Audio Classification based on Polynomial Interpolation

Dr. Nidaa F. Hassan
Computer Science Department, University of Technology/Baghdad
E-mail: nidaaalalousi_5@yahoo.com

Sarah Qusay Selah Alden
Computer Science Department, University of Technology/Baghdad
E-mail: sarahqusay19@yahoo.com

Abstract:
 - A new algorithm to classify gender based on autocorrelation is used, firstly the signal is interpolated (resample signal) to fill-in the gaps between the signal, and the signal is convolved to be Cross-Correlated (long signal for a shorter), so the gender is classified depending on pitch Cross-Correlation. Finally, the gender is classified to be male if the pitch value range between 85 Hz and 155 Hz, and the gender is female if the pitch value range between 165 Hz and 255 Hz.
Key words: - digital audio, Pitch, Autocorrelation, Cross- Correlation and polynomial interpolation.

المستخلص :
تم استخدام خوارزمية جديدة للتصنيف بين صوت الجنسين على أساس الارتباط الذاتي. في البداية يعمل اقحام للاشارة (اعادة تشكيل الاشارة) لملء الفراغات بين الاشارات، ثم يتم عمل التفاف للاشارة (convoluted signal) إلى اشارة (Cross-Correlated) (إشارة طويلة لأقصر). واخيرا يتم التصنيف بين الجنسين اعتمادا على حدة الصوت عبر مرتبطة (Cross-correlation). ان الصوت يحدد على انه صوت ذكر إذا كان النطاق في قيمة حدة الصوت بين 85 هرتزو 155 هرتز، اما تحديد صوت الأنثى يتم إذا كان النطاق في قيمة حدة الصوت بين 165 هرتزو 255 هرتز.

1. Introduction

Multi-media data-bases or file systems are capable of easily having thousands of audio records. Libraries like those are usually in efficiently indexed or named. Finding a specific file or class of audio (like the ringing of the doorbell, a laugh, or a dog barking) might become quite a difficult job. Therefore, the generalized audio classifying approach takes a significant part in multi-media content retrieving [1].

Audio signal classification (ASC) is made up of the extraction properties from a sound, and then these properties are used in the identification of which of a group of classes the audio file is mostly going to belong to. The extracting of properties and grouping methods that are implemented can be very different according to the application's domain of classification [2].

2. Applications of Audio Signal Classification (A.S.C.)

The applications of ASC are quite reaching and sufficient; several of those applications are in audio classification, data-base and auto-transcription. The applications have been classified into the following three areas:

i. Computing Tools

ASC can be implemented as front-end for some of already known computer applications to audio. Audio identification is more understand able application of the Audio Signal Classification, where signals are split into audios and after wards combined into words. Audio Signal Classification might be implemented for improvement on today's audio identification methods in variety ways. Audio includes a bigger amount of data than only words, like for example; sentimental content, idioms and tightening. Prosody is the variation that occurs in the audios pitch and loudness, for instance, a heighting pitch at the end of a question, so developing of Audio signal classification system could be used to exploit the and take benefit of prosody to create text notations automatically like punctuation and parentheses.

ii. Consumer Electronics

A big number of applications of ASC can be improved as market products. Government and industry that are monetizing the programs for ASC research. Embedded device in ASC applications includes: microchips such as cell phone, televisions and automobiles. An included ASC device in a cell phone could be used to notify the employer of the type of reacting signal when making a call. The included device could state the difference between a fax, and answering machine, a modem and human letter. Depending on the application, different replies would generate different activities from the included device, which could instruct the cell phone to hang up, redial, connect, wait, or navigate an automated cell phone menu.

iii. Automatic parallelization

Parallelization is applying, in parallel, a bank of filters to audio signal. The signal is changed liable on the comparative expansion or reduction of the signal in each filter range, or channel. The projected result of this procedure is that, the signal is equalized to equality with higher perceptual than the original signal [2].

3. Gender Classifier

Gender Classifier from speech, is a portion of auto-audio identification system for enhancing speaker's adaptability and a portion of the system of auto-audio identification. The necessity for gender classification from speech emerges in a variety of cases like sorting phone calls according to gender; moreover, it's a part of the in innovative audio passwords [3].

