

Adli Bilişim Açısından Ses İncelemeleri

Yunus KORKMAZ¹, Aytuğ BOYACI²

^{1,2}Fırat Üniversitesi Teknoloji Fakültesi Adli Bilişim Mühendisliği Bölümü, Elazığ, Türkiye
¹ykorkmaz@firat.edu.tr, ²aytugboyaci@firat.edu.tr

(Geliş/Received: 16.11.2017; Kabul/Accepted: 23.02.2018)

Özet

Geçmişte fiziksel araçlar ve yollarla işlenen suçlar dijitalleşmenin artmasıyla giderek yerini elektronik araçlarla işlenen suçlara bırakmaktadır. Elektronik araçlarla işlenen bu suçlar arkasında, bilgisayarlar, taşınabilir cihazlar, ağ cihazları ve depolama aygıtları gibi elektronik deliller bırakmaktadır. Ses delillerinin incelenmesi ya da ses adli bilişimi (audio forensic), elektronik delilleri ve içlerinde bulunan verileri kullanarak suçların aydınlatılmasına katkıda bulunan Adli Bilişim disiplininin bir alt dalı olarak bilinmektedir. Ses adli bilişimi için halihazırda kullanılabilen çok ses analiz yazılımı mevcuttur. Sayısal ses işleme teknikleri tabanlı yöntemleri kullanan bu yazılımların, ses kayıtlarında kurgu/montaj tespiti (kayıt bütünlüğünün doğrulanması), spektrum analizi, bozuk kayıtların anlaşılabilirliğinin artırılması (kayıt iyileştirme), konuşmacı profili (yaş, cinsiyet vs.) belirleme, ortam gürültüleriyle maskelenmiş ses analizi, adli takipterde gerçek zamanlı ses analizi ve konuşmacının etnik yapısı ile ilgili olan dil ve aksan öğelerini belirleme gibi ses adli bilişimi işlemlerini otomatik olarak gerçekleştirebilmesi beklenmektedir. Bu makalede, konuşma üretimi ve ses bilimi, ses incelemeye kullanılan temel konuşma işleme teknikleri, adli bilişim açıdan ses incelemeleri ve bu incelemelerde kullanılan yazılımların teknik ve performans kriterleri göz önünde bulundurularak karşılaştırılması yapılmıştır.

Anahtar Kelimeler: Adli Bilişim; Ses İnceleme; Konuşma İşleme; Ses Analiz Araçları; Ses Adli Bilişimi.

Audio Analysis in Terms of Digital Forensics

Abstract

Day by day, the crimes committed by physical means in the past give way to the crimes committed by electronic means with the increase of digitization. These crimes committed by electronic devices leave electronic evidences such as computers, mobile devices, network tools, storage devices behind. The job of examining audio evidences, mostly called as “audio forensic”, is a sub-branch of digital forensic area which contributes clarification of crimes by using electronic devices and data included in these devices. There are already plenty of audio analysis tools that can be used for audio forensic. These tools, which basically use DSP-based methods, are expected to automatically implement audio forensic tasks such as verification of audio integrity (editing and montage detection), spectrum analysis, speech enhancement (increasing the intelligibility of distorted records), speaker profile (age, sex, etc.) identification, analysis of audio masked by ambient noise, real-time audio analysis in judicial tracking and identification of ethnic items like language and accent. In this article, speech production and phonetics, fundamental digital speech processing methods used for audio analysis, audio examinations in terms of digital forensics and technical and performance comparisons of audio analysis tools used in these examinations are covered in detail.

Keywords: Digital Forensics; Audio Examination; Digital Speech Processing; Audio Analysis Tools; Audio Forensics.

1. Giriş

Dijital çağ olarak adlandırılan günümüzde yaşamı kolaylaştıran teknolojilerin kullanımı giderek yaygınlaşmaktadır. Hızla gelişen teknoloji sayesinde her yeni gün hayatımıza farklı cihazlar entegre olmaktadır. Geçmişte fiziksel temas yoluyla işlenen suçlar günümüzde teknolojinin bu denli hızlı gelişimi sayesinde üretilen cihazlar yoluyla dijital olarak işlenmeye

başlamıştır. Dijital yollarla işlenen suçlarda kullanılan cihazlar ve bu cihazlar içerisinde bulunan verilere “dijital delil” adı verilmektedir. Adli bilişim (digital forensics), teknoloji ile hayatımıza giren dijital cihazlarla işlenen suçların dijital deliller kullanılarak aydınlatılması disiplnidir. Adli bilişim, incelenen cihazın türüne göre kendi içerisinde bilgisayar, mobil cihaz, ağ (network), veritabanı, RAM (Random Access Memory), görüntü ve ses adli bilişimi

gibi farklı alanlara ayrılmaktadır. Bu alanlardan ses adli bilişimi (audio forensics) adlı vakalarda ses delillerinin incelenip anlamlı sonuçlar elde edilmesi açısından önem arzettmektedir.

Ses adli bilişiminde;

- Konuşma çözümleme (transkripsiyon)
- Ses kayıtlarında kurgu/montaj tespiti için kayıt bütünlüğünün doğrulanması
- Ses kaydındaki konuşmacının tespiti için otomatik ses izi eşleştirme
- Spektrum analizi
- Ünsüz harfler bazında formant frekansları tespiti
- Bozuk kayıtların anlaşılabilirliğinin arttırılması (kayıt iyileştirme)
- Konuşmacı profili (yaş, cinsiyet vs.) belirleme
- Birden çok konuşmacının bulunduğu kayıtlarda konuşmacıları ayırt etme
- Ortam gürültüleriyle maskelenmiş ses analizi
- Adli takiplerde gerçek zamanlı ses analizi
- Kayıt cihazına yakın yerde bulunan radyo ve televizyon gibi cihazlardan kaynaklanan GSM gürültülerini yok etme
- Konuşmacının etnik yapısı ile ilgili olan dil ve aksan öğelerini belirleme

gibi ses inceleme işlemleri gerçekleştirilmektedir.

Günümüzde, ses kayıtlarının adlı vakalarda delil olarak kullanılmasının yanı sıra bilişim alanında çalışan firmaların sayısındaki artış, ses adli bilişimi konusunda analizler yapabilen açık kaynak kodlu veya lisanslı bir çok sistemin (yazılım/donanım) geliştirilmesine olanak sağlamıştır.

Ses incelemeye geliştirilen yazılımlar temelde dijital sinyal işleme (digital signal processing, DSP) tekniklerinden faydalananarak konuşma işleme gerçekleştirmektedir. Konuşma işleme genel olarak öznitelik çıkarma, konuşma tanıma, konuşmacı tanıma, ses iyileştirme ve ses sentezleme gibi alt başlıklara ayrılmaktadır.

Bu çalışmada konuşma üretimi ve ses bilimi, ses incelemeye kullanılan temel konuşma işleme teknikleri, adlı açıdan ses incelemeleri ve bu incelemelerde kullanılan yazılımların teknik ve performans kriterleri göz önünde bulundurularak karşılaştırılması yapılmıştır.

2. Konuşma / İşitme Fizyolojisi ve Ses Bilimi

2.1 Ses ve Konuşma

Ses, belirli frekans değeriyle titreyen bir cisim tarafından oluşturulan ve işitme hissi uyandıran basınç dalgası olarak tanımlanmaktadır. Ses, yayılma hızı, frekans, basınç, dalga boyu, desibel, genlik, şiddet, renk ve tını gibi fiziksel özelliklere sahip olabilmektedir [1].

İnsanlarda ses oluşturma sisteminin 3 ana bölümden oluşur [2]:

- Akciğer
- Gırtlak (ses telleri, ses kıvrımları)
- Ses Bölgesi (yutak,ağız)

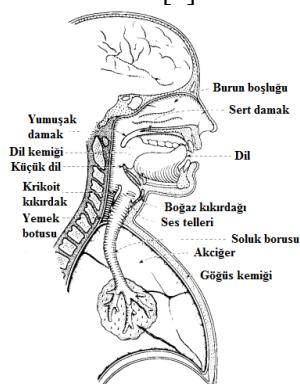
Bu sistemde, akciğerler hava kaynağı, ses telleri titreşim elemanı ve ağız rezonatör bölge olarak görev yapmaktadır.

2.2 Ses ve Konuşma Fizyolojisi

Ses dalgası, insanlarda ses üretim mekanizması tarafından üretilen akustik bir hava basıncı dalgası olarak ortaya çıkmaktadır. Konuşma sesleri genel olarak sesli kısım (voiced) ve sessiz kısım olmak (unvoiced) üzere ikiye ayrılmaktadır. Sesli kısım akciğerlerden gelen havanın gerilmiş ses tellerini titrettiği zaman oluşmaktadır. Sessiz kısım ise üretilen havanın ağızda ani ve düzensiz patlaması sonucu oluşur [3].

Ses Üretim Organları:

İnsan sesinde ayırt edici özellikler fiziksel ve öğrenilmiş olmak üzere iki genel başlığa ayrılmaktadır. Ses telleri ve yukarısında kalan organlar olarak tanımlanan vokal yolu, yapı itibarı ile insan sesini farklı kıلان fiziksel özelliklidir. Ses üretim organları alt yutak, orta yutak, ağız boşluğu, üst yutak ve burun boşluğu şeklinde sıralanmaktadır [4].



