International Journal of
Intelligent Systems and Applications in Engineering

IJISAE

ISSN:2147-6799 www.ijisae.org Original Research Paper

Detecting of Warning Sounds in the Traffic using Linear Predictive Coding

Cansu AKYÜREK ANACUR[1], Rıdvan SARAÇOĞLU[2*]

Submitted: 17/07/2019 Accepted : 14/10/2019

Abstract: Providing attendance of disabled persons into life and increasing it is an important social issue. In this context detecting sirens and sounds of those vehicles which have priority of way in traffic such as ambulance, fire-fighting vehicle and police car will enable to hearing disabled people to drive more comfortably. Recognising such warning sounds and detecting their direction have been aimed in this study. Sirens of ambulance, police car and fire-fighting vehicle in traffic have been classified as positive samples for application. Noises such as music and traffic noises have been classified as negative. Linear Estimator Coding has been made up by converting sound signals into digital data and qualities that reflect the sound in the best way has been determined by removing the qualities that do not reflect the sound in the data. By using principal component analysis method, meaningful and those qualities that represent the data in the best way have been classified by K Nearest Neighbor and Support Vector Machine Algorithm. After creating model about sound detection, model has been tested by performing new sound records. For the sound which has been detected to be warning sound, information about the sound direction has been given to user.

Keywords: *Linear Predictive Coding, Principal Component Analysis, Support Vector Machine, Sound Classification.*

1. Introduction

According to Turkey Health Survey of TUIK, 12,9 percent of 15-64 aged population of Turkey in 2016 is composed of hearing disabled people [1]. Traffic Registration and Supervision Bureau of Provincial Security Directorates started body language and traffic education for hearing disabled people during 2016-2017 academic year in order to increase the number of hearing disabled people in traffic and make them take part into life [2]. As hearing disabled people can not hear the sirens and warning sounds of ambulance, police car and fire-fighting vehicles which have priority of way, they may be late in terms of giving way. Especially for ambulance importance of one minute is crucial.

In this study for detecting sirens and warning sounds of vehicles and determining their directions such as ambulance, police car and fire-fighting vehicle which have priority of way in the traffic for contributing to driving of hearing disabled people more comfortably, an infrastructure has been tried to be built.

Sound is simple vibration that ear can hear. It can be defined as air molecules making hearing sense by affecting human ear or small air pressure changes whis these cause or auditory impression which this physical event leads to [3].

Recognising sound is a process which converts an acoustics signal into a character set [3]. It explains the the process in which a sound signal taken from outworld is converted into digital signs from analog sign in computer environment, that is the match with a sound known before thanks to necessary programming and software tecniques by converting it into digital data.

Sound detecting technology has been studied since 1950s. Its accuracy and competence are increasing depending on improving data set day by day. The better samples taken from digitised sound represent that sound and the better data set made up from samples taken is, the more successful sound recognising becomes.

Data set has been made up by converting sound signals into digital data and choosing the most meaningful and the best qualities which represent the data throwing unnecessary things away. Data set has been classified by Support Vector Machines (SVM) method. When any sound in data set was given to system, sound was detected whether it belonged to those vehicles having priority of way or not.

This study which essetially contains sound recognising and classification is a basis for setups which will enable hearing disabled people to drive more comfortably in traffic. It is an imprortant study both in terms of scientical and social aspect. As it is an example for upcomig studies, it will make benefit for more advanced systems to come out in terms of scientifical aspect. It will also create awareness about hearing disabled people who have troubles in traffic. If the model obtained by this study is transferred to a system, it will make positive effect on self-confidence of hearing disabled people driving. As a result of this, the number of hearing disabled people who drive will increase. Basing studies like this study which makes life easier for disabled people, producing new equipment will make life easier for disabled people and make them attend into life more.

