

Türkçe Sürekli Konuşma Tanıma Sistemlerinin Internet Protokolü Üzerindeki
Başarılarının Araştırılması ve Eniyilenmesi

İsmail Uzun

DOKTORA TEZİ

Elektrik-Elektronik Mühendisliği Anabilim Dalı

Ekim 2007

Investigation and Optimization of Turkish Continuous Speech Recognition Systems
Performance on the Internet Protocol

Ismail Uzun

DOCTORAL DISSERTATION

Department of Electrical-Electronics Engineering

October 2007

Türkçe Sürekli Konuşma Tanıma Sistemlerinin İnternet Protokolü Üzerindeki
Başarılarının Araştırılması ve Eniyilenmesi

İsmail Uzun

Eskişehir Osmangazi Üniversitesi
Fen Bilimleri Enstitüsü
Lisansüstü Yönetmeliği Uyarınca
Elektrik-Elektronik Anabilim Dalı
Telekomünikasyon Bilim Dalında
DOKTORA TEZİ
Olarak Hazırlanmıştır

Danışman: Yrd. Doç. Dr. Rifat Edizkan

Ekim 2007

İsmail Uzun'un DOKTORA tezi olarak hazırladığı "Türkçe Sürekli Konuşma Tanıma Sistemlerinin İnternet Protokolü Üzerindeki Başarımlarının Araştırılması ve Eniyilenmesi" başlıklı bu çalışma, jürimizce lisansüstü yönetmeliğinin ilgili maddeleri uyarınca değerlendirilerek kabul edilmiştir.

Üye : Yrd. Doç. Dr. Rifat Edizkan

Üye : Prof. Dr. Atalay Barkana

Üye : Prof. Dr. İdris Dağ

Üye : Yrd. Doç. Dr. Cüneyt Akınlar

Üye : Yrd. Doç. Dr. Erol Seke

Fen Bilimleri Enstitüsü Yönetim Kurulu'nun tarih ve sayılı kararıyla onaylanmıştır.

Prof. Dr. Abdurrahman KARAMANCIOĞLU

Enstitü Müdürü

TÜRKÇE SÜREKLİ KONUŞMA TANIMA SİSTEMLERİNİN INTERNET PROTOKOLU ÜZERİNDEKİ BAŞARIMLARININ ARAŞTIRILMASI VE ENİYİLENMESİ

İSMAİL UZUN

ÖZET

Bu çalışmada Türkçe sürekli konuşma tanıma sistemlerinin Internet protokolu (IP) üzerinde dağıtık yapı modeli kullanılarak paket kayıplı durumlarda başarımları araştırılmış, paket kayıp telafi yöntemlerinin başarımları karşılaştırılmış ve uyarlamalı kestirim yöntemi geliştirilerek başarımın zaman içerisinde iyileştirilmesi hedeflenmiştir.

IP üzerinden ses iletimi genellikle güvenli olmayan UDP protokolu kullanılarak paketler halinde iletilir. Paketlerin iletimi sırasında yönlendirici ve ağ geçitlerindeki sıkışıklık ve iş istasyonlarındaki yük fazlalığı gibi nedenlerden dolayı kayıplar oluşabilmektedir. IP üzerinden konuşma tanıma sistemlerinin başarımlarını arttırmak için kayıp paketlerin telafi edilmesi gerekir.

Bu çalışmada, sürekli konuşma tanıma sistemi olarak ses öznitelik vektörlerinin iletildiği dağıtık yapı modeli kullanılmış ve ses öznitelik vektörleri UDP paketleri şeklinde iletilmiştir. Alıcı tarafında paket kayıp telafi sistemi tanımlanmıştır. Telafi sistemi, en uygun yöntemin seçildiği ön işlem aşaması, telafi aşaması ve uyarlamalı yöntemler için parametrelerin güncellendiği ardışık aşamalarından oluşmaktadır. Kayıp telafi sistemi için dinamik tampon bellek tasarımı yapılmış ve uygulanacak telafi yöntemi seçimi için ölçütler belirlenmiştir. Telafi yöntemi olarak aradeğerleme ve kestirim yöntemleri uygulanmıştır. Kestirim yöntemlerinde mevcut uygulamalarda ses öznitelik vektörlerinin geniş anlamda durağan kabullenmesi yapılmakta ve eğitim amaçlı seçilen veriden elde edilen ortalama değer ve ortak değışinti değerleri telafi aşamasında kullanılmaktadır. Bu çalışmada, ses öznitelik vektörlerinin zaman içerisindeki ortalama değeri ve ortak değışinti değeri uyarlamalı olarak güncellenmektedir. Böylece, telafi sistemi her yeni konuşmada kendini güncelleyerek başarımını arttırmaktadır.

Anahtar Kelimeler: Türkçe sürekli konuşma tanıma, IP üzerinden ses iletimi, paket kayıp telafisi, uyarlamalı kestirim.

INVESTIGATION AND OPTIMIZATION OF TURKISH CONTINUOUS SPEECH RECOGNITION SYSTEMS PERFORMANCE ON THE INTERNET PROTOCOL

İSMAİL UZUN

SUMMARY

In this study, the performance of Turkish continuous speech recognition (TCSR) over the Internet protocol with distributed architecture is investigated, packet loss compensation (PLC) algorithms are compared and an adaptive estimation method where system performance increases adaptively is developed.

In voice over IP (VoIP), voice is usually transmitted as data packets using the unreliable UDP protocol. During transmission, packet losses may occur depending on congestion in routers and gateways. Packet losses should be compensated to increase the performance of continuous speech recognition over IP.

A distributed speech recognition architecture is used where only speech feature vectors are carried over UDP. For the receiver side, a PLC system, consisting of pre-processing, compensation and post-processing stages, is designed. In the pre-processing stage, speech feature vectors are buffered and compensation method is determined depending on packet loss type. The compensation method is then applied in the compensation stage. In the post processing stage, the mean and covariance of speech feature vectors are updated adaptively and the outputs are later used in estimation methods to increase the performance. A dynamic buffer is designed for the compensation and criteria to choose the method for the compensation of packet losses is defined.

In this study, interpolation and estimation techniques are investigated for the PLC algorithm. In the current estimation methods which are used for PLC, training data are assumed to be wide sense stationary. The estimation method is enhanced by adaptation of mean and covariance parameters with each new speech packet.

Keywords: Turkish continuous speech recognition, VoIP, packet loss compensation, adaptive estimation.

TEŞEKKÜR

Doktora çalışmalarında, gerek derslerimde ve gerekse tez çalışmalarında, bana danışmanlık ederek, beni yönlendiren ve her türlü olanağı sağlayan danışmanım Yrd. Doç. Dr. Rifat Edizkan'a, konu ile ilgili fikir ve önerilerini bizlerle paylaşan Prof. Dr. Atalay Barkana'ya ve tezimin tüm aşamalarında sabır ve ümitle bana destek olan sevgili aileme teşekkür ederim. Ayrıca, ses datalarının temini ve testlerimde yardımlarından dolayı Boğaziçi Üniversitesi'nden Yrd. Doç. Dr. Murat Saraclar ve Araş. Gör. Ebru Arısoy'a, akustik ve metin veri tabanı için ODTÜ ve Sabancı Üniversitelerine, çalışmalarında yer alan kaynakların temininde yardımlarını esirgemeyen arkadaşlarım Yrd. Doç. Dr. Ahmet Yazıcı, Dr. Kemal Özkan ve Yük. Müh. Erkan Elcik'e, burada ismini saymadığım ve çalışmalarında yardımcı olan tüm arkadaşlarıma da teşekkürü bir borç bilirim.

İÇİNDEKİLER

	<u>Sayfa</u>
ÖZET	v
SUMMARY	vi
TEŞEKKÜR	vii
ŞEKİLLER DİZİNİ	xi
ÇİZELGELER DİZİNİ	xiv
SİMGELER VE KISALTMALAR DİZİNİ	xv
1. GİRİŞ	1
2. İNTERNET PROTOKOLU VE SES İLETİMİ	6
2.1 İnternet Protokolünün Genel Yapısı.....	6
2.2 İnternet Protokolü Üzerinden Ses İletimi.....	8
3. KONUŞMA TANIMA	11
3.1 Yalıtık Konuşma Tanıma	11
3.2 Sürekli Konuşma Tanıma.....	12
3.3 Ses Analizi ve Öznitelik Vektörleri	15
3.3.1 Ön vurgulama.....	16
3.3.2 Pencereleme	17
3.3.3 Spektral analiz.....	18
3.3.4 Mel-Filtrebankası	18
3.3.5 Mel frekans cepstrum hesaplaması	19
3.4 Dağıtık Yapıda Konuşma Tanıma.....	20
4. PAKET KAYIPLARI VE TELAFİ YÖNTEMLERİ	24
4.1 Paket Kayıp Yapıları	24
4.1.1 Rasgele paket kayıpları	28
4.1.2 Seyrek paket kayıpları.....	28
4.1.3 Patlamalı paket kayıpları.....	29
4.2 Paket Kayıp Telafi Yöntemleri	29

4.3	Gönderici Tarafında Yapılan Paket Kayıp Telafi Yöntemleri	30
4.3.1	Gönderici tarafında hata düzeltimi.....	30
4.3.2	Serpiştirme	31
4.3.3	Yeniden İletim.....	33
4.4	Alıcı Tarafında Yapılan Paket Kayıp Telafi Yöntemleri	33
4.4.1	Yerine koyma	33
4.4.2	Aradeğerleme	34
4.4.2.1	Lagrange aradeğerleme	34
4.4.2.2	Spline aradeğerleme.....	35
4.4.3	Kestirim yöntemleri	38
4.4.3.1	En küçük ortalama kare hata kestirimi.....	41
4.4.3.2	En büyük olabilirlik kestirimi	42
4.4.3.3	En büyük sonsal kestirimi	42
5.	PAKET KAYIP TELAFİ SİSTEMİ	47
5.1	Ön İşlem	48
5.2	Dinamik Bellek Uzunluğunun Belirlenmesi	48
5.3	Telafi Algoritmasının Belirlenmesi.....	49
5.4	Ardışlem.....	50
6.	UYARLAMALI KESTİRİM YÖNTEMİ.....	51
6.1	Ortalama Değer Uyarlaması.....	53
6.2	Ortak Değişinti Uyarlaması.....	53
6.3	Çapraz Ortak Değişinti Uyarlaması	54
7.	DENEYSEL ÇALIŞMA VE SONUÇLAR.....	56
7.1	Eğitim ve Test Veritabanı	56
7.2	Test ve Analiz	57
7.3	Dağıtık Yapıda Ses Tanıma Sistemi ve Paket İletimi	57
7.4	Paket Kayıp Yapıları	58
7.5	Yalıtık Konuşma Tanıma ve Paket Kayıplarının Telafisi	60
7.6	Sürekli Konuşma Tanımada Aradeğerleme Yöntemleri ile Kayıp Telafisi	61
7.6.1	Lagrange Telafi Yöntemi	62
7.6.2	Spline Telafi Yöntemi	63

7.7	Kesitirim Yöntemi ile Kayıp Telafisi.....	69
7.8	Uyarlamalı Kestirim Yöntemi ile Kayıp Telafisi.....	75
8.	GENEL SONUÇLAR VE TARTIŞMA.....	81
	KAYNAKLAR DİZİNİ.....	83
	ÖZGEÇMİŞ	

ŞEKİLLER DİZİNİ

<u>Şekil</u>	<u>Sayfa</u>
Şekil 2.1 TCP/IP protokol katmanları ve paket formatları.	7
Şekil 2.2 UDP başlık ve veri formatı.	8
Şekil 2.3 Ses verisinin IP üzerinden iletimi	9
Şekil 3.1 Konuşma tanıma sistemi.	13
Şekil 3.2 Mel-frekans cepstral katsayılarının hesaplanması	15
Şekil 3.3 Ön-vurgu filtresinin genlik ve faz tepkisi	16
Şekil 3.4 Hamming penceresi	17
Şekil 3.5 Mel ve doğrusal frekans değerleri.....	19
Şekil 3.6 Dağıtık yapıda konuşma tanıma modeli	21
Şekil 3.7 Kablosuz konuşma tanıma için kullanılan üç yöntem, a) ses öznitelik parametrelerinin kod çözme işlemi sonrası elde edilmesi, b) ses öznitelik parametrelerinin kodlanmış bit katarından elde edilmesi, c) ses öznitelik parametrelerinin iletilmesi	22
Şekil 3.8 Ses öznitelik vektörlerinin kodlanmadan iletildiği dağıtık yapı modeli	23
Şekil 4.1 Test verisindeki örnek bir cümle için a-) MFCC (1), b-)Delta-MFCC (1), c-) Delta-Delta MFCC (1) değerleri.	27
Şekil 4.2 Ses öznitelik vektörlerinin IP üzerinden iletimi ve paket gösterimi	27
Şekil 4.3 Rasgele paket kayıplarının gösterimi.....	28
Şekil 4.4 Seyrek paket kayıplarının gösterimi	28
Şekil 4.5 Orta ve yüksek patlamalı paket kayıplarının gösterimi	29
Şekil 4.6 Serpiştirici Gösterimi a-) Serpiştirici b-) Geriserpiştirici	31
Şekil 4.7 4x4'lük serpiştirici örneği.....	32

Şekil 5.1 Paket kayıp telafi sistemi	47
Şekil 5.2 Ön işlem aşamasında dinamik tampon bellek uzunluğunun belirlenmesi	49
Şekil 6.1 Ses öznelik vektörlerinin test verisindeki 553 cümle için ortalama değer değişimi (MFCC : 5)	52
Şekil 7.1 Ses öznelik vektörlerinin UDP yükü olarak dağıtık yapıda iletimi için tanımlanmış basit bir UDP paket formatı.....	57
Şekil 7.2 Ses öznelik vektörlerinin IP üzerinden iletimi ve paket kayıpları	58
Şekil 7.3 Kayıp MFCC vektörleri üzerinde aradeğerleme uygulaması gösterimi	59
Şekil 7.4 Seyrek paket kayıplarının birden fazla merkezli olduğu durumlarda aradeğerleme yöntemlerinin uygulanması	59
Şekil 7.5 Patlamalı paket kayıp durumlarında kayıp vektörler ve telafi uygulanan vektörler	60
Şekil 7.6 Spline aradeğerleme yöntemi ile aradeğerleme uygulanmış MFCC (0) değerleri. Kayıp merkez oranı : 10, merkezdeki ardışık kayıp paket sayısı : 3	64
Şekil 7.7 Patlamalı kayıp durumlarında Lagrange telafisi ve hata oluşumu.....	67
Şekil 7.8 Patlamalı kayıp durumlarında Spline telafisinin etkili olmadığı durumlar (patlama büyüklüğü : 20), a) patlama merkezinde ortalama değer fazla değişmediği, b) patlama merkezinde ortalama değer değiştiği durum	68
Şekil 7.9 Kayıp vektörlerin gözlemlenen vektörlere göre çapraz ortak değişinti gösterimi.....	70
Şekil 7.10 Zaman farkına (τ) bağlı olarak MFCC (0) çapraz ortak değişinti değerinin değişimi	72
Şekil 7.11 Kayıp merkezindeki örneklerin kayıp etrafındaki örneklerle olan ilişkileri, a) bütün ortak çapraz değişinti değerleri, b) uzak mesafedeki örneklerle olan çapraz ortak değişinti değerleri (yaklaştırılmış). Kayıp sayısı : 20, kayıp merkezinin sağından ve solundan alınan örnek sayısı : 10.	73
Şekil 7.12 Uyarlamalı kestirim yöntemi testleri için kullanılan yapay veri.....	76

Şekil 7.13 Uyarlamalı ve uyarlamasız kestirim yönteminin yapay test verisi üzerinde gösterimi.....	77
Şekil 7.14 MFCC(0) için cümle sayısına göre hesap edilen ortak değişinti ($\tau = 0$) ve çapraz ortak değişinti ($\tau = 1$) değişimi.	78
Şekil 7.15 τ (0) değerinin farklı cümle sayılarına bağlı olarak değişimi.....	79

ÇİZELGELER DİZİNİ

<u>Çizelge</u>	<u>Sayfa</u>
Çizelge 7.1 Paket kayıplarında yerine koyma metodu ile yalıtık konuşma tanıma	61
Çizelge 7.2 Kayıplı (telafisiz) kelime hata oranları	62
Çizelge 7.3 Lagrange ile paket kayıp telafi yönteminde kelime hata oranları.....	63
Çizelge 7.4 Spline ile paket kayıp telafi yönteminde kelime hata oranları.....	64
Çizelge 7.5 Telafisiz ve aradeğerleme yöntemlerinin başarımlarının karşılaştırılması ..	65
Çizelge 7.6 Patlamalı paket kayıpları durumunda Spline aradeğerleme yönteminin başarımları	66
Çizelge 7.7 Patlamalı paket kayıpları durumunda Lagrange aradeğerleme yönteminin başarımları	66
Çizelge 7.8 Patlamalı paket kayıpları durumunda aradeğerleme yöntemlerinin SNR değerleri.....	67
Çizelge 7.9 Patlamalı paket kayıpları durumunda EBSK yönteminin SNR değerleri....	74
Çizelge 7.10 Eğitim ve test amaçlı kullanılan cümlelerdeki toplam ses öznitelik vektör sayısı.....	74
Çizelge 7.11 Patlamalı durumlarda kestirim ve aradeğerleme yöntemlerinin başarımları	75
Çizelge 7.12 EBSK yönteminin uyarlamalı ve uyarlamasız durumlarda başarımları.....	79

SİMGELER VE KISALTMALAR DİZİNİ

<u>Simgeler</u>	<u>Açıklamalar</u>
μ	Ortalama değer
σ^2	Varyans
$E\{ \cdot \cdot \}$	Beklenen değer
Σ_m	Ortak deęişinti matrisi
Σ_{mo}	Çapraz ortak deęişinti matrisi
$P(W)$	Olasılık yoğunluk fonksiyonu
$P(X_m; \mu, \Sigma)$	Olasılık yoğunluk fonksiyonu parametrik modeli
$P(X_m X_o; \Sigma)$	Koşullu olasılık dağılımı
Δ MFCC	Delta katsayıları
$\Delta\Delta$ MFCC	Delta-Delta katsayıları

<u>Kısaltmalar</u>	<u>Açıklamalar</u>
DSR	Distributed Speech Recognition
EBSK	En Büyük Sonsal Kestirim
GHD	Göndermede Hata Düzeltimi
IP	Internet Protocol (Internet Protokolü)
IP-PBX	IP Private Branch Exchange
ITU	International Telecommunication Union
KHO	Kelime Hata Oranı
MAP	Maximum a posteriori
MFCC	Mel-frequency Cepstral Coefficients
ODM	Ortak Deęişinti Matrisi
OYF	Olasılık Yoęunluk Fonksiyonu
PSTN	Public Switched Telephone Network
RTP	Real-time Transmission Protocol

SİMGELER VE KISALTMALAR DİZİN (devam)**Kısaltmalar****Açıklamalar**

TCP	Transmission Control Protocol
UDP	User Datagram Protocol
WER	Word Error Rate

BÖLÜM 1

GİRİŞ

İnternetin büyük bir hızla yaygınlaşması farklı teknolojilerin ve araştırma alanlarının doğmasına yol açmıştır. İlk başlarda laboratuvarlarda aynı ortamdaki bilgisayarları haberleştirmek için kullanılan yerel bilgisayar ağları, günümüzde kıtaları ve coğrafyaları içine alan evrensel bir bilgisayar ağı haline gelmiş ve hayatın hemen her alanına girmiştir.

Her ne kadar yeryüzünde toplulukların önemli bir kısmı en basit haberleşme sistemlerinden dahi mahrum olsa da, İnternet ve buna bağlı teknolojilerin kullanımı yeryüzünün diğer bölümünü hızlı bir şekilde sarmaktadır. İnternet ulaşabildiği yerleri daha küçük yapmakta, hayatı daha da hızlı kılmaktadır. Bazı topluluklardaki en temel haberleşme imkanlarının dahi kısıtlı olmasına, buna karşılık İnternet'in ulaşabildiği yerlerde akıl almaz hızla yaygınlaşmasının ve hayatın her alanına girmesinin toplumlara göre değişkenlik arz eden sosyolojik ve ekonomik sebepleri olabilir. Bu alandaki tartışmaları ve incelemeleri ilgili bilimlerin uzmanlarına bırakarak, bu çalışmada teknik açıdan en temel amacımız; ülkemizde son on-on beş yılda içine alan İnternet ve hızla yayılan İnternet teknolojilerini yararlı şekillerde kullanarak ülkemiz ve insanlığa katkıda bulunmaktır.

Hayatın her alanında kendini mecburen hissettiren İnternet, temel anlamda veri transferi yapmakla birlikte, son yıllarda ses ve görüntü ya da en genel anlamda medya transferinin sağlandığı ortam haline gelmiştir. Ses ve görüntü teknolojilerinin gelişimine paralel olarak İnternet sayesinde medya transferi ve ilgili teknolojilerin gelişimi de büyük oranda artmış ve hızla artmaya devam etmektedir.

İnternet teknolojileri ve ses teknolojilerindeki gelişmeler İnternet protokolu (IP) üzerinden konuşma tanıma uygulamalarının farklı alanlarda kullanılmasına yol açmıştır. Telefon haberleşmesi ve telefon santralleri, sesli yanıt sistemleri, IP üzerinden ev ve

endüstriyel komut kontrol uygulamaları da IP üzerinden konuşma tanımaya ihtiyaç duymaktadır.

Internet, uzun yıllardır kullanılagelen klasik telefonlar ve haberleşme sistemlerinin yerini almaktadır. Internet teknolojileri temelinde çalışan telefon santralleri (IP-PBX), sadece bilgisayarları değil coğrafi olarak farklı mekanlardaki telefon gruplarını da haberleşme anlamında birleştirmiş ya da yakınlaştırmıştır. Bu sayede, bir ofiste çalışan insanlar uzak mesafelerdeki başka bir ofis çalışanlarına iç hat numarası çevirerek IP üzerinden ulaşabilmektedir. Internet teknolojisinin bir ara ürünü niteliğini taşıyan ve genelde Internet'e eş değer tutulan global ağ (www), medya transferi ve uygulamalarını teknik ağırlıklı olmayan ev kullanıcıları için daha kullanışlı ve cazip kılmaktadır. İnsanların iş yerlerinden ağ tarayıcısıyla evini izleyebilmesi ve sesli komutlarla ev aletlerini kontrol edebilmesi, Internet'in insan hayatının her alanına girdiğinin önemli bir göstergesidir.

Endüstriyel anlamda düşünüldüğünde, üretim yapılan fabrika ortamlarında ya da mobil sistemlerin olduğu ve kontrol edilmesi gereken durumlarda Internet kullanılabilmekte ve medya transferi yapılarak, örneğin ses transferi ile gerekli kontrol işlemi gerçekleştirilebilmektedir (Larson, 2000).

Klasik telefon sistemlerinde ses verisi 64 kbit/s bit hızında transfer edilmektedir. Internet ortamında ses iletimi için band genişliğinin sınırlı olmasından dolayı ses verisi daha düşük bit hızlarında kodlanarak, örneğin 32, 16, 8 veya 4 kbit/s ile gönderilir. Alıcı tarafında ses, kod çözücülerden geçirilerek uygulamaya bağlı olarak gerekli işlem yapılır. Ancak, sesin düşük bit hızlarında kodlanması sesin kalitesinde bozulmalara sebep olmaktadır (Euler and Zinke, 1994; Lilly and Paliwal 1995; Besacier, et al., 2003).

Ses kalitesinin düşük olması konuşma tanıma başarımını etkilemektedir. Ses verilerinin iletimi için ağ katmanında IP üzerinde çalışan UDP (User Datagram Protocol) kullanılır. UDP paket tabanlı çalışan bir protokoldür ve ilgili veri paketlerinin gönderilen noktaya iletimini garanti etmez. Dolayısıyla, transfer sırasında ağdaki trafik yoğunluğuna göre paket kayıpları oluşabilmektedir. Konuşma tanıma uygulamaları için

Internet Protokolü (IP) üzerinden sesin kodlanarak gönderilmesi, özellikle yüksek bit hızlarının kullanılması, ağdaki veri trafiğini arttırmaktadır. Aynı ağda trafiğin fazla olması ses verisinin transfer edildiği UDP protokolden dolayı kayıplara yol açabileceğinden, veri trafiğinin fazla olması uygulama alanlarına bağlı olarak genellikle tercih edilmez.

Konuşma tanımada en yaygın olarak state-of-art yöntemi saklı Markov modeller (SMM) kullanılmaktadır. Bu algoritmaların bilgisayarda gerçekleştirilmesi belirli bir işlem gücü gerektirir. Bunun yanında tanıma algoritmasının başarımını arttırmak ve farklı ortamlarda daha gürbüz tanıma yapabilmek için kullanılan algoritmalar işlem gücü ve bellek gereksinimini arttırmaktadır.

Ağ veri trafiğini azaltmak, band genişliğinden en yüksek düzeyde yararlanmak ve konuşma tanıma için gerekli işlem gücünü paylaşmak için dağıtık yapıda konuşma tanıma yöntemleri geliştirilmiştir ve uygulanmıştır (Milner 2001; Demichelis et al., 2005; Hacıoğlu and Pellom, 2003; Palaz vd., 2005; ETSI Standard 2000, 2003).

Bu sistemlerde en genel anlamda, gönderici tarafında kodlanmış ses iletilir. Alıcı, ses verisi üzerine uygulanacak tanıma işlemlerini farklı donanım ve yazılım mimarileri kullanarak paylaşır. Bu şekilde gönderici ses verisi değil ses öznitelik vektörlerini de iletebilir (ETSI Standard 2000, 2003; Digalakis, et al., 1998; Ramaswamy and Gopalakrishnan, 1998). Dağıtık yapıda konuşma tanımada sesin öznitelik vektörlerinin gönderilmesi ile veri trafiğinin ve dolayısıyla paket kayıplarının azaltılması amaçlanmaktadır. Bu yapıda, alıcıya gelen ses öznitelik vektörleri tanımada kullanılan istemcilerle gönderilir ve istemcilerden tanıma sonuçları alınır. Dağıtık yapıda konuşma tanıma uygulamaları band genişliği kullanımının çok önemli olduğu gezgin telefonlar sistemlerinde de büyük önem arz etmektedir (ETSI Standard 2000, 2003).

Bu çalışmada Türkçe sürekli konuşma tanımada ses vektörlerinin IP üzerinden iletildiği durumlarda paket kayıplarının etkileri ve paket kayıp telafi yöntemlerinin başarımları incelenmiş ve paket kayıplarının etkisini azaltıcı paket kayıp sistemi

tasarımı ve yöntemlerin geliştirilmesi üzerinde çalışmalar yapılmıştır. Bu amaçla; mevcut sistemlerde paket telafisi için kullanılan bazı aradeğerleme ve kestirim yöntemleri ayrı ayrı incelenmiştir. Bu çalışmada, paket kayıp tipinin belirlenmesi, paket kayıp tipine göre en uygun telafi yönteminin uygulanması ve telafi uygulama sonrasında ardişlem uygulamasını kapsayan paket kayıp telafi sistemi önerilmiştir. Önerilen sistemde paket kayıpları ilk başta bir önışlemeden geçirilir ve kayıp tipi tesbit edilir. İkinci aşamada belirlenen kayıp yapısına göre telafi uygulanır ve ardişlem aşamasına geçilir. MAP kestirim yöntemi uygulanırken verilerin geniş anlamda durağan olduğu kabullenilmektedir. Bu şekilde eğitim verisinden elde edilen ortalama değer kayıp telafisinde kullanılmaktadır (Milner and James, 2006). Bu çalışmada MAP kestirim için yapılan kabullenmeden farklı olarak ortalama değer sürekli güncellenmektedir. Ayrıca ortak değışinti değerlerinin de güncellenmesi yapılarak sistem başarımının zaman içerisinde uyarlamalı olarak iyileşmesi sağlanmaktadır.

