Mathematical Modeling and Algorithm Application
ISSN: 3006-0842 | Vol. 5, No. 2, 2025

Research on Audio Encoding Cost-Performance Evaluation Model Based on Multi-dimensional Parameter Mapping

Zhenyu Wang[#, *], Naijie Gu[#], Jian Gao[#]

School of Information Engineering and Automation, Kunming University of Science and Technology, Kunming 650500, China

[*]**Corresponding author**: Zhenyu Wang (Email: 15896978961@163.com)

[#]**These authors contributed equally.**

Abstract: With the widespread application of digital audio technology, the balance between audio quality and file size has become an important research topic. This study establishes a systematic evaluation framework to quantitatively analyze how sampling rate, bit depth, compression algorithms, and audio duration collectively influence audio quality and storage efficiency through a Cost-Performance Index (CPI). The research first constructs a theoretical model, quantifying audio quality using Signal-to-Noise Ratio (SNR) and Mean Squared Error (MSE), and calculates CPI in combination with actual file size. Through analysis of "music" and "speech" content categories, the study finds that speech encoding parameter combinations offer a broader selection space in terms of "quality/volume" trade-offs, while music files generally have lower CPI values with a narrower distribution range. Experimental results show that for music files, 44.1kHz with MP3 64kbps offers the highest compression efficiency; while for speech applications, 16kHz downsampling combined with MP3 64kbps presents the best value proposition, delivering acceptable listening quality while keeping file size at an extremely low level. Correlation analysis confirms a strong positive correlation between CPI and quality score (r≈+0.85) and a significant negative correlation with file size (r≈-0.65), validating the rationality of the "high quality + small size" concept. This research provides clear parameter guidance for audio encoding and storage strategies, with practical application value for optimizing audio compression and storage solutions.

Keywords: Audio quality, File size, Cost-Performance Index, Compression algorithm

1. Introduction

With the rapid development of digital media technology and the widespread adoption of mobile internet, the acquisition, storage, and transmission of audio content have become fundamental needs in both daily life and professional fields[1, 2]. However, high-quality audio files typically occupy substantial storage space, creating a tension under limited bandwidth and storage resources[3]. Traditional audio encoding research has often focused on single-dimension optimization, such as sampling rate adjustments or compression algorithm improvements, lacking a systematic multi-parameter collaborative evaluation framework[4]. Particularly in different scenarios like voice communication and music appreciation, ideal parameter combinations differ significantly, yet objective quantitative decision-making criteria to guide selection are missing. As cloud storage costs rise and mobile device storage remains limited, how to minimize file size while ensuring acceptable audio quality has become a critical issue in digital audio processing.

This research aims to address this gap by establishing a cost-performance evaluation model for audio encoding based on multi-dimensional parameter mapping, systematically quantifying how core parameters such as sampling rate, bit depth, and compression algorithms affect audio quality and file size. The study innovatively designs a Cost-Performance Index (CPI) that unifies subjectively difficult-to-measure "audio quality" and objective "storage cost" into a comparable measurement framework. Additionally, the research conducts differentiated analysis on speech and music audio content, revealing optimal parameter combinations for different types of audio content and providing precise guidance for audio encoding strategies in specific scenarios.

2. CPI-Based Optimization Framework for Audio Encoding Parameters

This research data is sourced from http://mathorcup.org/. Combined with domain-specific prior knowledge, this study examines how sampling rate, bit depth, compression algorithms, and audio duration collectively influence audio quality and file size. The research establishes a cost-performance index to evaluate the trade-off between these factors[5].

The evaluation framework first defines essential parameters: sampling rate (fi measured in Hz) determines samples captured per second, while bit depth (bi measured in bits) establishes amplitude quantization precision. Compression ratio (CRi) represents the relationship between original and compressed file sizes—lossless encoding maintains CRi = 1, while lossy formats such as MP3 and AAC typically exhibit CRi > 1.

The model incorporates additional variables including audio duration (T measured in seconds) and actual file size (Si measured in MB). To objectively quantify audio quality degradation, established metrics such as Mean Squared Error (MSE) and Signal-to-Noise Ratio (SNR) are employed. These parameters collectively establish the foundation for a comprehensive evaluation methodology that objectively assesses how encoding choices impact both perceptual quality and storage efficiency.

