Proceedings of the World Congress on Engineering 2010 Vol I
WCE 2010, June 30 - July 2, 2010, London, U.K.

Statistical Analysis Of Speech Signals By Weighted ASMDF

Roudra Chakraborty[1], Sagnik Sinha[2] and Debapriya Sengupta. *

Abstract: **In previous two papers, we have demonstrated tracking of fundamental frequency of voiced signals in clean and noisy environment by ASMDF and compared them with two classical methods autocorrelation and AMDF. It was shown that ASMDF gives near error-free results in case of clean signal. But in case of noisy signals, the percentage of relative errors have increased. Also weighted autocorrelation method gives somewhat better results than ASMDF. In this paper, we propose a new method which is derived by refining ASMDF and show that the results using this method are better than ASMDF and weighted autocorrelation.**

Index terms: AMDF,ASMDF, autocorrelation,Pitch tracking, Signal processing.

1 Introduction:ASMDF

Extraction or determination of fundamental frequency (or pitch) of a speech signal is a fundamental problem in both speech processing and speaker recognition. The typical pitch range for a male human being is 80-200 Hz, and for females 150-350 Hz. Many methods to extract the pitch of speech signals have been proposed. Improvements in accuracy of performance, robustness against noise of these methods are still desired. As a whole, there are no reliable and accurate method for pitch extraction. Also measuring the period of a speech waveform, varying in and with the detailed structure of the waveform, can be quite difficult. Another problem is automatic selection of the window of the voiced speech segments.

Autocorrelation method [15] and the average magnitude difference function (AMDF) method [21] are known to be the most primitive standard time-domain methods to find pitch. Based on these two methods several refinements like auditory modeling [6], probabilistic AMDF modeling [9], real-time digital hardware pitch detector [17], semiautomatic pitch detector (SAPD) [18], automatic formant analysis [19], weighted autocorrelation [22], modified autocorrelation and AMDF [23], projection measure technique [24], pseudo-pitch synchronous analysis [25] and many more [16] are proposed. Some other ideas on pitch extraction have also been discussed in the paper [12] and some tutorials [3] and [7].

In all the propositions, different functions of speech signals have been proposed, optimized (peaks or dips of the function have been traced) and a sequence of harmonics have been found. Then a 'best fit' value or the most significant harmonics has been chosen from the sequence and the frequency corresponding to this value has been defined as the fundamental frequency.

For an idealized speech signal in a stationary noisy environment the following mathematical abstraction has been consistently assumed in this thesis.

$$\widehat{y_n} = \sum_j a_j \exp\{\mathbf{i}f(j)\,n\} + \sum_j c_j z_j \exp\{\mathbf{i}jn\}, \qquad (1)$$

where $\{z_j\}$ is an uncorrelated sequence of random variables or white noise and $\{c_j\}$ are coefficients of discrete spectrum of stationary noise. In case the main signal $\{\widehat{y_n}\}$ possesses a pure fundamental frequency (or pitch), which is again another idealized view, it is assumed in this thesis that $f(j) = jf_p$ for a suitable pitch value f_p. The actual situation can become complicated further if the idealized signal contains multiple pitch streams or is convoluted with a channel filter. Therefore, from a practical point of view, the estimation of the pitch of a signal is essentially a statistical problem. Here a new method for extraction of fundamental frequency of speech signal in clean and to some extent noisy environment using simple statistical techniques is proposed. Motivationally, the technique has similarity with Zero Crossing based techniques ([7] and [11]), however, the theory is much simpler and statistical in nature. The main novelty of this approach actually lies in formulating the problem in this manner and putting a certain number of standard measures (such as, Autocorrelation and AMDF) of pitch in the same framework with the

*Manuscript received: March 9, 2010. [1,2]Mathematics Department, JADAVPUR UNIVERSITY, KOLKATA: 700032, INDIA. [3]BIRU, INDIAN STATISTICAL INSTITUTE, KOLKATA: 700108, INDIA. email:[1]roudra@gmail.com,[2]sagnik62@yahoo.co.in,*dps@isical.ac.in

proposed measure (Average Squared Mean Difference Function or ASMDF). This way a more comprehensive approach is presented and left open for further refinements.

