

BOĞAZIÇI ÜNİVERSİTESİ TÜRKÇE DUYGULU KONUŞMA VERİTABANI ÜZERİNDE DENEYLER İÇİN PROTOKOL VE TABAN ÇİZGİSİ PROTOCOL AND BASELINE FOR EXPERIMENTS ON BOGAZICI UNIVERSITY TURKISH EMOTIONAL SPEECH CORPUS

Heysem Kaya, Albert Ali Salah, Sadık Fikret Gürgen
Bilgisayar Mühendisliği Bölümü
Boğaziçi Üniversitesi
{heysem,salah,gurgen}@boun.edu.tr

Hazım Ekenel
Bilgisayar ve Enformatik Fakültesi
İstanbul Teknik Üniversitesi
ekenel@itu.edu.tr

Özetçe —Bu çalışma, Boğaziçi Üniversitesi Elektrik Elektronik Mühendisliği bölümünde geçmişte toplanmış olan ancak sınıflandırmaya dönük sinyal işleme ve örüntü tanıma çalışmalarında kullanılmamış olan BÜ/EE Duygulu Konuşma Veritabanını ve bu veritabanı üzerinde ileriki deneylerde kullanılacak protokolü tanıtmak amaçlanmıştır. Duygusal veritabanı, 11 amatör aktörün 11 duygusal olarak tanımsız cümleyi Stanislavski etkisi kullanarak 4 ayrı duygu durumu ile canlandırdığı toplam 484 kayıttan oluşmaktadır. Alandaki en iyi sonuçları veren yaklaşıma paralel olarak, ses sinyallerini düşük düzeyli tanımlayıcıları özet fonksiyonlardan geçirerek sabit uzunlukta öznitelik vektörlerine dönüştürülmüştür. Bunun için openSMILE öznitelik sağaltıcı, INTERSPEECH 2013 paralinguistik müsabakası referans öznitelik kümesi ile kullanılmıştır. Veritabanı üzerinde eğitim, geçerleme ve test kümeleri tanımlanmıştır. Destek Vektör Makinaları ve Rassal Ormanlar ile elde edilen referans sonuçlar sunulmuştur.

Anahtar Kelimeler—*Duygulu Konuşma Veritabanı, Hesaplamasal Paralinguistik, İnsan-Makina Etkileşimi, Konuşmadan Duygu Tanıma*

Abstract—This study aims at presenting an emotional corpus collected at Boğaziçi University / Electrical and Electronics Department, on which no previous signal processing and machine learning study done for classification purposes. It also aims at providing the protocol for further experiments on this corpus. The emotional corpus consists of 484 speech utterances from 11 amateur actors acting 11 emotionally undefined sentences with 4 emotions using Stanislavsky effect. In-line with the state-of-the-art method in the field, functionals were passed over the Low Level Descriptors of the signal to obtain fixed length feature vectors. For this purpose, the openSMILE feature extractor was used with the baseline feature set from the INTERSPEECH 2013 Computational Paralinguistics Challenge. The training, validation and testing partitions are defined on the corpus. The baseline results obtained using Support Vector Machines and Random Forests are presented.

Keywords—*Emotional Speech Corpus, Computational Paralinguistics, Human-Computer Interaction, Speech Emotion Recognition*

I. GİRİŞ

İnsan-makina etkileşimi bilim dalı son bir kaç onyılıda özellikle kişi ve komutların tanınması alanında önemli gelişmeler

kaydetmiştir. Kısmen bu gelişmelere istinaden ve kısmen bağımsız olarak etkileşimi daha insansı bir hale getirmek amacıyla, komutların ötesinde kişinin durum, davranış ve özelliklerini konuşma, görü ve metin gibi kiplerden otomatik analiz etmeye yönelik çalışmalar son onyılıda ivme kazanmıştır [1], [2]. Kişinin duygu ve durumunu anlayabilen (hassas) robotlar, öğrencinin psikolojik durumu ve kişisel özelliklerine göre adapte olan akıllı öğrenme sistemleri, kuyruktaki kişileri müşteri memnuniyetini eniyileyecek şekilde otomatik olarak önceliklendiren karşılama sistemleri bu çalışmaların yakın gelecekteki ürünleri olacaktır. Psikoloji, bilgisayar bilimleri, işaret işleme, yapay öğrenme gibi farklı uzmanlık alanlarını bir araya getiren bu çalışma alanı bir özel ilgi alanından Hesaplamasal Paralinguistik (Computational Paralinguistics) adıyla bir bilim dalına evrilmektedir.

