Quality Correction for Songs of the War of Resistance Based on Computer Audio Technology

Ying Zhao*

Mudanjiang Normal University, Heilongjiang, China, 157000

*E-mail: 402407091@qq.com

Abstract. In audio coding, the use of audio signals with the minimum quality to perceive distortion and express them at the lowest possible coding rate are studied to facilitate the transmission or storage of the audio signals. In this paper, the VBR (VBR) audio coding algorithm for sound quality control is designed based on computer audio technology. Audio signals are quantized and encoded according to the requirements for different levels of sound quality. The songs of the War of Resistance are used for testing. The test results suggest that the proposed algorithm can minimize the coding rate at the set sound quality.

Keywords: Audio Coding, Sound Quality Model, VBR

1. Introduction

With the development of network technology and audio compression algorithms, various network digital music services, such as Internet webcast and online record stores, have been welcomed by more and more people. So far, the most widely used audio compression method is MP3 [1-2]. However, MP3 is not open. It is protected by copyright. For MP3 encoders and public release of music works in MP3 format; a certain royalty is charged for each product copy. The problem of audio coding (Audio Coding) research is to express audio signals at the lowest possible coding rate with the lowest quality distortion, so as to facilitate the transmission or storage of audio signals. Hence, the low rates are used to express audio signals digitally to design an audio coding or audio compression algorithm, which can minimize the signal distortion in the sense of hearing, rather than just minimize the mean square error of the traditional input and output waveforms [3]. Audio compression (Audio Compression) and audio encoding generally have the same meaning. Although audio coding and speech coding (Speech Coding) belong to the same source of compression coding, the coding algorithms of the two are different. The source of the speech signal is only one, that is, the human speech organ, and the basis of speech coding is the speech generation model, i.e., source model [4]. In contrast, the source of the audio signal includes all the sounds that the human ear can feel. There are many sound sources, and the signal is complex, and it cannot be processed with a unified sound source model. Nevertheless, various audio signals are finally accepted by the human ear. Hence, the features of the human auditory perception system can be used to study the method of audio coding, so the basis of audio coding is the auditory model or the sink model [5-6].

This paper relies on computer technology to design an audio coding algorithm that can reduce the coding rate as much as possible in the sense of auditory perception according to the desired audio
quality, thereby maximizing the coding compression ratio. The basis of the algorithm is wavelet transform, sound quality model, and adaptive bit allocation for sound quality control.

2. Perceptual Coding Algorithm for Audio Signals

Figure 1 describes the complete working process diagram of the audio coding system. Given the diversity of audio signal sources, the main factor to consider when designing audio encoding/decoding algorithms is the perceptual features of the human ear, i.e., auditory features. The primary means of low-rate high-fidelity audio coding is distortion masking or noise masking. Based on the method, the inevitable introduction of distortion or noise in the sound coding process can be shaped appropriately and redistributed to be masked by the original input signal. Masking can be partial or full. In this way, compared with the system without masking technology, the audio coding quality can be improved.

![Diagram of the audio coding and transmission system](image-url)

**Figure 1.** Schematic diagram of the audio coding and transmission system

The main problem with audio encoding is that the compression ratio is not high.

In general, relatively high rate of audio coding algorithms generally aim to maintain the original signal waveform and use the time-domain waveform of the decoded signal to approximate the time-domain waveform of the original signal. The audio codec design is based on the principle of the minimum mean square error of the signal waveform before and after encoding. In the low-rate audio coding algorithm, the number of allocated bits is limited, the time-domain waveform distortion is large, and the waveform mean square error distortion criterion can no longer reflect the perceptual features of the auditory system. The selection of the distortion criterion should be conducive to the high quality of hearing for thousands of people Reception, instead of simply requiring to approximate the original waveform, is not blindly pursuing a better signal-to-noise ratio. The final and highest criterion of audio signal quality is the human auditory system, and the audio signal encoding algorithm based on the least distortion of auditory perception is called perceptual encoding algorithm. The reason why the perceptual coding algorithm can complete the slot coding with high quality and high efficiency is that it can do:

1. The use of short-term stationary features of audio signals, such as the correlation between samples, the periodicity of the audio signal, the formant of the spectrum, etc., to separate the redundant components in the source signal, that is, the redundancy (Redundancy).

