Audio Engineering Society
Convention Paper

Presented at the 122nd Convention
2007 May 5–8 Vienna, Austria

Challenges of MP3 Player Testing

Steve Temme[1], Pascal Brunet [2], and Zachary Rimkunas[3]

[1] Listen, Inc., Boston, MA, 02118, USA
stemme@listeninc.com

[2] pbrunet@listeninc.com

[3] zrimkunas@listeninc.com

ABSTRACT

MP3 player audio performance is discussed including measurements of frequency response, phase response, crosstalk, distortion, sampling rate errors, jitter, and maximum sound pressure level with headphones. In order to make these measurements, several measurement techniques and algorithms are presented to overcome some of the challenges of testing MP3 players. We discuss test equipment requirements, selection of test signals and the effects of the encoding on these test signals. A new method for measuring non-coherent distortion using any test signal including music is also presented.

1. INTRODUCTION

MP3 players have been the 'must-have' electronic gadget for the past few years. Over 10 million players were sold in 2005, and this number is predicted to more than double by 2010. But how can manufacturers carry out QA tests on the production line, ensure excellent sound quality, and demonstrate their compliance with Sound Pressure Level regulations?

MP3 player testing is challenging as it combines traditional acoustic analysis techniques with some characteristics unique to MP3 players. Here, we examine the equipment and techniques that MP3 player manufacturers can use to test the sound quality of their products, and discuss the measurement methods and algorithms that can be used to overcome the challenges inherent to measuring MP3 players.

2. WHAT NEEDS TO BE MEASURED?

There are several different aspects of MP3 players that need to be tested. The most obvious is the audio quality – does the signal coming out of the MP3 player sound the same as the signal going into it? It is also becoming increasingly important to measure Maximum Sound Pressure Level (SPL) as this is now required in some European countries. Finally, to ensure a good user experience, the complete system including headphones and other ancillary components needs to be tested.

3. CHALLENGES OF MP3 PLAYER TESTING

Measuring MP3 players presents many challenges. MP3 files are by their nature compressed, which makes accurate measurements difficult, particularly at the extremes of the frequency spectrum. The wide dynamic range places some constraints on the hardware that can be used. In addition, synchronization of the signal playback and measurement, jitter, and sampling rate errors result in the necessity for specially designed measurement algorithms. In addition, manufacturers may need to test the complete system including headphones and possibly even a mobile telephone. These additional components require their own set of tests.

4. MEASUREMENT EQUIPMENT

The hardware is an important factor in MP3 player testing. MP3 players have a wide dynamic range as they use at least a 16 bit D/A converter. This means that any measurement system used must offer a dynamic range of at least 96dB. For software based systems, a high end sound card or data acquisition card is necessary. We have achieved excellent results using DAL sound cards (104dB), Lynx Soundcards (110dB) and NI PXI 4464 data acquisition cards (117dB). Standard built-in PC or laptop soundcards are usually around 80dB and do not offer sufficient sensitivity for MP3 player testing. If using a data acquisition card, it is important to ensure that it offers anti-aliasing filters for spectrum analysis. Whatever input device you use, care must be taken to ensure that the playback and acquisition are at similar levels in order to utilize the entire dynamic range. This is usually done using input attenuators on the signal.

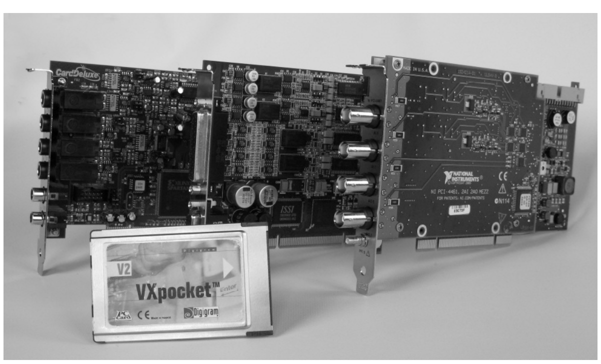

Figure 1 Selection of sound cards suitable for MP3 Player testing

Since an MP3 player has a stereo output, the fastest way of measuring it is with a system that offers at least 2 channels (a single channel system will work, but will take longer). A system with 4 channels will enable headphone output and line out to be simultaneously measured, and a six-channel system would enable the complete system (both outputs and the acoustic response of the headphones) to be measured in a single test.

