

# İstatistiksel Yöntemlerle Konuşma Aktivitesi Sezimi #

<sup>+</sup>Özgür Devrim Orman, \*Emin Anarım

<sup>+</sup>TÜBİTAK

Ulusal Elektronik ve Kriptoloji Araştırma Enstitüsü, Gebze, 41470 Kocaeli  
oorman@mam.gov.tr

\*Boğaziçi Üniversitesi

Elektrik-Elektronik Mühendisliği Bölümü, Bebek, 80815 İstanbul  
anarim@busim.ee.boun.edu.tr

## Özetçe

Gürültü ile kirlenmiş bir konuşma kaydında konuşma işaretinin aktif olduğu bölgelerin sezimi değişik hızlarda iletim yapan sistemler ve sayısal konuşma kaydı saklama sistemleri için çok önemlidir. Gürültü işaretinin durağan olmaması geleneksel yöntemlerle yapılan sezimlerin istatistiksel yöntemler kadar başarılı gösterememesine sebep olmaktadır. Bu çalışmada konuşma aktivitesi seziminde kullanılan istatistiksel yöntemler irdelenmiş ve altbant ile yapılan modellemede bir yenilik olarak permütasyon yöntemiyle tasarlanmış altbant süzgeç öbeği kullanılmıştır.

## 1. Giriş

Konuşma işaretinin aktivite gösterdiği bölgelerin sezimi, hem konuşma işaretinin değişik hızlarda iletimini gerçekleştiren paket iletim tabanlı sistemlerin gerçekleştirilmesinde hem de konuşma kayıtlarının sayısal ortamda saklanmasında önemli bir yer tutmaktadır. Uygulamada değişik konuşma aktivitesi sezim yöntemleri kullanılmaktadır. ITU-T standardı olan G.729 [1] konuşma bölgelerinin belirlenmesinde bir çok parametreyi kullanmakla birlikte durağan olmayan gürültü karşısında istatistiksel yöntemlerin gösterdiği başarıyı gösterememektedir. İstatistiksel yöntemlerle konuşma aktivitesi sezimi temel olarak iki farklı şekilde uygulanabilir. Bu çalışmada ilk yöntem olarak gürültü ile kirlenmiş konuşma işaretinin Ayırık Zaman Fourier Dönüşümü'ne (AZFD) dayanan bir yaklaşım kullanılmıştır [2,3]. İkinci yöntem de ise AZFD yerine birbirine dikgen süzgeçlerden oluşturulmuş altbant süzgeç öbeği ele alınmıştır [4].

Bu çalışmada bilgisayar ortamından alınan gürültüsüz kayıt üzerine NOISE-ROM-0 veri tabanından kabarcık gürültüsü olarak nitelenen gürültünün eklenmesi ile elde edilen işaret üzerinde her iki sezim yaklaşımı da ayrı ayrı kullanılarak konuşma aktif bölgeler elde edilmiştir. Konuşma aktivitesi seziminde kullanılan altbant süzgeç öbeği, daha önceki çalışmalarda gözlemlenmiş geleneksel yöntemlerin aksine, süzgeçlerin birbirine dikgenliğini kesin kılan permütasyona dayalı tasarım yöntemiyle gerçekleştirilmiştir. İkinci bölüm istatistiksel konuşma aktivitesi sezim metodolojisine ayrılmıştır. Üçüncü bölümde çalışmanın ayrıntıları verilmektedir. Sonuç bölümü olan dördüncü bölümde de bu çalışmanın sonuçları ile birlikte elde edilen veriler doğrultusunda araştırmanın ilerideki yöneliminden söz edilmektedir.

## 2. Metodoloji

### 2.1. Olabilirlik Oran Testi Kullanarak Konuşma Aktivitesi Sezimi

İstatistiksel konuşma aktivitesi seziminde AZFD'ye dayanan ilk yöntemde uygulanan temel aşamalar aşağıdaki gibi sıralanabilir.

