PITCH ESTIMATION

SMITA S. HANDE[1], DR. MILIND S. SHAH[2]

[1,2]EXTC Department, Fr. C R I T, Sector 9A Vashi, Navi Mumbai, Maharashtra State, India

Abstract- This paper describe three pitch estimation techniques using autocorrelation method, Cepstrum method and Praat's pitch estimation method using autocorrelation involving the preprocessing and the extraction of pitch contour. In speech processing the important but difficult is voiced/unvoiced detection. For voiced/unvoiced decision a method is introduced using energy and zerocrossing in autocorrelation method. Cepstrum uses part of end point detection technique for voiced/unvoiced detection. It also presents the implementation and the basic experiments and discussions. The performance of all pitch estimation methods are compared in terms of mean square error with the freely available Praat software using autocorrelation methods. A comparative evaluation of the pitch estimation results on test signals, synthesized vowels and recorded vowels for male and female voices shows the superiority of the Praat's method over other methods.

Index Terms- *Pitch, Pitch Detection Algorithm, Autocorrelation function, Cepstrum, Praat algorithm voiced/unvoiced.*

I. INTRODUCTION

The fundamental frequency of a sound, whose percept is called pitch, has great importance in many areas. Pitch detection is one of the oldest, yet unsolved topic among the researchers of speech and music [1]. Accurate pitch detection is essential to areas such as speech coding, speech synthesis, to more recent topic of speaker emotion recognition [10], gender recognition [13] as pitch range for female is more than male and kids have high value pitch [15]. The automatic tracking of pitch has multiple applications in the field of speech processing and speech technology [2]. Pitch contour is useful in assisting hearing impaired people. Consequently, a wide range of perceptual models and algorithms using a variety of techniques and a varying degree of accuracy to extract pitch exist [3]. However, the pitch detection algorithms (PDAs) face a real challenge in the presence of noise [4].

There are three types of PDAs in the literature: time-domain, frequency domain, and time-frequency domain. Time domain method includes the short-time average magnitude difference function (AMDF) [1]-[3], short-term autocorrelation function (ACF) [1]-[4], etc; frequency domain method includes harmonics enhancement based on instantaneous frequency, and cepstrum analysis [17]; while pitch detection based on Praat's algorithm [5] [18] falls in time domain method. Among all the methods, ACF-based algorithms are simpler to implement and robust against noise. In this paper, we have focused on ACF-based pitch detection algorithms such as modified autocorrelation method and Praat's autocorrelation method for its accuracy. Because of the periodic nature of voiced speech, its ACF is also periodic with period equal to the pitch value. ACF shows peaks at pitch and its harmonics locations. Natural speech is not absolutely periodic, rather it is quasi periodic. Hence ACF produces the highest peak at pitch period, and gradually decreasing peaks at its harmonics. Thereby, the highest peak other than zero location in ACF corresponds to the pitch period. However conditions like presence of noise, quasi periodic nature of speech signal, peaks due to detailed formant structure of vocal tract affect the location of the highest peak in ACF. A comparative performance study of seven pitch detection algorithms was conducted in [6]. Four of the pitch detectors (AUTOC, PPROC, LPC and AMDF) performed extremely well on the given input condition. The SIFT and DARD methods had somewhat poorer performance, while the CEP method had the worst performance. A weighted ACF method using AMDF has been proposed in [11], but this method suffers from double pitch error at low signal to noise ratio (SNR). An improved cepstrum-based voicing detection and pitch determination algorithm is presented in [17] where voicing decisions are made using a multifeature voiced/unvoiced classification algorithm based on statistical analysis of cepstral peak, zero-crossing rate, and energy. A low cost impedance glottograph and glottal pitch analyzer is developed in [12] which can be used for studying voice disorders as well as speech analysis and synthesis applications. A method where division of filtered signal by the normalized autocorrelation of the windowing function is done, and used by the Praat software for analysis and synthesis of speech signal is introduced in [18], which is both robust and better suited for the analysis of speech signals than previous methods used. A voiced/unvoiced decision based on end point detection is introduced in [16]. In the paper, the principles of the three pitch detection algorithms, preprocessing and the extraction of pitch pattern techniques are introduced. The implementation of them is described. Then the experiments and discussions are presented. Finally it's the conclusions of the paper.

