

**ANKARA ÜNİVERSİTESİ
FEN BİLİMLERİ ENSTİTÜSÜ**

DOKTORA TEZİ

**TIKANIKLIK GİZLENMESİ İÇİN
TERSİNİR ZAMAN SKALASI DEĞİŞİMİ**

Saadettin GÜLER

ELEKTRONİK MÜHENDİSLİĞİ ANABİLİM DALI

ANKARA

2011

Her hakkı saklıdır

TEZ ONAYI

Saadettin GÜLER tarafından hazırlanan “**Tıkanıklık Gizlenmesi İçin Tersinir Zaman Skalası Değişimi**” adlı tez çalışması 21/10/2011 tarihinde aşağıdaki jüri tarafından oy birliği ile Ankara Üniversitesi Fen Bilimleri Enstitüsü Elektronik Mühendisliği Anabilim Dalı’nda **DOKTORA TEZİ** olarak kabul edilmiştir.

Danışman : Prof. Dr. Hakkı Gökhan İLK

Jüri Üyeleri :

Başkan : Prof. Dr. Enis ÇETİN
Bilkent Üniversitesi, Mühendislik Fakültesi, Elektrik-Elektronik
Mühendisliği Bölümü

Üye : Doç. Dr. Çağatay CANDAN
Ortadoğu Teknik Üniversitesi, Mühendislik Fakültesi, Elektrik ve Elektronik
Mühendisliği Bölümü

Üye : Yrd. Doç. Dr. Murat Hüsni SAZLI
Ankara Üniversitesi, Elektronik Mühendisliği Anabilim Dalı

Üye : Yrd. Doç. Dr. Hakkı Alparslan ILGIN
Ankara Üniversitesi, Elektronik Mühendisliği Anabilim Dalı

Üye : Prof. Dr. Hakkı Gökhan İLK
Ankara Üniversitesi, Elektronik Mühendisliği Anabilim Dalı

Yukarıdaki sonucu onaylarım

Prof. Dr. Özer KOLSARICI
Enstitü Müdürü

ÖZET

Doktora Tezi

TIKANIKLIK GİZLENMESİ İÇİN TERSİNİR ZAMAN SKALASI DEĞİŞİMİ

Saadettin GÜLER

Ankara Üniversitesi
Fen Bilimleri Enstitüsü
Elektronik Mühendisliği Anabilim Dalı

Danışman: Prof. Dr. H. Gökhan İLK

Bu tez çalışmasında; sesin tıkanık kanallardan veya ağlardan iletimi için zaman skalası değişimi tekniklerinin sentezleme kaliteleri analiz edilmiştir. Sesin iletim öncesi hızlandırılması ses süresini düşürdüğünden daha az band genişliği ile iletilmesi sağlanır. İletilen ses karşı tarafta tekrar yavaşlatılması ile orijinal skalasına geri döndürülür. Paket anahtarlama ağılarında tıkanıklık gecikme ve ses çerçevesi kaybına da neden olabilir. Alıcı tarafına ulaşan seslerin zaman skalasının artırılması ile bu gecikme ve kayıp da gizlenebilir.

Sesin daha az bit oranı ile iletimi için; zaman alanı tekniklerinden OLA, SOLA, EM-TSM, GLS-TSM, WSOLA, PSOLA, AOLA, Suziki- Misaki ve PICOLA; frekans alanı tekniklerinden faz vokoder; sinyal modelleme tekniklerinden de sinüs modelleme ve spektral modelleme sentezi kullanılmıştır. Algoritmalar TIMIT ses veritabanından alınan 8 farklı aksanda 8 ayrı erkek ve 4 bayan tarafından seslendirilen fonotik dengeli cümle ile test edilmiştir. PESQ sonuçlarına göre en iyi sentez kalitesi WSOLA ve gerçek tepe enterpolasyonlu sinüs modelleme ile elde edilmiştir.

Uyarlanabilir çalma zamanlaması ve paket kaybı gizlemesi için SOLA, WSOLA ve tez çalışmasında önerilen yeni bir teknik kullanılmıştır. Bu algoritmalar örnek gecikme ve kayıp simülasyon modeline göre uygulanmıştır. Ses örneği olarak TIMIT ses veritabanından alınan aynı sesler kullanılmıştır. PESQ sonuçlarına göre bayan sesleri için WSOLA, erkek sesleri için ise önerilen yöntem en iyi sentez kalitesini sağlamıştır.

Tez çalışmasında yeni bir frekans alanı tekniği olarak DCT ile zaman skalası değişimi de önerilmiştir. DCT modelleme ile faz vokoderden daha kaliteli TSM sentezleme kalitesi elde edilmiştir. Kaybolan prototiplerin sentezlenmesi de PESQ sonuçlarına göre PSOLA'dan daha iyi gerçekleşmiştir.

Ekim 2011, 136 sayfa

Anahtar Kelimeler: Zaman skalası değişimi, uyarlanabilir çalma zamanlaması, paket kaybı gizlemesi, OLA, faz vokoder, DCT, sinüs modelleme, spektral modelleme sentezi

ABSTRACT

Ph.D. Thesis

REVERSIBLE TIME SCALE MODIFICATION FOR CONGESTION CONCEALMENT

Saadettin GÜLER

Ankara University
Graduate School of Natural and Applied Sciences
Department of Electronics Engineering

Supervisor: Prof. Dr. H. Gökhan İLK

In this thesis work; synthesis quality of time scale modification algorithms for delivery of voice/audio through congested channels or network are examined. Speeding audio/voice up is also shortening the duration of audio/voice and allows saving in transmission rate. The audio/voice is speed down at the receiver side to go back to the original scale. In packet switching network, congestion may increase the network delay or may cause to packet loss. The delay increment or packet loss may be concealed at the receiver side by increasing the duration of the received packets.

OLA, SOLA, EM-TSM, GLS-TSM, WSOLA, PSOLA, AOLA, Suziki- Misaki and PICOLA among time domain algorithms; phase vocoders among frequency domain algorithms and Sinusoidal modelling and spectral modelling synthesis among signal modelling algorithms are used for transmission rate decreament. Voices of a sentence spoken by 8 male and 4 female speakers each belongs to different dialect regions of the US are selected from TIMIT corpus to test the algorithms. WSOLA and Sinusoidal Modelling with quadratic spectral interpolation are the best performer according to the PESQ results.

SOLA, WSOLA and a novel proposed technique is used for adaptive playout and packet loss concealment. These algorithms are applied to selected samples of TIMIT corpus according to a sample playout scheduling and packet loss scheme. According to the PESQ results, the proposed technique is the best performer for male speakers while WSOLA is the best for female speakers.

Modified DCT synthesis is also used as novel frequency domain technique. Accordig to PESQ results, its synthesis quality is better than phase vocoder's for transmission rate decrement and beter than PSOLA for prototype interpolation.

October 2011, 136 pages

Key Words: Time scale modification, adaptive playout, packet loss concealment, OLA, phase vocoder, DCT, sinusodial modelling, spectral modelling synthesis.

TEŞEKKÜR

Danışman hocam Prof. Dr. H. Gökhan İLK'e (Ankara Üniversitesi, Elektronik Mühendisliği Anabilim Dalı) çok teşekkür ederim. Tez konusunun belirlenmesinden sonuçlanmasına kadar çalışmamın her safhasında yakın ilgi ve önerileri ile beni yönlendirmiş ve teşvik etmiştir.

Tez İzleme Komitesi üyeleri Prof. Dr. Enis ÇETİN (Bilkent Üniversitesi, Mühendislik Fakültesi, Elektrik-Elektronik Mühendisliği Bölümü) ve Yrd. Doç. Dr. M. Hüsnü SAZLI'ya (Ankara Üniversitesi, Elektronik Mühendisliği Anabilim Dalı) tez çalışmam boyunca değerli yorum ve önerileriyle yaptıkları katkıdan dolayı teşekkür ederim.

Tez jürisi üyeleri Doç. Dr. Çağatay CANDAN (Ortadoğu Teknik Üniversitesi, Mühendislik Fakültesi, Elektrik ve Elektronik Mühendisliği Bölümü) ve Yrd. Doç. Dr. H. Alparslan ILGIN'a (Ankara Üniversitesi, Elektronik Mühendisliği Anabilim Dalı) tezi okuyup yorumlamalarından dolayı teşekkür ederim.

Son olarak tez çalışmam boyunca bana tahammül eden ailem ve arkadaşlarıma teşekkür ederim.

Saadettin GÜLER

Ankara, Ekim 2011

İÇİNDEKİLER

ÖZET.....	i
ABSTRACT.....	ii
TEŞEKKÜR	iii
KISALTMALAR DİZİNİ	vi
ŞEKİLLER DİZİNİ	viii
ÇİZELGELER DİZİNİ	x
1. GİRİŞ	1
1.1 Amaç ve Kapsam	4
2. KURAMSAL TEMELLER.....	7
2.1 Zaman Skalası Değişimi İçin Zaman Alanı Teknikleri.....	7
2.1.1 OLA	8
2.1.2 SOLA	10
2.1.3 GLS-TSM	12
2.1.4 EM-TSM	14
2.1.5 TD-PSOLA.....	17
2.1.6 WSOLA	19
2.1.7 AOLA	20
2.1.8 Suzuki-Misaki yöntemi	23
2.1.9 Laroch yöntemi.....	25
2.1.10 PICOLA	26
2.2 Zaman Skalası Değişimi İçin Frekans Alanı Teknikleri.....	29
2.2.1 Kısa Zaman Fourier Dönüşümü (STFT)	30
2.2.2 Faz vokoderleri	33
2.2.3 Faz kilitli vokoderler	38
2.3 Zaman Skalası Değişimi İçin Sinyal Modelleme Teknikleri.....	39
2.3.1 Sinüs modelleme	40
2.3.2 Spektral Modelleme Sentezi (SMS)	50
2.3.3 Geçiş Modelleme Sentezi (TMS)	53
2.4 Paket Anahtarlama Ağılarda Tıkanıklık İçin Zaman Skalası Değişimi	55
2.4.1 Alınan sesin çalma zamanlaması	56
2.4.2 Gecikme tahmin yöntemleri	58
2.4.3 Uyarlanabilir çalma ve zaman skalası değişimi	61
2.4.4 Paket kaybı saklama	63
2.5 PESQ.....	67
2.5.1 PESQ algoritması	67
2.5.2 PESQ tahminlerinin kabul edilebilirliği	73
3. MATERYAL VE YÖNTEM.....	75
3.1 TSM Algoritmalarıyla Sesin Sıkıştırılması	75
3.1.1 TSM ile değişken bit oranlı ses sıkıştırması	76
3.1.2 TSM algoritmalarının sentezleme kalitesinin karşılaştırma modeli	77
3.1.3 Algoritmaların ses sıkıştırılması için uygulanması	79
3.2 Veri Kaybı Gizlemesinde Gecikmenin Azaltılması	83
3.2.1 Paket kaybı gizlemesi için örtüşmeli ekleme limitleri.....	83
3.2.2 WSOLA ile paket kaybının gizlenmesinde gecikmenin azaltılması	85
3.2.3 SOLA	86

3.2.4	Paket içerisinde örtüşmeli ekleme	86
3.3	DCT ile Zaman Skalası Değişimi.....	87
3.3.1	DCT ve değiştirilmiş ters DCT dönüşümü.....	88
3.3.2	Spektral genliklerin enterpolasyonu.....	90
3.3.3	Perde dönüşümü ve enterpolasyonu	91
4.	BULGULAR VE TARTIŞMA.....	94
4.1	Ses Sıkıştırma Performanslarının Değerlendirilmesi.....	94
4.1.1	OLA ile ses sıkıştırma öznel değerlendirmesi.....	95
4.1.2	SOLA ile ses sıkıştırma öznel değerlendirmesi.....	96
4.1.3	GLS-TSM ile ses sıkıştırma öznel değerlendirmesi.....	98
4.1.4	EM-TSM ile ses sıkıştırma öznel değerlendirmesi	99
4.1.5	TD-PSOLA ile ses sıkıştırma öznel değerlendirmesi	100
4.1.6	WSOLA ile ses sıkıştırma öznel değerlendirmesi.....	101
4.1.7	AOLA ile ses sıkıştırma öznel değerlendirmesi	102
4.1.8	Suzuki-Misaki yöntemi ile ses sıkıştırma öznel değerlendirmesi.....	103
4.1.9	PICOLA ile ses sıkıştırma öznel değerlendirmesi	105
4.1.10	Faz vokoderi ile ses sıkıştırma öznel değerlendirmesi	107
4.1.11	Sinüs modelleme ile ses sıkıştırma öznel değerlendirmesi.....	108
4.1.12	SMS ile ses sıkıştırma öznel değerlendirmesi	111
4.1.13	PESQ skorlarına göre nesnel kalite değerlendirmesi	112
4.2	Uyarlanabilir Çalma Zamanlaması ve Paket Kaybı Gizlemesi Değerlendirmeleri.....	116
4.2.1	WSOLA ile paket kaybının gizlenmesinde gecikmenin azaltılması	117
4.2.2	SOLA paket kaybının gizlenmesinde gecikmenin azaltılması	119
4.2.3	Paket içerisinde örtüşmeli ekleme	120
4.2.4	Uyarlanabilir çalma zamanlaması ve paket kaybı PESQ değerlendirmesi	122
4.3	DCT Modelleme Bulguları.....	122
5.	SONUÇ.....	126
5.1	Ses Sıkıştırma İçin TSM Algoritmasının Seçimi.....	126
5.2	Uyarlanabilir Çalma Zamanlaması ve Paket Kaybı Gizlemesi için TSM Tekniğinin Seçimi	127
5.3	Değiştirilmiş DCT ile Sentezleme	128
	KAYNAKLAR	130
	ÖZGEÇMİŞ.....	136

KISALTMALAR DİZİNİ

AMDF	Ortalama Mutlak Fark (Average Magnitude Difference Function)
AOLA	Uyarlanabilir Örtüşmeli Ekleme (Adaptive OverLap Add)
AR	Kendi kendine Azalan (AutoRegressive)
DCT	Ayrık Kosinüs Dönüşümü (Discrete Cosine Transform)
DRT	Anlaşılabilirlik Teşhisi Testi (Diagnostic Rhyme Test)
DTW	Dinamik Zaman Katlama (Dynamic Time Warping)
EM-TSM	Zarf Eşlemeli Zaman Skalası Değişimi (Envelop Matching for Time Scale Modification)
FFT	Hızlı Fourier Dönüşümü (Fast Fourier Transform)
FIR	Sonlu Dürtü Tepkisi (Finite Impulse Response)
GLS-TSM	Genel Yerel Aramalı Zaman Skalası Değişimi (Global Local Search Time Scale Modification)
HF	Yüksek Frekans (High Frequency)
IDCT	Ayrık Kosinüs Ters Dönüşümü (Inverse Discrete Cosine Transform)
IEEE	Elektrik-Elektronik Mühendisleri Birliği (Institute of Electrical and Electronics Engineers)
IFFT	Hızlı Fourier Ters Dönüşümü (Inverse Fast Fourier Transform)
ISO	Uluslararası Standartlaştırma Organizasyonu (International Organization for Standartization)
ITU	Uluslararası Telekomünikasyon Birliği (International Telecommunication Union)
KHz	Bin Hertz
LPC	Doğrusal Tahminleme Kodlaması (Linear Predicting Coding)
MELP	Mixed Excitation Linear Prediction (Karışık Uyarmalı Doğrusal Tahminleme)
MOS	Ortalama fikir skoru (Mean opinion Score)
MPEG	Hareketli Resim Uzman Grubu (Moving Picture Experts Group)
ms	mili saniye
MSE	Ortalama Hata Kare (Mean Square Error)
NLMS	Normalleştirilmiş En Düşük Ortalama Kare (Normalized Least Mean Square)
NTSC	Ulusal Televizyon Sistem Komitesi (National Television System Committee)

OLA Örtüşmeli Ekleme (OLA- OverLap Add)

PAL Faz Dalgalı Çizgi (Phase Alternating Line)

PESQ Ses Kalitesinin Algısal Değerlendirmesi (Perceptual Evaluation of Speech Quality)

PICOLA İşaretleyici Aralığı Kontrollü Örtüşmeli Ekleme (Pointer Interval Controlled OverLap Add)

PSOLA Perde Senkron Örtüşmeli Ekleme (Pitch Synchronous OverLap Add)

SMS Deterministik artı Rastgele Çözümleme (Spectral Modelling Synthesis)

SNR Sinyal Gürültü Oranı (Signal to Noise Ratio)

SOLA Örtüşmeli Ekleme (OLA- OverLap Add)

SPL Ses Basınç Seviyesi (Sound Pressure Level)

STFT Kısa Zaman Fourier Dönüşümü (Short Time Fourier Transform)

TD-PSOLA Zaman Alanında Perde Senkron Örtüşmeli Ekleme (Time Domain - Pitch Synchronous OverLap Add)

TMS Geçiş Modelleme Sentezi (TMS-Transient Modeling Synthesis)

TSM Zaman Skalası Değişimi (Time Scale Modification)

VoIP İnternet Protokolü üzerinden Ses (Voice over İnternet Protocol)

WSOLA Dalgabiçimi Benzerlikli Örtüşmeli Ekleme (Waveform Similarity OverLap Add)

ŞEKİLLER DİZİNİ

Şekil 2.1	Kes yapıştır yönteminde orijinal ses, yavaşlatılmış ses ve hızlandırılmış ses için çerçevelerin gösterimi	8
Şekil 2.2	Örtüşmeli ekleme (Bonada 2002)	9
Şekil 2.3	SOLA işleminde örtüşme tahmini	12
Şekil 2.4	PSOLA algoritmasında 2 perdelik segmentler ve bu segmentler ile örtüşmeli ekleme (Bonada 2002)	19
Şekil 2.5	AOLA ile zamanda skalası değişimi (Lawlor, 1999)	22
Şekil 2.6	AOLA yavaşlatması (Lawlor, 1999)	22
Şekil 2.7	Zaman skalası değişimi ile yavaşlatma (Suziki ve Misaki 1992)	24
Şekil 2.8	PICOLA ile hızlandırma (ISO/IEC 14496-3)	27
Şekil 2.9	PICOLA ile yavaşlatma (ISO/IEC 14496-3)	28
Şekil 2.10	Zaman frekans düzlemi (Bonada 2002)	34
Şekil 2.11	Faz vokoderlerinin analiz ve sentezi (Sethares 2007)	36
Şekil 2.12	Faz vokoderlerinde analiz ve sentez	37
Şekil 2.13	İnsan sesinin modellemesi	40
Şekil 2.14	Sinüs modelleme analizi (McAuley ve Quatieri 1986a)	42
Şekil 2.15	McAuley ve Quatieri (1986b) sinüs modellemesinde analiz	44
Şekil 2.16	Gerçek frekans tepelerini bulmak için parabolik frekans enterpolasyonu (Smith ve Serra 1987)	45
Şekil 2.17	Sinüs fazlarının enterpolasyonunda kesiklik	47
Şekil 2.18	SMS (deterministik artı rastgele çözümleme) analizi	51
Şekil 2.19	SMS sentezi	52
Şekil 2.20	TMS analizi (Verma vd. 1997)	54
Şekil 2.21	TMS sentezi	55
Şekil 2.22	Çalma zamanlama yöntemlerine göre uçtan uca gecikme ve ağ gecikmesi (Liang vd. 2003)	58
Şekil 2.23	Uyarlanabilir tahminleyici adımı (DeLeon ve Sreenan 1999)	61
Şekil 2.24	Çalma zamanlamaları (Liang vd. 2003)	62
Şekil 2.25	Kayıp paketin önceki paketlerden sentezlenme gösterimi (Stenger vd. 1996)	65
Şekil 2.26	Kayıp paketin sonraki paketlerden sentezlenme gösterimi (Sanneck vd. 1996)	65
Şekil 2.27	Kayıp paketin en az tampon gecikmesi ile gizlenmesi (Liang vd. 2003)	66
Şekil 2.28	PESQ modelinin yapısı	68
Şekil 3.1	Yarı zaman çift yönlü haberleşme kanalından tam zaman çift yönlü haberleşme için WSOLA'nın kullanım	75
Şekil 3.2	Uyarlanabilir WSOLA ile sabit bit oranlı kodeklerin değişken bit oranlı kodeklere dönüştürülmesi	76
Şekil 3.3	Uyarlanabilir WSOLA'da analiz sesindeki benzerlik segmentleri ve asimetrik pencereler	77
Şekil 3.4	DCT katsayılarından ters dönüşüm ile sentezlenen periyodik sinyalin aynasal simetri özelliği (İlk ve Güler 2011)	89
Şekil 3.5	Değiştirilmiş Ters DCT ile periyodik sinyal sentezleme	90
Şekil 3.6	Aynı perde periyodu ve spektrumuna sahip iki analiz segmenti ile enterpolasyon	91

Şekil 3.7	Değiştirilmiş DCT ile perde enterpolasyonu.....	93
Şekil 4.1	OLA sentez sesleri	95
Şekil 4.2	OLA sentez sinyal spektrumları	96
Şekil 4.3	SOLA sentez sesleri	97
Şekil 4.4	SOLA sentez sinyal spektrumları.....	97
Şekil 4.5	GLS-TSM sentez sesleri.....	98
Şekil 4.6	GLS-TSM sentez sinyal spektrumları	98
Şekil 4.7	EM-TSM sentez sesleri	99
Şekil 4.8	EM-TSM sentez sinyal spektrumları	100
Şekil 4.9	PSOLA sentez sesleri	100
Şekil 4.10	PSOLA sentez sinyal spektrumları	101
Şekil 4.11	WSOLA sentez sesleri	101
Şekil 4.12	WSOLA sentez sinyal spektrumları	102
Şekil 4.13	AOLA sentez sesleri.....	103
Şekil 4.14	AOLA sentez sinyal spektrumları	103
Şekil 4.15	Suziki-Misaki yöntemi ile sentez sesleri	104
Şekil 4.16	Suziki Misaki yöntemi ile sentez sinyal spektrumları.....	105
Şekil 4.17	PICOLA sentez sesleri	106
Şekil 4.18	PICOLA sentez sinyal spektrumları.....	106
Şekil 4.19	Faz vokoderi sentez sesleri.....	107
Şekil 4.20	Faz vokoderi sentez sinyal spektrumları	107
Şekil 4.21	Sinüs tepelerinin yolları	108
Şekil 4.22	Sinüs modelleme sentez ses parçaları	109
Şekil 4.23	Sinüs modelleme sentez sinyal spektrumları	109
Şekil 4.24	MQ sinüs modelleme sentez ses parçaları.....	110
Şekil 4.25	MQ sinüs modelleme sentez sinyal spektrumları.....	110
Şekil 4.26	Spektral modelleme sentezi sesleri	111
Şekil 4.27	Spektral modelleme sentez sinyal spektrumları	112
Şekil 4.28	Simülasyonlar için veri modeli	116
Şekil 4.29	WSOLA ile uyarlanabilir çalma zamanlaması ve ağ kayıplarının gizlenmesi spektrumları	118
Şekil 4.30	Gecikmeye ve paket kayıplarına göre WSOLA ile uyarlanmış sesler	118
Şekil 4.31	SOLA ile uyarlanabilir çalma zamanlaması ve ağ kayıplarına göre zaman skalası değişimi simülasyon spektrumları	120
Şekil 4.32	Gecikmeye ve paket kayıplarına göre SOLA ile uyarlanmış sesler.....	120
Şekil 4.33	Paket içerisinde örtüşmeli ekleme ile uyarlanabilir çalma zamanlaması ve ağ kayıplarına göre zaman skalası değişimi simülasyon spektrumları..	121
Şekil 4.34	Gecikmeye ve paket kayıplarına göre paket içerisinde örtüşmeli ekleme ile uyarlanmış sesler	121
Şekil 4.35	DCT modelleme sentez sesleri.....	123
Şekil 4.36	DCT modelleme sentez sinyal spektrumları	123
Şekil 4.37	DCT modelleme ile kayıp paketlerin sentezlenmesi.....	124
Şekil 4.38	Orijinal ve prototipleri tekrar sentezlenen sinyallerin yakınsanmış parçaları.....	125
Şekil 4.39	Yaklaşık 40 ms'lik kaybın olduğu sinyal bölgesine yakınsama	125

ÇİZELGELER DİZİNİ

Çizelge 2.1 PESQ'in kabul edilebilir doğruluğunun gösterildiği faktörler, kodlayıcılar ve uygulamalar	73
Çizelge 2.2 PESQ'in tahminleme doğruluğuna güvenilemeyecek faktörler, kodlayıcılar ve uygulamalar	74
Çizelge 2.3 PESQ'in kabul edilebilir doğruluğunun henüz gösterilmediği faktörler, kodlayıcılar ve uygulamalar	74
Çizelge 4.1 TIMIT Veritabanından alınan seslerin dizin ve dosya adları	94
Çizelge 4.2 Algoritmaların hızlandırma ve yavaşlatma sentezi PESQ sonuç istatistikleri	113
Çizelge 4.3 Beş kez hızlandırma-yavaşlatma sentezinden sonraki PESQ sonuç istatistikleri	113
Çizelge 4.4 5ms segment büyüklüğüne göre WSOLA PESQ sonuç istatistikleri	116
Çizelge 4.5 Uyarlanabilir çalma zamanlaması ve paket kaybı gizlemesi algoritmalarının PESQ sonuç istatistikleri	122

1. GİRİŞ

Zaman skalası değişimi (TSM) ses sinyalinin yerel frekans içeriğini değiştirmeden süresini değiştirmekte böylece orijinal ses kayıtlarında perde ve tınıyı değiştirmeden hızlı veya yavaş onlarak dinlemeyi sağlamaktadır. Bu işlem konuşma seslerinde konuşmacının hızlı veya yavaş konuşması, müzik seslerinde temponun artması veya azalması etkilerini gerçekleştirir.

Sesi başka bir zaman skalasına dönüştürmek popüler ve faydalı bir sayısal ses işleme tekniği olduğundan birçok ses işleme uygulamalarında kullanılmaktadır (Dorran 2006). Bunlardan bazıları;

- Görüntü ve sesin post-senkronizasyonu: Genelde bir film çekiminde önce görüntüler çekilir daha sonra seslendirmeler yapılır. Birleştirme aşamasında görüntü ve sesin birbiriyle senkronizasyonları düzeltilir (Bonada 2002).
- Sinema /video formatına çevirmede: Sinema formatında saniyede 24 görüntü var iken PAL ve NTSC gibi formatlarda saniyede 25 ve 30 görüntü olmaktadır. Bu nedenle çevrim sırasında sesde de bazı işlemler gerekebilmektedir (Pallone vd. 1999).
- Ses sentezlemede: Önceden kaydedilmiş ses parçalarının (örneğin difonlar) birleştirilerek konuşma veya sesin bütünü oluşturulurken sentez esnasında bazı parçaların süresinin değiştirilmesinde kullanılmaktadır (Moulines vd. 1989, Charpentier ve Moulines 1990, Macon ve Clements 1996).
- Dil veya müzik eğitiminde: Sesin veya konuşmanın yavaşlatılması anlaşılabilirliğini arttırmaktadır (Eroğul ve Karagöz 1998, Donnellan vd. 2003, Demol vd. 2004). Müzik ritminin öğrencinin enstrümanı çalma yeteneğine göre yavaşlatılması çalışmayı kolaylaştırmaktadır.

- Telesekreter mesajlarının veya diktafonların yavaş veya hızlı dinlenmesi: Bırakılan mesajları hızla taramak zaman kazandırmakta; seçilen mesajı yavaş dinlemek de anlaşılabilirliği arttırmaktadır (Hejna 1990).
- Sayısal kütüphaneler ve uzaktan eğitimde ses materyallerinin hızlı taranması: Kullanıcıların ses verilerini hızlı taramasına imkan sağlamaktadır (Wong 1998, Omoigui vd. 1999).
- Müzik aranjmanında: Önceden kaydedilmiş sesler için tempo değiştiren bir araç olarak kullanılmaktadır. Buna ek olarak birbiri ardına çalınan müzik parçalarında ritim geçişini yumuşatmak için de kullanılmaktadır (Bonada 2002).
- İletişim kanallarındaki çerçeve kayıplarının ve gecikmelerin gizlenmesi: Gerçek zamanlı uygulamalarda iletişimde kaybolan veya geciken ses çerçevelerinin dinleyiciye hissettirmeden giderilmesi için diğer ses çerçeveleri uzatılmaktadır (Liang vd. 2001).
- Ses sıkıştırma: Sesi iletişim kanalından göndermeden önce hızlandırmak ve karşı tarafta yavaşlatmak iletişim bandının daha uygun kullanımını sağlamaktadır (Eroğul vd. 2001, İlk ve Güler 2006).

Sesin zaman skalası değişimi için literatürde çok fazla teknik bulunmaktadır. Uygulanacak tekniğin seçiminde üretilen sesin kalitesi ve algoritmanın işlem basitliği en önemli kriterlerdir. Zaman skalası değişimi teknikleri, genel olarak zaman alanı teknikleri, frekans alanı teknikleri ve sinyal modelleme teknikleri olmak üzere üç ana kısımda kategorize edilmektedir (Bonada 2002).

Zaman alanı teknikleri, analiz ve sentez işlemlerini zaman alanında gerçekleştirmektedir. Analiz sinyalinin alınan ses parçalarının benzerliği aranmakta, bulunan benzerliklere göre örtüşmeli ekleme yapılarak sentezleme gerçekleştirilmektedir. Benzerlik araması analiz sesinden alınan segmente göre SOLA

ve AOLA’da olduđu gibi sentez sesi üzerinde veya WSOLA, PSOLA ve PICOLA’da olduđu gibi analiz sesi üzerinde yapılabilir. Örtüşmeli ekleme segment uzunluđu da kullanılan tekniđe göre sabit veya deđişken olabilir. Benzerlik aramalarda ilintinin en yüksek deđeri, ortalama mutlak farkın (AMDF) en küçük deđeri, fark kare ortalamasının en küçük deđeri, tepe noktalarının eşleştirilmesi, perdelerin bulunması gibi yöntemler kullanılmaktadır. Zaman skalasının deđişimi; benzerlik arama ve/veya sentez segmenti büyüklüđu, ardışık analiz segmentleri arasındaki ortalama uzaklık ve ardışık sentez segmentleri arasındaki ortalama uzaklık parametreleri deđiştirilerek sağlanmaktadır. Skalanın düşürülmesi olarak ifade edilen konuşmanın hızlandırılması veya müzik seslerinde temponun artırılması; analiz sinyalinin alınan segmentlerin arasındaki uzaklığın artırılması, sentez segmenti boyutunun düşürülmesi, ardışık sentez segmentleri arasındaki uzaklıkların düşürülmesi veya iki benzer analiz segmentinin örtüştürülerek tek sentez segmenti sentezleme ile sağlanmaktadır. Skalanın artırılması ise azaltılmasında yapılan işlemlerin tersi ile mümkün olmaktadır.

Frekans alanı teknikleri ise analiz sinyalini segmentlere bölerek Kısa Dönem Fourier dönüşümü (STFT) ile spektrumu elde etmekte ve ardışık segmentler arasındaki spektrumun deđişimini modellemektedir. Faz vokoderleri ve türevleri olan bu tekniklerde zaman skalası deđişimini spektrum deđişim modelinin zamanlamasını artırarak veya düşürerek frekans alanında gerçekleştirmekte ve sentez spektrumlarını ters STFT ile zaman alanına çevirerek sentez sinyalini oluşturmaktadır. STFT için ayrık zaman dönüşümü Hızlı Fourier dönüşümü (FFT) ile gerçekleştirilmektedir.

Sinyal modelleme teknikleri ise analiz aşamasını frekans alanında gerçekleştirmekte sentezi ise genelde zaman alanında gerçekleştirmektedir. Tez kapsamında incelenen teknikler genelde sinüs modelleme tabanlı olup sesin üretilmesini (yol-kaynak modeline göre) modellemektedir. Sinüs modellemenin analiz aşaması faz vokoderlerine benzer şekilde analiz sesinin STFT spektrumunu çıkarmaktadır. Spektrumun genlik tepelerine göre baskın sinüsleri genlik, frekans ve faz kaymaları ile belirlenmektedir. Ardışık analiz segmentleri arasındaki baskın sinüsler eşleştirilerek genlik, frekans ve faz deđişimleri modellenmektedir. Sentez aşamasında ise bu analiz modeli parametrelerinin enterpolasyonu ile skalası deđiştirilmiş sentez modeli oluşturulmakta ve sentez sinyali

bu sentez modelinin sinüslerinin genlik, frekans ve faz modellerine göre zaman alanında üretilmektedir.

1.1 Amaç ve Kapsam

Bu tez çalışmasının amacı iletişim kanallarında olabilecek tıkanıklık veya bant kapasitesi yetersizlikleri durumlarında ses iletişiminin sağlanabilmesi veya devamlılığının ve kalitesinin korunması için TSM tekniklerinin uygulanmasıdır.

Devre anahtarlama ağılarda önceden tesis edilmiş haberleşme oturumları bazı kısımlardaki haberleşme bant genişliğinin tamamını doldurabilir. Bu durumda ses iletişimi için yeni bir haberleşme oturumu başlatılmak istendiğinde yeni tesis edilecek devre bağlantısı bu tıkanık kısımların da kullanımını gerektirebilir. Tıkanık kısımlardan akan diğer haberleşme verileri sıkıştırılamaz ise oturum başlatılamamaktadır.

Paket anahtarlama ağlardan ses iletiminde ise oturumunun başlangıcında tıkanıklık olmaz ise bağlantı yapılmakta, fakat daha sonraki tıkanma durumlarında gecikme veya paket kayıpları artabilmektedir. Bu durum ses iletişiminin kalitesini düşürmektedir.

Tıkanıklık durumu önceden öngörülebiliyor ise devam eden ve yeni gelen ses oturumlarında iletilen seslerin zaman skalaları hızlandırılarak kodlanması ile iletişim kanalının bant genişliği daha verimli kullanılabilir. Bu durumda özellikle VoIP veya telefon gibi gerçek zamanlı iletişim için alıcı tarafından alınan hızlandırılmış sentez sesinin kod çözümü yapıldıktan sonra tekrar orijinal skalaya döndürülmesi gerekmektedir. Kullanılacak TSM algoritmasının hem hızlandırma hem de yavaşlatma için sentezleme kalitesinin yüksek, algoritma gecikmesi düşük ve skala değişim oranı kontrol edilebilir olması önemlidir.

Özellikle paket anahtarlama ağılarda gerçek zamanlı ses iletişimde kanal gecikmesi oturum esnasında artabilmekte veya kanal kapasitesinin yetersiz kalmasından dolayı veri kaybı oluşabilmektedir. Kanal gecikmesinin artması veya kanalda veri kaybının oluşması, kanalın alıcı tarafında alınmış olan tüm ses verilerinin sentezlenerek

çalınmasına ve çalma belleğinin boşalmasına neden olabilir. Bu durum yeni ses paketi ulaşınca kadar çalınan sinyalde bir boşluğa neden olmaktadır. Dinlenen sesin anlık kopmalar nedeniyle kalitesi ve anlaşılabilirliği azalmaktadır. TSM algoritmaları, gecikmelerde çalma belleğini mevcut ses verilerinin zaman skalasını artırarak sentezlenen ses ile doldurmakta, gecikme azalmalarında çalma belleğindeki mevcut ses verilerinin skalasını düşürerek uçtan uca gecikmeyi düşürebilmektedir. Kanalda veri kaybı oluşması durumunda da kaybolan veriler TSM algoritmaları ile mevcut verilerden sentezlenebilmektedir.

Tezin KURAMSAL TEMELLER Bölümü altında TSM teknikleri ve uyarlanabilir çalma zamanlaması ve paket kaybı gizlemesi için literatürdeki çalışmalar anlatılmıştır. İlk üç alt bölümde sırasıyla zaman alanı, frekans alanı ve sinyal modelleme TSM teknikleri algoritma detayları ile yer almaktadır. Son alt bölüm ise bu tekniklerin uyarlanabilir çalma zamanlaması ve paket kaybı için uygulanmasına yönelik teknikler detaylandırılmıştır.

MATERYAL VE YÖNTEM Bölümü'nde ise tez kapsamında yapılan çalışmalar üç alt bölüm altında anlatılmıştır. İlk altbölüm TSM algoritmaları ile iletişim kanallarından iletilecek sesin sıkıştırılmasına yönelik olup daha önceki bölümde yer alan TSM tekniklerinin hızlandırma ve yavaşlatma modeline göre uygulanmasına yönelik uyarlamalarını ve algoritma parametrelerinin belirlenmesini içermektedir. İkinci alt bölümde ise uyarlanabilir çalma zamanlaması ve paket kaybı gizlemesi için WSOLA, SOLA ve tez kapsamında önerilen teknik ile uygulanmasına yönelik modeller anlatılmaktadır. Son alt bölümde ise yeni bir frekans alanı TSM tekniği olarak DCT modelleme sunulmaktadır.

BULGULAR VE TARTIŞMA Bölümü'nde bir önceki bölüme paralel üç alt bölüm altında gerçekleştirilen testler ve bu testlerin değerlendirilmesini içermektedir. Testler TIMIT ses veritabanından alınan sesler kullanılarak gerçekleştirilmiştir. Sentezlenen seslerin dinlenmesi ve sinyal grafiklerinin incelenmesi ile gerçekleştirilen öznel kalite değerlendirmelerinin yanında ITU-T P.862'de tanımlanan PESQ skorları da kullanılmıştır. Ses sıkıştırılmalarına yönelik testlerde alınan orijinal sese önce

hızlandırma uygulanmış, hızlandırılan bu sese daha sonra yavaşlatma uygulanmıştır. Uyarlanabilir çalma zamanlaması ve paket kaybı için ise ağ gecikmesi ve uçtan-uca gecikme simülasyonu için veri modeli kurulmuş, algoritmalar bu veri modeline uygun olarak çalıştırılmıştır. DCT modellemenin sıkıştırma performansı hızlandırma-yavaşlatma sentezlemesi ile değerlendirilmiş, prototip enterpolasyonu performansı da PSOLA ile karşılaştırılarak verilmiştir.

2. KURAMSAL TEMELLER

2.1 Zaman Skalası Değişimi İçin Zaman Alanı Teknikleri

Konuşmanın hızını değiştirmek için sesin kayıt hızı ile çalma hızını farklı tutmak yeterli değildir. Örneğin hızı kontrol edilebilen bir kayıt hızlı olarak çalındığında seslerin oldukça değiştiği hatta hız iki katına çıktığında seslerin anlaşılmadığı, dinleyen herkes tarafından fark edilebilir. Orijinal (analog) ses $x_a(t)$ ile bu sesin zaman alanında α katı kadar hızlandırılmış hali $y_a(t)$ ile gösterildiğinde aralarındaki bağıntı Eşitlik 2.1 ile ifade edilmektedir (Verhelst 2000).

$$y_a(t) = x_a(\alpha t) \quad (2.1)$$

Bu eşitliğin Fourier dönüşümleri alındığında frekans alanındaki gösterimi Eşitlik 2.2 ile ifade edilir (Verhelst 2000).

$$Y_a(\Omega) = \frac{1}{|\alpha|} x_a\left(\frac{\Omega}{\alpha}\right) \quad (2.2)$$

Eşitlik 2.2’de açıkça belirli olduğu gibi zaman alanındaki bu tür bir sıkıştırma veya genişletme frekans alanında da sırasıyla genişletme veya sıkıştırmaya neden olmaktadır. Bu nedenle konuşmanın hem kısa dönem (zarfı veya LPC katsayıları) hem de uzun dönem (perde periyodu ve diğer uzun zaman parametreleri) ilişkileri bu işlem ile değiştirilmektedir. İnsanın ses algılaması frekans alanında olduğu için anlaşılabilirlik bozulmaktadır.

Bu nedenle ilk TSM uygulamalarında, istenen zaman skalasını elde etmek için TSM sisteminin giriş sesi birbirleriyle örtüşmeyen 20-30 ms zamanlı çerçevelere ayrılmakta ve bu çerçevelerden uygun olanlar atılıp veya tekrarlanıp giriş sesi hızlandırılmakta veya yavaşlatılmaktadır (Lee 1972). Kes yapıştır olarak adlandırılan ve şekil 2.1’de de gösterilen bu yapı algoritmik olarak basit ve hızlı olmakla birlikte seste bazı

bozukluklar meydana getirmektedir. Bu bozukluklar sesin kes yapıştır noktalarındaki sürekliliklerin ve en önemlisi perde periyotlarının bozulmasından meydana gelmektedir.



Giriş sesi ve çerçeveleri



Yavaşlatılmış zaman skalası (%150 süre)



Hızlandırılmış zaman skalası(%75 süre)

Şekil 2.1 Kes yapıştır yönteminde orijinal ses, yavaşlatılmış ses ve hızlandırılmış ses için çerçevelerin gösterimi

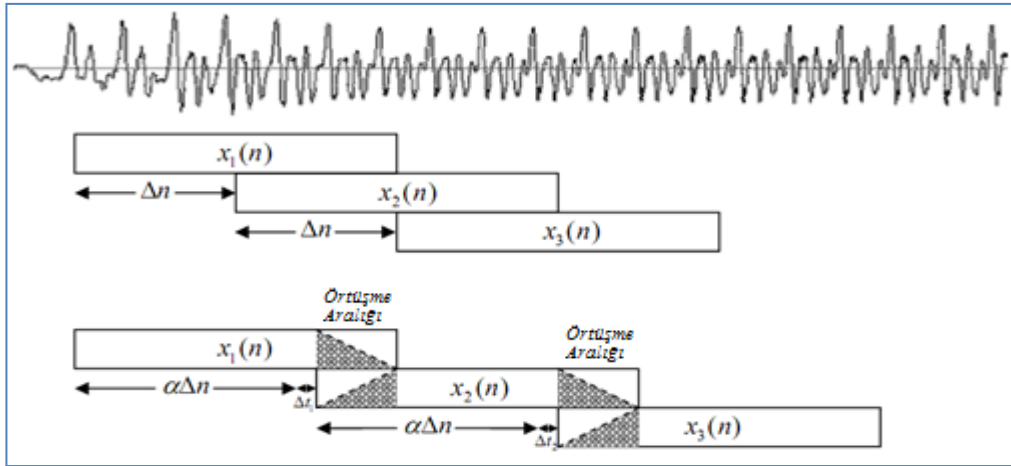
Bu bozulmaları gidermek için sesi çerçevelere ayırırken sürekliliği dikkate alan (lokal tepeler veya perde periyoduna göre çerçeveler) kes yapıştır algoritmaları geliştirilmiş ve zaman skalası değiştirilen seslerin kalitesi arttırılmıştır (Roucos ve Wilgus 1985, Malah 1979).

2.1.1 OLA

Kes yapıştır yöntemleri perde periyodu sürekliliğine dikkat etse dahi bir çerçevenin eklenmesi veya çıkarılması sinyalde sürekliliği istenen düzeyde sağlamayabilmektedir. Çünkü normal bir konuşmada sesler aynı gibi görünse de ton ve vurgu ardışık perde periyotlarında değişebilmektedir. Ton ve vurgu değişimi geçişlerde (sesli-sessiz veya sessiz-sesli geçişleri) daha da fazla olduğundan bu bozukluk daha da belirgin hale gelmektedir.

Bu bozuklukları düzeltmek için ekleme/çıkarma yerine örtüşmeli ekleme (OLA-OverLap Add) kullanılmaktadır. OLA yönteminde skalası değiştirilecek ses

birbirlerinden Δn uzaklıkta boyutu N olan çerçeveler kullanılmaktadır (Roads 1978). Daha sonra ses sentezlenirken çerçeveler arasındaki uzaklıklar $\alpha\Delta n$ olarak değiştirilir. Önceki çerçevelerin son kısımları ile ardındaki çerçevelerin ilk kısımlarında örtüşme olacağından, örtüşen noktalarda önceki çerçevenin değerlerine giderek azalan ardındaki çerçevenin örtüşen değerlerine giderek artan bir ağırlık verilir ve ağırlıklı toplam ile örtüşen yerlerin değerleri bulunur. Bu işlemde ardışık iki parçanın örtüşen tüm noktalarında kullanılan ağırlıkların toplamı 1'e eşit olacak, bu ağırlıklar önceki çerçevenin örtüşen kısmı için monoton azalan diğeri için monoton artan olacaktır. Ağırlık olarak doğrusal ağırlık (Barlet penceresinin ilk ve ikinci yarısı) kullanılabileceği gibi, Hamming, Hanning vb. pencereler de kullanılabilir. Bu OLA işlemi Şekil 2.2'de gösterilmektedir.



Şekil 2.2 Örtüsmeli ekleme (Bonada 2002)

Şekil 2.2'de de gösterilen örtüsmeli ekleme, $f(j)$ fonksiyonu katlama ağırlıkları, L ise örtüşen örneklem sayısı olmak üzere Eşitlik 2.3 ile ifade edilir.

$$y(m.\alpha\Delta n + j) := (1 - f(j))y(m.\alpha\Delta n + j) + f(j)x(m\Delta n + j) \quad ; 0 \leq j \leq L-1 \quad (2.3.a)$$

$$y(m.\alpha\Delta n + j) = x(m\Delta n + j) \quad ; L \leq j \leq N-1 \quad (2.3.b)$$

OLA işlemi diğer birçok başarılı zaman skalası değişimi tekniklerine temel oluşturmakla birlikte örtüşen kısımlar birbirleriyle benzer olamayacağı için süreklilik ve perde periyodunun kaymasından dolayı sentezlenen seste bozukluklar oluşmaktadır.

2.1.2 SOLA

OLA tekniğindeki ardışık çerçeveler arasındaki uzaklığın değişmesinden dolayı sinyal sürekliliği bozulduğundan sinyal sürekliliğini sağlamak için katlanan ardışık çerçevenin birbirleriyle uyumlu hale getirilmesi bu problemleri ortadan kaldırabilir. Bu nedenle SOLA’da benzer yerlerin katlanması amaçlanmaktadır (Roucos ve Wilgus 1985).

SOLA’da giriş sesi x ; S_a (analiz adımı) örneklem aralıklarıyla N uzunluğundaki çerçevelere ayrılmaktadır ($\Delta n = S_a$). Sentez sesinde ise çerçeveler birbirlerinden S_z (sentez adımı) örneklem aralıklarının benzerliği sağlayan k komşuluğunda örtüştürülür. ($\alpha \Delta n = S_z$) Tüm çerçeveler için k değerleri ayrı ayrı hesaplandığından ardışık iki sentez çerçevesi arasındaki uzaklık (“ m ” ve “ $m-1$ ” çerçevesi) $S_z + k_m - k_{m-1}$ olarak hesaplanır. Optimum k_m değeri $k_{min} \leq k_m \leq k_{max}$ aralığında aranır. Optimum k_m değerini hesaplamak için;

Eşitlik 2.4’te ifade edilen normalleştirilmiş ilintinin en yüksek değeri (Roucos ve Wilgus 1985),

Eşitlik 2.5’de ifade edilen basitleştirilmiş ilintinin en yüksek değeri (Verhelst ve Roelands 1993, Lawlor ve Fagan 1999),

Eşitlik 2.6’da ifade edilen ortalama mutlak farkın en düşük değeri (Verhelst ve Roelands 1993)

Eşitlik 2.7’de ifade edilen fark kare ortalamasının en düşük değeri gibi yöntemler kullanılır (Wong vd. 1998).

$$R_m(k) = \frac{\sum_{j=0}^{L_k-1} y(mS_s + k + j)x(mS_a + j)}{\sqrt{\sum_{j=0}^{L_k-1} y^2(mS_s + k + j) \sum_{j=0}^{L_k-1} x^2(mS_a + j)}} \quad (2.4)$$

$$R_m(k) = \frac{\sum_{j=0}^{L_k-1} y(mS_s + k + j)x(mS_a + j)}{\sqrt{\sum_{j=0}^{L_k-1} |y(mS_s + k + j)| \sum_{j=0}^{L_k-1} |x(mS_a + j)|}} \quad (2.5)$$

$$AMDF(k) = \frac{\sum_{j=0}^{L_k-1} |y(mS_s + k + j) - x(mS_a + j)|}{L_k} \quad (2.6)$$

$$E_m(k) = \frac{\sum_{j=0}^{L_k-1} (y(mS_s + k + j) - x(mS_a + j))^2}{L_k} \quad (2.7)$$

Yukarıdaki eşitliklerde L_k ; k değerine göre örtüşen örneklem sayısıdır. Bu eşitliklere göre k_m değeri seçildikten sonra zaman skalalı ses sentezi Eşitlik 2.8'e göre yapılmaktadır.

$$y(mS_z + j) := (1 - f(j))y(mS_z + j) + f(j)x(mS_a + j), \quad 0 \leq j \leq L_k - 1 \quad (2.8.a)$$

$$y(mS_z + j) = x(mS_a + j), \quad L_k \leq j \leq N - 1 \quad (2.8.b)$$

Eşitlik 2.8.a'da yer alan "işareti $y(mS_z + j)$ değerinin güncelleneceğini göstermektedir. İşaretin sol tarafındaki yeni değer işaretin sağ tarafındaki eski değerini de kullanarak güncellenecektir. $f(j)$ ise doğrusal olarak artan bir fonksiyondur ve değeri Eşitlik 2.9'da verilmektedir.

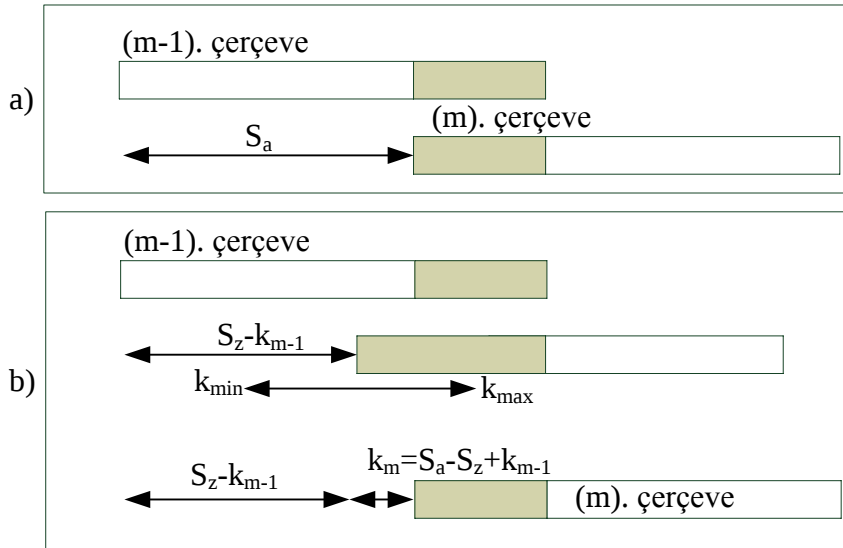
$$f(j) = \begin{cases} 0, & j < 0 \\ j/(L_k - 1), & 0 \leq j < L_k \\ 1, & j \geq 1 \end{cases} \quad (2.9)$$

SOLA ile sentezlenen sesin kalitesi parametrelerin seçimine de bağlıdır. İlk uygulamalarda N ; konuşma sesi için 30 ms, müzik sesi için 40 ms seçilirken, $S_a = N/2$ $k_{min} = -N/2$ $k_{max} = N/2$ seçilmiştir (Roucos ve Wilgus 1985). Makhoul vd. (1986), S_a

değerini hızlandırma için $N/2$, yavaşlatma için ise $N/2\alpha$ olarak önermiştir. Wong vd. (2003) ise S_a değerini hem hızlandırma hem de yavaşlatma için $N/4\alpha$ olarak seçmiştir.

