EURASIP Journal on Advances in Signal Processing
a SpringerOpen Journal

RESEARCH Open Access

Transient noise reduction in speech signal with a modified long-term predictor

Min-Seok Choi* and Hong-Goo Kang

Abstract

This article proposes an efficient median filter based algorithm to remove transient noise in a speech signal. The proposed algorithm adopts a modified long-term predictor (LTP) as the pre-processor of the noise reduction process to reduce speech distortion caused by the nonlinear nature of the median filter. This article shows that the LTP analysis does not modify to the characteristic of transient noise during the speech modeling process. Oppositely, if a short-term linear prediction (STP) filter is employed as a pre-processor, the enhanced output includes residual noise because the STP analysis and synthesis process keeps and restores transient noise components. To minimize residual noise and speech distortion after the transient noise reduction, a modified LTP method is proposed which estimates the characteristic of speech more accurately. By ignoring transient noise presence regions in the pitch lag detection step, the modified LTP successfully avoids being affected by transient noise. A backward pitch prediction algorithm is also adopted to reduce speech distortion in the onset regions. Experimental results verify that the proposed system efficiently eliminates transient noise while preserving desired speech signal.

Keywords: speech enhancement, transient noise reduction, long-term prediction, median filter

1 Introduction

Reducing noise from noise-corrupted speech is essential for communication or recording devices. Spectral subtractive noise reduction algorithms have been widely developed under the assumption that input noise is stationary or slowly varying [1-3]. Therefore, the linear filtering methods cannot remove transient noise easily which has abruptly varying characteristic [4-6]. In general, transient noise is generated by tapping a recording device or an object near it. Since transient noise randomly occurs in time and has a time-varying unknown impulse response, the characteristic of the noise is not easy to estimate. In other words, both the occurrence time and the impulse response of transient noise are unpredictable. The good thing is that transient noise usually is a fast varying signal with short duration and high amplitude thus its activity is relatively easy to detect [4-8].

Transient noise can be removed by utilizing a nonlinear filter such as a median filter or a power limiter [4-7,9]. The nonlinear power limiter suppresses input segments which have enormous magnitude compared to a pre-assigned value. Since it only cuts down the high amplitude portion of transient noise, some noise component still remains in the output. Moreover, if transient noise is added to speech, determining the amount of the signal power reduction is difficult because the level of the speech waveform varies rapidly. Consequently, the power limiter is not efficient to eliminate transient noise in speech [5,7,9]. A median filter is a signal dependent filter which removes the fast varying components while preserving slowly varying components of the input signal [4,6,7,10]. The median filter does not require any pre-defined threshold during the filtering process. Since the median filter only preserves the slowly varying components of input signal, however, it may distort the characteristic of fast varying region of speech, i.e., around pitch epoch. Therefore, an additional pre-processing step to keep the speech characteristic before applying the median filter is needed. For example, a short-term linear prediction (STP) filter and a long-term prediction (LTP) filter which are parametric approaches to model speech signal can be utilized as a pre-

* Correspondence: zzugie@gmail.com
School of Electrical and Electronic, Yonsei University, 134 Shinchon-dong, Seodaemun-gu, Seoul 120-749, Korea

processor [11]. The purpose of the pre-processor is passing transient noise components but keeping speech information by utilizing the speech modeling filter not to be affected by the median filtering afterwards.

Typical speech modeling methods such as STP and LTP are good candidates for the pre-processing module. The STP filter represents the short-term characteristic of speech, and the LTP filter does the long-term periodic components. If the STP or the LTP filter extracts all speech components from input and leaves all transient noise components in the residual signal, the median filter may be successfully applied to remove the transient noise at the residual signal. It has been reported that applying both STP and LTP to speech is effective to represent the characteristic of the speech [10-12].

After removing transient noise from the residual signal, the speech component extracted by the STP filter or the LTP filter should be re-synthesized. Please note that the pre-filter should not keep the characteristic of transient noise not to bring any residual noise. In general, transient noise lasts for the certain amount of time, e.g., up to 50 ms, and has short-term correlation. Therefore, the STP filter which models the short-term characteristic of signal is not appropriate for our purpose. On the contrary, transient noise component which generally has short duration would not affect an LTP result [7,8,10,11,13].

Figure 1 depicts residual signals after the STP analysis and the LTP analysis. The input signal of the analysis contains both speech and transient noise to show the influence of the speech modeling filters. Figure 1a represents a transient noise segment which is added to speech signal. Figure 1b,c are residual signals after performing the STP and the LTP analysis, respectively. Note that the residual signal in Figure 1c is not processed by the STP filter but only processed by the LTP analysis filter. As shown in Figure 1b, the STP analysis removes the transient noise component. It indicates that the STP filter somewhat models the characteristic of the transient noise. However, the residual signal after the LTP analysis, Figure 1c, is almost same as the input transient noise, which indicates that the LTP filter does not keep the transient noise component. Consequently, applying the median filter to the LTP residual should be quite effective to remove the transient noise. Table 1 represents the normalized cross-correlation (NCC) between the input transient noise and the residual signal after the STP or the LTP analysis [14]. The NCC results also verify the efficiency of the LTP filter as the speech preserving pre-processor of the transient noise reduction system[1] [10].

