Avrupa Bilim ve Teknoloji Dergisi
Özel Sayı 43, S. 36-40, Kasım 2022

Araştırma Makalesi

www.ejosat.com ISSN:2148-2683

European Journal of Science and Technology
Special Issue 43, pp. 36-40, November 2022

Research Article

Comparison of the Effect of MFCC and GTCC Features on Determining the Ideal Recording Time for Body Sound Location Identification

Osman Balli

Department of Computer Engineering, Malatya Turgut Ozal University, Turkey, ORCID: 0000-0001-6757-3717), osman.balli@ozal.edu.tr

(4th International Conference on Applied Engineering and Natural Sciences ICAENS 2022, November 10 - 13, 2022)

(**DOI:** 10.31590/ejosat.1202062)

ATIF/REFERENCE: Balli, O. (2022). Comparison of the Effect of MFCC and GTCC Features on Determining the Ideal Recording Time for Body Sound Location Identification, *European Journal of Science and Technology*, (43), 36-40.

Abstract

One of the fastest ways to get information about the state of the human body is to analyse body sounds. The ability to transfer sounds to a digital medium facilitates this analysis. In this study, zone detection was performed from the sound data obtained from the heart, lung, and abdominal regions. 20s data with a sampling frequency of 4000 from 12 men were used in the training. The data was analysed in 9 different seconds. All data for each second is divided and prepared for training. Features were extracted using MFCC and GTCC and these features were trained in CNN model. The effect of MFCC and GTCC coefficients on the results was compared. In training, the best result was obtained from the MFCC coefficient obtained from 1.5-second recordings with 98%, and in validations, the best result was obtained from MFCC coefficients of 1-second recordings with 85%. Looking at the general validation results, it was seen that the MFCC results were more successful.

Keywords: Artificial intelligence, body sounds, convolution nneural network, gammatone cepstral coefficients, mel-frequency cestrum coefficients

Vücut Seslerinden Bölge Tanımlanması için İdeal Kayıt Süresinin Belirlenmesinde MFCC ve GTCC Özniteliklerinin Etkisinin Karşılaştırılması

Öz

İnsan vücudunun durumu hakkında bilgi almak için yapılabilecek en hızlı yöntemlerden birisi vücut seslerini analiz etmektir. Seslerin dijital ortama aktarılabilmesi bu analizi kolaylaştırmaktadır. Bu çalışmada kalp, akciğer ve karın bölgelerinden alınan ses verilerinden bölge tespiti yapılmıştır. Eğitimde 12 kişiden alınan 4000 örnekleme frekansına sahip 20s lik veriler kullanılmıştır. Veriler 9 farklı saniyede incelenmiştir. Her bir saniye için tüm veriler bölünmüş ve eğitim için hazırlanmıştır. MFCC ve GTCC kullanılarak öznitelikler çıkarılmış ve bu öznitelikler CNN modelinde eğitilmiştir. MFCC ve GTCC katsayılarının sonuçlar üzerindeki etkisi kıyaslanmıştır. Eğitimde en iyi sonuç %98 ile 1,5 saniyelik kayıtlardan alınan MFCC katsayısından, validationlarda ise en iyi sonuç %85 ile 1 saniyelik kayıtların MFCC katsayılarından elde edilmiştir. Genel validation sonuçlarına bakıldığında MFCC sonuçlarının daha başarılı olduğu görülmüştür.

Anahtar Kelimeler: Yapay zeka, vücut sesleri, evrişimli sinir ağı, gammatone kepstral katsayısı, mel-frekansı kepstral katsayıları

1. Introduction

The process of listening to body sounds is called "auscultation". Through the process of auscultation, doctors can obtain general information about human health. The most important part of this procedure is the stethoscope. Since the invention of the stethoscope in 1816, it has shown a lot of progress (Tamas et al., 2016). Especially the developments in the technological field have also affected the health field, and electronic stethoscopes have started to be used more often. Body sound data obtained thanks to these stethoscopes can be stored and analysed electronically.

The most basic material required for the use of artificial intelligence is “data”. Artificial intelligence, which has begun to play an important role in the field of health, can diagnose many diseases such as COPD and asthma. At the same time, with the use of artificial intelligence methods, early diagnosis, which is an important parameter for treatment, can be made more accurately and quickly. Artificial intelligence needs to work with meaningful features to give better results. MFCC and GTCC are generally used to generate meaningful coefficients from audio data. In recent years, there have been several publications on lung sound recognition using MFCC (Pittner and Kamarthi, 1999; Mayorga et al., 2010; Bahoura and Pelletier, 2004; Bahoura and Ezzaidi, 2013). Looking at the literature, Aziz et al first pre-processed auscultatory signals through Empirical Mode Decomposition (EMD), which separates the original signal into constituent components known as intrinsic mode functions (IMFs).

