Çukurova Üniversitesi Mühendislik Mimarlık Fakültesi Dergisi, 30(2), 103-113 ss., Aralık 2015
Çukurova University Journal of the Faculty of Engineering and Architecture, 30(2), pp. 103-113 , December 2015

Konuşma/Müzik Ayrıştırması için Kesikli Dalgacık Dönüşümü Tabanlı Öznitelik Çıkarımı

Timur DÜZENLİ[*1], Hatice DOĞAN[2], Nalan ÖZKURT[3]

[1]*Dokuz Eylül Üniversitesi, Mühendislik Fakültesi, Elektrik-Elektronik Müh. Bölümü, İzmir*
[2] *Dokuz Eylül Üniversitesi, Mühendislik Fakültesi, Elektrik-Elektronik Müh. Bölümü, İzmir*
[3]*Yaşar Üniversitesi, Mühendislik Fakültesi, Elektrik-Elektronik Müh. Bölümü, İzmir*

Geliş tarihi:08.09.2015 *Kabul tarihi:30.12.2015*

Özet

Bu çalışmada konuşma ve müzik işaretlerinin birbirinden ayrıştırılabilmesi için kesikli dalgacık dönüşümü tabanlı bir öznitelik seti önerilmiştir. Öznitelik setinde dalgacık katsayılarının ortalamaları, varyansları ve altbandlar arası değişim oranları kullanılmıştır. Dalgacık dönüşümünün sinyalleri iyi ifade edebilmesi sayesinde, 0,5 saniyelik pencerelerde dahi yüksek doğruluklu bir sınıflandırma sağlanabilmiştir. Veri seti olarak internet radyolarından kaydedilmiş çeşitli bayan-erkek konuşmaları ve farklı türlerden müzik işaretleri kullanılmıştır. Daubechies-8 dalgacığının yok etme moment sayısı ve dikgenliği dikkate alındığında bu ailenin diğer üyeleri arasında en iyi performansa sahip olduğu gözlenmiştir. Öznitelikler çıkarıldıktan sonra, ilintili öznitelikleri yok etmek için temel bileşen analizi kullanılmıştır. Sınıflandırma hem yapay sinir ağları hem de destek vektör makineleri ile yapılmış ve önerilen özniteliklerin, klasik özniteliklerden çok daha iyi performans gösterdiği gözlenmiştir.

Anahtar Kelimeler : Konuşma/müzik ayrıştırma, Ayrık dalgacık dönüşümü, Destek vektör makineleri

Discrete Wavelet Transform Based Feature Extraction for Speech/Music Discrimination

Abstract

In this study, a discrete wavelet transform based feature set has been proposed for discrimination of music and speech. The feature set is constructed using the mean and variances of discrete wavelet coefficients and ratio of the change between the wavelet subbands. Due to the good representation ability of the wavelets, a high accuracy classification can be obtained even for a short window of 0,5 seconds.

[*] Yazışmaların yapılacağı yazar: Timur Düzenli, *Dokuz Eylül Üniversitesi, Mühendislik Fakültesi, Elektrik-Elektronik Müh. Bölümü, İzmir. timur.duzenli@deu.edu.tr*

A database which contains a wide variety of radio recordings from internet radios with different male and female speakers and various genres of musical pieces is constructed. The best performance is obtained with Daubechies-8 wavelet among the other members of the Daubechies family, considering the number of vanishing moments and orthogonality. The principal component analysis has been applied to eliminate the correlated features. The classification has been accomplished using both artificial neural networks and support vector machines and according to the results the proposed feature set outperforms the traditional ones.

Keywords: Speech/music discrimination, Discrete wavelet transform, Support vector machines

1. GİRİŞ

Konuşma ve müzik işaretlerinin ayrıştırılması (KMA), ses işleme uygulamaları arasında özellikle çoklu ortam verilerinin artışıyla önem kazanmaya başlamış bir alandır. Farklı spektrum ve zaman karakteristiklerine sahip konuşma ve müzik işaretlerine uygulanacak sıkıştırma algoritmaları farklı olabileceği gibi, bu iki işaretin bant genişliği ihtiyaçları da farklıdır. Ayrıca KMA algoritmalarına ses ve/veya konuşmacı tanıma uygulamalarında önişlemci olarak ihtiyaç duyulmaktadır. Bu nedenlerden ötürü bu iki tip verinin sınıflandırılması bir problem olarak karşımıza çıkmaktadır [1-4]. Konuşma/müzik ayrıştırmasında aşılması gereken zorluklar; uygun uzunlukta pencerenin belirlenmesi, sınıflandırmaya en çok katkı yapabilecek ve hesaplanması kolay özniteliklerin seçilmesi ve uygun sınıflandırıcının kullanılmasıdır. Pencere uzunluğu, özniteliklerin sınıflandırma performansını doğrudan etkilerken fazla uzun pencereler hesaplama süresinin uzamasına sebep olacak, dolayısıyla gerçek zamanlı uygulama yapılmasını zorlaştıracaktır. Bununla bağlantılı olarak öznitelik seçimi konusunda birçok çalışma yapılmıştır. Bu çalışmalarda sıfır geçişlerinin sayısı [5]; 4 Hz modülasyon enerjisi, düşük enerji oranı, izgesel merkez, izgesel düşüş noktası ve izgesel akış gibi özniteliklerden oluşan bir set [6]; entropi ve dinamizm öznitelikleri [7], efektif değer tabanlı sıfır geçişlerinin ortalama yoğunluğu ve ortalama frekans [8]; harmonik özellikler [9]; izgesel entropi [10]; süzgeç bankası çıkışlarının ortalaması ve varyansları [11] gibi öznitelikler farklı sınıflandırıcılarda kullanılmıştır.

