Afyon Kocatepe Üniversitesi Fen ve Mühendislik Bilimleri Dergisi

Afyon Kocatepe University Journal of Science and Engineering

AKÜ FEMÜBİD **20** (2020) 067101 (1147-1155) AKU J. Sci. Eng. **20** (2020) 067101 (1147-1155)

DOI: 10.35414/akufemubid.803547

Araştırma Makalesi / Research Article

Yalıtık Sözcüklü bir Türkçe Konuşma Tanıma Sisteminin Yapay Veri Artırımı ile Tasarımı ve Gerçekleştirimi

İbrahim Baran USLU[1*], Hakan TORA[2], Emre SÜMER[3], Mustafa TÜRKER[4]

[1] *Atılım Üniversitesi, Mühendislik Fakültesi, Elektrik ve Elektronik Mühendisliği Bölümü, Ankara.*

[2] *Atılım Üniversitesi, Sivil Havacılık Yüksekokulu, Uçak Elektrik-Elektronik Bölümü, Ankara.*

[3] *Başkent Üniversitesi, Mühendislik Fakültesi, Bilgisayar Mühendisliği Bölümü, Ankara.*

[4] *Hacettepe Üniversitesi, Mühendislik Fakültesi, Geomatik Mühendisliği Bölümü, Ankara.*

**Sorumlu yazar e-posta: baran.uslu@atilim.edu.tr ORCID ID: http://orcid.org/0000-0001-5116-779X*

hakan.tora@atilim.edu.tr ORCID ID: http://orcid.org/0000-0002-0427-483X

esumer@baskent.edu.tr ORCID ID: http://orcid.org/0000-0001-8502-9184

mturker@hacettepe.edu.tr ORCID ID: http://orcid.org/0000-0001-5604-0472

Geliş Tarihi: 01.10.2020 Kabul Tarihi: 14.12.2020

Anahtar kelimeler
Konuşma tanıma; Veri artırımı; Ses aktivitesi tespiti; MFCC katsayıları; Destek vektör makinesi

Öz

Bu çalışmada toplamda doksan iki adet sesli komuttan oluşan bir yalıtık sözcüklü Türkçe konuşma tanıma sistemi tasarlanmış ve gerçekleştirilmiştir. Sistem, destek vektör makinesi (SVM) tabanlı olup, eğitimde kullanılan veri kümesi kaydedilen konuşmaların yapay olarak çeşitlendirilip artırılmasıyla elde edilmiştir. Farklı yapay veri oranlarının tanıma başarımı üzerindeki etkisi incelenmiştir. Akustik öznitelik olarak, mel frekansı kepstral katsayıları (MFCC) kullanılmıştır. Ayrıca, ses aktivitesi tespitinin ve MFCC katsayılarının tanıma başarımına etkileri de irdelenmiştir. Sonuçta doksan iki yalıtık komut için ortalama %92.6'lık doğrulukla çalışan bir konuşma tanıma sistemi geliştirilmiştir.

Design and Implementation of an Isolated-word Turkish Speech Recognition System with Data Augmentation

Keywords
Speech recognition; Data augmentation; Voice activity detection; MFCC; Support vector machine

Abstract

In this study, an isolated-word Turkish speech recognition system comprising of ninety-two voiced commands has been designed and implemented. The system is support vector machine (SVM) based and the data set used in training has been obtained by augmenting the original recordings artificially. The effect of different augmented data amounts on recognition performance has been examined. As acoustic features, mel frequency cepstral coefficients (MFCC) were used. Moreover, the effects of voice activity detection and MFCCs on recognition performance have also been investigated. In the end, 92.6% recognition accuracy on average has been obtained for ninety-two isolated commands.

