

See discussions, stats, and author profiles for this publication at: <https://www.researchgate.net/publication/269303283>

# Voiced-unvoiced classification of speech using autocorrelation matrix

Conference Paper · April 2014

DOI: 10.1109/SIU.2014.6830601

CITATIONS

0

READS

88

3 authors, including:



Ömer Emre Yetgin

Anadolu University

4 PUBLICATIONS 10 CITATIONS

SEE PROFILE



Ozgul Salor

Gazi University

82 PUBLICATIONS 542 CITATIONS

SEE PROFILE

Some of the authors of this publication are also working on these related projects:



Unmanned Air Systems [View project](#)



National Power Quality Project [View project](#)

# ÖZİLİNTİ MATRİSİ KULLANILARAK ÖTÜMLÜ-ÖTÜMSÜZ KONUŞMA SINIFLANDIRMA

## VOICED-UNVOICED CLASSIFICATION OF SPEECH USING AUTOCORRELATION MATRIX

Zekeriya Şentürk<sup>1,2</sup>, Ömer Emre Yetgin<sup>1</sup>, Özgül Salor<sup>2</sup>

1. Elektrik-Elektronik Mühendisliği Bölümü  
KARA HARP OKULU  
zsenturk@kho.edu.tr, oeyetgin@kho.edu.tr

2. Elektrik Elektronik Mühendisliği Bölümü  
GAZİ Üniversitesi  
salorduma@gazi.edu.tr

### ÖZETÇE

*Bu bildiri, konuşma sinyalinin ötümlü-ötümsüz sınıflandırılması amacıyla geliştirilen hızlı bir yöntem anlatılmıştır. Önerilen yöntemde, sinyalin enerjisinin, özilinti fonksiyonunun tepe değerlerinin farkının, sıfır geçiş sayısının ve özilinti fonksiyonunun bir birimlik gecikmedeki değerinin birlikte kullanılması ile ötümlü-ötümsüz kararı vermektedir. Yöntem, 1 kadın ve 2 erkek konuşmacının laringograf sinyali ile steryo olarak yapılmış ses kayıtları üzerinde denenmiştir. Laringograf sinyalinden çıkartılan ötümlü-ötümsüz kararı elle etiketlendikten sonra, önerilen yöntem ile elde edilen sonuçlarla karşılaştırılmıştır. Kadın sesi için % 100, erkek sesleri için % 98 doğrulukta sonuçlar elde edilmiştir.*

### ABSTRACT

*In this paper, a fast method for voiced-unvoiced classification of speech signals is introduced. The suggested method makes the V-UV decision, using signal energy, the peak-to-peak difference of the autocorrelation function, number of zero crossings of the autocorrelation function and the unit delay autocorrelation coefficient all together. This method has been tested on speeches of three speakers, one woman and two men, which include both the speech waveform and the laryngograph signal in stereo form. Having labeled the speech using the laryngograph signal manually, comparison of the hand-labelled decisions and those of the proposed method is achieved. The accuracy of the proposed method is found to be 100% for woman and 98% for men.*

### 1. GİRİŞ

Günümüzde konuşma sinyalleri için çok sayıda uygulama bulunmaktadır ve birçok konuşma çözümü ve sentez sisteminde, sesin ötümlü ya da ötümsüz olmasının tespiti kritik öneme sahiptir [1]. Ötümlü-ötümsüz ses tespitiyle, konuşma sisteminde ses tellerinin titreşip titreşmediği tespit edilebilir [2]. Bu tespit, konuşma modellemesinde, konuşmacı tanıma ve konuşma tanıma uygulamalarında, sinyalin özniteliklerinin belirlenmesi gibi uygulamalarda büyük öneme sahiptir [1]. Bu yüzden ses işleme sürecinin başlangıcında da ötümlü-ötümsüz ayrımı ve perde periyodu tahmininin doğru ve eksiksiz bir şekilde yapılması gerekmektedir.

