Jianjun HUANG, Yafei ZHANG, Xiongwei ZHANG, Zou XIA

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

A Nonnegative Subspace Approach for Packet Loss Concealment

***Abstract**. This paper presents a nonnegative subspace approach for packet loss concealment problem. The magnitude spectrogram of speech signal is projected onto nonnegative subspace using nonnegative matrix factorization algorithm. Consequently, packet loss concealment problem is transformed to linear interpolation of the projective coefficients in nonnegative subspace. Simulation examples, objective tests show that packet loss concealment in the nonnegative subspace results in improved perceptual quality of speech compared to popular packet loss concealment algorithms.*

***Streszczenie.** Zaprezentowano metodę subprzestrzeni dla rozwiązania problemu straty pakietu. Spektrogram amplitudowy sygnału mowy .jest poddawany projekcji do nieujemnej podprzestrzeni przy wykorzystaniu macierzy faktoryzacji. W rezultacie problem staje się możliwy do liniowej interpolacji. Osiągnięto dostrzegalną poprawę jakości przetwarzania sygnału mowy. (**Metoda nieujemnej podprzestrzeni stosowana w przypadku straty pakietu**)*

Keywords: packet loss concealment; nonnegative subspace; nonnegative matrix factorization
Słowa kluczowe: strata pakietu, nieujemna przestrzeń.

1 Introduction

With the widespread usage of the Internet, voice over packet network has become increasingly popular. Since the delivery of packets is not guaranteed in the IP network, packet loss could occur if the packets in the network are misrouted, or if the packets arrive at the destination too late to be useful. Voice quality of service is subject to packet loss. To prevent from serious quality degradation, various Packet Loss Concealment (PLC) algorithms have been proposed.

The basic idea behind PLC is to recover the speech signal content of a lost packet from its neighbors. Most of the approaches to PLC are formulated in a heuristic manner. However, the procedures can roughly be categorized into two classes: 1) *predictive or extrapolative*, where the PLC attempts to recover the lost speech signal segment based solely on previous samples, and 2) *estimative or interpolative*, where the PLC attempts to recover the missing speech segment based on both past and future samples [1].

The simplest predictive method is to replicate the previous speech packet instead of the lost packet. This method has low computational complexity and performs better than muting the signal, but it is not sufficient for high quality applications.

An even better method is waveform substitution. This method selects a portion of the previous speech which can best approximate the lost samples, and uses these samples to fill the lost packet. In [2] D. J. Goodman presented the predictive case of waveform substitution by the use of pitch cycle repetition. Y. J. Liang extended the work to estimative case through waveform similarity overlap-add (WSOLA) [3]. The waveform substitution based techniques have proved to be very popular due to its simplicity. However, continual loss packets lead to perceptually metallic sounding artifacts [4].

A more sophisticated technique is to use a Hidden Markov Model (HMM) to track the evolution of speech signal parameters in a statistical signal processing framework. The HMM-based PLC transcends the conventional methods of variations of repetition and interpolation. But as C. A. Rødbro described in [1], HMM is an imperfect model of speech signals. Although differing from conventional PLC algorithm, HMM-based PLC algorithm is still a “local” algorithm which is determined by the number of model states. Therefore more sophisticated models of speech for PLC are welcome to provide more natural sounding speech.

In this paper, we present an approach to perform PLC in nonnegative subspace using the well-known Nonnegative Matrix Factorization (NMF) algorithm. Although extensively used in speech separation and polyphonic music separation, NMF have not been previously used in PLC. In applying NMF to PLC, the magnitude spectrogram of speech signal is decomposed into a sum of basis function, which forms the nonnegative subspace and time-varying gains. Figure 1 shows how the spectrogram is projected onto nonnegative subspace. The projective coefficients named time-varying gains are derived after projection. Consequently, the PLC problem is transform into linear interpolation of the time-varying gains in nonnegative subspace. When a packet is lost, the PLC system performs linear interpolation of the time-varying gains on both sides. In fact, we show that PLC in nonnegative subspace provides more natural sounds than conventional PLC, including the HMM-based PLC algorithm.

