A NEGENTROPY BASED ADAPTIVE LINE ENHANCER FOR SINGLE-CHANNEL NOISE REDUCTION AT LOW SNR CONDITIONS

Jalal Taghia and Rainer Martin

Institute of Communication Acoustics, Ruhr-Universität Bochum, Bochum, Germany
Email: {Jalal.Taghia,Rainer.Martin}@rub.de

ABSTRACT

In this paper, we propose an adaptive line enhancer based on negentropy for single-channel noise reduction. Our proposed approach can be integrated in a speech enhancement system as a preprocessor to be combined with other noise reduction approaches. The proposed method performs the noise reduction by splitting the noisy speech components into the deterministic and the stochastic parts through the minimization of negentropy in an adaptive manner. We consider the negentropy as a cost function, and we derive a learning rule via Newton's method to minimize the negentropy of the error signal. By the experimental results, we demonstrate that exploiting the proposed approach can be potentially useful as a preprocessor for improving the performance of conventional single-channel noise reduction approaches at low signal-to-noise ratio (SNR) conditions. Moreover, it is shown that our approach by itself can also enhance the noisy speech in an adverse noisy environment.

Index Terms**—** Single-channel noise reduction, speech enhancement, adaptive line enhancer, negentropy

1. INTRODUCTION

Single-channel noise reduction algorithms aim to improve both quality (e.g. speech pleasantness and naturalness) and intelligibility of noisy speech signals in many speech communication systems. Although single-channel noise reduction approaches may improve the speech quality, some recent studies (e.g. [1], [2]) show that most of these algorithms such as spectral subtractive, sub-space, statistical model based and Wiener-type algorithms [3] cannot provide significant improvements in speech intelligibility. The performance of these algorithms highly depends on the accurate estimation of the background noise statistics and the estimation of the *a priori* signal-to-noise ratio (SNR). Thus, it is of great interest to design novel single-channel noise reduction schemes which can improve both the quality and intelligibility of noisy speech in adverse environments (involving non-stationary noise and low input SNR conditions).

In this paper, we introduce a new approach for single-channel noise reduction at low SNR conditions which does not require knowledge of the noise statistics. Our approach is able to enhance the noisy speech as a stand-alone method, and moreover it can be regarded as a preprocessor to be combined with a conventional single-channel noise reduction system. The proposed approach incorporates the negentropy [4], [5] in an adaptive line enhancer (ALE) [6]. As shown in Fig. 1, typically in ALE, a filter is adapted based on minimizing a cost function of the error (residual) signal $e(n) = x(n) - \tilde{x}(n)$ where $x(n)$ is a discrete time domain signal at the input, and $\tilde{x}(n)$ is the filtered signal. Traditionally, the mean square error (MSE) is the criterion which is chosen in ALE as the cost function to be minimized [7], [8], [9].

In the work that we present here, the negentropy is exploited as a cost function of the error signal to split the noisy speech into the deterministic and the stochastic parts. The deterministic part, which mainly includes the stationary and periodic components of noisy speech, appears in the filtered signal while the error signal contains the stochastic (or fluctuating) part, and it is more random than the filtered signal. The negentropy is an information-theoretic criterion which has been widely used as a contrast function for independent component analysis (ICA) e.g., [4], [5], and [10]. On the other hand, in novel applications, the negentropy has been considered as a criterion for measuring the randomness of signals in the voice-activity detection [11] or as a non-Gaussianity criterion in beamforming [12]. For example, in [11], it is argued that since the segments of noisy speech which contain speech are relatively more structured than the noise-only (or speech-pause) segments, the negentropy of these frames (i.e., speech-active segments) should be higher than the negentropy of noise-only frames.

In the proposed approach, the filter coefficients are found through a learning rule based on Newton's method to minimize the negentropy of the error signal in an adaptive manner. The minimization of negentropy implies that the error signal is more random than the filtered signal. In the experiments, we introduce scenarios in which our approach can be integrated for single-channel noise reduction. The evaluation framework is based on instrumental quality and intelligibility measures. We demonstrate that exploiting the proposed method can be potentially useful as a preprocessor for improving the performance of single-channel noise reduction approaches at low SNR conditions. Furthermore, it is shown that this approach by itself is able to enhance the noisy speech in an adverse noisy environment.