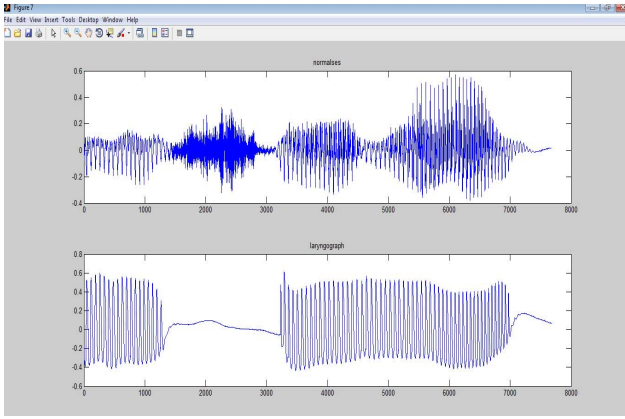
Ötümlü ve ötümsüz sınıflandırma problemi, üzerinde yıllardır çalışılan önemli bir problemdir [3]. Ötümlü ötümsüz sınıflandırmada doğrusal öngörü kodlama (LPC) çözümü

[4], kepsral zirve [5], bazı mel-frekans kepsral katsayıları [6], sıfır-geçiş oranı [7], sinüzoidal modeldeki sinyalin gürültüye oranı [8], özilinti tepe noktası [9] gibi özellikler kullanılabilir. Her bir özelliğin kendine has avantajları ve dezavantajları olduğu için, geliştirilen yeni yöntemlerde genellikle, ötümlü ve ötümsüz ses tespiti için birden fazla özellik kullanılmaktadır [10].

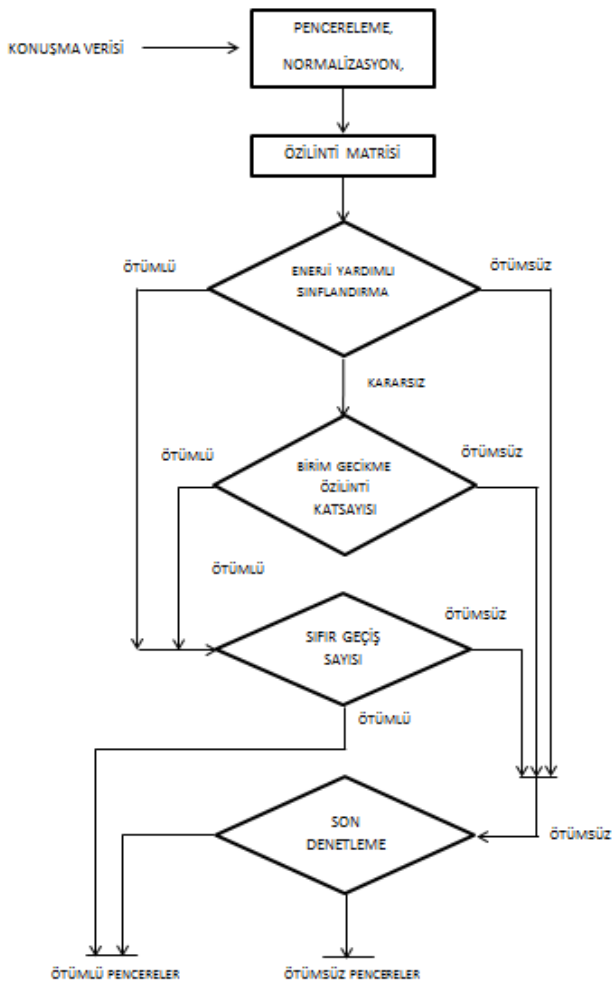
Bu çalışmada, sinyalin enerjisinin, özilinti fonksiyonunun tepe değerlerinin farkının, özilinti fonksiyonunun sıfır geçiş sayısının ve özilinti fonksiyonunun bir birimlik gecikmedeki değerinin hep birlikte değerlendirilmesi ile ötümlü-ötümsüz kararı vermektedir. Doğrudan sinyalin kendisi üzerinde değil, sinyalden oluşturulmuş olan özniteliklerin belirli bir sıra ile değerlendirilmesi ile çok yüksek ötümlü-ötümsüz karar başarımları elde edilmiştir.

