DUALISTIC THRESHOLD BASED MIN-MAX METHOD FOR VOICE SIGNAL ENHANCEMENT

Pushpraj Tanwar and Ajay Somkuwar

ECE Department Maulana Azad National Institute of Technology Bhopal, India

ABSTRACT

In this article, unwanted transitory assessment is specified by averaging previous phantom power standards and through a levelling stricture that is in step by the symbol existence likelihood in sub-bands. A two threshold based min-max method is proposed that we compare with the recursive averaging (RA) approach for unwanted transient assessment with SAL (spectral amplitude with logarithm assessment) in voice signal enrichment .The SAL procedure includes a transient assessment and the unwanted transient PSDs in virtual stationary form, and enrichment of voice signal. It is proficient of tracing the incoming signal rapid variations in spectra and supports to assessment of PSD for transients effectively. The unwanted transient assessment is much efficient by proposed method as compared to SAL method and RA in terms of the voice signal power to noise ratio (SNR). And finally by simulation modelling we show the results.

KEYWORDS

Aural noise, signal assessment, spectral exploration, transient noise assessment

1. INTRODUCTION

The plain constituent of a real-world voice signal improvement scheme is the unwanted transient power spectrum assessment. The most collective tactic is to normal the voice signal with noise over non-voice signal segments. Voice signal silence revealing is realised on a frame foundation [1] and also assessment independently for separate sub-bands via *posteriori* SNR [2-3]. Conversely, the recognition consistency harshly worsens for week voice signal constituents and small incoming signal SNR. Also, the range of presumable non-voice signal fragments of signal possibly will not be enough, which confines the tracing proficiency of the unwanted transient assessment in circumstance of changeable unwanted transient spectrum. Otherwise, histograms may also be used to assessment of unwanted transient power spectral realm [3]–[4]. Martin [5] has suggested an unwanted transient assessment algorithm centred on minutes measurements. The unwanted transient assessment is achieved from minima standards of a levelled power assessment of the unwanted transient signal, grown thru a reason that reimburses the bias. A computationally new competent minimum tracking pattern is accessible from [7]. Its main hitch is the sluggish update speed of the unwanted transient assessment for an abrupt rise in unwanted transient vigour level with its trend to abandon the voice signal [8]. The existing *recursive averaging* (RA) approach for unwanted transient assessment is specified by around standards of previous spectral power, spending a levelling factor that is accustomed via the voice signal existence likelihood in sub-bands. It shows the presence of voice signal in a certain structure of a

sub-group can be resulted by the quotient among the native vigour of the voice signal with unwanted transients with its minimum inside a definite time frame. The proportion is equated to a certain onset worth, where the smaller ratio specifies absenteeism of voice signal. After that a time-based smooth out is carried to lessen fluctuations amongst voice signal and no voice signal segments, so that exploiting a solid association of voice signal occurrence in adjacent frames.

In the SAL procedure, to correct the unwanted transient PSD component, trace quicker PSD deviates and to keep the smooth components (mutually voice signal and circumstantial blare) seem such as "pseudo-stationary". At that time, calculating the SAL assessment centred onPSD assessment of the flat constituents allows improving the transitory portion and suppressing the voiced signal and circumstantial noise. It usages small STFT structures which diminish the voice signal variation amongst successive time frames. It is eminent that the tradition of tiny spell structures weakens the regularity perseverance. In the realistic trials, a spell structure size of 64 sections was exposed to be appropriate and returned good performance [17]. The subsequent unwanted transient assessment is calculation competent, stouten terms of input voice signal SNR in addition variety of principal additive unwanted transient and categorized with the capability to hastily track rapid fluctuations in the unwanted transient spectrum.