Bu alanda 2009 yılından itibaren [3] ses kipinde INTERSPEECH konferansı kapsamında Computational Paralinguistics (ComParE) müsabakaları, 2011 yılından itibaren de ses-görüntü kiplerinde çeşitli konferanslara bağlı çalıştaylar halinde AVEC (Audio-Visual Emotion Corpus and Challenge) müsabakaları düzenlenmektedir [4]–[6]. Bu müsabakalardaki temel amaç, farklı uzmanlık alanlarındaki bilimsel gelişmeler ile çalışmaların mukayese edilebilirliği arasındaki uçurumu kapatmaktır. Müsabakalar duygu tahmininden [3], [4] kişilik özelliği tahminine [7], depresyon düzeyi tahmininden [6], otizm ve kahkaha tespitine [8] kadar geniş bir yelpazede problemler sunmuştur.

Müsabakalar mukayese edilebilir şekilde bölümlenmiş veritabanı yanında her sene nicel ve nitel olarak geliştirilen referans öznitelik kümeleri de sunulmaktadır. Sözelimi 2009 yılında INTERSPEECH bağlamında gerçekleştirilen ilk paralinguistik müsabakasında 384 referans öznitelik varken, 2013 yılındaki müsabakada bu sayı 6373'e çıkarılmıştır. [3], [8]. Müsabakalarda sağlanan referans öznitelikleri genellikle açık kaynak kodlu yazılımlar (örn. openSMILE [9]) ile çıkarılıp yine genel kullanıma açık (örn. WEKA [10]) veri madenciliği/yapay öğrenme araçları ile uyumlu biçimlerde sunulmaktadır.

Technische Universität München (TUM) tarafından geliştirilip bakımı yapılan openSMILE [9] öznitelik çıkarıcı, son yıllardaki müsabakaların zorlu referans özniteliklerini çıkarmakta kullanılmaktadır. Hesaplamasal paralinguistik alanındaki en iyi sonuçlar, ses işaretininin 10-30 milisaniyelik çerçevelerinden

çıkarılan Düşük Düzeyli Tanımlayıcıların-DDT (Low Level Descriptors - LLD) özet fonksiyonları (örn. ortalama, standart sapma gibi istatistikler) ile sabit uzunlukta bir vektöre dönüştürüldüğü işlem süreci ile elde edilmektedir.

En iyi sonuçların elde edildiği bu işlem sürecine paralel olarak, bu makalede openSMILE aracı kullanılarak en güncel INTERSPEECH Paralinguistik Müsabakası öznitelik kümesi (2013) çıkarılmıştır. Nispeten az sayıda örneğe karşın çok yüksek boyutlu (6373) verinin aşırı öğrenmeden sakınarak sınıflanması için bu probleme dayanıklılığı ile bilinen Destek Vektör Makinaları (DVM) ve Rassel Ormanlardan (RO) faydalanılmıştır.

Konuşmadan duygu/durum tanıma alanında Türkçe kaynaklar üretilmekle beraber, bunlar diğer dillerdeki materyallere oranla oldukça kısıtlı kalmaktadır [11]. Bu çalışmanın amaçlarından biri en üst düzeyde mukayese edilebilirlik ile bu kaynakların çeşitlenmesini sağlamaktır. Makalenin kalanı şu şekilde organize edilmiştir. Bir sonraki bölümde veritabanı tanıtılmakta, üçüncü bölümde referans öznitelik kümesi sunulmaktadır. Dördüncü bölümde RO ve DVM ile elde edilen referans sonuçları verilmiştir. Tartışma ve Görünüm bölümünde vargılar müsabaka çağrısı ile sonlanmaktadır.

