The Numerically-Identical CD Mystery: A Study in Perception versus Measurement

Ian Dennis and Julian Dunn
Prism Sound Ltd.
Cambridge, UK

Doug Carson
DCA
Cushing, Oklahoma, USA

The authors have for some time been involved in an investigation into claims that numerically-identical CDs can sound noticeably different on replay. It has been suggested that these differences are attributable to variables in certain processes in the CD production chain. The work has centred on the controlled pressing of a set of test CDs encompassing the suspected variations. The discs contain test signals as well as music, so that listening test results can be compared with objective measurements from the same physical discs. This paper describes efforts so far in comparing listening test results with 'conventional' audio measurements. The work is on-going.

INTRODUCTION

The work described in this paper was begun early in 1996 as a result of reports, which abounded at that time, of different CD pressings of the same material exhibiting markedly different sound qualities on replay. An interim report on the progress of the work was published in November 1996 [7].

The usual scenario was that pre-mastering engineers or their clients would notice that a particular CD pressing would sound different from the original 'master tape' in the studio. Often, the same material would be re-pressed (perhaps in a different plant) and would sound better.

The affected discs reportedly had been verified to contain identical data to the master. However, it soon became clear that it was difficult to establish whether the comparison had been performed 'scientifically' in many cases.

There were a number of specific questions which needed to be answered in order to further our understanding of the problem, and this proved difficult - a lot of the evidence was more than second-hand, and many of the tales were probably apocryphal. It was therefore determined to go 'back to square one' - to take nothing on trust!

SOMETHING NASTY IN THE PRESSING PLANT?

It was suspected among pre-mastering engineers that certain processes in the CD mastering and replication chain were responsible for the sonic degradation. A brief summary of the CD mastering process is necessary to explain this.

[Note: in this paper, the final studio preparation of the recording is referred to as 'pre-mastering' in order to distinguish it from the preparation of the glass master by the CD manufacturer, who refers to this process as 'mastering'.]

On receipt of the master recording from the pre-mastering studio, the pressing plant prepares a 'glass master' disc, from which all subsequent pressing dies are made. The glass master is made by a laser-beam recorder (LBR). The final digital data feeding the LBR is not precisely the PCM audio data from the pre-mastering studio; the PCM data is pre-encoded as 'EFM' (eight-fourteen modulation) data which contains clocking information to allow the pits of the final disc to be successfully tracked and read by the player. The EFM coding process is not strictly deterministic, i.e. the same PCM data can be represented by different EFM patterns; the selection of the most appropriate pattern is determined by an algorithm in the encoder.

Originally, glass mastering was performed in 'real-time', i.e. during preparation, the glass master was rotated at the same speed as the final CD would be played. More recently, the glass master can be produced at twice or four-times real-time, or even faster; this is clearly more economic and so the plants are trying to speed the process up as much as possible.

The disc production variable which has fallen most under the suspicion of the music industry is the glass mastering speed. It has become widely suspected that the faster speeds latterly employed may be responsible for the sonic inferiority of the final disc. The choice of LBR and EFM coding electronics are also suspected, but apparently to a lesser extent.

The choice of pre-mastering medium (i.e. the format on which the master data is carried to the CD plant), is also widely cited, but this may also be bound up with the glass mastering speed, since the traditional UMATIC (PCM1630) cassettes have to be transferred to the LBR in real time, whereas newer media such as recordable CD (CD-R) and 8mm EXABYTE DDP cartridges, which are suspected of engendering inferior sound, can be transferred faster.

Some are suspicious of the use of AES3 interfaces in the chain, which are claimed to 'sound worse' than the original SDIF interfaces, even where the final data is received intact. It is hard to offer an explanation for this, although SDIF interfaces have tended to be used specifically with UMATIC media, so this may actually be another side-effect, ultimately linked to the glass mastering speed. The fact that these variables are so often connected makes it difficult to distinguish cause and effect in normal production situations.

Perhaps because of the imagined causes, or perhaps because of the sound of the degradations, 'jitter' is often cited as the culprit. Experiments have been conducted whereby the EFM data is closely reclocked at the input to the LBR, as described in [1]. The authors do not know the results of these experiments.

POLITICS

The issue has become a bone of contention between the music production business and the CD replicators, as is well documented in [4]. The pressing plants have been accused by artists and producers of destroying their product, whereas replicators have taken the view that as long as the data on the disc is correct and the physical properties of the discs are within limits laid down in the 'red-book' standard (and nowadays they are normally substantially better) then their commitment has been met.

Particular CD productions have become renowned within the industry as the subjects of disputes, and several have had to be reprocessed at the insistence of artist or record company. There is a rumour that one of the larger record companies, which also controls replication facilities, has 'A-list' artists whose discs are mastered in real-time, whilst the rest are mastered at the faster, cheaper speed.

But the CD manufacturers are not so disinterested as the music business would have us believe: after all, they need to attract customers, and these rumours are potentially damaging to their business. Even if the replay system turns out to be the cause of the degradations, replicators currently need to press discs which are acceptable on current players.

REPRODUCTION BASICS

The most interesting question surely is: "How can the discs sound different if the numbers on them are the same?". It is indeed hard to offer a rational explanation, though many explanations have been offered. It seems from a technical standpoint that the fault must, by definition, lie in the reproduction system.

A common theory is that physical imperfections in the 'pits' of the CD cause timing variations in the recovery of the EFM data, which are carried through to the player's digital-to-analogue converter (DAC). This 'sampling jitter' could cause significant audible degradation, as has been documented in [2], [3].

However, this theory misunderstands the basic architecture of the player: the timing clock of the DAC does not commonly depend on the recovered EFM data timing, but is derived from a free-running and stable crystal oscillator. The data is buffered in memory between the disc-reading electronics and the DAC. This buffer is emptied at the crystal oscillator rate, and kept half-full by varying the rotation speed of the disc; therefore jitter in the recovered EFM data should not affect the clocking of the DAC.

Another suggestion is that sonic degradation is the result of uncorrectable data errors on replay. This seems initially unlikely, since uncorrectable errors are very infrequent (at least as evidenced by the VALID flag at the player's output), and the sonic degradation described by listeners is continuous in character.

Interestingly, usual descriptions of the degradation involve confusion of stereo image, loss of 'life' etc. which are often associated with spurious aharmonic low-level components such as may be produced by incorrect dithering, sampling jitter or overambitious perceptual coding.

Initially, the authors were suspicious that the affected discs may not, in fact, carry identical data to the original. We had already experience that a considerable number of supposedly-transparent items of digital studio equipment were not actually precisely transparent to audio samples. This is usually the result of gratuitous re-dithering, truncation, or processing not advertised as a property of the unit. This is especially common in older equipment designs.

A preliminary experiment was conducted wherein a set of analytical test signals was prepared on DAT, and submitted to pre-mastering studios for transfer onto CD-Rs. In some cases the results were not numerically identical, and this could be traced to one or more items of non-transparent equipment in the transfer chain. Most common seems to be the inability to defeat the re-dithering action of some digital audio workstations. In these cases, sample values are different every time audio is played off the workstation, although this should not adversely affect the sound quality.

AIMS OF THE INVESTIGATION

In consideration of the reports detailed above, the aims of the work were defined as follows:

1) To verify or refute claims of consistent sonic differences between numerically identical discs;

2) To relate any consistent sonic differences to processes in the CD manufacturing cycle;

3) To relate any consistent sonic differences to objective measurements of test signals on the same discs;

4) To relate any consistent sonic differences to particular types of replay system;

TEST DISC SET TD-2

It was decided that, to achieve correlated measurements and listening test results, a 'custom set' of numerically-identical discs would have to be prepared incorporating analytical measurement signals as well as musical excerpts for listening tests. Each disc in the set would be manufactured using a different set of key variables.

The variables were:

Pre-mastering medium:	UMATIC, Exabyte DDP, CD-R
EFM coder:	Four manufacturers: V,W,X,Y
Laser-beam recorder:	Four manufacturers: V,W,X,Z
Glass-mastering speed:	1x, 2x real-time

It was also deemed desirable to keep the number of discs in the set as small as possible in order to facilitate listening tests, even though this prevented independent variation of all manufacturing variables. In fact, the disc set was limited to thirteen differently-manufactured (but numerically identical) discs.

The contents of the TD-2 disc are listed in figure 1.

Five musical excerpts were included covering a variety of genres. All the material had apparently been subject to adverse criticism when originally pressed. It was felt that this choice was desirable in case the sonic degradations were limited to particular types of music.

A variety of different analytical test signals were included, some of them being repeated near the beginning and end of the disc. These test signals were intended to allow investigation of different distortion phenomena.

The DSA-1 confidence test signal is a pseudo-random sequence that can be recognised by the Prism Sound DSA-1 analyzer. This signal was included to allow verification that no changes to the data had been made.

The J-test signal is described in [2]. It is primarily used to stimulate pathological jitter behaviour in digital audio interfaces, and is generally useful in revealing intersymbol interference (or 'data jitter').

The single and multiple tones were generated using the Prism Sound DScope system. They were designed to repeat over exactly 16384 samples so that a window-free FFT analysis could be conveniently performed using a frequency locked analysis system. This method allows the maximum frequency resolution to be achieved over the whole audio band.

The last two tracks contain values that do not allow an EFM modulator to maintain a zero DC content. These 'awkward' values were intended to show differences between EFM modulator algorithms used in the manufacturing process, but also proved useful as a source of DC modulation.

The TD-2 disc set represents a unique opportunity for comparison of controlled listening data with measurements of well-behaved analytical tests signals from the same physical discs.

LISTENING TEST STRATEGY

The principal decision to be made concerning the nature of the listening tests was whether to allow the subjects to listen using their own players in their own rooms, or whether to 'import' the listeners to a controlled replay system and environment.

It was decided to allow listeners to use their own replay systems, for a number of reasons:

1) It was desirable to use (at least) the listeners who had previously noticed the effect; these listeners were typically recording and pre-mastering engineers, who are busy people and are spread across the world.

