Joint Coding Rate Control for Audio Streaming in Short Range Wireless Networks

Jelena Kovacevic, Dragan Samardzija, *Member, IEEE* and Miodrag Temerinac, *Member, IEEE*

Abstract— *Streaming high quality audio/video (AV) from home media sources to TV sets and audio speakers over a wireless local area network (WLAN) is a challenging problem because of the fluctuating bandwidth caused by interference and fading. Retransmission and data buffering, as common techniques in data transfer over error-prone channels, are not adequate for delay sensitive applications such as high quality audio streaming. Our approach is to adjust the audio data rate dynamically in order to improve the perceptual quality of audio according to wireless channel quality. In this paper we specifically focus on robust audio streaming over short range wireless channels. The major strength of the proposed algorithm is in its ability to match the source and channel coding parameters to the given channel conditions, thus providing better quality and lower latency of audio in home networking*[†].

Index Terms — audio, wireless, streaming, joint source and channel coding.

I. Introduction

Wireless networks are becoming a key enabler to a rapidly growing number of consumer applications. Short range wireless communication systems are of a particular interest. For example audio and video streaming, gaming, as well as integrated multimedia entertainment centers rely heavily on wireless communications. In recent years a number of short range wireless communication standards have emerged targeting consumer applications (802.11a, b, g and n, Ultra-Wideband (UWB) …).

Unlike the data applications, audio and video streaming are intolerant of bandwidth fluctuations due to the delay constraints. Guaranteed bandwidth and Quality-of-Service (QoS) are essential requirements in order to satisfy customer expectations for “wire-like” performance. Assuring high bandwidth is essential but not sufficient. When several applications try to access the same bandwidth, the ones that are intolerant to time delays and bandwidth fluctuations will not function properly. MAC layer and cross-layer optimization is well investigated [1].

A number of wireless technologies and their corresponding standards have introduced strong QoS mechanisms. For example, 802.11e defines the QoS mechanism for wireless local area networks [2], while similar solutions are applied in cellular (e.g., Universal Mobile Telecommunications System (UMTS) / High Speed Packet Access (HSPA) and Mobile WiMAX) [3, 4] and fixed wireless networks (Fixed WiMAX) [5].

However, methods for providing good quality of audio in such environments are still an open issue. A number of proprietary solutions are being developed. For example, the Wireless Home Digital Interface (WHDI) [6] and WirelessHD [7] technologies transmit high quality uncompressed audio and video signals. Both technologies pose significant spectrum requirements. The WHDI and WirelessHD technologies use the 5 GHz and 60 GHz band, respectively.

We propose an optimized spectrally efficient algorithm in the source-channel coding domain. Since the early days of digital communications, channel coding and modulation have been separated from source coding. According to the information theoretic results optimal source and channel coding can be performed separately for the stationary additive white Gaussian noise channel (AWGN), when no latency constraint is imposed. The result is known as *Separation Theorem* [8].

Audio and video streaming (including broadcast and unicast services) over time-varying wireless channels and finite block data lengths (implicitly latency-constrained) cannot be optimally transmitted when source and channel coding are separated.

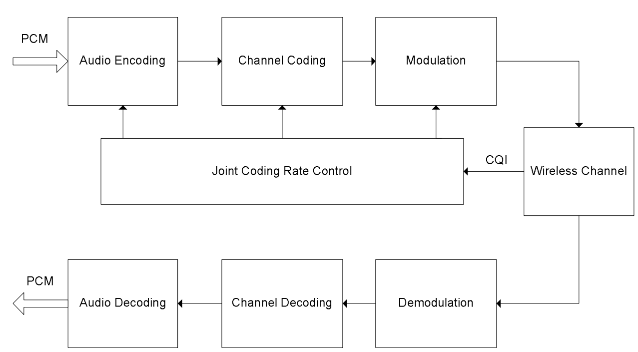

Fig. 1: Functional block diagram of the wireless communication system for audio streaming.