### 2. ÖTÜMLÜ-ÖTÜMSÜZ SINIFLANDIRMA YÖNTEMİ

Özilinti fonksiyonu sinyalin kendine ne kadar benzediğinin bir ölçüsüdür. Oluşturulan sistemde tüm işlemler konuşma sinyalinden elde edilen özilinti matrisi üzerinden yapılmaktadır. Bu matrisin her bir satırı bir pencereleşmiş ses verisinin özilinti fonksiyonunu vermektedir. Bu matris üzerinde yapılacak olan değerlendirmeler, temel olarak enerji değeri, özilinti maksimum sapması, birim gecikmeli özilinti katsayısını ve özilintinin sıfır geçiş sayılarını kullanılmaktadır. Bu amaçla kullanılan konuşma kayıtları iki kanallı olarak alınmış ve bir kanala laringograf sinyali ve diğer kanala da konuşma sinyali kaydedilmiştir. Şekil-1'de kullanılan kayıtlardan örnek bir kısım gösterilmiştir. Konuşmanın ötümlü olarak adlandırılan kısmı için ses tellerinin periyodik olarak açılıp kapandığı ve laringograf sinyalinin, zaman alanında tam periyodik bir yapı gösterdiği bilgisinden faydalanılarak, konuşma sinyali laringograf sinyali yardımıyla ötümlü, ötümsüz ve geçiş pencereleri olarak 3 sınıfa ayrılıp, elle etiketlenmiştir. 1 kadın, 2 erkek toplam 3 farklı kişinin aynı cümleleri söylerken alınan ses kayıtları üzerinde etiketleme işlemi yapılmış, geçiş pencereleri hesaba katılmadan algoritma uygulanmış ve sonuçlar elde edilmiştir. Bu çalışmada, sinyalin kendisinin yerine özilinti değerlerinin kullanılmasının nedeni, bize sinyal hakkında daha ayrıntılı ve daha sağlam bilgiler vermesinin yanında perde periyodu hesabı için sıklıkla kullanılması [11] ve bu matrisin oluşturulmasının bu açıdan herhangi bir işlem yükü getirmeyeceğindedir.



Şekil 1. Eşzamanlı kaydedilmiş konuşma ve laringograf sinyalleri



Şekil 2. Önerilen yöntemin akış şeması

Bu işlemi, zamanda sınırlı bir sinyal için olası tüm gecikmeler için tekrarladığımızda bize sağladığı çok etkin bilgiler vardır. Şekil-2'de, önerilen yöntemin akış şeması verilmiştir.

Yukarıda bahsedilen özilinti matrisinin oluşturulması için öncelikle ses sinyali 20 ms'lik parçalara ayrılmıştır. Daha sonra her bir 20 ms'lik veri için kaydırma miktarını sıfırdan

başlayarak birer birer arttırıp penceredeki örnek sayısı kadar yükseltilecek şekilde özilinti bulunur. Her bir kaydırmadan elde edilen özilinti değeri sırasıyla kaydedilir ve bu 20ms'lik veri için özilinti vektörünü oluşturur. Ses verisinde bölütlenmiş olan her 20 ms'lik pencere için bu işlem tekrarlanır ve oluşan vektörler alt alta sıralanarak özilinti matrisi oluşturulur.

$$y_1(k) = \sum_{t=1}^N x_1(t) \cdot x_1(t-k)$$

$$y_2(k) = \sum_{t=1}^N x_1(t) \cdot x_1(t-k)$$

$$y_L(k) = \sum_{t=1}^N x_1(t) \cdot x_1(t-k) \quad (1)$$

(1)'de,  $N$ , penceredeki örnek sayısını;  $L$  ses verisindeki toplam pencere sayısını göstermektedir. Bu işlem sonunda  $(N \times L)$  boyutunda matris elde edilir. Bu aşamadan sonra yapılacak bütün işlemler bu matris üzerinden yürütülerek sonuçlar elde edilecektir.