2) It was suspected that the degradations would vary with different replay systems. By incorporating results from a wide range of replay systems, we might be able to understand which systems are more susceptible.

3) It was felt that since differences were likely to be (at best) small, a large number of results would be needed for analysis. This could be attained more quickly by circulating many sets of discs to the listeners' facilities.

Since the listening tests would cover a wide range of replay systems and environments, it was important that detailed information about the listeners' systems was gathered.

Another major issue was how best to ask respondents to define the relative qualities of the discs in the set. It was decided that one disc would be presented as the 'reference', and all the others in the set would be compared to it. Although the disc selected to be the reference was produced in the most 'traditional' fashion, there was no implied suggestion that it was the best sounding disc in the set. In fact, the whole idea of 'good' and 'bad' sound was eliminated from the questionnaire, which instead was couched in terms of the degree of difference from the reference disc.

It was clear that some numerical ranking system should be employed to allow convenient statistical analysis of the results. The listener was asked to place the discs in order of their difference from the reference. Discs identical to the reference are ranked '0', the most similar (but identifiably different) disc is ranked '1' etc. It was decided to use a 'forced' ranking scheme, where no tied results (other than '0's) are allowed. This is intended to improve result resolution (at the expense of increased 'noise') in a way which will be familiar to digital audio engineers.

The listening test questionnaire is shown in figure 2.

The disc numbers were scrambled so that each set was differently numbered. This was to reduce the possibility of distortion of results by conferring amongst listeners or by a consistent playing order. A duplicate copy of the reference disc was anonymously included in the set to allow another means of evaluating the result sets; thus the final disc set circulated amounted to 14 discs.

A number of sets of the test discs have been in circulation amongst critical listeners since autumn 1996. At the same time, laboratory analysis of the test signals has been performed using a variety of replay systems.

DATA INTEGRITY CHECKS

Laboratory measurements were performed using a variety of stand-alone ('one-box') players, and also systems with external DACs ('two-box' players).

Firstly, it was important to verify that the data on the pressed CDs was correct - early tests with CD-Rs had revealed that correct duplication of test data in pre-mastering facilities (i.e. without gratuitous processing, dithering or truncation) was difficult to guarantee, as described above. The DSA-1 confidence test showed that all discs were numerically identical to the master data.

In examining the reliability of the various players in playing the 'awkward' EFM patterns, it was initially assumed that replay errors could be monitored using the 'VALID' flag in the digital output bitstream of the players. However, this was not the case with all players; among those which failed to reproduce the patterns reliably, some did not set the 'VALID' flag at all and others did so only sometimes. The following table shows how many of the 13 different discs in the set were NOT reliably playable on various players on the awkward EFM tracks, and which players flagged these errors on their 'VALID' bit:

Player	No of discs	Flags V=1?
Philips CD624	2	Rarely
Philips CD930	1	Never
Marantz CD-63	13	Never
Technics SL-PS670A	4	Usually
Sony CDP-761E	3	Sometimes

Interestingly, there was quite a strong correlation between the problem discs and the LBR used in manufacture, although not with the EFM coder.

The failure of some players to play the awkward EFM patterns reliably was interesting, but was felt not to be relevant in explaining sonic differences, since sustained awkward patterns do not occur frequently in real music, and occasional uncorrected errors (even if they did occur) would not correspond with the sort of continuous degradation in sound quality which had been described. However, it was disappointing that the 'VALID' flag could not be relied upon, since it necessitated checking all discs in all players using the DSA-1 confidence test in order to assure general data integrity for all disc/player combinations. This test was found to be reliable in all cases.

MEASUREMENT OF SPURIAE

The main part of the laboratory tests involved identification of player-related and disc-related spuriae, and investigation of their causes.

Plots of the various experiments described above, made using the Prism Sound DScope analysis system, are shown in figures 3 through 12. For brevity, one-box systems are illustrated using a Marantz CD-63 player, and jitter-rejecting two-box systems are illustrated by addition of a Prism Sound DA-1 external DAC.

The 1kHz sine@-60dBFS track was used to investigate the noise floors of the various player/disc combinations. It was necessary to use this signal in preference to digital silence, since the DACs in many players effectively turn off in the absence of signal. The -60dBFS stimulus is generally of sufficiently low amplitude that modulation components do not obscure the noise floor itself.

Figure 3. shows the noise floor of a typical one-box player. Spurious components in this test are characteristic of each player, and appear to be the result of electrical interference adding to the player's output. The strongest in this case is a

component just below 5kHz. The group of components around 7350Hz are found in many one-box players, and are related to the EFM block rate which occurs at one sixth of the sample rate (fs/6). The completion of an EFM data block is a time of heightened electronic activity in the player, since data is usually transferred, and servo drive signals often updated, at this time.

Figure 4. shows the same test with an external DAC. Any spuriae here would normally be characteristic of similar additive electrical interference in the external DAC, since modulation components are excluded by the low signal level. In this case, the DAC exhibits no significant interference components.

Next, the J-test signal was used to investigate data-pattern induced jitter. At first, the results of this test in one-box players appeared to be very exciting, since discrete distortion sidebands, a few hundred Hz apart and rising towards the fs/4 stimulus, were observed. The J-test produces these sidebands where data-pattern-induced sampling jitter is present, as described in [3]. However, it was later noticed that the sidebands were typically 390Hz apart, rather than fs/192 apart (about 230Hz at fs=44.1kHz) which is the separation of the J-test sidebands. It was further observed that the same sidebands were present when the stimulus was a simple tone, rather than the intermodulating J-test.

Figure 5. shows the modulation components for a single fs/4 (11.025kHz) tone. The test signal is actually the odd-phases of the J-test signal, hence the -3dBFS amplitude. The 390Hz-separated sidebands are clearly visible. Note also the modulation sidebands at the EFM block rate (fs/6)and half this rate (fs/3). This result seemed initially to confirm what we had been told to expect, because these sideband components suggest classic sampling jitter, as described in [2].

Figure 6. shows the same test, but using an external DAC with stringent jitter rejection capabilities. The modulation sidebands are absent, at least down to very low frequencies. This would have been consistent with the sidebands being caused by clock jitter.

However, doubt was cast on the jitter theory by the absence of similar components when a non-jitter-rejecting external DAC was used. Possibly the jitter was on the clock of the player's DAC but not on its digital output; this seemed doubtful since most players use the same clock source for both.

To investigate the jitter theory further, a DC signal was applied, which was fortunately available on the TD-2 disc disguised as one of the 'awkward EFM' signals.

Figure 7. shows the result for a one-box player. The 390Hz-separated modulation components are still clearly visible, at a slightly higher level even than before. This result shows that the components cannot be the result of jitter, but suggests that they are caused instead by amplitude modulation. Amplitude modulation produces similar sidebands to the phase modulation action of sampling jitter, as described in [2]. Jitter sidebands increase with the frequency of the test stimulus and are absent for a DC stimulus; this is not the case with amplitude modulation, which affects low and high stimulus frequencies alike.

In some charge-transfer DAC architectures, jitter can cause combined amplitude *and* phase modulation; the amplitudes of the combined sidebands are related to sideband frequency, so the sidebands at 390Hz would be 30dB lower than those at 11.025kHz ± 390Hz.

Figure 8. shows the removal of the modulation components in the case of an external DAC.

The next step was to investigate the cause of the discrete amplitude modulation components.

Figures 9. and 10. show the sidebands produced, in a one-box player, by two identical test signals from different parts of the disc. The darker shaded areas mark the 390Hz-separated sidebands. The lighter shaded area marks a second set of lower-frequency modulation components.

In comparing figures 9. and 10., note the difference in the amplitude of the darker-shaded components, especially the 'fundamental' modulation component closest to the stimulus. the amplitude change in this component was found to correspond repeatably with the track position of the stimulus on the disc, and did not vary from disc to disc within the set.

The lower-frequency components in the lighter-shaded area were similar between the two track positions, but not identical.

Figures 11. and 12. show the same two test tracks but using a different disc from the set. Note that the behaviour of the 390Hz-spaced components is the same as for the other disc, and varies identically with track position. The low-frequency components, on the other hand, are characteristic of the disc and vary only slightly with the track position.

Neither component type was reproducible with an external DAC, regardless of the jitter-rejection capabilities of the DAC.

The track-position-dependent and disc-dependent components were investigated further using the DC test track, with the analogue output of the player connected through an analogue band-pass filter to an oscilloscope.

By tuning the band-pass filter to 390Hz, the modulation source was seen to be a 390Hz signal which was being rapidly switched on and off. The duty-cycle of the switching was directly dependent on the track position on the disc. It was conjectured that the 390Hz-spaced modulation components were related to the player's motor-control servo.

By tuning the band-pass filter over the 10Hz-100Hz region, it was possible to investigate the disc-dependent components. In this case it was necessary to capture the output on a storage oscilloscope. The technique employed was to start the DC track and allow a fixed period of time to elapse for the system to stabilise after the beginning of the DC output, before capturing a few seconds of output on the oscilloscope. This was found to produce a characteristic pattern for each disc. It was inferred that the low-frequency modulation components were related to the player's tracking and/or focusing servos.

The extent and precise characteristics of the servo-related modulation components varied considerably from player to player.

It appears on the basis of superficial investigation that the nature of the disc-related components does not correlate noticeably with the manufacturing variables; i.e. they are not repeatable from set to set.

None of the disc-related or servo-related artifacts was present for any two-box system: jitter-rejecting DACs were generally clean, as shown in the figures, whereas the spectra of non-jitter-rejecting DACs were swamped with interface-jitter-induced components which were large in comparison to the artifacts noted above in one-box players.

The existence of the above-mentioned amplitude modulation components in one-box players have since been independently corroborated by other investigators [5].

LISTENING TEST ANALYSIS

The principal objective of the listening tests was to assess 'concordance' among listeners, i.e. to establish to what degree they agreed in their ranking of the discs.

Early listening test results showed no obvious concordance, and it was clear that statistical analysis of a large number of tests would be required to uncover underlying concordance.

