

# Unwanted Transients Reduction in Voice Signal by Applying a Predictor and Spectral Subtraction Process

Pushpraj Tanwar  
Research Scholar in ECE Dept. Maulana Azad  
National Institute of Technology Bhopal, India

Ajay Somkuwar  
Professor in ECE Dept. Maulana Azad National  
Institute of Technology Bhopal, India

## ABSTRACT

This work introduces an efficient median filter based algorithm to remove unwanted transient in voice signal. The projected Spectral subtraction process implements a modified predictor (MP) for long term as the mainframe of the unwanted transient reduction process to reduce voice distortion due to nonlinear nature of median strainer. To minimize residual unwanted transients and voice distortion after the unwanted transient reduction, MP process estimates the features of voice more accurately. By ignoring unwanted transient presence regions in the pitch lag finding phase, the MP successfully evades being influenced by unwanted transient. A Spectral subtraction algorithm is compared with Modified predictor to reduce voice distortion in the inception regions. Investigational results show the system effect how much they eliminate transient noise while preserving desired voice signal.

## General Terms

Speech Signal Processing, Voice Quality Analysis, Signal Prediction, Pattern Analysis.

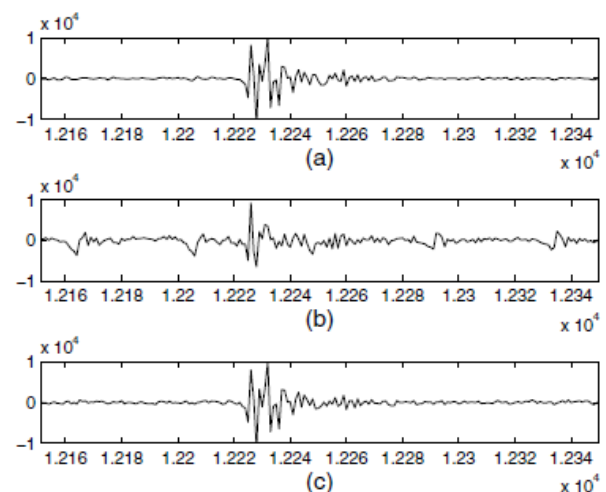
## Keywords

Voice Enhancement, Transient Noise Reduction, Modified Predictor, Norm Filter.

## 1. INTRODUCTION

Reducing noise from noise-corrupted voice 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 features of the unwanted transients 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 quiet remains in outcome. Furthermore, if noisy transient is summed up to voice, determining the extent of signal power reduction is difficult because the level of the voice waveform varies rapidly. Consequently, the power limiter is not efficient to eliminate transient noise in voice [5,7,9]. A median filter is a signal dependent filter which removes the high frequency components while preserving slowly varying components of incoming signal [4,6,7,10]. The basic median filtering 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 features of fast varying region of voice, i.e., around pitch epoch. Therefore, an additional pre-processing step to keep the voice characteristic before applying the basic median filter is needed. The purpose of the pre-processor is passing transient noise components but keeping voice information by using voice modeling filter not to be showing affect by median filtering afterwards. Typical voice modeling methods such as STP and the LTP are good candidates for the pre-processing module. The STP filter represents the short-term characteristic of voice, and the LTP filter does the long-term periodic components. If the STP or the LTP filter extracts all voice components from input and leaves all transient noise components in the remaining signal, the basic median filter may be successfully used to eliminate the noisy transient at the remaining signal. There it is testified that applying both STP and LTP to voice is effective to represent the features of the voice [10-12]. After removing noisy transient from remaining signal, the voice component extracted by STP filter or LTP filter should be re-synthesized. Please note that the pre-filter should not keep the features of transient noise not to bring any residual noise. The 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 voice and transient noise to show the influence of voice modeling filters.