## 2.1. Enerji Sınıflandırması

### 2.1.1. Sinyal Enerjisi

Sinyal enerjisi, ötümlü-ötümsüz (Ö-ÖSZ) sınıflandırma için kullanılan etkin bir yöntemdir. Ötümlü konuşma sinyali periyodik bir yapı gösterip yüksek genlik değerlerine sahip olabilmekte iken, ötümsüz ses sinyali genellikle düşük genlikli ve periyodik olmayan yapıdadır [12]. Bu özelliklerinden ötürü ötümlü bir sinyalin enerjisi, ötümsüz bir sinyalin enerjisinden daha yüksek olur. Oluşturulan özilinti matrisinin ilk sütunu  $k=0$ 'da her bir pencerenin enerjisini vermektedir. Bu bilgi, Ö-ÖSZ sınıflandırma için bize değerli bir bilgi sağlamaktadır:

$$y_1(k) = \sum_{t=1}^N x_1(t) \cdot x_1(t-k) \quad (2)$$

(2),  $k=0$  için

$$y_1(0) = \sum_{t=1}^N (x_1(t))^2$$

olur. Bu sınıflandırma yöntemi için sınırdeğer belirleme işlemi kritik öneme sahiptir. Bu değeri daha kolay belirlemek ve her sinyale uygulanabilmesi adına, ilk başta konuşma verisini içerdiği maksimum genlik değerine bölerek normalizasyon yapılmaktadır. Yapılan gözlemlerle bu değer 0,5 alınması uygun görülmüştür.

### 2.1.2. Özilinti Fonksiyonunun Tepe Değerlerinin Farkı

Bir veri için özilinti değerlerinin ne kadar sapma gösterdiği, sinyalin genlikte ne kadar değiştiğinin bir göstergesidir. Düşük genlikte seyreden ve periyodik olmayan sinyaller için bu değer düşük kalmakta iken, yüksek genliğe çıkan ve periyodik bir yapı gösteren sinyalde bu değer belirgin bir şekilde artmaktadır. Ö-ÖSZ ayrımında da, aslında yüksek genlikli ve yaklaşık olarak periyodik bir sinyal, düşük genlikli ve periyodik yapı göstermeyen bir sinyalden ayrılmaya çalışılmaktadır. Bu açıdan bakıldığında özilinti fonksiyonunun tepe noktalarının saptması, Ö-ÖSZ ayrımında değerli bir bilgi

sunmaktadır. Burada uygulanan sınır değeri, enerji yardımcı sınıflandırma için kullanılan 0,5 değerinin iki katı olacak şekilde 1 seçilmiştir.

## 2.2. Birim Gecikmeli Özilinti Katsayısı

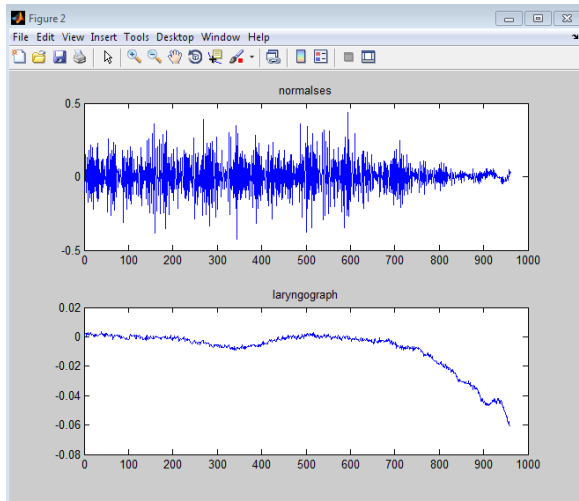
Özilinti, sinyalin kendine ne kadar benzediğinin bir ölçütüdür. Bir sinyali bir birim kaydırıp özilintisini tespit etmek aslında bu sinyalin ne derece periyodik olduğunu tespit etmek demektir.

$$j_1 = \frac{\sum_{t=1}^N x_1(t) \cdot x_1(t-1)}{\sum_{t=1}^N (x_1(t))^2} \quad (3)$$

Bu amaçla (3)'te verilen özilinti katsayısı kullanılmaktadır. Bu yöntem, geçiş pencerelerinde iyi çalışması nedeniyle kurulan algoritmada enerji yardımcı iki sınıflandırma yöntemi olan sinyal enerjisi ve özilinti fonksiyonunun tepe değeri farkı yöntemlerinde oybirliği olmadığı durumda, kullanılacak bir nicelik olarak seçilmiştir. Bu yöntemde, sinyalin kendisine % 97 ve daha fazla benzemesi durumunda, sinyal ötümlü ses sınıfına dahil edilmiştir.

