Telecommunications and Radio Engineering, 72(6):495-504 (2013)

PROPOSED A NEW APPROACH FOR VOICED / UNVOICED DECISION OF SPEECH FILE USING LAGRANGE TECHNIQUE

Nidaa F. Hassan ♦ ***& Hala Bahjat Abdul Wahab***

University of Technology, Baghdad, Iraq

♦Address all correspondence to Nidaa F. Hassan E-mail: nidaaalalousi@yahoo.com

In speech analysis, voiced/unvoiced decision is usually performed in extracting the information from the speech signals. This paper presented a new approach for voiced/unvoiced (V/UV) decision; the approach is used Lagrange polynomial interpolation for voiced/ unvoiced signals speech decision, depending on the advantage of the characteristics of the interpolation techniques. Speech file is divided in different size of blocks, each block is evaluated by Lagrange technique, and the results are evaluated by one of most popular voiced/unvoiced measured which is the short time energy classifier. The results show that Lagrange technique worked well as classifier and give an accepted decision to analysis the speech to voice and unvoiced sample, especially with small size non overlapping blocks.

KEY WORDS: *Speech Recognition, Lagrange interpolation, voiced/unvoiced decision*

1. INTRODUCTION

The problem of voiced/unvoiced speech determination is an important one and has been worked on extensively by researchers [2,6] during the last three decades. In [1-3] a statistical parametric method was proposed whereas in [4-6] non-parametric methods based on linear discrimination functions, multi-layer feed forward and recurrent neural networks were adopted. In [7], a two channel approach which made use of the speech and electroglottogram signals was pursued. Most of the above methods proposed for voiced/unvoiced classification were implemented and tested in quiet.

Voiced/unvoiced classification in noise, however, is a far more challenging task since the noise can potentially mask low-energy speech segments such as fricatives (e.g., /f/, /th/) and stop-consonants (e.g., /b/, /d/). Also, most of the above methods utilized long analysis frames with some [8] using as large as 40-ms duration frames.

In this paper, a new approach for voiced/unvoiced speech classifier is proposed; the classification is exceeded the difficulties in the classification of small blocks that we had encountered with the short time energy classifier.

0040-2508/13/$35.00

2. LAGRANGE POLYNOMIAL INTERPOLATION

The Lagrange interpolating polynomial is the polynomial that passes through all pre-defined points. The formula was first published by Waring (1779), rediscovered by Euler in 1783, and published by Lagrange in 1795 (Jeffreys and Jeffreys 1988) [8]. For example, in mathematics, it is used in the construction of the Newton-Cotes formulas [9]. The constant problem in Lagrange interpolation is a tradeoff between having a better fit and having a smooth well-behaved fitting function. Also the number of data points must be optimal. It is well known that when constructing interpolating polynomials, the more data points that are used in the interpolation cause the higher degree of the resulting polynomial and the greater Oscillation in interpolation function between the data Points. In that way, a high-degree Lagrange interpolation may predict the function between points with greater error, although the accuracy at the data points will be perfect, and for this reason simple Lagrange polynomial interpolation is adapted in this paper to overcome any predictable errors.

2.1 Definition

First we will define mathematical fundamentals of the Lagrange interpolation[1]. Let us suppose that A is a field and B is a vector space over C. Elements of B is vectors and elements of C are scalars. If $a_1,...,a_n$ are scalars and $b_1,...,b_n$ are vectors, then the *linear combination* of those vectors with those scalars as coefficients is:

$$a_1b_1 + a_2b_2 + a_3b_3 + ... + a_nb_n . \quad (1)$$

A and B are specified explicitly. A linear combination of the vectors $b_1,...,b_n$ with the unspecified coefficients (that must be in space C) are often in practice [1,3]. If S is a subset of space C, we may have a linear combination of vectors in S, where both the coefficients and the vectors are unspecified, except that the vectors must belong to the subset S. By definition, a linear combination contains only finite number of vectors. The subset S that the vectors are taken from can still be infinite. Each individual linear combination will only involve finite number of vectors. Also, number n could be zero. In that case, result convention is declared of the linear combination is the zero vector in B. If we have a set of $k+1$ data points:

$$(x_0, y_0),...,(x_k, y_k) . \quad (2)$$

Where no two xj are the same, the interpolation polynomial in the Lagrange form is a linear combination:

$$l(x) = \sum_{j=0}^{k} Yjlj(x), \tag{3}$$

where lj represents the Lagrange basis polynomials:

$$\prod_{lj(x)=i=0,\ j\neq i}^{k} \frac{x - xi}{xj - xi}. \tag{4}$$

A Lagrange method of interpolation using the polynomial fits to the available values to interpolate between those values could be very effective in digital signal processing. If there are N data values, a polynomial of degree N-1 can be found that will pass through all the points. The Lagrange polynomials provide a convenient alternative to solving the simultaneous equations that result from requiring the polynomials to pass through the data values.

Example

Suppose we start with a polynomial, say, $f(x) = (x-1)^2$. And suppose we have two values, at $x_0 = 0$ and $x_1 = 1$, $f(0) = 1$ and $f(1) = 0$ [13].

The Lagrange interpolation formula takes the form $p_1(x) = f_0 l_0(x) + f_1 l_1(x) = l_0(x)$, since $f_0 = f(0) = 1$ and $f_1 = f(1) = 0$. Next, we have

$$L_0(x) = \frac{x - x_1}{x_0 - x_1} = \frac{x-1}{0-1} = 1 - x.$$

As seen from the figure, $p(x)$ does not interpolate $f(x)$ well. Now let us add the value of f at a third point, say $f(-1) = 4$. So we have $x_0 = 0,\ f_0 = 1,\ x_1 = 1,\ f_1 = 0,\ x_2 = -1,\ f_2 = 4$.

The Lagrange polynomials can be computed as:

$$l_0(x) = x - x_1 / x_0 - x_1 \cdot x - x_2 / x_0 - x_2 = -(x-1)(x+1),$$

$$l_2(x) = x - x_0 / x_2 - x_0 \cdot x - x_1 / x_2 - x_1 = 1/2x(x-1).$$

We do not need to compute $l_1(x)$ above since $f_1 = 0$. Now we have

$$p_2(x) = f_0 \cdot l_0(x) + f_1 \cdot l_1(x) + f_2 \cdot l_2(x) = -(x-1)(x+1) + 2x(x-1) =$$
$$= (x-1)(2x - x - 1) = (x-1)^2.$$

So, as one would expect, this approximation is exact. The goodness of an approximation depends on the number of approximating points and also on their location. One problem with the Lagrange interpolating polynomial is that we need n additions, $2n^2+2n$ subtractions, $2n^2+n-1$ multiplications, and $n+1$ divisions to evaluate $p(\varepsilon)$ at a given point ε. Even after all the denominators have been calculated once and for all and divided into the a_1 values, we still need n additions, n^2+n subtractions, and n^2+n multiplications. Another problem is that in practice, one may be uncertain as to how many interpolation points to use. So one may want to increase them over time and see whether the approximation gets better. In doing so, one would like to use the old approximation, in that way, a high-degree Lagrange interpolation may predict the function between points with greater error, although the accuracy at the data points will be perfect, for this reason the simple Lagrange polynomial interpolation is adapted in this paper to overcome any predictable errors. Figure 1 illustrated the Lagrange interpolation approximation.

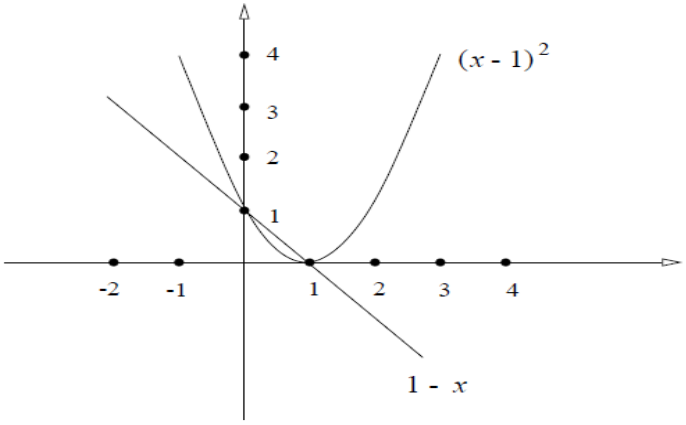


FIG. 1: Lagrange interpolation approximation

3. VOICED/ UNVOICED DETECTORS

Speech can be divided into numerous voiced and unvoiced regions. The classification of speech signal into voiced, unvoiced provides a preliminary acoustic segmentation for speech processing applications, such as speech synthesis, speech enhancement, and speech recognition [14].

