

KARADENİZ TEKNİK ÜNİVERSİTESİ
FEN BİLİMLERİ ENSTİTÜSÜ

İSTATİSTİK VE BİLGİSAYAR BİLİMLERİ ANABİLİM DALI

SPEKTRAL TERS ENTROPİ ÖZNETLİĞİNİN AĞIRLIKLI BULANIK
KÜMELENMESİ İLE SES VERİLERİNİN BÖLÜTLENMESİ

YÜKSEK LİSANS TEZİ

Nasim JABBARİ

EKİM 2014
TRABZON

KARADENİZ TEKNİK ÜNİVERSİTESİ
FEN BİLİMLERİ ENSTİTÜSÜ

İSTATİSTİK VE BİLGİSAYAR BİLİMLERİ ANABİLİM DALI

**SPEKTRAL TERS ENTROPİ ÖZNETELİĞİNİN AĞIRLIKLI BULANIK
KÜMELENMESİ İLE SES VERİLERİNİN BÖLÜTLENMESİ**

Nasim JABBARI

**Karadeniz Teknik Üniversitesi Fen Bilimleri Enstitüsünde
"YÜKSEK LİSANS (İSTATİSTİK)"
Unvanı Verilmesi İçin Kabul Edilen Tezdir.**

Tezin Enstitüye Verildiği Tarih : 16.09.2014
Tezin Savunma Tarihi : 09.10.2014

Tez Danışmanı : Yrd. Doç. Dr. Orhan KESEMEN

Trabzon 2014

Karadeniz Teknik Üniversitesi Fen Bilimleri Enstitüsü
İstatistik ve Bilgisayar Bilimleri Anabilim Dalında
Nasim JABBARI tarafından hazırlanan

SPEKTRAL TERS ENTROPİ ÖZNETELİĞİNİN AĞIRLIKLI BULANIK
KÜMELENMESİ İLE SES VERİLERİNİN BÖLÜTLENMESİ

başlıklı bu çalışma, Enstitü Yönetim Kurulunun 23 / 09 / 2014 gün ve 1571 sayılı
kararıyla oluşturulan jüri tarafından yapılan sınavda
YÜKSEK LİSANS TEZİ
olarak kabul edilmiştir.

Jüri Üyeleri

Başkan : Prof. Dr. Funda KARAÇAL

Üye : Yrd. Doç. Dr. Orhan KESEMEN

Üye : Yrd. Doç. Dr. Halil İbrahim ŞAHİN

Prof. Dr. Sadettin KORKMAZ
Enstitü Müdürü

ÖNSÖZ

Bu tez, Karadeniz Teknik Üniversitesi Fen Bilimleri Enstitüsü İstatistik ve Bilgisayar Bilimleri Anabilim Dalı, Yüksek Lisans Programı'nda hazırlanmıştır. Bu tez, "Spektral Ters Entropi Özniteliğinin Ağırlıklı Bulanık Kümelenmesi ile Ses Verilerinin Bölütlenmesi" üzerine yapılmıştır. Tez çalışması sırasında değerli yardım ve katkılarıyla beni yönlendiren danışman hocam Yrd. Doç. Dr. Orhan KESEMEN 'ne, teşekkürlü bir borç bilirim.

Çalışmalarım boyunca maddi ve manevi desteklerini esirgemeyen anne ve babama teşekkür ederim. Bu tezin, bundan sonraki çalışmalara katkı sağlamasını temenni ederim.

Nasim JABBARI

Trabzon 2014

TEZ BEYANNAMESİ

Yüksek Lisans Tezi olarak sunduğum “Spektral Ters Entropi Özniteliğinin Ağırlıklı Bulanık Kümelenmesi ile Ses Verilerinin Bölütlenmesi” başlıklı bu çalışmayı baştan sona kadar danışmanım Yrd. Doç. Dr. Orhan KESEMEN’in sorumluluğunda tamamladığımı, verileri/örnekleri kendim topladığımı, deneyleri/analizleri ilgili laboratuvarlarda yaptığımı/yaptırdığımı, başka kaynaklardan aldığım bilgileri metinde ve kaynakçada eksiksiz olarak gösterdiğimi, çalışma sürecinde bilimsel araştırma ve etik kurallara uygun olarak davrandığımı ve aksinin ortaya çıkması durumunda her türlü yasal sonucu kabul ettiğimi beyan ederim. 12/09/2014

Nasim JABBARI

İÇİNDEKİLER

	<u>Sayfa No</u>
ÖNSÖZ	III
TEZ BEYANNAMESİ.....	IV
İÇİNDEKİLER.....	V
ÖZET	VII
SUMMARY	VIII
ŞEKİLLER DİZİNİ	IX
SEMBOLLER DİZİNİ	XII
1. GENEL BİLGİLER.....	1
1.1. Ses	1
1.1.1. Ses Dalgası	2
1.1.2. Dalga Boyu.....	2
1.1.3. Dalga Periyodu.....	3
1.1.4. Genlik	3
1.1.5. Sesin Frekansı.....	4
1.1.6. Sesin Hızı	4
1.1.7. Sesin Şiddeti	5
1.2. Ses Tanıma	6
1.3. Ses Sinyali Bölütleme	7
1.4. MFCC Öznitelik Çıkarma	10
1.4.1. Çerçeveleme	12
1.4.2. Pencereleme.....	13
1.4.3. Hızlı Fourier Dönüşümü.....	14
1.4.4. Mel Frekans Kepstrum Katsayıları.....	15
1.4.5. Kepstrum	16
1.5. Sıfır Geçiş Oranı.....	17
1.6. Fark Fonksiyonunun Ortalama Büyüklüğü	18
1.7. Enerji Eşik Değeri Yöntemi	20

1.7.1.	Çerçeve Enerjisi	20
1.7.2.	Eşikleme Değeri	20
1.8.	Entropik Bölütlemesi.....	20
1.9.	Ayrık Fourier Dönüşümü	22
1.10.	Kümeleme Analizi.....	24
1.10.1.	K- Ortalama Kümeleme Algoritması	26
1.10.2.	Bulanık C-Ortalama Algoritması	27
1.10.3.	Bulanık Kümeleme Algoritmasında Optimal Küme Sayısı	29
2.	YAPILAN ÇALIŞMALAR	32
2.1.	Ön İşlemler	32
2.2.	Ortalama Kayması (DC) Düzeltmesi	32
2.3.	Ortalama Silme Yöntemi.....	33
2.4.	Genlik Eşitleme	33
2.5.	Normalleştirme.....	34
2.6.	Spektrogram	35
2.7.	Ses Sinyali Spektrogramı	36
2.8.	Olabilirliğin Elde Edilmesi.....	37
2.9.	Ağırlıklı Bulanık C-Ortalama Kümelemesi	38
2.10.	Uygulama Örneği 1	38
2.11.	Uygulama Örneği 2	47
3.	BULGULAR	56
4.	ÖNERİLER	57
5.	KAYNAKLAR.....	58

ÖZGEÇMİŞ

Yüksek Lisans

ÖZET

SPEKTRAL TERS ENTROPİ ÖZNETELİĞİNİN AĞIRLIKLIL BULANIK
KÜMELENMESİ İLE SES VERİLERİNİN BÖLÜTLENMESİ

Nasim JABBARI

Karadeniz Teknik Üniversitesi
Fen Bilimleri Enstitüsü
İstatistik ve Bilgisayar Bilimleri Anabilim Dalı
Danışman: Yrd. Doç. Dr. Orhan KESEMEN
2014, 61 Sayfa

Müzik sesinden elde edilen sayısal ses verileri çok karmaşık yapılar ve çeşitli stiller içermektedir. Bu çalışmadaki amaç, bulanık kümeleme yöntemiyle ses verisini içeriğine göre bölütlere ayırarak her notanın bölgesini kestirmektir. Literatürde bölütleme için geliştirilen birçok yöntem bulunmaktadır. Bu yöntemlerden bazıları; Mel Frekans Kepstrum Katsayıları, Sıfır Geçiş, Fark Fonksiyonların Ortalama Büyüklüğü, Enerji Eşik Değeri Yöntemi, Entropi ve Ayrık Fourier dönüşümü yöntemleridir. Bu yöntemlerden en çok kullanılan yöntem eşik değeri yardımıyla yapılan bölütleme işlemidir. Fakat bu yöntem birçok ses verisi için uygun değildir. Eğer verideki genlik değerleri artan bir fonksiyon oluşturuyorsa, nota bölütleri art arda çalınmaya başlanmışsa ya da notada herhangi bir boğumlanma olmuşsa bu yöntem istenilen verimde bir bölütleme yapamaz.

Bu çalışmada bölütleme için uygulanabilecek yeni bir yöntem önerilmiştir. Bu yöntemde ses verisi üzerinde küçük çerçeveler alınarak bir çeşit öznitelik çıkarımı yapılır. Bu öznitelik katsayılarına göre bir bulanık kümeleme uygulanarak bölütleme bölgeleri belirlenir. Böylece seslilik bölgeleri belirlenmiş olur. Bu durumda sessizlik bölgeleri düşük üyelik değerlerinde seslilik kümelerine katılır. Her seslilik bölgesi bulanık bir kümeyle bölütlere ayrılmış olur.

Anahtar Kelimeler: Ses sinyali bölütleme, Bulanık bölütleme, Öznitelik çıkarıma

Master Thesis

SUMMARY

SEGMENTATION OF AUDIO DATA WITH WEIGHTED FUZZY CLUSTERING TO
ATTRIBUTE OF SPEKTRAL INVERSE ENTROPY

Nasim JABBARI

Karadeniz Technical University
The Graduate School of Natural and Applied Sciences
Statistical and Computer Science Graduate Program
Supervisor: Assist. Prof. Orhan KESEMEN
2014, 61 Pages

The numerical audio data obtained from the music volume involve very complex structures and various styles. The purpose of this paper, is both to separate the audio data into segments based on the content of them using fuzzy clustering method and also to predict the region of each note. In the literature, there are many methods developed for segmentation. Some of these methods are: Mel Frequency Campestral Coefficients, Zero-Crossing, Average Magnitude Difference Function, Energy Threshold Value, Entropy and Discrete Fourier transform methods. The most commonly used method is segmentation process using threshold value. This method cannot segment efficiently in some situations provided that the amplitude of values create an increasing function or musical notes and start to play in succession or any made musical notes fade out. This method cannot desire segmentation efficiency. In this paper, a new method is proposed that uses segmentation.

In this method, various feature coefficient extraction is performed by taking small frameworks on the audio data. The regions of the segment are determined using fuzzy clustering according to these feature coefficients, therefore, the silence regions are determined. In this case, silence zones join silence clusters taking low membership values. Therefore, each silence region becomes separated from the segment with a fuzzy cluster.

Key Words: Audio segmentation, Fuzzy segmentation, Feature extraction

ŞEKİLLER DİZİNİ

	<u>Sayfa No</u>
Şekil 1. Bir ses sinyali üzerinde dalga boyu gösterimi	3
Şekil 2. Bir ses sinyali üzerinde dalga periyot ve dalga genliğini gösterimi	3
Şekil 3. Bir ses sinyali üzerinde frekans gösterimi	4
Şekil 4. Zayıf bir ses ile şiddetli bir ses sinyalinin gösterimi	6
Şekil 5. Öznitelik çıkarma.....	11
Şekil 6. Bir konuşmanın çerçevelere bölünmesi.....	13
Şekil 7. MFCC özniteliklerinin çıkarılma işleminin blok diyagramı	16
Şekil 8. (a) Örnek bir konuşma sesi sinyali; (b) (a)'daki sesin sıfır geçiş oranının grafiği.....	18
Şekil 9. (a) Örnek bir konuşma sesi sinyali; (b) (a)'daki sesin fark fonksiyonunun ortalama büyüklük grafiği.....	19
Şekil 10. Enerji Entropi diyagramı	22
Şekil 11. /a/ harfi konuşma sinyali; (a) 1 saniyelik /a/ harfi konuşma sinyali; (b) 0.2sn-0.3sn aralıktaki /a/ harfi konuşma sinyali; (c) 1 saniyelik /a/ harfi konuşma sinyalinin ayrık Fourier dönüşümü	23
Şekil 12. Genlik eşitlemesi; (a) İşlenmeye hazır ses sinyali; (b) Genliği dengelenmiş bir ses sinyali	34
Şekil 13. Normalize edilmiş bir sinyal.....	35
Şekil 14. Spektrogram gösterimi, (a) frekansı zamanla artan yapay ses sinyali, (b) spektrogram gösterimi	36
Şekil 15. Flütle çalınmış bir notanın spektrogramı	37
Şekil 16. Spektrogramın ters entropi uygulanarak olabirlikleri elde edilmiş ses sinyali.....	38
Şekil 17. Örnek ses sinyali.....	40
Şekil 18. Ortalama kaymansı uygulanmış ses sinyali.....	40
Şekil 19. Düşük frekans bileşenleri elenmiş ses sinyali	41
Şekil 20. Şekil 19'da verilen ses sinyalinin enerji düzeyleri ve eşik seviyesi	41
Şekil 21. Eşikleme yapılmış ses sinyali	42
Şekil 22. Genlik eşitlemesi uygulanması	42

Şekil 23.	Genlik eşitlemesi uygulanmış sinyalinin normalize edilmesi	43
Şekil 24.	Şekil 23'te verilen genlik eşitlemesi yapılmış sinyalin enerji düzeyi.....	43
Şekil 25.	Eşitlenmiş sinyalin spektrogramı.....	44
Şekil 26.	Olabilirlik değerleri	44
Şekil 27.	Olabilirlik değerlerine göre sinyalin kümelenmesi.....	45
Şekil 28.	Şekil 19'da verilen sinyal ile birinci kümenin üyelik değerlerinin çarpımıyla yapılan pencereleme sonucu elde edilen bulanık bölütlenmiş sinyal.....	45
Şekil 29.	Şekil 19'da verilen sinyal ile ikinci kümenin üyelik değerlerinin çarpımıyla yapılan pencereleme sonucu elde edilen bulanık bölütlenmiş sinyal.....	46
Şekil 30.	Şekil 19'da verilen sinyal ile üçüncü kümenin üyelik değerlerinin çarpımıyla yapılan pencereleme sonucu elde edilen bulanık bölütlenmiş sinyal.....	46
Şekil 31.	Şekil 19'da verilen sinyal ile dördüncü kümenin üyelik değerlerinin çarpımıyla yapılan pencereleme sonucu elde edilen bulanık bölütlenmiş sinyal.....	47
Şekil 32.	Örnek ses sinyali.....	48
Şekil 33.	Ortalama kaymansız uygulanmış ses sinyali	49
Şekil 34.	Düşük frekans bileşenleri elenmiş ses sinyali	49
Şekil 35.	Şekil 29'da verilen ses sinyalinin enerji düzeyleri ve eşik seviyesi	50
Şekil 36.	Eşikleme yapılmış ses sinyali	50
Şekil 37.	Genlik eşitlemesi uygulanması	51
Şekil 38.	Genlik eşitlemesi uygulanmış sinyalinin normalize edilmesi	51
Şekil 39.	Şekil 34'te verilen genlik eşitlemesi yapılmış sinyalin enerji düzeyi.....	52
Şekil 40.	Eşitlenmiş sinyalin spektrogramı.....	52
Şekil 41.	Olabilirlik değerleri	53
Şekil 42.	Olabilirlik değerlerine göre sinyalin kümelenmesi.....	53
Şekil 43.	Şekil 19'da verilen sinyal ile birinci kümenin üyelik değerlerinin çarpımıyla yapılan pencereleme sonucu elde edilen bulanık bölütlenmiş sinyal.....	54

Şekil 44.	Şekil 19’da verilen sinyal ile birinci kümenin üyelik değerlerinin çarpımıyla yapılan pencereleme sonucu elde edilen bulanık bölütlenmiş sinyal.....	54
Şekil 45.	Şekil 19’da verilen sinyal ile birinci kümenin üyelik değerlerinin çarpımıyla yapılan pencereleme sonucu elde edilen bulanık bölütlenmiş sinyal.....	55
Şekil 46.	Şekil 19’da verilen sinyal ile birinci kümenin üyelik değerlerinin çarpımıyla yapılan pencereleme sonucu elde edilen bulanık bölütlenmiş sinyal.....	55

SEMBOLLER DİZİNİ

dB	: Ses şiddeti ölçü birimi (desi Bell)
MFCC	: Mel Frekansı Kepstral Katsayısı
DFT	: Ayrık Fourier Dönüşümü
FFT	: Hızlı Fourier Dönüşümü
ZCR	: Sıfır Geçiş Oranı
AMDF	: Fark Fonksiyonun Ortalama Büyüklüğü
FCM	: Bulanık c-Ortalama Kümeleme Algoritması
PC	: Bezdek Bölünme Katsayısı
PE	: Bezdek Bölünme Entropisi
XB	: Xie-Beni kriteri
FS	: Fukuyama ve Sugeno kriteri

1. GENEL BİLGİLER

Son yıllarda mobil teknolojilerinin gelişmesiyle birlikte, video ve ses verilerine olan talep çığ gibi büyüdü. Bu talep sadece iletim aşamasında değil işleme ve saklama aşamasında da gelişmeleri kendisiyle birlikte getirdi. Ses işleme, ses tanıma, ses birleştirme, ses oluşturma, ses arama gibi işlemler sayısal hale getirilen veri üzerinde rahatlıkla uygulanmaya başlandı.

Ses tanıma uygulamalarında, sesin öncelikle karakteristiğine göre bölgelere ayrılması gerekir. Ancak ses sinyalleri birçok karmaşık yapıyı içerilerinde barındırdığından dolayı literatürde birçok yöntem olmasına rağmen halen tam anlamıyla bölütleme yapan bir yöntem geliştirilememiştir.

1.1. Ses

Esnek cisimlerin titreşmesi sonucu periyodik basınç değişimleri meydana gelir. Bu basınç değişimlerine ses adı verilir. Sesler bir çeşit enerji türüdür ve atmosferde dalgalar halinde yayılırlar (Eskidere & Ertaş, 2009). Ses üreten varlıklara ses kaynağı denir. Titreşen bir cismin duyulabilir bir ses kaynağı oluşturabilmesi için titreşimin saniyede en az 20 kez olması gerekir. Ses kaynaklarını doğal ve yapay olmak üzere ikiye ayırabiliriz. Doğal ses kaynakları insan sesi, su sesi, hayvan sesi gibi ses üreten varlıklardan oluşurken; yapay ses kaynakları, insanların ürettikleri müzik aletleri, hoparlör, taşıtlar gibi varlıklardır.

Canlılar sesleri ciğer, nefes borusu, gırtlak, boğaz, ağız boşluğu ve burun boşluğu gibi bölümlerden oluşan ses sistemi sayesinde oluştururlar. Bu ses sisteminde ses telleri nefes ile titreşir ve böylece ses ortaya çıkmış olur. Yapay ses kaynaklarından olan hoparlörlerdeki ses ise hoparlörün içindeki koninin ileri – geri titreşmesinden oluşur. Bu koni ileri dışa doğru hareket ettiğinde önündeki hava moleküllerini iterek sıkıştırır. İleri doğru hareket ettiğinde önündeki hava moleküllerinin arasını açar. Bu şekilde sürekli sıkışan ve ayrılan hava molekülleri ses dalgalarını oluşturur.

Seslerin işitilmesi ise ses oluşturmaktan daha zorlu bir süreçtir. Öncelikle kulak kepçesi ile kemikleri bu titreşimi oval pencereye iletir. Oval pencereden ileriye iletilen titreşimler buradan da salyangoza geçer. Ses, salyangozdaki işitme almaçları tarafından

algılanır ve işitme sınırları aracılığıyla beyindeki işitme merkezine iletilir. Böylece işitme olayı gerçekleşmiş olur. Fakat her canlının bir duyma eşiği vardır. Bu yüzden her canlı her sesi duyamaz. Örneğin yunuslar insanların duyamayacağı çok ince sesleri algılayabilir, köpekler fay hatlarının kırılma seslerini duyabilirler (Ses Dalgası ve Sesin Özellikleri, 2014).

1.1.1. Ses Dalgası

Bir kaynak tarafından oluşturulan titreşimin bir noktadan diğer noktalara iletilmesiyle ortaya çıkan olaya dalga denir. Dalga hareketinde yayılan ortam değil enerjidir. Ses kaynaklarından çıkan sesler yayılırken hava taneciklerini çarpıştırarak onları titreştirirler. Böylelikle ses dalgaları oluşur.

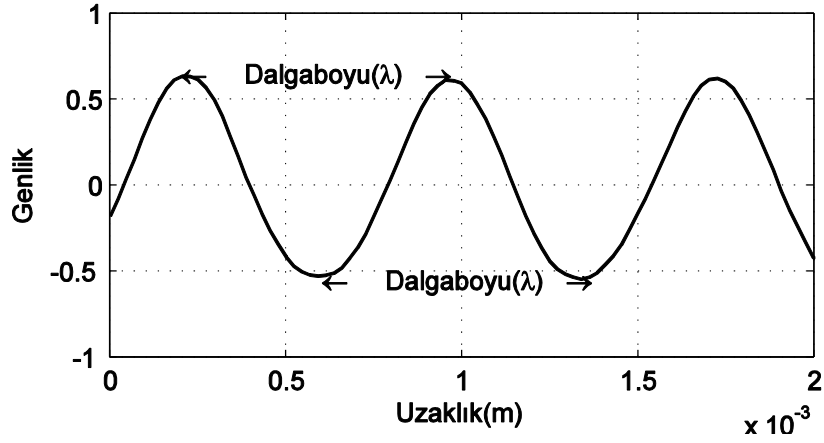
Ses dalgaları katı, sıvı ve gaz ortamlarda yayılırken boşlukta yayılmaz. Çünkü boşlukta sesin yayılabileceği maddesel bir ortam yoktur. Titreşimin oluşturduğu sesi bizim kulağımıza taşıyan ortam genellikle havadır.

Ses dalgaları bir yüzeye çarptıklarında gösterdiklerin basıncın ölçülmesiyle elde edilen veriler zamana göre sinüzoidal şekilde gösterilebilirler (Şekil 1).

Ses dalgalarının yayılması su dalgalarına benzer bir yapıdadır. Bu nedenle tüm dalgalarda olan özellikler ses dalgalarında da mevcuttur. Bu özellikler: Dalga boyu, Periyot, Frekans, Genlik ve Hız'dır.

1.1.2. Dalga Boyu

Dalga boyu, bir ses dalgasının oluşması için sesin aldığı yoldur (Eskidere & Ertaş, 2009). Ardışık iki dalga çukuru ya da iki dalga tepesi arasındaki yatay uzaklığa dalga boyu denir (Şekil 1). Dalga boyu (λ) sembolü ile gösterilir. Dalga boyunun birimi metre (m) dir.



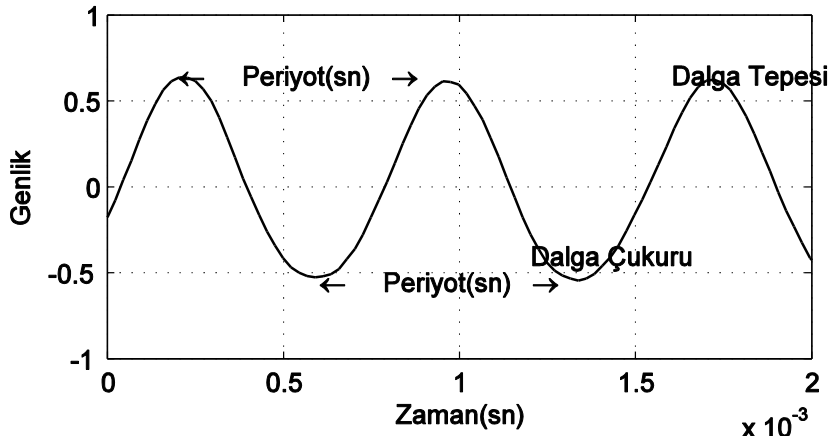
Şekil 1. Bir ses sinyali üzerinde dalga boyu gösterimi

1.1.3. Dalga Periyodu

Bir tam dalganın oluşması için geçen süreye veya bir noktadan ardışık iki dalga çukuru ya da iki dalga tepesinin geçmesi için gereken süreye periyot denir (Şekil 2). Periyot T ile gösterilir. Birimi saniye(s) dir.

1.1.4. Genlik

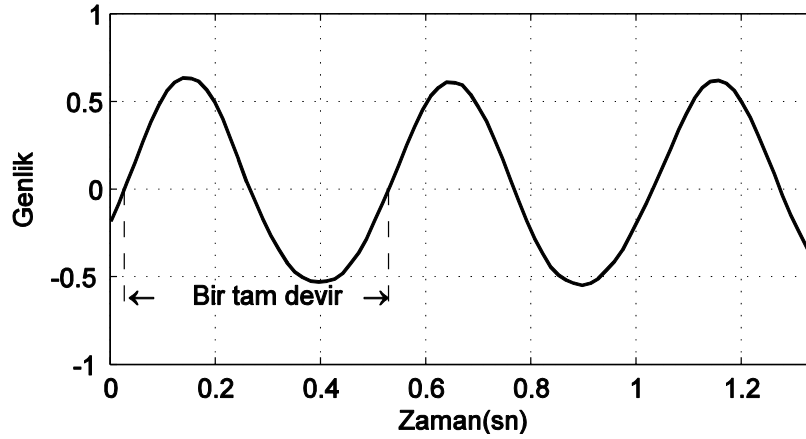
Bir dalga tepesinin ya da çukurunun durgun konumuna olan uzaklığına genlik denir. Dalgaların taşıdığı enerji genlikleriyle orantılıdır. Bir ses dalgasının genliği ne kadar büyükse sesin de şiddeti o kadar büyük olur. Şekil 2’de bir ses sinyali üzerinde dalga periyodu ve dalga genliği gösterilmektedir.



Şekil 2. Bir ses sinyali üzerinde dalga periyodu ve dalga genliğini gösterimi

1.1.5. Sesin Frekansı

Birim zamanda oluşan dalga sayısına frekans denir. Farklı şekilde anlatılırsa frekans bir saniyedeki titreşim sayısıdır ve birimi s^{-1} (Hz) dir. Sesin frekansı sadece ortama bağlıdır. Sesin frekansı arttıkça üretilen dalga sayısı artacağından ses kesintisiz gibi duyulur ve frekans azaldıkça ses daha kısa aralıklarda kesintili gibi duyulur. İnsan kulağı frekansı 20 Hz ile 20.000 Hz arasındaki sesleri duyabilir. Duyulmayan seslerde 20 Hz'den küçük ve 20.000 Hz'den büyük frekanslı seslerdir. İnsan kulağının duyabileceği sesin frekansı diğer canlılara göre farklıdır. Sesin frekansı sesin ince ve ya kalın olmasını belirleyen özelliktir. 20 Hz'e doğru yaklaşan sesler kalın ses ve 20.000 Hz e doğru çıkan sesler de ince sesler olarak adlandırılır. Frekansı belirleyen ve aynı zamanda periyodu değiştiren tek unsur ses dalgasını oluşturan dalga kaynağıdır. Örneğin konuşma sırasında ses telleri belli aralıklarla titreşirler. Bu titreşim erkeklerde saniyede ortalama 130kez kadınlarda ise 220 kez tekrarlanır. Titreşim sayısı sabit olmayıp çıkarılacak sese ve vurguya göre değişebilmektedir. Şekil 3.'de saniyede 2 Hz'lik bir ses sinyali ve üzerindeki 1 Hz frekansı gösterilmektedir.



