

WHICH OF THE TWO DIGITAL AUDIO SYSTEMS BEST MATCHES THE QUALITY OF THE ANALOG SYSTEM?

Wieslaw Woszczyk*, Jan Engel**, John Usher*, Ronald Aarts***, Derk Reefman***

* Centre for Interdisciplinary Research in Music, Media and Technology (CIRMMT) & Graduate Program in Sound Recording, Schulich School of Music of McGill University, Montreal, QC, Canada H3A 1E3

** Centre for Quantitative Methods CQM BV, 5611 BK Eindhoven, the Netherlands

*** Philips Research, High Tech Campus, 5656 AA Eindhoven, the Netherlands

Emails: wieslaw@music.mcgill.ca, Jan Engel al_engel@natlab.research.philips.com, John Usher jusher@po-box.mcgill.ca, Ronald Aarts ronald.m.aarts@philips.com, Derk Reefman derk.reefman@philips.com

Abstract

In this listening test, two digital audio systems (X and Y), and one analog system (A) were tested by 10 test persons who listened to a surround sound scene “live” (without recording). The main question to be answered was: “Which of the two digital systems best matches the audio quality of the analog system?” Both digital versions had 24bit dynamic resolution but differed in sampling rate with which the analog signal was sampled. One version (Y) was sampled with a CD rate of 44.1 kHz, the other (X) 8 times faster. There were also two test conditions, where in one condition there was a bandwidth cut off at 20 kHz instead of the 100 kHz that was possible with special 100 kHz microphones and added super-tweeters. For each subject, the experiment was replicated six times in random order, in each of the two conditions. The outcome of each experiment was a 0 or 1, where the 1 means that the, technically best, digital system X has been chosen as meeting the analog quality. The paper describes the test and the outcome.

Introduction

The audio archiving community responsible for the preservation of our sonic cultural heritage is interested in adopting a digital conversion and storage format that can be considered transparent by listeners skilled in the art of audio. Therefore, a digital medium having high degree of fidelity to the analog reference is needed. While there is a considerable debate in the audio community about the relative merit of high-resolution audio formats currently in use, compared to the CD format, only a few comprehensive listening tests were performed to verify the perceived sonic value of high-resolution audio. The opponents of high-resolution enlist the increased cost of storing high-resolution data and the lack of human ability to hear high frequencies. However, the international recording and mastering community reveals a growing appreciation of high-resolution audio

provided by high sampling one-bit and multi-bit systems. This appreciation must relate to a number of perceived characteristics other than high-frequency response partly because microphones and loudspeakers normally used in studios do not have a substantially wider response than the audible range. This listening experience is based on many hours of repeated sound evaluations performed in working, not double-blind, experimental conditions of a professional studio.

In the last few years, a number of complex listening tests were performed at the NHK Science and Technical Research Laboratories in Tokyo, Japan [1,2] to identify the importance of very high frequencies (over 21 kHz) in the discrimination of high sampling rate recordings. These tests were performed using two-channel stereo recordings of many different program types some containing large magnitudes of high-frequency energy. In double-blind tests, casual and professional listeners could not reliably identify high-bandwidth and high-resolution (192 kHz 24bit versus 48 kHz 24bit) conditions.

The listening tests arranged for this investigation described here were conducted at Philips Research in Eindhoven and were aimed to answer additional questions:

- (1) Is there a possible benefit of high-resolution audio at lower frequencies without the necessity for the reproduction of supersonic components?

- (2) Is a noticeable benefit of high-resolution revealed in surround sound listening?

The test – experimental setup

The tests consisted of a real-time comparison of an audio scene captured in an anechoic chamber and presented live to each subject in surround sound over six loudspeakers in a listening room. The subjects were comparing live analog reference with two digitized presentations of the scene made with two separate analog-to-digital and digital-to-analog conversion systems (X, Y) working simultaneously: (X) with high, and (Y) with low sampling rates.

Two identical high quality 8-channel **Analog-to-Digital-to-Analog Converters (ADAC)** were used in addition to a high quality 6-channel analog monitor selector and level controller having four groups of six-channel inputs and one six-channel output. The converters each generated a 5 bit delta-sigma modulated signal sampled at $128 \times F_s$ (128 times 44.1 kHz) from which 24 bit resolution signals were down-sampled to DXD format of 352.8 kHz ($8 \times F_s$) and to 44.1 kHz ($1 F_s$) by decimation. DXD is a PCM format jointly developed by Merging Technologies and Denmark Audio Digital to allow multi-bit editing of DSD 1-bit signals, and to offer low-noise recording at high frequencies, wide dynamic range (120dB), flat frequency response to above 100 kHz, and a narrow impulse response due to

the relaxed ADC anti-aliasing roll-off filter.

The EMM Labs Switchman 6-channel monitor level controller and selector was used to adjust audio levels with digitally controlled analog circuits via a wired remote and to select among 4 sets of 6-channel inputs (the fourth set was a mute with no input signal).