2. The use of auditory features to eliminate components inconsistent with the subjective perception of the auditory system, i.e., irrelevance. Irrelevance is shown as the useless high resolution of amplitude or frequency, i.e., the signals can be masked by the hearing part.

The VBR (VBR) audio coding algorithm for sound quality control mainly refers to the input audio signal is a two-channel stereo high-fidelity audio signal, its amplitude resolution is 16bit, the sampling frequency is 44.1kHz, and its PCM coding rate is 1411.2kbit/s/Decoding algorithm is carried out in frames, the frame length is taken as 512 points, which is equivalent to 11.61ms. After the original audio signal enters the audio encoder, each channel audio signal is divided into two channels, one channel signal is first sent to the positive The cross-wavelet transform filter bank performs wavelet transformation and decomposition to decompose the audio signal with a bandwidth of 22.05kHz into 22 wavelet sub-bands with different bandwidths continuously distributed in the frequency domain according to the in-ear hearing features. These 22 bandwidths are different The wavelet sub-band has a higher frequency resolution in the low frequency band and a higher time resolution in the high-frequency band. This aspect reflects the features of the audio signal and is also the requirement of the auditory features of the human ear. The dynamic range of the audio signal components decomposed from thousand to each wavelet subband is very large. To effectively quantize and encode, the scale factor of each wavelet subband must be measured first, and it is used to reduce the sample.
points of each subband One-time processing, so that the dynamically allocated bits can quantize and encode each sub-band auspicious point on a uniform scale. While time-frequency analysis, another audio signal is used to estimate the sound quality model, and the output of the sound quality model is to reflect hearing Signal masking ratio (SMR) of the masking feature. The final operation of the audio encoder is to assemble the encoded value and bit allocation information into a data packet, and add the frame header information to form the encoded data frame and send it to the channel or storage medium.

Different from other audio encoding algorithms, two input control parameters are set in the audio encoder, i.e., rate control and sound quality control. The rate control parameter can be feedback information from the channel, indicating the current channel capacity; it can also be audio encoding. The information of the device itself is used to control the maximum output rate. The sound quality control parameters can be used to control the desired subjective quality of the audio encoding. In this way, the signal masking ratio, rate control, and sound quality control of the sound quality model are three parameters together to determine the rate and quality of audio encoding.

The audio decoder algorithm does not need to calculate the sound quality model, so it is much simpler than the audio encoder algorithm. The audio decoder receives the audio coded bitstream from the channel or storage medium, decodes the scale factor and wavelet subband auspices based on the bit allocation information after frame synchronization, and then performs scale transformation to restore its dynamic range before encoding. Finally, after the wavelet inverse transform of the wavelet synthesis filter bank, the two-channel stereo audio signal with a reconstruction rate of 1411.2kbit/s is output.

The masking effect is a frequency-domain auditory phenomenon, which is manifested as a sound signal with a large energy that can make another sound signal with a lower energy that cannot be heard by the human ear as long as the frequency difference between the two signals is sufficiently small. The signal with higher energy is known as the masked signal, and the signal with lower energy is known as the masked signal. The masked signal may include lower energy audio signals, quantization noise, overlap distortion, or distortion due to transmission errors, etc.. The masking effect plays a more significant role in the critical frequency band where the masking signal is located, while it rapidly decreases in its adjacent critical frequency band. The critical frequency band is a frequency-domain psychoacoustic or sound quality measurement, which reflects the frequency selectivity of the human ear. The unit of the critical band is bark, which represents a non-linear frequency scale, which is related to a physical distance in the cochlear basement membrane and expresses the minimum bandwidth that can distinguish the masked audio signal.

In the critical band, the audio signal below the masking threshold (shaded area) will be masked. The difference between the sound pressure level of the masking signal and the masking threshold is called the signal masking ratio (SMR), and its maximum value is in the critical band On the left boundary (point A in Figure 3), the minimum value is near the frequency of the masked signal. Assuming that the audio signal is quantized in m bits, as long as the signal-to-noise ratio (SNR) in the critical band is greater than the signal masking ratio (SMR), the quantization noise will be unheard.