In the ideal scenario without header overhead, the size of uncompressed PCM audio should satisfy:

$$S_i^{theo} = \frac{f_i \times b_i \times T}{8 \times 1024^2}\ (MB) \quad (1)$$

Actual measured file size is recorded as Si, which typically differs from the theoretical value only by the overhead of header information.

For the audio preprocessing phase, the Interquartile Range (IQR) method was employed to eliminate outlier samples that fell outside the specified range based on the CPI metric. This statistical approach ensured that extreme values which could potentially skew the analysis results were removed from the dataset prior to further processing. Any CPI value below or above the threshold will be classified as an outlier and subsequently removed from the dataset.

This filtration process targeted two specific categories of anomalous data. Samples with extremely low CPI values typically represented highly compressed lossy encoded files with significant audio quality degradation, such as "Speech_8000Hz_MP3_32kbps" and "Music_32000Hz_AAC_64kbps". These samples achieved artificially favorable denominators due to their exceptionally small file sizes, yet their excessive distortion resulted in disproportionately low CPI values that deviated substantially from the overall distribution.

Conversely, samples exhibiting exceptionally high CPI values often corresponded to duplicated lossless encodings or samples with abnormal header information, exemplified by "Music_48000Hz_24bit_duplicate" and "Speech_44100Hz_PCM_test". These samples produced unusually large numerators due to their theoretical absence of distortion, resulting in CPI values that significantly exceeded the normal range.

For audio quality measurement, the evaluation metrics were calculated using the original audio samples x_{orig}[n] and the reconstructed samples x_i[n] from the i-th test file. The mean squared error was utilized as one of the primary quantitative measures to assess the degree of distortion between the original and processed audio signals.

$$MSE_i = \frac{1}{N}\sum_{n=1}^{N}(x_{orig}[n] - x_i[n])^2 \quad (2)$$

The signal-to-noise ratio and its conversion to a quality score on a linear scale are shown in the following equations:

$$SNR_i = 10\,log_{10}\frac{\sum x_{orig}\,[n]^2}{\sum(x_{orig}[n] - x_i[n])^2} \quad (3)$$

$$Q_i = 10^{SNR_i/10}\ \ or\ \ Q_i = \frac{1}{MSE_i} \quad (4)$$

Both definitions improve as the distortion decreases, which can intuitively reflect the quality of sound. The CPI design was developed to simultaneously account for "audio quality obtainable per unit of storage." Two common measurement approaches were proposed for this purpose. First is the ratio-based type, as shown in the equation below:

$$CPI_i = \frac{Q_i}{S_i} \quad (5)$$

The "quality gain" per MB, with higher index values indicating better performance. Secondly, there is the weighted normalization type, which first dimensionlessly processes both quality and storage:

$$\widetilde{Q_i} = \frac{Q_i - Q_{min}}{Q_{max} - Q_{min}}, \qquad \widetilde{S_i} = \frac{S_i - S_{min}}{S_{max} - S_{min}} \quad (6)$$

Then construct:

$$CPI_i{}' = \alpha\,\widetilde{Q_i} + (1-\alpha)\,(1-\widetilde{S_i}),\quad \alpha \in [0,1] \quad (7)$$

Where α controls the weight of "audio quality" in the cost-performance ratio, typically set to α=0.5 to achieve a symmetrical compromise.

For the classification, sorting, and recommendation process, all test files were categorized into two groups based on "content type": "music" and "speech." The CPI or CPI′ was calculated separately for each group. Within each group, all parameter combinations were ranked according to their cost-performance ratio in descending order, with the optimal parameter combination becoming the recommended configuration.

The underlying principle of this model integrates physical computation (theoretical file size), objective quantification (conversion of MSE/SNR to quality score Q), and multi-attribute decision making (using either ratio-based or weighted CPI). This comprehensive approach enables systematic evaluation of how parameters such as sampling rate, bit depth, and compression algorithm collectively impact audio quality and storage efficiency.

Taking "ratio-type CPI" as an example, this paper will elaborate on the modeling steps from data acquisition to recommended output.

For data preprocessing and parameter extraction, this study will scan all audio files in dataset, reading filenames to parse parameters: sampling rate, bit depth, compression algorithm identifier (such as PCM, MP3, AAC) and estimated compression ratio. The duration T of each file will be extracted (either read from audio headers or manually set uniformly). The actual file size will be measured.

For audio quality quantification, the corresponding "original music" and "original speech" will serve as quality benchmarks. Each test file will be compared and the sum calculated using the aforementioned formula. The quality score will be calculated based on the selected model.