The audio scene was captured in the anechoic chamber using high quality microphones and microphone preamplifiers. Each amplified microphone signal was delivered to two electronically balanced and isolated outputs (XLR and DB25, see Appendix 1) of Grace Design Model m802, 8-channel preamplifier, and each output was connected to two loads in parallel to feed the following devices equally:

- EMM Labs Switchman master volume controller (A)
- DAD Sphynx 2 Converter set to DXD sampling (X)
- DAD Sphynx 2 Converter set to 44.1 kHz sampling (Y)

The Switchman was arranged to allow the selection of the above sources as: A, B, C. Selection D was not assigned and provided a complete mute when activated.

The position of Converter X and Converter Y was not assigned permanently to B or C, but it was

switched between them randomly according to a predetermined pattern. The switching was done by a human test-facilitator (test manager) from an isolated control room using DADman PC GUI software controlling the converter remotely between each repeat, and for each listener. It effectively disassociated the selector position (B, or C) from specific sampling frequencies, and from a subjective quality judgment. Position A was always the live reference. Position B and C was always digital and unknown sampling rate. There was no perceived latency difference between the three versions so listeners could not use latency to determine which of the sampling frequencies was used.

Listeners compared two digital versions of the scene with the reference analog version, considered to be a “live reference”. Their task was to specify which of the two digital versions had the most identical sound to the analog reference. The listeners were not to express preference for liking any particular version, they were simply asked to decide which of the two digital representations sounded most like the specified analog reference.

Two test conditions: C1 and C2 bandwidth

There were two separate tests conducted, each providing a different bandwidth condition. Test Condition C1 applied a 100 kHz audio bandwidth, whereas Test Condition C2 had a bandwidth cut-off at 20 kHz by virtue of microphones and

loudspeakers used that limit audio bandwidth to 20 kHz. In condition C1, Sanken CO-100k microphones (flat to 100 kHz) were used to capture the sources, and ribbon super-tweeters to reproduce them. In condition C2, measurement microphones limited to 20 kHz, and standard studio loudspeakers were used, with super-tweeters not active.

The test subjects

Ten listeners, male and female, age between 20 and 30 years, took part in the test on a voluntary basis. All expressed strong interest in the test outcome, and all had prior listening experience making evaluations of sound in high-resolution or low-bit-rate systems. Some subjects were accomplished musicians, or had musical and technical training, and all were on staff in the Acoustics and Signal Processing group at Philips Research.

Each listener specified which one of the two digital presentations sounded most like the analog reference A. Each time a decision was made, the test facilitator would change the assignment of the sampling rate to each of the converters, and would indicate to the subject using an intercom that the next sound setting was ready for the evaluation.

The sound scene - mechanical performance

The continuous sound scene created in anechoic chamber was mechanical in nature, powered by an electric motor driving a bicycle wheel. The intention

was to provide a sustainable constant level of audio “performance” over as long as is required to conduct the test. By not depending on live musicians, mechanical performance was able to maintain the same drive and intensity of execution for several days of testing, and for all listening subjects.

The mechanical nature of the sound scene and the repeating mechanism powering the scene caused the sounds to be regularly spaced, although not identically each time. This required listeners to focus their attention on various sounds and contexts presented in the scene as their relationships were constantly changing.

The emotionally neutral mechanical character of the music was intended to suspend the “musical brain” of listeners from influencing their evaluations. Only the assessment of sound quality (its fidelity to the known reference) was asked for, hence, musical preferences of test subjects were not to play a dominant role in the listening comparison of various sounds. As there is evidence of strong influence of content and expectations on the evaluation of audio quality (Gustavino and Katz, 2004), our test scene was to provide little support of these characteristics. Listeners’ preference of the music and its interpretative delivery was not to affect the judgment of the archiving quality of the recording medium.

The sound sources and the musical program material

The sound scene was made up of several distinct auditory objects distributed in space enveloping the listener. The rather quiet sound sources were arranged in the anechoic chamber that has an extremely low level of background noise. The following sound sources were used:

- Small copper bell suspended on a rubber band and activated by the wheel
- Bamboo wood chime moved by the air motion from a fan
- Loose strips of aluminum foil moved by the air flow from a fan
- Plastic wires installed near the spokes scratching the misaligned wheel
- Low-frequency pulsed sound from repeated silent groove of an LP record reproduced by a pair of subwoofers
- Water bubbling in a metal can being pushed by the air from a noiseless pump

The electric motor used to activate the sounds had its speed adjusted to create a relaxing musical tempo of the ever-evolving soundscape, which could hold the interest of listeners without being annoying or boring.

Presentation of the sound sources in the scene

The test scene was rich in transients, broad in bandwidth, and consisting of many events with distinct tonal, spatial,

and dynamic character. Each loudspeaker channel carried a dominant image of a specific source from the scene with some spatial (stereophonic) overlap between them. Altogether the listening impression was that of being immersed in a wide surrounding image filled with sound objects, some of which were elevated.

