SOURCE AUDIO CODING IN DIGITAL RADIO MONDIALE

Marko Horvat*[1], Hrvoje Domitrovic[1], Maja Kurjak[1]

[1]Department of Electroacoustics, Faculty of Electrical Engineering and Computing, University of Zagreb

Unska 3, 10 000 Zagreb, Croatia
marko.horvat@fer.hr

Abstract

Digital Radio Mondiale (DRM) is a new digital radio broadcasting system originally intended for use at frequencies below 30 MHz, while efforts are made to extend the use of the system to the frequency range up to 120 MHz. The purpose of developing and introducing this system is to replace all existing analog radio broadcasting systems operating in the fore mentioned frequency range. Furthermore, the system is fully incorporated in the existing frequency plan for easier implementation, thereby using standard AM channels of 9 (10) kHz bandwidth or half-channels of 4.5 (5) kHz width. In some cases, when allowed by the frequency plan, double channels of 18 (20) kHz width can be used. In any case, the usable bandwidth is rather narrow, which represented a big obstacle for realization of DRM for a long time. This was due to the lack of audio coding methods efficient enough to compress the source audio signal to a size acceptable for transmission in the mentioned bandwidth, while maintaining a reasonable sound quality. Finally, the development of MPEG-4 Audio standard opened up the possibility to overcome this problem and offer interference-free listening at near-FM quality, covering great distances with a single transmitter. The methods of source coding for audio signals defined in the MPEG-4 standard and modified for specific use in DRM system will be presented in this paper.

INTRODUCTION

Up till a few years ago, the radio broadcasting systems operating in the frequency range below 30 MHz were all analog, without exception. Their analog nature, however, offered poor quality of transmitted audio signal. Unfortunately, there has not been room for improvement until recently, when the development of computer and digital technology in general made it possible to commence working on a digital

broadcasting system which is to replace all existing analog systems operating below 30 MHz. Besides the lack of appropriate technology having the desirable characteristics, the main obstacle for developing this system was the lack of efficient audio coding algorithms capable of compressing the signal so that it can fit in the standard AM channel. For these reasons, the work on the Digital Radio Mondiale (DRM) system began in the late 90s, to be precise, on March 3, 1998 in Guangzhou, China. For the first time, interference-free listening at a near-FM quality in the frequency range below 30 MHz became a reality. Furthermore, the digital nature of the system offered new types of data services and information. The latest developments[1] include the extension of the DRM system to the frequency range up to 120 MHz with the intention to replace the existing analog FM broadcasting in the near future.

In order to implement the DRM system in the existing frequency plan, each DRM digital channel is required to occupy either half the standard AM channel, having 4,5(5) kHz bandwidth, or one standard AM channel having 9(10) kHz bandwidth. Exceptionally, the DRM channel can expand over two standard AM channels where permitted by the frequency plan. The bit rates available through mentioned channels range from 8 kbps for half-channel, 24 kbps for a standard channel up to 72 kbps for double channels. The bit rates available in half and standard channels represent a serious challenge for an audio coding algorithm, which is required to compress speech or music and prepare it for transmission at the stated bit rates while maintaining reasonable quality. Standardized audio coding algorithms capable of meeting these requirements did not exist until the development of MPEG-4 Audio standard, the first widespread and established standard to offer audio compression of sufficient efficiency.

Three subsets of MPEG-4 Audio standard, modified for specific use, are used for the DRM source audio coding, with the addition of two more coding procedures used to increase coding efficiency. General audio signals are coded with an MPEG-4 version of the Advanced Audio Coding (AAC) algorithm, while pure speech signals use Code eXcited Linear Prediction (CELP) or Harmonic Vector eXcitation Coding (HVXC). Additionally, the Spectral Band Replication (SBR) technique can be used in conjunction with any of the three basic coding methods in order to extend the frequency range of the transmitted signal. Finally, the Parametric Stereo (PS) can be used as an addition to the AAC in order to enable the coding of stereo signals at even lower bit rates without degrading the quality.