Şekil 1. Ses üretim mekanizması.

Ses Oluşumu:

Şekil 1'de de görüldüğü gibi akciğerlerin uyarım kaynağı olarak ürettiği hava akışının nefes borusu yardımıyla ses tellerine geçmesiyle fiziki olarak ilk ses dalgası üretim işlemi başlamaktadır. Uyarımın fonasyon, fisıldama, friksyon (sürtünme), basınçlandırma, titreşim gibi çeşitleri bulunmaktadır [5]. Ses telleri titreştiği anda altında ve üstünde yankılanmalar (sublottal resonance) oluşturur. Nefes borusunun yapısına göre değişebilen bu yankılanmalar konuşmacıya bağlı fiziksel bir özelliktir [6].

2.3 Ses Bilimi (Fonetik)

Bir dilin yapısını oluşturan seslerin fizyolojik olarak nerede ve nasıl çıkarıldıklarını, zamanla değişimlerini, konuşma organlarının işleyişini inceleyen, sesleri söyleyiş (ürtim), akustik (iletim) ve dinleyiş (algılama) özelliklerine göre sınıflandıran bilim dalına ses bilimi (fonetik) denmektedir [7]. Ses bilimi genel olarak fonem (phoneme) adı verilen seslerin oluşumları ve bu oluşum sırasında artikülatörlerde meydana gelen değişimleri konu alır. Örneğin /s/ foneminin çıkarılışı sırasında dudakların, dilin, akciğerlerden gelen basınçlı havanın oral kavitedeki hareketi, bu basınçlı havanın zamanla dil ve girtlak sayesinde nasıl kesildiğini araştırmaktadır [8].

Türkiye Türkçe'sinde Sesler:

Konuşma organlarının bir arada uyumla ve düzenli çalışmasıyla anlam ihtiva eden sözcükler ve tümceler oluşturmak için ağızdan çıkarılan birimlere ses (phon) denmektedir [9]. Seslerin kümelenmesiyle anlam ayırcı özelliği bulunan fonemler meydana gelmektedir. Türkçe'de fonemler parçalı ve parçalarüstü (bürün) olmak üzere ikiye ayrılmaktadır [10].

Parçalı Fonemler:

Türkçe'de parçalı fonemler ünlü, ümsüz ve kayan ünlü olmak üzere 3 başlıkta incelenmektedir.

Ünlüler ve Özellikleri:

Ünlü sesler akciğerlerden gelen havanın hiçbir sürtünme veya engele maruz kalmadan dilin üstünden geçerken çıkardığı seslerdir. Türkçe'de 8 adet ünlü harf bulunmakta ve tüm ünlüler 8 adet fonem ile temsil edilmektedir. Bu fonemler:

/a/, /e/, /o/, /ö/, /u/, /ü/, /ı/, /i/

şeklindedir. Ünlü sesler, dilin ağızındaki konumuna, dudakların biçimine, genizin açık/kapalı durumuna göre sınıflandırılmaktadır:

- Çene açıklığı (dar, geniş)
- Dil konumu (ön, arka)
- Dudakların biçimi (yuvarlak, düz)
- Geniz (açık, kapalı)

Türkçe'de geniz dışındaki durumlar önem arz etmektedir. Tablo 1'de ünlü fonemler çene açıklığı, dil konumu ve dudakların biçimine göre gösterilmiştir.

Tablo 1. Türkçe'deki ünlü fonemlerinin dilin konumu, dudak biçimini ve çene açıklığına göre sınıflandırılması.

	Dilin Konumu				Dudak Biçimi		Çene Açıklığı	
	Düz Ön	Yuvarlak Ön	Orta	Arka	Yuvarlak	Düz	Dar	Geniş
a					+		+	+
e	+					+		+
o					+			+
ö		+			+			+
ı			+			+	+	
i	+					+	+	
u					+		+	
ü		+			+		+	

Bir ünlü fonemin diğer bir ünlü fonem ile yer değiştirmesi sonucu anlam değişikliğinin meydana gelip gelmediğinin kontrol edildiği sözcük çiftlerine yalnız sözcük çiftleri denmektedir. Türkçe'deki ünlü fonemleri dilin konumu, dudak biçimini ve çene açıklığını esas alınarak tespit etmede yararlanılan yalnız sözcük

çiftleri Tablo 2, Tablo 3 ve Tablo 4'te verilmiştir.

Tablo 2. Çene açıklığı bazında yalnız çiftler.

Geniş / Dar	ı	i	u	ü
a	kar/kır	kar/kır	kar/kur	sar/sür
e	kes/kış	tez/tız	bez/buz	ses/süs
o	koş/kış	sos/sis	koş/kuş	son/sün
ö	söz/sız	söz/siz	son/sun	söz/süz

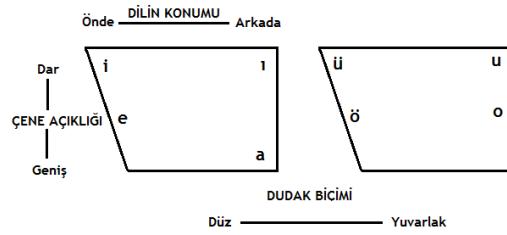
Tablo 3. Dudakların biçimini bazında yalın çiftler.

Düz / Yuvarlak	o	ö	u	ü
a	kar/kor	kar/kör	kar/kur	sar/sür
e	kel/kol	gel/göl	ser/sur	kel/kül
i	kış/koş	sır/sör	tır/tur	tır/tür
ı	kir/kor	kir/kök	kir/kur	kir/kür

Tablo 4. Dil konumu bazında yalın çiftler.

Arka / Ön	e	ö	i	ü
a	kas/kes	kar/kör	kar/kir	sar/sür
o	kol/kel	kor/kör	kor/kir	kor/kür
u	sur/ser	sun/sön	kur/kir	kul/kül

Ünlü fonemlerini sınıflandırmada kullanılan bir başka yöntem ise Şekil 2'de gösterilen ünlü dörtgenidir (vowel diagrams) [11].



Şekil 2. Türkçe'de ünlü dörtgeni.

Ünsüzler ve özellikleri:

Ünsüz fonemlerin çıkışında akciğerden gelen hava konuşma organlarının herhangi bir yerinde engele rastlamaktadır. Bu engeller dudaklar, dişler, alt ve üst damak ve dil ile oluşmaktadır. Ünsüz fonemle çıkış biçimlerine, çıkış yerlerine ve ses tellerinin durumuna göre Tablo 5'te gösterildiği gibi sınıflandırılmaktadır.

Tablo 5. Ünsüz fonemlerin sınıflandırılması.

	Özellik	b	c	ç	d	f	g	h	j	k	l	m	n	p	r	s	ş	t	v	y	z
Çıkış Biçimlerine Göre	Patlamalı	+			+		+							+				+			
	Genizden												+	+							
	Çarpmalı																+				
	Yan daralmalı												+								
	Sızmalı	+	+		+		+	+								+	+	+	+	+	+
Çıkış Yerlerine Göre	Çift dudak	+											+		+						
	Alt dudak- Üst diş					+													+		
	Dil ucu- Diş ardi					+													+		
	Dil ucu- Diş eti													+		+	+				+
	Dil ucu- Ön damak											+									
	Dil- Ön damak		+	+					+									+			+
	Dil- Damak sonu						+			+											
	Ses telleri							+													
Ses Telleri Titreşimine Göre	Ötümülü	+	+		+		+		+		+	+	+	+	+	+			+	+	+
	Ötümstüz				+		+		+		+			+		+	+	+			

Parçalarüstü Fonemler:

Parçalarüstü fonemler (bürünler) sesli ifade tanıma kapsamında, sesli ifade örüntüsünün yakalanmasına yardımcı olmaktan çok sesli ifadeden anlam çıkarma aşamasında önem taşımaktadır.

Süre:

Süre, bir sesin söylenirken kısa ya da uzun zaman alması ile ilgilidir. Örneğin; “düğün” sözcüğü söylenirken süre kısa tutulursa “dün” sözcüğü ile anlam karışıklığı oluşturabilir.

Türkçe'de süre seslenme ve buyurma biçimlerinde de ortaya çıkabilmektedir.

Ton (Perde değişimi):

Bir sesteki sikliğin yüksek ya da düşük olması ton olarak tanımlanmaktadır. Örneğin; “aferin” sözcüğü ile ya begeni ya da yanlış yapılan bir işe serzeniş dile getirme ton (perde değişimi) ile sağlanmaktadır.

Kavşak ve Durak:

Ünsüz ile biten bir kök ünlü ile başlayan ek aldığımda hece düzenin değişmesi kavşak olarak tanımlanmaktadır. Örneğin; “ki-lim”, “ki-li-min”

sözcüklerinde olduğu gibi. Durak, bir tümcede anlamda farklılık oluşturabilecek şekilde duraklama yapılmasıdır. Örneğin; "Kara, deniz, hava yolları" ve "Karadeniz havayolları" tümceleri arasındaki fark gibi.