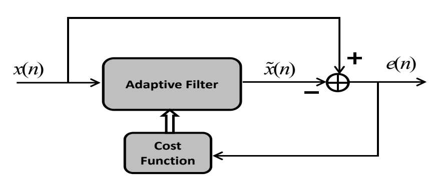

Fig. 1. The structure of a typical adaptive line enhancer (ALE).

2. PROPOSED APPROACH

Let $s(n)$ denote a clean speech signal which is degraded by an additive background noise $\xi(n)$ and produces the noisy speech signal $x(n)$. It is assumed that $s(n)$ and $\xi(n)$ are statistically independent. Our aim is to enhance the noisy speech $x(n)$ only based on the observed signal without the estimation of noise power spectral density (PSD).

The differential entropy is a fundamental information-theoretic concept for continuous-valued random variables. For a random vec-

978-1-4799-2893-4/14/$31.00 ©2014 IEEE

tor $\mathbf{x}$ with the probability density $f_x(.)$ the differential entropy h is defined as [13]

$$h(\mathbf{x}) = -\int f_x(\mathbf{x}) \, \log(f_x(\mathbf{x})) \, d\mathbf{x}. \tag{1}$$

In the concept of independent component analysis, the negentropy was introduced instead of differential entropy for minimization of mutual information [5], [14]. The negentropy (or the negative normalized entropy) J of a random vector $\mathbf{x}$ is defined as [5]

$$J(\mathbf{x}) = h(\mathbf{x}_{gauss}) - h(\mathbf{x}), \tag{2}$$

where $\mathbf{x}_{gauss}$ is a Gaussian random vector with the same covariance matrix as $\mathbf{x}$. The negentropy can be approximated using the cumulants of the probability distribution, however, it has been shown that the cumulant-based approximations are not robust [5]. For a random variable u, a robust approximation of its negentropy can be derived as follows [5], [14]

$$J(u) \propto \left[E\left\{G\left(u\right)\right\} - E\left\{G\left(\nu\right)\right\}\right]^2, \tag{3}$$

where $G(.)$ is a non-quadratic function, $E\left\{\cdot\right\}$ denotes the expectation operator, and ν denotes a normally distributed random variable of zero mean and unit variance. The approximation in (3) is robust and has a low computational complexity [14].

The negentropy has some interesting properties [5], [14]. Unlike the differential entropy, the negentropy is invariant for invertible linear transformations, and it is always positive. The negentropy can be interpreted as a measure of non-Gaussianity where it is zero for a normally distributed random variable. The important property of negentropy which we take into account in our approach is that it can be regarded as a criterion to measure the amount of structure in the probability distribution of a random variable. When a random variable has a sparse distribution or is clearly clustered, its distribution is concentrated on certain values, and hence its negentropy is higher than the negentropy of a random variable that is unpredictable and unstructured.

We incorporate the negentropy in an adaptive line enhancer (ALE) (shown in Fig. 1) as the cost function. Given the noisy speech $x(n)$ at the input of ALE, we minimize the negentropy to estimate the error signal $e(n)$ such that it mostly contains the stochastic (random) part of noisy speech $x(n)$ whereas the filtered signal $\tilde{x}(n)$ contains the deterministic (periodic) part. ALE predicts the input signal $x(n)$ as a linear combination of $\mathbf{x}(n) = [x(n-\tau), \ldots, x\,(n-\tau-L+1)]^{\mathrm{T}}$ such that the filtered signal $\tilde{x}(n)$ is derived as

$$\tilde{x}(n) = \mathbf{w}^{\mathrm{T}}(n)\mathbf{x}(n), \tag{4}$$

where T denotes the transpose operator, τ the delay parameter which is a significant parameter, L the number of samples in each segment (equal to the number of tap coefficients in the filter), and $\mathbf{w}(n) = [w_1(n), \ldots, w_L(n)]^{\mathrm{T}}$ denotes a vector of filter coefficients. Hence, the error signal $e(n)$ can be expressed as

$$\begin{aligned} e(n) &= x(n) - \tilde{x}(n) \\ &= x(n) - \mathbf{w}^{\mathrm{T}}(n)\mathbf{x}(n). \end{aligned} \tag{5}$$

