Jianjun HUANG, Yafei ZHANG, Xiongwei ZHANG, Tao ZHU

Institute of Command Automation, PLA University of Science and Technology.

A Data Field method for speech enhancement incorporating Binary Time-Frequency Masking

***Abstract**. A data field approach coupled with binary time-frequency masking is presented for the speech enhancement problem. In this proposed approach, data field method is employed to model the time and frequency dependencies of speech. This formulation has proved to be very helpful in enhancing speech quality by exploiting the correlation of speech both in time and in frequency. The experimental results demonstrate that the proposed algorithm offers improved signal to noise ratio and less spectral distortion.*

***Streszczenie.** Do poprawy jakości dźwięku mowy zastosowano metodę pola danych (Data field) połączoną z binarnym maskowanie czasowo-częstotliwościowym. Pozwoliło to znacząco poprawić jakość dźwięku przez wykorzystanie korelacji czasowej i częstotliwościowej. Uzyskano poprawę stosunku sygnału do szumu i zmniejszenie poziomu zniekształceń. (**Metoda pola danych oraz maskowania czasowo-częstotliwościowego wykorzystana do poprawy jakości dźwięku**)*

Keywords: speech enhancement; data field; time-frequency masking; noise estimate
Słowa kluczowe: poprawa jakości dźwięku mowy, pole danych.

1 Introduction

This paper addressed the problem of suppressing the background noise in noisy speech. Speech can be corrupted by noise in various situations, such as trains, cars, airport etc. When noisy speech is transmitted over communication channel its quality at the other end can be significantly degraded, while the performance of services that use automatic speech recognition can deteriorate. Moreover, noise causes annoyance to listeners and induces fatigue on them. Hence, speech enhancement is essential to improve the performance of communication systems in noisy environments.

The objective of this work is to develop a novel speech enhancement algorithm based on data field theory that will account for the time and frequency dependencies of speech. The Short Time Fourier Transformation (STFT) matrices of speech signal are known to have rich structure. Consecutive samples within a frequency bin are highly correlated, as it was shown by Cohen [1]. Additionally, frequency dependencies within the same time frame exist due to the spectral leakage caused by the windows used in the calculation of the STFT and also due to the common modulation of the frequency bins [2,3]. The information that is encapsulated in the above attributes of a speech STFT matrix can prove very helpful in enhancing speech quality.

In conventional speech enhancement algorithms, such as spectral subtraction [4,5], minimum statistics [6] and wiener filter [7,8,9] etc., the speech STFT matrix samples in frequency axis are assumed to be mutually independent. Although the time dependencies of speech are exploited in traditional speech enhancement algorithm, little work has been done to incorporate the frequency dependencies of speech.

In this work, we propose the modelling of speech spectral amplitudes using Data Field that is capable of modelling both time and frequency dependencies of speech signal. Several techniques of the proposed algorithm are explained in detail. They include definition of potential function in noisy speech, influence factor optimization of field function, noise estimate in Data Field and binary Time-Frequency Masking calculation.

The rest of this paper is organized as follows. After a brief review of the Data Field theory in section 2, the new proposed speech enhancement algorithm is explained in section 3. In section 4 we present our simulation examples, and we conclude the paper in section 5.

2 Data Field and Perception of Speech Signal

The idea of field was first proposed by an English physicist, Michael Faraday, in 1837. In his opinion, the non-contact interaction between particles, such as gravitational force, electrostatic force, and magnetic force, is mediated through a certain medium called “field.” Inspired by the knowledge in physics discussion above, Deyi LI [10,11] introduce the interaction of particles and their field model into the data space in order to describe the relationship among data points and reveal the general characteristics of the underlying data distribution. With the development of the field theory, Deyi LI establish a virtual cognitive field named “Data Field” to model the interactions of data objects, and visualize the human minds perceptual, memorial, and thinking processes. In data field theory, each data object is considered as a mass point or atom with a certain field around it and the interaction of all data objects will form a data field through space. Starting with the wonderful ideas, we would consider all time-frequency points in the spectrum of speech as interacting objects by means of the data field.

