Interspeech 2018
2-6 September 2018, Hyderabad

Phase-Locked Loop (PLL) Based Phase Estimation in Single Channel Speech Enhancement

Priya Pallavi[1], *Ch V Rama Rao*[1]

[1]Department of E C E, National Institute of Technology Meghalaya

priyapallavi1993@gmail.com, chvramarao@nitm.ac.in

Abstract

Conventional speech enhancement techniques are based on the modification of noisy spectral magnitude. In the reconstruction of the enhanced signal, noisy phase is combined with the modified noisy spectral magnitude. Recent studies on the importance of phase in enhancement process shows that the clean speech phase improves the quality of the enhanced signal. This work focused on Phase-Locked Loop (PLL) based time-domain approach for estimating the clean speech phase from noisy speech signal. The proposed technique is compared with the conventional approaches where noisy phase is used in the reconstruction of the enhanced signal. Here, Log-Likelihood Ratio (LLR), Weighted Spectral Slope (WSS) distance and Perceptual Evaluation of Speech Quality (PESQ) are used as performance measures. From experimental results, it is observed that the speech quality and intelligibility improved significantly with the proposed method over existing methods.

Index Terms: speech enhancement, phase estimation

1. Introduction

Speech enhancement is the process of improving the quality and intelligibility of speech. Here, quality refers to how the speaker conveys an utterance while intelligibility concerns with what the speaker had said. Speech corrupted with backgroundnoise not only loses its information but also increase the communication barriers.The enhancement of a single-channel recorded speech corrupted with the background noise is one of the most active fields of research. The single-channel speech enhancement methods are getting so much attention because of their versatile applications viz. hearing devices, telephony speech transmission, smart phones and many more.In all these applications, the device performance should be robust for an adverse noise condition. To achieve this we have to find an approach for the noise reduction.

Conventional speech enhancement techniques are mainly focused on how to modify the noisy speech spectral amplitude [1], [2], [3]. Because the spectral amplitude is known to have higher contribution in the perceived quality of speech than spectral phase [4], [5]. But the recent studies on the importance of phase in speech enhancement reveals that the intelligibility of speech is also associated with its phase information [6]. The estimation of phase has a potential to improve the performance of existing magnitude based methods.For better quality of speech, phase information should be taken into account. In [7],the authors discussed about the importance of phase in perceived quality of enhanced speech. They estimate the clean speech phase using relative phase shift representation of phase and used instead of noisy phase for reconstructing the enhanced signal. Pejman Mowlaee et.al [8] proposed the phase estimation method based on fundamental frequency and estimated phase is smoothed based on apriori Signal-to-Noise Ratio (SNR). Zhen Li et.al [9] proposed a method for improving the quality of reconstructed speech based on the modified speech spectral amplitude and the modified compensated phase spectrum.

In this work the concept of PLL is used to estimate the clean speech phase. In general, PLL is used for tracking the phase of received signal in communication systems. Here, the phase information obtained from PLL is combined with the modified spectral amplitude for the reconstruction of signal in order to increase the quality of synthesis speech.

2. Conventional Speech Enhancement

Let y(n) denote the discrete time signal of noisy speech as given in (1)

$$y(n)=x(n)+d(n) \qquad (1)$$

where x(n) and d(n) represents the discrete time signals of clean speech and additive noise respectively.

The spectrum of noisy speech signal is obtained by applying short-time Fourier transform (STFT) to (1). Mathematically the complex spectrum of noisy speech signal Y^c is given by

$$Y^c(k,l)=X^c(k,l)+D^c(k,l) \qquad (2)$$

where X^c and D^c are the complex spectrum of clean speech and noise signals;l and k refers to frame and frequency bin index. In conventional speech enhancement methods [1,2,3,10,11], enhancement has been done by multiplying the noisy speech signal magnitude with gain function G(k,l), which is mathematically represented by

$$X^c(k,l)=G(k,l)|Y^c(k,l)|e^{j\angle Y^c(k,l)} \qquad (3)$$

where $X^c(k,l)$ and $\angle Y^c(k,l)$ represent the modified spectral amplitude and phase of the noisy speech signal respectively. In these methods only magnitude of the noisy speech signal is modified. Some researchers, modeled the gain function G(k,l)based on spectral subtraction [10] and filtering approaches [11]. The enhanced signal is reconstructed by combining the estimated clean speech signal amplitude with noisy speech phase component, $\angle Y^c(k,l)$.