Taking into account (3), (4), and (5), we estimate the filter $\mathbf{w}(n)$ in an adaptive manner by minimizing the negentropy of the error signal as follows

$$J_G(\mathbf{w}(n)) = \left[E\left\{G\left(x(n) - \mathbf{w}^{\mathrm{T}}(n)\mathbf{x}(n)\right)\right\} - E\left\{G\left(\nu\right)\right\}\right]^2, \tag{6}$$

where, similarly to (3), $G(.)$ is a non-quadratic function that in our work is chosen as

$$G(u) = -\frac{1}{a}\exp(\frac{-a\,u^2}{2}) \tag{7}$$

where $a \approx 1$. The cost function $J_G(\mathbf{w}(n))$ is convex (concave-up) and twice-differentiable with respect to $\mathbf{w}(n)$. Hence, we can use Newton's method to derive a learning rule for the minimization of the cost function (6). To obtain the learning rule based on Newton's method, we need to compute the first and the second derivatives of $J_G(\mathbf{w}(n))$ with respect to $\mathbf{w}(n)$ which are denoted by $J'_G(\mathbf{w}(n))$ and $J''_G(\mathbf{w}(n))$, respectively. Thus, we consider the following learning rule,

$$\mathbf{w}(n+1) \leftarrow \mathbf{w}(n) - \mu\frac{J'_G(\mathbf{w}(n))}{J''_G(\mathbf{w}(n))} \tag{8}$$

where μ is the learning rate (or the adaptation rate) parameter, and $J'_G(\mathbf{w}(n))$ is computed as

$$\begin{aligned} &J'_G(\mathbf{w}(n)) = -c\,\mathbf{x}(n)\,E\left\{G'(x(n) - \mathbf{w}^{\mathrm{T}}(n)\mathbf{x}(n))\right\}, \\ &c = 2\left(E\left\{G(x(n) - \mathbf{w}^{\mathrm{T}}(n)\mathbf{x}(n))\right\} - E\left\{G(\nu)\right\}\right). \end{aligned} \tag{9}$$

where $G'(.)$ is the first derivative of $G(.)$ defined in (7) i.e., $G'(u) = u\exp(\frac{-a\,u^2}{2})$. To simplify the expression (9), we assume that the term c is a constant which does not change the stationary points of the learning rule although its sign affects the stability of learning rule. A similar assumption was considered in [5] to derive the negentropy based adaptive neural algorithms for ICA. Following this assumption, we derive $J''_G(\mathbf{w}(n))$ as

$$J''_G(\mathbf{w}(n)) = c\,\mathbf{x}^{\mathrm{T}}(n)\mathbf{x}(n)\,E\left\{G''(x(n) - \mathbf{w}^{\mathrm{T}}(n)\mathbf{x}(n))\right\}, \tag{10}$$

where $G''(.)$ is the second derivative of $G(.)$ i.e., $G''(u) = \exp(\frac{-a\,u^2}{2}) - au^2\exp(\frac{-a\,u^2}{2})$. Finally, we obtain the following learning rule by replacing (9), (10) and (5) in (8),

$$\mathbf{w}(n+1) \leftarrow \mathbf{w}(n) + \mu\frac{\mathbf{x}(n)\,E\left\{G'(e(n))\right\}}{\mathbf{x}^{\mathrm{T}}(n)\mathbf{x}(n)\,E\left\{G''(e(n))\right\}}, \tag{11}$$

where the first moments in (11) are estimated by the exponential smoothing, i.e.,

$$\overline{G'(e(n))} = \alpha\,\overline{G'(e(n-1))} + (1-\alpha)\,G'(e(n)), \tag{12}$$

where $0 < \alpha < 1$ is the smoothing factor. Similarly, we estimate $E\left\{G''(e(n))\right\}$ according to (12).

As we mentioned before, the proposed adaptive filter separates the observed noisy speech $x(n)$ into the deterministic part which is included in the filtered signal $\tilde{x}(n)$ and the stochastic part which is included in the error signal $e(n)$. Minimization of negentropy guarantees that $e(n)$ is less structured than $\tilde{x}(n)$. Depending on the type of background noise, the target speech may be captured in $\tilde{x}(n)$ as the deterministic part or may be captured in $e(n)$ as the stochastic part. In other words, the separation of target speech from the noisy signal is performed with a permutation ambiguity. We resolve the aforementioned permutation problem by computing the kurtosis of both $e(n)$ and $\tilde{x}(n)$ and making the decision based on the fact that speech signals are highly super-Gaussian and generally their kurtosis is larger than that of the background noise. Thus, either the error signal or the filtered signal can be selected as the estimated target speech if its kurtosis is the largest.

Table 1. Performance of different speech enhancement scenarios (shown in Fig. 2) when clean speech is degraded by WGN-Harmonics at low SNR conditions -15, -10, -5, 0, and 5 dB.

