2007 - 230

TÜBİTAK

TÜRKİYE BİLİMSEL VE TEKNOLOJİK ARAŞTIRMA KURUMU

THE SCIENTIFIC AND TECHNOLOGICAL RESEARCH COUNCIL OF TURKEY

Elektrik, Elektronik ve Enformatik Araştırma Grubu

Electrical, Electronical and Informatics Research Group

Konuşmacı Tanımada MPEG-7 Ses Özniteliklerinin Kullanılabilirliği

PROJE NO: 104E142

Doç.Dr. Tolga Çiloğlu
Doç.Dr. Yunus Hakan Altınçay
Y. Doç. Dr. Cem Ergun

Mart 2007
ANKARA

Önsöz

Konuşmacı onaylama/doğrulama ve konuşmacı tanıma güncel uygulama alanları olan araştırma ve mühendislik konularıdır. Genel olarak biyometriklere dayalı tanıma ile ilgilidir. Konuşmacı onaylama süreçleri, konuşma sinyalinin incelenmesini ve sinyalin değişik özelliklerine ilişkin sayısal bilgilerin elde edilmesini içerir. Farklı konuşmacılara ait bu tür bilgiler karşılaştırılarak bir karar verilir. Diğer örüntü tanıma konularında olduğu gibi bu konuda da iki temel eksenden söz edilebilir; öznitelikler ve sınıflandırıcılar. "İyi" bir öznitelik tanımı ve bu özniteliklerin kişiler (sınıflar) arasındaki farklılaşmasını "iyi" kavrayabilen bir sınıflandırıcı olduğunda başarılı sonuçlar elde edilmektedir. Bu çalışma ağırlıklı olarak özniteliklerle ilgili bir araştırmadır. "Mel Frequency Cepstral Coefficients" olarak bilinen öznitelikleri destekleyici farklı öznitelikler üzerinde çalışılmıştır. Bu özniteliklerle konuşma sinyalinin harmonik içeriğine ait ek bilgiden yararlanılması ele alınmıştır. Özniteliklerin tanımı için MPEG-7 ses niteleyicilerinden yararlanılmıştır. Bu niteleyiciler içinde yer alan harmonik niteleyiciler kullanılmıştır. Öznitelik ve sınıflandırıcı bireştirme çalışmaları yapılmış, öngörüye uygun olarak harmonik bilginin kullanılması ile tanıma oranlarında kayda değer artış sağlanmıştır. Harmonik bilginin kullanımına yönelik daha önce yapılmış çalışmalar bulunmaktadır. MPEG-7 niteleyicilerini kullanarak bu kapsamda dünyada yapılmış ilk çalışma olmuştur. MPEG-7 yaygın olarak bilinen kavram olduğu için bu çalımada elde edilen sonuçlar önem taşımaktadır. Çalışma TÜBİTAK tarafından desteklenmiştir (proje no: 104E142).

İÇİNDEKİLER

ÖNSÖZ **3**

TABLO LİSTESİ **5**

ŞEKİL LİSTESİ **5**

ÖZET **6**

ABSTRACT **6**

1. GİRİŞ **7**

2. İLK ALTI AYLIK DÖNEMDE YAPILAN ÇALIŞMALAR **10**

2.1. YALNIZCA MFCC VE YALNIZCA MPEG-7 NİTELEYİCİLERİNİ KULLANAN KONUŞMACI ONAYLAMA SİSTEMLERİ VE DENEYLER. 10

2.2. MFCC VE MPEG-7 NİTELEYİCİLERİNİN ÖZNİTELİK DÜZEYİNDE BİRLEŞTİRİLMESİNE DAYALI KONUŞMACI ONAYLAMA SİSTEMİ VE DENEYLER. 11

2.3. YALNIZCA MFCC VE YALNIZCA MPEG-7 NİTELEYİCİLERİNİ KULLANAN İKİ FARKLI KONUŞMACI ONAYLAMA SİSTEMİNİN ÜRETTİKLERİ SONUÇLARIN BİRLEŞTİRİLMESİ İLE KONUŞMACI ONAYLAMA VE DENEYLER 13

3. İKİNCİ ALTI AYLIK DÖNEMDE YAPILAN ÇALIŞMALAR **15**

3.1. MPEG-7 NİTELEYİCİLERİNİN SINIFLANDIRILMASI 16

3.2. KULLANILAN VERİTABANI VE MODELLER 17

3.3. YALNIZCA MFCC VE YALNIZCA MPEG-7 NİTELEYİCİLERİNİ KULLANAN KONUŞMACI ONAYLAMA SİSTEMLERİ VE DENEYLER. 18

3.4. MFCC VE MPEG-7 NİTELEYİCİLERİNİN ÖZNİTELİK DÜZEYİNDE BİRLEŞTİRİLMESİ 19

3.5. YALNIZCA MFCC VE YALNIZCA MPEG-7 NİTELEYİCİLERİNİ KULLANAN FARKLI KONUŞMACI ONAYLAMA SİSTEMLERİNİN ÜRETTİKLERİ SONUÇLARIN BİRLEŞTİRİLMESİ (SKOR DÜZEYİNDE BİRLEŞTİRME) 20

3.4.1 Destek Vektör Sınıflandırıcı (SVM) Yöntemi *21*

3.4.2 Çok Yönlü Sarsım *21*

3.6. SKOR DÜZEYİNDE BİRLEŞTİRME SONUÇLARI 22

4. SON ALTI AYLIK DÖNEMDE YAPILAN ÇALIŞMALAR **24**

4.1. MPEG-7 HARMONİK NİTELEYİCİLERİ ARASINDAN EN İYİ ALTKÜME SEÇİMİ 24

5. SONUÇ **29**

TEŞEKKÜR **30**

KAYNAKLAR **30**

Tablo Listesi

TABLO 1: YALNIZCA MFCC VE YALNIZCA MPEG-7 NİTELEYİCİLERİNİ KULLANAN KONUŞMACI ONAYLAMA SİSTEMLERİ VE DENEYLERE AİT EŞİT HATA ORANLARI (EER). (HER KÜME 50 KONUŞMACI, 10 SANİYELİK 200 TEST PARÇASI İÇERİR) 11
TABLO 2: YALNIZCA MFCC, YALNIZCA MPEG-7 VE MFCC+MPEG-7 KÜMESİNDEN LDA İLE ELDE EDİLEN ÖZNİTELİKLERİN VERDİĞİ SONUÇLAR (EER). 12
TABLO 3: YALNIZCA MFCC, YALNIZCA MPEG-7 VE MFCC+MPEG-7 KÜMESİNDEN LDA İLE ELDE EDİLEN ÖZNİTELİKLERİN VERDİĞİ SONUÇLAR (EER). 13
TABLO 4: YALNIZCA MFCC, YALNIZCA MPEG-7 VE MFCC+MPEG-7 KÜMESİNDEN LDA İLE ELDE EDİLEN ÖZNİTELİKLERİN İLE ELDE EDİLEN EER DEĞERLERİ (%). 14
TABLO 5: TEMEL SİSTEMLERİN EER DEĞERLERİ (%). 18
TABLO 6: ÖZNİTELİK DÜZEYİNDE BİRLEŞTİRİLMİŞ SİSTEMLERİN EER DEĞERLERİ (%) 19
TABLO 7: SKOR DÜZEYİNDE BİRLEŞTİRİLMİŞ SİSTEMLERİN EŞİT HATA ORANLARI (%) 23
TABLO 8: NİTELEYİCİ EKSİLTME DENEYİNDE ALTKÜME TANIMLARI. 25
TABLO 9: F_1, F_2, F_3, F_4, F_5, F_6, F_7, F_8 VE F_9 KÜMELERİ İLE ELDE EDİLEN SONUÇLARIN EN KÜÇÜKTEN EN BÜYÜĞE DOĞRU SIRALAMALARI. 26
TABLO 10: NİTELEYİCİ EKSİLTME DENEYİNİN İKİNCİ AŞAMASINDA ALTKÜME TANIMLARI. 27
TABLO 11: F_8^1, F_8^2, F_8^3, F_8^4, F_8^5, F_8^6, F_8^7 VE F_8 KÜMELERİ İLE ELDE EDİLEN SONUÇLARIN KÜÇÜKTEN BÜYÜĞE DOĞRU SIRALAMALARI. 28

Şekil Listesi

ŞEKİL 1: KÜME1 İÇİN MFCC VE MPEG ÖZNİTELİKLERİNİ BİRLEŞTİREREK OLUŞTURULAN DET EĞRİLERİ. 11
ŞEKİL 2: KÜME1 İÇİN LDA İLE ELDE EDİLEN DET EĞRİSİ 12
ŞEKİL 3: KÜME2 İÇİN LDA İLE ELDE EDİLEN DET EĞRİSİ. 13
ŞEKİL 4: KÜME1 İÇİN, ÖNERİLEN BİRLEŞTİRME YÖNTEMİ KULLANILARAK ELDE EDİLEN DET EĞRİSİ. 15
ŞEKİL 5: KÜME2 İÇİN, ÖNERİLEN BİRLEŞTİRME YÖNTEMİ KULLANILARAK ELDE EDİLEN DET EĞRİSİ. 15
ŞEKİL 6: BİREYSEL MFCC VE MPEG SİSTEMLERİ İLE KÜME1 İÇİN ELDE EDİLEN DET EĞRİLERİ. 18
ŞEKİL 7: BİREYSEL MFCC VE MPEG SİSTEMLERİ İLE KÜME2 İÇİN ELDE EDİLEN DET EĞRİLERİ. 19
ŞEKİL 8: KÜME1(A) VE KÜME2 (B) İÇİN MFCC, MPEG VE SPHR SİSTEMLERİNİN SKOR DÜZEYİNDE BİRLEŞTİRİLMESİ 23
ŞEKİL 9: SEKİZ NİTELEYİCİDEN OLUŞAN HR KÜMESİNDEN HER SEFERİNDE BİR NİTELEYİCİ ÇIKARARAK ELDE EDİLEN YEDİ NİTELEYİCİLİ KÜMELERLE (F_1, F_2, F_3, F_4, F_5, F_6, F_7, F_8) VE *HR* (F_9) VE YALNIZCA MFCCLERLE (F_{10}) İLE ELDE EDİLEN SONUÇLAR. 26
ŞEKİL 10: YEDİ ELEMANLI F_8 KÜMESİNDEN HER SEFERİNDE BİR NİTELEYİCİ ÇIKARARAK ELDE EDİLEN YEDİ NİTELEYİCİLİ KÜMELER, (F_1^2, F_2^2, F_3^2, F_4^2, F_5^2, F_6^2, F_7^2), F_9 VE F_{10} KÜMELERİ İLE ELDE EDİLEN SONUÇLAR. 28

Özet

Konuşmacı onaylama probleminde konuşma sinyalinin harmonik içeriğine ait bilginin kullanımı ele alınmıştır. Harmonik içerik, seçilen MPEG-7 niteleyicileri ile temsil edilmiştir. Bu bilginin MFCC gibi bilinen öznitelikleri ne ölçüde destekleyeceği incelenmiştir. Temel sınıflandırıcı olarak Gauss karışım modelleri kullanılmıştır. Deneyler NIST 99 veritabanı ile gerçekleştirilmiştir. Öznitelik ve sınıflandırıcı birleştirme çalışmaları yapılmıştır. Sınıflandırıcı birleştirmede farklı yöntemler denenmiş, birleştirme sisteminin eğitiminde "AdaBoost" ve çok yönlü sarsım gibi teknikler kullanılmıştır. Öznitelik düzeyinde birleştirmede öznitelik seçimi üzerinde durulmuştur. Yapılan deneyler, MFCC ve harmonik niteleyicilerin öznitelik düzeyinde birleştirilmesi ile yalnızca MFCC'lerin kullanımı ile elde edilen EER değerlerine göre % 11 azalma sağlandığını göstermiştir. Sınıflandırıcı birleştirme ile bu azalma %17'ye ulaşmıştır. Azaltarak seçme yöntemi kullanılarak MPEG-7 niteleyicileri arasından seçilen harmonik niteleyicilerin hemen hemen tamamının başarıma katkısını olduğu belirlenmiştir.

Anahtar Sözcükler: Konuşmacı onaylama, konuşmacı tanıma, MFCC, MPEG-7, harmonik öznitelikler, öznitelik birleştirme, sınıflandırıcı birleştirme.

Abstract

The use of information about the harmonic content of speech signal in the problem of speaker verification has been considered. Harmonic content was represented by a selecetd set of MPEG-7 audio descriptors. The extent to which such information would support MFCC fetaures has been investigated. Gaussian mixture models were used as the basic classifier. NIST 99 speaker verification database was used in the experiments. Feature and classifier combination studies have been performed. Various techniques have been worked out for classifier combination; methods such as "AdaBoost" and multimodal perturbation were used for the training of the fusing system. Feature selection and feature extraction were considerd along with feature fusion. The combinaton of MFCCs and harmonic features yields a reduction of 11 % in equal error rates compared to that obtained by using MFCCs alone. The reduction is about 17 % with classifier combination. After a backward selection process, it has been found that almost all harmonic descriptors selected form the MPEG-7 set contribute to the increase of the performance.

Keywords: Speaker verification, speaker identification, MFCC, MPEG-7, harmonic features, feature fusion, classifier fusion.

1. Giriş

Bu çalışmanın amacı MPEG-7 standardı [MPEG7 2001] çerçevesinde tanımlanmış olan ses niteleyicilerin ("audio descriptors") konuşmacı onaylamada ("speaker verification") ne kadar yararlı olabileceklerini, aynı çerçeve içinde söz konusu niteleyicilerden hangilerinin daha yararlı olabileceğini incelemektir. Bu amaç doğrultusunda yapılacak çalışmalarda, MPEG-7 niteleyicilerinin belirli alt kümelerinin tek başlarına öznitelik grubu olarak kullanılması öngörülmemiştir. Çünkü, MFCC ("Mel Frequency Cepstral Coefficients") niteleyicileri ile konuşmacı tanımada/onaylamada ulaşılan başarım değerleri, dünyada MFCClerin akustik öznitelik olarak güvenilirliğini kanıtlamıştır [Davis1980] [Reynolds 1995] [Reynolds 1997] [Campbell 1997] [Reynolds 2000] [Campbell 1997]. MFCC ve MPEG-7 niteleyecilerini kavramsal olarak karşılaştırdığımızda MPEG-7 niteleyicilerinin, MFCClere göre çok daha iyi sonuçlar verebileceği yönünde beklentimiz bulunmamaktadır; tutarlı olarak elde edilecek böyle bir sonuç şaşırtıcı olur. Beklentimiz MPEG-7 niteleyicilerini, MFCClerle birlikte kullanarak daha iyi sonuçlar elde edebilmektir. Bunun temel nedeni MPEG-7 niteleyicilerinden bazılarının MFCClerden farklı bilgi taşıyabileceği doğrultusundaki öngörümüzdür. MFCC niteleyicilerinin esas olarak sinyal izge genliğinin tepe hattına ("envelope") ait bilgi verdiği bilinir [Davis 1980]. Tepe hattı bilgisi seslerin ayırt edilmesinde belirleyici öneme sahiptir; ancak sesleri niteleyen (karakterize eden) tek bilgi değildir. Örneğin, sesin sesli ya da sessiz olması hakkında doğrudan bilgi vermez (öte yandan, konuşma sinyallerinde bir sesin sesliliği ve tepe hattı bilgisinin tamamen ilintisiz olmadığını da biliyoruz). Tepe hattının içermediği başka bir bilgi de sesli ve sessiz bölgelerin frekans bantlarına dağılımıdır. Bu dağılımın hem konuşmacıya özgü hem de sese özgü niteliklerinin olduğu düşünülmektedir. Buna uygun bir örnek olarak NATO standardı olarak kabul edilmiş olan MELP kodlama yöntemi verilebilir. Bu yönteme göre konuşma kodlanırken ara bantlarda seslilik analizi yapılır ve kodlama bilgisi olarak kullanılır. Bir başka deyişle seslilik dağılımının sesleri tanımlayıcı özelliği olduğu kabul edilmiştir.

Kim ve meslektaşları MPEG-7 niteleyicileri ile konuşmacı tanıma amacıyla birtakım deneyler yapmışlardır [Kim 2003, Kim 2004]. Bu deneylerde logaritmik frekans ölçeğinde hesaplanmış MPEG-7 grubundan "Audio Spectrum Envelope" (ASE) niteleyicileri kullanılmıştır. MPEG-7 niteleyicileri cinsiyet tanıma probleminde de kullanılmış ve MFCC niteleyicileri ile karşılaştırılmıştır. 22,05 kHz örnekleme hızı ile elde edilmiş ses sinyalleri üzerinde yapılan çalışmalarda çoğunlukla MFCC niteleyicilerinin daha yüksek başarım sağladığı gözlemlenmiştir. Bildiğimiz kadarıyla, literatürde var olan çalışmalarda sadece ASE niteleyicileri kullanılmıştır. Oysa, MPEG-7 standardında konuşmacı tanımada etkili olabilecek başka niteleyiciler de tanımlanmıştır. Bu çalışmada MPEG-7 standardında yer alan daha geniş bir niteleyici gurubu dikkate alınmaktadır.

MPEG-7 standardında tanımlanmış olup tepe hattı bilgisi dışında bilgiler içerdiğini ve konuşmacı tanımada etkili olabileceğini düşündüğümüz niteleyiciler sinyalin harmonik içeriği hakkında bilgi veren niteleyicilerdir. MPEG-7 ses niteleyicilerinin tamamına bakıldığında MFCClerin verdiği bilgiye benzer bilgi veren bir kümenin yanı sıra sinyalin harmonik özellikleri hakkında bilgi veren bir küme olduğu görülür. Çalışmanın amacı MPEG-7 standardı [MPEG-7 2001] çerçevesinde tanımlanmış olan ses niteleyicileri içinde ("audio descriptors") ağırlıklı olarak sinyalin harmonik özelliklerini betimleyen bir kümenin konuşmacı onaylamada ("speaker verification") ne kadar yararlı olabileceğini, aynı çerçeve içinde söz konusu niteleyicilerden hangilerinin daha yararlı olabileceğini incelemektir.

Çalışmalarda MPEG-7 öznitelikleri üç grup olarak ele alınmıştır. Bir grup, "audio spectrum envelope coefficients" (ASE) parametreleridir. Bunlar MFCC parametreleri ile benzer özelliklere sahiptir. Aralarındaki temel fark, farklı frekans ölçeklemelerinden kaynaklanmaktadır. Diğer grup, izgenin harmonik içeriğine ait bazı özellikleri ortaya çıkaran parametrelerdir ve bu raporda *Hr* olarak anılmaktadır. Bu gruptaki parametrelerin konuşma sisteminin kaynak kısmına (ses telleri) ait bilgi içerdiği düşünülmektedir. Son grup, izgenin tepe hattına ait özet bilgi içeren üç değerden oluşmaktadır; bu raporda *Sp* olarak anılmaktadırlar.

Son iki grup MFCC parametreleri ile öznitelik düzeyinde birleştirilmişlerdir. Öznitelik düzeyindeki birleştirme, farklı öznitelik kümelerinin aynı vektör içinde toplanması ve sınıflandırıcının ya da modelin bu vektörlerle eğitilmesidir. Bu çalışmada GMM modeller eğitilmiştir.