- Gürültü ile kirlenmiş konuşma işareti bölütlere ayrılır.
- Bölütler bir pencere işlevi ile çarpılır (genel olarak Hamming penceresi kullanılmaktadır).
- Pencerelemiş her bir bölüt AZFD'den geçirilir.
- Bölütlerin Fourier katsayılarından, her bir bölütün sıklık düzlemi güç yoğunluk katsayıları hesaplanır.

# Bu çalışma, Boğaziçi Üniversitesi Araştırma Fonu tarafından 99A207 numaralı proje kapsamında desteklenmiştir.

e. Bulunan güç yoğunluğu katsayıları kullanılarak bölütlerin işaret-gürültü kararı verilir ve oluşturulmuş istatistiksel model bu doğrultuda güncellenir.

Elimizdeki işaretin bulunabileceği olası iki kuramsal konum vardır:

$$H_0: y[n] = x[n] + w[n]$$

$$H_1: y[n] = w[n]$$

Bu gösterimde  $x[n]$  ses işaretini ve  $w[n]$ 'de eklenmiş olan durağan olmayan gürültüyü temsil etmektedir. Merkezi Limit Teoremi (MLT) doğrultusunda yukarıdaki işaretlerin AZFD katsayılarının birbirlerinden bağımsız, Gauss dağılımlı birer rasgele değişken oldukları varsayılabilir [5]. Varsayım doğrultusunda  $L$  uzunluklu  $\underline{Y}$  katsayı vektörünün gözlemlenmesinin sözü edilen kuramlar altındaki koşullu olasılıkları aşağıdaki gibi yazılabilir:

$$p(\underline{Y}/H_0) = \prod_{k=0}^{L-1} \frac{1}{\pi \lambda_w(k)} \cdot \exp \left\{ -\frac{|Y_k|^2}{\lambda_w(k)} \right\} \quad (1)$$

$$p(\underline{Y}/H_1) = \prod_{k=0}^{L-1} \frac{1}{\pi (\lambda_x(k) + \lambda_w(k))} \cdot \exp \left\{ -\frac{|Y_k|^2}{(\lambda_x(k) + \lambda_w(k))} \right\} \quad (2)$$

Yukarıdaki denklemlerde görülen  $\lambda_x(k)$  ve  $\lambda_w(k)$  değerleri sırası ile  $X_k$  ve  $W_k$ 'nin saçınımını göstermektedir.  $\lambda_w(k)$  değeri salt gürültü içeren bölgelerde güncellenip gürültü ve sesin birlikte bulunduğu bölgelerdeki işlemlerde en son güncellenmiş değeri kullanılır. Ekli bulunan gürültü durağan olduğunda saçınımın başlangıçtaki salt gürültü içeren bölgede bir defa kestirimi yeterli olacaktır [5], fakat bu karakteristikte bir gürültü ile gerçek uygulamalarda karşılaşmak imkansızdır.  $\lambda_w(k)$  değerleri analiz için oluşturulan ses dizgelerinin bir pencere işleviyle çarpılmasından sonra periodogram yönteminin uygulanmasıyla kestirilmektedir. Koşullu olasılık ifadelerini kullanarak  $\underline{Y}$  yöneyinin  $k$ . elemanının olabirlik oran testini aşağıdaki gibi gösterebiliriz.

$$\Lambda_k \stackrel{\Delta}{=} \frac{p(Y_k/H_1)}{p(Y_k/H_0)} = \frac{1}{1 + \xi_k} \cdot \exp \left\{ \frac{\gamma_k \cdot \xi_k}{1 + \xi_k} \right\}$$

$$\xi_k = \lambda_x(k) / \lambda_w(k)$$

$$\gamma_k = |Y_k|^2 / \lambda_w(k)$$

$\xi_k$  değeri ençok olabirlik kestirim yöntemiyle aşağıdaki gibi bulunur.