4. Polynomial Interpolation

Polynomial Interpolation methods are used to estimate a massing function value via obtaining weighted mean of known function value of adjacent points. Up-sampling is interpolation procedure, used for the digital signal processing and sample rate converting. When this procedure is applied on a series of samples of a continuous function or signal, it results in an approximation of the series which would've been extracted by sampling that signal at a bigger rate (or at higher density, like the case of a photograph). For instance, in case that a compact audio is up sampled using a factor of 5/4, the obtained sample-rate rises from 44,100 Hz to 55,125 Hz [4].

5. Linear interpolation:

Linear interpolation utilizes a line segment which passes through 2 distinct points:$(x_{0,}y_0)$and $(x_{1,}y_1)$and it's the same as the approximation of a function f for which$f(\mathrm{x}_0) = y_0$, and $f(\mathrm{x}_1) = y_1$ by means of 1^{st}degree polynomial interpolating.
The slop between$(x_{0,}y_0)$, and $(x_{1,}y_1)$ is

$\text{slop} = \text{m} = \frac{y_1 - y_0}{x_1 - x_0}$ ……… (1)

Where

$y = m(x - x_0) + y_0$

$y = p(x) = m(x - x_0) + y_0 = \frac{y_1 - y_0}{x_1 - x_0}(x - x_0) + y_0$

$p_1(x) = y_0 \frac{x - x_1}{x_0 - x_1} + y_1 \frac{x - x_0}{x_1 - x_0}$ …………………… (2)

Every one of the terms of the right side of (1) involves a linear factor; therefore the summation is a polynomial which is of degree ≤1.

$L_{1,0}(x) = \frac{x - x_1}{x_0 - x_1}$, and $L_{1,1}(x) = \frac{x - x0}{x1 - x0}$ .... (3)

When $x = x_0$, $L_{1,0}(x_0) = 1$ and $L_{1,1}(x_0) = 0$. When, $x = x_1$, $L_{1,0}(x_1) = 0$ and $L_{1,1}(x_1) = 1$.

In terms $L_{1,0}(x)$ and $L_{1,1}(x)$ in Ep. (2) called **Lagrange** coefficient of polynomial hazed on the nodes x_0 and x_1.

$p_1(x_0) = y_0 = f(x_0)$, and $p_1(x_1) = y_1 = f(x_1)$ using this notation in Eq (4), can be written in summation

$p_1(x) = y_0 L_{1,0}(x_0) + y_1 L_{1,1}(x_1)$ …… (4)

$$p_1(x) = \sum_{k=0}^{1} \text{ykL1k(x)}$$

Suppose that the ordinates

$y_k = f(x_k)$

If $p_1(x)$ is uses to approximate $f(x)$ over intervals $[x_0\ x_1]$.

6. Autocorrelation

The study of signal processing systems was dominated via the term of convolution which is held usually between a signal and a filter, therefore it might be thought of as a system that has only one input and stored co-efficient. Correlation is typically between two signals, therefore it might be considered as a system has two inputs without stored coefficients. Autocorrelation is the Cross-Correlation of a signal with itself. In other words, it’s the similarity between observations as a function of the time lag between them. It’s a mathematical tool used to find redundant patterns, like the existence of periodic signal distorted by noise, or to identify the missing basic frequency in a signal implied by its harmonic frequency. It’s usually utilized in signal processing in order to analyze functions or groups of values [5].

7. Cross-Correlation

The Cross-Correlation Function (CCF) between two signals performs description of the degree by which the two signals resemble (correlate) one to another. The auto correlation function (ACF) is identical to the CCF, except for the fact that ACF makes a correlation of a signal with itself. The idea of obtaining a quantitative measure of similarity between two signals is fundamentally the same as statistical correlation of two random variables, except for the fact that it is not quantified explicitly by statistical measures like the covariance and standard deviations. Correlation, when referred to in a signal processing sense, is considered as direct comparison between the signals based on their actual values at different time instant, although it is definitely possible to interpret this operation in a statistical sense. Fundamentally, the comparison between any two signals or waveforms may be quantitatively based upon the amount of the component of one waveform contained in the other waveform [6].

8. Pitch Detection

The high or low audio being played by a musical tool or vocal is described by a Pitch. The pitch of audio relays on the frequency of the digital audio file; the frequency is measured in Hertz (Hz), with one vibration per second being equivalent to one hertz (1 Hz). A high pitched of audio is produced by high frequency and low pitched note is produced by low frequency.