**Fig1: signal when applying voice modeling filter to noisy voice signal. Waveforms in Time-domain of (a) Noise signal, (b) Resultant signal when STP analysed, and (c) Resultant signal when LTP analysed [22]**

Figure 1a represents a transient noise segment which is added to voice signal. Figure 1b,c are residual signals after performing the STP and the LTP analysis, respectively. Note

that the remaining signal in Figure 1c is not processed by the STP filter but only processed by the LTP analysis filter. As the Figure 1b shows, the STP analysis removes the noisy transient component. It indicates that the STP filters somewhat models the features of noisy transient. However, the remaining 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 noisy transient component. Consequently, applying the basic median filter to the LTP residual should be quite effective to remove the noisy transient. 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 search range. In such case, the LTP filter models the features of noisy transient and brings residual noise in synthesized voice. Another problem of the straight LTP method is that the LTP filter cannot preserve pitch information at the onset and the transition region of voice because a reference pitch does not exist. As a result, the straight LTP method needs to be modified to accurately model the pitch related voice 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 noisy transient region while searching a reference pitch. However, skipping the noisy transient region occasionally results in failure in the pitch prediction when the noisy transient is located where the reference pitch is present. Consequently, we spread the pitch hunt range to cover multiple pitch periods. The pitch estimation problem at the onset and the transition region of voice 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 voice after the noisy transient reduction [22].

The rest of this article is organized as follows. In the following section, the basic median filter for removing transient noise is briefly described. The straight LTP method which is generally used for voice coding is given in Section 3. The noisy transient reduction system with the modified LTP method is basic in Section 4. Signal sampling in section 5, Spectral subtraction in section 6, Experimental results and parameter in Sections 7 and finally conclusion in sector 7.

## 2. MEDIAN FILTERING FOR TRANSIENT NOISE REDUCTION

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

$$x(n) = s(n) + d(n). \quad (1)$$

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

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

Where  $T_k$  defines the occurrence time of the  $k_{th}$  transient noise.  $h_k(n)$  And  $g_k(n)$  denote the impulse response and the amplitude of the  $k_{th}$  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 basic median filter because it efficiently removes transient noise while preserving the slowly varying constituent in input signal. In other words, the slowly varying component of desired voice remains at output of median filter. Moreover,

the basic median filter is easy to implement because it does not need any pre-defined threshold. Though the basic median filter is effective for eliminating transient noise, however, it may also distort the features of desired voice while removing the very fast changing constituent. Therefore, the filter should be applied only to noisy transient occurrence region to diminish the voice distortion problem.

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

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 extent 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 noisy transient presence which becomes one when the noise exists and vice versa. It can be determined by comparing 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 basic in [16] shows 99.3% of detection accuracy while making only 1.49% of false-alarm. Employing the noisy transient detection result, the basic median filter can be applied only to the noise presence region. However, the voice 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 voice. Especially, the fast varying components in voice 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 voice by scaling a voice segment at one pitch period before. It efficiently estimates periodic and stationary component in signal [10-12].

$$\tilde{x}(m, l) = g_p(l)x(m - \tau_p(l), l); \quad 0 \leq m \leq M - 1, \quad (4)$$

where  $l$  and  $M$  denote the frame index and the extent frame, respectively. The index  $(m, l)$  represents the  $m$ th sample in the  $l$ th 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:

$$\tau_p(l) = \underset{\tau_{\min} \leq \tau \leq \tau_{\max}}{\text{argmax}} \frac{\sum_{m=0}^{M-1} x(m, l)x(m - \tau, l)}{\sqrt{\sum_{m=0}^{M-1} x^2(m - \tau, l)}} \quad (5)$$

where the range of  $\tau$  is determined by considering the general pitch period of human's voice, e.g.,  $2.5 \text{ ms} \leq \tau \leq 18 \text{ ms}$ . Since  $\tau_p(l)$  in Eq. (5) is the integer multiple of sampling duration 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 mend LTP accuracy [12]. The gain,  $g_p(l)$ , to minimize the signal modeling error is defined as:

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

Nevertheless, LTP gain is normally restricted to a constant to escape over assessment of 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 LTP system [12]. Utilizing the estimated pitch lag and gain, the LTP analysis filter extracts the pitch component from the input voice.

