Phantom Source Perception in 24 Bit @ 96 kHz Digital Audio 4561 (I-2)

Bernd Theiß
University of Essex, Colchester, Essex CO4 3SQ, UK
Audio Physic Gerhard GmbH, 59929 Brilon, Germany

Malcolm O. J. Hawksford
University of Essex, Colchester, Essex CO4 3SQ, UK

Presented at the 103rd Convention 1997 September 26–29 New York

AES

AN AUDIO ENGINEERING SOCIETY PREPRINT

Phantom source perception in 24 bit @ 96 kHz digital audio

Bernd Theiß* and Malcolm O. J. Hawksford**

Abstract

The increased data capacity of DVD over CD has created an interest in extending the performance envelope of consumer digital audio to 96 kHz @ 24 bit. Two experiments to measure imaging are reported based upon cross-correlation of two noise sources and an estimate of minimum audible angle (MAA) for a range of stimuli. Results derived from 48 kHz and 96 kHz tests are presented to demonstrate whether the extended performance envelope improves the reproduction of phantom images.

* Centre for Audio Research Engineering, University of Essex, Colchester, Essex CO4 3SQ, United Kingdom and Audio Physic Gerhard GmbH, 59929 Brilon, Germany

** Centre for Audio Research Engineering, University of Essex, Colchester, Essex CO4 3SQ, United Kingdom

0 Introduction

The question as to whether sampling rates and resolution higher than red-book consumer standard of 44.1 kHz @ 16 bit offer a perceivable sonic advantage is currently a controversial topic attracting considerable debate. While one group advocates that the human hearing capability seldom approaches more than 18 kHz and that a 96 dB dynamic range is more than sufficient for typical music listening environments, the other group argues that the given limits are a contributory factor as to why music recorded on CD exhibits an unnatural character. In referring to a higher resolution standard we shall assume here the standard is 24 bit @ 96 kHz and refer to it as "high-definition digital audio", HDDA.

One of the areas which is often cited as exposing the deficiencies of restricted time and amplitude resolution is in the reproduction of the spatial attributes of the auditory event. These spatial attributes can be roughly divided into two categories, spaciousness[1,2] and localization accuracy[3]. Localization (accuracy) can again be divided into localization of single sound sources, localization of multiple sound sources radiating coherent signals and localization of multiple sound sources radiating partially coherent or incoherent signals.

To investigate the possible benefits of HDDA for music reproduction, the restricted case of two sound sources is of greatest experimental interest, where this study considers the azimuthal accuracy of localization as a function of sampling frequency.

1 General experimental conditions

To perform an investigation into the possible advantage of 96 kHz sampling rate over 48 kHz on human perception it is essential that the equipment and its interconnection is the same for all conditions under evaluation. This experimental strategy is important as there may be changes in perceived sound quality for a number of reasons where it is imperative that the hardware configuration and data transfer rates remain invariant so that the only performance changes are due to bandlimitation. Because of this restriction, computer generated signals (e.g. noise and pulses) with or without contributions above 24 kHz were used for most of the experiments. All signals were replayed at 96 kHz irrespective of the signal bandwidth. Consequently, secondary effects such as a change in jitter performance or a change in DAC hardware, amplification, cabling etc. used in a second equipment configuration to run at 48 kHz rate have been eliminated. However, in addition a flexible sampling rate converter with various word lengths and noise shaping options was used for those cases where a limitation of the frequency content of the stimulus was otherwise impossible.

1.1 Equipment

Digital signal sources consisted of a 96 kHz @ 16 bit DAT-recorder (Pioneer D9601) and a 96 kHz @ 24 bit tape recorder (Nagra D) running into a digital-to-digital sampling and bit-rate converter (dCS 972). A 48/96 kHz @ ≤ 24 bit digital-to-analogue converter (DAC), a dCS Elgar, served for transforming the data in the analogue domain. The analogue signal was directly fed into the experimental hardware necessary to control stimuli during the experiments. The output of the experimental hardware was subsequently fed into a power amplifier which drove the loudspeakers.

Harmonic distortion of the experimental set-up was more than 80 dB below the fundamental and consisted mainly of second harmonic at -93 dB and third harmonic at -86 dB for most combinations of frequency and output level. Bandwidth for the analogue electronics, except the DAC, was 20 Hz to > 100 kHz.

The loudspeakers used for this experiment were 1.06 m high and 0.26 m wide two-way, floor-standing (Audio Physic Tempo II), a two-way design using a 7.5" paper-cone bass-midrange driver and a 1" aluminium dome tweeter. To extend their bandwidth above the limits of the dome tweeter, a Technics EAS-10TH1000 ribbon tweeter was crossed over at 10 kHz / 12 dB per octave. Measurements of the resulting third-octave smoothed frequency responses at different heights (at -20 cm, -10 cm, 0 cm, 10 cm and 20 cm referenced to the height of an average seated subject) are shown in Figure 1. The measurements confirm that there is sufficient high-frequency output above 24 kHz for both above and below average height subjects, a prerequisite for the study of the influence of bandwidth limitation at 24 kHz on auditory perception. Harmonic distortion figures of the loudspeaker were 0.1% at 84 dB in the mid-band rising to 0.2% at higher frequencies, again mainly second and third harmonic.

