PROtek: Jurnal Ilmiah Teknik Elektro
https://ejournal.unkhair.ac.id/index.php/protk/index
Volume 11. No 2, May 2024
e-ISSN: 2527-9572 / ISSN: 2354-8924

Mathematical Representation for speech signal based on polynomial equation

Entesar Alasaad
Computer Science Dept./ College of Computer Science and Mathematics/
University of Mosul, 41002 Mosul, Iraq
Intesar.csp39@student.uomosul.edu.iq

Khalil I. Alsaif
Computer Science Dept./ College of Computer Science and Mathematics/
University of Mosul, 41002 Mosul, Iraq
Khalil_alsaif@uomosul.edu.iq

***Abstract* – The sound is an important vital information that it relies to recognize the character when lessening to it. Therefore, the audio signal adopted into many important applications. Sound forming and synthesizing in addition to distinguishing the speaker are so important in fields of digital signal processing.In this paper, work is done to represent acquired acoustic signal based on mathematical techniques. Mathematical representation provides deal with the sound signal which lead to smoothing, amplification or compression, in addition to the sound filtering process. In the proposed algorithm, polynomials of various degrees were adopted as a mathematical representation for speech signal, then the retrieved speech was studied based on level of clarity and the possibility of adopting it as an alternative signal in terms of the proximity to the original sound and the amount of noise added to it. The results shows that the proposed algorithm with degree of polynomial 20 and segment length 25 had the best sound representation and so closed to the original, which clearly seen from the evaluation parameters (Correlation=0.9993, Mean squared error(MSE) =1.32e-06, Standard deviation(STD)=1.80e-05 and the Euclidean dimension(ED) =0.1703).**

Keywords: curve fitting, sound representation, digital speech processing, digital signal processing.

I. INTRODUCTION

Sound is a continuous analog wave with an amplitude (representing the degree of loudness) and a frequency (representing the intensity of the sound), and it is a one-dimensional signal (in terms of time) representing the air pressure in the ear[1]. Work is done on the audio signal with digital data, where separate values of analog waves are captured within small fields in a process known as (sampling) using an analog-to-digital converter (A/DC) in the sound card, then the samples are then converted to Binary numbers, where each number represents the height of the sound wave in binary coding. To be stored and then dealt with again through a loudspeaker, for example, where it is transformed in the same way from digital signals to analog waves using an (D/AC) [2]. As shown in Figure (1)[3].

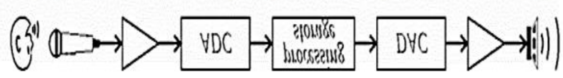

Figure 1. A Complete Digital Audio Processing System

Digital processing is now the preferred method of dealing with sound and speech as new applications and sound systems are mostly digital in nature. This revolution from analog to digital has mostly occurred during the last century.[4]. Researcher Oded Ghitza et al. In 1994 presented an assessment of the suitability of current peripheral auditory models to mimic human performance in the context of speech recognition and speech coding. Drawing on principles derived from the properties of the auditory system, and thus the ability to simulate human performance in the same task for the purpose of speech recognition[5]. A study was conducted by Anandthirtha. B. GUDI and H. C. Nagara in 2009 on children (5 to 8 years old) native speakers of the Kannada language. In this work, the curve fitting technique was applied and obtaining suitable constant values for the curve by fitting the amplitude profile of the speech data taken using the sum of the sine functions with a confidence level of over 90% [6]. Heba Abdul Nabi and Dr. Khalil Ibrahim Al-Saif proposed in 2011 an algorithm that classifies the speaker's age into two categories (young and old) based on his speech indication. Fit the curve by adopting a polynomial. After applying the proposed algorithm to 50 people of both sexes, it was found that the algorithm succeeded in classifying the age of people by 80% and failed in 20% of them[7]. Aditi Pareek and Dr. Lata Gidwani in 2015 used curve fitting methods for solar irradiance pressure and prediction based modeling. Where the curve fitting methods were used to derive the modeling equation to estimate the horizontal global solar irradiance. The polynomial data fit method was used for data homogeneity and was tested with different degrees of polynomial curve combinations, ranging from degree I (linear) to degree [8]. In 2017, Adedayo M. Farayola et al. Presented an approach to design MPPT (Maximum Power Point Tracking) fast tracking using a predicted grade VI polynomial curve compatible with MPPT technology. The results indicate that the use of a sixth order polynomial curve and Fuzzy