Şekil 3. Bir ses sinyali üzerinde frekans gösterimi

1.1.6. Sesin Hızı

Ses dalgalarının birim zamandaki yer değiştirmesine dalga hızı denir. Ses dalgaları ortama bağlı olarak yayılır. Ses dalgaları katı, sıvı ve gaz maddelerinin hepsinde yayılma

özelliğine sahiptir. Fakat her maddeden geçme hızları farklıdır. Ses dalgaları hava gibi gaz cisimlerde yayılma hızı düşüktür. Su gibi sıvılarda ise hızı biraz daha fazladır beton veya kayaç gibi katı maddelerde hızı oldukça yüksektir. Bu ifadelerden yola çıkılarak (1) eşitsizlikler verilebilir.

$$V_{katı} > V_{sıvı} > V_{gaz} \quad (1)$$

Sesin hızı, ortamın sıcaklığından, basıncından ve ortamın yoğunluğundan etkilenir. Sıcaklık arttıkça sesin hızı da artar. Bunun nedeni sıcaklık artışının moleküllerin titreşim hızlarını arttırmasıdır. Sesin hızı normal koşullarda; havada 340 m/s, suda 1400-1600 m/s, çelikte ise 8000 m/s dir (Ses Dalgası ve Sesin Özellikleri, 2014). Ses hızı, sesin dalga boyu ve frekansına göre bulunabilir. (2) eşitliğinde bu durum verilmektedir,

$$v = \lambda \cdot f \quad (2)$$

burada v dalga hızını verirken λ ve f sırasıyla dalga boyunu ve dalga frekansını vermektedir. Diğer bir yaklaşımda ise dalga boyu ile periyot arasındaki ilişki

$$v = \frac{\lambda}{T} \quad (3)$$

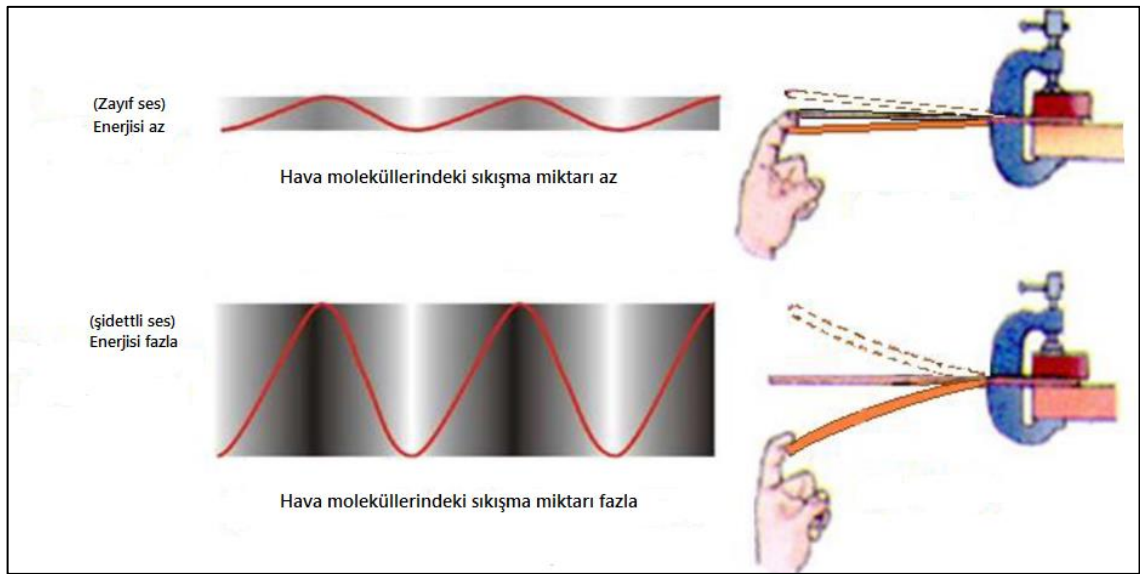
eşitliğini vermektedir. Burada T dalga periyodunu vermektedir.

1.1.7. Sesin Şiddeti

Bir sesin işitilebilmesini sağlayan, özellikle kulak tarafından duyulma yüksekliğine sesin şiddeti denir. Ses dalgalarının genliği ne kadar büyük ise taşıdığı enerjinin bir göstergesi olarak daha şiddetli dalgalar elde edilir. Ses dalgalarının şiddeti kulak zarına yaptığı basınçla orantılıdır. Bu nedenle işitme yeteneği zayıf olan insanlar kulaklarına işitme cihazı takarak sesin şiddetini arttırmaya çalışırlar.

Ses farklı madde veya ortamlarda farklı yayılması nedeniyle ses kaynakları aynı olsa bile farklı ortamlarda oluşan aynı sesin şiddetleri farklıdır. Yoğun ortam, ses şiddetini azaltır. Böylece sesin daha az işitilmesine sebep olur. Hafif sesler kulağa az şiddette

kuvvetli sesler ise yüksek şiddette gelmektedir. Sesin şiddet birimi desibel'dir (dB) . İşitilebilen en hafif ses 0 dB'dir. Buna işitme eşiği denir. Normal şiddetteki sesler 30-60 dB arasındadır. Şiddetli ve rahatsız eden sesler 60 dB'den fazla olan seslerdir. Ses kaynağına yapılan etkinin artması ile birlikte çıkardığı sesin şiddeti de artar. Ses kaynağından uzaklaştıkça sesin şiddeti azalır ve ses kaynağına yakın yerlerde şiddet artar. Sesin şiddeti, ses dalgalarının enerji ve genliği ile doğru orantılıdır. Yani sesin şiddetini arttırdığında sadece taşıdığı enerji artar sesin diğer özelliklerinde değişme olmaz. Şekil 4'de zayıf bir ses ile şiddetli bir ses sinyalinin gösterimi.



Şekil 4. Zayıf bir ses ile şiddetli bir ses sinyalinin gösterimi

1.2. Ses Tanıma

Ses tanıma, konuşmacı tarafından söylenen sözleri tanımlamak için işitsel sinyallerin analizi işlemidir. Ses tanıma süreci, sesin kaydedilmesi ile başlar, sesin işlenmesi, öz niteliklerinin çıkarılıp kaydedilmesi, karşılaştırma yapılarak sesin tanınması ile son bulur. Ses tanıma teknolojisinin kökeni 18. yüzyılın ikinci yarısına rastlamaktadır. O sıralarda konuşmayı yazıya çevirmekten çok “konuşan” makineler yapılmaya çalışılmıştır. 1773'te Rus bilim adamı Kratzenstein, kilise orglarının boruları ve rezonans tüplerini bağlayarak sesli harfler çıkarmayı başarmıştır. Ardından Viyanalı von Kempelen, Akustik-Mekanik Ses Makinesi icat etmiştir. Ses tanıma adına daha büyük adımlar 20. yüzyılda, birçok teknolojinin doğduğu Bell Laboratuvarları'nda atılmıştır. Homer Dudley ve Harvey

Fletcher konuşmada çıkan sesi tanımlamada sinyal aralığının önemli olduğunu bulmuşlardır. Modern ses tanıma algoritmalarının temeli de onların yöntemlerine dayanmaktadır. 1962'de yine Bell Laboratuvarları'nda yalıtılmış sözcük tanıma yapılmıştır. Bu noktadan sonra kademeli olarak ses tanıma kapasitesi daha çok ve daha karmaşık söz dizilerini çözümlmeye kadar gidilebilir (Mandıracıoğlu, 2014).

Muscle Fish LLC tanımlama ve sınıflandırma sistemi Erling Wold (Wold, 1996) tarafından 1996 yılında geliştirildi. Bu çalışma daha önceki çalışmalarla kıyasladığında çok farklı ve daha geçerlidir. Bu yöntemde, ses sinyallerinin analiz ve sınıflandırılması için sesin akustiğiyle ilgili dört özellik kullanılmaktadır. Bunlar: ses perdesi, genlik, ton kalitesi ve bant genişliğidir. Yöntem, normalleştirilmiş öklid uzaklığı ve en yakın komşu sınıflandırması kurallarını kullanarak, değişik ses sinyallerini uygun sınıflara ayırır (Wold, 1996; Stan, 2000).

Başka bir çalışma 1997 yılında Foote tarafından ortaya atılmıştır. Bu yöntem müzik ve sessiz seslerden, konuşmayı ayıran bir yöntemidir (Foote, 1997). Bu yöntem Mel Frekans Kepstrum Katsayıları özniteliğinin ağaç yapısını kullanarak, her örneğin ses frekanslarına göre bir histogramını oluşturur ve bunu bir özellik vektörü olarak kullanır. Daha sonra özellik vektörlerinin alanı, istenen sayıda ayırık bölgeye ayrılır ve sonunda, sınıflandırma kosinüs mesafe ve yakın komşular yöntemleriyle gerçekleştirilir.

1.3. Ses Sinyali Bölütleme

Nesnenin birbirinden bağımsız parçalara ayrılmasına bölütleme denir. İnsanoğlu kesintisiz konuşulan bir konuşmada, kelimeleri bölütlere ayırmada doğuştan bir yeteneğe sahiptir. Bu yetenek 8 aylık olduğundan itibaren bebeklerde gelişmeye başlar. Burada bölütleme denetimsiz bir şekilde gerçekleştirilir (Masataka, 2003).

Ses sinyallerinde bölütlemenin temel hedefi giriş ses akışını akustik homojen bölütlere bölmek ve ses sinyal yapısı hakkında bilgiler çıkartmaktır. Ses bölütleme, çeşitli ses tanıma uygulamalarda önemli etkisi olan bir ön işlem adıdır.

Ses bölütleme konusunda çeşitli yöntemler geliştirilmiştir. Chen 2005 yılında, iki tür bölütleme tanımlamıştır. Bunlar sınıflandırmaya bağımlı bölütleme (CDS) ve sınıflandırmadan bağımsız bölütleme (CIS). CIS yöntemleri, hangi ses özelliklerini kullandığı, zaman ve frekans alanında gözetimli veya gözetimsiz yöntemleri, önceki ses bölütleme sonuçlarına bağlı olup olmadığına göre gruplara ayrılabilirler. Bundan ziyade

CIS yöntemi, modele-dayalı olan bir yöntem veya modele dayalı olmayan bir yöntem olarak tanımlanabilir (Cheng, Tan, & Chen , 2005).

Panagiotakis ve Tziritas 2005 yılındaki çalışmalarında zaman ortamlı yöntemlere ağırlık vermişlerdi. Karekök ortalamasını (RMS) ve sıfır geçiş oranını (ZCR) müziği konuşma bölümünden ayırmak için kullanmışlardır (Panagiotakis & Tziritas, 2005).

Tzanetakakis ve Cook 1999 yılında çoklu özelliklere dayalı zamansal ortamlı bölütleme için genel bir metodoloji tanımlamışlardır (Tzanetakakis & Cook, 1999).

Mohammad A. Haque ve Jong-Myon Kim 2013 yılında gelişmiş bir yaklaşım önerdiler Bu yöntemi, korelasyon yoğunluklu bulanık c-ortalama (CIFCM) algoritması olarak adlandırdılar. Önerilen CIFCM algoritması, komşu çerçeveler veya ses akısının bölütlerinin etkisini içermektedirler. Bu yöntemde sinyal, soldan boğumlama, sağdan boğumlama ve çapraz boğumlama gibi ses efektleri içeriyor bile olsa ses kesintileri etkin bir şekilde algılanabilir. Bu algoritma farklı seslerin türleri arasındaki sınırları tespit ederek çalışır ve sessizlik, konuşma, müzik, müzik ile konuşma ve gürültü ile konuşma gibi kümeler halinde sınıflandırabilir (Hague & Myon, 2013).

Michael M. Goodwin ve Jean Laroche tarafından 2004 önerilen yöntemlerinde bir ses sinyalini, küme benzerlik özelliğine dayalı karakteristik bölgelere ayırmaktadır. Bunun için öncelikle sinyalin bir özellik-uzayı oluşturulur. Özellik-uzayının doğrusal diskriminant dönüşüm (LDA) analizi ile diskriminant altuzayı kullanılarak özelliklerin kümelere ayrılması için tasarlanmıştır. Algoritma doğrusal diskriminant analizi (LDA), dinamik programlama (DP) ve özellik-uzayı koşulu altında veri kümeleri tanımlamak için kullanılmıştır. Beklenen veri kümelerinin yapısı hakkında ön bilgilere dayanmaksızın etkin bir maliyet fonksiyonu elde edildi (Goodwin & Laroche, 2004, May).

Peiszer ve diğ. tarafından 2008 yılında önerdikleri yöntemde otomatik ses bölütleme ile bir şarkının yapısı üzerinde, bilgi çıkarma amaçlamaktadırlar. Burada özellik algoritması ile son yıllardaki popüler pop şarkıların yapılarını ve bölüt sınırları ortaya koymaktadır (Peiszer, Lidy, & Rauber, 2008).

Kulkarni ve diğ., 2001 yılında, ses parçalarının yapısal bileşenlerine bölütlemek için yeni bir algoritma önerdiler. Bu algoritma, homojen bölgelerin sınırlarını belirlerken zaman ve frekans ortamında çeşitli özelliklerine göre karar verir. Bu algoritma, iki aşamada tasarlanmıştır. Birinci aşamasında, sessiz ve sesli bölge sınıflandırma çok terimli softmax regresyon kullanılarak yapılır. İkinci aşamada önceki çıkış verileri üzerinde gizli

Markov modeli (smooth) düzgünleştirilerek kullanılmıştır (Kulkarni, Iyer, & Sridharan, 2001).

David ve diğ., 2009 yılında farklı bilinen bölümlenme yöntemlerin avantajlarını birleştiren yeni bir yapıyı tanıttılar. Otomatik logaritmik-doğrusal bölüt tahmin edici modeli ses akısının bütünsel şekilde kodlama stratejiyle bölütlenmesini gerçekleştirmek için yerel değişim noktaları sınıflandırması kullanılmıştır (Davis & Mermelstein, 1980).

Başka bir çalışma da Zhang ve diğ., tarafından 2009 yılında ortaya atılmıştır. Yöntemin ilk aşamasında, sinyal zaman ortamında sessizlik algılama ile bölütlere ayrılır. Bu ilk bölütlemeden sonra, başka bölütlemeler frekans ortamında gerçekleştirilir. Bu yöntem, ses özelliklerinin homojen olmasından yararlanılarak yapılır. Yöntemde ses özelliklerindeki herhangi birinin ani değişikliği, yeni bir ses sinyalinin başlangıcını gösterdiğini varsayar (Zhang, Whalley, & Brooks, 2009).

Ajmera ve diğ., 2003 yılında pratik uygulamalar üzerinde yüksek performanslı konuşma / müzik ayrımı konusunda yeni bir yaklaşım önerdiler . Önerilen yöntem kolayca diğer konuşma ve konuşma olmayan kısımları ayrımı konularındaki uygulamalara uyarlanabilir. Ayrıca düzgün konuşma üzerinde eğitilmiş yapay sinir ağları (YSA), sadece her 10 ms ölçülecekte olan bir kanal model ile çıkışında bir entropi değeri verecek şekilde kullanılmıştır (Ajmera, McCowan, & Bourlard, 2003).

Shen, Jia-lin, Jieih-weih Hung 1998 yılında konuşma tanıma alanında, entropiye dayalı bitiş noktaları doğru ve güçlü gürültülü ortamlarda tesbit etmek için bir algoritma tanımlamışlar. Bu yöntemde geleneksel enerjiye bağlı olan özellikleri kullanmak yerine, doğru bir konuşma bölütleme için spektral entropiyi geliştirmişler . Bu algoritma enerjiye bağlı algoritmalara göre gürültülü ortamda doğru algılama ve tanıma performansı ve aynı zamanda % 16 den daha az hata ortalamasıyla daha yüksek performansa sahip olmuştur (Shen, Hung, & Lee, 1998).

Wu ve diğ., 2005 yılında çalışmalarında bir özellik parametresi önerdiler. Bu çalışma bant-bölümlemeli entropi spektrumu (BSE) olarak adlandırılmaktadır. Spektrogram'ın üzerindeki bant yapısının kullanılması ile gerçekleştirilmektedir. Bu çalışmada, artırılmış uyarlamalı bant seçimi yöntemi ile BSE parametresi birleştirilerek yeni ve güçlü bir özellik parametresi olan uyarlamalı bant bölütleme spektrum entropisi (ABSE) elde edilmektedir. Bu yöntem istenmeyen ortamların sınır noktalarının belirlenmesinde başarıyla uygulanmaktadır (Wu & Wang, 2005).

Nitanda ve Haseyama tarafından 2006 yılında önerilen yöntemde MPEG Audio ile kodlanmış olan ses sinyali için bir bölütleme ve sınıflandırma yöntemi önermişlerdi. Yöntemde ilk olarak iki farklı ses sinyalin arasındaki sınırı algılıyor ve bunları ses kesitleri olarak adlandırılıyor. Daha sonra bu ses kesitleri bulanık C-Ortalama kümeleme algoritması kullanarak kümelere ayrılıyor. Önerilen yöntemle, bulanık C-Ortalama kümeleme kullanılarak, [0,1] arasındaki gerçek sayılarla ses kesitlerinin olasılık değerleri elde ediliyor. Sınırları belirtilen ses kesitleri: sessizlik, konuşma, müzik, müzik arka plan ile konuşma ve gürültü arka plan ile konuşma gibi beş ses kümesine bölünmektedir. Kümeleme sonuçları kullanılarak, gereksiz ses kesitler çıkarılabilir (Nitanda, Haseyama, & Kitajima, 2006).

Jean-Julien Aucouturier ve Mark Sandler tarafından 2001 yılında önerilen yöntemde ise, gizli Markov modeli kullanılarak, akustik müzik sinyalleri için bir bölütleme algoritması önerilmiştir. Burada sabit istatistiksel özelliklerini sahip olan müzik bölgeleri belirlenerek farklı çalışmalarla birlikte performansları ölçülmüş ve karşılaştırılmıştır (Aucouturier & Sandler, 2001).

Ying ve diğ., 1993 yılında Teager enerji algoritmasına dayalı bir enerji ölçümü geliştirdiler. Bunu uç nokta algılamada uygulamışlar. Tigran enerji algoritmasında, konuşma sinyali hem genlik ve hem frekanstan etkilenmektedir (Ying, 1993).

Ses sinyallerinin sınıflandırılması iki boyutlu bir süreçtir. Bunlar özellik seçimi ve seçilen özelliklerine göre sınıflandırma yapılmasıdır. Özellik vektörü seçtikten sonra, sorun özelliklere göre sınıflandırma yöntemleridir. Aşağıda en yaygın bölütleme yöntemleri verilmiştir. Bunlar,

- Mel Frekans Kepstrum Katsayıları
- Sıfır Geçiş Oranı
- Fark Fonksiyonların Ortalama Büyüklüğü.
- Enerji Eşik Değeri
- Entropy Değeri
- Ayrık Fourier Dönüşümü

olarak listelenebilir.

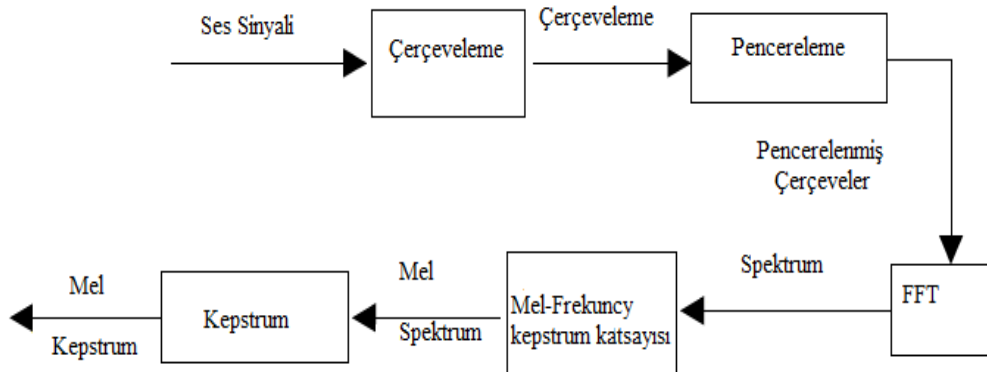
1.4. MFCC Öznitelik Çıkarma

Konuşma tanıma sistemlerinin en önemli parçalarından birisi öznitelik çıkarmadır. Ses sinyalindeki akustik özelliklere öznitelik çıkarma denir. Ses parçalarının içerisinde

taşıdığı bilgiyi temsil edecek daha az miktarda veriye dönüştürülmesi gerekir. Bu veriye “öznitelik” adı verilir. Konuşma sinyal analizinde farklı özellikler kullanılmaktadır. Ancak bu özelliklerin uygulama alanları farklı olduğundan özellik seçimi uygulamaya bağlı olarak yapılır. Örneğin, gürültülü ortamlarda bazı özellikler daha iyi cevap verirler ya da daha az hesaplama gerektirirler. Bu özellikler, genelde sinyal özellikleri pencereleme ile analiz edilirken, tüm konuşma sinyalinde de analiz edilebilir. Analiz edilen bu özellikler sayı, enerji, konuşma sinyalinin pencere ortalaması veya bir vektör olabilir. Örneğin, zaman uzayında konuşma sinyalinin boyutu konuşma penceresinin uzunluğunda bir boyutlu bir vektör olabilir. Bu nedenle, bu özellikler öznitelik değerleri veya özellik vektörü olarak adlandırılır (Gonzalez & Woods, 2002). Öznitelik çıkarma hem kaydolma hem de saptama aşamasında kullanılır ve aşağıdaki aşamalardan oluşur (Wold, 1996).

- Çerçeveleme
- Pencereleme
- Hızlı Fourier Dönüşüm (FFT)
- Mel Frekans Kepstrum Katsayıları
- Kepstrum

Öznitelik çıkarma işlemleri Şekil 5’te diyagram olarak gösterilmiştir.



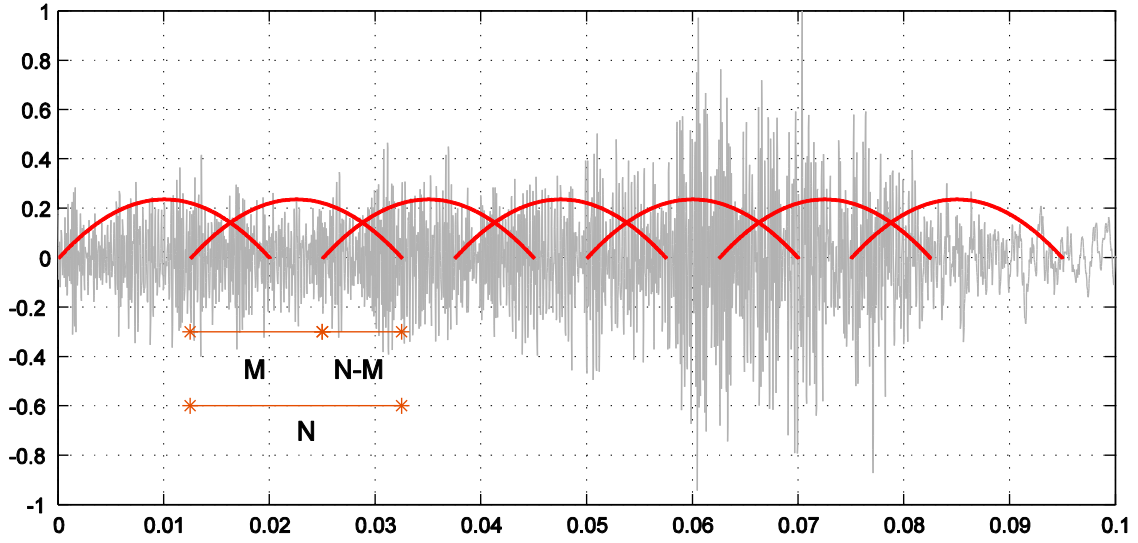
Şekil 5. Öznitelik çıkarma

1.4.1. Çerçeveleme

Mel frekans kepstrum katsayıların (MFCC) hesaplamasında, Bir ses sinyalinin çerçevelere bölmek ilk adımdır, Genelde bu iş sabit aralıklarla bir pencere fonksiyonu uygulanarak gerçekleşir. Bir ses sinyali esasında durağan olmayan bir sinyaldir. Buradaki amaç istatistiksel olarak durağan bir sinyali küçük bölümlerle modellemektir. Bu parçalar içerisinde ses sinyali durağan akustik karakteristiklere sahip olduğu kabul edilir. Genellikle 20 ve 100 milisaniye arasında seçilen bir uzunlukta çerçevelere bölünürler. Çoğu durumda en etkili zaman aralığı 20-30 ms arasındadır (Atal, 1976).

Titreşimlerin kelimelere bağlı olarak değişmesinden dolayı konuşma sinyali de sürekli olarak değişir. Konuşma sinyali, parametrelerin sabit kaldığı kabul edildiği çerçeve olarak küçük parçalara ayrılmalıdır çünkü tüm sinyal boyunca Fourier dönüşümü hesaplanırsa, farklı fonemlere ait spektral bilgilerin tutulmasında kayıplar oluşur. Tüm sinyalin Fourier dönüşümünü almak yerine çerçevenin Fourier dönüşümü hesaplanır (Eskidere & Ertaş, 2009).

Her bir çerçeveye örtüşme uygulanır. İlk çerçeve N örnekten oluşur sonraki çerçeve ilk çerçeveden M ($M < N$) örnek sonra başlar ve böylece N-M örnek örtüşür yani her çerçeve kendisinden bir önceki çerçevenin belli bir kısmını örter. Çerçevelerin örtüşme oranı, çerçeve uzunluğunun % 30'u ile % 75' i arasında alınır (Kinnunen, 2003). Örtüşme uygulanması ile çerçeve sonundaki sinyalin önemli kısımlarının kaybolmaması sağlanır. Yani her çerçeve kendisinden bir önceki çerçevenin belli bir kısmını örter. Örtme yönteminin amacı bir çerçeveden diğerine geçişin yumuşak olmasını sağlamaktır. Şekil 6'da bir konuşma işareti üzerinde çerçeveleme işlemi görülmektedir.



Şekil 6. Bir konuşmanın çerçevelere bölünmesi

1.4.2. Pencereleme

Mel frekansı kepstrum katsayılarını elde etmek için ikinci yapılan işlem pencerelemedir. Pencerelemenin amacı çerçeveleme işlemi sonucunda oluşan spektral etkilerin azaltılmasıdır ve aynı zamanda çerçeve başında ve sonundaki süreksizlikleri ortadan kaldırmasıdır. Ses sinyalinin her bir çerçevesini dikdörtgen pencereler ile çarpma işlemine pencereleme denir. Pencereleme ile çerçevelerde süreksizliğin önüne geçilir (Rabiner & Juang, 1993). Bu sayede sesin orta bölgeleri güçlendirilirken kenar bölgeleri zayıflatılır.

Pencere fonksiyonu $w(n)$ kabul edilirse, sinyalin pencerelenmiş hali aşağıdaki verilen eşitlik gibidir.

$$y(n) = x(n).w(n), \quad 0 \leq n \leq N - 1 \quad (4)$$

Burada “N” her çerçevedeki örnek sayısıdır.

Pencerelemenin bir çok çeşidi vardır. Yaygın olarak kullanılan Hamming, Hanning, Blackman, Gauss, Dikdörtgen ve Üçgen pencereleme fonksiyonlarının matematiksel ifadeleri aşağıdaki fonksiyonlarında olduğu gibidir.

Hamming pencereleme:

$$w[k + 1] = 0.54 - 0.46 \cos\left(\frac{2\pi k}{N - 1}\right), \quad k = 0, 1, \dots, N - 1 \quad (5)$$

Hanning pencereleme:

$$w[k + 1] = 0.5 \left(1 - \cos\left(\frac{2\pi k}{N - 1}\right)\right), \quad k = 0, 1, \dots, N - 1 \quad (6)$$

Blackman pencereleme:

$$w[k + 1] = 0.42 - 0.5 \cos\left(\frac{2\pi k}{N - 1}\right) + 0.8 \cos\left(\frac{4\pi k}{N - 1}\right), \quad k = 0, 1, \dots, N - 1 \quad (7)$$

Gauss pencereleme:

$$w[k + 1] = e^{-\frac{1}{2} \left(\frac{\alpha \left(k - \frac{N}{2}\right)}{\frac{N}{2}}\right)^2}, \quad 0 \leq k \leq N \quad \text{ve} \quad \alpha \geq 2 \quad (8)$$

Dikdörtgen pencereleme:

$$w[k + 1] = 1, \quad k = 0, 1, \dots, N - 1 \quad (9)$$

Üçgen pencereleme:

N tek için;

$$w[k] = \begin{cases} \frac{2k}{N + 1}, & \text{and } 1 \leq k \leq \frac{N + 1}{2} \\ \frac{2(N - k + 1)}{N + 1}, & \text{and } \frac{N + 1}{2} \leq k \leq N \end{cases} \quad (10)$$

N çift için;

$$w[k] = \begin{cases} \frac{2k - 1}{N}, & \text{and } 1 \leq k \leq \frac{N}{2} \\ \frac{2(N - k + 1)}{N}, & \text{and } \frac{N}{2} + 1 \leq k \leq N \end{cases} \quad k = 0, 1, \dots, N - 1 \quad (11)$$

1.4.3. Hızlı Fourier Dönüşümü

Bir sonraki aşamada her çerçevenin Hızlı Fourier Dönüşümü alınmaktadır. Bir konuşma sinyalini analiz etmede kullanılan en yaygın ve en çok bilgi verici yol, Fourier dönüşümü yöntemini kullanarak kısa zamanlı güç spektrumunun kestirimini yapmaktır. Bu

dönüşüm Ayrık Fourier Dönüşümünün hızlı bir şeklidir ve tanım kümesini zaman uzayından frekans uzayına geçirir (Cooley & Tukey, 1965).

