Shenghuan Zhang* and Ye Cheng

Masking and noise reduction processing of music signals in reverberant music

https://doi.org/10.1515/jisys-2022-0024
received October 14, 2021; accepted December 07, 2021

Abstract: Noise will be inevitably mixed with music signals in the recording process. To improve the quality of music signals, it is necessary to reduce noise as much as possible. This article briefly introduces noise, the masking effect, and the spectral subtraction method for reducing noise in reverberant music. The spectral subtraction method was improved by the human ear masking effect to enhance its noise reduction performance. Simulation experiments were carried out on the traditional and improved spectral subtraction methods. The results showed that the improved spectral subtraction method could reduce the noise in reverberant music more effectively; under an objective evaluation criterion, the signal-to-noise ratio, the de-reverberated music signal processed by the improved spectral subtraction method had a higher signal-to-noise ratio; under a subjective evaluation criterion, mean opinion score (MOS), the de-reverberated music signal processed by the improved spectral subtraction method also had a better evaluation.

Keywords: reverberant music, musical noise, masking effect, spectral subtraction method

1 Introduction

Music will be inevitably mixed with nonmusic signal noises in the recording process [1]. These noises will have different degrees of impact on the normal music signal after mixing, and the specific size of the impact depends on the application scenario of the music signal, but regardless of the scenario, the quality of music signals will be reduced [2]. To improve the quality of music signals, they need to be processed for noise reduction [3]. Kohlberg et al. [4] evaluated the impact of noise reduction algorithms on music enjoyment and found that reducing the noise of music helped improve the music enjoyment. Deng and Han [5] proposed an adaptive denoising algorithm based on the sparsity of the first-order differential overlap group of heart sound signals. The experiments found that the algorithm performed better in denoising heart sound signals than traditional “db10,” “db5,” and “bior5.5” wavelet methods. Chen et al. [6] explored the correlation between wavelet transform and heartbeat noise performance and the adaptation of heartbeat classification using bispectral estimation. By comparing the denoising effect of four wavelets, including haar, db6, sym8, and coif6, it was found that the db6 wavelet had the best denoising effect on heartbeat signals. The aforementioned studies all focus on music noise reduction. Some of the aforementioned studies started from the improvement of music quality by noise reduction algorithms and verified and the improvement of musical experience by noise reduction algorithms; some started from the specific algorithms of noise reduction and explored the advantages of different noise reduction algorithms. All the aforementioned studies provide effective references for music signal noise reduction. In this article, noises in the reverberant music were rapidly reduced based on the simplicity of spectral subtraction, and the human ear
masking effect was used to further improve the quality of music after spectral subtraction. The comparison between the traditional and improved spectral subtraction methods verified the effectiveness of the improved method. The novelty of this article lies in reducing noise in reverberant music fast by the spectral subtraction method and improving the quality of music after noise reduction by taking advantage of the insensitivity of the human ear to the phase change of music signals. This study provides an effective reference for improving the noise reduction of music signals.

2 Noise and masking effects

Objectively speaking, noise refers to all irregular signals [7], and in this article, it refers to irregular sound signals. Subjectively speaking, sounds that cause physiological and psychological discomfort and interfere with normal hearing are all noise. Even music with a regular pattern can be classified as noise if it interferes with the sound of people's normal communication or their normal routine [8].

Influenced by different environments, noise can be divided into additive noise and nonadditive noise [9]. Additive noise includes impulse noise, broadband noise, periodic noise, and interference from other voices; nonadditive noise mainly includes reverberation, circuit noise, etc. [10]. Impulse noise is characterized by the sudden appearance of a short period of high-amplitude band spectrum; broadband noise is characterized by its overlap with pure sound signals in time and frequency domains; speech noise is characterized by the common transmission of other sound signals and pure sound signals; circuit noise is inherent to circuit operation, but can be handled after being transforming to additive noise through homomorphic conversion [11].