In this work, Data field is employed to model the relationship or the dependencies of data points in time-frequency representation of speech to exploit the information of its neighbors. Fig.1 shows the transformation of speech signal to data field. Fig.1(a) is the time domain waveform of speech signal, which contains five Chinese words. In Fig.1(b) the equipotential lines plot of speech data field exhibit five local maximums of the associated potential field, which can be nearly considered as a data field produced by five “virtual point sources.” Obviously, these virtual sources are explicitly depicted in three-dimensional view of potential field in Fig.1(c). In this work, these virtual point sources are corresponding to acoustic unit, such as phonemes of speech. Due to the attraction of the “virtual sources,” all the time-frequency points of speech will tend to converge toward each other and exhibit certain features of self-organization. These features are also exploited by the recent research of Computational Auditory Scene Analysis (CASA) [3,12,13] which is used in sound segregation task. In CASA framework, the perceptual process that underlies auditory organization in human listeners is believed to involve two main stages: the segmentation stage and the grouping stage. In the grouping stage, time-frequency points from the same source are grouped according to a set of grouping principles. This grouping process is similar to the function of data field discussed above.

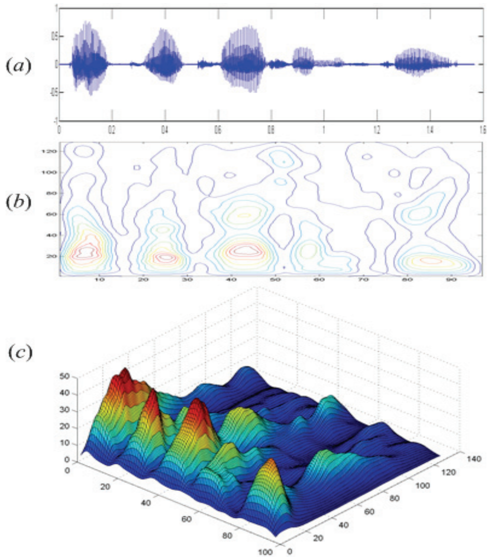

Fig.1. Transformation of speech signal to data field. (a) speech signal in time domain, (b) equipotential lines plot of speech data field, and (c) three-dimensional view of a two-dimensional potential field.

3 Speech enhancement based on Data Field theory

The proposed speech enhancement algorithm is based on Data Field theory. In this new framework, we would consider all time frequency points in the spectrum of speech as interacting objects by means of the Data Field. Data Field is used to model the relationship or the dependencies of data points in time-frequency representation of speech to exploit the information of its neighbors. A block diagram outlining the framework is shown in Fig.2.