Vurgu:

Bir sözcükteki herhangi bir hecenin diğer hecelere göre daha baskılı bir şekilde söylemenesine vurgu denmektedir. Örneğin; "varmış" sözcüğü için yapılan vurgu "Epeyce parası varmış" ve "eve varmış" cümlelerinde farklı yerlerde bulunmakta olmasından anlam ayırt edici özellik sağlamaktadır.

Ezgi:

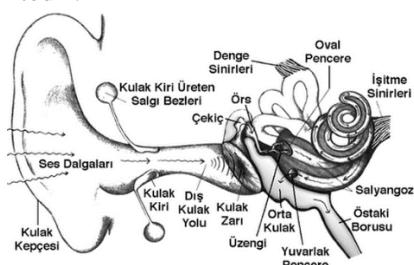
Sesli ifadedeki hece, durak ve vurguya bağlı ortaya çıkan ton değişimlerine ezgi denmektedir. Ruhsal etkenler, hece, durak ve vurgu değişimleri yoluyla, tümce içinde kimi sözcüklerdeki ton yükselmesi ya da alçaltılması, anlatılmak istenende belirli anlam ayrimı sağlamaktadır.

2.4 İshitme

Canlılar arasında iletişim, konuşma ve işitme olarak iki yönlü gerçekleşmektedir. İletişimde önemli bir role sahip olan işitme sisteminde, kulak işitmenin ilk adımı olan organdır [12].

Kulak Yapısı ve Görevi:

Şekil 3'te gösterildiği gibi hem harici kulak kanalını koruma görevini üstlenmekte hem de duyulabilir yüksek frekanslarda çok yönlü ses alabilen yapısıyla gelen sesin konumunu belirlemektedir.



Şekil 3. Kulağın yapısı [13].

Kulak genel olarak sesi toplayıp yönünü belirleyen ve kulak kepçesiyle başlayıp kulak zarı ile sonlanan dış kulak, ses dalgalarını mekanik dalgalara dönüştüren (empedans dönüşümü) çekici, örs, üzengi kemikçiklerinin bulunduğu orta kulak ve mekanik titreşimlerin sinirsel uyarılarına dönüştüğü yer olan

salyangozun bulunduğu orta kulak olmak üzere 3 kısma ayrılmaktadır.

Mekanik'ten Sinirsel Yola Geçis:

Mekanik dalgaın sinirsel aktiviteye dönüştüğü reseptörlerle organa corti organı denmektedir. Reseptörler salyangoz zarının içerisinde yerleşik halde bulunup işitsel sinir hücrelerinin de aralarında bulunduğu bir kısım sinir hücresi (nöron) ile bağlantı halinde bulunmaktadır. Bir nöron, belirli bir eşik değerinin üzerinde elektriksel girdi ile uyarıldığında, bir milisaniye süren elektrik sinyali (vuruş) üretir. Bu işlem nöron sinyalleşmelerinin beyne taşınmasını sağlamaktadır [14].

3. Ses İncelemede Kullanılan Teknikler

3.1 Ön İşlemler

Cerçeveleme (Frame Blocking):

Konuşma sinyalinin her biri X adet konuşma örneği içeren parçalara ayrılması işlemi cerçeveleme olarak tanımlanmaktadır. Cerçeveleme yapılrken birinci çerçeve X adet konuşma örneği içeriyorsa ikinci çerçeve birinci çerçeveden Y adet örnek sonra başlatılır. Cerçeveleme yapılrken dikkat edilmesi gereken husus Y değerinin X değerinden daha küçük bir değer olarak seçilmesi gerektidir ($Y < X$). Bu şekilde her çerçeve kendisinden sonra gelen çerçevenin bir kısmını örtmüştür. Örtme işleminde amaç bir çerçeveden diğerine geçişteki keskinliği ortadan kaldırılmaktır.

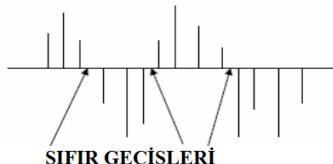
Pencereleme (Windowing):

Pencereleme, sinyalin N adet örneğinin (sample) pencere olarak tabir edilen matematiksel bir ifadeyle çarpılması işlemi olarak bilinmektedir. Pencereleme işlemi uygulanacak N adet örnek çerçeve (frame) adı verilen sinyal bölümünü oluşturur. Bir ses sinyalinden, o sinyale özgü özellikler çıkarılmak istendiğinde sinyal çerçevelere bölünüp bu çerçeveler üzerinden özellik çıkarılır. Bilinen pencereleme yöntemleri arasında Hamming, Hann (Hanning), Triangular, Gauss, Welch, Blackman ve Bartlett gibi yöntemler yer almaktadır.

Sıfır Geçiş Oranı (Zero Crossing Rate):

Sıfır geçiş oranı ayrık zamanlı bir sinyalin ardışık örneklerinin farklı matematiksel işaretle

sahip olma oranı olarak tanımlanmaktadır. Şekil 4'te gösterildiği gibi sıfır geçişlerinin gerçekleşme oranı sinyalin frekans içeriğinin basit bir ölçümünü yansımaktadır. Ses sinyallerinde sıfır geçiş oranı, belli bir zaman aralığında ya da bir çerçeveye içerisinde, ses sinyalinin genlik (amplitude) değerinin sıfır değerini kaç defa geçtiği ile ölçülmektedir.



Şekil 4. Ayrık zamanlı bir sinyaldeki sıfır geçişleri.

Enerji:

Ses sinyalinin kısa zamanlı enerji değeri sinyal genliğindeki değişimleri yansıtmaktadır. Genel olarak ses sinyallerine ait özellikler zaman uzayında önemli ölçüde çeşitlilik göstermektedir. Ses sinyalinin enerji değeri sesli-sessiz bölge ayrimı yapmaktadır da kullanılmaktadır. Bir ses sinyalinin enerji değeri sesli bölgelerde, sıfır geçiş oranı değeri ise sessiz bölgelerde yüksektir.

Hızlı Fourier Dönüşümü (Fast Fourier Transform, FFT):

Hızlı Fourier Dönüşümü, N adet konuşma örneği içeren her çerçeveyi orjinal uzayından frekans uzayına çevirmek için kullanılmaktadır. Buradaki amaç ses sinyalinin spektrum özelliklerini görmektir. Eşitlik 1'deki gibi tanımlanan Hızlı Fourier Dönüşümü, Ayrık Fourier Dönüşümü hesaplamasındaki zaman karmaşıklığı olan $O(n^2)$ 'yi azaltmak için kullanılan hızlı bir algoritma olarak bilinmektedir.

$$X_k = \sum_{n=0}^{N-1} x_n e^{-i2\pi kn/N} \quad k = 0, 1, \dots, N-1 \quad (1)$$

Özellik Çıkarma Yöntemleri:

Konuşma sinyali içerisinde ortam sesleri, kullanılan donanımdan kaynaklanan gürültüler ve konuşmaciya ait akustik özellikler bulunmaktadır. Mel Frekansı Kepstrum Katsayıları (Mel Frequency Cepstrum Coefficients, MFCC), Doğrusal Öngörülü Kodlama (Linear Predictive Coding, LPC) ve Algısal Doğrusal Öngörü (Perceptual Linear Predictive, PLP) yöntemleri özellikle konuşmacı tanıma işlemlerinde konuşmaciya özgü akustik

model çıkarmada kullanılan yöntemler arasında bulunmaktadır. Mel Frekansı Kepstrum Katsayıları (MFCC), insan kulağı referans alınarak oluşturulan Mel ölçüği kullanılarak frekans uzayında en yaygın kullanılan özellik çıkarma yöntemlerinden biridir. Frekans uzayında özellikler olarak da bilinen MFCC, zaman uzayında işlem yapılmasıdan daha net sonuçlar üretmektedir [15]. MFCC, FFT sonucu üretilen pencerelenmiş kısa zamanlı sinyalin gerçek katsayılarını göstermektedir. Bir konuşma sinyaline MFCC uygulamak için sırasıyla aşağıdaki adımlar uygulanmaktadır [16].

- Ses sinyali her çerçevede aynı sayıda örnek bulunacak şekilde çerçevelere ayrılr.
- Her çerçeveye pencereleme yöntemlerinden biri (ör: Hamming) uygulanır.
- Zaman uzayında bulunan ses sinyalinin frekans değerlerini çıkarmak için her çerçeveye FFT uygulanır.
- FFT uygulanan her çerçeveye 1 kHz'e kadar doğrusal 1 kHz'den yüksek frekanslarda logaritmik olan mel ölçekli filtre bankası uygulanır.
- Filtre bankası uygulanan çerçevelere daha sonra Ayrık Kosinüs Dönüşümü (DCT) uygulanır.

Konuşma sinyali içindeki her çerçeve için MFCC aşamaları uygulanarak bu işlemler sonunda bir MFCC seti elde edilmektedir. Bu set, sesin fonetik olarak önemli karakteristigi içeren akustik vektör seti olarak da tanımlanmaktadır.

