

Embedded Reconfigurable Hardware Platform for Real-Time Audio and Speech Signal Processing

Michael Livshitz, Alexey Petrovsky, Andrey Stankevich, Mikhail Kachinsky,
Alexander Petrovsky

Computer Engineering Department, Belarusian State University of Informatics and Radioelectronics, P. Brovki, 6,
220013, Minsk, Belarus, mlivshitz@tut.by, petrovsky@bsuir.by, stankevich@bsuir.by, kachinsky@bsuir.by,
palex@bsuir.by

Abstract – This paper deals with reconfigurable hardware platform for different purposes real-time speech and audio signal processing. A design conception and turnkey solution are described. Much attention is paid to reconfigurable peripheral processor meant for external interface realization, pre- and post- data processing as well as digital signal processing algorithms implementation with the object of the DSP unloads. Moreover, three applications implemented on the considered platform are demonstrated.

Keywords – DSP, FPGA, DISC, Reconfigurable Structure

I. INTRODUCTION

The effectiveness of hardware platforms traditionally based on DSP executing user-written program can be significantly increased using FPGA co-processor. At the same time, increasing of effectiveness is achieved by means of inherent FPGA flexibility that provides implementation of not suitable operations for DSP freeing last one for data processing task. Loading these instructions into FPGA and leaving target instructions that need high-speed processing for DSP performance and cost of the DSP platform are optimized. Moreover, DSP platform that includes FPGA co-processor can easily be reorganized to meet the requirements of the current task by algorithms distribution among DSP and FPGA.

FPGA co-processor can realize function of peripheral processor, implementing hardware platform interface with specific external equipment in audio signal processing tasks.

II. HARDWARE PLATFORM DESIGN CONCEPTION

The suggested approach is strongly aligned with DISC (Dynamic Instruction Set Computer). The Core Processor (See Fig. 1) presents a static component of the system, the main task of which is control: managing I/O operations, initiating of application-specific instructions, and providing the reconfiguration sequences in so-called Instruction Space.

The latter component is a dynamic part of the system usually built on reconfigurable FPGA(s). Configurations

implementing the circuitry needed by the DSP algorithm are being loaded into the instruction space. These predefined configurations (instructions from the point of view of the Core Processor) individually respond to the initiation signals that depend on the DSP algorithm that being implemented by the processor.

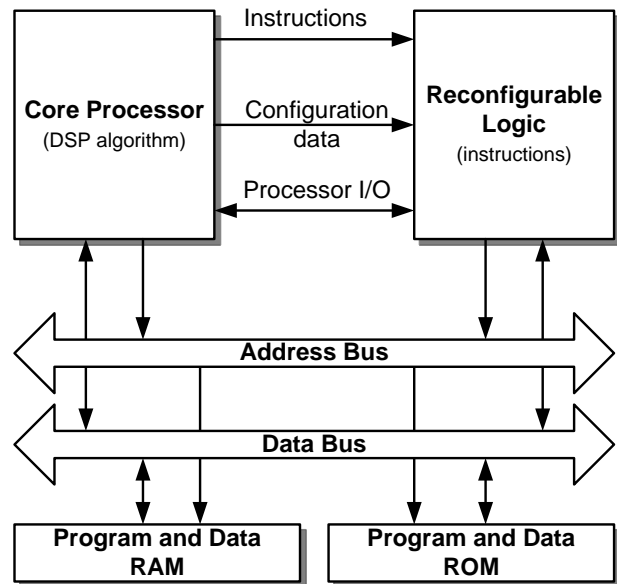


Fig. 1. DISC-processor structure

Taking into account the increasing abilities of modern FPGA for real-time reconfiguration, it is well justified to use DISC-approach in various DSP applications by using specific circuitry libraries that boost the performance of the currently used algorithm. At the same time, the structure of the DISC-processor remains unchanged, that yields a significant flexibility particularly suited for embedded processors.

III. SPECIFICATION

The proposed hardware platform is based on a dual-core processor consisting from TI TMS320C6713 DSP [1] and FPGA Xilinx Spartan-3 XC3S200 [2] realizing peripheral processor functions. DSP provides executing complex algorithms of digital processing of audio

signals.

