NTERNATIONAL CONGRESS ON ACOUSTICS
MADRID, 2-7 SEPTEMBER 2007

MP3 STIMULI IN ROOM ACOUSTICS

PACS: 43.20.Ye

Hak, C.C.J.M[1]; Vertegaal, J.S.[2]
[1]Technische Universiteit Eindhoven, Department of Building and Architecture, Unit Building Physics and Systems, Laboratorium voor Akoestiek, P.O. Box 513, NL-5600 MB, Eindhoven, The Netherlands; c.c.j.m.hak@tue.nl
[2]Acoustics Engineering; Groenling 43-45, NL-5831 MZ, The Netherlands; h.vertegaal@acoustics-engineering.com

ABSTRACT
In some situations it is very inconvenient or even impossible to perform an impulse response measurement where both the excitation source and the receiving microphone have to be wired to the same measurement system. Think for instance of speech intelligibility measurements through public address systems in railway stations or large airport halls. This is usually solved by open loop measurements playing prerecorded stimulus wave files from a CD player or laptop PC. Given the growing popularity of the MP3 format, which is roughly 10 times smaller in size than the uncompressed PCM format, the question has arisen whether it is also possible to use MP3 stimulus files. To investigate this, PCM and MP3 stimulus based measurements were performed with DIRAC, a room acoustics measurement software tool that supports open loop measurements.

It is shown that the investigated room acoustical parameters are not significantly affected by MP3 compression of sweep and MLS stimuli. Using an open loop measurement system, such as DIRAC, it is therefore possible to perform room acoustical measurements using MP3 stimuli, played from any portable device (e.g. iPod, mobile phone, etc.).

INTRODUCTION
Many room acoustic parameters can be derived from the room impulse responses [1][2]. Impulse response measurements are often based on convolution techniques, using maximum length sequences (MLS) [3] or sweeps [4] as stimulus signals. These techniques basically require the room response to be captured synchronously with the generated excitation signal. In some cases it is necessary to play the stimulus signal and capture the response asynchronously. For instance with speech intelligibility measurements in railway stations, the source may be a CD player in an announcer’s booth in one city, while the receiver is a microphone connected to a PC on a platform in another city. The slight difference between stimulus playback speed and response recording speed requires special open loop measurement techniques.

MP3 is now the most widely supported digital music format for media players. MP3 players can hold the equivalent of many CDs, and MP3 files can easily be stored and transferred via email. All these properties make MP3 an attractive format for stimuli in measurement setups as described above. The question now is whether the inherent lossy compression in MP3 will affect the outcome of the room acoustic measurements.

MP3 BACKGROUND

History
The MP3 format was developed in the late 1980's and early 1990's by the Fraunhofer Society (Germany), IRT (Germany), CCET (France) and Philips (the Netherlands). The development was financed by the EUREKA research program, formalized as MPEG-1 layer 3 (1992) and published in ISO/IEC 11172-3 (1993) [5].

The first software MP3 encoder was released in 1994 by the Fraunhofer Society. In the second half of the 1990's the MP3 format spread over the Internet, partly due to the popularity of the free WinAmp player (1997).

The successor to MP3 is the MPEG-2 Advanced Audio Coding (AAC) standard which was first defined in MPEG-2 part 7 and published in ISO/IEC 13818-7 [6]. The AAC format is used in Apple's iPod and iTunes and in Sony's PSP and PS3. AAC is better at compressing audio, especially at bit rates lower than 128 kbps and provides more flexibility than MP3 (e.g. Sample rates up to 192 kHz, up to 48 channels and more encoder implementation options). The popularity of MP3 remains strong which is in part due to the absence of DRM (Digital Rights Management) in MP3 as opposed to many AAC implementations.

MP3 internals

MP3 uses perceptual coding, based on knowledge from psychoacoustics, to achieve inaudible compression [7]. MPEG-1 details the MP3 file format and the decoder, but leaves a lot of room for different encoder implementations. Each of these encoders will use different perceptual coding algorithms. These perceptual coding algorithms make use of some of the limitations of human hearing such as temporal masking and frequency masking. Block diagrams of an MP3 encoder and decoder are depicted in Figure 1.