## 2.3. Sıfır Geçiş Oranı

Sadece enerji ile yapılan ötümlü ötümsüz sınıflandırmalar yanıltıcı sonuçlar verebilmektedir. Çünkü, aslında ötümsüz olan ancak yüksek genlikli ve yüksek frekanslı bazı sesler vardır (Şekil 3). Bu tür konuşma sinyallerinin ayırımında sinyalin kendi sıfır geçişinin yanında, özilinti fonksiyonunun sıfır geçişleri de kullanılabilir. Sinyalin özilinti fonksiyonunun sıfır geçişi ağırlıklandırılmış olduğundan gürültü kaynaklı hatalardan daha az etkilenir ve yüksek genlikli, yüksek frekanslı konuşma sinyalleri için sinyalin sıfır geçişinden daha iyi çalışır. Özilinti fonksiyonunun sıfır geçiş oranı, ötümlü bir sinyalin gerçekten ötümlü olup olmadığının tespitinde kullanılabilirken, ötümsüz bir sinyalin doğrudan tespitinde kullanılması doğru olmaz. Çünkü, ötümsüz olan bölgelerin frekans değerlerinin çok yüksek olması gibi bir durum her zaman geçerli değildir.



Şekil 3. Yüksek enerjili ötümsüz konuşma kesiti

## 2.4. Son Denetleme Bloğu

Sıfır geçiş bloğu, ötümlü olarak karar verilen pencerelerden yüksek frekanslıların süzülerek, olması gerektiği gibi, ötümsüz olarak sınıflandırılması görevini yerine getirmektedir. Son denetleme bloğu da sistemin son aşamasına gelene kadar bir şekilde elenerek ötümsüz olarak belirlenen pencerelerin, tekrar gözden geçirilmesi ve sesin genliğinin düştüğü ancak ötümlü yapının devam ettiği bazı pencereleri ötümsüz olarak etiketlemekten kurtarılması amacıyla oluşturulmuştur. Bu aşamada sınır değerler %10 oranında oynatılarak, pencerenin hangi özellikleri cinsinden ötümlü hale geldiklerinin analizi yapılarak, bir karar verilmektedir.

## 3. DENEYSEL SONUÇLAR

Laringograf sinyali yardımıyla etiketlenen konuşma verisi önerilen yöntemle çalıştırıldığında Tablo-1'de verilen sonuçlar elde edilmiştir. Bu sonuçlara bakıldığında, yöntemin bu bildiriye kullanılan kadın sesi için hatasız çalıştığı, erkek sesleri için yaklaşık % 2'lik bir hata verdiği görülmektedir.

		K	E1	E2
ETİKET	ÖTÜMLÜ	313	246	329
	ÖTÜMSÜZ	346	353	396
	TOPLAM	659	599	725
SONUÇ	DOĞRU	659	585	715
	YANLIŞ	0	14	10
	DOĞRULUK %	100	97,66	98,62

