

# INTERSPEECH 2009 Duygu Tanıma Yarışması Değerlendirmesi

## INTERSPEECH 2009 Emotion Recognition Challenge Evaluation

Elif Bozkurt<sup>1</sup>, Engin Erzin<sup>1</sup>, Çiğdem Eroğlu Erdem<sup>2</sup>, A. Tanju Erdem<sup>3</sup>

1. Elektrik ve Bilgisayar Mühendisliği Bölümü  
Koç Üniversitesi, İstanbul  
{ebozkurt,erzin}@ku.edu.tr
2. Elektrik Elektronik Mühendisliği Bölümü  
Bahçeşehir Üniversitesi, İstanbul  
cigdem.eroglu@bahcesehir.edu.tr
3. Elektrik Elektronik Mühendisliği Bölümü  
Özyeğin Üniversitesi, İstanbul  
tanju.erdem@ozyegin.edu.tr

### Özetçe

Bu makalede INTERSPEECH 2009 Duygu Tanıma Yarışması sonuçlarını değerlendiriyoruz. Yarışmanın sunduğu problem doğal ve duygu bakımından zengin FAU Aibo konuşma kayıtlarının beş ve iki duygu sınıfına en doğru şekilde ayrılmasıdır. Bu problemi çözmek için bürün ilintili, spektral ve SMM-temelli (saklı Markov model) öznitelikleri Gauss Bileşen Model (GBM) sınıflandırıcıları ile inceliyoruz. Spektral öznitelikler, Mel frekans kepsral katsayılarını (MFKK), doğru spektral frekans (DSF) katsayılarını ve bunların türevlerini içerirken, bürün öznitelikleri perde, perdenin birinci türevi ve enerjiden oluşuyor. Bürün ilintili özniteliklerin zamanla değişimini tanımlayan SMM özniteliklerini, güdümsüz eğitilen SMM yapıları ile elde ediyoruz. Ayrıca, konuşmadan duygu tanıma sonuçlarını iyileştirmek için farklı özniteliklerin veri kaynaşımını ve farklı sınıflandırıcıların karar kaynaşımını da inceliyoruz. İki aşamalı karar kaynaşımı yöntemimiz beş ve iki sınıflı problemler için sırasıyla, % 41.59 ve % 67.90 başarı oranını ve tüm yarışma sonuçları arasında 2. ve 4. sırayı elde etti .

### Abstract

In this paper we evaluate INTERSPEECH 2009 Emotion Recognition Challenge results. The challenge presents the problem of accurate classification of natural and emotionally rich FAU Aibo recordings into five and two emotion classes. We evaluate prosody related, spectral and HMM-based features with Gaussian mixture model (GMM) classifiers to attack this problem. Spectral features consist of mel-scale cepstral coefficients (MFCC), line spectral frequency (LSF) features and their derivatives, whereas prosody-related features consist of pitch, first derivative of pitch and

intensity. We employ unsupervised training of HMM structures with prosody related temporal features to define HMM-based features. We also investigate data fusion of different features and decision fusion of different classifiers to improve emotion recognition results. Our two-stage decision fusion method achieves 41.59 % and 67.90 % recall rate for the five and two-class problems, respectively and takes second and fourth place among the overall challenge results .

### 1. Giriş

Konuşmadan duygu tanıma çalışmalarında kullanılan veri tabanlarının içerdikleri duygu çeşitliliği, toplam kayıt süresi, doğal ya da rol yapılarak kaydedilmiş olması gibi farklılıklar göz önüne alındığında, bu çalışmalardaki yöntemlerin sağlıklı bir şekilde karşılaştırılması güçleşir. INTERSPEECH 2009 Duygu Tanıma Yarışmasının sunduğu FAU Aibo duygu yüklü konuşma veri tabanının beş ve iki duygu sınıfına ayrılması problemleri farklı yöntemlerin kıyaslanması olanağını sunuyor [1]. Bu yarışmada kullandığımız yeni öznitelikler ve iki-aşamalı karar kaynaşımı yöntemi ile beş ve iki sınıflı problemler için sırasıyla, % 41.59 ve % 67.90 başarı oranları ile yarışmada ikinci ve dördüncü sırayı elde ettik.