Tezin en önemli katkıları şunlardır:

- IP üzerinde Türkçe sürekli konuşma sisteminin dağıtık yapı modelinde ve çeşitli paket kayıp durumlarında başarımları elde edilmiştir. Paket kayıplarının telafisi için paket kayıp telafi sistemi tasarlanmıştır. Mevcut telafi yöntemleri uygulanmış ve yeni yöntemler üzerinde çalışmalar yapılmıştır.. Tezde elde edilen sonuçlar, bu konuda çalışan araştırmacılara yol gösterecektir.
- Kullanılan en büyük sonsal dağılım kestirim yönteminde eğitim verisinin geniş anlamda durağan kabul edilerek hesap edilen ortalama değer telafi başarımının düşük çıkmasına sebep olmuştur ve ortalama değer bu şekilde seçilemeyeceği anlaşılmıştır. Bununla birlikte, ortak değışinti ve çapraz ortak değışinti değerlerinin farklı sayıda eğitim cümlesi için değışkenlik gösterdiği belirlenmiş ve uyarlama gereksinimi gözlemlenmiştir. Bu amaçla, ortak ve çapraz ortak değışinti değerleri, gözlemlenen her ses öznitelik vektörü için uyarlamalı olarak

güncellenmiştir. Böylece telafi sistemi zaman içerisinde başarımını arttırmaktadır.

- Paket kayıp telafi sistemi tanımlanarak paket kayıpları önilem, telafi ve ardişlem olmak üzere üç farklı aşamada telafi edilmiş ve telafi yöntemi parametreleri güncellenmiştir.
- Dinamik bellek tanımlanarak kayıpların oluştuğu durumlarda uygun aradeğerleme yöntemi seçilerek değişik platformlarda özellikle gerçek zamanlı uygulamalar için en uygun telafi yönteminin uygulanması sağlanmıştır.

Tezin genel yapısı şu şekildedir:

Internet protokolünün yapısı ve IP üzerinden ses iletimi ikinci bölümde incelenmiştir. Üçüncü bölümde, konuşma tanıma yöntemleri üzerinde genel bilgiler verilerek bu tezde konuşma tanıma için kullanılan parametreler ve dağıtık yapıda konuşma tanıma sistemi anlatılmıştır. IP üzerinden ses iletimi uygulamaları geliştirmek ve paket kayıplarını telafi etmek için oluşan paket kayıplarının yapısını anlamak ve istatistiksel olarak kayıp tiplerini bilmek gerekmektedir. Paket kayıplarının yapısı ve paket kayıp telafi yöntemleri dördüncü bölümde incelenmiştir. Beşinci bölümde, tasarlanan paket kayıp telafi sistemi anlatılmıştır. Altıncı bölümde uyarlamalı kestirim yönteminin avantajları sıralanarak yöntem detaylı olarak anlatılmıştır. Yapılan deneysel çalışmalar, testlerle ilgili kabullenmeler ve sonuçlar yedinci bölümde verilmiştir. Genel olarak sonuçların değerlendirilmesi ve tartışmalar sekizinci bölümde ele alınmıştır.

BÖLÜM 2

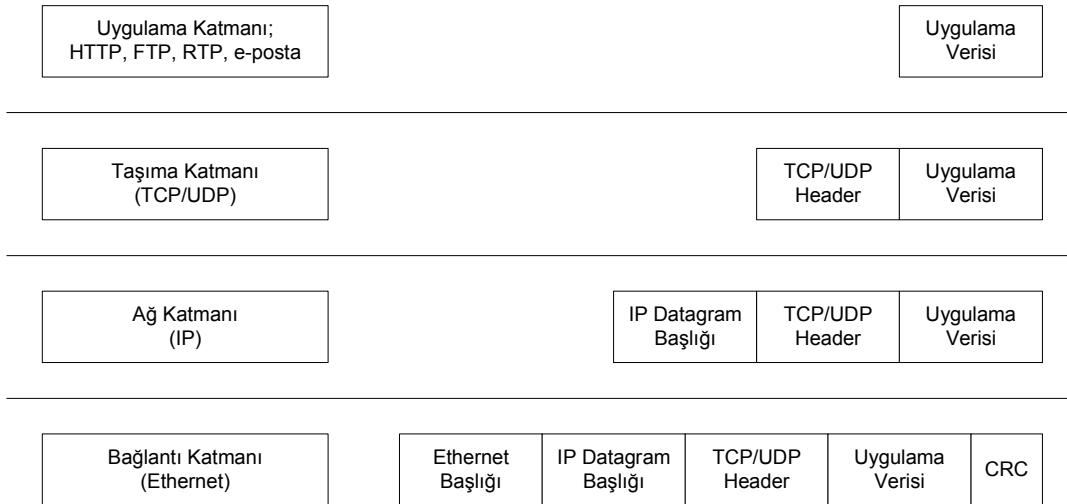
INTERNET PROTOKOLU VE SES İLETİMİ

Internet Protokolu (IP) üzerinden ses iletimi çalışmaları 1970'li yıllara dayanmaktadır (Perkins, 2003). İlk olarak günümüzde kullanılan TCP/IP benzeri bir yapıya sahip olan ARPANET üzerinde çalışan Ağ Ses Protokol (Network Voice Protocol) (Cohen, 1977) standardı tanımlanmıştır. Bu standardın, son kullanıcılar arasında paketlerin tekrar gönderilmesini önlemek, ses paketlerini kontrol paketlerinden ayrı tutmak ve alt seviye protokollerden bağımsız olmak gibi temel hedefleri vardı. UDP'nin (Postel, et al., 1980) ve ağ katmanı protokolünün (Postel, et al., 1981 a) ve daha sonra ağ katmanı üzerinde çalışan TCP (Postel, et al., 1981 b) protokolünün tanımlanması ile günümüz Internet yapısı şekillenmiş ve veri transferleri taşıma katmanı olan TCP ve UDP üzerinden yapılmaya başlanmıştır.

2.1 Internet Protokolünün Genel Yapısı

Internet protokolu temel olarak bağ katmanı, ağ katmanı, taşıma katmanı ve uygulama katmanı olarak dört ana katmandan oluşmaktadır (Stevens, 1994). Taşıma katmanında veri (ses, görüntü vs.) TCP ya da UDP ile taşınmaktadır.

Şekil 2.1'de TCP/IP katman yapısı ve her katmana ait paket formatları gösterilmektedir. Ethernet başlığı 14 bayt, CRC (*cyclic redundancy check*-çevrimsel artıklık kodu) 4 bayt, IP başlığı 20 bayt, TCP/UDP başlığı 20/12 bayttan oluşmaktadır. Uygulama katmanında, özellikle ses ya da görüntü uygulamalarında, alınacak ya da gönderilecek veriye ek olarak bu başlıkların getireceği yük hesap edilmelidir.



Şekil 2.1 TCP/IP protokol katmanları ve paket formatları.

TCP, güvenli, bağlantılı ve bayt katarları şeklinde çalışan protokoldür. Ancak, TCP bağlantılı olması ve güvenilirliği sağlaması sebebiyle veri transferinde gecikmelere ve veri trafiğinin artmasına sebep olur. TCP'nin bağlantılı olması sebebiyle ilk önce bağlantıyı kurmak için oluşan paket trafiği uygulamalar için makul görülebilir. Ancak, TCP'de alıcı tarafına ulaşmayan paketler tekrar gönderilir. Tek bir paketin bile alıcı tarafına ulaşmaması durumunda paketlerin tekrar gönderilmesi çok fazla gecikmeye sebep olur. Bununla birlikte, TCP bağlantılı olması ve güvenilirliği sağlaması sebebiyle gecikmeleri arttırdığı ve fazla işlem gücü gerektirdiği için IP üzerinden ses iletiminde, özellikle gerçek zamanlı uygulamalar için, tercih edilmez. Bayt katarı şeklinde transfer gerçekleştiği için ses verisi gibi paketler halinde iletilen veriler için uygulama katmanında daha fazla işlem gerektirdiği için uygun değildir.

UDP, bağlantısız ve güvenli olmayan bir protokoldür (Stevens, 1994). Gönderilecek paketlerin iletimi için ilk aşamada herhangi bir bağlantı kurulumu ve paket alış verişi olmaz. Bununla birlikte, alıcı tarafına paketlerin ulaştığını da garanti etmez. Ses uygulamaları gibi paket kayıplarının telafi edilebileceği durumlar için UDP tercih edilir. UDP paketleri toplam 8-Bayt'lık başlık ile iletilir. Şekil 2.2'de UDP paketi başlığı gösterilmektedir.

16-bit gönderici port numarası	16-bit alıcı port numarası	16-bit UDP uzunluğu	16-bit UDP sağlama toplamı	UDP Yüğü / Uygulama Verisi
--------------------------------------	----------------------------------	------------------------	----------------------------------	-------------------------------

Şekil 2.2 UDP başlık ve veri formatı.

UDP'nin en önemli özeliği paket tabanlı bir protokol olmasıdır. Paket olarak işlem gören ses verisinin yine paketler halinde IP üzerinden taşınması alıcı ve gönderici tarafında daha az işlem gücü harcanmasını sağlar.

UDP yükü olarak herhangi bir uygulama verisi gönderilebileceği gibi veri olarak tanımlanan yükte yeni başlıklar tanımlanarak uygulama katmanında yeni protokol geliştirilebilir.

2.2 İnternet Protokolü Üzerinden Ses İletimi

Kamusal anahtarlmalı telefon ağında (PSTN) devre anahtarlmalı sistem kullanılırken, IP üzerinden ses iletiminde paket anahtarlmalı sistem kullanılır. Paket anahtarlmalı sistemlerin en önemli farkı haberleşme hatları için özgül hattın olmaması ve bütün haberleşmelerin ortak hat üzerinden yapılmasıdır.

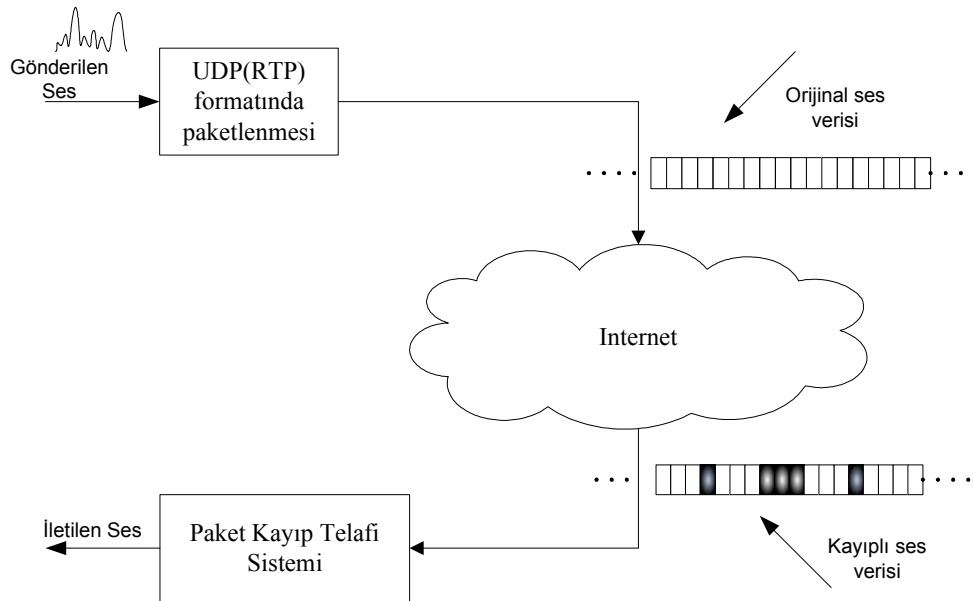
İnternet alt yapısının genişlemesi ve IP üzerinden ses iletimi teknolojilerinin gelişmesi neticesinde telefon haberleşmesi de PSTN yerine IP üzerinden yapılar hale gelmiştir. Bu uygulamalar için web tabanlı çağrı merkezleri ve IP telefonlar en iyi örneklerdir (Hassan, et al., 2000). Ne var ki, telefon haberleşmesi yüksek güvenilirlik gerektirmektedir.

Sesin IP üzerinden güvenli olmayan, kayıplara izin veren protokol üzerinden taşınması, haberleşmenin PSTN teknolojisindeki ses kalitesi ve güvenilirliğini sağlamamaktadır. PSTN hatlarda 8 kHz örnekleme hızında ve 64 kbit/s bit hızında veri

iletimi yapılır. Buna karşın IP üzerinden ses iletiminde band genişliğinden dolayı ses kodlayıcılar kullanılmaktadır. (ITU-T Recommendation G.729, 1996; ITU-T Recommendation G.722.2, 2003).

IP üzerinde ses paketleri 10-40ms gibi küçük paketler halinde transfer edilir (ITU-T Recommendation G.729; Adersen, et al., 2002). İnsan kulağı sesteki gecikmelere hassastır ve ses paketlerindeki gecikmeler kolaylıkla fark edilebilir. Ses verisinin iletiildiği paketlerin gecikmesinden kaynaklanan etkileri ve iletim sırasında oluşan paket kayıplarının etkisi hem alıcı hem de gönderici tarafında kullanılan yöntemlerle telafi edilmeye çalışılır (Perkins, et al., 1998; Bolot, et al., 1999).

UDP üzerinden ses verisinin iletiminde ses verisini doğrudan UDP başlığı ve UDP verisi olarak taşımak mümkündür. Ancak, gerçek zamanlı olarak ses verisini aktarmak ve hem alıcı hem de gönderen tarafında kayıpların telafisini mümkün kılabilmek ve alıcı ile gönderici arasında istatistikleri temin edebilmek için RTP standardı tanımlanmıştır (Schulzrinne, et al., 2003). Ses verisinin IP üzerinden iletimi Şekil 2.3'te gösterilmektedir.



Şekil 2.3 Ses verisinin IP üzerinden iletimi

Gönderici tarafında UDP ya da RTP paketleri olarak gönderilen ses verisin Internet protokolü üzerinden iletimi sırasında kayıplar oluşur ve alıcı tarafında paket kayıp telafi yöntemleri ile kayıp verinin etkisi giderilmeye çalışılır.

BÖLÜM 3

KONUŞMA TANIMA

Konuşma tanıma uygulamaları genel olarak yalıtık konuşma tanıma, bağlantılı konuşma tanıma ve sürekli konuşma tanıma olarak üç temel gruba ayrılabilir. Tanıma işlemi için uygulanan yöntemlerin farklı olması sebebiyle bu çalışmada yalıtık ve sürekli konuşma tanıma çalışmaları iki farklı bölümde anlatılmıştır.

Tez çalışmasında paket kayıpları için temel olarak sürekli konuşma tanıma sistemi kullanılmıştır. Bunun yanında, paket telafi yöntemlerinin yalıtık konuşma tanımadaki başarımları da incelenmiştir.

Konuşma tanıma çalışmalarında, işlem gücü ve bellek gereksinimlerini paylaşmak ve gezgin telefon sistemlerinde daha yüksek verim ve daha az band genişliği kullanmak için dağıtık yapıda konuşma tanıma sistemleri (ETSI Standard 2000, 2003; Milner 2001; Demichelis et. al., 2005) kullanılır. Dağıtık yapıda IP üzerinden ses iletimi ve konuşma tanıma sistemi de bu bölümde anlatılmıştır.

3.1 Yalıtık Konuşma Tanıma

Yalıtık konuşma tanıma çalışmalarında şablon uyumlama (template matching), akustik-fonetik tanıma (acoustic-phonetic recognition) ve olasılıksal (stochastic) yöntemler kullanılmaktadır (Rabiner and Juang, 1993).

Şablon uyumlama yönteminde her kelime özellik ya da parametre vektörleri ile ifade edilir ve eğitim aşamasında her kelime için şablon elde edilir. Tanıma aşamasında tanınacak kelime ve mevcut şablonlar arasında seçilen uzaklık işlevine göre

karşılaştırma yapılır (Markowitz,1996). Bilinmeyen kelime, en küçük uzaklık veren sınıfa atanır.

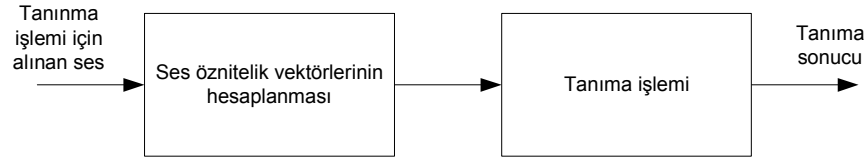
Akustik-fonem tanıma fonem seviyesinde tanıma işlemi yapar. Fonemlerin tanımlanması ve kullanılması karşılaştırma sayısını sınırlı tuttuğu için tanıma işleminde hız olarak daha çok tercih edilebilir. Akustik-fonetik yöntemde fonemleri belirlemek ve kelimenin başlangıç ve bitişini iyi belirlemek gerekmektedir. Belirlenen fonemler, fonem örgüsü şeklinde ifade edilir ve tanıma işlemi bu örgü kullanılarak gerçekleştirilir.

Olasılıksal yöntemlerde şablon yöntemlerindeki gibi tanıma aşamasında kullanmak üzere önceden bazı vektörler ya da parametreler hesap edilir. Tanıma aşamasında doğrudan şablon karşılaştırması yapılmaz ancak olasılık yöntemlerine göre karar verilir. Olasılık temelli yöntemlerden saklı Markov modeli (HMM: Hidden Markov Model) en yaygın kullanılan yöntemdir.

Yalıtık konuşma tanıma işleminde her yalıtık kelime SMM ile modellenir. Tanıma işlemi için akustik gözlem vektörü $O = \{o_1, o_2, \dots, o_T\}$ her SMM modelinden geçerek ileri (forward) ve geri (backward) işlemlerle olası değerleri hesap edilir. Bu işlem ile her modelin bu gözlem dizisini üretme olasılığı elde edilir ve en yüksek olasılığa sahip kelime tanınan kelime olarak belirlenir.

3.2 Sürekli Konuşma Tanıma

Sürekli konuşma tanıma sistemi genel olarak tanıma işlemine tabii tutulan ses sinyalinin analiz ve öznitelik vektörlerinin hesaplanması, olasılık ve dil modellerinin eğitilmesi ve tanıma işlemi (kod-çözme) olarak farklı adımları içerir. Örnek bir konuşma tanıma sistemi Şekil 3.1'de gösterilmiştir.



Şekil 3.1 Konuşma tanıma sistemi.

Sürekli konuşma tanıma yönteminde yalıtık konuşmadan farklı olarak kelimeler arasında duraklama olmaz. Bu özelliği nedeniyle doğrudan şablon kullanımı gibi yöntemler uygun olmazken, akustik-fonetik yöntemlerde bir dil için gerekli fonetik bilginin tüm özelliklerini içeren fonetik bilginin kayıt edilmesindeki zorluklar nedeniyle yetersiz kalmaktadır. Başarım açısından günümüzde en yaygın olarak kullanılan sürekli konuşma yöntemi saklı Markov modelidir.

Sürekli konuşma tanıma problemi verilen akustik model, dil modeli ve akustik gözlemlere göre en yüksek olasılıklı dizinin seçimi şeklinde ifade edilebilir. Burada kelime sınırlarının belirli olmaması sürekli konuşma tanıma işlemi için önemli bir problemdir. Tanımlanan problemi matematiksel olarak ifade edecek olursak, bir cümle farklı kelimelerin dizilişleri ile ifade edilir, $W = w_1, w_2, \dots, w_n$. Sürekli konuşma tanıma sistemi gözlenen X akustik vektörlerine göre sonsal olasılığı en yüksek (maximum *a posteriori*) kelime dizisini (\hat{W}) seçer (Renals and Hochberg, 1999) ve Eşitlik 3.1'deki gibi ifade edilir.

$$\hat{W} = \arg \max_w P(W / X) \quad (3.1)$$

Eşitlikte $P(W / X)$, W kelime dizisinin verilen X akustik vektörüne göre elde edilen en büyük sonsal olasılığıdır. En yüksek olasılıklı kelimeyi seçmek için Bayes kuralı uygulanarak Eşitlik 3.1 şu şekilde yazılabilir.

$$\hat{W} = \arg \max_w P(W / X) = \arg \max_w \frac{P(W)P(X/W)}{P(X)} \quad (3.2)$$

Eşitlik 3.2’de, $P(W)$ kelime dizisinin önsel olasılık (*priori* probablity) değerini göstermektedir. Bu değer gözlemlenen ses sinyalinden bağımsızdır ve dil modeline göre hesap edilmektedir (Young, 1996). $P(X/W)$ ise akustik model olabilirliğini (likelihood) göstermektedir (Renals and Hochberg, 1999).

Her tanıma işlemi için, eğitim ve kod çözme aşamalarında ses sinyali bir ön-ışlemden (front-end) geçirilerek ses sinyali akustik öznitelik vektörleri $X = x_1, x_2, \dots, x_T$ dizisi şeklinde ifade edilir. Her bir akustik vektör ses izgesinin (spectrum) genellikle 10ms’lik kısa süreli ifadesidir (Young, 1996). En yaygın kullanılan ses öznitelik vektörleri Mel-frekans kepstrum katsayıları (Mel-frequency cepstrum coefficients) (Davis and Mermelstein, 1980) ve algısal doğrusal öngörücü (perceptual linear predictive) katsayılarıdır (Hermansky, 1990).

Eğitim aşamasında, seçilen model parametreleri istatistiksel yöntemlerle belirlenir. Eğitimde SMM’in her durumunda (state) ortaya çıkan akustik ses öznitelik vektörlerinin dağılımları Gaussian olasılık yoğunluk fonksiyonu ile gösterilir.

Sürekli konuşma tanıma sistemlerinde dil modellemesi yapılarak doğal konuşma yapısına en uygun model elde edilmeye çalışılır. En yaygın kullanılan istatistiksel yöntem n-gram modelleridir (Rosenfeld, 2000). n-gram modeller, cümle içindeki bir kelimenin önce gelen n-1 kelimeye bağlı olarak sözdizimi (syntactic) ve anlamsal (semantic) bağını belirler (Gauvain and Lamel, 2000) ve

$$P(w_i | h_i) \approx P(w_i | w_{i-n+1}, \dots, w_{i-1})$$

şeklinde ifade edilir. Eşitlikte gösterildiği gibi kestirim problemi n-1 kelime ile sınırlandırarak problemin boyutunu azaltır (Rosenfeld, 2000).

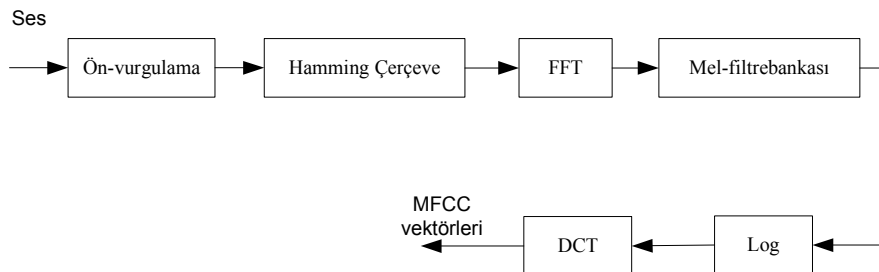
Kod çözme aşamasında bilinmeyen ve sistemin girdisi olan ses ve eğitilen modeller arasında karşılaştırma yapmak için arama (search) işlemi uygulanır. Arama

işlemi akustik ve dil bilgileri kullanılarak gerçekleştirilir. Viterbi kod-çözme yöntemi en iyi arama (optimal search) yapan yöntemdir (Forney, 1973).

3.3 Ses Analizi ve Öznitelik Vektörleri

Konuşma tanıma sistemlerinde ses gösterimi için değişik öznitelikler kullanılır. En yaygın kullanılan ses gösterimi MFCC (Mel-frequency cepstral coefficients) parametreleridir (Davis and Mermelstein, 1980). Mel, sesin algılanan frekansının veya perdesinin (pitch) ölçüm birimidir. Mel, bilinen anlamda frekanstan farklıdır. İnsanın işitsel sistemi perdeyi lineer şekilde algılamaz. Mel ölçüsüne göre insanın işitsel algılaması 1 kHz 'e kadar doğrusal, 1 kHz'nin üzerinde ise logaritmiktir. Bu algılama ölçüsüne göre ses sinyali Mel ölçüsüne göre yerleştirilmiş N adet filtreden geçirilir. Filtre enerjileri üzerine DCT (Discrete Cosine Transform) uygulanarak MFCC parametreleri elde edilir.

Tez çalışmasında kullanılan ses öznitelik hesaplama işlemi Şekil 3.2'te gösterilmektedir (ETSI 2000, 2003). Şekilde gösterildiği gibi ses sinyali Mel filtre bankasından geçirilmeden önce ön-vurgulama, Hamming pencereleme ve Fourier dönüşümü işlemlerinden geçirilir.



Şekil 3.2 Mel-frekans cepstral katsayılarının hesaplanması

3.3.1 Ön vurgulama

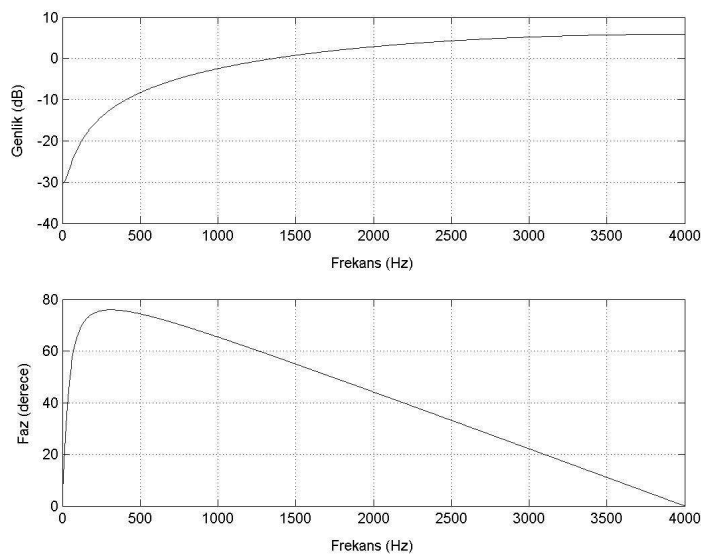
Ses üretimi sırasında ses sinyallerinin yüksek frekansları daha fazla bastırılır. Bu frekans bileşenlerinin genliğini yükseltmek için ses sinyali üzerine ön-vurgulama yapılır. Ön-vurgulama, ses sinyalinin transfer fonksiyonu Eşitlik 3.3'te gösterilen birinci dereceden FIR (Finite Impulse Response) filtreden geçirilmesi ile gerçekleşir (Oppenheim and Shafer, 1989).

$$H(z) = 1 - \alpha \cdot z^{-1} \quad (3.3)$$

Ön-vurgulama işleminde ses sinyalinin zamana bağlı gösterimi ise

$$y(n) = x(n) - \alpha \cdot x(n-1) \quad 0 \leq \alpha \leq 1 \quad (3.4)$$

şeklinde ifade edilir. α değeri 1'e yakın bir değerdir ve genellikle 0,97 olarak seçilir (ETSI Standard 2000, 2003). Şekil 3.3'te ön-vurgu filtresinin genlik ve faz tepkisi gösterilmiştir.