To address the above problem in this study we consider a practical communication system for audio streaming that is depicted in Fig. 1. The transmitter consists of an audio encoder, channel encoder, modulator and joint coding rate

† J. Kovacevic and M. Temerinac are with the Department of Computer Engineering and Communications, Faculty of Technical Sciences, University of Novi Sad, Serbia (e-mail: jelena.kovacevic@rt-sp.com, miodrag.temerinac@rt-rk.com)

D. Samardzija is with Bell Laboratories, Alcatel-Lucent, USA (e-mail: dragan@alcatel-lucent.com)

control (JCRC). Parameters of source and channel coding are set dynamically depending on the channel quality (quantified by the channel quality indicator (CQI)) and the size of the data block that is to be transmitted.

In Section II we present the system that is considered in this study. We present audio coding, channel coding, modulation and joint coding rate control. Furthermore, in Section III we present the system performance methodology, considering the audio performance and physical layer abstraction. The simulation results are presented in Section IV, and we conclude in Section V.

II. System Outline

A. Audio encoding and decoding

Audio encoder implemented in this study is based on perceptual coding principles [9]. It encodes two audio channels into a bit-stream ranging between 64 and 384 kbps. The sampling rates supported are 32, 44.1 and 48 kHz. The frame size corresponds to 1536 samples, 16 bits per sample. At 48 kHz a frame covers a time interval of 32 ms. A block diagram of the encoding process is shown in Fig. 2.

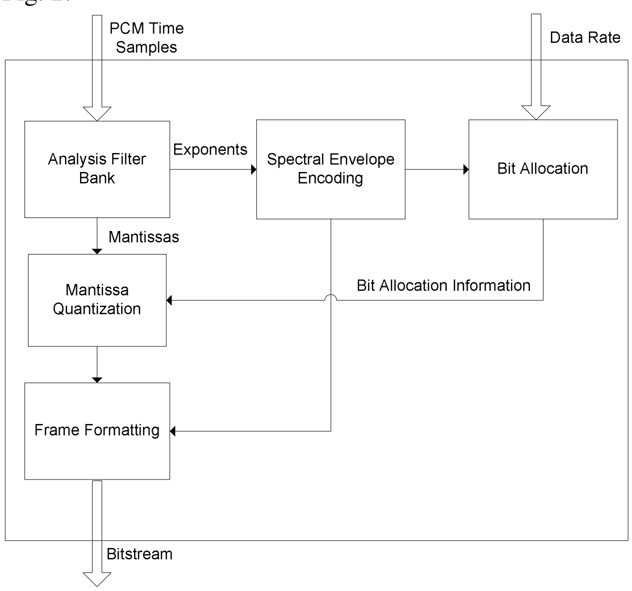

Fig. 2: Audio encoder block diagram.

The encoder implementation allows per-frame data rate adaptation. The three-bit bit-stream variable frame size conveys the encoded data rate to the decoder.

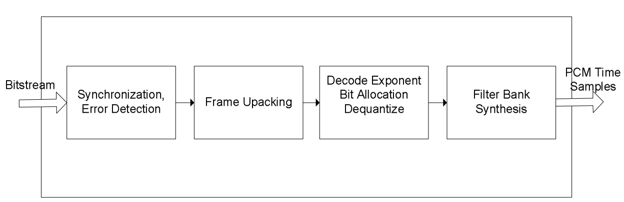

Fig. 3: Audio decoder block diagram.

Decoding is a reverse operation of the encoding process. First, the input bit-stream is demultiplexed and errors are detected. The concealment process repeats previously received blocks of data, or, for worse error conditions it causes muting for the duration of one or more frames.

The perceptual quality of decoded samples depends on data rate. Average audio signal-to-noise ratio (SNR) and peak SNR (PSNR) measures for music test streams for all rates supported by the encoder are given in Table I.

