Real Time Noise Suppression in Social Settings Comprising a Mixture of Non-stationary and Transient Noise

Pei Chee Yong, Sven Nordholm

Department of Electrical and Computer Engineering
Curtin University, Kent Street, Bentley, WA 6102, Australia
P.Yong@curtin.edu.au, S.Nordholm@curtin.edu.au

***Abstract*—Hearable is a recently emerging term that describes a wireless earpiece that enhances the user's listening experience in various acoustic environment. Another important feature of hearable devices is their capability to improve speech communication in difficult social settings, which usually consist of a mixture of different non-stationary noise. In this paper, we present techniques to suppress a combination of non-stationary noise and transient noise. This is achieved by employing a combined noise suppression filter based on prediction and masking to achieve impulsive noise suppression. Experimental results highlight the robustness of the proposed algorithm in suppressing the transient noise while maintaining the speech components, without requiring any prior information of the noise.**

I. Introduction

For wearable ear-mounted listening devices, recently termed hearables, speech enhancement is one of the essential modules required to improve the quality of the speech signals from the external acoustic environment that are often contaminated by different types of noise and interference. With the space and power constraints in the earbuds, single microphone noise reduction systems remain the preferred framework over the multi-channel structures, with only the spectral-temporal structure of the signals being exploited. Even when multi microphones setups are used, a well-formed single-channel method can serve as a post filter to further suppress unwanted noise and to improve the speech signal-to-noise ratio (SNR) [1]–[4].

Numerous single-channel speech enhancement algorithms have been developed over the decades aiming at estimating the power spectrum of the background noise and obtaining the desired clean speech signal estimate [2], [5]–[8]. In particular, these approaches work well when the power spectral density (PSD) of the noise signal during the observation time interval is more stationary than the speech. A common practice for estimating the noise PSD is to recursively average the noisy observation in short-time intervals by using an estimation of speech presence probability (SPP) [9]–[12]. The computation of SPP is however mainly based on the estimation of SNR, which is often inadequate to distinguish speech from noise in environment with highly non-stationary and transient noise such as restaurant, office or worksite. In these environments some noise components may vary even faster than the speech signal.

Due to the sparse characteristics of the transient noise in the time signal, several time domain algorithms have been developed to identify and remove transient noise, which include threshold-based approaches [13], [14] and statistical-based approaches [15], [16]. These time domain methods produce sample by sample based transient noise detection and apply identical suppression weight to all frequencies. In order to provide better detection, time-frequency domain methods have been proposed to exploit the spectral-temporal characteristics of speech and transient noise [17], [18]. However, these algorithms do not provide information about the position of the transient within the observation interval. This can be improved by reducing the frame size, which increases the time resolution, but lowers the frequency resolution. Alternatively, wavelet-based [19] and phase-based detections [20] have also been studied to exploit more properties of the transient noise. Another group of research focused on developing supervised transient noise reduction methods, where speech enhancement is done by utilising the noise learnt from training datasets [21], [22]. This type of processing requires prior information such as the repetition frequencies of the transient noise to achieve the desired performance.

In this paper, we present an algorithm that suppresses transient interferences for speech enhancement particularly for social settings. The algorithm mainly consists of three stages: (1) a linear prediction procedure to enhance the difference between transient noise and other signal components, (2) a speech masking threshold based on the predicted signal, and (3) a noise PSD estimation function that differentiates the transient noise from the more-stationary background noise. The transient noise suppression gain function is then applied to a speech enhancement framework as shown in Fig. 1, based on the structure in [8]. Experimental results show that the proposed algorithm is capable of tracking and suppressing the transient noise, which enables a similar speech quality and maintains the essence of the speech intelligibility when compared to the approach without the transient noise suppression. The paper also demonstrates that the proposed algorithm

ISBN 978-0-9928626-7-1

does not require any prior knowledge about the temporal or spectral structure of the transient noise, and is suitable for on-line hearable applications.

