

EGE ÜNİVERSİTESİ FEN BİLİMLERİ ENSTİTÜSÜ

(DOKTORA TEZİ)

**AMPİRİK KİP AYRIŞTIRMASI YÖNTEMİ İLE
UYARLANIR EŞİKLEME TABANLI KONUŞMA
İYİLEŞTİRME VE SESLİ ETKİNLİK ALGILAMA**

Özkan ARSLAN

Tez Danışmanı : Dr. Öğr. Üyesi Erkan Zeki ENGİN

Elektrik - Elektronik Mühendisliği Anabilim Dalı

Sunuş Tarihi : 07.08.2018

**Bornova-İZMİR
2018**

Özkan ARSLAN tarafından doktora tezi olarak sunulan “AMPİRİK KİP AYRIŞTIRMASI YÖNTEMİ İLE UYARLANIR EŞİKLEME TABANLI KONUŞMA İYİLEŞTİRME VE SESLİ ETKİNLİK ALGILAMA” başlıklı bu çalışma EÜ Lisansüstü Eğitim ve Öğretim Yönetmeliği ile EÜ Fen Bilimleri Enstitüsü Eğitim ve Öğretim Yönergesi’nin ilgili hükümleri uyarınca tarafımızdan değerlendirilerek savunmaya değer bulunmuş ve 07.08.2018 tarihinde yapılan tez savunma sınavında aday oybirliği/oyçokluğu ile başarılı bulunmuştur.

Jüri Üyeleri:

İmza

Jüri Başkanı : Dr. Öğr. Üyesi Erkan Zeki ENGİN

Raportör Üye : Prof. Dr. Mehmet ENGİN

Üye : Dr. Öğr. Üyesi Nalan ÖZKURT

Üye : Doç. Dr. Yücel KOÇYİĞİT

Üye : Prof. Dr. Aydoğan SAVRAN

.....

EGE ÜNİVERSİTESİ FEN BİLİMLERİ ENSTİTÜSÜ

ETİK KURALLARA UYGUNLUK BEYANI

EÜ Lisansüstü Eğitim ve Öğretim Yönetmeliğinin ilgili hükümleri uyarınca Doktora Tezi olarak sunduğum “AMPİRİK KİP AYRIŞTIRMASI YÖNTEMİ İLE UYARLANIR EŞİKLEME TABANLI KONUŞMA İYİLEŞTİRME VE SESLİ ETKİNLİK ALGILAMA” başlıklı bu tezin kendi çalışmam olduğunu, sunduğum tüm sonuç, doküman, bilgi ve belgeleri bizzat ve bu tez çalışması kapsamında elde ettiğimi, bu tez çalışmasıyla elde edilmeyen bütün bilgi ve yorumlara atıf yaptığımı ve bunları kaynaklar listesinde usulüne uygun olarak verdiğimi, tez çalışması ve yazımı sırasında patent ve telif haklarını ihlal edici bir davranışımın olmadığını, bu tezin herhangi bir bölümünü bu üniversite veya diğer bir üniversitede başka bir tez çalışması içinde sunmadığımı, bu tezin planlanmasından yazımına kadar bütün safhalarda bilimsel etik kurallarına uygun olarak davrandığımı ve aksinin ortaya çıkması durumunda her türlü yasal sonucu kabul edeceğimi beyan ederim.

07/08/2018



Özkan ARSLAN

ÖZET**AMPİRİK KİP AYRIŞTIRMASI YÖNTEMİ İLE UYARLANIR
EŞİKLEME TABANLI KONUŞMA İYİLEŞTİRME VE SESLİ
ETKİNLİK ALGILAMA**

ARSLAN, Özkan

Doktora Tezi, Elektrik-Elektronik Mühendisliği Anabilim Dalı
Tez Danışmanı: Dr. Öğr. Üyesi Erkan Zeki ENGİN
Ağustos 2018, 151 sayfa

Bu çalışmada, gürültülü konuşma sinyallerinin içkin kip fonksiyonu katsayılarına Teager enerji operatörünün uygulanması ile çıkartılan enerji dağılımının Gamma istatistiksel modeline dayanan yeni bir sesli/konuşma etkinlik algılama ve konuşma iyileştirme algoritması önerilmiştir. Gamma dağılım fonksiyonu, gürültülü konuşma ve kestirilen gürültünün Teager enerji uygulanmış içkin kip fonksiyonu katsayılarından uyarlanır eşik değerlerinin belirlenmesinde kullanılmıştır. Simetrik Kullback-Leibler uzaklığı ile konuşma ya da gürültü olma olasılığı üzerinden uyarlanır eşik değerleri belirlenmiş ve her kip için hesaplanmıştır. Gürültülü konuşma sinyallerinin içkin kip fonksiyonu katsayılarının bu uyarlanır eşik değerlerinden geçirilmesi ile konuşma ve konuşma olmayan bölgeler belirlenerek sesli etkinlik algılama gerçekleştirilmiştir. Konuşma iyileştirme algoritması için ise bu katsayıların uyarlanır eşik değerleri kullanılarak oluşturulan bir yarı-yumuşak eşikleme fonksiyonundan geçirilmesi ile iyileştirilmiş konuşmalar elde edilmiştir.

Yöntemlerin test edilmesinde İngilizce için NOIZEUS ve Türkçe için ise ODTÜ veri tabanı kullanılmıştır. Yöntemler yedi farklı gürültü tipi ve beş farklı sinyal gürültü oranı seviyesi koşullarında değerlendirilmiştir. Önerilen sesli etkinlik algılama yöntemi; konuşma doğruluk oranı, konuşma dışı doğruluk oranı ve genel doğruluk oranı ölçütlerine göre literatürdeki yöntemlere üstünlük sağlamıştır. Benzer şekilde, önerilen konuşma iyileştirme yöntemi nesnel kalite ve anlaşılabilirlik ölçütleri ile değerlendirilmiş ve literatürde sıklıkla kullanılan diğer yöntemler ile karşılaştırılmıştır. Tüm sonuçlar göz önünde bulundurulduğunda önerilen yöntem hem gürültü azaltmada hem de anlaşılabilirliğin sağlanmasında diğer yöntemlere göre daha başarılı bulunmuştur.

Anahtar sözcükler: Konuşma iyileştirme, sesli etkinlik algılama, ampirik kip ayrıştırması, Gamma olasılık dağılımı, Teager enerji, Kullback-Leibler uzaklığı.

ABSTRACT**SPEECH ENHANCEMENT AND VOICE ACTIVITY DETECTION
BASED ON ADAPTIVE THRESHOLDING USING EMPIRICAL
MODE DECOMPOSITION**

ARSLAN, Özkan

Ph.D. in Electrical-Electronics Eng.
Supervisor: Dr. Öğr. Üyesi Erkan Zeki ENGİN
August 2018, 151 pages

In this study, a new voice/speech activity detection and speech enhancement algorithms based on Gamma statistical model of the energy distribution extracted by Teager energy operator applied to intrinsic mode functions coefficients of noisy speech signals are proposed. The Gamma distribution function is used to determine the adaptive threshold values from Teager energy operated intrinsic mode function coefficients of noisy speech and estimated noise signal. Adaptive threshold values are determined and calculated for each mode based on the probability of speech or noise with Symmetric Kullback-Leibler divergence. The speech and non-speech regions are obtained by applying adaptive thresholding in these coefficients of noisy speech. For speech enhancement algorithms, the enhanced speech signals are obtained by a semi-soft thresholding function which is utilized by thresholded intrinsic mode coefficients of noisy speech.

The proposed methods are tested on NOIZEUS for English and on METU for Turkish speech database and evaluated for seven kinds of noises across five different SNR levels. The proposed voice activity detection method is superior to other literature methods in terms of speech hit rate (HR1), non-speech hit rate (HR0) and overall accuracy rate. Similarly, the proposed speech enhancement method is compared to other methods frequently used in the literature in terms of objective quality and intelligibility measures. Simulation results show that the proposed methods are effective and outperformed to provide in terms of noise reduction and intelligibility compared with other speech enhancement methods for different SNR levels.

Keywords: Speech enhancement, voice activity detection, empirical mode decomposition, Gamma distribution, Teager energy, Kullback-Leibler divergence.

TEŞEKKÜR

Tez çalışmamda her zaman destek olan ve yardımlarını esirgemeyen danışman hocam Dr. Öğr. Üyesi Erkan Zeki ENGİN'e ve hocamız Prof. Dr. Mehmet ENGİN'e,

Tezin başlangıcından bitimine kadar tüm süreçlerde katkısı bulunan Yaşar Üniversitesi, Elektrik-Elektronik Mühendisliği Bölümünden Dr. Öğr. Üyesi Nalan ÖZKURT'a,

Tez kapsamında kullanmış olduğumuz Türkçe konuşma veri tabanını oluşturan ve bizimle paylaşarak çalışmamıza katkıda bulunan Gazi Üniversitesi, Elektrik-Elektronik Mühendisliği Bölümünden Prof. Dr. Özgül SALOR DURNA'ya,

Tez çalışmamda her türlü kolaylığı ve imkanı sağlayan Namık Kemal Üniversitesi, Elektronik ve Haberleşme Mühendisliği Bölümünden Prof. Dr. Hafız ALİSOY ile bilimsel katkılarından ve desteklerinden dolayı Doç. Dr. Reşat MUTLU hocalarıma,

Her zaman yardım, destek ve arkadaşlıkları ile yanımda bulunan Ege Üniversitesi, Elektrik-Elektronik Mühendisliği Bölümünden Arş. Gör. Erman SELİM, Müh. Ozan GÖKKAN ve İzmir Biyotıp ve Genom Merkezi'nden Dr. İbrahim AKKAYA'ya,

Tüm eğitim hayatım boyunca yanımda bulunan ve destekleyen herkese çok teşekkür ederim.

İÇİNDEKİLER

	<u>Sayfa</u>
ÖZET	vii
ABSTRACT	ix
TEŞEKKÜR	xi
ŞEKİLLER DİZİNİ	xviii
ÇİZELGELER DİZİNİ	xxiii
SİMGELER VE KISALTMALAR DİZİNİ	xxv
1. GİRİŞ	1
1.1 Genel Gürültü Kaynakları ve Çeşitleri	2
1.1.1 Toplamsal gürültü	3
1.1.2 Çarpımsal Gürültü	5
1.1.3 Dürtüsel Gürültü	5
1.1.4 Durağan ve Durağan Olmayan Gürültü	5
1.2 Hedefler	6
1.3 Katkılar ve Özgünlükler	6
1.4 Tezin Organizasyonu	8
2. LİTERATÜR TARAMASI	9
2.1 Konuşma İyileştirme Literatür Taraması	9

İÇİNDEKİLER (devamı)

Sayfa

2.2 Sesli Etkinlik Algılama Literatür Taraması	18
3. LİTERATÜRDE KULLANILAN KONUŞMA İYİLEŞTİRME YÖNTEMLERİNİN GÖZDEN GEÇİRİLMESİ.....	20
3.1 Zaman Alanı Yöntemleri.....	20
3.1.1 Alt-Uzay Tabanlı Algoritmalar.....	20
3.2 Frekans Alanı Yöntemleri.....	22
3.2.1 Temel İzgesel Çıkarım.....	22
3.2.2 Çoklu Bant İzgesel Çıkarım.....	24
3.2.3 En Küçük Ortalama Kareler Kestiricisi	25
3.2.4 Logaritmik En Küçük Ortalama Kareler Kestiricisi	26
3.2.5 Sinyal Varlık Belirsizliğine Dayalı Logaritmik En Küçük Ortalama Kareler Kestiricisi	26
3.2.6 Wiener Süzgeçleme	27
4. NESNEL KALİTE VE ANLAŞILABİLİRLİK ÖLÇÜTLERİ	29
4.1 Nesnel Kalite Ölçütleri.....	29
4.1.1 Sinyal Gürültü Oranı (SNR) ve Bölütsel SNR	29
4.1.2 Ağırlıklı İzgesel Eğim.....	30
4.2 Nesnel Anlaşılabilirlik Ölçütleri	30

İÇİNDEKİLER (devamı)

	<u>Sayfa</u>
4.2.1 Konuşma Kalitesinin Algısal Değerlendirmesi	30
4.2.2 Kısa-Zamanlı Nesnel Anlaşılabilirlik	31
5. KONUŞMA VE GÜRÜLTÜ VERİ TABANLARI	34
6. LİTERATÜRDEKİ KONUŞMA İYİLEŞTİRME YÖNTEMLERİNİN İNGİLİZCE VE TÜRKÇE İÇİN DEĞERLENDİRME SONUÇLARI	38
6.1 İngilizce İçin Yöntemlerin Nesnel Kalite ve Anlaşılabilirlik Ölçütleri İle Değerlendirilmesi	38
6.2 Türkçe İçin Yöntemlerin Nesnel Kalite ve Anlaşılabilirlik Ölçütleri İle Değerlendirilmesi	47
6.3 Tartışma	54
7. KONUŞMA/SESLİ ETKİNLİK ALGILAMA	56
7.1 Standart AMR1, AMR2 ve G.729B Yöntemleri	57
7.2 SOHN Sesli Etkinlik Algılama	58
7.3 Uzun-Süreli İzgesel Ayrılma	58
7.4 Önerilen Konuşma/Sesli Etkinlik Algılama Yaklaşımı	59
7.4.1 Ampirik Kip Ayrıştırması	61
7.4.2 Teager Enerji Operatörü	64
7.4.3 TEO Uygulanmış İKF Katsayıları için Önerilen İstatistiksel Model	65
7.4.4 Uyarlanır Eşik Değerlerinin Hesaplanması	68

İÇİNDEKİLER (devamı)

	<u>Sayfa</u>
7.5 Sesli Etkinlik Algılama Performans Değerlendirme Ölçütleri	70
7.6 Uygulama Koşulları ve Deneysel Sonuçlar	71
7.6.1 İngilizce için Sesli Etkinlik Algılama Yöntemlerinin Performans Değerlendirmesi	75
7.6.2 Türkçe için Sesli Etkinlik Algılama Yöntemlerinin Performans Değerlendirmesi	81
7.7 Tartışma.....	87
8.AMPİRİK KİP AYRIŞTIRMA TABANLI KONUŞMA İYİLEŞTİRME YAKLAŞIMI	89
8.1 Zaman-Frekans Alanı Yöntemleri	90
8.1.1 Ayırık Dalgacık Dönüşümü	90
8.1.2 Ayırık Dalgacık Paket Dönüşümü	92
8.1.3 Ampirik Kip Ayrıştırması	92
8.2 Önerilen Konuşma İyileştirme Yöntemi	96
8.2.1 Gürültü Kestirimi	96
8.2.2 Teager Enerji Operatörü	98
8.2.3 TEO Uygulanmış İKF Katsayıları için Önerilen İstatistiksel Model	100
8.2.4 Önerilen Uyarlanır Eşik Değerlerinin Hesaplanması	103
8.2.5 Önerilen Yöntem için Kullanılan Eşikleme Fonksiyonu	105

İÇİNDEKİLER (devamı)

	<u>Sayfa</u>
8.2.6 Ters Ampirik Kip Ayrıştırması	107
8.3 Yöntemlerin Kalite ve Anlaşılabilirlik Bakımından Değerlendirilmesi	107
8.3.1 İngilizce İçin Önerilen Yöntemin Kalite ve Anlaşılabilirlik Ölçütleri İle Değerlendirilmesi	108
8.3.2 Türkçe İçin Önerilen Yöntemin Kalite ve Anlaşılabilirlik Ölçütleri İle Değerlendirilmesi	120
8.4 Tartışma	132
9. GENEL SONUÇLAR, ÖNERİLER VE GELECEK ÇALIŞMALAR	135
KAYNAKLAR DİZİNİ.....	138
ÖZGEÇMİŞ	150

ŞEKİLLER DİZİNİ

<u>Şekil</u>	<u>Sayfa</u>
1.1 Konuşma iyileştirme işlemi	1
1.2 Beyaz gürültünün genlik-zaman, öz-ilişki ve güç izgesel yoğunluk gösterimi .3	
1.3 Gürültülerin zaman ve güç izgesel yoğunluk gösterimleri: (a) araba, (b) mırıltı ve (c) hava alanı	5
3.1 Optimum süzgeçleme.....	27
6.1 İngilizce için tüm gürültü ve tüm SNR seviyelerinde BSNR İyileştirme (dB) değerleri	39
6.2 İngilizce için tüm gürültü ve tüm SNR seviyelerinde yöntemlerin AİE değerleri	41
6.3 İngilizce için beyaz Gauss gürültüsünün farklı SNR seviyelerinde KKAD sonuçları: (a) -10 dB, (b) 0 dB ve (c) 10 dB	43
6.4 Beş farklı SNR seviyesinde Beyaz Gauss gürültüsü koşullarında tüm yöntemlere ait KKAD sonuçları	43
6.5 Türkçe için tüm gürültü ve tüm SNR seviyelerinde BSNR iyileştirme (dB) değerleri	48
6.6 Türkçe için tüm gürültü ve tüm SNR seviyelerinde yöntemlerin AİE değerleri.....	50
7.1 LTSD SEA yönteminin akış diyagramı	59
7.2 Önerilen yöntemin blok diyagramı	60
7.3 AKA ile ayrıştırılan gürültülü konuşma sinyali için ilk 5 İKF ve artık bileşen.....	63

ŞEKİLLER DİZİNİ (devamı)

Şekil

Sayfa

7.4 Araba gürültüsü koşullarında gürültülü konuşma sinyalinin İKF katsayılarının enerjisinin (a) -10 dB, (b) 0 dB ve (c) 10 dB SNR seviyeleri için ampirik histogram, Gauss, Rayleigh ve Gamma dağılımları	66
7.5 Mırıltı gürültüsü koşullarında gürültülü konuşma sinyalinin İKF katsayılarının enerjisinin (a) -10 dB, (b) 0 dB ve (c) 10 dB SNR seviyeleri için ampirik histogram, Gauss, Rayleigh ve Gamma dağılımları	67
7.6 Araba gürültüsü koşullarında (a) -10 dB, (b) 0 dB ve (c) 10 dB SNR seviyeleri için uyarlanır eşik değerleri	72
7.7 Mırıltı gürültüsü koşullarında (a) -10 dB, (b) 0 dB ve (c) 10 dB SNR seviyeleri için uyarlanır eşik değerleri.....	73
7.8 Her kip için eşikten geçirilerek etiketlenen sesli bölgeler	74
7.9 Mırıltı gürültüsü koşullarında 5 dB SNR için önerilen yöntemin SEA sonucu.....	75
7.10 Tüm gürültü ve SNR seviyeleri için SEA yöntemlerine ait HR1 sonuçları ..	76
7.11 Tüm gürültü ve SNR seviyeleri için SEA yöntemlerine ait HR0 sonuçları ..	77
7.12 Tüm gürültü ve SNR seviyeleri için SEA yöntemlerine ait genel doğruluk oranı (GDO) sonuçları	78
7.13 Tüm gürültü için SNR seviyelerinin ortalaması alınarak bulunan HR1, HR0 ve Genel Doğruluk yüzdeleri.....	79
7.14 Her bir SNR seviyesinde tüm gürültü tiplerinin ortalama alınarak bulunan HR1, HR0 ve genel doğruluk yüzdeleri	80
7.15 Tüm gürültü ve SNR seviyeleri için yöntemlere ait HR1 değerleri	82

ŞEKİLLER DİZİNİ (devamı)

<u>Şekil</u>	<u>Sayfa</u>
7.16 Tüm gürültü ve SNR seviyeleri için yöntemlere ait HR0 değerleri.....	83
7.17 Tüm gürültü ve SNR seviyeleri için yöntemlere ait genel doğruluk oranı değerleri	84
7.18 Tüm gürültü tipleri için SNR seviyelerinin ortalaması alınarak bulunan HR1, HR0 ve Doğruluk yüzdeleri	85
7.19 Her bir SNR seviyesinde tüm gürültü tiplerinin ortalama alınarak bulunan HR1, HR0 ve genel doğruluk yüzdeleri.....	86
8.1 Eşikleme fonksiyonları: (a) eşiklemesiz, (b) sert eşikleme, (c) yumuşak eşikleme	91
8.2 AKA ile ayrıştırılan gürültülü konuşma sinyali için ilk 5 İKF ve artık bileşen....	95
8.3 Önerilen konuşma iyileştirme yönteminin blok diyagramı.....	97
8.4 Gürültülü konuşma sinyalinin İKF katsayılarının enerjisinin (a) -10 dB, (b) 0 dB ve (c) 10 dB SNR seviyeleri için ampirik histogram, Gauss ve Gamma dağılımları	101
8.5 Gürültülü sinyalinin İKF katsayılarının enerjisinin (a) -10 dB, (b) 0 dB ve (c) 10 dB SNR seviyeleri için ampirik histogram, Gauss ve Gamma dağılımları	102
8.6 Sert, yumuşak ve yarı-yumuşak eşikleme fonksiyonları için giriş-çıkış ilişkisi.....	107
8.7 İngilizce için tüm gürültülü koşullara ait Bölütsel SNR iyileştirme miktarları.....	110
8.8 Gürültülerin ortalaması alınarak SNR seviyesine göre yöntemlerin Bölütsel SNR iyileştirme miktarları	111

ŞEKİLLER DİZİNİ (devamı)

<u>Şekil</u>	<u>Sayfa</u>
8.9 Gürültülerin ve SNR seviyelerinin ortalaması alınarak yöntemlere ait Bölütsel SNR iyileştirme miktarları.....	112
8.10 İngilizce için tüm gürültülü koşullara ait AİE değerleri.....	113
8.11 Gürültülerin ortalaması alınarak SNR seviyesine göre yöntemlerin AİE değerleri	114
8.12 Gürültülerin ve SNR seviyelerinin ortalaması alınarak yöntemlere ait AİE değerleri	115
8.13 İngilizce için; temiz sinyalin, 5 dB araba gürültüsü ile bozulmuş gürültülü sinyalin, önerilen ve diğer yöntemler ile iyileştirilmiş sinyallerin spektrogramları.....	120
8.14 Türkçe için tüm gürültü koşullarına ait Bölütsel SNR iyileştirme miktarları.....	122
8.15 Gürültülerin ortalaması alınarak SNR seviyesine göre yöntemlerin Bölütsel SNR iyileştirme miktarları.....	123
8.16 Gürültülerin ve SNR seviyelerinin ortalaması alınarak yöntemlere ait Bölütsel SNR iyileştirme miktarları	124
8.17 Türkçe için tüm gürültülü koşullara ait AİE değerleri.....	125
8.18 Gürültülerin ortalaması alınarak SNR seviyesine göre yöntemlerin AİE değerleri	126
8.19 Gürültülerin ve SNR seviyelerinin ortalaması alınarak yöntemlere ait AİE değerleri	127

ŞEKİLLER DİZİNİ (devamı)ŞekilSayfa

8.20 Türkçe için; temiz sinyalin, 5 dB araba gürültüsü ile bozulmuş gürültülü sinyalin, önerilen ve diğer yöntemler ile iyileştirilmiş sinyallerin spektrogramları..... 132



ÇİZELGELER DİZİNİ

<u>Çizelge</u>	<u>Sayfa</u>
4.1 Ses kalite kategorisi ve FDO için beş maddelik ölçek	31
5.1 NOIZEUS veri tabanına ait cümle listesi	34
5.2 Kullanılan Türkçe veri tabanına ait cümle listesi	37
6.1 İngilizce konuşma sinyalleri için farklı yöntem ve gürültülere ait KKAD sonuçlarından elde edilen anlaşılabilirlik sonuçları.....	44
6.2 İngilizce konuşma sinyalleri için farklı yöntem ve gürültülere ait KZNA sonuçlarından elde edilen anlaşılabilirlik oran tahminleri (%).....	46
6.3 Türkçe konuşma sinyalleri için farklı yöntem ve gürültülere ait KKAD sonuçlarından elde edilen anlaşılabilirlik sonuçları.....	51
6.4 Türkçe konuşma sinyalleri için farklı yöntem ve gürültülere ait KZNA sonuçlarından elde edilen anlaşılabilirlik oran tahminleri (%).....	53
7.1 İngilizce konuşmalar için yedi farklı gürültü tipi ve beş farklı SNR seviyesinin ortalaması alınarak yöntemlere ait ortalama başarı yüzdeleri	81
7.2 Türkçe konuşmalar için yedi farklı gürültü tipi ve beş farklı SNR seviyesinin ortalaması alınarak yöntemlere ait ortalama başarı yüzdeleri	87
8.1 İngilizce konuşma sinyalleri için farklı yöntem ve gürültülere ait KKAD sonuçlarından elde edilen anlaşılabilirlik sonuçları.....	116
8.2 İngilizce konuşma sinyalleri için farklı yöntem ve gürültülere ait KZNA sonuçlarından elde edilen anlaşılabilirlik oran tahminleri (%).....	118
8.3 Türkçe konuşma sinyalleri için farklı yöntem ve gürültülere ait KKAD sonuçlarından elde edilen anlaşılabilirlik sonuçları.....	128
8.4 Türkçe konuşma sinyalleri için farklı yöntem ve gürültülere ait KZNA sonuçlarından elde edilen anlaşılabilirlik oran tahminleri (%).....	130

SİMGELER VE KISALTMALAR DİZİNİ

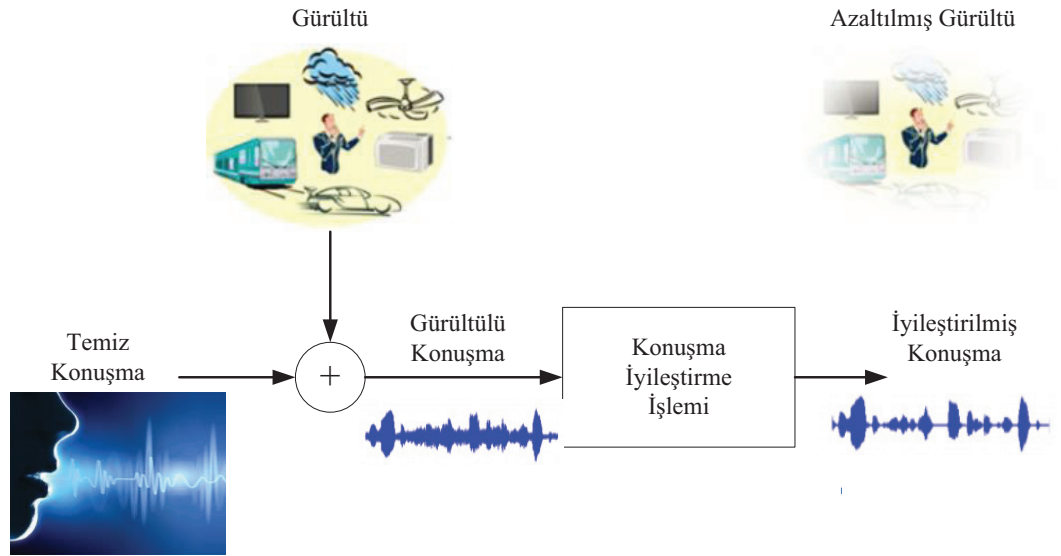
<u>Simgeler</u>	<u>Açıklama</u>
ξ	Önsel SNR
Ψ_d	Teager enerji operatörü
d_j	Orta anlaşılabilirlik ölçütü
p_i	Konuşma olma olasılığı
q_i	Gürültü olma olasılığı
r_i	Temiz konuşma olma olasılığı
$\Gamma(.)$	Gama olasılık fonksiyonu
b	Ölçek parametresi
a	Şekil parametresi
σ_s	Gürültülü konuşma varyansı
σ_n	Gürültü varyansı
τ_k	Eşik parametresi
λ	Eşik değerleri

SİMGELER VE KISALTMALAR DİZİNİ (devamı)

<u>Kısaltmalar</u>	<u>Açıklama</u>
ADD	Ayrık Dalgacık Dönüşümü
AKA	Ampirik Kip Ayırıştırması
İKF	İçkin Kip Fonksiyonu
İÇ	İzgesel Çıkarım
ÇBİÇ	Çoklu-Bant İzgesel Çıkarım
EOKK	En Küçük Ortalama Kareler Kestricisi
KLD	Karhunen-Loeve Dönüşümü
AKLD	Algısal Karhunen-Loeve Dönüşümü
SKL	Simetrik Kullback-Leibler
TEO	Teager Enerji Operatörü
SNR	Sinyal-Gürültü Oranı
BSNR	Bölütsel Sinyal-Gürültü Oranı
AİE	Ağırlıklı İzgesel Eğim
KKAD	Konuşma Kalitesinin Algısal Değerlendirmesi
KZNA	Kısa-Zamanlı Nesnel Anlaşılabilirlik

1. GİRİŞ

Konuşma yoluyla iletişim, insanların temel işlevlerinden biridir. İnsanlar dış dünyadan bilgi almak veya birbirleriyle iletişim kurmak için çeşitli yollara sahiptir ve en önemli üç bilgi kaynağı; konuşma, görüntü ve yazılı metinlerdir. Birçok amaç için konuşma en verimli ve uygun yol olarak öne çıkmaktadır. Konuşma sadece dilsel içeriği değil, aynı zamanda konuşmacının ruh hali gibi diğer yararlı bilgileri de iletmektedir. Konuşmacı ve dinleyici, sessiz bir ortamda birbirine yakın olduğunda, iletişim genellikle kolay ve doğru olmaktadır. Ancak uzaktan veya gürültülü bir arka planda, dinleyicinin anlama kabiliyeti azalmaktadır. Birçok konuşma iletişim sisteminde konuşmanın kalitesi ve anlaşılabilirliği, bilgi alışverişinin kolaylığı ve doğruluğu açısından büyük önem taşımaktadır. Konuşmayı iletme veya depolama için kullanılan konuşma işleme sistemleri genellikle gürültüsüz bir ortam için tasarlanmaktadır. Ancak gerçek ortamda arka plan veya kanal gürültüsünün varlığı, bu sistemlerin performansını önemli ölçüde düşürmekte ve bilgi değişimi gibi yanlışlıklara neden olmaktadır. Şekil 1.1’de gürültülü konuşmaların oluşumu ve iyileştirilmiş konuşmaların elde edilmesi gösterilmektedir.



Şekil 1.1 Konuşma iyileştirme işlemi

Konuşma iyileştirme, gürültülü konuşma sinyallerinin işlenmesiyle, insanların duyma algılarını veya makineler tarafından doğru kod çözme amaçlamaktadır. Konuşma iyileştirme algoritmaları, konuşma sinyalleri gürültü

ile bozulduğunda iletişim sistemlerinin performansını geliştirmeye çalışmaktadırlar. Arka plan gürültüsünün varlığı, konuşmanın kalitesini ve anlaşılabilirliğini bozmaktadır. Burada konuşmanın niteliği, konuşmacının bir konuşmayı nasıl ilettiğini ve içerik olarak doğallık ve konuşmacı tanınabilirliği gibi özellikleri ifade etmektedir. Anlaşılabilirlik, konuşmacının ne söylediği ile yani kelimelerin arkasındaki anlam veya bilgi içeriğiyle ilgilidir. Bu nedenle, gürültülü ortam konuşmacı ile dinleyici arasındaki iletişim yeteneğini azaltmaktadır. Bu sorunun etkisini azaltmak için konuşma iyileştirme yapılması gerekmektedir. Konuşmanın kalite ve anlaşılabilirliğinin arka plan gürültüsünden olumsuz etkilenmesi; konuşma kodlama, konuşma ve konuşmacı tanıma, işitme cihazları ve mobil telefon haberleşme gibi alanlarda konuşma iyileştirme algoritmalarının geliştirilmesi ihtiyacını doğurmaktadır. Konuşmayı bozmadan gürültüyü azaltmak genellikle zordur ve bu nedenle konuşma iyileştirme sistemlerinin performansı, konuşma bozulma (distortion) ve gürültü azaltma arasındaki farkla sınırlı olmaktadır. Konuşma iyileştirmenin amacı, dinleyicinin yorgunluğunu azaltmak, genel konuşma kalitesini arttırmak, anlaşılabilirliği sağlamak ve ses iletişim cihazının performansını iyileştirmek gibi belirli uygulamalara göre değişmektedir.

Son yıllarda, araştırmacılar bozulmuş konuşmadan temiz konuşmayı elde etmek için birçok yöntem önermiş ve geliştirmişlerdir. Ancak konuşma sinyalinin karmaşık yapısından dolayı, konuşma ve arka plan gürültüsü karışımından istenen konuşma sinyalinin yeniden elde edilmesi, konuşma işleme ve iletişim sistemi araştırmasında hala önemli bir sorun teşkil etmektedir.

1.1 Genel Gürültü Kaynakları ve Çeşitleri

Gürültü tipi, konuşma iyileştirmede kullanılacak yönteme karar vermede önemli bir etkidir. Bu nedenle iyi bir gürültü modeli çıkartmak konuşma iyileştirme sisteminin performansı için önemlidir. Bir konuşma iyileştirme algoritmasının ya da modelinin farklı gürültü türlerinde gösterdiği başarıyı yöntemin kabul edilebilirliği açısından belirleyici olmaktadır.

Gürültüler; çeşitli istatistiksel, izgesel veya uzaysal (spatial) özelliklerine göre farklılıklar gösterebilir. Gürültü kaynaklarının niteliğine ve özelliklerine bağlı olarak gürültü sinyali, arka plan gürültüsü, hoparlörden gelen gürültü gibi konuşmalar, dürtü gürültüsü, evrimsel gürültü ve çarpımsal gürültüler olarak sınıflandırılabilir. Genel olarak, gürültünün özellikleri hakkında önceden bir bilgi

bulunmayan durağan olmayan (non-stationary) gürültüler ile uğraşmak zordur. Durağan olmayan gürültüler zamana göre değiştiği için gürültüyü konuşmanın başlangıç zaman aralıklarından tahmin etmek iyileştirme başarısını düşürmektedir.

1.1.1 Toplamsal gürültü

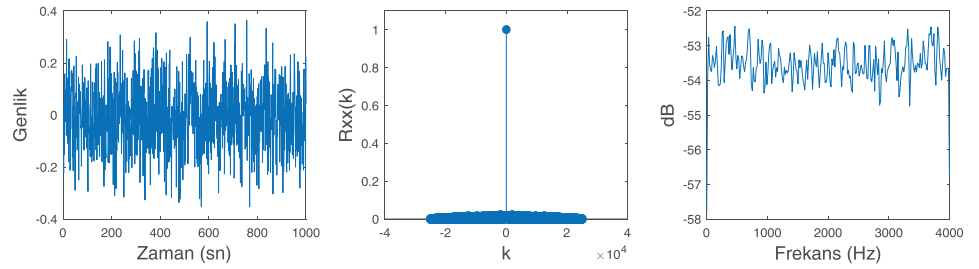
Toplamsal gürültüler, istenen sinyale toplamsal olarak eklenebilen sinyallerdir. Temiz konuşma sinyali $x(n)$ ve toplamsal gürültü sinyali $v(n)$ olması durumunda gürültülü konuşma sinyali $y(n)$ şu şekilde ifade edilir:

$$y[n] = x[n] + v[n] \quad (1.1)$$

Beyaz ve renkli olarak adlandırılan farklı türde toplamsal gürültüler bulunmaktadır.

Beyaz Gürültü

Beyaz gürültü tüm frekanslarda eşit güce sahip ilintisiz sinyaller olarak tanımlanmaktadır. Örneğin, 10 kHz bant genişliğine sahip bir ses sistemi için, 10 kHz'den büyük bir bant genişliğine sahip herhangi bir düz izge (spectrum) gürültüsü beyaz gürültü olarak görünmektedir. Şekil 1.2'de sıfır ortalamalı bir beyaz gürültünün öz-ilişki (autocorrelation) işlevinin bir delta işlevi ve güç izgesinin sabit olduğu görülmektedir.

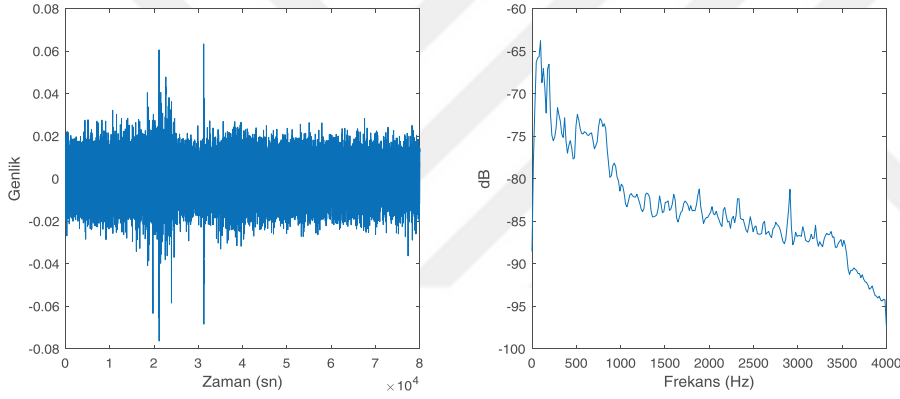


Şekil 1.2 Beyaz gürültünün genlik-zaman, öz-ilişki ve güç izgesel yoğunluk gösterimi

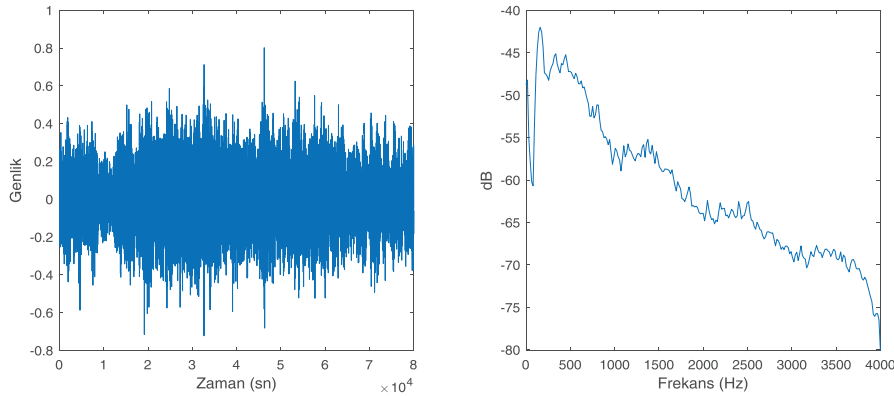
Renkli Gürültü

Beyaz gürültü kavramı, ses iletim sistemlerinde karşılaşılan bazı baskın gürültü süreçlerine matematiksel olarak uygun ve kullanışlı bir yaklaşım sağlayabilmektedir. Ancak birçok gürültü süreci beyaz değildir.

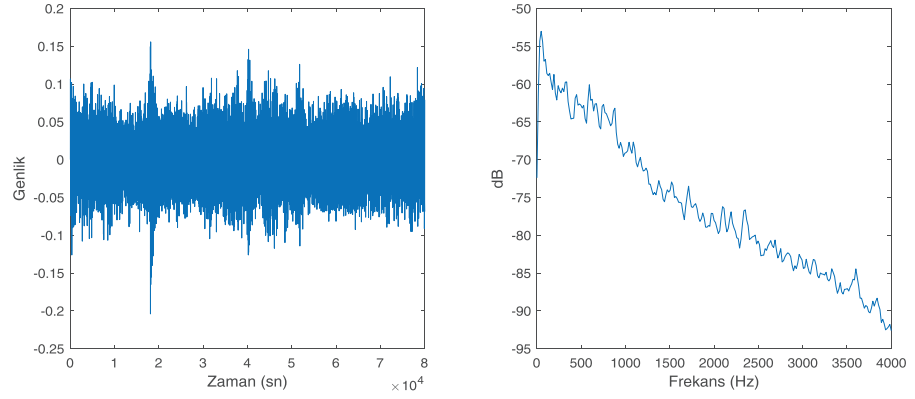
Renkli gürültü terimi, beyaz olmayan herhangi bir geniş bant gürültüsünü ifade eder. Örneğin, hareketli arabaların gürültüsü, tren gürültüsü, restoran gürültüsü, bilgisayar fanlarının gürültüsü, elektrikli matkap gürültüsü ve arka planda konuşan insan gürültüsü, baskın olmayan düşük frekanslı izgelere sahiptir. Ayrıca, bir kanaldan geçen beyaz bir ses, kanal izgesinin şekliyle renklendirilmektedir. Şekil 1.3’de araba, mırıltı ve hava alanı gürültüsü için genlik-zaman ve güç izgesel yoğunluk gösterimleri verilmiştir.



(a)



(b)



(c)

Şekil 1.3 Gürültülerin zaman ve güç izgesel yoğunluk gösterimleri: (a) araba, (b) mırıltı ve (c) hava alanı

1.1.2 Çarpımsal Gürültü

Sinyal işlemede çarpımsal gürültü terimi, yakalama, aktarma veya diğer işlemler sırasında bazı ilgili sinyale çarpan istenmeyen bir rastgele sinyale karşılık gelir. Konuşma sinyallerinin gürültü ile çarpımsal olarak bozulması iyileştirme işlemini zorlaştırmaktadır.

1.1.3 Dürtüsel Gürültü

Dürtüsel gürültü, anahtarlama gürültüsü, iletişim sistemindeki olumsuz kanal ortamı, ses kayıtlarının bırakılması veya yüzey bozulumu, bilgisayar klavyelerindeki tıklamalar gibi çeşitli kaynaklardan gelen kısa süreli “açma/kapama” darbelerinden oluşur.

1.1.4 Durağan ve Durağan Olmayan Gürültü

Matematik ve istatistiklerde, durağan bir süreç (tam olarak durağan bir işlem ya da güçlü bir durağan süreç), zaman içinde kaydırıldığında ortak olasılık dağılımı değişmeyen stokastik bir süreçtir. Durağan olmayan gürültüler zamanla değişmektedir ve bu durum iyileştirme işlemini zorlaştırmaktadır. Durağan gürültüler ise zamanla değişmez ve konuşma olmayan zaman aralıklarından tespit edilmesi daha kolay olmaktadır. Gürültünün durağan olması iyileştirme sürecinde

daha başarılı sonuçlar alınmasını sağlamaktadır.

1.2 Hedefler

Konuşma iyileştirmenin temel amacı, konuşmanın kalite ve anlaşılabilirliğini artırarak iletişim sistemlerinde, makina kod çözmede, konuşma ve konuşmacı tanıma sistemlerinde performansı arttırmak ve konuşmacı ile dinleyici arasında sağlıklı bir iletişim kurulmasına olanak sağlamaktır. Bu araştırma için ana odak, en sık karşılaşılan sorunlardan biri olan dinleyiciler için toplamsal gürültülerin bulunduğu konuşmaların iyileştirilmesi olacaktır. Bu nedenle araştırmanın temel amaçları şunlardır:

Literatürde sıklıkla kullanılan tek-kanal konuşma iyileştirme yöntemlerini ve standart haline gelmiş Sesli Etkinlik Algılama (SEA) yöntemlerini karşılaştırmak ve Türkçe ile İngilizce dilleri arasındaki farklılıkları tespit etmek,

Sesli etkinlik algılama ve konuşma iyileştirme için farklı gürültü tiplerinde ve farklı SNR (sinyal gürültü oranı) seviyelerinde literatürdeki yöntemlerden başarılı olacak bir yöntem geliştirmek.

1.3 Katkılar ve Özgünlükler

Bu tezin ana katkıları ve özgünlükleri şunlardır:

Literatürde sıklıkla kullanılan alt-uzay gibi zaman alanı, temel izgesel çıkarım, çoklu-bant izgesel çıkarım, en küçük ortalama kareler kestiricisi (EOKK) ve Wiener süzgeçleme gibi frekans alanı, Dalgacık dönüşümü ve Ampirik Kip Ayırıştırması gibi zaman-frekans alanı yöntemlerinin hem İngilizce hem de Türkçe için iyileştirilmiş konuşmaların kalite ve anlaşılabilirlik değerlendirmesinin yapılması ile bu yöntemlerin üstünlükleri ve sakıncaları belirlenmiştir.

İngilizce cümleler için bilimsel araştırmalarda kabul görmüş ve onaylanmış olan TIMIT veri tabanından fonetik olarak dengeli ve zorluk düzeyleri eşit olarak seçilmiş 30 cümleden oluşan NOIZEUS veri tabanı kullanılmaktadır. Konuşma iyileştirme alanında Türkçe için kullanılan bir veri tabanı bulunmadığından, 2006 yılında ODTÜ’de Türkçe dili için sıklıkla kullanılan kelimelerin bir araya getirilmesiyle oluşturulan cümlelerden 3 erkek ve 3 kadından alınan 5’er cümleden bir veri tabanı oluşturulmuştur. Oluşturulan bu veri

tabanı NOIZEUS veri tabanına benzer şekilde Türkçe dili için fonetik olarak dengeli ve zorluk düzeyleri eşit cümlelerden seçilmiştir.

Çalışmadaki ilk ana özgünlüğümüz, konuşma işleme alanında etkili ve doğru bir şekilde kullanılabilecek bir Sesli Etkinlik Algılama (Voice Activity Detection-VAD) algoritmasının önerilmesidir. Farklı gürültü tipi ve SNR seviyelerinde gürültülü konuşmadan konuşma bölgelerini ve konuşma olmayan bölgeleri algılayabilen ve bu bölgeleri etiketleyen Ampirik Kip Ayırıştırması (Empirical Mode Decomposition-EMD) tabanlı bir yaklaşım kullanılmıştır. Zaman-frekans alanı yöntemlerinden dalgacık dönüşümünde sabit bir temel dalgacık işlevi kullanılmakta ve eşik değerleri yalnızca enerji üzerinden çıkarılmaktadır. Eşik değerlerinin bu şekilde bulunması gürültülü ortamlarda konuşma olmayan bölgelerin tespitini zorlaştırmaktadır. Bizim önerdiğimiz yaklaşımda, Ampirik Kip Ayırıştırma (AKA) yöntemi kullanılarak ayrıştırılan gürültülü konuşmaların içkin (intrinsic) kip fonksiyonu katsayılarının Teager enerjisi bulunarak Gamma dağılım işlevi yardımıyla uyarlanabilir eşik değerleri hesaplanmaktadır. Uyarlanabilir eşik değerlerinin konuşmanın yapısına uygun olarak Gamma dağılım işlevi ile hesaplanması sesli etkinlik algılama yöntemleri bakımından bu çalışmanın özgünlüğüne katkıda bulunmaktadır. Ayrıca, literatürdeki diğer yöntemlerden farklı olarak önerilen yöntemde konuşma ve konuşma dışı bölgelerin enerji olasılık dağılımı üzerinden belirlenmesi konuşmanın özellikle yüksek frekanslı bileşenlerinin tespitini kolaylaştırmaktadır.

Konuşma iyileştirmede, literatürde sıklıkla kullanılan zaman, frekans ve zaman-frekans alanı yöntemleri gürültü giderildikten sonra iyileştirilmiş konuşmanın kalitesini belli ölçüde artırırken anlaşılabilirliği sağlamada yetersiz kalmaktadır. İngilizce dili için tasarlanmış konuşma iyileştirme yöntemlerinin dillerin yapısı gereği Türkçe için aynı performans ve başarıyı göstermesi zor olmaktadır. Bu nedenle, hem İngilizce hem de Türkçe için kalite ve anlaşılabilirliği sağlamak için zaman-frekans alanı yöntemlerinden olan AKA kullanılmıştır. Önerdiğimiz konuşma iyileştirme yönteminde, gürültülü konuşmanın ve gürültünün içkin kip fonksiyonu katsayılarının Teager enerjileri hesaplanarak bu enerji dağılımına uygun bir dağılım olan Gamma dağılım işlevi ile temiz konuşma olma olasılığı ve gürültü olma olasılığı üzerinden uyarlanabilir eşik değerleri hesaplanmıştır. Uyarlanabilir eşik değerlerinin konuşmanın yapısına uygun olarak Gamma dağılım işlevi ile hesaplanması konuşma iyileştirme yöntemleri bakımından bu çalışmanın özgünlüğüne katkıda bulunmaktadır. Bu nedenle dil yapıları farklı olan İngilizce ve Türkçe için önerilen yöntemin başarılı

olması literatüre bir yenilik katmıştır.

Önerilen Sesli Etkinlik Algılama seçilmiş bir kaç güncel yöntem ile birlikte konuşma ve konuşma dışı doğruluk oranı ölçütleri ile, önerilen konuşma iyileştirme yaklaşımı ise nesnel kalite ve anlaşılabilirlik ölçütleri ile değerlendirilmiştir. Ayrıca, tüm yöntemlere ait sonuçlar farklı gürültü tipi ve farklı SNR seviyeleri için hem İngilizce hem de Türkçe için ayrıntılı olarak açıklanmıştır.

1.4 Tezin Organizasyonu

Bu tez şu şekilde düzenlenmiştir; Konuşma iyileştirmenin temelleri ve uygulamaları, gürültü türleri ve kaynakları birinci bölümde anlatılmıştır. İkinci bölüm konuşma iyileştirme ve sesli etkinlik algılama algoritmaları için kapsamlı bir literatür incelemesi sunmaktadır. Üçüncü bölümde, literatürde sıklıkla kullanılan konuşma iyileştirme yöntemleri açıklanmaktadır. Değerlendirmede kullanılan nesnel kalite ve anlaşılabilirlik ölçütleri dördüncü bölümde tanımlanmıştır. Beşinci bölümde kullanılan konuşma ve gürültü veri tabanları verilmiştir. Literatürdeki konuşma iyileştirme algoritmalarının İngilizce ve Türkçe dilleri için tüm koşullarda nesnel kalite ve anlaşılabilirlik değerlendirme sonuçları altıncı bölümde ayrıntılı olarak verilmiştir. Yedinci bölümde önerilen Ampirik Kip Ayırıştırma tabanlı yöntem açıklanarak yönteme ait başarımlar diğer yöntemler ile karşılaştırılarak verilmiştir. Sekizinci bölümde önerilen Ampirik Kip Ayırıştırma tabanlı konuşma iyileştirme yöntemi tüm aşamaları ile açıklanarak hem İngilizce hem de Türkçe için kalite ve anlaşılabilirlik sonuçları ayrıntılı olarak açıklanmıştır. Son olarak, dokuzuncu bölümde tez ile ilgili genel sonuçlar, öneriler ve gelecek çalışmalar sunulmuştur.

2. LİTERATÜR TARAMASI

2.1 Konuşma İyileştirme Literatür Taraması

Konuşma sinyallerinin arka plan gürültüler ile bozulması, konuşmanın kalite ve anlaşılabilirliğini olumsuz etkilemektedir. Bu durum; konuşma kodlama, konuşma tanıma, işitme cihazları ve mobil telefon haberleşme gibi alanlarda konuşma iyileştirme algoritmalarının geliştirilmesi ihtiyacını doğurmaktadır. Son yıllarda, konuşma iyileştirme problemlerinde konuşmayı bozmadan gürültüyü azaltmak için çeşitli yöntemler geliştirilmiştir. Bu yöntemler; zaman alanı, frekans alanı ve zaman-frekans alanı olmak üzere üç temel sınıfa ayrılmaktadır. Zaman alanı yöntemleri olarak alt-uzay (Loizou et al., 2005), frekans alanı yöntemleri olarak izgesel çıkarım (İÇ) ve çoklu-bant izgesel çıkarım (ÇBİÇ) (Lu and Loizou, 2008; Paliwal and Schwerin, 2009), en küçük ortalama kareler kestiricisi (EOKK) (Yi Hu and Loizou, 2007) ve Wiener süzgeçleme (Chehrehse and Savoji, 2014) gibi yaklaşımlar ve zaman-frekans yöntemleri olarak da Dalgacık Dönüşümü (Cohen et al., 2001; Ghanbari and Karami-Mollaei, 2006) ve Ampirik Kip Ayırıştırması (Empirical Mode Decomposition-EMD) yaklaşımları (Khaldi et al., 2008; Chatlani and Soraghan, 2009; Hamid et al., 2013; Rudresh et al., 2016) konuşma iyileştirmede kullanılmaktadır.

Konuşma iyileştirmede, alt-uzay yaklaşımı gibi zaman alanı yöntemleri, konuşma bozulması ve artık (residual) gürültülerin oluşmasına neden olmakla birlikte hesaplama yükü bakımından gerçek zamanlı kullanılması çok zor olmaktadır. Frekans alanı yöntemlerinden izgesel çıkarım yaklaşımı, tarihsel olarak gürültü azaltma için önerilen ve basit bir prensibe dayanan ilk algoritmalarından biridir. Basit bir algoritma olan bu yöntemde “müzikal gürültü” olarak bilinen rahatsız edici bir ses, konuşma sinyalini bozmaktadır. Birçok algoritma, bu istenmeyen gürültünün giderilmesini amaçlamıştır (Lu and Loizou, 2008). İzgesel çıkarım algoritmasında, gürültünün durağan ya da yavaş değiştiği kabul edilerek gürültü izgesinin güncellenen periyotlar arasında önemli bir değişikliğe uğramadığı düşünülür. Bu yöntemde iyileştirilmiş sinyal, gürültülü sinyalin fazı kullanılarak kestirilen sinyalin izgesi ters ayırık Fourier dönüşümü ile elde edilmektedir. Algoritma yalnızca ileri ve ters Fourier dönüşümü kullanılmaktadır ve hesaplanabilirlik açısından oldukça basittir. Bu yöntemin en büyük sakıncalarından biri fazla izge çıkarımında konuşmanın içinde bulunan bazı bilgilerin kaybolması veya az izge çıkarımında gürültünün yeterince atılmamış olmasıdır. Kalamani et al. (2014) çalışmalarında konuşma iyileştirme için izgesel

çıkarım algoritması tabanlı değiştirilebilir modülasyon büyüklüğü kestirimi (Modified Modulation Magnitude Estimation-MMSE) ile gürültü kestirimi yaparak gürültü varyansını tahmin etmişlerdir. Bu çalışmada Chi-square dağılım tabanlı gürültü kestirim algoritması bir iyileştirme yöntemi olarak önerilmiştir.

Frekans alanı yöntemlerinden istatistiksel model tabanlı algoritmalar ise konuşma sinyalinde gürültü azaltma yöntemleri arasında en yaygın kullanılan yöntemlerden biridir (Ephraim and Malah, 1985; Gui and Kwan, 2005; Hu and Loizou, 2007; Naidu and Srinivasan, 2014). Temiz konuşma sinyalinin elde edilmesinde, dönüşüm katsayıları ya da büyüklükleri bilinen konuşma ve gürültü sinyali istatistikleri bir Bayes kestirimi işleminden geçirilir. Birçok kestirici, gürültü ve konuşma dağılımları için farklı varsayımlar altında elde edilmiştir. Bir istatistiksel yöntem olan Wiener süzgeç algoritması, gürültülü konuşma izgesinden ortalama karesel hatanın minimize edilmesi ile bir en iyi (optimal) süzgeç kestirimi yardımıyla yapılmaktadır (Chehrehse and Savoji, 2014; Hooda and Aggarwal, 2012; Gui and Kwan, 2005; Hu and Loizou, 2007). Frekans alanı yöntemlerinden en küçük ortalama kareler kestiricisi (EOKK) ve Logaritmik EOKK gibi istatistiksel model tabanlı konuşma iyileştirme algoritmalarında (Ephraim, 1992) çeşitli istatistiksel modeller ve eniyileme kriterleri kullanılarak sinyalin karmaşık izgesinden ziyade ayrık Fourier katsayılarından genliğin doğrusal olmayan kestirimi üzerinde durulmaktadır. Bu yöntemin en büyük sakıncalarından biri izgesel katsayıların varyansının oldukça büyük olmasıdır. Konuşma iyileştirme yöntemlerinde, bu eksikliklerin üstesinden gelmek için, Dalgacık Dönüşümü ve Ampirik Kip Ayırıştırması gibi doğrusal olmayan zaman-frekans alanı yöntemleri önerilmiştir.

Bir diğer konuşma iyileştirme algoritması olan Dalgacık Dönüşümü tabanlı yöntemler, çeşitli ses ve konuşma uygulamalarında sıklıkla kullanılmıştır. Dalgacık analizinde konuşma iyileştirmenin temel prensibi; gürültülü konuşma sinyalinden temiz konuşma sinyali ve gürültü bileşenlerini ayırmada ayrık dalgacık katsayılarının eşikten (threshold) geçirilmesine dayanmaktadır. Ancak, konuşma sinyalinin dalgacık katsayılarına sabit eşik (fixed threshold) uygulandığında gürültü ile birlikte bazı sessiz sesler de (unvoiced speech) kaybolmaktadır ve bu durum iyileştirilmiş konuşma sinyalinin anlaşılabilirliğini olumsuz etkilemektedir. Bu problemi çözmek için eşik değerleri uyarlanabilir olmalıdır. Dalgacık Dönüşümünün en büyük sakıncası ise frekans bandı sayısının sınırlı olmasıdır. Ayrıca, gürültülü konuşma sinyallerinin sessiz (unvoiced) kısımlarının, dalgacık büzülmesi (shrinking) bakımından bir problem olduğu

kanıtlanmıştır. Konuşma sinyallerinin gürültüye benzeyen birçok yüksek frekanslı bileşen içeren sessiz (unvoiced) kısımlarından dolayı gürültünün dalgacık alanında giderilmesi, iyileştirilmiş konuşmanın anlaşılabilirliğini ve kalitesini düşürmektedir (Messaoud et al., 2015).

Dalgacık Dönüşümü tabanlı konuşma iyileştirme yöntemlerinin birçoğunda dalgacık katsayılarının birbirinden bağımsız olduğu varsayılmaktadır. Ancak dalgacık katsayılarının ortak (joint) histogramına bakıldığında ise bu katsayıların bazılarının bağımlılıklarının olduğu görülmektedir (Sendur and Selesnick, 2002). Thodypour ve Ahadi (2015) tarafından yapılan çalışmada, iki değişkenli büzülmeye dayalı konuşma iyileştirme için en uygun dalgacık yapısının bulunması amaçlanmış olup, iki değişkenli büzülme işlevi konuşma iyileştirme için optimize edilmiştir. Bahoura ve Rouat (2006) çalışmalarında konuşma iyileştirmede gürültünün bilinmesine gerek olmadan dalgacık paket katsayılarının zaman-ölçek eşik tabanında kullanılmasına dayalı yeni bir öneri sunmuşlardır. Burada eşik değerleri, gürültülü konuşmalarda zaman ve frekans alanını yansıtan doğrusal olmayan bir maske ile modüle edilmiştir. Gürültülü konuşma sinyalleri TIMIT veri tabanından alınan temiz konuşma sinyallerine farklı sinyal-gürültü oranlarında (SNR) beyaz gürültü eklenmesi ile oluşturulmuştur. Dalgacık paket dönüşümü kullanılarak oluşturulan 32 alt banttaki dalgacık katsayıları zaman-uyarlanmış (time-adapted), ölçek-uyarlanmış (scale-adapted) ve zaman-ölçek-uyarlanmış (time and scale-adapted) eşikleme uygulanarak iyileştirme başarıları karşılaştırılmıştır. Burada yüksek SNR’larda zaman-uyarlanmış (ZU), zaman-ölçek-uyarlanmış (ZÖU) ve EMF (Ephraim and Malah, 1984)’den daha iyi bir iyileştirme sağlamıştır. ZU ve ZÖU eşikleme birlikte kullanıldığında ise EMF’den daha iyi sonuçlar verdiği görülmüştür.