The peripheral processor meant for hardware platform external interface realization, pre- and post- data processing, supplying with necessary interfaces data formats, external interface packets forming and disassembling as well as digital signal processing algorithms implementation with the object of DSP unload.

The DSP TMS320C6713 is a highest-performance floating-point digital signal processor (DSP) based on Advanced Very Long Instruction Word (VLIW) architecture: Eight 32-Bit Instructions/Cycle; 32/64-Bit Data Word; 200-MHz (PYP) Clock Rates; 3.3-, 4.4-, 5-, 6-Instruction Cycle Times; 1800/1350 MIPS/MFLOPS; rich peripheral set optimized for audio; highly optimized C/C++ compiler; extended temperature devices available.

The major features of DSP are following:

Eight Independent Functional Units: 2 ALUs (Fixed-Point); 4 ALUs (Floating-/Fixed-Point); 2 Multipliers (Floating-/Fixed-Point); Load-Store Architecture With 32 32-Bit;

General-Purpose Registers: Instruction Packing Reduces Code Size; All Instructions Conditional.

Instruction Set Features: Native Instructions for IEEE 754; Single- and Double-Precision; Byte-Addressable (8-, 16-, 32-Bit Data); 8-Bit Overflow Protection; Saturation; Bit-Field Extract, Set, Clear; Bit-Counting; Normalization.

L1/L2 Memory Architecture: 4K-Byte L1P Program Cache (Direct-Mapped); 4K-Byte L1D Data Cache (2-Way); 256K-Byte L2 Memory Total: 64K-Byte; L2 Unified Cache/Mapped RAM, and 192K-Byte Additional L2 Mapped RAM.

Device Configuration: Boot Mode: HPI, 8-, 16-, 32-Bit ROM Boot; Endianness: Little Endian, Big Endian;

32-Bit External Memory Interface (EMIF) [3]: Glueless Interface to SRAM, EPROM, Flash, SBSRAM, and SDRAM; 512 MB Total Addressable External Memory Space.

The Enhanced Direct-Memory-Access (EDMA) Controller [4] (16 Independent Channels), 16-Bit Host-Port Interface (HPI), two McASPs, two Inter-Integrated Circuit Bus (I2C Bus), two Multichannel Buffered Serial Ports, two 32-Bit General-Purpose Timers, Dedicated GPIO Module with 16 pins (External Interrupt Capable), Flexible Phase-Locked-Loop (PLL) Based, Clock Generator Module IEEE-1149.1 (JTAG), Boundary-Scan-Compatible, 0.13- μ m/6-Level Copper Metal Process, CMOS Technology, 3.3 V I/Os, 1.2 V Internal.

The external memory and peripheral processor (FPGA) are connected to the TMS320C6713 through the 32-Bit EMIF. The EMIF has four address spaces CE0-CE3 (see Fig. 2). The SDRAM is mapped on the CE0, FLASH memory – CE1, FPGA – CE2. The SDRAM memory with total capacity of 8 MB is connected to

processor via internal controller that is a part of the EMIF and operates on the maximum clock frequency available for the memory of this type. Flash memory with total capacity of 2 MB intended for storing DSP program and constant coefficients that are required by signal processing algorithms. The processor uses flash-memory in boot-mode. The program is automatically loaded into DSP's internal memory and external SDRAM by means of second-order boot loader after module initialization that is prepared by first-order boot-loader. The program load and its debug are realized with the help of PC that wired to the external DSP JTAG interface (see Fig. 3).