The ear is the human organ that receives sound. In judging the noise reduction effect of music and other sound signals, although the signal-to-noise ratio can be used to objectively judge the noise reduction effect, it ultimately depends on the subjective perception of the human ear [12]. Usually, the music sound waveform subjectively perceived by the human ear is not the same as the actual music sound waveform. In addition, the human auditory system has a good noise suppression effect when processing the reverberant sound mixed with background noise. Therefore, when noise reduction is applied to reverberant music signals, the perceptual characteristics of the human ear can also be taken into consideration to make the noise reduction effect closer to the subjective perception of the human ear [13].

The human ear can perceive sound at frequencies between 20 and 20,000 Hz, and there is a masking effect when the human ear processes the sound signal. In short, a sound is more difficult to hear by the human ear due to the presence of another sound. When strength of the masking sound can make the masked sound just audible to the human ear, then that strength is the masking threshold [14].

3 Noise reduction algorithm

In this article, the spectral subtraction method is used to remove noise from the reverberant music, i.e., to remove the background noise mixed in the music signal [15].

3.1 The traditional spectrum subtraction method

The principle of removing noise from reverberant music by spectral subtraction is to set noise in the reverberant music that is not related to the music and then to subtract the noise from the reverberant music to get the de-reverberated music signal [16]. The process of reducing noise from the reverberant music with spectral subtraction is shown in Figure 1. The basic steps are as follows.
The reverberant music signal $Y(n)$ is processed by fast Fourier transform (FFT) [17] to obtain $Y(\omega)$. The signal frames without pure music in the reverberant music signal are counted to obtain the estimated noise signal, and Fourier transform is also performed to obtain $\hat{D}(\omega)$.

(3) The reverberant music signal and the estimated noise signal are subtracted to obtain the de-reverberated music signal:

$$\hat{X}(\omega) = \frac{1}{1+\gamma \cdot \text{diff}} X(\omega) - Y(\omega),$$

where $\hat{X}(\omega)$ is the estimated value of the music signal after de-reverberation.

(4) After adding the phase information to $\hat{X}(\omega)$, the inverse FFT is performed to obtain the de-reverberated music signal $X(n)$. The phase signal comes from the original reverberant music signal. The reason why the phase of reverberant music signals containing noise is directly used is that the human ear is not sensitive to changes in the phase of music signals.

As mentioned in the aforementioned steps, it is the basic process of spectral subtraction method for de-reverberation of reverberant music. The spectral subtraction method is simple and convenient when de-reverberating the reverberant music signals, but musical noise is generated in the de-reverberated music [18]. The musical noise is generated because the noise signal subtracted from the reverberant music signal in the spectral subtraction method is not the real noise in the reverberant music, but the estimated noise mean value obtained by statistics. Therefore, when the estimated noise signal is subtracted from the reverberant music signal, the real noise signal in the reverberant music signal whose real noise amplitude is lower than the estimated noise mean will be removed substantially, while the real noise signal whose real noise amplitude is higher than the mean of the estimated noise will have more residuals. These residual noise signals tend to be clearer than the original noise, affecting the de-reverberation effect of the spectral subtraction method.
3.2 The improved spectrum subtraction method

To enhance the de-reverberation effect of the spectral subtraction method on reverberant music and strengthen the de-noising performance of the method, this article improves the spectral subtraction method by using the human ear masking effect. The complete flow of the improved spectral subtraction method is shown in Figure 1. The improved spectral subtraction method for reverberant music de-reverberation and noise reduction can be divided into two stages. The first stage is consistent with the basic spectral subtraction method, i.e., subtracting the estimated noise from the reverberant music to obtain the de-reverberated music signal. The second stage is the postprocessing of the de-reverberated music signal by using the human ear masking effect [19] to eliminate the musical noise in the de-reverberated music.

The basic steps of the second stage are as follows.

(5) Both the original reverberant music $Y(n)$ and the de-reverberant music $X(n)$ are processed by FFT to obtain $Y(\omega)$ and $X(\omega)$.