FAU Aibo veri tabanı eğitim ve test bölümlerinden oluşuyor. İki farklı okulda, yaşları 10-13 arasında değişen 51 çocuğun (30 kız, 21 erkek) Aibo adlı bir robotla etkileşimi sırasında kaydedilen, doğal duygu yüklü Almanca dilindeki kayıtlar *öfke*, *olumlu*, *empatik*, *nötr* ve *diğerleri* olarak beş, *olumsuz* ve *olumsuz olmayan* olarak iki sınıfa ayrılmış [1]. Sonuçların sıralaması ise bu beş ya da iki duygunun tanıma oranlarının aritmetik ortalamasına (AO) göre hesaplanıyor. İkinci bir ortalama değeri ise ağırlıklı aritmetik

Tablo 1: INTERSPEECH 2009 Konuşmadan Duygu Tanıma Yarışması Sonuçları

Makale Başlığı - Yazarlar	Yöntem	Tanıma Oranı (Sıralama)	
		2 sınıf	5 sınıf
<i>Brno University of Technology System for Interspeech 2009 Emotion Challenge</i> Kockmann, Burget ve Cernocky	Birinci ve ikinci türevleriyle Mel frekans kepstral katsayıları (MFKK $\Delta\Delta$ ) öznitelik vektörleri birleşik etmen analiz yöntemi ile modellendi.	68.3 (3)	41.7 (1)
<i>Improving Automatic Emotion Recognition from Speech Signals</i> Bozkurt, Erzin, E.Erdem ve Erdem	MFKK $\Delta\Delta$ öznitelik vektörleri ve DSF öznitelikleri ile Gauss bileşen modeli (GBM) eğitilip kaynaştırıldı. SMM GBM ile ikinci aşama kaynaşım uygulandı.	67.9 (4)	41.59 (2)
<i>GTM-URL Contribution to the Interspeech 2009 Emotion Challenge</i> Planet, Iriondo, Socoro, Monzo ve Adell	Sıfır kesiş oranı (SKO), etkin enerji değeri, bürün, harmonik gürültü oranı (HGO), MFKK istatistiksel verileri Naive Bayes ile modellendi	-	41.16 (5)
<i>Acoustic Emotion Recognition Using Dynamic Bayesian Networks and Multi-space Distributions</i> Chicote, Fernandez, Lutfi, Cuesta, Guarasa, Montero, Segundo ve Pardo	MFKK $\Delta\Delta$ , log enerji ve log bürün öznitelikleri dinamik Bayes ağ ile modellendi.	67.06 (7)	38.24 (8)
<i>Emotion Recognition Using A Hierarchical Binary Decision Tree Approach</i> Lee, Mower, Busso, Lee ve Narayanan	SKO, enerji, bürün, HGO, MFKK istatistiksel verileri z-normalleştirmeden sonra Bayes biçimsel bağlanım ile modellendi.	-	41.57 (3)
<i>Combining Spectral and Prosodic Information for Emotion recognition in the Interspeech 2009 Emotion Challenge</i> Luengo, Navas ve Hernaez	Mel ölçekli kısa süreli log frekans güç katsayıları GBM ile; tonlama, güç, ritim, bağlanım, ses kalitesi istatistikleri radyan temel fonksiyonlu – karar destek makinesi (RTF-KDM) ile modellendi, kaynaştırıldı.	67.19 (6)	41.38 (4)
<i>Emotion Classification in Children's Speech Using Fusion of Acoustic and Linguistic Features</i> Polzehl, Sundaram, Ketabdar, Wagner ve Metze	Konuşma tanıma ile duygu belirten kelimeler tanındı ve RTF-KDM ile sınıflandırıldı. Enerji, bürün, biçimlendirici, SKO, süre ve HGO istatistikleri RTF-KDM ile modellendi. Bu iki sistem kaynaştırıldı.	67.55 (5)	-
<i>Cepstral and Long-term Features for Emotion Recognition</i> Dumouchel, Dehak, Attabi, Dehak ve Boufaden	MFKK $\Delta\Delta$ vektörleri ile GBM eğitildi, sonra her duyguya uyarlandı*. GBM süper vektör öznitelikleri KDM ile modellendi. Bu iki sistem kaynaştırıldı**.	*69.72 (2) **70.2 (1)	39.4 (6)
<i>Exploring the Benefits of Discretization of Acoustic Features for Speech Emotion Recognition</i> Vogt ve Andre	Bürün, enerji, spektral, kepstral ses segmentleri, ses kalitesi, seyirme öznitelikleri istatistikleri Naive Bayes ile modellendi.	66.4 (8)	39.4 (7)

ortalama (AAO) ki bu değer bir duygunun test kayıtlarının tüm test kayıtlarına oranı dikkate alınarak belirleniyor; ancak sıralama için bir ölçüt olarak kabul edilmiyor [1].