Listening room conditions

Listeners were presented with 6-channel surround sound projected from 4 loudspeakers in the horizontal plane (Left, Right at $\pm 30^\circ$; Surround Left, Surround Right at $\pm 110^\circ$) arranged as per BS.1116 without the Center loudspeaker, and with two additional Wide Left and Wide Right channels at azimuth of $+60^\circ$ and -60° that were elevated to $+30^\circ$ above the horizontal (2.35m above the floor). Listeners' ears were approx. 1.2m above the floor. All loudspeakers were at equal distance of 2.25m to the listener. The elevated loudspeakers were custom made flat ribbon drivers and the four horizontal loudspeakers were B&W802 Nautilus. The front loudspeakers were 1.6m and the surround loudspeakers 1.2m away from the nearest walls.

The listening room measured 9.0 x 5.7 x 3.5 (LxWxH) meters and had its width tapered along the front-back dimension, being narrower in the front (5.1m) and wider in the rear (6.6m). The surround listening set up was arranged slightly towards the narrower front portion of the room, 0.4m from the room's center. Having a fairly dry acoustics, the room

was originally designed to represent an average carpeted living room of generous dimensions.

Listening evaluation

Listeners were able to choose any one or any number of sonic elements included in the scene in order to make decisions about the matching of sound quality between the three transmission systems. This approach allowed each listener to develop a personal point of interest focusing attention towards specific items within the complex spatial soundscape. Each listener could set the loudness level according to a personal preference, and could listen as long as needed in order to make judgment with a degree of certainty. Listeners had to make a choice (forced choice) even if they were uncertain that the choice was correct or appropriate.

In this test, we have made a general assumption that the high-sampling version was closer in quality to the reference system than the low-sampling version, simply on the basis of it having 8 times greater number of data points in the conversion from analog, and without regard for the perceived difference between these digital formats.

The questionnaire presented to the listeners in the test is included in Appendix 2

Analysis of experimental results

Factors in the experiment

In this section, we analyse the data obtained from the listening test. Firstly, a summary of conditions and systems (factors) that are included in the listening test is given in Table 1.

Factor	Values
1. System	A, X, Y
2. Test condition	C1, C2
3. Subject	10 subjects
4. Replicate	6 replicates

Table 1. Factors included in the test.

Most important factors in Table 1 are System and Test condition. In this text, system A represents the analog system, X is the high-sampling digital system and Y is the low-sampling digital system. Furthermore, condition C1 applies a 100 kHz bandwidth, where condition C2 has a cut-off at 20 kHz bandwidth by virtue of microphones and loudspeakers used that limit audio bandwidth to 20 kHz.

Subjects and Replicates are needed for data gathering and are not of much interest by themselves. However, the variability observed within and between subjects is important for the conclusions we draw from the data, as it affects statistical testing results and length of confidence intervals.

Statistical model for system choice

The statistical model for the experimental data is fully presented in Appendix 3 and here we only give the main results. To find the statistical model, the following two assumptions have been made:

Assumption 1

There is a one-dimensional latent (non-measured) variable that denotes the perceived quality of the audio systems. This variable determines the outcomes of the listening test for the subjects [3].

Assumption 2

Both digital systems X and Y have a performance that is not better than the performance of the analog system A, if we measure this performance on the one-dimensional latent variable.

The following may serve as an argument for Assumption 2. The subjects were told that A is an analog clean and direct feed of the signal, while both X and Y were processed digital versions. The processing was only an A-to-D-to-A conversion of data, not a purpose made enhancement of the signal.

In the Appendix 3, we firstly find a statistical model for the latent variable from the Assumption 1. This model includes all factors in the listening test. Secondly, from this first model we obtain a second model and this is for the response data that has been obtained from the test.

We define p_k as the probability that digital system X is judged as *most alike* to the analog system A, under test condition Ck, for $k = 1, 2$. The statistical model for p_k from Appendix 3 is:

$$\Phi^{-1}(p_k) = \Delta + \Omega_k$$

for test condition Ck, $k = 1, 2$. We shall reformulate the above model as:

$$\Phi^{-1}(p_k) = \Delta + \Omega * Con_k,$$

(1)

where the variable Con_k has the value -1 for test condition C1, and $+1$ for test condition C2. This formulation simplifies the interpretation of the model parameters. In the model, Φ is the cumulative normal distribution function. Then, Δ is the mean value of the transformed probability $\Phi^{-1}(p_k)$ over the two conditions, and Ω is the effect of the test condition on this transformed probability $\Phi^{-1}(p_k)$. In the next step, Δ and Ω are estimated from the data from the listening test, and the statistical significance of Δ and Ω is tested. The results are summarized in Table 2. Conclusions are presented in the next section.

