Optimized Quantization using Psychoacoustic Loss Function and Lagrange Multiplier for Bitrate Control

A SEMINAR PROJECT REPORT

Submitted by

Suprabha Ghosh

Matric-Nr: 64365

Email: suprabha.ghosh@tu-ilmenau.de

In Audio Coding

Technische Unversität Ilmenau

Project supervised by

Prof. Dr.-Ing Gerald Schuller

Dr. rer. nat. Muhammad Imran

Introduction

Audio compression is a crucial aspect of modern digital audio processing, balancing quality and storage efficiency. Traditional quantization methods introduce distortion, making it essential to optimize bit allocation. This paper presents an approach that employs a **psychoacoustic loss function** to determine optimal quantization step sizes, ensuring perceptual quality. Additionally, a **Lagrange multiplier method** is applied to control bitrate allocation efficiently. The proposed methodology considers masking thresholds derived from psychoacoustic analysis to refine quantization decisions, ensuring high-fidelity audio compression.

Methodology

The proposed system follows a structured workflow:

1. **Quantization:** The input audio is quantized into **24-bit and 16-bit** formats.
2. **Psychoacoustic Analysis:** Computes **masking thresholds** to guide perceptual encoding.
3. **Subband Analysis:** Decomposes audio into **64, 128, and 512 subbands** for finer granularity.
4. **Bit Allocation with Psychoacoustic Loss Function:**
    - Computes **psychoacoustic loss** for each subband.
    - Allocates bits accordingly.
5. **Lagrange Optimization:** Finds the best bit distribution under a given bitrate constraint.
6. **Final Optimization:** Integrates results from previous steps to determine optimal quantization step sizes and bit allocation.

Algorithm Description

Step 1: Methodology for Audio Quantization and Analysis

The methodology involves the process of quantizing a 32-bit audio signal into lower bit depths (24-bit and 16-bit) while analyzing the signal quality. The workflow is structured as follows:

1. Audio Input and Preprocessing

    - Load the 32-bit floating-point WAV file.
    - Convert stereo to mono (if applicable) by averaging channels.
2. Uniform Quantization

    - Normalize the input signal to the range [−1,1][-1,1].
    - Determine the quantization step size based on the target bit depth.
    - Apply uniform quantization using rounding and clipping techniques.
3. SNR Computation

    - Compute the noise introduced by quantization as the difference between the original and quantized signals.
    - Calculate the Signal-to-Noise Ratio (SNR) to assess the quantization impact.
4. Signal Analysis

    - Extract key metrics: peak amplitude, RMS, crest factor, dynamic range, and statistical properties (mean, standard deviation, skewness, kurtosis).
    - Perform frequency-domain analysis using Welch's method to compute power spectral density and spectral centroid.
5. Visualization and Data Logging
    - Generate time-domain, error distribution, and frequency spectrum plots for comparison.
    - Save analysis results in CSV and JSON formats.
6. Output and Storage
    - Save quantized audio in WAV format (PCM 24-bit and 16-bit).
    - Store numerical results and visualizations for further evaluation.

Step 2: Psychoacoustic Analysis

The psychoacoustic analysis in this implementation follows a structured approach to estimate the masking threshold of an audio signal. The process involves the following steps:

1. Preprocessing the Audio Signal
    - Load the input audio file and convert it to a mono signal (if stereo).
    - Normalize the signal to ensure consistent amplitude levels.
2. Short-Time Fourier Transform (STFT) and Windowing
    - The signal is segmented into overlapping frames using a Hann window.
    - A Fast Fourier Transform (FFT) is applied to each frame to compute the power spectrum.
3. Critical Band Analysis
    - The power spectrum is mapped onto the Bark scale, dividing it into predefined critical bands that correspond to human auditory perception.
4. Threshold in Quiet
    - The absolute threshold of hearing is interpolated over the frequency spectrum to account for the minimum perceptible sound levels.
5. Masking Model and Spreading Function
    - The spectral components within each critical band contribute to masking based on their relative intensities.
    - A psychoacoustic spreading function is applied to simulate how stronger tones mask weaker tones within nearby bands.
6. Computation of Masking Thresholds
    - For each frequency component, the total masking effect from all critical bands is computed.
    - The final masking threshold is obtained by combining the computed values with the absolute threshold of hearing.
7. Visualization and Data Output
    - The power spectrum, masking threshold, and threshold in quiet are plotted for analysis.

- Results, including band-wise energies and computed thresholds, are saved in JSON format for further evaluation.