1. Giriş

Konuşma tanıma teknolojileri, sesli komutlarla iş yapmayı sağlamakta; bu da özellikle engelli kişilerin yaşamını kolaylaştırmaktadır. En doğal iletişim yöntemi olan konuşma ile bir sistemi kumanda etmenin; sürekli ve yalıtık konuşmayla olmak üzere iki türü bulunmaktadır. Sürekli konuşma tanıma sistemleri, kullanıcının normal konuşma ritminde söylediği komutları algılayıp tanıyabilen sistemlerdir. Yalıtık sözcüklü konuşma tanıma sistemlerinde ise komutlar tek tek ve tane tane söylenir. Her iki konuşma tanıma sisteminde de komutların doğru şekilde tanınması için kullanılacak sınıflandırıcıların yeterli miktarda veri ile eğitilmesi zorunluluğu bulunmaktadır.

Konuşma tanıma sistemleri, konuşmacıdan bağımsız ve konuşmacı bağımlı olarak da ikiye ayrılır. Herkesin söylediği komutu doğru şekilde tanıyan bir sistem, konuşmacıdan bağımsız iken; belirli bir kişinin söylediği komutu tanıyan ve diğer kullanıcıların verdiği komutları tanımayan sistemler, konuşmacı bağımlı olarak adlandırılır.

Literatürde Türkçe konuşma tanıma ile ilgili son dönem yapılan bazı çalışmalar incelendiğinde; Coşkun ve Karadaş (2014), okul öncesi eğitime yönelik bir eğitici sistem geliştirmiştir. Ayrık sözcük tanıma tabanlı olan bu sistem; renkler, sayılar, şekiller, hayvanlar ve bazı kavramların sesli olarak tanınması ve kumanda edilmesi için tasarlanmıştır. Büyük (2018), yaptığı çalışmada mobil platformlar (cep telefonu ve tablet) için bir Türkçe konuşma tanıma sistemi geliştirmiştir. Hem sınırlı sayıda komut, hem de geniş dağarcıklı sürekli konuşma tanıma (LVCSR: Large Vocabulary Continuous Speech Recognition) amaçlı bu sistem, televizyonu sesle kumanda etmede (50 komut için) %98'lik bir sözcük tanıma başarımı göstermiştir. Diğer yandan genel metin yazdırma uygulamasında %61'lik bir sözcük tanıma oranı elde edilmiştir. Bu çalışmanın önemli bir özelliği Türkçe için mobil araçlarla bir konuşma veri tabanının (600 cümlelik) hazırlanmış olmasıdır. Gelegin ve Bolat (2011), yaptıkları çalışmada 20 adet sesli komut ile bilgisayarı kumanda etmeyi amaçlamışlardır. MFCC katsayılarını HMM (Hidden Markov Model) yapısı ile sınıflandırmışlardır. Sistem ortalama %98.2 doğruluk ile çalışmıştır.

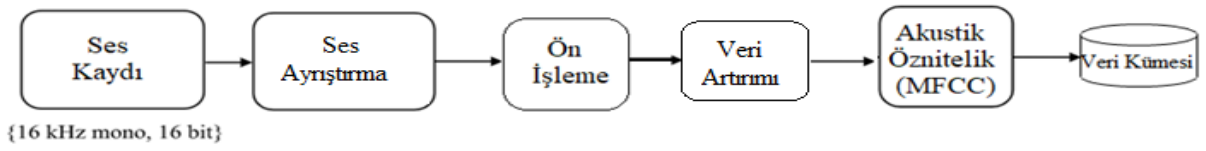
Konuşma tanıma sistemlerinin yüksek başarımla çalışması için, mümkün olduğunca fazla veriyle eğitilmesi gerekmektedir. Veri sayısının artması ile makine öğrenme yöntemleri daha başarılı şekilde genelleme ve ayrımsama yapabilmektedir. Konuşma tanıma sistemlerinde veri artırımı için, ilgili kullanıcılardan, farklı zamanlarda ve koşullarda, mümkün olduğunca fazla sayıda ses kaydı almak gerekmektedir. Fakat bu zahmetli bir yöntemdir. Onun yerine, alınan doğal ses kayıtlarını kullanarak veri artırımı yapmak mümkündür.

