Optimally Using the Bluetooth Subband Codec

Christian Hoene and Mansoor Hyder
Universität Tübingen, WSI-ICS
72076 Tübingen, Germany
Email: hoene|mansoor.hyder@uni-tuebingen.de

***Abstract*—The Bluetooth Special Interest Group (SIG) has standardized the subband coding (SBC) audio codec to connect headphones via wireless Bluetooth links with the A2DP profile. SBC compresses audio at high fidelity while having an ultra-low algorithm delay and very low complexity. SBC is configurable to a very large extend. Six parameters can be changed including sampling frequency, bit rate and algorithmic delay. In total, more than 60000 different configuration can be selected. In this publication, we objectively measure the audio and speech quality, if SBC runs under those configurations to find the most suitable parameter under given rate and delay constrains. Finally, we present objective tests comparing the audio codecs SBC, CELT, APT-X and ULD.**

***Index Terms*—Bluetooth SBC, audio quality evaluation, VoIP**

I. Introduction

The Bluetooth SIG, the standardization body for Bluetooth related technologies, published a specification to support high quality audio distribution to Bluetooth devices called A2DP [2]. It is intended to connect wireless headsets and headphones via Bluetooth to an audio source. The A2DP profile defines which audio codecs should be used over a Bluetooth link. Multiple codecs are supported the mandatory SubBand Codec (SBC). The SBC comes with a couple of properties which make it worthwhile to consider it beyond its original usage scenario of connecting wireless headphones. More precisely, we use SBC for phones supporting adaptive Audio over IP .

One of the nice features of SBC is that it is configurable to a large extend. Most encoding parameters such as the sampling rate, the number of frequency bands it compresses, the bit rate and frame size can be freely selected at run-time to cope with changed requirements. For example, if only speech has to be transmitted, its bit rate can be reduced or even the bit rate can be further reduced during silence periods [3]. Rate reductions help to save energy in case if it is required. Also, on the Internet, if the bandwidth is too low, then both frame and bit rate can be changed.

Another feature of SBC is the algorithm delay which is in the order of a few milliseconds. Thus, the SBC codecs can be used for musician playing over the Internet, which require a total one-way acoustic transmission delay of about 25 ms [4].

However, to optimally take advantage of SBC, one has to know which of SBC parameter value sets provide an optimal quality, bit rate and latency tradeoff. Therefore, we conducted formal subjective listening-only tests following the MUSHRA method and various objective tests using instrumental methods standardized such as ITU-R BS.1387-1 [5] and ITU-T P.862 [6]. In light of the results of this work, SBC can now be optimally used for an audio and wide speech transmissions and for variable bit and frame rate transmission over the Internet.

Finally, we compared SBC with other codecs having similar properties, which includes the ultra-low delay (ULD) codec [7], [8], the APT-X codec [9] and the Constrained-Energy Lapped Transform (CELT) [10] codec.

This publication continues with a brief description of SBC, we then describe subjective and objective audio test methodology before presenting the results on how well SBC encodes audio and speech. Before summarizing, we show preliminary objective results comparing SBC with other similar low-delay audio codecs.

II. The A2DP Subband Codec

The Bluetooth's low Complexity Subband Coding (SBC) is defined in A2DP specification version 1.0 [2][1] and is based on work of Frans de Bont [11] and Bernard et al. [12]. The SBC encoders take, as an input, signed 16-bit PCM coded audio signals having a *sampling frequency* f_s of 16, 32, 44.1 or 48 kHz. SBC can run in a one channel *mono* mode or in the two channel *stereo*, *joint-stereo* or *dual channel* modes.

The SBC encoder converts the stereo audio signal into multiple subbands which are equally spaced. The subband coder uses 4 or 8 subbands. A polyphase quadrature filters converts $n = subbands$ audio samples into n single subband samples. These n samples form one *block*. SBC collects 4, 8, 12 or 16 blocks before using these blocks to calculate the maximal loudness of each subband. The loudnesses are then rounded up to the next power of two. Using scale factors, the subband audio signals are normalized to values ranging between $[-1;1]$. The normalized subband samples are not transmitted in full resolution but are quantized.

SBC supports two different algorithms for calculating how many bits should be allocated to each subband. The two modes are called SNR and LOUDNESS. The SNR mode is simple and calculates the number of bits needed using $(\log_2 scalefactor) - 1$, where $scalefactor$ is calculated during normalization. The LOUDNESS mode calculates the bit needed in a way similar

This work has been funded by the DAAD/HEC and the Universität Tübingen. This paper is partly based on a technical report [1].

