

Çalgı Seslerinin LS Frekansları Kullanılarak Tanınması

Bora Mutluer¹

İbrahim Baran Uslu²

Sıtkı Çağdaş İnam³

Elektrik-Elektronik Mühendisliği Bölümü, Başkent Üniversitesi, Ankara

¹e-posta: bmutluer@gmail.com

²e-posta: ibuslu@baskent.edu.tr

³e-posta: inam@baskent.edu.tr

Özetçe

Bu çalışmada, farklı çalgıların (flüt, obua, keman, çello) seslerinin tanınması amaçlı bir uygulama yapılmıştır. Tanımda özellik olarak LS (Line Spectral: Çizgi Tayf) frekansları kullanılmaktadır. İki adet nefesli (flüt ve obua) ve iki adet yaylı (keman, çello) çalgıya ait tek sesli (tonlu) ses örnekleri, K-NN (K Nearest Neighbor: En Yakın K Komşu) yöntemine göre sınıflandırılmıştır. Sonuçta %97,52 başarı ile çalgı tanıma; %98,40 başarı ile çalgı ailesi tanıma gerçekleştirilmiştir.

1. Giriş

Müzik içerik analizi; müziğin, bilgisayar ortamında otomatik olarak notaya dökülmesine ve kataloglandırılmasına olanak sağlar. Bu tip uygulamalara özellikle, müzik eğitiminde ve son yıllarda sayıları gittikçe artan çevrimiçi müzik mağazalarında ihtiyaç duyulmaktadır. Çalgıların otomatik olarak tanınması, müzik içerik analizinin en önemli bileşenlerinden birisidir.

Bu alanda yapılan güncel çalışmalar bulunmaktadır. Brown yaptığı çalışmada, iki çalgının (obua ve saksofon) ayırt edilmesinde cepstral katsayıları ve örüntü tanıma tekniklerini kullanmış, %84,1 başarımla elde etmiştir. Sınıflandırma için GMM (Gaussian Mixture Models) yönteminden yararlanmıştır [1].

Marques ve Moreno, yaptıkları çalışmada, 8 farklı çalgı (klarnet, flüt, klavye, org, piyano, trombon, keman) için doğrusal öngörü (LPCC), cepstral ve Mel frekans cepstral katsayıları kullanarak, GMM ve Support Vector Machines (SVM) sınıflandırıcıları ile %70 başarımla elde etmiştir [2].

Kitahara, Goto ve Okuno, F0 temel frekans parametresini dikkate alarak yaptıkları çalışmada, 19 çalgı için, %79,73 çalgı tanıma; %90,65 çalgı ailesi tanıma başarımla elde etmişlerdir [3].

Krishna ve Sreenivas, çalışmalarında özellik olarak LS frekanslarını; sınıflandırıcı olarak da GMM ve K-NN (K Nearest Neighbor: En Yakın K Komşu) yöntemlerini kullanmış, 14 çalgı için yaklaşık %90 başarımla elde etmişlerdir. Aile tanıma başarımları ise yaklaşık %95'tir [4].

Bolat ve Küçük, yaylı çalgı seslerini tanımda LPC katsayıları ve olasılıksal sinir ağı kullanarak %76,47 başarımla elde etmişlerdir [5].

Özbek ve Savacı, yeni çalışmalarında, dalgacık (wavelet) analizi ile elde edilen altbant katsayılarını özellik olarak kullanıp, genelleştirilmiş Gauss yoğunluk modellemesi yöntemiyle çalgı sınıflandırması yapmışlardır [6].

Bu çalışmada, özellik olarak LS frekansları, sınıflandırma için de K-NN algoritması kullanılmaktadır. Makalenin 2. ve 3. bölümlerinde LS frekansları ve K-NN sınıflandırması hakkında temel bilgiler verilmiştir. 4. bölümde, uygulama aşamasından bahsedilmiş, 5.bölümde de ulaşılan sonuçlar sunulmuştur.

2. LS Frekansları

Ses sinyallerinin frekans karakteristikleri, insanların bu sesleri algılaması ve ayırt edebilmesinin temelini oluşturmaktadır. Konuşma sıkıştırma uygulamalarında da kullanılan LS frekansları; sinyallerinin çeşitli frekans özelliklerini özetleyebildiği için, tını modelleme çalışmalarında da kullanılabilir.

LS frekansları, doğrusal öngörü kodlama (LPC) katsayılarından elde edilen ve LPC katsayılarının bir gösterim şekli olan LS çiftlerinin (LSP) açısal frekanslarına karşılık gelir. Bu bölümde, LPC katsayılarının çıkartılması ve bu katsayıların LS çiftleri ve LS frekanslarıyla olan ilişkisi anlatılmıştır.