Müzik ve konuşma işaretlerinin durağan olmayan frekans karakteristiklerine sahip olması nedeniyle birçok çalışmada ayrıştırma amacıyla zaman-frekans dönüşümleri de kullanılmıştır [12-15]. Farklı ana dalgacıklar ve dalgacık katsayılarının ortalaması, varyansı ya da bu ortamda hesaplanan enerji katsayılarından oluşturulmuş öznitelikler kullanılarak, 20ms'den 3 saniyeye kadar farklı zaman penceresi uzunluklarıyla sınıflandırmalar yapılmış ve zaman ortamı, frekans ortamı ya da Mel frekansı kepstrumu öznitelikleri ile karşılaştırmalar yapılmıştır [15]. Diğer bir çalışmada ise ayrık dalgacık dönüşümleri ve karmaşık dalgacık dönüşümlerinden elde edilen öznitelikler yapay sinir ağları kullanılarak sınıflandırılmış ve literatürde sıklıkla kullanılan özniteliklerle sınıflandırma performansı karşılaştırılmış [16], önerilen özniteliklerin gerçek zamanlı uygulaması geliştirilmiştir [17]. Bu çalışmalarda Daubechies dalgacıkları ile yapılan ayrık dalgacık dönüşümünün en iyi performansı gösterdiği gözlenmiştir.

Bu çalışmada, önceki çalışmalardan farklı olarak, örnekleme penceresi uzunluğunun, seçilen temel dalgacıkların ve bu dalgacıklardan oluşturulan özniteliklerin sınıflandırma başarımına etkisi, yapay sinir ağları (YSA) ve destek vektör makineleri (DVM) gibi farklı sınıflandırıcılar kullanılarak daha ayrıntılı şekilde incelenmektedir. Bu amaçla farklı örnekleme penceresi uzunlukları, farklı ana dalgacıklar ve dalgacık katsayılarından elde edilmiş farklı öznitelik setleri önce temel bileşen analizi ile incelenmiş, daha sonra bu özniteliklerle yapılan sınıflandırmalar klasik özniteliklerin performansı ile karşılaştırılmıştır. Sınıflandırma için yapay sinir ağlarının yanı sıra

destek vektör makineleri de kullanılmıştır. Veri seti internet radyolarından çeşitli bayan ve erkek konuşmacıların sesleri ve farklı türlerden müzik parçalarının kaydedilmesi ile oluşturulmuştur.

2. KONUŞMA/MÜZİK AYRIŞTIRMA (KMA) İÇİN KULLANILAN ÖZNİTELİKLER

Bu kısımda sırasıyla, KMA sistemlerinde genel olarak kullanılan öznitelikler ve bu çalışmada önerilen kesikli dalgacık dönüşümü (KDD) tabanlı öznitelik çıkarım yöntemleri anlatılmaktadır.

2.1. KMA Sistemlerinde Genel Olarak Kullanılan Öznitelikler

Sıfır geçişlerinin sayısı gibi zaman ortamından elde edilen bilgiler ile düşük enerji oranı, izgesel merkez, izgesel düşüş noktası ve izgesel akış gibi frekans ortamından alınan bilgiler KMA sistemlerinde sıklıkla kullanılmaktadır [6]. Bunlara ek olarak, Mel frekansı kepstrum katsayılarının da konuşma ve müzik seslerinin sınıflandırılmasında başarılı olduğu görülmektedir [8]. Önerilen yöntemin karşılaştırılması amacıyla, literatürde genel olarak kullanılan bu yöntemlerden bir öznitelik vektörü oluşturulmuştur. Kullanılan öznitelik vektörünü oluşturan parametreler şu şekilde sıralanmaktadır:

2.1.1. Sıfır Geçişlerinin Sayısı

Zaman düzleminde çıkartılan bir öznitelik olup, bir bölüt içerisinde gerçekleşen sıfır geçişlerinin sayısını ifade etmektedir. İşaretteki baskın frekansın bir göstergesi olduğundan, müzik ve konuşma ayrımında kullanılan önemli özniteliklerden birisidir [6]. Sıfır geçişlerinin sayısı, her bölüt için

$$SGS_i = \frac{1}{2}\sum_{n=2}^{N}\left|\text{sign}\left(x[n]\right)-\text{sign}\left(x[n-1]\right)\right| \qquad (1)$$

şeklinde hesaplanabilmektedir. Denklemde x[n], işaretin N uzunluğundaki i. bölütünün n. elemanını ifade etmektedir. "sign" ise işaret fonksiyonudur.

2.1.2. Düşük Enerji Oranı

Etkin ya da ortalama karekök enerjisi, işaretin ortalama enerjisinden düşük olan bölütlerin sayısını vermektedir. Ortalama karekök enerjisi her bölüt için şu şekilde tanımlanmaktadır:

$$X_{OKE} = \sqrt{\frac{1}{K}\sum_{k=1}^{K} X_k^2} \qquad (2)$$

Denklem (2)'de, X_k, k. frekans bileşeninin genliğine karşılık gelmektedir. Konuşma sesleri için enerji dağılımı müzik seslerine göre sola eğimli olduğundan, bu özniteliğin değeri konuşma sesleri için daha büyük değerler almaktadır.

2.1.3. İzgesel Merkez

İzgesel merkez, adından da anlaşılacağı üzere, frekans izgesinin "kütle merkezi" olarak tanımlanmaktadır ve

$$\dot{I}M = \frac{\sum_{k=1}^{K} f_k X_k}{\sum_{k=1}^{K} X_k} \qquad (3)$$

şeklinde hesaplanmaktadır. Denklemde, f_k, izgesel dağılımdaki k. frekans değeri ve X_k ise, bu frekansa karşılık gelen izgesel genlik değeridir [6].