Model parameter	Estimated value	Significance probability in testing hypothesis H_0 : model parameter = 0
Δ	-0.05	64%
Ω	0.24	4%

Table 2. Estimation and testing results for Δ and Ω .

From the estimated values of Δ and Ω we can find model predictions and 95% confidence intervals for the probabilities p_k that system X is judged as most alike to the system A, under both test conditions C1 and C2. The results are in table 3.

Test condition	Model prediction of p_k	95% confidence interval for p_k
C1	0.40	(0.32, 0.50)
C2	0.60	(0.50, 0.68)

Table 3. Model prediction and 95% confidence interval for p_k .

A graphical presentation of the 95% confidence intervals is finally given in Figure 1.

Software used for computation is S Plus version 5 for the Unix platform. The approach follows [4].

Conclusions from data analysis

1. The probability p_k under test condition C_k measures the chance that the *high* sampling digital system X is judged as *most alike* to the analog system A. The effect of test condition on this probability is statistically significant (4% significance, testing at the 5% level; see Table

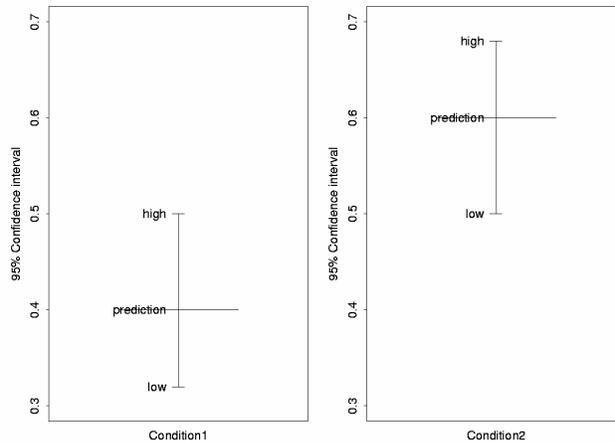


Figure 1. The 95 % confidence interval for p_k under the two test conditions (see also Table 3).

- 2). This means that probability p_k is different for both test conditions. We shall thus give results per each test condition.
2. *Results for test condition C1 (100 kHz bandwidth).* The probability p_1 is significantly *smaller* than 0.5; see Table 3. It is predicted as $p_1 = 0.40$. The 95% confidence interval for this probability is (0.32, 0.50); see Figure 1. This means that at the 100 kHz bandwidth the *low sampling* system Y is more often than X judged to be most like the analog system A.
 3. *Results for test condition C2 (cut-off at 20 kHz bandwidth).* The probability p_2 is significantly *larger* than 0.5; see Table 3. It is predicted as $p_2 = 0.60$. The 95% confidence interval for this probability is (0.5, 0.68); see Figure 1. This means that with the cut-off at 20 kHz bandwidth the *high sampling* system X is more often than Y judged to be most like the analog system A.
 4. *On average* over both test conditions, probability p is not significantly different from 0.5.
 5. Although in both test conditions the deviation of p from 0.50 is significant, the question is whether or not this deviation, which e.g. p is 0.68 at the maximum in test condition C2, is of *practical relevance*.

What is particularly interesting in these results is the indication that the higher bandwidth, beyond 20 kHz or so, may not be necessary to make the sound appear “analog” in quality. In the 100 kHz case (C1), people judged the low-sampling version as being more like analog. In the 20 kHz case (C2), they liked the high-sampling version, which would have an effect on other aspects of technical quality than the high frequency response. This may indicate that we do not need 100 kHz microphones and 100 kHz loudspeakers, but we do need 100 kHz capable recorders.

Conclusions

The results of this test indicate that listeners more often than not identify high-resolution audio as being similar in quality to the unprocessed analog audio. This conclusion, based on listening to the audio scene captured and reproduced with microphones and loudspeakers limited to 20 kHz bandwidth, indicates that high-sampling conversion system seems to be more transparent and provides a higher degree of fidelity to the analog reference.

Our listeners also reported that when they were listening to the wider bandwidth, up to 100 kHz, of analog audio converted to digital, they would choose the low-sampling rate of digital audio as sounding more like the analog transmission.

These results seem to indicate that the ultra high-frequency content may not be

necessary to reproduce audio that sounds more transparent. Supersonic content may contain noise-like artefacts that interfere with the perceived transparency of audio. However, to achieve a higher degree of fidelity to the live analog reference, we need to convert audio using high sampling rate even when we do not use microphones and loudspeakers having bandwidth extended far beyond 20 kHz. Listeners judge high sampling conversion as sounding more like the analog reference when listening to standard audio bandwidth.

These results suggest that the archiving community should consider using high-sampling conversion to ensure transparency even if the recording is made with standard audio-bandwidth transducers, and when digitizing older recordings made with bandwidth-limited analog systems.

The amplifiers in our tests system were always broadband, and no filtering of audio was done other than by using microphones and loudspeakers with bandwidth limited to approximately 20 kHz.