Yarışma katılımcılarının kullandıkları yöntemler ve elde ettikleri sonuçlar Tablo 1’ de özetleniyor [2]. Bu çalışmalarda en sık kullanılan öznitelik olarak birinci ve ikinci türevleriyle Mel frekans kepstral katsayıları (MFKK) göze çarpıyor. Ayrıca, enerji, bürün ve harmonik gürültü oranı ve sıfır kesiş oranı (SKO) istatistiksel verileri de sık kullanılan öznitelikler arasında sayılabilir. Bu öznitelik vektörlerinin sınıflandırılması içinse Gauss bileşen modeli (GBM), karar destek makinesi (KDM) ve Naive Bayes sınıflandırıcıları en çok tercih edilen yöntemler arasında yer alıyor.

Konuşmadan duygu tanıma sistemimizde, MFKK ve bürün özniteliklerine ek olarak, daha önce konuşmadan duygu tanıma problemi için kullanılmamış olan Doğru Spektral Frekans (DSF) ve SMM-temelli öznitelikleri de kullandık. Sınıflandırıcı olarak GBM seçtik ve iki

aşamalı karar kaynaşımı yöntemi ile hem beş hem de iki sınıflı problemler için en yüksek AO tanıma oranlarımızı elde ettik.

Makalenin içeriği şu şekilde düzenlenmiştir: Bölüm 2’ de, kullandığımız öznitelikleri detaylandırıyoruz. Bölüm 3’ te sınıflandırma için kullandığımız GBM yapılarını açıklıyoruz. Deneysel sonuçlarımızı Bölüm 4’te sunuyoruz. Son olarak, Bölüm 5’ te sonuçlarımızı değerlendiriyor ve tartışıyoruz.

## 2. Duygu Tanıma için Öznitelik İfadeleri

Konuşmacının duygusal durumunu sınıflandırmak için ses sinyalinin hem bürün ilintili hem de spektral özniteliklerini kullandık. Bürün özniteliklerini perde, perdenin birinci türevi ve enerji olarak tanımladık ve  $f_b$  ile ifade ettik. Perde özniteliklerini öz ilinti yöntemi ile hesapladık [3]. Nitekim, perde değerleri konuşmacıya bağlı olduğundan, sıfırdan farklı perde değerleri ortalamasını her perde değerinden çıkararak konuşmacılar arası eşitliği sağladık.

Spektral öznitelik olarak Mel-frekans kepsral katsayılarını (MFKK) ve Itakura tarafından tanımlanan doğru spektral frekans (DSF) katsayılarını kullandık [4]. Toplam 12 kepsral katsayı ve enerji teriminden oluşan 13 boyutlu MFKK vektörlerini  $f_M$  ile ifade ettik. 16 dereceli DSF özniteliklerini  $f_D$  ile simgeledik.

Ses tayfındaki zamana bağlı değişimleri dinamik öznitelikler ile ifade edebiliriz. MFKK öznitelik vektörünü birinci ve ikinci türev öznitelikleri de içerecek şekilde  $f_{MA} = [f_M', \Delta f_M', \Delta \Delta f_M']$  olarak genişlettik. Aynı şekilde, DSF özniteliklerine dinamik özniteliklerini ekleyip  $f_{DA}$  ile simgeledik.

Ayrıca, sıfırdan farklı perde segmentleri için perde-enerji ve MFKK özniteliklerini birleştirerek  $f_{BM}$  öznitelik vektörünü ve dinamik özellikleri de ekleyerek  $f_{BMA}$  öznitelik vektörünü elde ettik.

SMM-temelli öznitelikler için farklı duygulara ait duygu yüklü konuşma kayıtlarını kullanarak, her dalında soldan sağa  $N$  tane durumu bulunan,  $B$  paralel dallı  $\Lambda$  SMM yapısını bürün temelli öznitelikler ile güdümsüz eğittik. Daha sonra, her dalın bir duyguya ait bürün örüntüsünü modellediğini varsayarak, paralel dallı  $\Lambda$  SMM yapısını, tek dallı  $\Lambda_1, \Lambda_2, \dots, \Lambda_B$  yapılarına ayırdık. SMM yapısının  $i$ 'nci dalı için  $U$  cümlesinin kestirim değerini aşağıdaki gibi ifade edebiliriz,

$$p_i = P(U | \lambda_i) \quad (1)$$

Sigmoid normalleştirmesini her cümle için kestirim değerini  $[0,1]$  aralığına eşlemek için kullandık [5]. Cümlelerin bu yeni kestirim değerleri ile SMM-tabanlı öznitelikler kümesini oluşturduk ve  $f_S$  ile ifade ettik.