Hejna (1990) ile Wong vd. (2003) bazı özel durumlarda optimum k değeri için hesaplamalara gerek olmadığını belirtmişlerdir. $S_a - S_s + k_{m-1}$, k_{min} ile k_{max} arasına denk geldiğinde $k_m = S_a - S_s + k_{m-1}$ için ardışık iki pencerenin örtüşen kısımları analiz sesinin aynı parçası olmakta ve en yüksek benzerlik sağlanmaktadır. Bu durum Şekil 2.3'te de gösterilmektedir.

$$k_m = S_a - S_s + k_{m-1} \quad (2.10)$$



Şekil 2.3 SOLA işleminde örtüşme tahmini

a. Orijinal sesteki çerçeveler arası uzaklık, b. SOLA sentez sesindeki çerçeveler arası uzaklık

2.1.3 GLS-TSM

GLS-TSM (Genel, yerel arama- zaman skalası değişimi) SOLA tabanlı bir tekniktir (Yim ve Pawate 1996). SOLA'da ardışık çerçevelerde örtüşecek benzer noktaları bulmak için ilinti hesapları çok fazla işlem gerektirdiğinden bu algorithmada bunun yerine önce genel benzerliğe, daha sonrada bu genel benzerlikte bulunan değer komşuluğunda yerel benzerliğe bakılır.

Genel benzerlikte sentez sesi ile analiz sesinin örtüşen kısımları için sıfır kesme noktalarının sayısına bakılmaktadır. Ardışık iki nokta farklı işaretlerde ise burada bir sıfır kesişi vardır demektir. Sıfır kesişleri sinyalin frekans içeriğiyle alakalı olduğundan benzer yerlerde sıfır kesiş sayıları hemen hemen aynıdır. Sıfır kesiş noktaları sayısı Eşitlik 2.11'deki gibi hesaplanmaktadır.

$$z = \sum_{m=0}^{L-1} |sg(x[k+m]) - sg(x[k+m+1])| \quad (2.11)$$

Burada $x(m)$ pozitif ise $sg(x(m)) = 1$; negatif ise $sg(x(m)) = 0$ 'dır. k değeri ise benzerlik arama aralığında ($k_{min} \leq k \leq k_{max}$) herhangi bir değerdir. L ise örtüşen noktalardaki örneklem sayısıdır. Analiz ve sentez seslerinin örtüşme noktalarında Eşitlik 2.11'e göre hesaplanan z değerleri farklarının en küçük olduğu k indisi genel benzerlik sonucudur ve $k_{golbalmin}$ olarak adlandırılır.

Genel benzerlik değeri bulunduktan sonra daha net benzerliğe ulaşmak için yerel benzerliğe bakılır. Yerel benzerlikte sıfır noktaları için 11 boyutlu bir vektör tanımlanır. Bu vektörün elemanları Eşitlik 2.12'de verilmektedir.

$$\begin{aligned} f_1 &= x(i) - x(i+1) \\ f_2 &= |x(i)| \\ f_3 &= |x(i+1)| \\ f_4 &= (x(i) - x(i+2)) / 2 \\ f_5 &= |x(i+2)| \\ f_6 &= (x(i) - x(i+3)) / 3 \\ f_7 &= |x(i+3)| \\ f_8 &= (x(i-1) - x(i+1)) / 2 \\ f_9 &= |x(i-1)| \\ f_{10} &= (x(i-2) - x(i+1)) / 3 \\ f_{11} &= |x(i-2)| \end{aligned} \quad (2.12)$$

Yerel benzerliğe bakılırken x sinyalinin örtüşen değerlerinde sıfırı en büyük açıyla kesen nokta bulunur. Bu noktalarda yerel benzerlik vektörü f_x hesaplanır. Sentez

sinyali y 'nin $k_{golbalmin}$ komşuluğundaki sıfır noktaları için yerel benzerlik vektörleri $f_{y,i}$ hesaplanır. Daha sonra f_x vektörü ile $f_{y,i}$ vektörleri arasındaki mesafelere bakılır. Mesafenin en küçük olduğu noktaya göre en iyi örtüşme indisi k bulunur. Mesafenin hesaplanması için Eşitlik 2.13 kullanılmaktadır.

$$d_{k,i} = \frac{1}{11} \sum |f_x[j] - f_{y,i}[j]| \quad (2.13)$$

Yim ve Pawate (1996) GLS-TSM'i elektronik klavye, üflemeli çalgı ve yaylı çalgının da aralarında bulunduğu çeşitli enstrümanların sesleriyle test etmişler ve yüksek kaliteli zaman skalalı ses elde etmişler. 44.1 KHz'deki sesler için SOLA'dan kırk kat daha hızlı olduğunu belirtmektedirler.

Yim ve Pawate (1996), yerel benzerlik aramasının $k_{globalmin}$ komşuluğunda arandığı belirtilmiştir. Fakat bunun ne kadar komşuluğunda olduğu açık değildir. Bu konu başka kaynaklardan da araştırılmış fakat bu değer bulunamamış hatta bazı kaynaklarda bu sorun dile getirilerek kendileri tarafından bu değerler seçilmiştir. Wong vd. (1998); $k_{globalmin}$ komşuluğunda 7 adet benzerlik arama sıfır kesiş noktasında mesafe ölçümü yapmıştır. Dorran vd. (2006) $k_{golbalmin}$ komşuluğunda 10 adet benzerlik arama sıfır kesiş noktasında mesafe ölçümü yapmıştır. Her iki kaynakta da sentezlenen sesin kalitesinin kötü olduğu belirtilmiştir. Pawate'nin de yer aldığı bir makalede GLS-TSM'in genel aramaları 400 ms'lik çerçevelerde yapıldığı belirtilmektedir (Agnihotri vd. 2002). 400 ms'lik ses çerçeveleri başta konuşma olmak üzere birçok ses için oldukça büyüktür.

2.1.4 EM-TSM

EM-TSM (Zarf Eşlemeli Zaman Skalası Değişimi) SOLA'da kullanılan normalleştirilmiş ilinti araması yerine analiz ve sentez seslerinin işaretlerinin ilintisini kullanmaktadır (Wong vd. 1998). Ses örneklerin işaretleri için Eşitlik 2.14 - 2.15'teki fonksiyonlar tanımlanmaktadır (Wong vd. 2002).

$$x_2(i) = \text{sign}(x(mS_a + i)) = \begin{cases} 1, & x(mS_a + i) \geq 0 \\ -1, & x(mS_a + i) < 0 \end{cases} \quad (2.14)$$

$$y_2(i) = \text{sign}(y(mS_s + i)) = \begin{cases} 1, & y(mS_s + i) \geq 0 \\ -1, & y(mS_s + i) < 0 \end{cases} \quad (2.15)$$

$y_{2,k}(i) = y_2(k + i)$ olarak bir fonksiyon tanımlansın. Bu fonksiyonun örtüşme noktalarındaki örnekler kümesi $y_{2,k} = \{y_{2,k}(i), i = 0, 1, \dots, L_{k-1}\}$ olacaktır. Aynı şekilde $x_{2,k}(i) = x_2(i)$ olarak bir fonksiyon tanımlansın. Bu fonksiyonun örtüşme noktalarındaki örnekler kümesi $x_{2,k} = \{x_{2,k}(i), i = 0, 1, \dots, L_{k-1}\}$ olacaktır. Bu iki fonksiyonun örtüşme noktalarındaki çarpımı $z_{2,k}(i) = x_{2,k}(i)y_{2,k}(i)$ olacaktır ve bu fonksiyonun örnekler kümesi $z_{2,k} = \{z_{2,k}(i), i = 0, 1, \dots, L_{k-1}\}$ olacaktır (Wong vd. 2002). Buna göre işaretlerin normalleştirilmiş ilintisi (EMF / Zarf eşleme fonksiyonu) Eşitlik 2.16'da ifade edildiği gibi olacaktır.

$$R_m(k) = \frac{\sum_{j=0}^{L_k-1} \text{sign}\{y(mS_s + k + j)\} \text{sign}\{x(mS_a + j)\}}{\sqrt{\sum_{j=0}^{L_k-1} \text{sign}\{y(mS_s + k + j)\}^2 \sum_{j=0}^{L_k-1} \text{sign}\{x(mS_a + j)\}^2}} = \frac{\sum_{j=0}^{L_k-1} z_{2,k}(j)}{L_k} \quad (2.16)$$

A_k adında bir küme tanımlansın. Bu küme örtüşme alanındaki analiz sesinin sıfır kesiş noktalarını versin. Ardışık 2 örneğin çarpımları -1 olan değerler bu kümenin elemanları olacaktır. $A_k = \{j : x_{2,k}(j-1)x_{2,k}(j) = -1, \quad 0 < j \leq L_k - 1\}$ kümesinin eleman sayısı M_k ile gösterilsin.

B_k adında bir küme tanımlansın. Bu küme örtüşme alanındaki sentez sesinin sıfır kesiş noktalarını versin. Ardışık 2 örneğin çarpımları -1 olan değerler bu kümenin elemanları olacaktır. $B_k = \{j : y_{2,k}(j-1)y_{2,k}(j) = -1, \quad 0 < j \leq L_k - 1\}$ kümesinin eleman sayısı N_k ile gösterilsin.

C_k adında $C_k = A_k \oplus B_k = (A_k \cap B_k^c) \cup (A_k^c \cap B_k)$ olan bir küme tanımlansın. $A_k \cap B_k$ kümesinin eleman sayısı r_k ile gösterilirse, $r_k \leq M_k$ ve $r_k \leq N_k$ olacaktır. C_k kümesinin eleman sayısı $c_k = M_k + N_k - 2r_k$ olacaktır. Bu üç kümenin elemanları aşağıdaki gibi sıralanır.

$$A_k = \{a_{k,1}, a_{k,2}, \dots, a_{k,M_k}\}$$

$$B_k = \{b_{k,1}, b_{k,2}, \dots, b_{k,N_k}\}$$

$$C_k = \{c_{k,1}, c_{k,2}, \dots, c_{k,M_k+N_k-2r_k}\}$$

Burada C_k kümesi; $z_{2,k}$ fonksiyonu için sıfır kesiş noktalarını gösterdiği Wong vd. (2002) tarafından ispatlanmaktadır.

$\beta_{z,k} = z_{2,k}(0)$ olursa $\beta_{z,k} = 1$ ise $R(k)$ hesaplamasının payında $c_{k,1}$ noktasına kadar hep artı gelecek daha sonra işaret değişeceği için – değerler gelmeye başlayacak, daha sonra tekrar işaret değişirse ($c_{k,2}$ için) tekrar artı değerler gelmeye başlayacaktır. Bu nedenle $R_m(k)$ Eşitlik 2.17’deki gibi yazılabilir.

$$R_m(k) = \frac{1}{L_k} \sum_{j=0}^{L_k-1} z_{2,k}(j) = \frac{\beta_{z,k}}{L_k} \left[2 \sum_{j=1}^{M_k+N_k-2r_k} (-1)^{j+1} c_{k,j} + (-1)^{M_k+N_k} L_k \right] \quad (2.17)$$

Eşitlik 2.17, Eşitlik 2.4’e göre hesaplanması çok daha kolay olmasına rağmen $k_{min} \leq k \leq k_{max}$ aralığındaki tüm k değerleri için ilinti bulunmalıdır. Bu hesaplama tekrarlarını azaltmak için sınırlı sayıda k değerleriyle ilintinin hesaplaması durumunda diğer k değerleri için ilintinin bu sınırlı sayıda hesaplanan ilintilere göre daha basit işlemlerle hesaplanabilmektedir (Wong vd. 2003). Bu sınırlı sayıdaki k değerlerinin kümesi K_o ile gösterilirse bu küme Eşitlik 2.18’de verilmektedir.

$$K_o = \{k : A_k \cap B_k \neq \emptyset\} \cup \{k : b_{k-1,1} = 1 \cup b_{k-1,N_k} = N-1\} \cup \{k : a_{k-1,M_k} = L_k - 1 \cap L_k < N\} \cup \{k_{min}, k_{max}\} \quad (2.18)$$

$K_o = \{k_1, k_2, k_3, \dots, k_{Q-1}\}$ elemanları için ilinti Eşitlik 2.17'ye göre bulunduktan sonra diğer k elemanları için normalize edilmiş ilinti değeri Eşitlik 2.19'da verilen eşitliğe göre hesaplanmaktadır (Wong vd. 2003).

$$R_m(k_i + 1) = \frac{L_k}{L_k - (k_{i+1} - k_i)} R_m(k_i) + (k_{i+1} - k_i) \frac{2\beta_{z,k_i} \xi_{k_i} + \beta_{z,k_i} (-1)^{M_k + N_{k_i} + 1}}{L_k - (k_{i+1} - k_i)} \quad (2.19)$$

Eşitlik 2.19'da yer alan $\xi_{k_i} = \sum_{j=1}^{N_k} (-1)^{g(k_i, j)}$ olup $g(k_i, j)$ ise $b_{k_i, j}$ lerin $C_{k_i, j}$ setindeki yerleridir.

2.1.5 TD-PSOLA

Ses sinyalleri (monofon enstrüman veya konuşma) uzun dönem ilişkisini sağlayan perdelere sahiptir. PSOLA (Perde senkronlu örtüşmeli ekleme) ses sinyalinin perdelerini bulup 2 perde uzunluğundaki kısımlara ayırmakta ve örtüşmeli eklemeleri tespit edilen perdelere göre yapmaktadır (Moulines vd. 1989, Moulines ve Charpentier 1990). PSOLA algoritması ilk olarak analiz sesinin ötümlü ve ötümsüz kısımlarını bulmaktadır. Ötümlü kısımları için sesin perdesini bulup işaretler, ötümsüz kısımlar içinse perde işaretlerini sabit uzaklıklarla koyar.

Zaman alanında PSOLA (TD-PSOLA) perde tespitini zaman alanında yapar. Örtüşmeli ekleme prosedürü perde senkronlu yapıldığından sentezlenen sesin kalitesi yüksektir. PSOLA algoritması analiz ve sentez olmak üzere iki kısımdan oluşmaktadır. Analiz kısmında giriş sesi çerçevelere ayrılır. Sentez kısmında ise bu çerçeveler birleştirilerek zaman skalası değiştirilmiş ses sentezlenir.

Analizde ilk adım sesin periyodik kısımlarının maksimum noktalarına ve ötümsüz kısımlarının sabit uzaklıklı örneklemelerine perde işaretini t_i koymaktır. Perde periyodu $P(t_i)$, zamana bağlı olarak işaretlenmiş örneklemelerin arasındaki uzaklıktan $P(t_i) = t_i - t_{i-1}$ olarak hesaplanır. Daha sonra giriş sesi perde işaretlerinin olduğu örneklemeler

ortalananak şekilde parçalara ayrılır. Her parça 2 perde periyotluk Hanning penceresi ile pencerelenir.

Sentezde hızlandırma/yavaşlatma faktörü α , sentez sinyalinin perde periyotlarının uzunluğu $\tilde{P}(\tilde{t})$ 'yi belirler (Eşitlik 2.20).

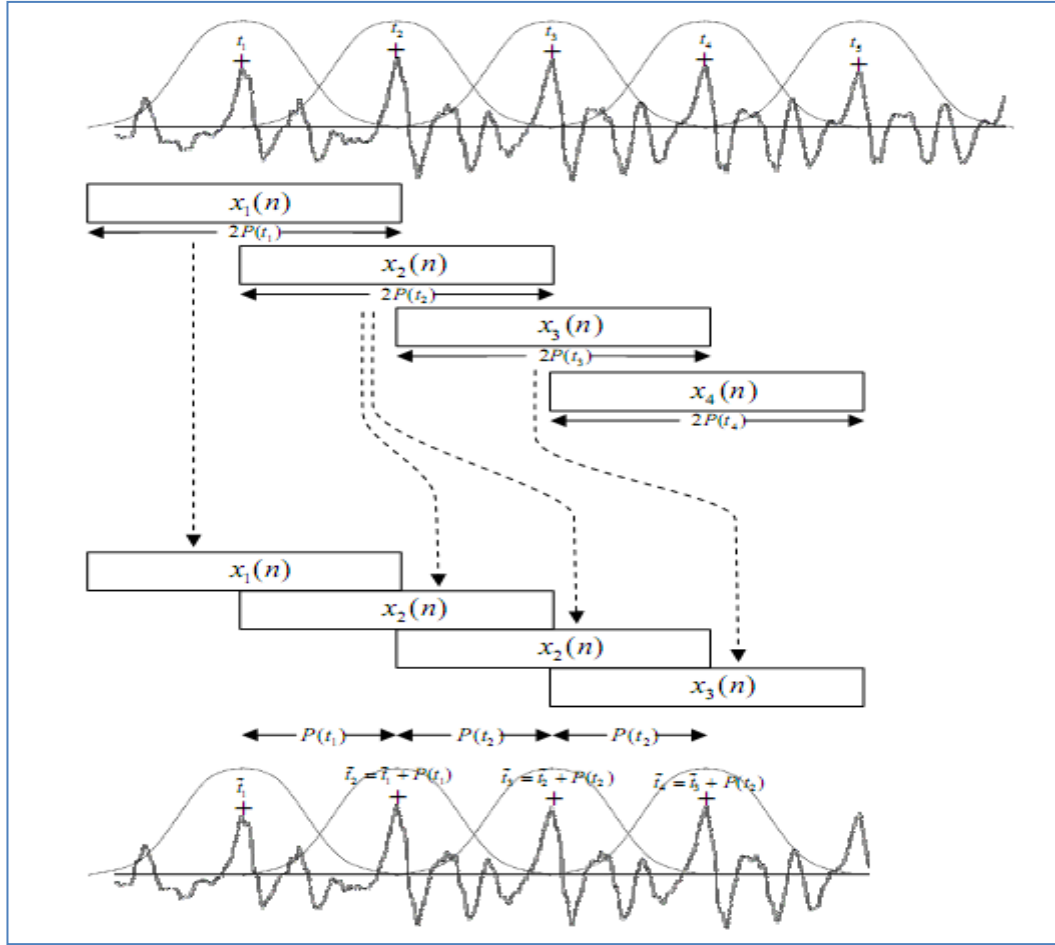
$$\tilde{P}(\tilde{t}) = \tilde{P}(\alpha t) = P(t) \quad (2.20)$$

Sentez perde zamanları önceki perde periyotlarından Eşitlik 2.21'e göre hesaplanır.

$$\tilde{t}_{k+1} = \tilde{t}_k + \tilde{P}(t_k) = \tilde{t}_k + P(t_i) \quad (2.21)$$

Eşitlik 2.21'de yer alan $\tilde{t}_k$ sentez sesinde yer alan k . perde işaretini t_i ise analiz sesinde yer alan i . perde işaretini simgelemektedir. Yani senteze eklenecek k . parça analizdeki i . parçanın örtüşmesiyle elde edilmektedir. Burada t_i , $|\alpha t_i - \tilde{t}_k|$ yı en küçük yapacak şekilde seçilir.

Eşitlik 2.20 - 2.21 hızlandırma/yavaşlatma faktörü α , sabittir. Fakat bu $\tilde{t} = \int_0^t \alpha(\tau) d\tau_k$ ile sabit olmayan hızlandırma/yavaşlatma faktörü α için de genellenebilir. Şekil 2.4'ten de görülebildiği gibi Eşitlik 2.21'e göre $\alpha > 1$ için bazı analiz parçaları birden fazla örtüşmeli eklemede kullanılabilir. Yine aynı şekilde $\alpha < 1$ için bazı analiz parçaları kullanılmayabilir.



Şekil 2.4 PSOLA algoritmasında 2 perdelik segmentler ve bu segmentler ile örtüşmeli ekleme (Bonada 2002)

2.1.6 WSOLA

WSOLA (Dalgabiçimi benzerlikli örtüşmeli ekleme), analiz sesi (giriş sesi) $x(n)$ ile zaman skalalı sentez sesi $y(n)$ arasında $n = \tau(m)$ için en çok yerel benzerliği bulur. Bu ifade Eşitlik 2.22 ile gösterilir;

$$\forall m: y(n + \tau(m)) \cdot w(n) (=) x(n + m) \cdot w(n) \quad (2.22)$$

Eşitlik 2.22’de yer alan $w(n)$ ağırlık penceresi, $(=)$ işareti en çok benzerlik anlamında ve $\tau(m)$ ise zaman skalasını değiştiren zaman katlama fonksiyonudur. Eşitlik 2.22’yi sağlamak için WSOLA algoritması katlanacak çerçevelerin en çok benzerliği sağlayacak şekilde $\Delta_{\min} \leq \Delta \leq \Delta_{\max}$ komşuluğunda Δ kadar kaymasına izin verir. Buna

göre WSOLA zaman skalası Eşitlik 2.23’de ifade edilmektedir (Verhelst ve Roelands 1993, Verhelst 2000).

$$y(n) = \frac{\sum_k w^2(n - S_k) \cdot x(n + \tau^{-1}(S_k) - S_k + \Delta_k)}{\sum_k w^2(n - S_k)} \quad (2.23)$$

Bu formülde yer alan Δ_k k segmenti için en çok benzerliği sağlayan sapmadır. S sentez segmentinin uzunluğu olup; S_k ise sabit S için $k.S$ ’e sabit olmayan S içinse $\sum_{i=0}^k S_i$ toplamına eşittir. Eşitlik 2.23’de hesaplama kolaylığı için $\sum_k w^2(n - S_k) = 1$ seçilir ve paydadan kurtulunur. Sabit S için $S_k = k.S$ ettiğinden β skala faktörü için Eşitlik 2.24 yazılabilir.

$$\tau^{-1}(S_k) = k.S / \beta \quad (2.24)$$

WSOLA’daki pencere fonksiyonu $w(n)$ ’in pencere uzunluğu sentezlenen sinyalin her bir segmenti için katlama ile kullanılacak olan analiz segmenti sayısına bağlıdır. Verhelst %50 katlamanın başarılı sonuç verdiğini söylemektedir ki o zaman pencere uzunluğu $2S$ olacaktır. %50 katlamada $y(n)$, $(k-1)S < n \leq k.S$ için Eşitlik 2.25 ile ifade edilir.

$$y(n) = w^2(n - (k-1)S) \cdot x(n + (k-1)S / \beta - (k-1)S + \Delta_{k-1}) + w^2(n - kS) \cdot x(n + kS / \beta - kS + \Delta_k) \quad (2.25)$$

Bu eşitlikte Δ_k $k=0$ için sıfır alınır. Diğer Δ_k ’lar ise İlinti, AMDF veya normalleştirilmiş ilintiye göre bulunabilir (Verhelst ve Roelands 1993, Verhelst 2000). SOLA’da kullanılan diğer benzerlik indisi hesaplama yöntemleri de Δ_k için kullanılabilir.

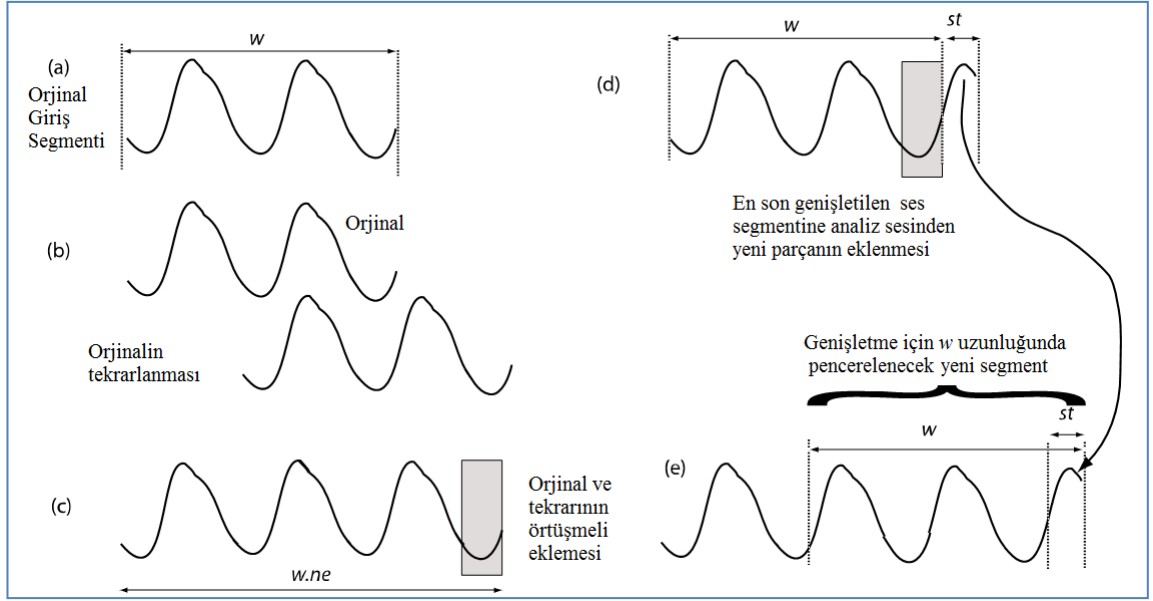
2.1.7 AOLA

AOLA (Uyarlanabilir Örtüşmeli Ekleme) algoritmasında giriş (analiz) sesine uzunluğu w olan bir pencere uygulanarak bir ses segmenti elde edilmektedir (Lawlor 1999). Şekil 2.5.a'da da görüldüğü gibi pencere (dolayısıyla ses segmenti) uzunluğu ötümlü sesler için en az 2 tane en düşük frekans (temel frekans veya perde) periyodu içerecek şekilde seçilmektedir. Ötümsüz segmentler için ise bir önceki ötümlü seste kullanılan segment uzunluğu ile aynı seçilebilir (Lawlor 1999). Daha sonra bu segmentin aynısı kopyalanarak, şekil 2.5.b'de görüldüğü gibi tepe noktaları birbirleriyle denk gelecek şekilde sağa doğru kaydırılır. Daha sonra bu iki segmentin örtüşen noktaları için ağırlıklı ekleme (örtüşmeli ekleme) yapılarak şekil 2.5.c'de görüldüğü gibi “doğal genişletilmiş” (yavaşlatılmış) bir sinyal parçası oluşturulur. Bu oluşturulan genişletilmiş ses parçasının uzunluğu doğal genişletme oranı ne için $w.ne$ 'dir.

Daha sonra analiz sesinden st uzunluğunda (daha önce alınan w uzunluğundaki segmentten sonraki örnekler) bir parça şekil 2.5.d'de görüldüğü gibi alınır. Yeni alınan bu parçanın uzunluğu w , ne , de (istenen genişletme oranı) parametrelerine bağlıdır.

Analiz sesinden alınan bu yeni st parçası sentezlenen sese eklenir (şekil 2.5.e). Sentezlenen seste şekil 2.5.e'de görüldüğü gibi st parçasını da içerecek yeni bir w uzunluğunda segment alınır ve bu yeni segmentin kopyası daha önceki işlem gibi kaydırılarak tepe noktaları örtüştürülecek şekilde örtüşmeli ekleme yapılır. Yeni segmentin bir kısmı eski sentezden gelmekte yalnızca st parçası analiz (orjinal giriş) sesinden gelmektedir. Bu segmentin genişlemesi de ne oranında olacağı için bu yeni segmentte yer alan st parçasının genişlemesi de $ne.st$ olacaktır.

Analiz sesinden tekrar bir st parçası alınıp sentez sesine eklenirse, yeni oluşturulacak w uzunluğundaki segmentte bu yeni alınan st parçası ve bir önceki tekrarda alınmış ve uzatılmış $ne.st$ parçası olacaktır. Bu segment de uzatıldığında sentezde st uzunluğu $ne.st$ ve bir önceki tekrardan gelen $ne.st$ uzunluğu da $ne^2.st$ uzunluğuna sahip olacaktır.

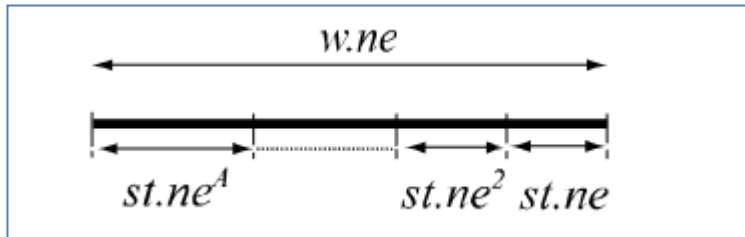


Şekil 2.5 AOLA ile zamanda skalası değişimi (Lawlor, 1999)

Bu tekrarlar $A+1$ kere sürerse sentez sesinin sonundaki $w.ne$ uzunluğundaki parça şekil 2.6'da da gösterildiği gibi $ne^A.st$ 'yi de içerecektir.

Şekil 2.6'da görüldüğü gibi $w.ne$ Eşitlik 2.26'daki gibi açılabilir.

$$w.ne = st.ne + st.ne^2 + \dots + st.ne^A \quad (2.26)$$



Şekil 2.6 AOLA yavaşlatması (Lawlor, 1999)

Eşitlik 2.26'da $ne^A = de$ seçilirse Taylor serisi açılımından Eşitlik 2.27'de ifade edilen değere ulaşılır.

$$st = w.(1 - ne)/(1 - ne^A) = w.(1 - ne)/(1 - de) \quad (2.27)$$

Eşitlik 2.27'de yer alan de istenen genişleme faktörüdür.

Doğal genişleme faktörü ne segment içinde yer alan en düşük frekansa bağlı olacağından her tekrarda değişebilirler. Bu nedenle Eşitlik 2.26 - 2.27 aslında bir yakınsama eşitliğidir. ne ve st de her genişleme için baştan hesaplanmalıdır.

Zaman skalası sıkıştırması (hızlandırma) işleminde de yavaşlatmada olduğu gibi analiz sesinden w uzunluğunda bir segment alınır. Aynı şekilde bu segmentin bir kopyası kaydırılarak örtüşmeli ekleme yapılır. Fakat örtüşmeli eklenen örneklemelerin sağında kalan ve örtüşmeyen örneklemeler doğal hızlandırılan segmenti bırakmak için bırakılır. Doğal hızlandırma oranı nc ve istenen hızlandırma oranı dc ise Eşitlik 2.28 ile ifade edilir.

$$st = w.(1 - nc)/(1 - dc) \quad (2.28)$$

2.1.8 Suzuki-Misaki yöntemi

Suzuki ve Misaki (1992) zaman skalası değişimi ile yavaşlatma için analiz sesinden ardışık iki adet T_s uzunluğunda çerçeve almaktadır. Daha sonra bu iki ardışık çerçevenin $S_- \leq \tau \leq S_+$ için Eşitlik 2.29'daki gibi ilintileri hesaplanmaktadır.

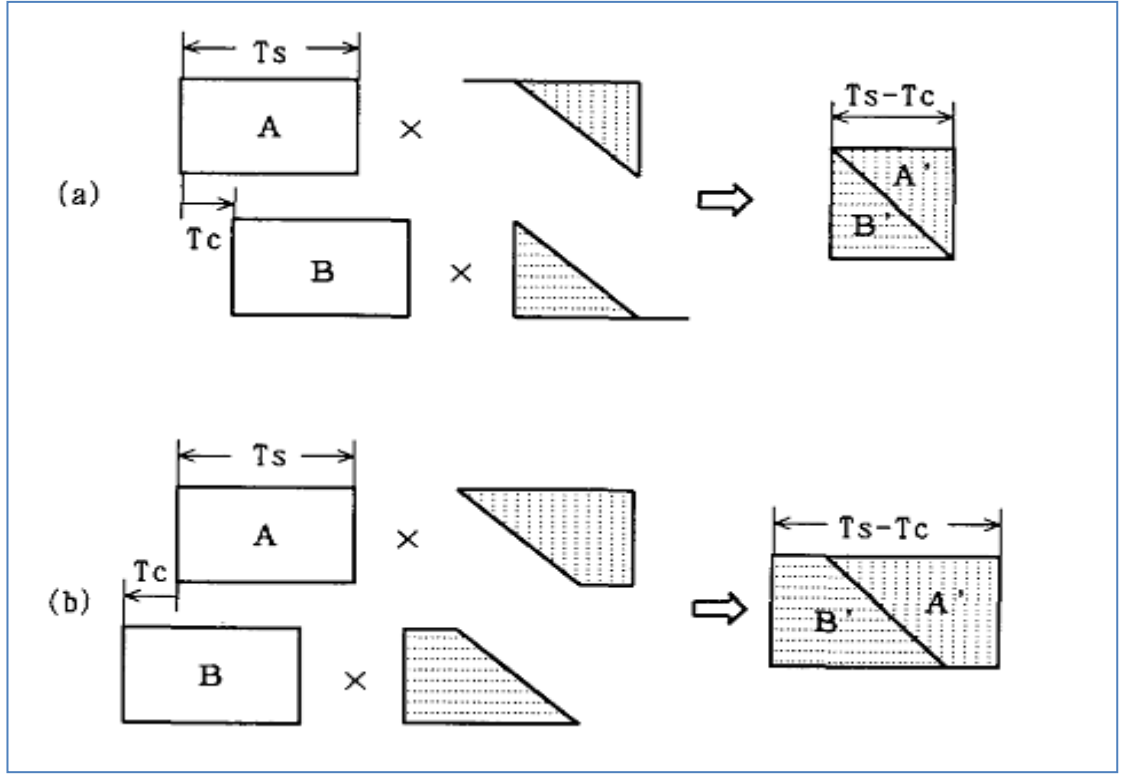
$$R(\tau) = \sum_{m=0}^{T_s-1} x(i + m + \tau)x(j + m) \quad (2.29)$$

Eşitlik 2.29'da yer alan i ilk çerçevenin başlangıcı, j ise ikinci çerçevenin başlangıcıdır.

$$T_c = \max_{\tau} R(\tau) \quad (2.30)$$

İlintisi en yüksek çıkan τ değeri T_c için iki çerçeve arasında örtüşmeli ekleme yapılır (Eşitlik 2.30). Böylece sentezlenen sesin uzunluğu $T_s - T_c$ olur. Bu durum $T_c > 0$ ve $T_c < 0$ için Şekil 2.7'de gösterilmektedir.

Şekil 2.7'den de anlaşılabileceği gibi Eğer $T_c > 0$ için sentezlenen ses Eşitlik 2.31.a ile ifade edilir. $T_c < 0$ için sentezlenen ses ise Eşitlik 2.32.b ile ifade edilir.



Şekil 2.7 Zaman skalası değişimi ile yavaşlatma (Suziki ve Misaki 1992)

$$y(k) = w(k)x(i + k + T_c) + (1 - w(k))x(j + k) \quad 0 \leq k \leq T_s - T_c - 1 \quad (2.31.a)$$

$$y(k) = \begin{cases} x(j + k) & 0 \leq k \leq -T_c - 1 \\ w(k)x(i + k + T_c) + (1 - w(k))x(j + k) & -T_c \leq k \leq T_s - 1 \\ x(i + k + T_c) & T_s \leq k \leq T_s - T_c - 1 \end{cases} \quad (2.31.b)$$

Eşitlik 2.31.a için ağırlık penceresi $w(k)$ değerleri Eşitlik 2.32.a'da, Eşitlik 2.31.b için ağırlık penceresi $w(k)$ değerleri ise Eşitlik 2.32.b'de yer almaktadır.

$$w(k) = (k + 1)/(T_s - T_c + 1) \quad (2.32.a)$$

$$w(k) = (k + T_c + 1)/(T_s - T_c + 1) \quad (2.32.b)$$

Eşitlik 2.30 veya 2.31 ile sentez sesi oluşturulduktan sonra analiz sesinden alınan ilk çerçeveden sonraki örneklerden $T_x = T_s / (\alpha - 1) + T_c$ kadar bir sinyal alınır ve sentez sinyalinin sonuna eklenir. Böylece önceki sentezlenen kısım $T_s - T_c$ olduğundan toplam sentez uzunluğu $T_s - T_c + T_x = \alpha \cdot T_s$ olur. Zaman skalası faktörü α büyük olduğunda T_s küçük

olacaktır. Daha sonra genişletme işlemine devam etmek için ardışık segmentlerin indisleri Eşitlik 2.33 - 2.34'deki gibi güncellenir.

$$i \leftarrow i + T_s / (\alpha - 1) \quad (2.33)$$

$$j \leftarrow i + \alpha T_s / (\alpha - 1) + T_c \quad (2.34)$$

Suziki ve Misaki zaman skalasını hızlandırmak için yavaşlatmada kullanılan aynı yöntemi kullanmaktadır. Fakat fark olarak ağırlık penceresi yavaşlatmadaki gibi monoton azalan değil monoton artan bir fonksiyon olarak seçilmektedir. İlinti hesaplamasında ilk çerçeve sabit tutulup 2. çerçeve kaydırıldığından, sinyal ekleme ve indislerin güncellenmesi formülleri değişmektedir. Buna göre sentez sinyali en çok benzerliği sağlayan τ için Eşitlik 2.31 ile benzer şekilde olacaktır. Fakat ağırlık penceresi değerleri Eşitlik 2.35 - 2.36'deki gibi olacaktır.

$$w(k) = (T_s + T_c - k) / T_s + T_c + 1 \quad (2.35)$$

$$w(k) = (T_s - k) / T_s - T_c + 1 \quad (2.36)$$

İlk çerçevelerden sentez sesinin ilk kısmı oluşturulduktan sonra $T_x = (2\alpha - 1)T_s / (1 - \alpha) - T_c$ uzunluğundaki parça giriş sinyalinden alınır ve sentez sinyalinin sonuna eklenir. Daha sonra genişletme işlemine devam etmek için ardışık segmentlerin indisleri Eşitlik 2.37 - 2.38'deki gibi güncellenir.

$$i \leftarrow j + \alpha T_s / (1 - \alpha) - T_c \quad (2.37)$$

$$j \leftarrow j + T_s / (1 - \alpha) \quad (2.38)$$

Suziki ve Misaki (1993), yaptıkları çalışmada T_s uzunluğunu 12 ms S_- ve S_+ değerlerini ise -10 ve +10 ms seçmişlerdir.

2.1.9 Laroche yöntemi

Laroche (1993) AOLA'da olduğu gibi analiz sesinden aldığı bir çerçeveyi kendisiyle örtüştürüyor. Bu yöntemde analiz sesinden uzunluğu $T_a=40$ ms olan bir segment

alınmaktadır. Daha sonra bu segmentin $\tau_{\min} \leq \tau \leq \tau_{\max}$ aralığında Eşitlik 2.39 ile özilintisine bakılır.

$$R(\tau, t_1) = \frac{1}{T_a - \tau} \sum_{k=1}^{T_a - \tau} x(t_1 + k)x(t_1 + k + \tau) \quad (2.39)$$

Öz ilintisinin maksimumunu veren τ değeri için kendisiyle örtüşen kısımlarda örtüşmeli ekleme yapılır. Zaman skalasında genişleme için $\tau_{\min}$ ve $\tau_{\max}$ negatif; sıkıştırma için ise pozitif seçilmektedir. $\tau_{\min}$ ve $\tau_{\max}$ en az 10 ms en çok 25 ms uzunluğunda bir segmentin tekrarlamasını/silinmesini sağlayacak büyüklükte seçilir.

Laroche(1993), analiz sesini uzunluğu $T_a=40$ ms olan ve birbirleriyle çakışmayan segmentlere ayırmaktadır. Bu segmentlerin her biri daha sonra ya sentez sesine eklenmekte veya yavaşlatıldıktan (veya hızlandırıldıktan) sonra sentez sesine eklenmektedir. Bir segmentin hızlandırılacağı/yavaşlatılacağı çıktının beklenen uzunluğunun belirli bir toleransın altında veya üstünde olmasına göre karar verilmektedir. Laroche çalışmasında bu toleransı da 40 ms olarak kullanmıştır.

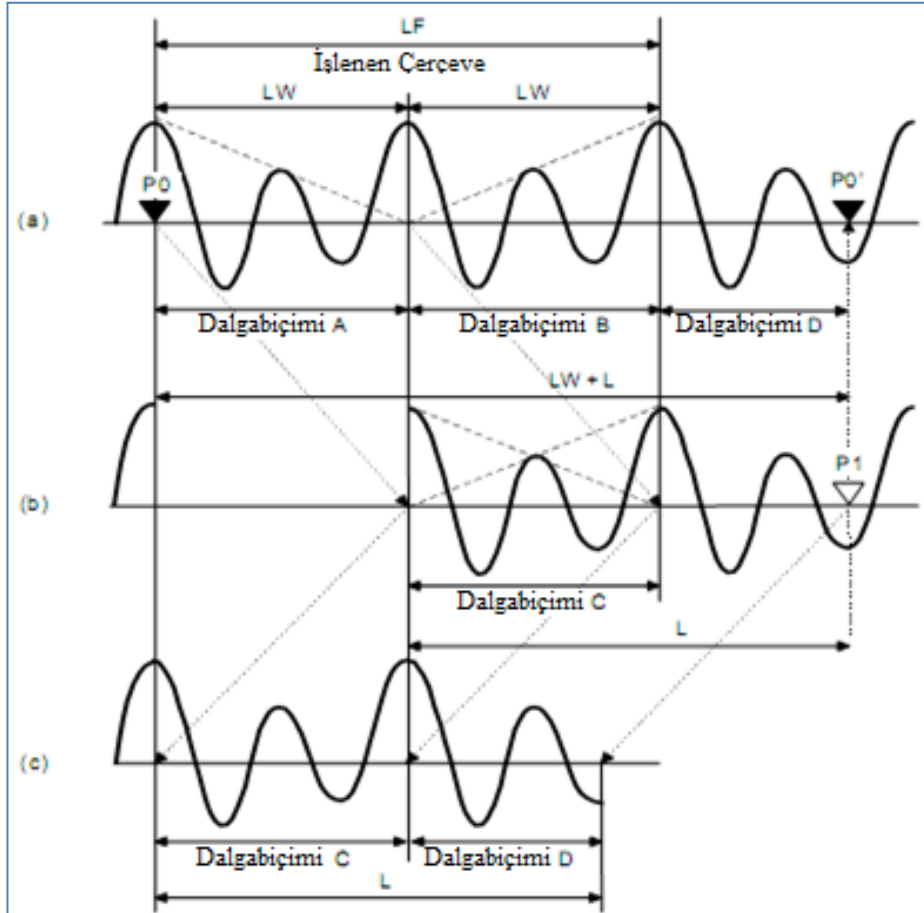
2.1.10 PICOLA

PICOLA (İşaretlenmiş aralık kontrollü örtüşmeli ekleme) ISO/IEC 14496-3'de (MPEG-4 audio) önerilen zaman skalası değişimi yöntemidir. PICOLA ile hız kontrolü MPEG-4'de zorunlu değil kullanılabilir bir özelliktir. PICOLA 8 KHz veya 16 KHz'de örneklenmiş mono seslerde kullanılabilir. Hız kontrolü giriş sinyali örtüşmeli eklemeli sentez sinyali ile değiştirilerek veya örtüşmeli eklemeli dalgabıçımı ekleyerek sağlanır. MPEG-4'de PICOLA için zaman skalası değişimi oranı ve zaman skalası değişiminin olup olmadığını gösteren bayraklar (flag) kullanılır. İletilecek sesin çerçevelerinde zaman skalası uygulanıp uygulanmadığı uygulandıysa oranı bu şekilde belirtilir.

PICOLA ile hızlandırma şekil 2.8'de gösterilmektedir. P_0 işaretleyicisi (pointer) mevcut işlenen çerçevenin başlangıç örneğini göstermektedir. Çerçeve uzunluğu LF örnek boyutundadır ve hemen ardındaki çerçeveler LW uzunluğundadır (Morita ve Itakura 1986). LF uzunluğundaki çerçevenin ilk yarısı (*Dalgabıçımı A*) ile ikinci yarısı

(Dalgabiçimi B) arasındaki örnek başına ortalama bozulma Eşitlik 2.40 ile hesaplanmaktadır.

$$D(LW) = \frac{1}{LW} \sum (x(n) - y(n))^2 \quad P_{MIN} \leq LW \leq P_{MAX} \quad (2.40)$$



Şekil 2.8 PICOLA ile hızlandırma (ISO/IEC 14496-3)

a. Orijinal sinyal, b. Örtüşmeli eklenmiş sinyal, c. Hızlandırılan sentez sinyali

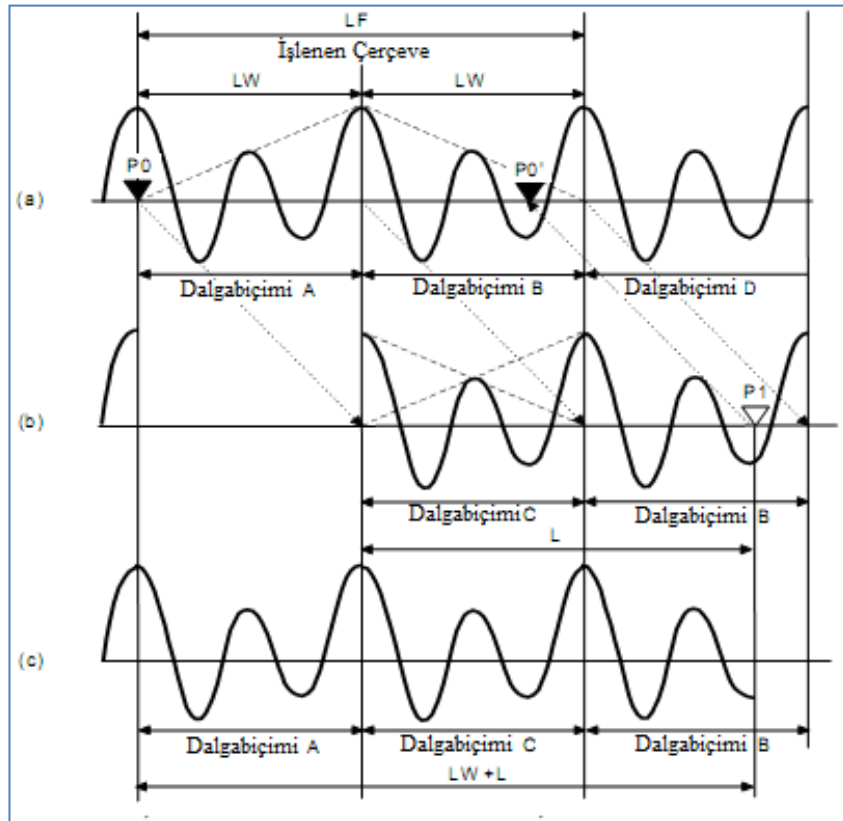
Eşitlik 2.40'da $D(LW)$ dalgabiçimi uzunluğu LW olduğunda ardışık iki dalgabiçimi arasında örnek başına ortalama bozulmayı, $x(n)$ Dalgabiçimi A'yı; $y(n)$ Dalgabiçimi B'yi P_{MIN} ve P_{MAX} 'da LW parçası için en az ve en çok dalgaboyu uzunluğunu göstermektedir. Tipik olarak 8 KHz ile örneklenmiş ses için $P_{MIN} = 32$ $P_{MAX} = 160$; 16 KHz ile örneklenmiş ses için $P_{MIN} = 80$ $P_{MAX} = 320$ örneklem seçilir (Morita ve Itakura 1986).

PICOLA’da LW uzunluğu $D(LW)$ bozulmayı en düşük seçecek şekilde seçilir(Morita ve Itakura 1986). Daha sonra katlama ağırlıkları *Dalgabiçimi A* için doğrusal azalan (Barlet’in ikinci yarısı) *Dalgabiçimi B* için doğrusal artan ve toplamaları 1 olacak şekilde seçilerek örtüşmeli ekleme yapılır ve şekil 2.8.b’deki gibi *Dalgabiçimi C* oluşturulur. Daha sonra işaretleyici P_0 ’dan P_1 noktasına kaydırılır. *Dalgabiçimi C*’nin başlangıcı ile P_1 noktası arasındaki uzaklık L ’dir ve uzunluğu Eşitlik 2.41’de verilmektedir.

$$L = LW \frac{1}{HIZORANI - 1} \quad 1 < HIZORANI \leq 2 \quad (2.41)$$

L , LW ’den büyük ise analiz sesinde *Dalgabiçimi B*’yi takip eden *Dalgabiçimi D* sentez sesine eklenir. Böylece analiz sesinden alınan $LW+L$ uzunluğundaki bir parça için L uzunluğunda daha küçük bir parça sentezlenmiş olur.

PICOLA ile yavaşlatma şekil 2.9’da gösterilmektedir.



Şekil 2.9 PICOLA ile yavaşlatma (ISO/IEC 14496-3)

a. Orijinal sinyal, b. Örtüşmeli eklenmiş sinyal, c. Yavaşlatılan sentez sinyali

Şekil 2.9.a'da P_0 işaretleyicisi (pointer) mevcut işlenen çerçevenin başlangıç örneğini göstermektedir. Çerçeve uzunluğu LF örnek boyutundadır ve hemen ardındaki çerçeveler LW uzunluğundadır. LW uzunluğu Eşitlik 2.40'a göre belirlendikten sonra *Dalgabiçimi A* aynen sentez sesine eklenir ve daha sonra *Dalgabiçimi A* ile *Dalgabiçimi B* hızlandırma kısmında olduğu gibi örtüşmeli ekleme yapılarak sentez sesinin sonuna eklenir. Daha sonra işaretleyici şekil 2.9.b'de gösterildiği gibi P_0 'dan P_1 noktasına kaydırılır. P_1 noktası *Dalgabiçimi C*'nin başlangıcından L örneklem uzaklıktadır. Böylece L uzunluğundaki bir parçadan $L+LW$ uzunluğunda bir parça sentezlenmiş olmaktadır (Morita ve Itakura 1986).

L uzunluğu Eşitlik 2.42'ye göre seçilmektedir.

$$L = LW \frac{HIZORANI}{1 - HIZORANI} \quad 0,5 \leq HIZORANI < 1 \quad (2.42)$$

2.2 Zaman Skalası Değişimi İçin Frekans Alanı Teknikleri

Frekans alanı tekniklerinin temeli Homer Dudley'in (1939) kanal vokoderine dayanmaktadır. Kanal vokoderinde ses sinyali analiz edilerek temel frekansı tespit edilir ve spektral bilgileri bir filtre ile kodlanır. Kodlanmış olan bu ses sinyali tekrar sentezlenirken temel frekansın enerjisi kullanılarak ötümlü/ötümsüz ayrımı yapılır ve belirlenen filtre ötümlü sesler için temel frekanstaki periyodik sinyal ile ötümsüz sesler için ise rastgele gürültü ile sürülür. Böylece ses sinyali tekrar oluşturulmuş olur.

Flanagan ve Golden (1966) bu modeli ses sinyalinin kısa zaman genlik ve faz spektrumunu alarak geliştirmiştir. Faz vokoder olarak alınan bu teknik frekans alanı zaman skalası değişimi tekniklerinin temeli olmuştur.

Faz vokoderleri Kısa-Zaman Fourier Dönüşümüne (STFT) dayanmaktadır. Bu nedenle STFT de zaman skalası değişimi için kullanılabilir. Faz vokoderleri tanımlandıktan sonra zaman skalası değişimi kalitesini arttırmak için Faz vokoderleri üzerine çeşitli geliştirmeler yapılmıştır. Geliştirilen bu teknikler faz vokoderlerinin

sahip olduğu yatay (zamansal) faz uyumluluğuna ek olarak dikey (frekans bölgeleri arası) faz uyumluluğunu da (eş fazlılık) sağlamaktadır.

