Article **Journal of Technology and Innovation**
December 2023, Vol.10 No.27 1-7

MFCC feature extraction for COPD detection

Extracción de características MFCC para la detección de la EPOC

VILLAFUERTE-LUCIO, Diego Ángel´†*, YAÑEZ-VARGAS, Juan Israel´, QUINTANILLA-DOMINGUEZ, Joel´,´´ and TREJO-FRIAS, Alejandra´

´Universidad Politécnica de Juventino Rosas. Maestría en Ingeniería. Hidalgo 102, Comunidad de Valencia, Santa Cruz de Juventino Rosas, Guanajuato, México.
´´Universidad Virtual del Estado de Guanajuato, UVEG, Hermenegildo Bustos 129-A, Purísima del Rincón, Guanajuato, México.

ID 1st Author: *Diego Ángel, Villafuerte-Lucio* / **ORC ID:** 0000-0002-7559-7847, **CVU CONACYT ID:** 1141208

ID 1st Co-author: *Juan Israel, Yáñez-Vargas* / **ORC ID:** 0000-0001-5749-8442, **CVU CONACYT ID:** 295711

ID 2nd Co-author: *Joel, Quintanilla-Domínguez* / **ORC ID:** 0000-0003-2442-2032, **CVU CONACYT ID:** 165237

ID 3rd Co-author: *Alejandra, Trejo-Frías* / **ORC ID:** 0000-0003-1582-1971, **CVU CONACYT ID:** 1152060

DOI: 10.35429/JTI.2023.27.10.1.7 Received July 10, 2023; Accepted December 30, 2023

Abstract

Mel Frequency Cepstral Coefficients (MFCC) are acoustic features, which are based on human auditory perception, focus on spectral properties and capture relevant features of the audio signal related to vocal tract shape and energy in different frequency bands. The methodology for feature extraction involves several stages. First, the audio signals go through a preprocessing stage in which they are initially normalized, then background noise is reduced, and finally, they pass through a median filter. In the next block of MFCC feature extraction, the first task was to split the audio signal into small frames. Then, a hamming window is applied to smooth the edges of each frame and the short time Fourier transform of each frame is calculated. Next, Mel filters are applied, which adjusts the representation of the frequency spectrum to human auditory perception. Finally, cepstral coefficients are calculated from the frequency spectrum. The MFCC coefficients are then used as input features for classifiers and machine learning algorithms.

COPD Detection, Features extraction, coefficients MFCC

Resumen

Los coeficientes Mel Frequency Cepstral Coefficients (MFCC) son características acústicas, que se basan en la percepción auditiva humana, se centran en las propiedades del espectro y capturan características relevantes de la señal de audio relacionadas con la forma del tracto vocal y la energía en diferentes bandas de frecuencia. La metodología para la extracción de características implica varias etapas. Primeramente, las señales de audio pasan por una etapa de preprocesamiento en la cual inicialmente se normalizan, después se disminuye el ruido de fondo, por último, pasan a través de un filtro de medianas. En el siguiente bloque de extracción de las características MFCC la primera tarea que se realizo fue dividir la señal de audio en pequeñas tramas. Luego, se aplica una ventana hamming para suavizar los bordes de cada trama y se calcula la transformada de Fourier de tiempo corto de cada trama. Enseguida, se aplican los filtros de Mel, que ajusta la representación del espectro de frecuencia a la percepción auditiva humana. Finalmente, se calculan los coeficientes cepstrales a partir del espectro de frecuencia. Los coeficientes MFCC se utilizan luego como características de entrada para clasificadores y algoritmos de aprendizaje automático.

Detección EPOC, Extracción de características, MFCC coeficientes

Citation: VILLAFUERTE-LUCIO, Diego Ángel, YAÑEZ-VARGAS, Juan Israel, QUINTANILLA-DOMINGUEZ, Joel and TREJO-FRIAS, Alejandra. MFCC feature extraction for COPD detection. Journal of Technology and Innovation. 2023. 10-27:1-7.

* Correspondence of the Author (Email: M21030005@upjr.edu.mx)
† Researcher contributing as first author.

