Journal of Theoretical and Applied Information Technology
15th January 2023. Vol.101. No 1

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

VOICE PRODUCTION FROM THE MOVEMENTS OF A HAND

RODOLFO ROMERO-HERRRERA[1], JESUS YALJA MONTIEL PEREZ[2]

[1]IPN, ESCOM, Department of Computer Science and Engineering, México

[2]IPN. CIC, Robotics and mechatronics laboratory, México

E-mail: [1]rromeroh@ipn.mx, [2]yalja@ipn.mx

ABSTRACT

The system presented allows the interpretation of the movements of a hand. With sensors integrated with a Micro Bit card and the use of recorded voice, the concatenation of phonemes is generated. The result allows relating each degree of movement of a hand with the X and Y axes to reproduce a specific phoneme. Time-series analysis is performed in a range from -90o to 90o for both the X and Y axes; in such a way that by comparing the data generated and those previously stored, it is possible to relate a movement to a specific phoneme. In the concatenation of phonemes, recursion was used through graphs. Therefore, audio output is delivered. Intelligibility greater than 84% is obtained for a total of 350 phonemes. The system was developed using recursive functions in graphs.

Keywords: *BBC Micro Bit; accelerometer; degrees of hand positions; phoneme; Cartesian x, y-axes; recursive*

1. INTRODUCTION

Image processing is normally used for the recognition of the human body [1]; however, it is not always possible to have a camera in front of you. Due to this difficulty, it is necessary to implement other techniques that allow greater mobility. Hands are an essential part of the human being, and therefore a basic part of communication [2]. So, it is feasible to use them for voice generation from hand movements [3]. This article shows the design of a system that allows interpreting movements measured in degrees of inclination of the hand for the generation of basic phonemes [4], the project proposes the generation of signs with one hand. A sensor is characterized to obtain an equation that allows translating positions to degrees for each X, and Y-axis. The management of the inclination of the limb facilitates the interpretation, by using degrees and relating it to a natural movement of the hand [5]. An embedded Micro Bit system [6] and a card are used to transmit information wirelessly to the phoneme player, to improve mobility. In this way, a voice reproduction system based on phonemes is designed that uses the technique of graphs.

General objective

Develop a system that allows the generation of words through concatenated phonemes from the movement of a hand.

Specific objectives

- Establish a database of 350 phonemes of spoken Spanish.
- Specify utilizing graphs the design of the software that concatenates the phonemes and implements it in a hardware design.
- Establish the words that correspond to the movements of the hand.

2. RELATED WORKS

In [7] the detection of manual signals through an electronic device such as a Tablet and external sensors was proposed. It can detect up to 20 different hand signals, with IMU sensors that are placed on the fingers. In the system [8] a solution for deaf-mute people is proposed through the recognition of hand gestures and the use of a deaf-mute language. Some researchers use sophisticated algorithms such as the use of convulsive neural networks that require training [9]; and there are projects with image processing that overlook the difficulties of having a camera in front of the person or that it is required to hold it with the other hand [10]; where there are also image processing techniques that do not solve the segmentation before different scenarios and the characterization of different cameras. In [11] and [12] the recognition of manual signals using muscle tension sensors for a specific language such as Hindi is highlighted. In the

area of pattern recognition, many deals with motion recognition [13]; However, designing mechanisms is difficult [14], since it requires a great mastery of the tools; Despite this, with the appearance of block programming, programming [15] and hardware design are facilitated. These blocks are like functions that, given an input, deliver an output, which requires a way to describe them, one of these ways being recursion, which in the present case depends on probabilities [15][16]. It is here where we can relate these blocks with graphics and thus simplify the design and consequently improve the application [17] [18] [19]; thus, it is possible to attack the conversion of motion to speech based on how the signal changes in the time simply and efficiently.

The design of systems based on directed graphs or not allows establishing a correspondence even between the design of combinational, sequential, or probability-based systems and the algebra of graph theory [18].

So, a question arises. Is it feasible to develop a movement-to-speech conversion system to support people who are deprived of speech that can be carried without affecting the mobility of the user?