For detecting pitches, the approach of Auto-Correlation appeared to be efficiently precise; it's utilized the Pitch for classifying genders (male or female). The gender is classified as male in the case where the value of the pitch is between 85 Hz and 155 Hz and it is classified as female when the value of the pitch ranges between 165 Hz and 255 Hz [7, 8].

9. The Proposed Algorithm

In this proposed algorithm the audio signal is passed through number of process to classify gender sound (male from female).The gender classification process is applied on CMU_ARCTIC (Carnegie Mellon University) audio dataset with format Wave [9].
A figure 1 show the flow chart of gender classification process, the audio gender classification consists of the following steps:

1. Input the original audio signal
2. Interpolated the audio signal.
3. Cross-Correlated the interpolated audio signal (convolved).
4. Autocorrelation audio signal process.
5. Gender classification process based on Pitch value.

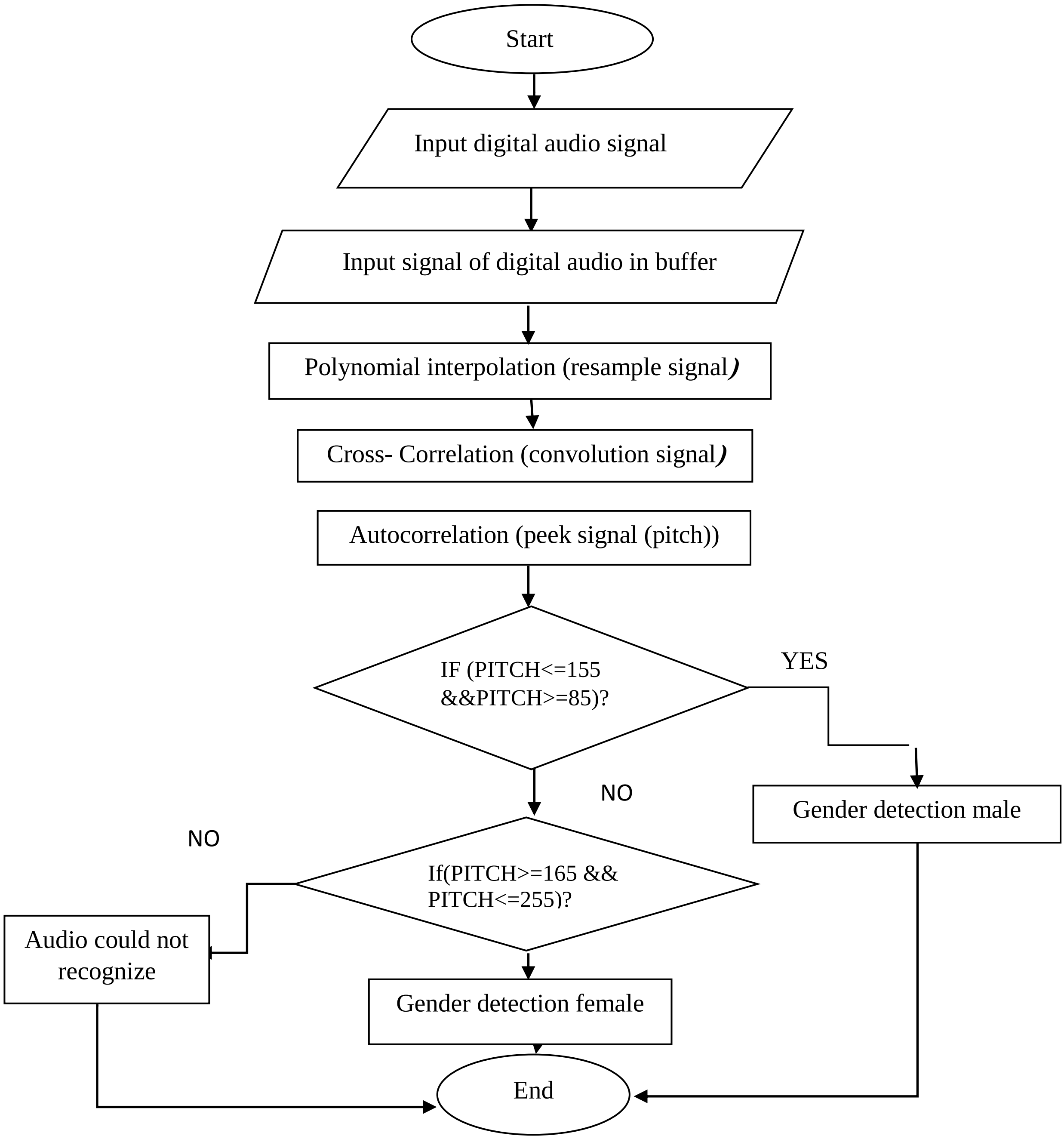


