Estimating MP3PRO Encoder Parameters From Decoded Audio

Paul Bießmann[1], Daniel Gärtner[1], Christian Dittmar[1], Patrick Aichroth[1], Michael Schnabel[2], Gerald Schuller[1,2], and Ralf Geiger[3]

[1]Semantic Music Technologies, Fraunhofer Institute For Digital Media Technology IDMT, Ehrenbergstr. 31, 98693 Ilmenau, daniel.gaertner@idmt.fraunhofer.de
[2]Institute Of Media Technology, Ilmenau University of Technology, Helmholtzplatz 2, 98693 Ilmenau, gerald.schuller@tu-ilmenau.de
[3]International Audio Laboratories Erlangen, Am Wolfsmantel 33, 91058 Erlangen, ralf.geiger@iis.fraunhofer.de

Abstract: We present an approach to estimate encoder parameters from previously MP3PRO-compressed, then decompressed audio material. The algorithm has been designed to identify the presence of spectral band replication (SBR), and for bit-rate detection. Furthermore, MP3 compression parameters like frame offset and block type are detected. As evaluation results show, the approach is able to identify SBR with a high accuracy, while the compression bit-rate detection is prone to errors, especially for higher bit-rates.

1 Introduction and Motivation

Today, lots of music is stored and distributed digitally. Large hard-drives, fast internet connections and the wide-spread use of audio compression techniques are supporting this development. Even though audio codecs like MP3 have been designed to preserve the audio quality for a human listener, lossy audio compression techniques discard irrelevant information in the audio signal in order to achieve low compression bit-rates and small file-sizes. This might not necessarily be an important issue for the average music listener. However, several scenarios exist, where loss-less distribution of audio content can be critical, e.g., when dealing with high quality production recordings or if the user payed extra for loss-less quality. The quality of distributed digital audio is important. Usually, low quality versions encoded with a low bit-rate are used to enable the preview for the customers. Unfortunately these reduced versions can become an object of a fraud. For example, a freely distributed low quality preview MP3 can be transcoded at a higher bit-rate and sold online as a high quality product.

An audio file does not reveal directly whether the data has already been compressed with a lossy audio coder, and therefore might not be suitable for further processing. We thus present methods that help to identify if the audio data has undergone previous lossy compression. Additionally, our goal is to determine the encoder parameters automatically.

Knowledge of the encoder parameters of a previous encoding process can help to re-compress an audio signal with less artifacts by using the same encoder parameters again.

Furthermore, if different encoder parameters are determined over the duration of an audio signal, it can be assumed that the audio material has been concatenated from different audio signals. Thus, the method presented here can also be applied to (semi-) automatically detect audio tampering [Gup12].

The remainder of this paper is organized as follows. Section 2 presents the basics of the investigated compression algorithms, followed by a section on related work in compressed audio analysis (3). Next, our approach is presented in Section 4. An evaluation is carried out in Section 5 and finally in Section 6 conclusions and an outlook are presented. Following previous papers in this area, we use the term *inverse decoder* for the overall process.

2 MP3 And MP3PRO Basics

In this section, we outline components of typical MP3 and MP3PRO encoders, with a focus on those that are of particular interest in the presented inverse decoder approach.

2.1 MP3

MP3 (the common short name for MPEG-1/2 Audio Layer III) is an algorithm for lossy compression of audio data. In principle, a share of the data reduction is achieved by removing information from the signal that is assumed to be irrelevant to human listeners. Figure 1 shows an overview of the components of an MP3 encoder system. The components of an MP3 encoder are summarized in [Bra99].

- Psychoacoustic Model. The model estimates the time- and frequency-dependent masking threshold of the current input signal. The term masking in the frequency domain describes the fact that some spectral components can not be perceived by the human auditory system in the presence of another signal at a nearby frequency (masker). The masker can be narrow-band noise or a tonal component. Hence neighboring sub-band signals can have masking effects on each other resulting in frequency-dependent increases of the hearing threshold. Signals at frequencies with a level below this threshold cannot be perceived, i.e., they are masked, and can therefore be heavily quantized or set to zero. If the quantization noise in each sub-band stays under the masking threshold, no difference between the original and compressed signal is audible [Bra99, BFF95].
  The psychoacoustic model further determines whether the audio signal is stationary or transient at a given point of time, and choses an appropriate window. Four different window types (also referred to as block types) can be used: Long window

