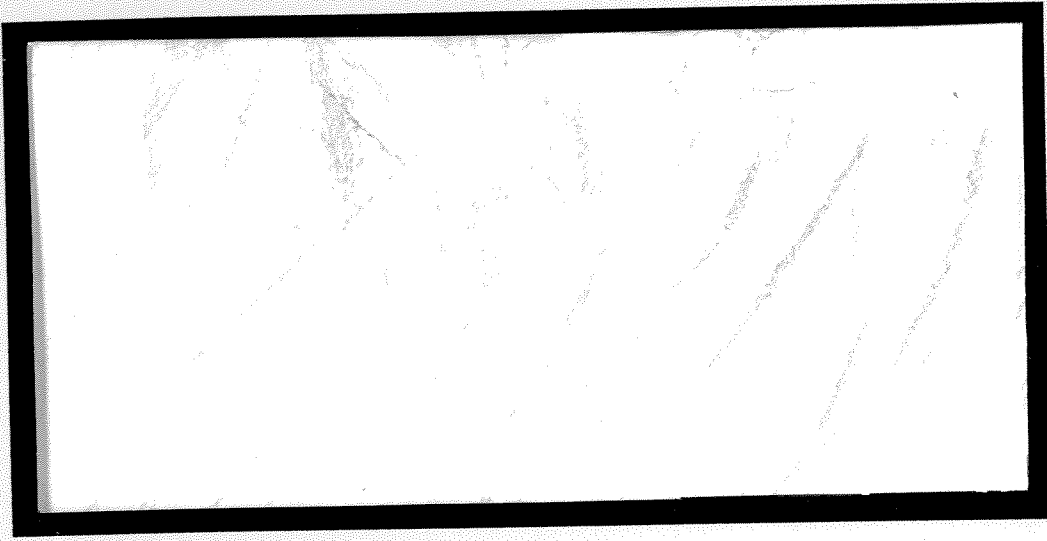




TÜRKİYE BİLİMSEL VE  
TEKNİK ARAŞTIRMA KURUMU

THE SCIENTIFIC AND TECHNICAL  
RESEARCH COUNCIL OF TURKEY

2002/95



Elektrik, Elektronik ve Enformatik Araştırma Grubu

Electric, Electronics and Informatics Research  
Grant Committee

**KONUŐMANIN GÜRÜLTÜDEN  
ARINDIRILMASI**

**PROJE NO : 100E004**

**Doç.Dr. Tolga ÇİLOĐLU**

**ŐUBAT 2002  
ANKARA**

## Önsöz

Günümüzde konuşma iletimine artan talep bu alandaki teknolojik gelişimi de hızlandırmıştır. Konuşmanın iletimden önce gürültü ile kirlenmesi duyum kalitesine ve anlaşılabilirliğe zarar vermektedir. Gürültülü konuşma eğer sıkıştırılarak iletiliyorsa bu tür olumsuz etkiler daha da belirginleşmektedir. Bu kapsamda gürültü etkilerinin azaltılması güncel bir sorun olarak ele alınmakta ve bir çok çalışma yapılmaktadır. İnsanın duyma işlevinin karmaşık ve tam olarak bilinmeyen yapısına bağlı olarak “güzel” ya da “iyi nitelikli” konuşmayı yeterince tanımlayan matematiksel modellerin olmayışı, bu alanda yapılan çalışmalardaki çeşitliliği ve yeni yaklaşımlara duyulan gereksinimleri de beraberinde getirmektedir.

Bu çalışmada, özellikle, sıkça karşılaşılabilen güçlü tonal bileşenlere sahip gürültülerin konuşma üzerindeki bozucu etkisinin azaltılması ele alınmıştır. Anında (“real-time”) çalışmaya elverişli yöntemlerin bu tür gürültülere karşı daha başarılı olmalarını sağlayacak yeni bir yaklaşım düşünülmüş ve uygulamaya yönelik ayrıntıları belirlenmiştir. Önerilen yaklaşımla elde edilen sonuçlar informal dinleme testleri ve objektif kalite ölçütleri kullanılarak değerlendirilmiştir. Önerilen yaklaşımın yarar sağladığı gözlenmiştir.

Sonuçların değerlendirilmesi aşamasında objektif ölçütlerin yazılımlarını bize sağlayan The Center for Spoken Language Research, University of Colorado, Boulder’dan Bryan Pellom’a teşekkür ederiz.

Bu çalışma Türkiye Bilimsel ve Teknik Araştırma Kurumu bünyesindeki Elektrik, Elektronik ve Enformatik Araştırma Grubunca desteklenmiştir.

<b>ÖNSÖZ.....</b>	<b>1</b>
<b>ÇİZİM LİSTESİ.....</b>	<b>3</b>
<b>TABLO LİSTESİ.....</b>	<b>4</b>
<b>ÖZ.....</b>	<b>6</b>
<b>ABSTRACT.....</b>	<b>7</b>
<b>KISALTMALAR .....</b>	<b>8</b>
<b>1 GİRİŞ .....</b>	<b>9</b>
<b>2 KONUŞMAYI GÜRÜLTÜDEN TEMİZLEME SİSTEMİ.....</b>	<b>12</b>
2.1 TANIMLAR VE GÖSTERİM .....	12
2.2 ÇERÇEVELERE AYIRMA VE PENCERELEME .....	13
2.3 LOGARİTMİK-EN KÜÇÜK-KARELER-ORTALAMA-İZGESEL-HATAYA GÖRE GENLİK KESTİRİMİ .....	14
2.4 GÜRÜLTÜ DEĞİŞİNTİSİNİN BULUNMASI .....	16
2.5 KONUŞMA ETKİNLİĞİNİN SAPTANMASI.....	17
2.6 ÖNCEL VE ARDIL SGO'LARIN BULUNMASI .....	20
2.7 KAZANÇ ALT SINIRININ GÜRÜLTÜ DEĞİŞİNTİSİNE BAĞLI OLARAK HESAPLANMASI... 21	
<b>3 SONUÇLAR .....</b>	<b>23</b>
3.1 OBJEKTİF DEĞERLENDİRME.....	32
3.2 İLERİYE YÖNELİK ÇALIŞMALAR .....	35
<b>KAYNAKÇA .....</b>	<b>35</b>



## Çizim Listesi

Çizim 1: Basitleştirilmiş akış şeması .....	12
Çizim 2: Konuşma var/yok kararı verme süreci .....	19
Çizim 3: Temiz konuşma örneği .....	25
Çizim 4: Gürültülü (SGO = -5 dB) konuşma örneği .....	25
Çizim 5: Gürültülü konuşmanın (SGO = -5 dB) Ephraim-Malah Yöntemi ile temizlenmiş hali .....	26
Çizim 6: Gürültülü konuşmanın (SGO = -5 dB) Ephraim-Malah Yöntemi ile önerilen yöntemin birlikte kullanılarak temizlenmiş hali .....	26
Çizim 7: Gürültülü konuşmanın (SGO = -5 dB) Ephraim-Malah Yöntemi ile temizlendikten sonra MELP'ten geçmiş hali .....	27
Çizim 8: Gürültülü konuşmanın (SGO = -5 dB) Ephraim-Malah Yöntemi ile önerilen yöntemin birlikte kullanılarak temizlendikten sonra MELP'ten geçmiş hali .....	27
Çizim 9: Temiz konuşma örneği .....	28
Çizim 10: SAAB gürültüsü ile kirlenmiş konuşma. SGO = -5 dB .....	29
Çizim 11: SAAB gürültüsü ile kirlenmiş konuşmanın MELP'ten geçmiş hali. SGO = -5 dB .....	29
Çizim 12: SAAB gürültülü konuşmanın (SGO = -5 dB) Ephraim-Malah yöntemi ile temizlenmiş hali .....	30
Çizim 13: SAAB gürültülü konuşmanın (SGO = -5 dB) Ephraim-Malah yöntemi ile birlikte önerilen yöntemin de kullanılmasıyla temizlenmiş hali .....	30
Çizim 14: SAAB gürültülü konuşmanın (SGO = -5 dB) Ephraim-Malah yöntemi ile Temizlendikten sonra MELP'ten geçmiş hali .....	31
Çizim 15: SAAB gürültülü konuşmanın (SGO = -5 dB) Ephraim-Malah yöntemi ile birlikte önerilen yöntemin de kullanılmasıyla temizlendikten sonra MELP'ten geçmiş hali .....	31

## Tablo Listesi

- Tablo 1: Temizlenmiş konuşmaların objektif olarak karşılaştırılmaları. SAAB gürültüsü; B-SGO = -5 dB. A: Ephraim-Malah yöntemi, B: Ephraim-Malah yöntemine önerilen yöntemin eklenmesi .....33
- Tablo 2: Temizlenmiş ve MELP kodlayıcı/çözücüden geçirilmiş konuşmaların objektif olarak karşılaştırılmaları. SAAB gürültüsü; B-SGO = -5 dB. A: Ephraim-Malah yöntemi, B: Ephraim-Malah yöntemine önerilen yöntemin eklenmesi .....33
- Tablo 3: Temizlenmiş konuşmaların objektif olarak karşılaştırılmaları. SAAB gürültüsü; B-SGO = 0 dB. A: Ephraim-Malah yöntemi, B: Ephraim-Malah yöntemine önerilen yöntemin eklenmesi .....33
- Tablo 4: Temizlenmiş ve MELP kodlayıcı/çözücüden geçirilmiş konuşmaların objektif olarak karşılaştırılmaları. SAAB gürültüsü; B-SGO = 0 dB. A: Ephraim-Malah yöntemi, B: Ephraim-Malah yöntemine önerilen yöntemin eklenmesi .....33
- Tablo 5: Temizlenmiş konuşmaların objektif olarak karşılaştırılmaları. SAAB gürültüsü; B-SGO = 5 dB. A: Ephraim-Malah yöntemi, B: Ephraim-Malah yöntemine önerilen yöntemin eklenmesi .....33
- Tablo 6: Temizlenmiş ve MELP kodlayıcı/çözücüden geçirilmiş konuşmaların objektif olarak karşılaştırılmaları. SAAB gürültüsü; B-SGO = 5 dB. A: Ephraim-Malah yöntemi, B: Ephraim-Malah yöntemine önerilen yöntemin eklenmesi .....33
- Tablo 7: Temizlenmiş konuşmaların objektif olarak karşılaştırılmaları. F16 gürültüsü; B-SGO = 5 dB. A: Ephraim-Malah yöntemi, B: Ephraim-Malah yöntemine önerilen yöntemin eklenmesi .....34
- Tablo 8: Temizlenmiş ve MELP kodlayıcı/çözücüden geçirilmiş konuşmaların objektif olarak karşılaştırılmaları. F16 gürültüsü; B-SGO = 5 dB. A: Ephraim-Malah yöntemi, B: Ephraim-Malah yöntemine önerilen yöntemin eklenmesi .....34
- Tablo 9: Temizlenmiş konuşmaların objektif olarak karşılaştırılmaları. F16 gürültüsü; B-SGO = 10 dB. A: Ephraim-Malah yöntemi, B: Ephraim-Malah yöntemine önerilen yöntemin eklenmesi .....34
- Tablo 10: Temizlenmiş ve MELP kodlayıcı/çözücüden geçirilmiş konuşmaların objektif olarak karşılaştırılmaları. F16 gürültüsü; B-SGO = 10 dB. A: Ephraim-Malah yöntemi, B: Ephraim-Malah yöntemine önerilen yöntemin eklenmesi .....34

Tablo 11: Temizlenmiş konuşmaların objektif olarak karşılaştırılmaları. F16 gürültüsü; B-SGO = 15 dB. A: Ephraim-Malah yöntemi, B: Ephraim-Malah yöntemine önerilen yöntemin eklenmesi .....	35
Tablo 12: Temizlenmiş ve MELP kodlayıcı/çözücüden geçirilmiş konuşmaların objektif olarak karşılaştırılmaları. F16 gürültüsü; B-SGO = 15 dB. A: Ephraim-Malah yöntemi, B: Ephraim-Malah yöntemine önerilen yöntemin eklenmesi .....	35

## Öz

Bu çalışmanın konusu gürültülü konuşmanın temizlenmesi, iyileştirilmesidir. Gürültü temizleme, konuşma kodlayıcıların başarımını da olumlu yönde etkiler. Bu çalışmada, özellikle güçlü tonal bileşenlere sahip gürültüler için gürültü temizleme sistemlerinin başarısının artırılması amaçlanmıştır. Buna yönelik olarak bir yöntem önerilmiş ve uygulanmıştır. Önerilen yöntem, tonal enerjinin dar bir frekans bandını kaplıyor olması ve bu bandın tamamen çıkarılması durumunda önemli ölçüde gürültü azaltımı sağlanması ile birlikte konuşma üzerinde önemli bir kayıp yaratmıyor olması gözlemine dayanmaktadır. Yöntem, sinyal-gürültü-oranının düşük olduğu yerlerde, yüksek gürültüyü daha fazla söndürme düşüncesini gerçekleştirmektedir. SAAB 340B pervaneli uçağının tonal bileşenleri güçlü olan gürültüsü ile F16 jet uçağı gürültüsü kullanılarak elde edilen sonuçların dinleme testlerine ve objektif ölçütlere göre değerlendirilmesi sonucunda yöntemin yararlı olduğu gözlenmiştir.

Anahtar sözcükler: Konuşma iyileştirme, konuşma temizleme, konuşma kalitesi, konuşma işleme, gürültü azaltımı, konuşma kodlama/sıkıştırma.