$$f_S(i) = \left[ 1 + e^{-\left(\frac{p_i - \bar{p}}{2\sigma} + 1\right)} \right]^{-1} \quad (2)$$

Yukarıdaki denklemde eğitim veri tabanında  $p_i$  kestirim değerlerinin ortalama ve standart sapma değerlerini sırasıyla  $\bar{p}$  ve  $\sigma$  ile belirtiyoruz.  $f_{BM}$  ve  $f_{BMA}$  bürün-temelli öznitelik vektörlerinin SMM ile güdümsüz eğitimi sonunda oluşan  $B$  boyutlu SMM tabanlı öznitelik vektörlerini sırasıyla  $f_{SB}$  ve  $f_{SBM}$  olarak ifade ediyoruz.

### 3. GBM Temelli Duygu Tanıma

Öznitelik olasılık yoğunlukları her duygu için, köşegen kovaryans matrisi kullanan GBM ile modellendi. GBM ile tanımlanan olasılık yoğunluk fonksiyonu

$$p(f) = \sum_{k=1}^K w_k p(f|k) \quad (3)$$

ile ifade edilebilir ve  $K$  bileşenin ağırlıklı birleşimidir. Formüle  $f$  gözlem öznitelik vektörüne,  $w_k$   $k$ 'inci Gauss bileşeninin birleşimdeki ağırlığına karşılık gelmektedir. Ağırlıklar,

$$0 \leq w_k \leq 1 \quad \text{ve} \quad \sum_{k=1}^K w_k = 1 \quad (4)$$

koşullarını sağlamalıdır. Koşullu olasılık  $p(f|k)$  bileşen ortalama vektörü  $\mu_k$  olan ve köşegen kovaryans matrisi  $\Sigma_k$

olan Gauss dağılımı ile modellendi. Her duygunun GBM modeli, duyguyu betimleyen eğitim öznitelik vektörleri kullanılarak ve beklenen değerin en büyütmesi yöntemiyle kestirildi. Duygu tanıma aşamasında ise, verilen konuşmanın özniteliklerinin ardıl olasılıkları tüm GBM yoğunlukları üzerinden en büyütüldü.

### 3.1. Karar Kaynaşımı

Karar kaynaşımı için ağırlıklı toplama metodunu kullandık [5]. GBM sınıflandırıcılarının her duygu ve konuşma için ürettiği tanıma kestirim değerlerini, karar kaynaşımı öncesi  $[0,1]$  aralığına indirgedik [5]. Her  $e$  duygusu için  $\log$  olabirlik değerlerini sırasıyla  $\bar{p}_{\gamma_e}$  ve  $\bar{p}_{\lambda_e}$  ile ifade edelim.

İki sınıflandırıcının birbirinden bağımsız olduğunu varsayarak, ağırlıklı toplama,

$$\rho_e = \alpha \bar{p}_{\gamma_e} + (1 - \alpha) \bar{p}_{\lambda_e} \quad (5)$$

olarak tanımlıyoruz. Bu toplamda  $\alpha$  katsayısını GBM sınıflandırıcısının ağırlığı olarak ve  $[0,1]$  aralığında seçtik.

## 4. Deneysel Sonuçlar

Sistemimizde INTERSPEECH 2009 Konuşmadan Duygu Tanıma yarışmasında dağıtılan FAU Aibo konuşma veri tabanını kullandık [1, 6]. Bu veritabanının eğitim ve test bölümleri, iki farklı okuldaki farklı konuşmacılardan doğal olarak kaydedilerek oluşturulmuş [1, 6]. Kesin bir şekilde ayrılan toplam 9959 eğitim ve 8257 test kayıtlarından 30 ms'lik pencerelerde ortalanmış 20 ms kaydırmalı çerçeveler üzerinden DSF özniteliklerini; 25 ms'lik pencerelerde ortalanmış 10 ms kaydırmalı çerçeveler üzerinden ise diğer öznitelikleri hesapladık. Daha sonra, eğitim bölümü konuşmacılarının üçte ikisinin kayıtları ile sınıflandırıcıları eğitip, kalan üçte bir konuşmacının kayıtları üzerinde test ederek her öznitelik için model parametrelerini belirledik. Son olarak, en yüksek başarımları sağlayan parametreler için tüm eğitim kayıtlarını kullanarak modelleri yeniden eğittik ve test kayıtları sınıflandırmasını gerçekleştirdik.