The remainder of this paper is organized as follows. In section II, the signal model of a single channel speech enhancement framework is formulated. Section III demonstrates the proposed algorithm. Section IV presents the graphical and objective experimental results and Section V concludes the paper.

II. Signal model

Let the observed noisy signal be expressed in discrete-time domain as

$$y(n) = x(n) + v(n) \tag{1}$$

where $x(n)$ is the clean speech signal and $v(n) = t(n)+\nu(n)$ contains the additive highly non-stationary transient noise $t(n)$ and the background noise $\nu(n)$ with relatively less time-varying statistics. By using the short-time Fourier transform (STFT), the spectral coefficients of the observed signal $Y(k,m)$ can be obtained by

$$Y\left(k,m\right) = \sum_{n=1}^{N} y\left(mR+n\right) w_a\left(n\right) \exp\left(\frac{-j2\pi kn}{N}\right) \tag{2}$$

where $k = [1,...,K]$ is the frequency bin index, $m = [1,...,M]$ is the frame index, R is the STFT frame rate and $w_a(n)$ is an analysis window function. The observed signal in Eq. (2) can be written as

$$Y(k,m) = X(k,m) + T(k,m) + \mathcal{V}(k,m) \tag{3}$$

where $X(k,m)$, $T(k,m)$ and $\mathcal{V}(k,m)$ represent the STFTs of $x(n)$, $t(n)$ and $\nu(n)$, respectively. Assume that all components in Eq. (3) are uncorrelated with each other, the power spectral density (PSD) of the observed signal can be defined as

$$\lambda_y(k,m) = \lambda_x(k,m) + \lambda_t(k,m) + \lambda_\nu(k,m) \tag{4}$$

where

$$\begin{aligned} \lambda_x(k,m) &= \mathbb{E}\left\{|X(k,m)|^2\right\}, \\ \lambda_t(k,m) &= \mathbb{E}\left\{|T(k,m)|^2\right\}, \\ \lambda_\nu(k,m) &= \mathbb{E}\left\{|\mathcal{V}(k,m)|^2\right\} \end{aligned} \tag{5}$$

denote the periodograms of the clean speech, the transient noise and the background noise, respectively.

III. Proposed algorithm

A. Transient noise estimation

The first stage of the proposed algorithm is to distinguish the difference between the transient noise and speech from the observed signal. Consider an auto-regressive (AR) model for the speech signal $x(n)$ as defined by

$$x(n) = \sum_{l=1}^{L} \alpha_l x(n-l) + w(n) \tag{6}$$

where $\{\alpha_l\}_{l=1,...,L}$ are L AR parameters and $w(n)$ is a zero-mean white noise excitation signal with σ_w^2 variance. The value of the parameter L has to be large enough to represent both

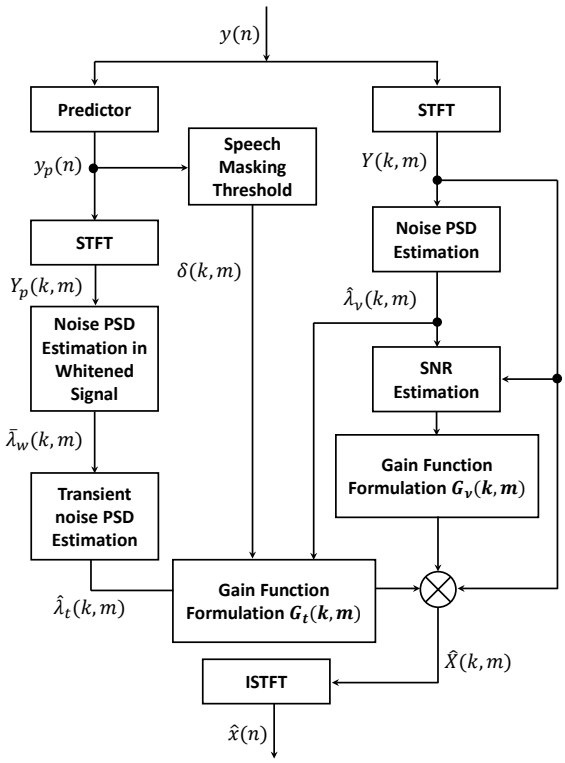

Fig. 1. Block diagram of transient and background noise suppression framework.

