Kurdish Spoken Letter Recognition Based on K-NN and SVM Model

Zrar Khalid Abdul

Department of ApplyingComputer , College of Medicals and Applied Science ,Charmo University, Chamchamal, Kurdistan Region, Iraq.

E-mail: zrar.abdul@charmouniversity.org

Abstract:

Automatic recognition of spoken letters is one of the most challenging tasks in the area of speech recognition system. In this paper, different machine learning approaches are used to classify the Kurdish alphabets such as SVM and k-NN where both approaches are fed by two different features, Linear Predictive Coding (LPC) and Mel Frequency Cepstral Coefficients (MFCCs). Moreover, the features are combined together to learn the classifiers. The experiments are evaluated on the dataset that are collected by the authors as there as not standard Kurdish dataset. The dataset consists of 2720 samples as a total. The results show that the MFCC features outperforms the LPC features as the MFCCs have more relative information of vocal track. Furthermore, fusion of the features (MFCC and LPC) is not capable to improve the classification rate significantly.

Keywords: K-Nearest Neighbor, Support Vector Machine, Confusion Matrix, LPC, MFCC.

1. Introduction:

Kurdish is an Indo-European language and it is one of the oldest languages in the world. Currently, there are approximately 60 million individuals employed Kurdish as native speaker. The language consists of four main dialects which are (Sorani, Zaza, Hawrami and Kurmanji)(Sheyholislami, 2011,p.47). Almost all Kurdish alphabets are used in several languages such as Arabic, Persian and Urdu. Basically, standard Kurdish language consists of 34 phonemes, of which are seven vowels, and twenty seven are consonants(McCarus, 2013,p.663).

In the last three decays, more attention has focused on creating a model to build speech recognition for most of the languages especially English(Amodei et al., 2016,p. 173)(Saon et al., 2017,p.132). However, for the Kurdish language, the efforts of researchers have just begun to build the requirement of proper Kurdish speech recognition for example in In (Omer, 2019,p.78) authors was developed a model to Kurdish digit recognition and Kurdish dialect recognition Was studied in () as it might be helpful of Kurdish speech recognition . The vocal track characteristics and behavior of each uttered letters signal is a dealt with in any speech recognition system because of the spoken alphabets for various languages were studied by automatic speech recognition researchers(Cole et al., 1990,p.122)(SAYEM, 2014,p.88)(Meyer et al., 2011,p.129).

Early work, in(Jan, 1943,p.2), matching technique was studied . The model used template matching with dynamic time warping to manipulate frame by frame matching of an input utterance to basis templates. The speaker-dependent recognition rates as obtained almost 60% on the alphabet recognition. A speaker-independent spoken English alphabet recognition system was designed by Cole et al(Cole et al., 1990,p.122) .The proposed model, which was neural network model, trained on one token of each letter from 120 speakers. Performance was 95% when it tested on a new set of 30 speakers and was increased to 96% when it tested on a second token of each letter from the original 120 speakers.

In the mid-1980's, a biggest enhancement were achieved in speech recognition system by (Vlontzos and Kung, 1989,p.1729), which used the Hidden Markov models (HMM) that can deal with sequence of data. For that time, HMM is become a well-known and widely used statistical method for describing the spectral features of speech frame. In (Alotaibi et al., 2010,p.122) and (Cited and Hamdani, 2018,p.3) HMM proposed with some stochastic features and their results illuminated that HMM can be beneficial way of recognizing speech for a wide range of applications.

Arabic researchers also used HMM for Arabic language, for instance, in(Cole et al., 1990,p.122) developed HMM for isolate utterance Arabic letter and the result achieved 64.06% overall correct alphabets recognition using mixed training. Other machine learning approaches have been used in the speech analysis area. For example, English Alphabetic Recognition (EAR) system has been developed in (B.Adam and Salam, 2012,p.21) which had the highest accuracy rate achieved for speaker-dependent witch was 79.17%.

The biggest challenge in this field is how to extract efficient feature from the speech signal. Consequently, different techniques are used by researchers to extract features from the data. Most common techniques are Mel Frequency Cepstral Coefficients (MFCCs), Linear Predictive Coding (LPC) and Formant frequency extraction(Dhameliya and Bhatt, 2015,p.1)(Hibare and Vibhute, 2014,p.1).