Donoho (1995) ve Donoho ve Johnstone (1994) konuşma sinyallerine eklenmiş beyaz gürültüyü gidermek için geniş kapsamlı bir dalgacık eşikleme önermişlerdir. Bir diğer yaklaşımda ise Johnstone ve Silverman (1997) eklenen beyaz gürültüyü gidermede zaman-bağımlı eşikleme önermişlerdir. Ancak tüm bu eşikleme yöntemleri, bazı konuşma bileşenlerini bastırması sebebi ile konuşma iyileştirmede çok başarılı olamamışlardır (Bahoura and Rouat, 2006). Ghanbari ve Karami (2006) çalışmalarında dalgacık paket katsayılarının kullanılması ile uyarlanabilir eşikleme tabanlı konuşma iyileştirme yapmışlardır. Önerilen algoritma TIMIT veri tabanından alınan bir İngilizce cümle ile test edilmiştir. Dalgacık işlevi olarak 8. dereceden Daubechies (db8) kullanılmış olup, diğer dalgacık ailelerinin kullanılması sonuçlar üzerinde önemli değişiklikler

yaratmamıştır. Beyaz Gauss gürültüsü, pembe ve mırıltı gürültü tipleri için farklı SNR seviyeleri (-5, 0, 5, 10, 15 dB) kullanılmış olup, yöntemin performansı konuşma tanıma sistemi ile test edilmiştir. Sesli Etkinlik Algılaması ile gürültü tespitinin yapıldığı bu yöntem daha önce yapılmış temel izgesel çıkarım algoritması (Boll, 1979), çoklu bant izgesel çıkarım (Kamath and Loizou, 2002), dalgacık alanında izge çıkarımı (Ghanbari and Karami, 2004), Donoho (1995) tarafından önerilen ve Biyonik Dalgacık Dönüşümü (Bionic Wavelet Transform-BWT) olarak bilinen temel dalgacık eşikleme algoritması, gelişmiş dalgacık tabanlı konuşma iyileştirme (Sheikhzadeh and Abutalebi, 2001) ve dalgacık alanı gürültü bileşeni azaltma ile konuşma iyileştirme (Seok and Bae, 1997) olarak isimlendirilen altı farklı algoritma ile karşılaştırılmıştır. Önerilen yöntemin diğer yöntemlerden daha iyi sonuçlar verdiği spektrogram üzerinden gösterilmiştir.

Messaoud et al. (2015) çalışmalarında; Temel Bileşenler Analizinin (TBA) gelişmiş bir versiyonu ve Dalgacık Paket Dönüşümünün (DPD) bir birleşimi olan tek kanallı konuşma iyileştirme yöntemini önermişlerdir. Bu yöntem konuşma iyileştirme için kullanılan öznitelik vektörlerini elde etmek için DPD analizi ile doğrusal bir ilişki çıkararak katsayıları ilintisizleştirmede TBA'nın yeteneklerini bütünleştirmektedir. Bu durum konuşma sinyalini bozmadan gürültü atmada yeni katsayılar üzerinde uygun bir büzülme (shrinking) işlevi ile çalışmasına olanak sağlamaktadır. Daha sonra, iyileştirilmiş konuşma sinyali Ters Dalgacık Paket Dönüşümü (TDPD) ile ayrıştırılarak düşük rank, aralık (sparse) ve gürültü bileşeni olmak üzere üç alt uzay elde edilmiştir. Sonuç olarak, bu alt uzay bileşenleri gürültünün hesaplanmasında kullanılarak önerilen yöntemin performansı test edilmiştir. Daubechies dalgacık işlevinin kullanıldığı bu yöntemde oldukça yüksek gürültülü koşullarda bozucu yapılar (artifacts) oluşmadan düşük sinyal bozulması ve yüksek gürültü azaltma sağlanmıştır. Ayrıca, önerilen yöntem beyaz, mırıltı, fabrika, araba ve sokak gibi çeşitli gürültülü konuşmalarda test edilmiştir. Rissanen'e (2000) göre Donoho ve Johnstone (1994) tarafından önerilen eşikleme tekniği çok fazla dalgacık katsayısının bastırılmasına yol açmaktadır. Bu nedenle gürültü gidermede Minimum Tanımlama Uzunluğu (Minimum Description Length) model seçme kriteri ile eşikleme değerinin belirlenmesi daha uygun bulunmuştur (Whitmal et al., 1996; Messaoud et. al., 2015). Burada kullanılan minimum tanımlama uzunluğu kriteri kovaryans matrisi tabanlı bir yöntem olup, geri çatma işlemi ise örtüşme ve ekleme metodu ile yapılmıştır. Nesnel değerlendirme olarak Kör Sinyal Ayırma Yöntemlerinin Değerlendirilmesi (Evaluation of Blind Singal Separation Methods-BSS EVAL) araç kutusunda (Vincent et. al., 2006) bulunan

ve konuşma iyileştirme performansını ölçen Konuşma-Bozulma Oranı (Speech-to-Distortion Ratio - SDR), Konuşma-Girişim Oranı (Speech-to-Interference Ratio - SIR) ve Konuşma-Bozucu yapılar Oranı (Speech-to-Artifacts Ratio - SAR) olmak üzere üç ölçüt kullanılmıştır. Burada SDR ölçütü; konuşma iyileştirme yönteminin tüm kalitesini (overall quality), SIR ölçütü; gürültü azaltma oranını gösterirken SAR ise iyileştirme yöntemi tarafından tanımlanan bozucu etkilerin yansımaları göstermektedir. Nesnel ve öznel değerlendirme sonuçları, kullanılan yöntemin diğer üç yöntemle göre (izgesel çıkarım, istatistiksel model tabanlı ve alt-uzay yöntemleri) daha iyi sonuçlar verdiğini göstermiştir.

Sanam ve Shahnaz (2013) tarafından yapılan çalışmada, dalgacık paket alanında değiştirilebilir sert eşikleme işlevi ve uyarlanabilir eşikleme kullanılarak konuşma iyileştirme yapılmaktadır. Bu çalışmada, eşik değerleri ve eşikleme fonksiyonu belirlendikten sonra katsayılar eşikten geçirilmekte ve her çerçeve için standart örtüşme ve ekleme yöntemi ile geri çatma yapılmaktadır. Benzer çalışmalarda eşik değerleri doğrudan Teager enerji ile belirlenirken bu çalışmada eşik değerleri her bir alt banttaki dalgacık katsayılarının Teager enerjisi bulunarak elde edilmektedir. Katsayıları doğrudan eşikten geçirmek yerine, enerjinin olasılık yoğunluk işlevinden elde edilen uyarlanabilir eşik değerleri ile eşikleme işlemi yapılmaktadır. Teager enerji operatörü (TEO) zamanla değişen sinyalin enerjisinin kestirilmesini sağlarken konuşmanın olduğu ya da olmadığı kısımları elde etmede kullanılmaktadır. Ancak gürültülü konuşma sinyali üzerinde doğrudan TEO uygulamak, istenmeyen bozucu etkilerin oluşmasına yol açmaktadır. Bu nedenle, gürültülü konuşmalara doğrudan TEO uygulamak yerine gürültülü konuşmanın dalgacık paket katsayıları üzerinden TEO uygulamak daha makul bulunmuştur. Bu yöntemle bulunan sonuçlar nesnel ölçütler olan Sinyal Gürültü Oranı (SNR), Konuşma Kalitesinin Algısal Değerlendirmesi (Perceptual Evaluation of Speech Quality - PESQ) ve Ağırlıklı İzge Eğim (Weighted Spectral Slope - WSS) ölçütü ile test edilmiştir. Aynı zamanda, öznel olarak dinleme testi yapılmış olup, temiz, gürültülü ve iyileştirilmiş konuşmaların spektrogramlarına bakılarak yöntemin başarısı test edilmiştir. Geleneksel eşikleme tabanlı yöntemlerin aksine, kullanılan bu yöntemde dalgacık paket katsayılarının değiştirilebilir ve uyarlanabilir sert eşikleme ile gürültü gidermede daha etkili olduğu sonucuna varılmıştır. Simülasyon sonuçları bu yöntemin yüksek SNR, yüksek konuşma kalite algısı ve düşük ağırlıklı izgesel eğim bakımından diğer yöntemlere göre konuşma iyileştirmede başarılı bir yöntem olduğunu göstermiştir.

Swami et. al. (2015) tarafından yapılan çalışmada, algısal olarak ölçeklenmiş sürekli dalgacık katsayılarının kullanılması ile elde edilen eşikleme tabanlı konuşma iyileştirme yönteminde öncelikle gürültü seviyesi kestirilmektedir. Kestirilen gürültü seviyesinden ölçek sayısının bulunması ile bu ölçeklerdeki katsayılar eşikten geçirilmektedir. Bu yöntemde sürekli dalgacık katsayıları logaritmik olarak ölçeklenmiştir ve kullanılacak ölçek sayısı önerilen yöntemin temelini oluşturmaktadır. Gürültü tipi olarak beyaz Gauss gürültüsü, pembe, mırıltı, araba ve F16 kokpit kullanılmıştır. Önerilen yöntem, Ephraim ve Malah algoritması, Stein'in tarafsız risk tahmini (Stein's unbiased risk estimate-SURE), Biyotik dalgacık dönüşümü (BDD), BDD-SURE, Wiener süzgeç, Algısal dalgacık paket dönüşümü (Perceptually Wavelet Packet Transform), çok modellenli Wiener süzgeç ve çoklu model seyrek kod büzülme (multi-model sparse code shrinkage- MultiSCS) iyileştirme yöntemleri ile karşılaştırılmıştır. Çalışmada elde edilen sonuçlar SNR, bölütsel SNR, izgesel bozulma (spectral distortion) ve konuşma kalitesinin algısal değerlendirilmesi (KKAD) gibi nesnel kalite ölçütleri ile test edilmiştir. Beyaz Gauss gürültüsünün olduğu durumlarda tüm gürültü seviyelerinde, önerilen yöntemin SNR ve bölütsel SNR sonuçları diğer tüm yöntemlerden daha iyi bulunmuştur. İzgesel bozulma ve KKAD sonuçlarına bakıldığında ise 10 dB SNR'da önerilen yöntem çoklu model seyrek kod büzülme yöntemine göre daha kötü bulunurken 15 ve 20 dB SNR'larda daha iyi bulunmuştur. Pembe gürültünün kullanıldığı durumlarda ise önerilen yöntem diğer yöntemlere göre en iyi sonuçları vermiştir.

Konuşma iyileştirme algoritmalarının en önemli amacı normal ya da işitme bozukluğu olan kişilerde gürültüden arındırılmış konuşmaların anlaşılabilirliğini arttırmaktır. İlk anlaşılabilirlik çalışması 1970'lerin sonunda Lim (1978) tarafından yapılmıştır. Bu çalışmada -5, 0 ve 5 dB sinyal-gürültü oranlarında beyaz Gauss gürültü ile bozulan sesler için izgesel çıkarım algoritması ile anlaşılabilirlik yeterince sağlanamamıştır. Bu çalışmadan 30 yıl sonra, Hu ve Loizou (2007) tarafından yapılan anlaşılabilirlik çalışmasında kullanılan sekiz farklı konuşma iyileştirme algoritmasından hiçbirinin anlaşılabilirliği geliştiremediği öne sürülmüştür.

Gürültü azaltma algoritmalarının anlaşılabilirlik değerlendirmesi üzerine yapılan çalışmada (Hu and Loizou, 2007); IEEE veri tabanındaki cümleler ve ünsüzler (consonants), 0 ve 15 dB olmak üzere iki farklı SNR seviyesinde mırıltı, araba, sokak ve tren gibi dört farklı gürültü tipi ile bozulmuştur. Elde edilen gürültülü sesler, izgesel çıkarım, alt-uzay, istatistiksel model tabanlı ve Wiener-

tipi algoritma olmak üzere dört sınıfta bulunan sekiz iyileştirme metodu ile işlenmiştir. Bu algoritmalar ile iyileştirilen konuşmalar, yöntemlerin performansını ölçmek amacıyla duyma bozukluğu olmayan dinleyiciler ile test edilmiştir. Tek gürültünün kullanıldığı durumlarda hiçbir algoritma konuşma anlaşılabilirliğini geliştirmede önemli bir katkı sağlayamamıştır. Sonuçlara göre hem cümle hem de ünsüz ses tanımda önsel SNR tabanlı Wiener (Wiener-önsel SNR) algoritması en iyi iyileştirme yöntemi olarak bulunmuştur. Wiener algoritmasını Karhunen-Loeve Dönüşümü (KLD), çoklu bant izgesel çıkarım ve Logaritmik en küçük ortalama kareler kestiricisi (LogEOKK) algoritmaları takip etmektedir. Algısal KLD algoritması gürültü azaltmada iyi sonuçlar verirken anlaşılabilirliği olumsuz yönde etkilemektedir. Ünsüz karışım (confusion) matrisi için algoritmaların çoğu düşük SNR'da düşük enerjili ünsüzlerde yeterli iyileştirme sağlayamamaktadır. 0 dB mırıltı gürültüsünün kullanıldığı durumlarda ise alt-uzay algoritması iyi sonuçlar verirken diğer durumlarda başarılı olamamıştır. 0 dB tren gürültüsünün eklenmesi durumunda ise hiçbir algoritma gürültü azaltmada iyi sonuçlar verememiştir. Tek kanal mikrofon iyileştirme algoritmalarını içeren bu çalışma normal dinleyiciler ve işitme bozukluğu olan dinleyicilerde anlaşılabilirliği geliştirmemektedir. Bu yüzden koklear implantlı kişilerde bu algoritmaların kullanılması önerilmemiştir. Loizou et. al. (2005) çalışmasında KLD algoritmasının duyma bozukluğu olan kişilerde kullanılabileceğini öne sürmüştür.

Li et al. (2011) çalışmalarında; Çince, Japonca ve İngilizce olmak üzere üç farklı dilde oluşturulan gürültülü konuşma sinyallerine tek kanallı çeşitli gürültü azaltma algoritmaları uygulamışlardır. Temiz konuşma sinyalleri (Çince ve Japonca kelimeler) iki farklı SNR seviyelerinde (0 ve 5 dB) üç çeşit gürültü tipi (beyaz Gauss, mırıltı ve araba) ile bozulmuştur. Bozulan bu sesler beş farklı gürültü azaltma algoritması (KLD, Wiener-önsel SNR, LogEOKK, LogEOKK-SVB ve ÇBİÇ) ile işlenmiştir. İşlenmiş konuşma sinyalleri tanıma testi için işitme bozukluğu olmayan dinleyicilere dinlettirilmiştir. Bu çalışmada yapılan anlaşılabilirlik değerlendirmesine göre, gürültü azaltma algoritmalarının çoğunun anlaşılabilirliği geliştirmediğini (sağlamadığını) göstermiştir. Çince (Mandarin) için anlaşılabilirlik değerlendirmesinde, fonetik dengesi olan ve her bir tabloda 75 adet ünsüz-ünlü (consonant-vowel) yapısında heceler olmak üzere toplam 10 farklı tablo seçilmiştir. Her bir tablodan rastgele üç hece olmak üzere toplam 25 anlamlı ya da anlamsız cümle oluşturulmuştur. Testten önce test prosedürü ile ilgili her dinleyiciye birkaç cümle dinlettirilmiştir. Test sürecinde ise dinleyicilerden duydukları kelimeleri yazmaları istenmiştir. Dinleme testinde her

dinleyici 900 (25 cümle x 36 durum: 2 SNR seviyesi x 3 gürültü tipi x 6 algoritma) cümleyi üç oturumda dinlemiştir. Her bir oturumda dinleyicilere verilen cümleler ve koşullar her bir denek için rastgele yapılmıştır. Çalışmanın sonuçlarına göre mırıltı gürültü tipi kullanılan durumlarda tüm dillerde anlaşılabilirlik oranı düşerken araba gürültü durumunda ise anlaşılabilirlik oranı daha yüksek bulunmaktadır. Ayrıca sessiz seslerin (ünsüzler) ünlü seslere göre gürültüde maskelendiğinden anlaşılabilirlik oranı düşmektedir. Çince dili için anlaşılabilirlik skoruna bakıldığında Wiener-önsel SNR algoritması diğer algoritmalara göre daha iyi sonuçlar vermektedir. KLD algoritması ise İngilizce’de anlaşılabilirliği artırırken Çince ve Japonca’da iyi bir performans gösterememektedir. LogEOKK-SVB algoritması bütün dillerde tüm durumlar için kötü sonuçlar vermektedir. Ayrıca tüm diller için anlaşılabilirlik skoru bakımından en iyi sonuçları Wiener-önsel SNR algoritması vermektedir.

Konuşma iyileştirme algoritmalarının konuşma anlaşılabilirliğini sağlayamamasının birkaç nedeni vardır. Birinci neden, çoğu algoritmanın uygulanması için gerekli olan arka plan gürültü güç izgesinin yeterince iyi kestirilememesidir. İyi bir gürültü kestirimi için doğru bir Ses Etkinlik Algılama (Voice Activity Detection - VAD) algoritması gereklidir. Gürültü kestirim algoritmalarının araba gürültüsü gibi durağan arka plan gürültülerinde iyi bir performans gösterdikleri söylenebilir. Hu ve Loizou (2007) tarafından yapılan çalışmada araba gürültüsünün bulunduğu ortamda yapılan konuşma iyileştirmede anlaşılabilirlik oranı mırıltı gürültü ortamında yapılabildiğine göre daha iyi bulunmuştur. Bu durum durağan gürültülerin olduğu seslerde gürültü kestiriminin daha doğru yapıldığını göstermektedir ve kestirimin doğru yapılması durumunda anlaşılabilirliğin sağlanmasına katkıda bulunabileceğini göstermektedir.

Dalgacık dönüşümü tabanlı önerilen yaklaşımlarda (Donoho and Johnstone, 1994; Donoho, 1995) gürültülü konuşma sinyaline dalgacık ayrıştırması uygulanmakta ve konuşma sinyalinin iyileştirilmiş versiyonunun yeniden yapılandırılmasından önce en az bozulmuş bileşenleri bir karar kriteri olan eşik değerleri ile belirlenmektedir. Tüm dalgacık katsayılarına sabit bir eşikleme uygulamak, gürültünün yanında konuşma bileşenlerinin de yok edilmesine yol açacağından uyarlanır (adaptive) eşikleme yöntemleri tercih edilmektedir. Ancak, dalgacık dönüşümünün sabit temel dalgacık işlevi kullanması ve algısal dalgacık paket dönüşümü yöntemlerinde kullanılacak alt bantların sayısının değişkenlik göstermesi gibi bazı sakıncaları bulunmaktadır.

Geleneksel Ayırık Dalgacık Dönüşümü (ADD) tabanlı analizde, ölçek alanı ayrıştırılırken dalgacık alanı ayrıştırılmamaktadır. Gürültülü konuşmaya iyileştirme amacıyla ADD uygulandığında böyle bir analizin önemli bir eksikliği, bozulmamış bir konuşma kalitesine yol açan birçok ses-benzeri konuşma bileşenini içeren işlenmemiş konuşma çerçevelerinin büzülmesidir. ADD tabanlı analizden farklı olarak, dalgacık paket tabanlı analizde, dalgacık alan ayrıca ayrıştırılır ve böylece daha yüksek frekans bandı ayrışması mümkün olmaktadır. Hem yaklaşım hem de ayrıştırma dalgacık paket katsayıları, her ayrıştırma seviyesinde iki parçaya ayrıldığı için, üstün frekanslı lokalizasyona sahip tam bir ikili ağaç elde edilebilir. Ayırık Dalgacık Paket Dönüşümü (ADPD), sinyali daha fazla sayıda alt bantta parçalara ayırır ve yüksek frekanslarda standart dalgacık dönüşümünden daha yüksek frekans çözünürlüğüne sahip olabilen çok işlevli bir çerçeve oluşturur (Ghanbari and Karami-Mollaei, 2006; Xu et al., 2007; Messaoud et al., 2015)

Bu nedenle, son yıllarda yeni bir zamansal sinyal ayrıştırma yöntemi olan Ampirik Kip Ayrışımı (AKA) durağan olmayan ve doğrusal olmayan zaman serilerini analiz etmek için önerilmiştir. Bu yeni uyarlanabilir yöntem bir sinyali içkin kip fonksiyonları (Intrinsic Mode Functions-IMFs) olarak adlandırılan salınımlı bileşenleri halinde parçalara ayırmaktadır. AKA yönteminin dalgacık dönüşümüne göre en büyük üstünlüğü temel işlevlerinin sabit olmayıp sinyalden türetilmesidir. Bu nedenle analiz, temel işlevlerin sabitlendiği geleneksel yöntemlerin aksine uyarlanabiliridir.

Son yıllarda, konuşma iyileştirmede zaman-frekans alanı yöntemlerinden AKA sıklıkla kullanılmaktadır. Konuşma iyileştirmede gürültü azaltmak için AKA tabanlı süzgeçleme (Khaldi et al., 2008; Sharma and Mahadeva Prasanna, 2016) ya da eşikleme yaklaşımları (Hamid et al., 2013; Kemiha, 2014) önerilmiştir. Ancak AKA yönteminde, sabit bir eşik değeri belirlenerek gürültülü konuşmanın eşiklenmesi gürültünün yanında konuşmaya ait bilgilerin de kaybolmasına neden olabilmektedir. Bu nedenle gürültü atmada sabit eşik kullanmak yerine İKF katsayılarından uyarlanabilir eşik değerleri belirlenmesi iyileştirmede daha başarılı sonuçlar vermektedir (Chatlani and Soraghan, 2009; Khaldi et al., 2014b). Uyarlanabilir eşik değerlerinin belirlenmesi, dalgacık dönüşümü tabanlı yöntemlerde olduğu gibi Teager Enerji Operatörü kullanılarak belirlenebilmektedir (Sanam and Shahnaz, 2012, 2013; Islam et al., 2015).

2.2 Sesli Etkinlik Algılama Literatür Taraması

Konuşma ya da sesli etkinlik algılama (Speech/Voice Activity Detection - VAD); gürültülü konuşma sinyallerinde, konuşmanın etkin olduğu bölgelerin konuşma dışı bölgelerden ayırt edilmesini sağlayan bir yöntemdir. Konuşma dışı bölgeler; sessizlik (silence), gürültü ve diğer ilgisiz akustik bilgi parametrelerini içermektedir. Sesli etkinlik algılama (SEA), konuşma kodlama (Freeman and Cosier, 1989; Beritelli et al., 1998; Enqing et al., 2002; Zhang et al., 2005), konuşma sinyallerinin bölütlenmesi ve konuşma tanıma sistemleri (Karray and Martin, 2003), sayısal işitme cihazlarında gürültü azaltma, mobil iletişim ve konuşma iyileştirme (Woo et al., 2000) gibi birçok konuşma ve ses işleme uygulamalarında sıklıkla kullanılmaktadır.

Geleneksel SEA algoritmalarının birçoğu basit yapılarından dolayı enerji, perde periyodu (pitch) tespiti, sıfır-geçiş oranı (zero-crossing rate), doğrusal öngörü katsayıları ve kepsstral özellikler gibi ölçütlerin eşiklenmesine dayanmaktadır (Greenwood and Kinghorn, 1999; Harsha, 2004; Babu and Vanathi, 2009; Bachu et al., 2010). Ancak, bu ölçütler ile yapılan SEA uygulamalarında, düşük sinyal gürültü oranlarında ya da durağan olmayan arka plan gürültülerinde konuşma bölgelerinin tespiti zorlaşmaktadır.

Son yıllarda, çok sayıda araştırmacı, sesli etkinlik algılamanın konuşma ve ses işleme uygulamalarının performansı üzerindeki etkisini belirlemek için farklı yaklaşımlar araştırmışlardır. Bu yaklaşımlardan istatistiksel modellere dayanan olasılık oranı testini (Likelihood Ratio Test) kullanan gürbüz (robust) SEA algoritmaları konuşma bölgelerinin belirlenmesinde, geleneksel yöntemlere göre daha iyi bir performans göstermektedir (Sohn, 1999; Chang and Shin, 2003). Bir diğer yaklaşımda, Benyassine et al. (1997) çizgisel izgesel frekanslar (Line Spectral Frequencies - LSF), düşük bant enerjisi, sıfır geçiş hızı ve tam-bant enerjisini içeren bir dizi ölçütü kullanarak oluşturulan G.729B algoritmasını önermişlerdir. G.729B algoritmasını geliştirmek amacıyla, bu algoritma için kullanılan akustik özellikler seri olarak birleştirilmiş ve destek vektör makinesinin SEA'ya uygulanması önerilmiştir (Enqing et al., 2002). Böylece önerilen algoritmanın G.729B SEA üzerinde dikkate değer bir gelişme elde ettiği görülmüştür.

Yukarıdaki SEA özelliklerinin tümü kısa süreli analiz çerçevelerinden çıkarılır ve SEA kararı bu kısa süreli çerçeveler üzerinden yapılır. SEA için kısa

sürekli çerçevelerin kullanılmasının aksine, uzun süreli izgesel ayrılma (LTSD) algoritması önerilmiş ve uzun süreli izgesel zarfın ortalama gürültü izgesi ile karşılaştırılmasıyla konuşma ve konuşma-dışı karar kuralı formülize edilmiştir (Ramírez et al., 2004). Bir diğer uzun süreli yöntem olan uzun süreli izgesel düzlük ölçümünde ise düşük varyans izge tahmini ve uyarlanır eşik değerleri kullanılmıştır (Ma and Nishihara, 2013).

Temiz konuşma sinyallerine ek olarak gürültülü konuşma sinyallerinin ses veya konuşma etkinlik bölgelerini farklı gürültülü koşullar altında tespit etmek için çeşitli yöntemler geliştirilmiştir. Bu yöntemler, zaman alanı, frekans alanı ve zaman-frekans alanı olmak üzere üç temel gruba ayrılmaktadır. Zaman alanı yöntemleri arasında enerji, sıfır geçiş hızı ve düşük frekanslı bant gücü bulunmaktadır. Frekans alanı yöntemleri olarak izgesel tabanlı enerji-entropi özelliği (Khoa, 2012; Zhang et al., 2014), kepsral özellikler (Haigh and Mason, 1993; Chung and Oh, 2015), Teager enerji özelliği (Stegmann and Schroder, 1997; Chen et al., 2007), ses sinyallerinin uyarlamalı modellemesi gibi çeşitli izgesel enerji tabanlı özellikler (Aibinu et al., 2012) ve periyodiklik ölçütü (Tucker, 1992; Li et al., 2006) bulunmaktadır. Ayrıca, zaman-frekans alanı yöntemleri ise dalgacık dönüşümü (Chen et al., 2007, 2010; Eshaghi and Karami Mollaei, 2010) ve Ampirik Kip Ayırıştırması (Kanai et al., 2013; Chong Feng, 2014) gibi yöntemleri içermektedir.

Gürültülü koşullar altında, enerji, sıfır-geçiş oranı ve düşük frekans bant gücü gibi zaman alanı, izge ve periyodiklik/a-periyodiklik gibi frekans alanı ve dalgacık dönüşümü gibi zaman-frekans alanı yöntemleri konuşma/ses bölgelerinin tespitinde yetersiz kalmaktadır. Bu yöntemlerin birçoğunda, tüm gürültü tipi ve tüm SNR seviyelerinde eşik değerleri tam ve doğru bir şekilde hesaplanamamaktadır. Bu nedenle, son yıllarda, konuşma bölgelerinin belirlenmesinde konuşma sinyallerini yüksek frekanslardan düşük frekanslara doğru içkin (intrinsic) kiplere ayırtıran Ampirik Kip Ayırıştırması yöntemi kullanılmaktadır.

3. LİTERATÜRDE KULLANILAN KONUŞMA İYİLEŞTİRME YÖNTEMLERİNİN GÖZDEN GEÇİRİLMESİ

Konuşma iyileştirme yöntemlerini sınıflandırmak için birçok yol mevcuttur. Tipik bir algoritma için tüm gürültü türlerinde homojen bir performans göstermek genellikle zor olmaktadır. Bu nedenle konuşma iyileştirme sistemleri genellikle uygulama ve çevresine bağlı olan belli varsayım ve kısıtlamalara dayanmaktadır. Genel olarak, bir konuşma iyileştirme algoritmasının performansı mevcut gürültü kaynaklarının sayısı ve bozulan sinyal ile ilgili ön bilginin kullanılması gibi etkenlerle sınırlanmaktadır. Konuşma iyileştirme algoritmaları literatürde genellikle zaman alanı, frekans alanı ve zaman-frekans alanı yöntemleri adı altında çeşitli konuşma işleme stratejilerinden oluşmaktadırlar. Zaman alanı yöntemleri olarak alt-uzay tabanlı algoritmalar, frekans alanı yöntemleri olarak temel izgesel çıkarım, çoklu bant izgesel çıkarım, Wiener süzgeçleme, en küçük ortalama kareler kestiricisi (EOKK), Logaritmik EOKK (LogEOKK), sinyal varlık belirsizliği altında logaritmik EOKK (LogEOKK-SVB) kestiriciler ve zaman-frekans alanı yöntemleri olarak da Ayırık Dalgacık Dönüşümü, Ayırık Dalgacık Paket Dönüşümü, Algısal Dalgacık Paket Dönüşümü ve Ampirik Kip Ayırıştırması yöntemleri kullanılmaktadır.

3.1 Zaman Alanı Yöntemleri

3.1.1 Alt-Uzay Tabanlı Algoritmalar

Karhunen-Loeve Dönüşümü

Alt uzay tabanlı konuşma iyileştirme tekniklerinde gürültülü sinyal M-boyutlu sinyal ve gürültü alt uzaylarına ayrıştırılmaktadır. Gürültü tüm uzayı kapsarken konuşma sinyalinin yalnızca sinyal alt uzayı içerisinde olduğu varsayılmaktadır. Sinyal alt uzay bilgisi orijinal konuşma sinyalinin kestiriminde kullanılmaktadır. Alt uzay yaklaşımı izge kestirimi; sistem tanımlama, görüntü işleme, dizi işleme, ses tanıma ve iyileştirme uygulamalarında sıklıkla kullanılmaktadır. Konuşma iyileştirme uygulamalarında genellikle Karhunen-Loeve Dönüşümü (KLD) tabanlı algoritmalar kullanılmaktadır (Rezayee and Gazor, 2001; Hu and Loizou, 2007). KLD yaklaşımında; sinyalin kovaryans matrisi öz vektörler matrisi olarak ifade edilmektedir (Quatieri, 2001).

$$R_y = R_x + R_d \quad (3.1)$$

R_d gürültü sinyalinin, R_x temiz sinyalin, R_y ise gürültülü sinyalin kovaryans matrisini göstermektedir.

Sinyalin kovaryans matrisinin öz vektörlerini doğru bir şekilde kestirebilmek için KLD tabanlı alt uzay yöntemi kritik öneme sahiptir. En basit yaklaşım ile kestirilen temiz sinyalin kovaryans matrisi:

$$\hat{R}_x = R_y - \hat{R}_d \quad (3.2)$$

ile ifade edilmektedir.

Alt uzay tabanlı konuşma iyileştirme aşağıda birbirini izleyen altı adımda yapılmaktadır:

Adım 1. Gürültülü sinyalin kovaryans matrisi R_y hesaplanır ve $\sum R_d^{-1} R_y - I$ matrisi kestirilir. Gürültü kovaryans matrisi R_d , ses olmayan bölütler boyunca gürültü örnekleri kullanılarak hesaplanır.

Adım 2. $\sum$ 'nin öz-ayrıştırması (eigen-decomposition) yapılır.

$$\sum V = V \Lambda_\Sigma \quad (3.3)$$

Burada V öz-vektörleri, Λ_Σ ise öz-değerler ile oluşturulan köşegen matrisi temsil etmektedir.

Adım 3. Konuşma sinyalinin alt uzayının boyutu kestirilir. λ_Σ öz-değerler olmak üzere;

$$M = \arg \max\{\lambda_\Sigma(k) > 0\} \quad (3.4)$$

Adım 4. μ Lagrange çarpanı değeri hesaplanır:

$$\mu = \begin{cases} \mu_0 - (\text{SNR}_{dB})/s & , \quad -5 \leq \text{SNR}_{dB} \leq 20 \\ 1 & , \quad \text{SNR}_{dB} \geq 20 \\ 5 & , \quad \text{SNR}_{dB} \leq -5 \end{cases} \quad (3.5)$$

Burada $\mu_0 = 4,2$ ve $s = 6,25$ alınmaktadır.

Adım 5. Bir optimal doğrusal kestirici hesaplanır:

$$g(k) = \begin{cases} \frac{\lambda_{\Sigma}(k)}{\lambda_{\Sigma}(k) + \mu} & k = 1, 2, \dots, M \\ 0 & k = M + 1, \dots, K \end{cases} \quad (3.6)$$

$$G_1 = \text{diag}\{g(1), \dots, g(M)\} \quad (3.7)$$

$$H_{\text{opt}} = R_d V \begin{bmatrix} G_1 & 0 \\ 0 & 0 \end{bmatrix} V^T = V^{-T} \begin{bmatrix} G_1 & 0 \\ 0 & 0 \end{bmatrix} V^T \quad (3.8)$$

Adım 6. İyileştirilmiş konuşma sinyali elde edilir:

$$\hat{x} = H_{\text{opt}} \cdot y \quad (3.9)$$

Algısal Karhunen-Loeve Dönüşümü

Konuşma iyileştirmede algısal tabanlı alt uzay yönteminde artık gürültüyü önlemek için maskeleme eşiğin (masking threshold) altında kalan artık gürültü izgesi kestirilir (Hu and Loizou, 2007). Bu yaklaşım ile yapılan KLD yönteminde, ilk olarak temiz sinyal kovaryans matrisinin eigen-ayrıştırması yapılmakta ve maskeleme eşiği, eigen-frekans alanı dönüşümünden elde edilen izgeden kestirilmektedir. Son olarak maskeleme eşikleri, frekans-öz (eigen) alan dönüşümü kullanılarak öz alanında ifade edilir ve daha sonra sinyal alt uzay yaklaşımı ile birleştirilir.

3.2 Frekans Alanı Yöntemleri

3.2.1 Temel İzgesel Çıkarım

İzgesel çıkarım algoritması tarihsel olarak gürültü azaltma için önerilen ve basit bir prensibe dayanan ilk algoritmalarından biridir. Gürültünün toplamsal olduğu göz önünde bulundurularak, gürültülü konuşma izgesinden gürültü izgesinin kestirimi ile temiz konuşma izgesinin kestirimi elde edilebilir. Gürültü izgesi, sinyalin olmadığı periyotlar boyunca kestirilebilir ve güncellenebilir (Lu and Loizou, 2008). Basit bir algoritma olan bu yöntemde “müzikal gürültü” olarak bilinen rahatsız edici bir ses, konuşma sinyalini bozmaktadır. Birçok algoritma, bu istenmeyen gürültünün giderilmesini amaçlamıştır (Petrovsky et al., 2004; Lu and

Loizou, 2008). İzgesel çıkarım algoritmasında, gürültünün durağan ya da yavaş değiştiği kabul edilerek gürültü izgesinin güncellenen periyotlar arasında önemli bir değişikliğe uğramadığı düşünülür. Bu metotta iyileştirilmiş sinyal, gürültülü sinyalin fazı kullanılarak, kestirilen sinyalin izgesi Ters Ayrık Fourier dönüşümü ile elde edilmektedir. Algoritma yalnızca ileri ve ters Fourier dönüşümü kullanmaktadır ve hesaplanabilirlik açısından oldukça basittir. Bu yöntemin en büyük sakıncalarından biri, fazla izge çıkarımında konuşmanın içinde bulunan bazı bilgilerin kaybolması veya az çıkarım yapıldığında gürültünün yeterince atılmamış olmasıdır (Boll, 1979; Kamath and Loizou, 2002).

İzgesel Çıkarım Yöntemi İlkesi

Temiz konuşma sinyaline istatistiksel olarak bağımsız bir gürültü eklendiği düşünüldüğünde gürültülü konuşma sinyali şu şekilde ifade edilebilir (Quatieri, 2001):

$$y[n] = s[n] + d[n] \quad (3.10)$$

Burada $y[n]$, $s[n]$ ve $d[n]$ sırasıyla gürültülü konuşma, temiz konuşma ve eklenmiş gürültüyü temsil etmektedir. Eklenmiş gürültünün sıfır ortalamalı ve temiz konuşma sinyali ile ilintisiz olduğu varsayılmıştır. Konuşma sinyali durağan olmayan ve zamanla değişen olduğundan, gürültülü konuşma sinyali genellikle bölüt olarak işlenmektedir. Gürültülü konuşma sinyalinin Kısa Zamanlı Fourier Dönüşümü (KZFD) şu şekilde verilmektedir (k bölüt sayısını göstermektedir):

$$Y(w, k) = S(w, k) + D(w, k) \quad (3.11)$$

Konuşma sinyalinin arka plan gürültüsü ile ilintisiz olduğu kabul edildiğinde, gürültülü konuşma sinyalinin kısa zamanlı güç izgesinde çapraz terimler bulunmamalıdır. Bu nedenle gürültülü sinyalin güç izgesi aşağıdaki gibi ifade edilebilir:

$$|Y(w)|^2 = |S(w)|^2 + |D(w)|^2 \quad (3.12)$$

Temiz konuşma sinyali, gürültülü konuşma sinyalinden bir gürültü sinyali kestirimi ile elde edilebilir.

$$|\hat{S}(w)|^2 = |\hat{Y}(w)|^2 - |\hat{D}(w)|^2 \quad (3.13)$$

Gürültü izgesinin kestirimi $|\widehat{D}(w)|^2$, konuşmanın olmadığı bölütlerin ortalaması alınarak elde edilir:

$$|\widehat{D}(w)|^2 = \frac{1}{M} \sum_{j=0}^{M-1} |Y_{SP_i}(w)|^2 \quad (3.14)$$

Burada M konuşmanın olmadığı (SP-Speech Pauses) bölüt sayısını göstermektedir. Arka plan gürültüsünün durağan olması durumunda daha uzun bir ortalama alınarak optimum gürültü güç izgesi tahmini yapılabilir.

İzgesel çıkarım, bir süzgeç olarak düşünülebilir ve gürültülü konuşma sinyali ile izgesel çıkarım süzgecinin (Spectral Subtraction Filter-SSF) çarpımı olarak ifade edilebilir:

$$|\widehat{S}(w)|^2 = \left(1 - \frac{|\widehat{D}(w)|^2}{|\widehat{Y}(w)|^2}\right) |\widehat{Y}(w)|^2 = H^2(w) |\widehat{Y}(w)|^2 \quad (3.15)$$

Burada $H(w)$ izgesel çıkarım süzgeci olarak bilinen bir kazanç işlevidir. $H(w)$ sıfır fazlı bir süzgeç olup, genlik cevabı 0 ile 1 arasında değişmektedir.

Temiz konuşma sinyalini geri elde etmek için konuşma sinyalinin faz kestirimine ihtiyaç duyulmaktadır. İnsan kulağının algılamasında faz nispeten önemsiz olduğundan temiz konuşma sinyalinin fazının, gürültülü konuşma sinyalinin fazına eşit olduğu kabul edilmektedir. Daha sonra konuşma sinyali bölüt bazında şu şekilde hesaplanmaktadır:

$$\widehat{S}(w) = |\widehat{S}(w)| e^{j\angle Y(w)} = H(w) Y(w) \quad (3.16)$$

Temiz konuşma sinyali ekleme-örtüşme yöntemi kullanılarak ters Fourier dönüşümü ile zaman alanında elde edilebilir.

3.2.2 Çoklu Bant İzgesel Çıkarım

Çoklu bant yaklaşımında konuşma sinyalinin izgesi birbiri ile örtüşmeyen N sayıda banda bölünmektedir ve izgesel çıkarım birbirinden bağımsız her bir bantta gerçekleştirilmektedir. Konuşma sinyalinin farklı bantlara bölünmesi işlemi zaman alanında bant geçiren süzgeçlerin kullanılması ya da frekans alanında uygun pencerelerin kullanılması ile gerçekleştirilmektedir. İkinci yöntem

hesaplanabilirlik olarak daha ekonomik olduğundan tercih edilmektedir. Temiz konuşma sinyalinin i . bandının genlik izgesini kestirme işlemi şu şekilde yapılmaktadır (Quatieri, 2001):

$$|\hat{S}_i(w)|^2 = \begin{cases} |\hat{Y}_i(w)|^2 - \alpha_i \delta_i |\hat{D}_i(w)|^2, & \text{eğer } |\hat{S}_i(w)|^2 > 0 \\ \beta |\hat{Y}_i(w)|^2, & \text{diğer} \end{cases} \quad (3.17)$$

Burada α_i i . bandın aşırı-çıkarım faktörünü δ_i ise her bantta gürültü çıkarma seviyesini kontrol altında tutmak için ayrı ayarlanabilen ek bant çıkarma faktörünü göstermektedir.

3.2.3 En Küçük Ortalama Kareler Kestiricisi

İstatistiksel model tabanlı konuşma iyileştirme algoritmalarında, çeşitli istatistiksel modeller ve optimizasyon kriterleri kullanılarak sinyalin karmaşık izgesinden (Wiener süzgeç) ziyade Ayırık Fourier katsayılarından genliğin doğrusal olmayan kestirimi üzerinde durulmaktadır. Bu doğrusal olmayan kestiriciler, konuşma sinyalinin Ayırık Fourier katsayılarını ve gürültü sinyalinin olasılık yoğunluk fonksiyonlarını kullanmaktadır. Literatürde, doğrusal olmayan kestiriciler ve Bayesian kestiriciler olmak üzere çeşitli konuşma iyileştirme teknikleri bulunmaktadır (Martin, 2002; Hooda and Aggarwal, 2012).

Birçok araştırmada, gerçek izge büyüklüğü ile kestirilen izge büyüklüğü arasındaki ortalama karesel hatayı en aza indirmeye dayalı yöntemler önerilmiştir.

$$e_k = E \{ (\hat{X}_k - X_k)^2 \} \quad (3.18)$$

Burada $\hat{X}_k$ k . frekans bölmesi için kestirilen büyüklüğü, X_k ise temiz sinyal için k . bölmenin büyüklüğünü göstermektedir. Optimal katsayılar (Ortalama Karesel Hata (OKH) anlamında) Bayesian OKH hatasının en aza düşürülmesi ile elde edilebilir;

$$B_{OKH}(\hat{X}_k) = \iint (\hat{X}_k - X_k)^2 p(Y, X_k) dY dX_k \quad (3.19)$$

$\hat{X}_k$ 'ya göre Bayesian OKH'nın en aza indirilmesi optimum En Küçük Ortalama Kareler Kestirici'yi (EOKK) vermektedir:

$$\hat{X}_k = E[X_k | Y(w_0) Y(w_1) \dots Y(w_{N-1})] \quad (3.20)$$

Burada $Y = [Y(w_0) Y(w_1) \dots Y(w_{N-1})]$ gürültülü konuşma sinyalinin Ayırık Fourier Dönüşümü katsayılarını içeren vektörü temsil etmektedir.

3.2.4 Logaritmik En Küçük Ortalama Kareler Kestiricisi

Optimal en küçük ortalama kareler kestiricilerde izgesel büyüklüklerde hatanın en aza indirilmesi hesaplanabilirlik olarak uygun olabilir, ancak öznel olarak anlamlı olmayabilir. Konuşma sinyali işleme için, logaritmik bir büyüklük olan karesel hataya dayalı bir ölçütün uygun olduğu ileri sürülmüştür (Hu and Loizou, 2006). Bu yöntemde, logaritmik izgelerin büyüklüğünün karesel hatasını en aza indirgeyen bir kestirici kullanılmaktadır (Quatieri, 2001):

$$E \left\{ (\log \hat{X}_k - \log X_k)^2 \right\} \quad (3.21)$$

Optimum logaritmik en küçük ortalama kareler kestiricisi (LogEOKK) $\log X_k$ 'nin koşullu ortalamasının değerlendirilmesi ile elde edilebilir:

$$\begin{aligned} \log X_k &= E\{\log X_k | Y(w_k)\} \\ X_k &= \exp(E\{\log X_k | Y(w_k)\}) \end{aligned} \quad (3.22)$$

3.2.5 Sinyal Varlık Belirsizliğine Dayalı Logaritmik En Küçük Ortalama Kareler Kestiricisi

Konuşmanın etkinliği için iki durumlu model; matematiksel olarak ikili hipotez modeli kullanılarak ifade edilebilir:

$$\begin{aligned} H_0^k: & \text{konuşma yok} & |Y(w_k)| &= |D(w_k)| \\ H_1^k: & \text{konuşma var} & |Y(w_k)| &= |X(w_k)| + |D(w_k)| \end{aligned} \quad (3.23)$$

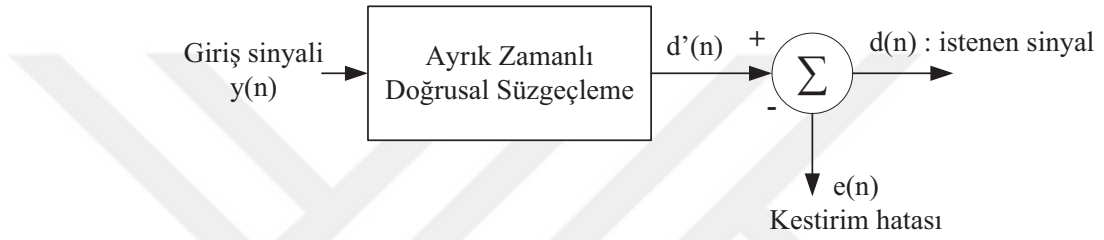
Burada H_0^k , k. frekans bölgesinde konuşmanın olmadığı geçersiz hipotezi, H_1^k ise konuşmanın olduğu hipotezi göstermektedir.

Bu yöntemde, sinyal-varlık belirsizliğini (signal-presence uncertainty-SPU) dikkate alan LogEOKK kestiricileri kullanılmaktadır. Sinyal-varlık belirsizliğini (SVB) hesaba katan yeni EOKK şu şekilde ifade edilebilir:

$$\begin{aligned}
X_k &= E(X_k | Y(w_k), H_1^k) P(H_1^k | Y(w_k)) \\
\log \hat{X}_k &= E[(\log X_k | Y(w_k), H_1^k)] P(H_1^k | Y(w_k))
\end{aligned}
\tag{3.24}$$

3.2.6 Wiener Süzgeçleme

Şekil 3.1’de doğrusal ve zamanla değişmeyen istatistiksel süzgeçleme blok diyagramı verilmektedir. Optimum süzgeç, Wiener süzgeç olarak adlandırılan ve kestirim hatasını en aza indirgeyen bir sistemden oluşmaktadır.



Şekil 3.1 Optimum süzgeçleme

Konuşma iyileştirme uygulamalarında, Şekil 3.1’de giriş sinyali $y(n)$ gürültülü konuşma sinyalini göstermektedir:

$$y(n) = x(n) + n(n) \tag{3.25}$$

Burada $x(n)$ temiz konuşma sinyalini ve $n(n)$ ise gürültü sinyalini temsil etmektedir. İstenen sinyal $d(n)$, gürültüden arındırılmış temiz $x(n)$ sinyalini vermektedir. Wiener süzgecinin kullanım amacı temiz sinyalin üretilmesine olanak sağlamaktır. Wiener süzgeç zaman ya da frekans alanında gürültü atmada kullanılabilmektedir. Zaman alanında öncelikle öz-ilişki matrisi R_{yy} hesaplanır:

$$R_{yy} = E[yy^T] = E[(x + n)(x + n)^T] = R_{xx} + R_{nn} \tag{3.26}$$

Konuşma sinyali ve gürültü ilintisiz olduğundan çapraz-ilişki vektörü r_{yd}^- , temiz sinyalin öz-ilişki işlevi r_{xx} ’e eşit olacaktır. Bu nedenle zaman alanında Wiener süzgecin çıkışı şu şekilde ifade edilir:

$$h^* = (R_{xx} + R_{nn})^{-1} r_{xx} \tag{3.27}$$

Frekans alanında Wiener süzgecin kullanılması durumunda gürültülü sinyalin Fourier dönüşümü alınarak süzgeç şu şekilde ifade edilir:

$$Y(w_k) = X(w_k) + N(w_k) \quad (3.28)$$

Burada öncelikle istenen ve gürültülü sinyalin çapraz-güç izgesi $P_{dy}(w_k)$ ve gürültülü konuşma sinyalinin güç izgesi $P_{yy}(w_k)$ hesaplanır. $D(w_k) = X(w_k)$ olduğuna göre yeni eşitlik şu şekilde ifade edilebilir:

$$P_{dy}(w_k) = E[X(w_k)\{X(w_k) + N(w_k)\}^*] = P_{xx}(w_k) \quad (3.29)$$

Sonuç olarak frekans alanında Wiener süzgeç:

$$H(w_k) = \frac{P_{xx}(w_k)}{P_{xx}(w_k) + P_{nn}(w_k)} \quad (3.30)$$

Önsel (priori) SNR'ın ξ_k ile ifade edilmesi durumunda Wiener süzgeç şu şekilde ifade edilebilir:

$$\xi_k = \frac{P_{xx}(w_k)}{P_{nn}(w_k)} \quad H(w_k) = \frac{\xi_k}{\xi_k + 1} \quad (3.31)$$

Düşük SNR bölgelerinde $\xi_k \rightarrow 0$ olduğundan $H(w_k) \approx 0$ olmaktadır. Yüksek SNR bölgelerinde ise $\xi_k \rightarrow \infty$ olduğundan $H(w_k) \approx 1$ olmaktadır.

4. NESNEL KALİTE VE ANLAŞILABİLİRLİK ÖLÇÜTLERİ

Farklı gürültüye sahip ortamlarda yapılan konuşma iyileştirme algoritmalarının performansları nesnel kalite ve anlaşılabilirlik ölçütleri ile değerlendirilmektedir. Algoritma başarılarının ölçülmesinde sıklıkla kullanılan ölçütler şu şekilde verilmektedir.

4.1 Nesnel Kalite Ölçütleri

Nesnel konuşma kalite ölçütleri, bazı matematiksel formülleri kullanarak genellikle gürültülü konuşma sinyali ve orijinal (temiz) konuşma sinyali arasında hesaplanır. Bu ölçütler herhangi bir dinleyici gerektirmediğinden pahalı olmayıp, zaman alıcı değildir. Nesnel kalite ölçütleri genellikle kalitenin kabaca hesaplanmasında kullanılır.

4.1.1 Sinyal Gürültü Oranı (SNR) ve Bölütsel SNR

Sinyal Gürültü Oranı (Signal-to-Noise Ratio-SNR) en eski ve yaygın olarak kullanılan nesnel ölçütlerden biridir. Hesaplanması matematiksel olarak basit olup, hem gürültülü hem de temiz konuşma örnekleri gerektirir. Konuşma sinyali durağan bir sinyal olmadığından klasik SNR'ın konuşma kalitesi ile iyi bir ilinti içinde olmadığı gözlenmiştir.

Konuşma sinyalinin enerjisi zamanla değişmektedir ve bu zaman dilimlerinin çoğunda sinyalin enerjisi büyük olup, gürültü nispeten duyulamaz haldedir. Bu nedenle SNR kısa zaman dilimlerinde hesaplanır ve sonra ortalaması alınır. Bu şekilde elde edilen ölçüt bölütsel SNR olarak adlandırılır. Bölütsel SNR ölçütü, konuşma kodlama ve iyileştirme uygulamalarında ses kalitesinin değerlendirilmesinde sıklıkla kullanılmaktadır (Grancharov and Kleijn, 2008; Quackenbush et. al., 1988). Daha önce yapılmış çalışmalar (Ma et al., 2009; Hu and Loizou, 2008); zaman alanında SNR nesnel kalite ölçütünün, konuşmanın anlaşılabilirliği ya da kalitesi ile yüksek ilintiye sahip olmadığını göstermiştir. Buna karşın, bölütsel SNR ölçütünün frekans alan versiyonu, konuşmanın hem kalitesi hem de anlaşılabilirliği ile yüksek bir ilintiye sahiptir (Tribolet et al., 1978). Bölütsel SNR (BSNR) şu şekilde hesaplanmaktadır:

$$\text{BSNR} = \frac{10}{M} \sum_{j=0}^{M-1} \log_{10} \left[\sum_n^L \frac{s^2(n)}{[s(n) - \hat{s}(n)]^2} \right] \quad (4.1)$$

Burada $s(n)$ temiz konuşma sinyalini, $\hat{s}(n)$ iyileştirilmiş konuşma sinyalini, L çerçeve uzunluğunu ve M ise toplam çerçeve sayısını temsil etmektedir. Çerçeve uzunluğu normal şartlarda 15-20 ms arasında seçilmektedir.

4.1.2 Ağırlıklı İzgesel Eğim

Ağırlıklı İzgesel Eğim (Weighted Spectral Slope-WSS) ölçütü, her bir frekans bandında izgesel eğimler arasındaki ağırlıklı farkı hesaplamaktadır. İzgesel eğim, desibel olarak bitişik izgesel büyüklükler arasındaki fark olarak elde edilmektedir. Ağırlıklı İzgesel Eğim (AİE) ölçütü şu şekilde tanımlanmaktadır:

$$d_{\text{AİE}} = \frac{1}{M} \sum_{m=0}^{M-1} \frac{\sum_{j=1}^K W(j, m) (S_c(j, m) - S_p(j, m))^2}{\sum_{j=1}^K W(j, m)} \quad (4.2)$$

$W(j, m)$ j . frekans bandındaki ağırlığı, K bant sayısını, M sinyaldeki çerçevelerin toplam sayısını, $S_c(j, m)$ ve $S_p(j, m)$ sırası ile temiz ve iyileştirilmiş konuşma sinyalinin m . çerçevesinin j . frekans bandındaki izgesel eğimi vermektedir. İyileştirilmiş konuşma orijinal konuşma sinyaline yaklaştıkça AİE değeri azalmaktadır.

4.2 Nesnel Anlaşılabilirlik Ölçütleri

4.2.1 Konuşma Kalitesinin Algısal Değerlendirmesi

Kullanılan anlaşılabilirlik ölçütlerinden konuşma kalitesinin algısal değerlendirmesi (Perceptual Evaluation of Speech Quality-PESQ), iyileştirilmiş konuşma sinyalinden Fikir Değerlendirme Oranı-FDO (Mean Opinion Score-MOS) kestirimi için kullanılan uluslararası bir standarttır. Konuşma kalitesinin algısal değerlendirmesi (KKAD) ölçütü, 2001 yılında P.862 olarak Uluslararası Telekomünikasyon Birliği-Telekomünikasyon Standardizasyon Sektörü (ITU-T) tarafından resmi olarak standartlaştırılmıştır.

Fikir değerlendirme (Opinion Rating), konuşma örneklerinin genel olarak algılanan kalitesini değerlendirmek için kullanılmaktadır. Dinleyiciler Çizelge 4.1'de gösterildiği gibi test altındaki konuşma örneklerini beş kalite kategorisinde

derecelendirmektedirler. Ancak denekler üzerinde bu testlerin yapılması maliyetli olup oldukça zaman almaktadır. Ayrıca testlerin defalarca yapılmasının gerekmesi deneklerde ezberlemeye sebep olabilmektedir. Bunun engellenmesi içinde fonetik olarak dengeli çok sayıda cümle listesi oluşturulmalıdır.

Çizelge 4.1 Ses kalite kategorisi ve FDO için beş maddelik ölçek

Oran	Ses Kalite Kategorisi	Bozulma
5	Çok İyi	Algılanamaz
4	İyi	Sadece algılanabilir fakat rahatsız edici
3	Normal	Algılanabilir fakat biraz rahatsız edici
2	Kötü	Rahatsız edici
1	Çok Kötü	Çık rahatsız edici

4.2.2 Kısa-Zamanlı Nesnel Anlaşılabilirlik

Kısa-zamanlı nesnel anlaşılabilirlik ölçütü (Short-Time Objective Intelligibility-STOI), gürültü ile bozulan konuşmaların bir konuşma iyileştirme yöntemi kullanıldıktan sonra öznel anlaşılabilirlik değerlendirmelerinin yapılmasını sağlayan korelasyon tabanlı bir yöntemdir (Taal et al., 2010).

Son yıllarda yapılmış birçok çalışmada, kısa-zamanlı nesnel anlaşılabilirlik (KZNA) skorlarının, gürültü azaltma algoritmaları ile iyileştirilen konuşmalardan elde edilen öznel anlaşılabilirlik değerleri ile çok yakın ve yüksek bir korelasyona sahip olduğu görülmektedir (Zao et al., 2014). Kısa-zamanlı nesnel anlaşılabilirliğin hesaplanmasında, konuşma sinyalinin temiz ve gürültülü versiyonları kısa zaman çerçevelerine ayrılır ve 15 üçte bir (1/3) oktav bandında gruplandırılır. Her çerçeve ve her bant için orta anlaşılabilirlik ölçütü $d_j(m)$, temiz ve gürültülü konuşma sinyallerinden elde edilen zamansal zarf (temporal envelop) vektörleri arasındaki korelasyon katsayısı olarak tanımlanır. Sonuç olarak, KZNA ölçeği ara değerlerin 15 üçte biri oktav bandı ve tüm konuşma çerçeveleri üzerinden ortalaması alınarak verilir.

KZNA ölçütü sırasıyla x ve y ile gösterilen temiz ve işlenmiş konuşmanın bir fonksiyonudur. Bu yöntemde model, konuşma anlaşılabilirliği için ilgili frekans aralığını kapsayacak şekilde 10 kHz'lik bir örnekleme oranı için tasarlanmıştır. Diğer örnekleme oranlarındaki tüm sinyaller yeniden örneklendirilmelidir. Ayrıca, hem temiz hem de işlenmiş sinyalin zaman hizalı (time-aligned) olduğu varsayılmaktadır. Temiz ve işlenmiş konuşma sinyallerinin

zaman-frekans gösterimi şu şekilde ifade edilebilir:

$$X_j(m) = \sqrt{\sum_{k=k_1}^{k_2(j)-1} |\hat{x}(k, m)|^2} \quad Y_j(m) = \sqrt{\sum_{k=k_1}^{k_2(j)-1} |\hat{y}(k, m)|^2} \quad (4.3)$$

Burada k_1 ve k_2 üçte bir oktav bandlarını temsil etmektedir.

Orta anlaşılabilirlik ölçütü $d_j(m)$ temiz ve işlenmiş sinyal arasındaki doğrusal korelasyon katsayısının bir tahmini olarak tanımlanmaktadır ve şu şekilde ifade edilebilir (Taal et al., 2010):

$$d_j(m) = \frac{\sum_n \left(X_j(n) - \frac{1}{N} \sum_l X_j(l) \right) \left(Y'_j(n) - \frac{1}{N} \sum_l Y'_j(l) \right)}{\sqrt{\sum_n \left(X_j(n) - \frac{1}{N} \sum_l X_j(l) \right)^2 \sum_n \left(Y'_j(n) - \frac{1}{N} \sum_l Y'_j(l) \right)^2}} \quad (4.4)$$

Burada, $l \in M$ şeklindedir. Sonuç olarak, nesnel anlaşılabilirlik (KZNA) ölçütü, tüm çerçeve ve bantlar üzerindeki orta anlaşılabilirlik değerlerinin ortalamasıyla hesaplanmaktadır ve şu şekilde ifade edilmektedir:

$$KZNA = \frac{1}{JM} \sqrt{\sum_{j,m} d_j(m)} \quad (4.5)$$

Burada M toplam çerçeve sayısını ve J ise üçte-bir oktav bant sayısını vermektedir.