3.2 Konuşma / Konuşmacı Tanıma

Dinamik Zaman Eşleştirme (Dynamic Time Warping):

Dinamik zaman eşleştirme, iki özellik vektörü dizisini zaman ekseniinde tam bir örtüşme sağlanana kadar daraltıp genişleterek eşleştirmeye çalışan döngüsel bir algoritma olarak tanımlanmaktadır. Dinamik zaman eşleştirme algoritması genel olarak zamanla değişiklik gösteren iki dizi arasındaki uzaklığı hesaplamak için kullanılmaktadır. Bu algoritma ayrıca gerçek zamanlı bir konuşma tanıma sisteminde konuşmacının farklı hızlarda ifade

ettiği aynı kelimeyi tanımak için kullanılmaktadır. Başka bir deyişle dinamik zaman eşleştirme algoritması yaklaşık şablonu olan iki sinyal arasındaki benzerliği arayan en uygun algoritma olarak bilinmektedir [17].

Örütü Eşleştirme (Pattern / Template Matching):

Konuşma tanıma sistemi için kullanılacak olan örütü eşleştirme yönteminde tanınacak konuşmacının sistemi kullanmadan önce birden fazla konuşmaörneğini sisteme kaydetmesi gerekmektedir. Örütü eşleştirme yöntemi çoğunlukla konuşmacı tanımada kelime sayısının kısıtlı olduğu sistemlerde kullanılmaktadır. Bir konuşmacı tanıma sisteminin öğrenme aşamasında her tanınacak kelime için birden fazla referans örütü tutulur. Örütü karşılaştırma öklid uzaklığı, kare uzaklığı ve kepstral uzaklık gibi ölçüm yöntemleri ile bulunan uzaklıkların en küçüğünü seçerek gerçekleştirilmektedir [18].

Vektör Niceleme Kaynak Modeli (Vector Quantization Source Model):

Vektör niceleme yönteminde temel amaç aynı sınıfı dahil vektörlerin birbirlerine daha çok yaklaşmasını farklı sınıfı dahil vektörlerin ise birbirlerinden uzaklaşmasını sağlamaktır. Bu yöntem, temel olarak en yakın komşular (nearest neighbours) algoritmasını kullanmaktadır. Konuşmaciya ait öznitelik vektörü çıkarma işleminden sonra vektör M sayıda bölgeye bölünüp gruplandırılarak konuşmacı modeli oluşturulur. Her birine "kod" adı verilen bu bölgeler bir araya geldiğinde "kod kitabı (codebook)" oluşturup "merkez (centroid)" ile gösterilmektedir. Sistem eğitilirken her konuşmaciya özgü kod kitabı tüm konuşmacılara ait öznitelik vektörlerinin birleştirilmesiyle oluşmaktadır. Kişiye özgü özellik vektörünü kod

$$A = [a_{ij}], \\ a_{ij} = P(q=j \mid q_{t-1}=i), \quad i, j = 1, \dots, X \quad (2)$$

kitabındaki en yakın kod kelimesine uzak olması durumuna "vektör niceleme bozulması (VQ-distortion)" denmektedir. Konuşmacı tanınırken bilinmeyen konuşmacı ve arasında minimum vektör niceleme bozulması olan kullanıcı eşleştirilir. Böylece tanıma gerçekleşmiş olmaktadır [19].

En Yakın Komşular (Nearest Neighbours):

En yakın komşular yöntemi vektör niceleme yönteminin aksine bir kod defteri üretmek için kaydedilen eğitim verilerini kümelememektedir. Bunun yerine bütün eğitim verilerini tutarak geçici olacak şekilde kullanır. En yakın komşu uzaklığı test verisi ile kaydedilmiş veriler arasındaki uzaklığın minimum olana olarak da ifade edilmektedir. Büttün test verilerinden (her birini çerçevelere bölgerek) hesaplanan en yakın komşu değeri ile eşleşme skoru üretilmektedir. Sonuç olarak bu eşleşme skoru ile benzerlik oranı yaklaşımı saptanmaktadır. En yakın komşu algoritması hesaplama yoğunluğu ve dolayısıyla hafıza gerektiren konuşmacı tanıma algoritmalarından biri olmasına rağmen en güçlü algoritma olarak bilinmektedir.

Saklı Markov Modeli (Hidden Markov Model, HMM):

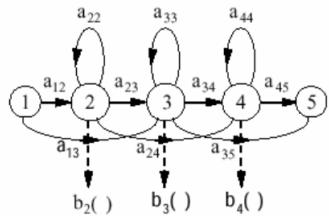
Saklı markov modeli (SMM), konuşma/konuşmacı tanıma alanında en çok kullanılan istatistiksel yöntemlerden biridir. X sayıda durum bulunan bir HMM'de önceki durum göz önünde bulundurularak oluşturulan geçiş olasılığına göre sonraki duruma geçilmektedir [20]. SMM, konuşma/konuşmacı tanıma alanında art arda gelen kısa süreli sesli ifade kesimleri için model oluşturmak ve bu modele dayanarak uzun süreli sesli ifade kesimlerini tanıma amacıyla kullanılmaktadır [21]. Ayrik bir t zamanındaki X adet durumdan birine Markov işlemi yada zinciri denmektedir. Durum değişkenleri q_t ile belirtilmektedir. Sistemin anlık durumu bir önceki duruma bağlı ise birinci dereceden markov modelinden bahsedilebilir. Durum geçiş matrisi Eşitlik 2'deki gibidir.

Sistemin ilk durumu q_0 olarak tanımlanmaktadır. Bundan sonra gelecek herhangi bir durum dizisi $q = (q_0, q_1, \dots, q_r)$ 'nın bir markov işlemi ile üretilebilme olasılığı Eşitlik 3'te gösterildiği gibi tanımlanmaktadır.

$$P(q \mid A) = \pi_{q_0} a_{q_0 q_1}, a_{q_1 q_2}, \dots, a_{q_{T-1} q_r} \quad (3)$$

Eğer durum dizisi q doğrudan gözlemlenmemiyorsa markov işlemi gizli olarak adlandırılmalıdır. Bu durumda durumların dolaylı olarak gözlemlenebilmesi söz konusudur. Bir gözlem ve bir durum arasında her ne kadar birebir bir gereklik olmazsa bile her bir

durumun Şekil 5'teki gibi belirli bir olasılıkla gözlenmesi gerekmektedir.

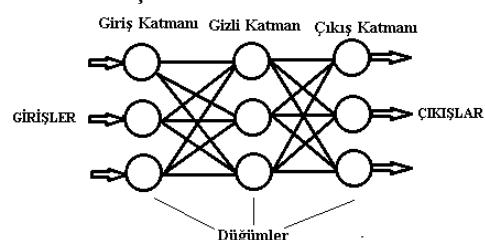


Şekil 5. 5 durumlu soldan sağa HMM örneği.

Bir konuşma tanıma sisteminde gözlemler özellik vektörleri, durumlar ise sesli ifade birimlerine denk gelmektedir. Bu durumda saklı olan durum dizisini gözlemleri kullanarak bulmak temel amaçtır. Eğer konuşma tanıma sistemi fonem tabanlı bir sistem ise, fonem dizisi durum dizisi olarak kabul edilmektedir.

Yapay Sinir Ağları (Artificial Neural Network):

İnsan beyninin çalışma prensibi üzerine kurulmuş olan Yapay Sinir Ağları yöntemi giriş ve çıkış veri kümelerini kullanarak sistem davranışını öğrenebilen yapay sistemler olarak bilinmektedir [22]. Şekil 6'da YSA'nın genel modeli verilmiştir.



Şekil 6. Yapay sinir ağının genel modeli.

Konuşma tanıma sistemlerinde kesin sınırlar belirli olmadığından YSA yaklaşımına dayalı geriye yayılma algoritması (perceptron), çok katmanlı ağlar (multilayer networks) ve kendi kendini organize eden Kohonen haritaları (Kohonen self-organizing maps) gibi yöntemlerin kullanılması uygun görülmektedir.

YSA yöntemi, eğitme ve kullanma olmak üzere iki temel aşamadan oluşmaktadır. Eğitme aşamasında düğümlerin birbirlerine bağlantılarını ifade eden yüzdeler elde edilmektedir. Bu aşamada elde edilen yüzdeler kullanma aşamasında verilen girişlere çıkışlar hesaplanması için kullanılmaktadır. Bu yüzden kullanma aşaması eğitme aşamasından daha hızlı çalışmaktadır. [23].

3.3 Ses İyileştirme

Spektral Çıkarma Algoritmaları:

Spektral çıkarma algoritması (SCA) ses sinyallerinde gürültü azaltımı için önerilen en eski algoritmalarдан biridir. Şekil 7'de görselleştirildiği gibi gürültü içeren ses sinyalinden tahmini bir gürültü spektrumunun çıkarılmasıyla tahmini gürültüsüz bir sinyal elde edilmektedir. Ses sinyali içerisinde sesli ifadenin olmadığı yerlerde gürültü spektrumu tahmin edilmiş gürültülü sinyalin fazı kullanılarak tahmin edilmiş sinyalin ters ayrik Fourier dönüşümü (inverse DFT) hesaplanmasıyla elde edilmektedir. SCA'nın, doğrusal olmayan spektral çıkarma, çoklu bant spektral çıkarma, en küçük ortalama karesel hata tabanlı spektral çıkarma, genişletilmiş spektral çıkarma, uyarlanabilir ortalama kazanımı tabanlı spektral çıkarma, seçici spektral çıkarma ve algısal özellik tabanlı spektral çıkarma gibi algoritmaları bulunmaktadır [24].



