ISSN 1870-4069

Automatic Recognition of Indigenous Languages from Different Mexican Geographic Regions Using Long-Term Average Spectrum

Luis David Huerta-Hernández[1], Ricardo Melchor[1],
Julio Cesar Ramírez-Pacheco[3], Homero Toral-Cruz[3], Khalid S. Aloufi[2],
José Antonio León-Borges[3]

[1] Universidad del Istmo,
Mexico

[2] Taibah University,
College of Computer Science and Engineering,
Saudi Arabia

[3] Universidad de Quintana Roo,
Mexico

{luisdh2, ricardo.melchor}@bianni.unistmo.edu.mx,
alawfikhalid@gmail.com,
{jleon, juliocr, htoral}@uqroo.edu.mx

Abstract. In this paper, a method for automatic recognition of Mexican Indigenous Languages (MIL): Maya, Mixteco, Zapoteco, Mixe, Nahuatl, Tarahumara, Mazahua, Tseltal, Chichimeco, and Huichol is proposed. The Long-Term Average Spectrum (LTAS) is employed as a feature for the recognition process. Moreover, the performance of classifiers such as Multi-Layer Perceptron (MLP), Sequential Minimal Optimization (SMO), Random Forest (RF), Naive Bayes, and K-Nearest Neighbor (KNN) are also highlighted. In order to reduce speech vector features, LTAS sequences, extracted from audios are first passed on *BestFirst* filters. In the experiments, high performance for MIL recognition was achieved, using a simplified codification scheme of the voice, with vectors features with a low number of values. The method is remarkable for its simplicity and effectivity, bringing away on no-tested languages in the speech processing area.

Keywords: Speech processing, recognition, data mining, long term average spectrum, classifiers, Mexican indigenous languages.

1 Introduction

The continuous development of technologies requests for more robust applications capable to generalize tasks, including most possible cases found in the real world. In this sense, the Automatic Speech Recognition (ASR) area, since some years ago, awoke the interest between researchers for the Automatic Dialect Identification (ADI) [1].

Many countries have been dedicating efforts to the research of ADI [1, 2], recent studies have been presented for a variety of dialects [2], Arabic [3], and Malayalam [4].

Many are the benefits for integrate with ASR the languages representatives for each country, which the importance is high too. Many people could be benefited from more robust ASR, including dialect languages. Applications as the automatic call center and improves enriched indexed of spoken documents are discussed in [1], improving and enriching of ASR engines, as well as characterizing of speaker traits and Data Mining processes [2].

In socio-political context, the United Nations declares about the importance and promoting of multilingualism: “it is urgent to *take action to promote multilingualism*, in other words, to encourage the development of coherent regional and national language policies which give opportunity for the appropriate and harmonious use of languages in a given community and country" [5].

In Mexico, the efforts to research new technologies and their endemic languages are scarce. Although there was focused research on the textual analysis of indigenous languages as *Nahuatl* [6], there are not enough research aimed to integrate the indigenous languages with the ASR.

On the other hand, most of the literature makes the ADI using the typical features as MFCC, leading in well-detailed information extracted from the speech, but on the other side, the number of parameters is high. These values of the speech scheme codification must be processed by filters and classifiers, affecting the time of values extraction, filtering, and training, as well as the precision of the classification [7, 8], for mention some disadvantages. As well, many efforts are done to reduce speech parameters [9] [10]. So, we need to test alternative speech codifications. Speech codifications scheme as LTAS have been scarcely explored on language recognition, hence, in this work, a method that uses LTAS as a speech codification scheme is presented. One direction of this study is to determine if LTAS is highly useful to make the ADI at the same time the number of extracted values is reduced.

One motivation of this study is to start the basis for ADI of Mexican Languages, encouraging this research area in the country, due to the limited or null studies in this sense, at the same time to propose a new method, using a low number of speech values for the classification, reducing the complexity of data and timely response, as well as in-crease the probability to enhance the classifiers accuracy. As well, is important to experiment with the recognition based on the vocal tract, the way the speech is produced, where the LTAS fits in this sense.