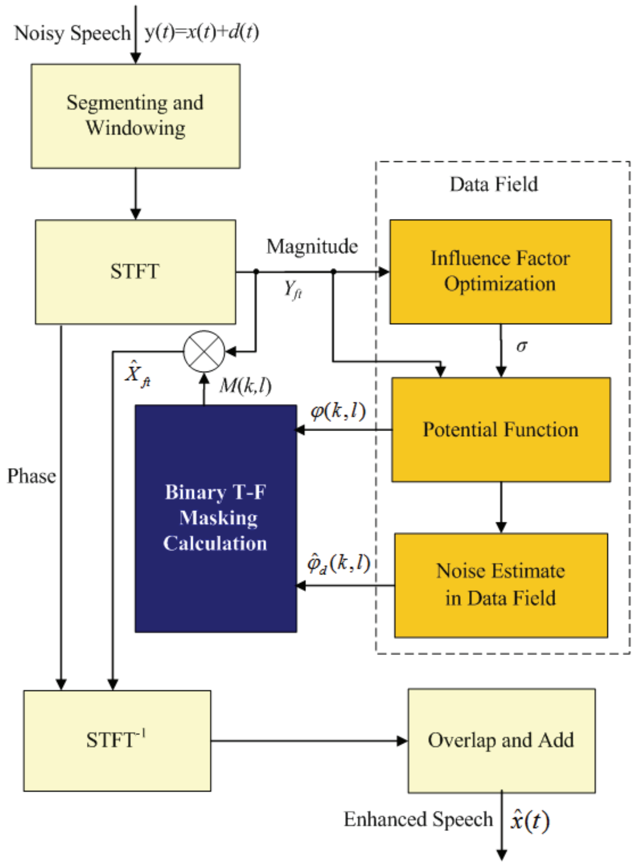

Fig.2. Block diagram of the proposed speech enhancement method

The noisy speech signal $y(t)$ is represented using the magnitude spectrogram, which is calculated as follows. Firstly, the time-domain speech signal is divided into frames and windowed. In our implementation a fixed 20ms frame size is used with 50% overlap between frames since it provides a good compromise between time and frequency resolutions. Each frame is then transformed into frequency domain using Short Time Fourier Transform (STFT), the length of the STFT being equal to the frame size. Only positive frequencies are retained. Phase are discarded by taking the absolute values of the STFT spectra to result in the magnitude spectrogram y_{ft}, where f=1…F is the discrete frequency index and t=1…T is the frame index. F is the number of frequency bins and T is the number of frames. Matrix Y_{ft} is used to denote the noisy speech magnitude spectrogram. The phase spectrogram is stored since it is needed in the synthesis. After the magnitude spectrogram y_{ft} is computed, binary time-frequency masking can be obtained in the following steps. First, influence factor σ is optimally chosen to make the distribution of potential field as consistent with the underlying distribution of speech as possible. Then, the potential $\varphi(k,l)$ at point (k,l) can be calculated using the optimal choice of σ and Matrix Y_{ft}. Thereafter, minimum statistics technique is employed to estimate the power spectral density (PSD) of noise $\hat{\varphi}_d(k,l)$ from noisy speech data field instead of frequency domain. Finally, Binary masks are estimated by comparing the values of the noisy speech potential with that of the estimated noise potential at each time-frequency (T-F) unit. The element of the mask matrix M is assigned 1 if the value of the noisy speech potential is larger than that of the estimated noise potential at that T-F unit and 0 otherwise. We use the mask matrix M and the STFT Matrix Y_{ft} of noisy speech to obtain the estimated clean speech $\hat{X}$. The estimated speech signal can then be transformed back to the time domain using inverse STFT and synthesis with the overlap and add method.

3.1 Definition of Potential Function in Noisy Speech

Field describes the distribution of a physical variable or mathematical function throughout space. There exit "scalar field" and "vector field" in physics. The potential field is the most frequently discussed vector field in fundamental physics, where potential means the work done to move a unit object, such as a mass point of the unit mass in a gravitational field or a point charge with unit charge in a static electric field, from some position to the reference point. In a potential field, the distribution of potential energy is determined uniquely by the relative position of interactive particles in space. Since the potential energy difference between some position and the reference point is a definite value, it is clear that potential function is only a function of the coordinates, regardless of whether objects exist.

Let x(n) and d(n) denote the speech and noise process respectively, the observed signal y(n) is given by y(n)=x(n)+d(n),where n denotes the time index. We further assume that x(n) and d(n) are statistically independent and zero mean. The noisy speech is transformed into frequency domain by applying a window h(n) to a frame of L samples of y(n) and by computing the short time Fourier transform (STFT) of size L on the windowed data. The window is shifted by R samples before the next STFT computation. The STFT analysis results in a set of frequency domain signal which can be written as

$$(1)\quad Y(k,l)=\sum_{n=0}^{L-1} y(lR+n)h(n)e^{-j2\pi kn/L}, \qquad 0\le k\le L-1,$$

where k represents the frequency bin index, and l the frame index.

