

Ayrık Dalgacık Tabanlı Öznitelikler İle Ses Hastalıkları Tanısı Voice Disorders Identification Using Discrete Wavelet Based Features

Ömer Eskidere¹, Ömer Aktas¹, Cevat Ünal¹

¹ Elektrik Elektronik Mühendisliği Bölümü Mühendislik Fakültesi
Bursa Orhangazi Üniversitesi
{omer.eskidere, omer.aktas, cevatal}@bou.edu.tr

Özetçe

Ses bozuklukları günümüzde en çok karşılaşılan rahatsızlıkların başında gelmektedir. Bu çalışma ses örneklerinin analiz edilmesi ile kişinin ses patolojisi içerip içermediğinin tespit edilmesi amaçlanmaktadır. Bu amaçla öznitelik vektörü olarak ayrık dalgacık dönüşüm tabanlı doğrusal öngörülü kepsral katsayıları (DÖKK) ve istatistiksel ölçümlerin birlikte kullanılması önerilmiştir. Deneylerde 5 farklı vokal fold hastalık grubu ve sağlıklı kişileri içeren toplam 304 kişiye ait uzatılmış /a/, /i/ ve /u/ sesleri kullanılmıştır. Deneysel sonuçlarda %99 üzerinde doğru tanıma başarımı gözlenmiştir.

Anahtar kelimeler: Ses hastalıkları, ayrık dalgacık dönüşümü, doğrusal öngörülü kepsral katsayıları

Abstract

Voice disorders are currently one of the most common diseases. This study aims to determine whether a person has voice pathology by analyzing his/her sound samples. For this purpose, co-utilizing the discrete wavelet transform based the linear predictive cepstral coefficients and their statistical parameters is proposed as feature vector. In the experiments, five different vocal fold disease groups and healthy individuals out of 304 people were employed using sustained /a/, /i/ and /u/ sounds. Experimental results show over 99% correct recognition performance.

Key words: Voice disorders, discrete wavelet transform, linear predictive cepstral coefficients

1. Giriş

Günümüzde klinik olarak ses hastalıklarının tanısında fiziksel muayene, videolaringoskopi, indirek laringoskopi gibi yöntemleri kullanılmaktadır. Bu yöntemler her ne kadar hastalığın tanısı hakkında önemli bilgiler verse de invaziv tanı koyma yöntemleridir. Akustik ses analizi, noninvaziv, hızlı ve etkili bir yöntem olup, ilk tanı koyma, tedavi takibi ve sonuçların objektif değerlendirilmesinde yardımcı olarak kullanılmaktadır. Bu tekniklerin sınıflandırma metodları ile

birlikte kullanımı ses patolojisi belirleme için uzman destek sistemlerin geliştirilmesine olanak tanımaktadır. Bu sistemler profesyonel olarak sesini kullanan (şarkıcı, spiker vb.) kişiler için önlüyci tıpta taşınabilir bir araç olarak yararlanılabilir. Ayrıca tele tıp ortamlarında bu sistemler uzaktan otomatik olarak ses bozukluğu belirlenmesinde kullanılabilir.

Sesi etkileyen hastalıkları bazen vokal fold gerginlik ve koordinasyonunda değişime yol açarken, bazen de ses üretim yolunda patolojik değişiklikler oluşturur. Özellikle ses tellerinde oluşan nodül, polip, kist, veya ödem, ses tellerinin tam kapanmasını engelleyerek ses tellerinin düzensiz titreşimler oluşturmasına neden olabilir. Bu ses hastalıklarının en yaygın belirtisi disfoni, ses kısıklığıdır. Disfoninin derecesini sesin spektrogramındaki harmonik bileşenlerin kaybı, gürültü bileşenlerinin miktarı ve temel frekanstaki aperiodyklik miktarı ile gözlenebilir [1]. Literatürde de bu konuda yapılan çalışmalarda disfoni derecesinin artması ile ses sinyalinde periyodik olmayan gürültü artışı görülmektedir [2]. Bazen de ses hastalığı öyle şiddetlenir ki ses sinyalinde herhangi bir periyodiklik gözlenmez. Bu durumlarda akustik ses analizlerinde yaygın olarak kullanılan temel frekansın periyot ve genliğindeki değişimler (jitter ve shimmer), harmoniklerin gürültüye oranı ve normalize gürültü enerjisi gibi parametreler yetersiz kalmaktadır. Bu parametrelerin doğruluğu, ses sinyalindeki temel frekansın doğru kestirimi ile yakından ilişkilidir. Sinyal periyodik yapıda olmayıp, alt harmonik bileşenler içerdiği veya periyodik olmayan titreşimler içerip random bir hal aldığı ses bozukluklarında, bu kestirimde hatalar ortaya çıkmaktadır. Bu durum farklı akustik analiz metodların araştırılması gerekliliğini ortaya çıkarmıştır. Bu amaçla disfonik sesin kaotik özellikte olduğu varsayımından hareketle bazı doğrusal olmayan dinamik analiz teknikleri önerilmiştir [3]. Bunlar arasında en yaygın olanları Lyapunov üstelleri ve korelasyon katsayıları yöntemleridir[4]. Diğer bir yaklaşım ise dalgacık transformu temelli yöntemler kullanmaktır. Literatürde alt bant kepsral katsayıları [5], ayrık dalgacık parametreleri [6-7], çok bantlı doğrusal öngörü katsayıları [8] ve dalgacık temelli MFCC [9], ayırt edici öznitelik olarak kullanılmaktadır.

