

Araştırma Makalesi

Ses Sinyalinde Gürültü Saptama İçin Özgün Bir Yaklaşım

Ramazan Çolak¹, Rafet Akdeniz^{2,*}

¹ TRT İstanbul Televizyonu, Ortaköy, İstanbul

² Elektronik ve Haberleşme Mühendisliği Bölümü, Çorlu Mühendislik Fakültesi, Tekirdağ Namık Kemal Üniversitesi, Tekirdağ, Türkiye

Geliş: 14.09.2018

Kabul: 09.11.2018

Özet: Ses etkinliği algılama, genelde konuşma tanıma, konuşma sıkıştırma gibi konuşma işleme süreçlerinin başında kullanılan bir uygulamadır. Konuşma sesinin olup olmadığını tespit etmede kullanılır ve buna göre uygulamanın devamına yön verir. Sesin varlığını tespit etmede kullanılan belli başlı özellikler vardır. Kullanılan özelliklerin fazla olması algoritmanın verimliliği ile doğrudan ilişkilidir. Klasik VAD algoritmaları genelde STE kullanılarak oluşturulduğundan, düşük sinyal gürültü oranı değerlerinde çok hassastır, bu yüzden istenilen sonuçları veremeyebilir. Çözümde kullanılan özellikler için gerçek zamanlı sesler kullanarak sesli bölge ve gürültülü bölge ayrt edilmeye çalışılmıştır. Bu çalışmada sesin varlığını tespit etmek için STE, periyodiklik ve Spektral düzlük gibi üç özellik kullanılmıştır, kullanılan bu üç özellik ile düşük SNR değerlerinde de istenilen sonuçlar elde edilmiştir. Bu yöntemin, özellikle düşük SNR değerlerinde klasik metotlara göre daha iyi performans elde ettiği gözlemlenmiştir.

Anahtar kelimeler: En baskın frekans bileşeni, Kısa zamanlı enerji, Spektral düzlük, Ses etkinliği algılama

A Novel Approach to Noise Reduction in Audio Signal

Abstract: Voice activity detection is often used at the beginning of speech processes, such as speech recognition, and speech compression. It is used to detect the presence of a speaking voice, and it directs the execution of the application accordingly. There are certain features that are used to detect the presence of the voice. The efficiency of the algorithm is directly related to the number of the features used. Since the classical Voice activity detection algorithms are usually developed using Short Time Energy, they are very sensitive to the low signal-to-noise ratio values, therefore they may not provide the desired results. The parts with a speech and the parts with noise were attempted to be distinguished by using real-time sounds for the features used in the solution. In this study, three features, such as Short Time Energy, Periodicity, and the Spectral Flatness, were used to detect the voice. The desired results have been obtained by using these three features, even at low SNR values. This method has been observed to achieve better performance especially at low SNR values than conventional methods.

Keywords: Most dominant frequency component, Short-time energy, Spectral flatness, Voice activity detection

Received: 14.09.2018

Accepted: 09.11.2018

* Sorumlu yazar.

E-posta adresi: rakdeniz@nku.edu.tr (R. Akdeniz)

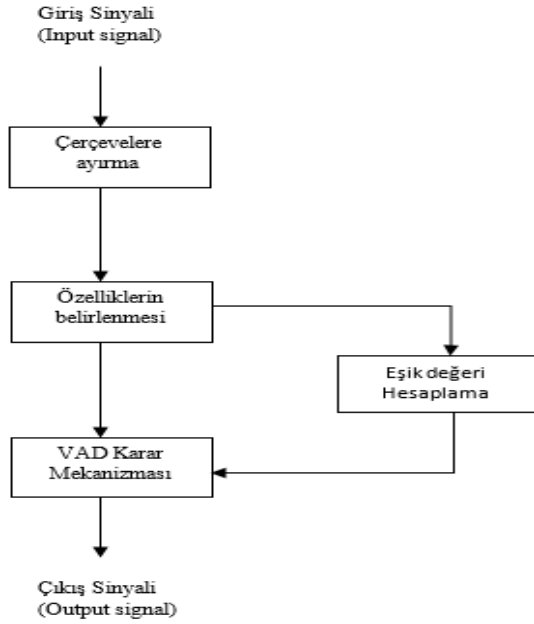
1. Giriş

Temel prensip olarak VAD (Voice activity detection - Ses etkinliği algılama), ses sinyalinin özelliklerinin çıkarılması ve bu özelliklerin ses sinyalinin yalnızca gürültülü bölümlerinden belirlenen eşik değeri ile karşılaştırılmasıyla sesli bölümü ve gürültülü bölümü tespit etmeye yarayan bir uygulamadır. Bu uygulama; konuşma iyileştirme (speech enhancement), konuşma tanıma (speech recognition) ve konuşma kodlama (speech coding) gibi sinyal işleme süreçlerinin ilk aşamasında kullanılır ve sürecin performansını doğrudan etkiler.

VAD algoritmalarında, algoritmanın verimliliği, sağlamlığı, basitliği gibi kavramlar algoritmanın kullanılabilirliği ile doğrudan bağlantılıdır. Sesin varlığını ya da yokluğunu tespit etmek için kullanılan özelliklerden bazıları aşağıda sıralanmıştır:

- STE (short time energy - kısa süreli enerji),
- Periodicity (periyodiklik),
- ZCR (zero crossing rate - sıfır geçiş oranı),
- SF (spectral flatness - frekans spektrumundaki düzlük),
- Most dominant frequency component (en baskın frekans bileşeni),
- High-low frequency rate (yüksek-alçak frekans oranı).

Şekil 1'de temel bir VAD'ın blok şeması gösterilmiştir.