Skor düzeyinde birleştirmede üç MPEG-7 grubu da kullanılmıştır. Skor düzeyinde birleştirmede her öznitelik grubu için bir GMM topluluğu (kişilere özgü modeller ve genel ortak model (UBM)) eğitilmiştir. Böylece bir konuşmacı onaylama testinde, verilen konuşmaya karşılık her öznitelik türü için iki skor (olabilirlik değeri) elde edilmiştir. Farklı öznitelik türleri için elde edilen bu değerler yeni bir öznitelik vektörü olarak tanımlanmış ve buna uygun bir sınıflandırma sistemi geliştirilmiştir. Konuşmacı onaylama probleminde bulunan sınıf dengesizliği (yanıltıcı verisinin doğrucularadan çok daha fazla olması) göz önüne alınarak bir sınıflandırıcılar toplulukları eğitilmiştir.

Hem öznitelik düzeyinde hem skor düzeyindeki birleştirmelerde yalnızca MFCC parametreleri ile elde edilen sonuçlara göre iyileşme sağlanmıştır. Skor düzeyindeki birleştirmede ASE parametrelerinin de yararlı olduğu gözlenmiştir.

Hr grubu öznitelikleri ağırlıklı olarak harmoniklerle ilgili ölçümleri içermektedir. Harmoniklere ait farklı özellikleri açığa çıkaran bu özniteliklerin insan konuşma üretim sisteminin kaynak kısmına (ses telleri) ait bilgi içerdiği düşüncesinden yola çıkılmaktadır. Ses telleri olarak

adlandırılan doku yapısı kişiden kişiye değişen özellikler gösterir. Ses telleri sisteminin titreşen kısmında kaslar ve çevreleyen yumuşak dokular bulunmaktadır. Bu dokuların, insanın diğer biyometrik özellikleri gibi, kişiye özgü olduğu bilinmektedir. Doku yapısının kalınlığı, sıkılığı/sertliği ve sarkıklığı gibi özellikleri titreşim sırasında gırtlakta oluşan basınç dalgasının harmonik özelliklerini ve gürültü içeriğini etkilemektedir. Konuşma sinyalindeki harmonik bileşenlerin yanısıra gürültünün frekans bandı üzerindeki dağılımı, harmonik yapının frekans ekseni boyunca yer yer bozulmasına neden olabilmektedir. Buna bağlı olarak, harmoniklerin ağırlık merkezi ve diğer özellikleri hakkında bilgi taşıyan *Hr* parametrelerinin etkileneceği düşünülmektedir.

Konuşma seslerinin kavranmasında, sinyallerin yapısının harmonik ve harmonik olmayan kısımlar olarak ele alınması yerleşmiş yaklaşımlardan birisidir. Konuşma seslerinin üretiminde kişinin fizyolojisi ve artikülasyon biçimi harmonik ve harmonik olmayan bileşenlerin katkısını belirlemektedir. Harmonik içerik temel olarak ses tellerinin açılıp kapanması sonucu oluşur. Harmonik olmayan içerik, havanın ses telleri bölgesinden ve ses yolunun diğer kısımlarından geçişi sırasında, bu yol üzerindeki değişik geometrik oluşumlara, daralmalara, kıvrımlara bağlı olarak oluşan türbülans ve benzer olayların sonucudur. Harmonik içeriğin yapısı, harmoniklerin birbirlerine göre düzeyleri, hem harmonik oluşumunu sağlayan ses telleri sisteminin dinamik yapısına hem de akustik süzgeç olarak çalışan ses yolunun frekans tepkisine göre oluşur.

MFCC gibi parametrelerle izgenin tepe hattı hakkında bilgi sağlanmaktadır. Tepe hattının oluşumuna, harmonik ve harmonik olmayan içeriğin, kabaca, toplamsal bir birleşimle katkıda bulunduğu düşünülebilir. MFCC ve harmonik içeriğin birlikte değerlendirilmesi ile hem sonuç (tepe hattı) hem de sonucu oluşturan bileşenler değerlendirilmiş olur. Harmonik içerik ile ilgili bilgiler, dolaylı olarak harmonik olmayan içerik hakında da bilgi vermektedir. Aynı sonuç, bileşenlerin farklı katkıları ile elde edilebileceği için, MFCC ve harmonik bilgisi, yalnızca MFCC bilgisinden daha fazla bilgi taşımaktadır.

Bu durumun konuşmacı tanımada da yararlı olacağı belirtilebilir. Konuşmacıların benzer fonetik dizileri birbirleri ile akustik olarak da benzerlikler taşır. Konuşmacıları ayırt etmeyi sağlayan farklılıklar ses tellerinin ve ses yolunun statik ve dinamik özelliklerinin farklılaşması sonucunda ortaya çıkar. Bu farklılaşma tepe hattında olabileceği gibi, harmonik ve harmonik olmayan içeriğin farklılaşmasından da kaynaklabilir. Bu tür farklılaşmaları algılamak için harmonik özelliklerin MFCCler ile birlikte kullanılması anlamlıdır.

Projenin ilk iki altı aylık süreleri sonunda birer adet gelişme raporu hazırlanmıştır. Bu raporlarda çalışma konusu ile ilgili literatür, projede kullanılan yöntemler, yapılan deneyler ve elde

edilen sonuçlar ayrıntılı olarak açıklanmıştır. Bu sonuç raporunda, ilk iki dönemde yapılanlar ve elde edilen sonuçlar bütün olarak özetlenecek, son dönemde yapılan çalışmalar ve sonuçlar anlatılacaktır. Raporda literatürde yaygın olarak geçen deyimlerin kısaltmaları İngilizcedeki haliyle kullanılmıştır.

2. İlk Altı Aylık Dönemde Yapılan Çalışmalar

İlk altı aylık dönemde yapılan çalışmalar üç başlık altında tanımlanmıştır:

- Yalnızca MFCC ve yalnızca MPEG-7 niteleyicilerini kullanan konuşmacı onaylama sistemleri ve deneyler.
- MFCC ve MPEG-7 niteleyicilerinin öznitelik düzeyinde birleştirilmesine dayalı konuşmacı onaylama sistemi ve deneyler.
- Yalnızca MFCC ve Yalnızca Mpeg-7 Niteleyicilerini Kullanan İki Farklı Konuşmacı Onaylama Sisteminin Ürettikleri Sonuçların Birleştirilmesi İle Konuşmacı Onaylama ve Deneyler

Bunlar aşağıda sırasıyla açıklanmaktadır.

2.1. Yalnızca MFCC ve yalnızca MPEG-7 niteleyicilerini kullanan konuşmacı onaylama sistemleri ve deneyler.

Bu sistemler Gauss karışım modelleri üzerine kurulmuştur. Her konuşmacı için bir adet "konuşmacıya özgü" model, ayrıca bir adet "genel ortak model" üretilmiştir. Bir onaylama testi, verilen konuşma sinyalinin, bildirilen konuşmacı kimliğine ait modelden ve genel ortak modelden elde edilen olabilirlik değerleri karşılaştırılarak yapılmaktadır.

Deneylerde her kişisel model eğitimi için yaklaşık 30 saniyelik konuşma kullanılmıştır. MFCC sisteminde 16 niteleyici, MPEG-7 sisteminde, 20 niteleyici kullanılmıştır. Konuşmacı onaylama deneyleri ellişer konuşmacıdan oluşan iki farklı küme, (KÜME1 ve KÜME2) üzerinde yapılmıştır. Her konuşmacıya ait kayıt, yaklaşık 10 saniyelik, yalnızca konuşma içeren dört parça kullanılarak her küme için 200 test parçası elde edilmiştir. Her parça hem doğrucu, hem de yanıltıcı olarak diğer konuşmacı modellerine karşı kullanılmıştır.

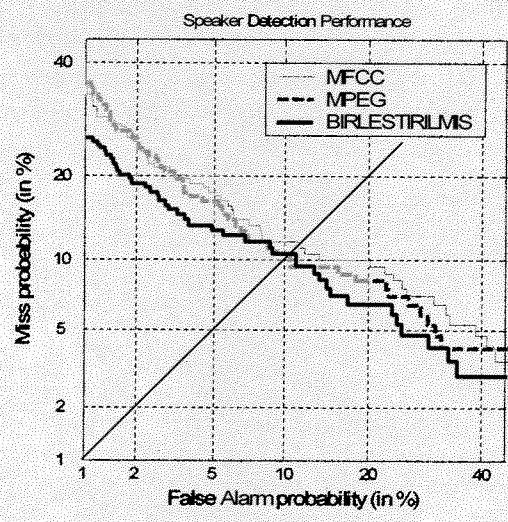
Tablo 1: Yalnızca MFCC ve yalnızca MPEG-7 niteleyicilerini kullanan konuşmacı onaylama sistemleri ve deneylere ait eşit hata oranları (EER). (Her küme 50 konuşmacı, 10 saniyelik 200 test parçası içerir)

	Küme-1	**Küme-2**	**Ortalama**
MFCC	11.24	9.49	10.37
MPEG-7	10.54	9.49	10.02

Tablo 1'deki sonuçlara göre iki küme üzerindeki ortalama başarım MFCClerle %10.37, MPEG-7lerle %10.02dir. Bu EERde %3.4 oranında küçük bir azalmaya karşılık gelmekle birlikte herhangi bir genelleme yapmak için yeterli olmadığını düşünülmektedir.

2.2. MFCC ve MPEG-7 niteleyicilerinin öznitelik düzeyinde birleştirilmesine dayalı konuşmacı onaylama sistemi ve deneyler.

Öznitelik düzeyinde birleştirme MFCC ve MPEG-7 niteleyicilerini birlikte içeren bir vektör oluşturularak yapılmıştır. Toplam 36 elemanlı (16 MFCC ve 20 MPEG-7) öznitelik vektörü kullanılmıştır. Küme-1 için elde edilen DET (detection error tradeoff) eğrisi Şekil 1'de verilmiştir.



Şekil 1: KÜME1 için MFCC ve MPEG özniteliklerini birleştirerek oluşturulan DET eğrileri.

Şekil 1'de görüldüğü gibi MFCC ve MPEG-7 niteleyicileri birlikte kullanıldığında düşük ve yüksek yanlış alarm bölgelerinde başarım artışı sağlanmıştır. ASE niteleyicileri ile MFCC niteleyicileri birbirleri ile ilintisi yüksek parametrelerdir (İki grup da izge tepe hattına ait bilgi vermektedir.) Bu tür durumlarda elde edilen başarının veri miktarına bağlı olduğu bilinmektedir. Dolayısıyla, başarının belirgin hale getirilmesinde öznitelik seçmenin etkili olabileceği düşünülerek, üç farklı yöntem ile öznitelik seçimi yapılmıştır. Bu yöntemler şunlardır:

a. Doğrusal destek vektör makinası ("Support Vector Machine", SVM) ile öznitelik seçimi
b. Doğrusal farklılaştırma ("Linear Discriminant Analysis", LDA) ile öznitelik seçimi
c. Ana bileşenler incelemesi ("Principal Component Analysis", PCA) ile öznitelik seçimi

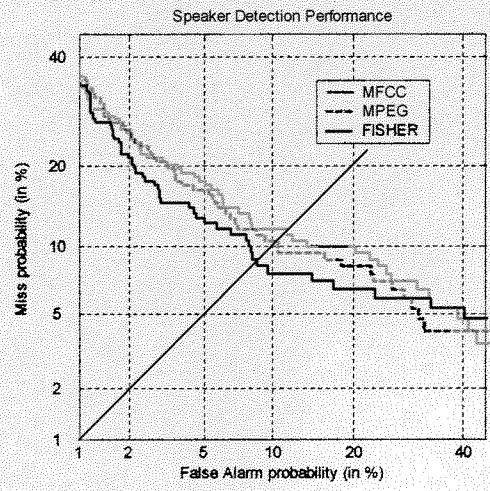
İlk yöntemle (SVM) konuşmacı çiftlerini sınıflandıran doğrusal sınıflandırıcılar bulunmuş ve bu sınıflandırıcaların gösterdiği en baskın 20 öznitelik seçilmiştir. Toplam 36 öznitelik arasından seçilen bu 20 öznitelik ile başarımda artış görülmemiştir.

İkinci yöntem, bilinen LDA yöntemidir. Bu yöntem ile başarımda artış sağlanmıştır. Elde edilen sonuçlar EER olarak Tablo 2'de görülmektedir. Ortalama olarak, yalnızca MFCC kullanan sisteme göre EER %13.1 azalmıştır.

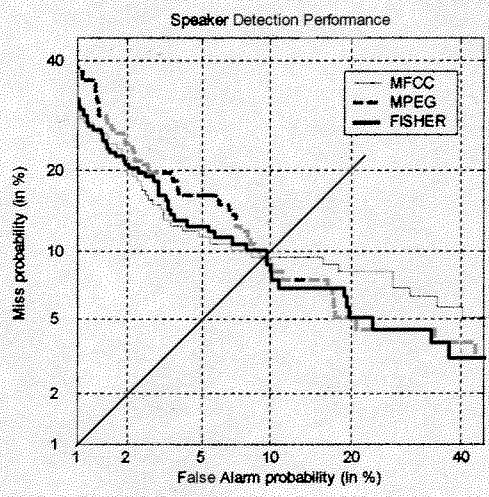
Tablo 2: Yalnızca MFCC, yalnızca MPEG-7 ve MFCC+MPEG-7 kümesinden LDA ile elde edilen özniteliklerin verdiği sonuçlar (EER).

	Küme-1	**Küme-2**	**Ortalama**
MFCC	11.24	9.49	10.37
MPEG-7	10.54	9.49	10.02
MFCC+MPEG-7 LDA	8.43	9.58	9.01

Aynı deneye ait DET eğrileri Şekil 2'de ve Şekil 3'de görülmektedir. Bu sonuçlara dayanarak MFCC ve MPEG-7 niteleyicilerinin belli ölçüde birbirlerini tamamlayıcı özelliği olduğu söylenebilmektedir.



Şekil 2: KÜME1 için LDA ile elde edilen DET eğrisi



Şekil 3: KÜME2 için LDA ile elde edilen DET eğrisi.

Üçüncü yöntem olan PCA ile, 36 olan vektör boyutu 20ye indirilmiştir. Bu yöntemle başarım artışı sağlanmamıştır.

2.3. Yalnızca MFCC ve Yalnızca Mpeg-7 Niteleyicilerini Kullanan İki Farklı Konuşmacı Onaylama Sisteminin Ürettikleri Sonuçların Birleştirilmesi İle Konuşmacı Onaylama ve Deneyler

MFCC ve MPEG-7 sistemlerinin ayrı ayrı ürettikleri sonuçları birleştirmek için iki yol izlenmiştir. Bunlardan birisi, iki sistemin ürettiği olabilirlik değerlerinin ortalamasının alınmasıdır. Diğeri ise bir üst-sınıflandırıcılar topluluğunun oluşturulmasıdır.

Ortalama yöntemi ile, olabilirlik oranı testi iki sistemin ürettiği olabilirlik oranlarının ortalaması kullanılarak yapılmıştır. Buna göre elde edilen EER değerleri Tablo 3'de verilmiştir.

Tablo 3: Yalnızca MFCC, yalnızca MPEG-7 ve MFCC+MPEG-7 kümesinden LDA ile elde edilen özniteliklerin verdiği sonuçlar (EER).

	Küme-1	**Küme-2**	**Ortalama**
MFCC	11.24	9.49	10.37
MPEG-7	10.54	9.49	10.02
Ortalama ile sonuç birleştirme	8.89	8.10	8.50

KÜME1 ve KÜME2 için EER değerleri % 8.89 ve % 8.10 olmuştur. Ortalama değer % 8.50 olup, MFCC sistemine göre EER % 18.0 azalmıştır.

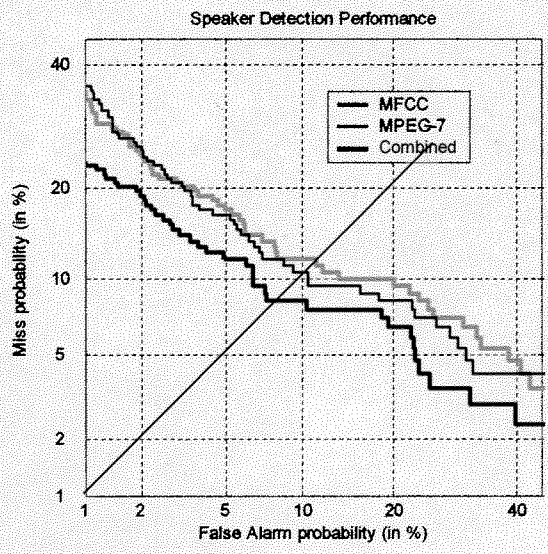
Sınıflandırıcılar topluluğu yöntemi ile yapılan çalışmada herbiri 25 doğrusal sınıflandırıcı içeren 15 adet altsınıflandırıcı toplulukları eğitilmiştir. Bu sistemde herhangi bir doğrusal sınıflandırıcının girdisi, "konuşmacıya özgü" model ve "genel ortak model" tarafından üretilen olabilirlik değerlerini içeren dört boyutlu vektörlerdir. Her altsınıflandırıcı topluluğu, eğitim verisinin farklı bir altkümesi ile eğitilmiştir. Sınıflandırıcı topluluklarının ürettiği sonuçların ortalaması son değer olarak alınmıştır; bu değer sıfırdan büyükse "doğrucu", küçükse "yanıltıcı" kararı verilmişir. 15 adet altsınıflandırıcı topluluğunun herhangi birinin ürettiği değer, o topluluğun elemanlarının ürettiği değerlerin ağırlıklı ortalaması hesaplanarak bulunur. Ağırlıklar AdaBoost algoritması ile saptanmışlardır. KÜME1 (KÜME2) için yapılan deneylerde yukarıdaki sistemi eğitmek için KÜME2 (KÜME1) verilerinden elde edilen çıkış değerleri kullanılmıştır.

Bu yöntemle yapılan deneyler sonucunda KÜME1 ve KÜME2 için EER değerleri Tablo 4'de görülmektedir. Ortalama EER, yalnızca MFCCler ile elde edilen EER değerine göre %23.44 oranında azalmıştır. Basit ortalama ile yapılan üst-sınıflandırmada elde edilen %8.50 EER değerine göre ise azalma % 6.59 olmuştur.

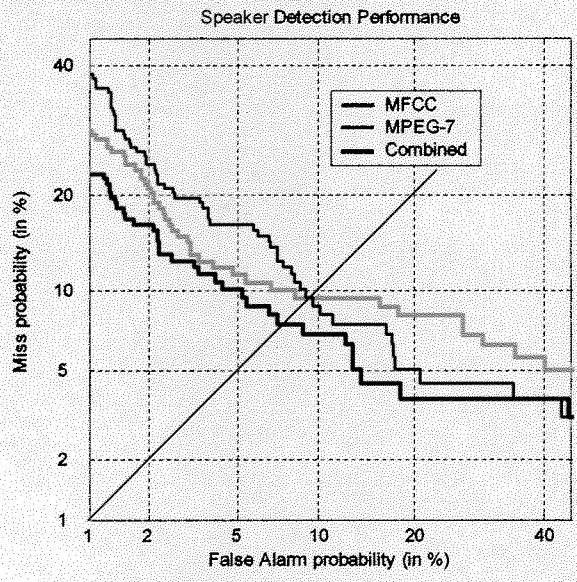
Tablo 4: Yalnızca MFCC, yalnızca MPEG-7 ve MFCC+MPEG-7 kümesinden LDA ile elde edilen özniteliklerin ile elde edilen EER değerleri (%).