There are both hardware and software based systems on the market, and even systems that are a combination of the two. Software systems generally offer considerably more flexibility and a lower initial purchase price than hardware systems. Since they use the processing power of the computer, IT upgrades can make your system run faster whereas a hardware based system will offer the same performance over its entire life. A Windows software based system offers the added advantage that it works directly with WAV files, thus avoiding the additional step (and possible introduction of errors / noise) of having to convert the test signal from the hardware system's native format. The main advantage of hardware-based systems is that they can offer extremely high test accuracy due to the data acquisition device. This can be important for testing some high-precision audio electronic devices but in our experience, a professional sound card is more than adequate for testing MP3 players.

Another important thing to consider when selecting a test platform is speed. Will the system be used on a production line and what is the necessary throughput? Systems vary in the time taken to run a test, and in their ability to be integrated with automated production lines, offer simple pass/fail results, etc. If high throughput is important, a multi-channel system, although it has a higher initial purchase price, may prove less expensive in the long run as more parameters can be measured simultaneously, enabling faster testing. If it is important to you to use the same system in R&D as on the production line to facilitate test development, make sure that it also offers the test design flexibility that product designers will need.

5. TEST SIGNAL

MP3 players can be tested using WAV files (preferred) or MP3 files which restrict somewhat the accuracy of the tests.

MP3 (or, to give it its full name, MPEG1/2-Layer 3) is a lossy compression algorithm as well as a file format. It was the result of European research (as part of the MPEG group) around 1990 and it became an ISO/IEC standard in 1993. It is able to greatly reduce the size of an audio file while preserving the sound quality of the content. An MP3 file will be typically 10 times smaller than a WAV file for the same content. For example, a 16 bit stereo WAV file at 44,100 Hz is about 1400 kbits per second, whereas a good quality MP3 can be achieved with only 128kb/s. This compression is achieved mainly by applying psychoacoustic models on both temporal and frequency analysis of the signal to remove the audio features that are not audible. For example, MP3 encoding uses extensive frequency masking to remove weak tones that are inaudible when strong tones are present (see Figure 2). Time domain masking effects are also used. Good explanations of the techniques used can be found in [1, 3].

The quality of the result depends on the quality of the encoder. Because the encoding algorithm is not fully specified by the standard, and because a high level of acoustic expertise is required to implement it, all encoders are not equal in terms of quality and performance.

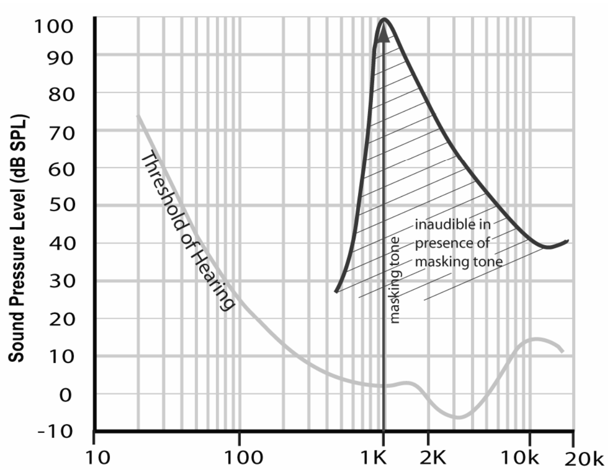

Figure 2 Masking curve used to eliminate 'unheard' sounds

Table 1 shows how the distortion + noise and the frequency range change with the chosen encoding. It compares commonly used compression algorithms (at 2 different bit rates) to a WAV file. AAC (which is part of the MPEG2 standard) and WMA are technical successors to MP3. They provide higher compression than MP3 for the same quality (e.g. 96 kb/s instead of 128 kb/s) and a greater number of sampling rates to choose from... Other common types of sound defects found also in MP3 files are time-frequency varying errors and pre-echoes.

Effects of Encoding on a Multitone Stimuli			
Format	Bit Rate (kbits/s)	-3 dB Freq Range (kHz)	Distortion+ Noise (%)
WAV	1411	20	.002
MP3	96	15.8	39.9
MP3	320	20	8.4
AAC	96	15	21.9
AAC	320	19.8	4.6
WMA	96	15.4	14.6
WMA	192	18.6	6.2

Table 1 Frequency range and distortion vs. Encoding

These numbers look poor compared to standard hi-fi audio performances. Even with greater distortion + noise, an MP3 encoded signal does not sound significantly degraded because, all the frequency responses are flat within the frequency range, and the noise is carefully distributed along the frequencies as to be inaudible.