Unfortunately, Kurdish speech analysis is not being interested yet by many researchers. However, there are some works can be seen, for instance, a model was developed for identifying Kurdish dialect and neighboring language by using one dimensional local binary pattern. Their result showed that the proposed model improves the accuracy

of classification (Al-Talabani et al., 2017,p.20). Formant and LPC features were used in (Dinler and Karabiber, 2017,p.12) to analyze vowel letters in Kurdish language. The result shows that the model can be beneficial for the Kurdish vowel letters and acoustic property of the language. In (Omer, 2019,p.78) reported fusion of some features like LPC, MFCC and Formant can lead to classify the digit Kurdish letter from (0-9). Convolutional neural network was fed by MFCC frames for Kurdish speaker identification(Abdul, 2019,p.566).

In this paper, we study the capability of MFCC and LPC separately to extract features, then, both of them are concatenated to get extent feature. After that k-Nearest Neighbors (k-NN) and Support Vector Machine (SVM) are used to classify the extracted feature MFCC, LPC and their fusion. The performance of SVM and KNN are compared to the performance of Artificial Neural Network (ANN) as it was adopted by (B.Adam and Salam, 2012,p21). The proposed model can be helpful for building speech recognition system as it is not built yet properly and we do not aware of any work that has done for this purpose in Kurdish language.

The rest of this paper is organized as follows: section two contains the data description, then section three provides methodology of the model, and section four contains details about result and discussion, and finally section five presents the conclusion and future work.

2. Data Description

As there was not available data, the authors' collected dataset with considering all standard rules such as statistical distribution of region, age, gender. In addition, the dataset was designed to be rich in terms of the number of the speech samples. Therefore, 40 people are involved and said each Kurdish letter twice in different environment. As a result, there are 80 samples per letter. As we have 34 Kurdish letters. So, we have 2720 samples as a total. Each wave signal is recorded in 0.5 second. Overall length is 1,360 ms.

3. Methodology

A complete proposed model based on feature-based was developed to carry out the goals of this research. The model is divided into two modules according to classification methods SVM and KNN. Each of the method is fed by two common features which are LPC and MFCC as shown in the Figure (1). The feature extraction process is transforming the raw signal into most effective information in the vector form.

Linear predictive coding is the one of the powerful and useful techniques used in speech processing for extracting the spectral envelope of speech in compacted form which represent the information of a linear predictive model of the signal(Dhameliya and Bhatt, 2015,p.1)(Bäckström, 2017,p.4). In this paper, 60 statistics values are extracted from 12 coefficients of each frame of the signal and used for the training stage for both SVM and k-NN.

MFCC is another powerful technique in audio signal processing, which are commonly used for speaker identification as it transforms the raw signal into most relative information for vocal tract of the speech(Hibare and Vibhute, 2014,p.1)(Fernandes et al., 2018,p.200). Both classifiers approaches are fed by 60 coefficients of the MFFC. Praat software was used for isolating the speech into clear letter and extracting features as well.

K-nearest neighbors is a non-parametric classification and non-linear technique which based on the distance. There is some advantage of using k-NN which takes low execution time and it is easy for interpretation(Jadhav and Channe, 2016,p.1842). Choosing value of k is very essential and it affects the accuracy result. So, in general, value

of k is between 3 and 10. However, in this paper, the k value is evaluated to the optimum value which is 5 as a range of k is applied from 1-10. The best result is obtained when k is 5.

SVM is linear classification method which has the objective to find the "best" hyperplane in dimensional space. The hyper-plane here, is the boundary separator of the two classes (Nayak et al., 2015,p.169). In this pairwise SVM is used to support multiclass problem as there are more than two classes.