$$\xi_k^{(ML)} = \gamma_k - 1$$

$\underline{Y}$  yöneyinin tüm elemanları gözönüne alınarak bulunan olasılık oran testi Denklem 3'te gösterildiği gibidir.

$$\frac{1}{L} \sum_{k=0}^{L-1} \left\{ \gamma_k - \log \gamma_k - 1 \right\} \underset{H_0}{\overset{H_1}{>}} \eta \quad (3)$$

İkinci bir olasılık oran testi de doğrusal öngörü yöntemi yardımıyla bulunabilir [2]. GSM sistemlerinde DTX durumunda, doğrusal öngörü süzgeci ile gürültünün sıklık düzlemi güç dağılımı ve farksal enerjisinin kestirimi yapıldıktan sonra, ters doğrusal öngörü süzgeci çıkışındaki farksal enerji ile kestirilmiş bulunan gürültünün farksal enerjisi karşılaştırılarak konuşma aktivitesi sezimi gerçekleştirilir. Bu yaklaşım sonucu ortaya çıkan olabirlik oran testi Denklem 4'te ifade edilmiştir.

$$\frac{1}{L} \sum_{k=0}^{L-1} \frac{|Y_k|^2}{\lambda_N(k)} \underset{H_0}{\overset{H_1}{>}} 1 + \alpha \quad (4)$$

Oluşturulabilecek üçüncü bir olabirlik oran testi de  $H_0$  ve  $H_1$  durumları arasındaki geçişin Saklı Markov Modeller (SMM) kullanılarak yumuşatılmasıyla elde edilebilir [3]. Elimizdeki konuşma işareti bölütlerinin birbirleriyle oldukça ilintili oldukları göz önüne alındığında, bir sonraki durumun olasılığının bir önceki durum tarafından belirleneceği varsayılabilir. Söz konusu durum geçişleri birinci dereceden bir Markov modelle belirlenebilir. Oran testinin açık ifadesi gösterilecek evrelerden sonra elde edilir.

$$L(n) \stackrel{\Delta}{=} \frac{P(Y_n/q_n = H_1)}{P(Y_n/q_n = H_0)} = \frac{P(H_0)}{P(H_1)} \cdot \frac{P(q_n = H_1/Y_n)}{P(q_n = H_{01}/Y_n)} \stackrel{H_1}{>} \eta \stackrel{H_0}{<} \quad (5)$$

$$Y_n = [Y(1), Y(2), \dots, Y(n)]$$

$Y_n$  n. bölüme kadar olan gözlemleri göstermektedir. Denklem 5'te görülen soncul (a posteriori) olasılık oranı  $\Theta(n)$  ile gösterilmiştir.

$$\Theta(n) = \frac{P(q_n = H_1/Y_n)}{P(q_n = H_{01}/Y_n)}$$

Bu oranın hesaplanmasını kolaylaştırmak için "i" değeri durum değişkenini göstermek üzere  $\alpha_n(i)$  gibi bir ileri doğru değişken tanımlarız.

$$\alpha_n(i) = P(q_n = H_i, Y_n)$$

Tüm bu adımlardan sonra olabilirlik oran testini aşağıdaki şekilde ifade edebiliriz.

$$L(n) = \frac{P(H_0)}{P(H_1)} \cdot \Theta(n) = \frac{P(H_0)}{P(H_1)} \cdot \frac{a_{01} + a_{11}\Theta(n-1)}{a_{00} + a_{10}\Theta(n-1)} \cdot \Lambda(n) \quad (6)$$

Denklem 6'da görülen  $a_{ij}$  değerlerinde "i" bir önceki durumu, "j" değeri de bulunulan durumu göstermektedir. SMM'in konuşma aktivitesi sezimine eklenmesiyle olabilirlik oran testinde meydana getirilmiş değişim Denklem 6'da görülmektedir.