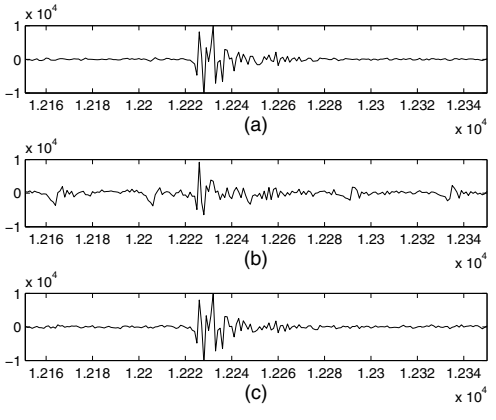

Figure 1 Residual signal after applying speech modeling filter to noisy speech. Time-domain waveforms of **(a)**: Noise signal, **(b)**: Residual signal after STP analysis, and **(c)**: Residual signal after LTP analysis.

The LTP filter generally searches the most similar signal segment to the current signal segment within a predefined search range [11,12]. If transient noise component exists in the search range, however, a transient noise segment in the current frame can be predicted by the other transient noise in the search range. In such case, the LTP filter models the characteristic of the transient noise and brings residual noise in synthesized speech. Another problem of the conventional LTP method is that the LTP filter cannot preserve pitch information at the onset and the transition region of speech because a reference pitch does not exists. As a result, the conventional LTP method needs to be modified to accurately model the pitch related speech component without being affected by transient noise. To solve the first problem on having transient noise component within a pitch search interval, we need to skip the transient noise region while searching a reference pitch. However, skipping the transient noise region occasionally results in failure in the pitch prediction when the transient noise is located where the reference pitch exists. Therefore, we extend the pitch search range to cover multiple pitch periods. The pitch estimation problem at the onset and the transition region of speech can be solved by adopting a look-ahead memory and a backward pitch estimation method. The modified LTP significantly reduces the residual noise in an enhanced signal and successfully reconstructs desired speech after the transient noise reduction.

Table 1 NCC between transient noise and residual signals.

	Residual after STP analysis	Residual after LTP analysis
NCC	0.8267	0.9908

The NCCs between transient noise and residual signals after speech modeling process, e.g., STP and LTP analysis.

The rest of this article is organized as follows. In the following section, the median filter for removing transient noise is briefly described. The conventional LTP method which is generally used for speech coding is given in Section 3. The transient noise reduction system with the modified LTP method is proposed in Section 4. Experimental results and conclusions are followed in Sections 5 and 6, respectively.

2 Median filtering for transient noise reduction

We assume that an input signal, $x(n)$, is the summation of a clean speech signal, $s(n)$, and a transient noise signal, $d(n)$, such as:

$$x(n) = s(n) + d(n). \tag{1}$$

The transient noise randomly occurs in time and has a time-varying unknown impulse response and variance [7].

$$d(n) = \sum_k (h_k(n) * \delta(n - T_k))g_k(n), \tag{2}$$

where T_k defines the occurrence time of the kth transient noise. $h_k(n)$ and $g_k(n)$ denote the impulse response and the amplitude of the kth transient noise, respectively. Note that T_k, $h_k(n)$, and $g_k(n)$ are unpredictable in general.

A relatively easy way to remove transient noise is to apply a time-domain median filter or a nonlinear power limiter to transient noise presence region [4-6,9]. This article adopts the median filter because it efficiently removes transient noise while preserving the slowly varying component in the input signal. In other words, the slowly varying component of desired speech remains in the output of the median filter. Moreover, the median filter is easy to implement because it does not need any pre-defined threshold. Though the median filter is effective for eliminating transient noise, however, it may also distort the characteristic of desired speech while removing the fast varying component. Therefore, the filter should be applied only to transient noise presence region to minimize the speech distortion problem.

$$y(n) = \begin{cases} x(n), & H_T(n) = 0 \\ \text{med}_w[x(n)], & H_T(n) = 1, \end{cases} \tag{3}$$

where $\text{med}_w[x(n)]$ defines the median filtering operator of which output is the median value of input samples from $x(n - w)$ to $x(n + w)$. The length of the median filter, $2w + 1$, should be long enough to cover the length of transient noise [4]. $H_T(n)$ in Eq. (3) denotes the detection flag of transient noise presence which becomes one when the noise exists and vice versa. It can be determined by comparing the time-domain energy, the frequency-domain energy, or the cross-correlation of input signal [4,6,15,16]. For example, a time-frequency domain transient noise detector proposed in [16] shows 99.3% of detection accuracy while making only 1.49% of false-alarm. Employing the transient noise detection result, the median filter can be applied only to the noise presence region. However, the speech distortion still exists in the region where the median filtering is performed.

3 Conventional long-term predictor

The nonlinear waveform suppression filter, e.g., the median filter, not only reduces noise but also distorts speech. Especially, the fast varying component in speech such as pitch epoch are notably removed during the median filtering. Therefore, an additional step is needed to preserve the pitch component before removing the noise.