Şekil 1. VAD Blok Şeması

STE özelliği yalnız başına özellikle düşük SNR değerlerinde istenilen sonuçları vermemektedir, bu nedenle STE'nin dışında frekans düzleminde SF ve most dominant frequency component kullanılarak düşük SNR değerlerinde daha iyi sonuçlar veren bir algoritma geliştirilmiştir [1]. K. Sakhnov ve arkadaşları tarafından geliştirilen farklı bir algoritmada ise, sesin periyodik olma özelliği ve yüksek-düşük frekans oranı kullanılarak verimli bir VAD algoritması geliştirilmiştir [2]. K. Sakhnov ve E. Verteletskaya tarafından geliştirilen başka bir algoritmada ise periyodik olma özelliği, yüksek düşük frekans oranının yanı sıra STE özelliği de kullanılarak

yeni uygulanması kolay bir algoritma oluşturulmuştur [3]. Peak valley difference (maximum minimum farkı) olarak isimlendirilen farklı bir özellik ile birlikte STE, maksimum frekans bileşeni ve SF özellikleri kullanılarak [3]'e benzer bir algoritma ile yeni bir VAD algoritması sunulmuştur [4]. Bir başka çalışmada zaman düzleminde ve frekans düzleminde VAD kavramları kullanılarak lineer enerji temelli detektör, fuzzy logic ve yapay sinir ağları temelli VAD performansları karşılaştırılmıştır [5]. T.H. Zaw ve N. War tarafından yapılmış olan bir çalışmada; STE, ZCR, spectral entropy ve linear prediction error (doğrusal tahmini hata) gibi özellikler kullanılarak ses sinyalinin her çerçevesi için bir D bileşik parametresi hesaplanarak ve her ses çerçevesinin D/D_{max} olarak belirlenen eşik seviyesinin üstünde olup olmamasına göre konuşmanın varlığı tespit edilmiştir [6]. Kulaklık teknolojilerinde de gürültü azaltma teknolojileri giderek artan bir şekilde kullanılmaya başlanmıştır. N. Lezzoum ve arkadaşları tarafından yapılan bir çalışmada akıllı kulaklık sistemleri tarafından kullanılan bir VAD algoritması geliştirilmiştir. Bu algoritmada kulaklıkta bulunan harici bir mikrofon ile dış ortam sesi toplanmış ve frekans düzleminde 3 banda ayrılıp insan sesinin frekans aralığı göz önünde bulundurularak enerji temelli bir algoritma oluşturulmuştur [7]. M. Kumari ve I. Ali tarafından yapılan bir çalışmada ise KVD olarak isimlendirilen ses sinyalinin örneklerinin standart sapmaya bağlı olarak altı farklı kümeye ayrılması ve en düşük değerin bulunduğu kümenin belirlenen bir eşik değeriyle karşılaştırılması ile sinyalin konuşma olup olmadığını belirleyen farklı bir VAD algoritması geliştirilmiştir [8]. A. Pasad ve arkadaşları tarafından yapılan bir çalışmada enerji temelli ve harmonik temelli iki özellik kullanılarak otomatik konuşma tanıma (Auto Speech Recognition-ASR) uygulamalarında kullanılan bir VAD algoritması tasarlanmıştır [9]. İnsan sesinin düşük frekansındaki enerjinin yüksek frekanstaki enerjisinden yüksek olduğundan hareketle frekans spektrumunun düşük frekans bandındaki enerjisinden yüksek frekans bandındaki enerjisinin çıkarılmasıyla sesli bölümü tespit eden ve -5 dB SNR değerine kadar iyi sonuçlar veren bir VAD algoritması geliştirilmiştir [10]. I. Almajai ve B. Milner tarafından yapılan bir çalışmada bir konuşma sinyalinin hem ses özellikleri hem de sinyalin görsel özellikleri kullanılarak hem temiz ses sinyalinde hem de gürültülü ses sinyalinde güvenilir bir VAD algoritması geliştirilmiştir [11].

Bu çalışmada ise gürültülü ortamda gürültülü konuşma ve sadece gürültülü çerçeveleri ayırt etmek için; STE, periyodiklik ve SF özelliklerini kullanarak uygulanması kolay bir VAD algoritması geliştirilmiştir. II. bölümde kullanılan özellikler, klasik enerji temelli metotlar ve III. bölümde önerilen algoritma detaylarıyla açıklanmıştır. IV. bölümde deneysel sonuçlar bulunup klasik metotlarla verimlilikleri karşılaştırılmıştır.

2. Short time (kısa süreli) Özellikler

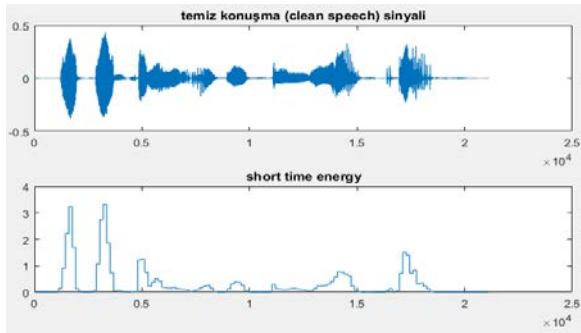
Konuşma sinyallerinde yapılan işlemler, seslerin birbirlerine benzerlik göstermesinden ötürü çok küçük zaman aralıklarında yapılır. Önerilen algoritmada her ses çerçevesi (frame) için üç farklı özellik kullanılmıştır.