Bu çalışma, doğal ses kayıtlarını veri artırma yöntemleri ile çoğaltarak 92 adet yalıtık (ayrık) komutu tanıyan bir sistemin gerçekleştirilmesini ele almaktadır. Doğal seslerden yapay yöntemlerle elde edilen ses kayıtları (veri artırma) ve gerçek doğal seslerden oluşan eğitim kümesi bir SVM sınıflandırıcısını eğitmek için kullanılmıştır. Yoğun şekilde yapılan denemeler sonucunda 92 komutun kabul edilebilir bir oranda başarıyla tanındığı görülmüştür.

Çalışmanın ikinci bölümünde Konuşma Tanıma Sistemi Tasarımı, üçüncü bölümünde Deneysel Bulgular, dördüncü ve son bölümünde de Sonuçlar verilmektedir.

2. Konuşma Tanıma Sistemi Tasarımı

Bu bölümde tasarlanan konuşma tanıma sisteminin özellikleri anlatılacaktır. Yalıtık sözcüklü Türkçe konuşma tanıma sistemimizin konuşmacı-bağımlı olmasına karar verilmiştir. Çünkü tasarlanan bu sistem ileride kullanıcıya özel bazı işlevleri sesli komutlarla yerine getirecek daha büyük bir sistemin parçası olacaktır. Şekil 1'de tasarlanan sistemin eğitim bölümünün blok şeması görülmektedir.



Şekil 1. Tasarlanan sistemin *eğitim* bölümünün blok şeması

Aşağıdaki alt bölümlerde yapılan çalışmaların detayları anlatılmaktadır.

2.1 Ses kayıtlarının alınması

Konuşma tanıma sisteminin eğitim ve test aşamalarında kullanılacak ses kayıtları; her bir konuşmacıdan kapalı ve ortam gürültüsü düşük bir mekânda, bir dizüstü bilgisayar mikrofonu aracılığıyla ve tüm komutların yer aldığı tek bir oturumda kaydedilmiştir. Kayıtlar WavePad ses editörü ve Praat programı yardımıyla toplanmıştır.

Örnekleme frekansı 16 kHz seçilmiştir. 5 konuşmacıdan 92 komutun her birini ikişer kez seslendirmeleri istenmiştir. Böylelikle toplam 920 ses kaydı elde edilmiştir.

2.2 Ses ayrıştırma

Her bir kullanıcıdan tek bir oturumda kaydedilen ses kayıtları operatörler tarafından manuel olarak WavePad ses editörü ile komutlarına ayrılarak "wav" uzantılı dosyalar şeklinde saklanmıştır. Bu çalışma kapsamında Ek-A'da verilen doksan iki adet komut incelenmektedir.

2.3 Ön işleme

Ön-işlemede konuşma kayıtları bir ön-vurgu (yüksek geçiren) süzgecinden geçirilmektedir. Bu süzgeç, konuşmanın yüksek frekanslı bileşenlerini de dikkate almayı amaçlar. Çerçeveler 30 ms uzunluğunda ve %75 örtüşmeli olarak alınmıştır. Pencere fonksiyonu olarak dikdörtgen pencere kullanılmıştır.

2.4 Veri artırımı

Bu bölümde, doğal ses kayıtlarından veri artırma yöntemleri (data augmentation) irdelenmektedir. Sırasıyla, konuşma oranı (speech rate), hız (speed), VTLP (Vocal Tract Length Perturbation), perde frekansı ve ses şiddeti değiştirme yöntemleri incelenecektir. Güneş ve Bicakcı (2018)'de belirtildiği gibi, Türkçe için genel kullanıma açık bir veri seti bulunmamaktadır. Kanda vd. (2013)'e göre veri kümesi kaynakları az olan dillerde, var olan az sayıdaki verinin artırılması yoluna gidilmektedir. Ancak Ko vd. (2015)'te belirtildiğine göre veri artırımında kullanılan yöntemlerin doğal sesin frekans içeriğine yan etkileri söz konusudur. Dolayısıyla seçilecek yöntemin, konuşma tanıma başarımını eniyileyen yöntem olması hedeflenmektedir.

