Compression Of Wideband Sound For Multimedia Application

Adrian Romiński, Zygmunt Ciota[1]

Abstract – **Two methods of sound compression: MPEG and Ogg-Vorbis have been discussed and compared. The objective criteria has been used during test procedure. Mixed hardware-software approach can improve the speed and also the quality of the compression. Audible noise cancellation in high quality speech processing has been also presented. 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.**

Keywords – **Multimedia, GPL, compression, audio, ASIC, FIR, DSP**

I. INTRODUCTION

GPL (General Public License) made up by Richard Stallman is completely new idea of software and intellectual property distribution. GPL guarantees free of charge possibility of coping, using and access to the source code of the program. Idea of free software written by a little more experienced users and given to whole community had been ignored long time until Linux started up as the professional and competitive product to the most powerful operational systems which cost a lot of work and money and are created by the most powerful firms of the world.

We can observe two strategies of big firms being active in this new reality: combating against this new trend with all means, while the second strategy is a trial of cooperation and using this new experience in own products, vide Sun corporation.

The second part of the article deals with the hardware implementation of algorithms for sound processing. The earlier mentioned GPL licensed solution are also very attractive in this context.

II. ORBIS-OGG AS THE NEW GPL COMPRESSION FORMAT

An oppressive license policy in time of CD-writers, printers, Internet and other technical possibilities of coping and spreading any kind of media but also in time of huge resources of created programs, multimedia contents, seems to be a little bit out of date.

An equivalent of Linux system in audio compression formats is vorbis-ogg format [1]. It is based on GPL and is available absolutely for free. 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 MP3) and comparable to lately created MP3pro, but without any fees also for coders, coding and using in hardware players at any bitrate [2]. 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 strongly promoted 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 strong 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 channels) 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

III. SOME OTHER FEATURES WHICH WILL BE USEFUL IN THE FUTURE

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,

[1]Technical University of Lodz Department of Microelectronics and Computer ScienceAl. Politechniki 11, 90-924 Lodz, Poland

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.

Xiphomorphous intend to create also moving picture coding (Tarkin project), it could be very interesting, specially with new restrictive licensing policy of MPEG 4, when many firms tries to find an alternative.

IV. COMPARISON TEST

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.

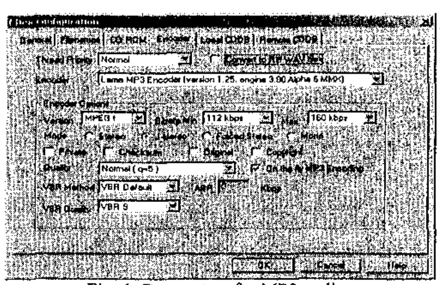

Fig. 1: Parameters for MP3 coding

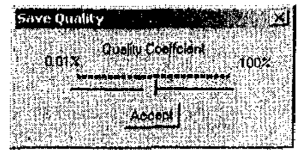

Fig. 2: Parameter set for Vorbis codec

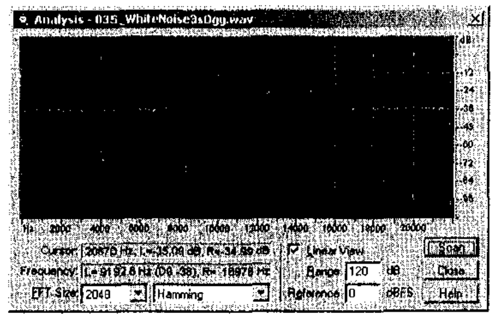

Fig. 3: Bandwidth of encoded sound - Vorbis encoder

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 sec. 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).

