Multi-stage Declipping of Clipping Distortion Based on Length Classification of Clipped Interval *

☆ Chenlei Li, Shigeki Miyabe, Takeshi Yamada, Shoji Makino (University of Tsukuba)

1 Introduction

In order to record high-quality sounds with low noise such as quantization noise, electric hum etc, the gain level should be suitable when signal input. However, the high input gain often causes an over-level distortion in recorded signal. In the instant when the maximum input range of a digital acquisition system is exceeded, the over range amplitude of the original signal is truncated by the limit of highest acceptable level, and the recorded signal is shown in Fig. 1. This phenomenon is called clipping effect, and the resultant is annoying for hearing. Since the automatic gain adjustment is not an easy problem, our research focuses on the restoration of the clipped signal.

It is difficult to restore a waveform of clipped signal to the original one perfectly, especially the clipped signal is deteriorated severely from the original. However, it is possible to estimate the original amplitude partially by using the prior knowledge on the target signal. Classical approaches to audio interpolation and declipping by using prior knowledge include autoregressive (AR) modeling[1], signal matching with bandwidth constraints by utilizing the property of oversampling[2], and statistical assumption on signal can also be used[3].

The problem of restoration for generic clipped signal without prior knowledge on the target signal is much harder, and there have been few works on this problem. However, each of these methods have particular problems. For example, the time domain method based on orthogonal matching pursuit (OMP)[4] cannot restore the clipped signal with high quality when the clipping level is low (i.e, the clipping is severe), and Miura's declipping method[5] in the time-frequency domain causes large noise in the high frequencies.

We focus on the characteristics of the waveforms in different length of clipped intervals, most of declipped waveforms of conventional methods in short clipped intervals are too complicated, they are not

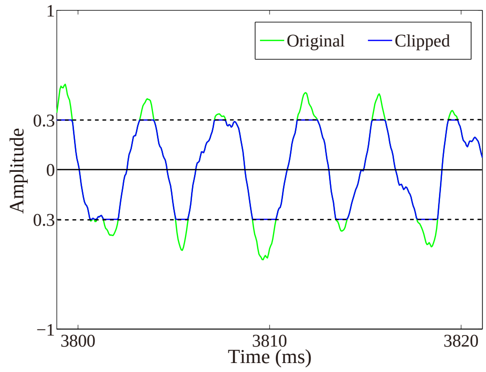

Fig. 1 A signal and its clipped version, the clipping level is 0.3 and the maximum amplitude is 1.

as simple and smoothness as the original ones. For this reason, we propose a new restoration scheme to restore the short and the long clipped intervals separately using different methods after the classification of the lengths. We restore the short intervals by a simple spline interpolation, and restore the long intervals by the conventional methods. After the evaluation by SNR, we improve the performance of Miura's method in all clipping levels, and we also improve the performance of the OMP in low clipping level from 0.1 to 0.2 for speech and 0.1 to 0.4 for music.

2 The Conventional Method

2.1 Orthogonal Matching Pursuit

The OMP is a novel sparse representation based approach for the restoration of clipped audio signals, it is a fitting of short-time-frame signal to limited number of cosine bases. The restoration estimate lost samples based on consistency with the clean signal in the neighboring samples. By fitting the bases to the neighboring samples of the clipped ones, signal restoration of the clipped samples is obtained as a by-product of the fitting. Since the restoration is estimated to be consistent with the neighboring samples, the restored signal is free from the clicking noise. But in low clipping level, as Fig. 2 shows, the performance of OMG is not as good as it is in

* クリッピング区間の長さ分類に基づいたクリッピング歪み多段除去、李晨雷、宮部滋樹、山田武志、牧野昭二 (筑波大)