© ECORFAN-Bolivia

December 2023, Vol.10 No.27 1-7

Introduction

Studies have shown that lung diseases are one of the leading causes of death in the world. Of the various lung diseases, Chronic Obstructive Pulmonary Disease (COPD) stands out.

According to the World Health Organization (WHO), it is estimated that around 384 million people worldwide are affected by COPD, it is the third leading cause of death worldwide, causing 3.23 million deaths in 2019, almost 90% of COPD deaths in people under 70 years of age occur in low- and middle-income countries (WHO, 2023). Over time, different techniques have been developed for the diagnosis of lung diseases, some of which include thoracic imaging examination, thoracic bronchoscopy, pulmonary auscultation, among others. Lung auscultation is a method used to examine and diagnose the behaviour of the respiratory system by listening to the sounds of breathing in and out (WHO, 2023).

Lung auscultation is usually performed on the posterior walls of the thorax and trachea, as shown in Figure 1. This provides useful information about the lungs, commonly when a respiratory condition is present, the first thing that is performed is a lung auscultation (Bohadana, 2014).

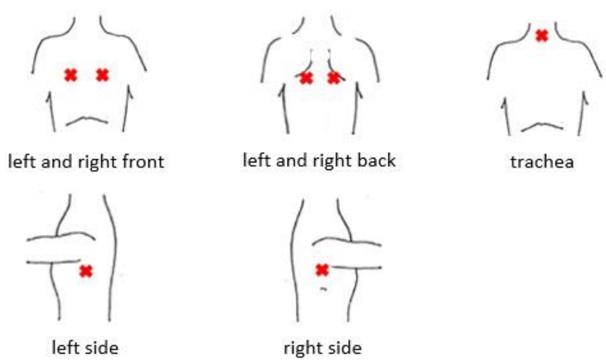

Figure 1 Key points for lung auscultation
Source: Own source

Dataset

The dataset used was the Int. Conf. on Biomedical Health Informatics (ICBHI) of breath sounds, this dataset contains audio samples collected from two research teams in two different countries over several years. The dataset includes 920 recordings of varying lengths (from 10 to 90 seconds).

These recordings were taken from 126 patients for a total of 5.5 hours. The dataset is made up as shown in Figure 2 of different lung diseases such as: COPD, pneumonia, bronchiolitis, bronchiectasis, URTI, LRTI, asthma, and healthy (Rocha BM, 2019).

Diseases															
URTI		Healthy		Asthma		COPD		LRTI		Bronchiectasis		Pneumonia		Bronchiolitis	
M	W	M	W	M	W	M	W	M	W	M	W	M	W	M	W
6	8	13	13	0	1	48	15	2	0	2	5	4	2	4	2
14		26		1		64		2		7		6		6	

Figure 2 Number of patients for each disease
Source: Own source

Each of the audio signals has the following attributes, as shown in Figure 3:

Sample rate	Number of channels	Duration	Format
44100 Hz	1	20 seg	WAV

Figure 3 Attributes of the audios
Source: Own source

Methodology

Figure 4 presents the proposed methodology for MFCC feature extraction consisting of data acquisition, preprocessing and feature extraction blocks. Each of the presented blocks plays an important role in the effective processing of the audio signal to obtain MFCC features that will later be used for sound classification..

Figure 4 oposed methodology for MFCC feature extraction
Source: Own source

Data acquisition

For the acquisition of the data, a segmentation was performed by taking only the EPOC and Sanos audios of the male gender. Ten patients were selected for each type of audio, taking one audio for each patient as shown in Figure 5.

ISSN: 2410-3993
ECORFAN® All rights reserved.