The models have been implemented on five different set of labels. First, classification is done among 7 vowel letters. Second, the proposed models are implemented by 27 consonant letters in order to classify only the consonant letters. The third model is isolated two classes which are vowel and consonant. The fourth model is classified all the letters together so in this case, there are 34 labels as Kurdish letters are 34 letters. the final model is classified E_set letter which consist B, C, D, E, G, P, T, V and Z and the result of this model is compared with (B.Adam and Salam, 2012,p21). 10-fold cross-validations is used to evaluate the performance of both machine learning approaches.
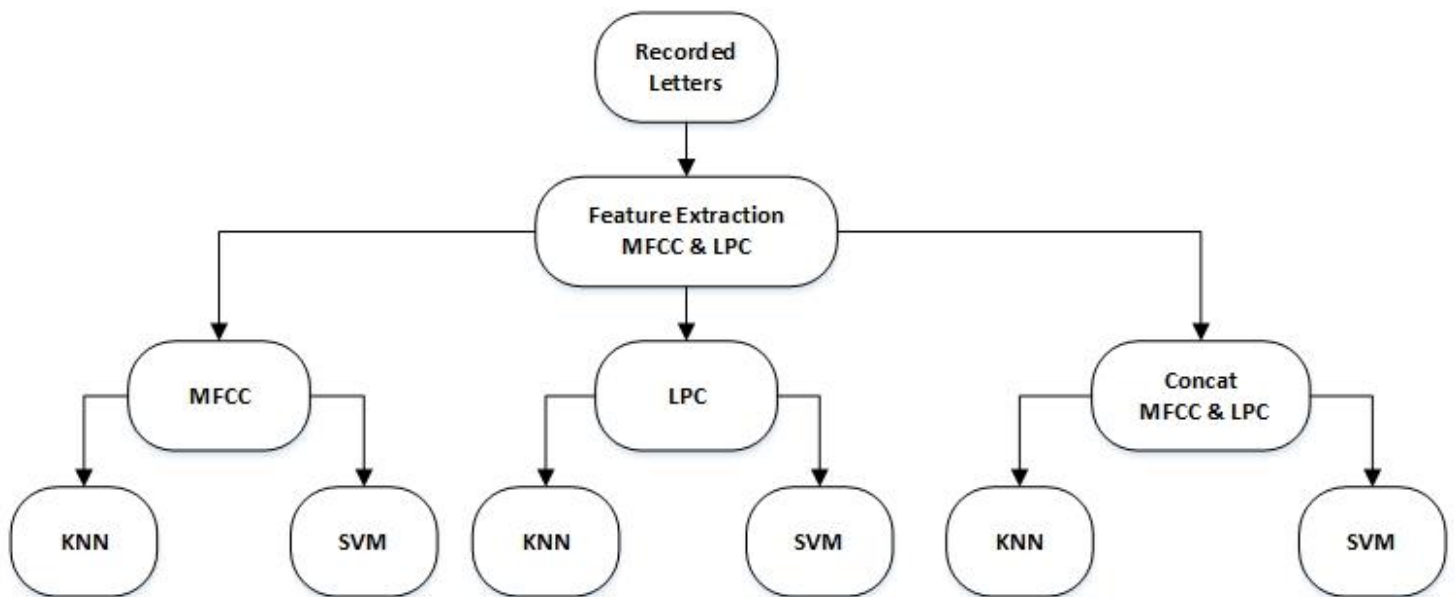

Figure 1 Classification Procedure

4. Result and discussion

We conducted the models in three different experiments and obtained three confusion matrices. The k-NN and SVM are trained and tested by MFCC, LPC and the combined of these two features (MFCC and LPC) named as fusion of the features, the total experiments are 6 experiments per case MFCC-KNN,MFCC-SVM,LPC-KNN,LPC-SVM, fusion feature-KNN and fusion feature-SVM. In order to organize the cases, we divided the process into three classes:

- 7 vowel letters
- 27 consonant letters
- Two classes of letters (vowel and consonant)
- 34 letters which includes consonants and vowel letters title
- E_set letters

4

4.1 Vowel letters (7 letters):

Kurdish letters include seven vowel letters which are ئە - و - ا- وو- ۆ -ی-ئ . The result of the experimental is shown in the Table (1). MFCC features represent high classification rate compare to LPC. For instance, SVM-MFCC and KNN-MFCC is outperform to the SVM-LPC and k-NN-LPC by almost 25 percentage of accuracy rate which is significant. However, there is not significant change in the success rate for fusion features compare to the MFCC features as it seems that MFCC obtain those information that the LPC carry it out.

Table 1 Accuracy Result for Vowel Letters

Method	MFCC	LPC	MFCC&LPC
k-NN	72.5	40.5	73.7
SVM	73.6	50.6	77.6

By focusing on the confusion matrix, deeper analysis can be made, for example, Table (2) shows the success rate and misclassification between them. There is also a huge misclassification rate between وو and و which is almost 50%. It means that there are more similarities between both letters in term of their features.

