

Alt-bant İşlemeye Dayalı Bir Ses Sınıflandırma Sistemi

Oytun Türk

Ömer Şaylı

Helin Dutağacı

Levent M. Arslan

Boğaziçi Üniversitesi
Elektrik-Elektronik Mühendisliği Bölümü
Bebek, İstanbul

<turkoytu, sayliome, dutagach, arslanle>@boun.edu.tr

Özetçe

Bu çalışmada, farklı kaynaklarca üretilen ses sinyallerini kaynağına göre otomatik sınıflandıran bir sistem tasarlanmıştır. Kullanılan ses sinyalleri konuşma, müzik, gürültü ve sessizlik içermektedir. Her ses sınıfının ayırt edilebilmesi için uygun alt-bant işleme yöntemleri kullanılmıştır. Algoritma, sınıflandırma işlemini temel olarak dört aşamada gerçekleştirmektedir : Eğitim, başlangıç-bitiş zamanlarının belirlenmesi, akustik parametrelerin hesaplanması ve sınıflandırma. Harmonik yapının ve başlangıç-bitiş zamanlarının belirlenmesi için alt-bant işlemeye dayalı yeni bir yöntem denenmiştir. Başarının artırılması için farklı akustik parametreler kullanılmıştır. Sınıflandırma, LBG algoritması kullanılarak tasarlanan vektör veri dosyalarıyla gerçekleştirilmektedir. Sistem, farklı ses karışımları ile sınanmış ve eş-zamanlı olmayan ses karışımlarında %88 otomatik sınıflandırma başarısı elde edilmiştir.

1. Giriş

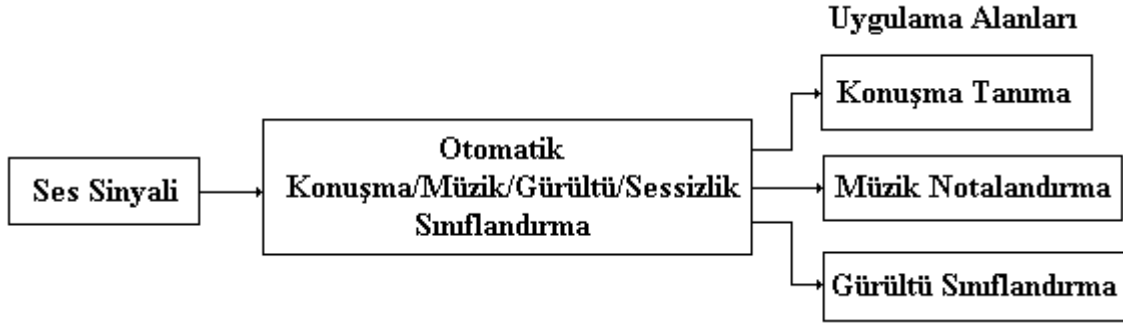
İnsan işitme sistemi, günlük yaşamda karşılaştığı pek çok karmaşık ses sinyalini başarılı bir şekilde birbirinden ayırt edebilmektedir. İşitsel Görünüm Çözümleme (Auditory Scene Analysis), bu davranışı taklit edebilen yapay sistemler tasarlamayı amaçlar. Ses sinyallerinin otomatik sınıflandırılması, pek çok alanda uygulamaları bulunan bir konudur. Uygulamalar arasında değişik kaynaklardan gelen ses sinyallerinin çözümlenmesi, konuşma/konuşmacı tanıma, müzik enstrümanı sınıflandırma, tek sesli/çok sesli müzik notalandırma ve çoklu ortam verilerinin içeriklerine göre sınıflandırılması sayılabilir. Bu çalışmada, ses sinyallerinden psiko-akustik kurallar ve alt-bantlardan elde edilen bilgileri kullanarak akustik parametreler hesaplayan yöntemlerin tasarlanması, farklı ses sinyalleri içeren veri tabanlarının akustik parametre hesaplama yöntemleriyle analizi, analiz sonuçlarının sınıflandırma için en kullanışlı parametre kümelerinin belirlenmesinde kullanılması; otomatik ses sınıflandırma sisteminin gerçekleşmesi ve testi amaçlanmıştır. İşitsel Görünüm Çözümleme konusundaki çalışmalar 1970'lerde başlamış [1], çeşitli çalışmalarda akustik parametreler kullanılarak ses sinyallerinin sınıflandırılması konusu işlenmiştir. Bu çalışmalara örnek olarak radyo/televizyon haberlerinin çözümlenmesi [7], görüntü sınıflandırmada ses sinyallerinden faydalanılması [6], otomatik müzik notalandırma [5] ve müzik enstrümanı tanıma [2] verilebilir.

