



Konuşma Tabanlı Duygu Tanımda Ön İşleme ve Öznitelik Seçim Yöntemlerinin Etkisi

Turgut ÖZSEVEN*

Tokat Gazosmanpaşa Üniversitesi, Bilgisayar Mühendisliği Bölümü, Tokat
turgut.ozseven@gop.edu.tr ORCID: 0000-0002-6325-461X, Tel: (356) 252 16 16 (2938)

Geliş: 18.12.2018, Kabul Tarihi: 01.12.2019

Öz

Konuşma tabanlı duygu tanımda sayısal sinyal işleme ile sayısal hale getirilen ses sinyalinden elde edilen öznitelikler kullanılmaktadır. Konuşmadan çıkartılan özniteliklerin tamamı tek boyut altında ele alınabildiği gibi boyutsal veya yapı bakımından gruplandırılarak da ele alınmaktadır. Bu çalışmada, öznitelik seçim ve ön işleme yöntemlerinin duygu tanımdaki etkisi araştırılmıştır. Bu amaçla, EMO-DB veri seti ve üç farklı sınıflandırıcı kullanılmıştır. Elde edilen sonuçlara göre, en yüksek başarı çok katmanlı algılayıcı ve yüksek geçiren filtre ile %90.3 olarak elde edilmiştir. Spektral öznitelikler prosodik özniteliklerden daha yüksek başarı sağlamıştır. Ayrıca, bayanlar erkeklere göre ve 20-29 yaş aralığındaki bireyler 30-35 yaş aralığındaki bireylere göre duygularını seslerine daha fazla yansıtmaktadır. Çalışmada ele alınan ön işleme yöntemlerinden yüksek geçiren filtreler sınıflandırıcı başarısını artırırken alçak geçiren, bant geçiren filtreler ve gürültü giderme başarı oranını düşürmüştür.

Anahtar Kelimeler: Konuşmadan duygu tanıma; ön işleme; öznitelik seçimi; filtreleme;duygu sınıflandırma;

* Yazışmaların yapılacağı yazar

Giriş

Duygu tanıma psikoloji, psikiyatri ve bilişsel bilim gibi alanlarda her geçen yıl artarak ihtiyaç duyulan bir çalışma alanıdır. Duygu tanıma için kullanılan iki temel yöntem konuşma ve yüz ifadelerinin analizidir (Bayrakdar, Akgün, ve Yücedağ 2017). Konuşma işleme bir sinyal işleme süreci olup ses kalitesinin artırılması, kişi tanıma, konuşma tanıma, duygu tanıma ve iletişim gibi birçok alanda kullanılmaktadır. Konuşma işleme yöntemleri öncesi konuşma üzerinden duygu tanıma (KDT) için algısal analiz kullanılmaktadır. Algısal analiz konuşmanın katılımcılara dinletilmesi ve onların yorumları ile konuşmacının içerisinde bulunduğu duygunun tanınmasıdır. Ancak, algısal analiz kişiden kişiye değişiklik göstermesi nedeni ile objektif bir yöntem değildir. Konuşmanın objektif olarak değerlendirilmesi için akustik analiz kullanılmaktadır. Akustik analiz konuşmanın içerdiği sesin üretilmesi ve sinyalin yapısı ile ilgili özniteliklerin sayısal sinyal işleme ile tespit edilmesidir. Hangi duyguların hangi akustik öznitelikler üzerinde etkili olduğunun tespiti yaş ve cinsiyet dahil olmak üzere sesteki bireysel farklılıklardan dolayı zorluk oluşturmaktadır (Zupan vd. 2009). KDT’de bir diğer zorluk ise duyguların konuşmaya tam olarak yansıtıldığı verinin elde edilmesidir. Bu nedenle çalışmalarda geçerliliği kabul görmüş EMO-DB (Burkhardt vd. 2005) ve SUSAS (Hansen vd. 1997) gibi hazır veri setleri kullanılabilirdiği gibi araştırmacı tarafından toplanan geçerliliği test edilmiş veriler de kullanılabilir. KDT’de konuşma sinyalinin kalitesini artırma, konuşma dışındaki seslerin bastırılması, vurgunun ön plana çıkartılması veya sinyal normalizasyonu gibi süreçler ön işleme adımıyla gerçekleştirilmektedir. Bu adımda filtreleme yöntemleri, örnekleme değerini düşürme, ön vurgulama, gürültü azaltma ve DC-ofset kaldırma en çok kullanılan yöntemlerdir. Ayrıca, sinyal küçük zaman aralıklarında kararlı kaldığı için ön işleme adımıyla çerçeveleme ve

pencereleme kullanılmaktadır. Çerçeveleme sinyalden küçük kesitler alınmasıdır. Pencereleme sinyalden alınan çerçevelerdeki spektral sızmayı, başta ve sonda yer alan süreksizlikleri kaldırmaktadır (Tarnag vd. 2010). Rabiner ve Schafer (Rabiner ve Schafer 1978)’a göre çerçeve boyutu 10-20ms arasında seçilmesi uygundur.

Konuşma sinyallerinden öznitelik çıkartımı için Praat (Boersma ve Weenink 2010) ve OpenSMILE (Eyben, Wöllmer, ve Schuller 2010) gibi hazır araçlar kullanılabilirdiği gibi araştırmacılar tarafından geliştirilen kodlar da kullanılmaktadır. Sinyalden çıkartılan öznitelikler dört kategoride gruplandırılabilir (Huang, Xue, ve Mao 2015). Bunlar, akustik öznitelikler, dilsel öznitelikler, bağlamsal bilgi ve farklı öznitelik kümelerinin birleşiminden elde edilen hibrit öznitelikler. Çalışmalarda çoğunlukla akustik öznitelikler kullanılmış olup bunlarda kendi içerisinde prosodik ve spektral olmak üzere iki gruba ayrılır. Çoğunlukla kullanılan akustik öznitelikler perde, formant frekansları, jitter, shimmer, enerji, sıfır geçiş oranı, konuşma oranı, ses kalitesi, MFCC, LPCC, dalgacık ve TEO tabanlı özniteliklerdir.

Bu çalışmada, ön işleme yöntemleri, öznitelik kümesi ve öznitelik seçim yöntemlerinin KDT üzerinde etkisi üç farklı sınıflandırıcı ile karşılaştırılmıştır.

Çalışmanın bir sonraki bölümünde mevcut çalışmaların özeti, 3. bölümde çalışma sürecince kullanılan veri seti ve kullanılan yöntemler detaylandırılmıştır. 4.bölüm çalışma sonucu elde edilen sonuçlarını içermektedir. Çalışmanın son bölümünde de elde edilen sonuçlar yorumlanmıştır.

Literatür Taraması

KDT ile ilgili literatürde yapılmış çok sayıda çalışma bulunmaktadır. Bu çalışmalar genel hatlarıyla, duygu sınıflandırma, öznitelik tespiti, hibrit sınıflandırıcılar ve duygular üzerinde etkili

olan özneliklere odaklanmıştır.