Biz bu çalışmada ses kayıtlarından kaynak cihazı tanıma problemi için önerdiğimiz [10] ayrık dalgacık dönüşüm tabanlı

DÖKK ile istatistiksel ölçümlerin birlikte öznel olarak kullanılmasını, ses hastalıkları tespiti problemine uyguladık. Önerilen yöntem kişinin vokal patolojisi olup olmadığının otomatik olarak tespitinde yüksek başarımlar vermektedir.

2. Veri

Önerilen ses hastalığı belirleme sisteminin performansını test etmek için, Barry&Putzer tarafından Saarland üniversitesi Phonetics enstitüsünde oluşturulan Saarbrücken Ses Veri tabanı (SVD), kullanılmıştır. Veri tabanı 1320 ses hastası ve 650 normal konuşmacıdan alınan ses örneklerini içermektedir. Ses örnekleri uzatılmış /a/, /i/, /u/ sesleri ve "Guten Morgen, wie geht es Ihnen?" (Günaydın, nasılsınız?) cümlesini içermektedir [11]. Ses örnekleri düşük, orta ve yüksek olmak üzere 3 farklı perde frekanslarında kaydedilmiştir. Uzatılmış seslerin süresi 1-3 saniye arasında değişmekte olup kayıtların örnekleme frekansı 50 kHz dir.

Biz bu çalışmada veri tabanından 152 hasta ve 152 sağlıklı kişiye ait /a/, /i/ ve /u/ ses örneklerini kullandık. Her bir ses örneği 16 kHz örnekleme frekansı ile tekrar örneklendi. Tablo 1'de bu çalışmada kullanılan SVD'den seçilen ses örnekleri görülmektedir.

Tablo 1: Patolojik ve sağlam kişilere ait SVD veri tabanından seçilen konuşmacıların dağılımı

Teşhis	Konuşmacı Sayısı		
	Erkek	Kadın	Toplam
Vokal Nodül	3	13	16
Reinke's Ödem	7	60	67
Vokal Kanser	21	1	22
Vokal Polip	25	16	41
Vokal Kist	1	5	6
Normal	76	76	152
Toplam	133	171	304

3. Yöntem

3.1 Ayrık dalgacık dönüşümleri (ADD)

Ses ve konuşma analizinde, kısa süreli Fourier dönüşüm yöntemi ile eş zamanlı zaman-frekanstaki ani değişimlerin tespitinde yetersiz kalınması durumunda ADD kullanılmaktadır. ADD, değişen pencere boyutları kullanımına imkân tanımasından dolayı tercih edilmektedir. Ayrıca ADD, gürültüye karşı gürbüz olup ses sinyalindeki ani değişimleri takibine imkân vermektedir. Ayrık zamanda ADD katsayıları, bir sürekli zaman sinyali $s(t)$ için şu şekilde tanımlanır;

$$ADD_{j,k} = \langle s(t), \psi_{j,k}(t) \rangle = \int_{-\infty}^{+\infty} s(t) \psi_{j,k}(t) dt \quad (1)$$