2.1 Short Time Energy (STE-Kısa Süreli Enerji)

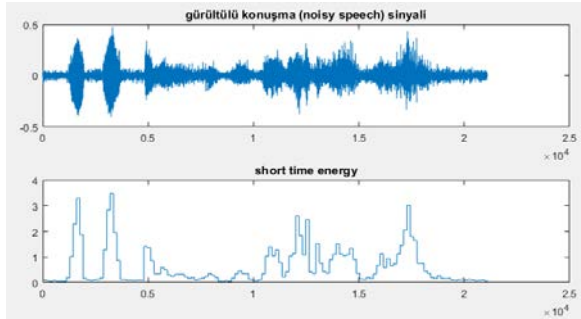
Konuşma sinyallerinde zamana bağlı olarak konuşmanın genliği değişiklik göstermektedir. Konuşmanın olduğu bölgelerde STE, konuşmanın olmadığı bölgelere göre daha yüksektir. Bu da bize konuşma bölgeleri ile ilgili önemli ipuçları verir. STE'nin formülü aşağıdaki gibidir:

$$E(n) = \sum_{i=1}^N X_n^2(i) \quad (1)$$

Burada N , Ses çerçevesinin uzunluğunu, $X(i)$; Orijinal konuşma sinyalini ve $E(n)$ ise ses çerçevesinin enerjisini göstermektedir. Görüldüğü üzere her ses örneğinin karelerinin toplamı alınarak, bir sesli çerçevesinin enerjisi hesaplanabilir. Şekil 2 ve şekil 3'te sırasıyla temiz bir konuşma sinyali ve bu konuşma sinyalinin STE'si ve gürültülü bir konuşma sinyali ve onun STE'si gösterilmiştir.



Şekil 2. Temiz bir konuşma sinyali ve STE'si



Şekil 3. Gürültülü bir konuşma sinyali ve STE'si

2.1.1 Lineer enerji temelli VAD (LED)

LED (Linear Energy Based Detector-Linear enerji temelli dedektör), ses sinyalinin her bir çerçevesinin belirli bir STE eşik seviyesi ile karşılaştırılması ve buna göre çerçevesinin sesli ya da gürültülü olduğuna karar verilen bir VAD algoritmasıdır.

$$\begin{aligned} \text{Eğer } (E_j > K X E_r), & \quad \text{Çerçeve: Sesli} \\ \text{Değilse,} & \quad \text{Çerçeve: Gürültülü} \end{aligned} \quad (2)$$

Burada, K ölçekleme faktörü ($K > 1$), E_j çerçeve enerji seviyesi ve E_r ise gürültülü çerçeve enerji seviyesi KXE_r ifadesi çerçevesinin sesli ya da gürültülü olduğunu belirlemek için kullanılan eşik değeridir. Farklı LED algoritmalarında farklı şekilde gürültü eşik değeri güncellenmektedir. Gürültülü çerçevesinin güncellenme formülü aşağıdaki şekilde

gösterilmiştir [12].

$$E_{ryeni} = (1-p) \cdot E_{reski} + p \cdot E_{noise}, \quad (0 < p < 1) \quad (3)$$

Burada E_{ryeni} güncellenmiş enerji eşik seviyesidir. E_{reski} ise önceki enerji eşik seviyesidir. E_{noise} ise o anki gürültülü çerçevesinin enerjisidir. p parametresi ise sıfır ile bir arasında seçilen bir sabittir. Gelen çerçeve gürültü eşik seviyesinin altında ise eşitlik (3) aracılığıyla eşik seviyesi en son gelen gürültünün enerji seviyesine yakın olacak biçimde güncellenir.

2.1.2 Adaptif Lineer Enerji Temelli VAD (ALED)

Bu metod bir önceki LED metoduna benzer bir metottur. LED metodunda gelen enerji sinyali E_j , gürültü eşik değerinin altında kaldığında gürültü eşik değeri E_r , eşitlik (3) ile güncellendiği yukarıda bahsedilmiştir. ALED uygulamasında da gürültü eşik değeri eşitlik (3) ile belirlenir. LED uygulamasından farklı olarak LED algoritmasında p sabit bir değer alırken, ALED uygulamasında o anki ve önceki gürültülü çerçevelerin enerjilerinin değişimi oranlarına göre değişir [13]. Bu algoritmada öncelikle tüm ses çerçevesinin enerjilerinin değişimleri hesaplanır.

$$\sigma^2 = \text{var}(E_j) \quad (4)$$

Daha sonra gelen ses çerçevesinin gürültülü konuşma ya da gürültü olduğu belirlenir. Eğer gelen ses çerçevesi gürültü ise o ses çerçevesinin enerjisinin değişimi ile bir önceki gürültü olarak belirlenen ses çerçevesinin enerjisinin değişimi oranı $\sigma^2_{yeni} / \sigma^2_{eski}$ şeklinde hesaplanır. Gelen ses çerçevesi gürültü olarak belirlendiğinde bu oran değerine göre p sabiti değişir. p sabitinin değişimi oranlarına göre değişim tablosu aşağıda verilmiştir.

Tablo 1. p sabitinin değişimi oranına göre değişimi

$\sigma^2_{yeni} / \sigma^2_{eski} \geq 1.25$	$p=0.25$
$1.25 > \sigma^2_{yeni} / \sigma^2_{eski} \geq 1.10$	$p=0.20$
$1.10 > \sigma^2_{yeni} / \sigma^2_{eski} \geq 1$	$p=0.15$
$\sigma^2_{yeni} / \sigma^2_{eski} < 1$	$p=0.10$

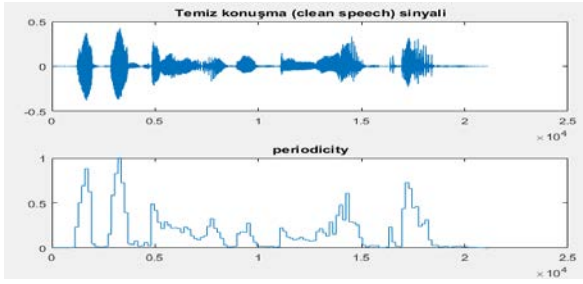
2.2 Periodicity (Periyodiklik)

Periyodiklik konuşma, müzik gibi ses işaretlerinde sesli bölgeyi tespit etmede önemli bir özelliktir. Sinyalin periyodik olma özelliği, pitch estimation (en kısa tekrarlanabilir aralık) ile belirlenebilir. Pitch estimation değerinin hesaplanabilmesi için normalize edilmiş özilinti fonksiyonu $R(\tau)$ aşağıdaki formülle hesaplanabilir.