Bu çalışma kapsamında yapılan çalışmalar incelendiğinde; KDT ön işleme adımında çoğunlukla ön vurgulama (Tarnng vd. 2010; Chen, Mao, Wei, vd. 2012; Chen, Mao, Xue, vd. 2012; Joshi ve Zalte 2013), DC-ofset kaldırma (Kandali, Routray, ve Basu 2009; Ntalampiras ve Fakotakis 2012), örnekleme değerini düşürme (Grimm vd. 2007; Sezgin, Günsel, ve Kurt 2012; Orlandi vd. 2013) ve filtreleme (alçak geçiren, yüksek geçiren ve bant geçiren filtreler) (Orlandi vd. 2013; Tao, Kang, ve Li 2006; MacCallum vd. 2011; Ludeña-Choez ve Gallardo-Antolín 2015) kullanılmıştır. Bu çalışmalarda bu yöntemler, ön işleme sürecinin bir gerekliliği olarak kabul edilmiş ve sonuçlar üzerindeki etkisi incelenmemiştir. Sayısal filtrelerin akustik parametreler, cinsiyet yaş ve duygu durumu üzerindeki etkilerinin araştırıldığı çalışmada ise sadece bu yöntemlerin parametreler ve demografik özellikler üzerindeki etkisi incelenmiş olup duygu tanıma başarısı üzerindeki etkileri ele alınmamıştır (Özseven ve Dügenci 2017).

Özneliklerdeki birim farklılıklarını ortadan kaldırmak için normalizasyon ve öznelik kümesinin boyutunu azaltmak için öznelik seçim yöntemleri kullanılmaktadır. En çok kullanılan normalizasyon yöntemleri z-puanı (Scherer vd. 2015; Goudbeek ve Scherer 2010; Patel vd. 2011; Laukka vd. 2011; Polzehl vd. 2011; Lee vd. 2011; Sundberg vd. 2011; Truong, van Leeuwen, ve de Jong 2012; Mariooryad ve Busso 2013; Bänziger, Patel, ve Scherer 2014) ve min-maks (Clavel vd. 2008), en çok kullanılan özellik seçim yöntemleri ise temel bileşenler analizi (TBA), doğrusal ayırmacılık analizi (DAA) ve ileri doğru seçim (IDS)'dir (Özseven, Dügenci, ve Durmuşoğlu 2018). Öznelik seçimi konuşma sinyalinden elde edilen öznelik kümesinin boyutunu azaltmak ve duyguları en iyi ifade eden özneliklerin tespiti için kullanılır. Ancak, öznelik seçim yöntemlerinin kullanımı her zaman sınıflandırıcı başarısını artırmamaktadır. Grimm vd., (2007) TBA ve

tüm özneliklerin kullanımının KDT üzerindeki etkisi incelemiş ve tüm özneliklerin kullanımında daha yüksek başarı elde etmiştir (Grimm vd. 2007). Hoque vd. (2006) yaptıkları çalışmada TBA ve DAA kullanarak öznelikleri düşük boyutlu uzayda yansıtarak ve bu yöntemlerin birlikte kullanımının ayrı ayrı kullanıma göre daha iyi sonuçlar verdiğini tespit etmiştir (Hoque, Yeasin, ve Louwerse 2006). Fisher ve TBA öznelik seçim yöntemleri ve iki sınıflandırıcı ile gerçekleştirilen çalışmada öznelik seçiminde Fisher TBA'dan daha iyi sonuçlar üretmiştir (Chen, Mao, Xue, vd. 2012).

Gerek KDT gerekse de diğer çalışmalar için literatürde kullanılabilir çok sayıda sınıflandırıcı yer almaktadır. Ancak, bir öznelik verisi üzerinde başarılı sonuçlar elde eden sınıflandırıcı her öznelik kümesi üzerinde bu başarıyı elde edemeyebilir. Bu nedenle, sınıflandırma başarısı kullanılan sınıflandırıcı kadar kullanılan öznelik kümesi ve öznelik kümesi üzerinde yapılan işlemlerle de ilgilidir. KDT içeren çalışmalar incelendiğinde çoğunlukla gauss karışım modeli (GKM), destek vektör makineleri (DVM), saklı markov modeli (SMM), çok katmanlı algılayıcı (ÇKA) ve k-en yakın komşu (k-EK) gibi tekil sınıflandırıcılar veya hybrid sınıflandırıcılar kullanılmıştır (Özseven, Dügenci, ve Durmuşoğlu 2018).

DVM tabanlı ardışıl ileri yönlü öznelik seçim yönteminin sunulduğu ve çeşitli sınıflandırıcının kullanıldığı çalışmada 276 olan öznelik kümesi 75'e düşürülmüş ve sınıflandırma başarısı DVM sınıflandırıcı için duygu tanıma oranı %84.84'den %87.5'e yükselmiştir. Diğer sınıflandırıcılar üzerindeki etkisinin de incelendiği çalışmada tüm sınıflandırıcılarda başarı oranında artış sağlanmıştır (Schuller vd. 2005). Ardışıl ileri yönlü öznelik seçim yöntemi kullanılan bir diğer çalışmada 3713 olan öznelik kümesi 150'ye düşürülmüş ve tüm öznelik kümeleri üzerinde %63 başarı elde edilirken öznelik seçimi sonrası başarı oranında azalma olmuştur (Batliner vd. 2011). Aynı öznelik seçim yönteminin kullanıldığı bir diğer

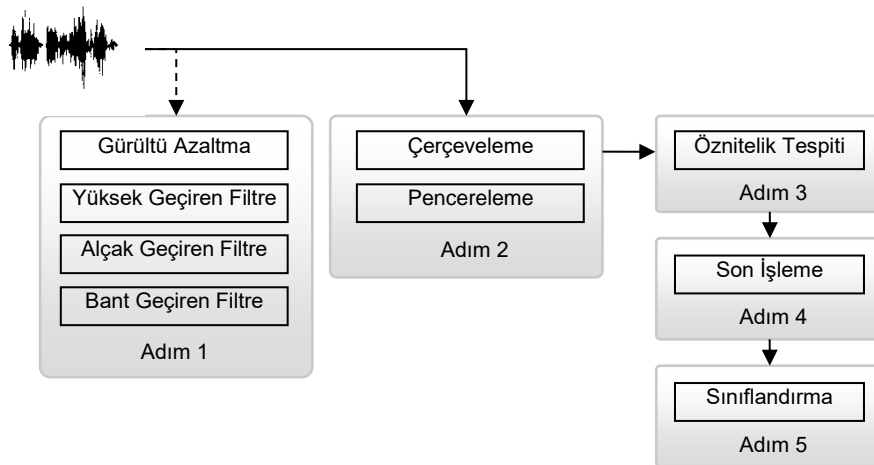
çalışmada ise 58 öznitelik kullanılmıştır. Tüm öznitelikler DVM ile sınıflandırıldığında %79.9, öznitelik seçimi sonrası 15 öznitelik kullanıldığında ise %76.9 başarı elde edilmiştir. Yazarlar, ardışıl ileri yönlü öznitelik seçim yöntemi üzerinde değişiklik yaparak başarı oranını %82.8 olarak elde etmiştir (Altun ve Polat 2009). DVM sınıflandırıcı ve mRMR öznitelik seçim yöntemi kullanılarak yapılan çalışmada 380 olan öznitelik sayısı 121'e düşürülürken başarı oranı da %78.3'den %76.8'e düşmüştür (Luengo, Navas, ve Hernaez 2010). Korelasyon tabanlı öznitelik seçimi kullanılan çalışmada 204 olan öznitelik sayısı 90'a düşürülmüştür (Zhao, Zhang, ve Lei 2014). Yazarlar, DVM sınıflandırıcı kullanıldığında %80.4 başarı elde ederken Enhanced-SRC kullandıklarında %83.2 başarı elde etmiştir.