2.2.1 Kısa Zaman Fourier Dönüşümü (STFT)

Ses sinyali kısa süre içerisinde hemen hemen periyodik bir yapıdadır. Bununla birlikte bu süre yeterince uzadığında bu periyodiklik azalmakta ve tükenmektedir. Ses ve benzeri sinyallerin kısa süreli frekans spektrumunu almak için sinyal kısa zaman periyodunu belirleyen pencere fonksiyonu ile çarpılarak sesin sinyalin sadece belirli bir zaman aralığındaki değerleri bırakılır. Daha sonra elde edilen sinyale Fourier dönüşüme uygulanarak sinyalin STFT'si alınmış olur. Herhangi bir sinyalin STFT'si Eşitlik 2.43 ile ifade edilir.

$$X(mS, \omega) = \sum_{l=-\infty}^{\infty} w(mS-l)x(l).e^{-j\omega l} \quad (2.43)$$

Bu eşitlikte m , Fourier dönüşümü yapılan m . çerçeve; S , ardışık çerçeveler arasındaki uzaklık; $\omega = 2\pi k / N$, $k=0,1,2,\dots,N-1$ olmak üzere frekans değeri, N pencere uzunluğu ve $w(n)$, pencere fonksiyonudur. $X(mS, \omega)$ karmaşık sayı olup Eşitlik 2.44'te ifade edildiği gibi faz ve genlik değerlerine sahiptir.

$$X(mS, \omega) = |X(mS, \omega)| \angle X(mS, \omega) \quad (2.44)$$

Eşitlik 2.43'te yer alan $w(mS-l)x(l)$ $x(n)$ sinyalinin m . sinyal çerçevesi (şekil 2.10.c'de gösterilmektedir) olup bu sinyalin Fourier dönüşümü Eşitlik 2.45 ifadesinden dolayı Eşitlik 2.46 ile gösterilir.

$$x_w(mS, l) = w(mS-l)x(l) \quad (2.45)$$

$$X(mS, \omega) = \sum_{l=-\infty}^{\infty} x_w(mS, l).e^{-j\omega l} \quad (2.46)$$

Eşitlik 2.46’da gösterildiği gibi herhangi bir $x(n)$ sinyalinin pencerelemiş parçasının STFT’si ardışık pencerelemiş segmentlerin Fourier dönüşümü olur.

Pencere büyüklükleri $2S$ seçildiğinde pencereler %50 örtüşme sağlamış olurlar. Örtüşen segmentler için pencerelerin zaman alanındaki örtüşen değerleri toplamı 1 olacak şekilde seçilirse $x(n)$ sinyalinin Fourier dönüşümü tüm çerçevelerin STFT’lerinin toplamı olarak ele alınabilir.

2.2.1.1 Tekrarlı STFT

Griffin ve Lim (1984) sesin haberleşme kanalından iletiminde genlik değerleri yerine STFT parametrelerini kullanmayı önermişlerdir. STFT parametreleri genlik ve faz spektrumu değerlerinden oluştuğundan hem genlik hem faz parametreleri iletmek veri miktarını arttırdığından sadece genlik değerlerini iletmeyi önermişlerdir.

Sadece STFT genlik spektrumu parametreleri kullanılarak yapılan sentezleme ardışık çerçevelerde faz kaymasına neden olmakta bu da sesin bozulmasına neden olmaktadır. Bu nedenle faz bilgisinin karşı tarafta tekrar elde edilmesi gerekmektedir.

$$x(n) \xrightarrow{STFT} X(mS, \omega) \quad (2.47)$$

Eşitlik 2.47’de ifade edildiği gibi $x(n)$ sinyalinin STFT’si alınmış olsun. $X(mS, \omega)$ komşuluğunda herhangi bir STFT gösterimli sinyalden $x(n)$ sinyali belirli bir hata toleransı ile elde edilebilmektedir.

$$Y(mS, \omega) \xrightarrow{?} x(n) \quad (2.48)$$

$Y(mS, \omega)$ ile $X(mS, \omega)$ hata kare yakınsamasında benzerdir (Griffin ve Lim 1984).

$$D[x(n), Y(mS, \omega)] = \sum_{m=-\infty}^{\infty} \sum_{l=-\infty}^{\infty} [x_w(mS, l) - y_w(mS, l)]^2 \quad (2.49)$$

Eşitlik 2.49'da $x_w(mS, l) = w(mS - l)x(l)$ ve $y_w(mS, l) = w(mS - l)y(l)$ olup, hatayı en aza indirmek için bu ifadenin $x(n)$ 'e göre gradyanı alınarak sıfıra eşitlenir. Bu gradyan ifadesinden $x(n)$ çekilirse Eşitlik 2.50 elde edilir.

$$x(n) \cong \frac{\sum_{m=-\infty}^{\infty} w(mS - n)y_w(mS, n)}{\sum_{m=-\infty}^{\infty} w^2(mS - n)} \quad (2.50)$$

Yukarıda da belirtildiği gibi alıcı tarafa sadece genlik bilgileri gönderildiğinden genlik spektrumu $|X(mS, \omega)|$ biliniyor, faz spektrumu $\angle X(mS, \omega)$ bilinmiyor. $x(n)$ sinyalini tekrar elde etmek için aşağıdaki algoritma kullanılır.

- Bilinmeyen faz sapması için ilk faz tahmini yapılır (gerekirse faz kayması sıfır alınır) ve $|X(mS, \omega)|$ ile $y_w(mS, n) = \text{IFFT}\{ |X(mS, \omega)| \angle X^0(mS, \omega) \}$ hesaplanır.
- Eşitlik 2.50 kullanılarak $x(n)$ tahmini $x^1(n)$ bulunur.
- Hesaplanan $x^1(n)$ için STFT uygulanarak $X^2(mS, \omega) = |X(mS, \omega)| \angle X^1(mS, \omega)$ bulunur.
- Hesaplanan bu yeni STFT'nin faz bilgileri yeni faz tahmini olarak alınır ve ilk 3 adım tekrarlanarak $X^{k+1}(mS, \omega) = |X(mS, \omega)| \angle X^k(mS, \omega)$ olarak bulunur.

25 ile 100 adımda fazda çok fazla değişim olmayacağından hemen hemen doğru tahmin edilerek $x^{k+1}(n)$ ile ses sinyali kabul edilebilir kalitede bulunur.

2.2.1.2 Tekrarlı STFT ile zaman skalası değişimi

Griffin ve Lim (1984), STFT'nin zaman skalası değişimi içinde kullanılabileceğini belirtmişler. STFT algoritmasına göre Eşitlik 2.43 her bir ses çerçevesi için spektrum

bilgisini vermektedir. Spektrum bilgisinden sinyalin zamandaki değerlerini elde etmek için Ters STFT uygulanır ve N adet örneklem değeri bulunur. Ters STFT uygulanırken STFT frekans değerlerinin sayısı interpolasyon ve eksiltme ile değiştirilip, ters STFT bu yeni STFT adetlerine göre yapılırsa zaman skalası değişimi gerçekleştirilmiş olur.

Zaman skalası değişiminin analiz aşamasında STFT alınırken ardışık pencereler arası uzaklık için $S = S_a$ seçilir. Sentez için ise $S = S_s$ seçilir. Bu sayede $S_s < S_a$ için zamanda sıkıştırma, tersi için zamanda genleştirme uygulanmış olur.

Ters STFT ile elde edilen sinyal çerçevelerinin arasındaki uzaklık değiştiği için bu çerçevelerin sınırlarında faz kayması oluşmaktadır. Faz uyumluluğunu tekrar sağlamak için tekrarlı STFT yöntemi kullanılır. Yeterince tekrardan sonra Eşitlik 2.51 ile sentezlenen zaman skalası değişmiş çerçeveler arası faz uyumu sağlanmış sinyal elde edilir.

$$x^{k+1}(n) = \frac{\sum_{m=-\infty}^{\infty} w(mS_s - n) y_w(|Y(mS_a, \omega)| \angle X^k(mS_s, \omega))}{\sum_{m=-\infty}^{\infty} w^2(mS_s - n)} \quad (2.51)$$

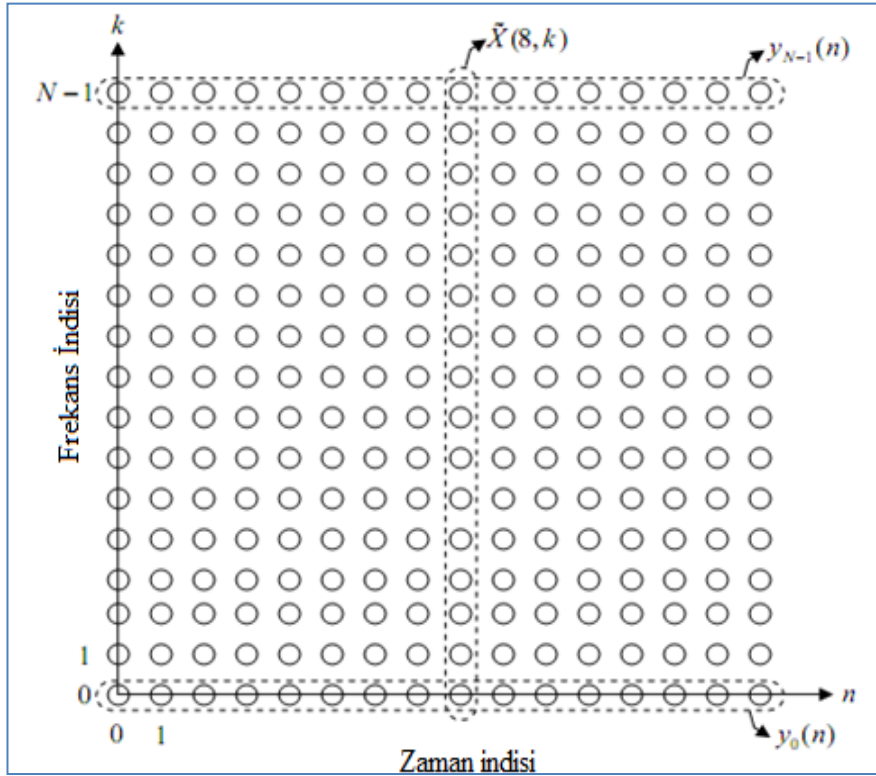
2.2.2 Faz vokoderleri

Faz vokoderleri Flanagan ve Golden (1966) tarafından konuşma sesini haberleşme kanalından iletmeden önce sıkıştırmak amacıyla geliştirilmiştir. Bu modelde konuşma sinyali; kısa zamanlı segmentlere bölünerek sinüs bileşenlerinin frekans ve genlik parametreleri ile tekrar yaratılabilir şekilde modellenmektedir. Flanagan ve Golden faz vokoderinin zaman skalası değişimi veya perde skalası değişimi için de kullanılabileceğini belirtilmişlerdir. Zaman skalası ve perde skalası değişimi Portnoff (1976) tarafından FFT kullanılarak gerçekleştirilmiştir.

Faz vokoderleri Bölüm 2.2.1’de anlatılan Kısa Zaman Fourier Dönüşümünü kullanmaktadır. Faz vokoderi matematiksel ifadeleri aynı olan iki ayrı bakış açısına göre yorumlanmaktadır (Dolson 1986).

1. Filtre bankaları yorumu
2. Fourier dönüşümü yorumu

Bu bakış açıları, şekil 2.10'da zaman frekans düzleminde yatay ve dikey çizgiler olarak görülebilir. Yatay eksene paralel tüm çizgiler sesin o frekans haznesinde zamana bağlı olarak değişimini (filtre bankası) dikey çizgiler ise herhangi bir kısa zaman aralığında sesin frekans spektrumunu göstermektedir.



Şekil 2.10 Zaman frekans düzlemi (Bonada 2002)

Faz vokoderlerinin filtre bankaları ile gösterimde giriş sinyalinin zamanla değişen spektrumu Eşitlik 2.52 - 2.53'te ifade edilen N bant-geçirgen paralel filtre çıktıları toplamıdır. Yani frekansları N frekans haznesinde toplanarak ayrılır. N arttıkça frekans hazneleri daralarak çözünürlüğü artar, zaman çözünürlüğü azalır.

$$h_k(n) = h(n)e^{-j\frac{2\pi kn}{N}}, \quad k = 0, 1, \dots, N-1 \quad (2.52)$$

$$H_k(e^{j\Omega}) = H_k(e^{j(\Omega - \Omega_k)}), \quad \Omega_k = \frac{2\pi k}{N} \quad (2.53)$$

Her bir filtrenin çıktısı (sinyalin frekans haznelerindeki değerleri) Eşitlik 2.54 ile ifade edilebilir.

$$\begin{aligned} y_k(n) &= \sum_{m=-\infty}^{\infty} x(m)h_k(n-m) = \sum_{m=-\infty}^{\infty} x(m)h(n-m)e^{j\frac{2\pi(n-m)k}{N}} \\ &= e^{j\frac{2\pi nk}{N}} \sum_{m=-\infty}^{\infty} x(m)e^{-j\frac{2\pi mk}{N}} h(n-m) = e^{j\frac{2\pi nk}{N}} X(n,k), \quad k = 0,1,\dots,N-1 \end{aligned} \quad (2.54)$$

Faz vokoderinin toplam çıktısı Eşitlik 2.54'te verilen her bir filtrenin çıktılarının toplamıdır (Eşitlik 2.55).

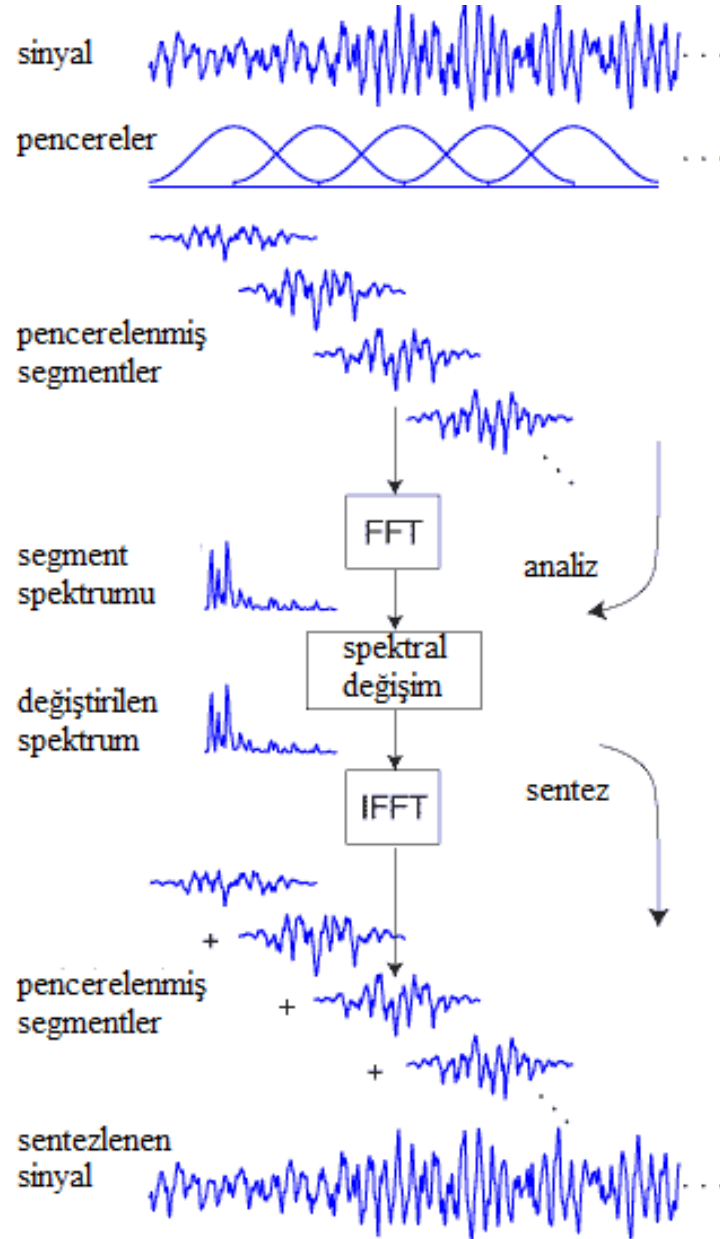
$$y(n) = \sum_{k=0}^{N-1} y_k(n) = \sum_{k=0}^{N-1} X(n,k)e^{j\frac{2\pi nk}{N}} \quad (2.55)$$

Faz vokoderlerinin Fourier Dönüşümü ile gösteriminde ise şekil 2.11'de gösterildiği gibi STFT'den faydalanılır. Analiz adımı Bölüm 2.2.1'de anlatıldığı gibi Eşitlik 2.43-ile gerçekleştirilir.

Sinyalin Fourier dönüşümü yapıldıktan sonra istenmesi durumunda spektral özelliklerde de değişime gidilebilir. Spektral özellikleri değiştirilirken her bir kısa dönem çerçevesi de değiştirilebilir. $t_s^s = sR_s$, R_s ardışık çerçeveler arası uzaklık olmak üzere spektrali değiştirilmiş sinyal $Y(t_s^s, \Omega_k)$ ile gösterilsin. Değiştirilen Fourier dönüşümünden sentez sinyali üretilirken sinyalin kısa zamanlı ters Fourier dönüşümü alınır. Elde edilen sinyalin zaman alanında gösterimi Eşitlik 2.56'da gösterilmektedir. Eşitlik 2.57 ise filtre bankası gösterimine göre her bir filtreye tekabül eden ve Eşitlik 2.55 ile de ifade edilen sentez sinyalidir.

$$y(n) = \sum_{n=-\infty}^{\infty} w(n-t_s^s)y_s(n-t_s^s) \quad (2.56)$$

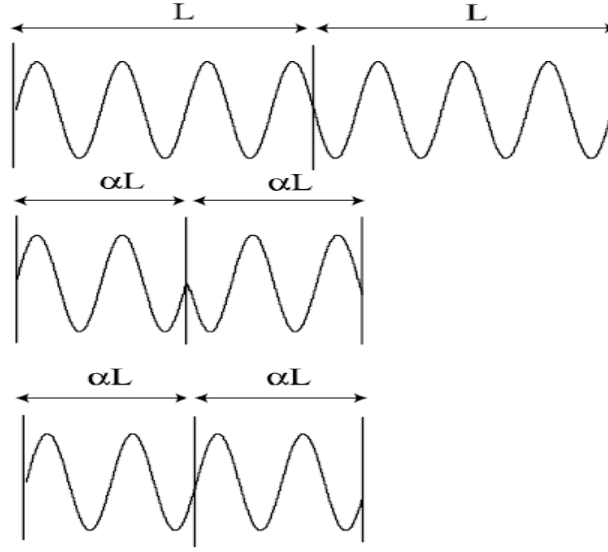
$$y_s(n) = \frac{1}{N} \sum_{k=0}^{N-1} Y(t_s^s, \Omega_k) e^{j\Omega_k n} \quad (2.57)$$



Şekil 2.11 Faz vokoderlerinin analiz ve sentezi (Sethares 2007)

2.2.2.1 Faz vokoderi ile zaman skalası değişimi

Zaman skalası değişimi sentez adımını analiz adımından farklı tutarak gerçekleştirilir. $R_s \neq R_a$ için, Sentez sinyalinin genlik spektrumu ve faz spektrumu analiz sinyali ile aynı tutulursa ardışık segmentler arasında faz kayması meydana gelir. Bu durum şekil 2.12’de gösterilmektedir.



Şekil 2.12 Faz vokoderlerinde analiz ve sentez

a. Analiz sinyalinin zamanda gösterimi, b. Aynı faz ve genlik için sentez sinyalindeki faz bozulması, c. Fazı düzeltilmiş sentez sinyali

Zaman skalası değişimi faz kaymasını düzeltmek için spektral değişikliği gerektirir. Zaman skalası değişiminde sentez sinyalinin genlik spektrumu Eşitlik 2.58’de ifade edildiği gibi analiz sinyalinin genlik spektrumu ile aynı seçilmektedir. Frekans spektrumu Eşitlik 2.59’da gösterildiği gibi güncellenir.

$$|Y(t_s^s, \Omega_k)| = |X(t_a^s, \Omega_k)| \quad (2.58)$$

$$\angle Y(t_s^s, \Omega_k) = \angle Y(t_s^{s-1}, \Omega_k) + R_s \hat{\omega}_k(t_a^s) \quad (2.59)$$

Eşitlik 2.59’da ifade edildiği gibi faz düzeltmesi için sentez sinyalinin bir önceki segmentteki faz spektrumu değerleri baz alınır. $\hat{\omega}_k(t_a^s)$ k . filtre haznesi (filtre gösterimindeki) için birim örneklem başına anlık faz değişimi olup iki ardışık segment arasındaki örneklem sayısı R_s olduğundan bu süre için güncelleme yapılır.

Anlık faz değişimi $\hat{\omega}_k(t_a^s)$ tespiti için iki ardışık analiz segmenti arasındaki faz değişimine bakmak gerekir. Eğer herhangi bir frekans haznesi içinde sadece tek bir frekans (sapma yok) olsaydı ardışık segmentler arasındaki faz değişimi $R_a \Omega$ olurdu.

Bununla birlikte frekans haznelerine düşen sinyal haznenin frekans aralığı içinde değişim gösterdiğinden sapma Eşitlik 2.60 ile bulunur.

$$\Delta\Phi_k^s = \angle X(t_a^s \Omega_k) - \angle X(t_a^{s-1} \Omega_k) - R_a \Omega_k \quad (2.60)$$

Eşitlik 2.60'dan birim örneklem başına düşen sapmayı bulmak için ardışık analiz segmentleri arasındaki birim örneklem başına düzen faz farkı Eşitlik 2.61'de yer almaktadır.

$$\hat{\omega}_k(t_a^s) = \Omega_k + \frac{1}{R_a} \Delta_p \Phi_k^s \quad (2.61)$$

Eşitlik 2.61'de yer alan $\Delta_p \Phi_k^s$ temel frekans sapması olup değer $-\pi$ ile $+\pi$ arasında yer alabilmesi için 2π katlarının eklenmesiyle tespit edilir.

2.2.3 Faz kilitli vokoderler

Faz vokoderi, STFT faz spektrumunda herbir frekans kanalı için zaman ekseninde yatay faz uyumluluğunu sağlamaktadır. Fakat faz vokoderlerinde farklı frekans hazneleri farklı faz katlama çözümlerine sahip olabilmektedir. Faz çözülümü algoritması ölçülen faz ile 2π katları toplamına eşittir. Fakat 2π katları tüm frekans hazneleri için eşit olmayabilmektedir. Bu da zaman germe oranı ile çarpıldığında fazlarda bozulmalara yol açabilmektedir.

Faz kilitli vokoderler frekans kanalları arasında (dikey eksen) faz uyumluluğunu sağlamak için geliştirilmiştir. Gevşek ve katı kilitlemeli olarak ayrılmaktadır. Katı kilitlemeli de kendi arasında Identity ve Skalarlı faz kitleme olarak ikiye ayrılmaktadır.

2.2.3.1 Gevşek kitleme

Faz vokoderi yatay faz uyumluluğunu sağlarken frekansları filtre bankasının filtrelerinin ortasına toparlamaktadır. Bu da bazen yankılara neden olabilmektedir.

Puckete (1995) tarafından önerilen gevşek kilitlemede faz vokoderlerinin faz yakınsama eşitliğinde (Eşitlik 2.59) yer alan $\angle Y(t_s^{s-1}, \Omega_k)$ yerine komşu frekans haznelerinin faz kayması etkisini de ifade etmek için $\angle Y(t_s^{s-1}, \Omega_k) - \angle Y(t_{s-1}^{s-1}, \Omega_k) - \angle Y(t_{s+1}^{s-1}, \Omega_k)$ ortalaması konur. Bu eşitlikte komşu haznelerin negatif olarak toplanmasının nedeni ardışık frekans hazneleri arasında yer alan 180 derecelik faz farkıdır.

2.2.3.2 Katı kilitleme

Gevşek kilitlemede faz kaymalarını göreceleme tüm frekans haznelerine uygulanmaktadır. Laroche ve Dolson (1999) tarafından önerilen katı kilitlemede ise sadece yerel tepelere göre göreceleme uygulanır. Herhangi bir frekans haznesinin genliği komşu 4 haznenin genliğinden daha yüksek ise o frekans haznesi bölgenin tepesi olarak alınır. Tepe frekans haznelerinin fazları hesaplanır ve çevresindekilerin sentez fazı bu tepe frekansın analiz fazına göre Eşitlik 2.62 ile ifade edildiği gibi hesaplanır.

$$\angle Y(t_s^s, \Omega_k) = \angle Y(t_s^s, \Omega_{kl}) + \angle X(t_a^s, \Omega_k) - \angle X(t_a^s, \Omega_{kl}) \quad (2.62)$$

Eğer tepe frekans haznesi ardışık zaman segmentlerinde değişirse yani k_1 iken k_2 olursa çevre frekans kanallarının fazları Eşitlik 2.63'te ifade edildiği gibi olur.

$$\angle Y(t_s^s, \Omega_k) = \angle Y(t_s^s, \Omega_{k_2}) + \beta [\angle X(t_a^s, \Omega_k) - \angle X(t_a^s, \Omega_{k_2})] \quad (2.63)$$

Eşitlik 2.63'te $\beta = 2/3 + \alpha/3$ ve $\angle Y(t_s^s, \Omega_{k_2}) = \angle Y(t_s^{s-1}, \Omega_{k_1}) + R_s \omega_{k_2}(t_a^s)$ olarak yerine konur.

2.3 Zaman Skalası Değişimi İçin Sinyal Modelleme Teknikleri

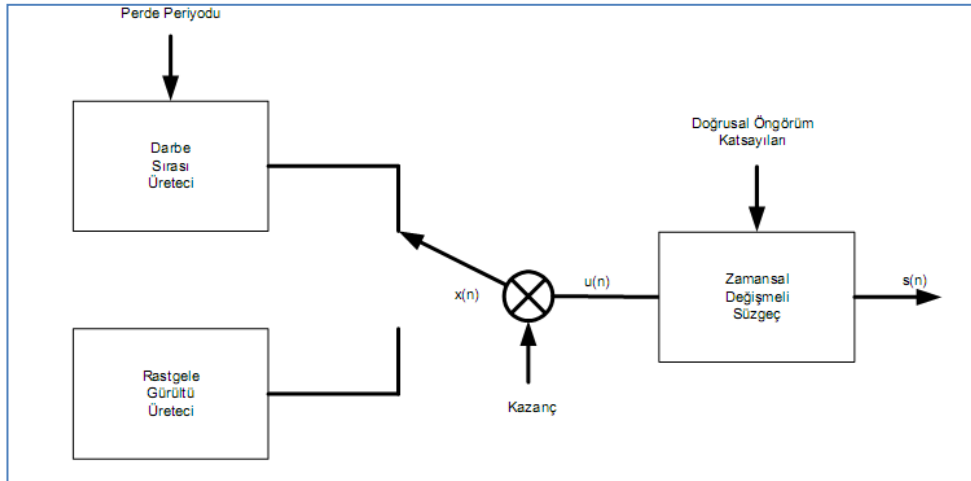
TSM için en temel sinyal modelleme tekniği Sinüs modelleme olup diğer teknikler genelde bu model üzerine kurulmuştur. Sinüs modelleme özellikle ötümlü sesler için oldukça başarılıdır. Bu model algoritmasının sinyalin ötümsüz veya ötümlü ötümsüz geçişlerini (veya ötümsüz ötümlü geçişleri) modellemesindeki problemleri düzeltmek

için Spektral Modelleme Sentezi (SMS) ve Geçiş Modelleme Sentezi (TMS) gibi teknikler geliştirilmiştir.

2.3.1 Sinüs modelleme

McAulay ve Quatieri (1986) tarafından geliştirilen Sinüs modelleme tekniği ses sinyalinin zaman ile değişen spektral karakterini zaman ile değişen sinüsler toplamı olarak ifade etmektedir.

Sinüs modelleme temel olarak sesin oluşmasını modellemektedir. Sesin oluşumu için iki önemli yapı kaynak ve yol olarak tanımlanmaktadır. Kaynak belirli frekanslara göre titreşimdir. İnsan sesinde kaynak ses telleri (ötümlü) veya ciğerlerden basınç ile çıkan havayı sıkıştıran organlar (ötümsüz ses) olabilir. Yol ise boğaz, ağız ve burundaki nefes yolu ve bu yoldaki gırtlak dil, dişler gibi organların yapısıdır. Şekil 2.13'te insan sesinin modellemesi yer almaktadır. Müzik enstrümanlarından çıkan sesler veya diğer sesler de belirli frekanslarda titreşen teller(örneğin telli çalgılar), materyaller (örneğin vurmali çalgılar) veya havanın sıkışması ile oluşur.



Şekil 2.13 İnsan sesinin modellemesi

Sesin kaynağı $e(t)$, zamana göre değişen yol modeli ise $h(t,t)$ ile gösterilirse ses Eşitlik 2.64 ile ifade edilir (McAuley ve Quatieri 1986a).

$$s(t) = \int_0^t h(t, t - \tau) e(\tau) d\tau \quad (2.64)$$

Bu ifadede sesin oluşumu için belirli frekanslarda titreşim gerektiği için $e(\tau)$ farklı frekans, faz ve genliklere sahip sinüsler toplamı olarak ifade edilebilir (Eşitlik 2.65).

$$e(t) = \sum_{l=1}^{L(t)} a_l(t) \cos \Omega_l(t) \quad (2.65)$$

Bu ifadede $L(t)$, t anında sesi oluşturan sinüslerin sayısı; $a_l(t)$, bu sinüslerin her birinin genliğini ifade eder. $\Omega_l(t)$ ise frekans ve o frekanstaki fazını içerdiğinden Eşitlik 2.66 ile ifade edilebilir (McAuley ve Quatieri 1986a).

$$\Omega_l(t) = V_l(t) + \phi_l \quad (2.66)$$

$$V_l(t) = \int_{t_1}^t w_l(\sigma) d\sigma \quad (2.67)$$

Eşitlik 2.66'da yer alan ϕ_l l . sinüsün faz sabitini ifade eder. Kaynağın frekansı azda olsa zamana bağlı değişebildiğinden her bir sinüsün frekansı $w_l(t)$ ile gösterilir ve l . sinüsün t anındaki değişken faz etkisi $V_l(t)$ Eşitlik 2.67'deki gibi ifade edilir.

Eşitlik 2.64'te yer alan yol modelinin Fourier dönüşümü ise Eşitlik 2.68'te gösterildiği gibidir (McAuley ve Quatieri 1986a).

$$H(w, t) = M(w, t) e^{j\Phi(w, t)} \quad (2.68)$$

Eşitlik 2.68'de $M(w, t)$ ve $\Phi(w, t)$ zamana bağlı filtre (yol modeli) genliği ve fazıdır. Yolun her bir sinüs kaynağına etkisi ayrı olarak $M_l(t) = M_l(w, t)$ ve $\Phi_l(t) = \Phi_l(w, t)$ ile gösterildiğinde Eşitlik 2.64'te yer alan ses modeli aşağıdaki gibi ifade edilir.

$$s(t) = \sum_{l=1}^{L(t)} A_l(t) \cos(\theta_l(t)) \quad (2.69)$$

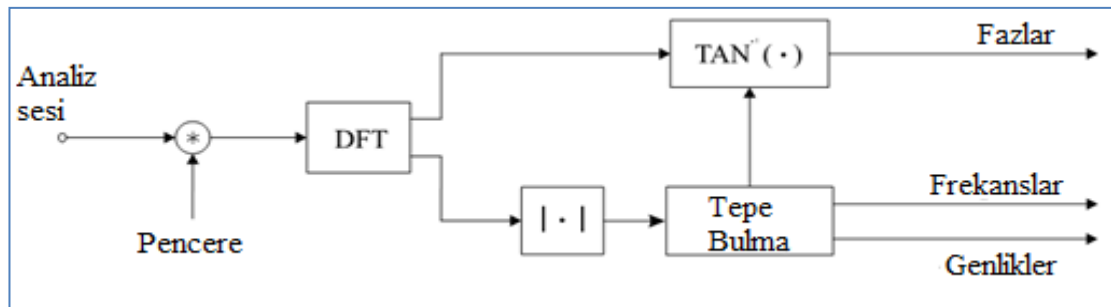
$$A_l(t) = a_l(t) M_l(t) \quad (2.70)$$

$$\theta_l(t) = \Omega_l(t) + \Phi_l(t) = V_l(t) + \phi_l(t) + \Phi_l(t) \quad (2.71)$$

Eşitlik 2.69 sesin sürekli zaman alanında Sinüs modelleme eşitliğidir. Bu modellemenin genlik ve fazı ise sırası ile Eşitlik 2.70 - 2.71’de ifade edildiği gibi kaynak ve yolun genlik ve fazlarını ihtiva etmektedir.

2.3.1.1 Analiz

Sinüs modellemenin analizi (sesin sinüs modelleme parametrelerin bulunması), şekil 2.14’te gösterildiği gibi analiz çerçevelerinin Kısa Zaman Fourier Dönüşümünden (STFT) frekans değerleri $w_l(t)$, genlik parametreleri $A_l(t)$ ve sabit faz farkları $\phi_l(t)$ değerlerinin tahmini ile başlar. McAuley ve Quatieri (1986) insan sesinin sinüs modellemesini önerdikleri için analizin 2. aşamasında kaynak ve yol parametrelerini çıkartmaktadır.



Şekil 2.14 Sinüs modelleme analizi (McAuley ve Quatieri 1986a)

Şekil 2.14’te sesin pencere fonksiyonu ile çerçevelere ayrıldığı görülmektedir. Pencere fonksiyonu olarak Hamming pencere seçilmektedir. McAuley ve Quatieri (1986a) pencere büyüklüğü olarak perde periyodunun en az 2,5 katının seçilmesini önermektedir. Ötümsüz çerçevelerde spektral çözünürlüğün sağlanması için de 20 ms ile 30 ms arasında bir uzunlukta çerçeve kullanımını önermektedir.

McAuley ve Quatieri (1986b), sinüs modelleme işlemini her çerçeve için ayrı olarak gerçekleştirdiğinden analiz ve sentez için Eşitlik 2.69 - 2.71'de yer alan t yerine kR ve örneklem veri notasyonu kullanmaktadır. Bu notasyonda R çerçevede yer alan örneklem sayısını k ise hangi çerçeve olduğunu göstermektedir. Bu notasyona göre STFT Eşitlik 2.72 ile hesaplanır.

$$S(w, kR) = \sum_m w(kR - m) s(m) e^{-jmw} \quad (2.72)$$

Fourier dönüşümü gerçekleştirildikten sonra çerçeve spektrumunun ardışık iki frekans noktası arasındaki eğiminin pozitiften negatife döndüğü spektral noktaları tepe olarak alınır. Yani $S(w_l, kR) - S(w_{l-1}, kR) > 0$ ve $S(w_{l+1}, kR) - S(w_l, kR) < 0$ sağlayan w_l frekansları yerel tepe noktaları olduğundan tepe sinüslerin frekansı olarak alınır.

$\hat{\omega}_l^k = \hat{\omega}_l(kR)$ olmak üzere $\hat{\omega}_l^k$ k . çerçevenin l . sinüs tahmininin frekans değeri olmak üzere genlik ve faz tahminleri Eşitlik 2.73 - 2.74 ile ifade edilir.

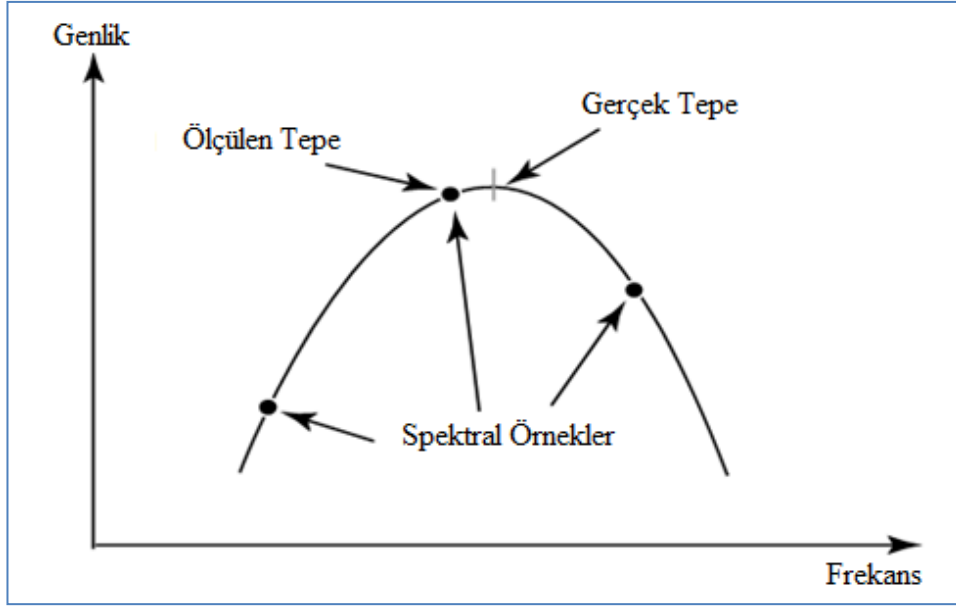
$$\hat{A}_l^k = |S(\hat{\omega}_l^k, kR)| \quad (2.73)$$

$$\hat{\theta}_l^k = \angle S(\hat{\omega}_l^k, kR) \quad (2.74)$$

McAuley ve Quatieri bu değerleri hesaplandıktan sonra Eşitlik 2.70 - 2.71'de yer alan sesin kaynak ve yol parametrelerinin şekil 2.15'de de gösterildiği gibi hesaplanmasını önermiştir.

Ses yolunun genliği $\hat{M}_l^k = \hat{M}(\hat{\omega}_l^k, kR)$ ve fazı $\hat{\Phi}_l^k = \hat{\Phi}(\hat{\omega}_l^k, kR)$ çıkarılarak kaynak parametreleri de $\hat{a}_l^k = \hat{A}_l^k / \hat{M}_l^k$ ve $\hat{\Omega}_l^k = \hat{\theta}_l^k - \hat{\Phi}_l^k$ ile hesaplanmaktadır. Ses yolunun genliği $\hat{M}_l^k$ değerini hesaplamak için homomorfik dekonvolüsyon veya hep-kutup yöntemlerinden biri kullanılır.

Ses yolunun fazını hesaplamak ise daha zordur. Fakat ses yolunun minimum faz filtre modeli olduğu varsayılarak sistem fazının $(\hat{\Omega}_l^k)$ ve genliğinin $(\hat{S}_l^k)$ logaritmaları



Şekil 2.16 Gerçek frekans tepelerini bulmak için parabolik frekans enterpolasyonu (Smith ve Serra 1987)

Gerçek tepe örnek frekansı ile ölçülen tepe örnek frekansı arasındaki fark Eşitlik 2.75 ile hesaplanır (Smith ve Serra 1987).

$$p = \frac{1}{2} \frac{\alpha - \gamma}{\alpha - 2\beta + \gamma} \quad (2.75)$$

Bu eşitlikte β ölçülen tepe örneğinin frekans değerini α ve γ ise ölçülen tepe örneğinden önceki ve sonraki örneklerin frekans değerlerini ifade etmektedir. Böylece gerçek frekans değeri Eşitlik 2.76'da ifade edildiği gibi hesaplanır.

$$w^* = w_\beta + p \quad (2.76)$$

Bu frekansdaki genliğin karmaşık sayı değeri ise Eşitlik 2.77 ile hesaplanarak genliği ve fazı bulunur.

$$A = \beta - \frac{1}{4}(\alpha - \gamma)p \quad (2.77)$$

Ninness (2000) yaptığı çalışma sonucunda McAuley ve Quatieri (1986a, 1986b) ve Smith ve Serra (1987) tarafından önerilen iki yöntem ile analizi yapılan seslerin sentezlenmesinde pek fazla algısal bir fark olmadığını belirtmiştir.

2.3.1.2 Sentez

Sinüs modellemenin analizi çerçeve çerçeve yapıldığından analiz ile hesaplanan sinüs bileşenlerinin parametrelerinden sinyalin sentezlenmesi tekrar çerçeve çerçeve gerçekleştirilirse çerçeve geçişlerinde kesiklik meydana gelebilir. Bu kesikliğin engellenmesi için sinyal geçişlerinin yumuşatılması gerekir (McAuley ve Quatieri, 1986b). Ses yolu ve kaynağın genlik ve faz farkları frekanslarına bağlı olduğundan analizde k . çerçeve için ölçülen tepe frekanslarının $k+1$. çerçevede ölçülen tepe frekansları ile ilişkilendirilmesi gerekir. Özellikle ses sinyalinin ötümlü-ötümsüz veya ötümsüz-ötümlü geçişlerinde (müzik enstrümanlarında da nota değişimleri) bazı sinüs bileşenleri eşleştiremeyebilir. $k+1$. çerçevede yer alan herhangi bir sinüs bileşeni k . çerçevedeki sinüs bileşenleri ile ilişkilendirilemiyor ise bu sinüs bileşeninin (ilk $k+1$. çerçevede oluştuğundan) DOĞUM noktası $k+1$. çerçevedir. Benzer şekilde k . çerçevede yer alan herhangi bir sinüs bileşeni $k+1$. çerçevedeki sinüs bileşenleri ile ilişkilendirilemiyor ise bu sinüs bileşeninin (en son k . çerçevede yer aldığından) ÖLÜM noktası k . çerçevedir.

McAuley ve Quatieri(1986b) tarafından önerilen yöntemde; ardışık çerçeveler arasındaki sinüs bileşenleri ilişkilendirildikten sonra sinüs bileşenlerine ait yol ve kaynak modeli için genlik geçişlerinin yumuşatılması k . sentez çerçevesinin her bir örneği için aşağıdaki doğrusal enterpolasyon eşitliği ile basitçe hesaplanmaktadır (Eşitlik 2.78 - 2.79).

$$\hat{M}_l(t) = \hat{M}_l^k + (\hat{M}_l^{k+1} - \hat{M}_l^k)t/T \quad (2.78)$$

$$\hat{a}_l(t) = \hat{a}_l^k + (\hat{a}_l^{k+1} - \hat{a}_l^k)t/T \quad (2.79)$$

Smith ve Serra (1987) tarafından daha çok müzik için önerilen Sinüs modellemeye yol ve kaynak kullanılmadığından geçişlerin yumuşatılması ardışık çerçevelerde ilişkilendirilen sinüs bileşenlerinin genlikleriyle doğrudan uygulanmaktadır (Eşitlik 2.80).

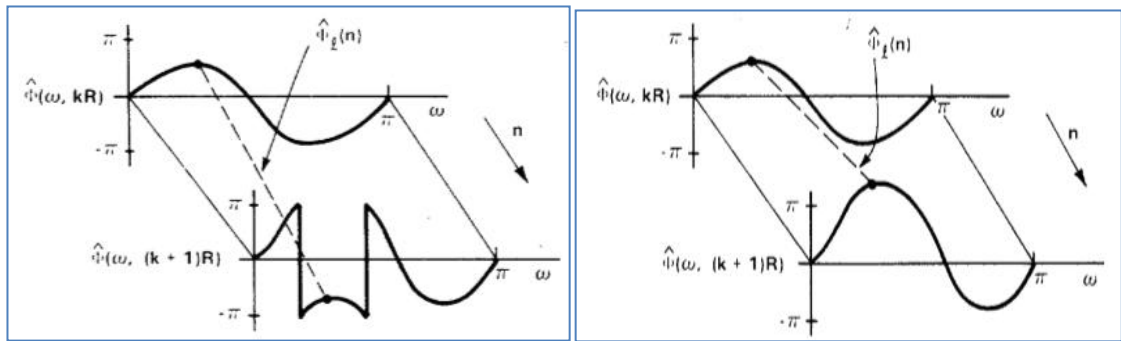
$$\hat{A}_l(t) = \hat{A}_l^k + (\hat{A}_l^{k+1} - \hat{A}_l^k)t/T \quad (2.80)$$

Eşitlik 2.78-2.80’de yer alan T çerçeve boyutu olup t ise $0 \leq t < T$ olmak üzere zamandır. $\hat{A}_l^{k+1}$ ve $\hat{A}_l^k$ $k+1$. ve k . çerçevelerde ölçülen ve birbirleriyle ilişkilendirilen sinüs bileşenlerine ait genliklerdir.

Eşitlik 2.69 ile modellenen ses yolunun filtre parametreleri yavaş değiştiğinden fazı da aynı şekilde yavaş değişir. Bu nedenle yol fazının da lineer enterpolasyon ile hesaplanması yeterlidir (Eşitlik 2.81).

$$\hat{\Phi}_l(t) = \hat{\Phi}_l^k + (\hat{\Phi}_l^{k+1} - \hat{\Phi}_l^k)t/T \quad (2.81)$$

Sinüs bileşenlerinin frekanslarının ve kaynak fazlarının geçiş yumuşatması ise daha karmaşık bir yöntem olup kübik enterpolasyon gerektirir. Bir başka zorluk ise faz değeri π ile $-\pi$ arasında ölçüldüğünden enterpolasyon hesabında şekil 2.17’de gösterildiği gibi dikkate alınmalıdır.



Şekil 2.17 Sinüs fazlarının enterpolasyonunda kesiklik

a. Mod 2π ile ölçümden dolayı kesiklik, b. Kesikliğin düzeltilmesi ve gerçek faz farkı

Eşitlik 2.69’da yer alan sinüslerin faz için kübik polinom $0 \leq t < T$ olmak üzere Eşitlik 2.82 ile ifade edilmektedir. Fazın türevi de frekansı vermektedir (Eşitlik 2.83)

$$\hat{\theta}_l(t) = \zeta + \gamma t + \alpha t^2 + \beta t^3 \quad (2.82)$$

$$\hat{\omega}_l(t) = \gamma + 2\alpha t + 3\beta t^2 \quad (2.83)$$

Bu eşitlikteki polinom sabitleri, $t=0$ ve $t=R$ için ardışık iki çerçevedeki faz ve frekans değerleri yerine konarak elde edilir (Eşitlik 2.84).

$$\alpha = \frac{3}{R^3}(\hat{\theta}_l^{k+1} - \hat{\theta}_l^k - \hat{\omega}_l^k R + 2\pi M) - \frac{1}{R}(\hat{\omega}_l^{k+1} - \hat{\omega}_l^k) \quad (2.84.a)$$

$$\beta = \frac{-2}{R^3}(\hat{\theta}_l^{k+1} - \hat{\theta}_l^k - \hat{\omega}_l^k R + 2\pi M) - \frac{1}{R^2}(\hat{\omega}_l^{k+1} - \hat{\omega}_l^k) \quad (2.84.b)$$

$$\zeta = \hat{\theta}_l^k \quad (2.84.c)$$

$$\gamma = \hat{\omega}_l^k \quad (2.84.d)$$

Tüm bu parametreler hesaplandıktan sonra sinüs modelleme parametrelerinden Eşitlik 2.85 ile sentezleme gerçekleştirilir.

$$\hat{s}(n) = \sum_{l=1}^{L(n)} \hat{A}_l(n) \cos(\hat{\theta}_l(n)) \quad (2.85)$$

Smith ve Serra (1987) ses sentezinde tepe sinüslerinin fazını kullanmamıştır. Hem analiz hem de sentezde faz hesaplamaları kullanılmadığından özellikle kübik enterpolasyondan kaynaklanan hesap yüklerinden kurtulunmuştur. Sadece genlikle yapılan bu sentezlemede sentezlenen ses orijinal sesten farklı olmakla birlikte birçok uygulama için algısal olarak aynı sesi vermektedir

2.3.1.3 Sinüs modelleme ile zaman skalası değişimi

Sinüs modelleme sesin oluşumunu modellediği için zaman skalası değişiminde de sesin oluşumunun değişimi modellenmektedir. Sabit bir oranda değişim için değişim oranı ρ ile ifade edilirse orijinal hızdaki sesin t_0 anındaki kısmı sentezlenen sesteki $t'_0 = \rho t_0$ anında sentezlenir. $\rho > 1$ için ses yavaşlatılmış $\rho < 1$ için ses hızlandırılmış olur. Sentezlenen sesin ifadesi Eşitlik 2.86'daki gibidir (McAuley ve Quatieri 1986b).

$$s'(t') = \sum_{l=1}^{L(t')} A_l(t') \cos(\theta_l(t')) \quad (2.86)$$

Sentezlenen sesin t' anındaki kısmı, orijinal sesin $\rho^{-1}t'$ anındaki kısmını modelleyeceğinden genliği de bu şekilde modellenir ve Eşitlik 2.87 ile ifade edilir.

$$A_l(t') = A_l(\rho^{-1}t') = a_l(\rho^{-1}t') M_l(\rho^{-1}t') \quad (2.87)$$

Fazın modellenmesi için ses yolu fazı da ρ ile aynı oranda değişeceği için sentez sesinin fazı Eşitlik 2.88 ile ifade edilir.

$$\theta_l(t') = \Omega_l(t') + \Phi_l(\rho^{-1}t') \quad (2.88)$$

Sentez sesinin kaynağının frekansı da aynı şekilde orijinal sesin $\rho^{-1}t'$ anındaki frekans ile aynı olacağından Eşitlik 2.89 - 2.90 ile ifade edilir.

$$\Omega_l(t') = \int_{t_i}^{t'} \omega_l(\rho^{-1}\tau) d\tau + \phi_l \quad (2.89)$$

$$\Omega_l(t') = \int_{t_i}^{\rho^{-1}t'} \omega_l(\sigma) d\sigma / \rho^{-1} + \phi_l = V_l(\rho^{-1}t') / \rho^{-1} + \phi_l \quad (2.90)$$

2.3.2 Spektral Modelleme Sentezi (SMS)

Spektral modelleme sentezinde (SMS) sinüs modellemeye gürültü bileşeni eklenmektedir. Serra ve Smith (1990) tarafından önerilen bu modelde sinüsler ses sinyalinin sadece rasgele olmayan hemen hemen periyodik kısımlarını modellerken, kalan gürültü bileşenleri de ayrı olarak modellenir. SMS’de sinyal Eşitlik 2.91’deki gibi ifade edilir.

$$s(t) = \sum_{R=1}^R A_r(t) \cos(\theta_r(t)) + e(t) \quad (2.91)$$

Eşitlik 2.91’de yer alan $A_r(t)$ ve $\theta_r(t)$ sinüs modellemeye olduğu gibi r . sinüs bileşeninin genliği ve fazı, $e(t)$ ise gürültü bileşenidir. Sinyalin R sinüs bileşeni toplanarak rastgele olmayan (deterministik) kısmı oluşturulur. $\theta_r(t)$ fazı ise anlık frekansın (radyan cinsinden) integralinin ilk sabit faz farkı ile toplanmasıyla elde edilir (Eşitlik 2.92).

$$\theta_r(t) = \int_0^t \omega_r(\tau) d\tau + \theta_r(0) \quad (2.92)$$

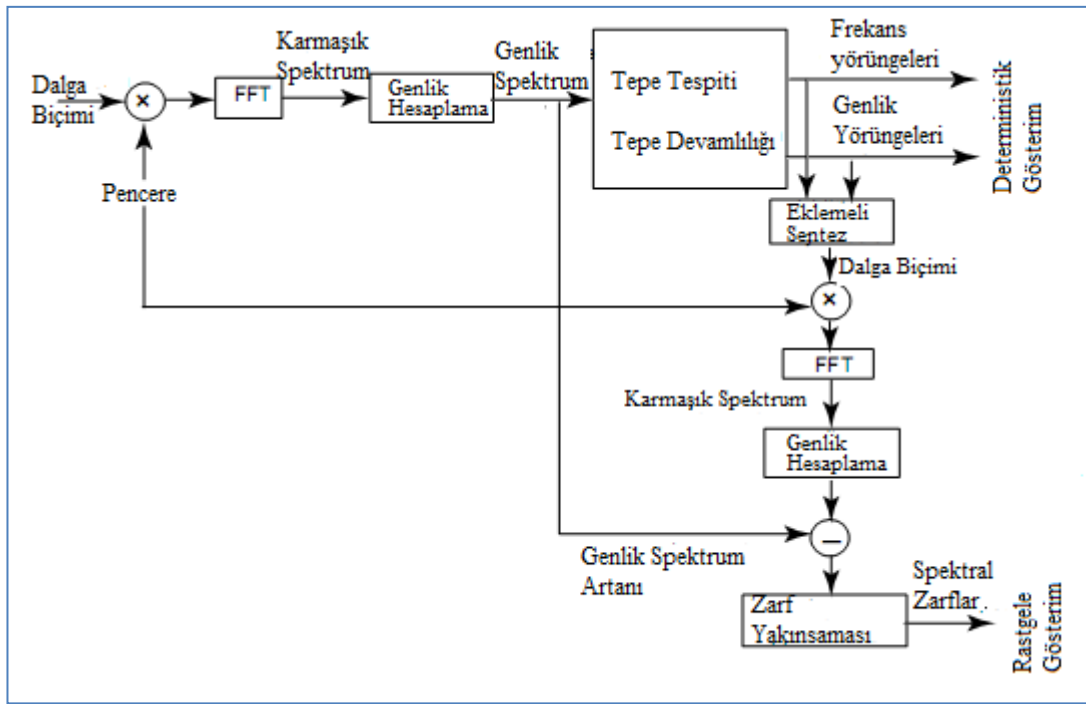
Sinyalin gürültü bileşeni ise filtrelenmiş beyaz gürültü olarak modellenir ve Eşitlik 2.93 ile ifade edilir.

$$e(t) = \int_0^t h(t, \tau) u(\tau) d\tau \quad (2.93)$$

Eşitlik 2.93’te $u(\tau)$ beyaz gürültü ve $h(t, \tau)$ zamana bağlı değişen filtrenin dürtü tepkisi.