1. MPEG-4 ADVANCED AUDIO CODING

As stated before, the MPEG-4 AAC coding method is used for coding of general audio signal. The core of this method is the MPEG-2 AAC, enhanced with new tools indigenous to MPEG-4. When the source coding of audio signal in DRM is concerned, the AAC permits two sampling frequencies, 12 and 24 kHz, thereby operating with narrowband or broadband audio signal, respectively. Although the bit rate is not fixed, but can be chosen in an arbitrary fashion, the length of the

transformation block is fixed to 960 samples. The reason for this is that 960 samples taken at the stated sampling frequencies yield elementary audio frames of 80 or 40 ms length. Since a DRM audio super frame has a length of 400 ms, an integer number of elementary audio frames can be used to form this audio super frame.

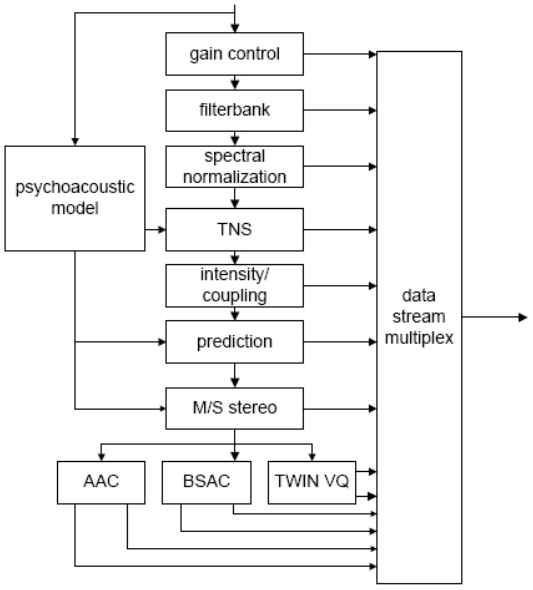

Figure 1 – Block diagram of DRM MPEG-4 AAC general audio encoder

The AAC coding procedure, shown in *Figure* 1, starts with feeding the input signal to the gain control stage and to the psychoacoustic model responsible for yielding the parameters required for different processing stages contained in the spectral processing block. The psychoacoustic model takes advantage of the human auditory system, thus improving the coding efficiency. Upon passing through the gain control stage, the signal undergoes the transformation to the frequency domain by processing in the filterbank stage. After that, the block of spectral processing begins with spectral normalization. In the Temporal Noise Shaping stage, the quantization noise is manipulated in the frequency domain in order to shape it in time domain, following the shape of the signal. In the Intensity/Coupling stage, the stereo signal is coded using the Joint Intensity Stereo method, thus further improving the coding efficiency. The long-term frame-wise prediction tool, indigenous to MPEG-4, predicts the spectral values of the current frame from the values of previous frames. The last tool in the spectral processing block is the Mid/Side Stereo coding tool, where the stereo signal is represented with the sum (mid channel) and the difference (side channel) of the left and the right stereo channel. The quantization and coding stage is improved with new methods which enable fine bit rate scalability. Apart from standard AAC quantization and coding, Bit Sliced Arithmetic Coding (BSAC) and Transform-domain Weighted INterleaved Vector Quantization (TWIN VQ) are also applied in the quantization and coding stage. Finally, all data is multiplexed into a single

MPEG-4 audio data stream. The Perceptual Noise Substitution (PNS) tool is not used in DRM MPEG-4 AAC coding because the Spectral Band Replication (SBR) technique, to be explained in section 4, does the same job more effectively.

2. MPEG-4 CODE EXCITED LINEAR PREDICTION

The Code Excited Linear Prediction (CELP) coding method is used as one possible way of coding speech signals in the DRM system. The method allows for two fixed sampling frequencies; the lower one is 8 kHz, in which case the usable frequency range is set to 100 - 3800 Hz and the coding method is referred to as narrowband CELP, and the higher one is 16 kHz, in which case the frequency range is 50 -7000 Hz and the method has been named broadband CELP. The DRM system allows for a number of specific bit rates ranging from 3 850 – 12 200 bps for narrowband CELP and from 10 900 – 23 800 bps for broadband CELP. The lengths of elementary audio frames are fixed to 40, 20 and 10 ms, in a fashion similar to AAC and for the reasons already stated.