	Küme-1	**Küme-2**	**Ortalama**
MFCC	11.24	9.49	10.37
MPEG-7	10.54	9.49	10.02
Sınıflandırıcı Topluluğu ile sonuç birleştirme	8.28	7.59	7.94

KÜME1 ve KÜME2 üzerinde yapılan bu deneylerin DET eğrileri sırasıyla Şekil 4'de ve Şekil 5'de gösterilmiştir.



Şekil 4: KÜME1 için, önerilen birleştirme yöntemi kullanılarak elde edilen DET eğrisi.



Şekil 5: KÜME2 için, önerilen birleştirme yöntemi kullanılarak elde edilen DET eğrisi.

3. İkinci Altı Aylık Dönemde Yapılan Çalışmalar

Projenin ilk altı aylık döneminde yapılan çalışmalarda MPEG-7 niteleyicileri bütün olarak ele alınmış, bir ön değerlendirmeye bağlı olarak ayrıştırma yapılmamıştır. İkinci altı aylık dönemde MPEG-7 niteleyicileri kategorik olarak sınıflandırılmış ve bu çerçevede öznitelik ve karar düzeyinde birleştirme çalışmaları yapılmıştır. Deneyler daha büyük konuşmacı kümeleri

üzerinde yapılmıştır. Yapılan çalışmalar ilk altı aylık kısımda olduğu gibi üç başlık altında sürdürülmüştür:

1. Yalnızca MFCC ve yalnızca MPEG-7 niteleyicilerini kullanan konuşmacı onaylama sistemleri ve deneyler.
2. MFCC ve MPEG-7 niteleyicilerinin öznitelik düzeyinde birleştirilmesine dayalı konuşmacı onaylama sistemi ve deneyler.
3. Yalnızca MFCC ve Yalnızca Mpeg-7 Niteleyicilerini Kullanan İki Farklı Konuşmacı Onaylama Sisteminin Ürettikleri Sonuçların Birleştirilmesi İle Konuşmacı Onaylama ve Deneyler

Aşağıda, sırasıyla, MPEG-7 niteleyicilerinin kategorik olarak sınıflandırılması, ikinci altı aylık dönemde kullanılan genişletilmiş veritabanı ve üç başlık altındaki çalışmalar anlatılmaktadır.

3.1. MPEG-7 Niteleyicilerinin Sınıflandırılması

İkinci dönemdeki çalışmalarda MPEG-7 öznitelikleri üç grup olarak ele alınmıştır. Bir grup, "audio spectrum envelope coefficients" (ASE) parametreleridir. Bunlar MFCC parametreleri ile benzer özelliklere sahiptir. Aralarındaki temel fark, farklı frekans ölçeklemelerinden kaynaklanmaktadır. Bu çalışmanın başlamasında birinci sırada öneme sahip değillerdir.

Başka bir grup, izgenin harmonik içeriğine ait bazı özellikleri ortaya çıkaran parametrelerdir (bu raporda *Hr* olarak anılmaktadır):

1. "audio fundamental frequency" ve "related confidence measure"
2. "audio harmonicity" ve "upper limit of harmonicity"
3. "harmonic spectral centroid"
4. "harmonic spectral deviation"
5. "harmonic spectral spread"
6. "harmonic spectral variation"

1. niteleyici sinyalin temel frekansını ve frekans hesaplamasının güvenilirliğini ifade etmektedir. 1. niteleyiciye bağlı olarak hesaplanan diğer niteleyiciler sinyalin harmonik içeriğini ifade etme amaçlıdır. "*harmonic spectral centroid*" sinyalin harmoniklerinin bulunduğu bölgenin ağırlık merkezini verir. "*harmonic spectral deviation*" harmonik tepelerinin ortalama tepe hattından ne kadar saptığını belirtir. "*harmonic spectral spread*" harmonik frekanslarının, harmonik genlikleri ile ağırlıklandırılmış standart sapmasıdır. "*harmonic spectral variation*" ardışık iki çerçeve sinyalinin harmonik tepeleri arasındaki ilintidir ("correlation"). Bu küme izgenin tepe hattının altında kalan yapının özellikleri hakkında bilgi vermektedir. Bu araştırmanın başlamasında belirleyici olmuşlardır.

Son grup, izgenin tepe hattına ait özet bilgi içeren üç değerden oluşmaktadır (bu raporda *Sp* olarak anılmaktadır):

1. "audio spectrum centroid"
2. "spectral centroid"
3. "audio spectrum spread"

Bu parametreler isimlerinden anlaşıldığı gibi, izgenin tepe hattının ağırlık merkezini ve yaygınlığını ifade ederler. İlk ikisi ağırlık merkezi olmakla birlikte, ilkinde logaritmik, ikincisinde doğrusal frekans ölçeği kullanılmaktadır.

Raporda, *Sp* ve *Hr* niteleyicilerinin birleştirilmesi ile oluşan küme *SpHr* olarak anılmaktadır.

3.2. Kullanılan Veritabanı ve Modeller

İkinci dönemdeki deneylerde daha fazla konuşmacı kullanılmıştır. İlk altı aylık dönemde kullanılan 50şer konuşmacıdan oluşan iki küme yerine 150 konuşmacıdan oluşan iki küme oluşturulmuştur. Deneyler NIST99 veri tabanından seçilmiş M =150 konuşmacı içeren iki farklı küme, KÜME1 ve KÜME2 ile gerçekleştirilmiştir. Kayıtlar 8 KHz örnekleme hızında 16 bit çözünürlüktedir. KÜME1de ve KÜME2de olmayan 46 erkek ile 46 bayan konuşmacı kullanılarak UBM eğitilmiştir. Bu model 2048 bileşenden oluşan bir GMMdir. Konuşmacılara ait modeller UBM'den Bayesçi uyarlama yoluyla oluşturulmuştur. Her konuşmacı modeli için yaklaşık 60 saniyelik konuşma kullanılmıştır.

3.3. Yalnızca MFCC ve yalnızca MPEG-7 niteleyicilerini kullanan konuşmacı onaylama sistemleri ve deneyler.

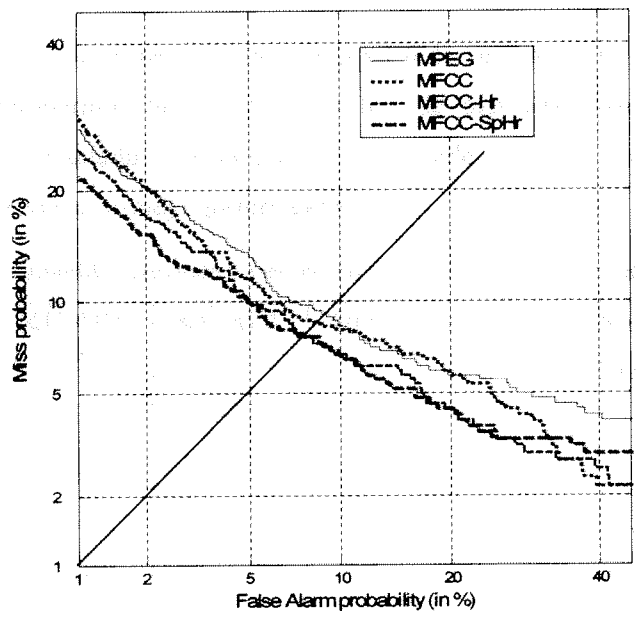
İlk altı aylık dönemdeki çalışmalarda olduğu gibi, birleştirilmiş sistemlerin başarımlarının değerlendirilebilmesinde referans olarak kullanılmak üzere iki sistem geliştirilmiştir. Bu tekrarın nedeni veritabanının genişletilmiş olmasıdır. Bunlardan birisinde 16 MFCC parametresi, diğerinde ise *ASE* ve SpHr gruplarındaki toplam 24 öznitelik kullanılmıştır. İkinci sistem MPEG olarak anılacaktır.

Yalnızca MFCC ve yalnızca MPEG-7 parametrelerini kullanan bireysel sistemlerin Eşit Hata Oranları (EER, "Equal Error Rate") Tablo 5'de verilmiştir. Bu deneylerde elde edilen EER değerleri, ilk atı aylık dönemde yapılan aynı türdeki deneylerde elde edilenlerden daha düşüktür, başarım artmıştır.

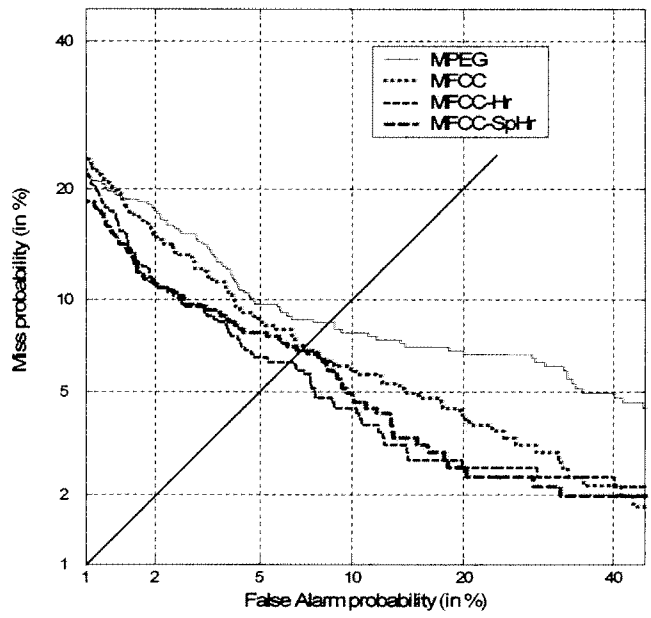
Tablo 5: Temel sistemlerin EER değerleri (%).

	KÜME1	**KÜME2**	**Ortalama**
MFCC	8.58	7.13	7.86
MPEG	9.00	8.53	8.77

Bireysel MFCC ve MPEG sistemlerinin başarımlarını gösteren DET eğrileri KÜME1 ve KÜME2 için sırasıyla Şekil 6'da ve Şekil 7'de görülmektedir.



Şekil 6: Bireysel MFCC ve MPEG sistemleri ile KÜME1 için elde edilen DET eğrileri.



Şekil 7: Bireysel MFCC ve MPEG sistemleri ile KÜME2 için elde edilen DET eğrileri.

3.4. MFCC ve MPEG-7 Niteleyicilerinin Öznitelik Düzeyinde Birleştirilmesi

Öznitelik düzeyinde birleştirmede yukarıda tanımlanan *SpHr* grubu ile *Hr* grubu MFCC parametreleri ile ayrı ayrı birleştirilerek kullanılmıştır. Bu sistemler MFCC_*SpHr* ile MFCC_*Hr* olarak adlandırılmıştır. MFCC_*SpHr* sistemi toplam 16+11=27 öznitelik, MFCC_*Hr* ise 16+8=24 öznitelik kullanmaktadır. *ASE* grubu, MFCC parametreleri ile benzer oldukları ve daha önceki deney sonuçlarına göre başarımı artıracakları yönünde bir beklenti olmadığı için bu deneylerde kullanılmamıştır.

Öznitelik düzeyinde birleştirme sonuçları da Şekil 6'da ve Şekil 7'de verilmiştir. Görüldüğü gibi *SpHr* ve *Hr* parametrelerinin MFCC parametreleri ile öznitelik düzeyinde birleştirilmeleri başarımı belirgin olarak artırmaktadır. Ortaya çıkan EER değerleri Tablo 6'de verilmiştir. Son sütun MFCC'ye göre sağlanan yüzde iyileşmeyi göstermektedir.

Tablo 6: Öznitelik düzeyinde birleştirilmiş sistemlerin EER değerleri (%)

	KÜME1	**KÜME2**	**Ortalama**	**MFCC Sistemine Göre EERde Azalma**
MFCC	8.58	7.13	7.86	
MPEG	9.00	8.53	8.77	
MFCC_*SpHr*	7.78	6.94	7.36	(7.86→7.36) %5.93
MFCC_*Hr*	7.78	6.35	7.07	(7.86→7.07) %10.07

Tablo 6'ya göre MFCC_*Hr* ile MFCC_*SpHr*'den daha iyi sonuçlar elde edilmiştir. Bu durumda öznitelik düzeyindeki birleştirmede *Sp* parametrelerinin kullanılmasının yarar sağlamadığı ortaya çıkmaktadır.

Eğitilen modelin, modellenmeye çalışılan verinin dağılımına uygun biçime sahip olması ve eğitim verisinin "yeterli" olması durumunda öznitelik artırımının başarımı düşürmemesi, eklenen özniteliklerle, en kötü durumda, önceki başarımın elde edilmesi beklenir; ancak uygulamada bunun doğrulanmadığı durumlarla sık sık karşılaşıldığı da bilinen bir gerçektir [Duda 2001]. Yukarıda verilen sonuçlara göre öznitelik sayısının artırılması başarımın azalmasına neden olmaktadır. Bu durumda modelin (burada GMM) yeni veri dağılımına yeterince uymadığı ve/veya eğitim verisinin yetersiz olduğu düşünülebilir. Eğitim verisinin yeterliliği hakkında MFCC ve MFCC-*Hr* öznitelikleri ile bir kuşku oluşmamıştır. Kullanılan veri miktarı literatürde yer alan benzer çalışmalar ile uyumludur. Bu nedenle modelin veri dağılımına uygun olup olmadığı sorgulanabilir. Ancak *Sp* parametreleri kategorik olarak önemli bir katkı sağlaması beklenen parametreler olmadığı için bu yönde çalışma yapılmamıştır.

3.5. Yalnızca MFCC ve Yalnızca MPEG-7 Niteleyicilerini Kullanan Farklı Konuşmacı Onaylama Sistemlerinin Ürettikleri Sonuçların Birleştirilmesi (Skor Düzeyinde Birleştirme)

Bu çalışmada, her konuşmacı onaylama sistemi karar aşamasında kullanılmak üzere iki değer üretmektedir. Bunlardan birisi iddia edilen konuşmacıya ait GMMden elde edilen olabilirlik değeri, diğeri de UBMden elde edilen olabilirlik değeridir. N farklı onaylama sistemi kullanıldığında $2N$ boyutlu olabilirlik uzayı oluşturulmakta ve tanıma veya reddetme kararı bu uzayda verilmektedir. Bu uzayın oluşturulmasında iki (N=2) ve üç (N=3) farklı doğrulama sistemi kullanılmıştır. Bunlar MFCC, MFCC_SpHr ve MFCC_Hr sistemleridir.

$2N$ boyutlu birleşik olabilirlik uzayında karar üretmek için, bu uzayda çalışan SVM yöntemine dayalı bir ardıl-sınıflandırıcı (post-classifier) geliştirilmiştir.

Konuşmacı onaylamada iki farklı sınıf vardır. Bunlar iddia edilen konuşmacı kimliğinin doğru olduğu (doğrucu) ve doğru olmadığı (yanıltıcı) durumlardır. Toplam M konuşmacının kullanıldığı ve her konuşmacı için bir ses sinyali olduğu durumda, toplam M adet doğrucu konuşmacı testi ve Mx(M-1) adet yanıltıcı konuşmacı testi ("impostor test") mümkündür. Bu durumda, ardıl-sınıflandırıcının eğitilmesinde kullanılabilecek yanıltıcı konuşmacı verisi, doğrucu konuşmacı verisinden çok daha fazladır. Bu durum sınıf dengesizliği ("class-imbalance") olarak bilinmektedir. Sınıf dengesizliği, ardıl-sınıflandırıcının eğitme verisi daha az olan sınıfta çok hata yapmasına neden olabilmektedir. Sınıf dengesizliği sorununa karşı tek sınıflandırıcı yerine bir sınıflandırıcı topluluğu kullanılmıştır. Bu toplulukta yer alan sınıflandırıcıların birbirlerine göre farklılaşması ("diversity") sistemin başarısını

artırıcı bir etkendir. Sınıflandırıcılar arası farklılaşma, sınıflandırıcıların yaptıkları hataların ilintisinin ("correlation") düşük olması anlamına gelir. Buna yönelik olarak çok yönlü sarsım ("multimodal perturbation") [Zhou 2005] kullanılmıştır.

3.4.1 Destek Vektör Sınıflandırıcı (SVM) Yöntemi

SVM yöntemi, temel olarak farklı sınıfların karar yüzeyine en yakın elemanlarının uzaklığınıençoklamaya dayalı bir karar yüzeyi yaratmaktadır. Bu amaçla aşağıdaki karar fonksiyonu kullanılmaktadır:

$$f(\mathbf{o}) = \operatorname{sgn}\left(\sum_{i=1}^{S} \alpha_i y_i K(\mathbf{o}_i, \mathbf{o}) + b\right)$$

Verilen bir gözlem, **o**, için karar,

$f(\mathbf{o}) = +1 \rightarrow$ "doğrucu",

$f(\mathbf{o}) = -1 \rightarrow$ "yanıltıcı"

olarak verilmektedir. Yukardaki ifadede S, toplam destek vektörü sayısını; **o**, gözlem vektörünü; $\mathbf{o}_i$, i'inci destek vektörünü; y_i, i'inci destek vektörünün sınıf etiketini (-1,+1) göstermektedir; $K(\cdot,\cdot)$ simetrik çekirdek fonksiyonu olarak bilinmekte ve iki farklı vektörün benzerliğini ölçmektedir. Bu fonksiyon yüksek boyutlu uzayda iç çarpımı gerçekleştirecek şekilde seçilmektedir. α_i parameterleri, $0 \le \alpha_i \le C$ koşulu altında, eğitim kümesi üzerinde karesel bir ifadenin enazlanması ile hesaplanmaktadır. SVM sınıflandırıcılarının eğitilmesinde, kullanıcının belirlediği en önemli parametreler C değeri ile kernel fonksiyon tipidir. Bu çalışmada doğrusal çekirdek fonksiyonu kullanılmıştır; $K(\mathbf{o}_i, \mathbf{o}) = \mathbf{o}_i^T \mathbf{o}$. C değerinin seçimi bir sonraki bölümde anlatılmaktadır. SVM sınıflandırıcıların hazırlanmasında PRTools yazılımı kullanılmıştır **[PRTools]**.

