

# Fundamental Frequency Extraction by Utilizing Accumulated Power Spectrum based Weighted Autocorrelation Function in Noisy Speech

## **Nargis Parvin**

Department of Computer Science and Engineering, Bangladesh Army International University of Science and Technology (BAIUST), Cumilla, Bangladesh  
E-mail: [nargis@baiust.edu.bd](mailto:nargis@baiust.edu.bd)  
<https://orcid.org/0009-0002-7469-2057>

## **Moinur Rahman**

Department of Information and Communication Technology, Comilla University, Cumilla, Bangladesh  
E-mail: [pranta7907@gmail.com](mailto:pranta7907@gmail.com)  
<https://orcid.org/0009-0000-0077-6093>

## **Irana Tabassum Ananna**

Department of Information and Communication Technology, Comilla University, Cumilla, Bangladesh  
E-mail: [tabassumananna3@gmail.com](mailto:tabassumananna3@gmail.com)  
<https://orcid.org/0009-0000-5110-7332>

## **Md. Saifur Rahman\***

Department of Information and Communication Technology, Comilla University, Cumilla, Bangladesh  
E-mail: [saifurice@gmail.com](mailto:saifurice@gmail.com)  
<https://orcid.org/0009-0008-0629-486X>  
\*Corresponding author

Received: 16 September 2023; Revised: 08 November 2023; Accepted: 21 January 2024; Published: 08 June 2024

**Abstract:** This research suggests an efficient idea that is better suited for speech processing applications for retrieving the accurate pitch from speech signal in noisy conditions. For this objective, we present a fundamental frequency extraction algorithm and that is tolerant to the non-stationary changes of the amplitude and frequency of the input signal. Moreover, we use an accumulated power spectrum instead of power spectrum, which uses the shorter sub-frames of the input signal to reduce the noise characteristics of the speech signals. To increase the accuracy of the fundamental frequency extraction we have concentrated on maintaining the speech harmonics in their original state and suppressing the noise elements involved in the noisy speech signal. The two stages that make up the suggested fundamental frequency extraction approach are producing the accumulated power spectrum of the speech signal and weighting it with the average magnitude difference function. As per the experiment results, the proposed technique appears to be better in noisy situations than other existing state-of-the-art methods such as Weighted Autocorrelation Function (WAF), PEFAC, and BaNa.

**Index Terms:** Accumulated Power Spectrum, Fundamental Frequency Extraction, Power Spectrum, Weighted Autocorrelation.

## **1. Introduction**

Extracting fundamental frequency also known as pitch has numerous uses in fields related to speech and plays a significant role in speech processing. A pitch extraction technique resistant to random noise is critical for the functioning of speech processing systems which include voice encoding, recognizing speaker, emotion detection, speech synthesis, text to speech conversion, gender identification, speaker verification and voice morphing. For speech separation, pitch is a robust feature with invariant properties [1]. In speech enhancement, the pitch extraction of noisy

signals are much difficult for speech's unstable features [2]. Due to this, numerous techniques for determining the fundamental frequency of speech signals have been developed [3-6]. However, in real world noisy circumstances, we must extract the fundamental frequency of speech signals, and an enhanced efficiency in noisy conditions is still expected. Unfortunately, fundamental frequency extraction in noisy situations, especially when the SNR level is low, lacks sufficient trustworthy and precise methods. Pitch extraction methods have previously been relied on speech signals characteristics like the periodic pattern within the time domain or the harmonic structure within the frequency domain.

In the time domain, various pitch extraction algorithms are employed in speech signal, like Autocorrelation function (ACF), Average magnitude difference function (AMDF), Average squared mean difference Function, Weighted autocorrelation function (WAF) and YIN [7-12]. Autocorrelation function (ACF) is that the most extensively utilized technique of pitch extraction which measures the similarity between two segments of a speech signal and finds the amount that provides the closest distance. This measurement of correlation between various delays of the input speech is completed by using ACF. Another correlation between various delays of the input speech is AMDF, which computes a symptom by computing the absolute value of the space among the lagged and current speech signal. Deep nulls appear at integral multiples of the pitch period of voiced sounds within the resulting value. The inverse of an AMDF is employed to weight an ACF in the WAF. Leading to that WAF is very effective to suppress noise. The YIN method improves the accuracy of the ACF by employing a function which calculates the differences of cumulative mean normalized squares of the speech signal. It also uses the post processing techniques. ACF-based pitch extraction techniques work well in noise and unaffected by phase distortion in waveform. On the opposite side, the effectiveness of ACF-based algorithms for pitch extraction is mostly reduced when the clean speech is masked by noise. The characteristics of the vocal tract also have an impact on the ACF.

