SUST J Sci Tech, Vol 33(1) 2023, 13-21 ISSN- 2959-3522

Group Delay Moment of Cepstrum for Formant Estimation of High-Pitched Noisy Speech

(Submitted: 8.11.2022; Accepted: 08.06.2023)

Husne Ara Chowdhury*

Department of Computer Science and Engineering, Shahjalal University of Science and Technology, Sylhet 3114, Bangladesh
*Corresponding Author's Email: husna-cse@sust.edu

Abstract

The estimation task of formant frequencies is challenging for some spectral estimation issues. However, it is significant in the female or child speech-processing arena. This paper proposes a method for calculating the formant frequencies of high-pitched speech employing third-order group delay moment (GDM) of cepstrum. The GDM is a time domain equivalent signal estimated using the inverse discrete Fourier transform (DFT) of group delay spectrum (GDS). The GDS is calculated from the cepstrum. The stabilized spectral root cepstrum (SRC) is used in place of log-based cepstrum to obtain better control of the noisy speech spectrum. The resultant GDM becomes a vocal tract-dominated signal with noise-robust as well. The efficiency of the proposed method has been shown by calculating the formant values of some synthetic vowels against different fundamental frequency variations from 100 Hz to 400 Hz. Additionally, standard F2-F1 plots obtained from the natural vowel sounds of male and female speakers are demonstrated. An utterance from the TIMIT corpus has been utilized to plot the formant contours on the respective spectrogram. The results are likened to two related sophisticated methods. The proposed technique outperforms both approaches, especially when high-pitched speaking in the presence of ambient noise.

Keywords: Deconvolution; Group delay; Spectral root cepstrum; Stabilization; Noise

1. Introduction

A signal utterance is emanated by the convolution of the originating signal passing through the glottis with the vocal tract filter. Formant estimation is essential for some practical situations such as speech recognition, synthesis, speech coding, voicing detection, dysphonic severity detection, etc. Recently deep learning techniques improve the accuracy of these tasks. However, such type of methods pays more computational complexity. Thus for light-weight applications traditional formant extraction task is still demandable.

The linear predictive coding (LPC) methods [1-2] are extensively used for source-filter separation. However, the accuracy relies on the appropriate model order selection. The cepstrum method [3-4] has a long history to achieve successful segregation of the filter from the source. Most of these methods consider a noise-free environment. It was found that the SRC method is more effective in a noisy case than the log-based cepstrum. Naturally, different noises corrupt the speech source. The methods described in [5-7] consider the noisy environment for speech analysis. However, these methods do not have special analyses for high-pitched cases. For high-pitched speech, original formant peaks are distorted or masked by the nearest harmonics, even in the case of clean speech. So, the model-based or peak-picking method alone is not perfect for high-pitched speech analysis. Rahman and Shimamura [8] identified the aliasing effect in high-pitched speech as a barrier. In my previous research [9], the high-resolution GDS [10] is estimated in place of the magnitude spectrum as an excellent tool to improve the accuracy of the high-pitched speech analysis. The higher-order GDS as a higher-order statistical (HOS) tool has an attractive property [11] to suppress the Gaussian noise. The SRC [12] method is used for deconvolution to gain control when estimating the spectrum. Firstly, the SRC is stabilized yielding reduced noise effect. After truncating it by the liftering window, the inverse SRC is applied to convert it back to the time domain equivalent signal. Then the third-order GDM signal is calculated, which reduces the noise noticeably.

The accuracy of the proposed method has been shown by comparing it with the two state-of-the-art methods: the modified magnitude spectrum (MMS) [13] method and the WORLD formant estimator. In the MMS method, the magnitude spectrum is modified by the GDS. The Cheap Trick process of [14], described in the WORLD vocoder [15], provides a cepstrally smoothed spectral envelope. The auto-regressive (AR) coefficients are calculated from this spectrum for formant estimation, which is referred to as a WORLD formant estimator. The analysis conducted by choosing different analysis parameters showed that the proposed method outperforms the above two techniques in case of noisy and high-pitched utterances.

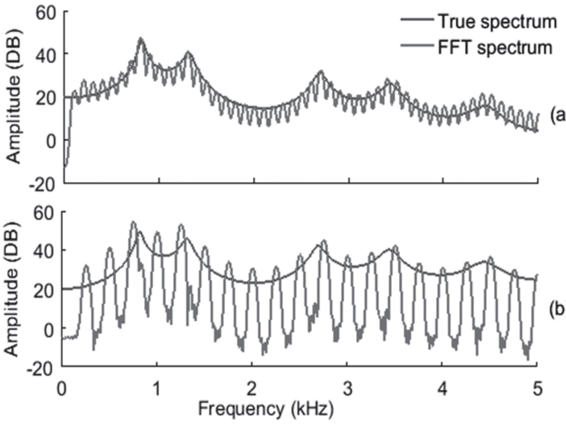


Fig. 1: A speech spectrum of a synthetic vowel /a/ compared with the 'true' spectrum (a) at F0=100 Hz and (b) at F0=250 Hz.

2. Problem Analysis

A voiced speech segment is *x(n)*. The DFT of *x(n)* can be expressed as

$$X(k) = \frac{1}{N}\sum_{n=0}^{N-1} x(n)e^{-j\frac{2\pi}{N}kn}, \qquad (1)$$

where N is the number of DFT points and $0\leq k\leq N-1$.