Given a discrete time signal $y = (y_1, y_2, \ldots, y_n)$ (which is the real part of the complex idealized signal (1)), the autocovariance function is defined as

$$r_y(k) = \frac{1}{n-|k|} \sum_{i=0}^{n-|k|} (y_i - \overline{y})(y_{i+k} - \overline{y}), \tag{2}$$

and the autocorrelation function is defined by

$$\rho_y(k) = \frac{r_y(k)}{r_y(0)} \tag{3}$$

defined for all n and lag k.

A variation of autocorrelation analysis for measuring the periodicity of voiced speech uses the average magnitude difference function (AMDF), defined by the relation

$$D_y(k) = \frac{1}{n-|k|} \sum_{j=1}^{n-|k|} |y_{j+k} - y_j| \, .$$

In case of AMDF, $D_y(k)$ has been approximated by a scalar multiple of

$$\frac{1}{n-|k|} \left\{ \sum_{j=1}^{n-|k|} (y_{j+k} - y_j)^2 \right\}^{1/2}$$

which again has been approximated by the scalar multiple of $[2\{r_y(0) - r_y(k)\}]^{1/2}$, where $r_y(k)$ is the autocovariance defined as above. These approximations may suppress calculations important for pitch extraction.

Also AMDF can be considered as a replica of Ginis mean difference formula ([8], page 233-234). But the new method described below gives more optimal results as it is a replica of variance of the data set. Consider the voiced segment $y = (y_1, y_2, \ldots, y_n)$ in a digital speech signal. Since most speech signals can be viewed as quasi-periodic sequences the fundamental frequency may not be uniquely defined mathematically. In this approach the fundamental frequency is estimated, statistically enhancing the most significant harmonics present in y.

The algorithm for estimation of the pitch in this voiced segment is described below. For $1 \le i \le n$, and $k \ge 1$ consider the downsampled subsets (windows) of the original signal,

$$y_{i,k} = (y_{i+pk} : p = 0, \pm 1, \pm 2, \ldots).$$

Note that due to finiteness of the data stream, for each (i, k) pair one has to consider only those values of p so that y_{i+pk} is within the range. Furthermore, for several (i, k) pairs, $y_{i,k}$ will become singleton and they will not come under further considerations. One of the basic assumption is that the fundamental frequency of the voiced part is estimable for the given signal.

Next, let for each $k \ge k_0$, define

$$S_k = \{i : 1 \le i \le n;\; y_{i,k} \, \textbf{has at least two elements}\},$$

and, let q_k be the number of elements in S_k. This automatically sets an upperbound for $k \le k_{max} < ([(n+1)/2] - 1)$ ([x] being the greatest integer less than x). Finally define for k values with $q_k > 0$,

$$g(k) = \frac{1}{q_k} \sum_{i \in S_k} \mathbf{Var}(y_{i,k}), \tag{4}$$

where $\mathbf{Var}(y_{i,k})$ denotes the sample variance of the signal values in the subset $y_{i,k}$. It is interesting to note that aliasing, statistical precision of the sample variances (requiring larger values of q_k over a large range of k) of downsampled signals and robustness of the method under noisy environments considered are largely related issues.We call g the 'Average Squared Mean Difference Function' (ASMDF) [26],[27].

Let f_0 be the sample rate of the original speech signal and $f(k) = \frac{f_0}{k}$, where $k_0 \le k \le k_{max}$. In view of (4), g can be thought of as a function of f. Also g can be thought of as a mean squared mutual difference function which can be approximated with the standard autocorrelation function.

Let i be the index of the second minimum of the components of $g(k)$, i.e. $g(i) = min_k g(k)$ Then $f_p = f(i)$ is referred as the estimated fundamental frequency of the speech signal of y.

2 Proposed Method: WEIGHTED ASMDF

Let us assume that $\hat{y_n}$, is a speech signal given by

$$\hat{y_n} = A_n + Z_n, \tag{5}$$

where

$$A_n = \sum_j a_j \exp\{\mathbf{i} f(j)\, n\} \tag{6}$$

is clean and

$$Z_n = \sum_j c_j z_j \exp\{\mathbf{i} j n\} \tag{7}$$

is additive white Gaussian noise. In this case, the autocovariance function is given by,