II. BACKGROUND INFORMATION:

A. Autocorrelation Method:

A commonly used method to estimate pitch is based on detecting the highest value of the autocorrelation function in the region of interest. Given a discrete time signal x(n), defined for all n , a short-time auto-correlation function is generally defined in (1):

$$R[n] = \sum_{m=0}^{N-n-1} s(m)]s(m+n) \qquad (1)$$

where, N is total number of samples in a window, and n is the lag index. The choice of window length N for calculating R[n] has conflicting requirements:

- N should be as small as possible to show time variation;
- N should be large enough to cover at least 2 periods so that periodicity can be captured by R[n].

R[n] has same periodicity as the s[m], maximum value at n = 0 and R[0] is equal to energy of deterministic signal and if s[m] is periodic with period of P samples, R[n] has maximum at n = 0,± P, ±2P,….

B. Cepstrum Method:

The cepstrum, defined as the real part of the Inverse Fourier transform of the log-power spectrum, has a strong peak corresponding to the pitch period of the voiced speech segment being analyzed [4].

$$c(n) = IDFT\ (\log |S(\omega)|) = IDFT\ (\log |E(\omega)| + \log |H(\omega)|) \qquad (3)$$

The cepstral peaks corresponding to the voiced segments are clearly resolved and quite sharp and cepstral peaks decrease in amplitude with increasing quefrequency. Hence, the peak picking scheme is to determine the cepstral peak in the interval [2.5–15 ms], corresponding to pitch frequencies between 60–400 Hz, which exceeds some specified threshold.

C. Praat algorithm:

The analysis of periodic curves is best suited when working with continuous, stable curves. However, the speech signals are unstable, which makes analysis more difficult. Praat overcomes this by assuming that speech is sufficiently stable when looking at small enough fragments of it. Each window is filtered to make sure there are no intensity peaks on the edges, which facilitates analysis. According to the Praat documentation, a Hanning window is more responsive when working with 3 periods per analysis window, while a Gaussian window is better when working with a larger analysis window (the Gaussian window is twice as large as the Hanning window). Praat indeed uses both these windows as default depending on the task and the degree of precision that is required. When we work with a complex sound wave the autocorrelation curve shows a false peak before the time lag i.e. actual fundamental frequency. In order to correct for this, the division of filtered signal by normalized autocorrelation of windowing function is used in Praat's algorithm. This estimate gets increasingly unreliable after roughly half the length of the analysis window. And by finding the maximum at a time lag>0 in this estimated curve, pitch of the original signal converting from samples to Hz is calculated.

f0 = (1 / time.lag.of.max) / sample.rate (3)

D. Preprocessing and Center Clipping:

The speech signal, sampled at 11025Hz, is analyzed at 5 sec intervals using a 40 ms Hamming window. An optional bandpass noise-suppression filter (i.e., a ninth-order Butterworth filter with lower cutoff frequencyof 200 Hz and upper cutoff frequency of 3400 Hz) is applied to deemphasize the out-of-band noise when the input speech is contaminated with additive noise as well as providing an appropriate high-frequency spectral roll-off. After this preprocessing stage center clipping is done for autocorrelation method. The ACF may contain too much information, most of which is not related to the fundamental frequency. For pitch detection, speech signal is usually pre-processed to make the periodicity more prominent and to suppress other distracting features. Such techniques are often called spectrum flattening. Center clipping is the most popular spectrum flattening technique, and infinite peak clipping can be expressed as Eq. (4). A choice of clipping level should fulfill the following criterion: - should be high enough to eliminate all distracting peaks, but - cannot be too high so as not to lose desirable peaks.

$$y(n) = \mathrm{sgn}[\, s(m)] = \begin{cases} +1 & s(m) > +Cl \\ 0 & -Cl \le s(m) \le +Cl \\ -1 & s(m) < -Cl \end{cases} \qquad (4)$$

Usually, the clipping level is chosen to be 60% to 80% of the maximum amplitude and is adaptively adjusted according to the signal level. For center clipping, the minimum of the maximum amplitudes of the first one-third samples and the last one-third samples in a frame is determined. Then the clipping level is set to 68% of that minimum value.

The erroneous voiced/unvoiced decisions and inaccurate voiced pitch hypotheses can lead to noisy and undependable feature measurements. Then a smoothing stage is necessary in improving the performance of the system. The most common smoothing techniques includes: median filter, linear smoothing and dynamic programming technique. According to the reliability of pitch tracking algorithm, generally the median-filter is used. In the method of medianfilter, it uses a moving window with the length L. The value at point n is determined by the data from point n-L to point n+L. Then the median value in these 2L+1 points is chooses as the value the point.