Figure 1: Flow chart of the proposed Gender audio classification

1. Input the original audio signal

The first step in gender audio classification is selecting sample of audio signal; this signal is considered to be input signal as shown in the following figure.

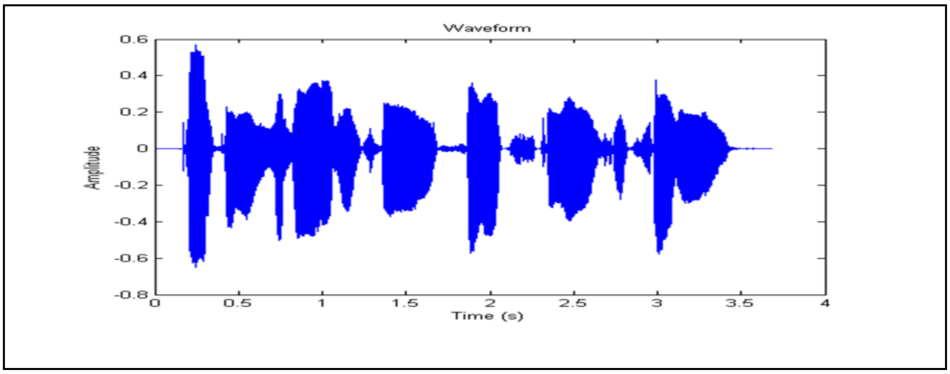


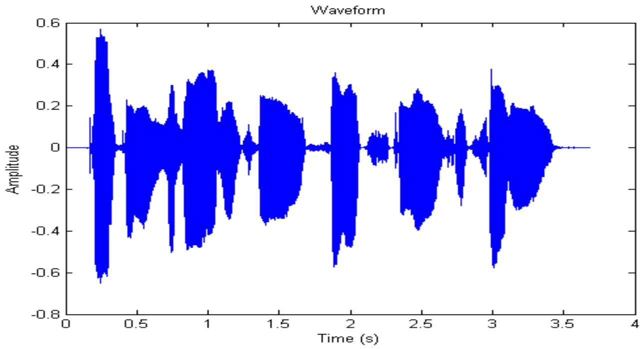
Figure 1: Sample of audio signal

2. Polynomial Interpolation

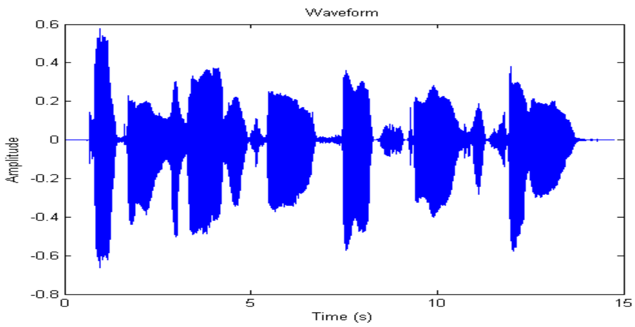
Lagrange polynomial interpolation is applied to increase sampling rate and re-full gaps between audio signals, this will help to predict the exact sound, so using Lagrange polynomial function made gaps between signal and sound processing filled. For example, when a word in voice consists of two sections such as "**Salsabel**", after apply Lagrange function on this word, the gaps between these two sections is filled and become a full section. Algorithm 1 illustrates the filling gaps between two sections of word by using Lagrange function; figure 2 shows the audio signal before and after filling gaps between words.

