşenleri dar bantlı konuşmaya eklenip yüksek kaliteli konuşma elde edilmiştir. Şekil 5.4'de bu yöntemle genişletilmiş ve orijinal konuşmalara ait izgegrafikler (spectrogram) görülmektedir.



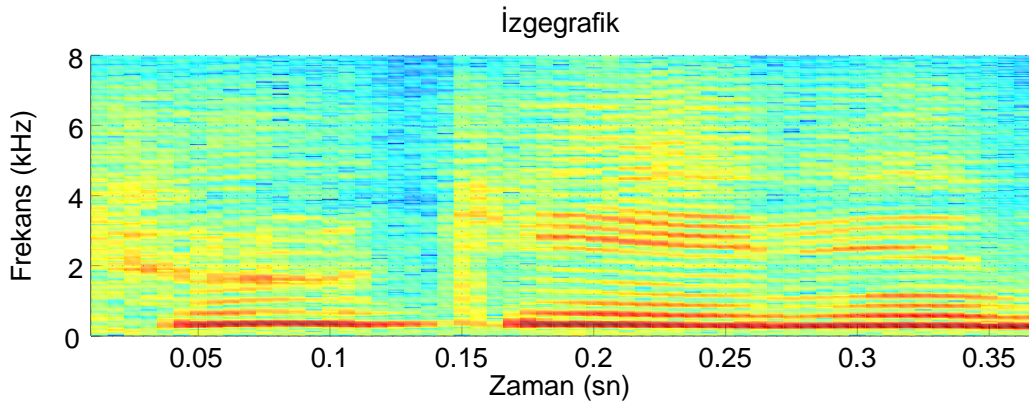
Şekil 5.3. Dar bantlı konuşmanın kaynak-süzgeç modeliyle genişletilmesi

#### 5.3.1 Uyarım Sinyalinin Genişletilmesi

Konuşma kalitesi artırmak için eklenecek yüksek frekans bileşenlerinin elde edilmesi için TK parametrelerinin yanında uyarım sinyalinin de genişletilmesi gerekmektedir. Bunun için öncelikle uyarım sinyali hakkında önbilgiye sahip olmak gerekmektedir. Şekil 5.5'de 16 kHz ile örneklenmiş konuşma parçası ve izgegrafığı (spectrogram) görülebilir. Buradan da anlaşılacağı gibi sesli konuşmanın 5.5 kHz'ten sonra hemen



a. Orijinal konuşma

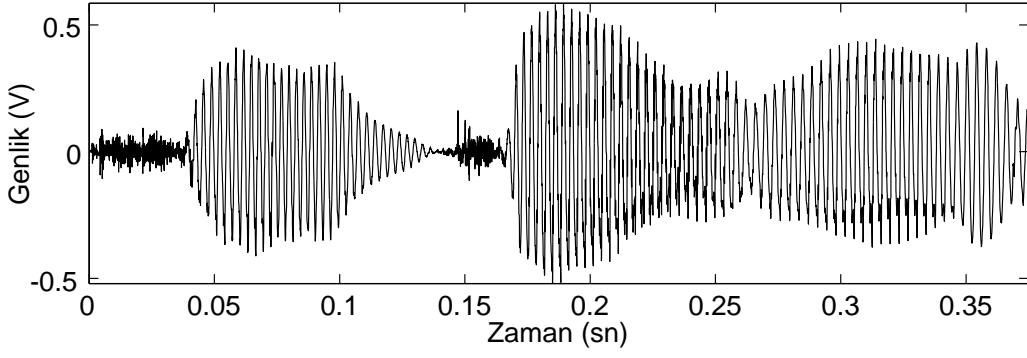


b. Genişletilen konuşma

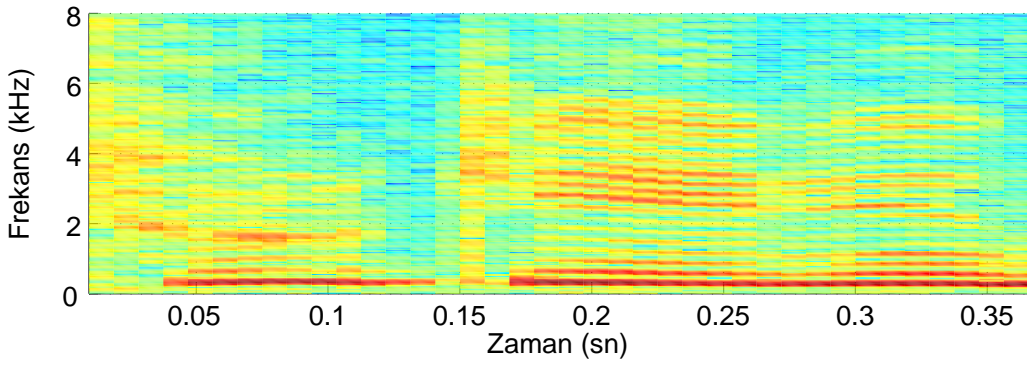
Şekil 5.4. 16 kHz ile örneklenmiş konuşma parçaları

hemen hiç harmonik bileşeni yoktur. Ayrıca uyarım sinyalinin izgel zarfının da tüm frekans bandı boyunca sabit olduğu düşünülebilir.

Yüksek bant genişliğine sahip izgel zarfı elde etmek için, orijinal konuşmadan elde edilen uyarım sinyaline yüksek frekanslı bileşenlerin toplanması gerekmektedir. Bunun için öncelikle dar bantlı uyarım sinyali yukarı örneklenip (upsample) aradeğerlenerek örnek sayısı artırılmıştır. Daha sonra, bu sinyale yüksek frekans bileşenlerine sahip sinyal eklenerek geniş bantlı uyarım sinyali elde edilmiştir. Bu çalışma sırasında geniş bantlı uyarım sinyali birbirine benzeyen iki yöntemle elde edilmiştir. Birinci yöntemde geniş bantlı uyarım sinyali, dar bantlı uyarım sinyalinin yüksek frekanslı harmonik sinüslerle toplanmasıyla elde edilmiştir. İkinci yöntemde ise dar bantlı uyarım sinyali için bulunan sinüsler yüksek frekans bandına kaydırılarak dar bantlı uyarım sinyaline eklenmiş ve böylece geniş bantlı uyarım sinyali elde edilmiştir.



a. Konuşma parçası 'Türkiye'nin'



b. Konuşma parçasının izgeografiği

Şekil 5.5. 16 kHz ile örneklenmiş konuşma ve izgeografiği

### 5.3.1.1 Uyarım Sinyalinin Harmonik Sinüslerle Genişletilmesi

Konuşmanın 2.3'de anlatıldığı gibi değişik frekans, genlik ve fazdaki sinüslerin toplamı şeklinde ifade edilebildiğinden yola çıkıldığında uyarım sinyalinin de sinüslerin toplamı biçiminde ifade edilebileceği kolayca görülebilir. Konuşmanın sesli olduğu durumda sinüslerin frekansları, perde frekansının tam katları olarak ifade edilebilir. Uyarım sinyalinin yüksek frekanslı bileşenleri genişletileceği frekans bölgesinde, perde frekansının katlarındaki sinüsler toplanarak

$$E_H[n] = \sum_{l=\ell_{min}}^{\ell_{max}} A_H^k \cos(l \omega_0^k n + \theta^k) \quad (5.1)$$

biçiminde elde edilir. Burada  $\omega_0^k$ ,  $A_H^k$  ve  $\theta^k$  sırasıyla k çerçevesindeki perde frekansını, genliği ve evreyi belirtir. Ayrıca yukarıdaki eşitlikteki sinüslerin frekansı

$$W_{nb} \leq \ell_{min} \omega_0^k \leq l \omega_0^k \leq \ell_{max} \omega_0^k \leq W_{hb} \quad (5.2)$$

olacak biçimde seçilmelidir ve burada  $w_{nb}$  ve  $w_{hb}$  sırasıyla dar ve geniş bantlı konuşmaların bant genişlikleridir.