The LTP is a method for representing the current pitch component of speech by scaling a speech segment at one pitch period before. It efficiently estimates periodic and stationary component in the signal [10-12].

$$\begin{aligned} &\tilde{x}(m,\ l) = g_p(l)x(m - \tau_p(l),\ l) \\ &0 \le m \le M - 1, \end{aligned} \tag{4}$$

where l and M denote the frame index and the length of the frame, respectively. The index (m, l) represents the mth sample in the lth frame such as $(m + (l - 1)M)$. The optimum time lag, $\tau_p(l)$, which denotes the pitch interval at the current frame is a value that maximizes the cross-correlation of the input such as:

$$\tau_p(l) = \underset{\tau_{\min} \le \tau \le \tau_{\max}}{\arg\max} \frac{\sum_{m=0}^{M-1} x(m,\ l)x(m - \tau,\ l)}{\sqrt{\sum_{m=0}^{M-1} x^2(m - \tau, l)}}, \tag{5}$$

where the range of τ is determined by considering the general pitch period of human's speech, e.g., 2.5 ms $\le \tau \le$ 18 ms. Since $\tau_p(l)$ in Eq. (5) is the integer multiple of the sampling period of the input signal, the estimation error of the pitch period depends on the sampling frequency. Therefore, interpolating the cross-correlation and finding a fractional pitch period is helpful to improve the LTP accuracy [12]. The gain, $g_p(l)$, to minimize the signal modeling error is defined as:

$$\hat{g}_p(l) = \frac{\sum_{m=0}^{M-1} x(m,\ l)x(m - \tau_p(l),\ l)}{\sqrt{\sum_{m=0}^{M-1} x^2(m - \tau_p(l), l)}}. \tag{6}$$

However, the LTP gain is generally limited to a certain constant to avoid the over-estimation of the pitch.

$$g_p(l) = \begin{cases} \hat{g}_p(l), & \hat{g}_p(l) < g_{p\max} \\ g_{p\max}, & \text{otherwise.} \end{cases} \quad (7)$$

We restrict the gain to 1.2 in the proposed system [12]. Utilizing the estimated pitch lag and gain, the LTP analysis filter extracts the pitch component from the input speech.

$$r(m,\ l) = x(m,\ l) - \tilde{x}(m,\ l), \quad (8)$$

where $r(m, l)$ denotes the residual signal after the LTP analysis. To synthesize the desired speech from the residual signal, the pitch period, the gain, and the previously synthesized speech segment are needed. Assuming that they are exactly known, the synthesizing process becomes:

$$y(m,\ l) = r(m,\ l) + g_p(l)y(m - \tau_p(l), l). \quad (9)$$

Note that the synthesis process is an iterative method thus the quality of the currently synthesized speech segment depends on the quality of the previous pitch. In other words, the pitch synthesis error at the previous frame can be propagated to the next frame [12].

4 Proposed algorithm

The proposed algorithm employs the LTP as a pre-processor of the median filter, but note that the STP filter which is usually used in speech analysis systems is not utilized because the STP filter may model not only speech component but also the characteristic of transient noise. As a result, applying the STP filter results in the residual noise to the re-synthesized speech after the noise reduction [7,8,10].

The conventional LTP method predicts a speech segment by utilizing a previous speech segment at one pitch period before [10-12]. Unlike the STP filter, the LTP filter is not affected by the short-term characteristic of transient noise. However, the LTP filter also models transient noise component if the transient noise exists within the search range of the pitch lag. One way of reducing the problem is to skip the transient noise region while searching the pitch lag. Note also that, the conventional LTP method cannot estimate pitch at the onset or the transition region of vowel because the reference pitch segment does not exists. The proposed method utilizes look-ahead samples to predict the current speech segment more accurately thus it becomes more appropriate for preserving the speech component in transient noise environment.

In this section, we firstly propose the transient noise reduction system based on the median filter which utilizes the LTP as a pre-processor. The proposed system adopts a non-predictive speech synthesis method thus the error caused by the median filter is not propagated to future speech samples. In Section 4.2, the modified LTP method is proposed to efficiently estimate speech component while not being affected by transient noise.

4.1 Median filter by utilizing the LTP with non-predictive pitch synthesis

If transient noise does not exist, the noise reduction process is not necessary. Therefore, we perform the median filtering depending on the activity of transient noise.

$$y(m,\ l) = \begin{cases} x(m,\ l)\ H_T(m,\ l) = 0 \\ \hat{y}(m,\ l)\ H_T(m,\ l) = 1, \end{cases} \quad (10)$$

where $\hat{y}(m,\ l)$ represents the synthesized speech after the median filtering. In the proposed system, the median filter is applied to the residual signal after the LTP analysis given in Eq. (8).

$$\hat{r}(m,\ l) = \text{med}_w[r(m,\ l)], \quad (11)$$

where $\hat{r}(m,\ l)$ defines the output of the median filter. The speech can be restored by re-synthesizing the pitch to the output of the median filter.

$$\hat{y}(m,\ l) = \hat{r}(m,\ l) + \tilde{x}(m,\ l). \quad (12)$$

Note that we directly use $\tilde{x}(m,\ l)$ which is estimated during the LTP analysis for the speech synthesis. The predictive synthesis method in Eq. (9) is very efficient in the speech compression aspect because it requires a little information for restoring speech. However, it propagates the prediction error in the past to the currently synthesizing segment, which degrades speech quality [12]. In the proposed method, the non-predictive synthesis method given in Eq. (12) is introduced to prevent from propagating the error caused by the median filter. Figure 2 shows the block diagram of the proposed transient noise reduction system [10].