[1] *Electrical and Electronics Eng. Van Yüzüncü Yıl University, Van – 65080, TURKEY*
[2] *Electrical and Electronics Eng. Van Yüzüncü Yıl University, Van – 65080, TURKEY*
ORCID ID : 0000-0002-5853-4545
** Corresponding Author Email: ridvansaracoglu@yyu.edu.tr*

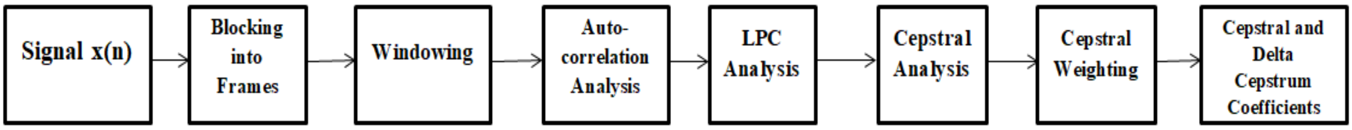

Fig. 1. Feature Extraction Steps [12]

2. Literature Review

Sound recognising, sound processing has been among the study subject of academicians in recent years. They have been often used in fields such as smart housing system and car control systems. Researchers have studied especially on speaker recognising systems. It has been seen from the source scaanning that researchers generally use feature extraction and its methods to process sound signal. İnal [4] has studied on text-dependent and text-independent speaker recognising system by using Artificial Neural Nets (ANN) algorithmn. He made recognising process avaliable extracting unnecessary qualities from data by means of ANN algorithmns by using feature extraction algorithmn which is an essential part of speaker recognising and sound detecting systems.
Campbell and his collegues [5] has carried out feature extraction in their studies and made classification after converting sound signals into vectors. They suggested a performable polynomial speaker recognising method together with hidden Markov models and other techniques.
Ertaş and Hanilçi [6] has suggested a new classifying for text-free speaker recognising system in their papers. In their studies they compared Vector Quantising (VQ) based on feature extraction principle to class non-principal component analysis algorithmn which benefits from feature extraction but are treated by different processes later.
Hanilçi [7] has studied Hidden Markov Models (HMM) and Vector Quantising (VQ) algorithmns which are used frequently in terms of text-free speaker determing in his thesis study. He has aimed to detect the most ideal size of feature vectors. By studying feature extraction methods and HMM algorithmn he revealed a text-free speaker system. He revealed a system which has the highest number of speakers among the text free speaker systems created before.
Başaran [8] has tried to detect speaker by ANN and back-propagation algorithmn by realizing reducing data size via VQ method after feature vectors groups extracting Mel Frequency Kepstrum Parameters of sound signals.
Campbell et al. [9] have applied SVM for speaker and language recognising. There is a core using approach producing similiarity measure and sequence comparing of vectors. This study has been proved to be a strong technique for SVM model classification.

3. Material and Method

For a voice recognition system to recognize the sound, the audio signal must be correctly expressed. In other words, the elements that define only that sound in the whole sound must be defined by specific signal processing techniques [10]. Data that is ready to be recognized can be recognized by the voice recognition system. In order to be able to recognize the received sounds, the voice has to go through some operations in the system that enables the recognition of warning sounds for the people with hearing loss. In order to be able to classify the warning sounds with higher accuracy, feature selection methods have been applied. Some pre-treatments have been applied to enable the data to be feature selected. The following part describes the structure of the system from raw data collection to recognition:

3.1. Collection of Raw Data

With the help of a personal computer, the siren sounds of the vehicles with pass priority and other sounds such as noise, music, car horn, etc., which can be observed in the traffic environment, are recorded. In this system, the warning sounds of the vehicles with the priority of traffic in the traffic are considered as positive sounds, while the other sounds recorded outside these sounds are considered as negative voices.
Sounds, which are in ".wav" format on the computer, are called from the coding environment and transforms into a matrix with numerical values. The sound in .wav format that we can listen here is now expressed in numerical values.