An improvement over this multi-object scene would be to arrange a single-object scene that uses multiple surround channels to enhance the perception of a single source to improve scene's readability. When listening to different transmission systems, decisions should be based on the amount of detail listeners are able to recover from the

scene. The future experiment will thus use a simpler surround scene presenting a live musical performance of a single sound source in a live room.

Discussion

The results indicate that in some cases low-pass filtering of audio data stream may produce an improvement in perceived quality of the content. Our listeners have more frequently correctly identified the higher resolution digital conversion system when the audio bandwidth was filtered down (by using standard bandwidth microphones and loudspeakers) than when it was extended. This is not entirely unusual if the ear considers the high frequency information as noise or artefacts, which do not contribute to the recognition of the qualitative aspects of the scene.

There was also the evidence of listeners' expectation influencing their sound quality judgment. The subjects knew that audio A was the analog reference and X and Y were the unknown digital versions. Some listeners commented that both X and Y were inferior to the analog reference A, and that this has made it difficult for them to specify which of the digital versions sounded like the analog version, as they seemed different from it. We do not know whether their judgment would be the same if they did not know that A was the reference (hidden reference). There were also listeners who commented that all versions sounded almost the same and that it was difficult to hear any difference. We

attribute this to the lack of experience and the difficulty level of the test.

The subjects generally appreciated the surround quality of the music and sound, and the freedom to choose any sound in the scene and any direction of listening, as long as they stayed in the sweet spot, for making the evaluation. They were also free to choose their preferred listening level, as long as it remained constant during all selections in the test, and to take as long as needed to make the evaluation. This accommodation aimed to make listening conditions more personalized and familiar to each subject. It usually took our subjects between 30 and 50 minutes to run 6 independent comparisons. They had no contact with the test facilitator during that time except when using the intercom to indicate readiness for the next test, and to verify the test number. Listeners seemed to use similar comparison techniques, taking longer to listen to the reference, and switching often between the unknown digital versions, as well as between the analog and the digital versions. The test manager could see this by monitoring the lights of the selected inputs on the monitor switcher.

A few listeners felt that the assessment would have been easier to perform if the music had exact repeats and predictable sonic events appearing always within the expected temporal context. As mentioned, the music and sound events of the audio scene were always evolving and not appearing exactly as predicted,

or at the predetermined regular time. Listeners had to memorize the quality of the objects but could not rely on the memory of the context and exact repeatable flow of the events. We felt that we did not want the judgments to be made based on a looped recording that is exactly replayed each time around, because this is hardly the condition found in normal listening to music. We wanted the music to have the element of freshness and surprise, and avoid perceptual adaptation that lowers listening sensitivity.

One can argue that providing unfamiliar music or sonic environment has made it more difficult for listeners to effectively judge audio quality (and the archiving quality of the digital medium) because they could not use the experience of and reliance on cognitive aspects of music. The archiving quality (fidelity to a live reference) could perhaps be evaluated more accurately when subject could appreciate the contribution to, or distraction from, the values inherent in communication of music. In this experiment, we decided to allow only the sound quality judgment, not the musical judgment, to play a role in the evaluation. The extreme case in this direction would be to make the music so repeatable and boring that it would inadvertently help subjects to focus on the inspection of sound quality. This may happen in tests allowing subjects to loop indefinitely a short fragment of previously recorded music until they are able to detect the slightest audible difference. By suppressing musical

listening, the method amplifies subjects' attention to technical parameters of sound.

In summary, the evaluation of the sound quality of digital conversion (its sonic fidelity to the analog reference) using the arranged scene was difficult. It could be that the scene provided too much information for listeners to process, requiring them to detect and recognize unfamiliar sources, memorize contexts, compare. [5] showed cases where "too much information" reduced the capacity of human listener and affected test results. This could have been the case here since the scene presented several sonic events occurring simultaneously. Additional channels in this case did not help to articulate the source but to detract from it, obscuring it with unwanted information. This might have been a case of providing too much information, or too much noise and too little information. There was perhaps less integration of information into a unified object that could guide the listener's attention and simplify their search and evaluation process. To verify this hypothesis, the subsequent test will involve a single musical instrument providing coordinated spatial, temporal, and timbral cues in surround mode of presentation with height.

ACKNOWLEDGMENTS

The authors would like to acknowledge the help of Signal Processing and Acoustics Group at Philips Research, including researchers and students who participated in

the listening tests, and the engineers Jan Witkamp, Leo van den Besselaar, and Theo Loring who were very helpful in constructing the auditory objects used as sound scene for the listening test. Also acknowledged is the assistance of these companies: Merging Technologies, Digital Audio Denmark, Grace Design, and Sanken Microphones. The first author is grateful for the financial support from VRQ, NSERC, CFI and James McGill Chair.