II. BOĞAZIÇI ÜNİVERSİTESİ DUYGULU KONUŞMA VERİTABANI

A. Sık Kullanılan Konuşma Veritabaları

Hesaplamasal paralinguistik alanında en sık kullanılan duygulu konuşma veritabanlarının özet istatistikleri Tablo I'de verilmiştir. Genel olarak dilde Almanca ve İngilizcenin, kayıt ortamlarında ise stüdyonun ağırlığı göze çarpmaktadır. Konuşma sabit olabileceği gibi, Vera Am Mittag (VAM) veritabanındaki gibi tamamen serbest ve gürültülü de olabilmektedir. Tablodaki tüm kolonu çok sınıflı sınıflandırmanın ağırlıksız ortalama tanıma performansı, son iki kolon da sırasıyla aktivasyon ve değerlik için aynı ölçüyü göstermektedir.

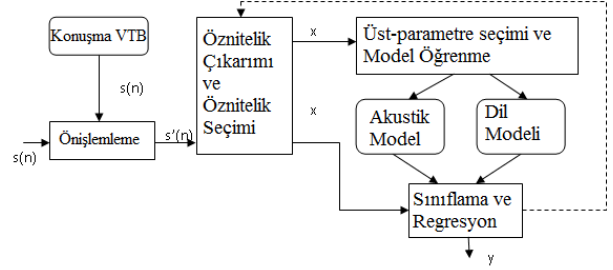
B. Boğaziçi Üniversitesi Duygulu Konuşma Veritabanı

Duygusal veritabanı Meral ve diğerleri tarafından 2003 yılında 11 amatör tiyatrocu tarafından Stanislavski etkisi ile canlandırma yapılarak toplanmıştır [12]. Stanislavski etkisi, aktörün bir duyguyu uyaran durum ve koşulları zihninde canlandırarak yarattığı etkidir [13]. Tümceler aktivasyon/değerlik eksenlerinde dengeli olarak dağılmış dört temel duygu ile canlandırılmıştır: sevinç (+/+), olağan (-/+), kızgınlık (+/-), ve üzüntü (-/-). Veritabanında kullanılan tümceler sözdizimsel açıdan değişik yapılara sahip, olumlu-bildirici tümcelerdir [12]:

- Dışarıda kar yağıyor.
- Sınavdan yetmiş aldım.
- Hoca bana yetmiş verdi.
- Galatasaray maçı iki sıfır kazandı.
- Beni çok şaşırttı.
- Telefonum çalıyor.
- Kurs yarın bitiyor.
- Yarın kar yağacakmış.

Tablo II: Konuşmacı Ayrık Veri Bölümlemesi. K: Kadın, E:Erkek

	Eğitim	Geçerleme	Test
Cinsiyet Dağılımı	2K+2E	2K+1E	3K+1E
Örnek Sayısı	176	132	176
Aktivasyon (poz/neg)	88/88	66/66	88/88
Değerlik (poz/neg)	88/88	66/66	88/88



Şekil 1: Konuşmadan Durum/Özellik Tanıma İşlem Süreci [1]

- Dersler iki hafta ertelendi.
- Kapı açık kalmış.
- Dersi sadece iki kişi geçemedi.

Toplamda 4 duygu x 11 aktör x 11 tümce = 484 örnek toplanmıştır. Aktörlerin 7'si kadın, 4'ü erkektir. Bu bildiride ve ileriki çalışmalarda kullanılmak üzere konuşmacı ayrımı sağlanarak veri eğitim, geçerleme ve test kümelerine bölünmüştür (bkz. Tablo II). Bir sonraki bölümde konuşma işaretinin işlenmesi ve çıkarılan öznitelik kümesi hakkında bilgiler verilmiştir.