2.1. Doğrusal Öngörü Kodlama (LPC)

LPC, farklı konuşma işleme uygulamalarında, konuşmanın zarf ve kısa vadeli güç spektrumunun gösteriminde kullanılır. Kesikli bir sinyalin güncel değerlerinin, sinyalin geçmiş değerlerinin doğrusal bir fonksiyonu olarak ifade edilmesine, doğrusal öngörü denir. Şu şekilde ifade edilir:

$$\hat{x} = -\sum_{i=1}^M a_i x(n-i) \quad (1)$$

(1) eşitliğinden Levinson-Durbin algoritması kullanılarak hesaplanan a_i katsayıları ile ;

$$A(z) = 1 + \sum_{i=1}^M a_i z^{-i} \quad (2)$$

analiz süzgeci tanımlanır. $1/A(z)$ ise sentez süzgeci olarak adlandırılır.

2.2. LS Çiftleri (LSP)

LPC katsayılarını doğrudan kullanmak, sistemleri hatalara karşı duyarlı hale getirmektedir. Katsayılardaki ufak değişimler sentez filtresinin tüm frekans cevabını değiştirebilmekte, hatta sentez filtresini kararsız hale getirebilmektedir. Bu duyarlılığın üstesinden gelmek amacıyla çeşitli LPC katsayı gösterimleri geliştirilmiştir. Bu gösterimlerden en çok bilineni, Itakura [7] tarafından geliştirilen LS çiftleri gösterimidir. Burada, analiz süzgeci: $A(z)$,

$$\begin{aligned} P(z) &= A(z) + z^{-(M+1)} A(z^{-1}) \\ Q(z) &= A(z) - z^{-(M+1)} A(z^{-1}) \end{aligned} \quad (3)$$

polinomlarının oluşturulmasında kullanılır. Bu polinomlar simetrik ve kökleri birim çemberin üzerinde $P(z)$ 'in köküyle başlamak üzere sıralı olarak yer alır. Böylece, bu polinomlar köklerinin açısız konumuyla tanımlanabilir. Birim çemberin üst kısmında yer alan LS çiftlerinin açısız konumları, onların LS frekanslarını verir. LS frekanslarının hesaplanmasında Çebişev polinomları kullanılabilir [8].

3. Sınıflandırma

K-NN; veri madenciliği (data mining), örüntü tanıma (pattern recognition) ve imge işleme (image processing) gibi pek çok alanda kullanılan denetimli öğrenme (supervised learning) tekniklerinden biridir.

Sorgulanan örneğin sınıfını, örneğe en yakın K adet eğitim örneğinin sınıflarına göre belirler. Mesafe ölçüsü için genellikle Euclidean uzaklığı kullanılır. K adet eğitim örneği içinde en çok elemana sahip olan sınıf, sorgulanan örneğin de sınıfı kabul edilir.

4. Uygulama

4.1. Çalgı Veritabanı

Çalışma süresince, the University of Iowa'nın müzik enstrümanları örnekleri (MIS) [9] kullanılmıştır. Yaylı ailesinden çello ve keman, nefesli ailesinden flüt ve obua olmak üzere, toplam 4 adet çalgı incelenmiştir. Öncelikle, veritabanında yer alan bu çalgılara ait ses örnekleri tek sesli (tonlu) toplam 849 adet dosyaya ayrılmıştır. Keman ve çello için sadece yayla çalınan notalar incelenmiş, pizzicato (parmakla çekilen) olanlar kullanılmamıştır. Veritabanında yer alan bütün çalma şiddetlerindeki (pianissimo, mezzo-forte, fortissimo) ses örnekleri çalışmaya dahil edilmiştir. Bu dosyalar, atak ve sönüm kısımları atılarak, 1'er saniyelik örnekler haline getirilmiştir. Tablo 1'de çalgılara göre örnek dağılımı görülebilir.

Tablo 1 – Enstrümanlara göre örnek sayıları

Yaylı		Nefesli	
Çello	351	Flüt	113
Keman	281	Obua	104
Toplam	632	Toplam	217

Her çalgı için, örneklerin 3'te 2'si eğitim, 3'te 1'i ise test için kullanılmıştır. 3 farklı eğitim ve test kümesi oluşturulmuş, genel başarımlar bu 3 kümenin başarımlarının ortalamasının alınmasıyla elde edilmiştir.

4.2. Eğitim Evresi

Eğitim evresinde, ses örneklerinden LSF vektörleri çıkartılmakta ve bu veriler, test evresinde kullanılmak üzere kaydedilmektedir.

LSF vektörleri çıkartılırken 23 ms'lik Hamming pencereleri kullanılmaktadır. Pencereler arasında 6 ms örtüşme vardır. K-NN algoritmasının sınıflandırma için tüm eğitim verisine ihtiyaç duyması, eğitim verisini azaltma zorunluluğunu beraberinde getirmektedir. Bu amaçla; her örnek için, çıkartılmış tüm LSF vektörlerinin ortalaması alınır. Sonuçta, her ses örneğinden bir adet ortalama LSF vektörü elde edilmektedir.