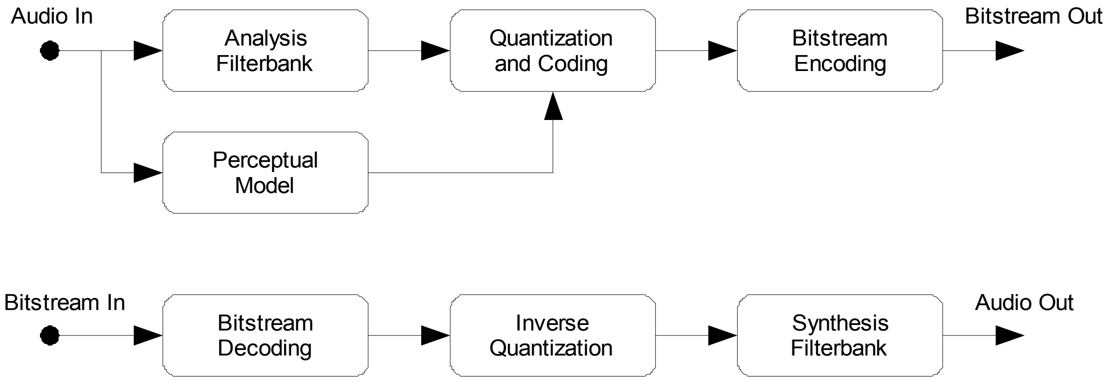

Figure 1.- Block diagram of MP3 encoder and decoder.

The analysis filterbank decomposes the input signal into spectral components using a modified discrete cosine transform (MDCT). These spectral components are then quantized with input from the perceptual model to decide the quantization threshold and coarseness. The quantized data is then compressed by a Huffman coder which assigns shorter codes to more frequently appearing values. Finally, the coded signal is packed into a standard format as specified by MPEG-1.

The course quantization of more or less irrelevant spectral components is what causes the algorithm to lose information. Insofar as this lost information only pertains to that part of the signal that is masked or otherwise below the hearing threshold, it should not affect our perception of the decoded audio. The higher the available MP3 bit rate, the fewer spectral components have to be discarded or coarsely quantized, and the better the decoded signal will approximate the unencoded signal.

With this knowledge about the MP3 encoding process, it does not seem far fetched to assume that insofar room acoustic parameters relate to audible qualities of a room, these parameters should not be influenced by the absence of otherwise imperceptible signals.

IMPACT ON ROOM ACOUSTIC PARAMETERS

Stimulus selection

To gain insight into the effect of using different coders and bit rates, several tests were performed whereby PCM coded stimuli were converted to MP3 format. These MP3 coded signals were then converted to PCM signals. The original signal was used to deconvolve the decoded MP3 signal. The result of this deconvolution is the impulse response of the MP3 coder. The impulse responses were evaluated using the INR [8] as a quality measure.
No significant difference was found between different encoders (e.g. Fraunhofer and Lame), and subsequent tests were performed with the Lame [9] encoder. Figure 2 shows the relation between the bit rate and the INR.

