Nonlinear Echo Cancellation Based on Polyphase Filter Bank

Meng Liang*, Zhong-Hua Fu*, Xiang Zhao‡, Jinglei Zhou†, Haikun Wang‡
* School of Computer Science and Technology, Northwestern Polytechnical University, Xi'an, Shaanxi, China
† School of Electronics and Information, Xi'an Polytechnic University, Xi'an, China
‡ Xi'an IFLYTEK Hyper Brain Information Technology Co., Ltd.
E-mail: {mengliangccc@mail, mailfzh@}nwpu.edu.cn, xiangzhao5@iflytek.com, jlzhou@xpu.edu.cn, hkwang@iflytek.com

***Abstract*—In hands-free telephone systems and mobile communication devices, it is often desirable for the devices to operate at large sound volume, which can result in obvious nonlinear acoustic echoes due to overload of small loudspeaker.These nonlinear echoes can't be fully eliminated by linear AEC (Acoustic Echo Cancellation) algorithm, so the conversation quality is affected seriously. Since the nonlinear echoes contain additional harmonics in high frequency, which breaks the linear relation required for fullband linear AEC, these harmonics, however, becomes additive noise in subband system. Therefore, in this paper, a subband AEC method based on polyphase filter-bank is proposed. It is found that ERLE (Echo Return Loss Enhancement) of the proposed method outperforms the fullband counterpart constantly in nonlinear situation, especially with tone-like signals, where ERLE improves more than 15 dB. The results are validated through both simulation and real signals.**

I. Introduction

Hand-free interaction with mobile smart devices is very common nowadays. Those devices, due to size limitation, are usually equipped with small loudspeakers. To produce sound with sufficient level, these loudspeakers often work near their saturation range, which results in obvious nonlinear acoustic echoes that are very difficult to eliminate using typical linear AEC algorithm. The general set-up of an acoustic echo cancellation system is illustrated in fig.1. The signal sent to the loudspeaker is $x(n)$ and the signal recorded by microphone is $y(n)$. Even if there is unnoticed nonlinear distortion between the two signals, the convergence of the linear adaptive filter will be significantly affected [1][2]. Thus how to deal with the nonlinear echo problem becomes a challenge issue in these applications.

The typical solution to handle nonlinear echo problem is to model the nonlinear echo path and transform the nonlinear problem into linear form, such as the Volterra series model [3][4] and the neural network (NN) model [5][6]. The Volterra series represents the nonlinear relation in polynomial form, which generally requires sufficient higher-order filter to effectively model the nonlinear echo path. So the computation load increases hugely. The NN model, as a data-driven method, requires a large amount of data, including both the excitation signal and the echo signal, to train adequately. Since the nonlinearity of the electric-acoustic system is coupled with acoustic echo path, the generalization capability of the model is a challenge issue. The most arguable issue of these solutions is that the characteristics and the causes of the nonlinear echo are complicated and lack of proof for universal representation, so the solution of nonlinear echo problem is still on its way.

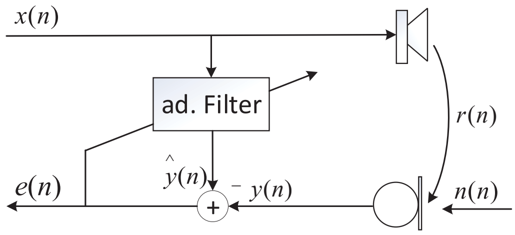

Fig. 1. General set-up of the acoustic echo cancellation problem.

Instead of trying to model the nonlinear echo path, in this paper, we transfer the fullband AEC problem to subband AEC problem. It is known that the nonlinearity of loudspeaker leads to intermodulation distortion, which results in additional harmonics appeared in high frequency [7][8]. For the fullband adaptive filter, the cost function is defined based on the fullband error, and the nonlinear relation exists in any part of the frequency domain will affect the full-band error, thus the adaptive filter convergence. However, for the subband adaptive filter, since each band is updated independently according to the error signal defined on that band, the harmonics occur in higher band simply becomes additive "noise" of that band. Therefore those bands without harmonics can converge normally and the convergence on those bands with harmonics depend on the level of the additive "noise" [9].