3. METHODS AND IMPLEMENTATION

3.1 Sections and Subsections

The BBC Micro bit is a 4x5 cm programmable embedded system [6]. It incorporates sensors such as light, temperature, movement, etc. The card enables wireless communication via Bluetooth and radio. See figure 1a. ISD devices are used for audio storage [20]. See Figure 1b. Of these devices, the main characteristic to consider is their storage capacity for store a few seconds to minutes. See figure 1b. Enough time to save many phonemes, with ranges that go from 1 to a maximum of 3 seconds. A PICAXE microcontroller is a family based on PICs. They have pre-programmed firmware that enables code to boot directly from a PC, simplifying the development of embedded systems. The PICAXE is used to control access to the phonemes stored in the integrated ISD. For wireless transmission, you can use any of the embedded modules for Arduino or, as in our case, an HC05 for Bluetooth. See figure 1c. Bluetooth included in the Microbit card could be used.

a) Micro Bit card.

b) Integrated ISD for audio storage.

c) HC 05 module

Figure 1: Main components of the system

3.2 Methodology for the development of the system

A general sketch of the system is shown in figure 2. It is made up of two elements, the Micro Bit card, and the Wi-Fi card. The Micro Bit card contains the accelerometer. As the prototypes were developed, they were modified to improve their performance. Also shown in the figure are a PICAXE microprocessor and the audio module used to concatenate the phonemes.

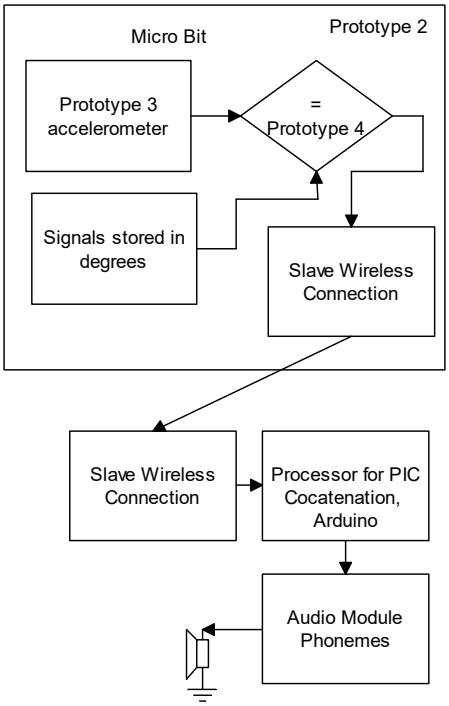


Figure 2: Diagram of the Prototype 1 system.

Prototype 1 is capable of extracting data obtained during the realization of the manual signs, the data obtained are the accelerometer values in the X and Y axes. A code corresponding to the signal is sent through the wireless connection. The codes are interpreted by the processor, so they are related to

Journal of Theoretical and Applied Information Technology
15th January 2023. Vol.101. No 1

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

the corresponding phoneme, and the sounds are concatenated to form a word.

In figure 3 you can see the procedure used by prototype 2. The variable "I" indicates when the signal processing starts. It consists of assigning values depending on the acceleration.

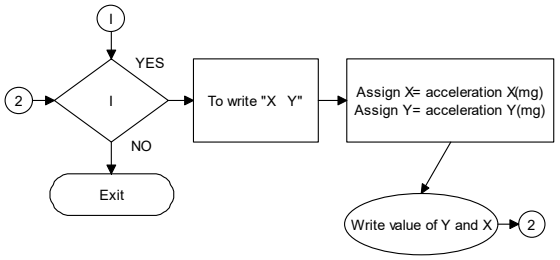


Figure 3: Prototype design 2.

Prototype 3 converts the signals from the sensor (accelerometer) into degrees of inclination for both X, and Y axes, which are sent to the comparator. See figure 4. To achieve the objective, it is not enough to use functions for linear sensors, since they move away from reality. For this reason, it is necessary to obtain a mathematical equation for said action. The output is sent to the comparator "Prototype 4".

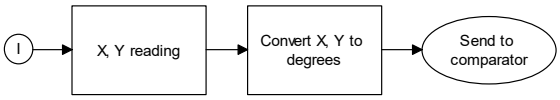


Figure 4: Prototype design 3.

Prototype 4 oversees the data analysis; The objective is to identify the signal that the user is making; which it compares with the previously stored data and in this way assigns the corresponding code to send it through the wireless connection. See figure 5.