3. PLL based phase estimation

A PLL is a closed loop system that consists of a voltage controlled oscillator (VCO), a phase detector and a low pass filter. The purpose of PLL is to generate a signal which is synchronized in frequency and phase with that of input signal.The phase detector is a multiplier circuit which compares the phase of input signal with respect to the signal which is generated by the VCO and produces an error signal

 10.21437/Interspeech.2018-1950

V_e. The error signal is proportional to the difference of phasesφ_{out} and φ_{in} respectively withK_D as the gain of phase detector. Mathematical representation of error signal is given in (4)

$$V_e = K_D(\varphi_{out} - \varphi_{in}) \tag{4}$$

The phase difference actsas the control voltage to the input of the VCO. In PLL application, VCO is treated as linear, time-invariant system. The output frequency $\omega_{out}$of VCO depends on the input controlled voltage and is given in (5)

$$\omega_{out} = K_O V_{con} \tag{5}$$

whereV_{con} is the control voltage and K_O is the gain coefficient of VCO. The output of the VCO can be taken as frequency and phase depending on the applications. The mathematical relation of phase and frequency is given in (6)

$$\varphi_{out} = \int_0^t \omega_{out} dt \tag{6}$$

The PLL goes through three different states free running, capture and phase-locked. Before the signal is applied, PLL remains in free running state. As, noisy speech signal is given to PLL, it goes to capture mode. Phase detection, minimization of phase error and updating of VCO take place in capture mode. The signal is processed within the loop of PLL until the phase-locked state reaches. The PLL at this phase-locked state signifies that the output signal is synchronized in phase and frequency with that of input signal. This estimation procedure reduces the phase deviation between the input and output signal. The synchronized phase is extracted as the output variable of PLL. In spite of noisy phase which is used in conventional techniques for the reconstruction of signal, synchronized phase of the signal is used.The block diagram of a PLL based speech enhancement system is as shown in Figure 1. Single-channel speech enhancement composed of three stages: amplitude modification, phase estimation and synthesis stage.

In amplitude modification stage, conventional noise reduction schemes are used so as to obtain an estimate for the clean speech spectral amplitude.The estimated phase which is the output of PLL along with estimated speech spectral amplitude is given to the synthesis stage.In synthesis stage, the reconstruction of signal takes place. The reconstructed signal is the enhanced version of the noisy speech signal which is better in quality as well as in intelligibility.

4. Results

This section deals with performance evaluation of proposed method for clean speech phase estimation in speech enhancement process. This work is carried out for improving the performance of the speech enhancement methods by reconstructing the enhanced signal using estimated clean speech phase. The interesting observations from [12] that the conventional noise reduction approaches relying only on the modification of the noisy speech signal amplitude reduce the speech intelligibility. In this evaluation, enhanced signal is obtained by combining noisy phase (NP) and estimated clean speech phase using PLL with existing magnitude based enhancement methods. Here, Spectral Subtraction (SS) [1], Minimum Mean Square Error Short Time Spectral Amplitude Estimator (MMSE-STSA) [2], Non-negative Matrix Factorization (NMF) based speech enhancement [3] are considered as existing methods. Performance of proposed method is also compared with the enhancement through modifying noisy phase by using Phase Spectrum Compensation (PSC) approach [9].

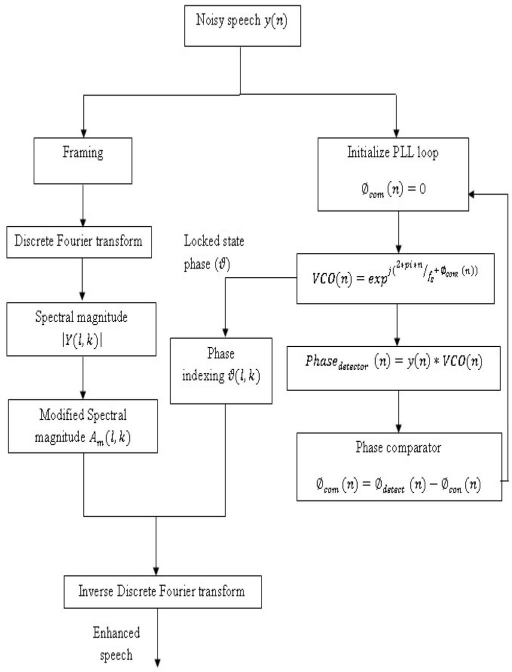

Figure 1: *Block diagram of PLL based speech enhancement*

4.1 Database used

For demonstrating the potential of proposed method, NOIZEUS database [13] is considered. This database contains sentences uttered by 3 male and 3 female speakers. These speech samples are corrupted by eight different real-world noises at different SNR levels.

4.2 Performance measures

The following performance measures are used in the performance evaluation: log-likelihood ratio (LLR), weighted spectral slope distance (WSSD) andperceptual evaluation of speech quality (PESQ) [14].

