Compression of wideband sound for multimedia application

Conference Paper · March 2003

DOI: 10.1109/CADSM.2003.1255089 · Source: IEEE Xplore

2 authors:

Adrian Rominski
Lodz University of Technology

Zygmunt Ciota
Lodz University of Technology

Wideband Sound Compression for Multimedia Application

ADRIAN ROMINSKI, ZYGMUNT CIOTA
Department of Microelectronics and Computer Science
Technical University of Lodz
Al. Politechniki 11, 90-924 Lodz
POLAND

Abstract: - Two basic methods of sound compression: MPEG and Ogg-Vorbis have been discussed and compared. Mixed hardware-software approach can improve the speed and also the quality of the compression. Furthermore, speech coding system has to have real-time capabilities, so the hardware-software co-design permits to achieve low cost and high-speed performances. Predictive method of speech compression has been also presented. Seeking for more efficient solutions, systolic structures using fuzzy-neural predictors seems to be a smart realization of an effective speech processor. Mixed analog-digital realizations of integrated filter banks have been also discussed.

Key-Words: - MPEG codec, Vorbis codec, speech processing, voice compression

1 Introduction

An equivalent of Linux system in audio compression formats is vorbis-ogg format. Now state of the project could be described as Release Candidate 3, it means that offered codecs, also in source code version, for the most popular operational systems, machines and programs (Linux/UNIX, Windows, Macintosh, BeOS, Winamp and other players plug-ins) are fully functional and ready to use, moreover there is no discovered errors in the program operation.

Users tests proved that quality of sound at the same bitrate is better than for MPEG audio layer 3 (famous and most popular MP3) and comparable to lately created MP3pro. Of course, the popularity of Ogg is still less than widely spread MP3, but Internet query showed up, that it is the second most used format for compression after MP3, but more often used than Microsoft Windows Media Audio 8.

Not only music is a place for audio compression. Sound tracks in films are very promiscuous; tandem DivX – Ogg seems to work now only in 2 channels, but multi-channel surround sound is coming and it will be alternative for Dolby Digital, DTS and MPEG. Also other branches of electronic entertainment started to use the new format. Many computer games were released with sound in Vorbis-Ogg, one of them Schizm were awarded with Seal Of Excellence – Adrenaline Vault.

Ogg format was not created in the vain and utilizes experiences of yet created formats. Some main features, characterized good, modern audio codec were established:

- Effective compression based on MDCT, but with opportunity of using other methods
- Variable Bitrate (VBR), which accommodates bitrate to signal characteristics, but with possibility to obtain restricted bitrate in order to utilize transmission channels with fixed bitrate for a signal.
- Multi-channel (2 and much more than 2) feature with possibility of using correlation between channels in compression process
- GPL license
- Ogg is predefined for using in streamable digital media as Internet radio, Voice Over IP telephony, communication between players in computer games, etc.
- Tags system useful for description of files with predefined tags and user definable tags

Today we stand in front of a big problem, growing rapidly in multimedia, but also other information and contents. These problems, mostly derives from extraordinary amount of multimedia information. One of the Peer-To-Peer file exchange client offers hundreds of terabytes compressed music, it equivalent of human life duration. In that situation is obvious, that one man cannot listen all these files, so it is necessary to index and describe all pieces of the information. Contemporary tag system works now, but one day may not be enough, even set of headers will overcome limits of human perception.

One of the solutions could be splitting data base between users using different contents and then let them interchange information. Maybe, some information should be also incrementally deleted to make room for new things (how to decide which?).

Another challenge are new methods of coding which will allow algorithmic coding of sound, like sound synthesis with type of voice choosing, coding of music with vocal with newer equivalent of MIDI format.