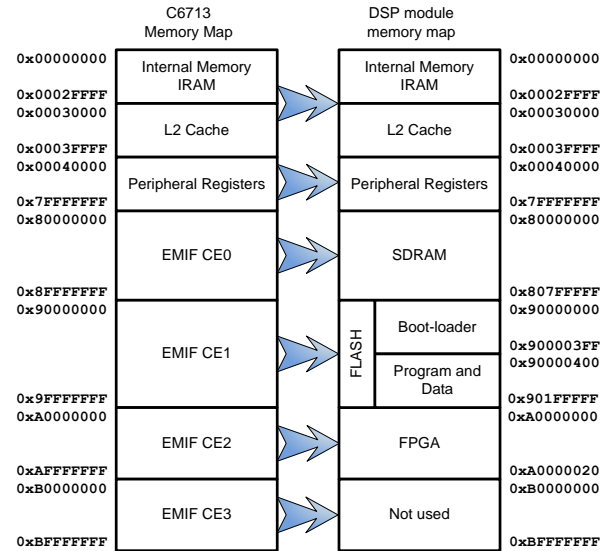


Fig. 2. DSP Module Memory Map

The TLV320AIC23 codec [5] is used for stereo audio signal input and output purposes. Last one is wired to the DSP via two serial channels: McBSP0 (control channel), McBSP1 (receive/transmit channel). The synchronize clock frequency can be adjusted by the Digital Clock Manager (DCM) blocks (FPGA) (see Fig. 3).

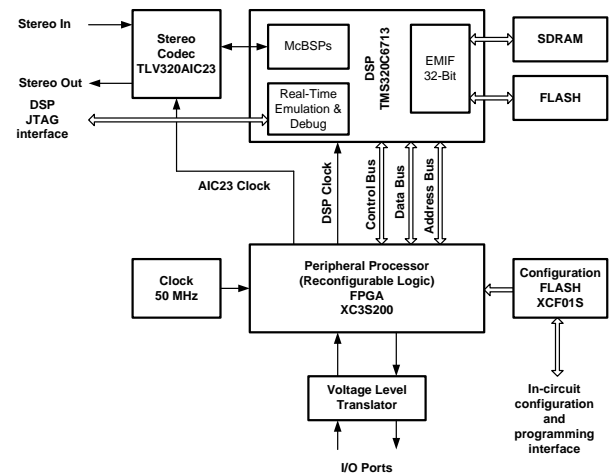


Fig. 3. Module Structure

For system external interface implementation, FPGA's I/O ports are used (for package PQ208 the number of user outputs are 141). The Spartan-3 family FPGAs supports 26 I/O standards 8 of which are differential (LVDS, LVPECL etc.). In the part of I/O ports, voltage level translators provide connection capability with systems with power level more than 3.3 V. Additional coupling capabilities are provided by Schmitt triggers that allows stable operating with long rising and negative-going slopes signals.

The I/O ports can be used as one or several parallel data buses of necessary capacity as well as serial lines. In addition to logic cells, FPGA includes triggers, memory blocks, matrix multipliers and phased-locked loop schemes. Twelve matrix multipliers blocks with 18-bits operands and 36-bits result can be used to implement digital signal processing functions on the peripheral processor.

Realization of peripheral processor project can be done with the help of CAD Xilinx ISE® [6]. In this case, FPGA configuration sequence is generated by means of iMPACT utility and loaded through JTAG interface into the platform's flash-memory. Every time after power up FPGA reads this sequence from flash-memory.

Described architectural decision (see Fig. 4) allows taking following hardware platform features:

- 1) power computational floating-point core allowing up to 1800/1350 MIPS/MFLOPS;
- 2) task paralleling possibility in dual-core system;
- 3) platform reconfiguration via FPGA configuration sequence and DSP-core software change;
- 4) platform embedding capability into various equipment due to flexible setting of external interface by peripheral processor;
- 5) standard configuration and programming means [6],[7].

Supply voltage of the module is +5 V DC, current consumption is less than 300 mA.

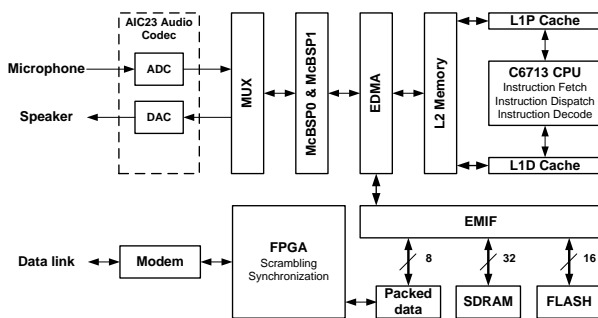


Fig. 4. Block diagram of the proposed hardware platform

IV. APPLICATIONS

A. Speech enhancement system

The complete implementation details of the WDFT based noise reduction system were described in [8]. The

block diagram of the WDFT based system is depicted in Fig. 5. In that implementation, a common psychoacoustically motivated spectral weighting technique is used for noise reduction.