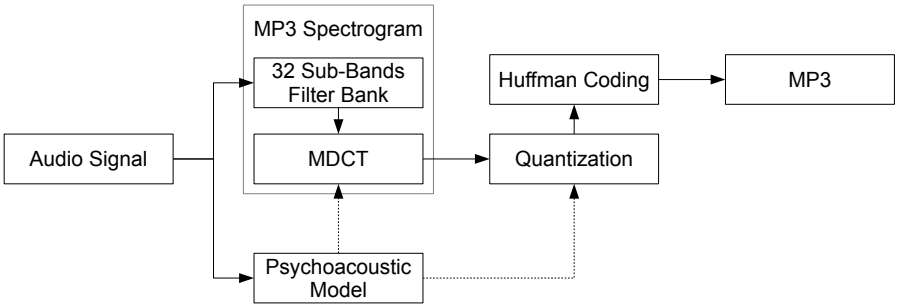


Figure 1: Overview of the MP3 algorithm

(type 0), start window (type 1), short window (type 2), and stop window (type 3). The long window is the most common window. The short window is used to process transient regions, which would cause pre-echos on the decoder side when using a long window. The remaining windows are used as transition from long to short (start) and short to long (stop). The determination of the frame offset and the used block-type for each window is described in Section 4.2.

- MP3 Spectrogram. In the encoder, the signal is first filtered by a quadrature mirror filter bank (QMF) with 32 sub-bands [Rot83]. Next, each sub-band signal is decomposed by a Modified Discrete Cosine Transform (MDCT) [PJB87] with 6 or 18 sub-bands. The consecutive transforms also reduce redundancy in the signal, based on the so-called transform coding gain. Depending on whether the signal is stationary or transient, one of the two different MDCT sub-band configurations is switched to, resulting in either 576 or 192 spectral bins in total [Edl89]. The spectral coefficients resulting from this process are further grouped into Scale-factor Bands (SFB).

- Quantization and Coding. With MP3, quantization and coding are combined. The output of the psychoacoustic model (masking threshold) is used to adapt the quantization step size such that the quantization error ideally stays below the computed masking threshold for each sub-band. Two interleaved iteration loops are used in the encoder. The outer loop is referred to as the noise control loop and reduces the quantization noise until it is under the masking threshold for every spectral band. Reducing the quantization noise requires a higher bit rate. This means that each time the outer loop calculates a new so-called scale-factor (which denotes the resulting quantization step-size), the estimated bit rate has to be updated. This is done in the inner loop, the so-called rate loop. The SFBs are encoded using entropy coding, resulting in variable word lengths. This makes the estimation of the resulting bit rate necessary. A number of 2- or 4-dimensional Huffman tables are employed to

assign frequent values to shorter code words, while less frequent values are assigned to longer code words. By varying the quantization step sizes, the rate loop can vary the bit rate until it meets the requirements of the noise control loop [Bra99].

Based on this MP3 encoder principle it is obvious, that multiple encoding and decoding will lead to artifacts, since each encoding and decoding adds noise that will be considered as a part of the signal of interest for the next encoding step.

2.2 MP3PRO

In MP3PRO [DLKO02], the MP3 format is extended by Spectral Band Replication (SBR, [ZEEL02]). The basic idea of SBR is to express the high frequency content of an audio signal as a modified copy of the lower frequency content. The lower frequency content is transmitted via the core codec (in this case MP3 at half the original sample rate). To reconstruct the higher frequency content, only the modification parameters (the ancillary data) are needed. During encoding, the SBR component and the core encoder are connected and important information can be exchanged. Therefore it is possible that for the system to handle situations where the characteristics of the low-band and the high-band differ a lot, e.g. a strong harmonic series in one band and a noise-like signal in the other one.