In the frequency domain, numerous pitch extraction approaches are developed to cut back the vocal tract effect. One of the foremost prominent ways is that the cepstrum (CEP) method [13, 14]. The CEP is produced by performing an inverse Fourier transform on the log-amplitude spectrum. CEP uses the logarithmic function within the speech signal to isolate the periodic portions out from the vocal tract properties. During a noiseless environment, the CEP performs well, but in a very noisy one, the CEP's performance suffers greatly. Improved versions of the modified CEP (MCEP) and also the ACF of the log spectrum (ACLOS) were designed to handle this issue [15, 16]. In the frequency domain, a filter bank technique is also employed [17]. The windowless ACF (WLACF) based CEP (WLACF-CEP) makes use of both WLACF and CEP features [18]. The WLACF is efficient in reducing the noise effect from a speech signal without changing its periodicity. Within the WLACF-CEP the noise affect is curtailed from a loud speech signal so as to combine it with the CEP to enhance pitch extraction accuracy. The WLACF-CEP is demonstrated to be resilient against various sorts of noise.[18] Two advanced techniques have recently been discussed [19, 20]. The PEFAC is a frequency domain pitch extraction approach that employs log frequency domain sub-harmonic summations [19, 21]. So as to enhance its noise robustness, the PEFAC uses an amplitude compression technique. The BaNa technique, takes under consideration noisy speech spikes and chooses the primary five spectral spikes within the speech signal's amplitude spectrum [20]. Using an integrated pitch extraction algorithm the BaNa reliably extracts the pitch of the speech stream by calculating the fractions of the frequencies of the harmonic spikes with specified limit.

There are also various pitch extracting algorithms relying on the Hidden Markov Model (HMM) which construct pitch tracks by measuring the hidden patterns from sightings [22]. The use of multiple characteristics, neurons, and network topologies in neural networks (NN) is another novel strategy for pitch extraction [23]. Another pitch estimator is proposed in which speech signals are decomposed into subbands by an acoustic filter bank based on the time-frequency sparsity assumption [24]. Furthermore, the subband signals are encoded using an encoding model to get separate and trustworthy fundamental frequency assessments of potentially noisy and semi periodic voice sounds [25]. The standardized autocorrelation values of the captured subband signals are then used to select pitch estimates. But these methods are ineffective in real time scenario as they are time consuming, costly and based on complex post processing.

In this article, we propose to utilize the accumulated power spectrum (APS) instead of power spectrum of noisy speech to extract the fundamental frequency [26]. The APS approach is superior to conventional ACF because it is very simple to suppress the noise components and delivers outstanding fundamental frequency extraction efficiency in noisy scenarios. In the proposed method, the APS is used instead of the power spectrum in the WAF method. However, the proposed method utilizes the accumulated power spectrum and weighting it with the average magnitude difference function to display distinct harmonics and highlights the pitch peak while reducing the noise elements present in noisy speech.

## 2. Accumulated Power Spectrum

The time division component divides the framed speech signal  $y(j)$ ,  $0 \leq j \leq L-1$  into three subframes  $y_1(j)$ ,  $y_2(j-s)$  and  $y_3(j-2s)$  as follows:

$$y_1(j) = y(j), 0 \leq j \leq L - 1 \quad (1)$$

$$y_2(j - s) = y(j), S \leq j \leq S + M - 1 \quad (2)$$

$$y_3(j - 2s) = y(j), 2S \leq j \leq 2S + M - 1 \quad (3)$$

Here, the length of the subframe is represented by a number M. and S is a sample of the frame shift. Generally, it is preferred to set  $2S+M-1$  to be equal to L. Each power spectrum is determined as  $P_1^y(i)$ ,  $P_2^y(i)$  and  $P_3^y(i)$  out of each subframe  $y_a(n)$ ,  $a = 1, 2, 3, \dots, 0 \leq j \leq M - 1$ . For each frequency bin, the three power spectra are added together as follows:

$$P_1^y(i) = \sum_{a=1}^3 P_{y,jj}(i) \quad (4)$$

With the exception of the passband from 50[Hz] to 900[Hz], the collected power spectrum  $P_y(i)$  undergoes band pass filtering by forcing to zeros. The IFFT then performs an inverse Fourier transformation on the filtered power spectrum, resulting the APS which will be used in our proposed method. Fig. 1 shows a simplified representation of the APS technique.

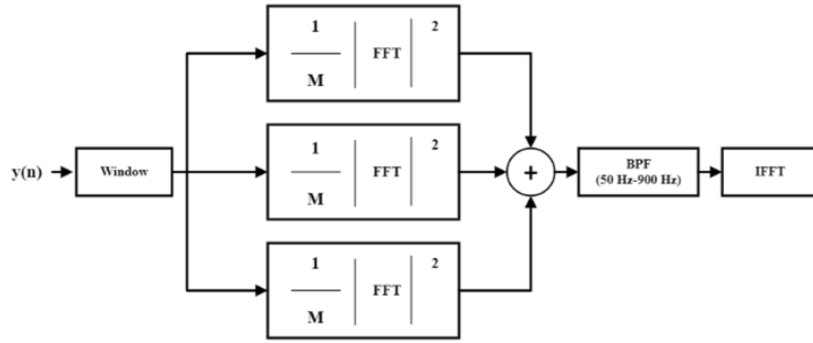


Fig.1. Visual representation of APS.

### 3. Materials and Methods

As demonstrated in Fig. 2, we present a novel method in this study for noise-resilient fundamental frequency extraction.