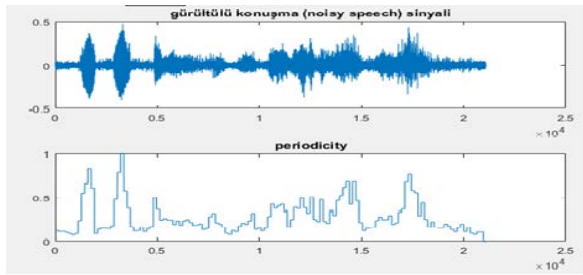
$$R(\tau) = \frac{\sum_{n=0}^{N-m-1} x(n)x(n+\tau)}{\sqrt{\sum_{n=0}^{N-m-1} x_n^2(n+\tau)}}, \quad T_{\min} < \tau < T_{\max} \quad (5)$$

$$C = \text{Max}(R(\tau))$$

Burada τ gecikme değerini göstermektedir. Yukarıdaki formülde $x(n)$; $n=0,1,\dots,N$ aralığında giriş sinyal çerçevesinin uzunluğunu temsil etmektedir. τ değerinin T_{min} 'den T_{max} 'a kadar değerlerini kullanarak özilinti fonksiyonu hesaplanır. Bu algoritmada bu değerleri sırasıyla 16 örnek (2 msn) ve 64 örnek (8 msn) seçilmiştir. Her ses çerçevesi için maksimum $R(\tau)$ değeri bize o ses çerçevesinin periyodikliğini (C) verir. $C=1$ ise tamamen periyodik, $C=0$ ise sinyal rastgele bir sinyaldir diyebiliriz. Şekil 4 ve şekil 5'te sırasıyla temiz bir konuşma sinyali ve bu konuşma sinyalinin periyodikliğı ve gürültülü bir konuşma sinyali ve onun periyodikliğı gösterilmiştir.



Şekil 4. Temiz bir konuşma sinyali ve periyodiklik grafiği



Şekil 5. Gürültülü bir konuşma sinyali ve periyodiklik grafiği

2.3 Spectral Flatness (SF-Spektral düzlük)

SF frekans düzleminde kullanılan bir özelliktir. Sadece gürültünün olduğu bölgelerin SF değeri, konuşmanın olduğu bölgelere göre daha yüksektir. Öncelikle ses sinyalinin çerçevelere ayırıp her çerçeve için ayrı ayrı FFT (Fast Fourier Transform) alınır. Aşağıdaki formül ses çerçevelerinin FFT'sini almak için kullanılır.

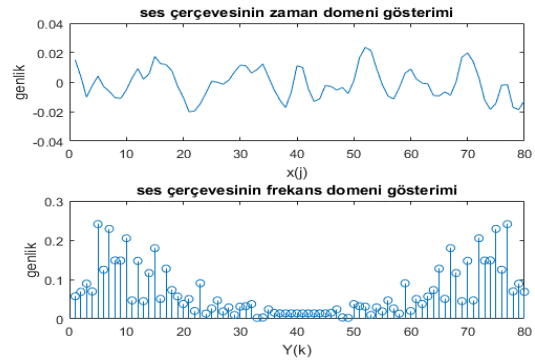
$$Y(k) = \sum_{j=1}^n x(j)W_n^{(j-1)(k-1)} \quad (6)$$

$$W_n = e^{(-2\pi i)/n}$$

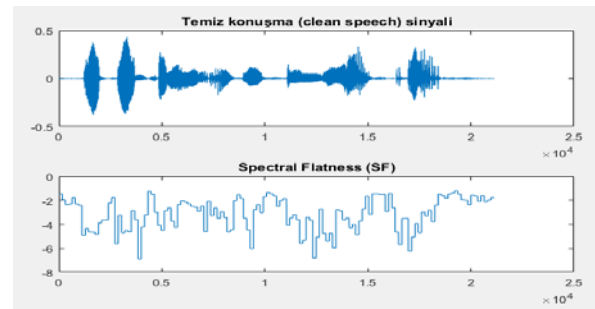
Burada $x(j)$ ses çerçevesi, $Y(k)$ ise $x(j)$ ses sinyalinin frekans düzleminde gösterimidir.

Şekil 6' da örnek bir ses çerçevesi ve onun FFT grafiği gösterilmiştir. Her ses çerçevesi için buna benzer bir şekilde FFT alındıktan sonra aşağıdaki formülü kullanarak dB cinsinden SF hesaplanır.

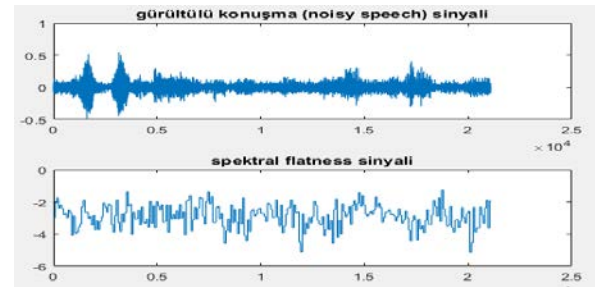
$$SF_{dB} = 10 * \log_{10} \left(\frac{\sqrt{\prod_{k=0}^{N-1} Y(k)}}{\sum_{k=0}^{N-1} Y(k)} \right) \quad (7)$$