KDT çalışmalarının mevcut durumları incelendiğinde, çok farklı oranlarda duygu tanıma başarıları elde edildiği görülmektedir. Bu durum, kullanılan veri seti, sınıflandırıcı, ön işleme yöntemleri veya öznitelik seçim yöntemine bağlı olarak oluşmaktadır. Bir diğer önemli husus ise aynı veri seti üzerinde aynı yöntemle farklı başarı oranlarının elde edilmesidir. Bunun nedeni de, kullanılan veri setinin içerdiği verilerin dağılımı dengesiz olduğunda ya sınıf sayısı azaltılmakta ya da dağılımı dengelemek için yeni veriler

eklenmektedir. Araştırmacıların KDT alanındaki eğilimleri incelendiğinde ise çalışmaların duygu tanıma başarısını artırmak için yeni öznitelik kümelerinin tespitine, hiyerarşik sınıflandırıcıların kullanımına ve ses sinyallerinin görüntüye dönüştürülerek görüntü işleme yöntemlerinin kullanımına yöneldiği görülmektedir. Ayrıca, mevcut sınıflandırma veya öznitelik seçim yöntemlerini temel alan yeni yöntemler geliştirilmektedir. İncelenen çalışmalardan da görüleceği üzere aynı koşullar altında farklı öznitelik seçim yöntemlerinin duygu tanımda ki başarısı detaylı olarak incelenmemiştir. İncelenen çalışmalar ve araştırmacıların eğilimleri göz önüne alındığında ön işleme ve öznitelik seçim yöntemlerinin duygu tanıma üzerindeki etkisinin araştırılması hem literatüre katkı sağlayacak hem de araştırmacılara yol gösterici olacaktır.

Materyal ve Yöntem

KDT süreci kullanılan veri, elde edilmek istenen bulgular ve kullanılacak yöntemlere farklı süreçler içermektedir. Genel olarak ele alınacak olursa Şekil 1'de gösterildiği gibi 5 adımdan oluşmaktadır. Bu adımlardan Adım 2, Adım 3 ve Adım 5 zorunlu olup Adım 1 ve Adım 4 sınıflandırma doğruluğunu veya sınıflandırıcı performansını artırmak için kullanılmaktadır.



Şekil 1. KDT süreci akış diyagramı

Adım 1’de konuşma sinyali sayısal sinyal işleme sürecinden geçirilerek konuşma sinyalinin tamamı üzerinde işlemler gerçekleştirilmektedir. Gürültü azaltma, sinyalin içerdiği gürültülerin bastırılarak sinyal gürültü oranı değerinin artırılmasını sağlar. Yüksek, alçak ve bant geçiren filtreler belirli frekans aralığındaki sinyallerin işleme alınması veya bastırılması için kullanılır. Adım 2 ön işleme sürecinin bir parçası olup sinyalin çerçevelere bölünmesi ve pencereleme uygulanması için kullanılır. Adım 3 her konuşma sinyali üzerinden öznitelik çıkartımını içermektedir. Adım 4 çıkartılan özelliklere normalizasyon veya öznitelik seçimi gibi yöntemlerin uygulanmasını içermektedir. Adım 5 ise elde edilen öznitelik kümesi kullanılarak duyguların sınıflandırmasını gerçekleştirmektedir. Bu akış diyagramı doğrultusunda çalışma kapsamında

gerçekleştirilen işlemler aşağıdaki alt başlıklarda verilmiştir.

Duygu Durumlarına Göre Kategorize Edilmiş Veri Seti

Bu çalışmada Berlin Database of Emotional Speech (Emo-DB) veritabanı kullanılmıştır. Emo-DB Technical University of Berlin tarafından geliştirilen proje kapsamında aktörler tarafından konuşulan yedi farklı duygusal ifadeden oluşan ses kayıtları içermektedir. Veri seti 20-35 yaş aralığında 10 aktör tarafından 10 farklı cümle duygusal olarak ifade edilerek oluşturulmuştur. Ses kayıtları 16 kHz örnekleme frekansına sahip olup 16 bit mono’dur (Burkhardt vd. 2005). Çalışmada 535 adet ses kaydı incelenmiş olup kullanılan verilere ait özelliklerin dağılımı Tablo 1’de verilmiştir.

Tablo 1. Çalışmada kullanılan ses kayıtlarına ait özellikler

Duygu	Erkek	Bayan	Yaş1 (30-35)	Yaş2 (20-29)	Toplam
Öfke	60	67	90	37	127
Anksiyete/Korku	36	33	51	18	69
Can Sıkıntısı	35	46	60	21	81
İğrenme	11	35	31	15	46
Mutluluk	27	44	53	18	71
Üzüntü	25	37	47	15	62
Nötr	39	40	50	29	79
Toplam	233	302	381	154	535

Konuşma Kayıtlarının Ön İşlenmesi

Çalışma kapsamında, ön işleme adımında gürültü azaltma, yüksek, alçak ve bant geçiren filtreleme kullanılmıştır. MacCallum vd. (2011) yaptıkları çalışmada alçak geçiren filtrenin en uygun kesim frekansının 300Hz ve üstü olduğu sonucuna ulaşmıştır. Perde değeri erişkin erkekler ve kadınlarda sırası ile ortalama 100-150 Hz ve 150-250 Hz arasındadır (MacCallum vd. 2011). Filtre türü olarak Praat içerisinde yer alan Hann filtreleme kullanılmıştır. Bu bilgiler doğrultusunda çalışmada kullanılan filtrelere ait özellikler Tablo 2’de verilmiştir.

Tablo 2. Kullanılan filtrelere ait özellikler

Filtre Türü	f_1	f_2	w
Alçak Geçiren	0 Hz	300 Hz	20 Hz
Yüksek Geçiren	100 Hz	0 Hz	20 Hz
Bant Geçiren	100 Hz	300 Hz	20 Hz

Tabloda verilen f_1 geçiş bandının alt kenarını, f_2 geçiş bandının üst kenarını ve w geçiş ve durdurma arasındaki bölgenin genişliğidir.

Konuşma sinyalindeki arka plan gürültüsü konuşma sıkıştırma, tanıma ve doğrulama gibi uygulamalarda performansı düşürülebilir. Gürültü azaltma sayısal veya analog konuşmadaki dalga biçiminden gürültünün ayrıştırılmasını sağlar (Boll 1979). Çalışmada gürültü azaltma için Praat üzerinde yer alan spektral çıkarım yöntemi kullanılmıştır. Spektral çıkarım konuşmanın spektrumundan tahmin edilen gürültü spektrumunun çıkartılması ile sinyalin iyileştirilmesi işlemidir.

Öznitelik Kümesinin Elde Edilmesi

Konuşma sinyali durağan olmayan bir sinyaldir ama kısa zaman aralıklarında durağan olduğu kabul edilir. Kısa zaman aralıklarını elde etmek için de sinyal çerçevelere bölünmektedir. Çalışma kapsamında, çerçeve boyutu 20ms, hamming pencereleme ve %50.0 örtüşme oranı kullanılmıştır. Ön işleme sonrası akustik analiz

gerçekleştirilerek öznitelik kümesi elde edilmiştir. Elde edilen öznitelikler prosodik ve spektral olmak üzere 2 gruba ayrılmıştır. Her bir özelliğin istatistiksel varyasyonları da öznitelik kümesine dahil edilerek kullanılan tüm öznitelikler Tablo 3’de verilmiştir.