Şekil 7. SCA'nın genel yapısı.

Wiener Filtreleme:

Wiener Filtreleme (WF)'de, spektral çıkarma algoritmalarında olduğu gibi sezgisel tabanlı ilkeler bulunmamaktadır. En düşük karesel hata filtresi teorisini ilk olarak zaman uzayında Andrei Kolmogorov ve frekans uzayında Norbert Wiener geliştirmiştir [25]. Wiener filtresi katsayılarının hesaplanması için girişte kullanılacak sinyalin otokolerasyon fonksiyonunun ve çıkışta elde edilmek istenen sinyalin çapraz kolerasyon fonksiyonunun tahmin edilmesi gerekmektedir. Eşitlik 4'te Wiener filtresinin giriş-çıkış işlemi gösterilmiştir [26].

$$x(m) = \sum_{k=0}^{P-1} w_k y(m-k) \quad (4)$$

Bu denklemde;

m : zaman indeksini,

$y^T = [y(m), \dots, y(m-P-1)]$: giriş sinyalini,

$x(m)$: Wienerfiltresi çıkışını,
 $w^T = [w_0, w_1, \dots, w_{P-1}]$: Wienerfiltresi katsayı vektörünü ifade etmektedir.

İstatistiksel Model Tabanlı Yöntemler:

Ses iyileştirme yapılrken istatistiksel model tabanlı yöntemlerde sinyal büyülüğünün (ayrik fourier dönüşümü katsayıları) doğrusal olmayan kestirimini kullanılmaktadır. Doğrusal olmayan bu kestirimler, gürültüye ait olasılık yoğunluk fonksiyonunu ve konuşma sinyalinin ayrık fourier dönüşümü katsayılarını kullanarak işlem yapmaktadır. En fazla benzerlik kestirimini, Bayesian kestirimini, en az ortalama karesel hata kestirimini, yönlendirilmiş karar yaklaşımı üzerinde iyileştirmeler, müzikal görüntüyü eleme yöntemleri, logaritma tabanlı en az ortalama karesel hata kestirimini, n 'inci güç spektrumunun en az ortalama karesel hata kestirimini, Gaussian olmayan dağıtımlar tabanlı en az ortalama karesel hata kestirimini, MAP kestirimini, genel Bayesian kestirimini, algısal odaklı Bayesian kestirimini ve konuşma iyileştirmede konuşma bölgesi olmayan yerleri birleştirme gibi yöntemler istatistiksel model tabanlı yöntemler olarak bilinmektedir.

Altuzay Algoritmaları:

Ses iyileştirmede kullanılan altuzay algoritmaları çoğunlukla lineer cebir tabanlıdır. Altuzay algoritmaları temiz sinyalin gürültülü öklid uzayının bir altuzayıyla sınırlanabileceği ilkesine dayanmaktadır. Bu algoritmalar ile gürültülü sinyalin vektör uzayının, çoğunlukla temiz sinyalden oluşan bir altuzay ve çoğunlukla gürültülü sinyalden oluşan bir altuzay olarak iki ana parçaya ayrıştırılması amaçlanmaktadır. Daha sonra gürültülü altuzayda bulunan gürültü vektörünün elemanlarına "null" (boş) değer atanarak temiz sinyal tahmin edilmektedir. Gürültülü sinyalin vektör uzayının "sinyal" ve "gürültü" altuzaylarına ayrıştırılması tekil değer ayışı (TDA) ve özvektör (eigenvector) - özdeğer (eigenvalue) çarpanlara ayırma teknikleri gibi lineer cebirdeki en iyi bilinen ortogonal (dikey) matris çarpanlara ayırma teknikleri ile gerçekleştirilmektedir.

Gürültü Tahmin Algoritmaları:

Bir konuşma sinyali içerisinde gürültü tahmininin çok düşük seviyede gerçekleştirilmesiyle, ses içerisinde gürültüden kalan rahatsız edici artıkların duyulabilmesi,

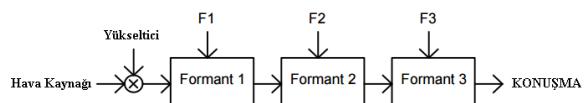
yüksek seviyede gerçekleştirilmesiyle de konuşma verisinin bozulup anlaşılabilirliğin yok olması dikkate alınması gereken iki önemli sorunu oluşturmaktadır. Bu bağlamda gürültü tahmini için en basit yaklaşımlarından biri ses hareketliliği tespiti (SHT) algoritması kullanarak, konuşmadaki her durma esnasında gürültü spektrumunu hesaplayıp güncellemektedir. Bu yaklaşım sabit ortam gürültüsü (white noise) olan ses sinyallerinde başarılı olmasına rağmen, gürültünün sürekli değiştiği daha gerçekçi ortamdaki (ör. bir kırathane'deki) seslerde tam verimli olmamaktadır.

Gürültü tahmin algoritmaları, minimal takip algoritmaları, zaman özyinelemeli ortalama algoritmaları ve histogram tabanlı teknikler gibi temel başlıklara ayrılmaktadır.

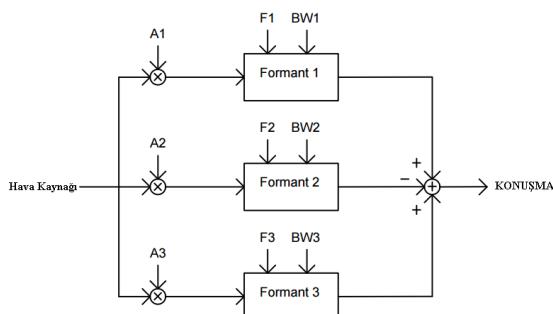
3.4 Konuşma Sentezleme

Formant Sentezleme:

Periyodik ve aperiodik sinyaller üretilip bu sinyallerin vokal yolun modeli gibi olan bir filtre veya rezonatör devresiyle beslenildiği kaynak-filtre modeli olarak tasarlanmıştır. Kabaca basitleştirilmiş bir uygulamada kaynak sinyali rasgele gürültü elemanları yüklenilmiş bir atış zinciri (impulse train) yada testere ucuna benzeyen şekilde bir sinyal olarak düşünülebilir. Bu yöntemde daha anlaşılr bir konuşma sinyali üretmek için azami 3 adet en düşük formant dikkate alınması gerekmekte olup sinyal ne kadar fazla formant içerirse o kadar yüksek kalitede olmaktadır. Vokal yol modellemesi yankılayıcının (resonators) Şekil 8'deki gibi basamak halinde (cascade) veya Şekil 9'daki gibi yatay olarak bağlanmasıyla gerçekleştirilmektedir [27]. Ses sentezleyici, formantları modelleyen yankılayıcıların yanı sıra girtlagen yapısı ve dudak titreşimlerini modelleyen filtreleri ve burundan çıkan sesleri modellemek için anti-yankılayıcılar (anti-resonators) barındırmaktadır.



Şekil 8. Basamak halinde formant sentezlemenin genel yapısı.



Şekil 9. Paralel formant sentezlemenin genel yapısı.

İfadesel Sentezleme:

İfadesel sentezlemede girtlaktan başlayıp dudakta sona eren vokal yolun kesit alanı değişimini modellemek için genelde bir dizi alan fonksiyonu kullanılmaktadır. İfadesel modelde dudak pozisyonunu, dil konumunu, akciğer hava basıncını ve ses telleri gerilimini detaylı bir şekilde ayarlamak için çok sayıda kontrol parametresi kullanılmaktadır. Modellemeye esas olarak kullanılan veriler ise doğal konuşmanın X-Ray analizi ile elde edilmektedir.

Bitiştirerek Sentezleme:

“Kes ve yapıştır” sentezlemesi olarak da bilinen bitiştirerek sentezlemede, daha önce veritabanında saklanmış ses kayıtlarından kısa bölümler kesilip yanyana eklenecek istenilen ses oluşturulmaktadır. Yapay sesin oluşturulması için gerçek ses verilerinin kullanılması teoride yüksek ses kalitesi sağladığı bilinsede pratikte başta böyle bir sistemin gerektireceği yüksek kapasitede hafiza olmak üzere birçok kısıtlama mevcuttur. Bitiştirerek sentezlemede önemli sorulardan biride kesilecek/seçilecek ses parçasının uzunluğu ne olması gerektidir. Yeteri kadar esnekliği sağlamak ve hafiza gereksinimini karşılamak için en yaygın olarak fonemler ve difonlar seçilmektedir. Bitiştirme fonem ile yapılyorsa her fonemin merkez noktasında yapılır. Bu nokta fonemin en istikrarlı/sabit noktası olduğundan birleşmenin sınırlarında meydana gelebilecek bozulmanın (distortion) en düşük olması beklenmektedir.

