

MÜZİK VE KONUŞMA İŞARETLERİNİN DALGACIK ÖZNETELİKLERİ İLE SINIFLANDIRILMASI

Timur Düzenli¹

Nalan Özkurt²

^{1,2} Elektrik-Elektronik Mühendisliği Bölümü, Dokuz Eylül Üniversitesi, İzmir

¹e-posta: timurduzenli@gmail.com

² e-posta: nalan.ozkurt@deu.edu.tr

Özetçe

Bu çalışmada müzik ve konuşma seslerinin, farklı öznetelikler kullanılarak yapay sinir ağları ile sınıflandırılması gerçekleştirilmiştir. Dalgacık özneteliklerinden elde edilen başarımın karşılaştırılması için dört öznetelik grubuyla sınıflandırma yapılmıştır. İlk yöntemde zaman ve frekans ortamına ait öznetelikler ile beraber mel-kepstrum katsayıları kullanılmıştır. İkinci yöntemde, tamamı ayrık dalgacık dönüşümü (Discrete Wavelet Transform-DWT) tabanlı parametrelerden yararlanılmıştır. Üçüncü gruptaki öznetelikler, yine dalgacık dönüşümü sonucu elde edilen anlık ve teager enerji bileşenlerini içermektedir. Son yöntemde ise daha önceki çalışmalardan farklı olarak, karmaşık dalgacık dönüşümü (Complex Wavelet Transform-CWT) kullanılmıştır. Yapılan sınıflandırmalar sonucunda 3. ve 4. yöntemlerin diğer yöntemlere göre daha başarılı oldukları gözlemlenmiştir.

Anahtar Kelimeler: Müzik, Konuşma, Sınıflandırma, Ayrık ve Karmaşık Dalgacık Dönüşümü, Enerji

1. Giriş

Günümüzde konuşma ve müzik seslerinin ayrıştırılması büyük önem taşır hale gelmiştir. Örnek olarak, ses sıkıştırma tekniklerinin bazıları konuşma, bazıları ise müzik sesi için daha verimlidir. Gelen bir veriden konuşma ve müzik kısımları doğru şekilde belirlenebildiği takdirde buna uygun olan sıkıştırma tekniği uygulanabilmekte ve daha etkin sıkıştırma oranları elde edilebilmektedir. Aynı zamanda, radyo istasyonlarında da gönderilecek olan verinin konuşma veya müzik olarak belirlenmesi durumunda band genişliği daha etkin olarak kullanılabilir.

Konuşma ve müzik sesinin sınıflandırılması üzerine birçok çalışma yapılmıştır. Önceki çalışmalarda genel olarak zaman ve frekans ortamındaki temel özellikleri kullanan özneteliklerden yararlanılmıştır [1] [2]. Başka bir çalışmada

da Saunders [3] sıfır-geçişlerine dayanan bir konuşma-müzik ayrıştırıcısı önermiştir. Bu çalışma radyo yayınlarında, reklamlar ve programların ayrıştırılması için gereken bir uygulama için önerilmiştir. En güncel çalışmalardan birinde [4] ise, öznetelikler dalgacık dönüşümü kullanılarak elde edilmiştir. Burada, durağan olmayan sinyaller için zaman ve frekans özelliklerinin birleşiminden yararlanılmıştır.

Daha önceki çalışmalardan farklı olarak, yaptığımız çalışmada karmaşık dalgacık dönüşümü (CWT) kullanılmıştır. Bu dönüşümde, ayrık dalgacık dönüşümündeki (DWT) kaydırma etkisi ve yön bilgisinin azlığı gibi problemlerle karşılaşmamaktadır [5].

Bildirinin bundan sonraki bölümlerinde öznetelik çıkartımlarının anlatılmasının ardından, yapay sinir ağları tanıtılacak ve deneysel çalışmalar anlatılıp sonuçlar aktarılacaktır. Tartışma bölümünde ise başarımlar değerlendirilecektir.