The classification of the speech signal into voiced, unvoiced, and silence (V/U/S) provides a preliminary acoustic segmentation of speech, which is important for speech analysis. The nature of the classification is to determine whether a speech signal is present and, if so, whether the production of speech involves the vibration of the vocal

folds. The vibration of vocal folds produces periodic or quasi-periodic excitations to the vocal tract for voiced speech whereas pure transient and/or turbulent noises are periodic excitations to the vocal tract for unvoiced speech. When both quasi-periodic and noisy excitations are present simultaneously (mixed excitations), the speech is classified here as voiced because the vibration of vocal folds is part of the speech act. The mixed excitation, however, could also be treated as an independent category [15].

The most important methods of (V/UV) classification are:

1. Zero Crossing Rate method
2. Energy of Speech

3.1 Zero Crossing Rate Method

Zero crossing detection is the most common method for measuring the frequency or the period of a periodic signal. When measuring the frequency of a signal, usually the number of cycles of a reference signal is measured over one or more time periods of the signal being measured. Measuring multiple periods helps to reduce errors caused by phase noise by making the perturbations in zero crossings small relative to the total period of the measurement. The net result is an accurate measurement at the expense of slow measurement rates [16]. The zero crossing rates are calculated by using the formula given below [17]:

$$ZCR = \sum_{i=1}^{j-1} \frac{\operatorname{sgn}\left[y(i)\right] - \operatorname{sgn}\left[y(i-1)\right]}{2}, \tag{5}$$

where, $\operatorname{sgn}\left[y(i)\right]$ stands for the sign function, i.e.,

$$\operatorname{sgn}\left[y(i)\right] = \begin{cases} 1; & y(i) > 0, \\ 0; & y(i) = 0. \\ -1; & y(i) < 0. \end{cases} \tag{6}$$

3.2 Energy of Speech

Energy of a speech is another parameter for classifying the voiced/unvoiced parts. The voiced part of the speech has high energy because of its periodicity and the unvoiced part of speech has low energy [14].

The short time energy is said to be the sudden increase in energy signal. For calculating the short time energy the signal is split into *s* windows and the windowing function is calculated for each window. The short time energy is calculated using the equation given below [17]:

$$Short\ Time\ Energy = \sum_{r=-\infty}^{\infty} y(r)^2 \cdot h(s-r). \tag{7}$$

4. THE PROPOSED CLASSIFIER

In this section a new approach of V/UV blocks decision is presented, the approach used the properties for the Lagrange Polynomial Interpolation. The main advantage of this method is getting good results when it's applied on small size blocks of speech signal.

The Lagrange Polynomial Interpolation formula fits a curve that passes through the data points and matches the slopes at those points and the advantage of this curve fitting is exploited to decide blocks of speech file is voiced or not, the main steps of the proposed method is illustrated as follows:

1. Parse Speech Audio File.
2. Isolated header from each Speech Audio File.
3. Divided Speech file into different size of non- overlapped windows (blocks).
4. Each window is estimated by Lagrange Polynomial Interpolation formula. The result of estimation is used as parameter for (V/UV) blocks decision. The decision of whether a block is voiced or unvoiced is evaluated as follows:

$$f_n(x) = \sum_{i=0}^{n} L_n f(x_i), \tag{8}$$

where n in $f_n(X)$ stands for n^{th} order Polynomial approximates the function $y = f(x)$ given at $n+1$ data point as:

$$(x_0, y_0), (x_1, y_1), ..., (x_{n-1}, y_{n-1}), (x_n, y_n)$$

$$L_i(x) = \prod_{\substack{i=0 \\ j=1}}^{n} \frac{x - x_j}{x - x_j}. \tag{9}$$

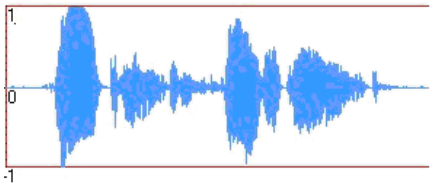
$L_i(x)$ is a weighting function that includes a products of n-1 terms with terms of $j = 1$.