Given *M* frames of frequency domain signal of the noisy speech $Y(k,l)(0 \le k \le L-1, 0 \le l \le M-1)$, they form a $L \times M$ two-dimensional time-frequency space. If each time-frequency point is considered as a data object in two-dimensional speech space and the associated amplitude $\rho_{kl} = \|Y(k,l)\|$ is considered as the mass of associated data objects, the interaction between all the time-frequency points will generate a data field in the speech space. The potential at any point can be calculated as

$$\varphi(z) = \sum_{k=0}^{L-1}\sum_{l=0}^{M-1}\varphi_{kl}(z) = \sum_{k=0}^{L-1}\sum_{l=0}^{M-1}\left(\rho_{kl} \times e^{-\left(\frac{\|z-z_{kl}\|}{\sigma}\right)^2}\right), \quad (2)$$

where z_{kl} denotes the position of the time-frequency point at (*k,l*),then $\|z-z_{kl}\|$ is Euclidean distance between point z and point z_{kl}. $\sigma \in (0, \infty+)$ is the influence factor that indicates the range of interaction.

3.2 Influence Factor Optimization for Field Function

Once the form of potential function is fixed, the distribution of the associated data field is primarily determined by the influence factor σ. If σ is too small, the range of interaction is very short, which means that each time-frequency point can only influence few points around it. If σ is too large, there is strong interaction between the time-frequency points and each time-frequency point will affect the points faraway. It sounds no meaningful that words said now would affect words said in a near few seconds. Obviously, the wrong choice of σ cannot produce a meaningful estimate of the underlying distribution of speech. Thus, the choice of σ should make the distribution of potential field as consistent with the underlying distribution of speech as possible.

In [10], the authors used potential entropy to measure the uncertainty about the distribution of potential field. If the potentials at the positions of all objects are equal to one another, we are most uncertain about the underlying distribution. And Shannon's entropy is the highest in this case. Conversely, if the distribution of the potential field is highly skewed, the uncertainty and the entropy will be low. The potential entropy can also be used in speech enhancement task. When the distribution of speech and noise are very asymmetrical, we can easy to separate noise from noisy speech.

If $\phi(k,l)$ is the potential at the position of object (*k,l*), the potential entropy can be defined as

$$H = -\sum_{k=0}^{L-1}\sum_{l=0}^{M-1}\left(\frac{\varphi(k,l)}{\Theta}\log\left(\frac{\varphi(k,l)}{\Theta}\right)\right), \quad (3)$$

where $\Theta = \sum_{k=0}^{L-1}\sum_{l=0}^{M-1}\phi(k,l)$ is the normalization factor.

Optimal choice of σ is a minimization problem of a univariate, non-linear function *H*, i.e., min *H*(σ), which is described as follows.

$$\min\ H(\sigma) = \min -\sum_{k=0}^{L-1}\sum_{l=0}^{M-1}\left(\frac{\phi(k,l)}{\Theta(\sigma)} * \log\left(\frac{\phi(k,l)}{\Theta(\sigma)}\right)\right). \quad (4)$$

3.3 Noise Estimate in Data Field

R.Martin has proposed a method based on minimum statistics to estimate the Power Spectral Density (PSD) of noise from the noisy speech [6]. This method can be combined with any speech enhancement algorithm which requires a noise power spectral density estimate. Here we also use minimum statistics to estimate the PSD of noise from noisy speech data filed instead of frequency domain. The algorithm for estimate the PSD of noise in data field is given here.

1) calculate the smoothed potential as a estimate of signal *PSD*. The smoothing parameter is typically set to values between 0.9~0. 95

$$\hat{\varphi}(k,l+1) = \alpha\hat{\varphi}(k,l) + (1-\alpha)\,|\,\varphi(k,l)\,|^2 . \quad (5)$$

2) find the minimum of M consecutive PSD estimate $\hat{\varphi}(k,l)$ *as the estimate of noise PSD*. Here we typically use *M*=12.

$$\hat{\varphi}_{\min}(k,l) = \min\{\hat{\varphi}(k,l-M-1),..,\hat{\varphi}(k,l-1),\hat{\varphi}(k,l)\} . \quad (6)$$