2.4.1 Tempo (konuşma oranı) ile veri artırımı

Konuşma oranı; tempo olarak da isimlendirilebilir. Burada WSOLA (Waveform Similarity Overlap Add) veya faz ses-kodlayıcı (phase vocoder) gibi yöntemlerle konuşmanın süresinin değiştirilmesinden bahsedilmektedir. Yani, sözlü komutun daha kısa veya daha uzun süreli türevleri üretilir. Konuşmacılar komutları o günkü ruhsal durumlarına göre daha önce kaydedilenden daha kısa veya daha uzun söyleyebilirler. Dolayısıyla tempo ile veri artırmak iyi bir seçenektir. Süresi değiştirilen ses kaydının frekans içeriği değişmez.

2.4.2 Hız değiştirme ile veri artırımı

Burada adı geçen hız parametresi, konuşmanın yeniden örneklenmesine karşılık gelmektedir. Yani, $x(t)$ sinyalinden $x(kt)$ elde edilmekte; $k>1$ için komut sıkıştırılmakta, $k<1$ için ise komut uzatılmaktadır. Bu işlem, zaman ekseninin bükülmesine (time warping) karşılık gelmekte ve hem süreyi hem de frekans içeriğini değiştirmektedir. Bu yöntem, konuşmacının ses tonundaki ve söyleme hızındaki değişiklikleri bir arada ele almaya karşılık gelir.

2.4.3 VTLP ile veri artırımı

VTLP (Vocal Tract Length Perturbation) yöntemi, konuşmacılar arası değişimin sebebi olan ses yolu uzunluğu farkını kullanır. Konuşmacı normalizasyonu için kullanılan bu yöntem, mevcut çalışmamızda verilerin çeşitlendirilmesi amacıyla kullanılmıştır. Yöntem, frekans ekseninin bükülmesi (frequency warping) tabanlıdır. Jaitly ve Hinton (2013)'te, bu yöntemin konuşma tanıma sisteminin başarımını artırdığı belirtilmektedir.

2.4.4 Perde frekansı değiştirme ile veri artırımı

Bu yöntemde, konuşmanın ötümlü (voiced), yani periyodik özellik gösteren bölümlerinde perde frekansı olarak isimlendirilen temel frekans belirli oranlarda (± %10 gibi) artırılıp azaltılmaktadır. Frekansın artırılması sesin incelmesine, azaltılması ise sesin kalınlaşmasına sebep olmaktadır.

2.4.5 Ses şiddeti değiştirme ile veri artırımı

Bu da sesin şiddetinin değiştirilmesiyle çeşitliliğin sağlanmasıdır.

Bu çalışmada, yukarıdaki yöntemlerden tempo ve perde frekansı değiştirmeyle veri artırımı yapılmıştır. Konuşma oranı ± %20, perde frekansı ise

± %10 oranlarında ve rastgele şekilde değiştirilmiştir. Bu işlemlerin gerçekleştirimi, Matlab programlama ortamı üzerinde geliştirilen bir uygulama aracığıyla otomatik olarak yapılmıştır.

2.5 MFCC katsayılarının hesaplanması

Akustik öznitelik olarak MFCC (Mel Frequency Cepstral Coefficients – Mel Frekansı Kepstral Katsayıları) kullanılmıştır. Logaritmik enerji parametresi olan ilk katsayı alınmamış, geri kalan 13 katsayı irdelenmiştir.