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

where  $r(m, l)$  denotes the remaining signal after the LTP analysis. To synthesize the desired voice from the remaining signal, the pitch period, the gain, and the previously synthesized voice segment are needed. Assuming that they are just recognized, synthesizing method 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 worth of the currently synthesized voice segment depends on the worth of the previous pitch. In other words, the pitch synthesis error at the previous structure can be spread to the next structure [12].

#### 4. THE MODIFIED PREDICTOR FOR LONG TERM

The basic algorithm employs the LTP as a mainframe of median filter, but note that the STP filter which is usually used in voice analysis systems is not utilized because STP filter may model not only voice component but also the features of transient noise. As a result, applying STP filter results in the residual noise to re-synthesized voice after the noise reduction [7,8,10]. The straight LTP method predicts a voice segment by utilizing a previous voice segment at one pitch period before [10-12]. Unlike STP filter, LTP filter does not have affect by short-term characteristic of transient noise. However, the LTP filter also models transient noise component if noisy transient exists within the search range of the pitch lag. One way of reducing the problem is to skip the noisy transient region while searching the pitch lag. Note also that, the straight LTP method cannot estimate pitch at the onset or the transition region of vowel because the reference pitch segment does not exists. The basic method utilizes look-ahead samples to predict the current voice segment more accurately thus it becomes more appropriate for preserving the voice component in transient noise environment [22]. In this segment, we firstly propose the noisy transient reduction system based on the basic median filter which utilizes the LTP as a pre-processor. The basic system adopts a non-predictive voice synthesis method thus the error caused by the basic median filter is not propagated to future voice samples. In Section 4.2, the modified LTP method is basic to efficiently estimate voice constituent however not affected by noisy transients [22].

#### 5. NOISY VOICE SIGNAL SAMPLING

Let us assume the clean voice  $S(t)$  corrupted by transient noisy  $N(t)$  and revealed in Fig. 1. Noisy voice signal is a continuous-time function that is transformed to an electrical voltage signal  $X(t)$  using a transducer as microphone linked to digital recorder structure. The transducer makes this translation by sensing the varying air force pressure from audio signal. The A-to-D transformer converts continuous-time noisy signal into a discrete-time noisy voice signal  $X[n]$ . An unbroken signal is experimented at correspondingly spread out impulses at time  $tn = nTs$  as follows

$$X[n] = X(nTs), \quad (10)$$

Where  $T_s$  represents the sampling duration or constant time amongst every sample, every impulse worth of  $X[n]$  is termed sample of discrete signal. Now the sampling duration can be embodied as a fix sampling period rate:

$$f_s = 1/T_s \text{ Hz} \quad (11)$$

Here with in this work, a clean voice signal graph  $S[n]$  was logged with the help of software named sound recorder which was set up on a computer. The waveform of voice data is given in Fig. 4(a). The voice signal recorded up to extent of 640 ms. The Shannon theorem for sampling states that one continuous-time signal with extreme spectrum frequency  $f_{max}$  can be remodelled faithfully from its samples  $X[n] = X(nTs)$  if samples taken are larger than  $2 f_{max}$ . Since auditory incidences of audible jingles collection from 20Hz to 20KHz, hence, in required applications,  $f_{max}$  is around 20000 Hz. The rate of sampling was routinely calculated in Matlab, and has rate 44.1 kHz, that is higher than twofold of 20 kHz, plus satisfies Shannon sampling formula. There are complete 28,444 samples through period interval of 22.68  $\mu$ s between each sample [23]