2.2.1 Decoder

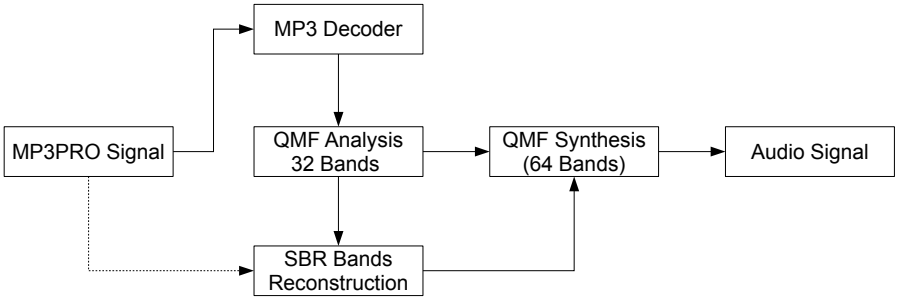


Figure 2: Overview of an MP3PRO decoder

3 Related Work

Figure 2 gives an overview of the decoder process in MP3PRO. The core MP3 bit-stream is decoded by an MP3 decoder. At the same time, ancillary data is read from the bit-stream. The MP3 decoded time-signal is then forwarded to the SBR-decoder in conjunction with the ancillary data. Next, frequency analysis of the time-signal is performed using a QMF filter bank. From the spectrum frames resulting from filter bank analysis and the ancillary data, the SBR part of the spectrum frames is formed. The MP3 frames and the SBR frames are then combined into a 64 bin frame and re-synthesized using a 64 band QMF synthesis filter bank. I The idea of the inverse decoder has been first discussed in [HS00]. The authors address the generic audio coding structure and present the details of the inverse decoder algorithm based on the MPEG-2/4 AAC coder. In addition, possible applications for inverse decoding are presented. The authors suggest the use for the detection of known coding tools used to encode a given audio signal. Also, characteristics of unknown encoders can be investigated. Finally, distortion can be reduced when a signal is repeatedly encoded and decoded. The analysis of the encoding parameters is done in the following step-by-step process:

- The decoder framing grid is determined.
- For coding schemes with flexible coder filter banks, the filter bank parameters are recovered.
- The quantization information is estimated from the quantized spectral values.
- Other parameters, such as joint stereo coding modes, can be recovered if applicable.

The experimental results show that the recovery of several basic encoding parameters is feasible. An inverse decoder specialized for MP3 coding was presented in [MHG02]. Here the framing structure is determined by trying different frame-offsets and then detecting at which offset quantized amplitude distributions occur. Then the quantization step sizes, and many more parameters used in MP3 coding are estimated, with the goal of a complete inverse decoder. The experimental results include the evaluation of the overall reconstruction precision of the system. Here the decompressed, inverse decoded audio signal is compared to the decompressed original bit-stream by means of the standardized PEAQ (Perceptual Evaluation of Audio Quality) measurement method. From these prior works, especially the frame offset, block type and quantization are of further interest for this work and will by further explained in Section 4.
D'Alessandro and Shi [DS09] show the bit-rate of decoded MP3 files can be detected from the PCM audio signal by analyzing high frequency components. They define five classes of bit rates (CBR 128 kbps, 192 kbps, 256 kbps, 320 kbps, and VBR-0) and apply a Support Vector Machine for the classification. In their evaluation, the average success rate reaches 97%.
Yang et al. [YSH09] propose a method to recover the original bit-rate of decoded MP3 by analyzing the number of the MDCT coefficients with small values. In particular, they show that there are more MDCT coefficients of small values in normal MP3 than in MP3

files, that have been previously encoded with a lower bit-rate than their actual one. These findings are supported by the theoretical analysis on the quantization artifacts during the double-compression (first time with the low bit-rate and second time with the high bit-rate). The accuracy of the fake-quality detection method exceeds 97%.
In [YQH08], frame offsets are determined over the complete duration of an audio signal. If an audio signal is encoded, the frame offset is expected to stay constant. The authors use this information to determine locations in the audio signal where the signal has been manipulated, e.g. parts have been added to or removed from the signal.

4 Approach

In this section, we describe the approach used for MP3PRO analysis. Some components can be directly adapted from the mechanisms described in [HS00] and [MHG02], the inverse MP3 decoder. In this paper, we are only investigating mono audio data.
In MP3PRO, mp3 serves as core codec for SBR, operating at half the original sample rate. An estimation of the core audio signal can be obtained by sample rate conversion by factor two of the audio signal. Based on the core audio signal, MP3 Frame Offsets, Block Types, Quantization and Bitrate are estimated, following the approaches of the inverse MP3 decoder. The presence of SBR is determined on the original audio signal.