Artificial Neural Inference System (ANFIS) techniques can trace the highest maximum power point from the lower order curve techniques[9]. Gabor Kees et al., 2017, consider a multilingual examination of prospects for expecting misery based on the discourse process. The test was conducted in three European dialects: German, Hungarian and Italian. These vocal selections were mapped by a mathematical vector of cursor-specific information that relate with the seriousness of sorrow in comparable dialect independent way. The technique is even fit for foreseeing the seriousness of despondency inside the instance of a dialect not utilized all through the training of the model[10].

Fenghao Cui and others in 2019 presented a study whose purpose was to create an RWQM system to illustrate new calibration procedures using field river water quality data. The extensive application of numerical curve fitting techniques of cubic homogeneity, polynomial curve fitting, and nonlinear least squares provided a robust estimation procedure to support data analysis and parameter calibration when estimating unknown parameters[11]. Speech is a form of communication in daily life that has been present in humans since the beginning of civilizations [12]. It is a complex signal that features variable distributions of energy in time as well as in frequency depending on the specific sound being produced. It also consists of an organized group of continuous sounds by virtue of its production mechanism [13]. Speech signals consist of a series of sounds and the sequence of sounds results from the acoustic excitation of the vocal tract when air is expelled from the lungs. Speech begins with lung contractions and carries a sound roughly Gaussian frequency distribution. This air is pushed up through the airways, bypassing a group of muscle folds at the top of the windpipe called the vocal cords. Then the air enters the back of the oral cavity, where it follows one of two paths to the outside. The first path is above and around the tongue, passes through the teeth and exits through the mouth. The second route passes through the nasal cavity - this is the only pathway possible when Vellum is closed [1].

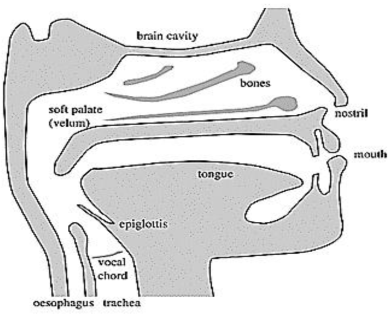

Figure 2. A Cross-Sectional Diagram Of The Human Vocal Apparatus, Showing The Main Hinges, Resonators, And Properties Of The Vocal And Nasal Tracts.

Figure 2 [2] shows a diagram of a speech production device (also known as a human head). The actual sound that is produced depends on many parameters including lung strength and pressure adjustment, glottis constriction, tension in the vocal cords, shape of the mouth, and position of tongue and teeth [2]. There is nothing special about speech from a phonological perspective (because it is simply like any other sound). It is only when we hear it that our brains begin to interpret this signal as speech. Regardless of this explanatory behavior of the human brain, there are acoustic properties in music and other sounds that are similar in nature to speech in their spectral and temporal characteristics [2]. Speech processing is one of the fastest growing topics, and its applications are also expanding very quickly, in addition to the rapid growth of computational capabilities with regard to digital machines, which leads to speed and ease of application. The goal of digital speech processing is to take advantage of digital computing techniques to process speech signal to increase comprehension, improve communication, and increase efficiency and productivity associated with speech activities[14]. There are three main stages of speech processing [13]: speech analysis, speech recognition and speech coding.