1.2 Room and loudspeaker set-up

The experiments were conducted in a 5.78 m by 3.96 m by 2.4 m listening room, built after IEC recommendations. The acoustic treatment of the room consisted of nine bass absorbers (1.2 m by 0.3 m by 0.09 m elements of glasswool) to reduce room modes, four triangular cushions (0.38 m side length and 0.05 m thick) mounted in the upper corners to suppress a flutter echo and a 1.2 m wide and 2.4 m high coverage of the wall behind the listener with highly absorptive material. The frequency-dependent reverberation time T60 of the treated room, measured at the head position of the subject, is given in Figure 2.

To achieve optimum phantom image sharpness, the loudspeakers were placed according to guidelines given in[4], with an angle of ± 37 degree from centre front. The matching of the direct field of the loudspeakers at the subject's head position is demonstrated in Figure 3 as sum and difference of the impulse responses of both loudspeakers stimulated together. It is shown that not only the direct field but also early reflections cancel to a high degree in the left minus right impulse response. Hence both loudspeakers have exactly mirror imaged positions to the listener within the room.

The complete set-up of the room is shown in Figure 4. The electronic equipment necessary to conduct the experiment, except for the loudspeakers, was located externally in an office adjacent to the listening room.

2 Experiment 1: Partially coherent sound sources

The degree of coherence k between two signals can be described as the maximum of the normalized cross-correlation of the two signals. When k = 1 the signals are coherent, if they are reproduced over the loudspeaker layout shown in Figure 4, a sharp central image should be perceived. For k = 0 the signals are incoherent, consequently neglecting initially the influence of the room and interaural crosstalk, a subject should perceive two completely independent images under the same circumstances. For $0 \leq k \leq 1$ the size of the phantom image is expected to increase with decreasing k until at some value ($k < 0.4$ for headphone presentation) the phantom image splits in two[5]. Experiment 1 was designed as an interactive experiment where subjects were asked to adjust the degree of coherence between two noise sources to the point where a single but extended auditory event split in two. Figure 5 shows the signal processing hardware necessary to conduct the experiment. Two completely incoherent white-noise signals L and R are played back from the DAT-recorder. They are converted to a sum signal L + R and a difference signal L - R using a MS-matrix[6]. By means of a motor driven potentiometer, the amplification *a* of the difference signal can be varied between zero and one. The attenuated difference signal and the sum signals are converted to loudspeaker feeds L' and R' with an inverse connected MS-Matrix to give:

$$L' = (1+a)\ L + (1-a)\ R\ , \qquad (1)$$
$$R' = (1+a)\ R + (1-a)\ L \qquad for\ \ 0 \leq a \leq 1 \qquad (2)$$

For $a = 0$ the speaker feeds are totally coherent while for $a = 1$ they are totally incoherent.

2.1 Procedure

The subjects were seated in the totally dark listening room, having been instructed to adjust the degree of coherence from an initial value of one, via two push-buttons, to the point where the single phantom image split in two distinguishable images and then press an "OK" button after the task has been performed. Afterwards, the coherence was automatically reset to k = 1 and the experiment repeated a further nine times.

2.2 Results

Regrettably, no results can be published about the outcome of this particular experiment. It was discovered that significant variations spanning k = 0.01 to k = 0.07 could readily be experienced with a single subject in a single experiment sequence. Also, repeating this experiment did not show any kind of learning effect. Consequently, because of this unexpected variation requiring an unreasonable auditory memory, it was considered inappropriate to continue an experiment which probably would swamp any effect related to sampling frequency due to its inherent poor resolution.

3 Experiment 2: Minimum audible angle

A fundamental question in the research of spatial hearing is: What is the minimum displacement of a sound source that is detected as a displacement of the auditory event by a majority of subjects within a given group? To investigate this question of minimum displacement a two-interval forced choice procedure is a particularly useful strategy. A stimulus is presented to a subject first from a fixed position and then from a second position which has randomly changed. The change should be selected from one out of two alternative positions, e.g. to the left / right or to the front / back. The magnitude of displacement of the two alternative positions should be the same, directions should be to opposing sides of the initial position. A displacement of the second auditory event is accepted when 75% of the reports indicate the right direction in displacement, this is half-way between a random guess and certainty. The term used to describe this displacement is called "localization blur"[7] or in case of horizontal angular displacements "minimum audible angle" (MAA)[8]. The MAA is dependent on a number of factors, including the position of the sound source in relation to the subject, the type of signal radiated, the radiation pattern of the sound source in reflective environments, the subject itself, the mental state in which the subject is during the experiment, etc.. In case of two-channel sound reproduction a change of position of a phantom source can be introduced, for example, by a change in relative gain with which a signal is fed to the loudspeakers. It is not strictly correct to speak of the MAA in the case of phantom source reproduction, as a fixed change in gain can result in different perceived angles depending on the signal being transduced (see for instance[9]). If we refer to angles in subsequent text they should be, unless otherwise noticed, understood as derived from gain changes by means of the "law of sines"[10]:

$$\alpha = \arcsin\left(\frac{a_L - a_R}{a_L + a_R}\sin(\vartheta_0)\right) \tag{3}$$

where a_L and a_R are the gains in the left and right channel and $\pm\,\vartheta_0$ is the subtended angle between subject and loudspeaker. We do not imply that the law of sines results in the most accurate predictions of phantom image position, but it is the longest established standard for image control.