4.1 Spectrogram Calculation

Two different spectral representations are used in the presented analysis.

4.1.1 QMF Spectrogram

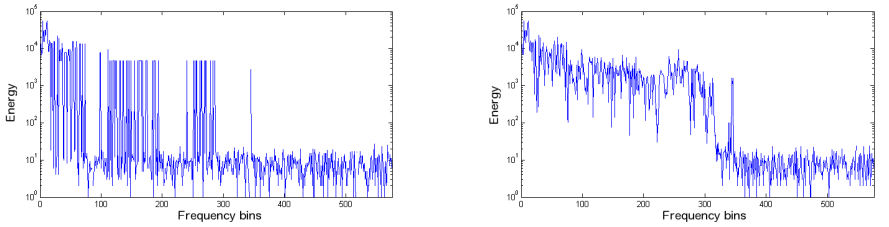
A QMF spectrogram is needed for the analysis of the presence of SBR. The QMF spectrogram calculation is the inverse of the QMF synthesis in Figure [DLKO02]. The audio signal is filtered with a QMF filter bank with 64 sub-bands.

4.1.2 Core Codec Spectrogram

In order to obtain the core codec spectrogram, the audio signal is first down-sampled by factor 2, using an FIR multi-rate filter. Next, following an MP3 encoder, the down-sampled core audio signal is processed with an 32 sub-bands QMF, followed by an MDCT.

4.2 Core Data Frame Offset and Block Type Detection

It is essential that the frame offset and the block type are correctly determined. This is done on the core codec spectrogram. Figure 3a shows a spectrum extracted with the original offset and the correct window, while Figure 3b shows the spectrum extracted with just one sample offset to the correct framing position. In 3a, the bins whose energy has



(a) Spectral coefficients when using the correct frame offset

(b) Spectral coefficients for a frame offset shifted by one sample

been discarded or heavily quantized according to the perceptual model can be clearly seen, while 3b does not reveal these typical characteristics.
Two measures are used to obtain a detection function for the frame offset. The first measure is given by the number of energy bins having values close to zero (they are not necessarily zero due to quantization noise). The second measure quantifies the fluctuation in the spectrogram frame. It is calculated as the sum of the absolute differences between neighboring bins. At this stage of the algorithm, the block-type for the current block is not known. Since type 0 blocks are the most common, this approach is performed assuming type 0 blocks for several frames in the beginning of the signal. Calculating both measures while shifting the window sample by sample over the signal, both measures will exhibit extreme values when the correct offset is found. As long as an audio file has not been manipulated, the frame offset stays the same over the whole duration of the audio file.
Knowing the correct offset, the same procedure can be repeated for different block types for all frames. Incorporating knowledge about which block-types can be successors of a given block-type, the analysis process can be sped up.

4.3 MP3PRO SBR Detection

In this study, two cues are investigated to detect the presence of SBR. As with MP3, for the analysis of a potential MP3PRO signal we also need to know the right framing offset and block types. Since MP3 is used as the core codec in MP3PRO, we down-sample the audio signal by a factor of two, and use the MP3 inverse decoder algorithms (4.2) to determine the core coder's parameters. Finding a framing offset with high confidence in this sub-sampled domain is already a hint for the use of MP3 as core codec in MP3PRO.
Next, the potentially replicated high band is analyzed. For each QMF spectrogram frame, low band and high band are compared using the Pearson's correlation coefficient [Web02].

At this point of analysis, the border between low-band and high-band is not yet known, so all possible border frequencies have to be tested in order to estimate a splitting frequency. To the knowledge of the authors, in the MP3PRO encoding process, the splitting frequency is set only based on the target bit-rate and the sample rate. Therefore, a correct determination of the splitting frequency could already give hints on the target bit-rate. However, this is not further investigated in this paper.

