Audio Engineering Society
Convention Paper 6086
Presented at the 116th Convention
2004 May 8–11 Berlin, Germany

DVD-Audio versus SACD

Perceptual Discrimination of Digital Audio Coding Formats

Listening Comparison Test between *DSD* and *High Resolution PCM (24-bit / 176.4 kHz)*

by

Dominik Blech and **Min-Chi Yang**

Erich-Thienhaus-Institute (Tonmeisterinstitut), University of Music Detmold, Germany
http://www.hfm-detmold.de/hochschule/eti.html

ABSTRACT

To study perceptual discrimination between two digital audio coding formats, "Direct Stream Digital" and high-resolution (24-bit, 176.4 kHz) PCM, subjective listening comparison tests were conducted with specially recorded sound stimuli in stereo and surround.

To guarantee their reliability, validity and objectivity, the double-blind ABX tests followed three main principles: The signal chain should be based on identical audio components as far as possible; these components should be able to convey very high audio frequencies; and the test population should consist of various groups of subjects with different listening expectations and perspectives.

The results showed that hardly any of the subjects could make a reproducible distinction between the two encoding systems. Hence it may be concluded that no significant differences are audible.

1. INTRODUCTION

Two currently coexisting systems for digital recording—"**P**ulse **C**ode **M**odulation" (PCM) and "**D**irect **S**tream **D**igital" (DSD)—have aroused considerable controversy with regard to both technical and sonic issues. These systems have spawned two corresponding Compact Disc formats: **DVD-A**udio, which is based on PCM, and the **S**uper **A**udio **C**ompact **D**isc (SACD), which is based on a 1-bit signal with 64x oversampling relative to the original 44.1 kHz CD (= 2.8224 MHz). The creators of these systems, many audio professionals using these systems and consumers listening to the resulting products have claimed that audible, distinctly perceptible differences exist between them. But there is no consensus about this, and it is still unclear which of these competing systems will succeed in the long term.

The present investigation undertakes to determine the degree to which test subjects can perceive a difference between DSD and high-resolution (176.4 kHz / 24-bit) PCM in an ABX test. The experience of carrying out

these listening tests has shown ever more clearly the importance of using double-blind ABX tests, since only by this means—free of suggestion and subconscious prejudice—can it be shown what is or is not perceivable on a repeatable basis. Otherwise, it is well known that the transition zone between auditory perception and imagination can become quite narrow.

This work is intended to redress the imbalance which currently exists between the abundance of theoretical data and the much smaller amount of reliable, valid and objective scientific evidence concerning these systems, as well as to stimulate further investigation and thought.

2. PRELIMINARY CONSIDERATIONS

The listening test is intended to reflect each underlying digital encoding system, not the format or medium which carries it. Thus it is necessary to set up recording and playback signal paths which are based, insofar as possible, on identical audio components, so that the two recording methods are being compared sonically rather than the equipment.
The unavoidable weak point is the A/D and D/A converters. To minimize differences, converters of the same make and model which support both encoding systems should be used exclusively.

Furthermore, a workable test routine must be established in which the selections are played back with straightforward, accurate synchronization to let the user switch between DSD and PCM at any desired moment. Audible differences in latency (which might occur if two independent audio workstations were used) must be avoided; otherwise the listener might be able to differentiate between the sources on the basis of timing alone. The "AES data bit-mapping" approach offers a solution to this problem by allowing lossless "packing" of the data after A/D conversion, then "unpacking" the data on the D/A side with a corresponding algorithm. This arrangement also permits the use of a single multi-channel audio workstation to record all the digitized audio channels for each example simultaneously and synchronously.
Listening tests should be conducted with both stereophonic and surround recordings, to take advantage of the average listener's greater familiarity with stereo and to allow the possibility of hearing any effects which might alter spatial perception alone.
It is essential that the specially recorded music and sound samples be absolutely identical. To this end, all processing of the recorded material such as level changes or editing must be completely avoided, since any such processing would require temporary conversion of the recorded DSD material to a multi-bit format—fundamentally upsetting the test conditions. For the same reason any mixing that would influence the sound quality must be dispensed with; accordingly, the audio signals from two (or in surround, five) microphones must be recorded without falsification and later, routed correspondingly to an equal number of loudspeakers.
Each person who performs and/or listens to music has an individual set of listening experiences, expectations and focal points. Thus the population of test subjects and the range of available sample recordings should be as wide as possible. Furthermore, due to the individuality of each test subject, it is vital to test everyone separately.