The proposed method solves the transient problem in two ways .the min and max values of threshold for transient reduction. The min of the threshold value is computed by taking the one measurement in the forward and backward direction. In the similar way the max values are also computed and the transient are suppressed. The further paper is organized as. The second Sector covers the basic unwanted transient spectrum model with sector III the minima precise assessment for the voice signal presence model. In Sector IV, contain the SAL assessment procedure and the proposed method in section V and the deliberate trial results in VI section and finally segment VII concludes precisely.

2. UNWANTED TRANSIENT SPECTRUM MODEL

Let $x(n)$ and $d(n)$ represent voice signal and unallied additive unwanted transient signals, correspondingly, where n is a discrete time index. The perceived signal $y(n)$, assumed by $y(n)=x(n)+ d(n)$, is spread into overlying frames through the frame task and scrutinized via short-time FT (STFT).Specifically

$$Y(k,l) = \sum_{n=0}^{N-1} y(n + lM)h(n)e^{-j(2\pi/N)nk} \tag{1}$$

Where k represents the rate bin catalogue, l represents the spell structure key, be situated an analysis frame of size N, and M stands the frame fill in stage in time. Known two assumptions, $H_0(k,l)$ and $H_1(k,l)$ indicate, respectively, the voice signal absence and existence in the lth structure of the kth subgroup, it has

$$H_0(k,l):Y(k,l) = D(k,l)H_1(k,l):Y(k,l) = X(k,l) + D(k,l) \tag{2}$$

Where, $X(k,l)$ and $D(k,l)$ denote STFT of voiced clean and unwanted transient signals, respectively. Let, $\lambda_d(k,l) = E[,|D(k,l)|^2]$ denote the discrepancy of the unwanted transient in the kth subgroup. And then a general technique to achieve unwanted transient assessment is to smear

a iterative levelling to the unwanted transient depth throughout periods of voice signal nonappearance. In individual,

$$H'_0(k,l): \hat{\lambda}_d(k,l+1) = \alpha_d \hat{\lambda}_H(k,l) + (1-\alpha_d)|Y(k,l)|^2$$

$$H'_1(k,l): \hat{\lambda}_d(k,l+1) = \hat{\lambda}_d(k,l) \quad (3)$$

Where $\alpha_d (0 < \alpha_d < 1)$ is a levelling factor H'_0 and H'_1 andlabel hypothetical voice signal nonappearance and appearance, respectively. Here, it make a discrepancy among the theory specified in(2), intended for assessment of unwanted transient free voice signal and the postulates in(3), which govern the self-edition of the unwanted transient spectrum. Clearly, the decisive voice signal is inattentive(H_0) when voice signal is existent ((H_1)is additional harsh when doing assessment of voice signal with the unwanted transients. Hereafter, different judgment instructions are working and be likely to adopt with (H_1) a greater confidence than$H'_1, P(H_1|Y) \geq, P(H'_1|Y)$. Let , $P'(k,l) = P(H'_1(k,l)|Y(k,l))$ designate the likelihood of signal presence with condition. The equation (3) infers

$$\hat{\lambda}_d(k,l+1) = \alpha_d(k,l)\lambda_d(k,l) + [1-\alpha_d(k,l)Y|(k,l)|^2 \quad (4)$$

$$\text{Where } \tilde{\alpha}_d(k,l) = \alpha_d + (1-\alpha_d)p'(k,l) \quad (5)$$

And $\tilde{\alpha}_d(k,l)$ represents a time changing levelling parameter, thus, the unwanted transient range can be predictable by an average of previous signal spectral power standards, using a levelling constraint that is accustomed by the voice signal occurrence likelihood.