4.4 Core Data Quantization Estimation

As a next step, the quantization parameters are determined. During encoding, the spectral coefficients are weighted, where the weighting factor can be determined using the corresponding scale-factor. Afterwards, a power law quantization is applied. During determination of the quantization, both steps are reversed.
In the non-uniform quantization domain, the coefficients are situated on the quantization grid, which is the main idea behind the determination of the scale-factors. Therefore, the compressed coefficients in a scale-factor band will have a greatest common divisor (which relates to the quantization step-size), from which the actual scale-factor can be calculated. Since the components that are used to obtain the MP3 Spectrogram are not perfectly reconstructing, the used QMF analysis in the inverse decoder is only an approximation of the reverse synthesis step, which introduces additional noise. As a consequence the determined values might not have an exact integer value. Values that are far away from their closest integer neighbor further do occur, if the determined quantizer step is a multiple of the correct quantizer step. Then, it should be replaced a smaller one, obtained by dividing the current one by an appropriate integer factor.
It is important to remove the noise floor (those bins that have been set to zero during encoding, but are now exhibiting small noisy values) before the quantization steps are determined. In order to get an estimate for the noise floor, the spectral coefficients per detected block are sorted in descending order. A significant drop of energy can be observed between the larger, information-carrying coefficients and the smaller noise coefficients. The value corresponding to this drop corresponds to the threshold that can be used to suppress the noise floor.
When dealing with MP3PRO, an extension of this original algorithm for MP3 quantization step determination is necessary. Since the SBR part of the data is not quantized, it has a negative impact on quantization estimation and should be excluded in the quantization step analysis. Therefore, the splitting frequency has to be exactly determined first.

4.5 MP3PRO Splitting Frequency Determination

Already in the MP3PRO SBR detection step, the splitting frequency can be determined, since it corresponds to the correlation coefficient with the highest value. However, this method is not always reliable. On the other hand, exact knowledge of the splitting frequency is necessary to obtain the quantization steps. To exactly determine the splitting

frequency, the algorithm for quantization steps detection in MP3 can be used again. Since the SBR data is not bound to quantization steps, this algorithm will return very small values for the greatest common divisor during quantization step determination, as long as unquantized SBR coefficients are part of the data that is investigated. Starting with the full frequency range, quantization steps determination is performed until the calculation of the greatest common divisor. As long as SBR data is still contained in the spectral frames, this greatest common divisor will be very small. As a consequence, coefficients from the upper frequency range are successively removed, and after every removal, the corresponding greatest common divisor is determined. Removing all SBR data results in abrupt increase of greatest common divisor. As soon as this can be observed, the splitting frequency is determined, and the quantization steps can be finally obtained from the greatest common divisor.

4.6 MP3PRO Bit-rate Estimation

Since the quantization steps and zero bins (those with values around the noise threshold) are known, the spectral frames can be further processed using quantization and Huffman coding. The corresponding tables can be found in the standard [ISO]. Finally, the amount of data needed to code each frame can be determined. We are limiting this work on constant bit-rates. For SBR data, about 1.5 kbit/s has to be added. The final bit-rate can be determined by averaging over the data requirement in the frames, and rounding towards the next higher possible bit-rate.

5 Evaluation

A first evaluation has been conducted, testing the capabilities of the proposed algorithm. 15 music files from different genres have been encoded to MP3 and MP3PRO in different bit-rates, and then decoded back to PCM. Two experiments have been conducted. In the first experiment, both the decoded MP3 and the decoded MP3PRO files are analyzed for the presence of SBR. In a second experiment, the decoded MP3PRO files are subjected to bit-rate estimation.

5.1 SBR Detection

The results of the SBR detection experiment can be seen in Figure 3. The presence and absence of SBR is identified in most of the test-cases. However, for higher bit-rates, the SBR detection fails in some cases. This is especially true for signals that exhibit a limited bandwidth. There, the encoder can spend more bits on the limited frequency range. This results in untypical spectral coefficients, which makes the framing detection fail.