2. Özneteliklerin çıkartılması

2.1. Zaman-frekans temelli parametreler ve mel-kepstrum katsayıları

İlk kullanılan yöntemde kullanılan öznetelik vektörü, sıfır geçişlerinin sayısı gibi zaman ortamından ve izgesel akış gibi frekans ortamından alınan bilgilerle oluşturulmuştur. Bunlara ek olarak, Mel frekansı kepstrum katsayılarından (MFCC-Mel Frequency Cepstrum Coefficients) yararlanılmıştır. Kullanılan öznetelik vektörünü oluşturan parametreler şu şekilde sıralanabilir [2]:

2.1.1 İzgesel merkez (Spectral Centroid)

İzgesel merkez adından da anlaşılacağı üzere sayısal işaret işlemede izgenin “kütle merkezi” ni bulmak için kullanılır ve

$$\frac{\sum_{n=0}^{N-1} f(n)x(n)}{\sum_{n=0}^{N-1} x(n)} \quad (1)$$

olarak verilir. Bu şekilde, genliğiyle ağırlıklandırılan işaretin ortalaması bize izgesel merkezi vermektedir. Denklemden, $f(n)$, izgesel dağılımdaki n . frekans değeri ve $x(n)$ ise, bu frekansa karşılık gelen izgesel genlik değeridir.

2.1.2 İzgesel düşüş (Spectral roll-off)

İzgesel şekli ifade eden önemli parametrelerden bir diğeri de izgesel düşüş noktasıdır. İzgesel düşüş noktası, izgenin genlik dağılımının %85 e ulaştığı frekans değeri (R_t) olarak kabul edilir:

$$\sum_{n=1}^{R_t} M_t[n] = 0.85 * \sum_{n=1}^N M_t[n] \quad (2)$$

Denklemden, M_t izgesel genlik dağılımını ifade eder ve R_t noktasındaki frekans değeri $f(R_t)$, izgesel düşüş noktası olarak kabul edilir.

2.1.3 İzgesel akı (Spectral flux)

İzgesel akı, izgenin şeklindeki değişimleri ifade eder. İşaretin her bölütü arasındaki izgesel fark, izgesel akı yardımıyla bulunur ve

$$F_t = \sum_{n=1}^N (N_t(n) - N_{t-1}(n))^2 \quad (3)$$

şeklinde verilir. Burada, N_t ve N_{t-1} , sırasıyla, t an üzerinde çalışılan bölütün izgesel dağılımı ile bir önceki bölütün izgesel dağılımını ifade eder. İzgesel akı bulunurken iki bölütteki bütün noktalar arasındaki fark hesaplanır ve bu farkların kareleri toplanır.

2.1.4 Sıfır geçişlerinin sayısı (Zero Crossings)

Zaman temelli bir özneliktir. Bir bölüt içerisinde gerçekleşen sıfır geçişlerinin sayısını ifade eder. Bu öznelik, işarettaki gürültü oranını ifade eden bir ölçüt olarak kullanılabilir. Sıfır geçişleri örnekler arasında işaret farkı olduğu zaman meydana gelir.

$$Z_t = 0.5 * \sum_{n=1}^N |\text{sgn}(x(n)) - \text{sgn}(x(n-1))| \quad (4)$$

2.1.5 Düşük etkin(RMS) enerji oranı (Low RMS Energy Ratio)

Düşük etkin enerji oranı, RMS enerjisi genel ortalamasının altında olan bölütlerin sayısını verir. Her bölütün enerjisi RMS değeri bulunarak hesap edilebilir.

$$x_{RMS} = \sqrt{\frac{1}{n} \sum_{i=1}^n x_i^2} = \sqrt{\frac{x_1^2 + x_2^2 + \dots + x_n^2}{n}} \quad (5)$$

Denklemden x_i , her bir frekans değeri için o bölütün izgesel dağılımındaki genliğe karşılık gelmektedir.