Duyum özelliklerimize paralel olarak mel ölçeği:

$$M(f) = 1127 \ln(1 + f/700) \qquad (1)$$

$S_i(k)$, $s_i(n)$ konuşma çerçevesinin hızlı Fourier dönüşümü (FFT) olmak üzere:

$$S_i(k) = \sum_{n=1}^{N} s_i(n)h(n)e^{-j2\pi kn/N},\ 1 \le k \le K, \qquad (2)$$

(*K*: FFT uzunluğu)

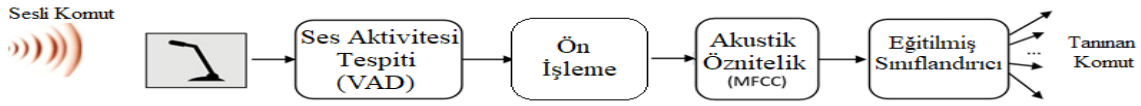
Burada $h(n)$ analiz penceresini (Hanning, Hamming gibi) temsil etmektedir. *N*; pencere uzunluğudur.

$$P_i(k) = \frac{1}{N}|S_i(k)|^2 \qquad (3)$$

$P_i(k)$; periodogram olarak adlandırılır ve güç spektral yoğunluğunu temsil eder.

Mel frekansı filtre bankaları (%50 örtüşen üçgen filtreler) ile periodogram çarpılır ve filtre bankalarının altında kalan toplam enerji hesaplanır. Bu enerji değerlerinin logaritması alınır (insan kulağının duyumsal modeli gereği) ve DCT (Discrete Cosine Transform – Ayrık Kosinüs Dönüşümü) ile fazlalık bilgilerden kurtularak kepstral katsayılar elde edilir.

Tasarlanan sistemin test bölümünün blok şeması Şekil 2'dedir.



Şekil 2. Tasarlanan sistemin *test* bölümünün blok şeması

Eğitimde kullanılan kayıtların başı ve sonu elle kesilebilmektedir. Ancak gerçek zamanlı testlerde ses aktivitesi tespiti (SAT = VAD: Voice Activity Detection) önem kazanmaktadır. Eğer verilen sesli komutun başı ve sonu doğru tespit edilmezse tanıma başarımı düşmektedir. Ses aktivitesi tespitinin doğruluğunun konuşma tanıma başarımındaki önemi Oyucu vd. (2020)'lerinin çalışmasında da incelenmiştir.

3. Deneysel Bulgular

Bu bölümde, konuşma tanımada, doğal ve artırılmış verilerin akustik özelliklerinin sınıflandırılmasında kullanılan SVM (Support Vector Machine) yapısı anlatılacaktır. Sınıflandırmada ikinci dereceden (quadratic) SVM tercih edilmiştir.

Eğitim kümesinin doğal ve yapay ses oranları ve tanıma başarımları bu bölümde incelenmektedir. Yalnızca doğal seslerle eğitilen SVM yapay sesleri hiç tanımamıştır; yalnızca yapay seslerle eğitilen SVM de doğal sesleri hiç tanımamıştır. Çizelge 1.'de, her bir konuşmacıdan alınan ikişer adet (92 komutluk – Komut listesi Ek-A'de verilmiştir) kayıttan, ilki eğitim kümesinde olduğu gibi 20'şer kez tekrarlanmıştır.

Bu kayıtlardan 20 adet perde frekansı ve 20 adet de tempo değiştirmeyle yapay veri elde edilerek eğitim kümesinde kullanılmıştır. Eğitim kümesinde toplamda 40 adet yapay ve 20 adet doğal ses yer almaktadır. Testte ise eğitimde kullanılmayan diğer doğal kayıtlar kullanılmıştır. Bu şekilde 5 farklı konuşmacıyla yapılan testlerin sonucunda ortalama %92.6'lık tanıma oranı elde edilmiştir.