If $f(x) <$ Threshold then the block is unvoiced, Else the block is voiced.

5. Each window is estimated by Short Average Energy Computation, The result of estimation of Short Average Energy Computation is used as parameter for (V/UV) blocks decision. The decision of whether a block is voiced or unvoiced is evaluated as follows:

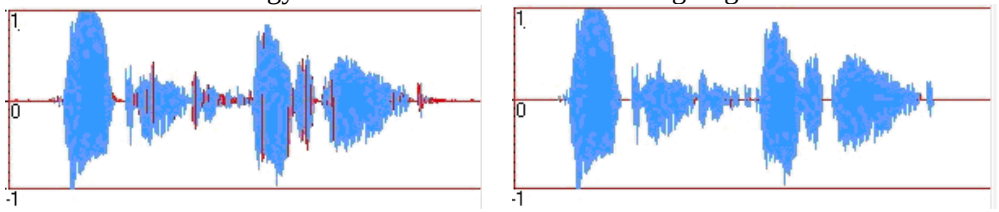
$$AE(i) = \sqrt{\frac{1}{w} \sum_{j-i}^{j-i+w} \left(Wav(j) - 128 \right)^2}. \tag{10}$$

If $AE(i) <$ Threshold then the block is unvoiced, Else the block is voiced.

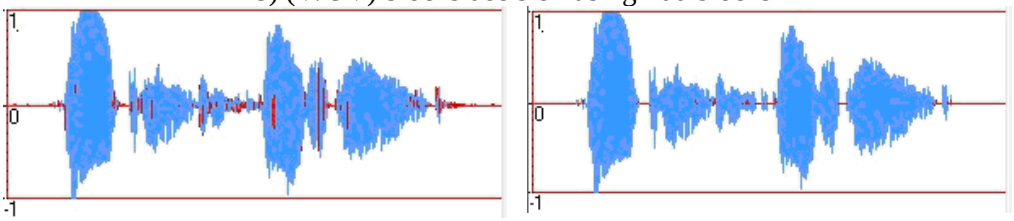


a) Sample_1 , Original Speech Signal

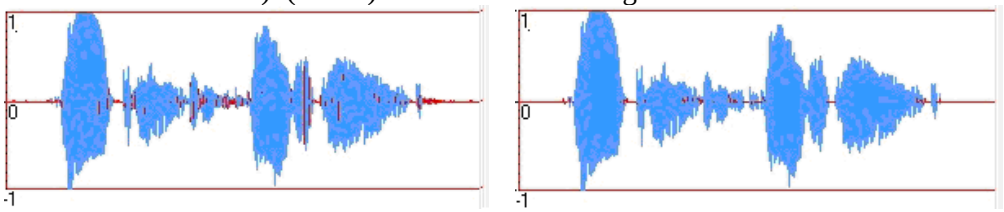
Short Time Energy Classifier Lagrange Classifier



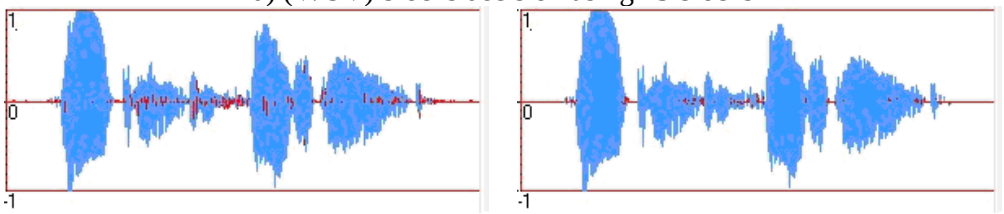
b) (V/UV) blocks decision using 100 blocks