3) compute the final estimate of signal potential by multiplying a bias correction factor $B_{\min}$ which is typically set to 1.5

$$\hat{\varphi}_d(k,l) = B_{\min}\hat{\varphi}_{\min}(k,l) . \quad (7)$$

3.4 Binary Time-Frequency Masking

In this section, the human perceptual auditory masking effect is incorporated into the enhancement algorithm in order to achieve an optimal effectiveness for both high noise suppression and low speech distortion. There are many proposed speech enhancement algorithm which exploit the properties of the human auditory system. Recent studies in [13,14] show that binary time-frequency masking can be used to effectively suppress interference in speech signal processing task, such as blind speech separation, due to its superior performance in interference rejection. In this work, audible noise is masked according to auditory masking effect which is represented using binary time-frequency mask. If noisy speech potential $\hat{\varphi}(k,l)$ is larger than estimated noise potential $\hat{\varphi}_d(k,l)$, then the binary mask *M*(*k,l*)=1. Otherwise *M*(*k,l*)=0. The binary mask is formulated as follows.

$$M(k,l) = \begin{cases} 1, & if\ \hat{\varphi}(k,l) > \hat{\varphi}_d(k,l) \\ 0, & otherwise \end{cases} . \quad (8)$$

The binary mask determines whether speech is present or absent in a time-frequency point. Observations show that human auditory system is not sensitive to phase. After the binary time-frequency mask is calculated, the estimated speech spectral can be formulated as:

$$\hat{X}(k,l) = M(k,l)\cdot|\,Y(k,l)\,|\cdot\angle Y(k,l) , \quad (9)$$

where $|\,Y(k,l)\,|$ is noisy spectral amplitude, $\angle Y(k,l)$ is noisy spectral phase. The estimated clean speech signal can then be transformed back to the time domain using inverse STFT and synthesis with the overlap and add method.

4 Performance Evaluation

In this section, detailed performance evaluations are performed for the proposed enhancement algorithm. The quality of the enhanced speech is assessed in terms of Signal to Noise Ratio (SNR) as well as Log spectral distortion (LSD) objective speech quality measures. Obviously, higher values for the SNR reflect better signal-to-noise ratio, and smaller LSD measures indicate better distortion performance. Finally, subjective study of time waveform, spectrograms and potential in data field of speech signal is performed. We compared the performance obtained with the proposed enhancement algorithm against the performance obtained with minimum statistics estimation in frequency domain [6] and enhancement algorithm based on markov random field theory [2]. The abbreviation DF, MM and MRF are used to refer to these algorithms, respectively.

The sample frequency is 8 kHz and the following parameters have been chosen:

1) frame size L=256 with 50% overlap;
2) Hanning window is chosen as the window function;
3) M=12 frames is used to form data field;
4) smoothing parameter α=0.95
5) bias correction factor $B_{\min}$ =1.5

The NOIZEUS database [15] is used here to compare speech enhancement algorithms. The noisy database contains 30 IEEE sentences (produced by three male and three female speakers) corrupted by eight different real-world noises. The sentences were originally sampled at 25 kHz and downsampled to 8 kHz. The noise power is adjusted so as to give different input SNR (signal noise ratio) in the range of 0 to 15 dB. The performance of the speech enhancement algorithm is subject to some quality measures. Hence, it is necessary to employ different criteria to evaluate the performances of all algorithms. Here we used SNR and LSD (log spectral distortion) measurements.

1) *Signal to Noise Ratio (SNR)*: input SNR and output SNR are defined in time domain by

$$\text{SNR}_{in} = 10\lg \frac{\sum_{n=1}^{N}\sum_{l=1}^{L} x_n^2(l)}{\sum_{n=1}^{N}\sum_{l=1}^{L} d_n^2(l)}, \tag{10}$$

$$\text{SNR}_{out} = 10\lg \frac{\sum_{n=1}^{N}\sum_{l=1}^{L} x_n^2(l)}{\sum_{n=1}^{N}\sum_{l=1}^{L} [\hat{x}_n(l) - x_n(l)]^2}, \tag{11}$$

where $x_n(l)$, $d_n(l)$ and $\hat{x}_n(l)$ denote the time domain signal of clean speech, noise and enhanced speech, respectively.