## Abstract

The subject of this work is the enhancement of noisy speech. Speech enhancement improves the performance of speech coders also. It has been aimed to improve the performance of speech enhancement systems especially under the presence noisy with strong tonal components. A method has been proposed and applied for this purpose. Proposed method is based on the fact that tonal energy exists in a narrow band and the observation that the complete removal of this band significantly lowers the residual noise while not yielding a considerable loss in speech. The method realizes the idea of providing more noise attenuation at higher noise levels in the regions of low signal-to-noise ratio. The method has been observed to be useful based on the listening tests and objective quality measures applied on the results obtained by using the noise of SAAB 340B propeller aircraft with strong tonal components and F16 noise.

Keywords: Speech enhancement, speech quality, noise reduction, speech coding.

## Kısaltmalar

AFD : Ayrık Fourier Dönüşümü (“Discrete Fourier Transform”)

YFD : Yerel Fourier Dönüşümü (“Short Time Fourier Transform”)

KES : Konuşma Etkinliği Saptayıcı (“Voice Activity Detector”)

SGO : Sinyal Gürültü Oranı (“Signal-to-Noise Ratio”)

P-SGO: Parçalarda Sinyal Gürültü Oranı (“Segmental SNR”)

WSS: “Weighted Spectral Slope”

LAR: “Log Area Ratio”

# 1 GİRİŞ

Bu araştırmanın konusu değişik kaynaklardan (motor, araç, pervane,...) gelen gürültü ile kirlenmiş konuşmaların gürültüden arındırılmasıdır. Gürültüden arındırmanın amacı duyum kalitesini ve anlaşılabilirliği artırmaktır. Gürültüden arındırmaya duyulan gereksinimin nedenlerinden bir diğeri de artan iletişim etkinliğinin ortaya çıkardığı sıkıştırılmış konuşma iletimidir. Gürültülü konuşmanın iletim için sıkıştırılması durumunda, sıkıştırma algoritmalarının ileri düzeydeki doğrusal olmayan etkileri nedeniyle iletilen konuşma kalitesi son derece yetersiz olabilmektedir. Sıkıştırılan gürültülü konuşma çözüldüğünde hem konuşma içeriği zayıflamakta hem de gürültünün özelliği değişerek genellikle daha kötüleşmektedir. Bu nedenle sıkıştırma işleminden önce konuşmanın temizlenmesi, sıkıştırma işleminden beklenen başarımın sağlanması bakımından önem taşımaktadır.

Bu çalışmanın temel amacı, işlem yükü, anında ('real time') çalışmaya elverişli olan yöntemlerin özellikle tonal gürültüye karşı başarımlarının artırılmasıdır. Güçlü tonal bileşenlere sahip gürültü türleri ile uygulamada oldukça sık karşılaşılmaktadır. Özellikle pervaneli uçaklar, kara araçlarının motor gürültüsü, üretim ortamlarındaki motorların, havalandırma düzeneklerinin gürültüleri güçlü tonal bileşenler içermektedir. Tonal gürültünün etkin olarak azaltılması, gürültü enerjisinin önemli bir kısmının bu bileşenlerde toplandığı durumlarda gürültü azaltımının temel amacı olarak düşünülebilir. Bu çalışmada tonal bileşenlerin ön plana çıkmasında, daha önce yapılan çalışmalara ilişkin bazı saptamalar ve gözlemler belirleyici olmuştur. Literatürde yer alan çalışmalarda, konuşma kalitesinin daha iyi ve kalan gürültünün daha düzenli/eşdağılımlı olması için gürültü azaltımını belirleyen izgesel kazanç ('gain') için bir alt sınır uygulaması benimsenmiştir. Böylece kalan gürültü bilinçli olarak belirli bir düzeyin üstünde tutulmaya çalışılmakta ve düzensizlik gösteren artık gürültü psikoakustik nedenlere bağlı olarak maskelenmektedir. Yaptığımız deneyler sırasında, gürültü belirli frekanslarda diğerlerine göre çok daha fazla enerjiye sahip olduğunda (tonal bileşenli gürültülerde olduğu gibi), uygulanan alt sınır nedeniyle bu güçlü bileşenlerin belirgin olarak varlıklarını sürdürebildikleri görülmüştür. Oysa, bir kaç harmonikten oluşan tonal gürültü, konuşma sinyaline ait izgenin ('spectrum') oldukça küçük bir kısmını kapladığı için, tonal gürültüyü oluşturan bileşenler çakıştıkları konuşma bileşenleri ile birlikte atılsalar bile konuşmada önemli bir kayıp oluşmamaktadır. Bu saptamadan yola çıkarak, çakışan konuşma

bileşenlerine verilen zararı en azda tutmaya çalışarak tonal gürültüyü oluşturan bileşenlerin daha fazla azaltılması amaçlanmaktadır.

Bu projenin bir başka amacı da etkin çalışan bir konuşma etkinliği saptayıcı (KES) biriminin oluşturulmasıdır. Bu konuda çalışmalar yapılmıştır.

Yapılan çalışmalar literatürdeki yöntemlerin taranmasını, bunlardan amaca uygun olanlara ilişkin yazılımların gerçekleştirilerek bu yöntemlerin karşılaştırılmasını, değişik türdeki gürültülerin internet ortamından ve yapılan kayıtlarla elde edilmesini içermektedir.

Bu projenin önerisinde yer almamış olmakla birlikte, paralel bir çalışma olarak, gürültü azaltma sisteminin TMS320C54 sabit-nokta işlemcisi üzerinde çalışan yazılımları da hazırlanmıştır.

Konuşma temizleme alanındaki temel yaklaşımlardan birisi izge katsayılarına ait genlik değerlerinin, gürültü etkisini azaltacak şekilde değiştirilmesidir. Değişik ve geliştirilmiş halleri ile 'spectral subtraction' [1] ve diğer izgesel genlik kestiriciler bu sınıfa girerler. Bu yöntemler izge genliklerini değiştirirken taban gürültüsüne gereksinim duyarlar. Bu bilgi genellikle bir KES birimi aracılığıyla saptanır. Bu çalışmanın tamamlandığı sıralarda bir KES birimine gereksinim duymadan taban gürültüsü hakkında bilgi toplayabilen bir yöntem de önerilmiştir [2]. Sözkonusu yöntemler düşük işlem yükleri ve yeterli başarımları nedeniyle anında çalışan uygulamalar için çekici olmuşlardır.

Diğer taraftan, konuşma sinyallerine özgü özelliklerin öncel bilgi ('a priori information') olarak kullanılması fikri model tabanlı yöntemleri ortaya çıkarmıştır. Hansen ve Clements [3], Lim ve Oppenheim yaklaşımına [4] göre elde edilen yalnızca-kutuplu model parametreleri üzerinde genel konuşma sinyali özelliklerini göz önüne alarak bazı kısıtlamalardan yararlanmışlardır. Hansen ve Arslan, daha ayrıntılı bir yol izleyerek, fonem ya da fonem sınıflarına özgü kısıtlamaları/özellikleri 'Hidden Markov Model' (HMM) tekniği ile konuşma temizleme sürecine yerleştirmişlerdir [5]. Sreenivas [6], aynı amaç için, farklı bir yöntem olarak, yalnızca-kutuplu model parametrelerini bulabilmek için bir 'codebook'tan ve vektör nicemlemeden yararlanmışlardır.



Özellikle, yukarıda ilk grupta yer alan yöntemlerde ‘müziksel gürültü’ olarak bilinen sorun ortaya çıkmaktadır. Bu gürültü zaman ve frekansta rasgele dağılmış dar bantlı enerji kümecikleridir. Bu oluşum ardışık çerçevelere ait izgeler arasındaki tutarsızlıklardan kaynaklanmaktadır. Çerçeveler üzerinde ortalama alınması bu tutarsızlığı azaltmakta ancak konuşmada bozulmaya neden olmaktadır [1]. Öncel SGO’yu kestirmek için Ephraim ve Malah’ın önerdiği karardan-yönlendirmeli (‘decision-directed’) yaklaşım [7] (bir tür doğrusal olmayan ortalama işlemi, [8]) müzik gürültüsü sorununu önemli ölçüde hafifletmiştir. Ayrıca, kalan gürültü miktarının bilinçli olarak belirli bir düzeyin üzerinde bırakılması da müziksel gürültüye karşı basit ve etkili bir yol olmuştur [7]-[9]. Uygulamada, izgesel genlikleri değiştirmek için kullanılan kazanca alttan sınırlama uygulanarak bu amaca ulaşılmaktadır.

Daha önce de belirtildiği gibi, gürültünün güçlü tonal bileşenlere sahip olması durumunda, kazancın alttan sınırlanması bu bileşenlerin belirginliklerini korumalarına yardımcı olmaktadır. Tonal bileşenler ‘notch’ süzgeç ile yok edildiğinde, konuşmanın da aynı frekanslardaki bileşenleri yok olmakta ancak konuşmanın yaygın frekans izgesi nedeniyle bu eksiklik konuşma kalitesi ve anlaşılabilirliği yönünden önemli bir fark yaratmamaktadır. Bu durumda uyarlamalı ‘notch’ süzgeç [10] ile soruna çözüm bulunabilir ancak farklı gürültü türleri için bu yöntem başarısız olacaktır. ‘Notch’ süzgecin yalnızca tonal gürültü ağırlıklı durumlarda kullanılması durumunda da, konuşmanın harmonikli izgesel yapısı nedeniyle yakınsama sorunları öne çıkacaktır. Bu sorun, farklı konuma sahip ikinci bir mikrofondan alınan sinyal ile giderilebilir ama bu da uygulama açısından önemli bir fiziksel güçlük doğurur.

Bu çalışmada önerilen yöntem, genel amaçlı gürültü azaltma yaklaşımını korurken, tonal gürültü bileşenlerini, konuşmanın aynı frekanstaki bileşenlerini ‘notch’ süzgeç durumunda olduğu gibi yok etmeden, daha güçlü olarak bastırabilmektedir. İzgesel genlik kazancı için sabit bir alt sınır kullanmak yerine, kazancı belirli bir eşik değerinin altında kalan bileşenler için, o frekanslardaki gürültü düzeyine bağlı bir kazanç saptanmaktadır. Bu yaklaşımda gürültü tonları ile çakışan konuşma bileşenleri ‘notch’ süzgeçte olduğu kadar zarar görmemektedir. Ayrıca gürültü tonlarının izlenmesi (‘tracking/adaptation’) için ayrı bir işleme hiç gerek duyulmamaktadır.

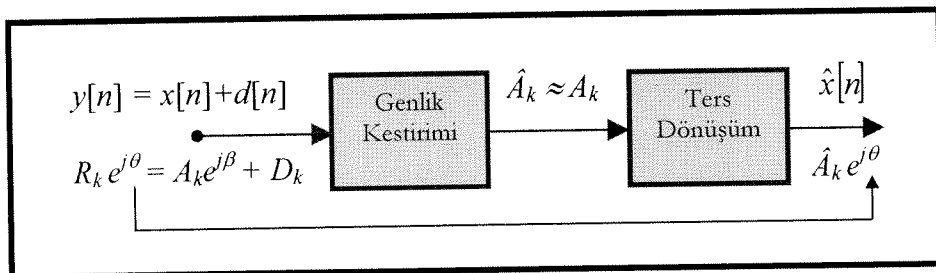
Konuşmanın gürültüden arındırılması amacıyla kullanılan izgesel zayıflatma yöntemleri, gürültülü konuşmanın sinyal-gürültü-oranı'na (SGO) bağlı olarak değişen bir izgesel kazanç fonksiyonuna göre konuşmanın gürültüden arındırılmasını sağlamaktadır. Söz konusu yöntemlerin işleyişi şöyle özetlenebilir; bir konuşma etkinliği saptayıcı (KES) yardımıyla konuşma olmayan bölgelerde gürültüye ait ortalama izgesel enerji hesaplanır. Bu bilgi kullanılarak tüm frekans bileşenlerindeki anlık (bir çerçeve için) SGO bulunur. Her frekans bileşeni SGO'ya bağlı bir fonksiyonun belirlediği ölçüde değişikliğe (azaltıma) uğrattıldıktan sonra çerçeve sinyali yeniden oluşturulur. SGO'ya bağlı azaltım fonksiyonu için Wiener [11], en küçük ortalama karesel hata [11]-[13], en büyük olabilirlik [11] gibi farklı ölçütler kullanılmıştır. Proje çalışmalarının ilk aşamalarında bu yöntemleri gerçekleştiren yazılımlar hazırlanmış ve denemeler yapılmıştır. Proje ara raporunda bu gelişmeler sunulmuştur.

Bu raporda, ortaya çıkan konuşma temizleme sisteminin bütünü anlatılmakta ve önerilen yenilikler sunulmaktadır. Yapılan çalışmalar başarılı olmuş, konuşma kalitesinde azalmaya neden olmadan kalan gürültü düzeyinde azalma sağlanmıştır. Sonuçlar spektrogram görüntüleri ile karşılaştırmalı olarak sunulmaktadır.

## 2 KONUŞMAYI GÜRÜLTÜDEN TEMİZLEME SİSTEMİ

### 2.1 Tanımlar ve Gösterim

Yazımda kullanılan semboller Çizim 1'deki basitleştirilmiş akış şeması üzerinde görülmektedir.  $y[n]$  gürültülü giriş sinyalini,  $x[n]$  temiz konuşma sinyalini,  $d[n]$  gürültüyü;  $R_k$  ve  $\theta$ , sırasıyla,  $y[n]$ 'nin YFD genliğini ve evresini,  $A_k$  ve  $\beta$ , sırasıyla, temiz konuşmanın YFD genlik ve evresini;  $D_k$  gürültünün YFD değerini,  $\hat{A}_k$ ,  $A_k$ 'nin kestirilmiş değerini,  $\hat{x}[n]$  temizlenmiş konuşma sinyalini ve  $\hat{A}_k e^{j\theta}$  onun YFD değerini göstermektedir. Bu gösterimde  $k$ , YFD'nin frekans indeksidir.