Tablo 3. Çalışmada kullanılan öznitelik kümesi

	Akustik Öznitelik	Değer	Ort.	Medyan	Maks	Min.	Std. Sap.	Aralık
PF (Prosodik Öznitelikler)	F0		√	√	√	√	√	√
	F1		√	√	√	√	√	√
	F2		√	√	√	√	√	√
	F3		√	√	√	√	√	√
	F1 bant genişliği			√				
	F2 bant genişliği			√				
	F3 bant genişliği			√				
	HNR		√		√	√	√	√
	Sessiz çerçeve sayısı	√						
	Duraksama derecesi	√						
	Enerji		√	√	√	√	√	√
	ZCR	√						
	Jitter local	√						
	Jitter rap	√						
	Shimmer local	√						
Shimmer apq3	√							
SF (Spektral Öznitelikler)	MFCC1...13		√	√			√	
	ΔMFCC1...13		√				√	
	LPCC1...13		√	√			√	

Çalışmada Tablo 3’de verilen toplam 149 öznitelik kullanılmış olup bu özniteliklerin 45’i PF ve 104’ü SF ‘dir. ΔMFCC öznitelikleri openSMILE (Eyben, Wöllmer, ve Schuller 2010) ile diğer öznitelikler Praat (Boersma ve Weenink 2010) ile elde edilmiştir.

Normalizasyon, öznitelik kümesinin içerdiği değerlerin dağılımını belirli bir aralıkta sabit tutarak verideki anormal dağılımı düzenlemekte ve ölçü birimi farklılıklarını ortadan kaldırmaktadır. Çalışmamızda z-puan normalizasyon kullanılmıştır (El Ayadi, Kamel, ve Karray 2011).

Sınıflandırıcı

KDT sistemlerinde kullanılan çok sayıda sınıflandırıcı mevcuttur. Bu sınıflandırıcılar bağımsız kullanılabilir gibi birden fazla sınıflandırıcının birleşimini içeren hibrit

sınıflandırıcı olarak da karşımıza çıkmaktadır (Özseven, Düğenci, ve Durmuşoğlu 2018). Bu çalışmada, DVM, ÇKA ve k-EK sınıflandırıcılar ve sınıflandırma için de WEKA (Hall vd. 2009) paket programı kullanılmıştır. DVM istatistiksel öğrenme teorisine dayalı bir yöntemdir. Temel amaç sınıfları birbirinden en iyi şekilde ayıran karar fonksiyonun başka bir ifadeyle hiperdüzlemin tanımlanması esasına dayanır. ÇKA, sınıflandırma amacı ile geri yayımlı eğitilmiş ileri beslemeli bir ağdır (Khanchandani ve Hussain 2009). Çalışma kapsamında DVM sınıflandırıcıda öznitelik sayısı yüksek olduğu için doğrusal çekirdek tercih edilmiştir. ÇKA için giriş katmanında öznitelik sayısı kadar nöron, gizli katmanda (öznitelik+sınıf sayısı)/2 adet katman ve çıkış katmanında 7 nöron kullanılmıştır.

Bulgular

Çalışmada kullanılan EMO-DB 'de yer alan her bir ses kaydı için 45'i PF ve 104'ü SF olmak üzere toplamda 149 öznitelik kümesi oluşturulmuştur. 535 veri ve 149 öznitelik 10-katlamalı çapraz doğrulama ile eğitim ve test kümesine ayrılmıştır. Sınıflandırma işlemi öznitelik kümeleri ve ön işleme yöntemlerinin her birisi için ayrı ayrı gerçekleştirilmiştir. Elde edilen sonuçlar normalizasyon, öznitelik kümesi,

ön işleme yöntemleri, yaş/cinsiyet ve duygu bazlı olarak verilmiştir.

Öznitelik Kümesinin Sınıflandırma Başarısı Üzerindeki Etkileri

PF ve SF içeren 535x149 (PF+SF), PF içeren 535x45 ve SF içeren 535x104 boyutlarındaki matrisler analizlerde kullanılmıştır. Elde edilen öznitelik kümesi ile gerçekleştirilen sınıflandırma başarısı Tablo 4'de verilmiştir.

Tablo 4. Tüm verinin sınıflandırma başarısı

Öznitelik Kümesi	DVM		ÇKA		k-EK	
	GS (sn)	TO (%)	GS (sn)	TO (%)	GS (sn)	TO (%)
PF+SF	327.77	66.0	60.63	89.2	0.01	70.1
PF	421.46	55.9	7.26	75.3	0.0	56.5
SF	0.92	79.1	35.07	84.1	0.01	70.7

GS: Gerçekleşme süresi, TO: Duygu tanıma başarısı

Tablo 4'e göre SF'nin başarıya etkisi PF'den daha yüksektir. Ayrıca verideki ölçü birimi farklılıkları DVM sınıflandırıcının performansında ciddi düşmelere neden olmuştur. Mevcut durumda en yüksek başarı ÇKA sınıflandırıcıda PF+SF birleşiminde elde

edilmiştir. DVM sınıflandırıcı için gerçekleşme süresi PF kümesinde çok yüksek olup SF kümesinde çok düşüktür ve SF öznitelik kümesindeki başarı daha yüksektir. Öznitelik kümesine z-puan normalizasyon uygulandıktan sonra elde edilen sonuçlar Tablo 5'de verilmiştir.

Tablo 5. Normalize edilmiş öznitelik kümesinin sınıflandırma başarısı

Öznitelik Kümesi	DVM		ÇKA		k-EK	
	GS (sn)	TO (%)	GS (sn)	TO (%)	GS (sn)	TO (%)
PF+SF	0.78	86.7	62.73	89.2	0.0	70.1
PF	0.5	71.9	9.84	75.3	0.0	56.5
SF	1.02	83.0	40.29	84.1	1.0	70.7

GS: Gerçekleşme süresi, TO: Duygu tanıma başarısı

Tablo 5'e göre normalizasyon ÇKA ve k-EK sınıflandırıcıda hiçbir değişikliğe neden olmazken DVM sınıflandırıcıda hem gerçekleşme süresi hem de doğruluk anlamında ciddi artışlar sağlamıştır. Bunun nedeni DVM sınıflandırıcıların istatistiksel öğrenme temeline dayanmasıdır.

Prosodik ve spektral özniteliklerin birleşiminin başarısı daha yüksek olduğu için öznitelik seçiminin sınıflandırmaya başarısının etkisi bu öznitelik kümesi üzerinde araştırılmıştır. Öznitelik seçim yöntemi olarak ileri doğru seçim (IDS), temel bileşen analizi (TBA) ve aykırı değer algılama (ADA) kullanılmıştır.

Tablo 6. Normalize edilmiş öznitelik kümesinde öznitelik seçimi sonrası elde edilen sonuçlar

Öznitelik Seçim Yöntemi	Öznitelik Boyut	DVM			ÇKA		k-EK	
		DO(%)	TO(%)	DO(%)	TO(%)	DO(%)	TO(%)	
ADA	120	↓ 19.5	88.4	↑ 1.7	89.2	↔ 0.0	71.2	↑ 1.1
IDS	55	↓ 63.1	80.8	↓ 5.9	79.0	↓ 10.2	71.0	↑ 0.9
TBA	70	↓ 53.0	85.1	↓ 1.6	83.6	↓ 5.6	61.9	↓ 8.2

TO: Duygu tanıma başarısı, DO: Değişim oranı

Tablo 6'ya göre ADA yöntemi tüm sınıflandırıcılar için en yüksek duygu tanıma oranına sahiptir. ÇKA sınıflandırıcı için başarı oranı değişmemiş olsa bile öznelik sayısı azaltıldığı için sonuç başarılı sayılabilir. ADA kullanıcı kontrolünde gerçekleştirildiği için yoruma açıktır. Sonuçlar özellik seçim yöntemine göre yorumlandığında IDS yöntemi sadece k-EK sınıflandırıcıda başarıyı artırmış olup ÇKA ve DVM'de düşüşe neden olmuştur. TBA yöntemi en yüksek başarıyı DVM sınıflandırıcı da elde etmiştir.