2.1.4. İzgesel Düşüş Noktası

Bu öznitelik, frekans izgesinin şeklini belirlemek amacıyla kullanılan önemli bir özniteliktir. İzgesel düşüş noktası R_k, izgesel gücün %95 ini içeren frekans sınırı olarak tanımlanmakta ve aşağıdaki şekilde hesaplanmaktadır:

$$\sum_{k=1}^{R_K} X_k^2 = 0{,}95\sum_{k=1}^{K} X_k^2 \qquad (4)$$

Konuşma sesleri için enerji dağılımı sola eğimli olduğundan bu özniteliğin değeri konuşma sesleri için daha düşüktür [6].

2.1.5. İzgesel Akı

Komşu bölütler arasındaki izgesel değişimleri ifade etmektedir. İşaretin her bölütü arasındaki izgesel fark, izgesel akı yardımıyla bulunur ve aşağıdaki şekilde hesaplanabilir:

$$\dot{I}A = \sum_{k=1}^{K} \left(X_k^h - X_k^{h-1} \right)^2 \quad (5)$$

Denklem (5)'te, X_k^h ve X_k^{h-1}, sırasıyla, o 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

2.1.6. Mel Frekansı Kepstrum Katsayıları

Mel frekansı kepstrum katsayıları, ses işaretinin; mel-frekansı ölçeğinde ifade edilen kısa-zaman enerji izgesinin logaritması alındıktan sonra, ayrık kosinüs dönüşümü uygulanması ile elde edilir [20].

Mel ölçeklendirme, frekans aralıklarının doğrusal olmayan şekilde yapılandığı insan kulağının işitsel özellikleri göz önüne alınarak oluşturulmuş bir ölçeklendirme şeklidir. Bu sayede sesler daha iyi ifade edilebilmektedir.

Yukarıda anlatılmış olan özniteliklerin ortalama ve varyans değerleri ile bunlara ek olarak 12 Mel Kepstral katsayısının kullanılması sonucunda oluşturulan öznitelik vektörü 21 elemanlıdır. Literatürde bu öznitelikler, pencerelenen işaretin durağanlığını koruyabilmesi amacıyla, 11025-44100 Hz frekans aralığında örneklenen yaklaşık 20 ms' lik bir bölüt uzunluğu kullanılarak hesaplanmaktadır [8]. Ancak bu uzunluk, işaretlerin ayrıştırıcı özelliklerini sınırlayacak şekilde kısa olabilmektedir. Bu yüzden, bu çalışmada, öznitelikler %12,2'si (512 örneği) örtüşen 4196 örnek uzunluğunda bölütler üzerinden hesaplanmıştır. Bu da 95 ms'lik bir bölüt uzunluğuna karşılık gelmektedir. Karşılaştırma yapılması amacıyla, KMA için öznitelik çıkartımında her iki bölüt uzunluğu da (20 ms ve 95 ms) kullanılmıştır. Yapılan deneysel çalışmalar sonucunda, seçilen bölüt uzunluğunun (95 ms) %2,4 oranında performans iyileşmesi sağladığı görülmüştür.

2.2. Kesikli dalgacık dönüşümü (KDD) temelli önerilen öznitelikler

Çoklu çözünürlük analizi, durağan olmayan işaretler için uygun bir zaman- frekans gösterimi sağlamaktadır. Bu yöntem sayesinde işaretler, dalgacık ve ölçeklendirme fonksiyonlarının gerdiği bir uzayda, farklı frekans ve çözünürlüklerde ayrıştırılıp analiz edilebilmektedirler. Dalgacık dönüşümü aşağıdaki şekilde, işaretin bu uzaydaki izdüşümünün hesaplanmasıyla gerçekleştirilir:

$$DT(s,r) = \frac{1}{\sqrt{s}} \int x(t) \psi^* \left(\frac{t-r}{s} \right) dt \quad (6)$$

Denklem (6)'da, $\psi(t)$ ana dalgacık olup, *s* ve *r* katsayıları ise, sırasıyla ölçek ve kaydırma katsayılarıdır [18]. r ve s değerleri değiştirilerek ana dalgacık fonksiyonu kaydırılabilir ve ölçeklendirilebilir. Hesaplama yükünü azaltmak için ölçek ve kaydırma aralıkları tamsayıların katları olarak seçildiğinde, kesikli dalgacık dönüşümü (KDD); x[m], m=0…N-1 olmak üzere,

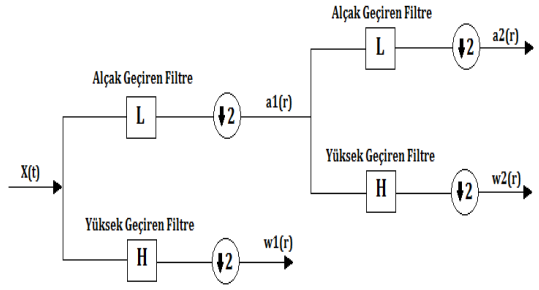
$$KDD\left[n, 2^j \right] = \sum_{m=0}^{N-1} x[m] \psi_{2^j}^* [m-n] \quad (7)$$

şeklinde verilmekte olup, 2^j ölçekleme katsayısını ifade etmektedir. Burada $\psi_{2^j}[n]$,

$$\psi_{2^j}[n] = \frac{1}{\sqrt{2^j}} \psi \left(\frac{n}{2^j} \right) \quad (8)$$

olarak hesaplanmaktadır.

Zaman ortamında, alçak ve yüksek geçiren süzgeçlerden oluşan bir süzgeç bankası kullanılarak KDD katsayılarının elde edilmesi mümkündür [13], [18]. Bu yaklaşım Şekil 1'de gösterilmektedir.