Kişilerle yapılan öznel dinleme testlerinde elde edilen doğru kelimelerin yüzdesini tahmin etmek için KZNA sonuçlarına bir monotonik doğrusal haritalama uygulanması önerilmiştir (Khaldi et al., 2014). Bu yöntemde, iki veri tabanından gelen iyileştirilmiş konuşma sinyalleri göz önünde bulundurulduğunda sonuçların anlaşılabilirliğinin elde edilmesinde iyi bir hassasiyet gösterdiği anlaşılmaktadır. Tahmini anlaşılabilirlik skorları, KZNA sonuçlarına bir monotonik haritalama fonksiyonu uygulanarak şu şekilde elde edilebilir (Taal et al., 2011):

$$f(KZNA) = \frac{100}{1 + \exp(aKZNA + b)} \quad (4.6)$$

Burada $a = -17,4906$ ve $b = 9,6921$ alınmıştır. Konuşma iyileştirme yöntemlerinin anlaşılabilirliğini incelemek için mutlak KZNA değeri yerine

$f(KZNA)$ deęerini kullanmak daha pratik bir yol saęlamaktadır. $f(KZNA)$ deęerinin yksek olması anlaşılabiliirlięin daha iyi saęlandıęını gstermektedir (Zao et. al., 2014).



5. KONUŞMA VE GÜRÜLTÜ VERİ TABANLARI

TIMIT (Texas Instruments and Massachusetts Institute of Technology) veri tabanı; Amerikan İngilizce'sinin sekiz büyük lehçesini içeren, fonetik yönden zengin 630 cümleden oluşmaktadır. Bu cümleler, Tucker Davis Teknolojileri (TDT) cihazı ile 25 kHz örnekleme frekansı ve 16 bit çözünürlük ile kaydedilmiştir.

NOIZEUS veri tabanı, TIMIT veri tabanından nispeten fonetik dengeli ve zorluk düzeyi eşit olarak seçilmiş 30 cümleden (Çizelge 5.1) oluşmaktadır. Bu veri tabanında kullanılan cümleler, her biri beş cümle seslendirmek üzere üç erkek ve üç kadın konuşmacı tarafından oluşturulmuştur. Bu çalışmada İngilizce dili için NOIZEUS veri tabanı kullanılmış olup, örnekleme frekansı 8 kHz'e düşürülmüştür.

Çizelge 5.1 NOIZEUS veri tabanına ait cümle listesi

1	The birch canoe slid on the smooth planks.	16	A pot of tea helps to pass the evening.
2	Glue the sheet to the dark blue background.	17	Smoky fires lack flame and heat.
3	It's easy to tell the depth of a well.	18	The soft cushion broke the man's fall.
4	These days a chicken leg is a rare dish.	19	The salt breeze came across from the sea.
5	Rice is often served in round bowls.	20	The girl at the booth sold fifty bonds.
6	The juice of lemons makes fine punch.	21	The small pup gnawed a hole in the sock.
7	The box was thrown beside the parked truck.	22	The fish twisted and turned on the bent hook.
8	The hogs were fed chopped corn and garbage.	23	Press the pants and sew a button on the vest.
9	Four hours of steady work faced us.	24	The swan dive was far short of perfect.
10	A large size in stockings is hard to sell.	25	The beauty of the view stunned the young boy.
11	The boy was there when the sun rose.	26	Two blue fish swam in the tank.
12	A rod is used to catch pink salmon.	27	Her purse was full of useless trash.
13	The source of the huge river is the clear spring.	28	The colt reared and threw the tall rider.
14	Kick the ball straight and follow through.	29	It snowed, rained, and hailed the same morning.
15	Help the woman get back to her feet.	30	Read verse out loud for pleasure.

Son on yılda, Türkçe dili için konuşma işleme alanında önemli gelişmeler yaşanmıştır. Eklemeli bir dil olan Türkçe, metin-konuşma sentezi ve konuşma (cümle) tanıma ve iyileştirme alanlarında birçok zorluklar çıkarmaktadır. Bu nedenle, Türkçe ses ve metin yapısı oluşturmak için bazı araştırmalar yapılmıştır. Interactive Systems Laboratories, 100 konuşmacı tarafından okunan ve her biri 20 dakikalık olan Türk gazetelerinden alınan konuşma bölümlerini içeren çok dilli bir ses veri tabanı toplamıştır (Schultz and Waibel, 2001). Orta Doğu Teknik Üniversitesi Enformatik Enstitüsü, Sabancı Üniversitesi ile birlikte morfolojik ve sözdizimsel olarak ek açıklama yapan bir Türkçe metin yapısı geliştirmeye yönelik bir Türk Ağaç-bankası veri yapısı çalışması üzerinde çalışmışlardır (Ofłazer et al., 2003). Türkiye Bilimsel ve Teknik Araştırma Kurumu'nun UEKAE Konuşma Teknolojileri bölümü, 93 konuşmacı tarafından okunan tamamen fonetik dengeli 15 cümle ve 373 kelimeden oluşan bir telefon konuşma veri tabanı oluşturmuşlardır (Doğan and Arslan, 2002). Bir diğer veri tabanı çalışması ise Orien-Tel projesi kapsamında bir Türk telefon konuşma veri tabanıdır (Çilöglu et al., 2004).

Yukarıda bahsedilen çalışmaların tümü uygulamaya özel olup, fonetik açıdan dengeli bir metin kullanılmamaktadır ve bazıları yalnızca telefon konuşmalarına odaklanmıştır. Bu eksikliğe bir çözüm bulabilmek adına 2007 yılında Salor ve arkadaşları (Salor et al., 2007) tarafından ODTÜ'de Türkçe için konuşma tanıma alanlarında kullanılabilecek fonetik dengeli bir veri tabanı oluşturma çalışması yapılmıştır. Bu çalışma kapsamında, Amerikan İngilizcesi için oluşturulan TIMIT veri tabanına benzer, 193 konuşmacı tarafından seslendirilen fonetik olarak dengeli Türkçe konuşma cümleleri veri tabanı oluşturulmuştur. Oluşturulan bu veri tabanı, çalışmanın tamamlanmasının ardından birçok çalışmada kullanılmıştır (Hacıöglu et al., 2003; Salor et al., 2002 ; Salor, 2005).

Modern standart Türk dili, Fince ya da Japonca gibi sesli bir dildir ve bu durum fonemlerin yazılı dilde harflerle temsil edildiği anlamına gelmektedir. Yazılı metin ile telaffuzu arasında neredeyse birebir eşleşmenin olduğu söylenebilir. Bununla birlikte, bazı sesli ve sessiz harfler ses yollarında üretildikleri yere bağlı olarak değişime uğramaktadırlar. Bu nedenle Türkçe'deki 29 harf, 45 fonetik sembolle ifade edilmektedir (Ergenç, 1995). ODTÜ, Ergenç'in çalışmasında açıklanan fonetik sembol setine dayanan yeni bir metin oluşturma kuralı geliştirmiştir. Bu kurallar sözlükteki harflerin fonetik kayıtlarını gözlemleyerek ve ortaya çıktıkları koşulları belirleyerek geliştirilmiştir.

Salor ve arkadaşlarının ODTÜ’de oluşturduğu veri tabanı Türk gazetelerinin web sayfalarından derlenmiştir. Metin verileri, sayıları ve kısaltmaları genişletmek üzere normalize edilmiştir. Diğer dillerden gelen özel isimler gibi kelimeler silinmiştir. Normalleştirilmiş metin 2.539.850 kelime içermekte ve bu metinler METUbet simgelerine dönüştürülmüştür. Bu metin verileri kullanılarak Türkçe dilinde üçlü seslerin (trifon) görülme sıklığına ilişkin bazı bilgiler verilmektedir. Bu veri tabanında METUbet karakterleri bakımından üçlü ses sayısı 29.266 olarak belirlenmiştir. Bu veri tabanında Türkçe dili için en sık kullanılan 10 üçlü ses şu şekildedir: ‘evlerini’, ‘atlarını’, ‘evleri’, ‘**bir**’, ‘yanında’, ‘evini’, ‘evlerinde’, ‘atları’, ‘**ile**’, ‘atınm’ (Salor et al., 2002).

ODTÜ’de veri tabanı oluşturma için yapılan fonetik dengeli cümleler, Amerikan İngilizcesi TIMIT veri tabanının ilk 2000 cümlesinin Türkçe’ye çevrilmesi ile geliştirilmiştir. Daha sonra bu cümleler METUbet sembollerine dönüştürülmüş ve bu cümlelerin üçlü ses frekansları gazete web sitelerindeki metin yapıları ile karşılaştırılmıştır. Bu cümlelerde en az bir kez bulunan üçlü ses sayısı 9.492 olarak bulunmuştur. Türkçe için en çok kullanılan 5000 üçlü sesin kapsamasını sağlamak için 462 cümle eklenmiştir. Arttırılmış liste 11.033 üçlü ses içermekte ve ortaya çıkan 2462 cümle konuşma veri tabanını geliştirmek için kullanılmıştır. Oluşturulan veri tabanı 193 konuşmacı (89 kadın ve 104 erkek konuşmacı) tarafından rastgele seçilen 40 cümle seslendirilerek kaydedilmiştir. Yaş dağılımı 19 ile 50 arasında olup, yaş ortalaması 23,9’dur. Konuşma cümleleri 16 kHz örnekleme hızında Sennheiser mikrofonu ve bir Sound Blaster ses kartı ile sayısal olarak kaydedilmiştir.

Türkçe için yöntemlerin nesnel kalite ve anlaşılabilirlik değerlendirmesinde kullanılan, Middle East Technical University Turkish Microphone Speech v1.0 veri tabanından (Salor et al., 2007) alınan 30 cümle Çizelge 5.1’de verilmiştir.

Çizelge 5.2 Kullanılan Türkçe veri tabanına ait cümle listesi

1	Yüksek güvenli hapisane dikenli tellerle çevrilmişti.	16	Çığ küçük bir depremi tetikledi.
2	Her sene birçok yeni öğrenci ağırlıyoruz	17	Sekiz yaşındayken keman kurslarına başladı.
3	Hayat boyu arkadaş ve ortak kaldılar.	18	Vazelin diye yanlışlıkla jazelin yazmışlar.
4	İki hafta önce sen burada nöbetçi miydin?	19	Araba kullanırken planını düşündü.
5	Zafer öncesi hepimiz heyecanlıyız.	20	O gün Kaya okula gitmedi.
6	Deri bir el çantası uygun bir hediye olurdu.	21	Anlayışla beraber işbirliği tartışmaları en aza indirir.
7	Hala ruh hastalığı diye bir şey var.	22	Araba kapısı çarparak kapandı.
8	Derece yaparken sıkıntıya ve engellemelere maruz kalıyoruz.	23	O zaman belki de işler daha kötü.
9	Birinin düşünmeye başlaması gerekecek.	24	Pudra tenekesini kaptı ve dışarı koştu.
10	Vidayı sol elinle sık.	25	Fakat bu bildiri inandırıcılıktan tamamen uzak.
11	Sadece duyuruları öneri kutusuna atınız.	26	Davranışını itiraf etti ve savundu.
12	Baskın stres genellikle dikkat ve duygunun yoğunlaştığını gösterir.	27	Gençlerin çoğu sabah erken kalkar.
13	Küçük oğlan çocuğu solucanı olta iğnesine taktı.	28	Dansları şarkılı ve sözlü şiirler izledi.
14	Birçok turnikeli otoyol için başarı zor geldi.	29	Orhan hayatını kazanmak için midye ayıklıyor.
15	Kışkırtıcı bir şekilde bana baktı.	30	Sofrasında bulunanlara, izninizle diyerek kalktı.

İngilizce ve Türkçe konuşma cümlelerine farklı SNR seviyelerinde toplamsal olarak eklenen gürültüler, NOISEX-92 (Varga and Steeneken, 1993) gürültü sinyali veri tabanından alınan beyaz Gauss, araba, mırıltı, sergi, hava alanı, sokak ve tren gürültü tiplerinden oluşmaktadır. Bu gürültüler SNR seviyesi -10 dB'den 10 dB'e kadar değişen beş farklı SNR seviyesinde temiz cümlelere eklenerek gürültülü konuşma cümleleri elde edilmiştir.

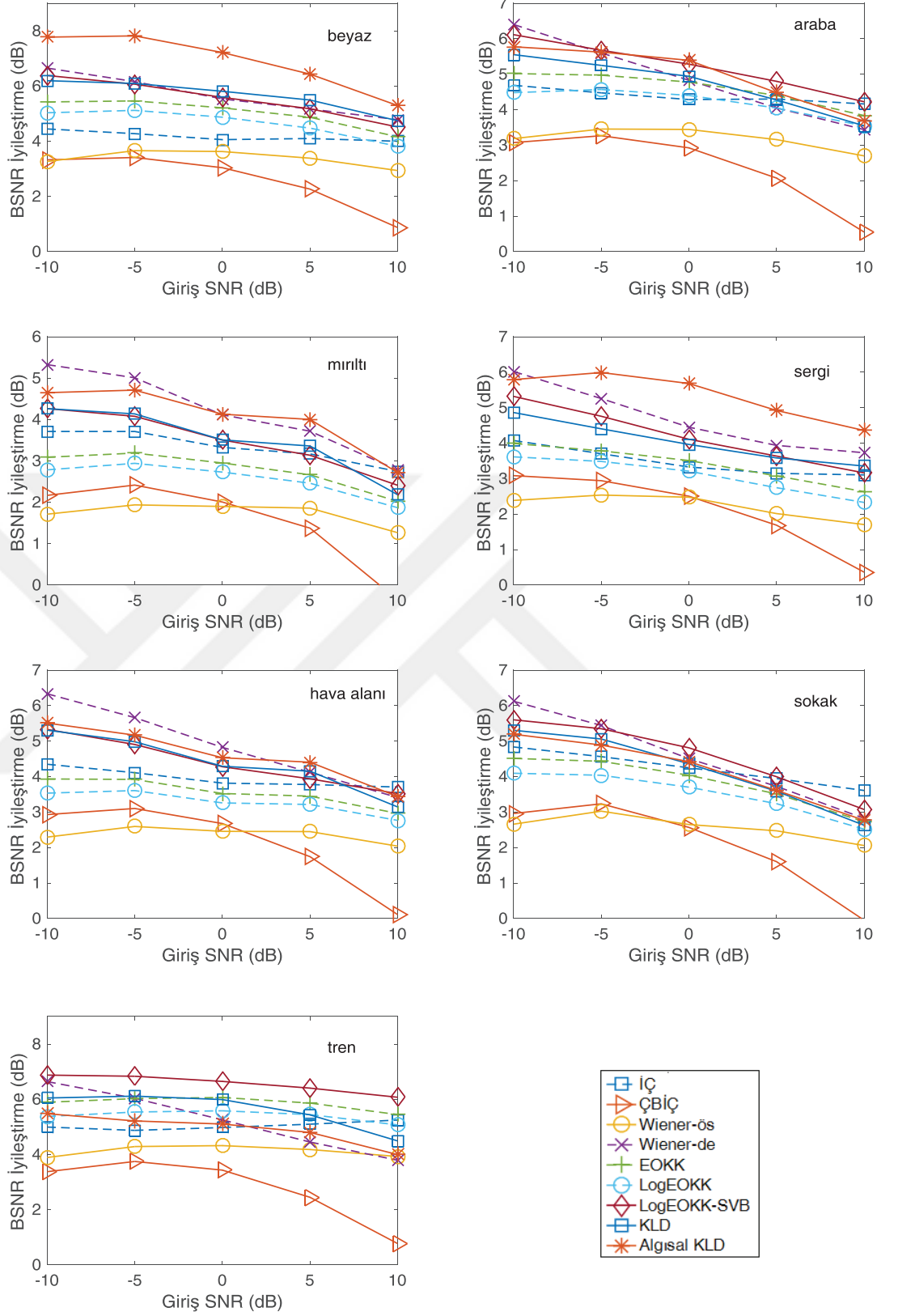
6. LİTERATÜRDEKİ KONUŞMA İYİLEŞTİRME YÖNTEMLERİNİN İNGİLİZCE VE TÜRKÇE İÇİN DEĞERLENDİRME SONUÇLARI

Türkçe ve İngilizce konuşma sinyalleri için, literatürde sıklıkla kullanılan konuşma iyileştirme algoritmalarının nesnel kalite ve nesnel anlaşılabilirlik değerlendirmesinin yapıldığı bu çalışmada, yedi farklı gürültü tipi ve beş farklı SNR seviyesi kullanılmıştır. İngilizce için NOIZEUS veri tabanı ve Türkçe için Middle East Technical University Turkish Microphone Speech v1.0 (ODTÜ) veri tabanındaki (Salor et al., 2007) 30 temiz cümle, NOISEX-92 gürültü veri tabanından alınan beyaz Gauss, mırıltı, araba, sergi, hava alanı, sokak ve tren gürültüleri ile -10, -5, 0, 5 ve 10 dB SNR seviyelerinde bozulmuştur.

Kullanılan iyileştirme algoritmaları, dört sınıfta kategorize edilen dokuz farklı yöntemden oluşmaktadır. Bu yöntemler; İzgesel Çıkarım (İÇ ve ÇBİÇ), İstatistiksel model tabanlı (EOKK, LogEOKK ve LogEOKK-SVB), Wiener tipi (Wiener-önsel SNR ve Wiener-dalgacık eşikleme), ve alt-uzay tabanlı (KLD ve AKLD) konuşma iyileştirme yaklaşımlarıdır. Konuşma iyileştirme algoritmaları; bölütsel SNR iyileştirme miktarı (dB) ve Ağırlıklı İzgesel Eğim (AİE) gibi klasik nesnel ölçütler ile değerlendirilmiştir. Ayrıca, konuşma kalitesinin algısal değerlendirmesi (KKAD) ve kısa-zamanlı nesnel anlaşılabilirlik (KZNA) ölçütleri ile nesnel anlaşılabilirlik değerlendirmesi yapılmıştır.

6.1 İngilizce İçin Yöntemlerin Nesnel Kalite ve Anlaşılabilirlik Ölçütleri İle Değerlendirilmesi

Beyaz Gauss, araba, mırıltı, sergi, hava alanı, sokak ve tren gürültüsü eklenerek beş farklı SNR (-10, -5, 0 5 ve 10 dB) seviyesinde bozulan konuşma sinyallerinin iyileştirilmesinde kullanılan dokuz farklı konuşma iyileştirme algoritmasının bölütsel SNR iyileştirme (BSNR) değerleri Şekil 6.1’de gösterilmektedir.



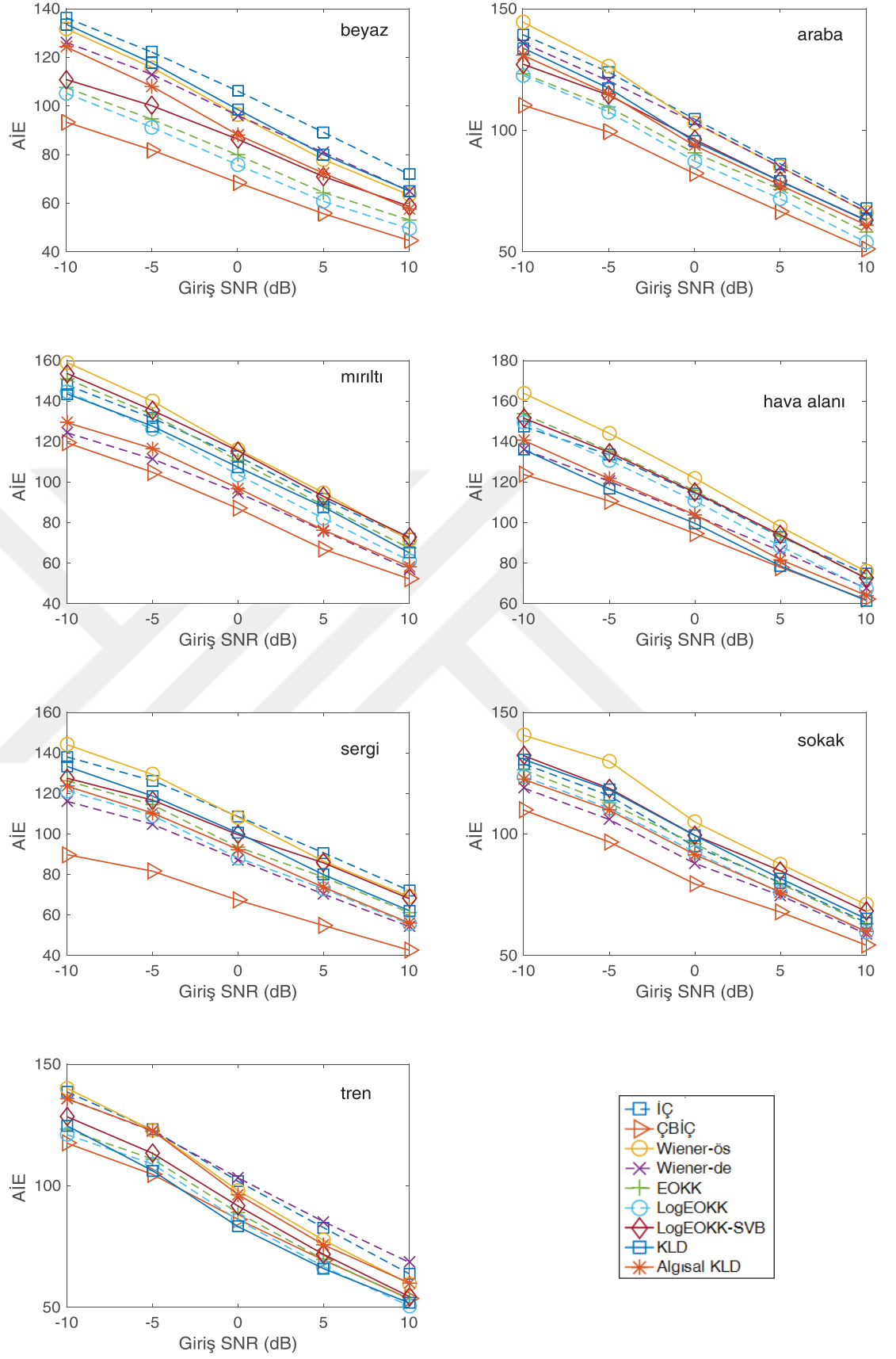
Şekil 6.1 İngilizce için tüm gürültü ve tüm SNR seviyelerinde BSNR İyileştirme (dB) değerleri

Şekil 6.1'den görülebileceği üzere, beyaz Gauss ve sergi gürültü tiplerinde tüm SNR seviyeleri için en iyi Bölütsel SNR (BSNR) iyileştirme miktarını Algısal KLD (AKLD) iyileştirme yöntemi sağlamaktadır. Araba gürültüsü için LogEOKK-SVB ve Algısal KLD yöntemi, mırıltı ve hava alanı gürültü tiplerinde ise düşük SNR seviyelerinde Wiener-dalgacık eşikleme (Wiener-de) iken yüksek SNR seviyelerinde ise Algısal KLD yöntemi daha fazla BSNR iyileştirme göstermektedir. Sokak gürültüsü için düşük SNR'larda Wiener-dalgacık eşiklime (Wiener-de) iken yüksek SNR'larda ise LogEOKK-SVB ve KLD iyileştirme yöntemleri daha fazla BSNR iyileştirme sağlamaktadır. Tren gürültüsü koşullarında LogEOKK-SVB yöntemi diğer iyileştirme yöntemleri ile kıyaslandığında daha fazla bölütsel SNR iyileştirme miktarı vermektedir.

Tüm gürültü koşullarının beş farklı SNR seviyeleri göz önünde bulundurulduğunda, beyaz Gauss, araba ve tren gürültü tipleri için en düşük BSNR iyileştirme miktarını (dB) Çoklu-Bant İzgesel Çıkarım (ÇBİÇ) yöntemi vermektedir. Benzer şekilde, mırıltı, sergi, hava alanı ve sokak gürültüleri için düşük SNR seviyelerinde (-10, -5 ve 0 dB) Wiener-önsel SNR (Wiener-ös) yöntemi, yüksek SNR seviyelerinde (5 ve 10 dB) ise ÇBİÇ yöntemi nesnel kalite ölçütü olan BSNR iyileştirme miktarı bakımından en kötü konuşma iyileştirme yöntemleri olarak belirlenmiştir.

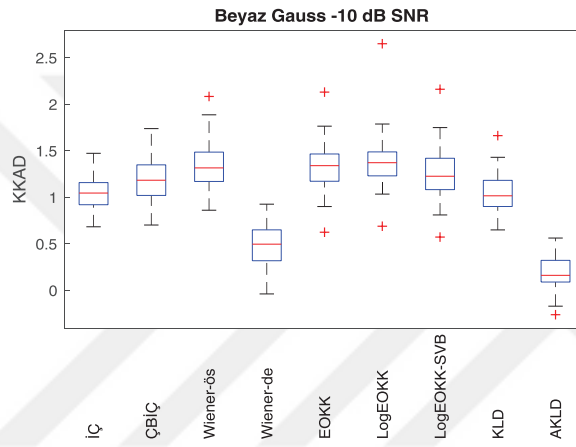
Kullanılan konuşma iyileştirme yöntemlerinin nesnel kalite ölçütü olan Ağırlıklı İzgesel Eğim (AİE) bakımından tüm gürültü çeşitleri ve beş farklı SNR seviyesi için performans değerlendirmesi yapılmıştır. AİE değerinin düşük olması nesnel kalite bakımından yöntemin başarısını göstermektedir. İngilizce konuşma sinyalleri kullanılarak tüm gürültü ve SNR seviyeleri için elde edilen AİE sonuçları Şekil 6.2'de verilmiştir. Elde edilen AİE sonuçlarına göre, İngilizce konuşmaların iyileştirilmesinde tren gürültüsü dışındaki tüm gürültülerde ve tüm SNR seviyelerinde ÇBİÇ yöntemi diğer yöntemlere göre daha iyi bir performans göstermektedir. Tren gürültüsünde ise düşük SNR seviyelerinde ÇBİÇ yöntemi daha düşük değerler verdiği için dolayı daha fazla iyileştirme yaparken yüksek SNR seviyelerinde tüm yöntemlerin AİE değerleri birbirine çok yakın bulunmaktadır.

AİE değerlerine göre en kötü performansı gösteren yöntem, gürültü tipi ve SNR seviyesine göre değişmekle birlikte genellikle Wiener-ös yöntemi daha yüksek AİE değerleri vermektedir. Bu yüzden, AİE sonuçlarına göre Wiener-ös yönteminin kullanılması iyileştirme bakımından sakıncalar yaratmaktadır.

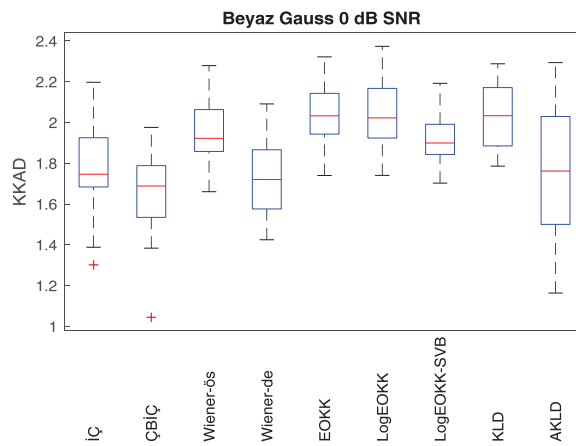


Şekil 6.2 İngilizce için tüm gürültü ve tüm SNR seviyelerinde yöntemlerin AİE değerleri

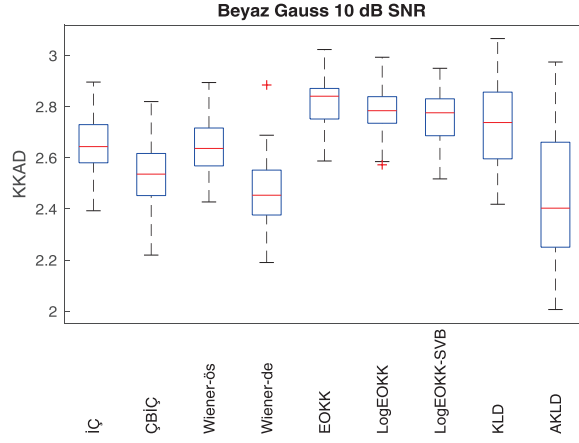
Konuşma iyileştirme yöntemlerini anlaşılabilirlik bakımından nesnel olarak değerlendirebilmek amacıyla KKAD ölçütü kullanılmaktadır. KKAD ölçütü kişilerle öznel değerlendirme yapmaya gerek kalmadan iyileştirilmiş konuşmaların anlaşılabilirliği hakkında bilgi vermektedir. Şekil 6.3’de beyaz Gauss gürültüsü koşullarında -10, 0 ve 10 dB SNR seviyeleri için tüm yöntemlere ait KKAD değerlerinin istatistiksel dağılımı verilmiştir. Tüm SNR seviyelerinde KKAD değerleri ise Şekil 6.4’de verilmiştir. Beyaz Gauss gürültüsü koşullarında KKAD sonuçlarına göre düşük SNR seviyelerinde (-10 ve -5 dB) LogEOKK yöntemi ve yüksek SNR seviyelerinde ise EOKK yöntemi anlaşılabilirliği diğer yöntemlere göre daha iyi sağlamaktadır.



(a)

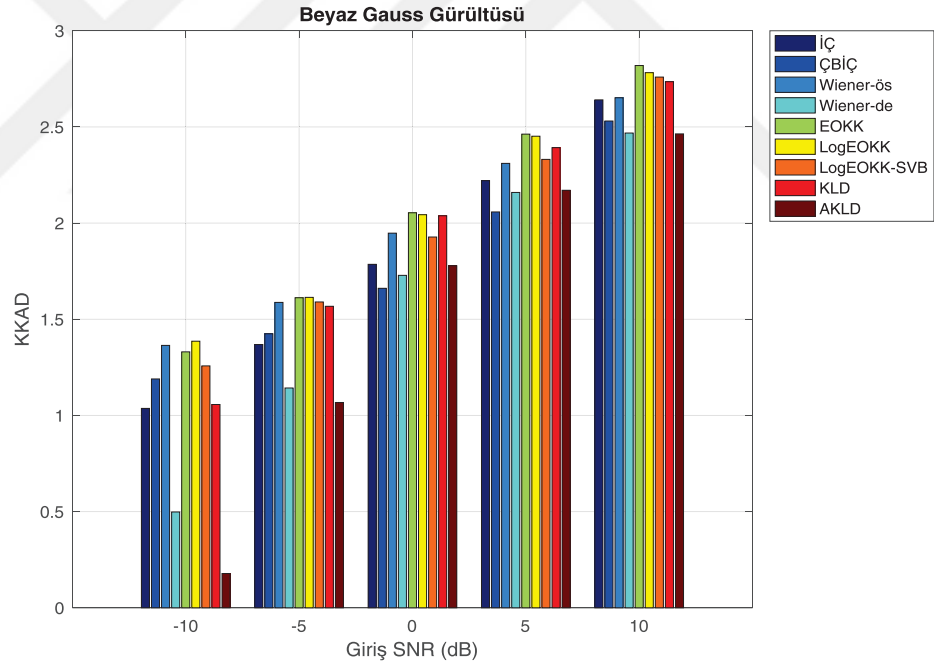


(b)



(c)

Şekil 6.3 İngilizce için beyaz Gauss gürültüsünün farklı SNR seviyelerinde KKAD sonuçları: (a) -10 dB, (b) 0 dB ve (c) 10 dB



Şekil 6.4 Beş farklı SNR seviyesinde Beyaz Gauss gürültüsü koşullarında tüm yöntemlere ait KKAD sonuçları

Tüm gürültü koşullarında farklı beş SNR seviyesi ve bu SNR seviyelerinin ortalaması alınarak elde edilen tüm yöntemlere ait KKAD değerleri Çizelge 6.1’de verilmiştir. Çizelge 6.1’den görülebileceği üzere; beyaz, araba, mırıltı, hava alanı, sergi, sokak ve tren gürültü koşullarında EOKK ve LogEOKK en iyi KKAD

değerlerine sahiptir. Her gürültü tipi için farklı SNR seviyelerinde KKAD değerleri değişkenlik gösterdiğinden beş SNR seviyesi için bulunan KKAD skorlarının ortalaması alınarak en iyi konuşma iyileştirme yönteminin bulunması amaçlanmıştır.

Çizelge 6.1 İngilizce konuşma sinyalleri için farklı yöntem ve gürültülere ait KKAD sonuçlarından elde edilen anlaşılabilirlik sonuçları

1-İÇ, 2-ÇBİÇ, 3-Wiener-ös, 4-Wiener-de, 5-EOKK, 6-LogEOKK, 7-LogEOKK-SVB, 8-KLD, 9-AKLD

Ortamlar		Yöntemler								
	SNR	1	2	3	4	5	6	7	8	9
beyaz	10	2,64	2,53	2,65	2,46	2,81	2,78	2,75	2,73	2,46
	5	2,22	2,05	2,31	2,15	2,46	2,45	2,33	2,39	2,17
	0	1,78	1,66	1,94	1,72	2,05	2,04	1,92	2,03	1,77
	-5	1,36	1,42	1,58	1,14	1,61	1,61	1,59	1,56	1,06
	-10	1,03	1,18	1,36	0,49	1,33	1,38	1,25	1,05	0,17
	Ort.	1,81	1,77	1,97	1,59	2,06	2,05	1,97	1,95	1,53
araba	10	2,73	2,72	2,75	2,40	2,91	2,87	2,90	2,67	2,40
	5	2,30	2,38	2,43	1,99	2,57	2,55	2,46	2,31	2,04
	0	1,87	1,97	2,05	1,54	2,16	2,16	1,98	1,89	1,54
	-5	1,35	1,62	1,68	0,97	1,77	1,79	1,60	1,42	0,92
	-10	1,06	1,19	1,33	0,45	1,44	1,47	1,32	0,96	0,46
	Ort.	1,86	1,97	2,05	1,47	2,18	2,17	2,05	1,85	1,47
mırıltı	10	2,53	2,64	2,57	2,34	2,65	2,65	2,60	2,47	2,32
	5	2,14	2,35	2,25	1,89	2,30	2,32	2,21	2,10	1,93
	0	1,73	1,98	1,86	1,50	1,93	1,95	1,86	1,74	1,54
	-5	1,32	1,64	1,51	0,97	1,61	1,61	1,53	1,34	0,99
	-10	0,95	1,13	1,20	0,41	1,24	1,36	1,22	0,85	0,42
	Ort.	1,73	1,95	1,88	1,42	1,95	1,98	1,86	1,70	1,44
hava alanı	10	2,69	2,76	2,70	2,51	2,79	2,77	2,79	2,64	2,47
	5	2,27	2,44	2,37	2,02	2,43	2,43	2,36	2,25	2,07
	0	1,87	2,12	2,03	1,64	2,06	2,09	1,97	1,90	1,63
	-5	1,40	1,72	1,64	1,09	1,67	1,70	1,56	1,39	1,03
	-10	1,18	1,42	1,30	0,98	1,44	1,46	1,38	1,26	0,99
	Ort.	1,88	2,09	2,01	1,65	2,08	2,09	2,01	1,89	1,64

sergi	10	2,48	2,55	2,49	2,26	2,60	2,59	2,54	2,42	2,30
	5	2,06	2,14	2,15	1,93	2,23	2,23	2,12	2,08	1,97
	0	1,67	1,75	1,80	1,40	1,85	1,87	1,67	1,64	1,46
	-5	1,19	1,44	1,38	0,77	1,42	1,44	1,29	1,08	0,66
	-10	0,81	1,21	1,03	0,43	1,06	1,15	1,02	0,68	0,47
	Ort.	1,64	1,82	1,77	1,36	1,83	1,86	1,73	1,58	1,37

sokak	10	2,81	2,84	2,91	2,46	2,98	2,95	2,88	2,68	2,42
	5	2,40	2,51	2,51	2,06	2,62	2,61	2,53	2,32	2,02
	0	2,00	2,21	2,16	1,65	2,26	2,28	2,15	1,99	1,61
	-5	1,56	1,84	1,78	1,21	1,89	1,87	1,73	1,45	1,02
	-10	1,19	1,37	1,34	0,72	1,48	1,51	1,36	1,16	0,47
	Ort.	1,99	2,15	2,14	1,62	2,25	2,24	2,13	1,92	1,51

tren	10	3,14	3,06	3,16	2,92	3,36	3,26	3,33	3,00	2,70
	5	2,68	2,78	2,82	2,49	2,96	2,92	2,91	2,63	2,32
	0	2,27	2,44	2,48	2,11	2,61	2,60	2,48	2,27	1,92
	-5	1,88	2,09	2,15	1,73	2,27	2,28	2,10	1,89	1,47
	-10	1,50	1,76	1,76	1,22	1,95	1,94	1,79	1,40	0,78
	Ort.	2,29	2,43	2,48	2,09	2,63	2,60	2,52	2,24	1,84

Ortalama Sonuçlar	1,88	2,02	2,04	1,60	2,14	2,14	2,03	1,87	1,54
--------------------------	-------------	-------------	-------------	-------------	-------------	-------------	-------------	-------------	-------------

İngilizce konuşma sinyalleri için KKAD anlaşılabilirlik ölçütü bakımından en iyi ve etkili yöntemin tespit edilmesi amacıyla tüm gürültü tipi ve tüm SNR seviyeleri için ortalama KKAD değerlerine bakılmıştır. Elde edilen sonuçlara göre 2,14 KKAD skoruna sahip olmalarından dolayı EOKK ve LogEOKK yöntemleri konuşma iyileştirmede anlaşılabilirlik bakımından daha başarılı bulunmuştur. KKAD sonuçlarına göre en iyi yöntemler olarak belirlenen EOKK ve LogEOKK yöntemlerini sırasıyla 2,04, 2,03 ve 2,02 KKAD skorlarına sahip Wiener-ös, LogEOKK-SVB ve ÇBİÇ yöntemleri izlemektedir. KKAD skoru bakımından iyileştirmede anlaşılabilirliğin sağlanmasında en kötü yöntem olarak 1,54 KKAD skoruna sahip Algısal KLD (AKLD) yöntemi tespit edilmiştir.

Tüm gürültü tipleri ve tüm SNR seviyeleri için kullanılan yöntemlere ait KZNA nesnel anlaşılabilirlik değerleri Çizelge 6.2’de verilmektedir. Tüm gürültüler için ayrı ayrı KZNA değerleri incelendiğinde ÇBİÇ, KLD ve Wiener-ös yöntemleri anlaşılabilirliği sağlamada ön plana çıkmaktadır. Tüm gürültülerin tüm SNR seviyeleri için KZNA ortalamalarına bakıldığında ise ÇBİÇ yöntemi

%74,31 anlaşılabilirlik oranı ile en iyi yöntem olarak bulunmaktadır. Anlaşılabilirliğin sağlanmasında ÇBİÇ yöntemini %72,06 oranına sahip KLD, %71,72 oranına sahip Wiener-ös ve %70,44 oranına sahip LogEOKK yöntemleri takip etmektedir.

Çizelge 6.2 İngilizce konuşma sinyalleri için farklı yöntem ve gürültülere ait KZNA sonuçlarından elde edilen anlaşılabilirlik oran tahminleri (%)

1-İÇ, 2-ÇBİÇ, 3-Wiener-ös, 4-Wiener-de, 5-EOKK, 6-LogEOKK, 7-LogEOKK-SVB, 8-KLD, 9-AKLD

Ortamlar		Yöntemler								
	SNR	1	2	3	4	5	6	7	8	9
beyaz	10	99,33	99,34	99,49	99,38	99,40	99,36	99,19	99,61	99,48
	5	95,46	96,81	98,05	97,21	97,47	97,42	96,16	98,56	98,08
	0	62,56	85,77	89,47	84,95	86,12	86,60	80,60	92,26	89,78
	-5	11,28	54,66	52,80	40,96	49,52	50,94	43,37	57,10	46,70
	-10	1,60	21,51	15,02	5,76	15,69	17,58	13,63	10,79	4,67
	Ort.	54,05	71,61	70,97	65,65	69,64	70,38	66,59	71,66	67,74

araba	10	99,34	99,38	99,43	99,26	99,35	99,32	99,08	99,50	99,43
	5	96,60	97,37	97,86	96,84	97,23	97,24	95,67	98,27	97,59
	0	75,63	88,16	88,59	81,80	85,01	84,89	78,83	90,44	83,08
	-5	19,36	57,09	54,95	32,71	51,39	52,12	43,90	54,16	22,62
	-10	2,58	20,73	13,33	4,27	14,30	15,78	13,26	9,61	1,86
	Ort.	58,70	72,55	70,83	62,98	69,46	69,87	66,15	70,40	60,92

mırıltı	10	99,22	99,29	99,26	99,19	99,16	99,14	98,87	99,39	99,33
	5	95,01	96,79	96,31	95,85	94,97	95,12	93,17	97,20	97,56
	0	61,39	82,11	77,60	76,26	71,54	72,78	66,61	83,06	79,83
	-5	11,04	34,89	25,79	22,24	21,55	23,59	18,30	30,59	17,75
	-10	1,72	8,35	5,11	2,43	4,23	4,99	2,88	3,44	1,14
	Ort.	53,68	64,28	60,81	59,19	58,29	59,12	55,97	62,74	58,92

hava alanı	10	99,43	99,53	99,54	99,39	99,47	99,46	99,30	99,58	99,53
	5	96,96	98,23	98,29	97,28	97,61	97,68	96,51	98,47	97,89
	0	78,42	91,80	89,35	87,25	86,13	86,58	82,01	92,73	86,57
	-5	23,86	64,50	54,02	43,68	47,10	48,62	39,35	57,16	27,30
	-10	3,56	23,81	12,04	6,84	11,30	12,45	9,26	10,21	1,71
	Ort.	60,44	75,57	70,65	66,89	68,32	68,96	65,29	71,63	62,60

sergi	10	99,35	99,38	99,45	99,40	99,24	99,21	98,93	99,58	99,54
	5	95,45	96,78	97,49	97,01	96,32	96,25	93,93	98,18	98,04
	0	59,87	82,93	83,17	76,21	75,72	76,46	66,01	86,38	80,20
	-5	6,27	33,27	27,55	15,64	21,47	23,01	14,86	27,16	15,17
	-10	0,91	11,50	4,57	1,37	4,02	4,87	2,93	2,51	1,05
	Ort.	52,37	64,77	62,45	57,92	59,36	59,96	55,33	62,76	58,80

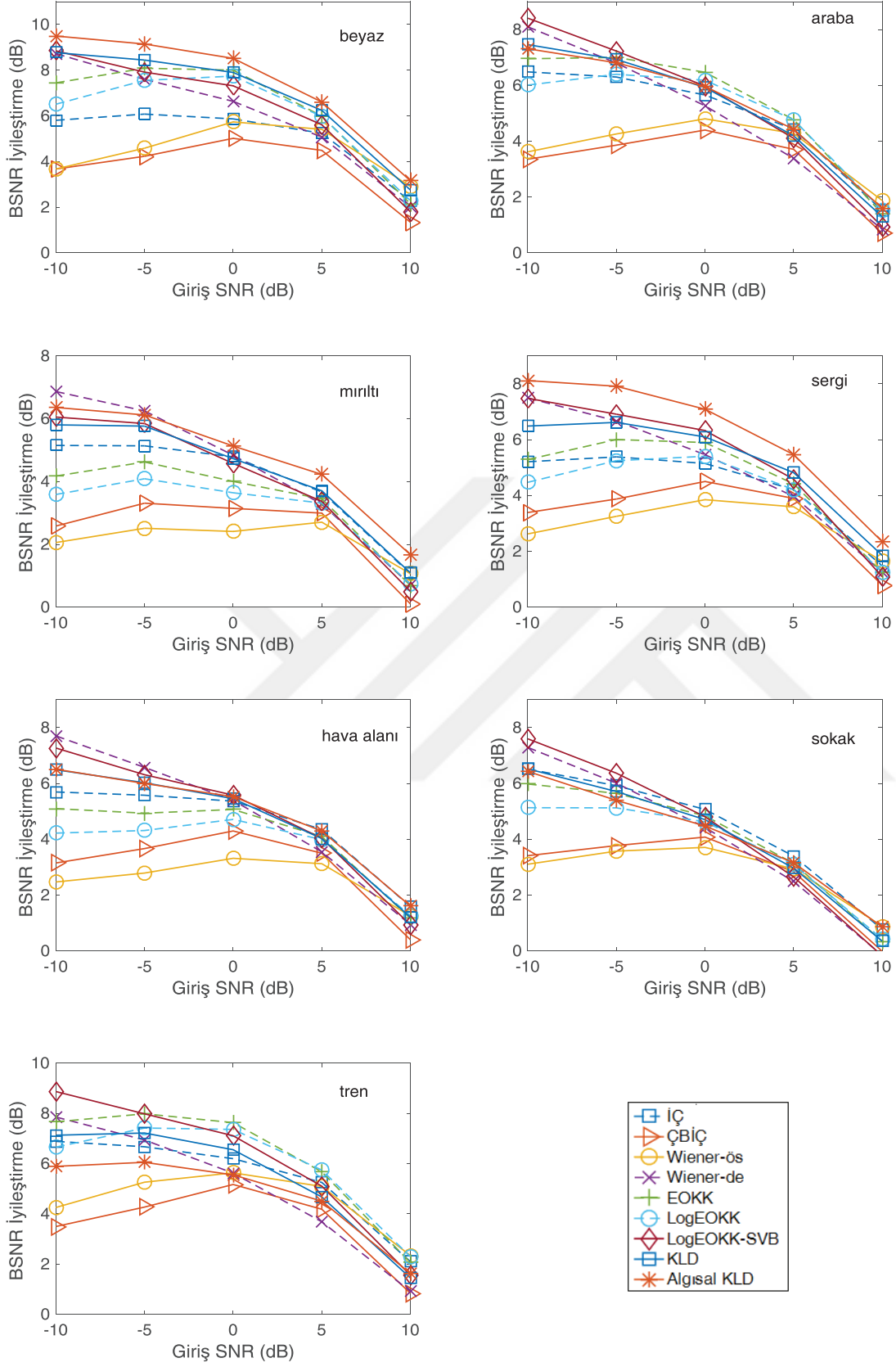
sokak	10	99,38	99,45	99,49	99,21	99,29	99,32	99,01	99,41	99,31
	5	97,28	98,10	98,10	97,13	97,40	97,49	96,22	98,19	97,34
	0	84,73	92,14	92,14	88,66	90,98	91,00	87,78	93,50	87,56
	-5	40,53	68,02	68,02	47,11	65,83	66,02	58,38	63,82	31,32
	-10	8,08	24,72	24,72	10,14	23,63	24,34	17,94	21,19	4,37
	Ort.	66,00	80,15	76,49	68,45	75,43	75,63	71,87	75,22	63,98

tren	10	99,74	99,78	99,81	99,68	99,79	99,78	99,72	99,79	99,74
	5	98,96	99,40	99,51	98,98	99,35	99,36	98,94	99,49	99,14
	0	94,22	97,87	98,02	96,11	97,23	97,27	95,56	98,30	95,92
	-5	65,16	91,65	90,94	84,01	88,94	88,67	84,66	92,62	73,85
	-10	20,27	67,52	61,02	43,19	61,53	60,95	54,46	60,11	11,51
	Ort.	75,67	91,24	89,86	84,39	89,37	89,21	86,67	90,06	76,03

Ortalama Sonuçlar	60,13	74,31	71,72	66,49	69,98	70,44	66,83	72,06	64,14
--------------------------	--------------	--------------	--------------	--------------	--------------	--------------	--------------	--------------	--------------

6.2 Türkçe İçin Yöntemlerin Nesnel Kalite ve Anlaşılabilirlik Ölçütleri İle Değerlendirilmesi

Türkçe konuşma sinyalleri için beş farklı SNR (-10, -5, 0 5 ve 10 dB) seviyesinde beyaz Gauss, araba, mırıltı, sergi, hava alanı, sokak ve tren gürültüsü eklenerek bozulan konuşma sinyallerinin iyileştirilmesinde kullanılan dokuz farklı konuşma iyileştirme algoritmasının BSNR iyileştirme değerleri Şekil 6.5’de gösterilmektedir.

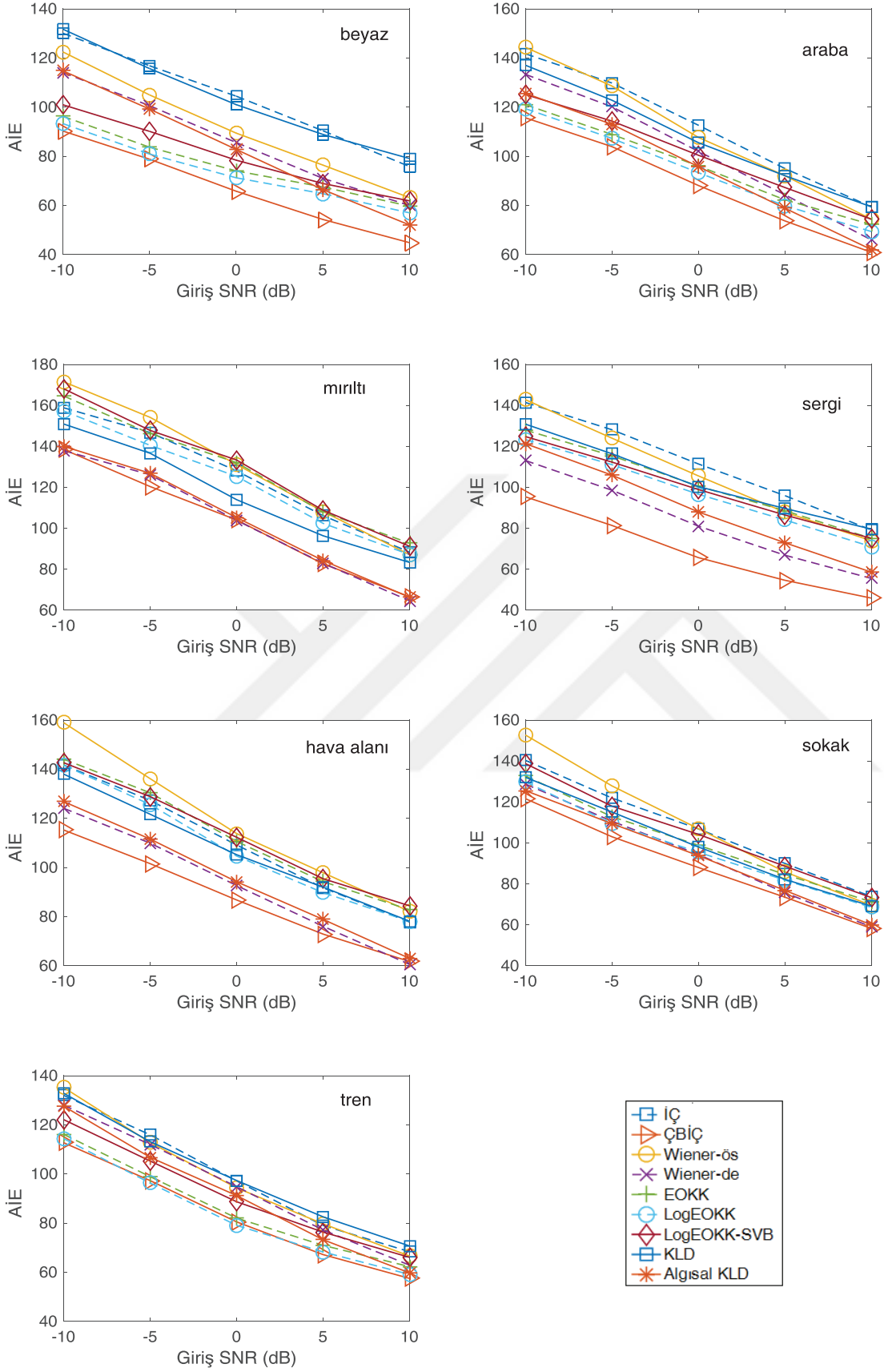


Şekil 6.5 Türkçe için tüm gürültü ve tüm SNR seviyelerinde BSNR iyileştirme (dB) değerleri

Şekil 6.5'den görülebileceği üzere, beyaz Gauss, mırıltı ve sergi gürültü tipi koşullarında tüm SNR seviyelerinde en iyi BSNR iyileştirme miktarını (dB) Algısal KLD (AKLD) yöntemi sağlamaktadır. Araba, sokak ve tren gürültüleri için düşük SNR (-10 ve -5 dB) seviyelerinde LogEOKK-SVB yöntemi diğer yöntemlere göre daha fazla iyileştirme miktarı sağlarken, yüksek SNR'larda tüm yöntemlerin BSNR iyileştirme miktarı birbirine çok yakın değerlerde bulunmuştur. Hava alanı gürültü koşullarında ise -10 dB ve -5 dB SNR seviyelerinde Wiener-dalgacık eşikleme (Wiener-de) yöntemi daha fazla iyileştirme miktarı verirken, yüksek SNR seviyelerinde tüm yöntemlerin iyileştirme miktarı bakımından performansı benzerlik göstermektedir.

Tüm gürültü koşullarının BSNR iyileştirme miktarı performansı değerlendirildiğinde, özellikle düşük SNR seviyelerinde beyaz, araba ve tren gürültü tiplerinde ÇBİÇ yöntemi en kötü sonuçları vermektedir. Mırıltı, sergi, sokak ve hava alanı gürültü tiplerinde ise düşük SNR seviyelerinde Wiener-ös iyileştirme yöntemi diğer yöntemlere göre daha düşük iyileştirme miktarı vermektedir.

Kullanılan konuşma iyileştirme yöntemlerinin nesnel kalite ölçütü olan Ağırlıklı İzgesel Eğim (AİE) bakımından tüm gürültü çeşitleri ve beş farklı SNR seviyesi için performans değerlendirmesi yapılmıştır. AİE değerinin düşük olması nesnel kalite bakımından yöntemin başarısını göstermektedir. Türkçe konuşma sinyalleri kullanılarak tüm gürültü ve SNR seviyeleri için elde edilen AİE sonuçları Şekil 6.6'da verilmiştir. Elde edilen AİE sonuçlarına göre, İngilizce konuşmaların iyileştirilmesinde mırıltı ve tren gürültüleri dışındaki tüm gürültülerde her SNR seviyesinde Çoklu-Bant İzgesel Çıkarım (ÇBİÇ) yöntemi diğer yöntemlere göre daha iyi bir performans göstermektedir. Mırıltı gürültü tipinde Algısal KLD (AKLD) ve ÇBİÇ yöntemlerinin AİE bakımından iyileştirme performansı birbirine yakın olmaktadır. Tren gürültüsünde ise tüm SNR seviyelerinde EOKK, LogEOKK ve ÇBİÇ yöntemleri AİE değerleri bakımından diğer yöntemlere göre daha iyi sonuçlara sahiptir.



Şekil 6.6 Türkçe için tüm gürültü ve tüm SNR seviyelerinde yöntemlerin AİE değerleri

AİE değerlerine göre en kötü performansı gösteren yöntem gürültü tipi ve SNR seviyesine göre değişmekle birlikte genellikle Wiener-ös yöntemi daha yüksek AİE değerleri vermektedir. Bu yüzden, AİE sonuçlarına göre Wiener-ös yönteminin kullanılması iyileştirme bakımından sakıncalar yaratmaktadır.

Konuşma iyileştirme yöntemlerini anlaşılabilirlik açısından değerlendirmek amacıyla KKAD ölçütü kullanılmaktadır. KKAD ölçütü kişilerle öznel değerlendirme yapmaya gerek kalmadan iyileştirilmiş konuşmaların anlaşılabilirliği hakkında bilgi vermektedir. Tüm gürültü koşullarında farklı beş SNR seviyesi ve bu SNR seviyelerinin ortalaması alınarak elde edilen tüm yöntemlere ait KKAD değerleri Çizelge 6.3’de verilmiştir. Çizelge 6.3’den görülebileceği üzere, beyaz, araba, mırıltı, hava alanı, sergi, sokak ve tren gürültü koşullarında EOKK ve LogEOKK en iyi KKAD değerlerine sahiptir. Her gürültü tipi için farklı SNR seviyelerinde KKAD değerleri değişkenlik gösterdiğinden beş SNR seviyesi için bulunan KKAD skorlarının ortalaması alınarak en iyi konuşma iyileştirme yönteminin bulunması amaçlanmıştır. Tüm gürültü ve SNR seviyelerinin ortalama KKAD skorlarına göre 2,13 ve 2,11 değerlerine sahip EOKK ve LogEOKK yöntemleri Türkçe konuşmalar için en iyi anlaşılabilirlik değerlerine sahiptir. Bu yöntemleri sırasıyla Wiener-ös, ÇBİÇ ve LogEOKK-SVB yöntemleri izlemektedir. Türkçe konuşmalar için en düşük KKAD skoruna sahip Wiener-de yöntemi anlaşılabilirliğin sağlanmasında diğer yöntemlere göre en kötü iyileştirme performansını göstermektedir.