Şekil 6. Örnek ses sinyali ve frekans düzlemi gösterimi



Şekil 7. Temiz bir konuşma sinyali ve SF grafiği



Şekil 8. Gürültülü bir konuşma sinyali ve SF grafiği

3. Önerilen Algoritma

Önerilen algoritmada öncelikle ses sinyali 10 msn süreli çerçevelere ayrılır. İlk sesli beş çerçevenin STE değerlerinin ortalaması gürültünün enerjisinin başlangıç eşik seviyesi olarak kabul edilir ($STE_{eşik}$). Her çerçeve için yukarıda açıklanan üç özellik hesaplanır. Eğer bu üç özellikten bir veya daha fazlası belirlenen eşik seviyesinin üstünde ise konuşma olduğu, bu özelliklerden her hangi birinin eşik seviyesinin üstünde olmaması durumunda konuşma olmadığı kabul edilir. Aşağıda önerilen algoritmanın betik döngüsü verilmiştir.

1. Ses çerçevesinin boyutu 10 msn olarak alınmıştır.
2. Kullanılan özellikler için eşik değerleri harici olarak belirlenmiştir.
 - STE için K sabiti ($K>1$)
 - Periyodiklik eşik değeri
 - SF_eşik değeri

3. for i=1 den çerçeve sayısına kadar
 - 3.1. STE(i)'yi hesapla
 - 3.2. Periyodikliği(i)'yi hesapla
 - 3.3 Her çerçeve için FFT uygula.
 - 3.3.1. Her çerçeve için SF değerini hesapla.
 - 3.4. STE, Periyodiklik ve SF eşik değeri belirlendikten sonra, ilk 5 çerçevenin enerjisinin ortalaması gürültü eşik değeri olarak, yine ilk beş çerçevenin maksimum SF değeri ise gürültünün SF'si olarak değerlendirildi ve STE_{noise} ile SF_{noise} değeri hesaplanır.
 - 3.4.1. SF_{noise}=max(SF(1,1:5)); olarak girildi
 - 3.4.2. STE_{noise}= STE(1,1:5)/5;
 - 3.5. Sayaç=0
 - Eğer STE(i)>=K.STE_{noise} değeri, sayaç=sayaç+1;
 - Eğer SF(i)-SF_{noise}<=SF_{esik} ise sayaç=sayaç+1;
 - Eğer Periodicity(i)>=0.3 ise sayaç=sayaç+1;
 - 3.6. Eğer sayaç>=1 ise; Ses çerçevesi gürültülü konuşmadır; Eğer sayaç>=1 değilse; Ses çerçevesi gürültüdür;
 - 3.7. Eğer ses çerçevesi gürültü ise aşağıdaki formülle STE'nin eşik değeri güncellendi. P=0.15 kabul edildi. STE_{esikyeni}=(1-p).STE_{esikeski}+p.STE_{noise}
4. Ardışık 10 çerçeveden daha az sessiz çerçeve varsa bir olarak değiştir.
5. Ardışık beş çerçeveden daha az konuşma çerçeve varsa sıfır olarak değiştir.

4. Deneysel Sonuçlar

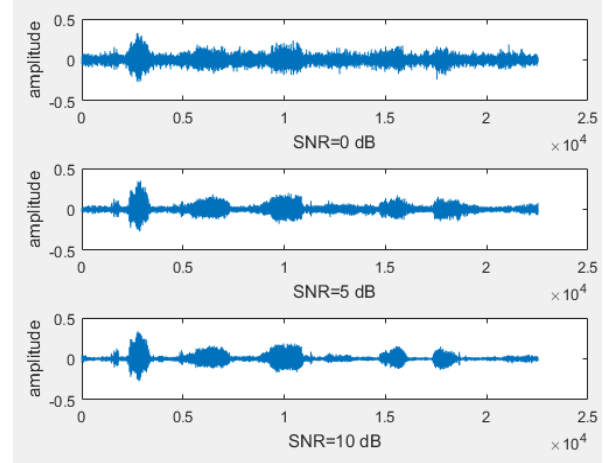
Bu çalışmada; Intel Core i5-5200U CPU 2.2 GHz işlemcili, 8 GB RAM bellek ve 64 bit Windows 10 işletim sistemine sahip bir bilgisayar, MATLAB R2016a programı kullanılmıştır. Bir kadın ve bir erkek konuşmacı sesleri [14], AURORA veri tabanından alınan farklı SNR değerlerindeki gürültüye eklenerek oluşturulan test sinyalleri elde edilmiştir [15]. Ses dosyaları 8 KHz örnekleme frekansı ile örneklenmiştir.

SNR, VAD performansını incelerken kullanılan önemli kavramlardan biridir. SNR değerinin düşük olması ses sinyalinin daha fazla gürültüye maruz kaldığını belirtir ve bu da VAD karar mekanizmasının hata yapma olasılığını artırır. Şekil 9'da farklı SNR değerlerine sahip ses sinyalleri verilmiştir.

VAD algoritmasının verimliliğini değerlendirmek için aşağıdaki parametreler kullanılmıştır.

1. *Doğru gürültülü konuşma tespiti (HR1)*: Bu parametre ses sinyalinde doğru tespit edilen gürültülü konuşma içeren ses çerçevelerini belirtir.
2. *Doğru gürültü tespiti (HR0)*: Bu parametre ses sinyalinde doğru tespit edilen gürültülü ses çerçevelerini belirtir.
3. *Hatalı gürültülü konuşma tespiti (HR0_hata)*: Bu parametre ses sinyalinde yalnızca gürültü olmasına rağmen VAD karar mekanizmasının yanlış tespiti ile gürültülü konuşma tespit edilmesidir.
4. *Hatalı gürültü tespiti (HR1_hata)*: Bu parametre ses sinyalinin gürültülü konuşma olmasına rağmen VAD karar mekanizmasının yanlış tespiti ile ses çerçevesinin yalnızca

gürültü olarak tespit edilmesidir.