Table I
Average audio PSNR and SNR for various data rates

Data rate [kbps]	PSNR [dB]	SNR [dB]
64 kbps	40.412451	15.106117
96 kbps	47.999052	22.692717
128 kbps	56.448934	31.142599
192 kbps	65.751588	40.445253
256 kbps	68.10504	42.798705
384 kbps	68.299529	42.993194

B. Channel encoding and decoding

In this study we implement a convolutional encoder according to [10]. It is a rate-1/2 channel encoding, with the constraint length 9 and the generator polynomials 753 and 561 in octal representation. The convolutional encoder is followed by a repetition and/or puncturing enabling different coding rates. The overall channel encoding is depicted in Fig. 4.

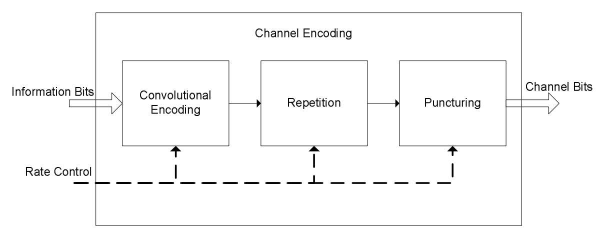

Fig. 4: Channel coding block scheme.

On the receiver side, the soft Viterbi decoder is implemented [11, 12]. The inputs to the decoder are 3-bit-wide log-likelihood ratios (LLRs), i.e., soft-bits. The soft bits are generated by the demodulator and processed by the de-repeater and/or de-puncturing.

C. Modulation and demodulation

In this study we implement an adaptive modulator which supports QPSK, 16-QAM, 64-QAM and 256-QAM. These modulation formats are also defined by the 802.11a/g specifications (except 256-QAM). The demodulator generates soft bits

$$LLR = \log \frac{p(b=1)}{p(b=0)} \tag{1}$$

where $p(b = 1)$ and $p(b = 0)$ is the probability that the received channel bit is 1 and 0, respectively.

A particular combination of the modulation format and channel coding rate is defined as modulation and coding scheme (MCS). The implemented MCSs and the corresponding data rates are given in Table II.

TABLE II
MODULATION, CODING RATE AND TRANSMISSION DATA RATE

MCS	Modulation	Coding Rate	Data Rate [kbps]
1	QPSK	½	64
2	QPSK	¾	96
3	16-QAM	½	128
4	64-QAM	½	192
5	256-QAM	½	256
6	256-QAM	¾	384

D. Joint coding rate control

To achieve good perceptual quality, audio encoders rely on higher data rates (up to 384 kbps for the 2.0 audio format implemented in this study). This approach performs well in a case of wired channel or good wireless channel conditions (e.g., a short distance to the access point and low levels of interference). Unlike the above case, poor channel conditions lead to lower throughput, higher frame error and drop rate and consequently to lower perceptual quality.

In this paper, we propose a JCRC algorithm that optimizes the transmission data rate according to the channel quality, thus keeping the perceptual quality as high as the channel allows. The optimization is done in the source-channel domain, providing the best match of the source data rate, channel coding and modulation for the given channel quality. Basic idea of the proposed algorithm is to increase the quantization noise of the source when channel bandwidth decreases, which will result in lower source data rates. Quantization noise could be seen as a smart error introduced by the source which will decrease perceptual quality of audio, but on the other hand unpredictable errors of the channel that lead to frame losses and audible artifacts of the reproduced signal, are becoming less probable.

The basic block diagram of proposed algorithm is given in Fig. 5. The JCRC algorithm is performed dynamically, on every audio frame. In this study audio frame is 32 ms long (at 48 kHz) which results in 31.25 rate adaptations per second. Before the audio encoding, CQI is required. Highest MCS which meets the bit error rate (BER) threshold for the current CQI is chosen. The BER values are pre-calculated.