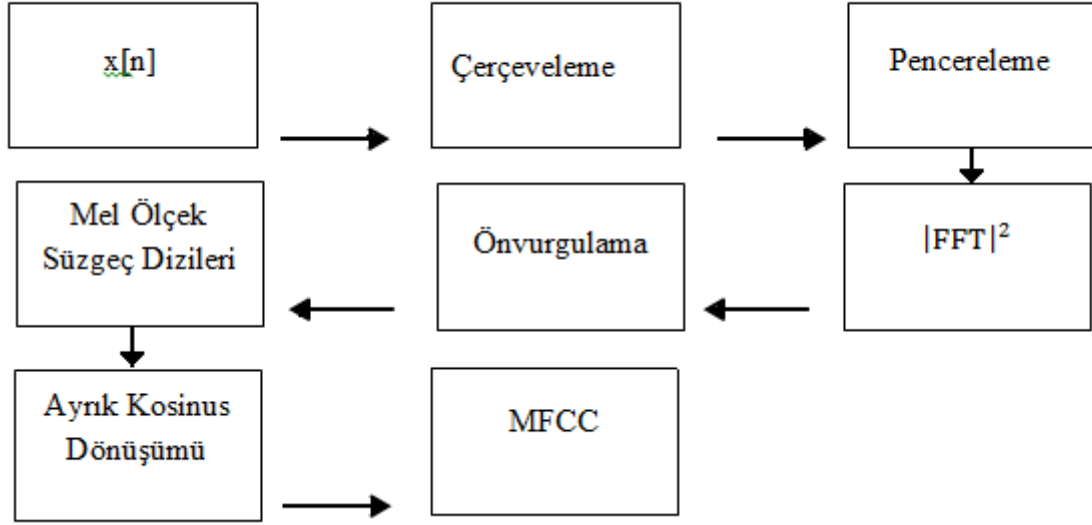
1.4.4. Mel Frekans Kepstrum Katsayıları

Mel frekans kepstrum katsayıları otomatik konuşma tanıma (ASR- Automatic Speech Recognition) sistemlerinde kullanılan en yaygın özelliklerinden biridir bunun nedeni, MFCC insan kulağının frekans seçiciliğini taklit ederek iyi bir şekilde konuşmacıları ayırt edici değerler elde etmesidir (Eskidere & Ertaş, 2009). Bu katsayıların ile en temel dezavantajı, gürültüye karşı yüksek hassasiyetidir. Öznitelik çıkarma yöntemlerde, özellikle sinyal spektrumu elde etmek için, Fourier analizi kullanan yöntemlerde, enerji spektrumu hesaplamak için genelde filtre bankası kullanılmaktadır. Filtrelerinin sayısı sistemlere göre değişiklik gösterir. Ama genelde 16-20 arasında filtre kullanılır. Filtrelerin artırılması genelde sınıflandırmanın kalitesinin artmasına sebep olur. Fakat filtrelerin bant genişliğini azaltıldığında sınıflandırma kalitesi düşer. Filtrelerin arasındaki mesafeyi doğrusal olmayan şekilde ayarlaması mel ölçeğiyle yapılır.

Mel-Frequency kepstrum katsayıları, algı temelli sesi temsil eden katsayılardır. Fourier Dönüşümü veya Ayrık Kosinüs dönüşümünden türetilir. Mel frekans kepstrum katsayıları, ilk olarak Davis ve Mermelstein tarafından tanımlanmıştır (Davis & Mermelstein, 1980). MFC katsayıları konuşma tanıma için ilk olarak Young ve diğ. (Young, Woodland, & Byrne, 1993) tarafından kullanılmıştır. Onların başarısı konuşmanın genlik spektrumunun iyi bir biçimde temsil edebilmesinden kaynaklanmıştır. Bir frekans mel-frekansına çevirmek için aşağıdaki eşitlik kullanılır,

$$Mel(f) = 2595 \log (1 + f/700) \quad (12)$$

Mel ölçeği, 1 kHz e kadar doğrusal, ve daha yüksek değerlerde ise logaritmiktir. MFCC vektörü çıkartımı blok diyagramı Şekil 7'de görülmektedir:



Şekil 7. MFCC özniteliklerinin çıkarılma işleminin blok diyagramı

1.4.5. Kepstrum

Öznitelik çıkarmanın son aşaması olarak, her çerçeve ters Fourier dönüşümüne sokulur ve frekans uzayından tekrar zaman uzayına döndürülür. Bu işlemin sonucu olarak Mel-Frekansı Kepstral Katsayıları elde edilir. MFCC aşağıdaki gibi hesaplanır:

$$MFCC_i = \sum_{k=1}^K X_k \cos\left(i(k - 0.5) \left(\frac{\pi}{K}\right)\right) \quad i = 1, 2, \dots, M \quad (13)$$

Bu eşitlikte M, kepstrum katsayılarının sayısını, X_k , $k = 1, 2, \dots, K$, ise k. filtrenin log-enerji çıktısını göstermektedir.

Mel Frekans Kepstrum Katsayılar (MFCC)'lerin çıkarımında şu temel adımlar atılır: Her bir ses çerçevesine Ayrık Fourier dönüşümü uygulanarak frekans spektrumu bulunur. Ayrık Fourier Dönüşümü katsayıları Mel filtre bankasının genlik frekans cevabına göre ağırlıklandırılır. Logaritmik Enerjileri Hesaplanır ve Ayrık Kosinüs Dönüşümü uygulanır. Sonuçta ulaşılan Ayrık Kosinüs Dönüşümün katsayılarının genlikleri MFCC değerlerini vermektedir. Öznitelik olarak alınan MFCC sayısı, spektrumun gösterimindeki hassasiyeti belirler. MFCC sayısı arttıkça hassasiyet artar (Gonzalez & Woods, 2002).

1.5. Sıfır Geçiş Oranı

İlk olarak, ZCR ile sinyalin temel frekansı arasında basit bir ilişki kurulabilir. Sinyal tek bir bileşenden oluşuyorsa, bir periyot boyunca iki kez işaret değiştirecektir. Böylece ZCR, sinyalin temel frekansının iki katına eşit olur ancak, sinyal izgesel olarak zengin ise bir periyottaki ZCR değeri yükselecektir. ZCR, ölçümü çok kolay olduğu için yaygın olarak kullanılan bir özneliktir. Literatürde sıfır geçiş sayısından yararlanılarak hesaplanabilen çeşitli öznelikler tanımlanmıştır (Bolat, Küçük, & Yıldırım, 2004).

Sıfır geçiş elektronik, matematik, ses ve görüntü işleme sık kullanılan bir terimdir. herhangi bir elektriksel sinyalin sıfır noktasından geçtiği anın tespit edilmesidir. Bir ses sinyalinde genliğin işaretinin değişme veya örneklerin negatiften pozitif ya da tam tersi geçişlerini sıfır geçiş noktası denir. Bu geçişlerin fazla olması ortamda bir konuşmanın olduğuna kanıt olabilmektedir. Bir konuşmada, ses şiddetinin ani artması ve düşmesi ile tespit edildikten sonra tam olarak nerede başladı nerede bitti gibi bir kontrolün yapılabilmesi için bu geçişlerin sayısına bakılır. Bu yöntem yüksek gürültülü olan sinyallerde zayıf sonuçlar elde edilebilir.

ZCR hesaplamak için aşağıdaki eşitlik kullanılır.

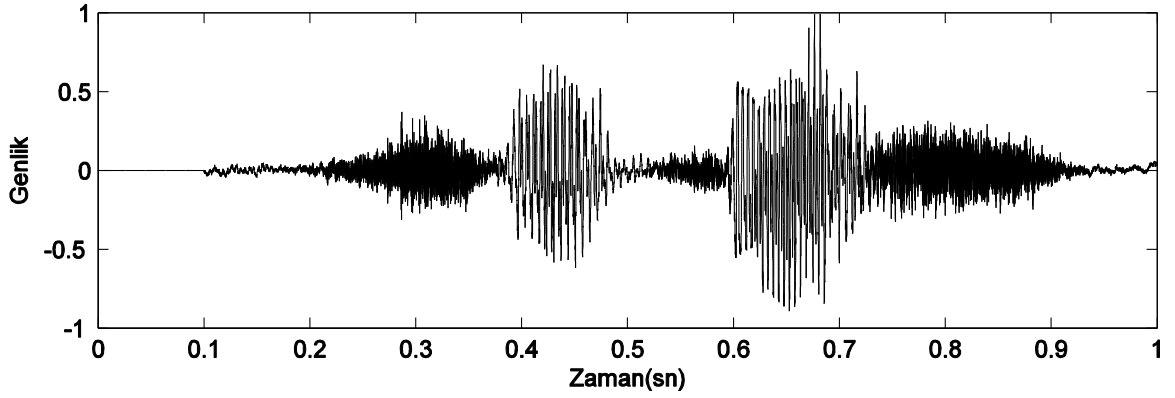
$$Z_t = \frac{1}{2} \sum_{n=1}^N |\text{sign}(x[n]) - \text{sign}(x[n-1])| \quad (14)$$

Burada

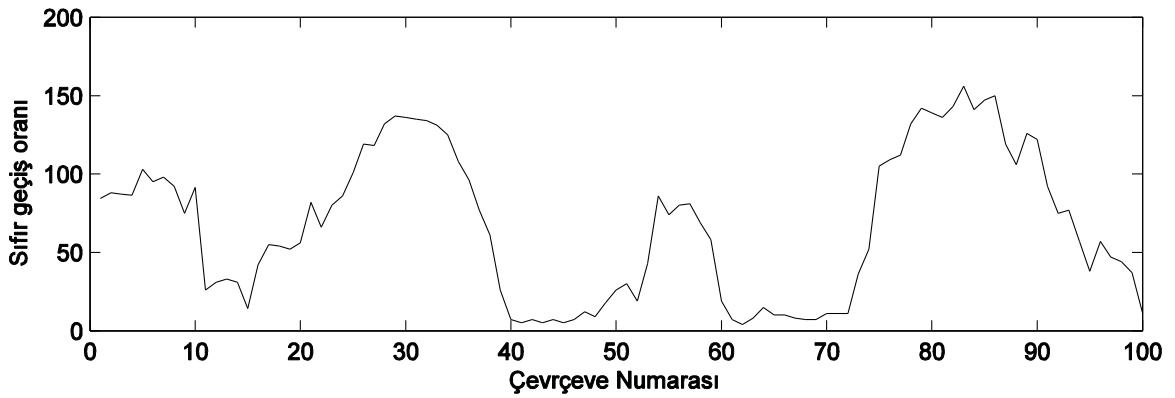
$$\text{sign}(x) = \begin{cases} -1, & x < 0 \\ 0, & x = 0 \\ 1, & x > 0 \end{cases}$$

biçiminde gösterilmektedir.

Sıfır geçiş oranı bir sinyal ne kadar gürültülü bir sinyal olduğunu ölçüyor. Bu oran önemli bir konuşma sinyali özelliğidir. Gürültülü sinyalde veya sessiz seslerde örneğin: (/t/, /s/, /p/) harflerde yüksek ZCR değere sahip ve gözlenebilir temel frekans sesli sesler örneğin (/a/, /e/, /o/) düşük ZCR değere sahiptir. Şekil 8(a), bir örnek sinyali ve Şekil 8(b)'de sinyalin ZCR grafiği gösterilmektedir.



(a)



(b)

Şekil 8. (a) Örnek bir konuşma sesi sinyali; (b) (a)'daki sesin sıfır geçiş oranının grafiği

1.6. Fark Fonksiyonunun Ortalama Büyüklüğü

Fark Fonksiyonlarının Ortalama Büyüklüğü, sesli konuşma seslerde perde periyodunu tahmin etmek için kullanılır. AMDF özilişki analizinde kullanılan bir tekniktir. Giriş konuşmanın değişik gecikmeleri arasındaki ilişkiyi araştıran bir yöntemdir. Gecikmiş bir konuşma sinyali ile orijinal sinyal arasındaki ilişkinin derecesini bulur. Her bir gecikme de, farkın mutlak büyüklüğü alınarak toplanır (Ross, Shaffer, Cohen, Freudberg, & Manley, 1974).

AMDF nin bazı avantajları:

- Perde periyodun için iyi ve basit bir tahmin ölçüsü.
- Hiçbir çarpma işlemi içermez.
- Dinamik aralık özellikleri, bir 16-bitlik makine uygulanması için uygundur.
- İşlemin doğası gereği, programlanabilir bir işlemci üzerinde ya da özel amaçlı donanım uygulama için uygun yapıdadır.

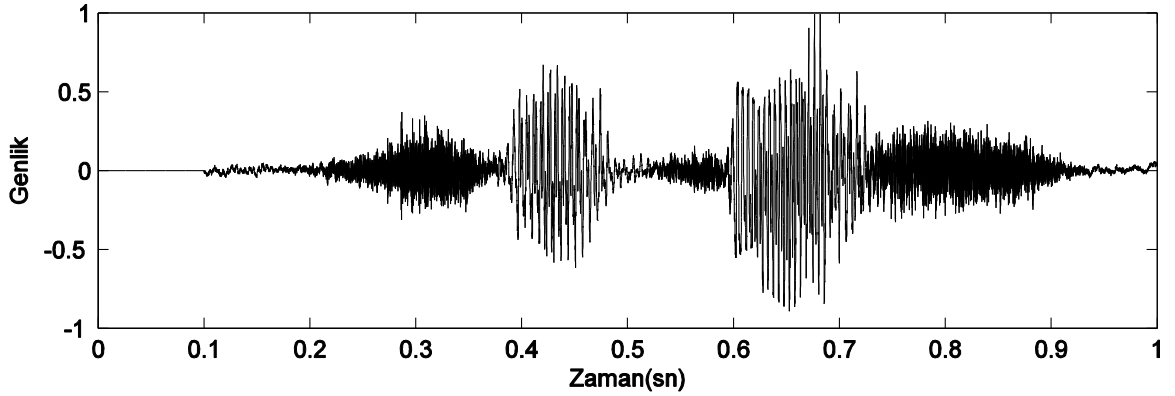
Kısa-sürelili AMDF fonksiyonu aşağıdaki eşitlikteki gibi tanımlanır.

$$D_{AMDF}(m) = \frac{1}{N} \sum_{n=0}^{N-1} |x(n) - x(n+m)| \quad (15)$$

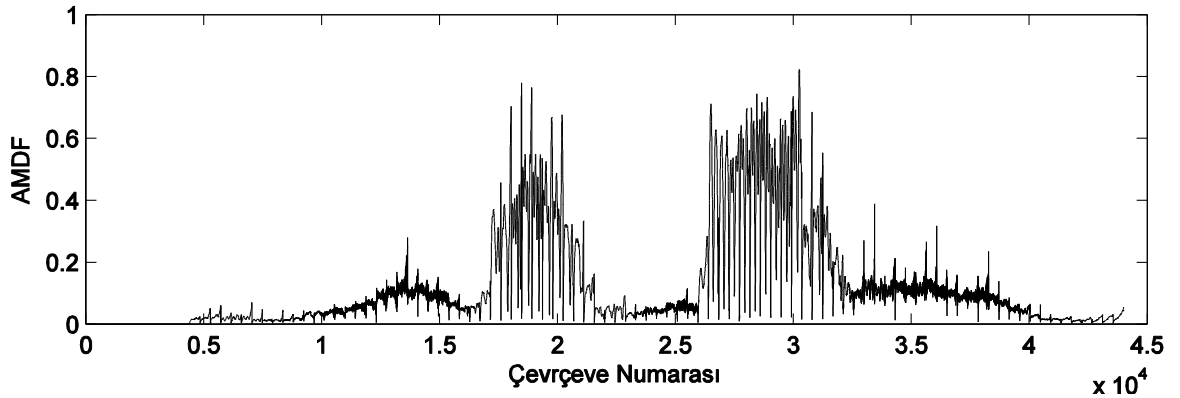
burada $x(n)$ bir konuşma sinyalinin örnekleridir. Periyodik bir sinyal için T_0 periyot ile gecikme m indeksi T_0 ya eşit olduğunda bu fonksiyon en düşük değerde olması beklenir. Şekil 9(a), bir örnek sinyali ve Şekil 9(b)'de sinyalin AMDF grafiği gösterilmektedir. Perde süresi, genel olarak aşağıdaki gibi tahmin edilmektedir.

$$T_0 = \min(D_{AMDF}(m)), \quad m = m_{\min}, \dots, m_{\max}$$

burada m_{\min} ve m_{\max} değerleri perde-aralığını kapsayacak şekilde seçilen deneme aralıklarıdır (Abdullah-Al-Mamun, Sarker, & Muhammad, 2009).



(a)



(b)

Şekil 9. (a) Örnek bir konuşma sesi sinyali; (b) (a)'daki sesin fark fonksiyonunun ortalama büyüklük grafiği

1.7. Enerji Eşik Değeri Yöntemi

Enerji eşik değeri, ses sinyallerinin zamansal ortamda göstermiş oldukları yüksek genlikli etkinliklerine dayalı bir yöntemdir (Sakhnov, Verteletskaya, & Simak, 2009). Bu yöntemin gerçekleşmesinde öncelikli ses sinyalinin kısa zamanlı çerçevelere bölünmesidir. Böylelikle her bölge bir enerji seviyesine sahip olmaktadır.

$$f_j = \{x_i\}_{i=(j-1)N+1}^{jN} \quad (16)$$

1.7.1. Çerçeve Enerjisi

$$E_j = \frac{1}{N} \sum_{i=(j-1)N+1}^{jN} x_i^2 \quad (17)$$

burada, E_j , j . çerçevenin enerjisidir.

1.7.2. Eşikleme Değeri

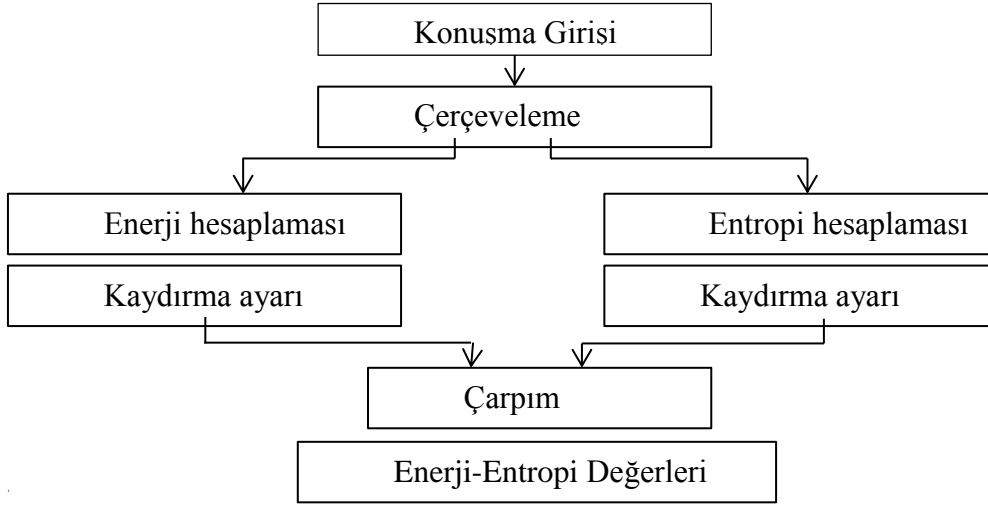
Enerji değerlerinin ayıklanmasında eşik değeri önemli bir yer tutmaktadır. Bu eşik değeri sinyal içerisindeki ses aktivitesi ile gürültüyü bir birinden ayıran bir ölçüttür. Bu ölçütün belirlenmesinde yaygın bir yöntem olan Otsu algoritması tercih edilmiştir (Otsu, 1979). Eşikleme sonucunda ses sinyalinin enerjisi ikil enerji sinyaline dönüşmektedir. Bu dönüşüm sayesinde gürültü bölgeleri sıfıra, ses bölgeleri 1 indirgenmiş olmaktadır. Böylelikle ayırım tamamlanmış olmaktadır.

1.8. Entropik Bölütlemesi

Entropi belirli bir dağılımda belirsizlik ya da bozukluğun bir ölçüsüdür. Shen Entropi teorisi aşağıdaki denkleme dayanmaktadır,

$$H(X) = - \sum_{i=1}^N p(x_i) \log_2 p(x_i). \quad (18)$$

Burada x_i , i zaman da ses vektörü temsil etmektedir. Eğer x_i dizisi düzgün dağılımdan rastgele değişkenler olursa, oluşan maksimum belirsizliğe karşılık gelen bir durumda çıkış entropi sonucu maksimum oldur. Aksine, eğer eğride sadece bir keskin tepe görünür ise o zaman entropi minimumdur. Enerji kullanım nedeni, sınır algılamada, konuşmada sesli bölümler enerji veya güçleri sessiz bölümlere göre daha fazladır. Bu yöntemde, ilk olarak sabit uzunlukta bir pencere tanımlanır. Bu pencereyi, konuşma sinyalinde hareket ederek ne zaman pencere enerjisi enerji eşik noktasını aştığında, başlangıç noktası olarak belirlenir. Benzer şekilde, bitiş noktası, eşik değeri bölge enerjisinden daha az olduğu nokta belirlenir. Enerjiye dayalı sınır algılama algoritmalarının başarımı gürültülü ortamlarda yetersiz kalmaktadır. Entropi en yaygın kullanım alanı bilgi teorisi ve metin kodlamadır. Shen entropisi ilk olarak, sınırları algılamak için kullanmıştır. Shen sonuçlarına göre, sesli entropi spektrumu, sessiz entropi spektrumdan çok farklıdır. Bu özellik, konuşma ve sessizlik sınırını algılamak için kullanılabilir. Entropiye dayalı yöntem özellikle durağan olmayan ve gürültü olduğu durumlarda, enerjiye dayalı yöntemden daha çok güvenlidir. Enerji ve entropi her biri bazı gürültülü ortamlarda kendi kısıtlamaları vardır. Bu iki yöntemin kombinasyonu her iki yöneme göre daha sağlıklı sonuçlar vermektedir. Bu yöntemde her pencere için hem enerji ve hem de entropi hesaplanır. Enerji ve entropi değerleri birbiriyle çarpılarak tek bir veri kümesi ile değerlendirme yapılır. Bu iki yöntemi başarımı sessizlik bölgesinin konuşmadan ayrılması, özellikle gürültülü ortamlarda oldukça iyi sonuçlar vermektedir (Abdullah-Al-Mamun, Sarker, & Muhammad, 2009). Şekil 10 enerji entropi diyagramını göstermektedir.



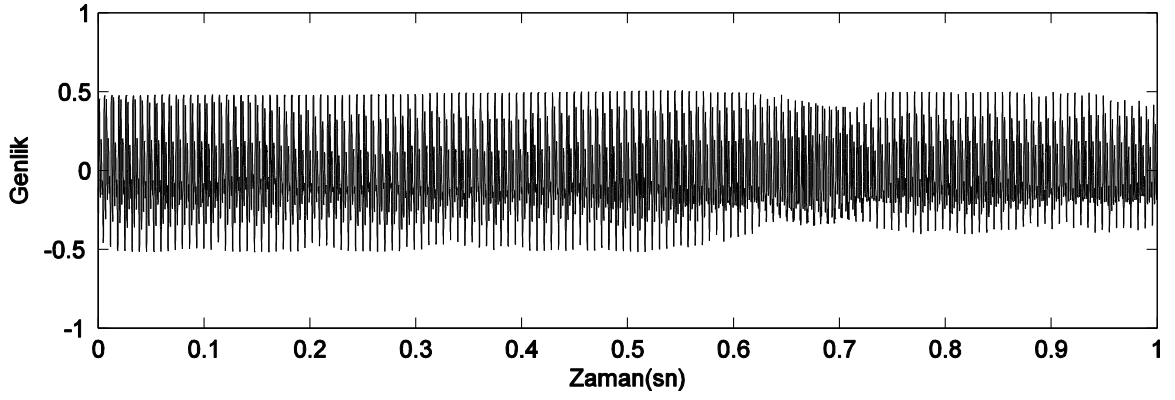
Şekil 10. Enerji Entropi diyagramı

1.9. Ayrık Fourier Dönüşümü

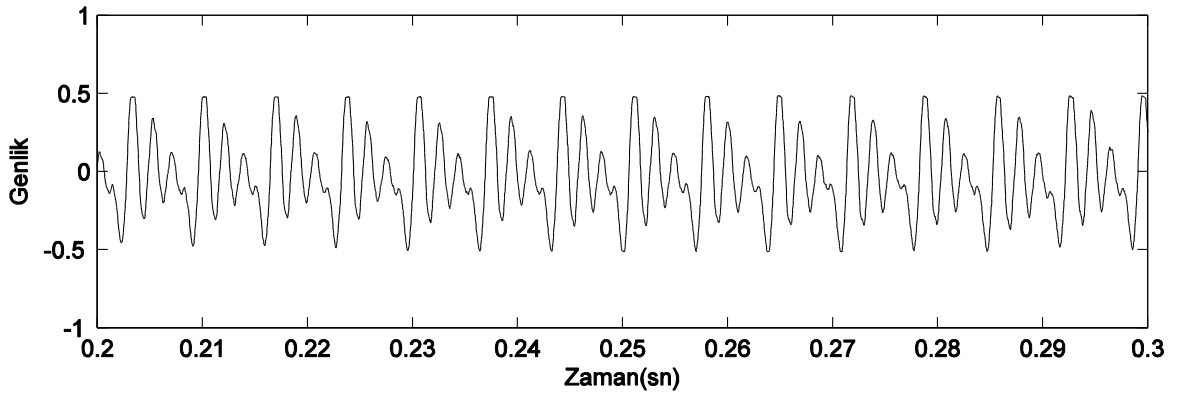
Fourier dönüşümü, ses işlemede en yaygın kullanılan dönüşümlerden birisidir. Fourier dönüşümü her frekanstaki genlik ve evreyi gösteren bir fonksiyondur. Dönüşüm iki matematiksel yöntemle hesaplanır, birincisi fonksiyon sürekli ise ikincisi ise fonksiyon ayrıkça uygulanır. Eğer fonksiyon ayrıkça, yani fonksiyon ayrık zaman aralıklarına ait değerlerden oluşuyorsa, ayrık frekanslardaki sinüzoidal fonksiyonların serisi şeklinde gösterilebilir. Bu frekanslar temel frekansın iki, üç ve daha büyük katlarıdır. Ses sinyalleri bilgisayar ortamında sayısal olarak tutulup işlemlerde kullanıldığından Ayrık Fourier Dönüşümü (DFT) daha uygun bir dönüşüm olmaktadır.

Her harfin temel ayırt edici özelliği, o harfin oluşturan frekanslardır. Örneğin, /a/ harfi frekansları üç ana frekansı 750, 1150 ve 2400 Hz dir ve aynı frekanslar /o/ harfi için 400 , 1150, ve 2300 Hz dir. Bu nedenle, /a/ ve /o/ harflerinin ayrılmasını sağlayan şey harfleri oluşturan temel frekans ve harmonilerin genlikleridir. Dolayısıyla konuşma tanıma uygulamalarında, konuşma sinyalinin temel frekanslarını bulmak çok önemlidir.

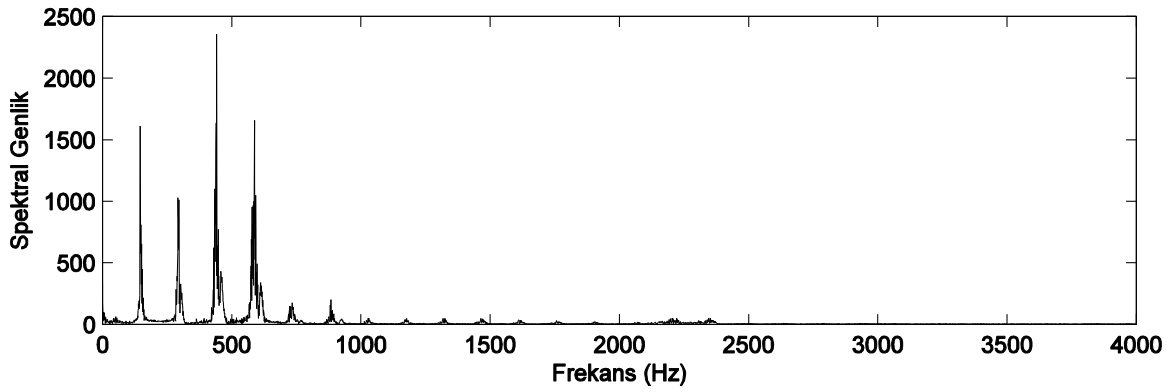
Genelde konuşma sinyalleri zaman ortamında sinyallerdir ve bir sinyalin zaman ortamında temel frekansı tespit etmek oldukça zordur. Örneğin aşağıdaki Şekil 11(a)'da /a/ harfinin konuşma sinyalinin zaman ortamında gösterimi verilmiştir. Sinyalin bir parçası alınıp yakınlaştırılmış halini Şekil 11 (b) gösterilmektedir. Aynı sinyalin Fourier dönüşümü alındıktan sonra elde edilen genlik spektrumu Şekil 11(c) verilmiştir.