The statistical technique is as follows:

First, the 'null hypothesis', H_0, is proposed - that there is no basis for concordance between the listeners' results. The 'significance probability' of H_0, SP_{H0} is then calculated; this is the probability, given H_0, that the result set could have occurred by chance. SP_{H0} is a number between 0 and 1, and lower the value the better is the concordance.

SP_{H0} is calculated using 'Kendall's coefficient of concordance':

For N discs, m listeners, where T_n is the sum of the normalized rankings per disc, Kendall's coefficient is calculated as:

$$= \frac{\mathit{Varianceof}\ T_n}{\mathit{Maximumpossiblevarianceof}\ T_n}$$

$$= \frac{12\sum_m T_n^2}{m^2 N(N^2-1)} - \frac{3(N+1)}{N-1}$$

The SP_{H0} quantile occurs on the ^{2} distribution at:

$$\chi^2_{N-1} = m(N-1)W$$

The secondary objective was to quantify correlations between listening test rankings and disc manufacturing variables.

The statistical technique is as follows:

For N discs, where $r_1..r_N$ are the re-ranked listener rank totals, and $d_1..d_N$ is the mask defining the manufacturing parameter of interest, 'Spearman's rank correlation coefficient', R_S, is calculated:

$$R_S = \frac{1}{N-1}\sum_{i=1}\left(\frac{r_i \quad r}{s_r}\right)\left(\frac{d_i \quad d}{s_d}\right)$$

where are the mean values, and s_r , s_d the standard deviations of r and d respectively.

R_S describes the direction and extent of the correlation, assuming it is significant.

To assess significance, as before, the 'null hypothesis', H_0, is proposed - in this case that there is no correlation between the listeners' rankings and a particular manufacturing variable.

In this case, the SP_{H0} quantile occurs on the t_{N-2} distribution at:

$$t_{N-2} = R_S\sqrt{\frac{N-2}{1-R_S^2}}$$

Another method of assessing the result sets involves the anonymous inclusion of the duplicate reference disc. Presumably, this disc should receive a low ranking if we are to infer anything further from the results.

As well as analyzing the result set as a whole, various subsets were analyzed in order to minimize spurious concordances. The listener set definitions were as follows:

Global: All submitted results were included.

Individual: Where an individual had submitted more than one questionnaire, for example covering multiple tracks or players, the results for that individual were combined and presented as one result.

Unconferred: Where several individuals' tests were completed together at the same session, the results for all individuals were combined and presented as one result.

Professional: As for 'Unconferred', but only audio-industry professionals' results were included.

Within each listener set, separate analyses were made for 'one-box' and 'two-box' replay systems, as well as for all replay systems combined.

LISTENING TEST RESULTS

It became clear from examination of early responses that listeners had found it difficult to perform the tests as instructed. Only a minority of respondents had adhered to the rules laid down for forced ranking. Most commonly, questionnaires contained tied rankings, and many contained positive as well as negative rankings, based on which discs sounded better or worse than the reference, rather than expressing a unified ranking of the degree of difference. These problems necessitated re-normalizing of many of the responses. Even amongst those who responded according to instructions, it was clear that the task of comparing 13 different discs with a reference had proved a very difficult and time-consuming task. This is not surprising, given that any differences observed might be expected to be extremely subtle.

At the time of writing, 50 responses are included in the listening results set. These are analyzed in figure 13.

A surprising and important point is that of the 50 responses included, only three claimed that all the discs sounded identical to the reference. Despite this fact, it can be seen from figure 13. that concordance among listeners was poor, and that ranking of the anonymous copy of the reference disc was astonishingly high: it was generally rated as the MOST DIFFERENT from the primary reference disc in the global/all players set, principally as a result of professional listeners' responses, many of whom were using two box players. However, even in the one-box player analysis, it is mid-ranked.

Despite the poor concordance, some general trends are visible: firstly that the concordance is generally worse for two-box players and, surprisingly, among professional listeners within that group. Within the better-agreeing one-box listeners, the progression of significance through the listener groups is perhaps more logical - the concordance falls as individual and then unconferred groups are considered, then rises among professional listeners.

Upon inspecting rankings of the various discs in the set, however, some moderately consistent results do emerge; possibly the concordance as expressed by SP_{H0} is worsened by disagreement among the mid-ranking discs.

Discs A, C and G were ranked high among one-box listeners, with disc E being ranked consistently low. In the two-box case, discs D (the reference copy!) and H were ranked high, and discs B, I and E (again) were ranked low.

Among all listener groups and player categories, the correlation of the massed rankings with the manufacturing variables was investigated. Figure 14. shows the result for all listener groups for one-box players. For clarity, the table only contains entries for the cases of highest significance, i.e. where the probability of the correlation being random is low.

The observed trend of EFM encoder 'V' producing a disc which greatly varies from the reference, and the CD-R pre-mastering medium producing discs tending to be similar to the reference, seems to be reasonably consistent. EFM encoder 'V' was only used to produce the highly-ranked disc A.

It should be stressed that the significance of correlating disc rankings with manufacturing variables is doubtful so long as concordance among listeners is poor.

'GOLDEN-EARS' LISTENING TESTS

The random nature of the massed listening tests has been criticised by those who felt that results were likely to be compromised by inclusion of amateur listeners and uncontrolled listening environments. Critics felt that blind testing among expert listeners would be more likely to produce meaningful results.

To test this theory, blind tests with expert listeners have recently been started. At the time of writing, two expert listeners have been tested. Both are respected listeners of considerable experience, who are accustomed to serving on listening panels auditioning 'high-end' consumer equipment. The two tests were performed separately, listener 'A' using a one-box player and listener 'B' using a two-box player.

In each case, only two discs from the set were used; these were selected from the massed test results as being markedly different. One track, selected from the five music tracks by the listener, was played to the listener repeatedly from each disc in turn. The discs identified to the listener as '1' or '2' at the start of each play. After a number of plays, both listeners felt that they could identify the discs with some degree of confidence.

Ten trials were then performed using the selected track. Each trial consisted of one disc being played, then the other, after which the first was repeated. For each trial, the order of the discs was selected randomly on the toss of a coin, and the listener was asked to state whether '1' or '2' had been played first.

Listener 'A' was correct four times and wrong six times. The probability of getting four or more trials correct would have been 0.83 if responses had been offered at random.

Listener 'B' was correct three times and wrong seven times. The probability of getting three or more trials correct would have been 0.95 if responses had been offered at random.

CONCLUSIONS

Measurements have confirmed that amplitude modulation of the analogue outputs of many one-box CD players by motor and servo-related interference occur. Furthermore, the resulting spuriae are of a character and at a level which would be consistent with noticeable sound degradation for a critical listener. The disc-dependent modulations are low-frequency, and so produce distortion sidebands close to the stimulus frequency. Masking theory suggests that these would be inaudible. The track-position-dependent modulations are generally higher in frequency and amplitude and, as such, may be noticeable to a critical listener. These effects have not been identified in two-box players, which is not surprising since they appear to be caused by modulation of the reference voltage of the internal DAC by the servo and motor electronics.

The effects of disc-related or servo-related sampling jitter have NOT been found in either two-box or one-box players. Sampling jitter has been widely cited as a significant artifact in CD players by writers in both the consumer and professional audio fields. It seems possible that the sidebands produced by amplitude modulation may have been mistaken for the sidebands characteristic of sampling jitter modulation.

Listening tests have so far failed to produce convincing evidence for consistent sonic differences among the TD-2 disc sets. However, respected listeners maintain that the differences are present and reliably audible. There could be a number of reasons for this: most compelling is the idea that physical disc differences do affect the sound of one-box players in the manner described above, but that these physical differences are dominated by factors other than the manufacturing variables exercised by the TD-2 set. Several possibilities have been suggested:

1) The TD-2 sets were pressed in quite small quantities. CD production experts have subsequently suggested that the physical quality of discs is very variable as the machinery 'warms up', and that a large run would be required to achieve consistency.

2) The TD-2 sets have deliberately never been cleaned. It is quite possible that the cleanliness of the disc surface would have a significant effect on the level or character of servo activity during playing.

3) The TD-2 discs have paper labels. It is possible that variations in the mass and eccentricity of the labels may dominate servo behaviour on replay.

4) Other physical effects, such as the build up of static electricity on the disc surface, have been suggested as relevant. These have yet to be investigated.

However, the blind tests, whilst as yet too small in number to be conclusive, suggest that differences may actually be too small to be audible, even amongst expert listeners. If so, then other psychological factors must be responsible for the assertions by nearly all listeners tested that they can actually hear a difference.

The fact that so many professional listeners are hearing disc-dependent differences with jitter-rejecting external DACs is especially interesting. If this can be reliably confirmed, there is almost certainly a flaw in our understanding of the limits of perception. The only obvious cause would be that low frequency interface clock jitter related to the player's servos (although no sign of this has been discovered in tests so far), somehow passes through to the external DAC and manifests itself as sampling jitter. Sampling jitter at frequencies which would survive the rejection of a good-quality external DAC should be comfortably inaudible according to current masking theory. Perhaps a new audibility mechanism awaits discovery.

The measurement of servo-related modulation at the output of many one-box players is an important message to player manufacturers. It would be relatively inexpensive to reduce these effects considerably by improving isolation between the servo/digital electronics and the DAC within the player.

It seems that manufacturers must respond quickly if we are to avoid large numbers of DVD players reaching the consumer with the same problems. Audio performance expectations of DVD are high, with 24-bit, 96kHz operation supported for audiophile applications. The modulations measured in this work are comfortably manifested in the 16-bit DACs of current CD players.

The sensitivity of the reference terminal of most DAC chips is not sufficiently respected by most designers. It seems that this problem may actually be getting worse as we approach the DVD age. The application notes of some (so-called) 24-bit/96kHz DAC chips for DVD use, for example [6], show the reference fed directly from a power-rail shared by digital (and possibly servo) electronics with only minimal filtering, which would be ineffective at low frequencies.