2 Comparison Tests

We have made some test based on objective (not the best in case of subjective lossless formats, but always objective) criteria. Because Vorbis codec is VBR (Variable Bitrate), the same option was used in MPEG layer-3 codec. The coding was made with Vorbis RC3 encoder and LAME MPEG layer-3 encoder considered as one with the best quality, mp3 encoder. The codecs were set up to obtain the same average bitrate of about 128 kbit/s, coding was made from source file of CD-audio quality (44100Hz, 16 bit/sample, stereo). The set of parameters for mp3 coding is shown in Figure 1, while the parameter for Vorbis codec is shown in Figure 2. Because Vorbis is made as VBR from the beginning, the average bitrate is derived from the quality factor, but there is a method to set lower and upper limit for the bitrate.

The first criteria, very often used, is bandwidth of encoded sound, 3 seconds of white noise was used to examine the codecs, as it was shown in Figure 3. Vorbis encoder performs all the bandwidths, while mp3 encoder (see Figure 4) has a threshold at about 12 kHz. It is worth to say that subjectively high frequency sound are weakly audible.

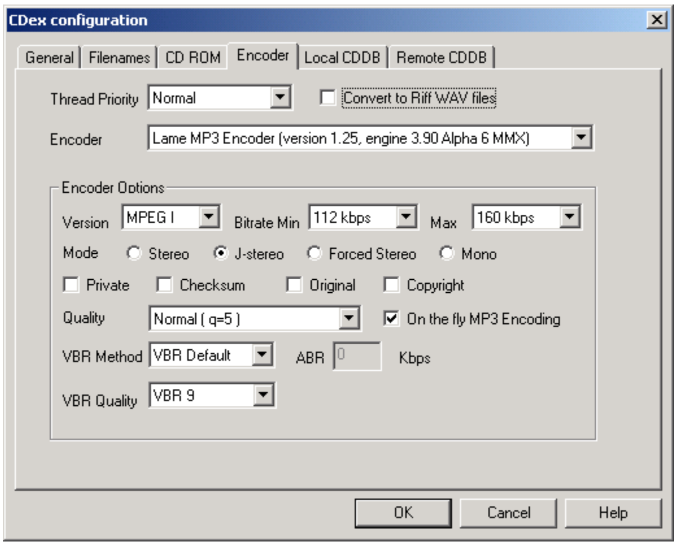

Fig. 1. Parameters for MP3 coding

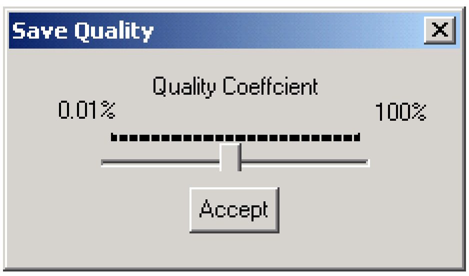

Fig. 2. Parameter set for Vorbis codec

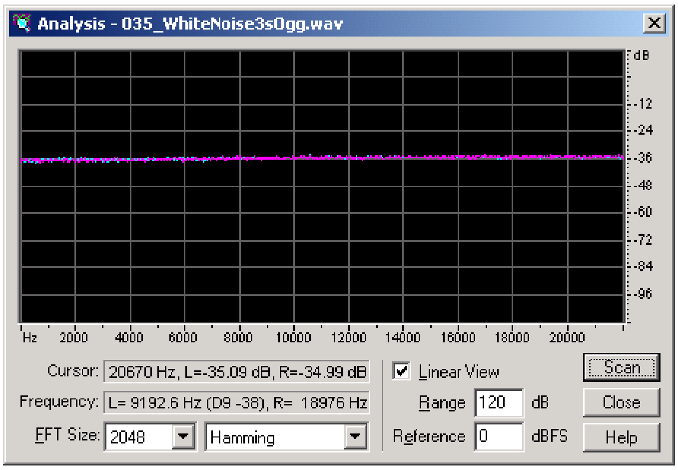

Fig. 3. Bandwidth of encoded sound - Vorbis encoder

Channel separation was very good in both cases, so this test is not worth of publishing.