II. METHODS

The most important stage is speech recognition. The main goal of speech recognition is to obtain effective methods for humans to communicate with computers due to the complexity that exists in human languages. This makes automatic computer-assisted speech recognition a difficult problem. Curves fitting can be defined as a method in which a mathematical combination is made through mathematical equations for a number of points through which a curve is created that passes through the specified data, thus forming the best equation that can pass the points [15]. Ideally, a function will be found that best represents this data to allow for making predictions about how the data series will behave in the future [8]. Curves are classified in two types [16]:

1. You meet exactly the specified points.
2. Approximate passes through most of the points or close to some of them.

Least squares is an important Curve fitting concept that is used to find the values of the coefficients of the equations used for the purpose of matching and with the lowest estimate of error.

$$\sum_{i=0}^{n}\left(y_i - F(x_i)\right)^2 \quad \ldots\ldots\ldots (1)$$

In eqn.1 y_i represents the experimental data and $F(x_i)$ is the curve-fitted model[15]. Polynomials are an important class of functions with one (independent) variable. It is the sum of a set of terms with non-negative powers of the independent variable. The polynomial has the following form:

$$\mathbf{y = a_0 + a_1x + a_2x^2 + a_3x^3 + \ldots + a_nx^n} \quad \ldots\ldots. (2)$$

Where a_i are the coefficients of the polynomial and may be real numbers, the highest power of the equation gives the degree of the equation. First-degree polynomials produce a straight line when represented in the coordinate plane and its formula $\mathbf{y = ax + b}$, which we need to represent at least two points related to a straight line, while second-degree polynomials generate a class of curves called a parabola whose formula is **y = ax2 + bx + c** and the simplest three-point curve. Third-degree polynomials generate cubic curves, and their simplest representation consists of four points, and their formula is **y = ax3 + bx2 + cx + d**. Thus, as the degree of the equation increases, it will give a series of curves with an increase in concave or convex**[4]**. For example, linear curve fit occurs when the data fit a straight line. Although there may be a move away from some of these data (not all of them fall on the line), the straight line provides a reasonable enough fit to make predictions[17]. In addition, other types of curves can also be used, such as Exponential Models, Fourier Series, Gaussian, Power Series, Sum of Sines, …etc[17]. Various fitting methods can evaluate the input data to find appropriate curve model parameters. Each method has its own criteria for evaluating composition residues in finding a suitable curve. By understanding the criteria of each method, then it is possible to choose the most appropriate method to apply to the data set and fit the curve as shown in Figure 3.

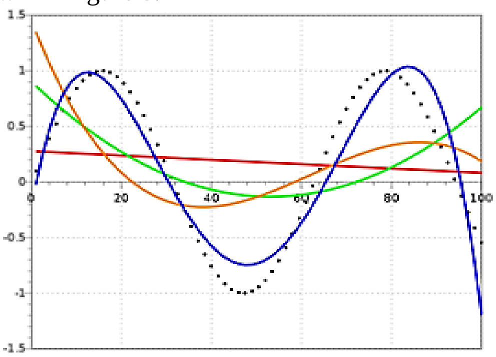

Figure 3. Set Of Points Was Created Using The Sine Function

As shown in Figure 3, a set of points was created using the sine function. The real data is represented by the black dotted line, the red line is the first degree polynomial, the green line is the second degree, the orange line is the third degree, and the blue line is the fourth degree. It is evident from the previous example that the best of these equations gives a graph or curve close to the original points is a fourth degree polynomial. In general, approximation is used to obtain results and curves fitting the desired curve [8]. Curve fitting can be used to perform the following tasks:

1. Noise reduction and data smoothing
2. Find a mathematical relationship or a function that connects the variables and use this function to perform further manipulations on the data, such as error compensation, velocity, acceleration calculation, etc.
3. Estimate the variable value between samples of data
4. Estimating the value of the variable outside of the sample data range

In this paper, an algorithm was proposed in order to find a mathematical representation of the speech signal based on polynomials of different degrees, and then test them to find the best coefficients of the polynomial equation in addition to the degree of the polynomial. The steps of the algorithm can be represented by :

1. Speech is recorded by a single-channel 'mono' audio capture at a sample rate of 8000Hz and represented by 16bit.
2. Storing the recorded voice data for the purpose of dealing with it, after dividing it into sections of different lengths ranging from 25 to 1000 beats.
3. Inserting each of these sections on functions that find the adopted mathematical equation parameters and then using the results to find the polynomial equation with degrees from 5 to 25.
4. Carrying out a normalization operation on the resulting data in step (3) for the purpose of matching and expanding the range in a range between [-1, +1].
5. Finding the extent of convergence between the original and the final data, as well as the Euclidean dimension, the mean squared error, and the standard deviation. Through these measures, the extent to which the proposed equation matches the original speech signal is determined.