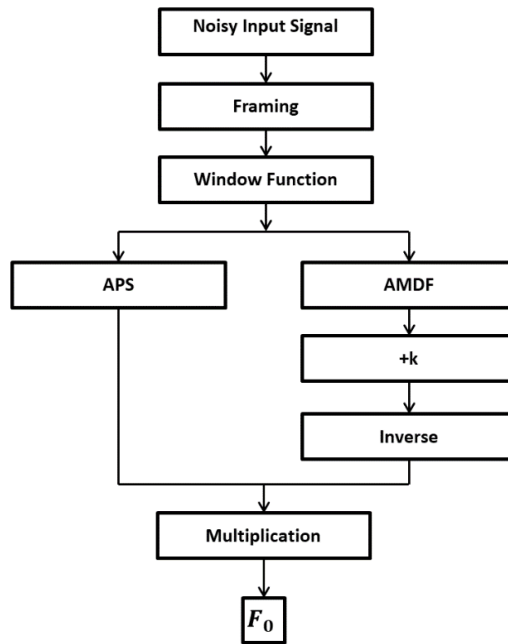


Fig.2. Block diagram of proposed method

Here we use the accumulated power spectrum based weighted autocorrelation function, where the ACF is replaced with APS. The weighted autocorrelation function is represented as:

$$\zeta(z) = \alpha(z) \delta(z) + k \quad (5)$$

Where,  $\alpha(z)$  = the APS;  $\delta(z)$  = the AMDF and  $k$  = a fixed number. The AMDF shares same properties as the autocorrelation function. The AMDF results in a notch, whereas the autocorrelation function gives a peak. However, the periodicity of both functions is practically the same. This technique makes use of the fact that, the noise portions of the autocorrelation function and AMDF respond separately in a noisy environment. The autocorrelation function's peak is highlighted when paired with the inversed AMDF in a noisy environment because of these uncorrelated properties. This concept led to the development of the weighted autocorrelation function for pitch detecting, which addresses the challenge of detecting the pitch from noisy speech signal. The accumulation-based pitch detection approaches use power spectra and ACFs from shorter subframes and a filter bank, respectively. They are adept to suppress noise elements while maintaining voice harmonics. As the ACF is improved to the APS method, which employs shorter subframes, it becomes noise-resistant even in circumstances with high noise levels.

Fig. 3 demonstrates that the frequent use of time division processing in the APS technique contributes to noise reduction, particularly in low SNR circumstances.

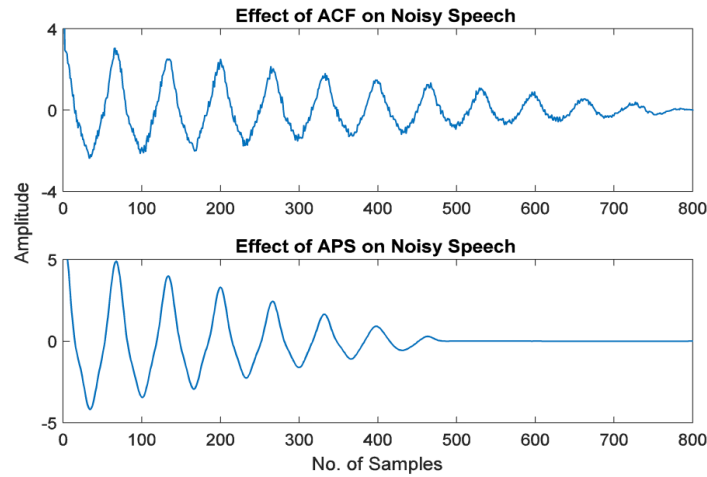


Fig.3. Effect of APS and ACF on noisy speech.

From fig. 4, we can observe that the second largest peak in traditional WAF may occur at  $z < T$  or  $T < z \leq 2T$  or elsewhere as noise increases. As a result, at low SNRs, WAF cannot be utilized to determine the pitch duration accurately. But in the proposed method while computing the WAF, APS is deployed to the noisy speech signal instead of ACF to mitigate the impact of noise. Pitch extraction accuracy is improved by the proposed way over the traditional WAF methodology.

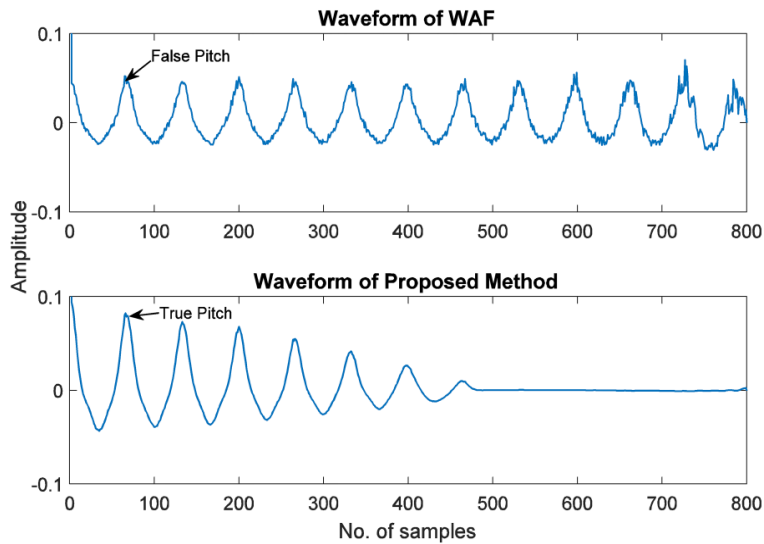


Fig.4. Waveform of WAF and Proposed method

## 4. Results and Discussion

### 4.1. Experimental Details

For experiments KEELE [27] database and NTT [28] database were used. KEELE database contains ten English speaker's speeches having duration of 6 [m] and sampling rate of 16 [KHz]. NTT database contains eight Japanese speaker's speeches having duration of 10 seconds each with 10 [kHz] sampling rate and 3.4 [kHz] band restriction. This research suggests an efficient idea that is better suited for speech processing applications for retrieving the accurate pitch from speech signal in noisy conditions. White Gaussian noise, Train noise, HF Channel noise, and Babble noise was added to each speech signal. Also, we used SNR levels from -5 to 20 [dB] and the additional experimental parameters such as frame length, window function, and DFT (IDFT) points are 50 [ms], rectangular, and 2048 except PEFAC and BaNa. On the other hand, frame shift is 10 [ms].