Şekil 3.3 Ön-vurgu filtresinin genlik ve faz tepkisi

3.3.2 Pencereleme

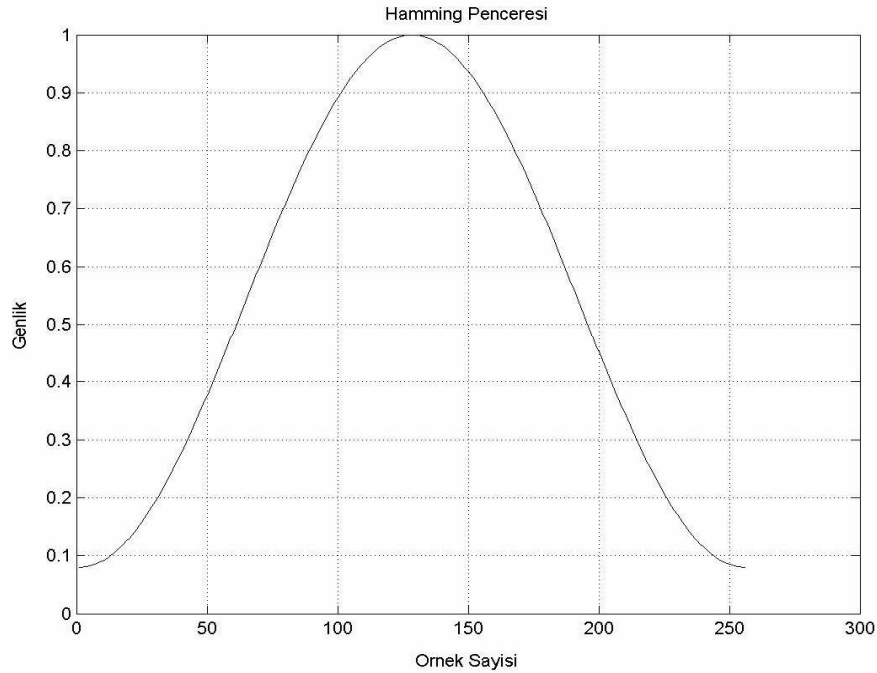
Konuşma tanımada ses sinyalleri analiz edilirken belirli sürelerle çerçevenir ve her bir çerçeve Hamming pencereden geçirilir.

$$y(n) = w(n)x(n) \quad (3.5)$$

Eşitlik 3.5'te $x(n)$ giriş sinyalini, $y(n)$ pencerelemiş sinyali ve $w(n)$ ise pencere sinyalini göstermektedir. Hamming penceresi Eşitlik 3.6'daki gibi ifade edilir ve genlik spektrumu Şekil 3.4'teki gibidir.

$$w(n) = \begin{cases} 0.54 - 0.46 \cos\left(\frac{2\pi n}{N-1}\right), & 0 \leq n \leq N \\ 0 & \text{diğer} \end{cases} \quad (3.6)$$

Eşitlikte N pencere uzunluğunu göstermektedir.



Şekil 3.4 Hamming penceresi

3.3.3 Spektral analiz

Ön-vurgulama ve pencereleme işlemlerinden geçirilen ses sinyalinin bütün çerçevelerine mel-filtrebankası hesap edilmeden önce frekans bilgisini elde etmek için Fourier dönüşümü uygulanır.

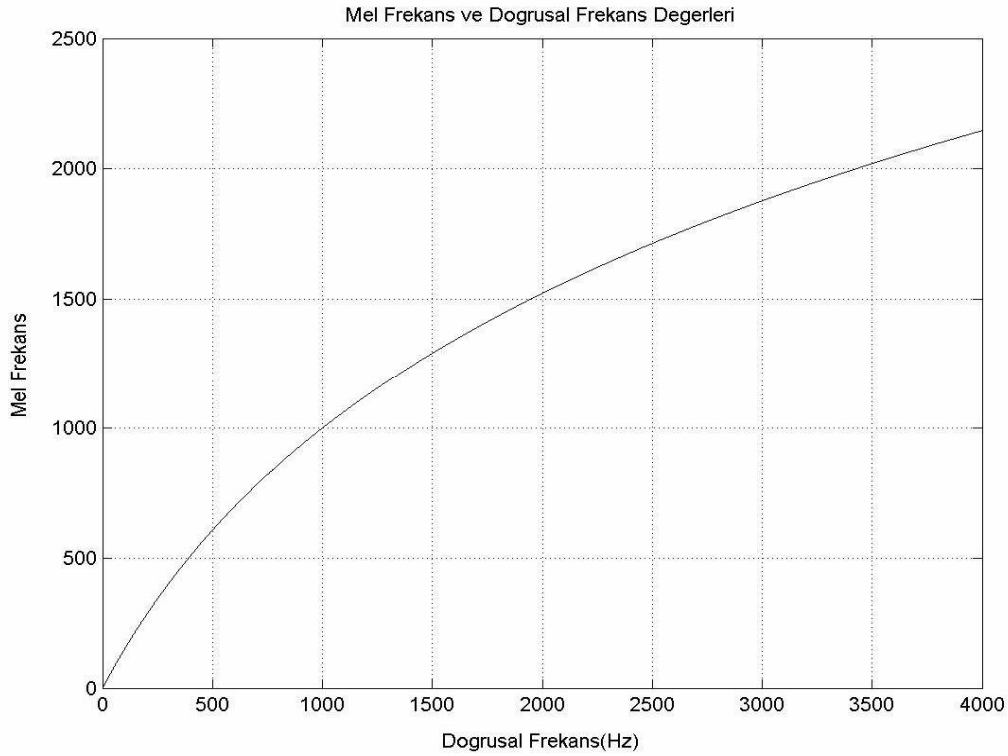
$$Y(k) = y(n) \left(e^{j2\pi kN} \right), \quad k = 0, 1, \dots, N-1 \quad (3.7)$$

Eşitlik 3.7'de, $y(n)$ ses sinyalinin zaman eksenindeki n . değerini, $Y(k)$ ise ses sinyalinin frekans eksenindeki değerini göstermektedir.

3.3.4 Mel-Filtrebankası

Mel-filtrebankası üçgensel band-geçiren filtrelerden oluşur ve her üçgensel filtre mel ölçümünde eşit band genişliğine sahiptir (Davis and Mermelstein, 198). Mel frekansı doğrusal frekanstan aşağıdaki formül yardımı ile elde edilir.

$$Mel\{x\} = 2595 \log_{10} \left(1 + \frac{x}{700} \right) \quad (3.8)$$



Şekil 3.5 Mel ve doğrusal frekans değerleri

Şekil 3.5'te Mel frekansı ve doğrusal frekans arasındaki ilişki gösterilmektedir. Yatay eksen doğrusal, dikey eksen Mel frekanslarını karşılık gelmektedir.

3.3.5 Mel frekans cepstrum hesaplaması

Frekans tanım kümesindeki veriyi cepstral tanım kümesine transfer etmek için aşağıdaki eşitlik kullanılır.

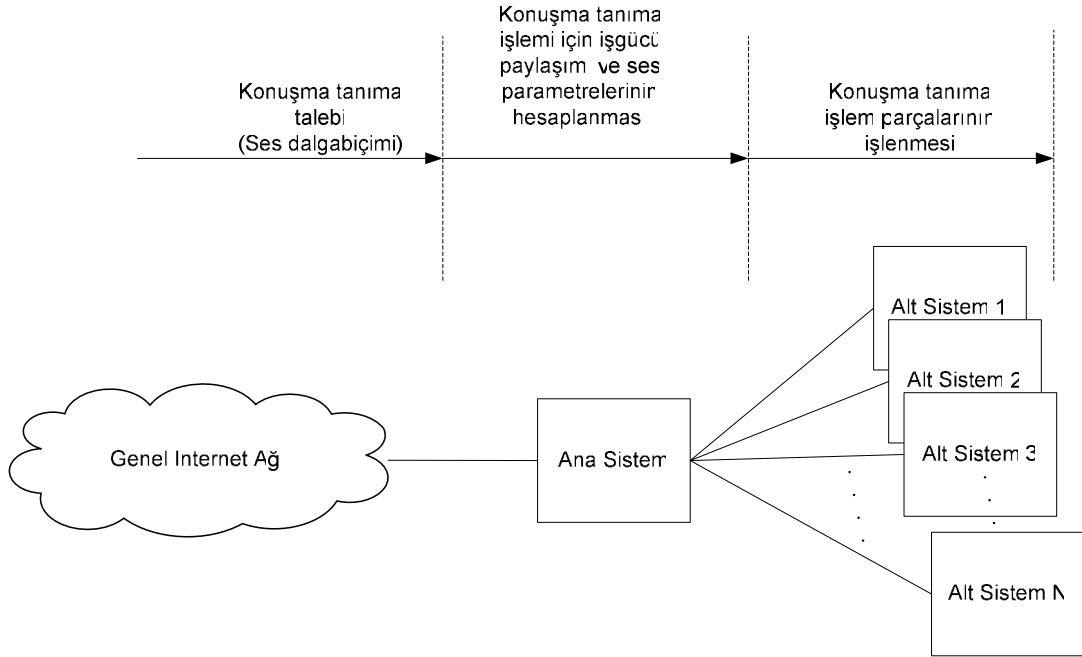
$$c^m(n) = \sum_{m=1}^M \log\{|Y(k)|\} \cos\left(n\left(m - \frac{1}{2}\right) \frac{\pi}{M}\right), \quad (3.9)$$

Eşitlik 3.9’da, M filtre bankası sayısını (23), $c(n)$ n . ceptral katsayısını ve $Y(k)$ log mel-filtrebankasının k . elemanını göstermektedir.

3.4 Dağıtık Yapıda Konuşma Tanıma

Internet ve konuşma tanıma uygulamalarının yaygınlaşması neticesinde çok farklı platformlarda konuşma tanıma algoritmalarının çalıştırılmasına ihtiyaç duyulmuştur. Bunun yanında konuşma tanıma algoritmalarının tanıma oranlarını arttırmak ve tanıma algoritmalarını daha gürbüz yapmak için karmaşık algoritmalar kullanılmaktadır. Hem gerekli işlem gücü hem de daha fazla bellek gereksinimleri konuşma tanıma adımlarının değişik işlemci ve platformlarda paylaşılmasını mecbur kılmaktadır.

Dağıtık yapıda konuşma, tanıma uygulamadaki mevcut donanım ve yazılım altyapısına bağlı olarak ses işlemlik sezimi, ses öznitelik vektörlerinin belirlenmesi, ses iyileştirme gibi çok farklı kısımlara ayrılabilir ve her bir kısmın verisi ortak bir hafıza üzerinden erişilip paylaşılabilir (Hacıoğlu and Pellom, 2003). IP üzerinden ise merkezi bir sunucu gerekli ses analizini yapıp tanıma işlemleri için en uygun uzak sunuculara tanıma verisini içeren paketleri gönderebilir (Palaz vd., 2005). Şekil 3.6’da dağıtık yapıda konuşma tanıma sistemi modeli gösterilmiştir. Şekil 3.6’da, Internet ortamından genel bir konuşma tanıma talebi ana sisteme gelmektedir. Ana sistemde, konuşma tanıma işlemleri farklı yöntemlerle IP üzerinden alt sistemlere paylaştırılabilmektedir. Bu tür sistemlerde ana sistem alt sistemlerin mevcut yüküne göre görev dağılımı yapabilir ve mevcut işlem gücü verimli bir şekilde değerlendirilerek hızlı bir şekilde konuşma tanıma sonucu elde edilir.

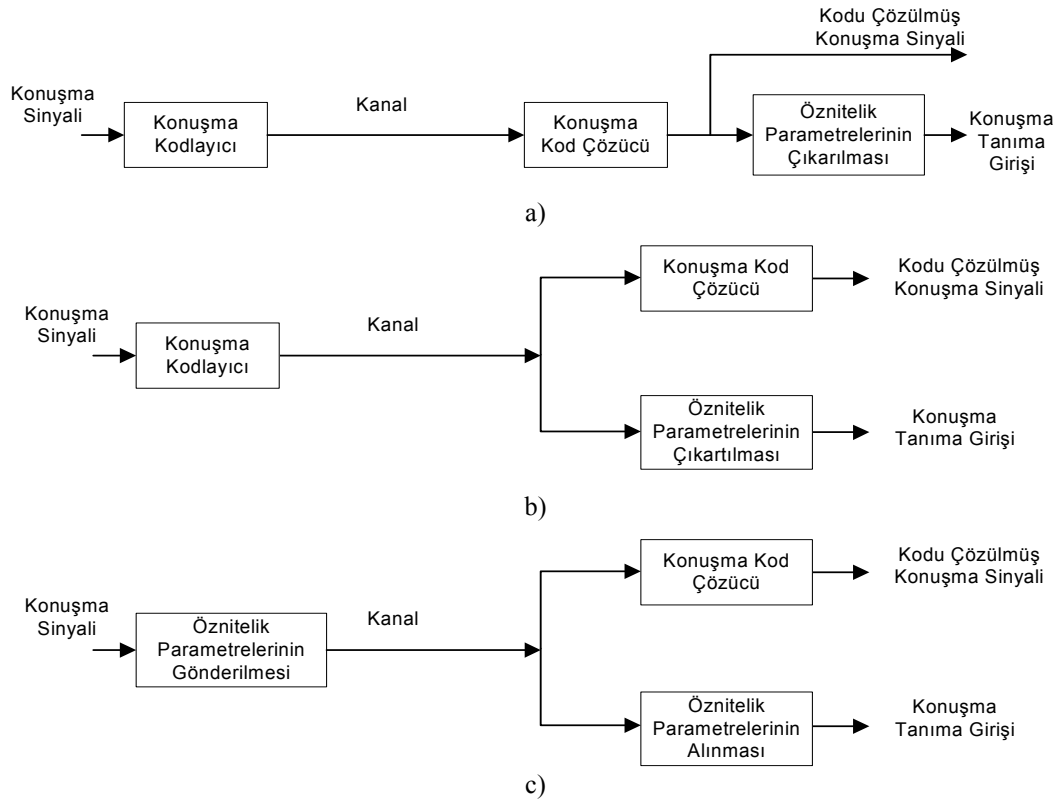


Şekil 3.6 Dağıtık yapıda konuşma tanıma modeli

Dağıtık yapıda konuşma tanıma sistemleri için uygulama alanı olarak sesle kontrol edilebilen web sayfaları ve gezgin telefon sistemlerinde ses kontrollü uygulamaları da gösterilebilir (Bernard and Alwan, 2002; Srinivasamurthy, et al., 2000). Özellikle web tarayıcıları üzerinden dağıtık yapıda uygulama geliştirilmesi konuşma tanıma çalışmalarının web tarayıcıları aracılığı ile dağıtık yapıda gerçekleşmesinin yolunu açmıştır (Goddeau, 1997; Bayer, 1996). Gezgin telefon sistemlerinde dağıtık yapıda konuşma tanıma uygulaması için karmaşıklığın azaltılması hedeflenmektedir. Dağıtık yapıda iletilen veri miktarı azalmaktadır (Milner 2001; Demichelis et. al., 2005; Bernard and Alwan, 2002; Srinivasamurthy, et al., 2000).

Kablosuz ya da gezgin telefon sistemlerinde dağıtık yapı modeli kullanılarak yapılan konuşma tanıma sistemleri Şekil 3.7'de gösterilen model üzerinde yoğunlaşmaktadır (Kim and Cox, 2001). Şekil 3.7(a)'da ses verisi kodlanarak gönderilir ve alıcı tarafında kod çözme işlemi yapıldıktan sonra ses öznelik vektörleri çıkartılır ve konuşma tanıma işlemi gerçekleştirilir. Bu yöntemde alıcı tarafında hesaplanan ses öznelik vektörleri kullanılan kodlayıcı ve kod çözücünden etkilenecektir.

Kod çözücülerin ses öznitelik vektörleri ve konuşma tanımaya olan etkiler değişik çalışmalarda incelenmiştir (Euler and Zinke, 1994; Lilly and Paliwal 1995; Besacier, et al., 2003). Şekil 3.7(b)'de gösterilen yöntemde gönderici tarafında ses kodlanarak gönderilmektedir ve alıcı tarafında kod çözme işlemi yapmadan ses öznitelik vektörleri kodlanmış parametrelerden elde edilerek konuşma tanıma işlemine sokulmaktadır. Doğrudan kodlayıcı parametrelerinden konuşma tanıma yapılması tanıma başarımını olumsuz etkilemektedir (Huerta and Stern, 1998). Buna karşılık, düşük bit hızlarına sahip kodlayıcılarda kod çözme işlemi yapmadan doğrudan kodlayıcı parametrelerinden konuşma tanıma yapıldığında tanıma performansının daha üstün olduğu da gösterilmiştir (Tuğaç ve İlk, 2004).

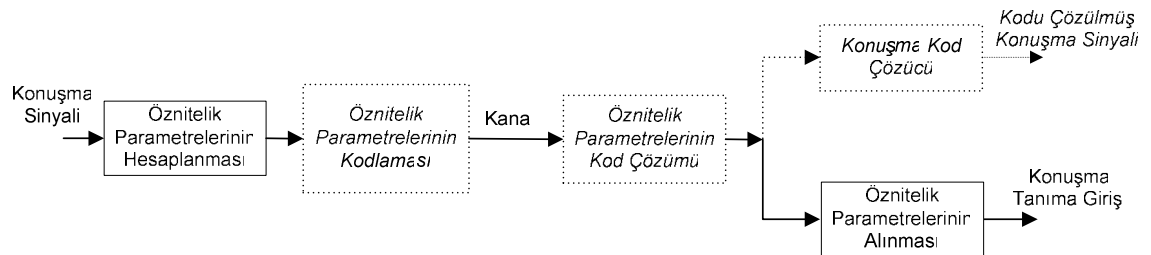


Şekil 3.7 Kablosuz konuşma tanıma için kullanılan üç yöntem, a) ses öznitelik parametrelerinin kod çözme işlemi sonrası elde edilmesi, b) ses öznitelik parametrelerinin kodlanmış bit katarından elde edilmesi, c) ses öznitelik parametrelerinin iletilmesi

Şekil 3.7(c), kablosuz telefon sistemlerinde kodlanmış ses yerine ses öznitelik parametrelerinin iletildiği durumu göstermektedir. Avrupa Haberleşme Standartları Enstitüsü (ETSI)'nün tanımlamış olduğu Aurora DSR standardı dağıtık yapıda konuşma tanıma için temel teşkil eder ve dağıtık yapıda ses paketleri yerine ses öznitelik vektörleri gönderilir (ETSI Standard 2000, 2003). Dağıtık yapıda konuşma tanıma sisteminde veri trafiğini azaltmak için ses öznitelik vektörleri değişik vektör nicemleme yöntemleri ile kodlanarak gönderilebilir (Digalakis, et al., 1998; Ramaswamy and Gopalakrishnan, 1998).

Bu çalışmada ses öznitelik vektörlerinin konuşma tanıma amaçlı olarak IP üzerinden iletildiği dağıtık yapı modeli kullanılmıştır. Ses öznitelik vektörleri 25ms çerçeve için elde edilmiştir. Öznitelik vektörü, 32-bit kayan noktalı formatta toplam 39 MFCC değerinden oluşmaktadır ve bu vektör verici tarafından kodlanmadan iletilmektedir. Bu çalışmada amaç, konuşma tanıma için en çok kabul gören ses öznitelik vektörleri olan MFCC parametrelerinin paket halinde IP üzerinden iletimi ve kayıplı durumların benzetimini yaparak, bu parametreler üzerinde paket kayıp telafi uygulamalarının IP üzerinden Türkçe sürekli konuşma tanıma başarımına etkilerini incelemek ve tanımadaki başarımı eniyileyecek paket kayıp telafi sistemi ve yöntemleri geliştirmektir.

Şekil 3.8’de öznitelik parametrelerinin kodlanmadan iletildiği model gösterilmiştir. Şekil 3.7(c)’deki durumdan farklı olarak öznitelik vektörleri hesaplandıktan sonra kodlanmadan IP paketi halinde iletilir. Alıcı tarafında ya ses tanıma yapılır ya da ses öznitelik parametrelerinden konuşma sinyali geri elde edilir.



Şekil 3.8 Ses öznitelik vektörlerinin kodlanmadan iletildiği dağıtık yapı modeli

BÖLÜM 4

PAKET KAYIPLARI VE TELAFİ YÖNTEMLERİ

4.1 Paket Kayıp Yapıları

Paket anahtarlama sistemlerinde İnternet protokolü ile ses verisinin UDP üzerinden taşınması sırasında kodlanmış ses verisinin ya da dağıtılmış konuşma tanıma yapılarında ses öznitelik parametrelerinin kaybolması kaçınılmazdır. Bu paket kayıpları, IP üzerinden ses tanıma başarımını düşürmektedir (Boulis et. al., 2002).

İnternet ortamında kullanılan protokoller gereği paket kayıplarının oluşması mevcut veri trafiğine çok bağlıdır ve zaman içerisinde kullanıcı sayısı ve uygulamalara bağlı olarak değişkenlik arz etmektedir. Her geçen gün artan İnternet kullanıcıları ve İnternet tabanlı uygulamalar, özellikle ses ve görüntü uygulamaları, paket trafiğini çok fazla arttırmakta bu da paket gecikmeleri ve kayıpları gibi parametreleri etkilemektedir.

İnternet ortamında paket kayıpları temelde şu sebeplerden dolayı oluşmaktadır (Hardman, et al., 1995):

- Yönlendirici ve ağ geçitlerindeki sıkışıklık
- Paket gecikmelerinin çok fazla olması
- İş istasyonlarındaki yük fazlalığından ve çoklu-görev işletim sistemlerindeki zorluklar.

Birçok uygulama, paketlerin çok fazla gecikmesi durumunda sistemdeki gecikmeleri önlemek için çok fazla geciken paketleri kayıp paket kabul eder. Bu durumda, varsa paket kayıp telafi yöntemi uygulanır ya da mevcut paketlerle uygulama çalışmaya devam eder.

IP üzerinden ses uygulamalarında, paket kayıplarının telafisi amacıyla telafi yöntemleri üzerinde çalışmak için Internet üzerinde oluşan kayıpların istatistiklerini bilmek ya da genel kayıp yapılarını anlamak gerekmektedir. Internet ortamında oluşan kayıpların yapısını ya da karakteristik özelliklerini bilmek telafi yöntemlerinin belirlenmesi konusundan önemlidir.

Internet ortamında paketlerin alıcıdan göndericiye kadar gittiği yol ve geçtiği yönlendiriciler her zaman aynı değildir. Bununla birlikte, iki farklı noktadaki alıcı ve verici konumundaki sistemde aynı anda alıp gönderdikleri paketlerin gecikmeleri ve kayıp oranları bile farklılık göstermektedir (Chua and Pheanis, 2006). Bu belirsizlik kesin bir kayıp tipi belirlemeye olanak vermesede, genel Internet ağı ve dağıtık yapıda paket trafiğini ve mevcut band genişliğini göz önünde bulundurarak test amaçlı model tanımlanabilir. UDP için değişik yerlerde ve değişen zaman dilimlerinde Internet ortamında oluşan paket gecikmeleri ve kayıplarının analizi, genel olarak meydana gelen kayıplar içinde patlamalı paket kayıplarının seyrek paket kayıplarına oranla daha fazla olduğunu göstermiştir (Shanghi, et al., 1993; Bolot, 1993; Borella, 2000; Jiang and Schulzrinne, 1999).

Paket kayıpları zamansal bağımlılık (temporal dependency) gösterir (Yajnik, et al., 1999; Jiang and Schulzrinne, 1999). Eğer paket P_n kayıp ise bu durumda P_{n+1} 'in de kayıp olma olasılığı yüksektir. Kayıpların zamansal bağımlılıklarının modellenmesi üzerinde de çalışmalar yapılmıştır (Sanneck, et al., 2000; Bolot, J., Fosse-Parisis, S., Towsley, D., 1999).

Tezde yapılan çalışmada dağıtık yapıda konuşma tanıma sistemi üzerinde durulduğu için, paket kayıp yapıları olarak patlamalı kayıplar yanında seyrek paket kayıpları durumları da incelenmiştir. Tez çalışmasında göz önüne alınan paket kayıp yapıları ve tanımlamaları şu şekildedir:

- *Rasgele Paket Kayıpları*: Herhangi bir zamanda tek bir paketin kayıp olduğu durumlar,

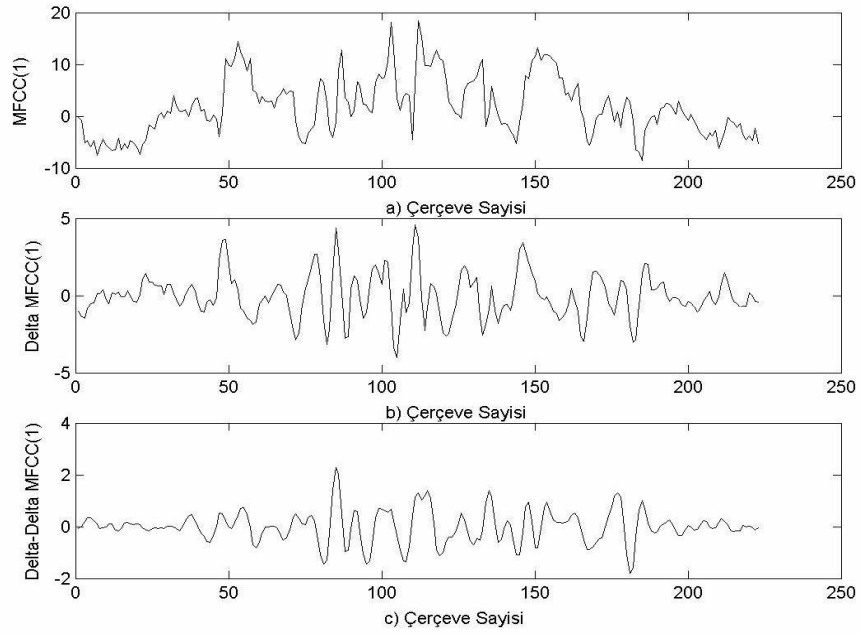
- *Seyrek Paket Kayıpları*: En fazla 3 paketin arka arkaya kaybolduğu ve zaman içerisinde 1, 2 ve 3 paketin farklı zaman ve kayıp aralıkları ile oluştuğu durumlar,
- *Patlamalı Paket Kayıpları*: 3'ten fazla paketin ardışıl olarak (aynı patlama anında) kayıp olduğu durumlar. Patlamalı paket kayıpları 10 ve 20 ardışıl paket kayıplarının oluştuğu durumlar olarak incelenmiştir.

Internet protokolu üzerinden iletilecek ses özniteliklerini taşıyan paketler 39 elemandan oluşmaktadır. Paketteki vektör elemanları kayan noktalı sayıdır. Ses öznitelik vektörüne ait paket 39x1'lik vektör formatında şu şekilde gösterilebilir:

$$P = [c_0; c_1; \dots; c_{12}; cd_0; cd_1; \dots; ca_0; ca_1; \dots; c_{12}].$$

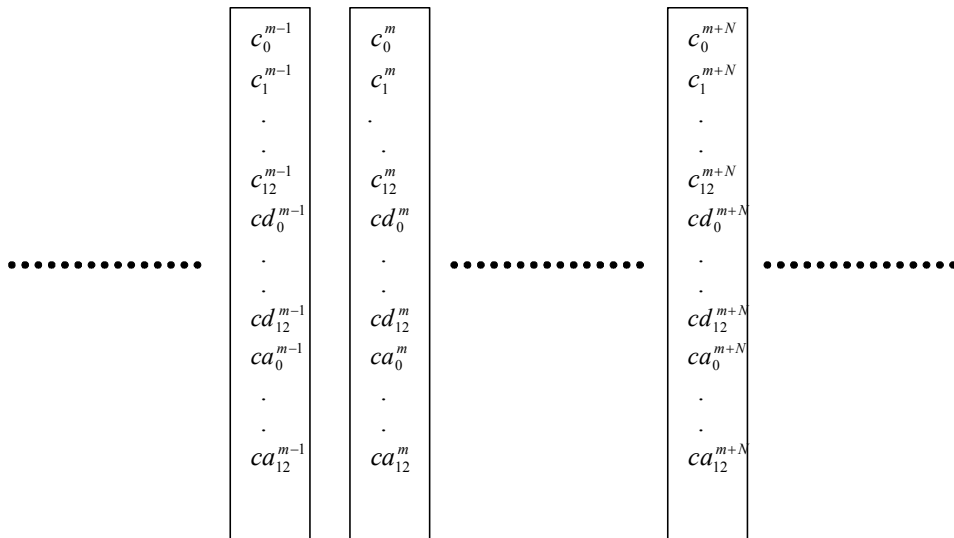
Eşitlikte, c_1, \dots, c_{12} MFCC katsayılarını; c_0 , çerçeve enerjisi (E); cd_0, \dots, cd_{12} , sesin dinamik özelliklerini yakalayabilmek için MFCC katsayıları ve LogE'nin birinci türevlerini (Δ MFCC, Δ LogE); ve ca_0, \dots, ca_{12} , MFCC katsayıları ve LogE'nin ikinci türevlerini ($\Delta\Delta$ MFCC, $\Delta\Delta$ LogE) göstermektedir.