The rest of this paper is organized as follows. After a brief review of the nonnegative matrix factorization in section 2, the new proposed packet loss concealment algorithms in nonnegative subspace are explained in section 3. In section 4 we present our simulation examples, objective tests result, and we conclude the paper in section 5.

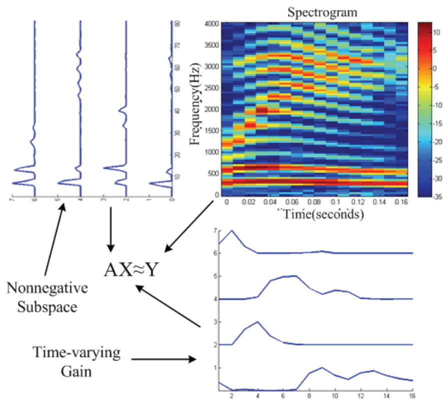

Fig.1. Projecting the spectrogram onto Nonnegative Subspace

2 Nonnegative Matrix Factorization

Nonnegative Matrix Factorization was firstly introduced by Lee and Seung [5,6]. NMF approach has widely adopted as a very useful and promising technique for decomposing high-dimensional datasets into a lower-dimensional nonnegative subspace. The use of NMF in acoustic signal processing [7,8,9,10,11] has gradually increased in recent years due to their capability of providing new insights and relevant information on complex latent relationships in nonnegative data sets. NMF achieves its success as a result of a main effect caused by nonnegative constraints. Since no elements are allowed to be negative and, thus, all combinations are additive, the factorization often leads to a part based representation with physical or physiological interpretations. The part we hope to find here are basic feature vectors which span the nonnegative subspace.

Its formulation is as follows. Given a nonnegative matrix $Y \in \mathrm{R}^{\geq 0, F\times T}$, the goal is to factorize it as a product of nonnegative matrix $A \in \mathrm{R}^{\geq 0, F\times r}$ and $X \in \mathrm{R}^{\geq 0, r\times T}$:

(1) $$Y \approx AX \quad s.t. \quad A, X \geq 0$$

The decomposing process is shown in Figure 1 when matrix *Y* represents a spectrogram. The factorization is performed by minimizing the error of reconstruction of *Y* by *A*×*X*.

(2) $$\min_{A,X\geq 0} D(Y, AX)$$

We formulate the NMF problem within the context of a noise model as stated in equation (3)

(3) $$Y = AX + N$$

The matrix $N \in \mathrm{R}^{F\times T}$ is the residual noise.

In this paper, we propose to build prior knowledge into the solution of the NMF problem in the maximum a posterior (MAP) estimate framework. The posteriori probability is calculated using Bayes rule as follows.

(4) $$P_{A,X|Y}(A,X\mid Y) = \frac{P_{Y|A,X}(Y\mid A,X)P_{A,X}(A,X)}{P_Y(Y)}$$

Assume that *A* and *X* is independent of each other, equation (4) can be reformulated as equation (5)

(5) $$P_{A,X|Y}(A,X\mid Y) = \frac{P_{Y|A,X}(Y\mid A,X)P_A(A)P_X(X)}{P_Y(Y)}$$

Since the denominator is constant, the negative log posterior is the sum of a likelihood term that penalizes model misfit, and two prior terms that penalize solutions that are unlikely under the prior.

(6) $$L_{A,X|Y}(A,X) \propto L_{Y|A,X}(Y\mid A,X) + L_A(A) + L_X(X)$$

where $L_{A,X|Y}(A,X) = -\log(P_{A,X|Y}(A,X\mid Y))$,

$L_{Y|A,X}(Y\mid A,X) = -\log(P_{Y|A,X}(Y\mid A,X))$,

$L_A(A) = -\log(P_A(A))$,

$L_X(X) = -\log(P_X(X))$.

The MAP estimate of *A* and *X* is

(7) $$\{A_{map}, X_{map}\} = \arg\min_{A,X\geq 0} L_{A,X|Y}(A,X)$$

and it can be calculated using any appropriate optimization algorithm.