Then, according to following contents:

$$S_i^{theo} = \frac{f_i b_i T}{8 \times 1024^2} \times \frac{1}{CR_i} \quad (8)$$

Verify the deviation between the actual measurement Si and the theoretical value, and decide whether to remove the outliers or use the correction coefficient. For price-performance ratio calculation, this study will directly calculate C/P for each file in a similar manner.

$$CPI_i = \frac{Q_i}{S_i} \quad (9)$$

Adjust the weighting coefficient for different application scenarios and construct a weighted normalized CPI. For classification and sorting, this study will categorize all "speech" files and "music" files separately according to their filename labels. Within each category, files will be ranked from high to low based on CPIi to form a ranking sequence. Similar processing will be applied.

$$CPI_{(1)} \geq CPI_{(2)} \geq \ldots \quad (10)$$

The parameter combination ranked first will be considered the best recommendation for "that category."

For robustness and sensitivity analysis, this study will compare the impact of different definitions of Q_i (SNR vs. 1/MSE) on the recommendation results. The paper will examine the patterns of α variation in the weighted CPI and observe whether the "optimal combination" remains stable under different α values. The contribution of each parameter (fi, bi, CRi) to the final CPI will be analyzed using local partial derivatives or regression analysis.

For recommendation output, this study will present the optimal parameter combination {f*, b*, CR*} and the corresponding CPI_max for the "speech" category. Similarly, for the "music" category, the optimal parameter combination (f_i, b_i, CR_i) and the corresponding CPI_max will be provided.

Through this process, the paper will systematically map multidimensional parameters such as sampling rate, bit depth, and compression algorithm to a unified price-performance index, and identify the optimal configurations for both speech and music content separately. This will provide clear parameter guidance for subsequent audio encoding and storage strategies.

3. Quantitative Analysis of Audio Compression Efficiency Using CPI

The CPI distribution characteristics are shown in Figure 1. From the box plot, it can be seen that the CPI of the "speech" category has a median of approximately 0.8×10^{-6}, while the CPI of the "music" category is concentrated around 0.1×10^{-7}. The whisker ranges on both sides are notably different: the speech distribution is wider, indicating that speech encoding parameter combinations offer a broader selection space in terms of "quality/volume" trade-offs; whereas music files generally have lower CPI values with a narrower distribution, suggesting that music requires higher fidelity and compression efficiency is more difficult to significantly improve.

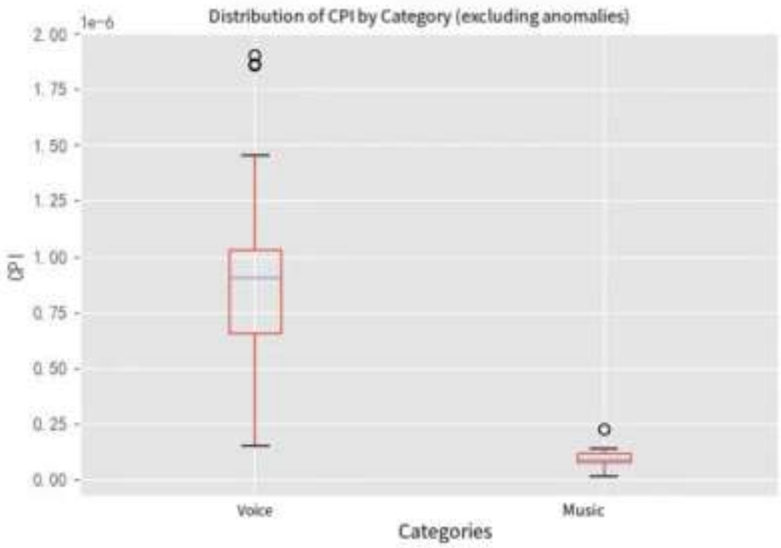

Figure 1: The box plot of CPI

The top 5 rankings by the index are shown in Figure 2, with side-by-side horizontal bar charts displaying the Top 5 parameter combinations for both categories. For the Music Top 5, the best performer is Music_44100Hz_MP3_64kbps, with a CPI of approximately 2.2×10^{-7}, leading the second place by about 30%. This is followed by 16 kHz/128 kbps, 44.1 kHz/128 kbps (constant bit rate), and other combinations.