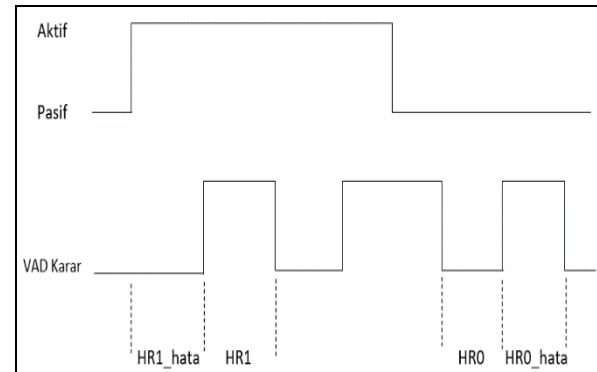
Şekil 9. SNR=0 dB, SNR=5 dB, SNR=10 dB değerlerine sahip gürültülü konuşma sinyalleri

HR1_hata parametresinin yüksek olarak bulunması konuşma kalitesini ve anlaşılabilirliğini olumsuz anlamda etkileyecektir. HR0_hata parametresinin yüksek bulunması ise VAD verimliliğini olumsuz yönde etkileyecektir. Bu parametrelerden HR1 (Speech hit rate) ve HR0 (Silence hit rate) doğruluk oranları yüzde olarak aşağıdaki gibi hesaplanmıştır:

$$\%HR1 = \frac{HR1}{HR1 + HR1_hata} \cdot 100 \quad (8)$$

$$\%HR0 = \frac{HR0}{HR0 + HR0_hata} \cdot 100 \quad (9)$$

Yukarıdaki parametrelerin ses sinyalinde gösterimi şekil 10'daki gibidir.



Şekil 10. VAD değerlendirme parametrelerinin gösterimi

Önerilen algoritma, ALED ve LED algoritmaları için HR1 (sesli bölümün doğruluk oranı), HR0 (sessiz bölümün doğruluk oranı) ve bu ikisinin ortalaması A (Average) kullanılarak yapılan değerlendirme sonuçları woman.wav ve man.wav sinyalleri için Tablo 2'den Tablo 7'ye kadar sırasıyla verilmiştir. Kadın ve erkek seslendirmesindeki cümleler aşağıdaki gibidir:

Örnek 1 - man.wav: "The birch canoe slid on the smooth planks".

Örnek 2 - woman.wav: "Let's all join as we sing the last chorus".

Tablo 2. Önerilen algoritmada man.wav dosyasının farklı gürültü ve SNR değerleri için deneysel sonuçlar

ÖNERİLEN ALGORİTMA					
Ses	Gürültü	SNR	HR1(%)	HR0(%)	A (%)
woman.wav	Airport	0 dB	96.67	46.49	71.57
		5 dB	96	49.12	72.56
		10 dB	96.67	71.05	83.85
	Babble	0 dB	96	53.51	74.75
		5 dB	94	72.81	83.40
		10 dB	100	76.32	88.15
	Exhibition	0 dB	98	54.39	76.19
		5 dB	99.33	57.89	78.61
		10 dB	100	78.95	89.47

Tablo 3. Önerilen algoritmada woman.wav dosyasının farklı gürültü ve SNR değerleri için deneysel sonuçlar

ÖNERİLEN ALGORİTMA					
Ses	Gürültü	SNR	HR1 (%)	HR0(%)	A (%)
man.wav	Airport	0 dB	95.14	42.03	68.58
		5 dB	96.53	46.38	71.45
		10 dB	97.92	61.59	79.75
	Babble	0 dB	97.22	52.90	75.06
		5 dB	95.14	65.94	80.54
		10 dB	96.53	70.29	83.40
	Exhibition	0 dB	86.11	39.13	62.62
		5 dB	93.06	43.48	68.26
		10 dB	97.92	88.41	93.16

Tablo 4. ALED algoritmasında man.wav dosyasının farklı gürültü ve SNR değerleri için deneysel sonuçlar

ALED					
Ses	Gürültü	SNR	HR1(%)	HR0(%)	A(%)
man.wav	Airport	0 dB	88.19	53.62	70.90
		5 dB	95.14	50.00	72.56
		10 dB	95.14	71.74	83.43
	Babble	0 dB	93.06	36.96	65.00
		5 dB	97.22	54.35	75.78
		10 dB	96.53	49.28	72.90
	Exhibition	0 dB	84.72	52.90	68.81
		5 dB	97.92	47.83	72.87
		10 dB	98.61	86.23	92.42

Tablo 5. ALED algoritmasında woman.wav dosyasının farklı gürültü ve SNR değerleri için deneysel sonuçlar

ALED					
Ses	Gürültü	SNR	HR1(%)	HR0(%)	A(%)
woman.wav	Airport	0 dB	93.33	58.77	76.05
		5 dB	95.33	74.56	84.94
		10 dB	98.67	85.09	91.87
	Babble	0 dB	91.33	42.11	66.71
		5 dB	97.33	55.26	76.29
		10 dB	99.33	70.18	84.75
	Exhibition	0 dB	96.67	54.39	75.52
		5 dB	98.67	58.77	78.71
		10 dB	100.00	90.35	95.17

Tablo 6. LED algoritmasında man.wav dosyasının farklı gürültü ve SNR değerleri için deneysel sonuçlar