The aim of this study is to propose a method to make ADI, using some representative Mexican Indigenous Languages of different geographical regions, answering the following questions: Is LTAS useful to recognize MIL? The use of LTAS promotes the use of a reduced number of values extracted from speech, in ADI?, Which of the classifiers are best suited in order to classify MIL?

The paper is organized as follows: Section 2, exposes the related works citing the classifiers and features used. Section 3, a brief explanation of the common features used on ADI, is presented. Section 4 exposes the database description and the geographical representation of the experimental dialects. Section 5, the experimental results of MIL recognition, are presented. Section 6, the Discussion, presenting the similarity of the classifiers used in state-of-art works and the present study; remarking the low number of values used to perform recognition, for the presented method. Section 7. Conclusions and Future works.

2 Related Works

In this section, related works about ADI are exposed. In respect to Mexican Indigenous Languages recognition, there are not works reported in the literature of the state of art.

The works will be revised briefly, with respect to the methods, the features, and the number of values used to recognize the dialects. These works will be cited too, in the Discussion section. The methods used above the 300 values per second and MFCC features, Mel energies, energy, pitch, or a mix of them.

2.1 ADI based on MFCC as the Main Feature

A comparative study for classifiers to recognize Malayalam dialect has been reported in [4], using Thrissur and Kozhikode corpora. Malayalam dialect is spoken in the south of India, concretely, in Kerala state. The study is based on two of the fourteen Malayalam dialects. Speech files of 15 speakers for each dialect were used, who read 30 sentences with 3 three or a maximum of six words.

The features used were MFCC (detailed in *Speech Features* section), energy, and pitch. The speech signals were pre-processed before the feature extraction, to enhance the accuracy and efficiency of the extraction process. In this work, the MFCC extraction is exposed and citing the use of 10-30 *ms* size frames. In each frame several coefficients are extracted, typically 12-13 coefficients. Since the precise number of coefficients and frame size are not detailed, assuming the use of 12 coefficients and 20 *ms* frame size, for each second there are 50 coefficients vectors with length 12. Simple calculating, 12×50, having 600 values of MFCC per second.

This method, additionally uses energy and pitch features, resulting in above those 600 values per second. Artificial Neural Net classifier produced a recognition accuracy of 90.2 %, Support Vector Machine of 88.2%, and Naive Bayes of 84.1%.

The study reported in [2], is approached to analyze the differences between reading versus spontaneous speech Arabic dialect. To analyze the differences, a Gradient Mixture Model (GMM) was used. The method is based on MFCC features, using a 20 *ms* frame size, estimating 600 values per second.

2.2 ADI based on Mel Filter Banks as the Main Feature

Automatic Arabic dialect identification was reported in [3], were proposed methods to discriminate between the five most widely used dialects of Arabic, with an accuracy of 59.2%. Using classifiers as Naïve Bayes, Support Vector Machines, and Deep Neural Net. Different dimension vectors in the range 300-1600 were used. For instance, DNN with 5 hidden sigmoid layers, where the first layer used 23 critical band energies obtained from Mel scaled filter-bank. Since the Mel scale uses frames too, and frame size is about 20-40 *ms*, if a 40 *ms* frame is used, then there is 23×25, 575 values per second (estimated).

Chinese dialect recognition was reported in [7], identifying the 10 most wide spreads Chinese dialects. Recurrent Neural Nets and 40-dimensional Mel filter bank coefficients with a frame size of 25 *ms* (40 per second), were used. Estimating 40×40, 1600 coefficients per second. The dialect recognition achieved approximately 90% of accuracy.

In this section, was observed the common use of MFCC and Mel filter bank coefficients to perform ADI, an approach based on the auditory human system. At the difference, the study reported here was done using the LTAS, which models the vocal tract and a simple and reduced number of values per signal, detailed in the Experiments section.

3 Speech Features

The extracted features from audios are fundamental in the ASR performance, for this reason, it takes a great relevance to know and make mention about the benefits and usage of them.