In this paper, we firstly analyse the nonlinear component in the signal path in AEC system, its characteristics, and how this nonlinearity becomes linearity in the subband scheme. Then the subband AEC method using polyphase filter bank is proposed. To verify the effectiveness of the subband method in nonlinear echo situations, we conduct AEC experiments using both simulation and real recording data. Comparing with the fullband method, the ERLE of the subband method is constantly better, and as the nonlinearity becomes more obvious, the improvement is more significant, especially using tone-like signal.

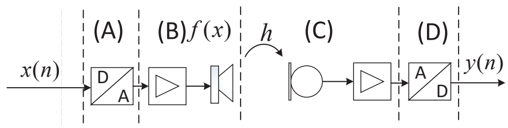

Fig. 2. Model of the nonlinear echo path.

II. Nonlinearity in AEC system

In this part, we discuss three issues: the components in the signal path in AEC system, the factors that influence the nonlinearity of loudspeaker, and why subband scheme can work, especially with tone-like signal.

A. Signal path in AEC system

The basic idea of linear AEC is to simulate the echo signal sent from loudspeaker and recorded by microphone using a linear adaptive filter, and then subtract it from the real recording signal to cancel the real echo. The crucial requirement is the signal path from D/A converter in play-out chain to the A/D converter in the recording chain must be linear. As shown in fig.2, the whole signal path contains four parts from (A) to (D) [10], where (A) is D/A converter, (B) loudspeaker unit (power amplifier and loudspeaker), (C) acoustic echo path, and (D) A/D converter.

In modern digital communication systems, the accuracy of D/A and A/D converters is getting higher and higher, and the nonlinear distortion caused by them is negligible. The acoustic echo path in part (C) can be modeled as a linear time-varying system, which can be identified using linear adaptive filter. Therefore, the major cause of nonlinearity is the loudspeaker unit, which is a physical transducer to converse electric signal to acoustic signal. It generally has its linear work range and out-off that range, nonlinearity will be significant.

B. Factors that influence the nonlinearity of loudspeaker

As a commonly used electric-to-acoustic transducer, loudspeaker's input-to-output response is correlated with signal volume and frequency as well. The typical rules, as demonstrated in the experiment section, are as follows:

(1) The louder the volume, especially close to the saturation region, the overload of the loudspeaker shows more nonlinearity. We test the loudspeaker of a laptop computer and find that when the SPL (Sound Pressure Level) exceeds 80 dB, more significant harmonic distortions can be found.

(2) The nonlinearity is related with signal frequency and is different between different types of loudspeakers. Generally, with panel speaker or MEMS speaker, the response of lower frequency and higher frequency, especially the former, are worse. For example, using sinusoid signals played by the laptop computer with same SPL, the nonlinear characteristics of low-frequency signal are more obvious than high-frequency signal. More and denser harmonics are observed with the low-frequency signal.

C. Why subband method

For simplicity, let's consider the reference signal $x(n)$ has two distinct frequency components, f_a and f_b. The frequency signal components are $X(f_a)$ and $X(f_b)$. If the signal path is linear and modeled as $H(f)$, then the recording signal $y(n)$ in the frequency domain, i.e. $Y(f)$ also has two distinct frequency component, $Y(f_a)$ and $Y(f_b)$, and they satisfy that

$$Y(f_a) = X(f_a) \cdot H(f_a), Y(f_b) = X(f_b) \cdot H(f_b). \quad (1)$$

By taking $X(f)$ as input and $Y(f)$ as output signal, the linear model $H(f)$ at f_a and f_b can be easily identified.