2.3.2.1 SMS Analiz

SMS analizi sinüs modelleme analizi gibi başlar. Ses sinyali pencere fonksiyonu ile çerçevelere ayrılıp FFT ile kompleks (karmaşık) spektrumu hesaplanır. Çerçeveler arası sinüs ilişkilendirmeleri ile tepe sinüs bileşenlerin frekans ve genlik spektrumu yörüngeleri bulunur. Bu sinüs bileşenlerin toplamı ile sentezlenen sinyalin deterministik kısmı sinyalden çıkartılarak sinyalin gürültü bileşeni (rastgele kısmı) bulunur. Daha sonra sinyalin gürültü bileşeninin spektral yakınsaması hesaplanır (şekil 2.18).



Şekil 2.18 SMS (deterministik artı rastgele çözümleme) analizi

Genlik spektrumu N 'lik FFT ile hesaplandığında; spektrum noktalarının çözünürlüğü, ses sinyalinin örnekleme frekansının FFT boyutuna oranı olacaktır. Çerçeve uzunlukları spektral çözünürlüğü arttırmakta yetersiz kaldığından çözünürlüğü artırmak için pencere fonksiyonu ile ayrılan sinyal çerçevelerine sıfır örneklemleri eklenerek (zero padding) frekans çözünürlüğü artırılır. Frekans çözünürlüğünün binde bir hassasiyetinde olması verimlilik açısından aranmaktadır (Serra ve Smith 1990).

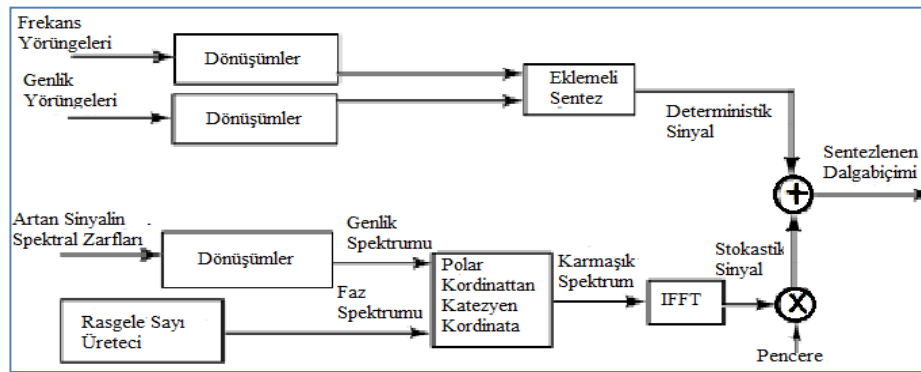
Sinüs bileşen gürültü çözümümü için sinüs tepe sürekliliği oldukça önemlidir. Sinyalin temel frekansının hesaplanması tepelerin sürekliliğinin kararlı olarak sağlanmasına imkan vermektedir.

Spektral yakınsamada ses sinyalinin gürültü bileşeni pencere fonksiyonu (ses sinyalinin sinüslerini bulmak için kullanılan pencerenin aynısı) ile çerçevelere ayrıldıktan sonra genlik spektrumu hesaplanır. Gürültü bileşeninin rastgeleye yakın (quasi-periodic) olduğu varsayımıyla sesin karakteristiğine sadece gürültünün şekli etki ettiğinden genlik spektrumları sadece spektrum zarfı ile ifade edilebilir. Bunun için aşağıdaki teknikler kullanılabilir (Serra ve Smith 1990);

- Spline Enterpolasyon,
- En küçük kare metodları,
- Düz çizgi yakınsamaları,
- LPC (Doğrusal tahmin kodlaması- Linear Predicting Coding).

2.3.2.2 SMS sentez

SMS sentezi için; hesaplanan analiz parametrelerinde istenilen değişikliğe göre yapılan dönüşümlerden sonra sinyalin deterministik ve gürültü bileşenleri eklenir (Şekil 2.19).



Şekil 2.19 SMS sentezi

Gürültü bileşeninin analizinde sadece gürültü spektrumunun genliği hesaplandığından faz spektrumu için rastgele değerler atanır. Bu sayede gürültü spektrumunun ters FFT'si gürültü bileşenini verir.

2.3.2.3 SMS ile zaman skalası değişimi

SMS zaman skalasını değiştirmek için uygulandığında sesin perdesi ve gürültü bileşeninin gürültüsü korunur. Hem sinüs bileşenlerinin hem de gürültü bileşeninin sentez ve analiz adımlarının oranı zaman skalası değişim oranını verir. Fakat bu yöntem geçişlerde ses kalitesi açısından sıkıntı yaratır (Serra ve Smith 1990).

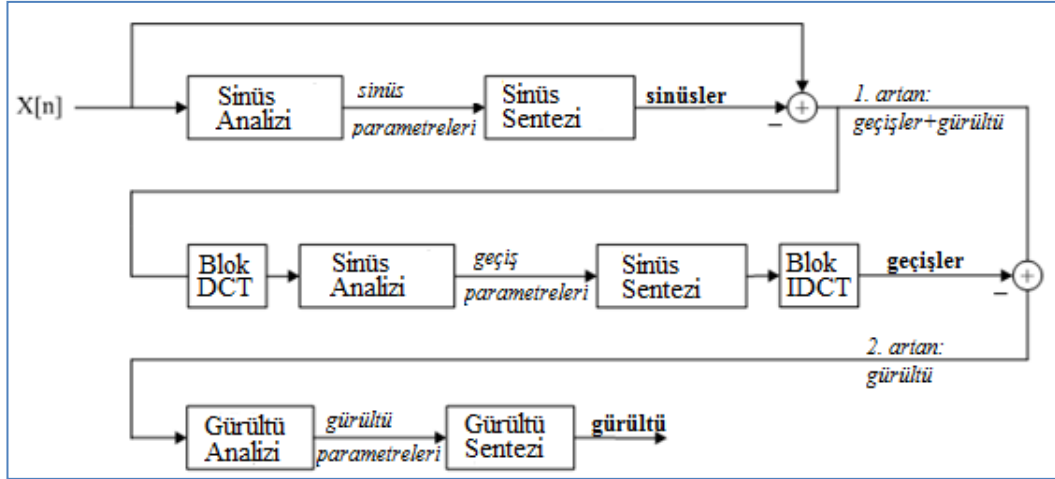
2.3.3 Geçiş Modelleme Sentezi (TMS)

Sesin SMS tekniği ile modellenmesinde ses sinyali deterministik ve rastgele bileşenlerinden oluşurken, Geçiş Modelleme Sentezi (TMS-Transient Modeling Synthesis) bu modeli geliştirerek geçişlerini de ayrı olarak hesaba katmaktadır (Verma vd. 1997).

Geçişlerin açık bir şekilde ayrı olarak ele alınması daha gerçekçi ve sağlam bir sinyal modeli kurulmasını sağlamakta ve sinüs modelleme ile SMS'e göre esneklik kazandırmaktadır. SMS'de geçişler ve gürültü birlikte filtrelenmiş beyaz gürültü olarak modellenirken TMS'de geçişler gürültüden ayrı modellenmekte, böylece sinüs modelleme ve SMS'in sesin keskin geçişlerindeki handikapları azaltılmaktadır.

2.3.3.1 TMS analiz

Şekil 2.20'de görüldüğü gibi TMS analizinde ilk önce diğer sinyal modelleme tekniklerinde olduğu gibi sesin baskın sinüs bileşenleri modellenir. Sinüs bileşenleri ile sentezlenen ses orijinal sesten çıkartıldıktan sonra kalan artık sinyal geçişleri ve gürültüyü içermektedir. Bu artık sinyalden geçiş bileşenleri modellenip çıkartıldığında kalan sinyal SMS'de olduğu gibi gürültü olarak modellenir.



Şekil 2.20 TMS analizi (Verma vd. 1997)

Sinüs bileşenlerinin ve gürültü bileşeninin modellenmesi SMS modellemesine benzer yapılmaktadır. Geçişin modellenmesi için sinüsler ile geçişler arasındaki dönüşüm ilişkisinden faydalanılır.

Sinüs bileşenlerinin analizinde zaman bölgesi sinyalinin spektral tepelerinin izlenmesi bulunurken, geçişlerin analizinde frekans bölgesi sinyalin spektral tepelerinin izlenmesi gerçekleştirilir. Bunun için DCT (Ayrık Kosinüs Dönüşümü) kullanılır. DCT zaman bölgesindeki geçişleri sinüslere çevirir.

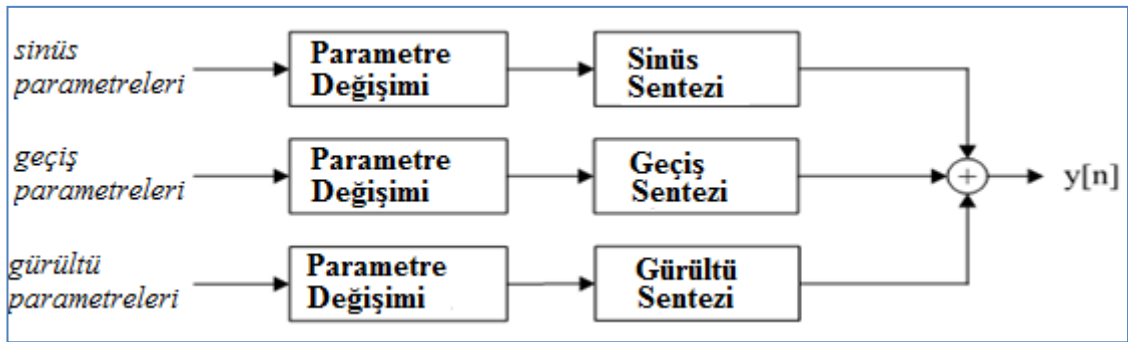
Zaman bölgesinde yavaşça değişen bir sinüs sinyali (salınım) DCT bölgesinde darbe olarak görülür. Zaman bölgesindeki darbeler ise frekans bölgesinde salınım olarak görülür. Zaman bölgesinde bir kosinüs sinyalin frekansı ne kadar büyük ise DCT bölgesindeki dönüşümü başlangıçtan o kadar uzak olduğundan zaman bölgesindeki darbe başlangıçtan ne kadar uzaksa DCT bölgesinde o kadar yüksek frekans kosinüs olarak görülür. Bu nedenle DCT bölgesinde yapılacak bir sinüs analiz ile tespit edilen spektral tepelerin genliği, frekansı ve fazı zaman bölgesi için tespit edilen geçişlerin özelliklerini belirler.

TMS analizinde; DCT'nin bu özelliklerinden faydalanarak şekil 2.20'de de gösterildiği gibi önce birbirleriyle örtüşmeyen blokların DCT'si alınır, daha sonra da elde edilen sinyalin sinüs bileşenlerinin parametreleri hesaplanarak geçiş parametreleri tespit edilir.

Tespit edilen bu parametrelerden sinüs model sentezi yapıp ters DCT alınarak bulunan geçiş sinyali ilk artık sinyalden(geçişler ve gürültüyü içeren) çıkartılır.

2.3.3.2 TMS sentez

TMS modelleme sentezinde analiz aşamasında tespit edilen parametrelerden sinüs bileşenleri, geçiş bileşenleri ve gürültü bileşeni ayrı ayrı modellenerek birleştirilir ve sinyal sentezlenmiş olur.



Şekil 2.21 TMS sentezi

2.3.3.3 TMS ile zaman skalası değişimi

Zaman skalası değişimi ile TMS sentezi için sinüs bileşenlerinin ve gürültü bileşeninin değişimi Sinüs modelleme ve SMS'deki gibi gerçekleştirilir. Geçişlerin de sinüs ve gürültüyle paralel oranda değiştirilmesi gerektiğinden DCT blok uzunlukları aynı oranda değiştirilir. Böylece DCT blokda daha çok veya daha az sinüs dalgası görülür ve geçişler zaman bölgesindeki uygun lokasyonlarda sağlanır.

2.4 Paket Anahtarlama Ağılarda Tıkanıklık İçin Zaman Skalası Değişimi

Ses ve verinin bir arada iletildiği İnternet gibi paket anahtarlama ağılar üzerinden iletişimde ses paketlerin gecikmesi değişebilmekte hatta bu paketler alıcıya hiç ulaşmadan ağda (yok olabilmekte) kaybolabilmektedir. İnternet telefon gibi gerçek zamanlı sesli iletişimin yapıldığı durumlarda alıcılar ses paketlerini uzun süre bekleyemeyeceği için zamanında ulaşmayan paketler için tekrar gönderme gibi

yöntemler kullanılamamaktadır. Bu nedenle herhangi bir paket kaybı alıcı tarafındaki ses sinyalinin anlaşılabilirliğini veya kalitesini bozmaktadır.

Gerçek zamanlı ses iletişimi için paket kaybının iki temel sebebi vardır. İlki paketin ağda yok olması ve alıcıya hiç ulaşmaması, diğeri ise paketin alıcıya zamanında ulaşamamasıdır. Paketin ağda tamamen kaybolmasını engellemek için sadece alıcı tarafında uygulanabilecek bir yöntem bulunmamaktadır. Gecikme ve jitterden (gecikme değişiminden) kaynaklanan paket kayıpları ise alıcının tampon bellekleri geniş tutularak azaltılabilmektedir. Alıcı tampon belleklerinin geniş tutulması ise tampon gecikmesini dolayısıyla uçtan uca gecikmeyi artırmaktadır. Gecikme, telefon gibi karşılıklı iletişimde interaktifliği engellediğinden karşılıklı iletişimin kalitesini bozmaktadır (ITU-T G114 2003). Bu nedenle jitteri telafi etmek için kullanılan tampon gecikmesi (playout time) ve paket kaybı birbirleriyle çelişen ve iletişim kalitesi için optimize edilmesi gereken parametrelerdir.

2.4.1 Alınan sesin çalma zamanlaması

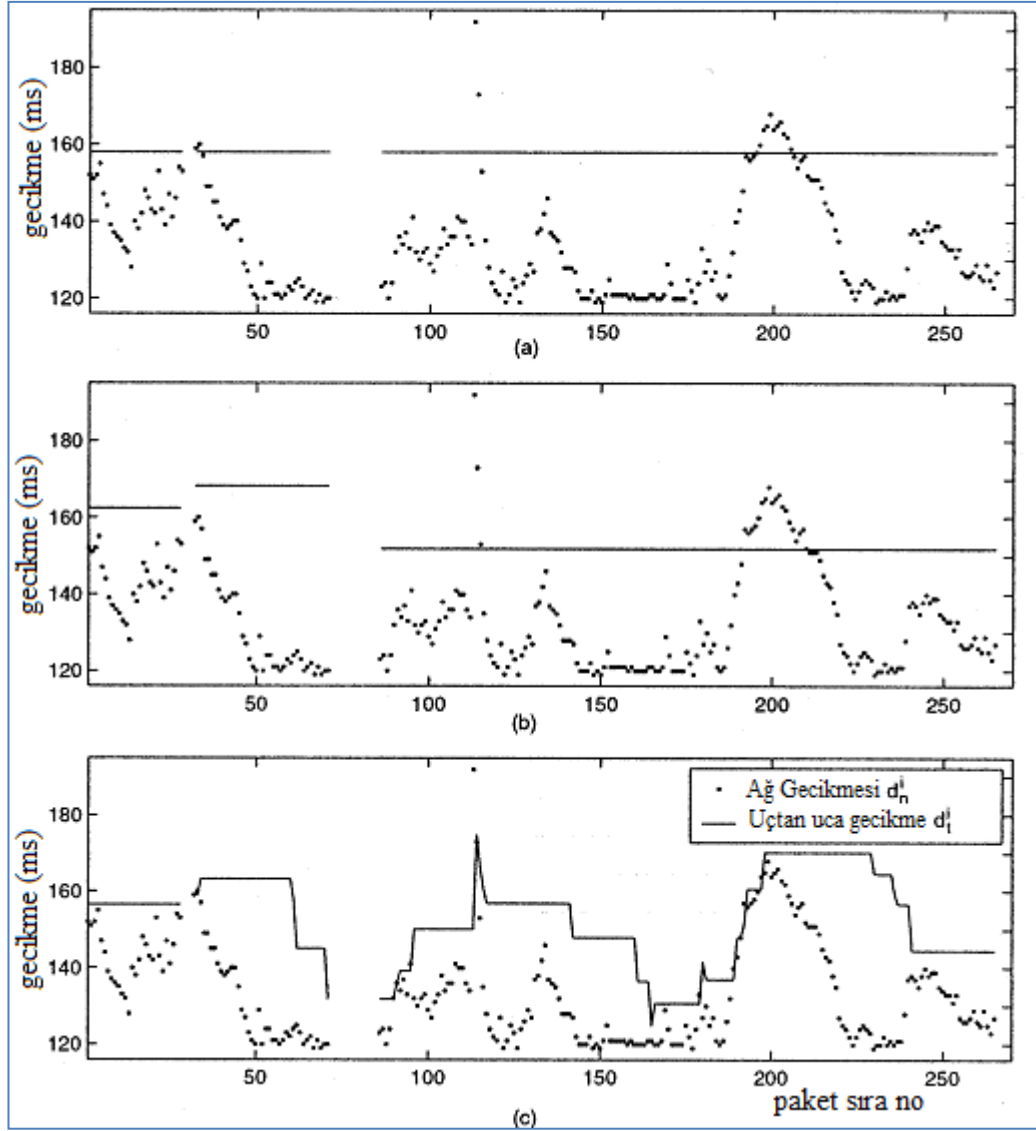
Paket anahtarlama ağı üzerinden ses iletişimde, analog kaynaktan alınan ses sayısal hale çevrilir ve uygun uzunlukta parçalara ayrılarak belirlenen kodlayıcı ile kodlanır. Gönderen ağ noktası tarafında yapılan bu işlemler ile elde edilen tüm kodlanmış ses parçaları doğrudan paketlenerek alıcıya ulaştırılmak üzere ağdan gönderilir. Genelde her pakette gönderilen ses parçası uzunluğu sabit olduğundan gönderen taraftaki ardışık iki paketin ağa sunulma zamanları arasındaki fark paketin içerdiği sesin süresi kadardır.

Ağ üzerinden iletilen ve alıcı tarafından alınan ses paketleri jitter telafisi için tampon belleğe aktarılır. Bu jitter ayarlama tekniğinde alınan ses paketi hemen çalınmayarak planlanan çalma zamanı gelene kadar tampon bellekte bekletilir. Bu nedenle belirlenen çalma zamanında tampon bellekte tutulan ses çalınmaya başladığında, bu sesin çalma işlemi tamamlanmadan yeni paketin gelmiş olması ve tampon belleğe aktarılması gerekmektedir. Aksi durumda ulaşmayan veya geciken paketin gecikme süresi kadar bir çalma boşluğu olacak ve ses kalitesi bozulacaktır.

En az paket kayıplı ses kalitesini sağlamak için tampon bellek büyüklüğünü en yüksek ağ gecikmesi ile en düşük ağ gecikmesi arasındaki süre farkı kadar sesi saklayabilecek büyüklükte tutmak gerekmektedir. Bu durum gerçek zamanlı konuşmada sesin alıcıya geç ulaşmasına neden olduğundan tampon bellek çok yüksek tutulamamaktadır.

Uçtan uca gecikmenin sabit tutulmaya çalışıldığı sabit zamanlamalı çalıda tampon bellek büyüklüğü genelde paketlerin geç ulaşmasından dolayı kayıplara da az da olsa izin verecek şekilde belirlenir. Çalma zamanlamasının değişebildiği uyarlanabilir çalma ise konuşma aralarındaki sessizlik boşlukları esnasında uyarlanabilen veya konuşma esnasında uyarlanabilen olmak üzere 2 ayrı yöntem ile gerçekleştirilebilir. Şekil 2.22 sabit ve uyarlamalı çalma yöntemlerine göre uçtan uca gecikmelerini ve paket kayıplarını örnek bir konuşmanın paketlerinin ağ gecikmelerine göre göstermektedir (Liang vd. 2003).

Şekil 2.22.a’da, tüm iletişim boyunca uçtan uca gecikmenin sabit tutulduğu ve tampon bellek genişliğinin buna göre belirlendiği örnek çalma zamanlamasıdır. Ortalama gecikme 160 ms olup paket kayıpları da fazla olduğundan verimli bir yöntem değildir. Ortadaki grafikte ise uçtan uca gecikmeye göre çalma zamanlaması konuşma boşluklarında uyarlanmaktadır. Gecikmeden kaynaklanan paket kaybı açısından avantaj sağlamasa da ortalama uçtan uca gecikmesi daha iyidir. En alttaki grafikte ise çalma zamanlamasının konuşma esnasında da dinamik olarak değiştirilebildiği uyarlanabilir çalma zamanlaması tekniğini göstermektedir. Bu grafikte hem gecikmeden dolayı paket kaybının azaltıldığı hem de ortalama gecikmenin 150 ms’nin altına düşürüldüğü görülmektedir.



Şekil 2.22 Çalma zamanlama yöntemlerine göre uçtan uca gecikme ve ağ gecikmesi (Liang vd. 2003)

a. Sabit çalma zamanlaması, b. Konuşma periyotları arasındaki boşluklarda uyarlanabilir çalma zamanlaması, c. Konuşma esnasında uyarlanabilir çalma zamanlaması

2.4.2 Gecikme tahmin yöntemleri

VoIP gibi paket anahtarlama ağılar üzerinden gerçek zamanlı konuşma veya telekonferans iletiminde uçtan uca gecikmenin ayarlanması (uyarlanabilir çalma zamanlamasının uygulanabilmesi) için ilk önce paketlerin ağ gecikmelerinin tahmini gerekmektedir. Bu tahmine yönelik birçok algoritma bulunmaktadır.

Ramjee vd. (1994), uçtan uca gecikmenin belirlenmesinde ağ gecikmesi ve jitter tahmini için AR (autoregressive) model kullanmışlardır. Bu modelde uçtan uca gecikmeler Eşitlik 2.94 - 2.95 ile hesaplanan tahminlere göre belirlenmektedir.

$$\hat{d}_i = \alpha \hat{d}_{i-1} + (1-\alpha)n_i \quad (2.94)$$

$$\hat{v}_i = \alpha \hat{v}_{i-1} + (1-\alpha)|\hat{d}_i - n_i| \quad (2.95)$$

Bu eşitliklerde yer alan $\hat{d}_i$ AR modeline göre ses paketleri için ağ gecikmesi tahmini olup $\hat{v}_i$ gecikme değişimi (jitter) tahminidir. n_i ise en son alınan i . paketin ağ gecikmesi olup α uyarlama oranıdır. Bu tahminlere göre uçtan uca gecikme D_i Eşitlik 2.96'ya göre belirlenir.

$$D_i = \hat{d}_i + \mu \hat{v}_i \quad (2.96)$$

Ramjee vd. (1994) ağ gecikmesi ve varyans tahminlerini alan her pakete göre yapılmasını fakat D_i güncellemesinin konuşmalar arasındaki seslik anlarında güncellenmesini modellemişlerdir. Yapılan çalışmada 4 ayrı durum için ayrı algoritmalar AR modeline göre önerilmiştir. İlk algoritmada üssel ağırlıklı ortalama kullanılmış olup Eşitlik 2.94'te yer alan α 0,998002 olarak seçilmiştir. İkinci algoritmada gecikme artışlarına daha hızlı tepki vermek için α değeri azalan gecikme ($n_i \leq \hat{d}_{i-1}$) için ilk algoritma ile aynı seçilmiş olmasına rağmen artan gecikme için 0,75 olarak belirlenmiştir. Üçüncü algoritma ise gecikmeyi en aza indirmeye yönelik olup bir önceki konuşma periyodunda alınan paketlerin en düşük ağ gecikmesi gecikme tahmini olarak belirlenir. Dördüncü algoritma ise ani gecikme artmalarını tespit ve takip etmeye yöneliktir. Algoritma normal ve spike mod olmak üzere 2 ayrı moda sahiptir. Spike moda ağ gecikmesi tahmini bir önceki paketin gerçekleşen gecikmesine göre takip edilmektedir. Artışın olmadığı normal durumlarda ise α . 0,875 olarak alınmaktadır.

Jung ve Atwood (2006) AR modelini uygularken uçtan uca gecikme yerine jitteri kullanmışlardır. Bu yöntemde ardışık iki paketin ulaşma gecikmeleri arasındaki fark

anlık jitter olarak belirlenmektedir. Daha sonra Eşitlik 2.94'deki gecikme tahmini yerine yumuşatılmış jitter tahmini ve alınan paketin ağ gecikmesi yerine anlık jitter konarak Eşitlik 2.97'deki gibi güncellenmektedir.

$$\hat{J}_i = \alpha \hat{J}_{i-1} + (1 - \alpha) |J_{i-1}| \quad (2.97)$$

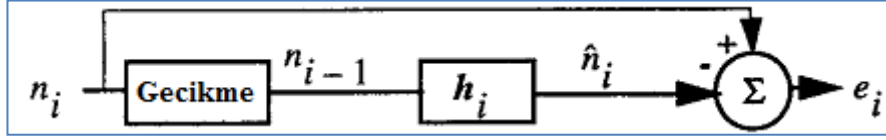
Ani gecikme artmalarının tespitine ve takibine yönelik Moon vd. (1998) AR modelini geliştirmek için histogram tabanlı tahminleme kullanmışlardır. Bu modelde uçtan uca gecikmenin tahmini, önceki konuşma periyodundaki ağ gecikmesi istatistikleri kullanılarak hedeflenen paket düşme oranına göre belirlenmektedir. Pinto ve Christensen (1999) tarafından önerilen optimum aralık tekniğinde de yeni başlayan bir konuşma periyodunun gecikmesi aynı benzer yöntemle belirlenmekte ve konuşma periyodları arasındaki sessiz aralık bu güncellenen uçtan uca gecikmeye göre güncellenmektedir.

Liang vd. (2003) uçtan uca gecikmeyi dinamik olarak değiştirebildiklerinden geçmiş konuşma periyodları (talkspurt) yerine en son alınan w adet paketin gecikmelerini küçükten büyüğe sıralayıp sıralamaktadır. Bu paketlerin standart sapmasını hesapladıktan sonra sıralamanın en sağına bu w adetteki en büyük gecikme ile hesaplanan standart sapmanın iki katını toplayarak sıralamanın en sağına eklemişlerdir. Benzer şekilde sıralamanın en soluna olabilecek en küçük ağ gecikmesini eklemişlerdir. Hedeflenen paket kaybını sağlayan bu sıralamadaki en küçük gecikme noktası ve bu noktanın sağındaki gecikme noktasını tespit edip interpolasyon yaparak uçtan uca gecikmeyi hesaplamaktadır.

NLMS (normalleştirilmiş en küçük hata kare ortalaması) gecikme değişimlerine en iyi tepki veren yöntemlerden biridir (DeLeon ve Sreenan 1999). NLMS ile beklenen ve gerçek gecikme farkının karelerinin toplamı en küçük olacak şekilde belirlenerek tahmin hatası en aza indirilmeye çalışılır.

NLMS yöntemine göre ağ gecikmesinin hesaplanması şekil 2.23'te yer almaktadır. NLMS'de son paketten önce alınan paketler FIR (sonlu dürtü tepkisi) filtreden

geçirilerek ağ gecikmesi tahmini yapılır. Daha sonra uyarlanabilir FIR filtrenin katsayılarını güncellemek için MSE (Ortalama Hata Kare) kullanılır.



Şekil 2.23 Uyarlanabilir tahminleyici adımı (DeLeon ve Sreenan 1999)

Uyarlanabilir NLMS filtrenin katsayılarının tahmini Eşitlik 2.98'e göre gerçekleştirilir.

$$\mathbf{h}_{i+1} = \mathbf{h}_i + \frac{\tilde{\mu}}{\mathbf{n}_i^T \mathbf{n}_i + a} \mathbf{n}_i e_i \quad (2.98)$$

Bu formülde $\mathbf{h}_i$ uyarlanabilir filtre katsayılarının $N \times 1$ vektörü, $\tilde{\mu}$ adım büyüklüğü ve $\mathbf{n}_i$ en son alınan ağ gecikmelerinin $N \times 1$ vektörü, T vektör transpozudur. e_i tahmin hatası olup Eşitlik 2.99 ile hesaplanır.

$$e_i = \mathbf{h}_i^T \mathbf{n}_i - \hat{n}_i \quad (2.99)$$

LMS algoritması aşağıdaki adımlara göre gerçekleştirilir (DeLeon ve Sreenan, 1999).

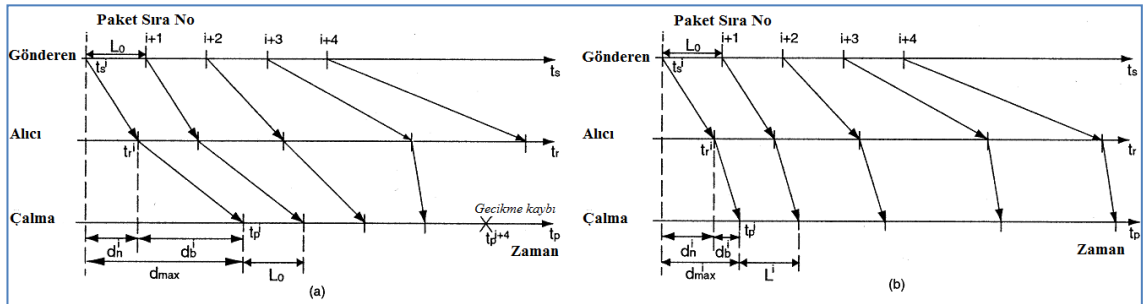
- Daha önce gelmiş olan N paketin ağ gecikmeleri NLMS algoritmasında kullanılarak gelecek olan paketin ağ gecikme tahmini yapılır. (Eşitlik 2.99'a göre)
- Eşitlik 2.95'e göre varyans hesaplanır.
- Eşitlik 2.96'da $\hat{d}_i$ yerine $\hat{n}_i$ konarak uçtan uca gecikme hesaplanır.

Shallwani ve Kabal, (2003) normalize edilmiş LMS (least mean square- en düşük ortalama kare) yöntemini kullanmış ve hızlı gecikme değişimlerini (spike) hızla ayarlayabilmek için spike tespit modu geliştirmiştir.

2.4.3 Uyarlanabilir çalma ve zaman skalası değişimi

Gerçek zamanlı sesler paketlenirken ardışık iki paketlerin ağa sürülme zamanlamaları arasındaki fark paketin içerdiği sesin süresi kadardır. Sabit kodeklerin kullanıldığı paketlerde bu süre sabittir. Örneğin internet ortamında paketlerin içerdiği ses uzunluğu olarak 20 ms sıkça kullanılan bir süredir. Bu durumda gönderen taraf her 20 ms’de bir sesi paketleyip ağa salmaktadır. Uçtan uca gecikmenin sabit tutulduğu sabit çalma zamanlamasında alınan her ses paketi zamanı geldiğinde kod çözümü yapılarak aynı süre olarak çalınmakta zamanında yetişmeyen paketler için o paketlerin ses süreleri kadar bir boşluk oluşmaktadır.

Uyarlanabilir çalma zamanlamasında uçtan uca gecikmeler dinamik olarak değişince tampon bellekte seslerin çalma sürelerinin yeniden uyarlanabilmesi gerekmektedir. Alıcı tarafta paket sürelerinin ayarlanması için zaman skalası değişimi yöntemleri kullanılmaktadır. Şekil 2.24’te sabit çalma zamanlamalı ve uyarlanabilir çalma zamanlamalı teknikler için paketlerin gönderme, alma ve çalma zamanları gösterilmektedir.



Şekil 2.24 Çalma zamanlamaları (Liang vd. 2003)

Şekil 2.24’te kullanılan notasyon aşağıdaki gibidir.

- t_s^i i. ses paketinin ağa gönderilme zamanı
- t_r^i i. ses paketinin alıcıya ulaşma zamanı
- t_p^i i. ses paketinin çalma zamanı
- d_n^i i. ses paketinin ağ gecikmesi
- d_b^i i. ses paketinin tampon bellekte bekleme süresi

- L_o Paketleme süresi
- L^i i. ses paketinin çalma süresi

Şekil 2.24.a’da tüm paketler için paketleme süresi ve bu paketlerin çalma süreleri eşittir ($t_s^{i+1} - t_s^i = t_p^{i+1} - t_p^i = L_o$). Uçtan uca gecikme sabit tutulup paketin ağ gecikmesi ile tampon bellekte bekleme süresinin toplamı belirlenen bu sabit değere ($t_p^i - t_s^i = d_n^i + d_b^i = d_{\max}$) eşittir.

Uyarlanabilir uçtan uca gecikmenin gösterildiği şekil 2.24.b’de paketleme süresi sabit olmasına rağmen alıcı tarafta bu paketlerin çalma süresi değişmektedir. Bu paketlerin çalma süreleri bir sonraki paket için belirlenen çalma zamanı ile o paketin çalma zamanı arasındaki fark kadardır.

$$L^i = t_p^{i+1} - t_p^i \quad (2.100)$$

L^i , uçtan uca gecikme bir önceki ile aynı ise L_o ile aynı olup, bu gecikme daha düşük ise L_o ’dan küçük daha yüksek ise L_o ’dan büyük olacaktır. Küçültme ve büyütme zaman skalasında sesin sıkıştırılması veya genişletilmesi ile gerçekleştirilmektedir.

Liu vd. (2001), uçtan uca gecikmenin belirlenmesinde AR model tabanlı tahminleme, zaman skalası değişimi için SOLA kullanmışlardır. Liang vd. (2003) ise uçtan uca gecikmenin belirlenmesinde histogram yöntemini, zaman skalası değişimi için ise WSOLA kullanmışlardır. Her iki çalışmada da AR model veya histogram modele göre belirlenen gecikme değerleri hedeflenmiş gerçek gecikme değerleri ise bu hedefe göre zaman skalası değişimi ile elde edilen süre olarak güncellenmiştir. Bir başka OLA tabanlı çalışma ise Gournay ve Anderson (2006) tarafından PSOLA kullanılarak gerçekleştirilmiştir. Rodbro ve Jensen (2002), zaman skalası değişimi olarak OLA teknikleri yerine sinüs modelleme kullanmışlardır.

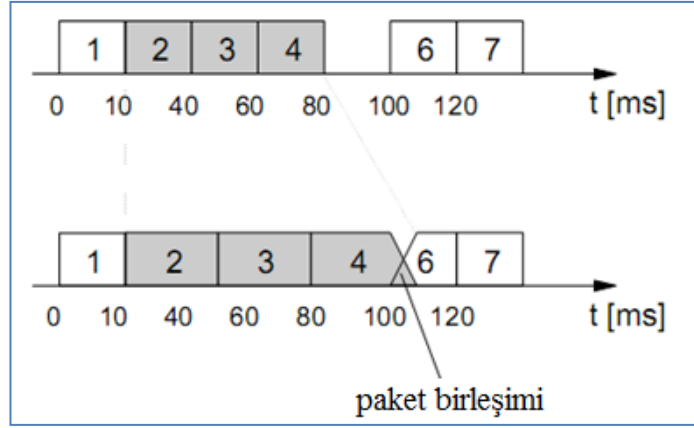
2.4.4 Paket kaybı saklama

Paket anahtarlamaalı ağlardan ses iletiminde alıcı tarafta uyarlanabilir alma tekniğinin kullanımı ge paket kayıplarını büyük oranda engellemektedir. Paket anahtarlamaalı ağlarda kayıplar ağda da oluşabilmektedir. Özellikle ağdaki yoğun hatlardan paketi yönlendirmeye alışan ağ noktalarının tamponlarının taşması bunun en büyük nedenidir.

Paket kayıplarının gizlenmesi için ilk kullanılan tekniklerden biri patern eşlemedir. Bu yöntemde kayıp olan pakete benzeyen paternler daha önceki paketlerin arasından seçilerek kayıp paket süresi kadar tekrarlanır (Goodman vd. 1986). Aynı paket paterni en az bir kez daha tekrarlandığından sistemde ekolara neden olabilmektedir. Bu ekoları gidermek için bir önceki paketten bir perde boyu kadar örnek alınarak bu prototip perdenin kayıp paket süresi kadar tekrarlanması gerçekleştirilmiştir (Goodman vd. 1986, Goodman vd. 1988). Bu durumda sentezlenen kayıp paketin bir önceki ve bir sonraki paket uyumsuzluğu seste ani “tık”lara neden olmaktadır. Kalitenin artırılması için Valenzuela ve Animalu (1989) bu perde paternlerinin eklenmesinde önceki ve sonraki paketlerle senkronizasyonu sağlayarak faz uyumunu da sağlamışlardır. Bununla birlikte seste mekanikleşmelerin olduğu belirtilmektedir (Sanneck vd 1996).

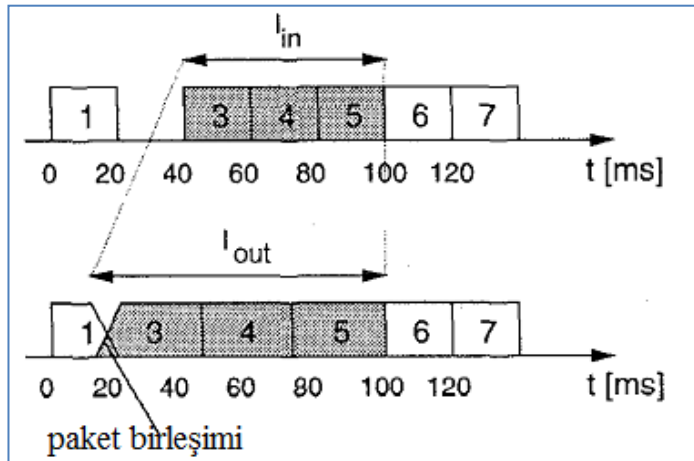
Zaman skalası değışimi algoritmaları da paket kayıplarının giderilmesi için en sık kullanılan tekniklerden biridir (Sanneck vd. 1996, Liu vd. 2001, Liang vd. 2003, Merazka 2008, Mojia vd. 2009).

Kayıp paketin sentezlenmesinde zaman skalası değışimi uygulanırken Stenger vd. (1996), kaybolan paketten önceki iki veya üç paketi WSOLA ile kayıp paketin süresini biraz geçecek şekilde uzatmış (1,25 ms), kayıp paketten sonra alınan ilk paketin ön kısmı ile örtüşmeli ekleme yaparak diğerk uçta oluşabilecek faz uyumsuzluğunu gidermeye alışmışlardır (şekil 2.25). Bu yöntem kayıp paketten önceki iki üç paketin de tampon bellekte beklemesine yol açtığından gecikmeyi artırmaktadır.



Şekil 2.25 Kayıp paketin önceki paketlerden sentezlenme gösterimi (Stenger vd. 1996)

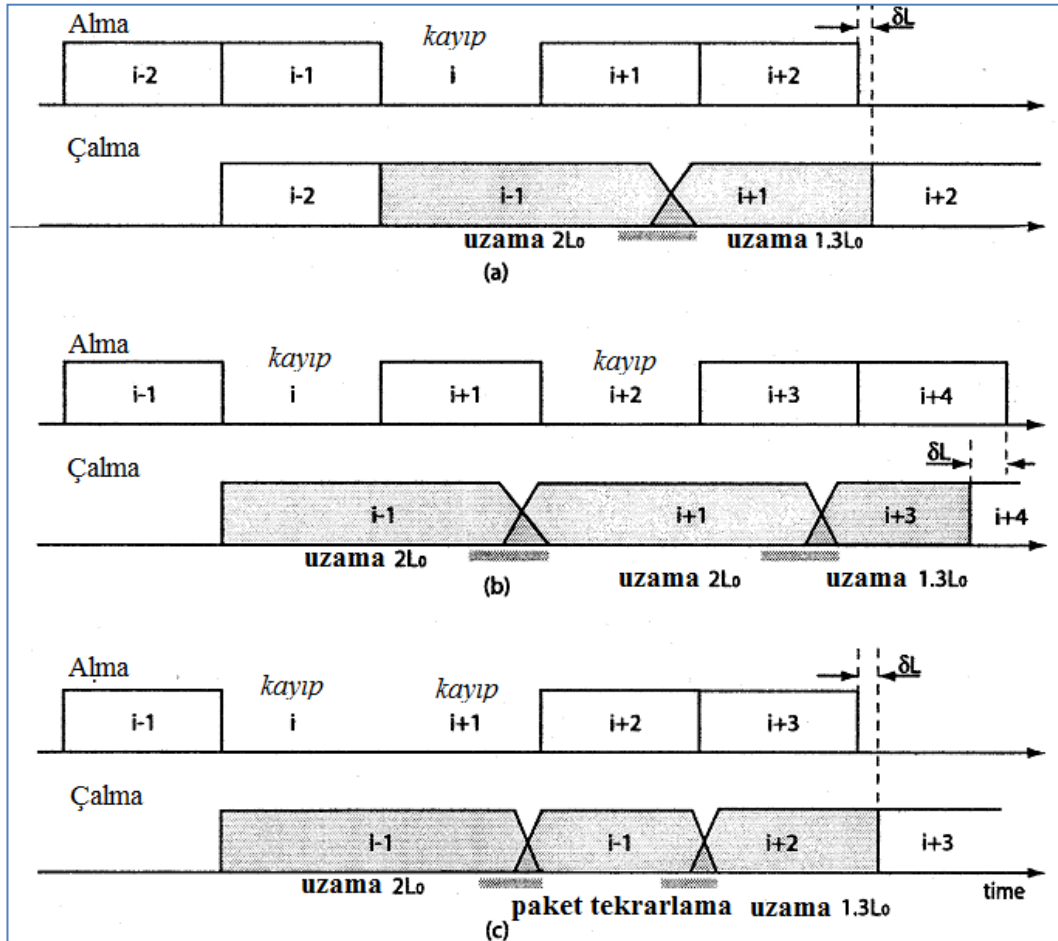
Sanneck vd. (1996) bu yöntemin aksine paket uzatmayı kayıp paketten sonra alınan paketlerde yapmışlardır. Bu yöntemde kayıp paketten önceki paketin son 2 ms'lik kısmı hariç aynen kullanılmaktadır (şekil 2.26). Kayıp paketten sonra alınan paketlerin uzatılması ile kayıp paket sentezlenmiş olur. Kayıp paketten önceki paket ile uyumsuzluğu azaltmak için de 2 ms'lik örtüşmeli ekleme kullanılmıştır. Her ne kadar bu yöntem kayıp paketten sonraki paketlerin uzamasını kullansa da örtüşmeli ekleme yapılabilmesi için bu paketlerin boyutunun da tampona eklenmesine ihtiyaç duymaktadır. Bu nedenle tampon gecikmesini düşürmediği iddia edilmektedir (Sanneck vd. 1996).



Şekil 2.26 Kayıp paketin sonraki paketlerden sentezlenme gösterimi (Sanneck vd. 1996)

Liang vd. (2003) tampon gecikmesini en aza indirmek için uyarlanabilir çalma zamanlaması kullandıklarından kayıp paket gecikmesini de düşürmek istemişlerdir. Bu nedenle kayıp paketi sentezlemek için bir önceki ve bir sonraki paketi kullanmışlardır. Uyarlanabilir çalma zamanlaması da kullandıklarından önceki ve sonraki paketin senkronizasyonunu en çok benzer kısımları örtüştürerek sağlamışlardır. Bu yöntemde kayıp paketin öncesindeki paketin uzunluğunun 2 katına, sonraki paketin uzunluğunun ise 1,3 katına çıkarılması hedeflenmiştir.

Hedeflenen ses uzaması ile gerçekleşen ses uzamaları aynı olmadığından, şekil 2.27'deki grafiklerde alınan paket ile çalınan paketlerin çalma zamanlamaları sonrasındaki süre farkları δL ile gösterilmektedir.



Şekil 2.27 Kayıp paketin en az tampon gecikmesi ile gizlenmesi (Liang vd. 2003)

a. Tek paket kaybı, b. Aralıklı iki paketin kaybı, c. Ardışık iki paketin kaybı

2.5 PESQ

Modern haberleşme ağları; düşük bit oranlı kodlayıcılar, kanal hataları/kayıpları, değişken gecikme ve filtreleme nedeniyle bozulmalara sahiptir. Bu ağlar üzerinden iletilen sesin algılama kalitesi, SNR (signal-to-noise ratio-sinyal gürültü oranı) gibi klasik metrikler ile değerlendirilememektedir. Kalite değerlendirmesinin güvenilir bir yolu, çok sayıda dinleyici örneklerinin kişisel kalite değerlendirme yaptığı testlerden elde edilen istatistiklere göre puanlama oluşturmaktır. Fakat bu öznel testler zaman almakta ve çok sayıda değerlendirme gerektirdiğinden yüksek maliyetli olmaktadır.

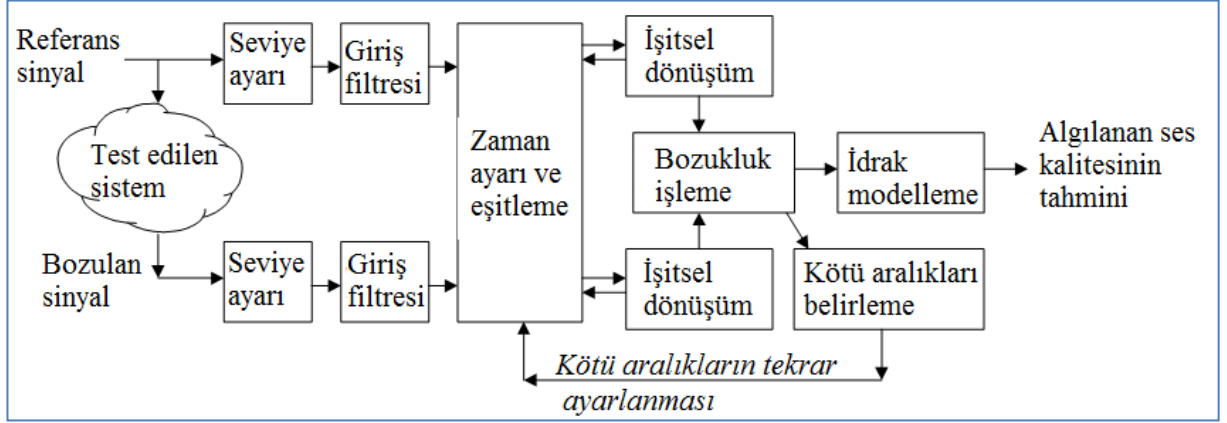
PESQ (Perceptual Evaluation of Speech Quality - Konuşma Kalitesinin Algısal Değerlendirmesi) ITU-T Recommendation P.862 ile tanımlanan konuşma ses kalitesinin ölçümü için nesnel bir yöntemdir. PESQ öznel dinleme testleri ile elde edilen MOS skor tahmini için geliştirilmiş olup; ses kodlayıcılar ve uçtan uca haberleşme kalite ölçümleri için kullanılmaktadır. PESQ tek yönde iletilen konuşmadaki ses bozulmalarını ve sesteki gürültüleri değerlendirmektedir (ITU-T P.862). Ses şiddeti seviyesindeki değişimleri ve çift yönlü konuşma iletişimine ait uçtan uca gecikme süresi gibi diğer algısal rahatsızlıkları puanlamamaktadır.

2.5.1 PESQ algoritması

PESQ; orijinal referans sinyali $X(t)$ 'yi, haberleşme kanallarından geçtikten sonra elde edilen $Y(t)$ bozulan sinyali ile karşılaştırmaktadır. Bu bozulma sesin kodlayıcı ile kodlanmasından, kanal gürültüsünden, gecikme ve gecikme değişiminden, paket kaybından veya kanalın doğrusal olmayan frekans tepkisinden kaynaklanabilmektedir.

Şekil 2.28'de PESQ algoritmasının basit model yapısı gösterilmektedir. Algoritma önce referans sinyal ve bozulan sinyalin algısal ses seviyelerini eş baza getirmektedir. Elektriksel sinyalden sesin tekrar üretilmesindeki algısal kayıpları bir filtre ile modelleyip iki ses arasındaki gecikmeleri de tahmin etmektedir. Psikoakustik bir modele göre işitsel dönüşümü modelleyen adımlar ile algılanabilecek bozuklukları

belirlemektedir. Bu bozuklukların etkilerinin puanlarını da toplayarak MOS tahminini hesaplamaktadır.



Şekil 2.28 PESQ modelinin yapısı

2.5.1.1 Seviye ayarı

Bozulan sinyal ile referans sinyalin ses seviyeleri aynı olmayabilmektedir. PESQ algoritmasında ilk adım; her iki sinyalin kulak referans noktasına göre algısal ses şiddeti 79dB SPL seviyesine getirilmesidir. Farklı frekans tonlarındaki seslerin şiddetsel algılanması aynı olmadığından seviye ayarı aşağıdaki adımlarla gerçekleştirilmektedir.

- Orijinal ve bozulan sinyal bir filtreden geçirilir. Bu filtre 250Hz'den küçük frekansları bloklamakta; 250-2000Hz arasındaki frekansları aynen geçirmekte; 2000 Hz'de 0dB, 2500 Hz'de -5dB, 3000Hz'de -10dB, 3150 Hz'de -20dB, 3500Hz'de -50 dB, 4000 Hz ve üstünde -500 dB olacak şekilde parçalı doğrusal olarak zayıflatmaktadır. Bu filtre çıkışları sadece kazanç hesaplaması için kullanılmaktadır.
- Bu filtre çıkışındaki ses örneklerinin karelerinin ortalaması bulunur.
- Bulunan ortalama değerlere göre kazanç katsayısı hesaplanır. Her iki ses için de ayrı hesaplanan bu kazanç katsayılarına göre orijinal ses $X(t)$ ve $Y(t)$ 'nin seviye ayarları yapılarak $X_s(t)$ ve $Y_s(t)$ elde edilir.

2.5.1.2 Giriş Filtresi

PESQ algoritması sesin dinlemesinin telefon avizesiyle yapılacağı varsayımıyla $X_S(t)$ ve $Y_S(t)$ sinyallerini bir filtreden geçirir. Telefon avizesini modelleyen bu filtre FFT ile gerçekleştirilerek $X_{IRSS}(t)$ ve $Y_{IRSS}(t)$ sinyalleri elde edilir.