To keep the test subjects' "performance anxiety" to a minimum, a pleasant, neutral room arrangement and personal atmosphere should be established (without, however, influencing the candidates' choices). It should also be possible for the subject to opt for a pause in the testing procedure at any time.

3. DESIGN OF THE EXPERIMENT

As previously stated, one fundamental requirement for an objective, technically valid listening comparison is that the source material which is to be compared must be completely identical and "unprocessed"—it must not be altered in level, subjected to artificial reverberation, edited or otherwise "treated." Since such material, if it exists at all, was not available, original samples in both two-channel stereo and five-channel surround were recorded by the authors before the start of the listening tests. This was done with the help of instrumentalists from the University of Music in Detmold (Hochschule für Musik Detmold) in the "Neue Aula" concert hall, under optimal conditions and with the air conditioning system deactivated.
To avoid any influence of a mixer on the sound quality, the stereo music examples were recorded with two microphones and the surround examples with five. All the microphones had extended frequency response to 40 or 50 kHz (Schoeps MK 2S, MK 4 and MK 41 capsules with CMC 6-- xt amplifiers, and Sennheiser MKH 800); one microphone was simply assigned to each playback loudspeaker. The microphones were connected to microphone preamplifiers (Lake People F/35 II) which raised the signals to line level, then these signals were

sent to the control room via 50-meter low-capacitance cables (Klotz M1 series). At that point the five analog signals were split via "Y" adapters and converted to digital, with one set of three two-channel dCS 904 units used for DSD and another such set used for 176.4 kHz, 24-bit PCM. The resulting digital signals were then stored on a "Pyramix Virtual Studio System" (Merging Technologies) as "non-audio" files by using the "data bitmapping" system of the converters to generate 24-bit, 44.1 kHz files (*i.e.* two channels of DSD were stored as six channels on the workstation).

For playback, the audio signals were converted back to analog again using dCS converters (a separate pair of two-channel dCS 954 for the L, R, LS and RS of each encoding system, plus separate two-channel dCS 955s for each system's center channel signal), and sent through a high-quality stereo and surround monitor control unit developed by the Emil Berliner Studios (type MU 2000). The listener could switch between DSD and PCM signals by using the ABX software (also developed by the Emil Berliner Studios) to operate this monitor control unit. A software-controlled delay is introduced at the moment of switching between these signals to prevent any accidental overlap. Loudspeakers by Manger, distinguished by their very precise impulse response and frequency response up to 35 kHz, were used for playback. If the test subject opted for a stereo listening sample, he or she could furthermore listen on a pair of Stax headphones. All connections were carried out exclusively with new, high-quality analog and digital cables from Klotz.

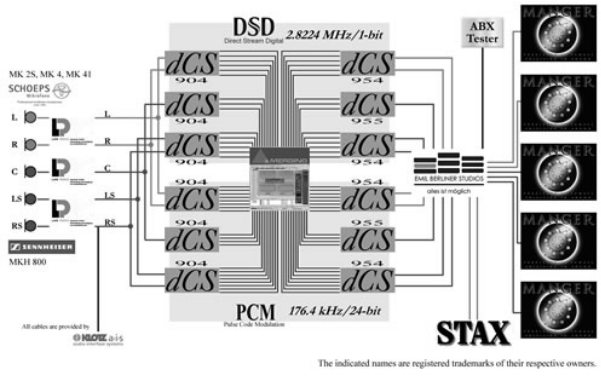

Figure 1: Signal flow diagram for the testing setup

To do justice to the diverse experience, expectations and listening focuses of the broadest possible range of listeners, it was felt that a correspondingly broad selection of musical samples should be made available. The following table gives an overview of the recorded music samples:

Stereo	Surround
Harpsichord F. Couperin – *Rondeau* (C minor) 3:21	**Harpsichord** F. Couperin – *Rondeau* (C minor) 3:25
Vocal W. A. Mozart – *Le nozze di Figaro*, Susanna's aria, *"Deh vieni, non tardar"* 3:11	**Vocal** J. Strauss – *Die Fledermaus*, Adele's Song, *"Mein Herr Marquis"* 1:32
Guitar E. Clapton – *Signe* 2:06	**Guitar** Unknown – *Romance* 2:20
	Jazz Trio M. Manieri – *Sarah's Touch* 4:32
Oboe G. Ph. Telemann – *Phantasie Nr. 8, 2nd Mvt. (Spirituoso)* 1:11	**Oboe** G. Ph. Telemann – *Phantasie Nr. 8,* *1st Mvt. (Largo)* 2:22 *2nd Mvt. (Spirituoso)* 1:14
	Organ M. Reger – *Fugue in D Minor, Op. 135b* 5:10
Percussion Solo *Maracas* 1:30 *Guiro* 1:30 *Wind chime* 1:30 *Castanets* 1:30	**Percussion Solo** N. J. Zivkovic – from *Danza Barbara* Tutti Section No.1 2:33 Tutti Section No.2 1:33
Piano D. Scarlatti – *Sonata, K. 188* (A Minor) 2:38	**Piano** Fr. Chopin – *Études, Op. 25, No. 11* (A Minor) 4:04
Speech – (Russian) A. Pushkin – from *Eugen Onegin* 2:08	**Speech – (Russian)** A. Pushkin – from *Eugen Onegin* 2:09
	String Orchestra E. Rautavaara – *Pelimannit "Fiddlers"* *2nd Mvt. (Presto)* 1:08 *5th Mvt. (Presto)* 1:16
Trumpet *Blues Improvisation* 2:31	**Trumpet** *Blues Improvisation* 2:36
	Violin J. S. Bach – *Sonate No. 1, BWV 1001, Adagio* 4:19

Figure 2: Overview of the available music and sound samples

The music and sound samples were available to be played in their full length. The listeners had complete

operational control over the ABX software by means of a control unit, so they could determine the course and timing of the listening comparison process. This ability was an important factor in minimizing the previously mentioned risk of performance anxiety in the test subjects.

All testing was performed with careful attention paid to the exact matching of levels between the two signal paths. All A/D and D/A converters were measured and set carefully to precisely equal levels before the tests, so that they were operating under identical conditions.

4. LISTENING ENVIRONMENT

The listening room was measured and brought acoustically into conformance with EBU [1] and ITU [2 and 3] guidelines for monitoring rooms with respect to reverberation time, background noise level and reference listening level.

The Manger loudspeaker systems were arranged in a circle in accordance with ITU [3] recommendations for multi-channel loudspeaker arrangements (L, C, R, LS and RS). In this arrangement the stereo basis width (B) between L and R loudspeakers—and consequently the radius of the circle—amounted to 2.20 meters. The LS and RS loudspeakers were positioned on the circle at points 110° equidistant from the center speaker.

5. COURSE OF THE EXPERIMENT

ABX testing is a method which permits a "blind" comparison to be made between two differing signals, "A" and "B." An ABX software program randomly assigns A or B temporarily to "X," and the listener's goal is then to identify "X" correctly as either A or B by comparing its sound with that of the two original signals. After each such decision, the software records the result and again randomly reassigns "X" to either A or B so that the next independent choice can be made. Before registering a decision, the listener can switch among A, B and "X" freely and as often as desired so that there is no time pressure. At least 16 such decisions must be recorded before it becomes possible mathematically to rule out chance decision-making and achieve statistical significance. On the basis of statistical analysis it can be determined whether a difference between A and B was perceived or not. The ITU [3] recommends the use of "double blind" testing when carrying out such listening comparisons.

In the present study either "A" was DSD and "B" was PCM or vice versa; this was set at random for each new listener by one of the two authors conducting the tests, but the identity of "A" and "B" was naturally kept constant throughout any one subject's test procedure. To increase the evidence value of the results, each test subject was given 20 rather than 16 comparisons to decide.

The listening tests were divided into two phases: Following a precise explanation of the test procedure and a technical briefing from one of the authors, there was a learning phase during which the subject could become accustomed to the relatively simple operation of the control module for the ABX software, to the software itself and, to the extent of his or her interest, the musical material (stereo and surround) that was available for use. The subject could listen repeatedly to predefined musical segments. Unlike the actual testing phase, however, in the learning phase the test subject was told after each choice whether he or she had correctly identified "X" or not. In this way a simulated version of the testing situation was offered.