Aynı test bu defa yapay sesleri tanıma başarımını ölçmek için tekrarlanmıştır. Çizelge 2.'de elde edilen sonuçlar verilmiştir. Bu deneyde, eğitimde kullanılmayan 10 adet yapay veri testte kullanılmış ve tüm kullanıcılarda tam başarımla tanınmıştır. Çizelge 3.'te ise eğitim kümesinin ideal içeriğinin belirlenmesi için yapılan testlerin sonuçları

görülmektedir. İki konuşmacı için yapay ve doğal verilerin farklı kombinasyonları için test başarımına bakılmıştır. En uygun eğitim kümesinin, hesaplama karmaşıklığı ve başarım sonucu dikkate alındığında, 5 adet yapay veriyle 20 kez tekrarlanan doğal verinin karışımından elde edildiği değerlendirilmektedir

Çizelge 1. Doğal seslerle yapılan test sonuçları (92 komut için).

Eğitim Kümesi	Test Kümesi	Kullanıcı	Eğitim Başarımı	Test Başarımı
40'ar adet Yapay + 20'şer adet (20 defa tekrarlanan) Doğal	1 adet Doğal (92 komut)	K1	%100	%100
		K2	%100	%92.4
		K3	%100	%100
		K4	%100	%88
		K5	%100	%82.6
		Ortalama:	**%100**	**%92.6**

Çizelge 2. Yapay seslerle yapılan test sonuçları (92 komut için).

Eğitim Kümesi	Test Kümesi	Kullanıcı	Eğitim Başarımı	Test Başarımı
40'ar adet Yapay + 20'şer adet (20 defa tekrarlanan) Doğal	10 adet Yapay	K1	%100	%100
		K2	%100	%100
		K3	%100	%100
		K4	%100	%100
		K5	%100	%100
		Ortalama:	**%100**	**%100**

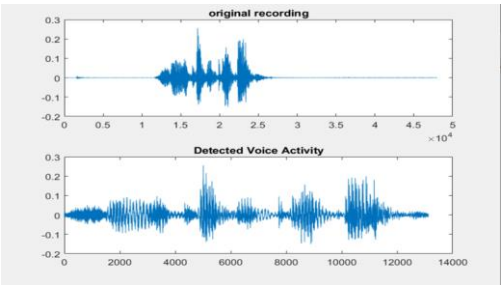
Çizelge 3. Yapay seslerin tanıma başarımına etkisi (92 komut için).

Eğitim Kümesi	Test Kümesi	Kullanıcı	Eğitim Başarımı	Test Başarımı
40'ar adet Yapay + 20'şer adet (20 defa tekrarlanan) Doğal	1 adet Doğal (92 komut)	K4	%100	%88
		K5	%100	%82.6
30'ar adet Yapay + 20'şer adet (20 defa tekrarlanan) Doğal	1 adet Doğal (92 komut)	K4	%100	%82.6
		K5	%100	%82.6
20'şer adet Yapay + 20'şer adet (20 defa tekrarlanan) Doğal	1 adet Doğal (92 komut)	K4	%100	%88
		K5	%100	%83.7
10'ar adet Yapay + 20'şer adet (20 defa tekrarlanan) Doğal	1 adet Doğal (92 komut)	K4	%100	%88
		K5	%100	%83.7
5'er adet Yapay + 20'şer adet (20 defa tekrarlanan) Doğal	1 adet Doğal (92 komut)	K4	**%100**	**%88**
		K5	**%100**	**%83.7**
1 adet Yapay + 20'şer adet (20 defa tekrarlanan) Doğal	1 adet Doğal (92 komut)	K4	%100	%85.9
		K5	%100	%80.4
20'şer adet Yapay + 10'ar adet (10 defa tekrarlanan) Doğal	1 adet Doğal (92 komut)	K4	%100	%88
		K5	%100	%82.6

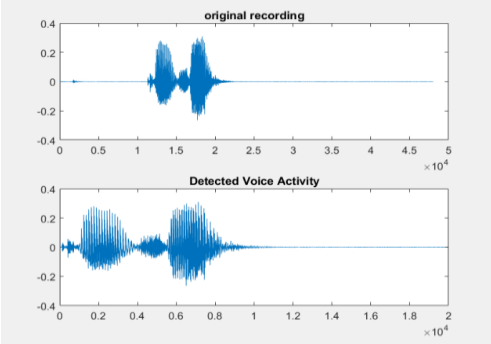
Şekil 3. ve Şekil 4.'te SAT (Ses Aktivitesi Tespiti) ile elde edilen başarılı ve hatalı tespitler gösterilmiştir.