3.1 Time-Domain Speech Features

Energy is a common and frequently used speech time-domain feature. The energy is obtained by the summary of the signal amplitude squares, and its measure unit is given in *Pa2 s* [8]. being *x(t)* the amplitude of the sound, given on Pascal's *Pa*, then the energy is defined in (1):

$$\int_{-\infty}^{\infty} x^2(t). \tag{1}$$

The fundamental frequency (F0) is another time domain feature, also known as pitch, is the vibratory frequency of the vocal cords. The number of air pressure oscillations per second determines the sound pitch. The pitch is sampled into many frames centered on equally spaced time.

3.2 Mel Spectrums

The Mel spectrums vector is obtained as a result of passing a signal through a filter bank. Every spectrum contained in the vector is the result of filtering the input spectrum through an individual filter, being the length of the vector equal to the number of filters. The triangular filters are centered over the frequency axis, distributed on the no lineal Mel scale; this scale was initially proposed by Stevens and Volkman in 1940 [9].

The filters banks emulate the perceptual critical band (Fig 1.), accentuating the low frequencies. The edges of the filters coincide with the axis of the adjacent filters. A common model for the relations of Mel frequencies and lineal scales is expressed in (2):

$$Mel\ Frequency = 2595 \times log_{10}\left(1 + \frac{Lineal\ Frequency}{700}\right). \tag{2}$$

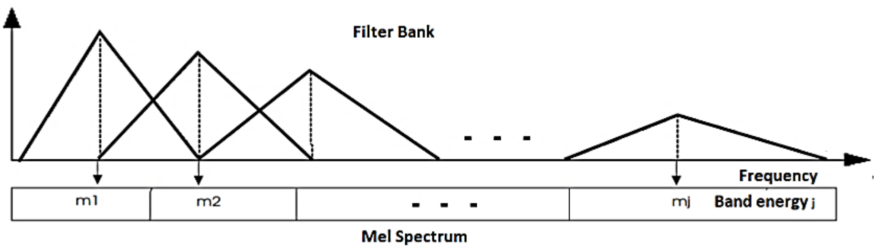

Fig. 1. Mel filter bank.

The audio codification scheme has been used in phonetic speech segmentation [10], also the Mel scale is the base for another widely used scheme in speech processing, the Mel Frequency Cepstral Coefficients (MFCC).

3.3 Mel Frequency Cepstral Coefficients

Mel Frequency Cepstral Coefficients is one of the most used sound representations in ASR and audio engineering. The process to obtain the MFCC starts applying a Hamming window, to avoid spectral distortions, fragmenting the continual signal in frames, where these frames typically are of 20 or 30 *ms*, because in this length the signal assumes to be stationary.

The logarithm of the energy in each filter is calculated and accumulates before to be applied the discrete cosine transform, producing the vector features of MFCC. The cepstral analysis denotes an unusual treatment of the signal in the frequency domain as if it were on the time domain [11]. The unit measure in the cepstral domain is in seconds, but they represent the spectral variations of frequencies.

Most of the audio processing studies have used this scheme codification: in emotion recognition, speech, and speakers. Studies for ADI, MFCC were used too in [2, 4].

3.4 Long Term Average Spectrum

The Long-Term Average Spectrum (LTAS) represents the power spectral density (PSD), expressed on *dB*/*Hz* relative to 2x10^{-5} *Pa* [8].

The spectrum is computed based on the PSD which is obtained from overlapped FFT series. Typically, the FFT length is 4096, step 2048. The segments of X are obtained applying the Hann-Window. The PSD average is smoothed with Gaussian function [12]. The average sound power between a range of time (t_1, t_2), is obtained by (3).

$$\int_0^F PSD(f)df = \frac{1}{T}\int_{t_1}^{t_2} |x(t)^2| \cdot dt. \tag{3}$$

The LTAS brings a representation for the location of the vocal tract resonances and the glottal source [13], which is more suitable to study the way that the people speak. On the other hand, the MFCC represents a human auditory system, the way that we perceive sounds.