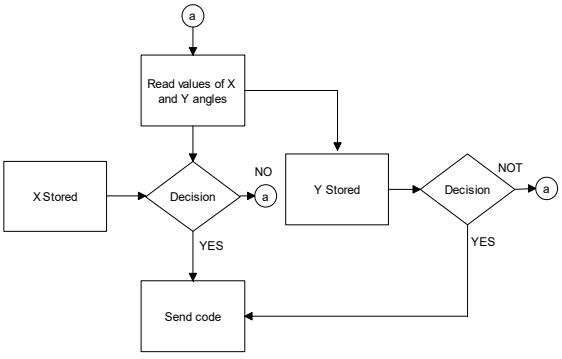


Figure 5: Prototype design 4.

The discrete function approximation problem arises from a finite set of data.

$$(x_i, Y_i)\ to\ i = 1 \ldots . N \quad (1)$$

Equation (1) represents exact or approximate values, to a function Yi. With these data, an attempt is made to construct a piecewise polynomial function r (x), such that

$$rr(x) \approx Y_i \quad (2)$$

It is called a polynomial because generally, its algebraic expression is of the form:

$$Y_i = a_0 + a_1 x + a_2 x^2 + a_3 x^3 + \cdots + a_n x^n \quad (3)$$

In figure 6 you can see the graph of the X-axis from -90 ° to 90 °. From the graph, the characteristic equation (4) is obtained as a polynomic of the 6th degree [21]. The Red line indicates the original data while the blue line is the polynomial curve.

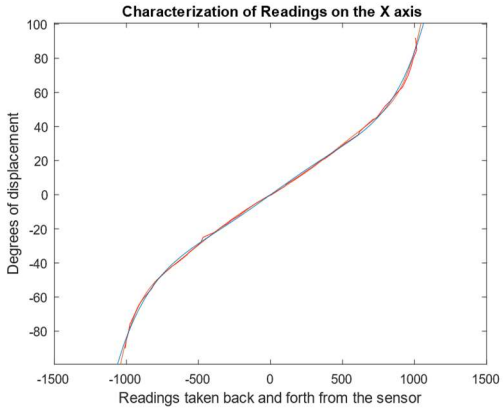


Figure 6: Plotted data of the X-axis.

Equation (4) allows converting the readings obtained by the accelerometer to degrees of inclination. The Red line indicates the original data while the blue line is the polynomial curve. See figure 6.

$$X = p_1 x^6 + p_2 x^5 + p_3 x^4 + p_4 x^3 + p_5 x^2 + p_6 x + p_7 \quad (4)$$

Where the coefficients are:

- p_1=4.8007e-18
- p_2=4.8668e-14
- p_3=-5.6517e-12
- p_4=-2.5461e-08
- p_5=9.3549e-07
- p_6=0.060532

- p_7=-0.016071.

Figure 7 shows the graph of the Y-axis from -90 ° to 90°. The characteristic equation (5) is a 6th-degree polynomial for the Y-axis.

$$Y = p_1 y^6 + p_2 y^5 + p_3 y^4 + p_4 y^3 + p_5 y^2 + p_6 x + p_7 \quad (5)$$

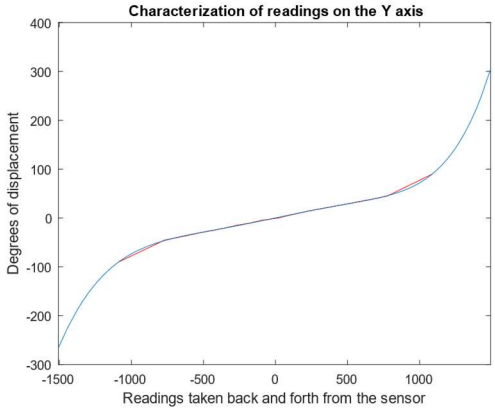


Figure 7: Y-axis plotted data.

Where the coefficients are:

- p_1=5.148e-18
- p_2=3.727e-14
- p_3=-7.232e-12
- p_4=-2.656e-8
- p_5=1.251e-6
- p_6=0.06188
- p_7=-0.1247

With equations (4) and (5) is possible to implement the conversion of acceleration to degrees of inclination that are transformed into codes associated with a phoneme.