2.1.6 Mel frekans keppstrum katsayıları (Mel frequency cepstrum coefficients-MFCC)

Mel frekans keppstrum katsayıları, ses işaretinin; mel-frekans ölçeğinde ifade edilen kısa-zaman enerji izgesinin logaritması alındıktan sonra, ayrık kosinüs dönüşümü yapılması ile elde edilir[6]. Mel ölçeklendirme, insan kulağının işitsel özellikleri göz önüne alınarak oluşturulmuş bir ölçeklendirme şeklidir. Mel frekans keppstrum katsayıları hesaplanırken aşağıdaki şekilde bir yol izlenmektedir:

- (1) Pencerelemiş işaret, ayrık Fourier dönüşümü (DFT) ile frekans ortamına aktarılır ve her bölüt için anlamlı enerji $P(f)$ elde edilir.
- (2) Elde edilen güç izgesi $P(f)$, frekans ekseninden alınarak 6. denklem yardımıyla mel-frekans ölçeğine aktarılır[6]. Burada, $M(f)$ her bir frekans değerine karşılık gelen mel-frekans değeridir.

$$M(f) = 2595 * \log\left(1 + \frac{f}{700}\right) \quad (6)$$

- (3) $P(f)$, üçgen band-geçiren filtreler ($\psi(M)$) ile evrilerek, $\theta(M)$ katsayıları elde edilir. Denklemden M , mel frekans değerlerini ifade etmektedir.

$$\theta(M) = \sum_M P(M - M_k) \psi(M), \quad k=1 \dots K \quad (7)$$

Bu işlemler sonucunda K adet çıkış $X(k) = \ln(\theta(M_k))$, $k=1 \dots K$ şeklinde elde edilir.

- (4) Mel Frekans keppstrum katsayıları ise

$$MFCC(d) = \sum_{k=1}^K X_k \cos\left(\frac{d(k-0.5)\pi}{k}\right),$$

$d=1 \dots D$ (8)

denklemlerle elde edilir.

2.2 Ayrık Dalgacık Dönüşümü (Discrete Wavelet Transform-DWT)

Ayrık dalgacık dönüşümü; işaret işlemede, bilginin zaman-frekans ortamında tam olarak ifade edilebilmesi açısından sıklıkla tercih edilen bir yöntemdir[9]. Sürekli dalgacık dönüşümünden kolaylıkla türetilir. Verilen herhangi bir $x(t)$ işareti için sürekli dalgacık dönüşümü aşağıdaki gibi tanımlanır:

$$CWT(r, s) = \frac{1}{\sqrt{|s|}} \int x(t) \psi^* \left(\frac{t-r}{s} \right) dt \quad (9)$$

Burada, $\Psi(t)$ ana dalgacık olarak isimlendirilen zaman fonksiyonunu, r üzerinde işlem yapılan pencerenin zamansal olarak konumunu ve s ise ölçeği ifade eder. r ve s değerleri değiştirilerek ana dalgacık fonksiyonu kaydırılabilir ve ölçeklendirilebilir. Hesaplama yükünü azaltmak için ölçek ve kaydırma aralıkları tamsayıların katları olarak seçildiğinde, ayrık dalgacık dönüşümü; $x[m]$, $m=0..N-1$ olmak üzere,

$$DWT \left[n, 2^j \right] = \sum_{m=0}^{N-1} x[m] \psi_{2^j}^* [m-n] \quad (10)$$

şeklinde yapılır. Burada, $\psi_{2^j}^*$

$$\psi_{2^j}^* [n] = \frac{1}{\sqrt{2^j}} \psi \left(\frac{n}{2^j} \right) \quad (11)$$

olarak tanımlanır.