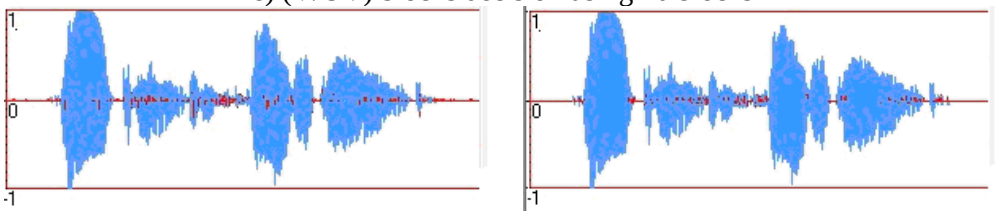
c): (V/UV) blocks decision using 50 blocks



d) (V/UV) blocks decision using 25 blocks



e) (V/UV) blocks decision using 10 blocks



f) (V/UV) blocks decision using 5 blocks

FIG. 2: (V/UV) Decision (Speech_1)

4.1 Test the Proposed Classifier Approach

In this section some speech samples are used to assess the performance of the proposed decision approach. For example, Sample -1 is used as test sample, 68.2KB size, WAVE type, PCM format, one channel (mono), and 8-bit sample size.

Two parameters are used as control parameters to implement the (V/UV) blocks decision, the first is block or window size which represents the size of speech data block that should be tested as voiced/unvoiced block, while the second is the threshold which represents the level value used to distinguish unvoiced blocks from voiced blocks.

An illustrative example of (V/UV) blocks decision is shown in Fig. 2, (V/UV) blocks decision are applied by two approaches; first decision is made using Short Time Energy Classifier , and the second decision is made using Lagrange Polynomial Interpolation. The voiced segments are blurred by Blue color and the unvoiced segments blurred by Red color. The original speech signal is shown in (a); (V/UV) blocks classifier shown in (b) where the block size = 100 byte, (V/UV) blocks classifier shown in (c) where the block size = 50 byte, (V/UV) blocks classifier shown in (d) where the block size = 25 byte, (V/UV) blocks classifier shown in (e) where the block size = 10 byte and (V/UV) blocks classifier shown in (f) where the block size = 5 byte.

One of the speech signal used in this paper is given with Fig. 2, proposed voiced/unvoiced classification approach uses short energy computation and the Lagrange technique of the speech signal. The results of (V/UV) decision using approach are presented in Table 1 and Table 2.

TABLE 1: V/UV blocks decision using Short Average Energy Computation

Window size	No. of Blocks	No. of Voiced Blocks	No. of Unvoiced Blocks
100	697	236	461
50	1395	489	906
25	2792	1012	1780
10	6982	2677	4305
5	13965	5600	8365

TABLE 2: V/UV blocks decision using Lagrange Polynomial Interpolation

Window size(byte)	No. of Blocks	No. of Voiced Blocks	No. of Unvoiced Blocks
100	697	250	447
50	1395	503	892
25	2792	988	1804
10	6982	2466	4516
5	13965	4878	9087

Tables 1 and 2 are illustrated clearly that the proposed Lagrange classifier approach is produced convergence decisions results when compared with Short Time Energy classifier, and being more accurate based decision-making especially when it's applied on small size of blocks samples.

5. DISCUSSIONS

The proposed scheme provide another method to classify (V/UV) blocks, by applying new application for Lagrange Polynomial Interpolation formula to detect voiced segment from unvoiced, there are some issues to be discussed here:

1. The Lagrange Polynomial Interpolation formula produced a good result when its apply on short a non-overlapping frame of samples.
2. (V/UV) blocks detectors can be used as a tool to quantize voiced or unvoiced segments.
3. The Lagrange Polynomial Interpolation formula is computationally simple, so it's faster than other methods.

6. CONCLUSION

We have presented a new approach for separating the voiced /unvoiced part of speech in a simple and efficient way using proposed Lagrange classifier. The proposed classifier results shows that Lagrange polynomial interpolation formula is best (V/UV) blocks classifier for short a non-overlapping frame of samples.

The ideas of applying Lagrange Polynomial Interpolation formula for (V/UV) blocks detector can be extended in future work along several interesting directions, among these are:

1. Removal noise.
2. Use Hermit Interpolation Polynomial.
3. Use Interpolation Polynomial formula for gender clustering and classification of speech signal.
4. Improve our result classifier for voiced/unvoiced discrimination in noise.