### 4.2. Evaluation Criteria

Here, the effectiveness of the suggested approach is evaluated by determining the error rate in accordance with Rabinar's rule [3] as follows:

$$E_r(z) = F_{est}(z) - F_{true}(z) \quad (6)$$

Where,  $z$  is the frame number,  $F_{est}(z)$  is the extracted fundamental frequency of the  $z$ -th frame and  $F_{true}(z)$  is the true fundamental frequency of the  $z$ -th frame. If  $|e(z)| > 10\%$  of the  $F_{true}(z)$ , then the error was identified as Gross Pitch Error (GPE) and the percentage of this error was determined throughout the entire voiced frames in the speech data. For the fundamental frequency extraction, we solely recognized and evaluated voiced elements of sentences. To extract the fundamental frequency, we utilized the search range of  $f_{min} = 50\text{Hz}$  and  $f_{max} = 400\text{Hz}$ , this range represents the most common range of a human's fundamental frequency.

### 4.3. Performance Comparison and Result Discussion

The effectiveness of pitch extraction in noisy situations was assessed between the proposed method (PROP) and the conventional methods (WAF, PEFAC, and BaNa). Here, we consider four forms of noise, namely white, babble, train, and HF channel noises. With the exception of length of the frame, window function and quantity of DFT(IDFT) points for PEFAC and BaNa, all the factors of the existing techniques were identical to those of the proposed technique. Hamming window function was used in both BaNa and PEFAC. The frame duration for BaNa was configured to 60 [ms], and  $2^{16}$  points were used for the DFT (IDFT) points. The source code of BaNa was implemented in this environment which is ideal for BaNa [20, 29]. For PEFAC, Hamming window function and 90 [ms] was utilized as the window function and frame length respectively. The source code uses  $2^{13}$  as the value for the DFT (IDFT) points. The source code of PEFAC was implemented in this environment which is ideal for BaNa [19, 30].

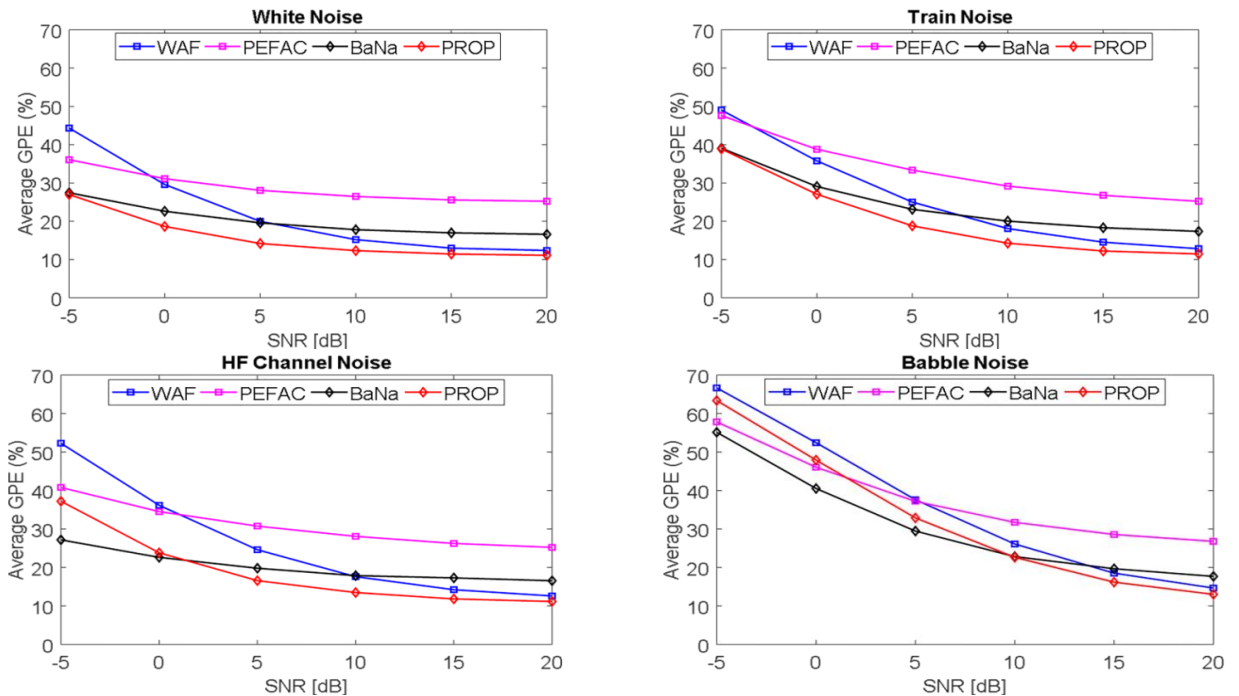


Fig.5. In different noises, the average gross pitch error rate (GPE) for the KEELE database