This coding method is a good choice for implementation of multiple speech transmissions, e.g. if three speech signals are coded at 8 kbps, then it is possible to incorporate all three coded streams into a single channel operating at 24 kbps and thus transmit three different information simultaneously. Furthermore, the use of CELP is desirable for adding a speech service to an audio service, as well as for robust and reliable speech transmissions through error prone channels, where and when necessary. Finally, CELP can be exploited when simulcast transmissions are required, meaning that half the channel bandwidth is occupied by the DRM signal and the other half by the remaining AM (actually SSB or VSB) signal.

The block diagram of a CELP encoder is shown in *Figure* 2.

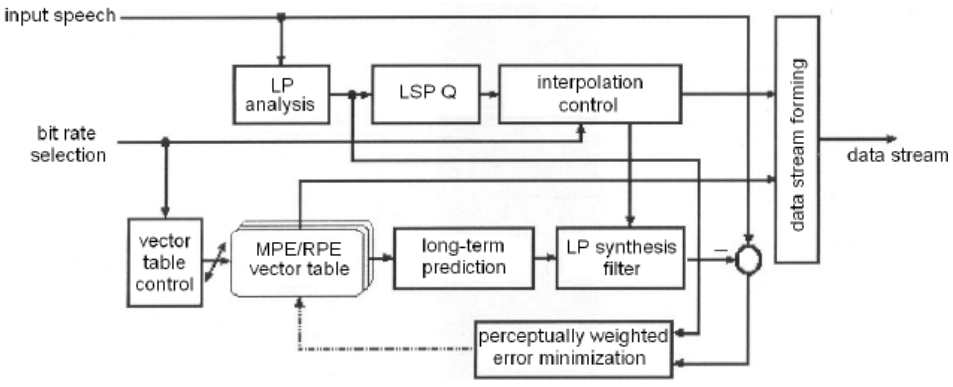

Figure 2 – Block diagram of MPEG-4 CELP encoder

The CELP coding method uses the well-known method of analysis-by-synthesis. The input speech undergoes the linear prediction analysis in order to extract the parameters of the speaker’s vocal tract. These coefficients are then quantized and

interpolated if necessary, depending on the chosen bit rate, and transmitted as a part of the information required for the reconstruction of the speech signal at the receiving side. On the other hand, these coefficients are fed to the synthesis block, where the same synthesis process is carried out as it would be in the receiver. The purpose of this decoder-in-the-coder scheme is to find a particular excitation vector from the vector table which will, upon passing through the synthesis block, yield the synthesized signal as similar to the original speech signal as possible. The search itself is guided by the error signal which is perceptually weighted in order to ascertain whether (and how much) the synthesized signal is audibly different than the original signal. The vector table itself is also controlled by the chosen bit rate. Once the right vector is found, its index in the table is transmitted to the receiving side as the other part of the useful information. The receiver itself contains the same vector table as the one found at the transmitting side.

3. MPEG-4 HARMONIC VECTOR EXCITATION CODING

The Harmonic Vector eXcitation Coder (HVXC), is a parametric encoder intended for speech coding at very low bit rates, namely 2 and 4 kbps, which makes it the ideal choice for application in the DRM system. However, this coding method is able to process only narrowband speech sampled at 8 kHz, which gives the usable frequency range between 100 and 3800 Hz. Given the mentioned bit rates, it is now possible to form a multilingual or a multi-speech service capable of providing several different information or single information on different languages at the same time in the same DRM channel. Its parametric nature also enables independent changing of the parameters themselves, e.g. the pitch and tempo. Since the information encoded with this codec requires very little space, the HVXC is also the perfect choice for building speech databases. For the reasons already stated, the length of an elementary audio frame in DRM HVXC coding is fixed to 20 ms.

The block diagram of the HVXC encoder is shown in *Figure* 3.