(a)



(b)



(c)

Şekil 11. /a/ harfi konuşma sinyali; (a) 1 saniyelik /a/ harfi konuşma sinyali; (b) 0.2sn-0.3sn aralıktaki /a/ harfi konuşma sinyali; (c) 1 saniyelik /a/ harfi konuşma sinyalinin ayrık Fourier dönüşümü

Şekil 11'de zaman ortamındaki sinyalde sinyali oluşturan frekansları bulunması oldukça zordur. Ancak frekans bileşenlerini bulma amacına ulaşmak için başka bir araca gerek duyulur. Joseph Fourier gösterdi ki herhangi bir periyodik fonksiyonu bir sinüs (veya kosinüs) dalgaların kombinasyondan oluşturulabilmektedir. Bu ise Fourier serisi olarak

bilinmektedir. Her sinüs dalgası, ses şeklinde hoparlörlerin çıkışından duyulabilir. Sinüs dalgasının frekansı, üretilen ses modülasyonu belirler.

Fourier analizi kullanılarak, bir sinyali oluşturan sinüs dalgalarını ve sinüs dalgası ile konuşma sinyalin oluşturan frekansların genlikleri elde edilebilir. Ancak Fourier serileri periyodik fonksiyonları üzerinde tanımlanır bu nedenle direkt olarak Fourier serilerinden bu amaç için yararlanamayız. Ama sinyaller üzerinde çalıştığımızda, konuşma sinyal için hiçbir fonksiyon tahmin edilemez. Bu sorunu çözmek için, Fourier dönüşümü adında bir araç sayısal verilere (sinyal) üzerinde uygulamak için tanıtıldı (Signal Fourier Concept). Karmaşık bileşenli ayrık Fourier dönüşümü aşağıdaki denklem ile tanımlanabilir,

$$X(m) = \sum_{n=0}^{N-1} x(n)e^{-j2\pi\frac{nm}{N}}, \quad (m = 0,1,2, \dots, N - 1). \quad (19)$$

Burada N , giriş sinyalinin uzunluğu; $x(n)$, giriş sinyalini; $X(m)$ karmaşık bir sayı dizisi olan Fourier dönüşümünü vermektedir. Bu yüzden m inci frekans büyüklüğünü ölçmek için, bu karmaşık sayı dizinin genliğini ölçmek yeterlidir. Genlik spektrum değeri (20) eşitliği yardımıyla elde edilir.

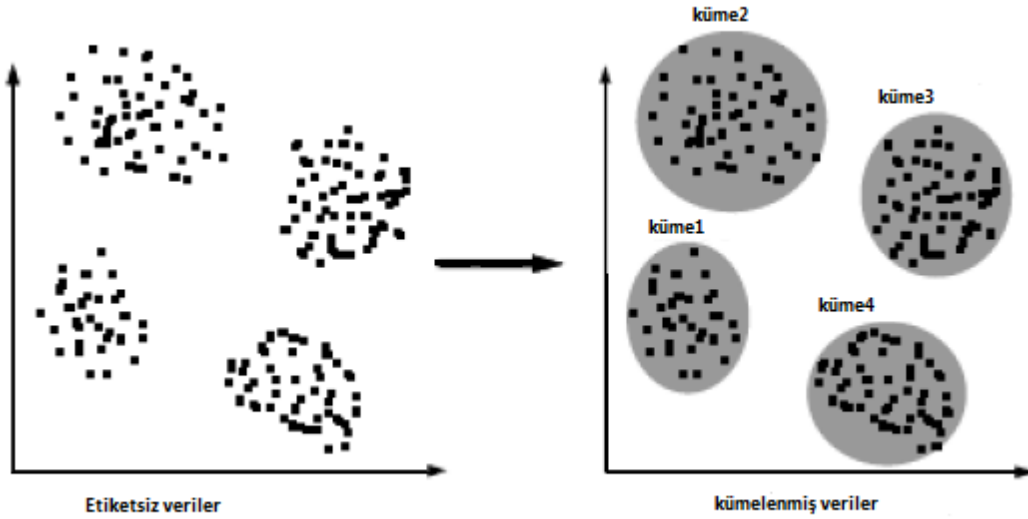
$$A(m) = \sqrt{\left(X_{real}^2(m) + X_{imag}^2(m)\right)} \quad (20)$$

burada $A(m)$ genlik spektrum değerlerini, X_{real}^2 sinyalin Fourier dönüşümünün gerçel bileşenini, X_{imag}^2 sinyalin Fourier dönüşümünün sanal bileşenini vermektedir.

1.10. Kümeleme Analizi

Kümeleme analizi veri madenciliğinde çok önemli bir araştırma konusudur ve pek çok alanda pratik uygulamalar bulunmaktadır. Kümeleme analizi, son zamanlarda iş ve bilim alanında sıkça kullanılmaya başlanan çok değişkenli veri analiz yöntemlerinden biridir. Bu yöntemin temel amacı, veri setindeki veri noktalarını benzerliklerine göre alt gruplara ayırmaktır. Başka bir ifade ile veri noktalarının aynı grupta veya kümede toplanmasını sağlamaktır (Güler, 2006). Çeşitli nesnelere oluşan topluluğa klasik

mantıkta küme denir. Küme teorisinin temeli bir elemanın kümeye ait olması veya olmamasına dayanır. Kümedeki elemanlar birbirlerine benzer gruplara ayrılır ve bu nedenle kümeler birbirine benzer nesnelere topluluğu diye tanımlanır. Bir kümenin geometrik çizimini yaparken nesnelere küme içerisinde birbirine çok yakın, kümeler ise birbirinden uzak olacaktır. Verilerin gruplandırılması sırasında yapılacak ilk iş kullanılacak benzerlik ölçütünün belirlenmesidir. Bu benzerlik ölçütü aynı küme içerisindeki verilerin maksimum oranda benzerlik göstermesini sağlarken diğer gruptaki verilerle maksimum oranda farklılık göstermesini sağlamalıdır. En yaygın olarak kullanılan benzerlik ölçütü mesafeye dayalı benzerliktir (Ortakçı & Göloğlu, 2012). Şekil 12'da birçok sayıda elemandan oluşan bir veri seti mesafeye dayalı benzerlik ölçütüne göre dört farklı kümeye ayrılmıştır.



Şekil 13. Veri kümeleri (Ortakçı & Göloğlu, 2012)

Kümeleme algoritmasında genellikle önceden giriş örneklere ait bir temsilci dizisi tanımlanır. Sonra örnekler ve bu temsilci dizinin arasında benzerlik oranıyla, örnek hangi kümeye ait olduğu belirlenir. Bu aşamadan sonra, her küme için yeni bir temsilci dizi hesaplanır ve tekrar bu örnekler yeni temsilciler dizisiyle karşılaştırılır ve bu iş kümelerin temsilcileri sabit olana kadar tekrarlanır.

1.10.1. K- Ortalama Kümeleme Algoritması

En eski kümeleme algoritmalarından olan K-Ortalama algoritması (KO), 1967 yılında MacQueen (MacQueen, 1967) tarafından geliştirilmiştir. En yaygın kullanılan gözetimsiz öğrenme yöntemlerinden birisidir. K-Ortalama algoritmasının atama mekanizması, her verinin sadece bir kümeye ait olabilmesine izin verir. Bu nedenle, keskin bir kümeleme algoritmasıdır. Merkez noktanın kümeyi temsil etmesi ana fikrine dayanır (Işık & Çamurcu, 2007; Han & Kamber, 2001).

KO veri setini istenen sayıda alt kümeye ayırabilen bir yöntemdir. K-Ortalama yönteminin kullandığı algoritmanın temel amacı c adet küme ve bu kümelere ait k adet merkez belirlemektir.

- Bunu gerçekleştirmek için ilk önce n adet veri seti içerisinde rastgele seçilmiş k adet veriyi merkez olarak belirler. $x = \{x_1, x_2, \dots, x_n\}$, $v = \{v_1, v_2, \dots, v_c\}$, ($c \leq n$) sonra kalan diğer veriler $n-c$ Küme merkezi olmayan birimler her bir verinin c adet merkeze olan öklid uzaklığını hesaplayıp en yakın olan merkeze ait kümeye atanır.
- Daha sonra k adet başlangıç kümeyle ait elemanların ortalamaları alınarak, yeni küme merkezleri oluşturulur böylece küme merkezleri güncellenir ve tekrar birim-merkez uzaklıkları incelenir ve elemanlar en yakın yeni küme merkezlere atanırlar.
- Bu işlemler, her bir küme merkezinde ve kümeye ait elemanlar değişiklik meydana gelmeyinceye kadar devam eder ve aynı zamanda, küme içi varyans minimum ve kümeler arası varyans maksimuma ulaşıncaya kadar tüm birimler c kümeye atanmaya ve merkez ötelenmeye devam eder (Kılıç, 2013; Atalay & Tortum, 2010).

K-ortalama kümeleme yönteminin uygulanmasında yaygın olarak karesel hata kriteri SSE kullanılır. En düşük SSE değerine sahip kümeleme sonucu en iyi sonucu verir. Elemanların buldukları kümenin merkez noktalarına olan uzaklıklarının karelerinin toplamı (21) nolu eşitlik ile hesaplanmaktadır (Pang-Ning, Steinbach, & Kumar, 2006).

$$SSE = \sum_{j=1}^c \sum_{i=1}^n dist^2(v_j, x_i) \quad , \quad J = \sum_{j=1}^c \sum_{i=1}^n \|x_i^j - v_j\|^2 \quad (21)$$

Burada n veri sayısını , c merkez sayısını. , $\|x_i^j - v_j\|^2$ formülü ise x_i^j verisinin v_j küme merkezine yani i . verinin j . küme merkezine olan Euclidean uzaklığını belirtir. v_i , j -inci küme merkezi ve aynı zamanda v_j birimlerin ortalamasıdır.

K- Ortalama Kümeleme Algoritma Avantajları:

- Uygulanabilirliği kolaydır.
- Büyük veri kümelerinde hızlı çalışabilir.
- Veri sayısı çok fazla olan hesaplamalarda, k-ortalama, küme sayısı küçük ise hesaplamalar, hiyerarşik kümelemeden daha hızlı yapar.

K- Ortalama Kümeleme Algoritma Dezavantajları:

- K-ortalama algoritması k küme sayısını tespit edememektedir. Bu nedenle uygun k sayısını bulana kadar bir deneme yanılma süreci gerçekleşmektedir.
- Gürültülü verilere duyarlıdır. Bu veriler de kümelere dahil edilir.

1.10.2. Bulanık C-Ortalama Algoritması

Bulanık c-ortalama (FCM) algoritması, bulanık kümeleme tekniklerinden en yaygın kullanılan yöntemdir. Bulanık c-ortalama algoritması 1973 yılında Dunn tarafından ortaya atılmış ve 1981' de Bezdek tarafından geliştirilmiştir. Bulanık c-ortalama yöntemi, nesnelerin iki veya daha fazla kümeye ait olabilmesine izin verir. Bulanık mantık prensibi gereği her veri, $[0,1]$ arasında değişen birer üyelik değeri ile kümelerin her birine aittir. Yani FCM algoritmasının sonuçları 0-1 arası tüm değerleri alabilir. FCM, kesin aitlik belirtmek yerine aitlik derecesi belirtir. Dolayısıyla bulanık yöntemler klasik yöntemlere göre çıkış verisinde daha çok bilgi taşımaktadır (Bezdek, Ehrlich, & Full, 1984). FCM algoritmasında, bir verinin tüm sınıflara olan üyelik değerleri toplamı "1" olmalıdır. Nesne hangi küme merkezine yakın ise o kümeye ait olma üyeliği diğer kümelere ait olma üyeliğinden daha büyük olacaktır (Işık & Çamurcu, 2007). Kümeleme yöntemi için kullanılan amaç fonksiyonu şu şekildedir:

$$J_m(U, V) = \sum_{i=1}^n \sum_{j=1}^c u_{ij}^m \|X_i - V_j\|^2, \quad 1 < m < \infty. \quad (22)$$

Bu fonksiyon ağırlıklandırılmış en küçük kareler fonksiyonudur. Burada, n parametresi gözlem sayısını, c ise küme sayısını gösterir. u_{ij} ise c . kümedeki, x_i nin üyeliği, $J_m(U, V)$ değeri ise tüm ağırlıklandırılmış hata karelerinin toplamının bir ölçüsüdür (Alpaslan, Erilli, Yolcu, Eğrioğlu, & Aladağ, 2011). $X_k = \{x_1, x_2, \dots, x_n\}$ ve $V_i = \{V_1, V_2, \dots, V_c\}$ sırayla gözlem değerleri ve V_i küme merkezlerini simgelemektedir. $U = (u_{ki})_{n \times c}$ gözlemlerin üyelik değerlerin matrisini göstermektedir ve 0 ile 1 arasında herhangi bir değer olabilir. Bu tekniğin uygulanabilmesi için küme sayısının ve bireylerin üyelik derecelerinin önceden bilinmesi gerekmektedir. Bu algorithmada her $i = 1, 2, \dots, n$ için $\sum_{j=1}^c u_{ij} = 1$, $u_{ij} \geq 0$ ve $m > 1$ genellikle bu parametreye bulanıklaştırma parametresi denir. (23) nolu eşitliği minimize ederek u_{ij} üyelik matrisi aşağıdaki gibi hesaplanır:

$$u_{ij} = \left(\left(\sum_{k=1}^c \frac{\|X_i - V_j\|^2}{\|X_i - V_k\|^2} \right)^{\frac{2}{m-1}} \right)^{-1} \quad (23)$$

FCM algoritması uygulandıktan sonra hangi bireyin hangi kümeye gireceğine karar vermek için üyelik dereceleri kullanılır. Her bir bireyin hangi kümeye olan üyeliğinin en büyük olduğuna bakılır ve bu bireyler o kümeye dahil edilir. Ancak her bir birey diğer kümelere de belli bir üyelik dereceleri ile girebilir. Eğer $J_m(U, V)$ fonksiyonu c nin her değeri için minimize edilecek olursa, diğer bir deyişle V_i 'lere göre 1. dereceden türevi alınıp 0'a eşitlenirse FCM algoritması bu şekilde olacaktır ve küme merkezlerini bu (24) eşitliğiyle elde edilir. Her $i = 1, 2, \dots, n$ ve $j = 1, 2, \dots, c$ için

$$v_j = \frac{\sum_{i=1}^n u_{ij}^m x_i}{\sum_{i=1}^n u_{ij}^m} \quad (24)$$

u_{ij} üyelik matrisi rastgele atanarak algoritma başlatılır. İkinci adımda ise merkez vektörleri hesaplanır. Merkezler (24) deki formüle göre hesaplanır. Hesaplanan küme

merkezlerine göre U_{ij} matrisi (23) formülü kullanılarak yeniden hesaplanır. Eski U_{ij} matrisi ile yeni U_{ij} matrisi karşılaştırılır ve (25) eşitliğine fark ϵ 'dan küçük olana kadar işlemler devam eder (Yıldız, Çamurcu, & Doğan, 2010). Bulanık c-ortalama algoritmasının en önemli özelliği olan üyelik matrisinin kümeleme üzerinde olumlu etkileri olmasıdır. Bu matris belirsiz durumların tanımlanmasını kolaylaştırır (Azem, 2003). Ayrıca üyelik dereceleri düşük olduğundan sıra dışı verilerin etkisi azdır. Esnek bir yapıya sahiptir. Örtüşen kümeleri bulma kabiliyeti diğer bölünmeli algoritmalara göre daha fazladır.

$$\max_{i,j} \left\{ \left\| u_{ij}^{(t+1)} - u_{ij}^{(t)} \right\| \right\} < \epsilon \quad (25)$$

1.10.3. Bulanık Kümeleme Algoritmasında Optimal Küme Sayısı

Bulanık kümeleme analizinde en iyi küme sayısının belirlenmesi önemli bir problemdir. Literatürde en iyi küme sayısının belirlenmesi için çeşitli küme geçerlilik kriterleri önerilmiştir. Bununla birlikte birçok kümeleme algoritmasında küme sayısının önceden bilinmesi gerekir. Birçok çalışmada, araştırmacının küme sayısı hakkında ön bilgisinin olmaması, bulunan küme sayısının gerçek küme sayısından az ya da çok olup olmadığının bilinmemesine yol açmaktadır. Eğer bulunan küme sayısı gerçek küme sayısından az çıkarsa, mevcut kümelere bir veya birkaçı birleşmeli, çok çıkarsa mevcut kümeler birbirlerinden ayrılmalıdır. Optimal küme sayısının belirlenmesi işlemlerine genel olarak Küme Geçerliliği (Cluster Validity) adı verilmektedir. Veriler iki boyutlu uzayda olduğunda küme sonuçları görsel olarak yorumlanarak küme sayısına karar verilebilir. Ancak uzaydaki boyut sayısı arttıkça görsellik zorlaşmakta ve geçerlilik kriterlerine ihtiyaç duyulmaktadır. Sonuç olarak, en uygun küme sayısının belirlenmesi için iki kriterden bahsedilebilir.

- Yoğunluk: Küme elemanlarının birbirlerine yakınlıklarını ölçer.
- Ayrılma: İki küme arasındaki mesafeyi ölçerek, kümelerin birbirlerinden ne kadar ayrıldıklarını gösterir (Alpaslan, Erilli, Yolcu, Eğrioğlu, & Aladağ, 2011).

Küme geçerliliği problemi, kümeleme algoritmaları ile bulunan sonuçların ne kadar iyi olduğunun belirlenmesi için bir amaç kriteri bulunmasına dayanır. Aşağıdaki hedeflere ulaşılması için bu kriterin bulunması önemlidir (Erilli, Tunç, Öner, & Yolcu, 2009):

- Veri seti için en iyi küme sayısının belirlenmesi

- Veri setinin kümelenmesi için alternatif algoritma sonuçlarının karşılaştırılması
- Veri setinin herhangi bir doğal gruplama yapısının olup olmadığının belirlenmesi

Bu hedeflerden en önemlisi en iyi küme sayısının belirlenmesidir. Bu konuda oldukça farklı yöntemler literatürde bulunmaktadır. Optimal küme sayısını belirlemek için aşağıda verilen yöntemler önerilmiştir.

- Bezdek Bölünme Katsayısı (PC)
- Bezdek Bölünme Entropisi (PE)
- Xie-Beni Katsayısı
- Fukuyama ve Sugeno Katsayısı

gibi sık kullanılan yöntemler olarak karşımıza çıkmaktadır (Bezdek J. , 1974).

1.10.3.1 Bezdek Bölünme Katsayısı (PC)

c küme sayısı, μ üyelik değerleri olmak üzere bölünme katsayısı aşağıdaki gibi tanımlanır:

$$v_{PC}(c) = \frac{\sum_{j=1}^n \sum_{i=1}^c \mu_{ij}^2}{n} \quad (26)$$

burada bölünme katsayısını maksimum yapan küme sayısı, $k = \max\{v_{PC}(\mu, c_i)\}$ şeklinde bulunmaktadır (Bezdek J. C., 1975).

1.10.3.2 Bezdek Bölünme Entropisi (PE)

c küme sayısı, μ üyelik değerleri olmak üzere bölünme entropisi aşağıdaki gibi tanımlanır:

$$v_{PE}(c) = -\frac{1}{n} \sum_{j=1}^n \sum_{i=1}^c [\mu_{ij} \log_a(\mu_{ij})] \quad (27)$$

Burada bölünme entropisini minimum yapan küme sayısı, $k = \min\{v_{PE}(\mu, c_i)\}$ optimal küme sayısını vermektedir (Bezdek J. , 1974).

1.10.3.3 Xie-Beni kriter (XB)

Xie - Beni tarafından (1991) (Xie & Beni, 1991) geliştirilen bu kriter, yoğunluk ve ayrılma geçerlilik fonksiyonu olarak da bilinir. c küme sayısı, μ üyelik değerleri, v_i i . küme merkezi olmak üzere Xie-Beni kriteri aşağıdaki gibi tanımlanır:

$$v_{XB}(c) = \frac{\sum_{j=1}^c \sum_{i=1}^n \mu_{ij}^2 \|x_i - v_j\|^2}{n \left(\min_{i \neq j} \|x_i - v_j\|^2 \right)} \quad (28)$$

Burada Xie-Beni kriterini minimum yapan küme sayısı, $k = \min\{v_{XB}(\mu, c_i)\}$ optimal küme sayısını vermektedir.

1.10.3.4 Fukuyama ve Sugeno kriterleri (FS)

c küme sayısı, μ üyelik değerleri, v_i i . küme merkezi ve \bar{v} küme merkezlerinin ortalaması olmak üzere Fukuyama ve Sugeno kriterleri aşağıdaki gibi tanımlanır.

$$v_{FS}(c) = \sum_{i=1}^n \sum_{j=1}^c \mu_{ij}^m \|x_i - v_j\|^2 - \sum_{i=1}^n \sum_{j=1}^c \mu_{ij}^m \|v_j - \bar{v}\|^2 \quad (29)$$

$$1.4 < m < 2$$

Burada Fukuyama ve Sugeno kriterleri minimum yapan küme sayısı, $k = \min\{v_{FS}(\mu, c_i)\}$ optimal küme sayısını vermektedir.

2. YAPILAN ÇALIŞMALAR

2.1. Ön İşlemler

Sinyal işleme sürecinden geçmeden önce bazı ön işlemlerin uygulanması gerekir. Bu ön işlemler yapılmazsa sinyal işleme sırasında istenmeyen durumlarla karşılaşılabilir. Birçok sinyal işleme uygulaması sinyalin içeriğindeki bazı istenmeyen olayların varlığından dolayı büyük hatalara neden olabilmektedir. Bu istenmeyen olaylar ve bunların yok edilmesi konularına aşağıda değinilmiştir.

Gerçek dünyada ses dalgaları sürekli bir fonksiyon olmasına rağmen bilgisayar ses sinyali kayıt yaparken, ayrık değerler biçiminde saklanmaktadır. Aşağıdaki eşitlik kullanılarak ses sinyali Δt aralıklarda örneklenerek ayrık veriler haline dönüştürülmektedir.

$$x[n] = x(n \cdot \Delta t), \quad n = 0, 1, \dots, N - 1 \quad (30)$$

burada, $x[n]$ ayrık ses sinyalini, $x(t)$ sürekli ses sinyalini göstermektedir. Δt ardışık iki örnek arasındaki zaman (örnekleme aralığı) farkıdır. Örnekleme aralığının tersi (31),

$$f_s = \frac{1}{\Delta t} \quad (31)$$

örnekleme frekansı veya örnekleme oranı olarak adlandırılmaktadır. Birimi Herzt (Hz) veya s^{-1} olarak gösterilmektedir.

2.2. Ortalama Kayması (DC) Düzeltmesi

Genelde ses sinyali ilk kayıt ederken, sinyalin genel ortalaması tam sıfır olmamaktadır. Kayıt sistemin kalitesiyle ilişkili olan bu durum veri işlem aşamasında sorunlara yol açmaktadır. Bu sorunu ortadan kaldırmak için tüm sinyalin ortalaması alınır ve her veriden tek tek çıkarılır. Bu işlemin matematiksel gösterimi

$$x[n] = x[n] - \frac{1}{N} \sum_{m=0}^{N-1} x[m] \quad (32)$$

biçiminde verilmektedir.

2.3. Ortalama Silme Yöntemi

Bazı sinyaller çevre etkenlerden kaynaklanan düşük frekansa sahip bileşenler içerebilir. Bu sinyaller veri işlem aşamasında istenmeyen sorunlara neden olabileceğinden bunların sinyalden ayıklanması gerekir. Bunun için sinyaldeki düşük frekanslı bileşenleri silecek basit ve kullanışlı yöntem, ortalama silme yöntemidir. Bu yöntemin matematiksel gösterimi,

$$x[n] = x[n] - \frac{1}{M} \sum_{m=n-\frac{M}{2}}^{n+\frac{M}{2}} x[m], \quad n = 0, 2, \dots, N - 1 \quad (33)$$

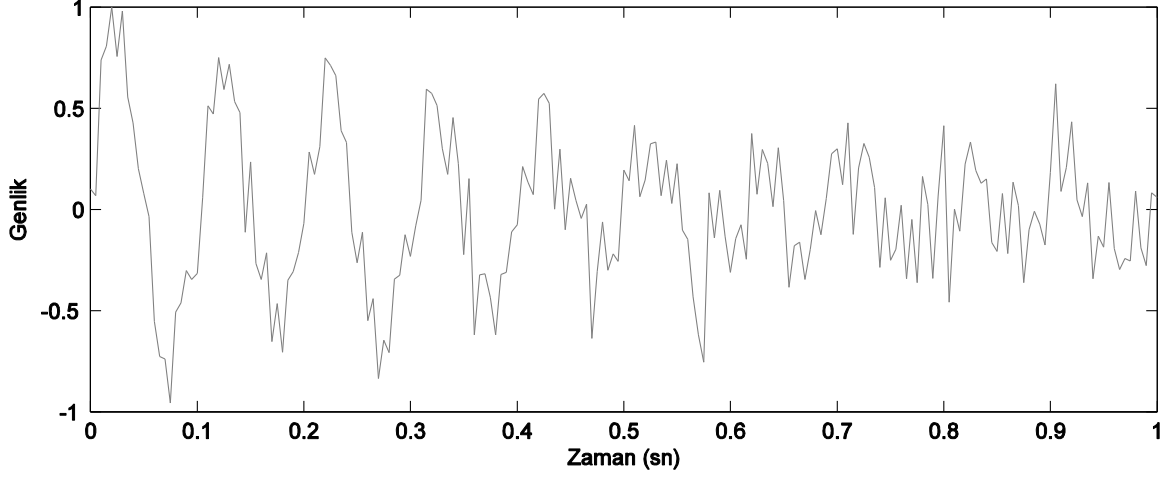
biçiminde verilmektedir. Bu işlemde M büyüklüğünde bir pencere seçilir ve bu M veriden alınan ortalama seçilen pencerenin ortasındaki veriden çıkartılır. Aynı işlem N defa tekrar edilerek işlem tamamlanmış olur. Böylece verideki düşük frekanslı bileşenler yok edilmiş olur. Ses sinyali işlenebilir hale gelir. Bu yöntem ortalama kayması düzeltmesinin de işlevini yerine getirmektedir.