REFERENCES

- [1] Nishiguchi, T., Iwaki, M., Hamasaki, K., Ando, A., Perceptual Discrimination between Musical Sounds with and without Very High Frequency Components, Convention Paper #5876, Presented at the 115th Convention of the Audio Engineering Society, September 2003.
- [2] Hamasaki, K. and Nishiguchi, T., Differences of Hearing Impressions Among Several High Sampling Digital Recording Formats, Convention Paper #6469, Presented at the 118th Convention of Audio Engineering Society May 2005.
- [3] Thurstone, L.L. A law of comparative judgment. *Psychological Review*, 34, pp. 273-286, 1927.
- [4] Rajae-Joordens, R. and Engel, J. (2005). Paired comparisons in visual perception studies using small sample sizes. *Displays*, 26, pp. 1-7.

[5] Guastavino, C and Katz, B.F.G., Perceptual evaluation of multi-dimensional spatial audio reproduction, The Journal of the Acoustical Society of America, Volume 116, Issue 2, pp. 1105-1115, August 2004.

- FIGURES -

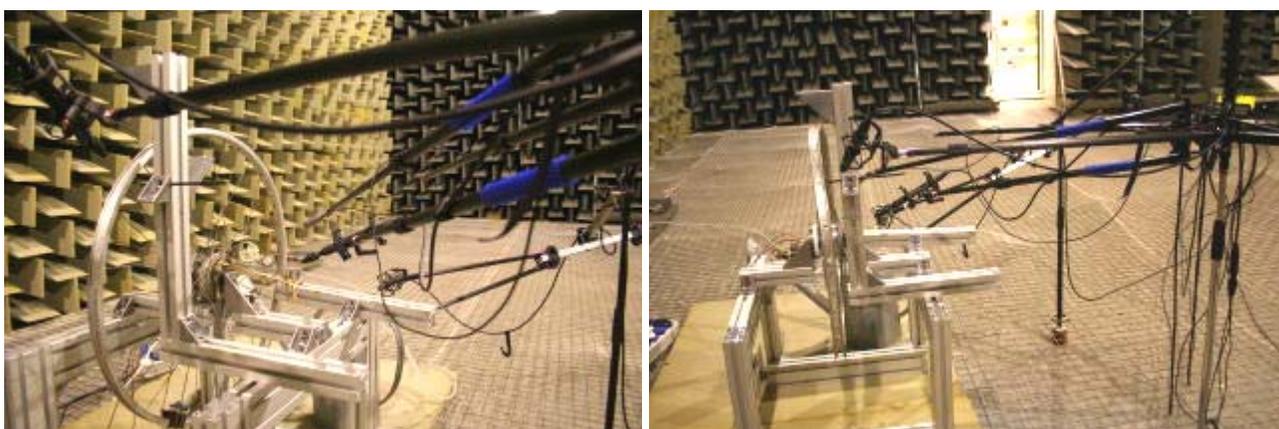


Figure 1

- Left: Bicycle wheel and sound making objects with an array of microphones in the anechoic chamber.
- Right: Slightly twisted shape of the wheel triggers sounds at time intervals related to the speed of rotation

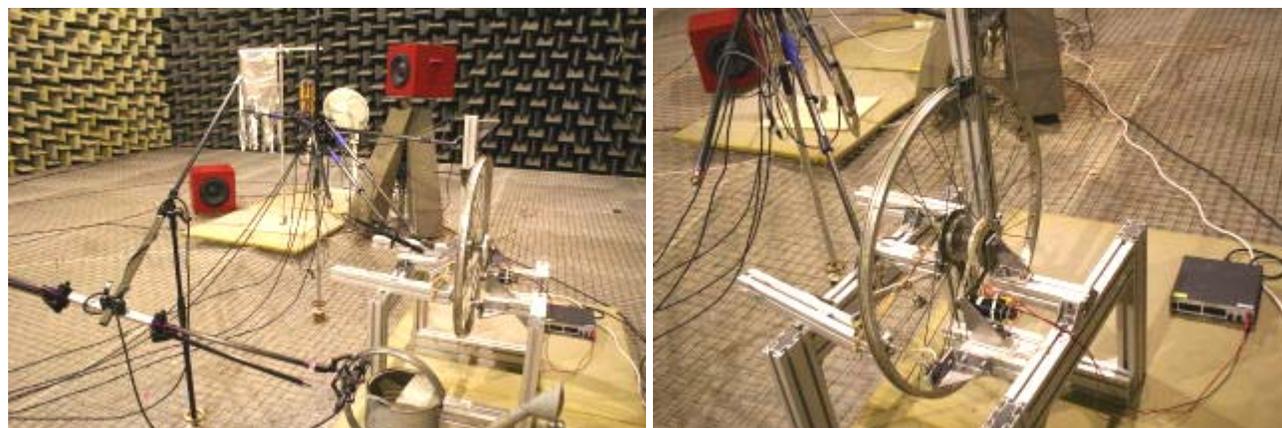


Figure 2 –

-- Left: The collection of sound-making devices arranged in the anechoic chamber to produce the sound-scene used in the two tests. From near to far: the water-can, the bicycle wheel assembly, two woofers, an aluminum sheet and a wooden wind-chime with an air-fan. Microphone systems are also shown. Not shown is the turntable.