Algorithm 1: filling gaps between two sections of word by using Lagrange function.
Input: Digital audio signal (wave).
Output: Filling gaps between two sections of word.

```
Step 1:Input digital audio signal
Step 2 : Initialize variable Pz=0        // Initial value of polynomial
Step 3:Set n to the number of samples contains pairs of points (x, y).
Step 4:Set L to be the all ones vector of length n.
Step5:Input signal of digital audio in buffer.
                  // Polynomial interpolation (resample signal)
Step 6: For i = 1 to n do…
           For j =1to n do Step 5.
             If i ≠ j then Li = (z –xj )/(xi –xj ) ∗ Li
              Pz = Li ∗Yi + P z
             output Pz     //Fill gaps between two signals
           End For j
          End For i
```



(a) Original audio signal.



(b) Audio signal (polynomial interpolation).

Figure 2: Audio signal before and after filling gaps between words

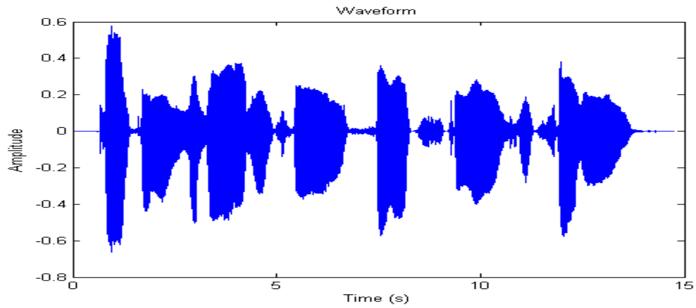
3. Cross-Correlation Process (Convolution)

Cross-Correlation is used to measure the similarity of two waveforms as a function of a time lag applied to one of them. Cross-Correlation coefficient function between the original and the center clipped signal is computed over the range of human pitch frequency (50 – 500 Hz). So in order to detect weather a digital audio is a male or female, long signal for a shorter is applied as a feature by using Cross-correlation. The Cross-Correlation at zero delay is computed for pitch frequency, Cross-correlation function is then searched for its maximum frequency value and the pitch delay is computed as the time delay between consecutive peaks.

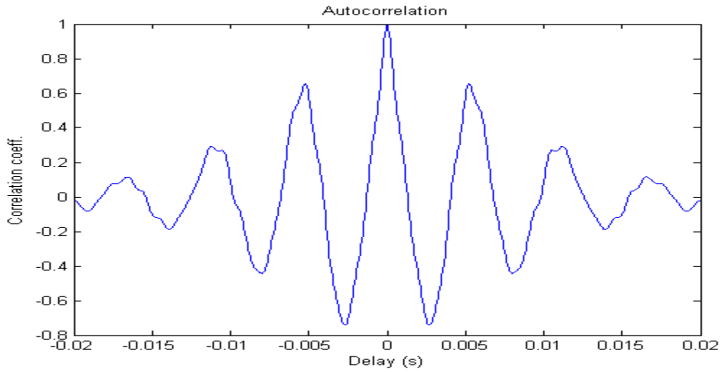
4. Autocorrelation Process

Gender classification is mainly based on the pitch Cross-Correlation. Autocorrelation is used to extract the peak of audio signal. The gender is classified to be male if the pitch value range between 85 Hz and 155 Hz and the female if the pitch value range between 165 Hz and 255 Hz. Algorithm 2 the classify audio signal by using Autocorrelation.

```
Algorithm 2: Classify audio signal by using Autocorrelation.
Input:   Digital audio signal after Lagrange interpolation signal (wave).
Output: Classify audio signal in two classes (male or female).
-----------------------------------------------------------------
Step 1: Input digital audio signal after interpolation signal.
Step2:  Input interpolated digital signal in buffer a( )
Step 3 : Input interpolated digital signal in buffer b( )
Step 4:    For I = 1 to n // Cross-correlation as convolution two signals.
               For  J = 1 to n
                       a= array 1, b= array 2.
                       double[] Xcorr(double[] a, double[] b)
                       Int Len = a .length.
                       if(b. length > a. length).
                               Len = b. length.
                               Return Xcorr (a, b, len-1).
                       End if.
                   Next I.
Next J.
Step5: Auto-correlation specific signals by using (pitch).
// The autocorrelation value reflects the similarity between the frame s[n]
and the time-shifted version s[n-l]
                  n= [m–N+ 1 ,m]
                  l is a positive integer representing a time lag.
Step6: Pitch (m, N)// calculate the pitch of audio signal.
              Peak = 0.
                     For t = 20 to 150
                             autoc= 0.
                         For n= m – N+ 1 to m
                           autoc= autoc+ s[n]s[n– l].
                           if autoc> peak
                             Peak = autoc.
                              Lag= t.
                            End if.
                          Next n.
                      Next t.
Step5: Fx= the value of pitch specify the gender   //classify signal into
                                                          (male or female).
            If  Fx in [50..250] then audio signal is male
                   If  Fx in [120..500] then audio signal is female
                 Else
                          "Audio signal could not recognize"
                End If
```



(a) Audio signal (polynomial interpolation).