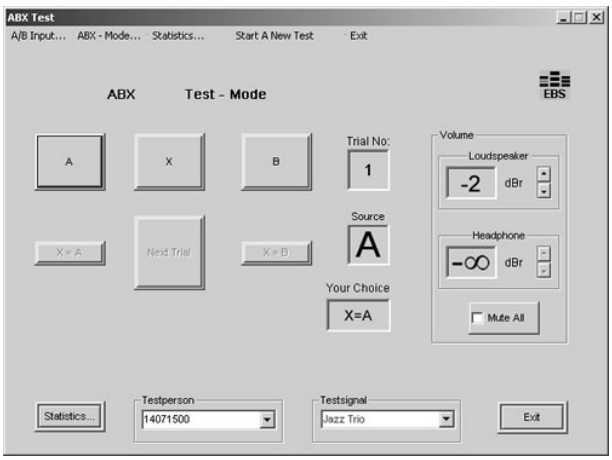

Figure 3: Screen image of the ABX software which the test subjects operated via control module (see below)

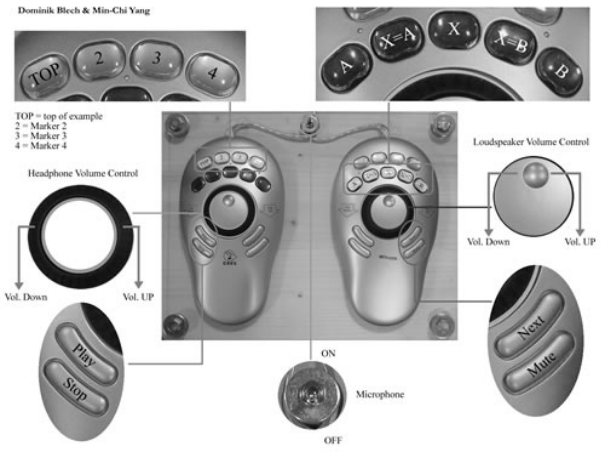

Figure 4: Control module for operating the ABX software, using two modified "Shuttle Pro" panels (Contour) mounted on a tray

During the learning phase, the subject had to choose one music or sound sample and decide between two-channel stereo or surround playback; if stereo playback was chosen, loudspeaker or headphone playback was also decided upon. The statistical measures would be valid only if the conditions resulting from these choices were maintained for all 20 of the test decisions to follow. To avoid ear fatigue, the learning phase was limited to *ca.* 20 – 25 minutes including the introductory discussion.

Following a brief, optional intermission, the second phase (the actual listening test) was carried out with the previously chosen music or sound example. Results were given to the test subject only after completion of the entire listening test and the filling out of two questionnaires concerning his or her personal characteristics, music listening habits, and reactions to the testing experience.

6. EVALUATION OF THE EXPERIMENTAL DATA

The present study is intended to find out whether test subjects can demonstrably differentiate between the two digital encoding systems DSD and PCM (176.4 kHz / 24-bit). The mathematical evaluation of the test data is based on the stochastic model of binomial distribution. The ITU [3] recommends using a significance threshold of 5%, *i.e.* $p <= 0.05$. With 20 trials per test run, the percentage score required for $p <= 0.05$ is 75%; thus the test subject must give at least 15 correct answers. The probability of achieving such a score by guessing at random is $p = 0.021$ or 2.1%, while for example there would be nearly a 6% chance of guessing 14 of 20 trials correctly ($p = 0.0591$). Thus in accordance with international standard practice, the threshold of critical probability was set at 15 correct responses per test for this experiment.

The listening tests took place over a period of 28 days. During this span of time 145 tests could be carried out with 110 test subjects. Some participants carried out the test twice, either consecutively or on separate days, using different music examples. ITU guidelines [3], which state that conclusions may be drawn on the basis of results from 20 persons or more, were clearly met.

The testing population consisted of 43 female and 67 male subjects, whose age distribution is shown in Figure 5. The mean age was 32.9 years.