Şekil 4.1'de test verisindeki “*bin ladin iCin geri sayIm*” cümlesine ait MFCC (1) değeri, birinci türevleri ve ikinci türevleri gösterilmektedir. Şekilde görüldüğü gibi cümleye ait MFCC değerlerinin birinci türevi ikinci türevinden daha fazla genliğe sahip, buna karşılık MFCC değerleri birinci ve ikinci türevlerinden oldukça fazla genlik değerine sahiptir.



Şekil 4.1 Test verisindeki örnek bir cümle için a-) MFCC (1), b-)Delta-MFCC (1), c-) Delta-Delta MFCC (1) değerleri.

IP üzerinden ses öznitelik vektörleri Şekil 4.2’de gösterildiği gibi paketler halinde iletilmektedir.

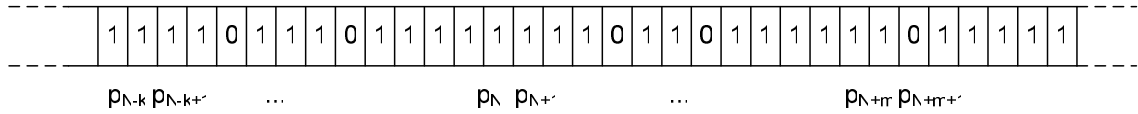


Şekil 4.2 Ses öznitelik vektörlerinin IP üzerinden iletimi ve paket gösterimi

Paket kayıpları tanımlanırken gözlemlenen vektör 1, kayıp vektörlerde ise 0 ile gösterilmiştir. Örneğin, zaman ekseninde N . paket gelmişse bu vektörü $P_N = [1]$ şeklinde gösterebiliriz, eğer $N+4$. paket kayıp ise $P_{N+4} = [0]$ şeklinde ifade edebiliriz. Paket kayıplarının gösteriminde bu format kullanılacaktır.

4.1.1 Rasgele paket kayıpları

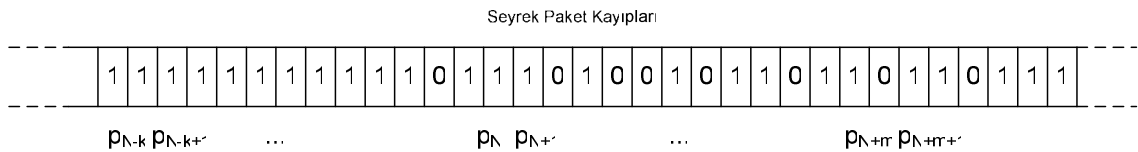
Rasgele paket kayıplarında kayıplar t_0 ve t_1 zaman aralığında bir kez olur (Zhang, et al., 2005). Şekil 4.3'te rasgele paket kayıpları gösterilmektedir. Şekilde her bir paket ses öznelik vektörlerine karşılık gelmektedir.



Şekil 4.3 Rasgele paket kayıplarının gösterimi

4.1.2 Seyrek paket kayıpları

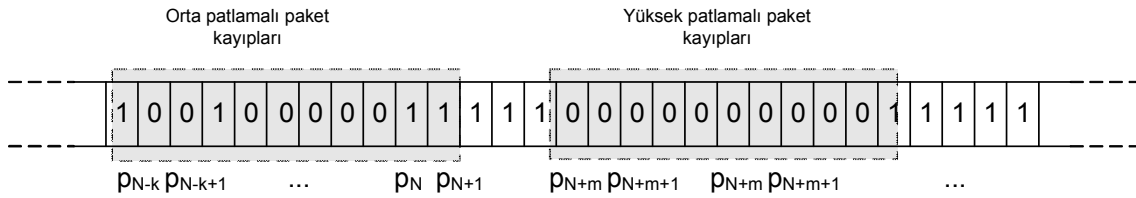
Seyrek paket kayıplarında t_0 ve t_1 arasında değişken kayıplar oluşacağı kabul edilmiştir. Şekil 4.4'te örnek seyrek paket kayıpları gösterilmektedir.



Şekil 4.4 Seyrek paket kayıplarının gösterimi

4.1.3 Patlamalı paket kayıpları

Patlamalı paket kayıp durumları t_0 ve t_1 zaman aralığında ardışıl paket kayıplarının çok fazla olduğu durumlardır. Patlamalı paket kayıpları, patlamanın derecesini göstermek için orta ve yüksek patlama şeklinde de ifade edilebilir. Şekil 4.5'te orta patlamalı ve yüksek patlamalı paket kayıpları gösterilmektedir.



Şekil 4.5 Orta ve yüksek patlamalı paket kayıplarının gösterimi

Paket kayıpları Şekil 4.3, 4.4 ve 4.5'deki herhangi kayıp durumundan birisi olabileceği gibi, bu üç durumun birleşimi şeklinde de olabilecektir. Gerçek Internet ortamında paket kayıplarını bu kayıpların birleşimi şeklinde düşünmek yararlı olabilir.

4.2 Paket Kayıp Telafi Yöntemleri

IP üzerinden ses iletiminde band genişliğinin fazla olmaması ve klasik telefonlardaki gibi özgül (dedicated) hattın olmaması sebebiyle klasik telefon haberleşmesinden farklı olarak ses iletilirken kodlanır ve bu şekilde daha az band genişliği kullanılır.

Telefon sistemlerinde 64 kbit/s bit hızı kullanılırken, IP üzerinden ses iletiminde daha düşük bit hızında kodlama yapan ses kodlayıcılar geliştirilmiştir (ITU-T

Recommendation G.729, 1996). Kodlanan ses verisi deęişik çerçeve uzunluklarında UDP üzerinde çalışan RTP ile iletilir.

Paket kayıplarının etkisini gidermek için hem gönderen hem de alan tarafında telafi işlemleri yapılabilmektedir. Ancak, gönderen tarafında yapılan her işlemin alıcı tarafın bilinmesi ve bunun tanımlanan standartlara uygun olması gerekir. Aksi durumda, alıcı tarafında gelen paket anlamsız paket olarak değerlendirilecek ve üzerinde işlem yapılmadan yitik (drop) paket işlemi görecektir.

4.3 Gönderici Tarafında Yapılan Paket Kayıp Telafi Yöntemleri

Gönderici tarafında yapılan paket kayıp telafileri aslında doğrudan telafi mekanizması deęil, alıcı tarafında telafinin daha kolay yapılabilmesi için artık verinin deęişik şekillerde gönderilmesini içerir. Gönderici tarafında artık veri gönderilen verinin özelliklerine baęlı ya da tamamen baęımsız olabilir, örneęin deęerlerin YA (OR)'lanması şeklinde olabilir (Perkins, et al., 1998).

4.3.1 Gönderici tarafında hata düzeltimi

Göndermede hata düzeltimi (FEC : Forward Error Correction) yönteminde amaç, paket kayıpları olduęu durumlarda alıcı tarafının paket kayıplarını telafi etmesini kolaylaştırmaktır. Bu amaçla asıl gönderilmek istenen paketlere ek olarak bazı paketler tekrarlanabildięi gibi paketler arasındaki istatistiksel bilgilerinde olabileceęi fazladan paketler gönderilir (Bolot, et al., 1999; Perkins, et al., 1998).

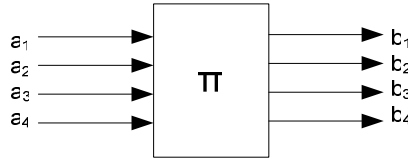
4.3.2 Serpiştirme

Serpiştirme (interleaving) tekniđi, tek girdili ve tek çıktıli sonlu-durum aygıtı olarak tanımlanabilir (Andrews, et al, 1997). Serpiştirme yönteminde gönderilecek paketlerin yerleri deđiştirilerek (Perkins, et al., 1988), paketler belirli bir permütasyonda, adeta serpiştirilerek, sistemin çıktısı olan paketlerin farklı sırada gönderilmesi sağlanır.

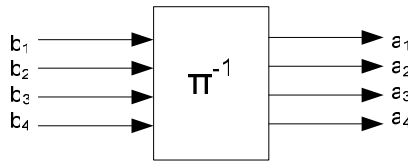
Şekil 4.6'de gösterildiđi gibi her a_i dizisi Π serpiştirme fonksiyonundan geçerek yeni b_i dizisi elde edilir.

$$a_i = b_{\Pi(i)} \quad (4.1)$$

Eşitlikte Π_i , a_i 'nin serpiştirme sonrası indis deđerini verir.



a. Serpiştirici



b. Geriserpiştirici

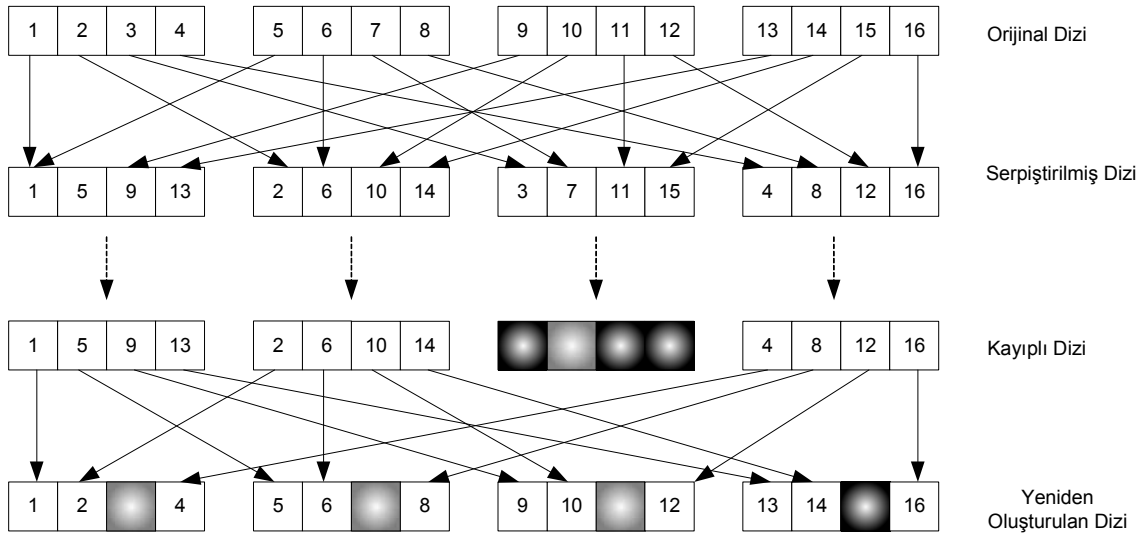
Şekil 4.6 Serpiştirici Gösterimi a-) Serpiştirici b-) Geriserpiştirici

Her serpiştirme işlemi için geri-serpiştirme tanımlanır (Andrews, et al, 1997). Şekil 4.6.b’de gösterildiği gibi serpiştirme işlemi sonucu elde edilen $\{b_n\}$ dizisinden geri-serpiştirme işlemi ile $\{a_n\}$ dizisi elde edilir. Geri-serpiştirme işlemi sonucu aynı a_i değeri elde edilir (Eşitlik 4.2).

$$\Pi^{-1}(i)$$

$$\Pi(\Pi^{-1}(i)) = i \quad (4.2)$$

4x4’lük serpiştirici örneği Şekil 4.7’de verilmiştir (Perkins, et al., 1998). Andrew (1997), yaptığı çalışmada değişik serpiştiriciler ve serpiştiricilerin neden olduğu gecikmeleri ve bellek kullanımını detaylı olarak incelemiştir. Serpiştiricilerde gecikme ve bellek kullanımı tanımlanan serpiştirme yöntemine ve seçilen dizi uzunluğuna bağlı olarak değişkenlik gösterir. Serpiştiriciler, paket kayıp telafi yöntemi olarak kullanıldığında gecikmeler ve bellek kullanımı çok önemli olmaktadır.



Şekil 4.7 4x4’lük serpiştirici örneği.

Serpiştirme yöntemi aslında paket kayıp telafi yöntemi değildir. Patlamalı kayıp durumlarında ardışık paketlerin kaybolmasını, böylece paket kayıp etkisini azaltmayı hedefler. Bu yöntem başarımlı olarak iyi olmasına karşın (James and Milner, 2004), patlamalı paket kayıplarında kayıp paketlerin sayısına bağlı olarak gecikme süresini ve bellek gereksinimini arttırmaktadır. Olasılık açısından serpiştirme sonucunda seyrek kayıplar olup geri serpiştirme sonucunda patlamalı bir kayıp bölgesi de oluşabilir.

4.3.3 Yeniden İletim

Yeniden iletim (retransmission) yönteminde, gönderilen paketler belirli bir kurala göre tekrarlanır. Alıcı tarafında tekrarlanan paketin (eğer kaybolan paket ile aynı ise) kaybolan paketin yerine konulması sağlanır. Bu yöntemde ne kadar tekrarlanan paket olacağı belli olmadığı için gerekirci olmayacak (non-deterministic) ve fazla tekrarlayan paketin olması İnternet trafiğini çok fazla arttırabilecektir. Ses kodlayıcılar için “redundant audio” yeniden iletim tekniği önerilmiş (Hardman, et al., 1995) ve standart olarak tanımlanmıştır (Perkins, et al., 1997).

4.4 Alıcı Tarafında Yapılan Paket Kayıp Telafi Yöntemleri

4.4.1 Yerine koyma

Yerine koyma yöntemi en basit ve işlemci gücü açısından en kolay yöntemdir. Dalgabiçimi ses kodlayıcılar için alıcı tarafında seyrek ve rasgele paket kayıpları için kullanılır. Ses kodlayıcıları için telafi yöntemi olarak yerine koyma yönteminde kayıp paket bir önceki paketle telafi edilebileceği gibi, kayıp paketin yerine sessizlik ya da beyaz gürültü (white noise) de konulabilir (Perkins, et al., 1998).

Dağıtık konuşma tanımada ise kayıp olan ses vektörlerinin yerine ETSI AURORA standardında bir önceki ses öznelik vektörü konulmaktadır (ETSI Standart, 2000; 2003).

4.4.2 Aradeğerleme

Aradeğerleme (interpolation) yöntemleri ses kodlayıcılarda kod çözücülerin başarımını arttırmak için, dalgabıçımı kodlayıcılarda dalgabıçımı aradeğerlemesi (Perkins, et al., 1998), ya da kodlanmış parametrelerin aradeğerlemesi için (Whang and Gibson, 2001) sık kullanılan bir yöntemdir. Aradeğerleme yöntemi yerine koyma metoduna göre daha fazla işlem gücü gerektirmektedir. Bu yöntem, kayıplı verinin karakteristiklerine göre cevap verebildiği için başarımı yüksektir (Perkins, et al., 1998).

Dağıtık yapıda konuşma tanıma çalışmalarında da ses öznelik vektörlerinin iletildiği durumlarda aradeğerleme yöntemleri kullanılmıştır (Mayorga, et al., 2002; Milner 2000; Milner and James, 2001)

4.4.2.1 Lagrange aradeğerleme

Aradeğerleme yöntemi çok genel ve eski bir yöntemdir (Meijering, 2002). Paket kayıp telafi yöntemi olarak aradeğerleme yöntemlerinin kullanılmasında aradeğerleme yönteminin aradeğerleme yapılacak verilerin, örneğin x -ekseninde, eşit olarak dağılmadığını kabul etmek gerekir. Lagrange polinomları bu şekilde düzgün dağılmayan (nonuniformly spaced) veriler için iyi bir aradeğerleme yöntemidir.

Basit bir aradeğerleme fonksiyonu aşağıdaki şekilde gösterilmiştir (Conte, 1965).

$$p(x) = L_0(x)f_0 + L_1(x)f_1 + L_2(x)f_2 + \dots + L_N(x)f_N = \sum_{k=0}^n L_k(x)f_k \quad (4.3)$$

Eşitlik 4.3'te $L_k(x)$ ($k = 0, 1, \dots, n$) derecesi n 'i geçmeyen polinomu ifade etmektedir. Her bir $L_x(k)$ polinomu Eşitlik 4.4'teki gibi ifade edilir.

$$L_k(x) = \frac{(x - x_0)(x - x_1) \dots (x - x_{k-1})(x - x_{k+1}) \dots (x - x_n)}{(x_k - x_0)(x_k - x_1) \dots (x_k - x_{k-1})(x_k - x_{k+1}) \dots (x_k - x_n)} \quad (4.4)$$

$L_x(k)$ polinomu n . dereceden bir polinom olduğu için $p(x)$ 'in derecesi de n 'e eşit ya da n 'den küçük olacaktır. Eşitlik 4.4 şu şekilde daha basit halde gösterilebilir:

$$L_k(x) = \frac{\prod_{\substack{j=0 \\ j \neq k}}^n (x - x_j)}{\prod_{\substack{j=0 \\ j \neq k}}^n (x_k - x_j)} \quad (4.5)$$

Lagrange aradeğerlemede $n=1$ seçilirse $p(x)$,

$$p(x) = f_0 + \frac{f_1 - f_0}{x_1 - x_0} (x - x_0) \quad (4.6)$$

şeklinde gösterilir ve doğrusal (linear) aradeğerleme olarak ifade edilir.

4.4.2.2 Spline aradeğerleme

Spline, belirli bir ağırlık katsayısı ile tanımlanan her noktadan geçen esnek bir şerit şeklinde düşünülebilir. Aradeğerleme uygulamalarında sıklıkla kullanılan bir yöntemdir. Cubic spline, her $[a, b]$ aralığı için farklı 3. derece polinom tanımlanarak elde edilir ve her i . ve $i+1$. noktalardaki 3. derece fonksiyonların birinci türevi (eğimi) ve ikinci türevi (eğim derecesi) aynı olmalıdır (Gerald, 1970).

Genel anlamda n . dereceden $S(x)$ Spline fonksiyonlar şu şartları sağlamalıdır:

- $S(x)$, her $[x_i, x_{i+1}]$ alt-aralığında n . dereceden ya da daha küçük bir polinomdur.
- $S(x)$, $S'(x)$, $S''(x)$, ..., $S^{n-1}(x)$ $[x_i, x_{i+1}]$ aralığında süreklidir.
- $S(x)$, $[x_i, x_{i+1}]$ aralığında her noktada tanımlıdır.

$[x_i, y_i]$ ve $[x_i, y_{i+1}]$ noktaları arasında i . aralık için 3. derece bir fonksiyonu

$$y = a_i(x - x_i)^3 + b_i(x - x_i)^2 + c_i(x - x_i) + d_i \quad (4.7)$$

şeklinde tanımlayabiliriz. Bu durumda, $[x_i, x_{i+1}]$ aralığı için her iki noktata 3. derece polinomlarını

$$y_i = a_i(x - x_i)^3 + b_i(x - x_i)^2 + c_i(x - x_i) + d_i = d_i, \quad (4.8)$$

$$\begin{aligned} y_{i+1} &= a_i(x_{i+1} - x_i)^3 + b_i(x_{i+1} - x_i)^2 + c_i(x_{i+1} - x_i) + d_i \\ &= a_i h_i^3 + b_i h_i^2 + c_i h_i + d_i, \end{aligned} \quad (4.9)$$

şeklinde elde ederiz. Eşitlik 4.8 ve 4.9'da h_i , i . aralık için Δx değerini göstermektedir. Her iki noktada sürekliliği sağlamak ve birinci ve ikinci derece türevlerinin aynı olduğunu sağlamak için Eşitlik 4.7'nin birinci ve ikinci türevlerini

$$y' = 3a_i(x - x_i)^2 + 2b_i(x - x_i) + c_i, \quad (4.10)$$

$$y'' = 6a_i(x - x_i) + 2b_i \quad (4.11)$$

şeklinde ifade edebiliriz. Matematiksel olarak kolay hesaplayabilmek için $[x_i, x_{i+1}]$ noktalarındaki 2. türevlerini S_i ve S_{i+1} şeklinde gösterirsek, Eşitlik 4.10 ve 4.11'deki gösterimi elde edebiliriz.

$$S_i = 6a_i(x_i - x_{i-1}) + 2b_i = 2b_i \quad (4.12)$$

$$S_{i+1} = 6a_i(x_{i+1} - x_i) + 2b_i = 6a_i h_i + 2b_i \quad (4.13)$$

Bu durumda,

$$b_i = S_i / 2, \quad (4.14)$$

$$a_i = (S_{i+1} - S_i) / 6h_i, \quad (4.15)$$

şeklinde yazılabilir.

a_i , b_i ve d_i katsayılarının S ve y 'ye bağlı ifadelerini Eşitlik 4.8, 4.14 ve 4.15'teki gibi kullanarak Eşitlik 4.9'dan

$$c_i = \frac{y_{i+1} - y_i}{h_i} - \frac{2h_i S_i + h_i S_{i+1}}{6}, \quad (4.16)$$

katsayıları elde edilebilir.

Eşitlik 4.10'da x_i ve x_{i-1} noktalarında eğim aynı olmalıdır. Buna göre a , b , c ve d katsayılarını eşitlikte yerine koyarsak

$$h_{i-1} S_{i-1} + (2h_{i-1} + 2h_i) S_i + h_i S_{i+1} = 6 \left(\frac{y_{i+1} - y_i}{h_i} - \frac{y_i - y_{i-1}}{h_{i-1}} \right) \quad (4.17)$$

elde ederiz.

Eşitlik 4.17'yi n adet nokta için matris şeklinde yazabiliriz.

$$\begin{array}{cccccccc|c} h_2 & -(h_1 + h_2) & h_1 & 0 & 0 & \cdot & \cdot & \cdot & 0 & S_1 \\ h_1 & 2(h_1 + h_2) & h_2 & 0 & 0 & & & & 0 & S_2 \\ 0 & h_2 & 2(h_2 + h_3) & h_3 & 0 & & & & 0 & S_3 \\ 0 & 0 & h_3 & 2(h_3 + h_4) & h_4 & & & & 0 & S_4 \\ \cdot & & & & & & & & & \cdot \\ \cdot & & & & & & & & & \cdot \\ \cdot & & & & & & & & & \cdot \\ 0 & \cdot & 0 & 0 & & h_{n-1} & -(h_{n-2} + h_{n-1}) & h_{n-2} & & S_n \end{array}$$

$$= 6 \begin{vmatrix} \frac{y_3 - y_2}{h_2} - \frac{y_2 - y_1}{h_1} \\ \frac{y_4 - y_3}{h_3} - \frac{y_3 - y_2}{h_2} \\ \frac{y_5 - y_4}{h_4} - \frac{y_4 - y_3}{h_3} \\ \cdot \\ \cdot \\ \cdot \\ 0 \end{vmatrix} \quad (4.18)$$

Eşitlik 4.18'de S_i değerleri hesap edildikten sonra a_i, b_i, c_i ve d_i katsayıları elde edilir.

4.4.3 Kestirim yöntemleri

İstatiksel yöntemler kayıp veri bilgisi için eldeki verinin dağılımı hakkında önsel (*prior*) bilgiyi kullanırlar. Kayıp verinin elde edilmesi (kestirimi) için eldeki veriye ait olasılık dağılımı ya da dağılıma ait bazı parametrelerin bilinmesi gerekir.

Rassal değişkenler olasılık yoğunluk fonksiyonları (oyf) ile ifade edilirler. Rassal değişkene ait olasılık yoğunluk fonksiyonlarının parametreleri o dağılımın karakteristiğini belirtir. Rassal değişkenler içeren uygulamalarda teorik geliştirmeleri basite indirgemek için çok yararlı özelliklere sahip Gaussian dağılımı kullanılır.

Ortalama değeri μ ve varyansı σ^2 olarak ifade edilen X rassal değişkeni

$$X \sim \theta(\mu, \sigma^2)$$

şeklinde ifade edilebilir. Bu durumda, sayıl (scalar) Gaussian olasılık yoğunluk fonksiyonu Eşitlik 4.19'daki gibi gösterilir.

$$p_x(x) = \frac{1}{\sigma\sqrt{2\pi}} \exp\left[-(x - \mu)^2 / 2\sigma^2\right] \quad (4.19)$$

Benzer şekilde, k boyutlu Gaussian rassal vektör X , ortak değışinti matrisi Σ olarak gösterilirse, rassal vektöre ait oyf

$$P_X(X) = \frac{1}{(2\pi)^{k/2} |\Sigma|^{1/2}} \exp\left[-\frac{1}{2}(X - \mu)^T R^{-1}(X - \mu)\right] \quad (4.20)$$

ile ifade edilir. Eşitlikte μ ortalama vektörü göstermektedir:

$$\mu = E[X].$$

Σ ise $k \times k$ boyutlu ortak değışinti matrisidir (ODM) ve řu şekilde ifade edilir;

$$\Sigma = E[(X - \mu)(X - \mu)^T] = E[XX^T] - \mu\mu^T.$$

İki farklı rassal değışken X ve Y kabul edildiğinde bunların ortak olarak Gaussian (jointly Gaussian) olabilmeleri için ařağıdaki şekilde gösterilmeleri gerekir:

$$P_{x,y}(X,Y) = A \exp\left[-\frac{1}{2(1-\rho^2)} \left\{ \frac{(X - \mu_x)^2}{\sigma_x^2} + \frac{(Y - \mu_y)^2}{\sigma_y^2} - \frac{2\rho(X - \mu_x)(Y - \mu_y)}{\sigma_x\sigma_y} \right\}\right] \quad (4.21)$$

Eşitlikte $A = \frac{1}{2\pi\sigma_x\sigma_y\sqrt{1-\rho_{xy}^2}}$ olarak ifade edilir ve ρ değışinti değerini gösterir.

Kestirim yöntemlerinde X 'e ait herhangi bir değeri kestirilmek istendiğinde en basit yöntem olarak kestirilen değeri

$$\hat{x} = \mu_x$$

olacak şekilde seçilebilir ve bu yöntem önsel kestirim olarak adlandırılır. Eğer, X 'e ait bir değer kestirileceği zaman X ile değışintiye sahip Y biliniyorsa koşullu (conditional) olasılık kullanılabilir. Koşullu olasılık şu şekilde tanımlanır.

$$P_{x|y}(X | Y) = P(X | Y) = \frac{P(X, Y)}{P(Y)}.$$

Bu ifade Eşitlik 4.19 ve 4.21'in sayıl halini kullanarak

$$\begin{aligned} P(x | y) &= \frac{\frac{1}{2\pi\sigma_x\sigma_y\sqrt{1-\rho_{xy}^2}} \exp\left[-\frac{1}{2(1-\rho^2)} \left\{ \frac{(x-\mu_x)^2}{\sigma_x^2} + \frac{(y-\mu_y)^2}{\sigma_y^2} - \frac{2\rho(x-\mu_x)(y-\mu_y)}{\sigma_x\sigma_y} \right\}\right]}{\frac{1}{\sigma_y\sqrt{2\pi}} \exp\left[-\frac{1}{2\sigma_y^2}(y-\mu_y)^2\right]} \\ &= \frac{1}{2\pi\sigma_x\sigma_y\sqrt{1-\rho_{xy}^2}} \exp\left[-\frac{1}{2\sigma_x^2(1-\rho)} \left(x - \left(\mu_x + \frac{\sigma_x}{\sigma_y}\rho(y-\mu_y)\right)\right)^2\right] \end{aligned} \quad (4.22)$$

şeklinde yazılabilir. Kare tamamlama yöntemi kullanılarak elde edilen Gaussian dağılıma sahip y 'nin ortalama değeri

$$E[X | y] = \mu_x + \frac{\sigma_x}{\sigma_y}\rho(y - \mu_y) \text{ olarak hesaplanır.}$$

Bu durumda kestirilecek X değeri için

$$\hat{x} = \mu_x + \frac{\sigma_x}{\sigma_y}\rho(y - \mu_y) \quad (4.23)$$

yazılabilir. Eğer değışinti değeri $\rho = 0$ ise bu durumda kestirilen değeri öncül ortalama değeri (prior mean) olacaktır.