IMFs carrying redundant and noisy data are rejected and thus the pre-processing has become more efficient. After pre-processing, MFCC is applied with the obtained coefficients. Support Vector Machines (SVM) classifiers are trained and tested through 5-fold cross validation. It is performed on various classifiers on a self-collected dataset containing 480 auscultation signals from normal and pneumonic subjects. different classifiers to classify pneumonia and normal subjects SVM-linear (SVM-L), SVM-Quadratic (SVM-Q), Ensemble–Boosted trees (En-B. T), Ensemble-KNN (En-KNN) and K-nearest neighbor-coarse (KNN-C) for performance evaluation to extract the highest accuracy. The highest success was achieved in SVM-Quadratic with 99.7% (Aziz et al., 2019). In their study to measure asthma severity, Shaharum et al obtained the best performance using MFCC features using the KNN classifier. As a result of training with a total of 250 breathing data from 50 men, the highest result is 95.92%, 96.33% and 98.42% average accuracy, sensitivity, and specificity (Shaharum et al., 2019). Cheng et al. collected the sleep sounds of 33 patients and 10 normal men in their study for the diagnosis of obstructive sleep apnea hypopnea syndrome. Then Mel-frequency cestrum coefficients (MFCC), Mel Filter Banks (Fbanks), Short-Term Energy and Linear Prediction Coefficient (LPC), characteristic features of snoring, which represent different characteristics of snoring, were extracted. To identify snoring and synthesize information, an input model based on LSTM, which can capture various audio features, is designed. In the experiment, snoring related to the respiratory event and normal snoring were classified with 95.3% accuracy (cheng., 2022). In Lella and Pja's study on the diagnosis of Covid 19, the features were obtained with De-noising Auto Encoder (DAE) technique, GFCC (Gamma-tone Frequency Cepstral Coefficients), and IMFCC (Improved Multi-frequency Cepstral Coefficients) methods, and CNN trained in the model. With the data set consisting of Breath, Cough, Sample Voice data, it achieved the highest efficiency with a value of 95.45% (Lella and Pja, 2022). Dar et al designed an effective Covid-19 detection model using designed JHBO-based DNFN. Gaussian Filter was applied to remove the noises and then feature extraction was performed with MFCC. accuracy, sensitivity, and specificity of 0.9176, 0.9218 and 0.9219 were obtained (Dar et al., 2022). Bardou et al., in their work on the classification of breathing sounds, determined the features to be used for classification as MFCC, spectrum and LBP (local binary pattern). It achieved &91.12 in training with MFCC features and SVM classifier. Using the same features and the CNN classifier, 91.67% success was achieved (Bordou et al., 2018). In another study conducted with the MFCC and CNN model, 93% success was achieved in the respiratory dataset containing 6 classes (COPD (Chronic Pulmonary Obstructive Disease), Healthy, URTI (Upper Respiratory Tract Infection), Bronchiectasis, Pneumonia, Bronchiolitis.) (Mridha et al., 2021). In the work done by Jayalakshmy and Sudha, first, feature vectors of lung sounds were extracted with intrinsic mode function (IMF). In the next step, Gammatone filters were applied to the best combination IMF features and Gammatone cepstral coefficients (GTCC) were calculated. results show that the proposed GTCC of the third IMF component applied to the clustered BiLSTM framework outperforms the competing Convolutional Neural Network classification method in terms of accuracy, specificity, and sensitivity (Jayalakshmy and Sudha, 2021). In Kutlu and Karaca's study, feature selection was done using Mel frequency cepstral coefficients (MFCC) and gammatone cepstral coefficients (GTCC) methods. With the obtained results, the classification process was not done with the traditional CNN model. The results obtained are between 82.9% and 85.83%. The results obtained using the MFCC method were determined to be more successful (Kutlu and Karaca, 2022).

In this study, a comparison of MFCC and GTCC feature extraction methods on regional sound detection was conducted. Correct labelling of data, which is one of the most important materials for artificial intelligence, is very important for training. The fact that labelling is done by people and the importance of experience makes this study important. A model design that can prevent mislabelling has been considered.