4.2 Non-causal pitch estimation without being affected by transient noise

In the pitch lag estimation algorithm given in Eq. (5), the search range to estimate the optimum pitch period needs to be pre-defined. As we already mentioned in Section 3, it is generally determined by considering the characteristic of the human's voice. However, transient noise can be modeled by the LTP if some of the transient noise component exists within the search range. In the proposed system, we discard the transient noise presence region during the pitch lag estimation step.

$$\tau_p(l) = \underset{\substack{\tau_{\min}\le\tau\le\tau_{\max} \\ \sum_{m=0}^{M-1} H_T(m-\tau,\ l) = 0}}{\arg\max} \frac{\sum_{m=0}^{M-1} x(m,\ l)x(m-\tau,\ l)}{\sqrt{\sum_{m=0}^{M-1} x^2(m-\tau, l)}}. \quad (13)$$

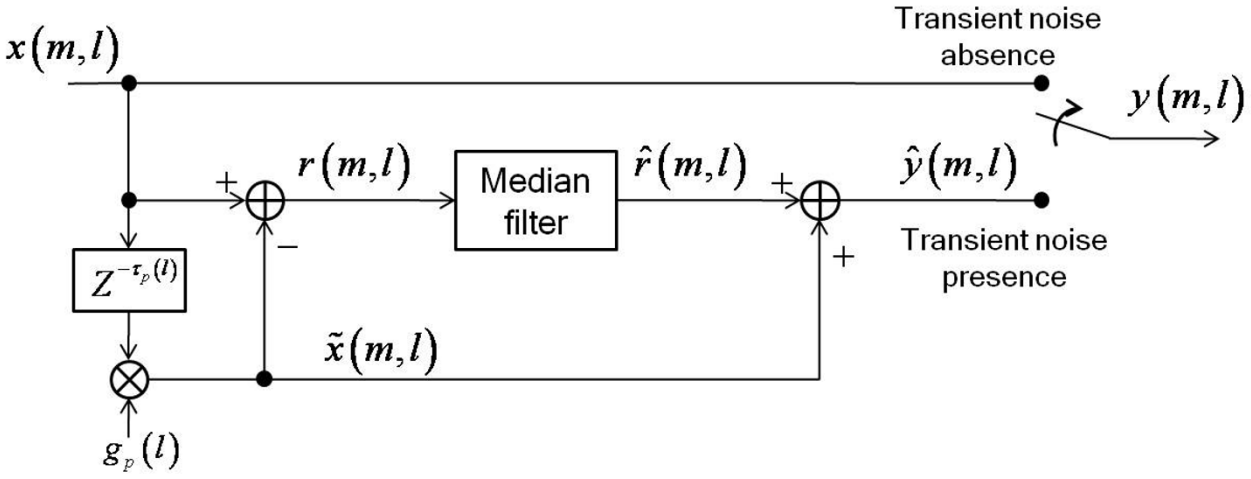

Figure 2 A block diagram of proposed transient noise reduction system. A median filtering system after the LTP analysis. The transient noise reduction process is applied only in noise presence region.

If the sum of H_T (m - τ, l) with any τ where $0 \le m \le M$ - 1 is bigger than zero, the system skips the τ while searching the pitch period because some of $x(m - \tau, l)$ with the τ may contain transient noise component. The method in Eq. (13) is helpful for reducing the residual noise in the synthesized speech because the LTP employing the pitch lag detector in Eq. (13) does not preserve transient noise even when the transient noise exists in the search range of the pitch lag.

However, if we adopt the method in Eq. (13), the pitch of the current frame cannot be estimated when transient noise exists at the location of the previous pitch. To save the pitch more efficiently, we need to expand the pitch search range so that the range contains multiple candidate pitches. Note that we do not need to find an exact pitch period, but we should find the most similar pitch to the current pitch. If the previous pitch is contaminated by transient noise, pitch epoch that is located at farther from the current frame can be an alternative candidate of the current pitch. In the proposed system, we set $\tau_{\min}$ and $\tau_{\max}$ to about 2.5 ms and 36 ms, respectively. It is twice as wide as the range of usual pitch searching range, which includes at least two pitches [11,12].

Figure 3 depicts the output waveforms of the noise reduction system which utilize the conventional pitch lag estimation algorithm and the modified method given in Eq. (13). Figure 3a,b represent the desired speech and the input signal, respectively. Figure 3c is the enhanced output adopting the conventional LTP method, and Figure 3d is the output with the modified pitch lag detection algorithm. As shown at the shaded region in Figure 3c, the conventional pitch lag estimator results in much higher residual noise in the noise reduction result because the LTP filter keeps and re-synthesizes transient noise component. When we utilize the modified pitch lag estimator in Eq. (13), the amount of the residual noise is reduced as depicted in Figure 3d.