2.5.1.3 Zaman ayarı

Bozulan sesin referans sese göre gecikmesi; tüm ses için ortalama gecikmenin hesaplandığı, sessiz periyotlar arasındaki konuşma periyotlarındaki gecikmelerin hesaplandığı ve konuşma periyotlarının alt çerçevelere de bölünerek anlık gecikmelerin hesaplandığı adımlarla hesaplanır. İşitsel modellemede bozuklukların büyük olması durumunda da yanlış belirlenen gecikmelerin düzenlenmesi için de gecikmeler tekrar hesaplanır.

- **Zarf tabanlı gecikme tahmini:** $X_{IRSS}(t)$ ve $Y_{IRSS}(t)$ sinyalleri 4ms'lik çerçevelere ayrılarak tüm çerçevelerin enerjilerinin eşik değere (aktif ses tespiti seviyesine göre) oranları ile k çerçeve numarası olmak üzere $X_{ES}(t)_k$ ve $Y_{ES}(t)_k$ zarfları hesaplanır. Bu zarfların çapraz ilintisinden bozulan sesin referans sese göre gecikmesi kabaca bulunur.
- **Sessiz periyotlardaki gecikme değişimi:** Aktif ses tespiti seviyesine göre sessiz bölge olarak tespit edilen ses kısımlarının 200ms'den fazla olduğu yerlerden sesler ayrılır. Her bir alt zaman aralığı için zarfların çapraz ilintisiyle bozulan sesteki gecikme değerleri 4ms hassasiyetle kabaca bulunur. Gecikme değerlerinin daha hassas bulunması için $X_{IRSS}(t)$ ve $Y_{IRSS}(t)$ seslerindeki 4ms gecikme hassasiyeti ile kabaca ilintilendirilen alt zaman aralıklarından %75 örtüşmeli hanning pencereli, 64 mslik çerçeveler alınır. Referans ses ile bozulan ses alt parçalarındaki her bir çerçeve için çapraz ilinti uygulanır. Çapraz ilintinin en yüksek seviyesi güven derecesiyle birlikte gecikmenin başlangıç ve bitiş noktalarını tespit eder. Her bir alt aralıktaki çerçevelerin gecikmeleri ve gecikme güven ölçümü değerlerinin histogramı çıkarılır. Alt

aralıkların histogramlarının, 1ms'lik simetrik üçgen pencere ile konvolüsyonları hesaplanır. Konvolüsyonun maksimum değeri kullanılarak her bir alt aralık için gecikme ve güven derecesi güncellenir.

- **Aktik konuşma periyotlarındaki gecikme değişimleri:** Konuşma periyotları sırasındaki gecikme değişimleri alt aralıkların bölünmesi ve her bir bölüm için gecikmelerinin tekrar düzenlenmesi ile gerçekleştirilir. Her bir bölümün gecikme tahmini için önce zarf tabanlı zaman ayarlamasından sonra çapraz ilinti ile hesaplanarak güven ölçümleriyle birlikte hesaplanır. Farklı bölümlere göre gecikme ve güven dereceleri bulunarak güven derecesinin en yüksek olduğu nokta bulunana kadar değişik bölümler işlemi tekrarlanır. Böylelikle tüm alt aralık bölümleri için gecikmeler hassas olarak hesaplanmış olur. Bölümleme yapılarak hesaplanan yeni güven derecelerinin bölümleme yapılmadan hesaplanmış olan alt aralık güven derecelerinden daha yüksek olması aranır. Bölümleme sayısı konuşma periyotundaki gecikme değişimi sayısı kadardır.

2.5.1.4 İşitsel dönüşüm

PESQ algoritmasındaki; işitsel dönüşüm; sinyalleri zaman ve frekansa bağlı işitsel algılama seviyesine göre ifade eden bir psikoakustik modeldir. İşitsel dönüşüm aşağıdaki aşamaları içermektedir.

- **Bark spektrum:** Hamming pencereli FFT ile 32ms süreli ve %50 örtüşmeli her bir çerçevenin anlık güç spektrumu hesaplanır. Bu güç spektrumundaki frekans değerleri değiştirilmiş Bark skalasındaki algısal frekanslar için eş aralıklı 42 kutuya gruplanır.
- **Frekans eşitleme:** Aktif konuşma çerçeveleri için Bark kutularının değerlerinin zamansal ortalaması ile ortalama Bark spektrumu hesaplanır. Referans sinyal ile bozulmuş sinyalin ortalama Bark spektrum oranlanmasıyla test edilen sistemin sabit frekans tepkisi tahmin edilir. Bu tahmine göre

referans ses çerçevelerindeki Bark kutularındaki değerler bu frekans tepkisi oranları ile çarpılarak bozulmuş sese göre eşitlenir. Bu eşitleme oranları her bir Bark kutusu için $\pm 20\text{dB}$ ile sınırlıdır.

- **Kazanç sapmalarını eşitleme:** Kazanç sapmalarını belirlemek için her bir referans sinyal çerçevesi ve ilgili bozulmuş sinyal çerçevesinin duyulabilir (sesli) güçlerinin oranları kullanılır. Bu oranlar birinci dereceden alçak geçirgen filtre ile filtrelenir, sınırlandırılır ve bozulmuş sinyal referansa göre eşitlenir.
- **Ses şiddeti eşitleme:** Frekansa bağlı eşik değerleri ve üsleri de kullanılarak Bark spektrumları algısal ses şiddetine göre eşitlenir. Böylece her bir zaman-frekans hücresi için algılanan ses şiddeti hesaplanmış olur.

2.5.1.5 Bozukluk işleme ve idrak modelleme

İşitsel dönüşüm ile kazanç sapmaları ve ses şiddetleri eşitlenen referans sinyal çerçevelerinin ilintili bozulmuş sinyal çerçeveleri ile Bark spektrumlarının mutlak farkı işitsel hata ölçümünü verir. PESQ skor tahmini yapılırken, gecikme değişimlerinden kaynaklanan sapma değişimleri (silme) ve algılanamayacak kadar küçük mutlak farklar (maskeleme) da dikkate alınır. Bazı kodlayıcılar seste oluşan yeni zaman-frekans bileşeninin modellenmesi durumunda daha fazla algısal bozukluk da meydana gelebildiğinden daha fazla bozukluk meydana getirebilmektedir. Bu yeni bileşenin hiç modellenmemesi algısal olarak daha başarılı sonuç verebildiğinden, PESQ asimetri faktörlü bozukluk parametresi de hesaplamaktadır. PESQ'in idrak modellemesinde kullanılan bu silme, maskeleme ve asimetri adımları aşağıdaki gibidir.

- **Silme:** Silme (negatif gecikme değişimi) bozulan seste örtüşmeli eklenen bir kısım oluşturmaktadır. Eğer bu gecikme değişimi yarım çerçeve uzunluğundan büyük ise örtüşmeli eklenen kısımlar atılır.

- **Maskleme:** Referans ve bozulmuş sinyalin zaman-frekans hücrelerindeki duyulan ses şiddetlerinin en küçüğünün dörtte biri maske eşiği olarak baz alınır. Her bir zaman-frekans hücresi için referans ve bozulmuş sinyal ses şiddetlerinin farkı bu eşiğe göre karşılaştırılır. Fark maske eşiğinden daha küçük ise maskelenir ve sıfır olarak alınır.
- **Asimetri:** PESQ algoritması asimetri faktörlü olan ve olmayan iki farklı hata ortalaması hesaplamaktadır. Asimetri faktörü her bir zaman-frekans hücresi için bozulmuş sinyal ile referans sinyalin Bark spektral yoğunluklarının istikrarlı oranlarından hesaplanmaktadır. Bu oranların 1,2 üzerinden üsleri alınır. Bu üssel değerler 12,0'dan büyük ise sınırlandırılır, 3,0'dan küçük ise 0,0 olarak değiştirilir. Asimetrik ağırlıklı bozukluk bu değerlerin çarpımıyla elde edilerek sadece eklemeli bozulmaların ölçümü yansıtılır.

İdrak modelleme ile çıkarılan Bark spektum mutlak farkları büyük olan çerçeveler için zaman ayarlama adımlarına geri dönülür ve kötü aralıkların tekrar ayarlanması sağlanır.

2.5.1.6 PESQ skorunu hesaplama

Frekans ve Zaman alanındaki asimetri faktörlü ve asimetri faktörsüz bozukluklar ayrı değerlendirilir. Herbir çerçeve için Bark spektrum mutlak farklarının frekans değerlerinin ortalaması asimetri faktörlü bozukluk değerleri için L1 norm, simetrik bozukluklar için L3 norm ile hesaplanır. Her çerçeve için elde edilen bu değerler kullanılarak ardışık 20 çerçeve (örtüşmeli eklemekten dolayı yaklaşık 320 ms) için hem asimetrik hem simetrik faktörler için L6 norm ile zamansal ortalama elde edilir. Elde edilen bu değerlerin her konuşma periyodu için L2 norm ortalaması ile de asimetrik ve simetrik bozukluk puanları hesaplanır. Asimetrik ve simetrik bozukluk puanları 4,5 sayısından çıkarılarak PESQ skoru elde edilir.

2.5.2 PESQ tahminlerinin kabul edilebilirliđi

ITU-T Study Group 12 alıřmalarında, PESQ'in MOS skorlarına gre kabul edilebilirliđinin dođrulaması geniř bir ses veritabanı ile yapılan testler ile yapılmıřtır. Bu testlerden elde edilen sonulara gre PESQ, izelge 2.1'de yer alan test faktrleri, kodekler ve uygulamalar iin kabul edilebilir dođrulukta tahminleme yapabilmektedir. PESQ'in MOS'a gre dođru tahmin yapamayacađı faktrler, kodlayıcılar ve uygulamalar izelge 2.2'de, tahmin dođruluđu henz ıspatlanmamıř olanlar ise izelge 2.3'te yer almaktadır.

izelge 2.1 PESQ'in kabul edilebilir dođruluđunun gsterildiđi faktrler, kodlayıcılar ve uygulamalar

TEST FAKTRLERİ
Kodlayıcı iin konuřma sesi giriř seviyesi
Haberleřme kanalı hataları
CELP kodlayıcılar ile Paket kaybı ve paket kaybı gizlemesi
Kodlayıcı farklı bit oranı kiplerini destekliyor ise bit oranları
Kodlayıcı deđiřimleri
Konuřan taraftaki evresel grlt
Dinleme testlerindeki gecikme deđiřimi etkileri
Ses sinyallerinin uzun-dnem zaman katlaması
Ses sinyallerinin kısa-dnem zaman katlaması
KODLAMA TEKNOLOJİLERİ
G.711, G.726, G.727 gibi dalgabiimi kodlayıcıları
G.728, G.729, G.723.1 gibi CELP ve hibrid kodlayıcılar
GSM-FR, GSM-HR, GSM-EFR, GSM-AMR, CDMA-EVRC, TDMA-ACELP, TDMA-VELP, TETRA gibi diđer kodekler
UYGULAMALAR
Kodlayıcı seimi veya deđerlendirmeleri
Analog veya sayısal ađ bađlantılarının canlı ađ testleri
Prototip ve benzetim ađların testi

Çizelge 2.2 PESQ'in tahminleme doğruluğuna güvenilemiyecek faktörler, kodlayıcılar ve uygulamalar

TEST FAKTÖRLERİ
Dinleme ses seviyeleri (PESQ seviyeleri 79 dB SPL'e normalize etmektedir)
Karşılıklı konuşmadaki gecikme etkileri
Konuşmacı ekoları
KODLAMA TEKNOLOJİLERİ
Aktif konuşma seslerinin %25'ten fazla sürekli kısımlarının sessizlik ile değiştirilmesi
UYGULAMALAR
Çift yönlü konuşma performansı

Çizelge 2.3 PESQ'in kabul edilebilir doğruluğunun henüz gösterilmediği faktörler, kodlayıcılar ve uygulamalar

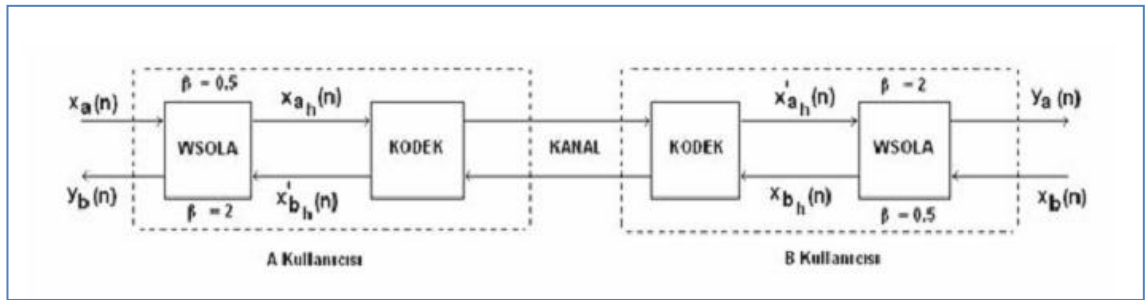
TEST FAKTÖRLERİ
PCM tipi kodlayıcılar ile paket kaybı ve paket kaybı gizlemesi
Konuşma zamanlarının kırılması
Konuşma genliğinin kırılması
Çok sayıda anlık konuşmacı
Kodlayıcılar için yapay konuşma sinyalleri
Kodlayıcıya girdi olarak müzik verilmesi
Alıcı ekosu
Eko bastırıcıların etkileri
Gürültü azaltma algoritmalarının etkileri
KODLAMA TEKNOLOJİLERİ
4 Kbps'den daha düşük oranlı CELP ve hibrid kodekler
MPEG4 HVXC
UYGULAMALAR
Akustik terminal/avize testi

3. MATERYAL VE YÖNTEM

3.1 TSM Algoritmalarıyla Sesi Sıkıştırılması

Bölüm 2’de literatürde yer alan çeşitli zaman alanı zaman skalası değişimi algoritmaları incelenmiştir. Zaman skalası değişimi sesi sıkıştırmak amacıyla kullanılacağından zaman skalası değişimi ile hızlandırılan sesin aynı oranda yavaşlatılarak orijinal skalasına geri dönülebilir olması önemlidir.

Zaman skalası değişiminin sesi sıkıştırılması için kullanımı daha önce Serinken vd. (1997) tarafından yarı çift yönlü (half duplex) bir HF kanaldan tam çift yönlü (full duplex) bir iletişim için önerilmiştir. Serinken vd. bu çalışmalarında zaman skalası değişim algoritması olarak WSOLA, kodek olarak da MELP kullanmıştır. Bu teknikte iletişim kanalının her iki ucundaki noktalara gelen giriş sesleri $\beta=0,5$ kullanarak yarı zamana düşecek şekilde hızlandırıldıktan sonra kodlanarak kanala verilmekte, alıcı noktada kod çözümü yapılarak ve $\beta=2$ kullanılarak orijinal zaman skalasına dönüştürülmektedir. Böylece birim zamanda $2x$ uzunluğunda veri iletilebilen yarı çift yönlü haberleşme kanalının her iki yöndeki kodlanmış ses girişleri x uzunluğuna düşürülerek tam zamanlı çift yönlü olarak da kullanılabilir.



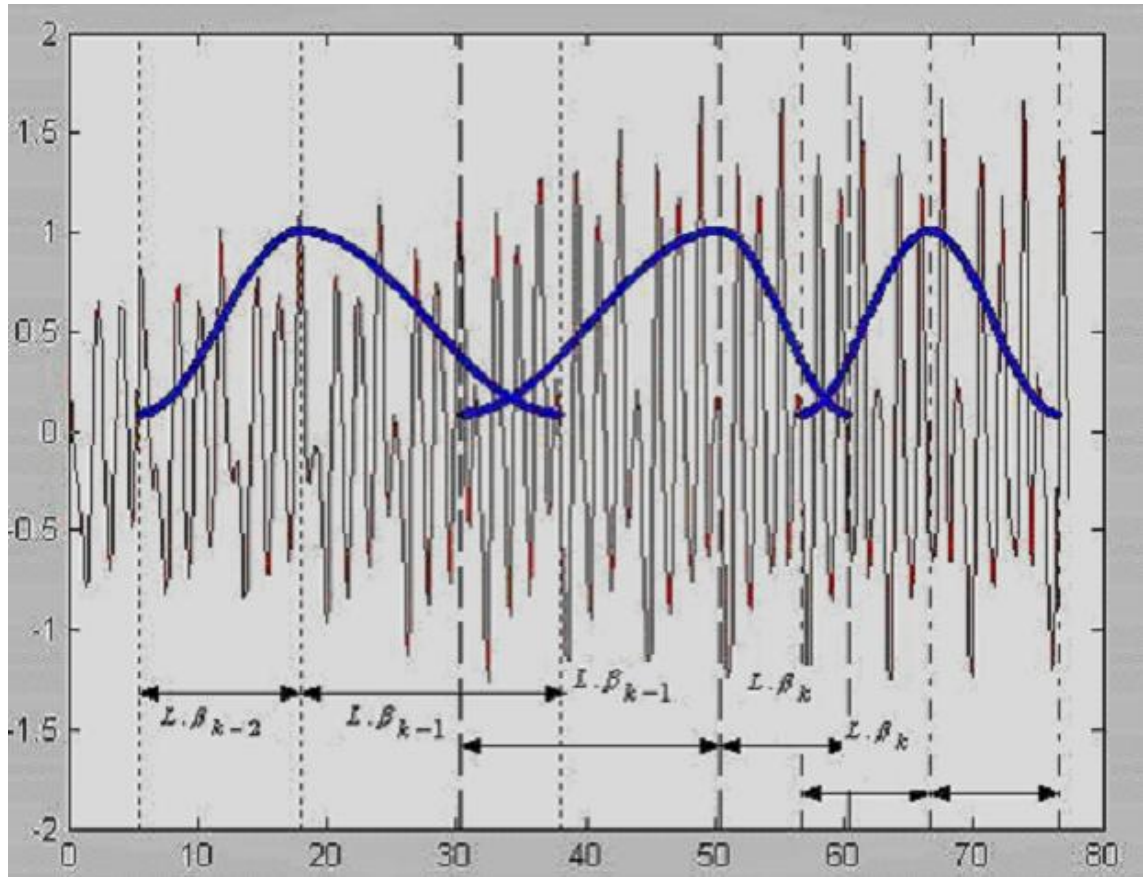
Şekil 3.1 Yarı zaman çift yönlü haberleşme kanalından tam zaman çift yönlü haberleşme için WSOLA’nın kullanım

Eroğul vd. (2001) şekil 3.1’de yer alan yapıyı sesin daha az bit ile sıkıştırılabilmesi için kullanmışlardır. Haberleşme kanalının kapasitesine göre kodek sıkıştırma oranının istenilen oranda düşürülebildiği bu yöntemde WSOLA, DTW (dinamik zaman katlama) ve PSOLA’dan daha iyi kalitede ses sentezleyebildiğinden tercih edilmiştir.

$$S_k = \sum_{m=1}^k \beta_m L \quad (3.2.a)$$

$$\tau^{-1}(S_k) = kL \quad (3.2.b)$$

Asimetrik pencerelerin yarı uzunlukları ise $\beta_k L$ olacaktır. Şekil 3.3’de yavaşlatma için analiz sesi benzerlikleri ve asimetrik pencereler görülmektedir.



Şekil 3.3 Uyarlanabilir WSOLA’da analiz sesindeki benzerlik segmentleri ve asimetrik pencereler

3.1.2 TSM algoritmalarının sentezleme kalitesinin karşılaştırma modeli

ITU-T P.800’de tanımlanan MOS (Mean Opinion Score, Ortalama Fikir Skoru) ses sıkıştırılmasında en çok tercih edilen subjektif kalite değerlendirme yöntemidir. MOS puanı çok sayıda insanın sentezlenen sese orijinal sesi baz alarak verdikleri öznel

puanlarının ortalaması olarak alınır. Anlaşılabilirliğin ölçüldüğü DRT (Dynamic Rhyme Test-Dinamik uyum testi) ITU-T P.800’de tanımlanan diğer bir öznel değerlendirmedir.

Öznel değerlendirmelere göre yapılan puanlama algoritma başarısının yanında seçilen deneklerin işitsel yeteneklerine de bağlıdır. Sentezlenen sesin MOS puanlaması yapılırken denek seçimi için kriterler olsa dahi özellikle yeni ses kodekleri ve benzer sentez seslerinin değerlendirmelerinde denek gruplarının farklı olması skorların tartışılabilir olmasına neden olmaktadır. Sentezlenen sesin kalitesinin nesnel değerlendirmesi için SNR ve segmentsel SNR gibi yöntemler de algısalıktan ziyade sinyalin biçimine baktığından MOS skorlarına göre; özellikle sesin kaynağının ve yolunun modellenmesine göre çalışan kodekler için kötü puanlama verebilmektedir. ITU-T P.862’de tanımlanan PESQ (Perceptual Evaluation of Speech Quality, Ses Kalitesinin Algısal Değerlendirmesi) insanların işitsel algısının modellenmesine göre çalışarak objektif bir skor oluşturmaktadır.

Geçmişte yapılan çalışmalarda; SOLA, WSOLA, PSOLA, faz vokoder, sinüs modelleme gibi TSM algoritmalarının ilk kalite değerlendirmeleri genelde öznel olarak gerçekleştirilmiştir. Bu değerlendirmelerde sesin belirli bölgelerinde algılanan eko, mekanikleşme gibi problemler de nitel olarak tanımlanmıştır. Algoritmaların sentezleme kalite başarısını vurgulamak için orijinal ve sentezlenen seslerin zaman ve spektrum grafikleri ile de desteklenmiştir. Kalite karşılaştırılmasına yönelik skorsal değerlendirmesine yönelik çalışmalar genelde 1990’ların ikinci yarısından sonra başlamıştır. Sanneck vd. (1996) kayıp paketleri sentezledikleri sesin değerlendirmesi için MOS kullanmışlardır. Eroğul ve Karagöz (1998) WSOLA ve sinüs modelleme ile Seneff(1982) tarafından önerilen TSM algoritmalarının kalite değerlendirmesi için MOS ve DRT testleri kullanmıştır.

Dorran vd. (2006) zaman skalası değişimi tekniklerinin nesnel kalite değerlendirmesi için benzerlik aramada kullanılan yöntemleri kullanarak toplam skora ulaşmışlardır. SOLA ve türevlerinin (AMDF’li, Normalleştirilmiş ilintili ve basit ilintili SOLA; EM-TSM; MEM-TSM; GLS-TSM) değerlendirildiği bu çalışmada orijinal ses ile sentezlenen ses segmentlerinin arasındaki benzerliklere göre puanlama yapılmıştır.

Değerlendirme benzerliklerinde kullanılan yöntem, örtüşmeli ekleme için benzerlik aramasında kullanılan yöntem ile aynı olması durumunda yanlış olacağı için 4 ayrı benzerlik yöntemi kullanılarak ortalaması alınmıştır. Bu benzerlik yöntemleri normalleştirilmiş ilinti, basitleştirilmiş ilinti, AMDF ve fark kare ortalamasıdır.

WSOLA ve uyarlanabilir WSOLA algoritmasının ses kodekleri ile kullanımının (şekil 3.2’de gösterilen yapı) kalite değerlendirmesinde DRT ve MOS skorları kullanılmıştır (İlk ve Güler 2006). Bu çalışmada orijinal zaman skalasına dönüldüğü için WSOLA algoritmasının hızlandırma ve yavaşlatma kalite başarımları birlikte değerlendirilebilmiştir (İlk ve Güler 2006).

TSM algoritmalarının ses sıkıştırma amacıyla kullanımında orijinal skalaya dönüldüğü için orijinale göre skorsal olarak kıyaslamak daha kolay olmakla birlikte MOS ve DRT gibi yöntemler çok fazla denek tarafından değerlendirme gerektirmekte, çok sayıda algoritmanın çok fazla örnek ile birbirlerine karşı kalite kıyaslaması zorlaşmaktadır. Bu nedenle bu çalışma kapsamında algoritmaların kalite değerlendirme yöntemi olarak aşağıdakiler kullanılmaktadır.

- Her algoritmanın hızlandırma ve yavaşlatma sentezlemesi için sinyal ve spektrum grafikleri orijinale göre sunulmaktadır. Bu grafiklerde sentez başarısı ve problemleri görülmektedir.
- Hızlandırma/yavaşlatma sentez döngüsü bir kere ve beş kere tekrarlandıktan sonra kalite değerlendirmeleri için ITU-T P.862 PESQ kullanılmaktadır.
- Beş tekrarda sentezleme kalitesi iyi olan yöntemler ile sentezlenen sesler dinlenerek PESQ sonuçlarının öznel değerlendirmesi yapılmaktadır.

3.1.3 Algoritmaların ses sıkıştırılması için uygulanması

TSM algoritmaları olarak OLA, SOLA, GLS-TSM, EM-TSM, WSOLA, PSOLA, AOLA, Suziki-Misaki yöntemi, PICOLA, faz vokoder, sinüs modelleme ve SMS kullanılmıştır. Faz vokoder, sinüs modelleme ve SMS algoritmaları, internet üzerinden hazır bulunan MATLAB kodların algoritmaları uyarlanarak, diğerleri ise ilgili

kaynaklarda yer alan algoritmalara göre MATLAB’da kodlanarak uygulanmıştır. Seslerin zaman ve frekans alanında görsel incelemeleri MATLAB’ın grafik gösterimleri ile bölgesel olarak da gerçekleştirilebilmektedir.

Bu algoritmalar genelde literatürdeki çalışmalara paralel uygulanmıştır. TD-PSOLA ve SMOLA için benzerlik arama tekniği değiştirilmiştir.

3.1.3.1 TD-PSOLA ile ses sıkıştırma

PSOLA algoritmasında en büyük problem perdelerin belirlenmesidir. Perdeler belirlendikten sonra örtüşmeli eklemeler Bölüm 2.1.5’te anlatıldığı gibi yapılmaktadır. Özilinti, AMDF veya frekans alanı teknikleri (FFT, Cepstral analiz) perde işaretlerini çok iyi bulmakla birlikte çok fazla işlem karmaşıklığı gerektirmektedir. Bu nedenle perde işaretleri tespit edilirken öncelikle yerel segmentlerin tepelerinden faydalanılmıştır. Hem erkek hem de bayan seslerinin perdelerinin bulunabilmesi için yaklaşık 35 ms’lik segmentin en büyük tepe noktası bulunduktan sonra, bu tepe noktasının 3 ms ile 1,4 ms komşuluklarında bu tepenin genliğinin en az %80 değerine sahip noktalar aranmıştır. Bu şekilde ötümlü bölgelerin perdeleri büyük oranda bulunabilmekle birlikte ötümsüz bölgelerde problem yaşanmaktadır. Moulines ve Charpentier (1988), ötümsüz bölgelerin gürültü özellikleri taşıdığından benzerlik aranmadan eşit uzunlukla işaretlenebileceğini belirtmişlerdir. Bununla birlikte örtüşmeli eklemelerden kaynaklanan sönümlenmenin en aza indirilmesi için perdelerin işaretlenmesinde sıkıntı yaşanan bölgeler için özilintiden de faydalanılmıştır. Bu algoritma aşağıdaki gibi uygulanmıştır:

- Yerel tepelerden perde işaretleme tahminlemesi yapıldı.
- Bulunan işaretlere göre perde peryodlarından medyan noktası bulundu. Bu medyan perde uzunluğunun 1,3 katından daha büyük perde uzunlukları medyan perde uzunluğuna bölünerek aday işaretleme noktaları bulundu.
- Bu aday işaretleme noktalarının 10 örneklem komşuluklarına göre özilintiye bakılarak işaretlemeler tamamlandı.

3.1.3.2 Suziki-Misaki yöntemi ile ses sıkıştırma

Suziki Misaki Yöntemi tekniği uygulamasında segment büyüklüğü T_s 12 ms benzerlik arama aralığı ise -10 ms ve 10 ms seçilmiştir. Hızlandırma için istenen sıkıştırma $0,5$ alınmıştır. Yavaşlatma için ise istenen genişletme 2 alınmıştır.

Suziki ve Misaki benzerlik aramada Eşitlik 2,29'da yer alan özilinti eşitliğini kullanmıştır. Bununla birlikte örtüşmeli ekleme yapılırken maksimum özilinti hesaplamasında kullanılan tüm örnekler örtüştürülmemiştir. Farklı τ değerleri için özilinti örneklem çarpım sayıları da değişmektedir. Bu nedenle benzerlik aramada sadece τ değerine göre örtüşecek örneklemeler için özilinti hesaplanmış, bu özilinti de normalleştirilmiştir.

3.1.3.3 Diğer TSM teknikleri ile ses sıkıştırma

OLA tekniğinin uygulanmasında segment büyüklüğü 30 ms alınmıştır. Hızlandırma için $\alpha = 0,5$, analiz aralığı S_a 15 ms, sentez aralığı S_s ise $7,5$ ms alınmıştır. Yavaşlatma için ise $\alpha = 2$, analiz aralığı S_a $7,5$ ms, sentez aralığı S_s ise 15 ms alınarak ses orijinal zaman skalasına getirilmiştir.

GLS-TSM ve EM-TSM; SOLA tabanlı algoritmalar olup sadece benzerlik arama tekniğinde değişiklik içermektedir. Bu nedenle bu üç TSM algoritması için Wong vd. (2002) çalışması baz alınarak aynı parametreler kullanılmıştır. Segment büyüklüğü 30 ms alınmıştır. Hızlandırma için $\alpha = 0,5$, analiz aralığı S_a 15 ms, sentez aralığı S_s ise $7,5$ ms alınmıştır. Yavaşlatma için ise $\alpha = 2$, analiz aralığı S_a $7,5$ ms, sentez aralığı S_s ise 15 ms alınarak ses orijinal zaman skalasına getirilmiştir. Benzerlikler $-7,5$ ms ile $7,5$ ms arasında aranmıştır. Fakat örtüşme uzunluğu L_k $3,75$ ms'den ($N/4$) daha düşük olmaması için sınırlandırılmıştır.

WSOLA kodlamasında segment büyüklüğü 20 ms seçilmiş, benzerlik aramaları 10 ms'lik ses parçalarıyla yapılmıştır. Analiz sesinde ardışık segmentler arası uzaklık

$S_a = S_s / \alpha$ olduğundan ($S_s = N/2 = 10$ ms) hızlandırma için 20 ms yavaşlatma için ise 5 ms alınmıştır. Benzerlikler ise -5 ms ile 5 ms sapma aralığında aranmıştır.

AOLA algoritması benzerliği sinyalin tepe noktalarına bakarak bulmaktadır. Orijinal TD-PSOLA algoritmasında da kullanılan bu yöntem benzerliğin yanlış bulunabilmesi nedeniyle kalite bozulmasına neden olabilmektedir. Bu nedenle benzerlik AMDF ile aranmıştır. AOLA kodlamasında segment büyüklüğü en düşük frekanstan en az iki periyot içerecek şekilde 20 ms olarak kullanılmıştır. Hızlandırma için istenen sıkıştırma $dc=0,5$ yavaşlatma için ise istenen genişletme $de=2$ alınmıştır.

PICOLA kodlamasında P_{min} ve P_{max} standartta bahsedilen değerlere göre 8 KHz için sırasıyla 32 ve 160 örneklem, 16 KHz için de sırasıyla 80 ve 160 örneklem seçilmiştir.

Faz vokoder hesaplamasında FFT ve pencere uzunluğu 512, ardışık FFT segmentleri arası uzaklık 128 örneklem seçilmiştir. Ardışık analiz segmentlerin frekans haznelerindeki genlik değerleri sentezlenen segmente yakınlığı (skalasal) oranda ağırlıklar ile çarpılarak eklenmiştir. Tüm frekans hazneleri için ilk segment hazneleri sıfır olarak alınmış, daha sonra güncellemeler Eşitlik 2.59 - 2.61'e göre gerçekleştirilmiştir.

Sinüs modelleme için McAuley ve Quatieri (1986) ile Smith ve Serra (1987) tarafından önerilen her iki model de uygulanmıştır. Her iki yöntemde de analiz aşamasında pencere uzunluğu ve FFT boyutu 512, ardışık segmentler arası uzaklık 128 örneklem seçilmiştir. Sesin algılanmasında faz önemli olmadığından, Smith ve Serra (1987) tarafından önerilen yöntemin uygulanmasında sadece frekans genlikleri kullanılmış, gerçek sinüs tepesi tahminlemesi için parabolik enterpolasyon kullanılmıştır. Sinüs yollarının başlangıç (doğum) fazları sıfır alınmıştır. McAuley ve Quatieri (1986) tarafından önerilen yöntemde faz bilgileri de kullanıldığından sinüs tepelerinin ardışık komşularından enterpolasyonu uygulanmamıştır. Sentez aşamasında ardışık segmentler arasında genlik ve frekans enterpolasyonu yanında, faz için kübik enterpolasyon da uygulanmıştır. Sinüs yollarının başlangıç fazları karmaşık FFT değerlerinin tanjantı ile belirlenmektedir.

Spektral Modelleme Sentezi için yapılan testlerde MQ sinüs modelleme kodlarından faydalanılmıştır. Sinüs modelleme analizinden çıkan parametreler ile orijinal skala ile sentezlenen ses, orijinal sesten çıkarılarak kalan sinyalin LPC parametreleri bulunmuştur. Daha sonra sesi yavaşlatmak için Sinüs modelleme analiz parametreleri ve LPC parametreleri ile daha geniş pencereler kullanılmıştır. Testin ikinci aşamasında yavaşlatılan ses aynı yöntemle hızlandırılarak orijinal zaman skalasına dönmüştür.

3.2 Veri Kaybı Gizlemesinde Gecikmenin Azaltılması

Bölüm 2.4'te TSM algoritmalarının uyarlanabilir çalma zamanlaması ve paket kaybı gizlemesi için kullanımına yönelik literatürdeki çalışmalar anlatılmıştır. Literatürdeki mevcut çalışmalarda TSM algoritması olarak genelde WSOLA ve SOLA kullanılmaktadır. Tekrarlı STFT, faz vokoderi vb. frekans alanı TSM teknikleri ile sinyal modelleme TSM teknikleri işlem karmaşıklığı yüzünden genelde az tercih edilmektedir.

Uyarlanabilir çalma zamanlaması ve paket kaybı gizlemesi için TSM algoritmaları olarak WSOLA, SOLA ve tek pakette uyarlama yapabilmek için alternatif olarak önerilen yeni bir örtüşmeli ekleme yöntemi kullanılmıştır. WSOLA ve SOLA'nın uygulanmasında uçtan uca gecikmenin düşürülmesi için Bölüm 2.4'te anlatılan literatürdeki çalışmalara göre farklılıklar da bulunmaktadır. Tüm algoritmalar Bölüm 3.2.1'de belirtilen limitlere uygun olarak uygulanmıştır.

3.2.1 Paket kaybı gizlemesi için örtüşmeli ekleme limitleri

Temel frekansı f_{min} ile f_{max} arasında değişebilen hemen hemen periyodik bir ses sinyali olsun. Herhangi bir t anında temel periyot, $1/f_{min} > P(t) > 1/f_{max}$ olacaktır. Zaman alanı TSM algoritmaları ile bu sinyal üzerinden alınan bir segmentin en benzer komşu segmenti özilinti veya diğer yöntemler ile aranacak olursa benzerlik arama uzaklığı Eşitlik 3.3'teki aralığı içeriyor ise en benzer segment bulunabilecektir.

$$[1/f_{max}, 1/f_{min}] \quad (3.3)$$

Herhangi bir pozitif tamsayı K için; $f_{max} > K.f_{min}$ ise arama alanı $[1/(K.f_{min}), 1/f_{min}]$ aralığını kapsayacak şekilde daraltılabilir. Bu durumda $1/(K.f_{min}) > P(t) > 1/f_{max}$ eşitsizliğini sağlayan yüksek frekanslar için benzerlik K periyot uzaklıkta bulunabilecektir. Hemen hemen periyodik bir sinyal için bu uzaklık $P(t)+P(t+P(t))+P(t+P(t+P(t)))+....$ olacaktır.

İnsan sesinin perde periyodu erkekler için 100-146 Hz bayanlar için 185-221 Hz arasında değişebilmektedir (Gelfer ve Mikos 2005). Bu nedenle bayan sesleri için iki perde uzaklıktaki en benzer noktasının da kabul edilebilir olması durumunda arama uzaklık aralığının 5 ms ile 10 ms arasında olması yeterli olacak, 201-221 Hz temel frekanslar için iki periyot uzaklıkta; diğerleri için ise bir periyot uzaklıkta en benzer segment bulunabilecektir. Yüksek frekanslar için bir periyot uzaklıktaki segmentin benzerliğine de bakılmak istenirse arama uzaklığı 4,5 ms ile 10 ms arasını kapsaması yeterli olacaktır.

Benzerlik aramada segment büyüklüğü Δt olarak seçilirse tek bir sinüs bileşeni içeren bir sinyal için Δt sıfıra yakınsarken arama aralığının Eşitlik 3.3'te yer alan aralığı içermesi durumunda benzerlik (eğim ve örneklem işaretinden) doğru olarak bulunabilecektir. Ayırık zaman sinüs sinyalinde 2 adet örneklem yeterli olabilecektir. Normal bir ses sinyalinde birden fazla sinüs bileşeni bulunmakta ve tam periyotluk sağlanamamaktadır. Bu nedenle benzerlik aranacak segment büyüklüğünün en az bir periyoda yakın olması bulunacak benzerliği garanti edecektir. Yapılan deneylerde bir periyottan daha düşük segmentler ile de benzerliğin başarılı olarak bulunabildiği görülse de buna yönelik bir model kurulamadığından ve farklı ses prototipleri ile desteklenemediğinden alt sınır belirlenememiştir.

Uyarlanabilir çalma zamanlaması ve paket kaybı gizlemesi için algoritmaların kullanımında arama aralığı ve segment büyüklüğü limitlerine uyulması algoritmaların başarılı olarak sentezlemesini sağlayabilecektir. Uyarlanabilir çalma zamanlamasında ağ gecikmesinin azalması nedeniyle hızlandırma yapmak (uçtan uca gecikmenin veya çalma zamanlaması gecikmesinin düşürülmesi) için sesin bir periyotluk düşüşü ile bu sağlanabilir. Bu nedenle ağdaki gecikmesindeki düşüşün bir periyottan küçük olması

durumunda uçtan uca gecikme düşürölmek istenir ve düşüş eğer sesin ötümlü bölgesindeki benzerlik ile sağlanır ise ileride tekrar genişletme ihtiyacı doğabilecektir. Segmentin uçtan uca genişletilmesi veya paket kaybının gizlenmesi için yapılacak yavaşlatmada da artış bu oranda olacaktır.

Diğer yandan özellikle tek paketlik ses ile paket kaybının gizlenebilmesi için yapılacak örtüşmeli eklemelerin paketteki ses segmentinin boyutunu ikiye katlaması hedeflenmektedir. TSM ile yavaşlatmada aynı segmentlerin örtüşmeli ekleme yapması veya tekrar etmesi eko yapma gibi etkilere neden olabilmektedir. Yiang vd. (2003) bu problemi aşmak için WSOLA'yı tekrarlı olarak kullanmışlardır. Arama aralığının minimum uzaklığı $1/(K \cdot f_{min})$ veya $1/f_{min}$ olabileceğinden benzerlik aramanın yapıldığı segmentten önceki sesler çalınmış olsa dahi geriye doğru yapılacak arama ile yavaşlatma için ilk örtüşmeli ekleme segmenti gerçekleştirilebilir. Bu durum tek pakette yapılacak örtüşmeli ekleme için ihtiyaç duyulacak genişletmenin daha aşağıya düşmesini sağlayacaktır.

3.2.2 WSOLA ile paket kaybının gizlenmesinde gecikmenin azaltılması

WSOLA uyarlanabilir çalma zamanlaması ve paket kaybının gizlenmesi için literatürde en yaygın kullanılan tekniktir. Sanneck vd. (1996) tarafından paket kaybının gizlenmesi için önerilen kayıptan önceki son üç paketin veya kayıptan sonraki ilk üç paketin genişletilmesi yöntemi gecikmeyi en az 60ms daha fazla artırdığı için tercih edilmemiştir. Bunun yerine Liang vd. (2003) tarafından önerilen ve uyarlanabilir çalma zamanlamasının da kullanıldığı sadece bir paketlik gecikmeye ihtiyaç duyan yöntemin kullanılması tercih edilmiştir. Bununla birlikte Liang vd. (2003) tarafından önerilen modelde paket genişletme son paketin başında ve bir önceki paket de kullanılarak tekrarlı WSOLA ile gerçekleştirilmiştir. Bu nedenle sentezlenen paket istenilen genişliğe ulaştıktan sonra çalınabilmekte bu da uçtan uca gecikmeyi bir paketteki ses süresine ek olarak algoritmanın çalışma süresi kadar uzatmaktadır. Genişletme Bölüm 3.2.1'de anlatıldığı gibi paket öncesini de kullanarak yapılırsa ilk örtüşmeli ekleme ile bir kısım ses çalınabilir hale gelir ve gecikme düşürölebilir. Ayrıca; tekrarlı WSOLA'nın her tekrarında örtüşmeli ekleme segment sayısı artırılmaktadır. Tüm

geniřleme sentezleri iin hep aynı rneklemeler ile sentez yapılması olasılığı da dřrlmektedir.

3.2.3 SOLA

WSOLA, PSOLA, PICOLA vb. birok rtřmeli ekleme teknikleri, ekleme noktalarını analiz ařamasında belirlemekte sentez ařamasında ise sadece ağırlık noktalarıyla arpıp sesi retmektedir. SOLA ve EM-SOLA, GLS-TSM gibi SOLA trevlerinin diğerk OLA tekniklerinden (AOLA da sentez ařamasında) en temel farkı ekleme noktalarını sentez ařamasında belirlemesidir. Analiz ařamasında ardışık segmentlerin birbirlerinden sabit uzaklık aralıklar ile belirlenebilmesi tek bir pakette zaman skalası değıřimi iin algoritma kolaylığı ve esnekliğı saėlamaktadır.

SOLA'nın uygulanmasında benzerlik arama sentezde yapıldığından rtřmeli ekleme yapılabilecek tm segmentler bařtan aralıklar ile belirlenebilmiřtir. Bu sayede zellikle geniřlemeler kolaylıkla saėlanabilmiřtir.

SOLA tekniğinin uygulanmasında paketten seilen ardışık segmentler arası uzaklık geniřletme iin 0,2 ms, daraltma iin ise 10 ms olarak seilmiřtir. Segment byklkleri de yine geniřletme iin 12 ms daraltma iin ise 10 ms seilmiřtir. Geniřletme iin paketin ncesinde ve sonrasında paket kaybı olup olmadığına da dikkat edilmiřtir. Paket kaybından sonra ucuna ekleme iin yapılan geniřlemede segmentlerin benzerlik araması ve rtřmesi paketin bařından sonuna doėru ilerlemektedir. Diğerk durumlarda benzerlik araması paketin sonundan bařına doėru yapılmaktadır.

3.2.4 Paket ierisinde rtřmeli ekleme

Algoritma Laroch yntemindeki gibi paketin kendi ierisinde benzerlik aramaktadır. Bununla birlikte Laroch ynteminde rtřen tm rneklemeler benzerlik zilintisinde kullanılmakta ve zilintide kullanılan rneklem sayısı zilinti sapması τ deėerine gre azalmaktadır. Bu algorithmada benzerlik arama boyutu WSOLA'daki gibi sabit tutulmuřtur. Bununla birlikte WSOLA'da sentez sesinin tamamı rtřmeli eklemelelerden

oluşmaktadır. Bu algoritmada SOLA'daki gibi örtüşmeli eklemesiz kısımlar da bulunabilmektedir.

Benzerlik arama ve örtüşmeli ekleme boyutu olarak 8 ms seçilmektedir. Aramada kullanılacak şablonlar için 8 ms'lik 2 ayrı segment belirlenmektedir. Aramanın ilk aşamasında; ilk segment paketin 4. ms'sinde sabittir. İkinci segment 8. ms'den 12. ms'ye kadar hareketlendirilerek özilintiye bakılır. Daha sonra ise 2. segment 12. ms'de sabit tutulurken ilk segment paketin başı ile 4. ms arasında hareket ettirilmekte ve özilintiye bakılmaktadır. Benzerlik aranmasında iki segmentin noktaları birbirlerinden en az 4 ms en çok 12 ms uzaktadır. Bu sayede Bölüm 3.2.1'de belirtilen limitlere uyulmaktadır. Daha sonra özilintisi en büyük olan kayma sapmaları için örtüşmeli ekleme yapılmaktadır. Örtüşmeli eklemenin sağ ve solu SMOLA'daki gibi doldurulur. Bununla birlikte hem hızlandırma hem de yavaşlatma için ilk segmentin ağırlığı hamming penceresinin monoton azalan kısmı seçilir.

Yavaşlatma ve hızlandırma için hedef boyutun 2,5 ms (paket süresinin sekizde biri) komşuluğunda bir değere ulaşmak yeterli olarak ele alınmıştır. Özellikle paket kaybının da gizlenmesi istenen durumlarda hedef boyut ses paketi boyutunun en az 2 katı olmaktadır. Bu durumda bir örtüşmeli ekleme yetmeyecektir. Bu nedenle 2. benzerlik araması için Bölüm 3.2.2'deki gibi önceki paketin son kısımları kullanılır. Daha önceki paketin son kısımları ile de örtüşmeli ekleme ve sentez gerçekleştirildikten sonra paket genişletmenin yetersiz kalması durumunda WSOLA'daki gibi aynı adımlar tekrarlanır.

3.3 DCT ile Zaman Skalası Değişimi

Sinüs modellemeye ses sinyali baskın sinüs bileşenlerin toplamı olarak, faz vokoderlerinde ise Fourier genlik spektrumu ile modellenmektedir. Bu yöntemler ile zaman skalası değişiminde sentez sinyalinin kalitesi (yankı, mekanikleşme veya benzeri bozulmalar) ardışık çerçeveler arasında fazların uyumlandırılmasının başarısına bağlıdır.

DCT dönüşümü, sinyali tamamen gerçel sayılar ile ifade edilebilen bir spektral bölgede ifade etmektedir. Bu sayede fazın uyumlandırılması için kübik spline enterpolasyonu ve faz katlama katsayısının bulunması gibi karmaşık hesaplamalara gerek kalmamaktadır. DCT dönüşümünün diğer bir avantajı ise enerji paketleme verimliliğidir. DCT ortogonal bir dönüşüm olup özellikle ses sinyalleri için spektrum başlangıç noktasından uzaklaştıkça azalmaktadır. Bu özelliği sayesinde DCT katsayılarının kırılması ile çözünürlüğünün azalması durumunda enerji kaybı sinüs modelleme veya Fourier dönüşümüne göre daha azdır.

3.3.1 DCT ve değiştirilmiş ters DCT dönüşümü

DCT dönüşümü Eşitlik 3.4 - 3.5'te ifade edilmektedir. DCT'nin ters dönüşümü ise $n=0,1,2,...,N-1$ olmak üzere Eşitlik 3.6'da ifade edilmektedir.

$$S[k] = DCT\{s[n]\} = \frac{2}{N} f_k \sum_{n=0}^{N-1} s[n] \cos\left(\frac{(2n+1)k\pi}{2N}\right) \quad (3.4)$$

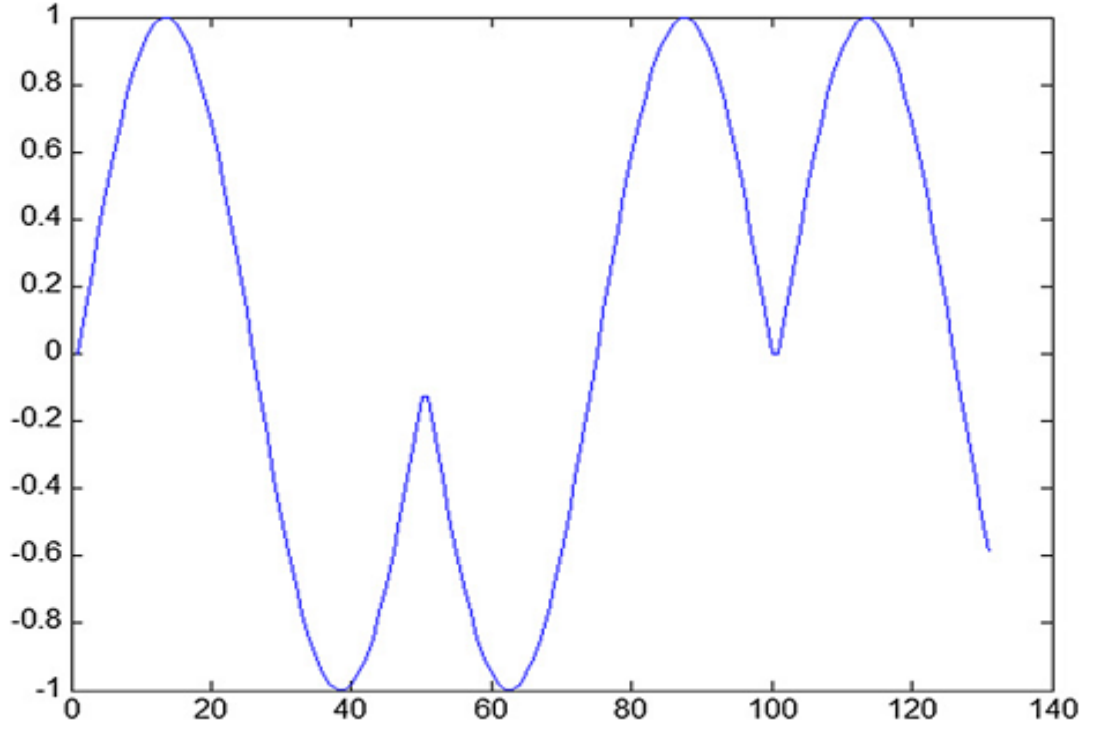
$$f_k = \begin{cases} \frac{1}{\sqrt{2}} & k = 0 \\ 1 & k = 1, 2, \dots, N-1 \end{cases} \quad (3.5)$$

$$s[n] = IDCT\{S[k]\} = \sum_{k=0}^{N-1} f_k S[k] \cos\left(\frac{(2n+1)k\pi}{2N}\right) \quad (3.6)$$

$s[n]$ sinyali bant sınırlı ve periyodik olması durumunda $s[N]=s[0]$, $s[N+1]=s[1]$,... olacaktır. Eşitlik 3.5 ile bir periyottan daha uzun bir sinyal sentezlenmek istenirse l ve m herhangi bir tam sayı olmak üzere $n=m.N+l$ için kosinüs terimi Eşitlik 3.7'deki gibi olacaktır.

$$\cos\left(\frac{(2(mN+l)+1)k\pi}{2N}\right) = \cos\left(\frac{2Nm k\pi}{2N} + \frac{(2l+1)k\pi}{2N}\right) = \cos\left(mk\pi + \frac{(2l+1)k\pi}{2N}\right) \quad (3.7)$$

Kosinüs işlemi π 'ye göre aynasal simetrik olduğundan sentezlenen örnekler, $m.k$ 'nın alacağı tek değerlerde $n=l$ 'nin kosinüs değerinin negatifi, mk 'nın alacağı çift değerlerde $n=l$ 'nin kosinüs değerinin aynısı olur. Örnek bir kosinüs sinyalinin Eşitlik 3.7'ye göre ters DCT dönüşümü ile birden fazla periyot için sentezlenmesi şekil 3.4'te gösterilmektedir.