Şekil 1. İki ayrıştırma seviyesi için kesikli dalgacık dönüşümü

Şekil 1'e göre, her filtreleme katmanından sonra filtrelerin çıkışlarındaki işaretlerin örnek sayısı yarı yarıya azaltılmaktadır. Alçak geçiren (L) filtrelerin çıkışlarındaki katsayılar yaklaşım katsayıları ve yüksek geçiren (H) filtrelerin çıkışındaki katsayılar ayrıntı veya dalgacık katsayıları olarak isimlendirilmektedir. Yaklaşım katsayıları işaretin yerel ortalamalarını verirken, ayrıntı katsayıları ise bu yerel ortalamalar arasındaki farkı ifade etmektedir.

Durağan olmayan bir işareti kapsayacak kadar uzun, ancak gerçek zamanlı bir işlem yapabilmek için de yeterince kısa bir süre olduğundan, analiz penceresinin uzunluğu 0,5 sn olarak alınmıştır. Milisaniye mertebesindeki pencere uzunluklarının yüksek derecede birbirleriyle ilişkili katsayılar ürettiği ve bunun da sınıflandırma sonuçlarını olumsuz olarak etkilediği gözlenmiştir.

Ana dalgacıkların seçiminde iki etken ön plana çıkmaktadır: Dalgacık ailesi ve yok etme moment sayısını gösteren dalgacık numarası. Bir dalgacığın fazla sayıda yok etme momentine sahip olması, onun yüksek dereceden polinomlar gibi karmaşık dalga şekillerini iyi bir şekilde ifade edebilmesi anlamına gelmektedir. Yok etme moment sayısı arttığında, dalgacık keskin olmayan daha yumuşak geçişlere sahip olmaktadır. Bu tip dalgacıklar, müzik sesleri gibi zamanla yavaş olarak değişim gösteren işaretlerde yüksek değerli katsayılar üretmektedir. Diğer taraftan hızlı geçişlerin ağırlıkta olduğu konuşma seslerinde, beklendiği üzere, daha düşük değerli katsayılar elde edilmektedir. Bu, konuşma-müzik ayrıştırımasında ayırt edici bir özellik olarak kullanılabilir. Buna ek olarak, dalgacık numaraları da filtre katsayılarının sayısını doğrudan etkilediklerinden işlem yükü açısından önem taşımaktadırlar.

Literatürde, temel dalgacık olarak Daubechies ailesinin konuşma ve müzik seslerinin ayrıştırılmasında başarılı olduğu ifade edildiğinden [16-17], bu çalışmada da Daubechies ailesi (db2, db8, db15 ve db20) tercih edilmiş olup, karşılaştırma amacıyla, yok etme moment sayısı düşük olan Haar dalgacığı (db1) kullanılmıştır.

Öznitelik çıkartımı aşamasında, kesikli dalgacık dönüşümü katsayıları, seçilen ana dalgacık kullanılarak her bölüt için hesaplanmıştır. İşaretler, 12 ayrıntı ve 1 yaklaşım katsayısı oluşturacak şekilde 12 seviyeye ayrıştırılmıştır. Ayrıştırma sonucu oluşan bu katsayıların ortalama, varyans, medyan gibi istatistiksel ölçümleri kullanılarak elde edilen öznitelik vektörünün uzunluğu 26 dır. Farklı frekanslardaki enerji dağılımı bilgisini elde etmek için komşu altbandların ortalamalarının birbirine oranı ilave öznitelik olarak bu vektöre eklenmiş olup, bu durumda oluşan öznitelik vektörünün uzunluğu 38' dir. Eklenen bu 12 adet özniteliğin performansta %1,5'luk bir iyileştirme sağladığı gözlemlenmiştir.

3. DENEYSEL SONUÇLAR

3.1. Veri Seti

Deneysel çalışmalar için internet radyolarından alınan; klasik müzik, pop ve rock gibi değişik müzik türleri ile farklı bay ve bayan konuşmacılara ait sesleri içeren kayıtlar kullanılmıştır. Bütün kayıtlar 44100 Hz örnekleme frekansı ile örneklenmiş olup; veritabanında, her biri 0,5 saniye uzunluğundaki olan toplam 2194 müzik sesi ve 2624 konuşma sesi bulunmaktadır. Zaman tabanlı ve FFT tabanlı yöntemler için 0,5 sn'lik pencereler içinden %12,2 (512 örnek) örtüşümlü olmak üzere 4196 örnek uzunluğunda bölütler kullanılmıştır. KDD özniteliklerinde ise 0,5 saniyelik pencerelerin dalgacık dönüşümü hesaplanmış ve öznitelikler elde edilmiştir.

3.2. Sınıflandırma Algoritmaları

KMA işlemi için yapay sinir ağları ve destek vektör makineleri (DVM) olmak üzere iki farklı sınıflandırıcı incelenmiştir. Sınıflandırma aşamasında ilk olarak ölçeklendirilmiş gradyan geriye yayılım yordamıyla eğitilen çift katmanlı yapay sinir ağları kullanılmıştır. Eşlenik gradient algoritmalar, çok sayıda ağırlık katsayısı içeren ağlarda yüksek başarım göstermektedir. Aynı zamanda düşük hafıza gereksinimi ve hızlı bir algoritma olmasıyla da ön plana çıkmaktadır. Yapılan deneysel gözlemler ışığında gizli katmandaki nöron sayısı 40 olarak seçilmiş, hedeflenen ortalama hata 0,001 olarak alınmıştır.