2.1 Issues of High-Pitched Spectral Estimation

According to Eq. (1), more harmonics exist in low-pitched speech, and its spectrum conforms with the 'true' spectrum, shown in Figure 1 (a), where the 'true' spectrum means the spectrum calculated from the vocal tract impulse response. The pitch period decreases for high-pitched speech. All harmonics are congested to the narrow region yielding overlapping of harmonics of speech components. The resulting spectrum shows a smaller number of wider harmonics. Therefore, the information on some components is lost in the high-pitched speech spectrum. The harmonics, which may conform with the true resonance peaks are masked or distorted by the nearest harmonics, as observed in Figure 1 (b). Thus, analyzing high-pitched speech becomes a challenging task.

2.2 Group Delay Spectrum

An alternative representation of the Eq. (1) is

$$X(k) = |X(k)|e^{\varphi_{X(k)}}, \text{ where}$$

$$\varphi_{X(k)} = arctan\frac{X_I(k)}{X_R(k)}$$

The $|X(k)|$ denotes the magnitude spectrum, whereas $\varphi_{X(k)}$ is the phase spectrum, and $X_R(k)$ and $X_I(k)$ indicate the real and imaginary parts of the spectrum *X(k)*. The GDS is the other delegation of the phase spectrum. It is calculated using any one of the followings

$$\tau(k) = -\frac{d\{\arg(X(k))\}}{dk}, \text{ or alternatively} \qquad (2)$$

$$\tau(k) = -\frac{X_R(k)X'_I(k)-X_I(k)X'_R(k)}{|X(k)|^2},$$

where $'$ indicates the derivative with respect to k. The primary concern about the GDS is its spikiness. Bozkurt [16] identified the cause of it as the lack of synchronization window with the glottal closure instant. Along with this, the GDS estimation besets lots

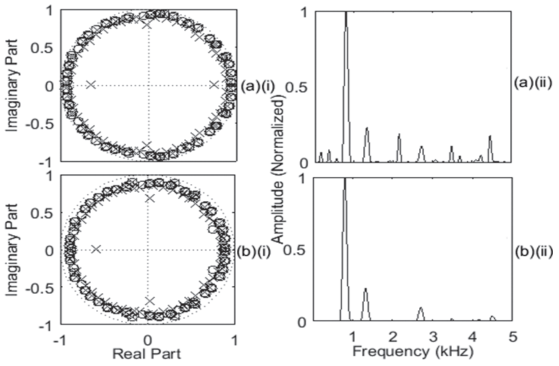


Fig. 2: The pole-zero plots and the third-order GD spectrum of a noisy synthetic vowel /a/ at Pink noise, SNR=15 are compared (a) at r = 1 and (b) at r = 0.95.

of issues. These are discussed in my previous research [9]. Unfortunately, this GDS and the higher-order GDS become spiky for noisy speech, which will become discernible from the next subsection.

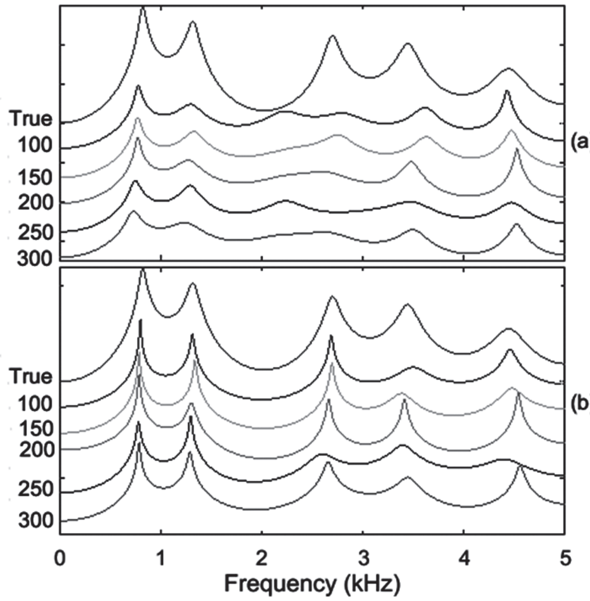


Fig. 3: The LP spectrum, computed using the proposed method, on a noisy vowel /a/ (Pink noise, SNR=15), at different pitch values are compared with the 'true' spectrum, (a) at r = 1 and (b) at r = 0.95.

2.3 Stability Violation in Noisy Speech

I calculated the higher-order GDS from the time-domain equivalent signal obtained from the spectral root (real) cepstrum. There are no issues when computing the GDS from such a causal and stable clean signal. When the signal becomes noisy, the roots near the unit circle produce spurious peaks on that spectrum, violating the system's stability. The third-order GDS, named group delay (GD) bi-spectrum computed from such a signal, fails to emphasize the formant peaks, as illustrated in Figure 2 (a). Bozkurt et al. [16] proved that the roots away from the unit circle emphasize the formant peaks in the GDS, which is shown in Figure 2 (b). The AR coefficients estimated at pole/zero radii $r = 1$ provide confusing formant information. Again, $r = 0.95$ conforms to the original formant values, as shown in Figure 3.

3. Proposed Method

For successful deconvolution, I employed the stabilized spectral root cepstrum after adjusting the appropriate γ value. A GD bi-spectrum is then calculated from the inverse spectral root cepstrum for noise reduction. The overall method is illustrated briefly in the block diagram, shown in Figure 4. The next few subsections describe all of these relevancies.