[1] The more recent version 1.2 has a couple of editorial errors and thus is incomplete.

to the SNR mode but it uses a weighting based on subband position and the sampling rate. More bits are allocated to the lowest band whereas the higher bands require a lower number of bits. Also, subbands with a medium loudness get more bits due to the sacrifice of quiet bands.

If the requested number of bits is calculated, a limited number of bits are distributed to the bands. Typically, the number of bits given the *bitpool* parameter is constant. These bits are distributed among all subbands. The bits from a given *bit pool* are distributed in proportion to the relative number of demanded bits. Subbands that need more bits, get more bits but not necessarily all the bits they have requested for.

Depending on the SBC coding parameters, the length of an SBC frame, the coding rate, the frame rate and the algorithmic delay varies to a large extend. Also, the SBC's algorithmic delay is variable. The encoder reads $blocks * subbands$ samples and introduces a delay of $blocks * subbands - 1$ samples. The analysis and synthesis filters add a delay of of $10 * subbands - 1$ samples. Thus, the total algorithmic delay is calculated as:

$$delay = \frac{((blocks + 10) * subbands - 2)}{f_s} \tag{1}$$

III. Human and Objective Audio Assessments

ITU Recommendation BS.1534 [13] describes a procedure on how to judge the impact of intermediate audio degradations. In listening tests, those degraded audio samples are rated relative to a reference signal. Typically, a continuous scale called subjective difference grade (SDG) having as anchors the values 0 (Imperceptible), -1 (Perceptible but not annoying), -2 (Slightly annoying), -3 (Annoying), and -4 (Very annoying) is used. These tests have to be done repeatedly with multiple listeners and the results are then averaged.

The signal items used for our MUSHRA tests are based on the audio items given in ITU-R BS.1387 and the „Kiel Corpus Vol. 1“. We generated anchors consisting of IRS48 filter for narrow-band, P341 filter for wideband, a super-wideband filtering at 14 kHz (all made with the ITU-T G.191 software), and a version sampled at 8000 Hz frequency. Distorted samples contain SBC encoded samples and samples distorted by packet loss and packet loss concealment. We conducted listening-tests and noted responses of questions from 12 participants/subjects. In total, we got 584 assessment values, each ranging from 0 to 100.

Beside the MUSHRA values obtained from the listening-only tests, we also applied computational methods for perceptually assessing the quality of speech and audio transmission. The ITU developed an improved algorithm that is called perceptual evaluation of audio quality (PEAQ) [5]. PEAQ is intended to predict the quality rating of low-bit-rate coded audio signal. Two different versions of PEAQ are provided: a basic version with lower computational complexity and an advanced version with higher computational complexity. In addition to PEAQ, we used ITU P.862 (PESQ) for evaluating speech quality. We used PESQ for narrow and wide band assessment of the down-sampled but not IRS filtered sample items. Throughout this publication we will use the raw calculation results of PEAQ denoted as Objective Difference Grade (ODG).

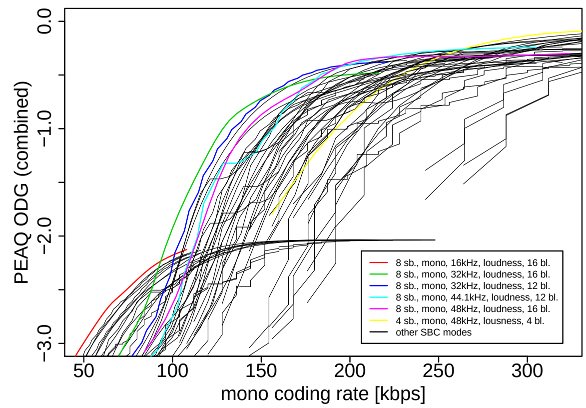

Figure 1. Using SBC for mono audio

Because objective audio quality evaluation is not as good as the human interrogation, we need to compare the results of subjective and objective assessment to figure out the precision, weaknesses and strength of the objective assessment algorithms. The results of this comparison have been published in [1]. If one combines both the results of the basic and advance versions of PEAQ, the prediction performance compared to the subjective rating increases. For example, averaging both ODG values and mapping them to MUSHRA yields a goodness of fitness of $R^2 = 0.907^2$, which results in a good correlation between subjective and objective results. Thus, we will use throughout this publication for objective rating the combined results of PEAQ-AV and BV.