Çalışmamızda, farklı frekans aralıklarından elde edilen akustik parametreler, sınıflandırma için kullanılmaktadır. Ses olaylarının başlangıç ve bitiş zamanları, zaman ve frekans uzayında maskeleyme özellikleri göz önünde bulundurularak tasarlanmış, alt-bant işlemeye dayalı bir yöntemle hassas biçimde belirlenmektedir. Her sıklık bandındaki harmonikler, harmonik yapıyı

betimlemek için hesaplanmaktadır. Ses sinyalinin farklı sıklık bantlarına ayrılması için doğrusal fazlı sonlu dürtü yanıtı bir süzgeç bankası tasarlanmış, bant genişlikleri ve merkez frekanslar insan işitme sistemi özellikleri göz önünde bulundurularak belirlenmiştir [9]. Tasarlanan ses sınıflandırma sisteminin şeması 2. bölümde verilmiştir. 3. bölüm, sınıflandırmada kullanılan akustik parametreleri ve sistemin ana parçalarını açıklamaktadır. 4. bölümde deneylerde kullanılan ses karışımları anlatılıp deneysel sonuçlar sunulmaktadır. Son bölümde sonuçlar tartışılmış ve sistemin geliştirilmesi için gelecekte yapılacak çalışmalar belirtilmiştir.

2. Sistem

Tasarlanan ses sınıflandırma sisteminin şeması Şekil 1’de verilmiştir. Sessiz kısımlar, enerji ve ortalama sıfır kesme oranı ölçüleriyle belirlenir. Ses olaylarının başlangıç ve bitiş zamanları bulunduğundan her ses olayı için akustik parametreler hesaplanıp sınıflandırma algoritması çalıştırılır. Çevrimdışı eğitim sırasında elde edilen ses sınıflarına ait bilgiler sınıflandırmada kullanılır.



Şekil 1. Ses Sınıflandırma Sistemi Şeması

3. Yöntem

3.1 Sessizlik ve Başlangıç/Bitiş Zamanlarının Belirlenmesi

Sinyaldeki sessiz kısımların belirlenmesi için sinyal 20 ms.’lik parçalar halinde işlenerek enerji, ve ortalama sıfır kesme oranı ölçümleri yapılır. Ölçümler medyan süzgeçleme ve doğrusal süzgeçleme kullanılarak düzlenir. Otomatik bölütleme gerektiren tüm uygulamalarda olduğu gibi, İşitsel Görünüm Çözümleme’deki en önemli adımlardan biri başlangıç/bitiş zamanlarının belirlenmesidir. Akustik olayların başlangıçları genel olarak ses yüksekliği, ses perdesi, ses tınısı ve sıklık dağılımı gibi faktörlerin en az birinin değiştiği anlara denk düşmektedir. Başlangıç/bitiş zamanlarının belirlenmesindeki başlıca zorluklar başlangıç/bitiş zamanlarının sinyaldeki kısmi değişimlere bağlı olması ve bu değişimlere bağlı olarak yapılacak karar verme hataları, bulunan başlangıç zamanlarının otomatik olarak düzeltilmesinin zorluğu ve gürültü olarak sıralanabilir. Bu zorluklar, alt-bant işlemeye dayalı sistemlerin kullanılması ile aşılabilmektedir [3]. Farklı frekans bantlarının kullanılmasındaki en önemli etken, insan işitme sisteminin ses sinyallerini farklı frekans bantlarına ayrıştırarak işlemesidir. Başlangıç zamanlarının belirlenmesi için aşağıdaki yöntem kullanılmıştır:



Şekil 2. Başlangıç Zamanlarının Belirlenmesi

Sinyallerin alt-bantlara ayrıştırılması için 6. derece eliptik süzgeçlerden oluşan 5 bantlı bir süzgeç bankası tasarlanmıştır. Merkez frekansları olarak 400, 800, 1200, 1600 ve 3200 Hz seçilmiştir. Bu seçim insan işitme

sisteminin özelliklerini yansıtmaktadır [9]. Başlangıç/bitiş zamanlarının belirlenmesindeki temel adımlar aşağıda özetlenmiştir:

- Ön-işleme (genlik normalizasyonu, süzgeç bankasıyla bantlara ayırıştırma, ve yarım dalgalı doğrultma).
- 200 ms.'lik Hanning penceresiyle evrişim ve seyreltme
- Türev ve göreceli fark fonksiyonu hesaplanması, eşik değeri kullanılarak başlangıç zamanlarının belirlenmesi.