ISBN: 978-988-17012-9-9
ISSN: 2078-0958 (Print); ISSN: 2078-0966 (Online)

$$\begin{aligned} r_y(k) = & \frac{1}{n-|k|}\sum_{i=0}^{n-|k|}(y_i-\overline{y})(y_{i+k}-\overline{y}) \\ = & \frac{1}{n-|k|}\sum_{i=0}^{n-|k|}(A_i+Z_i-\overline{A}-\overline{Z}) \\ & (A_{i+k}+Z_{i+k}-\overline{A}-\overline{Z}) \\ = & \frac{1}{n-|k|}\sum_{i=0}^{n-|k|}(A_i-\overline{A})(A_{i+k}-\overline{A}) \\ & +\frac{1}{n-|k|}\sum_{i=0}^{n-|k|}(A_i-\overline{A})(Z_{i+k}-\overline{Z}) \\ & +\frac{1}{n-|k|}\sum_{i=0}^{n-|k|}(Z_i-\overline{Z})(A_{i+k}-\overline{A}) \\ & +\frac{1}{n-|k|}\sum_{i=0}^{n-|k|}(Z_i-\overline{Z})(Z_{i+k}-\overline{Z}) \\ = & r_A(k)+r_{AZ}(k)+r_Z(k), \end{aligned} \tag{8}$$

where $r_{AZ}(k)$ is the crosscovariance function of A_n and Z_n.

For large n, $r_{AZ}(k) = 0$.

Therefore, $r_y(k) = r_A(k) + r_Z(k)$.

Now,

$$g_y(k) = \frac{1}{q_k}\sum_{i\in S_k}\mathbf{Var}(y_{i,k}),$$

where $q_k = |S_k|$ and $y_{i,k} = (y_{i+pk} : p = 0, \pm1, \pm2, \ldots)$. This leads to

$$g_y(k) = \frac{1}{q_k}\sum_{i\in S_k}\mathbf{Var}(A_{i,k}+Z_{i,k}),$$

where $A_{i,k} = (A_{i+pk} : p = 0, \pm1, \pm2, \ldots)$. and $Z_{i,k} = (Z_{i+pk} : p = 0, \pm1, \pm2, \ldots)$. Hence

$$\begin{aligned} g_y(k) \leq & \frac{1}{q_k}\sum_{i\in S_k}\mathbf{Var}(A_{i,k}) + \frac{1}{q_k}\sum_{i\in S_k}\mathbf{Var}(Z_{i,k}) \\ = & g_A(k)+g_Z(k). \end{aligned} \tag{9}$$

Hence the additive noise behaves independently with that included in $g_y(k)$.

Now since g was being minimized to get pitch of clean speech, $1/g_y(k)$ should give a peak and hence can be used to weight $r_y(k)$. It can be expected that the true peak can be found and errors can be minimized if the following function is used to get pitch of noisy speech :

$$\eta(k) = \frac{r_y(k)}{g_y(k)+\beta}. \tag{10}$$

where β is a constant. This η is maximized for different windows to get corresponding pitch values.

3 Experiments

One male (RL) and one female speaker (SB) each spoke 50 sentences, out of which, fifteen speeches of each speaker (with known pitch) were taken from FDA Evaluation Database [2] for experiment. Taking window size of 400 samples, data sets of pitch and the following graphs (where sample rate of each is given to be 20000Hz) of such speeches are found using autocorrelation, AMDF, ASMDF and weighted ASMDF.To save space, only few of such graphs are presented here. All the graphs have time (in ms) as horizontal axis and pitch (in Hz) as vertical axis.

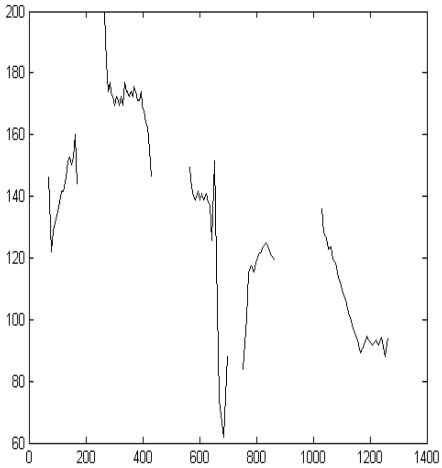

Fig. 1(a): Graph of true pitch values of speech RL001.

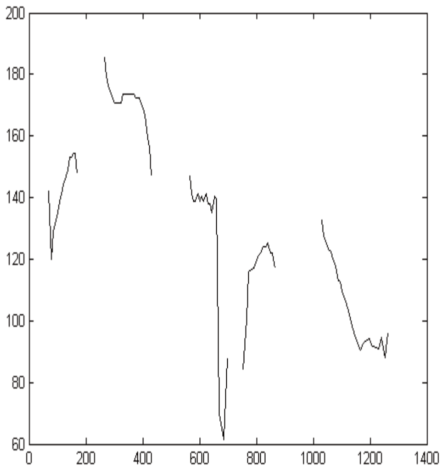

Fig. 1(b): Graph of ASMDF pitch values of speech RL001.