2.4. Genlik Eşitleme

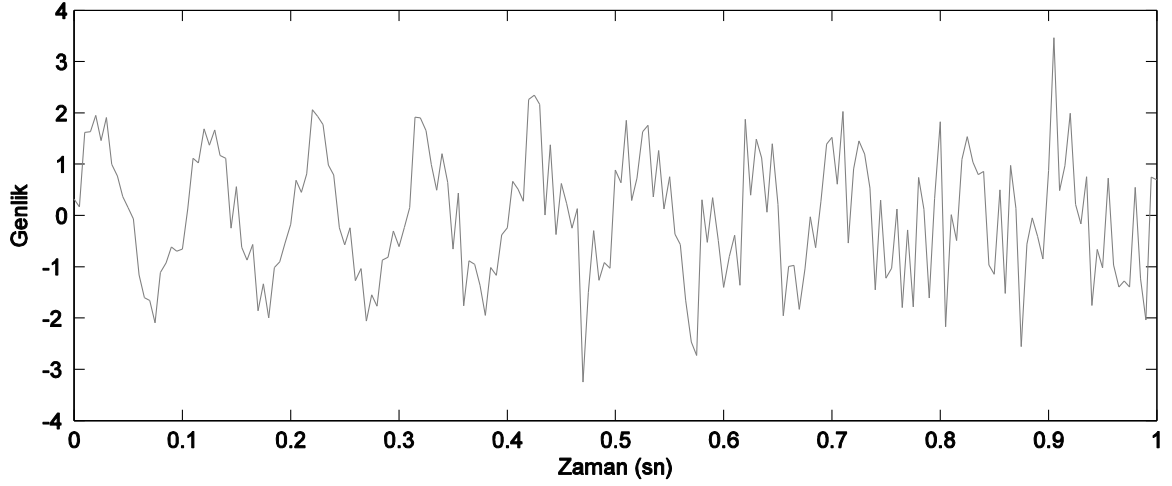
Genlik eşitleme, ses sinyalindeki farklı enerji seviyelerine sahip bölgeleri aynı enerji seviyesine çekmek için kullanılmaktadır. Bu işlemde M büyüklüğünde bir pencere seçilir ve bu M veriden alınan ortalama seçilen pencerenin ortasındaki veriyi bölerek işlem gerçekleştirilir. Aynı işlem N defa tekrar edilerek işlem tamamlanmış olur (34).

$$x[n] = \frac{x[n]}{\frac{1}{M} \sum_{m=n-\frac{M}{2}}^{n+\frac{M}{2}} x[m]}, \quad n = 0, 1, \dots, N - 1 \quad (34)$$

Genliđi dengelenmiř bir sinyal řekil 12(b)'de gstermektedir.



(a)



(b)

řekil 12. Genlik eřitlemesi; (a) İřlenmeye hazır ses sinyali; (b) Genliđi dengelenmiř bir ses sinyali

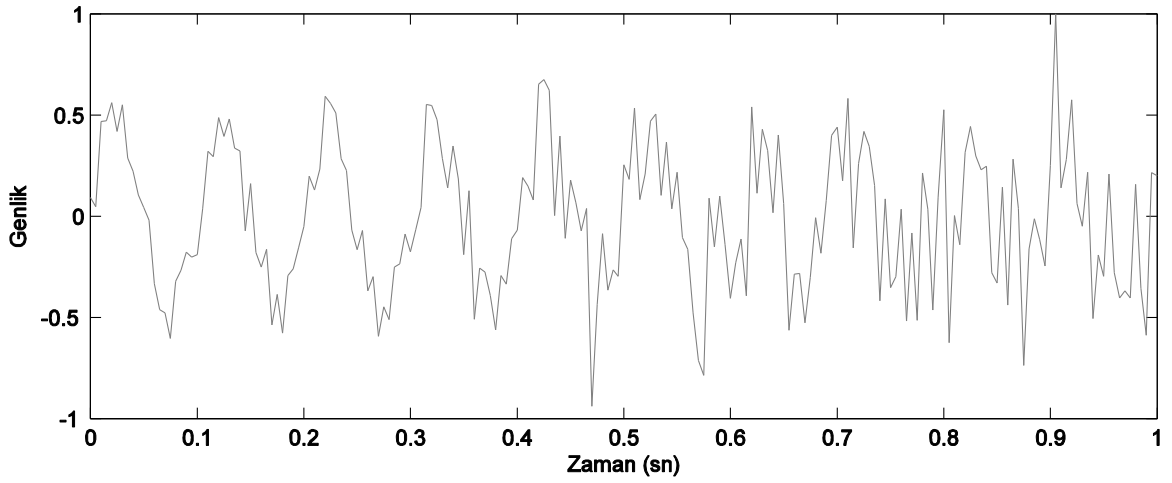
2.5. Normalleřtirme

Ses sinyalleri mikrofona deđiřik uzaklıkta veya deđiřik biçimlerde kayıt edildiklerinde, ses dzeyleri (enerjileri) eřit olmayabilir. Ses sinyalin verilerini genlik

dengeleme uygulayarak verilerin genlik aralığı değişebilir ve tüm ses sinyallerine aynı işlemi uygulamak için sinyaller $[-1,1)$ aralığına ölçeklendirilir. Bunun için,

$$x[n] = \frac{x[n]}{\max_m(|x[m]|)}, \quad n, m = 0, 1, \dots, N - 1 \quad (35)$$

eşitliği kullanılır.



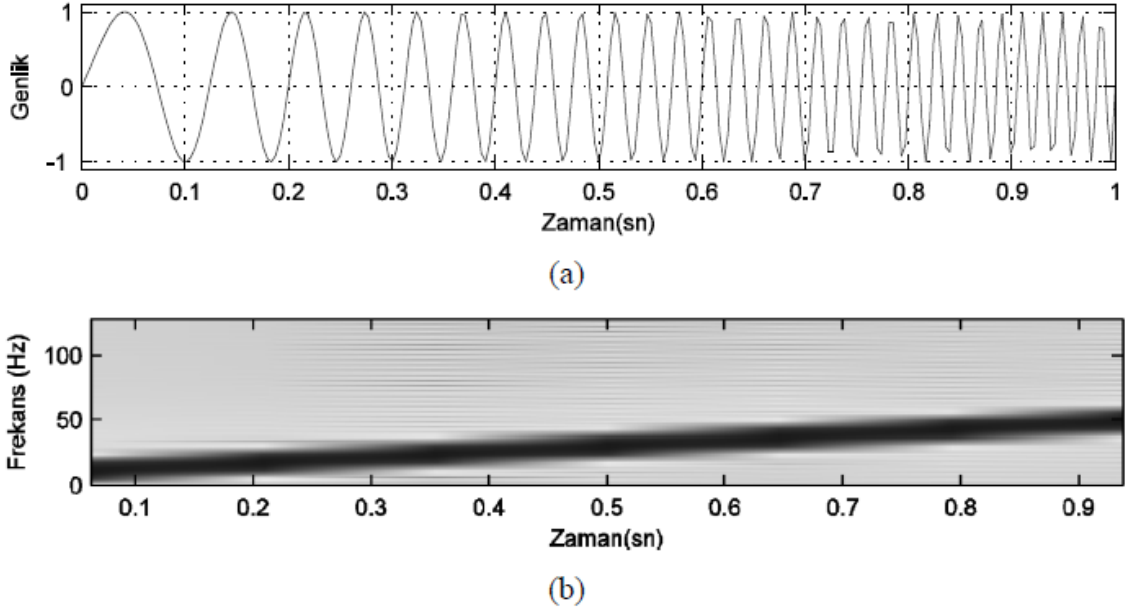
Şekil 13. Normalize edilmiş bir sinyal

2.6. Spektrogram

Spektrogram bir sinyalin spektral içeriğinin zamana göre değişimini gösteren frekans ve zamana bağlı iki boyutlu bir spektrum fonksiyonudur. STFT kullanarak elde edilen iki boyutlu zaman-frekans fonksiyonundan

$$spec[k, l] = |X[k, l]|^2 \quad (36)$$

biçiminde hesaplanır. Spektrogram, ses ölçer, ses izi gibi adlar da almaktadır (Spectrogram, 2014). Şekil 19(a)'da zamanla artan frekansa (5-50 Hz) sahip 1 saniyelik yapay bir ses sinyali verilmektedir. Bu sinyalin spektrogramı da Şekil 19(b)'de verilmektedir. Frekansta ani bir değişimin olmaması spektrogramda düzgün bir geçişin olduğu görülmektedir (Karpuz, 2011).

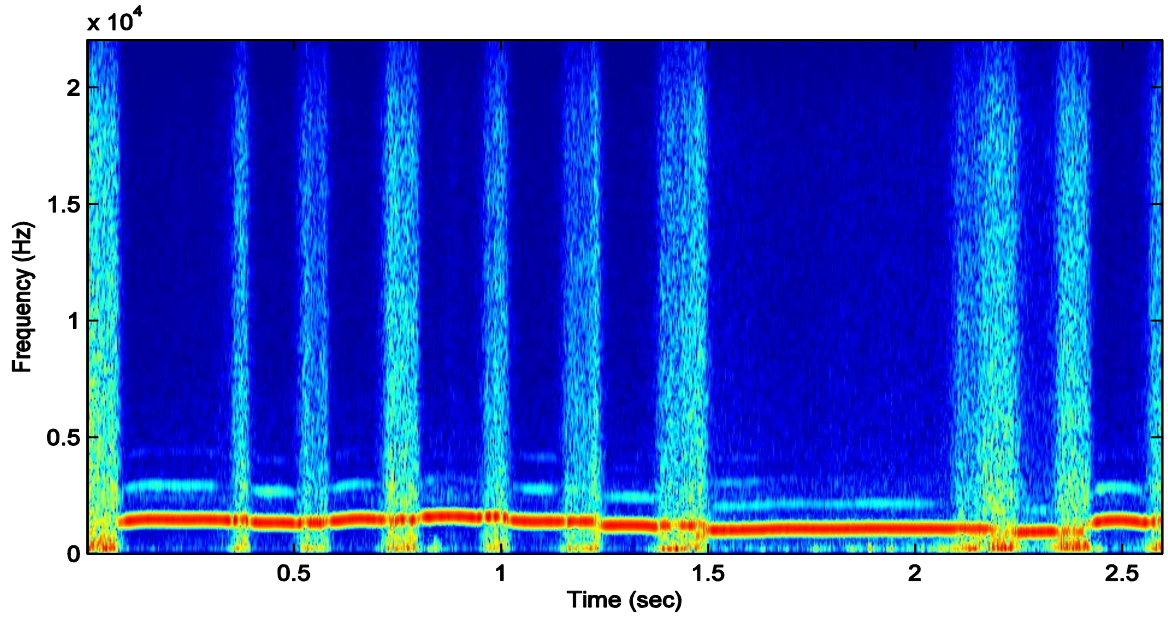


Şekil 14. Spektrogram gösterimi, (a) frekansı zamanla artan yapay ses sinyali, (b) spektrogram gösterimi

2.7. Ses Sinyali Spektrogramı

Şekil 15'teki spektrogram, flütle çalınmış bir müzik parçasının spektrogramını göstermektedir. Ses sinyali zamana göre frekansın nasıl değiştiğini görmek için veriler üzerinde kısa süreli Fourier dönüşümü uygulanır. Öncelikle verileri M boyutta küçük çerçevelere ayrılarak gerçekleştirilir (37). Burada açık mavi rengi olan kısımlar sesiz bölgeleri göstermektedir bu bölgeler genelde çok düşük genliğe sahip bölgelerdir ama genlik eşitleme uygulandığında bu bölgeler güçlü genliğe sahip olurlar. Şekil 15'den görüldüğü gibi ses sinyali, sessizlik kısımlara rağmen yine devam etmekte ve tamamen yok olmamaktadır. Dolayısıyla bu ses sinyal bir bulanık yapıya sahip olduğunu sonucuna varılmaktadır.

$$S_j[k] = \sum_{m=0}^{M-1} s_j[m] e^{-i2\pi \frac{m.k}{N}}, \quad k = 0,1,2, \dots, M-1 \quad (37)$$

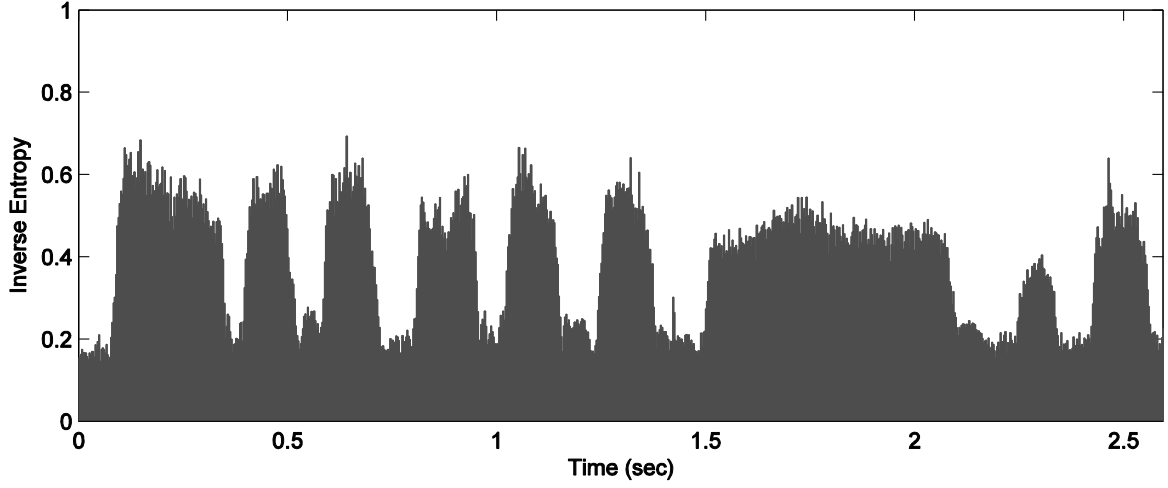


Şekil 15. Flütle çalınmış bir notanın spektrogramı

2.8. Olabilirliğin Elde Edilmesi

Entropy rastgeleliği, belirsizliği ve beklenmeyen durumun ortaya çıkma olasılığını gösterir. Bu aşamada ses sinyalinin spektrogramı kısa zamanlı Fourier dönüşümüyle elde edildikten sonra ters entropisi alınarak spektrogramın zamana göre değişen olabilirliği bulunmuş olur. Şekil 16'da spektrogramın ters entropi uygulanarak elde edilmiş olabilirlikleri gösterilmektedir. Burada olabilirliğin yüksek olduğu bölgeler ses aktivitesinin bulunduğu bölgeleri göstermektedir.

$$P_s = \frac{1}{Entropy\{Spektrogram\{x\}\}} \quad (38)$$



Şekil 16. Spektrogramın ters entropi uygulanarak olabirlikleri elde edilmiş ses sinyali

2.9. Ağırlıklı Bulanık C-Ortalama Kümelemesi

Bölüm 1.10.2’de sözü edilen bulanık c-Ortalama (FCM) algoritmasının olabirliği bulunmuş ses sinyallerine uygulanması için önerilen bir yöntemdir. Bu yöntemde tüm işlem adımları aynen FCM algoritmasında olduğu gibidir. Ancak tek farklılık küme merkezlerin bulunmasında üyelik değerlerinin ses sinyalinin olabirlik değerleri ile (39). eşitlikte görüldüğü gibi çarpılmasıdır.

$$v_j = \frac{\sum_{i=1}^N P_{S_i} u_{ij}^m t_i}{\sum_{i=1}^N P_{S_i} u_{ij}^m}, \quad t_i = n. \Delta t, \quad n = 0,1,2, \dots, N \quad (39)$$

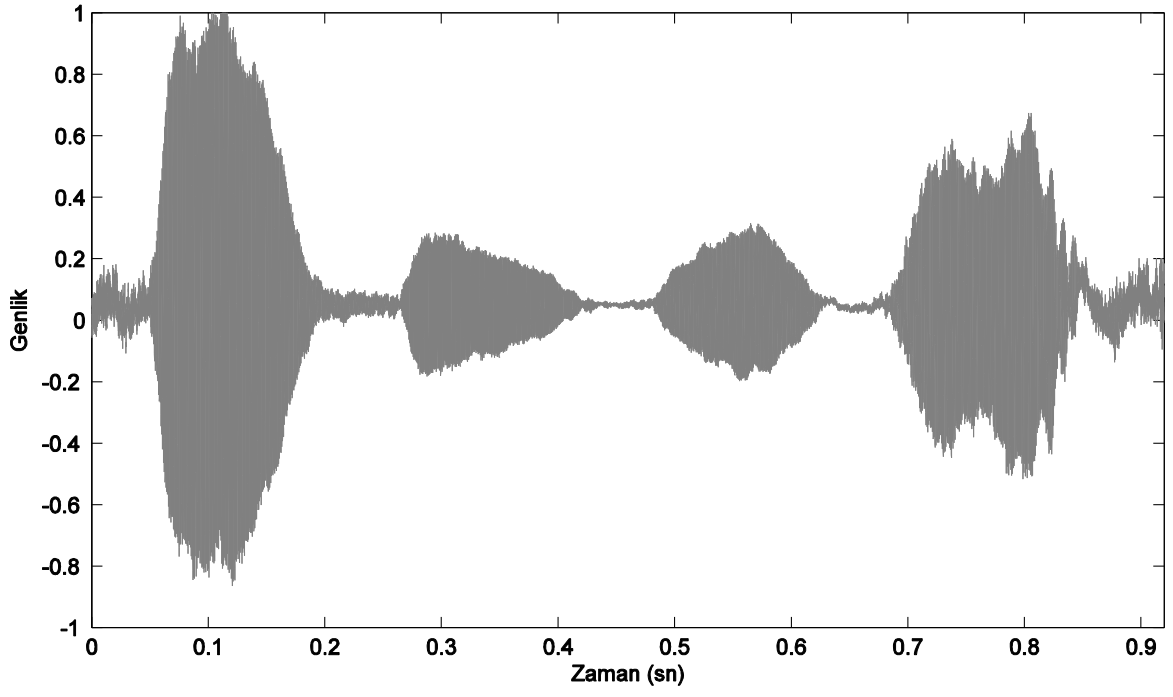
burada t_i , olabirlik değerlerine karşılık gelen ayırık zaman değerleridir.

2.10. Uygulama Örneği 1

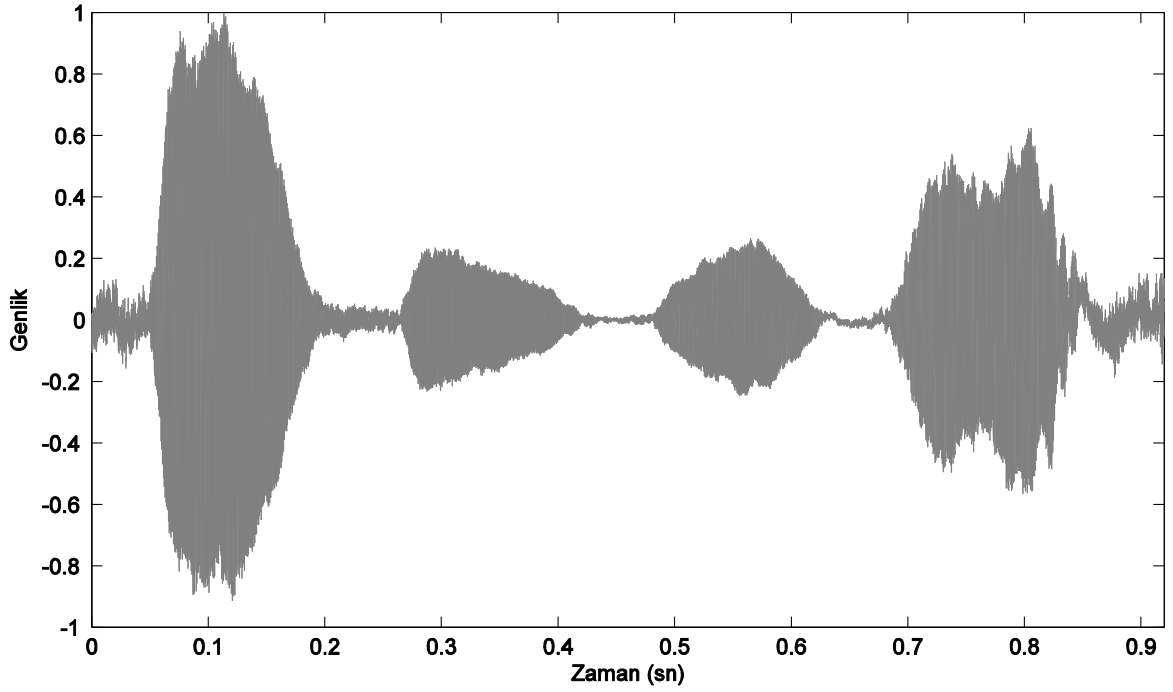
Orta kısımda düşük genlikli 4 seslilik bölgesinden oluşan bir ses sinyali örnek olarak alınmıştır (Şekil 17). Şekil 18’de örnek sinyalin ortalaması ortalama kayma düzeltmesi ile sıfır seviyesine çekilmiştir. Şekil 19’da ortalama çıkarma yöntemi ile sinyalin düşük frekanslı bileşenleri sinyalden atılmıştır. Şekil 20’de sinyalin enerji seviyeleri belirlenmiş ve bu seviyelere göre Otsu algoritması (Otsu, 1979) ile optimal eşik seviyesi belirlenmiştir.

Şekil 21’de belirlenen eşik seviyesine göre sinyalin eşiklemesi ve bulunan iki seslilik bölgesi gösterilmektedir.

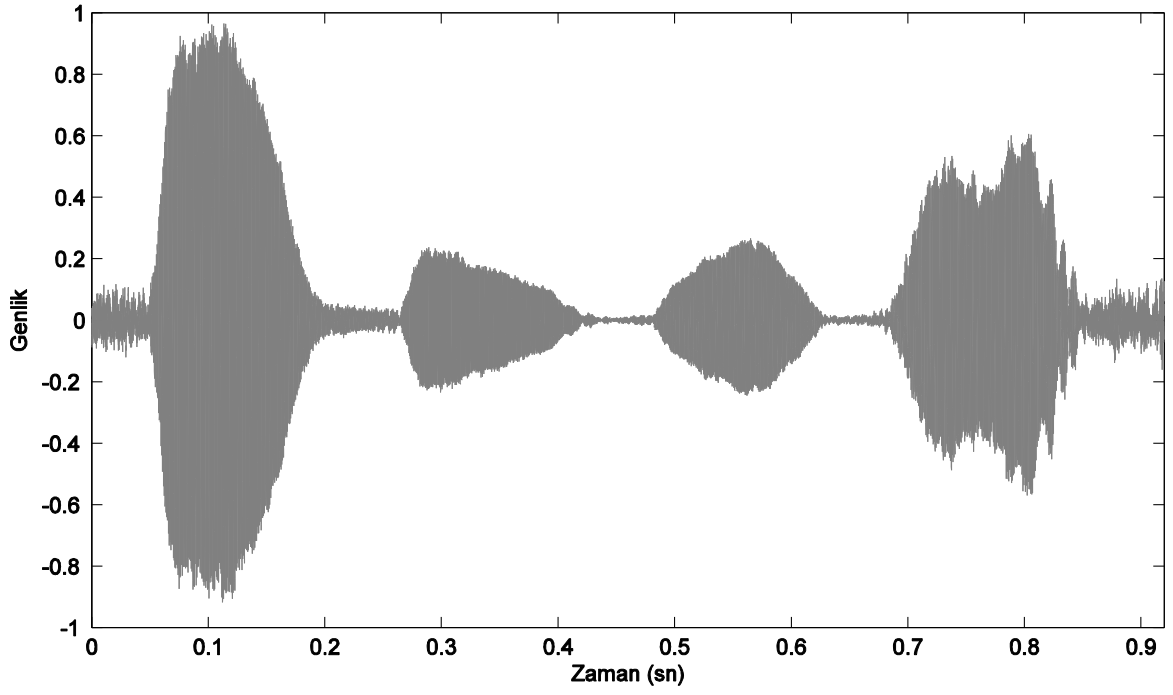
Şekil 22’de enerjinin düşük olmasından kaynaklanan yetersizliği ortadan kaldırmak için ses sinyaline genlik eşitlemesi uygulanmıştır. Genlik eşitlenmesinden sonra sinyalin tanım aralığı genişlediğinde bu sinyal Şekil 23’te görüldüğü gibi normalize edilmiştir. Normalize edilen ses sinyalinde enerji düzeyleri aynı konuma geldiğinden eşikleme ile bölütleme yapmak imkansızlaşmıştır. Bu sonucu Şekil 24’deki enerji düzeylerine bakıldığında daha net görülmektedir. Öte yandan sinyalin spektral ortamda bölütlenmesi gerçekleştirilebilir (Şekil 25). Şekil 25’te sinyalin spektrogramına göre seslilik mavi tonun bulunduğu bölgede olmasına rağmen diğer bölgelerde de sürekliliği devam etmektedir. Bu devamlılık olabilirlik değerlerinde de kendini göstermektedir (Şekil 26). Dolayısıyla seslilik bölgelerinin sessizlik bölgesinde de devam ettiği görülmektedir. Bu durumda her seslilik bölgesinin keskin sınırlarını belirlemek yerine bu sınırlar bulanık alınırsa ses tanıma sürecinde olumlu sonuçların alınacağı görülmektedir (Şekil 27). Şekil 27’de olabilirlik değerleriyle küme merkezleri ve bunların üyelik değerleri bulunmuştur. Şekil 19’da verilen sinyal ile tüm kümelerin üyelik değerleri teker teker çarpılarak yapılan pencereleme sonuçları Şekil 28-31 verilmiştir. Bu sonuçlara göre dört ayrı bulanık bölütlenmiş sinyal elde edilmiştir.