2) *Log spectral distortion (LSD)*: LSD is defined by

$$LSD = \left(\frac{1}{K}\sum_{l=0}^{K-1}\frac{1}{L}\sum_{k=0}^{L-1}\left(10\lg\frac{|\hat{X}(k,l)|^2}{|X(k,l)|^2}\right)^2\right)^{1/2}, \tag{12}$$

where $X(k,l)$ and $\hat{X}(k,l)$ denote the STFT of the clean signal and the estimated signal, respectively. LSD measures the distortion of the estimated signal, where smaller value of LSD corresponds to better speech quality. The output SNR performance and log spectral distortion measure obtained at various noise levels for different speech enhancement algorithms are listed in Table 1 and Table 2, respectively.

Table 1 Performance evaluation using output SNR

Methods	Input SNR(dB)			
	0	5	10	15
MM	10.24	14.57	18.67	21.73
MRF	12.05	16.90	19.68	22.89
DF	13.74	17.94	20.79	23.04

Table 2 Performance evaluation using LSD measures

Methods	Input SNR(dB)			
	0	5	10	15
MM	7.73	6.15	5.24	4.60
MRF	6.14	5.17	4.25	4.03
DF	5.36	4.45	3.61	3.24

The results shown in Table 1 and Table 2 are the average results obtained from 240 experiments on 30 sentences corrupted by 8 noise types. We can see from these tables that the proposed enhancement algorithm always outperforms the other methods throughout the whole input SNR range. Compared to other algorithms, the proposed algorithm produces an increased SNR at all noise levels. At about 1 dB improvement of the proposed algorithm in SNR value is clearly visible as compared to MRF enhancement algorithm. Values of the LSD measure in various noise levels and for different enhancement algorithms are presented in Table 2. For LSD measure, the proposed algorithm achieves a significant improvement over the other two algorithms. The LSD values is 0.73 lower than the MRF algorithm on average. As evident, the enhancement algorithm based on minimum statistics produces the worst results for lack of exploiting the frequency dependencies.

To analyze structure of the residual noise and speech distortions, the time domain waveforms, spectrograms and speech data fields are used. Signal examples resulting from the proposed enhancement algorithm and Martin's enhancement algorithm at 6dB noise level are shown in Fig.3-5. The test speech is corrupted by Gaussian noise. We see that the structure of the speech signal is well recovered by ours proposed algorithm. In the proposed algorithm, the residual noise is better reduced, which leads to the best results as compared to other enhancement methods.

As expected, the proposed method produces lower residual noise and less speech distortion in some speech segments. This confirms the values of the output SNR and LSD measure of Table 1 and Table 2 and it is validated by informal listening tests.

Segment A1 and A2 of Fig.3 and listening tests show that speech enhanced with the proposed method is more pleasant and the residual noise is almost inaudible. However, the residual noise of the Martin enhancement algorithm is not well reduced which can be perceived as a perceptually white quality. Informal listening tests and examination of speech waveforms in Fig.3 also reveal that the proposed enhancement algorithm achieves the recovery of some weak speech spectral components which are suppressed by the Martin's algorithm. This is obvious as shown in segment B1 and B2 of Fig.3.

Speech enhancement results obtained by the two algorithms in the frequency domain and speech data field representation are also presented in Fig.4 and Fig.5, respectively. As expected, similar results are illustrated in these figures.