3. VOICE SIGNAL PRESENCE MODEL

Voice signal existence in an indicated structure of a sub group is firmed by the fraction concerning the native vigour of the voice signal with unwanted transient and its least inside a stated time gap. Let the native vigour of the unwanted transiency voice signal be achieved by levelling, the square of magnitude of the aforementioned STFT in frequency and time. In frequency, it takes a frame task of length $2w+1$

$$S_f(k,l) = \sum_{i=w}^{W} B(I)|Y(k-i,L)|^2 \quad (6)$$

To take in interval, the levelling is done through an iterative mean of firstly mandate, specified by

$$S_f(k,l) = \alpha_s S(k,l-1) + (1-\alpha_s)S_f(k,l) \quad (7)$$

Where$\alpha_s(0 < \alpha_s < 1)$ is a parameter, The lowest of then a tivevigour, $S_{min}(k,l)$, is searched via a easy formula of the system anticipated in [5]. Firstly, the lowest and a momentary mutable $S_{tmp}(k,l)$, re initialized by $S_{min}(k,0) = S(k,0)$ and $S_{tmp}(k,0) = S(k,0)$. now, a section wise evaluation of the native vigour. The least standard of the preceding structure produces the minutest standard for the present structure

$$S_{min}(k,l) = min\{S_{min}(k,l-1), S(k,l)\} \qquad (8)$$

$$S_{tmp}(k,l) = min\{S_{tmp}(k,l-1), S(k,l)\} \qquad (9)$$

Whenever L structures have been considered, i.e. dividable by, the momentary mutable is hired and reset by

$$S_{min}(k,l) = min\{S_{tmp}(k,l-1), S(k,l)\} \qquad (10)$$

$$S_{tmp}(k,l) = S(k,l) \qquad (11)$$

Let$S_r(k,l) = S(k,l) / S_{min}(k,l)$ signify the quotient between the native vigour of the unwanted transiently voice signal and its consequent least standard. A Bayes lowest charge pronouncement tenet is specified by

$$\frac{P(S_r|P(H_1))}{P(S_r|P(H_0))} \lessgtr \frac{c_{10}|P(H_0)}{c_{10}|P(H_1)} \qquad (12)$$

where $P(H_0)$and $P(H_1)$ be a priori likelihoods for voice signal nonappearance and existence, separately, and $c_{i,j}$is the charge for determining $H'_i$when H'_j.

Meanwhile the likelihood fraction $P(S_r/H_1)/P(S_r/H_0)$ is a monophonic utility, the judgment tenet via (12) has been stated as

$$S_r(k,l) \lessgtr \delta \qquad (13)$$

This deal the resulting appraisal parameter for $p'(k,l)$:

$$\hat{p}_a(k,l) = \alpha_p \hat{p}'(k,l-1) + (1-\alpha_p)I(k,l) \qquad (14)$$

Where$\alpha_p(0 < \alpha_p < 1)$ is a levelling constraint$I(k,l)$ and designates an pointer utility for the outcome via (13), i.e. $I(k,l) = 1$ is S$S_r(k,l) > \delta$ and $I(k,l) = 0$ otherwise. The excellence of this appraisal is approx. three crinkle. Initially δ, do not affect by type and strength of ecological unwanted transient. Second, the likelihood of $Y^2 > \lambda_d$is precisely minor when $S_r < \delta$.

4. SAL ALGORITHM

The SAL assessment for transient intervention subdual is divided in double phases. First it exploits the various deviation amounts of voice signal, unwanted variations and circumstantial unwanted transients[15]. Then, calculating the SAL assessment centred on PSD assessment of the clear voiced components empowers to boost the transient fragment and quash the voiced signal and circumstantial unwanted transients.

It uses tiny STFT structures which diminish the voice signal variation amongst succeeding time frames. It is eminent that the practice of tiny spell structures diminishes the rate perseverance. In the empirical experiments, a time structure extent of samples with count 64 was exposed to be appropriate and produced noble enactment. This specific length relates to 4 ms for a 16000 Hz sampler frequency [15]. It modifies the iterative averaging method [13] to assessment the specified PSD of levelled signal constituents.