voiced and unvoiced phonemes [21]. As the speech signal can be viewed as a periodic and rather stationary signal in a short time interval, linear prediction can be used to predict the speech from the observed signal $y(n)$. Let $y_p(n)$ be the whitened signal obtained from

$$y_p(n) = y(n) - \sum_{l=1}^{L} \alpha_l y(n-l), \tag{7}$$

which produces the background noise $t(n)$, excitation signal $w(n)$ and the residual speech. In order to reduce the amount of speech in the linear prediction, a lattice filter is used to improve the estimation accuracy of the vocal tract filter. The structure of a lattice filter consists of a forward prediction error $f_i(n)$ and a backward prediction error $b_i(n)$, which are given, respectively, by

$$\begin{aligned} f_i(n) &= f_{i-1}(n) + \kappa_i(n) b_{i-1}(n-1) \\ b_i(n) &= b_{i-1}(n-1) + \kappa_i(n) f_{i-1}(n). \end{aligned} \tag{8}$$

The reflection coefficient in the lattice filter $\kappa_i(n)$ is updated by using Burg's algorithm as defined by

$$\kappa_i(n) = \frac{n_i(n)}{d_i(n)} \tag{9}$$

where

$$\begin{aligned} d_i(n) &= \lambda_{\mathrm{p}} d_i(n-1) + (1-\lambda_{\mathrm{p}}) \left[f_{i-1}^2(n) + b_{i-1}^2(n-1)\right] \\ n_i(n) &= \lambda_{\mathrm{p}} n_i(n-1) + (1-\lambda_{\mathrm{p}}) \left(-2\right) \left[f_{i-1}(n) b_{i-1}(n-1)\right]. \end{aligned} \tag{10}$$

ISBN 978-0-9928626-7-1 © EURASIP 2017

In this work, the I-th tap forward prediction residual $f_I(n)$ is used as the whitened signal $y_p(n)$.

The next step is to to further exploit the character difference between the transient noise and the residual by utilising the spectral-temporal features of the transient. The transient noise PSD estimate can be computed by employing a spectral gain function to the STFT of $y_p(n)$, as given by

$$\hat{\lambda}_t(k,m) = G_{\rm ss}(k,m) Y_p(k,m) \tag{11}$$

where

$$G_{\rm ss}(k,m) = 1 - \beta_{\rm ss} \frac{\bar{\lambda}_w(k,m)}{\bar{\lambda}_{yp}(k,m)} \tag{12}$$

is the spectral subtraction function with over-subtraction factor $\beta_{\rm ss}$, with $\bar{\lambda}_w(k,m)$ and $\bar{\lambda}_{yp}(k,m)$ denote the smoothed periodograms of the residual noise and predicted signal, respectively. The smoothed periodogram of the whitened signal can be obtained by a first-order recursive averaging of the spectral amplitude as follows

$$\bar{\lambda}_{yp}(k,m) = \alpha_{yp} \bar{\lambda}_{yp}(k,m-1) + (1-\alpha_{yp}) |Y_p(k,m)|^2 \tag{13}$$

where $\alpha_{yp} = \exp\left((-2.2R)/(t_{yp} f_{\rm s})\right)$ denotes the smoothing factor, with $f_{\rm s}$ denotes the sampling frequency. By assigning a smaller value to α_{yp} leads to better capability to capture faster PSD variations of the observation.