III. KONUŞMA İŞARET İŞLEME VE ÖZİNTELİKLER

Konuşmadan durum/özellik tanıma uygulamalarında kullanılan işlem süreci canlı konuşma yahut veritabanından alınan işaretin ön işlemden geçirilmesi ile başlayıp, öznitelik çıkarımı ile devam eder. Konuşmadan çıkarılan potansiyel özniteliklerin arasından ilgili probleme en uygun olanları seçilir ve yapay öğrenme algoritmasına model öğrenimi için verilir. Varsa, bu algoritmanın üst-parametreleri geçerleme kümesi yardımıyla eniyenir. İşlem dizisinin özeti Şekil 1'de gösterilmiştir. Konuşma işaret işlemede amaç konuşma tanıma ise genellikle çerçeve bazında DDT ler (temel sıklık-F0, Mel Frekanslı Kepstral Katsayıları- MFCC, enerji vs) ve bunların türevleri çıkarılıp Saklı Markov Modelleri (Hidden Markov Models- HMM) tabanlı sekans sınıflayıcılar ile modellenir. Eğer amaç bu bildirideki gibi konuşmacı durumu tanımak ise, kısa süreli akustik özniteliklerin sırasal modellenmesi yerine bunların istatistikleri üzerinden elde edilen sabit uzunluklu vektörler kullanılmaktadır. Bu şekilde sabit uzunluklu vektörlerle modellenmenin daha başarılı olduğu tespit edilmiştir [1]. Bu başarının nedeni konuşmacı durum/özelliğinin konuşmaya yayılmış olmasıdır. Yakın tarihlerde gerçekleştirilen müsabakalarda da bu yaklaşımla elde edilen çok yüksek boyutlu öznitelik kümeleri kullanılmaktadır.

Tablo I: Konuşma Duygu Tanıma Alanında En Sık Kullanılan Veritabanları Özet İstatistikleri. Dil A: Almanca, İ: İngilizce, D:Danca; Canl(andırma) E:Evet, H:Hayır; #Top: Toplam kayıt sayısı; Tüm: Tüm duygu sınıfları; Aktiv(asyon), Değer(lik) [1]

Veritabanı	Dil	Konuşma	Canl	#Top.	ss:dd	#E	#K	Kayıt	kHz	Tüm	Aktiv	Değer
ABC	A	sabit	E	430	01:15	4	3	stüdyo	16	56.1	69.3	79.6
AVIC	İ	serbest	H	3002	01:47	11	10	stüdyo	44	59.9	75.6	75.2
DES	D	sabit	E	419	00:28	2	2	stüdyo	20	59.9	90.0	71.7
EMOD	A	sabit	E	494	00:22	5	5	stüdyo	16	84.6	97.6	82.2
eNTER	İ	sabit	H	1277	01:00	34	8	stüdyo	16	72.5	78.1	78.6
SAL	İ	serbest	H	1692	01:41	2	2	stüdyo	16	35.9	65.1	57.7
Smart	A	serbest	H	3823	07:08	32	47	gürültülü	16	25.0	55.2	52.2
SUSAS	İ	sabit	H	3593	01:01	4	3	gürültülü	8	61.4	68.2	74.4
VAM	A	serbest	H	946	00:47	15	32	gürültülü	16	39.3	78.4	52.4

Bu müsabakaların sonucusu olan ve aynı zamanda şu ana kadarki en geniş öznitelik yelpazesini tanıtan (6373 boyutlu) INTERSPEECH 2013 Paralinguistik Müsabakası referans öznitelik kümesi, 64 düşük DDT üzerinden uygun özet fonksiyonlarının geçirilmesi ile elde edilmiştir [8]. DDT ve özet fonksiyonlarının detayları Eyben ve diğerlerinin çalışmasında mevcuttur [14]. Bir sonraki bölümde bahsi geçen deneylerde kullanılan öznitelikler, konuşma işaretinden openSMILE aracında INTERSPEECH 2013 ComParE konfigürasyon dosyası kullanılarak elde edilmiştir.

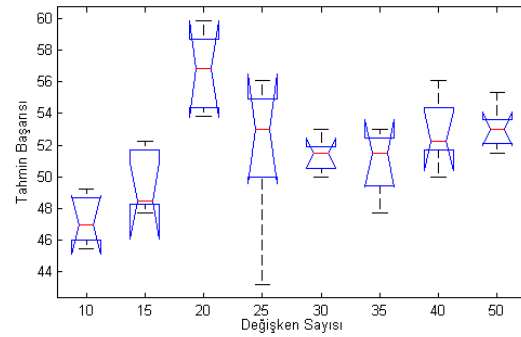
IV. DENEYSEL SONUÇLAR

Deneylerimizde önceki bölümlerde ayrımı verilen veritabanının eğitim kümesinden öğrenip geçерleme kümesine uyguladığımız modellerin performansını ve bu modelleri en iyileyen kurgulardaki üst-parametreleri eğitim+geçerleme kümelerinin birleşimine uygulayarak elde edilen modellerin performanslarını irdelemekteyiz. Konuşma işaretinin sabit uzunlukta vektör ile temsili çok yüksek boyutsallığa neden olduğundan, boyutsallık problemlerine karşı sağlam, iki popüler yöntemi, DVM ile Roları, referans sonuçları elde etmekte kullandık. Her iki algoritmanın WEKA veri madenciliği aracındaki uygulamalarından faydalandık. [10]