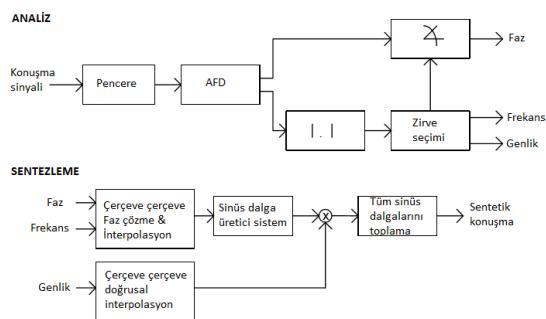
Yapay seste doğal ölçüyü yakalamak bugünkü teknikler ile imkansız olarak görülmekte olup ses parçacıkları arasındaki devamsız bölgeleri yok etmek için yoğunluğu ve temel frekans tabanlı birçok teknik önerilmektedir [28].

Sinusoidal sentezleme:

Sinusoidal sentezleme, bir konuşma sinyalin zamana bağlı değişen farklı genlik ve frekanslardaki birden fazla sinüs dalgasının birleşimiyle oluşturduğu varsayıma dayanmaktadır. Eşitlik 5'te gösterildiği gibi, sinusoidal sentezlemede $s(n)$ konuşma sinyali L adet sinusoid'in toplamı olarak modellenmektedir.

$$s(n) = \sum_{i=1}^L A_i \cos(w_i n + f_i) \quad (5)$$

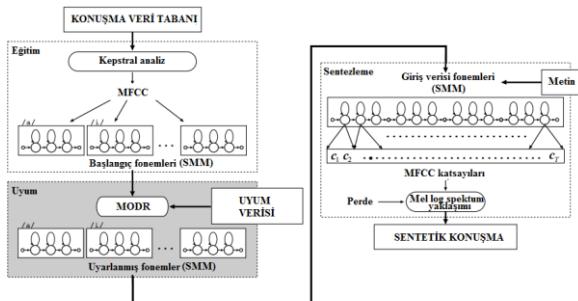
Bu denklem'de, A_i her bir sinusoidal'in sinyalin genlik değerini, f_i faz değerini ve w_i frekans değerini göstermektedir. A_i ve f_i parametrelerini bulmak için pencerelenmiş sinyal çerçevelerinin ayrik fourier dönüşümü (AFD) değeri hesaplanıp, Şekil 10'da gösterildiği gibi her çerçeveden spektral büyüğünün zirve noktası seçilmektedir [29].



Şekil 10. Sinusoidal konuşma sentezleme/analiz sistemi.

Saklı Markov Modeli Tabanlı Sentezleme:

Saklı markov modeli tabanlı bir konuşma sentezleme sistemi Şekil 11'de gösterildiği gibi genel olarak eğitim, uyum ve sentezleme olmak üzere 3 aşamadan oluşmaktadır. Eğitim aşamasında konuşma veritabanından alınan örneklerin MFCC katsayılarıyla işlem yapılmaktadır. Uyum aşamasında verilen uyum bilgisine göre özellik vektörü hesaplanmaktadır. Sentez aşamasında ise sentezlenmesi istenen giriş verisi fonem dizisine dönüştürilmektedir [30].

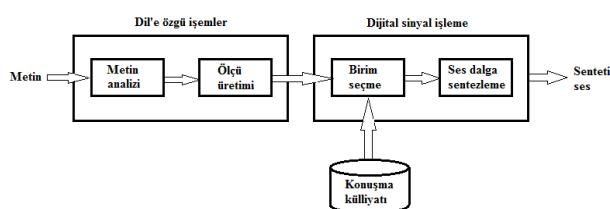


Şekil 11. SMM tabanlı konuşma sentezleme sistemi.

Birim Seçme Sentezleme:

Külliyat tabanlı sentezleme olarak da bilinen birim seçme sentezlemede geniş veritabanı kullanmaktadır. Veritabanı oluşturulurken, kaydedilmiş her konuşma örneği fonlara, hecelere, kelimelelere, ifadelere ve cümlelere bölünmektedir. Genel yapısı Şekil 12'de gösterilen birim seçme sentezleme diğer metodlarla karşılaştırıldığında sentetik konuşmada daha çok doğallık sağladığını görülmektedir [31].

Birim seçme sentezlemede birim uzunluğunun seçimi önemli bir konu olarak görüldüğünden bu konu üzerinde birçok çalışma bulunmaktadır. Kısa birim uzunluğu seçimi daha az hafıza gerektirirken örnek toplayıp etiketleme daha zor ve karmaşık hale gelmektedir. Uzun birim seçimi, bölümme noktaları çok olmadığı için daha doğal ve eklemelik etkilerden uzak iken daha fazla hafıza gerektirmektedir [32].



Şekil 12. Birim seçme sentezleme sistemi.

4. Adli Açıdan Ses İncelemeleri

Adli Ses İnceleme Konuları:

Ses incelemelerinde kullanılabilen yazılımlarda sayısal ses işleme teknikleri tabanlı yöntemler uygulanmaktadır. Kullanılan bu yöntemlerin adli ses inceleme yöntemi olarak sayılabilmesi için adli niteliğe sahip olması gerekmektedir. Örneğin bir metnin konuşmaya çevrilmesi (ses sentezleme) ve bir müzik kaydından vokal seslerin alınıp sadece müzik

aletlerine ait seslerin ortaya çıkarılması gibi işlemler ses işleme yöntemi olarak görülebilir ancak adli niteliğe sahip değildir. Ses inceleme yazılımlarının adlı açıdan değerlendirilebilmesi için bu yazılımların aşağıdaki özelliklerini barındırması beklenmektedir.

Konuşma Çözümleme (Transkripsiyon):

İncelenenek ses kaydı üzerinde kayıtta geçen konuşmaların yazıya dökülmesi işlemi olarak tanımlanmaktadır. Bu çözümleme, kullanıcının ses kaydını segmentlere ayırip her bir segmente karşılık o dildeki ifadeleri (fonem, hece, harf, kelime vs.) segment ile birlikte kaydetme işlemidir. Konuşma kaydı çözümleme işleminden sonra kayıt içerisinde istenilen ifadenin aratılması, konuşma kaydının dökümünün alınması gibi işlemler mümkün hale gelmektedir.

Kayıt Bütünlüğünün Doğrulanması:

Adli vakalarda delil olarak kullanılabilen ses kayıtları bütünlük açısından değerlendirildiğinde dikkate alınması gereken noktalar, o ses kaydının tek bir parça halinde olması ve üzerinde herhangi bir düzenleme yapılmamış olmasıdır. Günümüzde ses kayıtları ücretsiz birçok yazılım ile kolayca düzenlenebilmekte ve ortaya yeni bir ses kaydı çıkarılabilirmektedir. Bir ses inceleme yazılımının adlı ses inceleme yazılımı olarak kullanılabilmesi için adli vakalara konu olmuş ses kayıtları üzerinde kurgu/manipülasyon tespiti yapıp kayıt bütünlüğünü doğrulayabilmesi gerekmektedir.

Otomatik Ses İzi Eşleştirme:

Otomatik ses izi eşleştirme, içerik olarak niteliği belirsiz olan ses kaydının bilinen bir ses kaydı ile eşleştirilmesi işlemi olarak tanımlanmaktadır. Birçok problem adlı açıdan otomatik ses izi eşleştirme analizi ile çözülmektedir. Örneğin incelenenek ses kaydında bulunan bir silah sesine ait ses izi üzerinden silahın marka/model tespitinin yapılması veya ses kaydındaki konuşmacıların tehis edilmesi mümkün hale getirebilmektedir.

Spektrum Analizi:

Konuşma sinyalleri kendi içerisinde farklı frekanslarda bulunan birçok sinyal barındırdığı için karmaşık sinyaller olarak bilinmektedir. Konuşma sinyalleri açısından spektrum, konuşma sinyalindeki frekans değerlerinin zamanla değişimi anlamına gelmektedir. Spektrum analizi, zamana göre frekans

yoğunluklarının renklendirilmiş grafikle gösterildiği ve sesin görselleştirilmiş hali olarak da bilinen spektrogram üzerinden gerçekleştirmektedir. Bir ses kaydındaki konuşmada kulak yoluyla anlaşılmayan hece, harf, kelime vs. birimlerinden herhangi biri varsa bu birimlerin formant frekans değerleri göz önünde bulundurularak anlaşılmayan yerler belirlenebilmektedir.

Otomatik Formant Frekansları Tespiti:

Formant, ses dalgası enerjisinin belli bir frekansta yoğunlaşması olarak tanımlanmaktadır. Formant frekansı ise bu yoğunlaşmanın gerçekleştiği frekans değerini belirtmektedir. Her insanın hatta ses dalgası yayan her nesnenin ses yolu farklı olduğu için ürettiği ses dalgası içerisindeki formant frekansları farklıdır. Sesli harflerin seslendirilmesi sırasında oluşan kişiye özgü formant frekanslarının tespiti, spektogram üzerinden görsel olarak yapılmaktadır. Formant frekans değerlerinin otomatik olarak belirlenmesi, seslendirme yapılrken formant frekans değerlerinin sesli harflerden önce ve sonra gelen harflerden etkilenmemesi açısından önem taşımaktadır.