The periodogram of the residual noise is obtained by

$$\begin{aligned} \bar{\lambda}_w(k,m) = \; & p(k,m)\, \bar{\lambda}_w(k,m-1) \\ & + (1 - p(k,m)) |Y_p(k,m)|^2 \end{aligned} \tag{14}$$

where $p(k,m)$ denotes the transient presence probability (TPP). The TPP is estimated by using a soft decision based estimator with conditional averaging smilar to [11], as given by

$$p(k,m) = \begin{cases} \mathcal{P}_1, & \text{if } p_{\rm s}(k,m) \le 0.3 \\ \mathcal{P}_2, & \text{if } 0.3 < p_{\rm s}(k,m) \le 0.6 \\ \mathcal{P}_3, & \text{if } 0.6 < p_{\rm s}(k,m) \le 0.8 \\ \mathcal{P}_4, & \text{if } p_{\rm s}(k,m) > 0.8 \end{cases} \tag{15}$$

where $\mathcal{P}_i = \exp\left((-2.2R)/(t_i f_{\rm s})\right)$ indicates the exponential smoothing constant, with $i = [1,2,3,4]$, and $p_{\rm s}(k,m)$ denotes a sigmoid function given by

$$p_{\rm s}(k,m) = \left\{1 + \exp\left(-\sigma_{\rm post}\left(\hat{\gamma}_t(k,m) - \epsilon_{\rm post}\right)\right)\right\}^{-1} \tag{16}$$

where $\hat{\gamma}_t(k,m) = |Y_p(k,m)|^2 / \bar{\lambda}_w(k,m)$ denotes the estimate of the *a posteriori* SNR, while $\sigma_{\rm post}$ and $\epsilon_{\rm post}$ denote the slope and the mean of the sigmoid curve, respectively. Both the slope and the mean in [11] are computed based on the *a priori* speech presence uncertainty estimate and the SNR estimate. In this paper, the noise PSD estimate $\bar{\lambda}_w(k,m)$ tracks only the more stationary noise instead of the short bursts of transient interference. Thus, the values of $\sigma_{\rm post}$ and $\epsilon_{\rm post}$ are chosen such that the TPP estimate is quick enough to track the variation of the transient noise.

B. Speech enhancement with masking

The aim of speech enhancement in this paper is to obtain the clean speech spectrum estimate $\hat{X}(k,m)$ from the observed signal $Y(k,m)$, which is given by

$$\hat{X}(k,m) = G(k,m) Y(k,m) \tag{17}$$

where $G(k,m) = G_t(k,m) G_\nu(k,m)$ is a multiplicative nonlinear gain function consists of a gain function $G_t(k,m)$ for transient noise suppression and a gain function $G_\nu(k,m)$ mapped with the *a priori* SNR estimate or the *a posteriori* SNR estimate $\hat{\gamma}_\nu(k,m)$. The gain function can usually be optimally derived in the MMSE sense [2], [6]. As an alternative gain function a modified sigmoid (MSIG) function [8] has been used in this work, which is given by

$$\begin{aligned} G_\nu(k,m) = & \frac{1-\exp\left[-a_1 \hat{\xi}_\nu(k,m)\right]}{1+\exp\left[-a_1 \hat{\xi}_\nu(k,m)\right]} \\ & \times \frac{1}{1+\exp\left(-a_2\left[\hat{\xi}_\nu(k,m) - c\right]\right)} \end{aligned} \tag{18}$$

where a_1, a_2 and c are parameters to control the shape of the sigmoid curve. The *a priori* SNR estimate $\hat{\xi}_\nu(k,m)$ and the noise PSD estimate $\hat{\lambda}_\nu(k,m)$ are obtained from [8] and [11], respectively. The objective of the transient noise suppression is to reduce the power of the transient interferences in the noisy speech without introducing audible speech and noise distortions. However, the transient noise PSD estimate $\hat{\lambda}_t(k,m)$ contains the speech residual which may result in speech distortion after suppression. With this in mind, a speech masking threshold and a spectral gain function are proposed to suppress the transient noise in the noisy signal by utilising a noise-to-transient ratio (NTR). The gain function can be written as

$$G_t(k,m) = \min\left\{ \frac{\hat{\lambda}_\nu(k,m) + \delta(k,m)}{\hat{\lambda}_\nu(k,m) + \beta_t \hat{\lambda}_t(k,m)},\; 1 \right\} \tag{19}$$

where β_t denotes a transient noise suppression weight and $\delta(k,m)$ denotes the speech masking threshold that masks the residual speech components at higher frequencies in $\hat{\lambda}_\nu(k,m)$ with a frequency dependent floor by utilising the variance of the whitened signal and a first-order low pass filter. The gain function takes a value close to 0 when $\hat{\lambda}_t(k,m)$ is larger than $\hat{\lambda}_\nu(k,m)$, indicating that transient noise with high volume being suppressed.