III. IMPLEMENTATION:

A. Modified Autocorrelation Method:

According to the discussion above, the modified autocorrelation pitch detector based on the center-clipping method and infinite-clipping is used in our implementation.

Step by step implementation:

1. Sectioning of signal:

Sectioning of signal is done by deciding the window length which depends on minimum and maximum pitch values (50Hz and 500Hz resp.). Window length is the length of the analysis frame chosen to be the three times the minimum pitch period as in Eq. (5). Time step used is of 15ms.

$$window\_length = \frac{3 * sampling\_frequency}{minimum\_pitch} \quad (5)$$

The number of frames used given by Eq. (6),

$$no.of.frames = \frac{speech\_length - window\_length}{window\_step} \quad (6)$$

Where window step is given by Eq. (7),

$$window\_step = time\_step * sampling\_frequency \quad (7)$$

2. For voiced and unvoiced decision:

The low pass butter worth filter is used for filtering the signal gaussian window is used for windowing of signal. Find normalized energy and normalized zerocrossing. Find the multiplication function as in Eq. (8),

$$MultValue = NormalizedEnergy * (1 - NormalizedZerocrossing) \quad (8)$$

For unvoiced signal MultValue is less because energy is less and zerocrossing is high giving (1-zerocrossing) very low value. For voiced signal MultValue is very high because energy is high and zerocrossing is low giving (1-zerocrossing) very high value. Make unvoiced decision if MultValue of speech is less than threshold value, which is found by Eq. (9),

$$threshold = mean\left(\frac{variance(MultValue)}{3} > MultValue > MinMultValue\right) \quad (9)$$

The speech with MultValue less than threshold is made zero i.e unvoiced sections are replaced with no speech signal.

3. For each frame steps used are,

- Center clipping operation: In this part for each frame select first one third samples and last one third samples. Find the max of each one third samples denote them by ipk1 and ipk2 and then select the segment whose peak value is minimum. Finally find the clipping level (cl) i.e. (0.68*min(ipk1,ipk2). Once the clipping levels is known then do following operation for that particular frame.

Case I: if sample value > cl replace it by +1
Case II: if sample value < -cl replace it by –1
Case III : replace remaining sample value by 0

- Autocorrelation: After doing the center clipping operation on the section perform the autocorrelation operation on the center clipped section. Now with this we can find max autocorrelation peak at 0th interval. Store the max value in some parameter and exactly next highest peak which arrive in the autocorrelation samples gives us the time period of peak. For getting exact position of peak use spline interpolation and then find fundamental frequency or pitch of the signal.

4. Use window steps to find pitch contour. Calculate the jump cost associated with a transition from F1 to F2 by Eq. (10),

$$jump\_\cos t = \log_2\left(\frac{F1}{F2}\right) \quad (10)$$

If jump cost is very high make the pitch value zero. This procedure is used because when there is transition from voiced to unvoiced section or vice versa there is a chance of getting false pitch value which is either very high or very low value from the previous consecutive pitch value.

5. Plot in the output window.
6. Get output from praat software for the same audio file and find mean square error for above algorithm and output of praat.

B. Cepstrum Method:

Step by step implementation:

1. Sectioning of signal:

Sectioning of signal is done by deciding the window length which depends on minimum and maximum pitch values (50Hz and 500Hz resp.). Window length is the length of the analysis frame chosen to be the three times the minimum pitch period as in Eq. (5). Time step used is of 15ms. The number of frames used given by Eq. (6), where window step is given by Eq. (7).