By studying the results of the previous steps, the best data segment and the best degree of the equation are determined, which gives us the best representation of the speech signal.

III. Results and Discussion

The proposed algorithm was applied to a set of variables that control a polynomial equation in the representation of a mathematically recovered sound. Among those factors: The degree of the polynomial: - The degree of the polynomial represents the extent of the complexity represented by the equation, which leads to the passage within the majority of the target points, which differ clearly in the speech signal (which changes randomly when the sound is represented on the time axis) so the degree of the polynomial is important Therefore, a set of degrees was adopted when representing the sound, ranging between degree (5) and degree (25). It was clearly and logically shown that the high degree of the polynomial gives positive results in terms of accuracy, but at the expense of time. Segment length: - The stability of the algorithm is studied when a variable number of phoneme segment length is adopted.

Table 1 includes the values of the statistical measures adopted when applying the polynomial of different degrees, starting with (5) and ending with (25) on the recorded sound data (which number 40,000 values) after dividing it into sections of different lengths ranging from (25) to (1000) Value for each segment. Figure (4) represent the speech signal with optimal segment with polynomial of degree equal 20 and segment of 25.

Table 1. Statistical Measures For Different Degrees Of Polynomial

Effect of the polynomial degree on different segment length					
Polynomial Degree	Segment length	Scales			
		Corr.	MSE	STD	ED
5	25	0.9291	0.0001	-0.0016	4.0467
	50	0.6099	0.0007	-0.0037	33.7986
	200	0.1241	0.0023	-0.0012	108.0186
	400	0.0606	0.0023	-0.0005	81.2703
	800	0.0341	0.0026	-0.0055	95.2108
	1000	0.0284	0.0044	-0.0124	183.5536
10	25	0.9919	1.06e-05	-0.0004	2.2558
	50	0.9114	0.0001	-0.0021	5.0093
	200	0.2921	0.0008	0.0048	9.6431
	400	0.1194	0.0013	0.0050	29.2499
	800	0.0567	0.0019	0.0025	46.1559
	1000	0.0455	0.0030	-0.0007	55.8499
15	25	0.9979	2.67e-06	-0.0001	0.0823
	50	0.9745	3.62e-05	-0.0011	1.3309
	200	0.5153	0.0006	0.0005	4.9041
	400	0.2093	0.0008	0.0074	7.4828
	800	0.0799	0.0012	0.0066	16.8451
	1000	0.0549	0.0020	0.0049	31.0944
20	25	0.9993	1.32e-06	1.80e-05	0.1703
	50	0.9910	1.18e-05	-0.0005	0.5613
	200	0.6224	0.0005	-0.0001	4.3057
	400	0.2934	0.0007	0.0069	3.2570
	800	0.1291	0.0008	0.0087	5.2100
	1000	0.1076	0.0008	0.0086	7.5576
25	25	0.1180	5.43e+40	-1.64e+19	6.56e+23
	50	0.7827	0.0003	0.0001	48.4295
	200	0.6117	0.0005	0.0019	10.8511
	400	0.3005	0.0008	0.0060	18.4682
	800	0.1462	0.0008	0.0090	26.1115
	1000	0.1071	0.0009	0.0094	19.8016