3 The new packet loss concealment algorithm

PLC in nonnegative subspace algorithm is based on the principle that the received speech signal can be effectively modelled as basic feature vectors confined in a lower dimensional nonnegative subspace. Thus, the lost speech samples can be recovered by projecting the received speech signal onto the signal nonnegative subspace and performing linear interpolation of the time-varying gains. The information about the signal nonnegative subspace is obtained using nonnegative matrix factorization. A block diagram outlining the framework is shown in Fig. 2.

The received 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. The packet length is equal to the frame size in the packet loss model which will be described in Section 4. 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. Phases 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 received magnitude spectrogram. The phase spectrogram is stored since it is needed in the synthesis.

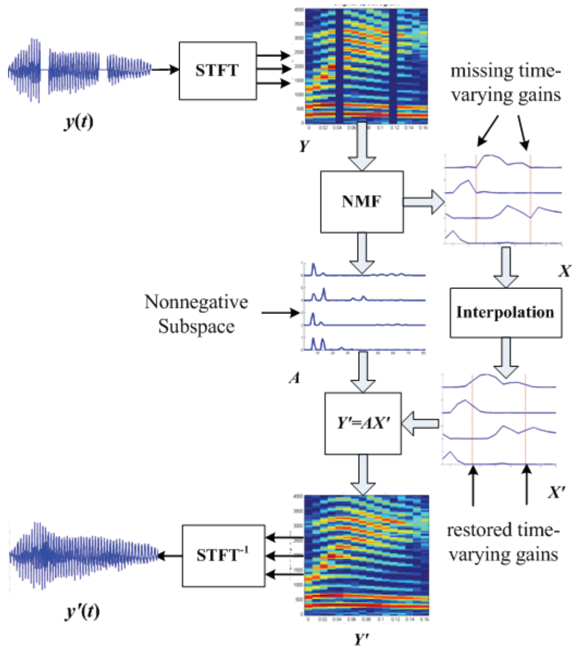

Fig. 2. Block diagram of the new PLC framework.

Since all the entries in *Y* are not negative, the magnitude spectrogram matrix *Y* can be decomposed into the product of nonnegative matrix *A* and *X* by Equation (1). The column vectors of *A* form the nonnegative subspace, while rows of *X* are time-varying gains. When packet losses occur, the corresponding columns of the time-varying gain matrix *X* become zero. In this scenario, linear interpolation of the time-varying gains is performed between the previous and the last receive packet. The output *X′* of Linear interpolation module is the estimates of the time-varying gains in nonnegative subspace *A*, which is mapped back into the magnitude spectrogram representation, *Y*=*AX′*. We could simply combine the phase information of the received speech signal to compute the time domain signal by the inverse STFT. Since human auditory perception is rather insensitive to phase, the phase of lost speech signal is estimated simply using the previous phase.

As is described in section 2, decomposition of *Y* into *A* and *X* is done by minimizing the negative log posterior, which is a sum of three terms: likelihood term $L_{Y|A,X}(Y|A,X)$, prior term $L_A(A)$and $L_X(X)$. In this paper, We assume that the prior on each data point Y_{ft} ($1\leq f\leq F$,$1\leq t\leq T$) is i.i.d Poisson distribution, the prior on A_{fi} ($1\leq f\leq F$,$1\leq i\leq r$) is positive, and the prior on X_{it} ($1\leq i\leq r$, $1\leq t\leq T$) is temporally smooth.

3.1 Likelihood Term

In the MAP framework, we can state our knowledge of the distribution of the residual in terms of a likelihood function. The priors are chosen in accordance with our beliefs about the distribution of the speech. It is reasonable to assume that speech signals are made up of a sparse combination of basis elements at each discrete time point [12]. This sparsity allows to make a model that each data point Y_{ft} is generated by a Poisson distribution with the mean value $(AX)_{ft}$. The likelihood of the Poisson distribution is

$$(8)\quad \begin{aligned} P_{Y|A,X}(Y \mid A, X) &= \prod_{ft} P_{Y|A,X}(Y_{ij} \mid A, X) \\ &= \prod_{ft} \exp\left[-(AX)_{ft}\right] \frac{(AX)_{ft}^{Y_{ft}}}{Y_{ft}!} \end{aligned}$$