3.1 Spectral Root Cepstrum

The convolution of impulse train with the vocal tract impulse response forms the speech signal *x(n).* Consider a revertible system such that,

$$\hat{X}(k) = X^{\gamma}(k).$$

Here *X(k)* is the DFT of *x(n)*. For real spectral root cepstrum

$$\hat{x}(n) = \frac{1}{N}\sum_{k=0}^{N-1} |X(k)|^{\gamma} e^{j\frac{2\pi}{N}kn}, \qquad (3)$$

where $0 \leq n \leq N - 1$ and $\hat{x}(n)$ in Eq. (3) indicates the spectral root cepstrum. That means the spectral root cepstrum method converts the convolutional vector space to another vector space. Thus, the filter information can be easily separated from the source by truncating it by a liftering window.

3.2 Stabilization

The stabilized spectral root cepstrum plays a significant role when emphasizing the vocal tract impulse response. The system represented by Eq. (3) is stabilized by shifting the roots inside the unit circle ($r<1$), yielding the stabilized signal $\hat{x}_s(n)$ as

$$\hat{x}_s(n) = \hat{x}(n)(r)^n,$$

where $0.5 \leq r \leq 1$. Selecting the lower value of *r*, near *0.5*, diminishes the spectral peaks along with the possible formant peaks. Again, an unstabilized signal impacts the overall system performance, as discussed in the previous section.

3.3 Choice of Liftering Window for Truncating

A fixed-sized liftering window would be the simplest form of truncation. Since the underlying system is unpredictable as male or female speech, complexity arises to define such an independent cepstral window. Thus the window size varies according to the pitch values. From the study of [17], I have used *0.5T* as a length of cepstral window *w(n)* for low-pitched speech.

$$w(n) = \begin{cases} 1, & if\ 0 < n \leq 0.5T \\ 0, & otherwise. \end{cases}$$

Here, T is the signal period. After truncating, the signal becomes as

$$\hat{x}_c(n) = \hat{x}_s(n)w(n).$$

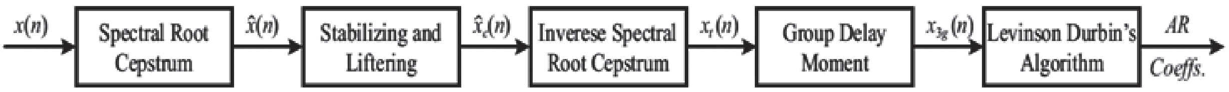
The high-pitched speech shows less information within 0<n≤0.5T. So, it is convenient to select the window length of 0.7 T as the cepstral window for the high-pitched speech.

3.4 Inverse Spectral Root Cepstrum

Application of inverse SRC operation converts the truncated signal $\hat{x}_c(n)$ to $\hat{x}_s(n)$ as

$$x_t(n) = \frac{1}{N}\sum_{k=0}^{N-1} \left|\hat{X}_c(k)\right|^{\frac{1}{\gamma}} e^{j\frac{2\pi}{N}kn}, \qquad (4)$$

where $\hat{X}_c(k)$ is the DFT of $\hat{x}_c(n)$. The signal $\hat{x}_t(n)$ is time domain equivalent signal.