The LTP cannot model the pitch at the onset and the transition region of vowel because the reference pitch does not exist in previous samples. If we allow to

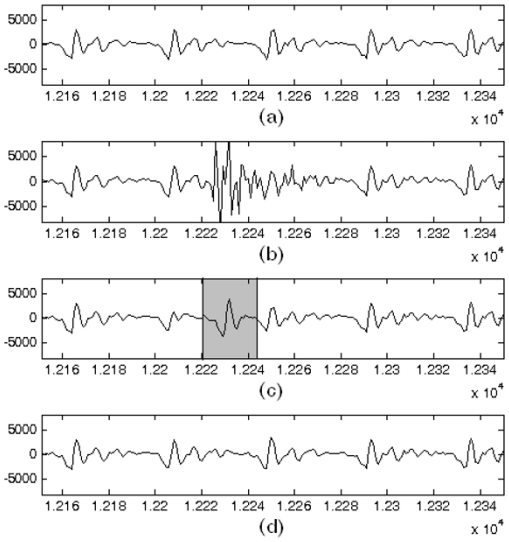

Figure 3 Results of transient noise reduction. Time-domain waveforms of **(a)**: Clean speech, **(b)**: Noise corrupted speech, **(c)**: Output signal utilizing the conventional LTP method in Eq. (5), and **(d)**: Output signal utilizing the modified LTP method in Eq. (13) which discards the transient noise presence region during the pitch prediction.

estimate the current pitch by utilizing the pitch in the future, the pitch at the onset also can be preserved and restored. Consequently, the pitch lag estimator in the proposed system is designed as follow:

$$\tau_p(l) = \underset{\substack{\tau_{\min}\le|\tau|\le\tau_{\max} \\ \sum_{m=0}^{M-1} H_T(m-\tau,l)=0}}{\arg\max} \frac{\sum_{m=0}^{M-1} x(m,l)x(m-\tau,l)}{\sqrt{\sum_{m=0}^{M-1} x^2(m-\tau,l)}}. \quad (14)$$

The proposed method detects the pitch lag which is the best estimation of the current pitch among previous samples, $\tau_{\min} \le \tau \le \tau_{\max}$, and future samples, $-\tau_{\max} \le \tau \le -\tau_{\min}$, while skipping samples that include transient noise component. Referring the future pitch for the pitch estimation improves the capability of preserving speech information, However, the system delay increases somehow due to the look-ahead memory.

A method to find a fractional pitch lag can be also applied to Eq. (14), which may further improve the pitch estimation accuracy. The optimum pitch gain for the estimated pitch lag is calculated by using Eqs. (6) and (7). Finally, we can extract the pitch component from input speech, and generate a residual signal, $r(m, l)$. The results of the transient noise reduction utilizing the causal and the non-causal LTP filters are depicted in Figure 4. Figure 4a-c represent the desired speech, the output signal utilizing the causal LTP filter, and the output utilizing the non-causal LTP filter, respectively. The result with the non-causal LTP can recover the speech at the onset of vowel after the median filtering. When we use the causal LTP filter, it cannot model the pitch at the onset of vowel thus the pitch epoch remains in the residual signal. Therefore, the pitch at the onset is removed during the noise reduction process such as shaded region in Figure 4b.

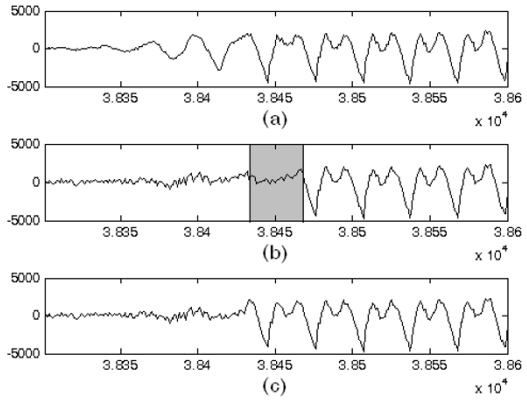

Figure 4 Results of transient noise reduction utilizing the causal and non-causal LTP methods. Time-domain waveforms of **(a)**: Clean speech, **(b)**: Output signal utilizing the causal LTP method in Eq. (13), and **(c)**: Output signal utilizing the non-causal LTP method in Eq. (14).

5 Performance evaluation

To evaluate the performance of the proposed system, we apply it to recorded speech signals which contain transient noise. Every speech signals and transient noise signals are recorded in real environment, separately. The transient noise signals are acquired by using mobile recoding devices while clicking buttons on the recording devices or tapping the body of the recording devices. We add the transient noise segments to the random points of time of the speech signals. More than one hundred transient noise sequences are added to eight sentences of speech signals. Speech database is recorded by four male and four female speakers, and the total length of the speech signals is about sixteen seconds. The sampling frequency of the speech is 8 kHz. Since the transient noise is recorded in real environment, additive background noise such as fan noise is also included in the recoded noise signal. In other words, the test signals contain clean speech, transient noise, and background noise. The signal-to-noise ratio (SNR) between the desired speech and the background noise is around 15 dB.

The median filter and the LTP filter are applied only at transient noise presence region by utilizing the hand-marked result of the noise presence. However, the transient noise presence region can be detected by measuring the time- or the frequency-domain energy of the input signal with a certain threshold [4,15,16]. Experimental results utilizing the transient noise detector proposed in [16] are almost same as results with the hand-marked noise detection result shown in this article. The length of the median filter, $2w + 1$, used for the experiments is 101 samples, and the frame size for the LTP, M, is 32 samples. The minimum and the maximum bounds of the pitch lag search range, $\tau_{\min}$, $\tau_{\max}$, is 20 and 143 samples for the conventional pitch lag detection in Eq. (5), and the maximum bound is doubled to 286 samples for the modified pitch lag detectors in Eqs. (13) and (14). The maximum bound of the pitch gain, $g_{p\ \max}$, is set to 1.2. The interpolation of the cross-correlation for the pitch lag detection is performed to find a fractional pitch period. As a result, the resolution of the pitch lag, $\tau_p(l)$, is the triple of the sampling frequency [12]. Note that the LTP performance can be degraded by background noise. Therefore, an optimally modified minimum mean-square error log-spectral amplitude (OM-LSA) estimator with an improved minima controlled recursive averaging (IMCRA) noise estimator is applied to remove background noise before the transient noise reduction process [17-19]. Since the OM-LSA

estimator and the IMCRA noise estimator are designed to remove only stationary noise, they do not affect the transient noise.