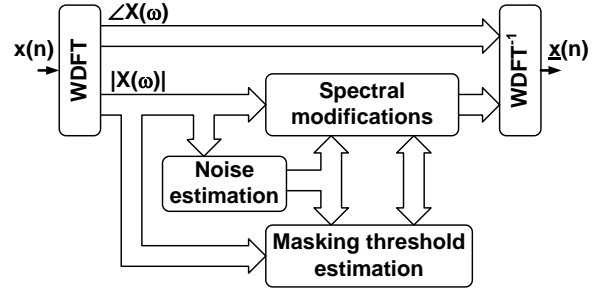


Fig. 5. Block diagram of the WDFT based perceptual noise reduction system

For experiments was used the set of eight speech sentences with strong high frequency components. The sentences were about a 5-8s long. As a degradation signal, the colored noise was added to the clean speech such that the segmental signal to noise ratio (SEGSNR) was between -5 dB and 20 dB. An objective performance evaluation was based on the instrumental and perceptual measurement (see Fig. 6).

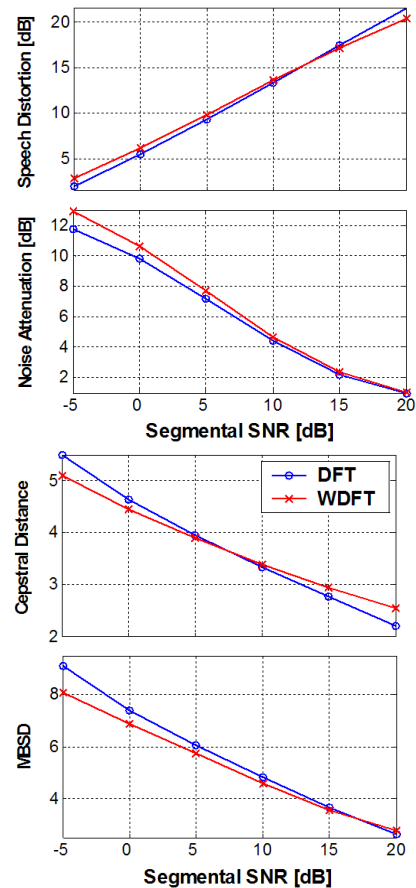


Fig. 6. Performance evaluation of the WDFT speech enhancement system

Speech distortions were measured using segmental signal to noise ratio (SEGSNR) where the noise is interpreted as a difference between original and enhanced speech. The higher value of the segmental SNR indicates the weaker speech distortions.

The cepstral distance was also used as a speech distortion measure. The distance between clean and enhanced speech cepstra was calculated. In this case, the higher cepstral distance reflects the stronger speech distortions. The amount of noise reduction was measured using noise attenuation factor (NA) which is defined as the mean ratio between the input noise power and output noise power. The instrumental measurement is relatively low correlated with the speech perception. Thus perceptual distortion defined as audible difference between clean and enhanced speech was evaluated using Modified Bark Spectral Distortion (MBSD) measure [9].

The experimental results depicted in Fig. 6 show that proposed approach outperforms both conventional DFT and ordinary WDFT solutions. Extended WDFT system has significantly better noise attenuation performance than conventional DFT system and comparable to pure WDFT method.

B. Combined System of Perceptual Wideband Speech Coder with Noise Reduction Preprocessor

The system works in the following way. We used a technique where two buffers (referred to as the PING buffer and the PONG buffer) are used for sampled data transfer by means of EDMA [4] via McBSP port from and to TLV320AIC23 audio codec (See Fig. 4).