3.2. Feature Extraction

Digitized audio data gets subject to feature extraction operation.
Feature extraction is the process of converting the audio signal into vectors in which the parts that do not contain information about the sound are removed and the sound-specific parameters are highlighted. Feature extraction can also be defined as the reduction of mathematically high-dimensional vectors to lower-dimensional vectors [11]. Feature extraction is one of the important steps of the recognition system because it affects the performance of the classification [12]. The LPC is based on the principle that it can be modeled with the output of a linear and time-varying system induced by periodic impulse or random noise [20]. In this method, sound samples are estimated by looking at the past samples [12]. LPC method is one of the most preferred methods because it takes up little memory. In this study, feature vectors are extracted by using Linear Predictive Coding (LPC) method as a feature extraction method. The LPC provides a model for good estimation of the characteristic of the audio signal. The LPC calculation method is simple and straightforward to implement in either software or hardware. The steps of feature extraction are shown in Figure 1. The incoming audio signal is first divided into frames by passing through the frame blocking step. Each frame is windowed to minimize discontinuous sections at the beginning and end of the signal. Each windowed signal is subjected to autocorrelation analysis. Then, from the autocorrelation of each frame, the LPC parameter set is calculated. Then, Cepstral Analysis is done and LPC coefficients are converted to cepstral coefficients. Using Cepstral and delta cepstrum coefficients, a 10-dimensional observation vector is obtained.

3.3. Dimension Reduction

Data size is reduced to increase classification success after creating quality vector made up of qualities which contain important features of sound by removing unnecessary data digitalising sound signal given to system. Dimention reduction of data both increases quality of data set and improves classification success. Principal Component Analysis (PCA) method was used for dimension reduction. PCA is a multivariate statistical method used in the fields of recognition, classification, image compression, explaining the variance covariance structure of a set of variables by means of linear combinations of these variables. In this method, (p) number of variables showing the interdependence structure and the number of measurements (n); linear, orthogonal (orthogonal) and being independent of each other (k) are converted into new variables. PCA is a very effective method of revealing the necessary information in the data. By finding the general characteristics of high dimensional data, it decreases the number of dimensions and compresses the data [21]. PCA is uncontrolled linear converting method used frequently for dimention reduction. PCA carries data set to a new data space by protecting the deviation in data distribution [13]. In the new created data set the best qualities are placed at first.

Certain features are used in data set which is extracted by PCA and the best qualities are places at the first. These features make up data set whose size has been reduced. Training and test sets are ready to be created by these features.

3.4. Classification

Dimention reduction is done and can be done setting training and testing data sets have been set then subjected to classification process.

3.4.1. SVM

The SVM suggested by Cortes and Vapnik [14] uses the principle of structural risk minimization. SVM is a method of machine learning that divides data into two classes with the help of hyperplane [15].

SVM is one of the classifiers used for many different tasks, especially in recent years [16]. In this method, it is tried to find a high plane where the distances of the nearest samples between the two classes are maximized. This method can often be used for linearly separating data; besides, it can also be used for non-linearly separable data because it can make the data linearly separable with the help of the kernel functions [15]. The working principle of the classifier can be explained by an example; the data is classified by labeling the class to be detected as 1, and all other data in the other class as -1. This process is repeated on all models for each class separately. Finally, the results obtained from the models are combined to reveal the real class knowledge.

3.4.2. KNN Algorithm

KNN is one of the simplest lattices recognising methods classifiying objects basing on the nearest training samples in data set [17]. According to all avaliable states, it is a simple algorithmn classifying new states in accordance with a similiarity measurment (distance functions). By looking at the distance between data to be classified and its neighbors, classification process is carried out [18]. Three distance functions which can be commonly used in KNN classification are Minkowsky, Euclidean and Manhattan methods [19].

After performing distance calculation by means of any of these methods, the class it will be included is determined according to a k value according to the nearest k neighbor. Sample to be tested is compared one by one to each sample in training group. In order to determine class of sample to be tested, the nearest k neighbor to this sample in training group is studied and data to be classified is included to the group which has the highest number of samples in terms of k neighbours [17].