To evaluate the performance of the transient noise reduction systems, we measure SNR, segmental signal-to-noise ratio (SSNR), and log-spectral distance (LSD) between output signals and a clean speech such as [20]:

$$SNR = 10\log_{10}\left(\frac{E_{m,l}\{s(m,l)^2\}}{E_{m,l}\{(s(m,l)-\gamma(m,l))^2\}}\right)$$
$$SSNR = E_l\left\{10\log_{10}\left(\frac{E_m\{s(m,l)^2\}}{E_m\{(s(m,l)-\gamma(m,l))^2\}}\right)\right\} \quad (15)$$
$$LSD = E_l\left\{\sqrt{E_f\left\{\left(20\log_{10}\frac{|S(f,l)|}{|Y(f,l)|}\right)^2\right\}}\right\},$$

where $E_{m,l}$, E_m, and E_l define the mean of whole samples, a frame, and all frames, respectively. Similarly, E_f represents the mean of frequency bins in a frame. $S(f, l)$ and $Y(f, l)$ denote the frequency responses of desired speech and system output, respectively.

Tables 2 and 3 show the evaluation results of the proposed systems. Note that we measure the objective scores only when transient noise exists. The results in Table 2 are measured without regard for speech presence, and the results in Table 3 are measured only in speech presence region. To prove the efficiency of the proposed system, the output signals of the median filter employing various pre-processing techniques are tested. The first column in the tables represents the methods of the pre-processor. "STP" denotes that the STP filter is used as a pre-processor. The result utilizing both the STP filter and LTP filter is given in the "STP and LTP" row. The frame size and the filter length of the STP analysis is 120 samples and 16 taps, respectively.

The experimental results given in Tables 2 and 3 verify that utilizing the STP filter before the transient noise reduction is not good for preserving speech because it models transient noise component thus it brings the residual noise problem in the synthesized signal. Oppositely, utilizing only the LTP filter before the median filtering preserves only speech component. Consequently, the median filter can successfully remove transient noise while not distorting the speech. If we discard transient noise presence region during the pitch lag estimation process given in Eq. (13), the residual noise in the enhanced speech becomes much smaller than the system with the conventional LTP. Both the SSNR and the LSD are improved by utilizing the LTP with the modified pitch lag detector in Eq. (13). Sometimes it cannot estimate the pitch component correctly when the transient noise is located at the onset or the transition region of the vowel. However, the pitch estimation problem in the onset and the transition region can be solved by adopting the proposed non-causal LTP method. The results with the non-causal pitch lag estimation, "LTP with Eq. (14)", show the best performance in all objective quality measurements because of improved pitch modeling accuracy.

Table 2 Objective quality evaluation results of enhanced signals.

Algorithm	Without OM-LSA			With OM-LSA		
	SNR	SSNR	LSD	SNR	SSNR	LSD
Input	-4.97	-11.15	23.10	-2.67	-8.67	21.03
STP	-2.74	-10.49	22.12	7.70	-0.78	13.31
STP and LTP	-2.65	-10.51	22.44	-1.25	-8.35	20.53
LTP with Eq. (5)	5.96	-3.31	15.16	7.70	-0.78	13.31
LTP with Eq. (13)	5.88	-3.14	14.82	7.58	0.64	12.29
LTP with Eq. (14)	6.68	-2.52	14.26	9.06	0.50	12.74

The SNRs, SSNRs, and LSDs between enhanced signals and desired speech which are measured in both speech presence and absence regions.

Table 3 Objective quality evaluation results of enhanced signals measured only in speech presence region.

Algorithm	Without OM-LSA			With OM-LSA		
	SNR	SSNR	LSD	SNR	SSNR	LSD
Input	-3.71	-4.14	17.48	-1.31	-1.80	15.57
STP	-1.57	-4.21	16.95	-0.04	-2.55	15.50
STP and LTP	-1.42	-3.96	17.09	0.09	-1.64	15.25
LTP with Eq. (5)	6.27	2.37	10.74	7.94	4.38	9.55
LTP with Eq. (13)	6.17	2.47	10.44	7.67	5.07	9.26
LTP with Eq. (14)	7.04	3.15	9.90	9.29	5.52	9.07

The SNRs, SSNRs, and LSDs between enhanced signals and desired speech which are measured in speech presence region only.

The results with and without the OM-LSA estimator show same tendency. When the background noise exists, the speech modeling accuracy of the LTP filter is degraded by the background noise. However, the LTP analysis and synthesis process does not amplify the background noise component because the LTP method prevents the over-estimating of the signal. Since the pitch prediction gain is restricted to a certain constant, e.g., 1.2, the synthesized signal does not become much larger than the input [12]. The results utilizing the OM-LSA estimator show much higher objective scores because the background noise reduction process improves the output quality and pitch estimation efficiency. Though the proposed system works well even when background noise exists as shown in Tables 2 and 3, we recommend to remove the background noise before the LTP analysis and the transient noise reduction process.