-- Right: The bicycle wheel assembly with the power supply regulating the rotation speed. The stand supporting the microphone array is mounted independently from the wheel assembly in the anechoic chamber. The woofer, the wiring, and the microphone assembly visible in the background in the anechoic chamber.



Figure 3

- Left: The surround listening room at Philips Research with 5 B&W802 Nautilus loudspeakers (4 were used) and two elevated ribbon speakers (one is visible behind the subject).
- Right: The super-tweeter mounted on the top of the B&W802 Nautilus loudspeaker to extend the bandwidth of reproduction to 100 kHz.

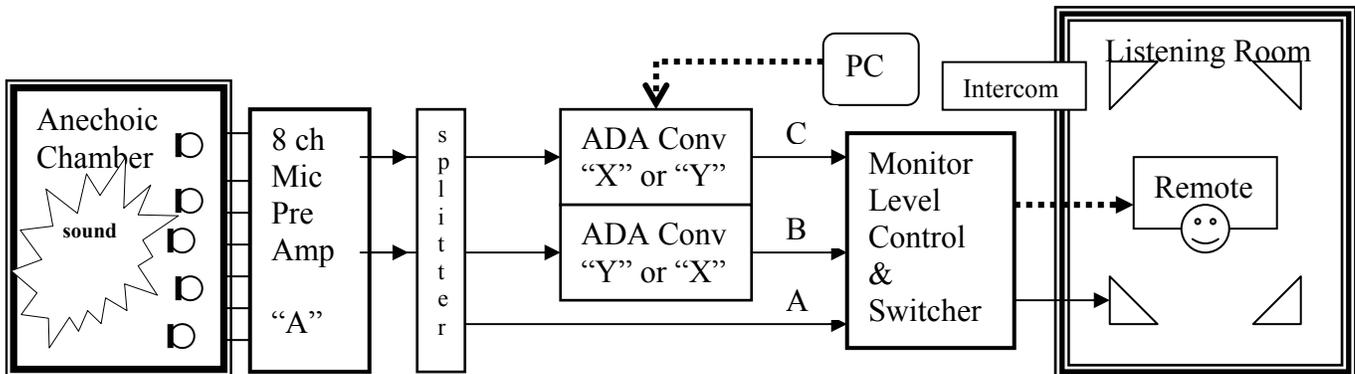


Figure 4

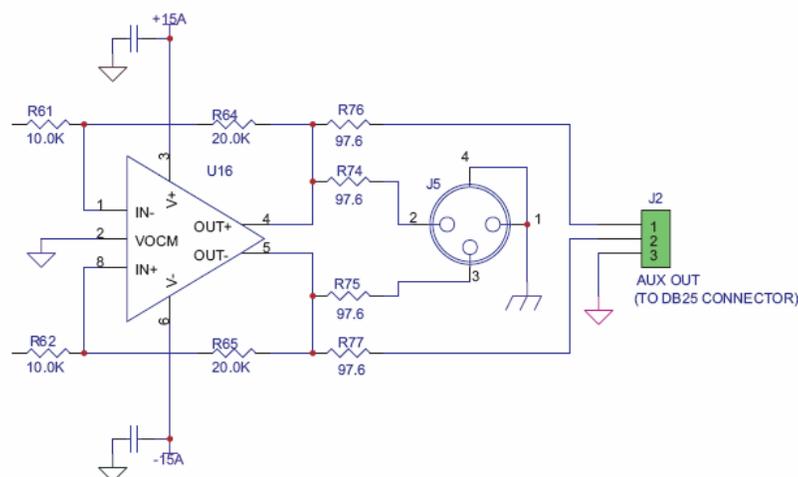
The block diagram of the test set up at Philips Research providing a listening comparison of three presentations (A, B, C) of a sound scene captured in the anechoic chamber and reproduced in the listening room in 6-channel surround with elevation.



Figure 5

– Left: The front view of Merging Technologies (and DAD) Sphynx II A-D-A 24 bit converter used in two sampling frequencies, 44.1 kHz, and 352.8 kHz, during the listening tests.
 -- Right: The GUI of DADman, a PC software developed by Digital Audio Denmark used to remotely control the sampling frequency of the Sphynx II converters.

Appendix 1



m802 OUTPUT SECTION

Appendix 2

Questionnaire

Dear Listener:

You are presented with a sound scene played by mechanical instruments. This scene will continue as long as you desire. You can choose to listen to it directly via the analog reference path, or via either one of the two digital paths. Your task is to identify which of the two digital paths sounds the closest to the analog reference in every detail of sound quality. Please take your time to learn the aspects of sound quality of the reference and of the unknown path in order to make an accurate choice.