ISBN: 978-988-17012-9-9
ISSN: 2078-0958 (Print); ISSN: 2078-0966 (Online)

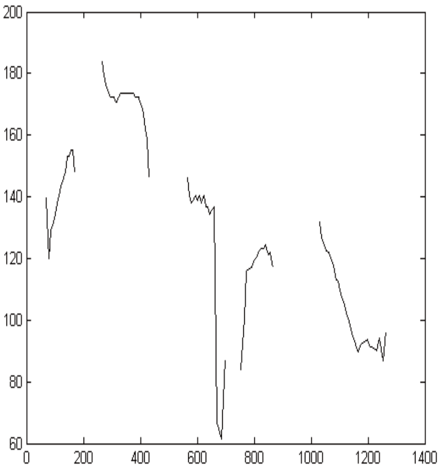

Fig. 1(c): Graph of autocorrelation pitch values of speech RL001.

Now Gaussian white noises of -5, -10, -15, -20, -25 and -30 decibels are added with all the thirty files. Then weighted ASMDF and later weighted autocorrelation have been applied and compared with the true pitch of clean signals. The results are in next section. One of the graphs of speech RL001 with -5DB of white noise is given below.

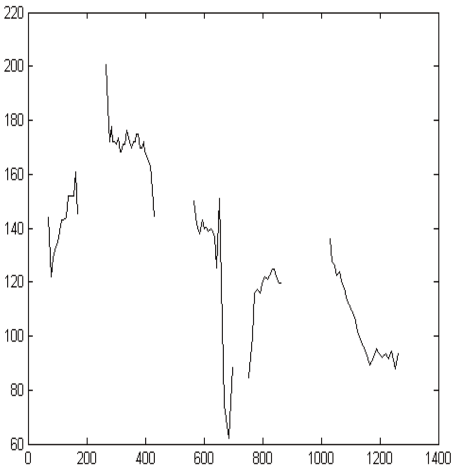

Fig. 2: Pitch graph of speech RL001 with -5 DB white noise using weighted ASMDF.

Independent discrete random points z_i with Gaussian distribution on interval (-1, 1) make the white noise. Correlated random points N_i are obtained by averaging of white noise in radius R_c sphere, i.e.

$$N_i = \sum_{j=-R_c}^{R_c} z_{i+j}. \tag{11}$$

Correlated noises of -10, -20, -30, -50 and -70 decibels and radii (R_c) 5, 25, 50 and 100 are added with these signals. Then weighted ASMDF and later weighted autocorrelation have been applied and compared with the true pitch of clean signals. The results are in next section. One graph of speech RL001 added with correlated noise of -50 decibel with R_c value of 100 using weighted ASMDF is given below.

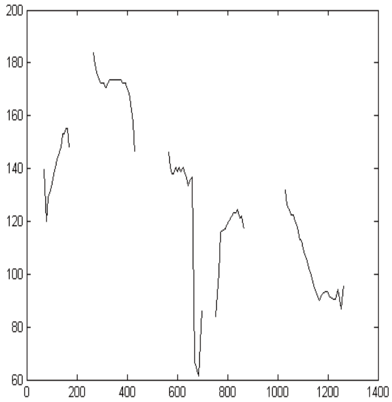

Fig. 3: Pitch graph of speech RL001 with correlated noise of -50 DB and $R_c = 100$ using weighted ASMDF.

Comparison

Now let us compare this η with that in the paper "Weighted Autocorrelation for Pitch Extraction of Noisy Speech" by Shimamura and Kobayashi [22]. The graphs of the same files using Shimamura and Kobayashi's method are found and one graph of speech RL001 added with correlated noise of -50 decibel with R_c value of 100 using weighted autocorrelation is shown below.