Method	Measure	-15dB	-10dB	-5dB	0dB	5dB
Case 1	STOI-in	0.63	0.70	0.77	0.84	0.90
	STOI-out	0.79	0.87	0.92	0.95	0.97
	ΔPESQ	0.61	1.00	1.13	1.26	1.32
	ΔSSNR	10.4	11.6	12.1	11.8	11.0
Case 2	STOI-out	0.80	0.88	0.93	0.95	0.97
	ΔPESQ	0.21	0.62	0.79	0.94	1.02
	ΔSSNR	7.06	8.93	9.03	7.47	2.11
Case 3	STOI-out	0.84	0.90	0.94	0.96	0.97
	ΔPESQ	0.85	1.24	1.34	1.45	1.48
	ΔSSNR	11.7	12.6	11.7	9.60	3.02

Table 2. Performance of different speech enhancement scenarios (introduced in Fig. 2) when clean speech is degraded by traffic noise at low SNR conditions -15, -10, -5, 0, and 5 dB.

Method	Measure	-15dB	-10dB	-5dB	0dB	5dB
Case 1	STOI-in	0.41	0.51	0.62	0.74	0.84
	STOI-out	0.40	0.52	0.66	0.78	0.87
	ΔPESQ	-1.61	-0.09	0.11	0.36	0.44
	ΔSSNR	2.66	3.70	4.31	4.54	4.32
Case 2	STOI-out	0.47	0.57	0.67	0.75	0.80
	ΔPESQ	-1.28	0.05	0.17	0.15	0.10
	ΔSSNR	2.70	3.36	2.95	1.49	-0.88
Case 3	STOI-out	0.46	0.58	0.68	0.76	0.82
	ΔPESQ	-1.29	0.21	0.40	0.46	0.45
	ΔSSNR	4.90	5.23	4.41	2.54	0.01

However, it is noteworthy that according to our pilot experiments, the target speech is obtained in the error signal (i.e., the stochastic part) for different types of noise except the white Gaussian noise (WGN). In other words, we do not require making the decision based on kurtosis since the only exceptional case in which the target speech is found in the filtered signal (i.e., the deterministic part) is when the target speech is degraded by WGN.

3. EXPERIMENTAL RESULTS

In this paper, the number of tap coefficients L for the filter is set to 300 (or equivalently the segment length is 18.7 ms for signals sampled at 16 kHz sampling frequency). We set the delay parameter τ equal to 1 ($\tau = 1$) and the smoothing factor α equal to 0.3 ($\alpha = 0.3$) according to our pilot experiments. Later on, the proposed approach is indicated by "NegALE" in the paper.

The sampling frequency of all signals used in this work is 16 kHz. The clean speech signal has a total duration of 60 s taken from the TIMIT database [15] including two male and two female speech signals (of four different speakers) where each one has a duration of 15 s. In the experiments, two types of noise signals are considered. One is a synthetic noise which is a white Gaussian noise combined with some harmonics at six certain frequencies from 500 Hz to 3 kHz in 500 Hz steps (called WGN-Harmonics). The other one is a real-world noise in a traffic environment involving a lot of horn sounds (called traffic noise). To simulate low SNR conditions, we consider input overall SNRs from -15 to 5 dB in 5 dB steps.

Our approach can be used to enhance the noisy speech as a stand-alone method; furthermore, it can be considered as a preprocessor for a conventional single-channel noise reduction (SCNR) method. In Fig. 2, we show the usage of a conventional noise reduction method (indicated by SCNR) as well as two different scenarios which are introduced to employ our approach (NegALE). In this figure, Case 1 shows the usage of SCNR for enhancing the noisy speech $x(n)$ where the estimated target speech is denoted by $\hat{s}(n)$. For SCNR, we perform the spectral gain calculation in the frequency domain based on the decision-directed approach [16] (with smoothing factor equal to 0.9) and the minimum mean-square error log-spectral amplitude estimator (MMSE-LSA) [17]. The noise PSD is derived by the MMSE based noise power estimator [18] that has a robust performance according to the study in [19]. SCNR was implemented in a discrete Fourier transform (DFT)-based spectral analysis-synthesis system using overlapping square-root Hann windows. The window length as well as the DFT length are 512 samples (32ms) and the amount of overlap between the frames is 256 samples (16ms).