The third test has been performed as pre- and post-echo test. Encoders often doesn't work properly leaving/adding some artifacts before and after original sound. We used a sound consisted of 1 second of silence, 3 seconds of white noise and then again 1 second of silence. As it is shown in Figure 5 the sound has delay, because the noise should finished at 4th second, and then there is a long period of post-echo fading - about 10ms in comparison with 2ms of Ogg's post-echo (Figure 6).

We can also observe that there is no sound delay in the case of Ogg Vorbis. Pre-echo is similar as post-echo for both codecs.

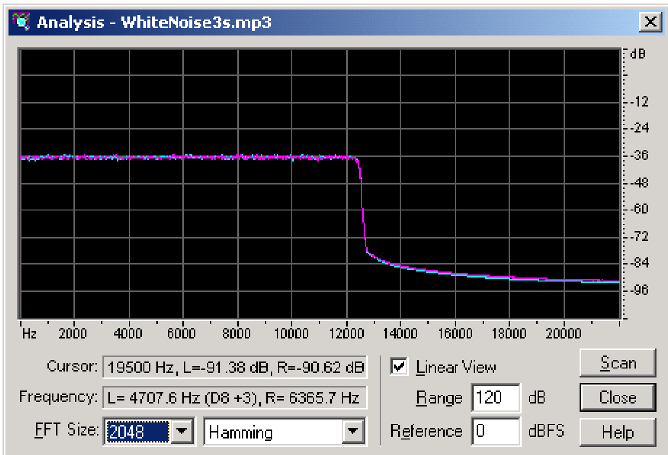

Fig. 4. Bandwidth of encoded sound – MP3 encoder

This 3 tests revealed an advantage in the objective criteria, but it should be mentioned that it is not a "full truth" about codecs. The most important is a people opinion based on subjective criteria but sometimes also connected with customs, prejudices and economic criteria. One of the important criteria that could be taken into consideration is encoder speed. Because Vorbis is rather fresh and under development, so there was no possibility to optimize it, like LAME, which uses MMX and has strongly optimized code. One of the way of optimization could be through the hardware acceleration. For example, such often repeated operations, like MDCT transform or subband synthesis based on the bank of digital FIR filters, are predisposed to realize them in hardware form.

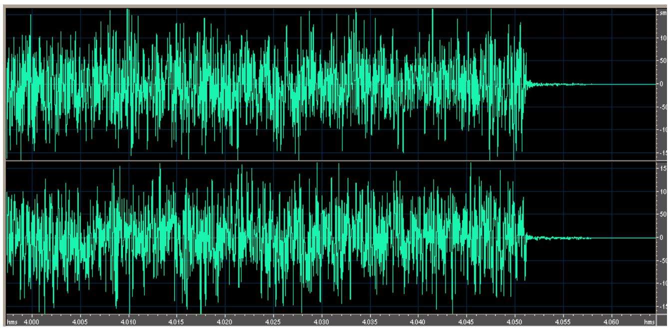

Fig. 5. MP3 codec: pre- and post-echo test

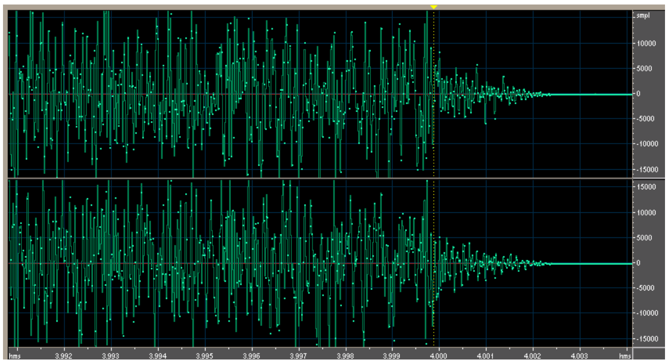

Fig. 6. Vorbis codec: pre- and post-echo test

3 Compression in Speech Processing

The latest development of multimedia systems demands new methods of speech processing like effective coding, recognition and synthesis [1,2,3]. The vocal communication between computer and man has nowadays a lot of gaps, because language is very essential evidence of humanity. As a consequence, it is necessary to improve the transformation methods of speech signal for obtaining more efficient transmission procedures and also for an enhancement of intelligibility. New digital signal processing permits to apply a lot of efficient algorithms (using deterministic and probabilistic approaches) for speech processing.