Ön İşlemenin Sınıflandırma Başarısı Üzerindeki Etkileri

Bu bölümde gürültü azaltma, yüksek, alçak ve bant geçiren filtrelerin sınıflandırıcı başarısı üzerindeki etkisi araştırılmıştır. Analizlerde üç sınıflandırıcı ve ADA öznelik seçim yöntemi kullanılmıştır. Ön işleme sonrası elde edilen sonuçlar Tablo 7'de verilmiştir.

Tablo 7. Ön işleme sonrası elde edilen sonuçlar

Ön İşleme Yöntemi	TO (%)								
	DVM			ÇKA			k-EK		
	PF+SF	PF	SF	PF+SF	PF	SF	PF+SF	PF	SF
-	88.4	73.3	84.9	89.2	67.2	87.9	71.2	60.5	74.3
Yüksek Geçiren Filtre	↑ 88.5	↑ 72.2	↓ 84.0	↑ 90.3	↑ 72.4	↓ 86.0	↑ 72.9	↓ 60.3	↓ 72.8
Alçak Geçiren Filtre	↓ 75.4	↔ 73.3	↓ 68.6	↓ 80.7	↔ 69.2	↓ 72.0	↓ 61.3	↓ 56.6	↓ 62.6
Bant Geçiren Filtre	↓ 73.1	↓ 72.9	↓ 67.7	↓ 79.8	↔ 78.7	↓ 71.3	↓ 68.0	↓ 55.7	↓ 61.6
Gürültü Azaltma	↓ 69.8	↓ 68.1	↓ 61.7	↓ 74.1	↓ 64.5	↓ 67.7	↓ 54.7	↓ 53.0	↓ 49.5

TO: Duygu tanıma başarısı

Yüksek geçiren filtre başarı oranında artış sağlarken alçak geçiren, bant geçiren ve gürültü azaltma başarı oranında düşüşe neden olmuştur. Bu nedenle KDT sistemlerinde alçak geçiren, bant geçiren ve gürültü azaltma yöntemlerinin kullanılması uygun değildir. Sonuçlar öznelik kümelerine göre incelendiğinde başarı oranındaki değişim özellikle SF'den kaynaklanmaktadır. Yüksek geçiren filtre başarı oranında en yüksek %5.2 artış ile PF'de sağlamıştır. Gürültü azaltma hem PF hem de SF'de başarı oranında düşüşe neden olmuştur. Özet olarak KDT sistemlerinde ön işleme

yöntemlerinde sadece yüksek geçiren filtreleme başarıyı olumlu yönde etkilemektedir.

Cinsiyet ve Yaşın Sınıflandırma Başarısı Üzerindeki Etkileri

EMO-DB 'nin içerdiği her ses kaydı konuşmacının yaş ve cinsiyet bilgisini içermektedir. Yaş ve cinsiyetin KDT üzerindeki etkisini araştırmak için en yüksek başarıya sahip olan ÇKA sınıflandırıcı, yüksek geçiren filtre ve ADA öznelik seçim yöntemi kullanılmıştır. Elde edilen sonuçlar Tablo 8'de verilmiştir.

Tablo 8. ÇKA sınıflandırıcı ile yaş ve cinsiyete bağlı sonuçlar

	TO (%)		
	PF+SF	PF	SF
Erkek	↔ 90.3	↑ 82.2	↓ 83.1
Bayan	↑ 91.8	↑ 78.6	↑ 87.5
Yaş1 (30-35)	↓ 89.2	↑ 80.9	↓ 84.7
Yaş2 (20-29)	↔ 90.3	↑ 73.3	↓ 84.9

TO: Duygu tanıma başarısı

Tablo 8'e göre PF ile erkek ve bayan içeren sınıflandırmanın tek bir cinsiyet içeren sınıflandırmaya göre daha düşük başarıya sahip olduğu görülmektedir. PF, sınıflandırma

başarısını erkek için %9.8 ve bayan için %6.2 artırmıştır. SF'de ise cinsiyet gözetmeksizin sınıflandırma sadece erkek içeren sınıflandırmaya göre %2.9 daha fazla başarılı ve

sadece bayan içeren sınıflandırmaya göre %1.5 daha başarısız olduğu görülmektedir. PF+SF özellik kümesinde elde edilen sonuçlara göre bayanlar duygularını seslerine erkeklerden daha fazla yansıtmaktadır. Aynı durum yaş için de geçerli olup 20-29 yaş aralığındaki bireyler duygularını seslerine daha fazla yansıtmaktadır.

Her Bir Duygu İçin Elde Edilen Sınıflandırma Başarısı

Elde edilen tüm sonuçlar doğrultusunda z-puan normalizasyon, ADA öznitelik seçim yöntemi ve yüksek geçiren filtre ile en yüksek başarı %90.3 ile ÇKA sınıflandırıcı da elde edilmiştir. Bu sonuca ait başarı içeren çalışmanın karışıklık matrisi Tablo 9'da verilmiştir.

Tablo 9. ÇKA sınıflandırıcı ile tanıma oranları

Duygu	Mutluluk	Nötr	Öfke	Üzüntü	Korku	Can Sık.	İğrenme	TO (%)
Mutluluk	55	1	12	0	3	0	0	77.5
Nötr	1	73	0	0	2	2	1	92.4
Öfke	10	0	117	0	0	0	0	92.1
Üzüntü	0	0	0	60	0	2	0	96.8
Korku	2	1	4	1	61	0	0	88.4
Can sıkıntısı	0	3	0	1	0	75	2	92.6
İğrenme	0	0	2	0	1	1	42	91.3

TO: Duygu tanıma başarısı

Duyguların tanıma oranları incelendiğinde en yüksek başarı üzüntü ve en düşük başarı mutluluk duygularında elde edilmiştir. Öfke duygusuna en yakın duygu mutluluk, üzüntü

duygusuna en yakın duygu can sıkıntısı olarak tespit edilmiştir. Her bir duygunun cinsiyet ve yaşa göre analizi ile elde edilen sonuçlar Tablo 10'da verilmiştir.