3.1 Procedure

A block diagram of the experimental set-up used to conduct this experiment is shown in Figure 7. The experimental procedure was designed to be straight-forward so as not to give undue stress to the subjects and experiments were conducted individually in a completely darkened room [1]. The signal under evaluation was

[1] One subject assumed to reach better results with the lights switched on in the room. A control experiment run with this subject clearly showed that this is not the case. Any visual distraction leads to a) much higher MAA's and introduces b) another time dependent variable, the relationship of brain processing power divided between listening and seeing.

played continuously from centre front for 30 second to allow subjects time to adapt to the localization task. The signal was then attenuated gradually to -80 dB, the direction changed, and the signal slowly returned to its nominal loudness. The fading sequence lasted about 1 s so as to avoid additional localization cues due to rapid changes in signal[11]. In each test, the subject was given 7 s to judge the change in direction and to record the results using a 2-push button interface to the computer. After a total of 61 sequence repetitions or less than 9 minutes the experiment was terminated.

The sequence of angular displacements commenced at 0° (centre front) for the 30 seconds adaption time followed by a displacement of ±1.6° for the first judgement which corresponds to 0.8 dB gain difference between the left and right channel, where ± means plus or minus. A 0.8 dB change was found to be audible during an informal study of level matching for stereophonic loudspeakers. The complete sequence of displacements as shown in Figure 6 are: 0°, ±1.6°, 0°, ±0.0518°, 0°, ±1.5482°, 0°, ±0.1036°, 0°, ±1.4964°........, ±0.8774°, 0°, ±0.7742°, 0°, ±0.8258°, 0°. This sequence allows every angle between 0° and 1.6° in 0.0518° steps to be evaluated independently. The interleaving of 0° phantom images ensures that the MAA is investigated for centre front even for a series of displacements in one direction, otherwise this could cause the whole image to drift asymmetrically. The interleaving of monotonically decreasing and increasing sequences minimises subject fatigue by reducing the number of imperceptible displacement changes within a given time.

The general direction of displacement (right or left) was determined by a pseudo-random noise generator and judgements were compared with the actual displacement changes and stored on a computer as correct {1} or wrong {-1} answers. If no judgement was made in the experiment window, this was stored as {0} and later considered as a 50% correct answer; fortunately this occurred only three times in the worst sequence and less than 0.5 times per 62 judgements on average. The experiment was repeated five times for each stimulus and subject, the first time being an "accommodation" or training experiment. The results from the subsequent four experiments were then recorded as data using either a sampling-rate sequence of {96 kHz - 48 kHz - 48 kHz - 96 kHz} for one half of the subjects or {48 kHz - 96 kHz - 96 kHz - 48 kHz} for the other half of the subjects. This strategy is designed to cancel out effects of learning.

To reduce experiment stress, subjects were allowed to decide individually how long the recovery period between the trials should be, where the average time was about five minutes. Head movements were not blocked but subjects were advised to hold their heads still during the experiment. To minimise recording a "no result" in those cases where subjects felt insecure about a directional change, they were advised to judge on instinct.

3.2 Subjects

Seven subjects participated in the experiments, six were men aged between 26 and 49 years and one was a women aged 30 (as of June 1997). Four subjects are employees of Audio Physic and have prior experience in listening tests[12]. The remaining three were experienced listeners who were invited to take part in this experiments. On a self-certification basis, none of the subjects reported hearing damage.

3.3 Stimuli

Three types of stimuli have been used for this investigation. The first was a computed pulse train with 0.1 s repetition rate calculated from,

$$p(k) = \sum_{k=0}^{9599} \sum_{l=0}^{n} (-1)^{l} \sin\left((2l+1)\,\pi \frac{k}{9600}\right) \tag{4}$$

$$p(k+i*9600) = p(k) \qquad \text{for } i \in \mathbb{N}+ \tag{5}$$

$$\text{with } n = 2398 \text{ (for 48 kHz) or } n = 4796 \text{ (for 96 kHz)},$$

and played back at an average level of 46 dBA, measured in the

24 kHz bandwidth limited mode. The second stimulus was a 1 s sequence of white noise, which was recorded in repetitive mode and calculated from,

$$wn_{48}(k) = \sum_{k=0}^{95999} \sum_{l=0}^{11999} \sin\left((2l+1)\,\pi\,\frac{k}{48000} + \varphi(l)\right) \quad (6)$$