Fig. 2. The study area of Mexican indigenous languages.

4 Database and Study Area

The audio samples were obtained from [14]. The audios have a length between 1 and 3 seconds, speaking phrases as: "come in", "good morning", "good afternoon", "good night", "welcome", "thank you for visiting us", "enjoy this meeting", "excuse me", "come back soon". Maya, Mixteco, Tseltal, Mixe, and Mazahua were spoken by female natives. Nahuatl, Zapoteco, Huichol, Tarahumara and Chichimeco were spoken by males' natives. The audios were recorded with a sampling frequency of 22100 Hz. The samples used were: 12 samples of Maya, Mixteco and Mixe respec-tively; 14 samples of Nahuatl, Tseltal, Huichol, Tarahumara, Mazahua and Chi-chimeco for each one; and 15 samples of Zapoteco. The phrases were spoken by five men and five women (one speaker per language). The indigenous languages used in this study, cover every zone of the Mexican territory. The geographical study area of Mexican indigenous languages is shown in Fig. **2**: Maya, Mixteco, Zapoteco, Mixe, Nahuatl, Tarahumara, Mazahua, Tseltal, Chichimeco, and Huichol.

5 Experiments

This study is based on the premise, every language adjusts the vocal tract in a particular way, to express its words and elocutions. This idea is supported by previous studies, for example, because "aerodynamic and anatomical properties of the vocal tract, influence in shape and patterning of the speech sounds" [14]. In this sense, the use of the LTAS fits as the main feature to recognize the patterns of languages, in our case, the Mexican Indigenous Languages.

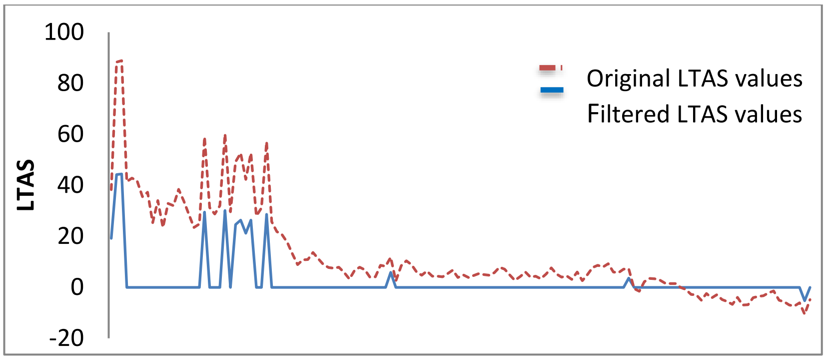

Fig. 3. LTAS values *vs* LTAS filtered values.

Table 1. Classifier's performance using five classes and *bin width* variations.

Classifier	80 Hz 276 LTAS values	100 Hz 221 LTAS values	140 Hz 158 LTAS values
J48	95.52	95.52	77.61
KNN3	91.04	89.55	88.05
KNN	89.55	94.02	94.02
MLP	94.02	95.52	94.02
RF	89.55	89.55	89.55
Random Tree	58.2	59.7	68.65
SMO	94.02	94.02	95.52
Naive Bayes	88.05	85.07	88.05
Bayes Net	79.1	79.1	77.61

The Long-Term Average Spectrum (LTAS) was extracted from speech signals using PRAAT software version 6.0.39 [8]. The classifiers are provided by the Waikato Environment for Knowledge Analysis version 3.8.5, as well as the attribute filter called *BestFirst* detailed in [15].

5.1 Experimental Classifiers Settings

The settings of the classifiers with the best performance, used in these experiments are presented. The same settings were used for five and 10 classes recognition.

The KNN was set using *Euclidean distance*, *no distance weighting*, *with 1 and 3 Neighbors* denoted as KNN and KNN3 respectively.

The MLP was set 0.3 on *learning rate, 0.2* on *momentum*, and *500 epochs.* One single hidden layer was used in the experiments. The *(classes + input size)/2* criteria, in order to determine the number of neurons in the hidden layer was used. So, for instance, using five classes and 140 *Hz* for LTAS extraction, there were 13 features and 5 classes (Table 2), in this case, the neurons in hidden layers were 9.