Rassal deęişken olan $X = \{X_1, X_2, \dots, X_n\}$ 'e ait parametre ifadesini ϕ ile gösterirsek, Gaussian daęılımlı verinin parametrik modeli $P(X | \phi)$ şeklinde ifade edilebilir.

Paket kayıplarının telafisi için kayıp ses öznitelik vektörlerini X_m , gözlemlenen ses öznitelik vektörlerini X_o ile gösterirsek, bu durumda $P(X | \phi)$ 'i kullanarak koşullu olasılık daęılımını $P(X_m | X_o; \phi)$ ve koşullu daęılımını $P(X_o | X_m; \phi)$ şeklinde elde edebiliriz.

İstatistiksel kestirim yöntemleri, spektogram kayıplı ses sinyalinin kestirimi (Raj, 2000) ve IP üzerinden ses iletiminde kayıp ses öznitelik vektörlerinin hesap edilmesi için uygulanmıştır (Milner and James, 2006).

Bu çalışmada kestirim yöntemleri olarak en küçük ortalama kare hata, en büyük olabilirlik ve en büyük sonsal kestirim yöntemi incelenmiş ve en büyük sonsal kestirim yöntemi uygulamaları yapılmıştır.

4.4.3.1 En küçük ortalama kare hata kestirimi

En küçük ortalama kare hata (minimum mean squared error) kestirimi yönteminde kayıp veri, kestirilen ve özgün değerler arasındaki beklenen ortalama kare hatayı enküçültecek (minimize) şekilde seçilir.

$$\hat{X}_m = \arg \min_{X_m} \left\{ E \left[\left\| X_m - X_m^t \right\|^2 \mid X_o \right] \right\} \quad (4.24)$$

Eşitlikte \hat{X}_m , kestirilen kayıp paketleri; ve X_m^t , kayıp paketlerin özgün değerlerini göstermektedir.

4.4.3.2 En büyük olabilirlik kestirimi

En büyük olabilirlik kestirimi (*maximum likelihood*) yönteminde kestirilecek değerler gözlemlenen değerlere göre en büyük koşullu olabilirlik değerine sahip olacak şekilde seçilir:

$$\hat{X}_m = \arg \max_{X_m} \{P(X_o | X_m, \phi)\} \quad (4.25)$$

Bu yöntemde sadece gözlemlenen vektörler kullanılır ve olasılık dağılımına ait, $P(X_m | \phi)$, herhangi bir önsel bilgi kullanılmaz (Moon and Stirling, 1999).

4.4.3.3 En büyük sonsal kestirimi

En büyük sonsal kestirim (*maximum a posteriori*) yönteminde kestirilecek değerler gözlemlenen değerlere göre en büyük olabilirlik değerine sahip olacak şekilde seçilir.

$$\hat{X}_m = \arg \max_{X_m} \{P(X_m | X_o, \phi)\} \quad (4.26)$$

$P(X_m | X_o; \phi)$ Gaussian olduğu durumlarda en küçük ortalama kare kestirimi ve en büyük sonsal kestirim aynı olacaktır (Raj, 2000).

IP üzerinden dağıtık yapı modelinde ses iletiminde paket kayıplarının patlamalı olduğu durumlarda kestirim yöntemleri aradeğerleme yöntemlerine göre daha başarılı sonuçlar vermektedir (Milner and James, 2006).

Ses öznitelik vektörlerinin Gaussian dağılıma sahip olduğu ve dağılım parametreleri μ , ortalama değer ve Σ , ortak değişinti matrisi olduğu kabul edilirse, gözlemlenen ses öznitelik vektörleri (X_o) için parametrik model,

$$P(X_m; \mu, \Sigma),$$

kayıp ses öznitelik vektörleri (X_m) için parametrik model,

$$P(X_o; \mu, \Sigma)$$

şeklinde ifade edilebilir.

Gözlemlenen ve kayıplı ses öznitelik vektörlerini $X = [X_o, X_m]$ olarak gösterdiğimizde, sistemin ortalama vektörlerini

$$\mu = [\mu_o, \mu_m]$$

ve ortak ilişinti matrislerini

$$\Sigma = \begin{bmatrix} \Sigma_{oo} & \Sigma_{om} \\ \Sigma_{mo} & \Sigma_{mm} \end{bmatrix}$$

şeklinde ifade edebiliriz. Bu ifadede Σ_{mo} , X_m ve X_o arasındaki çapraz ortak değişinti matrisin göstermektedir ve $\Sigma_{mo} = \Sigma_{om}^T$ 'dir (Raj, 2000). Bu durumda koşullu olasılık dağılımı Eşitlik 4.27'deki gibi olacaktır.

$$P(X_m | X_o, \Sigma) = C \exp(-0.5(X_m - \mu_m - \Sigma_{mo}\Sigma_{oo}^{-1}(X_o - \mu_o))^T (\Sigma_{mm} - \Sigma_{mo}\Sigma_{oo}^{-1}\Sigma_{om})^{-1})$$

$$(X_m - \mu_m - \Sigma_{mo}\Sigma_{oo}^{-1}(X_o - \mu_o)) \quad (4.27)$$

Eşitlik 4.27’de C normalize etme katsayısıdır (normalizing constant). Eşitlik 4.26 ve 4.27’yi kullanarak kestirilen vektör

$$\hat{X}_m = \arg \max_{x_m} \{P(X_m | X_o, \mu, \Sigma)\} = \mu_m + \Sigma_{mo} \Sigma_{oo}^{-1} (X_o - \mu_o) \quad (4.28)$$

şeklinde ifade edilebilir (Raj, 2000).

Kestirilecek kayıp vektör bulunurken, kayıp vektörler ve gözlemlenen vektörler arasındaki değişinti ve kayıp vektörlerin ortalama değerlerinden yararlanılır.

Paket kayıplarının patlamalı olduğu durumlarda ses öznitelik vektörleri üzerinde kestirim yöntemi uygulamaları, eğitim verisinin geniş anlamda durağan olduğu kabul edilerek yapılmıştır (Milner and James, 2006). Eğitim verisinin ortalama değeri,

$$\mu_G = E[x_n] \quad (4.29)$$

Eşitlik 4.28 ve 4.29 kullanılarak test verisinde paket kayıplarının kestirimi için 3. derece ortak değişinti tensörü, Σ_G , tanımlanabilir ve bu tensörün her elemanı $\sigma^{i,j,\tau}$, i . ve j . ses öznitelik vektörleri arasında τ zaman farkıyla ortak değişinti değerini gösterir (Milner and James, 2006).

$$\sigma^{i,j,\tau} = E[(x_n^i - \mu_G^i)(x_{n+\tau}^j - \mu_G^j)] \quad (4.30)$$

Eşitlik 4.30’da ses öznitelik vektörünün elemanları arasındaki ortak değişinti de ele alınmaktadır. Ancak, paket kayıp telafilerinde ses öznitelik vektörünün zaman ekseninde diğer vektörlerle olan değişinti ilişkisi, kendi elemanları arasındaki ilişitiden daha çok önem kazanmaktadır. Bu durumda, Eşitlik 4.30,

$$\sigma^{k,\tau} = E[(X_n^k - \mu_G^k)(X_{n+\tau}^k - \mu_G^k)] \quad (4.31)$$

şeklinde ifade edilir. Eşitlikte k , quefrecny değerini, τ ise çapraz ortak değışintinin (cross-covariance) hesap edildiđi zaman farkını göstermektedir.

Eşitlik 4.28’de elde edilen en büyük sonsal kestirim (EBSK) ifadesi aslında bir doğrusal bağlanım (linear regression) ifadesi olarak düşünülebilir. Bir doğrusal bağlanım probleminde elde olan veriler arasındaki ilişki fonksiyonel bir ilişki ile ifade edilmeye çalışılır.

İki boyutlu bir doğrusal bağlanım probleminde, n noktada değer alan x_i değerleri Eşitlik 4.32’deki gibi bir doğru olarak ifade edilebilir.

$$y_i \approx ax_i + b, \quad i = 1, 2, \dots, n \quad (4.32)$$

Eşitlik 4.32 hata terimi ile beraber matris formunda Eşitlik 4.33’teki gibi ifade edilebilir.

$$\begin{bmatrix} y_1 \\ y_2 \\ \cdot \\ \cdot \\ y_n \end{bmatrix} = \begin{bmatrix} ax_1 + b \\ ax_2 + b \\ \cdot \\ \cdot \\ ax_n + b \end{bmatrix} + \begin{bmatrix} e_1 \\ e_2 \\ \cdot \\ \cdot \\ e_n \end{bmatrix} \quad (4.33)$$

$Y = [y_1, y_2, \dots, y_n]^T$, $\varepsilon = [e_1, e_2, \dots, e_n]^T$, $A = [a \ b]^T$ ve $X = [x_1 \ 1; x_2 \ 1, \dots, x_n \ 1]^T$ olarak ifade edilirse, Eşitlik 4.33 şu şekilde yazılabilir:

$$Y = AX + \varepsilon \quad (4.34)$$

En küçük kareler kestirimine göre A ’nın kestirimi,

$$\hat{A} = (X^T X)^{-1} X^T Y \quad (4.35)$$

olarak ifade edilir. X ve Y 'nin ortalama değeri sıfır olarak kabul edilirse, $X^T X = n\Sigma_{XX}$ ve $X^T Y = n\Sigma_{XY}$ olacaktır. Bu durumda, Eşitlik 4.35,

$$\hat{Y} = \hat{A}X + \varepsilon = (\Sigma_{XY})^T (\Sigma_{XX})^{-1} X \quad (4.36)$$

olarak yazılacaktır. Eğer, X ve Y 'nin ortalamaları μ_X ve μ_Y olarak kabul edilirse,

$$\hat{Y} = \mu_Y + (\Sigma_{XY})^T (\Sigma_{XX})^{-1} (X - \mu_X) \quad (4.37)$$

olur.

Eşitlik 4.37, Eşitlik 4.28'teki EBSK yöntemi ile aynı etkiye sahiptir.

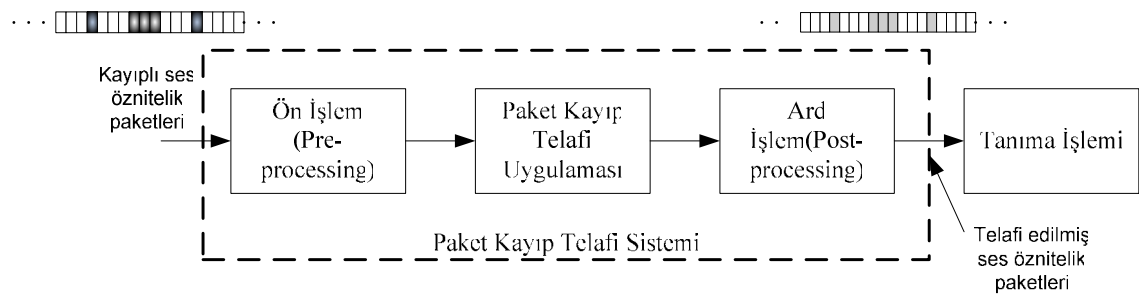
BÖLÜM 5

PAKET KAYIP TELAFİ SİSTEMİ

Paket kayıp telafisinde, paket kayıp tipine göre telafi yönteminin kullanılması hem işlem gücü hem de elde edilecek başarımlar açısından önemlidir. Örneğin, seyrek ya da rasgele paket kayıp durumlarında kestirim tekniklerinin kullanılması gerekli işlem gücü ve başarımlar oranı açısından tercih edilmeyebilir. Benzer şekilde, patlamalı kayıp durumlarında yerine koyma ya da aradeğerleme yöntemlerinin başarımları iyi olmayabilir.

Paket kayıp telafisinde işlem gücü ve sistem başarımlarının yanında kullanılacak tampon belleğin uzunluğuda çok önemlidir. Tampon bellek, işlem gücü ve hafıza gereksinimi açısından en iyi (optimum) uzunluk değerinde olmalıdır.

Şekil 5.1’de kayıp telafisinin işlem gücü, başarımlar ve bellek uzunluğunun eniyilenecek yapılabileceği sistem yapısı gösterilmektedir. Bu sistem ön işlem, telafi uygulaması ve ardışlemeden oluşmaktadır.



Şekil 5.1 Paket kayıp telafi sistemi

Tasarlanan sistemde, uygulanacak telafi algoritması öncesinde bir ön işlem yapılarak dinamik bellek uzunluğu ve uygulanacak telafi yöntemi belirlenir. Ön

işlemden sonra telafi yöntemi uygulanır ve telafi uygulanmış bellek üzerinde ardışık işlem yapılır. Ardışık işlemden geçen paketler tanıma işlemi için konuşma tanıma ünitesine aktarılır.

Temel olarak önerilen paket kayıp telafi sisteminin hedefleri şunlardır:

- Hız ve başarımlar açısından en uygun algoritmanın seçilmesi
- Algoritmanın ve belleğin en iyi şekilde kullanılması açısından en uygun bellek miktarının belirlenmesi
- Uyarlamalı iyileştirme yöntemlerini kullanarak zaman içinde sistem başarımlarının artırılması

5.1 Ön İşlem

Ön işlem aşamasında gelen UDP paketleri tampon bellekte toplanır ve paket başlıklarına (tanımlanan sistemde bu sadece paket numarasında olabilir) göre kayıp paketler ve kayıp yapısı belirlenir. Paket kayıp yapısına göre tampon bellek uzunluğu belirlenir.

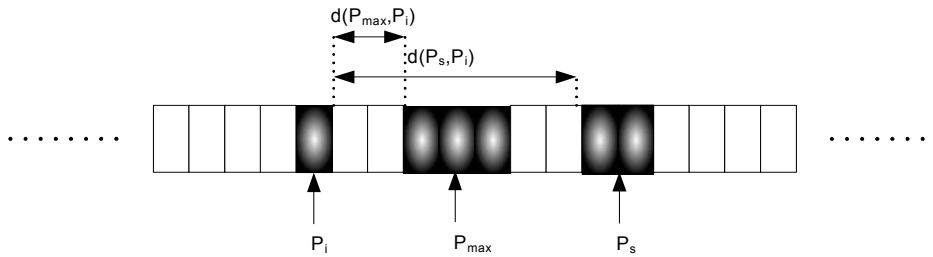
5.2 Dinamik Bellek Uzunluğunun Belirlenmesi

Ön işlem aşamasında paket kayıplarının durumuna göre tampon bellekteki paketler ya daha fazla kayıt için bekletilir ya da telafi yöntemi başlatılır. Kayıpların patlamalı olduğu durumlarda kayıplardan sonra yeteri kadar paketin gelmesi gerekmektedir. Bu nedenle gerekli paket sayısı arttığı için tampon bellek uzunluğu artacaktır. Kayıpların rasgele ya da seyrek olduğu durumlarda paket kayıplardan sonra daha az paket gerekir. Bu durumda bellek uzunluğunun kısa tutulması yeterli olacaktır.

Bellek uzunluğu (BU) aşağıdaki formül ile belirlenir:

$$BU = \begin{cases} J + d(P_s, P_i) + M & , P_i \neq 0 \text{ ve } P_s \neq 0 \\ J + M + P_{\max} & , \text{diğer} \end{cases} \quad (5.1)$$

Eşitlik 5.1'de P_i ve P_s tampon bellekteki ilk ve son paket kayıp merkezini göstermektedir. Eşitlikte J ilk paket kayıp merkezinden önce gözlemlenen paket sayısıdır ve $J \geq P_i$ şartını sağlamalıdır. $d(P_s, P_i)$ değeri iki paket kayıp merkezi arasındaki uzaklığı göstermektedir. M ise kayıp merkezinden sonra gözlemlenen ve bellekte tutulan paket sayısıdır. Şekil 5.2'de (Eşitlik 5.1)'deki ifadenin uzaklık ölçümlerini göstermektedir.



Şekil 5.2 Ön işlem aşamasında dinamik tampon bellek uzunluğunun belirlenmesi

5.3 Telafi Algoritmasının Belirlenmesi

Uygulanacak kayıp telafi yöntemi için patlama oranlarının hesap edilmesi gerekmektedir. Her kayıp merkezi için kayıp oranı

$$\frac{P_{\max}}{10^{P_{\max}}} \quad (5.2)$$

şeklinde gösterilebilir. Bu durumda çerçevedeki toplam kayıp oranı (%)

$$\frac{1}{L} \sum_{i=1}^P P_i \quad (5.3)$$

ile ifade edilir. Eşitlik 5.3'te L toplam çerçeve sayısını vermektedir.

Uygulanacak telafi yöntemi (TY) belirlenirken Eşitlik 5.4'ten yararlanır. Eşitlikte J değeri kayıpların başladığı zamana kadar gözlemlenen vektör sayısını göstermektedir. N değeri ise kayıp merkezi sayısını göstermektedir. Kayıp merkezlerindeki vektör sayısının toplamı gözlemlenen vektör sayısına göre çok fazla ise yerine koyma yöntemi, eşit veya az ise kestirim yöntemi uygulanır. Kayıp vektör sayısı gözlemlenen vektör sayısından biraz fazla ya da yakın ise aradeğerleme yöntemleri uygulanır.

$$TY = \begin{cases} \text{yerine koyma} & , \sum_{i=1}^N \frac{P_N}{10^{P_N}} \gg \frac{J}{10^J} \\ \text{aradeğerleme} & , \sum_{i=1}^N \frac{P_N}{10^{P_N}} > \frac{J}{10^J} \\ \text{kestirim} & , \sum_{i=1}^N \frac{P_N}{10^{P_N}} \leq \frac{J}{10^J} \end{cases} \quad (5.4)$$

5.4 Ardişlem

Ardişlem aşamasında telafi edilmiş ses öznitelik vektörleri dışında her yeni gelen öznitelik vektörü ile ortalama değer ve ortak değişinti değerleri güncellenir. Sistemde güncelleme işlemine telafi edilmiş ses öznitelik değerleri katılmaz.

Ortalama değer ve ortak değişinti değerlerinin gözlemlenen ses öznitelik paketleri üzerinden yeniden hesap edilmesi için τ değerine bağlı olarak bellek ayrılır. Kullanılan belleğin en büyük değeri patlama merkezindeki kayıp ses öznitelik vektörü sayısı ve kestirim yönteminde kullanılan kayıptan önce ya da sonra gözlemlenen ve bellekte tutulan ses öznitelik vektör sayısının toplamına eşittir ($\tau + 1$).

BÖLÜM 6

UYARLAMALI KESTİRİM YÖNTEMİ

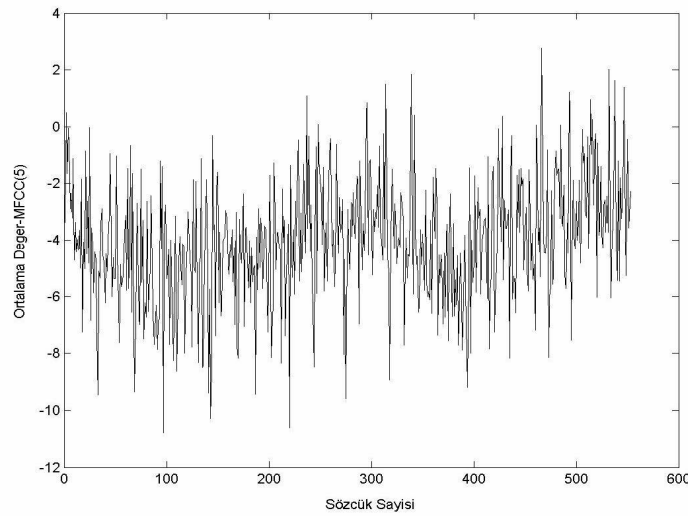
En büyük sonsal kestirim (EBSK) yöntemi gürültülü ortamlarda, eksik spektrogram olduğu durumlarda konuşma tanıma amaçlı kullanılmış (Raj, 2000) ve patlamalı paket kayıp telafi yöntemi olarak ses öznitelik vektörlerinin geniş anlamda durağan olduğu kabullenmesi yapılarak uygulanmıştır (Milner and James, 2006).

EBSK yöntemine göre, kestirilecek kayıp vektör, kayıp vektörler ve gözlemlenen vektörler arasındaki çapraz ortak değişimliliğe (cross-covariance), gözlemlenen vektörler arasındaki ortak değişimliliğe (covariance) ve kayıp vektörlerin ortalama değerlerine bağlıdır. Geniş anlamda durağan kabullenmesi altında eğitim amaçlı seçilen ses öznitelik vektörlerinin ortalama değerleri, ortak değişimlilik değerleri ve farklı zaman kayması ile elde edilen çapraz ortak değişimlilik değeri evrensel değer olarak kabul edilir ve bütün test verisi üzerinde kullanılır. Bu durumda, zaman içerisinde ses öznitelik vektörlerinin ortalamalarındaki değişimler ve istatistiksel olarak değişimlilik değerlerinin değişmesi paket kayıp telafisine etki etmez.

Ses öznitelik vektörlerinin ortalama değerleri zaman içerisinde değişkenlik göstermektedir. Şekil 6.1’de MFCC (5) için ortalama değer değişimini gösterilmiştir.

Uyarlamalı kestirim yöntemi uygulaması, ses öznitelik vektörlerinin ortalama değerlerinin ve değişimlilik değerlerinin zaman içerisinde değişimini göz önünde bulundurarak paket kayıp telafi sisteminin zaman içerisinde uyarlamalı olarak başarımının artırılmasını amaçlamaktadır.

Paket kayıp telafi sisteminde (Bkz. Şekil 5.1) ardışık aşamalarında her test sözcüğüne (*utterance*) ait her ses öznitelik vektörü için eğitim aşamasında elde edilen evrensel ortalama değer, ortak değişimlilik ve çapraz ortak değişimlilik değerleri güncellenir.



Şekil 6.1 Ses öznitelik vektörlerinin test verisindeki 553 cümle için ortalama değer değişimi (MFCC : 5)

Uyarlamalı kestirim yönteminin avantajı şunlardır:

- Ortak değişinti ve çapraz ortak değişinti değerleri farklı sayıda eğitim cümlesi için değişkenlik göstermektedir. Uyarlamalı yöntemle test aşamasında yeni cümlelerin ortalama değer, ortak ve çapraz ortak değişinti değerleri güncellenecektir.
- Eğitim amaçlı seçilen cümle sayısı çok az tutulabilir. Bu durumda sistem her gözlemlenen yeni cümle ya da konuşmalarla başarımını arttıracaktır ve çok fazla eğitim verisi kullanılarak elde edilen ortalama değer ve ortak değişinti değerlerine yakınsayacaktır.
- Eğitim verisi tek bir konuşmacıya ait cümlelerden seçilebilir. Yeni konuşmacılara ait karakteristikler kısa sürede sistemin parametrelerini (ortalama değer ve ortak değişinti değerleri) güncelleyecektir.

6.1 Ortalama Değer Uyarlaması

Ses öznitelik değerlerine ait ortalama değer uyarlanmasının amacı eğitim aşamasında elde edilen evrensel ortalama değer olarak kabul edilen değere her yeni ses öznitelik vektörünün ortalama değere katkısını bulmaktır.

N adet örnek veriye ait ortalama değer Eşitlik 6.1'deki gibi ifade edilir.

$$\mu = \frac{1}{N} \sum_{i=1}^N x_k(i) \quad , \quad k = 1, \dots, 39 \quad (6.1)$$

Eşitlikte $x_k(i)$, k . ses öznitelik vektörünün i . değerini göstermektedir. Eşitlik 6.1'deki ifadeyi $N-1$ örnek için düşünersek bu durumda $N-1$ 'den sonra N . örneğin sistemin ortalama değere katkısını şu şekilde ifade edebiliriz.

$$\mu_N^* = \frac{N-1}{N} \left(\frac{1}{N-1} \sum_{i=1}^{N-1} x(i) \right) + \frac{1}{N} x(N) \quad (6.2)$$

Eşitlik 6.2 daha genel ifade ile Eşitlik 6.3'teki şekilde gösterilebilir.

$$\mu_N^* = \frac{N-1}{N} \mu_{N-1} + \frac{1}{N} x(N) \quad (6.3)$$

6.2 Ortak Değişinti Uyarlaması

N adet örnek üzerinde τ zaman farkı ile elde edilen çapraz ortak değişinti değeri

$$\sigma_N^\tau = \frac{1}{N-1} \sum_{i=1}^N (x(i) - \mu)(x(i + \tau) - \mu) \quad (6.4)$$

ile ifade edilir. Eşitlik 6.4'te $\tau = 0$ kabul edersek, elde edilen eşitlik N adet örnek için ortak değişinti ifadesini gösterir ve Eşitlik 6.5'teki gibi yazılabilir.

$$\sigma_N^0 = \frac{1}{N-1} \left(\sum_{i=1}^{N-1} (x(i) - \mu)(x(i) - \mu) + (x(N) - \mu^*)(x(N) - \mu^*) \right) \quad (6.5)$$

Eşitlik 6.5'te, μ^* yeni gözlemlenen örneklerin sayı dizisine eklenmesi ile elde edilen ve Eşitlik 6.3'te ifade edilen yeni ortalama değeri göstermektedir. Eşitlik 6.5'te μ^* yerine Eşitlik 6.3'te elde edilen yeni ortalama değer ifadesini yazar ve eşitliğin ilk kısmına $\frac{N-2}{N-2}$ çarpanını eklersek, eşitlik şu şekilde yazılabilir:

$$\sigma_N^0 = \frac{N-2}{N-1} \left(\frac{1}{N-2} \sum_{i=1}^{N-1} (x(i) - \mu)(x(i) - \mu) \right) + \frac{1}{N-1} \left(\frac{N-1}{N} \right)^2 ((x(N) - \mu)(x(N) - \mu)) \quad (6.6)$$

Eşitlik 6.6'da ilk terim ($\tau = 0$), $N-1$ örneğe ait ortak değişintiyi göstermektedir. İkinci terim ise $x(N)$. örneğin $\tau = 0$ için katkısını göstermektedir. Eşitlik sadeleştirme işleminden sonra aşağıdaki şekilde ifade edilir:

$$\sigma_N^0 = \frac{N-2}{N-1} \sigma_{N-1}^0 + \frac{N-1}{N^2} ((x(N) - \mu)(x(N) - \mu)) \quad (6.7)$$

Eşitlik 6.7'de, σ_{N-1}^0 , $N-1$ örneğe ait ortak değişinti değerini göstermektedir.