## 2.2 Dikgen Süzgeç Öbeği Kullanarak Konuşma Aktivitesi Sezimi

Altbant süzgeç öbeği kullanarak konuşma aktivitesi seziminde öncelikle dikkat edilmesi gereken unsur altbant süzgeçlerin birbirine dikgen olmasıdır. Süzgeç tasarımı seçilmiş bir eniyileme ölçütüne göre yapılmaktadır. Bu ölçütlerden birisi de kodlama kazancının ençoklanmasıdır [6]. Bir altbant kodlayıcının başarımlı ölçütü Denklem 7'deki gibi gösterilebilir.

$$G_{SBC} = \frac{\frac{1}{M} \sum_{k=0}^{M-1} \sigma_k^2}{\left[ \prod_{k=0}^{M-1} \sigma_k^2 \right]^{\frac{1}{M}}} \quad (7)$$

Denklem 7'de görülen  $\sigma_k$ , k. süzgeç çıkışının standart sapmasını göstermektedir. Önceden sözünü ettiğimiz gibi birinci dereceden bir Markov modeli kullanarak bölütler arası ilişkiyi ifade edersek, özilişki işlevi Denklem 8'deki gibi yazılabilir [6]. Bu çalışmada kullanılan altbant süzgeç modeli sürecin birinci derece Markov olduğu varsayımı altında kodlama kazancını ençoklayan, her biri 8 katsayılı 8 süzgeç bantından oluşan bir yapıdır.

$$R(k) = \rho^{|k|}, \quad k = 0, \pm 1, \dots \quad (8)$$

Permütasyon yöntemiyle, birbirine dikgen altbant süzgeçlerden oluşmuş süzgeç öbeği tasarımı temel olarak Hadamard dönüşüm matrisindeki permütasyon koşullamasıyla temellendirilebilir. Bu yöntemle yukarıdaki gibi değişik ölçütlere göre eniyilenecek bulunmuş bir kaynak süzgeçinden tüm süzgeçlerin dürtü yanıtları birbirine dikgen olan bir altbant süzgeç öbeği tasarlanabilir. Söz konusu süzgeçlerin sınırlı dürtü yanıtı olması bu şekilde bir altbant dönüşüm matrisinin tanımlanmasına olanak sağlamaktadır. Örnek olarak dört katsayılı bir alçak geçiren süzgeci parametrik olarak aşağıdaki gibi gösterebiliriz.

$$h_0(n) = [a \quad b \quad c \quad d] \quad (9)$$

Süzgeçimizin doğrusal fazlı olması halinde dürtü yanıtı simetrik yapıda olur. Bu özelliği kullanarak elimizdeki herhangi N katsayılı alçak geçiren süzgeci kullanarak 2N uzunluklu dürtü yanıtı olan doğrusal fazlı bir alçak geçiren süzgeç kolaylıkla meydana getirilebilir. Söz konusu 2N uzunluklu süzgeçle 2N bantlı bir altbant süzgeç öbeği tasarlayabiliriz. Denklem 9'daki dürtü yanıtını kullanarak permütasyon yöntemiyle dört bantlı bir altbant dönüşüm matrisi Denklem 10'da görüldüğü şekilde oluşturulabilir [7].

$$H = \begin{bmatrix} a & b & c & d \\ -b & a & d & -c \\ -c & -d & a & b \\ -d & c & -b & a \end{bmatrix} \quad (10)$$

Bu yöntemle konuşma aktivitesi sezimi gerçekleştirilirken kullandığımız gürültü modeli konuşma işaretinin olmadığı bölgelerde güncellenir. Konuşma aktif bölgelerde önceden oluşturulmuş gürültü modeli korunur. Konuşma var yok kararı, oluşturulmuş gürültü modeline dayanan bir işaret-gürültü oranı testiyle gerçekleştirilir [4]. Yöntemin detayları aşağıda gösterilmiştir.