Let's take a digital process (digital quantization of time-varying speech signal) of speech compression and voice reconstruction using linear prediction approach. Applying digital simulation we can very easy observe two kinds of correlation. Each sample of voiced speech is highly correlated with the corresponding sample, which occurred one pitch period earlier.

Additionally, each sample is also correlated with neighboring samples. For unvoiced speech signal the correlations are inferior, but fortunately, the shape of invoiced signal is irregular like noise, so distortions caused by quantization noise are very low (an explication of this phenomena is simple: because noise plus noise gives also noise, the influence of sampling process is significantly low). The construction of speech synthesis procedure corresponds to the design of a synthesis filter described in the z domain by the following transfer function: H(z) = 1/A(z), where A(z) is a predictive filter, built as finite impulse response (FIR) filter [5]. The coefficient of FIR filter can be calculated using predictive approach, it means we approximate a current sample by linear combination of immediately preceding samples.

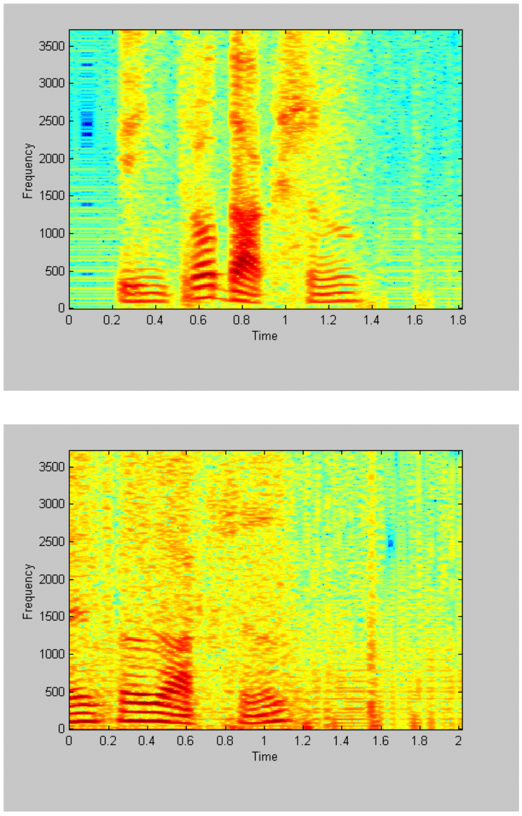

Fig. 7. Two spectrograms of the same word Introduction (two different speakers)

Designing a speech processing system is inherently a complex task involving human expertise as well as aids intended to accelerate the process. Furthermore, such efficient system has to have real-time capabilities, so the hardware-software co-design permits to achieve low cost and high-speed performances. While microcontrollers and microprocessors are inherently digital components, some functions can be executed in analog or digital form.

Since the language is ambiguous, then it will be useful to include fuzzy approach to obtain more effective methods of speech processing. Application of fuzzy logic can be useful not only for speech synthesis, but also in the case of speech recognition. The simplification of speech process in computer systems gives a lot of redundancy, so fuzzy logic approach to speech prediction seems to be promising solution [1]. An example presenting spectrograms of the same word Introduction as a fuzzy process is shown in Fig. 7.

On the other hand, in the case of fuzzy system, we have to use in the beginning of the design process, pre-defined membership functions and a linguistic model, applying an expert knowledge. Unfortunately, the system must work correctly with different users and very often with different languages. So, we have very limited possibility to adapt our speech processor to strongly varying conditions.