Selector A -- Known Reference: analog sound

Selector B -- Unknown Digital version X

Selector C -- Unknown Digital version Y

You may listen as long as needed, and switch as many times between the sounds as required, before you make the judgment. It is recommended that you listen for a while and learn the characteristics of various sounds, then make a switch to identify the difference. You may compare the reference to B and to C, and you may compare B to C.

Please indicate your choice by writing your selection of B or C on the response

sheet, corresponding to the letter of the monitor selector you consider closest to the analog A.

You may add a comment about how easy or difficult it was to make a particular choice, or comment on the sound quality in general.

Thank you very much for your participation.

Appendix 3

Statistical Model

To derive the statistical model for the data analysis, we recall the Assumptions 1 & 2 introduced in the main text. These will be applied later in this appendix.

The first step is to find the statistical model for the assumed latent variable. This model is quite similar to the Analysis of Variance (Anova) model that is applied in experimental design when a variable is measured at the interval scale.

We now propose this model for the latent variable L_{ijkl} as follows:

$$L_{ijkl} = \mu + S_i + T_j + ST_{ij} + C_k + SC_{ik} + TC_{jk} + e_{ijkl} \quad (A1)$$

The terms in the model have the following interpretation:

μ = general mean

S_i = effect of subjects, a random term

T_j = contribution of audio system, $j = 1, 2, 3$ for the three systems, a fixed term.

We shall interpret system Y as $j = 1$, X as $j = 2$, A as $j = 3$.

ST_{ij} = interaction of subjects and system, a random term with variance σ_{ST}^2 .

C_k = effect of (test) condition; $k = 1, 2$ (fixed)

SC_{ik} = interaction of subject and test condition (random)

TC_{jk} = interaction of system and test condition (fixed)

e_{ijkl} = residual error of replicates, $l = 1, \dots, 6$, with variance σ_e^2 .

We assume all random terms in the model (A1) as normally distributed. The residual errors will be assumed to be independent, also from the other random terms. This is the usual Anova assumption.

The model (A1) has three main contributions (subject, system, condition). But also interactions may appear. It may happen that subjects have different opinions on the difference of the systems (interaction ST_{ij}). The perceived difference of the systems may depend on condition (interaction TC_{jk}).

In the second step, we find the model for the probability p_k that system X is perceived as being most alike system A, under test condition C_k , $k = 1, 2$. We find, under the Assumptions 1 & 2,

$$p_k = P_k("X = A") = P_k(A-X < A-Y) = P_k(X > Y) = P(L_{i2kl} > L_{i1kl}), \text{ where all computations are at the latent scale. This}$$

final probability is worked out as follows, using model (A1),

$$P(L_{i2kl} > L_{i1kl}) = P(T_2 + ST_{i2} + TC_{2k} + e_{i2kl} > T_1 + ST_{i1} + TC_{1k} + e_{i1kl}), \quad (A2)$$

where we only keep terms that are not the same at the left and right side of the $>$ sign. Using normality of the distributions of the random effects, we get the result

$$P(L_{i2kl} > L_{i1kl}) = \Phi\left(\frac{T_2 + TC_{2k} - (T_1 + TC_{1k})}{\sqrt{(\sigma_e^2 + \sigma_e^2 + \sigma_{ST}^2 + \sigma_{ST}^2 - 2\text{cov}(ST_{i1}, ST_{i2}))}}\right) =$$

$= \Phi\left(\frac{T_2 + TC_{2k} - (T_1 + TC_{1k})}{\text{constan } t}\right)$. We set the constant = 1 without losing generality and have the final result:

$$p_k = \Phi(T_2 + TC_{2k} - (T_1 + TC_{1k})).$$

So that

$$\Phi^{-1}(p_k) = T_2 + TC_{2k} - (T_1 + TC_{1k}) = (T_2 - T_1) + (TC_{2k} - TC_{1k}) = \Delta + \Omega_k \quad (A3)$$

The probability p_k of perceiving system X as the most alike of system A depends only on two parameters.

1. The parameter $\Delta = T_2 - T_1$ for the (*mean*) *difference* of systems X and Y;
2. The parameter $\Omega_k = TC_{2k} - TC_{1k}$ related to test condition Ck, k = 1,2. This parameter is completely determined by the *interaction* of audio system (levels T_1 and T_2 , so Y and X) and test condition (levels C_1 and C_2) in the model for the latent variable.

The model (A3) has been applied at the data analysis section in this paper.

Appendix 4

Experimental data

Subject	Test condition	Number of times system X has been chosen (out of 6 replicates)
1	C1	1
2	C1	2
3	C1	2
4	C1	2
5	C1	3
6	C1	2
7	C1	3
8	C1	3
9	C1	1
10	C1	4
1	C2	5
2	C2	4
3	C2	3
4	C2	4
5	C2	3
6	C2	0
7	C2	missing
8	C2	3
9	C2	5
10	C2	4