Çizelge 6.3 Türkçe konuşma sinyalleri için farklı yöntem ve gürültülere ait KKAD sonuçlarından elde edilen anlaşılabilirlik sonuçları

1-İÇ, 2-ÇBİÇ, 3-Wiener-ös, 4-Wiener-de, 5-EOKK, 6-LogEOKK, 7-LogEOKK-SVB, 8-KLD, 9-AKLD

Ortamlar		Yöntemler								
	SNR	1	2	3	4	5	6	7	8	9
beyaz	10	2,69	2,52	2,65	2,49	2,79	2,75	2,73	2,78	2,62
	5	2,29	2,05	2,31	2,18	2,45	2,43	2,40	2,49	2,49
	0	1,76	1,63	1,98	1,77	2,09	2,07	1,99	2,14	2,23
	-5	1,37	1,44	1,62	1,14	1,65	1,65	1,56	1,70	1,86
	-10	1,09	1,26	1,36	0,58	1,37	1,42	1,28	1,08	1,19
	Ort.	1,84	1,78	1,99	1,63	2,07	2,06	1,99	2,04	2,08

araba	10	2,71	2,75	2,71	2,41	2,88	2,83	2,80	2,64	2,46
	5	2,36	2,37	2,42	2,03	2,58	2,54	2,44	2,33	2,17
	0	1,86	1,90	2,05	1,59	2,17	2,15	1,99	1,96	1,83
	-5	1,38	1,52	1,66	0,95	1,69	1,68	1,52	1,45	1,21
	-10	1,09	1,10	1,31	0,53	1,42	1,44	1,31	0,98	0,74
	Ort.	1,89	1,93	2,03	1,51	2,15	2,13	2,01	1,87	1,68

mırlıtı	10	2,57	2,67	2,55	2,37	2,63	2,61	2,58	2,53	2,42
	5	2,20	2,37	2,24	1,95	2,31	2,31	2,22	2,17	2,14
	0	1,71	1,98	1,88	1,46	1,92	1,93	1,81	1,78	1,73
	-5	1,27	1,62	1,48	0,84	1,49	1,53	1,42	1,25	1,13
	-10	1,12	1,23	1,08	0,54	1,36	1,28	1,20	0,76	0,68
	Ort.	1,77	1,97	1,85	1,43	1,94	1,93	1,84	1,70	1,62

hava alanı	10	2,68	2,77	2,67	2,49	2,75	2,72	2,70	2,63	2,51
	5	2,32	2,46	2,36	2,10	2,42	2,41	2,36	2,30	2,21
	0	1,86	2,13	2,03	1,68	2,03	2,05	1,93	1,92	1,82
	-5	1,41	1,75	1,65	1,07	1,62	1,65	1,53	1,49	1,26
	-10	1,29	1,34	1,26	0,98	1,42	1,33	1,28	1,06	1,04
	Ort.	1,91	2,09	1,99	1,66	2,05	2,03	1,96	1,88	1,76

sergi	10	2,54	2,57	2,51	2,31	2,65	2,61	2,61	2,52	2,42
	5	2,12	2,16	2,16	1,91	2,26	2,24	2,22	2,14	2,16
	0	1,66	1,70	1,80	1,44	1,84	1,84	1,75	1,70	1,76
	-5	1,22	1,41	1,38	0,78	1,32	1,38	1,27	1,08	1,23
	-10	1,02	1,13	1,01	0,42	1,10	1,08	1,05	0,86	1,05
	Ort.	1,71	1,79	1,77	1,37	1,84	1,83	1,78	1,66	1,72

sokak	10	2,85	2,88	2,89	2,49	2,96	2,92	2,79	2,72	2,55
	5	2,48	2,59	2,56	2,09	2,64	2,62	2,48	2,38	2,23
	0	2,05	2,24	2,23	1,58	2,30	2,30	2,15	2,03	1,81
	-5	1,63	1,84	1,82	1,04	1,85	1,87	1,75	1,54	1,24
	-10	1,27	1,28	1,30	0,86	1,42	1,38	1,48	1,18	0,90
	Ort.	2,06	2,17	2,16	1,61	2,23	2,22	2,13	1,97	1,75

tren	10	3,14	3,09	3,15	2,87	3,33	3,24	3,22	3,02	2,75
	5	2,79	2,84	2,83	2,53	3,00	2,93	2,87	2,71	2,48
	0	2,34	2,49	2,53	2,14	2,68	2,64	2,49	2,40	2,21
	-5	1,93	2,06	2,18	1,68	2,31	2,30	2,09	2,02	1,73
	-10	1,53	1,70	1,77	1,08	1,92	1,90	1,84	1,56	1,11
	Ort.	2,35	2,44	2,49	2,06	2,65	2,60	2,50	2,34	2,06

Ortalama Sonuçlar	1,93	2,03	2,04	1,61	2,13	2,11	2,03	1,92	1,81
--------------------------	-------------	-------------	-------------	-------------	-------------	-------------	-------------	-------------	-------------

Türkçe konuşmaların nesnel anlaşılabilirlik değerlendirmesinde tüm gürültü tipleri ve tüm SNR seviyeleri için kullanılan yöntemlere ait KZNA nesnel anlaşılabilirlik değerleri Çizelge 6.4’de verilmektedir. Tüm gürültüler için ayrı ayrı KZNA değerleri incelendiğinde ÇBİÇ, Wiener-ös ve KLD yöntemleri anlaşılabilirliği sağlamada ön plana çıkmaktadır. Tüm gürültülerin tüm SNR seviyeleri için KZNA ortalamalarına bakıldığında ise ÇBİÇ yöntemi % 68,21 anlaşılabilirlik oranı ile en iyi yöntem olarak bulunmaktadır. KZNA anlaşılabilirlik oranlarına göre ÇBİÇ yöntemini % 65,16 oranına sahip Wiener-ös ve % 64,47 oranına sahip KLD yöntemi takip etmektedir.

Çizelge 6.4 Türkçe konuşma sinyalleri için farklı yöntem ve gürültülere ait KZNA sonuçlarından elde edilen anlaşılabilirlik oran tahminleri (%)

1-İÇ, 2-ÇBİÇ, 3-Wiener-ös, 4-Wiener-de, 5-EOKK, 6-LogEOKK, 7-LogEOKK-SVB, 8-KLD, 9-AKLD

Ortamlar		Yöntemler								
SNR		1	2	3	4	5	6	7	8	9
beyaz	10	98,03	97,39	97,90	98,02	97,70	97,59	97,66	98,71	98,35
	5	89,54	90,18	92,27	92,97	92,46	92,30	92,78	95,50	95,19
	0	40,09	71,44	75,74	71,66	75,74	76,05	72,31	80,63	83,65
	-5	7,73	48,31	40,67	24,10	40,02	42,49	36,47	40,23	54,46
	-10	1,72	24,53	13,84	3,76	15,66	17,86	14,15	6,62	6,72
	Ort.	47,42	66,37	64,08	58,10	64,32	65,26	62,68	64,34	67,67

araba	10	98,63	98,51	98,61	98,57	98,41	98,34	98,16	98,90	98,73
	5	94,04	93,98	95,05	94,63	94,28	94,09	91,86	96,39	95,48
	0	55,84	74,96	77,85	71,03	74,04	74,35	66,12	83,08	77,74
	-5	10,14	39,34	33,21	15,93	33,34	34,55	27,43	31,80	20,72
	-10	2,07	16,13	8,32	1,89	10,56	11,72	9,26	4,37	1,36
	Ort.	52,14	64,58	62,61	56,41	62,13	62,61	58,57	62,91	58,81

mırıltı	10	98,23	98,37	98,20	98,44	97,91	97,95	97,65	98,64	98,62
	5	91,46	93,20	92,27	92,80	90,52	90,76	88,44	94,31	94,87
	0	39,77	67,19	61,23	59,45	52,63	54,49	46,46	67,77	68,22
	-5	4,08	22,52	14,43	5,72	11,32	13,08	8,42	10,36	6,97
	-10	0,86	5,07	2,45	0,62	2,08	2,55	1,40	1,04	0,56
	Ort.	46,88	57,27	53,72	51,41	50,89	51,77	48,47	54,42	53,85

hava alanı	10	98,83	98,90	98,88	98,99	98,72	98,74	98,62	99,09	98,94
	5	94,99	96,49	96,21	96,08	95,36	95,47	94,61	97,08	96,75
	0	62,28	84,45	83,77	77,36	77,74	78,89	69,54	84,78	80,02
	-5	12,67	45,75	39,59	17,91	30,61	33,08	21,77	33,42	16,14
	-10	1,94	13,68	6,57	1,32	5,98	6,77	4,32	3,14	0,83
	Ort.	54,14	67,86	65,00	58,33	61,68	62,59	57,77	63,50	58,54

sergi	10	97,90	97,37	97,62	98,09	97,01	96,99	96,85	98,41	98,51
	5	87,45	90,05	89,52	91,81	87,05	87,07	87,21	93,18	94,30
	0	40,46	71,96	70,96	64,77	65,40	66,72	60,21	74,80	78,18
	-5	5,65	37,57	27,02	13,73	22,78	25,25	16,62	22,55	26,79
	-10	1,29	16,95	7,47	1,73	6,32	7,82	4,80	3,92	2,78
	Ort.	46,55	62,78	58,52	54,03	55,71	56,77	53,14	58,57	60,11

sokak	10	99,12	99,09	99,14	98,92	98,85	98,86	98,46	99,24	99,03
	5	96,42	97,36	97,51	95,95	96,31	96,48	94,22	97,49	96,45
	0	75,29	88,77	87,37	73,70	84,73	85,05	77,61	86,60	74,65
	-5	23,14	56,86	46,32	17,50	45,86	46,38	34,45	28,16	10,32
	-10	5,35	23,22	11,78	3,03	14,37	15,13	10,67	5,91	1,30
	Ort.	59,87	73,06	68,42	57,82	68,02	68,38	63,08	63,48	56,35

tren	10	99,50	99,53	99,58	99,48	99,54	99,54	99,47	99,60	99,43
	5	98,67	98,90	99,06	98,67	98,87	98,85	98,51	99,16	98,61
	0	90,75	96,17	96,82	94,77	95,86	95,82	93,78	97,57	94,96
	-5	51,69	83,42	83,54	68,06	81,03	80,91	73,75	86,04	56,36
	-10	13,41	49,90	39,84	17,75	43,93	43,70	37,28	38,12	6,55
	Ort.	70,81	85,58	83,77	75,75	83,85	83,76	80,56	84,10	71,18

Ortalama Sonuçlar	53,97	68,21	65,16	58,83	63,80	64,44	60,61	64,47	60,93
--------------------------	--------------	--------------	--------------	--------------	--------------	--------------	--------------	--------------	--------------

6.3 Tartışma

İngilizce ve Türkçe konuşma sinyallerinin iyileştirilmesinde literatürde sıklıkla kullanılan yöntemlerin nesnel kalite ve nesnel anlaşılabilirlik ölçütleri ile değerlendirmesinin yapıldığı bu çalışmada, Bölütsel SNR iyileştirme miktarı (dB), Ağırlıklı İzgesel Eğim (AİE), konuşma kalitesinin algısal değerlendirmesi (KKAD) ve kısa-zamanlı nesnel anlaşılabilirlik (KZNA) ölçütü sonuçları göz önünde bulundurulmuştur. Elde edilen sonuçlara göre, Bölütsel SNR (BSNR) iyileştirme miktarına bakıldığında İngilizce ve Türkçe konuşma sinyalleri için Algısal KLD (AKLD) yöntemi diğer yöntemlere göre daha fazla BSNR

iyileştirme miktarı sağlamaktadır. İyileştirilmiş ve orijinal konuşma sinyalleri arasındaki izgesel eğim farkını veren AİE ölçütüne göre, hem İngilizce hem de Türkçe sinyaller için en düşük değerlere sahip ÇBİÇ yöntemi diğer yöntemlerle karşılaştırıldığında daha iyi bir iyileştirme olanağı sağlamaktadır.

İyileştirilmiş konuşmaların nesnel anlaşılabilirliğinin değerlendirilmesi amacıyla kullanılan KKAD ölçütüne göre, İngilizce konuşmalar için 2,14 KKAD skorlarına sahip EOKK ve LogEOKK yöntemleri diğer yöntemlere göre anlaşılabilirliğin sağlanmasında daha etkili olmaktadır. Bu yöntemleri sırasıyla 2,04, 2,03 ve 2,02 KKAD skorlarına sahip Wiener-ös, LogEOKK-SVB ve ÇBİÇ yöntemleri izlemektedir. KKAD sonuçlarına göre bu yöntemlerin anlaşılabilirlik skorları çok kötü ve çok rahatsız edici kategorisinde bulunmaktadır. Türkçe konuşma sinyallerinin anlaşılabilirlik ölçümünde ise İngilizce’de olduğu gibi EOKK ve LogEOKK yöntemleri sırasıyla 2,13 ve 2,11 KKAD skorları ile en iyi yöntemler olarak bulunmuştur. Ancak, bu KKAD skorları anlaşılabilirliğin sağlanmasında yetersiz kalmaktadır.

Kısa-zamanlı nesnel anlaşılabilirlik (KZNA) değerlendirmesinde ise ÇBİÇ yöntemi İngilizce için ortalama % 74,31 ve Türkçe için ise ortalama % 68,21 oranı ile en iyi skora sahiptir. ÇBİÇ yöntemi KZNA ölçütüne göre Türkçe konuşmaların anlaşılabilirliğini sağlamada İngilizce’ye göre daha düşük bir performans göstermektedir. Bu durum, dillerin yapısal farklılığından kaynaklanabilir.

Tüm sonuçlar göz önünde bulundurulduğunda, hem İngilizce hem de Türkçe konuşma sinyallerinin iyileştirilmesinde, literatürde sıklıkla kullanılan zaman alanı ve frekans alanı yöntemleri ölçütlere göre farklılıklar göstermektedir. Ayrıca ortalama KKAD ve ortalama KZNA skorları anlaşılabilirliğin sağlanmasında yetersiz kalmaktadır. Tüm gürültü ve SNR seviyeleri için yöntemlerin kalite ve anlaşılabilirlik performansları değişkenlik göstermektedir. Bu durum, hem İngilizce hem de Türkçe konuşma sinyallerinin iyileştirilmesinde kalite ve anlaşılabilirliği sağlayabilecek yeni bir yöntem ihtiyacı olduğunu göstermektedir. Bu nedenle, zaman alanı ve frekans alanı yöntemlerinin yerine bir zaman-frekans alanı yöntemi olan Ampirik Kip Ayırıştırması (AKA) tabanlı bir konuşma iyileştirme yaklaşımı önerilmiştir.

7. KONUŞMA/SESLİ ETKİNLİK ALGILAMA

Konuşma ya da sesli etkinlik algılama (Speech/Voice Activity Detection - VAD); gürültülü konuşma sinyallerinde, konuşmanın etkin olduğu bölgelerin konuşma dışı bölgelerden ayırt edilmesini sağlayan bir yöntemdir. Konuşma dışı bölgeler; sessizlik (silence), gürültü ve diğer ilgisiz akustik bilgi parametrelerini içermektedir. Sesli etkinlik algılama (SEA), konuşma kodlama (Freeman and Cosier, 1989; Beritelli et al., 1998; Enqing et al., 2002; Zhang et al., 2005), konuşma sinyallerinin bölütlenmesi ve konuşma tanıma sistemleri (Karray and Martin, 2003), sayısal işitme cihazlarında gürültü azaltma, mobil iletişim ve konuşma iyileştirme (Woo et al., 2000) gibi birçok konuşma ve ses işleme uygulamalarında sıklıkla kullanılmaktadır.

Geleneksel SEA algoritmalarının birçoğu basit yapılarından dolayı enerji, perde periyodu (pitch) tespiti, sıfır-geçiş oranı (zero-crossing rate), doğrusal öngörü katsayıları ve kepsstral özellikler gibi ölçütlerin eşiklenmesine dayanmaktadır (Greenwood and Kinghorn, 1999; Ilarsha, 2004; Babu and Vanathi, 2009; Bachu et al., 2010). Ancak bu ölçütler ile yapılan SEA uygulamalarında, düşük sinyal gürültü oranlarında ya da durağan olmayan arka plan gürültülerinde konuşma bölgelerinin tespiti zorlaşmaktadır. Bu problemi çözmek için izge, öz-ilişki, sınırlı frekans bantlarının gücü, periyodiklik ve yüksek mertebe istatistikler gibi gürbüz akustik öznitelikler SEA için önerilmiştir. Bu yöntemlerin birçoğunda, arka plan gürültüsünün belirli periyotlar boyunca durağan özellik gösterdiği varsayılmaktadır. Bu nedenle bu yöntemler, işlenen konuşma sinyalindeki SNR değişimlerine oldukça duyarlıdır.

Son yıllarda, çok sayıda araştırmacı, SEA'nın konuşma ve ses işleme uygulamalarının performansı üzerindeki etkisini belirlemek için farklı yaklaşımlar araştırmışlardır. Bu yaklaşımlardan istatistiksel modellere dayanan olabilirlik oranı testini kullanan gürbüz (robust) SEA algoritmaları konuşma bölgelerinin belirlenmesinde geleneksel yöntemlere göre daha iyi bir performans göstermektedir (Sohn, 1999; Chang and Shin, 2003). Bir diğer yaklaşımda (Benyassine et al., 1997), çizgisel izgesel frekanslar (Line Spectral Frequencies - LSF), düşük bant enerjisi, sıfır geçiş hızı ve tam-bant enerjisini içeren bir dizi ölçütü kullanarak sezgisel olarak belirlenmiş bölgeler ve sınırlar ile birlikte kullanılan bir SEA olan G.729B algoritmasını önermişlerdir. G.729B SEA algoritmasını geliştirmek amacıyla, bu algoritma için kullanılan akustik özellikler seri olarak birleştirilmiş ve destek vektör makinesinin SEA'ya uygulanması

önerilmiştir (Enqing et al., 2002). Böylece önerilen algoritmanın G.729B SEA üzerinde dikkate değer bir gelişme elde ettiği görülmüştür.

Gürültülü koşullar altında, enerji, sıfır-geçiş oranı ve düşük frekans bant gücü gibi zaman alanı, spektrum ve periyodiklik/a-periyodiklik gibi frekans alanı ve dalgacık dönüşümü gibi zaman-frekans alanı yöntemleri konuşma/ses bölgelerinin tespitinde yetersiz kalmaktadır. Bu yöntemlerin birçoğunda, tüm gürültü tipi ve tüm SNR seviyelerinde eşik değerleri tam ve doğru bir şekilde hesaplanamamaktadır. Bu nedenle, son yıllarda, konuşma bölgelerinin belirlenmesinde konuşma sinyallerini yüksek frekanslardan düşük frekanslara doğru içkin (intrinsic) kiplere ayırtıran Ampirik Kip Ayırtırması (AKA) yöntemi kullanılmaktadır (Kanai et al., 2013; Chong Feng, 2014).

7.1 Standart AMR1, AMR2 ve G.729B Yöntemleri

AMR seçenek 1 (AMR1), AMR seçenek 2 (AMR2) ve G.729B sesli etkinlik algılama (SEA) algoritmaları özellikle haberleşme alanında sesli ve konuşma bölgelerini belirlemede kullanılan ve bir standardı olan üç SEA yöntemidir. AMR1 daha yüksek frekans bantlarına vurgu yapan ve süzgeç kenarlarını kullanarak sinyali dokuz alt banda ayırtıran bir sesli etkinlik algılama algoritmasıdır. Her bir alt bant için enerji ve sinyal-gürültü oranı (SNR) tahminlerini hesaplamaktadır. Sinyal-gürültü oranlarının toplamı daha sonra SEA kararını vermek için uyarlamalı eşik değerleri ile karşılaştırılmaktadır. AMR2, AMR1'e yapı olarak benzetmekle birlikte süzgeç kenarları yerine Fourier dönüşümü kullanılmaktadır. AMR2 algoritması sinyali 16 banda ayırtırmakta ve konuşma olmayan çerçeveler boyunca her bant için arka plan gürültü enerjisini uyarlamaktadır. Genel olarak AMR algoritmaları değişen gürültü koşullarında iyi çalışmaktadır. Bununla birlikte, algoritmadaki sınırlamalar sebebiyle konuşma dışı bölgelerin tespitinde yetersiz kalmaktadır (Torre et al., 2006).

Bir ITU-T standart SEA algoritması olan G.729B ise çizgisel izgesel frekanslar, düşük bant enerjisi, sıfır geçiş hızı ve tam-bant enerjisini içeren bir dizi ölçütten oluşmaktadır. Bu algoritmada, sezgisel olarak belirlenmiş bölgeler ve sınırlar ile birlikte kullanılan ve her 10 ms için karar veren bir yapı kullanılmaktadır (Benyassine et al., 1997). Arka plan gürültü modelini kullanan bu SEA yönteminde karar vermede aşağıdaki dört parametre kullanılmaktadır:

- Giriş sinyali ve gürültü modeli arasında tam bant enerji farkı
- Giriş sinyali ve gürültü modeli arasında düşük bant enerji farkı
- İzgesel bozulma
- Gürültü sinyali ve gürültü modeli arasında sıfır geçiş oranı farkı

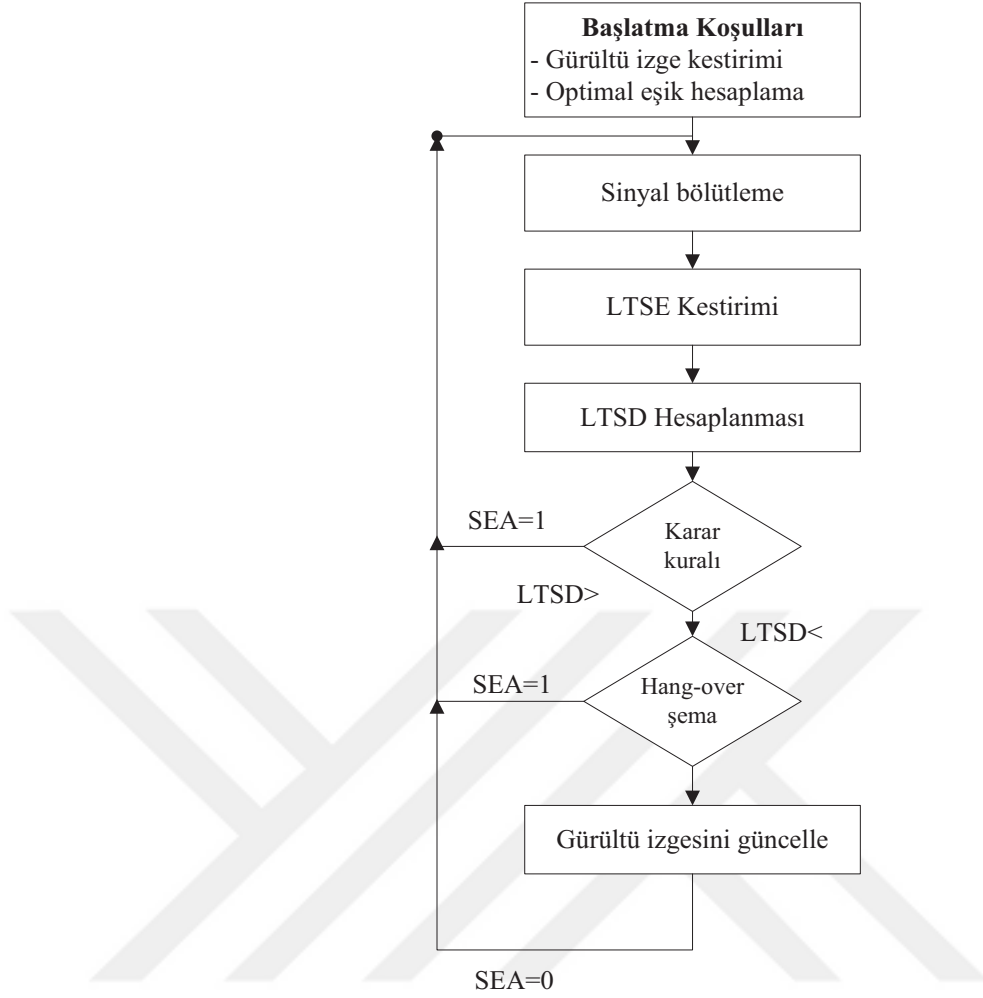
G.729B SEA için kullanılan akustik özellikler seri olarak birleştirilmiş ve destek vektör makinesinin SEA'ya uygulanması önerilmiştir (Enqing et al., 2002). Böylece önerilen algoritmanın G.729B SEA üzerinde dikkate değer bir gelişme elde ettiği görülmüştür.

7.2 SOHN Sesli Etkinlik Algılama

Bir sesli etkinlik algoritmasının (SEA) karar kuralı, genellikle gürültü ve gözlemlenen sinyal istatistikleri arasındaki farkın ölçümü olan bir büyüklük ve deneysel olarak belirlenen bir karar eşiğinden oluşmaktadır. Bu yöntemde, sağlam bir karar kuralı, gürültü istatistiğinin öncelikli olduğu varsayılarak, geliştirilmiş olabilirlik oranı testinden elde edilmektedir. Zamanla değişen gürültü istatistiklerini kestirmek için, konuşma sinyalinin bulunduğu bölgelerde, yumuşak karar bilgisini kullanan yeni bir gürültü izgesi uyarlama algoritması geliştirilmiştir. Temiz konuşma sinyallerinin sesli bölgelerinin belirlenmesi için tasarlanmış olan bu algoritma, gürültülü konuşmalarda konuşma ve konuşma dışı bölgelerin belirlenmesinde aynı performansı gösterememektedir (Sohn, 1999; Chang and Shin, 2003).

7.3 Uzun-Süreli İzgesel Ayrılma

Standart SEA algoritmalarının çoğu kısa süreli analiz çerçevelerinden çıkarılmaktadır ve SEA kararı bu kısa süreli çerçeveler üzerinden yapılmaktadır. SEA için kısa süreli çerçeve kullanılmasının aksine uzun süreli bir SEA analizi olan uzun süreli izgesel ayrılma (Long-Term Spectral Divergence-LTSD) önerilmiş ve uzun süreli izgesel zarfı ortalama gürültü izgesi ile karşılaştırılarak konuşma ve konuşma dışı karar kuralı oluşturulmuştur (Ramírez et al., 2004). Şekil 7.1'de LTSD SEA algoritmasının akış diyagramı verilmiştir.

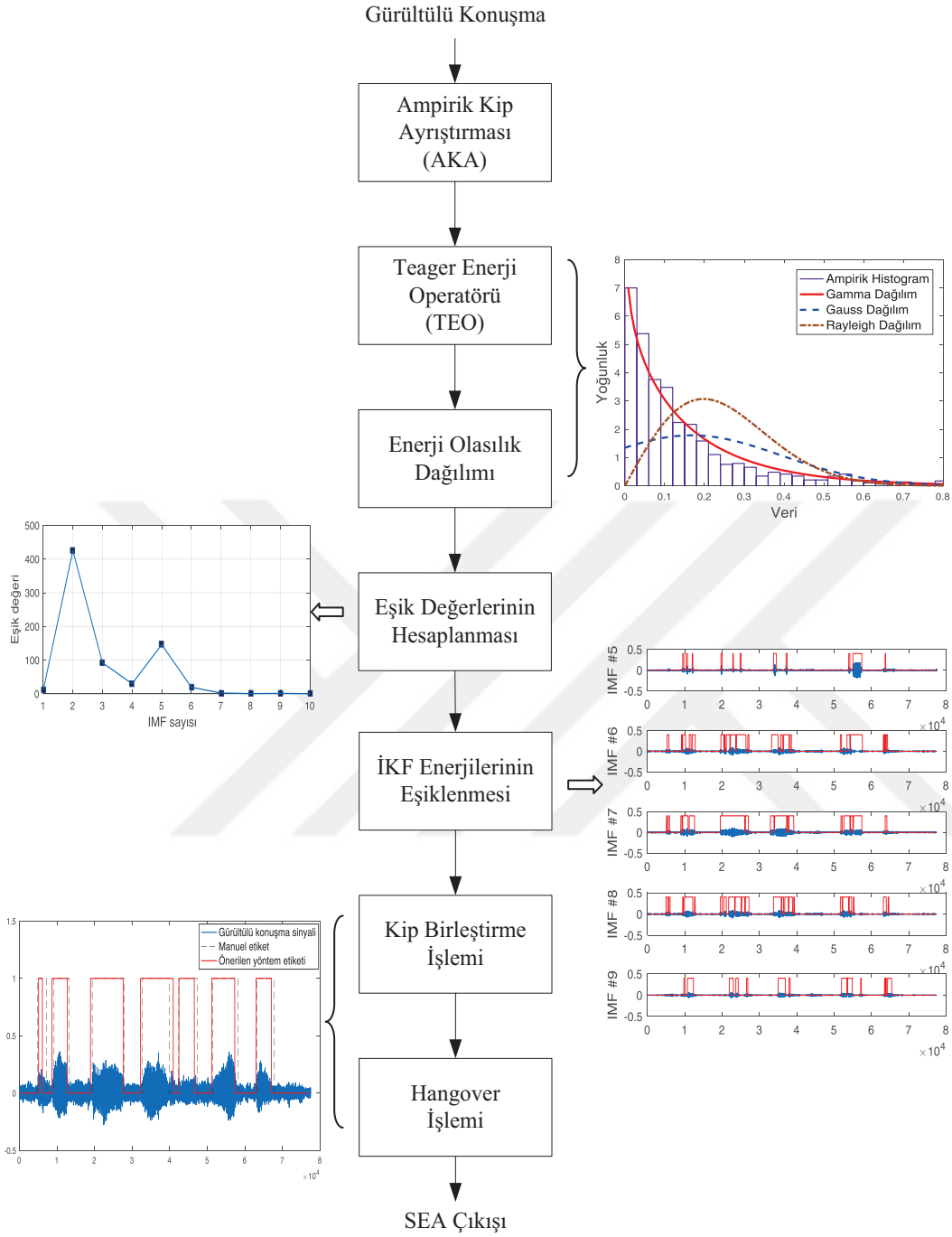


Şekil 7.1 LTSD SEA yönteminin akış diyagramı

Şekil 7.1’de görüldüğü gibi kısa bir başlatma periyodunda gürültü izgesi, gürültü izge büyüklüklerinin ortalaması ile kestirilmektedir. Başlatma periyodundan sonra, uzun süreli izgesel zarf hesaplamasına dayanan LTSE (Long-Term Spectral Envelope) SEA algoritması sinyalin izgesini örtüşmeli bantlara ayırmaktadır. LTSD ise LTSE değerlerinin gürültü izge ortalamasına bölünmesi ile elde edilmektedir. LTSD değerinin kara kuralından büyük olması durumunda SEA=1 aksi durumda ise SEA=0 olmaktadır.

7.4 Önerilen Konuşma/Sesli Etkinlik Algılama Yaklaşımı

Bu çalışmada, gürültülü koşullar altında AKA yöntemi ile elde edilen İKF katsayılarından elde edilen uyarlanır eşik değerlerinin kullanılmasına dayalı yeni bir konuşma/sesli bölge belirleme yaklaşımı geliştirilmiştir. Önerilen yöntemin blok diyagramı Şekil 7.2’de gösterilmektedir.



Şekil 7.2 Önerilen yöntemin blok diyagramı

Şekil 7.2’den görülebileceği üzere, önerdiğimiz yöntemde, diğer sesli etkinlik algılama yöntemlerinden farklı olarak konuşma ve konuşma dışı bölgelerin İKF katsayılarının enerjisinin uygun bir istatistiksel modeli

çıkartılmaktadır. Çıkartılan istatistiksel dağılım yardımıyla, gürültülü konuşma sinyallerinin konuşma olma olasılığı veya gürültü olma olasılığı üzerinden uyarlanabilir eşik değerlerinin belirlenmesi amaçlanmıştır. Önerilen bu yaklaşımda, gürültülü konuşma sinyali AKA yöntemi ile İKF katsayılarına ayrıştırılmış ve bu İKF katsayılarına Teager Enerji Operatörü (TEO) uygulanmıştır. TEO uygulanmış İKF katsayılarından bir istatistiksel model olan Gamma dağılım fonksiyonu yardımıyla hem gürültülü sinyalin hem de gürültünün olasılık dağılım fonksiyonu elde edilmiştir. Simetrik Kullback-Leibler (SKL) uzaklığı kullanılarak konuşmanın ve gürültünün olduğu kiplere göre uygun eşik değerleri belirlenmiştir. Her kip için uygun eşik değerleri belirlendikten sonra, ilgili kipteki İKF katsayıları uyarlanabilir eşik değerlerine göre konuşma ve konuşma dışı bölgeler olarak etiketlenmektedir. Son olarak, tüm bu kiplerden elde edilen etiketler ‘veya (or)’ işlemi ile birleştirilir ve hangover işlemi ile yumuşatılarak gürültülü konuşma sinyalinin konuşma ve konuşma dışı bölgeleri belirlenmektedir.

7.4.1 Ampirik Kip Ayrıştırması

Hilbert Huang Dönüşümü’nün (HHD) temel bir bileşeni olarak Ampirik Kip Ayrıştırması (AKA), doğrusal olmayan ve durağan olmayan sinyallerin işlenmesi için uyarlanabilir bir yöntemdir (Khaldi, 2012). AKA yöntemi, önceki sinyal işleme yöntemlerinden (Fourier, Dalgacık ve Dalgacık Paket Dönüşümü) farklı olarak sezgisel, doğrudan ve uyarlanabilir bir yöntemdir. Ayrıştırmanın bütün işlemlerinde, tüm temel fonksiyonlar sinyalin kendisinden türetilmiştir. Bu nedenle, AKA yöntemi konuşma sinyali gibi doğrusal ve durağan olmayan sinyalleri işlemek için oldukça uygundur. Bir $x(t)$ sinyali göz önüne alındığında, AKA yöntemi, "eleme" (sifting) işlemi yoluyla elde edilen ve İçkin Kip İşlevleri (İKF) olarak adlandırılan bir dizi salınımsal bileşene uyarlanabilir biçimde ayrıştırmaktadır. Elde edilen her İKF, orijinal sinyalden gelen frekans bileşenlerinin bir alt kümesinden oluşan bir salınımsal sinyaldir.

Eleme (sifting) süreci aşağıdaki gibi tanımlanabilir:

Adım 1: $x(t)$ 'nin tüm ekstremaları (aşırılıkları) (maksima / minima) bulunur.

Adım 2: Sırasıyla bir üst zarf $e_{max}(t)$ ve bir alt zarf $e_{min}(t)$ elde etmek için ‘kübik spline interpolasyon’ ile maksima ve minimal noktaları hesaplanır.

Adım 3: Üst ve alt zarfların ortalaması $m_1(t) = (e_{max}(t) + e_{min}(t))/2$ hesaplanır ve $h_1(t) = x(t) - m_1(t)$ elde etmek için orijinal sinyalden çıkarılır.

Adım 4: $h_1(t)$ 'nin aşağıdaki iki İKF koşulunu karşılayıp karşılayamadığı değerlendirilir:

- Tüm veri parçalarında ekstremaların sayısı sıfır geçişlerinin sayısına eşit olmalı ya da en fazla bir fark olmalıdır.
- Her noktada, maksimum zarfın ve minimum zarfın ortalaması sıfır olmalıdır (sinyal zaman ekseninde simetrik olmalıdır).

$h_1(t)$, bir İKF olma koşulunu sağlıyorsa, ilk $\dot{IKF}_1(t)$, $\dot{IKF}_1(t) = h_1(t)$ olarak kabul edilir.

Adım 5: Eğer $h_1(t)$ iki koşulu sağlamazsa, $h_1(t)$ yeni bir sinyal olarak kabul edilir ve Adım 1-4 $h_2(t)$ 'yi oluşturmak için $h_1(t)$ üzerinden tekrarlanır. Eğer $h_2(t)$ iki koşulu sağlayamıyorsa, eleme işlemini sonlandırmak için bir standart sapma olan durdurma kriteri (Standart Sapma-SS) hesaplanır. Durdurma kriteri şu şekilde verilmektedir:

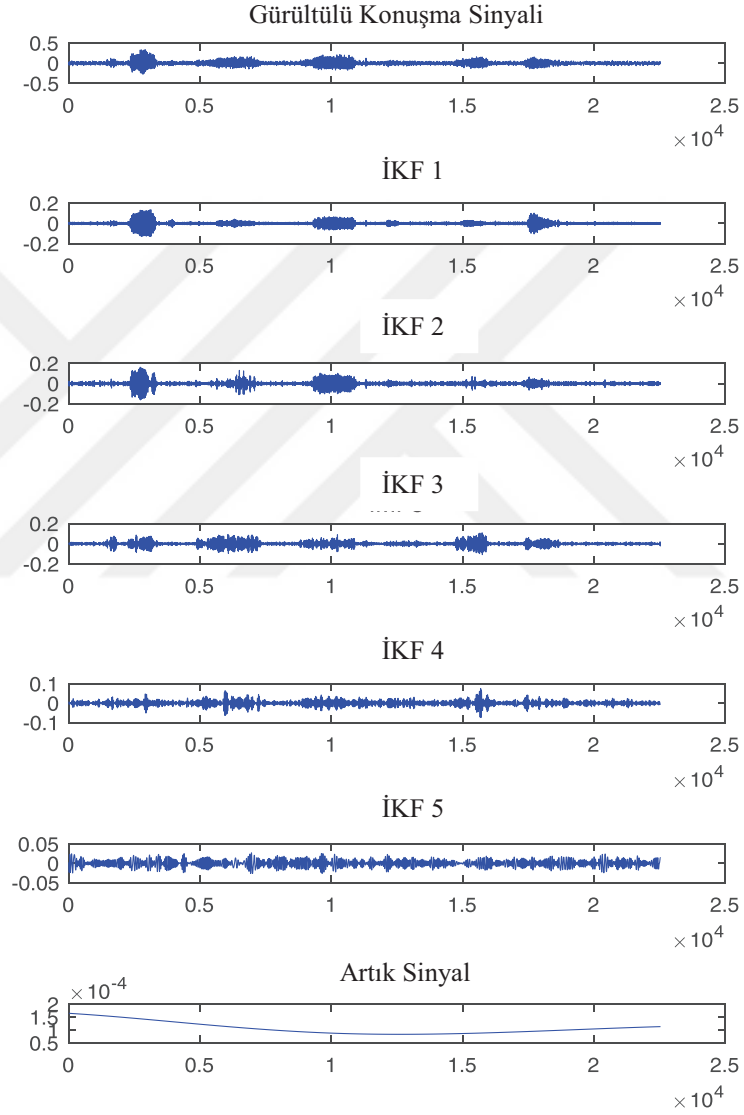
$$SS(i) = \sum_{t=0}^N \frac{|h_{i-1}(t) - h_i(t)|^2}{h_{i-1}^2(t)} \quad (7.1)$$

Genellikle, SS (standart sapma) değer aralığı 0,2 ile 0,3 arasındadır. $h_2(t)$ SS'yi karşılırsa, $\dot{IKF}_1(t) = h_2(t)$ olur. Eğer $h_2(t)$ durdurma kriterini karşılamazsa, $h_2(t)$ yeni bir sinyal olarak kabul edilir ve Adım 1-5 boyunca, $h_i(t)$ SS ya da İKF olmanın iki koşulunu sağlayıncaya kadar $h_i(t)$ 'yi oluşturmak için $h_2(t)$ üzerinden işlemler tekrarlanır. Daha sonra, $\dot{IKF}_1(t) = h_i(t)$ olarak hesaplanır.

Adım 6: $IMF_1(t)$ oluşturulur ve artık (residual) $r_1(t)$ 'yi elde etmek için orijinal sinyalden çıkarılır: $r_1(t) = x(t) - \dot{IKF}_1(t)$. Artık sinyal orijinal sinyal olarak ele alınır ve sonraki artık sinyali almak için 1'den 5'e kadar adımlar tekrarlanır. Dolayısıyla artık sinyal, $r_n(t) = r_{n-1}(t) - \dot{IKF}_n(t)$ olarak ifade edilir. Bu noktada, $r_n(t)$ monoton bir dizidir. Eleme işlemi sonrasında, orijinal sinyal $\dot{IKF}_1(t)$, $\dot{IKF}_2(t)$, ..., $\dot{IKF}_n(t)$ gibi birçok İKF bileşenine ve artık bir dizi $r_n(t)$ 'ye ayrıştırılmaktadır. Bu nedenle, orijinal sinyal şu şekilde ifade edilebilir:

$$x(t) = \sum_{i=1}^n \dot{I}KF_i(t) + r_n(t) \quad (7.2)$$

Şekil 7.3’de 5 dB SNR seviyesinde araba gürültüsü eklenmiş konuşma sinyali ve bu sinyalin AKA yöntemi ile ayrıştırılan ilk 5 İKF ve artık bileşeni gösterilmektedir.



Şekil 7.3 AKA ile ayrıştırılan gürültülü konuşma sinyali için ilk 5 İKF ve artık bileşen

7.4.2 Teager Enerji Operatörü

Teager Enerji Operatörünün (TEO) sürekli formu aşağıdaki gibi verilmektedir:

$$\Psi_c[y(t)] = \left(\frac{d}{dt} y(t) \right)^2 - y(t) \frac{d^2}{dt^2} y(t) \quad (7.3)$$

Burada $\Psi_c[.]$ ve $y(t)$ sırasıyla sürekli Teager enerji operatörünü ve sürekli bir sinyali temsil etmektedir. Belirli bir bantta sınırlandırılmış $y[n]$ ayrık sinyali için ayrık zamanlı TEO şu şekilde ifade edilmektedir:

$$\Psi_d(y[n]) = y[n]^2 - y[n+1]y[n-1] \quad (7.4)$$

Ayrık zamanlı Teager Enerji Operatörü, her zaman anında enerji hesaplanması için yalnızca üç örnek gerektirdiğinden neredeyse anlık bir ölçüttür. Çok iyi bir zaman çözünürlüğü sayesinde bir Teager Enerji Operatörü'nün çıkışı bize enerji dalgalanmalarını yakalama ve dolayısıyla sinyal üretmek için gereken enerjinin tahminini verme imkânını sağlamaktadır.

TEO, zamanla sinyalin enerjisinin bir tahminini sağladığından, konuşma veya konuşma dışı etkinlik fikrini elde etmek için kullanılabilir ve daha sonra konuşma ya da konuşma dışı bölgeler için uygun bir eşik değerinin belirlenmesini sağlayabilmektedir. Ancak TEO, gürültülü konuşmada doğrudan kullanıldığında, istenmeyen bozucu etkiler (artifact) ve güçlendirilmiş gürültüler ortaya çıkabilmektedir. Bu nedenle gürültülü konuşma sinyaline doğrudan TEO uygulamak yerine, TEO'nun gürültülü konuşmanın içkin kip fonksiyonları katsayıları üzerine uygulanması eşik değerlerinin daha uygun bulunmasını sağlamaktadır (Khaldi et al., 2014a).

Gürültülü konuşma sinyali AKA yöntemi ile ilk önce İKF'lere ve artık (residual) sinyale ayrıştırılır ve daha sonra TEO'nun her kipe aşağıdaki gibi uygulanması sağlanır:

$$E_{k,m} = \Psi[\text{IKF}_{k,m}(t)] \quad (7.5)$$

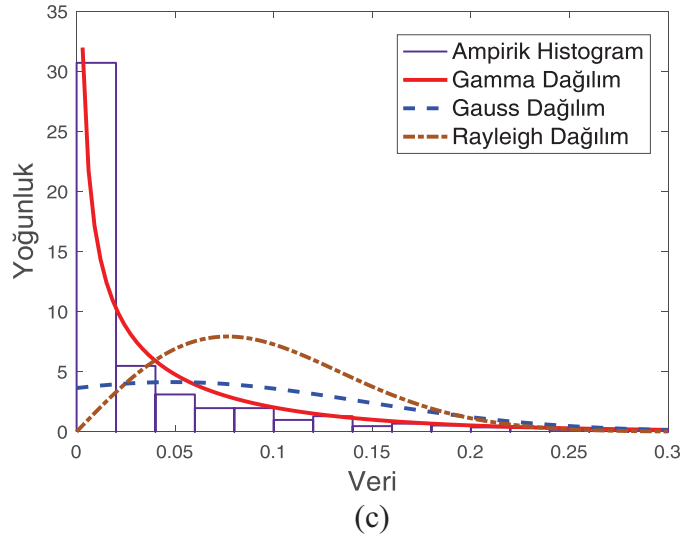
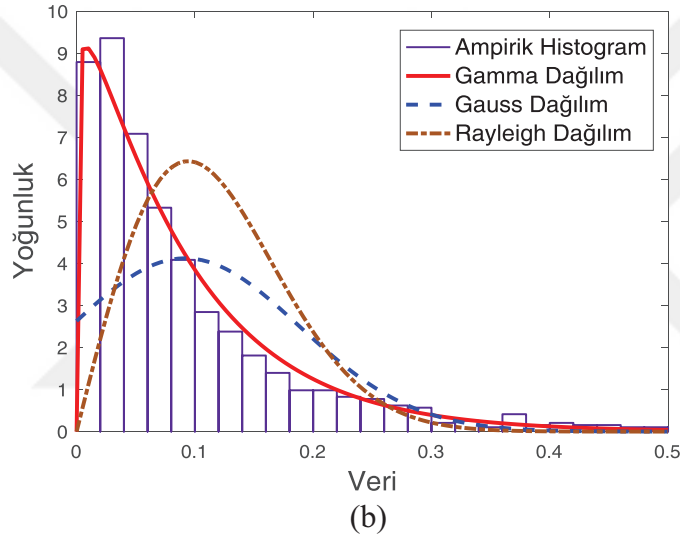
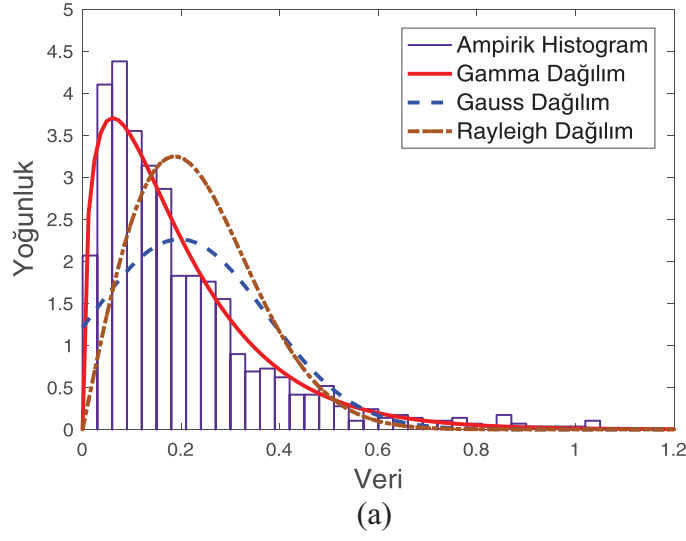
Burada k ve m sırası ile her bir kipi ve çerçeveleri göstermektedir.

7.4.3 TEO Uygulanmış İKF Katsayıları için Önerilen İstatistiksel Model

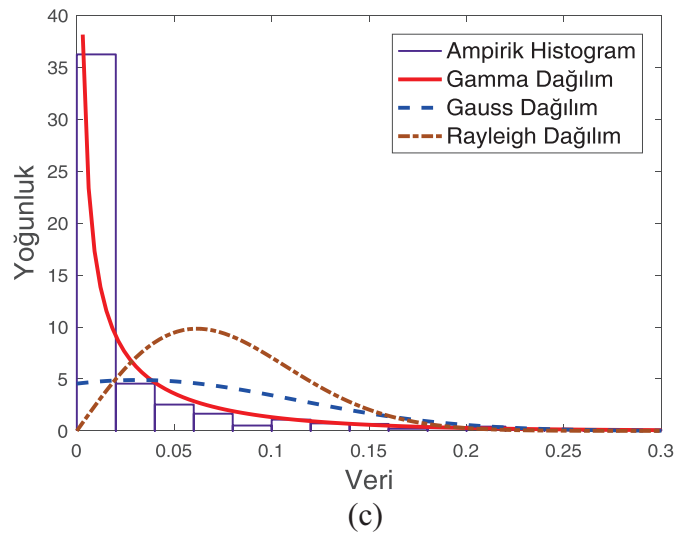
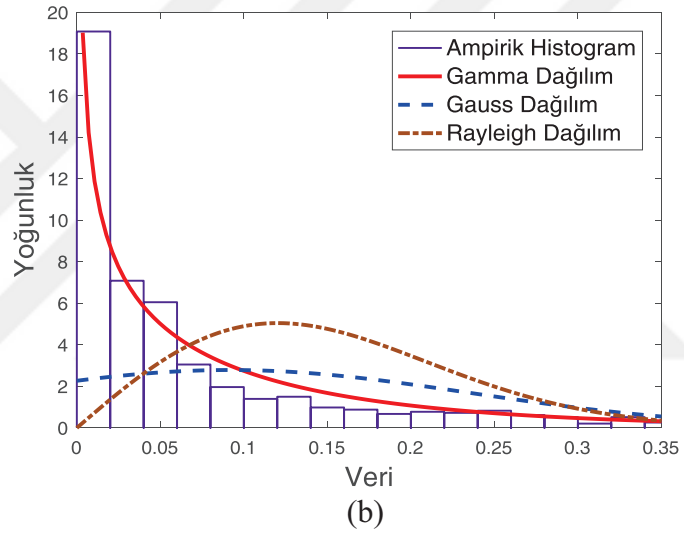
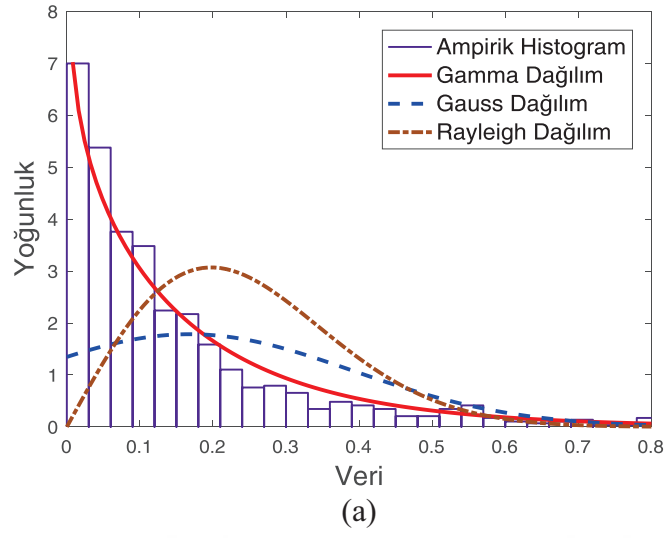
Bir dönüşüm ya da analiz alanında gürültülü konuşma sinyallerinin sesli etkinlik algılama performansı, temel olarak eşik değerlerine bağlı olmaktadır. Gürültülü konuşma sinyallerinden konuşma ve konuşma dışı bölgelerinin belirlenmesinde tüm AKA kipleri için tek bir eşik değeri kullanmak hata payını arttırmaktadır. Bu nedenle, eşik değerleri her bir kip için uyarlanabilir olmalıdır. Gürültülü konuşma sinyalinin $E_k(t)$ enerjisinin olasılık dağılımı göz önünde bulundurulduğunda, uygun bir örüntü eşleme ya da benzerlik ölçütü kullanılarak daha doğru ve net uyarlanabilir eşik değerleri elde edilebilir (Islam et al., 2015).

Konuşma sinyali zamanla değişen bir sinyal olduğundan, konuşmanın ya da enerjisinin doğru olasılık dağılım fonksiyonunu elde etmek zordur. Bu nedenle doğrudan konuşmanın enerjisinin olasılık dağılım fonksiyonunu (probability density function-PDF) çıkarmak yerine enerjinin ampirik histogramı üzerinden uygun olasılık dağılımını çıkarmak daha doğru bir yaklaşım olacaktır. Konuşma sinyalinin enerjisinin ampirik histogramını kapsayacak en uygun olasılık dağılım fonksiyonunun belirlenmesi, eşik değerlerinin hesaplanmasında önemli bir rol oynamaktadır.

Şekil 7.4’de gürültülü konuşmanın -10, 0 ve 10 dB SNR seviyelerinde araba gürültüsü için Teager Enerji Operatörü uygulanmış İKF katsayılarının ampirik histogramı, Gaussian, Rayleigh ve Gamma dağılımları verilmiştir. Benzer şekilde, aynı SNR seviyelerinde mırıltı gürültüsü için kullanılan bu dağılımlar Şekil 7.5’de gösterilmektedir. Şekil 7.4 ve Şekil 7.5’den görülebileceği üzere Gamma dağılımının, gürültülü konuşmanın enerjisinin ampirik histogramını Gaussian ve Rayleigh dağılıma göre daha doğru bir şekilde kapsadığı (fit) görülmektedir. Bu nedenle her kip için uyarlanabilir olan eşik değerlerinin belirlenmesinde Gamma dağılım fonksiyonu önerilmiştir.



Şekil 7.4 Araba gürültüsü koşullarında gürültülü konuşma sinyalinin İKF katsayılarının enerjisinin (a) -10 dB, (b) 0 dB ve (c) 10 dB SNR seviyeleri için ampirik histogram, Gauss, Rayleigh ve Gamma dağılımları



Şekil 7.5 Mırlı gürültüsü koşullarında gürültülü konuşma sinyalinin İKF katsayılarının enerjisinin (a) -10 dB, (b) 0 dB ve (c) 10 dB SNR seviyeleri için ampirik histogram, Gauss, Rayleigh ve Gamma dağılımları

7.4.4 Uyarlanır Eşik Değerlerinin Hesaplanması

Gürültülü konuşmalarda sesli bölgelerin belirlenmesinde, AKA yöntemi ile ayrıştırılan her kipteki İKF katsayılarının Teager enerjisini doğrudan kullanmak yerine bu katsayıların enerjisinin olasılık dağılımı çıkarılarak eşik değerlerinin belirlenmesi önerilmektedir. Gürültünün belirlenmesinde, gürültülü konuşma sinyalinin her bir kipindeki ilk 5 çerçevesi alınmaktadır. Çünkü tüm kiplerin ilk birkaç çerçevesi sesin olmadığı bölgelerdir ve bu çerçevelerde yalnızca gürültü bulunmaktadır. Gürültülü konuşmanın enerjisinin olasılık dağılımının belli bir aralıkta gürültünün enerjisinin olasılık dağılımına benzer olması beklenmektedir. Böylece, gürültülü konuşmanın ve gürültünün enerjisinin olasılık dağılımları göz önüne alınarak olasılık dağılımları arasında uygun bir model eşleştirme şeması veya benzerlik düzeyi kullanılarak daha doğru eşik değerleri elde edilebilir. İki dağılım arasındaki mesafenin belirlenmesinde yaygın olarak kullanılan Kullback-Leibler uzaklık ölçütü kullanılmaktadır. İki dağılım arasındaki farkları doğru bir şekilde tahmin etmek eşik değerlerinin belirlenmesinde oldukça önemlidir.

Kullback-Leibler (K-L) uzaklığı, gürültülü konuşmanın olasılık dağılımının gürültünün olasılık dağılımına yaklaşması durumunda her zaman sıfır değerini almaktadır. K-L uzaklığı simetrik bir ölçüt olmadığından, gürültülü konuşma ve gürültünün Gamma dağılım fonksiyonları arasında simetrik bir uzaklık olması amacıyla Simetrik Kullback-Leibler (SKL) kullanılmıştır. SKL uzaklığı şu şekilde ifade edilebilir:

$$SKL(p, q) = \frac{KL(p, q) + KL(q, p)}{2} \quad (7.6)$$

Burada p ve q her biri N sayıda bin'e sahip ilgili histogramlardan hesaplanan gürültülü konuşma enerjisinin ve gürültü enerjisinin yaklaşık Gamma olasılık dağılım fonksiyonlarını göstermektedir. $KL(p, q)$ şu şekilde hesaplanmaktadır:

$$KL(p, q) = \sum_{i=1}^N p_i(E_{k,m}(t)) \ln \frac{p_i(E_{k,m}(t))}{q_i(E_{k,m}(t))} \quad (7.7)$$

$p_i(E_{k,m}(t))$, gürültülü konuşma sinyalinin Teager enerjisinin yaklaşık Gamma olasılık dağılım fonksiyonu olarak ifade edilir ve şu şekilde hesaplanır:

$$p_i(E_{k,m}(t)) = \frac{n_i}{N_c} \quad (7.8)$$

Burada n_i , i. bin'deki katsayıların sayısını, N_c ise her bir kipteki katsayıların toplam sayısını vermektedir.

Gürültülü konuşma sinyalinin enerjisinin belli bir λ değeri altında, $p_i(E_{k,m}(t))$ ve $q_i(E_{k,m}(t))$ arasındaki Simetrik K-L ayrıştırması yaklaşık olarak sıfır olmaktadır.

$$SKL(p_i(E_{k,m}(t)), q_i(E_{k,m}(t))) \approx 0 \quad (7.9)$$

$p_i(E_{k,m}(t))$ ve $q_i(E_{k,m}(t))$ için genelleştirilmiş Gamma olasılık dağılım fonksiyonu şu şekilde ifade edilebilir (Erkelens et al., 2007; Sun, 2016).

$$p_i(E_{k,m}(t)) = \frac{1}{b^a \Gamma(a)} \left| \frac{x}{\sigma_s} \right|^{a-1} \exp\left(-\frac{1}{b} \left| \frac{x}{\sigma_s} \right| \right) \quad (7.10)$$

$$q_i(E_{k,m}(t)) = \frac{1}{b_n^{a_n} \Gamma(a_n)} \left| \frac{x}{\sigma_n} \right|^{a_n-1} \exp\left(-\frac{1}{b_n} \left| \frac{x}{\sigma_n} \right| \right) \quad (7.11)$$

Burada $\Gamma(.)$ Gamma fonksiyonunu, b ve a sırasıyla gürültülü konuşmanın önsel SNR ile ilişkili olan ölçek parametresini (scale parameter) ve şekil parametresini (shape parameter), b_n ve a_n ise sırasıyla gürültü sinyalinin önsel SNR ile ilişkili olan ölçek parametresini ve genelleştirilmiş Gamma fonksiyonunun şekil parametresini temsil etmektedir. Ayrıca, σ_s ve σ_n sırasıyla gürültülü konuşma ve gürültünün gücünü göstermektedir.

Eşitlik (7.10) ve (7.11) için SKL yöntemi uygulanacak olursa eşik değeri λ şu şekilde hesaplanabilir:

$$\int_0^\lambda \left[\frac{1}{b^a \Gamma(a)} \left(\frac{x}{\sigma_s} \right)^{a-1} \exp\left(-\frac{x}{b\sigma_s}\right) - \frac{1}{b_n^{a_n} \Gamma(a_n)} \left(\frac{x}{\sigma_n} \right)^{a_n-1} \exp\left(-\frac{x}{b_n\sigma_n}\right) \right] I_1 dx = 0 \quad (7.12)$$

$$I_1 = \ln\left(\frac{\sigma_s}{\sigma_n}\right) \exp\left(-\frac{x}{\sigma_s} + \frac{x}{\sigma_n}\right) \quad (7.13)$$

Eşitlik (7.12) çözülerek λ eşik değerleri Eşitlik (7.14) ile türetilir:

$$\lambda(k) = \frac{\sigma_n(k) \sqrt{b^a - b_n^{a_n}} \sqrt{1 + \gamma(k)}}{\ln \sqrt{1 + \gamma(k)}} \quad (7.14)$$

Burada $\gamma(k)$, k. AKA kiplerinin bölütsel SNR değerlerini vermektedir ve şu şekilde ifade edilir:

$$\gamma(k) = \frac{\sigma_s^2(k) - \sigma_n^2(k)}{\sigma_n^2(k)} \quad (7.15)$$

7.5 Sesli Etkinlik Algılama Performans Değerlendirme Ölçütleri

Bu çalışmada önerilen SEA algoritmasının performansını değerlendirmek amacıyla bir dizi gürültülü konuşma sinyali için konuşma (speech) doğruluk oranı $HR1$, konuşma-dışı doğruluk oranı $HR0$ ve genel doğruluk oranı (GDO) ölçütleri kullanılmıştır. İdeal bir Sesli Etkinlik Algılama (SEA) için, konuşma doğruluk oranı, konuşma dışı doğruluk oranı ve toplam doğruluk oranı yüksek olmalıdır. $HR1$ ve $HR0$ yüzdeleri elde etmek amacıyla temiz konuşma sinyallerinin konuşma (voice) ve konuşma dışı (sessizlik veya gürültü) bölgeleri öncelikle manuel olarak etiketlenmektedir. Konuşma doğruluk oranı $HR1$, manuel olarak işaretlenmiş konuşma bölgelerinin SEA algoritması tarafından doğru algılanma yüzdesi olarak hesaplanırken, konuşma-dışı doğruluk oranı $HR0$ ise manuel olarak işaretlenmiş konuşma dışı bölgelerin doğru algılanma yüzdesi olarak hesaplanmaktadır. Konuşma doğruluk oranı (speech hit rate - $HR1$) ve konuşma dışı doğruluk oranı (non-speech hit rate - $HR0$) aşağıdaki şekilde hesaplanır:

$$HR1 = \frac{N_{1,1}}{N_1^{ref}} \quad HR0 = \frac{N_{0,0}}{N_0^{ref}} \quad (7.16)$$

Burada $N_{1,1}$ ve $N_{0,0}$ sırasıyla doğru bir şekilde sınıflandırılan konuşma ve konuşma dışı örnek sayısını gösterirken, N_1^{ref} ve N_0^{ref} ise sırasıyla veritabanından alınan gerçek konuşma ve konuşma-dışı örnek sayılarını göstermektedir. Genel doğruluk oranı (GDO) ise şu şekilde tanımlanmaktadır:

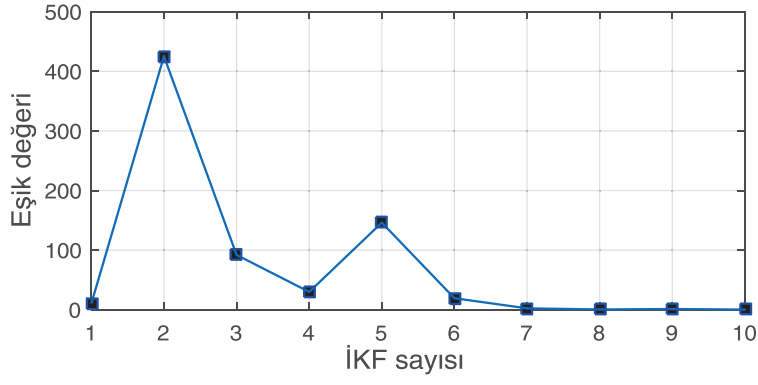
$$GDO = \frac{N_{1,1} + N_{0,0}}{N_1^{ref} + N_0^{ref}} \quad (7.17)$$

SEA algoritmasından en iyi performansı elde etmek için kullanılan bu üç ölçütten elde edilen sonuçların maksimum düzeye çıkarılması gerekmektedir.