The negative logarithm of both sides is taken to transform the equation into

$$(9)\quad \begin{aligned} L_{Y|A,X}(Y \mid A, X) &= -\log(P_{Y|A,X}(Y \mid A, X)) \\ &= \sum_{ft}\left[(AX)_{ft} - Y_{ft}\log(AX)_{ft} + \log(Y_{ft}!)\right] \end{aligned}$$

Because the value for $\log(Y_{ft}!)$ is constant with respect to A and X, we can drop the term $\log(Y_{ft}!)$, and we will get the cost function as in equation (10)

$$(10)\quad L_{Y|A,X}(Y \mid A, X) \propto \sum_{ft}(AX)_{ft} - \sum_{ft} Y_{ft}\log(AX)_{ft}$$

This corresponds to the KL-divergence cost function as depicted in [5]. Among the tested likelihood term (including those proposed in [5] and [6]), the KL-divergence produced the best results.

3.2 Temporal Smoothness Constraint

The lost speech samples can be accurately estimated using linear interpolation if the time-varying gains of speech are temporally smooth. This means that the prior on X_i should be temporally smooth (X_i is the ith row vector of the time-varying gain matrix X). Here we choose a temporal smoothness constraint term the same as in paper [13]. Suppose each time-varying gain represents a temporal signal with length T, we then write the ith time-varying gain X_{it} as $X_i(t)$. When $X_i(t)$ is temporally locally smooth, its short-term variance is relatively small compared to a larger long-term variance. Let $\hat{X}_i(t)$ denote the short-term exponentially weighted average of $X_i(t)$, namely

$$(11)\quad \hat{X}_i(t) = a\hat{X}_i(t-1) + (1-a)X_i(t), \quad (1 \le i \le r)$$

where $0<a<1$ is forgetting factor that determines the local smoothness range. The temporal smoothness constraint under consideration is measured by the ratio of short-term variance against long-term variance in time domain which is proposed by Stone J. V. [14].

$$(12)\quad \begin{aligned} R &= \log\frac{\sum_t (X_i(t) - \hat{X}_i(t))^2}{\sum_t (X_i(t) - m_i)^2} \\ &= \log\frac{\sum_t a^2(X_i(t) - \hat{X}_i(t-1))^2}{\sum_t (X_i(t) - m_i)^2} \end{aligned}$$

where $m_i = \frac{1}{T}\sum_t X_i(t)$ denotes the mean value of X_{it} given the total T observations. Hence, the smoother the time-varying gain $X_i(t)$, the smaller the ratio value R. We reformulate equation (12) in vector notation in order to impose temporal smoothness constraint on NMF cost function. If we constrain the variance of the time-varying gain as 1, then equation (12) may be equivalently rewritten as

$$(13)\quad \begin{aligned} &R = \frac{1}{T}\left\|X_i - \hat{X}_i\right\|^2 = \frac{1}{T}(X_i - \hat{X}_i)(X_i - \hat{X}_i)^T, \\ &s.t.\ \mathrm{Var}[X_i]{=}1 \end{aligned}$$

The template operator P defined in paper [13] transforms time-varying gain $X_i(t)$ into its smooth counterpart $\hat{X}_i(t)$ as shown in equation (14). P is a $T\times T$ Toeplitz matrix which is only determined by forgetting factor a.

$$(14)\quad \hat{X}_f^{\,T} = PX_f^{\,T}$$

Then R can be reformulated as

$$(15)\quad R = \frac{1}{T}\left\|X_f^{\,T} - PX_f^{\,T}\right\|^2 = \frac{1}{T}\left\|(I-P)X_f^{\,T}\right\|^2$$

Here I denotes $T\times T$ identity matrix. Simulation result shows that the performance of PLC is insensitive to the prior on A. Therefore we only constrain A as positivity.