Pratikte, filtre bankalarından sıfır çıkan MFCC katsayıları logaritma işleminde hata vermektedir. Bu da tanıma başarımını düşüren bir diğer faktördür. Bunun önlenmesi için konuşma kayıtlarına SNR = 50 dB olacak şekilde beyaz gürültü eklenmiştir. Böylece katsayıların sıfır olması önlenmekte ve bu da tanıma başarımındaki olumsuz etkiyi ortadan kaldırmaktadır (Tüfekci and Disken, 2019).

Hem SAT'ın (Ses Aktivitesi Tespiti), hem de MFCC katsayılarına beyaz gürültü eklemenin tanıma başarımına olan etkilerini incelemek için, sıfır-on bir arası rakamlarla (toplam 12 komut) yapılan testlerin sonucunda Çizelge 4.'teki sonuçlar elde edilmiştir. Burada komutların başına elle yapay boşluklar eklenmiştir ve tanıma başarımına bakılmıştır. Boşluk miktarı arttıkça, yani SAT hatası büyüdükçe tanıma başarımı düşmektedir. MFCC katsayılarına beyaz gürültü eklemek ise başarımı artırmaktadır.



Şekil 3. Başarılı SAT (Ses Aktivitesi Tespiti) sonucu.



Şekil 4. Hatalı SAT sonucu.

Çizelge 4. Doğru tanıma oranlarının SAT durumuna ve MFCC katsayılarına göre değişimi (12 adet rakam için).

Çok paylı SAT	Az paylı SAT	Paysız SAT	Özellik
3/12	5/12	7/12	MFCC katsayıları
6/12	7/12	11/12	MFCC + beyaz gürültü (SNR = 50 dB)

4. Tartışma ve Sonuç

Bu çalışmada yalıtık sözcüklü bir Türkçe konuşma tanıma sistemi tasarlanmış ve gerçekleştirilmiştir. Sınıflandırıcı eğitiminde, konuşmacılardan toplanan sınırlı sayıdaki doğal kayıtların yapay yöntemlerle artırılması ile oluşturulan veri kümesi kullanılmıştır. Böylece zahmetli olan tekrar tekrar kayıt almak yerine kayıtlar üzerinde tempo ve perde frekansı değişiklikleri yapılarak veri artırımı gerçekleştirilmiştir. Yapay verilerin tanıma başarımı üzerindeki etkisinin incelenmesi bu çalışmanın özgün tarafıdır. Ayrıca MFCC katsayılarının hesaplanmasında sıfırın logaritmasından kaynaklı olumsuz durumun giderilmesi için SNR = 50 dB olacak şekilde komutlara beyaz gürültü ilave edilmiş, bu işlemin de tanıma başarımındaki etkisi incelenmiştir. Sonuç olarak, 92 yalıtık komut için ortalama %92.6 doğrulukla çalışan bir Türkçe konuşma tanıma sistemi tasarlanmış ve gerçekleştirilmiştir.

5. Kaynaklar

Boersma, P. "Praat, a system for doing phonetics by computer". Glot International **5:9/10** (2001): 341-345.

Büyük, O., 2018. Mobil araçlarda Türkçe konuşma tanıma için yeni bir veri tabanı ve bu veri tabanı ile elde edilen ilk konuşma tanıma sonuçları, *Pamukkale Üniversitesi Mühendislik Bilimleri Dergisi*, **24(2)**, 180-184.

Coşkun, A. ve Karadaş, İ., 2014. Okul öncesi eğitime yönelik ses kontrollü eğitim yazılımı, *Pamukkale Üniversitesi Mühendislik Bilimleri Dergisi*, **20 (2)**, 36-41.