3.4.2 Çok Yönlü Sarsım

Çoğul sınıflandırıcılı sistemlerin uygun bireysel sınıflandırıcılar kullanıldığında daha iyi kararlar üretebildikleri bilinmektedir. Bu çalışmada, birden fazla SVM tipi sınıflandırıcının birleştirilerek ardıl-sınıflandırıcı olarak kullanılması düşünülmüştür. Bu tür güçlü sınıflandırıcılar etkin bir şekilde bir arada kullanılabilmeleri ancak uygun koşullar sağlandığında mümkün olabilmektedir. Bireysel olarak sınıflandırma başarımı açısından güçlü sınıflandırıcıların birbirleri ile uyumlu bir küme olabilmeleri için farklılığa (diversity) sahip olmalarının yararlı olduğu kabul edilir. Farklılaşma sağlamak için değişik yöntemler ayrı ayrı veya birlikte kullanılabilir [Kuncheva 2003]. Bu çalışmada eldeki probleme uygun bazı yöntemler önerilmektedir. Yukarıda da belirtildiği gibi ardıl-sınıflandırma probleminin en belirgin özelliği sınıflar arası dengesizliktir. Hem dengesizliğin neden olabileceği

olumsuzlukları önleyebilmek hem de farklı sınıflandırıcılar yaratmak için uygulanmış olan işlemler şunlardır:

a) Her sınıflandırıcı için yanıltıcı konuşmacı eğitim verisinin sayısı $p \in \{1,2,3,4\}$ olmak üzere, pM dir. p, her sınıflandırıcı için rastlantısal olarak seçilmektedir. Her sınıflandırıcı için pM adet yanıltıcı konuşmacı eğitim verisi, toplam $M \times (M-1)$ veri içinden rastlantısal olarak seçilmektedir.

b) Her sınıflandırıcı için seçilen $M+pM$ adet toplam verinin %30u ile %70i arasında bir sayıda eleman içeren bir alt küme rastlantısal olarak seçilir. Böylece, her sınıflandırıcı için kullanılan doğrucu ve yanıltıcı konuşmacı kümelerinin eleman sayıları farklı olur. Her sınıflandırıcı için eğitim verisinin bu şekilde seçimi "bagging" türü bir yaklaşımdır [Duda 2001].

c) Her sınıflandırıcı boyutu ve elemanları rastlantısal olarak seçilen bir altuzayda çalışmaktadır. Birleştirilecek öznitelik grubu sayısı ψ olarak tanımlanırsa, toplam 2ψ adet olabilirlik değeri, $\{o_1, o_2, \ldots, o_{2\psi}\}$, bulunmaktadır. Sınıflandırıcının üzerinde çalışacağı öznitelik vektörünün boyutu M, $\lfloor 1.5\psi \rfloor \leq M \leq 2\psi$ aralığından rasgele seçilmektedir ($\lfloor x \rfloor$, x'den küçük en büyük tamsayıdır.). M adet öznitelik de $\{o_1, o_2, \ldots, o_{2\psi}\}$ kümesi içinden rasgele seçilmektedir.

d) Ardıl-sınıflandırıcı parametrelerinden C değeri her sınıflandırıcı için farklı olmak üzere [0.1,15] aralığından rastlantısal olarak seçilmektedir.

Skor düzeyindeki birleştirmede 15 adet SVMden oluşan sınıflandırıcı topluluğu kullanılmıştır. Yukarıda açıklanan sarsım işlemleri bu topluluk üzerinde uygulanmıştır. Test aşamasında 15 SVMden 15 karar elde edilmiş, son karar bunların ortalaması alınarak bulunmuştur.

Sınıflandırıcılar arasındaki farklılaşmayı ölçmek için önerilmiş ölçütler vardır. Farklılaşmanın başarımı artırıcı etkisi olduğu kabul edilmektedir. Ancak farklılaşma ile başarım arasında ilişki kuran ölçütler bulunmamaktadır. Sınıflandırıcı topluluğu tasarımı bu çalışmanın ana temaslarından olmadığı için farklılaşmayı ölçmeye yönelik bir çalışma yapılmamıştır.

3.6. Skor Düzeyinde Birleştirme Sonuçları

Skor düzeyinde birleştirme sistemi, KÜME1 test edilirken SVM ardıl sınıflandırıcısının eğitilmesi için KÜME2'den elde edilen test skorları kullanılmıştır. Benzer şekilde KÜME2 test edilirken SVM için eğitme verisi olarak KÜME2'den elde edilen test skorları kullanılmıştır. Skor düzeyinde SVM sınıflandırıcı topluluğu ile gerçekleştirilen deneylerde farklı öznitelik kümeleri ile eğitilmiş bireysel sistemler kullanılmıştır. Burada üç birleştirmenin sonuçları verilmektedir:

- MFCC - *SpHr*
- MFCC - MPEG

- MFCC - MPEG - *SpHr*

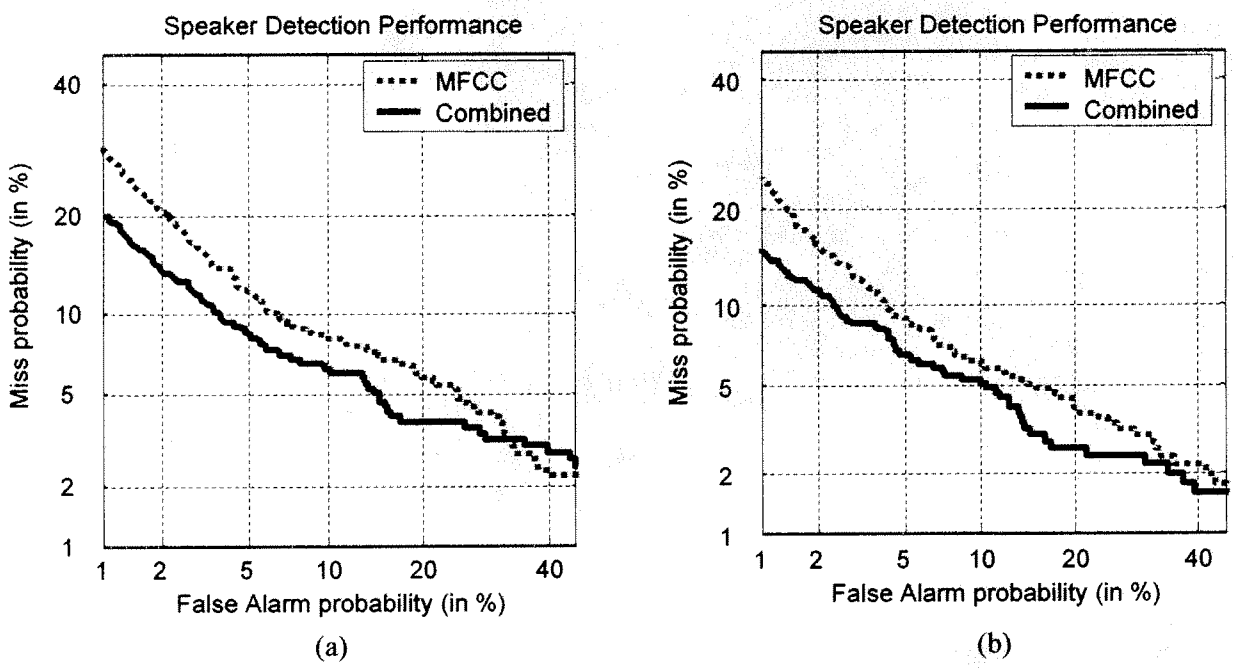
Yukarıda sıralanan sarsım teknikleri 15 kez tekrarlanarak elde edilen 15 sınıflandırıcının verdiği sonuçların ortalaması alınarak son karar verilmiştir. Her bileşen rastlantısal olarak 4 veya 5 boyutlu *altuzayda*, veya 6 boyutlu *özgün uzayda* eğitilmiştir.

Skor düzeyinde birleştirme sonuçları Tablo 7'de yer almaktadır. KÜME1 ve KÜME2 için test sonuçları ve son sütunda MFCC sistemine göre EERdeki iyileşmeler verilmiştir.

Tablo 7: Skor düzeyinde birleştirilmiş sistemlerin Eşit Hata Oranları (%)

Bireysel Sistemler	KÜME1	KÜME2	Ortalama İyileşme
MFCC *SpHr*	7.19	6.78	10.49
MFCC MPEG	6.99	6.35	14.67
MFCC MPEG *SpHr*	6.99	5.95	17.48

SpHr parametreleri tek başına kullanıldığında, yanlış alarm hatasının yüksek olduğu bölgede MPEG sisteminden daha iyi sonuçlar sağlandığı için bu sistem de birleştirmede kulanıllarak üçlü birleştirme yapılmıştır. Bunun sonucunda EERde de azalma görülmüştür. MFCC, MPEG ve SpHr sistemlerinin üçünün birden skor düzeyinde birleştirilmesinden elde edilen eğriler Şekil 8'de verilmiştir.



Şekil 8: KÜME1(a) ve KÜME2 (b) için MFCC, MPEG ve SpHr sistemlerinin skor düzeyinde birleştirilmesi

4. Son Altı Aylık Dönemde Yapılan Çalışmalar

Proje çalışmaları, proje önerisinde belirtilen plana göre daha hızlı ilerlemiştir. Projenin ilk iki altı aylık döneminde, MPEG-7 harmonik ses niteleyicilerinin MFCCler ile sağlanan konuşmacı onaylama başarımını artırıp artırmayacağına yönelik deney sistemleri tamamlanmış ve deneyler önemli ölçüde yapılmıştır. Elde edilen sonuçlar, MPEG-7 harmonik niteleyicilerinin anlamlı ölçüde katkı sağladığını göstermiştir. Başarım artışı hem öznitelik düzeyinde birleştirme ile hem de farklı sınıflandırıcılarla skor düzeyinde birleştirmede söz konusu olmaktadır. Proje konusunun ortaya çıkışındaki beklentilerin doğrulandığı görülmüştür. Öznitelik düzeyinde birleştirme ile eşit hata oranında, yalnızca MFCC niteleyicileri ile elde edilen değere göre yaklaşık %10 azalma sağlanmıştır. Farklı sınıflandırıcıların birleştirilmesi ile bu iyileşme %17'ye çıkmaktadır. İlk iki altı aylık dönemde yapılan deneylerde bu *Hr* niteleyicileri topluca kullanılmışlardır. Bunlar arasında en iyi altküme seçimi yapılmamıştır. İkinci altı aylık dönemin sonunda hazırlanan raporda, son dönemde yapılması planlanan çalışmalar şöyle belirtilmiştir: "*Çalışmanın son altı aylık döneminde Hr kümesi içinde yer alan parametrelerin konuşmacı onaylamaya katkılarının daha ayrıntılı incelenmesi düşünülmektedir. Hr kümesi içinde yer alan parametrelerin sağladıkları katkı bakımından sıralanması ele alınacaktır. Bu kapsamda işlem yükü bakımından uygun olduğu ölçüde tanıma oranlarından yararlanılacaktır.*" Bu plana uygun olarak, tanıma oranlarına dayalı öznitelik seçimi çalışması yapılmıştır.

4.1. MPEG-7 Harmonik Niteleyicileri Arasından En İyi Altküme Seçimi

Yapılan çalışmalar sonucunda MPEG-7 niteleyicileri arasından *Hr* kümesi olarak adlandırılan aşağıdaki sekiz harmonik niteleyicinin yararlı olacağı görülmüştür.

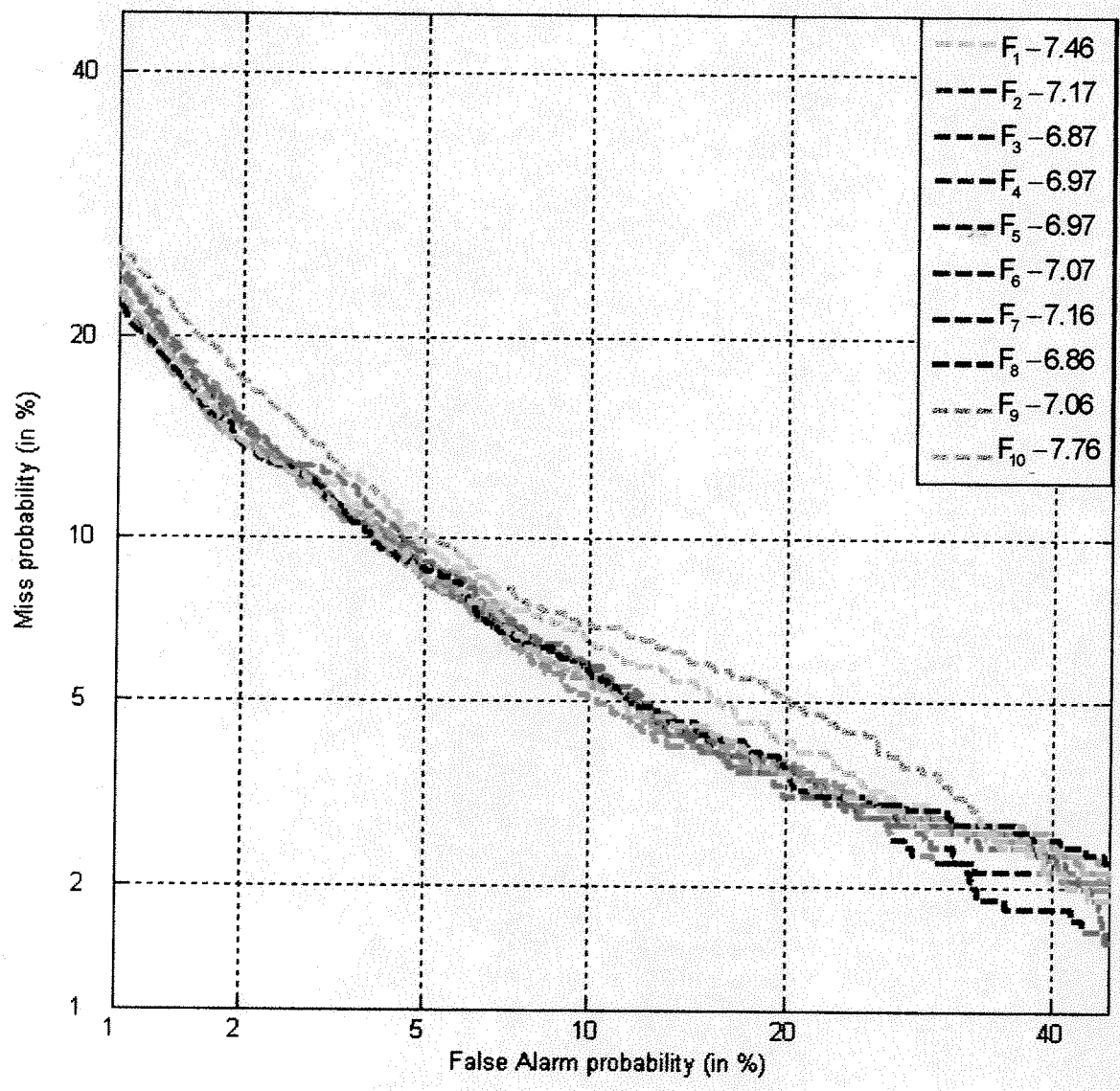
"audio fundamental frequency" ve "related confidence measure"
"audio harmonicity" ve "upper limit of harmonicity"
"harmonic spectral centroid"
"harmonic spectral deviation"
"harmonic spectral spread"
"harmonic spectral variation"

En iyi altküme seçimi tam olarak yapılırsa, toplam 8 farklı niteleyici ile, 255 deney yapılmasını gerektirmektedir. Her deneyde eğitimin ve testlerin yeniden yapılması gerekmesi ve bu sürelerin günler düzeyinde olması nedeniyle, tüm altkümelerin denenmesi gerçekçi değildir. Bir çok sınıflandırma probleminde benzer sorunlarla karşılaşıldığı için, bu durumda izlenebilecek yöntemler de ortaya çıkmıştır. Bu yöntemler arasında en uygun olanları, artırarak seçme ("forward selection") ve azaltarak seçme ("backward selection") yöntemleridir. Bu çalışmada azaltarak seçme yöntemi uygulanmıştır. Buna göre, en geniş küme ile başlanıp, her aşamada bir niteleyici kümeden çıkarılarak deneyler tekrar edilmiştir. Ele alınan durumda, önce sekiz niteleyici ile deney yapılmıştır. Daha sonra tüm yedi niteleyici içeren altkümelerle deneyler yapılmıştır. Bu sekiz deneyde, en iyi sonucun elde edildiği altkümede olmayan niteleyici, daha sonraki aşamalarda kullanılmamak üzere terkedilmiştir. Birinci aşamada elde edilen "en iyi" yedi niteleyici ikinci aşamada başlangıç kümesi olmuştur. Yine bir niteleyici eksilterek elde edilen, yedi tane altı niteleyicili altkümeler ile deneyler yapılmıştır. Elde edilen sonuçlar incelendiğinde, kullanılan yöntemle daha fazla eksiltme yaparak başarım artışı olmayacağı sonucuna varıldığından deneyler sonlandırılmıştır.

Deneylerin ilk aşamasında yedi niteleyiciden oluşan altkümeler F_1, F_2, ... F_8 olarak adlandırılmıştır. Bu kümelerin her birinde sekiz *Hr* niteleyicisinden birisi eksiktir. Kümelerin tanımı Tablo 8'de verilmiştir. Yedi niteleyiciden oluşan altkümelerle elde edilen DET eğrileri Şekil 9'da görülmektedir. Şekil 9'da *Hr* kümesinin tamamı (F_9) ve yalnızca MFCClerle elde edilen sonuçlar da gösterilmiştir.

Tablo 8: Niteleyici eksiltme deneyinde altküme tanımları.

Altküme	**Eksik niteleyici**
F1	F0
F2	Periodicity
F3	Audio harmonicity
F4	Up_Lmt_Harmo
F5	Harmonic Spectral Centroid
F6	Harmonic Spectral Spread
F7	Harmonic Spectral Deviation
F8	Harmonic Spectral Variation



Şekil 9: Sekiz niteleyiciden oluşan Hr kümesinden her seferinde bir niteleyici çıkararak elde edilen yedi niteleyicili kümelerle (F_1, F_2, F_3, F_4, F_5, F_6, F_7, F_8) ve *Hr* (F_9) ve yalnızca MFCClerle (F_{10}) ile elde edilen sonuçlar.

EER değerleri en küçükten en büyüğe doğru sıralanmış olarak Tablo 9'da verilmiştir. Sekiz Hr niteleyicisi birlikte kullanıldığında elde edilen sonuç, F_9, koyu zeminde görülmektedir.

Tablo 9: F_1, F_2, F_3, F_4, F_5, F_6, F_7, F_8 ve F_9 kümeleri ile elde edilen sonuçların en küçükten en büyüğe doğru sıralamaları.