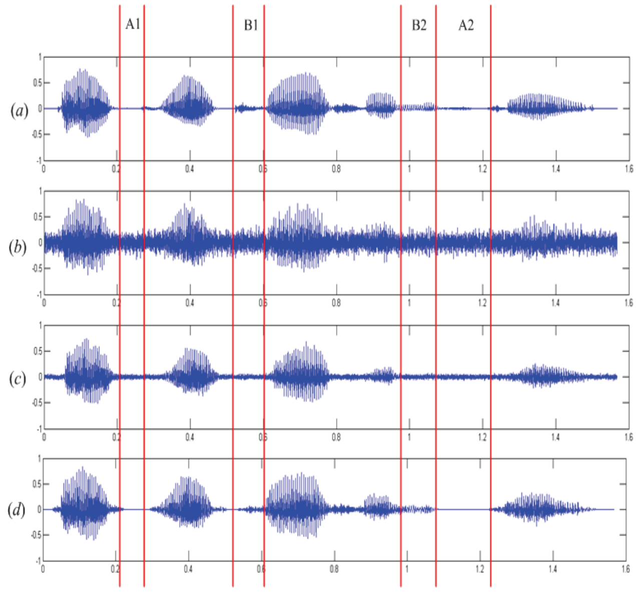

Fig.3. Time waveforms of speech. (a) Original speech. (b) Noisy speech. (c) Speech enhanced with Martin's algorithm. (d) Speech enhanced with proposed algorithm

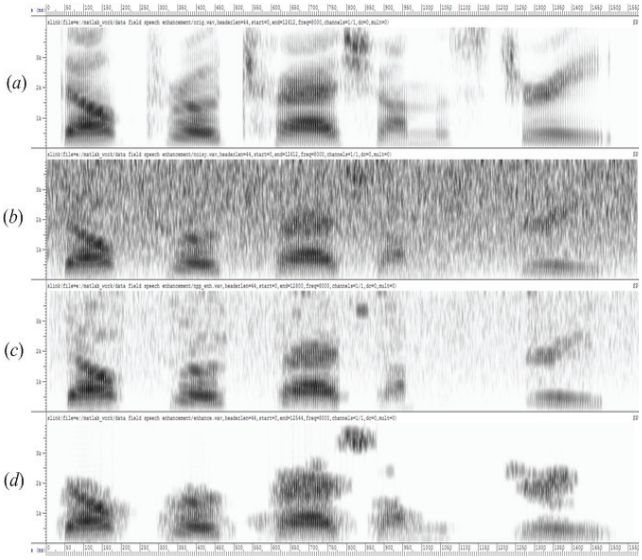

Fig.4. Speech spectrograms: (a) Original speech; (b) Noisy speech; (c) Speech enhanced with Martin's algorithm; (d) Speech enhanced with proposed algorithm.

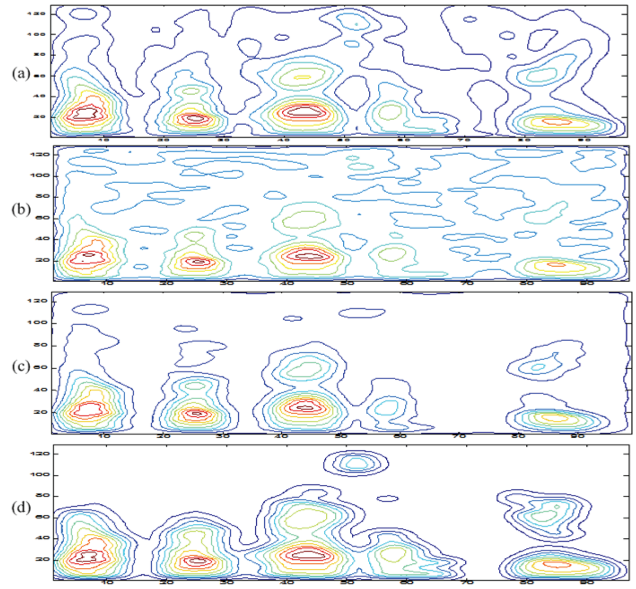

Fig.5. Speech signal in Data Field: (a) Original speech; (b) Noisy speech; (c) Speech enhanced with martin's algorithm; (d) Speech enhanced with proposed algorithm.

5 Conclusion and Future work

We have developed a novel speech enhancement algorithm based on data field which is capable of modeling both time and frequency dependencies of speech. It offers a new insight into the organization of speech signal. Data field method shows its superior quality by exploiting the correlation of speech both in time and in frequency. The information that exit in the time and frequency correlations of a speech STFT matrix is proved very helpful in the restoration of speech degraded by background noise. The proposed algorithm has been tested and compared to classical methods, in various noise type and SNR levels. The objective evaluation has been completed by the description of Signal to Noise Ratio and Log Spectral Distortion. Results show that the residual noise produced by the proposed algorithm is less than that of the compared methods, while the distortion of speech is smaller. This confirms that data field is suitable to model the time-frequency structure of speech and allows for a significant improvement over classical methods.