A. Rassal Ormanlar ile Deneyler

Breiman tarafından tanıtilan Rolar rassal karar ağaçlarının tümleşimidir [15]. Karar ağaçlarının her biri rassal olarak yerine koyma ile seçilmiş bir örnek alt kümesi ve yine rassal olarak seçilmiş bir değişken alt kümesi ile eğitilerek farklılaşır. Temel sınıflayıcıların farklı ve bağımsız olmaları tümleşirme performansını artırmak için gerekli koşulları sağlamaktadır [16]. RO'da verilen karar yerel ağaçların oylaması ile elde edilmektedir.

Çalışmalarımızda farklı ağaç ve yerel (rassal seçilen) değişken sayısı ile deneyler yaptık. Ağaç sayısı (A) için 10 dan 50 ye kadar 10'ar adımlarla 5 farklı değer, değişken sayısı (d) içinse 10,15,20,25,30,35,40,50 değer kümesini kullandık. Yapılan deneyler sonucunda elde edilen sonuçlar tek yönlü ANOVA ile hem değişken sayısı hem de ağaç sayısı için değerlendirilmiştir. Tek yönlü ANOVA testi ağaç sayısına göre anlamlı bir fark tespit etmemiştir ($p=0.298$). Bununla beraber değişken sayısına göre anlamlı bir fark bulunmuş ($p<0.0005$) ve bunun üzerine post-hoc Tukey-HSD testi yapılmıştır. ANOVA test sonucunun resmedildiği Şekil 2'den görülebileceği üzere, en iyi sonuçlar 20 yerel değişken ile elde edilmiştir. 20 değişken ile elde edilen sonuçlar 10 ve 15 değişkenli modellerin sonuçlarından $\alpha = 0.01$ ile anlamlı bir



Şekil 2: Değişken Sayısına Göre Tek Yönlü ANOVA Çizgesi

farka sahiptir. Benzer şekilde bu sonuçlar, 30 ve 35 değişkenli modellerin sonuçlarından da $\alpha = 0.05$ anlamlılık ile istatistiki olarak daha iyidir. RO larla elde ettiğimiz en iyi geçerleme kümesi sonucu 40 ağaç ve 20 değişken ile erişilen %59.85 tahmin başarısıdır.

B. Destek Vektör Makinaları ile Deneyler

DVMler ile testlerde Liner çekirdek iki ayrı normalizasyon yöntemi denenmiştir: z-normalizasyon (z-norm) ve minimum-maksimum (min-maks) normalizasyonu. Z-norm, her değişkenin ayrı ayrı ortalamasının çıkarılıp standart sapmasına bölünmesiyle, min-maks normalizasyonu ise değişkenlerin minimum değerinin çıkarılıp sonra yeni maksimum değerine bölünmesiyle elde edilir. İlkinde değişkenler sıfır ortalı ve 1 standart sapmalı standart normal dağılıma dönüştürülürken ikincisi tüm değişkenleri 0-1 aralığına getirir. Geçerleme ve test kümesi normalize edilirken eğitim kümesindeki parametreler kullanılır. Liner çekirdek kullanılmakta amaç 1) ComParE müsabakalarındaki taban çizgisi takdim yöntemi ile uyumluluk ve 2) taban çizgisinin belirlenmesinde çok fazla parametrenin eniyilenmesiyle ortaya çıkacak bir aşırı öğrenme durumundan sakınmaktır.

DVM testleri bir dizi DVM karmaşıklık parametresi (C) ile denenmiştir. DVMlerin parametre ve normalizasyona göre geçerleme kümesi performansı Tablo III'de listelenmiştir. Z-norm için en iyi değer $C = 10^{-4}$ ile %60.6 bulunurken daha yüksek C değerlerinde hep aynı sonuç (%57.58) elde edilmiştir.