(b) Audio signal (Autocorrelation & Cross-correlation)

Figure 3: Audio signal before and after (Autocorrelation & Cross-correlation)

5. Gender classification process based on Pitch value.

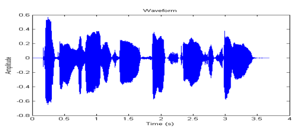
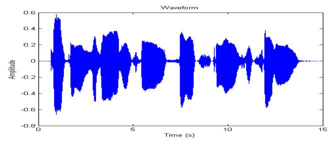
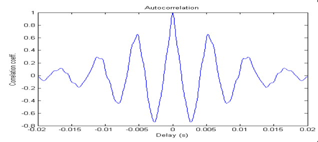
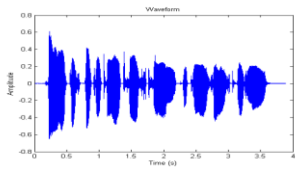
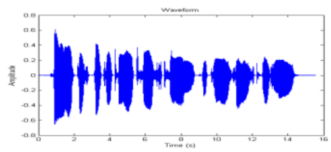
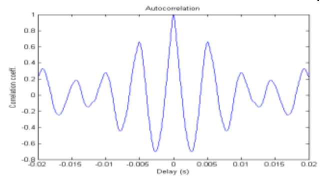
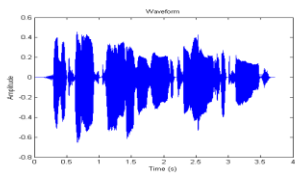
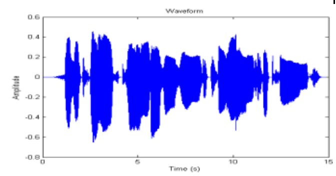
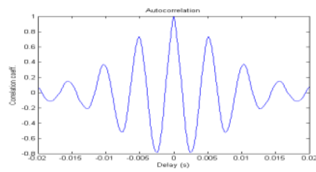
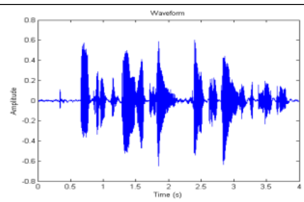
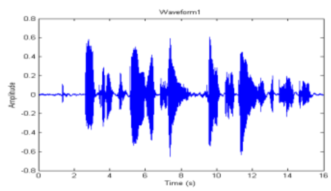
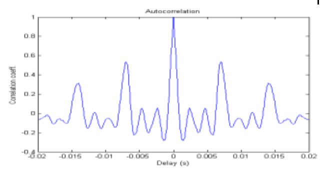
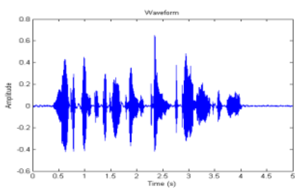
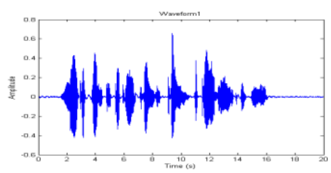
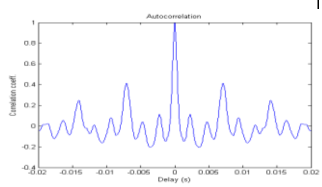
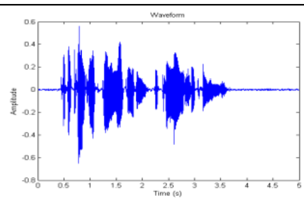
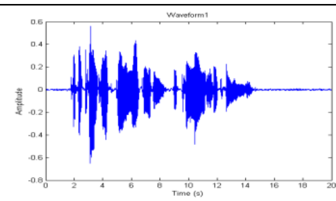
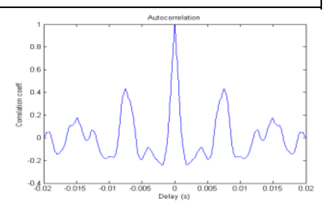
The following table shows the gender classification of six speech samples, theses sample are with wave format with (one channel, 16000 sapling rate and 16 valid bits). Table 1 shows each of six samples with its number of frames; Mix-mum speech (Is the resulting frequency ratio).