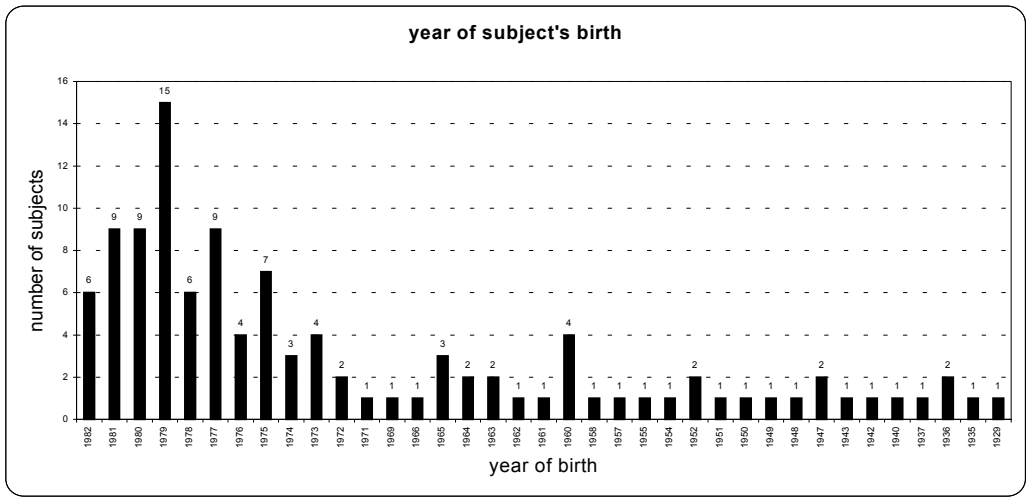

Figure 5: Age distribution of the test subjects

Figure 6 profiles the occupations of the test subjects. Since nearly all were trained as performing musicians, Figure 7 details the subjects' major instruments. These charts make clear that this was a test population in which the majority was well accustomed to musical and critical/analytical listening.

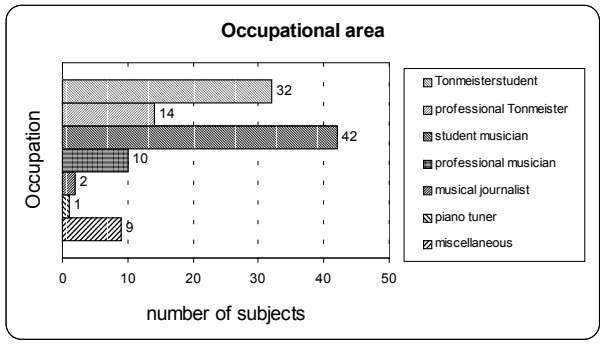

Figure 6: Distribution of the test subjects by occupation

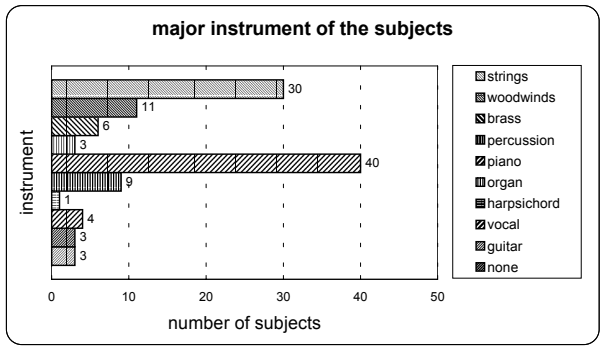

Figure 7: Distribution of the test subjects by major instrument

The 145 completed tests consisted of 45 stereo examples (30 of which were auditioned through headphones) and 100 surround examples, for a ratio of 1:2.2. Figures 8a and 8b show which of the 20 available music samples were selected by test subjects for stereo and for surround playback respectively.