Input wideband speech ($f_s=16kHz$) goes to the noise reduction unit where clean speech spectrum estimation and masking threshold evaluation is performed. This information is used to estimate spectral weighting coefficients and further processing of noised speech. Using the masking threshold estimates in each bark (twenty one in all for 8 kHz speech bandwidth), subband masking threshold evaluation is implemented by grouping barks according to subband decomposition scheme accepted in the coder [10]. Evaluation of the subband perceptual entropy (SPE) is used by codebook structure monitoring unit with the object of multiband codebook reconfiguration. When the coder's structure is tuned and speech enhancement is done, speech coding process is realized.

Noise reduction system partitions input signal frames of length $N=320$ samples (20 ms) with 50% overlap and multiplies by Hamming window. To provide embedding capability we have chosen analysis window size of extended WDFT [11],[12] $N=320$ samples and overcomplete basis size $M=512$. The noise reduction unit adds a half-frame algorithmic delay (10 ms) which results in a total algorithmic delay of 30 ms in combined speech coding system. For the perceptual weighting rule

the residual noise level was set to -26 dB. The LSA rule was configured to achieve maximum noise reduction of 20 dB and to work with noise at SNR higher than -10 dB. The factor for decision directed approach was set to $\alpha=0.05$. Encoder operates at variable bit-rates (from 4.3 kbps up to 24.2 kbps) and provides following subjective speech quality estimates: syllabic legibility – 98%, MOS – 4.2, speaker's voice recognizability – 95%. The objective estimates of the reconstructed speech quality are following: SNR=11.59 dB, Noise-to-Mask Ratio (NMR) [13] -6.76dB, BSD=0.05, MBSD=0.02. The overall computational complexity distribution among the parts of the encoder system is depicted in Fig. 7.

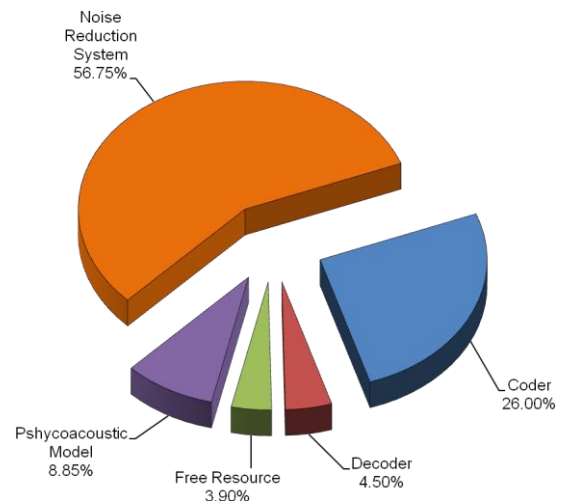


Fig. 7. Computational complexity distribution

The memory requirements for combined speech coding system are represented in Table I.

TABLE I
COMBINED SPEECH CODING SYSTEM MEMORY REQUIREMENTS

Memory Type	Used Memory, KB			Available, KB
	Program Memory	Data Memory	Total	
IRAM	178	10	188	192
SDRAM	67	3156	3223	8192
EEPROM	–	–	1895	2048

In the real-time implementation, such complex algorithms and plentiful hardware devices needs to be managed by task manager. For example, in this application four hardware, two software interrupts and three tasks are used (see Fig. 8) [14]. In common case, number of interrupts and background tasks can be different to meet the requirements of an application.