The valuation of PSD is centred on a levelled period gram attained by time iterative mean with phantom amplitude as follows $S(k, l) = \alpha_s S(k, l-1) + (1 - \alpha_s)|Y(k, l)|^2$. Thus, handover a lesser worth to the levelling factor α_s enables quicker tracing PSD. The worse αs is, the additional heaviness is given to the present spell mount, and consequence quicker discrepancies in specified PSD of voiced signal or background unwanted transients can be caught. To improve further the standard of αs is reduced starting a specified choice of $0.91 - 0.98$ upto 0.7.
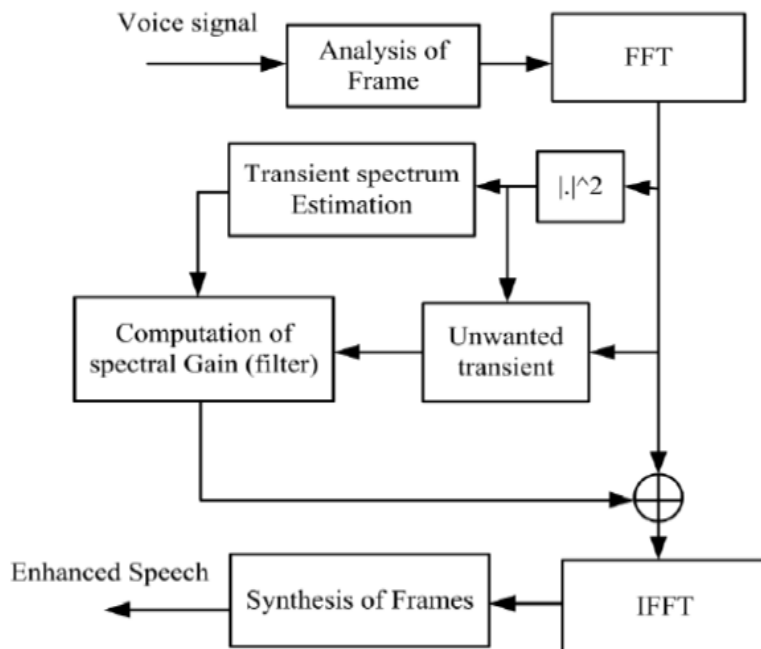

Fig. 1 Diagrammatic illustration of the SAL algorithm [15]

The transitory occurrence likelihood is organized by the least tenets of the levelled periodogram [14], which is acquired from a finite causative frame of size L

$$S_{min}^L(k, l) = min \ \{S(k, l), S(k, l-1), ... S(k, l+1)\} \tag{15}$$

Then, the transitory occurrence judgment is prepared using the succeeding tenet

$$S_r(k,l) \equiv \frac{S(k,l)}{S_{min}^L(k,l)} \gtrless \delta \qquad (16)$$

Where δ represents an experimental onset. As $S_r(k,l) > \delta$ the recent slot that is striking as holding a transitory. Else, it will beheld as comprising voiced signal and circumstantial unwanted transients. Assume $I(k,l)$ designate the transient occurrence constraint, defined as

$$I(k,l) = \begin{cases} 1, & S_r(k,l) > \delta, \\ 0, & \text{otherwise.} \end{cases}$$

Let $p(k,l)$ be the transitory existence likelihood, which is levelled agreeing to

$$p(k,l) = \alpha_p p(k,l-1) + (1-\alpha_p)I(k,l) \qquad (17)$$

Here α_p ($0 < \alpha_p < 1$) represents a levelling factor. Lastly, the specified PSD of levelled signal constituents is predicted by taking mean of previous voiced spectral influence standards with a levelling constraint that is accustomed via the transient existence likelihood:

$$\hat{\lambda}(k,l+1) = \hat{a}\hat{\lambda}(k,l) + [1 - \hat{a}(k,l)]|Y(k,l)|^2 \qquad (18)$$