Finally, the enhanced speech signal $\hat{x}(n)$ is obtained by using an inverse STFT to transform $\hat{X}(k,m)$ back to the time domain.

IV. Experimental results

In this section, the performance evaluation was done for the aforementioned speech enhancement framework with and without the proposed transient noise suppression algorithm, defined as MSIG-PRED and MSIG, respectively. The parameters for the algorithms were selected based on empirical studies as follows: for prediction, $I = 25$, $\lambda_{\rm p} = 1 - (1/160)$; for transient noise PSD estimation, $\beta_{\rm ss} = 1.3$, $t_{yp} = 0.01$s, $t_1 = 0.8$, $t_2 = 3$, $t_3 = 6$, $t_4 = 12$, $\sigma_{\rm post} = 6$, $\epsilon_{\rm post} = 1.5$;

ISBN 978-0-9928626-7-1 © EURASIP 2017

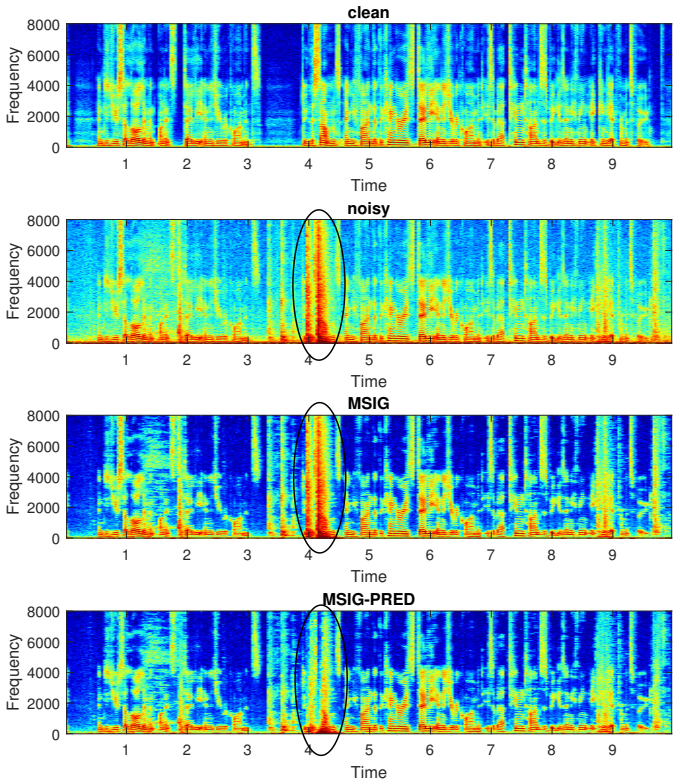

Fig. 2. Spectrograms of clean speech, noisy speech, and enhanced signals, with door closing interference at 0 dB SNR. The figure highlights the transient noise suppression at around 4.2 second using the proposed method.