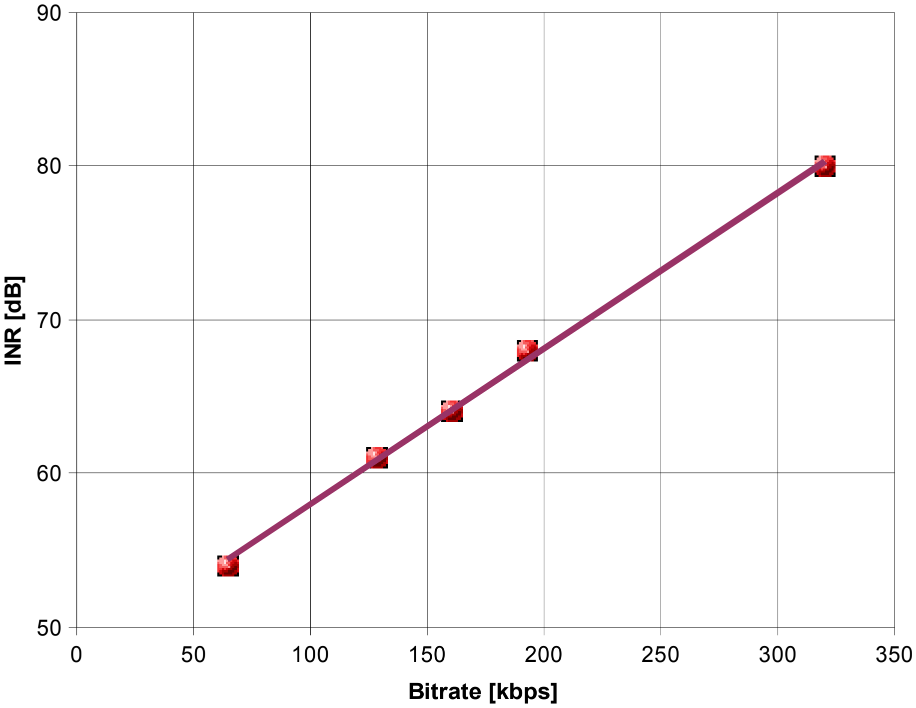

Figure 2. - INR vs. MP3 bitrate in the 1 kHz octave band.

As would be expected, the INR improves with increasing bit rate. Due to the linear relationship between the INR and the bit rate there is no 'sweet point' where further increases in the bit rate would not result in corresponding increases in the INR. The optimum bit rate with respect to the INR is therefore simply the maximum bit rate allowed in the MP3 format, which is 320 kbps.
In practical field measurements with PCM coded stimuli, the obtained INR values are most often lower than 60 dB. One could therefore argue that a bit rate resulting in an INR higher than 60 dB in a loopback measurement should have a negligible negative impact on practical INR values.

Further experiments are based on a 160 kbps bit rate. This speed is supported by all MP3 players, has a loopback INR well above 60 dB and provides a reasonable compression factor of 4.41 compared to uncompressed 16-bit PCM at 44.1 kHz.

At this point it should be noted that MP3 encoding limits the signal frequency range to a bandwidth of approximately 19 kHz, because higher frequencies are inaudible an therefore suppressed by the perceptual model. This means that the stimulus no longer excites the full 16 kHz octave band. Parameter values can therefore only be calculated up to the 8 kHz octave band.

Measurements

To evaluate the impact of MP3 encoded stimuli on the measured room acoustic parameters, measurements were performed in the concert hall of “Muziek Centrum Eindhoven” with a volume of approx. 14400 m^3 and $T_{empty} \approx 2$ s.

The measurement equipment consisted of the following components:

- *signal source:* DIRAC 4.0 (B&K/Acoustics Engineering Type BZ5449) on laptop PC 1;
- *output:* USB audio device 1 (Acoustics Engineering - Triton);
- *power amplifier:* (Acoustics Engineering - Amphion);
- *sound source:* omni-directional (B&K Type 4292);
- *microphone:* omni-directional, sound level meter (Rion - NL21);
- *input:* USB audio device 2 (Acoustics Engineering - Triton);
- *receiver:* DIRAC 4.0 (B&K/Acoustics Engineering Type BZ5449) on laptop PC 2.

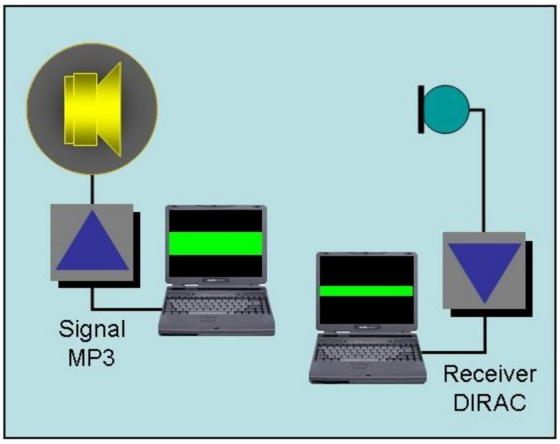

Figure 3.- Measurement setup.