$$wn_{48}(k+i*96000) = wn_{48}(k) \qquad for\ \ i \in \mathbb{N}+ \quad (7)$$

$$wn_{96}(k) = wn_{48}(k) + \sum_{k=0}^{95999} \sum_{l=12000}^{23998} \sin\left((2l+1)\,\pi\,\frac{k}{48000} + \varphi(l)\right) \quad (8)$$

$$wn_{96}(k+i*96000) = wn_{96}(k) \qquad for\ \ i \in \mathbb{N}+ \quad (9)$$

$$with\ \varphi(l) \in [0,2\pi],$$

where $\varphi(l)$ is random variable with a uniform probability distribution over a range 0 and 2π. It should be noted that the sequence containing frequency contributions above 24 kHz (wn_{96}; re 96 kHz sampling rate) is exactly the same below 24 kHz as the restricted bandwidth sequence (wn_{48}; re 96 kHz sampling rate). This approximation to white noise was played back at 68 dBA, measured in 24 kHz bandwidth limited mode. The third and final stimulus was a short excerpt from a piano concert of Brahms, played back at levels ranging from 50 dBA to 85 dBA. This HDDA Nagra D sourced recording was transferred to 96 kHz @ 16 bit and 48 kHz @ 16 bit DAT using triangular dither. The reason the Nagra D / dCS 972 combination was not used directly for this stimulus is discussed in Section 4.3.

Both the impulse train and white noise sequences were calculated on a computer and transferred to DAT-tape at half speed (48 kHz) from a professional SPDIF sound card (CardD from Digital Audio Labs) without any kind of dithering or noise-shaping.

3.4 Experiments

To explore whether there are statistically significant differences in localization accuracy between 48 kHz and 96 kHz sampled sounds, a t-test can be used. If the mean score of correct answers for each sequence yields a significance of t = 0.1 (i.e. a 90% probability that the result is not caused by the randomness of the distribution), we would consider the result convincing. The t-test is particularly suitable for experimental data based only on a limited number of subjects that are not selected to be a representative group of the population, as it is not very sensitive to distributions which are not normal-distributed[13].

3.4.1 Impulse train

All subjects participated in the experiment with impulses. Results of the experiment, i.e. percentage of correct estimations as function of angular displacement, are shown in Figure 8 both for 48 kHz and 96 kHz sampling rates, the average over both sampling rates is shown in Figure 9. Straight lines have been fitted using a least-square-error algorithm to obtain the values of MAA, which result in 1.08° for 48 kHz and 1.03° for 96 kHz. Mean scores of correct answers have been calculated as 69.6% at 48 kHz and 69.7 % at 96 kHz. The MAA of 1.05° is considered an extremely low value for the intensity-stereo arrangement in the reflective environment used. In the overview about "localization blur" given in[7], MAA's for impulses range from 0.75° to 2° where these values are obtained using natural sound sources in free-field environments. The hypothesis that a difference in localization accuracy exists due to differences in sampling rate was then tested using the t-test. The mean values of correct answers in each sequence for each subject at the different sampling rates were calculated and compared. However, a t-score of 0.0328 at 26 degree of freedom (m = 26; seven subjects times 4 sequences minus 2) shows that the difference between sampling rates is insignificant and the hypothesis is invalid.

In many psychoacoustic experiments subjects improve their discrimination skills over time[14], on the other hand in tests that are too long the stress induced by the task can also impair the performance of the subject. To evaluate if there are any differences due to learning effects or stress, the average of all sequences has been tested against each sequence taken on its own. With respective t-scores of 0.28, 0.52, 0.56 and 1.00 at m = 12 it can be concluded that a single training sequence before the session was sufficient to prepare the subjects for the experiment and that the experiment with five sequences was not too long as to impair the results. Another interesting question is, whether long-term experience in listening to stereo has improved the subject's localization ability? With a t-score of 2.48 at m = 18 this latter hypothesis is acceptable at a 0.02 confidence level. For the experienced listeners, analysis showed that the MAA improved to 0.83° (1.21°) and the mean percentage of correct answers also improved to 75% (65.7%), where the values for the inexperienced listeners are given in brackets.

3.4.2 White noise

Four subjects (no experienced listeners) participated in the experiment with white noise as stimulus. However, the hypothesis that a bandwidth limit of 24 kHz has a negative influence on localization must again be rejected. With a t-score of 1.1 at 6 degree of freedom this is well below the required t = 1.9 needed for a 0.1 confidence level. The mean score of correct answers was 61.1%, 4.6% lower than for the same group of subjects in the first experiment. The MAA, at 1.55°, is correspondingly higher than in the previous experiment, see Figure 10 for details. This is an indication that noise is more difficult to localize than a sequence of impulses, although at t = 1.32 and m = 30 not with a 0.1 confidence level, which would require t = 1.7. Interestingly, this MAA is significantly better than the 3.2°-result reported for broadband noise[7].

3.4.3 Music

Four subjects participated in the experiment using music as a stimulus, three were selected from the group of employees of Audio Physic, while the fourth was from the group of experienced listeners. The results of this experiment showing percentage of correct answers versus angular displacement, are presented in Figure 11. Again there was no confirmation for the hypothesis that 96 kHz sampling improves localization (t = 1.1 at m = 6).