Çizim 1: Basitleştirilmiş akış şeması.

$$\lambda_x(k) = E\{A_k^2\} \quad \lambda_d(k) = E\{|D_k|^2\} \quad (1)$$

$\lambda_x(k)$  ve  $\lambda_d(k)$ , sırasıyla, konuşma ve gürültüye ait  $k$ 'inci bileşenlere ait değişinti değerlerini göstermektedir.

$$\eta_k = \frac{\lambda_x(k)}{\lambda_d(k)} \quad (2)$$

$$\gamma_k = \frac{R_k^2}{\lambda_d(k)} \quad (3)$$

$\eta_k$  ve  $\gamma_k$ , sırasıyla, öncel ve ardıl SGO olarak anılır.  $\lambda_x(k)$  ve  $\lambda_d(k)$ , sırasıyla, konuşma ve gürültüye ait  $k$ 'inci bileşenlere ait değişinti değerlerini göstermektedir.

## 2.2 Çerçevelere Ayırma ve Pencereleme

8 KHz'de örneklenmiş olan gürültülü konuşma sinyali,  $y[n]$ , 256 örnekten oluşan çerçeveler halinde,  $\mathbf{y}$ , işleme alınır. Ardışık çerçeveler arasında %50 ya da % 75 oranında örtüşme için denemeler yapılmıştır. Genel olarak, artan çakışma oranı kaliteyi artırmakla birlikte, uygulamada bu seçim işlem yükü kapasitesine göre yapılır. Belirli bir pencere işlevi,  $\mathbf{w}$ , ile pencerelenen giriş sinyali,  $\mathbf{y}_p$ ,

$$\mathbf{y}_p = \mathbf{w}^T \mathbf{y} \quad (4)$$

izleyen bölümde açıklandığı gibi genlik kestirme kuralına göre işlendikten sonra elde edilen 256 örneklilik çerçeve,  $\mathbf{z}$ , ile kendinden önce ve sonra gelen çerçevelere, kullanılan örtüşme oranına göre, örtüştür-topla işlemi uygulanarak  $\hat{x}[n]$  elde edilir. Örtüştür-topla işlemi, genellikle,  $\mathbf{z}$  çerçevelerinin örtüşük toplamının elemanlarının, pencere işlevinin örtüşük toplamının elemanlarına bölünmesi ile gerçekleştirilir. En küçük-ortalama-karesel-hatayı en aza indirmeye dayalı örtüştür-topla işlemi ise,  $\mathbf{z}$  çerçevelerinin pencere işlevi ile çarpıldıktan sonra elde edilen örtüşük toplamının, pencere işlevinin karesinin örtüşük toplamına bölünmesi ile gerçekleştirilmektedir [14]. İşlem yükü bakımından, geleneksel yaklaşım kullanıldığında pencere işlevinin örtüşük toplamının sabit olması istenirken, diğer yaklaşım kullanıldığında pencere işlevinin karesinin örtüşük toplamının sabit olması yeğlenmektedir.

Sıklıkla kullanılan pencere işlevlerinden Hamming ve Hanning pencereler, ne karelerinin örtüşük toplamı ne de kendilerinin örtüşük toplamı için (%50 ve %75 örtüşmeler için) sabit bir değer üretmezler. Bu pencere işlevlerinin örtüşük toplamlarının ortalama değerden en büyük sapmaları Hamming ve Hanning için sırasıyla, %75 örtüşme için %0.05 ile %0.06, ve %50 örtüşme için % 0.26 ile % 0.3 olmaktadır. Küçük sayılabilecek bu dalgalanma nedeniyle örtüşük toplamlar istenirse sabit olarak da kabul edilebilirler.

Öte yandan değiştirilmiş ('Modified') Hamming ya da değiştirilmiş Hanning pencerelerin örtüşük toplamı her iki örtüşme oranı için de sabit bir değer üretmekte, karelerinin örtüşük toplamı ise %75 örtüşme için sabit olurken %50 örtüşme için sırasıyla, %27 ve %33 gibi büyük dalgalanmalar göstermektedirler.

Yukarıdaki göstergelerin ışığında, bu çalışmada değiştirilmiş Hamming pencere işlevi geleneksel örtüşük-topla yöntemi ile kullanılmıştır. Ayrıca, geleneksel örtüşük-topla yöntemi ile en küçük-karesel-hataya göre yapılan örtüşük-topla arasında algılanabilir bir farklılık gözlenmemiştir.

### 2.3 Logaritmik-En Küçük-Kareler-Ortalama-İzgesel-Hataya Göre Genlik Kestirimi

Temiz konuşmaya ait genlik izgesinin bulunmasında en büyük olabilirlik, Wiener, en küçük-ortalama-karesel-hata ya da logaritmik-en küçük-karesel-ortalama-izgesel-hata gibi ölçütler kullanılmıştır. Bu tür ölçütlere göre elde edilen genlik kestiriciler proje ara raporunda incelenmiştir. Burada yalnızca, yapılan denemelere göre bu projede temel olarak alınan logaritmik-en-küçük-karesel-ortalama-izgesel-hatayı enazlayan genlik kestirici ele alınacaktır.

Konuşma için benzerlik ölçütü olarak en küçük-kareler-ortalama-izgesel-hata yerine *logaritmik-en küçük-kareler-ortalama-izgesel-hatanın* (LEKÖİH) konuşmadaki bozulmanın insan kulağı tarafından algılanan halini daha iyi modellediği bilinmektedir [12],[13]. Bu ölçüte göre temiz konuşma genliği  $A_k$ ,

$$E\{(\log A_k - \log \hat{A}_k)^2\} \quad (5)$$

ile tanımlanan bozulma ölçütünü en küçük yapan  $\hat{A}_k$ 'nin bulunması ile kestirilmektedir.  $H_k^1$ 'i  $k$ 'inci frekansta konuşmanın var olduğu ve  $H_k^0$ 'ı  $k$ 'inci frekansta konuşma olmadığı

hipotezleri olarak tanımlayalım. (25)'te ifade edilen ölçüte göre temiz konuşmanın genlik kestirimi

$$\hat{A}_k = \exp\left(E\left\{\log \hat{A}_k | Y_k, H_k^1\right\} G_e(k)\right) = \left(G_{\log}(k) R_k\right)^{G_e(k)} \quad (6)$$

olmaktadır [12]. Burada  $G_{\log}(k)$  LEKOH genlik kestiricisi olup (7)'deki ifadeye sahiptir

$$G_{\log}(k) = \frac{\xi_k}{1 + \xi_k} \exp\left(\frac{1}{2} \int_{v_k}^{\infty} \frac{e^{-t}}{t} dt\right) \quad (7)$$

$$\xi_k = \frac{\eta_k}{1 - q_k}, \quad v_k = \frac{\xi_k}{1 + \xi_k} \gamma_k, \quad q_k: k'inci fekansta konuşma olmama olasılığı;$$

$G_e(k)$  esnek karar ('soft decision') çarpanıdır,

$$G_e(k) = P(H_k^1 | Y_k) \quad (8)$$

ve Bayes kuralı kullanılarak, olabilirlik oranı,  $\Lambda$ ,

$$\Lambda(k) = \mu_k \frac{P(Y_k | H_k^1)}{P(Y_k | H_k^0)} \quad (9)$$

$$\mu_k = \frac{1 - q_k}{q_k}$$

ile bağlantılı olarak

$$G_e(k) = \frac{\Lambda(k)}{1 + \Lambda(k)} \quad (10)$$

ifadesi elde edilebilmektedir. Olabilirlik oranı  $\Lambda$ 'nın açık ifadesi ise

$$\Lambda(k) = \mu_k \frac{\exp(v_k)}{1 + \xi_k} \quad (11)$$

olarak elde edilmektedir [12].

(6)'daki ifadenin aşağıdaki gibi değiştirilmesinin daha iyi sonuçlar verdiği deneysel olarak gözlenmiştir [13].

$$\hat{A}_k = \left(G_{\log}(k) G_e(k)\right) R_k = G(k) R_k \quad (12)$$

Buradaki  $G(k)$ , izleyen kısımlarda LEKOİH genlik kestiricisi olarak anılacaktır.

Yukarıdaki ifadeler incelendiğinde, temiz konuşma izgesinin genliğinin hesaplanmasında her frekans için şu üç parametrenin bulunması gerektiği görülmektedir:

- gürültünün her frekanstaki değışintisi,  $\lambda_k^d$ ,
- öncel SGO,  $\eta_k$ ,
- $k$ 'inci fekansta konuşma olmama olasılığı,  $q_k$ .

Bu parametrelerin bulunması izleyen bölümlerde açıklanmaktadır.

## 2.4 Gürültü Değişintisinin Bulunması

Gürültü değışintisi, problemin tanımı gereği yalnızca gürültü sinyaline ait bilgi sağlayabilen bir kaynak olmaması nedeniyle gürültülü konuşmadan elde edilmek durumundadır. Burada, konuşmanın süreklilik göstermeyen bir etkinlik olmasından yararlanılmaktadır. Bir çerçeve sinyali içinde, özellikle düşük SGO düzeylerinde, konuşma sinyaline ait enerjinin az olduğu aralıklarda konuşma bulunup bulunmadığının saptanması ayrı bir araştırma konusudur. Akan çerçevelerde konuşmaya ait var/yok kararı verildikten sonra konuşma olmayan çerçeveler gürültü değışintisinin hesaplanması için kullanılırlar. Hesaplama yönteminin değışen ('nonstationary') gürültü özelliklerini izlemeye elverişli olması aranan özelliklerden biridir.

Bir konuşma etkinliği saptayıcı (KES) yardımıyla yalnızca gürültü içeren çerçevelerin saptandığını varsayalım.  $m$ 'inci gürültü çerçevesi için  $k$ 'inci frekanstaki gürültü değışintisi  $\lambda_k^d(m)$  olsun. Birinci dereceden bir yineleme ile  $\lambda_k^d(m)$  aşağıdaki gibi elde edilebilir:

$$\lambda_k^d(m) = \alpha_d \lambda_k^d(m-1) + (1 - \alpha_d) R_k^2(m); \quad 0 < \alpha_d < 1 \quad (13)$$

Bu işlem alçak geçiren süzgeçlemeye karşılık gelmektedir.  $\alpha_d$  sabiti süzgecin kesim frekansını belirler.  $\alpha_d$ 'nin bire yakın değeri gürültü değışintisinin yenilenme hızını yavaşlatır. Özellikleri yavaş değışen bir gürültü için  $\alpha_d$  bire yakın seçilebilir ancak herhangi bir anda gürültünün değışmeye başlaması durumunda, bu değışimin izlenemeyecek ve gürültü azaltımındaki hatalar artacaktır. Gürültüdeki değışimi daha hızlı izleyebilmek için  $\alpha_d$ 'nin küçük seçilmesi durumunda ise YFD'nin yakın çerçeveler için ürettiği tutarsız değeri

nedeniyle ardışık  $\lambda_k^d(m)$  değerlerindeki dalgalanma artmakta, bu da  $\lambda_k^d(m)$ 'ye bağlı olarak elde edilen SGO değerlerinin dalgalanmasına ve müziksel gürültünün artmasına yol açmaktadır. Bu çelişen gereksinimleri karşılayabilmek için nedeniyle  $\alpha_d$ 'nin uyarlamalı olarak değiştirilmesi için bir yöntem [13]'de önerilmiştir. Amaç, gürültü özelliklerinin değişim gösterdiği yerlerde  $\alpha_d$ 'nin küçültülmesi buna karşılık gürültü özelliklerinin durağanlaştığı yerlerde de  $\alpha_d$ 'yi tekrar artırmaktır. Buna uygun bir  $\alpha_d$  değiştirme kuralı olarak

$$\alpha_d(m) = 1 - F|1 - \bar{\gamma}(m-1)| \quad (14)$$

düşünülmüştür [13]. Bu ifadede  $F$  bir sabit ve  $\bar{\gamma}(m-1)$  yalnızca gürültü içeren bir önceki çerçeve sinyalinden tüm frekanslarda elde edilen ardıl SGO'ların ortalamasıdır. Gürültü özellikleri önemli ölçüde değişiklik göstermediği zaman, yalnızca gürültü içeren çerçevelerde ardıl SGO değerleri bire yakın değerler alırlar. Böylece  $\alpha_d$  bire yakın bir değere eşitlenir ve (14)'deki yenileme yavaşlar. Buna karşılık, gürültü özelliklerinde ortaya çıkan bir değişiklik  $\bar{\gamma}(m-1)$ 'in birden uzaklaşmasına ve böylece  $\alpha_d$ 'nin küçülmesine neden olur. Küçülen  $\alpha_d$  gürültüdeki değişimin daha hızlı izlenmesini sağlar.

(14) ifadesinde  $F$  sabiti için 0.2 uygun bir değer olarak belirlenmiştir. Ayrıca bu ifadede  $\alpha_d$  için elde edilen değerlerin [0.8 , 0.98] aralığının dışına çıkmasına izin verilmemektedir.  $\alpha_d$ 'nin 0.8'den küçük değerleri, yukarıda açıklanan nedenlerden dolayı elde edilen çıktının kalitesini düşürmektedir.