Kayıt İyileştirme (Filtreleme):

Adli ses inceleme yöntemlerinden biride ses kaydının kalitesini ve anlaşılabilirliğini artırmak için yararlanılan kayıt iyileştirmektedir. Kayıt iyileştirme ile ses kaydı yapılan cihazdan kaynaklanan gürültüler önemli ölçüde temizlenmektedir. Gürültü azaltma ve sesli kısımları vurgulama gibi kayıt iyileştirme işlemlerinden sonra konuşma çözümleme, konuşmacı teşhis ve tanıma, konuşmacı profili belirleme gibi işlemler daha başarılı sonuç vermektedir.

Konuşmacı Profili Belirleme:

Konuşmacı profili belirleme işleminde kişinin yaş, cinsiyet ve sağlık gibi fiziksel özelliklerini ortaya çıkarmaktadır. Bu işlemde işitsel yol ile yapılan analizlerle birlikte adli ses inceleme yazılımının fiziksel özelliklerini otomatik tespit edebilmesi sağlıklı sonuç elde etmede başarı sağlamaktadır. Dünyada birçok hukuk sisteminin önemle üzerinde durduğu sorunlardan olan çocuk istismarı, tehdit ve şantaj gibi vakalarda yararlanılacak herhangi bir ses delilinde yaş ve cinsiyet tespiti yapılması önem arz etmektedir.

Birden Çok Konuşmacıyı Ayırt Etme:

Adli vakalarda dijital delil olarak kullanılan ses kayıtlarının incelenmesi sırasında, bu kayıtlarında geçen konuşmalarda işitsel olarak birbirlerine benzeyen birden fazla konuşmacının var olması bu konuşmacıların seslerinin yazılım yol ile ayırtılması gerekliliğini ortaya çıkarmaktadır. Birden çok konuşmacı varlığının söz konusu olduğu bir ses kaydı üzerinde belirli kişiye ait sesin mevcut olup olmadığını tespiti, mevcut ise bu kişinin seslendirdiği zaman aralığının belirlenmesi ve ses kaydındaki konuşmacıların seslendirdiği bölümlerin ayrı ayrı segmente edilmesi işlemleri gerçekleştirilebilmektedir.

Maskelenmiş Ses Analizi:

Ses maskeleme, bir ortamda kayıt altına alınan konuşmaların yüksek müzik veya televizyon sesleriyle fiziksel maskelemeye maruz kalması olarak bilinmektedir. Özellikle yasal dinlemelerde, dinlendiğinden şüphelenen kişi tarafından kaydın kalitesini bozacak maskeleme yapılyorsa bu ses kaydına gerçek zamanlı veya sonradan maskelenmiş ses analizi uygulanıp ses kaydında sadece konuşmanın olduğu kısımlar vurgulanabilmektedir.

Gerçek Zamanlı Ses Analizi:

Adli ses analizleri, analizin yapıldığı zaman açısından, sesin kayıt altına alındığı esnada ve ses kaydı alındıktan sonra analiz yapılması şeklinde temel olarak ikiye ayrılmaktadır. Gerçek zamanlı ses analizi, ses kaydının olduğu esnada anlık analiz yapılması işlemidir. Bu tür analizde kayıt cihazı yazılıma bağlanıp yazılım ve donanımın eşzamanlı çalışması sağlanmaktadır. Kayıt iyileştirme, spektrum analizi, maskelenmiş ses analizi gibi işlemlerin gerçek zamanlı analiz yoluyla yapılması konuşmayı net anlaşırlır hale getirmekte ve konuşmacı tanımayı kolaylaştırmaktadır.

GSM Gürültüsü Yok Etme:

IS-95 (Interim Standard 95) ile birlikte günümüzün en çok kullanılan hücresel iletişim teknolojilerinden biri olan GSM (Global System for Mobile) [33], kendi sinyallerine yakın olan elektronik bir cihaz üzerinde parazit oluşturabilmektedir. Cep telefonları tarafından üretilen bu titreşim sesi, cep telefonu yakınındaki bir ses kayıt cihazı ile alınan sesi anlaşılmaz hale getirerek adli ses incelemeyi zorlaştırmaktadır. Bu nedenle ses kayıtlarında bulunabilen GSM

gürültülerinin temizlenmesi, bu kayıtların daha sağlıklı bir şekilde analiz edilmesine olanak sağlamaktadır.

Dil / Aksan Tanıma:

Küresel güvenlik açısından terörizmle mücadele ve istihbarat alanında dil ve aksan tanımanın önemi gittikçe artmaktadır [34]. Dil/Aksan tanıma analizi, incelemeye konu olan ses kaydındaki konuşmacıların hangi ülkeden ve daha da ötesi hangi yörenen olduğunu tespit etmeye yönelik çalışmaları içermektedir. İşitsel yol ile yapılan dil/aksan tanımanın otomatik dil/aksan tanıma ile gerçek zamanlı ses analizinde bütünlendirilip yapılması adli ses analistlerine başta zaman yönetimi olmak üzere her açıdan fayda sağlamaktadır.

Adli Ses İncelemelerinde Zorluklar:

Ses delillerinin incelenmesinde diğer delil türlerinde (disk, ram, cep telefonu vs.) olduğu gibi bir takım zorluklarla karşılaşılabilir. Adli ses incelemelerinde:

- Mevcut ses inceleme yazılımlarının, otomatik ses analizlerinde eksiklikler barındırması nedeniyle adli ses incelemelerinin çoğunlukla incelemecinin işitsel kabiliyetine bağlı olarak gerçekleştirilmesi,
- Dijital ses kayıtları üzerinde yazılımsal olarak yapılan manipülasyonların analog kayıtlar üzerinde yapılan fiziksel değişikliklere göre daha zor tespit edilmesi,
- Ses benzerlikleri yoluyla ses sahteciliği,

- Alışkanlık (sigara, alkol vs.), yaş, ve hastalık gibi etmenlerin ses üzerinde doğrudan değişiklik etkisinin bulunması

gibi zorluklarla karşılaşılmaktadır.

Adli Ses İnceleme Yazılımları:

Ses incelemeleri için kullanılan yazılımlarda sayısal ses işleme tabanlı teknikler kullanılmaktadır. Bu yazılımların adlı açıdan değerlendirilebilmesi için adli ses inceleme konularını uygulayabilmesi beklenmektedir. Bu makalede, önceki bölümde anlatılan adli ses inceleme konuları ve bazı ilave performans kriterleri göz önünde bulundurularak adli ses inceleme yazılımlarının değerlendirilmesi gerçekleştirılmıştır. Ses incelemelerinde kullanılabilecek birçok yazılım mevcuttur. Bu makalede, DC/Live Forensics (A.B.D.), SIS II (Rusya), SESTEK (Türkiye), Acu-Expert Audio Forensic (İrlanda), CEDAR Audio Forensic (İngiltere), Audacity, PhonEdit (Fransa), SFS/WASP (İngiltere), Agnitio SIFT (İspanya), SIL Speech Analyzer (A.B.D.), Praat (Hollanda), UCL Enhance (İngiltere), CoolEdit (A.B.D.), Acoustica 7 (Norveç), WaveSurfer (İsviçre), Adobe Audition (A.B.D.), IKAR Lab (Rusya), LingWaves (Almanya), WinPitch 10 (Kanada), TrueRTA (A.B.D.), GoldWave (Kanada), WavePad (A.B.D.), Raven (A.B.D.), Sound Forge Pro 11 (Almanya), SoundRuler (A.B.D.), SpectraPLUS (A.B.D.), QuickEnhance Plug-in (A.B.D.) ve Foenics (Almanya) yazılımları detaylı bir şekilde incelenmiştir. Bu yazılımların karşılaştırımları Şekil 13'te gösterilmiştir.