December 2023, Vol.10 No.27 1-7

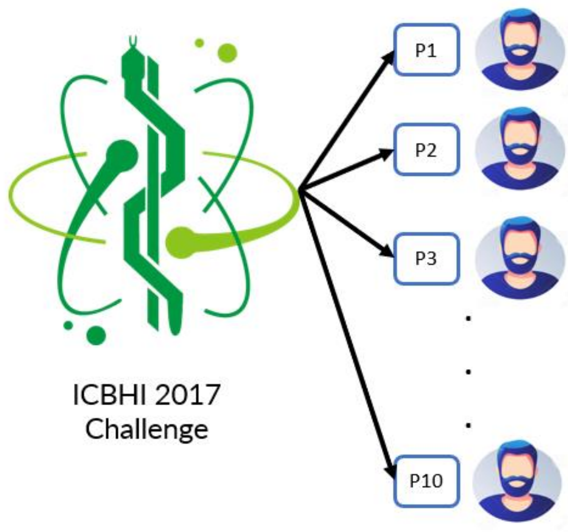

Figure 5 Data acquisition
Source: Own source

Preprocessing

In the pre-processing stage, operations are performed with the intention of improving the audio signal and eliminating unwanted noise. The operations performed in this stage are represented in Figure 6 and are the stages of normalisation 0-1, Wiener Filter and Median Filter.

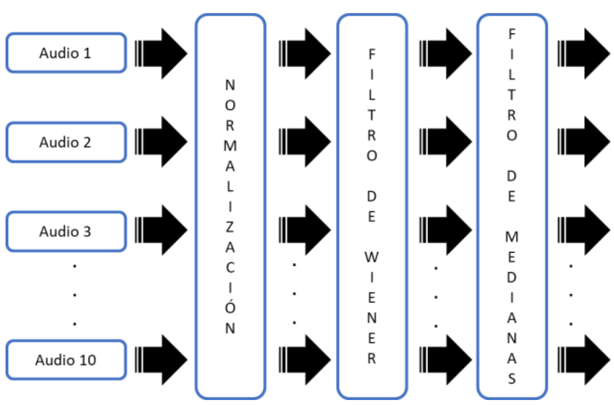

Figure 6 Pre-processing stage
Source: Own source

Normalisation

Normalisation is quite useful when seeking to equalise amplitudes or to adapt values to specific ranges. In this stage, the aim is to scale the values of the audio signal so that they are in the range 0 to 1, this is done by means of the formula 1:

$$x\_out = \frac{(x - min\,(x))}{(max(x) - min\,(x))} \qquad (1)$$

Where:

x is the input value to be normalised.

min(x) is the minimum value in the input matrix or vector x.

max(x) is the maximum value in the input matrix or vector x.

x_out is the normalised value that will be in the range [0 1].

Wiener filter

The Wiener filter is used to perform background noise filtering of the audio signal, providing an estimate of the signal of interest as used in (Pascal). This process is depicted in Figure 7.

Where:

w [n] = is the input signal.

x [n] = is the output of the filter.

s [n] = is the reference signal.

e [n] = is the filter error.

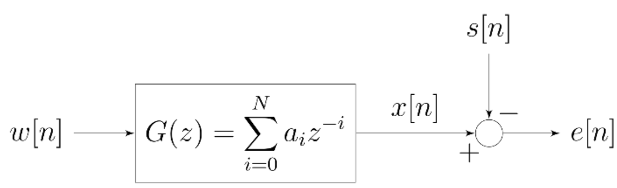

Figure 7 Schematic of the Wiener filter
Source: Own source

Median filter

The median filter is quite useful for removing noise in audio signals, such as clicks, pops or sudden interference. This type of noise is usually more noticeable compared to background noise. By applying the median filter, the amplitude of the noise is reduced without smoothing the audio signal of interest too much, which helps to improve the quality of the audio signal by eliminating or significantly reducing the noise.

MFCC Feature Extraction

MFCC feature extraction is a process used in audio signal processing, audio classification and speech recognition. MFCC is based on human auditory perception and is used to represent the acoustic features of an audio signal more efficiently this process can be seen reflected in (Winursito, A) and (Alodia Yusuf).

ISSN: 2410-3993
ECORFAN® All rights reserved.

The feature extraction process performed is represented by Figure 8.