Sessiz konuşma parçaları için perde frekansı 100 Hz olarak alınabilir (McAulay and Quatieri, 1986). Uyarım sinyali için genlik, sabit kabul edilebilir ve dar bantlı kısmın ortalaması alınarak bulunabilir; evre ise yüksek frekanslı bölgeler için  $[-\pi, \pi]$  aralığında rastgele seçilebilir. Daha sonra elde edilen yüksek frekans bileşenli sinyal dar bantlı uyarım sinyaline eklenerek geniş bantlı uyarım sinyali

$$E_{EX}^k[n] = E_L[n]^k + E_H[n]^k \quad (5.3)$$

elde edilir.

### 5.3.1.2 Uyarım Sinyalinin Sinüslerle Genişletilmesi

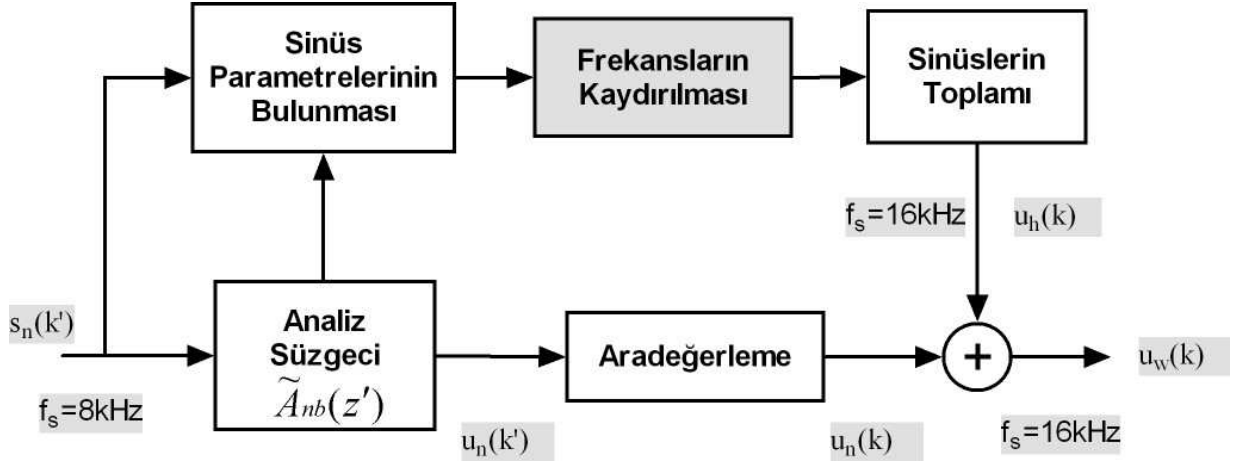
Geniş bantlı uyarım sinyalini elde etmenin bir başka yolu ise dar bantlı uyarım sinyalinin frekans bandında kaydırılıp tekrar kendisiyle toplanmasıdır. Konuşmanın sinüslerin toplamı biçiminde ifade edilebilmesi aslında uyarım sinyalinin sinüslerin toplamı biçiminde ifade edilebilmesine dayanmaktadır. Burada bu iki temelden yola çıkarak dar bantlı uyarım önce sinüslerin toplamı

$$E_L[n]^k = \sum_{\ell=1}^{L^k} A_{\ell}^k \cos(n\omega_{\ell}^k + \theta_{\ell}^k) \quad (5.4)$$

biçiminde modellenmiş ve tüm sinüs parametreleri bulunmuştur (McAulay and Quatieri, 1986), (Bulu, 2001). Daha sonra bu sinüsler yüksek frekans bandına kaydırılarak toplanmış ve böylece uyarım sinyalinin yüksek frekanslı bölgeleri

$$E_H[n]^k = \sum_{\ell=1}^{L^k} A_{\ell}^k \cos(n(\omega_{\ell}^k + \omega_s) + \theta_{\ell}^k) \quad (5.5)$$

kullanılarak elde edilmiştir. Burada  $\omega_s$ , frekans eksenindeki kaymayı gösterir ve genelde dar bantlı konuşmanın bant genişliği kadar seçilir. Örneğin 8 kHz ile örneklenmiş konuşma için  $\omega_s$ , 3.5 kHz civarında seçilebilir. Bu çalışmada frekansların devamlılığı sağlamak için,  $\omega_s$  dar bantlı konuşmanın kestirilen en yüksek frekansı



Şekil 5.6. Dar bantlı uyarım sinyalinin sinüslerle genişletilmesi

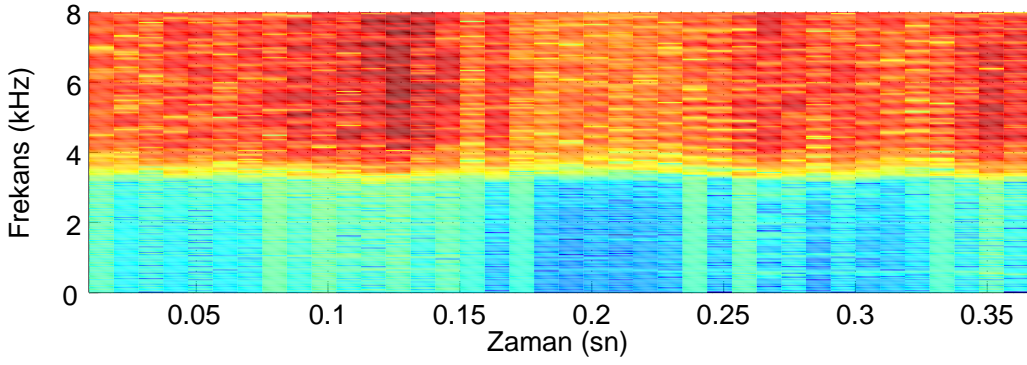
olarak seçilmiştir. Şekil 5.7'te orijinal ve sinüslerin toplanarak elde edilen yüksek frekans bileşenlerine sahip uyarım sinyalleri görülebilir. Ardından bu yüksek frekanslı sinyale dar bantlı uyarım sinyali eklenerek,  $k$  çerçevesi için geniş bantlı uyarım sinyali

$$E_{EX}^k[n] = E_L[n]^k + E_H[n]^k \quad (5.6)$$

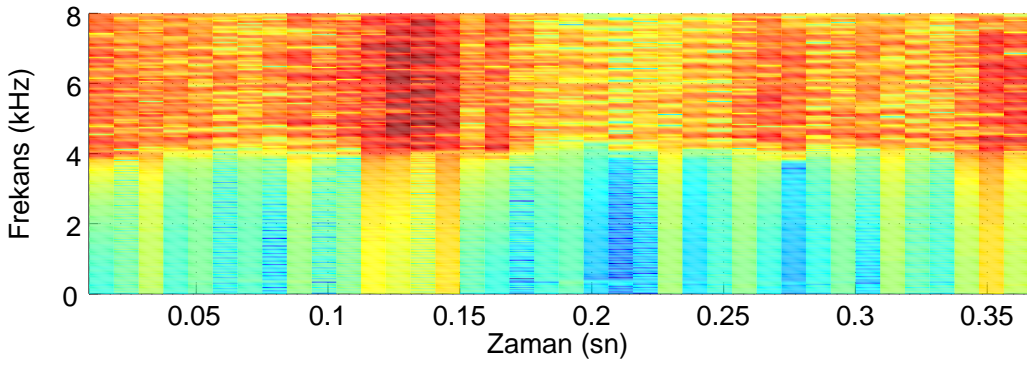
biçiminde bulunur. Elde edilen genişletilmiş uyarım sinyali Şekil 5.8'te görülebilir. Uyarım sinyalinin sinüslerin kaydırılmasıyla genişletildiği bu yöntemin genel işleyiş biçimi Şekil 5.6'de görülebilir. Sessiz konuşma bölgeleri için elde edilen bu genişletilmiş uyarım sinyaline (uyarım sinyalinin gücüyle orantılı olacak biçimde) gürültü de eklenebilir. Eğer konuşma 3.4 kHz'ten 7 kHz'e genişletilecekse eklenecek gürültü [5.5 – 7] kHz bandında olmalıdır.