Burada j , k tamsayılar olup, a ve b ayrıklaştırma parametreleri

$$a = a_0^j, \quad b = kb_0 a_0^j \quad (2)$$

ADD temeli $\psi_{j,k}(t)$ aşağıdaki gibi tanımlanabilir:

$$\psi_{j,k}(t) = a_0^{-j/2} \psi(a_0^{-j} t - kb_0), \quad (3)$$

Literatürde [12] sürekli bir sinyalinin hızlı piramit algoritması ile ayrıştırılması önerilmiştir. Bu algoritmaya göre sinyal, alçak frekans ve yüksek frekans bilgilerini içeren iki bileşene ayrıştırılmaktadır. Sinyal alçak ve yüksek geçiren filtreler kullanılarak yaklaşık ve ayrıntı katsayılarına ayrılır. Bu filtreler aşağıdaki şartı sağlaması gerekir [13]:

$$f_1(N-1-n) = (-1)^n \cdot f_0(n), \quad (4)$$

Burada N filtre genişliği, $f_0(n)$ alçak geçiren filtre, $f_1(n)$ ise yüksek geçiren filtredir. Daha sonra, ardışıl şekilde yaklaşık katsayıları yeni yaklaşık ve ayrıntı katsayılarına bölünür ve bu süreç iteratif olarak yaklaşık katsayıları üzerinde devam ettirilir. Sonuçta farklı seviyelerde yaklaşık ve ayrıntı katsayıları elde edilir. Her bir skalada ardışıl filtre ve alt örnekleme işlemi aşağıdaki gibi tanımlanabilir [13]:

$$ADD_{yak.} = \sum_n s[k] \cdot f_0[-n+2k], \quad (5)$$

$$ADD_{ay.} = \sum_n s[k] \cdot f_1[-n+2k], \quad (6)$$

3.2 ADD-DÖKK öznelik parametresi

Ayrık dalgacık dönüşümü ve doğrusal öngörü kepstum katsayıları (ADD-DÖKK) kullanarak öznelik parametresi elde etmek için ilk olarak ses sinyali belirli bir uzunlukta çerçevelere bölünür. Bu çerçevelere % 50 örtüşme ile ilerletilir. Daha sonra çerçevelenen sinyal parçasına Bölüm 3.1'de belirtildiği şekilde çok çözünürlük ayrışımı uygulanarak alt-bant katsayıları elde edilir. ADD için, Daubechies 3 (db3) dalgacığı seçilerek alt bantlara ayrıştırma yaklaşık ve ayrıntı katsayıları elde edilir. Şekil 1 ve Şekil 2'de sırasıyla sağlıklı ve patolojik bir kişinin /a/ ses örnek parçası ve 4 seviyeli çok çözünürlük ayrıştırma sonucu oluşan yaklaşık (a_4) ve ayrıntı ($d_1 - d_4$) katsayıları görülmektedir. ADD ayrıştırma ile oluşturulan bu alt-bant sinyalleri için doğrusal öngörü katsayıları (DÖK) kestirilir. DÖK katsayıları kestiriminde Levinson-Durbin algoritması kullanılmıştır. Daha sonra DÖK katsayıları kepstal katsayılarına (DÖKK) çevrilir ve her bir alt sinyal için elde edilen katsayılar birleştirilir. Bu şekilde sinyal spektrumundaki değişim daha iyi bir şekilde gösterilebilir. Ayrıca ADD uygulanmış alt bant sinyalinin istatistiksel olarak değişimi de anlamlı bilgiler içermektedir. Bu amaçla ortalama, standart sapma ve entropi gibi ölçümler kullanılabilir. Son olarak her bir alt bant için elde edilen istatistiksel ölçümler alt bantların DÖKK parametreleri birleştirilip tek bir öznelik haline getirilir. Bu öznelik çıkartma yöntemi ile ilgili ayrıntılı bilgi [10] nolu çalışmada bulunabilir.