2. For each frame steps used are,

- Filtering: Low pass filter is used with cut off frequency 900Hz which is used to avoid the first formant.
- Windowing: In this part for each frame hamming window is used for windowing of signal.
- Use end point decision algorithm [16]. Find the energy and zerocrossing rate for signal. Set the threshold values as,

$Zth = 40$ is threshold for initial 10ms silence period.

$$IMN = mean(Initial10ms\_SilencePeriod\_energy) \quad (11)$$

$$IMX = \max(speech\_energy) \quad (12)$$

$$I1 = 0.03 * (IMX - IMN) + IMN \quad (13)$$

$$I2 = 4 * IMN \quad (14)$$

$$ITL = \min(I1, I2) \quad (15)$$

$$ITU = 5 * ITL \quad (16)$$

$$IZC = mean(Initial10ms\_SilencePeriod\_zerocross) \quad (17)$$

$$stdev = std(Initial10ms\_SilencePeriod\_zerocross) \quad (18)$$

$$IZCT = \min(Zth, IZC + 2 * stdev) \quad (19)$$

- Cepstrum: After doing windowing operation on the section perform the cepstrum operation on section.
- Liftering: Low_time liftering is used for each section. Now with this we can find max peak. Store the position and value of the maximum peak.
- Voiced/Unvoiced decision: If zerocross rate is less than IZCT and value of peak larger than a threshold (found by experiment and not fix) section is voiced. Find F0 by period of peak and for unvoiced section F0 is zero.

3. Use window steps to find pitch contour. Calculates the jump cost associated with a transition from F1 to F2 by Eq. (10). If jump cost is very high make the pitch value zero. This procedure is used because when there is transition from voiced to unvoiced section or vice versa there is a chance of getting false pitch value which is either very high or very low value from the previous consecutive pitch value.
4. Plot in the output window.
5. Get output from praat software for the same audio file and find mean square error for above algorithm and output of praat.

C. Praat Method:

Step by step implementation:

Computer program Praat is a research, publication and productivity tool for phonetician. With it we can analyze, synthesize and manipulate speech and create high quality picture for our articles and thesis.

A summary program praat, is given here:

1. Preprocessing: Low pass filter the entire signal with cutoff of 4KHz.
2. Compute the global absolute peak value of the signal.
3. Because the method is a short-term analysis method, the analysis is performed for a number of small segments (frames) that are taken from the signal in steps given by the Time Step parameter (0.015sec is used). For every frame, look for at most MaximumNumberOf CandidatesPerFrame (default is 4) lag-height pairs that are good candidates for the periodicity of this frame. This number includes the unvoiced candidate, which is always present.

The following steps are taken for each frame:

3.1. Take a segment from the signal. The length of this segment (the window Length) is determined by the minimum pitch parameter, which stands for the lowest fundamental frequency that you want to detect given by Eq.(5). The window should be just long enough to contain three periods (for pitch detection) of minimum pitch. E.g. If minimum pitch is 75 Hz, the window length is 40 ms for pitch detection. The number of frames used given by Eq. (6), where window step is given by Eq. (7).

3.2. Subtract the local average.

3.3. The first candidate is the unvoiced candidate, which is always present. The strength of this candidate is computed with two soft threshold parameters. e.g., if voicing threshold is 0.4 and silence threshold is 0.05, this frame bears a good chance of being analyzed as voiceless if there are no autocorrelation peaks above approximately 0.4 or if the local absolute peak value is less than approximately 0.05 times the global absolute peak value, which was computed in step 2.

3.4. Multiply by the window function.

3.5. Append half a window length of zeroes (because autocorrelation values add up to half a window length for interpolation is required).

3.6. Append zeroes until the number of samples is a power of two.

3.7. Find autocorrelation of window and speech segment.

3.8. Divide autocorrelation of speech segment by the autocorrelation of the window.

$$r_x(\tau) \approx \frac{r_a(\tau)}{r_w(\tau)} \quad (20)$$

This gives a sampled version of $r_x(\tau)$.

3.9. Find the places and heights of the maxima of the continuous version of $r_x(\tau)$. The only places considered for the maxima are those that yield a pitch between minimum pitch and maximum pitch. The only candidates that are remembered, are the unvoiced candidate, which has a local strength equal to

$$R \equiv VoicingThreshold + \max\left(0, 2 - \frac{LocalAbsolutePeak / GlobalAbsolutePeak}{SilenceThreshold / (1 + VoicingThreshold)}\right) \quad (21)$$

and the voiced candidates with the highest values of the local strength

$$R \equiv r(\tau_{\max}) - OctaveCost * \log(MinimumPitch * \tau_{\max}) \quad (22)$$

The octave cost parameter favors higher fundamental frequencies. One of the reasons for the existence of this parameter is that for a perfectly periodic signal all the peaks are equally high and we should choose the one with the lowest lag. Other reasons for this parameter are unwanted local downward octave jumps caused by additive noise, to minimize the number of incidental voiced/unvoiced decisions and large frequency jumps.