Table 2 shows the difference between (Original, Polynomial interpolation and Auto-correlation) signals.

Table 1: Explains the information to samples of waves.

Sample of Wave	Number of frame	Mix-mum speech	Gender classification
Sample_1	58960	190.4	Female
Sample_2	62160	197.5	Female
Sample_3	59440	195.1	Female
Sample_4	6400	141.5	Male
Sample_5	80000	141.5	Male
Sample_6	80000	133.3	Male

Table 2: Explains the difference between (Original, Polynomial interpolation and Auto-correlation)

Sample of Wave	Original audio	Polynomial interpolation	Auto-correlation
Sample_ 0 (female)			
Sample_ 1 (female)			
Sample_ 2 (female)			
Sample_ 3 (male)			
Sample_ 4 (male)			
Sample_ 5 (male)			

10. Conclusion

1- In this paper, a new algorithm is applied on audio dataset with format Wave, the main idea of the classification is based on polynomial interpolation Autocorrelation and Cross-Correlation process and Pitch feature.
2- Lagrange linear is used fill gaps between signals in the words when contain two sections and this helps in increasing of accuracy rate.
3- Autocorrelation and Cross-correlation process are used to estimate pitch features. Pitch feature could be considered as the main parameter which gender classification is depending on to reach optimality classification.

11. Future Work

1- The proposed algorithm can be used to recognize audio to determine to whom its belong
2- The proposed algorithm can be further modified for determining the classification for other audios format such as mp3 and in another environment such as cell phone.
3- The audio classifier could be used for other types of audio signal such as animal voices by analysis samples of audio data set and supplemented another features descriptors.
4- Gender classification from audio could be applied on digital video after this media is prepared for further processing.

References

[1] Stan Z. Li, and Guo G., "Content-Based Audio Classification and Retrieval Using SVM Learning", IEEE TRANSACTIONS ON NEURAL NETWORKS, Microsoft Research China, 5/F Beijing Sigma Center, Beijing 100080, China, (2003).

[2] David G., "Audio Signal Classification: History and Current Techniques", Technical Report TR-CS 2003-07 November, Department of Computer Science University of Regina, Saskatchewan, CANADA S4S 0A2 ,ISSN 0828-3494 (2003).

[3] Bhagya L. and Abhishek M., Beda P., "Gender Classification by Speech Analysis", Students of Electronics & Tele-communication Branch, Silicon Institute of Technology, (2014).

[4] Tamanna I., April, "Interpolation of Linear Prediction Coefficients for Speech Coding" Department of Electrical Engineering McGill University Montreal, Canada, (2000).

[5] Hema K. and Limaye S., "Autocorrelation of a Sound signal", IOSR Journal of Electrical and Electronics Engineering (IOSR-JEEE) e-ISSN: 2278-1676, p-ISSN: 2320-3331 PP 50-53, (2014).

[6] Ir. A. Gisolf, and Ir. D. de Vries, "Implementation of A Source tracking System Using Cross Correlation of Speech Signals", Department of Applied Physics Faculty of Applied Sciences Delft University of Technology, Delft, September (2002).

[7]Kapil K.,"GENERATIONOFCONTROLSIGNALSUSINGPITCHANDONSET DETECTIONFORANUNPROCESSEDGUITARSIGNAL", Center for Computer Research in Music and Acoustics Stanford University, (2006).

[8] Preetir R. and Sumantra R.,"AUDIO SIGNAL CLASSIFICATION", **IEEE Xplore**, M. Tech. Credit Seminar Report, Electronic Systems Group, EE. Dept, IIT Bombay, Submitted November(2004).

[9]http://www.speech.cs.cmu.edu/cmu_arctic/cmu_us_bdl_arctic/wav/?C=D;O=A.