#### 5.4 Geniş Bantlı Konuşma İçin Kazancın Bulunması

Konuşma Eş. 2.24'teki gibi TK modellendiğinde  $g$ , kazanç parametresi olarak tanımlanır ve bu kazanç aslında konuşmanın TK modellemeye uygunluğunun da bir göstergesidir. Konuşmanın sessiz olduğu bölgede kazanç yüksek, sesli olduğu bölgelerde ise kazanç düşük değerler alır. Geniş bantlı konuşmayı elde etmek için izgel zarf ve uyarım sinyaline ek olarak işte bu kazanç parametresinin de geniş bantlı konuşma için bulunması gerekmektedir.



a. Orijinal uyarım sinyalinin yüksek frekanslı bölgesinin izge grafiği



b. Genişletilen yüksek frekanslı bölgenin izge grafiği

Şekil 5.7. 16 kHz ile örneklenmiş konuşma için orijinal ve genişletilen uyarım sinyalinin yüksek frekanslı bölgeleri

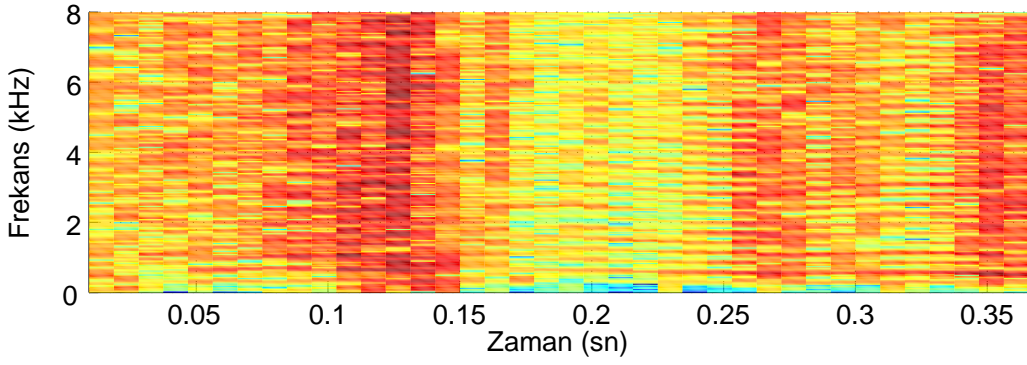
Burada öncelikle  $m$ 'inci çerçeve için dar bantlı konuşmaya ait ortalama izgel güç yoğunluğu

$$\bar{\Phi}_n(\omega; m) \approx \frac{g_n^2}{|\bar{A}_n(\omega; m)|^2} \quad (5.7)$$

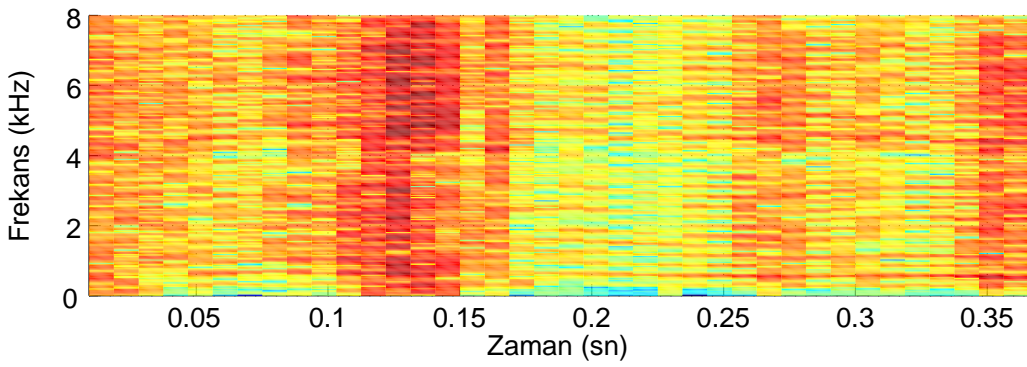
biçiminde ifade edilebilir. Ardından da yine aynı çerçeve için geniş bantlı (genişletilmiş) konuşmaya ait ortalama izgel güç yoğunluğu

$$\bar{\Phi}_e(\omega; m) \approx \frac{g_e^2}{|\bar{A}_e(\omega; m)|^2} \quad (5.8)$$

şeklinde ifade edilir. Daha sonra orijinal dar bantlı konuşmaya ait ortalama izgel güç yoğunluğu ile genişletilmiş konuşmaya ait izgel güç yoğunlunun düşük bantlı kısmı



a. Orijinal uyarım sinyalinin izgeografiği



b. Genişletilen uyarım sinyalinin izgeografiği

Şekil 5.8. 16 kHz ile örneklenmiş konuşma için orijinal ve genişletilmiş uyarım sinyallerinin izgegrafikleri

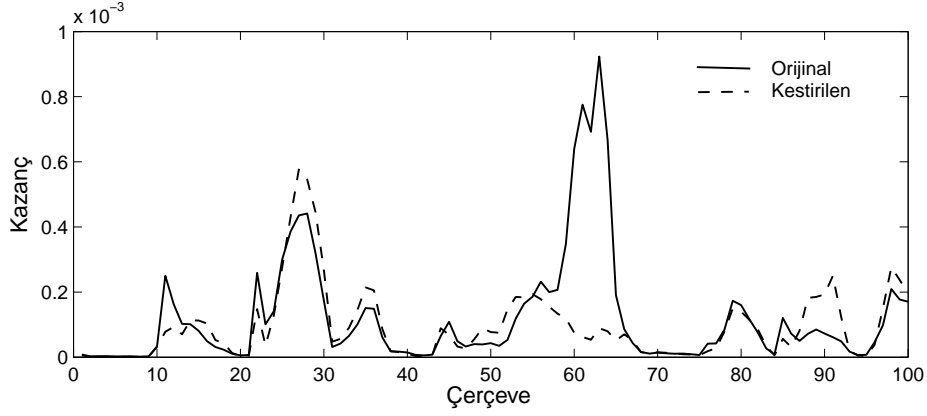
birbirine eşitlenip  $g_e$  parametresi

$$g_e = \sqrt{\frac{|\bar{A}_{en}(\omega; m)|^2}{|\bar{A}_n(\omega; m)|^2}} g_n = \sqrt{\frac{|\bar{\Phi}_{nnorm}(\omega; m)|^2}{|\bar{\Phi}_{enorm}(\omega; m)|^2}} g_n \quad (5.9)$$

biçiminde bulunabilir. Burada  $\bar{\Phi}_{nnorm}(\omega; m)$  ve  $\bar{\Phi}_{enorm}(\omega; m)$  darbantlı ve genişletilmiş zarfın dar bant kısmını gösteren düzleştirilmiş ortalama izgel güç yoğunluklarıdır.

