Identification of the English Accent Spoken in Different Countries by the k-Nearest Neighbor Method

Yavuz Selim TASPINAR, Murat KOKLU*², Mustafa ALTIN³

Abstract: Sound is the pressure wave created by an object vibrating with a certain frequency. 3 organs are needed for the formation of voice in humans. These are lungs, vocal cords and mouth. Due to the structure of these organs and the similarity of the person with their current language, they can speak another language with different accent. A language can be spoken in different parts of the same country and in different countries. The second most widely used language in the world is English, has numerous accents around the world. In this study, it is aimed to determine which country the English accent spoken in different regions belongs to. In the dataset used, there are 330 sound samples including English accents spoken in Spain, France, Germany, Italy, England and America. Classification has been made with 12 features obtained by Mel Frequency Cepstrum Coefficients feature extraction method. k-Nearest Neighbor (kNN) were used in the classification and 87.2% success was achieved.

Keywords: Accent, region spoken, English, language,

1. Introduction

In order for a language to be understood, speakers need to transfer the words in that language accurately and fluently. In order for the communication between the listener and the speaker to be complete, they must use the same meaning on both sides of the words they speak. However, as in many languages in the world, a word belonging to the same language can have different meanings, spelling and accents according to regions. This situation may cause some disruptions in communication between people in different regions [1]. It can also be said that migration between countries plays an important role in the formation of different accents [2]. Recently, significant progress has been made in speech recognition with the invention of new machine learning techniques. There are various acoustic properties of the voice that gives information about accent, gender, stress, and emotional states [3]. The accent of a language is an acoustic feature that differentiates the speech of that language. People who speak a language other than their native language speak in an accent that is not under their control. Although many factors affect this accent, it is more related to the age at which the foreign language is learned. In a foreign language learned at an early age, accent is less common [4]. The rate of recognizing the accent of a native speaker of English and the accent of a non-native English speaker is different. It is easier to identify the accents of speakers whose mother tongue is different. Studies have been conducted to define the accents of people with the same mother tongue that differ according to regions [5]. There have also been studies using stress in words in order to define the spoken accent of people in different regions with the same mother tongue [6]. In addition to these studies, accent recognition studies were also carried out using the voices of native speakers who have a

different mother tongue and have no similarities with the English language [7]. Since there are many accents in countries with a large population such as India, accent recognition studies have been carried out in Indian [8]. In a study in which telephone conversations were used to define English accents spoken in different regions of England, 14 different accents could be defined [9]. The target audience for accent recognition studies are babies and adults. Among the reasons for this is that babies' accents quickly adapt to the new region within 5 months, and adults change their accent after a long time [10]. Artificial neural networks were used to describe the English accents spoken by people of different origins in Malaysia, and very high results were obtained [11]. This study focuses on defining the English accent spoken in different regions. The study was planned as follows. In Chapter 2, the used data set and kNN method are explained. Test results are given in Chapter 3 and results in Chapter 4.

2. Material and Methods

In this section, the content and features of the dataset used to define the English accent, the kNN method used for classification, the performance metrics required to compare the results of the experiments and measure the performance of the classification method are explained. The steps of extracting features from audio data and classifying using these features are shown in Fig. 1. The pre-processing steps and feature extraction steps on the dataset were previously performed during the dataset creation stage.

¹ *Doganhisar Vocational School, Selcuk University, Konya, TURKEY*
ORCID ID : 0000-0002-7278-4241

² *Technology Faculty, Computer Eng., Selcuk University, Konya, TURKEY*
ORCID ID : 0000-0002-2737-2360

³ *Vocational School of Technical Sciences, Konya Technical University, Konya, TURKEY*
ORCID ID : 0000-0003-3292-4788

* *Corresponding Author Email: ytaspinar@selcuk.edu.tr*

Fig. 1. Identification steps using sound data

2.1. Dataset

The dataset used in the study includes 165 English speech sound data with American accents and 165 voice data of English accents spoken in other regions. A total of 330 audio data were taken from 22 speakers. 11 of these talks are men and 11 of them are women. There is no noise in the background of sound data [12]. Table 1 shows how many data were collected by gender. There are 12 properties in the dataset to be used in classification. These features have been extracted from speech data with the MFCC method before.