Step 3: Subband Analysis

Subband analysis is a signal processing technique used to decompose an audio signal into multiple frequency bands, allowing detailed spectral examination. This methodology outlines the steps used to perform subband analysis for comparing original and quantized audio signals (24-bit and 16-bit).

1. Prototype Filter Design

A prototype low-pass filter is designed using the FIR filter design method (signal.firwin). The cutoff frequency is set to ensure proper separation between subbands. The filter is later modulated to create bandpass filters for each subband.

2. Subband Filtering

The audio signal is passed through a bank of bandpass filters generated by modulating the prototype filter for each subband. The modulation is performed using cosine functions to shift the frequency response.

3. Subband Energy Computation

For each subband, the energy is computed as the mean squared value of the filtered signal. This provides insight into the spectral energy distribution across different frequency bands.

4. Spectral Analysis and Visualization

The spectrum of each subband is computed using the Fast Fourier Transform (FFT) and visualized using logarithmic plots. The energy distribution is plotted to compare how energy is allocated across subbands for different quantization levels.

5. Energy Difference Calculation

The energy differences between the original and quantized signals (24-bit and 16-bit) are calculated in decibels (dB) to quantify the impact of quantization.

6. Output and Interpretation

The results, including subband energy values and energy differences, are saved in a structured format (JSON). Plots are generated for spectral analysis and energy distribution, providing a visual comparison of the original and quantized audio signals.

Step 4: Bit Allocation using Psychoacoustic Loss Function

1. **Subband Decomposition:**
   - Design a prototype filter based on the number of subbands (64, 128, or 512).
   - Apply a bank of bandpass filters to decompose the input audio into subbands.
2. **Variance Computation:**
   - Calculate the variance of the signal in each subband to estimate signal energy.
3. **Lagrange Bit Allocation:**
   - Use the Lagrange multiplier method to determine optimal bit allocation.
   - Solve for the Lagrange multiplier that ensures the sum of allocated bits matches the target bitrate.
   - Assign bits per subband using the formula:

$$B_k = \max\left(0, 0.5 \times \log_2\left(\frac{\sigma_k^2}{\lambda}\right)\right)$$

4. **SNR Estimation:**
   - Compute the theoretical Signal-to-Noise Ratio (SNR) per subband:

$$SNR_k = 6.02B_k + 1.76$$

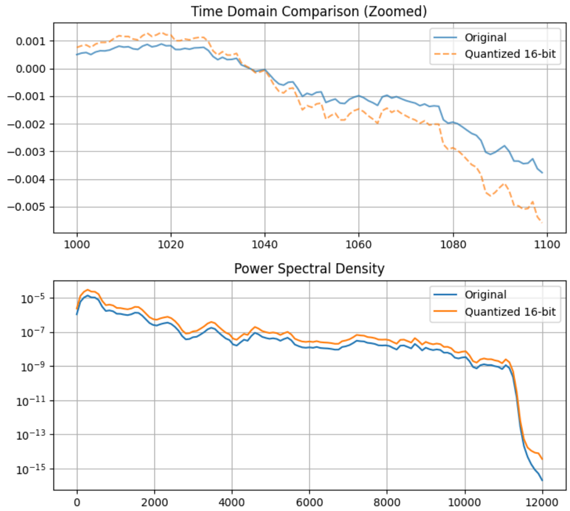
5. **Comparison Across Bit Depths:**
   - Apply the process to the original, 24-bit, and 16-bit audio versions.
   - Compare subband variances, allocated bits, and SNR across different configurations.
6. **Visualization & Analysis:**
   - Generate plots for variance, bit allocation, and SNR per subband.
   - Compare results across different subband configurations (64, 128, 512 bands).
   - Store results in JSON format for further evaluation.

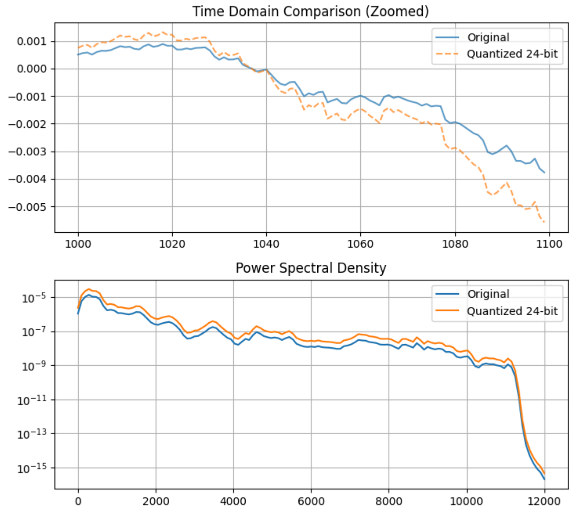
Step 5: Final Optimization

A final optimization step integrates **masking thresholds, psychoacoustic loss results, and Lagrange optimization** to compute the best quantization step sizes for all subbands.