If SNR (m) is used to represent the SNR during mbit quantization, then the subjectively perceptible distortion in a given frequency band can be measured by the noise masking ratio (NMR)

\[ NMR(m) = SNR(m) - SMR(dB) \]  

NMR (m) describes the ratio between SNR and SMR determined by m-bit quantization, and also shows the difference between the level of distortion that will be perceived by hearing and the amount of quantization noise in a given frequency band. m) The larger the NMR(m), the less subjective perception of quantization coding noise. When NMR (m) is sufficiently large, the quantization noise will not be perceived by the auditory system.

The sound quality model is based on the short-term frequency spectrum of the input audio signal and uses the masking features to calculate the signal masking ratio. It is the foundation of low-rate high-fidelity audio coding. In the sound quality model designed in this paper (as shown in Figure 1),
high-resolution spectrum analysis is first conducted. Subsequently, the modulated (similar to a thousand periodic sine wave) and unmodulated (similar to noise) components are extracted. The masking threshold of each single masking signal is determined according to the frequency band, amplitude and modulated or unmodulated type of each signal component as follows:

\[
M_{\text{tonal}}(z_j, z_i) = X_{\text{tonal}}(z_j) + V_{\text{tonal}}(z_j) + V_j(z_j, z_i) (dB)
\]

\[
M_{\text{non-tonal}}(z_j, z_i) = X_{\text{non-tonal}}(z_j) + V_{\text{non-tonal}}(z_j) + V_j(z_j, z_i) (dB)
\]

Where \( M_{\text{tonal}}(z_j, z_i) \) is the separately generated masking threshold at \( z_i \) (unit is bark) with a modulated component frequency of \( X_{\text{tonal}}(z_j) \) (unit is dB) and frequency \( z_j \) (unit is bark); \( M_{\text{non-tonal}}(z_j, z_i) \) is \( X_{\text{non-tonal}}(z_j) \) A separately generated masking threshold at the frequency of the unmodulated component of frequency \( z_j \) at \( z_i \). In (2) and (3), \( V_{\text{tonal}}(z_j) \) and \( V_{\text{non-tonal}}(z_j) \) represent the masking index of the modulated component and the unmodulated component with frequency \( z_j \), that is, the degree of the masking effect changing with frequency \( z_j \).

3. Sound quality control for songs of the war of resistance
The primary purpose of audio coding is to resolve the contradiction between sound quality and rate. Low-rate audio coding inevitably produces coding errors or quantization noise. To achieve high-fidelity quality, we have to keep the quantization error within the range that subjective hearing cannot perceive. The coding rate sound quality control algorithm designed in this paper is to control the coding rate with sound quality. In other works, under the given sound quality conditions (even if the global noise-mask ratio in the frame is not less than a certain set value NMRset), the perceptual adaptive dynamic bit allocation algorithm is used to reduce the audio coding rate as much as possible, and make the quantization coding distortion audible. The increase is minimal. Under this condition, the coding rate generally changes with time.

The VBR bit allocation of sound quality control is a process of minimizing the number of bit allocations. The result of bit allocation \( B(t) \), \( i = 0, 1, \ldots, M - 1 \) defines the word length or resolution of the transmitted samples or parameters of the ith frequency band at time t. The operation process shall meet the constraint condition \( \sum_{i=0}^{M-1} b_i(t) = B_{\text{total}}(t) \). The process of variable-rate bit allocation is an iterative process, and the resolution of the auspicious point in the wavelet sub-band with the minimum noise-mask ratio (NMR) will increase by 1 bit for each iteration completed. The specific steps are as follows:

(1) Initialization. If \( B_{\text{header}} \) is used to indicate the number of bits required to encode the frame header information, \( B_{\text{error}} \) is used to indicate the number of error control words, and \( B_{\text{bit-allocate}} \) is used to indicate the number of bits required for bit allocation information. The encoding quality threshold or quality level is set.

Let \( B_{\text{sample}} \) and \( B_{\text{scale}} \) represent the number of bits allocated to the wavelet subband auspicious point and scale factor, respectively. During initialization, the bits assigned to each wavelet subband sample vector and each scale factor are all zero. Calculate the noise masking ratio NMR of each wavelet sub-band.