6.3 Çapraz Ortak Değişinti Uyarlaması

$N-1$ örnek için $\tau > 0$ olduğu durumlarda çapraz ortak değişinti değeri Eşitlik 6.8'den elde edilir.

$$\sigma_N^\tau = \frac{1}{N-\tau-1} \left(\sum_{i=1}^{N-1} (x(i) - \mu)(x(i+\tau) - \mu) \right) \quad (6.8)$$

N . örneğin ilgili τ için katkısını belirlemek için Eşitlik 6.8 aşağıdaki şekilde yazılabilir:

$$\begin{aligned} \sigma_N^\tau = & \frac{1}{N-\tau-1} \left(\frac{N-\tau-2}{N-\tau-2} \sum_{i=1}^{N-1} (x(i) - \mu)(x(i+\tau) - \mu) \right) + \\ & \frac{1}{N-\tau-1} ((x(N-\tau) - \mu^*)(x(N) - \mu^*)) \end{aligned} \quad (6.9)$$

Eşitlikte ilk terimde sadeleştirme yapar ve yeni ortalama değeri yerine koyarsak,

$$\sigma_N^\tau = \frac{N-\tau-2}{N-\tau-1} \sigma_{N-1}^\tau + \left(\frac{N-1}{(N-\tau-1)N} \right) (x(N-\tau) - \mu)(x(N) - \mu) \quad (6.10)$$

eşitliğini elde ederiz. Eşitlikte, σ_{N-1}^τ ifadesi ilgili τ değeri için $N-1$ örneğe ait ortak değışinti değerini göstermektedir. Eşitlik 6.10'daki ifadede ikinci terimdeki çarpanın belirli bir değerden sonra etkisi azalacağı için şu şekilde kabullenme yapılabilir:

$$\left(\frac{N-1}{(N-\tau-1)N} \right) \approx \frac{1}{N}$$

Buna göre Eşitlik 6.10

$$\sigma_N^\tau = \frac{N-\tau-2}{N-\tau-1} \sigma_{N-1}^\tau + \frac{1}{N} (x(N-\tau) - \mu)(x(N) - \mu) \quad (6.11)$$

ile ifade edilir.

BÖLÜM 7

DENEYSEL ÇALIŞMA VE SONUÇLAR

Tezde kullanılan eğitim ve test veritabanı ile sonuçların analizinde kullanılan araçlar, tanımlanan paket kayıp yapıları ve dağıtık yapı modeli ile test sonuçları bu bölümde anlatılmıştır.

7.1 Eğitim ve Test Veritabanı

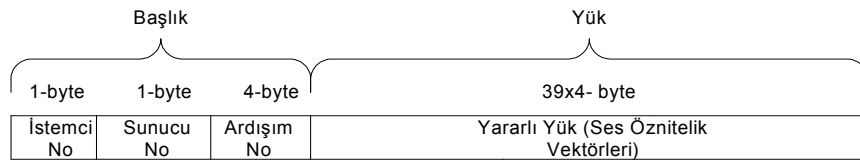
Bu tez çalışmasında, veri tabanı olarak 1,373,249 tane ayrı (özgün) kelime içeren toplam 96,439,401 (96.4M) kelime kullanılmıştır. Veritabanı edebiyat, hukuk, politika, sosyal bilimler, popüler bilim, bilişim teknolojileri, tıp, gazete, magazin ve spor haberleri olmak üzere değişik alanlardaki kelimeleri kapsamaktadır. Akustik modellerin eğitimi için yaklaşık 5000 SMM (HMM) durumlu karar-ağaç durumu ile gruplandırılmış çaprazkelime 3'lü fon modeli kullanılmıştır. Akustik model, harften fonemlere kuralı yerine doğrudan harflere bağlıdır. Her konuşmacıdan bağımsız Saklı Markov Modeli 6 karışım bileşenli Gaussian Karışım Modeli (GKM)'ne sahiptir. Test aşamasında ses öznitelik vektörleri Hidden Markov Toolkit (HTK) kullanılarak hesaplanmıştır (Young, et al., 2005). Belirtik fon süre modelleri uygulanmamıştır. Eğitim verisi 550 farklı konuşmacıya ait 39.44 saatlik konuşmadan oluşmaktadır. Sözlük için metin veritabanında en sık geçen 50,000 kelime seçilmiştir. Akustik karıştırma, sistem karmaşıklığı ve sözlük dışı kelimelerin artması gibi ödünleşimlerden dolayı sözlükteki her kelimenin eklenmesi mümkün değildir. Seçilen en sık geçen 50,000 kelime'de sözlük dışı kelime oranı eğitim verisinde %10,34'tür. 3-gram dil modelleri SRILM aracı kullanılarak üretilmiştir. Test cümleleri bir kadın konuşmacıya ait yaklaşık bir saatlik konuşmadan oluşmaktadır. Test sonuçlarının analizleri için NIST "wgscore" araçları kullanılmıştır (NIST-Score, 2000-2007).

7.2 Test ve Analiz

Test sonuçlarını karşılaştırmak için ilk önce orijinal MFCC vektörleri üzerinde tanıma işlemi gerçekleştirilerek referans değerler elde edilmiştir. Daha sonra, paket kayıpları oluşturularak telafisiz ve telafili değerler elde edilerek referans değerlerle karşılaştırılmıştır. Tanıma sonuçları, NIST araçlarından “wgscore” kullanılarak analiz edilmiştir. Analiz araçları, kayıplar ve telafi işlemleri sonrası oluşan kelime hata oranlarını (WER : word error rate) yüzde olarak listelemektedir. Kelime hata oranları, araya sokulan (inserted), silinen ve değişen kelimelerden (substituted) oluşmaktadır.

7.3 Dağıtık Yapıda Ses Tanıma Sistemi ve Paket İletimi

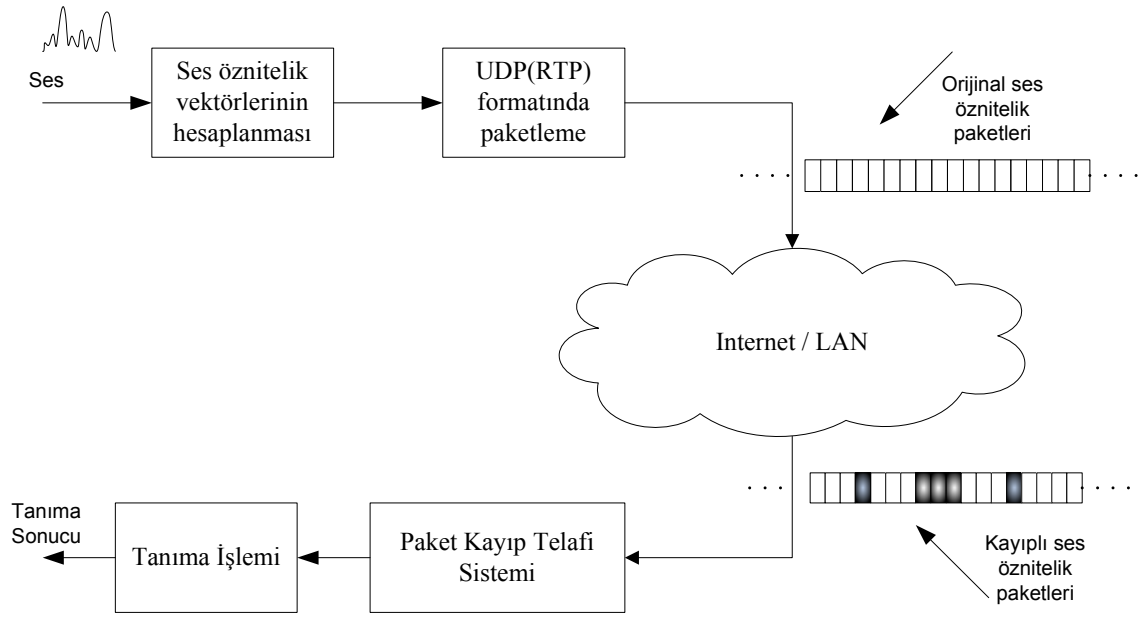
Dağıtık yapıda ses tanıma için tanımlanmış RTP protokolu kullanılabileceği gibi UDP yükü olarak amaca özel bir paket başlığı ve yükü tanımlanabilir. Bu çalışmada, ses öznitelik vektörlerinin iletimi için özel amaçlı UDP paketi tanımlanmıştır.



Şekil 7.1 Ses öznitelik vektörlerinin UDP yükü olarak dağıtık yapıda iletimi için tanımlanmış basit bir UDP paket formatı.

Şekil 7.1’de ses öznitelik vektörlerinin UDP yükü olarak dağıtılmış yapıda iletimi için tanımlanmış UDP paket formatı gösterilmiştir. Şekil 7.2’de istemci ve sunucu numaralarını gösteren ilk iki paket dağıtık yapı modelinde paketin iletileceği istemci ve kaynak sunucuların belirlenmesi için kullanılabilir. Ardışım sırası ise gönderilen ya da gözlemlenen ses öznitelik vektörlerinin sırasını belirlemektedir ve kayıp paket yapısının tespiti için kullanılmaktadır. Gönderilen UDP paketinde tanımlanan başlık toplam 6

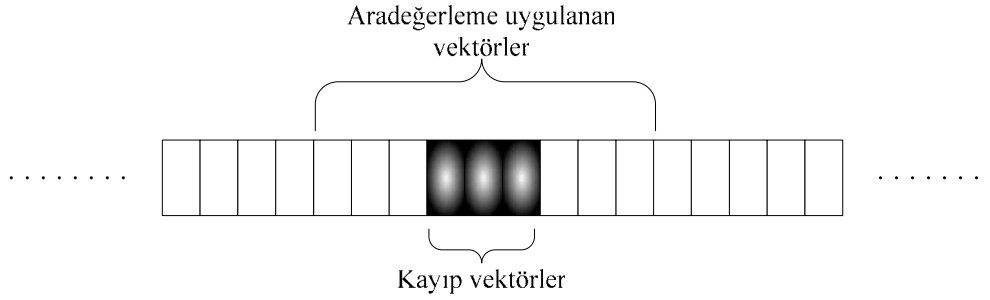
bayttan oluşmaktadır. Ses öznitelik vektörleri ise her biri 4 bayt olan toplam 39 değerden oluşmaktadır (Bkz. Şekil 4.2).



Şekil 7.2 Ses öznitelik vektörlerinin IP üzerinden iletimi ve paket kayıpları

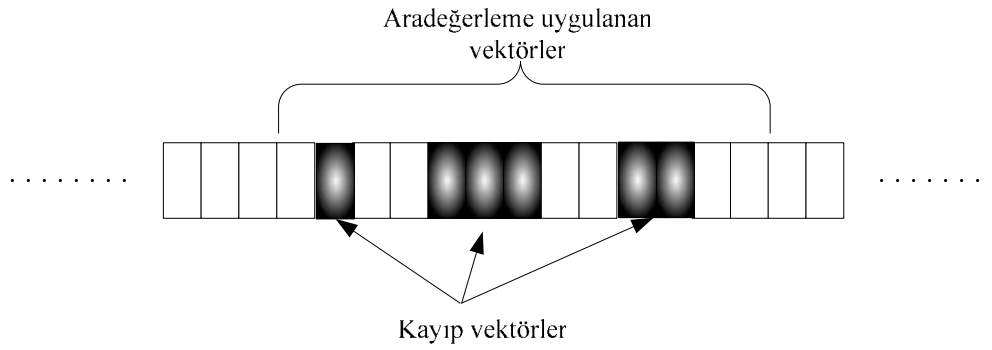
7.4 Paket Kayıp Yapıları

Sürekli konuşma tanıma çalışmaları için hem seyrek hem de patlamalı kayıplar analiz edilirken, yalıtık konuşma tanıma için sadece patlamalı kayıp durumları incelenmiştir. Kayıplar her cümlede eşit aralıklarla oluşturulmuştur. Örneğin, kayıp merkez sayısı 10 olarak uygulanan testlerde, her cümlede cümle uzunluğuna bağlı olarak eşit aralıklarla kayıplar oluşturulmuştur. Bu şekilde, paket kayıplarının bizim isteğimiz dışında bir noktada birikmesi önlenmiştir. Seyrek kayıplarda en yüksek kayıp aralığı 3 olarak tanımlanmıştır ve kayıplar 1, 2 ve 3 olacak şekilde testler yapılmıştır. Şekil 7.3'te kayıp vektör bölgesi ve aradeğerleme uygulanacak vektörler gösterilmiştir.



Şekil 7.3 Kayıp MFCC vektörleri üzerinde aradeğerleme uygulaması gösterimi

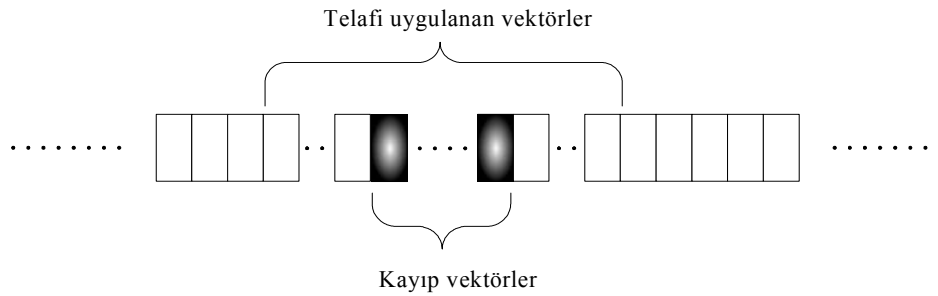
Şekilde her çerçeve 1×39 'luk MFCC vektörlerini ifade etmektedir ve siyah çerçeveler kayıp vektörleri göstermektedir. Şekil 7.4'te aradeğerleme yapılacak seyrek paket kayıplarının tek bir merkezde değil daha geniş alanda uygulandığı durum gösterilmiştir.



Şekil 7.4 Seyrek paket kayıplarının birden fazla merkezli olduğu durumlarda aradeğerleme yöntemlerinin uygulanması

Kayıp bölgesinde ardışıl paket kaybı 3 ya da daha az ise belirlenen uzunluğa göre aradeğerleme yöntemi uygulanır. P kayıp vektör için $2xP$ paket daha kullanılmıştır ve toplam aradeğerleme uygulanacak katsayı $3xP$ kadardır. Gerçek ortamda kayıplar birden fazla oluşabilir ancak testlerimizde rasgele paket kayıpları olarak t_1 anında tek bir paket kayıp olacak şekilde düşünülmüştür (Bkz. Şekil 7.3).

Patlamalı paket kayıp durumlarında, kayıp vektörlerin öncesinden ve sonrasında 5'er vektör alınır (Milner and James, 2006) ve telafi uygulanan toplam vektör sayısı $P+10$ ile ifade edilir. Patlamalı paket kayıp telafisi için kullanılan paketler Şekil 7.5'te gösterilmiştir. Kestirim yöntemleri için patlamalı kayıp durumlarında kayıp merkezinin öncesinden ve sonrasında 5 yerine 3 vektör alınarak da testler yapılmıştır.



Şekil 7.5 Patlamalı paket kayıp durumlarında kayıp vektörler ve telafi uygulanan vektörler

7.5 Yalıtık Konuşma Tanıma ve Paket Kayıplarının Telafisi

Tez çalışmasında, paket kayıplı durumda yalıtık konuşma tanıma için telafi yöntemi olarak yerine koyma yöntemi uygulanmıştır. Testler, Türkçe yalıtık konuşma veri tabanı kullanılarak (Barkana vd., 1995) yapılmıştır. Tanıma sonuçları Çizelge 7.1'de gösterilmiştir.

Yalıtık konuşma durumunda, paket kayıpları kelimenin başında, sonunda ve ortasında olacak şekilde üç farklı durumda incelenmiştir. Testler, patlamalı paket kayıplı durumlar için elde edilmiştir. Seyrek paket kayıplarının etkisi yok denecek kadar az olmuştur (Uzun ve Edizkan, 2007). Kayıplar özellikle kelime başlarında olduğunda etkilenme daha fazla olmuştur. Yalıtık kelimelerdeki ses öznitelik vektörü sayısı en az 25 en fazla 88 adettir.

Çizelge 7.1 Paket kayıplarında yerine koyma metodu ile yalıtık konuşma tanıma

Kayıp Paket Sayısı	Kayıp Yeri (Kelimenin)	Ortalama Tanıma	
		Kayıp Telafisiz	Kayıp Telafili
0	-	99.52	
5	Başında	99.29	99.29
	Ortasında	99.35	99.29
	Sonunda	99.35	99.52
10	Başında	98.99	99.11
	Ortasında	99.17	99.46
	Sonunda	99.11	99.52
15	Başında	97.92	99.11
	Ortasında	98.75	99.11
	Sonunda	98.75	99.46
20	Başında	93.45	99.05
	Ortasında	96.96	98.87
	Sonunda	98.33	99.46

7.6 Sürekli Konuşma Tanımda Aradeğerleme Yöntemleri ile Kayıp Telafisi

Gerçek zamanlı sistemler için sistemin ve yöntemlerin başarımı kadar harcanan işlem gücü ve kullanılan bellek miktarı çok önemlidir. Bu çalışmada, paket kayıp telafi sistemi ve dinamik bellek sayesinde paket kayıp durumuna göre başarımı ve işlemci gücü açısından en uygun yöntem seçilmiştir. Seyrek paket kayıpları için aradeğerleme yöntemleri, patlamalı paket kayıpları için kestirim yöntemleri uygulanmıştır.

Aradeğerleme yöntemleri olarak Lagrange ve Cubic Spline yöntemleri kullanılmış ve Türkçe sürekli konuşma tanıma için başarımları karşılaştırılmıştır. Sonuçları karşılaştırmak için elde edilen orijinal ses öznelik vektörleri (MFCC) ile elde edilmiş tanıma oranları ve kayıp sonrası telafisiz tanıma oranları Çizelge 7.2'de gösterilmiştir. Aradeğerleme yöntemlerinin karşılaştırılmasında 39 adet MFCC üzerinde kayıplar oluşturulmuş ve telafi uygulanmıştır. Kayıp yapıları sütununda ilk değer cümledeki kayıp merkez sayısını, ikinci değer her kayıp merkezinde kaybolan ardışık paket sayısını vermektedir. Kayıpların seyrek ve cümle üzerinde kayıp merkez sayısının az

olduğu durumlarda tanıma başarımı çok etkilenmemektedir. Ancak, kayıp merkez oranı arttıkça özellikle ardışıl kayıplar üç olduğu durumlarda tanıma başarımı etkilenmeye başlamaktadır (Uzun vd., 2007).

Çizelge 7.2 Kayıplı (telafisiz) kelime hata oranları

Kayıplı(Telafisiz) Sonuçlar(%)					
Kayıp Yapısı	Doğru Cümle	Değişen kelime (Substituted)	Silinen kelime (Deleted)	Eklenen kelime (Inserted)	Kelime Hata Oranı (KHO) (WER)
-	71.9	24.6	3.5	4.0	32.1
5-3	71.4	24.8	3.9	3.7	32.3
10-1	71.6	24.7	3.7	3.7	32.2
10-2	71.4	24.8	3.9	3.7	32.3
10-3	70.5	25.4	4.2	3.5	33.0
20-1	71.4	24.8	3.8	3.8	32.4
20-2	70.6	25.2	4.2	3.5	32.9
20-3	69.0	26.1	4.9	3.2	34.2

7.6.1 Lagrange Telafi Yöntemi

Lagrange telafi yönteminin başarımı Çizelge 7.3'te gösterilmektedir. Kayıp merkezleri sözcük uzunluğuna orantılı olarak eşit mesafelerde seçilmiştir. Kayıpların olduğu bölgede ardışıl kayıp paket sayısı seyrek paket kayıp yapısında seçilmiştir ve 1, 2 ve 3 paket kayıplı durumlar incelenmiştir. Lagrange yöntemi ile telafi edildiğinde seyrek kayıpların etkisi büyük oranda giderilmektedir. Ancak, kayıp merkezi arttıkça ve ardışıl kayıplar 3 olunca tanıma başarımı etkilenmeye başlamaktadır (Uzun vd., 2007).

Çizelge 7.3 Lagrange ile paket kayıp telafi yönteminde kelime hata oranları

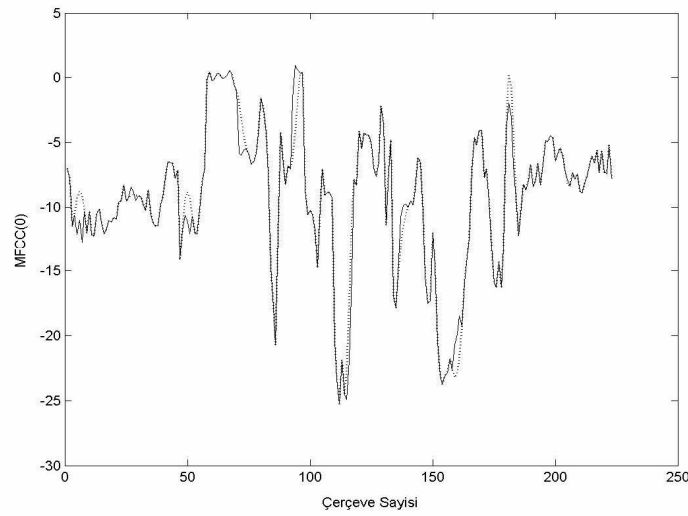
Lagrange Telafili Sonuçlar(%)					
Kayıp Yapısı	Doğru Cümle	Değişen Kelime	Silinen Kelime	Eklenen Kelime	KHO
-	71.9	24.6	3.5	4.0	32.1
5-3	71.7	24.7	3.6	4.1	32.3
10-1	71.9	24.6	3.5	4.0	32.1
10-2	71.9	24.6	3.5	4.0	32.1
10-3	71.8	24.7	3.5	4.0	32.2
20-1	71.9	24.6	3.5	3.9	32.0
20-2	71.9	24.5	3.5	3.9	32.0
20-3	71.4	24.9	3.7	3.9	32.6

7.6.2 Spline Telafi Yöntemi

Spline aradeğerleme yöntemi, MFCC vektörlerinin paketler halinde iletiildiği dağıtık konuşma tanıma sisteminde paket kayıplarının telafisi uygulamalarında seyrek paket kayıplarının telafisi için kullanılmıştır (Milner and James, 2006). Ses öznelik vektörleri toplam olarak 39 adettir ve aradeğerleme her vektör elemanı için uygulanmaktadır. Genel olarak Spline eşitliğini

$$h^k S^k = Y^k$$

şeklinde ifade edebiliriz. Eşitlikte k değeri MFCC vektöründeki katsayı sayısını (39) sayısını göstermektedir. Şekil 7.5'te paket kayıpları için spline aradeğerleme yöntemi uygulayarak telafi edildiği durumlarda geri elde edilen ses öznelik vektörlerini göstermektedir.



Şekil 7.6 Spline aradeğerleme yöntemi ile aradeğerleme uygulanmış MFCC (0) değerleri. Kayıp merkez oranı : 10, merkezdeki ardışık kayıp paket sayısı : 3

Spline telafi yönteminin başarımı Çizelge 7.4'te gösterilmektedir. Lagrange yönteminin uygulandığı durumla aynı şekilde kayıp merkezleri sözcük uzunluğuna orantılı olarak eşit mesafelerde seçilmiştir (Uzun vd., 2007). Kayıpların oluştuğu bölgede ardışık kayıp paket sayısı seyrek paket kayıp yapısında seçilmiştir ve 1, 2 ve 3 paket kayıplı durumlar incelenmiştir.

Çizelge 7.4 Spline ile paket kayıp telafi yönteminde kelime hata oranları

Spline Telafili Sonuçlar(%)					
Kayıp Yapısı	Doğru Cümle	Değişen kelime	Silinen kelime	Eklenen kelime	KHO
-	71.9	24.6	3.5	4.0	32.1
5-3	71.8	24.7	3.5	4.0	32.3
10-1	71.9	24.6	3.5	4.0	32.1
10-2	71.9	24.6	3.5	4.0	32.1
10-3	71.7	24.8	3.5	4.0	32.2
20-1	71.9	24.6	3.5	3.9	32.0
20-2	71.9	24.5	3.5	3.9	32.0
20-3	71.6	24.8	3.6	4.0	32.4

Spline yöntemi ardışıl kayıplar 3 olduğu durumda Lagrange yöntemine göre 0.2 puan iyileşme göstermektedir. Telafisiz başarımlar ile yerine koyma ve aradeğerleme yöntemleri ile telafi edilmiş paket kayıplarının başarımları Çizelge 7.5'te karşılaştırılmıştır. Çizelgeye göre seyrek paket kayıplarında ardışık kayıp sayısı 3 ve üzerinde ise Spline telafi yöntemini uygulamak daha doğru olacaktır. 1 ve 2 kayıplı durumda gerçek zamanlı sistemlerde telafi uygulaması başarımlar ve işlem gücü açısından gerekli görülmeyebilir. Yerine koyma yöntemi işlem gücü açısından en uygun olan yöntem olmasına karşılık tanıma başarımları açısından en kolay etkilenen yöntemdir. Seyrek kayıplarda Spline yöntemi yerine koyma yöntemine göre 0.7 puan daha iyi sonuçlar vermektedir.

Çizelge 7.5 Telafisiz ve aradeğerleme yöntemlerinin başarımlarının karşılaştırılması

Kayıp Yapısı	Telafi Yöntemi / Kelime Hata Oranları (%)			
	Telafisiz	Yerine Koyma	Lagrange	Spline
-	32.1	-	-	-
10-1	32.2	32.3	32.1	32.1
10-2	32.3	32.3	32.1	32.1
10-3	33.0	32.8	32.2	32.2
20-1	32.4	32.1	32.0	32.0
20-2	32.9	32.3	32.0	32.0
20-3	34.2	33.1	32.6	32.4

Patlamalı paket kayıp durumlarında aradeğerleme yöntemleri kayıp merkezinden önce ve sonra gelen 5 vektör ve kayıp vektörler üzerine uygulanmıştır. Böylece patlamalı paket kayıpları durumlarında aradeğerleme uygulanacak verinin çok fazla olmasının önüne geçilmiştir.

Patlamalı paket kayıpları durumunda aradeğerleme yöntemlerinde paket kayıpları yapıları ve başarımların değerleri Çizelge 7.6 ve 7.7’de gösterilmiştir. Patlamalı paket kayıpları durumunda Lagrange aradeğerleme yönteminin başarımlarını patlama merkezinin 10 ve daha fazla olduğu durumlarda çok hızlı bir şekilde düşmektedir. Lagrange aradeğerleme yöntemi için kelime hata oranının patlama merkez büyüklüğüne bağlı olarak çok fazla artışının bir sebebi de, telafi sırasında kayıp merkezinde oluşabilecek aşırı artış ya da düşüşler olabilir. Bu durumlarda konuşma tanıma sistemi cümlelerin tanınmasını tamamlayamıyor ve hata oranı çok artıyor. Şekil 7.7’de Lagrange aradeğerleme sonucu oluşabilen tepe oluşumu gösterilmektedir.

Çizelge 7.6 Patlamalı paket kayıpları durumunda Spline aradeğerleme yönteminin başarımlarını

Spline Telafili Sonuçlar(%)					
Kayıp Yapısı	Doğru Cümle	Değişen kelime	Silinen kelime	Eklenen kelime	KHO
-	71.9	24.6	3.5	4.0	32.1
10-5	71.1	25.0	3.9	3.9	32.8
10-10	56.9	34.9	8.2	3.2	46.3

Lagrange yöntemine karşılık Spline aradeğerleme yönteminin başarımlarını daha yavaş düşmektedir. Patlama merkezinin 5 olduğu durumlarda Spline aradeğerleme yönteminin başarımlarını Lagrange aradeğerleme yöntemine göre belirgin bir şekilde daha yüksektir. Ancak, genel olarak aradeğerleme yöntemlerinin başarımlarını patlama merkezi büyüdükçe düşmektedir ve paket kayıp telafileri için yeni yöntemler gerekmektedir.