Hence we can derive negative log posterior $L_{A,X|Y}(A,X \mid Y)$ as follows.

$$(16)\quad \begin{aligned} L_{A,X|Y}(A,X) &\propto L_{Y|A,X}(Y \mid A,X) + L_A(A) + L_X(X) \\ &= \sum_{f=1}^{F}\sum_{t=1}^{T}\left[(AX)_{ft} - Y_{ft}\log(AX)_{ft}\right] + \frac{\lambda}{2T}\sum_{i=1}^{r}\left\|(I-P)X_i^T\right\|^2 \end{aligned}$$

where λ is a small positive weight coefficient that balances the trade-off between the likelihood term and temporal smoothness constraint term.

It has been proved that the negative log posterior function (16) is monotonically non-increasing under the learning rules [13].

$$(17)\quad A_{fi} = A_{fi}\frac{\sum_t X_{it}Y_{ft}/(AX)_{ft}}{\sum_t X_{it}}$$

$$(18)\quad X_{it} = \frac{-\sum_{i=1}^{r}A_{fi} + \sqrt{(\sum_{i=1}^{r}A_{fi})^2 + 4\lambda(\sum_{n=1}^{T}q_{tn})(\sum_{f=1}^{F}Y_{ft}\frac{A_{fi}X_{it}}{\sum_{i=1}^{r}A_{fi}X_{it}})}}{2\sum_{n=1}^{T}q_{tn}}$$

where the denominator of equation (18) $\sum_{n=1}^{T}q_{tn}$ is the column sum of each row of the matrix Q which is defined as $Q = \frac{1}{T}(I-P)^T(I-P)$.

3.3 Interpolation of Time-Varying Gain

In this work the problem of recovery of lost speech packets is transformed into interpolation of the time-varying gains. Assuming only one packet of speech is available before and after the lost packets, the lost time-varying gains can be estimated by using linear interpolation. We define the general linear interpolating function as

$$(19)\quad X_{it} = \frac{X_{ib}(T_{lost} - t + 1) + X_{ia}t}{T_{lost} + 1}, \quad 1 \le i \le r, 1 \le t \le T_{lost}$$

where X_{ib} is the first value before the lost packets, X_{ia} is the final value after the lost packets, X_{it} is the estimated value of time-varying gain and T_{lost} is the consecutive lost packets number. This is showed in Fig.3. Note that the subscripts i, t represent the source index and packet index, respectively. Time-varying gain X_{it} is estimated using linear interpolation of the values on both sides as shown in Fig.4.

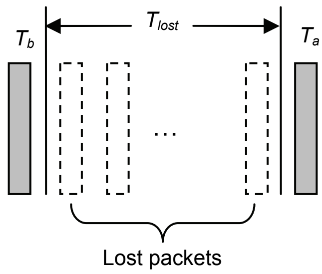

Fig.3. Illustration of T_{lost} missing packets of speech signal.

The linear interpolation function is derived in two instances.

For $X_{ia}>X_{ib}$ We obtain

$$X_{it}=X_{ib}+\frac{X_{ia}-X_{ib}}{T_{lost}+1}t=\frac{X_{ib}(T_{lost}-t+1)+X_{ia}t}{T_{lost}+1},\quad 1\le i\le r, 1\le t\le T_{lost} \tag{20}$$

For $X_{ia}\le X_{ib}$ We have

$$X_{it}=X_{ib}-\frac{X_{ib}-X_{ia}}{T_{lost}+1}t=\frac{X_{ib}(T_{lost}-t+1)+X_{ia}t}{T_{lost}+1},\quad 1\le i\le r, 1\le t\le T_{lost} \tag{21}$$

In this two instances the linear function is identical to Equation (19).

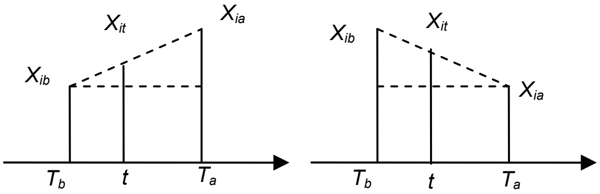

Fig.4. Linear interpolation of lost time-varying gains.