(6) After normalizing $Y(\omega)$ and $X(\omega)$, the energy difference is obtained by subtracting them and is expressed as follows:

$$\text{diff} = \begin{cases} 0, & X_N(\omega) - Y_N(\omega) < 0, \\ X_N(\omega) - Y_N(\omega), & X_N(\omega) - Y_N(\omega) \geq 0 \end{cases}$$

(2)

where $Y_N(\omega)$ and $X_N(\omega)$ are the signals obtained by normalizing $Y(\omega)$, and $X(\omega)$ and diff is the energy difference between $X_N(\omega)$ and $Y_N(\omega)$.

(7) The masking factor is calculated based on the calculated energy difference diff and used to postprocess the de-reverberated music. The corresponding formula is:

$$\text{Mask} = \frac{1}{1 - \gamma \cdot \text{diff}},$$

(3)

$$X_m(\omega) = X_N(\omega) \cdot \text{mask},$$

where mask is the masking factor used to postprocess the de-reverberated music, $\gamma$ is the weighting factor whose value is adjusted according to the actual situation to achieve the best masking effect, and $X_m(\omega)$ is the postprocessed de-reverberated music signal.

4 Simulation experiments

4.1 Experimental environment

In this study, the de-reverberation performance of spectral subtraction was simulated by MATLAB software [20]. The experiments were conducted on a laboratory server with relevant configurations of the Windows 7 operating system, 32G RAM, and Core i7 processor.

4.2 Experimental methods

The audio data used were all from the music information collected in the soundproof room: the testers sang the songs, and the music information samples were collected in the soundproof room under quiet conditions. Each sample was 10 s long, and a total of 50 samples were collected, with a sampling rate of 5,000 Hz. Then, samples were added with Gaussian white noise to obtain reverberant music signal samples.

After obtaining the reverberant music samples, they were de-reverberated and noise reduced using the traditional and improved spectral reduction methods, respectively. In this article, the signal-to-noise ratio
method was used to objectively evaluate the performance of the two de-reverberation noise reduction algorithms. The signal-to-noise ratio is expressed as follows:

$$\text{SNR} = 10 \log_{10} \frac{\sum_n s(n)}{\sum_n (s(n) - \hat{s}(n))^2},$$

where $s(n)$ is the pure music signal and $\hat{s}(n)$ is the music signal after de-reverberation. As the ultimate goal of various noise reduction algorithms in the actual use process is to eliminate or reduce the noise in the music signal so as to improve the listening quality of the music signal, no matter how excellent the objective signal-to-noise ratio is after denoising, it ultimately depends on the subjective perception of the human ear.

Therefore, in addition to objectively evaluating the de-reverberation noise reduction performance with the signal-to-noise ratio, this article also used the mean opinion score (MOS) method. In this study, five volunteers with good hearing were selected to evaluate the noise reduction performance of the two algorithms, and the samples before and after de-reverberation were played to the volunteers one by one. The volunteers scored the de-reverberated samples for one to five points.

The MOS method divided the quality of the processed music signal into five levels: five points meant excellent, four points meant good, three points meant moderate, two points meant poor, and one point meant bad. A score of five points meant that the human ear was not aware of the noise in the music; a score of four points meant that the human ear was slightly aware of the noise; a score of three points means that the human ear was aware of the noise and slightly disgusted with it; a score of two points means that the human ear was obviously aware of the noise, disgusted with it, but is tolerable. One point meant that the human ear could not tolerate the noise.

### 4.3 Experimental results

Due to the limitation of space, this article only shows the time domain waveform of the relevant music signal for 1 of the 50 samples, as shown in Figure 2. It contained the pure music signal, the reverberant music after adding Gaussian white noise, and de-reverberated music signals after being processed by the two spectral subtraction methods. Figure 2 shows that after adding Gaussian white noise, the time domain waveform of the sample was significantly interfered with compared with the pure music signal; after processing by the traditional spectral subtraction, the white noise part of the time domain waveform of the de-reverberated music signal was significantly converged, but the noise still remained compared with the pure music signal; in the music signal processed by the improved spectral subtraction method, the white noise part has been further removed so that it was much closer to the pure music signal.