## 5.5 Özel Testler

Bundan önceki bölümlerde önce konuşmanın izgel zarfı genişletilmiş daha sonra ise bu zarf kullanılarak konuşma kalitesi artırılmaya çalışılmıştır. Ancak kalitesi artırılan konuşmaların başarımının ölçülmesi için nesnel (objective) bir başarıım ölçütü bulunmamaktadır. Nesnel başarıım ölçütü olarak sadece izgel zarfların birbirine ya-



Şekil 5.9. 16 kHz ile örneklenmiş konuşma için orijinal ve kestirilen kazançlar

kınlıklarını belirten izgel bozulum ölçütü kullanılmaktadır. Bu ölçütünde konuşma kalitesiyle ilişkisi olmakla beraber aralarında tam olarak bire bir bir ilişki yoktur. Bunun sebebi ise konuşma kalitesinin kullanılan sentez yöntemiyle ve dinleyiciden dinleyiciye değişmesidir. Günümüzde konuşmanın kalitesini ölçmek için Çizelge 5.1’de de görülen Ortalama Kanı Puanı (Mean Opinion Score, MOS) testi kullanılmaktadır (Jayant and Noll, 1984). Bu çalışma sırasında elde edilen sonuçları değerlendirmek için kullanılacak testlerden bir tanesi OKP’dir (Ortalama Kanı Puanı). Sonuçları değerlendirmek için kullanılan bir diğer testte ise dinleyicilere, dinledikleri konuşma parçaları arasındaki tercihleri sorulmuştur.

Çizelge 5.1. Ortalama kanı puanlama cetveli

Puan	Konuşma Kalitesi	Bozulum Seviyesi
5	Mükemmel	Hissedilmez
4	İyi	Çok az hissedilir ve rahatsız etmez
3	Orta	Hissedilir ve çok az rahatsız eder
2	Yetersiz	Rahatsız eder fakat katlanılabilir
1	Kötü	Rahatsız eder ve katlanılamaz

### 5.5.1 Ortalama Kanı Puan Testleri

Bu test sırasında dinleyicilere ikisi bayan ikisi bay olmak üzere toplam dört farklı konuşma parçasının herbiri için yedi tane olmak üzere toplam 28 tane test cümlesi dinletilmiştir. Bu konuşma parçalarından üçü televizyondan biri ise elde bulunan ses kayıt sistemiyle elde edilmiştir. Televizyondan kayıt edilen seslerde bir miktar gürültü



mevcuttur diğer konuşma ise çok kalitelidir. Her bir konuşma parçası için dinletilen sesler Çizelge 5.2'de görülmektedir. Bu tabloda görülen 4 kHz ve 8 kHz ile örneklenen işlenmemiş konuşmalar, 16 kHz ile örneklenmiş geniş bantlı konuşmadan örtüşme önler süzgeçle süzülüp örnekleme frekansları düşürülerek elde edilmişlerdir. Diğer test konuşmaları ise yarı örnekleme hızındaki konuşmalardan son sütünde yazan yöntemler kullanılarak kalitesi artırılan konuşmalardır. Testlerde kullanılan tüm izgel zarflar sinir ağları yöntemi kullanılarak genişletilmişlerdir.

Çizelge 5.2. Ortalama kanı puanlaması için kullanılan konuşma parçalarının elde edilmiş yöntemleri ve örnekleme frekansları

Konuşma	Örnekleme Frekansı (kHz)	Kullanılan Yöntem
$S_4$	4	Yok
$S_8$	8	Yok
$S_{8-ks}$	8	Kaynak-Süzgeç
$S_{16}$	16	Yok
$S_{16-ks}$	16	Kaynak-Süzgeç
$S_{16-sin}$	16	Sinüs
$S_{16-g}$	16	Gürültü

Dinleyicilere, gürültülü olmayan ofis ortamında sessiz çalışan bir dizüstü bilgisayardan Philips SBC HP200 kulaklığı kullanılarak konuşma parçaları dinletilmiş ve Çizelge 5.1'e göre konuşmalara 0.25 hassasiyette puanlar vermeleri istenmiştir. Bir kişi için OKP testinin yapılması ortalama 20 dakika sürmüştür. 18 dinleyiciyle yapılan OKP testinin sonucu Çizelge 5.3'da görülmektedir.

Çizelge 5.3'ün son sütunu tüm konuşma parçaları için ortalama puanı göstermektedir. Bu sütun dikkatle incelendiğinde 1.8 kHz'e bant sınırlı 4 kHz ile örneklenmiş konuşmanın izgel zarfı genişletilip kalitesi artırıldığında elde edilen konuşmanın 1.8 kHz'e bant sınırlı konuşmaya göre daha tercih edilir olduğu söylenebilir. Aynı işlemler uygulanarak kalitesi artırılmaya çalışılan 3.4 kHz'e bant sınırlı konuşma için ise aynı şeyi söylemek her zaman mümkün değildir. Çünkü; 3.4 kHz'e bant sınırlı işlenmemiş konuşmanın OKP'si zarfı genişletilerek kalitesi artırılmaya çalışılmış konuşmalarinkinden daha yüksektir. Ancak gürültüsüz ortamda kaydedilen konuşma için, kalitesi kaynak-süzgeç modeli ile artırılan konuşmanın kalitesi, 3.4 kHz'e bant sınırlı konuşmanın kalitesinden daha yüksek bulunmuştur.

Çizelge 5.3. Test konuşmaları için OKP sonuçları

Konuşma	1. Cümle	2. Cümle	3. Cümle	4. Cümle	Ortalama
S <sub>4</sub>	2.29	2.43	2.24	2.53	2.37
S <sub>8</sub>	3,93	3,71	3,69	4,03	3,84
S <sub>8-ks</sub>	2,56	2,17	2,49	3,06	2,57
S <sub>16</sub>	3,89	3,92	3,88	4,57	4,06
S <sub>16-ks</sub>	3,40	3,38	2,96	4,13	3,47
S <sub>16-sin</sub>	3,69	3,44	3,13	3,86	3,53
S <sub>16-g</sub>	3,31	3,63	2,96	3,38	3,32

18 dinleyici ile yapılan test sonuçlarından sadece en kötü üç tanesinin çıkarıldıktan sonra elde edilen OKP sonuçları Çizelge 5.4'te görülmektedir. Bu tablo dikkatle incelendiğinde sonuçların Çizelge 5.4'e göre iyileştiği görülebilir. Bu durumda kaynak-süzgeç modeli ile kalitesi artırılan konuşmaların, orijinal 3.4 kHz'e bant sınırlı (dar bantlı) konuşmalara göre dinleyiciler tarafından daha tercih edilir oldukları sonucuna varılır.

Çizelge 5.4. 15 dinleyici için için OKP sonuçları

Konuşma	1. Cümle	2. Cümle	3. Cümle	4. Cümle	Ortalama
S <sub>4</sub>	2,38	2,47	2,25	2,53	2,41
S <sub>8</sub>	3,68	3,52	3,65	3,90	3,69
S <sub>8-ks</sub>	2,63	2,25	2,50	3,12	2,63
S <sub>16</sub>	4,01	3,95	3,92	4,60	4,12
S <sub>16-ks</sub>	3,65	3,58	3,40	4,32	3,74
S <sub>16-sin</sub>	3,68	3,48	3,10	3,97	3,56
S <sub>16-g</sub>	3,40	3,70	2,93	3,38	3,35

Dinleyicilerle testler bittikten sonra yapılan değerlendirmelerde, televizyondan kaydedilen konuşmalardaki gürültünün kendilerini rahatsız ettiğine ve ayrıca bu gürültünün, kalitesi artırılan konuşmalara olumsuz etki ettiği sonucuna varılmıştır. Ancak tüm bunlara rağmen elde edilen OKP sonucu (Gustafsson *et al.*, 2006)'daki sonuçlarla paralel çıkmıştır.