## 2.5 Konuşma Etkinliğinin Saptanması

Yalnızca gürültü içeren çerçevelere karar vermek için logaritmik olabilirlik oranı kullanıldığında Itakura-Saito bozulma ölçütünün ayırık *yaklaşık* değeri elde edilmektedir. Bunun için, yalnızca gürültünün ve gürültülü konuşmanın Gauss yoğunluk işlevlerine sahip oldukları varsayıldığında aşağıdaki karşılaştırma ortaya çıkmaktadır, [15],

$$\frac{1}{N} \log \frac{p(\mathbf{R}|H_1)}{p(\mathbf{R}|H_0)} = \frac{1}{N} \sum_k \left\{ \frac{R_k^2}{\lambda_d^k} - \log \frac{R_k^2}{\lambda_d^k} - 1 \right\} \begin{matrix} H_1 \\ > \\ < \\ H_0 \end{matrix} \sigma \quad (15)$$

Burada  $\mathbf{R}$  giriş sinyalinin izge katsayılarını içeren vektörü ve  $N$  AFD uzunluğudur. (14)'deki logaritma terimi ancak  $\frac{R_k^2}{\lambda_d^k} = \gamma_d^k$  değerinin birden çok küçük olması durumunda etkili olacağı ve yalnızca gürültüden oluşan çerçevelerde bile bunun mümkün olmaması nedeniyle yok sayılabilir; (14)'deki ifade yaklaşık olarak (15)'deki gibi kabul edilebilir.

$$D = \frac{1}{N} \sum_k \frac{R_k^2}{\lambda_d^k} \begin{matrix} > \\ < \end{matrix} \begin{matrix} H_1 \\ H_0 \end{matrix} 1 + \sigma \quad (16)$$

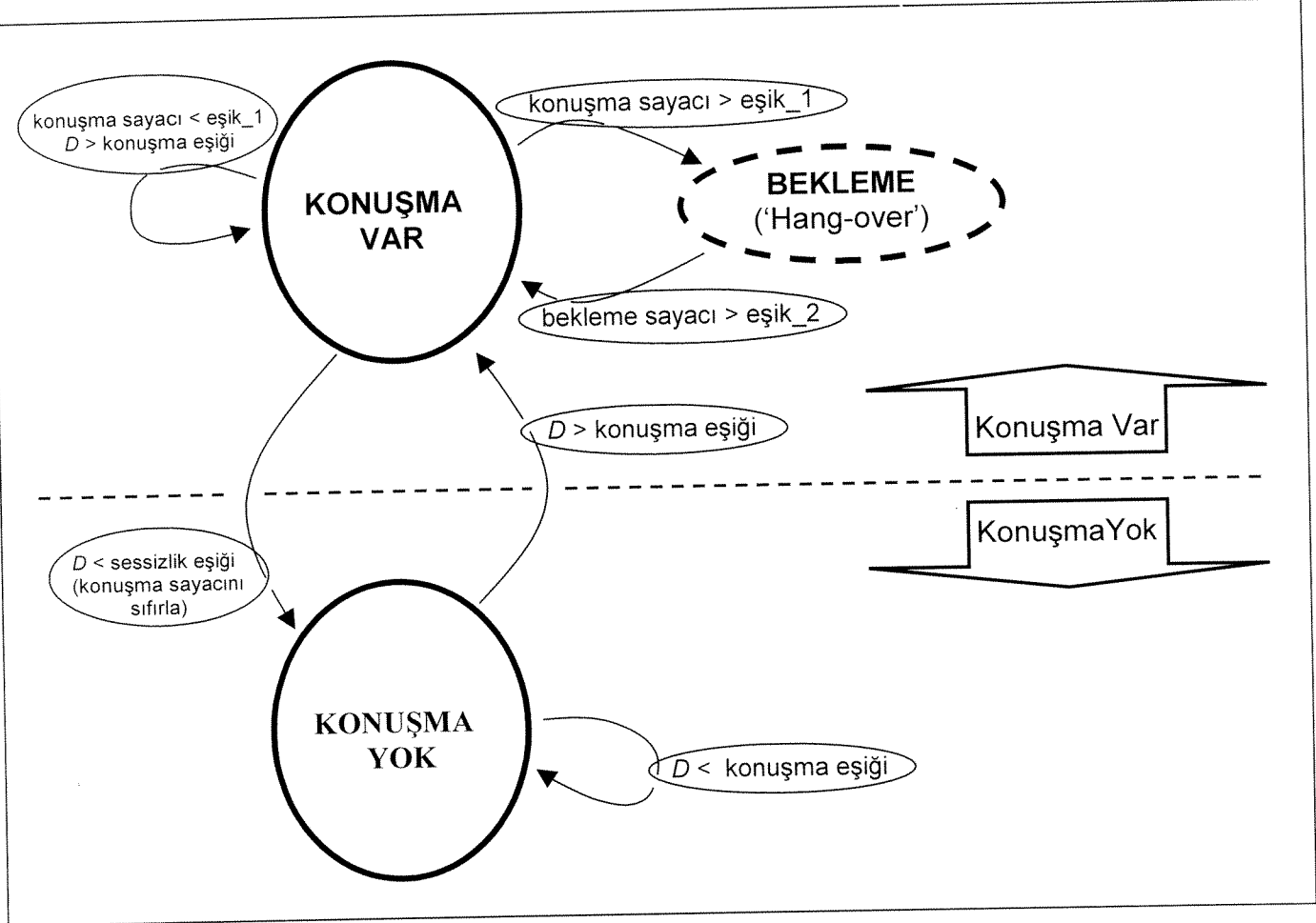
Eşik değerinin 1.2-2 arasındaki değerlerinin iyi sonuçlar verdiği saptanmıştır.

Bir çerçevede konuşma var/yok kararı verilirken yalnızca o çerçeveye ait bilginin kullanılması yerine çerçeveler arasında bağlantılı olarak karar vermenin yararlı olacağı düşüncesiyle buna uygun bir yaklaşım geliştirilmiştir. Özellikle düşük enerjiye sahip konuşma içeren çerçevelerde hata yapma olasılığı yüksektir. Konuşma etkinliği sırasında düşük enerjili konuşma bölgeleri ile yüksek enerjili konuşma bölgelerinin genellikle ard arda geliyor olmasından dolayı, (15)'deki testte kullanılan ardıl SGO'nun eşige yakın değerler aldığı çerçevelerde hemen karar vermek yerine önceki çerçevelere ait ardıl SGO'lar da dikkate alınmaktadır. Karar için önceki çerçevelerin yanısıra sonraki çerçevelerin de kullanılmasının yararlı olması beklenir. Ancak sonraki çerçevelerin kullanılması durumunda işlem gecikmeli olarak tamamlanabilir. Anında işleyen sistemler için sakıncalı olduğundan bu yaklaşım tercih edilmemiştir.

Bu çalışmada ortaya çıkan konuşma var/yok kararı verme yöntemi Çizim 2'de görülmektedir. Çizim 2'de 'KONUŞMA VAR', 'KONUŞMA YOK' ve 'BEKLEME' adlı üç durum görülmektedir. Sistem, 'BEKLEME' durumunda 'KONUŞMA VAR' durumundaki gibi davranmaktadır; gürültü bilgisi toplanmamaktadır. Sistem 'KONUŞMA VAR' durumunda verdiği her 'konuşma var' kararından sonra konuşma sayacını artırır. 'KONUŞMA VAR' durumunda belli bir süreden daha fazla kalırsa (konuşma sayacı > eşik\_1) 'BEKLEME' durumuna geçer ve başka herhangi bir koşula bakmadan belirli bir süre (bekleme sayacı > eşik\_2) bu durumda kalır. Böylece konuşma sırasında yüksek enerjili bölgeleri izleyen düşük enerjili bölgelerde verilebilecek yanlış kararlara bağlı hatalı (konuşma bileşenleri içeren) gürültü bilgisi toplanmamış olur. 'BEKLEME' durumunda konuşma yok ise gürültü bilgisinin yenilenmesi aksar. Genellikle



olduğu gibi özellikleri çok hızlı değişmeyen gürültü türlerinde bunun bir sakıncası bulunmamaktadır. 'BEKLEME' durumundan 'KONUŞMA VAR' durumuna dönen sistem (15)



Çizim 2: Konuşma var/yok kararı verme süreci.

yine 'konuşma var' kararı verirse tekrar beklemeye döner; bu sırada konuşma sayacı sıfırlanmaz.

'KONUŞMA VAR' durumundayken (15)'teki testin konuşma yok kararı vermesi ile ( $D < \text{sessizlik eşığı}$ ) sistem 'KONUŞMA YOK' durumuna geçer. Bu durumdayken konuşma var kararı verilirse ( $D > \text{konuşma eşığı}$ ) sistem 'KONUŞMA VAR' durumuna geçer. 'KONUŞMA VAR' ve 'KONUŞMA YOK' durumları arasındaki geçişte iki farklı eşik (konuşma eşığı ve sessizlik eşığı) kullanılmaktadır. Sessizlik eşığı, konuşma eşığı'ndan daha küçük bir değere sahiptir. Böylece, 'KONUŞMA VAR' ve 'KONUŞMA YOK' durumları arasında küçük değişimlerden dolayı oluşacak salınımlar engellenmektedir.

## 2.6 Öncel ve Ardıl SGO'ların Bulunması

SGO değerleri giriş sinyalinin izgesel katsayılarının genliklerine uygulanacak zayıflatmaların saptanması için gerekmektedir. Ardıl SGO,  $\gamma_k$ ,

$$\gamma_k = \frac{R_k^2}{\lambda_d(k)} \quad (17)$$

olarak tanımlanmıştır. Burada  $R_k^2$  doğrudan giriş sinyalinden,  $\lambda_d(k)$  ise Bölüm XX'de açıklandığı gibi bulunmaktadır.

Öncel SGO,  $\eta_k$ ,

$$\eta_k = \frac{\lambda_x(k)}{\lambda_d(k)} \quad (18)$$

tanımında ise temiz konuşma sinyalinin değişintisine,  $\lambda_x(k)$ , gereksinim duyulmaktadır. Giriş sinyalinden doğrudan elde edilemeyen bu değer için bir kestirme yöntemine gerek vardır. Bu kestirim için literatürde yaygın olarak kullanılan ölçüt en büyük olabilirlik olmuştur. Buna göre, öncel SGO'nun en büyük olabilirlik kestirimi

$$\eta_k = \gamma_k - 1 \quad (19)$$

biçiminde elde edilebilmektedir [12]. Uygulamada (18)'deki anlık değerlerin kullanılması, Fourier dönüşümünün ardışık çerçevelerde tutarsız değerler üretmesi nedeniyle tercih edilmemektedir. Öncel SGO için de birinci dereceden yinelemeli bir türetim yararlı olmaktadır. Buna göre,  $m$  çerçeve indeksi olmak üzere,

$$\bar{\gamma}_k(m) = \alpha \bar{\gamma}_k(m-1) + (1-\alpha)\gamma_k(m) \quad ; \quad 0 < \alpha < 1 \quad (20)$$

ile ardıl SGO'nun yumuşatılmasından sonra

$$\eta_k(m) = \begin{cases} \bar{\gamma}_k(m) - 1 & \bar{\gamma}_k(m) - 1 \geq 0 \\ 0 & \bar{\gamma}_k(m) - 1 < 0 \end{cases} \quad (21)$$

ile öncel SGO kestirilebilir.

(20)'deki yaklaşımın yanısıra 'karara dayalı kestirim' sınıfına giren ve daha çok sezgisel olarak [12]'de ortaya atılmış olan yöntem müziksel gürültünün azaltılmasına önemli bir katkı sağlamıştır. Bu yöntemin ana fikri, bir çerçevedeki öncel SGO değerlerinin bulunmasında bir önceki çerçevede kestirilmiş olan temiz konuşma bilgisinden yararlanmaktır. Buna bağlı olarak, öncel SGO

$$\eta_k(m) = \alpha \frac{A_k^2(m-1)}{\lambda_k^d(m)} + (1-\alpha)P[\gamma_k(m)-1] \quad ; \quad 0 < \alpha < 1 \quad (22)$$

ile türetilmektedir;  $P[\cdot]$  işlemcisi

$$P[x] = \begin{cases} x & x \geq 0 \\ 0 & x < 0 \end{cases} \quad (23)$$

olarak tanımlıdır. (21)'in sağ tarafında yer alan toplamdaki ilk terim bir önceki ((m-1)'inci) çerçevede elde edilen temiz konuşma kestiriminin genliğini içermektedir. Bu genliğin gürültü değişintisine oranı öncel SGO için karara dayalı kestirimi oluşturur; ikinci terim ise öncel SGO'nun en büyük olabirlik kestirimidir. (21)'deki ifade bu iki farklı sonucu birleştirmektedir. Yapılan denemeler sonucunda  $\alpha$ 'nın bire oldukça yakın (~0.95) seçilmesinin iyi sonuçlar verdiği saptanmıştır. Bu durumda öncel SGO daha çok karara dayalı terimden elde edilmiş olmaktadır.

Karara dayalı kestirimin müziksel gürültünün azalmasında neden etkili olduğu konusunda yapılan bir inceleme de bulunmaktadır [8]. (22)'deki ifadenin doğrusal olmayan bir ortalamaya karşılık geldiği bu çalışmada açıklanmaktadır.

## 2.7 Kazanç Alt Sınırının Gürültü Değişintisine Bağlı Olarak Hesaplanması

(12)'de  $G(k)$  olarak yer alan izgesel kazanç bir alt sınır uygulanması literatürde benimsenmiş olan bir işlemdir. Alt sınır, kalan gürültü miktarında belirli bir artışa neden olmakla birlikte müziksel gürültüyü oluşturan bileşenlerin duyumsal olarak maskelenmesini sağladığı için kullanılır.

Giriş bölümünde açıklandığı gibi tüm frekanslar için sabit bir alt sınır kullanmak yerine gürültü düzeyine ve SGO'ya bağlı olarak değişen bir alt sınır kullanılması bu çalışmada düşünülmüştür. Değişken alt sınır ile yüksek enerjili tonal gürültü bileşenlerinin kontrollu olarak daha da azaltılması amaçlanmıştır.