3.2. Akustik Parametreler

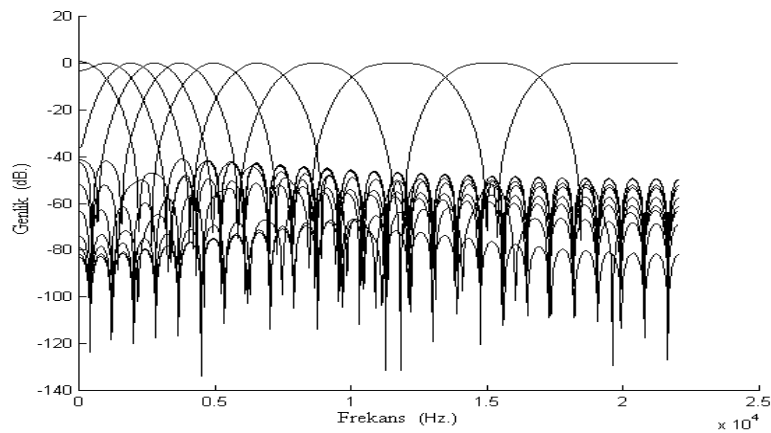
Tüm akustik parametreler, 20-30 ms.'lik ses sinyali parçalarından hesaplanıp akustik vektörler içinde toplanmaktadır. Alt-bant işlemeye dayalı her özellik, sinyalin süzgeç bankasından geçirilmesiyle elde edilen ilgili alt-bant sinyalinden hesaplanmaktadır. Kullanılan akustik parametreler aşağıda açıklanmıştır:

- Harmonikler: Bölüm 3.2.1'de açıklanan yöntem kullanılarak hesaplanan ilk 4 harmonik.
- Spektral Alçalma Frekansı: Tüm sinyal enerjisinin %95'ini içeren frekans üst sınırı.
- Alt-bant Enerjisi: Her alt-bant sinyalinin enerjisi
- Sıfır Kesme Oranı: Sinyalin, zaman uzayında genliğin sıfır olduğu noktayı ortalama kesme sayısı.
- Spektral Akı: Birbirini izleyen iki sinyal parçasının spektrumları arasındaki genlik farkı. Bu değer aşağıdaki bağıntıya göre hesaplanır:

$$(SF)_k = \| X_{k, n} - X_{k, n-1} \| \quad (1)$$

$X_{k, n}$: n'inci sinyal parçasının k'inci alt-bandı için spektrum genlik vektörü

- Spektral Kalıntı Enerjisi: Bu parametrenin hesaplanmasında, sinyalin LPC spektrumu ile gerçek spektrumu arasındaki fark kullanılır. Bu şekilde elde edilen kalıntı sinyali üzerinde her alt-bant için enerji hesaplanır. LPC model katsayılarının hesaplanması için konuşma içeren ses sinyallerinde 10'ar ms. arayla hesaplanan 25 ms.'lik sinyal parçaları Hamming penceresiyle çarpılarak kullanılır. Müzik için daha kısa ve birbirine yakın parçalar daha iyi sonuçlar elde edilmesine olanak sağlar (5'er ms. arayla hesaplanan 10-20 ms.'lik parçalar).
- Spektral Ağırlık Merkezi: Her alt-bant sinyalinin frekans uzayında Hz. türünden ağırlık merkezi



Şekil 3. Harmoniklerin hesaplanmasında kullanılan süzgeç bankası

3.2.1. Harmoniklerin Bulunması

Günlük yaşamda karşılaşılan en basit ses sinyali tek bir sinüs dalgası içeren saf tondur (pure tone). Çoğu zaman ses sinyali çok daha karmaşık yapıda olup birçok sinüs dalgası ve gürültü içerir. Birden fazla sinüs bileşeninden oluşan sinyaller karmaşık ton olarak adlandırılır. Eğer karmaşık tonu oluşturan sinüslerin temel frekansları ortak bir frekansın tamsayı katlarıysa, karmaşık ton harmonik kompleks olarak değerlendirilir [9]. Müzik ve konuşmada sıklıkla rastlanan harmonik kompleksler, otomatik tanıma, sınıflandırma ve sentez sistemlerinde akustik parametre olarak kullanılabilir.