However, if a nonlinear function $f(\cdot)$ is embedded into the signal path, then the recording signal will have more frequency components than in linear case, such as $Mf_a \pm Nf_b$, where M and N are certain integers that depend on the nonlinearity. As the result, for these new frequency components, there is no corresponding input signal in linear system identification scheme, and the fullband error cannot converge, so the full-band adaptive filter will fail.

If the fullband problem is transferred to subband problem, and let the bands be denoted as B_0, B_1, and so on, then the new frequency components, as well as the original frequency components, will be separated into different bands, especially in higher frequency range where the harmonics are more sparse. If the bands are narrow enough, then the bands will contain: (1) one of the original frequency component, such as f_a, or (2) one of the original frequency component and some distorted component related to other bands, such as f_b plus Mf_a or (3) some distorted component only, such as Nf_b or (4) no effective frequency components. Since the adaptive filters are performed in each band independently, the case (1) will converged normally, the case (2) becomes linear system identification with additive noise and its convergence will be influenced by the SNR in that band, the case (3) will fail and the case (4) has no information to learn but no residual echo will occur. Hence overall, the convergence of subband will be better than that of fullband.

Apparently, the subband method doesn't solute the nonlinear echo problem, but by separating the signal into different subbands, the linear relation between the input and output signals is kept in some bands, and echoes in those bands can be cancelled well. One more word, any solution conducted to the fullband nonlinear problem can be introduced to the subband method.

III. Subband AEC based on polyphase filter bank

A. Subband acoustic echo cancellation

The structure of a typical subband AEC system is shown in fig.3 [11], where each band has its own adaptive filter, and all adaptive filters are automatically updated to make the final error signal as small as possible. The input and the desired signals are decomposed into a number of subband signals using an analysis filter bank, then after filtering and cancellation, the subband residual echo signals are sent to a synthesis filter bank to reconstruct the final fullband residual error signal.

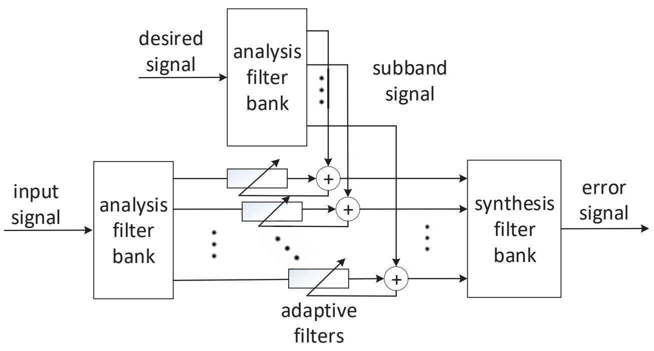

Fig. 3. Set-up of the subband echo cancellation.

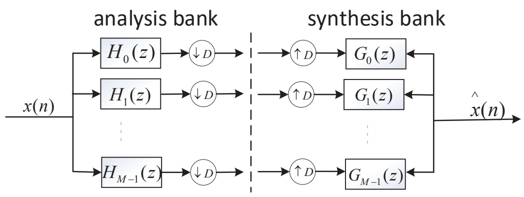

Fig. 4. Analysis and synthesis branch of a M-channel filter bank with subbands decimated by D.

Obviously, if this structure is directly deployed, the computational load is very high. Notice that the bandwidth of each subband signal is narrow, so a very low sampling frequency can be used in each band. An M-channel analysis and synthesis filter bank with decimator of D is shown in Fig.4. The key issues include how to reduce aliasing between neighbouring bands, and how to reduce the reconstruction error. The most popular filter bank with efficient computation cost is the polyphase structure [12].

B. Efficient implementation with polyphase structure

The polyphase filter bank used in this study is a typical uniform DFT-modulated filter bank, where a prototype lowpass filter is designed to satisfy several constraints and by modulating it to different frequency positions, a group of uniform subband filters are obtained. The special advantage of this kind of filter bank is that FFT can be involved to accelerate the calculations of subband analysis and synthesis.