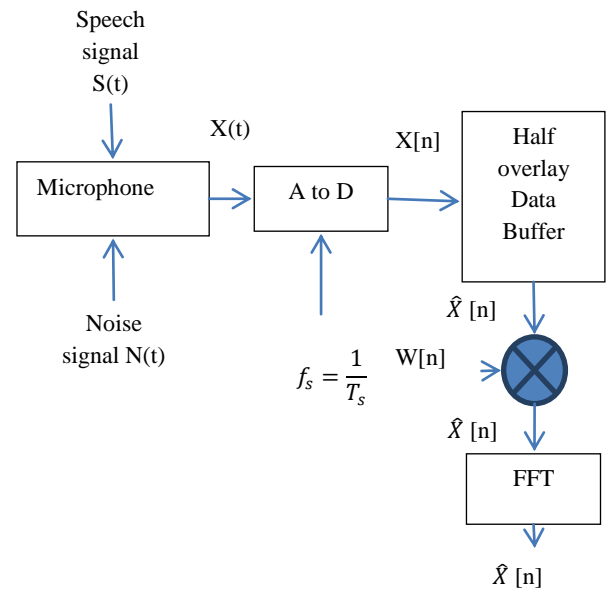


Fig 2: Illustration of noisy voice production and conversion in discrete data

the records in arithmetical worth, and are kept into pieces. Every segment contains 256 sections of noisy voice. Each segment is called a data-buffer  $X^{\wedge}[n]$ . All the data buffer 50 percent overlays with consecutive information buffer by overall 128 samples [23]. Our noisy voice has 221, 50 percent overlapped data buffers that include the whole extent of noisy voice data.

#### 6. NOISE DEDUCTION BY SPECTRAL SUBTRACTION

Here the sub unit covers, structures be around procedure was realised in demand to increase the enactment of spectral subtraction for noise deduction strategy [27]. Structures are an average of an optional stage between calculating the mediocre magnitude of noise spectrum then subtraction of this mediocre from extent of noisy voice structures. While applying, structures averaging comprise with magnitude mediocre of numerous structures of noisy voice rather than single structure at a stint. It is restricted this investigation to take averaging whichever three otherwise six successive structures. Greater numbers might outcome in lessening the voice eloquence [24]. Now the investigation effects the changing the sizes of half-overlapping information buffer and Hanning window for time improving transients removal design, The overlying of data buffers was changed between half and quarter overlapping and Hanning window span was speckled

beginning 256 points to half of it, and double of 256 points. This process is revealed with in Figure 3.

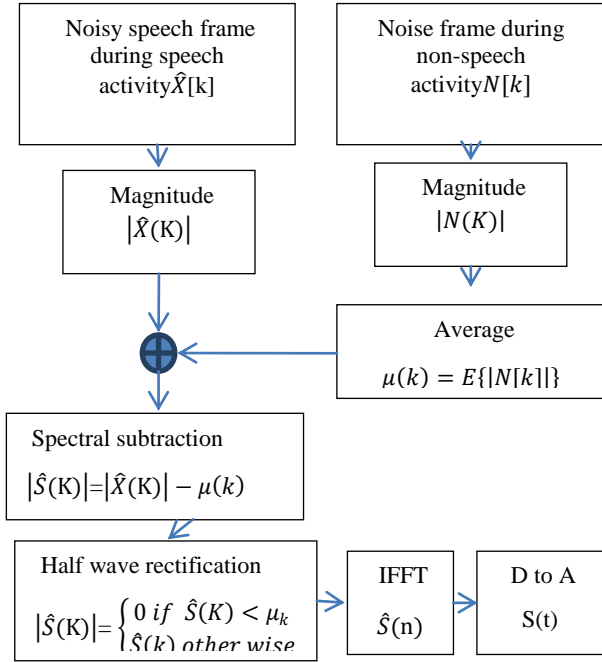


Fig 3: flowchart of unwanted transient removal Spectral subtraction

## 7. PERFORMANCE EVALUATIONS

To evaluate the enactment of the basic system, we apply it to recorded voice signals which contain transient noise. Every voice signals and transient noise signals are recorded in real environment, separately. The noisy transient signals are acquired by using mobile recording devices while clicking buttons on the recording devices or tapping the body of the recording devices. We add the noisy transient segments to the random points of time of the voice signals. More than one hundred transient noise sequences are added to eight sentences of voice signals. Voice database is recorded by four male and four female speakers, and the total extent voice signals is about sixteen seconds. The sampling frequency of the voice is 8 kHz. Since the noisy transient 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 voice, transient noise, and background noise. The signal-to-noise ratio (SNR) between the desired voice and the contextual noise is nearby 14 dB. The median filtering and LTP filtering are applied merely at transient disturbance presence section by using hand marked result of the noise presence. However, the noisy transient presence region can be detected by measuring the time- or the frequency-domain energy of incoming signal with certain threshold [4,15,16]. Experimental results utilizing the noisy transient detector basic in [16] are almost same as results with the hand marked noise detection result shown in this article.