Tasarladığımız harmonik yapı analizi birimi, insan işitme sisteminin özelliklerini göz önünde bulundurmaktadır. İnsan kulağında, iç kulak içinde bulunan kohlea yapısının farklı bölgelerinin farklı frekans aralıklarına duyarlı olduğu bilinmektedir. Bu özellik, duyulabilen frekans aralığı olan 20 Hz ile 20000 Hz arasındaki bölgeyi bant-geçirgen süzgeçlerle alt-bantlara ayırarak modellenabilir. Örnek bir süzgeç bankası Şekil 3'de verilmiştir. Her alt-bant sinyalinde özilinti işlevi hesaplanarak elde edilen temel frekans değerleri kullanılarak harmonikler aşağıdaki yöntemle belirlenebilir:

- Temel frekans değeri sıfır olmayan ilk alt-bant bulunur. Bu alt-banttaki temel frekans değeri f_0 ile gösterilecektir. Bu durumda f_0 , ilk harmonik bileşenin temel frekansıdır. Eğer bir sinyal için tüm alt-bantlarda temel frekans değerleri sıfır ise tüm harmonikler sıfır olarak değerlendirilir.
- Diğer alt-bantlardaki temel frekans değerleri ilk aşamada bulunan f_0 değerine bölünerek *normalize edilmiş harmonik değer* bulunur. n^{inci} alt-bant için bulunan temel frekans değerini f_n ile gösterirsek bu alt-bant için normalize edilmiş harmonik değer $h_n = f_n / f_0$ olacaktır.
- h_n değerine bağlı olarak karşılık gelen alt-banttaki sinyal (n^{inci} alt-bant sinyali) üç şekilde değerlendirilir:
 - $h_n = 0$ ise alt-bant gürültü içermektedir
 - $|h_n - K| \leq \varepsilon$ (K , h_n 'e en yakın tamsayı ve $\varepsilon < 0.05$) ise K^{inci} harmonik n^{inci} alt-bantta bulunmaktadır
 - $|h_n - K| > \varepsilon$ (K , h_n 'e en yakın tamsayı ve $\varepsilon < 0.05$) ise n^{inci} alt-bant harmonik içermemektedir

3.3. Sınıflandırma

Bu bölümde uygun akustik parametreler kullanılarak ses sinyallerinin sınıflandırılması amaçlanmaktadır. Eğitim verileri üzerinde hesaplanan akustik parametreler varyans analizinde (ANOVA) kullanılarak her parametrenin sınıflandırma açısından uygunluğu belirlenir. Bu aşamada çeşitli alt-bantlar için elde edilen f-oranı değerleri Tablo 1'de gösterilmektedir.

Parametre	Alt-bant	Konuşma/Müzik	Konuşma/Gürültü	Müzik/Gürültü
Sıfır Kesme Oranı	0-22KHz	25.39	2.15e+5	6.42e+5
Spek. Ağırlık Merkezi	0-2KHz	666.90	2.48e+4	1.28e+4
	2-6KHz	119.75	7351	1.88e+4
	6-12KHz	972.51	6010	2.42e+4
	12-22KHz	989.62	4743	1.51e+4
Spek. Alçalma Frek.	0-22KHz	55.70	1.94e+5	3.60e+5
Spek. Akı	0-2KHz	294.6	275.5	5.04e+3
	2-6KHz	67.75	6.62e+4	1.90e+5
	6-12KHz	43.49	1.25e+5	6.22e+5
	12-22KHz	14.24	1.01e+6	9.77e+5
Spek. Kalıntı Enerjisi	0-2KHz	0.37	190.70	415.20
	2-6KHz	56.93	89.56	55.15
	6-12KHz	71.78	123.60	171.01
	12-22KHz	71.76	147.51	448.42

Tablo 1. Akustik parametreler için f-oranı değerleri.