Min-maks normalizasyon ile en iyi sonuç $C = 0.01$ ile elde edilmiş olup, daha yüksek C değerleri aynı sonucu (%58.33) vermiştir.

Tablo III: Geçerleme Kümesi DVM Sonuçları

C	Z-norm	Min-maks
0.05	57.58	58.33
0.01	57.58	59.09
0.001	57.58	52.27
10^{-4}	60.6	36.36
10^{-5}	46.21	41.67

C. Test Kümesi Taban Çizgisi

Bağımsız test kümesi için ROLarda elde edilen en iyi sonucu veren kurguyu, DVMlerde her iki normalizasyon yöntemi ile elde edilen en iyi iki sonucu veren kurguları denedik. Test kümesinde eniyileme yapmamak için bu küme üzerindeki deney sayısını 5 ile sınırladık. DVMlerde aşırı öğrenme yapmamak için, geçerleme kümesinde en iyi ikinci sonucu veren üst parametreler arasında en küçük olanı seçtik. Test kümesi üzerindeki deneylerin sonuçları Tablo IV'da verilmiştir.

Tablo IV: Test Kümesi sonuçları

Yöntem	Ön İşleme	Üst-parametre	Doğruluk (%)
RO	z-norm	d=20, A=40	52.27
Liner SVM	z-norm	C= 10^{-4}	57.39
	z-norm	C=0.001	64.20
	min-max	C=0.01	62.50
	min-max	C=0.1	62.50

DVMler ile yapılan deneylerde test kümesinde (kısmen öğrenme kümesi daha fazla örnekliliği için) geçerleme kümesine oranla daha iyi sonuçlar gözlemlenmiştir. Bununla beraber test kümesi taban çizgisi olarak sunduğumuz %64.20 başarısı geçerleme kümesindeki en iyi ikinci C değeri ile sağlanmıştır. ROLarda geçerleme kümesinde en iyi sonucu veren kurgu test kümesinde %52.27 tahmin başarısı ile DVMlerin gerisinde kalmıştır.

V. TARTIŞMA VE GÖRÜNÜM

Son yıllarda gelişerek bir özel ilgi alanından, bir bilim dalına dönüşen Hesaplamasal Paralinguistik alanında Türkçe kaynaklar oldukça kısıtlıdır. Bu alandaki çalışmalara katkıda bulunmak ve mukayese edilebilirliği artırmak amacıyla daha önce yapay öğrenme alanında kullanılmayan Boğaziçi Üniversitesi Türkçe Duygulu Konuşma Veritabanı üzerinde ileriki çalışmalarda kullanılacak protokol ve taban çizgisi teşkil edecek deneyler tanıtılmıştır. Bu amaçla alanın en iyi sonuçlarını veren sabit öznitelik vektörü temsilî yöntemi INTERSPEECH 2013 ComParE müsabakası referans öznitelik kümesi üzerinde kullanılarak referans öznitelikleri elde edildi. Boyutsallık problemi dikkate alınarak çok yüksek boyutlu veride DVM ve ROLar kullanıldı. Eğitim, geçerleme ve bağımsız test kümesine bölünen veritabanının ilk iki bölümünde parametreleri eniyiledikten sonra az sayıda en iyi kurgu test kümesinde denendi. Bağımsız test kümesinde en iyi sonuç %64.20 ile Liner çekirdekli DVM ile elde edilmiştir. Geçerleme kümesinde en iyi sonucu veren RO kurgusunun test kümesi performansı DVMlerin altında kalmıştır. İleriye dönük olarak bu veri üzerinde bu bildiride tanımlanan protokol ile bir müsabaka gerçekleştirilmesi düşünülmektedir.

VI. TEŞEKKÜR

H. Kaya YÖK tarafından ÖYP kapsamında desteklenmektedir. A. A. Salah'ın çalışması Boğaziçi Üniversitesi BAP 6531 projesi kapsamında destelenmiştir.