FURTHER WORK

On the measurement side, there is more work to be done in investigating links between the disc-related servo artifacts and manufacturing processes. No firm links have yet been established, and it remains possible that random physical variations dominate.

Whilst it is unlikely that more massed listening tests will show anything further, there may be some useful benefit to be gained by more blind testing of expert listeners. It would be a breakthrough if any listener could be found who can reliably tell any two numerically-identical discs apart.

Further statistical analysis of the existing massed listening tests will be attempted. It is possible that good evidence of perceived disc differences may exist within the current data set which is being obscured by our present unsophisticated analysis.

ACKNOWLEDGEMENTS

This work has been a joint effort, made possible by the generous help of many colleagues in various branches of the industry. In particular, we are grateful for the advice and assistance of:

Robert Cleary
Ian Cooper at Metropolis Mastering, London
'Ez' Cousins at Studio Audio and Video, Cambridge
Crac Downes at Nimbus
Murray Harris at Chop'emOut, London
Hanson Hsu at Westlake, Los Angeles
Mike Martindell at Arcam, Cambridge
Angus McKenzie
Phil Mendham at Meridian Audio, Cambridge
David Smith at Sony Music, New York
Bernie Spratt at Chop'emOut, London
Robert Stuart at Meridian Audio, Cambridge
Don Tyler, and all at Precision Mastering, Los Angeles

The late Gabe Wiener of Quintessential Sound, New York, an uncompromising pursuer of ultimate sound quality, who achieved much in the short time allowed him.

Special thanks to Crispin Murray at Metropolis Mastering for meticulous preparation of the TD-2 disc masters.

We also wish to thank the various CD plants who generously made the test discs, and all the subjects who listened long and hard.

REFERENCES

[1] David Smith - Workshop presentation on ‘High-Resolution Audio Mastering’ given at the Copenhagen AES Convention, May 13 1996

[2] Julian Dunn - `Jitter and Digital Audio Performance Measurements' in `Managing the Bit Budget', the proceedings of the AES UK Conference 16-17 May 1994, pp 148-166.

[3] Julian Dunn and Ian Dennis - `The Diagnosis and Solution of Jitter-Related Problems in Digital Audio Systems' Pre-print 3868, presented at the 96th AES Convention, Amsterdam, February 1994.

[4] Patrick Stapley - 'The sound of music CD', published in One to One magazine, January 1997

[5] Franky K L Fan and Dr Sam M S Yeung - 'The sound of CD - Pt.2', published in One to One magazine, March 1997

[6] AK4321 and AK4324 Digital-to-Analog Converter data sheets, Asahi Kasei Microsystems, 1996

[7] Julian Dunn and Ian Dennis - `An Investigation of the Sonic Differences Between Numerically Identical Compact Discs' Pre-print 4399, presented at the 101st AES Convention, Los Angeles, November 1996.

Prism Sound Test Disc - TD-2

Track	Name	Signal Level	Start	Duration
Music				
1	Mariah Carey - Fantasy	(music)	00:00	01:20
2	Pink Floyd - Keep Talking	(music)	01:22	01:31
3	Haydn Piano Sonata #47	(music)	02:55	01:02
4	David Kikoski - Body and Soul	(music)	03:59	01:01
5	Original Cast - By Jeeves	(music)	04:59	01:01
****	Warning tone of increasing level	******	*****	*****
Test Signals				
6	Prism Sound DSA-1 Confidence test	Peaks to full scale	06:37	03:00
7	Prism Sound J-test	-3.01 dBFS rms	09:39	02:00
8	Odd phases of J-test	-3.01 dBFS rms	11:41	02:00
9	Even phases of J-test	-3.01 dBFS rms	13:43	02:00
10	IMD test 18kHz+20kHz	-3.01 dBFS rms	15:45	02:00
11	11kHz tone	0.00 dBFS	17:47	02:00
12	1kHz tone	0.00 dBFS	19:49	02:00
13	40Hz tone	0.00 dBFS	21:51	02:00
14	Multi-tone	-10.79 dBFS rms	23:53	03:00
15	11kHz tone	-60.00 dBFS	26:55	02:00
16	1kHz tone	-60.00 dBFS	28:57	02:00
17	440Hz tone	-60.00 dBFS	30:59	02:01
18	40Hz tone	-60.00 dBFS	33:02	02:00
19	Multi-tone	-70.79 dBFS rms	35:04	03:00
20	1kHz tone	-100.00 dBFS	38:06	03:11
21	Prism Sound DSA-1 Confidence test	Peaks to full scale	41:19	03:00
22	Prism Sound J-test	-3.01 dBFS rms	44:21	03:01

Track 23 DCA test signal 1: 22.05kHz tones with DC (Both channels same)

	Data values	DC level	AC level		
23:1	0005h, 7206h	-4.01 dBFS	-7.03 dBFS	47:24	02:00
23:2	0360h, FCEAh	-55.93 dBFS	-31.96 dBFS	+2:00	02:00
23:3	FB30h, 0101h	-33.54 dBFS	-32.87 dBFS	+4:00	02:00
23:4	8301h, 7230h	-24.48 dBFS	-0.59 dBFS	+6:00	02:00
23:5	85EAh, 85DFh	+2.60 dBFS	-75.50 dBFS	+8:00	02:00

Track 24 DCA test signal 2: Silence with DC (Both channels same)

	Data value	DC level		
24:1	7230h	+2.02 dBFS	57:26	02:00
24:2	0005h	-73.32 dBFS	+2:00	02:00

Total length: 61:32

Note: There are gaps with digital silence (zeroes) between tracks
All tracks are dithered except numbers 6,7,8,9,21,22,23,24
Signal levels do not include the dither component at -93.32dBFS
WARNING: Some tracks are at high level. Beware of the end of track 20

Figure 1: Contents of TD-2 test disc

Prism Sound CD Replication Investigation

Listening Test Questionnaire Instructions

General

Please fill in the listening test questionnaire according to the following guidelines. We will only be able to make valid statistical analyses of the test results if all listeners have tested in the same way.

If a number of listeners take part, please return separate questionnaires for each listener.

Please confine your listening tests initially to a common replay system for all the discs. If you repeat the tests using other replay systems, please return a separate questionnaire for each replay system used.

The five music tracks are at the start of the disc, followed by a warning tone of gradually increasing level. This is followed by the test signals, which you will NOT want to listen to! Be careful that you do not damage your equipment or yourself with these test signals, many of which contain extremes of level and frequency.

Please fax your results to +44 1223 425023.

'Basis' box

Please identify the disc set which has been issued to you. This is a letter ('A'...'L') marked on each disc in the set, and on their envelopes.

You may test as many or as few of the five music tracks as you wish. Please indicate which tracks you have listened to in the space provided.

'Listener' box

It will help us if individual testers' details are entered on their questionnaires; however **results of individuals' listening tests will remain completely confidential.**

'Replay System' box

Please describe the replay system in detail.

It is assumed that any external D/A converter is synchronised by the serial data from the transport. If this is NOT the case, please describe your synchronisation scheme in the 'Comments' box.

Figure 2: Listening test questionnaire, sheet 1 of 4

'Test Results' box

The purpose of these listening tests is primarily to compare the sound of each disc with that of the reference disc and, for those which noticeably differ, to rank the extent of the difference.

Rank any discs which are indistinguishable from the reference as '0', and leave the 'Characteristics' column blank.

For each disc which sounds different from the reference in any way, describe the nature of each distinct property of the disc's sound in the table on page 2, and write the designation letter ('A'...'F') in the 'Characteristics' column of the 'Test Results' box. Thus each disc may cite more than one 'characteristic' letter if multiple distinct sound properties are heard, and 'characteristic' letters may be cited by more than one disc if all such discs share common sound properties.

Additionally, each disc which IS distinguished from the reference should be ranked in order of degree of difference (i.e. the closest in sound is ranked '1', the next closest '2', etc.). It is important for our analysis that all non-zero ranked discs are ranked in order, with NO EQUAL RANKINGS (although all discs indistinguishable from the reference should be ranked '0').

Please attach a non-zero ranking to any disc which seems to you to differ in sound from the reference NO MATTER HOW SLIGHT THE DIFFERENCE MAY BE. Our analyses of the massed results are intended to decide whether the difference is significant or not.

Sound 'characteristics' can be attached to the reference if you feel that it has particular sound qualities. In this case, discs ranked '0' are implied to share any such characteristics, which need not be entered. Discs may differ from the reference by simply not sharing its characteristics, so it is possible to have a non-zero ranked disc with no characteristics cited.

'Description of Characteristics' box

This table is used to enter detailed descriptions of any sound differences identified during the listening tests, as described in the previous section.

If this table becomes full, please describe additional characteristics ('G', 'H' ... etc.) on a continuation sheet.

'Other Comments' box

This section is obviously for any other remarks which you might wish to make.

It is intended to acknowledge the participation of organisations and individuals in the study when it is finally printed. If you DO NOT WISH to be included in this list, please say so in the 'Other Comments' box.