In Figure 3 a 16 bit WAV file has, as expected, an absolutely flat frequency response and the noise is about -140 dB. As a matter of comparison MP3 encoding, with a bit rate of 320kb/s still ensures a flat frequency response but has a far worse signal-to-noise ratio (about -75 dB). However, at a bit rate of 96kb/s, the frequency response is compromised and the noise increases (about -60 dB). The WAV format is therefore the preferred format for testing an MP3 player. If the MP3 player being tested cannot play a WAV file and the test signal has to be encoded as an MP3 file, the highest available bit rate is preferred. It is clear that 96kb/s rate contains too much noise to use as a test signal for measuring distortion.

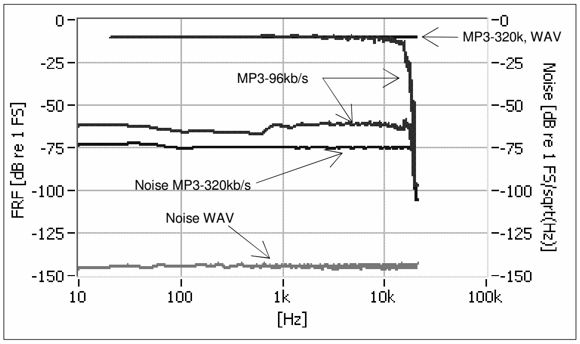

Figure 3 Frequency Response and Noise for MP3-96kb/s, MP3-320kb/s, WAV

The errors are only on the encoding side. The decoders, as long as they are compliant to the standard, have all the same performances [1].

For this paper it was decided to use primarily two different types of signals for testing an MP3 player, Sine Sweep and Multitone.

Sine sweep, the traditional signal for audio measurement, is slow but precise. It is not sensitive to sampling rate shift when sweeping from high to low frequencies, and it can be encoded to MP3 format without much damage: encoding at 320 kb/s yields 0.1 % THD+N. The disadvantage of using a sine wave is that it is very different from music or voice making it questionable how well the distortion measurements relate to human hearing.

Multitone is a fast test signal and therefore it is well suited for production test. The sampling rate shift is easy to correct, and statistically it is closer to music than sine waves, making the distortion measurement more realistic. Multitone also allows for easy measurement of cross-talk. However, MP3 encoding heavily distorts a multitone: even a 320 kb/s encoding yields about 9% of distortion + noise.

Noise (pink or white) is not suitable for testing MP3 players, although it is statistically very close to music, the sampling rate error cannot be measured and corrected with noise alone.

This means that if you are testing an MP3 player that can play WAV files, a multitone is the best test signal to use, and if your MP3 player can only play MP3 files, it is best to use a sine sweep encoded with the highest available bit rate.

6. MEASUREMENT DETAILS AND ALGORITHM DESIGN

6.1. Measurement Principle

For sound quality assessment, the linear distortion (frequency response amplitude and phase) and the non-linear distortion need to be measured. The cross-talk between left and right channels also needs to be measured.

The principle is first to create a stereo stimulus waveform and copy it on the MP3 Player to be tested.

This waveform is then played back, and the signal acquired and analyzed. The Left and Right channels are acquired simultaneously, using a trigger on one channel to start the measurement.

For a multitone WAV stimulus, the analysis compares the genuine stimulus and the playback response, using a cross-spectrum analysis that provides the frequency response, cross-talk and non-coherent distortion.

For a stepped sine sweep MP3 stimulus, a serial FFT analysis is used.

SoundCheck^{TM}, a software-based acoustic measurement and analysis system with a National Instruments 4461 Data Acquisition card was used for all measurements in this paper.

6.2. Sampling Rate Errors

Frequency shift and jitter are the effects of an unstable or skewed playback sampling rate. Frequency shift occurs when the playback sampling rate is offset from the desired sampling rate. For example, an MP3 file that is sampled at 44.1 kHz would actually be played back at a slightly higher or lower rate (44.09 kHz or 44.11 kHz). This causes every frequency in the waveform to be shifted as in Figure 4. The tones in the response spectrum are slightly higher than those in the stimulus spectrum.