Where

$$\hat{a}(k,l) + \alpha + (1-\alpha)p(k,l) \qquad (19)$$

And α represents a fixed levelling constraint ($0 < \alpha < 1$). The modified Iterative technique seizures maximum of the voiced signal and circumstantial unwanted transients fragments. It implements this distinction with an extra anti causative frame of span T. Assume $S_{min-\alpha}^T(k,l)$ be the least assessment within anti-causal structure, which is deliberated centred to

$$S_{min-\alpha}^T(k,l) = \min\{S(k,l), S(k,l+1) \ldots S(k,l+T-1)\} \qquad (20)$$

At that time, the maximal significance of the paired least spectra standards beginning causative and anti-causative frames is work out as

$$S_{min}(k,l) = \max\{S_{min}^T(k,l), S_{min-ac}^T(k,l)\} \qquad (21)$$

$S_{min}(k,l)$ Is regarded equally the assessment of the smooth voice signal spectral plane. The fig. 1 illustrates the flowchart of this method in simplified manner. It certifies the inceptions spectra are equated to the remaining phoneme than the fore going contextual unwanted transients. By substituting $S_{min}^L(k,l)$ with $S_{min}(k,l)$ in (16) it acquired the novel decision regulation [15]. The empirical tests recommend that a 40.0 ms extent anti-causality frame is suitable for several transients. Even though, the frame size should be centred to the specific transient interferences. It is also noted that an expected significance of expending an anti-

causality frame is the accumulation of period lag. Finally, the modified SAL output voice signal, which contains primarily the vigour of transient, for assessment the specified PSD of transitory as follows

$$\hat{\lambda}_t(k,l) = \left| \hat{T}(k,l) \right|^2 \qquad (22)$$

Where $\hat{T}(k,l)$ represents the modified SAL throughput transient amplitude assessment, this assessment is accomplished with simplicity of conversation. Alternatively, the transitory nosiness PSD could be assessed straight created on a-priory SNR [12].

5. PROPOSED METHOD

The proposed method is considering the multiple thresholds based for transient noise removal. It mainly provides the thresholds with adjustable values according to the coming frames. The transient noise is sudden changes in the coming sequence and hence it provides the adaptive nature for noise identification and it removal. The flowchart of the proposed algorithm is illustrated in fig. 2. It min max threshold calculation is the main part for noise detection. Then the noisy signal is replaced by the median of the considered window. The backward forward windows are taken for the min-max threshold calculation.
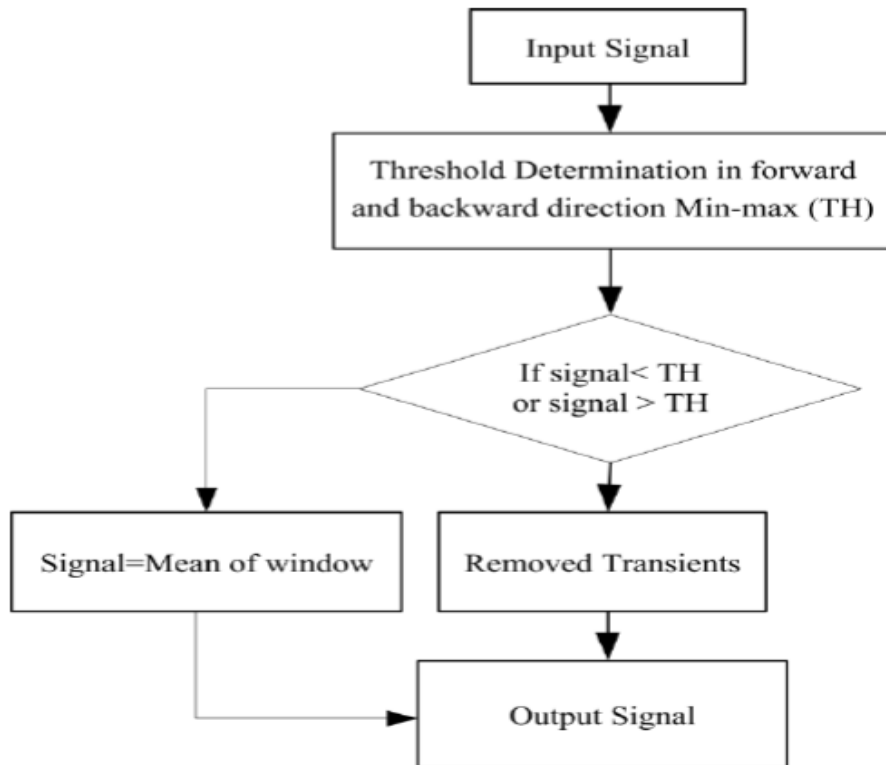