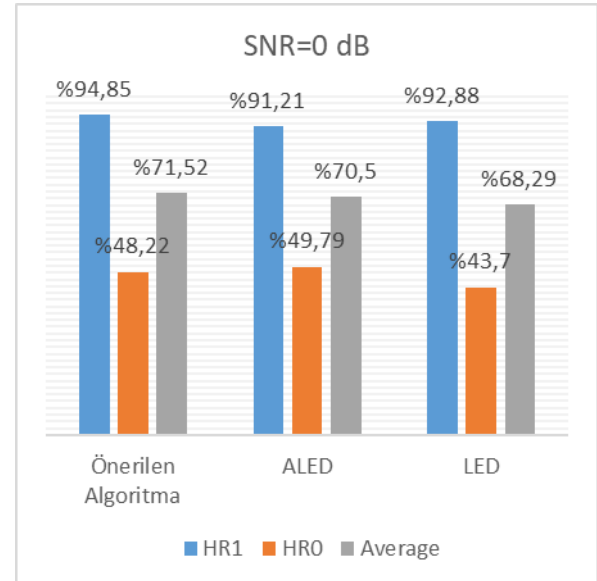
LED					
Ses	Gürültü	SNR	HR1(%)	HR0(%)	A(%)
man.wav	Airport	0 dB	88.19	50.00	69.09
		5 dB	93.75	61.59	77.67
		10 dB	95.14	73.91	84.52
	Babble	0 dB	92.36	39.86	66.10
		5 dB	97.92	49.28	73.58
		10 dB	98.61	43.48	71.04
	Exhibition	0 dB	86.11	47.83	66.96
		5 dB	98.61	44.20	71.40
		10 dB	98.61	85.51	92.05

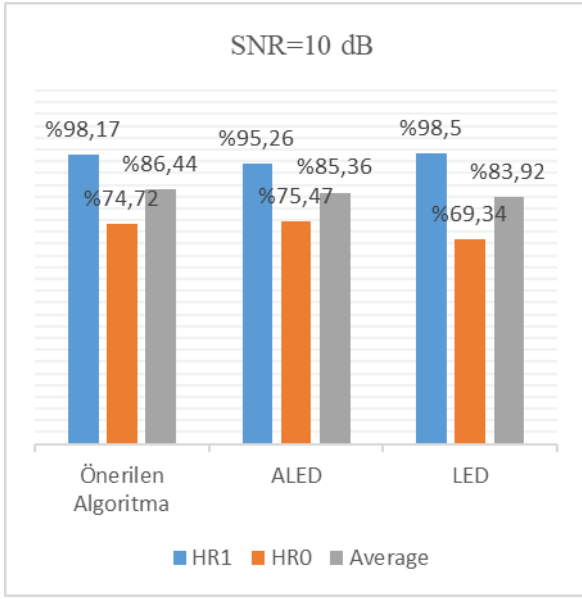
Tablo 7. LED algoritmasında woman.wav dosyasının farklı gürültü ve SNR değerleri için deneysel sonuçlar.

LED					
Ses	Gürültü	SNR	HR1(%)	HR0(%)	A(%)
woman.wav	Airport	0 dB	97.33	50.88	70.14
		5 dB	95.33	75.44	85.38
		10 dB	99.33	70.18	84.75
	Babble	0 dB	95.33	29.82	62.57
		5 dB	98.67	53.51	76.08
		10 dB	99.33	56.14	77.73
	Exhibition	0 dB	98.00	43.86	70.92
		5 dB	99.33	57.89	78.61
		10 dB	100.00	86.84	93.42

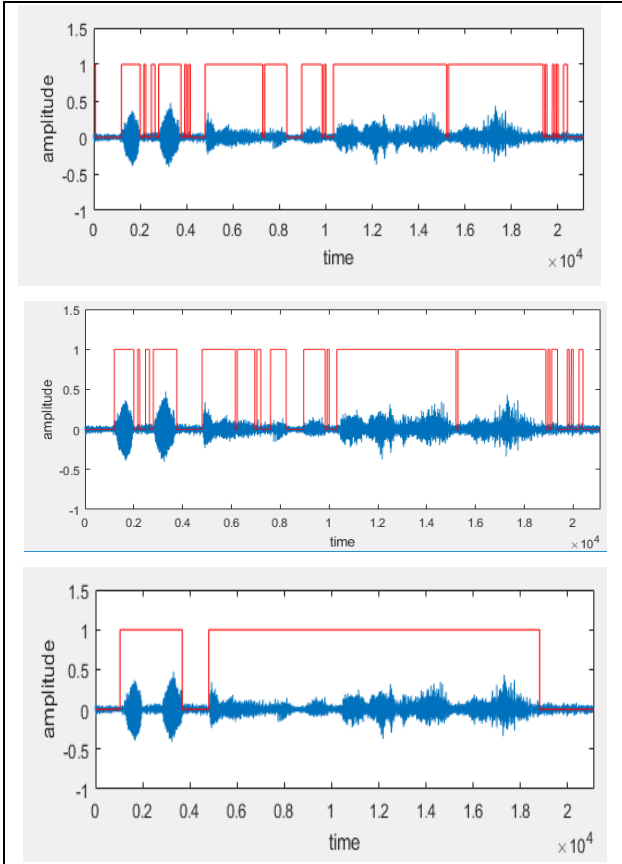
Yukarıda sonuçlara bakıldığında önerilen algoritmanın özellikle düşük SNR değerlerinde ALED ve LED metotlarıyla karşılaştırıldığında gürültülü konuşma sinyalinde konuşmanın olduğu sesli bölümün doğruluk oranında (HR1) daha iyi performans göstermiştir. Sadece gürültülü bölgelerin (HR0) tespit edilmesinde ALED algoritması önerilen algoritmadan biraz daha iyi sonuçlar vermiştir, yine HR0 tespitinde önerilen algoritma ile LED algoritmasından daha iyi sonuçlar elde edilmiştir.