Tablo 10. Cinsiyet ve yaşa göre duygu tanıma başarıları

	Mutluluk	Nötr	Öfke	Üzüntü	Korku	Can Sık.	İğrenme	TO (%)
Tüm veri	77.5	92.4	92.1	96.8	88.4	92.6	91.3	90.3
Erkek	↑ 83.2	↓ 87.8	↑ 99.8	↓ 92.5	↑ 88.9	↓ 80.5	↓ 63.9	90.3
Bayan	↑ 81.3	↑ 97.9	↓ 82.8	↑ 97.8	↓ 84.9	↑ 97.8	↑ 94.8	91.8
Yaş1	↓ 76.4	↓ 88.3	↓ 85.7	↑ 97.9	↓ 84.2	↓ 91.7	↑ 96.8	89.2
Yaş2	↑ 81.2	↑ 96.8	↑ 96.5	↓ 86.9	↓ 78.9	↓ 81.2	↓ 86.7	90.3

Tablo 10'a göre erkeklerin en iyi ifade ettiği duygu öfke, bayanların ki ise üzüntü ve can sıkıntısı duygularıdır. 30-35 yaş aralığındakilerin en iyi ifade ettiği duygu üzüntü ve 20-29 yaş aralığındakilerin ise öfkedir. Bu durum erkeklerin ve gençlerin öfkelerine çok fazla yansıttığını göstermektedir. Duygu bazlı analize göre mutluluk, öfke ve korku duyguları erkeklerde; üzüntü, can sıkıntısı ve iğrenme duyguları bayanlarda; üzüntü, korku, can sıkıntısı ve iğrenme 30-35 yaş aralığındaki bireylerde; mutluluk ve öfke 20-29 yaş aralığındaki bireylerde daha kolay tespit edilmektedir.

Uygulama Sonuçları

Sunulan bu çalışmada aktörler tarafından seslendirilen ve 7 duygu durumu içeren EMO-DB kullanılmıştır. EMO-DB üzerinden çıkartılan öznitelik kümesinin, ön işleme yöntemlerinin, yaş ve cinsiyet bilgisinin duyguyu ifade etme oranı başka bir ifadeyle KDT sisteminin başarısındaki etkisi araştırılmıştır.

Elde edilen sonuçlara göre en yüksek başarı ÇKA sınıflandırıcıda elde edilmiş olup bunu sırasıyla DVM ve k-EK takip etmiştir. Ancak, DVM sınıflandırıcının analizi gerçekleştirme süresi ÇKA sınıflandırıcıdan çok daha düşüktür. DVM istatistiksel öğrenme temeline dayandığı için veri normalize edildikten sonra sınıflandırıcının

başarı oranı artarken gerçekleştirme süresi önemli derecede düşmüştür.

Sınıflandırıcının başarısını artırmak, gerçekleştirme süresini kısaltmak veya daha az sayıda öznelikle duygu tanıma gerçekleştirmek için öznelikle kümesinde öznelikle seçim yöntemleri ile boyut indirgemesi yapılmaktadır. Öznelikle kümesinin çok büyük olduğu verilerde veya duyguları tam ifade etmeyen özneliklerin veride yer alması durumunda öznelikle seçim yöntemi sınıflandırıcının başarısını artırmaktadır. Çalışmamızda yer alan öznelikle seçim yöntemlerinden ADA ile normal dağılıma sahip olmayan 29 özneliklin çıkartılması tüm sınıflandırıcılarda başarı oranını artırmıştır. Öznelikle kümesi IDS 'de %63.1 ve TBA 'da

%53.0 azaltılmasına rağmen başarı oranları daha düşük oranlarda düşmüştür. Elde edilen sonuçlar göstermiştir ki öznelikle seçim yöntemleri ile boyut indirgemesi yüksek oranda gerçekleştirildiğinde sınıflandırıcının başarısı az da olsa düşmektedir. Ancak, ADA gibi sınırlı sayıda öznelikle indirgenmesi sınıflandırıcının performansında artışa neden olmaktadır.

Yapılan analizler sonrası elde edilen özet Tablo 11'de verilmiştir. Tabloda kullanılan "<<" sınıflandırma başarısının çok fazla azaldığını, "<" sınıflandırma başarısının azaldığını, "O" sınıflandırma başarısında değişiklik olmadığını, ">" sınıflandırma başarısının arttığını ve ">>" sınıflandırma başarısının çok fazla arttığını göstermektedir.

Tablo 11. Ön işleme ve demografik verilere bağlı özet sonuçlar

	DVM			ÇKA			k-EK		
	PF	SF	PF+SF	PF	SF	PF+SF	PF	SF	PF+SF
Yüksek Geçiren Filtre	>	<	>	>>	<	>	O	<	>
Alçak Geçiren Filtre	O	<<	<<	>	<<	<<	<	<<	<<
Bant Geçiren Filtre	O	<<	<<	>>	<<	<<	<	<<	<
Gürültü Azaltma	<<	<<	<<	<	O	<<	<	<<	<<
Erkek	>	<	<	>	<	O	>	<	>
Bayan	>	>	>	>>	>	>	>>	<	>
Yaş1	>	O	>	>	O	<	<	O	>>
Yaş2	<	>	<<	>	<	O	>>	>	>

"<<" çok düşük, "<" düşük, "O", ">" yüksek, ">>" çok yüksek

Tablo 11'de verilen özet incelendiğinde cinsiyet ve yaş durumunun duyguyu ifade etmede farklılık gösterdiği ve bu durumun sınıflandırıcı başarısına yansıdığı görülmektedir. Bayanların erkeklere göre ve 20-29 yaş aralığındaki bireylerin 30-35 yaş aralığındaki bireylere göre duygularını seslerine daha fazla yansıttığı görülmüştür. Bu durum 20-29 yaş aralığında bayan katılımcılarla gerçekleştirilecek çalışmalarda daha yüksek başarı elde edileceğini göstermektedir.

Ön işleme yöntemlerinin sınıflandırıcı başarısı üzerindeki etkileri incelendiğinde yüksek geçiren, bant geçiren ve alçak geçiren filtreler genel olarak prosodik öznelikle kümelerinde başarıyı artırırken, spektral öznelikle kümelerinde düşürmüştür. Prosodik öznelikle sesin üretilmesi ile ve spektral özellikler ses sinyalinin yapısı ile ilgili olduğu için elde edilen

sonuçlar beklendiği gibidir. Yüksek geçiren filtreler tüm sınıflandırıcılarda başarı oranını artırırken alçak geçiren ve bant geçiren filtreler başarı oranında düşüşe neden olmuştur. Gürültü azaltma tüm sınıflandırıcılarda başarı oranını önemli oranda düşürmüştür. Bunun nedeni konuşma içerisindeki her türlü gürültü anlamsız değildir. Örneğin konuşmadaki nefes alıp verme şekli sese gürültü olarak yansiyabilir ama duygusal durum ile ilgili ipucu sunmaktadır. Bu gürültünün konuşmadan atılması duygunun yok olmasına neden olacaktır. Gürültü azaltmanın başarı oranını düşürmesinin nedeni EMO-DB içerisindeki ses kayıtları aktörler tarafından özel olarak oluşturulmuştur ve ortam gürültüsü içermemektedir. Bu konuşma üzerinde gürültü azaltma yapılması konuşmadaki akustik parametrelerde değişime neden olmakta ve her duygu için yaklaşık değerler elde edilmektedir. Elde edilen yaklaşık değerler de farklı duyguları

ifade etmemektedir. Sonuç olarak ön işleme yöntemlerinden sadece yüksek geçiren filtreleme sınıflandırıcı performansını artırmaktadır.