Deney	F_8	F_3	F_4	F_5	F_9	F_6	F_7	F_2	F_1
Çıkarılan Niteleyici	Harmonic Spectral Variation	Audio harmonicity	Upper Limit Harmo	Harmonic Spectral Centroid		Harmonic Spectral Spread	Harmonic Spectral Deviation	Periodicity	F0
EER (%)	6.86	6.87	6.97	6.97	7.06	7.07	7.16	7.17	7.46
F_9 sonucuna göre (%) iyileşme	2.8	2.7	1.3	1.3	----	- 0.1	- 1.4	- 1.6	- 5.7

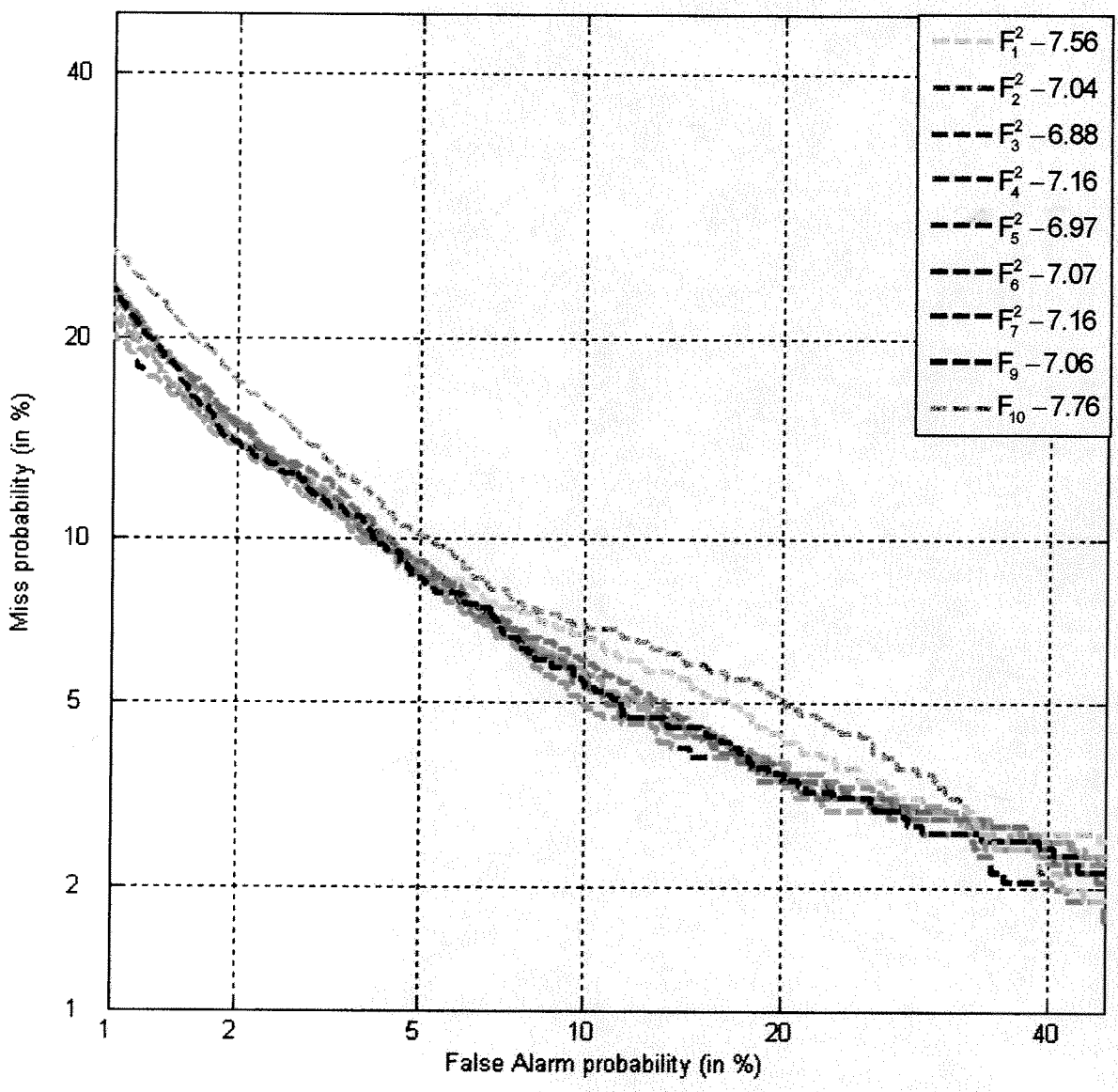
Tablo 9 incelendiğinde, F_8, F_3, F_4 ve F_5 deneylerinde, F_9 deneyinden daha iyi sonuçlar elde edildiği görülmektedir. En fazla iyileşme F_8 deneyinde % 2.8 olmuştur. F_3 deneyinde de buna yakın bir iyileşme, % 2.7, sağlanmıştır. Bu sonuçlara göre MFCCleri en iyi destekleyen yedi *Hr* niteleyicisi, "Harmonic Spectral Variation" ya da "Audio Harmonicity" niteleyicisinin çıkarılması ile elde edilmektedir. "Upper Limit of Harmonicity"in ve "Harmonic Spectral Centroid"in çıkarılması ile de (daha az) iyileşme sağlanmaktadır. Buna karşılık "F0"nun çıkarılması önemli başarım kaybı yaratmaktadır. Temel frekansın konuşmacı tanımada önemli rol oynadığı bilindiği için bu şaşırtıcı bir sonuç değildir. Daha az olmakla birlikte "Periodicity" ve "Harmonic Spectral Deviation" niteleyicilerinin çıkarılması da başarımda azalmaya neden olmaktadır. Niteleyicilerin, sağladıkları katkıya göre "iyi"den "kötü"ye doğru sıralanmışı aşağıdaki gibidir.

1. F0
2. Periodicity
3. Harmonic Spectral Deviation
4. Harmonic Spectral Spread
5. Harmonic Spectral Centroid
6. Upper Limit Harmo
7. Audio harmonicity
8. Harmonic Spectral Variation

İlk eksiltme deneyinde en iyi sonuç "Harmonic Spectral Variation" niteleyicisi çıkarıldığında elde edildiği için ikinci aşama eksiltme deneylerine bu niteleyici çıkarılarak devam edilmiştir. İkinci aşamada kullanılan kümelerin tanımları ve eksiltilen niteleyiciler Tablo 10'da görülmektedir.

Tablo 10: Niteleyici eksiltme deneyinin ikinci aşamasında altküme tanımları.

Altküme	**Eksik niteleyici**
F_1^2	F0
F_2^2	Periodicity
F_3^2	Audio harmonicity
F_4^2	Up_Lmt_Harmo
F_5^2	Harmonic Spectral Centroid
F_6^2	Harmonic Spectral Spread
F_7^2	Harmonic Spectral Deviation



Şekil 10: Yedi elemanlı F_8 kümesinden her seferinde bir niteleyici çıkararak elde edilen yedi niteleyicili kümeler, ($F_1^2, F_2^2, F_3^2, F_4^2, F_5^2, F_6^2, F_7^2$), F_9 ve F_{10} kümeleri ile elde edilen sonuçlar.

İkinci aşama deneylerde elde edilen DET eğrileri Şekil 10'da görülmektedir. EER değerleri küçükten büyüğe doğru sıralanmış olarak Tablo 11'de verilmiştir. F_8 kümesi ile elde edilen sonuç koyu zeminde görülmektedir. İkinci aşama deneylerinde var olandan daha iyi bir sonuç elde edilmemiştir.

Tablo 11: $F_8^1, F_8^2, F_8^3, F_8^4, F_8^5, F_8^6, F_8^7$ ve F_8 kümeleri ile elde edilen sonuçların küçükten büyüğe doğru sıralamaları.

Deney	F_8	F_8^3	F_8^5	F_8^2	F_8^6	F_8^4	F_8^7	F_8^1
Çıkarılan Niteleyici		Audio harmonicity	Harmo. Spectral Centroid	Periodicity	Harmo. Spectral Spread	Up_Lmt_Harmo	Harmonic Spectral Deviation	F0
EER (%)	6.86	6.86	6.97	7.04	7.07	7.16	7.16	7.56

Tablo 11'deki sayısal değerlere göre ikinci aşamada iyileşme sağlanamadığı anlaşılmaktadır. En iyi sonuç olarak, F_8^3 kümesi ile F_8 kümesi ile elde edilen sonucun aynısı elde edilmiştir; diğerleri ile daha büyük EER değerleri ortaya çıkmaktadır. İkinci aşamada başarım artışı sağlanamadığı için daha fazla eksiltme ile deney yapılmamıştır. DET eğrilerinin görünümü de, EER değerlerine göre yapılan değerlendirmelerle uyumludur. Bu sonuçlar *Hr* kümesinin sekiz ya da yedi elemanı ile birlikte MFCC parametrelerini destekleyici olarak kullanılabileceğini göstermektedir.

5. Sonuç

Bu araştırma projesinde konuşmacı onaylamada MPEG-7 ses niteleyicilerinin, MFCC niteleyicileri ile birlikte kullanılması ile ne ölçüde katkı sağlanabileceği incelenmiştir. Bu amaca yönelik olarak konuşmacı onaylama sistemleri hazırlanmış, NIST-99 veritabanı kullanılarak hazırlanan verilerle eğitim ve test çalışmaları yapılmıştır. Konuşmacı onaylama için hem özniteliklerin hem de bir kaç sınıflandırıcıdan elde edilen skorların birleştirilmesi üzerinde çalışılmıştır. Sınıflandırıcılar Gauss karışım modellerinden oluşturulmuştur. Skor düzeyinde birleştirmede kullanılan ardıl sınıflandırıcılar için değişik yöntemler denenmiştir.

Yalnızca MFCC parametreleri ile yapılan deneylerde EER değeri % 7.76 olarak elde edilmiştir. Bu değer *F0*, *Periodicity*, *Audio harmonicity*, *Up_Lmt_Harmo*, *Harmonic Spectral Centroid*, *Harmonic Spectral Spread*, *Harmonic Spectral Deviation* niteleyicilerinden oluşan yedi elemanlı kümenin kullanılması ile % 6.86 değerine inmiştir. Bu % 11.6 oranında bir azalmaya karşılık gelmektedir.

Daha önce yapılan skor düzeyinde birleştirme deneylerinde EER değerinde % 17.5 azalma sağlanmıştır. Bu azalmaya birden fazla sınıflandırıcı kullanılarak ulaşılmıştır.

Öznitelik düzeyinde ya da skor düzeyinde birleştirme ile sağlanan başarım artışları çok şaşırtıcı olmamakla birlikte anlamlı ve göz ardı edilemeyecek düzeylerdedir. Bu nedenle konuşmacı onaylamada MPEG-7 harmonik niteleyicilerinin kullanılması bir seçenek olarak düşünülmelidir. Sonuçlar harmonik içeriğe ait bilginin kullanımının yaralı olduğu göstermektedir. Bu tür bilgiyi elde etmenin MPEG-7 niteleyicilerinden farklı yolları da vardır. Daha önce de bu yönde çalışmalar yapılmıştır. Dolayısıyla harmonik içerik bilgisinin

kullanımından en fazla yararın nasıl sağlanılabileceği gelecekte ele alınabilecek bir araştırma konusudur.

Teşekkür

MPEG-7 niteleyicilerini hesaplama yazılımı veren TÜBİTAK-BİLTEN'e ve yazılımı üreten Banu Oskay'a ve Hacer Yalım'a teşekkür ederiz. Söz konusu yazılımın bazı eksikleri ve hataları bu çalışma sırasında giderilmiştir.

Kaynaklar

[MPEG7 2001] ISO Int. Standart, ISO/IEC CD 15938-4 (Multimedia Content Description Interface, Part 4: Audio), 2001.

[Campbell 1997] J. P. Campbell, "Speaker Recognition: A Tutorial", Proceedings of the IEEE, vol. 85, NO. 9, pp.1437-1462, Sept. 1997.

[Duda 2001] R. O. Duda, P. E. Hart, D. G. Stork, *Pattern Classification*, 2nd edition, Wiley, 2001.

[Davis 1980] S. B. Davis and P. Mermelstein, "Comparison of Parametric Representations for Monosyllabic Word Recognition in Continuously Spoken Sentences," *IEEE Trans. Acoustics, Speech and Signal Processing*, Vol. ASSP-28, No.4, pp.357-376, Aug 1980.

[Kim 2003] H. G. Kim, E. Berdahl, N. Moreau, and T. Sikora, "Speaker recognition using MPEG-7 descriptors", Proc. of EUROSPEECH (8th European Conf. on Speech Communication and Tech.), pp. 489-492, Geneva, Sept. 2003.

[Kim 2004] H. G. Kim and T. Sikora, "Comparison of MPEG-7 audio spectrum projection features and MFCC applied to speaker recognition, sound classification and audio segmentation", Proc. of IEEE Int. Conf. on Acoustics, Speech, and Signal Proc., vol. 5, pp. 925-928, May 2004.

[Kuncheva 2004] Ludmilla I. Kuncheva, "*Combining Pattern Classifiers, Methods and Algorithms*", John Wiley and Sons, 2004.

[PRTools] "PRTools - A Matlab toolbox for pattern recognition", Pattern Recognition Group, Delft University, Netherlands, http://www.prtools.org/.

[Reynolds 1995] D.A. Reynolds and R.C. Rose, "Robust Text-Independent Speaker Recognition using Gaussian Mixture Speaker Models," *IEEE Transactions on Speech and Audio Processing*, Vol. 3, pp. 72–83, January 1995.

[Reynolds 1997] D. A. Reynolds, "Comparison of background normalization methods for text-independent speaker verification," in *Proc. 5th European Conference on Speech Communication and Technology (Eurospeech '97)*, vol. 2, pp. 963–966, Rhodes, Greece, September 1997.

[Reynolds 2000] D. A. Reynolds, T. F. Quatieri and R. B. Dunn, "Speaker Verification using Adapted Gaussian Mixture Models," *Digital Signal Processing Review Journal*, Vol. 10, pp. 19-41, Jan. 2000.

[Yang 2005] Yang. P, Yang, Y. and Wu, Z., "Exploiting glottal information in speaker information in speaker recognition using parallel GMMs," Lecture Notes in Computer Science, Springer Verlag vol. 3546, p. 804, 2005.

[Zhou 2005] Zhou, Z.-H. and Yu. Y., "Ensembling local learners through multimodal perturbation," IEEE Transactions on Systems, Man, and Cybernetics - Part B: Cybernetics, 2005, 35(4): 725-735.

PROJE ÖZET BİLGİ FORMU

Proje Kodu: 104E142

Proje Başlığı: Konuşmacı Tanımada MPEG-7 Ses Özniteliklerinin Kullanılabilirliği

Proje Yürütücüsü ve Yardımcı Araştırmacılar:
Doç.Dr. Tolga Çiloğlu, Doç.Dr. Yunus Hakan Altınçay, Y. Doç. Dr. Cem Ergun

Projenin Yürütüldüğü Kuruluş ve Adresi:
Elektrik ve Elektronik Müh. Böl., ODTÜ, Ankara
Bilgisayar Müh. Böl., Doğu Akdeniz Üniv. G. Mağusa, KKTC.

Destekleyen Kuruluş(ların) Adı ve Adresi:

Projenin Başlangıç ve Bitiş Tarihleri: 01.07.2005 – 31.01.2007

Öz (en çok 70 kelime)

Konuşmacı onaylama probleminde konuşma sinyalinin harmonik içeriğine ait bilginin kullanımı ele alınmıştır. Harmonik içerik, seçilen MPEG-7 niteleyicileri ile temsil edilmiştir. Bu bilginin MFCC gibi bilinen öznitelikleri ne ölçüde destekleyeceği incelenmiştir. Temel sınıflandırıcı olarak Gauss karışım modelleri kullanılmıştır. Deneyler NIST 99 veritabanı ile gerçekleştirilmiştir. Öznitelik ve sınıflandırıcı birleştirme çalışmaları yapılmıştır. Sınıflandırıcı birleştirmede farklı yöntemler denenmiş, birleştirme sisteminin eğitiminde "AdaBoost" ve çok yönlü sarsım gibi teknikler kullanılmıştır. Öznitelik düzeyinde birleştirmede öznitelik seçimi üzerinde durulmuştur. Yapılan deneyler, MFCC ve harmonik niteleyicilerin öznitelik düzeyinde birleştirilmesi ile yalnızca MFCC'lerin kullanımı ile elde edilen EER değerlerine göre % 11 azalma sağlandığını göstermiştir. Sınıflandırıcı birleştirme ile bu azalma %17'ye ulaşmıştır. Azaltarak seçme yöntemi kullanılarak MPEG-7 niteleyicileri arasından seçilen harmonik niteleyicilerin hemen hemen tamamının başarıma katkısını olduğu belirlenmiştir.

Anahtar Kelimeler: Konuşmacı onaylama, konuşmacı tanıma, MFCC, MPEG-7, harmonik öznitelikler, öznitelik birleştirme, sınıflandırıcı birleştirme.

Projeden Kaynaklanan Yayınlar:

1) Hakan Altınçay, Cem Ergün ve Tolga Çiloğlu, "Konuşmacı Doğrulamada MFCC ve MPEG-7 Özniteliklerinin Birleştirilmesi", Sinyal İşleme ve Uygulamaları Kurultayı, Nisan 2006.

2) Hakan Altınçay, Cem Ergün and Tolga Çiloğlu, "Using MPEG-7 Audio Descriptors for Speaker Verification", 10th International Conference on Speech and Computer, Patras, Greece, October 2005.

3) Hakan Altınçay, Cem Ergün and Tolga Çiloğlu, "Using MPEG-7 Audio Descriptors for Speaker Verification", 13th International Workshop on Advances in Speech Technology, Maribor, Slovenia, July 2006.

Bilim Dalı:

Doçentlik B. Dalı Kodu:

USING MPEG-7 AUDIO DESCRIPTORS FOR SPEAKER VERIFICATION

Hakan Altınçay, Cem Ergün* and Tolga Çiloğlu***

*Computer Engineering Department
Eastern Mediterranean University

**Electrical and Electronics Engineering Department
Middle East Technical University

Problem Definition

- In this study, the use of MPEG-7 audio descriptors in speaker verification is addressed
- MPEG-7 audio descriptors contain harmonic-related features which are considered to carry glottal information.
- The effectiveness of feature level combination of MFCC parameters and a subset of MPEG-7 features is studied
- A hybrid system is proposed for score level combination of the verification systems that are based on either MFCC or MPEG-7 attributes.

Attributes in MPEG-7 standard

ASE	SpHr	Hr
13 Audio Spectrum Envelope parameters	• Audio spectrum centroid • Audio spectrum spread • Audio fundamental freq. and related confidence measure • Audio harmonicity and its upper limit • Harmonic spectral centroid • Harmonic spectral spread • Harmonic spectral deviation • Harmonic spectral variation • Spectral centroid	• Audio fundamental freq. and related confidence measure • Audio harmonicity and its upper limit • Harmonic spectral centroid • Harmonic spectral spread • Harmonic spectral deviation • Harmonic spectral variation

Feature Level Combination

- Two speaker verification systems are implemented where *SpHr* and *Hr* features are combined with MFCC separately.

- These systems are named as MFCC_SpHr and MFCC_Hr respectively. In MFCC_SpHr, there are totally 16+11=27 features and, in MFCC_Hr there are 16+8=24 features.

- The ASE feaures are not used for feature level combination due to to their similarity to MFCC parameters.

Score level combination of MFCC and MPEG-7 Systems

- The likelihood scores produced by MFCC, SpHr and ASE_SpHr based verification systems (2 dimensional) are combined to form feature vectors of the post-classification operation.
- Class imbalance where the target scores data is much less than those of the impostors occurs in this problem. The ratio of impostor attacks to target tests is one less than the total number of speakers involved.
- In order to avoid this problem, multi-modal perturbation described below is used.
- In this study, SVM classifier is used for post-classification.

