Robust Time-Varying Filtering of Speech Signals Corrupted by Mixed Gaussian and Impulse Noise

Nikola Zaric, Srdjan Stankovic and Irena Orovic
University of Montenegro, Faculty of Electrical Engineering, 20000 Podgorica, Montenegro
E-mails: zaric@ac.me, srdjan@ac.me, irenao@ac.me

Abstract

A robust time-varying filtering procedure for speech signals corrupted by mixed Gaussian and impulse noise is presented. It is based on the robust time-frequency distributions that can provide efficient representation of the noisy speech signals. The proposed approach has been compared with the time-varying filtering procedure based on the standard time-frequency distributions.
Keywords – *robust time-frequency distributions, time-varying filtering, speech signals*

1. Introduction

In the last two decades time-frequency distributions are intensively used in various applications dealing with nonstationary signals [1]-[4]. Since the speech signals are characterized with high nonstationarity, the time-frequency distributions represent an efficient tool for their analysis. The time-frequency distributions have been used in various speech signal processing applications, such as filtering, watermarking and coding, [5]-[8].

The time-varying filtering of the noisy speech signal, based on the time-frequency signal representation, has been proposed in [5], where the noise with Gaussian probability density function (pdf) has been assumed. According to the maximum likelihood (ML) estimation theory, the standard time-frequency distributions can provide an optimal representation for signals in the Gaussian noise environment. However, the speech signals are often corrupted by a kind of impulse noise or by the mixed Gaussian and impulse noise. In that case the standard time-frequency distributions produce poor results [9]. In order to overcome this problem the robust time-frequency distributions have been introduced [9]-[13]. It has been shown that the L-estimate forms of robust time-frequency distributions outperforms the other estimate forms, if the present noise is mixed: Gaussian and impulse noise, [12].

In this paper we propose the robust time-varying filtering procedure for speech signals corrupted by mixed Gaussian and impulse noise. It is based on using the robust time-frequency distributions. Namely, the L-estimate robust spectrogram is used to define the robust time-varying filter function that is further used to reconstruct the signal from the time-frequency domain. The experimental results confirm the advantages of using the robust time-varying filtering procedure. Also, it has been shown that, for speech signal corrupted by impulse noise, the L-estimate robust spectrogram provides much better signal representation compared to the standard spectrogram.

2. Theoretical background

The standard short-time Fourier transform (STFT) can be defined as, [1]:

$$\begin{aligned} STFT_S(n,k) &= \frac{1}{N}\sum_{m=-N/2}^{N/2-1} x(n+m)e^{-j2\pi mk/N} \\ &= mean\left\{x(n+m)e^{-j2\pi mk/N}, m\in[-N/2,N/2)\right\}. \end{aligned} \tag{1}$$

where N is the number of samples within the window, while $x(n)=s(n)+v(n)$ represents the noisy speech signal, where $s(n)$ is noiseless speech signal corrupted with noise $v(n)$. Generally, the STFT can be obtained as a solution of the following optimization problem, [9]:

$$STFT(n,k) = \arg\min_{\mu} \sum_{m=-N/2}^{N/2-1} F\left(e(n,k,m)\right), \tag{2}$$

where $F(e)$ is the loss function, while $e(n,k,m)$ is the error function, [9]:

$$e(n,k,m) = x(n+m)e^{-j2\pi mk/N} - \mu. \tag{3}$$

The parameter μ represents the optimization parameter. According to the ML estimation theory the optimal time-frequency representation will be obtained by using the loss function that corresponds to the assumed pdf of noise.

For the noise with Gaussian pdf the optimal loss function is $F(e)=|e|^2$. Thus, the standard STFT is obtained by using this function in the optimization problem given by (2). However, the ML estimate approach is very sensitive to the assumed noise distribution. Since in practical application the pdf of noise is not known or could not be precisely determined in advance, the ML estimate approach usually failed to provide good results. This problem could be solved by using Huber theory of robust estimates. Namely, the robust estimate is determined for a specific class of noises. The noise with heaviest tails, belonging to the specific class, is considered and its ML estimate is determined. That estimate is used as a robust estimate for all noises from the considered class. The noise with Laplacian pdf can be considered as a worst case noise, for various noise classes. For the noise with Laplacian pdf, the ML estimator produces the loss function $F(e)=|e|$. The direct solution of the optimization problem in (2) for the absolute loss function produces nonlinear equation that requires computationally demanding iterative procedure. It could be solved by using the loss function $F(e)=|\mathrm{Re}(e)|+|\mathrm{Im}(e)|$ that produces marginal median estimate, [10]:

$$\begin{aligned} STFT_M(n,k) &= median\left\{\mathrm{Re}\left(x(n+m)e^{-j2\pi mk/N}\right)\right\} \\ &+ j\cdot median\left\{\mathrm{Im}\left(x(n+m)e^{-j2\pi mk/N}\right)\right\}, \end{aligned} \tag{4}$$

where $m\in[-N/2,N/2)$. This form provides efficient results if the pure impulse noise is present. However, in practice the signals are often corrupted by mixed Gaussian and impulse noise. In that case the L-estimate approach provides better results compared to the standard and the median based approaches. The L-estimate robust STFT can be defined as, [12]:

$$\begin{aligned} & STFT_L(n,k)=\sum_{i=-N/2}^{N/2-1} a_i(r_i(n,k)+j\cdot i_i(n,k)), \\ & r_i(n,k)\in R(n,k), R(n,k)=\left\{\mathrm{Re}(x(n+m)e^{-j2\pi mk/N})\right\}, \\ & i_i(n,k)\in I(n,k),\ I(n,k)=\left\{\mathrm{Im}(x(n+m)e^{-j2\pi mk/N})\right\}. \end{aligned} \tag{5}$$

The elements: $r_i(n,k)$ and $i_i(n,k)$ are sorted in non-decreasing order as: $r_i(n,k)\le r_{i+1}(n,k)$ and $i_i(n,k)\le i_{i+1}(n,k)$, respectively. The coefficients a_i can be defined in analogy with α-trimmed mean in the non-linear digital filter theory, [14]:

$$a_i=\begin{cases} \dfrac{1}{N(1-2\alpha)+4\alpha}, & \text{for } i\in[(N-2)\alpha,\alpha(2-N)+N-1] \\ 0, & \text{elsewhere,} \end{cases} \tag{6}$$

where N is even number, while the parameter α takes values within the range [0,1/2]. Note that for α=0 and α=1/2 the standard STFT and the marginal median STFT are obtained, respectively. By using smaller value of the parameter α better spectral characteristics will be obtained, while using higher value of α provides better reduction of heavy-tailed noise. Consequently, the value of parameter α should be chosen to satisfy trade-off between these requirements.

By using the L-estimate robust STFT, the L-estimate robust spectrogram can be obtained as:

$$SPEC_L(n,k)=\mathrm{Re}\{STFT_L(n,k)\}^2+\mathrm{Im}\{STFT_L(n,k)\}^2. \tag{7}$$

In the case of multicomponent signals, such as speech signals, the spectrogram does not contain the cross-terms. Also, it is less affected by noise compared with quadratic time-frequency distributions. Thus, we will use the L-estimate robust spectrogram to define the robust time-varying filtering procedure.

3. Robust time-varying filtering procedure

The time domain or frequency domain filtering procedures cannot provide satisfactory noise reduction in the case of nonstationary multicomponent signals. In that case the time-varying filtering procedure should be used. An efficient procedure for time-varying filtering of speech signals corrupted by Gaussian noise has been proposed in [5].

However, using the same procedure in the presence of mixed Gaussian and impulse noise does not produce efficient noise reduction. Namely, the standard time-frequency distributions are very sensitive to the presence of impulse noise and do not provide appropriate signal representation. Therefore, for filtering of nonstationary signals corrupted by impulsive noise we will define the robust time-varying filtering procedure that is based on the robust time-frequency distribution.

For a given noisy signal x, the pseudo form of robust time-varying filtering, can be defined as, [5]:

$$Hx(n)=\sum_{m=-N/2}^{N/2-1} h(n+m/2,n-m/2)w(m)x(n+m) \tag{8}$$

where w is a lag window, while h represents the impulse response of the robust time-varying filter. Previous relation can be written in the form, [5]:

$$Hx(n) = \frac{1}{N} \sum_{k=-N/2}^{N/2} L_H(n,k) STFT(n,k), \quad (9)$$

that is more suitable for realization. The support function $L_H(n,k)$, has been defined as Weyl symbol mapping of the impulse response into the time-frequency domain, [5]:

$$L_H(n,k) = \sum_{m=-N/2}^{N/2-1} h(n+m/2, n-m/2) e^{-j2\pi mk}. \quad (10)$$

There are several ways to determine the appropriate support function. Here, we will use simple approach that is based on the robust time-frequency representation of the signal. Using a threshold valued, the robust support function can be defined as:

$$L_H(n,k) = \begin{cases} 1 & \text{if} \quad \mathrm{SPEC_L}(n,k) > \xi \\ 0 & \text{if} \quad \mathrm{SPEC_L}(n,k) < \xi \end{cases}. \quad (11)$$

The threshold value ξ can be determined as: $\xi = \lambda \max_{n,k}(\mathrm{SPEC_L}(n,k))$, where λ represent scaling parameter. Note that the support function will be zero in the regions where signal components are not present, or they are very weak. This form of the support function is used in (9) to reconstruct the signal from the time-frequency domain. In the next section it will be shown that support function given by (11) provides efficient reduction of mixed Gaussian and impulse noise.