Fig: 2 Flowchart of the proposed Algorithm

We find the min of the threshold value by taking the one measurement in the forward and backward direction. In the similar way the max values is also computed and the transient are suppressed.

6. EXPERIMENTAL RESULT

For the first phase of the transitory PSD assessment the STFT structures of size 64 with voice signal enhancement stage extended structures of extent 512 are used. For both the cases, 75% overlap between successive structures is used. The amounts of the voice signal and the variations are re-scrabbled to the identical leading amplitude standard. This offers a reasonable comparison amongst various voice signals and transients. Here the enactment of the SAL algorithm and RA approach is evaluated. The logged voice signal signals with various orators, in teamwork man and lady, with mean size of 3.0 s. We randomly add 3− 12 recorded transient. The RA unwanted transient assessment is combined with the modified spectral amplitude with logarithm (SAL) assessment [6] for finding a clean voice signal assessment in unwanted transitory environs.

TABLE 1 Sectional SNR enhancement for various unwanted transient types and levels, obtained using the RA and SAL unwanted transient estimators

Input Sec. SNR $|dB|$	White Gaussian unwanted transients RA SAL Proposed	Vehicle Ignitron unwanted transients RASAL Proposed
-5	9.62 9.85 9.88	8.22 10.16 11.04
0	7.23 7.84 8.67	6.21 7.93 8.12
5	5.25 6.15 7.98	6.04 6.88 7.02
10	4.02 4.87 5.08	3.21 5.31 5.87

The valuation is centred on an impartial enhancement in sectional SNR, a training of voice signal spectrograms and natural attending trials. There are various unwanted transient varieties, avail from Noise-x92 record [9], are coined in this assessment: white Gaussian unwanted transients, vehicle inside unwanted transients, and $F16$ battleground unwanted transients. Collectively voice signal is ruined by the several unwanted transients with sectional SNRs in 5-10 dB range. The sectional SNR is well-defined by [10]

$$segSNR = \frac{10}{|L|} \sum_{l=L} \log \frac{\sum_{k=0}^{N/2} |X(k,l)|^2}{\sum_{k=0}^{N/2} |D(k,l)|^2} \qquad (26)$$

Where L represents the traditional structures those contain voice signal and its fundamentality. The rate by which sampling is done is 41000 samples per second. Table 1 indicate the mean sectional SNR enhancement obtained for several unwanted transient forms and on several unwanted transient levels.