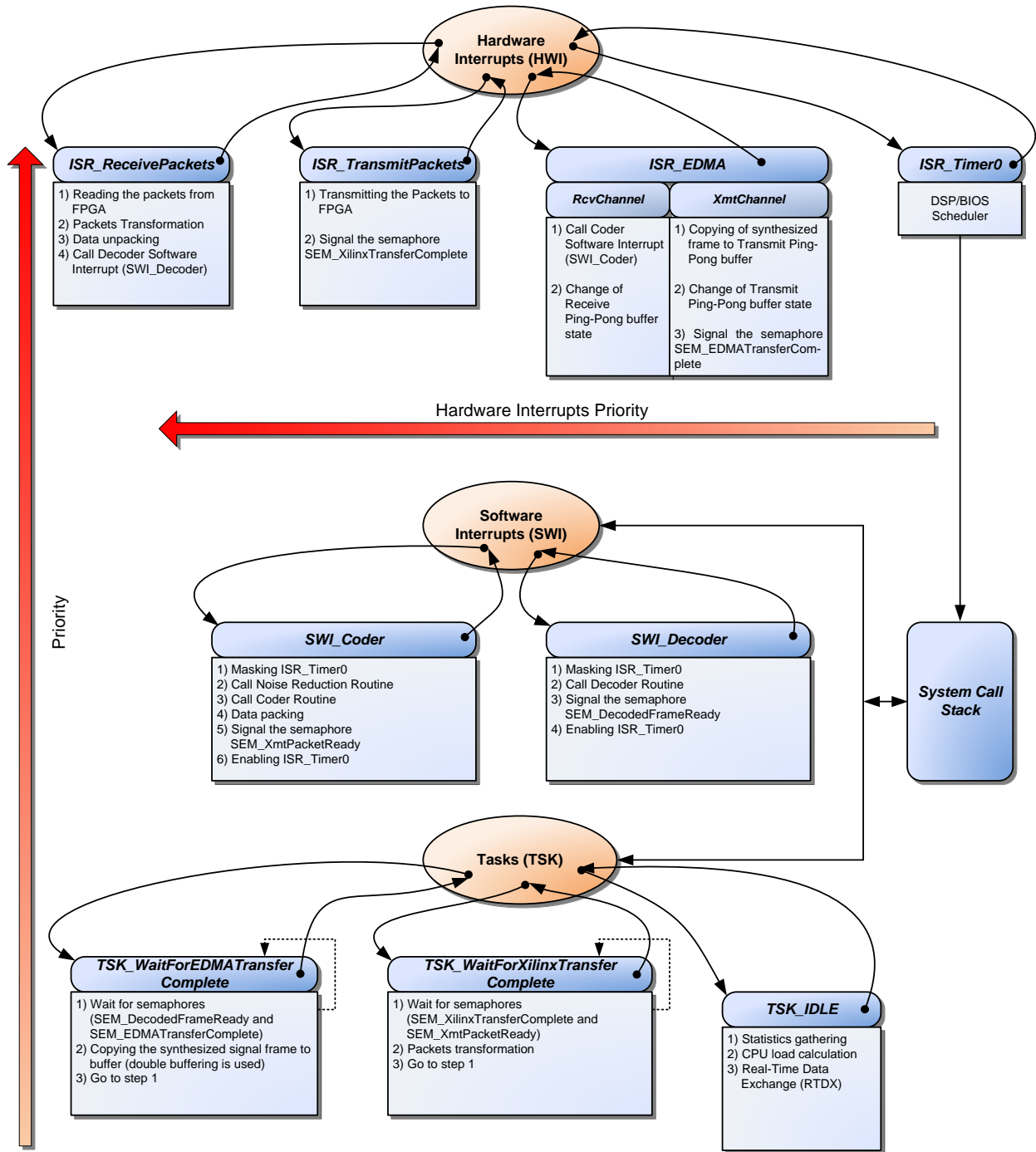


Fig. 8. Combined speech coding system task managing diagram

C. Transform audio encoder based on the WDFT

The encoder and decoder structures for wideband speech and audio signal perceptual coding based on the WDFT are shown in Fig. 9 [15]. The main components of the coder are forward and inverse WDFT, masking threshold estimation block, quantization/dequantization and lossless Huffman coding/decoding modules.

The performance of the wideband perceptual WDFT speech/audio coder (16 kHz sampling frequency), defined as the bit-rate, is not greater than 24 kbps. For

higher quality audio coding (32 kHz sampling frequency), it is 46 kbps, accordingly. Therefore, the coding gain using adaptive Huffman coding for Gaussian modeling is about 1.45 bit per sample for both band signals.