LLR is distance measure and is calculated from the linear prediction coefficient (LPC) vector of clean speech a_c, noisy speech a_d andauto-correlation matrix of the clean speech R_c.The low value of LLR shows the better quality of speech. Mathematical representation is given in (7)

$$d_{LLR}(a_d, a_c) = \log\left(\frac{a_d R_c a_d^T}{a_c R_c a_c^T}\right) \tag{7}$$

WSSDcomputes the weighted spectral slopes of clean signal and noisy signal in each frequency band. Low value of WSS is needed for the synthesis of better quality of speech.Mathematically, it is given in (8)

$$d_{wss} = \frac{1}{M}\sum_{m=0}^{M-1}\frac{\sum_{j=1}^{K} W_{WSS}(j,m)(S_c(j,m) - S_p(j,m))^2}{\sum_{j=1}^{k} W_{WSS}(j,m)} \tag{8}$$

where, $W_{WSS}(j,m)$ are the computed weights, M is the number of segments, $S_c(j,m)$ and $S_d(j,m)$ are the spectral slope of clean signal and noisy signal for the j^{th} frequency band.

PESQ is an objective method used for estimating the quality of speech. PESQ predicts subjective Mean Opinion Score (MOS) by comparing the degraded signal with the original version of the signal. Higher the value of PESQ indicates better quality of speech.

4.3 Performance evaluation

For evaluating the performance babble noise, restaurant, car and train noise environments are considered at 0 and 5 dB SNR levels. Obtained LLR values for considered enhanced methods are tabulated in Table 1. From Table 1, it is observed that lower values of LLR are obtained with estimated clean speech phase compared to noisy phase. Table 2 illustrates obtained WSSD values. Lower WSSD values are obtained with estimated clean speech phase over noisy phase. That means spectral slope deviation is reduced with estimated clean speech phase. Here, performance is also evaluated in terms of PESQ and obtained values are given in Table 3. From Table 3, higher PESQ values are obtained with estimated phase. In this work, performance of speech enhancement method by using estimated phase with PLL is compared with PSC based approach in terms of PESQ values. Table 4 illustrates obtained PESQ values for enhanced signal using estimated phase with PLL and also with PSC. Higher PESQ values are obtained with PLL based clean speech phase estimation. From experimental results, it is clear that PLL based estimation of clean speech phase improves performance of speech enhancement methods.

Table 1: *LLR values*

Noise type and SNR value		SS		MMSE-STSA		NMF based	
		NP	PLL	NP	PLL	NP	PLL
Babble	0	1.99	**1.84**	1.67	**1.47**	1.86	**1.26**
	5	1.77	**1.63**	1.45	**1.23**	1.40	**1.28**
Restaurant	0	1.87	**1.70**	1.78	**1.54**	1.74	**1.13**
	5	1.42	**1.38**	1.53	**1.42**	1.50	**1.42**
Car	0	1.85	**1.76**	1.75	**1.60**	1.74	**1.23**
	5	1.53	**1.34**	1.32	**1.26**	1.76	**1.63**
Train	0	1.90	**1.72**	1.77	**1.63**	1.75	**1.31**
	5	1.77	**1.56**	1.48	**1.30**	1.14	**1.06**

5. Conclusions

This work focused on improving the performance of single channel speech enhancement methods by estimating the clean speech phase. In conventional methods, speech enhancement is performed by modifying the magnitude of the noisy speech signal and enhanced signal is reconstructed using noisy speech phase. In this work clean speech phase is estimated based on PLL concept. Here, performance is evaluated by comparing the speech enhancement methods with noisy speech phase and estimated phase using PLL. From experimental results, it is observed that performance is improved with estimated phase.

Table 2: *WSSD values*

Noise type and SNR value		SS		MMSE-STSA		NMF based	
		NP	PLL	NP	PLL	NP	PLL
Babble	0	128	**113**	143	**106**	110	**103**
	5	119	**105**	105	**91**	128	**96**
Restaurant	0	132	**112**	147	**116**	106	**97**
	5	108	**102**	100	**92**	130	**103**
Car	0	124	**110**	117	**101**	110	**98**
	5	103	**98**	100	**88**	98	**89**
Train	0	110	**94**	124	**97**	110	**107**
	5	119	**89**	95	**93**	74	**69**

Table 3: *PESQ values*

Noise type and SNR value		SS		MMSE-STSA		NMF based	
		NP	PLL	NP	PLL	NP	PLL
Babble	0	1.48	**1.53**	1.37	**1.42**	1.53	**1.72**
	5	1.65	**1.79**	1.61	**1.73**	1.37	**1.92**
Restaurant	0	1.28	**1.35**	1.17	**1.29**	1.33	**1.66**
	5	1.63	**1.71**	1.68	**1.92**	1.70	**1.97**
Car	0	1.27	**1.35**	1.21	**1.42**	1.42	**1.58**
	5	1.52	**1.75**	1.58	**1.62**	1.50	**1.70**
Train	0	1.33	**1.49**	0.96	**1.21**	1.46	**1.52**
	5	1.68	**1.73**	1.84	**2.04**	1.73	**1.95**