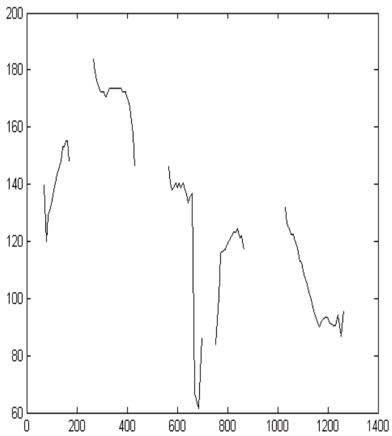

Fig. 4: Pitch graph of speech RL001 with correlated noise of -50 DB and $R_c = 100$ using wtd.autocorrln.

4 Comparative Analysis

Let P_c , P_m and P_s be the pitch contour found using autocorrelation, AMDF and ASMDF respectively calculated

ISBN: 978-988-17012-9-9
ISSN: 2078-0958 (Print); ISSN: 2078-0966 (Online)

over the same window for the above three speeches of the two speakers. Let P be the true pitch value over the same window. Let us denote by

$$e_s = (P - P_s)/P \text{ , } e_c = (P - P_c)/P \text{ and } e_m = (P - P_m)/P$$

the relative pitch error or percentage gross error. Let us define the standard deviation of the relative pitch error as

$$\sigma_e = \sqrt{\frac{1}{L_e - 1} \sum_{i=1}^{L_e} e^2(i) - \bar{e}^2}$$

where $e(i) = e_s(i), e_m(i), e_c(i)$; L_e being the length of each of the relative pitch error and $\bar{e} = \frac{1}{L_e} \sum_{i=1}^{L_e} e(i)$ is the mean pitch error.

Here, "$\sigma_e(s)$ for speech RL001 is 0.0248" means that gross pitch error for speech RL001 by ASMDF is 2.48% with respect to the true pitch value. If the standard deviation of gross pitch error is more than 20% (threshold) with respect to the true value, it's consistency is questionable.

The experimental values of σ_e for speeches with additive white noise and subsequently, additive correlated noise using ASMDF, weighted ASMDF and weighted autocorrelation are given in Tables I and II below:

Table I: Comparison of gross pitch errors in presence of White Noise.

Speeches	Decibels	ASMDF	Weighted ASMDF	Weighted Autocorrelation
RL001	-5	0.0626	0.0613	0.0667
RL001	-10	0.0565	0.0554	0.0654
RL001	-15	0.0633	0.0612	0.0692
RL001	-20	0.0652	0.0641	0.0679
RL001	-25	0.0578	0.0578	0.0752
RL001	-30	0.0585	0.0581	0.0667
RL002	-5	0.0615	0.0602	0.0602
RL002	-10	0.0572	0.0561	0.0592
RL002	-15	0.0581	0.058	0.0592
RL002	-20	0.0511	0.0512	0.0556
RL002	-25	0.0574	0.0567	0.0592
RL002	-30	0.0559	0.0558	0.0662
RL003	-5	0.0628	0.0621	0.0689
RL003	-10	0.0668	0.0643	0.0667
RL003	-15	0.0588	0.0588	0.0667
RL003	-20	0.0629	0.0628	0.0632
RL003	-25	0.0626	0.0613	0.0643
RL003	-30	0.0581	0.0572	0.0592
SB001	-5	0.1546	0.1541	0.1776
SB001	-10	0.1268	0.1262	0.1356
SB001	-15	0.1285	0.1281	0.1356
SB001	-20	0.1128	0.1116	0.1191
SB001	-25	0.1164	0.1154	0.1191
SB001	-30	0.1198	0.1192	0.1242
SB002	-5	0.1725	0.1725	0.198
SB002	-10	0.124	0.124	0.1291
SB002	-15	0.124	0.124	0.1377
SB002	-20	0.1297	0.1246	0.1355
SB002	-25	0.1179	0.1172	0.1285
SB002	-30	0.1179	0.1172	0.1289
SB003	-5	0.1428	0.1428	0.1778
SB003	-10	0.116	0.116	0.1291
SB003	-15	0.116	0.1158	0.1193
SB003	-20	0.1185	0.1158	0.1242
SB003	-25	0.1185	0.1182	0.1291
SB003	-30	0.124	0.121	0.1242

Table II: Comparison of gross pitch errors in presence of Correlated Noise.