The speech enhancement algorithm we proposed in this paper can be extended for other speech signal processing task, such as speech separation [14]. The research on how to solve this problem is the future work.

Acknowledgement

This paper is partially supported by Pre-Research Foundation of PLA University of Science and Technology (2009TX08) and Natural Science Foundation of Jiangsu Province, China (No. BK2009059).

REFERENCES

[1] I.Cohen, "Relaxed statistical model for speech enhancement and a priori SNR estimation," IEEE Transactions on Speech and Audio Processing, vol.13, no.5, pp.870-881, 2005.

[2] I.Andrianakis and P.R.White, "On the application of Markov Random Fields to speech enhancement," in Proc. IMA Int. Conf. Mathematics in Signal processing, pp.198-201, 2006.

[3] Yipeng Li, John Woodruff, and De Liang Wang, "Monaural musical sound separation based on pitch and common amplitude modulation," IEEE Transactions on Audio, Speech, and Language Processing, vol.17, no.7, pp.1361-1371, 2009.

[4] S. F. 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.

[5] Yang Lu and Philipos C. Loizou, "A geometric approach to spectral subtraction," Speech Communication, vol.50, pp. 453-466, 2008.

[6] R. Martin, "Noise power spectral density estimation based on optimal smoothing and minimum statistics," IEEE Transactions on Speech and Audio Processing, vol.9, no.5, pp. 504-512, Jul. 2001.

[7] Simon Doclo, Jan Wouters and Marc Moonen, "Speech enhancement with multichannel wiener filter techniques in multimicrophone binaural hearing aids," Journal of the Acoustical Society of America. vol.25, no.1, pp.360-371, 2009

[8] Plapous, C., Marro and C., Scalart, P., "Improved Signal-to-Noise Ratio Estimation for Speech Enhancement," IEEE Transactions on Audio, Speech and Language Processing, vol.14, no.6, pp.2098-2108, 2006.

[9] Gomez,R. and Kawahara,T., "Optimizing spectral subtraction and wiener filtering for robust speech recognition in reverberant and noisy conditions," Proceedings of ICASSP, Dallas, pp.4566-4569, 2010.

[10] D.Li and Y.Du, Artificial Intelligence with Uncertainty. National Defense Press, 2005.

[11] D.Li, K.Liu, Y.Sun and M.Han, "Emerging Clapping Synchronization From a Complex Multiagent Network With Local Information via Local Control," IEEE Trans. on Circuits and Syetems-II:Express Brief, vol.56, no.6, pp. 504-507, 2009.

[12] Hu G. and Wang D.L, "A tandem algorithm for pitch estimation and voiced speech segregation," IEEE Transactions on Audio, Speech, and Language Processing, vol.18, no.8, pp. 2067-2079, 2010.

[13] Y.Shao, S.Srinivasan, Z.Jin and D.L.Wang, "A computational auditory scene analysis system for speech segregation and robust speech recognition," Computer Speech and Language, vol.24, pp.77–93, 2010.

[14] Tao Xu and Wenwu Wang, "A block-based compressed sensing method for underdetermined blind speech separation incorporating binary mask," ICASSP, 2009.

[15] Y.Hu and P.Loizou, "Subjective comparison of speech enhancement algorithms," Proceedings of ICASSP, Toulouse, France, pp.153-156, May 2006.

***Authors**:*

Jianjun HUANG, Postgraduate Team 2, Haifu Xiang 1, Baixia District, Nanjing, China, 210007, E-mail: hjj954@gmail.com;

Yafei ZHANG, PLA University of Science and Technology.

Xiongwei ZHANG, PLA University of Science and Technology.

Tao ZHU, PLA University of Science and Technology.