Quantization level of audio data (data rate) is matched to the selected MCS to allow maximum throughput and minimum delay. This approach gives the constant transmission delay, as shown in Table III.

TABLE III
TRANSMISSION DELAY [SEC] FOR DIFFERENT MODULATION SCHEMAS AND AUDIO DATA RATES FOR CHANNEL BANDWIDTH W = 64KHZ

MCS	64 kbps	96 kbps	128 kbps	192 kbps	256 kbps	384 kbps
1	**0.032**	0.048	0.064	0.096	0.128	0.192
2	0.021	**0.032**	0.043	0.064	0.085	0.128
3	0.016	0.024	**0.032**	0.048	0.064	0.096
4	0.011	0.016	0.021	**0.032**	0.043	0.064
5	0.008	0.012	0.016	0.024	**0.032**	0.048
6	0.005	0.008	0.011	0.016	0.021	**0.032**

If channel conditions are favorable, the JCRC scheme may select the maximum source rate of 384 kbps, as opposed to a predefined bit rate which is typically done by a conventional solution.

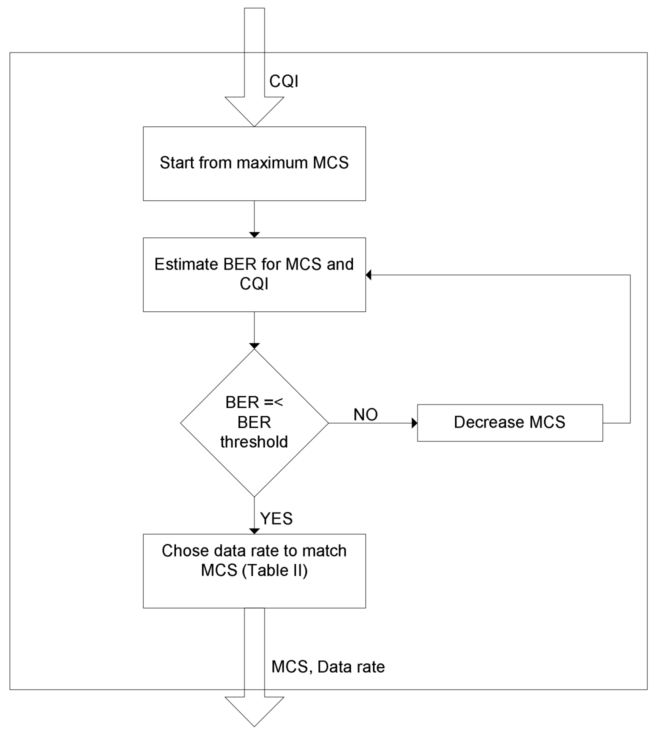

Fig. 5: Joint coding rate control block diagram.

E. Baseline solution

As a baseline solution that is compared against the above JCRC scheme we consider the equivalent communication system that employs a conventional automatic repeat-request (ARQ) mechanism. Namely, as a data frame is received, demodulated and decoded, and its consistency is checked using cyclic redundancy check (CRC). If the check fails, it is an indication that the errors are present in the received frame, and the receiver is issuing a request for the frame retransmission. The identical mechanism is used in 802.11a/g.

The throughput that is achieved using this scheme is inversely proportional to the average number of transmissions before the correct frame is received. The average number of transmission is

$$N = \sum_{k=1}^{\infty} k \; FER(SNR)^{k-1}(1 - FER(SNR)) \qquad (2)$$

where *FER* is the frame error rate as a function of the channel SNR. Therefore, the throughput is

$$T = \frac{R}{N} \qquad (3)$$

where *R* is the rate corresponding to a particular MCS (the fourth column in Table II).

In Section IV the above baseline solution is denoted as the baseline ARQ scheme. Its audio, throughput and latency performance will be investigated and compared against the JCRC scheme.

III. Methodology

A. *Audio quality*

Audio quality measurements are used to assess the performance of the above schemes. The following audio quality measures are used.