Ayrık dalgacık dönüşümünün önemli bir getirisi de farklı frekanslarda durağan olmayan güç bileşenleri içeren zaman serilerinde de etkili olmasıdır. Bu yöntemde; ilk adım, ana dalgacığın doğru bir şekilde seçilmesidir. Herhangi bir fonksiyonunun ana dalgacık kabul edilebilmesi için ortalamasının sıfır olması ve zaman-frekans ortamlarında sonlu enerjiye sahip olması gerekir. Sonraki adımda ise, bu ana dalgacık kullanılarak; işaret, alçak ve yüksek frekans bandlarına (yaklaşım-ayrıntı) ayrıştırılır. Bu işlem, ana dalgacığın kaydırılmış ve ölçeklendirilmiş versiyonlarıyla gerçekleştirilir. Birçok ana dalgacık türü olmakla beraber söz konusu çalışmamızda en başarılılardan biri olarak gösterilen[9] Daubechies dalgacık fonksiyonlarından faydalanılmıştır. Ayrık dalgacık dönüşümünün kullanıldığı diğer bir yöntem ise enerji temelli parametreleri içerir[4]. Bu yöntemde, dalgacık dönüşümünün sadece ayrıntı katsayılarından yararlanılır ve her band için anlık ve teager enerjileri şeklindeki gibi hesaplanır:

$$f_j^E = \log_{10} \left(\frac{1}{N_j} \sum_{r=1}^{N_j} (w_j(r))^2 \right) \quad (12)$$

$$f_j^{TE} = \log_{10} \left(\frac{1}{N_j} \sum_{r=1}^{N_j-1} (w_j(r))^2 - (w_j(r-1) * w_j(r+1)) \right) \quad (13)$$

Burada $w_j(r)$, r zamanı ve j frekans bandı için dalgacık katsayısını ifade etmektedir. N_j ise pencere uzunluğunu ifade etmektedir. Gerçekleştirilen çalışmada 5 band için ayrıntı katsayıları bulunmuş ve her band için anlık ve teager enerjileri hesaplanmıştır.

2. 3. Karmaşık dalgacık dönüşümü (Complex wavelet transform-CWT)

Dalgacık dönüşümünün, sinyallerin zaman-frekans ortamındaki ifadesinde önemli rolü olmakla beraber, kaydırmadan etkilenme, örtüşüm ve yön bilgisinin azlığı gibi yetersiz kaldığı noktalar da mevcuttur[5]. Karmaşık dalgacık dönüşümü(CWT), bu tür durumlara çözüm sunmaktadır. Bu dönüşümde, normal dalgacık dönüşümündekinden farklı olarak gerçel değerli ana dalgacık fonksiyonları yerine karmaşık fonksiyonlar kullanılır. Bu şekilde genlik ve faz bilgileri ayrı ayrı incelenebilir.

Karmaşık dalgacık fonksiyonu, gerçel ve sanal bileşenlerle beraber aşağıdaki şekilde ifade edilebilir:

$$\Psi_c(t) = \Psi_r(t) + j\Psi_i(t) \quad (14)$$

İşaretin üzerine iz düşümü gerçekleştirilerek, karmaşık dalgacık katsayısı

$$d_c(j, n) = d_r(j, n) + jd_i(j, n) \quad (15)$$

şeklinde ifade edilir. Buradan genlik,

$$|d_c(j, n)| = \sqrt{[d_r(j, n)]^2 + [d_i(j, n)]^2} \quad (16)$$

ve faz bilgisi,

$$\varphi_{d_c(j, n)} = \arctan \left(\frac{d_i(j, n)}{d_r(j, n)} \right) \quad (17)$$

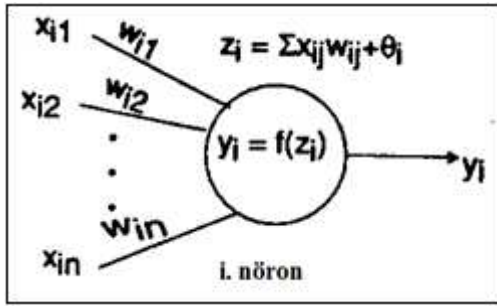
olarak hesaplanır^[5].

Ayrık karmaşık dalgacık dönüşümü, temel olarak iki yöntemde incelenir. İlk yöntemde, işaret için birimlik temel oluşturan $\Psi_c(t)$ dalgacığı aranır. Bu yöntemde, fonksiyonun geneli dikkate alınmakla beraber, diğer yöntemde ise bu işlem; ana dalgacığı oluşturan $\Psi_r(t)$ ve $j\Psi_i(t)$ bileşenleri kullanarak gerçekleştirilir. Bu yöntem çift ağaçlı dalgacık ayrıştırması (Dual-tree wavelet decomposition) olarak

isimlendirilir [5]. Yapılan çalışmada da bu yöntem tercih edilmiştir.