The calculation structure is shown in fig.5, where h_{proto} is the prototype filter, $\tilde{h}_{\text{proto}}$ is the time reversed copy of h_{proto} [9]. For subband analysis, the input signal is windowed by the prototype filter and reshaped before implementing FFT, then after FFT, the decimated subband signals are obtained. For subband synthesis, the subband signals are processed in opposite direction of the analysis steps.

Generally, we want as many as subbands as possible, no aliasing between neighbouring bands, and almost perfect reconstruction capability. However, these requirements are contradictory and need trade-off in practice. The performance of the filter bank depends on the characteristics of the prototype filter.

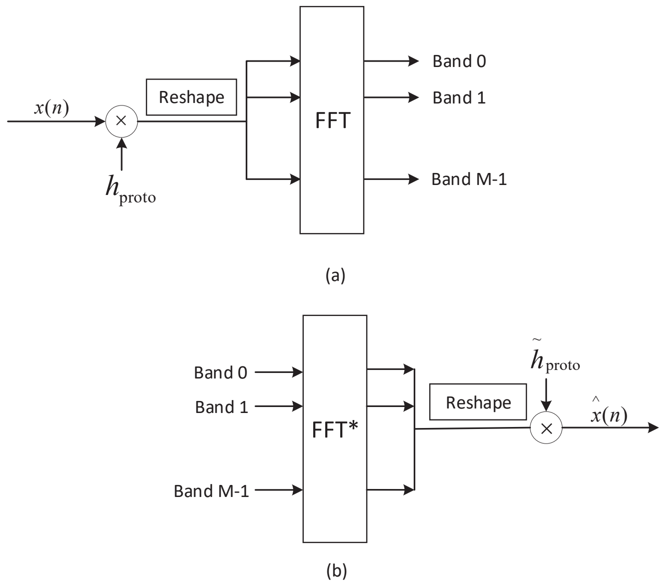

Fig. 5. The subband analysis (a) and synthesis (b) using DFT-modulated polyphase filter bank.

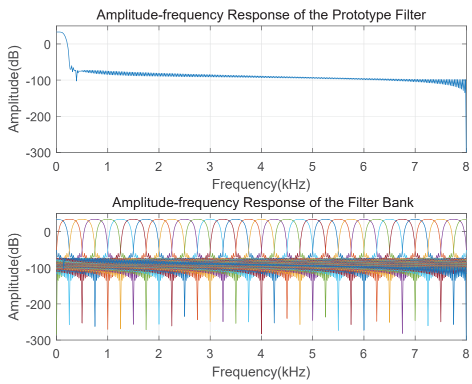

Fig. 6. Amplitude-frequency response of the prototype filter and filter bank.

C. Design of the prototype filter

There are lots of research works on the prototype filter design in the literature. It is a conditional optimization problem and the constraints are related to specific applications. In this paper, we use the iterative least squares method [13], which provides a flexible way to control the design parameters.

It is known that for subband AEC, the aliasing problem is very crucial [14]. To reduce the aliasing, the transition bandwidth should be very small, so the length of the prototype has to increased and the reconstruction error needs to be relaxed. Additionally, the stopband attenuation can also be relaxed to reduce the aliasing. The frequency responses of the prototype filter and the filter bank used in our experiments are shown in fig.6. There are 64 subbands from 0 Hz to the sampling frequency (16kHz in this paper), with decimator $D = 32$ and filter length of 512. The reconstruction error is about -100 dB, and the aliasing is about -92 dB.