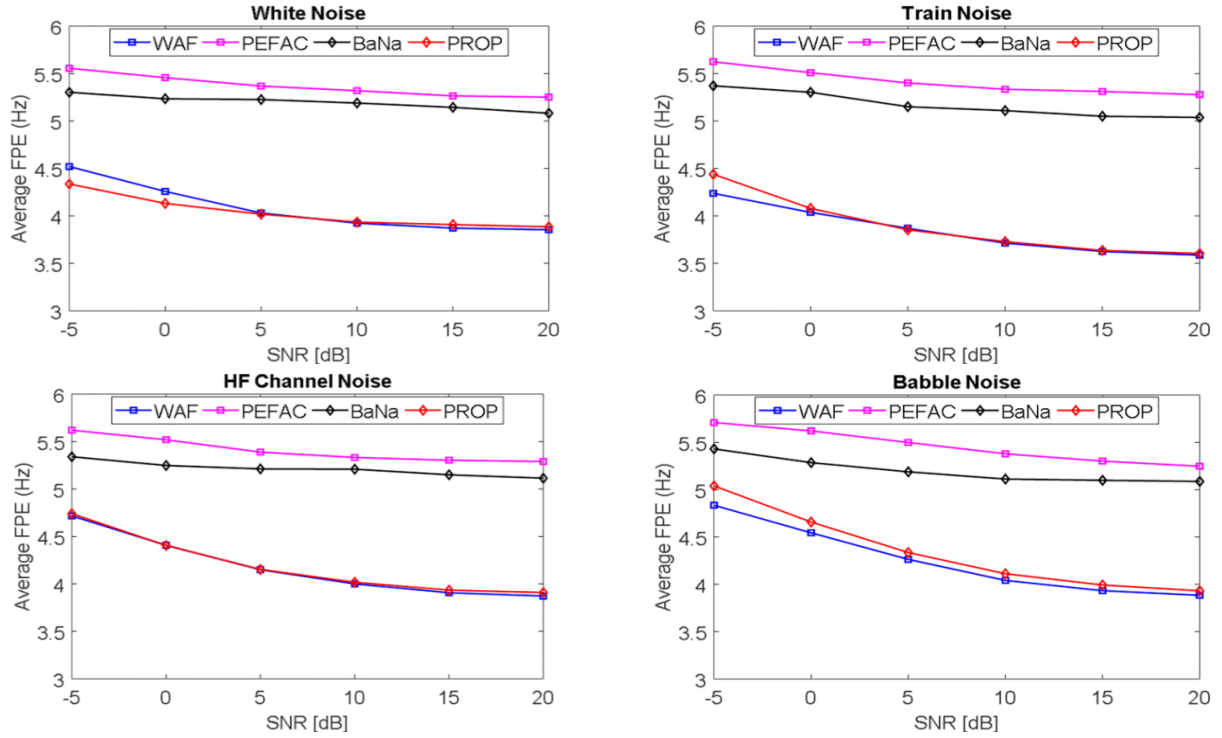


Fig.6. In different noises, the average fine pitch error rate (FPE) for the KEELE database

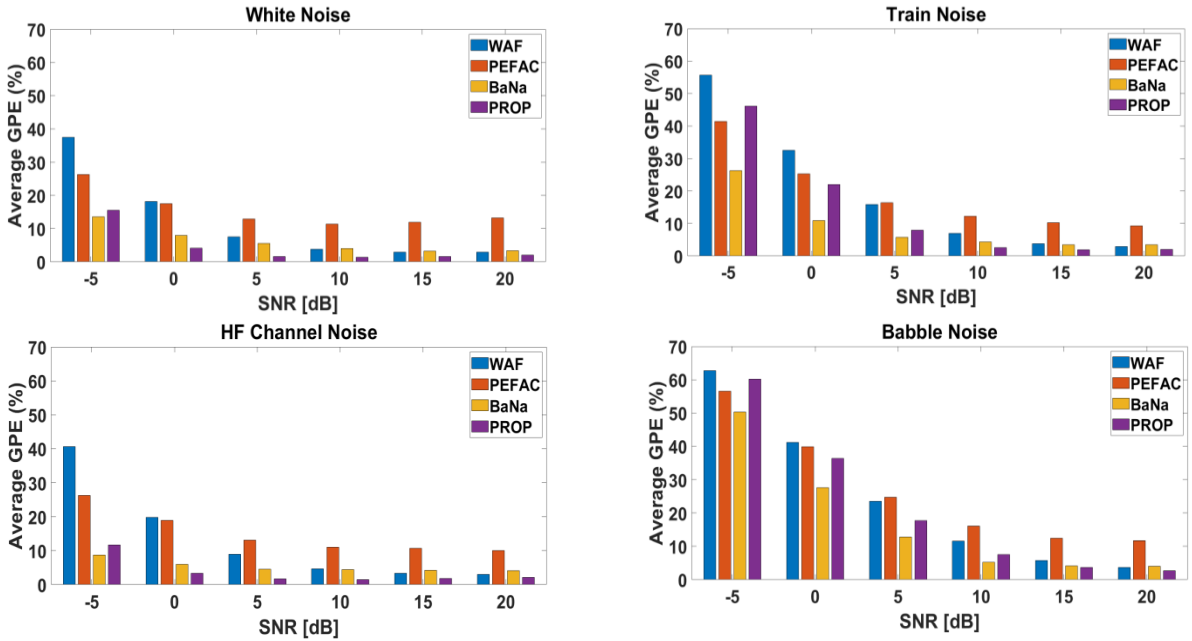


Fig.7. In different noises, the average gross pitch error rate (GPE) for the NTT database

Fig. 5 and Fig. 7 displays the average GPE rate in KEELE database with various types of noise. Fig. 6 and Fig. 8 shows the average fine pitch error (FPE) in KEELE database in different noises.