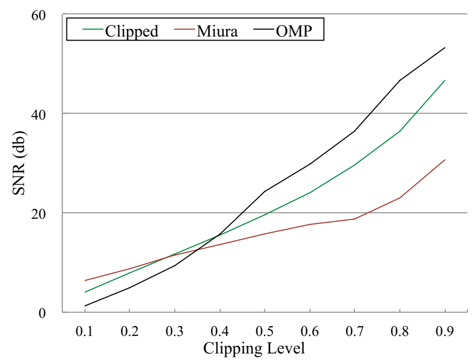

Fig. 2 Average SNR for 10 music and 10 speech signals of Miura's method, OMP and clipped signal。

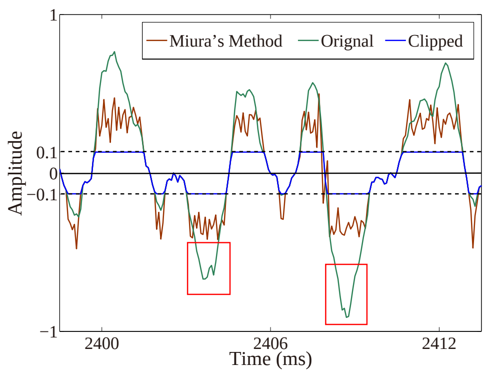

Fig. 3 A result sample of Miura's method, the clipping level is 0.1. The red rectangles is the simple clipped interval which contains two peaks.

high clipping levels, because the candidates selection becomes inaccurate when the referable neighboring samples become less and less with the decrease of clipping level, and the local consistency becomes weak too.

2.2 Miura's method

Miura's method is a restoration of clipped acoustic signal based on the consistency among time-frequency grids. It is a fitting of spectrogram to limited number of 2-dimensional cosine bases. First, the clipped signal is transformed into time-frequency domain by short-time Fourier transformation. Second, select the 2D discrete cosine basis from prepared candidates, and then multiply low-pass filter to obtain desired frequency characteristics. In this method, the global consistency is maintained by smooth frame transition, but it doesn't have a mechanism of maintaining the local consistency. As Fig. 3 shows, because of weak local consistency, there are lots of annoying high-frequency noise in the estimated signal.

3 The Proposed Method

As we know, the cue to estimate the missing clipped interval is the estimated waveform must be consistent with given waveform in unclipped interval. There are two consistency relationships, one is the local consistency, it means any part is consistent with adjacent parts on both sides, and the other one is global consistency, any two parts are consistent in the whole clipped signal. If the local and global consistency are kept in declipping result, we can get a continuous and smooth waveform which we demand. Miura's method maintain the global consistency by spectrum, but no mechanism for local consistency. The reason of insufficient local consistency of OMP is because the referable neighboring samples become less and less with the decrease of clipping level. In order to improve the local consistency for both Miura's method and the OMP, we focus on the possible impact of local consistency by characteristic of original waveforms in clipped intervals.

3.1 Number of Peaks in a Clipped Interval

As the Fig. 3 shows, Most shapes of original waveforms in clipped intervals are simple, and the clipped interval is shorter, the waveform is simpler. The interval length is the number of continuous sampling points in each clipped interval. The original waveform in clipped interval contains different number of peaks, if a original waveform only has two peaks, marked by the red rectangles in Fig. 3, the difference between two peaks is always not too much, and most of this kind of clipped intervals can be restored with one peak curve by a suitable simple method. In our paper, the intervals contain one or two peaks are considered as simple clipped interval.

The feasibility of using a simpler declipping method for simple clipped intervals depends on how many simple clipped intervals in a signal, if the number of simple clipped interval is large enough, the declipping performance can be improved by re-declipping the simple clipped intervals. So we calculate the percentage of simple clipped intervals in each clipped signal, and the result of average of 20 samples shows that, the percentage is increased from 86.01% to 100% with the increase of clipping level from 0.1 to 0.9 , and this result is just what we expected.

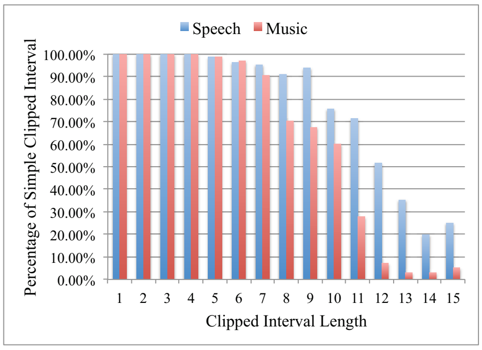

Fig. 4 Percentage of speech and music samples of simple clipped interval in each clipped interval length from 1 to 15, clipping level is 0.2.