3.3 DTW Dynamic time warp

Dynamic time warping (DTW) is a method that gets alignment between time series. When a series deforms non-linearly in time, the similarity among the time series is employed. Distance is measured by adding the distances between points matched by vertical lines. If the time series are identical, they will have DTW distances equal to zero [22]. The Euclidean distance is used in DTW. If you have time series X and Y of lengths | X | and | Y |, They are expressed by equations (6).

$$X = x_1, x_2, \dots, x_i, \dots, x_{|x|} \quad (6)$$

$$Y = y_1, y_2, \dots, y_j, \dots, y_{|y|}$$

construct a warped path W

$$W = w_i, w_2, \dots, w_k$$
$$\max(|X|, |Y|) \le K < |X| + |Y|$$

K is the measurement of the deformation route, kth is an element of the deformation path; i and j are the indexes of the time series of X and Y. The deformation way starts at w1 = (1, 1) and finished at w_K = (| X |, | Y |). The deformation lines do not overlap and all the indices from each time series must be used:

$$w_k = (i, j), \quad w_{k+1} = (i', j')$$
$$i \le i' \le i + 1, \quad j \le j' \le j + 1$$

The best trajectory occurs when there is a minimum deformation distance, which can be deduced by equation (7). Where Dist (w_{ki}, w_{kj}) is the separation of the two data point indices of time series in the k_{th} element of the deformation way.

$$Dist(W) = \sum_{k=1}^{k=K} Dist\ (w_{ki}, w_{kj}) \quad (7)$$

3.4 Voice processing – phonemes

During the test, the WAV structure was used at a sampling frequency of 11025 Hz in a single channel at 16 bits. When making the recording, the phoneme was filtered to eliminate input noise and delimit the size of the phoneme to adjust it to a time interval. It was necessary to select phonemes.

Simple and complex vowel phonemes were recorded, as well as separate vowels; phonemes were also stored with consonants in the coda position that can be isolated and recorded [23]. When there are coda phonemes with consonants p and t they are saved, along with the attack and the nucleus [23]; for Some others, it was not possible to isolate them.

3.5 Graphical modeling of the speech synthesizes

For the design, the elements that make up the system were first defined, and the mathematics used was justified, then the graphs were established, which have a degree of abstraction without affecting the clarity of the project. Figure 8 presents the graph diagram of the speech synthesizer where each node represents a function or instruction depending on the specific or generic degree with which a solution is to be expressed.

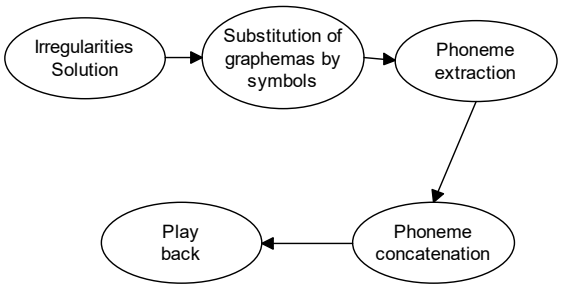


Figure 8: Graph for synthesizer components.

The Conca function has the task of concatenating the audio files into a single object, which is played. The tagged audio files are opened by reading each of the samples. These samples are then stored phonemes that you join. In figure 9, the graph of the Conca function is shown.

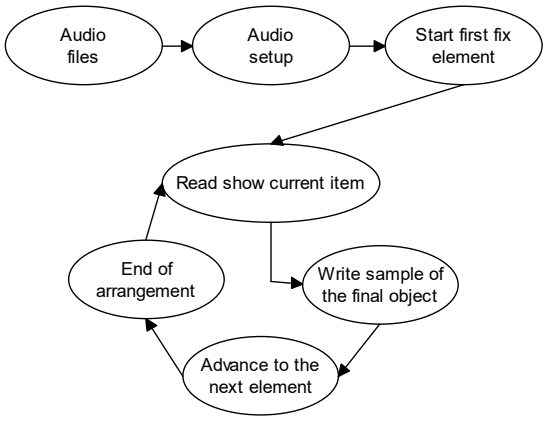


Figure 9: Conca graph.

The word or even phrase is reproduced with the possibility of adjusting the volume and balance using two variables of the module.

4. RESULTS

4.1 Obtaining data

Prototype 2 is capable of extracting data obtained during the realization of the manual signals, the data obtained is the value of the accelerometer in the X, Y, and Z axes. The signals obtained by the accelerometer are observed in figure 16 and their variation concerning the placement of the hand is verified. Only the X values in blue and Y in Red were considered. See figure 10.