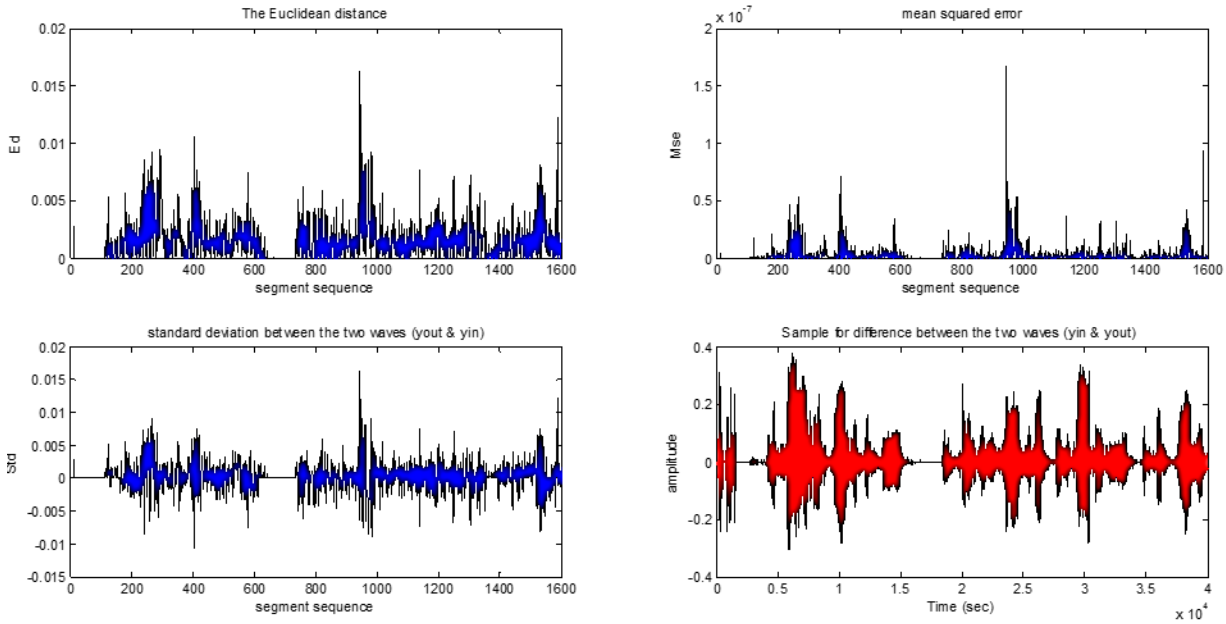

Figure 4. Speech Signal With Polynomial Degree of 20 and Segment of 25

The performance of the algorithm was evaluated by adopting several statistical measures to measure the distance between the incoming and outgoing signal:

1. Correlation coefficient, where the value of this parameter is limited to [-1,1]. When the value of this parameter approaches 1, this indicates the strength of the correlation between the two signals, but if its value is close to zero, then this indicates weak or no correlation between them[19].
2. The Euclidean dimension where its value is always positive, and the smaller the distance between two points of the incoming signal and two points of the outgoing signal, the better because it indicates the intensity of convergence between these two signals[2].
3. The mean square error measures the mean squares of errors - that is, the average square difference between the estimated values and the actual value - MSE is a measure of the quality of the estimator, its values are always positive, values close to zero are the best[20].
4. Standard deviation: used to calculate the extent of deviation of the external signal from the original (incoming) signal, or it measures the dispersion of the data and the amount of its difference from the arithmetic mean. Large values from the standard deviation indicate how far away the values are from the mean [20].

When applying the algorithm to different data segments (from 25 to 1000) with different degrees of the mathematical equation (from 5 to 25), it was observed that the best values obtained for the measures adopted in the previous step were when adopting the length for a segment of 25 values and the degree of equation 20. From the drawing, it can be seen the intensity of convergence between the incoming and outgoing signals, since the incoming signal is completely covered in red, which indicates the strength and intensity of convergence. In general, we conclude that the smaller the data segments, the better, and the greater the degree of the equation. This can be seen from the illustrations and written values, the higher the value of one segment, the more the scale values move away from the optimization and thus the departure from the signal outside the input signal, and the higher the degree of the equation, the greater the closeness between the two signals. Therefore, the curve closest to the recorded sound equation is section 25 and the degree of the polynomial mathematical equation is 20.