Fig3 illustrates the examples of speech signal acquired for investigation in (a) andAK−47 unwanted transient in (b). The noisy signal with combination of both above is shown in (c) and the denoised signal is shown in fig 3 (d).
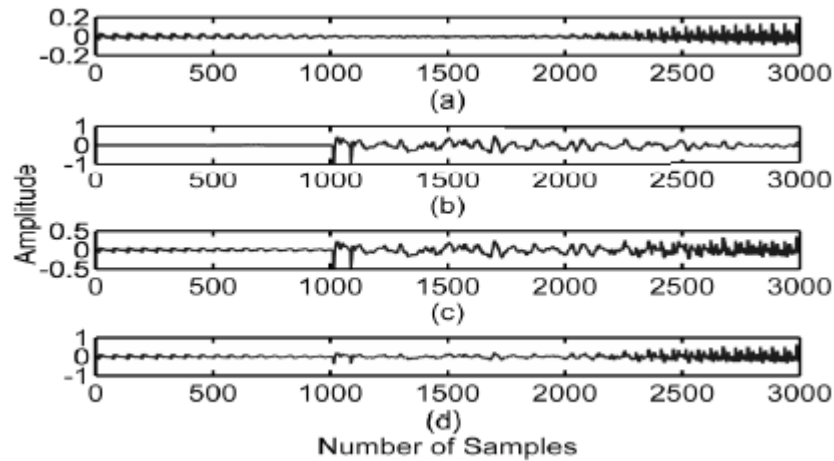
Fig. 3 (a) Original speech signal (b) Transient noise (c) Noisy signal as combination of both speech and transient noise, (d) De-noised signal through proposed method

Under varying unwanted transient conditions, strong unwanted transient components are misleadingly measured as voice signal constituents. This produces an underestimated unwanted transient, resulting in excessive level of melodious lasting unwanted transient. Meanwhile, small SNR produce an overemphasized unwanted transient, because of weak voice signal constituents that are partly handled as unwanted transient components.

Table 2 Voice signal assessment calculation for different transient interference

Unwanted Transient	Sec SNR Improvement by RA [dB]	Sec SNR Improvement By SAL [dB]	Sec SNR Improvement By Proposed Method [dB]	Delay [ms]
Metronome	3.9	4.7	4.9	40
Knocks	5.0	5.2	5.7	40
Shot	3.6	3.9	4.2	250
Scissors	5.1	5.4	6.1	100
Keyboard	4.1	4.8	5.2	40

Therefore, the SNR with weak voice signal components come to be straight poorer. This was established by a particular training of voice signal spectrograms and familiar listening trials. The updates of the SAL and RA assessments are précised by using the minima standards within a limited period frame. Hence, its reaction is not subtle with extent of unwanted transient variations. These algorithms are employed on voice signals tainted with 5 types of transients. The outcomes are valued using a joint impartial portion - the Sectional SNR (Sec. SNR) and concise in Table 2. It has been shown from result, the SAL unwanted transient assessment permits to defeat better transitory interventions and recovers greatly the Sec SNR of the voice signal as compared to RA unwanted transient assessment. We remark the size of the frame is empirically fixed to each variation of transitory to produce greatest performance; the obtained Sec SNR for individual diversity of transitory is sedate for different sizes of frames. Then, the interval that produces the maximal Sec SNR is fixed. The resolute extents of the frame structures are

prevailing in Table 2. Transient interferences of several varieties, intervals, and spectral structures are reduced willy-nilly of its locality nearby the voice signal and willy-nilly of the individual reciter.

7. CONCLUSION

Iterative mean is an ordinarily jumble-sale way for assessment of unwanted transient power spectrum. Nevertheless, computing a subjective mean centred on the rapid spectral extents of the ruined voice signal and assessment unwanted transients, both the methods use the levelling constraint in both time and the frequency in accordance with voice signal existence likelihood. The voice signal existence likelihood is manipulated through bare minimum standards of a levelled periodogram of the unwanted transient measurement. The proposed method shows approximately the 12-15 percent improvement in the Segmental SNR for the signals to be considered in the analysis .Equated to the RA unwanted transient assessment and the SAL assessment methods the response of the min-max method additionally rapid to unwanted transient deviations and, when assimilated into a voice signal enrichment scheme, yields good sectional SNR. The Implementation of the Algorithm is not considering the starting and final points in to the calculation. It can be enhanced in the future for up gradation of the method.

REFERENCES

[1]   R. Talmon, I. Cohen, S. Gannot, and R. R. Coifman "Graph-Based Bayesian Approach For Transient Interference Suppression" 2013.