KAYNAKÇA

- [1] S. Björn, *In Salah, A. A. and Gevers, T. (eds) Computer Analysis of human Behavior*. Springer, 2011, ch. Voice and Speech Analysis in Search of States and Traits, pp. 227–253.
- [2] G. Mohammadi and A. Vinciarelli, "Automatic personality perception: Prediction of trait attribution based on prosodic features," *IEEE Transactions on Affective Computing*, vol. 3, no. 3, pp. 273–284, 2012.
- [3] B. Schuller, S. Steidl, and A. Batliner, "The Interspeech 2009 Emotion Challenge," in *Proceedings INTERSPEECH 2009, 10th Annual Conference of the International Speech Communication Association*, ISCA, Brighton, UK: ISCA, September 2009, pp. 312–315.
- [4] B. Schuller, M. Valstar, R. Cowie, and M. Pantic, Eds., *Proceedings of the First International Audio/Visual Emotion Challenge and Workshop, AVEC 2011*, ser. Lecture Notes on Computer Science (LNCS), vol. 6975, Part II, HUMAINE Association. Memphis, TN: Springer, October 2011, held in conjunction with the International HUMAINE Association Conference on Affective Computing and Intelligent Interaction 2011, ACII 2011.
- [5] B. Schuller, M. Valstar, F. Eyben, R. Cowie, and M. Pantic, "AVEC 2012 – The Continuous Audio/Visual Emotion Challenge," in *Proceedings of the 14th ACM International Conference on Multimodal Interaction, ICMI*, L.-P. Morency, D. Bohus, H. K. Aghajan, J. Cassell, A. Nijholt, and J. Epps, Eds. Santa Monica, CA: ACM, October 2012, pp. 449–456.
- [6] M. Valstar, B. Schuller, K. Smith, F. Eyben, B. Jiang, S. Bilakhia, S. Schlieder, R. Cowie, and M. Pantic, "AVEC 2013–The Continuous Audio/Visual Emotion and Depression Recognition Challenge," in *Proceedings of the 3rd ACM international workshop on Audio/visual emotion challenge*, ser. AVEC '13, 2013, pp. 3–10.
- [7] B. Schuller, S. Steidl, A. Batliner, E. Nöth, A. Vinciarelli, F. Burkhardt, R. van Son, F. Weninger, F. Eyben, T. Bocklet, G. Mohammadi, and B. Weiss, "The INTERSPEECH 2012 Speaker Trait Challenge," in *Proceedings INTERSPEECH 2012*. Portland, OR, USA: ISCA, September 2012.
- [8] B. Schuller, S. Steidl, A. Batliner, A. Vinciarelli, K. Scherer, F. Ringeval, M. Chetouani, F. Weninger, F. Eyben, E. Marchi, M. Morfillaro, H. Salamin, A. Polychroniou, F. Valente, and S. Kim, "The INTERSPEECH 2013 Computational Paralinguistics Challenge: Social Signals, Conflict, Emotion, Autism," in *Proceedings INTERSPEECH 2013*. Lyon, France: ISCA, August 2013, pp. 148–152.
- [9] F. Eyben, M. Wöllmer, and B. Schuller, "Opensmile: the munich versatile and fast open-source audio feature extractor," in *Proceedings of the international conference on Multimedia*. ACM, 2010, pp. 1459–1462.
- [10] M. Hall, E. Frank, G. Holmes, B. Pfahringer, P. Reutemann, and I. H. Witten, "The weka data mining software: an update," *SIGKDD Explor. Newsl.*, vol. 11, no. 1, pp. 10–18, Nov. 2009.
- [11] C. Oflazoglu and S. Yildirim, "Turkish emotional speech database," in *Signal Processing and Communications Applications (SIU), 2011 IEEE 19th Conference on*, April 2011, pp. 1153–1156.
- [12] E. H. K. Meral, H. M. and A. Ozsoy, "Analysis of Emotion in Turkish," in *XVII. National Conference on Turkish Linguistics*, May 2003.
- [13] K. Stanislavsky, *Bir Aktör Hazırlanıyor*. Pegasus, 2013.
- [14] F. Eyben, F. Weninger, and B. Schuller, "Affect recognition in real-life acoustic conditions - a new perspective on feature selection," in *Proceedings INTERSPEECH 2013, 14th Annual Conference of the International Speech Communication Association*. ISCA, ISCA, 2013, pp. 2044–2048.
- [15] L. Breiman, "Random forests," *Machine learning*, vol. 45, no. 1, pp. 5–32, 2001.
- [16] E. Alpaydin, *Introduction to Machine Learning*, 2nd ed. The MIT Press, 2010.