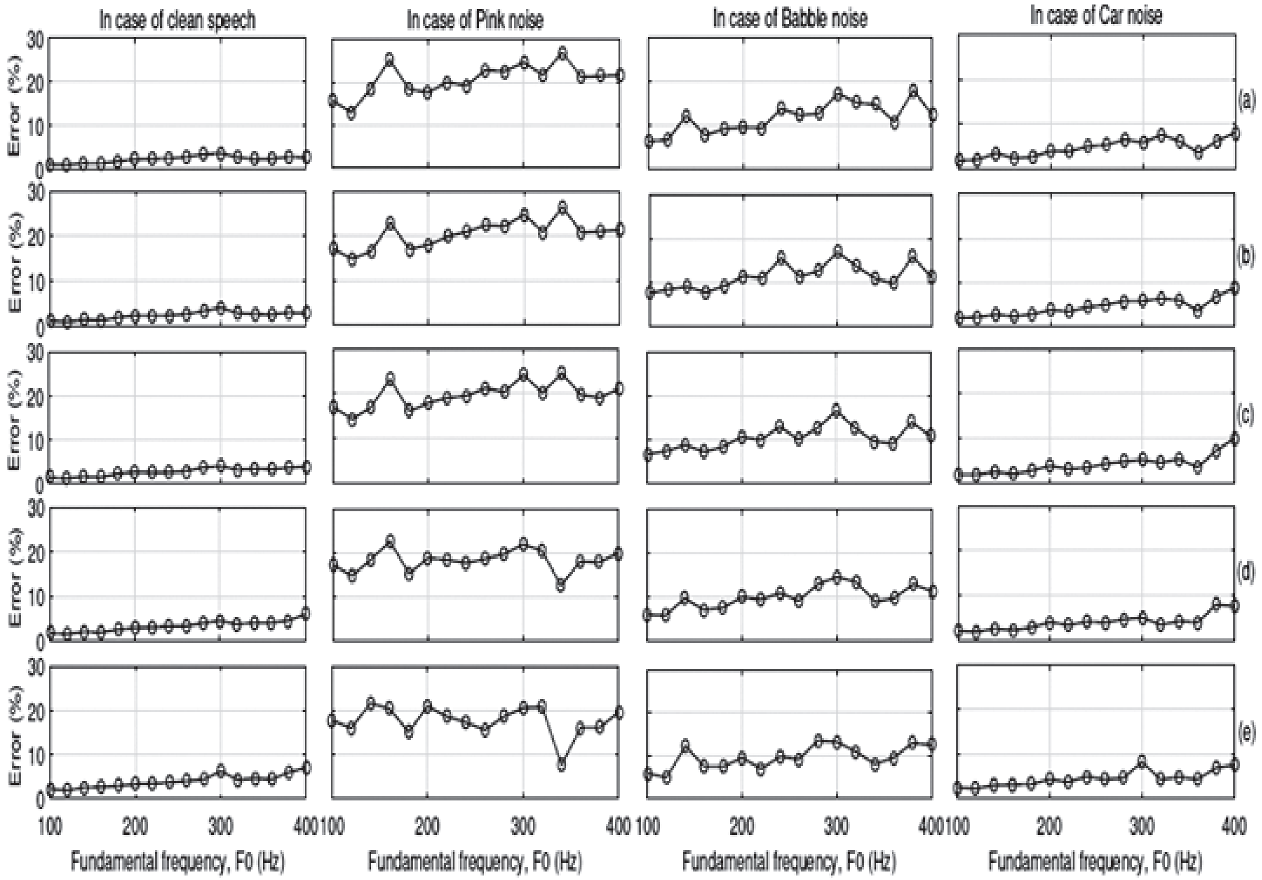


Fig. 5: The γ value impact on the synthetic speech. The result is shown by estimating the average of first three formants using the proposed method (a) at γ = 0.01, (b) at γ = 0.1, (c) at γ = 0.25, (d) at γ = 0.5, and (e) at γ = 0.75. The first column represents the results in the case of clean speech. The second, third, and fourth columns represent the results of noisy speech added with the Pink, Babble, and Car noises, respectively, at SNR = 10 dB.

3.5 Group Delay Moment

The GDS can be calculated effectively using the minimum phase signal, which avoids the related issues of GDS estimation [9]. Stability is ensured in $\hat{x}_s(n)$ by shifting the roots inside the unit circle. The truncated signal $\hat{x}_c(n)$ is a causal signal. Using Eq. (4), the calculated signal $\hat{x}_t(n)$, is a minimum phase signal appropriate for the improved GDS calculation. Using Eq. (2), the improved GDS is calculated as

$$\tau_t(k) = -\frac{d\{\arg(X_t(k))\}}{dk}.$$

The $X_t(k)$ is the DFT of $x_t(n)$. From the study of [18], the negative values contravene the system's causality

Thus, the GDS is half-wave rectified as follows

$$\tau_t(k) = \begin{cases} \tau_t(k), & if\ \tau_t(k) > 0 \\ 0, & Otherwise. \end{cases}$$

However, the above GDS is not noise-robust as expected. So, I used the GD bi-spectrum, which is calculated by

$$\tau_{3t}(k) = \tau_t^3(k).$$

Application of the inverse DFT converts $\tau_{3t}(k)$ to the third-order GDM as

$$x_{3g}(n) = \frac{1}{N}\sum_{k=0}^{N-1} \tau_{3t}(k)\, e^{j\frac{2\pi}{N}kn}.$$

The signal $x_{3g}(n)$ reflects the vocal tract impulse response with a reduced noise effect.

3.6 Formants Estimation

The signal $x_{3g}(n)$ is the third-order GDM, where the noise effect is minimized. Then the auto-regressive (AR) coefficients are calculated from it. Employing the Levinson–Durbin's algorithm [19] the said coefficients are estimated. Using the root-solving method, the formant frequencies are estimated. This method is dependent on appropriate model order selection.

Table 1 Formant frequencies to synthesize the vowels

Vowels	F1	F2	F3	F4	F5
/a/	813	1313	2688	3438	4438
/i/	375	2188	2938	3438	4438
/u/	375	1063	2188	3438	4438
/e/	438	1863	2688	3438	4438
/o/	438	1063	2688	3438	4438

4. Experimental Analysis and Discussion

The experiments are conducted by estimating the formant frequencies of some synthetic vowels and some natural vowels. A few utterances of male and female speakers from the TIMIT [20] database is also tested to show efficiency. The formant values are calculated for each speech segment using the following equation

$$\hat{F} = \frac{1}{N}\sum_{n=1}^{N} F_n, \tag{5}$$

where F_n is the formant frequency of n^{th} windowed speech segment.