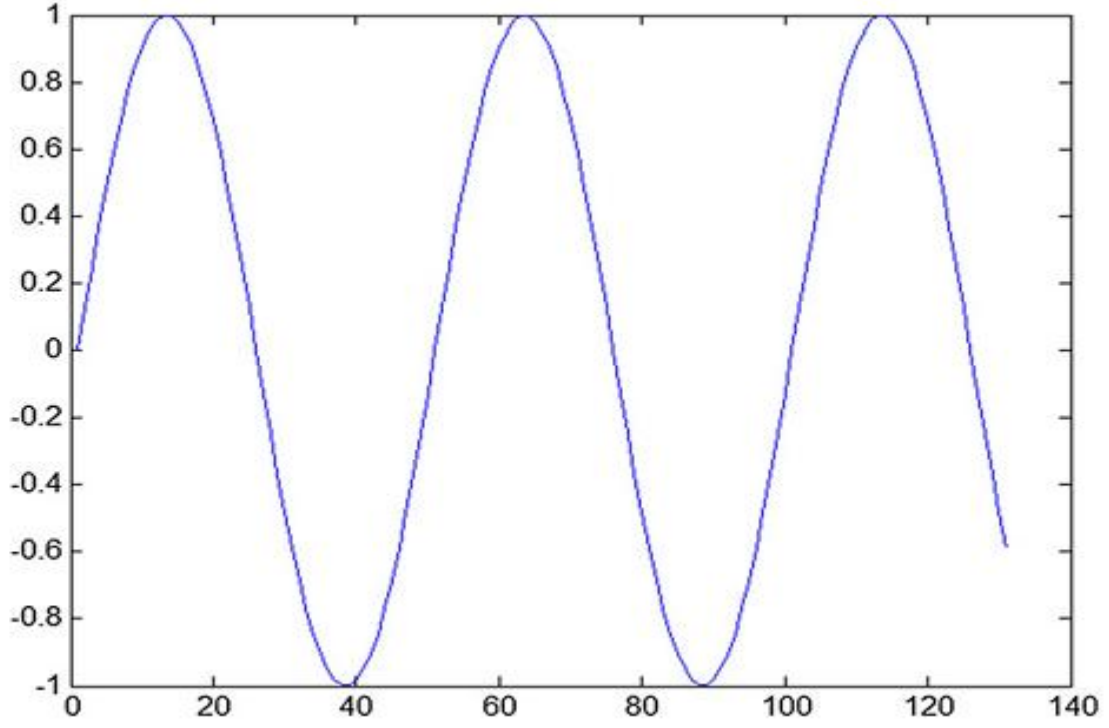
Şekil 3.4 DCT katsayılarından ters dönüşüm ile sentezlenen periyodik sinyalin aynasal simetri özelliği (İlk ve Güler 2011)

Ters DCT dönüşümünün aynasal simetrisi periyodik sinyal sentezlemesini engellemektedir. Aynasal simetriden kurtulmak için Eşitlik 3.8 aşağıdaki gibi güncellenir (İlk 1997).

$$s[n] = \sum_{k=0}^{N-1} f_k u^k S[k] \cos\left(\frac{(2n+1)k\pi}{2N}\right) \quad (3.8)$$

Birinci, üçüncü, beşinci vb. tek perdeler için $u=1$; ikinci, dördüncü, altıncı vb. perdeler için $u=-1$ olacaktır. Eşitlik 3.4'e göre DCT'si alınan bir periyotluk sinyalden Eşitlik 3.8

ile IDCT alınarak istenilen adette aynı periyoda sahip periyodik sentez sinyali elde edilebilir. Örnek sentezleme şekil 3.5’te gösterilmektedir.



Şekil 3.5 Değiştirilmiş Ters DCT ile periyodik sinyal sentezleme

3.3.2 Spektral genliklerin enterpolasyonu

Birden fazla perde periyotluk analiz sinyalinin mevcut olduğu durumda Eşitlik 3.8’de yer alan DCT katsayıları yerine ardışık perdelere ait DCT katsayılarının örtüşmeli eklemeleriyle oluşturulan değerler konulabilir. Bu durum Eşitlik 3.9 ile gösterilmektedir (İlk ve Güler 2011).

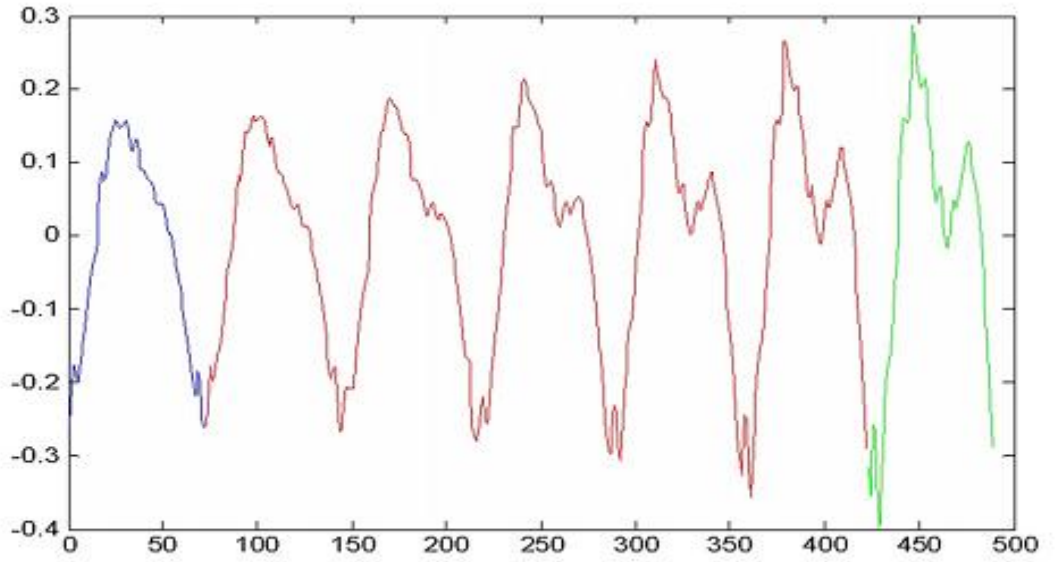
$$s[n] = \sum_{k=0}^{N-1} f_k u^k S_{\text{int}}[k] \cos\left(\frac{(2n+1)k\pi}{2N}\right) \quad (3.9)$$

Eşitlik 3.9’da yer alan $S_{\text{int}}[k]$, sinyalin iki ayrı perdesine ait DCT katsayılarından oluşur ve Eşitlik 3.10’a göre hesaplanır. Bu eşitlikte kullanılacak periyotların uzunluğu eşit değil ise kısa olan periyotun uzunluğunu artırmak için DCT katsayılarına sıfır ekleme

veya uzun olan periyotun uzunluğunu kısaltmak için kırpma kullanılır. Bu eşitlikte yer alan T iki ayrı periyot kullanılarak sentezlenen ses sinyali parçasının uzunluğunu verir.

$$S_{\text{int}}[k] = S_{\text{left}}[k](1 - \frac{n}{T})S_{\text{right}}[k](\frac{n}{T}) \quad n = 0, 1, 2, \dots, T-1 \quad (3.10)$$

Şekil 3.6’da aynı perde periyotlarına sahip iki prototip kullanılarak Eşitlik 3.9 - 3.10’a göre Değiştirilmiş DCT ile yapılan sentezleme gösterilmektedir. İlk prototip mavi ve son prototip yeşil renkle gösterilmiş olup sentezlenen prototipler kırmızı renkle gösterilmiştir.



Şekil 3.6 Aynı perde periyodu ve spektrumuna sahip iki analiz segmenti ile enterpolasyon

3.3.3 Perde dönüşümü ve enterpolasyonu

Zaman skalasına ek olarak perde skalası değişimi de birlikte gerçekleştirilmek istenebilir. Normal bir ses sinyalinde ardışık ses periyotlarının uzunlukları genelde az da olsa birbirlerinden farklıdır. Bu durumda perde periyodu N 'in zamana bağlı değişimi olarak modellenebilir. Perde periyodu değiştiği için bu perde periyodundaki sesin sentezlenmesinde kullanılacak DCT katsayısı değişiminin de modele eklenmesi gerekir.

Bu nedenle DCT ile skala ve perde değişimi için en genel sentez Eşitlik 3.11'e göre gerçekleştirilir.

$$s[n] = \sum_{k=0}^{N'(n)-1} f_k u^k S_{\text{int}}[k] \text{Cos}\left(\frac{(2a'n+1)k\pi}{2N(n)}\right) \quad (3.11)$$

Eşitlik 3.11'de yer alan $N(n)$ anlık perde modeli olup sentezde kullanılan iki analiz perdesi uzunluğuna (N_{left} ve N_{right}) göre bir eğri modeli ile hesaplanabilir. Değişimin doğrusal olduğu durumlarda $N(n) = \alpha n + \beta$ doğrusal enterpolasyon eşitliği kullanılabilir. Eşitlik 3.11'deki kosinüs terimi sinüs modellemeye olduğu gibi her bir DCT katsayısı için frekans ve o frekansdaki faz olmak üzere bileşenlerine ayrılırsa Eşitlik 3.12'deki gibi olacaktır.

$$\text{Cos}\left(\frac{(2a'n+1)k\pi}{2N(n)}\right) = \text{Cos}\left(\frac{a'k\pi}{N(n)}n + \frac{k\pi}{2N(n)}\right) = \text{Cos}(\phi_k(n) + k\theta) \quad (3.12)$$

Periyotlar arası geçişte faz korunması için ilk periyodun sonu baz alınır ise Eşitlik 3.12'de θ fazının paydasında yer alan $N(n)$ yerine N_{left} yazılabilir. $\phi_k(n)$ $k=2$ için temel frekans olduğundan anlık frekans $\frac{d\phi_2(n)}{d(n)}$ eşitliğinden bulunarak anlık perde

periyodu $N'(n) = \frac{2\pi}{\omega_2(n)}$ ile hesaplanır.

$\phi_2(0) = 0$ ve $\phi_2(T-1) = 2\pi c_0$, (c_0 tam sayı olmak üzere) olduğundan temel frekans

$\omega_2(n)$, $n=0$ için $\frac{2\pi}{P_{\text{left}}}$ ve $n=T-1$ için $\frac{2\pi}{P_{\text{right}}}$ olacaktır. Bu sınır koşullarına göre $N'(n)$

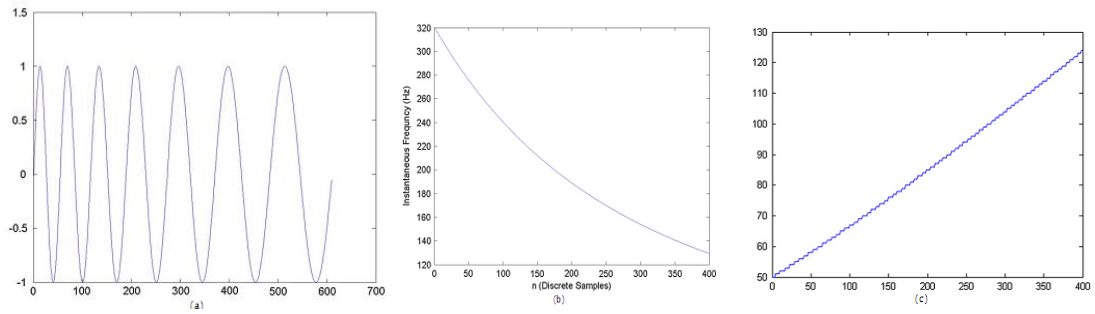
parabolik enterpolasyon ile modellenerek Eşitlik 3.13 veya Eşitlik 3.14 ile ifade edilebilir.

$$N'(n) = \frac{\alpha^2 n^2 - (\alpha^2 - 2\alpha\beta)n + \beta^2 - \alpha\beta}{\beta} \quad (3.13)$$

$$N'(n) = \left(\frac{a}{T^2} - \frac{1}{c_0 T} \right) n^2 + \left(\frac{P_{right}}{T} + \frac{1}{c_0} - \frac{2a}{T} \right) n + a \quad (3.14)$$

Eşitlik 3.14'de yer alan ifade $a = \frac{P_{right} P_{left} c_0}{T}$ ve $a' = \frac{P_{right} c_0}{T}$ ile hesaplanır.

Şekil 3.7'de 50 ve 120 Hz'lik iki sinüs sinyal prototipinden Eşitlik 3.11'e göre 400 örneklemlik değiştirilmiş DCT sentezlemesi gösterilmektedir (İlk ve Güler 2011).



Şekil 3.7 Değiştirilmiş DCT ile perde enterpolasyonu

a. Değiştirilmiş DCT ile sentezlenen sinyal, b. Doğrusal olmayan perde enterpolasyonu, c. Anlık frekans

4. BULGULAR VE TARTIŞMA

4.1 Ses Sıkıştırma Performanslarının Değerlendirilmesi

Algoritmaların karşılaştırmalarını yapmak için 16 KHz örnekleme frekansına sahip TIMIT ses veritabanından alınan 12 adet “She had your dark suit in greasy wash water all year.” cümlesinin sesleri analiz sesi olarak kullanılmıştır. Fonotik olarak dengeli olan bu cümle ile sentezleme algoritmalarının sesin farklı bölgelerinde yaratabileceği problemler hem işitsel hem de görsel (grafikler) olarak değerlendirilebilmektedir. Örnekler TIMIT veritabanında bulunan tüm lehçelerden (toplamda 8 adet) birer adet erkek sesi ve en çok kullanılan 4 lehçeden birer adet bayan sesi olarak seçilmiş olup bu seslerin CD’deki dosya yolları çizelge 4.1’de yer almaktadır.

Çizelge 4.1 TIMIT Veritabanından alınan seslerin dizin ve dosya adları

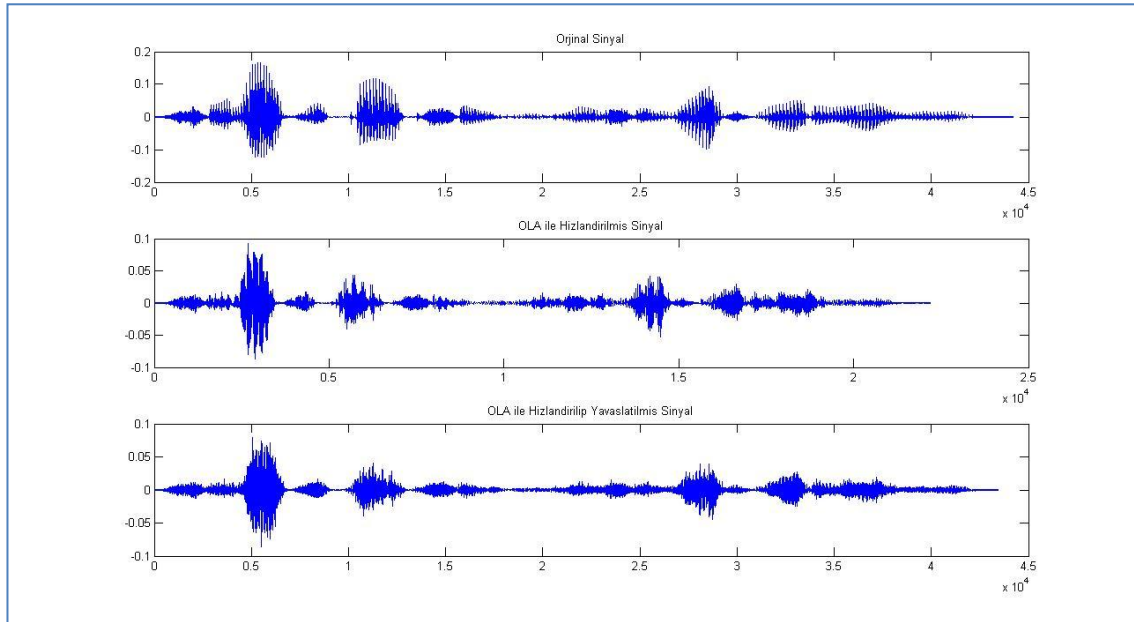
SES DOSYALARI
DR1\MDAB0\SA1.wav
DR2\FDRD1\SA1.wav
DR2\MTAS1\SA1.wav
DR3\MBDG0\SA1.wav
DR4\FEDW0\SA1.wav
DR4\MDLS0\SA1.wav
DR5\FHES0\SA1.wav
DR5\MDAC2\SA1.wav
DR6\MCMJ0\SA1.wav
DR7\FISB0\SA1.wav
DR7\MKJL0\SA1.wav
DR8\MJTC0\SA1.wav

TSM algoritmalarının bu analiz sesleri için nesnel kalite değerlendirmeleri Bölüm 4.1.13’te yer almaktadır. Nesnel değerlendirmelerden önce bu algoritmalar ile sentezlenen örnek bir sesin kişisel değerlendirmeleri Bölüm 4.1.1 ile 4.1.12 arasında yer

almaktadır. Sentezlenen sinyallerin zaman ve spektrum gösterimleri de ilgili algoritmaların değerlendirilmesinde ipucu verebilmektedir. Bu nedenle kullanılan örnek ses, bu sesin hızlandırılması ile sentezlenen ses ve hızlandırılan sesin orijinal skalaya dönüştürülmesi ile sentezlenen sesin zaman ve spektrum gösterimleri de kişisel değerlendirmeler ile birlikte alt bölümlerde yer almaktadır.

4.1.1 OLA ile ses sıkıştırma öznel değerlendirmesi

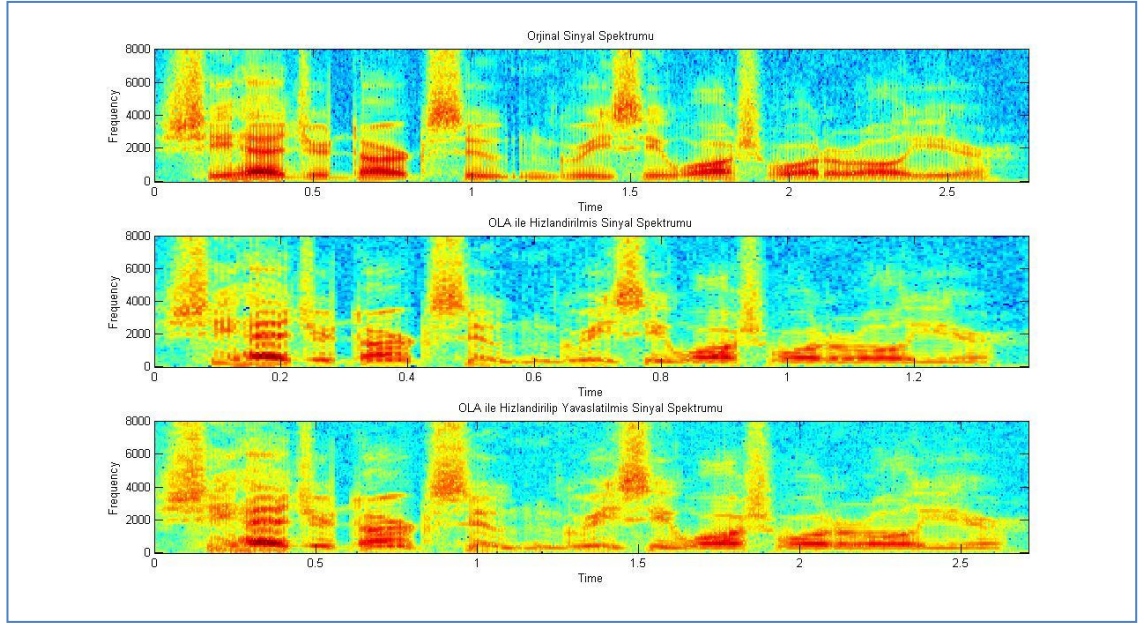
OLA algoritması benzerlik araması yapmadan örtüşmeli ekleme yapmaktadır. Bu nedenle sentezlenen sesler dinlendiğinde belirgin bir mekanikleşme bulunmaktadır.



Şekil 4.1 OLA sentez sesleri

a. Orijinal ses, b. Hızlandırılmış sentez sesi, c. Hızlandırılan sesin orijinal skalaya dönüştürüldüğü sentez sesi

Örnek çıktı grafikleri şekil 4.1 - 4.2’de görülmektedir. Sesteki bozulma grafiklerden de belli olmaktadır. Şekil 4.1’de en üstteki orijinal ses sinyalindeki tepe noktaları ayrık görünüyorken en alttaki OLA sentez sinyalinde ise bu ayrıklıklar bozulmuştur. Şekil 4.2’de de sesli-sessiz veya sessiz-sesli geçişlerindeki (özellikle 0,6 sn dolaylarında oldukça belli) ve sesli bölgelerdeki (0,2-0,3 sn aralığında oldukça) bozulmalar belli olmaktadır.

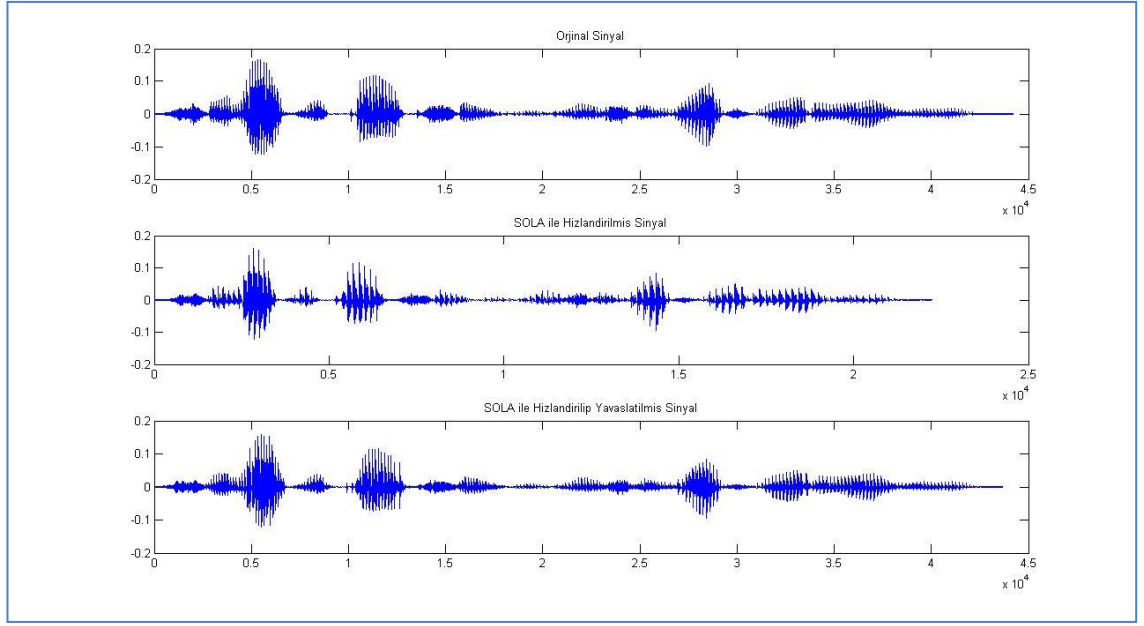


Şekil 4.2 OLA sentez sinyal spektrumları

a. Orijinal ses, b. Hızlandırılmış sentez sesi, c. Hızlandırılan sesin orijinal skalaya dönüştürüldüğü sentez sesi

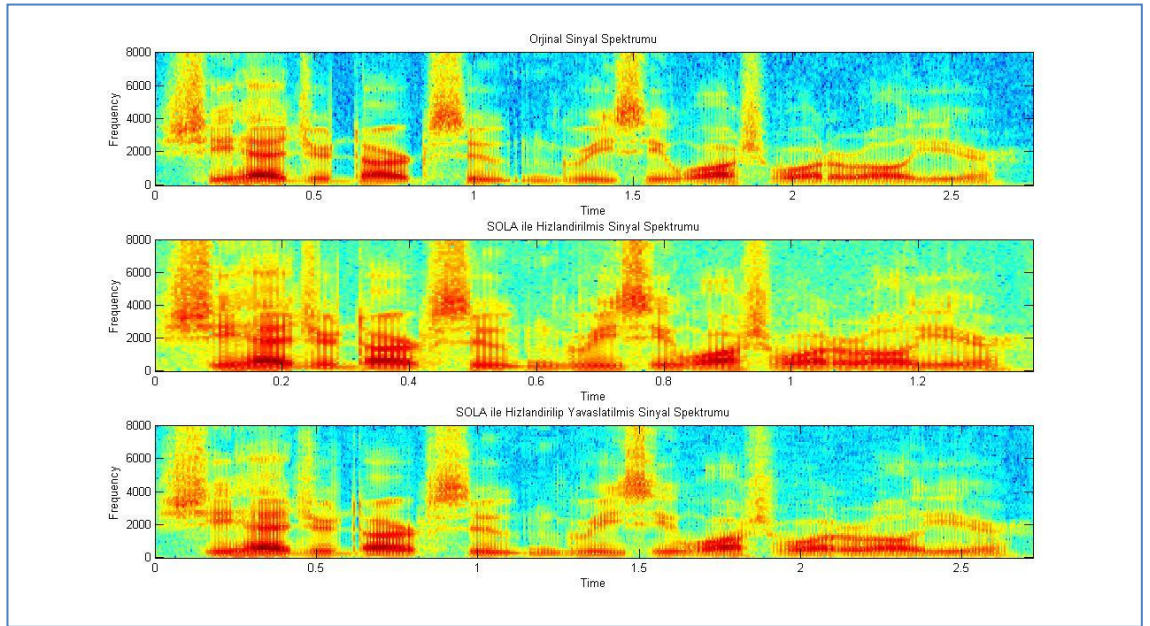
4.1.2 SOLA ile ses sıkıştırma öznel değerlendirmesi

SOLA benzerlik aramasını sentez sesinde yapmaktadır. SOLA ile sentezlenen ses dinlendiğinde sesin kalitesinin iyi olduğu görülmüştür. Bununla birlikte ses kalitesinde zor fark edilen bir bozulma bulunmaktadır. SOLA için deney çıktı grafikleri şekil 4.3 - 4.4'te görülmektedir. Sesteki bozulma grafiklerden çok belli olmamaktadır. Şekillerde SOLA sentez sesinin analiz sesinden daha kısa olması analiz sesinin son segmentinin işlenmemiş olmasındandır.



Şekil 4.3 SOLA sentez sesleri

a. Orijinal ses, b. Hızlandırılmış sentez sesi, c. Hızlandırılan sesin orijinal skalaya dönüştürüldüğü sentez sesi

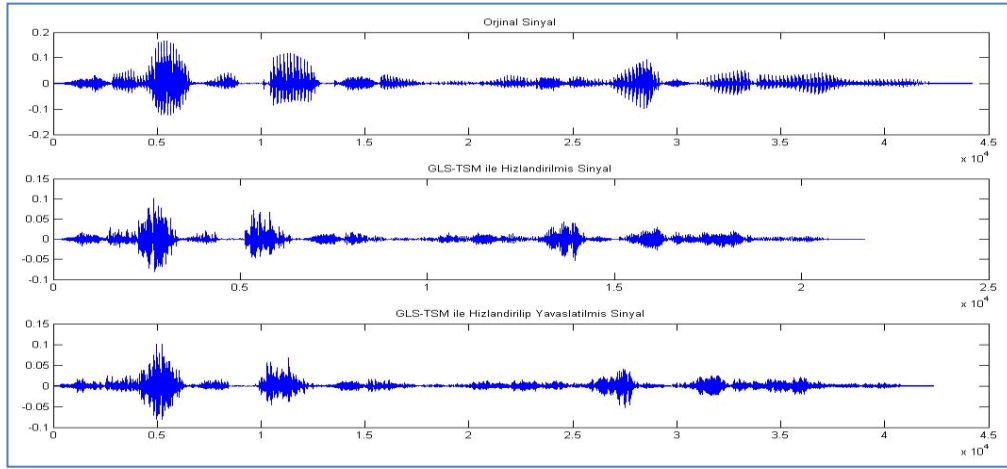


Şekil 4.4 SOLA sentez sinyal spektrumları

a. Orijinal ses, b. Hızlandırılmış sentez sesi, c. Hızlandırılan sesin orijinal skalaya dönüştürüldüğü sentez sesi

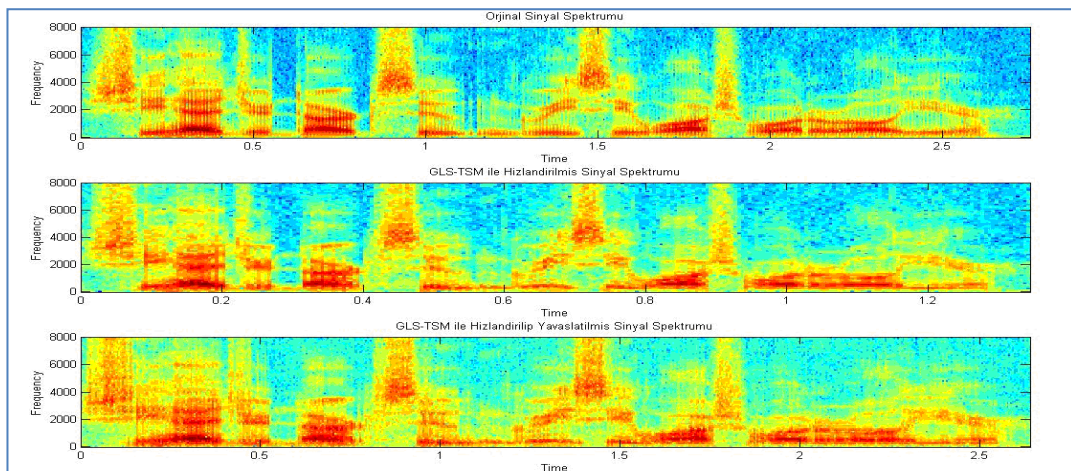
4.1.3 GLS-TSM ile ses sıkıştırma öznel değerlendirmesi

GLS-TSM algoritmasında hızlandırma ve yavaşlatma için SOLA parametrelerinin aynısı kullanılmıştır. Buna rağmen benzerlik aramada yaşadığı problemlerden dolayı SOLA algoritmasına göre oldukça kötü sonuç vermiştir. Yine de sesteki mekanikleşme OLA'daki kadar çok değildir. Benzerliği bulmadaki problemi ve sentezleme sesindeki bozulma şekil 4.5 - 4.6'daki grafiklerden de belli olmaktadır.



Şekil 4.5 GLS-TSM sentez sesleri

a. Orijinal ses, b. Hızlandırılmış sentez sesi, c. Hızlandırılan sesin orijinal skalaya dönüştürüldüğü sentez sesi

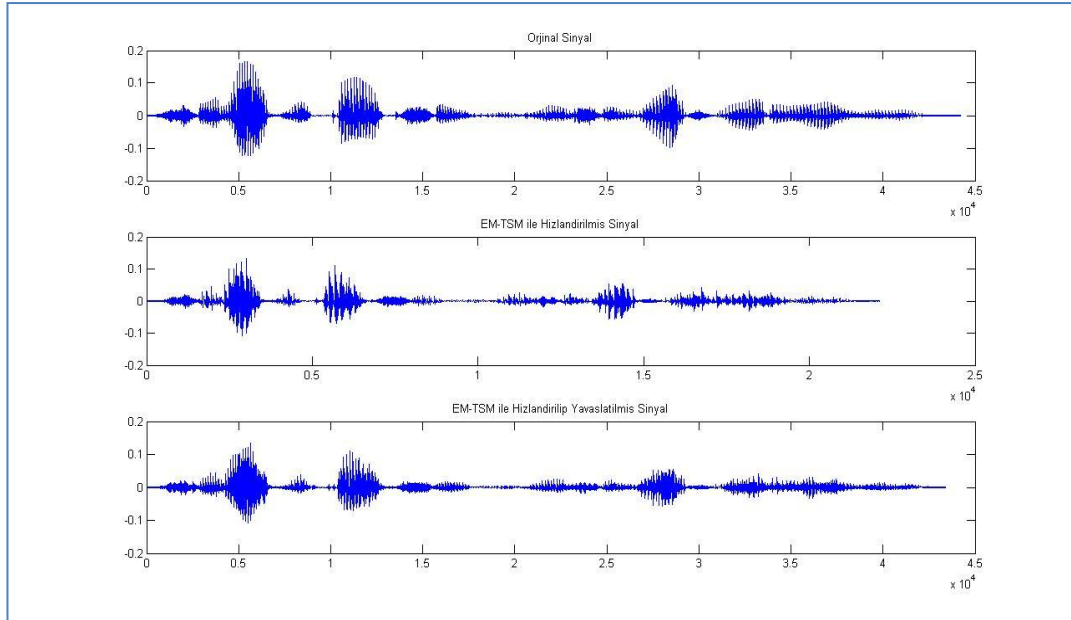


Şekil 4.6 GLS-TSM sentez sinyal spektrumları

a. Orijinal ses, b. Hızlandırılmış sentez sesi, c. Hızlandırılan sesin orijinal skalaya dönüştürüldüğü sentez sesi

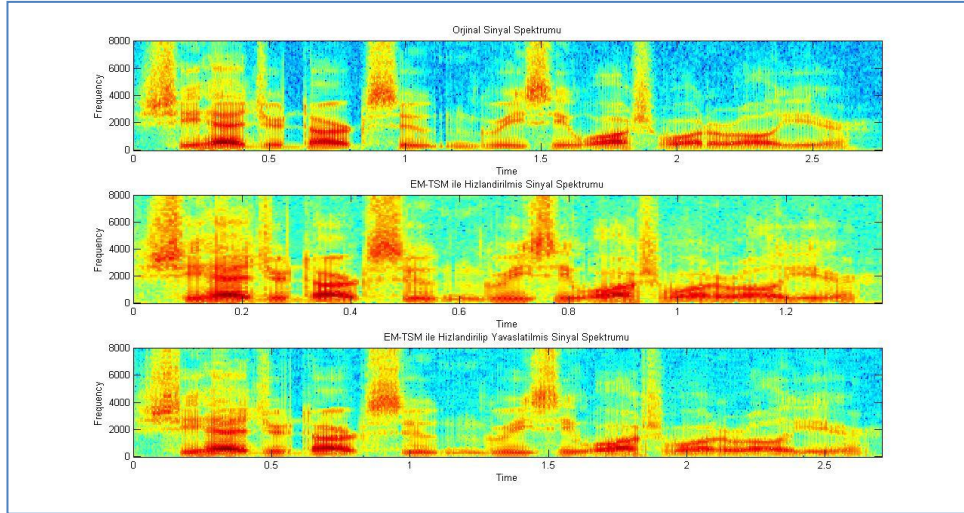
4.1.4 EM-TSM ile ses sıkıştırma öznel değerlendirmesi

EM-TSM benzerlik aramasını sinyal örneklerinin değerleri yerine işaretlerine göre yapmaktadır. Bu genelde doğru benzerliğin bulunmasını sağlasa da özellikle de sinyalin DC bileşen de içermesi durumunda yanlış benzerlikler bulunabilmektedir. Hızlandırılan ve yavaşlatılan ses dinlendiğinde sesin kalitesinin SOLA'dan biraz daha bozuk olduğu görülmüştür. EM-TSM için örnek grafikler şekil 4.7 - 4.8'de görülmektedir. Sesin son bölgelerinde ve sesli kısımlarında belirginleşen bozulma şekil 4.7.'deki orijinal ve sentez grafiklerinde de belirgindir.



Şekil 4.7 EM-TSM sentez sesleri

a. Orijinal ses, b. Hızlandırılmış sentez sesi, c. Hızlandırılan sesin orijinal skalaya dönüştürüldüğü sentez sesi

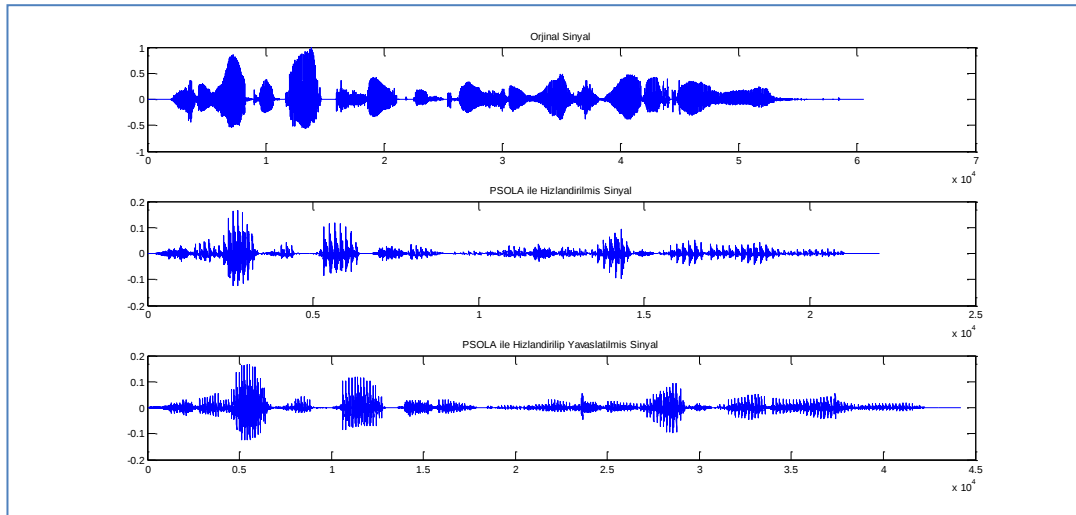


Şekil 4.8 EM-TSM sentez sinyal spektrumları

a. Orijinal ses, b. Hızlandırılmış sentez sesi, c. Hızlandırılan sesin orijinal skalaya dönüştürüldüğü sentez sesi

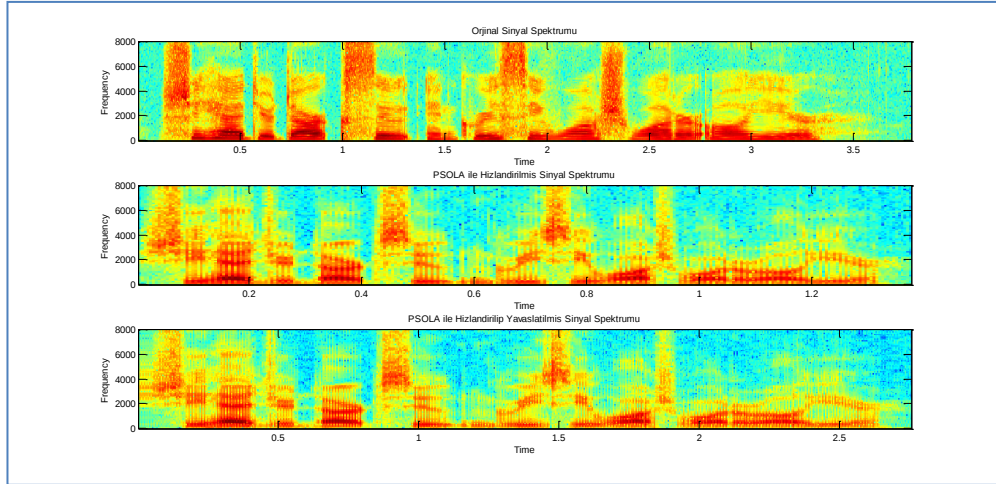
4.1.5 TD-PSOLA ile ses sıkıştırma öznel değerlendirilmesi

Ardışık perde periyotları farklı işaretlenebilmiş olmasına rağmen sentez kalitesi beklenenin aksine çok kötü gerçekleşmemiş hemen hemen SOLA'ya yakın ses kalitesi elde edilmiştir.



Şekil 4.9 PSOLA sentez sesleri

a. Orijinal ses, b. Hızlandırılmış sentez sesi, c. Hızlandırılan sesin orijinal skalaya dönüştürüldüğü sentez sesi

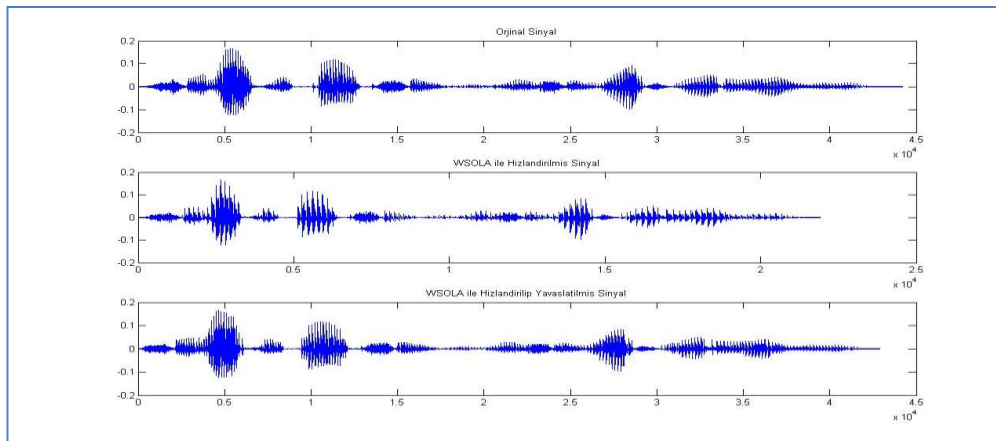


Şekil 4.10 PSOLA sentez sinyal spektrumları

a. Orijinal ses, b. Hızlandırılmış sentez sesi, c. Hızlandırılan sesin orijinal skalaya dönüştürüldüğü sentez sesi

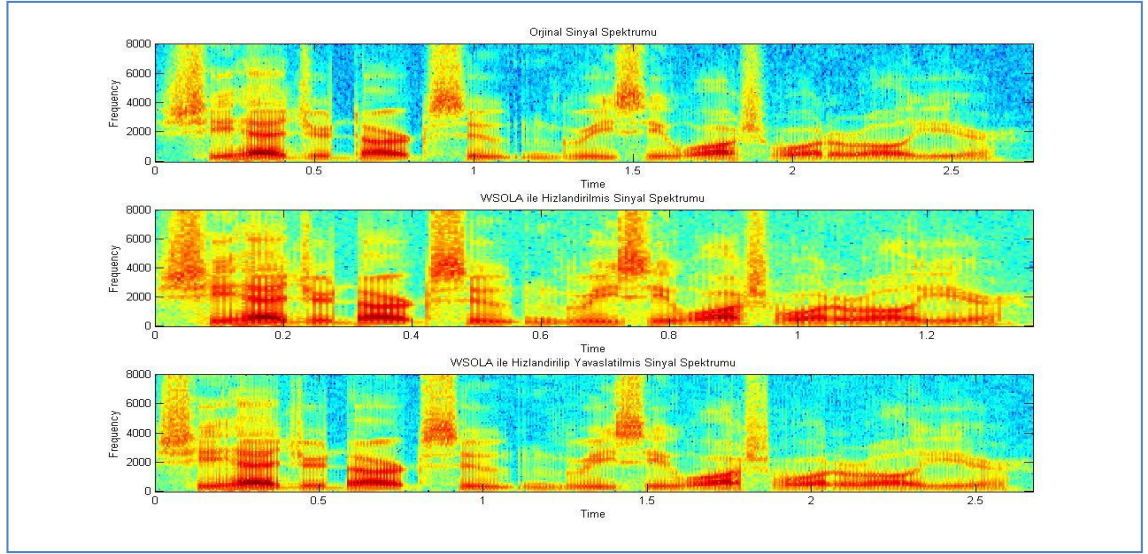
4.1.6 WSOLA ile ses sıkıştırma öznel değerlendirmesi

WSOLA algoritmasında çok daha düşük segmentler ve benzerlik arama aralığı kullanılmasına rağmen hızlandırılıp yavaşlatılan sesin öznel değerlendirmesinde diğerlerine göre genelde daha kaliteli ses elde edilmiştir. WSOLA için deney çıktı grafikleri şekil 4.11 - 4.12’de yer almaktadır.



Şekil 4.11 WSOLA sentez sesleri

a. Orijinal ses, b. Hızlandırılmış sentez sesi, c. Hızlandırılan sesin orijinal skalaya dönüştürüldüğü sentez sesi



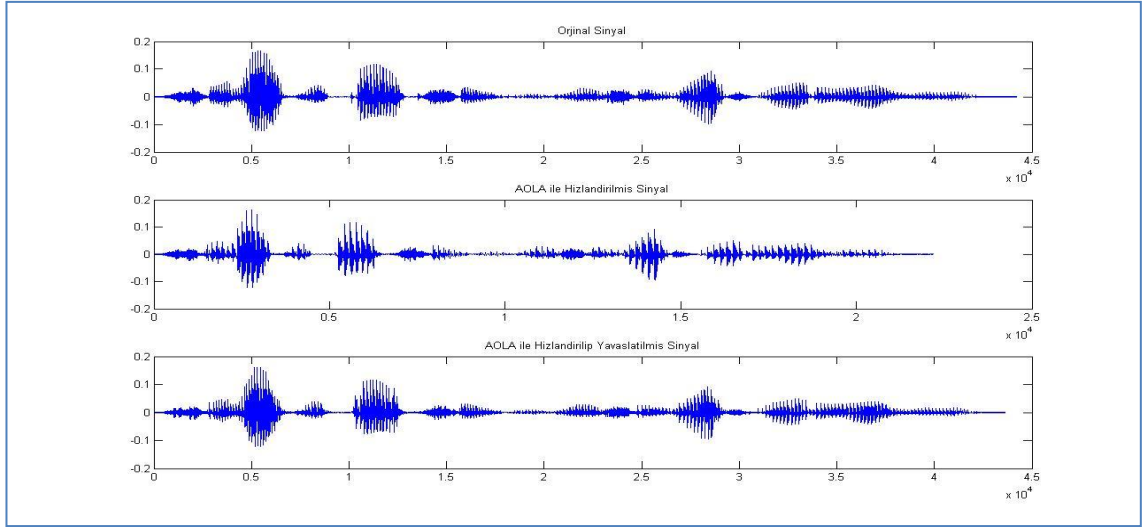
Şekil 4.12 WSOLA sentez sinyal spektrumları

a. Orijinal ses, b. Hızlandırılmış sentez sesi, c. Hızlandırılan sesin orijinal skalaya dönüştürüldüğü sentez sesi

Diğer algoritmaların grafiklerinde de olduğu gibi 0,5 saniyesindeki ses sinyalindeki anlık yükseliş düşme sentez sesinde sinyalin diğer yüksek genlikli kısmıyla iç içe girmiştir.

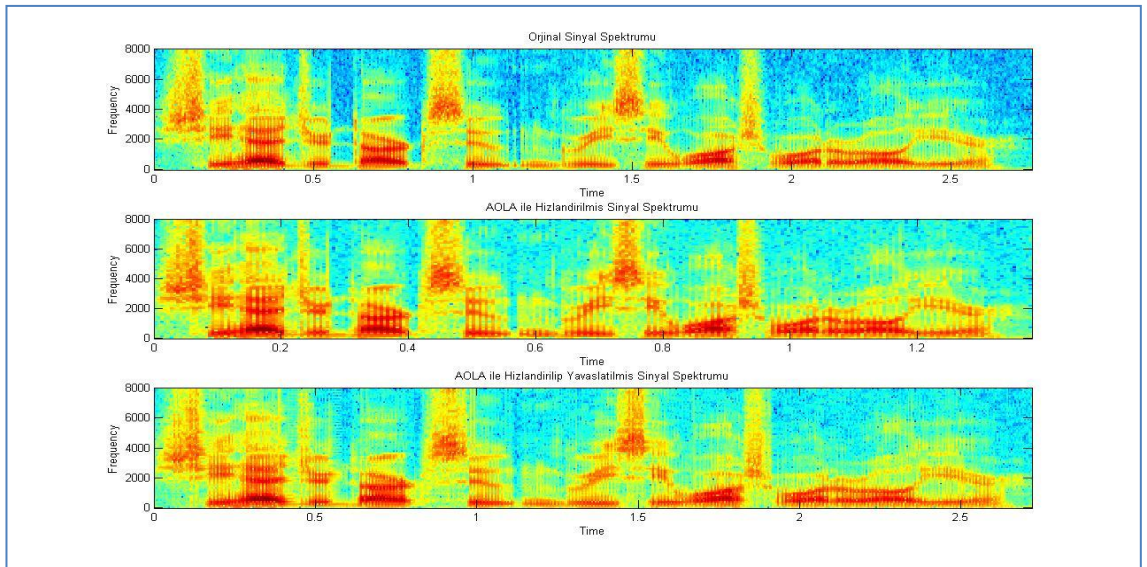
4.1.7 AOLA ile ses sıkıştırma öznel değerlendirilmesi

AOLA algoritmasında sentezlenen ses dinlendiğinde sesin kalitesinin WSOLA'ya çok yakın olduğu görülmüştür. Fakat orijinal AOLA algoritmasında segmentin kaydırılması tepe noktaları örtüşecek (peak alignment) şekilde tarif edilmektedir. Sadece yerel tepe noktalarına bakmak yanlış tepelerin örtüşmesine neden olabilecektir. Bu nedenle bunun yerine AMDF kullanılmıştır. AOLA için deney çıktı grafikleri şekil 4.13 - 4.14'te görülmektedir.



Şekil 4.13 AOLA sentez sesleri

a. Orijinal ses, b. Hızlandırılmış sentez sesi, c. Hızlandırılan sesin orijinal skalaya dönüştürüldüğü sentez sesi



Şekil 4.14 AOLA sentez sinyal spektrumları

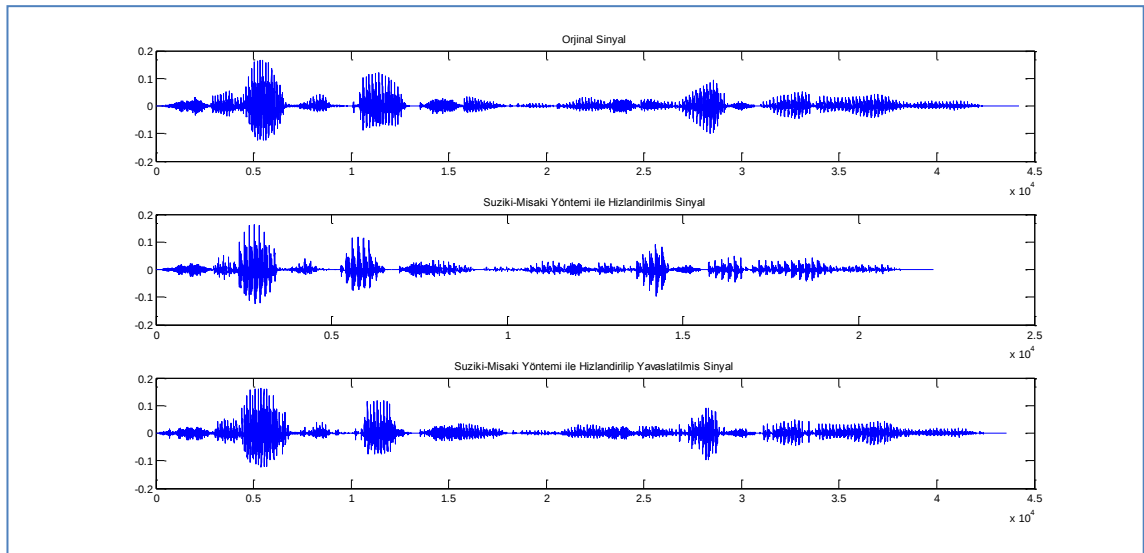
a. Orijinal ses, b. Hızlandırılmış sentez sesi, c. Hızlandırılan sesin orijinal skalaya dönüştürüldüğü sentez sesi

4.1.8 Suzuki-Misaki yöntemi ile ses sıkıştırma öznel değerlendirilmesi

Bu algoritma ile hızlandırılan ve yavaşlatılan ses dinlendiğinde sesin kalitesinin diğer yöntemlere göre daha kötü olduğu görülmüştür. Benzerlik arama yönteminin Bölüm

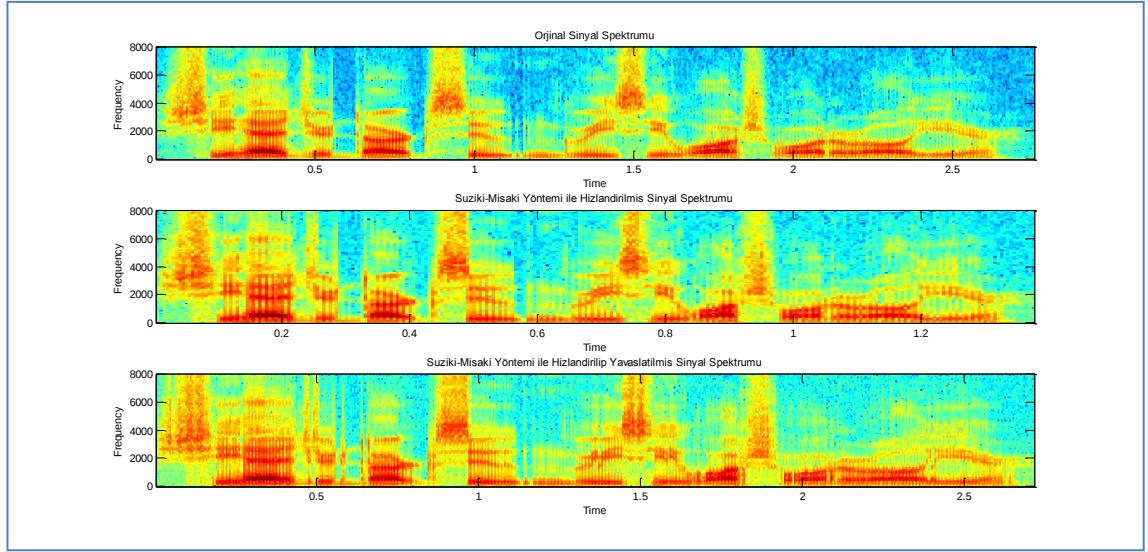
3.1.3.2’de belirtildiği gibi düzeltilmesi kısmi iyileştirme sağlamıştır. Bu yöntem için grafikler şekil 4.15 – 4.16’da görülmektedir.