Tablo 1. Doğruluk tablosu

## 4. SONUÇLAR

Bu bildiriye, sinyalin enerjisinin, özilinti fonksiyonunun tepe değerlerinin farkının, özilinti fonksiyonunun sıfır geçiş sayısının ve özilinti fonksiyonunun bir birimlik gecikmedeki değerinin hep birlikte değerlendirilmesi ile konuşma sinyalleri için ötümlü-ötümsüz kararı veren bir yöntem anlatılmıştır. Doğrudan sinyalin kendisi üzerinde değil, sinyalden oluşturulmuş olan özniteliklerin belirli bir sıra ile değerlendirilmesi ile çok yüksek ötümlü-ötümsüz karar başarımları elde edilmiştir. Bu amaçla laryngograph sinyali mevcut olan konuşma verilerinin el ile etiketlenmesi yoluyla bir veri tabanı oluşturulmuştur. Bu veri tabanından geçiş pencereleri süzülmesi ve oluşturulan algoritma bu konuşma verisine uygulanmıştır. Sistemin çalışması için gereksinim duyduğu işlemsel yük düşük olmasına rağmen, elde edilen doğruluk değerleri çok yüksek çıkmıştır. Bu çalışmanın devamı olarak, geçiş pencerelerinin yüzde olarak ötümlülük oranının tespiti hedeflenmektedir. Ayrıca, burada elde edilen özilinti matrisi özniteliklerini kullanarak gerçek-zamanlı uygulamalarda rahatlıkla kullanılacak gürültüye karşı sağlam çalışan bir perde periyodu algoritması oluşturulması da hedeflenmektedir.

## 5. KAYNAKÇA

- [1] Fisher, E., Tabrikian, J., Dubnov, S., "Generalized likelihoodratio test for voiced-unvoiced decision in noisy speech using the harmonic model," *IEEE Transactions on Audio, Speech, and Language Processing*, vol.14, no.2, pp. 502- 510, March 2006.
- [2] E. Fisher, J. Tabrikian and S. Dubnov, "Generalized Likelihood Ratio Test for Voiced-Unvoiced Decision in Noisy Speech Using the Harmonic Model," *IEEE Transactions on Audio, Speech, and Language Processing*, Vol. 14, No. 2, 2006, pp. 502-510. doi:10.1109/TSA.2005.857806
- [3] S. Ahmadi. and A. S. Spanias, "Cepstrum-based pitch detection using a new statistical V/UV classification algorithm," *IEEE Trans. Speech Audio Pro.*, vol. 7 No. 3, pp. 333-338, 1999
- [4] Nemer, E., Goubran, R., Mahmoud, S., "Robust voice activity detection using higher-order statistics in the LPC residual domain," *IEEE Trans. Speech Audio Process.*, vol. 9, no. 3, pp. 217 231, Mar. 2001.
- [5] B. Atal and M. Schroeder, "Predictive Coding of Speech Signals and Subjective Error Criteria," *IEEE Transactions on Acoustics, Speech and Signal Processing*, Vol. 27, No. 3, 2003, pp. 247-254. doi:10.1109/TASSP.1979.1163237
- [6] S. Imai, "Cepstral Analysis Synthesis on the Mel Frequency Scale," *IEEE International Conference on Acoustics, Speech, and Signal Processing*, Vol. 8, 2003, pp. 93-96.
- [7] B. Atal and L. Rabiner, "A Pattern Recognition Approach to Voicedunvoiced-Silence Classification with Applications to Speech Recognition," *IEEE Transactions on Acoustics, Speech and Signal Processing*, Vol. 24, No. 3, 2003, pp. 201-212. doi:10.1109/TASSP.1976.1162800
- [8] R. J. McAulay and T. F. Quatieri, "Pitch Estimation and Voicing Detection Based on a Sinusoidal Speech Model," *International Conference on Acoustics, Speech, and Signal Processing*, Vol. 1, 1990, pp. 249-252. doi:10.1109/ICASSP.1990.115585
- [9] L. Rabiner, "On the Use of Autocorrelation Analysis for Pitch Detection," *IEEE Transactions on Acoustics, Speech and Signal Processing*, Vol. 25, No. 1, 2003. pp. 24-33. doi:10.1109/TASSP.1977.1162905
- [10] Y. Qi and B. R. Hunt, "Voiced-Unvoiced-Silence Classifications of Speech Using Hybrid Features and a Network Classifier," *IEEE Transactions on Speech and Audio Processing*, Vol. 1, No. 2, 2002, pp. 250-255. doi:10.1109/89.222883
- [11] Li Hui, Bei-qian Dai, Lu Wei, "A Pitch Detection Algorithm Based on AMDF and ACF", *IEEE International Conference ICASSP 2006* , vol. 1, p. I-I, 2006
- [12] J. A. Marks, "Real time speech classification and pitch detection," *Communications and Signal Processing*, pp. 1-6, June , 1988.