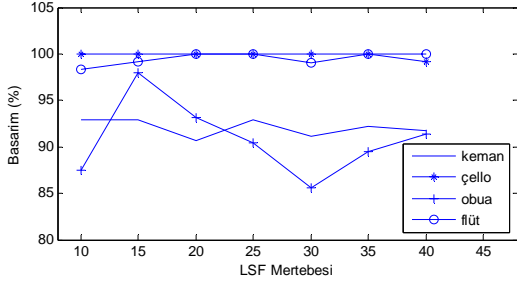
4.3. Test Evresi

Test evresinde, test için ayrılmış ses örneklerinden ortalama LSF vektörleri çıkartılmakta ve bu veri K-NN yöntemiyle, eğitim evresinde oluşturulan veriler kullanılarak sınıflandırılmaktadır. LSF vektörleri çıkartılırken, eğitim evresinde olduğu gibi 23 ms'lik ve 6 ms örtüşmeli Hamming pencereleri kullanılmaktadır. Sınıflandırmalar yapılırken K-NN'in derecesi 1 alınmıştır.

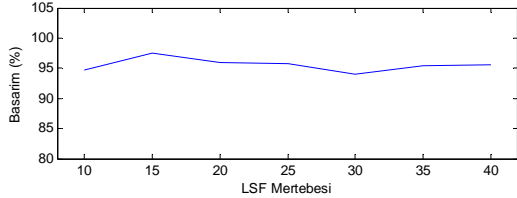
Eğitim ve test aşamalarında diğer pencere uzunluklarının ve örtüşme miktarının sonuçlara etkisi araştırılmış ve bu parametrelerdeki değişikliklerin başarımlara belirgin bir etkisinin olmadığı görülmüştür.

5. Sonuçlar ve Tartışma

LSF mertebesinin farklı değerleri için tanıma başarımının değişimine bakılmış, Şekil 1 ve Şekil 2'de gösterilen sonuçlar elde edilmiştir. Buna göre en yüksek ortalama başarımlar (%97,52) LSF mertebesinin 15 olduğu durumda elde edilmiştir. Sonuçlar Tablo 2'de listelenmiştir.



Şekil 1 LSF mertebesine göre her bir çalgı için tanıma başarımları.



Şekil 2 LSF mertebesine göre çalgı tanıma başarımları.

Tablo 2 Çalgı tanıma test sonuçları (%)

Küme-1	Keman	Çello	Flüt	Obua
Keman	91,40	2,10	5,40	1,10
Çello		100		
Flüt			100	
Obua	3,00%			97,00

Küme-2	Keman	Çello	Flüt	Obua
Keman	90,50	7,40	2,10	
Çello		100		
Flüt	2,60		97,40	
Obua				100,00

Küme-3	Keman	Çello	Flüt	Obua
Keman	96,80	1,10	2,10	
Çello		100		
Flüt			100	
Obua	2,90			97,10

Ort.	Keman	Çello	Flüt	Obua
Keman	92,90	3,53	3,20	0,37
Çello		100		
Flüt	0,87		99,13	
Obua	1,97			98,30

Çalışmada çalgı ailesinin doğru tanıma başarımları da incelenmiştir. Bunun için flüt ve obua'nun eğitim verileri "nefesli" ailesi için; keman ve çello'nun eğitim verileri de "yaylı" ailesi için birleştirilmiştir. Çalgı ailesi testleri, çalgı tanıma en yüksek başarımları elde ettiğimiz LSF mertebesi

olan 15 alınarak yapılmış ve %98,4 ortalama başarımları elde edilmiştir. Sonuçlar Tablo 3'te gösterilmiştir.

Gelecekteki çalışmalarda, çalgı sayısının artırılması, farklı şartlarda yapılan ses kayıtlarının da test edilmesi ve başka sınıflandırma tekniklerinin kullanılması düşünülmektedir.

6. Teşekkür

Yazarlar, Prof. Dr. Turhan Çiftçi'na ve Yrd. Doç. Dr. Mustafa Doğan'a yararlı tartışmalar için teşekkür eder.

Tablo 3 Çalgı ailesi tanıma test sonuçları

Küme-1	Yaylı	Nefesli
Keman	93,60	6,40
Çello	100,00	
Flüt		100,00
Obua	3,00	97,00

Küme-2	Yaylı	Nefesli
Keman	97,90	2,10
Çello	100,00	
Flüt	2,60	97,40
Obua		100,00

Küme-3	Yaylı	Nefesli
Keman	97,90	2,10
Çello	100,00	
Flüt		100,00
Obua	2,90	97,10

Ort.	Yaylı	Nefesli
Keman	96,47	3,53
Çello	100,00	0,00
Flüt	0,87	99,13
Obua	1,97	98,03

7. Kaynakça

- [1] Brown, J.C., "Computer Identification of Musical Instruments Using Pattern Recognition with Cepstral Coefficients as Features", J. Acoust. Soc. Amer. 105, 1933-1941, 1999.
- [2] Marques, J. and Moreno, P.J., "A Study of Musical Instrument Classification Using Gaussian Mixture Models and Support Vector Machines", Cambridge
- [3] Kitahara, T.Goto, M., Okuno, H.G., "Musical Instrument Identification Based on F0-Dependent Multivariate Normal Distribution", Proc. of ICASSP, pp. 421-424, 2003.