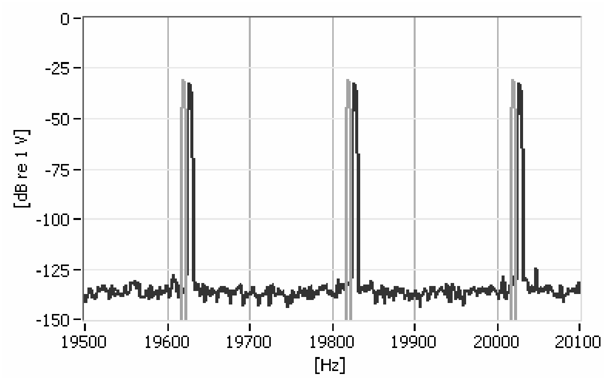

Figure 4 Response Spectrum (dark) vs. Stimulus Spectrum (light)

Jitter occurs when the sampling rate fluctuates with time. An example of jitter is shown in Figure 5. The sampling rate fluctuates around 44101.325 Hz.

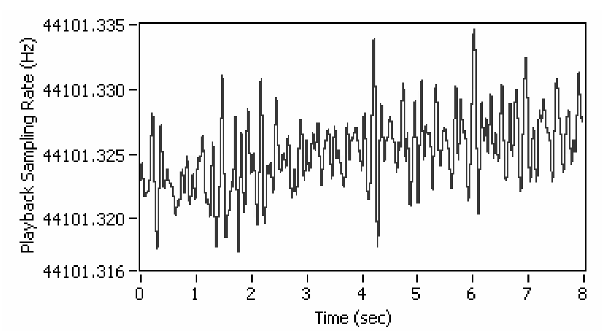

Figure 5 Jitter - Instantaneous Sampling Rate vs. Time

Sampling rate problems are caused by the relatively inexpensive crystals used in low cost consumer goods such as MP3 players. While these problems are not usually noticeable by the user, they are important to the manufacturer. More importantly, these characteristics negatively affect the quality of measurements such as frequency response and distortion.

Frequency shift negatively affects the coherence of the output versus input spectrum (Figure 6). Note: In Figure 6, the coherence was calculated only at the frequencies present in the stimulus waveform. The logarithmic frequency spacing and windowing used in the spectral analysis caused the tones to appear with different widths.

Measurement algorithms need to be designed to restore the coherence so that accurate frequency response and distortion measurements can be made. Jitter is interesting to the manufacturer but has negligible effect on the measurement quality.

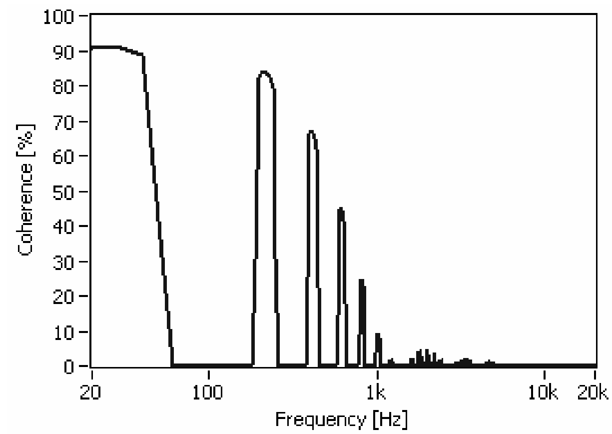

Figure 6 Coherence between Multitone stimulus and response before the frequency shift algorithm

This section will discuss the algorithm that has been designed to overcome the measurement challenge imposed by frequency shift.

6.2.1. Frequency Shift Algorithm

The goal of our algorithm was to 1) quantify and 2) correct the frequency shift in the response waveform. The algorithm contained a phase vocoder that very accurately quantified the frequency shift. The frequency shift was then corrected by adjusting the sampling interval (called dt) and resampling the response waveform.

Using the phase vocoder required prior knowledge of a sinusoidal tone (called f_i) that existed in the stimulus waveform. It was expected that in the response waveform there would be a tone slightly shifted from f_i (less than 1%). Determining the frequency shift of this one tone allowed the calculation of the playback sampling rate.