(2) The wavelet subband with the smallest noise-mask ratio NMR is determined;

(3) If the noise mask of the wavelet subband is larger than NMR by a thousand \( NMR_{\text{set}} \), the total number of bits that have been allocated is calculated, and the iteration ends. Otherwise, the iteration continues.

(4) For the wavelet subband with the smallest noise-mask ratio NMR, the sting resolution of
each auspicious point is increased by 1 bit.

(5) Cumulative subband auspicious points and scale factor bit allocation numbers.

If a wavelet subband auspicious point is assigned to a non-zero bit number for the first time, it is because the scale factor of the vector is quantized and encoded with 6 bits.

(6) The noise-mask ratio NMR of each wavelet sub-band is re-calculated. Subsequently, turn to step (2).

To test the relationship between coding rate, encoding quality and sound quality control settings, and audio encoding quality, this paper selects a form for the songs of the War of Resistance signal-to-noise ratio that is in good agreement with the subjective quality test results, the so-called segment signal Noise ratio (SegSNR) to objectively evaluate the audio coding quality. The audio signal used is a PCM signal with a sampling frequency of 44.1kHz, 16bit quantization accuracy, two-channel stereo, and a rate of 1411.2kbit/s. The relationship between the coding rate and the segment signal-to-noise ratio and the audio coding quality setting value NMR set is provided in Figure 2, which suggests that the average coding rate and the segment signal-to-noise ratio both increase with the increase of the quality setting value NMR set.

Figure 2. Relationship between the average coding rate and the segment signal-to-noise ratio and the set audio coding quality

From the Segment SNR calculation formula of the segment signal-to-noise ratio, it does not reflect the factors related to the auditory features. Therefore, it is difficult to determine to what extent the segment signal-to-noise ratio and the subjective sound quality test results can be consistent. Therefore, in addition to the SNR and other objective observation calculations, the audio coding quality ultimately needs to be measured by subjective perception. Subjective audition results suggest that for a two-channel stereo audio signal with the raw rate of 1411.2 kbit/s, the audio coding algorithm designed in this paper can achieve transparent high-fidelity quality when the average coding rate is above 180 kbit/s; where the average coding rate is about 120kbit/s, high-fidelity quality can also be obtained for most audio signals. Where the average coding rate is below 90kbit/s, as the average quantization coding accuracy is further reduced, the subjectively perceptible audio distortion will increase rapidly.

4. Conclusion

With the increasing popularity of audio compression, the requirement to ensure that the loss of sound quality should not be excessively large while ensuring the compression at the same time has become more and more clear. In this paper, the audio coding algorithm designed based on computer technology can minimize the coding rate in the sense of auditory perception based on the desired audio quality, thereby maximizing the coding compression ratio. The basis of the algorithm is wavelet transform, sound quality model, and adaptive bit allocation for sound quality control.
Acknowledgments
Project Name: A Study on the Era Significance of the Songs of the Mudanjiang Basin during the Fourteen Years of the Anti-Japanese War. Project No.: 1354ZD019

References
[1] Yoon C , Kang J , Han S , et al. Enhancement of photoacoustic image quality by sound speed correction: ex vivo evaluation[J]. Optics Express, 2018, 20(3):3082-3090.
[2] Barr, Richard G . Speed of Sound Correction and Quality Control[J]. Journal of Ultrasound in Medicine, 2017, 31(9):1471-1471.
[3] Mead Killion, John French, Steve Viranyi. A novel method of improving sound quality and reducing acoustic feedback in hearing aids[J]. journal of the acoustical society of america, 2016, 111(5):2381-2381.
[4] Lorente J , Ferrer M , De Diego M , et al. The frequency partitioned block modified filtered-x NLMS with orthogonal correction factors for multichannel Active Noise Control[J]. Digital Signal Processing, 2015, 43(3):47-58.
[5] Enamito A , Hiruma T . A Study on Acoustical Characteristics of Ear Canal by Earphones (A Derivation of Resonance Correction Filter by Using Transfer Matrix Method)[J]. Transactions of the Japan Society of Mechanical Engineers, 2012, 78(789):979-983.
[6] Lee S M , Back J , An K , et al. Design and Generation of a Target Sound to Achieve the Desired Sound Quality Inside a Car Cabin[J]. International Journal of Automotive Technology, 2020, 21(2):385-395.