The SMO used a *multinomial logistic regression,* the complexity parameter was set in 1, an *epsilon of 1.0E-12* for round-off error.

For the Naïve Bayes classifier, an especial setting is not required.

Table 2. Classifier's performance with filtered values, using five classes and *bin width* variations.

Classifier	**80 Hz 22 LTAS values**	**100 Hz 21 LTAS values**	**140 Hz 13 LTAS values**
J48	95.52	95.52	80.59
KNN3	98.5	98.5	**98.5**
KNN	98.5	97.01	**98.5**
MLP	98.5	98.5	**100**
RF	98.5	98.5	**98.5**
Random Tree	83.58	74.62	76.11
SMO	98.5	98.5	**98.5**
Naive Bayes	97.01	95.5	97.01
Bayes Net	98.5	98.5	95.52

Table 3. Classifier's performance over 10 classes.

Classifier	**140 Hz 158 LTAS values**	**140 Hz 18 LTAS filtered values**
J48	77.77	82.22
KNN3	85.18	91.11
KNN	90.37	90.37
MLP	92.59	**94.07**
RF	90.37	93.33
Random Tree	62.22	80.00
SMO	91.85	**94.07**
Naive Bayes	87.4	**94.07**
Bayes Net	74.81	88.14

5.2 Five Languages Classification

In this experimental phase, the LTAS from speech audios were extracted and used as the only feature in the classification process. In the first instance, the classification for indigenous languages as Maya, Zapoteco, Mixteco, Tsoltil, and Nahuatl, was experimented with. The previous languages selection was aleatory.

The performance of different classifiers tested with five classes and three sampling frequencies is observed in Table 1. The bandwidth (frequency step) used were 80Hz, 100Hz and 140 Hz, obtaining 276, 221, and 158 spectrums per signal, respectively. The classifier's performance difference with the variations of spectrums number was minimal.

To enhance the classification performance, a *BestFirst* filter was applied to the LTAS values of the signal. This filter gets the relevant values of LTAS, those forming the peaks, discarding those values with light changes (flat lines). The peaks are plotted on the blue line (down), the original values of LTAS are plotted on the red line (dotted), see Fig. 3.

Testing with filtered LTAS values, the Table 2 results were obtained, where a performance increase for all classifiers is observed, mainly Bayes Net followed by KNN. The *BestFirst* filter reduces the number of LTAS values (less than 10% respect the original number of them) down to 22, 21, and 13 LTAS values respectively. The MLP performance is notorious, classifying correctly the 100% of the samples, using only 13 values per signal, followed by the SMO, Random Forest with 98.5% of correct classification.

There is high performance for almost every one of the classifiers.

5.3 Ten Languages Classification

In this phase, the method was passed over a more rigorous test, to observe its behavior adding five classes over the existing.

The languages added were Huichol, Mixe, Tarahumara, Mazahua, Chichimeco. Starting on the previous results in Table 2, where high performance was using a 140 Hz *bin width*, and minimal values were required. The experiments with 10 classes were done only on this *bin width.* The results of this phase are shown in Table 3, where the best classifiers were MLP, SMO, and Naive Bayes with 94.07% of correct classification.

5.4 Overall Performance for Each Language

The languages were complicated to classify, we expose Table 3. The most recognizable language, using this set of classifiers was *Chichimeco* and the least recognizable was *Mazahua.* On the header of the columns, the total of samples and the letter assigned for each language is shown, so we obtained the performance per language with a simple division between the total value (bottom of the table) and the number of samples on the header, this las value multiplied by 10 (number of classifiers).

The idea of this simple overview is, to have an initial perspective about the complexity of recognition for each language.

6 Discussion

Similar research has been done to recognize different dialect languages. The discussion is about the methods used, although an exact comparative is difficult because of the different databases used.