Figure 2: Listening test questionnaire, sheet 2 of 4

Prism Sound CD Replication Investigation, Listening Tests
Disc Set TD-2

Basis

Disc set used (A..L):	Track(s) tested:	Date(s) of test:

Listener

Name:	Age:
Occupation:	Sex:
Have you previously noticed audible generation/pressing differences between CDs or any other digital media?:	

Replay System (fill in manufacturer and model details where appropriate)

CD player/transport:	Power amplifier:
Ext D/A converter:	Loudspeakers:
Digital interconnect:	Headphones:
Pre-amplifier:	Other:
Comments:	
Is mechanical/servo noise from the CD player audible at the listening position during track seeking?:	

Test Results

Disc	Ranking	Characteristics (detail in table, pg 4)	Disc	Ranking	Characteristics (detail in table, pg 4)
101			108		
102			109		
103			110		
104			111		
105			112		
106			113		

Figure 2: Listening test questionnaire, page 3 of 4

Prism Sound CD Replication Investigation, Listening Tests
Test Disc TD-2

Description of Characteristics (referred to in Test Results box on page 3)

Characteristic	Description
A	
B	
C	
D	
E	
F	

Other Comments

Figure 2: Listening test questionnaire, page 4 of 4

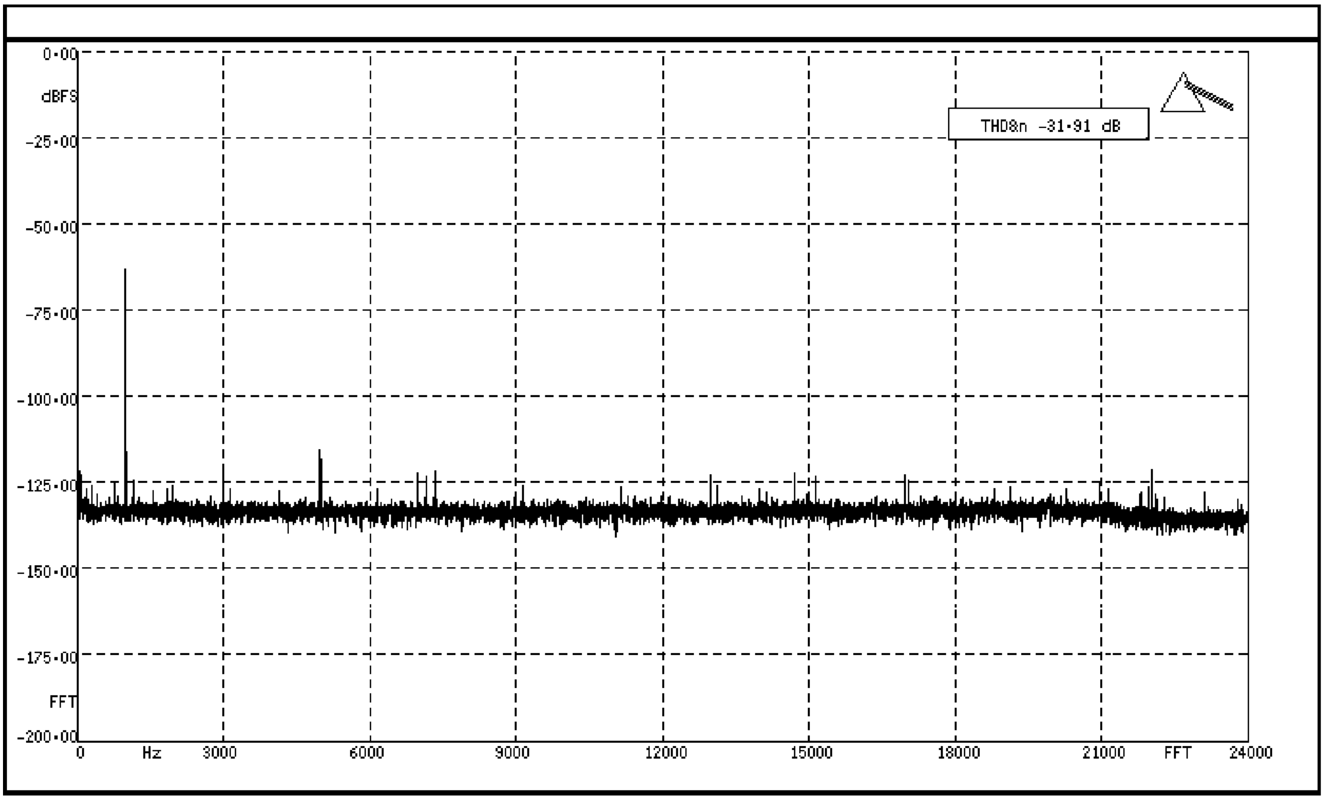

Figure 3: 1kHz @ -60dBFS, Disc A, Marantz CD-63, 32k FFT