In case of speech, we applied PESQ, which is known to have a correlation of about $R = 0.94$ for most distortions it has been designed for. We denote the raw results of PESQ described as PESQMOS-NB and PESQMOS-WB.

IV. Evaluation

As described in Section II, SBC can be parametrized in a wide range. Even though, the A2DP defines some recommended parameters to use, we are interested in verifying which parameter sets are best at a given bandwidth.

To address these questions, we run extensive simulations with PEAQ varying both the parameters and the reference samples. For all the reference sample files we calculated all coding modes varying the allocation mode (SNR, LOUDNESS), the number of subbands (4 and 8), the number of blocks (4, 8, 12, 16), the coding mode (mono, stereo, joint stereo) and the bit pool value (10, 12, 14, 18, 19, 25, 29, 31, 40, and 50). Overall, 4800 PEAQ ODG-BV and ODG-AV values have been calculated. In the one channel mode, we have compared the degraded files to the mono version of the references file. In the stereo modes, the degraded samples were compared with the original stereo reference file. In addition, we approximated the quality for remaining bitpool parameters between 11 and 49 with a natural spline function in order to save time.

In Figure 1, we plot the averaged ODG values versus the coding rate. The ODG results of parameter sets which differ only in the bitpool values are interconnected by lines. We also highlighted the best parameter sets with coloured lines. In the mono mode up to a rate of about 96 kbps the 16 kHz, 16 blocks, LOUDNESS coding mode is the best. Then between 96 and 72 kbps, the 32 kHz sampling rate should be chosen. Further up the axis multiple best coding alternates at fast pace.

In the stereo mode, choosing the right mode is simpler. Up to 106 kbps, the 16 kHz, 16 block, LOUDNESS mode is best. Both the stereo and joint-stereo mode seem to encode equally good. Then up to 237 kbps, the 32 kHz sampling rate is the best. At higher quality, the 44.1 kHz stereo encoding mode can be chosen.

Some of coloured lines match those recommended value in the A2DP standard. However, if a lower audio quality is required, the results indicate that it is better to use the 32 kHz coding mode instead of 44.1 and 48 kHz. Also, the joint stereo mode does not increase significantly the audio quality as compared to the stereo mode.

In packetized networks, speech frames are also transmitted in packets, which have packet headers. In the Internet, the size of packet headers can vary depending on the kind of protocol used and whether header compression is applied. In a typical scenario, one frame is transmitted with the RTP, UDP, IPv4 and IEEE 802.3 protocols and thus each packet contains packet headers having 12 bytes, 8 bytes, 20 bytes and 18 bytes respectively. In the end, the gross rate, as measured on the physical layer is much larger than the actual coding rate. Thus, we also consider this gross rate in addition to the coding rate. The gross rate calculates as

$$r_{gross} = r_{coding} + packetoverhead * framerate \qquad (2)$$

where coding rate gives the coding rate of the SBC codec, *packetoverhead* is the number of bits for protocol headers in each packet (typically 58*8=464), and the *framerate* is the number of packets/frames per second.

Considering the gross rate, the best coding mode for bandwidth constraint link is shown in Figure 2. As compared to the Figure 1, the best coding mode hardly changed.

The Bluetooth SIG standardization group currently considered to use SBC for wideband headsets to transmit the microphone signal in the upcoming Hands-Free Profile (HFP) version 1.5. We measured the mean ITU P.862 wideband MOS results for speech samples including the Kiel corpus samples and the ITU BS.1387 speech samples (English male, English female, German male, Suzanne Vega singing). For all SBC coding modes, the mode with 8 subbands, 16 kHz sampling rate, kHz sampling rate, LOUDNESS allocation mode, 16 blocks and mono provides the best speech quality. It performs slightly better than ITU G.722 and 48, 56, and 64 kbps.

In addition, we present the results of the SBC 16 kHz coding mode with samples that were shifted by one octave up. We refer to this mode as SBC 8 kHz sampling mode, which however is not standardized. The measured narrowband PESQ values for this mode are even better than of the wideband PESQ values at 16 kHz sampling mode.

The SBC might not perform equally well for all kind of acoustic contents. To avoid a content specific judgement in the previous tests, we have taken the objective ratings averaged over multiple sample files. This time, we take the average of all the sampling modes but keep the sample file to be fixed. The 16 kHz sampled speech and noisy instrument such as the snare drum can be compressed rather well. On the other side, single instruments having high tonal sounds such as the glockenspiel, the tambourine, the flute, the triangle and the clarinet are encoded relatively badly. If looking on the measured speech qualities, it is interesting to note that high female voices are encoded worse than low male ones. Music such as the opera, the piano and full band speech show an average compression efficiency. Expectingly these are also the most common contents which are transmitted via a hifi-phone.