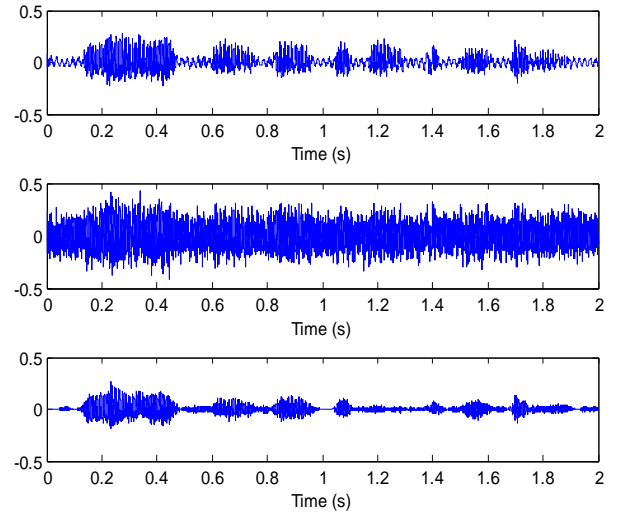


Fig 4: (a) Clean voice signal  $S(n)$ ; (b) signal with transient noise; (c) Signal after unwanted transient removing by spectral subtraction.

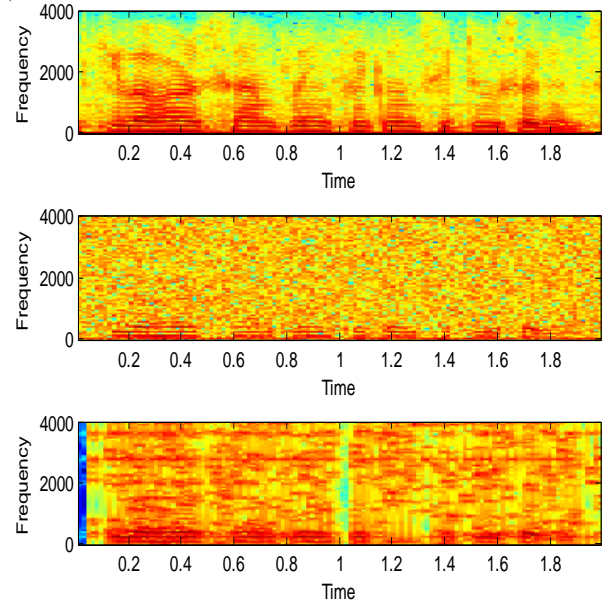


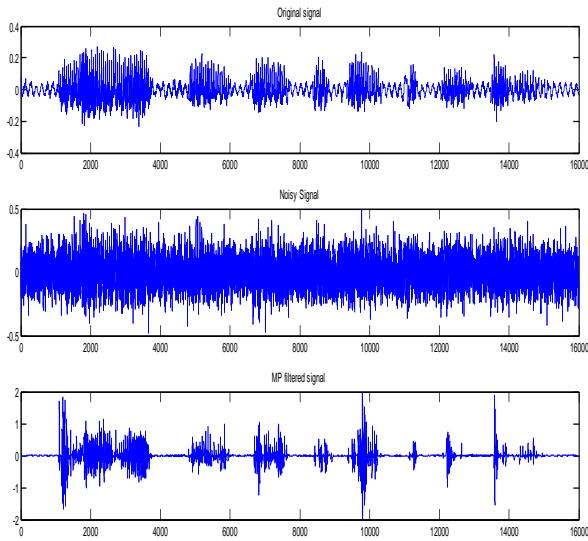
Fig 5: (a) Voice signal spectrum of clean speech; (b) spectrum of signal with transient noise; (c) Signal spectrum after unwanted transient removal by spectral subtraction

The extent median filter,  $2w + 1$ , used for the experiments are 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. (6) and (7). 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. To evaluate the enactment of noisy transient reduction systems, we measure SNR, segmental signal to noise ratio (SSNR), and log-spectral distance (LSD) between output signals and a clean voice such as [20]:

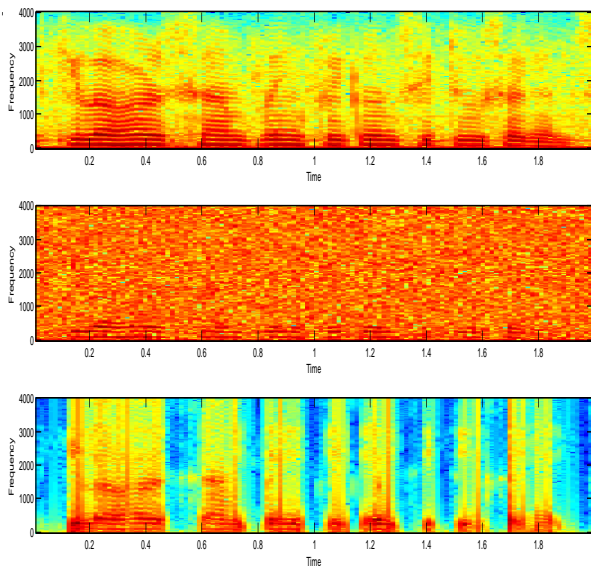
$$SNR = 10 \log_{10} \left( \frac{E_{m,l}\{s(m,l)^2\}}{E_{m,l}\{s(m,l)-y(m,l)\}^2} \right)$$

$$SSNR = E_l \left\{ 10 \log_{10} \left( \frac{E_{m,l}\{s(m,l)^2\}}{E_{m,l}\{s(m,l)-y(m,l)\}^2} \right) \right\} \quad (12)$$

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 voice and system output, respectively.



**Fig 6: Results of transient noise reduction utilizing LTP filters. Time-domain waveforms of (a) Clean voice, (b) Noise corrupted voice, (c) Median filter output utilizing the Modified Predictor filter**



**Fig 7: Spectrum of transient MP filters. (a): Clean voice spectrum, (b): Noise corrupted voice signal spectrum; (c): MP filtered voice signal spectrum**

**Table:-1 SNR and SSNR of Modified Predictor and Spectral Subtraction**

Method	SNR	SSNR
Modified predictor	3.842	-0.6861
Spectral subtraction	11.750	8.5941

## 8. CONCLUSIONS

We have systems for reducing transient noise in voice signal. The transient reduction filters as the pre-processor of the noise reduction filter to protect voice information from being removed while performing a noise reduction process. Noisy voice was generated digitally by debasing the data of the clean voice. Eradicating the noise wants an estimate of unwanted transient for the duration of voice bustle. The estimate of unwanted transients was found by considering average extent of unwanted transient spectrum for the period of non-voice bustle. The average extent of noise spectrum thru non-voice bustle was deducted from noisy voice spectrum for the duration of voice bustle. The early noise deduction design comprised no structure be an average with 50 percent overlapping of data buffers besides half of 512 points in Hanning windows. The resultant orientation signal to noise relation is 11.750 dB. The modified Predictor method is effective to preserve and restore voice information in noisy transient occurrence regions whereas it does not have the affect by noisy transient component. Since the modified Predictor process only preserves the pitch component, the consonant of voice can be misleading while noisy transient presents in the region. The spectral subtraction method from table 1 shows very good improvement in the SNR and SSNR as compared to the modified prediction filter. The future work by applying different filter combinations on speech signals to improve the transient removal performance.