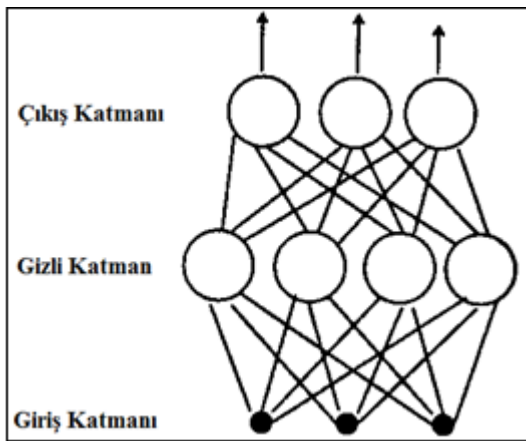
3. Yapay sinir ağları

Yapay sinir ağları, beyinsel hesaplamaları taklit ederek problemlere çözüm sunmayı amaçlayan araçlardan birisidir. Beyin, çok basit hesaplamalar gerçekleştiren küçük hesaplama birimlerinden (nöron) oluşur. Nöronlar, daha ağır problemler için, veriyi paralel olarak işleyebilen ağlar oluşturabilirler. Bir nöronun gerçekleştirdiği en basit işlem, genel olarak $y_i = f(z_i)$ şeklinde ifade edilebilir. Burada, z_i girişi, y_i i. nöronun çıkışı ve f ise doğrusal olmayan bir fonksiyonu ifade eder [10].



Şekil.1 Basit bir nöronun modellenmesi

Bir yapay sinir ağı; birinin çıkışı, diğerinin girişi olan birçok katman içerebilir. Yapay sinir ağları temel olarak; giriş, çıkış ve gizli katman şeklinde 3 ana bileşen içerir. Karşılaşılan problemlere göre bu katmanların ve bu katmanlardaki nöronların sayıları farklılık gösterebilir.



Şekil.2 Yapay sinir ağlarına genel bir bakış

Sınıflandırma yapılacak uygulamalarda genel olarak ileri beslemeli yapay sinir ağları tercih edilmektedir. İleri beslemeli ağlarda, veri akışı girişten çıkışa doğru tek yönde gerçekleşir. Bir yapay sinir

ağının tasarımında; kullanılacak katmanların sayısı, bu katmanlarda yer alacak nöron sayıları, her nöronda kullanılacak fonksiyonların tipi ve ağın çalışma şeklinin tanımlanması (ileri besleme-geri besleme) önemli rol oynar[10].

Tasarlanan ağın eğitimi, nöronlar arasındaki bağlantıları ifade eden ağırlıkların (w_{ij}) ve her nöron için eşik değerinin (θ_i) tanımlanması ile sağlanır. Bunun için, eğitim amacıyla hazırlanan örneklerden oluşan bir set girişe uygulanır. Nöronlar arasındaki ağırlıkların tanımlanması ($w_{ij}(t)$), uygulanan her eğitim girdisinden sonra, ağırlıkların Δw_{ij} kadar değiştirilmesiyle elde edilir:

$$w_{ij}(t) = w_{ij}(t-1) + \Delta w_{ij}(t) \quad (18)$$

Denklemden $\Delta w_{ij}(t)$, ağın başarımındaki hata oranını en aza indirecek ağırlık değerlerini veren değişim miktarıdır. Bu değişim miktarını bulmak için birçok yöntem mevcuttur, bunlardan en çok bilinenlerinden bir tanesi ise geri yayılım algoritmasıdır. Geri yayılım algoritması;

$$\Delta w_{ij}(t) = -\eta \left(\frac{\partial E}{\partial w_{ij}(t)} + \alpha \Delta w_{ij}(t-1) \right) \quad (19)$$

şeklinde verilir. Denklemden η , öğrenme hızı; α kazanç terimi, E ise istenen çıkışlar ile elde edilen çıkışlar arasındaki hata oranını ifade eder.