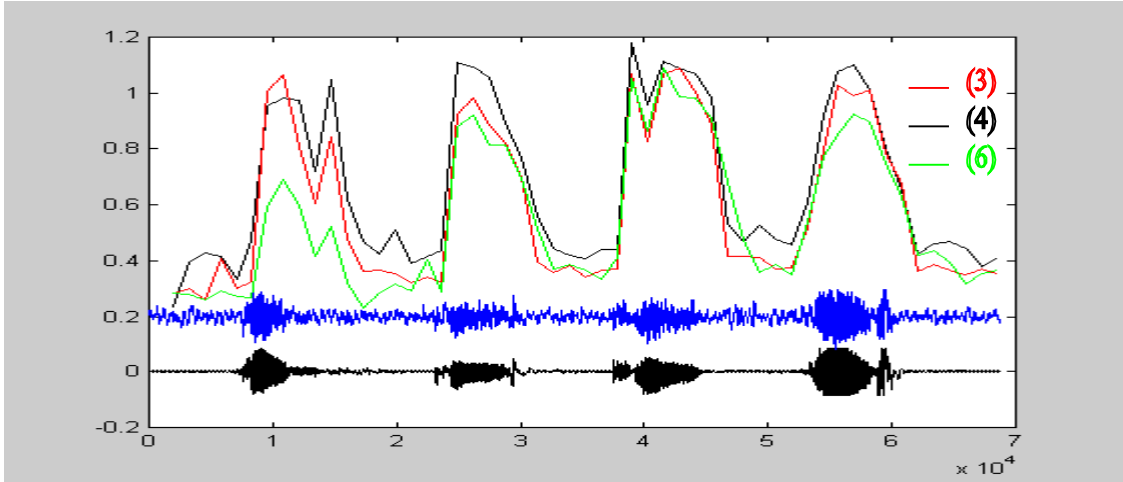
$$\frac{1}{M} \sum_{i=1}^M SNR \omega_i \begin{matrix} \geq T \\ < T \end{matrix} \begin{matrix} H_1 \\ H_0 \end{matrix}$$

$$SNR \omega_i = \max \left\{ \frac{X(j, \omega_i)^2 - \sigma(j, \omega_i)}{X(j, \omega_i)^2}, 0 \right\}$$

$$\sigma(j, \omega) = E[w(j, \omega)]$$

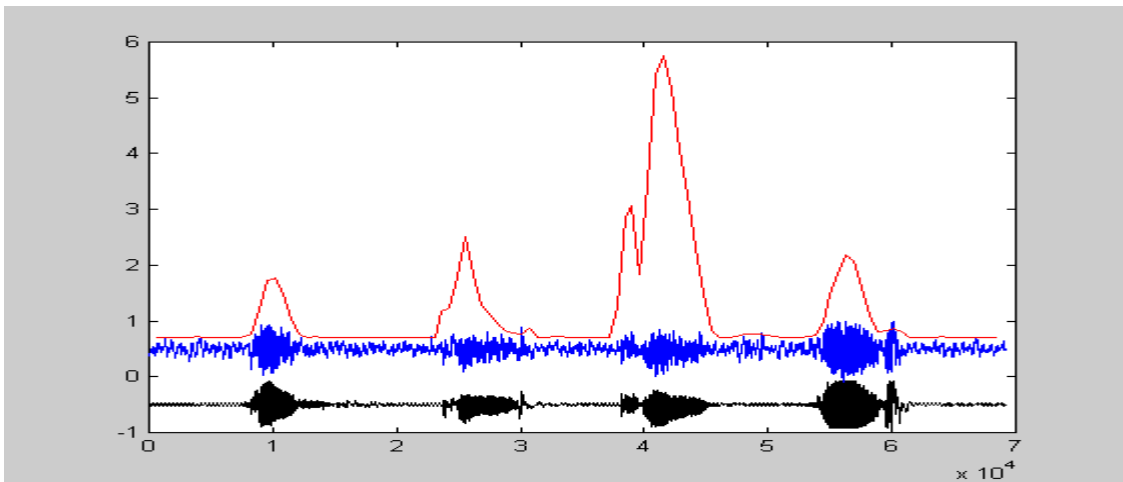
### 3. Deneysel Çalışmalar

Kullanılan iki farklı olabilirlik oran testinin karşılaştırmalı bir gösterimi Şekil 1’de verilmiştir. İki farklı varsayım ( $H_0$  ve  $H_1$ ) arasındaki geçişlerin saklı Gauss modeller (SGM) kullanarak istatistiksel olarak belirlenmesi yöntemi de Denklem 6’daki olabilirlik oran testi kullanılarak uygulanmış ve kestirime etkisi gözlenmiştir, sonuçlar Şekil 1’de gösterilmektedir.