It must be reported that the slope of the least-square-error fitted straight line for music is 8% per degree and much lower than the slopes for impulse trains or noise which are at 22% per degree and 18% per degree, which are remarkably similar. Comparing the mean values of correct answers (all subjects and sampling rates) between music and white noise gives at 61.6% versus 61.7% percent, nearly identical results. It is reasonable to conclude therefore, that although localization accuracy is generally higher with music than with white noise, either the dynamic changes in the music or the emotional involvement of the subjects heavily interferes with this particular testing scheme.

3.5 Discussion

The hypothesis that an increase in sampling rate will result in an increase in spatial resolution can be rejected from the experimental data given above. The estimated MAA's have been found to be comparable to those published in the literature[7], even when re-calculating the angles using the energy-vector theory[9], a value of 2.6° for impulse trains compared to 0.75° to 2° in the literature or a value of 3.9° for white noise compared to 3.2° in the literature are very low values, when the reflective environment or the fact that the sound was a phantom are taken into account.

4 Additional experiment into overall sound quality

As the relevant high-performance audio equipment was available, it was considered prudent to conduct an additional experiment to determine whether there is an observable change in sound quality between 48 kHz @ 16 bit compared to HDDA, and if so, whether there are common attributes recognized by all subjects. However, as there was insufficient equipment to run an ABX-test[15], the procedure was configured as follows.

4.1 Procedure

The subject was seated in the room and a piece of music (the Brahms piano concerto) was played at 48 kHz @ 16 bit with triangular dither. After the subject indicated an appropriate time had elapsed to acclimatise to the sound quality, a button was pressed to indicate to the operator that the next experiment phase could commence. The music was muted and the playback conditions changed to HDDA, the tape was rewound (all this took about thirty seconds with the Nagra D) and the same part of the music was played in HDDA format. With every press of the push button the condition of playback was changed. The next two changes to 48 kHz and HDDA were still known to the subject. Following this pre-conditioning sequence, a series of six trials followed in random order, for which the subject was advised to estimate and record the format in each trial on a piece of paper (with a electric torch because of the darkness) during the thirty seconds pause. After the sequence of six trials, the subject was advised to write down any overall observed differences before communicating with the operator. When the subject had recovered from the task, the whole sequence of four known formats and six random format changes was repeated. This task was repeated one further time.

It must be noted that this kind of test is more difficult for the subject than the commonly used ABX-test because it requires an acute aural memory. A trial where the subject is unsure about the actual stimulus (48 kHz or HDDA) can not be resolved by another quick comparison, but has to be either judged on the "feeling" or logged as "not sure".

4.2 Subjects

Choosing listeners for a difficult hearing experiment is not straightforward as there are no scientifically acceptable credentials for auditory discrimination capability or aural memory. The subjects were chosen here for their success and proven skill in setting up stereophonic equipment to very high standards without the help of measuring equipment; naturally this is a very subjective criteria. Two of the subjects were experienced Hi-Fi dealers with the nowadays rare practice of installing systems at their customers homes and using their extensive experience in judging sound quality. The other subject participating in this experiment had previously shown good discrimination capability with his contribution to beta-testing prototype loudspeakers.

4.3 Experiment

For the first subject the piano-concert of Brahms was played back on the Nagra D and either looped through the dCS digital-to-digital converter without re-sampling and bit-rate reduction or with reduction to 48 kHz @ 16 bit using triangular dither. The three sequences had the following experimental results:

Trial→ Sequence↓	1.	2.	3.	4.	5.	6.
1. Mode:	48 kHz	HDDA	HDDA	48 kHz	48 kHz	HDDA
1. Est.:	48 kHz	HDDA	HDDA	48 kHz	48 kHz	HDDA
2. Mode:	HDDA	HDDA	48 kHz	HDDA	48 kHz	HDDA
2. Est.:	HDDA	d. c.	48 kHz	HDDA	HDDA	HDDA
3. Mode:	48 kHz	HDDA	48 kHz	48 kHz	HDDA	HDDA
3. Est.:	48 kHz	HDDA	48 kHz	48 kHz	HDDA	HDDA

(d.c. = don't care / no estimation)

Neglecting the missing estimation in the second trial of the second sequence this gives sixteen correct answers out of seventeen. There is a chance of 0.014% to reach such a result by random guessing, so we can reasonably conclude that the subject was able to hear a difference. The subject recorded the following positive subjective attributes to HDDA:

> *"more open, less grain, more dynamic and better dynamic shadings, more focus, more body, more air, better ambience retrieval and the soundstage was more to the front."*

Attributed to 48 kHz were:

> *"graininess, the mid-range was more dulled and darker, the sustain of the piano was damped and the soundstage was narrower and more to the back."*