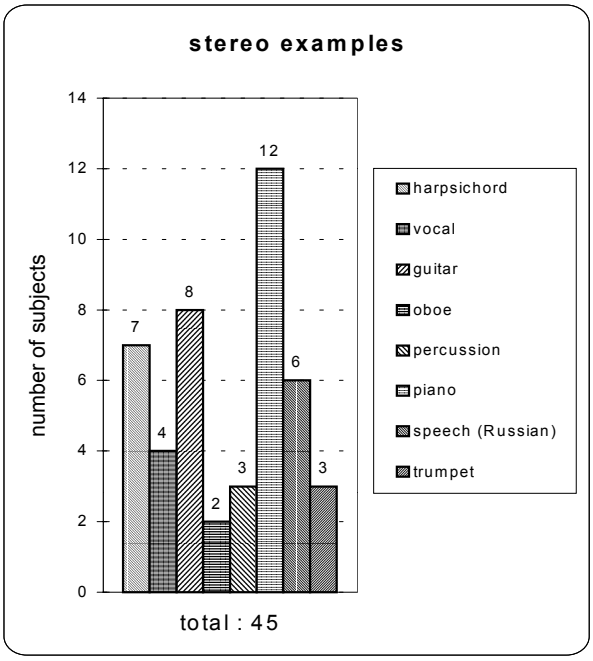

Figure 8a: Music selections chosen for the 45 completed stereo listening tests

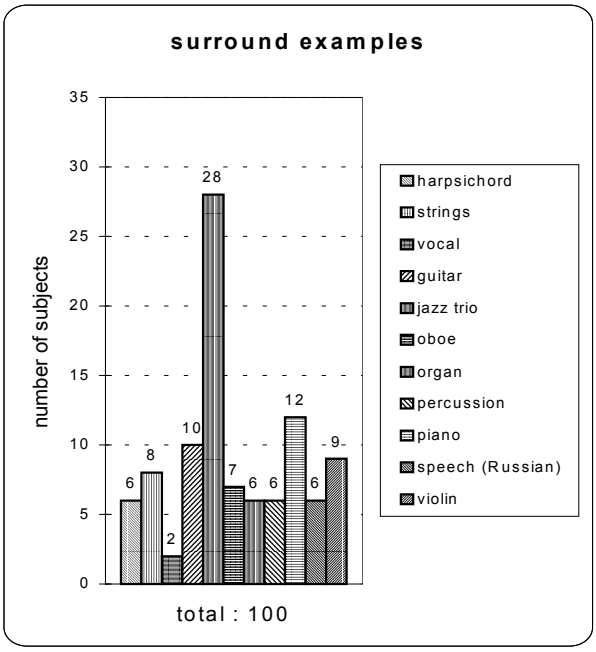

Figure 8b: Music selections chosen for the 100 completed surround listening tests

It is striking that the "Jazz Trio" example was chosen with above-average frequency. The test subjects explained that the recording seemed quite transparent with its clear panoramic layout of instruments: (piano ↔ L, R; bass ↔ C; percussion ↔ LS, RS) and that it contained a variety of both sonic and spatial aspects which offered good points of reference while listening.

The arithmetic mean of scores achieved for each recorded stereo example chosen are shown in Figure 9a, while the mean scores achieved with the various surround examples are shown in Figure 9b. The solid horizontal line indicates the score (75% or 15 correct answers) which would be required in order for a test result to achieve significance, given a threshold probability of 5% as mentioned previously.