To quantify the frequency shift, a complex heterodyne filter was applied to the response waveform. The filter shifted f_i to DC and filtered out all other tones:

$$y(n) = \sum_{k=0}^{M-1} x(n-k) \cdot W(k) \cdot e^{j2\pi f_i k} \qquad (1)$$

Where M was the FIR (finite impulse response) size and W was a time weighting window.

The result of the heterodyne filter, y(n), was complex and contained the phase information necessary to estimate the frequency shift. The segment of y that corresponded to the steady state output of the heterodyne filter was isolated by applying a thresholding algorithm. This was done to limit analysis to a segment of y that yielded a stable phase. It was important to avoid analyzing the phase of noise. The segment of y returned by the thresholding algorithm was called y'.

Next, the unwrapped phase of y' was calculated. This resulted in a curve representing the instantaneous phase with respect to time, $\varphi(n)$. The instantaneous frequency curve of the tone at f_i was then obtained by calculating the slope of $\varphi(n)$:

$$f_{inst}(n) = \frac{1}{2\pi}\frac{d\phi(n)}{dn} \quad (2)$$

The jitter curve was then obtained by scaling f_{inst} to lie about the playback sampling rate:

$$j(n) = \frac{f_{inst}(n) \cdot f_s}{f_i} \quad (3)$$

Where f_s was the stimulus sampling rate.

The playback sampling rate (called f_s') was then obtained by calculating the mean of j(n). With the playback sampling rate known, it was possible to correct the frequency shift. First, the dt of the response waveform was replaced with the playback sampling interval dt':

$$dt' = \frac{1}{f_s'} \quad (4)$$

Changing the sampling interval to dt' caused the duration of the stimulus and response waveforms to be identical. The response waveform was then resampled to the stimulus sampling rate, f_s. The result was a response waveform with the same duration and sampling rate as the stimulus waveform. The corrected response waveform exhibited no frequency shift and excellent coherence with the stimulus, as shown in Figure 7.

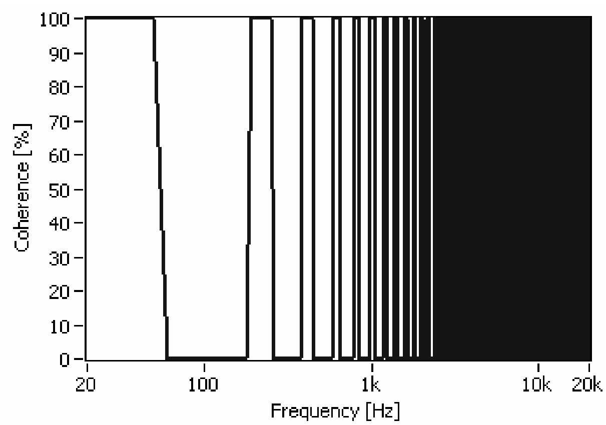

Figure 7 Coherence between Multitone stimulus and response after the frequency shift algorithm

The frequency shift algorithm was most reliable when using a Multitone stimulus. It was possible to apply the algorithm to a Stweep stimulus with less reliable results. The reliability was degraded because the duration of individual tones was shorter in the Stweep stimulus than in the Multitone. The playback frequency estimation was less accurate due to the shorter duration.

The frequency shift algorithm allowed for the testing of devices with skewed playback sampling rates. Without correcting for frequency shift, testing of such devices would be inaccurate.

6.3. Non-Linear Distortion

Unlike linear measurements, non-linear distortion measurements depend heavily on the amplitude distribution and spectral content of the stimulus. Traditional harmonic distortion measurements using a single test tone or sweeping tone are easy to calibrate and perform. However, they do not reveal intermodulation products and are not at all similar to music. In fact sine distortion is a poor predictor of the distortion that will occur on music.

Multitone is a popular test signal for fast frequency response measurements. It is also a more rigorous test signal for assessing system non-linearities because it excites many frequencies simultaneously and produces both harmonic and intermodulation distortion products. Statistically, it is closer to music than a sine wave or two tones.

To measure distortion + noise on multitone, a Non-Coherent Distortion metric is used:

$$\eta^2(\omega) = \frac{(1-\gamma^2_{(\omega)})\cdot G_{YY}(\omega)}{\sum_{\omega} G_{YY}(\omega)} = \frac{G_{NN}(\omega)}{\sum_{\omega} G_{YY}(\omega)} \tag{5}$$

Where: γ^2 is the coherence, G_{YY} is the output power spectrum and G_{NN} is Non-Coherent Power Spectrum. G_{NN} is the distortion + noise added to the output signal.