Case 2 shows the noise reduction only based on the proposed approach where the estimated target speech and the estimated noise are denoted by $\hat{s}(n)$ and $\hat{\xi}(n)$, respectively. As we mentioned in Section 2, the target speech depending on the type of background noise may be found in the error signal $e(n)$ (i.e., the stochastic part of noisy speech) or may be in the filtered signal $\tilde{x}(n)$ (i.e., the deterministic part of noisy speech). Our criterion for choosing the right part containing the target speech (i.e. resolving the permutation ambiguity) is based on the kurtosis of signals which in this work is computed by $kurtosis(.) - 3$ where $kurtosis(.)$ is the built-in function in MATLAB. However, as we mentioned it before, making the decision based on kurtosis for selecting the error or the filtered signal as the target speech is not necessarily required when WGN is not among the noise types which are considered in the experiments. In our experiments, for example since the target speech is degraded by either traffic noise or WGN-Harmonics, the target speech is always captured in the error signal (i.e., $\hat{s}(n) = e(n)$) as the stochastic part of noisy speech and the deterministic part of noise is captured in the filtered signal (i.e., $\hat{\xi}(n) = \tilde{x}(n)$) as the deterministic part of noisy speech.

Case 3 illustrates the proposed scenario of employing NegALE as a preprocessor for SCNR such that the noisy speech $x(n)$ is firstly processed by NegALE to separate the deterministic and stochastic parts of noisy speech. In this way, the signal $s_1(n)$ containing the target speech is provided at the input of SCNR for further enhancement. By applying the SCNR to the signal $s_1(n)$, the remaining part of noise can be suppressed to provide the final estimate of the target speech $\hat{s}(n)$.

We measure the performance of algorithms with respect to the speech quality enhancement by using the improvement in the segmental SNR [3] (indicated by ΔSSNR) and the improvement in Perceptual Evaluation of Speech Quality measure (denoted by ΔPESQ) as implemented in [3]. Furthermore, we evaluate the performance in terms of the speech intelligibility by means of the Short-Time Objective Intelligibility (STOI) measure [20]. The objective intelligibility scores for the unprocessed noisy speech $x(n)$ and the final processed speech $\hat{s}(n)$ are indicated by $\mathrm{STOI_{in}}$ and $\mathrm{STOI_{out}}$, respectively.

In the experiments, the suitable adaptation rate μ for NegALE is determined heuristically to separate deterministic noise components from the noisy speech while preserving speech components in the stochastic part of noisy speech (i.e., $\hat{s}(n)$ in Case 2 in Fig. 2). In Table 1, we show the results of different speech enhancement scenarios (introduced in Fig. 2) when the clean (target) speech is degraded by WGN-Harmonics. We choose the adaptation rate $\mu = 0.0003$ small enough to separate the periodic part of WGN-Harmonics from the noisy speech. In this case, the deterministic part of noisy speech contains the harmonics, and the stochastic part contains the estimated target speech plus WGN. In Fig. 3, the spectrograms of a clean speech degraded by WGN-Harmonics at -10 dB SNR, the noise estimated by NegALE, the estimated target speech derived by processing the noisy speech only with NegALE (i.e., Case 2 in Fig. 2), the estimated target speech derived by processing noisy speech via Case 3, and the estimated target speech derived by only employing the SCNR (i.e., Case 1) are presented for 4 s of signals. In this experiment, the values of kurtosis for the estimated target speech in Case 2 and the estimated noise are 11.69 and 0.616, respectively.

In Table 2, the results of different speech enhancement scenarios (introduced in Fig. 2) are presented when the clean speech is degraded by traffic noise. Here, the most suitable adaptation rate was found equal to 0.008 which is larger than the adaptation rate for WGN-Harmonics in the former experiment. The reason is that traffic noise is more non-stationary than WGN-Harmonics, and thus the adaptation speed needs to increase to capture more components of noise in the deterministic part of noisy speech. Here, similar to the previous experiment, by applying NegALE to the noisy speech, the target speech is derived in the stochastic part.

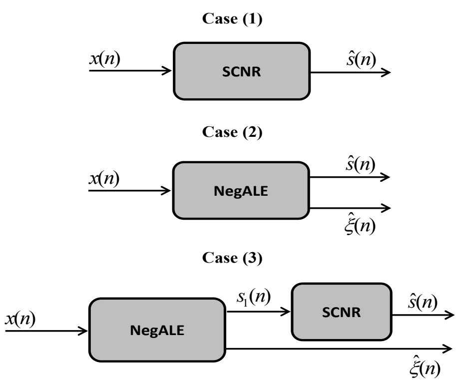

Fig. 2. Illustration for different scenarios of using the conventional single-channel noise reduction (SCNR) and the proposed approach (NegALE) in the experiments.