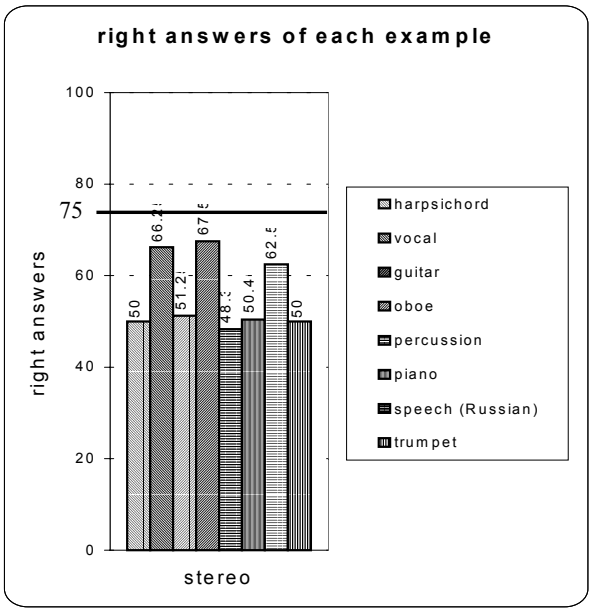

Figure 9a: Tabulation of mean scores per stereo music example

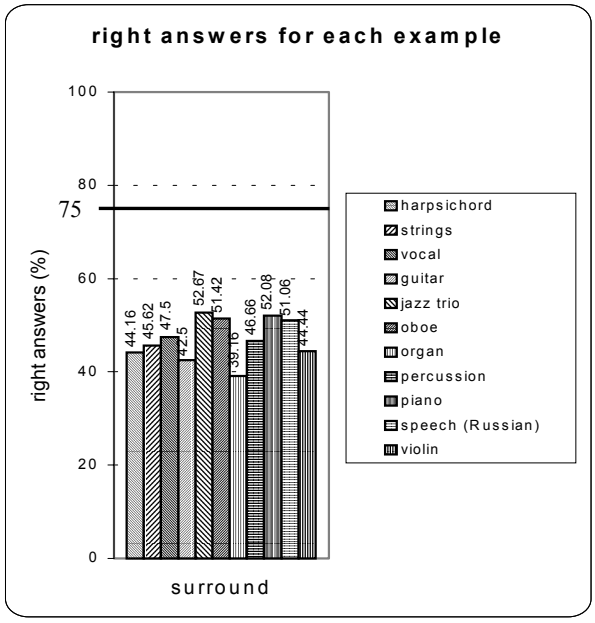

Figure 9b: Tabulation of mean scores per surround music example

Figure 10 shows the distribution of the 145 individual test scores.

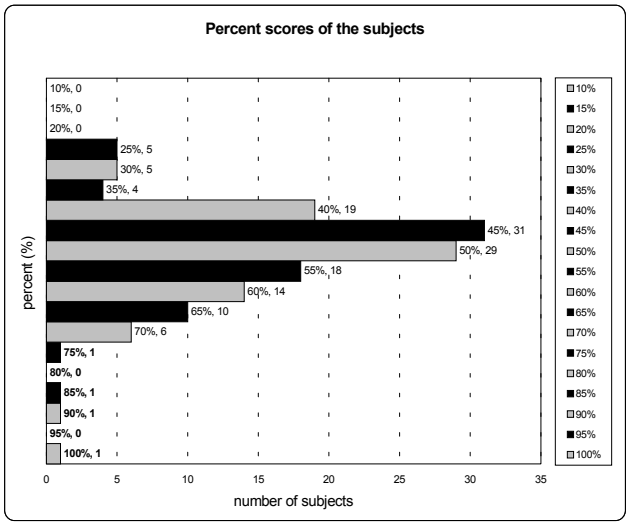

Figure 10: Distribution of the 145 individual test scores achieved by listeners

The four highest scores fell into the region of "critical probability." This amounted to only 2.76% of all the tests. These four tests were carried out by four separate listeners, all of whom chose stereo music examples, and

in all four cases headphones were used—thus excluding the influence of the listening environment to the greatest possible extent. Each of the four test subjects had chosen a different music example:

- **Oboe:** Stereo with headphones:
  75% correct responses → p = 0.0207
- **Speech:** Stereo with headphones:
  85% correct responses → p = 0.0013
- **Guitar:** Stereo with headphones:
  90% correct responses → p = 0.0002
- **Vocal:** Stereo with headphones:
  100% correct responses → p < 0.000001

All four of these tests occurred within the final four days of testing. By that point the testing schedule was full, so unfortunately these individuals could not be brought back for follow-up verification tests.

The full write-up of this investigation contains a detailed explanation of these four test results which lie outside the distinctly recognizable trend—particularly when one considers that in 100 surround tests, not even one result achieved a level of significance. Only a brief synopsis will be given here:

Because of its principle of operation, when a "stop" or "play" command was issued directly or indirectly, the DSD encoding in conjunction with the "non-audio format" which was used for file storage on the multi-channel audio workstation caused a very brief, low-level crackling sound. A similar sound also occurred in PCM mode at similar moments, but was subtly different sounding. It could probably have been avoided only by fading in and out in the digital domain—but any such fades would have required converting the DSD internally to multi-bit format at the most crucial stage, thus contradicting the fundamental design approach of the experiment, and in any event would have been extremely difficult given the non-audio data storage format. Despite intensive work on this problem, a restructuring of the originally planned computer arrangement and consultation with some of the firms which supported the listening tests, a solution by means of additional hardware or software was not obtainable at the time. Evaluation of the test subjects' questionnaires showed that this noise was not consciously perceived by any of them, and that the test does not lose any validity on account of it.