The output waveforms which utilize the STP or the LTP filter as the pre-processor of the median filter are depicted in Figure 5. Figure 5a,b denote the waveforms

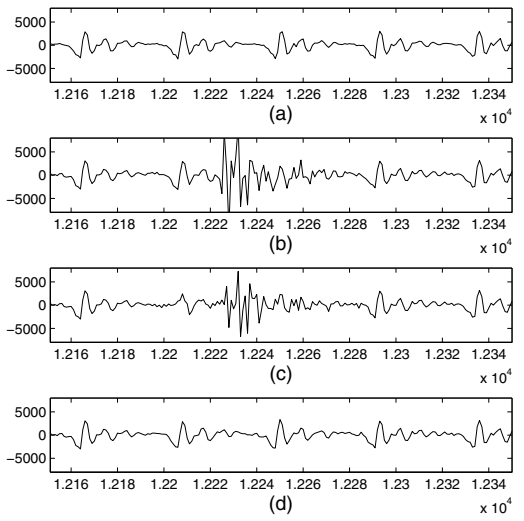

Figure 5 Results of transient noise reduction utilizing the STP and LTP filters. Time-domain waveforms of **(a)**: Clean speech, **(b)**: Noise corrupted speech, **(c)**: Median filter output utilizing the STP filter, and **(d)**: Median filter output utilizing the LTP filter.

of the desired speech and the noisy input, respectively. The enhanced output signals utilizing the STP pre-filter and the LTP pre-filter are represented in Figure 5c,d, respectively. The output with the proposed method, Figure 5d, successfully re-synthesizes the desired speech, but the output with the STP filter contains much residual noise. The perceptual evaluation of speech quality (PESQ) scores are also measured to compare the perceptual quality of output signals [21]. The PESQ scores for each speech sentence and the mean of the scores are represented in Tables 4 and 5. Tables 4 and 5 show the results with and without the OM-LSA estimator, respectively. The first columns in the tables denote the index of the speech signals where "Female" and "Male" indicate the gender of the speaker who pronounced the desired speech. The first rows in the tables denote the kind of the speech modeling pre-processor. The PESQ results show the same tendency with the objective evaluation results. However, the results adopting the non-causal LTP is not improved in some input signals comparing with the results with the modified causal LTP. In some input signals, transient noise does not exist at the onset and the transition region of the desired speech, thus the accuracy of the non-causal LTP and the causal LTP is not much different.

If we do not utilize the OM-LSA estimator before the transient noise reduction, the background noise somewhat disturbs the pitch estimation process thus the output quality improvement by adopting the modified LTP methods, i.e., Eqs. (13) and (14), is not enough as given in Table 4. On the contrary, the PESQ scores utilizing the modified LTP methods are notably improved when the background noise is removed before the LTP analysis because the accuracy of the LTP methods depends on input SNR. As a result, the PESQ scores utilizing the modified LTP methods become close to 3 which indicates that the output quality is in a perceptually fair category.

Table 4 PESQ scores without background noise reduction.

Algorithm	Input	STP	STP and LTP	LTP with Eq. (5)	LTP with Eq. (13)	LTP with Eq. (14)
Female 1	2.11	2.25	2.25	2.38	2.4	2.39
Female 2	1.22	1.50	1.50	2.12	2.12	2.14
Female 3	1.39	1.91	1.88	2.54	2.54	2.62
Female 4	1.63	1.67	1.72	2.22	2.21	2.25
Male 1	1.73	2.02	1.99	2.54	2.59	2.59
Male 2	1.38	1.77	1.74	2.30	2.31	2.34
Male 3	1.98	2.07	2.05	2.26	2.27	2.26
Male 4	1.40	1.76	1.78	2.40	2.41	2.44
Average	1.60	1.87	1.86	2.34	2.36	2.38

The PESQ scores of input and enhanced signals utilizing various speech modeling filters before the transient noise reduction. The input signals and the output signals contain background noise which become a reason of speech quality degradation. The first row represents the methods applied before median filtering. The first column denotes the kind of desired speeches.

6 Conclusion

We have proposed a system for reducing transient noise in speech signal. The proposed system utilizes a modified LTP filter as the pre-processor of the noise reduction filter to protect speech information from being removed while performing a noise reduction process.

Table 5 PESQ scores with background noise reduction.

Algorithm	Input	STP	STP and LTP	LTP with Eq. (5)	LTP with Eq. (13)	LTP with Eq. (14)
Female 1	2.57	2.76	2.75	3.11	3.17	3.17
Female 2	1.57	1.76	1.77	2.65	2.70	2.69
Female 3	1.44	1.83	1.82	2.74	2.70	2.86
Female 4	1.99	1.98	2.04	2.69	2.67	2.77
Male 1	1.86	2.15	2.14	2.89	3.09	3.10
Male 2	1.21	1.53	1.51	2.63	2.74	2.81
Male 3	2.44	2.57	2.55	3.13	3.16	3.15
Male 4	1.81	2.04	2.00	2.74	2.83	2.82
Average	1.86	2.08	2.07	2.82	2.88	2.92

The PESQ scores of input and enhanced signals utilizing various speech modeling filters before the transient noise reduction. The input signals are firstly processed by the OM-LSA estimator to remove the background noise. The first row represents the methods applied before median filtering. The first column denotes the kind of desired speeches.