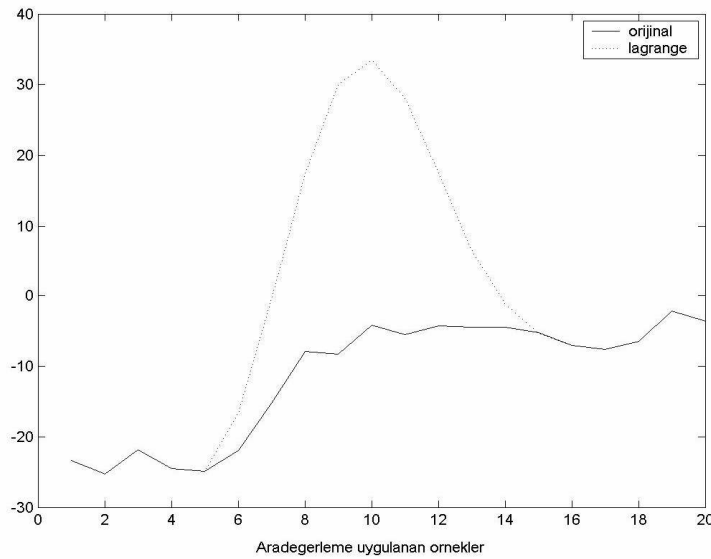
Çizelge 7.7 Patlamalı paket kayıpları durumunda Lagrange aradeğerleme yönteminin başarımlarını

Lagrange Telafili Sonuçlar(%)					
Kayıp Yapısı	Doğru Cümle	Değişen kelime	Silinen kelime	Eklenen kelime	KHO
-	71.9	24.6	3.5	4.0	32.1
10-5	66.9	28.2	4.9	4.0	37.1
10-10	17.6	69.7	16.2	10.6	93.0

Çizelge 7.8’de patlama büyüklüğü ve kayıp merkezi oranlarına göre aradeğerleme yöntemlerinin sinyal–gürültü oranları listelenmiştir. Görüldüğü gibi kayıp merkezlerindeki patlama değerleri büyüdükçe Lagrange aradeğerleme yönteminin SNR değeri çok hızlı bir şekilde düşmektedir.

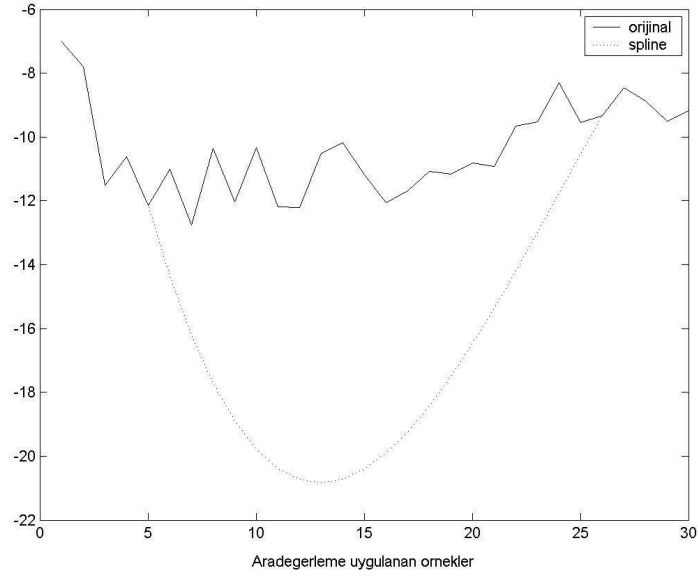
Çizelge 7.8 Patlamalı paket kayıpları durumunda aradeğerleme yöntemlerinin SNR değerleri

Patlama Büyüklüğü	Kayıp Merkezi	Kayıp (%)	Aradeğerleme Yöntemi	SNR (dB)
10	10	14.3	Spline	13.5295
			Lagrange	-1.3245
20	5	14.3	Spline	8.8685
			Lagrange	-17.5936

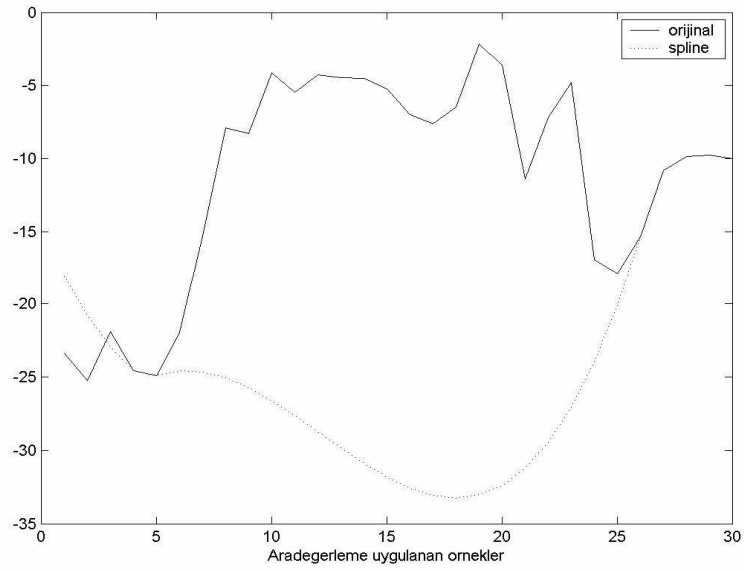


Şekil 7.7 Patlamalı kayıp durumlarında Lagrange telafisi ve hata oluşumu (patlama büyüklüğü : 10, MFCC : 0)

Aradeğerleme yöntemleri kayıplardan önce ve sonra alınan değerlere bağlı olarak yüksek başarımlar gösterebilecekleri gibi zaman zaman olumsuz sonuçlara sebep olabilirler. Şekil 7.8’de patlama merkezinin 20 olduğu durumlarda MFCC (0) üzerinde oluşan iki farklı durum için Spline telafisi gösterilmektedir. Şekillerde patlama merkezinin büyüdüğü durumlarda aradeğerleme yönteminin başarımlarının düştüğü açıktır.



a)



b)

Şekil 7.8 Patlamalı kayıp durumlarında Spline telafisinin etkili olmadığı durumlar (patlama büyüklüğü : 20), a) patlama merkezinde ortalama değerin fazla değişmediği, b) patlama merkezinde ortalama değerin değiştiği durum

7.7 Kesitirim Yöntemi ile Kayıp Telafisi

Ses öznelik vektörlerinin k . quefrensy değeri için EBSK yöntemi ile elde edilmiş kayıp vektörlerini Eşitlik 4.24 kullanarak Eşitlik 7.1'deki gibi yazabiliriz (Milner and James, 2000). Bu durumda, ses öznelik vektörüne ait her quefrensy değeri için bu işlem gerçekleştirilir.

$$\hat{\mathbf{x}}_m^k = \boldsymbol{\mu}_m^k + \boldsymbol{\Sigma}_{m_o}^k (\boldsymbol{\Sigma}_{o_o}^k)^{-1} (\mathbf{x}_o^k - \boldsymbol{\mu}_o^k) \quad (7.1)$$

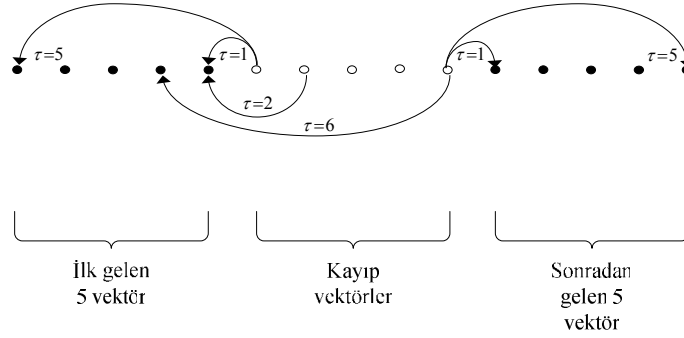
Eşitlikte $\boldsymbol{\mu}$, gözlemlenen ve kayıp vektörlere ait ortalama değer, $\boldsymbol{\Sigma}_{m_o}^k$ k . quefrensy değeri için gözlemlenen ve kayıp vektör elemanları arasındaki çapraz ortak değişinti matrisi, $\boldsymbol{\Sigma}_{o_o}^k$ ise gözlemlenen vektör elemanları arasındaki ortak değişinti matrisidir. Eşitlik 7.1'de $\hat{\mathbf{x}}_m^k$ k . quefrensy için kayıp merkezindeki kayıp sayısı kadar elemanı temsil eden vektördür.

Çapraz ortak değişinti ve ortak değişinti matrisleri Eşitlik 4.27 kullanılarak elde edilen evrensel matrisin içinden τ değerine göre seçilir. Evrensel matris,

$$\boldsymbol{\Sigma} = \begin{bmatrix} \sigma^{0,0} & \sigma^{0,1} & \sigma^{0,2} & \dots & \sigma^{0,\tau} \\ \sigma^{1,0} & \sigma^{1,1} & \sigma^{1,2} & \dots & \sigma^{1,\tau} \\ \cdot & \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot & \cdot \\ \sigma^{k,0} & \dots & \dots & \dots & \sigma^{k,\tau} \end{bmatrix}_{K \times \tau} \quad (7.2)$$

şeklinde ifade edilir ve boyutu $k \times \tau$ 'dur. Eşitlikte k değeri ses öznelik vektörlerinin boyutunu göstermektedir. Çapraz ortak değişinti matrisinin elde edilmesi Şekil 7.9'daki gibi gösterilebilir. Örneğin, 5 adet kayıp vektör için önceden ve sonradan gözlemlenen vektörün ortak değişinti matrisi hesaplanmak istendiğinde birinci kayıp değer için çapraz ortak değişinti değerleri ilk gözlemlenen (en baştaki $\tau=5$) değer ile aralarındaki

τ değerinden başlayarak beşinci ($\tau = 1$) değere kadar sıralanır, daha sonra kayıplardan sonraki gözlemlenen vektörlerle (sağ taraftakilerle) devam edilir.



Şekil 7.9 Kayıp vektörlerin gözlemlenen vektörlere göre çapraz ortak değişinti gösterimi.

Sonuçta, k . quefrençy için ilk kayıp değerini önce ve sonra gelen (5+5) gözlemlenen değere göre çapraz ortak değişinti matrisi şu şekildedir:

$$\Sigma_{mo}^k = \begin{bmatrix} \sigma^{1,5} & \sigma^{1,4} & \sigma^{1,3} & \sigma^{1,2} & \sigma^{1,1} & \sigma^{1,5} & \sigma^{1,6} & \sigma^{1,7} & \sigma^{1,8} & \sigma^{1,9} \\ \sigma^{2,6} & \sigma^{2,5} & \sigma^{2,4} & \sigma^{2,3} & \sigma^{2,2} & \sigma^{2,4} & \sigma^{2,5} & \sigma^{2,6} & \sigma^{2,7} & \sigma^{2,8} \\ \sigma^{3,7} & \sigma^{3,6} & \sigma^{3,5} & \sigma^{3,4} & \sigma^{3,3} & \sigma^{3,3} & \sigma^{3,4} & \sigma^{3,5} & \sigma^{3,6} & \sigma^{3,7} \\ \sigma^{4,8} & \sigma^{4,7} & \sigma^{4,6} & \sigma^{4,5} & \sigma^{4,4} & \sigma^{4,2} & \sigma^{4,3} & \sigma^{4,4} & \sigma^{4,5} & \sigma^{4,6} \\ \sigma^{5,9} & \sigma^{5,8} & \sigma^{5,7} & \sigma^{5,6} & \sigma^{5,5} & \sigma^{5,1} & \sigma^{5,2} & \sigma^{5,3} & \sigma^{5,4} & \sigma^{5,5} \end{bmatrix}$$

Eşitlik 7.2'deki evrensel matristen elde edilen ortak değişinti matrisi ve tersi şu şekilde gösterilir:

$$\Sigma_{oo}^k = \begin{bmatrix} \sigma^{k,0} & 0 & \dots & 0 \\ 0 & \sigma^{k,0} & 0 & \dots & 0 \\ \cdot & \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot & \cdot \\ 0 & \dots & \dots & \dots & \sigma^{k,0} \end{bmatrix}_{(K+M) \times (K+M)}, \quad (\Sigma_{oo}^k)^{-1} = \begin{bmatrix} \frac{1}{\sigma^{k,0}} & 0 & \dots & 0 \\ 0 & \frac{1}{\sigma^{k,0}} & 0 & \dots & 0 \\ \cdot & \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot & \cdot \\ 0 & \dots & \dots & \dots & \frac{1}{\sigma^{k,0}} \end{bmatrix}_{(K+M) \times (K+M)}$$

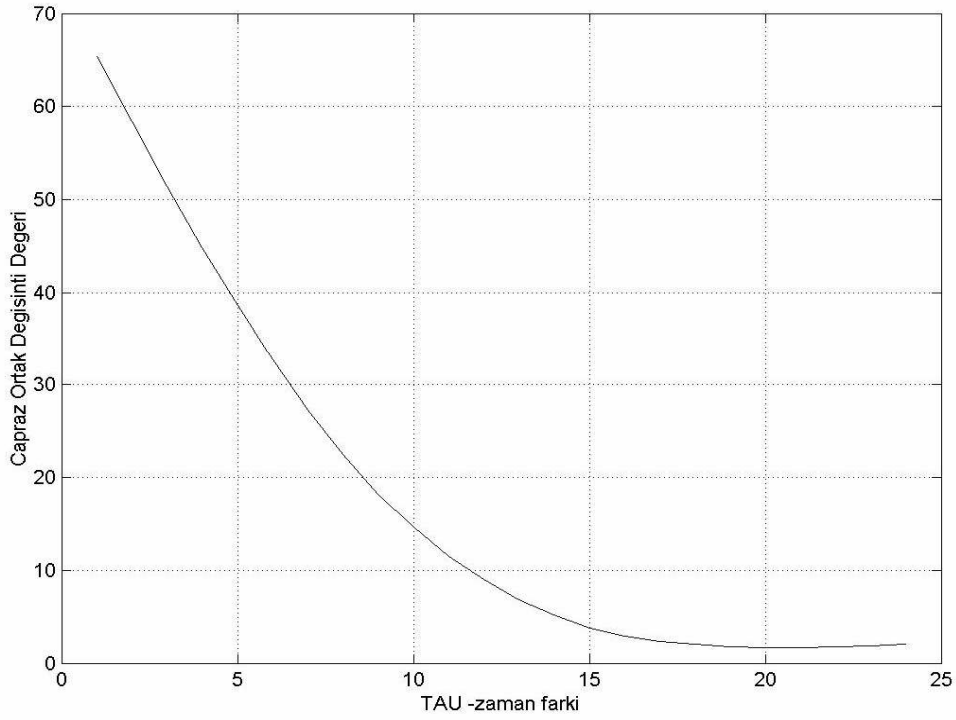
Ortak deęişinti matrisi $(K+M) \times (K+M)$ boyutludur. K kayıptan önceki ve M ise kayıptan sonraki gözlenen vektör sayısını göstermektedir. Bu durumda, kayıp vektörler

$$\hat{\mathbf{x}}_m^k = [\hat{x}_{m1}^k \ \hat{x}_{m2}^k \ \hat{x}_{m3}^k \ \hat{x}_{m4}^k \ \hat{x}_{m5}^k]$$

şeklinde kabul edilirse, Eşitlik 7.1'deki ifadeden kayıp vektörlere ait her eleman gözlemlenen vektörler ve ortak deęişinti ilişkisi kullanarak şu şekilde ifade edilebilir:

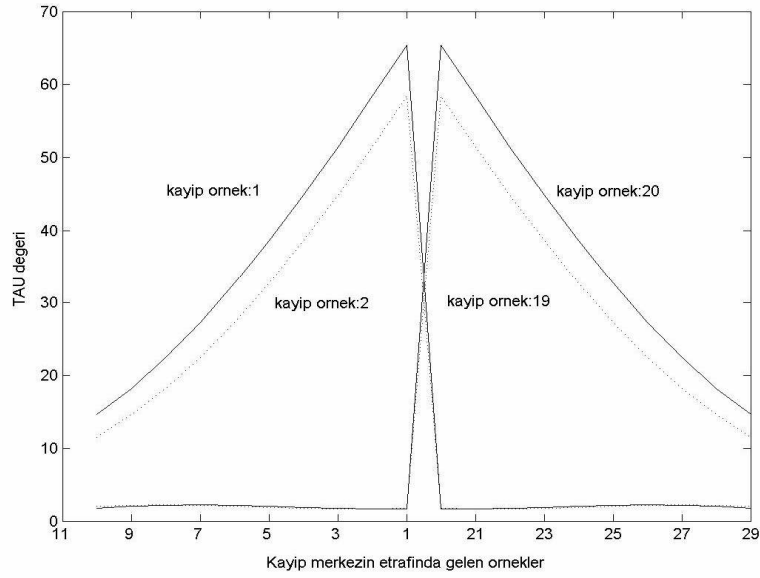
$$\begin{aligned} \hat{x}_{m1}^k &= \mu_{m1} + \frac{\sigma^{k,5}}{\sigma^{k,0}}(x_{o1} - \mu_{o1}) + \frac{\sigma^{k,4}}{\sigma^{k,0}}(x_{o2} - \mu_{o2}) + \dots + \frac{\sigma^{k,9}}{\sigma^{k,0}}(x_{o15} - \mu_{o15}) \\ \hat{x}_{m2}^k &= \mu_{m1} + \frac{\sigma^{k,6}}{\sigma^{k,0}}(x_{o1} - \mu_{o1}) + \frac{\sigma^{k,5}}{\sigma^{k,0}}(x_{o2} - \mu_{o2}) + \dots + \frac{\sigma^{k,8}}{\sigma^{k,0}}(x_{o15} - \mu_{o15}) \\ &\quad \cdot \\ &\quad \cdot \\ \hat{x}_{m5}^k &= \mu_{m5} + \frac{\sigma^{k,9}}{\sigma^{k,0}}(x_{o1} - \mu_{o1}) + \frac{\sigma^{k,8}}{\sigma^{k,0}}(x_{o2} - \mu_{o2}) + \dots + \frac{\sigma^{k,5}}{\sigma^{k,0}}(x_{o15} - \mu_{o15}) \end{aligned} \quad (7.3)$$

Eşitlik 7.3'te k deęeri MFCC vektörüne ait quefrensy indisini göstermektedir. Eşitlikte gözlemlenen ve kayıp vektörlerin ortalama deęerleri evrensel ortalama deęer olarak seçilebilir. Eşitlikten anlaşılacağı üzere, kayıp vektörün ilgili indisi gözlemlenen vektörün aynı indisindeki deęerlere τ derecesinde baęlıdır. Zaman ekseninde yakın olan (τ 'nin küçük deęerleri) gözlemlenen vektör elemanına baęlanım katsayısı daha yüksek, buna karşılık zaman ekseninde uzak olan elemanlarla olan baęlanım katsayısı daha düşük olmaktadır. Testlerde kullanılan ve haber programları için hazırlanmış olan tek bir konuşmacıya ait olan veri üzerinde τ zaman farkı ile elde edilmiş çapraz ortak deęişinti deęerleri ($\tau > 0$) Şekil 7.10'daki gibidir.

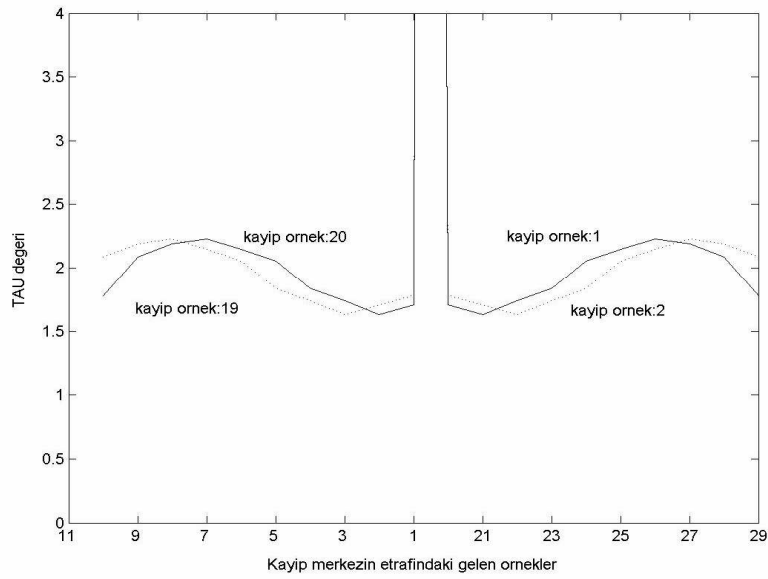


Şekil 7.10 Zaman farkına (τ) bağlı olarak MFCC (0) çapraz ortak değışinti değeri nin değışimi

Kayıp merkezindeki örneklerin kayıp etrafındaki örneklerle olan ilişkileri Şekil 7.11’de gösterilmiştir. Şekilde görüldüğü gibi uzak mesafedeki örneklerle olan çapraz değışinti yakın olanlara oranla çok daha azdır. Şekil 7.11 (b)’deki gösterim uzak mesafedeki örneklerle olan çapraz ortak değışinti değeri ni detaylı olarak göstermektedir.



a)



b)

Şekil 7.11 Kayıp merkezindeki örneklerin kayıp etrafındaki örneklerle olan ilişkileri, a) bütün ortak çapraz değişinti değerleri, b) uzak mesafedeki örneklerle olan çapraz ortak değişinti değerleri (yaklaştırılmış). Kayıp sayısı : 20, kayıp merkezinin sağından ve solundan alınan örnek sayısı : 10.

Kestirim yöntemlerinde kayıp merkezi büyük olsa da gözlemlenen veriler ve istatistik bilgisi kullanıldığı için aradeğerleme yöntemleri sonrasında oluşan tepeler

gözlenmemektedir. Farklı patlama büyüklükleri ve kayıp merkez oranları için EBSK yönteminin SNR değerleri Çizelge 7.9’da gösterilmiştir. Çizelgede eğitim ve test verisi için ayrı ayrı SNR değerleri gösterilmektedir. Test verisine ait SNR değerleri eğitim verisine oranla biraz daha düşük çıkmaktadır. Eğitim ve test amaçlı kullanılan cümle oranları Çizelge 7.10’da gösterilmiştir. Çizelge 7.9’da görüldüğü gibi patlama merkezinin büyüklüğüne bağlı olarak EBSK yönteminde SNR değerleri çok fazla değişmemektedir. Buna karşılık, Çizelge 7.8’de verilen aradeğerleme yöntemlerine ait SNR değerleri patlama merkezine bağlı olarak hızlı şekilde değişmektedir. Spline aradeğerleme yönteminin SNR değerinin kayıp merkez büyüklüğünün 10 olduğu durumda EBSK yönteminden biraz daha iyi çıkmasının sebebi aradeğerleme yöntemlerinde kayıp merkezinin iki ucunda daha yavaş bir geçiş olduğundan kaynaklanabilir.

Çizelge 7.9 Patlamalı paket kayıpları durumunda EBSK yönteminin SNR değerleri

Patlama Büyüklüğü	Kayıp Merkezi	Eğitim Cümlesi	Test Cümlesi	SNR (dB)	
				<i>Eğitim</i>	<i>Test</i>
20	5	400	153	10.5445	10.2037
10	10			11.1917	10.7670

Çizelge 7.10 Eğitim ve test amaçlı kullanılan cümlelerdeki toplam ses öznitelik vektör sayısı

Cümle Sayısı	Toplam Ses Öznitelik Vektör Sayısı
25	19440
400	281588
553	384794

Kayıp merkez büyüklüğüne bağlı olarak Lagrange aradeğerleme yönteminin başarımı hızlı bir şekilde düşmektedir. Bunun için patlamalı kayıp durumlarında aradeğerleme yöntemi olarak sadece Spline aradeğerleme yöntemi incelenmiş ve EBSK yöntemi ile karşılaştırılmıştır. Çizelge 7.11’de patlamalı kayıp durumlarında Spline ve EBSK yöntemlerinin karşılaştırılması verilmiştir. Spline ve EBSK yöntemi karşılaştırılırken ilk 13 MFCC üzerinde kayıplar oluşturulmuş ve telafi uygulanmıştır. Bundan dolayı, Çizelge 7.6’daki Spline aradeğerleme yönteminin aynı yapıdaki kayıplar için sonuçları farklı çıkmaktadır. Patlama merkez büyüklüğü 20 ve üzeri durumlarda Spline yönteminin başarımı hızlı bir şekilde düşmektedir. Buna karşılık EBSK yönteminde düşüş daha yavaştır.

Çizelge 7.11 Patlamalı durumlarda kestirim ve aradeğerleme yöntemlerinin başarımları

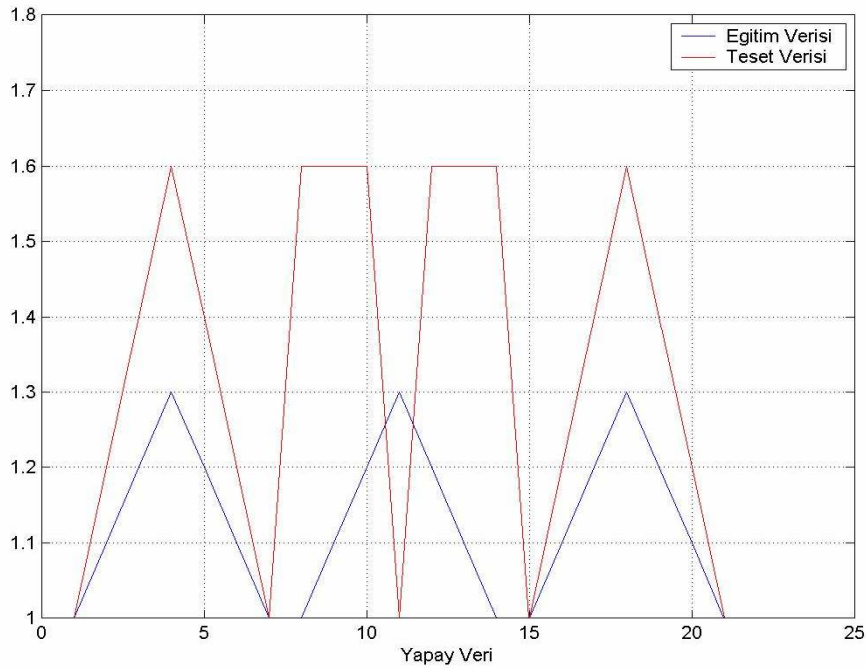
Patlamalı Durumlarda Telafi Yöntemlerinin Başarımı (%)						
Kayıp Yapısı	Telafi Yöntemi	Doğru Cümle	Değişen Kelime (Substituted)	Silinen Kelime (Deleted)	Eklene Kelime (Inserted)	Kelime Hata Oranı (WER)
	-	71,9	24,6	3,5	4,0	32,1
10-10	Spline	69,3	26,2	4,5	3,7	34,4
	EBSK	71,1	24,7	4,2	3,9	32,8
5-20	Spline	54,6	37,4	8,1	4,6	50,1
	EBSK	67,7	27,0	5,3	3,6	35,9

7.8 Uyarlamalı Kestirim Yöntemi ile Kayıp Telafisi

Uyarlamalı kestirim yöntemi ilk önce yapay veri üzerinde test edilmiştir. Yapay veri olarak eğitim aşamasında en az 1.0 ve en fazla 1.3 değerini alan üç adet üçgen oluşturacak sayı dizisi kullanılmıştır. Test aşamasında genlik olarak (ortalama değer etkisi) ve şekil olarak (ortak değişinti etkisi) farklı sayı dizisi uygulanmıştır. Şeklin üçgen olarak seçilmesinin sebebi, tekrar eden üçgen sayı dizisinin aradeğerleme

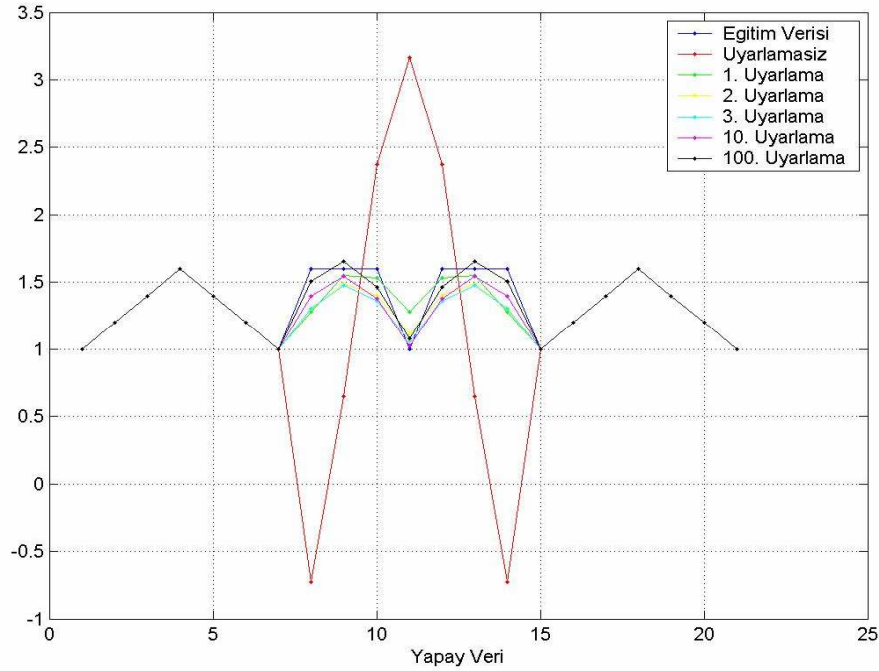
yöntemine göre kestirim yönteminin farkını daha açık ortaya koymasdır. Şekil 7.12’de eğitim ve test aşamasında kullanılan sayı dizisi gösterilmektedir.