Yapılan deneysel çalışmalar sonucunda, tasarlanan ağda gizli katman için 40 nöron seçilmiş ve hedeflenen hata oranı 0.001 olarak belirlenmiştir. Öğrenme algoritması olarak, geri yayılım kullanan eşlenik gradient algoritması tercih edilmiştir.

4. Deneysel Çalışmalar ve Sonuçlar

Yapılan çalışmada, müzik sesleri için Iowa Üniversitesi veritabanından ve konuşma sesleri için TIMIT veritabanından faydalanılmıştır. Kullanılan her yöntemde, yapay sinir ağlarının eğitimi için 78 adet müzik ve konuşma sesi kullanılmıştır. Test aşamasında ise, eğitimden setinden farklı olarak yine 78 adet müzik sesi ve 100 adet konuşma sesi kullanılmıştır. Bütün datalar 4196 uzunlukta, 512 örnek örtüşümlü olarak bölütlere ayrılmıştır. İlk yöntemde, her bölüt için çıkarılan zaman-frekans tabanlı parametrelerin ortalama ve varyans değerleri ile Mel frekansı keştrüm katsayıları kullanılarak 21 parametre içeren öznitelik vektörü oluşturulmuştur. 2. ve 3. yöntemlerde ise ayrık dalgacık dönüşümü kullanılmıştır. Ayrık dalgacık dönüşümü kullanan bu

iki yöntemden ilkinde ana dalgacık olarak daubechies8 fonksiyonu kullanılmış, dalgacık dönüşümü ise 12 bantta gerçekleştirilmiştir. Dolayısıyla her bölüt için 12 ayrıntı ve 1 yaklaşım katsayısı elde edilmiştir. Öznitelik vektörü oluşturulurken, bölütlerin ayrıntı ve yaklaşım katsayıları arasındaki ortalama, varyans ve standart sapma gibi değerler kullanılmıştır. Bu kısımda, her bir veri için elde edilen öznitelik vektörünün uzunluğu 38 dir. Ayrıntı dalgacık dönüşümünün kullanıldığı diğer yöntemde, sadece anlık ve teager enerjisinden yararlanılmıştır. Bu yöntemde, her bölüt için dalgacık dönüşümü gerçekleştirildikten sonra her band için ayrıntı katsayıları üzerinden anlık ve teager enerjiler hesaplanmaktadır. Bu yöntem için de dalgacık fonksiyonu olarak yine daubechies8 kullanılmıştır. Yapılan çalışmada, bu yöntem için ayrıştırma 5 band olarak gerçekleştirilmiştir ve elde edilen öznitelik vektörünün uzunluğu her veri için 10' dur. Son yöntem olarak ise karmaşık dalgacık dönüşümü kullanılmıştır. Bu yöntemde, 5 band olarak ayrıştırma gerçekleştirilmiş ve daha yüksek bandlarda başarımın çok etkilenmediği görülmüştür. Bulunan gerçel ve karmaşık katsayılar için ayrı ayrı ortalama, medyan, standart sapma ve enerji değerleri hesaplanmıştır. Oluşturulan öznitelik vektörünün uzunluğu 5 band için 40 tır. Elde edilen başarılar; Tablo.1, Tablo.2 ve Tablo.3 de verilmektedir. Tek bir işaret üzerinden öznitelik çıkartımı ve sınıflandırma için geçen süre Tablo.3' de genel başarı ile beraber verilmiştir. Tablolarda;

DP(Doğru Pozitif), Konuşma olarak sınıflandırılan konuşma seslerini;

DN(Doğru Negatif), Müzik olarak sınıflandırılan müzik seslerini;

YP(Yanlış Pozitif), Konuşma olarak sınıflandırılan müzik seslerini;

YN(Yanlış Negatif), Müzik olarak sınıflandırılan konuşma seslerini ifade etmektedir.