Table 2 Confusion Matrix of SVM Classifier for Vowel Letters using MFCC
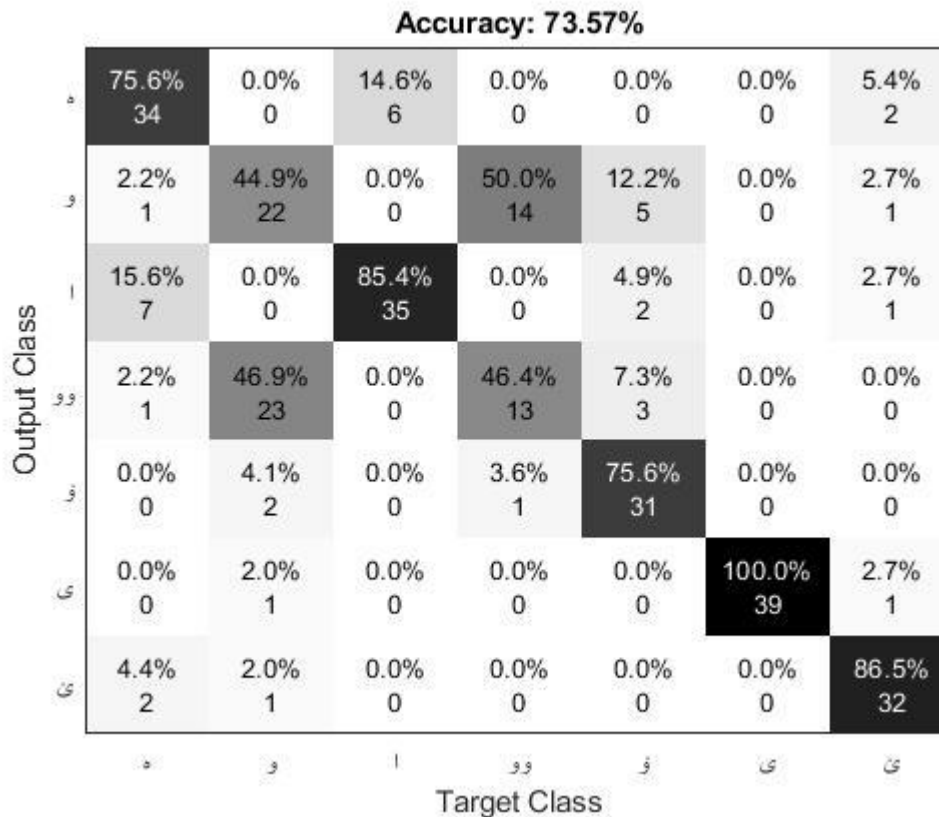
4.2 Consonant Letters (27 letters):

In this case, 27 classes are considered because there are 27 constant letters. Based on the result that shown in the Table (3). It is straight forward to know that the SVM with MFCC performs quite well compare to the rest of the experiments with p-value 0.046. However, the classification rate is decreased significantly compare to classification rate of classifying vowel letters. One of the main reasons here is difference in the number of classes,7 and 27 class of vowel and constant letter classification respectively.

Table 3 Accuracy Result of Consonant Letters

Model	LPC	MFCC	MFCC&LPC
KNN	86	95	95
SVM	86	93	93

Method	MFCC	LPC	MFCC&LPC
k-NN	43.8	10.6	43.8
SVM	53.6	26.5	55.5

4.3 Vowels vs. Consonant letters

As mentioned in previous sections, Kurdish letters include 7 vowel letters and 27 consonant letters. A model is implemented to classify vowel and consonant letters. Table (4) illustrates that there is not significant change between the performance of KNN and SVM. Moreover, combining MFCC and LPC features are not improved the accuracy of both classification methods.