Paket kayıpları için ortadaki parçanın kayıp olduğu varsayımı ile sağdan ve soldan gözlemlenen değerlere göre ortalama değer ve ortak değişinti değerlerinin değişimi incelenmiştir. Şekil 7.13’te uyarlamannın yapılmadığı ve kestirimin doğrudan eğitim aşamasında elde edilen ortalama değer ve ortak değişinti değerleri ile elde edilmiş telafisi *kırmızı* ile gösterilmektedir. Uyarlama farkını gözlemek için farklı sayıda uyarlamalar yapılmıştır. Farklı sayıda uyarlama, gelen her örnek sayısı değil aynı sayı dizisinin (kenarda iki üçgen ve ortadaki kayıp benzetimi yapılan şekil) farklı sayıda gelmesini ifade etmektedir.



Şekil 7.12 Uyarlamalı kestirim yöntemi testleri için kullanılan yapay veri.

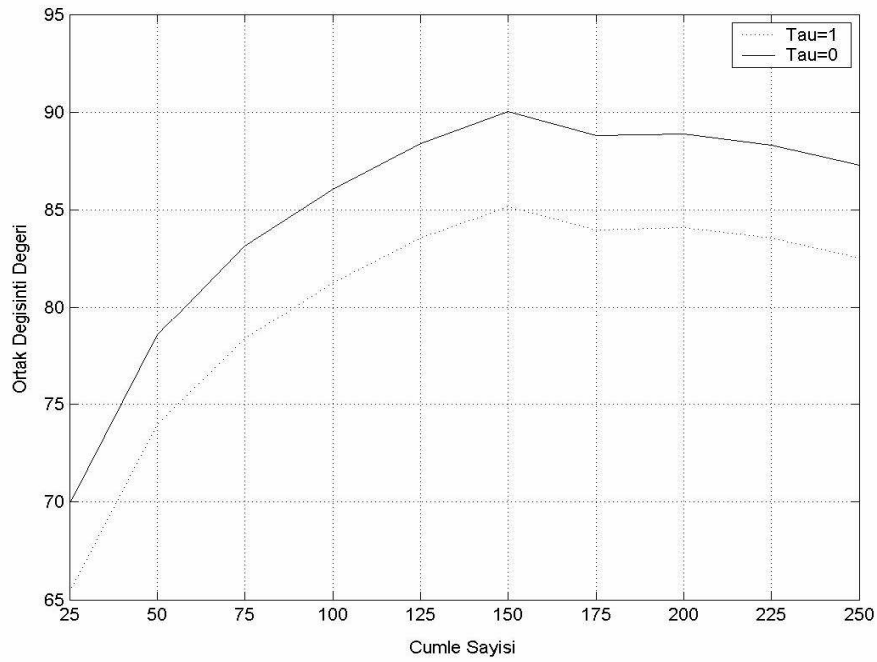
Şekil 7.13'te görüldüğü gibi uyarlama sayısı arttıkça yeni sayı dizisi üzerinde paket kayıp kestirimi daha doğru olmaktadır.



Şekil 7.13 Uyarlamalı ve uyarlamasız kestirim yönteminin yapay test verisi üzerinde gösterimi.

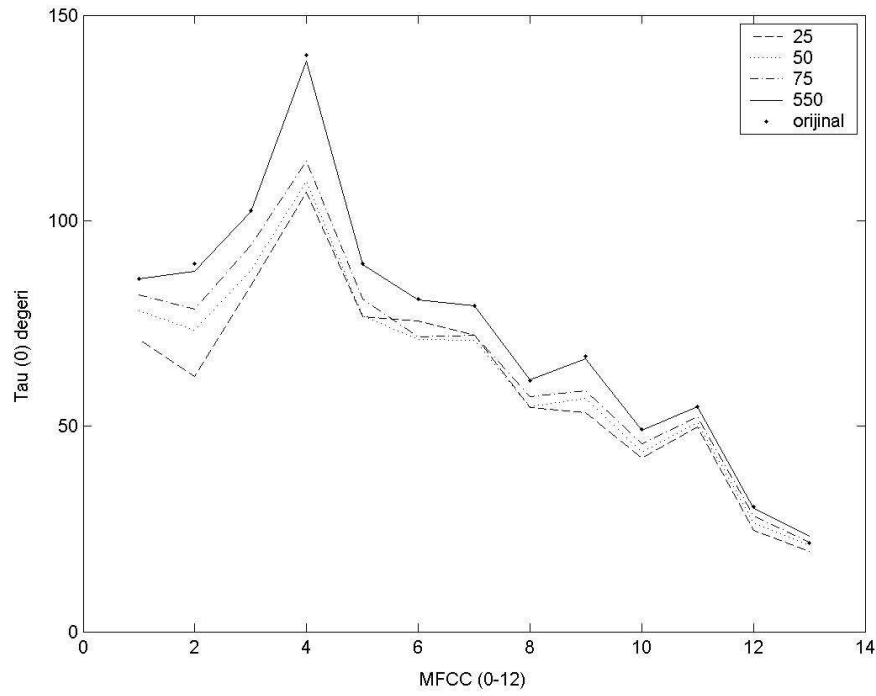
MFCC vektörleri ile birinci ve ikinci türevleri arasındaki genlik farkı yüksektir ve birinci ve ikinci türevlerin genlik değerleri çok düşük olduğu için yüksek sayıda ses öznitelik vektörü olduğu durumda güncellemeler sırasında uyarlamaya katkıları ihmal edilebilir. Bunun için, ortalama değer ve değişinti uyarlaması ses öznitelik vektörünün ilk 13 elemanı üzerinde yapılmıştır (MFCC (0)-MFCC (12)).

Şekil 7.14'te MFCC(0) için farklı cümle sayılarında eğitim sonucu elde edilen ortak değişinti ve çapraz ortak değişinti değerleri gösterilmektedir. Şekilde görüldüğü gibi eğitim cümle sayısına bağlı olarak ortak ve çapraz ortak değişinti değerleri değişmektedir.



Şekil 7.14 MFCC(0) için cümle sayısına göre hesap edilen ortak değışinti ($\tau = 0$) ve çapraz ortak değışinti ($\tau = 1$) değışimi.

Uyarlama sonucu ($\tau = 0$) değışimi Şekil 7.15'te gösterilmiştir. τ değeri ilk 13 MFCC değeri için hesap edilmiştir. İlk 25 cümle eğitim amaçlı kullanılmış ve farklı sayıda cümleler için (50, 75 ve 550) uyarlama sonucu değışen ortak değışinti değeri elde edilmiştir. Noktalı olan değeri ise uyarlama yapılmadan doğrudan 550 cümle üzerinden elde edilen değeri'dir. Şekilde görüldüğü gibi uyarlama sonucu orijinal değeri'ne yaklaşılmıştır.



Şekil 7.15 $\tau (0)$ deđerinin farklı cümle sayılarına bađlı olarak deđişimi.

Uyarlamalı ve uyarlamasız EBSK yöntemi ile telafi sonrası konuşma tanıma başarımları Çizelge 7.12’de gösterilmiştir.

Çizelge 7.12 EBSK yönteminin uyarlamalı ve uyarlamasız durumlarda başarımı

EBSK ve Uyarlamalı EBSK Başarımı				
Kayıp Yapısı	Eđitim Cümle Sayısı	Test Cümle Sayısı	Uyarlama	KHO (%)
5-20	25	528	Yok	37,7
			Var	37,6
	400	153	Yok	39,2
			Var	39,2

Eđitim amaçlı 25 cümle seçildiğinde eğitim sonrası geri kalan 538 cümle ile sistem zaman içerisinde uyarlamalı olarak geliştirilmiş ve sonuçta 0.1 puanlık bir gelişme sağlanmıştır. 400 cümle ile eğitim sonrası sadece 153 cümle kaldığı için

uyarlamanın etkisi gözlenmemiştir. Ancak, 400 cümle ile eğitim sonrası 153 test cümlesi üzerinden kelime hata oranı bir miktar fazla çıkmaktadır.

BÖLÜM 8

GENEL SONUÇLAR VE TARTIŞMA

Bu çalışmada dağıtık yapı modelinde Türkçe sürekli konuşma tanıma sistemlerinin kayıplı durumlarda başarımları araştırılmış ve telafi yöntemleri üzerinde geliştirmeler yapılmıştır.

Paket kayıplarının analizi ve telafi yönteminin seçildiği önışlem, telafinin uygulandığı telafi aşaması ve sistem parametrelerinin güncellendiği ardışlem aşamalarından oluşan paket kayıp telafi sistemi tasarlanmıştır. İşlem gücü ve hafıza kullanımının çok önemli olduğu gerçek zamanlı sistemler için dinamik bellek tanımlanmıştır. Genel amaçlı tasarlanan paket kayıp telafi sisteminde ardışlem aşaması çok farklı parametre ve güncellemeler ile telafi sonrası iyileştirmeler içinde kullanılabilir.

Paket kayıpları rasgele, seyrek ve patlamalı kayıplar olarak ayrı ayrı incelenmiştir. Telafi yöntemi olarak yerine koyma yöntemi rasgele kayıplar üzerinde uygulanmış ve 1 ya da 2 ardışıl kayıp olduğu durumlarda kullanılabileceği görülmüştür. Aradeğerleme yöntemleri seyrek kayıplarda kayıp merkezinin az olduğu durumlarda yüksek başarımlar göstermiştir. Ancak, Spline yöntemi Lagrange yöntemine göre daha yüksek başarımlar göstermektedir. Patlamalı kayıplarda Lagrange aradeğerleme yönteminin başarımlarını hızlı bir şekilde düşerken 5 ve 10 ardışıl paket durumlarında Spline yönteminin başarımlarını da belirgin bir şekilde düşmektedir. Kestirim yöntemleri patlamalı kayıplarda aradeğerleme yöntemlerine göre daha yüksek başarımlar göstermektedir.

Uyarlamalı kestirim yöntemi eğitim aşamasında az verinin seçilmesi ve zaman içerisinde değişen istatistiksel bilgilere göre sistem başarımlarının artması için uygulanabilir. Bu çalışmada tek bir kullanıcı ve sadece haber programlarına ait cümlelerden oluşan test verisine rağmen uyarlamalı kestirim yöntemi uyarlamasız yöntemine göre 0.1 puan daha yüksek başarımlar göstermiştir. Farklı kullanıcı ve farklı

alanlardaki konuşma cümleleri için uyarlamalı kestirim yönteminin başarımı çok daha fazla olabilir.

Test edilen kayıp durumları ve elde edilen sonuçlar doğrultusunda telafi yöntemlerinin ve sistemlerinin özellikle patlamalı kayıp durumları için geliştirilmesi gerektiği açıktır. Telafi yöntemlerini geliştirmek için kestirim ve uyarlamalı kestirim yöntemleri üzerinde çalışmalar yapılabileceği gibi ses öznitelik parametrelerinden ses işaretleri geri elde edilerek Türkçe kelime yapıları ve dil modellerine özgü çalışmalar da yapılabilir.

KAYNAKLAR DİZİNİ

- Andersen, S., Kleijn, W., Hagen, R., Linden, J., Murthi, M., and Skoglund, J., 2002, iLBC-a linear predictive coder with robustness to packet losses, in Proc. IEEE Speech Coding Workshop, pp. 23–25.
- Andrews, K., Heegard, C., Kozen, D., 1997, A theory of interleavers, Technical Report 97-1634, Computer Science Department, Cornell University.
- Barkana, A., Gülmezoğlu, M.B., Edizkan, R., Künkçü, Ü., Künkçü, Ş.E., 1995, Konuşma tanıma için veri tabanı ve yapılan çalışmalar, Konuşma Tanıma Çalıştayı, Ankara.
- Bayer, S., 1996, Embedding speech in web interfaces, Proceedings of ICSLP, pp. 1684-1687, Philadelphia, PA.
- Bernard, A., Alwan, A., 2002, Low-bitrate distributed speech recognition for packet-based and wireless communication, IEEE Transaction on Speech and Audio Processing, vol. 10, No. 8, pp. 570-579.
- Besacier, L., Mayorga, P., Bonastre, J.-F., Fredouille, C., Meignier, S., 2003, Overview of compression and packet loss effects in speech biometrics, IEE Proceedings, Vision, Image and Signal Processing, Vol. 150, Issue:6
- Bolot, J.C., 1993, End-to-end packet delay and loss behavior in the internet. In Proc. ACM Sigcomm'93, pages 189–199, San Francisco, CA.
- Bolot, J., Fosse-Parisis, S., Towsley, D., 1999, Adaptive FEC based error control for interactive audio in the Internet, Infocom
- Borella, M., 2000, Measurement and interpretation of internet packet loss, Journal of Communication and Networks, 2(2):93–102.

KAYNAKLAR DİZİNİ (devam)

- Boulis, C., Ostendorf, M., Riskin, E.A., Otterson, S., 2002, Graceful degradation of speech recognition performance over packet-erasure networks, *IEEE Trans. On Speech and Audio Proc.*, Vol. 10, No. 8.
- Chazan, D., Hoory, R., Cohen, G., Zibulski, M., 2000, Speech reconstruction from mel frequency cepstral coefficients and pitch, in *Proc. ICASSP'00*.
- Chua, T.K., Pheanis, D.C., 2006, Effects of loss characteristics on loss-recovery techniques for VoIP, *ICN/ICONS/MCL*, 204.
- Clark, A. D., 2001, Modeling the effects of burst packet loss and recency on subjective voice quality, *Proc. of IP Telephony Workshop*.
- Cohen, D., 1977, Specifications for the network voice protocol, Network Working Group, RFC 741.
- Conte, S.D., 1965, *Elementary numerical analysis: an algorithmic approach*, McGraw-Hill, New York, p. 278.
- Conte, S.D., Boor, C.D., 1972, *Elementary numerical analysis; an algorithmic approach*, Second Edition, McGraw-Hill, New York, p. 340.
- Cooke, M., Green, P., Crawford, M., 1994, Handling missing data in speech recognition, in *Int. Conf. On Spoken Language Processing (ICSLP)*.
- Demichelis, P., Rinott, A., Martin, J.C.D, 2005, Performance analysis of distributed speech recognition over 802.11 wireless networks on the TIMIT database, *Proceedings of IEEE 61st Semiannual Vehicular Technology Conference (VTC) Stockholm, Sweden*.
- Davis, S. B. and Mermelstein, P., 1980, Comparison of parametric representations of monosyllabic word recognition in continuously spoken sentences, *IEEE Trans. ASSP*, Vol 28, pp 357–366, 1980.

KAYNAKLAR DİZİNİ (devam)

- Digalakis, V.V., Neumeyer, L.G., and Perakakis, M., 1999, Quantization of cepstral parameters for speech recognition over the world wide web,” IEEE J. Select. Areas Commun., vol. 17, pp. 82–90.
- ETSI Standard, 2000, ES 201 108 V1.1.1 – STQ : DSR – Front-end feature extraction algorithm; compression algorithm.
- ETSI Standard, 2003, ES 202 212- STQ: DSR; Extended advanced front-end feature extraction algorithm; Compression algorithms; Back-end speech reconstruction algorithm.
- Euler S., and Zinke, J., 1994, The influence of speech coding algorithms on automatic speech recognition,” Proc. ICASSP, Vol 1, pp. 621-624.
- Forney, G.D., 1973, The Viterbi algorithm, Proc. IEEE, vol. 61, pp. 268-278.
- Gauvain, J., Lamel, L., 2000, Large-vocabulary continuous speech recognition: advances and applications, in Proc. IEEE, Vol. 88, No. 8.
- Gerald, C.F., 1970, Applied Numerical Analysis, Addison-Wesley, pp. 340.
- Goddeau, D., 1997, Deploying speech applications over the web, Proceedings of Eurospeech, pp. 685-688, Rhodes, Greece.
- Goodman, D.J., Lockhard, G.B., Wasem O.J., 1986, Waveform substitution techniques for recovering missing speech segments in packet voice communications, IEEE Transactions on Acoustics, Speech and Signal Processing, Vol. ASSP-34, Bo. 6.
- Gündüzhan, E., Momtahan, K., 2001, A linear prediction based packet loss concealment algorithm for PCM coded speech, IEEE Transactions on Speech and Audio Processing, Vol. 9, No. 8.

KAYNAKLAR DİZİNİ (devam)

- Hacioglu, K., Pellom, B., 2003, A distributed architecture for robust automatic speech recognition, IEEE ICASSP, Hong Kong.
- Hardman, V., Sasse, A., Handley, M., Watson, A., 1995, Reliable audio for use over the Internet, in Proc. INET'95, pp.171-178.
- Hassan, M., Nayandoro, A., and Atiquzzaman, 2000, M., Internet Telephony : Services, Technical Challenges and Products, IEEE Communications Magazine, Vol. 38, No. 4, pp. 96-103.
- Hermansky, H., 1990, Perceptual linear predictive (PLP) analysis of speech, J. Acoust. Soc. America, vol. 87, no. 4, pp. 1738-1752.
- Hochberg, M.M., Renals, S.J., Robinson, A.S., Kershaw, D., 1994, Large vocabulary continuous speech recognition using a hybrid connectionist-HMM system, in Proc., ICSLP'94, Yokohama, Japan, pp. 1499-1502.
- Huerta, J.M., Stern, R.M., 1998, Speech recognition from GSM codec parameters, in Proc. ICLSP, pp. 1463-1466.
- ITU-T Recommendation. G.729, 1996, Coding of speech at 8kbits/s using conjugate algebraic code-excited linear prediction (CS-ACELP).
- ITU-T Recommendation. G.722.2, 2003, Wideband coding of speech at around 16 kbit/s using adaptive multi-rate wideband (AMR-WB).
- James, A.B., Milner, B.P., 2004, An analysis of interleavers for robust speech recognition in burst-like packet loss, Proc. ICASSP.
- Jiang, W., and Schulzrinne, H., 1999, Qos measurement of internet real-time multimedia services. Technical Report CU-CS-015-99, Columbia University.

KAYNAKLAR DİZİNİ (devam)

- Jiang, W., Schulzrinne, H., 2000, Modeling of packet loss and delay and their effect on real-time multimedia service quality, in Proceedings of ACM NOSSDAV.
- Kim, H. K., Cox, R., 2001, Bitstream-based feature extraction for wireless speech recognition, in Proc. ICASSP, vol. 1, pp. 21-24.
- Larson, J.A., 2000, Speech-enabled Appliances, <http://www.larson-tech.com>
- Lilly, B.T., and Paliwal, K.K., 1996, Effect of speech coders on speech recognition performance, in ICSLP 96, (Philadelphia,PA), pp. 2344–2347.
- Markowitz, J.A., 1996, Using speech recognition, Prentice-Hall, New Jersey, p. 292.
- Mayorga, P., Lamy, R., and Besacier, L., 2002, Recovering of packet loss for distributed speech recognition, in Proc. EUSIPCO.
- Meijering, E., 2002, A chronology of interpolation: From ancient astronomy to modern signal and image processing, *Proc. of the IEEE*, vol. 90.
- Milner, B.P., Semnani, S., 2000, Robust speech recognition over IP networks, ICASSP 2000, İstanbul, Turkey.
- Milner, B.P., James, A.B., 2006, Robust Speech Recognition Over Mobile and IP Networks in Burst Like Packet Loss, *IEEE Transactions on Audio and Language Processing*, Vol. 14, No. 1.
- Milner, B.P., James, A.B. 2001, Analysis and compensation of packet loss in distributed speech recognition using interleaving, Proc. Eurospeech.
- Milner, B.P., 2001, Robust speech recognition in burst-like packet loss, Proc. ICASSP.

KAYNAKLAR DİZİNİ (devam)

- Moon, T.K., Stirling, W.C., 2000, Mathematical methods and algorithms for signal processing, Prentice Hall, NJ, 542-584.
- NIST-Score, 2000-2007, Speech Recognition Scoring Package (SCORE) Version 3.6.2, <http://www.nist.gov/speech/tools/index.htm>.
- Oppenheim, A.V., Shafer, R.W., 1989, Digital signal processing, Prentice Hall.
- Palaz, H., Kanak, A., Bicil, Y, Uğur Doğan, M., İslam, T., 2005, Tren-Turkish speech recognition platform, 13th European Signal Processing Conference, Sept 4-8.
- Perkins, C., Kouvelas, I., Hodson, O., Hardman, V., Handley, M., Bolot, J.C., Vega-Garcia, A., and Fosse-Parisis, A., 1997, RTP Payload for Redundant Audio Data, RFC 2198.
- Perkins, C., Hodson, O., Hardman, V., 1998, A survey of packet loss recovery techniques for streaming audio, IEEE Network, Vol. 12, Issue 5, pp. 40-48.
- Perkins, C., 2003, RTP audio and video for the Internet, Addison-Wesley, p. 414.
- Postel, J., 1980, User Datagram Protocol, RFC 768.
- Postel, J. (ed.), 1981, Internet protocol - DARPA Internet program protocol specification, RFC 791, USC/Information Sciences Institute.
- Postel, J. (ed.), 1981, Transmission control protocol, RFC 793.
- Rabiner, L.R., Juang, B.-J, 1993, Fundamentals of Speech Recognition., Prentice Hall Processing Series, New Jersey.
- Raj, B., 2000, Reconstruction of incomplete spectrograms for robust speech recognition, PhD Thesis, Carnegie Mellon University.

KAYNAKLAR DİZİNİ (devam)

- Ramaswamy, G.N., and Gopalakrishnan, P.S, Compression of acoustic features for speech recognition in network environments,” in Proc. ICASSP, Seattle, WA, 1998, pp. 977–980.
- Renals, S., and Hochberg, M.M., 1999, Start-synchronous search for large vocabulary continuous speech recognition, IEEE Transaction on Speech and Audio Processing, Vol. 7, No. 5.
- Rosenfeld, Ronald, 2000, Two decades of statistical language modeling: where do we go from here?, Proc. IEEE, vol. 88, pp.1270-1278.
- Sanghi, D., Agrawala, A.K., Gudmundsson, O., Jain, B.N., 1993, Experimental assessments of end-to-end behavior on Internet, in Proceedings IEEE INFOCOM’93, pp. 867-874
- Sanneck, H., Carle, G., and Koodi, R., 2000, A framework model for packet loss metrics based on loss run-lengths, In SPIE/ACM SIGMM Multimedia Computing and Networking Conference.
- Sayood, K., 2000, Introduction to data compression, Academic Press, Second Edition.
- Schulzrinne, H., Casner, S., Frederick, R., V. Jacobson, 2003, RTP: A transport protocol for real-time applications, RFC 3550.
- Shao, X., Milner, B.P., 2003, Clean speech reconstruction from noisy MFCC’s using sinusoidal model, in Proc., ICASSP’03, Hong Kong.
- Srinivasamurthy, N., Ortega, A., Zhu, Q., Alwan, A, 2000, Towards efficient scalable speech compression schemes for robust speech recognition applications, Proc. IEEE ICME, p. 249-52 Vol.1.
- Stevens, W.R., 1994, TCP.IP Illustrated, Volume 1: the protocols, Addison-Wesley, Massachusetts, pp. 576.

KAYNAKLAR DİZİNİ (devam)

- Tuğaç, S., İlk, H.G., 2004, Mixed excitation linear prediction (MELP) kodlayıcısı kullanan bit katarı tabanlı kablosuz konuşma tanıyıcı, IEEE Proceedings of Signal Processing and Communications Applications Conference, Kuşadası, Turkey.
- Uzun, İ., Edizkan, R., 2007, Dağıtılmış sistemlerde Internet protokolu üzerinden yalıtık ses tanıma, 15. Sinyal İşleme ve İletişim Uygulamaları Kurultayı, SIU'07, Eskişehir.
- Uzun, İ., Arısoy, E., Edizkan, R., Saraçlar, M., 2007, Dağıtık yapıda Türkçe sürekli konuşma tanıma sisteminde seyrek paket kayıplarının analizi ve telafisi, 15. Sinyal İşleme ve İletişim Uygulamaları Kurultayı, SIU'07, Eskişehir.
- Yajnik, M., Moon, S., Kurose, J., Towsley, D., 1999, Measurement and modelling of temporal dependence in packet loss, in Proc. IEEE Infocom'99, New York, NY.
- Young, S., 1996, A review of large-vocabulary continuous-speech recognition, IEEE Signal Processing Magazine.
- Young, S.J., Evermann, G., Hain, T., Kershaw, D., Moore, G.L., Odell, J.J., Ollason, D., Povey, D., Valtchev, V., Woodland, P.C., 2005, The HTK Book (Version 3.3), Documentation, HTK.
- Wang, J., Gibson, J. D. , 2001, Parameter interpolation to enhance frame erasure robustness of CELP coders in packet networks, in Proc. IEEE Int. Conf. Acoust., Speech, Signal Process., Salt Lake City, UT, vol. 2, pp. 721–724.
- Wasem, O. J., Goodman, D. J., Dvorak, C. A., 1988, The effect of waveform substitution on the quality of PCM packet communications, IEEE Trans. Acoust. Speech Signal Process., vol. 36, no. 3, pp. 342–348.
- Weerackody, V., Reichl, W., Potamianos, A., 2002. An errorprotected speech recognition system for wireless communications. IEEE Trans. Wireless Comm. 1 April, 282–291.

ÖZGEÇMİŞ

İsmail Uzun 2 Ocak 1976 Afyonkarahisar/Emirdağ doğumludur ve Türkiye Cumhuriyeti vatandaşıdır. İlköğrenimini köyünde tamamladıktan sonra eğitimine Eskişehir’de devam etmiştir. Eskişehir Osmangazi Üniversitesi’nde 1998 yılında lisans, 2000 yılında “Sinyal İşleme Teknikleri ile Gri Seviyeli Tek Resimden Üç Boyutlu Yüzey Elde Etme” konulu tezi ile yüksek lisansını tamamlamıştır. Aynı üniversitede 1998-2001 yılları arasında Telekomünikasyon Bilim Dalında araştırma görevlisi olarak çalışmıştır. İş hayatına özel sektörde devam etmektedir ve sayısal işaret işleme, IP üzerinden ses iletimi ve gömülü sistemler konularında çalışmaktadır. IP üzerinden ses iletimi ve konuşma tanıma konularında akademik çalışmalar da yapmakta ve doktora tezini yürütmektedir. Evli ve iki çocuk babasıdır.