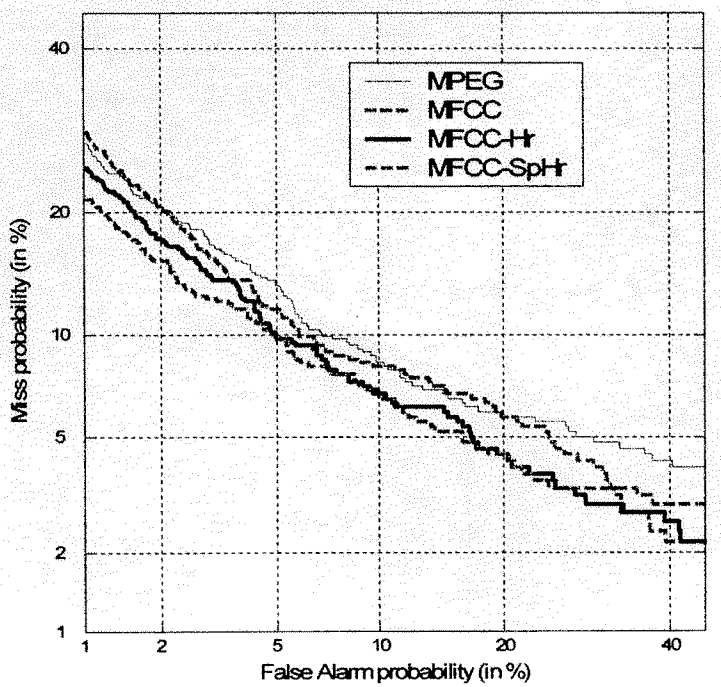
Multi-modal perturbation based post-classification

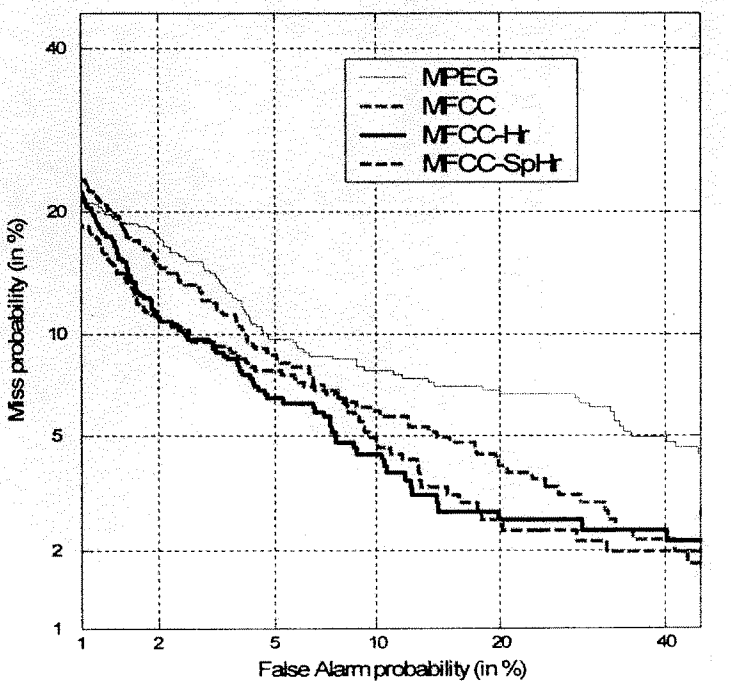
- To alleviate class-imbalance, an ensemble of 15 SVMs is trained.
- Each member classifier is trained on a randomly selected subset of the genuine and impostor scores. The training set size for impostors is p times that of genuine scores where $p \in \{1,2,3,4\}$ is randomly selected for each member SVM. The training set sizes for genuine scores are 501 and 504 for SET1 and SET2, respectively
- Score-based feature vectors of each member contain M log-likelihoods where $\lfloor 1.5\psi \rfloor \leq M \leq 2\psi$ and $\lfloor x \rfloor$ is the largest integer smaller than x. M and individual elements of a member are determined randomly. Hence, members operate in different subspaces.
- As a further reinforcement of diversity, the learning parameter C (penalty factor) of each SVM is also randomly selected in the interval [0.1, 15].

Baseline Speaker Verification Systems

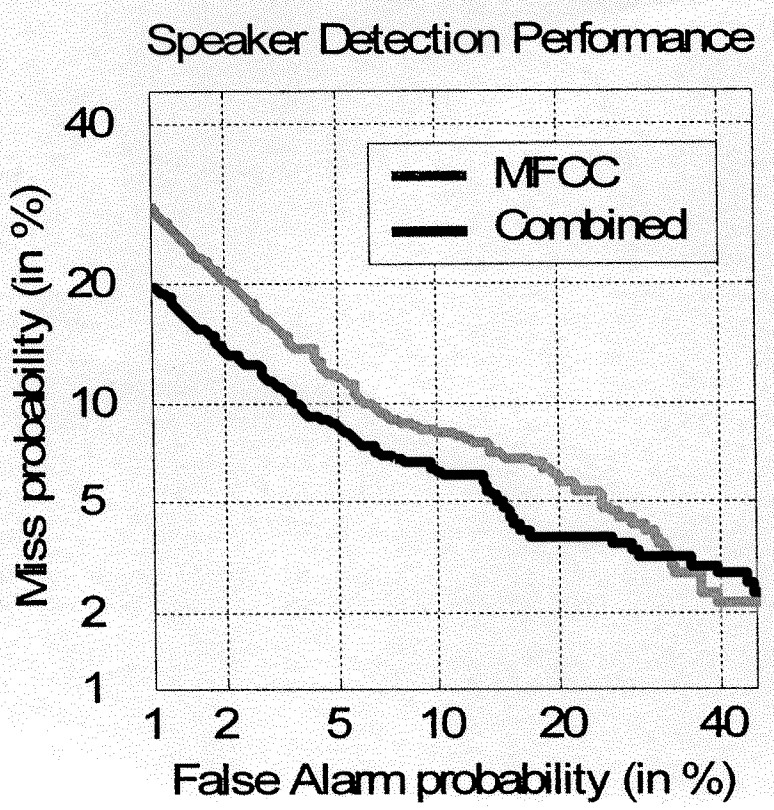
- The training of speaker verification system is based on estimating GMMs for each speaker obtained from a 2048-mixtures UBM using Bayesian adaptation.
- Each adapted GMM is trained for its mixture weights, mean and diagonal covariance matrices using the expectation-maximization algorithm.
- 16 MFCC are used as feature vectors using a frame length of 30ms with 50% overlap with the neighboring frames
- The experiments are conducted on the NIST99 database
- Experiments are performed on two randomly selected sets of 150 speakers, SET1 and SET2

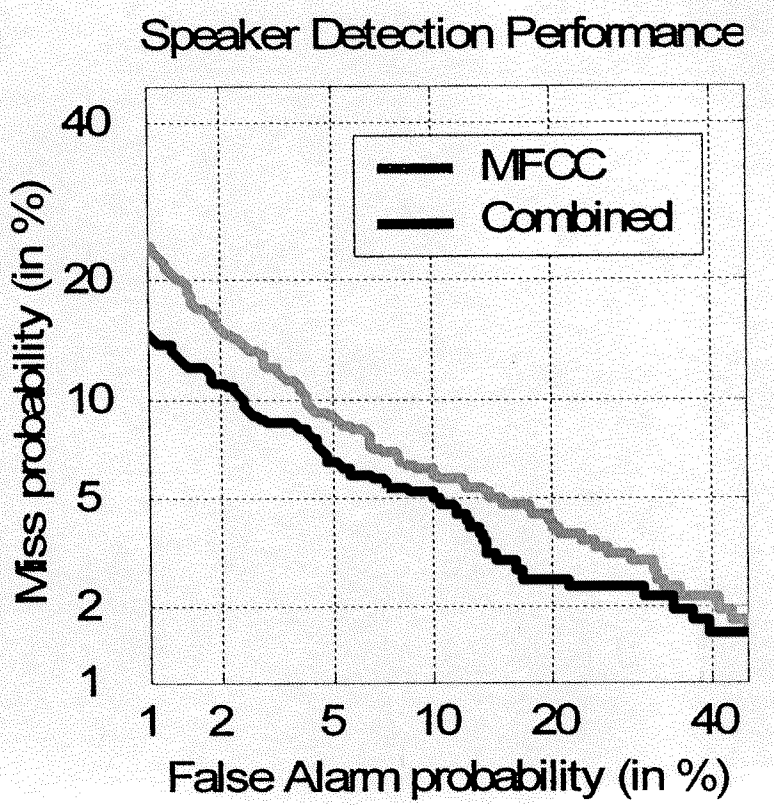
Feature level combination Results





Score level combination Results





Equal Error Rates

Individual systems	SET1	SET2
MFCC	8.58	7.13
MPEG	9.00	8.53

Individual systems	SET1	SET2	Average Improvement
MFCC_SpHr	7.78	6.94	5.93
MFCC_Hr	7.78	6.35	10.07

Individual systems r	SET1	SET2	Average Improvement
MFCC/MPEG/*SpHr*	6.99	5.95	17.48
MFCC/MPEG	6.99	6.35	14.67
MFCC/*SpHr*	7.19	6.78	10.49

Conclusions

- A subset of the audio descriptors defined in the MPEG-7 standard are investigated for speaker verification using telephone speech data sampled at 8 kHz
- Feature level combination of MFCC and MPEG-7 features are studied
- A decision combination system is proposed so as to combine the score level information provided by MFCC and MPEG-7 based systems
- The experiments performed on two different sets of speakers have shown that the combined system developed provides 18% reduction in the EER

Konuşmacı Doğrulamada MFCC ve MPEG-7 Özniteliklerinin Birleştirilmesi

Fusion of MFCC and MPEG-7 Attributes for Speaker Verification

Hakan Altınçay, Cem Ergün* ve Tolga Çiloğlu***

*Bilgisayar Müh. Böl., Doğu Akdeniz Üniv., KKTC
**Elektrik ve Elektronik Müh. Böl., ODTÜ, Türkiye
{hakan.altincay,cem.ergun}@emu.edu.tr,ciltolga@metu.edu.tr

Özetçe

Bazı MPEG-7 ses niteleyicilerininin MFCClerden daha farklı kaynak bilgisi ("glottal information") içerdiği öngörüsü doğrultusunda MPEG-7 niteleyicilerinin konuşmacı doğrulama probleminde kullanılması araştırılmıştır. MFCClerle hem öznitlelik hem de skor düzeyinde birleştirmeye dayalı deneyler yapılmıştır. Sonuçlar, birleştirme ile yalnızca MFCC parametreleri kullanan sisteme göre %18'e varan başarım artışı sağlandığını göstermiştir. Öngörüyle uyumlu bu sonuçlar ilgili alandaki çalışmalar için yeni ipuçları vermektedir.

Abstract

Following the anticipation that some of the MPEG-7 audio descriptors hold glottal information differing than those MFCCs hold, possible contribution of MPEG-7 descriptors to speaker verification has been investigated. Both feature level and score level fusion of MFCCs and MPEG-7 descriptors have been studied. Results indicate improvements up to 18 % compared to those obtained by using MFCCs alone; a justification of the anticipation and a novel indication to the community.

1. Giriş

Bu çalışmanın amacı MPEG-7 standardı çerçevesinde tanımlanmış olan ses niteleyicilerin ("audio descriptors") konuşmacı doğrulamada ("speaker verification") ne kadar yararlı olabileceklerini; söz konusu niteleyicilerden hangilerinin daha yararlı olabileceğini incelemektir. MPEG-7 niteleyicilerinin belirli alt kümelerinin tek başlarına niteleyici ya da öznitelik grubu olarak kullanılması öngörülmektedir. Çünkü, MFCC ("Mel Frequency Cepstral Coefficients") niteleyicileri ile konuşmacı tanımada/ doğrulamada ulaşılan başarım değerleri, dünyada MFCClerin akustik öznitelik olarak güvenilirliğini kanıtlamıştır [1]. Kavramsal karşılaştırmaya göre MPEG-7 niteleyicilerinin, MFCClere göre kayda değer düzeyde daha iyi sonuçlar verebileceği yönünde beklentimiz bulunmamaktadır; tutarlı olarak elde edilecek böyle bir sonuç ilginç, hatta şaşırtıcı olur. Beklentimiz MPEG-7 niteleyicilerini, MFCClerle birlikte kullanarak daha iyi sonuçlar elde edebilmektir. Bunun temel nedeni MPEG-7 niteleyicilerinden bazılarının MFCClerden farklı bilgi taşıdığı öngörümüzdür. MFCClerin esas olarak sinyal izge genliğinin tepe hattına ait bilgi verdiği bilinir. Bu parametreler ağırlıklı olarak insan ses yolunu modellemektedir. Öte yandan, konuşma sinyalinin oluşumunda kaynak olan "glottal" bilgiler de konuşmacıya has özellikler taşır [2]. MPEG-7 standardında yer alan bazı ses niteleyicilerin bu tür bilgileri içerdiği düşünülmektedir. Örneğin harmonik bileşenlerin ses sinyalindeki gürültü içeriğine baskınlığı, harmonik bileşenlerin yayılması ile harmonik izgenin dağılımı gibi niteleyiciler MPEG-7 standardında tanımlanmıştır.

Literatürde, MPEG-7 standardında yer alan Ses İzgesi Zarfı (audio spectrum envelope, ASE) ile konuşmacı tanıma deneyleri yapılmıştır [3,4]. Bu parametrelerin MFCC'lerden esas farkı kullanılan frekans ölçeğinin doğrusal olmasıdır. ASE parametrelerinin başarımı 22.05kHz örnekleme sıklığında MFCC parametrelerine göre biraz daha kötüdür.

Bu çalışmada MPEG-7 akustik niteleyicilerinin tümü incelenmektedir. İncelemenin bütünlüğü bakımından öncelikle MPEG-7 niteleyicilerinin bireysel başarımları elde edilmiştir. Daha sonra MFCC parametreleri ile hem öznitelik düzeyinde hem de farklı MPEG-7 altkümeleri kullanan konuşmacı doğrulama sistemleri ile skor düzeyinde birleştirilmeleri üzerinde deneyler yapılmıştır. Skor düzeyinde sınıflandırıcı birleştirme uygulamalarında çok-yönlü sarsım'ın (multi-modal perturbation) önemi son dönemlerde literatürde vurgulanmaktadır [5]. Bu çalışmada, ortaya çıkan birleştirme problemine uygun olan bir *çok-yönlü sarsım* yaklaşımı önerilmiş ve bu yaklaşım birden fazla destek vektör sınıflandırıcısı'nın ("support vector machines", SVM) skor düzeyinde birleştirme yapmaları amacıyla eğitilmelerinde kullanılmıştır. Yapılan deneyler, MPEG-7 standardında tanımlanmış olan parametrelerin MFCC'lere tamamlayıcı olduklarını ve %18'e yakın başarım artışı sağladıklarını göstermiştir. Bu sonuçlar literatürde yenidir.

2. Kullanılan MPEG-7 Öznitelikleri

MPEG-7 standardında tanımlanmış ses niteleyicileri bu çalışmada üç grup halinde ele alınmaktadır:

ASE:(toplam 13 öznitelik) Audio Spectrum Envelope parameters.
SpHr: (toplam 11 öznitelik) Audio spectrum centroid, Audio spectrum spread, Audio fund. freq. and related confidence measure, Audio harmonicity and its upper limit, Harmonic spectral centroid, Harmonic spectral spread, Harmonic spectral deviation, Harmonic spectral variation, Spectral centroid.
Hr: (toplam 8 öznitelik) Audio fund. freq. and related confidence measure, Audio harmonicity and its upper limit, Harmonic spectral centroid, Harmonic spectral spread, Harmonic spectral deviation, Harmonic spectral variation.

SpHr grubunun altkümesi olan *Hr* gurubu parametreleri ağırlıklı olarak harmoniklerle ilgili ölçümleri içermektedir.

3. MFCC ve MPEG-7 Öznitelikleri̇nin birlikte kullanılması

Birden fazla öznitelik vektörünün birlikte kullanılması için iki temel yaklaşım bulunmaktadır. Bunlardan birincisi öznitelik düzeyinde birleştirmedir. Bu yöntemde, farklı öznitelik vektörleri eklenerek birleşik vektör uzayı oluşturulmakta ve konuşmacı doğrulama sistemi bu uzayda geliştirilmektedir. İkinci yöntem ise her öznitelik vektörü ile farklı bir doğrulama sistemi geliştirmeyi öngörmektedir. Farklı sistemlerin sağladığı skorlar kullanılarak kararlar verilmektedir.

Öncelikle, birleştirilmiş sistemlerin başarımlarının değerlendirilebilmesinde referans olarak kullanılmak üzere iki sistem geliştirilmiştir. Bunlardan birincisinde 16 MFCC parametresi, ikincisinde ise *ASE* ve *SpHr* gruplarındaki toplam 24 öznitelik kullanılmıştır. İkinci sistem ASE_SpHr olarak anılacaktır.

3.1. Öznitelik Düzeyinde Birleştirme

Öznitelik düzeyinde birleştirmede yukarıda tanımlanan *SpHr* grubu ile *Hr* grubu MFCC parametreleri ile ayrı ayrı birleştirilerek kullanılmıştır. Bu sistemler MFCC_SpHr ile MFCC_Hr olarak isimlendirilmiştir. MFCC_SpHr sistemi toplam 16+11=27 öznitelik, MFCC_Hr ise 16+8=24 öznitelik kullanmaktadır. ASE gurubu, MFCC parametreleri ile olan benzerliklerinden dolayı bu amaçla kullanılmamıştır

3.2. Skor Düzeyinde Birleştirme

Bir konuşmacıya ait ses sinyali ile test edildiğinde, her konuşmacı doğrulama sistemi karar aşamasında kullanılmak üzere iki skor değeri üretmektedir. Bunlardan birincisi iddia edilen konuşmacıya ait modelden elde edilen skor, diğeri de referans modelden elde edilen skor değeridir. *N* farklı doğrulama sistemi kullanıldığında *2N* boyutlu birleşik skor uzayı oluşturulmakta ve tanıma veya reddetme kararı bu uzayda verilmektedir.

Bu uzayın oluşturulmasında üç (*N=3*) farklı doğrulama sistemi kullanılmıştır. Bunlar sırasıyla MFCC, MFCC_SpHr ve MFCC_Hr sistemleridir.

2N boyutlu birleşik skor uzayında kararlar üretmek için, bu uzayda çalışan bir ardıl-sınıflandırıcının (post-classifier) geliştirilmesi gerekmektedir. Böyle bir sınıflandırıcının esas amacı yanlış tanıma ve yanlış reddetme oranlarını enküçültmektir. Bu çalışmada, daha önce yapılan çalışmalara göre başarılı olduğu kabul edilen SVM yöntemi kullanılmıştır.

Bu sınıflandırma probleminde iki farklı sınıf oluşmaktadır; iddia edilen konuşmacı kimliğinin doğru olduğu durumlar ve doğru olmadığı durumlar. Toplam *M* konuşmacının kullanıldığı ve her konuşmacı için bir ses sinyali olduğu durumda, toplam *M* doğru kimlik testi ("target test") ve buna karşın *Mx(M-1)* yanlış konuşmacı testi ("impostor test") mümkündür. Bu durumda, ardıl-sınıflandırıcının eğitilmesinde kullanılmak üzere yanlış konuşmacı testine ait çok daha fazla veri elde edilebimektedir. Eğitme verisindeki sınıflar arası farklılık ("class-imbalance") ardıl-sınıflandırıcının eğitme verisi daha az olan sınıfda çok hata yapmasına neden olabilmektedir..

3.2.1. *Destek Vektör Sınıflandırıcı (SVM) yöntemi*

SVM yöntemi temel olarak boşluk (margin)ençoklamaya dayalı bir karar yüzeyi yaratmaktadır. Bu amaçla aşağıdaki karar fonksiyonu kullanılmaktadır:

$$f(\overline{o}) = sign\left(\sum_{i=1}^{S} \alpha_i y_i K(\overline{o}_i, \overline{o}) + b\right)$$

Yukardaki denklemde S, toplam destek vektörü sayısını, $\overline{o}_i$, i'inci destek vektörünü, y_i ise i'inci destek vektörünün etiketini göstermektedir. $K(.,.)$ ise simetrik kernel fonksiyonu olarak bilinmekte ve iki farklı vektörün benzerliğini ölçmektedir. Bu fonksiyon yüksek boyutlu uzayda iç çarpımı gerçekleştirecek şekilde seçilmektedir. α_i parameterleri karesel bir denklemin enazlanması ile hesaplanmaktadır. Bu hesaplamada kullanılan kısıt $\alpha_i \in [0, C]$ şeklindedir. b değeri ise α_i parametreleri hesaplandıktan sonra bulunmaktadır. SVM sınıflandırıcılarının eğitilmesinde kullanıcının belirlediği en önemli parametreler C değeri ile kernel fonksiyonudur. Bu çalışmada doğrusal kernel fonksiyonu kullanılmıştır.