Tablo 8. SNR=0 dB için ALED, LED ve önerilen algoritmanın % cinsinden karşılaştırması



Tablo 9. SNR=10 dB için ALED, LED ve önerilen algoritmanın % cinsinden karşılaştırması

Tablo 8 ve Tablo 9’da sonuçların daha iyi anlaşılabilmesi için tablolarda bulunan farklı gürültü tiplerine sahip kadın ve erkeğe ait SNR=0 dB ve SNR=10 dB’deki ortalama HR1, HR0 ve A (Average) değerleri, üç farklı yöntem için gösterilmiştir.

**Şekil 11.** LED ve ALED algoritması ile önerilen algoritmanın karşılaştırması:

a) Gürültülü sinyale LED algoritmasının uygulanmasının sonucu (SNR=0 dB)

b) Gürültülü sinyale ALED algoritmasının uygulanmasının sonucu (SNR=0 dB).

c) Gürültülü sinyale önerilen algoritmanın uygulanmasının sonucu (SNR=0 dB).

Şekil 11’ de önerilen algoritma, ALED ve LED algoritmasının woman.wav ses sinyaline SNR=0 dB olan havaalanı (airport) gürültüsü eklenmesiyle oluşturulmuş ses sinyaline uygulanmasıyla elde edilen sonuçlar gösterilmiştir. ALED ve LED algoritması konuşmanın olmadığı yerleri tespit etmede bizim önerdiğimiz algoritmaya göre biraz daha başarılı olsa da konuşma bölgelerinde kelimelerdeki başlangıç ve bitiş bölümlerini tespit etme açısından önerilen algoritmanın performansı daha iyidir, bu ise konuşmanın anlaşılabilirliği açısından oldukça önemlidir.

5. Sonuç

Bu çalışmada kolay uygulanabilir verimli bir VAD algoritması geliştirilmiştir. Bunun için üç temel kısa süreli özellik (STE, Periodicity ve SF) kullanılmıştır. Önerilen bu algortmada kadın ve erkek seslerine eklenmiş farklı tip ve SNR’de gürültüler ile oluşturulan dosyalar MATLAB programı yardımıyla test edilmiştir. Deneysel sonuçlar sesli çerçevelerin tespit edilmesinde oldukça iyi sonuçlar verirken, gürültülü çerçevelerin tespitinde algoritmanın geliştirilmesi gereken yönleri olduğunu göstermiştir. Bu çalışma çok kanallı ortamlarda gürültü azaltma çalışmalarında, gürültünün tespit edilmesi aşamasında fayda sağlayacaktır.

Kaynaklar

- [1] M. H. Moattar and M. M. Homayounpour, “A Simple But Efficient Real-Time Voice Activity Detection Algorithm”, 17th EUSIPCO, pp. 2549-2553, 2009.
- [2] K. Sakhnov, E. Verteletskaaya, B. Simak, “Low Complexity Voice Activity Detector Using Periodicity And Energy Ratio”, 16th International Conference on Systems, Signals and Image Processing IEEE, pp. 1-5, 2009.
- [3] E. Verteletskaaya, K. Sakhnov, “Voice Activity Detection for Speech Enhancement Applications”, ACTA POLYTECHNICA, Vol.50, No.4, 2010.
- [4] M. H. Moattar, M. M. Homayounpour, N.K. Kalantari “A New Approach For Robust Realtime Voice Activity Detection Using Spectral Pattern”, International Conference on Acoustic, Speech and Signal Processing IEEE, pp. 4478-4481, 2010.
- [5] Y. K Bharath, S. Veena, K. V. Nagalakshmi, Manjunath Darshan, Rohini Nagapadma, “Development of Robust VAD Schemes for Voice Operated Switch Application in Aircrafts”, 2nd International Conference on Applied and Theoretical Computing and Communication Technology (iCATccT), pp. 191-195, 2016.
- [6] T. H. Zaw, N. War “The Combination of Spectral Entropy, Zero Crossing Rate, Short Time Energy, and Linear Prediction Error for Voice Activity Detection”, International Conference of Computer and Information Technology (ICCIT), pp. 1-5, 2017.
- [7] N. Lezzoum, G. Gagnon, J. Voix, “Voice Activity Detection System for Smart Earphones”, IEEE Transaction on Consumer Electronics, Vol. 60, pp. 737-744, 2014.

- [8] M. Kumari, I. Ali, "An Efficient Un-Supervised Voice Activity Detector for Clean Speech", International Conference on Communication, Control and Intelligent Systems (CCIS), pp. 227-232, 2015.
- [9] A. Pasad, K. Sabu, P. Rao, "Voice Activity Detection for Children's Read Speech Recognition in Noisy Conditions", Twenty-third National Conference on Communications (NCC), pp. 1-6, 2017.
- [10] J. Pang, "Spectrum Energy Based Voice Activity Detection", IEEE 7th Annual Computing and Communication Workshop and Conference (CCWC), pp. 1-5, 2017.
- [11] I. Almajai, B. Milner, "Using Audio-Visual Features for Robust Voice Activity Detection in Clean and Noise Speech", 16th European Signal Processing Conference (EUSIPCO), pp. 1-5, 2008.
- [12] P. Pollak, P. Sovka and J. Uhlir, "Noise Suppression System For A Car", Third European Conference on Speech Communication and Technology, 3rd European Conference on Speech Communication and Technology-EUROSPEECH pp. 1073-1076, 1993.
- [13] K. Sakhnov, E. Verteletskaya, B. Simak, "Dynamical Energy-Based Speech/Silence Detector for Speech Enhancement Applications", in *Proc. of the World Congress on Engineering*, vol. 1, pp. 801, 2009.
- [14] IEEE Recommended Practice for Speech Quality Measurements. *IEEE Trans. Audio and Electroacoustics*, Vol. 17, pp. 225-246, 1969.
- [15] <http://ecs.utdallas.edu/loizou/speech/noizeus/>