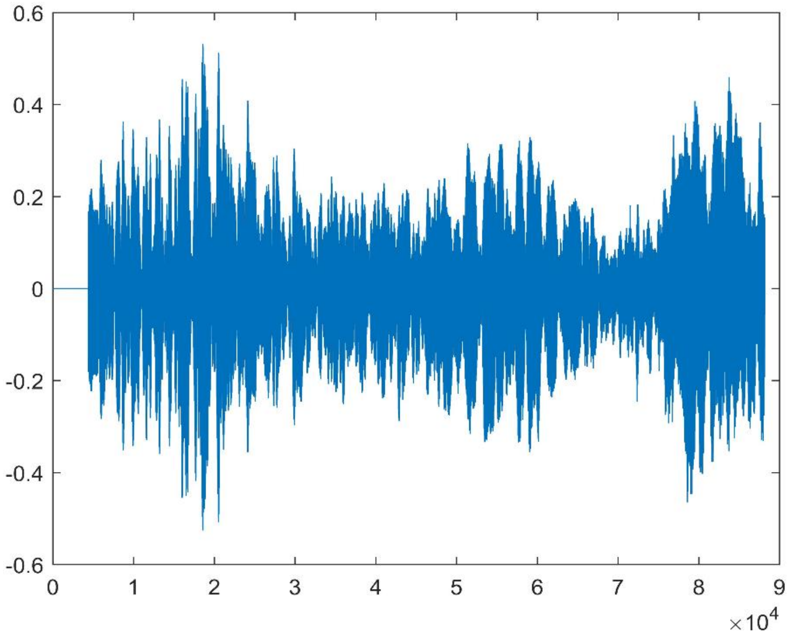

Fig. 2. A Sample Sound Signal

3.5. Data Set

Data set is composed of 47 positive and 47 negative sounds. While positive sounds consist of siren sounds of ambulance, police car and fire-fighting vehicle, negative sounds consist of street noise, horn sound, music sound etc. Each sound is 2 second long. A matrix made up of digital data of sound is obtained by digitalising two-second sounds in wav form in programming environment. This matrix is denoted by 88200 values for two second. Graph belonging to a- two -second sample is given in Fig.2.

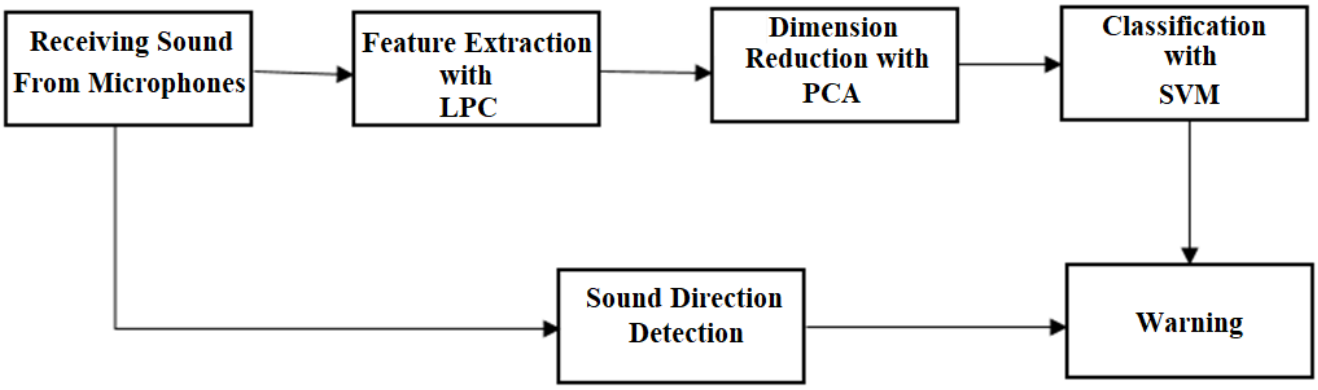

Fig. 3. System Model

4. Proposed Approach and Findings

For hearing disabled people suggested method for recognising siren sounds of vehicles such as ambulance, police car and fire-fightining vehicle which have priority of way is schematized in figure 3. Sounds from microphone is firstly taken in sound recognising process, is treated to feature choice by LPC and then is treated to dimension reduction by PCA. Data whose size is reduced by PCA is classified by SVM, after recognising sound warning is given. Following recognising process and the followed way for recognition of warning sounds are told:

First, to recognise the siren sounds of vehicles such as ambulance, police car and fire-fighting vehicle which have priority of way, obtained sound records have transferred to computer environment.