Speeches	Decibels	Radii	ASMDF	Weighted ASMDF	Weighted Autocorrelation
RL001	-10	5	0.0285	0.0285	0.0369
RL001	-20	5	0.0248	0.024	0.0407
RL001	-30	5	0.0221	0.0221	0.0368
RL001	-50	5	0.0216	0.021	0.0244
RL001	-70	5	0.0261	0.0261	0.0267
RL001	-10	25	0.0792	0.0792	0.1148
RL001	-20	25	0.0671	0.0671	0.1048
RL001	-30	25	0.0364	0.0364	0.0548
RL001	-50	25	0.0211	0.0211	0.0238
RL001	-70	25	0.0235	0.0235	0.0268
RL001	-10	50	0.1586	0.1531	0.1734
RL001	-20	50	0.111	0.1101	0.1404
RL001	-30	50	0.0345	0.0345	0.0369
RL001	-50	50	0.0242	0.0242	0.0259
RL001	-70	50	0.0243	0.0243	0.0268
RL001	-10	100	0.1859	0.1212	0.1215
RL001	-20	100	0.1208	0.116	0.116
RL001	-30	100	0.0305	0.0292	0.0295
RL001	-50	100	0.0255	0.0255	0.0265
RL001	-70	100	0.0238	0.0238	0.0268
SB001	-10	5	0.0493	0.0493	0.0758
SB001	-20	5	0.0541	0.0541	0.0779
SB001	-30	5	0.0833	0.0833	0.0918
SB001	-50	5	0.0891	0.0879	0.1079
SB001	-70	5	0.0906	0.0906	0.1018
SB001	-10	25	0.1188	0.1126	0.1728
SB001	-20	25	0.1366	0.1328	0.1728
SB001	-30	25	0.1204	0.1108	0.1508
SB001	-50	25	0.0909	0.0909	0.1028
SB001	-70	25	0.0907	0.0907	0.1018
SB001	-10	50	0.2174	0.1524	0.1528
SB001	-20	50	0.1842	0.1359	0.1359
SB001	-30	50	0.1766	0.1015	0.1019
SB001	-50	50	0.0905	0.0905	0.1019
SB001	-70	50	0.0905	0.0905	0.1019
SB001	-10	100	0.2346	0.1914	0.1918
SB001	-20	100	0.2431	0.1945	0.1949
SB001	-30	100	0.1029	0.1015	0.1219
SB001	-50	100	0.0885	0.0885	0.1024
SB001	-70	100	0.0905	0.0905	0.1018

5 Conclusion

Based on the experimental results it has been shown that autocorrelation weighted by inverse ASMDF (or, conversely, ASMDF, weighted by inverse autocorrelation) is more useful than weighted autocorrelation method in noisy environments. Also there is a scope of using smoothing technique(s) for evaluation of this method.

References

[1] *The IViE Corpus.* Phonetics Laboratory, University of Oxford and Department of Linguistics, University of Cambridge, 1997-2002.

[2] PAUL BAGSHAW, *Fundamental Frequency Determination Algorithm (FDA) Evaluation Database.* Centre for Speech Technology Research, University of Edinburgh, 1993.

[3] JOSEPH P. CAMPBELL JR. *Speaker Recognition: A Tutorial.* Proceedings of the IEEE, 85(9): 1437-1462, September 1997.

[4] CHRIS CHATFIELD, *The Analysis of Time Series: An Introduction.* Chapman & Hall/CRC, Sixth Edition, 2004.

[5] D. G. CHILDERS, *Speech Processing and Synthesis Toolboxes.* John Wiley & Sons, Inc. 2000.

[6] P. COSI, S. PASQUIN AND E. ZOVATO. *Auditory Modeling Techniques for Robust Pitch Extraction and Noise Reduction.*

Proc. of ICSLP 1998, paper 1053, 1998, http://citeseer.ist.psu.edu/cosi98auditory.html.

[7] DAVID GERHARD. *Pitch Extraction and Fundamental Frequency: History and Current Techniques.* Technical Report TR-CS 2003-06: 1-22, November 2003, http://citeseer.ist.psu.edu/gerhard03pitch.html.

[8] A. M. GOON, M. K. GUPTA and B. DASGUPTA. *Fundamentals of Statistics Vol. 1.* Sixth Edition. The World Press Private Ltd. July 1983.

[9] LEAH H. JAMIESON, GOANGSHIUAN S. YING and CARL D. MICHELL. *A Probabilistic Approach to AMDF Pitch Detection.* Proceedings of the 1996 International Conference on Spoken Language Processing, Philadelphia, PA: 1201-1204, October 1996.

[10] FREDERICK JELINEK, *Statistical Methods for Speech Recognition.* The MIT Press, Cambridge, Massachusetts,1998.