4 Simulation and Results

The proposed algorithm is evaluated both with and without temporal smoothness constraint of time-varying gains. The abbreviation NMFstandard and NMFsmooth are used to represent the proposed PLC algorithm with and without smoothness constraint, respectively. The results are compared to the other conventional PLC algorithms. One of these algorithms is the state-of-art HMM-based PLC algorithm proposed in [1]. The HMM-based PLC algorithm tracks the evolution of speech signal parameters and performs PLC within a statistical signal processing framework. The abbreviation HMM is used to refer to this algorithm. Another comparative PLC algorithm is a hybrid of time-scale modification and waveform substitution which is proposed by Y. J. Liang [3]. In this technique the packet loss concealment tries to cover loss packets by using Waveform Similarity Overlap-Add (WSOLA) of speech signal. Thus, WSOLA is used to refer to this algorithm. Furthermore, as it is widely used in the literature as a point of reference, the standard PLC algorithm in ITU-T G.711 [15] is also evaluated. The silence substitution (abbreviated to SS), also known as zero stuffing, is used as a baseline algorithm here. The results are presented using objective test and signal examples.

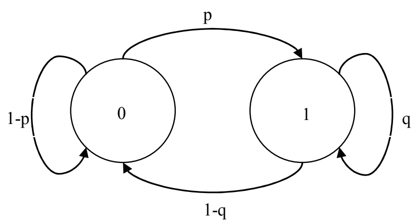

Fig.5. Gilbert packet loss model

4.1. Packet Loss Model

Packet losses are introduced by the Gilbert model [4] which is a 2-state Markov model as shown in Fig.4. p is the probability that a packet will be dropped given that the previous packet has been received, (1-q) is the probability that a packet will be received given that the previous packet has been lost. q is termed the “conditional loss probability” which serves as an indicator for the consecutive loss of traffic. Gilbert packet loss model emphasizes the burst-like nature of errors that might occur in some situation. The loss rate, P_{LR}, and the average consecutive packets lost number, L, of the model can be calculated as follow.

$$P_{LR}=\frac{p}{p+1-q}\ \text{and}\ L=\frac{1}{1-q}$$

4.2. Speech Quality Evaluation

After introducing lost packets in speech signals, each signal is recovered using the proposed algorithm, as well as conventional algorithms mentioned earlier. The assessment tool used to evaluate the performance of the PLC algorithms is the Perceptual Evaluation of Speech Quality (PESQ) standard P.862 developed by ITU-T[16]. PESQ is the most accurate tool in the perceptual-based standards that has shown to give reliable evaluation of the subjective quality. The score is given in the range [-0.5 4.5], similar to the standard mean opinion score (MOS) scale. A random packet loss (generated by Gilbert model) test was performed at loss rates between 5 and 40 %. Source number r=40, forgetting factor a=0.9 and weight of smooth constraint λ=0.1 are chosen in a heuristic manner in our experiments. The results are calculated and averaged for a test set of approximately 500 sentences randomly selected from the TIMIT database. Table 1 summarizes the averaged performance of the algorithms in recovery of lost packets.

Table 1.Performance of different PLC algorithms with different packet loss rates

	PESQ				
(p,q)	(.05,.05)	(.1,.1)	(.2,.2)	(.3,.3)	(.4,.4)
P_{LR}(%)	5	10	20	30	40
NMF_{smooth}	3.69	3.5	3.02	2.87	2.70
$NMF_{standard}$	3.65	3.46	2.95	2.81	2.62
HMM	3.54	3.31	2.90	2.67	2.31
WSOLA	3.51	3.15	2.8	2.39	1.89
G.711	3.27	3.11	2.54	2.03	1.52
SS	3.34	2.87	2.45	1.81	1.15

4.3. Discussion

It is obvious from the results presented in Table 1 that the PESQ score of the NMF_{smooth}-based PLC algorithm is generally higher than that of other algorithm for all percentage of packet losses, including the $NMF_{standard}$-based PLC algorithm. The results show that the temporal smoothness constraint of time-varying gains improves the quality of reconstructed speech signal. The reason for this is obvious because time-varying gains with temporal smoothness constraint have small variance than those without. Thus, linear interpolation that estimates the lost time-varying gains is more accurate. Therefore, NMF_{smooth}-based PLC algorithm provides more accurate reconstructed speech signal than $NMF_{standard}$-based PLC algorithm.