The Non-Coherent Distortion (NCD) is the Non-Coherent Power Spectrum normalized against the Total Output Power. Summing against frequency results in the ratio of the Total Non-Coherent Power to the Total Output Power.

$$\lambda = \sqrt{\sum_{\omega} \eta^2(\omega)} = \sqrt{\frac{\sum_{\omega} G_{NN}(\omega)}{\sum_{\omega} G_{YY}(\omega)}} \tag{6}$$

The Total Non-Coherent Distortion (TNCD) λ is an extension of the THD+N for a broadband signal. It can also be expressed in % . TNCD can be applied to a multitone as well as noise or music with consistent results. This method for TNCD measurement was introduced in a previous AES paper [2], which explains the theory and results in more detail.

As shown in Figure 8, the NCD spectrum obtained with a musical excerpt (Techno music) played on an MP3 player is close to the NCD obtained with encoded Pink Noise. The differences are due to the different spectral contents and levels.

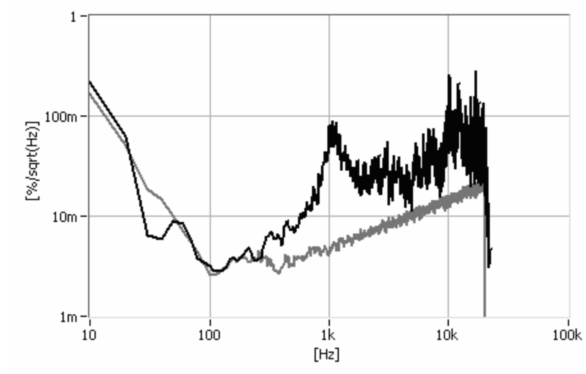

Figure 8 NCD Spectrum for Music (top) vs. Pink Noise (bottom)

In Figure 9, the THD+N curve and NCD spectrum both measured from an MP3 player is shown. The THDN was measured from a sine sweep and the NCD from a multitone. Both were at the same peak level. Although the THD+N seems higher, the Total NCD is in fact 0.07%, which, when compared to the average THD+N, is about 7 times higher.

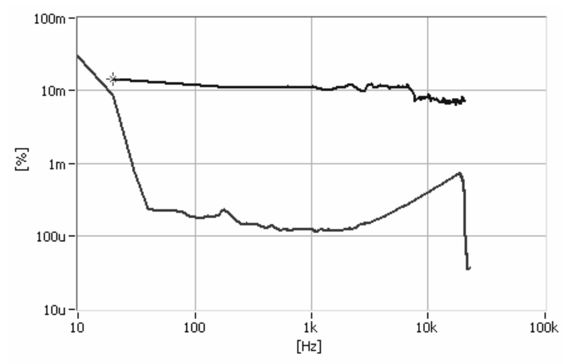

Figure 9 THD+N curve (top) vs. NCD spectrum (bottom). TNCD is 0.07%.

6.4. Cross-Talk

The cross-talk between channels is measured by simultaneously playing two different multitones for Left and Right channels with interleaved frequencies (see Figure 10), and making a cross-spectral analysis between channels (e.g. finding the contribution of the left stimulus in the right response).

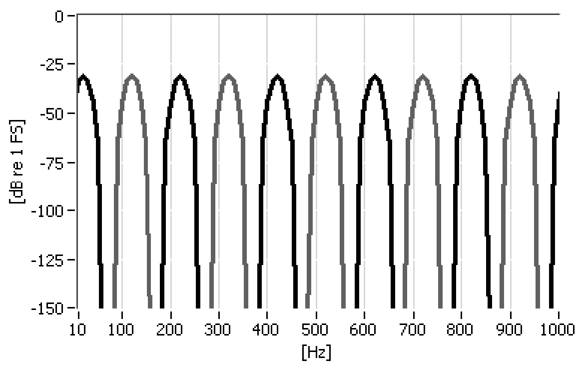

Figure 10 Stimulus Spectrum Left (light) and Right (dark)

Figure 11 shows a typical result. The flat frequency response is fairly typical of an MP3 player.