Exponential sweep (e-sweep) and MLS stimuli were used. The 16-bit 44.1 kHz stimuli were encoded by the Lame MP3 encoder at a fixed bit rate of 160 kbps. For reference, uncompressed stimuli measurements were performed. The differences in the calculated parameters between the uncompressed stimuli measurements and the MP3 stimuli measurements are compared with the JND values for the specific parameters. The JND values are partly taken from the ISO/DIS 3382-1:2006 draft [10], and partly based on estimations (these values are marked by an asterisk *). The measured maximum octave band differences (over 125 Hz to 4000 Hz) in calculated parameters based on measurements with MP3 vs. uncompressed stimuli are given in Table 1. The stage parameters (ST_{early} and ST_{late}) are measured at a distance of 1 m from the sound source, whereas the other parameters are measured at a listener position 10 m from the sound source.

Table 1.- Parameter value differences between MP3 and uncompressed stimuli.

Parameter	**Symbol**	**Measured difference MLS**	**Measured difference e-Sweep**	**JND**	**Unit**
Reverberation Time	T_{20}	< 1	< 1	**10***	**%**
Centre Time	T_S	< 2	< 0.5	**10**	**ms**
Clarity	C_{80}	< 0.1	< 0.1	**1**	**dB**
Definition	D_{50}	< 0.005	< 0.005	**0.05**	**-**
Early Support	ST_{early}	< 0.1	< 0.1	**1***	**dB**
Late Support	ST_{late}	< 0.2	< 0.1	**1***	**dB**
Speech Transmission Index	STI	< 0.005	< 0.005	**0.1***	**-**

*estimated

It was seen that preaveraging of the measurements resulted in a 3 dB improvement of the INR for every doubling of the preaveraging. This indicates that the INR is not limited by the MP3 compression.

CONCLUSIONS

- From the measurements it can be concluded that the differences between room acoustic parameter values measured using uncompressed PCM stimuli and using MP3 encoded stimuli are well below the parameter JND's.
- MP3 stimuli allow for increasing the INR by 3 dB for each doubling of the preaverage count.
- Bearing in mind the limited frequency range, MP3 encoded stimuli are a useful option when measuring room acoustic parameters.

REFERENCES

[1] M.R. Schroeder: Integrated-impulse method for measuring sound decay without using impulses. Journal of the Acoustical Society of America **66** (1979) 497-500
[2] **ISO 18233:2006** Acoustics - Application of new measurement methods in building and room acoustics
[3] D.D. Rife, J. Vanderkooy: Transfer-function measurements with maximum length sequences. Journal of the Audio Engineering Society **37** (1989) 419–444
[4] S. Müller, P. Massarani: Transfer-function measurements with sweeps. Journal of the Audio Engineering Society **49, No.6** (2001) 443–471
[5] **ISO/IEC 11172-3:1993** Information technology - Coding of moving pictures and associated audio for digital storage media at up to about 1,5 Mbit/s -- Part 3: Audio
[6] **ISO/IEC 13818-7:2006** Information technology -- Generic coding of moving pictures and associated audio information -- Part 7: Advanced Audio Coding (AAC)
[7] Karlheinz Brandenburg: MP3 and AAC Explained. Proceedings of the 17th AES International Conference: High Quality Audio Coding (August 1999)
[8] Acoustics Engineering: Impulse response to Noise Ratio INR. Technical Note 007, http://www.acoustics-engineering.com/files/TN007.pdf
[9] The LAME Project: http://lame.sourceforge.net/index.php
[10] **ISO/DIS 3382-1:2006 Draft** Acoustics - Measurement of room acoustic parameters – Part 1: Performance rooms