[11] BENJAMIN KEDEM. *Spectral Analysis and Discrimination by Zero-crossings.* Proceedings of the IEEE, 74(11): 1477-1493, November 1986.

[12] SYLVIAN MARCHAND. *An Efficient Pitch-tracking Algorithm Using a Combination of Fourier Transforms.* Proceedings of the COST G-6 Conference on Digital Audio Effects (DAFX-01), Limerick, Ireland: 170-174, December 2001, http://citeseer.ist.psu.edu/marchand01efficient.html

[13] RUDRA PRATAP, *Getting Started with Matlab 5: A Quick Introduction for Scientists and Engineers.* Oxford University Press, India, 1999.

[14] THOMAS F. QUATIERI, *Discrete-Time Speech Signal Processing: Principles and Practice.* Pearson Education, Inc. 2004.

[15] L. R. RABINER. *On the use of Autocorrelation Analysis for Pitch Detection.* IEEE Trans. Acoust., Speech, Signal Processing, ASSP-25(1): 24-33, February 1977.

[16] L. R. RABINER, M. J. CHENG, A. E. ROSENBERG and C. A. MCGONEGAL. *A Comparative Performance Study of Several Pitch Detection Algorithms.* IEEE Trans. Acoust., Speech, Signal Processing, ASSP-24(5): 399-417, October 1976.

[17] L. R. RABINER, J. J. DUBNOWSKI and R. W. SCHAFER. *Real-Time Digital Hardware Pitch Detector.* IEEE Trans. on Acoustics, Speech, and Signal Processing, ASSP-24(1): 2-8, February 1976.

[18] L. R. RABINER, C. A. MCGONEGAL and A. E. ROSENBERG. *A Semiautomatic Pitch Detector (SAPD).* IEEE Trans. on Acoustics, Speech, and Signal Processing, ASSP-23(6): 570-574, December 1974.

[19] L. R. RABINER and R. W. SCHAFER. *System for Automatic Formant Analysis of Voiced Speech.* Journal of the Acoustical Society of America, 47(2): 634-648, February 1970.

[20] LAWRENCE RABINER and BIING-HWANG JUANG, *Fundamentals of Speech Recognition.* Prentice Hall PTR, 1993.

[21] M. J. ROSS, H. L. SHAFFER, A. COHEN, R. FREUDBERG and H. J. MANLEY. *Average Magnitude Difference Function Pitch Extractor.* IEEE Trans. Acoust., Speech, Signal Processing, ASSP-22: 353-362, October 1974.

[22] T. SHIMAMURA and H. KOBAYASHI. *Weighted Autocorrelation for Pitch Extraction of Noisy Speech.* IEEE Trans. on Speech and Audio Processing, 9(7): 727-730, October 2001.

[23] L. TAN and M. KARNJANADECHA. *Pitch Detection Algorithm: Autocorrelation Method and AMDF.* Proceedings of the 3rd International Symposium on Communications and Information Technology, 2: 551-556, September 2003.

[24] KUO-HWEI YUO, TAI-HWEI HWANG and HSIAO-CHUAN WANG. *Comparison of Autocorrelation-based Features and Projection Measure Technique for Speaker Identification.* IEEE Trans. on Speech and Audio Processing, 13(4): 565-574, July 2005.

[25] R. D. ZILCA, B. KINGSBURY, J. NAVRATIL and G. N. RAMASWAMY. *Pseudo Pitch Synchronous Analysis of Speech with Applications to Speaker Recognition.* IEEE Trans. on Speech and Audio Processing, 14(2): 467-478, March 2006.

[26] ROUDRA CHAKRABORTY, DEBAPRIYA SENGUPTA and SAGNIK SINHA. *Pitch Tracking of Acoustic Signals based on Average Squared Mean Difference Function.* Signal, Image and Video Processing, 3(4): 319-327, December 2009.
http://dx.doi.org/10.1007/s11760-008-0072-5

[27] ROUDRA CHAKRABORTY, DEBAPRIYA SENGUPTA and SAGNIK SINHA. *Use of ASMDF for Pitch Tracking of Noisy Speech Signals.* Proceedings of Second National Conference on Mathematical Techniques: Emerging Paradigms for Electronics and IT Industries (MATEIT-2008). Deen Dayal Upadhyaya College, University of Delhi, New Delhi,September 26-28, 2008.