3.3. Parametre ayarları

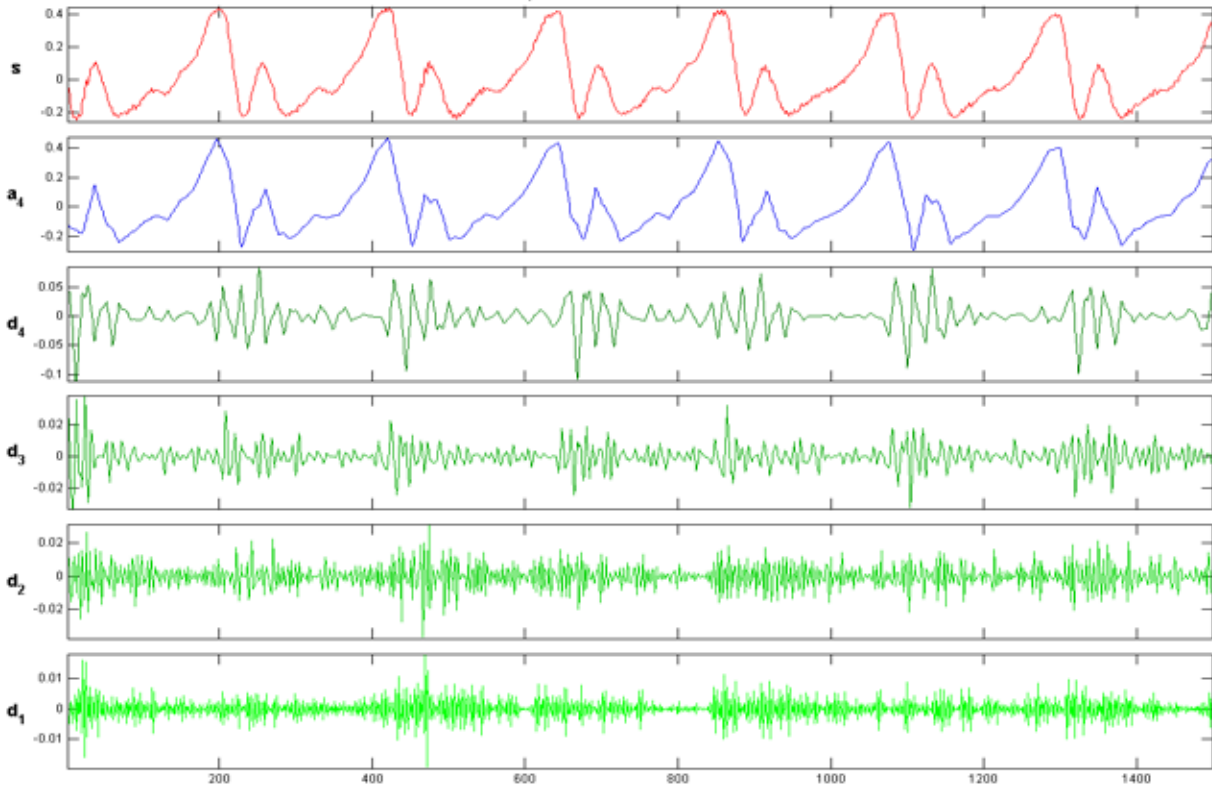
Ses bozukluğu belirleme deneylerinin değerlendirilmesi için sınıflandırıcı olarak en yakın komşu (k -NN) algoritması kullanıldı. Bölüm 3.2'de tanımlanan öznelikler kullanılarak, SVD veri tabanındaki ses örnekleri hasta veya sağlam kişiye ait olarak sınıflandırıldı. Sınıflandırma aşaması öncesi tüm öznelikler 0-1 arasına normalize edildi. Önerilen ses bozukluğu belirleme sisteminde, en yakın 3 komşu sayısı alındı. Özneliklerin yarısı eğitim kalan yarısı test için rastgele olarak ayrıldı ve deneyler 50 kez tekrar edildi her bir deneyin ortalama sonuçları hesaplandı.