According to Fig. 5 and Fig. 7, the suggested approach (PROP) provides the lowest average GPE rate of all the techniques under comparison at practically all SNRs for white noise, train noise, and HF channel noise. BaNa offers the reduced gross pitch error rate in HF channel noise at low SNR (-5 [dB]). In babble noise at low SNRs less than 10 [dB]

BaNa outperforms other approaches. The proposed approach (PROP), on the other hand, offers a decreased gross pitch error rate at high SNRs. The true peak is significantly impacted by the noise peaks at low SNRs with babble noise. Peak extraction becomes challenging as a result. BaNa benefits in this situation by choosing the first five spectral peaks and relies on a highly effective post-processing on the noisy speech spectrum to select the more correct pitch information.

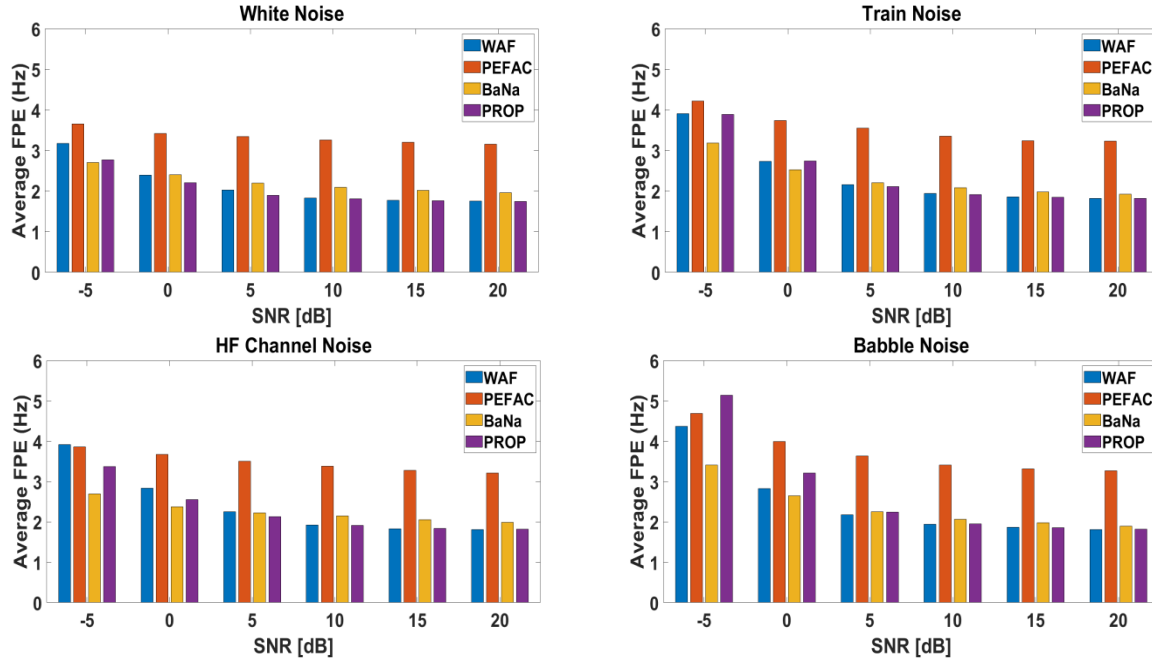


Fig.8. In different noises, the average fine pitch error rate (FPE) for the NTT database

From Fig. 6 and Fig. 8, we observed that the range of FPE from 1.8 [Hz] to 5.6 [Hz], where FPE represents the degree of fluctuation in pitch detection. We have also investigated that the proposed and WAF methods are more reasonable than PEFAC and BaNa in the case of FPE. On the other hand, innovative methods (PEFAC and BaNa) provide the higher FPE than that of proposed and WAF methods at almost all SNRs in all noise cases.

#### 4.4. Computational Time

Table 1 presents the computational time per second of speech for all methods in KEELE database. For generating the processing time, we have used all methods with 11<sup>th</sup> Gen Intel(R) Core (TM) i5-1135G7, CPU clock speed is 2.40GHz and memory is 8 (GB). Also, in order to calculate the processing time for accurate measurement, we took five trials for each approach into account. The processing time of proposed method is highly competitive with WAF and PEFAC methods. The PEFAC method is slightly improved due to the use of log frequency domain, which is highly affected by noise. As a result of using a big FFT size, BaNa, on the other hand, uses a high processing time.

Table 1. Processing time per second of speech

BaNa	PEFAC	WAF	Prop.
32.12	2.53	4.96	4.93

## 5. Conclusions

The Fundamental frequency extraction is an unavoidable task for speech signal processing functionalities which becomes challenging by being corrupted with noise in real world. The proposed method does better in separating noise from the waveform than other methods, especially in Gaussian white noise and train noise. For this characteristic it achieves less average GPE rate than other methods without any complicated post-processing. It also can overcome the vocal tract effect efficiently by making unnecessary ripples from the waveform equalized.