Şekil 1: Temiz ses, toplamsal gürültü ile kirlenmiş ses ve farklı olabilirlik oran testi sonuçları.

Yukarıdaki uygulamayla aynı konuşma kaydı üzerinde permütasyon yöntemiyle tasarlanmış altbant süzgeç öbeği kullanılarak gerçekleştirilen aktivite sezimi uygulamasının sonuçları Şekil 2’de gösterilmiştir.



Şekil 2: Temiz ses, toplamsal gürültü ile kirlenmiş ses ve altbant yöntemiyle belirlenmiş aktif bölgeler.

## 4. Vargılar

Şekillerden görüleceği gibi AZFD'ye dayalı yöntemin başarımı altbantlara dayalı yöntemden yüksek olmaktadır. Her iki yöntem de temel olarak sıklık düzleminde gürültü modellemesi yapmakla birlikte AZFD'li yöntemin içerdiği bant sayısı dönüşümdeki katsayı sayısının yarısına eşit değerlendirilebilir. Öte yandan altbant süzgeç öbeğinin kullanımı işlem yoğunluğunu azaltarak gerçek zamanlı uygulamaları daha olası kılmaktadır. Şekil 1'den görülebileceği gibi saklı Markov modellere dayalı yöntem, AZFD'li diğer yöntemlere göre daha doğruluklu bir sezim olanağı sunmaktadır.

Konuşma aktivitesinin seziminde kullanılan istatistiksel yöntemler ikinci bölümde de söz edildiği gibi konuşma işaretinin doğası hakkında iki temel varsayım içermektedir. Bu varsayımlar konuşma işaretinin zaman ve sıklık düzlemindeki istatistiksel dağılımının Gauss olduğudur [2]. Her iki varsayım da özde çok sağlam temellere dayanmamaktadır. Sözkonusu varsayımları kullanmak yerine konuşma işaretinin her iki düzlemdeki istatistiksel dağılımını daha iyi modelleyecek bir yöntem üzerinde çalışmalar sürdürülmektedir. Gelecekteki çalışmalarda altbantlara dayalı (M-bant) konuşma aktivitesi sezim yöntemi üzerinde yoğunlaşılacaktır.

## Kaynakça

- [1] ITU-T Recommendation G.729–Annex B, 11/96.
- [2] Sohn J., Sung W., “A voice activity detector employing soft decision based noise spectrum adaptation”, Proc. ICASSP, s. 365-368, 1998.
- [3] Sohn J., Kim N.S., “A statistical model-based voice activity detection”, IEEE Signal Processing Letters, Cilt. 6, No. 1, s. 1-3, 1999.
- [4] Yang J., “Frequency domain noise suppression approaches in mobile telephone systems”, ICASSP'93, Cilt. 2, s. 363-366, 1993.
- [5] Ephraim Y, Malah D, “Speech enhancement using a minimum mean-square error short-time spectral amplitude estimator”, IEEE Trans. Acoust., Speech and Signal Processing, Cilt. ASSP-32, s. 1109-1121, 1984.
- [6] Alkın O., Çağlar H., “Design of efficient M-band coders with linear-phase and perfect-reconstruction properties.”, IEEE Trans. Signal Processing, Temmuz 1995
- [7] Çağlar H., Güntürk S., Anarım E., Sankur B., “Permutation based design of filter banks and orthogonal block transforms”, Multidimensional Signal Processing'de Temmuz 2000'de basılacak.