IV. Experiments

In this section, we conduct three types of experiments to verify our analysis. The first experiment uses simulated data, where a typical nonlinear function, clipping, is considered, and a pure tone signal as well as white noise signal are involved. The purpose is to show the harmonics in higher frequency and the echo cancellation performance in this ideal situation. The second experiment uses real data recorded in an anechoic room to show the characteristics of loudspeaker, where a laptop computer is involved to examine the nonlinear AEC problem. In the third experiment, the recording is performed in an ordinary office, and the subband solution is compared with the fullband solution. The laptop computer is DELL G7 7588, and a sound-level meter AWA 5636 is used to measure the sound level of its loudspeaker.

Note that in this study, we just want to compare the performances of fullband and subband adaptive filters in nonlinear echo cancellation. So we don't use speech signal as the test signal, to avoid complicated AEC algorithm optimization and double-talk detection. Here, the basic NLMS algorithm is perform in fullband, or in each subband, with common learning factor 1. The AEC performance is evaluated using ERLE, which is defined as

$$ERLE(n) = \frac{E\left[|y(n)|^2\right]}{E\left[|e(n)|^2\right]}, \tag{2}$$

where $e(n)$ is the residual echo signal after cancellation. The higher ELRE, the smaller residual echo and hence the better AEC performance.

A. Experiment 1:Comparison using simulated data

The clipping function involved in the simulation is

$$f(x) = \begin{cases} 1 & ax > 1 \\ ax & -1 \le ax \le 1 \\ -1 & ax < -1 \end{cases} \tag{3}$$

where a is a factor that controls the amount of samples been clipped. The pure tone signal is a 1000 Hz sinusoid signal, with $a = 1.5$, and the white noise is with $a = 6$. For both of the signals, 40% samples are clipped. The clipped signals are convoluted with a simulated room impulse response to obtain the microphone signal $y(n)$, respectively. The sampling frequency is 16000 Hz. The durations are all 20s. As shown in fig.7, the clipped tone signal contains more harmonics in higher frequency, such as in 3000 Hz, 5000 Hz and 7000 Hz. This will seriously affect the convergence of the fullband AEC. When using subband solution, these frequency components are separately located in different subbands, and the original 1000 Hz component is located in the 5th subband isolated. The linear relation is maintained in that band.

The ERLEs are shown in fig.8. It is clear that the fullband NLMS is almost totally failed, and the subband NLMS works almost perfectly. Note that the length of all adaptive filters are long enough, so that in linear condition, the ELRE can almost approach infinite. However, with white noise, as shown in fig.9, the performance difference is not so significant while the subband solution is still better. The reason is that the white signal contains too many frequency components and even if in subband, the ideal linear case is almost not existent.

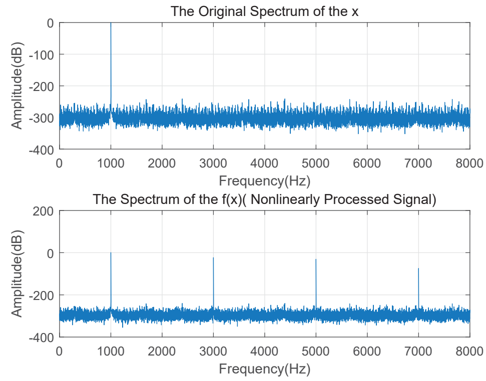

Fig. 7. The original spectrum of the x and the spectrum of the $f(x)$(nonlinearly processed signal).

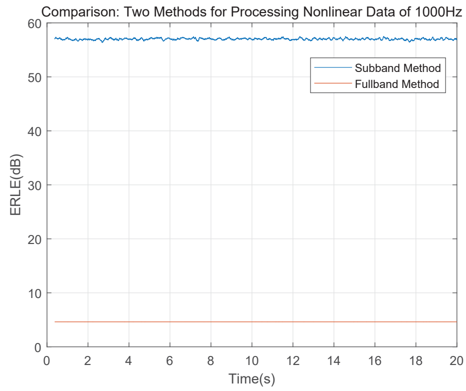

Fig. 8. Signal x is a 1000Hz single-frequency signal, ERLE comparison.