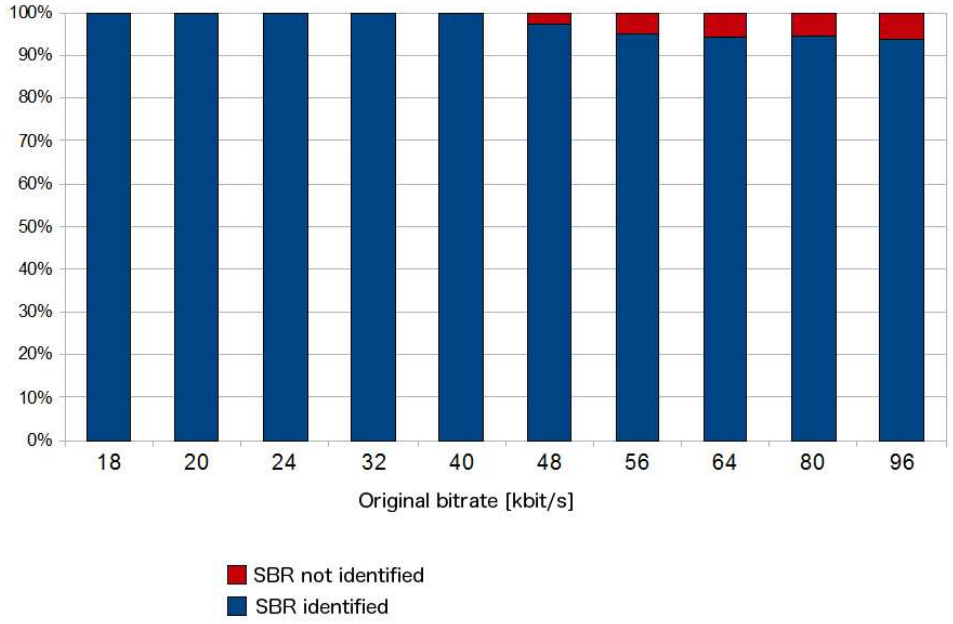


Figure 3: Evaluation results of the SBR detection component.

5.2 Bit-rate Detection

Figure 4 depicts the results of the bit-rate detection component. For lower bit-rates, the bit-rate is detected correctly or inside a tolerance of 8 kbit/s. For higher bit-rates, the number of correct detections increases. At the same time, bit-rates that differ more than 8 kbit/s are returned. The smaller errors can be explained by the possible bit-rates in the MP3PRO format, which are closer to each other in the lower bit-rate regions. For higher bit-rates, neighboring bit-rate candidates are well separated, but at the same time, more bits are used to encode spectral coefficients. This affects framing detection, noise floor detection and therefore also bit-rate detection.

6 Conclusion And Outlook

A method for analyzing decompressed audio signals has been presented. The method identifies whether the audio signal has been previously compressed using MP3PRO with SBR. Furthermore, if SBR is identified, the bit-rate of the file is estimated. While the accuracies returned from SBR detection are satisfying, the bit-rate detection needs further improvement.

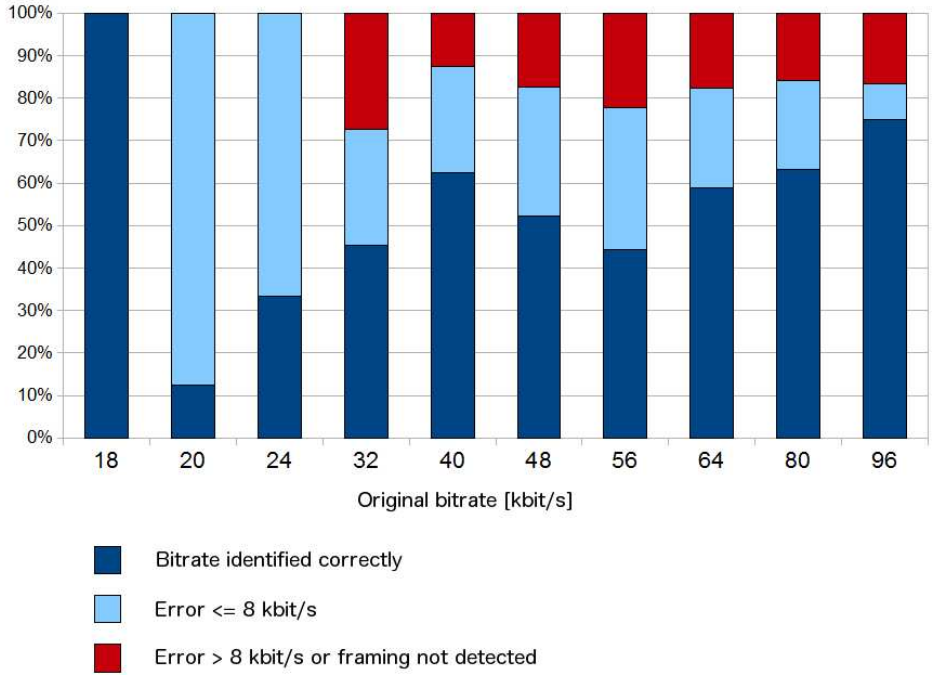


Figure 4: Results of the MP3PRO bit-rate detection component.