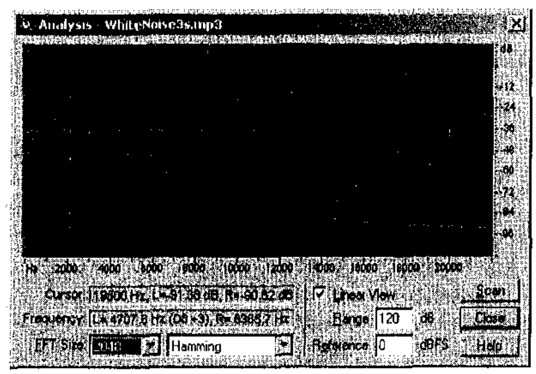

Fig. 4: Bandwidth of encoded sound – MP3 encoder

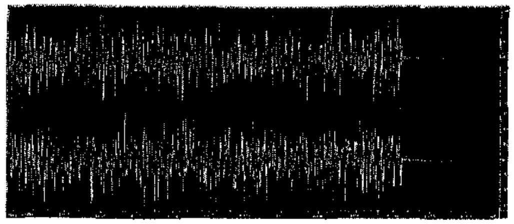

Fig. 4: MP3 codec: pre- and post-echo test

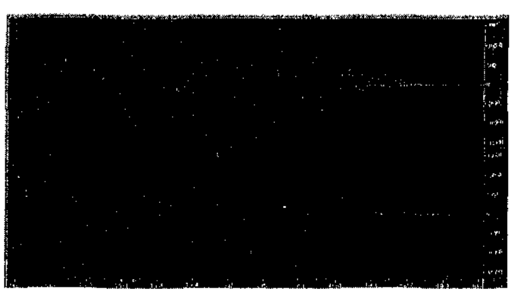

Fig. 5: Vorbis codec: pre- and post-echo test

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.

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 i 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 (see Figure 7 and 8).

V. HIGH QUALITY VOICE PROCESSING

The latest development of multimedia systems demands new methods of speech processing like effective coding, recognition and synthesis [3,4,5]. 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.

In this section an important problem has been discussed: how to improve the signal-to-noise ratio owning to digital structure of a speech signal. 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 [3, 4]. 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.

The sampling technique permits to obtain high quality speech under the condition, that noise spectrum has to be located under the three first formants (resonances) of the synthesized voice, according to Figure 6. The proper shape of noise spectrum can be achieved by using adaptive filters.

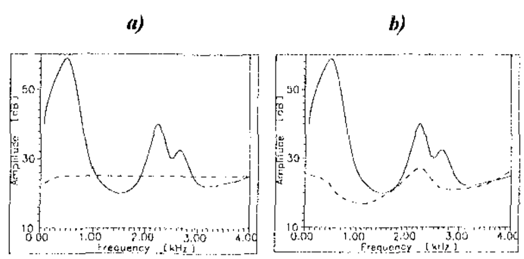

Fig. 6: The spectral representation of a vowel sound (solid line) and quantizing noise (dashed line): a) - audible noise, b) - audible distortions have been canceled after optimization

Adaptive filters are very important and responsible components in today's telecommunication systems. They are widely applied in different applications like full-duplex transmissions of telephone lines for acoustic echo cancellation, blind source separation, channel equalizers and noise cancellation. Since modern adaptive filters have to be suitable for real time applications, it is necessary to build appropriate fast algorithms. Such adaptive algorithms should be a compromise between high convergence, low steady-state error and fast tracking properties.

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 [3]. The simplicity of such approach is used very often in many practical implementations. Sometimes, LMS algorithm may be also applied in a non-stationary (but slowly varying only) environment.

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.

Application of mixed digital-analog realization to the design process of adaptive structures 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 [3,5], 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.

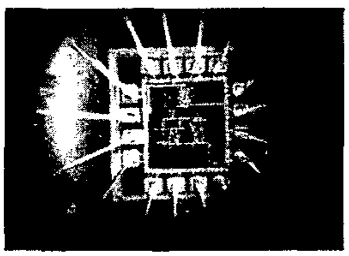

Fig. 7: Microphotograph of current mode CMOS components (A/D converter and transconductance amplifiers)

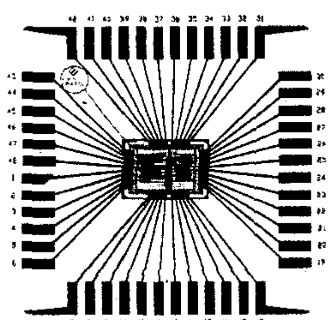

Fig. 8: The bank of switched current filters (CMOS 0.8μm technology)

We expect that hardware realization of some components (e.g. filters, neural network units) can be useful for smart real-time applicable .We have performed some preliminary works concerning current-mode CMOS components (see Fig. 7 and 8) [3, 4].

VI. CONCLUSION

Combining GPL licensed algorithms with effective hardware realization can be new way of audio signals processing. High efficiency of free software solutions like Ogg-Vorbis and their hardware realizations could be a base of new digital radio broadcasting, music distribution and the solution for modern audio signal processing at all.

VII. ACKNOWLEDGEMENTS

This work was supported by the EU programme Research and Training Action for System on Chip Design REASON IST-2000-30193.

VIII. REFERENCES

[1] www.vorbis.com

[2] www.mp3-tech.org

[3] 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

[4] 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

[5] Z. Goh, K.-C. Tan, B.T.G. Tan, Estimating time-varying noise introduced by CVSD for speech enhancement, *IEE Proc.-Vis. Image Signal Process.*, vol. 145, February 1998, pp. 23-29