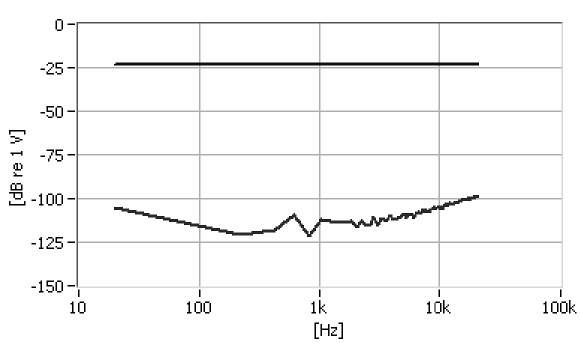

Figure 11 Frequency Response (top curve) and Cross-talk curve (bottom curve)

7. SYSTEM TEST

An MP3 player is usually part of a system which includes headphones. Generally speaking, the quality of the headphones is the limiting factor as headphones typically have much worse frequency response and distortion than the electronics.

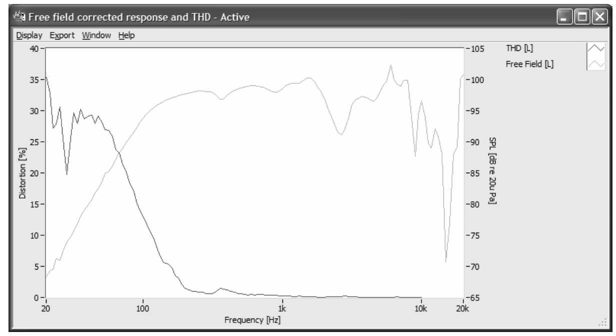

Figure 12 Headphones typically have much worse frequency response (light) and distortion (dark) than the MP3 player electronics.

Headphone testing has its own measurement techniques and industry standards that need to be considered. These include what kind of artificial ear to use, how exactly the headphones are positioned on the artificial ear, and how to apply the free field correction curve to the results. This is a detailed area for discussion, and a subject for a future paper.

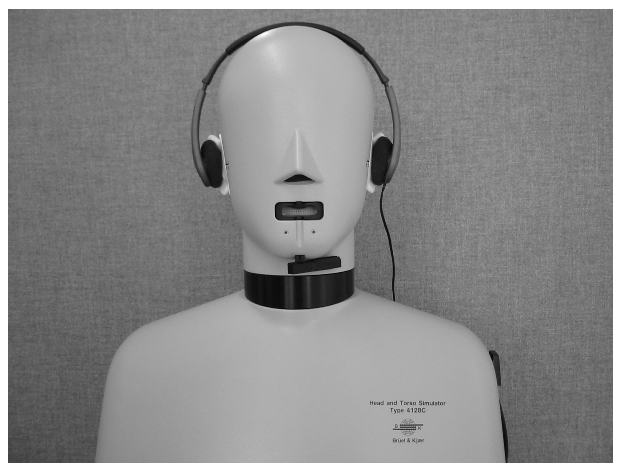

Figure 13 Most people listen to MP3 players through headphones so it makes sense to measure the headphones as well

In addition to testing the complete system for sound quality, it is also necessary to test the maximum SPL (Sound Pressure Level). MP3 players can be played at very high sound pressure levels (some over 110dB) which, together with the tendency of people to listen to MP3 players for long periods of time, can lead to high sound exposure levels. In France there is a law restricting portable audio devices to a maximum SPL of 100dB, and in the US at least one lawsuit has been filed against an MP3 player manufacturer for hearing damage caused by an MP3 player.

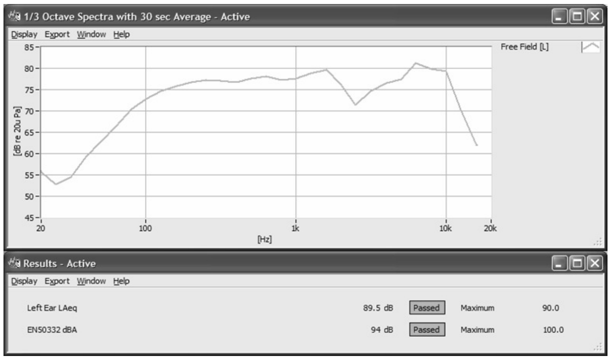

Figure 14 MP3 player tested according to British Standard BS EN 50332-1.