ANOVA’da sesler kaynaklarına göre ikili gruplar halinde ele alınmış ve her akustik parametre için ikili olarak grup ortalamalarının birbirine eşit olduğu hipoteziyle f-oranı değerleri elde edilmiştir. Sınıflandırmada kullanılacak vektör veri dosyalarının oluşturulması için her akustik parametre [0,1] aralığındaki değerlere eşlenir. Linde-Buzo-Gray algoritmasıyla [4] her ses sınıfına ait akustik vektörleri temsil edecek bir vektör kümesi bulunur. Her ses sınıfı için elde edilen vektör kümeleri, daha sonra test verilerini sınıflandırmada kullanılmak üzere veri dosyalarında saklanır. Çalışmamızda temel olarak üç ses sınıfı (konuşma, müzik ve gürültü) bulunduğundan eğitim sonrası üç farklı veri dosyası elde edilmektedir. Sinyallerdeki sessiz kısımlar bölüm 3.1’de açıklanan yöntemle tespit edilebildiğinden bu sinyaller ayrı bir ses sınıfı olarak ele alınmamıştır. Sınıflandırma için girdi sinyale ait akustik parametreler hesaplanır ve [0,1] aralığındaki değerlere eşlenir. Elde edilen akustik vektörlerle ses sınıfı veri dosyaları içindeki vektörler arasındaki Mahalanobis uzaklığı (MU) Denklem 2 ile hesaplanır. En düşük uzaklığın elde edildiği ses sınıfı, o andaki ses sınıfını oluşturur. Ses sınıfları konusunda verilen kararlar, komşu sinyal parçaları için verilen kararlarla karşılaştırılarak düzelir.

$$MU^2 = (X-E)^T S^{-1} (X-E) \quad (2)$$

X : girdi vektör, E : veri dosyası vektörü, S : tüm veri dosyası vektörleri için ortak değişinti matrisi

4. Deneyler

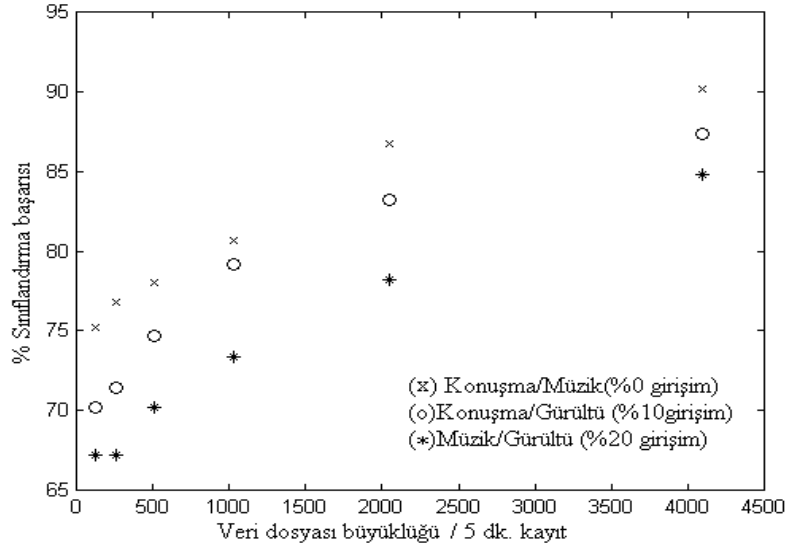
Ses sinyalleri 44.1KHz’de örneklenip 16 bit darbe kod kiplenimi (PCM) kullanılarak nicemlenerek eğitim ve test için iki ayrı ses veri tabanı hazırlanmıştır. Eğitim için hazırlanan veri tabanı konuşma, müzik ve gürültü içermektedir. Konuşma veri tabanı farklı yaşlardaki 30 değişik konuşmacı (15 bayan, 15 bay) tarafından söylenmiş cümlelerdir. Müzik veri tabanı, her müzik türü için 20 değişik örnek içerecek şekilde caz, klasik, rock ve pop müzik CD’lerinden yapılan kayıtlarla oluşturulmuştur. Beyaz gürültü ve ofis ortamındaki geri plan gürültüsü gürültü sınıfı veri tabanı hazırlanmasında kullanılmıştır. Her ses sınıfı, 25 dakikası eğitim 15 dakikası test için kullanılan yaklaşık 40 dakikalık kayıt içermektedir. Test veri tabanı, test için ayrılan verilerin hem eş-zamanlı olarak kaydedilmesi hem de her ses sınıfının ayrı ayrı kullanılmasıyla oluşturulmuştur. Bu durumda test veri tabanı, eş-zamansız ve eş-zamanlı iki değişik türde veri içermektedir. Eş-zamansız verilerde, farklı sınıflara ait ses örnekleri aynı anda bulunamazken eş-zamanlı veri tabanında farklı ses sınıflarının farklı enerji oranlarında karışımları kullanılmıştır. Tablo 2’de sınıflandırma sonuçları verilmiştir.