Table 4: Comparison of obtained *PESQ values for PLL and PSC*

Noise type and SNR value		SS		MMSE-STSA		NMF based	
		PSC	PLL	PSC	PLL	PSC	PLL
Babble	0	1.45	**1.53**	1.31	**1.42**	1.68	**1.72**
	5	1.70	**1.79**	1.62	**1.73**	1.96	**1.92**
Restaurant	0	1.21	**1.35**	1.21	**1.29**	1.53	**1.66**
	5	1.66	**1.71**	1.71	**1.92**	1.88	**1.97**
Car	0	1.28	**1.35**	1.27	**1.42**	1.44	**1.58**
	5	1.64	**1.75**	1.68	**1.62**	1.73	**1.70**
Train	0	1.38	**1.49**	1.19	**1.21**	1.43	**1.52**
	5	1.65	**1.73**	1.92	**2.04**	1.85	**1.95**

6. References

[1] S. Boll, "A spectral subtraction algorithm for suppression of acoustic noise in speech," in *IEEE International Conference on Acoustics, Speech, and Signal Processing*, pp.200-203,1979.

[2] Y. Ephraim and D. Malah, "Speech enhancement using a minimum-mean square error short-time spectral amplitude estimator," in *IEEE Transactions on Acoustics, Speech, and Signal Processing*, vol. 32, no. 6, pp. 1109-1121, Dec 1984.

[3] V. Anil Kumar, Ch. V. Rama Rao, “Unsupervised noise removal technique based on constrained NMF,” in *IET Signal Processing*,vol. 11, no. 7, pp. 788-795, 2017.

[4] D. Wang and Jae Lim, "The unimportance of phase in speech enhancement," in *IEEE Transactions on Acoustics, Speech, and Signal Processing*, vol. 30, no. 4, pp. 679-681, Aug 1982.

[5] A. V. Oppenheim and J. S. Lim, "The importance of phase in signals," in *Proceedings of the IEEE*, vol. 69, no. 5, pp. 529-541, May 1981.

[6] Kuldip Paliwal, Kamil Wojcicki and Benjamin Shannon, "The importance of phase in speech enhancement," in *Speech Communication*, vol. 53, no 4, pp. 465-494, 2011.

[7] I. Saratxaga, I. Hernaez, M. Pucher and I. Sainz, "Perceptual importance of the phase related information in speech," in *Thirteenth Annual Conference of the International Speech Communication Association,* 2012.

[8] P. Mowlaee and J. Kulmer, "Harmonic Phase Estimation in Single-Channel Speech Enhancement Using Phase Decomposition and SNR Information," in *IEEE/ACM Transactions on Audio, Speech, and Language Processing*, vol. 23, no. 9, pp. 1521-1532, Sept. 2015.

[9] Z. Li, W. Wu, Q. Zhang, H. Ren and S. Bai, "Speech enhancement using magnitude and phase spectrum compensation," in *15th International Conference on Computer and Information Science (ICIS),* pp. 1-4, 2016.

[10] H. Gustafsson, S. E. Nordholm and I. Claesson, "Spectral subtraction using reduced delay convolution and adaptive averaging," in *IEEE Transactions on Speech and Audio Processing*, vol. 9, no. 8, pp. 799-807, Nov 2001.

[11] P. Scalart and J. V. Filho, "Speech enhancement based on a priori signal to noise estimation," in *IEEE International Conference on Acoustics, Speech, and Signal Processing Conference Proceedings*, vol. 2, pp. 629-632, 1996.

[12] P. C. Loizou and K. Gibak, “Reasons why current speech-enhancement algorithms do not improve speech intelligibility and suggested solutions,” in *IEEE Trans. Audio, Speech, Lang. Process.,* vol. 19, no. 1, pp. 47- 56, Jan. 2011.

[13] Yi Hu and P. C. Loizou, "Subjective Comparison of Speech Enhancement Algorithms," *2006 IEEE International Conference on Acoustics Speech and Signal Processing Proceedings*, pp. 1, 2006.

[14] Y. Hu and P. C. Loizou, "Evaluation of Objective Quality Measures for Speech Enhancement," in *IEEE Transactions on Audio, Speech, and Language Processing*, vol. 16, no. 1, pp. 229-238, Jan. 2008.