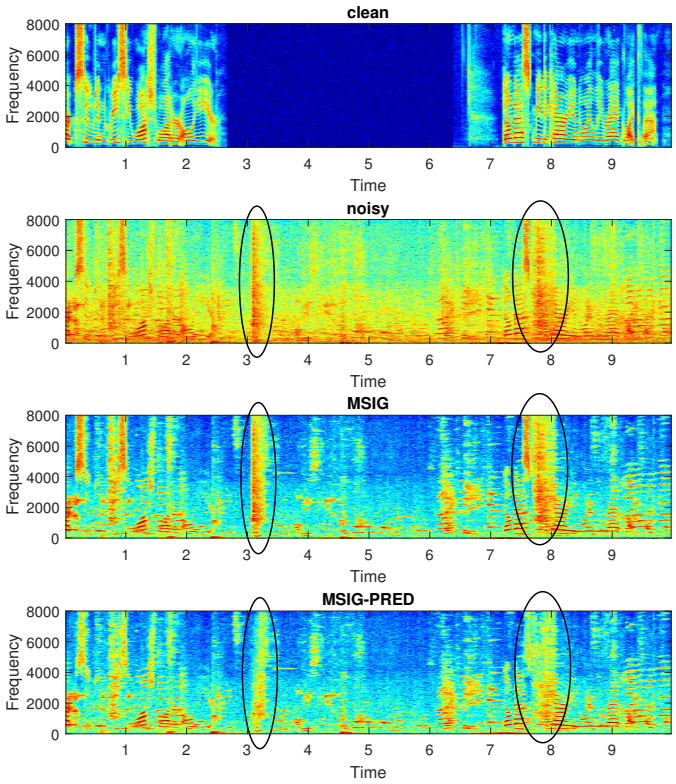

Fig. 3. Spectrograms of clean speech, noisy speech, and enhanced signals in a noisy cafeteria at 0 dB SNR. The proposed method reduced the impulsive noises while maintaining the harmonic structure of the speech.

for speech enhancement, $a_1 = 2.4$, $a_2 = 0.2$, $c = -1.7$, and $\beta_t = 1$. For objective measurement, the speech sequences were taken from NOIZEUS speech database, which contains 30 English sentences recorded from 3 male and 3 female speakers [2]. The evaluated noise was a recorded cafeteria noise comprising a mixture of non-stationary and transient noise. The signals were all sampled at $f_s = 16$ kHz. All speech utterances were contaminated by the noise with 4 levels of SNRs, -5 dB, 0 dB, 5 dB, and 10 dB. The results were generated with a square-root Hanning window and $K = 512$ frequency bins. Performance evaluation was done using the intrusive perceptual evaluation of speech quality (PESQ) measure [23] and short-time objective intelligibility (STOI) measure [24], where the former evaluates the speech quality from a score 0 to 4.5 and the latter rates the speech intelligibility from 0 to 1.

Figs. 2 and 3 depict the spectrograms of the noisy signals in two real-time social scenarios. Fig. 2 illustrates a speech sequence produced in a room with a door closing sound occurred at time instance around 4.2 second. It can be seen that MSIG-PRED was able to suppress the transient noise and maintain the speech components, while MSIG treated the sound as speech onsets. A more complicated noisy scenario has been shown in Fig. 3, which was recorded in a cafeteria with various non-stationary noise signals and transient noise. The figure shows that the proposed algorithm was capable of suppressing the banging sounds happening in the background of a social settings while maintaining the integrity of the speech components. This is an important features for a hearable device to preserve the speech and to prevent the transient noise from being accentuated after speech enhancement.

The objective measurement evaluates the noisy scenario as illustrated in Fig. 3 with speech sequences from NOIZEUS database. Fig. 4 shows the results of both the PESQ scores and the STOI scores for all the evaluated algorithms. It can be observed that both MSIG-PRED and MSIG have similar PESQ and STOI scores over the evaluated input SNRs. This indicates that the proposed transient noise suppression algorithm reduces the impact of the transient noise without affecting the quality and intelligibility of the speech. However, while the two processing methods improve the speech quality, they both lower the speech intelligibility. The benefit provided by the proposed transient noise suppression is that it does not reduce the intelligibility further.