Girişim	Karışım Türü	Konuşma/Müzik	Konuşma/Gürültü	Müzik/Gürültü	Ortalama
0%	Eş-zamansız	90 . 12%	89 . 15%	88 . 30%	89 . 19%
10%	Eş-zamanlı	80 . 04%	87 . 34%	87 . 20%	84 . 86%
20%	Eş-zamanlı	78 . 23%	83 . 25%	84 . 76%	82 . 08%
30%	Eş-zamanlı	60 . 21%	60 . 91%	55 . 37%	58 . 83%

Tablo 2. Farklı ses karışımları için sınıflandırma başarısı.

Örneğin Tablo 2’de 3. satır incelenirse, girişim miktarının %20 olarak belirtildiği görülür. Bu durumda Konuşma/Müzik karışımında eş-zamanlı olarak kullanılan konuşma ve müzik sinyalleri sırasıyla toplam enerjinin %80’ini ve %20’sini oluşturmaktadır. Farklı sınıflara ait veri dosyalarının içerdiği vektör sayısının sınıflandırma başarısı üzerindeki etkisi Şekil 3’de gösterilmiştir. Genel olarak veri dosyaları büyüdükçe (her ses sınıfı daha fazla vektörle temsil edildikçe) sınıflandırma

başarısı artmaktadır. Fakat veri dosyaları büyüdükçe artan arama ve karşılaştırma zamanı, sistemin gerçek zamanlı uygulamalarda kullanılabilmesine engel olabilir.



Şekil 4. Veri dosyası büyüklüğünün sınıflandırma başarısına etkisi

5. Sonuçlar

Bu çalışmada genel amaçlı bir ses sınıflandırma sistemi tasarlamayı amaçladık. Gerçeklenen sistem kullanarak eş-zamanlı olmayan ses karışımlarında %88 sınıflandırma başarısı elde edildi. Farklı ses sınıflarından örnekler içeren karışımlarda sistemin başarısının düştüğü gözlemlendi. En düşük sınıflandırma başarısı, eş-zamanlı karışımlarda farklı sınıflardan sesler birbirine yakın enerji oranlarında bulunduğu durumda elde edildi (%58). Sistem performansının artırılabilmesi için daha karmaşık sınıflandırma yöntemlerinin denenmesi (Saklı Markov Modelleri, Yapay Sinir Ağları gibi), yeni akustik parametrelerin çıkarılması ve denenmesi gelecekte yapılması planlanan çalışmaları oluşturmaktadır. Tasarlanan sistem gerekli değişiklik ve eklemelerle müzik notalandırma ve enstrüman tanıma gibi uygulamalarda kullanılabilir.

6. Kaynakça

- [1] Bregman A. Auditory Scene Analysis. Cambridge MA : MIT Press. 1990.
- [2] Brown J.C., "Computer Identification of Wind Instruments Using Cepstral Coefficients" *Proc. 16th Int. Congress. on Acoustics*, sf. 1889-1890, Seattle.
- [3] Klapuri, "Sound Onset Detection by Applying Psychoacoustic Knowledge", *IEEE ICASSP 1999*.
- [4] Linde Y., Buzo A. ve Gray R.M., "An Algorithm For Vector Quantizer Design", *IEEE Trans. Commun.*, COM-28, 1, sf.84-95.
- [5] Martin K.D., "A Blackboard System For Automatic Transcription of Simple Polyphonic Music", MIT Media Lab. Perceptual Computing Sect. Tech. Rep., No.385, 1996.
- [6] Saraceno C. ve Leonardi R., "Audio as a Support to Scene Change Detection and Characterization of Video Sequences", in *IEEE ICASSP 1997*, Cilt 4, sf. 2597-2600.
- [7] Spina M.S. ve Zue V.W., "Automatic Transcription of General Audio Data", Tech. Report, Spoken Language Systems Group, Lab. For Computer Science, MIT.
- [8] Spina M.S. ve Zue V.W., "Automatic Transcription of General Audio Data : Preliminary Analyses", in *Proc. of the ICSLP*, sf. 594-597.
- [9] Zwicker E. ve Fastl H. Psychoacoustics. Springer-Verlag Berlin - Heidelberg - New York. 1998.