DVM, istatistiksel öğrenme kuramında geçen, yapısal risk enküçültme prensibine dayalı bir ikili sınıflandırıcıdır. DVM sınıflandırıcısı $\boldsymbol{x}_j \in \mathbb{R}^n$ giriş verisini bir $\phi(\mathbf{x})$ fonksiyonu ile daha yüksek boyutlu bir H öznitelik uzayına eşlemekte ve paylarıençoklayan bir hiperdüzlem bulmaya çalışmaktadır. Pay, eniyi hiperdüzlem ile uzaydaki her sınıfın bu hiperdüzleme en yakın noktası arasındaki uzaklığı ifade etmektedir. $j = 1, \ldots, L$ ve $\mathbf{y} \in \{1, -1\}^L$ olmak üzere, $(\mathbf{x}_j, \mathbf{y}_j)$ giriş-çıkış çiftlerinden oluşan bir eğitim seti için DVM sınıflandırıcısı aşağıdaki şekilde tanımlanmaktadır:

$$g(\mathrm{x}) = \mathrm{sign}\left[\sum_{j=1}^{L} \alpha_j y_j \phi(\mathbf{x}_i)^T \phi(\mathbf{x}) + b\right] \quad (9)$$

Denklem (9)'da b, hiperdüzleme ait yanlılık miktarını göstermekte olup, α_j katsayıları ise aşağıda verilen dışbükey karesel programlama probleminin çözülmesi ile elde edilmektedir:

$$\max_{\alpha} -\frac{1}{2}\sum_{j=1}^{L}\sum_{i=1}^{L} y_j y_i \phi(\mathbf{x}_i)^T \phi(\mathbf{x}_j)\alpha_i \alpha_j + \sum_{j=1}^{L} \alpha_j \quad (10)$$

Bu problem için kısıtlar ise aşağıdaki gibi verilmektedir:

$$\sum_{j=1}^{L} \alpha_j y_j = 0 \quad (11)$$

$$0 \le \alpha_j \le C \quad (12)$$

Denklem (12)'de C, bir düzenlileştirme parametresi olup, pay ve yanlış-sınıflandırma hatası arasındaki dengeyi kontrol etmektedir. $\phi(\mathbf{x}_i)^T \phi(\mathbf{x}_j)$ terimi ise öznitelik uzayında iki vektörün çarpımını ifade etmektedir. Öznitelik uzayında eniyi hiperdüzlemin oluşturulması için, öznitelik uzayının belirtik şekli yerine iç çarpım çekirdeği $K(\mathbf{x}_i^T \mathbf{x}_j) = \exp\left(-\sigma \|\mathbf{x}_i - \mathbf{x}_j\|^2\right)$, $(\sigma > 0)$ kullanılabilmektedir [22].

Araştırmacılar tarafından yeni çekirdekler önerilmiş olmasına rağmen temel olarak 4 çekirdek tipi vardır: doğrusal, polinom, radyal baz fonksiyonu (RBF) ve sigmoid çekirdeği. Bu uygulamada, aşağıda listelenen sebeplerden dolayı

$K(\mathbf{x}_i^T \mathbf{x}_j) = \exp\left(-\sigma \|\mathbf{x}_i - \mathbf{x}_j\|^2\right)$, $(\sigma > 0)$ ile verilen RBF çekirdeği kullanılmıştır:

(i) RBF çekirdeği, örnekleri daha yüksek boyutlu bir uzaya doğrusal olmayan şekilde eşlediğinden, doğrusal çekirdeğe göre daha başarılıdır.
(ii) RBF çekirdeği, polinom çekirdeğine göre daha az parametreye (C ve σ) ihtiyaç duymaktadır.
(iii) RBF çekirdeği daha az sayısal zorluğa sahiptir [24].

C ve σ değerlerinin belirlenmesinde ızgara tarama yöntemi kullanılmaktadır. Bu yönteme göre C ve σ için belli bir sayısal değer aralığı (C=20, 40, … , 800 ve $\sigma=10^{-5}$, 10^{-4}, 10^{-3}, …, 10^{3}) taranmakta, sonrasında ise bu aralıkta en yüksek doğruluk oranını sağlayan C ve σ değerleri seçilmektedir.

3.3. Sınıflandırma Sonuçları

Başarım değerlendirmesi amacıyla veri seti, sırasıyla, 3212 ve 1606 örnek içerecek şekilde eğitim ve test gruplarına ayrılmıştır. Eğitim seti 1463 müzik ve 1479 konuşma sesi içermekte, test seti ise 731 müzik sesi ve 875 konuşma sesi içermektedir. Eğitim ve test setiyle oluşturulan öznitelik vektörleri, ortalamaları 0 ve varyansları 1 olacak şekilde normalize edilmiştir. Birbiri ile ilişkili özniteliklerin elenmesi için temel bileşen analizi (TBA) uygulanmıştır. Veri setinde toplam varyasyona %0,05'ten daha az katkıda bulunan temel bileşenler elenmiştir. Bu kısımda, sınıflandırma sonuçları doğruluk ölçütüyle sunulacaktır. Doğruluk, bir ikili sınıflandırıcının bir şartın sağlanıp sağlanmadığını ne oranda doğru olarak tespit edebildiğini gösteren bir ölçüt olup, doğru sonuçların (doğru pozitif veya doğru negatif) toplam örnek sayısına oranlanmasıyla elde edilir.

Zaman ve frekans dönüşümü tabanlı öznitelikler için 0,5 sn'lik ses örnekleri 4096 örneklik %12,2 (512 örnek) örtüşümlü bölütlere ayrılmış ve öznitelikler elde edilerek yapay sinir ağları ve DVM ile sınıflandırma yapılmıştır. Farklı bölüt uzunlukları ile yapılan deneyler sonucunda işlem yükü ve sınıflandırma başarısı yönünden 4096 örneklik bölütlerin daha başarılı olduğu gözlenmiştir.