4. DISCUSSION AND CONCLUSIONS

From the Table 1 and Table 2, one can observe that exploiting the proposed approach for speech enhancement can be useful at low SNR conditions when it is used as a preprocessor to be combined with a conventional noise reduction approach. In the experiment with WGN-Harmonics (see Table 1 and Fig.2), the proposed strategy in Case 3 showed a better performance than that of Case 2 and Case 1 in terms of ΔPESQ while preserving the intelligibility in terms of $\mathrm{STOI_{out}}$. In fact, employing NegALE was helpful to separate the harmonics from the noisy speech and preserve the speech components in the final estimated target speech. At higher SNR values i.e., 0 dB and 5 dB, we observe that the performance of Case 2 and Case 3 is lower than that of Case 1 in terms of ΔSSNR. This can be due to the attenuation which is introduced by NegALE to the speech signal's amplitude where the instrumental measure ΔSSNR is sensitive to it while the other instrumental measures ΔPESQ and $\mathrm{STOI_{out}}$ are not affected by this attenuation.

In the experiment with traffic noise (see Table 2), we observed that Case 3 provides the best results compared to the other speech enhancement scenarios in terms of ΔPESQ while preserving the intelligibility in terms of $\mathrm{STOI_{out}}$. At the SNR of -15 dB, we observe that all methods (Case 1, Case 2 and Case 3) fail to improve the quality of noisy speech in terms of ΔPESQ. At higher SNR values i.e., 0 dB and 5 dB, the performance of NegALE decreases such that the worst performance for both Case 2 and Case 3 is derived at 5 dB SNR. Thus, it is deduced from the results that our proposed approach can not perform well for high SNR values (i.e., SNR values larger than 5 dB). The interesting point about our experiment with traffic noise is that exploiting our proposed adaptive filter (NegALE) alone in Case 2 could provide a better quality and intelligibility results (in terms of $\mathrm{STOI_{out}}$ and ΔPESQ) at SNR conditions -15 to -5 dB than when employing only the conventional noise reduction algorithm in Case 1.

As the final point, we should note that the adaptation rate parameter plays an important role in the performance of our proposed method. In fact, having a prior information about the type of noise (e.g, if the noise is non-stationary or stationary) can be helpful to find the optimum adaptation speed. Of course, more investigations are required in a future work regarding to adjusting the adaptation rate in an optimal way.

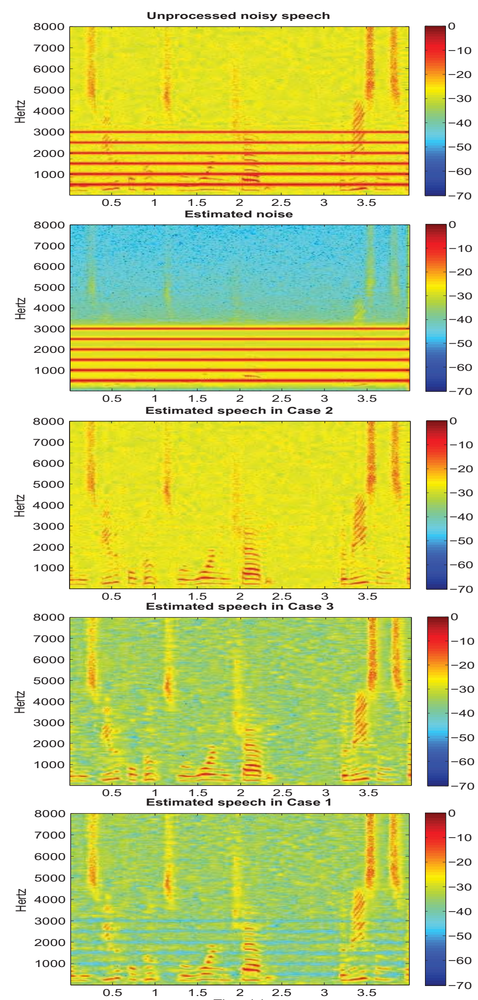

Fig. 3. The plots (from top to down) show the spectrograms of a clean speech degraded by WGN-Harmonics at -10 dB, the noise estimated by NegALE, the estimated target speech derived by processing the noisy speech with NegALE only (i.e., Case 2 in Fig. 2), the estimated target speech derived by processing noisy speech via Case 3, and the estimated target speech derived by employing the SCNR only (i.e., Case 1).