Şekil 4.16’da yer alan spektrumlar incelendiğinde orijinal sesin ve hızlandırılmış sesin spektrumları arasında çözünürlüğün yarıya inmesinden kaynaklananlar hariç önemli bir bozulma görülmemektedir. Bununla birlikte hızlandırılmış sesin yavaşlatılmasından sonraki spektrumda sesli-sessiz ve sessiz-sesli geçişlerinde bozulmalar (şekil 4.16.c’de 0,6 ve 0,7 saniye aralığındaki gibi) grafikten de belli olmaktadır. Bunun nedeni Bölüm 2.1.8’de yer alan şekil 2.7’deki yavaşlatma gösteriminden de açıkça belli olduğu gibi analizden alınan ardışık 2 pencerenin katlama ağırlıklarının ters olmasıdır. Sentez aşamasındaki katlamada önceki analiz segmentinin ağırlık pencere katsayıları sıfırdan bire doğru giderek artmakta, sonraki analiz segmentinin ağırlık pencere katsayıları ise giderek azalmaktadır. Bundan kaynaklı olarak önceki segment sesli ardışığı sessiz olduğunda sentez sesinin ilk kısımlarının sessiz son kısımlarının sesli ağırlıklı sentezlenmesine neden olmaktadır. Bu analiz sesindeki sesli-sessiz geçişlerinin sentezlenmesinin sesli- sessiz-sesli-sessiz olarak gerçekleşmesine neden olmaktadır.



Şekil 4.15 Suzuki-Misaki yöntemi ile sentez sesleri

a. Orijinal ses, b. Hızlandırılmış sentez sesi, c. Hızlandırılan sesin orijinal skalaya dönüştürüldüğü sentez sesi

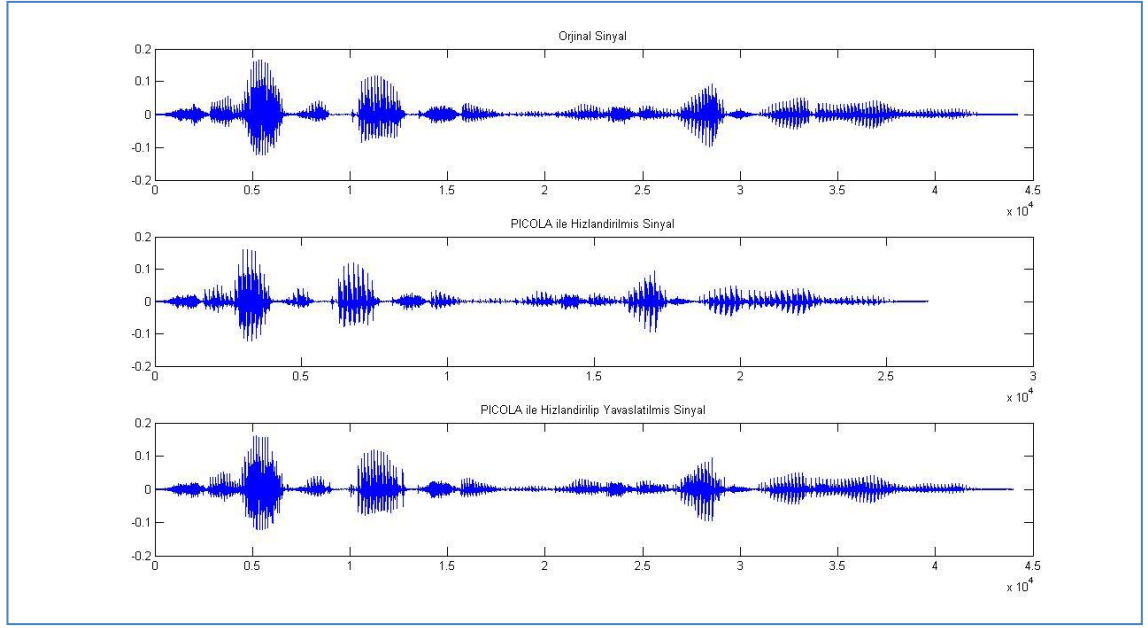


Şekil 4.16 Suziki Misaki yöntemi ile sentez sinyal spektrumları

a. Orijinal ses, b. Hızlandırılmış sentez sesi, c. Hızlandırılan sesin orijinal skalaya dönüştürüldüğü sentez sesi

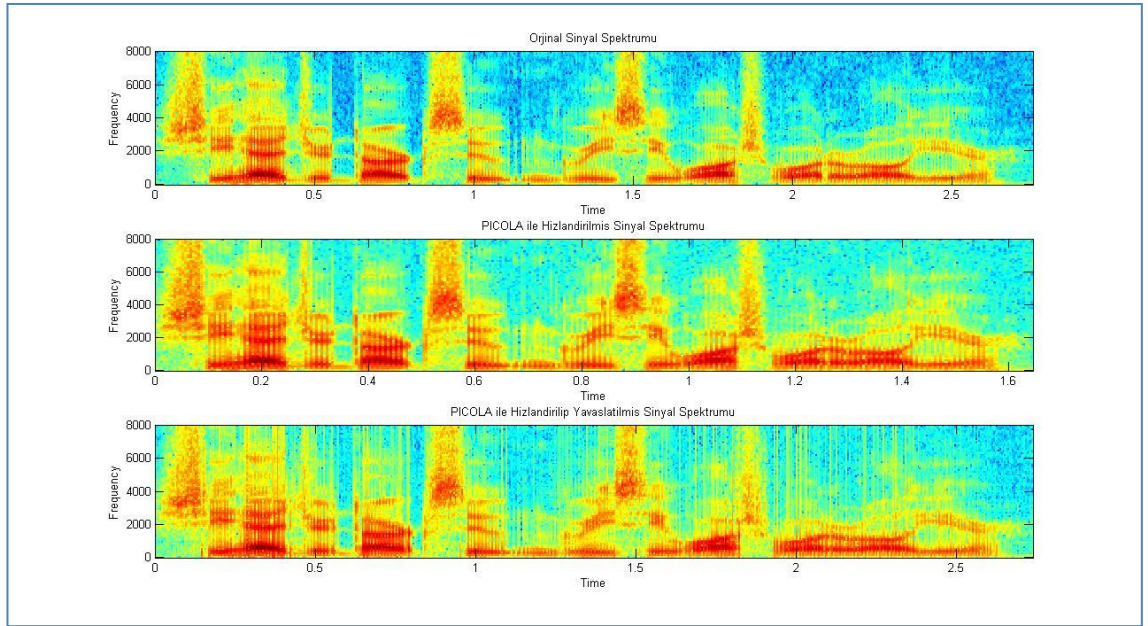
4.1.9 PICOLA ile ses sıkıştırma öznel değerlendirilmesi

PICOLA algoritması MPEG AAC’de belirtildiği gibi uygulanmıştır. PICOLA ile hızlandırılan ve yavaşlatılan ses dinlendiğinde sesin kalitesinin aşağı yukarı AOLA ve SOLA’daki gibi olduğu görülmüştür. PICOLA için grafikler şekil 4.17 - 4.18’de görülmektedir. Şekil 4.17’de WSOLA, SOLA, AOLA, PSOLA gibi sentezleme kalitesi nispeten iyi olan sinyallerdekinin aksine yaklaşık 0,65 ms’deki (yaklaşık 10000. örneklem dolayları) anlık yüksek genlik ses parçası bozulmamıştır. PICOLA’nın hızlandırma için o bölgede örtüşme yapmadığı hızlandırma grafiklerinden bellidir. Bununla birlikte diğer geçiş bölgelerinde de diğer algoritmaların aksine bozulmalar görülebilmektedir.



Şekil 4.17 PICOLA sentez sesleri

a. Orijinal ses, b. Hızlandırılmış sentez sesi, c. Hızlandırılan sesin orijinal skalaya dönüştürüldüğü sentez sesi

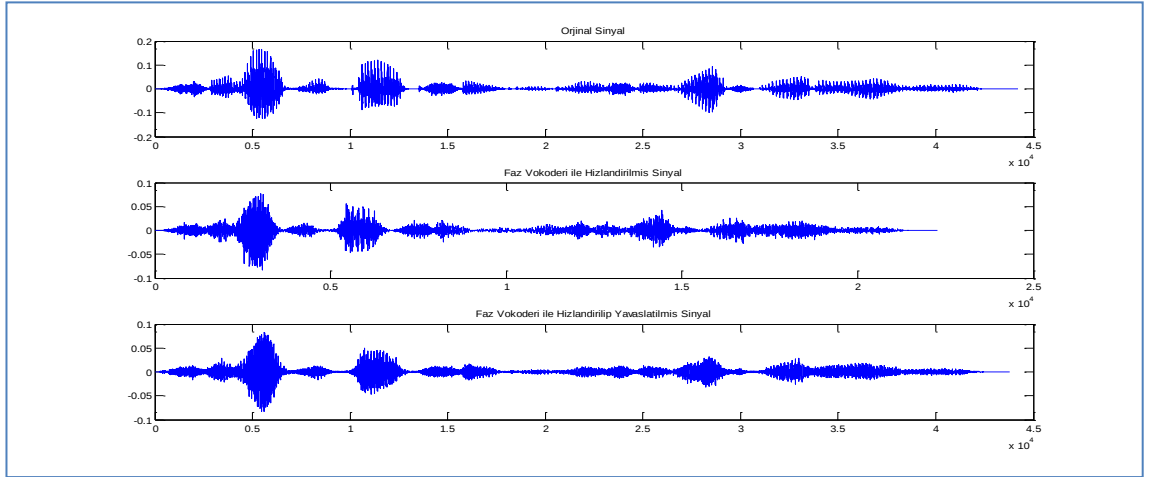


Şekil 4.18 PICOLA sentez sinyal spektrumları

a. Orijinal ses, b. Hızlandırılmış sentez sesi, c. Hızlandırılan sesin orijinal skalaya dönüştürüldüğü sentez sesi

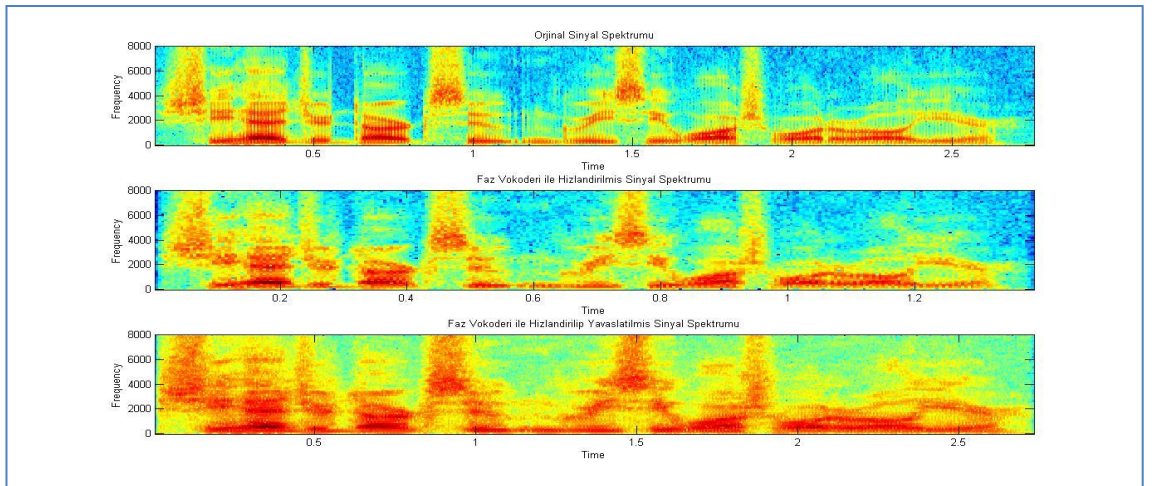
4.1.10 Faz vokoderi ile ses sıkıştırma öznel değeriendirilmesi

Faz vokoder ile sentezlenen sesin kalite bozulması genelde zaman alanı tekniklerinden daha fazladır. Ses sinyalinin enerji kaybı da şekil 4.19’da da görüldüğü gibi zaman alanında çalışan örtüşmeli ekleme tekniklerine göre daha fazladır. Şekil 4.20’de yer alan sentez spektrumundaki frekans yayılmaları da belirgindir.



Şekil 4.19 Faz vokoderi sentez sesleri

a. Orijinal ses, b. Hızlandırılmış sentez sesi, c. Hızlandırılan sesin orijinal skalaya dönüştürüldüğü sentez sesi

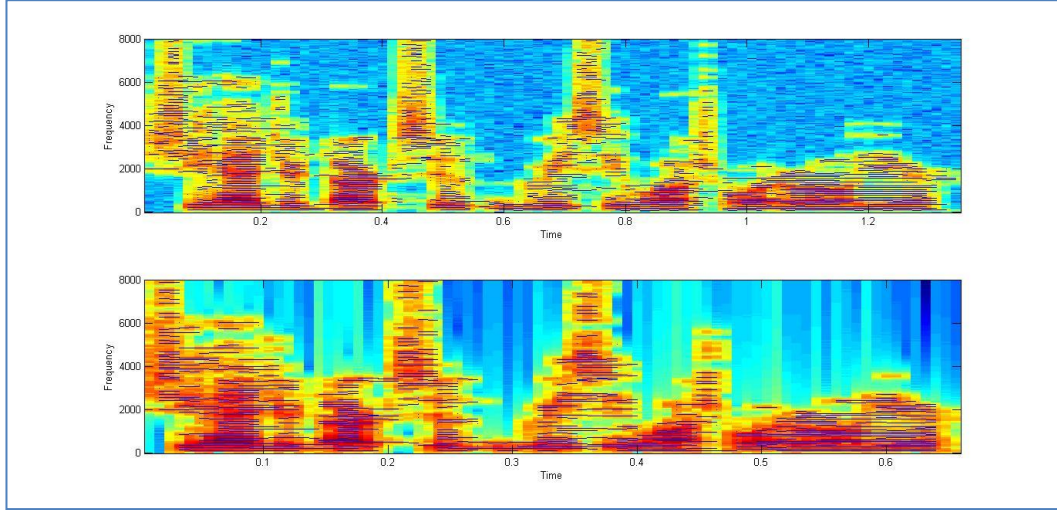


Şekil 4.20 Faz vokoderi sentez sinyal spektrumları

a. Orijinal ses, b. Hızlandırılmış sentez sesi, c. Hızlandırılan sesin orijinal skalaya dönüştürüldüğü sentez sesi

4.1.11 Sinüs modelleme ile ses sıkıştırma öznel değerlendirilmesi

Sinüs modellemede her iki yöntemde de tepelerin bulunması için aynı parametreler kullanılmıştır. Şekil 4.21’de orijinal sesin ve sinüs modelleme ile hızlandırılmış sentez (tepe enterpolasyonlu) sesinin sinüs tepe yolları mavi çizgiler ile gösterilmektedir.

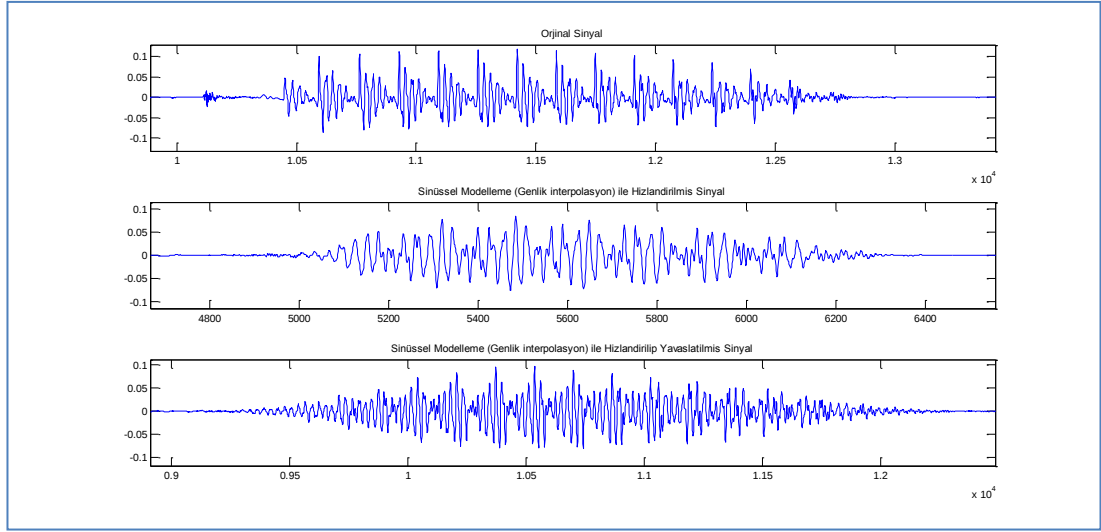


Şekil 4.21 Sinüs tepelerinin yolları

a. Orijinal sesin sinüs yolları(mavi çizgiler), b. Hızlandırılmış sesin sinüs yolları

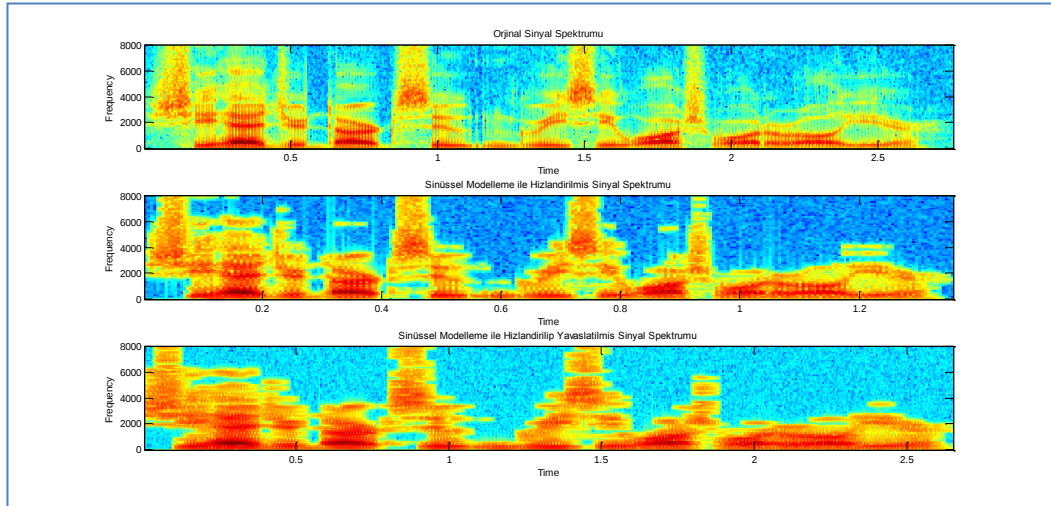
Smith ve Serra (1987) tarafından önerilen yöntemle göre sentezlenen seslerde sadece frekans genlikleri kullanılmış, gerçek sinüs tepesi tahminlemesi için parabolik enterpolasyon kullanılmıştır.

Şekil 4.22’de yer alan grafikte de görüldüğü gibi sinüs tepe yollarının başlangıç fazları hep sıfır alındığından, sentezlenen segmentlerin dalga biçimleri ilgili analiz sinyali segmentlerinin dalga biçimlerinden farklıdır. Frekansların yayvanlaşması şekil 4.23’de belirgin olarak görülmektedir. Zaman grafiğindeki sinyallerin orjinalden çok farklı olmasına ve spektrum grafiğinin oldukça bozuk olmasına rağmen ilgili ses dinlendiğinde sentez sesi kalitesi beklenenin aksine çok kötü değildir. Hemen hemen SOLA sentez sesi kalitesindedir.



Şekil 4.22 Sinüs modelleme sentez ses parçaları

a. Orijinal ses, b. Hızlandırılmış sentez sesi, c. Hızlandırılan sesin orijinal skalaya dönüştürüldüğü sentez sesi

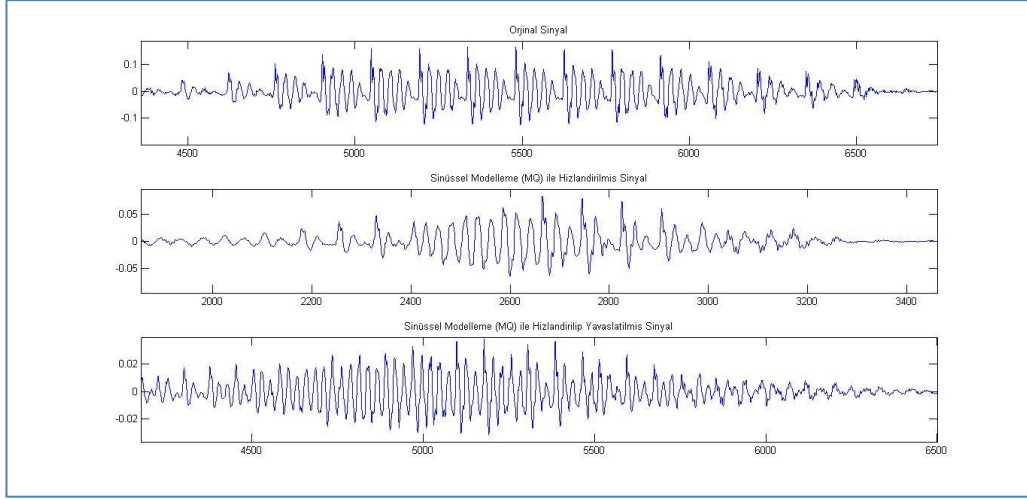


Şekil 4.23 Sinüs modelleme sentez sinyal spektrumları

a. Orijinal ses, b. Hızlandırılmış sentez sesi, c. Hızlandırılan sesin orijinal skalaya dönüştürüldüğü sentez sesi

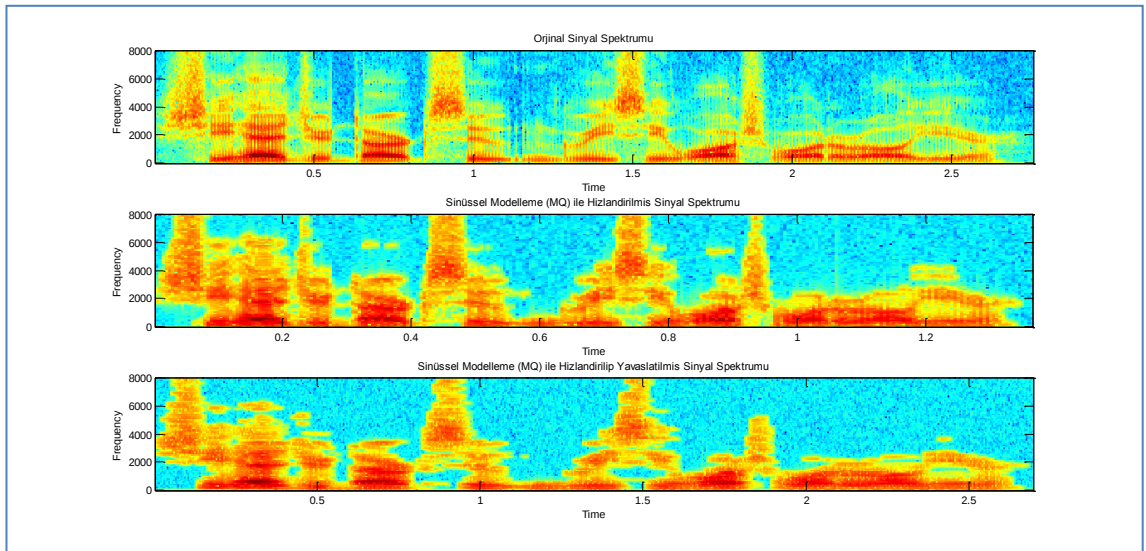
İkinci yöntemde faz bilgileri de kullanıldığından sinüs tepelerinin ardışık komşularından enterpolasyonu uygulanmamıştır. Sentez aşamasında ardışık segmentler arasında genlik ve frekans enterpolasyonu yanında, faz için kübik enterpolasyon da uygulanmıştır.

MQ sinüs modellemede sinüs yollarının başlangıç fazları karmaşık FFT değerlerinin tanjantı ile belirlenmektedir. Şekil 4.24'te de görüldüğü gibi sentez sesi model aldığı orijinal sese zaman alanı gösteriminde de benzemektedir. Şekil 4.25'de yer alan spektrumlarda da frekansların yayvanlaştığı görülmektedir. MQ sinüs modelleme beklenenin aksine kalite olarak daha kötü sonuç vermiştir.



Şekil 4.24 MQ sinüs modelleme sentez ses parçaları

a. Orijinal ses, b. Hızlandırılmış sentez sesi, c. Hızlandırılan sesin orijinal skalaya dönüştürüldüğü sentez sesi

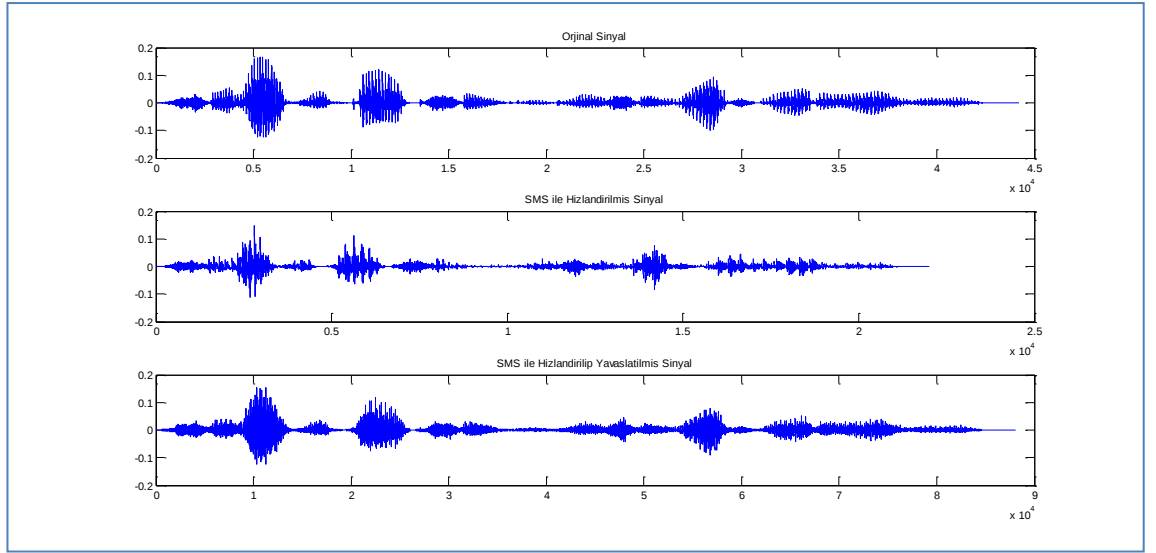


Şekil 4.25 MQ sinüs modelleme sentez sinyal spektrumları

a. Orijinal ses, b. Hızlandırılmış sentez sesi, c. Hızlandırılan sesin orijinal skalaya dönüştürüldüğü sentez sesi

4.1.12 SMS ile ses sıkıştırma öznel değerlendirilmesi

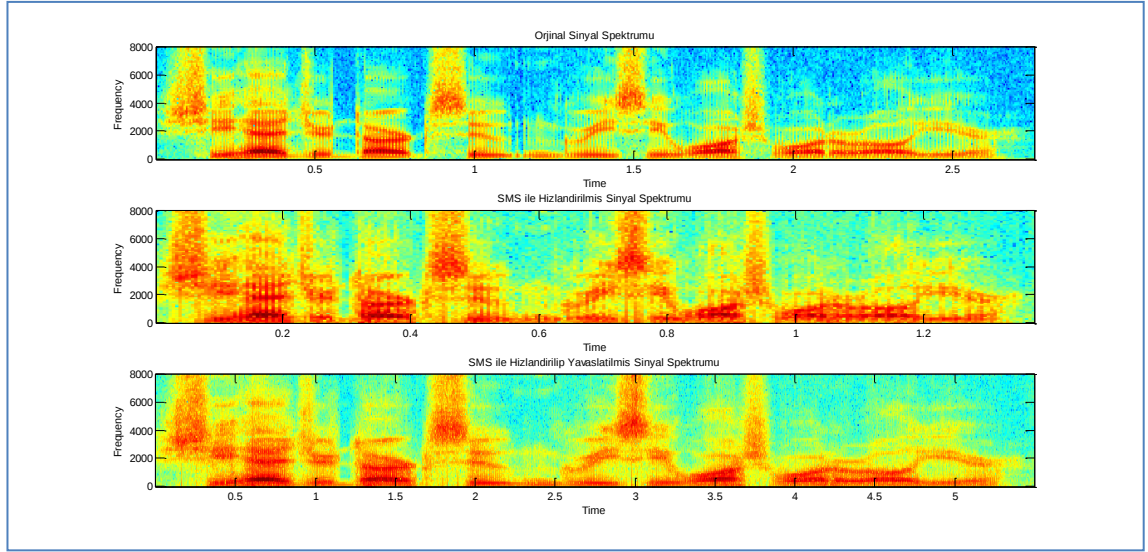
Spektral Modelleme Sentezi için yapılan testlerde yukarıda yer alan sinüs modelleme kodlarından faydalanılmıştır. Sinüs modelleme analizinden çıkan parametreler ile orijinal skala ile MQ ile sentezlenen ses orijinal sesden çıkarılarak kalan sinyalin LPC parametreleri bulunmuştur. Daha sonra sesi hızlandırmak için Sinüs modelleme (SS) analiz parametreleri ve LPC parametreleri ile daha geniş pencereler kullanılmıştır. Testin ikinci aşamasında hızlandırılan ses aynı yöntemle yavaşlatılarak orijinal zaman skalasına dönülmüştür.



Şekil 4.26 Spektral modelleme sentezi sesleri

a. Orijinal ses, b. Hızlandırılmış sentez sesi, c. Hızlandırılan sesin orijinal skalaya dönüştürüldüğü sentez sesi

Her ne kadar grafiklerdeki bozukluklar çok belirgin olmasa da SMS ile sentezlenen sesler dinlendiğinde çok fazla bozukluk olduğu görülmüştür. SMS'in deterministik kısmı MQ kullandığından MQ modelleme ile birlikte en kötü ses kalitesi elde edilmiştir.



Şekil 4.27 Spektral modelleme sentez sinyal spektrumları

a. Orijinal ses, b. Hızlandırılmış sentez sesi, c. Hızlandırılan sesin orijinal skalaya dönüştürüldüğü sentez sesi

4.1.13 PESQ skorlarına göre nesnel kalite değerlendirmesi

Analiz ses örneklerine göre algoritmaların hızlandırma-yavaşlatma PESQ skor istatistikleri çizelge 4.2’de yer almaktadır. Aynı analiz seslerine beş kez hızlandırma-yavaşlatma sentez çevrimi uygulandıktan sonra PESQ istatistikleri ise çizelge 4.3’te yer almaktadır.

Çizelgelerde 8 adet erkek ve 4 adet bayan analiz sesinin sentezleme PESQ sonuçlarının ortalaması ve standart sapması ayrı olarak da yer almaktadır. Çizelgenin satırları GENEL sütununun ORT. alt-sütununda yer alan 12 adet analiz sesi için PESQ skor ortalamasına göre sıralanmıştır.

Çizelge 4.2’de yer alan sonuçlara göre WSOLA en kaliteli ses sentezlemesi yapmaktadır. Smith ve Serra(1997) tarafından önerilen sinüs modelleme de PESQ sonuçlarına göre WSOLA hariç diğer zaman alanı tekniklerini geçmiştir. Bu sesler dinlendiğinde ise sinüs modellemenin bazı bölgelerinde çok az da olsa bazı mekanikleşmeler görülmek görülmektedir. WSOLA’da çok çok az olmakla birlikte zaman alanı teknikleri ile sentezlenen seslerde de az da olsa bazı geçişlerde pop sesleri

duyulmaktadır. AMDF ile benzerlik arayan AOLA, WSOLA hariç diğer tüm zaman alanı TSM tekniklerinden daha iyi sonuç vermektedir.

Çizelge 4.2 Algoritmaların hızlandırma ve yavaşlatma sentezi PESQ sonuç istatistikleri

TSM ALGORİTMASI	ERKEK SESİ		BAYAN SESİ		GENEL	
	ORT.	STD. SAPMA	ORT.	STD. SAPMA	ORT.	STD. SAPMA
WSOLA	3,26	0,28	3,30	0,10	3,28	0,23
Sinüs Modelleme(SS)	3,14	0,14	2,90	0,22	3,06	0,20
AOLA	3,10	0,24	2,93	0,14	3,04	0,22
SOLA	3,05	0,17	2,89	0,18	3,00	0,18
PSOLA	2,98	0,22	2,58	0,08	2,85	0,27
Faz Vokoder	2,80	0,12	2,59	0,05	2,73	0,14
PICOLA	2,74	0,21	2,48	0,12	2,66	0,22
EM-TSM	2,65	0,21	2,44	0,07	2,58	0,20
GLS-TSM	2,54	0,06	2,43	0,12	2,50	0,09
Suziki Misaki	2,38	0,31	2,18	0,14	2,31	0,28
OLA	2,28	0,22	2,11	0,10	2,23	0,20
Sinüs Modelleme(MQ)	2,01	0,23	1,78	0,09	1,93	0,22
Spektral Modelleme Sentezi	1,92	0,22	1,53	0,06	1,79	0,26

Çizelge 4.3'te ise sinüs modellemenin PESQ sonuçları en iyi çıkmıştır. Bununla birlikte öznel değirlendirmede WSOLA işitsel olarak daha kaliteli ses vermiş olmasına rağmen PESQ sonuçları daha düşüktür. Sinüs modelleme ile sentezlenen seslerde rahatsız edici bir cızırtı olmamakla birlikte ses sanki daha derinden çıkıyormuş, kelimeler yutuluyormuş gibi algılanmaktadır. SOLA ile sentezlenen seste de çok az bir cızırtı bulunmakta fakat onda da ses biraz derinden gelmektedir. AOLA ve PICOLA'da cızırtı SOLA'ya göre biraz daha fazladır. PSOLA'da ise cızırtılar AOLA'dan daha fazla olmakla birlikte cızırtı olmayan kısımlar oldukça net ve kalitelidir. Bu cızırtılar da

muhtemelen perde işaretlemenin bazı noktalarda yanlış yapılmasından kaynaklanmaktadır.

Çizelge 4.3 Beş kez hızlandırma-yavaşlatma sentezinden sonraki PESQ sonuç istatistikleri

TSM ALGORİTMASI	ERKEK SESİ		BAYAN SESİ		GENEL	
	ORT.	STD. SAPMA	ORT.	STD. SAPMA	ORT.	STD. SAPMA
Sinüs Modelleme(SS)	2,57	0,19	2,27	0,10	2,47	0,22
AOLA	2,50	0,17	2,20	0,14	2,40	0,22
SOLA	2,38	0,37	2,09	0,05	2,28	0,33
WSOLA	2,28	0,33	2,18	0,11	2,25	0,28
PSOLA	2,19	0,34	2,04	0,08	2,14	0,28
Faz Vokoder	2,07	0,28	1,95	0,20	2,03	0,25
PICOLA	1,96	0,20	1,76	0,08	1,89	0,19
EM-TSM	1,87	0,19	1,82	0,10	1,85	0,16
GLS-TSM	1,95	0,35	1,63	0,23	1,84	0,34
Suziki Misaki	1,69	0,15	1,51	0,06	1,63	0,15
OLA	1,58	0,12	1,52	0,05	1,56	0,10
Sinüs Modelleme(MQ)	1,27	0,20	0,98	0,13	1,17	0,22
Spektral Modelleme Sentezi	1,26	0,20	0,84	0,06	1,12	0,26

WSOLA'nın 5 kez sentezleme PESQ sonuçlarının diğerlerinden daha kötü çıkması muhtemelen sesin son kısımlarındaki kırılmadır. WSOLA örtüşmeli ekleme yapmak için benzerlik aramasını analiz sesi üzerinde gerçekleştirmektedir. Bu arama esnasında aday lokasyonun komşuluklarına bakılmaktadır. Her ne kadar arama komşuluğu 5 ms gibi kısa tutulmuş olsa da segment büyüklüğü 10 ms aday segmentler arası uzaklık hızlandırma için 20 ms yavaşlatma için ise 5 msdir. Sesin son kısımlarına gelindiğinde arama yapılacak alan kalmadığında sentez sonlandırılmakta ve kalan sesler skalası değiştirilemeden sentezlenen sesin sonuna eklenmektedir. 5 kez yavaşlatma hızlandırma

çevrimi için toplamda 10 kez sentez yapılmaktadır. Bu nedenle sesin yaklaşık son 15 ms'den öncesindeki yaklaşık 200 ms'lik kısımda eksilme görülmektedir.

Hızlandırma yavaşlatma çevriminin birden fazla uygulanması işitsel kalite değerlendirmesini kolaylaştırmaktadır. Diğer bir açıdan, iletişimin noktadan noktaya bağlantı ile değil de internet, GSM vb. ağlardaki farklı ağ kanallarından aktığı durumda hızlandırma yavaşlatma da ağ noktaları arasında sadece tıkanık kanallarda uygulanabilir. Ses iletişiminin yapıldığı ağ birden fazla tıkanık kanal içeriyor ise her kanalın gönderen noktasında hızlandırma yapılıp, alıcı noktasında yavaşlatma ile orijinal skalaya dönüleceğinden hızlandırma/yavaşlatma çevrimi sayısı tıkanık kanal sayısı kadar olacaktır. Çizelge 4.3'te yer alan PESQ skorlarına göre tıkanık kanal sayısının 5 adet olması (ara sayılarda değerlendirmeler yapılmamıştır) durumunda kalite açısından kabul edilebilir olmayacaktır.

PESQ istatistiklerinden çıkan bir başka sonuç da WSOLA haricindeki tüm algoritmalar yüksek perde periyotlu erkek sesini daha iyi sentezlemektedir. Zaman alanı teknikleri benzerlik araması ve örtüşmeli ekleme tekniklerini uygulamaktadır. Benzerlik aramasında kullanılan segment ve bulunan benzerliğe göre örtüşmeli ekleme yapılan segment büyüklüklerinin perde periyotlarına oranı bayan sesi için daha büyüktür. Bu oran bayan sesi için daha çok perdenin örtüşmesine neden olmaktadır. Frekans alanı tekniklerinde (modelleme teknikleri de dahil) FFT segmenti büyüklüğü 512 seçilmiş olup analiz segmenti büyüklüğü yaklaşık 30ms'dir. Bayan sesi bu segmentte de daha fazla perde periyotu içermektedir. WSOLA'da ise arama segmenti büyüklüğü ve örtüşmeli ekleme 10 ms olup daha az perde içermektedir.

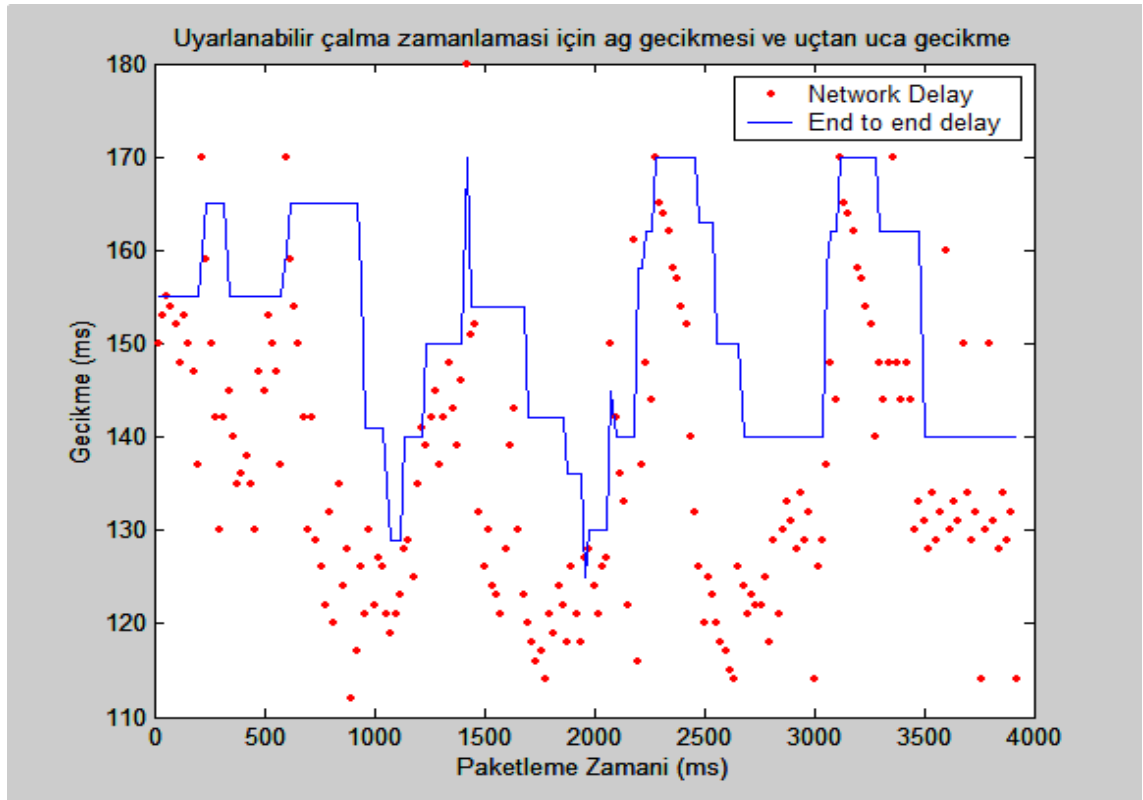
WSOLA'da segment büyüklüğü 10 ms'den 5 ms'ye düşürüldüğünde sentezlenen seslerin ortlamaları 1 çevrim için hemen hemen aynı çıkmıştır. Fakat 5 çevrim için bayan sesleri için PESQ ortalaması artmış erkek sesleri için PESQ ortalaması düşmüştür. Bu sonuçlar çizelge 4.4'te yer almaktadır.

Çizelge 4.4 5ms segment büyüklüğüne göre WSOLA PESQ sonuç istatistikleri

HIZLANDIRMA / YAVAŞLATMA SAYISI	ERKEK SESİ		BAYAN SESİ		GENEL	
	ORT.	STD. SAPMA	ORT.	STD. SAPMA	ORT.	STD. SAPMA
1 Çevrim	3,26	0,21	3,31	0,12	3,28	0,18
5 Çevrim	2,24	0,28	2,27	0,31	2,25	0,28

4.2 Uyarlanabilir Çalma Zamanlaması ve Paket Kaybı Gizlemesi Değerlendirmeleri

Testler gerçek zamanlı olarak uygulanmadığından ağ gecikmesi, uçtan uca gecikme tahminleri ve paket kayıpları için veri modeli oluşturulmuştur. Bu veri modeli şekil 4.28'deki grafikte görülmektedir.



Şekil 4.28 Simülasyonlar için veri modeli

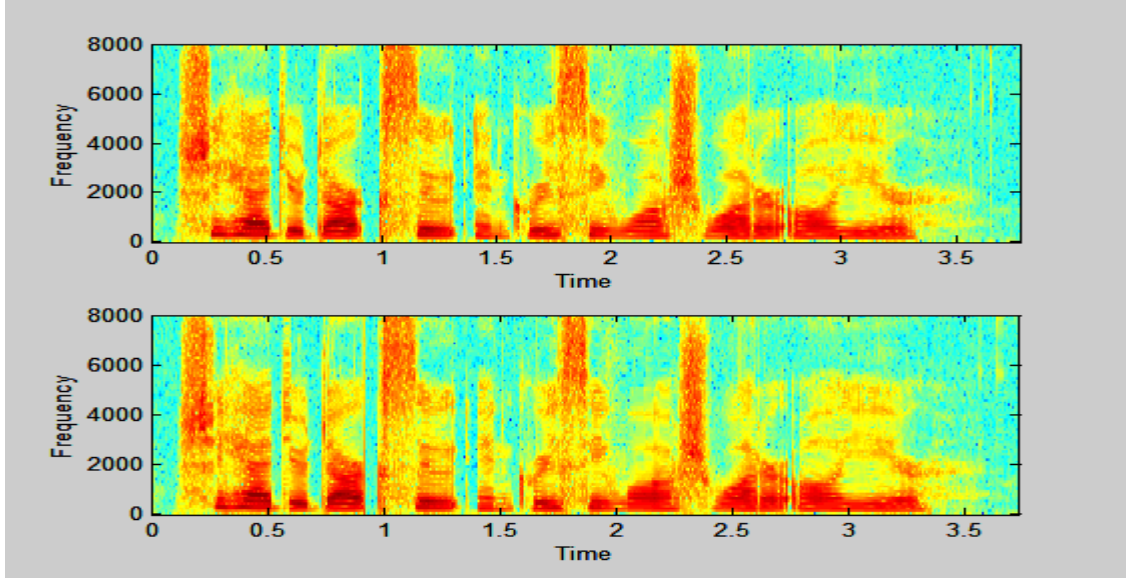
Şekil 4.28’de mavi çizgi ile gösterilen uçtan uca gecikme çizgisinin üstünde yer alan kırmızı noktalı ağ gecikmeleri paket kayıplarına neden olmaktadır. Bu veri modeli oluşturulurken uyarlanabilir çalma zamanlaması algoritmalarının ani gecikme artışlarında genelde ilk paketi kaçırmaması ve ağ kayıpları da kurgulanmıştır. Paket kaybından sonra gecikme tahmini artırılmıştır. Daha sonra ağ gecikmelerine göre düşürülmesi modellenmiştir.

Zaman skalası değişimi algoritmalarının karşılaştırması için simülasyon modeli aşağıdaki gibi uygulanmıştır.

- Ses dosyasından (TİMİT veritabanından) alınan ses 20 ms’lik paketlere bölünmüştür. 20 mslik paketleme internette en çok kullanılan ses paket büyüklüğüdür.
- Her paket için ağ gecikmesi ve bir sonraki paketin uçtan uca gecikme tahmini değerleri girilmiştir. Bu değerlerin girilmesinde gerçek veriler kullanılmamış, şekil 2.22.c örnek alınmıştır. Paket kaybı için çok yüksek ağ gecikmesi verilmiştir.
- Tüm zaman skalası algoritmalarında aynı gecikme ve kayıp verileri kullanılmıştır.
- Algoritmalar uygulanırken bu zamanlama ve kayıp modeline göre her bir paketin uzatılma ve kısaltılma süreleri bulunmuştur. Kayıp paketler için önceki ve sonraki paketlerden (Liang vd. (2003) tarafından önerilen modele göre) sentezleme gerçekleştirilmiştir.

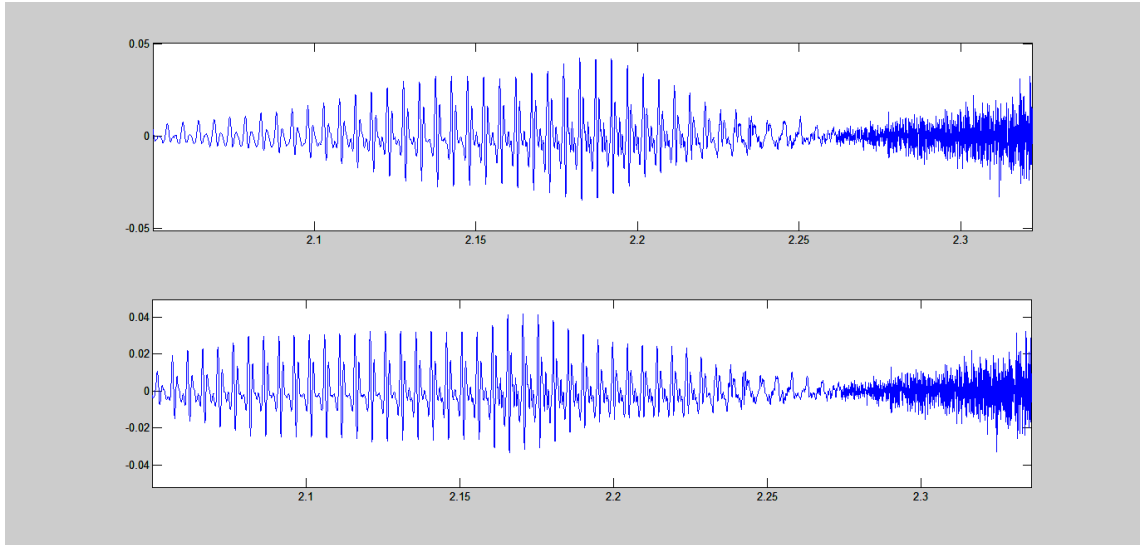
4.2.1 WSOLA ile paket kaybının gizlenmesinde gecikmenin azaltılması

Simülasyonda TIMIT veritabanından alınan bir bayan sesi kullanılmıştır. 16 KHz ile örneklenmiş olan bu ses üzerinde örneklem indirgeme yapılmamış, orijinali aynen kullanılmıştır.