Sonuç

KDT kişilerin konuşmalarından duygu durumlarının seslerine yansımaları ve içinde bulunan duygunun tespit edilmesini amaçlamaktadır. Bu işlemi gerçekleştirebilmek için konuşmalardan çeşitli özniteliklerin çıkartılması ve bunlardan kullanışlı olanların sisteme dahil edilmesi gerekmektedir. Bu çalışmada, öznitelik kümesinin, ön işleme yöntemlerinin, yaş ve cinsiyet bilgisinin duyguyu ifade etme oranı araştırılmıştır. Özellik seçim yöntemlerinin öznitelik boyutunu yüksek oranda azaltırken, KDT başarımı düşük oran düşürdüğü sonucuna ulaşılmıştır. Cinsiyet durumu göz önüne alındığında, bayan ve erkeklerin duygularını seslerine farklı şekilde yansıttığı görülmüştür. Benzer durum yaş, cinsiyet ve duygu ilişkisinde de geçerlidir. Filtreleme yöntemlerinin KDT üzerindeki etkisi öznitelik kümesinin türüne göre değişim göstermektedir. Ayrıca, konuşmanın içerdiği gürültünün duygunun sese yansımaları üzerinde etkili olduğu görülmüştür.

Kaynaklar

Altun, Halis, ve Gökhan Polat. 2009. "Boosting Selection of Speech Related Features to Improve Performance of Multi-Class SVMs in Emotion Detection". *Expert Systems with Applications* 36 (4): 8197-8203. <https://doi.org/10.1016/j.eswa.2008.10.005>.

Bänziger, Tanja, Sona Patel, ve Klaus R. Scherer. 2014. "The Role of Perceived Voice and Speech Characteristics in Vocal Emotion Communication". *Journal of Nonverbal Behavior* 38 (1): 31-52. <https://doi.org/10.1007/s10919-013-0165-x>.

Batliner, Anton, Stefan Steidl, Björn Schuller, Dino Seppi, Thuri Vogt, Johannes Wagner, Laurence Devillers, vd. 2011. "Whodunnit – Searching for the Most Important Feature Types Signalling Emotion-Related User States in Speech". *Computer Speech & Language* 25 (1): 4-28. <https://doi.org/10.1016/j.csl.2009.12.003>.

Bayrakdar, Sümeyye, Devrim Akgün, ve İbrahim Yücedağ. 2017. "Video dosyaları üzerinde yüz ifade analizi için hızlandırılmış bir yaklaşım." *Pamukkale University Journal of Engineering Sciences* 23 (5).

Boersma, Paul, ve David Weenink. 2010. *Praat: doing phonetics by computer [Computer program], Version 5.1. 44*.

Boll, Steven F. 1979. "Suppression of acoustic noise in speech using spectral subtraction". *Acoustics, Speech and Signal Processing, IEEE Transactions on* 27 (2): 113–120.

Burkhardt, Felix, Astrid Paeschke, Miriam Rolfes, Walter F. Sendlmeier, ve Benjamin Weiss. 2005. "A database of German emotional speech." İçinde *Interspeech*, 5:1517–1520. https://www.kw.tu-berlin.de/fileadmin/a01311100/A_Database_of_German_Emotional_Speech_-_Burkhardt_01.pdf.

Chen, Lijiang, Xia Mao, Pengfei Wei, Yuli Xue, ve Mitsuru Ishizuka. 2012. "Mandarin Emotion Recognition Combining Acoustic and Emotional Point Information". *Applied Intelligence* 37 (4): 602-12. <https://doi.org/10.1007/s10489-012-0352-1>.

Chen, Lijiang, Xia Mao, Yuli Xue, ve Lee Lung Cheng. 2012. "Speech Emotion Recognition: Features and Classification Models". *Digital Signal Processing* 22 (6): 1154-60. <https://doi.org/10.1016/j.dsp.2012.05.007>.

Clavel, C., I. Vasilescu, L. Devillers, G. Richard, ve T. Ehrette. 2008. "Fear-Type Emotion Recognition for Future Audio-Based Surveillance Systems". *Speech Communication* 50 (6): 487-503. <https://doi.org/10.1016/j.specom.2008.03.012>.

El Ayadi, Moataz, Mohamed S. Kamel, ve Fakhri Karray. 2011. "Survey on Speech Emotion Recognition: Features, Classification Schemes, and Databases". *Pattern Recognition* 44 (3): 572-87. <https://doi.org/10.1016/j.patcog.2010.09.020>.

Eyben, Florian, Martin Wöllmer, ve Björn Schuller. 2010. "Opensmile: the munich versatile and fast open-source audio feature extractor". İçinde *Proceedings of the international conference on Multimedia*, 1459–1462. ACM. <http://dl.acm.org/citation.cfm?id=1874246>.

Goudbeek, Martijn, ve Klaus Scherer. 2010. "Beyond Arousal: Valence and Potency/Control Cues in the Vocal Expression of Emotion". *The Journal of the Acoustical Society of America* 128 (3): 1322. <https://doi.org/10.1121/1.3466853>.

Grimm, Michael, Kristian Kroschel, Emily Mower, ve Shrikanth Narayanan. 2007. "Primitives-Based