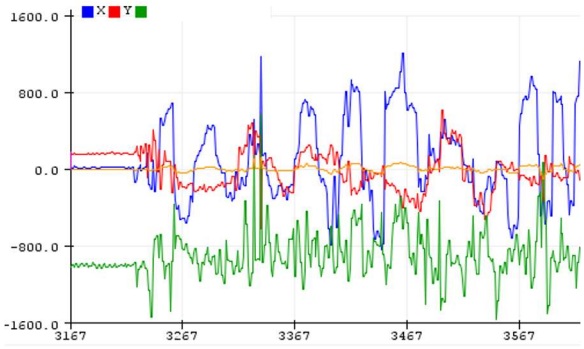


Figure 10: Prototype 2 plotted data.

4.2 Data time warping

For the comparison of signals, DTW was used. Once the degrees of the movement of the user that corresponds to a phoneme have been stored; later when a person moves these compares using DTW. In figure 11 two signals are observed two different signals (Upper graph). When performing the alignment, the Euclidean distance is large.

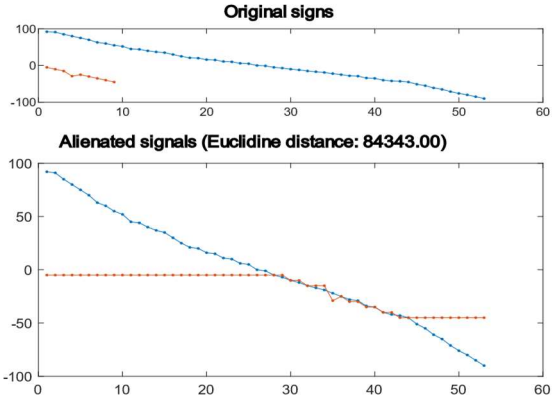


Figure 11: Different signals.

Figure 12 compares two-hand movement signals with a certain degree of similarity. The quadratic distance decreased.

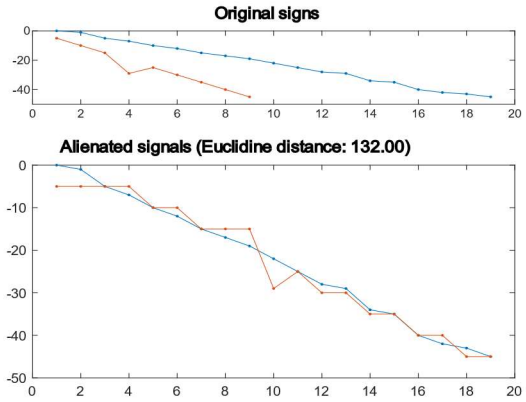


Figure 12: Almost equal signals

In figure 13 two of the same signals will be compared. It is observed that the lines overlap, and the Euclidean distance is zero. To validate it is not necessary for the distance to be zero. An acceptance threshold can be given.

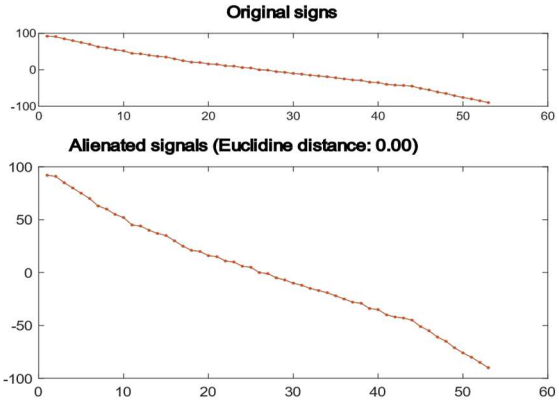
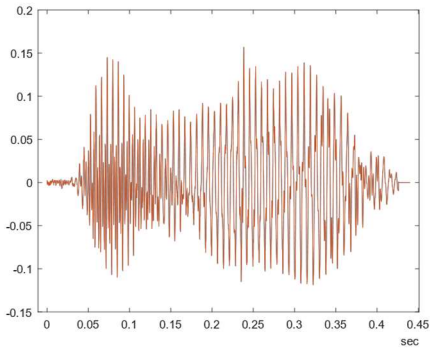


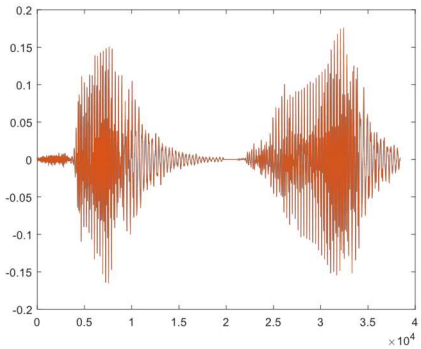
Figure 13. Equal Signals.