V. Related Codecs

Several encoding schemes can compress an audio signal with very low algorithmic delay. One of the simplest encoding techniques is to use a PCM coding at different sampling rates. Also, a logarithmic quantization of the samples [14] can be considered. The classic logarithmic quantization called µ-Law and A-Law has been standardized in ITU_T G.711 for 8 bits per sample at a sampling rate of 8000 Hz but the IETF RTP and SDP standards allow the usage of µ- and A-Law even at other sampling rates, for example at 48000 Hz. Thus, the use of logarithmic quantization (or PCM) allows the transmission of audio signals even with existing standards. The APT-X stereo codec has an algorithmic delay of 1.9 ms and a rate between 128 and 384 kbps but it is not available for free. Also Fraunhofer's Ultra Low Delay Encoding [8] compresses stereo audio to 96 kbps with a frame size of 2.7 ms and an algorithmic delay of 5.4 ms. Again this codec is not available as open source. Recently, the CELT codec has been developed by J.-M. Valin et al. [10]. It is open source and has a trade-off between very good quality and rate and also posseses very low algorithmic delays. We tested it at various sampling rates and frame sizes. We used it with an algorithmic delay of 150% of the reciprocal of the frame rate.

We tried to compare the performance of those coding schemes. However, we were not able to get a working implementation of ULD and APT-X. Thus we asked fellow researchers and a company to encode and decode a large sample file containing multiple samples files. The large sample file contained the shorter samples used throughout this work but kept them separated by one second of silence. After getting back the encoded and decoded large sample files, we removed aligned the file to the original and splitted it again into small files. Next we compared the original small samples with the degraded samples using the combined PEAQ metric. This step was done multiple times for different codecs for coding modes both in mono and in stereo conditions.

It is a general consensus that PEAQ is not capable of comparing different codecs because the kind of distortions

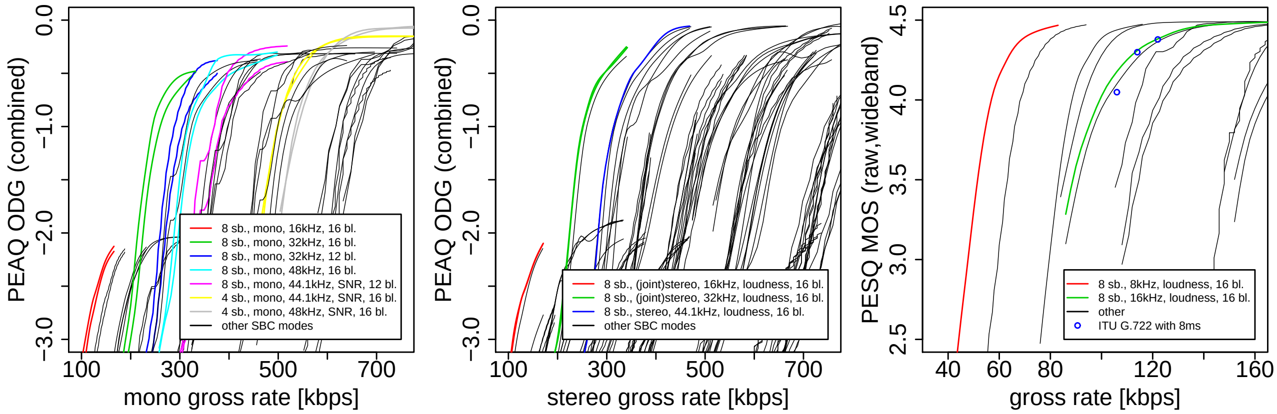

Figure 2. Using SBC for mono audio, stereo audio, and speech measuring the gross rate (including RTP, UDP, IPv4, and Ethernet packet headers)