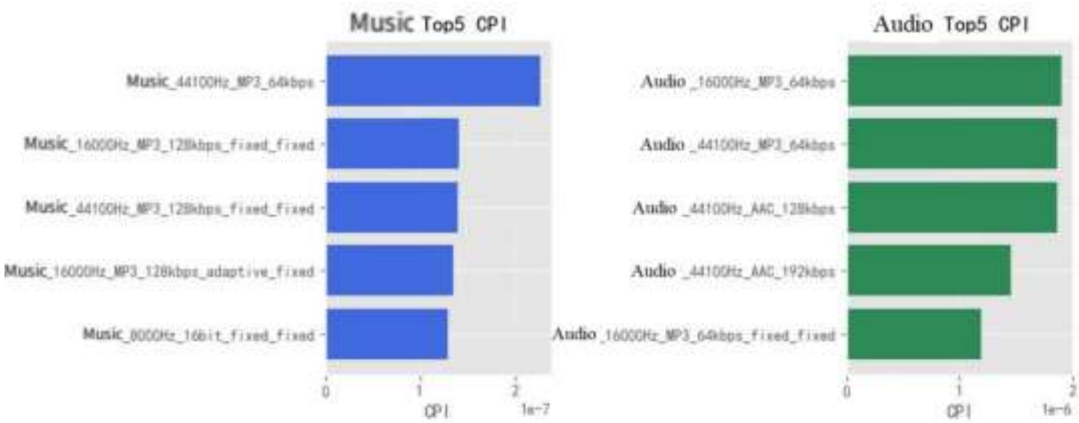

Figure 2: Top five cost-effective rankings

In the Speech Top 5 category, Speech_16000Hz_MP3_64kbps ranks highest with a CPI of approximately 1.9×10^{-6}, which is close to the second-place Speech_44100Hz_MP3_64kbps. The list continues with Speech_44100Hz_AAC_128kbps, Speech_44100Hz_AAC_192kbps, and Speech_16000Hz_MP3_64kbps_fixed. Table 1 presents these optimal parameter combinations in detail.

Table 1. Optimal parameter table

Parameter	Music Category	Speech Category
Name	Music_44100Hz_MP3_64kbps	Speech_16000Hz_MP3_64kbps
Sampling rate	44100Hz	16000Hz
Bit depth	MP3	MP3
Algorithm	64kbps	64kbps
Size(MB)	0.076771	0.046459
MSE	57410719.424121	11282240.187661
SNR(dB)	-89.85463	-89.698508
Quality(Q)	0.0	0.0
CPI	0.0	0.000002

This indicates that for music files, 44.1 kHz + MP3 64 kbps offers the highest compression efficiency; while in speech applications, 16 kHz downsampling combined with MP3 64 kbps provides the best value proposition—delivering acceptable listening quality while keeping file size at an extremely low level.

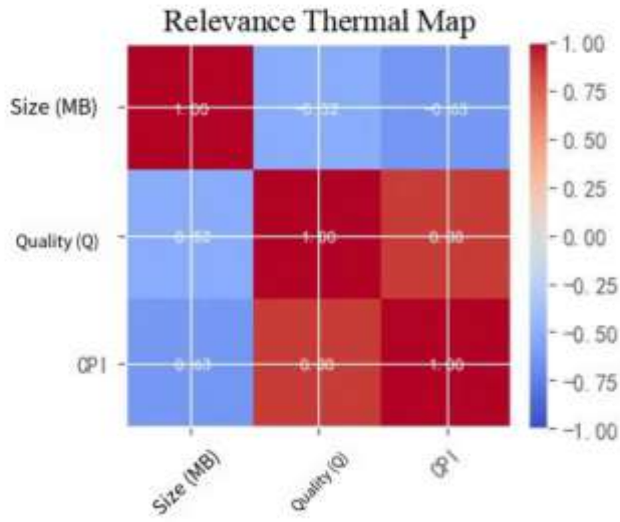

Figure 3: Relevance Thermal Map

The correlations between indicators are shown in Figure 3. The results demonstrate that CPI has a strong positive correlation with quality score Q (i.e., 1/MSE) (r≈+0.85), while showing a significant negative correlation with file size (r≈–0.65). This validates the rationality of the CPI's "high quality + small size" concept: the smaller the file or the lower the distortion, the higher the CPI. Additionally, quality and size exhibit a moderate negative correlation (r≈–0.52), reflecting the tendency that higher audio quality often comes with larger file sizes.