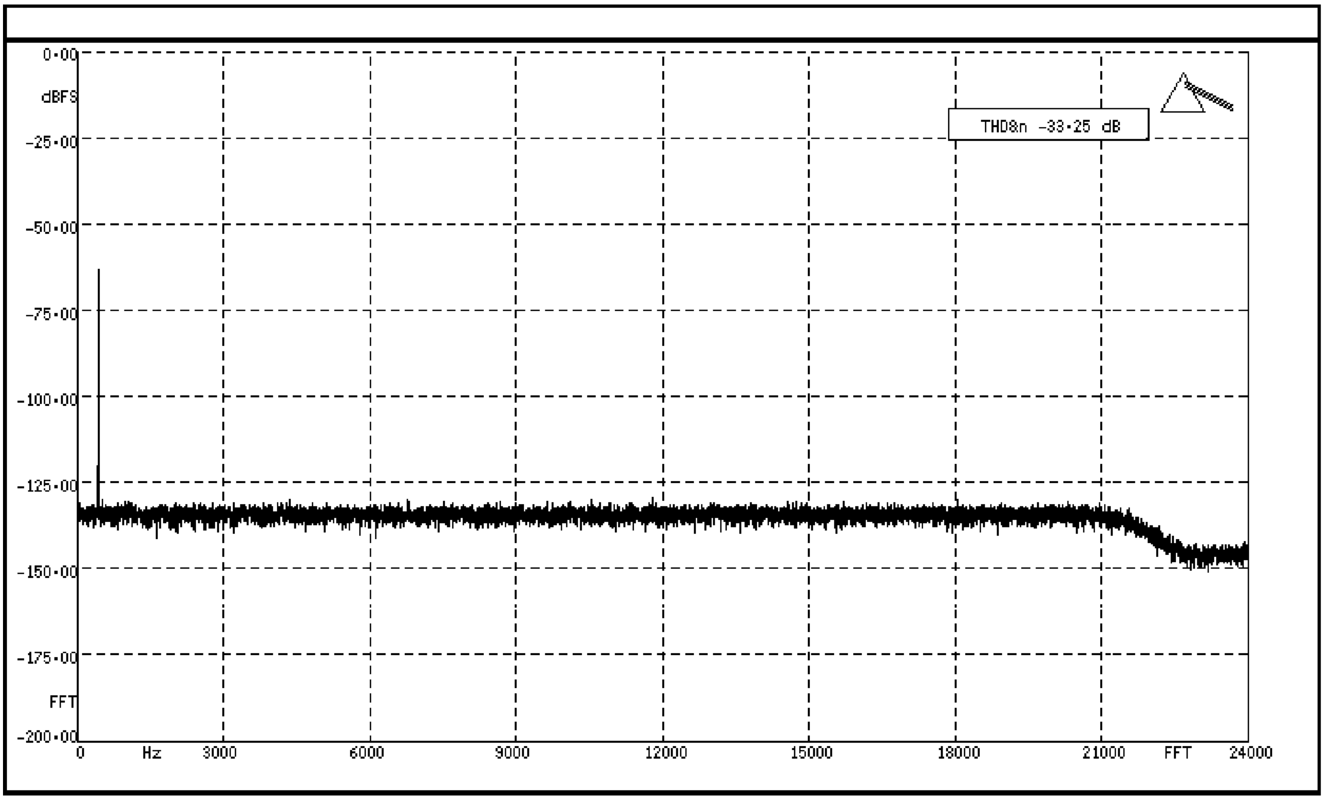

Figure 4: 1kHz @ -60dBFS, Disc A, External Prism DA-1, 32k FFT

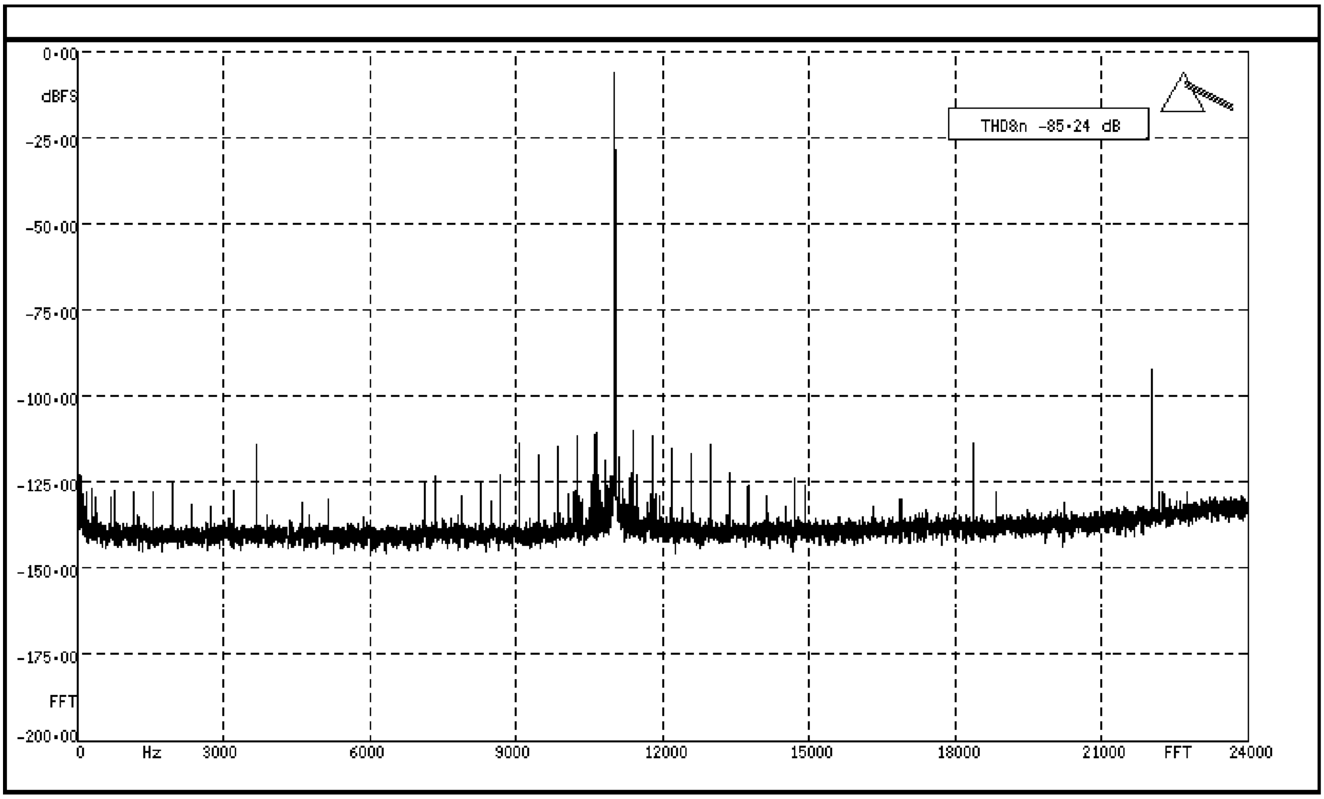

Figure 5: 11.025kHz @ -3dBFS, Disc A, Marantz CD-63, 32k FFT

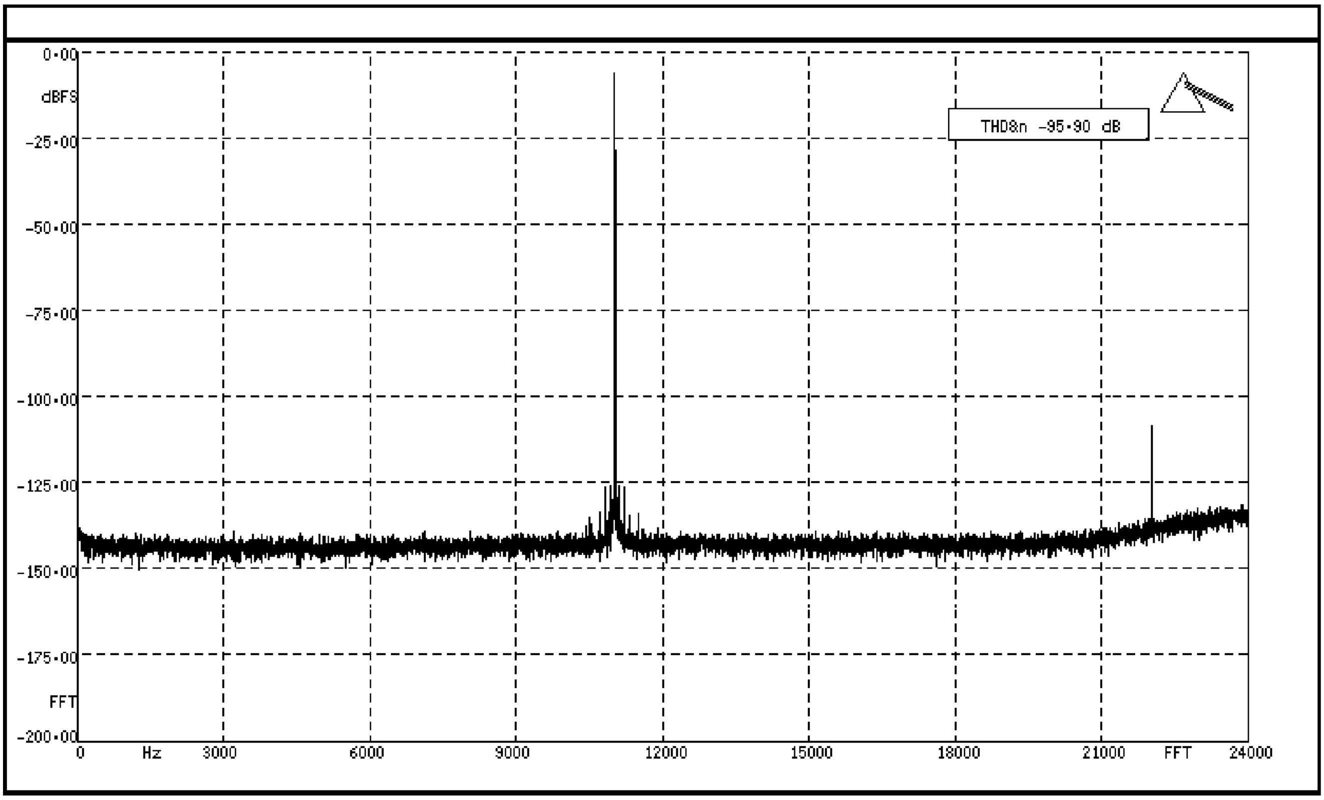

Figure 6: 11.025kHz @ -3dBFS, Disc A, External Prism DA-1, 32k FFT

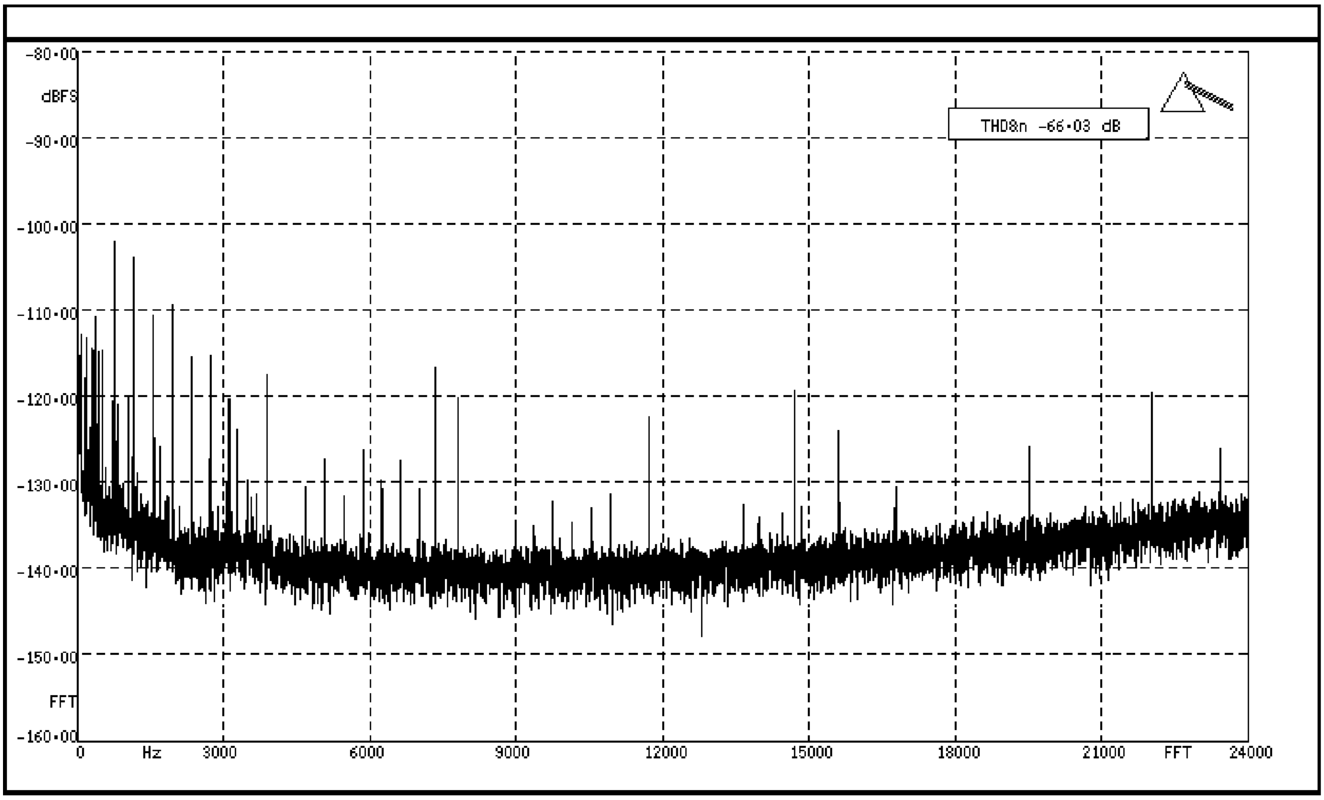

Figure 7: DC @ +2dBFS, Disc A, Marantz CD-63, 32k FFT

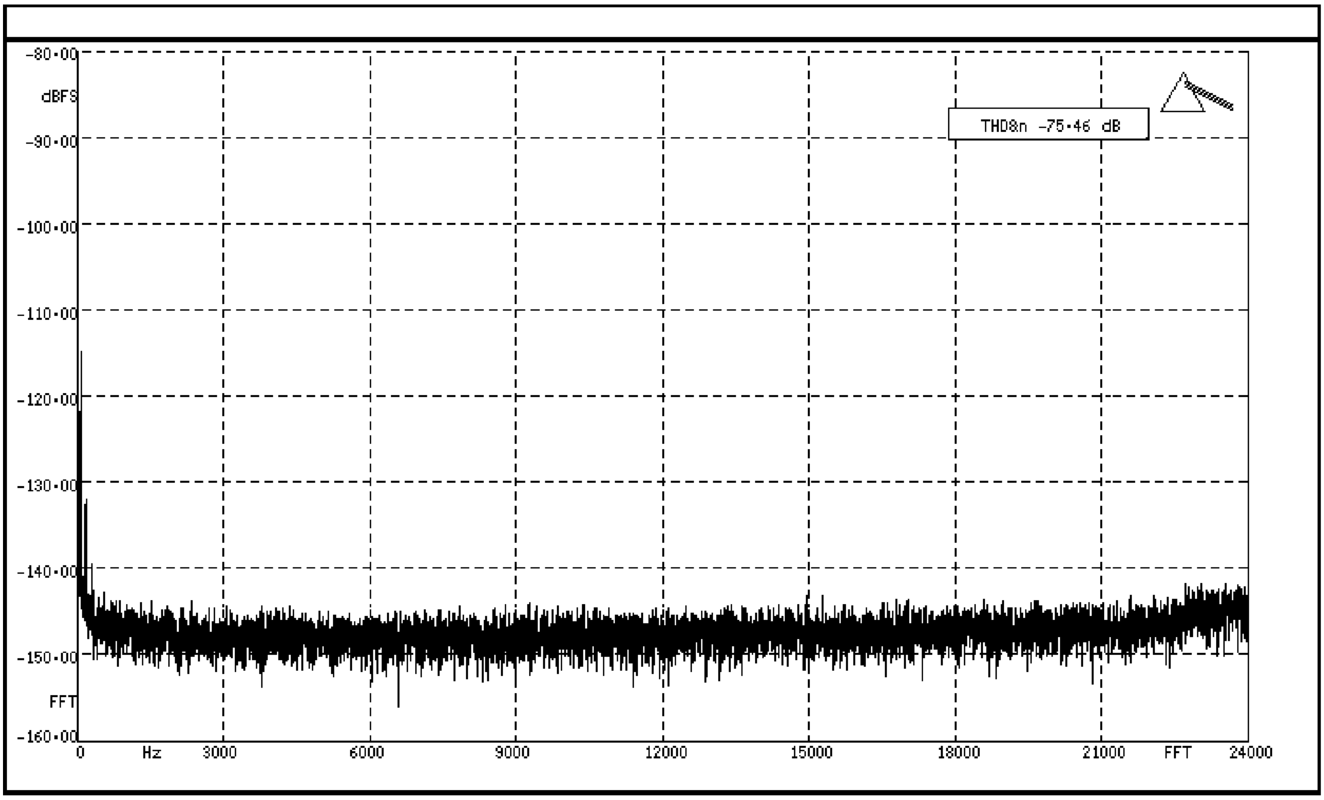

Figure 8: DC @ +2dBFS, Disc A, External Prism DA-1, 32k FFT

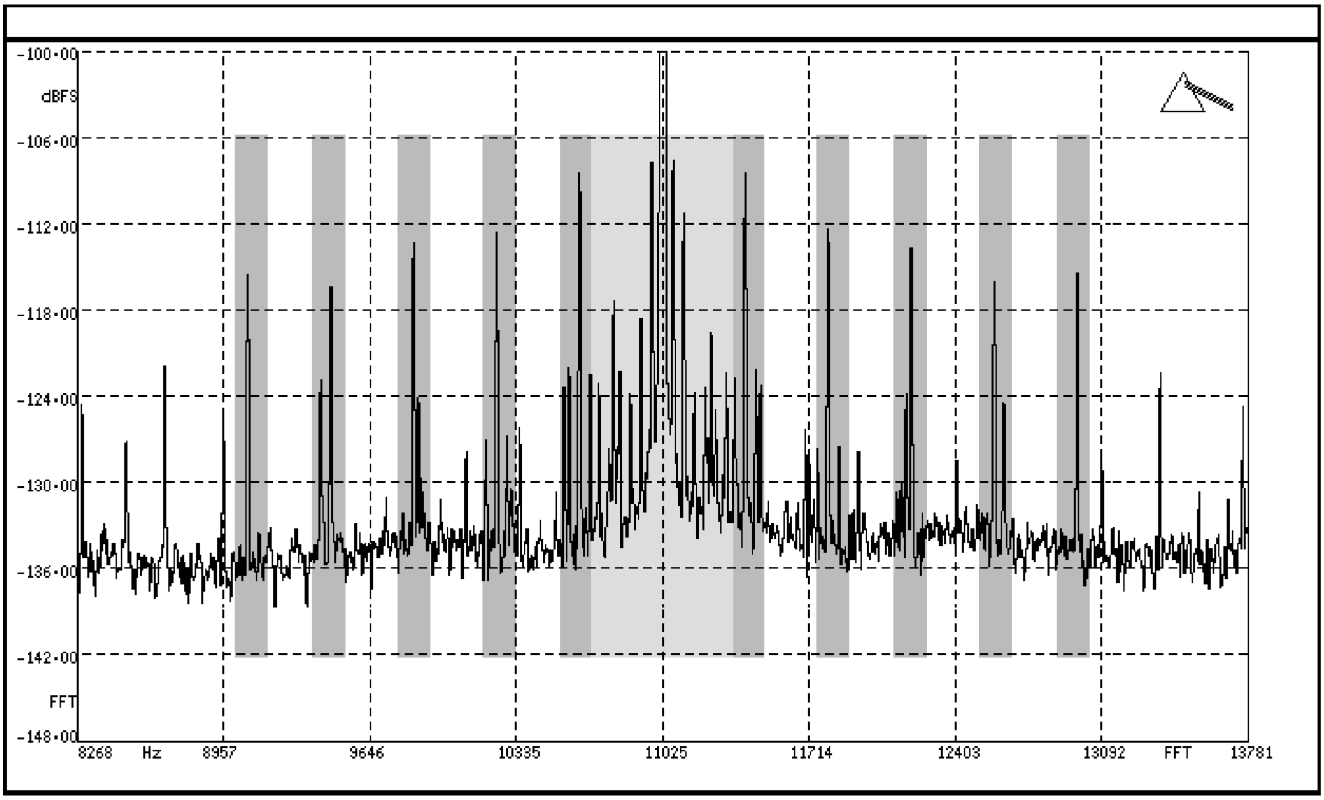

Figure 9: Track 7, Disc A, Marantz CD-63, 32k FFT

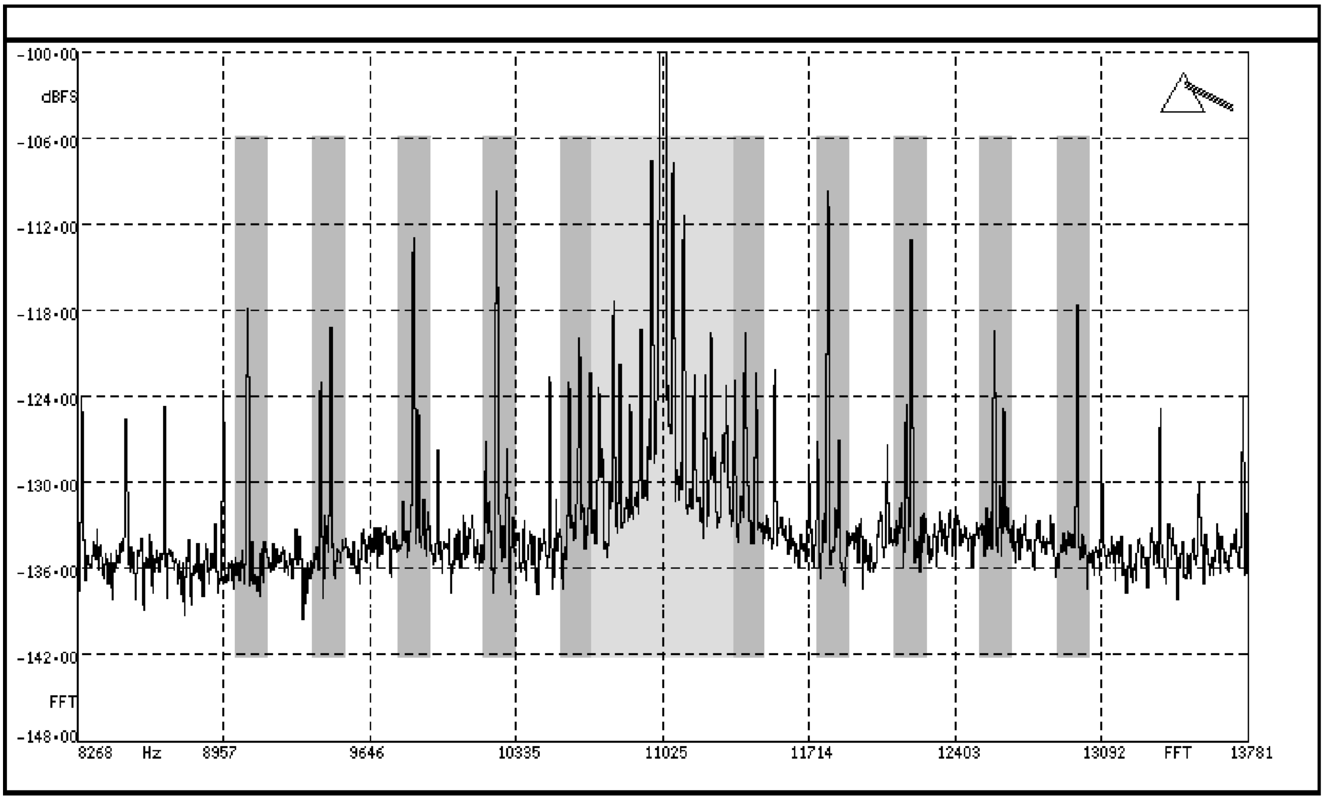

Figure 10: Track 22, Disc A, Marantz CD-63, 32k FFT

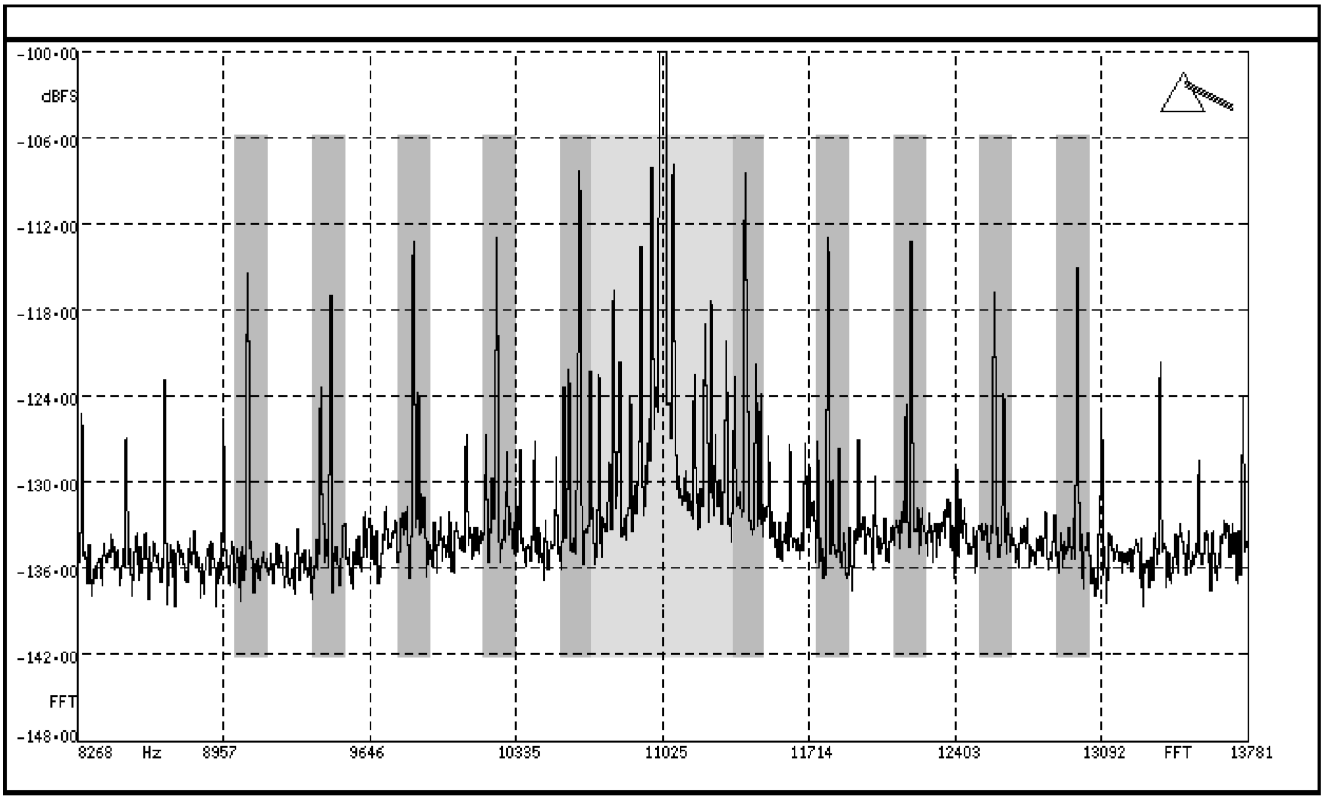

Figure 11: Track 7, Disc REF, Marantz CD-63, 32k FFT

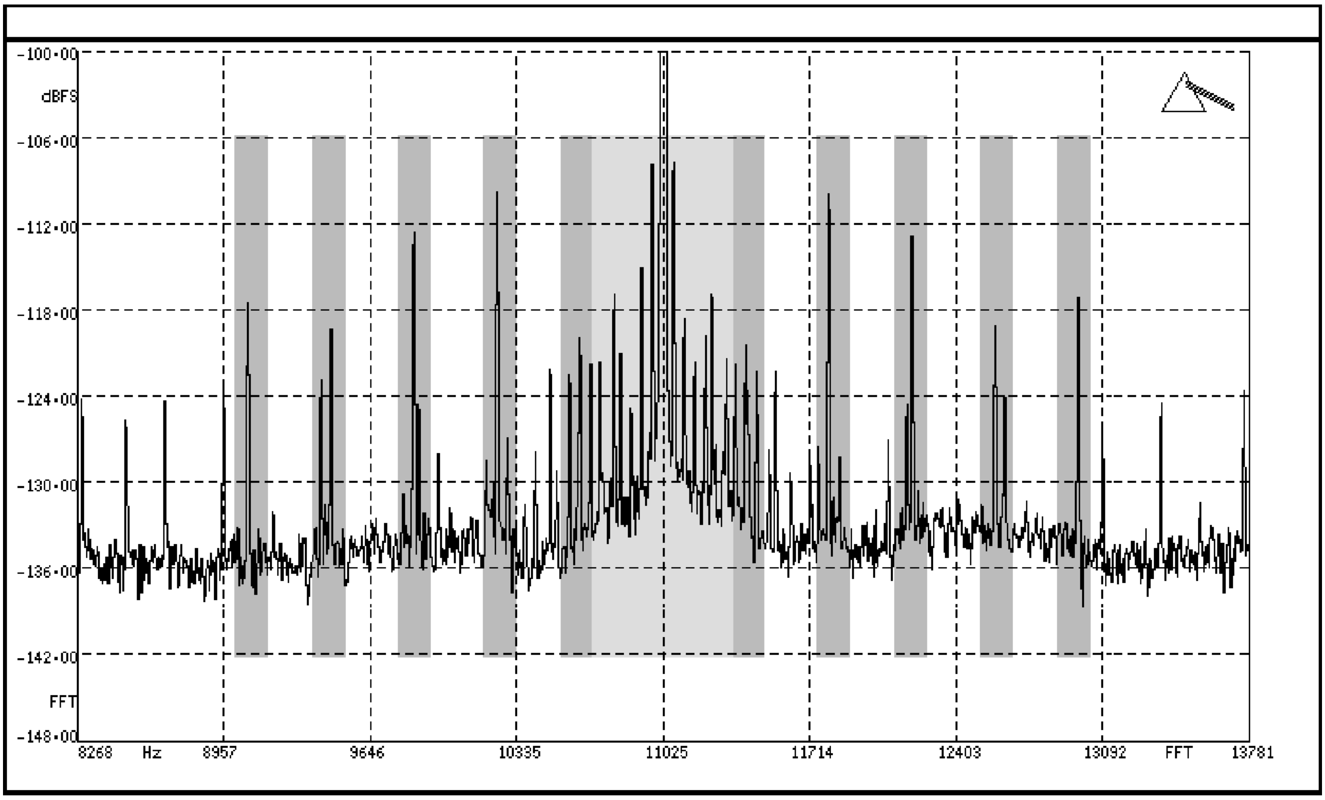

Figure 12: Track 22, Disc REF, Marantz CD-63, 32k FFT

	ALL PLAYERS							
	Average Rankings				Re-ranked Average Rankings			
	Global	Individual	Unconferred	Professional	Global	Individual	Unconferred	Professional
Disc A	7.59	8.17	8.15	7.17	12	13	13	7
Disc B	5.86	5.97	6.69	6.29	2	3	4	2
Disc C	7.66	7.42	7.3	8.01	13	9	8	13
Disc D (Ref)	7.53	7.87	7.57	7.19	11	12	9	8
Disc E	5.78	5.42	5.04	5.23	1	1	1	1