But, not all the clipped intervals are simple clipped interval, even though the length is very short. So we calculate the percentage of simple clipped interval in each clipped interval length. The Fig. 4 shows the part of results of interval length from 1 to 15, the percentage decreases with the length becomes longer, after length 15, the simple clipped interval is significantly diminished. In order to keep the most simple clipped intervals are declipping by another method in short classification, we select the clipped interval length which contains the simple clipped interval are much more than the others. After the analysis, we decide to use length from 1 to 10 as the short classification for music, and the length 1 to 12 for speech.

3.2 The Declipping Method for Simple Clipped Interval

According to the property of simple clipped interval, the declipped waveforms should be continuity and smoothness, just contains only one peak, and the estimated signal should cross all the knots in clipped interval. We find that the spline interpolation can satisfy the need of the condition of the waveform in short length classification. In our proposed method, the analysis interval is the convex which contains clipped interval and the known samples on the both adjacent sides, the new values generate by spline algorithm, and replace the values in clipped interval by new values.

3.3 The Flow of the Proposed Method

We restore the declipped interval in short length classification by spline interpolation, and the long length classification we use the declipping result of Miura's method or the OMP. The Fig. 5 shows the whole flow of our proposed method.

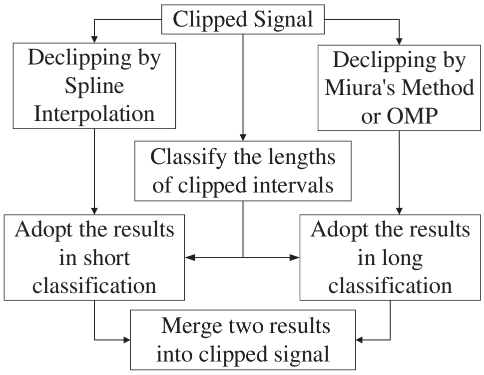

Fig. 5 The flow chart of the proposed method.

4 Experiments

4.1 Experiments Condition

The dataset of the experiments ten speech and ten music signals, all of them are sampled at 16 kHz, and the resolution of quantization is 16 bit. Each test signal is 5-second long and is part of the freely available material of the 2008 Signal Separation Evaluation Campaign[6]. The test data shows a large diversity of audio mixtures and isolated sources, including male and female speech from different speakers, singing voice, pitched and percussive musical instruments. All original signals are normalized so that the maximum amplitude is 1. Each sound is then artificially clipped with successive clipping levels, from 0.1 up to 0.9 with a 0.1-step. And we classify the clipped interval length from 1-12 as the short length classification for speech, the clipped interval length from 1-10 as the short length classification for music.

4.2 Evaluation Criteria

We conducted an objective evaluation SNR, The higher SNR the better sound quality. It is given by the following equation.

$$SNR(dB) = 10 \log \frac{\Sigma y(t)^2}{\Sigma(y(t) - \hat{y}(t))^2} \quad (1)$$

where, $y(t)$ is the original signal, and $\hat{y}(t)$ is the signal to be evaluated.

4.3 Result of Experiment and Consideration

The Fig. 6 and Fig. 7 shows the average SNR for 10 music signals and 10 speech over different clipping levels and methods respectively. We can see that, the result of spline interpolation and Miura's

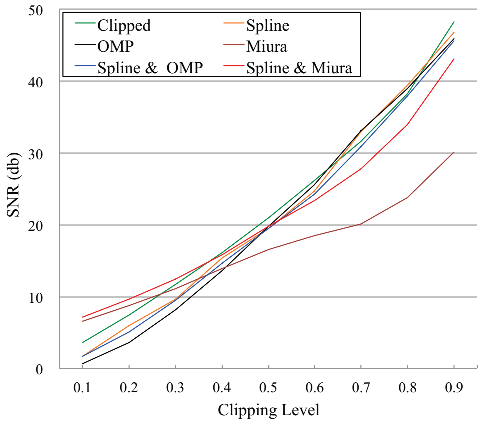

Fig. 6 Average SNR for 10 music signals over different clipping levels and methods.