3.2.2. *Çok-yönlü sarsım yöntemleri*

Çoğul sınıflandırıcılı sistemlerin uygun bireysel sınıflandırıcılar kullanıldığında daha iyi kararlar üretebildikleri bilinmektedir. Bu çalışmada, birden fazla SVM tipi sınıflandırıcının birleştirilerek ardıl-sınıflandırıcı olarak kullanılması hedeflenmiştir. Bu tür güçlü sınıflandırıcılar etkin bir şekilde bir arada kullanılabilmeleri ancak uygun koşullar sağlandığında mümkün olabilmektedir. Bireysel olarak sınıflandırma başarımı açısından güçlü sınıflandırıcıların birbirleri ile uyumlu bir küme olabilmeleri için yüksek çeşitlemeye (diversity) sahip olmaları gerekmektedir. Yüksek derecede çeşitleme elde etmek için ise değişik yöntemler ayrı ayrı veya birlikte kullanılabilir [6]. Bu çalışmada eldeki probleme uygun bir takım çeşitleme yaratma yöntemlerinin birlikte kullanılması önerilmektedir. Yukarıda da belirtildiği gibi ardıl-sınıflandırma probleminin en belirgin özelliği sınıflar arası dengesizliktir. Hem dengesizlğin neden olabileceği olumsuzlukları önleyebilmek hem de farklı sınıflandırıcılar yaratmak için uygulanan yöntemler şunlardır:

a) Yanlış konuşmacı testi verisinin toplam sayısı, *pM* dir. *p* katsayısı [1,4] aralığındadır ve her sınıflandırıcı için ayrı bir *p* değeri bu aralıktan rastlantısal olarak seçilmektedir. *pM* kadar yanlış konuşmacı testi verisi rastlantısal olarak her sınıflandırıcı için toplam *Mx(M-1)* veriden farklı olarak seçilmektedir.
b) Her sınıflandırıcı için seçilen *M+pM* verisi içinden *[%30,%70]* aralığındaki bir alt küme rastlantısal olarak seçilir. Böylelikle, her sınıflandırıcı için kullanılan doğru ve yanlış konuşmacı kümesi farklı olur.
c) Her sınıflandırıcı rastlantısal olarak seçilen bir altuzayda (bu altuzayı tam anlamadım) çalışmaktadır. Altuzay boyutu, esas uzayın %75'i ile tamamı arasında rastlantısal olarak seçilmektedir.

d) Ardıl-sınıflandırıcı parametreleri (C değeri) her sınıflandırıcı için ayrı olarak $[0.1,15]$ aralığından rastlantısal olarak seçilmektedir.

4. Konuşmacı Doğrulama Sistemleri

Bu çalışmada deneyler NIST99 veri tabanından seçilmiş $M=150$ konuşmacı içeren iki farklı küme, KÜME1 ve KÜME2 için gerçekleştirilmiştir. Öznitelik vektörleri her 10ms aralıkta Hamming süzgecinden geçirilmiş 30ms uzunluktaki ses çerçevelerinden hesaplanmaktadır. Konuşmacılar Gauss Karışım modelleme (GKM) yöntemi ile temsil edilmiştir. GKM, K bileşenli yoğunluğun ağırlıklı toplamıdır:

$$p(X) = \sum_{i=1}^{K} w_i N(X \mid \mu_i, \Sigma_i)$$

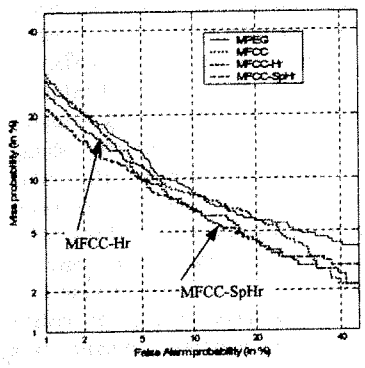
Bu kümeler dışında kalan 46 erkek ile 46 bayan konuşmacı kullanılarak Genel Arka-Plan Model'i (GAM) eğitilmiştir. Bu model *2048* bileşenden oluşan bir GKM'dir. Konuşmacılara ait modeller GAM'dan Bayesçi uyarlama tekniği kullanılarak oluşturulmuştur. Test aşamasında, verilen ses sinyalinin olabilirliği hem GAM hem de idia edilen konuşmacı modeli için hesaplanır. Bireysel sistemlerin başarımlarının hesaplanması için olabilirlik oranı testi aşağıdaki şekilde yapılmaktadır:

$$\Lambda(X) = \log p(X \mid \lambda_S) - \log p(X \mid \lambda_A)$$

Yukardaki denklemde $p(X \mid \lambda_S)$ iddia edilen konuşmacı modelinden elde edilen olabilirlik değeri, $p(X \mid \lambda_A)$ ise GAM'dan elde edilen olabilirlik değeridir.

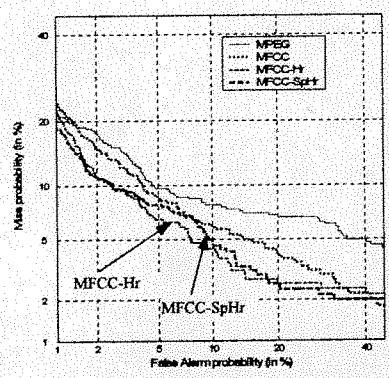
5. Deneysel Sonuçlar

Bireysel sistemlerin başarımlarını gösteren Algılama Hata Ödünleşimi (AHÖ, Detection Error Tradeoff) eğrileri KÜME1 ve KÜME2 için sırasıyla Şekil 1 ve Şekil 2'de görülmektedir.



Şekil 1: KÜME1 üzerindeki deneysel sonuçlar.

Şekillerde MPEG olarak gösterilen eğri *ASE* ile *SpHr* parametrelerinin (toplam 24) birlikte kullanıldığı sistemdir. Bu temel sistemlerin Eşit Hata Oranları (EHO) Tablo 1'de görülmektedir.



Şekil 2: KÜME2 üzerindeki deneysel sonuçlar.

Tablo 1: Temel sistemlerin Eşit Hata Oranları (%)

Konuşmacı Kümesi	KÜME1	KÜME2
MFCC	8.58	7.13
MPEG	9.00	8.53

5.1. Öznitelik Düzeyinde Birleştirme Sonuçları

Şekil 1 ve 2'den de görüleceği üzere *SpHr* ve *Hr* parametrelerinin MFCC parametreleri ile öznitelik düzeyinde birleştirmeleri sonucunda belirgin iyileşme sağlanmaktadır. Elde edilen EHO'ları Tablo 2'de verilmiştir. Son sütun MFCC'ye göre sağlanan yüzde iyileşmeyi göstermektedir.

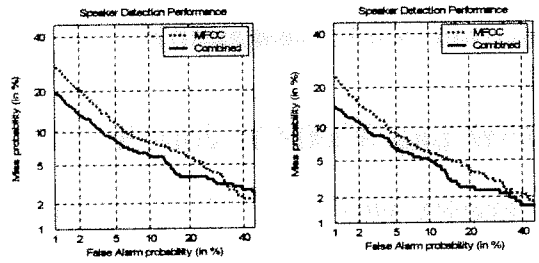
Tablo 2: Öznitelik düzeyinde birleştirilmiş sistemlerin Eşit Hata Oranları(%)

Bireysel Sistemler	KÜME1	KÜME2	Ortalama İyileşme
MFCC_SpHr	7.78	6.94	5.93
MFCC_Hr	7.78	6.35	10.07

5.2. Skor Düzeyinde Birleştirme Sonuçları

Skor düzeyinde SVM sınıflandırıcısı kullanılarak gerşekleştirilen deneyler farklı bireysel sistemler kullanılarak gerşekleştirilmiştir. Öncelikle MFCC, MPEG ve *SpHr* sistemleri birleştirilmiştir. Bu işlem 6 boyutlu uzayda sınıflandırma yapmaya eşdeğerdir. KÜME1 test edilirken SVM ardıl sınıflandırıcısının eğitilmesi için KÜME2'den elde edilen test skorları kullanılmıştır. Benzer şekilde KÜME2 test edilirken SVM için eğitme verisi olarak KÜME2'den elde edilen test skorları kullanılmıştır.

Yukarıda sıralanan sarsım teknikleri 15 kez tekrarlanarak elde edilen 15 bileşenin verdiği sonuçların ortalaması alınarak son karar verilmiştir. Her bileşen rastlantısal olarak 4 veya 5 boyutlu altuzayda ya da 6 boyutlu özgün uzayda eğitilmiştir. SpHr parametrelerinin tek başına kullanıldığı sistem yanlış alarm hatasının yüksek olduğu bölgede MPEG sisteminden daha iyi sonuçlar sağladığı için bu sistemin de bireysel olarak birleştirme işleminde kulanılması uygun bulunmuştur. Şekil 3 (a) ve (b) MFCC, MPEG ve *SpHr* sistemlerinin skor düzeyinde birleştirilmesinden elde edilmiştir.



(a) (b)

Şekil 3: KÜME1(a) ve KÜME2 (b) için MFCC, MPEG ve *SpHr* sistemlerinin skor düzeyinde birleştirilmesi

SpHr sisteminin yanlış alarm hatasının yüksek olduğu bölgede başarımının katkısını değerlendirmek için, bu sistem MFCC ile ikili olarak birleştirilmiştir. Şekil 4, birleştirilmiş sistemin KÜME1 üzerindeki AHÖ eğrisini göstermektedir. Şekilden de görüleceği üzere yanlış alarm hatasının yüksek olduğu bölgede Şekil 3 (a)'ya göre iyileşme sağlanmıştır.

Skor düzeyinde birleştirme sonuçları Tablo 3'de verilmiştir. İlk sütunda KÜME1 test sonuçları, son sütunda, MFCC sistemine göre EHO iyileştirmeleri verilmiştir.

Tablo 3: Skor düzeyinde birleştirilmiş sistemlerin Eşit Hata Oranları(%)

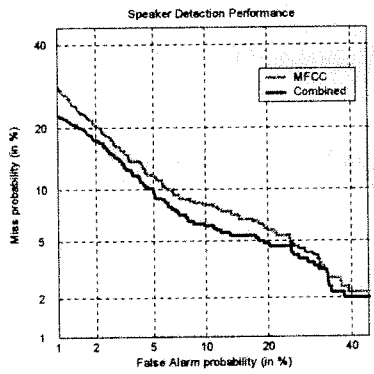
Bireysel Sistemler	KÜME1	KÜME2	Ortalama İyileşme
MFCC MPEG *SpHr*	6.99	5.95	17.48
MFCC MPEG	6.99	6.35	14.67
MFCC *SpHr*	7.19	6.78	10.49

6. Tartışma

MPEG-7 ses niteleyicilerinin bir grubunun, MFCC niteleyicilerine göre daha fazla kaynak bilgisi içerdiği öngörüsüne dayanarak, konuşmacı doğrulamada MPEG-7 niteleyicilerinin katkısı incelenmiştir. MPEG-7 niteleyicileri, MFCC niteleyicileri ile gerek öznitelik düzeyinde gerekse çoğul sınıflandırıcılarla skor düzeyinde birleştirildiğinde kayda değer iyileşme sağlamaktadır. Eşit hata oranı olarak bakıldığında iyileşme % 18'e kadar çıkmakta, en az % 10 olmaktadır. Bu durum çalışmayı başlatan öngörüyü doğrulamaktadır. İlginç olarak değerlendiriliebilecek bir sonuç, harmonik içeriği niteleyen MPEG-7 altkümesinin yanısıra izge zarfını niteleyen, MFCC benzeri MPEG-7 niteleyicilerinin de iyileşmeye katkıda bulunmasıdır. MPEG-7 niteleyicileri ile ilgili bu sonuçlar, dünyada bu alanda yeni bilgilerdir; bilgimiz dahilinde benzer yayınlanmış sonuçlar bulunmamaktadır.

7. Teşekkür

Bu çalışma TÜBİTAK-EEEAG tarafından desteklenmektedir (Proje no: 104E142). Ayrıca, MPEG-7 niteleyicilerini çıkarma yazılımı veren TÜBİTAK-BİLTEN'e ve yazılımı üreten Banu Oskay'a ve Hacer Yalım'a teşekkür ederiz.



Şekil 4: KÜME1 için MFCC ve *SpHr* sistemlerinin skor düzeyinde birleştirilmesi

8. Kaynakça

[1] Davies, S. B. and Mermelstein, P. "Comparison of Parametric Representations for Monosyllabic Word Recognition in Continuously Spoken Sentences," *IEEE Trans. Acoustics, Speech and Signal Processing*, Vol. ASSP-28, No.4, pp.357-376, Aug 1980.

[2] Yang. P, Yang, Y. and Wu, Z., "Exploiting glottal information in speaker information in speaker recognition using parallel GMMs," *Lecture Notes in Computer Science, Springer Verlag* vol. 3546, p. 804, 2005.

[3] Kim, H. G., and Sikora, T., "Comparison of MPEG-7 audio spectrum projection features and MFCC applied to speaker recognition, sound classification and audio segmentation", *Proc. of IEEE Int. Conf. on Acoustics, Speech, and Signal Proc.*, vol. 5, pp. 925-928, May 2004.

[4] Xiong, Z., Radhakrishnan, R., Divakaran, A., and Huang, T. S., "Comparing MFCC and MPEG-7 audio features for feature extraction, maximum likelihood HMM and entropic prior HMM for sports audio classification", *Proc. of Int. Conf. on Multimedia and Expo*, vol. 3, pp. 397-400, July 2003.

[5] Zhou, Z.-H. and Yu. Y., "Ensembling local learners through multimodal perturbation," *IEEE Transactions on Systems, Man, and Cybernetics* - Part B: Cybernetics, 2005, 35(4): 725-735.

[6] Kuncheva, L. I. and Whitaker, "Measures of diversity in classifier ensembles," *Machine Learning*, vol. 51, no. 2, pp.181-207, 2003.

Konuşmacı Doğrulamada MFCC ve MPEG-7 Özniteliklerinin Birleştirilmesi

Fusion of MFCC and MPEG-7 Attributes for Speaker Verification

Hakan Altınçay, Cem Ergün* ve Tolga Çiloğlu***

*Bilgisayar Müh. Böl., Doğu Akdeniz Üniv., KKTC
**Elektrik ve Elektronik Müh. Böl., ODTÜ, Türkiye
{hakan.altincay,cem.ergun}@emu.edu.tr,ciltolga@metu.edu.tr

Özetçe

Bazı MPEG-7 ses niteleyicilerininin MFCClerden daha farklı kaynak bilgisi ("glottal information") içerdiği öngörüsü doğrultusunda MPEG-7 niteleyicilerinin konuşmacı doğrulama probleminde kullanılması araştırılmıştır. MFCClerle hem öznitelik hem de skor düzeyinde birleştirmeye dayalı deneyler yapılmıştır. Sonuçlar, birleştirme ile yalnızca MFCC parametreleri kullanan sisteme göre %18'e varan başarım artışı sağlandığını göstermiştir. Öngörüyle uyumlu bu sonuçlar ilgili alandaki çalışmalar için yeni ipuçları vermektedir.

Abstract

Following the anticipation that some of the MPEG-7 audio descriptors hold glottal information differing than those MFCCs hold, possible contribution of MPEG-7 descriptors to speaker verification has been investigated. Both feature level and score level fusion of MFCCs and MPEG-7 descriptors have been studied. Results indicate improvements up to 18 % compared to those obtained by using MFCCs alone; a justification of the anticipation and a novel indication to the community.

1. Giriş

Bu çalışmanın amacı MPEG-7 standardı çerçevesinde tanımlanmış olan ses niteleyicilerin ("audio descriptors") konuşmacı doğrulamada ("speaker verification") ne kadar yararlı olabileceklerini; söz konusu niteleyicilerden hangilerinin daha yararlı olabileceğini incelemektir. MPEG-7 niteleyicilerinin belirli alt kümelerinin tek başlarına niteleyici ya da öznitelik grubu olarak kullanılması öngörülmektedir. Çünkü, MFCC ("Mel Frequency Cepstral Coefficients") niteleyicileri ile konuşmacı tanımada/ doğrulamada ulaşılan başarım değerleri, dünyada MFCClerin akustik öznitelik olarak güvenilirliğini kanıtlamıştır [1]. Kavramsal karşılaştırmaya göre MPEG-7 niteleyicilerinin, MFCClere göre kayda değer düzeyde daha iyi sonuçlar verebileceği yönünde beklentimiz bulunmamaktadır; tutarlı olarak elde edilecek böyle bir sonuç ilginç, hatta şaşırtıcı olur. Beklentimiz MPEG-7 niteleyicilerini, MFCClerle birlikte kullanarak daha iyi sonuçlar elde edebilmektir. Bunun temel nedeni MPEG-7 niteleyicilerinden bazılarının MFCClerden farklı bilgi taşıdığı öngörümüzdür. MFCClerin esas olarak sinyal izge genliğinin tepe hattına ait bilgi verdiği bilinir. Bu parametreler ağırlıklı olarak insan ses yolunu modellemektedir. Öte yandan, konuşma sinyalinin oluşumunda kaynak olan "glottal" bilgiler de konuşmacıya has özellikler taşır [2]. MPEG-7 standardında yer alan bazı ses niteleyicilerin bu tür bilgileri içerdiği düşünülmektedir. Örneğin harmonik bileşenlerin ses sinyalindeki gürültü içeriğine baskınlığı, harmonik bileşenlerin yayılması ile harmonik izgenin dağılımı gibi niteleyiciler MPEG-7 standardında tanımlanmıştır.

Literatürde, MPEG-7 standardında yer alan Ses İzgesi Zarfı (audio spectrum envelope, ASE) ile konuşmacı tanıma deneyleri yapılmıştır [3,4]. Bu parametrelerin MFCC'lerden esas farkı kullanılan frekans ölçeğinin doğrusal olmasıdır. ASE parametrelerinin başarımı 22.05kHz örnekleme sıklığında MFCC parametrelerine göre biraz daha kötüdür.

Bu çalışmada MPEG-7 akustik niteleyicilerinin tümü incelenmektedir. İncelemenin bütünlüğü bakımından öncelikle MPEG-7 niteleyicilerinin bireysel başarımları elde edilmiştir. Daha sonra MFCC parametreleri ile hem öznitelik düzeyinde hem de farklı MPEG-7 altkümeleri kullanan konuşmacı doğrulama sistemleri ile skor düzeyinde birleştirilmeleri üzerinde deneyler yapılmıştır. Skor düzeyinde sınıflandırıcı birleştirme uygulamalarında çok-yönlü sarsım'ın (multi-modal perturbation) önemi son dönemlerde literatürde vurgulanmaktadır [5]. Bu çalışmada, ortaya çıkan birleştirme problemine uygun olan bir *çok-yönlü sarsım* yaklaşımı önerilmiş ve bu yaklaşım birden fazla destek vektör sınıflandırıcısı'nın ("support vector machines", SVM) skor düzeyinde birleştirme yapmaları amacıyla eğitilmelerinde kullanılmıştır. Yapılan deneyler, MPEG-7 standardında tanımlanmış olan parametrelerin MFCC'lere tamamlayıcı olduklarını ve %18'e yakın başarım artışı sağladıklarını göstermiştir. Bu sonuçlar literatürde yenidir.

2. Kullanılan MPEG-7 Öznitelikleri

MPEG-7 standardında tanımlanmış ses niteleyicileri bu çalışmada üç grup halinde ele alınmaktadır:

ASE:(toplam 13 öznitelik) Audio Spectrum Envelope parameters.
SpHr: (toplam 11 öznitelik) Audio spectrum centroid, Audio spectrum spread, Audio fund. freq. and related confidence measure, Audio harmonicity and its upper limit, Harmonic spectral centroid, Harmonic spectral spread, Harmonic spectral deviation, Harmonic spectral variation, Spectral centroid.
Hr: (toplam 8 öznitelik) Audio fund. freq. and related confidence measure, Audio harmonicity and its upper limit, Harmonic spectral centroid, Harmonic spectral spread, Harmonic spectral deviation, Harmonic spectral variation.