- Evaluation and Estimation of Emotions in Speech”. *Speech Communication* 49 (10-11): 787-800.
<https://doi.org/10.1016/j.specom.2007.01.010>.
- Hall, Mark, Eibe Frank, Geoffrey Holmes, Bernhard Pfahringer, Peter Reutemann, ve Ian H. Witten. 2009. “The WEKA data mining software: an update”. *ACM SIGKDD explorations newsletter* 11 (1): 10–18.
- Hansen, John HL, Sahar E. Bou-Ghazale, Ruhi Sarikaya, ve Bryan Pellom. 1997. “Getting started with SUSAS: a speech under simulated and actual stress database.” İçinde *Eurospeech*, 97:1743–46.
https://catalog.ldc.upenn.edu/docs/LDC99S78/sus_sas_rev1b4.ps.
- Hoque, Mohammed E., Mohammed Yeasin, ve Max M. Louwerse. 2006. “Robust recognition of emotion from speech”. İçinde *Intelligent Virtual Agents*, 42–53. Springer.
http://link.springer.com/chapter/10.1007/11821830_4.
- Huang, Zheng-wei, Wen-tao Xue, ve Qi-rong Mao. 2015. “Speech emotion recognition with unsupervised feature learning”. *Frontiers of Information Technology & Electronic Engineering* 16: 358–366.
- Joshi, Dipti D., ve M. B. Zalte. 2013. *Recognition of Emotion from Marathi Speech Using MFCC and DWT Algorithms*. IJACECT.
http://www.irdindia.co.in/journal/journal_ijacect/pdf/vol2_iss2/11.pdf.
- Kandali, Aditya Bihar, Aurobinda Routray, ve Tapan Kumar Basu. 2009. “Vocal Emotion Recognition in Five Native Languages of Assam Using New Wavelet Features”. *International Journal of Speech Technology* 12 (1): 1-13.
<https://doi.org/10.1007/s10772-009-9046-4>.
- Khanchandani, K. B., ve Moiz A. Hussain. 2009. “Emotion recognition using multilayer perceptron and generalized feed forward neural network”. *Journal of Scientific and Industrial Research* 68 (5): 367.
- Laukka, Petri, Daniel Neiberg, Mimmi Forsell, Inger Karlsson, ve Kjell Elenius. 2011. “Expression of Affect in Spontaneous Speech: Acoustic Correlates and Automatic Detection of Irritation and Resignation”. *Computer Speech & Language* 25 (1): 84-104.
<https://doi.org/10.1016/j.csl.2010.03.004>.
- Lee, Chi-Chun, Emily Mower, Carlos Busso, Sungbok Lee, ve Shrikanth Narayanan. 2011. “Emotion Recognition Using a Hierarchical Binary Decision Tree Approach”. *Speech Communication* 53 (9-10): 1162-71.
<https://doi.org/10.1016/j.specom.2011.06.004>.
- Ludeña-Choez, Jimmy, ve Ascensión Gallardo-Antolín. 2015. “Feature extraction based on the high-pass filtering of audio signals for Acoustic Event Classification”. *Computer Speech & Language* 30 (1): 32–42.
- Luengo, Iker, Eva Navas, ve Inmaculada Hernaez. 2010. “Feature Analysis and Evaluation for Automatic Emotion Identification in Speech”. *IEEE Transactions on Multimedia* 12 (6): 490-501.
<https://doi.org/10.1109/TMM.2010.2051872>.
- MacCallum, Julia K., Aleksandra E. Olszewski, Yu Zhang, ve Jack J. Jiang. 2011. “Effects of low-pass filtering on acoustic analysis of voice”. *Journal of Voice* 25 (1): 15–20.
- Mariooryad, S., ve C. Busso. 2013. “Exploring Cross-Modality Affective Reactions for Audiovisual Emotion Recognition”. *IEEE Transactions on Affective Computing* 4 (2): 183-96.
<https://doi.org/10.1109/T-AFFC.2013.11>.
- Ntalampiras, S., ve N. Fakotakis. 2012. “Modeling the Temporal Evolution of Acoustic Parameters for Speech Emotion Recognition”. *IEEE Transactions on Affective Computing* 3 (1): 116-25.
<https://doi.org/10.1109/T-AFFC.2011.31>.
- Orlandi, Silvia, P. H. Dejonckere, Jean Schoentgen, Jean Lebacqz, N. Rruqja, ve Claudia Manfredi. 2013. “Effective pre-processing of long term noisy audio recordings: An aid to clinical monitoring”. *Biomedical Signal Processing and Control* 8 (6): 799–810.
- Özseven, Turgut, ve Muharrem Düğenci. 2017. “The effects of digital filters on acoustic parameters, gender, age and emotion”. *Pamukkale University Journal of Engineering Sciences* 23 (2): 144-48.
<https://doi.org/10.5505/pajes.2016.00922>.
- Özseven, Turgut, Muharrem Düğenci, ve Alptekin Durmuşoğlu. 2018. “A Content Analysis of the Research Approaches in Speech Emotion Recognition” 7 (1): 1-26.
- Patel, Sona, Klaus R. Scherer, Eva Björkner, ve Johan Sundberg. 2011. “Mapping Emotions into Acoustic Space: The Role of Voice Production”. *Biological Psychology* 87 (1): 93-98.
<https://doi.org/10.1016/j.biopsycho.2011.02.010>.
- Polzehl, Tim, Alexander Schmitt, Florian Metzke, ve Michael Wagner. 2011. “Anger Recognition in Speech Using Acoustic and Linguistic Cues”. *Speech Communication* 53 (9-10): 1198-1209.
<https://doi.org/10.1016/j.specom.2011.05.002>.

- Rabiner, Lawrence R., ve Ronald W. Schafer. 1978. *Digital processing of speech signals*. Prentice Hall.
- Scherer, Klaus R., Johan Sundberg, Lucas Tamarit, ve Gláucia L. Salomão. 2015. "Comparing the Acoustic Expression of Emotion in the Speaking and the Singing Voice". *Computer Speech & Language* 29 (1): 218-35. <https://doi.org/10.1016/j.csl.2013.10.002>.
- Schuller, Björn, Ronald Müller, Manfred K. Lang, ve Gerhard Rigoll. 2005. "Speaker independent emotion recognition by early fusion of acoustic and linguistic features within ensembles." İçinde *INTERSPEECH*, 805–808. Citeseer. <http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.364.5140&rep=rep1&type=pdf>.
- Sezgin, Mehmet Cenk, Bilge Günsel, ve Gunes Karabulut Kurt. 2012. "Perceptual audio features for emotion detection". *EURASIP Journal on Audio, Speech, and Music Processing* 2012 (1): 1–21.
- Sundberg, Johan, Sona Patel, Eva Bjorkner, ve Klaus R. Scherer. 2011. "Interdependencies among Voice Source Parameters in Emotional Speech". *IEEE Transactions on Affective Computing* 2 (3): 162-74. <https://doi.org/10.1109/T-AFFC.2011.14>.
- Tao, Jianhua, Yongguo Kang, ve Aijun Li. 2006. "Prosody conversion from neutral speech to emotional speech". *Audio, Speech, and Language Processing, IEEE Transactions on* 14 (4): 1145–1154.
- Tarnq, Wernhuar, Yuan-Yuan Chen, Chien-Lung Li, Kun-Rong Hsie, ve Mingteh Chen. 2010. "Applications of support vector machines on smart phone systems for emotional speech recognition". *World Academy of Science, Engineering and Technology* 72: 106–113.
- Truong, Khiet P., David A. van Leeuwen, ve Franciska M.G. de Jong. 2012. "Speech-Based Recognition of Self-Reported and Observed Emotion in a Dimensional Space". *Speech Communication* 54 (9): 1049-63. <https://doi.org/10.1016/j.specom.2012.04.006>.
- Zhao, Xiaoming, Shiqing Zhang, ve Bicheng Lei. 2014. "Robust Emotion Recognition in Noisy Speech via Sparse Representation". *Neural Computing and Applications* 24 (7-8): 1539-53. <https://doi.org/10.1007/s00521-013-1377-z>.
- Zupan, Barbra, Dawn Neumann, Duncan R. Babbage, ve Barry Willer. 2009. "The importance of vocal affect to bimodal processing of emotion: implications for individuals with traumatic brain injury". *Journal of Communication Disorders* 42 (1): 1–17.

The Impact of Pre-processing and Feature Selection Methods for Speech Emotion Recognition

Extended abstract

Speech emotion recognition uses features obtained from digital signal processing and digitized sound signal. All of the features extracted from the speech can be handled under one dimension as well as grouped in terms of dimensional or structure. In this study, the effects of feature selection and preprocessing methods on emotion detection were investigated. For this purpose, EMO-DB data set and three different classifiers are used.

EMO-DB includes voice recordings consisting of 7 different emotions (anger, boredom, disgust, anxiety/fear, happiness, sadness, neutral) spoken by actors within the project developed by department of communication science at Institute of Speech and Communication of Technical University of Berlin. The database was created by emotional expressions of 10 different sentences by 10 actors in 20-35 age interval. Voice records have 16 kHz sampling frequency and 16 bit mono features.

In this paper high-pass, band-pass and low-pass filter are used in pre-processing.

Speech signal is a nonstationary signal but it is accepted as stationary in short time intervals. In order to obtain short time intervals the signal is divided into frames. In our study signal processing was performed with 20ms frame length, hamming window and 50.0% overlapping pre-processing processes. Features are extracted from each frame. Features used in the study are divided into two groups namely prosodic and spectral. Statistical variations of each feature is included in the feature set.

According to the results obtained, the highest success was achieved with 90.3% with multi-layer perceptron and high-pass filter. Spectral features provide higher success than prosodic features. In addition, females compared to males and individuals in 20-29 age interval compared to individuals in 30-35 age interval reflect their emotions more to their voices. Among the filtering methods obtained in the study, high-pass filtering increased the success of classifier whereas low-pass filtering, band-pass filtering and noise reduction reduced it.

Keywords: *Speech emotion recognition, Pre-processing, Feature selection, Filtering, Emotion classification*