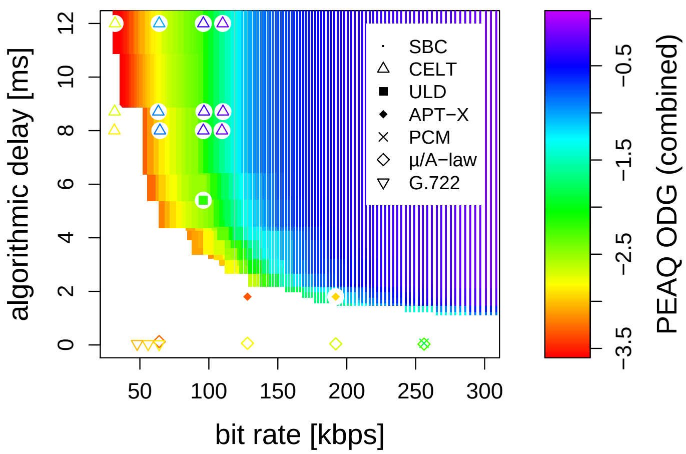

Figure 3. Objectively audio quality measured with combined PEAQ of samples encoded with different codecs and different coding modes. We display the ODG value versus algorithmic delay and bit/gross rate.

might vary to a large extend. PEAQ evaluates different kinds of distortions on different scales and therefore a comparison of different codecs without proper subjective verification has to be avoided. Being aware of these facts, we still included PEAQ comparison results into this document knowing that they are not means of suitable performance codec comparison but only provide an indication of quality. The PEAQ comparison results are shown in Figure 3.

The PEAQ-ODG ratings in the mono mode are clear. CELT v0.6 outperforms all other codecs at the tested rate vs. delay trade-offs. ULD is better as SBC if one considers just the bit rate and is equally good if looking at the more realistic gross rate.

VI. Summary and Conclusion

The main contribution of publication is on the optimal usage of SBC on the Internet transmission and on wireless (Bluetooth) connections. Now, we know, which, transmission mode of SBC to be chosen for given bandwidth and delay constraints and a Bluetooth A2DP device can operate more efficiently by using an optimal codec parameter set. Also, the results show the strength and weakness of SBC. Audio which are difficult to be coded with SBC are in general audio signals containing pure tones and stable harmonic series such as the harpsichord and the pitch pipe. On the other hand SBC is relatively good in coding audio signals with a high time resolution, e.g. castanets and applause.

References

[1] C. Hoene and M. Hyder, "Considering bluetooth's subband codec (sbc) for wideband speech and audio over internet," Universität Tübingen, Tübingen, Germany, Tech. Rep., 2009.

[2] "Advanced audio distribution profile (A2DP) specification version 1.0," http://www.bluetooth.org/, 2003, bluetooth Special Interest Group, Audio Video WG.

[3] M. Pilati, Laurent; Zadissa, "Enhancements to the SBC CODEC for voice communication in mobile devices," in *AES Convention 124*, no. 7347, Amsterdam, Netherland, May 2008.

[4] A. Carôt and C. Werner, "Network music performance - problems, approaches and perspectives," in *International School of new Media, Institute of Telematics, University of Lübeck*. Music in the Global Village - Conference, Sep. 2007.

[5] ITU-R, "Method for objective measurements of perceived audio quality," Recommendation BS.1387, Nov. 2001.

[6] *Perceptual Evaluation of Speech Quality (PESQ), an Objective Method for End-To-End Speech Quality Assessment of Narrowband Telephone Networks and Speech Codecs*, Recommendation P.862, ITU-T Std., Feb. 2001.

[7] G. Schuller, B. Yu, D. Huang, and B. Edler, "Perceptual audio coding using adaptive pre- and post-filters and lossless compression," *IEEE Transactions on Speech and Audio Processing*, vol. 10, no. 6, pp. 379–390, Sep. 2002.

[8] J. Hirschfeld, J. Klier, U. Kraemer, G. Schuller, and S. Wabnik, "Ultra low delay audio coding with constant bit rate," in *AES Convention 117*, no. 6197, October 2004.

[9] [Online]. Available: http://www.aptx.com/

[10] Xiph.Org Foundation, "The CELT ultra-low delay audio codec," http://www.celt-codec.org/, Mar. 2009.

[11] F. Bont, M. Groenewegen, and W. Oomen, "A high-quality audio coding system at 128 kb/s," in *AES 98th Convention*, no. 3937, Paris, February 1995.

[12] R. J. Bernard, D. Y. Francois, and R. J. Yves, "Digital transmission system using subband coding of a digital signal." European Patent, Publication number: EP0400755 (B1), Sep. 1995.

[13] ITU-R, "Method for the subjective assessment of intermediate quality levels of coding systems," Recommendation BS.1534-1, Jan. 2003.

[14] N. S. Jayant and P. Noll, *Digital Coding of Waveforms: Principles and Applications to Speech and Video*. Prentice Hall Professional Technical Reference, 1990.