REFERENCES

1. Vladan V. Vučković, (2008), The Reconstruction of the Compressed Digital Signal using the Lagrange Polynomial Interpolation, *Electronic, Engineering*, Aleksandra Medvedeva 14, 18000 Niš, Serbia. email:vld@elfak.ni.ac.yu.
2. Atal, B. and Rabiner, L., (1976), A Pattern Recognition Approach to Voiced-Unvoiced-Silence Classification with Applications to Speech Recognition, *IEEE Trans. on ASSP*, ASSP-24:201-212.
3. Ahmadi, S. and Spanias, A.S., (1999), Cestrum-Based Pitch Detection using a New Statistical V/UV Classification Algorithm, *IEEE Trans. Speech Audio Processing*, **7**(3):333-338.

4. Qi, Y. and Hunt, B.R. (1993), Voiced-Unvoiced-Silence Classifications of Speech using Hybrid Features and a Network Classifier, *IEEE Trans. Speech Audio Processing*, **1**(2):250-255.
5. Siegel, L., (1979), A Procedure for using Pattern Classification Techniques to obtain a Voiced/Unvoiced Classifier, *IEEE Trans. on ASSP*, ASSP-27:83- 88.
6. Burrows, T.L., (1996), Speech Processing with Linear and Neural Network Models, *Ph.D. thesis*, Cambridge University Engineering Department, U.K.
7. Childers, D.G., Hahn, M. and Larar, J.N., (1989), Silent and Voiced/Unvoiced/Mixed Excitation (Four-Way) Classification of Speech, *IEEE Trans. on ASSP*, **37**(11):1771-1774.
8. Wolfe, P.J., Godsill, S.J., and Dörfler, M., (2001), Multi-Gabor Dictionaries for Audio Time-Frequency Analysis, *Proc. IEEE Wkshp. on Appl. of Sig. Proc. to Audio and Acoust*., New Paltz, NY, pp. 43-46.
9. Guillaume, P., Schoukens, J., Pintelon, R., and Kollar, I., (1991), Crest-factor minimization using nonlinear Chebyshev approximation methods, *IEEE transactions on instrumentation and measurement*, **40**(6):982-989.
10. Jeffreys, H. and Jeffreys, B.S., (1988), *The Lagrange's Interpolation Formula. §9.011 in Methods of Mathematical Physics*, Cambridge, England: Cambridge University Press, - 260 p.
11. Séroul, R., (2000), *The Lagrange Interpolation. §10.9 in Programming for Mathematicians*. Berlin: Springer-Verlag, pp. 269-273
12. . Atal, B. and Rabiner, L., (1976), A Pattern Recognition Approach to Voiced-Unvoiced-Silence Classification with Applications to Speech Recognition, *IEEE Trans. on ASSP*, ASSP-24:201-212.
13. Polynomial Interpolation. pdf, Com S 477/577, 2007 URL: http://www.cs.iastate.edu/~cs577/handouts/interpolate.pdf.
14. Bachu, R.G., Kopparthi, S., Adapa, B., Barkana, B.D., (2008), *Separation of Voiced and Unvoiced using Zero crossing rate and Energy of the Speech Signal*, Electrical Engineering Department, School of Engineering, University of Bridgeport URL : http://www.asee.org/documents/zones/zone1/2008/student/ASEE12008_0044_paper.pdf
15. Bachu, R.G., Kopparthi, S., Adapa, B., Barkana, B.D., (2008) Voiced/Unvoiced Decision for Speech Signals Based on Zero-Crossing Rate and Energy, *IEEE International Joint Conferences on Computer, Information, and Systems Sciences, and Engineering (CISSE'08)*.
16. Yingyong, Qi and Bobby R. Hunt, (1993), Voiced-Unvoiced-Silence Classifications of Speech Using Hybrid Features and a Network Classifier, *IEEE Transactions on speech and audio processing*, **1**(2) URL : http://www.math.uci.edu/~yqi/ieee00222883.pdf.
17. Wall, R.W., *Simple Methods for Detecting Zero Crossing*, Moscow, ID 83844-1023 (e-mail: rwall@uidaho.edu).