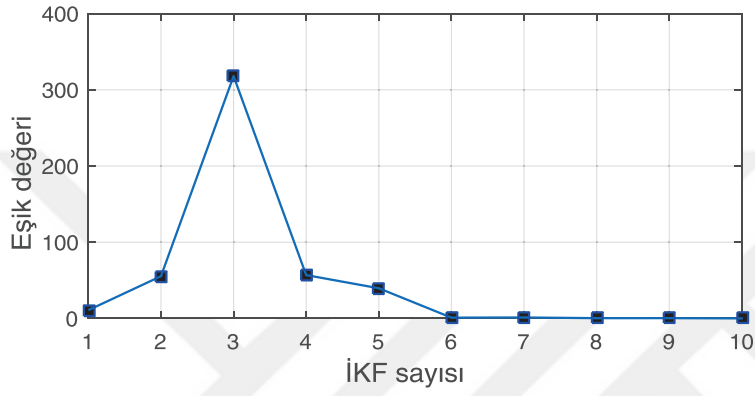
7.6 Uygulama Koşulları ve Deneysel Sonuçlar

Çalışmamızda İngilizce konuşmaların bulunduğu NOIZEUS veri tabanından ve Türkçe konuşmaların bulunduğu ODTÜ veri tabanından alınan 30 temiz konuşma sinyaline -10 ile 10 dB arasında değişen beş farklı SNR seviyelerinde beyaz, araba, mırıltı, sergi, hava alanı, sokak ve tren gürültüsü eklenerek gürültülü sinyaller elde edilmiştir. Elde edilen bu gürültülü konuşma sinyalleri AKA yöntemi ile kiplere ayrıştırılmış ve her kipte bulunan İKF katsayılarının çerçeveleme (framing) ile Teager enerjisi hesaplanmıştır. Çerçeveleme işleminde 32 ms örtüşmeli 64 ms pencere uzunluğu olan Hamming pencere işlevi kullanılmıştır. Çerçeveleme tüm kipler için Gamma olasılık işlevi uygulanarak her bir kip için bu fonksiyona ait b ölçek ve a şekil parametreleri kestirilmiştir. Bu işlemin her kip için tekrarlanması ile uyarlanabilir eşik değerleri hesaplanmıştır. Önerilen yöntemin performansını değerlendirmek amacıyla konuşma bölgelerinin doğru algılanma oranı $HR1$, konuşma dışı bölgelerin doğru algılanma oranı $HR0$ ve sistemin genel doğruluk oranı GDO hesaplanarak literatürde sıklıkla kullanılan AMR1, AMR2 ve G.729B gibi standart SEA yöntemleri ile SOHN ve LTSD SEA yöntemleri ile karşılaştırılmıştır.

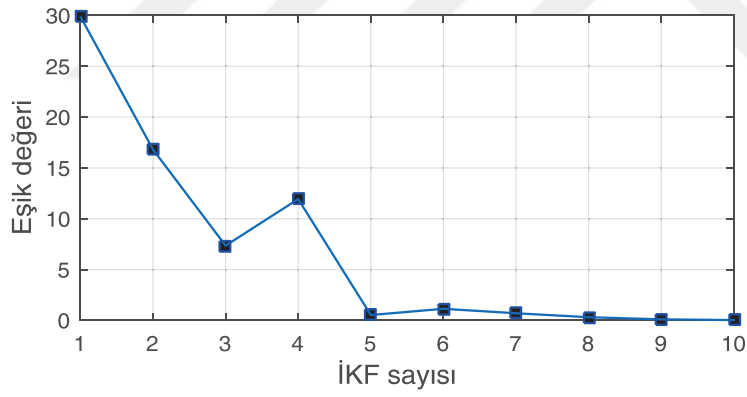
Şekil 7.6 ve Şekil 7.7’de sırasıyla araba ve mırıltı gürültüsü ile bozulan konuşma sinyalinin ilk 10 kipi için önerilen yöntemle bulunan uyarlanır eşik değerleri verilmektedir. Eşik değerleri; gürültü tipi, SNR düzeyi ve her kip için farklılık göstermektedir. Ayrıca, 10. kipten sonra gürültü azaldığından dolayı eşik değerleri düşmektedir ve neredeyse tüm bileşenler eşikten geçmektedir. Bu nedenle, sesli bölgelerin belirlenmesinde AKA yöntemi ile ayrıştırılan kiplerin ilk 10 tanesi kullanılmıştır.



(a)

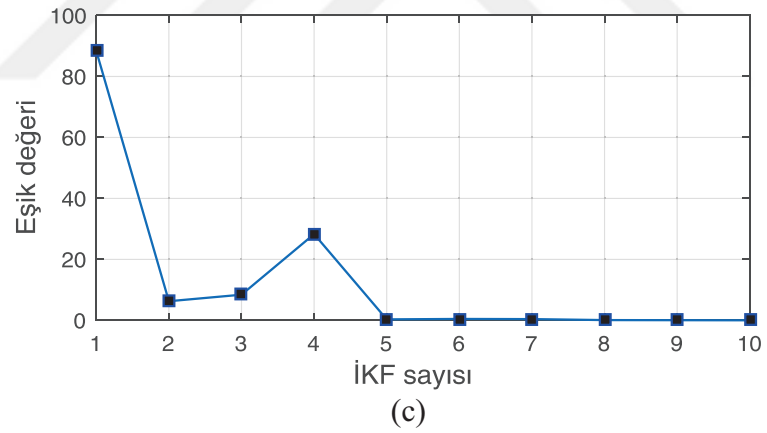
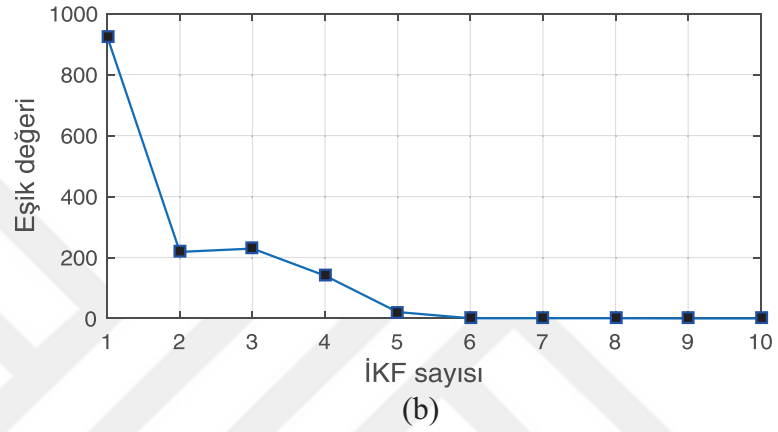
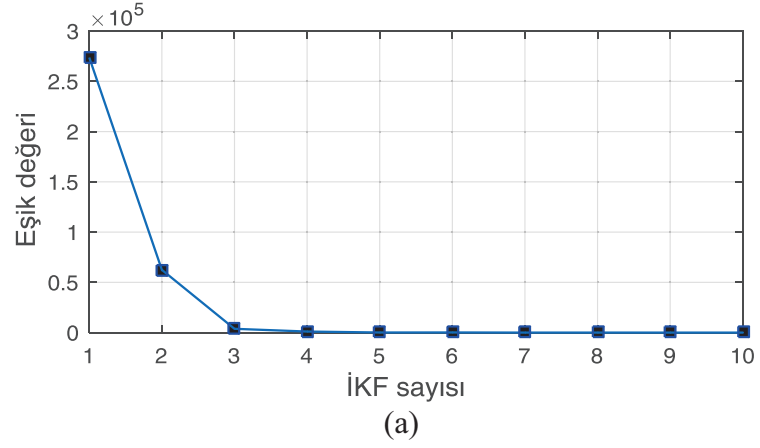


(b)



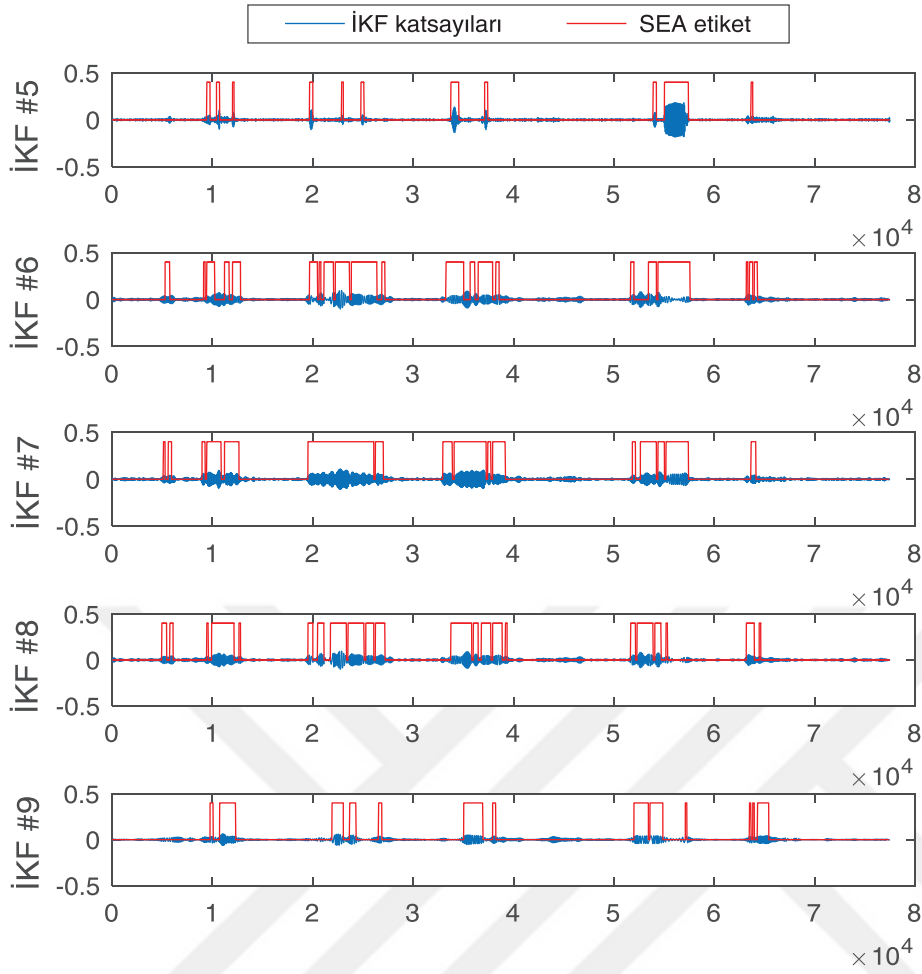
(c)

Şekil 7.6 Araba gürültüsü koşullarında (a) -10 dB, (b) 0 dB ve (c) 10 dB SNR seviyeleri için uyarlanırlar eşik değerleri



Şekil 7.7 Mırlı gürültü koşullarında (a) -10 dB, (b) 0 dB ve (c) 10 dB SNR seviyeleri için uyarlanır eşik değerleri

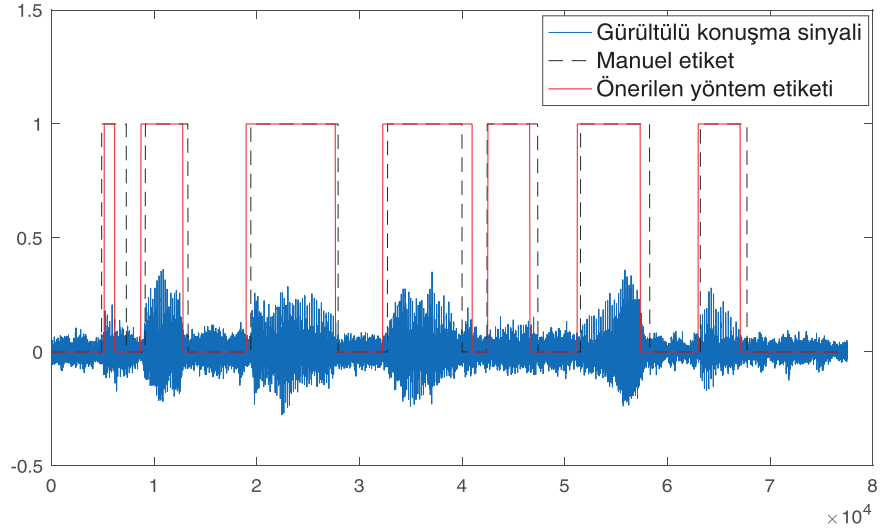
Farklı gürültü tipi ve SNR seviyesi koşullarında her kip için uyarlanabilir eşik değerleri belirlendikten sonra ilgili kipteki İKF katsayılarından elde edilen Teager enerji değerleri bu eşik değerlerinden geçirilmektedir. Şekil 7.8'de gürültülü bir konuşma sinyaline ait her bir kipi enerjisinin eşikten geçirilerek konuşma ve konuşma dışı bölgeler olarak etiketlenmesi gösterilmektedir.



Şekil 7.8 Her kip için eşikten geçirilerek etiketlenen sesli bölgeler

Her kip için etiketleme işlemi yapıldıktan sonra tüm bu kiplere ait etiketler bir ‘veya (or)’ işlemi ile birleştirilmektedir. Birleştirme işleminden sonra, etiketler arasında ses olamayacak kadar kısa etiketler oluşmaktadır. Bu kısa süreli etiketleri uygun etiketle birleştirmek için bir Hangover işlemi uygulanmıştır. Bu işlemde, bir sessiz bölgesinin en az 1000 örnek (12,5 ms) olması gerektiği düşünülerek bu değer altındaki etiketler sesli bölge olarak değiştirilmektedir. Böylece konuşma bölgeleri arada herhangi bir kayıp olmadan sesli bölge olarak etiketlenebilmektedir.

Şekil 7.9’da mırıltı gürültüsü eklenmiş 5 dB SNR’a sahip bir gürültülü konuşma sinyalinin (‘The birch canoe slid on the smooth planks’) konuşma ve konuşma dışı bölgeler olarak etiketlenmesi gösterilmektedir.



Şekil 7.9 Mırıltı gürültüsü koşullarında 5 dB SNR için önerilen yöntemin SEA sonucu

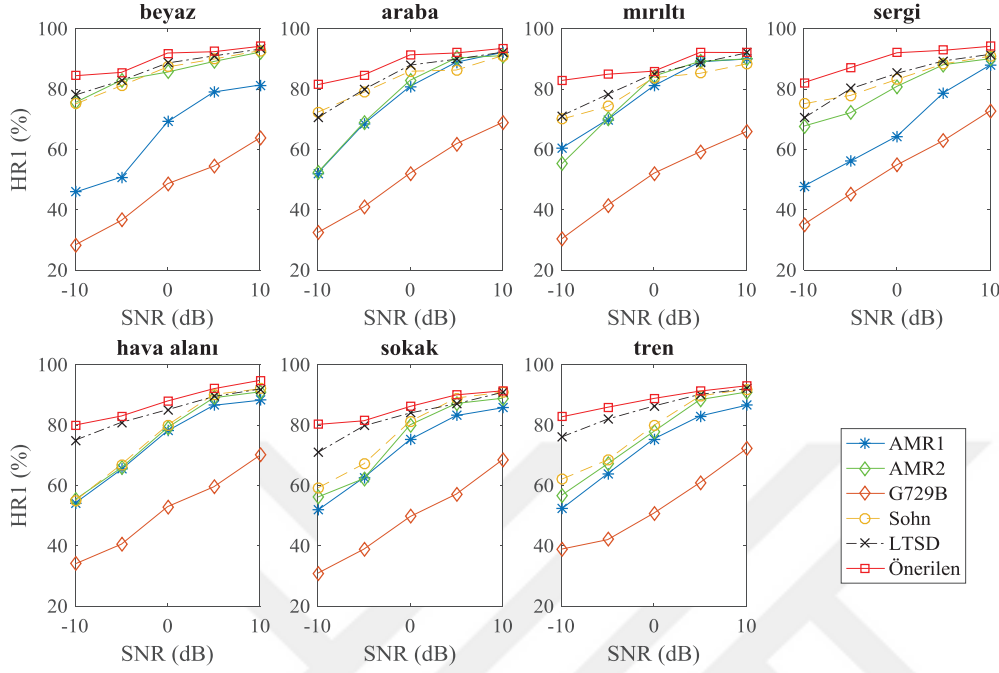
7.6.1 İngilizce için Sesli Etkinlik Algılama Yöntemlerinin Performans Değerlendirmesi

NOIZEUS veri tabanındaki İngilizce cümleler kullanılarak yapılan konuşma ve konuşma dışı bölgeleri algılama çalışmasında, önerilen AKA tabanlı SEA yöntemi, AMR1, AMR2 ve G.729B gibi üç standart SEA yöntemi ve son yıllarda sıklıkla kullanılan SOHN ve LTSD SEA yöntemleri ile HR1, HR0 ve genel doğruluk oranı bakımından karşılaştırılmıştır. Kullanılan yöntemlerin başarısı, beyaz, araba, mırıltı, hava alanı, sergi, sokak ve tren gibi yedi farklı gürültü tipi ve -10, -5, 0, 5 ve 10 dB SNR seviyeleri için test edilmiştir. Kullanılan bu yöntemlerin konuşma ve konuşma dışı bölgelerin belirlenmesinde sağladıkları üstünlükler ve sakıncalar gürültü tipi ve SNR seviyesine göre ayrıntılı olarak açıklanmıştır.

Yedi gürültü ve beş SNR seviyesi için başarı oranları

Önerilen AKA tabanlı SEA yöntemi, yedi farklı gürültü tipi ve -10 ile 10 dB arasında değişen beş farklı SNR seviyesi için AMR1, AMR2, G729B, SOHN ve LTSD SEA yöntemleri ile konuşma doğruluk oranı (HR1), konuşma-dışı doğruluk oranı (HR0) ve genel doğruluk oranı (GDO) bakımından karşılaştırılmıştır. Her gürültü tipi ve her SNR seviyesi için hesaplanan HR0, HR1 ve genel doğruluk oranı sonuçları sırasıyla Şekil 7.10, Şekil 7.11 ve Şekil 7.12’de verilmiştir. Şekil 7.10’dan görülebileceği üzere, HR1 sonuçlarına göre, önerilen AKA tabanlı SEA

yöntemi tüm gürültü tipi ve SNR seviyeleri için en iyi konuşma doğruluk oranını vermektedir.

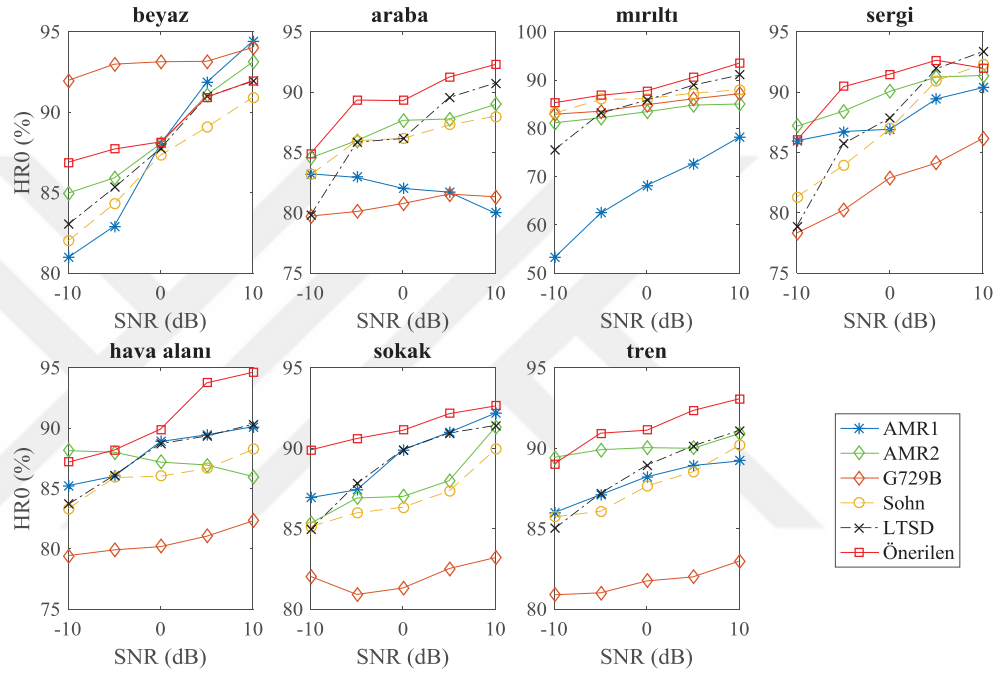


Şekil 7.10 Tüm gürültü ve SNR seviyeleri için SEA yöntemlerine ait HR1 sonuçları

Gürültü tipi ve SNR seviyesine göre önerilen yöntemin HR1 sonuçları şu şekilde bulunmaktadır; Beyaz gürültü için -10 dB'de % 81,46 ve 10 dB'de % 94,18; araba için -10 dB'de % 81,46 ve 10 dB'de % 93,48; mırıltı için -10 dB'de % 82,87 ve 10 dB'de % 92,12; sergi için -10 dB'de % 82,16 ve 10 dB'de % 94,18; hava alanı için -10 dB'de % 80,02 ve 10 dB'de % 94,81; sokak için -10 dB'de % 80,22 ve 10 dB'de % 91,35; tren için -10 dB'de % 82,72 ve 10 dB'de % 93,01. Önerilen yöntemden sonra, LTSD metodu her türlü gürültü tipi için en iyi HR1 sonuçlarını vermiştir. G.729B SEA yöntemi ise HR1 sonuçları bakımından en kötü sonuçları vermiştir. Bu nedenle, gürültülü koşullar altında önerilen SEA yöntemi, konuşma bölgelerinin belirlenmesinde diğer yöntemlere göre daha başarılı sonuçlar vermektedir.

Yedi farklı gürültü tipi ve beş farklı SNR seviyesi için tüm yöntemlere ait HR0 sonuçları Şekil 7.11'de gösterilmektedir. Şekil 7.11'deki HR0 sonuçlarına göre, önerilen yöntem beyaz gürültü dışında diğer tüm gürültülerde en iyi sonuçları vermektedir (araba için -10 dB'de % 84,92 ve 10 dB'de % 92,30; mırıltı için -10 dB'de % 85,32 ve 10 dB'de % 93,50; sergi için -10 dB'de % 86,82 ve 10 dB'de % 91,99; hava alanı için -10 dB'de % 87,18 ve 10 dB'de % 94,63; sokak

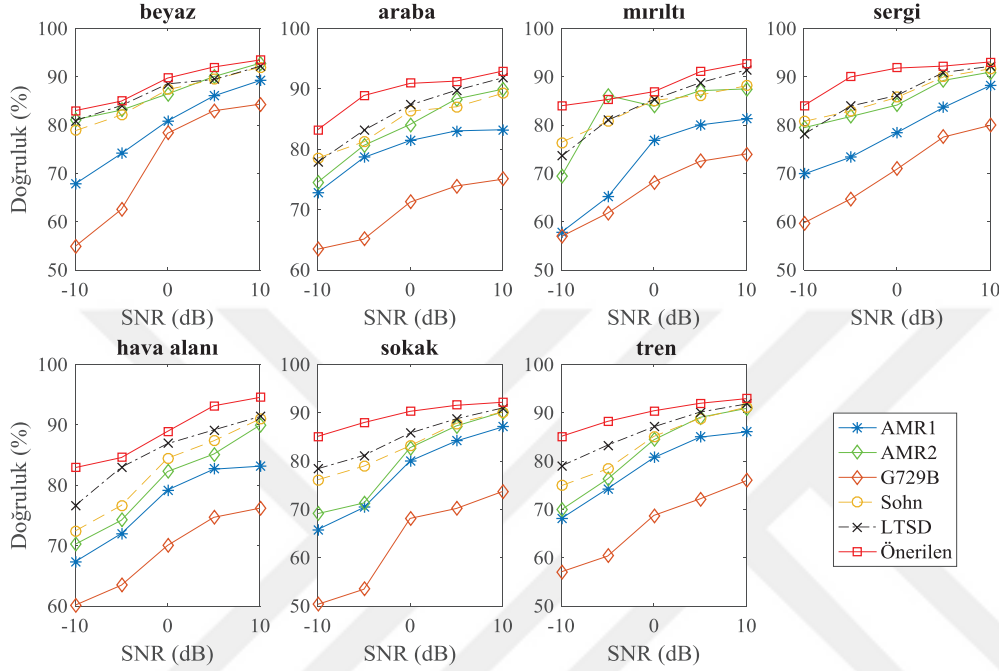
için -10 dB'de % 89,90 ve 10 dB'de % 92,63; tren için -10 dB'de % 89,04 ve 10 dB'de % 93,07). Bununla birlikte, beyaz gürültü koşullarında önerilen yöntemin HR0 sonuçları (-10 dB'de % 86,90 ve 10 dB'de % 91,97) tüm SNR seviyeleri için G.729B SEA (-10 dB'de % 91,98 ve 10 dB'de % 94,03) yöntemine göre daha düşük bulunmaktadır. Ayrıca, yüksek SNR seviyelerinde beyaz gürültü koşullarında önerilen yöntemin AMR1 (-10 dB'de % 80,98 ve 10 dB'de % 94,43) ve AMR2 (-10 dB'de % 84,98 ve 10 dB'de % 93,13) yöntemlerinden daha düşük HR0 değerlerine sahip olduğu anlaşılmaktadır.



Şekil 7.11 Tüm gürültü ve SNR seviyeleri için SEA yöntemlerine ait HR0 sonuçları

Tüm gürültü tipleri ve beş farklı SNR seviyesi koşullarında kullanılan yöntemlere ait genel doğruluk oranı (GDO) sonuçları Şekil 7.12'de gösterilmektedir. Şekil 7.12'den görülebileceği üzere, önerilen AKA tabanlı SEA yöntemi tüm gürültü ve SNR'lar için en iyi genel doğruluk oranı sonuçlarını vermektedir. Önerilen yöntem için gürültü tipine göre genel doğruluk oranı sonuçları şu şekilde bulunmuştur: beyaz için -10 dB'de % 83,01 ve 10 dB'de % 93,49; araba için -10 dB'de % 83,24 ve 10 dB'de % 92,95; mırıltı için -10 dB'de % 84,02 ve 10 dB'de % 92,85; sergi için -10 dB'de % 84,06 ve 10 dB'de % 93,09; hava alanı için -10 dB'de % 82,93 ve 10 dB'de % 94,56; sokak için -10 dB'de % 85,13 ve 10 dB'de % 92,19; tren için -10 dB'de % 85,15 ve 10 dB'de % 92,94.

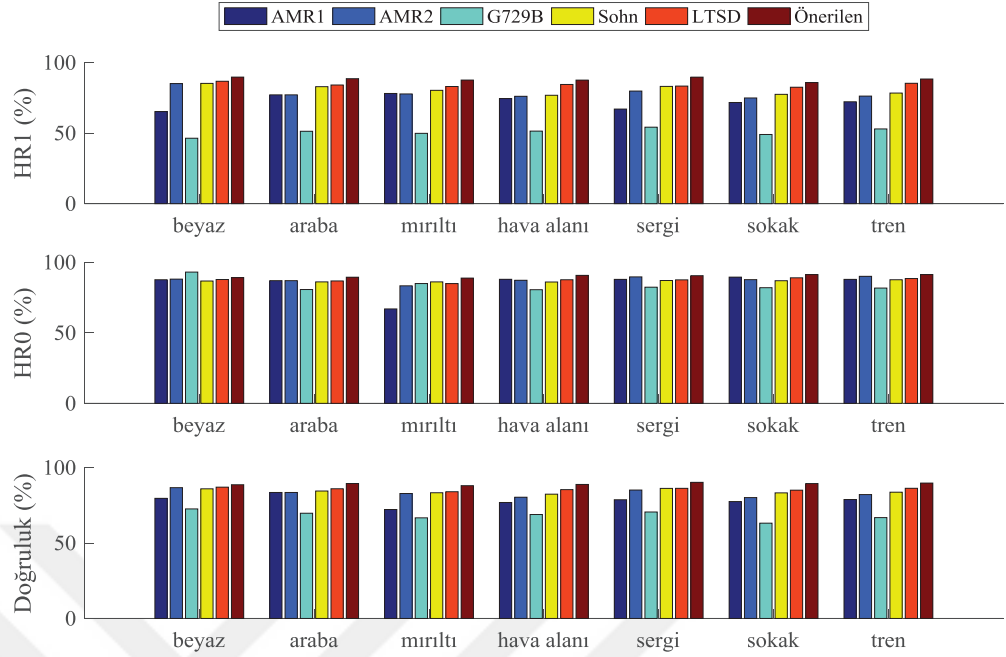
Elde edilen tüm sonuçlar göz önünde bulundurulduğunda, önerilen yöntemin artan SNR seviyeleri ile sürekli bir gelişme gösterdiği söylenebilir. Ayrıca, G.729B SEA yöntemi tüm gürültü tipi ve tüm SNR seviyelerinde en düşük genel doğruluk oranı değerlerini vermektedir.



Şekil 7.12 Tüm gürültü ve SNR seviyeleri için SEA yöntemlerine ait genel doğruluk oranı (GDO) sonuçları

Beş SNR seviyesi için ortalama başarı oranları

AMR1, AMR2, G729B, SOHN, LTSD ve önerilen AKA tabanlı SEA algoritmaları için yedi farklı gürültü tipi için beş SNR seviyesi (-10, -5, 0, 5 ve 10 dB) üzerinden ortalama alınarak elde edilen HR1, HR0 ve genel doğruluk oranı sonuçları Şekil 7.13’de gösterilmektedir. Şekil 7.13’de görülebileceği gibi, önerilen AKA tabanlı SEA yönteminin diğer SEA yöntemlerine kıyasla tüm gürültü tipleri için en iyi HR1 değerlerini sağladığı açık bir şekilde görülmektedir (beyaz gürültüde % 89,69; araba gürültüsünde % 88,59; mırıltı gürültüsünde % 87,60; hava alanı gürültüsünde %87,57; sergi gürültüsünde % 89,67; sokak gürültüsünde % 85,83 ve tren gürültüsü koşullarında % 88,34 HR1 sonuçları). Ayrıca, G.729B SEA yöntemi tüm gürültü çeşitlerinde en düşük HR1 değerlerine sahiptir.



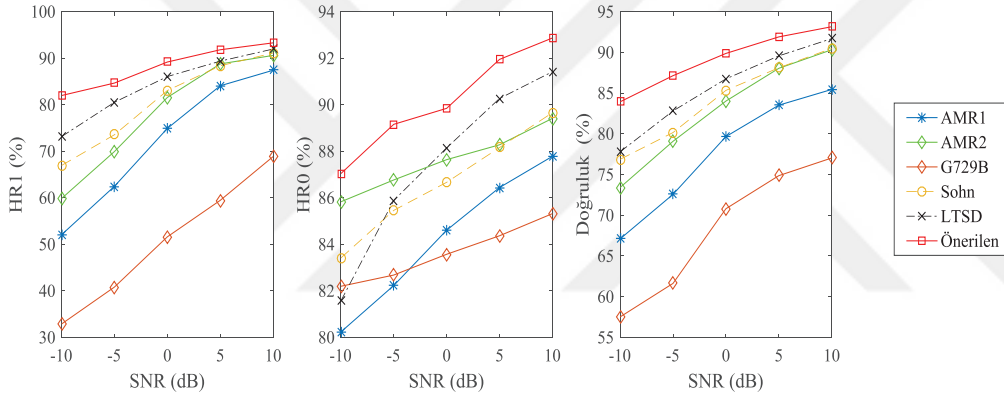
Şekil 7.13 Tüm gürültü için SNR seviyelerinin ortalaması alınarak bulunan HR1, HR0 ve Genel Doğruluk yüzdeleri

Benzer şekilde, önerilen yöntem beyaz gürültü (% 88,13 HR0 değeri) koşulları dışında tüm gürültü çeşitlerinde (araba % 89,13; mırıltı % 88,80; hava alanı % 90,73; sergi %90,52; sokak %91.28; tren % 91,30) en iyi HR0 sonuçlarını vermektedir.

Tüm SEA yöntemlerinin genel doğruluk oranı sonuçlarına göre değerlendirilmesinde, önerilen AKA tabanlı SEA yönteminin tüm gürültü çeşitleri için en iyi sonuçları verdiği görülmektedir. Gürültü çeşitlerine göre önerilen yöntemin genel doğruluk oranı dağılımı yüzdesel olarak şu şekilde hesaplanmıştır: Beyaz gürültüde % 88,65; araba gürültüsünde % 89,46; mırıltı gürültüsünde % 88.03; hava alanı gürültüsünde % 88,83; sergi gürültüsünde % 90,23; sokak gürültüsünde % 89,43 ve tren gürültüsü koşullarında % 89,74. Ayrıca, G.729B SEA yönteminin tüm gürültü çeşitlerinde diğer yöntemlere göre daha düşük genel doğruluk oranlarına sahip olduğu Şekil 7.13’de görülmektedir.

Yedi gürültü tipi için ortalama başarı oranları

Şekil 7.14’de, önerilen AKA tabanlı SEA yönteminin AMR1, AMR2, G.729B gibi üç standart SEA yöntemi ve son yıllarda sıklıkla kullanılan SOHN ve LTSD gibi SEA yöntemleri ile beş farklı SNR seviyesi için yedi gürültü tipinin ortalaması alınarak elde edilen HR1, HR0 ve genel doğruluk oranı sonuçları verilmektedir. HR1, HR0 ve genel doğruluk oranı tüm gürültülerin ortalamasından elde edilerek gürültü tipinden bağımsız olarak önerilen yöntemin başarı performansı ölçülmüştür. Şekil 7.14’den görülebileceği üzere, G.729B SEA yöntemi tüm SNR seviyelerinde en düşük HR1 (-10 dB’de % 32,88 ve 10 dB’de % 68,86) ve genel doğruluk oranı (-10 dB’de % 57,51 ve 10 dB’de % 77,03) sonuçlarına sahipken yüksek SNR seviyelerinde ise en düşük HR0 (0 dB’de % 83,57 ve 10 dB’de % 85,32) sonuçlarını vermektedir.



Şekil 7.14 Her bir SNR seviyesinde tüm gürültü tiplerinin ortalama alınarak bulunan HR1, HR0 ve genel doğruluk yüzdeleri

Elde edilen sonuçlara göre, AMR2 yöntemi HR1, HR0 ve genel doğruluk oranı bakımından AMR1 ve G.729B yöntemlerine göre daha iyi bir başarı performansı göstermektedir. Ayrıca, LTSD yöntemi tüm SNR seviyeleri için SOHN yönteminden daha iyi HR1 ve genel doğruluk oranı sonuçları vermektedir.

Önerilen AKA tabanlı SEA yöntemi tüm SNR seviyeleri için en iyi HR1 (-10 dB’de % 81,98 ve 10 dB’de % 93,30), HR0 (-10 dB’de % 87,04 ve 10 dB’de % 92,87) ve genel doğruluk oranı (-10 dB’de % 83,93 ve 10 dB’de % 93,15) sonuçlarını vermektedir. Ayrıca, artan SNR seviyeleri ile birlikte tutarlı bir ilerleme ve artış göstermektedir. Bu nedenle, önerilen yöntemin tüm gürültü çeşitlerinin ortalaması alındığında, tüm SNR seviyeleri için konuşma ve konuşma dışı bölgeleri tanımlamada iyi bir performans gösterdiği söylenebilir.

Önerilen ve diğer SEA yöntemleri, -10 ile 10 dB arasında değişen beş farklı SNR seviyesi ve yedi farklı gürültü tipi için ortalama alınarak HR1, HR0 ve genel doğruluk oranı açısından karşılaştırılmış ve bu sonuçlar Çizelge 7.1’de özetlenmiştir. Çizelge 7.1’den görülebileceği üzere, gürültü ve SNR seviyelerinin ortalaması alındığında önerilen yöntem en iyi genel doğruluk oranı (% 89,20), en iyi HR1 (% 88,19) ve en iyi HR0 (% 90,17) sonuçlarını sağlamaktadır. Önerilen yöntemi konuşma ve konuşma dışı bölgelerin belirlenmesi başarısı açısından LTSD SEA yöntemi izlemektedir. Ayrıca, gürültü tipi ve SNR seviyesi ortalamalarına göre en kötü sonuçları G.729B SEA yöntemi vermektedir.

Çizelge 7.1 İngilizce konuşmalar için yedi farklı gürültü tipi ve beş farklı SNR seviyesinin ortalaması alınarak yöntemlere ait ortalama başarı yüzdeleri

	SEA Yöntemi					
	AMR1	AMR2	G.729B	SOHN	LTSD	Önerilen
HR0 (%)	84,97	87,58	83,62	86,67	87,45	90,17
HR1 (%)	72,27	78,15	50,73	80,61	84,22	88,19
Doğruluk (%)	78,18	82,97	68,37	84,19	85,72	89,20

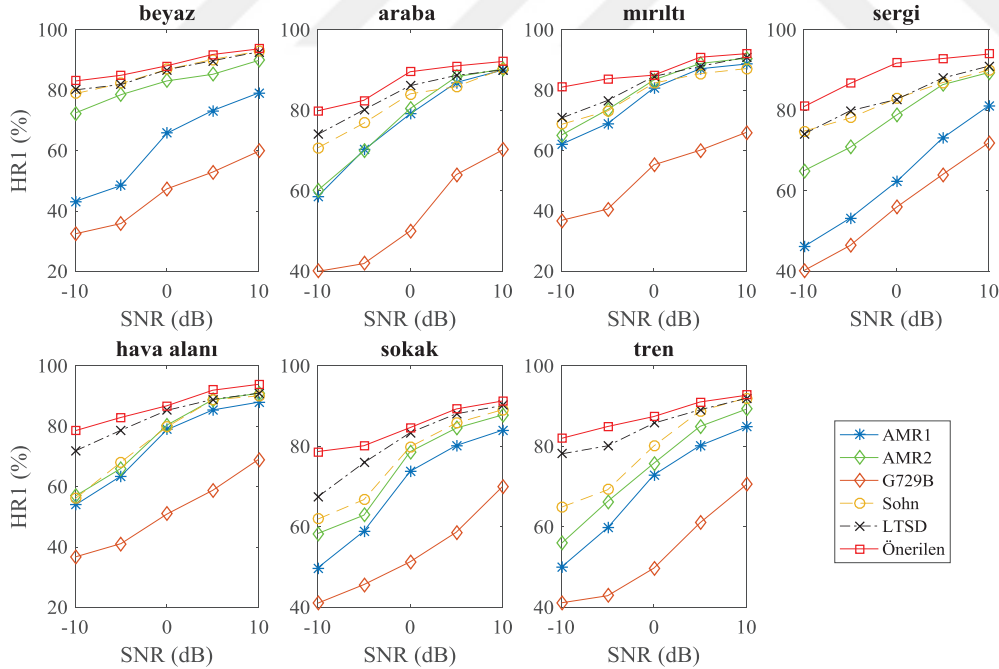
7.6.2 Türkçe için Sesli Etkinlik Algılama Yöntemlerinin Performans Değerlendirmesi

Middle East Technical University Turkish Microphone Speech v1.0 (ODTÜ) veri tabanındaki Türkçe cümleler kullanılarak yapılan konuşma ve konuşma dışı bölgeleri algılama çalışmasında, önerilen AKA tabanlı SEA yöntemi, AMR1, AMR2 ve G.729B gibi üç standart SEA yöntemi ve son yıllarda sıklıkla kullanılan SOHN ve LTSD SEA yöntemleri ile HR1, HR0 ve genel doğruluk oranı bakımından karşılaştırılmıştır. Kullanılan yöntemlerin başarısı, beyaz, araba, mırıltı, hava alanı, sergi, sokak ve tren gibi yedi farklı gürültü tipi ve -10, -5, 0, 5 ve 10 dB SNR seviyeleri için test edilmiştir. Kullanılan bu yöntemlerin konuşma ve konuşma dışı bölgelerin belirlenmesinde sağladıkları üstünlük ve sakıncalar gürültü tipi ve SNR seviyesine göre ayrıntılı olarak açıklanmıştır.

Yedi gürültü ve beş SNR seviyesi için başarı oranları

Önerilen AKA tabanlı SEA yöntemi, yedi farklı gürültü tipi ve -10 ile 10 arasında değişen beş farklı SNR seviyesi için AMR1, AMR2, G729B, SOHN ve LTSD SEA yöntemleri ile konuşma doğruluk oranı (HR1), konuşma-dışı doğruluk oranı (HR0) ve genel doğruluk oranı (GDO) bakımından karşılaştırılmıştır. Her gürültü tipi ve her SNR seviyesi için hesaplanan HR0, HR1 ve genel doğruluk oranı sonuçları sırasıyla Şekil 7.15, Şekil 7.16 ve Şekil 7.17’de verilmiştir. Şekil 7.15’den görülebileceği üzere, HR1 sonuçlarına göre, önerilen AKA tabanlı SEA yöntemi tüm gürültü tipi ve SNR seviyeleri için en iyi konuşma doğruluk oranını vermektedir.

Gürültü tipi ve SNR seviyesine göre HR1 sonuçları şu şekilde bulunmaktadır; Beyaz gürültü için -10 dB’de % 83,09 ve 10 dB’de % 93,72; araba için -10 dB’de % 79,91 ve 10 dB’de % 92,16; mırıltı için -10 dB’de % 81,04 ve 10 dB’de % 92,14; sergi için -10 dB’de % 80,96 ve 10 dB’de % 94,01; hava alanı için -10 dB’de % 78,54 ve 10 dB’de % 93,91; sokak için -10 dB’de % 78,73 ve 10 dB’de % 91,29; tren için -10 dB’de % 82,01 ve 10 dB’de % 92,71.

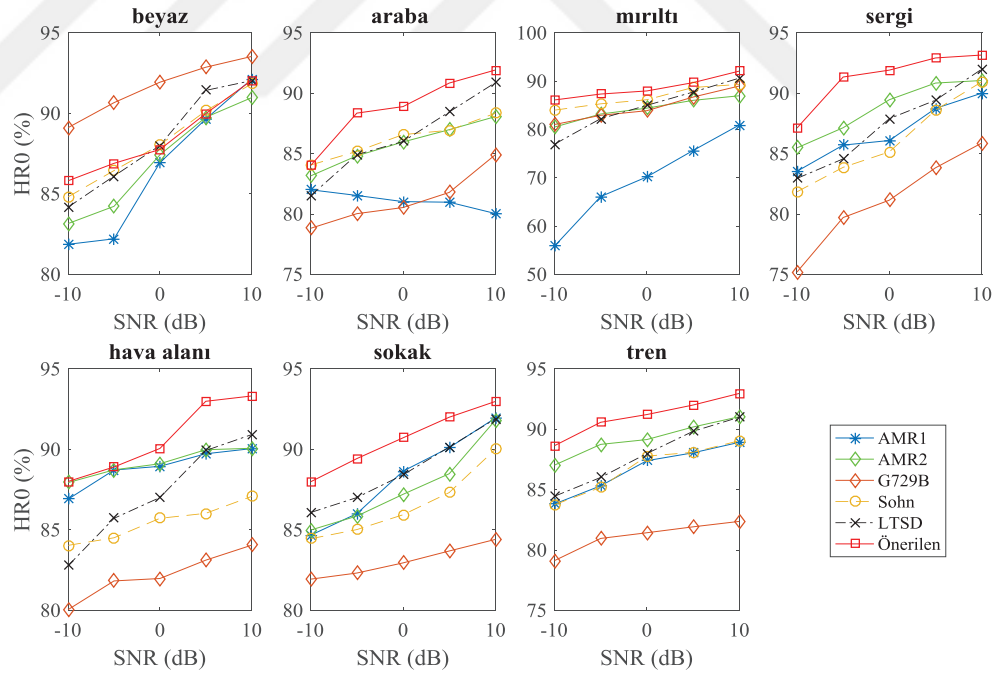


Şekil 7.15 Tüm gürültü ve SNR seviyeleri için yöntemlere ait HR1 değerleri

Önerilen yöntemden sonra, LTSD metodu her türlü gürültü tipi için en iyi HR1 sonuçlarını vermiştir. G.729B yöntemi ise HR1 sonuçları bakımından en kötü sonuçları vermiştir. Bu nedenle, gürültülü koşullar altında önerilen SEA

yöntemi, konuşma bölgesinin belirlenmesinde diğer yöntemlere göre daha başarılı sonuçlar vermektedir.

Yedi farklı gürültü tipi ve beş farklı SNR seviyesi için tüm yöntemlere ait HR0 sonuçları Şekil 7.16'da gösterilmektedir. Şekil 7.16'daki HR0 sonuçlarına göre, önerilen yöntem beyaz gürültü dışında diğer tüm gürültülerde en iyi sonuçları vermektedir (araba için -10 dB'de % 84,02 ve 10 dB'de % 91,93; mırıltı için -10 dB'de % 86,12 ve 10 dB'de % 92,10; sergi için -10 dB'de % 87,12 ve 10 dB'de % 93,17; hava alanı için -10 dB'de % 88,01 ve 10 dB'de % 93,32; sokak için -10 dB'de % 87,89 ve 10 dB'de % 92,98; tren için -10 dB'de % 88,64 ve 10 dB'de % 92,97). Bununla birlikte, beyaz gürültü koşullarında önerilen yöntemin HR0 sonuçları (-10 dB'de % 85,82 ve 10 dB'de % 92,05) tüm SNR seviyeleri için G.729B SEA (-10 dB'de % 89,12 ve 10 dB'de % 93,54) yöntemine göre daha düşük bulunmaktadır. Ayrıca, yüksek SNR seviyelerinde beyaz gürültü koşullarında önerilen yöntemin LTSD SEA (5 dB'de % 91,43 ve 10 dB'de % 92,04) yönteminden daha düşük ya da yakın HR0 değerlerine sahip olduğu anlaşılmaktadır.

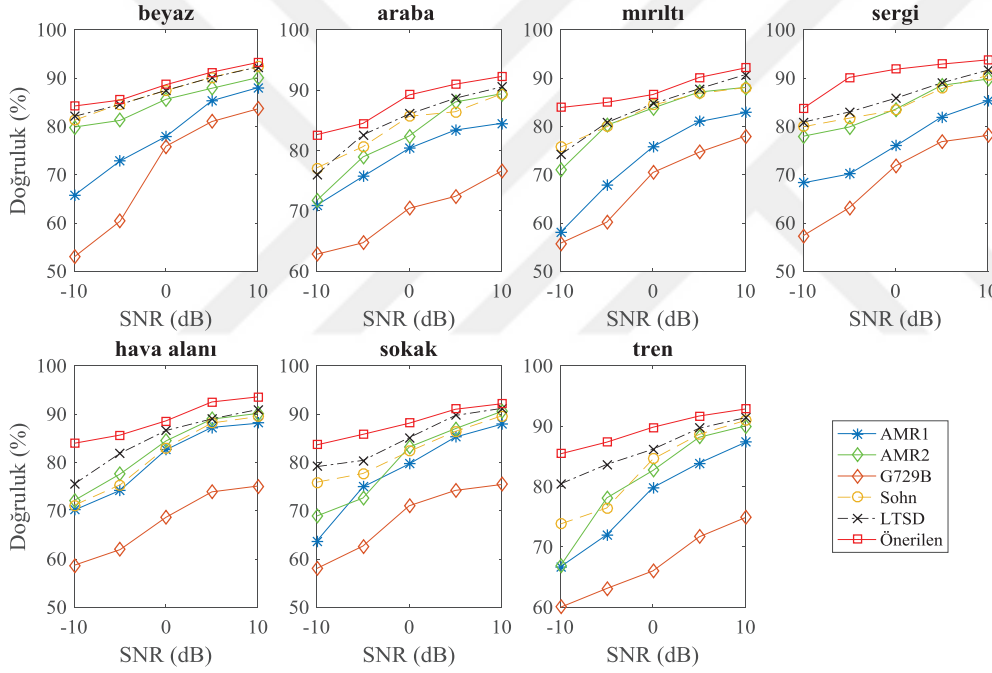


Şekil 7.16 Tüm gürültü ve SNR seviyeleri için yöntemlere ait HR0 değerleri

Tüm gürültü tipi ve beş farklı SNR seviyesi koşullarında kullanılan yöntemlere ait genel doğruluk oranı (GDO) sonuçları Şekil 7.17'de gösterilmektedir. Şekil 7.17'den görülebileceği üzere, önerilen AKA tabanlı SEA

yöntemi tüm gürültü ve SNR'lar için en iyi genel doğruluk oranı sonuçlarını vermektedir. Önerilen yöntem için gürültü tipine göre genel doğruluk oranı sonuçları şu şekilde bulunmuştur; beyaz için -10 dB'de % 84,23 ve 10 dB'de % 93,25; araba için -10 dB'de % 82,64 ve 10 dB'de % 92,28; mırıltı için -10 dB'de % 83,92 ve 10 dB 'de % 92,12; sergi için -10 dB'de % 83,73 ve 10 dB'de % 93,79; hava alanı için -10 dB'de % 83,97 ve 10 dB'de % 93,57; sokak için -10 dB'de % 83,63 ve 10 dB'de % 92,15; tren için -10 dB'de % 85,43 ve 10 dB'de % 92,85.

Elde edilen tüm sonuçlar göz önünde bulundurulduğunda, önerilen yöntemin artan SNR seviyeleri ile sürekli bir gelişme gösterdiği söylenebilir. Ayrıca, G.729B SEA yöntemi tüm gürültü tipi ve tüm SNR seviyelerinde en düşük genel doğruluk oranı değerlerini vermektedir.



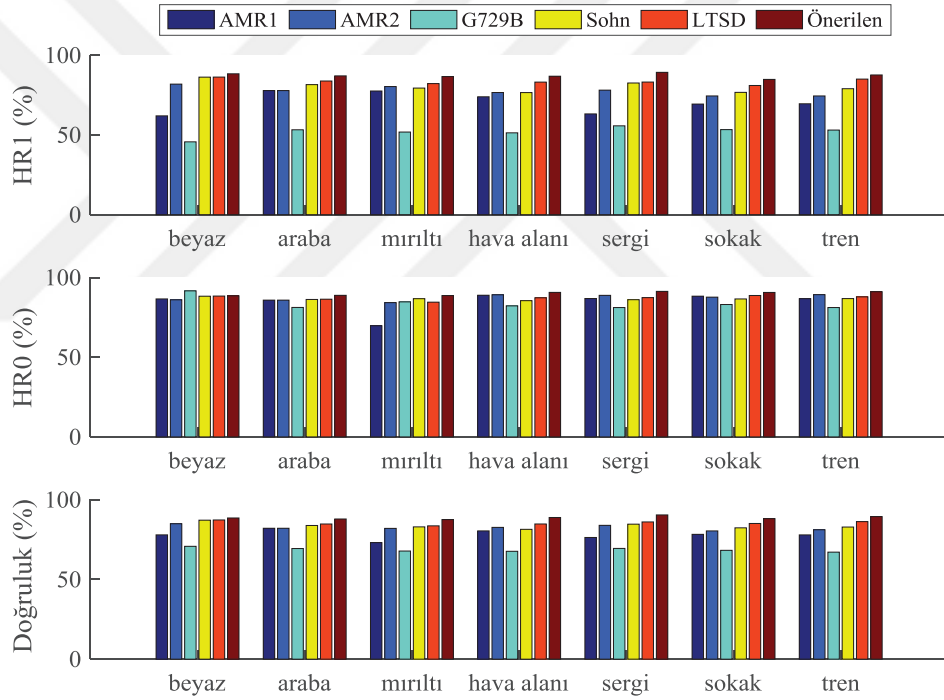
Şekil 7.17 Tüm gürültü ve SNR seviyeleri için yöntemlere ait genel doğruluk oranı değerleri

Beş SNR seviyesi için ortalama başarı oranları

AMR1, AMR2, G729B, SOHN, LTSD ve önerilen AKA tabanlı SEA algoritmaları için yedi farklı gürültü tipi için beş SNR seviyesi (-10, -5, 0, 5 ve 10 dB) üzerinden ortalama alınarak elde edilen HR1, HR0 ve genel doğruluk oranı sonuçları Şekil 7.18'de gösterilmektedir. Şekil 7.18'den görülebileceği gibi,

önerilen AKA tabanlı SEA yönteminin diğer SEA yöntemlerine kıyasla tüm gürültü tipleri için en iyi HR1 değerlerini sağladığı açık bir şekilde görülmektedir (beyaz gürültüde % 88,36; araba gürültüsünde % 87,05; mırıltı gürültüsünde % 86,60; hava alanı gürültüsünde % 86,83; sergi gürültüsünde % 89,26; sokak gürültüsünde % 84,72 ve tren gürültüsü koşullarında % 87,63 HR1 sonuçları). Ayrıca, G.729B yöntemi tüm gürültü çeşitlerinde en düşük HR1 değerlerine sahiptir.

Benzer şekilde, önerilen yöntem beyaz gürültü (%88,50 HR0 değeri) koşulları dışında tüm gürültü çeşitlerinde (araba % 88,81; mırıltı % 88,65; hava alanı % 90,65; sergi %91,29; sokak %90.62; tren % 91,08) en iyi HR0 sonuçlarını vermektedir.



Şekil 7.18 Tüm gürültü tipleri için SNR seviyelerinin ortalaması alınarak bulunan HR1, HR0 ve Doğruluk yüzdeleri

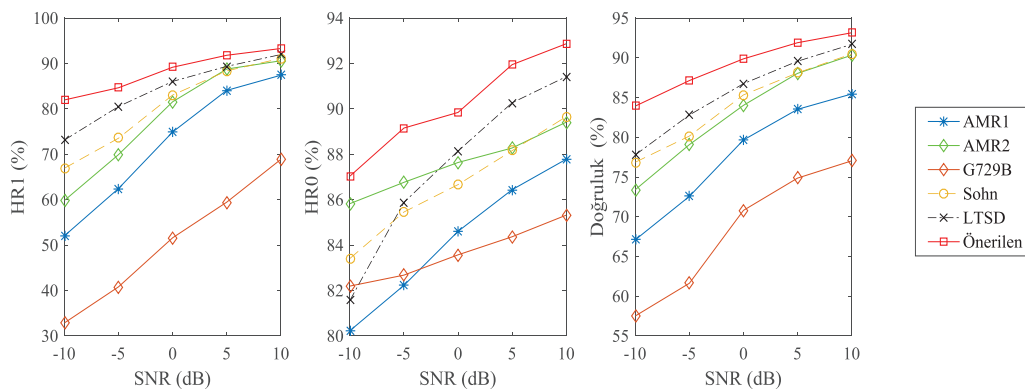
Tüm SEA yöntemlerinin genel doğruluk oranı sonuçlarına göre değerlendirilmesinde, önerilen AKA tabanlı SEA yönteminin tüm gürültü çeşitleri için en iyi sonuçları verdiği görülmektedir. Gürültü çeşitlerine göre önerilen yöntemin genel doğruluk oranı dağılımı yüzdesel olarak şu şekilde hesaplanmıştır: beyaz gürültüde % 88,56, araba gürültüsünde % 87,93, mırıltı gürültüsünde % 87,55, hava alanı gürültüsünde % 88,87, sergi gürültüsünde % 90,48, sokak

gürültüsünde % 88,14 ve tren gürültüsü koşullarında % 89,43. Ayrıca, G.729B SEA yönteminin tüm gürültü çeşitlerinde diğer yöntemlere göre daha düşük genel doğruluk oranlarına sahip olduğu Şekil 7.18’de görülmektedir.

Yedi gürültü tipi için ortalama başarı oranları

Şekil 7.19’da, önerilen AKA tabanlı SEA yönteminin AMR1, AMR2, G.729B gibi üç standart SEA yöntemi ve son yıllarda sıklıkla kullanılan SOHN ve LTSD gibi SEA yöntemleri ile beş farklı SNR seviyesi için yedi gürültü tipinin ortalaması alınarak elde edilen HR1, HR0 ve genel doğruluk oranı sonuçları verilmektedir. HR1, HR0 ve genel doğruluk oranı tüm gürültülerin ortalamasından elde edilerek gürültü tipinden bağımsız olarak önerilen yöntemin performansı ölçülmüştür. Şekil 7.19’dan görülebileceği üzere, G.729B SEA yöntemi tüm SNR seviyelerinde en düşük HR1 (-10 dB’de % 38,35 ve 10 dB’de % 68,21) ve genel doğruluk oranı (-10 dB’de % 57,96 ve 10 dB’de % 77,40) sonuçlarına sahipken yüksek SNR seviyelerinde ise en düşük HR0 (0 dB’de % 83,41 ve 10 dB’de % 86,31) sonuçlarını vermektedir.

Elde edilen sonuçlara göre, AMR2 yöntemi HR1, HR0 ve genel doğruluk oranı bakımından AMR1 ve G.729B yöntemlerine göre daha iyi bir performans göstermektedir. Ayrıca, LTSD yöntemi tüm SNR seviyeleri için SOHN SEA yönteminden daha iyi HR1 ve genel doğruluk oranı sonuçları, yüksek SNR seviyelerinde ise HR0 bakımından daha iyi sonuçlar vermektedir.



Şekil 7.19 Her bir SNR seviyesinde tüm gürültü tiplerinin ortalama alınarak bulunan HR1, HR0 ve genel doğruluk yüzdeleri

Önerilen AKA tabanlı SEA yöntemi tüm SNR seviyeleri için en iyi HR1 (-10 dB’de % 80,61 ve 10 dB’de % 92,84), HR0 (-10 dB’de % 86,81 ve 10 dB’de % 92,64) ve genel doğruluk oranı (-10 dB’de % 83,98 ve 10 dB’de % 92,85) sonuçlarını vermektedir. Ayrıca, artan SNR seviyeleri ile birlikte tutarlı bir ilerleme ve artış göstermektedir. Bu nedenle, önerilen yöntemin tüm gürültü çeşitlerinin ortalaması alındığında, tüm SNR seviyeleri için konuşma ve konuşma dışı bölgeleri tanımlamada iyi bir performans gösterdiği söylenebilir.