4. Examples

4.1. Example 1

The aim of this example is to illustrate the advantages of using the L-estimate robust spectrogram over the standard spectrogram, in the case of speech signals corrupted by mixed Gaussian and impulse noise.

The speech signal sampled at 8 KHz, corrupted by mixed Gaussian and impulse noise is considered. In order to obtain good frequency resolution the Hanning window of the width N=1024 is used. The parameter α=3/8 is used in the calculation of the L-estimate robust STFT. Namely, this value of parameter α provides satisfying trade-off between distribution concentration and noise reduction.

The standard spectrogram and the L-estimate robust spectrogram are shown in Figure 1. Observe that the standard spectrogram cannot provide good representation due to the presence of impulse noise. The influence of noise is significantly reduced in the case of the L-estimate robust spectrogram. Thus, the L-estimate robust spectrogram can be used as an efficient tool for analysis of speech signals corrupted by mixed Gaussian and impulse noise.

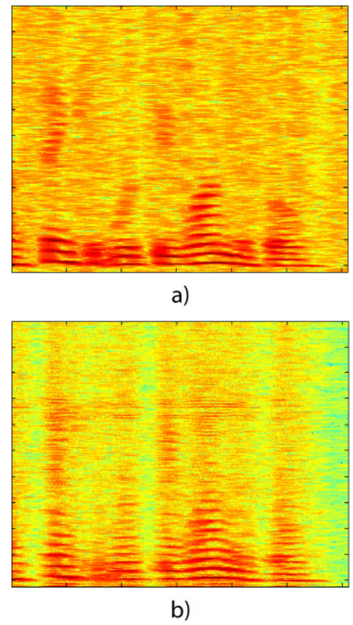

Figure 1. Time-frequency representations of speech signal corrupted by mixture of impulse and Gaussian noise: a) the standard spectrogram, b) the L-estimate robust spectrogram

4.2. Example 2

In this example, the proposed robust time-varying filtering procedure is applied on the speech signal corrupted by mixed Gaussian and impulse noise. The proposed procedure is compared with the standard time-varying filtering procedure based on the standard spectrogram. The same speech signal as in the previous example is considered. The STFT is calculated by using the rectangular window of the width N=256.

The standard spectrogram and the L-estimate robust spectrogram are shown in Figure 2.a and Figure 2.b, respectively. They are used in (11) to obtain the standard and the robust support functions that are shown in Figure 2.c and Figure 2.d, respectively. The value 0.05 for the parameter λ is used.

Note that the standard support function contains nonzero values in the regions where speech components are not present. Consequently, the noise will remain present in the signal filtered by using this support function. On the other hand, the robust support function has nonzero values only in the region where speech components are dominant. Thus, it will provide efficient reduction of mixed Gaussian and impulse noise.

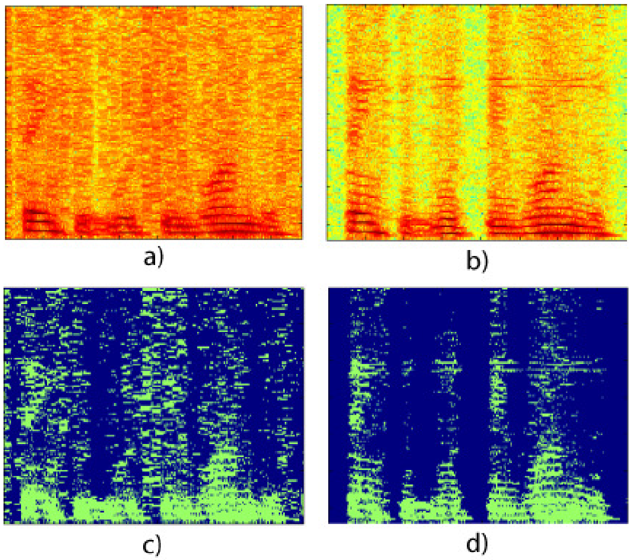

Figure 2. Time-frequency representations and corresponding support functions: a) the standard spectrogram, b) the L-estimate robust spectrogram, c) the standard support function, d) the robust support function.

The speech signal can be reconstructed from the time-frequency domain by using (9). The filtered speech signals, obtained by using the standard and the robust support functions, are shown in the last two rows in Figure 3, respectively. The original speech signal and non-filtered signal corrupted by mixed Gaussian and impulse noise are shown in the first two rows of Figure 3.