Table 4: Accuracy of the model for Vowel and Consonant Letters

6

4.4 All Kurdish letters (34 letters):

Kurdish alphabet consists of 34 letters which are distributed into two major parts, consonants and vowels letters. All letters are presented in the below:

ئ ـ ی ـ وو ـ ۆ ـ و ـ ە ـ ھ ـ ن ـ م ـ ڵ ـ ل ـ گ ـ ک ـ ق ـ ڤ ـ ف ـ غ ـ ع ـ ش ـ س ـ ژ ـ ز ـ ڕ ـ ر ـ د ـ خ ـ ح ـ چ ـ ج ـ ت ـ پ ـ ب ـ ا ـ ئـ

Table (5) shows that MFCC features are most effective feature for identify the letter. MFFC with both classifiers outperforms the LPC based models (p-value =0.049. Moreover, SVM with MFCC feature achieve high accuracy compare to the KNN with MFCC. Furthermore, the combination of the both MFCC and LPC do not increase the classification rate significantly as the p-value=0.24. However can change the accuracy slightly compare to MFCC features by using KNN and SVM.

Table 5 classification rate of all letters

Method	MFCC	LPC	MFCC&LPC
KNN	45.7	12.9	46.9
SVM	55.6	29	56.9

Overall, there are similarities between vowel letters as well as consonant letters that lead to have misclassification between some letters, for example, 33.3% of م is classified as ن letter. م is also recognize 12% of ل as م letter. Table (6) illustrates similarities percentage between some letters.

Table 6 Misclassification of letters

Target	Output	Percentage
ە	پ	17
گ	ڤ	13
ف	ش	14.7
ع	ئ	14.7
چ	ش	12.2
د	گ	23.8
ە	ف	12.5
ع	ق	11.8
ڤ	ل	22.8

4.5 E-Set

The E-set letter includes 9 letters which are B, C, D, E, G, P, T, V, Z. Beside the implementing SVM and KNN to classify the E-set, a previous adopted approach is re-employed which is Artificial Neural Network (ANN) and affair comparison are shown in the Table (7). The result illustrates that the SVM classifier has outperformance of the KNN and adopted ANN. However, the accuracy of ANN is higher than the KNN.

Table 7 classification rate for E-set

Model	MFCC
ANN (B.Adam and Salam, 2012,p21)	72
KNN	54
SVM	77

5. Conclusion

In this paper, a spoken Kurdish alphabets model is developed to investigate the process of automatic alphabets recognition based on their speech. Two global features are extracted from the raw signal, MFCC and LPC. The features are used to train classifiers, k-NN and SVM. Based on our experimental, SVM outperforms k-NN classifier to classify the Kurdish spoken independent alphabets. Moreover, most of the misclassification data occur between consonant letters. In term of features, the results show that the MFCC features have more efficient information for spoken alphabets than the LPC, as the MFCC features is almost represent the vocal track influence when uttering each the letters. The work can be extended to improve the classification rate using sequence classification methods such that hidden Markov model and long short term memory.

التعرف على الحروف الكردية المنطوقة على أساس نموذج SVMو NN-k

زرار خالد عبدول

قسم علوم احلاسوب التطبيقية، كلية الطب والعلوم التطبيقية، جامعة جارمو ، مججمال ، إقليم كوردستان ، العراق.

البريد الإلكتروني :　　　zrar.abdul@charmouniversity.org

خلاصة:

يعد التعرف التلقائي على الحروف المنطوقة أحد أكثر المهام صعوبة في مجال نظام التعرف على الكلام. في هذا البحث ، يتم استخدام مناهج مختلفة للتعلم الآلي لتصنيف الأبجديات الكردية مثل SVM و k-NN حيث يتم تغذية كلا الأسلوبين بميزتين مختلفتين ، الترميز التنبئي الخطي (LPC) ومعاملات Cepstral للتردد الخطي (.MFCCs) علاوة على ذلك ، يتم دمج الميزات معًا لتعلم المصنفات. يتم تقييم التجارب على مجموعة البيانات التي تم جمعها من قبل المؤلفين حيث لا توجد مجموعة بيانات كردية قياسية. تتكون مجموعة البيانات من 2720 عينة في المجموع. تظهر النتائج أن ميزات MFCC تتفوق في الأداء على ميزات LPC نظرًا لأن MFCCs لديها معلومات أكثر نسبيًا عن المسار الصوتي. علاوة على ذلك ، فإن دمج الميزات MFCC) و (LPC غير قادر على تحسين معدل التصنيف بشكل كبير.

الكلمة الرئيسية : k- أقرب الجار ، دعم آلة المتجهات ، مصفوفة الارتباك ، LPC، MFCC.