Disc F	7.5	7.57	7.64	7.95	10	10	11	12
Disc G	7.2	7.39	7.59	7.36	7	8	10	9
Disc H	7.37	7.81	7.77	7.42	9	11	12	10
Disc I	6.52	5.88	6.37	7	3	2	3	6
Disc J	6.78	6.77	6.23	6.33	5	5	2	3
Disc K	7.14	6.95	6.79	6.85	6	6	5	5
Disc L	6.76	6.63	6.91	6.59	4	4	7	4
Disc M	7.28	7.12	6.9	7.54	8	7	6	11
SP_{H0}	0.34	0.15	0.69	0.96				

	ONE-BOX PLAYERS							
	Average Rankings				Re-ranked Average Rankings			
	Global	Individual	Unconferred	Professional	Global	Individual	Unconferred	Professional
Disc A	7.74	8.4	8.34	7.49	12	13	13	8
Disc B	5.95	6.35	7.11	6.53	3	3	7	6
Disc C	8.52	8.11	7.93	9.08	13	12	11	13
Disc D (Ref)	7.17	7.09	6.75	5.67	6	7	5	4
Disc E	5.38	5.63	5.19	5.08	1	1	1	2
Disc F	7.34	7.09	7.55	8.17	10	7	9	10
Disc G	7.38	7.55	7.98	8.27	11	10	12	12
Disc H	7.33	7.7	7.6	7.46	8	11	10	7
Disc I	6.95	6.4	6.87	8.17	5	4	6	10
Disc J	6.79	6.65	6	5.48	4	5	3	3
Disc K	7.22	6.93	6.54	6.5	7	6	4	5
Disc L	5.86	5.89	5.75	4.86	2	2	2	1
Disc M	7.33	7.17	7.32	8.13	8	9	8	9
SP_{H0}	0.22	0.46	0.78	0.5				

	TWO-BOX PLAYERS							
	Average Rankings				Re-ranked Average Rankings			
	Global	Individual	Unconferred	Professional	Global	Individual	Unconferred	Professional
Disc A	7.32	7.17	7.49	6.8	9	7	10	5
Disc B	5.71	5.65	6.15	6.3	1	2	3	2
Disc C	6.21	6.52	6.79	7.14	3	4	6	9