B. Experiment 2:Comparison using data recorded in anechoic chamber

In this experiment, we want to examine the nonlinearity of the laptop loudspeaker. The test signal, 1000 Hz sinusoid signal and white noise are respectively recorded in an anechoic chamber, where the acoustic path influence can be neglected. The sound volume of the loudspeaker system is adjusted to achieve different sound levels, measured using the sound-level meter. The recorded signal is then convoluted with the simulated room impulse response to get the final microphone signal $y(n)$. The ERLEs with different sound levels are shown in fig.10 and fig.11, for the sinusoid signal and the white signal respectively.

In fig.10, one can find that from 40 dB to 80 dB, both of the fullband and the subband methods achieve steady improvement in ERLEs. The reason is that the echo to environment noise ratio is increased with the SPL. The environment noise is mainly from the laptop fan. As expected, when the SPL approaches 90 dB, which is almost the maximum volume of the laptop, the performance of the fullband method drops

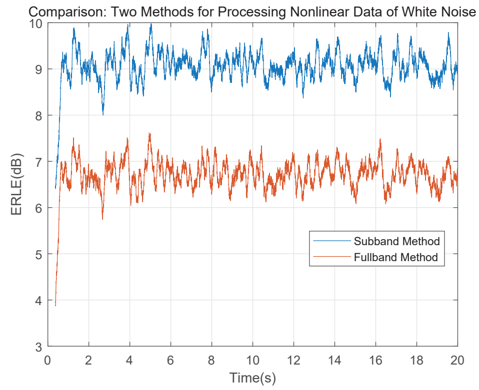

Fig. 9. Signal x is a white noise signal, ERLE comparison.

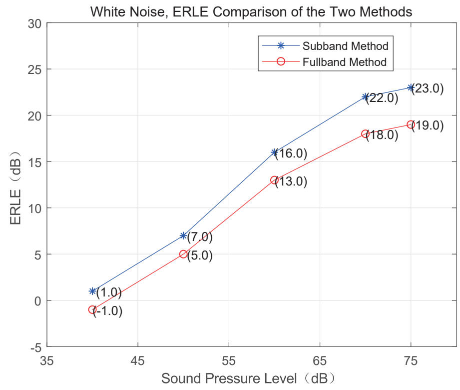

Fig. 11. Signal x is a white noise signal, ERLE values at different SPLs.

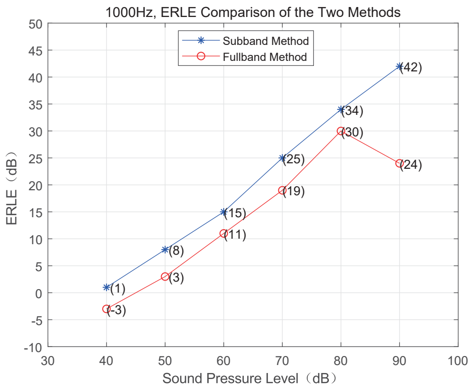

Fig. 10. Signal x is a 1000Hz single-frequency signal, ERLE values at different SPLs.

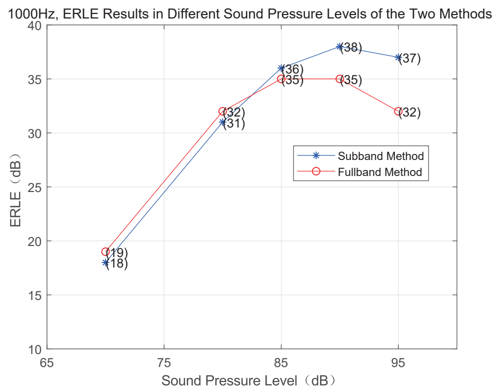

Fig. 12. Signal x is a 1000Hz single-frequency signal, ERLE values at different SPLs.

significantly due to saturation nonlinearity. But the performance of the subband methods keeps increasing. The results of white noise are slightly different. As examined in the first experiment, the subband method in this case is not as good as using tone signal. However, as the SPL increases, the ELRE difference between subband and fullband methods are slightly enlarged.