The signal-to-noise ratio was used in this study to objectively evaluate the de-reverberation effects of the two spectral subtraction methods. Average signal-to-noise ratios of the 50 samples before the noise reduction and after being processed by the two spectral subtraction methods are shown in Figure 3. Figure 3 shows that the amount of noise that was removed by the improved spectral reduction method was greater than that removed by the traditional spectral reduction method, thus making the noise-reduced music signal closer to the pure music signal, the difference between them smaller, and the signal-to-noise ratio larger.

In this study, in addition to using the signal-to-noise ratio to objectively evaluate the noise reduction performance of the two spectral reduction methods, the MOS method was also used to subjectively evaluate the performance in terms of the actual human ear perception of the noised-reduced music. The average MOS of the 50 music samples after noise reduction using the traditional and improved spectral subtraction methods is shown in Figure 4. It was seen from the comparison in Figure 4 that the human ear was perceptible to the noise in the de-reverberated music processed by the traditional spectral subtraction method and slightly disgusted with it, but could tolerate it; the human ear was occasionally perceptible to the noise in the de-reverberated music processed by the improved spectral subtraction method.
In the process of music signal acquisition, the mixed noise signal can affect the quality of music signals to varying degrees. Depending on the use of music signals, the processing also varies; however, the general rule is to reduce noise signals in music signals as much as possible to highlight music signals. This article used the spectral subtraction method for noise reduction of reverberant music. The noise reduction principle of the spectral subtraction method is to subtract noise signals from signals. In practical application, the first few frames without music signals are counted to estimate noise signals in music signals for the subsequent spectral subtraction processing.

The ideal state for noise reduction of reverberant music using traditional spectral subtraction is that the noise signal has been known, but in practice, the noise signal is not known and should be estimated from
the first few frames that do not contain the music signal, which makes the noise reduction effect unsatisfactory. Hence, this article used the human ear masking effect to improve the quality of music signals after noise reduction. Then, a comparison experiment between the traditional spectral subtraction method and the improved spectral subtraction method was also conducted, and the results have been shown earlier. The time domain waveform of the signals after processed by the improved spectral subtraction method was closer to the original signal than the signals processed by the traditional spectral subtraction method, and the comparison of the signal-to-noise ratio before and after the processing also showed that the improved spectral subtraction method had better noise reduction performance. In terms of subjective scoring, the music signals processed by the improved spectral subtraction method had a higher score.

The reason for the aforementioned results is as follows. The traditional spectral subtraction method subtracts noise signals to achieve noise reduction, but in the actual noise reduction process, the noise signal mixed in the music signal is unknown, so the noise signal needs to be estimated. The estimated noise signal is based on statistical samples, which may be statistically consistent with the actual noise, but cannot be guaranteed to be identical to the actual noise. The traditional spectral subtraction method uses the estimated noise signal for noise reduction, which may cause damages to music signals or incomplete removal of noise signals. The improved spectral subtraction method uses the human ear masking effect to further reduce noise from the music signal after noise reduction by the traditional spectral subtraction method. After noise reduction based on the human ear masking effect, the noise is masked by the music signal in the perception range of human ears to highlight the original music signal objectively and subjectively and to achieve noise reduction.

6 Conclusion

This article briefly introduces noise, masking effects, and the spectral subtraction method for reducing noise in reverberant music. The spectral subtraction method was improved using the human ear masking effect to enhance its noise reduction effect. Moreover, simulation experiments were carried out on the traditional spectral subtraction method and the improved spectral subtraction method. The research results showed that the time domain waveform of the pure music signal added with noise was significantly interfered with; the noise was significantly reduced after using the traditional spectral subtraction method and further reduced after using the improved spectral subtraction method, and the latter was closest to the pure music signal; the signal-to-noise ratio of the music de-reverberated by the improved spectral subtraction method was higher than that of the music de-reverberated by the traditional spectral subtraction method, which meant that the improved spectral subtraction method had better de-reverberation and noise reduction effects objectively; the MOS of the music de-reverberated by the improved spectral subtraction method was higher.
Conflict of interest: The authors state no conflict of interest.

References