5. REFERENCES

[1] Y. Hu and P. C. Loizou, “A comparative intelligibility study of single microphone noise reduction algorithms,” *J. Acoust. Soc. Amer.*, vol. 22, no. 3, pp. 1777–1786, 2007.

[2] P. C. Loizou and G. Kim, “Reasons why current speech-enhancement algorithms do not improve speech intelligibility and suggested solutions,” *IEEE Transactions on Audio, Speech, and Language Processing*, vol. 19, no. 1, pp. 47–56, 2011.

[3] P. Loizou, *Speech Enhancement: Theory and Practice.*, CRC, Boca Raton, Florida, 2007.

[4] P. Comon, “Independent component analysis, a new concept?,” *Signal Processing*, vol. 36, pp. 287–314, April 1994.

[5] A. Hyvärinen, “Independent component analysis by minimization of mutual information,” *Technical Report A46, Helsinki University of Technology, Laboratory of Computer and Information Science*, August 1997.

[6] R. M. Ramli, A. O. Abid Noor, and S. A. Samad, “A review of adaptive line enhancers for noise cancellation,” *Australian Journal of Basic and Applied Sciences*, vol. 6, no. 6, pp. 337–352, 2012.

[7] N. Sasaoka, Y. Itoh, and K. Fujii, “A noise reduction method for non-stationary noise based on noise reconstruction system with ALE,” *IEICE Trans. Fundamentals*, vol. E88-A, no. 2, pp. 593–596, Feb. 2005.

[8] N. Sasaoka, K. Shimida, S. Sonobe, Y. Itoh, and K. Fujii, “Speech enhancement based on adaptive filter with variable step size for wideband and periodic noise,” *in Proceedings of IEEE International Midwest Symposium on Circuits and Systems*, pp. 648–652, Mexico, Cancun, 2009.

[9] K. Shimada, N. Sasaoka, T. Takemoto, Y. Itoh, and K. Fujii, “Speech noise reduction system based on adaptive filter with variable step size using cross-correlation,” *Proceedings of Asia-Pacific Signal and Information Processing Association*, pp. 477–484, Sapporo, Japan, October 2009.

[10] A. Hyvärinen, J. Karhunen, and E. Oja, *Independent component analysis*, John Wiley and Sons, 2001.

[11] R. Prasad, H. Saruwatari, and K. Shikano, “Noise estimation using negentropy based voice-activity detector,” *in Proceedings of the 47th IEEE International Midwest Symposium on Circuits and Systems*, vol. 2, pp. II–149–II–152, July, 2004.

[12] K. Kumatani, J. McDonough, B. Rauch, D. Klakow, P. N. Garner, and W. Li, “Beamforming with a maximum negentropy criterion,” *IEEE Transactions on Audio, Speech, and Language Processing*, vol. 17, no. 5, pp. 994–1008, July 2009.

[13] T. M. Cover and J. A. Thomas, *Elements of information theory*, Wiley, 2006.

[14] A. Hyvärinen and E. Oja, “Independent component analysis: algorithms and applications,” *Neural Networks Research Centre*, vol. 13, no. 4-5, pp. 411–430, 2000.

[15] “TIMIT, acoustic-phonetic continuous speech corpus,” DARPA, NIST Speech Disc 1-1.1, Oct. 1990.

[16] Y. Ephraim and D. 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.

[17] Y. Ephraim and D. Malah, “Speech enhancement using a minimum mean-square error log-spectral amplitude estimator,” *IEEE Transactions on Acoustics, Speech, and Signal Processing*, vol. 33, no. 2, pp. 443–445, 1985.

[18] R. C. Hendriks, R. Heusdens, and J. Jensen, “MMSE based noise PSD tracking with low complexity,” *IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP)*, pp. 4266–4269, March 2010.

[19] J. Taghia, J. Taghia, N. Mohammadiha, J. Sang, V. Bouse, and R. Martin, “An evaluation of noise power spectral density estimation algorithms in adverse acoustic environments,” *IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP)*, pp. 4640–4643, Prague, May 2011.

[20] C. H. Taal, R. C. Hendriks, R. Heusdens, and J. 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, September 2011.