Bu çalışmada ortaya çıkan yaklaşım,  $G(k)$ 'nin belirli bir eşik değerinin altında kaldığı frekanslarda, o frekanslardaki göreceli gürültü düzeyine göre  $G(k)$ 'dan farklı bir kazanç değerinin bulunup  $G(k)$  yerine kullanılması biçimindedir. Burada,  $G(k)$ 'nin eşik altındaki değerleri dolaylı olarak SGO'nun da düşük değerlerini belirtmektedir. Söz konusu frekanslarda gürültü düzeyi ile azalan bir kazanç uygulanmaktadır. Kazanç ile gürültü düzeyi ile arasındaki ilişki sistemin frekans seçiciliğini de belirlemektedir.

$G(k)$ 'nin eşik değerinin altına indiği frekanslarda uygulanacak kazancın,  $L_k(\lambda_k^d, m)$ , (24)'deki ifadeye sahiptir

$$L_k(\lambda_k^d, m) = a(m)\lambda_k^{dC} + b(m) \quad (24)$$

Burada  $\lambda_k^d$ , (1)'de elde edildiği gibi gürültünün değışintisi ve  $C$  bir sabittir;  $a(m)$  ve  $b(m)$  katsayıları aşağıdaki eşitlikleri sağlayacak şekilde her çerçeve için hesaplanır.

$$L_{\max} = G(k) = a(m)\lambda_{\min}^C(m) + b(m) \quad (25)$$

$$L_{\min} = a(m)\lambda_{\max}^C(m) + b(m); \quad (26)$$

$$a(m) < 0, \quad 0 < C$$

$\lambda_{\max}$  ve  $\lambda_{\min}$ , sırasıyla,  $\lambda_k^d$ 'nin en büyük ve en küçük değerlerini;  $L_{\max}$  ve  $L_{\min}$  ise, sırasıyla, en büyük ve en küçük kazanç değerlerini göstermektedirler. Böylece gürültü enerjisinin en büyük değeri aldığı frekans bileşenine en küçük kazanç,  $L_{\min}$ , uygulanırken, gürültü enerjisinin en küçük değeri aldığı frekans bileşenine en büyük kazanç,  $L_{\max}$ , uygulanmaktadır.

Bu hesaplardan sonra  $k$ 'inci frekans bileşenine ait kazanç değeri

$$G(k) = \begin{cases} G(k) & G(k) > L_{\max} \\ L_k(\lambda_k^d, m) & G(k) < L_{\max} \end{cases} \quad (27)$$

kuralına göre saptanır.

LEKOİH ölçütüne göre elde edilen izgesel kazancın,  $G(k)$ , düşük değerler alması SGO'nun düşüklüğünü, dolayısıyla konuşma enerjisinin azlığını gösterir. (24)'de tanımlanan kazanç işlevine, LEKOİH kazancının belirli bir düzeyin altına inmesi durumunda işlerlik kazandırılmaktadır. Böylece konuşma enerjisinin yok olduğu ya da çok azaldığı yerlerde, gürültünün artışıyla azalan bir kazanç uygulanmış olmaktadır. Konuşma bölgelerine girildiğinde,  $G(k)$ 'nin artması ile birlikte,  $G(k)$  yeniden işlerlik kazanmakta ve konuşma bileşenlerinde bir kayıp oluşmamaktadır.

### 3 SONUÇLAR

Değişken kazanç alt sınırı, özellikle güçlü periyodik bileşenleri olan gürültünün bozduğu konuşma üzerinde son derece yararlı sonuçlar vermektedir. Bu yaklaşım ile temizleme işleminden sonra kalan gürültünün spektrumu kolayca şekillendirilebilmektedir. Önerilen yöntem gürültünün güçlü bileşenlerini kendiliğinden yakalayan bir "notch" süzgeç etkisi yaratacak biçimde de kullanılabilir. Ayrıca bu işlemin konuşmayı bozucu ek bir etkisi bulunmamaktadır.

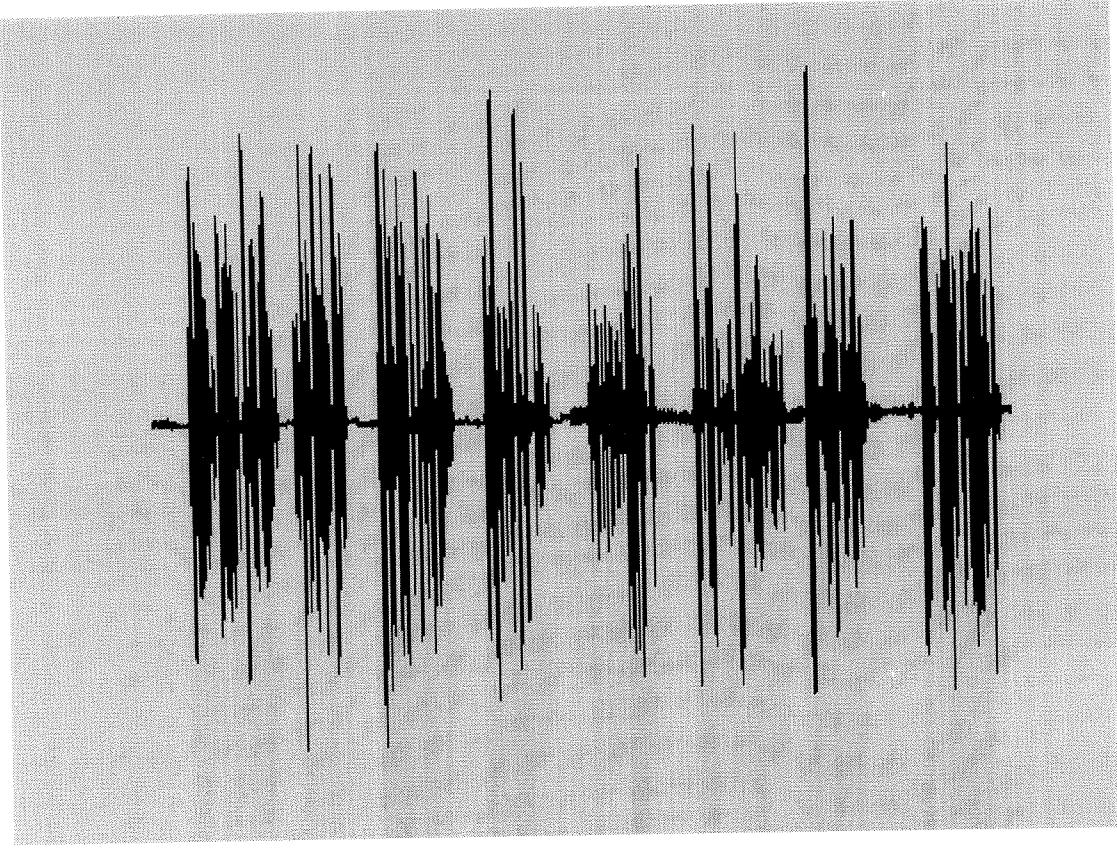
Yapılan çalışmalar sonucunda ortaya çıkan yöntemle elde edilen sonuçlar ve değişken kazanç alt sınırının etkisi formal olmayan dinleme testlerinin yanısıra objektif ölçütlerle de sınanmıştır. Formal test giderleri proje bütçesinde yer almadığı için bu tür testler yapılamamıştır.

Testlerde kullanılan konuşma kayıtları ODTÜ Elektrik ve Elektronik Mühendisliği Bölümünde yapılmıştır. Kayıt yapılan odada gürültü ve yankının azaltılması için bazı önlemler alınmış ve Senheiser MKE102 mikrofonu kullanılmıştır. Erkek ve kadın seslerinden oluşan bir dizi temiz konuşma Çizim 3'de; temiz konuşmanın bir kısmına ait spektrogram görüntüsü ise Çizim 9'da görülmektedir. Daha sonraki spektrogram görüntüleri de aynı bölgeyi içerecektir. Gürültü olarak SAAB 340B tipi bir pervaneli yolcu uçağı gürültüsü ve F16 savaş uçağı gürültüsü kullanılmıştır. SAAB gürültü kaydı, bu uçakla yapılan bir yolculuk sırasında kendi olanaklarımızla gerçekleştirilmiş, F16 gürültüsü ise SPEAR veritabanından

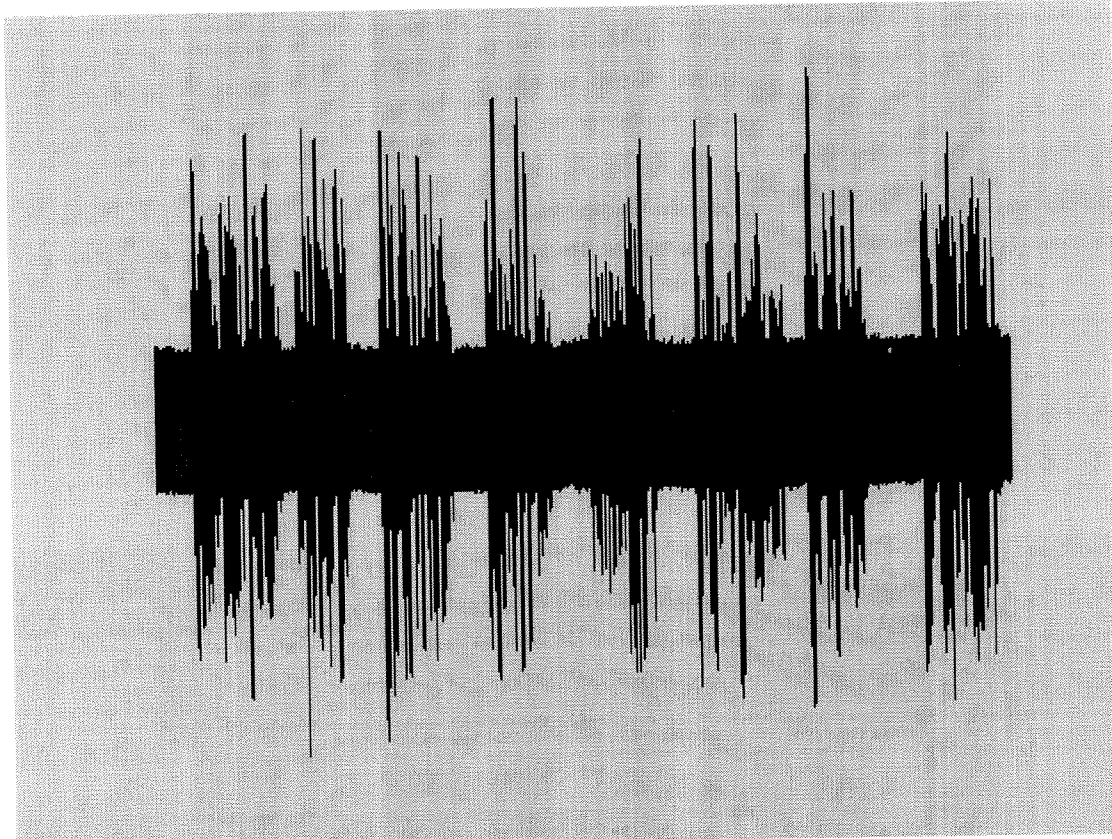
alınmıştır. SAAB gürültüsü çok güçlü tonal bileşenlere sahip bir gürültüdür. F16 gürültüsü ise bir kaç frekansta hafif yoğunlaşması olup e bant içinde daha düzenli dağılıma sahip bir gürültüdür.

Konuşma ile SAAB gürültüsü, -5dB SGO elde edilmek üzere toplanmışlardır. Gürültülü konuşma Çizim 4'de; spektrogramı ise Çizim 10'da görülmektedir. Çizim 11'de gürültülü konuşmanın MELP kodlayıcısından geçmiş hali bulunmaktadır.