Signal examples resulting from the NMF_{smooth}-based PLC algorithm and $NMF_{standard}$-based PLC algorithm under 20% packet losses are shown in Fig.6. We see that the structure of the missing speech segments is well recovered by ours proposed algorithm. Fig.7 shows the spectrograms of a sample speech signal, its distorted (with missing samples replaced with zeros) and restored versions.

It is also observed in Table 1 that the proposed packet loss concealment algorithms in nonnegative subspace show more robustness in high packet loss rate than other algorithms. In this new proposed PLC framework, NMF is introduced to reduce the dimensionality of data and to explain the whole speech signal by a few meaningful elementary objects which make up of the nonnegative subspace. Thanks to the nonnegative constraint, nonnegative subspace is able to provide a meaningful representation of the whole speech signal. Thus, PLC in nonnegative subspace can utilize more global information to reconstructed missing speech samples than conventional PLC algorithms. This reason contributes to the robustness of the new proposed PLC algorithm. The second more efficient PLC method was HMM, followed by WSOLA method and G.711 method. The PESQ score of the proposed algorithm is 0.15 higher than HMM-based PLC method on average. However, these PLC algorithms deteriorate as the loss rate increasing, especially the G.711 PLC algorithm. Although the performance of the G.711 algorithm is good at lower loss rate, it deteriorates greatly as the loss rate is increased. The G.711 PLC algorithm is only designed to deal with 60ms (3 packets) of loss speech samples [15] and it deteriorates very quickly as the packet losses is longer than 3 packets.

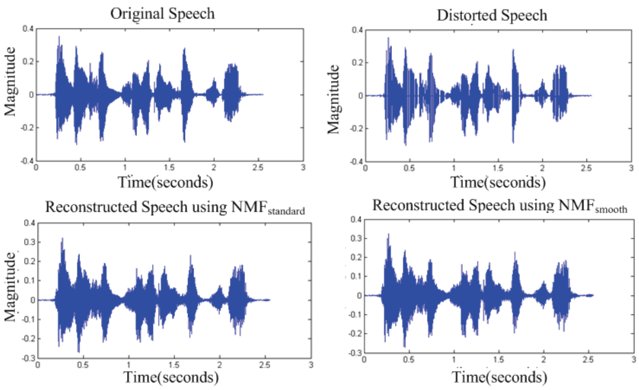

Fig.6. Waveform of a sample signal, with introduction of 20% packet loss and the restored versions.

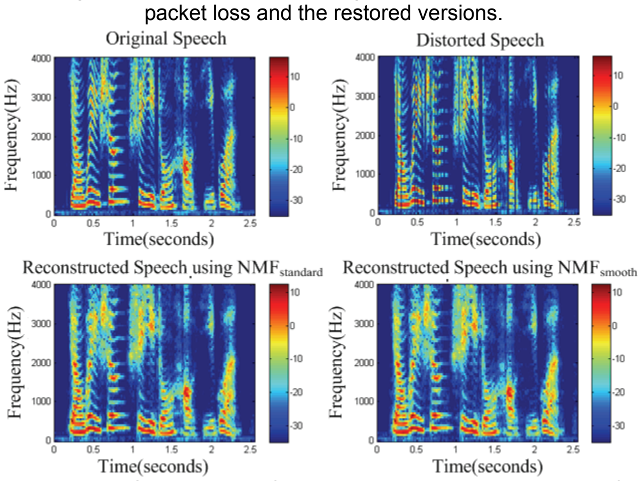

Fig. 7. Spectrogram of a sample signal, with introduction of 20% packet loss and the restored versions.

5 Conclusion

A new algorithm is introduced for packet loss concealment in nonnegative subspace by using interpolation of the trajectory of time-varying gains. It is shown that the interpolation of the time-varying gains, used for modeling the temporal activation of speech, results in more natural sound quality than conventional algorithms, including the state-of-art HMM-based PLC algorithm. Furthermore, temporal smoothness constraint is employed to make the interpolation more accurate to the estimation of time-varying gains. The performance of the proposed PLC algorithm may be improved by introducing more constraint conditions such as sparseness [7] or harmonic [8] in NMF technique which would reduce the reconstruction error for speech signal. This is being investigated for further improvement of the proposed PLC algorithm.