After sound being taken into computer environment, features which bear the qualities of sound best are detected. Features in sound but features that make it difficult to detect sound are removed. This study is feature extraction study.

LPC has been used as feature extraction method. It has been treated by feature extraction stages respectively. Coding created and sound signal have been divided into frames by firstly passing it into frame blocking stage. In order to minimize discontinuous units at the initial and last part of signal, each frame has been windowed by hamming windowing method. Signal windowed has been put to signal autocorrelation analysis. And then LPC parameter group in autocorrelation belonging to each frame has been estimated. Then LPC parameters have been turned into kepstral parameters by performing kepstral analysis. Ten-dimensional observing vector has been obtained by using kepstral and kepstrum parameters.

As a result of these processes, a sound signal denoted by 88200x1 matrix has changed into a 10x550 matrix and then each soung has changed into a single line 1x5500 vector denoted by 5500 value.

After carrying out these processes for each sound, 94x5500 quality vectors composing of all sounds.

The obtained data set has been firstly put to KNN and then to SVM classification processes.

While performing these classification processes, 5-Fold Cross Validation has been applied. After classification carried out by KNN, the success rate is 94,7 percent wheraas it is 93,6 percent after the classification process carried by SVM. Classification results performed by KNN and SVM are shown in the Table 1.

After feature extraction and quality vectors are obtained, PCA algorithmn is used as dimention reduction method to reduce size. Data set has been arranged in the best way in an order that the best features stand first by exracting features via PCA.

From data set in which the best qualities extracted by means of PCA are put at the outdet, a ceartain number of qualities has been choosen by starting initially.

A sound data denoted with 94x5500 matrix by feature extraction process carried out by LPC has been turned into 94x93 matrix after PCA process. By choosing 15 columns from data set denoted with 94x93 matrix in which the best ones are the first row by PCA, it has been turned into 94x15 matrix.

After dimention reduction process, the obtained data set has been firstly put to KNN and then to SVM classification processes.

While performing these classification processes, 5-Fold Cross Validation has been applied. After classification carried out by KNN, the success rate is 90,4 percent wheraas it is 92,60 percent after the classification process carried by SVM. Classification results performed by KNN and SVM are shown in the Table 2.

Table 1. Classifications Results

Classification Method	*Sensitivity*	*Specificity*	*Positive predictive value*	*Negative predictive value*	*Accuracy*
Cosine KNN	95,74%	93,61%	93,75%	95,65%	94,70%
Linear SVM	95,74%	91,48%	91,83%	95,55%	93,60%

Table 2. Classifications Results After PCA

Classification Method	*Sensitivity*	*Specificity*	*Positive predictive value*	*Negative predictive value*	*Accuracy*
Cosine KNN	100,0%	83,85%	83,92%	100%	90,40%
Linear SVM	97,87%	87,23%	88,46%	97,61%	92,60%

After classification, sound given to system has been detected whether it is one of those sounds which are among warning sounds having priority of way in traffic.

Direction of sound detected as to be warning sound has been tried to be detected. Even when trying to figuring out the direction of sound in a simple way, the first thing to consider is that sound belongs to the microphone in which sound is taken in a high way. Judging from this principlesounds in the microphones have been studied. As data set is composed of two-second-sounds, sound direction information has been tried to be found on two-second-sounds. In graph 4 and 5, the graphics of right and left microphones of a two-second-sound has been given. Studying the graphics, it is understood that amplitude values in raight and left microphones are different.