- PSNR and SNR are defined as

$$P = \frac{\sum_{n=0}^{N}(x(n) - r(n))^2}{N} \qquad (4)$$

$$PSNR = 10\log\left(\frac{2^{b-1}}{P}\right) \text{ [dB]} \qquad (5)$$

$$SNR = 10\log\left(\frac{\sum_{n=1}^{N} r(n)^2}{P}\right) \text{ [dB]} \qquad (6)$$

where *r* is referent signal, *x* is signal under test, *N* is frame size in samples and *b* is number of bits per sample.

In the case of perceptual audio coders, the above measures could not be used as an objective measure of audio quality due to the well know ‘13 dB miracle’, but they can give a quantitative measure for the proposed solution against the conventional one [13, 14].

B. *Physical layer modeling*

In order to optimize the transmission, the physical layer is abstracted as a ‘black box’. The methodology is depicted in Fig. 6. It is reporting the channel conditions in terms of the BER for the requested transmission data rate and data block size. Based on the reports the JCRC algorithm will determine the transmission data rate as well as the bock size. The selected transmission rate will directly correspond to a particular MCS.

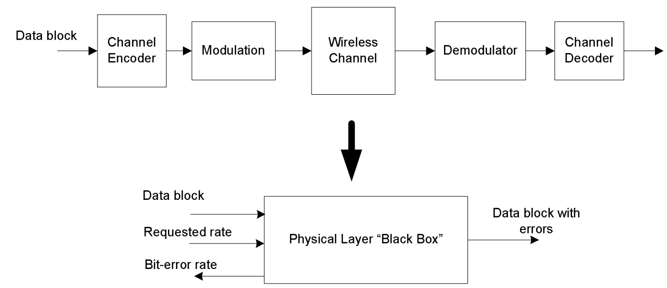

Fig. 6: Physical layer abstraction.

A block-fading wireless channel model is used. Specifically, the channel SNR is randomly generated, and it is kept constant for the duration of the frame transmission. The SNR is selected from a range spanning 0 to 36 dB. In realistic short-range wireless deployments, 0 to 5 dB corresponds to very unfavorable channel conditions (either large distances between the transmitter and receiver, or high levels of interference). On the other end of the SNR range, 30 to 36 dB would correspond to extremely good channel conditions. Both ends of the range are typically less probable while the most of realistic channel conditions will exists between 5 and 30 dB [15, 16].

IV. Experimental Results

To evaluate the effectiveness of the proposed algorithm we have performed extensive simulations with varying channel conditions.

A. *Audio SNR vs. time*

The audio SNR was calculated for different music test streams and compared to the baseline ARQ solution with 265 kbps and 192 kbps. The channel SNR is randomly generated between 5 and 30 dB.

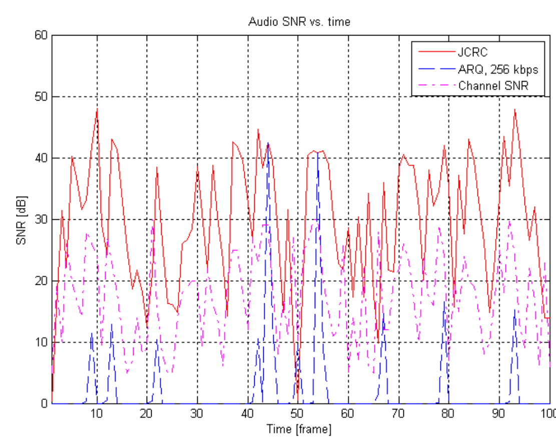

Fig. 7: Average audio SNR vs. time, baseline data rate 256 kbps.

The results are depicted in Fig. 7 and 8 and numerically shown in Table IV. The overall gain in the audio SNR of the proposed JCRC solution is approximately 18 and 27 dB versus the 192 kbps and 256 fixed data rate transmission with ARQ, respectively.