SpHr grubunun altkümesi olan *Hr* gurubu parametreleri ağırlıklı olarak harmoniklerle ilgili ölçümleri içermektedir.

3. MFCC ve MPEG-7 Özniteliklerinin birlikte kullanılması

Birden fazla öznitelik vektörünün birlikte kullanılması için iki temel yaklaşım bulunmaktadır. Bunlardan birincisi öznitelik düzeyinde birleştirmedir. Bu yöntemde, farklı öznitelik vektörleri eklenerek birleşik vektör uzayı oluşturulmakta ve konuşmacı doğrulama sistemi bu uzayda geliştirilmektedir. İkinci yöntem ise her öznitelik vektörü ile farklı bir doğrulama sistemi geliştirmeyi öngörmektedir. Farklı sistemlerin sağladığı skorlar kullanılarak kararlar verilmektedir.

Öncelikle, birleştirilmiş sistemlerin başarımlarının değerlendirilebilmesinde referans olarak kullanılmak üzere iki sistem geliştirilmiştir. Bunlardan birincisinde 16 MFCC parametresi, ikincisinde ise *ASE* ve *SpHr* gruplarındaki toplam 24 öznitelik kullanılmıştır. İkinci sistem ASE_SpHr olarak anılacaktır.

3.1. Öznitelik Düzeyinde Birleştirme

Öznitelik düzeyinde birleştirmede yukarıda tanımlanan *SpHr* grubu ile *Hr* grubu MFCC parametreleri ile ayrı ayrı birleştirilerek kullanılmıştır. Bu sistemler MFCC_SpHr ile MFCC_Hr olarak isimlendirilmiştir. MFCC_SpHr sistemi toplam 16+11=27 öznitelik, MFCC_Hr ise 16+8=24 öznitelik kullanmaktadır. ASE gurubu, MFCC parametreleri ile olan benzerliklerinden dolayı bu amaçla kullanılmamıştır

3.2. Skor Düzeyinde Birleştirme

Bir konuşmacıya ait ses sinyali ile test edildiğinde, her konuşmacı doğrulama sistemi karar aşamasında kullanılmak üzere iki skor değeri üretmektedir. Bunlardan birincisi iddia edilen konuşmacıya ait modelden elde edilen skor, diğeri de referans modelden elde edilen skor değeridir. *N* farklı doğrulama sistemi kullanıldığında *2N* boyutlu birleşik skor uzayı oluşturulmakta ve tanıma veya reddetme kararı bu uzayda verilmektedir.

Bu uzayın oluşturulmasında üç (*N=3*) farklı doğrulama sistemi kullanılmıştır. Bunlar sırasıyla MFCC, MFCC_SpHr ve MFCC_Hr sistemleridir.

2N boyutlu birleşik skor uzayında kararlar üretmek için, bu uzayda çalışan bir ardıl-sınıflandırıcının (post-classifier) geliştirilmesi gerekmektedir. Böyle bir sınıflandırıcının esas amacı yanlış tanıma ve yanlış reddetme oranlarını enküçültmektir. Bu çalışmada, daha önce yapılan çalışmalara göre başarılı olduğu kabul edilen SVM yöntemi kullanılmıştır.

Bu sınıflandırma probleminde iki farklı sınıf oluşmaktadır; iddia edilen konuşmacı kimliğinin doğru olduğu durumlar ve doğru olmadığı durumlar. Toplam *M* konuşmacının kullanıldığı ve her konuşmacı için bir ses sinyali olduğu durumda, toplam *M* doğru kimlik testi ("target test") ve buna karşın *Mx(M-1)* yanlış konuşmacı testi ("impostor test") mümkündür. Bu durumda, ardıl-sınıflandırıcının eğitilmesinde kullanılmak üzere yanlış konuşmacı testine ait çok daha fazla veri elde edilebimektedir. Eğitme verisindeki sınıflar arası farklılık ("class-imbalance") ardıl-sınıflandırıcının eğitme verisi daha az olan sınıfda çok hata yapmasına neden olabilmektedir..

3.2.1. Destek Vektör Sınıflandırıcı (SVM) yöntemi

SVM yöntemi temel olarak boşluk (margin)ençoklamaya dayalı bir karar yüzeyi yaratmaktadır. Bu amaçla aşağıdaki karar fonksiyonu kullanılmaktadır:

$$f(\overline{o}) = sign\left(\sum_{i=1}^{s} \alpha_i y_i K(\overline{o}_i, \overline{o}) + b\right)$$

Yukardaki denklemde S, toplam destek vektörü sayısını, $\overline{o}_i$, i'inci destek vektörünü, y_i ise i'inci destek vektörünün etiketini göstermektedir. $K(.,.)$ ise simetrik kernel fonksiyonu olarak bilinmekte ve iki farklı vektörün benzerliğini ölçmektedir. Bu fonksiyon yüksek boyutlu uzayda iç çarpımı gerçekleştirecek şekilde seçilmektedir. α_i parameterleri karesel bir denklemin enazlanması ile hesaplanmaktadır. Bu hesaplamada kullanılan kısıt $\alpha_i \in [0, C]$ şeklindedir. b değeri ise α_i parametreleri hesaplandıktan sonra bulunmaktadır. SVM sınıflandırıcılarının eğitilmesinde kullanıcının belirlediği en önemli parametreler C değeri ile kernel fonksiyonudur. Bu çalışmada doğrusal kernel fonksiyonu kullanılmıştır.

3.2.2. Çok-yönlü sarsım yöntemleri

Çoğul sınıflandırıcılı sistemlerin uygun bireysel sınıflandırıcılar kullanıldığında daha iyi kararlar üretebildikleri bilinmektedir. Bu çalışmada, birden fazla SVM tipi sınıflandırıcının birleştirilerek ardıl-sınıflandırıcı olarak kullanılması hedeflenmiştir. Bu tür güçlü sınıflandırıcılar etkin bir şekilde bir arada kullanılabilmeleri ancak uygun koşullar sağlandığında mümkün olabilmektedir. Bireysel olarak sınıflandırma başarımı açısından güçlü sınıflandırıcıların birbirleri ile uyumlu bir küme olabilmeleri için yüksek çeşitlemeye (diversity) sahip olmaları gerekmektedir. Yüksek derecede çeşitleme elde etmek için ise değişik yöntemler ayrı ayrı veya birlikte kullanılabilir [6]. Bu çalışmada eldeki probleme uygun bir takım çeşitleme yaratma yöntemlerinin birlikte kullanılması önerilmektedir. Yukarıda da belirtildiği gibi ardıl-sınıflandırma probleminin en belirgin özelliği sınıflar arası dengesizliktir. Hem dengesizlğin neden olabileceği olumsuzlukları önleyebilmek hem de farklı sınıflandırıcılar yaratmak için uygulanan yöntemler şunlardır:

a) Yanlış konuşmacı testi verisinin toplam sayısı, *pM*'dir. *p* katsayısı [1,4] aralığındadır ve her sınıflandırıcı için ayrı bir *p* değeri bu aralıktan rastlantısal olarak seçilmektedir. *pM* kadar yanlış konuşmacı testi verisi rastlantısal olarak her sınıflandırıcı için toplam *Mx(M-1)* veriden farklı olarak seçilmektedir.
b) Her sınıflandırıcı için seçilen *M+pM* verisi içinden *[%30,%70]* aralığındaki bir alt küme rastlantısal olarak seçilir. Böylelikle, her sınıflandırıcı için kullanılan doğru ve yanlış konuşmacı kümesi farklı olur.
c) Her sınıflandırıcı rastlantısal olarak seçilen bir altuzayda (bu altuzayı tam anlamadım) çalışmaktadır. Altuzay boyutu, esas uzayın %75'i ile tamamı arasında rastlantısal olarak seçilmektedir.

d) Ardıl-sınıflandırıcı parametreleri (C değeri) her sınıflandırıcı için ayrı olarak $[0.1,15]$ aralığından rastlantısal olarak seçilmektedir.

4. Konuşmacı Doğrulama Sistemleri

Bu çalışmada deneyler NIST99 veri tabanından seçilmiş $M=150$ konuşmacı içeren iki farklı küme, KÜME1 ve KÜME2 için gerçekleştirilmiştir. Öznitelik vektörleri her 10ms aralıkta Hamming süzgecinden geçirilmiş 30ms uzunluktaki ses çerçevelerinden hesaplanmaktadır. Konuşmacılar Gauss Karışım modelleme (GKM) yöntemi ile temsil edilmiştir. GKM, K bileşenli yoğunluğun ağırlıklı toplamıdır:

$$p(X) = \sum_{i=1}^{K} w_i N(X \mid \mu_i, \Sigma_i)$$

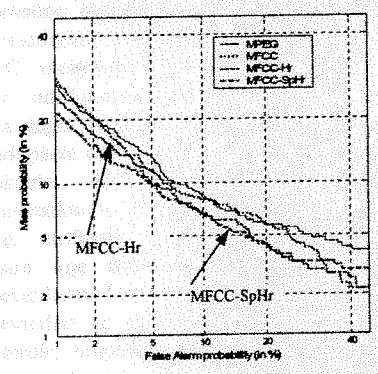
Bu kümeler dışında kalan 46 erkek ile 46 bayan konuşmacı kullanılarak Genel Arka-Plan Model'i (GAM) eğitilmiştir. Bu model *2048* bileşenden oluşan bir GKM'dir. Konuşmacılara ait modeller GAM'dan Bayesçi uyarlama tekniği kullanılarak oluşturulmuştur. Test aşamasında, verilen ses sinyalinin olabilirliği hem GAM hem de idia edilen konuşmacı modeli için hesaplanır. Bireysel sistemlerin başarımlarının hesaplanması için olabilirlik oranı testi aşağıdaki şekilde yapılmaktadır:

$$\Lambda(X) = \log p(X \mid \lambda_S) - \log p(X \mid \lambda_A)$$

Yukardaki denklemde $p(X \mid \lambda_S)$ iddia edilen konuşmacı modelinden elde edilen olabilirlik değeri, $p(X \mid \lambda_A)$ ise GAM'dan elde edilen olabilirlik değeridir.

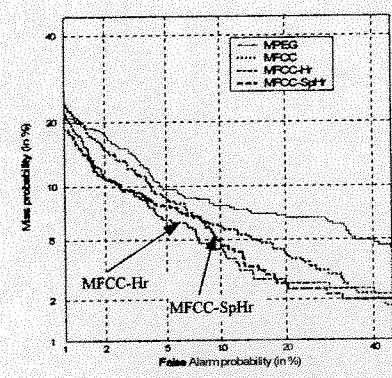
5. Deneysel Sonuçlar

Bireysel sistemlerin başarımlarını gösteren Algılama Hata Ödünleşimi (AHÖ, Detection Error Tradeoff) eğrileri KÜME1 ve KÜME2 için sırasıyla Şekil 1 ve Şekil 2'de görülmektedir.



Şekil 1: KÜME1 üzerindeki deneysel sonuçlar.

Şekillerde MPEG olarak gösterilen eğri *ASE* ile *SpHr* parametrelerinin (toplam 24) birlikte kullanıldığı sistemdir. Bu temel sistemlerin Eşit Hata Oranları (EHO) Tablo 1'de görülmektedir.



Şekil 2: KÜME2 üzerindeki deneysel sonuçlar.

Tablo 1: Temel sistemlerin Eşit Hata Oranları (%)

Konuşmacı Kümesi	KÜME1	KÜME2
MFCC	8.58	7.13
MPEG	9.00	8.53

5.1. Öznitelik Düzeyinde Birleştirme Sonuçları

Şekil 1 ve 2'den de görüleceği üzere *SpHr* ve *Hr* parametrelerinin MFCC parametreleri ile öznitelik düzeyinde birleştirmeleri sonucunda belirgin iyileşme sağlanmaktadır. Elde edilen EHO'ları Tablo 2'de verilmiştir. Son sütun MFCC'ye göre sağlanan yüzde iyileşmeyi göstermektedir.

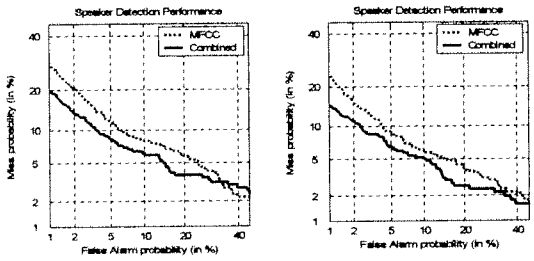
Tablo 2: Öznitelik düzeyinde birleştirilmiş sistemlerin Eşit Hata Oranları(%)

Bireysel Sistemler	KÜME1	KÜME2	Ortalama İyileşme
MFCC_SpHr	7.78	6.94	5.93
MFCC_Hr	7.78	6.35	10.07

5.2. Skor Düzeyinde Birleştirme Sonuçları

Skor düzeyinde SVM sınıflandırıcısı kullanılarak gerşekleştirilen deneyler farklı bireysel sistemler kullanılarak gerşekleştirilmiştir. Öncelikle MFCC, MPEG ve *SpHr* sistemleri birleştirilmiştir. Bu işlem 6 boyutlu uzayda sınıflandırma yapmaya eşdeğerdir. KÜME1 test edilirken SVM ardıl sınıflandırıcısının eğitilmesi için KÜME2'den elde edilen test skorları kullanılmıştır. Benzer şekilde KÜME2 test edilirken SVM için eğitme verisi olarak KÜME2'den elde edilen test skorları kullanılmıştır.

Yukarıda sıralanan sarsım teknikleri 15 kez tekrarlanarak elde edilen 15 bileşenin verdiği sonuçların ortalaması alınarak son karar verilmiştir. Her bileşen rastlantısal olarak 4 veya 5 boyutlu altuzayda ya da 6 boyutlu özgün uzayda eğitilmiştir. SpHr parametrelerinin tek başına kullanıldığı sistem yanlış alarm hatasının yüksek olduğu bölgede MPEG sisteminden daha iyi sonuçlar sağladığı için bu sistemin de bireysel olarak birleştirme işleminde kulanılması uygun bulunmuştur. Şekil 3 (a) ve (b) MFCC, MPEG ve *SpHr* sistemlerinin skor düzeyinde birleştirilmesinden elde edilmiştir.



(a) (b)

Şekil 3: KÜME1(a) ve KÜME2 (b) için MFCC, MPEG ve *SpHr* sistemlerinin skor düzeyinde birleştirilmesi

SpHr sisteminin yanlış alarm hatasının yüksek olduğu bölgede başarımının katkısını değerlendirmek için, bu sistem MFCC ile ikili olarak birleştirilmiştir. Şekil 4, birleştirilmiş sistemin KÜME1 üzerindeki AHÖ eğrisini göstermektedir. Şekilden de görüleceği üzere yanlış alarm hatasının yüksek olduğu bölgede Şekil 3 (a)'ya göre iyileşme sağlanmıştır.

Skor düzeyinde birleştirme sonuçları Tablo 3'de verilmiştir. İlk sütunda KÜME1 test sonuçları, son sütunda, MFCC sistemine göre EHO iyileştirmeleri verilmiştir.

Tablo 3: Skor düzeyinde birleştirilmiş sistemlerin Eşit Hata Oranları(%)

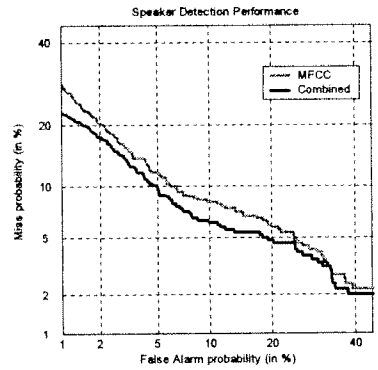
Bireysel Sistemler	KÜME1	KÜME2	Ortalama İyileşme
MFCC MPEG *SpHr*	6.99	5.95	17.48
MFCC MPEG	6.99	6.35	14.67
MFCC *SpHr*	7.19	6.78	10.49

6. Tartışma

MPEG-7 ses niteleyicilerinin bir grubunun, MFCC niteleyicilerine göre daha fazla kaynak bilgisi içerdiği öngörüsüne dayanarak, konuşmacı doğrulamada MPEG-7 niteleyicilerinin katkısı incelenmiştir. MPEG-7 niteleyicileri, MFCC niteleyicileri ile gerek öznitelik düzeyinde gerekse çoğul sınıflandırıcılarla skor düzeyinde birleştirildiğinde kayda değer iyileşme sağlamaktadır. Eşit hata oranı olarak bakıldığında iyileşme % 18'e kadar çıkmakta, en az % 10 olmaktadır. Bu durum çalışmayı başlatan öngörüyü doğrulamaktadır. İlginç olarak değerlendiriliebilecek bir sonuç, harmonik içeriği niteleyen MPEG-7 altkümesinin yanısıra izge zarfını niteleyen, MFCC benzeri MPEG-7 niteleyicilerinin de iyileşmeye katkıda bulunmasıdır. MPEG-7 niteleyicileri ile ilgili bu sonuçlar, dünyada bu alanda yeni bilgilerdir; bilgimiz dahilinde benzer yayınlanmış sonuçlar bulunmamaktadır.

7. Teşekkür

Bu çalışma TÜBİTAK-EEEAG tarafından desteklenmektedir (Proje no: 104E142). Ayrıca, MPEG-7 niteleyicilerini çıkarma yazılımı veren TÜBİTAK-BİLTEN'e ve yazılımı üreten Banu Oskay'a ve Hacer Yalım'a teşekkür ederiz.



Şekil 4: KÜME1 için MFCC ve *SpHr* sistemlerinin skor düzeyinde birleştirilmesi

8. Kaynakça

[1] Davies, S. B. and Mermelstein, P. "Comparison of Parametric Representations for Monosyllabic Word Recognition in Continuously Spoken Sentences," *IEEE Trans. Acoustics, Speech and Signal Processing*, Vol. ASSP-28, No.4, pp.357-376, Aug 1980.

[2] Yang. P, Yang, Y. and Wu, Z., "Exploiting glottal information in speaker information in speaker recognition using parallel GMMs," *Lecture Notes in Computer Science, Springer Verlag* vol. 3546, p. 804, 2005.

[3] Kim, H. G., and Sikora, T., "Comparison of MPEG-7 audio spectrum projection features and MFCC applied to speaker recognition, sound classification and audio segmentation", *Proc. of IEEE Int. Conf. on Acoustics, Speech, and Signal Proc.*, vol. 5, pp. 925-928, May 2004.

[4] Xiong, Z., Radhakrishnan, R., Divakaran, A., and Huang, T. S., "Comparing MFCC and MPEG-7 audio features for feature extraction, maximum likelihood HMM and entropic prior HMM for sports audio classification", *Proc. of Int. Conf. on Multimedia and Expo*, vol. 3, pp. 397-400, July 2003.

[5] Zhou, Z.-H. and Yu. Y., "Ensembling local learners through multimodal perturbation," *IEEE Transactions on Systems, Man, and Cybernetics* - Part B: Cybernetics, 2005, 35(4): 725-735.

[6] Kuncheva, L. I. and Whitaker, "Measures of diversity in classifier ensembles," *Machine Learning*, vol. 51, no. 2, pp.181-207, 2003.