## 9. REFERENCES

- [1] Boll S. F., 1979, Suppression of acoustic noise in speech using spectral subtraction. IEEE Trans Acoust Speech Signal Process. ASSP-27, 569–571.
- [2] Ephraim Y., Malah D., 1984, Speech enhancement using a minimum mean-square error short-time spectral amplitude estimator. IEEE Trans Acoust. Speech Signal Process. ASSP-32, 1109–1121.
- [3] Loizou P. C., 2007, Speech enhancement, Theory and practice, (CRC Press, Boca Raton, FL).
- [4] Kasparis T., Lane J., Suppression of impulsive disturbances from audio signals. Electron Lett. 29(22), 1926–1927 (1993). doi:10.1049/el:19931282
- [5] Kim S. R., Efron A., 1995. Adaptive robust impulse noise filtering. IEEE Trans Signal Process. 43(8), 1855–1866.
- [6] Kauppinen I., 2002, Methods for detecting impulsive noise in speech and audio signals, in Proc. IEEE Int. Conf. on Digital Signal Process. 2, 967–970.
- [7] Vaseghi S. V., 2000, Advanced Digital Signal Processing and Noise Reduction, 2nd edn, (John Wiley & Sons, Ltd, Chichester, UK).
- [8] Talmon R., Cohen I., Gannot S., 2010, Speech enhancement in transient noise environment using diffusion filtering, in Proc IEEE Int. Conf. on Acoust, Speech, Signal Process 4782–4785.
- [9] Efron A. J., Jeon H., 1994, Detection in impulsive noise based on robust whitening. IEEE Trans Signal Process. 42(6), 1572–1576.

- [10] Choi M.S., Kang H.G., Transient noise reduction in speech signal utilizing a long-term predictor. J Acoust Soc. Korea.
- [11] Kondoz A. M., 1994, Digital Speech - Coding for Low Bit Rate Communication Systems, (John Wiley & Sons, Ltd, Chinchester, UK.
- [12] ITU-T, 1996, ITU-T Recommendation G.729.
- [13] Quatieri T. F., 2001, Discrete-Time Speech Signal Processing, Prentice Hall, Inc., Upper Saddle River, NJ.
- [14] Papoulis A., Pillai S. U., 2002, Probability, Random Variables and Stochastic Processes, 4th edn., (McGraw Hill, New York).
- [15] Beh J., Kim K., Ko H., 2007, Noise estimation for robust speech enhancement in transient noise environment. in Proc KSCSP 2007 35–36.
- [16] Choi M. S., Shin H. S., Hwang Y. S., Kang H. G., 2011, Time-frequency domain impulsive noise detection system in speech signal. J Acoust Soc. Korea. 30(2), 73–79.
- [17] Cohen I., 2002, Optimal speech enhancement under signal presence uncertainty using log-spectral amplitude estimator. IEEE Signal Process Lett. 9(4), 113–116.
- [18] Cohen I., 2003, Noise spectrum estimation in adverse environments: improved minima controlled recursive averaging. IEEE Trans Speech Audio Process. 11(5), 446–475.
- [19] Cohen I., Berdugo B., 2001, Speech enhancement for non-stationary noise environments. Signal Process. 81, 2403–2418.
- [20] Benesty J., Makino S., Chen J., 2005, Speech Enhancement, Springer, Berlin.
- [21] ITU-T, ITU-T Recommendation P.862, 2001, Perceptual evaluation of speech quality (PESQ), an objective method for end-to-end speech quality assessment of narrowband telephone networks and speech codecs.
- [22] Choi M. S. and Kang H. G., 2011, Transient noise reduction in speech signals with a modified long-term predictor EURASIP Journal on Advances in Signal Processing, 2011:141.
- [23] Karam, M., Khazaal, H.F., Aglan, H. and Cole, C. 2014, Noise Removal in Speech Processing Using Spectral Subtraction, Journal of Signal and Information Processing, 5, 32-41.
- [24] Boll, S.F. 1979, Suppression of Acoustic Noise in Speech Using Spectral Subtraction. IEEE Transactions on Acoustic, Speech and Signal Processing, 27, 113-120.
- [25] Rabiner L.R, and Schafer R.W., 1978, Digital Processing of Speech Signals. Prentice Hall, Upper Saddle River.
- [26] Quantieri T.F. 2001, Discrete-Time Speech Signal Processing: Principles and Practice. Prentice Hall, Upper Saddle River.
- [27] Allen J. 1977, Short Term Spectral Analysis, Synthesis, and Modification by Discrete Fourier Transform. IEEE Transactions on Acoustic, Speech and Signal Processing, 25, 235-238.