4.3 Phoneme generation

The voice generated by the system has an intelligible medium quality. In Figure, 14a is the signal for the word "Hello", while in Figure 14b the concatenation is shown. In the concatenated phoneme there is a reduction in the amplitude of the signal at the junction of the phonemes, which causes a sound without fluency.



(a)



(b)

Figure 14: Engraved "Hello" word

When the tests were carried out to verify how understandable the voice is, only 15.66% did not understand the phrase or word. People had no knowledge of what words they would hear. Another test consisted of producing sentences with nothing in common, but coherent, and only 0.5% failed. The general results can be seen in tables 1 and 2.

Table 1: Sentence compressibility

Words understood correctly:	253
Words misunderstood:	47
Percentage of intelligibility:	84.34 %

Table 2: Understandable of the phrases without some in common

Phrases understood correctly:	224
Phrases misunderstood:	1
Percentage of intelligibility:	99.55%

In Figure 15 you can see the hardware of the voice player. The functions seen as graphs were implemented in a microprocessor and the phoneme storage system [24]. Although Wi-Fi transmission was also used, which allows the use of an application on a cell phone or computer to generate the sound, without using the internet.

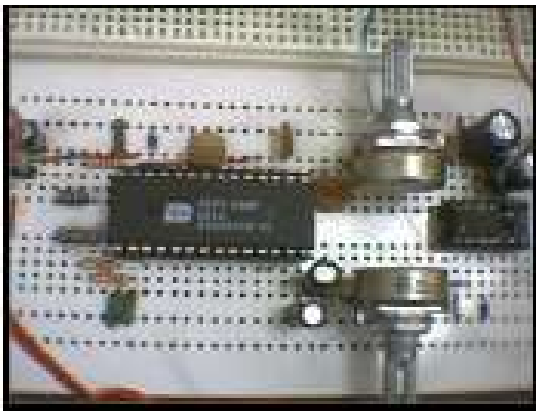


Figure 15: Physical voice reproduction system.

4. DISCUSSION

Comparing the results obtained concerning the reference articles, a system was obtained that made it possible to relate 350 movements with phonemes, a figure higher than that obtained in article 8 and that can be expanded if the gyroscope and the three coordinate axes of the accelerometer are considered. It allows mobility by not using a camera, which reduces complexity and avoids the problems generated by segmentation in a changing environment. Like the projects for the Hindu language, in our case the voice of Spanish is presented, a language that has many speakers in the world.

5. CONCLUSIONS

The main contribution of this work is to provide a system that allows people who for some reason have been deprived of their voice, to give them the means to communicate with the rest of the people, in this case in the Spanish language, where Like other countries, there are no jobs that cover this need. The project can be classified in the area of digital voice processing for the Spanish language.

Hardware modeling using graphs allows complex problems to be solved simply. This allows faster programming using blocks. This saves time in system design. In the present article, the Micro bit card was used for the programming and recognition of movements, since it has an integrated accelerometer, which simplifies the detection of movements.

When comparing the results obtained with the related works in terms of the number of identified signs, these far exceed, since up to 350 identified signs are reached. In addition to obtaining an application that gives the possibility to people with the impossibility of speaking to be able to communicate by producing phonemes through the movement of the hands.

The use of graphs and block programming allowed the generation of a system with few components and simplicity in its assembly. This way of developing projects is one more contribution to the project.

DTW turned out to be an efficient way of comparing hand movements. However, it is necessary to generate a motor code of hand or hand movements for the production of different words that expand the words that increase the dictionary used in this project.

Regarding the objectives, the database of 350 Spanish phonemes was established, using a design based on graphs that concatenate the phonemes, which generates words related to hand movements, thus meeting the objectives; however, it is necessary to generate a language based on movements that are easy to understand and use by people who lack speech, to achieve its universality not only for the deaf world but also for anyone.

FUTURE WORK

In future work; Use the Z axis of the accelerometer and the gyroscope to generate a greater number of phonemes and get the bases to obtain movements for a universal language.

ACKNOWLEDGMENT

The authors acknowledge the support received to carry out this project from the IPN - (Instituto Politécnico Nacional).

REFERENCES:

[1] K. Nishi and J. Miura, "Generation of human depth images with body part labels for complex human pose recognition," Pattern Recognition, vol. 71, pp. 402-413, 2017.

[2] Laín Entralgo, Pedro. "El cuerpo humano: Teoría actual.", Ed. Espasa Universida, 1989.