Önerilen ve diğer SEA yöntemleri, -10 ile 10 dB arasında değişen beş farklı SNR seviyesi ve yedi farklı gürültü tipi için ortalama alınarak HR1, HR0 ve genel doğruluk oranı açısından karşılaştırılmış ve bu sonuçlar Çizelge 7.2’de özetlenmiştir. Çizelge 7.2’den görülebileceği üzere, gürültü ve SNR seviyelerinin ortalaması alındığında önerilen yöntem en iyi genel doğruluk oranı (% 88,71), en iyi HR1 (% 87,22) ve en iyi HR0 (% 89,94) sonuçlarını sağlamaktadır. Önerilen yöntemi konuşma ve konuşma dışı bölgelerin belirlenmesi başarısı açısından LTSD SEA yöntemi izlemektedir. Ayrıca, gürültü tipi ve SNR seviyesi ortalamalarına göre en kötü sonuçları G.729B SEA yöntemi vermektedir.

Çizelge 7.2 Türkçe konuşmalar için yedi farklı gürültü tipi ve beş farklı SNR seviyesinin ortalaması alınarak yöntemlere ait ortalama başarı yüzdeleri

	SEA Yöntemi					
	AMR1	AMR2	G.729B	SOHN	LTSD	Önerilen
HR0 (%)	84,68	87,27	83,59	86,58	87,21	89,94
HR1 (%)	70,49	77,67	52,02	80,30	83,52	87,22
Doğruluk (%)	78,03	82,47	68,65	83,63	85,43	88,71

7.7 Tartışma

Bu çalışmada, Ampirik Kip Ayırıştırma kullanılarak içkin kip fonksiyonu katsayılarına Teager enerji operatörü uygulanması ve enerji dağılımının Gamma istatistiksel modeline dayanan yeni bir sesli/konuşma etkinlik algılama algoritması önerilmiştir. Gamma dağılım işlevi, gürültülü konuşma ve kestirilen gürültünün Teager enerji uygulanmış İKF katsayılarını kullanarak uyarlanabilir eşik değerlerinin belirlenmesinde kullanılmıştır. Uyarlanabilir eşik değerleri, Simetrik

Kullaback-Leibler uzaklığı kullanılarak konuşma ya da gürültü olma olasılığı üzerinden belirlenmiş ve bu eşik değerleri her kip için hesaplanmıştır. Her kip için katsayılardan elde edilen enerjilerin eşik değerlerinden geçirilmesi ile sesli ya da konuşma olan bölgeler etiketlenmiştir. Bu çalışma hem İngilizce hem de Türkçe gürültülü konuşma cümleleri için yapılmış ve önerilen yöntem için elde edilen sonuçlar literatürde sıklıkla kullanılan AMR1, AMR2, G.729B, SOHN ve LTSD gibi diğer sesli etkinlik algılama yöntemleri ile karşılaştırılmıştır. Elde edilen sonuçlara göre önerilen yöntem konuşma bölgelerini algılama oranı (HR1), konuşma olmayan bölgeleri algılama oranı (HR0) ve genel doğruluk oranı (GDO) bakımından diğer yöntemlere göre üstünlük sağlamıştır.

Önerilen sesli etkinlik algılama yönteminin Türkçe için konuşma olmayan bölgeleri algılama oranı (HR0) % 89,94 iken İngilizce için bu oran % 90,17 olarak bulunmuştur. Benzer olarak Türkçe için konuşma ya da sesli bölgeleri algılama oranı (HR1) % 87,22 iken İngilizce için bu oran % 88,19 ve hem konuşma hem de konuşma olmayan bölgeleri algılama oranı olan genel doğruluk oranı (GDO) Türkçe için % 88,71, İngilizce için ise % 89,20 olarak hesaplanmıştır.

Tüm sonuçlar göz önünde bulundurulduğunda, önerilen yöntemin performansı tüm gürültü ve SNR seviyelerinin ortalamasının alınması durumunda hem Türkçe cümleler hem de İngilizce cümleler için benzerlik göstermektedir. Sonuçların birbirine yakın olması önerilen yöntemin dillerin yapısından bağımsız bir şekilde sesli/konuşma bölgelerinin belirlenmesinde kullanılabileceğini göstermektedir.

8. AMPİRİK KİP AYRIŞTIRMA TABANLI KONUŞMA İYİLEŞTİRME YAKLAŞIMI

Yeni bir zamansal sinyal ayrıştırma yöntemi olan Ampirik Kip Ayrıştırması (AKA) durağan olmayan ve doğrusal olmayan zaman serilerini analiz etmek için önerilmiştir. Bu yeni uyarlanabilir yöntem, bir sinyali içkin kip fonksiyonları (Intrinsic Mode Functions-IMFs) olarak adlandırılan salınımlı bileşenleri halinde parçalara ayırmaktadır. AKA ayrıştırma yönteminin dalgacık dönüşümüne göre en büyük üstünlüğü temel fonksiyonların sabit olmayıp sinyalden türetilmesidir. Bu nedenle analiz, temel fonksiyonların sabitlendiği geleneksel yöntemlerin aksine uyarlanabildir.

Son yıllarda, konuşma iyileştirmede AKA yöntemi sıklıkla kullanılmaktadır. Konuşma iyileştirmede gürültü azaltmak için AKA tabanlı süzgeçleme (Khaldi et al., 2008; Sharma and Mahadeva Prasanna, 2016) ya da eşikleme yaklaşımları (Hamid et al., 2013; Kemiha, 2014) önerilmiştir. Ancak AKA yönteminde, sabit bir eşik değeri belirlenerek gürültülü konuşmanın eşiklenmesi gürültünün yanında konuşmaya ait bilgilerin de kaybolmasına neden olabilmektedir. Bu nedenle gürültü atmada sabit eşik kullanmak yerine İKF katsayılarından uyarlanır eşik değerleri belirlenmesi iyileştirmede daha başarılı sonuçlar vermektedir (Chatlani and Soraghan, 2009; Khaldi et al., 2014b). Uyarlanabilir eşik değerlerinin belirlenmesi, dalgacık dönüşümü tabanlı yöntemlerde olduğu gibi Teager Enerji Operatörü kullanılarak belirlenebilmektedir (Sanam and Shahnaz, 2012, 2013; Islam et al., 2015).

Bu çalışmada, zaman-frekans alanı yöntemlerinden olan AKA kullanılarak elde edilen İKF katsayılarını eşiklemeye dayalı yeni bir konuşma iyileştirme yaklaşımı geliştirilmiştir. Önerdiğimiz yöntemde, AKA tabanlı diğer konuşma iyileştirme yöntemlerinden farklı olarak konuşma ve konuşma dışı bölgelerin İKF katsayılarının enerjisinin uygun bir istatistiksel modeli çıkartılmıştır. Çıkartılan istatistiksel dağılım yardımıyla, gürültülü konuşma sinyallerinin konuşma olma olasılığı veya gürültü olma olasılığı üzerinden uyarlanabilir eşik değerlerinin belirlenmesi amaçlanmıştır. Önerilen bu yaklaşımda, gürültülü konuşma sinyali AKA yöntemi ile İKF katsayılarına ayrıştırılmış ve bu İKF katsayılarına Teager Enerji Operatörü (TEO) uygulanmıştır. TEO uygulanmış İKF katsayılarından bir istatistiksel model olan Gamma dağılım fonksiyonu yardımıyla hem gürültülü sinyalin hem de gürültünün olasılık dağılım fonksiyonu elde edilmiştir. Simetrik Kullback-Leibler (SKL) uzaklığı kullanılarak konuşmanın ve gürültünün olduğu

kiplere göre uygun eşik değerleri belirlenmiştir. Her kip için uygun eşik değerleri belirlendikten sonra, ilgili kipteki İKF katsayıları bir yarı-yumuşak eşikleme fonksiyonundan geçirilerek ters AKA yöntemi ile iyileştirilmiş konuşma sinyali elde edilmektedir.

8.1 Zaman-Frekans Alanı Yöntemleri

8.1.1 Ayırık Dalgacık Dönüşümü

Şimdiye kadar ele alınan tüm konuşma iyileştirme yöntemleri, hedef sinyalin Kısa Zamanlı Fourier Dönüşümü ile elde edilen izgesel bilgilere dayanmaktadır. Tüm bu frekans tabanlı yöntemler özellikle ünlü seslerin düşük frekanslı harmonikleri gibi konuşmanın yavaş değişen izgesel karakteristiklerini korumak için yeterli değildir. Ayrıca /p/, /t/, /k/, /b/, /d/, /g/ gibi patlamalı ünsüzlerin anlık özelliklerinin korunması için gereklidir. Dolayısıyla bir konuşma iyileştirme algoritması olarak Dalgacık Dönüşümü tabanlı yöntemler önerilmektedir. Dalgacık Dönüşümü, sinyalin düşük frekanslı bileşenleri için yüksek frekans çözünürlüğü (düşük zaman çözünürlüğü), yüksek frekanslı bileşenleri için ise düşük frekans çözünürlüğü (yüksek zaman çözünürlüğü) sağlamaktadır. Dalgacık Dönüşümünde sonlu etkili destek genişliği sağlayan bir dalgacık kullanılması durumunda, dalgacık analizi hem zaman hem de frekans alanında sinyalin yerel özelliklerini tespit etmede kullanılabilir (Quatieri, 2001).

Dalgacık Büzülme

Dalgacık analizi; ön-uç işitsel sistemi temel alan bir yapıya sahip olup, konuşma iyileştirme için kullanılan bir sinyal işleme aracıdır. En çok kullanılan yaklaşım, Donoho (1995) tarafından yapılan ve dalgacık katsayılarının doğrusal olmayan eşikleme ile yapılmaktadır.

Donoho (1995) aşağıdaki gibi dalgacık temelli bir yaklaşım önermiştir:

$$y = x + n \quad (8.1)$$

Burada y , sıfır ortalamalı beyaz Gauss gürültüsü n ile temiz sinyal olan x 'in oluşturduğu toplamsal bir gürültülü sinyali göstermektedir.

Dalgacık alanında eşitlik şu şekilde ifade edilebilir:

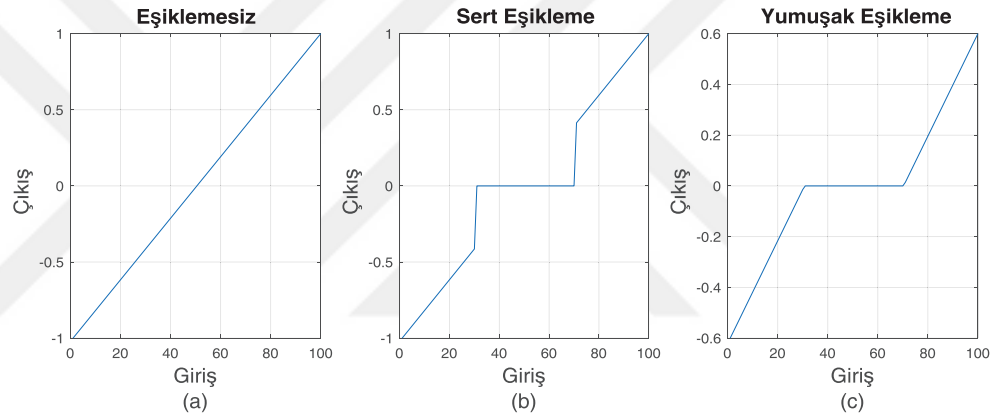
$$W_y = W_x + W_n \quad (8.2)$$

Temiz x sinyali Eşitlik (8.3) ile kestirilebilir:

$$x = W^{-1}X_{\text{kestirim}} = W^{-1}Y_{\text{eşik}} \quad (8.3)$$

Burada $Y_{\text{eşik}}$ eşiklemeden sonraki dalgacık katsayılarını, W^{-1} ise ters dalgacık dönüşümünü göstermektedir.

Genellikle iki türlü eşikleme fonksiyonu kullanılmaktadır. Şekil 8.1’de gösterildiği gibi bunlardan ilki sert eşikleme ve bir diğeri ise yumuşak eşikleme olarak adlandırılmaktadır.



Şekil 8.1 Eşikleme fonksiyonları: (a) eşiklemesiz, (b) sert eşikleme, (c) yumuşak eşikleme

$$THR_{\text{sert}}(X, T) = \begin{cases} X, & |X| > T \\ 0, & |X| < T \end{cases} \quad (8.4)$$

$$THR_{\text{yumuşak}}(X, T) = \begin{cases} \text{sgn}(X)(|X| - T), & |X| > T \\ 0, & |X| < T \end{cases} \quad (8.5)$$

Burada X eşiklemeden önceki dalgacık katsayılarını, T ise eşik değerini temsil etmektedir.

Eşikleme Seçimi

Eşik değerinin seçimi birçok şekilde belirlenebilir. Donoho (1995), eşiklemeyi beyaz Gauss gürültüsü varsayımına dayalı olarak Eşitlik (8.6) ile formüle etmiştir:

$$T = \sigma \sqrt{2 \log(N)} \quad (8.6)$$

Burada T eşik değerini, N gürültülü sinyalin uzunluğunu ve $\sigma = MAD/0,6745$ varyansı göstermektedir. MAD ise dalgacık katsayılarının bir ölçekteki mutlak medyanını göstermektedir.

Johnstone ve Silverman (1997) gürültü gidermede her frekans aralığında gürültünün standart sapması farklı olduğundan Eşitlik (8.7) ile verilen seviye bağımlı eşikleme yöntemini önermişlerdir.

$$\lambda_a = \sigma_a \sqrt{2 \log(N_a)} \quad (8.7)$$

Burada N_a , a ölçeğindeki örnek sayısını, $\sigma_a = MAD_a/0,6745$ a ölçeğindeki varyansı, MAD_a ise a ölçeğindeki mutlak medyanı göstermektedir.

8.1.2 Ayırık Dalgacık Paket Dönüşümü

Geleneksel ADD tabanlı analizde, sadece ölçek alanı ayrıştırılırken dalgacık alanı ayrıştırılmamaktadır. Gürültülü konuşmaya iyileştirme amacıyla ADD uygulandığında böyle bir analizin önemli bir eksikliği, bozulmamış bir konuşma kalitesine yol açan birçok ses-benzeri konuşma bileşenini içeren işlenmemiş konuşma çerçevelerinin büzülmesidir.

ADD tabanlı analizden farklı olarak, Dalgacık Paket tabanlı analizde, dalgacık alan ayrıca ayrıştırılır ve böylece daha yüksek frekans bandı ayrışması mümkün olmaktadır. Hem yaklaşım hem de ayrıştırma dalgacık paket katsayıları, her ayrıştırma seviyesinde iki parçaya ayrıldığı için, üstün frekanslı lokalizasyona sahip tam bir ikili ağaç elde edilebilir. Ayırık Dalgacık Paket Dönüşümü (ADPD), sinyali daha fazla sayıda alt bantta parçalara ayırır ve standart dalgacık dönüşümünden daha yüksek frekanslarda daha yüksek frekans çözünürlüğüne sahip olabilen çok işlevli bir çerçeve oluşturur (Ghanbari and Karami-Mollaei, 2006; Xu et al., 2007; Messaoud et al., 2015).

8.1.3 Ampirik Kip Ayrıştırması

Hilbert Huang Dönüşümü'nün (HHD) temel bir bileşeni olarak Ampirik Kip Ayrışımı (AKA), doğrusal olmayan ve durağan olmayan sinyallerin işlenmesi için uyarlanabilir bir yöntemdir (Khaldi, 2012). AKA yöntemi, önceki sinyal işleme

yöntemlerinden (Fourier, Dalgacık ve Dalgacık Paket Dönüşümü) farklı olarak sezgisel, doğrudan ve uyarlanabilir bir yöntemdir. Ayırıştırmanın bütün işlemlerinde, tüm temel işlevler sinyalin kendisinden türetilmiştir. Bu nedenle, AKA yöntemi konuşma sinyali gibi doğrusal olmayan ve durağan olmayan sinyalleri işlemek için oldukça uygundur. Bir $x(t)$ sinyali göz önüne alındığında, AKA yöntemi, "eleme" (sifting) işlemi yoluyla elde edilen ve İçkin Kip İşlevleri (İKF) olarak adlandırılan bir dizi salınımsal bileşene uyarlanabilir biçimde ayırştırmaktadır. Elde edilen her İKF, orijinal sinyalden gelen frekans bileşenlerinin bir alt kümesinden oluşan bir salınımsal sinyaldir.

Eleme (sifting) süreci aşağıdaki gibi tanımlanabilir:

Adım 1: $x(t)$ 'nin tüm ekstramaları (aşırılıkları) (maksima / minima) bulunur.

Adım 2: Sırasıyla bir üst zarf $e_{max}(t)$ ve bir alt zarf $e_{min}(t)$ elde etmek için kübik spline interpolasyon ile maksima ve minimal noktaları hesaplanır.

Adım 3: Üst ve alt zarfların ortalaması $m_1(t) = (e_{max}(t) + e_{min}(t)) / 2$ hesaplanır ve $h_1(t) = x(t) - m_1(t)$ elde etmek için orijinal sinyalden çıkarılır.

Adım 4: $h_1(t)$ 'nin aşağıdaki iki İKF koşulunu karşılayıp karşılayamadığı değerlendirilir:

- Tüm veri parçalarında ekstremaların sayısı sıfır geçişlerinin sayısına eşit olmalı ya da en fazla bir fark olmalıdır.
- Her noktada, maksimum zarfın ve minimum zarfın ortalaması sıfır olmalıdır. Yani, sinyal zaman eksenini etrafında simetrik olmalıdır.

$h_1(t)$, bir İKF olma koşulunu sağlıyorsa, ilk $IKF_1(t)$, $IKF_1(t) = h_1(t)$ olarak kabul edilir.

Adım 5: Eğer $h_1(t)$ iki koşulu sağlamazsa, $h_1(t)$ yeni bir sinyal olarak kabul edilir ve Adım 1-4 $h_2(t)$ 'yi oluşturmak için $h_1(t)$ üzerinden tekrarlanır. Eğer $h_2(t)$ iki koşulu sağlayamıyorsa, eleme işlemini sonlandırmak için bir standart sapma olan durdurma kriteri (Standart Sapma-SS) hesaplanır. Durdurma kriteri şu şekilde verilmektedir:

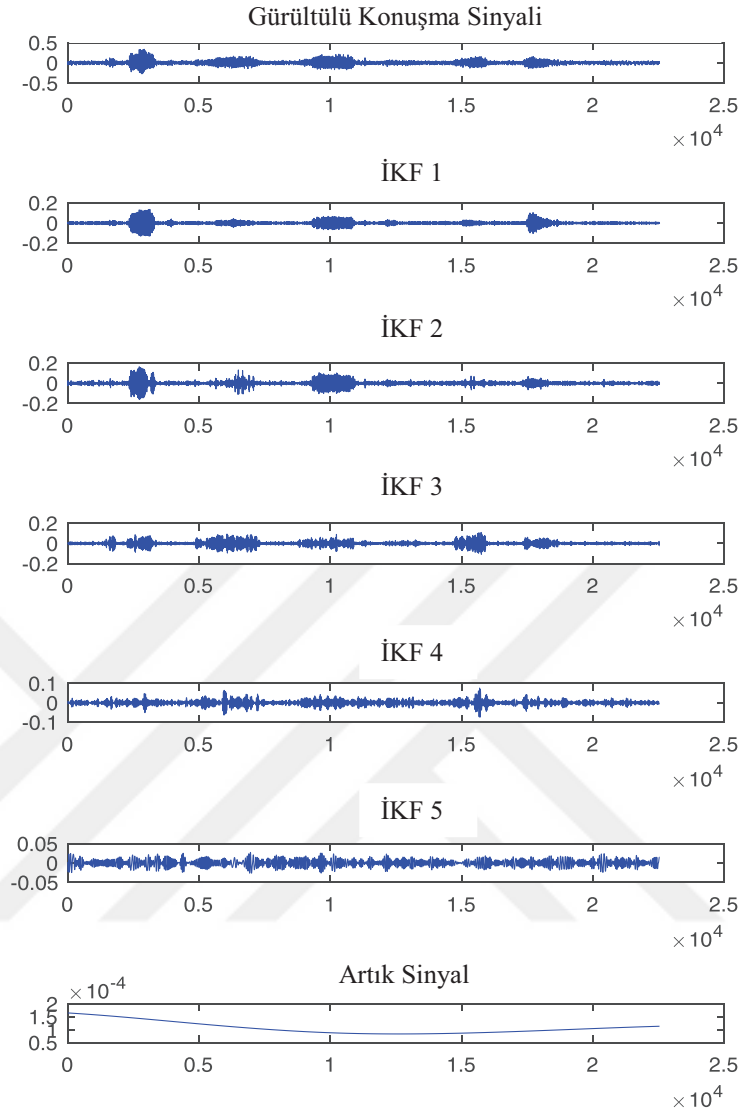
$$SS(i) = \sum_{t=0}^N \frac{|h_{i-1}(t) - h_i(t)|^2}{h_{i-1}^2(t)} \quad (8.8)$$

Genellikle, SS (standart sapma) değeri aralığı 0,2 ile 0,3 arasındadır. $h_2(t)$ SS 'yi karşılarsa, $\dot{IKF}_1(t) = h_2(t)$ olur. Eğer $h_2(t)$ durdurma kriterini karşılamazsa, $h_2(t)$ yeni bir sinyal olarak kabul edilir ve Adım 1-5 boyunca, $h_i(t)$ SS ya da $\dot{IKF}$ olmanın iki koşulunu sağlayıncaya kadar $h_i(t)$ 'yi oluşturmak için $h_2(t)$ üzerinden işlemler tekrarlanır. Daha sonra, $\dot{IKF}_1(t) = h_i(t)$ olarak hesaplanır.

Adım 6: $IMF_1(t)$ oluşturulur ve artık (residual) $r_1(t)$ 'yi elde etmek için orijinal sinyalden çıkarılır: $r_1(t) = x(t) - \dot{IKF}_1(t)$. Artık sinyal orijinal sinyal olarak ele alınır ve sonraki artık sinyali almak için 1'den 5'e kadar adımlar tekrarlanır. Dolayısıyla artık sinyal, $r_n(t) = r_{n-1}(t) - \dot{IKF}_n(t)$ olarak ifade edilir. Bu noktada, $r_n(t)$ monoton bir dizidir. Eleme işlemi sonrasında, orijinal sinyal $\dot{IKF}_1(t)$, $\dot{IKF}_2(t)$, ..., $\dot{IKF}_n(t)$ gibi birçok $\dot{IKF}$ bileşenine ve artık bir dizi $r_n(t)$ 'ye ayrıştırılmaktadır. Bu nedenle, orijinal sinyal şu şekilde ifade edilebilir:

$$x(t) = \sum_{i=1}^n \dot{IKF}_i(t) + r_n(t) \quad (8.9)$$

Şekil 8.2'de 5 dB SNR seviyesinde araba gürültüsü eklenmiş gürültülü konuşma sinyali ve bu sinyalin AKA yöntemi ile ayrıştırılan ilk 5 $\dot{IKF}$ ve artık bileşeni gösterilmektedir.



Şekil 8.2 AKA ile ayrıştırılan gürültülü konuşma sinyali için ilk 5 İKF ve artık bileşen

Ampirik Kip Ayrıştırma Büzülme

Gürültülü giriş konuşma sinyalinin iyileştirilmiş bir versiyonunda, İKF'ler sinyal geri çatılmadan önce eşiklenerek elde edilebilir. Bu durumda, her İKF'nin eşik parametresi aşağıdaki ifadeyle kestirilebilir:

$$\tau_k = \sigma_k \sqrt{2 \log(T)} \quad (8.10)$$

Burada T sinyal uzunluğunu, σ_k ise İKF'lerdeki kestirilen gürültü seviyesini göstermektedir. İlk İKF'nin gürültü seviyesi şu şekilde hesaplanmaktadır:

$$\tilde{\sigma}_1 = 1.4826 \times \text{Median}\{|\text{IKF}_1(t) - \text{Median}\{\text{IKF}_1(t)\}|\} \quad (8.11)$$

AKA tabanlı global eşiklemeye göre (Flandrin et al., 2004), k . İKF'nin gürültü seviyesi $\tilde{\sigma}_k$, Eşitlik (8.10)'da verilen ilk İKF'nin gürültü seviyesi $\tilde{\sigma}_1$ 'den elde edilebilir:

$$\tilde{\sigma}_k = \frac{\tilde{\sigma}_1}{\sqrt{2^{k-1}}} \quad (8.12)$$

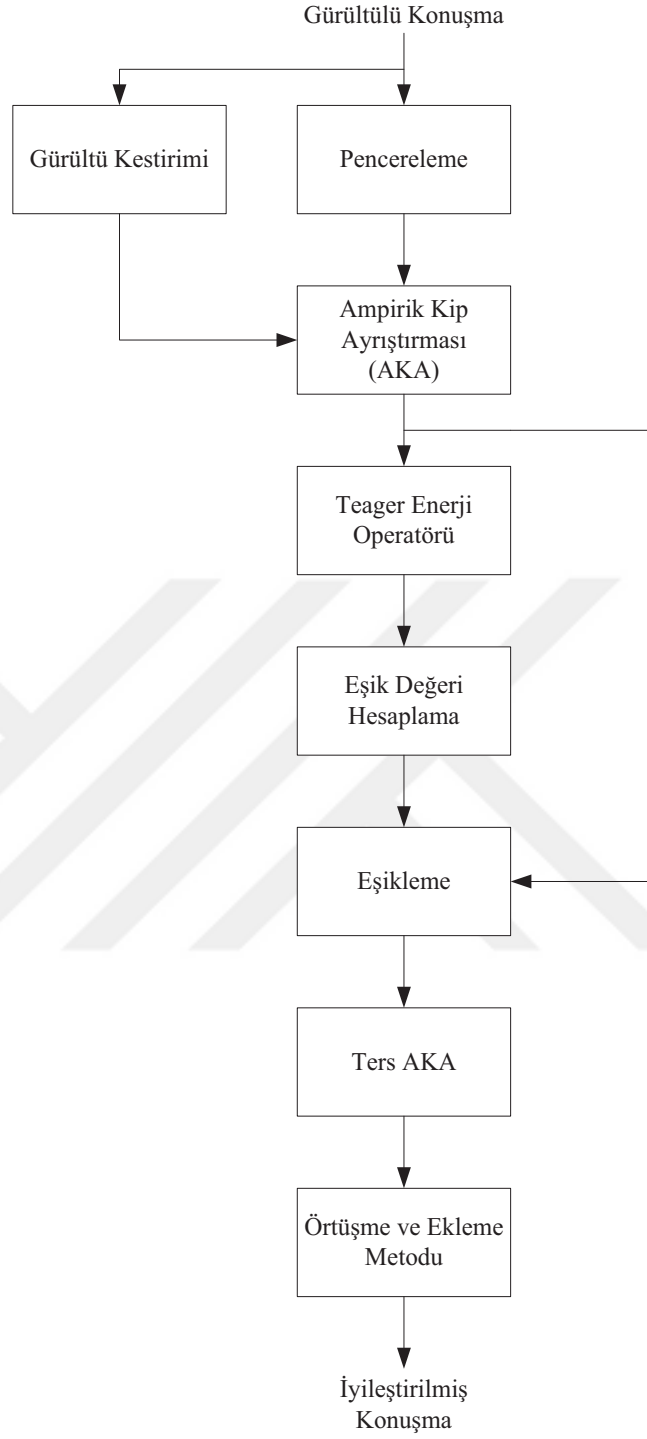
Farklı doğrusal olmayan büzülme (shrinkage) fonksiyonları vardır. Bu fonksiyonlardan sert ve yumuşak eşikleme fonksiyonları konuşma iyileştirmede en fazla kullanılan fonksiyonlardır.

8.2 Önerilen Konuşma İyileştirme Yöntemi

Önerilen yöntemin blok diyagram Şekil 8.3'de gösterilmektedir. Şekil 8.3'den görüleceği üzere, gürültülü giriş konuşma sinyalleri çerçevelere ayrılmaktadır. Gürültülü konuşma sinyallerinin iyileştirilmesinde gürültünün karakteristiği önemli bir rol oynadığından öncelikle gürültülü konuşma sinyalinden bir gürültü kestirimi algoritması ile gürültü sinyali elde edilmektedir. Daha sonra hem gürültülü konuşma hem de elde edilen gürültüye her bir çerçeve için AKA yöntemi uygulanmaktadır. Gürültülü konuşma ve gürültü sinyallerine Teager Enerji Operatörü (TEO) uygulanarak her kip için TEO'ya dayalı uyarlanabilir eşik değerlerinin belirlenmesi amaçlanmaktadır. Eşik değerleri belirlendikten sonra gürültülü konuşma sinyallerinin İKF katsayıları seçilen bir eşikleme fonksiyonundan geçirilmektedir. Eşikleme fonksiyonundan geçirilen çerçeve bazındaki katsayılar ters AKA yöntemi ile geri çatılarak Örtüşme ve Ekleme (Overlap-and-Add) yöntemi ile iyileştirilmiş konuşma sinyalinin elde edilmesi sağlanmaktadır.

8.2.1 Gürültü Kestirimi

Konuşma iyileştirmede, gürültünün karakteristiğinin belirlenmesi ve bu gürültünün tespit edilmesi iyileştirme performansı açısından önemli bir rol oynamaktadır. Son zamanlarda, konuşma iyileştirme çalışmalarında gürültü kestirim yöntemi olarak en uygun şekilde uyarlanmış logaritmik izgesel genlik kestirimi (Optimally Modified Log-Spectral Amplitude Estimator - OMLSA) algoritması sıklıkla kullanılmaktadır (Nsabimana et al., 2010; Flandrin et al., 2014; Nsabimana et al., 2009).



Şekil 8.3 Önerilen konuşma iyileştirme yönteminin blok diyagramı

Geleneksel konuşma iyileştirme teknikleri konuşma varlığı (speech presence) hipotezi altında temiz konuşma izgesini tahmin ederken, bir önsel konuşma yokluğu olasılığı (speech absence probability-SAP) içeren modifiye edilmiş bir kestirim, gürültünün belirlenmesinde daha iyi bir performans göstermektedir (Ephraim and Malah 1985; Soon et al., 1999; Kim et al., 2000).

Ephraim ve Malah (1985) tarafından geliştirilen Log-İzgesel Genlik (LSA) kestirimcisinin müzikal artık gürültü olaylarını azaltmada çok etkili olduğu kanıtlanmıştır. Sinyal varlığı belirsizliği altındaki modifikasyonu, her frekans bölmesi ve her çerçeve için tahmin edilen koşullu konuşma varlığı ihtimali ile izgesel kazanç çarpılarak elde edilmektedir.

Bu çalışmada, iyi bir konuşma iyileştirme için optimal olarak modifiye edilmiş LSA (OM-LSA) yöntemi gürültü kestiriminde kullanılmıştır. Optimal izgesel kazanç fonksiyonu, konuşma varlığı belirsizliğiyle ilişkili varsayımsal kazançların ağırlıklı geometrik ortalaması olarak elde edilmektedir. Her bir varsayımsal kazancın üssel ağırlığı, gözlemlenen sinyale bağlı olarak, buna karşılık gelen olasılıktır. Gürültü izgesi, konuşma varlığı olasılıkları alt bantlar tarafından ayarlanan yumuşatma parametresi kullanılarak geçmişteki izgesel güç değerlerinin tekrar ortalaması alınarak hesaplanmaktadır (Cohen and Berdugo, 2001; Cohen, 2002)

8.2.2 Teager Enerji Operatörü

Teager Enerji Operatörünün sürekli formu aşağıdaki gibi verilmektedir:

$$\Psi_c[y(t)] = \left(\frac{d}{dt} y(t) \right)^2 - y(t) \frac{d^2}{dt^2} y(t) \quad (8.13)$$

Burada $\Psi_c[.]$ ve $y(t)$ sırasıyla sürekli Teager enerji operatörünü ve sürekli bir sinyali temsil etmektedir. Belirli bir bantta sınırlandırılmış $y[n]$ ayrık sinyali için ayrık zamanlı TEO şu şekilde ifade edilmektedir:

$$\Psi_d(y[n]) = y[n]^2 - y[n+1]y[n-1] \quad (8.14)$$

Ayrık zamanlı Teager Enerji Operatörü, her zaman anında enerji hesaplanması için yalnızca üç örnek gerektirdiğinden neredeyse anlık bir ölçüttür. Çok iyi bir zaman çözünürlüğü sayesinde bir Teager Enerji Operatörü'nün çıkışı

bize enerji dalgalanmalarını yakalama ve dolayısıyla sinyal üretmek için gereken enerjinin tahminini sağlamaktadır.

Ampirik Kip Ayırıştırması analizi yardımıyla bir eşikleme kullanılarak gürültülü konuşmanın iyileştirilmesinde, sinyalde her zaman konuşma bölümleri olmadığından eşik değerleri zamanla uyarlanabilir olmalıdır. Eşik değerlerinin, konuşmanın olmadığı bölümlerde daha büyük ve konuşma olan bölümlerde daha küçük olması beklenmektedir. TEO, zamanla sinyalin enerjisinin bir tahminini sağladığından, konuşma veya konuşma dışı etkinlik fikrini elde etmek için kullanılabilir ve daha sonra konuşma ya da konuşma dışı bölümler için uygun bir eşik değerinin belirlenmesini sağlayabilir. Ancak TEO, gürültülü konuşmada doğrudan kullanıldığında, istenmeyen bozucu etkiler (artifact) ve güçlendirilmiş gürültüler ortaya çıkabilmektedir. Bu nedenle gürültülü konuşma sinyaline doğrudan TEO uygulamak yerine, TEO'nun gürültülü konuşmanın İKFs katsayıları üzerine uygulanması eşik değerlerinin daha uygun bulunmasını sağlamaktadır (Khaldi et al., 2014a). Eşik değerlerinin TEO yardımıyla bulunması, gürültülü konuşmanın iyileştirilmesinde konuşma kalitesini ve anlaşılabilirliği korurken mümkün olduğunca fazla gürültüyü gidermek için daha büyük bir potansiyel sağlamaktadır (Bahoura and Rouat, 2001).

Gürültülü konuşma sinyali AKA yöntemi ile ilk önce içkin kip fonksiyonu (İKFs) katsayılarına ve artık (residual) sinyale ayrıştırılır ve daha sonra Teager Enerji Operatörü'nün her kipe aşağıdaki gibi uygulanması sağlanır:

$$E_{k,m} = \Psi[\text{IKF}_{k,m}(t)] \quad (8.15)$$

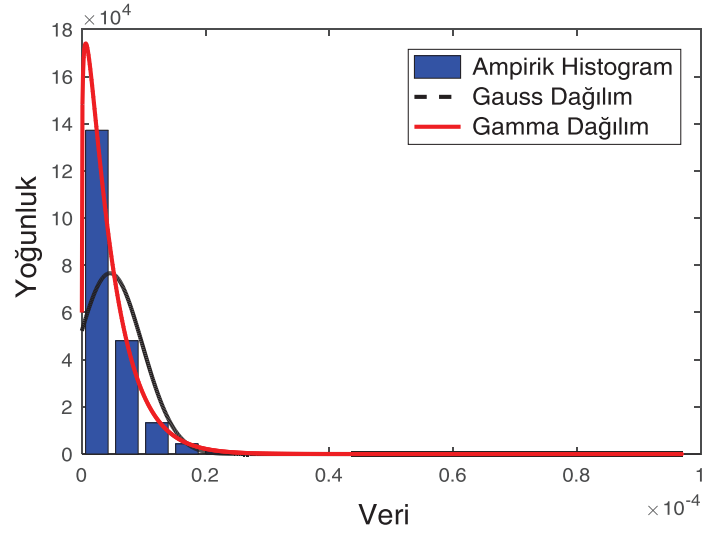
Burada k ve m sırası ile her bir kipi ve çerçeveleri göstermektedir.

8.2.3 TEO Uygulanmış İKF Katsayıları için Önerilen İstatistiksel Model

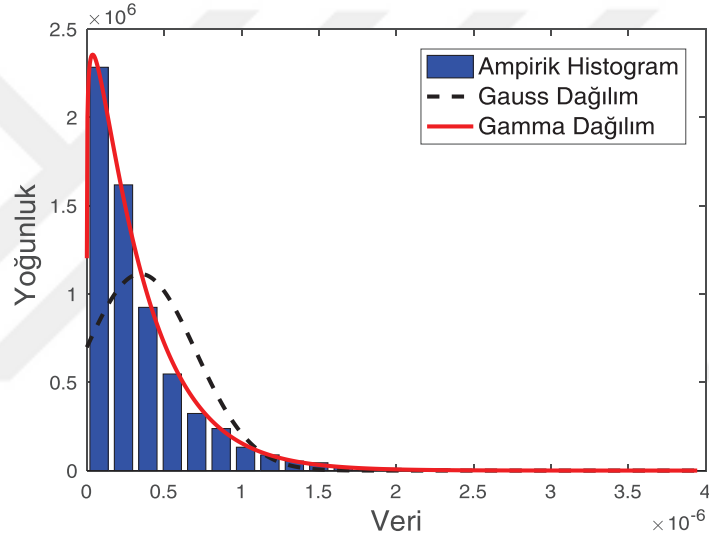
Bir dönüşüm ya da analiz alanında eşiklemeye dayanan bir konuşma iyileştirme yönteminin performansı, temel olarak eşik değeri ve eşikleme fonksiyonu gibi iki önemli faktöre bağlıdır. Gürültülü konuşmanın iyileştirilmesinde tüm AKA kipleri için tek bir eşik değeri kullanmak mantıklı değildir. Konuşma kalitesinin artırılması ve anlaşılabilirliğin sağlanmasında her bir kip için eşik değeri uyarlanabilir olmalıdır. Gürültülü konuşma, gürültü ve temiz konuşma sinyalinin $E_k(t)$ enerjisinin olasılık dağılımı göz önünde bulundurulduğunda, uygun bir örüntü eşleme ya da benzerlik ölçütü kullanılarak daha doğru ve net bir eşik değeri elde edilebilir (Islam et al., 2015).

Konuşma sinyali zamanla değişen bir sinyal olduğundan, konuşmanın ya da enerjisinin doğru olasılık dağılım fonksiyonunu elde etmek zordur. Bu nedenle doğrudan konuşmanın enerjisinin olasılık dağılım fonksiyonunu çıkarmak yerine enerjinin ampirik histogramı üzerinden uygun olasılık dağılımını çıkarmak daha doğru bir yaklaşım olacaktır. Konuşma sinyalinin enerjisinin ampirik histogramını kapsayacak şekilde en uygun olasılık dağılım fonksiyonunun belirlenmesi eşik değerlerinin hesaplanmasında önemli bir rol oynamaktadır.

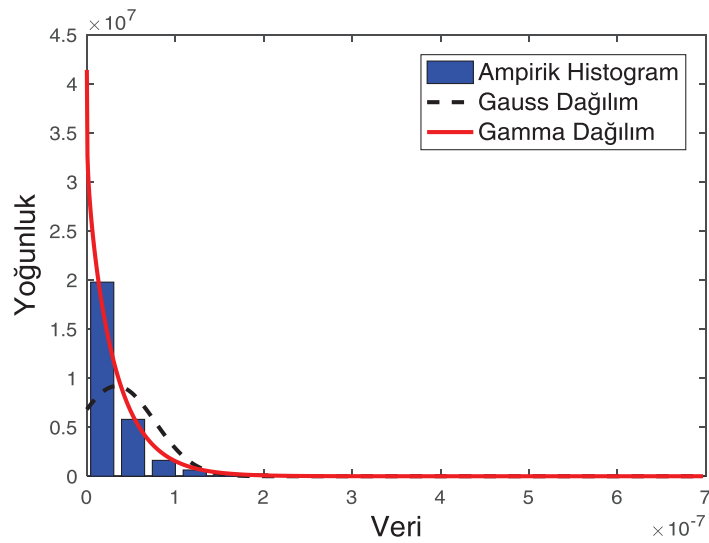
Şekil 8.4’de gürültülü konuşmanın -10, 0 ve 10 dB SNR seviyelerinde araba gürültüsü için Teager Enerji Operatörü uygulanmış İKF katsayılarının ampirik histogramı, Gauss ve Gamma dağılımları verilmiştir. Benzer şekilde aynı gürültü ve SNR seviyeleri için kestirilen gürültü sinyalinin TEO uygulanmış İKF katsayılarının ampirik histogramı, Gauss ve Gamma dağılımları Şekil 8.5’de verilmiştir. Şekil 8.4 ve Şekil 8.5’den görülebileceği üzere Gamma dağılımının, gürültülü konuşma ve gürültü sinyalinin ampirik histogramını daha doğru bir şekilde uydurduğu (fit) görülmektedir. Bu nedenle, AKA tabanlı konuşma iyileştirme yaklaşımında her kip için uyarlanabilir olan eşik değerlerinin belirlenmesinde Gamma dağılım fonksiyonu önerilmiştir.



(a)

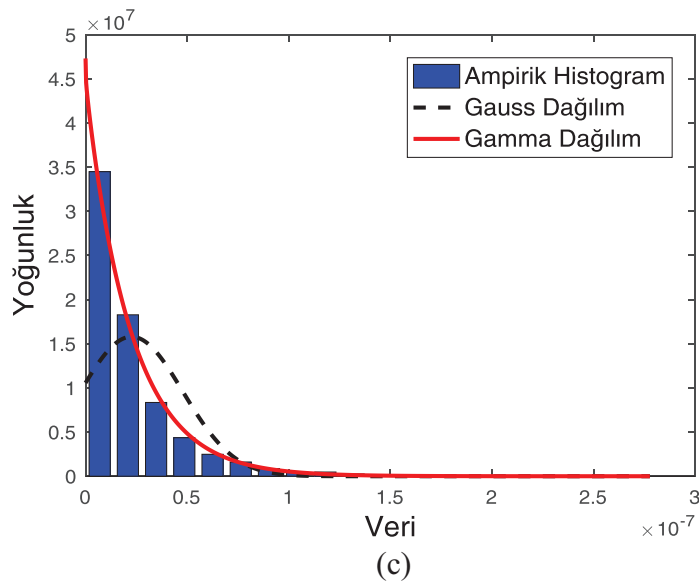
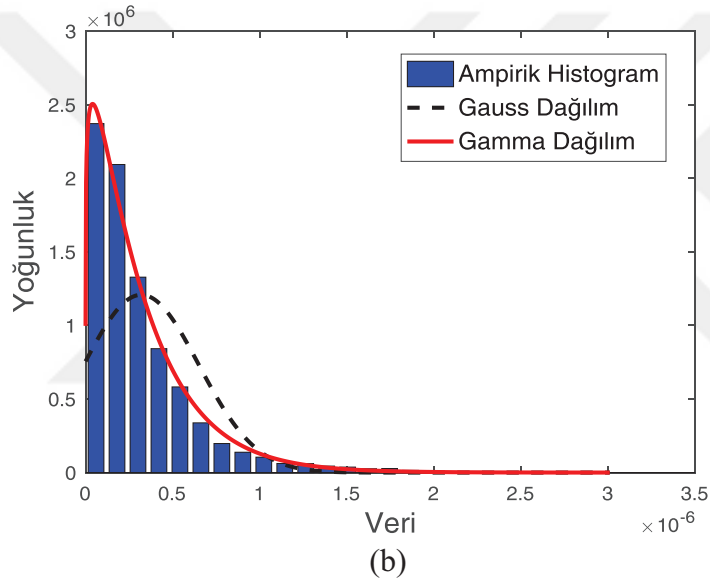
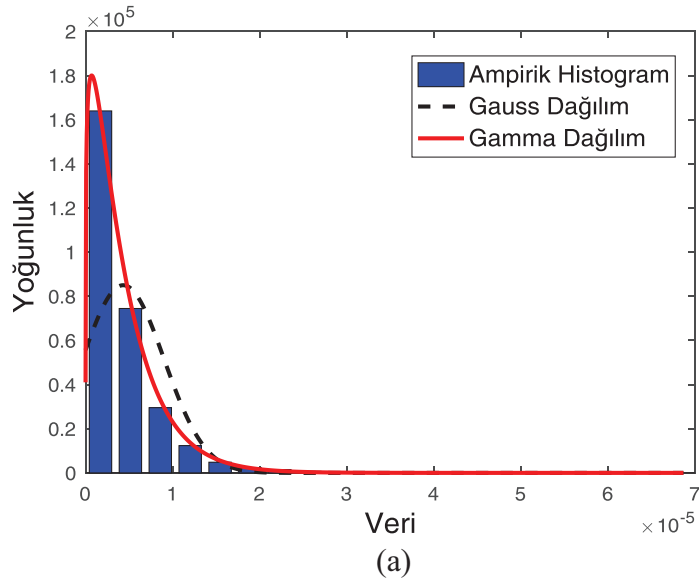


(b)



(c)

Şekil 8.4 Gürültülü konuşma sinyalinin İKF katsayılarının enerjisinin (a) -10 dB, (b) 0 dB ve (c) 10 dB SNR seviyeleri için ampirik histogram, Gauss ve Gamma dağılımları



Şekil 8.5 Gürültülü sinyalinin İKF katsayılarının enerjisinin (a) -10 dB, (b) 0 dB ve (c) 10 dB SNR seviyeleri için ampirik histogram, Gauss ve Gamma dağılımları

8.2.4 Önerilen Uyarlanır Eşik Değerlerinin Hesaplanması

Bu çalışmada, Ampirik Kip Ayırıştırması (AKA) yöntemi ile ayrıştırılan her kipteki İKF katsayılarının Teager enerjisi doğrudan kullanmak yerine bu katsayıların enerjisinin olasılık dağılımı çıkarılarak uyarlanabilir eşik değerlerinin belirlenmesi önerilmektedir. Gürültülü konuşmanın enerjisinin olasılık dağılımının belli bir aralıkta gürültünün enerjisinin olasılık dağılımına benzer olması beklenmektedir. Ayrıca bu aralığın dışında, gürültülü konuşmanın enerjisinin olasılık dağılımının temiz konuşmanın olasılık dağılımına eşit olması beklenir. Böylece, gürültülü konuşmanın, gürültünün ve temiz konuşmanın enerjisinin olasılık dağılımları göz önüne alınarak olasılık dağılımları arasında uygun bir model eşleştirme şeması veya benzerlik düzeyi kullanılarak daha doğru bir eşik değeri elde edilebilir. İki dağılım arasındaki mesafenin belirlenmesinde yaygın olarak Kullback-Leibler uzaklık ölçütü kullanılmaktadır. İki dağılım arasındaki farkları sağlam bir şekilde tahmin etmek eşik değerlerinin belirlenmesinde oldukça önemlidir.

Kullback-Liebler (K-L) uzaklığı, gürültülü konuşmanın olasılık dağılımının gürültünün olasılık dağılımına yaklaşması durumunda her zaman sıfır değerini almaktadır. K-L ayrıştırması simetrik bir ölçüt olmadığından, gürültülü konuşma ve gürültünün Gamma dağılım fonksiyonları arasında simetrik bir uzaklık olması amacıyla Simetrik Kullback-Leibler (SKL) uzaklığı kullanılmıştır. SKL uzaklığı şu şekilde ifade edilebilir:

$$SKL(p, q) = \frac{KL(p, q) + KL(q, p)}{2} \quad (8.16)$$

Burada p ve q her biri N sayıda bin'e sahip ilgili histogramlardan hesaplanan gürültülü konuşma enerjisinin ve gürültü enerjisinin yaklaşık Gamma olasılık dağılım fonksiyonlarını göstermektedir. $KL(p, q)$ şu şekilde hesaplanmaktadır:

$$KL(p, q) = \sum_{i=1}^N p_i(E_{k,m}(t)) \ln \frac{p_i(E_{k,m}(t))}{q_i(E_{k,m}(t))} \quad (8.17)$$

$p_i(E_{k,m}(t))$, gürültülü konuşma sinyalinin Teager enerjisinin yaklaşık Gamma olasılık dağılım fonksiyonu olarak ifade edilir ve şu şekilde hesaplanır:

$$p_i(E_{k,m}(t)) = \frac{n_i}{N_c} \quad (8.18)$$

Burada n_i , i. bin'deki katsayıların sayısını, N_c ise her bir kipteki katsayıların toplam sayısını vermektedir.

Benzer şekilde, gürültünün ve temiz konuşma sinyalinin $E_{k,m}(t)$ enerjisinin yaklaşık Gamma olasılık dağılım fonksiyonu Eşitlik (8.18) ile kestirilir ve sırası ile $q_i(E_{k,m}(t))$ ve $r_i(E_{k,m}(t))$ olarak sembolize edilir. Gürültülü konuşma sinyalinin enerjisinin belli bir λ değeri altında, $p_i(E_{k,m}(t))$ ve $q_i(E_{k,m}(t))$ arasındaki Simetrik K-L uzaklığı yaklaşık olarak sıfır olmaktadır. İki olasılık dağılımı arasındaki Simetrik K-L uzaklığı şu şekilde hesaplanır:

$$SKL(p_i(E_{k,m}(t)), q_i(E_{k,m}(t))) \approx 0 \quad (8.19)$$

$p_i(E_{k,m}(t))$ için genelleştirilmiş Gamma olasılık dağılım fonksiyonu şu şekilde ifade edilebilir (Erkelens et al., 2007; Sun, 2016)

$$p_i(E_{k,m}(t)) = \frac{\gamma\beta^v}{\Gamma(v)} \left| \frac{x}{\sigma_s} \right|^{v-1} \exp\left(-\beta \left| \frac{x}{\sigma_s} \right|^v\right) \quad (8.20)$$

Burada $\Gamma(.)$ Gamma fonksiyonunu, β önsel SNR ile ilişkili olan ölçek parametresini (scale parameter) ve v ise genelleştirilmiş Gamma fonksiyonunun şekil parametresini (shape parameter) temsil etmektedir.

Eşitlik (8.20)'de γ , β ve v parametreleri her bir kipteki İKF katsayıları için olasılık dağılımını önemli ölçüde etkilemektedir. γ parametresi genellikle 1 ya da 2 olarak seçilmektedir (Sun, 2016). Bu çalışma için γ ve β parametreleri 1, v şekil parametresi ise 2 olarak seçilmiştir. Bu durumda $p_i(E_{k,m}(t))$ ve $q_i(E_{k,m}(t))$ için Gamma olasılık dağılım fonksiyonu şu şekilde ifade edilebilir:

$$p_i(E_{k,m}(t)) = \frac{x}{\sigma_s} \exp\left(-\frac{x}{\sigma_s}\right) = \frac{x}{\sqrt{\sigma_r^2 + \sigma_n^2}} \exp\left(-\frac{x}{\sqrt{\sigma_r^2 + \sigma_n^2}}\right) \quad (8.21)$$

$$q_i(E_{k,m}(t)) = \frac{x}{\sigma_n} \exp\left(-\frac{x}{\sigma_n}\right) \quad (8.22)$$

Burada σ_r^2 temiz konuşma sinyalinin gücü, σ_n^2 gürültünün gücü ve σ_s^2 ise gürültülü konuşma sinyalinin gücü $\sigma_s^2 = \sigma_r^2 + \sigma_n^2$ olduğu göz önünde bulundurularak sembolize edilmiştir.

Eşik değeri λ , Eşitlik (8.21) ve (8.22) kullanılarak elde edilmektedir:

$$\int_0^\lambda \left[\frac{x}{\sigma_s} \exp\left(-\frac{x}{\sigma_s}\right) - \frac{x}{\sigma_n} \exp\left(-\frac{x}{\sigma_n}\right) \right] I_1 dx = 0 \quad (8.23)$$

$$I_1 = \ln\left(\frac{\sigma_s}{\sigma_n}\right) \exp\left(-\frac{x}{\sigma_s} + \frac{x}{\sigma_n}\right)$$

Eşitlik (8.23) çözülerek λ eşik değerleri şu şekilde türetilebilir:

$$\lambda(k) = \frac{\sigma_n(k) \sqrt{1 + \gamma(k)}}{\ln \sqrt{1 + \gamma(k)}} \quad (8.24)$$

Burada $\gamma(k)$, k. AKA kiplerinin bölütsel SNR değerlerini vermektedir ve şu şekilde ifade edilir:

$$\gamma(k) = \frac{\sigma_r^2(k)}{\sigma_n^2(k)} \quad (8.25)$$

8.2.5 Önerilen Yöntem için Kullanılan Eşikleme Fonksiyonu

Gürültülü konuşmanın iyileştirilmesinde kullanılabilen sert eşikleme, mutlak değeri eşik değerinin altında olan katsayıları sıfıra götürmektedir. Bu durumda, eşik değerinden daha büyük gürültü katsayılarının olabileceği göz ardı edilmekte ve böylece iyileştirilmiş konuşma izgesinde zaman-frekans süreksizlikleri meydana gelmektedir. Sert eşikleme fonksiyonunun aksine yumuşak eşikleme fonksiyonu, eşik değerine dayalı olarak işlenen ve sıfırlanan katsayılar arasında düzgün geçişler yaparak konuşma sinyallerini farklı bir şekilde işlemektedir. Eşitlik (8.24) ile belirlenen eşik λ_1 olarak tanımlanması ile yumuşak eşikleme fonksiyonu, Eşitlik (8.26)'da gösterildiği gibi $\hat{I}KF_{k,m}(t)$ 'nin k. kipinin m. çerçevesine uygulanmaktadır.

$$\tilde{f}_{k,m}(t)_s = \begin{cases} |\hat{I}KF_{k,m}(t)| - \lambda_1(k), & \text{eğer } |\hat{I}KF_{k,m}(t)| \geq \lambda_1(k) \\ 0, & \text{eğer } |\hat{I}KF_{k,m}(t)| < \lambda_1(k) \end{cases} \quad (8.26)$$

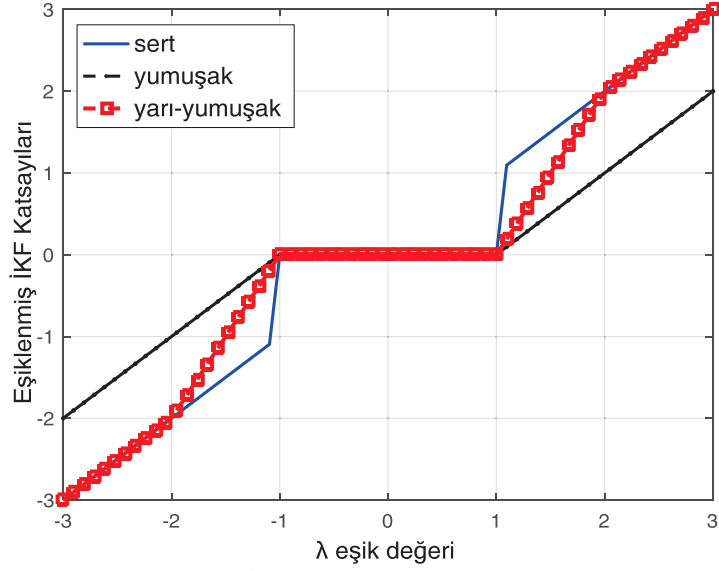
Yumuşak eşikleme, gürültünün bileşenlerini sıfırlamak ve gürültülü konuşma bileşenleri üzerinde bir büyüklük çıkarımı yapmaktadır. Yumuşak eşiklemenin zaman-frekans süreksizliğini ortadan kaldırdığı ve daha pürüzsüz sinyaller verdiği kanıtlanmıştır. Ancak yumuşak eşiklemede $\dot{IKF}_{k,m}(t)$ katsayıları $\lambda_1(k)$ kadar kaydırılmaktadır. $\dot{IKF}_{k,m}(t)$ gürültü seviyesi dışında olduğunda, böyle bir kaymanın meydana gelmesi iyileştirilmiş konuşmanın izgesinde gereksiz bir kayma (bias) yaratmaktadır. Meydana gelen kayma problemlerinin üstesinden gelebilmek için yarı-yumuşak eşikleme fonksiyonu kullanılmaktadır ve bu durumda katsayıların eşik değerinin miktarına göre kayması önlenmiş olmaktadır. Tüm eşikleme fonksiyonlarının üstünlükleri ve sakıncaları göz önüne alınarak, gürültülü konuşma sinyalinin $\dot{IKF}_{k,m}(t)$ katsayıları üzerinde yarı-yumuşak eşikleme fonksiyonu uygulanmıştır.

$\lambda_2(k) = \sqrt{2} \lambda_1(k)$ şeklinde alındığında yarı-yumuşak eşikleme fonksiyonu Eşitlik (8.27) ile ifade edilebilir:

$$\tilde{f}_{k,m}(t)_{SS} = \begin{cases} 0 & \text{eğer } |\dot{IKF}_{k,m}(t)| \leq \lambda_1(k) \\ \dot{IKF}_{k,m}(t) & \text{eğer } |\dot{IKF}_{k,m}(t)| > \lambda_2(k) \\ \text{sgn}(\dot{IKF}_{k,m}(t)) \cdot G & \text{diğer durumlarda} \end{cases} \quad (8.27)$$

$$G = \left\lceil \frac{\lambda_2(k) |\dot{IKF}_{k,m}(t)| - \lambda_1(k)}{\lambda_2(k) - \lambda_1(k)} \right\rceil \quad (8.28)$$

Burada $\tilde{f}_{k,m}(t)$, yarı-yumuşak eşiklenmiş AKA kiplerinin $\dot{IKF}_{k,m}(t)$ katsayılarını göstermektedir. Şekil 8.6'da sert, yumuşak ve yarı-yumuşak eşikleme fonksiyonları için giriş $\dot{IKF}$ katsayıları ve çıkış eşiklenmiş $\dot{IKF}$ katsayıları ilişkisi gösterilmektedir.



Şekil 8.6 Sert, yumuşak ve yarı-yumuşak eşikleme fonksiyonları için giriş-çıkış ilişkisi

8.2.6 Ters Ampirik Kip Ayırıştırması

İyileştirilmiş konuşma sinyali $\hat{s}[n]$, çerçeve bazında eşiklenmiş İKF katsayılarının ters AKA dönüşümü ile elde edilmektedir.

$$\hat{s}[n] = \text{AKA}^{-1} \left(\text{İKF}_{k,m}(t) \right)_{ss} \quad (8.29)$$

Son olarak, iyileştirilmiş konuşma sinyali standart örtüşme ve ekleme (overlap-and-add) yöntemi kullanılarak geri çatılır ve iyileştirme işlemi tamamlanmış olur.

8.3 Yöntemlerin Kalite ve Anlaşılabilirlik Bakımından Değerlendirilmesi

Bu bölümde, İngilizce ve Türkçe için önerilen konuşma iyileştirme yönteminin beyaz, araba, mırıltı, hava alanı, sergi, sokak ve tren gürültüsü altında -10 dB'den 10 dB'e kadar farklı SNR seviyeleri için nesnel kalite ve anlaşılabilirlik ölçütleri ile değerlendirmesi yapılmıştır. Nesnel kalite ölçütü olarak Bölütsel SNR iyileştirme miktarı (BSNR) ve Ağırlıklı İzgesel Eğim (AİE), nesnel anlaşılabilirlik ölçütü olarak da Konuşma Kalitesinin Algısal Değerlendirmesi (KKAD) ve Kısa-zamanlı Nesnel Anlaşılabilirlik (KZNA) ölçütleri kullanılmıştır. Önerilen yöntem, konuşma iyileştirme alanında sık kullanılan Dalgacık Dönüşümü ve Ampirik Kip Ayırıştırması tabanlı global

eşikleme yöntemleri, Çoklu Bant İzgesel Çıkarım (ÇBİÇ), ortalama karesel hatayı en aza indiren istatistiksel bir yöntem olan EOKK ve bir zaman alanı yöntemi olan Algısal KLD (AKLD) yöntemleri ile karşılaştırılarak performans değerlendirilmesi yapılmıştır.