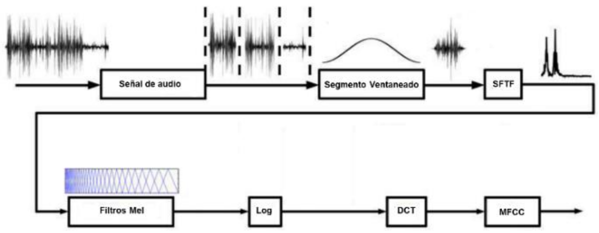

Figure 8 Stages of feature extraction
Source: Own source

Frames

For a better analysis of the lung sounds, 500 ms frames will be extracted, as this is where a better spectral analysis is possible.

First, the time of each of the frames is defined, as in Equation 2.

$$Tt = 0.5\, seg \quad (2)$$

Where:

Tt = is the time of the frame.

Once the time of the frames is known, the number of samples that will make up the frames is obtained, according to equation 3.

$$Nm = Tt\, x\, f \quad (3)$$

Where:

Nm = es the number of samples of each frame.

f = is the sampling frequency.

To avoid loss of information at the ends, we will overlap half the number of samples in each of the frames, as described in equation 4.

$$Op = \frac{Nm}{2} \quad (4)$$

Where:

Op = is the number of samples to be overlapped

Windowing

By dividing the signal into 500 ms frames, discontinuities occur at the beginning and end of each frame. This leads to errors in the spectral analysis.

To reduce discontinuities and minimise the introduction of spurious frequency components, a windowing technique, such as the Hamming window, is used.

This mathematical function described in equation 5 is multiplied point-by-point with the raster, which smooths the raster edges.

$$w(n) = 0.54 - 0.46\cos(2\pi\frac{n}{N}) \quad (5)$$

Short-Time Fourier Transform (STFT)

The short time Fourier transform of each frame is represented by the following equation 6.

$$X_m(f) = \sum_{n=-\infty}^{\infty} x(n)g(n-mR)e^{-j2\pi fn} \quad (6)$$

Where:

$g(n)$ = window length.

$Xm(f)$ = DFT of the data centred at time mR.

R = Jump size between successive DFTs. The skip size is the difference between the window length M and the overlap length L.

Mel filters

A set of Mel filters is used to map the linear scale of frequencies to a Mel scale, which better approximates human auditory perception. The Mel filters are superimposed and placed on the frequency scale, covering the whole range of frequencies relevant to human speech, as described in equation 7.

$$mel = 2595\, log_{10}(1 + \frac{f}{700}) \quad (7)$$

Log

The logarithm of the spectral power is calculated for each of Mel's filters. This is done to account for the non-linear response of the human ear to differences in sound intensity.

Discrete Cosine Transform (DCT)

Finally, the discrete cosine transform is applied to the logarithmic power spectral coefficients to reduce their dimensionality and obtain the cepstral coefficients, which represent the acoustic characteristics of the audio signal.

ECORFAN® All rights reserved.

Article
Journal of Technology and Innovation
December 2023, Vol.10 No.27 1-7

Feature selection

Feature selection using variance is a technique used to reduce the dimensionality of a set of features by selecting those features that have a high variance and discarding those features with a low variance.

Using variance it can be determined that if the variance value is small the content of that feature is very likely to be the same or very similar, so it will contribute little to the classifier.

Results

The results obtained by performing each of the processes are shown in Figure 9 and 10:

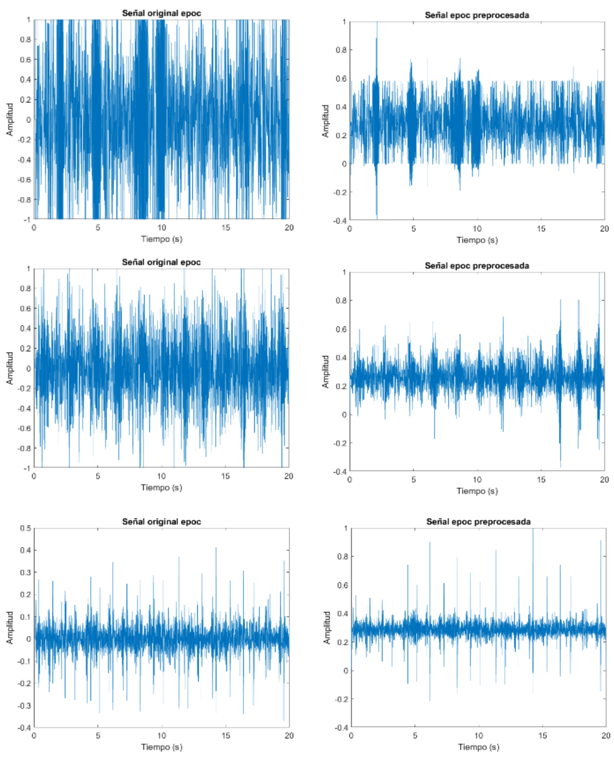

Figure 9 Graphical representation of the audios before and after pre-processing
Source: Own source

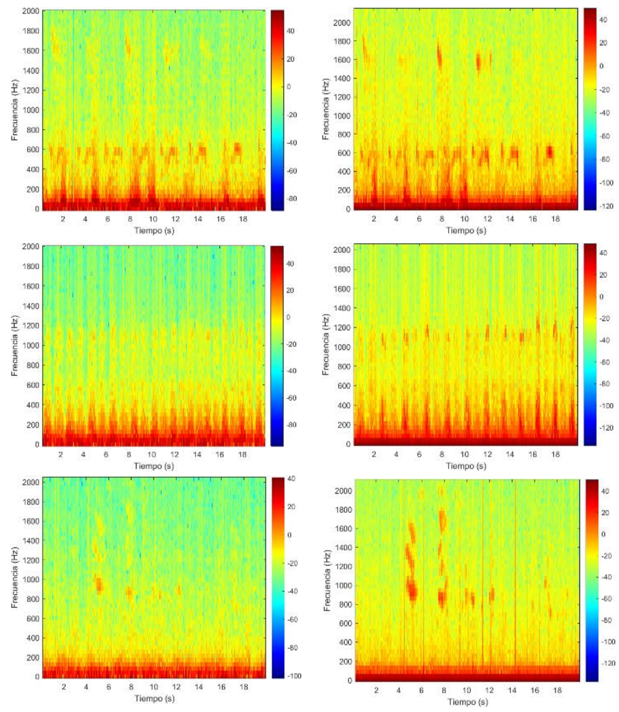

Figure 10 Graphical representation of the spectrograms before and after pre-processing
Source: Own source

After the MFCC feature extraction process, the result is 13 coefficients, described in Table 1 and 2, per audio type to be used as input data for an audio signal classification system.

	1	2	3	4
1	0.2488	-1.8851	-0.6121	-3.3223
2	1.4328	2.8509	2.1186	3.1375
3	0.4041	1.0396	-0.2124	-0.0243
4	0.6045	0.8364	1.9249	-0.1647
5	0.5093	0.5619	0.1245	0.3481
6	-0.5724	-0.0789	0.0635	1.0783
7	-0.0588	0.1585	0.0745	0.8590
8	0.7250	0.5187	0.2915	0.1677
9	0.2026	-0.0608	0.4754	-0.3561
10	0.7426	0.2865	0.3625	-0.1885
11	0.3227	0.1910	-0.0001	0.4146
12	-0.0980	0.0370	0.2977	0.4764
13	0.1760	0.2003	0.2746	0.0629

Table 1 MFCC coefficients of COPD patients
Source: Own source

	1	2	3	4
1	-10.9472	-5.8302	-4.0880	-9.3527
2	3.6684	3.0756	3.0943	5.1568
3	1.2175	0.8550	0.8539	1.3832
4	1.2603	1.0096	0.9853	0.9061
5	0.6564	0.6493	0.6354	0.4463
6	0.7117	0.6419	0.6213	0.4045
7	0.4813	0.4771	0.4561	0.2069
8	0.3474	0.4380	0.4435	0.1608
9	0.3499	0.3367	0.3569	0.2208
10	0.2725	0.3162	0.3516	0.2336
11	0.2770	0.2526	0.3002	0.1741
12	0.4101	0.2578	0.2899	0.2863
13	0.3601	0.2048	0.2398	0.2615

Table 2 MFCC coefficients of HEALTHY patients
Source: Own source

ISSN: 2410-3993
ECORFAN® All rights reserved.