### 4.1. Sınıflandırıcı Değerlendirmesi

Duyguların sınıflandırılması için Bölüm 3'te bahsedilen GBM modellerini kullandık. GBM bileşen sayısını ve karar kaynaşımı ağırlık katsayısı  $\alpha$  değerini en yüksek tanıma oranını verecek şekilde deneysel olarak hesapladık. Aşağıda, Tablo 2 çeşitli öznitelik vektörleri için GBM tanıma oranlarını özetliyor. İki ve beş sınıflı duygu tanıma problemleri için en yüksek aritmetik ortalama (AO) tanıma oranlarını sırasıyla 45 bileşenli  $f_{BMA}$  ve 50 bileşenli  $f_{MA}$  GBM modelleri ile % 66.39 ve % 39.94 olarak elde ettik [7].

Karar kaynaşımı için en yüksek AO tanıma oranına sahip tekil GBM modellerini kullandık. Karar kaynaşım sonuçlarını Tablo 3'te özetliyoruz. Tanıma oranları bakımından karar kaynaşımı tekil modellere kıyasla önemli bir gelişme sağlıyor.  $f_{BMA}$  ve  $f_{DA}$  karar kaynaşımı ile en yüksek beş sınıflı sınıflandırma sonucunu  $\alpha = 0.57$  değeri için % 40.90 olarak elde ettik. Ayrıca, iki sınıflı problem için  $f_{MA}$  ve  $f_{DA}$  karar kaynaşımı  $\alpha = 0.64$  değeri için % 67.52 başarımları sağladı [7].

Tablo 2: GBM sınıflandırma sonuçları

Öznitelik	Tanıma Oranı (%)			
	2 sınıflı		5 sınıflı	
	AO	AAO	AO	AAO
$\gamma(f_{MA})$	66.36	62.09	39.94	41.29
$\gamma(f_{DA})$	66.05	60.24	39.10	41.78
$\gamma(f_D)$	63.36	65.25	33.68	40.39
$\gamma(f_{BMA})$	66.39	60.70	39.10	46.66

Tablo 3: Karar kaynaşımı duygu tanıma sonuçları

Karar Kaynaşımı	Tanıma Oranı (%)			
	2 sınıflı		5 sınıflı	
	AO	AAO	AO	AAO
$\gamma(f_{MA}) \oplus \gamma(f_D)$	67.49	64.44	40.47	42.07
$\gamma(f_{MA}) \oplus \gamma(f_{DA})$	67.52	62.58	40.76	43.71
$\gamma(f_{BMA}) \oplus \gamma(f_{DA})$	67.44	61.64	40.90	47.83

#### 4.2. Öznitelik Değerlendirmesi

Paralel-dalı SMM yapısını bürün değışikliklerinin zamanla değışimini modellemek amacıyla kullandık. Farklı duygular farklı bürün örüntüleri oluşturduğundan, çok dalı SMM yapısının dallarının bu farklılıkları yakalayabileceğini düşündük.

Hem beş hem de iki sınıflı sınıflandırma problemleri için dal sayısını 5 olarak belirledik. Yaptığımız deneyler sonucunda,  $f_B$  ve  $f_{BM}$  öznitelik vektörlerini her dalı 3 durum, her durumu 12 bileşen içeren paralel-dalı SMM ile modelledik. Daha sonra, bu yapıyı tekil dallarına ayırıp beş kestirim değerini elde ettik ve sigmoid normalleştirme uygularak bu değerleri [0, 1] aralığına eşitledik. Beş boyutlu bu yeni  $f_{SB}$  ve  $f_{SBM}$  öznitelik vektörlerini GBM ile modelledikten sonra Tablo 4' te gösterildiği gibi beş sınıflı problem için sırasıyla % 24.56 ve % 29.53 AO tanıma oranlarını elde ettik. İki sınıflı problem içinse  $f_{SBM}$  GBM ile % 59.82 başarımlı kazandık [7].

Konuşmadan duygu tanıma problemi için, SMM öznitelik vektörleri tek başlarına en iyi başarımlı sağlamasalar da ikinci aşama karar kaynaşımı sırasında fayda sağladılar. Beş ve iki sınıflı problemler için en yüksek AO tanıma oranını, Tablo 4' te belirtildiği gibi, iki-aşamalı kaynaşım ile sırasıyla % 41.59 ve % 67.90 olarak elde ettik.