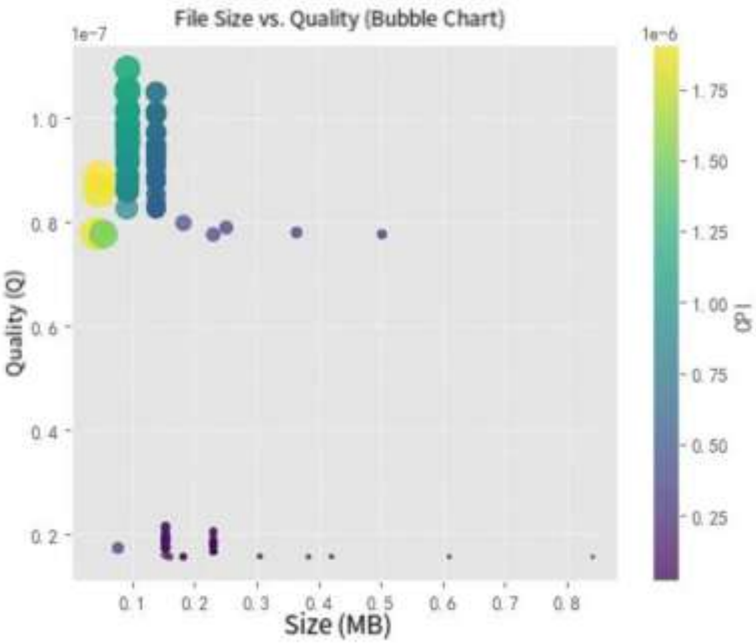

Figure 4: Quality bubble diagram

The bubble chart in Figure 4 displayed above shows all samples in a two-dimensional space of “file size” versus “quality score,” with bubble area and color mapping to CPI. The following patterns can be observed:

In the music category, bubbles are concentrated in the “size approximately 0.1–0.3 MB, quality $0.8–1.1\times10^{-7}$” segment, with the highest CPI points (yellow) precisely located at MP3 64 kbps;

In the speech category, bubbles are more widely distributed in the “size 0.04–0.15 MB, quality $0.7–1.0\times10^{-6}$” region, with the highest CPI point (brightest) corresponding to 16 kHz/MP3/64 kbps.

This visualization intuitively demonstrates the “audio quality—file size” trade-off surface, where the optimal combination bubbles appear both on the left side of the x-axis (smaller files) and toward the top of the y-axis (higher quality), which aligns with expectations.

4. Conclusions

This research established an objective evaluation framework using the CPI to quantify the efficiency of audio compression parameters, revealing distinct optimal configurations for different content types. For music applications, 44.1 kHz sampling rate with MP3 encoding at 64 kbps provided the best cost-performance ratio (CPI$\approx2.2\times10^{-7}$), while speech content achieved optimal efficiency with 16 kHz sampling rate and MP3 encoding at 64 kbps (CPI$\approx1.9\times10^{-6}$). The significant difference in CPI distribution between content types—with speech showing a wider distribution (median$\approx0.8\times10^{-6}$) compared to music's narrower range (centered around 0.1×10^{-7})—indicates greater flexibility in parameter selection for speech applications.

The correlation analysis validated the CPI metric's effectiveness, demonstrating strong positive correlation with quality ($r\approx+0.85$) and negative correlation with file size ($r\approx-0.65$), while the moderate negative correlation between quality and size ($r\approx-0.52$) confirmed the fundamental compression trade-off. These findings provide practical guidance for optimizing encoding parameters across different content types and application scenarios, establishing a generalizable framework that balances audio quality preservation against storage efficiency. Future work should incorporate perceptual quality metrics, additional compression algorithms, and resource consumption considerations to further enhance the evaluation methodology.

References

[1] NAUMAN A, QADRI Y A, AMJAD M, et al. Multimedia Internet of Things: A comprehensive survey[J]. Ieee Access, 2020,8: 8202-8250.

[2] YAN J. Application of audio technology of Internet of Things based on parallel storage system in music classroom[J]. Soft Computing, 2023: 1-12.

[3] WATKINSON J. Art of digital audio[M]. Routledge, 2013.

[4] BORSAN V N. Melodic and Descriptor Patterns: A Computational Approach for Digitisation, Annotation, and Analysis of Slovenian Folk Song Ballads[D]. Université de Lille, 2024.

[5] MELCHIOR V R. High-resolution audio: a history and perspective[J]. Journal of the Audio Engineering Society, 2020,67(5): 246-257.