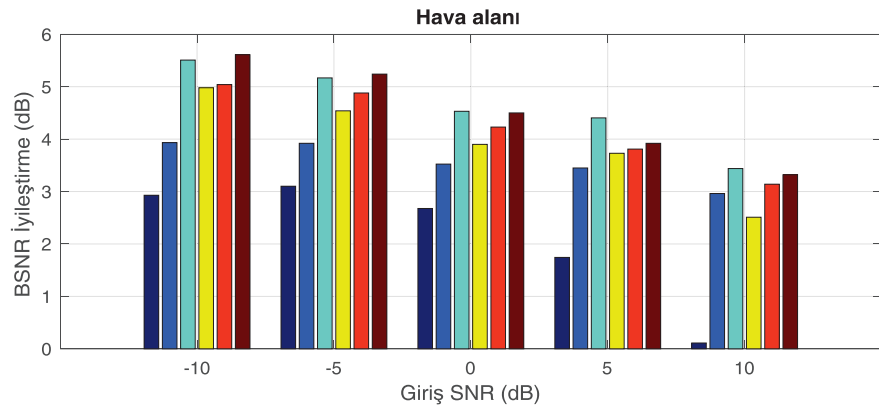
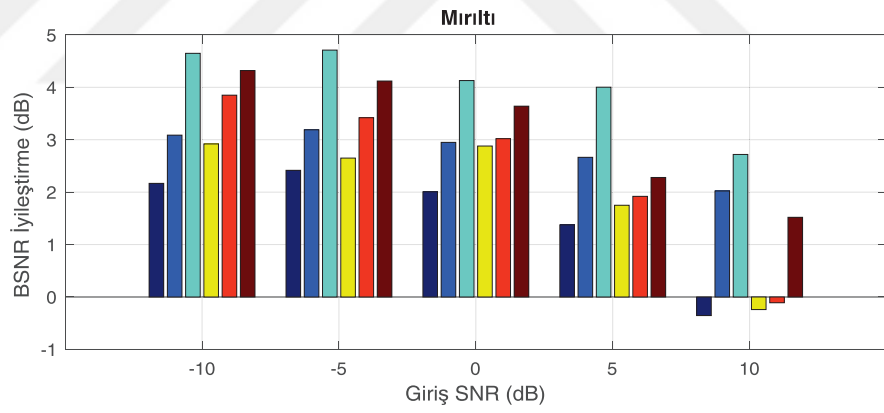
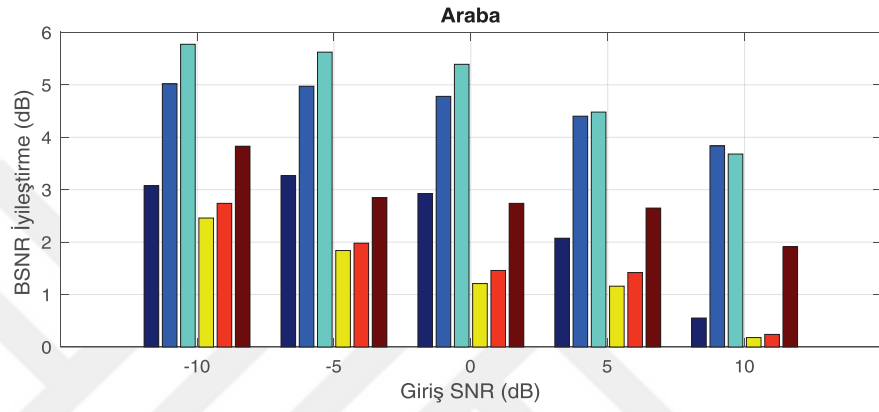
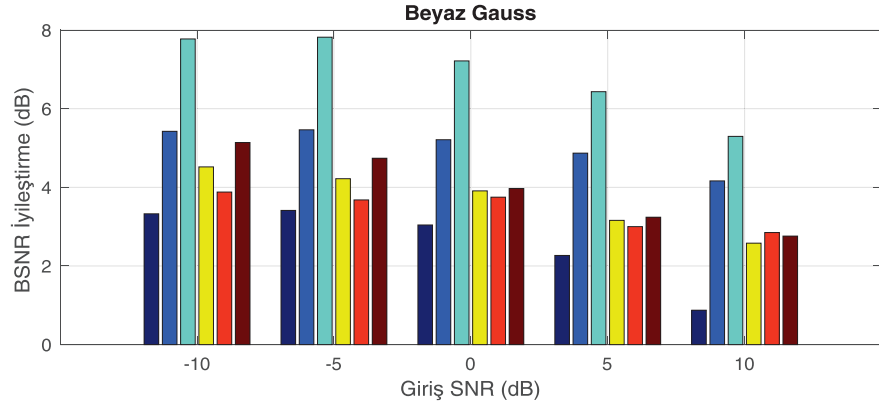
İngilizce için konuşma cümleleri NOIZEUS veri tabanından ve Türkçe için ise Middle East Technical University Turkish Microphone Speech v1.0 (ODTÜ) veri tabanından (Salor et al., 2007) alınan 30'ar cümleden oluşmaktadır. Gürültülü konuşma cümlelerini oluşturmak için temiz cümlelere -10 dB'den 10 dB'e kadar beş farklı SNR seviyelerinde beyaz, araba, mırıltı, hava alanı, sergi, sokak ve tren gürültüleri eklenmiştir. Temiz konuşma sinyallerini bozmak için kullanılan gürültüler NOISEX-92 (Varga and Steeneken, 1993) veri tabanından alınmıştır.

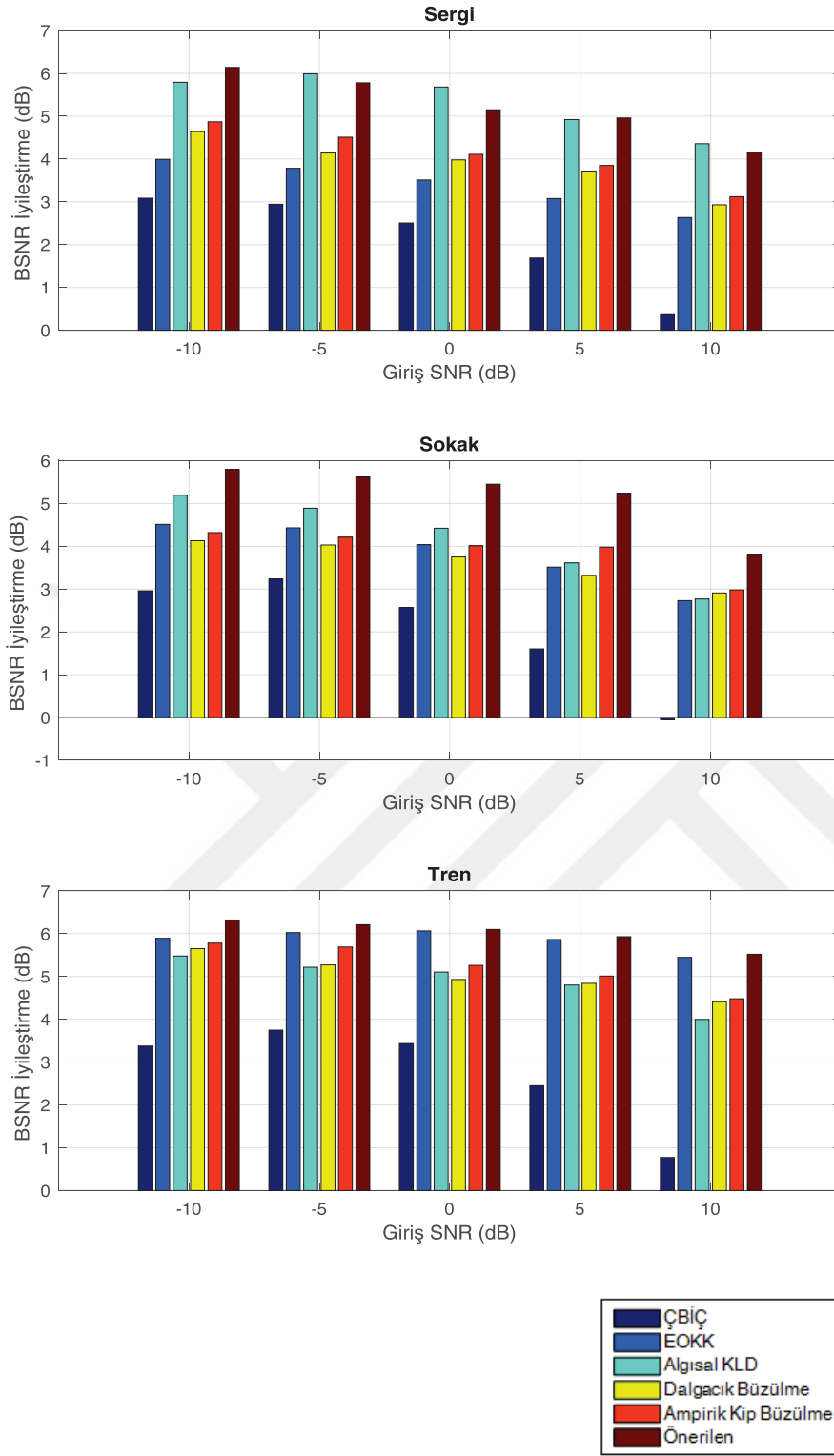
Çerçevelere ayrılan konuşma sinyallerinin analizinde %50 örtüşme olan 64 ms uzunluklu Hamming pencere işlevi kullanılmıştır. Hem gürültülü konuşma hem de gürültü sinyali Ampirik Kip Ayırıştırması yöntemi ile 7 içkin kip fonksiyonuna (İKF) ayrıştırılmıştır. Gürültülü konuşma ve gürültünün Gamma olasılık dağılım işlevinin elde edilmesinde γ ve β parametreleri 1, v şekil parametresi ise 2 olarak alınmıştır. Karşılaştıma amacıyla kullanılan bir diğer yöntem olan dalgacık büzülme yaklaşımında dalgacık işlevi olarak db8 dalgacığı kullanılmış olup, gürültülü konuşma sinyalleri sekiz seviye açılmıştır (Ghanbari and Karami-Mollaei, 2006; Xu, 2007). Gürültü bileşenlerinin eşiklenmesinde, önerilen yöntem için yarı-yumuşak, diğer yöntemler için ise yumuşak eşikleme fonksiyonu kullanılmıştır.

Bu çalışmada, önerilen konuşma iyileştirme yaklaşımı hem İngilizce hem de Türkçe konuşmalar için diğer yöntemler ile nesnel kalite ve anlaşılabilirlik ölçütleri bakımından karşılaştırılmıştır. Ayrıca, tüm yöntemlerin öznel olarak değerlendirilmesi amacıyla temiz, gürültülü ve iyileştirilmiş konuşma sinyallerinin spektrogram gösterimleri incelenmiştir.

8.3.1 İngilizce İçin Önerilen Yöntemin Kalite ve Anlaşılabilirlik Ölçütleri İle Değerlendirilmesi

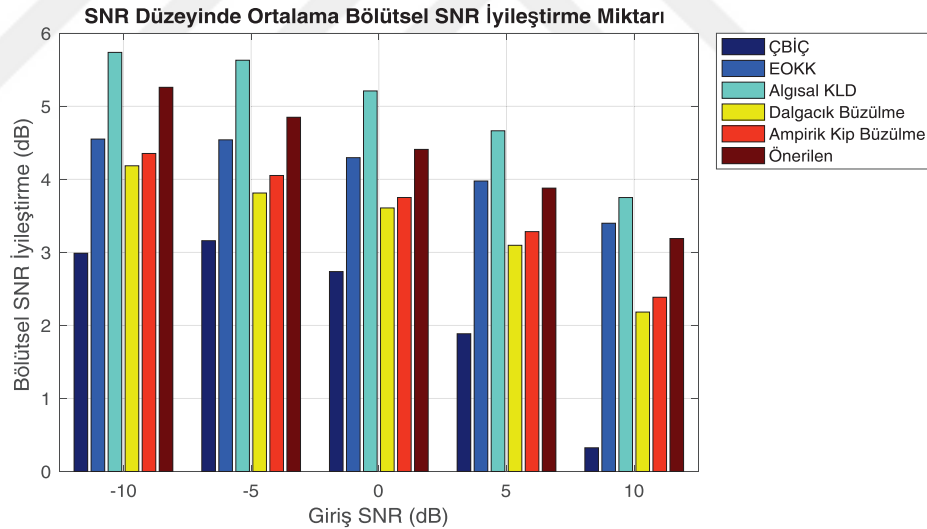
Farklı gürültü tipi koşullarında -10 dB'den 10 dB'e kadar beş farklı SNR seviyelerinde bozulan gürültülü İngilizce konuşma sinyallerinin iyileştirilmesinde, önerilen yöntemin performansı Bölütsel SNR (BSNR) iyileştirme miktarı bakımından diğer yöntemler ile Şekil 8.7'de karşılaştırılmıştır.





Şekil 8.7 İngilizce için tüm gürültülü koşullara ait Bölütsel SNR iyileştirme miktarları

Şekil 8.7'den görülebileceği üzere önerilen ve diğer yöntemlerin bölütsel SNR iyileştirme miktarları gürültü tipi ve SNR seviyesine göre değişkenlikler göstermektedir. Bu nedenle, öncelikle tüm gürültüler için her SNR seviyesi için BSNR iyileştirme miktarlarının ortalaması alınmıştır. Her bir SNR seviyesi için gürültülerin ortalaması alınarak elde edilen BSNR iyileştirme miktarı değerleri Şekil 8.8'de verilmiştir. Elde edilen ortalama sonuçlara göre, tüm SNR seviyelerinde Algısal KLD (AKLD) yöntemi en fazla bölütsel SNR iyileştirme miktarını sağlamaktadır. Algısal KLD yönteminin bölütsel SNR iyileştirme miktarları; -10 dB SNR için 5,73 dB, -5 dB SNR için 5,63 dB, 0 dB SNR için 5,21 dB, 5 dB SNR için 4,66 dB ve 10 dB SNR için 3,75 dB. Bölütsel SNR iyileştirme miktarı bakımından Algısal KLD yöntemini önerilen yöntem takip etmektedir. Önerilen yöntemin bölütsel SNR iyileştirme miktarları -10 dB SNR seviyesinde 5,26 dB, -5 dB SNR'da 4,85 dB, 0 dB SNR'da 4,41 dB, 5 dB SNR'da 3,88 dB ve 10 dB SNR'da 3,19 dB artış göstermektedir. Ayrıca SNR seviyesi yükseldikçe bölütsel SNR iyileştirme miktarları (dB) düşmektedir. Bölütsel SNR iyileştirme miktarı bakımından iyileştirilmiş konuşmaların kalitesini sağlamada Çoklu Bant İzgesel Çıkarım (ÇBİÇ) yöntemi en kötü performansı göstermektedir.

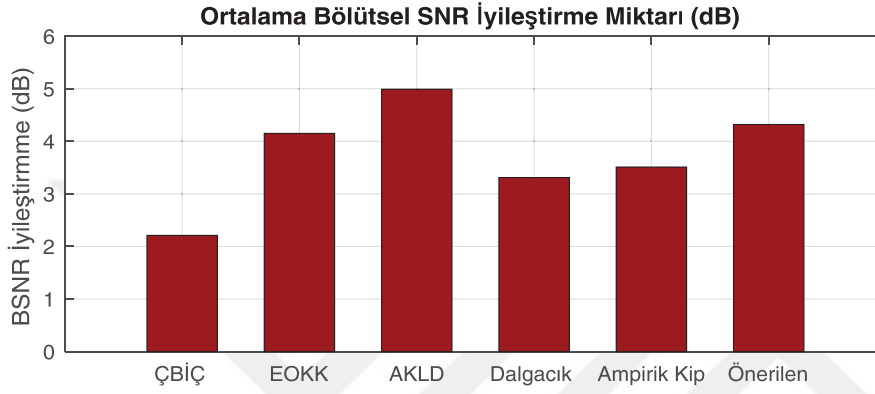


Şekil 8.8 Gürültülerin ortalaması alınarak SNR seviyesine göre yöntemlerin Bölütsel SNR iyileştirme miktarları

Konuşma iyileştirme yöntemlerinin nesnel kalite değerlendirmesinin gürültü tipi ve SNR seviyesinden bağımsız yapılması amacıyla tüm gürültü ve tüm SNR seviyelerinin ortalaması alınarak elde edilen bölütsel SNR iyileştirme miktarları Şekil 8.9'da verilmiştir. Şekil 8.9'dan görülebileceği üzere, Algısal KLD yöntemi 4,99 dB BSNR iyileştirme miktarı sağlarken, önerilen yöntem 4,32 dB iyileştirme

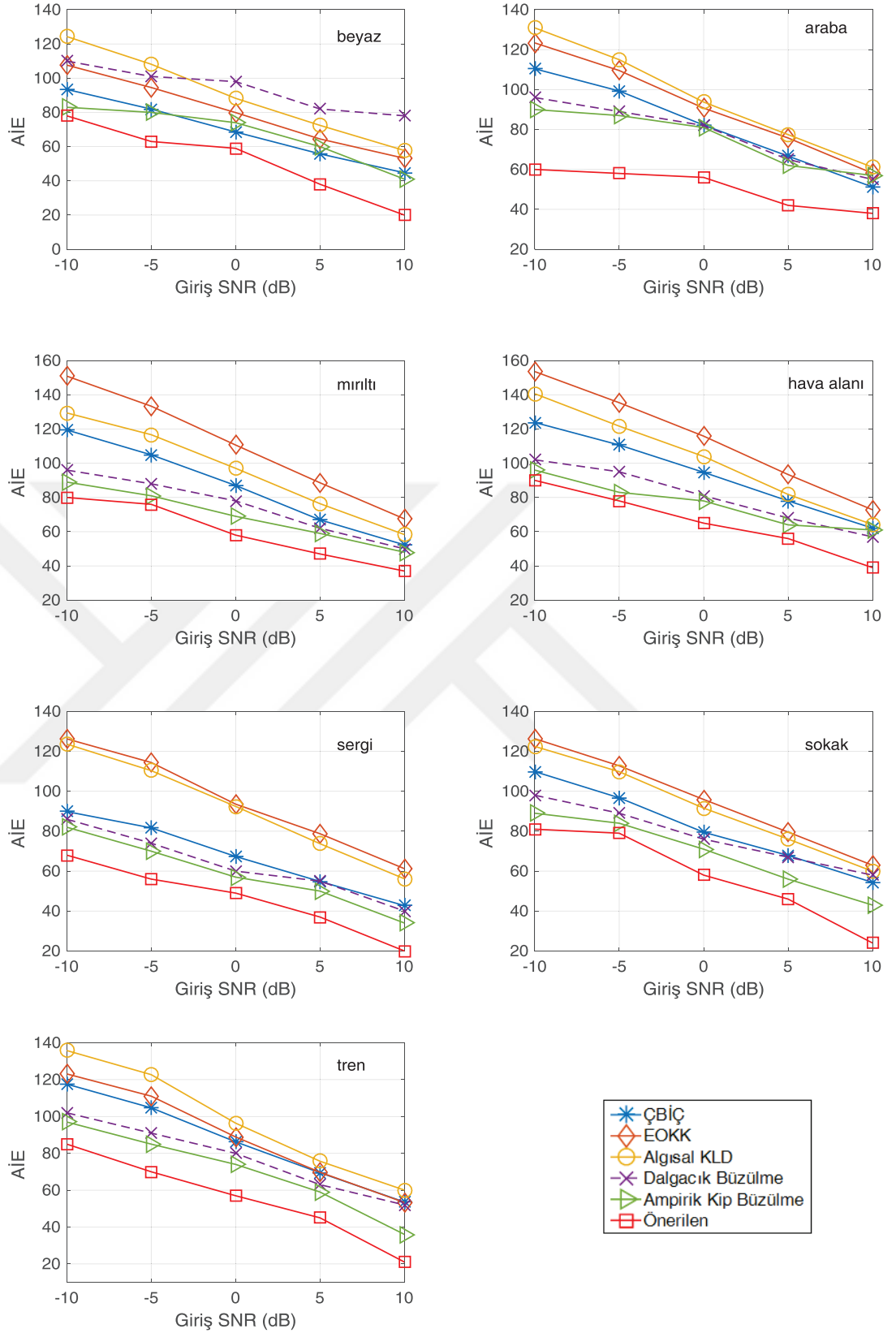
miktarı sağlamaktadır. Ayrıca, EOKK yöntemi 4,15 dB, Ampirik Kip Ayırıştırması büzülme yöntemi 3,51 dB, Dalgacık büzülme yöntemi 3,31 dB ve ÇBİÇ yöntemi 2,21 dB bölütsel SNR iyileştirme sağlamaktadır.

Tüm sonuçlar göz önünde bulundurulduğunda, önerilen konuşma iyileştirme yaklaşımı Algısal KLD (AKLD) yöntemi dışında diğer yöntemlere göre daha fazla bölütsel SNR iyileştirme miktarı sağlamaktadır ve bu sonuçlar önerilen yöntemin tatmin edici bölütsel SNR iyileştirmesi sağladığını göstermektedir.



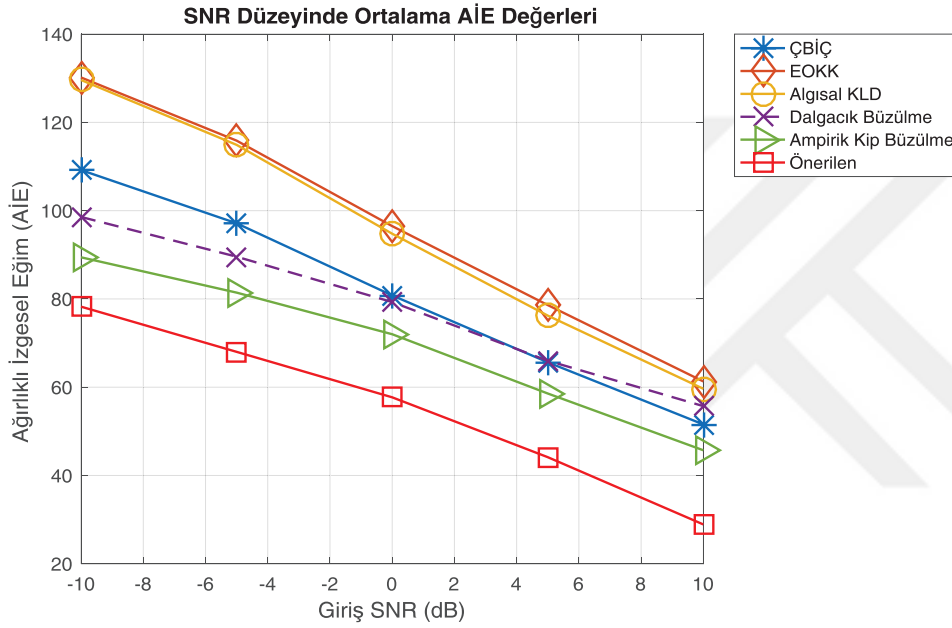
Şekil 8.9 Gürültülerin ve SNR seviyelerinin ortalaması alınarak yöntemlere ait Bölütsel SNR iyileştirme miktarları

İyileştirilmiş konuşma sinyallerinin nesnel kalite değerlendirmesinde kullanılan bir diğer ölçüt olan Ağırlıklı İzgesel Eğim (AİE) sonuçları tüm gürültü tipi ve tüm SNR seviyeleri için Şekil 8.10'da verilmiştir. Şekil 8.10'dan görülebildiği üzere, önerilen yöntem tüm gürültü tipi ve tüm SNR seviyelerinde en düşük AİE değerlerini vermektedir. Bu nedenle, önerilen yöntem diğer yöntemler ile karşılaştırıldığında AİE nesnel kalite ölçütüne göre konuşma iyileştirmede en etkili yöntem olarak bulunmaktadır. Kullanılan diğer yöntemlerin AİE kalite performansına bakıldığında ise EOKK ve AKLD yöntemleri en etkisiz yöntemler olarak tespit edilmiştir.



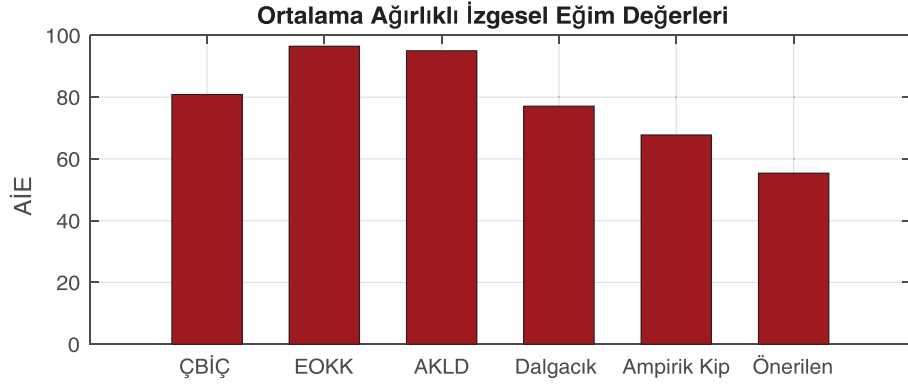
Şekil 8.10 İngilizce için tüm gürültülü koşullara ait AİE değerleri

Tüm gürültü tiplerinin her bir SNR seviyesi için ortalama AİE değerleri Şekil 8.11’de gösterilmektedir. Elde edilen ortalama sonuçlara göre, önerilen yöntem ile iyileştirilmiş konuşmaların AİE değerleri; -10 dB SNR seviyesinde 78,24, -5 dB SNR’da 68,01, 0 dB’de 57,70, 5 dB SNR’da 44,14 ve 10 dB SNR seviyesinde 28,91 şeklinde bulunmaktadır. Önerilen yöntem diğer yöntemler ile karşılaştırıldığında en düşük AİE değerlerine sahip olduğundan nesnel kalite bakımından en etkili yöntem olarak belirlenmiştir. En yüksek AİE değerlerine sahip EOKK yöntemi ise konuşma iyileştirmede kaliteyi sağlamada en kötü performansı göstermiştir.



Şekil 8.11 Gürültülerin ortalaması alınarak SNR seviyesine göre yöntemlerin AİE değerleri

Ağırlıklı İzgesel Eğim (AİE) nesnel kalite ölçütüne göre yöntemlerin genel performansını ölçmek amacıyla tüm gürültü ve tüm SNR seviyeleri için ortalama AİE değerlerine bakılmıştır. Şekil 8.12’de verilen ortalama AİE değerlerine göre, önerilen yöntem 55,40 ile en düşük AİE değerlerine sahip olmaktadır. Önerilen yöntemi sırasıyla Ampirik Kip büzülme (67,77), Dalgacık büzülme (77,08), ÇBİÇ (80,86), AKLD (95,02) ve EOKK (96,49) yöntemleri takip etmektedir. AİE nesnel kalite ölçütü bakımından en yüksek değerlere sahip EOKK yöntemi diğer yöntemlere göre en kötü performansı göstermiştir.



Şekil 8.12 Gürültülerin ve SNR seviyelerinin ortalaması alınarak yöntemlere ait AİE değerleri

Tüm gürültü koşullarında farklı beş SNR seviyesi ve bu SNR seviyelerinin ortalaması alınarak elde edilen tüm yöntemlere ait KKAD skorları Çizelge 8.1’de verilmiştir. Çizelge 8.1’den görülebileceği üzere beyaz, araba, mırıltı, hava alanı, sergi, sokak ve tren gürültü koşullarında önerilen AKA tabanlı konuşma iyileştirme yaklaşımı en yüksek KKAD değerlerine sahiptir. Her gürültü tipi için farklı SNR seviyelerinde KKAD değerleri değişkenlik gösterdiğinden beş SNR seviyesi için bulunan KKAD skorlarının ortalaması alınarak gürültü bazında en iyi konuşma iyileştirme yönteminin bulunması amaçlanmıştır. SNR seviyelerinin ortalaması alınarak elde edilen sonuçlara göre; beyaz gürültü koşullarında önerilen yöntemin KKAD skoru 2,11 iken araba için 2,39, mırıltı için 2,40, hava alanı için 2,80, sergi için 2,64, sokak için 2,86 ve tren gürültüsü için 3,17 olarak bulunmuştur. KKAD skorları bakımından önerilen yöntemi birçok gürültü koşulu için Ampirik kip büzülme ve Dalgacık büzülme yöntemleri takip etmektedir.

İngilizce konuşma sinyalleri için KKAD anlaşılabilirlik ölçütü bakımından en iyi ve etkili yöntemin tespit edilmesi amacıyla tüm gürültü tipi ve tüm SNR seviyeleri için ortalama KKAD değerlerine bakılmıştır. Elde edilen sonuçlara göre 2,62 KKAD skoruna sahip önerilen yöntem konuşma iyileştirmede anlaşılabilirlik bakımından daha başarılı bulunmuştur. KKAD sonuçlarına göre en iyi yöntem olarak belirlenen önerilen yöntemi sırasıyla 2,29, 2,11 ve 2,07 KKAD skorlarına sahip Ampirik kip büzülme, EOKK ve Dalgacık büzülme yöntemleri izlemektedir.

KKAD skoru bakımından iyileştirmede anlaşılabilirliğin sağlanmasında en kötü yöntem olarak 1,54 KKAD skoruna sahip Algısal KLD (AKLD) yöntemi tespit edilmiştir.

Çizelge 8.1 İngilizce konuşma sinyalleri için farklı yöntem ve gürültülere ait KKAD sonuçlarından elde edilen anlaşılabilirlik sonuçları

Ortamlar		Yöntemler					
	SNR	ÇBİÇ	EOKK	AKLD	Dalgacık	AKA	Önerilen
beyaz	10	2,53	2,82	2,46	1,95	2,25	2,97
	5	2,05	2,46	2,17	1,71	1,91	2,53
	0	1,66	2,05	1,77	1,48	1,54	1,91
	-5	1,42	1,61	1,06	1,28	1,33	1,73
	-10	1,18	1,33	0,17	1,13	1,30	1,40
	Ort.	1,77	2,05	1,53	1,51	1,66	2,11

araba	10	2,72	2,91	2,41	1,87	2,42	3,03
	5	2,38	2,57	2,04	1,61	2,11	2,94
	0	1,97	2,16	1,54	1,25	1,73	2,55
	-5	1,62	1,78	0,92	0,89	1,41	1,79
	-10	1,19	1,45	0,46	0,70	1,07	1,64
	Ort.	1,98	2,17	1,47	1,26	1,74	2,39

mırıltı	10	2,64	2,65	2,32	2,68	2,74	1,56
	5	2,35	2,30	1,93	2,52	2,63	1,93
	0	1,98	1,93	1,54	2,01	2,15	2,46
	-5	1,65	1,61	0,99	1,64	1,72	2,87
	-10	1,13	1,24	0,42	1,32	1,41	3,21
	Ort.	1,95	1,95	1,44	2,03	2,13	2,40

hava alanı	10	2,76	2,79	2,47	2,95	2,98	3,38
	5	2,44	2,43	2,07	2,56	2,85	3,12
	0	2,12	2,06	1,64	2,21	2,43	2,99
	-5	1,72	1,67	1,03	1,94	2,18	2,58
	-10	1,42	0,44	0,99	1,48	1,73	1,96
	Ort.	2,09	1,88	1,64	2,23	2,43	2,80

sergi	10	2,55	2,60	2,30	2,89	3,05	3,28
	5	2,14	2,23	1,97	2,71	2,86	3,07
	0	1,76	1,85	1,46	2,07	2,15	2,65
	-5	1,45	1,42	0,66	1,87	1,98	2,21
	-10	1,21	1,06	0,47	1,43	1,71	1,99
	Ort.	1,82	1,83	1,37	2,19	2,35	2,64

sokak	10	2,84	2,98	2,43	3,12	3,27	3,59
	5	2,51	2,62	2,02	2,96	3,04	3,18
	0	2,21	2,26	1,61	2,51	2,73	2,95
	-5	1,84	1,89	1,02	1,98	2,24	2,53
	-10	1,37	1,49	0,47	1,56	1,85	2,04
	Ort.	2,15	2,25	1,51	2,42	2,63	2,86

tren	10	3,06	3,36	2,71	3,29	3,72	3,85
	5	2,78	2,97	2,32	3,15	3,36	3,42
	0	2,44	2,61	1,92	2,86	2,99	3,04
	-5	2,09	2,28	1,47	2,64	2,85	2,93
	-10	1,76	1,95	0,78	2,29	2,56	2,61
	Ort.	2,43	2,63	1,84	2,84	3,09	3,17

Ortalama Sonuçlar	2,03	2,11	1,54	2,07	2,29	2,62
--------------------------	-------------	-------------	-------------	-------------	-------------	-------------

Farklı gürültü koşulları ve farklı SNR seviyeleri için iyileştirilmiş İngilizce konuşma sinyallerinin anlaşılabilirlik değerlendirmesi bir diğer nesnel anlaşılabilirlik ölçütü olan Kısa-zamanlı Nesnel Anlaşılabilirlik (KZNA) ölçütü kullanılarak yapılmıştır. Tüm gürültü tipleri ve tüm SNR seviyeleri için önerilen ve diğer konuşma iyileştirme yöntemlerine ait KZNA nesnel anlaşılabilirlik değerleri Çizelge 8.2’de verilmektedir. Elde edilen KZNA anlaşılabilirlik ölçütü sonuçlarına göre, tüm gürültü tipi ve SNR seviyeleri için önerilen yöntem en yüksek yüzdelere sahip olmaktadır. SNR seviyelerinin ortalaması alınarak gürültü bazında bulunan önerilen yönteme ait KZNA sonuçları şu şekildedir; beyaz Gauss gürültüsü koşullarında % 84,60, araba gürültüsü altında % 80,99, mırıltı gürültüsünde % 80,71, hava alanı % 86,06, sergi % 79,59, sokak % 86,76 ve tren gürültüsü koşullarında % 95,70.

Kullanılan yöntemlerin gürültü tipi ve SNR seviyesinden bağımsız olarak nesnel anlaşılabilirlik değerlendirmesinin yapılması amacıyla tüm gürültü ve SNR seviyeleri için hesaplanan KZNA skorlarının ortalaması alınmıştır. Elde edilen ortalama sonuçlara göre, % 84,92 KZNA anlaşılabilirlik yüzdesine sahip önerilen yöntem diğer yöntemlere göre anlaşılabilirliği sağlamada daha etkili ve başarılı olmuştur. Önerilen yöntemi, % 80,33 KZNA skoru ile Ampirik kip büzülme, % 77,86 ile Dalgacık büzülme, % 74,31 ile ÇBİÇ, % 69,98 ile EOKK ve son olarak % 64,14 ile Algısal KLD (AKLD) yöntemleri takip etmektedir.

Çizelge 8.2 İngilizce konuşma sinyalleri için farklı yöntem ve gürültülere ait KZNA sonuçlarından elde edilen anlaşılabilirlik oran tahminleri (%)

Ortamlar		Yöntemler					
	SNR	ÇBİÇ	EOKK	AKLD	Dalgacık	AKA	Önerilen
beyaz	10	99,33	99,40	99,48	99,53	99,64	99,88
	5	96,81	97,47	98,08	99,01	99,24	99,56
	0	85,77	86,12	89,78	92,18	93,47	98,83
	-5	54,66	49,52	46,70	73,67	76,65	81,03
	-10	21,51	15,69	4,67	32,39	35,72	43,68
	Ort.	71,62	69,64	67,75	79,36	80,94	84,60

araba	10	99,38	99,35	99,43	99,46	99,47	99,74
	5	97,37	97,23	97,59	98,12	98,24	98,85
	0	88,16	85,01	83,08	89,37	90,53	92,35
	-5	57,09	51,39	22,62	64,08	67,42	75,10
	-10	20,73	14,30	1,86	29,61	34,67	38,94
	Ort.	72,55	69,46	60,92	76,12	78,06	80,99

mırılta	10	99,29	99,16	99,33	99,36	99,49	99,83
	5	96,79	94,97	96,56	97,02	98,05	99,07
	0	82,11	71,54	79,83	80,95	85,14	91,67
	-5	34,89	21,55	17,75	42,83	53,29	70,85
	-10	8,35	4,23	1,14	17,61	21,05	42,17
	Ort.	64,28	58,29	58,92	67,55	71,40	80,71

hava alanı	10	99,53	99,47	99,53	99,52	99,63	99,85
	5	98,23	97,61	97,89	98,25	98,91	99,04
	0	91,80	86,13	86,57	91,87	92,17	98,27
	-5	64,50	47,10	27,30	72,28	76,38	82,03
	-10	23,81	11,30	1,71	28,62	32,86	51,14
	Ort.	75,57	68,32	62,60	78,10	79,99	86,06

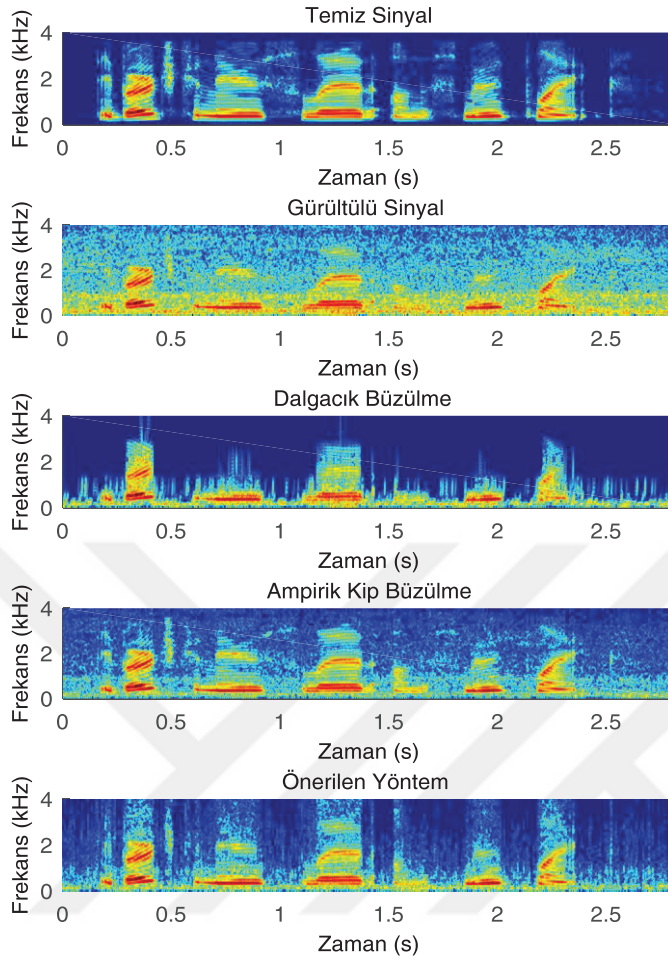
sergi	10	99,38	99,24	99,54	99,61	99,72	99,91
	5	96,78	96,32	98,04	98,64	99,03	99,52
	0	82,95	75,72	80,20	87,18	90,15	93,76
	-5	33,27	21,47	15,17	48,24	53,62	63,79
	-10	11,50	4,02	1,05	21,55	31,06	40,98
	Ort.	64,77	59,36	58,80	71,04	74,71	79,59

sokak	10	99,45	99,29	99,31	99,57	99,64	99,85
	5	98,10	97,40	97,34	98,90	99,07	99,30
	0	93,69	90,98	87,56	94,06	95,31	97,26
	-5	75,72	65,83	31,32	73,12	80,42	85,32
	-10	33,81	23,63	4,37	37,08	40,83	52,09
	Ort.	80,15	75,43	63,98	80,54	83,05	86,76

tren	10	99,78	99,79	99,74	99,80	99,89	99,96
	5	99,40	99,35	99,14	99,32	99,58	99,71
	0	97,87	97,23	95,92	98,75	99,05	99,47
	-5	91,65	88,94	73,85	92,50	93,81	96,35
	-10	67,52	61,53	11,51	71,27	78,43	83,02
	Ort.	91,25	89,37	76,03	92,32	94,15	95,70

Ortalama Sonuçlar	74,31	69,98	64,14	77,86	80,33	84,92
--------------------------	--------------	--------------	--------------	--------------	--------------	--------------

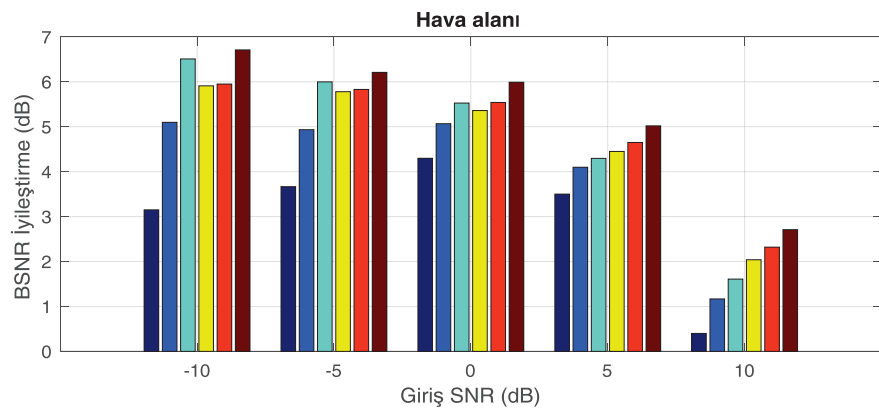
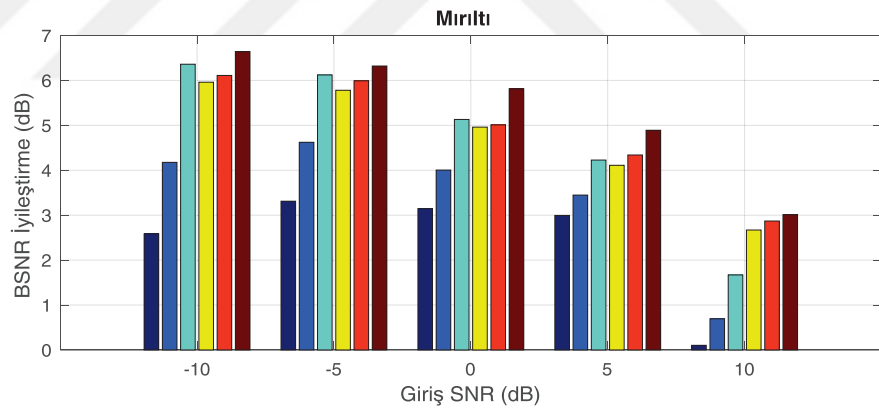
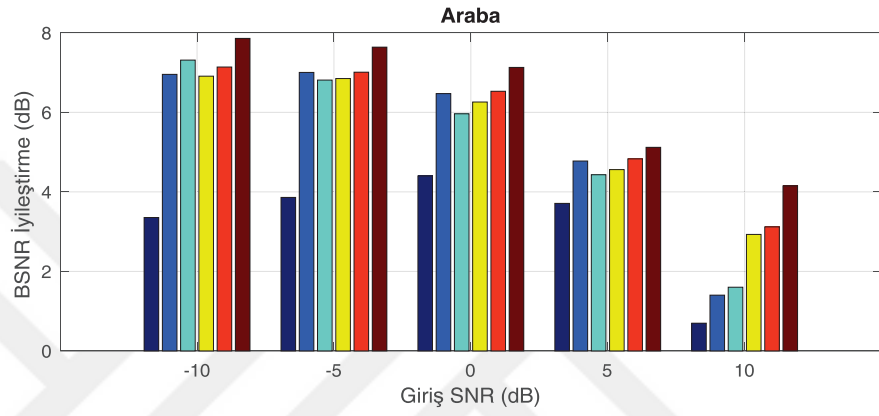
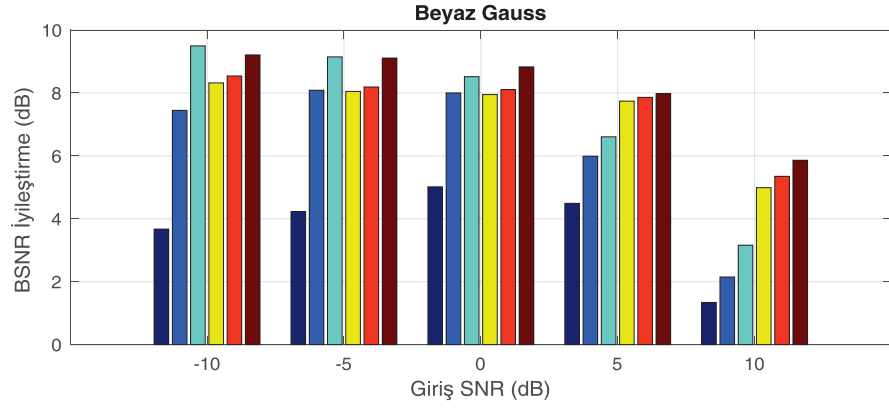
İyileştirilmiş konuşma sinyallerinin öznel kalite değerlendirmesinin yapılması amacıyla, temiz konuşma, araba gürültüsü ile bozulmuş 5 dB SNR seviyesinde gürültülü konuşma, önerilen ve diğer yöntemler kullanılarak elde edilen iyileştirilmiş konuşmaların spektrogramları Şekil 8.13’de verilmiştir. Konuşma iyileştirmede kullanılan Dalgacık büzülme yönteminde gürültünün yanında konuşmanın yüksek frekanslı bileşenlerinin de kaybolduğu görülmektedir. Konuşmanın yüksek frekanslı bileşenlerinin gürültü ile birlikte atılması konuşmanın anlaşılabilirliğini Çizelge 8.1 ve 8.2’de gösterildiği gibi olumsuz etkilemektedir. Kullanılan bir diğer yöntem olan Ampirik Kip büzülme yaklaşımında, konuşmanın tüm bileşenleri mümkün olduğunca korunurken gürültünün yeterince atılmadığı gözlemlenmektedir. Şekil 8.13’de verilen spektrogramlardan hareketle, önerilen yöntem diğer yöntemlere göre konuşma sinyalinin hem düşük hem de yüksek frekanslı bileşenlerini korurken, konuşmayı bozan gürültünün büyük miktarda azaldığı görülmektedir.

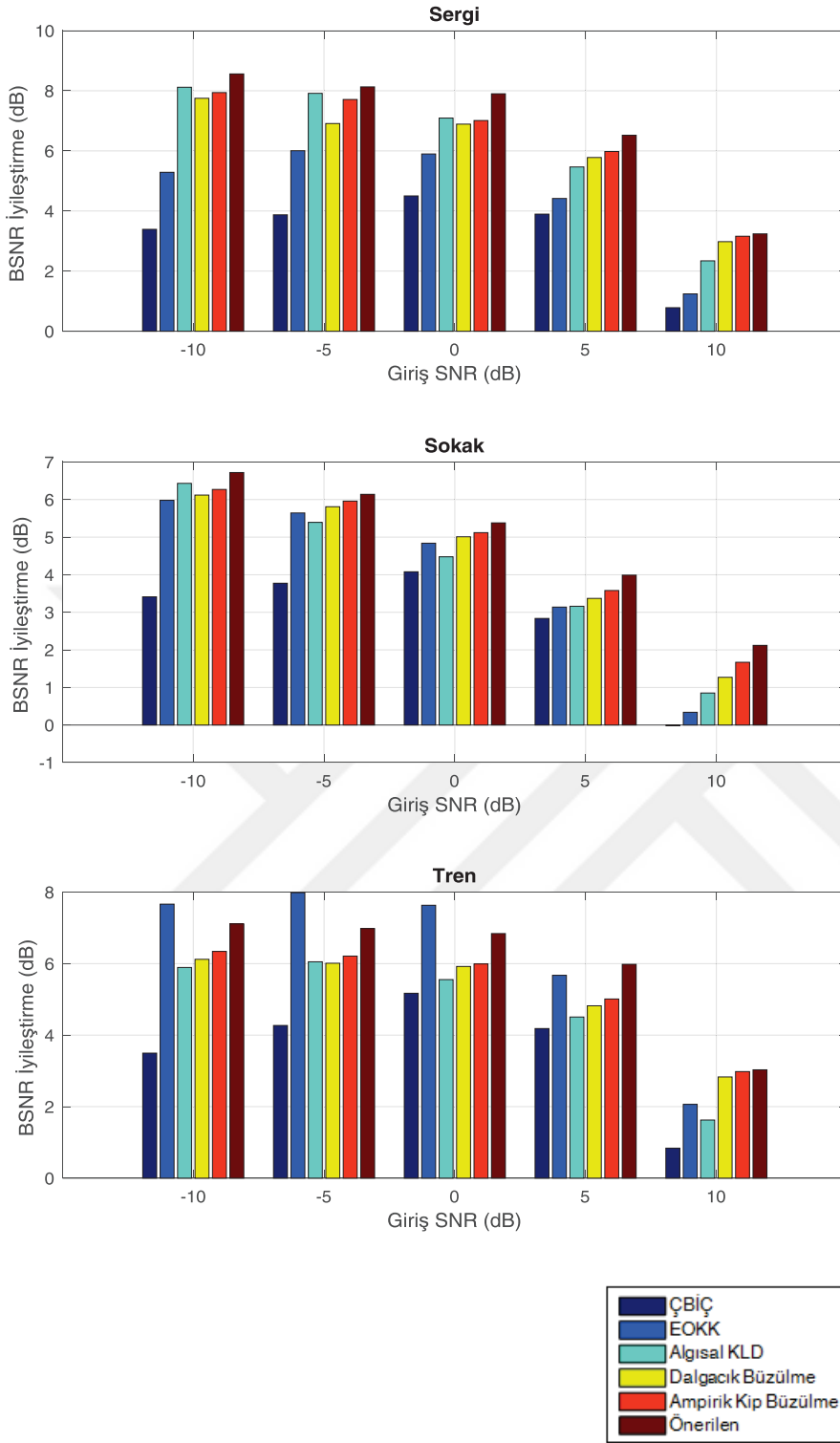


Şekil 8.13 İngilizce için; temiz sinyalin, 5 dB araba gürültüsü ile bozulmuş gürültülü sinyalin, önerilen ve diğer yöntemler ile iyileştirilmiş sinyallerin spektrogramları

8.3.2 Türkçe İçin Önerilen Yöntemin Kalite ve Anlaşılabilirlik Ölçütleri İle Değerlendirilmesi

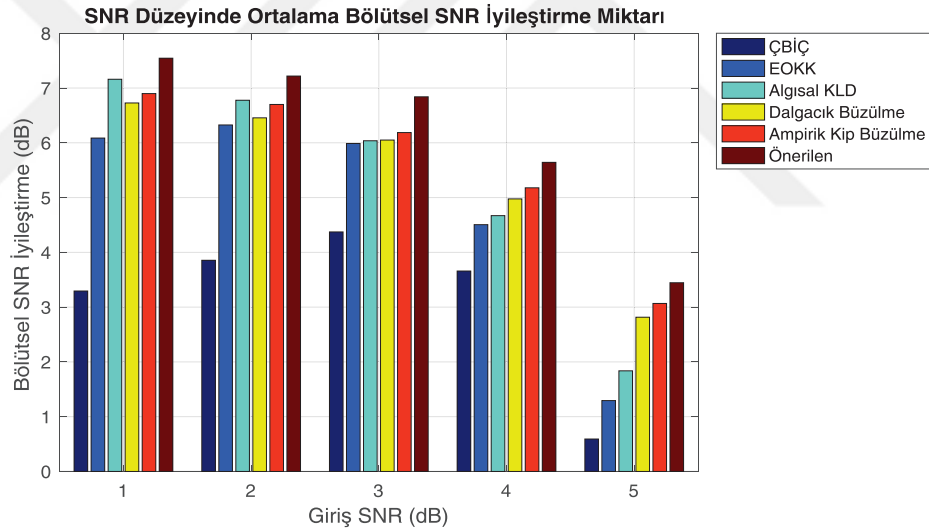
Farklı gürültü tipi koşullarında -10 dB'den 10 dB'e kadar beş farklı SNR seviyelerinde bozulan gürültülü Türkçe konuşma sinyallerinin iyileştirilmesinde, önerilen yöntemin performansı Bölütsel SNR (BSNR) iyileştirme miktarı bakımından diğer yöntemler ile Şekil 8.14'de karşılaştırılmıştır.





Şekil 8.14 Türkçe için tüm gürültü koşullarına ait Bölütsel SNR iyileştirme miktarları

Şekil 8.14'den görülebileceği üzere önerilen ve diğer yöntemlerin bölütsel SNR iyileştirme miktarları gürültü tipi ve SNR seviyesine göre değişkenlikler göstermektedir. Bu nedenle, öncelikle tüm gürültüler için her SNR seviyesi için BSNR iyileştirme miktarlarının ortalaması alınmıştır. Her bir SNR seviyesi için gürültülerin ortalaması alınarak elde edilen BSNR iyileştirme miktarı değerleri Şekil 8.15'de verilmiştir. Elde edilen ortalama sonuçlara göre, tüm SNR seviyelerinde önerilen konuşma iyileştirme yöntemi en fazla bölütsel SNR iyileştirmesini sağlamaktadır. Önerilen yöntemin bölütsel SNR iyileştirme miktarları; -10 dB SNR için 7,52 dB, -5 dB SNR için 7,16 dB, 0 dB SNR için 6,78 dB, 5 dB SNR için 5,68 dB ve 10 dB SNR için 3,43 dB. Bölütsel SNR iyileştirme miktarı bakımından önerilen yöntemi Ampirik Kip büzülme yöntemi takip etmektedir. Ayrıca SNR seviyesi yükseldikçe bölütsel SNR iyileştirme miktarları (dB) düşmektedir. Bölütsel SNR iyileştirme miktarı bakımından iyileştirilmiş konuşmaların kalitesini sağlamada Çoklu Bant İzgesel Çıkarım (ÇBİÇ) yöntemi en kötü performansı göstermektedir.

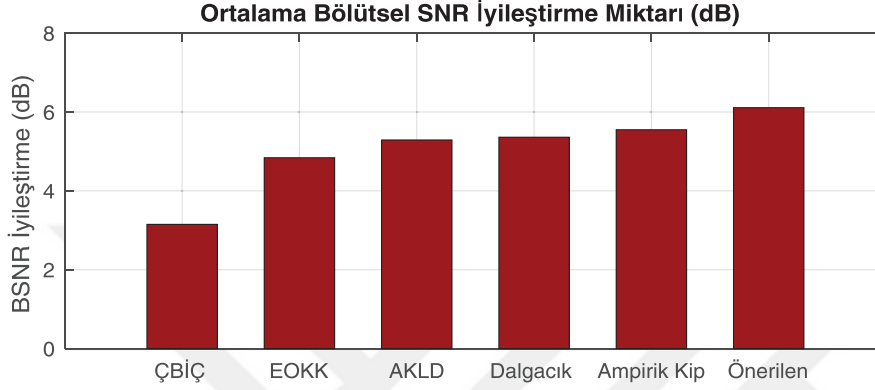


Şekil 8.15 Gürültülerin ortalaması alınarak SNR seviyesine göre yöntemlerin Bölütsel SNR iyileştirme miktarları

Konuşma iyileştirme yöntemlerinin nesnel kalite değerlendirmesinin gürültü tipi ve SNR seviyesinden bağımsız yapılması amacıyla tüm gürültü ve tüm SNR seviyelerinin ortalaması alınarak elde edilen bölütsel SNR iyileştirme miktarları Şekil 8.16'da verilmiştir. Şekil 8.16'dan görülebileceği üzere, önerilen konuşma iyileştirme yöntemi 6,11 dB BSNR iyileştirme miktarı sağlarken, önerilen yöntemi 5,55 dB iyileştirme miktarı ile Ampirik Kip büzülme yöntemi takip etmektedir. Ayrıca, Dalgacık büzülme yöntemi 5,36 dB, Algısal KLD yöntemi

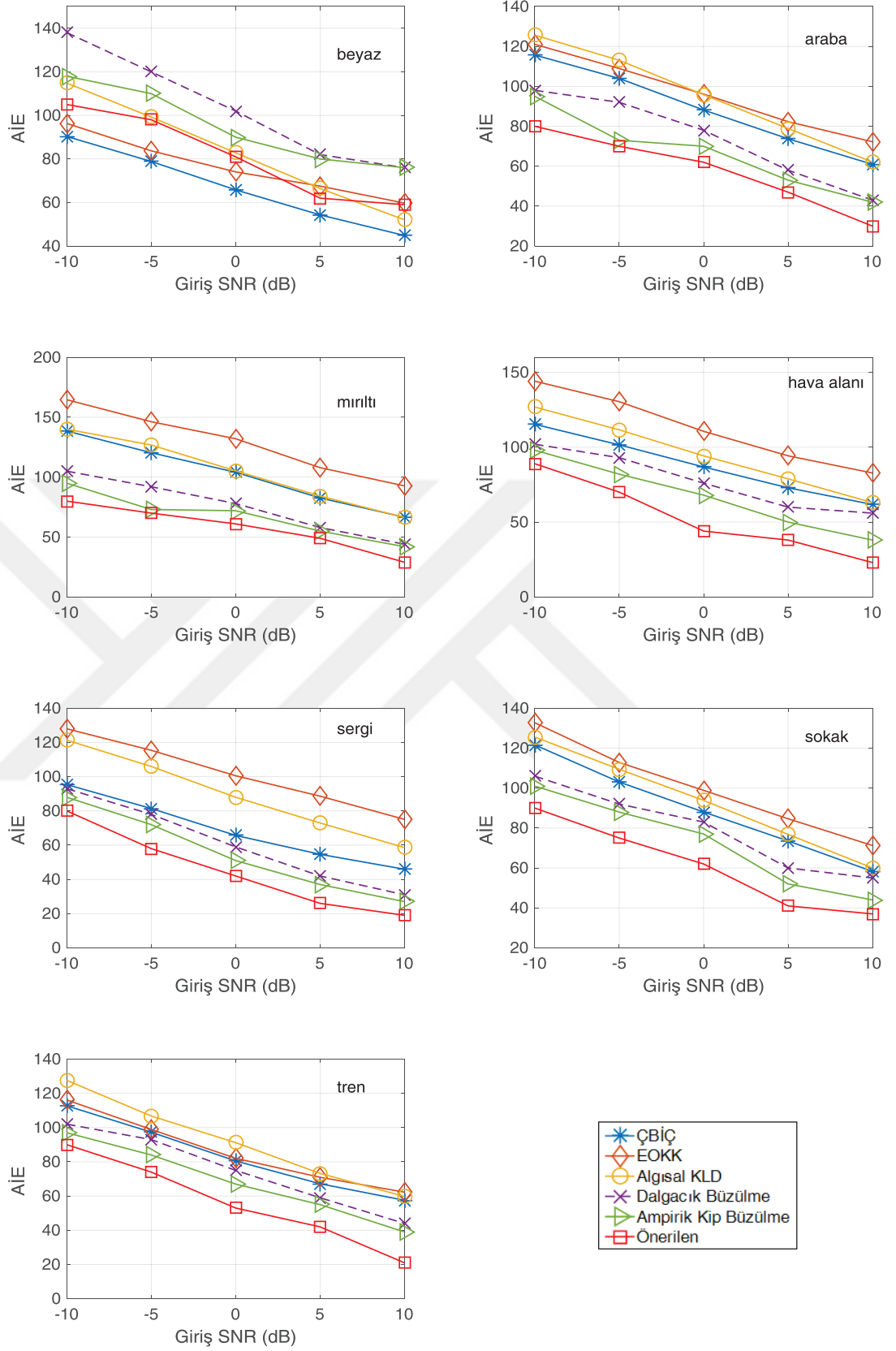
5,29 dB, EOKK yöntemi 4,84 dB ve ÇBİÇ yöntemi 3,15 dB bölütsel SNR iyileştirme sağlamaktadır.