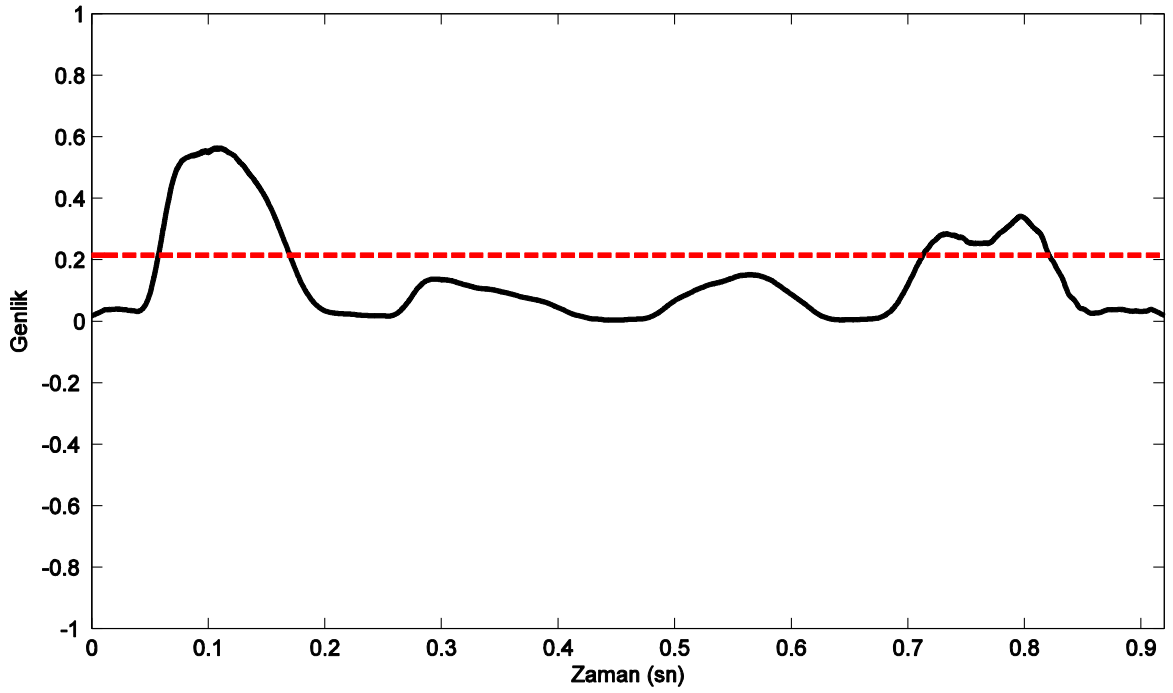
Şekil 17. Örnek ses sinyali



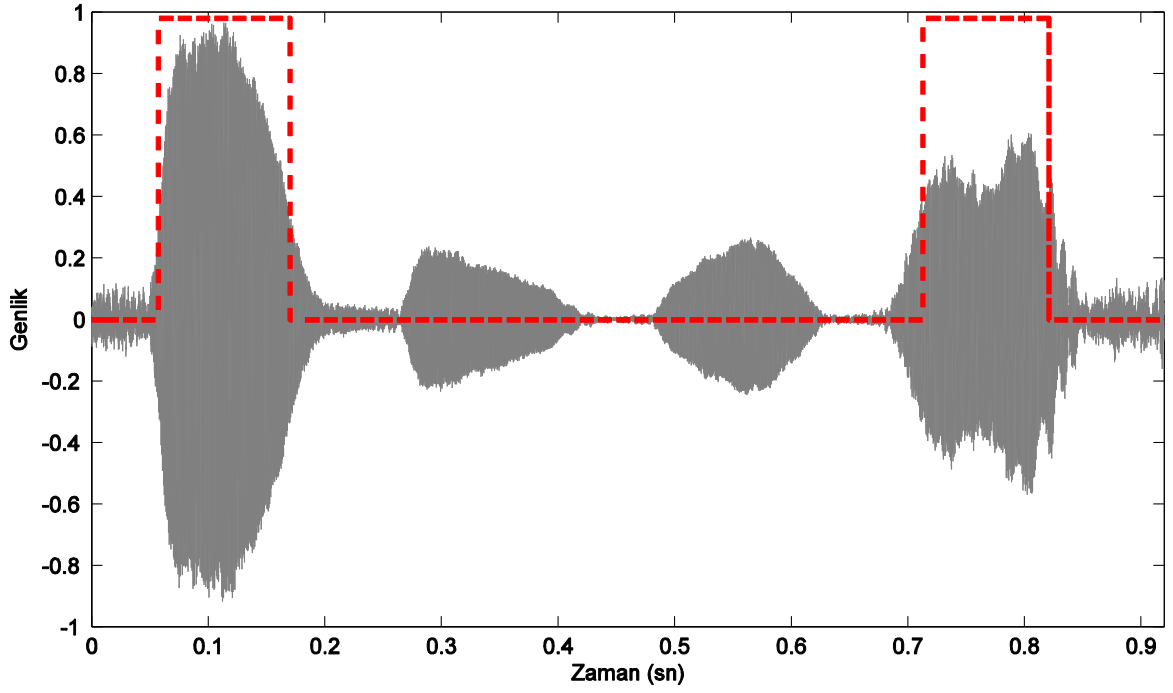
Şekil 18. Ortalama kaymansı uygulanmış ses sinyali



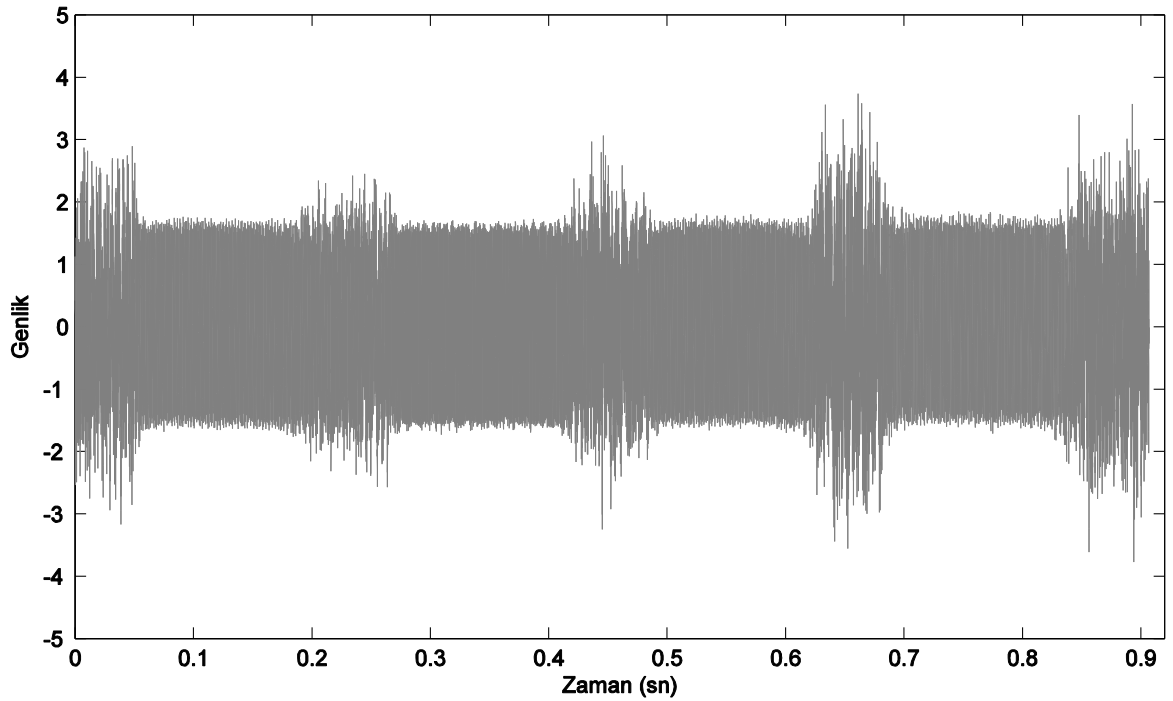
Şekil 19. Düşük frekans bileşenleri elenmiş ses sinyali



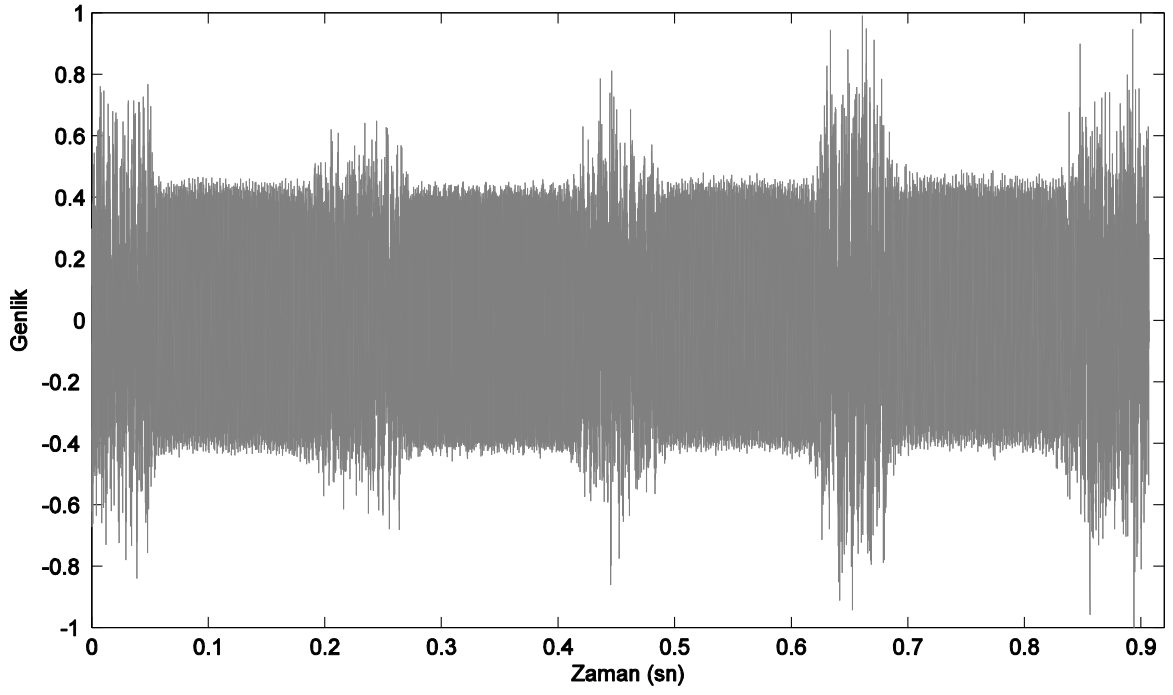
Şekil 20. Şekil 19'da verilen ses sinyalinin enerji düzeyleri ve eşik seviyesi



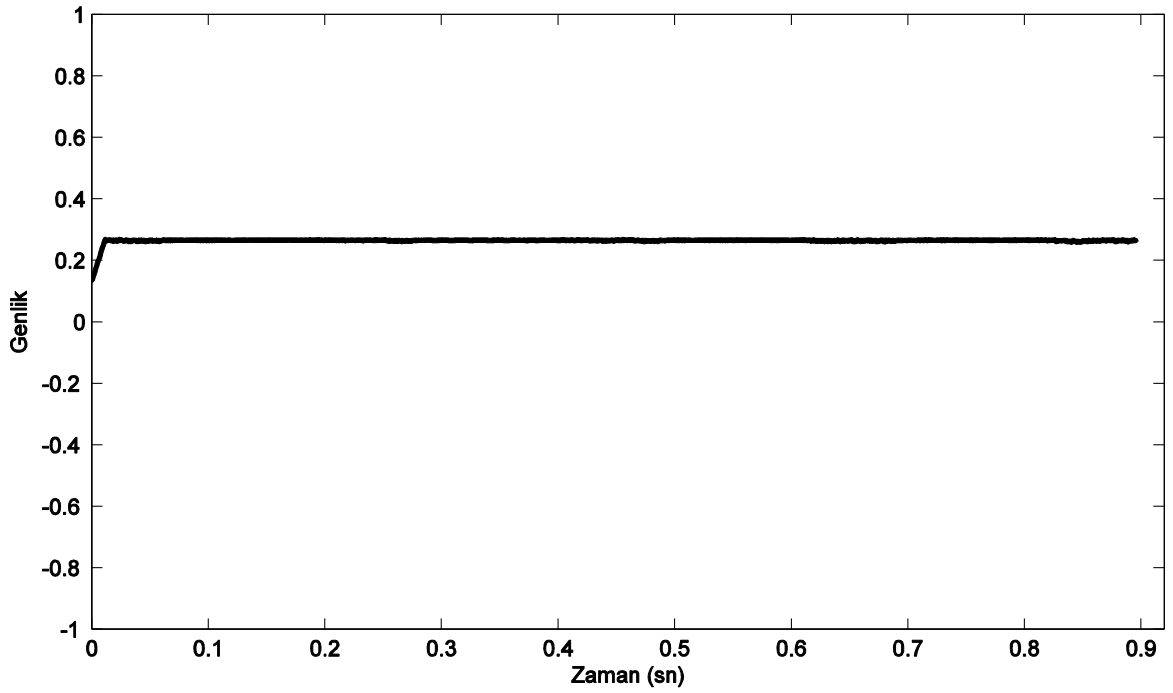
Şekil 21. Eşikleme yapılmış ses sinyali



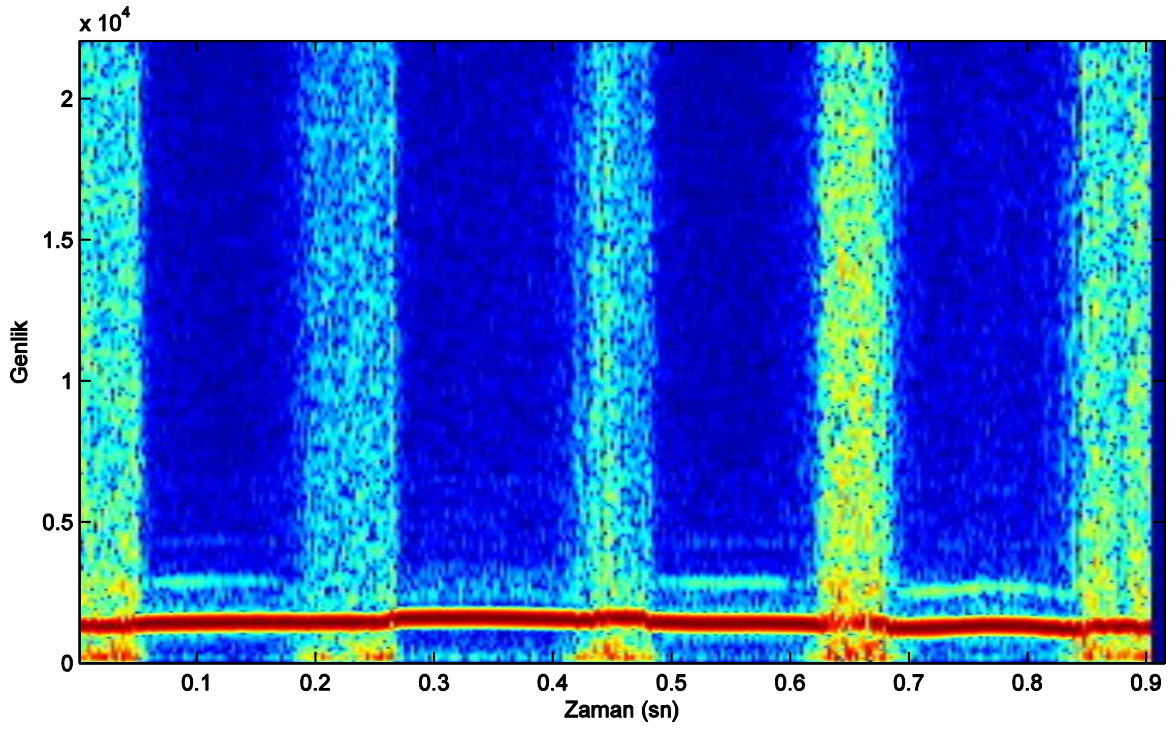
Şekil 22. Genlik eşitlemesi uygulanması



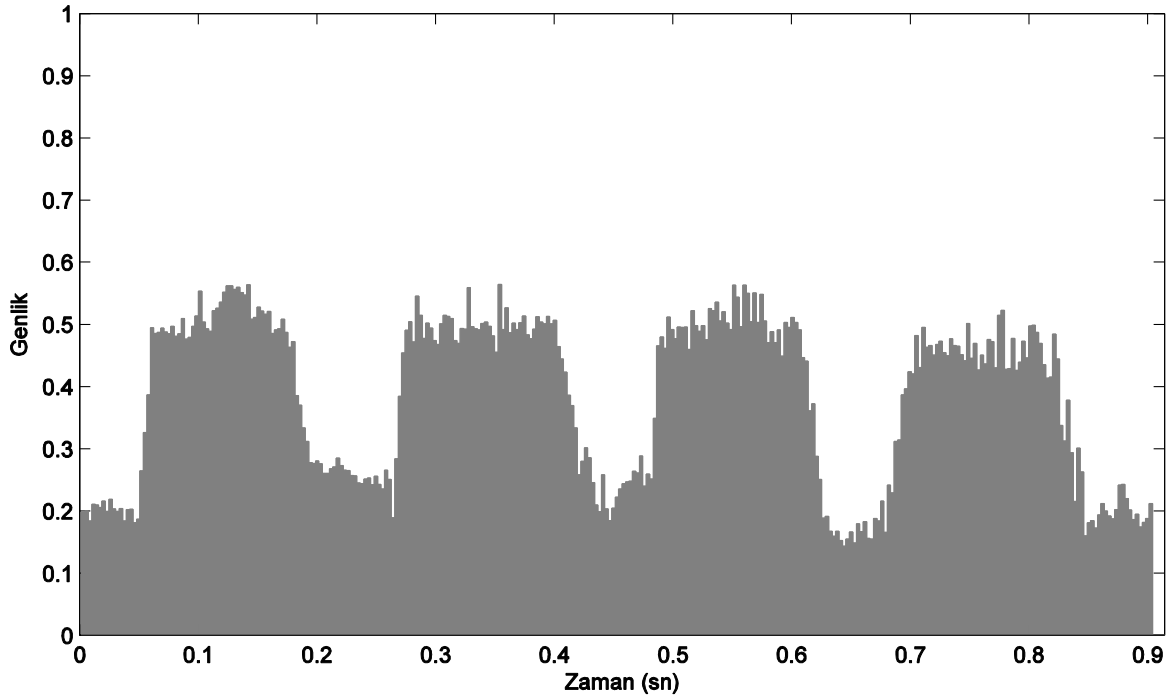
Şekil 23. Genlik eşitlemesi uygulanmış sinyali normalize edilmesi



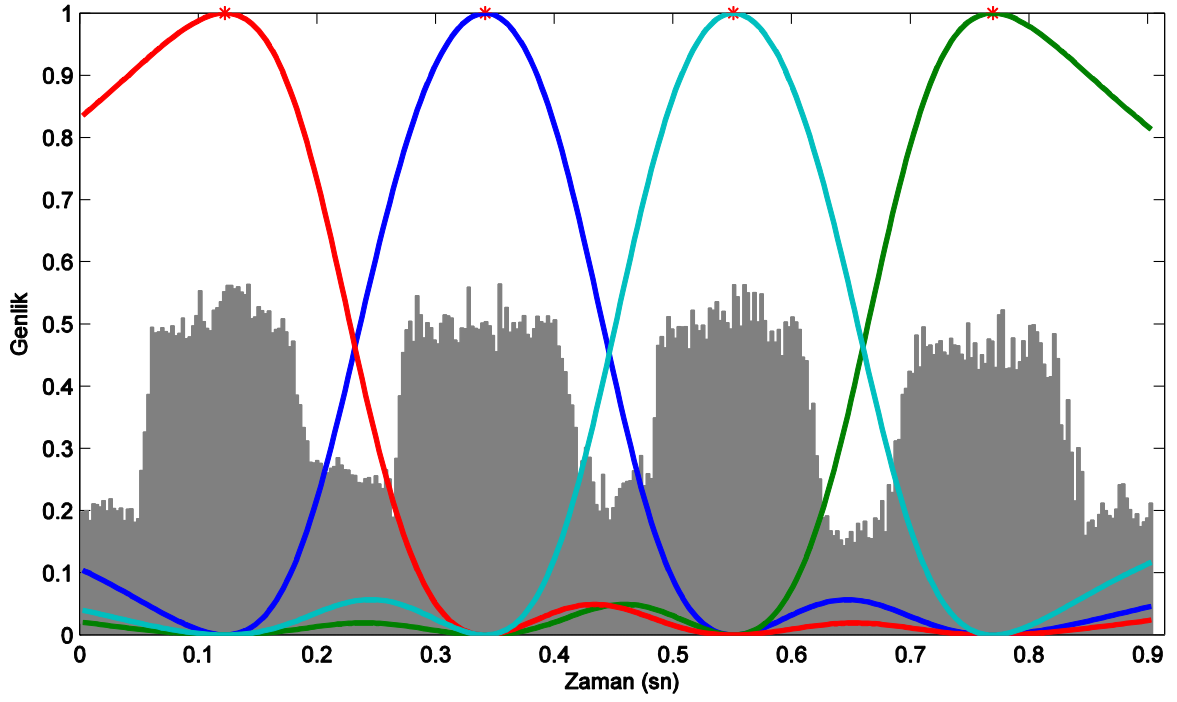
Şekil 24. Şekil 23'te verilen genlik eşitlemesi yapılmış sinyalin enerji düzeyi



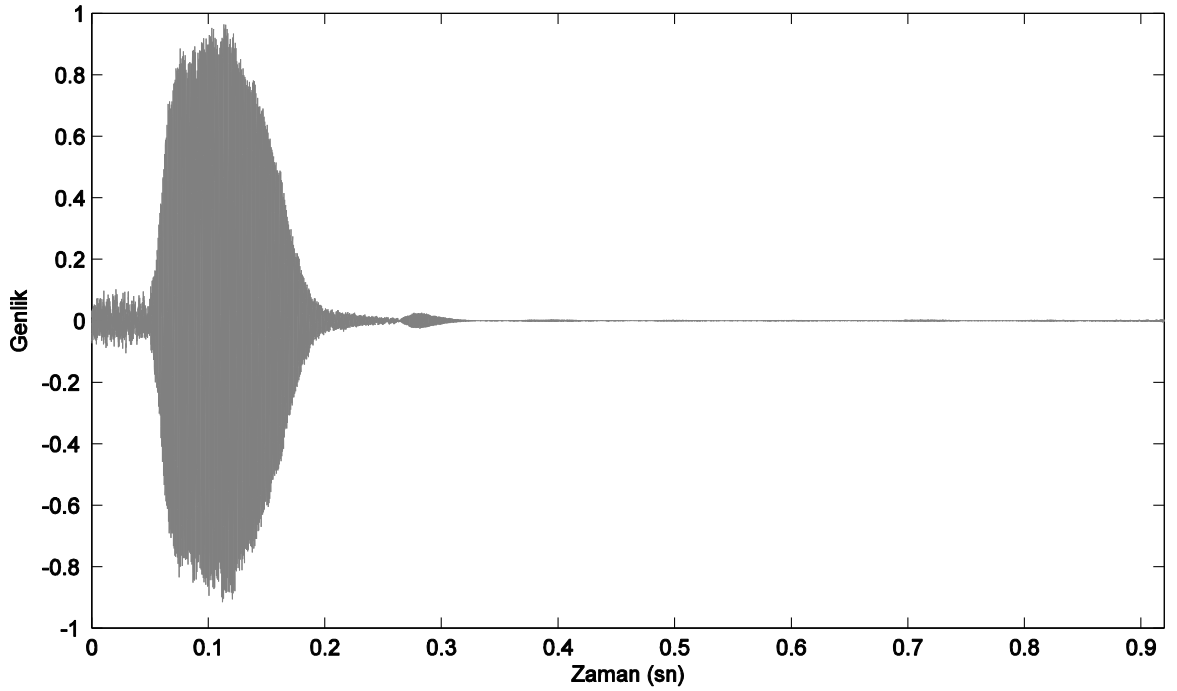
Şekil 25. Eşitlenmiş sinyalin spektrogramı



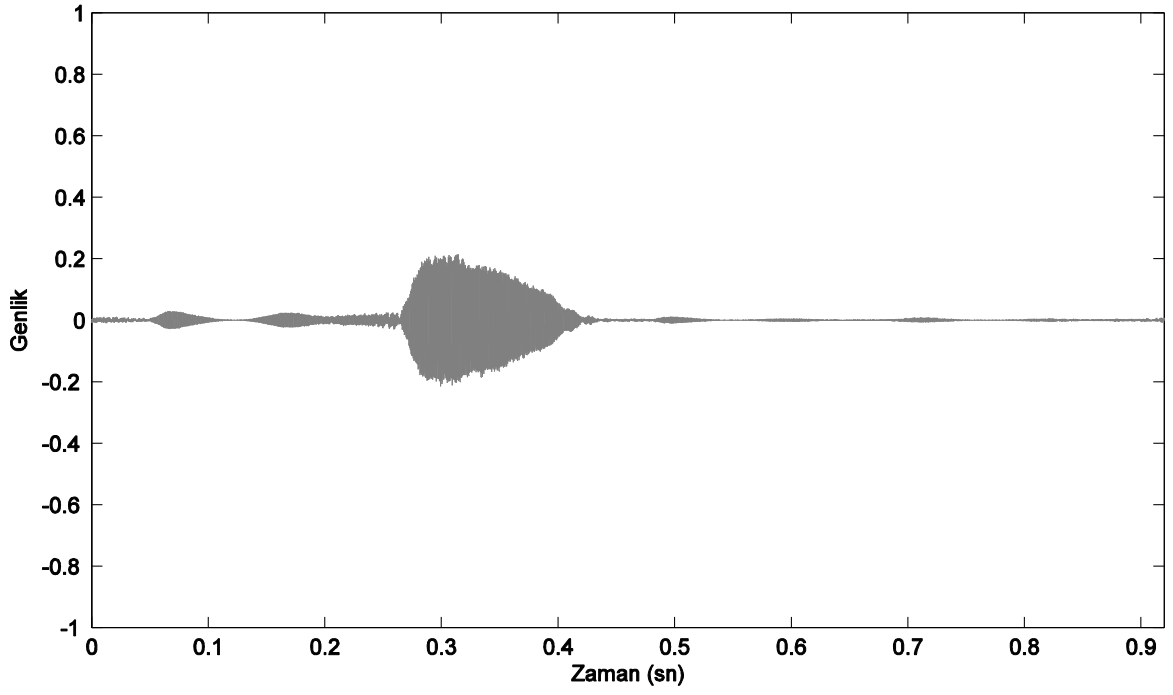
Şekil 26. Olabilirlik değerleri



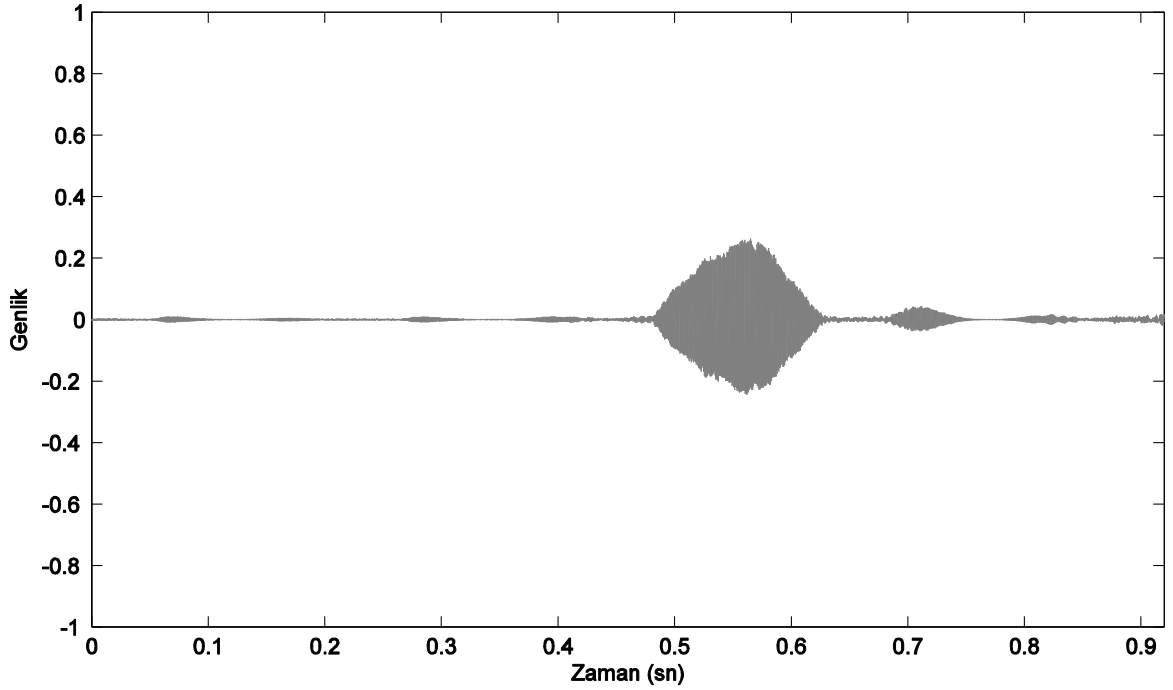
Şekil 27. Olabilirlik değerlerine göre sinyalin kümelenmesi



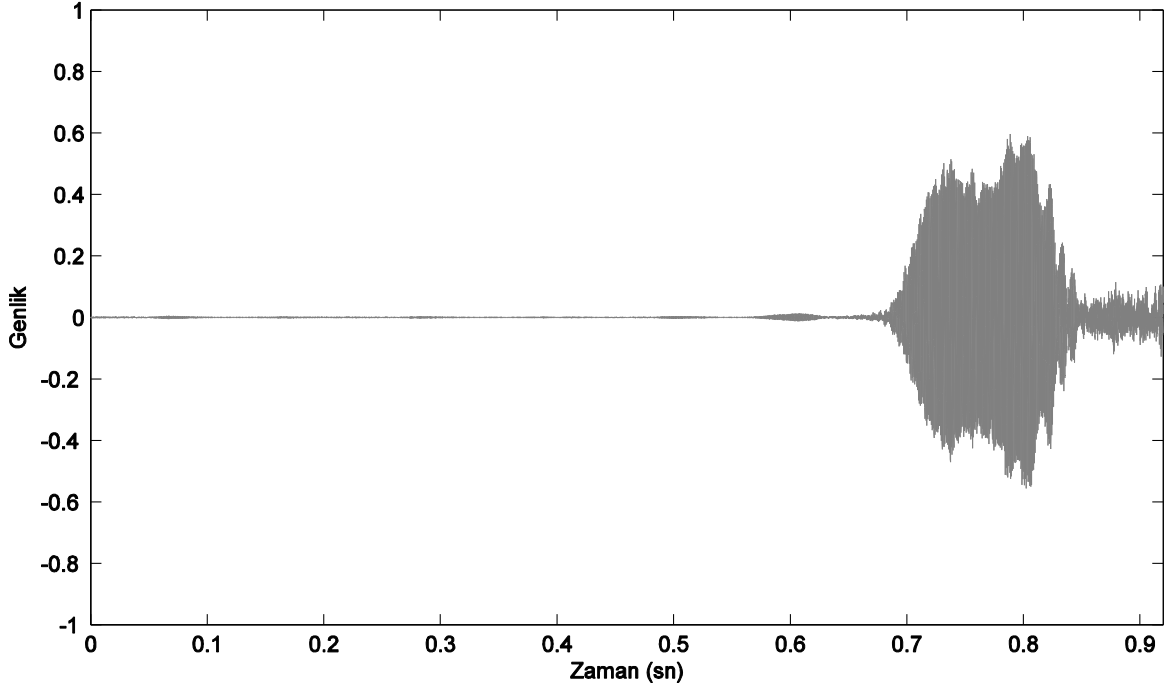
Şekil 28. Şekil 19'da verilen sinyal ile birinci kümenin üyelik değerlerinin çarpımıyla yapılan pencereleme sonucu elde edilen bulanık bölütlenmiş sinyal



Şekil 29. Şekil 19’da verilen sinyal ile ikinci kümenin üyelik değerlerinin çarpımıyla yapılan pencereleme sonucu elde edilen bulanık bölütlenmiş sinyal