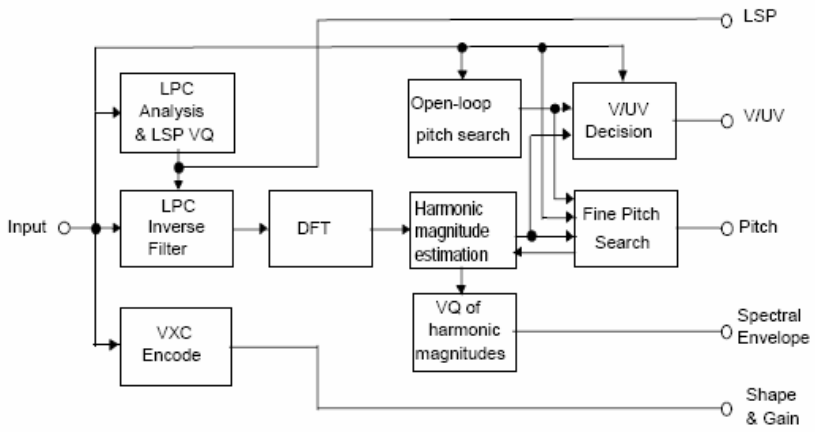

Figure 3 – Block diagram of MPEG-4 HVXC encoder[2]

The HVXC coding treats voiced and unvoiced speech differently. Voiced speech segments are harmonically coded, while unvoiced segments use a vector excitation coding method similar to CELP. The input speech signal undergoes the LPC analysis in order to extract the parameters of the speaker's vocal tract. These parameters are vector quantized and transmitted as a part of the useful information. At the same time, they are passed on to the inverse filter which represents the beginning of the coding procedure for voiced speech segments and removes all vocal tract information from the input signal, leaving only the excitation information. Since the voiced excitation is based on vibrations of the vocal cords, it is expected to consist of the fundamental frequency and a large number of harmonics. Therefore, after the transformation to the frequency domain, the harmonic magnitudes of this excitation are estimated, quantized and transmitted to the receiving end as the spectral envelope information. The other relevant parameter for a voiced speech segment is the pitch itself. The pitch search occurs in two steps, first a coarse search is conducted and then fine, if necessary. The V/UV decision is made based on the results of the pitch search and the harmonic magnitude estimation. If these steps do not produce satisfactory results, then the segment is declared to be unvoiced. As stated before, the unvoiced segments are coded using the coding scheme very similar to CELP.

4. SPECTRAL BAND REPLICATION

Since the DRM radio broadcasting system uses channels with very narrow bandwidth, the available bit rates are rather low. Therefore, further improvement can be achieved using the Spectral Band Replication (SBR) coding technique. The SBR offers a way of maintaining full bandwidth of the signal without increasing the bit rate by restoring the high frequency spectrum not coded by the source audio or speech encoder.

SBR processing is performed on speech or audio signal of full bandwidth prior to source encoding, after which the source encoder deals with the signal with truncated bandwidth at a lower sampling frequency. The main idea of SBR is to represent the high portion of the frequency band of the signal using the lower portion of the frequency band and some side information, taking advantage of the fact that the high frequency band of speech and audio signals contains either periodic components, i.e. harmonics related to fundamentals of musical instruments and human voice when voiced speech is concerned, or noise-like components such as unvoiced speech or the sounds of percussion instruments. The loss of the high band due to truncation has a great influence on the timbre of musical instruments and on the intelligibility of a speech signal. Therefore, the task of the SBR system is to code side information which contains the data on the high band, namely the shape of the spectral envelope, i.e. the coarse spectral distribution, and the general shaping of noise-like signal which are to guide the SBR decoder on the receiving side. This side information occupies only a small fraction of the total data stream.

On the receiving side, the SBR decoding takes place after the source decoding. An SBR enhanced DRM receiver will reproduce full bandwidth audio. On the other hand, a receiver with no SBR enhancement reproduces only the truncated audio

bandwidth signal. The side information on the spectral envelope, along with the inverse filtering based on harmonic information is required to restore the high band, and the information on the shaping of noise-like signals is fed to the noise generator in order to obtain correct spectral shaping of noise-like components. If necessary, sinusoidal components are added in order to obtain a more natural sound.

The SBR coding principle is shown in *Figure* 4.

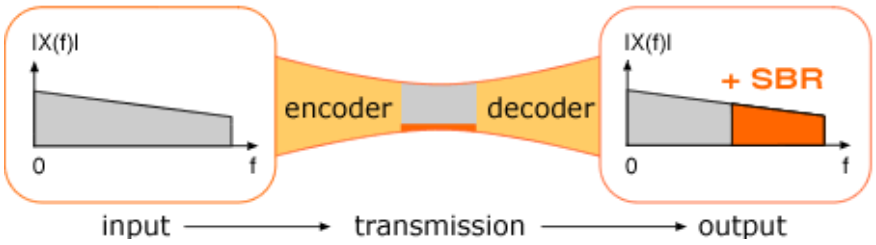

Figure 4 – The SBR coding principle[3]

The SBR coding technique was originally developed for use with AAC, forming a part of the MPEG-4 Audio standard known as AACplus. However, the DRM system also uses CELP and HVXC coding methods enhanced with SBR.