Journal of Technology and Innovation
December 2023, Vol.10 No.27 1-7

The results obtained from the selection of characteristics are shown in Table 3, where it is determined to take the first five coefficients from the variance study carried out on all the data as input for the classifier and to find their degree of dispersion.

	Variaance	
Coefficient	**EPOC**	**Healthy**
1	3.4455	9.0377
2	0.9791	0.8834
3	0.3008	0.7591
4	0.1665	0.1894
5	0.0953	0.1885
6	0.1406	0.0580
7	0.0561	0.0598
8	0.0478	0.0572
9	0.0724	0.0380
10	0.0418	0.0245
11	0.0318	0.0265
12	0.0318	0.0220
13	0.0233	0.0208

Table 3 Values resulting from the selection of characteristics
Source: Own source

Next, the normal distribution of the selected characteristics is shown by variance, through the Gaussian bell, quantile plot and box and whiskers diagram, this is done in order to analyse and model the distribution of values, it is worth mentioning that not all characteristics must necessarily follow a normal distribution, the data are represented in Figures 11, 12, 13, 14 and 15.

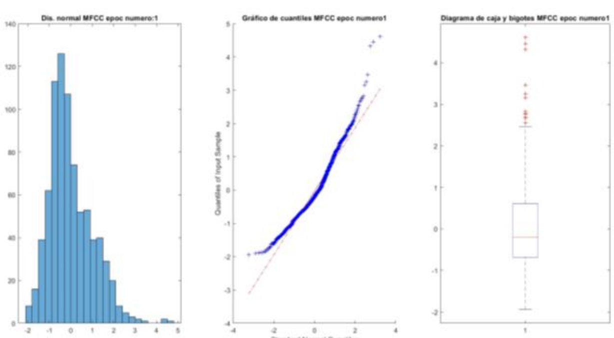

Figure 11 Normal distribution of coefficient 1 EPOC Gaussian Bell, Quantile Plot, Box and Whisker Plots
Source: Own source

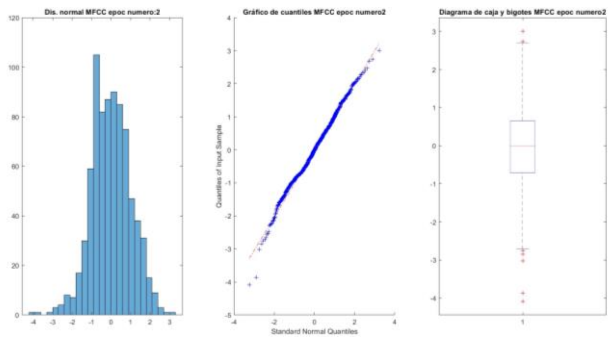

Figure 12 Normal distribution of coefficient 2 EPOC Gaussian Bell, Quantile Plot, Box and Whisker Plot
Source: Own source

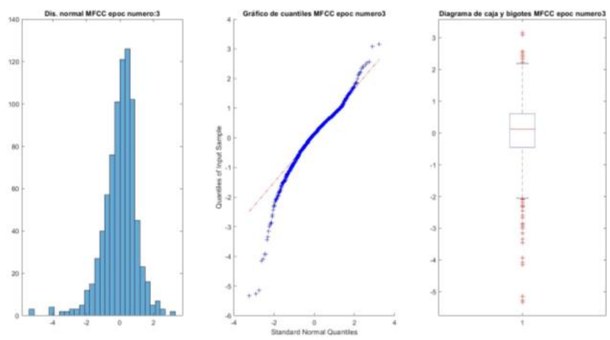

Figure 13 Normal distribution of coefficient 3 EPOC Gaussian Bell, Quantile Plot, Box and Whisker Plots
Source: Own source