C. Experiment 3: Comparison using data recorded in an office

In this experiment, we test the AEC performance in an ordinary quiet office. The test signals, beyond the previously used 1000 Hz sinusoid signal and white noise signal, another tone-like signal with two distinct frequency components, 1000 Hz plus 2500 Hz, is also involved. The purpose is to examine the influence of intermodulation distortion of the laptop loudspeaker. The system volume is also adjust to different levels using sound-level meter. The experimental results are shown in fig.12 - fig.14.

In fig.12, the 1000 Hz pure tone signal is used. The similar results as in the experiment 2 are obtained. When the SPL is lower than 80 dB, the ERLEs are increased with the SPL due to increase of SNR, where both methods works well. But when the SPL further increases above 85 dB, the performance of fullband method stop increasing and drops, and the subband shows better performance and drops slowly. The more convincing results are found in fig.13, where intermodulation distortion occurs. The total performances are worse than single pure tone, but the subband method show significantly better performance. It is interesting to found that even with modest sound level, the fullband method performance is very poor. That means the fullband method is more sensitive to the intermodulation distortion. In fig.14, as expected, with white noise, both methods fail in case of high nonlinearity.

V. Conclusions

In this paper we study the nonlinear echo cancellation problem. Instead of trying to model the complicated nonlinear system, we consider the subband AEC solution using polyphase filter bank. While the convergence of the fullband adaptive filter is controlled by the fullband error signal, the subband solution depends on the error signal in each narrow subband. We confirm that the nonlinearity, when transformed to subband, is either vanished or weakened in some subbands, and the adaptive filters of those bands can converge. Thus the AEC performance becomes better. We also analyzed the nonlinearity characteristics of loudspeaker. All the analysis results are verified in three types of experiments. It is found that nonlinearity of loudspeaker is more significant as the

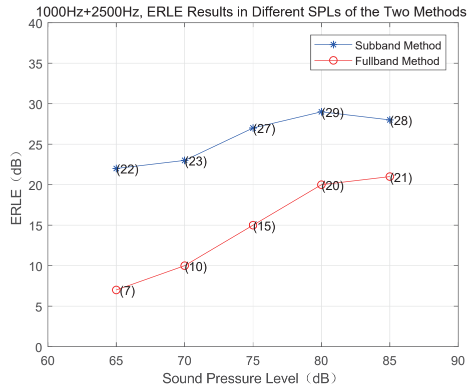

Fig. 13. Signal x is a 1000Hz+2500Hz signal, ERLE values at different SPLs.

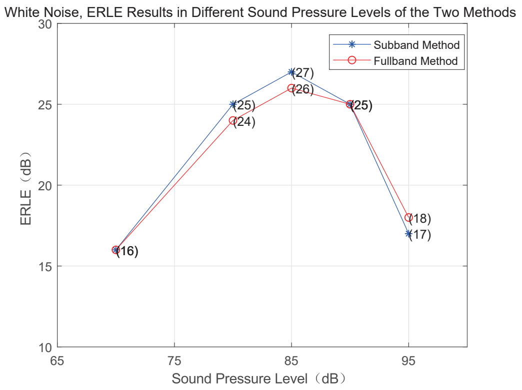

Fig. 14. Signal x is a white noise signal, ERLE values at different SPLs.

sound level approaches to its saturation range, and subband AEC constantly provides better ERLEs than the fullband method, especially with tone-like signals and intermodulation distortion.

Acknowledgment

The authors like to thank Prof.Jingdong Chen, Northwestern Polytechnical University, for fruitful discussions and suggestions,and the research work is supported financially by the National Key Research and Development Program of China under Grant No. 2017YFB1002102.