Table 1. Speakers demographic data (US: American english accent, Non-US: Non-American accent)

Accent	Gender		
	Female	Male	Total
US	90	75	165
Non-US	90	75	165
Total	180	150	330

Accents other than the American accent were collected from people of Spain, France, Germany, Italy, England descent. This situation makes the accent recognition process difficult. Because each word is voiced differently by people of different origin.

2.2. Performance Metrics

When a model is created to solve a classification problem, the success of the model is represented by the percentage of correct predictions among all the predictions. However, in order to be able to say that a model is a successful model, different calculations and the information resulting from these calculations are needed [13]. The confusion matrix is used to obtain this information. In order to evaluate the performance of the model over this matrix, information can be obtained by making various calculations. Table 2 shows the parameters in the confusion matrix.

Table 2. Confusion matrix

		Predicted	
		Positive	Negative
Actual	Positive	True positive (TP)	False negative (FN)
	Negative	False positive (FP)	True negative (TN)

The model can be evaluated by making calculations with confusion matrix data. Evaluation criteria can be listed as accuracy, error rate, specificity, precision, recall and F1-Measure [14]. The calculation of these values and their contribution to the evaluation are shown in Table 3.

ROC (Reciever Operator Characteristics Curve) curves are used to measure the performance of a classification problem. In the ROC curve, sensitivity is shown on the y-axis, specifity on the x-axis. These values are shown as dots and finally converge, resulting in the ROC curve. The area under the curve is called AUC (Area under the ROC curve) and the AUC value varies between 0 and 1. As this value approaches 1, the predictive value increases, as it approaches 0, the forecast value decreases [15]. Fig. 2 shows the ROC curve and the AUC area.

Table 3. Calculation of model evaluation performance criteria

| Measure	Formula	Declaration
Accuracy	(TP+TN)/Total	Returns the proportion of correctly predicted samples out of all samples
Error Rate	(FP+FN)/Total	Returns the proportion of samples incorrectly estimated from all samples
Recall (r)	TP/(TP+FN)	Returns the correct classification rate of positive samples
Specificity	TN/(TN+FP)	Returns the correct classification rate of negative samples
Precision (p)	TP/(TP+FP)	Returns the ratio of correctly classified positive samples to total positive samples
F1-Score	(2*p*r)/(p+r)	It is the harmonic mean of sensitivity and a value looked for for uneven class distribution.
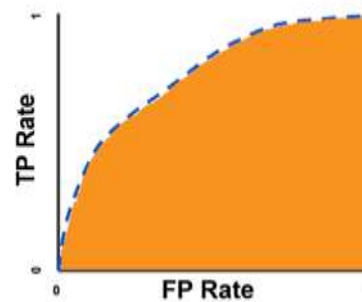
Fig. 2. ROC curve (blue dotted line) and AUC area (Orange region)

2.3. k-Nearest Neighbors

The k-Nearest neighborhood algorithm is a controlled machine learning algorithm. It can be used in regression and classification problems. It was used for classification purposes in this study. The kNN algorithm is frequently used in machine learning problems due to its compatibility with noisy data. However, in classification problems involving big data, it is not preferred because the distance calculation is performed each time, as memory shortage may be encountered [16]. The working steps of the kNN algorithm are shown in Fig. 3.
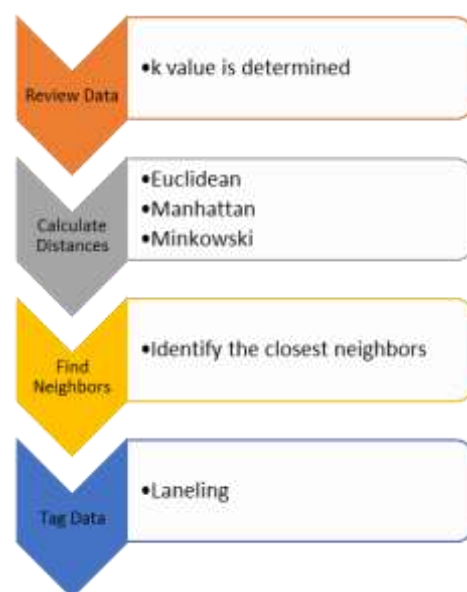

Fig. 3. The operation steps of the kNN algorithm

As can be seen in the process steps in Fig. 3, first the k parameter is determined. This value represents the number of most neighbors of a point. Since the number of classes is 2 in this study, the k value was determined as 2. After this step, the distance of a data to existing data is calculated with distance equations. Euclidean distance function has been used in the study and its formula is shown in equation (1). It is assigned to the nearest k neighbor class according to the distances determined. The specified class becomes the class of the data and the data is tagged.

$$d(a,b) = \sqrt{(x_1 - x_2)^2 + (y_1 - y_2)^2} \tag{1}$$

3. Experımental Results

Experiments have been carried out using a dataset containing features extracted from the audio signal data by MFCC method. The kNN method used for classification has been very successful. The number of neighbors used in the method was determined as 5. The confusion matrix obtained as a result of the classification made on 330 data is shown in Table 4.