[2]   D. Malah, R. V. Cox, and A. J. Accardi, "Tracking speech-presence uncertainty to improve speech enhancement in nonstationary noise environments"in Proc. 24th IEEE ICASSP'99, Phoenix, AZ, Mar. 15–19,1999, pp. 789–792.

[3]   H. G. Hirsch and C. Ehrlicher, "Noise estimation techniques for robustspeech recognition," in Proc. 20th IEEE ICASSP'95, Detroit, MI, May8–12, 1995, pp. 153–156.

[4]   R. J. McAulay and M. L. Malpass, "Speech enhancement using a softdecisionnoise suppression filter," IEEE Trans. Acoust., Speech, SignalProcessing, vol. ASSP-28, pp. 137–145, Apr. 1980.

[5]   V. Stahl, A. Fischer, and R. Bippus, "Quantile based noise estimationfor spectral subtraction and Wiener filtering," in Proc. 25th IEEEICASSP'2000, Istanbul, Turkey, June 5–9, 2000, pp. 1875–1878.

[6]   R. Martin, "Spectral subtraction based on minimum statistics," in Proc.7th EUSIPCO'94, Edinburgh, U.K., Sept. 13–16, 1994, pp. 1182–1185.

[7]   I. Cohen and B. Berdugo, "Speech enhancement for nonstationary noiseenvironments," Signal Process., vol. 81, pp. 2403–2418, Nov. 2001.

[8]   G. Doblinger, "Computationally efficient speech enhancement byspectral minima tracking in subbands," in Proc. 4th EUROSPEECH'95,Madrid, Spain, Sept. 18–21, 1995, pp. 1513–1516.

[9]   A. Varga and H. J. M. Steeneken, "Assessment for automatic speechrecognition: II. NOISEX-92: A database and an experiment to studythe effect of additive noise on speech recognition systems," SpeechCommun., vol. 12, no. 3, pp. 247–251, July 1993.

[10] S. Quackenbush, T. Barnwell, and M. Clements, Objective Measures ofSpeech Quality. Englewood Cliffs, NJ: Prentice-Hall, 1988.

[11] Israel Cohen, and Baruch Berdugo "Noise Estimation by Minima Controlled Recursive Averaging for Robust Speech Enhancement" IEEE Signal Processing Letters, Vol. 9, No. 1, January 2002.

[12] I Cohen and B. Berdugo, "Speech enhancement for non stationary noise environments," Signal Process., vol. 81, no. 11, pp. 2403–2418, Nov. 2001.

[13] I. Cohen, "Noise spectrum estimation in adverse environments: improved minima controlled recursive averaging," IEEE Trans. Speech, Audio Process., vol. 11, no. 5, pp. 466–475, Sep. 2003.

[14] R. Martin, "Noise power spectral density estimation based on optimal smoothing and minimum statistics," IEEE Trans. Speech and Audio Process., vol. 9, pp. 504–512, Jul. 2001.

[15] Ariel Hirszhorn, David Dov, Ronen Talmon and Israel CohenTransient Interference Suppression In Speech Signals Based On The OM-LSA AlgorithmInternational Workshop on Acoustic Signal Enhancement 2012, 4-6 September 2012.

[16] [Online]. Available: http://www.freesound.org.

[17] N. S. Kim and J.-H. Chang, "Spectral enhancement based on global soft decision," IEEE Signal Processing Lett. vol. 7, pp. 108–110, May 2000.

AUTHORS

Dr. AJAY SOMKUWAR

Designation: Professor, MANIT Bhopal

Qualification: B.E (Hons), M. Tech, Ph.D (IIT, Delhi)

Research Interest: Signal Processing, Bio Signal Processing, Statistical Signal Processing

PUSHPRAJ TANWAR Received MTech degree from MANIT Bhopal in 2009. He is currently pursuing the Ph.D. degree in Electronics engineering at MANIT. Bhopal. His research interest is signal processing, analysis of acoustic signals and speech enhancement.