Literatürde dalgacık dönüşümü kullanılan yöntemler 20 ms'den 2,4 saniye ye kadar farklılık göstermektedir. Farklı uzunluktaki analiz pencerelerinin başarımı, konuşma/müzik ayrıştırma deneylerimizde en yüksek başarımı gösteren Db8 ana dalgacığı için de değerlendirilmiştir. Kesikli dalgacık dönüşümü; 30, 40 ve 50 ms ve 0,5 sn uzunluğundaki pencereler için gerçekleştirilmiştir. Çizelge 1'de gösterildiği gibi, kısa pencerelerdeki örnekler birbirleriyle çok ilintili olduğu için başarım %80 civarında kalmaktadır. Bu sonuç, çalışmamızda önerilen saniyeler mertebesindeki pencereler için çıkartılan dalgacık tabanlı özniteliklerin daha ayırt edici özelliklere sahip olduğu iddiasını destekler niteliktedir. Db8 dalgacığı kullanılarak gerçekleştirilen kesikli dalgacık dönüşümünün ortalama işlem süresi kullanılan ekipmana ve işlenen verinin uzunluğuna göre değişebilmektedir. Bu süre, 0,5 sn pencere uzunluğu için yapılan simülasyonlarda yaklaşık olarak 50 msn olarak hesaplanmıştır. Dolayısıyla, gerçekli zamanlı bir uygulamada 50 msn'lik gecikme ile sınıflandırma yapılabilmektedir.

Çizelge 1. Farklı pencere uzunlukları kullanıldığı durumda, Db8 dalgacığı ile oluşturulan öznitelik vektörlerinin temel bileşen analizi uygulandıktan sonraki sınıflandırma başarımları

Pencere Uzunluğu	YSA (%)	DVM (%)
30 msn	83,30	89,47
40 msn	81,87	87,54
50 msn	81,62	86,11
0,5 sn	97,69	**98,38**

Zaman ve frekans ortamı tabanlı yöntemlerle oluşturulan öznitelik vektörü ve bu çalışmada önerilen öznitelik vektörünün uzunluklarının, temel bileşen analizinden önce ve sonra nasıl değiştiği Çizelge 2'de gösterilmektedir.

Çizelge 2. KMA sistemlerinde genel olarak kullanılan öznitelikler ile Haar, Db2, Db8, Db15 ve Db20 ana dalgacıkları kullanılarak elde edilen (altbantlar arası oranları da içeren) KDD tabanlı öznitelik vektörlerinin temel bileşen analizinden önceki ve sonraki boyutları

	Temel bileşen analizi uygulanmadan önce	Temel bileşen analizi uygulandıktan sonra
Genel Öznitelikler	21	20
Haar	38	19
Db2	38	19
Db8	38	22
Db15	38	21
Db20	38	21

Çizelge 3'te ise, komşu alt bantların oranlarına ait ortalama değerlerin hariç tutulduğu öznitelik vektörleri için temel bileşen analizi sonuçları gösterilmektedir. Çizelge 2 ve Çizelge 3 karşılaştırıldığında; bu ortalama değerlerin, öznitelik vektörüne ayırt edici özellik taşıyan 5 veya 6 ilave parametre eklediği görülmektedir.

Çizelge 3. KMA sistemlerinde genel olarak kullanılan öznitelikler ile Haar, Db2, Db8, Db15 ve Db20 ana dalgacıkları kullanılarak elde edilen (altbantlar arası oranları içermeyen) KDD tabanlı öznitelik vektörlerinin temel bileşen analizinden önceki ve sonraki boyutları.

	Temel bileşen analizi uygulanmadan önce	Temel bileşen analizi uygulandıktan sonra
Haar	26	13
Db2	26	13
Db8	26	16
Db15	26	16
Db20	26	16

Öznitelik çıkarım yöntemlerinin verimliliğini karşılaştırmak amacıyla, sınıflandırıcıların başarımı doğruluk ölçütü kullanılarak hesaplanmış,

Çizelge 4 ve Çizelge 5'de sunulmuştur. Çizelgelere göre, öznitelik vektörüne komşu alt bantlara ait oranların ortalamalarının eklenmesi ile sınıflandırıcı başarımlarında %1,5'luk bir iyileşme sağlandığı görülmektedir.

Çizelge 4. KDD tabanlı 26 öznitelik içeren öznitelik vektörlerinin sınıflandırma başarımı.

	YSA (%)	DVM (%)
Haar	94,58	95,52
Db2	95,95	96,32
Db8	97,69	98,38
Db15	97,88	98,13
Db20	98,32	98,07

En yüksek başarım sonuçları çizelgelerde koyu renkle vurgulanmıştır. Çizelge 5'teki sonuçlara göre dalgacık tabanlı öznitelikler, geleneksel yöntemlere göre daha yüksek sınıflandırma başarımı göstermektedir. Daha yüksek yok etme moment sayısına sahip dalgacıklar daha fazla ayrıştırıcı özelliklere sahipken, moment sayısı arttırıldığında buna paralel olarak sistemin karmaşıklığı da artmaktadır. Bu durumda konuşma/müzik ayrımında yeterince yumuşak geçişlere sahip Db8 dalgacığı, hesaplama açısından da sisteme çok yük getirmemektedir. Ayrıca, yapılan deneylerde Db8 dalgacık ailesinin oldukça başarılı olduğu gözlenmiştir.