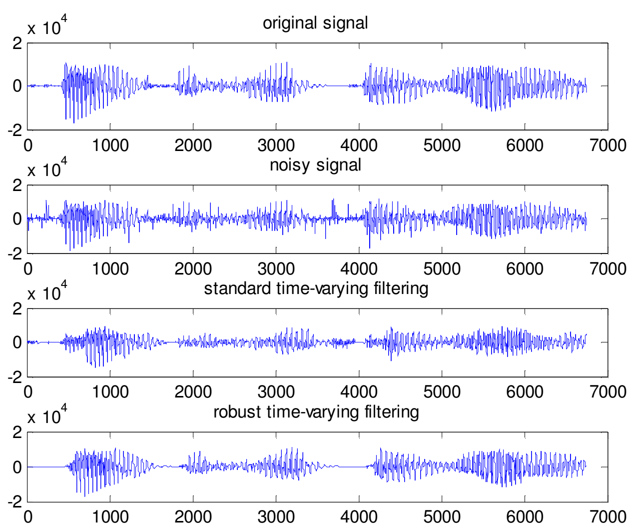

Figure 3. Time domain representations of: original signal, noisy signal, signal filtered with standard support function, and signal filtered with robust support function.

One may observe that significant amount of noise is still present in the case of signal filtered by using the standard support function. On the other side, the noise is almost completely removed from the signal that is filtered by using the robust support function.

5. Conclusion

The robust time-varying filtering procedure based on the robust time-frequency distributions is proposed. It is employed for filtering of speech signals corrupted by mixed Gaussian and impulse noise. For that purpose the appropriate time-frequency representation is obtained by using the L-estimate robust spectrogram. It has been shown that, for the considered type of noise, proposed procedure outperforms the standard time-varying filtering procedure.

6. References

[1] L. Cohen, Time-frequency analysis, Prentice-Hall, 1995.

[2] B. Boashash, Time-Frequency Analysis and Processing, Elsevier, 2003, Amsterdam.

[3] F. Hlawatsch, G.F. Boudreaux-Bartels, "Linear and quadratic time-frequency signal representations", *IEEE Signal Processing Magazine*, pp. 21-67, 1992.

[4] LJ. Stanković, "A method for time-frequency analysis", *IEEE Trans. Signal Processing*, vol. 42, no. 1, pp. 225-229, 1994.

[5] S. Stanković, "About Time-Variant Filtering of Speech Signals with Time-Frequency Distributions for Hands-Free Telephone Systems", *Signal Processing,* vol. 80, No. 9, pp. 1777-1785, Sept. 2000.

[6] K. Umapathy and S. Krishnan, *Audio signal coding using adaptive time-frequency transformation* in preparation, IEEE Trans. on Speech and Audio Processing, May 2001.

[7] C. Ioana, A. Jarrot, A. Quinquis, S. Krishnan, "A Watermarking Method for Speech Signals based on the Time-Warping Signal Processing Concept", in *Proc. IEEE ICCASP*-2007, vol. 2, pp. 201-204. April 2007.

[8] S. Stankovic, I. Orovic, N. Zaric, "Robust speech watermarking in the time-frequency domain," *EURASIP Journal on Advances in Signal Processing*, Vol. 2008, No. ID 519206, May 2008, Page(s) 9 pages.

[9] V. Katkovnik, "Robust M-periodogram," *IEEE Trans. Signal Processing*., vol. 46, no. 11, pp. 3104-3109, 1998.

[10] I. Djurović, V. Katkovnik, LJ. Stanković, "Median filter based realizations of the robust time-frequency distributions", *Signal Processing*, vol. 81, no. 7, pp. 1771-1776, 2001.

[11] V. Katkovnik, I. Djurović, LJ. Stanković, "Robust time-frequency distributions", in "Time-frequency signal analysis and applications", Elsevier, editor B.Boashash, pp. 392-399, 2003.

[12] I. Djurović, LJ. Stanković, J. F. Böhme, "Robust L-estimation based forms of signal transforms and time-

frequency representations", *IEEE Trans. Signal Processing*, vol. 51, no. 7, pp. 1753-1761, 2003.

[13] N. Zaric, I. Orovic, S. Stankovic, "Robust Time-Frequency Distributions with Complex-lag Argument Robust Time-Frequency Distributions with Complex-lag Argument," *EURASIP Journal on Advances in Signal Processing*, 2010.

[14] I. Pitas, A.N. Venetsanopoulos: Nonlinear digital filters: Principles and applications, Kluwer, 1990