## References

- [1] X. Zhang, H. Zhang, S. Nie, G. Gao and W. Liu, "A Pairwise Algorithm Using the Deep Stacking Network for Speech Separation and Pitch Estimation," IEEE/ACM Transactions on Audio, Speech, and Language Processing, Vol. 24, No. 6, pp. 1066-1078, 2016, doi: 10.1109/ICASSP.2015.7177969.
- [2] J. Stahl and P. Mowlaee, "A Pitch-Synchronous Simultaneous Detection-Estimation Framework for Speech Enhancement," IEEE/ACM Transactions on Audio, Speech, and Language Processing, Vol. 26, No. 2, pp. 436-450, 2018, doi: 10.1109/TASLP.2017.2779405.
- [3] L. Rabiner, M. Cheng, A. Rosenberg and C. McGonegal, "A comparative performance study of several pitch detection algorithms," IEEE Transactions on Acoustics, Speech, and Signal Processing, Vol. 24, No. 5, pp. 399-418, 1976, doi: 10.1109/TASSP.1976.1162846.
- [4] K. A. Oh and C. K. Un, "A performance comparison of pitch extraction algorithms for noisy speech," Proceedings under IEEE International Conference on Acoustics, Speech, Signal Processing, pp. 18B4.1-18B4.4, 1984, doi:



- 10.1109/ICASSP.1984.1172551.
- [5] L. Sukhostat and Y. Imamverdiyev, "A comparative analysis of pitch detection methods under the influence of different noise conditions," *Journal of voice*, Vol. 29, No. 4, pp. 410-417, 2015, doi: 10.1016/j.jvoice.2014.09.016.
- [6] W. J. Hess, "Pitch Determination of Speech Signals," Berlin, Germany: Springer-Verlag, 1983, doi: 10.1007/978-3-642-81926-1.
- [7] L. R. Rabiner, "On the use of autocorrelation analysis for pitch detection", *IEEE Transaction on Acoustics, Speech, Signal Processing*, Vol. ASSP-25, No. 1, pp. 24–33, 1977, doi: 10.1109/TASSP.1977.1162905.
- [8] M. Ross, H. Shaffer, A. Cohen, R. Freudberg and H. Manley, "Average magnitude difference function pitch extractor," *IEEE Transactions on Acoustics, Speech, and Signal Processing*, Vol. 22, No. 5, pp. 353-362, 1974, doi: 10.1109/TASSP.1974.1162598.
- [9] Un CK, Yang S, "A pitch extraction algorithm based on LPC inverse filtering and AMDF," *IEEE Transactions on Acoustics, Speech, and Signal Processing*, Vol. 25, No.6, pp. 353-362, 1977, doi: 10.1109/TASSP.1977.1163005.
- [10] R. Chakraborty, D. Sengupta, and S. Sinha, "Pitch tracking of acoustic signals based on average squared mean difference function," *Signal, image and video processing*, Vol. 3, No. 4, pp. 319–327, 2009, doi: 10.1007/s11760-008-0072-5.
- [11] T. Shimamura and H. Kobayashi, "Weighted autocorrelation for pitch extraction of noisy speech," *IEEE Transactions on Speech and Audio Processing*, Vol. 9, No. 7, pp. 727-730, 2001, doi: 10.1109/89.952490.
- [12] A. De Cheveigne and H. Kawahara, "Yin, a fundamental frequency estimator for speech and music," *The Journal of the Acoustical Society of America*, Vol. 111, No. 4, pp. 1917–1930, 2002, doi: 10.1121/1.1458024.
- [13] A. M. Noll, "Short-time spectrum and cepstrum techniques for vocal-pitch detection," *The Journal of the Acoustical Society of America*, Vol. 36, No. 2, pp. 296–302, 1964, doi: 10.1121/1.1918949.
- [14] 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, 1999, doi: 10.1109/89.759042.
- [15] Kobayashi H, Shimamura T., "A modified cepstrum method for pitch extraction," *Proceedings of IEEE Asia-Pacific International Conference on Circuits and Systems Microelectronics and Integrating Systems (APCCAS)*, 1998, doi: 10.1109/APCCAS.1998.743751.
- [16] Kunieda N, Shimamura T, Suzuki J, "Pitch extraction by using autocorrelation function on the log spectrum," *Electronics and Communications in Japan, Part 3*, Vol. 83, No.1, pp. 90–98, 2000, doi: 10.1002/(SICI)1520-6440(200001)83.
- [17] Lahat M, Niederjohn RJ, Krubsack DA. "A spectral autocorrelation method for measurement of the fundamental frequency of noise-corrupted speech," *IEEE Transactions on Acoustics, Speech, and Signal Processing*, Vol.35, No. 6, pp. 741-750, 1987, doi: 10.1109/TASSP.1987.1165224.
- [18] Hasan MAFMR, Rahman MS, Shimamura T. "Windowless autocorrelation-based cepstrum method for pitch extraction of noisy speech," *Journal of Signal Processing*, Vol. 16, No. 3, pp. 231-239, 2012, doi: 10.2299/jsp.16.231.
- [19] S. Gonzalez and M. Brookes, "PEFAC - A Pitch Estimation Algorithm Robust to High Levels of Noise," *IEEE/ACM Transactions on Audio, Speech, and Language Processing*, Vol. 22, No. 2, pp. 518-530, 2014, doi: 10.1109/TASLP.2013.2295918.
- [20] N. Yang, H. Ba, W. Cai, I. Demirkol and W. Heintzelman, "BaNa: A Noise Resilient Fundamental Frequency Detection Algorithm for Speech and Music," *IEEE/ACM Transactions on Audio, Speech, and Language Processing*, Vol. 22, No. 12, pp. 1833- 1848, 2014, doi: 10.1109/TASLP.2014.2352453.
- [21] Hermes DJ, "Measurement of pitch by subharmonic summation," *Journal of the Acoustical Society of America*, Vol.83, No.1, pp. 257–264, 1988, doi: 10.1121/1.396427.
- [22] D. Wang, C. Yu, and J. H. Hansen, "Robust harmonic features for classification-based pitch estimation," *IEEE/ACM Transaction on Audio, Speech, Language Processing*, Vol. 25, No. 5, pp. 952–964, 2017, doi: 10.1109/TASLP.2017. 2667879..
- [23] Y. Liu and D. Wang, "Speaker-dependent multi pitch tracking using deep neural networks," *The Journal of the Acoustical Society of America*, Vol. 141, No. 2, pp. 710–721, 2017, doi: 10.1121/1.4973687.
- [24] S. Lin, "Robust Pitch Estimation and Tracking for Speakers Based on Subband Encoding and The Generalized Labeled Multi-Bernoulli Filter," *IEEE/ACM Transactions on Audio, Speech, and Language Processing*, Vol. 27, No. 4, pp. 827-841, 2019, doi: 10.1109/TASLP.2019.2898818.
- [25] S. Lin, "A new frequency coverage metric and a new subband encoding model, with an application in pitch estimation," *Proceedings of Annual Conference of the International Speech Communication Association*, pp. 2147–2151, 2018, doi: 10.21437/Interspeech.2018-2590.
- [26] M. S. Rahman, Y. Sugiura, and T Shimamura, "Utilization of windowing effect and accumulated autocorrelation function and power spectrum for pitch detection in noisy environments," *IEEE Transactions on Electrical and Electronic Engineering*, Vol. 15, No. 11, pp. 1681–1690, 2020, doi: 10.1002/tee.23238.
- [27] Plante F, Meyer G, Ainsworth W, "A fundamental frequency extraction reference database," *Proceedings of the Eurospeech*, pp. 837–840, 1995, doi: 10.21437/Eurospeech.1995-191.
- [28] 20 Countries Language Database, NTT Advanced Technology Corp., Jpn, (1988)
- [29] Wcng, "Wireless communication networking group, [Online]. Available, <http://www.ece.rochester.edu/projects/wcng/code.html>".
- [30] M. Brookes, "Voicebox toolkit, [Online]. Available, <http://www.ee.ic.ac.uk/hp/staff/dmb/voicebox/voicebox.html>"