Çizelge 5. KDD tabanlı 38 öznitelik içeren öznitelik vektörlerinin sınıflandırma başarımı

	YSA (%)	DVM (%)
Genel Öznitelikler	94,00	95,95
Haar	96,51	97,95
Db2	97,69	98,07
Db8	99,19	99,13
Db15	98,63	99,00
Db20	98,69	99,00

Çalışmada kullanılan yöntemlerin ortalama hesaplama süreleri de ayrıca ölçülmüştür. 2 GHz hızında Intel Core Duo işlemcili bir bilgisayarda MATLAB R2008b kullanılarak ölçülen hesaplama süreleri Çizelge 6'da verilmektedir. Hesaplama süresi, filtre katsayılarının sayısı ile doğru orantılı olarak artmakta ve bu nedenle en hızlı öznitelik çıkartımı, en az sayıda süzgeç içerdiğinden Haar dalgacığı ile gerçekleşmektedir. DVM sınıflandırıcısında, Db15'de de başarımın yüksek olduğu gözlemlenmektedir. Ancak, öznitelik çıkarım süresi Db8'e göre daha uzundur.

4. SONUÇ

Konuşma/müzik ayrıştırımı üzerine uygulamalar birçok alanda kullanıldığından, daha verimli öznitelik çıkartımı üzerine yapılan çalışmalar son

zamanlarda artış göstermiştir. Literatürde birçok zaman, frekans ve zaman-frekans temelli yöntem önerilmiş olmasına rağmen, daha verimli yöntem ve sistemlerin geliştirilmesi için çalışmalar hala devam etmektedir. Bu yöntemlerden biri de, zamanla değişen frekans bileşenlerini analiz etmekte başarılı olan dalgacıkları kullanmaktır. Konuşma ve müzik seslerinin hiçbiri durağan olmadığından, kesikli dalgacık dönüşümü tabanlı teknikler daha ayırt edici nitelikte öznitelikler sağlayabilmektedir. Bu sebeple, yapılan bu çalışmada KDD tabanlı yöntemler incelenmiştir.

Çizelge 6. Öznitelik çıkarım yöntemleri için ortalama hesaplama süreleri

	Konuşma (sn)	Müzik (sn)
Genel Öznitelikler	0,2768	0,2745
Haar	0,0357	0,0382
Db2	0,0401	0,0400
Db8	0,0485	0,0462
Db15	0,1035	0,1034
Db20	0,1550	0,1547

Analiz yapılan bölüt uzunluğu, hem zaman hem frekans temelli yöntemler için önemli bir faktördür. Frekans temelli yöntemler için işaret durağan olarak kabul edildiğinden, uzun bir bölüt aralığı seçmek, bu varsayımın geçerliliğini ters yönde etkileyecektir. Bunun yanında, çok kısa pencere aralıkları seçmek de anlamlı öznitelik çıkartılmasını engelleyebilmektedir. Önceki çalışmalardan farklı olarak 0,5 sn'lik pencereler için 95 ms uzunluğundaki bölütlerin konuşma/müzik ayrımında daha yüksek başarım gösterdiği gözlenmiştir.

KDD tabanlı öznitelik vektörünün oluşturulması için; işaret, analiz penceresi boyunca önceden belirlenmiş ana dalgacık ve seçilen seviyeye göre dalgacık katsayılarına ayrıştırılmaktadır. Analiz penceresinin uzunluğu ilk önemli kriter olarak karşımıza çıkmaktadır. Literatürde, 20 msn'den 2,4 sn'ye uzanan bir aralıkta pencereler kullanılıyor olmasına rağmen, yapılan çok sayıda benzetim çalışmasından sonra 0,5 sn'lik bir pencere uzunluğunun konuşma/müzik ayrımı için yeterli olduğu sonucuna ulaşılmıştır.

İşaret işleme alanına yönelik çalışmalarda, uygulamanın türüne göre ihtiyaçlar farklılık gösterdiğinden hangi dalgacık ailesinin kullanılacağı konusunda ortak bir kanı yoktur. Ancak, Daubechies ailesi dikgenlik (orthogonality) özelliğinden ve yok etme moment sayısının seçilebilir olmasından dolayı, ses ve müzik sesi işleme ile ilgili çalışmalarda tercih edilmektedir. Bu yüzden bu çalışmada da yok etme momentleri ve karmaşıklığı göz önünde bulundurarak Daubechies ailesi kullanılmıştır. Daubechies ailesinden Db8 dalgacığı, müzik işaretlerinin analizinde Haar gibi yok etme moment sayısı düşük dalgacıklara göre çok daha güçlü bir araç olmaktadır. Yoketme moment sayısının çok fazla olması durumunda ise hem hesaplama yükü artmakta hem de öznitelik uzayı genişleyip daha karmaşık hale geldiğinden sınıflandırma performansı düşmektedir. Bu açıdan bakıldığında Db8 dalgacığının, performans ve karmaşıklık arasında bir denge sağladığı söylenebilir. Analiz edilen işaretlerin detaylarını yüksek çözünürlükle belirleyebilmek için 12 seviyede dalgacık ayrıştırması gerçekleştirilmiştir.

Komşu altbant oranlarının ortalamasının sınıflandırma başarımına etkisi ayrıca incelenmiştir. Bu öznitelik sayesinde komşu altbantlar arasında dalgacık enerjisinin değişimi de gözlenebilmektedir. Konuşma seslerinin enerji yoğunluğu genel olarak düşük frekans değerlerinde toplandığından bu öznitelik, konuşma/müzik ayrımında ayırt edici bir öznitelik olarak değerlendirilebilir.

Sonuç olarak, 0,5 sn uzunluğunda analiz penceresi ve Db8 dalgacığı kullanılarak elde edilen öznitelik vektörü, konuşma ve müzik ayrımında hem hesaplama süresi, hem de sınıflandırma başarımı açısından diğer yöntemlere göre daha üstün performans göstermektedir. Buna ek olarak, altbantların birbirine oranı, sınıflandırma başarımını arttırmıştır. DVM ve YSA performans açısından birbirine yakın sonuçlar üretmiştir.