### 5. Sonuç

INTERSPEECH 2009 Duygu Tanıma Yarışması sonuçlarını ve yöntemlerimizi sunduk. Yarışma katılımcılarının sundukları sistemlerde en çok MFKK, bürün, enerji özniteliklerini, GBM, KDM sınıflandırıcılarını ve karar kaynaşımı yöntemlerini tercih ettiklerini gözlemledik. MFKK öznitelikleri zengin spektral bilgiyi iyi modellediklerinden yüksek tanıma oranı sağladılar. Farklı duygu ipuçları taşıyan öznitelik vektörleri sınıflandırıcıları karar kaynaşımı da tanıma oranını artırdı.

Sistemimizde farklı olarak SMM-tabanlı öznitelikleri ve DSF özniteliklerini kullandık. Farklı öznitelik ve karar kaynaşımını test ettik. Tüm öznitelikler için dinamik öznitelikleri kullanmak fayda sağladı.

Tablo 4: Farklı öznitelik vektörleri için GBM duygu tanıma sonuçları

SMM Özniteliği ve Karar kaynaşımı	Tanıma Oranı (%)	
	5 sınıflı	
	AO	AAO
$\gamma(f_{SB})$	24.56	21.30
$\gamma(f_{SBM})$	29.53	27.48
$\gamma(f_{MA}) \oplus \gamma(f_{SB})$	40.22	41.37
$\gamma(f_{MA}) \oplus \gamma(f_{SBM})$	40.10	41.50
$(\gamma(f_{MA}) \oplus \gamma(f_{DA})) \oplus \gamma(f_{SB})$	40.69	43.33
$(\gamma(f_{MA}) \oplus \gamma(f_{DA})) \oplus \gamma(f_{SBM})$	41.59	44.17
2 sınıflı		
$\gamma(f_{SBM})$	59.82	57.43
$(\gamma(f_{MA}) \oplus \gamma(f_{DA})) \oplus \gamma(f_{SBM})$	67.90	63.03

Yarışma sonucunda en iyi iki ve beş sınıflı AO tanıma sonuçlarımız,  $f_{MA}$  ve  $f_{DA}$  GBM karar kaynaşımının  $f_{SBM}$  GBM ile ikinci aşama karar kaynaşımı için, sırasıyla % 67.90 ve % 41.59 başarımlı oranlarıyla tüm yarışma sonuçları arasında dördüncü ve ikinci oldu. En iyi iki ve beş sınıflı AAO tanıma sonuçlarımız ise sırasıyla,  $f_D$  GBM ile % 65.25 ve  $f_{BMA} - f_{DA}$  GBM karar kaynaşımı sonucu % 47.83 olarak elde ettik [7].

### 6. Teşekkür

Bu çalışma TÜBİTAK-TEYDEB 3070796, TÜBİTAK 106E201 numaralı projeler kapsamında, COST 2102 ve Bahçeşehir Üniversitesi Araştırma Fonu tarafından desteklenmiştir.

### 7. Kaynakça

- [1] B. Schuller, S. Steidl, and A. Batliner, "The interspeech 2009 emotion challenge," in Interspeech (2009), ISCA, Brighton, UK, 2009.
- [2] "Special session: INTERSPEECH 2009 emotion challenge," Interspeech 2009 Brighton, UK, 2009
- [3] J. Deller, J. Hansen, and J. Proakis, Discrete-Time Processing of Speech Signals. New York: Macmillan Publishing Company, 1993.
- [4] F. Itakura, "Line spectrum representation of linear predictive coefficients of speech signals," Journal of the Acoustical Society of America, vol. 57, no. Suppl. 1, p. S35, 1975.
- [5] E. Erzin, Y. Yemez, and A. Tekalp, "Multimodal speaker identification using an adaptive classifier cascade based on modality reliability," IEEE Transactions on Multimedia, vol. 7, no. 5, pp.840–852, Oct. 2005.
- [6] S. Steidl, Automatic Classification of Emotion-Related User States in Spontaneous Children's Speech. Berlin: Logos Verlag, 2009.
- [7] E. Bozkurt, E. Erzin, Ç. E. Erdem and A. T. Erdem, "Improving automatic emotion recognition from speech signals," in Interspeech (2009), ISCA, Brighton, UK, 2009