Şekil 30. Şekil 19’da verilen sinyal ile üçüncü kümenin üyelik değerlerinin çarpımıyla yapılan pencereleme sonucu elde edilen bulanık bölütlenmiş sinyal



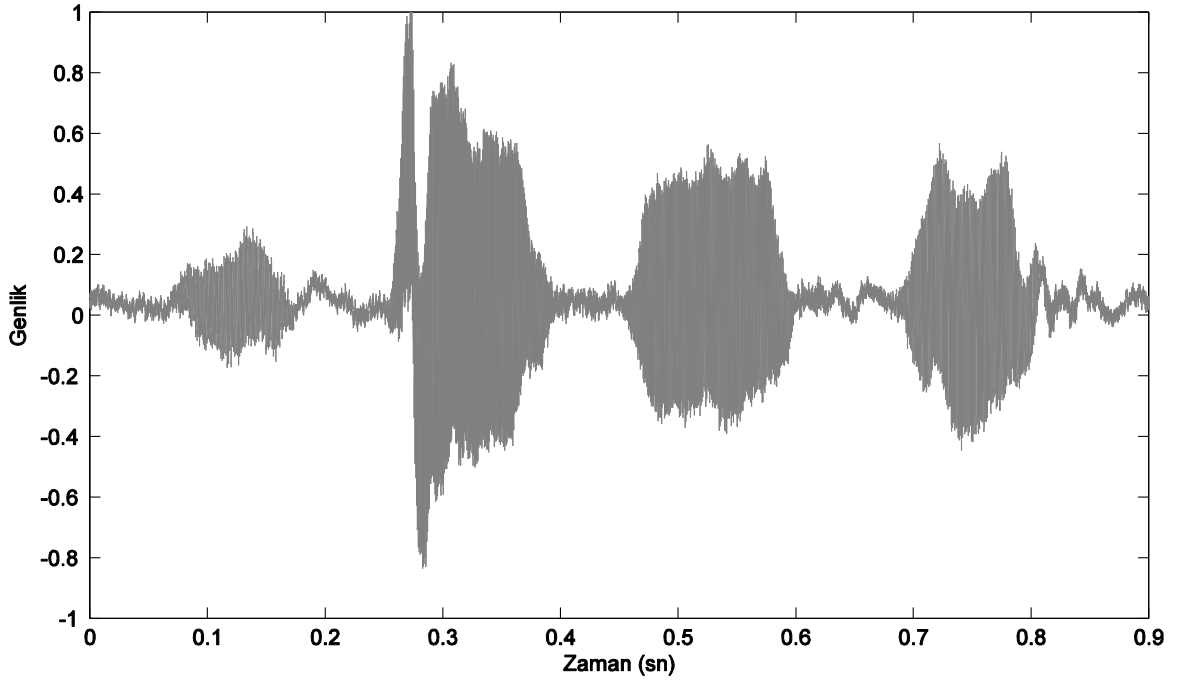
Şekil 31. Şekil 19’da verilen sinyal ile dördüncü kümenin üyelik değerlerinin çarpımıyla yapılan pencereleme sonucu elde edilen bulanık bölütlenmiş sinyal

2.11. Uygulama Örneği 2

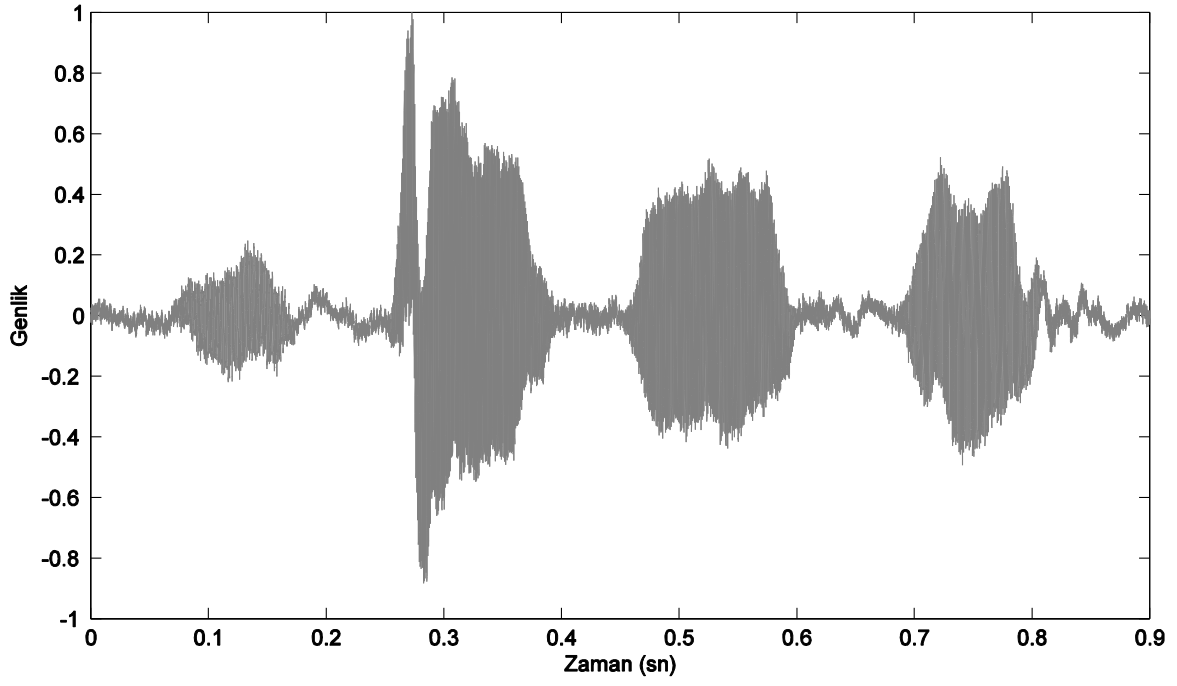
Sol kısımda düşük genlikli 4 seslilik bölgesinden oluşan bir ses sinyali örnek olarak alınmıştır (Şekil 32). Şekil 33’te örnek sinyalin ortalaması ortalama kayma düzeltmesi ile sıfır seviyesine çekilmiştir. Şekil 34’te ortalama çıkarma yöntemi ile sinyalin düşük frekanslı bileşenleri sinyalden atılmıştır. Şekil 35’te sinyalin enerji seviyeleri belirlenmiş ve bu seviyelere göre Otsu algoritması (Otsu, 1979) ile optimal eşik seviyesi belirlenmiştir. Şekil 36’da belirlenen eşik seviyesine göre sinyalin eşikleme ve bulunan üç seslilik bölgesi gösterilmektedir.

Şekil 37’de enerjinin düşük olmasından kaynaklanan yetersizliği ortadan kaldırmak için ses sinyaline genlik eşitlemesi uygulanmıştır. Genlik eşitlenmesinden sonra sinyalin tanım aralığı genişlediğinde bu sinyal Şekil 38’te görüldüğü gibi normalize edilmiştir. Normalize edilen ses sinyalinde enerji düzeyleri aynı konuma geldiğinden eşikleme ile bölütleme yapmak imkansızlaşmıştır. Bu sonucu Şekil 39’daki enerji düzeylerine bakıldığında daha net görülmektedir. Öte yandan sinyalin spektral ortamda bölütlenmesi

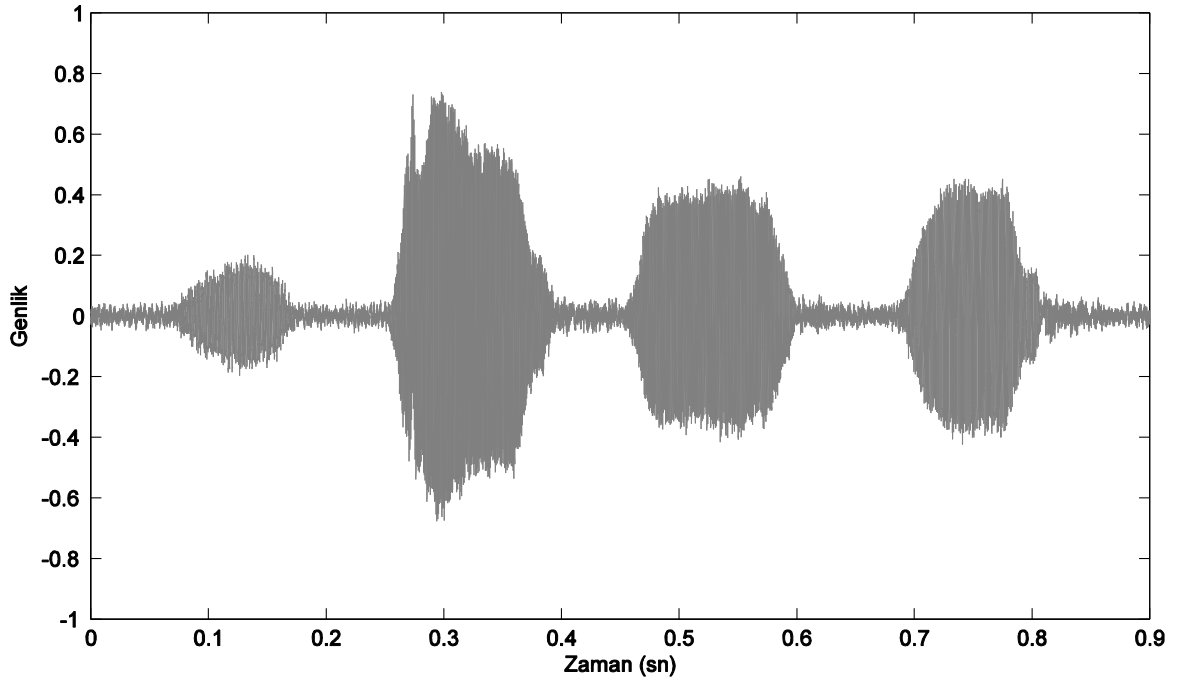
gerçekleştirilebilir (Şekil 40). Şekil 40'ta sinyalin spektrogramına göre seslilik mavi tonun bulunduğu bölgede olmasına rağmen diğer bölgelerde de sürekliliği devam etmektedir. Bu devamlılık olabilirlik değerlerinde de kendini göstermektedir (Şekil 41). Dolayısıyla seslilik bölgelerinin sessizlik bölgesinde de devam ettiği görülmektedir. Bu durumda her seslilik bölgesinin keskin sınırlarını belirlemek yerine bu sınırlar bulanık alınırsa ses tanıma sürecinde olumlu sonuçların alınacağı görülmektedir (Şekil 42). Şekil 42'de olabilirlik değerleriyle küme merkezleri ve bunların üyelik değerleri bulunmuştur. Şekil 34'de verilen sinyal ile tüm kümelerin üyelik değerleri teker teker çarpılarak yapılan pencereleme sonuçları Şekil 43-46 verilmiştir. Bu sonuçlara göre dört ayrı bulanık bölütlenmiş sinyal elde edilmiştir.



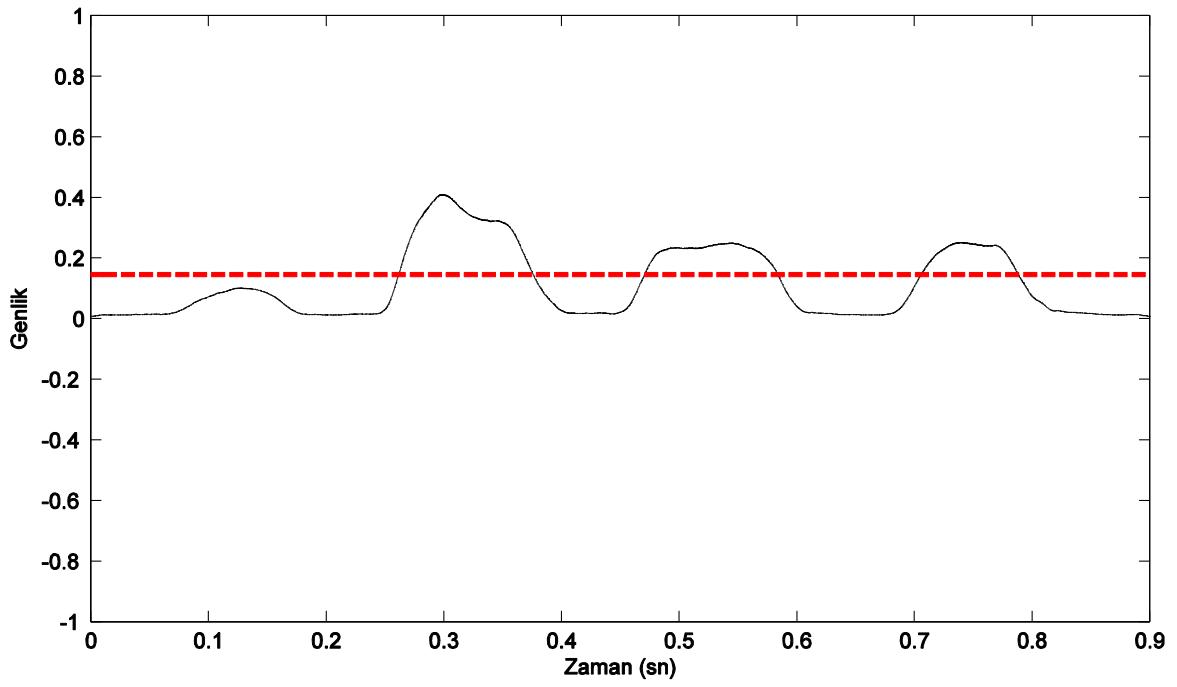
Şekil 32. Örnek ses sinyali



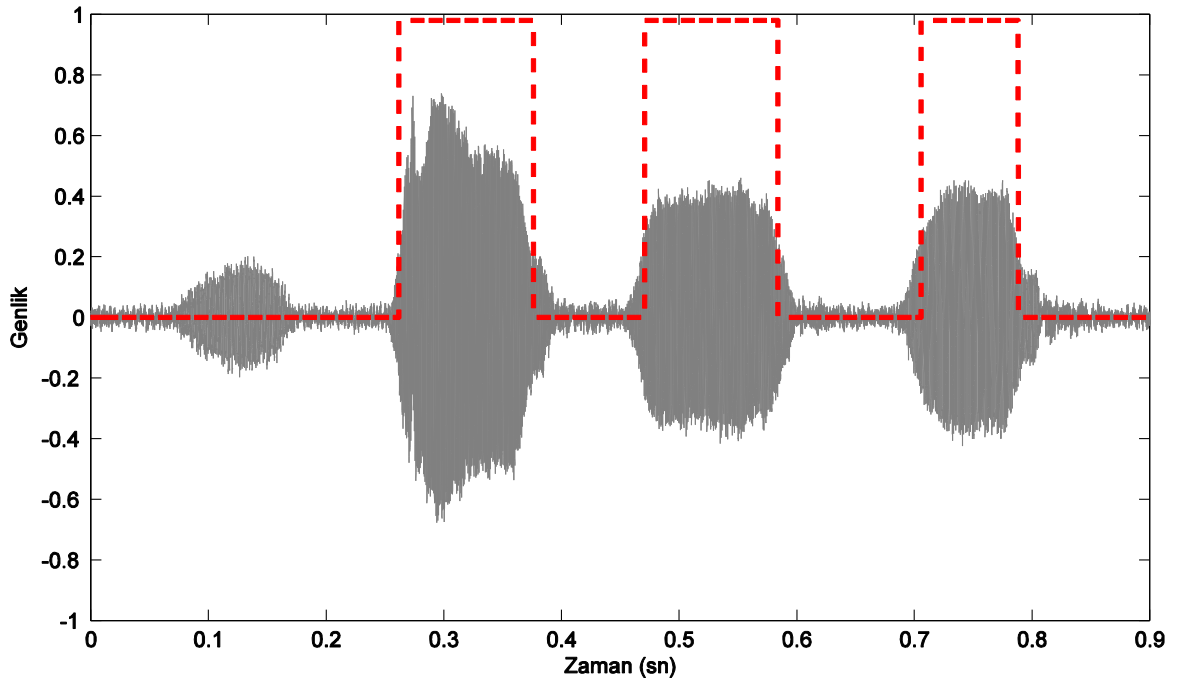
Şekil 33. Ortalama kaymansız uygulanmış ses sinyali



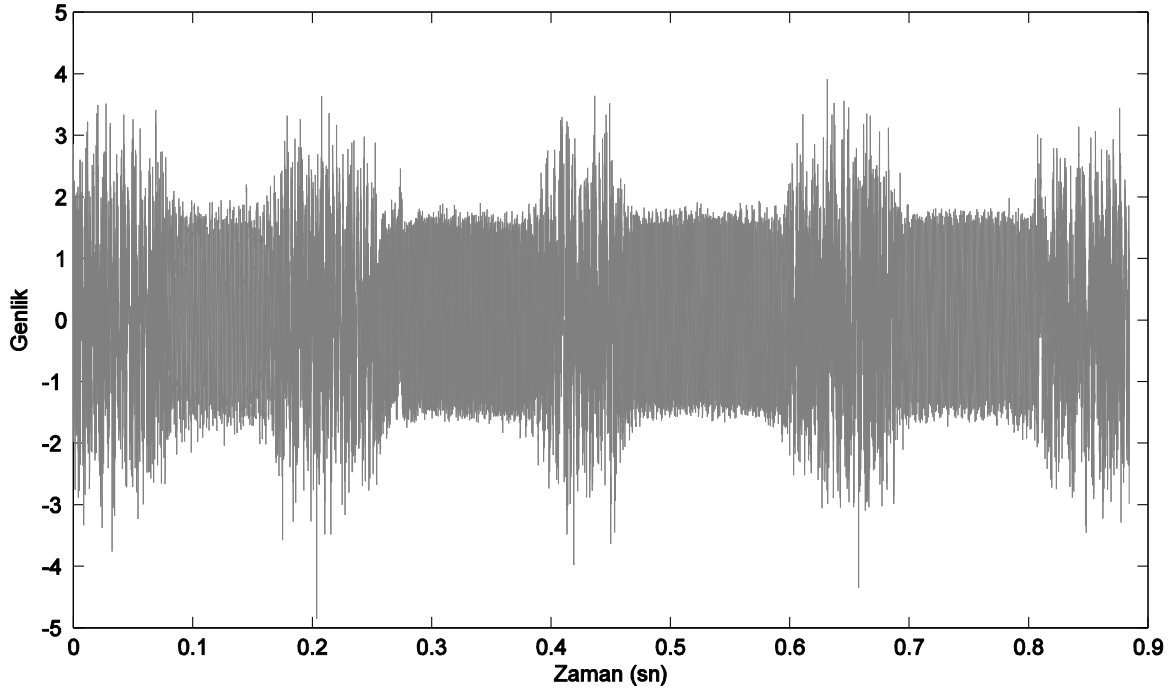
Şekil 34. Düşük frekans bileşenleri elenmiş ses sinyali



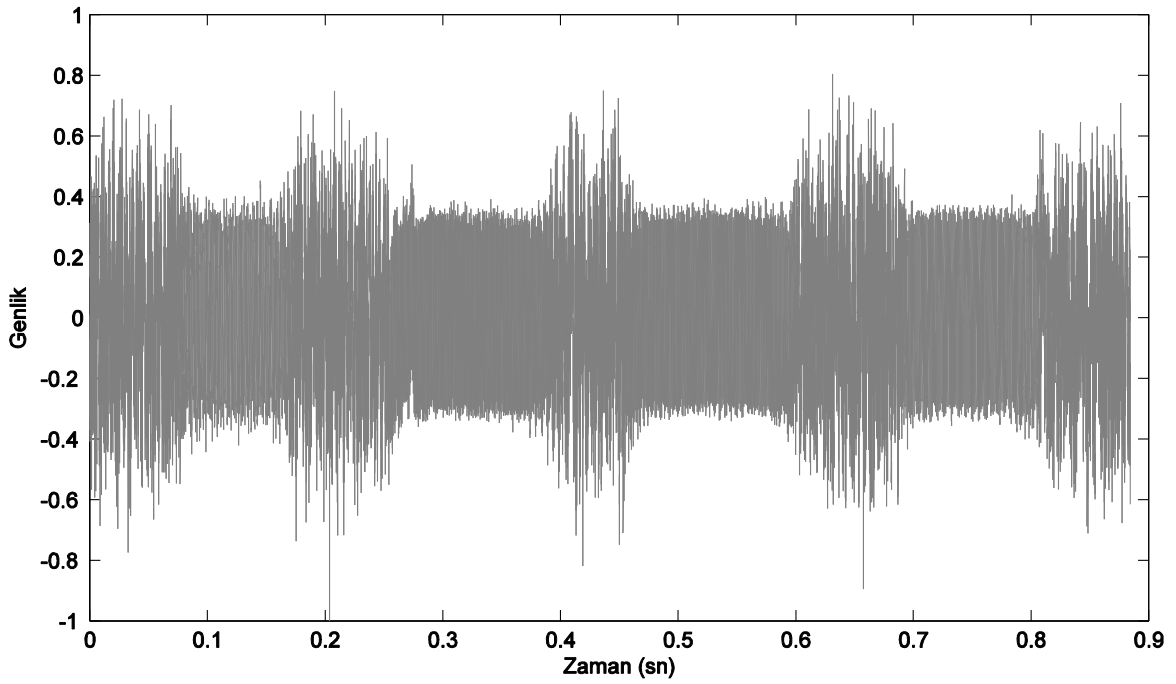
Şekil 35. Şekil 29’da verilen ses sinyalinin enerji düzeyleri ve eşik seviyesi



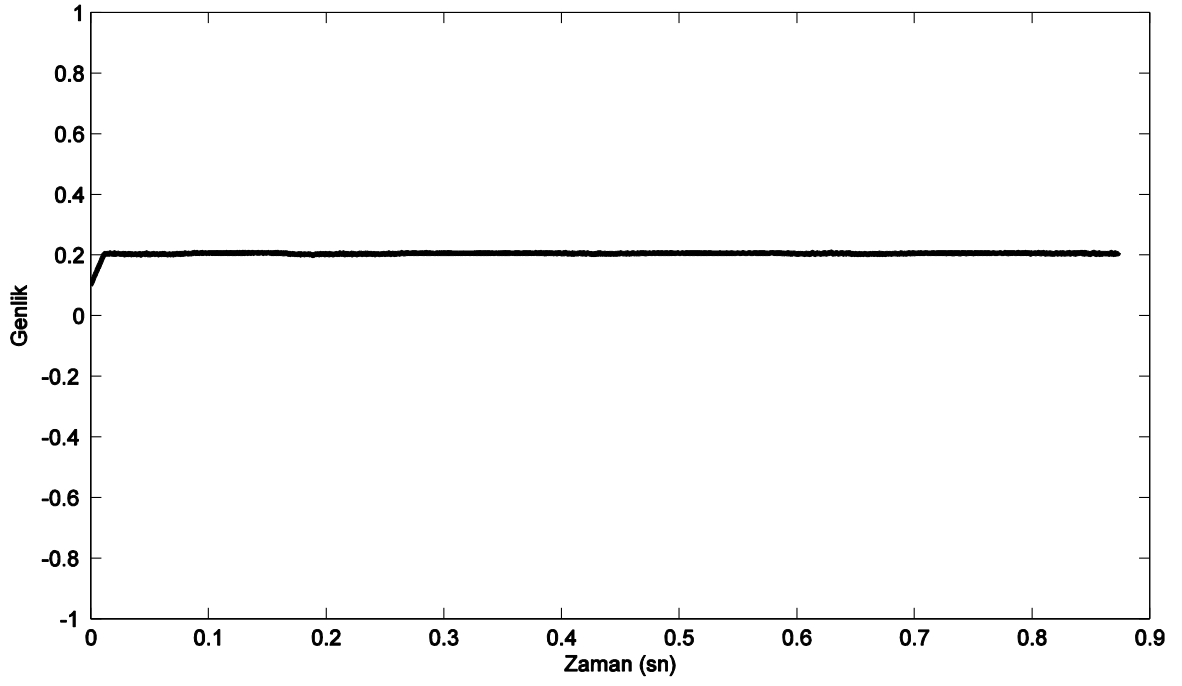
Şekil 36. Eşikleme yapılmış ses sinyali



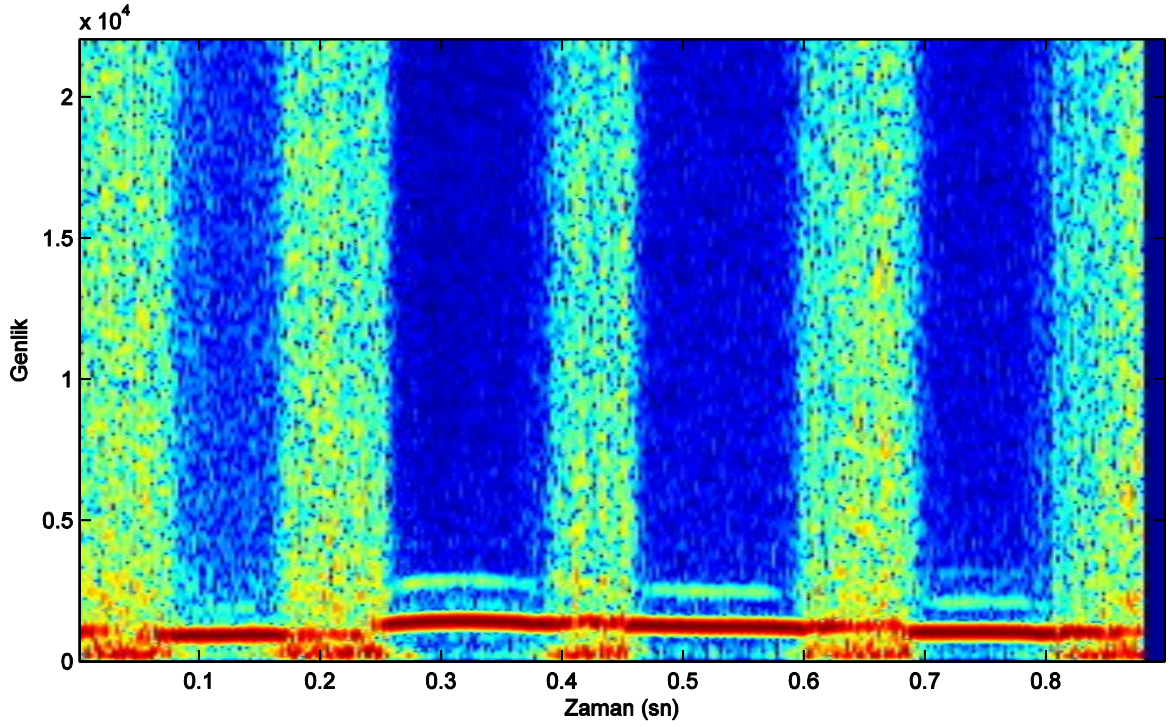
Şekil 37. Genlik eşitlemesi uygulanması



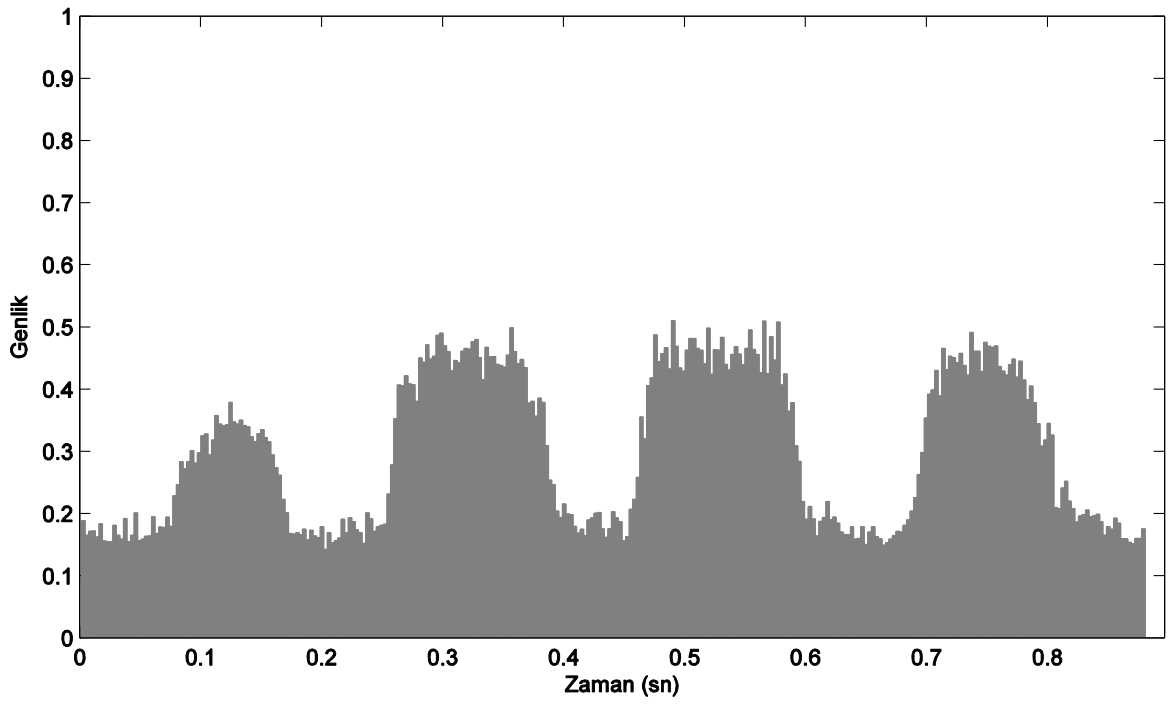
Şekil 38. Genlik eşitlemesi uygulanmış sinyali normalize edilmesi