3.10. For every frame n, pn is a number between 1 and the number of candidates for that frame. Then jump cost is calculated for all candidates and the candidates with low jump cost are stored.

$$\text{cost}(\{P_n\}) = \sum_{2}^{no.of.frames} transitionCost(F_{(n-1)p_{(n-1)}}, F_{np_n}) - \sum_{1}^{no.of.frames} R_{np_n} \quad (23)$$

IV. EXPERIMENTAL RESULTS

To perform the comparative study of all implemented methods the assessment of each algorithm is done, test signals with constant value sine waves (150Hz,200Hz, 250Hz, 300Hz, 400Hz and 470Hz), synthesized vowels synthesized by replacing pitch tier of vowels (A/a/, U/u/ and I/i/) using Praat software of female1 speaker, recorded vowels and speeches spoken by three female, three male, three female child and three male child speakers are examined. Speech materials are 5 sec-long sentences spoken by every speaker sampled at 11025 Hz rate, which are recorded by Praat Software in mono tone mode. The reference files for fundamental frequency of audio signals are constructed by saving fundamental frequency obtained for same audio files in Praat software as output. For the performance evaluation of all methods, criteria considered in experimental work is mean square error (MSE) as given by,

$$MSE = \frac{1}{Fv} \sum_{l=1}^{Fv} \left[Fv(l) - Fr(l) \right]^2 \quad (24)$$

Where Fv is pitch of voice section per frame of algorithm output, l is no. of. Frames and Fr pitch of voiced section per frame of reference file.

The experimental conditions are tabulated as,

Table I MSE for synthesized vowel A/a/ experiment

A/a/	autocorrelation	cepstrun	Praatimplemented	Praatsoftware
A/a/200female	0.365	12.132	16.775	0.095
A/a/250female	0.627	164.521	41.596	0.126
A/a/150-200-250	4.990	16.246	27.440	1.080
A/a/200-150	0.452	9.417	305.501	0.083
A/a/ramp	591.743	764.466	150.255	2.008

Table II MSE for recorded vowels female1

vowels	autocorrelation	Cepstrum	Praat Implemented
A/a/	0.961	29.618	43.678
E/e/	0.570	52.921	82.295
I/i/	4.056	49.771	73.394
O/o/	0.684	39.037	55.614
U/u/	0.538	509.43	110.861

Table III MSE for recorded vowels male1

vowels	autocorrelation	Cepstrum	Praat Implemented
A/a/	0.118	3.048	4.1252
E/e/	1.172	8.725	15.4081
I/i/	0.748	3.634	5.2689
O/o/	0.059	6.754	9.3506
U/u/	0.058	7.999	12.4884

Table IV MSE for recorded vowels female child1

female child1	autocorrelation	Cepstrum	Praat Implemented
A/a/	1.233	40.606	417.775
E/e/	1.633	134.107	246.357
I/i/	0.937	63.192	86.029
O/o/	0.681	41.252	55.986
U/u/	43.064	140.257	200.520

Table V MSE for recorded vowels male child1

male child 1	autocorrelation	Cepstrum	Praat Implemented
A/a/	3.3884	136.2123	195.4971
E/e/	5.8583	550.5291	323.346
I/i/	2.3521	140.4874	1475.0538
O/o/	4.2118	252.9389	251.5208
U/u/	5.3113	143.453	324.675

Similar results are observed for other synthesized signals, recorded vowels (2 female, 2 male, 2 female child and 2 male child), recorded sentances for above speakers. From the tabulated information conclusions can be made.

V. CONCLUSION

Three pitch estimation algorithms are implemented and bsed on mean square error a comparative study is done for three female, three male, three female child and three male child. For voiced/unvoiced detection in autocorrelation method a threshold technique is introduced and for cepstrum part of end point detection method is used. Praat algorithm has it's own techniques for voiced/unvoiced detection and abrupt changes of pitch in successive frames. Experimental results conclude the followings:

1. Mean square error of Praatsoftware is very less for test signals and resynthesized vowels so Praatsoftware output is used as the reference for mean square error (MSE) calculation of all implemented algorithms (AUTOC, Cepstrum and implemented PRAAT) for recorded vowels and recorded speech.
2. For all implemented algorithms it is observed that mean square error increases as frequency of test signal increases so MSE values observed are larger for female and kids than male vowels and speech. This is because as frequency increases peaks of the sine signals come closer and closer giving location of peak little away from where it is actually present in a window.
3. MSE for recorded vowels of autocorrelation algorithm is less than the cepstrum algorithm, cepstrum gives worst results so rarely used for pitch detection.
4. Large variation in the voiced-unvoiced sections of a speech affects the results. Rapid increase and decrease in pitch values gives more MSE. Here voiced-unvoiced detection technique fails.