IV. Conclusion

Through the results and their discussion, it is clear that the algorithm applied in many applications can be adopted, including: For the purpose of distinguishing people by building an information base that contains speech signal parameters for each person. The possibility of adopting the algorithm applied in the security systems. The algorithm can be applied to synthesize sounds with fixed parameters. The algorithm is good for improving the sound by changing the parameters of the polynomial equation. The proposed algorithm can study the letter outputs of languages.

V. References

[1] D. Jurafsky and J. H. Martin, "Speech and Language Processing: An Introduction to Natural Language Processing, Computational Linguistics, and Speech Recognition," ed, 2018.

[2] I. McLoughlin, *Applied speech and audio processing: with Matlab examples*. Cambridge University Press, 2019.

[3] A. E. Villanueva-Luna *et al.*, *De-noising audio signals using MATLAB wavelets toolbox*. INTECH Open Access Publisher, 2011.

[4] D. Rocchesso, *Introduction to sound processing*. Mondo estremo, 2013.

[5] O. J. I. T. o. s. Ghitza and a. processing, "Auditory models and human performance in tasks related to speech coding and speech recognition," vol. 2, no. 1, pp. 115-132, 2014.

[6] A. Gudi, H. J. I. J. o. C. S. Nagaraj, and I. Technology, "Optimal curve fitting of speech signal for disabled children," vol. 1, no. 2, pp. 99-107, 2019.

[7] K. J. I. J. O. S. S. I Alsaif, "Speaker age detection using eigen value," vol. 11, no. 20, pp. 271-290, 2011.

[8] A. Pareek and L. Gidwani, "Measured data of daily global solar irradiation using curve-fitting methods," in *2015 International Conference on Energy Systems and Applications*, 2015, pp. 561-565: IEEE.

[9] A. M. Farayola, A. N. Hasan, and A. Ali, "Curve fitting polynomial technique compared to ANFIS technique for maximum power point tracking," in *2017 8th International Renewable Energy Congress (IREC)*, 2017, pp. 1-6: IEEE.

[10] G. Kiss and K. Vicsi, "Investigation of cross-lingual depression prediction possibilities based on speech processing," in *2017 8th IEEE International Conference on Cognitive Infocommunications (CogInfoCom)*, 2017, pp. 000097-000102: IEEE.

[11] F. Cui, C. Park, and M. J. J. o. e. m. Kim, "Application of curve-fitting techniques to develop numerical calibration procedures for a river water quality model," vol. 249, p. 109375, 2019.

[12] H. I. Younes, "Speaker age detection using eigen values," Master's thesis 2011.

[13] S. Dixit, "Speech Processing: A Review," *International Journal of Advanced Research in Computer Engineering & Technology (IJARCET)*, vol. 3, no. 8, 2014.

[14] D. Mallie, "Voice processing using MATLAB as a tool," 2014.

[15] M. t. a. Eraky, "Curve fitting and interpolation techniques," *researchgate*, 2018.

[16] J. J. J. o. C. Rougier, "Ensemble averaging and mean squared error," vol. 29, no. 24, pp. 8865-8870, 2016.
[17] M. Tektas, "MATLAB EĞRİ UYDURMA (Curve Fitting)," 2020.
[18] "https://www.ni.com/en-lb/innovations/white-papers/08/overview-of-curve-fitting-models-and-methods-in-labview.html," Website 2020
[19] P. Vidyullatha, D. R. J. I. J. o. E. Rao, and C. Engineering, "Machine learning techniques on multidimensional curve fitting data based on R-square and chi-square methods," vol. 6, no. 3, p. 974, 2016.
[20] S. Ahn and J. A. J. E. D. Fessler, The University of Michigan, "Standard errors of mean, variance, and standard deviation estimators," pp. 1-2, 2013.