4. Deneysel Bulgular

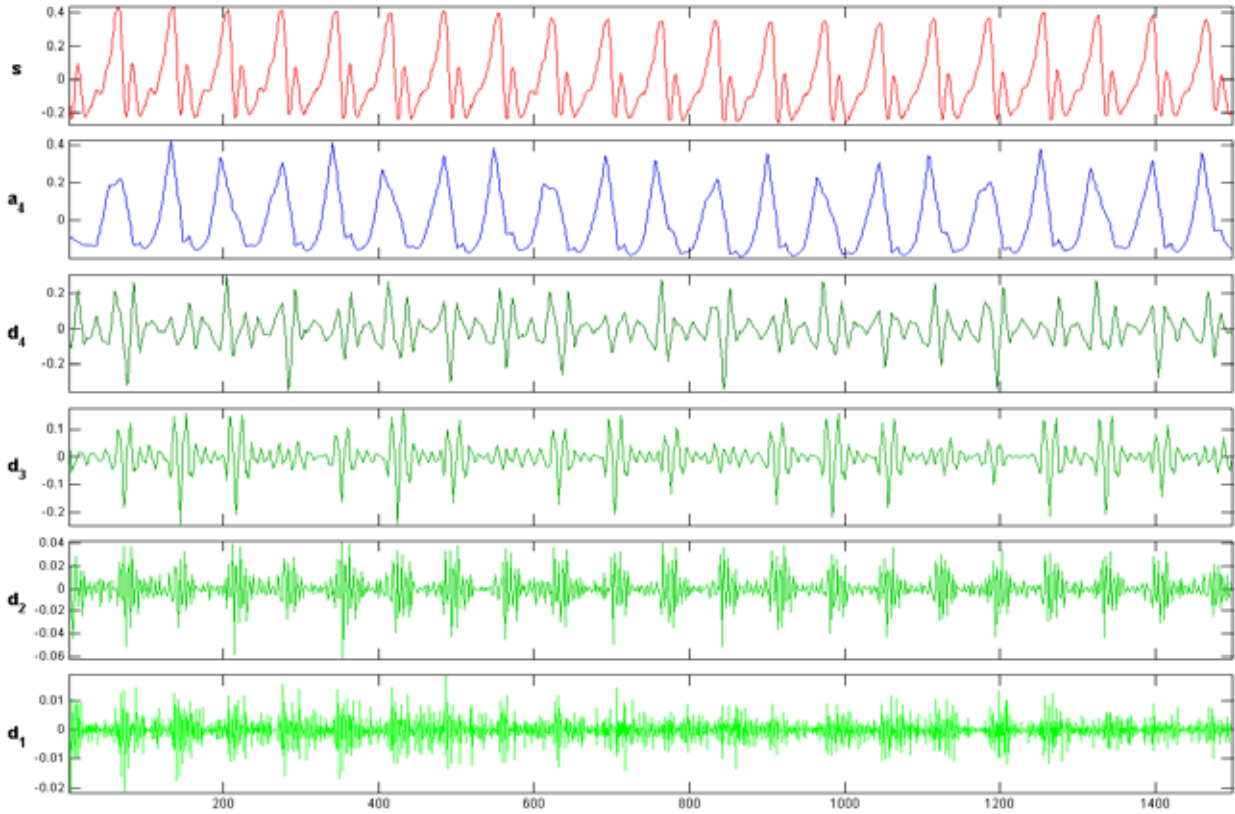
Deneysel değerlendirme için önerilen öznelik çıkartma yönteminin ses bozukluğu tanıma başarıları incelenmiştir. Bu amaçla 3 farklı ADD ayrıştırma seviyesi incelenmiştir. Öznelik çıkartımı için ilk olarak yalnız ADD-DÖKK, bu katsayılarla alt bant katsayılarının standart sapması ile birleştirilmiş hali (ADD-DÖKK-SS) incelenmiştir. Sürdürülmüş /a/, /i/ ve /u/ sesleri için doğru ses bozukluğu tanıma başarımları Şekil 3, Şekil 4 ve Şekil 5'de görülmektedir.

Sinyal İşleme 2

1. Gün / 15 Ekim 2015, Perşembe



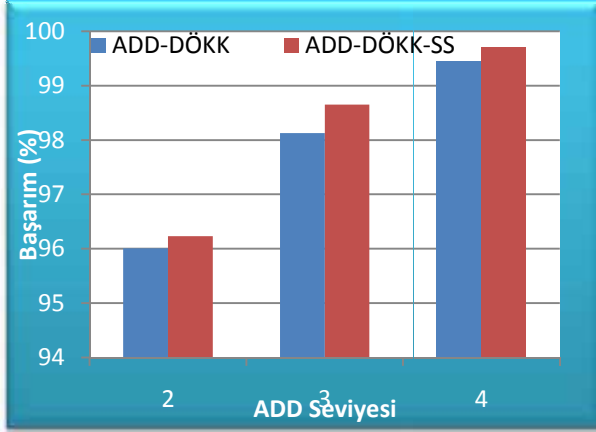
Şekil 1: Normal /a/ ses örneği ve 4. seviyeden yaklaşık ve ayrntı katsayıları.



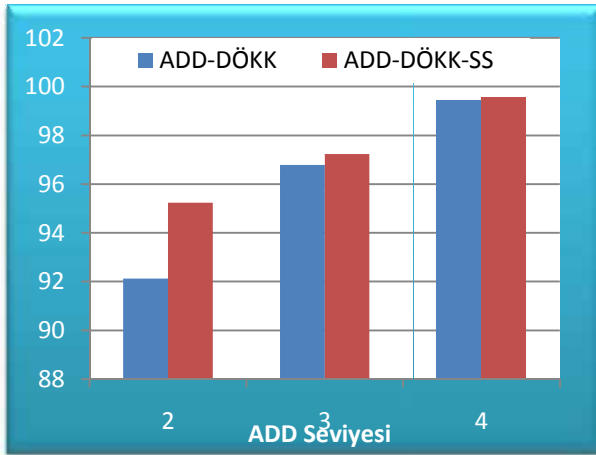
Şekil 2: Patolojik /a/ ses örneği ve 4. seviyeden yaklaşık ve ayrntı katsayıları.