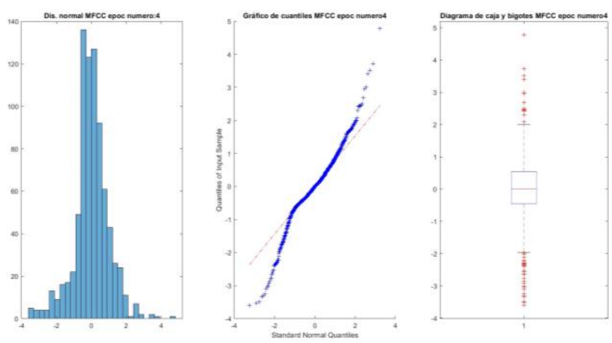

Figure 14 Normal distribution of coefficient 4 EPOC Gaussian Bell, Quantile Plot, Box and Whisker Plot
Source: Own source

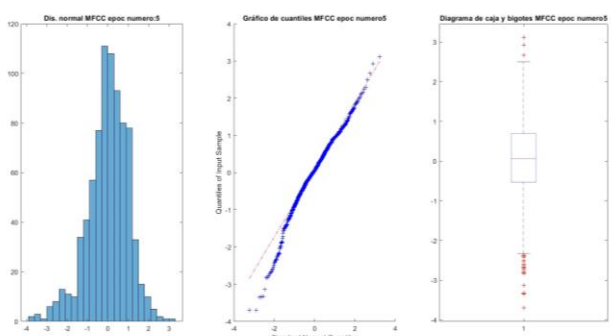

Figure 15 Normal distribution of coefficient 5 EPOC Gaussian bell, quantile plot, box and whisker plot
Source: Own source

ISSN: 2410-3993
ECORFAN® All rights reserved.

December 2023, Vol.10 No.27 1-7

Conclusion

MFCC feature extraction is a very efficient technique in audio signal processing. It provides an effective representation of the acoustic characteristics of an audio signal, allowing for more accurate analysis and classification.

MFCC provides several advantages, one of which is the ability to compact spectral information into a small number of coefficients, which reduces the dimensionality of the data and improves computational efficiency. In addition, MFCCs are relatively robust to noise and changes in acoustic conditions, which makes them suitable for applications in speech recognition, speech synthesis, sound classification and other audio signal processing tasks.

References

Alodia Yusuf, S. A., & Hidayat, R. (2019). "*MFCC Feature Extraction and KNN Classification in ECG Signals*". 6th International Conference on Information Technology, Computer and Electrical Engineering (ICITACEE). https://doi.org/10.1109/ICITACEE.2019.8904285

Bohadana, A., Izbicki, G., & Kraman, S. S. (2014). "*Fundamentals of Lung Auscultation*". New England Journal of Medicine, 370(8), 744–751. https://doi.org/10.1056/nejmra1302901

Dixit, A., Vidwans, A., & Sharma, P. (2016). "*Improved MFCC and LPC algorithm for bundelkhandi isolated digit speech recognition*". International Conference on Electrical, Electronics, and Optimization Techniques (ICEEOT). https://doi.org/10.1109/iceeot.2016.7755413

Madan, Akansha & Gupta, Divya. (2014). "*Speech Feature Extraction and Classification: A Comparative Review*". International Journal of Computer Applications. 90. https://doi.org/10.5120/15603-4392.

Pascal Scalart (2023). "*Wiener filter for Noise Reduction and speech enhancement*" (https://www.mathworks.com/matlabcentral/fileexchange/24462-wiener-filter-for-noise-reduction-and-speech-enhancement), MATLAB Central File Exchange. Recuperado Julio 14, 2023.

Rocha, BM. (2019) "*An open access database for the evaluation of respiratory sound classification algorithms*". https://doi.org/10.1088/1361-6579/ab03ea

Winursito, A., Hidayat, R., & Bejo, A. (2018). "*Improvement of MFCC feature extraction accuracy using PCA in Indonesian speech recognition*". International Conference on Information and Communications Technology (ICOIACT). https://doi.org/10.1109/ICOIACT.2018.8350748

World Health Organization (WHO). (March 2023). *Chronic obstructive pulmonary disease (COPD).* Recuperado de https://www.who.int/news-room/fact-sheets/detail/chronic-obstructive-pulmonary-disease-(copd)

ISSN: 2410-3993
ECORFAN® All rights reserved.