In the current system, only SBR detection and bit-rate estimation on constant bit-rate audio material is investigated. In an extended system, also stereo mode analysis could be incorporated, as well as variable bit-rate files. In addition, the experimental results need to be validated with a larger number of test files, allowing to study content depending effects such as with band limited original signals.

7 Acknowledgments

This research has been partially funded by the EU project REWIND. The project acknowledges the financial support of the Future and Emerging Technologies (FET) Programme within the Seventh Framework Programme for Research of the European Commission, under FET-Open grant number: 268478.

References

[BFF95] Frank Baumgarte, Charalampos Ferekidis, and Hendrik Fuchs. A Nonlinear Psychoacoustic Model Applied to ISO/MPEG Layer 3 Coder. In *Proceedings of the 99th Convention of Audio Engineering Society*, 10 1995.

[Bra99] Karlheinz Brandenburg. MP3 and AAC explained. In *Proc. AES 17th International Conference*, 1999.

[DLKO02] Martin Dietz, Liljeryd Lars, Kjörling Kristofer, and Kunz Oliver. Spectral Band Replication, a novel approach in audio coding. In *Proceedings of the 112th AES Convention*, 2002.

[DS09] Brian D'Alessandro and Yun Q. Shi. Mp3 bit rate quality detection through frequency spectrum analysis. In *Proceedings of the 11th ACM workshop on Multimedia and security (MM&Sec)*, pages 57–61, New York, New York, USA, 2009. ACM Press.

[Edl89] Bernd Edler. Codierung von Audiosignalen mit überlappender Transformation und adaptiven Fensterfunktionen. In *Kleinheubacher Tagung*, October 1989. [in german].

[Gup12] Kuo Gupta, Cho. Current Developments and Future Trends in Audio Authentication. *Multimedia in Forensics, Security and Intelligence*, pages 50 – 59, 2012.

[HS00] Jürgen Herre and Michael Schug. Analysis of Decompressed Audio - The "Inverse Decoder". In *Proceedings of the 109th AES Convention*, 2000.

[ISO] ISO/IEC 11172-3:1993. Information technology – Coding of moving pictures and associated audio for digital storagemedia at up to about 1.5 Mbit/s – Part 3: Audio. The Moving Picture Experts Group (MPEG).

[MHG02] Sascha Moehrs, Jürgen Herre, and Ralf Geiger. Analysing decompressed audio with the Inverse Decoder towards an operative algorithm. In *Proceedings of the 112th AES Convention*, 2002.

[PJB87] J. Princen, A. Johnson, and A. Bradley. Subband/Transform coding using filter bank designs based on time domain aliasing cancellation. In *Proceedings of the IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP)*, volume 12, pages 2161–2164, apr 1987.

[Rot83] Joseph H. Rothweiler. Polyphase quadrature filters – A new subband coding technique. In *Proceedings of the IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP)*, volume 8, pages 1280–1283, apr 1983.

[Web02] A.R Webb. *Statistical Pattern Recognition*. John Wiley and Sons Ltd, 2 edition, 2002.

[YQH08] Rui Yang, Zhenhua Qu, and Jiwu Huang. Detecting digital audio forgeries by checking frame offsets. In *Proceedings of the 10th ACM Workshop on Multimedia and Security (MM&Sec)*, 2008.

[YSH09] Rui Yang, Yun-Qing Shi, and Jiwu Huang. Defeating fake-quality MP3. pages 117–124, New York, New York, USA, 2009. ACM Press.

[ZEEL02] Thomas Ziegler, Andreas Ehret, Per Ekstrand, and Manfred Lutzky. Enhancing mp3 with SBR: Features and Capabilities of the new mp3PRO Algorithm. In *Proceedings of the 112th AES Convention*, 2002.