Şekil 4.29 WSOLA ile uyarlanabilir çalma zamanlaması ve ağ kayıplarının gizlenmesi spektrumları

a. Orjinal ses, b. Uyarlanmış ses



Şekil 4.30 Gecikmeye ve paket kayıplarına göre WSOLA ile uyarlanmış sesler

a. Orjinal ses, b. Uyarlanmış ses

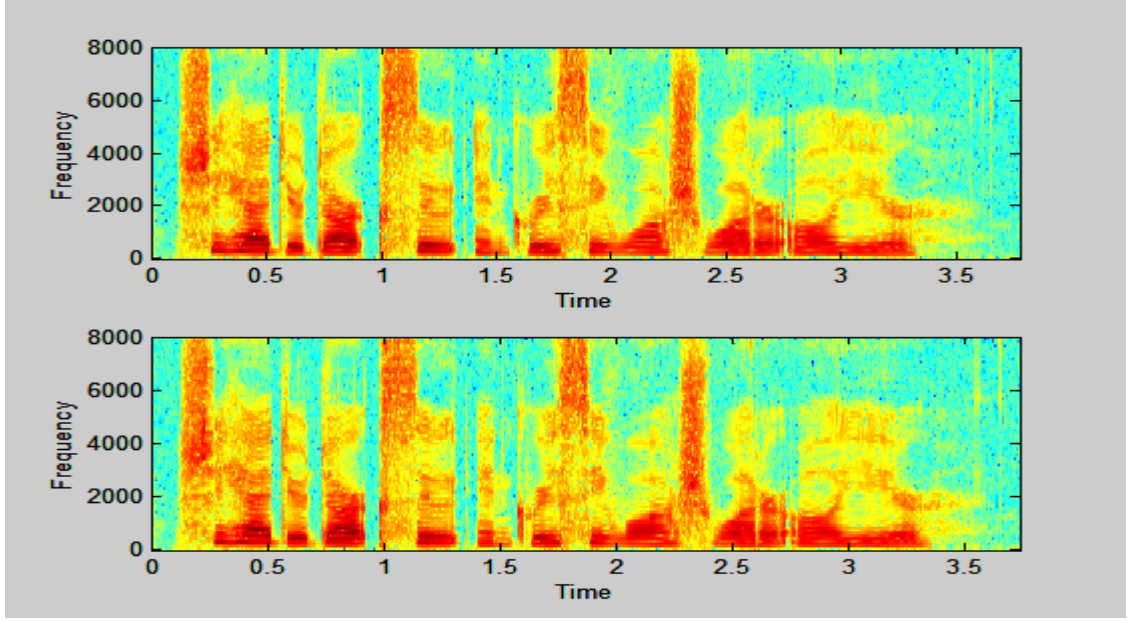
Şekil 4.29'daki spektrumlar incelendiğinde farklar zor fark edilmektedir. Bununla birlikte spektrumların x eksenlerine göre kaymalar fark edilmektedir. Şekil 4.30'da yer alan sesin zaman gösteriminde kaymalar görülmektedir. Veri modeline göre analiz sesinin en başında uçtan uca gecikme yaklaşık 155ms'dir. Bununla birlikte 2000 ms'de

gecikme yaklaşık 125 ms olup analiz sesine göre 30 ms sıkıştırma yapılmıştır. Daha sonra uçtan uca gecikme ilk başlarda 130 ms'ye daha sonra da 170 ms'ye çıkmaktadır. Paket kayıplarının dışında uçtan uca gecikme süresinin de zaman skalası değişimi ile sentezlenmesi yapılmaktadır. Bundan dolayı şekil 4.30'da görülen ses sinyallerinde orijinal sinyalin 2180ms ve 2280ms'deki paketler geç ulaştığından dolayı daha öncesi sentezlenmiştir.

WSOLA algoritmasında ardışık segmentler arasında benzerlik aranırken aday segment lokasyonlarının belirli bir komşuluk dâhilinde kayabilmesi gerekmektedir. Erkek ve kadın sesinin perde periyotları arasındaki farktan dolayı benzerlik arama benzerlik arama alanının en az 8 ms (erkek ve kadın perde periyotları arasındaki farktan dolayı); aday segmentlerin uzunlukları ve ardışık segment adayları arasındaki fark da dikkate alındığında 20 ms çok kısıtlı bir alandır. Bu nedenle fazla genişleme ve fazla daralma ancak tekrarlı olarak gerçekleştirilebilmektedir. En yüksek genişleme 2,3 kat, en düşük daralma 0,3 kat ile sınırlandırılmasına rağmen bu değerlerde bile genelde aynı örnekler örtüşmeli eklemede kullanılabilmektedir.

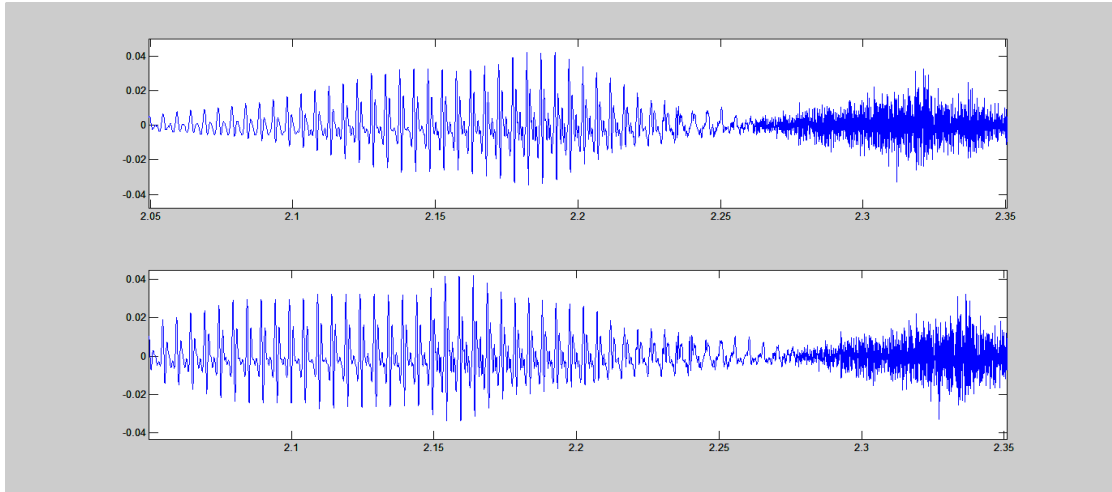
4.2.2 SOLA paket kaybının gizlenmesinde gecikmenin azaltılması

SOLA ile sentezlenen ses dinlendiğinde WSOLA'da olduğu gibi herhangi bir kalite bozukluğu tespit edilememiştir. Örnek spektrum ve yakınsanmış zaman alanı genlik grafikleri şekil 4.31 ve 4.32'de yer almaktadır.



Şekil 4.31 SOLA ile uyarlanabilir çalma zamanlaması ve ağ kayıplarına göre zaman skalası değişimi simülasyon spektrumları

a. Orjinal ses, b. Uyarlanmış ses



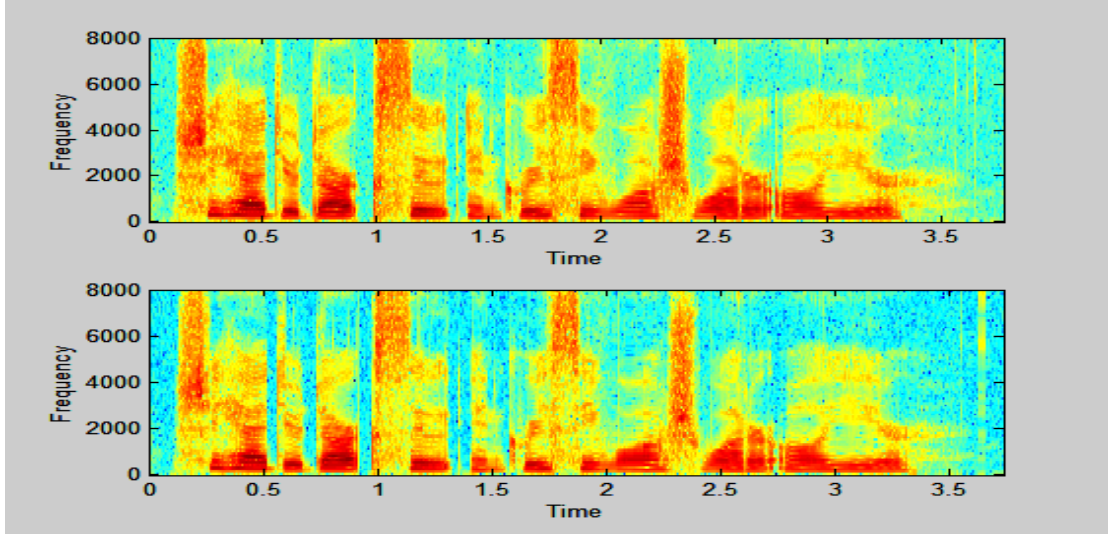
Şekil 4.32 Gecikmeye ve paket kayıplarına göre SOLA ile uyarlanmış sesler

a. Orjinal ses, b. Uyarlanmış ses

4.2.3 Paket içerisinde örtüşmeli ekleme

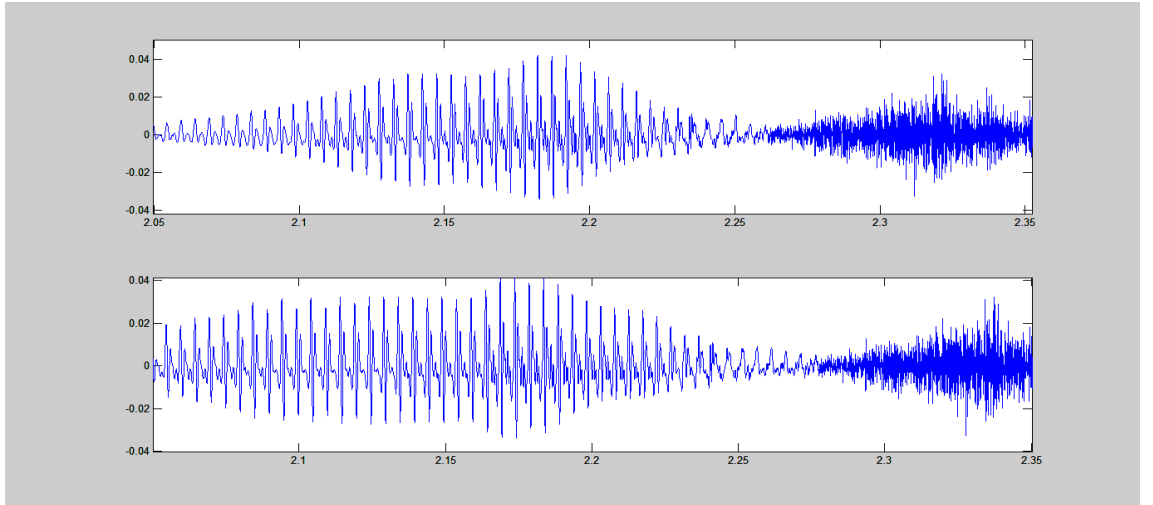
Bu algoritmaya göre yapılan testlerde de sesin algılanan kalitesinde bozukluk fark edilememiştir. Bu test ile sentezlenen sesin spektrumu şekil 4.33'te, zaman alanı örneği

olarak 2,15 ms ile 2,35 ms aralığında yakınsanmış genlik değerleri şekil 4.34'te yer almaktadır.



Şekil 4.33 Paket içerisinde örtüşmeli ekleme ile uyarlanabilir çalma zamanlaması ve ağ kayıplarına göre zaman skalası değişimi simülasyon spektrumları

a. Orjinal ses, b. Uyarlanmış ses



Şekil 4.34 Gecikmeye ve paket kayıplarına göre paket içerisinde örtüşmeli ekleme ile uyarlanmış sesler

a. Orjinal ses, b. Uyarlanmış ses

4.2.4 Uyarlanabilir alma zamanlaması ve paket kaybı PESQ deęerlendirmesi

Öznel deęerlendirmelerde orijinal ses ile sentezlenen sesler arasındaki kalite farkları ayırt edilememiştir. TIMIT veri tabanından alınan ve sıkıştırma kalite deęerlendirme testlerinde de kullanılan 12 adet örnek analiz sesi için bu üç algoritma da hazırlanan gecikme modeline göre test edilmiştir. Test sonuçları çizelge 4.5’te yer almaktadır.

Çizelge 4.5 Uyarlanabilir alma zamanlaması ve paket kaybı gizlemesi algoritmalarının PESQ sonuç istatistikleri

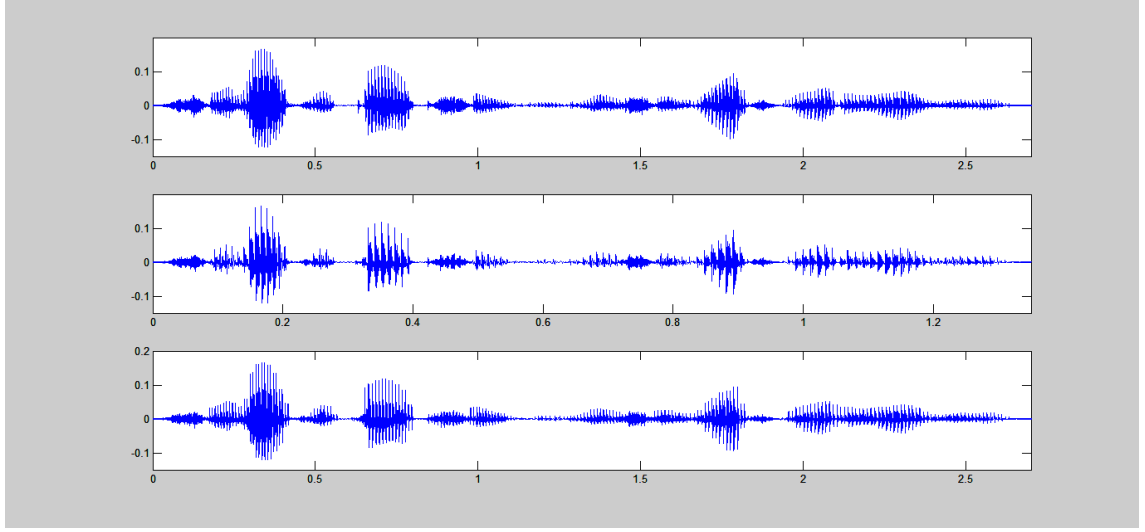
TSM ALGORİTMASI	ERKEK SESİ		BAYAN SESİ		GENEL	
	ORT.	STD. SAPMA	ORT.	STD. SAPMA	ORT.	STD. SAPMA
Önerilen Yöntem	2,82	0,34	2,55	0,19	2,73	0,32
WSOLA	2,75	0,26	2,66	0,29	2,72	0,26
SOLA	2,61	0,23	2,51	0,27	2,58	0,23

PESQ sonuçlarına göre önerilen yöntem erkek seslerinde daha başarılı iken WSOLA bayan seslerinde daha başarılı çıkmıştır.

4.3 DCT Modelleme Bulguları

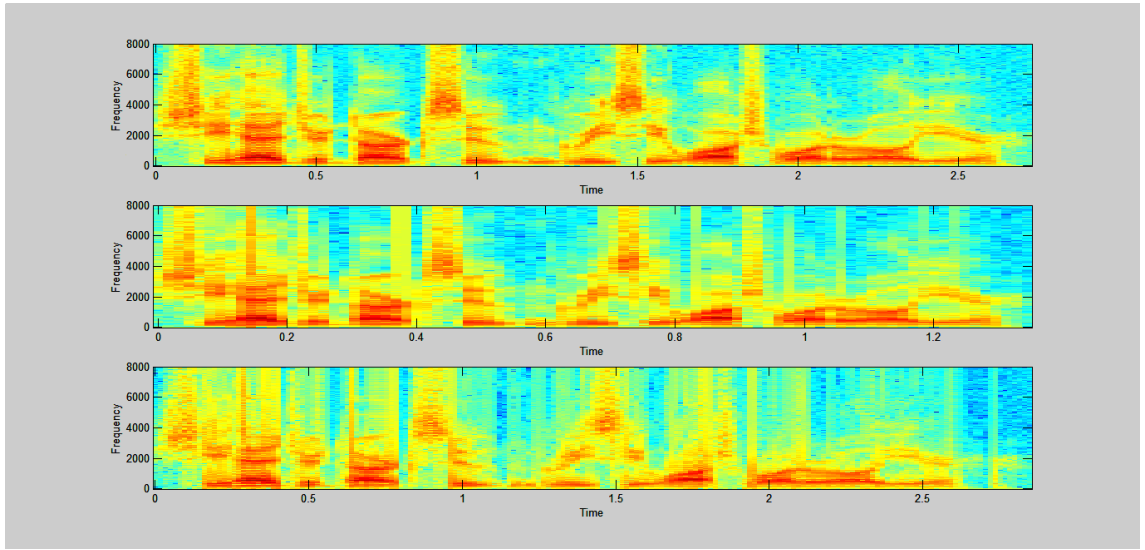
DCT modellemenin sentezleme performansı hızlandırma yavaşlatma sentezi ve paket kaybı gizlemesi için gerçekleştirilmiştir.

Bölüm 3.3’te yer alan örnek gösterimler ve şekillerde algoritmanın da açıklanabilmesi için sentez segmentinin boyutu genelde yüksek tutulmuştur. Bununla birlikte DCT modelleme sentezinin ses sıkıştırma amacıyla kullanılabilmesi için özellikle hızlandırma sentezini de başarılı yapabilmesi gerekir. Bu nedenle yapılan ilk testte örnek ses diğer algoritmalarda da olduğu gibi önce 2 kat hızlandırılmış daha sonra yavaşlatılmıştır. İlgili sentez sinyalleri ve spektrumları sırasıyla şekil 4.35 ve 4.36’da verilmektedir.



Şekil 4.35 DCT modelleme sentez sesleri

a. Orijinal ses, b. Hızlandırılmış sentez sesi, c. Hızlandırılan sesin orijinal skalaya dönüştürüldüğü sentez sesi



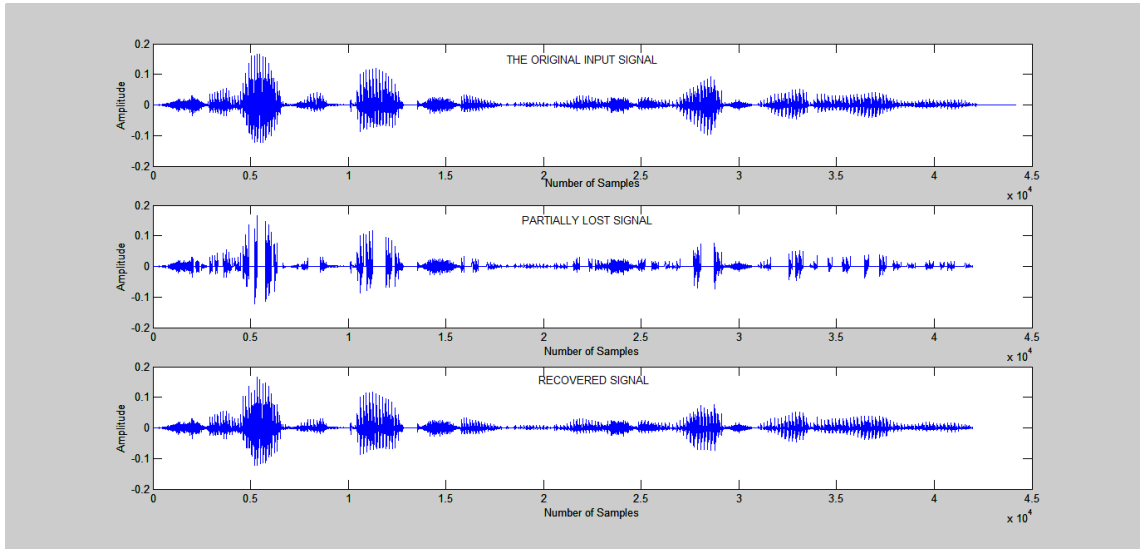
Şekil 4.36 DCT modelleme sentez sinyal spektrumları

a. Orijinal ses, b. Hızlandırılmış sentez sesi, c. Hızlandırılan sesin orijinal skalaya dönüştürüldüğü sentez sesi

Şekil 4.35’te özellikle 2,1 ve 2,2 saniye aralığındaki sessiz bölgenin sentezlenmesinde problem görülmektedir. Şekil 4.36’daki spektrum grafiğinden de bazı bozulmalar görülebilmektedir. Yine de sentezlenen ses dinlendiğinde mekanikleşmenin diğer frekans alanı tekniklerinden daha az olduğu algılanmaktadır. Algısal kalite olarak

PSOLA'ya yakın kalitede ses üretebilmektedir. Sentezlenen sesin PESQ test sonucu ise 2.836 olarak gerçekleşmiştir.

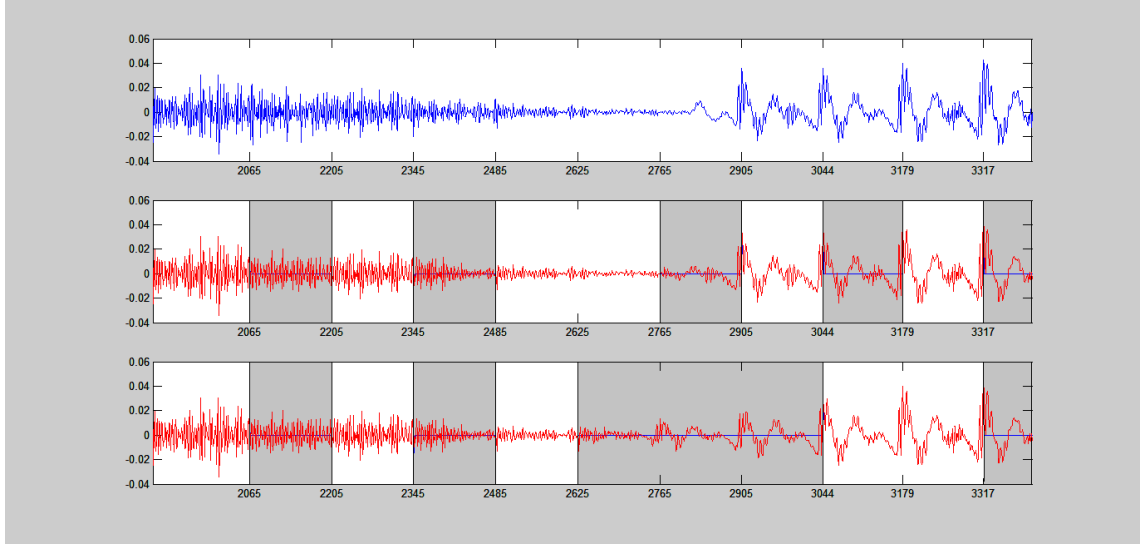
DCT modellemenin ikinci test senaryosu olarak kaybedilen perde prototiplerinin DCT modelleme ile tekrar sentezlenmesi gerçekleştirilmiştir. Bu test senaryosu için örnek grafikler şekil 4.37'de gösterilmektedir.



Şekil 4.37 DCT modelleme ile kayıp paketlerin sentezlenmesi

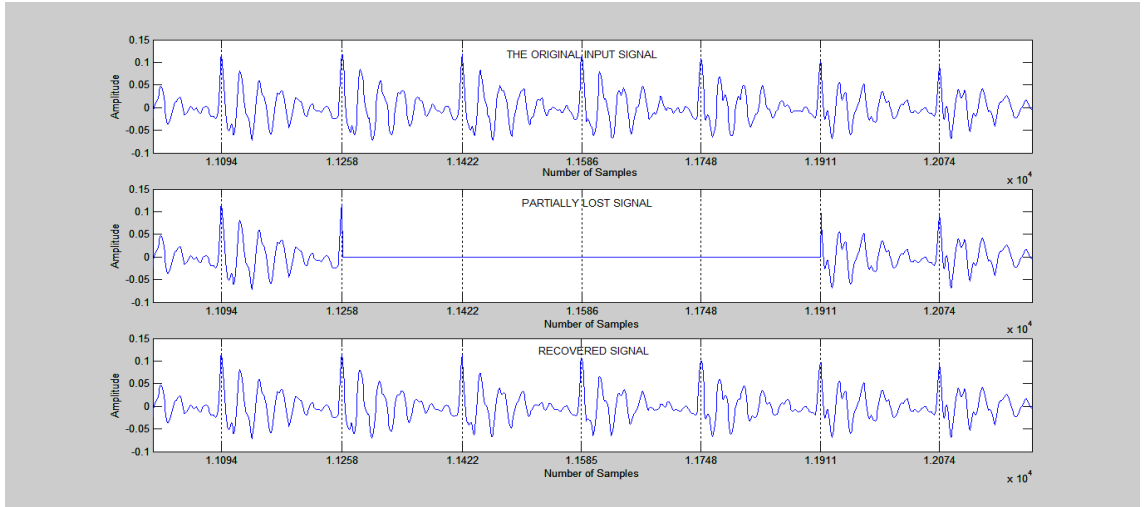
a. Orijinal ses, b. Paket kayıplı ses, c. DCT modelleme ile kayıp paketlerin sentezlenmesi

Algoritmanın prototip sentezleme başarısı PSOLA ile karşılaştırılmıştır. Şekil 4.37.b'de gösterilen kayıplı ses sinyalinin iyileştirilmesi için PSOLA PESQ sonucu olarak 2.667 vermekte iken değiştirilmiş DCT 3.222 vermiştir. Karşılaştırma için Bölüm 3.2 altında tanımlanan ve Bölüm 4.2 altında testleri gerçekleştirilen yöntemler yerine PSOLA seçilmesinin nedeni PSOLA'nın da hem zaman zaman hem perde skalasını değiştirebilmesidir. Bu testte, kaybolan prototiplerin perde boyutları komşu prototiplerin perde boyutlarından büyük ya da küçük olabildiğinden perde değişimi de gerçekleştirilmektedir. Her iki yöntemde de perde işaretlemesi aynı yöntemle gerçekleştirilmiştir.



Şekil 4.38 Orijinal ve prototipleri tekrar sentezlenen sinyallerin yakınsanmış parçaları

a. Orijinal ses sinyali, b. Ötümsüz, ötümsüz-ötümlü geçişi ve ötümlü bölgede tek prototiplerin sentezlenmesi, c. Ötümsüz-ötümlü bölgede 3 prototiplik sesin sentezlenmesi



Şekil 4.39 Yaklaşık 40 ms'lik kaybın olduğu sinyal bölgesine yakınsama

5. SONUÇ

Literatürde çok sayıda TSM algoritması mevcuttur. TSM algoritmaları genelde zaman alanı teknikleri, frekans alanı teknikleri ve sinyal modelleme teknikleri olarak gruplandırılmaktadır. Her ne kadar bu algoritmaların işlem karmaşıklığı, gecikmesi ve diğer birçok özellikleri biliniyorsa da sentezleme kalitelerinin birbirleri ile karşılaştırılması üzerine çalışmalar sınırlıdır.

Bu tez kapsamında TSM algoritmaların sesin tıkanık ağlardan veya haberleşme kanallarından iletimi ve algısal kalitesinin korunması için kullanımı ele alınmıştır.

5.1 Ses Sıkıştırma İçin TSM Algoritmasının Seçimi

TSM algoritmaları sesin hem LPC vb. yakın dönem ilişkilerini hem de perde peryodu, temel frekans gibi nispeten daha uzak dönem ilişkilerini korumaktadır. Bu özellikleri nedeniyle ses kodekleri ile birlikte kullanımı da mümkündür. Sesin ağ veya ses kanalından iletilmeden önce TSM algoritmaları ile hızlandırılması ses süresini azaltmaktadır. Ses süresinin azalması ile ihtiyaç duyulan kanal kapasitesi de azalmakta ve ses hızlandırma oranında sıkıştırılmaktadır. Sıkıştırılan ses karşı tarafa iletdikten sonra aynı oranda yavaşlatılarak gerçek zamanlı dinlenebilir hale getirilmektedir. Kodekler ile birlikte kullanımında kodeklerin sıkıştırma performansı da hızlandırma oranında artmaktadır. Bu yapıda bir kullanım, kodeklerin MOS değerlerini de TSM algoritmalarının hızlandırma ve yavaşlatma sentez kalitesi oranında düşürmektedir. Bu nedenle sentezleme kalitesi seçilecek TSM algoritması için en önemli kriterlerden birisidir. Çizelge 4.2 ve 4.3'e göre WSOLA en iyi sentez kalitesi vermektedir.

PESQ sonuçlarına göre Smith ve Serra (1987) tarafından önerilen sinüs modelleme ses sıkıştırması için WSOLA'dan sonra en iyi sonuç veren ikinci algoritma olarak görülmektedir. Bununla birlikte sinüs modelleme ile sentezlenen ses dinlendiğinde az da olsa bir koro etkisi olduğu hissedilmektedir. Yine de beklenenin aksine sinüs modellemenin ses kalitesi öznel değerlendirmeler ile de diğer bazı zaman alanı tekniklerinden daha iyi çıkmıştır.

TSM algoritmalarının sentezleme kalitesinin yanında algoritma gecikmesi ve işlem gücü ihtiyacı da önemli parametrelerdendir. Tüm TSM algoritmaları kodeklerin gecikmesini en az sıkıştırma oranı katı olarak artırmaktadır. Örneğin bir kodek 10 mslik paketleme yaptığında 10 mslik sesin TSM ile 2 kat hızlı olarak sentezlenebilmesi için en az 20 mslik bir analiz sesine ihtiyaç duyulmaktadır. Benzerlik arama ile çalışan zaman alanı tekniklerinde bu gecikme daha da artmaktadır. Frekans alanı ve sinüs modelleme teknikleri analizi frekans alanında gerçekleştirdiğinden gerekli frekans çözünürlüğüne ulaşmak için de çok fazla örnekleme ihtiyaç duyabilmektedir. Tez kapsamındaki bu tür algoritmalarda FFT boyutu olarak 512 kullanılmıştır ki bu da 30 ms'nin biraz üzerindedir. Bu nedenle sinüs modelleme, faz vokoder ve benzeri algoritmaların algoritma gecikmeleri de diğer zaman alanı tekniklerinden yüksek olmaktadır.

Zaman alanı tekniklerinde algoritma karmaşıklığı olarak da adlandırılabilen işlem gücü ihtiyacı benzerlik arama yöntemine bağlıdır. Basitleştirilmiş veya normalleştirilmiş özilinti gibi benzerlik arama yöntemlerinde de çok fazla işlem yapılmaktadır. Örneğin WSOLA algoritmasında normalleştirilmiş özilinti kullanıldığından tez çalışması kapsamında kullanılan parametrelere göre hızlandırma sentezi saniyede 2,6 milyon üzerinde, yavaşlatma sentezi ise saniyede 10,4 milyon üzerinde çarpma toplaması yapmaktadır (Güler 2004). Frekans alanı ve sinyal modelleme tekniklerinde ise frekans spektrumuna dönüş saniyede 144000 çarpma toplamasına eşittir. Bu nedenle faz vokoder sadece faz ve genlik enterpolasyonu yaptığından bu tez çalışmasında kullanılan parametrelere göre WSOLA'dan en az 5 kat daha hızlı sentezleme yapmaktadır. Spektral genlik modülasyonlu Sinüs modelleme ise WSOLA'dan daha yavaş çalışmaktadır. PSOLA da perde aramasını ağırlıklı olarak tepe noktalara göre yaptığından WSOLA'dan çok daha hızlı çalışmaktadır.

5.2 Uyarlanabilir Çalma Zamanlaması ve Paket Kaybı Gizlemesi için TSM Tekniğinin Seçimi

Gerçek zamanlı ses iletiminde çalma zamanlaması ağ gecikmesindeki değişimler dikkate alınarak hedeflenen uçtan uca gecikmeye göre belirlenir. Sabit çalma zamanlamasının kullanıldığı bir ses iletişiminde ağ gecikmesinin artması alıcı ucunda çalma zamanlaması tampon belleğinin boşalmasına ve dinlenen seste boşluk oluşmasına

neden olmaktadır. Bu dinlenen sesin kalitesini bozmaktadır. Uçtan uca gecikmenin çok yüksek tutulması ise karşılıklı iletişimde rahatsızlık verebilmektedir. Uyarlanabilir çalma zamanlamasında uçtan uca gecikme ağdaki gecikme değişimlerine göre değişmektedir. Uçtan uca gecikmenin artırılması veya azaltılması çalma zamanlaması tampon belleğinde yer alan seslerin daha yavaş veya daha hızlı çalınması ile mümkündür. Bu nedenle uyarlanabilir çalma zamanlaması için TSM teknikleri kullanılarak tampon belleğin istenilen hedef uçtan uca gecikmeye göre boşaltılması veya doldurulması sağlanır.

Tampon bellekte bulunan seslerin yeterince yavaşlatılması durumunda paket kayıpları da sentezlenebilmektedir. Sentezlenmiş olan ses ile kayıptan sonra alınacak paketteki sesin uyumlu olarak birleştirilmesi de gerekmektedir.

Uyarlanabilir paket çalması ve paket kaybının gizlenmesi için WSOLA ve SOLA en çok kullanılan TSM tekniklerindendir. Bu çalışma kapsamında tek paket ile sentezleme yapmak için alternatif bir yöntem de önerilmiştir. Oluşturulan kayıp ve gecikme modellerine göre TIMIT ses veritabanından alınan seslere WSOLA, SOLA ve önerilen teknik uygulanmıştır. PESQ sonuçlarına göre erkek sesleri için oluşturulan model daha iyi ses kalitesi sunmaktadır. Bayan sesleri için WSOLA daha iyi sonuç vermektedir.

WSOLA'nın benzerlik aramasında ardışık paketlerin her ikisi de aday lokasyonlarından sapabilmektedir. Bu nedenle benzerlik arama uzaklıkları kontrol edilememektedir. Önerilen teknikte benzerlik aranan segmentler arası uzaklık olabilecek en düşük perde peryodu ile en yüksek perde peryodunu içerecek şekilde seçilmektedir. WSOLA'dan daha az kayma ile benzerlik arayabildiğinden daha hızlıdır.

5.3 Değiştirilmiş DCT ile Sentezleme

DCT dönüşümü, Fourier dönüşümünün aksine gerçel sinyalleri, gerçel sayılar ile ifade etmektedir. Bu sayede fazın uyumlandırılması için kübik spline enterpolasyonu ve faz katlama katsayısının bulunması gibi karmaşık hesaplamalara gerek kalmamaktadır. DCT dönüşümünün diğer bir avantajı ise enerji paketleme verimliliğidir. Markov süreci

özellikleri gösteren ses gibi sinyaller için DCT, Karhunen-Loeve dönüşümüne yakınsayarak ortogonal dönüşüm özellikleri göstermektedir. Spektrum değerleri başlangıç noktasından uzaklaştıkça sifıra yaklaşmaktadır. Bu özelliği sayesinde DCT katsayılarının kırılması ile çözünürlüğünün azalması durumunda enerji kaybı, sinüs modelleme veya Fouirer dönüşümüne göre daha azdır. DCT katsayılarının kırılması veya sonuna sıfır değerlikli katsayıların eklenmesi ile sesin perdesi de değiştirilebilmektedir.

DCT'nin ters dönüşümü ile bir periyottan daha uzun sinyal sentezleme aynasal simetri göstermektedir. Bu aynasal simetrinin düzeltilmesi ile DCT ile zaman skalası değişimi de gerçekleştirilebilmektedir. İki ayrı prototipin DCT katsayıları örtüşmeli ekleme yapılarak sinyalin bir prototipten diğer prototipe geçişi istenilen uzunlukta bir sentezleme ile yumuşak olarak gerçekleştirilebilmektedir. DCT'nin bu özelliklerinden faydalanılarak oluşturulan modele göre ses skalası değiştirilebilmektedir.

DCT modelleme ile yapılan testlerde zaman alanı tekniklerine oldukça yakın kalitede ses sıkıştırması gerçekleştirilmiştir. Sinüs modellemenin aksine sentezlenen ses spektrumu yayvanlaştırmamakta ve dinlendiğinde koro etkisi yaratmamaktadır. Fakat ses sıkıştırma için PESQ sonucu 2.836 olarak PSOLA'nın hemen altında yer almıştır. Bununla birlikte prototip enterpolasyonu için yapılan testlerde PSOLA'dan çok daha iyi kalitede sonuç vermiştir.

KAYNAKLAR

- Agnihotri, S., Aravindhan, K., Jamadagni, H. S. and Pawate, B. I. 2002. A new technique for improving quality of speech in voice over IP using time-scale modification. In Proceedings of IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP 2002), Orlando.
- Anonymous. 1998. Information Technology - Very Low Bitrate Audio-Visual Coding. ISO/IEC-14496-3 standard.
- Anonymous. 1990. The DARPA TIMIT Acoustic-Phonetic Continuous Speech Corpus. NIST.
- Anonymous. 1996. ITU-T Recommendation P.800 Methods for subjective determination of transmission quality. ITU-T, 37 p.
- Anonymous. 2001. ITU-T Recommendation P.862 Perceptual evaluation of speech quality (PESQ): An objective method for end-to-end speech quality assessment of narrow-band telephone networks and speech codecs. ITU-T, 30 p.
- Anonymous. 2003. ITU-T Recommendation G.114 One-way transmission time. ITU-T, 20 p.
- Bonada, J. 2002. Time-scale modification of audio in the context of professional audio post-production. Ph. D. Thesis, Universitat Pompeu Fabra; 84 p., Barcelona.
- Charpentier, F. and Moulines, E. 1990. Pitch-synchronous waveform processing techniques for text-to-speech synthesis using diphones. Speech Communication, vol. 9(5/6); pp. 13-19.
- DeLeon, P. and Sreenan, C.J. 1999. An adaptive predictor for media playout buffering. In Proceedings of IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP 1999), vol. 6; pp. 3097-3100.
- Demol, M., Struyve, K., Verhelst, W., Paulussen, H., Desmet P. and Verhoeve, P. 2004. Efficient non-uniform time-scaling of speech with WSOLA for CALL applications. Proceedings of InSTIL/ICALL Symposium, Venice.
- Dolson, M. 1986. The phase vocoder: A tutorial. Computer Music Journal, vol.10; pp. 14-27.
- Donnellan, O., Elmar J. and Coyle, E. 2003. Speech-adaptive time-scale modification for computer assisted language-learning. The 3rd IEEE International Conference on Advanced Learning Technologies; pp. 165–169.
- Dorran D. and Lawlor, R. 2003. Time-scale modification of music using a subband approach based on the bark scale. IEEE Workshop on Applications of Signal Processing to Audio and Acoustics; pp. 173-176, Newyork.

- Dorran, D. 2006. Audio Time Scale Modification. Ph. D. Thesis, Dublin Institute of Technology, 188 p., Dublin.
- Dorran, D., Lawlor, R. and Coyle, E. 2006. A Comparison of Time-Domain Time-Scale Modification Algorithms. Audio Engineering Society, 120th Convention, Paris.
- Dudley, H. 1939. "The Vocoder". Bell Labs, Record 18; pp. 122–126
- Erogul, O., İlk, H.G. and İlk, O. 2001. A flexible bit rate switching method for low bit rate vocoders. Signal Processing, Vol. 81-8; pp.1737-1742.
- Erogul, O. and Karagoz, I. 1998. Time-scale modification of speech signals for language-learning impaired children. Proceedings of the 2nd International Conference Biomedical Engineering; pp. 33–35.
- Flanagan J. L. and Golden R. M. 1966. Phase vocoder. Bell System Technical Journal, vol. 45; pp. 1493-1509.
- Goodman, D.J., Lockhart, G.B., Wasem, O.J. and Wong, W. 1986. Waveform substitution techniques for recovering missing speech segments in packet voice communications. IEEE Transactions on Acoustics, Speech and Signal Processing, 34(6); pp. 1449-1464.
- Goodman, D.J., Wasem, O.J., Dvorak, C.A. and Page, H.G. 1988. The effect of waveform substitution on the quality of PCM packet communications. IEEE Transactions on Acoustics, Speech and Signal Processing, 36(3); pp. 342-348.
- Griffin, D. W. and Lim, J. S. 1984. Signal estimation from modified short-time Fourier transform. IEEE Transactions on Acoustics, Speech and Signal Processing, 32(2); pp. 236-243.
- Griffin, D. W. and Lim, J. S. 1985. A new model based speech analysis/synthesis system. IEEE International Conference on Acoustics, Speech, and Signal Processing; pp. 513-516.
- Güler, S. 2004. Paket anahtarlama ağı üzerinden konuşma iletim parametreleri. Master Tezi. Ankara Üniversitesi, 83 sayfa, Ankara.
- Hejna, D. J. 1990. Real-time time-scale modification of speech via the synchronized overlap-add algorithm. Ms. Thesis, M.I.T, 126 p., California.
- İlk, H.G. and Güler, S. 2006. Adaptive Time Scale Modification of Speech for Graceful Degrading Voice Quality in Congested Networks for VoIP Applications. Signal Processing, 86; pp. 127 –139.
- İlk, H.G. and Güler, S. 2011. Signal transformation and interpolation based on modified DCT synthesis. Digital Signal Processing, Article in Press, DOI:10.1016/j.dsp.2011.01.008.

- Jung, Y. and Atwood, J. W. 2006. Dynamic adaptive playout algorithm using interarrival jitter and dual use of alpha. IEE Proceedings on Communications, Vol. 153(2); pp. 279--287.
- Laroche, J. 1993. Autocorrelation method for high-quality time/pitch-scaling. IEEE Workshop on Applications of Signal Processing to Audio and Acoustics; pp. 131–134.
- Laroche, J. and Dolson, M. 1999. Improved phase vocoder. IEEE Transactions on Speech and Audio Processing, vol. 7(3); pp. 323 –332.
- Lawlor, B. and Fagan, A.D. 1999. A novel high quality efficient algorithm for time-scale modification of speech”, Eurospeech, 6th Conference on Speech Communication and Technology, Budapest.
- Lawlor, R. 1999. Audio time-scale and frequency-scale modification. PH. D. Thesis. Dublin University College, 313 p., Dublin.
- Lee, F. 1972. Time compression and expansion of speech by the sampling method. Journal of the audio engineering society; pp. 738 –742."
- Liang, Y.J., Farber, N. and Girod, B. 2001. Adaptive playout scheduling using time-scale modification in packet voice communications. IEEE International Conference on Acoustics, Speech and Signal Processing, Proceedings of ICASSP 01, vol. 3; pp. 1445-1448.
- Liang, Y.J., Farber, N. and Girod, B. 2003. Adaptive playout scheduling and loss concealment for voice communication over IP networks. IEEE Transactions on Multimedia, vol. 5(4); pp. 532 - 543.
- Lindblom, J. and Hedelin, P. 2002. Packet loss concealment based on sinusoidal extrapolation. IEEE International Conference on Acoustics, Speech, and Signal Processing, vol.1; pp. 173-176.
- Liu, F., Kim, J.W. and Kuo, C.-C.J. 2001. Adaptive delay concealment for Internet voice applications with packet based time-scale modification. IEEE International Conference on Acoustics, Speech, and Signal Processing, vol. 3; pp. 1461-1464.
- Macon, M.W. and Clements, M.A. 1996 “Speech concatenation and synthesis using an overlap-add sinusoidal model. IEEE International Conference on Acoustics, Speech, and Signal Processing, vol. 1; pp. 361 – 364.
- Makhoul, J. and El-Jaroudi, A., 1996 Time-scale modification in medium to low rate speech coding. IEEE International Conference on Acoustics, Speech, and Signal Processing., vol. 11; pp.1705 – 1708.
- Malah, D. 1979. Time-domain algorithms for harmonic bandwidth reduction and time scaling of speech signals. IEEE Transactions on Acoustics, Speech and Signal Processing, vol. ASSP-27(2); pp. 121–133.

- Marezka, F. 2008. Packet loss concealment using time scale modification for CELP based coders in packet network. 40th Southeastern Symposium on System Theory; pp. 84-87.
- McAulay, R. J. and Quatieri, T. F. 1986. Speech analysis/synthesis based on a sinusoidal representation. IEEE Transactions on Acoustics, Speech and Signal Processing, vol. ASSP-34(4); pp. 744 – 754.
- McAulay, R. J. and Quatieri, T. F. 1986. Speech transformation based on a sinusoidal representation. IEEE Transactions on Acoustics, Speech and Signal Processing, vol. ASSP-34(6); pp. 1449-1464.
- Mojia, L., Muqing, W., Dapeng, W., Lizhong, W. and Chunxiu, X. 2009. Packet loss concealment using enhanced waveform similarity overlap-and-add technique with management of gains. 5th International Conference on WiCom; pp 1-4.
- Moon, S.B. Kurose, J. Towsley, D. 1998. Packet audio playout delay adjustment: Performance bounds and algorithms. Multimedia Syst., vol. 6, no.1; pp.17–28
- Morita, N. and Itakura, F. 1986. “Time-scale modification algorithm for speech by use of pointer interval control overlap and add (PICOLA) and its evaluation”, Proc. Of Annual Meeting of Acoustical Society of Japan,
- Moulines, E. and Charpentier, F. 1990. Pitch-Synchronous Wave-form Processing Techniques for Text-to-Speech Synthesis Using Di-phones. Speech Communication, vol. 9(5/6); pp. 453-467.
- Moulines, E., Charpentier, F. and Hamon, C. 1989. A diphone synthesis system based on time-domain prosodic modifications of speech. ICASSP; pp. 238-241.
- Moulines, E. and Charpentier, F. 1988. Diphone synthesis using multipulse LPC technique. Proceedings 7th FASE symposium of Speech '88; pp. 47-53. Edinburgh.
- Ninness, B. and Henriksen, S. 2000. Time and frequency scale modification of speech signals. IEEE International Conference on Acoustics, Speech, and Signal Processing, vol. 3; pp. 1295 – 1298.
- Omoigui, N., He, L., Gupta A., Grudin, J. and Sanocki, E. 1999. Time-compression: Systems concerns, usage, and benefits. CHI Conference Proceedings; pp. 136-143.
- Pallone G., Boussard P., Daudet L., Guillemain P. and Kronland-Martinet R., 1999. A wavelet based method for audio-video synchronization in broadcasting applications. Proceedings of the International Conference on Digital Audio Effects, Norway.
- Pinto, J. and Christensen, K.J. 1999. An algorithm for playout of packet voice based on adaptive adjustment of talkspurt silence periods. Proc. 24th Conf. Local Computer Networks; pp. 224–231.

- Portnoff, M.R. 1976. Implementation of the digital phase vocoder using the fast Fourier transform. *IEEE Transactions on Acoustics, Speech and Signal Processing*, vol. ASSP-24; pp. 243-248.
- Puckett, M.S. 1995. Phase locked vocoder. *IEEE Workshop on Applications of Signal Processing to Audio and Acoustics*; pp. 222 –225.
- Ramjee, A., Kurose, J., Towsley, D. and Schulzrinne, H. 1994. Adaptive playout mechanisms for packetized audio applications in wide-area networks. In *Proceedings of the Conference on Computer Communications (IEEE Infocom)* Toronto, Canada, June 1994.
- Roads, C. 1978. Granular Synthesis of Sound. *Computer Music Journal*, vol. 2; pp. 61-62.
- Rødbrø, C.A., Christensen, M. G., Andersen, S.V. and Jensen, S.H. 2003. Compressed domain packet loss concealment of sinusoidally coded speech. *IEEE International Conference on Acoustics, Speech, and Signal Processing*, vol.1; pp. 104-107.
- Rødbrø, C.A. Jensen, S.H. 2002. Time-scaling of sinusoids for intelligent jitter buffer in packet based telephony. In *Proc. 2002 IEEE Workshop on Speech Coding*; pp.71-73.
- Roucos, S. and Wilgus, A.M. 1985. High quality time-scale modification for speech. *IEEE International Conference on Acoustics, Speech and Signal Processing*; pp. 493-496.
- Sanneck, H., Stenger, A., BenYounes, K. and Girod B. 1996. A new technique for audio packet loss concealment. In *IEEE GLOBE-COM*; pp. 48-52.
- Seneff, S. 1982. System to independently modify excitation and/or spectrum of speech waveform without explicit pitch extraction. *IEEE Transactions on Acoustics, Speech, and Signal Processing*, vol. 30(4); pp. 566 – 578.
- Serinken, N., Gagnon, B. and Eroğul, O. 1997. Full-duplex speech for HF radio systems", *IEEE Seventh International Conference of HF Radio Systems and Techniques*; pp. 281-284.
- Serra, X. and Smith, J. O. 1990. Spectral modeling synthesis: A sound analysis/synthesis system based on a deterministic plus stochastic decomposition. *Computer Music Journal*, vol. 14(4); pp. 12 – 24.
- Sethares, A.S. 2007. *Rhythm and Transforms*, Sprienger, 336p, Madison.
- Shallwani, A. and Kabal, P. 2003. An adaptive playout algorithm with delay spike detection for real-time VoIP. *IEEE CCECE*, vol 2; pp. 997-1000.

- Smith, J. and O. Serra, X. 1987. "PARSHL: An analysis/synthesis program for non-harmonic sounds based on a sinusoidal representation. Proceedings of the International Computer Music Conference; pp. 290 – 297.
- Stenger, A., BenYounes, K., Reng, R. and Girod, B. 1996. A new error concealment technique for audio transmission with packet loss. In Proc. European Signal Processing Conference, vol.3; pp.1965–1968.
- Suzuki, R. and Misaki, M. 1992. Time-scale modification of speech signals using cross-correlation functions, IEEE Transactions on Consumer Electronics, vol. 38(3); pp. 357 – 363.
- Suzuki, R. and Misaki, M. 1992. Time-scale modification of speech signals using cross-correlation", IEEE International Conference on Consumer Electronics; pp. 166 – 167.
- Valenzuela, R. A. and Animalu, C.N. 1989. In proceedings ICASSP; pp. 1334-1336.
- Verhelst, W. and Roelands, M. 1993. An Overlap-Add Technique based on Waveform Similarity (WSOLA) For High Quality Time Scale Modification of Speech. IEEE ICASSP 2; pp. 554-557.
- Verhelst, W. 2000. Overlap-Add Methods for Time-Scaling of Speech. Speech Communication, vol.30 (4); pp. 207-221.
- Verma, T. Meng, T. 1998. Time scale modification using a sines+transients+noise signal model. Proceedings of the Digital Audio Effects Workshop; pp. 49-52.
- Verma, T. S., Levine, S. N., and Meng, T. 1997. Transient Modeling Synthesis: a flexible analysis/synthesis tool for transient signals. Proceedings of the International Computer Music Conference.
- Wong P.H.W. 1998. Variable Speed Playback System for Speech and Audio Signals. Ms. Thesis. City University of HongKong. 145 p.
- Wong, J.W.C., Au, O.C. and Wong, P.H.W. 1998. Fast time scale modification using envelope-matching technique (EM-TSM)", Proceedings of the IEEE International Symposium on Circuits and Systems, vol. 5; pp. 550 -553.
- Wong, P.H.W. and Au, O.C. 2002. Fast SOLA-based time scale modification using modified envelope matching", IEEE International Conference on Acoustics, Speech and Signal Processing, vol. 3; pp. 3188-3191.
- Wong, P.H.W. and Au, O.C. 2003. Fast SOLA-based time scale modification using envelope matching", KAP Journal of VLSI Signal Processing-Systems for Signal, Image, and Video Technology, vol. 35(1); pp.75-90.
- Yim, S. and Pawate, B.I., "Computationally efficient algorithm for time scale modification (GLS-TSM)", IEEE International Conference on coustics, Speech, and Signal Processing , vol. 2; pp. 1009 -1012.

ÖZGEÇMİŞ

Adı Soyadı : Saadettin GÜLER

Doğum Yeri : İZMİR

Doğum Tarihi: 05/07/1974

Medeni Hali : Evli

Yabancı Dili : İngilizce

Eğitim Durumu (Kurum ve Yıl)

Lise : İzmir Atatürk Lisesi, 1991

Lisans : Bilkent Üniversitesi, Mühendislik Fakültesi, Elektrik-Elektronik Mühendisliği Bölümü, 1996

Yüksek Lisans : Ankara Üniversitesi, Fen Bilimleri Enstitüsü, Elektronik Mühendisliği Anabilim Dalı, 2004

Çalıştığı Kurum/Kurumlar ve Yıl

Bilkent Üniversitesi, Bilgisayar Merkezi, 1996-1998

Havelsan A. Ş. , 1998-2004

Hava Kuvvetleri Komutanlığı, 2004-2005

Havelsan A. Ş. , 2005-2008

A-TEL telekomünikasyon San. ve Tic. Ltd. Şti., 2009-

Yayınları (SCI ve diğer)

İlk, H. G. **Güler, S.** 2011. Signal Transformation and Interpolation Based on Modified DCT Synthesis. Digital Signal Processing, 21(6); 756-763.

İlk, H. G. **Güler, S.** 2006. Adaptive Time Scale Modification of Speech for Graceful Degrading Voice Quality in Congested Networks for VoIP Applications. Signal Processing 86(1); 127-139

Köprülü, T. **Güler, S.** and Genç, A. 2006. Security of Pipeline Systems. in Proceedings of The Second International Conference on Technologies for Homeland Security and Safety, Istanbul, Turkey.