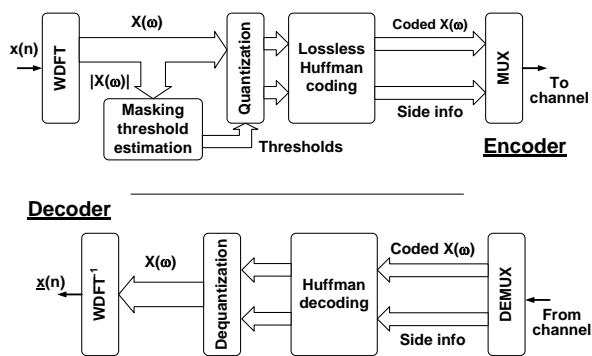


Fig. 9. The block diagram of the WDFT-based perceptual audio coder

D. Future Work

Such a flexible architecture allows designing reconfigurable platforms for different purposes real-time DSP-applications. However, in the future peripheral processor can be implemented with reconfigurable structure of arithmetical and logical units to meet the requirements of the current algorithm. Therefore, overall performance and flexibility of the system can be significantly improved.

REFERENCES

- [1] TMS320C6713B Floating-Point Digital Signal Processor, SPRS294B, Texas Instruments, Dallas, TX, June, 2006. – 152 p.
- [2] Spartan-3 FPGA Family: Complete Data Sheet (DS099) [Electronic resource] / Xilinx Inc., May, 2007. – Mode of access: <http://direct.xilinx.com/bvdocs/publications/ds099.pdf>. – Date of access: 25.05.2007.
- [3] TMS320C6000 DSP External Memory Interface (EMIF) Reference Guide, SPRU266E, Texas Instruments, Dallas, TX, February, 2006. – 146 p.
- [4] TMS320C6000 DSP Enhanced Direct Memory Access (EDMA) Controller Reference Guide, SPRU234B, Texas Instruments, Dallas, TX, March, 2005. – 269 p.
- [5] TLV320AIC23B Stereo Audio Codec, Data Manual, SLWS106H, Texas Instruments, Dallas, TX, February, 2004. – 56 p.
- [6] ISE Quick Start Tutorial, [Electronic resource] / Xilinx Inc., May, 2007. – Mode of access: http://www.xilinx.com/support/sw_manuals/xilinx7/download/qst.pdf. – Date of access: 25.05.2007.
- [7] Composer Studio User's Guide, SPRU328B, Texas Instruments, Dallas, TX, 2000.
- [8] A.Petrovsky, M.Parfieniuk, A. Borowicz, "Warped DFT based perceptual noise reduction system", Proc. AES 116th, Berlin, Germany, 8-11 May 2004, Conv. Paper #6035.
- [9] Yang W., Benbouchta M., Yantorno R. "Performance of a modified bark spectral distortion measure as an objective speech quality measure" // The Proc. of ICASSP, Seattle, USA, 1998, pp. 541–544.
- [10] Michael Livshitz and Alexander Petrovsky, "Perceptually Constrained Variable Bit Rate Wideband Speech Coder" // The Proc. of EUROCON, Serbia & Montenegro, Belgrade, November 22-24, 2005, pp.1296-1299.
- [11] Michael Livshitz and Alexander Petrovsky, "An Overcomplete WDFT-based Perceptually Constrained Variable Bit Rate Wideband Speech Coder with Embedded Noise Reduction System" // The Proc. of XI International Conference "SPEECH AND COMPUTER", St. Petersburg, Russia, June 25-29, 2006, pp.343-348.
- [12] Borowicz A., Petrovsky A., "An Overcomplete WDFT Sinusoidal

Basis for Perceptually Motivated Speech Enhancement" // The Proc. of 13th European Signal Processing Conference (EUSIPCO'2005), (CD-ROM), Antalya, 2005.

- [13] Brandenburg K., Sporer T., "NMR and masking flag: Evaluation of quality using perceptual criteria" // The 11th AES Proceedings, Portland, Oregon, USA, May 1992, pp. 169–179.
- [14] TMS320C6000 DSP/BIOS Application Programming Interface (API) Reference Guide, SPRU403F, Texas Instruments, Dallas, TX, 2003.
- [15] Petrovsky A., Borowicz A., Parfieniuk M., Petrovsky A. "Warped discrete Fourier transform in perceptual speech and audio processing" // X Symposium New trends in Audio and video, Wroclaw, Poland, 16-18 September 2004 (Prace Naukowe Instytutu telekomunikacji i akustyki Politechniki Wroclawskiej, 85, seria: konferencje 29). – pp.143-152.