[3] J. M. Palacios, C. Sagüés, E. Montijano, S. Llorente, “Human-Computer Interaction Based on Hand Gestures Using RGB-D Sensors” 2013.

[4] Jairo A. Vélez Pérez et al, "ASISTENTE DOMÓTICO A DISTANCIA MEDIANTE LA APLICACIÓN DE TÉCNICAS DE RECONOCIMIENTO DE VOZ," Revista De

Investigaciones Universidad Del Quindío, vol. 27, (2), pp. 48-53, 2015.

[5] Arias López, Luz Amparo2012, *"*Biomecánica y patrones funcionales de la mano*"*. Morfolia; Vol. 4, núm. 1 (2012) 2011-9860.

[6] Simon Monk, "Programming the BBC micro: bit, Getting Started with MicroPython", Ed. Mc Graw Hill, USA, 2018

[7] L. Hyunchul, J. Chung, C. Oh, S. Park, J. Lee y B. Suh, "Touch+Finger: Extending Touch-Based User Interface Capabilities with "Idle" Finger Gestures in the Air", Berlin, Germany, 2018.

[8] L. Jing, Y. Zhou, Z. Cheng y T. Huang, "Magic Ring: A Finger-Worn Device for Multiple Appliances Control Using Static Finger Gestures", 2012.

[9] J. Luan, T.-C. Chien, S. Lee y P. Chou, "HANDIO: A Wireless Hand Gesture Recognizer Based on Muscle-Tension and Inertial Sensing", California, 2016.

[10] OSTER, Elvis C. *Software Engineering: A Methodical Approach*. Auerbach Publications, 2021.

[11] E. G. Puerto Cuadros and J. L. Aguilar Castro, "Un algoritmo recursivo de reconocimiento de patrones", *Revista Técnica,* vol. 40, *(2)* pp. 95, 2017.

[12] B. F. Pimentel, "Synthesis of FPGA-Based Accelerators Implementing Recursive Algorithms", ProQuest Dissertations Publishing, 2009.

[13] J. J. Medel Juárez and M. T. Zagaceta Álvarez, "Estimación-identificación como filtro digital integrado: descripción e implementación recursiva", *Revista Mexicana De Física,* vol. 56, *(1),* pp. 1-8, 2010.

[14] León, Gorka García. "Programación en Scratch con «M» de matemáticas. *Aula de innovación educativa* 302 (2021): 71-72.

[15] Vidal, Cristian L., et al. "Experiencias prácticas con el uso del lenguaje de programación Scratch para desarrollar el pensamiento algorítmico de estudiantes en Chile." *Formación universitaria* 8.4 (2015): 23-32.

[16] Fagerlund, Janne, et al. "Computational thinking in programming with Scratch in primary schools: A systematic review.*Computer Applications in Engineering Education* 29.1 (2021): 12-28.

[17] Pratiwi, Umi, and Dwi Nanto. "Students' Strategic Thinking Ability Enhancement in Applying Scratch for Arduino of Block Programming in Computational Physics Lecture. *Jurnal Penelitian & Pengembangan Pendidikan Fisika* 5.2 (2019): 193-202.

[18] Gross, Jonathan L., Jay Yellen, and Mark Anderson. *Graph theory and its applications*. Chapman and Hall/CRC, 2018.

[19] JACOBSON, Lvar; BOOCH, James Rumbaugh Grady. The unified modeling language reference manual. 2021.

[20] López Vara Julio Sebastian, "Estudio analítico y experimental de los Circuitos integrados de voz ISD", Universidad Austrial de Chile, 2005.

[21] González, Héctor Jorquera, and Claudio Gelmi Weston. Métodos numéricos aplicados a la ingeniería: Casos de estudio en ingeniería de procesos usando MATLAB®. Alpha Editorial, 2016.

[22]Yadav, Munshi, and M. Afshar Alam. "Dynamic time warping (DTW) algorithm in speech: a review." *International Journal of Research in Electronics and Computer Engineering* 6.1 (2018): 524-528.

[23]Letowski, Tomasz R., and Angelique A. Scharine. *Correlational analysis of speech intelligibility tests and metrics for speech transmission*. US Army Research Laboratory Aberdeen Proving Ground United States, 2017.

[24] German Sarmiento, "Picaxe Manual de Programación: Como programar Microncontroladores Picaxe". CreateSpace Independent Publishing Platform, 2015.