4.1 Experimental Setups

The Liljancrants-Fant glottal model [21] is an effective tool for reproducing the signal source. For experimenting the synthetic speech, this model is used to simulate a signal source. Five formant frequencies from Table 1 are used to synthesize the five vowels, where a 10 *kHz* sampling frequency is employed. The speech signal is segmented using a Hamming window of 20 *ms* and analyzed by a 1 *ms* frame shift. A first-order differentiator $1 - z^{-1}$ is applied as a pre-emphasis filter to each speech segment. The five formants' bandwidths are 60, 100, 120, 175, and 281 *Hz* settled. The DFT size is fixed to 1024 with the analysis order 12. Then the formant values are calculated for each speech frame using Eq. (5). The formant estimation error of each vowel in a percentage is calculated by

$$EF_i = \left(\frac{1}{5}\sum_{i=1}^{5}\frac{|\hat{F}_{ij} - F_{ij}|}{F_{ij}}\right) * 100,$$

where F_{ij} indicates the i^{th} formant frequency of the j^{th} vowel from table 1 and $\hat{F}_{ij}$ is the calculated value. The average errors (%) of the first three formants are estimated using the following formula as

$$E_{avg} = \left(\frac{1}{5}\frac{1}{3}\sum_{j=1}^{5}\sum_{i=1}^{3}\frac{|\hat{F}_{ij} - F_{ij}|}{F_{ij}}\right) * 100.$$

4.2 Adjusting the γ Value

Utilization of the γ value on $|X(k)|^\gamma$ does not change the values of spectral peaks or pitch harmonics. The spectral root cepstrum method is consistent with the cepstrum method owing to the $()^{\gamma th}$ and $()^{\frac{1}{\gamma}th}$ power function is taken in place of the *logarithmic* and *exponential* function, respectively. Since there is no theoretical idea about the choice of γ, I have to depend on empirical observations. According to the study from [12], it should be $0 \leq \gamma \leq 1$ for the real cepstrum. The study about the γ value impact on formant estimation for both clean and noisy speech is shown in Figure 5. My analysis shows that a lower γ value is more reasonable for clean speech. However, this value does not apply to noisy speech. A higher γ value for the noisy speech shows better results for the source affected by the Pink, Babble, and Car noises.

4.3. Clean Synthetic Speech Analysis

The result of clean synthetic speech analysis is shown in Figure 6. The proposed method exhibits a reliable outcome for all three formants, even in a high-pitched region. Although the WORLD formant estimator shows improved accuracy in the low-pitched regions, it fails to enhance the accuracy of high-pitched locations.

4.4. Noisy Synthetic Speech Analysis

When the noises such as the Babble and Car noises at SNR=10, 20, and 30 dB corrupt the signal source, the

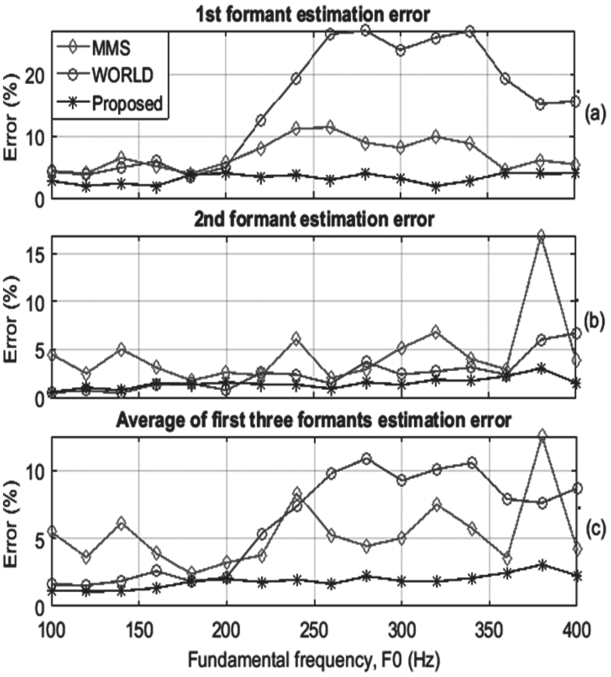


Fig. 6: The formant estimation errors (%) of five synthetic vowels are calculated at different fundamental frequencies up to 400 Hz for the (a) 1st formant, (b) 2nd formant, and (c) average of the first three formants.