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] C. A. Rødbro, M. N. Murthi, S. V. Andersen, and S. H. Jensen, “Hidden Markov model-based packet loss concealment for voice over IP,” *IEEE Trans. Audio, Speech, Lang. Process.*, 14(2006), No. 5, 1609-1623

[2] D. J. Goodman, G. B. Lockhart, O. J. Wasem, and W. C. Wang, “Waveform substitution techniques for recovering missing speech segments in packet voice communications,” *IEEE Trans. Acoustics, Speech, Signal Process.*, 34(1986), No. 5, 1440-1448

[3] Y. J. Liang, N. Färber, and B. Girod, “Adaptive playout scheduling and loss concealment for voice communication over ip networks,” *IEEE Trans. Multimedia*, 5(2003), No.2, 532-543

[4] Esfandiar Zavarehei and Saeed Vaseghi, “Interpolation of Lost Speech Segments Using LP-HNM Model With Codebook Post-Processing” *IEEE Trans. Multimedia*, 10(2008), No.3, 493-502

[5] D.D. Lee and H.S. Seung, “Learning the Parts of Objects by Nonnegative Matrix Factorization,” *Nature*, 401(1999), No. 6755, 788-791

[6] D. D. Lee and H. S. Seung, “Algorithms for nonnegative matrix factorization,” *Neural Inf. Process. Syst.*, 2001, 556-562

[7] T.Virtanen, “Monaural sound source separation by nonnegative matrix factorization with temporal continuity and sparseness criteria,” *IEEE Trans. Audio, Speech, Lang. Process.*, 15(2007), No. 3, 1066-1074

[8] E. Vincent, N. Bertin, and R. Badeau, “Harmonic and inharmonic nonnegative matrix factorization for polyphonic pitch transcription,” *Proc. ICASSP*, 2008, 109-112

[9] Alexey Ozerov and Cédric Févotte, “Multichannel nonnegative matrix factorization in convolutive mixtures for audio source separation,” *IEEE Trans. Audio, Speech, Lang. Process.*, 18(2010), No. 3, 550-563

[10] Nancy Bertin, Roland Badeau, and Emmanuel Vincent, “Enforcing harmonicity and smoothness in bayesian non-negative matrix factorization applied to polyphonic music transcription,” *IEEE Trans. Audio, Speech, Lang. Process.*, 18(2010), No. 3, 538-549

[11] Romain Hennequin, Roland Badeau and Bertrand David, “NMF with time-frequency activations to model non stationary audio events,” *Proc. ICASSP*, 2010, 445-448

[12] Asari, H., Pearlmutter, B. A., and Zador, A. M., “Sparse representations for the cocktail party problem”. *J Neurosci.* 26(2006), No.28, 7477–7490

[13] Zhe Chen, Cichocki, A., Rutkowski, T.M., “Constrained non-Negative Matrix Factorization Method for EEG Analysis in Early Detection of Alzheimer Disease,” *Proc. ICASSP*, 2006, 109-112

[14] Stone J. V. “Blind source separation using temporal predictability,” *Neural Computation*, 13(2001), 1559-1574

[15] Appendix I: A High Quality Low-Complexity Algorithm for Packet Loss Concealment with G.711. *ITU-T Recommendation- G.711*, 1999.

[16] Perceptual Evaluation of Speech Quality (PESQ), an Objective Method for End-to-End Speech Quality Assessment of Narrowband Telephone Networks and Speech Codecs, *ITU-T Recommendation-862*, 2001.

***Authors**:*
Jianjun HUANG, Postgraduate Team 2, Haifu Xiang 1, Baixia District, Nanjing, China, 210007, E-mail: hjj954@gmail.com;
Yafei ZHANG, Institute of Command Automation, PLA University of Science and Technology.
Xiongwei ZHANG, Institute of Command Automation, PLA University of Science and Technology.
Zou XIA, Institute of Command Automation, PLA University of Science and Technology.