### 5.5.2 Karşılaştırmalı Testler

Karşılaştırmalı testler, OKP testlerinde de kullanılan 4 test konuşma parçası kullanılarak gerçekleştirilmiştir. Karşılaştırma testleri, 1.8 *kHz* ile bant sınırlı konuşmanın izgel zarfı genişletilerek 3.4 *kHz*'e çıkarılması ve 3.4 *kHz*'e bant sınırlı işlenmemiş konuşmanın bant genişliğinin 7 *kHz*'e çıkarılması için ayrı ayrı yapılmıştır. Her iki testte de dört test cümlesi için dar bantlı konuşma, genişletilmiş konuşma ve işlenmemiş geniş bantlı konuşma, 16 dinleyiciye dinletilmiştir. 1.8 *kHz*'e bant sınırlı ve bu konuşmadan izgel zarf genişletilerek kalitesi artırılan konuşma için dinletilen toplam 64 test cümlesinden 50 (%78) tanesinde kalitesi artırılan konuşma, dar bantlı konuşmaya göre dinleyiciler tarafından tercih edilmiştir. Aynı test 3.4 *kHz*'e bant sınırlı konuşma için yapıldığında ise kalitesi artırılan konuşmanın tercih edilme sayısı ise 21'de (%32) kalmıştır. Ancak tercih edilen bu 21 cümlenin 10 tanesi sadece gürültüsüz konuşma için gerçekleşmiştir. Bu durumda gürültüsüz konuşma için frekans bandı genişletilen konuşmanın dar bantlı konuşmaya göre tercih edilme oranı %62 olmuştur.

## 6. SONUÇ

Bu çalışmada, telefon ve daha düşük bant genişliğine sahip konuşmaya ait izgel zarf genişletilerek konuşma kalitesi artırılmaya çalışılmıştır. İzgel zarf genişletme işlemi için başvuru çizelgesi ve sinir ağları yöntemi kullanılmıştır. Elde edilen sonuçlar literatürde dar bantlı konuşmanın izgel zarfını genişletme üzerine yapılan az sayıdaki diğer çalışmayla (Epps and Holmes, 1999), (Enbom and Kleijn, 1999), (Jax, 2004) benzer sonuçlar vermiştir. Bu çalışmanın esas olarak katkısı daha önce sinir ağlarıyla benzer bir çalışmanın yapılmamış olmasıdır. Çalışmada sinir ağlarıyla telefon kalitesindeki konuşma genişletilmekle birlikte bu yöntemin daha dar bantlı konuşmalarda da (örneğin  $1.8 - 2 \text{ kHz}$ ) izgel zarfın genişletilmesinde kullanılabileceği gösterilmiştir.

İzgel zarf genişletilirken performans ölçütü olarak ortalama karesel izgel bozulum kullanılmıştır ve telefon kalitesindeki konuşma için ortalama bozulum  $3.8 - 3.9 \text{ dB}$  aralığında bulunmuştur. Bulunan bu bozulum  $3 \text{ dB}$  olan teorik sınırdan büyüktür ancak kabul edilebilir sınırlar içerisinde ve benzer çalışmalardakilerle paralellik içindedir. Ayrıca, konuşmanın kalitesi elde edilen izgel zarfın orijinal zarfa yakınlığıyla orantılı olmakla birlikte kalite artırmada kullanılan sentezleme yöntemine de bağlıdır. Konuşma kodlama ve kalite artırma için yapılan çalışmalarda konuşmanın dar bantlı kısımları için izgel bozulumun  $1 \text{ dB}$  ve yüksek bantlı kısımları için  $2.5 - 3 \text{ dB}$  olmasının konuşma kalitesinde izgel zarftan kaynaklanan büyük bir kalite kaybına yol açmayacağı belirtilmiştir (ETSI EG 201 377-1).

İzgel zarf genişletmedeki amaç dar bantlı konuşmanın kalitesinin iyileştirilmesidir. Elde edilen bu yüksek frekans bileşenlerine sahip izgel zarflar üç ana yöntemle konuşma kalitesinin iyileştirilmesinde kullanılmıştır. Bu yöntemler basitten karmaşığa doğru izgel zarfın gürültüyle doldurulması, izgel zarfın sinüslerle doldurulması ve genişletilen izgel zarfın kaynak-süzgeç modelinde kullanılmasıdır. Kullanılan bu yöntemlerden göreceli olarak en başarılı olanı kaynak-süzgeç modeliyle konuşma kalitesinin artırılmasıdır. Yapılan öznel testlerde elde edilen yüksek frekans bileşenlerine sahip konuşma parçaları için konuşma kalitesi özellikle  $1.8 \text{ kHz}$ 'ten bant genişliği genişletilen konuşmalar için hissedilir derecede artmıştır. Telefon kalitesindeki konuşmalar için ise aynı şeyi söylemek her zaman mümkün değildir. Gürültülü konuşma bölgelerinde konuşma kalitesi bir miktar düşmüş gürültüsüz bölgelerde ise

kalitede bir miktar artma görülmüştür. Ayrıca dinleme testlerinde kulaklık kullanılmıştır. Kulaklıkla dinletilen konuşma parçalarındaki çok az gürültü bile dinleyicileri rahatsız etmiş ve test cümlelerine verdikleri puanlar düşmüştür. Aynı konuşma parçaları kaliteli bir hoparlör kullanılarak dinletildiğinde dinleyicilerin gürültüden daha az rahatsız oldukları görülmüş ve kalitesi artırılan konuşmaların dinleyicilerden aldıkları puanlar artmıştır.

Bu çalışmada önerilen konuşma kalitesi artırma yöntemlerinde herhangi bir ek bilgi iletimi (side information) gerekmediğinden iletim bant genişliğini artırıcı etkileri yoktur. Konuşmanın kodlanmış olup olmadığı, kodlandıysa da hangi yöntemle kodlandığı önem taşımamaktadır. Günümüzde telefon kalitesindeki konuşmaları kodlamak için geliştirilen kodlama tekniklerinde genelde 10. dereceden DÖK parametreleri kodlanarak alıcı tarafa iletilmektedir. 1.8 kHz'e bant sınırlı konuşmaların bant genişliğinin artırılması sırasında kullanılan izgel zarf genişletme yönteminin bu kodlama teknikleriyle beraber kullanılarak 10 yerine 6 tane DÖK parametresinin karşı tarafa gönderilebileceği ve böylece daha düşük bit hızlarında konuşma kodlamanın mümkün olabileceği düşünülmektedir. Yöntemin, tüm bu avantajları sayesinde dar bantlı konuşma iletiminde yaygın kullanım alanları bulabileceği değerlendirilmektedir.