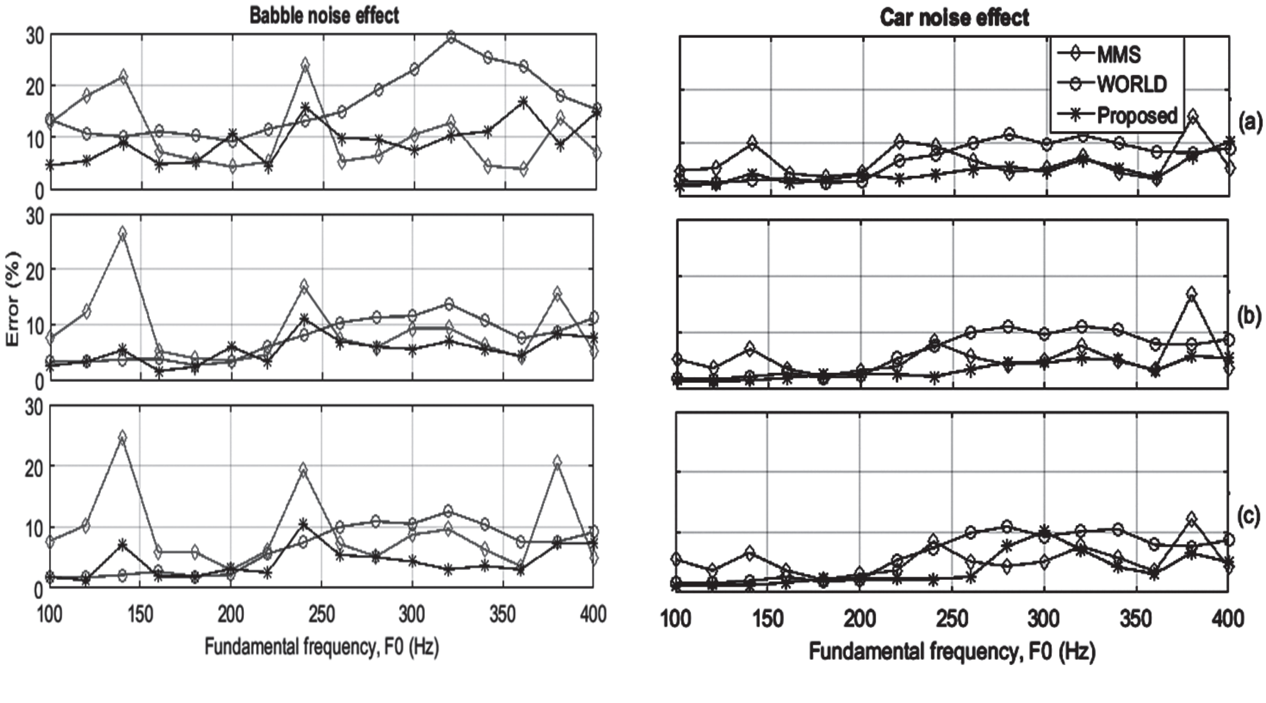


Fig. 7: The average error (%) of the first three formants, E_{avg} of a noisy (first and second column show the Babble and Car noise effect, respectively) synthetic speech signal with (a) SNR=10 dB, (b) SNR=20 dB, and (c) SNR=30 dB

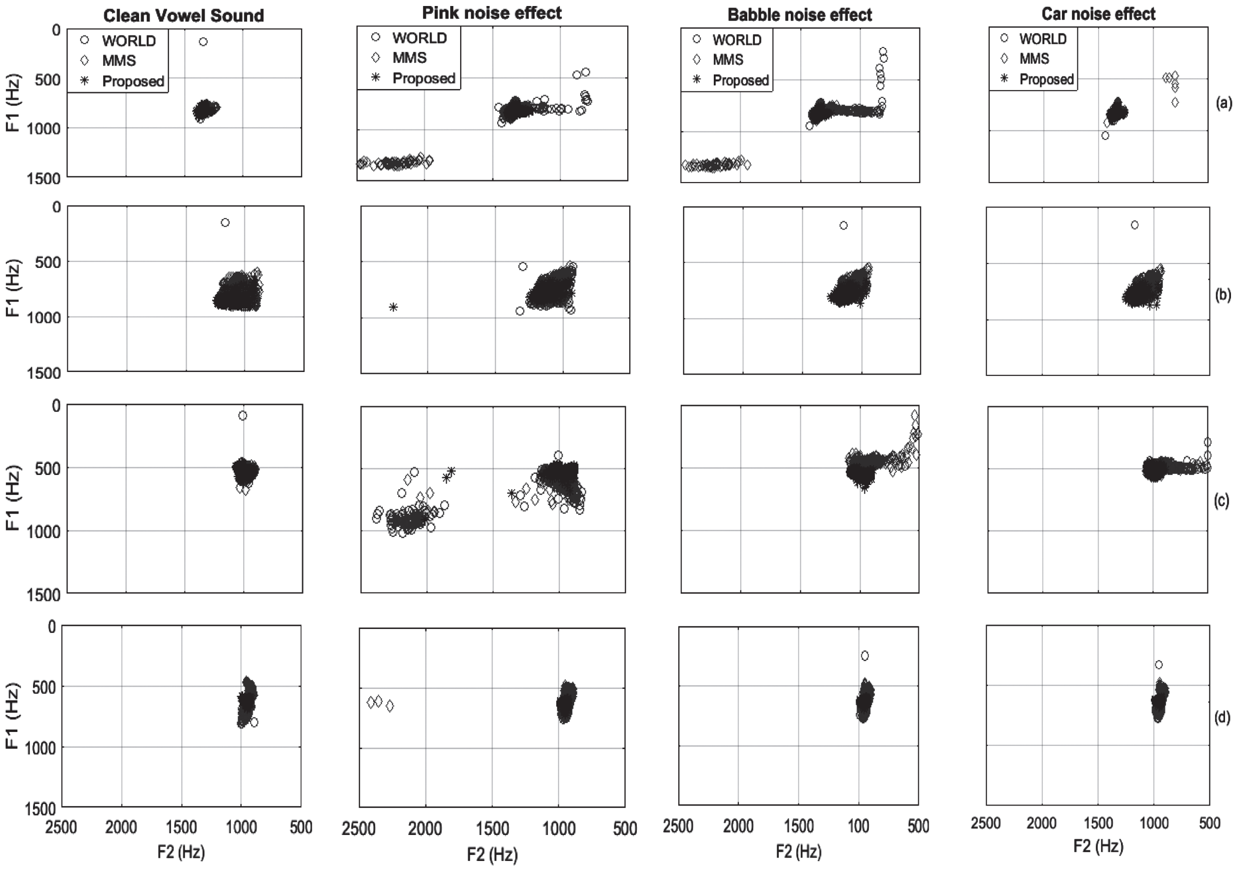


Fig. 8: F2 - F1 plot of the vowel sound /a/ pronounced by (a) a male speaker, (b) a female speaker and the vowel sound /o/ spoken by (c) a male speaker, (d) a female speaker. The first column represents the results in the case of clean speech. The second, third, and fourth columns represent the results of noisy speech added with Pink, Babble, and Car noise, respectively, at SNR = 20 dB.

proposed method outperforms the other two techniques. The noisy synthetic speech analysis result is shown by averaging the (%) errors of the first three formants in Figure 7. Both pitch estimation accuracy and the *log* function used to cepstrally smooth the spectrum influence the WORLD formant estimator. The MMS method utilizes the magnitude spectrum. So, these methods are inherently noise-sensitive.