Sinyal İşleme 2

Şekil 3, Şekil 4 ve Şekil 5'deki sonuçlardan görüleceği üzere en iyi tanıma başarımı ADD seviyesi 4 için elde edilmiştir. ADD-DÖKK-SS parametreleri ses patolojisi tanıma problemi için /a/, /i/ ve /u/ sesleri için %99'un üzerinde tanıma başarımı sağlamıştır.



Şekil 3: Sürdürülmüş /a/ sesi için doğru ses bozukluğu tanıma başarımları (%)



Şekil 4: Sürdürülmüş /i/ sesi için doğru ses bozukluğu tanıma başarımları (%)



Şekil 5: Sürdürülmüş /u/ sesi için doğru ses bozukluğu tanıma başarımları (%)

1. Gün / 15 Ekim 2015, Perşembe

5. Sonuçlar

Bu çalışmada akustik ses parametreleri kullanılarak kişinin vokal patolojiye sahip olup olmadığı araştırılmıştır. Bu amaçla yaygın olarak bilinen araç olan ayrık dalgacık dönüşümü temelli özellikler önerilmiştir. Ses sinyaline ADD uygulama ile elde edilen alt bantların DÖKK parametreleri ve istatistiksel ölçümleri birleştirilmesi önerilmiştir. Elde edilen sonuçlar ile önerilen metodun oldukça başarılı sonuçlar verdiği gözlenmiştir. Sağlıklı ve patolojik seslerin ayrılmasında %99 üzerinde başarımlar elde edilmiştir.

6. Kaynakça

- [1] Yanagihara N. "Significance of harmonic changes and noise components in hoarseness," J Speech Hearing Research 10: 531-541, 1967.
- [2] Titze IR: Workshop on acoustic voice analysis: Summary statement. 1995
- [3] J. B. Alonso, F. Díaz-de-María, C. M. Travieso, and M. A. Ferrer, "Using nonlinear features for voice disorder detection," in Proc. 3rd Int. Conf. Nonlinear Speech Process, Barcelona, Spain, Apr. 2005, pp. 94-106.
- [4] Y. Zhang and J. J. Jiang, "Nonlinear dynamic analysis in signal typing of pathological human voices," Electron. Lett. , vol. 39, pp. 1021-1023, Jun. 2003.
- [5] E. Erzin, A. E. Cetin, Y. Yardımcı, "Subband analysis for speech recognition in the presence of car noise", Proc. of ICASSP-95, Detroit, Vol. 1, pp. 417-420, 1995.
- [6] R. Sarikaya, B.L. Pellom, H.L. Hansen, "Wavelet packet transform features with application to speaker identification", Proc. of IEEE Nordic Signal Processing Symp., Visgo, Denmark, pp. 81-84, 1998.
- [7] R. Sarikaya, H.L. Hansen, "High resolution speech feature parameterization for monophone-based stressed speech recognition", IEEE Signal Processing Letters, Vol. 7 (7), pp. 182-185, 2000.
- [8] C. Wan-Chen, H. Ching-Tang, L. Eugene, "Multiband approach to robust text-independent speaker identification", Computational Linguistics and Chinese Language Processing, Vol. 9 (2), pp. 63-76, 2004.
- [9] I.A. Mahmoud, S.A. Hanaa, "Wavelet-based Mel-frequency cepstral coefficients for speaker identification using hidden markov models", Journal of Telecommunications, Vol. 1 (2), pp. 16-21, 2010.
- [10] Ö. Eskidere, "Identifying Acquisition Devices From Recorded Speech Signals Using Wavelet Based Features", Turkish Journal of Electrical Engineering Computer Sciences, DOI: 10.3906/elk-1312-193, 2015.
- [11] Barry, W.J., Putzer, M.: Saarbrücken Voice Database, Institute of Phonetics, Univ. of Saarland, <http://www.stimmdatenbank.coli.uni-saarland.de/>.
- [12] S. Mallat, "A theory for multiresolution signal decomposition: the wavelet representation", IEEE Trans. Pattern Anal. Mach. Intell., Vol. 11, pp. 674-693, 1989.
- [13] S. Mallat, "A wavelet tour of signal processing", Academic Press, San Diego, California, USA, 1998 .