## KAYNAKLAR DİZİNİ

- Bulu, G., 2001, Sinüs modeli ile düşük bit hızlarında konuşma kodlama, Yüksek lisans tezi, Hacettepe Üniversitesi.
- Bulu, G., ve Bingöl, A. S., 2007, Dar bantlı konuşmanın İzgel zarfının dışdeğerlenmesi, *SİU 2007 Sinyal İşleme ve Uygulamaları Kurultayı*.
- Carl, H., and Heute, U., 1994, Bandwidth enhancement of narrowband speech signals, In: *Proceedings of EUSIPCO, VII European Signal Processing Conference, Edinburg, Scotland*, pp. 1178–1181.
- Chan, C-F., and Hui, W-K., 1996, Wideband enhancement of narrowband coded speech using mbe re-synthesis, In: *Proceedings of ICSP, 3rd International Conference on Signal Processing*, pp. 667–670.
- Chennoukh, S., Gerrits, A., Miet, G., and Sluijter, R., 2001, Speech enhancement via frequency bandwidth extension using line spectral frequencies, In: *Proceedings of ICASSP, Salt Lake City*, pp. 665–668.
- Enbom, N., and Kleijn, W. B., 1999, Bandwidth expansion of speech based on vector quantization of the mel frequency cepstral coefficients, In: *Proceedings of IEEE Workshop on Speech Coding, Porvoo, Finland*, pp. 171–173.
- Epps, J., and Holmes, W. H., 1998, Speech enhancement using stc-based bandwidth extension, In: *Proceedings ICSLP, Sidney, Australia*, pp. 519–522.
- Epps, J., and Holmes, W. H., 1999, A new technique for wideband enhancement of coded narrowband speech, In: *Proceedings of IEEE Workshop on Speech Coding, Porvoo, Finland*, pp. 174–176.
- Epps, J., and Holmes, W. H., 2001, A new very low bit rate wideband speech coder with a sinusoidal highband model, In: *Proceedings of IEEE International Symposium on Circuits and Systems ISCAS 2001, Vol. 2*, pp. 349–352.
- Geiser, B., and Vary, P., 2007, Backwards compatible wideband telephony in mobile networks: Celp watermarking and bandwidth extension, In: *Proceedings of ICASSP, Honolulu*, pp. IV–533–IV–536.
- Gray, Jr. A. H., and Markel, J. D., 1976, Quantization and bit allocation in speech processing, *IEEE Transactions on Acoustics, Speech and Signal Processing* 24(6), 459–473.

- Gustafsson, H., Lindgren, U.A., and Claesson, I., 2001, Speech bandwidth extension, In: *IEEE International Conference on Multimedia and Expo, ICME*, pp. 809–812.
- Gustafsson, H., Lindgren, U.A., and Claesson, I., 2006, Low-complexity feature-mapped speech bandwidth extension, *IEEE Transactions on Audio, Speech and Language Processing* 14(2), 577–588.
- Hagen, R., 1994, Spectral quantization of cepstral coefficients, In: *Proceedings of ICASSP, Adelaide*, pp. 509–512.
- Itakura, F., 1975, Line spectrum representation of linear predictive coefficients of speech signals, *Journal of the Acoustical Society of America* 57, 35.
- Jax, P., 2004, Bandwidth extension for speech, In: *Audio Bandwidth Extension*, E. Larsen and R. M. Aarts, Eds.). Oxford University Press.
- Jax, P., and Vary, P., 2000, Wideband extension of telephone speech using a hidden markov model, In: *Proceedings of IEEE Workshop on Speech Coding, Delavan, WI*, pp. 133–135.
- Jax, P., and Vary, P., 2002, An upper bound on the quality of artificial bandwidth extension of narrowband speech signals, In: *Proceedings of ICASSP, Orlando*, pp. 237–240.
- Jayant, N. S., and Noll, P., 1984, *Digital Coding of Waveforms*, Prentice-Hall, New Jersey.
- Kay, S. M., 1988, *Modern Spectral Estimation: Theory and Application*, Prentice-Hall, New Jersey.
- Lahouti, F., Fazel, A. R., Safavi-Naeini, A. H., and Khandani, A. K., 2006, Single and double frame coding of speech lpc parameters using a lattice-based quantization scheme, *IEEE Transactions on Audio, Speech and Language Processing* 14(5), 1624–1701.
- Linde, Y., Buzo, A., and Gray, R. M., 1980, An algorithm for vector quantizer, *IEEE Transactions on Communications* 28(1), 84–95.
- Makhoul, J., 1975, Linear prediction: A tutorial review, *Proceedings of the IEEE* 63(4), 561–580.
- Markel, J. D., and Gray, A. H., 1976, *Linear Prediction of Speech*, Springer-Verlag, Berlin, Heidelberg, New York.

- McAulay, R. J., and Quatieri, T. F., 1986, Speechanalysis-synthesis based on a sinusoidal representation, *IEEE Transactions on Acoustics, Speech, and Signal Processing* 34(4), 744–754.
- Moon, T. K., and Stirling, W. C., 2000, *Mathematical Methods and Algorithms for Signal Processing*, Prentice-Hall, New Jersey, USA.
- Noll, A. M., 1967, Cepstrum pitch determination, *Journal of the Acoustical Society of America* 41(2), 293–309.
- Oppenheim, A. V., and Schaffer, R. W., 1989, *Discrete-Time Signal Processing*, Prentice-Hall, New Jersey.
- O’Shaughnessy, D., 1987, *Speech Communications - Human and Machine*, Addison-Wesley, Reading, Massachusetts.
- Paliwal, K.K., and Atal, B.S., 1993, Efficient vector quantization of LPC parameters at 24 bits/frame, *IEEE Transactions on Speech and Audio Processing* 1(1), 3–14.
- Patrick, P. J., Steele, R., and Xydeas, C. S., 1983, Frequency compression of 7.6 khz speech into 3.3 khz bandwidth, *IEEE Transactions on Communications* 31(5), 692–701.
- Quatieri, T. F., 2002, *Discrete-Time Speech Signal Processing*, Prentice Hall, Upper Saddle River, New Jersey.
- Rabiner, L. R., and Schaffer, R. W., 1978, *Digital Processing of Speech Signals*, Prentice-Hall.
- Rohde, N., and Vedstesen, S. A., 2007, Artificial bandwidth extension of narrowband speech, Master’s thesis, Aalborg University.
- Vaseghi, S., Zavarehei, E., and Qin, Y., 2006, Speech bandwidth extension: Extrapolations of spectral envelop and harmonicity quality of excitation, In: *Proceedings of ICASSP, Toulouse*, pp. III–844–III–847.
- Viswanathan, R., and Makhoul, J., 1975, Quantization properties of transmission parameters in linear predictive systems, *IEEE Transactions on Acoustics, Speech and Signal Processing* 23(3), 309–321.
- Yao, S., and Chan, C-F., 2005, Block-based bandwidth extension of narrowband speech signal by using cdhmm, In: *Proceedings of ICASSP, Philadelphia, PA*, pp. 793–796.



Yoshida, Y., and Abe, M., 1994, An algorithm to reconstruct wideband speech from narrowband speech based on codebook mapping, In: *Proceedings of ICSLP, Yokohama, Japan*, pp. 1591–1594.

Özen, Ş. S., 2002, Türkçe metinden konuşma sentezleme, Yüksek lisans tezi, Hacettepe Üniversitesi.

## **EKLER DİZİNİ**

EK 1. İZGEL BOZULUM

Ek 2. DAR BANTLI KONUŞMANIN İZGEL ZARFINI GENİŞLETMEDE TEORİK SİNİR

EK 3. İNGİLİZCE-TÜRKÇE TERİMLER SÖZLÜĞÜ

## EK 1. İZGEL BOZULUM

Dar bantlı konuşmadan, izgel zarfı genişletilerek elde edilen yüksek bant genişliğine sahip konuşmanın, orijinal yüksek bantlı konuşmaya yakınlığı yada bu konuşmanın düşük bant genişliğine sahip konuşmaya oranla kalitesinin ne ölçüde arttığı aslında dinleme testleri yapılarak ölçülmelidir. Ancak bu testleri yapmak çok uzun ve zahmetlidir. Ayrıca elde edilen yüksek bantlı konuşmanın kalitesi aynı zamanda kullanılan sentezleme yöntemlerine de bağlıdır.