Gelegin, İ. ve Bolat, B., 2011. Ayrık kelime tabanlı bir konuşma tanıma sistemiyle bilgisayar kontrolü, *Elektrik-Elektronik ve Bilgisayar Sempozyumu,* Elazığ: 5-7.

Güneş, H. ve Bicakcı, S., 2018. Akıllı evler için sesli komut algılama yöntemleri, *Balıkesir Üniversitesi Fen Bilimleri Enstitüsü Dergisi*, **20(2)**, 561-568.

Jaitly, N. and Hinton, G. E., 2013. Vocal tract length perturbation (VTLP) improves speech recognition, *In Proc. ICML Workshop on Deep Learning for Audio, Speech and Language*, **117**.

Kanda, N., Takeda, R., and Obuchi, Y., 2013. Elastic spectral distortion for low resource speech recognition with deep neural networks, *In 2013 IEEE Workshop on Automatic Speech Recognition and Understanding*, IEEE, 309-314.

Ko, T., Peddinti, V., Povey, D., and Khudanpur, S., 2015. Audio augmentation for speech recognition, *In Sixteenth Annual Conference of the International Speech Communication Association*.

Oyucu, S., Polat, H. ve Sever, H., 2020. Sessizliğin kaldırılması ve konuşmanın parçalara ayrılması işleminin Türkçe otomatik konuşma tanıma üzerindeki etkisi, *Düzce Üniversitesi Bilim ve Teknoloji Dergisi*, **8.1**, 334-346.

Tüfekci Z. and Dişken, G., 2019. Scale-invariant MFCCs for speech/speaker recognition, *Turk J Elec Eng & Comp Sci*, **27**, 3758–3762.

İnternet kaynakları

1-https://www.nch.com.au/wavepad/index.html, (17.09.2020)

Ek-A

(Komut Listesi)

1.	Sistemi Kapat	42.	Beş	83.	ş (şe)
2.	Ana Menü	43.	Altı	84.	t (te)
3.	Bekleme	44.	Yedi	85.	u
4.	Gözlem	45.	Sekiz	86.	ü
5.	Önceki Menü	46.	Dokuz	87.	v (ve)
6.	Başlat	47.	On	88.	y (ye)
7.	Geri	48.	Onbir	89.	z (ze)
8.	Çıkış	49.	Tamam	90.	Engelsiz Yaşam
9.	Kontrol	50.	Önceki	91.	Hey Gözlem
10.	Telefon	51.	Sonraki	92.	Sil
11.	Eğlence	52.	Kanal Artı		
12.	Televizyon	53.	Kanal Eksi		
13.	Müzik	54.	Ses Artı		
14.	Kitap	55.	Ses Eksi		
15.	İnternet	56.	Sessiz		
16.	Radyo	57.	Sayfa Aşağı		
17.	Haber	58.	Sayfa Yukarı		
18.	Youtube (yuutup)	59.	Dur		
19.	Işık	60.	Boşluk		
20.	Perde	61.	a		
21.	Klima	62.	b (be)		
22.	Yatak	63.	c (ce)		
23.	Aç	64.	ç (çe)		
24.	Kapat	65.	d (de)		
25.	Arttır	66.	e		
26.	Azalt	67.	f (fe)		
27.	İndir	68.	g (ge)		

28.	Kaldır	69.	ğ (yumuşak ge)
29.	Baş Yükselt	70.	h (he)
30.	Baş Alçalt	71.	ı
31.	Kabul Et	72.	i
32.	Reddet	73.	j (je)
33.	Mesaj	74.	k (ka)
34.	Ara	75.	l (le)
35.	Gönder	76.	m (me)
36.	Oku	77.	n (ne)
37.	Sıfır	78.	o
38.	Bir	79.	ö
39.	İki	80.	p (pe)
40.	Üç	81.	r (re)
41.	Dört	82.	s (se)