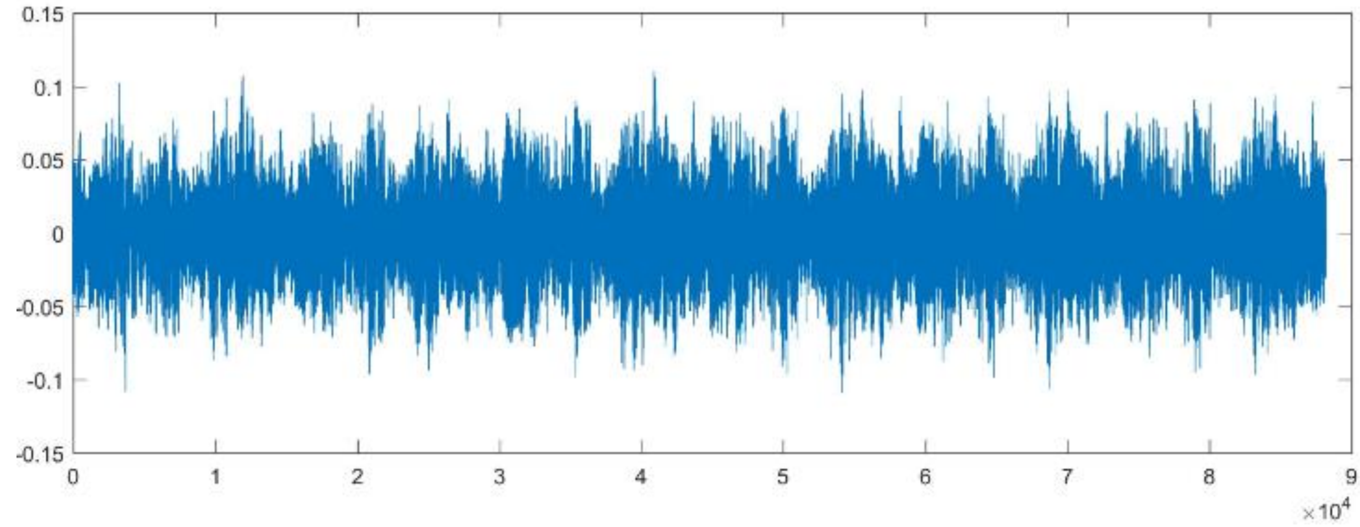

Fig. 4. Sound Signal of Microphone on Left Side

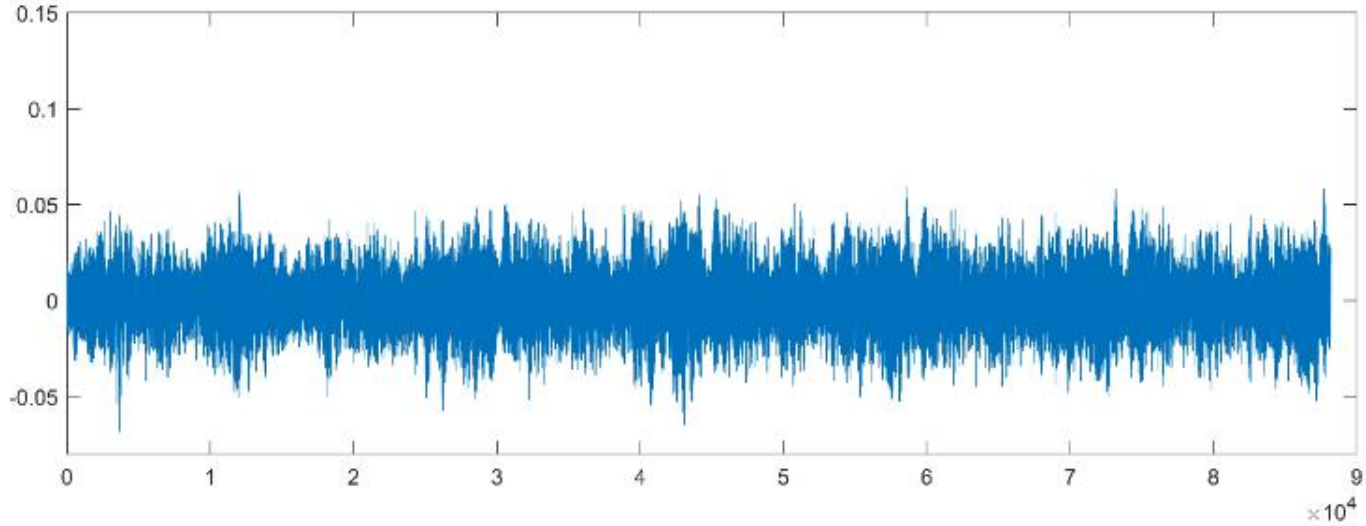

Fig. 5. Sound Signal of Microphone on Right Side

Sound record taken from two different microphones are denoted by two columns. While the first column shows the values of microphone on the left, the second column shows the values of microphone on the right. Total values have been compared with each other by accumulating amplitude values in sound signal of each microphones. It has been understood that sound is close to the microphone whose total amplitude value is higher, that is, sound comes from that direction. The direction of sound which has been detected as warning sound has been detected in this way.

There are no studies using the same data set in the literature. However, the accuracy rates of the studies that make voice recognition by similar methods are shown in Table 3.

5. Conclusion

Firstly, source scanning has been carried out while designing sound recognising system in attempt to enable hearing disabled people to drive in traffic comfortably. After studies carried out, it has been found that there is not a single system exactly the same.

Diffrent sounds and speaker recognising systems have been studied. It has been discovered that sound recognising system in the system is more similiar to multi-sound command recognising system in technical way. By determining positive sounds that is warned for hearing disabled people and negative sound which do

Table 3. Recognition Achievements

Input	***Method***	***Accuracy***
Speaker Sound[6]	Classwise Non-principal Component Analysis (CNPCA)	98%
Speaker Sound[6]	Vector Quantization (VQ)	100%
64-dimensional code book on the TIMIT database [7]	Vector Quantization (VQ)	99,6%
TIMIT database with 10 speakers [4]	Self-Organizing Network Model (SOM)	97,84%
Speaker Sound [10]	Artificial Neural Nets (ANN)	99%
Speaker Sound [9]	SVM	93,9%