Günümüzde yapılan çalışmalarda genişletilen izgel zarfların birbirine yakınlıklarını ölçmek için genelde izgel bozulum ölçütü (İBÖ, spectral distortion measure) kullanılmaktadır. Bu ölçüt izgel zarflar arasındaki bozulumu nesnel (objective) olarak ifade etmekle birlikte aynı zamanda sentezlenen konuşmanın kalitesiyle de ilişkilidir. İB k'ncı çerçevedeki konuşma parçası için dışdeğerlenen yüksek frekanslı izgel zarf bölgelerinde

$$D^k = \frac{1}{F_s/4} \int_{F_s/4}^{F_s/2} \left[ 10 \log_{10}(P_n^k(f)) - 10 \log_{10}(\hat{P}_{ex}^k(f)) \right]^2 df \quad (\text{Ek 1.1})$$

kullanılarak hesaplanabilir. Burada  $F_s$  örnekleme frekansı,  $P_n^k(f)$  ve  $\hat{P}_{ex}^k(f)$  ise sırasıyla k'nci çerçeve için orijinal ve dışdeğerlenmiş güç izgeleridir. Ancak bu ölçütte kazanç ve izgel parametrelerden kaynaklanan izgel zarfların birbirlerine göre güç kaymaları hesaba katılmamıştır. Bu çalışmada izgel zarflar arasındaki fark

$$D_M = \frac{1}{F_s/4} \int_{F_s/4}^{F_s/2} 20 \log_{10} \left[ G_c \frac{A_{ex}(f)}{\hat{A}_o(f)} \right] df \quad (\text{Ek 1.2})$$

kullanılarak hesaplanacaktır. Burada  $G_c$  güç kaymasını engellemek için kullanılan parametredir ve

$$G_c = \frac{1}{F_s/4} \int_{F_s/4}^{F_s/2} G_c \frac{A_o(f)}{\hat{A}_{ex}(f)} df \quad (\text{Ek 1.3})$$

biçiminde tanımlanır. Ayrıca yukarıdaki eşitliklerde çerçeve indisi,  $k$ , karmaşıklığı önlemek için düşürülmüştür. Daha sonra ortalama izgel bozulum

$$D_{avg} = \frac{1}{M} \sum_{k=1}^K D_M(k) \quad (\text{Ek 1.4})$$

biçiminde hesaplanabilir. Ancak bazı kaynaklarda izgel bozulum hesaplanırken ortalama karesel hata yöntemi de kullanılmıştır (Jax, 2004), (Rohde and Vedstesen, 2007).

$$D_{avg} = \sqrt{\frac{1}{M} \sum_{k=1}^K D_M^2(k)} \quad (\text{Ek 1.5})$$

## EK 2. DAR BANTLI KONUŞMANIN İZGEL ZARFINI GENİŞLETMEDE TEORİK SINIR

Bu bölümde izgel zarf genişletmenin teorik sınırı üzerinde durulacaktır. Ek-1'de izgel zarf genişletmede performans ölçütü olarak kullanılan izgel bozulum ölçütü hakkında kısa bilgi verilmişti. Burada bu ölçütün teorik sınırı bilgi teorisi kullanılarak bulunmaya çalışılacaktır.

Öncelikle yüksek kaliteli konuşma bir alçak geçirgen süzgeç bir de yüksek geçirgen süzgeçten geçirilerek darbant  $s_{nb}(k)$  ve yüksek bant (kayıp bant)  $s_{hb}(k)$  bileşenlerine ayrılır. Daha sonra bu iki sinyal 7 kHz örnekleme hızına düşürülür. Ardında bu iki sinyale ait TK parametreleri bulunur. Bu parametrelere göre yüksek banta ait karesel izgel bozulum Eş Ek1.1'dekine benzer şekilde

$$D^2 = \frac{1}{2\pi} \int_{-\pi}^{\pi} \left[ 20 \log_{10} \left( \frac{\rho_{rel}}{A_{hb}(e^{jw})} \right) - 20 \log_{10} \left( \frac{\hat{\rho}_{rel}}{\hat{A}_{hb}(e^{jw})} \right) \right]^2 df \quad (\text{Ek 2.1})$$

biçiminde tanımlanır. Burada  $\rho_{rel} = \rho_{hb}/\rho_{nb}$  biçiminde tanımlanabilir.  $A_{hb}(e^{jw})$  ve  $\hat{A}_{hb}(e^{jw})$  sırasıyla orijinal ve bulunan ters süzgeçlerdir. Daha sonra Parseval teoremi kullanılarak izgel bozulum kepsral katsayılar cinsinden

$$D^2 = \left( \frac{10}{\log_e 10} \right)^2 \left( (c_0 - \hat{c}_0)^2 - 2 \sum_{i=1}^{\infty} (c_i - \hat{c}_i)^2 \right) \quad (\text{Ek 2.2})$$

biçiminde bulunabilir (Hagen, 1994) . Burada  $c_0, c_1, \dots, c_{N_a+1}$  TK parametrelerinden ve kazançtan, bir çerçeve için hesaplanan kepsral katsayılarıdır (Markel and Gray, 1976). Bir konuşma parçası için ortalama karesel hata

$$D^2 = \frac{\sqrt{2}10}{\log_e 10} \sqrt{E \left\{ (c_0 - \hat{c}_0)^2 - 2 \sum_{i=1}^{\infty} (c_i - \hat{c}_i)^2 \right\}} \quad (\text{Ek 2.3})$$

biçiminde ifade edilebilir. Yukarıdaki eşitliği sadeleştirmek için

$$y_i = \begin{cases} \frac{1}{\sqrt{2}} c_i, & i = 0 \\ c_i, & 1 \leq i \leq d \end{cases} \quad (\text{Ek 2.4})$$

tanımlanabilir. Burada  $y_i$ 'ler scalardır ve bunlar d boyutlu  $\mathbf{y} = [y_0, y_1, \dots, y_{d-1}]^T$  vektörünü oluştururlar. Ardından yukarıdaki iki eşitlik kullanılarak izgel bozulum

$$D^2 \approx \frac{\sqrt{210}}{\log_e 10} \sum_{i=0}^{d-1} (y_i - \hat{y}_i)^2 \quad (\text{Ek 2.5})$$

biçiminde sadeleştirilebilir (Jax, 2004).

Sinyal işlemede iki sinyal arasındaki ilişki ilinti (correlation) katsayıları cinsinden ifade edilir. Fakat bu ilişkiye bilgi teorisi (information theory) perspektifiyle bakıldığında iki sinyal arasındaki ilişki karşılıklı bilgi miktarı (mutual information) cinsinden ifade edilebilir. Buradaki amacımız bir taraftaki ortalama karesel hata ile diğer taraftaki dar bantlı konuşma parametreleri  $\mathbf{x}$  ile geniş bant (kayıp) parametreleri  $\mathbf{y}$  arasındaki karşılıklı bilgi miktarı arasında bir üst sınır bulmaktır.