Çizim 5'de bu çalışmada önerilen yöntem kullanılmadan yapılan gürültü temizleme sonucunda elde edilen sinyal; Çizim 6'da ise önerilen yöntem kullanılarak yapılan temizleme sonucu görülmektedir. Çizim 5 ve Çizim 6'daki sinyallerin bir bölgesine ait spektrogramlar, sırasıyla, Çizim 12 ve Çizim 13'de yer almaktadır. Önerilen yöntemin sağladığı gürültü azaltımı bu çizimlerde kolaylıkla görülmektedir. Çizim 5 ve Çizim 6'daki sinyallerin MELP kodlayıcısından geçmiş halleri, sırasıyla, Çizim 7 ve Çizim 8'de; spektrogramları ise Çizim 14 ve Çizim 15'de verilmiştir. Çizim 14'deki spektrogramda sağdaki konuşma bölgesinin orta kesimlerinde harmonik yapının uğradığı bozulma görülmektedir. Önerilen yöntem ile temizlenen konuşmada gürültü düzeyinin daha düşük olmasının yanısıra, konuşma MELP kodlayıcısından geçirildiğinde, konuşma/sessizlik geçişlerinde oluşan bozulmaların da daha az olduğu, dinleme testi ile farkedilmektedir. Burada belirtilen sinyallere <http://www.eee.metu.edu.tr/~ciloglu/speech%20enhancement.html> adresinden erişilebilir

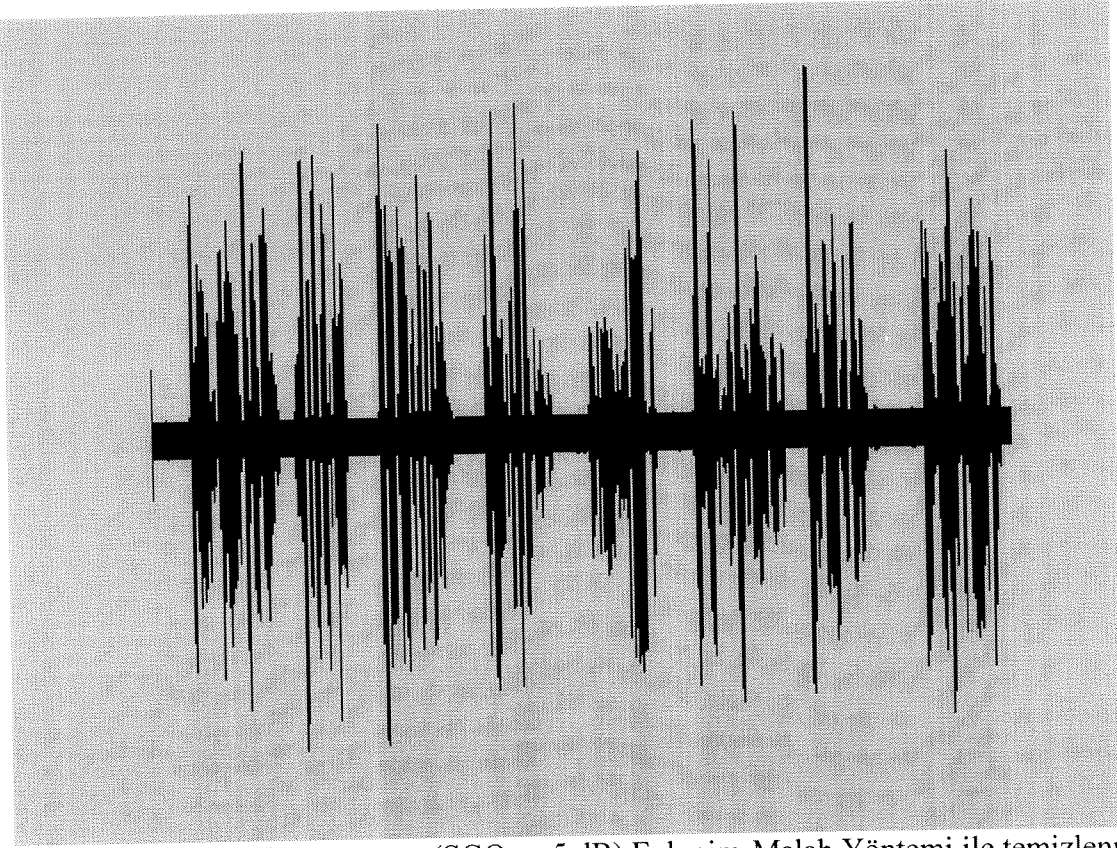


Çizim 3: Temiz konuşma örneği.

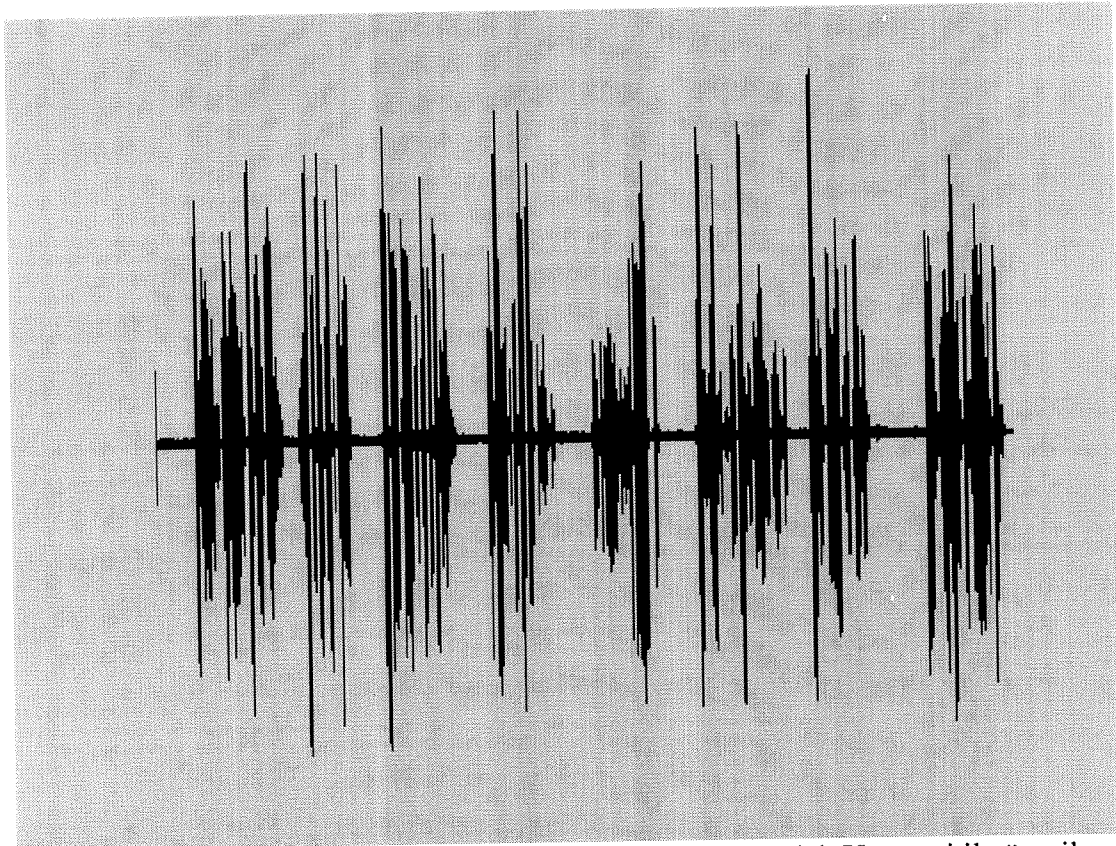


Çizim 4: Gürültülü (SGO = -5 dB) konuşma örneği



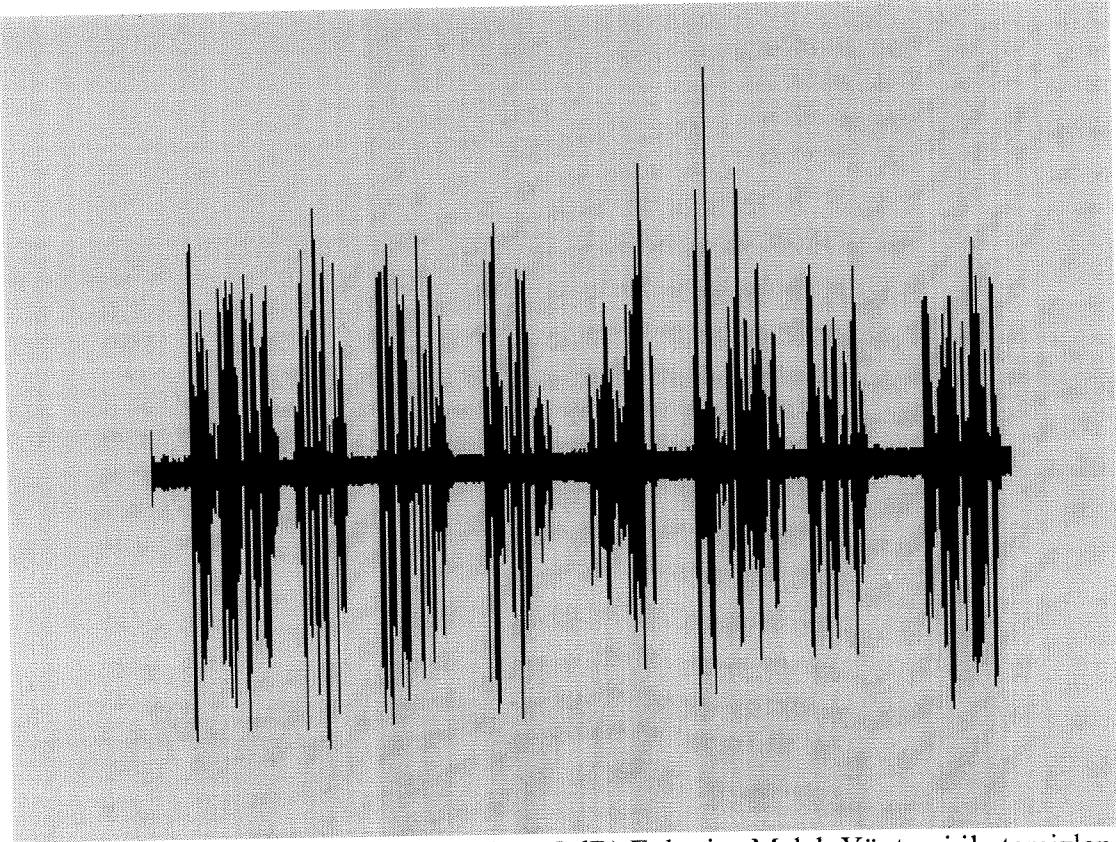


Çizim 5: Gürültülü konuşmanın (SGO = -5 dB) Ephraim-Malah Yöntemi ile temizlenmiş hali.

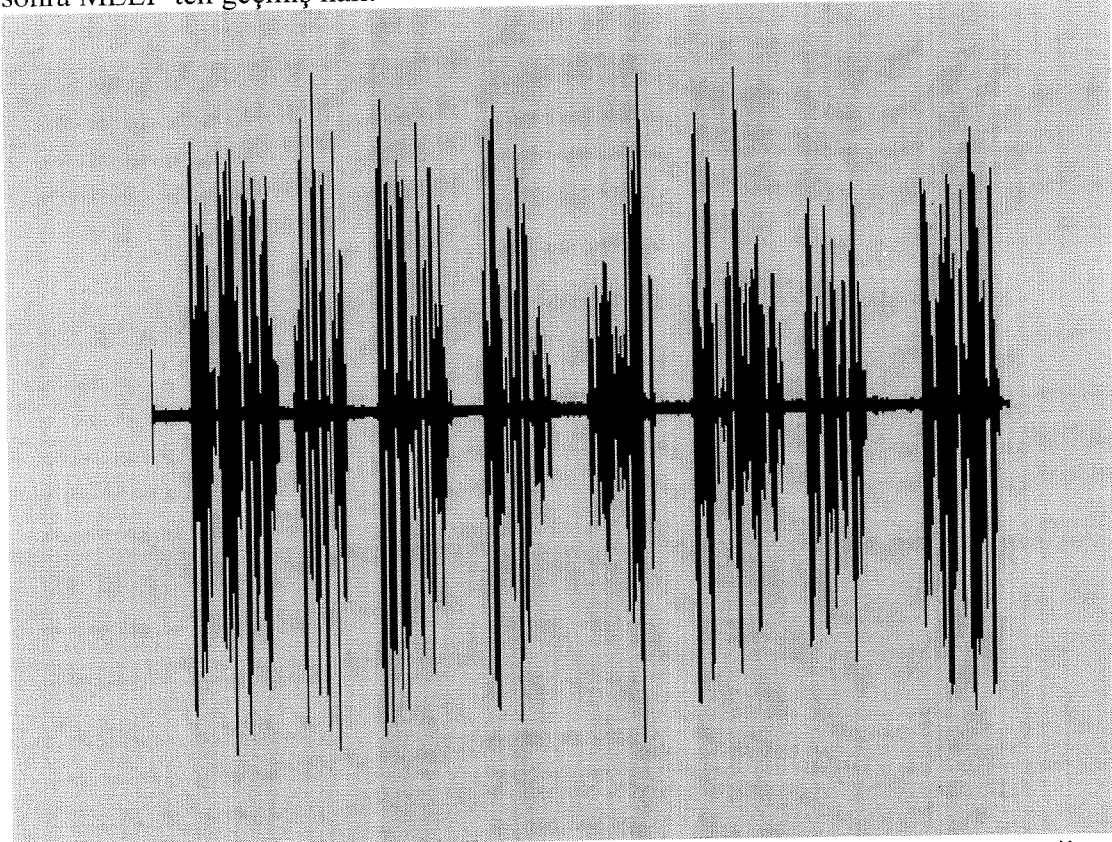


Çizim 6: Gürültülü konuşmanın (SGO = -5 dB) Ephraim-Malah Yöntemi ile önerilen yöntemin birlikte kullanılarak temizlenmiş hali.