4.5. Real Speech Analysis

For natural speech analysis, I used some isolated male and female vowels. An utterance of a high-pitched female speaker from the TIMIT database is taken for analysis.

4.5.1. Using Isolated Vowels

The vowels /a/ and /o/ uttered by male and female speakers are used for analysis. All preconditions are fixed to the same as the synthetic speech analysis. The technique DIO [22] of the WORLD vocoder is used to obtain the pitch values, which is employed for low-time-gating. The analysis was conducted on different window positions after shifting by 5 ms. The F2-F1 plot [23] represents the performance of the proposed method. In the case of clean vowel sound, the proposed technique shows the F2-F1 values more concentrated than the MMS and WORLD formant estimator as illustrated in Figure 8 (first column). In the noisy case, the proposed method shows higher accuracy. The MMS and the WORLD formant estimator can't tolerate the noise effect, which is evident in Figure 8 (second, third, and fourth column).

The measurement of standard deviation is shown in Table 2, where bold-marked numbers show the best result in that category. After investigating the measurements of distribution, it is evident that the proposed method shows more concentrated values than other methods both in clean and noisy cases.

Table 2. Standard deviation measurements of different vowels spoken by male and female speakers.

	Methods	Male Vowel /a/		Female Vowel /a/		Male Vowel /o/		Female Vowel /o/	
		F1	**F2**	**F1**	**F2**	**F1**	**F2**	**F1**	**F2**
Clean vowels	Proposed	14	**15**	**38**	76	20	**30**	**16**	12
	MMS	**12**	16	53	73	**19**	**30**	24	**10**
	WORLD	17	**15**	45	**54**	22	**30**	42	14
Noisy (Babble, SNR=20 dB) vowels	Proposed	**15**	**17**	**33**	**46**	18	49	**14**	13
	MMS	96	173	45	51	40	75	23	75
	WORLD	34	90	45	54	**17**	**37**	43	**12**

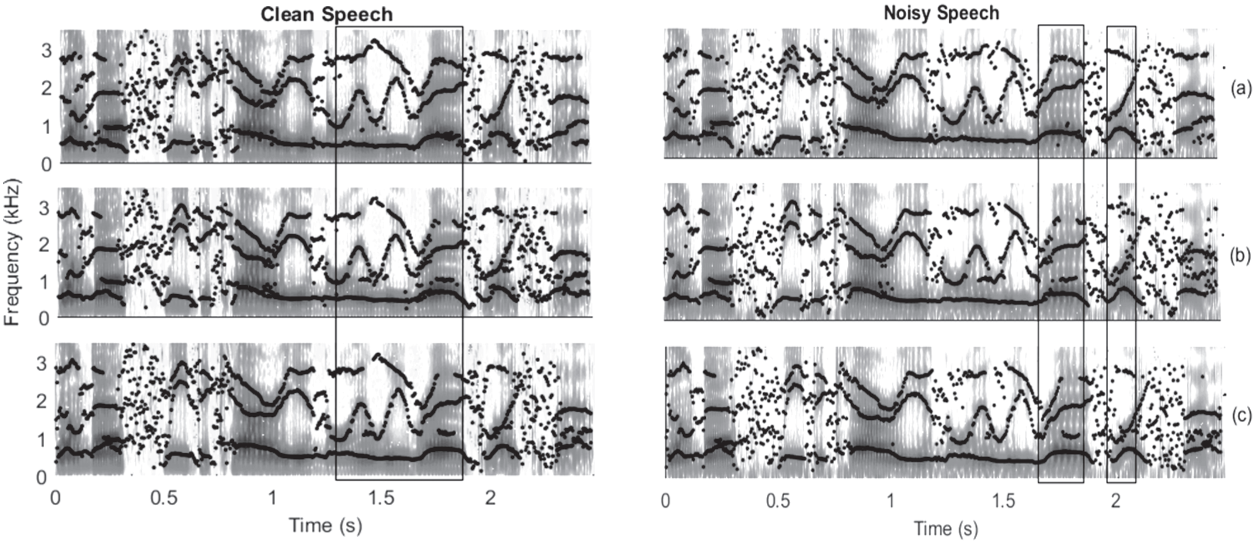


Fig. 9: The spectrogram estimated from a TIMIT utterance of a female speaker, "Don't ask me to carry an oily rag like that". First three formant contours are plotted on the spectrogram using (a) the proposed method, (b) the MMS method, and (c) the WORLD formant estimator.

4.5.2. Using Real Utterance

After estimating the formant values of a clean utterance, three formant contours are plotted on the spectrogram of the utterance as shown in Figure 9 (left column). The formant contours reliably match the selected region of spectrogram plot by the proposed method yielding higher accuracy than the other two methods.

To test the result of noisy utterance (Babble noise at SNR=20 dB), the estimated formant contours are plotted on the spectrogram as shown in Figure 9 (right column). The formant contours are more precise and reliable in the selected regions by the proposed method, whereas a few deviations are manifested in the MMS and the WORLD method. Although some spurious values are observed in the higher formants of all techniques, the proposed method improves the accuracy of formant values in the selected regions.