Eş. Ek 2.4'te tanımlanan  $\hat{\mathbf{y}}$  herhangi bir bant genişletme algoritmasıyla genişletilen konuşmadan elde edilen TK katsayılarından hesaplanabilir. Dolayısıyla burada hesaplanacak performans sınırı, izgel zarfın herhangi başka bir gösterim yöntemiyle hesaplandığı bant genişletme algoritmaları için de kullanılabilir. Herhangi bir şekilde genişletilen zarftan kestirilen ( $\hat{\mathbf{y}} = f(\mathbf{x})$ ) parametreler ile orijinal parametreler  $\mathbf{y}$  arasındaki karesel hata

$$E \{ \|\mathbf{y} - \hat{\mathbf{y}}\|^2 \} \geq \frac{d}{2\pi e} \exp \left( \frac{2}{d} (h(\mathbf{y}) - I(\mathbf{x}; \mathbf{y})) \right) \quad (\text{Ek 2.6})$$

biçiminde sadeleştirilebilir (Jax and Vary, 2002). Burada  $E \{ \|\mathbf{y} - \hat{\mathbf{y}}\|^2 \}$  Eş. Ek 2.5'teki kepsral uzaklıkta yerine konulursa

$$D \geq \frac{\sqrt{210}}{\log_e 10} \sqrt{E \{ \|\mathbf{y} - \hat{\mathbf{y}}\|^2 \}}$$

$$D \geq \frac{\sqrt{210}}{\log_e 10} \sqrt{\frac{d}{2\pi e} \exp \left( \frac{1}{d} (h(\mathbf{y}) - I(\mathbf{x}; \mathbf{y})) \right)} \quad (\text{Ek 2.7})$$

ortalama karesel izgel bozulum için karşılıklı bilgi miktarı  $I(\mathbf{x}; \mathbf{y})$  ve farksal entropi  $h(\mathbf{y})$  cinsinden bir alt sınır elde edilmiş olur. Burada  $h(\mathbf{y})$  sadece yüksek frekans bölgelerine ait kepsral katsayılar  $I(\mathbf{x}; \mathbf{y})$  ise düşük ve yüksek kepsral katsayıların ikisine birden bağlıdır. Entropi tanımından yola çıkarak

$$h(\mathbf{y}) = -E \{ \log p(\mathbf{y}) \} \quad (\text{Ek 2.8})$$

biçiminde tanımlanır. Bu eşitlik ancak olasılık yoğunluk fonksiyonu,  $p(\mathbf{y})$  tanımlı olduğu zaman hesaplanabilir. Bizim durumumuzda bu olasılık dağılım fonksiyonu,  $p(\mathbf{y})$  Gauss karışım modeli (GMM, Gaussian Mixture Model) ile modellenenabilir.

$$p_m(\mathbf{y}) = \sum_{i=0}^{L-1} \rho_i \mathcal{N}(\mathbf{y}; \mu_i; \Sigma_i) \approx p(\mathbf{y}) \quad (\text{Ek 2.9})$$

Burada  $\mu_i$  ortalama deęer vektörünü,  $\sigma_i$  ise kovaryans matrisini,  $\rho_i$  ise skalar aęırlığı belirtir. Bu olasılık daęılımına, Gauss karışım parametrelerinin 500000 çerçevelik konuşmaya ait parametrelerle eęitilerek yaklaşılmıştır (Jax and Vary, 2002). Daha sonra bulunan bu olasılık daęılım fonksiyonu yardımıyla Eş Ek 2.8'deki beklenti bir ortalama

$$h(\mathbf{y}) = -\frac{1}{N_v} \sum_{v=0}^{N_v-1} \log p_m(\mathbf{y}_m(v)) \quad (\text{Ek 2.10})$$

almaya dönüşebilir. Buradaki data vektörleri  $\mathbf{y}_m(v)$ ,  $v = 0, 1, \dots, N_v - 1$  olasılık daęılımlarına göre rastgele üretilmişlerdir ve  $N_v = 10^6$  seçilmiştir (Jax and Vary, 2002).

Daha sonra bu olasılık daęılım fonksiyonu yardımıyla hesaplanan  $h(\mathbf{y})$ , Eş. Ek 2.7'de yerine konulup izgel bozulum ile karşılıklı bilgi miktarı arasında bir ilişki bulunur ve bu ilişki Şekil 4.7'de gösterilmiştir.

### EK 3. İNGİLİZCE-TÜRKÇE TERİMLER SÖZLÜĞÜ

all pole	: tüm kutup
amplitude	: genlik
arcsine reflection coefficient	: arksinüs yansıma katsayıları
bandpass filter	: bant-geçirgen süzgeç
binary	: ikili
buzzy	: gürültülü
centroid	: ağırlık merkezi
code excited	: kod uyarımlı
coefficient	: katsayı
compress	: bastırmak
continuous-time	: sürekli zaman
converter	: çevirici
data	: veri
digital	: sayısal
discontinuity	: süreksizlik
discrete-time	: ayrık zaman
distortion	: bozulum
distortion measure	: bozulum ölçütü
entropy	: entropi
envelope	: zarf
filter	: süzgeç
filter bank	: süzgeç bankası
frame	: çerçeve
frame rate	: çerçeve hızı
frequency	: frekans, sıklık
function	: fonksiyon, işlev
gain	: kazanç
glottal excitation	: ses yolu uyarımı
input	: giriş
interpolation	: aradeğerleme
line spectral frequency	: çizgi izgel frekans
linear	: doğrusal



log area ratios	: logaritmik alan oranları
lowpass filter	: alçak geçiren süzgeç
minimum phase	: en küçük evreli
mutual information	: karşılıklı bilgi miktarı
narrowband	: darbant
neural network	: sinir ağları
normalize	: düzleştirmek
output	: çıkış
overlap-add	: örtüştür-topla
performance	: başarımlı
phase	: faz, evre
phase offset	: evre kayması
pitch frequency	: perde frekansı
prediction	: öngörüm
polynomial	: çokterimli
pulse	: vuru
quantization	: nicemleme
quantizer	: nicemleyici
recursion	: yineleme
reset	: başlangıçtaki (ilk) konuma getirmek
response	: tepki
sampling	: örnekleme
sampling period	: örnekleme aralığı
sensor	: duyargaç
sequence	: seri, dizi
set	: küme
set-point	: istenilen çıkış
signal	: sinyal
spectral distortion	: izgel bozulum
spectral tilt	: izgel kayma
spectrally flattened	: izgel düzleştirilmiş
spectrum	: izge
speech	: konuşma
stable	: kararlı
stationary	: durağan

steady-state	: durgun durum
steepest descent	: bayır iniři
time-varying	: zamanla deęiřen
time-weighted	: zaman aęırlılandırılmıř
transfer function	: aktarım iřlevi
transform	: dđnüşüm
unvoiced	: sessiz (ses)
unstable	: kararsız
voiced	: sesli (ses)
weight	: aęırlık

## ÖZGEÇMİŞ

Adı Soyadı : GÜRHAN BULU  
Doğum Yeri : Elazığ  
Doğum Yılı : 09.01.1977  
Medeni Hali : Evli

### Eğitim ve Akademik Durumu

Lise 1991-1994 : İstanbul Atatürk Fen Lisesi, İSTANBUL  
Lisans 1994-1998 : Hacettepe Üniversitesi  
Elektrik ve Elektronik Mühendisliği Bölümü,  
ANKARA  
Yüksek Lisans 1998-2001 : Hacettepe Üniversitesi  
Elektrik ve Elektronik Mühendisliği Bölümü,  
ANKARA

Yabancı Dil : İngilizce

### İş Tecrübesi

1998-2007 : Hacettepe Üniversitesi  
Elektrik ve Elektronik Mühendisliği Bölümü,  
ANKARA  
Araştırma Görevlisi