References

[1] M. H. Costa, J. C. M. Bermudez, and N. J. Bershad, "Statistical analysis of the lms algorithm with a zero-memory nonlinearity after the adaptive filter," in *1999 IEEE International Conference on Acoustics, Speech, and Signal Processing. Proceedings. ICASSP99 (Cat. No.99CH36258)*, vol. 3, March 1999, pp. 1661–1664 vol.3.

[2] N. J. Bershad, "On error-saturation nonlinearities in lms adaptation," *IEEE Transactions on Acoustics, Speech, and Signal Processing*, vol. 36, no. 4, pp. 440–452, April 1988.

[3] F. Kch and W. Kellermann, "Nonlinear line echo cancellation using a simplified second order volterra filter," in *2002 IEEE International Conference on Acoustics, Speech, and Signal Processing*, vol. 2, May 2002, pp. II–1117–II–1120.

[4] A. Stenger, L. Trautmann, and R. Rabenstein, "Nonlinear acoustic echo cancellation with 2nd order adaptive volterra filters," in *1999 IEEE International Conference on Acoustics, Speech, and Signal Processing. Proceedings. ICASSP99 (Cat. No.99CH36258)*, vol. 2, March 1999, pp. 877–880 vol.2.

[5] A. N. Birkett and R. A. Goubran, "Acoustic echo cancellation using nlms-neural network structures," in *1995 International Conference on Acoustics, Speech, and Signal Processing*, vol. 5, May 1995, pp. 3035–3038 vol.5.

[6] ——, "Nonlinear echo cancellation using a partial adaptive time delay neural network," in *Proceedings of 1995 IEEE Workshop on Neural Networks for Signal Processing*, Aug 1995, pp. 449–458.

[7] D. R. Morgan, J. L. Hall, and J. Benesty, "Investigation of several types of nonlinearities for use in stereo acoustic echo cancellation," *IEEE Transactions on Speech and Audio Processing*, vol. 9, no. 6, pp. 686–696, Sep. 2001.

[8] A. N. Birkett and R. A. Goubran, "Limitations of handsfree acoustic echo cancellers due to nonlinear loudspeaker distortion and enclosure vibration effects," in *Proceedings of 1995 Workshop on Applications of Signal Processing to Audio and Accoustics*, Oct 1995, pp. 103–106.

[9] J. Benesty, T. Gaensler, D. R. Morgan, M. M. Sondhi, and S. Grant, *Advances in Network and Acoustic Echo Cancellation*, 01 2001.

[10] F. Kuech, A. Mitnacht, and W. Kellermann, "Nonlinear acoustic echo cancellation using adaptive orthogonalized power filters," in *Proceedings. (ICASSP '05). IEEE International Conference on Acoustics, Speech, and Signal Processing, 2005.*, vol. 3, March 2005, pp. iii/105–iii/108 Vol. 3.

[11] S. Weiss, L. Lampe, and R. W. Stewart, "Efficient subband adaptive filtering with oversampled gdft filter banks," in *IEE Colloquium Adaptive Signal Processing for Mobile Communication Systems (Ref. No. 1997/383)*, Oct 1997, pp. 4/1–4/9.

[12] P. P. Vaidyanathan, "Multirate digital filters, filter banks, polyphase networks, and applications: a tutorial," *Proceedings of the IEEE*, vol. 78, no. 1, pp. 56–93, Jan 1990.

[13] M. Harteneck, S. Weiss, and R. W. Stewart, "Design of near perfect reconstruction oversampled filter banks for subband adaptive filters," *IEEE Transactions on Circuits and Systems II: Analog and Digital Signal Processing*, vol. 46, no. 8, pp. 1081–1085, Aug 1999.

[14] W. Kellermann, "Analysis and design of multirate systems for cancellation of acoustical echoes," in *ICASSP-88., International Conference on Acoustics, Speech, and Signal Processing*, April 1988, pp. 2570–2573 vol.5.