Using fuzzy logic approach we can recognize the proper shape of the vowel in the following manner. Each formant has been qualified using two parameters: dominant frequency and amplitude. The knowledge about the first three formants permits to recognize the whole vowel. The main task is to fit the current formants to the library patterns. For each formant we defined two membership functions. The fuzzy rule base has been filled using several empirically devised rules [1].

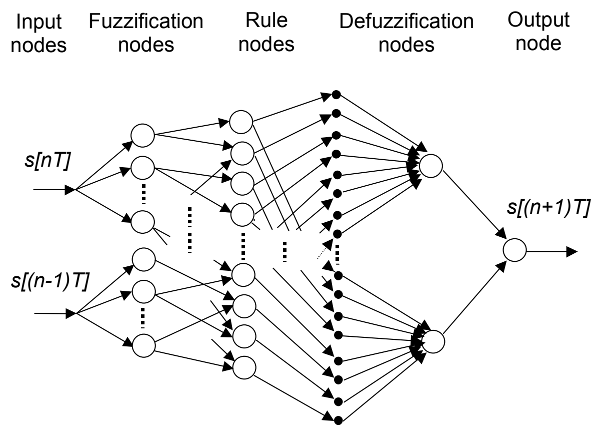

Fig. 8. Example of neural predictor

Better solution can be obtained combining fuzzy system and an artificial neural network (ANN). Such fuzzy-neural system permits to add two important developments: learning function and adaptive possibilities. The linear predictor can be realized by using multi-layer ANN. An example of such fuzzy-neural controller is shown in Fig. 8 [1,4].

Purely software realization of the system gives the principal contradiction: parallel calculations (neural network) are executed by serial paradigm (computer). Unfortunately, fully hardware realization is impossible, because physical connections between the neural units present a big technological problem. The better solution gives mixed parallel-serial approach. Using mixed digital-analog realization can increase the efficiency of such solution. Going over the nowadays solutions of the neural networks, the systolic structure gives the biggest possibility for smart realization of computer speech processor.

4 Conclusions

Application of mixed digital-analog realization to the design process of sound processors may be better in comparison with purely digital solution and very often we can achieve better results (decreasing of the chip surface and increasing of the speed parameter of the circuit). Different analog approaches of the programming of such systems have been presented [5], including time-sharing method. Advanced adaptive systems can work efficiently thanks to the additional control system, built as digital circuit, which applies the time-sharing principle.

The FIR filters are the most frequently used in adaptive systems. The application of the least-mean-square (LMS) algorithms and its modification to the control of FIR filter gives the approximation of the Wiener solution [5]. The simplicity of such approach is used very often in many practical implementations [1,2,5].

References:

[1] Z. Ciota, Efficient algorithm of adaptive filtering for real-time applications, Proc. Adaptive Systems for Signal Processing, Communications and Control (AS-SPPC'2000), Lake Louise, Canada, October 1-4, 2000, pp.299-303

[2] M. Szermer, M. Jankowski, Z. Ciota , A. Napieralski, M. Daniel, K. Slusarczyk, Efficient measurements of current mode circuits using LABVIEW environment, 8th International Conference MIXDES'2001, June 21-23, 2001, Zakopane, Poland

[3] Ch. I. Byrnes, P. Enqvist, A. Lindquist, “Cepstral coefficients, covariance lags, and pole-zero models for finite data strings”, IEEE Transactions on Signal Processing, vol. 49, no. 4, April 2001, pp.671-693

[4] W.Lin, S. N. Koh, X. Lin, “An 8.0-/8.4-kbps wideband speech coder based on mixed excitation linear prediction”, Signal Processing, vol. 81, 2001, pp. 1437- 1448

[5] S. Haykin, Adaptive filter theory, Prentice Hall, Englewood Cliffs, 1991