The industry is not yet agreed on what is an acceptable maximum sound pressure level (SPL) and sound exposure level (SEL). So far the only known standard for measurement is the British standard "BS EN 50332 Headphones and earphones associated with portable audio equipment – Maximum sound pressure level measurement methodology and limit considerations". Although it is more specifically a headphone test requirement, it applies to portable audio equipment so MP3 player manufacturers are wise to ensure compliance. Figure 15 shows the test set-up that recommended by this standard. This requires additional hardware such as a head and torso simulator (usually an expensive system component, see figure 13) which is important to also consider when specifying an MP3 player test system.

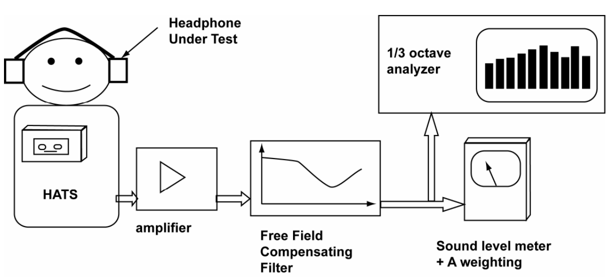

Figure 15 Test equipment requirements according to BS EN 50332

7.1. Impedance Measurement

The output impedance of an MP3 player influences the efficiency and frequency response of the connected headphones. This section describes a method for measuring the output impedance of the headphone and line out outputs.

The first step was to directly measure the RMS output voltage of the MP3 player. A pure sinusoid was played and the output voltage was measured with a 1MΩ input impedance voltmeter. The result (V_{out}) was the output voltage with no load on the device.

A resistor of known impedance (e.g. 100Ω) (z_r) was then placed in parallel with the MP3 player output.

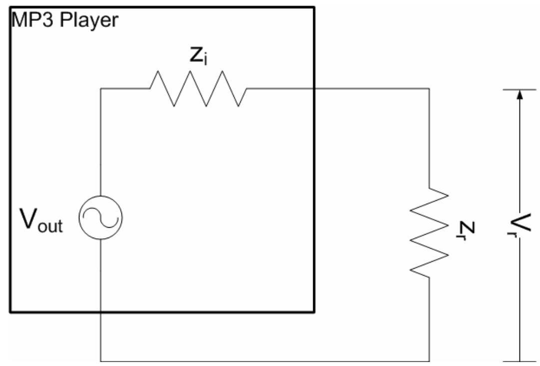

Figure 16 Circuit diagram of impedance measurement

The sinusoid was again played from the MP3 player and the voltage drop over z_r (V_r) was measured. Knowing V_{out}, V_{in} and z_r permitted calculation of z_i using a voltage divider:

$$z_i = z_r \cdot \left(\frac{V_{out}}{V_r} - 1 \right) \tag{7}$$

A common MP3 player was tested and the output impedances were:

- z_i = 5Ω for the headphone output
- z_i = 244Ω for the line out

8. CONCLUSIONS

MP3 player testing presents many challenges because it requires fast throughput, accurate testing and manufacturers need to be able to carry out a wide range of tests which may also include testing headphones and mobile telephones. A thorough understanding of how an MP3 player works and the characteristics that MP3 players display (jitter, sampling rate errors, etc.) is necessary to develop tests that quantify these errors. However, with careful test system selection, and awareness of the testing challenges, it is possible to accurately characterize the performance of MP3 players and the associated system components.

9. REFERENCES

[1] Karlheinz Brandenburg, "MP3 and AAC Explained", presented at AES17th International Conference on High-Quality Audio Coding, August 1999

[2] Steve Temme and Pascal Brunet, "A New Method for Measuring Distortion using a Multitone Stimulus and Non-Coherence", presented at AES 121st Convention, October 2006

[3] Richard J. Beaton et al., "Objective Perceptual Measurement of Audio Quality", AES Collected Papers on Digital Audio Bit-Rate Reduction, pp 126-152, May 1996

[4] J.L Flanagan and R.M. Golden, "Phase Vocoder", Bell System Technical Journal, July 1966

[5] Michael Betser et al., "Review and Discussion on Classical STFT-Based Frequency Estimators", Presented at AES 120th Convention, May 2006