2. Material and Method

2.1. Database

Breathing sounds from 12 men and 20 regions were used in this study. These sounds include 4 abdominal regions, 4 cardiac regions and 12 lung regions. 20s sounds with a sampling frequency of 4000 Hz were taken from each region. The sounds are divided into intervals of 1,1.5,2,2.5,3,2.5,4,4.5s and characterized by MFCC and GTCC. The default features of the MFCC and GTCC functions of the Scipy library of Python were used to create the feature. For training, the same number of random lung data as Abdomen and heart regions were obtained, and training was performed with the same amount of data for each region.

2.2.Mel Frequency Spectral Coefficient (MFCC)

The three main methods used when characterizing breathing sounds are: statistics in the time-frequency domain, wavelet coefficients, and cepstrum coefficients (Meng et al., 2020). When the literature is examined, it can be seen that MFCC gives good results in sound data in general (Kutlu and Karaca, 2022). MFCC relies on Mel frequency filtering and cepstrum analysis to extract features of the audio signal spectrum. MFCC is pre-emphasized and framed first. In the framing process, the signal is divided into short pieces. The length of split signals is about 20-40 ms. A windowing method is used to avoid discontinuity between signals. The FFT is then applied to each short-term analysis window. The mel function is used to create the Mel filter bank. Discrete Cosine Transform (DCT) is applied to the resulting frames and delta coefficients are obtained (Winursito et al., 2018). The number of features of each data as a result of MFCC is given in Table 1.

Table 1. Count of MFCC Coefficient obtained by seconds

Time (s)	Count of Coefficient
1	1287
1,5	1937
2	2587
2,5	3237
3	3887
3,5	4537
4	5187
4,5	5837
5	6487

2.2. Gammatone cepstral coefficients (GTCC)

Gammatone Cepstral Coefficients (GTCCs) have continued to be effective even though MFCCs have gained importance in audio data recognition in the last few years (Jayalakshmy and Sudha, 2021). GTCC feature extraction is similar to MFCC feature extraction. First, windowing is applied to audio signals. Gammaton filter banks are then applied to the fast Fourier transformed signal to highlight the most important frequencies present in the audio signal. Finally, similar to the steps involved in MFCC, a log function and a discrete cosine transform (DCT) are applied to the output of the filter banks to obtain the coefficients. The number of features of each data as a result of GTCC is given in Table 2.

Table 2. Count of GTCC Coefficient obtained by seconds

Time (s)	Count of Coefficient
1	1273
1,5	1924
2	2574
2,5	3224
3	3874
3,5	4524
4	5174
4,5	5824
5	6474

2.3. Convolution Neural Network (CNN)

Convolutional Neural Networks are a branch of deep learning. It is the most widely used deep learning method today. This method provides feature extraction using Convolution, Pooling, Dropmax and Full layer Connect layers. A neural network is then trained on these outputs. In this study, data consisting of MFCC and GTCC features are given as input data to CNN. The data were divided into 9 different periods and their achievements were evaluated.

When we look at the model used in this study, 3x3 filters of 64 grains were applied twice in the Convolution layer. After that, 3x3 max pooling was done.0.60 dropout added to these two operations. These 3 operations were repeated 3 times and a flattened layer was created. A 100-neuron neural network consisting of 1 Hidden Layer is added to the Flatten layer. It was realized with 500 epochs in trains. After 500 epochs, analyzes were made on the highest train results. 70% of the data was used for training and 30% for validation. The numbers of train and validation data for the times used in MFCC and GTCC are given in Table 3.

Table 3. Train and validation data numbers used in model training

Time (s)	Count of Train Data	Count of Validation Data
1	1872	1008
1,5	1310	706
2	936	504
2,5	748	404
3	655	353
3,5	561	303
4	468	252
4,5	468	252
5	374	202

The CNN model was applied to the generated data sets and the results were compared.

3. Results and Discussion

Results Auscultation sounds provide information about the patient's general condition. Today, with the rise of artificial intelligence, the importance of data has increased. Correct processing, recording, and labelling of data in the health field are very important. Since the use of artificial intelligence in the field of health will directly affect human health, the successful results obtained in this area will positively contribute to early diagnosis and treatment.

This study examined the effect of MFCC and GTCC on the ideal regional diagnosis recording time. The records are divided into 9 periods from 1 to 5 and analyzed. When the training results with MFCC were examined (Figure 1), it was seen that the highest performance was in the data of 1.5s duration. Looking at the test data obtained in the highest performance, it was found that the highest test performance is in the 1s data. In the data of 4s, it is seen that despite the high train result, the lowest validation success.