The conventional LTP sometimes models the information of transient noise thus it increases the amount of the residual noise. The modified LTP method proposed in this article is effective to preserve and restore speech information in transient noise presence regions while not being affected by the transient noise component. The non-causal way of the LTP further improves the pitch modeling accuracy thus it effectively recovers desired speech after the noise reduction process. Objective quality measurements and PESQ score verified the superiority of the proposed method. Since the LTP process only preserves the pitch component, the consonant of speech can be distorted when transient noise exists in the region. Especially, the burst of plosive speech is somewhat reduced when the median filter is applied to the burst region. However, the characteristic of plosive sound including the burst remains after the median filtering because the filter length is short enough. In other words, only the amplitude of the consonant is reduced and its characteristic is not much distorted. Consequently, the distortion of plosive speech does not degrade the intelligibility and perceptual quality of the speech.

Endnote

[1]The proposed LTP method explained in Section 4 is used to summarize the results given in Figure 1 and Table 1.

Authors' contributions

M-SC conceived and designed the study, builded up the system, designed and performed the evaluation, and wrote the manuscript. H-GK guided the study, designed the evaluation, and corrected the manuscript. All authors read and approved the final manuscript.

Competing interests

The authors declare that they have no competing interests.

Received: 23 March 2011 Accepted: 30 December 2011
Published: 30 December 2011

References

1. SF Boll, Suppression of acoustic noise in speech using spectral subtraction. IEEE Trans Acoust Speech Signal Process. **ASSP-27**, 569–571 (1979)
2. Y Ephraim, D Malah, Speech enhancement using a minimum mean-square error short-time spectral amplitude estimator. IEEE Trans Acoust Speech Signal Process. **ASSP-32**, 1109–1121 (1984)
3. PC Loizou, *Speech enhancement, Theory and practice*, (CRC Press, Boca Raton, FL, 2007)
4. T Kasparis, J Lane, Suppression of impulsive disturbaces from audio signals. Electron Lett. **29**(22), 1926–1927 (1993). doi:10.1049/el:19931282
5. SR Kim, A Efron, Adaptive robust impulse noise filtering. IEEE Trans Signal Process. **43**(8), 1855–1866 (1995). doi:10.1109/78.403344
6. I Kauppinen, Methods for detecting impulsive noise in speech and audio signals, in *Proc IEEE Int Conf on Digital Signal Process*. **2**, 967–970 (2002)
7. SV Vaseghi, *Advanced Digital Signal Processing and Noise Reduction*, 2nd edn, (John Wiley & Sons, Ltd, Chinchester, UK, 2000)
8. R Talmon, I Cohen, S Gannot, Speech enhancement in transient noise environment using diffusion filtering. in *Proc IEEE Int Conf on Acoust, Speech, Signal Process* 4782–4785 (2010)
9. AJ Efron, H Jeen, Detection in impulsive noise based on robust whitening. IEEE Trans Signal Process. **42**(6), 1572–1576 (1994). doi:10.1109/78.286980
10. MS Choi, HG Kang, Transient noise reduction in speech signal utilizing a long-term predictor. J Acoust Soc Korea (in press)
11. AM Kondoz, Digital Speech - Coding for Low Bit Rate Communication Systems, (John Wiley & Sons, Ltd, Chinchester, UK, 1994)
12. ITU-T, *ITU-T recommendataion G.729* (1996)
13. TF Quatieri, Discrete-Time Speech Signal Processing, (Prentice Hall, Inc., Upper Saddle River, NJ, 2001)
14. A Papoulis, SU Pillai, *Probability, Random Variables and Stochastic Processes*, 4th edn, (McGraw Hill, New York, 2002)
15. J Beh, K Kim, H Ko, Noise estimation for robust speech enhancement in transient noise environment. in *Proc KSCSP 2007* 35–36 (2007)
16. MS Choi, HS Shin, YS Hwang, HG Kang, Time-frequency domain impulsive noise detection system in speech signal. J Acoust Soc Korea. **30**(2), 73–79 (2011)
17. I Cohen, Optimal speech enhancement under signal presence uncertainty using log-spectral amplitude estimator. IEEE Signal Process Lett. **9**(4), 113–116 (2002). doi:10.1109/97.1001645
18. I Cohen, Noise spectrum estimation in adverse environments: improved minima controlled recursive averaging. IEEE Trans Speech Audio Process. **11**(5), 446–475 (2003)
19. I Cohen, B Berdugo, Speech enhancment for non-stationary noise environments. Signal Process. **81**, 2403–2418 (2001). doi:10.1016/S0165-1684(01)00128-1
20. J Benesty, S Makino, J Chen, *Speech Enhancement*, (Springer, Berlin, 2005)
21. ITU-T, *ITU-T Recommendation P.862*, Perceptual evaluation of speech quality (PESQ), an objective method for end-to-end speech quality assesment of narrowband telephone networks and speech codecs, (2001)

doi:10.1186/1687-6180-2011-141
Cite this article as: Choi and Kang: **Transient noise reduction in speech signal with a modified long-term predictor.** *EURASIP Journal on Advances in Signal Processing* 2011 **2011**:141.