Tüm sonuçlar göz önünde bulundurulduğunda, önerilen konuşma iyileştirme yaklaşımı diğer yöntemlere göre daha fazla bölütsel SNR iyileştirme miktarı sağlamaktadır ve bu sonuçlar önerilen yöntemin bölütsel SNR iyileştirmesi kriterine göre tercih edilebileceğini göstermektedir.



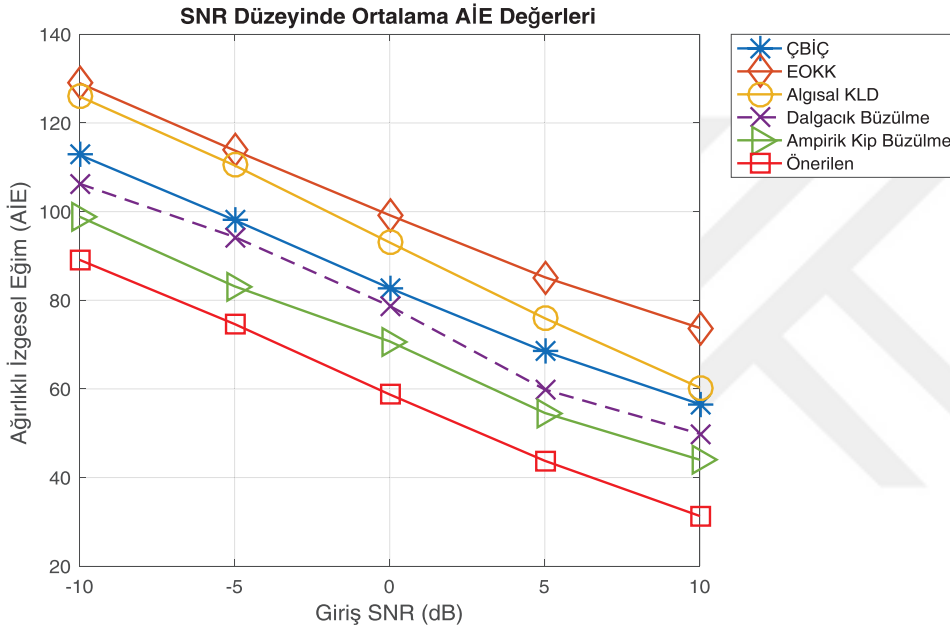
Şekil 8.16 Gürültülerin ve SNR seviyelerinin ortalaması alınarak yöntemlere ait Bölütsel SNR iyileştirme miktarları

İyileştirilmiş konuşma sinyallerinin nesnel kalite değerlendirmesinde kullanılan bir diğer ölçüt olan Ağırlıklı İzgesel Eğim (AİE) sonuçları tüm gürültü tipi ve tüm SNR seviyeleri için Şekil 8.17’de verilmiştir. Şekil 8.17’den görülebildiği üzere, önerilen yöntem beyaz Gauss gürültüsü dışında diğer tüm gürültü tipi ve tüm SNR seviyelerinde en düşük AİE değerlerini vermektedir. Bu nedenle, önerilen yöntem diğer yöntemler ile karşılaştırıldığında AİE nesnel kalite ölçütüne göre konuşma iyileştirmede en etkili yöntem olarak bulunmaktadır. Beyaz Gauss gürültü koşullarında ise ÇBİÇ yöntemi en düşük AİE değerlerini verdiği için en iyi yöntem olarak bulunmaktadır. Kullanılan diğer yöntemlerin AİE kalite performansına bakıldığında ise EOKK ve AKLD yöntemleri en etkisiz yöntemler olarak tespit edilmiştir.



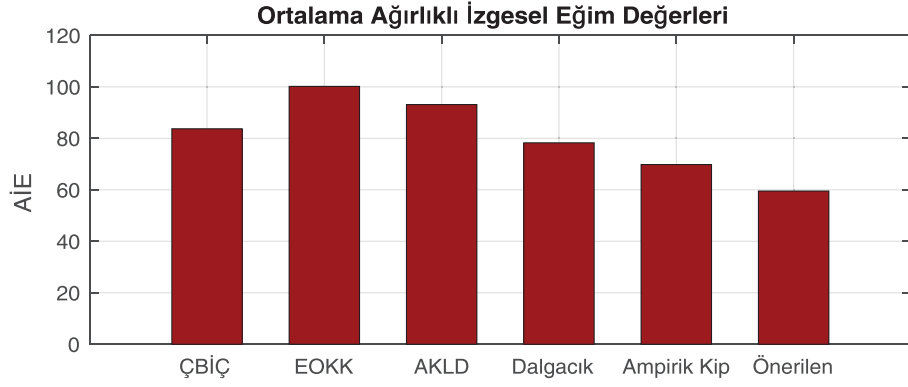
Şekil 8.17 Türkçe için tüm gürültülü koşullara ait AİE değerleri

Tüm gürültü tiplerinin her bir SNR seviyesi için ortalama AİE değerleri Şekil 8.18’de gösterilmektedir. Elde edilen ortalama sonuçlara göre, önerilen yöntem ile iyileştirilmiş konuşmaların AİE değerleri; -10 dB SNR seviyesinde 89,12, -5 dB SNR’da 74,63, 0 dB’de 58,74, 5 dB SNR’da 43,78 ve 10 dB SNR seviyesinde 31,31 şeklinde bulunmaktadır. Önerilen yöntem diğer yöntemler ile karşılaştırıldığında en düşük AİE değerlerine sahip olduğundan nesnel kalite bakımından en etkili yöntem olarak belirlenmiştir. En yüksek AİE değerlerine sahip EOKK yöntemi ise konuşma iyileştirmede kaliteyi sağlamada en kötü performansı göstermiştir.



Şekil 8.18 Gürültülerin ortalaması alınarak SNR seviyesine göre yöntemlerin AİE değerleri

Ağırlıklı İzgesel Eğim (AİE) nesnel kalite ölçütüne göre yöntemlerin genel performansını ölçmek amacıyla tüm gürültü ve tüm SNR seviyeleri için ortalama AİE değerlerine bakılmıştır. Şekil 8.19’da verilen ortalama AİE değerlerine göre, önerilen yöntem 59,52 ile en düşük AİE değerlerine sahip olmaktadır. Önerilen yöntemi sırasıyla Ampirik Kip büzülme (69,81), Dalgacık büzülme (78,26), ÇBİÇ (83,72), AKLD (93,12) ve EOKK (100,18) yöntemleri takip etmektedir. AİE nesnel kalite ölçütü bakımından en yüksek değerlere sahip EOKK yöntemi diğer yöntemlere göre en kötü performansı göstermiştir.



Şekil 8.19 Gürültülerin ve SNR seviyelerinin ortalaması alınarak yöntemlere ait AİE değerleri

Tüm gürültü koşullarında farklı beş SNR seviyesi ve bu SNR seviyelerinin ortalaması alınarak elde edilen tüm yöntemlere ait KKAD skorları Çizelge 8.3’de verilmiştir. Çizelge 8.3’den görülebileceği üzere, beyaz, araba, mırıltı, hava alanı, sergi, sokak ve tren gürültü koşullarında önerilen AKA tabanlı konuşma iyileştirme yaklaşımı en yüksek KKAD değerlerine sahiptir. Her gürültü tipi için farklı SNR seviyelerinde KKAD değerleri değişkenlik gösterdiğinden beş SNR seviyesi için bulunan KKAD skorlarının ortalaması alınarak gürültü bazında en iyi konuşma iyileştirme yönteminin bulunması amaçlanmıştır. SNR seviyelerinin ortalaması alınarak elde edilen sonuçlara göre; beyaz gürültü koşullarında önerilen yöntemin KKAD skoru 2,26 iken araba için 2,15, mırıltı için 2,38, hava alanı için 2,57, sergi için 2,51, sokak için 2,67 ve tren gürültüsü için 3,08 olarak bulunmuştur. KKAD skorları bakımından önerilen yöntemi araba gürültüsü dışında tüm gürültü koşulları için Ampirik kip büzülme ve Dalgacık büzülme yöntemleri takip etmektedir.

Türkçe konuşma sinyalleri için KKAD anlaşılabilirlik ölçütü bakımından en iyi ve etkili yöntemin tespit edilmesi amacıyla tüm gürültü tipi ve tüm SNR seviyeleri için ortalama KKAD değerlerine bakılmıştır. Elde edilen sonuçlara göre 2,52 KKAD skoruna sahip önerilen yöntem konuşma iyileştirmede anlaşılabilirlik bakımından daha başarılı bulunmuştur. KKAD sonuçlarına göre en iyi yöntem olarak belirlenen önerilen yöntemi sırasıyla 2,29, 2,14 ve 2,13 KKAD skorlarına sahip Ampirik kip büzülme, EOKK ve Dalgacık büzülme yöntemleri izlemektedir.

KKAD skoru bakımından iyileştirmede anlaşılabilirliğin sağlanmasında en kötü yöntem olarak 1,81 KKAD skoruna sahip Algısal KLD (AKLD) yöntemi tespit edilmiştir.

Çizelge 8.3 Türkçe konuşma sinyalleri için farklı yöntem ve gürültülere ait KKAD sonuçlarından elde edilen anlaşılabilirlik sonuçları

Ortamlar		Yöntemler					
	SNR	ÇBİÇ	EOKK	AKLD	Dalgacık	AKA	Önerilen
beyaz	10	2,52	2,79	2,63	2,71	2,78	2,96
	5	2,05	2,45	2,49	2,66	2,74	2,87
	0	1,63	2,09	2,24	2,29	2,42	2,01
	-5	1,44	1,65	1,86	1,66	1,87	1,94
	-10	1,27	1,37	1,19	1,29	1,34	1,52
	Ort.	1,78	2,07	2,08	2,12	2,23	2,26

araba	10	2,75	2,89	2,46	1,96	2,21	2,75
	5	2,38	2,58	2,17	1,72	1,95	2,41
	0	1,90	2,17	1,83	1,52	1,60	2,13
	-5	1,52	1,69	1,22	1,46	1,54	1,85
	-10	1,11	1,42	0,74	1,19	1,22	1,63
	Ort.	1,93	2,15	1,69	1,57	1,70	2,15

mırıltı	10	2,67	2,63	2,42	2,72	2,80	3,02
	5	2,37	2,31	2,14	2,34	2,52	2,78
	0	1,99	1,92	1,73	2,01	2,04	2,36
	-5	1,62	1,49	1,13	1,68	1,74	1,99
	-10	1,23	1,36	0,68	1,42	1,62	1,78
	Ort.	1,97	1,94	1,62	2,03	2,14	2,38

hava alanı	10	2,78	2,76	2,50	2,89	3,05	3,24
	5	2,46	2,42	2,21	2,61	2,84	3,01
	0	2,13	2,03	1,82	2,28	2,45	2,79
	-5	1,75	1,63	1,26	1,74	1,86	2,08
	-10	1,34	1,42	1,04	1,38	1,47	1,74
	Ort.	2,09	2,05	1,76	2,18	2,33	2,57

sergi	10	2,57	2,65	2,42	2,78	2,85	3,16
	5	2,16	2,26	2,16	2,45	2,72	2,99
	0	1,70	1,84	1,76	1,89	2,06	2,43
	-5	1,41	1,33	1,23	1,53	1,98	2,12
	-10	1,13	1,10	1,05	1,37	1,62	1,88
	Ort.	1,79	1,84	1,72	2,00	2,24	2,51

sokak	10	2,88	2,96	2,55	2,71	2,97	3,28
	5	2,60	2,64	2,23	2,58	2,82	3,03
	0	2,24	2,30	1,81	2,43	2,56	2,79
	-5	1,84	1,85	1,24	1,99	2,17	2,36
	-10	1,28	1,42	0,90	1,56	1,82	1,91
	Ort.	2,17	2,23	1,74	2,25	2,46	2,67

tren	10	3,09	3,32	2,75	3,26	3,48	3,74
	5	2,84	3,00	2,49	3,13	3,21	3,29
	0	2,49	2,69	2,20	2,85	2,97	3,08
	-5	2,06	2,31	1,73	2,62	2,75	2,96
	-10	1,70	1,92	1,11	2,05	2,16	2,35
	Ort.	2,44	2,65	2,06	2,78	2,91	3,08

Ortalama Sonuçlar	2,03	2,14	1,81	2,13	2,29	2,52
--------------------------	-------------	-------------	-------------	-------------	-------------	-------------

Farklı gürültü koşulları ve farklı SNR seviyeleri için iyileştirilmiş Türkçe konuşma sinyallerinin anlaşılabilirlik değerlendirmesi, bir diğer nesnel anlaşılabilirlik ölçütü olan Kısa-zamanlı Nesnel Anlaşılabilirlik (KZNA) ölçütü kullanılarak yapılmıştır. Tüm gürültü tipleri ve tüm SNR seviyeleri için önerilen ve diğer konuşma iyileştirme yöntemlerine ait KZNA nesnel anlaşılabilirlik değerleri Çizelge 8.4’de verilmektedir. Elde edilen KZNA anlaşılabilirlik ölçütü sonuçlarına göre, tüm gürültü tipi ve SNR seviyeleri için önerilen yöntem en yüksek yüzdelere sahip olmaktadır. SNR seviyelerinin ortalaması alınarak gürültü bazında bulunan önerilen yönteme ait KZNA sonuçları şu şekildedir; beyaz Gauss gürültüsü koşullarında % 86,90, araba gürültüsü altında % 80,43, mırıltı gürültüsünde % 75,98, hava alanı % 80,53, sergi % 80,02, sokak % 82,63 ve tren gürültüsü koşullarında % 92,79.

Kullanılan yöntemlerin gürültü tipi ve SNR seviyesinden bağımsız olarak nesnel anlaşılabilirlik değerlendirmesinin yapılması amacıyla tüm gürültü ve SNR seviyeleri için hesaplanan KZNA skorlarının ortalaması alınmıştır. Elde edilen ortalama sonuçlara göre, % 82,76 KZNA anlaşılabilirlik yüzdesine sahip önerilen yöntem diğer yöntemlere göre anlaşılabilirliği sağlamada daha etkili ve başarılı olmuştur. Önerilen yöntemi, % 77,18 KZNA skoru ile Ampirik kip büzülme, % 74,35 ile dalgacık büzülme, % 68,22 ile ÇBİÇ, % 63,80 ile EOKK ve son olarak % 60,93 ile Algısal KLD (AKLD) yöntemleri takip etmektedir.

Çizelge 8.4 Türkçe konuşma sinyalleri için farklı yöntem ve gürültülere ait KZNA sonuçlarından elde edilen anlaşılabilirlik oran tahminleri (%)

Ortamlar		Yöntemler					
	SNR	ÇBİÇ	EOKK	AKLD	Dalgacık	AKA	Önerilen
beyaz	10	97,39	97,70	98,35	99,05	99,28	99,61
	5	90,18	92,47	95,19	97,42	98,84	99,13
	0	71,44	75,74	83,65	89,01	91,25	94,50
	-5	48,13	40,02	54,46	68,46	74,57	83,08
	-10	24,53	15,66	6,72	30,49	37,16	58,20
	Ort.	66,37	64,32	67,68	76,88	80,22	86,90

araba	10	98,51	98,41	98,73	99,14	99,36	99,82
	5	93,98	94,28	95,48	97,47	98,25	99,16
	0	74,96	74,04	77,74	82,39	84,28	93,01
	-5	39,34	33,34	20,72	53,66	60,09	72,13
	-10	16,14	10,56	1,37	20,98	29,10	38,05
	Ort.	64,58	62,13	58,81	70,72	74,21	80,43

mırıltı	10	98,37	97,91	98,62	99,12	99,38	99,74
	5	93,20	90,52	94,87	97,40	98,01	99,14
	0	67,19	52,63	68,22	82,27	86,19	94,15
	-5	22,52	11,32	6,97	36,53	41,90	50,83
	-10	5,07	2,08	0,56	18,22	23,67	36,07
	Ort.	57,27	50,89	53,85	66,70	69,83	75,98

hava alanı	10	98,90	98,72	98,94	99,20	99,46	99,91
	5	96,47	95,36	96,75	98,23	98,95	99,26
	0	84,45	77,74	80,02	88,19	90,05	94,17
	-5	45,75	30,61	16,14	56,31	59,98	71,15
	-10	13,68	5,98	0,89	19,94	24,41	38,16
	Ort.	67,86	61,68	58,54	72,37	74,57	80,53

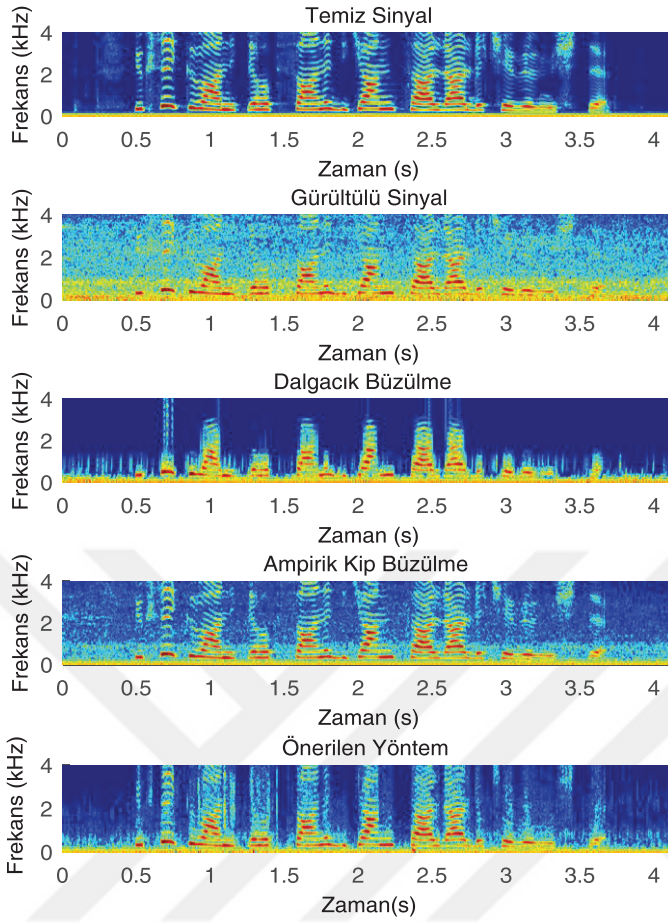
sergi	10	97,37	97,01	98,51	98,95	99,14	99,68
	5	90,05	87,05	94,30	97,26	98,10	99,35
	0	71,96	65,40	78,18	85,34	88,17	93,96
	-5	37,57	22,78	26,79	52,21	58,24	70,02
	-10	16,95	6,32	2,78	18,83	23,93	37,11
	Ort.	62,78	55,71	60,11	70,51	73,51	80,02

sokak	10	99,09	98,85	99,03	99,26	99,44	99,81
	5	97,36	96,31	96,45	98,19	98,57	99,22
	0	88,77	84,73	74,65	91,25	93,01	96,14
	-5	56,86	45,86	10,32	64,54	69,07	75,29
	-10	23,22	14,37	1,30	28,16	34,04	42,73
	Ort.	73,06	68,03	56,35	76,28	78,82	82,63

tren	10	99,53	99,54	99,43	99,49	99,71	99,93
	5	98,90	98,87	98,61	98,08	98,95	99,16
	0	96,17	95,86	94,96	97,16	98,03	98,95
	-5	83,42	81,03	56,36	85,21	88,17	92,63
	-10	49,90	43,93	6,55	54,84	60,59	73,32
	Ort.	85,58	83,85	71,18	86,95	89,09	92,79

Ortalama Sonuçlar	68,22	63,80	60,93	74,35	77,18	82,76
--------------------------	--------------	--------------	--------------	--------------	--------------	--------------

İyileştirilmiş konuşma sinyallerinin öznel kalite değerlendirmesinin yapılması amacıyla, temiz konuşma, araba gürültüsü ile bozulmuş 5 dB SNR seviyesinde gürültülü konuşma, önerilen ve diğer yöntemler kullanılarak elde edilen iyileştirilmiş konuşmaların spektrogramları Şekil 8.20’de verilmiştir. Konuşma iyileştirmede kullanılan dalgacık büzülme yönteminde gürültünün yanında konuşmanın yüksek frekanslı bileşenlerinin de kaybolduğu görülmektedir. Konuşmanın yüksek frekanslı bileşenlerinin gürültü ile birlikte atılması konuşmanın anlaşılabilirliğini olumsuz etkilemektedir. Kullanılan bir diğer yöntem olan Ampirik Kip büzülme yaklaşımında, konuşmanın tüm bileşenleri mümkün olduğunca korunurken gürültünün yeterince atılmadığı gözlemlenmektedir. Şekil 8.20’de verilen spektrogramlardan hareketle, önerilen yöntem diğer yöntemlere göre konuşma sinyalinin hem düşük hem de yüksek frekanslı bileşenlerini korurken, konuşmayı bozan gürültünün büyük miktarda azaldığı görülmektedir.



Şekil 8.20 Türkçe için; temiz sinyalin, 5 dB araba gürültüsü ile bozulmuş gürültülü sinyalin, önerilen ve diğer yöntemler ile iyileştirilmiş sinyallerin spektrogramları

8.4 Tartışma

Bu çalışmada, konuşma iyileştirmede istatistiksel bir model olan Gamma dağılımı tabanlı uyarlanabilir eşik değerlerinin belirlenmesi yaklaşımı önerilmiştir. Tüm katsayıların tek bir eşik değeri kullanılarak eşikleme fonksiyonundan geçirilmesi gürültünün yanında konuşmaya ait bileşenlerin kaybolmasına neden olabilmektedir. Bu nedenle, konuşma ve konuşma dışı bölgelerin İKF katsayılarının enerjisinin Gamma olasılık dağılımı kestirilerek uyarlanabilir eşik değerlerinin belirlenmesi sağlanmıştır. Bu yaklaşım, konuşma bölgelerinin korunup gürültü katsayılarının eşiklenmesini sağlamada önemli bir rol oynamaktadır. Ampirik Kip Ayırıştırması analizi kullanılarak İKF katsayılarına ayrıştırılan gürültülü konuşma sinyalleri yarı-yumuşak eşikleme fonksiyonundan

geçirilerek yüksek kalitede iyileştirilmiş konuşmanın elde edilmesi amaçlanmıştır. İngilizce ve Türkçe gürültülü konuşma sinyallerinin iyileştirilmesi için önerilen yöntem, Bölütsel SNR (BSNR) iyileştirme miktarı ve Ağırlıklı İzgesel Eğim (AİE) nesnel kalite ölçütleri ile KKAD ve KZNA nesnel anlaşılabilirlik ölçütleri bakımından literatürde sıklıkla kullanılan Ampirik Kip büzülme, Dalgacık büzülme, ÇBİÇ, EOKK ve Algısal KLD (AKLD) yöntemleri ile karşılaştırılmıştır. Nesnel kalite ve anlaşılabilirlik bakımından önerilen yöntemin diğer yöntemlere göre daha fazla bölütsel SNR iyileştirmesi, daha düşük ağırlıklı izgesel eğim değerleri, daha yüksek KKAD ve KZNA değerleri verdiği ve bu nedenle daha iyi bir konuşma iyileştirme performansı gösterdiği gözlemlenmiştir.

Önerilen yöntemin İngilizce ve Türkçe dilleri için performans değerlendirmesi yapıldığında, Türkçe konuşmalar için bölütsel SNR iyileştirme miktarı 6,11 dB bulunurken İngilizce konuşmalar için 4,32 dB bulunmaktadır. Ayrıca, Türkçe konuşmalar için ortalama AİE değeri 59,52 iken İngilizce konuşmalar için ortalama AİE değeri 55,40 bulunmaktadır. Bu sonuçlar, önerilen yöntemin Türkçe için İngilizce'ye göre daha fazla bölütsel SNR iyileştirmesi sağladığını göstermektedir. Ağırlıklı izgesel eğim bakımından ise İngilizce konuşmalarda iyileştirme performansı Türkçe konuşmalara göre daha iyi bulunmuştur. Böylece, önerilen yöntemin her iki dil için tatmin edici nesnel kalite sağladığı söylenebilir.

Türkçe ve İngilizce için anlaşılabilirlik değerlendirmesi yapıldığında ise, önerilen yöntem İngilizce konuşmalar için 2,62 KKAD skoruna sahip iken Türkçe konuşmalar için bu skor 2,52 olarak tespit edilmiştir. Bu skorlar, fikir değerlendirme oranı (Çizelge 4.1) açısından değerlendirildiğinde dinleyicinin iyileştirilmiş konuşmayı *algılanabilir fakat biraz rahatsız edici* olarak değerlendirmesine karşılık gelmektedir. Ayrıca, Türkçe için ortalama KZNA değeri % 82,76 iken İngilizce için bu oran % 84,92 olarak bulunmuştur. Bu anlaşılabilirlik sonuçlarına göre, önerilen yöntemin hem İngilizce hem de Türkçe konuşmaların anlaşılabilirliğini sağlamada etkili bir performans gösterdiği söylenebilir.

Önerilen ve diğer yöntemlerle elde edilen iyileştirilmiş konuşma sinyallerinin spektrogram gösterimlerine bakıldığında, önerilen yöntemin diğer yöntemlere göre İngilizce ve Türkçe konuşma sinyallerinin hem düşük hem de yüksek frekanslı bileşenlerini koruduğu ve gürültüyü daha fazla azalttığı görülmektedir.

Sonuç olarak, Ampirik Kip Ayrıştırması tabanlı önerdiğimiz yöntemin dillerin yapısından, gürültü tipinden ve SNR seviyesinden bağımsız olarak etkili bir konuşma iyileştirme yöntemi olarak kullanılabileceği söylenebilir.



9. GENEL SONUÇLAR, ÖNERİLER VE GELECEK ÇALIŞMALAR

Bu araştırmada doğrulanan önemli noktalardan biri, konuşma iyileştirme yöntemleri ile konuşmanın algısal kalite ve anlaşılabilirliğinde iyileşme sağlanabildiği ve gürültülü konuşmaya uygulanması gerektiğidir. Bununla birlikte, gürültü belli bir seviyeye kadar azaltılmasına rağmen tamamen ortadan kaldırılamamaktadır. Yüksek sinyal-gürültü oranı koşullarında, iyileştirilmiş konuşmanın kalitesi ve anlaşılabilirlik skoru çok iyi olabilir. Çok düşük sinyal-gürültü oranı koşullarında ise iyileştirilmiş konuşma sinyalinin kalitesinde bir iyileşme sağlanabilmesine karşın konuşma sinyalinde artık gürültü veya bozulma ya da her ikisi de hala bulunabilir. Bu tür durumlarda, özellikle anlaşılabilirliğin sağlanması için iyi bir gürültü kestirimi ve gürültü azaltmada farklı gürültü tipi koşullarında etkili olabilecek eşik değerlerinin belirlenmesi sağlayan iyi bir konuşma iyileştirme algoritmasına ihtiyaç duyulmaktadır.

Çalışmadaki ilk ana özgünlüğümüz, konuşma işleme alanında etkili ve doğru bir şekilde kullanılabilecek bir Sesli Etkinlik Algılama (Voice Activity Detection-VAD) algoritmasının önerilmesidir. Gürültülü konuşmalarda, konuşma ve konuşma olmayan bölgeleri algılayabilen ve bu bölgeleri etiketleyen Ampirik Kip Ayırıştırması (AKA) tabanlı bir yaklaşım kullanılmıştır. Zaman-frekans alanı yöntemlerinden dalgacık dönüşümünde sabit bir temel dalgacık işlevi kullanılmakta ve eşik değerleri yalnızca enerji üzerinden çıkartılmaktadır. Eşik değerlerinin bu şekilde bulunması gürültülü ortamlarda konuşma olmayan bölgelerin tespitini zorlaştırmaktadır. Önerdiğimiz yaklaşımda, tüm temel fonksiyonları sinyalin kendisinden türetilen ve sinyali farklı frekans bileşenlerine ayırtıran AKA yöntemi kullanılmıştır. Ayırıştırılan gürültülü konuşmaların içkin kip fonksiyonu katsayılarının Teager enerjisi bulunarak Gamma dağılım işlevi yardımıyla her kip için gürültü miktarına göre değişen eşik değerlerinin hesaplanması önerilen yöntemi etkin ve başarılı kılmaktadır. Uyarlanabilir eşik değerlerinin konuşmanın yapısına uygun olarak Gamma dağılım işlevi ile hesaplanması sesli etkinlik algılama yöntemleri bakımından bu çalışmanın özgünlüğüne katkıda bulunmaktadır. Ayrıca, literatürdeki diğer yöntemlerden farklı olarak önerilen yöntemde konuşma ve konuşma dışı bölgelerin enerji olasılık dağılımı üzerinden belirlenmesi konuşmanın özellikle yüksek frekanslı bileşenlerinin tespitini kolaylaştırmaktadır.

Konuşma iyileştirmede, literatürde sıklıkla kullanılan zaman, frekans ve

zaman-frekans alanı yöntemleri gürültü giderildikten sonra iyileştirilmiş konuşmanın kalitesini belli ölçüde artırırken anlaşılabilirliği sağlamada yetersiz kalmaktadır. İngilizce dili için tasarlanmış konuşma iyileştirme yöntemlerinin dillerin yapısı gereği Türkçe için aynı performans ve başarıyı göstermesi zor olmaktadır. Bu nedenle, hem İngilizce hem de Türkçe için kalite ve anlaşılabilirliği sağlamak için zaman-frekans alanı yöntemlerinden olan AKA kullanılmıştır. Önerdiğimiz konuşma iyileştirme yönteminde, gürültülü konuşmanın ve gürültünün içkin kip fonksiyonu katsayılarının Teager enerjileri hesaplanarak bu enerji dağılımına uygun bir dağılım olan Gamma dağılım işlevi ile temiz konuşma olma olasılığı ve gürültü olma olasılığı üzerinden uyarlanabilir eşik değerleri hesaplanmıştır. Uyarlanabilir eşik değerlerinin konuşmanın yapısına uygun olarak Gamma dağılım işlevi ile hesaplanması konuşma iyileştirme yöntemleri bakımından bu çalışmanın özgünlüğüne katkıda bulunmaktadır. Bu nedenle dil yapıları farklı olan İngilizce ve Türkçe için önerilen yöntemin başarılı olması literatüre bir yenilik katmıştır.

Hem önerilen sesli etkinlik algılama hem de konuşma iyileştirme yaklaşımı literatürde sıklıkla kullanılan yöntemler ile başarı ve iyileştirme oranları bakımından karşılaştırılmıştır. Sesli etkinlik algılama algoritmaları için konuşma bölgesi doğruluk oranı (HR1), konuşma olmayan bölge doğruluk oranı (HR0) ve genel doğruluk oranı (GDO) ölçütleri kullanılmıştır. Elde edilen sonuçlara AKA tabanlı önerilen sesli etkinlik algılama algoritması, AMR1, AMR2 ve G.729B gibi standart algoritmalar, SOHN ve LTSD SEA algoritmalarına göre HR1, HR0 ve GDO oranları bakımından üstünlük sağlamıştır. Konuşma iyileştirme yöntemlerinin değerlendirilmesinde ise nesnel kalite için bölütsel SNR iyileştirme miktarı (dB) ve ağırlıklı izgesel eğim (AİE), nesnel anlaşılabilirlik için konuşma kalitesinin algısal değerlendirmesi (KKAD) ve kısa-zamanlı nesnel anlaşılabilirlik (KZNA) ölçütleri kullanılmıştır. Nesnel kalite ve anlaşılabilirlik sonuçlarına göre önerilen AKA tabanlı konuşma iyileştirme yaklaşımı, literatürde sıklıkla kullanılan izgesel tabanlı, istatistiksel tabanlı, Wiener süzgeçleme ve alt-uzay algoritmaları ile zaman-frekans yöntemlerinden olan dalgacık büzülme ve ampirik kip büzülme yöntemlerine göre iyileştirilmiş konuşmada daha fazla gürültü atılmasını ve anlaşılabilirlik skorunun yükseltilmesini sağlamıştır.

İleriki çalışmalarda, aşağıda belirtilen hususlarda araştırma yapılabilir.

- Önerilen yöntemin Türkçe ve İngilizce dilleri için performans değerlendirmesi gerçekleştirilmiştir. Önerilen yöntemin performansı diğer

diller (Almanca, Japonca vb.) için de değerdendirilebilir.

- Önerilen konuşma iyileştirme yaklaşımında farklı gürültü kestirim algoritmaları ve eşikleme fonksiyonları kullanılarak yöntemin performansı üzerine etkileri incelenebilir.
- Önerilen sesli etkinlik algılama ve konuşma iyileştirme yönteminin, konuşma ve konuşmacı tanıma sistemlerinin performansına olan etkisi incelenebilir.



KAYNAKLAR DİZİNİ

- Aibinu, A. M., Salami, M. J. E. and Shafie, A.A.**, 2012, ‘Artificial neural network based autoregressive modeling technique with application in voice activity detection’, *Engineering Applications of Artificial Intelligence*. Elsevier, 25(6), pp. 1265–1276.
- Babu, C. and Vanathi, P.**, 2009, ‘Performance analysis of voice activity detection algorithms for robust speech recognition’, *International Journal of Computing Science and Communication Technologies*, 2(1), pp. 288–293.
- Bachu, R. G., Kopparthi, S., Adapa, B. and Barkana, B.D.**, 2010, ‘Separation of Voiced and Unvoiced using Zero crossing rate and Energy of the Speech Signal’, *Advanced Techniques in Computing Sciences and Software Engineering*, pp. 279–282.
- Bahoura, M. and Rouat, J.**, 2001, ‘Wavelet speech enhancement based on the Teager energy operator’, *IEEE Signal Processing Letters*, 8(1), pp. 10–12.
- Bahoura, M., and Rouat, J.**, 2006, Wavelet speech enhancement based on time-scale adaptation. *Speech Commun.* 48 (December (12)), pp. 1620–1637.
- Benyassine, A., Shlomot, E., Su, H. Y., Massaloux, D., Lamblin, C. and Petit, J. P.**, 1997, ‘ITU-T recommendation G.729 annex B: A silence compression scheme for use with G.729 optimized for V.70 digital simultaneous voice and data applications’, *IEEE Communications Magazine*, 35(9), pp. 64–73.
- Beritelli, F., Casale, S. and Cavallaro, A.**, 1998, ‘A robust voice activity detector for wireless communications using soft computing’, *IEEE Journal on Selected Areas in Communications*, 16(9), pp. 1818–1829.
- Boll, S.F.**, 1979, Suppression of Acoustic Noise in Speech using Spectral Subtraction, *IEEE Transactions on Acoustics, Speech, and Signal Processing*, vol. 27, no. 2, pp. 113–120, (1979).

KAYNAKLAR DİZİNİ (Devam)

- Chang, J.H., Shin, J.W., and Kim, N.S.**, 2003, 'Likelihood ratio test with complex Laplacian model for voice activity detection', *Eurospeech*, (4), pp. 1065–1068.
- Chatlani, D. and Soraghan, J.J.**, 2009, 'Speech Enhancement using Adaptive Empirical Mode Decomposition', *2009 16th International Conference on Digital Signal Processing (DSP)*, pp. 1–6.
- Chehrehsa, S. and Savoji, M.H.C.A.**, 2014, 'Speech Enhancement using Gaussian Mixture Models , Explicit Bayesian Estimation and Wiener Filtering', 10(3), pp. 168–175.
- Chen, S., Guido, R. C., Truong, T. and Chang, Y.**, 2010, 'Improved voice activity detection algorithm using wavelet and support vector machine', *Computer Speech & Language*. Elsevier Ltd, 24(3), pp. 531–543.
- Chen, S. H., Wu, H. Te, Chang, Y. and Truong, T. K.**, 2007, 'Robust voice activity detection using perceptual wavelet-packet transform and Teager energy operator', *Pattern Recognition Letters*, 28(11), pp. 1327–1332.
- Chong Feng, C. Z.**, 2014, 'Voice activity detection based on ensemble empirical mode decomposition and teager kurtosis', *Icsp2014*, pp. 455–460.
- Chung, K. and Oh, S. Y.**, 2015, 'Voice Activity Detection Using an Improved Unvoiced Feature Normalization Process in Noisy Environments', *Wireless Personal Communications*. Springer US, 89(3), pp. 1–13.
- Ciloglu, T. and ve Ahmet Tokath, D. A.**, 2016, 'OrienTel: Turkish telephone speech database', *Proceedings of the IEEE 12th Signal Processing and Communications Applications Conference, 2004.*, (January 2004), pp. 280–283.
- Cohen, I.**, 2002, 'Optimal speech enhancement under signal presence uncertainty using log-spectral amplitude estimator', *IEEE Signal Processing Letters*, 9(4), pp. 113–116.

KAYNAKLAR DİZİNİ (Devam)

- Cohen, I. and Berdugo, B.**, 2001, 'Speech enhancement for non-stationary noise environments', *Signal Processing*, 81(11), pp. 2403–2418.
- Cohen, I., Box, P. O. and Ilit, Y.**, 2001, 'Enhancement of Speech Using Bark-Scaled Wavelet Packet Decomposition', *Eurospeech 2001*, pp. 2–5.
- Donoho, D.L.**, 1995, 'De-noising by soft-thresholding', *IEEE Transactions on Information Theory*, 41(3), pp. 613–627.
- Donoho, D. L. and Johnstone, I.M.**, 1994, 'Ideal Spatial Adaption by Wavelet Shrinkage', *Biometrika*, pp. 425–455.
- Enqing, D., Heming, Z. and Yongli, L.**, 2002, 'Low bit and variable rate speech coding using local cosine transform', *TENCON'02. Proceedings.*, pp. 423–426.
- Ephraim, Y.**, 1992, 'Statistical-model-based speech enhancement systems', *Proceedings of the {IEEE}*, 80(10), pp. 1526–1555.
- Ephraim, Y., and Malah, D.**, 1985, 'Speech enhancement using a minimum mean-square error log-spectral amplitude estimator', *IEEE Transactions on Acoustics, Speech, and Signal Processing*, 33(2), pp. 443–445.
- Erkelens, J.S., Hendriks, R.C., Heusdens, R., and Jensen, J.**, 2007, 'Minimum mean-square error estimation of discrete fourier coefficients with generalized gamma priors', *IEEE Transactions on Audio, Speech and Language Processing*, 15(6), pp. 1741–1752.
- Eshaghi, M., and Karami Mollaei, M.R.**, 2010, 'Voice activity detection based on using wavelet packet', *Digital Signal Processing*. Elsevier Inc., 20(4), pp. 1102–1115.
- Flandrin, P., Rilling, G., and Goncalves, P.**, 2004, 'Empirical mode decomposition as a filter bank', *IEEE Signal Processing Letters*, 11(2), pp. 112–114.

KAYNAKLAR DİZİNİ (Devam)

- Freeman, D. and Cosier, G.,** 1989, 'The voice activity detector for the Pan-European digital cellular mobile telephone service', *Acoustics, Speech, and Language*, pp. 369–372.
- Ghanbari, Y., and Karami, M.,** 2004, Spectral subtraction in the wavelet domain for speech enhancement. *Internat. J. Software Inf. Technol. (IJSIT)* 1 (1), 26–30.
- Ghanbari, Y., and Karami-Mollaei, M.R.,** 2006, 'A new approach for speech enhancement based on the adaptive thresholding of the wavelet packets', *Speech Communication*, 48(8), pp. 927–940.
- Grancharov, V. and Kleijn, W.,** 2008, Speech quality assessment, in *Handbook of Speech Processing*, J. Benesty, M. Sondhi, and Y. Huang, Eds. Berlin, Germany: Springer Verlag, pp. 83–99.
- Greenwood, M. and Kinghorn, A.,** 1999, 'SUVing: Automatic silence/unvoiced/voiced classification of speech', *Coursework, Department of*, pp. 3–6.
- Haigh, J.A. and Mason, J.S.,** 1993, 'Robust voice activity detection using cepstral features', *Proceedings of TENCON '93. IEEE Region 10 International Conference on Computers, Communications and Automation*, (0), pp. 2–5.
- Hamid, E., Molla, K. I., Dang, X. and Nakai, T.,** 2013, 'Single channel speech enhancement using adaptive soft-thresholding with bivariate EMD'. *ISRN signal processing*, 2013.
- Harsha, B.V.,** 2004, 'A noise robust speech activity detection algorithm', *In Intelligent Multimedia, Video and Speech Processing, 2004. Proceedings of 2004 International Symposium on* (pp. 322-325) *IEEE*.

KAYNAKLAR DİZİNİ (Devam)

- Hooda, S. and Aggarwal, S.**, 2012, 'Review of MMSE Estimator for Speech Enhancement', *International Journal of Advanced Research in Computer Science and Software Engineering*, 2(11), pp. 419–424.
- Hu, Y., and Loizou, P.C.**, 2006, Subjective comparison of speech enhancement algorithms. *Speech Commun.* 49, 588–601. *Proc. IEEE Int. Conf. Acoust. Speech Signal Process.*, pp.153-156.
- Hu, Y. and Loizou, P.**, 2007, 'A comparative intelligibility study of speech enhancement algorithms', *International Conference on Acoustics, Speech and Signal Processing*, pp. 561–564.
- Hu, Y. and Loizou, P. C.**, 2007, 'A comparative intelligibility study of single-microphone noise reduction algorithms.', *The Journal of the Acoustical Society of America*, 122(3), p. 1777.
- Hu, Y. and Loizou, P.**, 2008, Evaluation of objective quality measures for speech enhancement, *IEEE Trans. Audio Speech Lang Process.*, vol. 16, no. 1, pp. 229–238.
- Islam, M.T., Shahnaz, C., Zhu, W.P. and Ahmad, M.O.**, 2015, 'Speech Enhancement Based on Student t Modeling of Teager Energy Operated Perceptual Wavelet Packet Coefficients and a Custom Thresholding Function', *IEEE/ACM Transactions on Speech and Language Processing*, 23(11).
- Johnstone, I. M., and Silverman., B.W.**, 1997, "Wavelet threshold estimators for data with correlated noise." *J. Royal Statistical Society* 59(2): 319-351.
- Kamath, S., and Loizou, P.**, 2002, A Multi-Band Spectral Subtraction Method for Enhancing Speech Corrupted by Colored Noise, *IEEE Int. Conf. on Acoustics, Speech, and Signal Processing*, Orlando, USA, vol. 4, pp. 4160–4164, May (2002).

KAYNAKLAR DİZİNİ (Devam)

- Kanai, Y., Morita, S. and Unoki, M.,** 2013, ‘Concurrent processing of voice activity detection and noise reduction using empirical mode decomposition and modulation spectrum analysis’, *Proceedings of the Annual Conference of the International Speech Communication Association, INTERSPEECH*, 1(August), pp. 742–746.
- Karray, L. and Martin, A.,** 2003, ‘Towards improving speech detection robustness for speech recognition in adverse conditions’, *Speech Communication*, 40(3), pp. 261–276.
- Kemiha, M.,** 2014, ‘Empirical mode decomposition and normal shrink thresholding for speech denoising’ *arXiv preprint arXiv:1405.7895*.
- Khalidi, K.,** 2012. ‘Processing and analysis of sound signals by Huang transform (Empirical Mode Decomposition : EMD)’ (Doctoral dissertation, Télécom Bretagne, Université de Bretagne Occidentale).
- Khalidi, K., Boudraa, A.-O. and Komaty, A.,** 2014a, ‘Speech enhancement using empirical mode decomposition and the Teager-Kaiser energy operator.’, *The Journal of the Acoustical Society of America*, 135(1), pp. 451–459.
- Khalidi, K., Boudraa, A.-O. and Komaty, A.,** 2014b, ‘Speech enhancement using empirical mode decomposition and the Teager–Kaiser energy operator’, *The Journal of the Acoustical Society of America*, 135(1), pp. 451–459.
- Khalidi, K., Boudraa, A. O., Bouchikhi, A. and Alouane, M.T.H.,** 2008, ‘Speech enhancement via EMD’, *Eurasip Journal on Advances in Signal Processing*, 2008.
- Khoa, P.,** 2012, Noise robust voice activity detection. *Nanyang Technological University*, 24.

KAYNAKLAR DİZİNİ (Devam)

- Li, Y., Wang, T., Cui, H. and Tang, K.,** 2006, 'Voice activity detection in non-stationary noise', *IMACS Multiconference on 'Computational Engineering in Systems Applications'*, CESA, 8(4), pp. 1573–1575.
- Li, J., Yang, L., Zhang, J., Yan, Y., Hu, Y., Akagi, M., and Loizou, P.,** 2011, Comparative intelligibility investigation of single-channel noise-reduction algorithms for Chinese, Japanese, and English. *J. Acoust. Soc. Am.* 129(5), pp. 3291–3301.
- Lim, J.,** 1978, Evaluation of a correlation subtraction method for enhancing speech degraded by additive noise, *IEEE Trans. Acoust., Speech, Signal Process.*, vol. ASSP-37, no. 6, pp. 471–472.
- Loizou, P.C., Lobo, A. and Hu, Y.,** 2005, 'Subspace algorithms for noise reduction in cochlear implants.', *The Journal of the Acoustical Society of America*, 118(5), pp. 2791–2793.
- Lu, Y. and Loizou, P.C.,** 2008, 'A geometric approach to spectral subtraction', *Speech Communication*, 50(6), pp. 453–466.
- Ma, J., Hu, Y. and Loizou, P.,** 2009, Objective measures for predicting speech intelligibility in noisy conditions based on new band importance functions, *J. Acoust. Soc. Amer.*, vol. 125, no. 5, pp. 3387–3405.
- Ma, Y. and Nishihara, A.,** 2013, 'Efficient voice activity detection algorithm using long-term spectral flatness measure', pp. 1–18.
- Martin, R.,** 2002, 'Speech enhancement using MMSE short time spectral estimation with gamma distributed speech priors', *IEEE International Conference on Acoustics Speech and Signal Processing*, p. I-253-I-256.
- Messaoud, M.A., Bouzid, A. and Ellouze, N.,** 2015, 'Speech enhancement based on wavelet packet of an improved principal component analysis', *Computer Speech & Language*. Elsevier Ltd, 35, pp. 58–72.

KAYNAKLAR DİZİNİ (Devam)

- Kim, N.S., and Chang, J.H.,** 2000, ‘Spectral enhancement based on global soft decision’, *IEEE Signal Processing Letters*, 7(5), pp. 108–110.
- Nsabimana, F. X., Subbaraman, V. and Udo, Z.,** 2009, ‘A Single Channel Speech Enhancement Techniques Using Psychoacoustic Principles’, (Eusipco), pp. 170–174.
- Nsabimana, F. X., Subbaraman, V. and Zalzer, U.,** 2010, ‘A single channel speech enhancement technique exploiting human auditory masking properties’, *Advances in Radio Science*, 8, pp. 95–99.
- Oflazer, K., Say, B., Hakkani-Tür, D.Z. and Tür, G.,** 2003, Building a Turkish treebank. In Treebanks (pp. 261-277). Springer, Dordrecht.
- Paliwal, K. and Schwerin, B.,** 2009, ‘Modulation Domain Spectral Subtraction for Speech Enhancement’, *Interspeech*, 10(4), pp. 1327–1330.
- Petrovsky, A., Parfieniuk, M., and Borowicz, A.,** 2004, Warped DFT based perceptual noise reduction system. In: Proceedings of the 116th Audio Engineering Society, Berlin, Germany.
- Quatieri,** 2001, Discrete-Time Speech Signal Processing, Principles and Practice, Prentice Hall PTR.
- Ramírez, J., Segura, J. C., Benítez, C., De la Torre, Á. and Rubio, A.,** 2004, ‘Efficient voice activity detection algorithms using long-term speech information’, *Speech Communication*, 42(3–4), pp. 271–287.
- Rezayee, A. and Gazor, S.,** 2001, ‘An adaptive KLT approach for speech enhancement’, *IEEE Transactions on Speech and Audio Processing*, 9(2), pp. 87–95.
- Rissanen, J.,** 2000, MDL denoising. *IEEE Trans. Inform. Theory* 46 (7), 2537–2543.

KAYNAKLAR DİZİNİ (Devam)

- Rudresh, M. D., Rajeshwari, K. M., Sujatha, S. and Suresh, M.,** 2016, 'EMD based speech enhancement using soft and hard threshold techniques', pp. 534–541.
- Salor, O., Pellom, B., Ciloglu, T., Hacıoglu, K. and Demirekler, M.,** 2002, 'On developing new text and audio corpora and speech recognition tools for the Turkish language', *International Conference on Spoken Language Processing (ICSLP02)*, 1, pp. 349–352.
- Salor, Ö., Pellom, B. L., Ciloglu, T. and Demirekler, M.,** 2007, 'Turkish speech corpora and recognition tools developed by porting SONIC: Towards multilingual speech recognition', *Computer Speech and Language*, 21(4), pp. 580–593.
- Sanam, T. F. and Shahnaz, C.,** 2012, 'A combination of semisoft and ??-law thresholding functions for enhancing noisy speech in wavelet packet domain', *2012 7th International Conference on Electrical and Computer Engineering, ICECE 2012*, (2), pp. 884–887.
- Sanam, T. F. and Shahnaz, C.,** 2013, 'A Semisoft Thresholding Method based on Teager Energy Operation on Wavelet Packet Coefficients for Enhancing Noisy Speech', *EURASIP Journal on Audio, Speech, and Music Processing*, 2013(25), pp. 1–15.
- Schultz, T. and Waibel, A.,** 2001, 'Language-independent and language-adaptive acoustic modeling for speech recognition', *Speech Communication*, 35(1–2), pp. 31–51.
- Seok, J., and Bae, K.,** 1997, Speech enhancement with reduction of noise components in the wavelet domain. *Proc. IEEE Internat. Conf. Acoust. Speech Signal Process. (ICASSP)* 2, 1323–1326.

KAYNAKLAR DİZİNİ (Devam)

- Sendur L, and Selesnick, I.W.**, 2002, Bivariate shrinkage functions for wavelet-based denoising exploiting interscale dependency. *IEEE Transactions on Signal Processing* 50: 2744–2756.
- Sharma, R. and Mahadeva Prasanna, S. R.**, 2016, ‘A better decomposition of speech obtained using modified Empirical Mode Decomposition’, *Digital Signal Processing: A Review Journal*. Elsevier Inc., 58, pp. 26–39.
- Sheikhzadeh, H., and Abutalebi, H.R.**, 2001, An improved wavelet- based speech enhancement system. In: *Proc. 7th Eur. Conf. Speech Comm. Technol. (EuroSpeech)*, Aalborg, Denmark, September 2001.
- Sohn, J.**, 1999, *A statistical model-based voice activity detection*, *IEEE Signal Processing Letters*. doi: 10.1109/97.736233.
- Soon, I. Y., Koh, S. N. and Yeo, C. K.**, 1999, ‘Improved noise suppression using self-adaptive estimator of probability of speech absence’, 75.
- Stegmann, J. and Schroder, G.**, 1997, ‘Robust voice-activity detection based on the wavelet transform’, *1997 IEEE Workshop on Speech Coding for Telecommunications Proceedings. Back to Basics: Attacking Fundamental Problems in Speech Coding*, pp. 99–100.
- Sun, P.**, 2016, ‘Wavelet Packet Transform based Speech Enhancement via Two-Dimensional SPP Estimator with Generalized Gamma Priors’, 4(4), pp. 579–590.
- Swami, P.D., Sharma, R., Alok, J., and Swami, D.K.**, 2015, Speech enhancement by noise driven adaptation of perceptual scales and thresholds of continuous wavelet transform coefficients. *Speech Communication*, Vol.70, pp.1-12.

KAYNAKLAR DİZİNİ (Devam)

- Taal, C. H., Hendriks, R. C., Heusdens, R. and Jensen, J.,** 2010, 'A Short-time Objective Intelligibility Measure for Time-Frequency Weighted Noisy Speech'. Signal Information & Processing Lab , 2628 CD Delft , The Netherlands Oticon A / S 2765 Smørum , Denmark', *Processing*, pp. 4214–4217.
- Taal, C. H., Hendriks, R. C., Heusdens, R. and Jensen, J.,** 2011, 'An Algorithm for Intelligibility Prediction of Time – Frequency Weighted Noisy Speech', 19(7), pp. 2125–2136.
- Tohidypour, H.R., and Ahadi, S. M.,** 2015, New features for speech enhancement using bivariate shrinkage based on redundant wavelet filter-banks. *Computer Speech & Language* 35:1–23.
- Torre, Á. de la, Ramírez, J., Benítez, M. C., Segura, J. C., García, L. and Rubio, A. J.,** 2006, 'Noise robust model-based voice activity detection.', *INTERSPEECH 2006 - ICSLP, Ninth International Conference on Spoken Language Processing, Pittsburgh, PA, USA, September 17-21, 2006*, (6), pp. 1954–1957.
- Tribolet, J., Noll, P., McDermott, B., and Crochiere, R. E.,** 1978, A study of complexity and quality of speechwaveform coders, In Proc. IEEE Int. Conf. Acoust., Speech, Signal Process., 1978, pp. 586–590.
- Tucker, R.,** 1992, 'Voice activity detection using a periodicity measure', *IEE Proceedings I (Communications, Speech and Vision)*, 139(4), pp. 377–380.
- Varga, A. and Steeneken, H. J. M.,** 1993, 'Assessment for automatic speech recognition: II. NOISEX-92: A database and an experiment to study the effect of additive noise on speech recognition systems', *Speech Communication*. North-Holland, 12(3), pp. 247–251.

KAYNAKLAR DİZİNİ (Devam)

- Vincent, E., Gribonval, R., and Fevotte, C.,** 2006, Performance measurement in blind audio source separation. *IEEE Trans. Audio Speech Lang. Process.* 14 (4), 1462–1469.
- Whitmal, N., Rutledge, J., and Cohen, J.,** 1996, Reducing correlated noise in digital hearing aids. *IEEE Eng. Med. Biol. Mag.* 15, 88–96
- Woo, K., Yang, T., Park, K. and Lee, C.,** 2000, ‘Robust voice activity detection algorithm for estimating noise spectrum’, *Electronics Letters*, 36(2), pp. 180–181.
- Xu, L.,** 2007, ‘Speech Enhancement by Perceptual Adaptive Wavelet De-Noising’, Master Thesis, Windsor, Ontario, Canada.
- Xu, Y., Wang, G., Gu, Y. and Liu, H.,** 2007, ‘A Novel Wavelet Packet Speech Enhancement Algorithm Based On Time-Frequency Threshold’, *Second International Conference on Innovative Computing, Informatio and Control (ICICIC 2007)*, pp. 492–492.
- Zao, L., Coelho, R. and Flandrin, P.,** 2014, ‘Speech enhancement with EMD and hurst-based mode selection’, *IEEE Transactions on Audio, Speech and Language Processing*, 22(5), pp. 899–911.
- Zhang, L., Gao, Y. C., Bian, Z. Z. and Lu, C.,** 2005, ‘Voice activity detection algorithm improvement in adaptive multi-rate speech coding of 3GPP’, *Proceedings - 2005 International Conference on Wireless Communications, Networking and Mobile Computing, WCNM 2005*, 2(1), pp. 1257–1260.
- Zhang, Y., Tang, Z., Li, Y. and Luo, Y.,** 2014, ‘A Hierarchical Framework Approach for Voice Activity Detection and Speech Enhancement’, *The Scientific World Journal*, 2014.

ÖZGEÇMİŞ

Adı Soyadı : Özkan ARSLAN
Doğum Yeri : Bismil/Diyarbakır
Doğum Tarihi : 23/11/1984

Öğrenim Durumu :

Derece	Üniversite	Yıl
Doktora	Ege Üniversitesi Elektrik-Elektronik Mühendisliği	2014-2018
Yüksek Lisans	Ege Üniversitesi Elektrik-Elektronik Mühendisliği	2012-2014
Lisans	Süleyman Demirel Üniversitesi Elektronik ve Haberleşme Mühendisliği	2005-2010

Akademik Unvanlar :

Arş. Gör.	Namık Kemal Üniversitesi, Çorlu Mühendislik Fakültesi, Elektronik ve Haberleşme Mühendisliği Bölümü	2018-Halen
Arş. Gör.	Ege Üniversitesi, Mühendislik Fakültesi, Elektrik-Elektronik Mühendisliği Bölümü	2012-2018
Arş. Gör.	Namık Kemal Üniversitesi, Çorlu Mühendislik Fakültesi, Elektronik ve Haberleşme Mühendisliği Bölümü	2011-2012

Yabancı Dil: İngilizce, ÜDS: 68.75 (2011 Güz)

Yayınlar:

Uluslararası hakemli dergilerde yayınlan makaleler

SCI-E

Pak, C., **Arslan, Ö.**, Jalali Pahnvar, A.,and Öztürk, Y., "Self-mixing frequency measurement using a liquid filled container with elastic bottom", *Optoelectronics and Advanced Materials – Rapid Communications (OAM-RC)*, Vol. 12, No. 3-4, March-April 2018, p. 146 – 150.

Uluslararası bilimsel toplantılarda sunulan ve bildiri kitabında (Proceedings) basılan bildiriler

Y. Efteli, E.Z. Engin, **Ö. Arslan**, M. Engin, ECG Signals Based Diagnosis Using Artificial Neural Networks, 19th Biomedical Science and Technology Symposium (BIOMED 2013), 12-15 November 2013, Kuşadası, Aydın.

Ulusal bilimsel toplantılarda sunulan ve bildiri kitabında basılan bildiriler,

Ö. Arslan ve E.Z. Engin, Tek-Kanal Konuşma İyileştirme Algoritmalarının Nesnel Kalite ve Anlaşılabilirlik Ölçütleri ile Değerlendirilmesi, 26. IEEE Sinyal İşleme ve İletişim Uygulamaları Kurultayı (SIU'18) / 2-5 Mayıs 2018 / Çeşme – İzmir

Ö. Arslan, E.Z. Engin, Y. Efteli, M. Engin, F. Ögüt, Normal ve Patolojik Seslerin Ayırık Dalgacık Dönüşümü Tabanlı Nesnel Analizi, TıpTekno'13, Antalya, sy. 84-87, 2013.

Ö. Arslan, M.A. Belen, M. Karhan, 2.4 GHz ISM Bandı Uygulamaları için Çift Katlı A Sınıfı Güç Kuvvetlendiricisi Tasarımı, 4. Mühendislik ve Teknoloji Sempozyumu, 28 Nisan 2011, Çankaya Üniversitesi, Ankara, Türkiye

M. Karhan, Z. Karhan, **Ö. Arslan**, RGB ve HSV Renk Uzaylarının Renk Değerlerini Kullanarak Morfolojik Görüntü İşleme Yardımıyla Yüz Tespiti, 4. Mühendislik ve Teknoloji Sempozyumu, 28 Nisan 2011, Çankaya Üniversitesi, Ankara, Türkiye

Projeler

Ege Üniversitesi, BAP (2016/MUH/016), **Araştırmacı**, Yüz ve Ses Tanıma Tabanlı Biyometrik Sistem Tasarımı, 2016-Devam ediyor.

Ege Üniversitesi, BAP (2012/MUH/061), **Araştırmacı**, Tıbbi Tanılama için Optiksel Biyomedikal Görüntüleme, Eylül 2012-2016.