TABLE IV
AUDIO PSNR AND SNR VS. TIME

Solution	PSNR [dB]	SNR[dB]
JCRC	54.679	29.393
Baseline ARQ, 256 kbps	27.388	2.102
Baseline AQR, 192 kbps	35.871	10.585

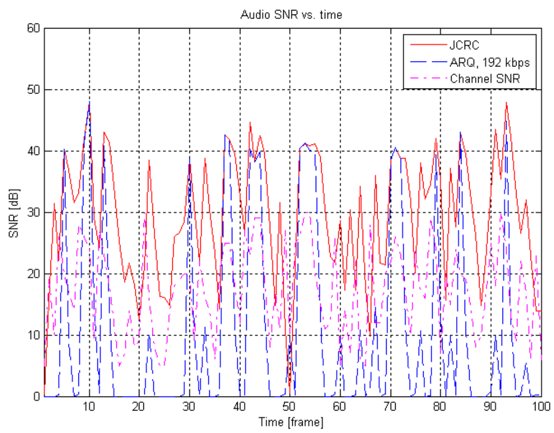

Fig. 8: Average audio SNR per time, baseline data rate 192 kbps.

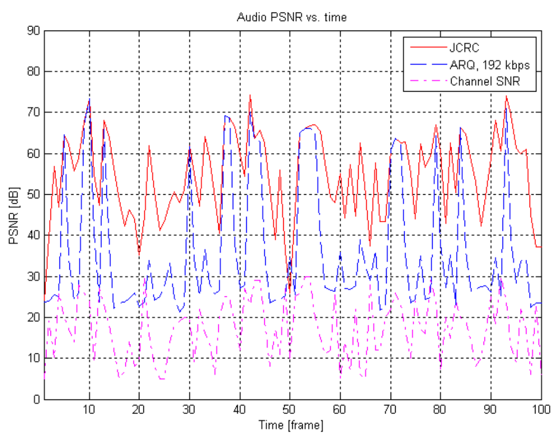

Fig. 9: Average audio PSNR per time, baseline data rate 192 kbps.

B. Audio SNR vs. channel SNR

We evaluate the average audio SNR for every channel SNR, ranging from 0 to 36 dB, as shown in Fig. 10.

It could be seen that the audio SNR for the JCRC scheme is gradually increasing, due to the adaptation to the channel conditions. For example, when the channel SNR reaches 15 dB, the proposed solution will achieve 128 kbps using 16-QAM, while the baseline ARQ solution will not be able to transmit any data.

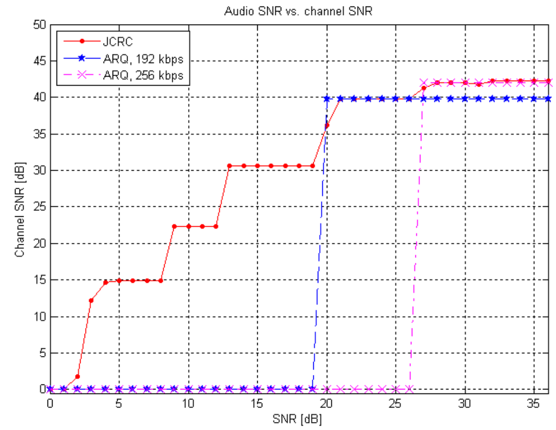

Fig. 10: Average audio SNR vs. channel SNR.

C. Audio data throughput vs. channel SNR

In this measurement we investigate the overall data rates, i.e., throughput as a function of the channel SNR. As it could be seen in Fig. 11, even under bad channel conditions (lower than 5 dB) the proposed JCRC solution is providing data transfer, while the baseline ARQ solution is failing to deliver any data. When the channel condition is extremely good (greater than 32 dB) the proposed solution will transmit at the maximum audio data rate thus providing the best perceptual audio quality.