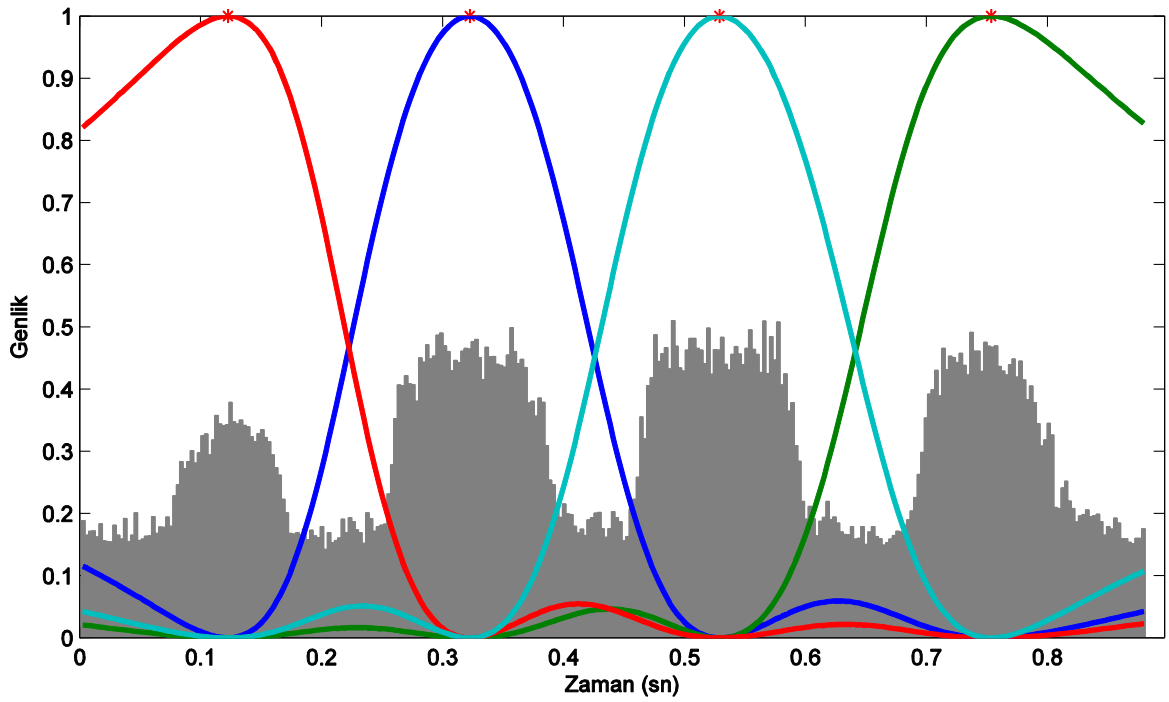
Şekil 39. Şekil 34’te verilen genlik eşitlemesi yapılmış sinyalin enerji düzeyi



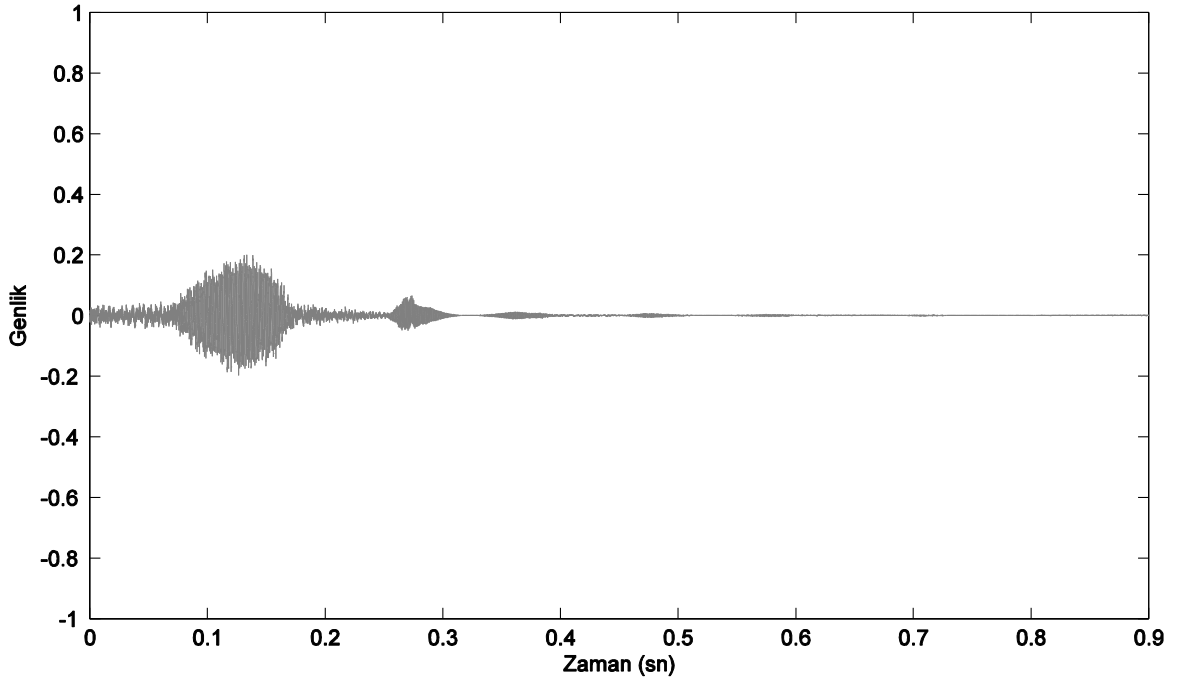
Şekil 40. Eşitlenmiş sinyalin spektrogramı



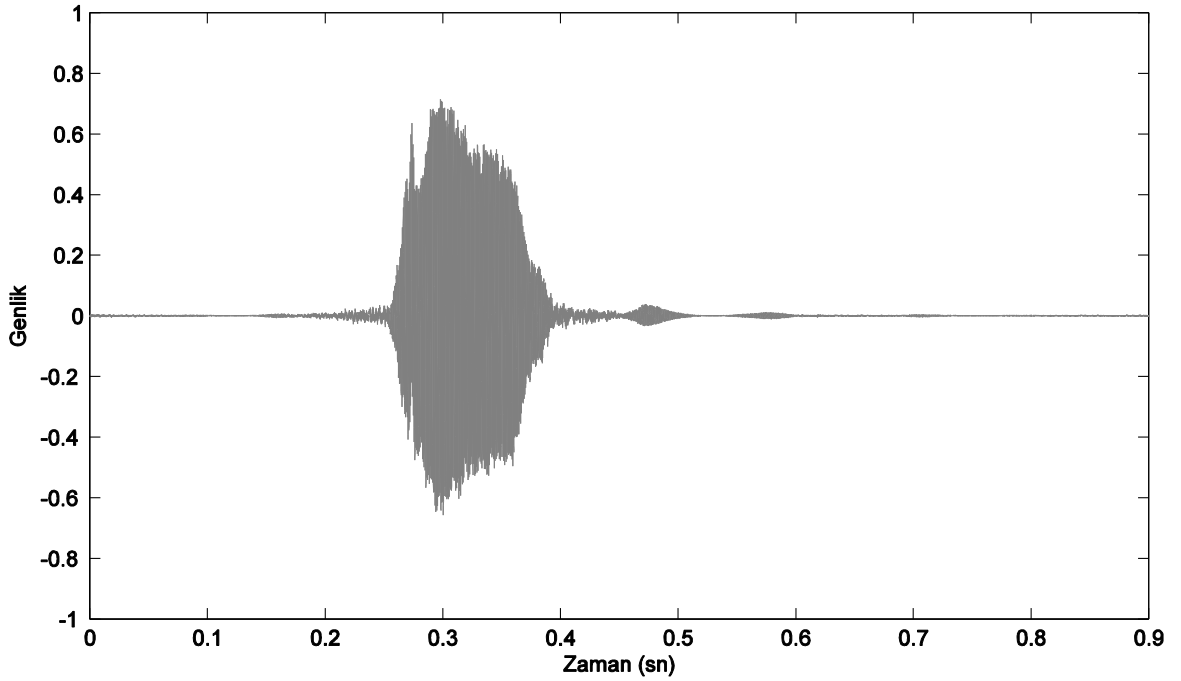
Şekil 41. Olabilirlik değerleri



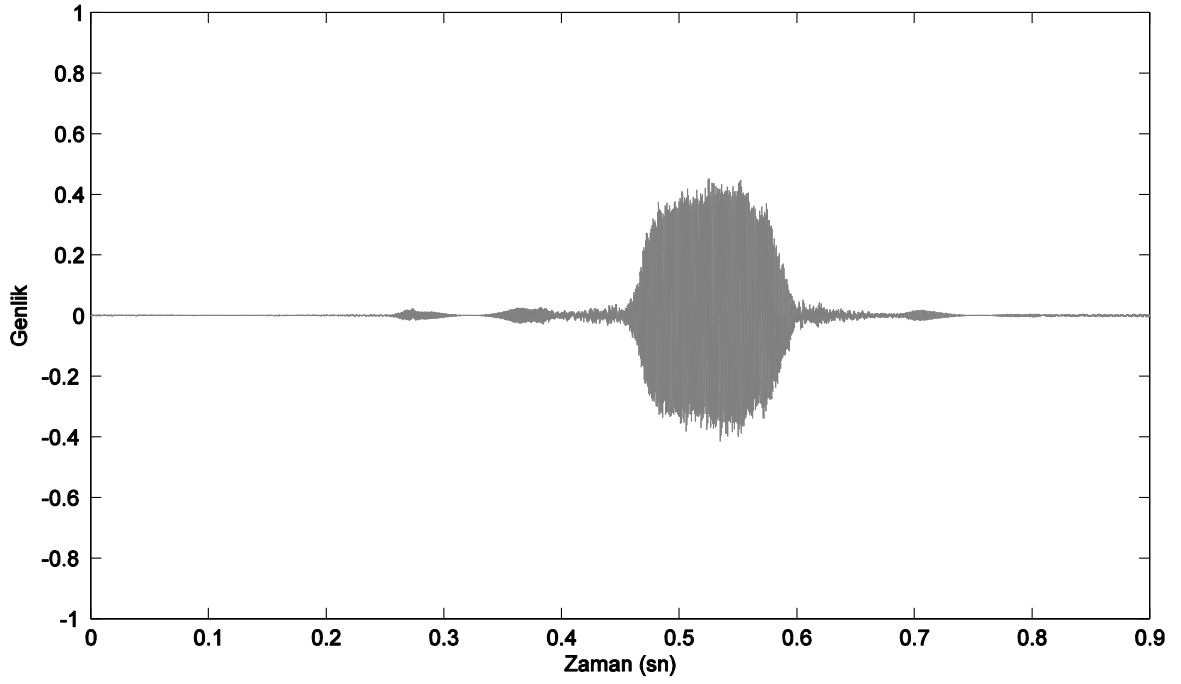
Şekil 42. Olabilirlik değerlerine göre sinyalin kümelenmesi



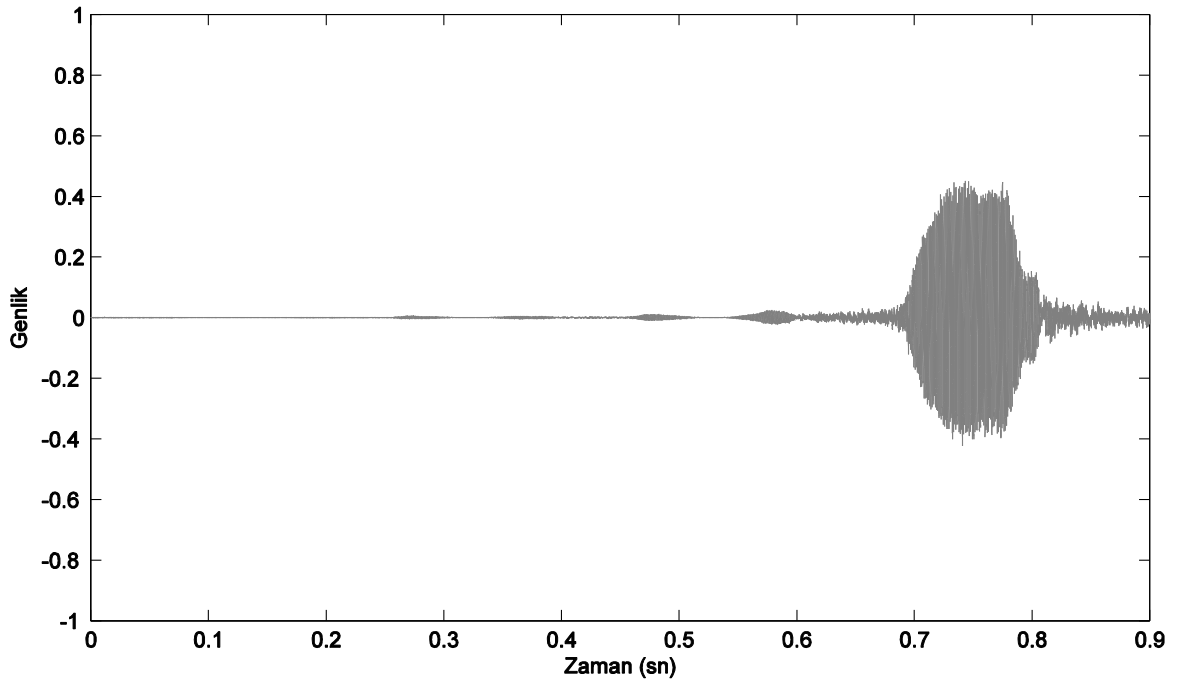
Şekil 43. Şekil 19’da verilen sinyal ile birinci kümenin üyelik değerlerinin çarpımıyla yapılan pencereleme sonucu elde edilen bulanık bölütlenmiş sinyal



Şekil 44. Şekil 19’da verilen sinyal ile birinci kümenin üyelik değerlerinin çarpımıyla yapılan pencereleme sonucu elde edilen bulanık bölütlenmiş sinyal



Şekil 45. Şekil 19’da verilen sinyal ile birinci kümenin üyelik değerlerinin çarpımıyla yapılan pencereleme sonucu elde edilen bulanık bölütlenmiş sinyal



Şekil 46. Şekil 19’da verilen sinyal ile birinci kümenin üyelik değerlerinin çarpımıyla yapılan pencereleme sonucu elde edilen bulanık bölütlenmiş sinyal

3. BULGULAR

İşlenebilir hale gelen ses sinyalinde kısa zamanlı enerji düzeyleri üzerinde eşikleme yaparak bölütler belirlenebilir. Ancak ses sinyalinin bazı enerji düzeyleri eşik değerlerin altında kaldığı için o bölgeler eşikleme sonucunda bölüte dahil edilmesi gereken bölgeler dışarıda kalabilmektedir. Şekil 21 yapılan eşikleme sonucunda dört seslilik bölgesi olmasına rağmen iki seslilik bölgesi ve Şekil 36'da yapılan eşikleme sonucunda ise dört seslilik bölgesi olmasına rağmen üç seslilik bölgesi bulunmuştur. Değişen enerji seviyesindeki bu tutarsızlık bölütleme işleminde sorunlara neden olmaktadır. Dolayısıyla sinyalin frekans içeriğine göre değerlendirme yapılsa bile aynı genlik değişkenliği frekans bandında da aynı düşük düzeyi göstereceğinden sorunun devam etmesini sağlamaktadır. Öte yandan önerilen yöntemde enerji düzeylerini dengelemek için genlik eşitleme kullanılmıştır. Ancak bu durumda eşiklemeye göre bölütleme sorunlu bir duruma düşmüştür. Bu durumda ise spektral ortamda sorun giderilebilmektedir.

Spektral ortamda bölütleme işleminde spektrumun doğrudan enerji düzeyiyle ilişki olması aynı sorunun burada da devam etmesine neden olacaktır. Oysa genlikteki bu dengesizlikler ortadan kaldırıldığında enerji seviyesine göre bölütleme yapılamasa da spektral ortamındaki bu sorun ortadan kaldırılmış olur.

Genliği eşitlenmiş ses sinyalinin spektrogramı incelendiğinde seslilik bölgelerindeki süreklilikler sessizlik bölgesinde de devam ettiği görülmektedir. Seslilik bölgeleri arasında değişik sessizlik bölgesinde yumuşak bir şekilde gerçekleşmektedir. Dolayısıyla her seslilik bölgesinin keskin sınırları sahip olmayıp bulanık bir geçiş göstermektedir.

Bu çalışmada, ele alınan her iki örnekte bölüm 1.10.3'te verilen tüm küme sayısı belirleme yöntemlerine göre optimal küme sayısının 4 olduğu görülmüştür.

4. ÖNERİLER

- Ses tanıma uygulamalarında, önerilen ağırlıklı bulanık c-ortalama yöntemiyle bölütlenmiş ses verileri kullanılabilir.
- Farklı ses sinyallerinde önerilen yöntem kullanılarak daha da geliştirilebilir.
- Sessizlik bölgelerindeki bulanık sürekliliği değerlendirmek için kullanılan entropik olabilirlik yaklaşımı yerine farklı olabilirlik yaklaşımları denenerek daha iyi sonuçların alınıp alınamayacağı bulunabilir.
- Bölütlenmiş ses sinyalleri farklı tanıma yöntemleri kullanılarak önerilen bölütleme işleminin diğer bölütleme yöntemlerine göre başarısı test edilebilir.

5. KAYNAKLAR

- Abdullah-Al-Mamun, K., Sarker, F., and Muhammad, G., 2009. A high resolution pitch detection algorithm based on AMDF and ACF. Journal of Scientific Research , 1,3, 508-515.
- Ajmera, J., McCowan, I., and Bourlard, H., 2003. Speech/music Segmentation Using Entropy and Dynamism Features in a HMM Classification Framework. Speech Communication, 40, 3, 351-363.
- Alpaslan, F., Erilli, N., Yolcu, U., Eğriođlu, E., and Aladađ, Ç., 2011. Bulanık Kümelemede En Uygun Küme Sayısının Yapay Sinir Ağları ve Diskriminant Analizi İle Belirlenmesi. Atatürk Üniversitesi İktisadi ve İdari Bilimler Dergisi, 25,10, 475-488.
- Atal, B. 1976. Automatic Recognition of Speakers from their Voices. 64, 4.
- Atalay, A., and Tortum, A., 2010. Türkiye'deki İllerin 1997-2006 Yılları Arası Trafik Kazalarına Göre Kümeleme Analizi. Pamukkale Üniversitesi Mühendislik Bilimleri Dergisi, 16(3), 335-343.
- Aucouturier, J., and Sandler, M., 2001. Segmentation of musical signals using hidden Markov models. Audio Engineering Society Convention, 110. Audio Engineering Society. Amsterdam.
- Azem, Z., 2003. A Comprehensive Cluster Validity Framework For Clustering Algorithms. MSc Thesis, Canada.
- Bezdek, J., 1974. Cluster validity with fuzzy sets. J. Cybernetics, 3, 3, 58-73.
- Bezdek, J. C., 1975. Mathematical models for systematics and taxonomy. Proceedings of eighth international conference on numerical taxonomy, 3.
- Bezdek, J. C., Ehrlich, R., and Full, W., 1984. FCM: The Fuzzy c-Means Clustering Algorithm. Computer & Geosciences, 10, 2-3, 191-203.
- Bolat, B., Küçük, Ü., ve Yıldırım, T., 2004. Aktif Öğrenen PNN ile Konuşma/Müzik Sınıflandırma. Akıllı Sistemlerde Yenilikler ve Uygulamaları Sempozyumu, Mayıs, İstanbul, 23-28.
- Cheng, G., Tan, H., and Chen , X.-M., 2005. Audio Segmentation Via the Similarity Measure of Audio Feature Vectors. 10, 5.
- Cooley, J. W., and Tukey, J. W., 1965. An algorithm for the machine calculation of complex Fourier series. Mathematics of computation, 19, 90, 297-301.

- Davis, S., and Mermelstein, P., 1980. Comparison of Parametric Representations for Monosyllabic Word Recognition in Continuously Spoken Sentences. IEEE Acousitcs, Speech, and Signal Processing, 28, 4, 357-366.
- Erilli, N., Tunç, T., Öner, Y., ve Yolcu, U. 2009. İllerin Sosyoekonomik Verilere Dayanarak Bulanık Kümeleme Analizi İle Sınıflandırılması. e-Journal of New World Sciences Academy (NWSA), 4, 1.
- Eskidere, Ö., ve Ertaş, F., 2009. Mel Frekansı Kepstrum Katsayılarındaki Değişimlerin Konuşmacı Tanımaya Etkisi. 14, 2.
- Foote, J., 1997. A Similarity Measure for Automatic Audio Classification.
- Gonzalez, R. C., and Woods, R. E., 2002. Digital Image Proceasing. Prentice Hall.
- Goodwin, M. M., and Laroche, J., 2004. A dynamic programming approach to audio segmentation and speech/music discrimination, . In Acoustics, Speech, and Signal Processing (ICASSP'04), 309.
- Güler, N., 2006. Bulanık Kümeleme Analizi ve Bulanık Modellemeye Uygulamaları. Yüksek Lisans Tezi, Muğla Üniversitesi, Fen Bilimleri Enstitüsü, Muğla.
- Hague, M. A., and Myon, K. J., 2013. An Enhances Fuzzy c-Means Algorithm for Audio Segmentation and Classification. Multimed Tools Application, 63, 485-500.
- Han, J., and Kamber, M. 2001. Data mining: concepts and technologies.
- Işık, M., ve Çamurcu, A., 2007. K-Means, K-Medoids ve Bulanık C-Means Algoritmalarının Uygulamalı Olarak Performanslarının Tespiti. İstanbul Ticaret Üniversitesi Fen Bilimleri Dergisi, 6, 11, 31-45.
- Karpuz, G., 2011. Melodik Işık Seslerindeki Notaların Otomatik Belirlenmesi. Yüksek Lisans Tezi, K.T.Ü., Fen Bilimleri Enstitüsü, Trabzon.
- Kılıç, İ., 2013. Genetik - LBG Algoritması ile Sayısal Görüntülerin Sıkıştırılması. C.B.Ü. Fen Bilimleri Dergisi, 9, 2, 45-53.
- Kinnunen, T., 2003. Spectral Features for Automatic Text-Independent Speaker Recognition.
- Kulkarni, A., Iyer, D., and Sridharan, S. R., 2001. Audio Segmentation. IEEE, International Conference on Data Mining, ICDM.
- MacQueen, J., 1967. Some Methods for Clssification and Analysis of Multivariate Observations. Proceeding of 5th Berkeley Symposium on Mathematical Statistics and Probability, University of California Press, 1, 281-297.
- Mandıracıoğlu, E., 2014. e-bergi. 07 14, 2014 tarihinde Ses Tanıma: <http://e-bergi.com/y/Ses-Tanima> adresinden alındı

- Masataka, G., 2003. A chorus-section detecting method for musical audio signals. Proceedings of the International Conference on Acoustics, Speech, and Signal Processing. IEEE, 5, 437-440.
- Nitanda, N., Haseyama, M., and Kitajima, H., 2006. An Audio Signal Segmentation and Classification Using Fuzzy c-means Clustering. Systems and Computers in Japan, 37, 4, 23-34.
- Ortakçı, Y., ve Göloğlu, C., 2012. Parçacık Sürü Optimizasyonu ile Küme Sayısının Belirlenmesi. Akademik Bilişim. Uşak.
- Otsu, N. A., 1979. A Threshold Selection Method from Gray-Level Histogram. IEEE Transactions on Systems, Man and Cybernetics, 9, 1, 62-66.
- Panagiotakis, C., and Tziritas, G., 2005. A Speech/Music Discriminator Based on RMS and Zero-crossing. 7, 1.
- Pang-Ning, T., Steinbach, M., and Kumar, V., 2006. Introduction to data mining. Library of Congress.
- Peiszer, E., Lidy, T., and Rauber, A., 2008. Automatic audio segmentation: Segment boundary and structure detection in popular music." . Proc. of LSAS.
- Rabiner, L. R., and Juang, B. H., 1993. Fundamentals of Speech Recognition. Englewood Cliffs: PTR Prentice Hall. <http://ai.amini.eu/tutorial/signal/signalFourierConcept.aspx>, 07.12.2014.
- Ross, M., Shaffer, H., Cohen, A., Freudberg, R., and Manley, H., 1974. Average magnitude difference function pitch extractor. IEEE Transactions on Acoustics, Speech and Signal Processing, 22, 5, 353-362.
- Sakhnov, K., Verteletskaia, E., and Simak, B., 2009. Approach for energy-based voice detector with adaptive scaling factor. IAENG International Journal of Computer Science, 36,4, 394.
- Ses Dalgası ve Sesin Özellikleri, <http://www.fenokulu.net> 07.10.2014.
- Shen, J.-l., Hung, J.-w., and Lee, L.-s., 1998. Robust Entropy-Based Endpoint Detection for Speech Recognition in Noisy Environments. International Conference on Spoken Language Processing. Sydney.
- Signal Fourier Concept., Tutorial Signal: <http://ai.amini.eu>, 08.05.2014.
- Spectrogram, <http://en.wikipedia.org/wiki/Spectrogram>, 27.07.2014.
- Stan, Z. L., 2000. Content-Based Audio Classification and Retrieval Using the Nearest Feature Line Method. 8, 5.
- Tzanetakis, G., and Cook, F., 1999. A Framework for Audio Analysis Based on Classification and Temporal Segmentation . 2.

- Wold, E., 1996. Content-Based Classification Search and Retrieval of Audio. 3, 3.
- Wu, B. F., and Wang, K.-C., 2005. Robust endpoint detection algorithm based on the adaptive band-partitioning spectral entropy in adverse environments. Speech and Audio Processing, IEEE Transactions on, 13, 5, 762-775.
- Xie, X. L., and Beni, G., 1991. A validity measure for fuzzy clustering. IEEE Transactions on pattern analysis and machine intelligence, 13, 8, 841-847.
- Yıldız, K., Çamurcu, Y., ve Doğan, B., 2010. Veri Madenciliğinde Temel Bileşenler Analizi ve Negatıfsız Matris Çarpanlarına Ayırma Tekniklerinin Karşılaştırmalı Analizi. Akademik Bilişim. Muğla.
- Ying, G. S., 1993. Endpoint detection of isolated utterances based on a modified Teager energy measurement. IEEE Acoustics, Speech, and Signal Processing, ICASSP-93, 732-735.
- Young, S., Woodland, P. C., and Byrne, W. J., 1993. HTK: Hidden Markov Model Toolkit V1. 5. Washington DC: Cambridge Univ. Eng. Dept. Speech Group and Entropic Research Lab. Inc.
- Zhang, J., Whalley, J., and Brooks, S., 2009. A Two Phase Method For General Audio Segmentation. IEEE International Conference on Multimedia and Expo, 626-629.

ÖZGEÇMİŞ

Nasim JABBARI, 20 Eylül 1987 tarihinde Urumiye (İran)'da doğdu. Lisans öğrenimini Urumiya Peyamnour Üniversitesinde, Fen Bilimleri Fakültesi, Uygulamalı İstatistik Bölümünü 2010 yılında tamamladı. 2012 yılında Karadeniz Teknik Üniversitesi, Fen Bilimler Enstitüsü, İstatistik ve Bilgisayar Anabilim dalında tezli yüksek lisans programına başladı. İngilizce, Farsça bilmektedir.