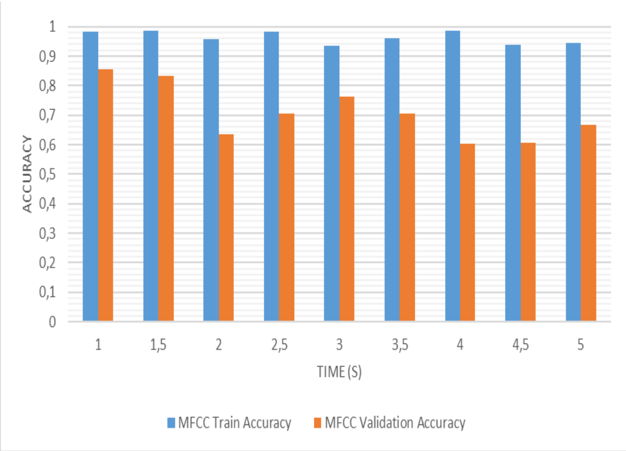

Fig. 1 MFCC performance results

When the results of the trains made with GTCC are examined (Figure 2), it is observed that the highest performance is in the data of 2s duration. It is seen that the highest successful validation results are in 1s time data, as in MFCC. It was seen that the lowest GTCC validation result was in the data with a duration of 4.5s.

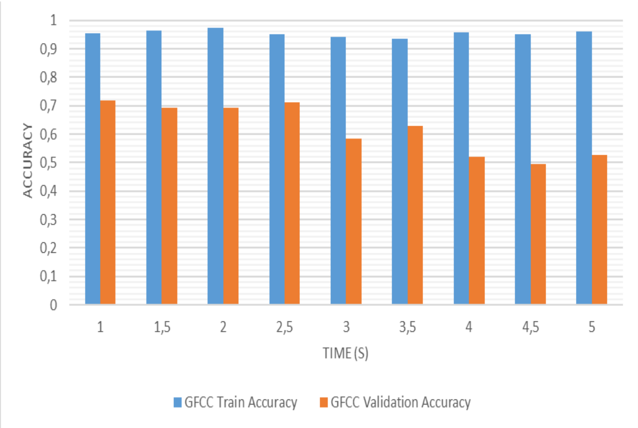

Fig. 2 GTCC performance results

When the train achievements of MFCC and GTCC are compared (Figure 3), it is generally seen that MFCC gives high results. The highest train achievements were achieved with MFCC. It is seen that the results of the 3 s data are also low in both methods.

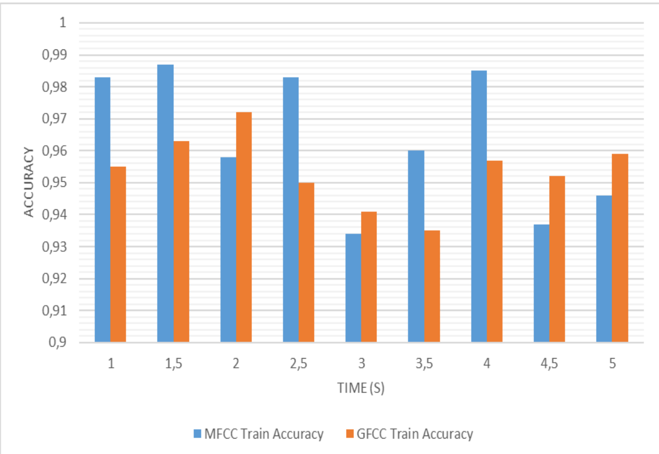

Fig. 3 GTCC Validation results

When looking at the comparison of validation results (Figure 4), it can be seen that MFCC gives better results in general. Especially in 1s time data, the validation values are high in both methods. It is noticeable that the data of 4.5s have the lowest values.

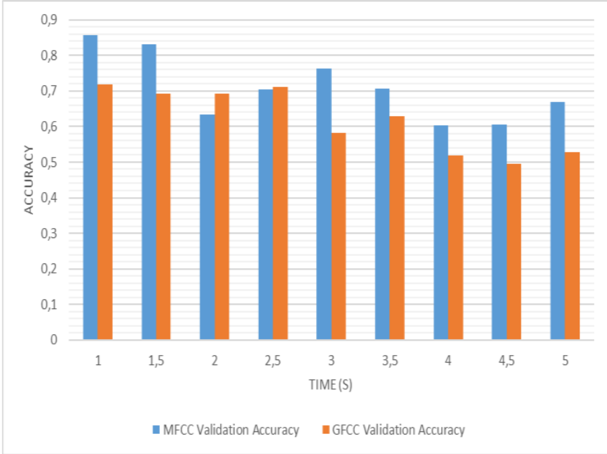

Fig. 4 GTCC validation results

According to the obtained results, it is seen that the results of the 1s duration data have high validation values. Also, it was seen that MFCC gave better validation results than GTCC. It is noticeable that the train achievements of MFCC and GTCC are close. However, since the ideal recording time was sought, it was concluded that the usable model is the 1s time model using MFCC. The most data in the train is 1s data. The validation value of the 1s time data has a high success rate and the train success rate is 98%, making this data ideal.