		YAZILIM																												
		DEĞERLENDİRME																												
Teknik	Performer	DC Forensics	STC SIS II	SESTEK	Act-Exp. Audio	Forensic Audio	CEDAR Cambr.	Audacity	PhoneEdit	SFSWASP	Agnitio SIFT	Speech Analyzer	Praat	UCL Enhance	CoolEdit	Acoustica 7	WaveSurfer	Adobe Audition	IKAR Lab	lingWaves	WinPitch 10	TrueRTA	GoldWave	WavePad	Raven	Sound Forge	SoundRuler	SpectralPLUS	QE Plug-in	C.T.FORENICS
Konuşma çözümleme (transkripsiyon)	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-		
Kayıt bütünlüğünün doğrulanması	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-		
Otomatik ses izi eşleştirme (Konuşmacı teşhis ve tanıma)	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-		
Spektrum analizi	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-		
Otomatik formant frekansları tespiti	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-		
Kayıt iyileştirme (filtreler)	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-		
Konuşmacı profili (yas, cinsiyet) belirleme	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-		
Birden çok konuşmacıyı ayırt edebilme	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-		
Maskelenmiş ses analizi	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-		
Gerçek zamanlı ses analizi	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-		
GSM gürültülerini yok etme	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-		
Dil-Aksan tanıma	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-		
Lisans gereksinimi	Ücretli												-	-	-	-	-	-	-	-	-	-	-	-	-	-	-			
Ücretli	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-		
Ücretsiz	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-		
USB Lisansı (dongle)	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-		
Eklenti desteği	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-		
Platform	Windows												-	-	-	-	-	-	-	-	-	-	-	-	-	-	-			
Windows	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-		
Linux	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-		
MAC	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-		
Backup	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-		
Ağ/Bulut'tan dosya aktarımı ve yönetme	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-		
Arayüz kişiselleştirme	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-		
Önceden kaydedilmiş (preset) yüklenenabilir ayarlar	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-		
Gelişmiş klavuz ve yardım	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-		
Raporlama desteği	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-		

Şekil 13. Ses incelemelerinde kullanılan yazılımların adlı açısından karşılaştırılması.

5. Sonuç ve Öneriler

Bu makalede, mevcut durumda yapılan ses incelemelerinin ses adlı bilişim (audio forensics) açısından değerlendirilmesi yapılmıştır. Konuşma fizyolojisi, işitme fizyolojisi, ses bilimi konuları detaylı bir şekilde açıklandıktan sonra ses incelemede kullanılan tekniklere ön işlemler, konuşma/konuşmacı tanıma, ses iyileştirme ve konuşma sentezleme gibi alt başlıklar bazında deñinmiştir. Yakın geçmişte dijital delil incelemeleri tek bir incelemeci tarafından yapılabiliyorken, inceleme alanlarının genişlemesi ile birlikte konusunda uzman incelemeçilere ve araçlara olan ihtiyacın artması

6. Kaynaklar

1. Kara, O. K., 2011. Adli amaçlı ses analizinde otomatik konuşmacı tanıma yazılımı kullanılarak kardeşler arası ses benzerliklerinin ve farklılıklarının belirlenmesi, Yüksek Lisans Tezi, İstanbul Üniversitesi Adli Tıp Enstitüsü, İstanbul.
2. M.E.B., Müzik aletleri yapımı ses oluşum uygulamaları, Ankara, 2006.
3. Makkook, M. A., 2007. A Multimodal Sensor Fusion Architecture for Audio-Visual Speech Recognition, Master Thesis, University of Waterloo, Canada.
4. Campbell, J. P., 1997. Speaker Recognition: A Tutorial, Proceedings of the IEEE, 85, 1437-1462.
5. Parsons, T. W., 1987. Voice and Speech Processing, McGraw-Hill, New York.
6. Pentz, A., 1990. Speech science (Spath 4313) class notes, Oklahoma State University, Stillwater.
7. Ergenç, İ., 1995. Konuşma Dili ve Türkçenin Söyleyiş Sözlüğü, Şafak Matbaacılık, Ankara.
8. Sever, H., 2008. Adli Ses İncelemeleri ve Hukuki Boyutu, Adalet Yayınevi, Ankara.
9. Selen, N., 1979. Söyleyiş sesbilimi Akustik sesbilimi ve Türkiye türkçesi, TDK yayınları, Ankara.

sonucu ortaya çıkan ses adlı bilişimde yer alan konuşma çözümleme, kayıt bütünlüğünün doğrulanması, konuşmacı profili belirleme ve kayıt iyileştirme gibi çalışmalarдан bahsedilmiştir. Bu çalışmada son olarak adlı ses incelemelerinde kullanılabilen yazılımlar teknik ve performans açısından değerlendirilip sonuçlar bir karşılaştırma tablosunda gösterilmiştir. Bu karşılaştırmanın en önemli amacı bir adlı ses uzmanına ses analizi yaparken hem analizin doğru ve güvenilir olmasını sağlamak hem de süreçleri optimum seviyede gerçekleştirmek için ihtiyaca uygun yöntem ve araçları seçme konusunda yardımcı olmaktadır.

10. Artuner, H., 1994. Bir Türkçe Fonem Kümeleme Sistemi Tasarımı ve Gerçekleştirimi, Doktora Tezi, Hacettepe Üniversitesi Fen Bilimleri Enstitüsü, Ankara..
11. Demircan, Ö., 1979. Türkiye Türkçesinin Ses Düzeni Türkiye Türkçesinde Sesler, TDK yayınları, Ankara.
12. Mengüoğlu, E., 2004. Confidence Measures for Speech/Speaker Recognition and Applications on Turkish LVCSR, PhD Thesis, University of Mons, Belgium.
13. <https://www.biyolojigunlugu.com/lys-kulak-konu-anlatimi-kulak-islevi-kulak-yapisi-kulak-bilesenleri/>, YS, Kulak Konu Anlatımı, Kulak İşlevi, Kulak Yapısı, Kulak Bileşenleri, 15 Eylül 2017.
14. Flanagan, J. L., 1972. Speech Analysis; Synthesis and Perception, Springer, U.S.A..
15. Xie, L. and Liu, Z. Q., 2006. A Comparative Study of Audio Features for Audio to Visual Conversion in MPEG-4 Compliant Facial Animation, Proceedings of the 5th International Conference on Machine Learning and Cybernetics, Dalian, August 13-16.
16. Dave, N., 2013. Feature Extraction Methods LPC, PLP and MFCC in Speech Recognition, International Journal for Advance Research in Engineering and Technology, 1, Issue VI.
17. Kavitha. R., Nachammai. N., Ranjani. R. and Shifali. J., 2014. Speech Based Voice Recognition System for Natural Language Processing (Voice Recognition using Dynamic Time Warping and Mel-Frequency Cepstral Coefficients Algorithms), International Journal of Computer Science and Information Technologies, 5, 5301-5305.
18. Gaudard, C., Aradilla, G. and Bourlard, H., 2007. Speech Recognition based on Template Matching and Phone Posterior Probabilities, IDIAP, Switzerland.
19. Karasartova, S., 2011. Metinden Bağımsız Konuşmacı Tanıma Sistemlerinin İncelenmesi ve Gerçekleştirilmesi, Yüksek Lisans Tezi, Ankara Üniversitesi Fen Bilimleri Enstitüsü, Ankara.
20. Gelegin, İ. ve Bolat, B., 2011. Ayırık Kelime Tabanlı Bir Konuşma Tanıma Sistemiyle Bilgisayar Kontrolü, Elektrik-Elektronik ve Bilgisayar Sempozyumu, Elazığ, 5-7 Ekim.
21. Yalçın, N., 2008. Konuşma Tanıma Teorisi ve Teknikleri, Kastamonu Eğitim Dergisi, 16, 249-266.
22. Rumelhart, D.E. and McClelland, J.L., 1986. Parallel Distributed Processing, MIT Press, Cambridge, 10-16.
23. Akçayol, M.A., 2001. Bir Anahtarlamalı Relüktans Motorun Sinirsel-Bulanık Denetimi, Doktora Tezi, Gazi Üniversitesi Fen Bilimleri Enstitüsü, Ankara.
24. Loizou, P.C., 2013. Speech Enhancement Theory and Practice, CRC Press, U.S.A.
25. Vaseghi, S.V., 2008. Advanced Digital Signal Processing and Noise Reduction, Wiley Publication, U.K.
26. Kutlu, C. ve Arserim, M. A., 2012. Spektral çıkarma tabanlı kalmanfiltresi ile ses sinyallerinin iyileştirilmesi, Dicle Üniversitesi Mühendislik Fakültesi, 3/1, 49-59.
27. Lemmetty, Sami., 1999. Review of Speech Synthesis Technology, Master Thesis, Helsinki University of Technology, Finland.
28. http://www.cs.tut.fi/courses/SGN-4010/puhesynteesi_en.pdf, Speech Synthesis. 7 Haziran 2017.
29. Macon, M. W., 1996. Speech Synthesis Based on Sinusoidal Modeling, PhD Thesis, Georgia Institute of Technology, U.S.A.
30. Tamura, M., Masuko, T., Tokuda, K. and Kobayashi T., 1999. Speaker Adaptation for HMM-Based Speech Synthesis System Using MLLR, The 3rd ESCA/COCOSDA Workshop on Speech Synthesis, Australia, November 26-29.
31. Kayte, S., Mundada, M. and Kayte C., 2015. A Review of Unit Selection Speech Synthesis, International Journal of Advanced Research in Computer Science and Software Engineering, 5, 5.
32. Gros, J. Z. and Mario, Z., 2008. An Efficient Unit-selection Method for Concatenative Text-to-speech Synthesis Systems, Journal of Computing and Information Technology, 16, 69-78.
33. <http://www.ni.com/tutorial/7107/en/>, IS-95 (CDMA) and GSM(TDMA) Overview, 4 Temmuz 2017.
34. <http://www.forensiclinguistics.net/POST-PN-0509.pdf>, Houses of Parliament, Forensic Language Analysis, 5 Temmuz 2017.