A Malayan dialect recognition system was exposed in [4]. The Thrissur and Kozhikode were the two kinds of Malayan dialects used in the recognizer task. The MFCC, energy, and pitch were the features extracted, and then processed for the classifiers Artificial Neural Net, Support Vector Machine, and Naive Bayes, reaching

Table 4. Classifier's performance over 10 classes.

Classifier	**a 12**	**b 12**	**c 14**	**d 14**	**e 15**	**f 14**	**g 12**	**h 14**	**i 14**	**j 14**
J48	9	11	9	12	13	11	9	13	10	14
KNN3	11	10	11	14	14	13	9	14	13	14
KNN	10	10	13	13	12	13	9	14	14	14
MLP	10	12	14	14	13	13	11	13	13	14
RF	10	11	13	13	15	13	10	14	13	14
RT	9	10	11	11	12	12	8	10	11	14
SMO	11	11	13	14	14	13	10	14	13	14
NBAYES	10	11	14	14	13	13	11	13	14	14
BNET	10	10	12	11	13	12	10	14	13	14
Total	90	96	110	116	119	113	87	119	101	126
Performance	75.00	80.00	78.57	82.86	79.33	80.71	72.50	85.00	72.14	90.00

a=Maya,b= Mixteco, c=Nahuatl, d=Tseltal, e=Zapoteco, f=Huichol, g=Mixe, h=Tarahumara, i=Mazahua, j=Chichimeco

a recognition accuracy of 90.2%, 88.2%, and 84.1% respectively. Our experiments match exactly with the same kind of classifiers: MLP, SMO, and Naive Bayes; citing SMO is an algorithm used to train Support Vector Machines.

Arabic automatic dialect detection was presented in [3], reported results where they discriminated between the five most widely used dialects of Arabic: namely Egyptian, Gulf, Levantine, North African, and MSA, with an accuracy of 59.2%. Experimenting with classifiers like Naive Bayes and SVM (between others), obtaining the best performances with the second, achieved 100% of accuracy for binary classification between English vs Modern Standard Arabic. Highlight the use of SVM and Naive Bayes, as classifiers selected for dialect recognition.

Also, a dialect database that contains 10 types of Chinese dialects was used in [7], using a Recurrent Neural Network with an accuracy of 90.04%.

On the side features, the LTAS has been useful in the recognition, in topics as recognition of speaker [13], gender, age, including diseases [16], and forensic usage [17], in a general way, LTAS are speech encoding related with the voice quality [18]. In this study, LTAS brings a high performance for MIL recognition, using 158 numeric values per signal, and only 18 values using the *BestFisrt* filter.

7 Conclusion

In this work a method based on Long Term Average Spectrum was presented. The proposed method can be used to recognize the main Mexican Indigenous Languages.

Our studies have revealed that LTAS can be considered as a promising feature of dialect or language recognition because used a low number of parameters to describe the speech audios, which leads to short vector inputs for classifiers. The LTAS parameters can be optimized by considering the peaks of its values, with the aim of reducing the information from 158 values per signal to 18 values and consequently decrease the load on the classifiers. There are various works related to languages recognition, however these works use above 300 values per second as input on the classifiers, while our method uses below of 30 values per signal. As a result of our experiments, we found that MLP, SMO, and Naïve Bayes bring the best performance with an accuracy of

94.07%, in the 10 MIL recognition task. Besides, Networks, Support Vector Machines and Naive Bayes were found as common classifiers in state of art studies presented here and our experiments, to perform dialect recognition. As future work we will test on a most robust database, increasing the number of samples per language, adding more languages and we will test other types of classifiers as deep learning, hybrid approaches, filters, and new codification scheme of the speech.