REFERENCES

[1] K. Abdullah-Al-Mamun, F. Sarker and G. Muhammad' "A High Resolution Pitch Detection Algorithm Based on AMDF and ACF", J. Sci. Res. 1 (3) ISSN: 2070-0237, pp.508-515, Aug.2009.

[2] L. R. Rabiner and R. W. Schafer , "Digital Processing of Speech Signals", Delhi, Pearson Education, 2004.

[3] D. O'Shaughnessey, "Speech Communication Human And Machine", Reading, Massachuselts: Addison-Wesley,1987.

[4] http://dea.brunel.ac.uk/ "Chapter13 : Speech Processing" downloaded on July 2012.

[5] P. Boersma, "Accurate short-term analysis of the fundamental frequency and the harmonics-to-noise ratio of a sampled sound", IFA Proceedings 17, pp. 97-110,1993.

[6] L. Rabiner, M. J. Cheng, A. E. Rosenberg, and C. A. McGonegal, "A comparative performance study of several pitch detection algorithms," IEEE Transactions on ASSP, vol. 24, No.5,pp. 399-417, October 1976.

[7] B. Gold and L. R. Rabiner, "Parallel processing techniques for estimating pitch period In Time domain", J. Acoust. Soc. Amer., vol. 46, No.2, Part 2, pp. 442-448, 1969.

[8] M. M Sodhi, "New Methods of Pitch Extraction," IEEE Tran. Audio Electroacoust., vol. 16, pp. 262-266, 1968.

[9] A. V. Patil, "Glottal Pitch Extractor ", IIT Bombay MTech. Diss., 2000.

[10] B. S. Atal and R. Rabiner," A Pattern Recognition Approach to Voiced-Unvoiced-Silence Classification with Applications to Speech Recognition", IEEE transactions on Acostics, speech, and signal processing, Vol. assp-24, no. 3, pp.201-212, June 1976.

[11] T. Shimamura and H. Kobayashi, "Weighted Autocorrelation for pitch Extraction of Noisy Speech", IEEE Transaction on Speech and Audio Processing, Vol. 9, No.7, PP.727-730. Oct.2001.

[12] P. K. Lehana and P. C. Pandey, "A Low Cost Impedance Glottograph and Glottal Pitch Analyzer", Proc. Bio Vision 2001 Int. Conf. Biomed. Eng., Bangalore, India, pp.1B. 33-37, Dec. 21-24, 2001.

[13] K. Rakesh, S. Dutta and K. Shama, "Gender Recognition using Speech processing techniques in lab view", International Journal of Advances in Engineering & Technology ISSN: Vol. 1, Issue 2, pp.51-63, May 2011.

[14] J. Chang, N. S. Kim and S. K. Mitra, "Pitch estimation of speech signal based on adaptive lattice notch filter", Elsevier -Signal Processing 85, pp.637–641, 2005.

[15] H. Traunmüller and A. Eriksson, "The frequency range of the voice fundamental in the speech of male and female adults," S-106 91 Stockholm, Sweden, downloaded on Sept. 2012.

[16] L. R. Rabiner and M. R. Sambur, " An Algorithm for determining the end points of isolated Utterances" , Bell System Tech. Jour., Vol. 54, No. 2, pp. 297-315, February 1975.

[17] S. Ahmadi and A. S. Spanias, "Cepstrum-Based Pitch Detection Using a New Statistical V/UV Classification Algorithm", IEEE transactions on speech and audio processing, vol. 7, no. 3, pp.333-338, May 1999.

[18] P. Boersma and D. Weenink, Praat: doing phonetics by computer (Version 4.3.01) [Computer program] 2005. Retrieved from http://www.Praat.org/

[19] L. R. Rabiner, "On the Use of Autocorrelation Analysis for Pitch Detection", IEEE transactions on acoustics, speech, and signal processing, vol. Assp-25, no. 1, pp.24-33, Feb. 1977.