Gelecekte yapılacak çalışmalarda müzik seslerinin içine gömülü konuşma sesleri için çoklu-sınıf sınıflandırmanın incelenmesi ve daha uzun pencere boyutları için başarımın ölçülmesi planlanmaktadır.

5. KAYNAKLAR

1. Mubarak, O. M., Ambikairajah, E., Epps, J., 2006. Novel Features for Effective Speech and Music Discrimination, IEEE Int. Conf. on Engineering of Intelligent Systems, Islamabad, Pakistan.

2. Exposito, J. E. M., Galan, S. G., Reyes N. R., Candeas, P., 2007. Audio Coding İmprovement Using Evolutionary Speech/Music Discrimination, IEEE Int. Conf. on Fuzzy Systems (FUZZ-IEEE), Londra, İngiltere.

3. El-Maleh, K., Klein, M., Petrucci, G., Kabal, P., 2000. Speech/music Discrimination for Multimedia Applications, IEEE Int. Conf. on Acoustics, Speech, Signal Processing (ICASSP'00), İstanbul, Türkiye.

4. Gedik, A., Bozkurt, B., 2010. Pitch Frequency Histogram Based Music Information Retrieval for Turkish Music, Signal Processing, Cilt 10, sayı 4, sayfa 1049–1063.

5. Saunders, J., 1996. Real Time Discrimination of Broadcast Speech/Music, IEEE Int. Conf. on Acoustics, Speech, and Signal Processing, Atlanta, ABD.

6. Scheier, E., Slaney, M., 1997. Construction and Evaluation of a Robust Multifeature Speech/Music Discriminator, IEEE Int. Conf. on Acoustics, Speech, Signal Processing (ICASSP'97), Münih, Almanya.

7. Ajmera, J., McCowan I., Bourlard, H., 2003. Speech/Music Segmentation Using Entropy and Dynamism Features in a Hmm Classification Framework, Speech Communication, Cilt 40, Sayı 3, 351-363.

8. Panagiotakis, C., Tziritas, G., 2005. A Speech/Music Discriminator Based on Rms And Zero-Crossings, IEEE Trans. on Multimedia, cilt 7, sayı 1, sayfa 155–166.

9. Wang, J., Wu, Q., Yan, Q., 2008. Real-time Speech/Music Classification with a Hierarchical Oblique Decision Tree, IEEE Int. Conf. on Acoustics, Speech, and Signal Processing (ICASSP'08), Las Vegas, ABD.

10. Pikrakis, T., Giannakopoulos, T., Theodoridis, S., 2008. A Speech/Music Discriminator of Radio Recordings Based on Dynamic Programming And Bayesian Networks, IEEE Trans. on Multimedia, Cilt 10, Sayı 5, sayfa 846-857.

11. Kos, M., Grasic, M., Kacic, Z., 2009. Online Speech/Music Segmentation Based on the Variance Mean of Filter Bank Energy, EURASIP Journal on Advances in Signal Processing, cilt 2009, sayfa 1–13.

12. Tzanetakis, G., Essl, G., Cook, P., 2001. Audio Analysis Using the Discrete Wavelet Transform, WSES Int. Conf. Acoustics and Music: Theory and Applications (AMTA 2001), Yunanistan.

13. Didiot, E., Illina, I., Fohr, D., Mella, O., 2010. A Wavelet-Based Parameterization for Speech /Music Discrimination, Computer Speech and Language, cilt 24, sayı 2, sayfa 341–357.

14. Ntalampiras, S., Fakotakis, N., 2008. Speech /Music Discrimination Based on Discrete Wavelet Transform, 5th Hellenic Conf. On Art.Int. (SETN'08), Yunanistan.

15. Khan, M. K. S., Al-Khatib, W. G., 2006. Machine-Learning Based Classification of Speech and Music, ACM Jour. on Multimedia Systems, Cilt 12, Sayı 1, Sayfa 55–67.

16. Düzenli, T., Özkurt, N., 2011. Comparison of Wavelet Based Feature Extraction Methods for Speech/Music Discrimination, IU-JEEE, cilt 11, sayı 1, sayfa 1355-1362.

17. Düzenli, T., Özkurt, N., 2011. Discrete and Dual TreeWavelet Features for Real-Time Speech/Music Discrimination, ISRN Signal Processing, cilt 2011, Article ID 269361.

18. Mallat, S., 1999. A wavelet tour of signal processing, Elsevier Academic Press, 3. Basım, Burlington, MA, ABD.

19. Joachims, T., 2002. Learning to Classify Text Using Support Vector Machines: Methods, Theory and Algorithms. Kluwer Academic Publishers Norwell, MA, ABD.

20. Zheng, F., Zhang, G., Song, Z., 2001. Comparison of Different Implemantations of mfcc, Arch. Rat. Mech. Anal., Cilt 16, Sayı 6, sayfa 582-589.

21. Haykin, S., 1999. Neural Networks A Comprehensive Foundation (2nd ed.), New Jersey: Prentice Hall.
22. Vapnik, V. N., 1999. The Nature of Statistical Learning Theory (2nd ed.), New York: Springer.
23. Moller, M. F., 1993. A Scaled Conjugate Gradient Algorithm for Fast Supervised Learning, Neural Networks, Cilt 6, Sayı 4, Sayfa 525-533.
24. Hsu, C.-W., Chang, C.-C. ve Lin, C.-J., 2010. A Practical Guide to Support Vector Classification, http://www.csie.ntu.edu.tw/ cjlin.
25. Duda, R. O., Hart, P. E., Stork, D. G., 2001. Pattern Classification and Scene Analysis (2nd ed.), New York: John Wiley & Sons Inc.