If one considers the test results with Treisman's "Suppression Theory of Selective Auditory Attention," which is based on perceptual psychology and is recognized today as the most far-reaching approach, sonic elements such as crackling sounds might be regarded by a Tonmeister as valid semantic content, and thus might influence a decision-making process either consciously or unconsciously. This level of importance could have been in effect particularly in this case: All four of the test subjects whose scores were in the range of critical probability were students in the Tonmeister course; all were aged 25 – 28; and tellingly, all auditioned their music examples over headphones.

It should be emphasized, however, that the validity of these four subjects' results is not in any question whatsoever, and the descriptive evaluation of this experiment does not depend on any special explanation of their scores.

Thus the only conclusion that can be drawn by evaluating the test results is that in four cases, within the critical range of probability the hypothesis "H" (that no perceptible differences exist between sources A and B) can be rejected on the basis of the previously stated decision rule, and the contrary hypothesis "G" (that there are perceptible differences between A and B) can be assumed. One could take the viewpoint that the test subjects in these cases perceived a difference between A and B.

On the other hand, for 141 of the 145 test scores (97.24%) hypothesis "H" cannot be rejected; in these cases one could assume that a difference between sources A and B was not perceived by the test subjects.

Figure 9a shows quite clearly how the score distribution for each music example hovers around the 50% (chance) level for the stereo examples. The surround examples (Figure 9b) show this even more clearly. This observation is reinforced if the total of all correct answers is compared with the total of all incorrect answers: Of a total 2,900 choices (145 test sequences × 20 choices per test sequence) there were 1,454 correct choices and 1,446 incorrect ones (see Figure 10). This result comes remarkably close to that which would be expected (arithmetic mean value of 1,450 correct and 1,450 incorrect responses) in a statistically "pure chance" experiment. The four extra correct choices (not to be confused with the four test subjects who attained critical probability with their test scores) represents a deviation of only 0.28%. Even with signals that had very short rise times (percussion and harpsichord), the digital encoding methods of the sources could not be distinguished from one another.

7. SUMMARY

These listening tests indicate that as a rule, no significant differences could be heard between DSD and high-resolution PCM (24-bit / 176.4 kHz) even with the best equipment, under optimal listening conditions, and with test subjects who had varied listening experience and various ways of focusing on what they hear. Consequently it could be proposed that neither of these systems has a scientific basis for claiming audible superiority over the other. This reality should put a halt to the disputation being carried on by the various PR departments concerned.
Only four of the 145 completed tests, or 2.76%, yielded results within the range of "critical probability" (i.e. less than 5% probability of guesswork). These four tests were conducted with two-channel listening material which was played back through headphones, while in the 100 completed tests using surround recordings, not a single test result achieved the critical probability level.

Though less readily formulated with mathematical equations, the high level of frustration felt by many subjects during their tests left quite a strong impression. These people, for the most part, were well accustomed to critical listening on a professional level, but they found that they could not even begin to recognize any sonic differences. A further frequent topic in personal conversations right after the test was the appearance of "would-be" differences—sonic illusions, so to speak. That is an issue which certainly has special importance to the Tonmeister; there needs to be a specific personal clarity concerning its causes. Developing a thorough understanding of this theme should profit both the musical and the sonic/technical aspects of a Tonmeister's work.

8. REFERENCES

[1] EBU European Broadcasting Union. *Listening conditions for the assessment of sound programme material: monophonic and two-channel stereophonic*. EBU Tech. 3276 – 2nd edition: 1998.

[2] ITU Radiocommunication Assembly. 1994. *Multichannel stereophonic sound system with and without accompanying picture*. Recommendation ITU-R BS. 775-1: 1992-1994.

[3] ITU Radiocommunication Assembly. 1997. *Methods for the subjective assessment of small impairments in audio systems including Multichannel Sound Systems*. Recommendation ITU-R BS. 1116-1: 1994-1997.

Detailed data as well as full-color versions of the graphics used in this paper can be found in the complete version of the Master's thesis from which it was drawn, on **http://www.hfm-detmold.de/hochschule/eti.html.**

Translated by **David Satz**.

Contacts: Dominik Blech – d.blech@gmx.de
Min-Chi Yang – mickyyang@msn.com