Audio Signal Compression Using the Wavelet Transform

Trịnh Thị Thu
Thai Nguyen University of Information and Communication Technology
Thai Nguyen University, Thai Nguyen, Vietnam

***Abstract*—This paper summarizes a 2007 undergraduate thesis on audio compression with a wavelet transform. The work reviews digital audio, MPEG perceptual coding, Fourier and short-time Fourier analysis, continuous and discrete wavelets, and wavelet packets. It then implements a multilevel wavelet decomposition with coefficient selection and an experimental psychoacoustic masking model. The reported system uses a five-level db4 decomposition and 2,048-sample frames for the psychoacoustic analysis. Across six tested music or speech categories, the thesis reports average compression ratios from 3.98 to 8.46, with subjective labels ranging from "transparent" to "original preferred." A separate spoken example reports an SNR of 17.45 dB after reconstruction. The document does not specify a formal listening protocol, uncertainty intervals, or a like-for-like codec baseline. Its results therefore demonstrate the implemented method and selected examples rather than establishing general superiority over MPEG or other codecs.**

***Index Terms*—audio compression, discrete wavelet transform, psychoacoustic masking, wavelet packets, signal-to-noise ratio**

I. Introduction

Audio compression aims to reduce storage and transmission cost while preserving acceptable reconstructed sound. Transform coding separates components that can be represented with fewer coefficients; perceptual coding further considers human hearing and masking. The source thesis compares the concepts behind MPEG audio coding with wavelet-based analysis and explores a wavelet approach to speech and music compression [1], [6], [2], [10].

The thesis first reviews the relevant signal-processing foundations, including sampling, discrete Fourier analysis, the short-time Fourier transform, and continuous, discrete, and packet wavelets [1], [2], [4], [5]. Its applied objective is to implement decomposition, coefficient removal, coding, and reconstruction, then examine compression ratio and recovered audio quality on several audio categories.

II. Methods

A. Transform and coefficient selection

The described implementation decomposes an input signal using the db4 wavelet to five levels. It represents the signal as an approximation component at level five and detail components at levels one through five. Coefficients outside a selected threshold are removed before coding and later reconstruction [1]. The source discusses threshold choice, decomposition level, and voiced and unvoiced speech as design considerations. The general multiresolution basis follows established wavelet and filter-bank methods [2], [3], [4], [5].

Supervisor: Phung Trung Nghia, Thai Nguyen University of Information and Communication Technology (ptnghia@ictu.edu.vn).

TABLE I
Reported compression results for six audio categories [1].

Audio category	Ratio	Reported quality
Drums (solo)	3.98	Transparent
Pop	5.12	Nearly transparent
Speech	4.46	Transparent
Piano (solo)	4.23	Transparent
Jazz	7.75	Original preferred
Rock	8.46	Original preferred

B. Psychoacoustic analysis and evaluation

The thesis adds an experimental psychoacoustic model. Each 2,048-sample frame is transformed to the frequency domain, divided into critical-band regions, and inspected for local spectral peaks that can mask neighboring components. The proposed rules use local differences in spectral level to identify masking points and guide quantization [1]. The implementation is illustrated with encoder and reconstruction modules.

For evaluation, the source reports average compression ratios and qualitative recovered-audio judgments for drums, pop, speech, piano, jazz, and rock. It also illustrates a short spoken signal and gives its original and reconstructed sizes and an SNR value. The qualitative labels were obtained by direct assessment in the system; the document does not describe listener recruitment, blinding, or a formal listening scale [1].

III. Results

The source presents waveform plots before and after compression and reports the ratios in Table I. The larger reported ratios for jazz and rock accompany the label "original preferred," which indicates that the source's qualitative judgment favored the original signal for those categories [1]. The document does not report bitrate, codec configuration, or objective distortion for each row, so the ratios should not be converted into a general perceptual-quality ranking.

For a separate spoken example ("one, two, three"), the thesis reports an original size of 18.8 kB, a reconstructed size of 18.6 kB, and SNR of 17.45 dB. These figures characterize that example only; they are not a mean over the six categories in Table I [1].

IV. Discussion

The results show a working wavelet-based compression and reconstruction workflow. They also highlight a trade-off: the highest compression ratios in the table coincide with a subjective preference for the uncompressed original. This is consistent with the need to choose thresholds and decomposition levels according to both compactness and perceived quality, but the thesis does not quantify that trade-off with a controlled listening study.

The psychoacoustic rules are described as experimental. A reproducible comparison would need the input audio files, wavelet and threshold settings, bitrate or file-size definition, a fixed baseline codec, and objective or blinded subjective quality measurements. The source itself proposes improving speech quality and compression ratio and extending wavelet methods to speech recognition [1].

V. Conclusion

The 2007 undergraduate work implemented and evaluated audio compression using multilevel wavelet analysis and a psychoacoustic component. Its reported ratios and one SNR measurement establish what the prototype achieved on the selected signals. The absence of a formal listening protocol and matched codec baseline limits stronger claims about relative performance [1].

References

[1] T. T. Thu, "Audio signal compression using the wavelet transform," undergraduate thesis (original in Vietnamese), Thai Nguyen University of Information and Communication Technology, n.d. [Online]. Available: ICTU repository record.

[2] G. Strang and T. Nguyen, *Wavelets and Filter Banks*. Wellesley, MA, USA: Wellesley-Cambridge Press, 1996.

[3] A. N. Akansu and R. A. Haddad, *Multiresolution Signal Decomposition: Transforms, Subbands, and Wavelets*. San Diego, CA, USA: Academic Press, 1992.

[4] I. Daubechies, *Ten Lectures on Wavelets*. Philadelphia, PA, USA: SIAM, 1992.

[5] S. G. Mallat, "A theory for multiresolution signal decomposition: The wavelet representation," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 11, no. 7, pp. 674–693, 1989.

[6] N. Q. Trung, *Signal Processing and Digital Filters* (in Vietnamese). Hanoi, Vietnam: Science and Technology Publishing House, 1998.

[7] N. V. Anh and N. H. Hải, *Wavelet Analysis Tools and Applications in MATLAB* (in Vietnamese). Hanoi, Vietnam: Science and Technology Publishing House, 2005.

[8] N. V. Anh and N. H. Hải, *MATLAB Programming and Applications* (in Vietnamese). Hanoi, Vietnam: Science and Technology Publishing House, 2005.

[9] M. Zanartu, "Audio compression using wavelet techniques," ECE 648 course report, Purdue University, 2005, as cited in the source thesis.

[10] D. Sinha and A. H. Tewfik, "Low bit rate transparent audio compression using adapted wavelets," *IEEE Trans. Signal Process.*, vol. 41, no. 12, pp. 3463–3479, 1993, doi: 10.1109/78.258086.