V. Conclusion

To conclude, an algorithm for transient noise suppression for speech enhancement is proposed. An adaptive linear prediction based on Burg's lattice algorithm is firstly utilised to enhance the transient noise from the speech components. Second, the power spectral density (PSD) of the enhanced transient noise is estimated by tracking and suppressing the residual noise with a soft-decision based estimator. A speech masking

ISBN 978-0-9928626-7-1 © EURASIP 2017

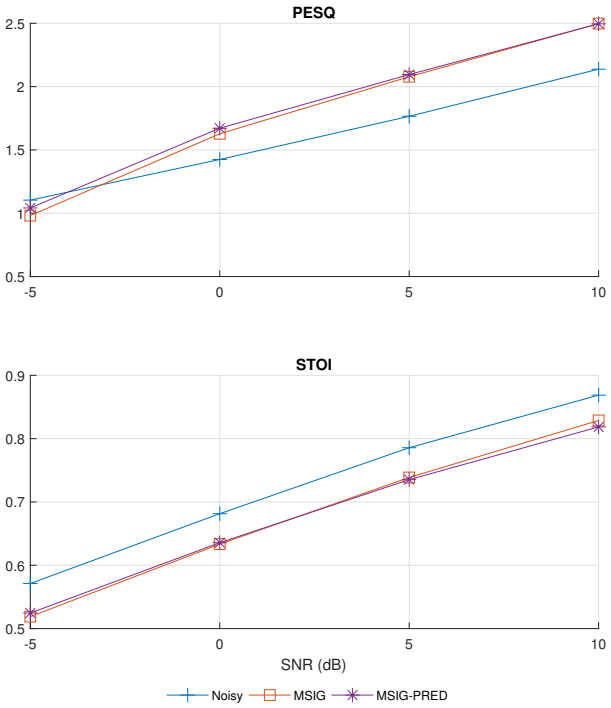

Fig. 4. Objective measurement with PESQ and STOI.

threshold is then utilised to avoid the suppression of the speech components at high frequencies. This filter is employed in a typical speech enhancement framework to realise a complete noise reduction scheme. Experimental results show that the proposed algorithm is capable of suppressing different types of transients, without affecting the speech. Based on the two examples shown, the proposed method reduced the PSD of the transients with low impact to the speech. This is supported by both objective measures, PESQ and STOI, which evaluate the speech quality and intelligibility, respectively. The algorithm also demonstrates its capability to be implemented for real-time applications without prior knowledge about the time position of the transient noise.

ACKNOWLEDGMENT

This research was sponsored by Nuheara through a research grant.

REFERENCES

[1] Joerg Meyer and Klaus Uwe Simmer, "Multi-channel speech enhancement in a car environment using wiener filtering and spectral subtraction," in *IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, Munich, Bavaria, Germany, May 1997, vol. 2, pp. 1167–1170.

[2] Philipos C Loizou, *Speech Enhancement Theory and Practice*, CRC Press, 2007.

[3] Pei Chee Yong, Sven Nordholm, and Hai Huyen Dam, "Effective binaural multi-channel processing algorithm for improved environmental presence," *IEEE/ACM Transactions on Audio, Speech, and Language Processing*, vol. 22, no. 12, pp. 2012–2024, 2014.

[4] Sven Nordholm, Alan Davis, Pei Chee Yong, and Hai Huyen Dam, "Assistive listening headsets for high noise environments: Protection and communication," in *IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, Brisbane, Queensland, Australia, April 2015, pp. 5753–5757.

[5] Steven Boll, "Suppression of acoustic noise in speech using spectral subtraction," *IEEE Transactions on acoustics, speech, and signal processing*, vol. 27, no. 2, pp. 113–120, 1979.

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

[7] Israel Cohen and Baruch Berdugo, "Speech enhancement for non-stationary noise environments," *Signal processing*, vol. 81, no. 11, pp. 2403–2418, 2001.

[8] Pei Chee Yong, Sven Nordholm, and Hai Huyen Dam, "Optimization and evaluation of sigmoid function with a priori SNR estimate for real-time speech enhancement," *Speech Communication*, vol. 55, no. 2, pp. 358–376, 2013.