Başarılar; konuşma sesleri için,

$$Başarı1 = \frac{DP}{DP + YN} \quad (20)$$

müzik sesleri için,

$$Başarı2 = \frac{DN}{DN + YP} \quad (21)$$

genel başarı için ise

$$GenelBaşarı = \frac{DN + DP}{DP + DN + YP + YN} \quad (22)$$

olarak verilmektedir.

	DN	YP	BAŞARI (%)
Zaman-Frekans +MFCC	67	11	87
Dalgacık Dönüşümü	77	1	98.7
Anlık ve Teager Enerji	78	0	100
Karmaşık Dalgacık Dönüşümü	75	3	96.15

Tablo.1: Müzik sesleri için başarı oranları

	DP	YN	BAŞARI (%)
Zaman-Frekans +MFCC	96	4	96
Dalgacık Dönüşümü	87	13	87
Anlık ve Teager Enerji	98	2	98
Karmaşık Dalgacık Dönüşümü	99	1	99

Tablo.2: Konuşma sesleri için başarı oranları

	Hesaplama Süresi (sn)	Genel Başarı (%)
Zaman-Frekans + MFCC	1.02	91.5
Dalgacık Dönüşümü	1.03	92.1
Anlık ve Teager Enerji	0.33	98.8
Karmaşık Dalgacık Dönüşümü	0.18	97.7

Tablo.3: Genel başarı ve tek bir işaret için hesaplama süresi

5. Tartışma

Gerçekleştirilen çalışmada, müzik ve konuşma seslerinin sınıflandırılması birçok yönden ele alınmıştır. Dalgacık tabanlı anlık ve teager enerji öznitelikleri ile karmaşık dalgacık dönüşümünden elde edilen özniteliklerin sınıflandırmayı daha iyi sonuçlandırdığı görülmüştür. Tablo.3' de görüldüğü gibi, karmaşık dalgacık dönüşümü anlık ve teager enerji özniteliklerine göre, hesaplama açısından daha hızlı sonuç vermekte; aynı zamanda yakın bir başarı oranı göstermektedir. Bu yüzden de gerçek zamanlı uygulamalarda kullanımı uygundur. Ayrıca, dalgacık dönüşümünden elde edilen öznitelikler, verinin özelliklerine bağımlı olduğu için farklı veri setlerinde

sonular deęiřebilmektedir. Gelecek alıřmalarda, gzlenen bařarıların kararlılıęı aısından, yapılan alıřma, daha geniř ve farklı bir veritabanı ile gerekleřtirilecektir.

6. Kaynaka

- [1] E.Scheier and M. Slaney, "Construction and Evaluation of a Robust Multifeature Speech/Music Discriminator", Proc.ICASSP 1997, say. 1331-1334.
- [2] E.M.Saad, M.I. El-Adawy, M.E.Abu-El-Wafa and A.A.Wahba, "A Multifeature Speech/Music Discrimination System", 19. Natioanal Radio Science Conference, Alexendria, 2002
- [3] John Saunders."Real Time Discrimination of Broadcast Speech/Music", Proc. ICASSP 1996, say. 993-996.
- [4]Didiot,E.et al., "A Wavelet-Based Parameterization for Speech /Music Discrimination", Computer Speech and Language(2009), doi:10.1016/j.csl.2009.05.03
- [5] Ivan W. Selesnick, Richard G. Baraniuk, and Nick G. Kingsbury, "The Dual-Tree Complex Wavelet Transform", IEEE Signal Processing Magazine, Kasım 2005, say. 123-151
- [6] Fand Zheng, Guoliang Zhang, Zhanjiang Song, " Comparison of Different Implemantations of MFCC", J. Computer Science & Technology, 16(6): 582-589, Eylil 2001
- [9] Stavros Ntalampiras, Nikos Fakotakis,"Speech /Music Discrimination Based on Discrete Wavelet Transform",SETN 2008, LNAI 5138, say. 205-211, 2008
- [10] Asim Roy, "Artificial Neural Networks- A Science in Trouble", SIGKDD Explorations, 1.Sayı, Ocak 2000, say. 33-38