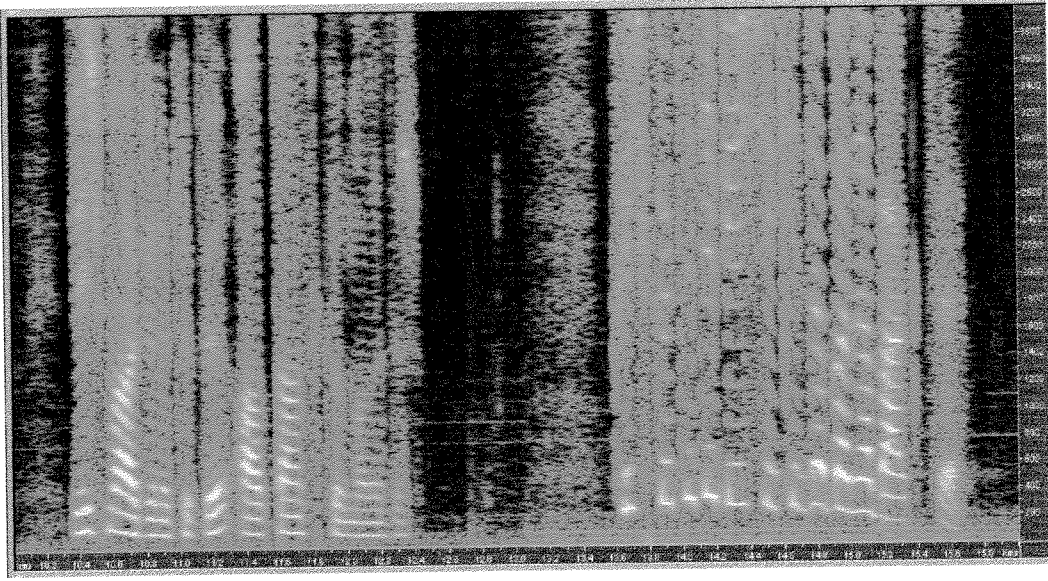




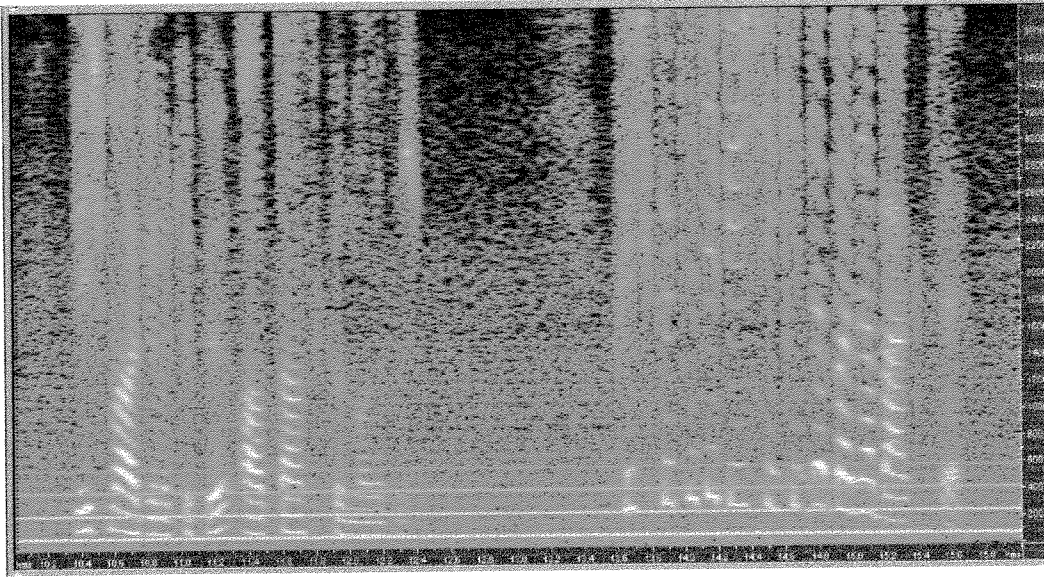
Çizim 7: Gürültülü konuşmanın (SGO = -5 dB) Ephraim-Malah Yöntemi ile temizlendikten sonra MELP'ten geçmiş hali.



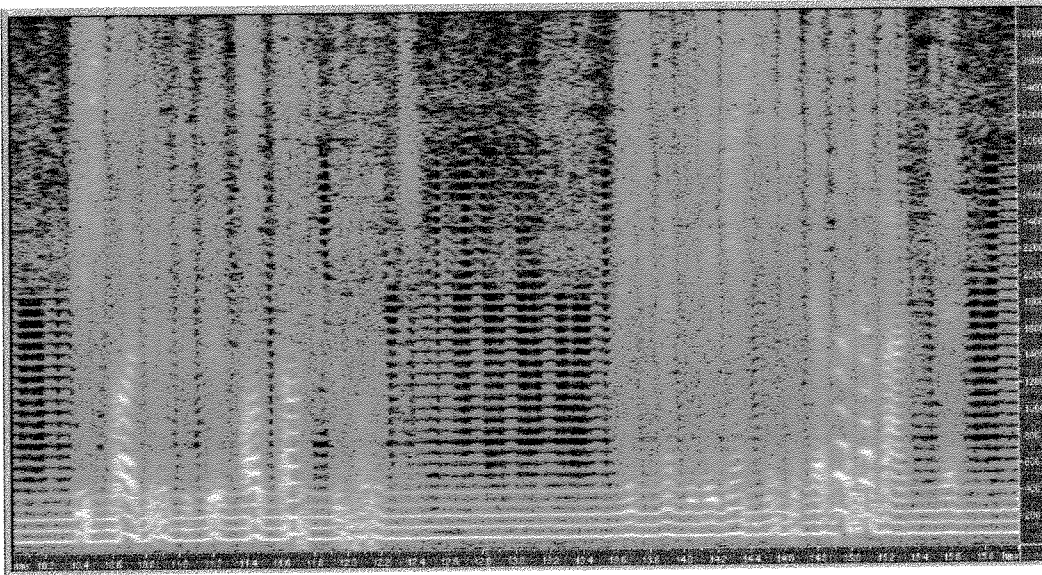
Çizim 8: Gürültülü konuşmanın (SGO = -5 dB) Ephraim-Malah Yöntemi ile önerilen yöntemin birlikte kullanılarak temizlendikten sonra MELP'ten geçmiş hali.



Çizim 9: Temiz konuşmanın spektrogramı.

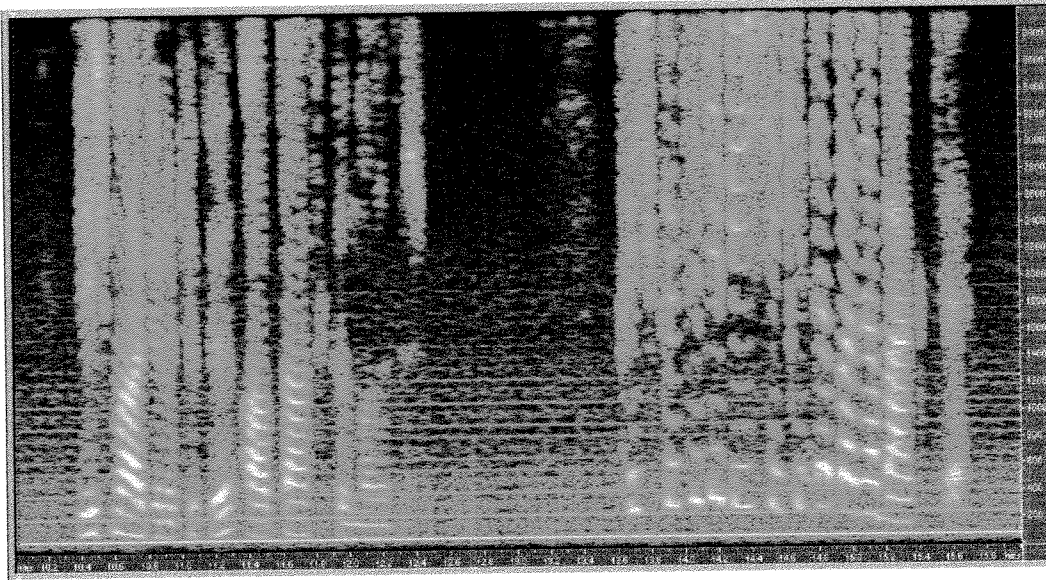


Çizim 10: SAAB 340B tipi pervaneli uçak gürültüsü eklenmiş konuşmanın spektrogramı (SNR = -5 dB)



Çizim 11: SAAB 340B tipi pervaneli uçak gürültüsü eklenmiş konuşmanın (SNR = -5 dB) MELP kodlayıcısından geçmiş halinin spektrogramı.

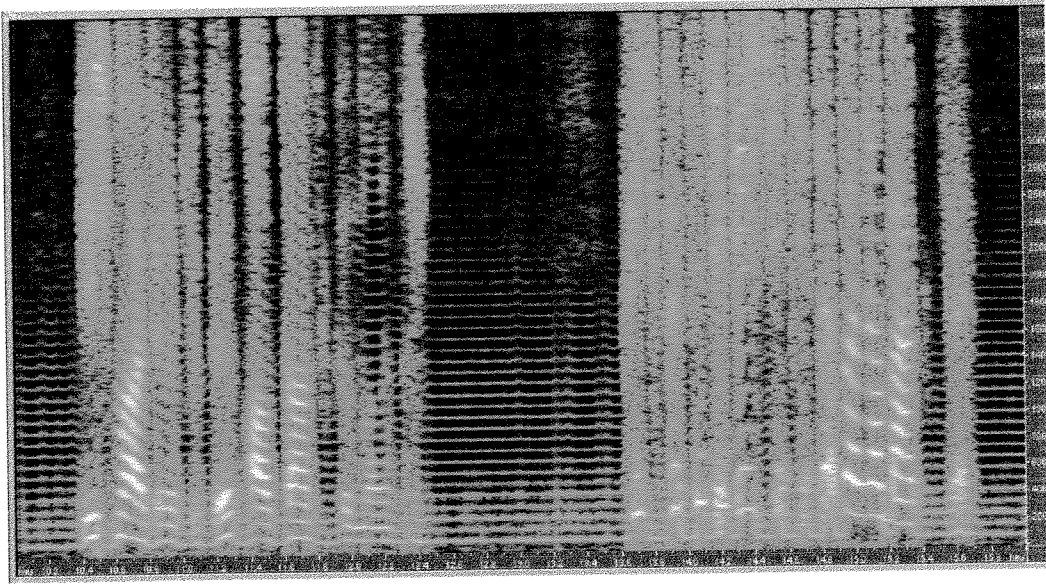




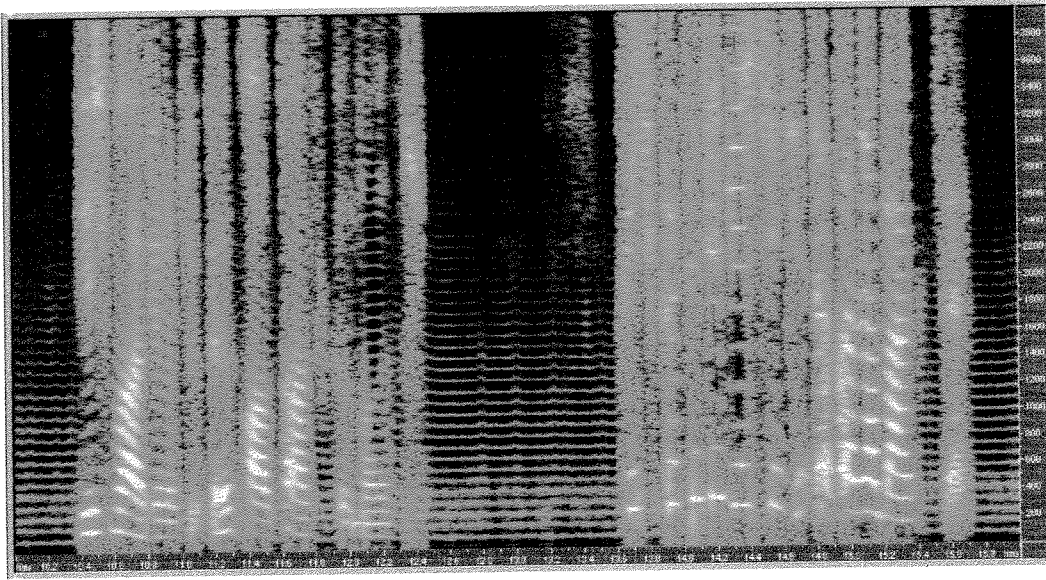
Çizim 12: Gürültülü konuşmanın (SGO = -5 dB) Ephraim-Malah Yöntemi ile temizlenmiş hali.



Çizim 13: Gürültülü konuşmanın (SGO = -5 dB) Ephraim-Malah Yöntemi ile önerilen yöntemin birlikte kullanılarak temizlenmiş hali.



Çizim 14: Gürültülü konuşmanın (SGO = -5 dB) Ephraim-Malah Yöntemi ile temizlendikten sonra MELP'ten geçmiş hali.



Çizim 15: Gürültülü konuşmanın (SGO = -5 dB) Ephraim-Malah Yöntemi ile önerilen yöntemin birlikte kullanılarak temizlendikten sonra MELP'ten geçmiş hali.

### 3.1 Objektif Değerlendirme

Objektif değerlendirme ölçütleri, [16],[17], ile konuşmanın duyumsal özellikleri arasında yakın ilişkiler kurmak zordur. İnsanın duyumsal algılama özelliklerinin ses ve gürültü türlerine göre değişiklik göstermesi, algılama sürecinin ileri derecede doğrusal olmayan ve zamanla değişen yapıda olması nedeniyle, objektif ölçütlerin tek yanlı sınıflandırması duyumsal testlerle çelişkili sonuçlar verebilmektedir. Ancak, dinleme testlerinin yapılmasındaki güçlük ve doğal olarak kişilere bağlı olması da objektif ölçütlere olan ilgiye zemin hazırlamıştır. Bugüne kadar önerilen objektif ölçütler arasında farklı temelde sınıflandırma yapan üç tanesi bu çalışmadaki sonuçlara da uygulanmıştır. Bunlar “log area ratio, LAR”, “B-SGO, (segmental SNR)” ve “weighted spectral slope, WSS”dir. Sinyaldeki tüm çerçeveler için elde edilen objektif değerlerin ortanca (‘median’) değerleri, farklı SGO değerleri için, hem SAAB gürültüsü hem de F16 gürültüsü ile aşağıdaki 12 tabloda görülmektedir. Bu tablolarda artan LAR ve WSS değerleri bozulmanın arttığını, artan B-SGO değerleri ise bozulmanın azaldığını gösterirler.