Table 4. Confusion matrix obtained by kNN method

		PREDICTED	
		US	NON-US
ACTUAL	US	150	14
	NON-US	28	138

According to the values given in Table 4, the value (TP) 150 in which the American accent is classified as correct, the accent (FN) 14, which is actually the American accent but is classified as incorrect, the value (FP) 28, which is not actually an American accent but is classified as an American accent, is not actually an American accent. The value classified as (TN) was found to be 138. The sum of these values is 330. By using these values, different data about the classification success can be reached. Table 5 shows the classification evaluation data obtained as a result of the calculations.

Table 5. Performance criteria required to evaluate classification success

| Measure	Formula	Results
Accuracy	(TP+TN)/Total	87.3%
Error Rate	(FP+FN)/Total	12.7%
Recall (r)	TP/(TP+FN)	84.1%
Specificity	TN/(TN+FP)	83.1%
Precision (p)	TP/(TP+FP)	84.3%
F1-Score	(2*p*r)/(p+r)	87.7%

The classification success of the kNN method was 87.3%. When evaluating the classification performance, other evaluation criteria should be taken into account. Correct classification value of positive samples is 84.1%, correct classification value of negative samples is 83.1%. The fact that these values are close to each other indicates that the classification has been made successfully.

According to the confisuon matrix data obtained from the kNN method, sensitivity and specifity values appear. Specificity shows the percentage value indicating how accurately the class was predicted by the Sensitivity classifier, and specificity shows the percentage value indicating how accurately the classes that are not included in the same class. Using these values, the ROC curve is drawn. The ROC curve is frequently used to measure the classification performance of methods in datasets containing classes that are not uniformly distributed. Fig. 4 shows the ROC curve for each class resulting from the data obtained from the classification.
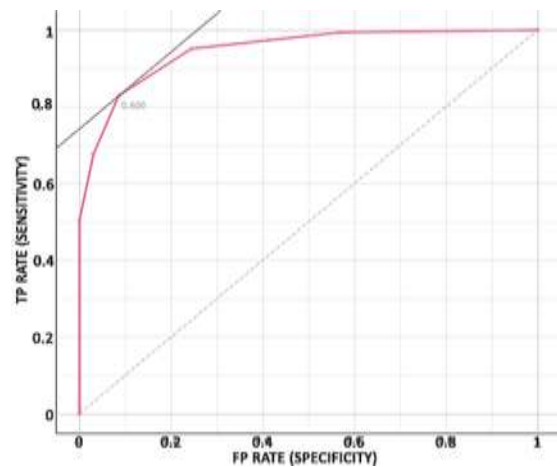
(a)
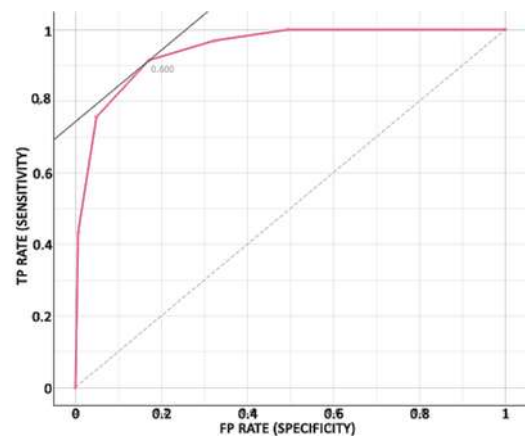
(b)

Fig. 4. ROC curves for US (a) and NON-US (b) class

The area under the ROC curves (AUC) shown in Fig. 4 and the movement of the curves on the coordinate plane contains information about the performance of the classifiers. The more the area under the curve, the higher the classifier's performance. The closer the TP rate (y coordinate) is to 1 and the closer the FP rate (x coordinate) to 0, the higher the performance of the classifier. In this direction, graphics like in Fig. 4 emerge. The classification performance of the kNN method is quite high compared to the graphs in Fig. 4. In Fig. 4 (a), the AUC value for the US class is 0.947, and the AUC value for the NON-US class is 0.947.