As this subject had taken part in an MAA experiment before but was interested to contribute further to this investigation, we decided to analyze whether the perceived changes could be attributed to sampling rate or to amplitude resolution or to a combination of both after a longer recovery break. Two experiments were planned for this second session: a comparison of 96 kHz @ 24 bit versus 48 kHz @ 24 bit and a comparison 96 kHz @ 24 bit versus 96 kHz @ 16 bit. Regrettably we can not give results on these experiments, nor on other experiments concerning higher than 16 bit resolution, as the dCS digital-to-digital converter intermittently refused to output data at 96 kHz, making any serious experiments impossible. However, the additional experiments were conducted with the same music but transferred to 16 bit DAT format using triangular dither at both the sampling rates under consideration. Subject 2 reached 61.1%, correct estimations in the experiment with no estimations missing, a result which can be reached or bettered by chance with a probability of 24 %. Attributed differences in favour of 96 kHz were,

> *"better retrieved onsets of quiet instruments coming into a melody line, more natural sounds from instrument, handles better retrieval of dynamic shadings and found it generally easier to follow melodic lines."*

Subject 3 reached 68.8 % correct answers, with two missing estimations, giving a 89 % confidence, and made comments on,

> *"enlarged body, the better floating melody lines and the greater spaciousness of the 96 kHz reproduction."*

It should be mentioned that subject 1 participated on this test on a Sunday morning, while the other subjects had only time in the evening after a day of work, and that pauses between two replays of the DAT-tape were longer because there was more winding and rewinding involved.

4.4 Discussion

While there is little doubt that subject 1 reliably heard a difference between HDDA and 48 kHz reproduction, results of subject 2 and 3 need closer evaluation. The probability that the results obtained from both subjects where randomly guessed is 6 %. It is reasonable to conclude that even in the 96 kHz to 48 kHz sampling rate comparison there was a perceivable difference.

Given the conversion accuracy of the dCS digital-to-digital converter (see Figure 12 for details), and taking into account that the digital equipment was of very high quality, it is fair to conclude that differences in the experiments can be attributed to the higher standard of HDDA (subject 1) and the higher frequency resolution of 96 kHz (subject 2 and subject 3).

5 Conclusions

Two experiments have been presented to investigate into "localization blur" as a function of the sampling rate of the stimuli. The first experiment which was concerned with localization of partially coherent signals did not reveal data exact enough for a quantitative evaluation. The second experiment to estimate MAA's in a standard stereophonic loudspeaker layout gave results stable enough for evaluation. Analyses of the data showed that the hypothesis that localization accuracy improves with higher sampling rates above the professional 48 kHz standard has to be rejected. Nonetheless, in a subsequent experiment, a task considered to be more difficult than the conventional ABX-test, one subject proved to be able to discern HDDA from 48 kHz @ 16 bit with a reliability not expected in such a difficult task. Two futher subjects were able to discern 96 kHz @ 16 bit from 48 kHz @

16 bit at more than 90 % confidence. The differences in sound quality between the two standards as noticed by the subjects, suggest that further experiments conducted in this area should focus on the finer structures of music. Dynamic shadings (contrast), ambience retrieval and distance perception seem to be good predictors worth further research. We can conclude by saying that the movement towards higher resolution in digital audio, at least at the recording studio level, is justified especially if we wish to prevent the finer nuances of contemporary music from being lost for future generations.

6 Acknowledgment

We would like to thank the many people whose contribution enabled us to conduct this investigation. Our special thanks go to Dr. Takeo Yamamoto for initiating this study and donating the 96 kHz DAT recorder, to Mr. Joachim Weber of Nagra Kudelsky GmbH for making the Nagra D available to us at a special rate, to Mr. Norman Chesky from Chesky records and to his German distributor Mr. Alex Manninger from Inakustik for making 96 kHz Nagra tapes available. We are indebted to Volker Ochsner and to Rudi Schurig, who travelled a long distance to participate in the experiments. Finally, we would like to thank Joachim Gerhard and Audio Physic Gerhard GmbH for the financial support of this work.

7 References

[1] Y. Ando, "Concert Hall Acoustics," Springer, New York, (1985)

[2] D. Griesinger, "Spaciousness and Localization in Listening Rooms and Their Effects on Recording Technique," J. Audio Eng. Soc., vol 34, pp 255-267, (1986 April)

[3] W.M. Hartmann, "Localization of Sound in Rooms," J. Acoust. Soc. Am., vol. 74, pp. 1380-1391, (1983 Nov.)

[4] B. Theiß, J. Gerhard & M. O. J. Hawksford, "Loudspeaker Placement for Optimised Phantom Source Reproduction," Preprint 4246 of the 100th Audio Engineering Society Convention, Copenhagen (1996 May)

[5] R.L. Chernyak & N.A. Dubrovsky, "Pattern of the noise images and the binaural summation of loudness for the different inter-aural correlation of noise," 6th Int. Congr. on Acoustics, vol. 1, Tokyo, A-3-12

[6] M.A. Gerzon, "Optimum Reproduction Matrices for Multispeaker Stereo," J. Audio Eng. Soc., vol. 40 no. 7/8, pp. 571-589 (1992 July/Aug.)