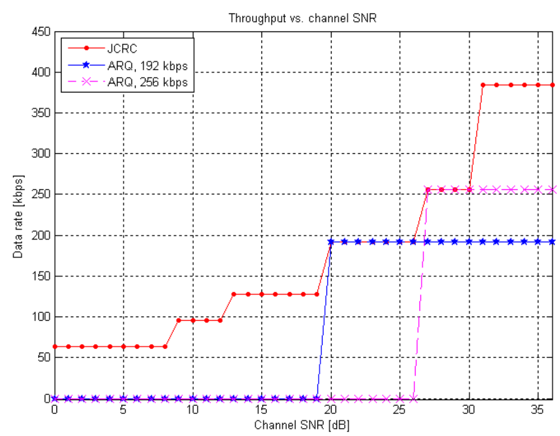

Fig. 11: Audio throughput vs. channel SNR.

Figures 10 to 11 clearly demonstrate the adaptive behavior of the JCRC scheme, and its ability to provide the best achievable quality of the audio transmission for the given channel conditions.

D. Transmission delay vs. channel SNR

In Fig. 12 we present the transmission delay versus the channel SNR. The delay corresponds to the time needed to deliver audio data frame to the receiver. It does not consider any processing delay, so it should be considered as a lower bound on the transmission latency.

Fig. 12 demonstrates that the JCRC scheme maintains constant latency over the whole channel SNR range, while the baseline ARQ solution fails at the lower values of SNR, resulting in infinite delay.

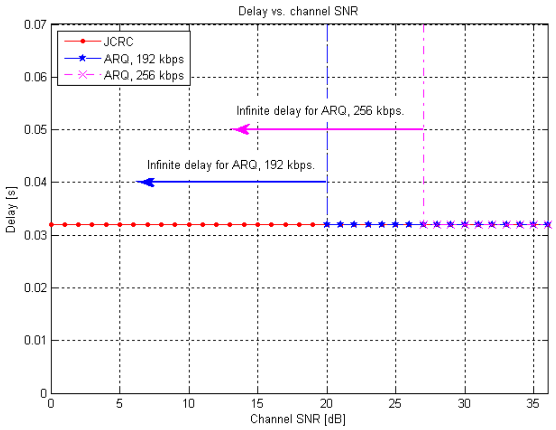

Fig. 12: Transmission delay vs. channel SNR.

E. Frame drop rate

To measure the drop frame rate, the channel SNR is randomly generated between 5 and 30 dB. The results are given in Table V. We note significantly better performance of the JCRC solution.

TABLE V
FRAME DROP RATE VS. TIME

Method	JCRC	ARQ, 192 kbps	ARQ, 256 kbps
Frame Drop Rate	1%	58%	88%

V. CONCLUSIONS

In this study we have proposed the joint source and channel coding algorithm that optimizes the transmission data rate according to the channel quality, thus keeping the perceptual quality as high as the wireless channel allows. The major strength of the proposed algorithm is in its ability to adaptively match the source and channel coding parameters to given channel conditions, thus providing better quality and lower latency of audio.

We have showed that the proposed JCRC scheme significantly outperforms the conventional baseline solution that is based on the ARQ mechanism. It provides significantly better audio quality and a constant low latency.

REFERENCES

[1] S. Khan, Y. Peng, E. Steinbach, M. Sgroi, W. Kellerer, "Application-driven cross-layer optimization for video streaming over wireless networks, " IEEE Communications Magazine, Vol. 44, pp. 122-130, January 2006.

[2] 802.11-2007, Wireless LAN Medium Access Control (MAC) and Physical Layer (PHY) Specifications, June 12 2007.

[3] 3GPP TS 23.107, Quality of Service (QoS) Concept and Architecture, Release 7, Jun 2007.