İlk altı tabloda SAAB gürültüsü ile elde edilen sonuçlara bakıldığında B-SGO ve WSS ölçütlerine göre önerilen yöntemin daha iyi sonuçlar, LAR ölçütüne göre ise daha kötü sonuçlar verdiği görülmektedir. Dinleme testlerinde ise temizlenmiş konuşma örneklerinde konuşma kalitesi bakımından bir farklılık gözlenmemekte buna karşılık önerilen yöntemin kullanılması durumunda kalan gürültü miktarında kayda değer bir azalma gözlenmektedir. Önerilen yöntemin gürültü azaltımına olumlu katkı sağlarken konuşma kalitesini bozmadığı yukarıda özetlenen objektif değerlendirme sonuçlarından da çıkartılabilir. Son altı tabloda yer alan F16 gürültüsü için elde edilmiş sonuçlara göre önerilen yöntemin etkisi B-SGO olarak olumlu görünmekte, WSS ve LAR olarak olumsuz görünmektedir. Buna karşılık temizlenmiş ve MELP kolayıcı/çözücüden geçen konuşma ile elde edilen sonuçlara göre ise önerilen yöntemin etkisi hem B-SGO hem de WSS olarak olumlu görünmekte, LAR olarak olumsuz görünmektedir. Bu sonuçlardan önerilen yöntemin özellikle kodlama ile birlikte kullanılmasının yararlı olacağı; ayrıca, genel olarak, tonal bileşenleri güçlü olan gürültünün temizlenmesinde yöntemin daha başarılı olduğu çıkartılmaktadır.

Tablo 1: Temizlenmiş konuşmaların objektif olarak karşılaştırılmaları.

SAAB gürültüsü; B-SGO = -5 dB

A: Ephraim-Malah yöntemi, B: Ephraim-Malah yöntemine önerilen yöntemin eklenmesi

	<b>A</b>	<b>B</b>
LAR	8.1893	8.5722
B-SGO	-8.0978	-2.2305
WSS	110.4893	102.5211

Tablo 2: Temizlenmiş ve MELP kodlayıcı/çözücüden geçirilmiş konuşmaların objektif olarak karşılaştırılmaları. SAAB gürültüsü; B-SGO = -5 dB

A: Ephraim-Malah yöntemi, B: Ephraim-Malah yöntemine önerilen yöntemin eklenmesi

	<b>A</b>	<b>B</b>
LAR	8.3220	8.8142
B-SGO	-4.8077	-3.0970
WSS	91.5600	87.5136

Tablo 3: Temizlenmiş konuşmaların objektif olarak karşılaştırılmaları.

SAAB gürültüsü; B-SGO = 0 dB

A: Ephraim-Malah yöntemi, B: Ephraim-Malah yöntemine önerilen yöntemin eklenmesi.

	<b>A</b>	<b>B</b>
LAR	7.9073	8.6046
B-SGO	-7.0401	-1.8420
WSS	106.8904	102.5317

Tablo 4: Temizlenmiş ve MELP kodlayıcı/çözücüden geçirilmiş konuşmaların objektif olarak karşılaştırılmaları. SAAB gürültüsü; B-SGO = 0 dB

A: Ephraim-Malah yöntemi, B: Ephraim-Malah yöntemine önerilen yöntemin eklenmesi.

	<b>A</b>	<b>B</b>
LAR	8.1089	8.7317
B-SGO	-5.6534	-2.0349
WSS	89.8084	87.0078

Tablo 5: Temizlenmiş konuşmaların objektif olarak karşılaştırılmaları.

SAAB gürültüsü; B-SGO = 5 dB

A: Ephraim-Malah yöntemi, B: Ephraim-Malah yöntemine önerilen yöntemin eklenmesi.

	<b>A</b>	<b>B</b>
LAR	7.6198	8.6237
B-SGO	-6.0283	-1.5325
WSS	102.8677	102.0089

Tablo 6: Temizlenmiş ve MELP kodlayıcı/çözücüden geçirilmiş konuşmaların objektif olarak karşılaştırılmaları. SAAB gürültüsü; B-SGO = 5 dB

A: Ephraim-Malah yöntemi, B: Ephraim-Malah yöntemine önerilen yöntemin eklenmesi.

	<b>A</b>	<b>B</b>
LAR	7.9589	8.8951
B-SGO	-5.3014	-2.2642
WSS	89.9873	86.6237

Tablo 7: Temizlenmiş konuşmaların objektif olarak karşılaştırılmaları. F16 gürültüsü; B-SGO = 5 dB  
A: Ephraim-Malah yöntemi, B: Ephraim-Malah yöntemine önerilen yöntemin eklenmesi.

	A	B
LAR	9.7811	10.0362
B-SGO	-5.3735	-1.5385
WSS	99.6305	105.8943

Tablo 8: Temizlenmiş ve MELP kodlayıcı/çözücünden geçirilmiş konuşmaların objektif olarak karşılaştırılmaları. F16 gürültüsü; B-SGO = 5 dB  
A: Ephraim-Malah yöntemi, B: Ephraim-Malah yöntemine önerilen yöntemin eklenmesi.

	A	B
LAR	9.6483	10.5494
B-SGO	-6.0642	-1.9684
WSS	93.5472	87.0605

Tablo 9: Temizlenmiş konuşmaların objektif olarak karşılaştırılmaları.  
F16 gürültüsü; B-SGO = 10 dB  
A: Ephraim-Malah yöntemi, B: Ephraim-Malah yöntemine önerilen yöntemin eklenmesi.

	A	B
LAR	9.4393	9.8179
B-SGO	-4.8361	-1.2023
WSS	97.1144	103.2270

Tablo 10: Temizlenmiş ve MELP kodlayıcı/çözücünden geçirilmiş konuşmaların objektif olarak karşılaştırılmaları. F16 gürültüsü; B-SGO = 10 dB  
A: Ephraim-Malah yöntemi, B: Ephraim-Malah yöntemine önerilen yöntemin eklenmesi.

	A	B
LAR	9.4239	10.4644
B-SGO	-5.5771	-1.7367
WSS	91.7084	84.7993

Tablo 11: Temizlenmiş konuşmaların objektif olarak karşılaştırılmaları.  
F16 gürültüsü; B-SGO = 15 dB  
A: Ephraim-Malah yöntemi, B: Ephraim-Malah yöntemine önerilen yöntemin eklenmesi.

	A	B
LAR	9.2148	9.8151
B-SGO	-4.2903	-1.0064
WSS	95.6147	102.1535

Tablo 12: Temizlenmiş ve MELP kodlayıcı/çözücünden geçirilmiş konuşmaların objektif olarak karşılaştırılmaları. F16 gürültüsü; B-SGO = 15 dB  
A: Ephraim-Malah yöntemi, B: Ephraim-Malah yöntemine önerilen yöntemin eklenmesi.

	A	B
LAR	9.2994	10.4478
B-SGO	-5.1157	-1.3693
WSS	91.6696	83.9733



### 3.2 İleriye Yönelik Çalışmalar

Giriş bölümünde belirtildiği gibi konuşma temizlemede konuşma sinyaline özgü özelliklerden yararlanılması üzerinde çalışılmış fikirlere biridir. Bu proje kapsamında yapılan çalışmanın da bu yönde geliştirilmesi yararlı olacaktır. Bu kapsamda ses modelleri için gerekli eğitim veritabanı oluşturulmalıdır. Bu veritabanından üretilecek şablonlar temizleme aşamasında bir uygunluk ölçütüne göre seçilerek çerçevelerin spektral özelliklerinin düzeltilmesinde kullanılacaktır. Şablonların seçiminde konuşulan dile özgü fonetik diziliş istatistiklerinin diğer yardımcı kurallar ile birlikte kullanılmasının da özellikle düşük enerjili olup anlaşılabilirlik üzerinde önemli etkisi olan seslerin ortaya çıkarılmasına olumlu katkı sağlayacağı kanısındayız. Bu düşüncenin bir adım ilerisinde, temizlemenin dil modellerini de kullanan bir ses tanıma sistemi yardımı ile birlikte yapılması gelmektedir. Ancak gerçek zamanda düşük işlem gereksinimi ile çalışma koşulu bir engel olarak karşımıza çıkmaktadır. Bu koşulu kaldırarak yapılacak çalışmaların yakın bir gelecekte gerçek zamanda çalışmaya elverişli hale dönüştürülebileceği de unutulmamalıdır.

#### Kaynakça

- [1] Boll, S.F., "Suppression of Acoustic Noise in Speech Using Spectral Subtraction", IEEE Trans. on Acoustics Speech and Signal Processing, c 27, s.113-120, Nisan 1979.
- [2] Martin, R., "Noise Power Spectral Density Estimation Based on Optimal Smoothing and Minimum Statistics", IEEE Trans. On Speech and Audio Processing, vol.9, no.5, s. 504-512, Temmuz 2001.
- [3] Hansen, J. H. L., Clements, M. A., "Constrained Iterative Speech Enhancement with Application to Speech Recognition", IEEE Trans. on Signal Processing, c. 39, no.4 s.795-805, Nisan 1991.
- [4] Lim, J. S., Oppenheim, A. V., "All-pole Modelling of Degraded Speech", IEEE Trans. on Acoustics Speech and Signal Processing, c. 26, s.197-210, Haziran 1978.
- [5] Hansen, J. H. L., Arslan, L., "Markov Model-Based Phoneme Class Partitioning for Improved Constrained Iterative Speech Enhancement", IEEE Transactions on Speech and Audio Processing, c.3, no. 1, s. 98-104, Ocak 1995.
- [6] Sreenivas, T. V., Kirnapure, P., "Codebook Constrained Wiener Filtering for Speech Enhancement", IEEE Transactions on Speech and Audio Processing, c.4, no. 5, s. 383-389, Eylül 1996.

- [7] Y. Ephraim, D. Malah, "Speech Enhancement Using a Minimum Mean-Square Error Short-Time Spectral Amplitude (MMSE-SSA) Estimator, IEEE Trans. on Acoustics Speech and Signal Processing, c. 32, no. 6, s. 1109-1121, Aralık 1984.
- [8] Cappe, O., "Elimination of the Musical Noise Phenomenon with the Ephraim and Malah Noise Suppressor", IEEE Transactions on Speech and Audio Processing, c.2, no. 2, s. 345-349, Nisan 1994.
- [9] Sim, B. L., Tong, Y. C., Tan, C. T., "A Parametric Formulation of the Generalized Spectral Subtraction Method", IEEE Transactions on Speech and Audio Processing, c.6, no. 4, s. 328-337, Temmuz 1998.
- [10] Soderstrand, M. A., Johnson, T. G., Strandberg, R. H., Loomis, H. H., Rangarao, K. V., "Suppression of Multiple Narrow-Band Interference Using Real-Time Adaptive Notch Filters", IEEE Transactions on Circuits and Systems-II, Analog and Digital Signal Processing, c. 44, no.3, s. 217-225, Mart 1997.
- [11] R. J. McAulay, M. L. Malpass, "Speech Enhancement Using a Soft-Decision Noise Suppression Filter", IEEE Trans. on Acoustics Speech and Signal Processing, c. 28, s. 137-145, 1980.
- [12] Y. Ephraim, D. Malah, "Speech Enhancement Using a Minimum Mean-Square Error Log-Spectral Amplitude (MMSE-LSA) Estimator, IEEE Trans. on Acoustics Speech and Signal Processing, c. 33, no. 2, s. 443-445, Nisan 1985.
- [13] D. Malah, R. V. Cox, A. J. Accardi, "Tracking Speech-Presence Uncertainty to Improve Speech Enhancement in Non-Stationary Noise Environments", IEEE Int. Conf. On Acoustics, Speech and Signal Processing, 1999.
- [14] D. W. Griffin, J. S. Lim, "Signal Estimation from Modified Short-Time Fourier Transform", IEEE Trans. Acoustics, Speech and Signal Processing, s. 236-243, Nisan 1984.
- [15] J. Sohn, W. Sung, "A Voice Activity Detector Employing Soft Decision Base Noise Spectrum Adaptation", Proc. IEEE Int. Conf. On Acoustics, Speech and Signal Processing, 1998.
- [16] J. H. L. Hansen, B. L. Pellom, "An Effective Quality Evaluation Protocol For Speech Enhancement Algorithms", Proc. Int. Conf. On Spoken Language Processing, 1998.
- [17] J. R. Deller, J. H. L. Hansen, J. G. Proakis, Discrete Time Processing of Speech Signals, IEEE Press, 2000.

## PROJE ÖZET BİLGİ FORMU

**Proje Kodu :** 100E004

**Proje Başlığı :** KONUŞMANIN GÜRÜLTÜDEN ARINDIRILMASI

**Proje Yürütücüsü ve Yardımcı Araştırmacılar :**

Doç. Dr. Tolga Çiloğlu  
Y. Müh. Cem Demirkır

**Projenin Yürütüldüğü Kuruluş ve Adresi :**  
Elektrik ve Elektronik Mühendisliği Bölümü, ODTÜ, Ankara, 06531

**Destekleyen Kuruluş(ların) Adı ve Adresi :**

**Projenin Başlangıç ve Bitiş Tarihleri :** 01 Ağustos 2000 – 28 Şubat 2002

**Öz :** Bu çalışmanın konusu gürültülü konuşmanın temizlenmesi, iyileştirilmesidir. Gürültü temizleme, konuşma kodlayıcıların başarımını da olumlu yönde etkiler. Bu çalışmada, özellikle güçlü tonal bileşenlere sahip gürültüler için gürültü temizleme sistemlerinin başarısının artırılması amaçlanmıştır. Buna yönelik olarak bir yöntem önerilmiş ve uygulanmıştır. Yöntem, sinyal-gürültü-oranının düşük olduğu yerlerde, yüksek gürültüyü daha fazla söndürme düşüncesini gerçekleştirmektedir. SAAB 340B pervaneli uçağının tonal bileşenleri güçlü olan gürültüsü ile F16 jet uçağı gürültüsü kullanılarak elde edilen sonuçların dinleme testlerine ve objektif ölçütlere göre değerlendirilmesi sonucunda yöntemin yararlı olduğu gözlenmiştir.

**Anahtar Kelimeler:** Konuşma temizleme, konuşma iyileştirme, konuşma işleme, gürültü azaltımı

**Projeden Kaynaklanan Yayınlar :** ---

**Bilim Dalı :**

**Doçentlik B. Dalı Kodu :**