4. Conclusions

When the experimental results obtained in this study were examined, the sensitivity of the classification was 83.1% and the recall value was 84.1%. It can be said that the reason why these values are lower than 87.3%, which is the classification success value, is the number of data in the used dataset and the number of features used. When the number of data and the number of features is increased, the success of the classification is expected to increase. When the ROC curves and AUC values are examined, it

can be said that the classifier has learned the training data well. However, due to the low number of data, it is thought that the success was lower than expected due to the low number of both training and test data. In addition, the fact that the sound data in the dataset is obtained from different regions of the accents other than the American accent also affects the success. Collecting audio data for diacritics other than the American accent from the same region is also likely to increase success. In addition to these, different results can be obtained if a different method is used in the feature extraction process from audio data. Different classification methods may also produce different results in the dataset used. It is planned to use different classification methods and to increase the classification success in future studies.

References

[1]   Cai, Z.G., et al., Accent modulates access to word meaning: Evidence for a speaker-model account of spoken word recognition. Cognitive Psychology, 2017. **98**: p. 73-101.

[2]   Dunton, J., C. Bruce, and C. Newton, Investigating the impact of unfamiliar speaker accent on auditorycomprehension in adults with aphasia. International Journal of Language & Communication Disorders, 2015: p. 1-11.

[3]   Ikeno, A. and J.H. Hansen, The effect of listener accent background on accent perception and comprehension. EURASIP Journal on Audio, Speech, and Music Processing, 2007. **2007**(1): p. 076030.

[4]   Asher, J.J. and R. García, The optimal age to learn a foreign language. The Modern Language Journal, 1969. **53**(5): p. 334-341.

[5]   Gupta, V. and P. Mermelstein, Effects of speaker accent on the performance of a speaker-independent, isolated-word recognizer. The Journal of the Acoustical Society of America, 1982. **71**(6): p. 1581-1587.

[6]   Pedersen, C. and J. Diederich. Accent classification using support vector machines. in 6th IEEE/ACIS International Conference on Computer and Information Science (ICIS 2007). 2007. IEEE.

[7]   Gaikwad, S., B. Gawali, and K. Kale, Accent recognition for indian english using acoustic feature approach. International Journal of Computer Applications, 2013. **63**(7).

[8]   Mannepalli, K., P.N. Sastry, and M. Suman, MFCC-GMM based accent recognition system for Telugu speech signals. International Journal of Speech Technology, 2016. **19**(1): p. 87-93.

[9]   Hanani, A., M.J. Russell, and M.J. Carey, Human and computer recognition of regional accents and ethnic groups from British English speech. Computer Speech & Language, 2013. **27**(1): p. 59-74.

[10]   Nazzi, T., P.W. Jusczyk, and E.K. Johnson, Language discrimination by English-learning 5-month-olds: Effects of rhythm and familiarity. Journal of Memory and Language, 2000. **43**(1): p. 1-19.

[11]   Ma, Y., et al. Speaker accent recognition through statistical descriptors of Mel-bands spectral energy and neural network model. in 2012 IEEE Conference on Sustainable Utilization and Development in Engineering and Technology (STUDENT). 2012. IEEE.

[12]   Ma, Z. and E. Fokoué, A comparison of classifiers in performing speaker accent recognition using MFCCs. arXiv preprint arXiv:1501.07866, 2015.

[13]   Koklu, M. and I.A. Ozkan, Multiclass classification of dry beans using computer vision and machine learning techniques. Computers and Electronics in Agriculture, 2020. **174**: p. 105507.

[14]   Hossin, M. and M. Sulaiman, A review on evaluation metrics for data classification evaluations. International Journal of Data Mining & Knowledge Management Process, 2015. **5**(2): p. 1.

[15]   Kannan, R. and V. Vasanthi, Machine learning algorithms with ROC curve for predicting and diagnosing the heart disease, in Soft Computing and Medical Bioinformatics. 2019, Springer. p. 63-72.

[16]   Liao, S., et al., Multi-object intergroup gesture recognition combined with fusion feature and KNN algorithm. Journal of Intelligent & Fuzzy Systems, 2020(Preprint): p. 1-11.