References

Aziz, S., Khan, M. U., Shakeel, M., Mushtaq, Z., Khan, A. Z. 2019. "An Automated System towards Diagnosis of Pneumonia using Pulmonary Auscultations". MACS 2019 - 13th International Conference on Mathematics, Actuarial Science, Computer Science and Statistics, Proceedings.

Bahoura, M. and Ezzaidi, H., 2013. "Hardware implementation of MFCC feature extraction for respiratory sounds analysis," in Proc. 8th Int. Workshop Syst., Signal Process. Appl. (WoSSPA), May 2013, pp. 226–229

Bahoura, M. and Pelletier, C., 2004. "Respiratory sounds classification using cepstral analysis and Gaussian mixture models," in Proc. 26th Annu. Int. Conf. IEEE Eng. Med. Biol. Soc., Sep. 2004, pp. 9–12.

Bardou, D., Zhang, K., Ahmad, S. M. 2018. "Lung sounds classification using convolutional neural networks". Artificial Intelligence in Medicine, 88, 58–69.

Cheng, S., Wang, C., Yue, K., Li, R., Shen, F., Shuai, W., … Dai, L. 2022. "Automated sleep apnea detection in snoring signal using long short-term memory neural networks". Biomedical Signal Processing and Control, 71(PB), 103238.

Dar, J. A., Srivastava, K. K., Lone, S. A. 2022. "Jaya Honey Badger optimization-based deep neuro-fuzzy network structure for detection of (SARS-CoV) Covid-19 disease by using respiratory sound signals". International Journal of Intelligent Computing and Cybernetics.

Jayalakshmy, S., Sudha, G. F. 2021. "GTCC-based BiLSTM deep-learning framework for respiratory sound classification using empirical mode decomposition". Neural Computing and Applications, 33(24), 17029–17040.

Kutlu Y, Karaca G. Recognition of turkish command to play chess game using cnn. Akıllı Sistemler ve Uygulamaları Dergisi

(Journal of Intelligent Systems with Applications) 2022; 5(1): 71-73.

Lella, K. K., Pja, A. 2022. "Automatic diagnosis of COVID-19 disease using deep convolutional neural network with multi-feature channel from respiratory sound data: Cough, voice, and breath". Alexandria Engineering Journal, 61(2), 1319–1334.

Mayorga, P., Druzgalski, C., Morelos, R. L., Gonzalez, O. H. and Vidales, J. 2010. "Acoustics based assessment of respiratory diseases using GMM classification," in Proc. Annu. Int. Conf. IEEE Eng. Med. Biol., Aug. 2010, pp. 6312–6316.

Meng, F., Shi, Y., Wang, N., Cai, M., & Luo, Z. (2020). Detection of Respiratory Sounds Based on Wavelet Coefficients and Machine Learning. IEEE Access, 8, 155710–155720.

Mridha, K., Sarkar, S., Kumar, D. 2021. "Respiratory Disease Classification by CNN using MFCC". 2021 IEEE 6th International Conference on Computing, Communication and Automation, ICCCA 2021, 517–523.

Pittner, S., and Kamarthi, S. V. 1999. "Feature extraction from wavelet coefficients for pattern recognition tasks," IEEE Trans. Pattern Anal. Mach. Intell., vol. 21, no. 1, pp. 83–88, Jan. 1999.

Shaharum, S. M., Sundaraj, K., Aniza, S., Palaniappan, R., Helmy, K. 2019. "A performance comparison of wheeze feature extraction methods for asthma severity levels classification". 2018 9th IEEE Control and System Graduate Research Colloquium, ICSGRC 2018 - Proceeding, (August), 145–150.

Tamas, W., Notton, G., Paoli, C., Nivet, M. L., Voyant, C. 2016. "Hybridization of air quality forecasting models using machine learning and clustering: An original approach to detect pollutant peaks". Aerosol and Air Quality Research, 16(2), 405–416.

Winursito, A., Hidayat, R. and Bejo, A. 2018. "Improvement of MFCC feature extraction accuracy using PCA in Indonesian speech recognition," 2018 International Conference on Information and Communications Technology (ICOIACT), 2018, pp. 379-383.