[7] J. Blauert, "Spatial Hearing," revised edition, MIT Press, Massachusetts, (1997)

[8] A. W. Mills, "On the Minimum Audible Angle," J. Acoust. Soc. Am., vol. 30, pp. 237-246, (1958)

[9] M.A. Gerzon, "General Metatheory of Auditory Localisation," Preprint 3306 of the 92nd Audio Engineering Society Convention, Vienna (1992 Mar.)

[10] H.A.M. Clark, G.F. Dutton and P.B. Vanderlyn, "The 'Stereo-sonic' Recording and Reproducing System," IEE Proc., vol. 104, pt. B, pp. 417-432, (1957)

[11] B. Rakerd & W.M. Hartmann, "Localization of Sound in Rooms, III: Onset and duration effects," J. Acoust. Soc. Am., vol. 80, pp. 1695-1706, (1986 Dec.)

[12] B. Theiß & M. O. J. Hawksford, "Localization Experiments in Three-Dimensional Sound Reproduction," Preprint 4156 of the 100th Audio Engineering Society Convention, Copenhagen (1996 May)

[13] I. N. Bronstein & K.A. Semendjajew, "Taschenbuch der Mathematik," 23rd edition, Verlag Harri Deutsch, Thun und Frankfurt/Main, (1987)

[14] R. E. Kirk, "Learning a Mayor Factor Influencing Preferences for High-Fidelity Reproducing Systems," J. Acoust. Soc. Am., vol. 28, pp. 1113-1116, (1956, Nov.)

[15] D. Clark, "High-Resolution Subjective Testing Using a Double-Blind Comparator," J. Audio Eng. Soc., vol. 30 no. 5, pp. 330-338 (1982 May)

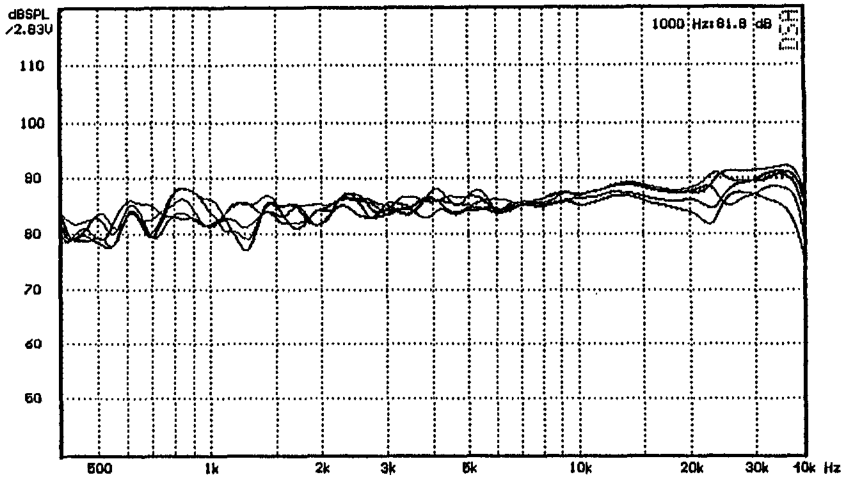

Figure 1 Third-octave smoothed frequency responses of the loudspeakers used for experiments at -20 cm, -10 cm, 0 cm, 10 cm and 20 cm referenced to the height of an average seated subject