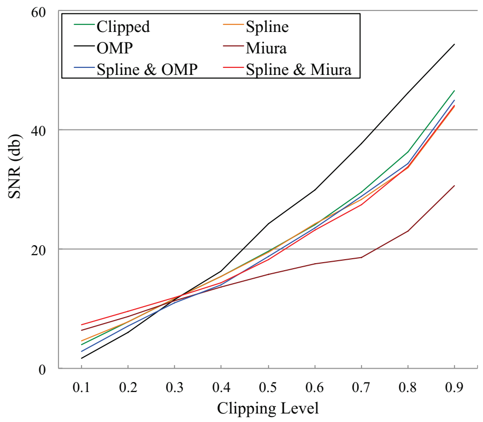

Fig. 7 Average SNR for 10 speech signals over different clipping levels and methods.

method is always better than just using Miura's method for both music and speech signal over all clipping levels. To the result of spline interpolation and the OMP, the improvement is appears in clipping level from 0.1 to 0.2 for speech, and 0.1 to 0,4 for music.

A sample declipped by spline interpolation and Miura's method is shown in Fig. 8, we use the result of spline interpolation in short interval (the red line) instead of result of Miura's method (the black dash), and the result of spline interpolation is much more compatible with the original waveform than the result of Miura's method.

5 Conclusion

We proposed a multi-stage restoration of clipping distortion based on length classification of clipped interval method in order to improve the quality of automatic restoration without prior knowledge. We analyze the property of the original waveforms in

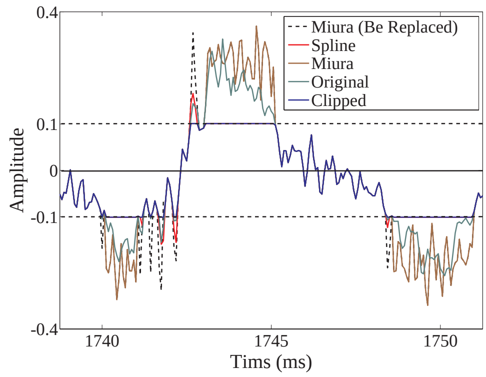

Fig. 8 A sample of declipping result by spline interpolation and Miura's method.

clipped interval, and find the most suitable boundary for classifying the short and long length classifications for both music and speech, we restore the short length classification by spline interpolation, restore long length classification by Miura's method or the OMP. To Miura's method, our proposed method is better in all clipping level for both music and speech samples, and we also improve the performance of the OMP in clipping level from 0.1 to 0.2 for speech, and 0.1 to 0.4 for music.

In future research, our approach is to investigate conventional global interpolation methods, and improve the declipping performance if possible.

References

[1] A. Janssen, R. Veldhuis, and L. Vries, Acoustics, Speech and Signal Processing, IEEE Transactions on, vol. 34, no. 2, 317330, 1986.

[2] J. S. Abel and J. O. Smith, Proceedings of the International Conference on Acoustics, Speech, and Signal Processing. IEEE, 1745-1748, 1991.

[3] S. J. Godsill, P. J. Wolfe, and W. N. W. Fong, Journal of New Music Research, vol. 30, no. 4, 2001.

[4] A. Adler, V. Emiya, M. Jafari, M. Elad, R. Gribonval and M. D. Plumbley, Proc. IEEE Int. Conf. Acoust., Speech, Signal Process., 329-332, 2011

[5] 三浦晋, 宮部滋樹, 山田武志, 牧野昭二, 中島弘史, 中臺一博, 日本音響学会講演論文集, 1-Q-10, 843-846, 2012.

[6] http://sisec2008.wiki.irisa.fr/