[9] Israel Cohen, "Noise spectrum estimation in adverse environments: Improved minima controlled recursive averaging," *IEEE Transactions on Speech and Audio Processing*, vol. 11, no. 5, pp. 466–475, 2003.

[10] Timo Gerkmann and Richard C Hendriks, "Unbiased MMSE-based noise power estimation with low complexity and low tracking delay," *IEEE Transactions on Audio, Speech, and Language Processing*, vol. 20, no. 4, pp. 1383 – 1393, 2012.

[11] Pei Chee Yong, Sven Nordholm, and Hai Huyen Dam, "Noise estimation based on soft decisions and conditional smoothing for speech enhancement," in *International Workshop on Acoustic Signal Enhancement (IWAENC)*, Aachen, Germany, September 2012.

[12] Pei Chee Yong and Sven Nordholm, "An improved soft decision based noise power estimation employing adaptive prior and conditional smoothing," in *International Workshop on Acoustic Signal Enhancement (IWAENC)*, Xi'an, China, September 2016, pp. 1–5.

[13] SV Vaseghi and PJW Rayner, "Detection and suppression of impulsive noise in speech communication systems," *IEE Proceedings I-Communications, Speech and Vision*, vol. 137, no. 1, pp. 38–46, 1990.

[14] Charu Chandra, Michael S Moore, and Sanjit K Mitra, "An efficient method for the removal of impulse noise from speech and audio signals," in *IEEE International Symposium on Circuits and Systems (ISCAS)*, Monterey, CA, USA, May 1998, vol. 4, pp. 206–208.

[15] Simon J Godsill and Peter JW Rayner, "A bayesian approach to the restoration of degraded audio signals," *IEEE Transactions on Speech and Audio Processing*, vol. 3, no. 4, pp. 267–278, 1995.

[16] James Murphy and Simon Godsill, "Joint bayesian removal of impulse and background noise," in *IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, Prague, Czech Republic, May 2011, pp. 261–264.

[17] SV Vaseghi and R Frayling-Cork, "Restoration of old gramophone recordings," *Journal of the Audio Engineering Society*, vol. 40, no. 10, pp. 791–801, 1992.

[18] Amarnag Subramanya, Michael L Seltzer, and Alex Acero, "Automatic removal of typed keystrokes from speech signals," *IEEE Signal Processing Letters*, vol. 14, no. 5, pp. 363–366, 2007.

[19] Rajeev C Nongpiur, "Impulse noise removal in speech using wavelets," in *IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, Las Vegas, NV, USA, May 2008, pp. 1593–1596.

[20] Akihiko Sugiyama and Ryoji Miyahara, "Tapping-noise suppression with magnitude-weighted phase-based detection," in *IEEE Workshop on Applications of Signal Processing to Audio and Acoustics (WASPAA)*, New Paltz, New York, USA, October 2013, pp. 1–4.

[21] Ronen Talmon, Israel Cohen, and Sharon Gannot, "Transient noise reduction using nonlocal diffusion filters," *IEEE Transactions on Audio, Speech, and Language Processing*, vol. 19, no. 6, pp. 1584–1599, 2011.

[22] Ronen Talmon, Israel Cohen, and Sharon Gannot, "Single-channel transient interference suppression with diffusion maps," *IEEE transactions on audio, speech, and language processing*, vol. 21, no. 1, pp. 132–144, 2013.

[23] Antony W Rix, John G Beerends, Michael P Hollier, and Andries P Hekstra, "Perceptual evaluation of speech quality (PESQ)-a new method for speech quality assessment of telephone networks and codecs," in *IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, Salt Lake City, Utah, USA, May 2001, vol. 2, pp. 749–752.

[24] Cees H Taal, Richard C Hendriks, Richard Heusdens, and Jesper Jensen, "An algorithm for intelligibility prediction of time–frequency weighted noisy speech," *IEEE Transactions on Audio, Speech, and Language Processing*, vol. 19, no. 7, pp. 2125–2136, 2011.

ISBN 978-0-9928626-7-1 © EURASIP 2017