6. Conclusion

The proposed technique achieves a noise-robust SRC from the speech signal, choosing the appropriate analysis parameters γ and r. To minimize the noise effect further, third-order GDS is calculated from it. The GDM signal estimated from GDS becomes a vocal-tract-dominated signal and overcomes most high-pitched related issues. Experimental analyses also prove the proposed method as a robust formant estimator for high-pitched noisy speech.

Acknowledgment

I acknowledge the ICT division of Bangladesh government for supporting such a research.

References:

1. Atal B. S. and Hanauer S. L. (1971). "Speech analysis and synthesis by linear prediction of the speech wave", J. Acoust. Soc. Am., 50(2B), 637-655. https://doi.org/10.1121/1.1912679.

2. Makhoul J. (1975). "Linear prediction: A tutorial review", Proceedings of the IEEE, 63(4), 561-580.

3. Noll A. M. (1967). "Cepstrum pitch determination", J. Acoust. Soc. Am., 41(2), 293-309. https://doi.org/ 10.1121 / 1.1910339.

4. Oppenheim A. V., R. W. Shafer R. W., and Stockham J. G. (1968). "Nonlinear filtering of multiplied and convolved signals," IEEE trans. audio electroacoustic. 16(3), 437-466.

5. Kaneko T. and Shimamura T. (2014). "Noise-reduced complex LPC analysis for formant estimation of noisy speech", Int. J. Electron. Elect. Eng., 2 (2), 90-94.

6. Gläser C., Heckmann M., Joublin F. and Goerick C. (2008). "Auditory-based formant estimation in noise using a probabilistic framework", Ninth Annu. Conf. Int. Speech Commun. Assoc.

7. Jameel A. S. M. M., Fattah S. A., Goswami R. and Zhu W. and Ahmad M. O. (2016). "Noise robust formant frequency estimation method based on spectral model of repeated autocorrelation of speech", IEEE/ACM Trans. Audio, Speech, Language Process., 25 (6), 1357-1370. https://doi.org/10.1109/TASLP.2016.2625423.

8. Rahman M. S. and Shimamura T. (2007). "Linear prediction using refined autocorrelation function", EURASIP J. Audio Speech Music Process., 1, 1-9. https://doi.org/10.1155/2007/45962.

9. Chowdhury H. A. and Rahman M. S. (2020). "Speech Signal Analysis in Phase Domain." Journal of Computer Science. Aug, 16(8), 1115-1127. https://doi.org/10.3844/jcssp.2020.1115.1127.

10. Murthy H. A. and Yegnanarayana B. (2011). "Group delay functions and its applications in speech technology", Sadhana, 36 (5), 745-782.

11. Mendel J. M. (1991). "Tutorial on higher order statistics (spectra) in signal processing and system theory: Theoretical results and some applications", Proc. IEEE., 79, 278-305.

12. Jae and Lim S. (1979). "Spectral root homomorphic deconvolution system"' IEEE Trans Audio Speech Lang Process., 27 (3), 223-233.

13. Chowdhury H. A. and Rahman M. S. (2021). "Formant -estimation from speech signal using the magnitude spectrum modified with group delay spectrum". Acoust. sci. & tech., 42 (2), 93-102. https://doi.org /10.1250/ast.42.93.

14. Morise M. (2015). "CheapTrick, a spectral envelope estimator for high-quality speech synthesis", Speech Commun., 67, 1-7. https://doi.org/10.1016/j.specom.2014.09.003.

15. Morise M., F. Yokomori and Ozawa, K., (2016). "WORLD: a vocoder-based high-quality speech synthesis system for real-time applications", IEICE Transactions on Information and Systems, 99,1877-1884. https://doi.org/10.1587/transinf.2015EDP7457.

16. Bozkurt B., Couvreur L. (2007). and Dutoit T., "Chirp group delay analysis of speech signals", Speech communication, 49 (3), 159-176. https://doi.org/ 10.1016/j.specom.2006.12.004.

17. Rahman M. S. and Shimamura T. (2005). "Formant frequency estimation of high-pitched speech by homomorphic prediction", Acoust. sci. & tech., 26 (6), 502-510. https://doi.org/10.1250/ast.26.502.

18. Loweimi E. (2018). "Robust phase-based speech signal processing from source filter separation to model based robust ASR.", Ph.D. dissertation, University of Sheffield.

19. Durbin J. (1960). "The fitting of time series models", Rev. Inst. Int. Stat., 233-243.

20. Zue V., S. Seneff and Glass J. (1990). "Speech database development at MIT: TIMIT and beyond", Speech Commun., 9 (4), 351-356.

21. Fant G., Liljencrants J., and Lin Q.-g. (1985). "A four -parameter model of glottal flow", STLQPSR., 4, 1-13.

22. Morise M., Kawahara H., and Katayose H. (2009). "Fast and reliable f0 estimation method based on the period extraction of vocal fold vibration of singing voice and speech", in Proc. AES 35th International Conference, CD-ROM Proceedings.

23. Watt D. and Fabricius A. (2002). "Evaluation of a technique for improving the mapping of multiple speaker's vowel spaces in the F1~ F2 plane", LWPLP., 9(9), 159-173.