Disc D (Ref)	8.15	9.08	8.5	8.56	12	13	13	13
Disc E	6.47	6.04	6.01	6.64	4	3	2	3
Disc F	7.76	7.72	7.11	6.88	11	10	8	6
Disc G	6.88	7.54	7.48	7.04	6	9	9	7
Disc H	7.44	8	7.95	7.44	10	12	11	11
Disc I	5.79	4.38	5.11	5	2	1	1	1
Disc J	6.76	6.62	6.25	6.78	5	5	4	4
Disc K	7	7.03	7.09	7.1	7	6	7	8
Disc L	8.29	7.89	8.39	8.19	13	11	12	12
Disc M	7.21	7.38	6.69	7.15	8	8	5	10
SP_{H0}	0.7	0.33	0.87	0.97				

Figure 13: Listener concordance results, from 50 responses

	Pre-Mastering Medium			EFM Encoder				Laser-Beam Recorder				LBR Rate	
	UMAT	EXA	CD-R	'V'	'W'	'X'	'Y'	'V'	'W'	'X'	'Z'	1x	2x
Global				0.4									
Individual				0.47									
Unconferred			-0.46	0.46									
Professional		0.4	-0.46							0.41			

The table is filled in wherever SP_{H0} is less than 0.1, i.e. where the probability of the correlation having occurred by chance is less than 0.1. The value tabulated is the Spearman Ranked Correlation Coefficient, R_S. This is an arbitrary indicator of the strength and polarity of the correlation. Positive numbers correspond to increasing rank, i.e. positive numbers correlate with greater disc differences, negative numbers with smaller disc differences.

Figure 14: Correlation of listening results with manufacturing variables for one-box players