Reference:

Abdul, Z.K., 2019. Kurdish speaker identification based on one dimensional convolutional neural network-Computational Methods for Differential Equations, , University of Tabriz,7, 566–572.

Al-Talabani, A., Abdul, Z., Ameen, A., 2017. Kurdish Dialects and Neighbor Languages Automatic Recognition. ARO-The Sci. J. Koya Univ. 5, 20–23.

Alotaibi, Y.A., Alghamdi, M., Alotaiby, F., 2010. Speech recognition system of Arabic alphabet based on a telephony Arabic corpus. Lect. Notes Comput. Sci. (including Subser. Lect. Notes Artif. Intell. Lect. Notes Bioinformatics) 6134 LNCS, 122–129.

Amodei, D., Ananthanarayanan, S., Anubhai, R., Bai, J., Battenberg, E., Case, C., Casper, J., Catanzaro, B., Cheng, Q., Chen, G., others, 2016. Deep speech 2: End-to-end speech recognition in english and mandarin. In: International Conference on Machine Learning. pp. 173–182.

B.Adam, T., Salam, M., 2012. Spoken English Alphabet Recognition with Mel Frequency Cepstral Coefficients and Back Propagation Neural Networks. Int. J. Comput. Appl. 42, 21–27.

Bäckström, T., 2017. Speech Coding: with Code-Excited Linear Prediction. Springer.

Cited, R., Hamdani, M., 2018. Discrete hidden markov model basis for arabic handwriting recognition 1.

Cole, R.A., Muthusamy, Y., Tutor, M.S., 1990. Speaker-independent recognition of spoken English letters Speaker-Independent Recognition of Spoken English Letters.

Dhameliya, K., Bhatt, N., 2015. Feature extraction and classification techniques for speaker recognition: A review. Int. Conf. Electr. Electron. Signals, Commun. Optim. EESCO 2015 3, 1–5.

Dinler, O.B., Karabiber, F., 2017. Formant analysis of vowels in Kurdish language. In: 2017 25th Signal Processing and Communications Applications Conference (SIU). IEEE, pp. 1–4.

Fernandes, V., Mascarehnas, L., Mendonca, C., Johnson, A., Mishra, R., 2018. Speech Emotion Recognition using Mel Frequency Cepstral Coefficient and SVM Classifier. In: 2018 International Conference on System Modeling & Advancement in Research Trends (SMART). pp. 200–204.

Hibare, R., Vibhute, A., 2014. Feature Extraction Techniques in Speech Processing: A Survey. Int. J. Comput. Appl. 107, 1–8.

Jadhav, S.D., Channe, H.P., 2016. Comparative Study of K-NN, Naive Bayes and Decision Tree Classification Techniques. Int. J. Sci. Res. 5, 1842–1845.

Jan, P., 1943. ' UNITED .sTAT-Es PATENT - 12, 2–4.

McCarus, E.N., 2013. Kurdish. In: The Iranian Languages. Routledge, pp. 663–709.

Meyer, B.T., Brand, T., Kollmeier, B., 2011. Effect of speech-intrinsic variations on human and automatic recognition of spoken phonemes. J. Acoust. Soc. Am. 129, 388–403.

Nayak, J., Naik, B., Behera, H.S., 2015. A Comprehensive Survey on Support Vector Machine in Data Mining Tasks: Applications & Challenges. Int. J. Database Theory Appl. 8, 169–186.

Omer, S.M., 2019. Uttered Kurdish digit recognition system 6, 78–85.

Saon, G., Kurata, G., Sercu, T., Audhkhasi, K., Thomas, S., Dimitriadis, D., Cui, X., Ramabhadran, B., Picheny, M., Lim, L.-L., others, 2017. English conversational telephone speech recognition by humans and machines. arXiv Prepr. arXiv1703.02136.

SAYEM, A., 2014. Speech Analysis for Alphabets in Bangla Language: Automatic Speech Recognition. Int. J. Eng. Res. 3, 88–93.

Sheyholislami, J., 2011. Kurdish Identity. In: Kurdish Identity, Discourse, and New Media. Springer, pp. 47–77.

Vlontzos, J.A., Kung, S.-Y., 1989. Hidden Markov models for character recognition. In: International Conference on Acoustics, Speech, and Signal Processing,. pp. 1719–1722.