References

1. Gray, S., Hansen, J.: An Integrated Approach to the Detection and Classification of Accents/Dialects for A Spoken Document Retrieval System. In: IEEE ASRU-2006, pp. 35–40 (2006)
2. Liu, G., Lei Y., Hansen, J.: Dialect Identification: Impact of Differences Between Read Versus Spontaneous Speech. In: EUSIPCO´10: European Signal Processing Conference, pp. 2003–2006 (2010)
3. Ali, A., Dehak, N., Cardinal, P.: Automatic Dialect Detection in Arabic Broadcast Speech. Proc. Interspeech, pp. 2934–2938 (2016)
4. Sunija, A.P., Rajisha, T., Riyas, K.: Comparative Study of Different Classifiers for Malayalam Dialect Recognition System. ScienceDirect, 24, pp. 1080–1088 (2016)
5. United Nations: United Nations (2008) https://www.un.org/en/events/iyl/multilingualism. shtml.
6. Martínez, C., Zempoalteca, A., Soancatl, V.: Computer Systems for Analysis of Nahuatl. Research in Computing Science, (47), pp. 11–16 (2012)
7. Pappu V., Pardalos, P. M.: High Dimensional Data Classification. Springer Optimization and Its Applications, pp. 34 (2013)
8. Othman, A., Hasan, T.: Impact of Dimensionality Reduction on the Accuracy of Data Classification. In: 3rd International Conference on Engineering Technology and its Applications (IICETA), pp. 128–133 (2020) doi: 10.1109/IICETA50496. 2020.9318955.
9. Hassan, M., Nath B., Bhuiya, M.: Bengali Phoneme Recognition: A New. In: 6th International Conference on Computer and Information Technology (2003)
10. Cheng, H., Ma X., Yugong, X.: A Study of Speech Feature Extraction Based on Manifold Learning. In: Journal of Physics: Conference Series, 1187(5), pp. 052021 (2019)
11. Zongze, R., Guofu Y., Shugong, X.: Two-stage Training for Chinese Dialect Recognition. Proc. Interspeech (2019)
12. Boersma P., Weenink, D.: Praat: Doing Phonetics by Computer [Computer program]. Version 6.1.50 (2020) url: http://www.praat.org/.
13. Volkmam J., Steven, S.: The Relation of Pitch to Frequency. American Journal of Psychology (1940)
14. Huerta, L., Huesca J., Contreras, J.: Speech Segmentation Algorithm based on Fuzzy Memberships. International Journal on Computer Science and Information Security, pp. 229–234 (2010)
15. Tukey, J., Bogert, P., Healy, M.: The Quefrency Analysis of Time Series for Echoes: Cepstrum, Psuedo-Autocovariance, Cross-Cepstrum and Saphe Cracking. Proceedings of the Symposium on Time Series Analysis (2006)

16. Hummersone, C.: Calculate the Long-Term Average Spectrum of a Signal (2021) https://github.com/IoSR-Surrey/MatlabToolbox.
17. Kinnunen, T., Hautmaki, V., Franti, P.: On the Use of Long-Term Average Spectrum in Automatic Speaker Recognition. In: International Synposium on Chinese Spoken Language Processing (ISCSLP) (2006)
18. Ohala, J.: The Origin of Sound Patterns in Vocal Tract Constraints. The Production of Speech, Springer, pp. 189–216 (1983)
19. Kohavi R., George, J.H.: Wrappers for Feature Subset Selection. Artificial Intelligence, 97 (1) pp. 273–324 (1997)
20. Cukier-Blaj, S., Camargo Z., Madureira, S.: Longterm Average Spectrum Loudness Variation in Speakers with Asthma, Paradoxical Vocal Fold Motion and Without Breathing Problems. In: Proceedings of the Fourth Conference on Speech Prosody, No. 9780616220, 44, pp. 41 (2008)
21. Rose, P.: Forensic Speaker Identification. CRC Press (2002)
22. Pittam, J.: Voice in Social Interaction: An interdisciplinary approach. London: SAGE Publications (1994)
23. Sunija, A., Rajisha T., Riyas, K.: Comparative Study of Different Classifiers for Malayalam Dialect. In: International Conference on Emerging Trends in Engineering, Science and Technology, 1088, pp. 1080 (2015)
24. Insituto Nacional de Lenguas Indígenas: Prontuarios de frases de cortersía de Lenguas Indígenas (2010) https://site.inali.gob.mx/Micrositios/Prontuarios/index.html.