not need to be warned, records have been obtained and categorized by help of PC. Sirens sounds of ambulance, police car and fire-fighting vehicles which have priority of way in traffic have been determined as positive sounds. Traffic noise, horn sound, music sounds and other sound that are not among positive sounds have been determined as negative sounds. After taking the sound records into computer environment, they have been put to feature extraction process by digitising them. LPC method which has been used commonly in many studies has been preferred for feature extraction process. After performing feature extraction by LPC approach, greatness of data size draws attention. In this state, dimension reduction process has been regarded as an appropriate approach. Feature extraction process has been applied for dimention reduction. The choosen method for this is PCA which has a common use domain. Data set obtained after putting it to PCA has been extracted by LPC. After putting it to SVM for classification, success rate was 92,6%. KNN and SVM success rates were higher in the classification process without PCA. However, it was decided that the system would be realized by applying PCA process as it shortened the PCA processing time. After classification process, by determining the groups of sounds which have been taken into system, it has been detected whether the sound is among the sounds which belong to those vehicles which have priority of way in traffic.

References

[1] Türkiye Sağlık Araştırması. Türkiye İstatistik Kurumu, Ankara [Online]. Available:http://www.tuik.gov.tr/PreTablo.do?alt_id=1017, Accessed on: Apr. 30, 2019.

[2] Harmancı, S., DHA, Amasya, Türkiye [Online]. Available: http://www.hurriyet.com.tr/isitme-engelli-ogrencilere-isaret-diliyle-trafi-40274838. Accessed on: Apr. 30, 2019.

[3] Ö. Aydın, "Yapay sinir ağlarını kullanarak bir ses tanıma sistemi geliştirilmesi" M.S. thesis, Computer Engineering, Trakya University, Edirne, Turkey, 2005.

[4] M. İnal, "Yapay Sinir Ağları Tabanlı Konuşmacı Tanıma," Ph.D. dissertation, Electric and Computer Education, Kocaeli University, Kocaeli, Turkey, 2001.