5. PARAMETRIC STEREO

Parametric Stereo (PS) coding further enhances the AAC + SBR (AACplus) stereo coding and the entire coding scheme is referred to as 'enhanced AACplus' or 'AACplus v2'. The basic principle of PS is to downmix the stereo signal into a single mono signal, while representing the stereo image with a set of parameters. The mono downmix then undergoes the SBR procedure, after which it is coded with a conventional AAC coder. The parameters that describe the stereo image are transmitted as side information occupying a small fraction of the total bit rate. This coding scheme provides higher quality of coded stereo signal than the conventional AAC plus. The listening test results [4] show that AAC plus v2 at 24 kbps has similar or even higher quality than the conventional AAC plus at 32 kbps.

The PS coding principle is shown in *Figure* 5.

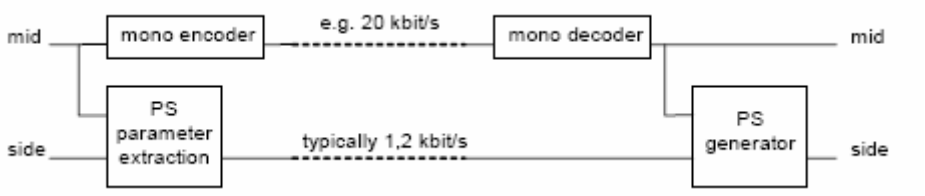

Figure 5 – The PS coding principle[2]

The PS used in DRM defines two stereo image parameters: the PANorama (PAN), representing the level differences between the left and the right stereo channel, and the Stereo Ambience (SA), containing information that cannot be described with just level differences (interchannel phase difference, interchannel coherence)[4]. Both parameters are changing over frequency and time and are dealt with individually. However, there is a certain degree of correlation between the two parameters which justifies their existence as a whole. For example, when the values of the PAN parameter are high, the SA tends to have a minor significance.

As complex stereo ambience factors, such as reverberation, tend to be more emphasized in the side channel than in the mid channel, much of this ambience information is lost in the process of downmixing to mono. The SA enables this complex ambience information to be reconstructed in a proper manner.

At the receiving side, the stereo image is recreated from the mono signal and the transmitted side information, but without affecting the total spectral energy. Therefore, no coloration is introduced into the final reconstructed stereo signal, compared to the source mono signal. Furthermore, PS displays the backward compatibility property, meaning that the decoder not supporting the PS feature will not be able to reconstruct the source stereo signal, but will decode the downmixed mono signal successfully.

The PS coding can easily be implemented into any system using the SBR extension, as the same filterbank is used in both procedures. Since both procedures benefit from the same time/frequency transformation, the increase in computational complexity when implementing the PS is very small.

6. CONCLUSION

The extreme demands the DRM source audio coding is faced with require the use of highly efficient coding techniques capable of representing the source audio signal with as small amount of information as possible, while providing acceptable quality. However, the coding techniques chosen for DRM audio coding have even surpassed these requirements, especially AAC coding enhanced with both SBR and PS, representing the most advanced, state-of-the-art stereo audio codec available today. Nowadays, great effort is put into extending this ‘AACplus v2’ onto multichannel audio coding. Furthermore, the development of DRMplus system intended for broadcasting at frequencies up to 120 MHz will not call for coding algorithms of such efficiency, but will be focused on transmitting high quality audio.

REFERENCES

[1| Lehpamer, H., “The Latest Developments in Digital AM Radio” (2005).

[2] “Digital Radio Mondiale (DRM); System Specification”, ETSI ES 201 980 V2.1.1 (2004-06).

[3] http://www.codingtechnologies.com/products/sbr.htm (Accessed March 24th, 2006).

[4] Breebaart, J. et al, “Parametric Coding of Stereo Audio”, EURASIP J. on Applied Signal Processing, 1305-1322 (2005:9).