[4] IEEE. Standard 802.16e-2005. Part16: Air Interface for Fixed and Mobile Broadband Wireless Access Systems—Amendment for Physical and Medium Access Control Layers for Combined Fixed and Mobile Operation in Licensed Band, December 2005.

[5] IEEE. Standard 802.16-2004. Part16: Air Interface for Fixed Broadband Wireless Access Systems, October 2004.

[6] M. Feder, "Enabling wireless uncompressed HDTV connectivity with a unique video-modem approach: A technical overview," White Paper, AMIMON Ltd., 2008.

[7] "WirelessHD specification version 1.0 overview", October 2007.

[8] T. M. Cover and J.A Thomas, *Elements of Information Theory*, Wiley-Interscience Publication, 2001.

[9] Marina Bosi and Richard E. Goldberg, *Introduction to Digital Audio Coding and Standards*, Springer, 2003.

[10] TIA/EIA Interim Standard, Physical Layer Standard for cdma2000 Spread Spectrum Systems, Release C, May 2002.

[11] A. Viterbi, "Error bounds for convolutional codes and an asymptotically optimum decoding algorithm," Information Theory, IEEE Transactions on, vol.13, no.2, pp. 260-269, Apr 1967.

[12] G. D. Forney, "The Viterbi algorithm," Proceedings of the IEEE , vol.61, no.3, pp. 268-278, March 1973.

[13] J. D. Johnston and K. Brandenburg, "Wideband coding - Perceptual considerations for speech and music," in Advances in Speech Signal Processing, Furui and Sondhi (Ed.), Marcel Dekker, 1991, Preprint 3011 A-4.

[14] Method for objective measurements of perceived audio quality, Recommendation ITU-R BS.1387-1, Geneva, Switzerland, 1998-2001.

[15] S. S. Ghassemzadeh and V. Tarokh, "The ultra-wideband indoor path loss model," IEEE P802.15 Working Group for Wireless Personal Area Networks, 2002.

[16] V. Erceg et al. IEEE P802.11 Wireless LANs Channel Models, IEEE 802.11-03/940r4, May 2004

Jelena Kovacevic received the B.S. and M.S. degree from the University of Novi Sad, Faculty of Technical Sciences, Department for Computer Engineering and Communications were she is presently a Ph.D. candidate. Her current research focuses on streaming video and audio over wireless channels. She is research and teaching assistant at the University of Novi Sad. She is also operations manager at RT-SP working on HD audio implementation on embedded systems.

Dragan Samardzija received the B.S. degree in electrical and computer engineering in 1996 from the University of Novi Sad, Serbia, and the M.S and Ph.D. degree in electrical engineering from Wireless Information Network Laboratory (WINLAB), Rutgers University, USA, in 2000 and 2004, respectively. Since 2000 he is with Bell Laboratories, Alcatel-Lucent, where he is involved in research in the field of the next generation wireless systems. He is also teaching at the University of Novi Sad. He authored more than 40 journal and conference publications. He has been a participant in a number of research projects sponsored either by the NSF or the European Union. As a member of the BLAST project team, he is a recipient of the 2002 Bell Labs President's Award. He received Central Bell Labs Teamwork Award for the HSDPA Demonstration Team, 2003.

Miodrag Temerinac received the Ph.D. degree in electronics and computer engineering from University of Belgrade, former Yugoslavia, in 1983. He is Alexander-von Humboldt fellow and he spent two years (1988-1990) having Humboldt fellowship at the University of Hanover in Germany. In 1992 he changed to industry joining Micronas GmbH in Freiburg, Germany working on the IC development and later switching to management. In 2005-2006 he spent one year as the director of system development in the new founded R&D Center in Shanghai. Today he is the managing director of the Information Technologies SW Center in Novi Sad and the head of Computer Engineering and Computer Communications at the Faculty of Technical Sciences in Novi Sad, Serbia. His fields of interest are DSP algorithms and architectures, hw/sw co-design of complex systems on chip and product development in consumer electronics.