[5] W. M. Campbell, K. T. Assaleh, C. C. Broun, "Speaker Recognition with Polynomial Classifiers," 2002.

[6] F. Ertaş, C. Hanilçi, "A New Classifier For Speaker Identification System," TMMOB Elektrik Mühendisleri Odası, Ankara, Turkey, 2007.[Online].Available:http://www.emo.org.tr/ekler/1c207d3fccca3a1_ek.pdf. Accessed on: May 04, 2019.

[7] C. Hanilçi, "Konuşmacı Tanıma Yöntemlerinin Karşılaştırmalı Analizi," M.S. thesis, Electronic Engineering, Uludağ University, Bursa, Turkey, 2007.

[8] S. Başaran, "Yapay Sinir Ağları Kullanarak Konuşmacı Tanıma," M.S. thesis, Electronic Engineering, Uludağ University, Bursa, Turkey, 2007.

[9] W. M. Campbell., J. P. Campbell, D. A. Reynolds, E. Singer, Torres-P. A. Carrasquillo, "Support vector machines for speaker and language recognition," Computer Speech & Language, 20(2-3), 210-229, 2006.

[10] G. Dede and M. H. Sazlı, "Biyometrik Sistemlerin Örüntü Tanıma Perspektifinden İncelenmesi Ve Ses Tanıma Modülü Simülasyonu," EEBM Ulusal Kongresi, 2010.

[11] C. Hanilçi, "Konuşmacı Tanımada Map Uyarlamalı Sınıflandırıcılar," Ph.D. dissertation, Electronic Engineering, Uludağ University, Bursa, Turkey, 2013.

[12] H. Uğuz, A. Arslan, "A new approach based on discrete hidden Markov model using Rocchio algorithm for the diagnosis of the brain diseases," Digital Signal Processing, 3(20), 923-934, 2010.

[13] C. Ünsalan and A. Erçil, "Comparison of Feature Selection Algorithms A New Performance Criteria for Feature Selection," Proceedings of IEEE SIU'98, pp. 60-65, 1998.

[14] C. Cortes and V. Vapnik, "Support-vector networks," Machine learning, 20(3), 273-297, 1995.

[15] O. Kaynar, H. Arslan, Y. Görmez, Y. E. Işık, "Makine Öğrenmesi ve Öznitelik Seçim Yöntemleriyle Saldırı Tespiti," Bilişim Teknolojileri Dergisi, *11*(2), 175-185, 2018.

[16] Ö. Eskidere, "A Comparison of Feature Selection Methods for Diagnosis Of Parkinson's Disease From Vocal Measurements," Sigma, *30*, 402-414, 2012.

[17] M. Karakoyun and M. Hacıbeyoğlu, "Biyomedikal Veri Kümeleri İle Makine Öğrenmesi Siniflandirma Algoritmalarinin İstatistiksel Olarak Karşilaştirilmasi," Dokuz Eylül Üniversitesi Mühendislik Fakültesi Fen ve Mühendislik Dergisi, *16*(48), 30-42, 2014.

[18] İ. F. Pilavcılar, "Metin Madenciliği İle Metin Sınıflandırma," M.S. thesis, Mathematics Engineering, Yıldız Teknik University, İstanbul, Turkey, 2007.

[19] E. Taşcı and A. Onan, "K-en yakın komşu algoritması parametrelerinin sınıflandırma performansı üzerine etkisinin incelenmesi," Akademik Bilişim, 2016.

[20] M. K. Baygün and A. K. Yaldir, "Linear predictive coding ve dynamic time warping teknikleri kullanilarak ses tanima sistemi geliştirilmesi."

[21] K. Yıldız, Y. Çamurcu and B. Doğan, "Veri madenciliğinde temel bileşenler analizi ve Negatifsiz matris çarpanlarına ayırma tekniklerinin karşılaştırmalı analizi," Akademik Bilişim, 10-12, 2010.