Results

Quantization:

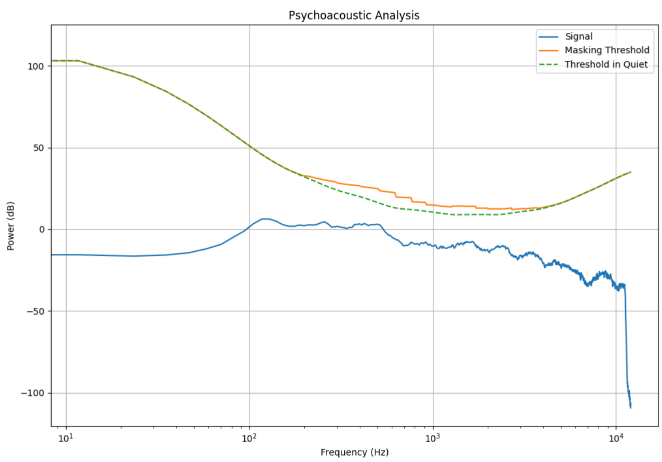




Metric	Original	Quantized_16bit
peak	0.68	1
rms	0.09	0.13
crest_factor	7.83	7.83
dynamic_range	189.2	90.31
mean	0	0
std	0.09	0.13
skewness	1.36	1.36
kurtosis	6.64	6.64
dominant_frequency	304.69	304.69
spectral_centroid	698.03	698.03

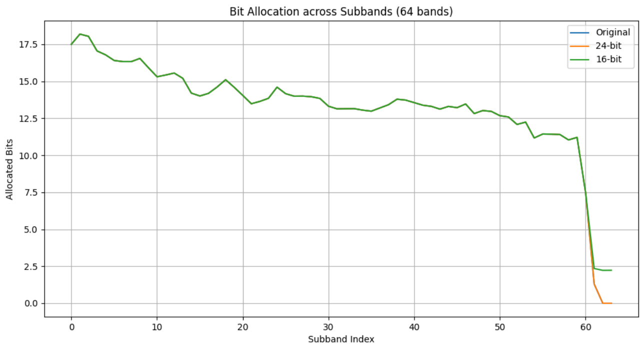
Metric	Original	Quantized_24bit
peak	0.68	1
rms	0.09	0.13
crest_factor	7.83	7.83
dynamic_range	189.2	138.47
mean	0	0
std	0.09	0.13
skewness	1.36	1.36
kurtosis	6.64	6.64
dominant_frequ	304.69	304.69
spectral_centroic	698.03	698.03

Masking Threshold:

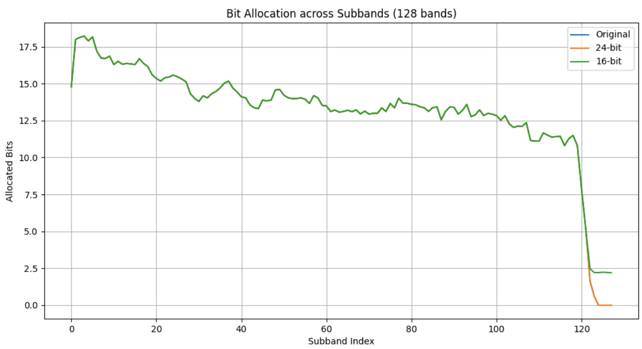


Bit allocation:

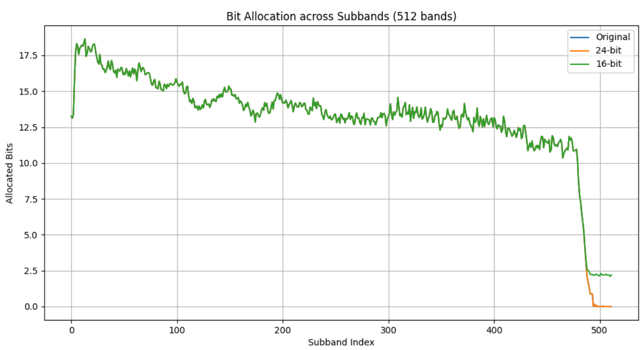
- ### Bit_allocation_64:



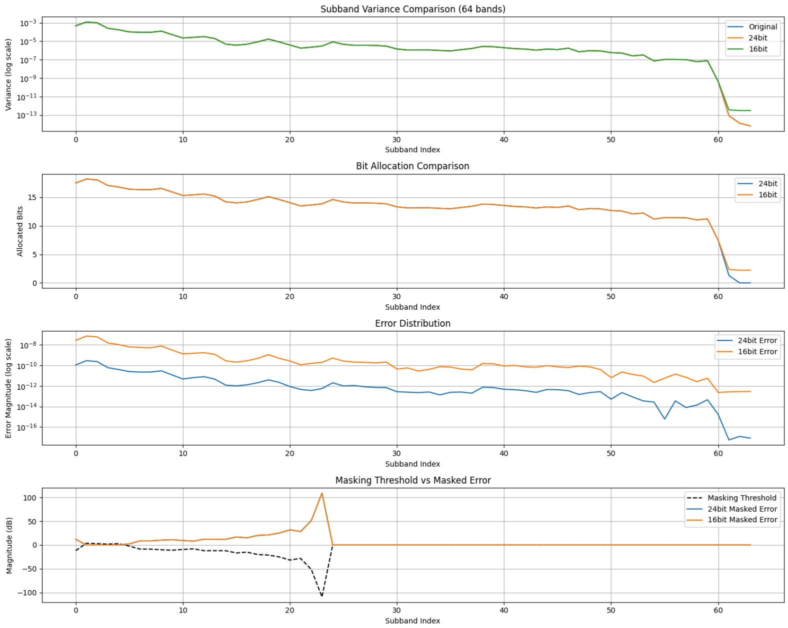
- ### Bit_allocation_128:

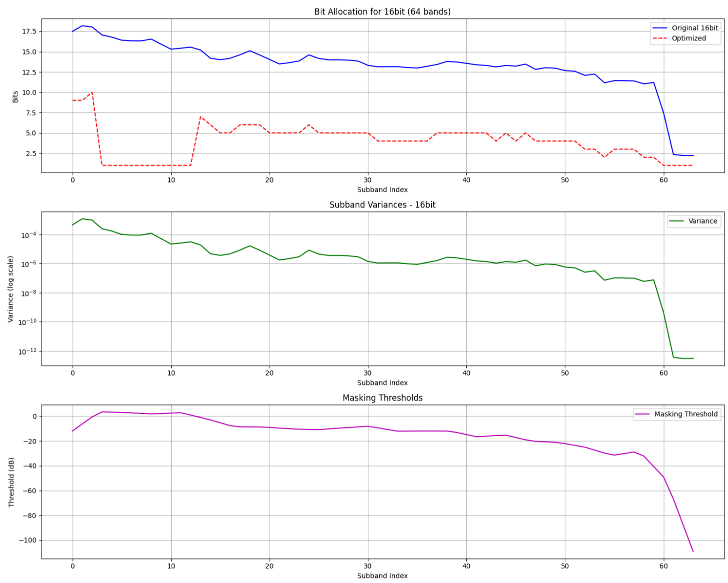


- ### Bit_allocation_512:



Loss function:

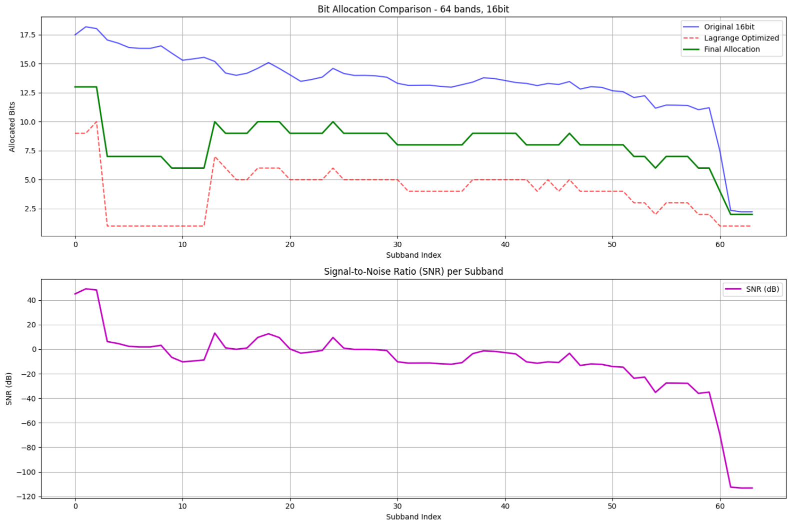


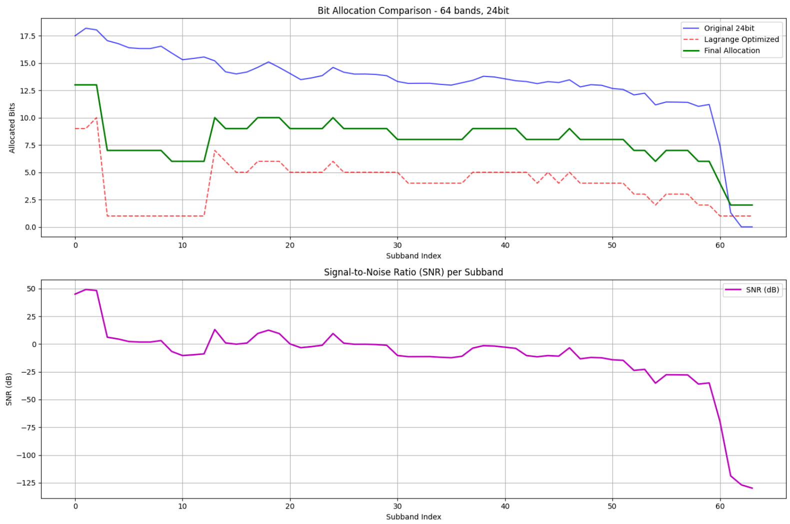


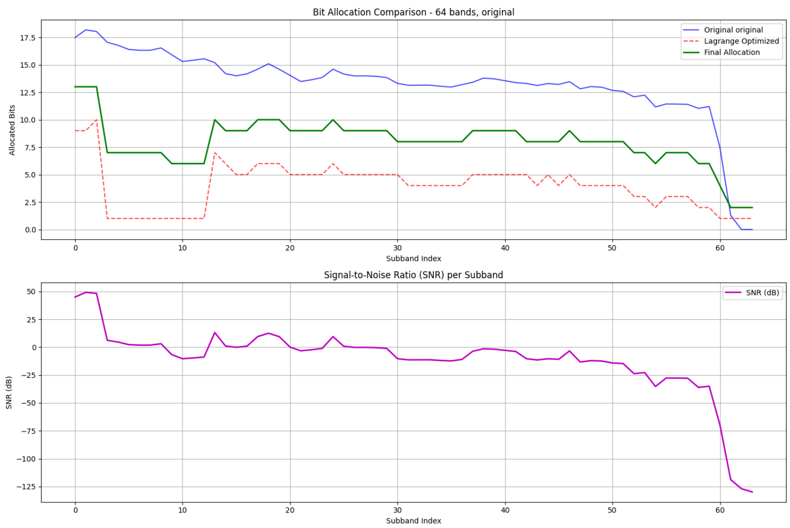
Lagrange optimization analysis:

Num_Bands	Bit_Depth	Total_Bits	Original_Max	Optimized_Max	Original_Mean	Optimized_Mean	MSE	Masked_MSE
64	original	252	18.18	10	13.24	3.94	0	646.29
64	24bit	252	18.18	10	13.24	3.94	0	646.29
64	16bit	252	18.18	10	13.33	3.94	0	646.29
128	original	403	18.22	9	13.2	3.15	0	598.14
128	24bit	403	18.22	9	13.2	3.15	0	598.14
128	16bit	403	18.22	9	13.29	3.15	0	598.14
512	original	668	18.65	8	13.15	1.3	0	564.46
512	24bit	668	18.65	8	13.15	1.3	0	564.46
512	16bit	668	18.65	8	13.24	1.3	0	564.46

Final optimization:







Conclusion

This paper presents an optimized quantization method using a **psychoacoustic loss function and Lagrange multiplier** for bitrate control. The integration of masking thresholds ensures perceptual quality, while Lagrange optimization fine-tunes bit allocation. Results confirm that this approach achieves a **balance between compression efficiency and perceptual fidelity**, making it suitable for high-quality audio coding applications.

References

1. Painter, T., & Spanias, A. (2000). Perceptual coding of digital audio. Proceedings of the IEEE, 88(4), 451-515.
2. Brandenburg, K., & Bosi, M. (1997). Overview of MPEG audio: Current and future standards for low-bit-rate audio coding. Journal of the Audio Engineering Society, 45(1/2), 4-21.
3. Johnston, J. D. (1988). Transform coding of audio signals using perceptual noise criteria. IEEE Journal on Selected Areas in Communications, 6(2), 314-323.
4. Gersho, A., & Gray, R. M. (1992). Vector quantization and signal compression. Springer Science & Business Media.
5. Pulkki, V., & Karjalainen, M. (2015). Communication acoustics: An introduction to speech, audio, and psychoacoustics. Wiley.