## Authors' Profiles



**Miss. Nargis Parvin** received her B.Sc. (Honours) and M.Sc. degrees in Information and Communication Engineering from University of Rajshahi, Rajshahi, Bangladesh, in 2006 and 2007, respectively.

In 2013, she joined as a Lecturer in the Department of Computer Science and Engineering, Bangladesh Army International University of Science and Technology (BAIUST), Cumilla Cantonment, Cumilla, Bangladesh, where she is currently serving as an Assistant Professor. She pursued her Ph.D. degree in the field of speech processing at the Graduate School of Science and Engineering at Saitama University, Japan. Her research interests include speech analysis and digital signal processing. She can be contacted at email: nargis@baiust.edu.bd.



**Moinur Rahman** received his B.Sc. (Engineering) and M.Sc. (Engineering) degrees in Information and Communication Technology from Comilla University, Cumilla, Bangladesh, in 2018 and 2019, respectively.

In 2022, he joined as a Lecturer in the Department of Computer Science and Engineering, The People's University of Bangladesh, 3/2 Asad Avenue, Dhaka, Bangladesh. Now he is currently serving as a lecturer in the Department of Information Technology, University of Information Technology and Sciences, Baridhara, Dhaka, Bangladesh from March 2023. His current research interests include speech analysis and digital signal processing. He can be contacted at email: pranta7907@gmail.com.



**Irana Tabassum Ananna** received her B.Sc. (Engineering) and M.Sc. (Engineering) degrees in Information and Communication Technology from Comilla University, Cumilla, Bangladesh, in 2019 and 2020, respectively. Her current research interests include speech analysis and digital signal processing. She can be contacted at email: tabassumananna3@gmail.com.



**Md. Saifur Rahman** received his B.Sc. (Honours) and M.Sc. degrees in Information and Communication Engineering from University of Rajshahi, Rajshahi, Bangladesh, in 2006 and 2007, respectively.

In 2012, he joined as a Lecturer in the Department of Information and Communication Technology, Comilla University, Cumilla, Bangladesh, where he is currently serving as an Associate Professor. He pursued his Ph.D. degree in the field of speech processing at the Graduate School of Science and Engineering at Saitama University, Japan. His research interests include speech analysis and digital signal processing. He can be contacted at email: saifurice@gmail.com.

**How to cite this paper:** Nargis Parvin, Moinur Rahman, Irana Tabassum Ananna, Md. Saifur Rahman, "Fundamental Frequency Extraction by Utilizing Accumulated Power Spectrum based Weighted Autocorrelation Function in Noisy Speech", International Journal of Information Technology and Computer Science(IJITCS), Vol.16, No.3, pp.52-60, 2024. DOI:10.5815/ijitcs.2024.03.05