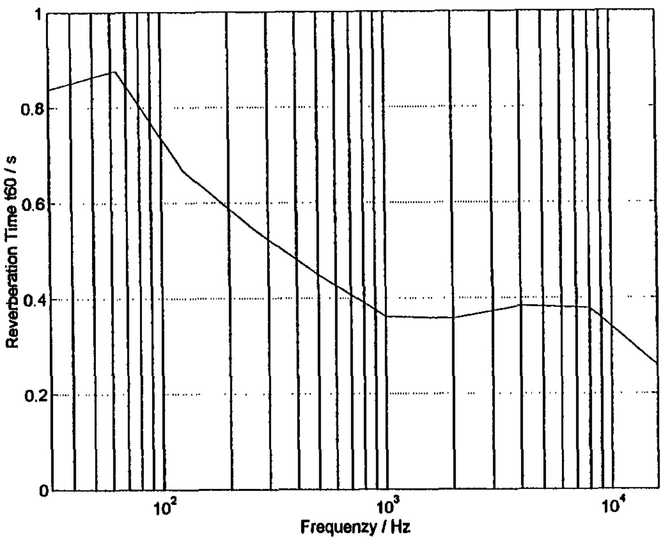

Figure 2 Reverberation time t_{60} of the room used for the experiments

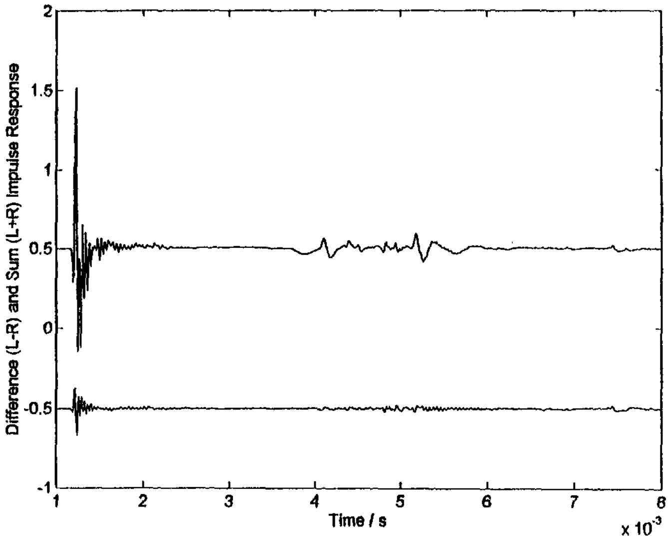

Figure 3 Difference- (lower trace) and sum- (upper trace) impulse responses of the loudspeakers when stimulated together

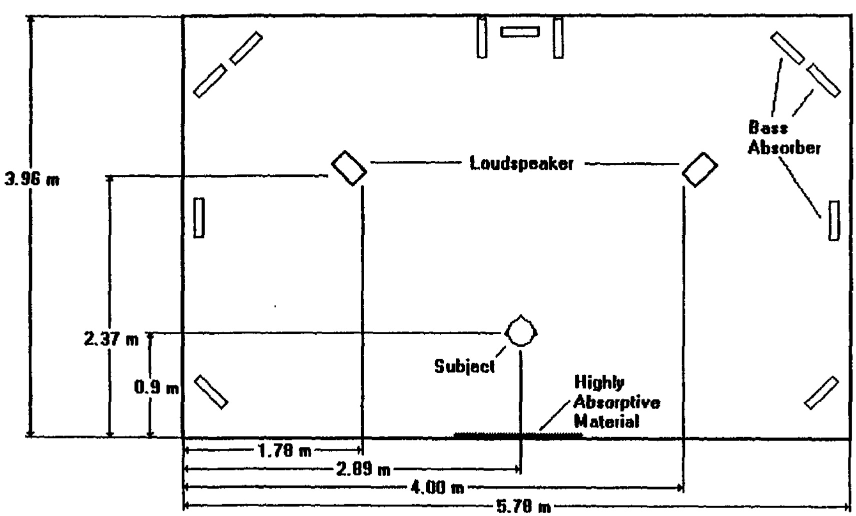

Figure 4 Complete set-up of the room used for the investigation

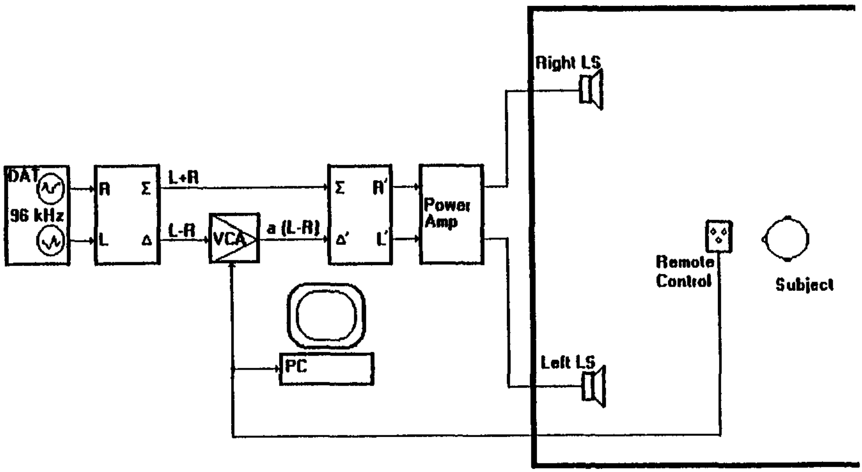

Figure 5 Block diagram of the experimental set-up used to conduct experiment 1 (partially coherent sound sources)

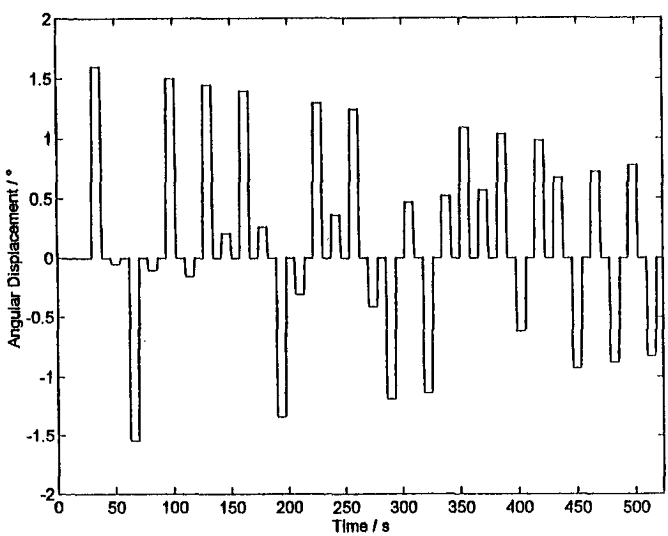

Figure 6 A typical sequence of the angular deviation of the phantom image from centre front as function of time

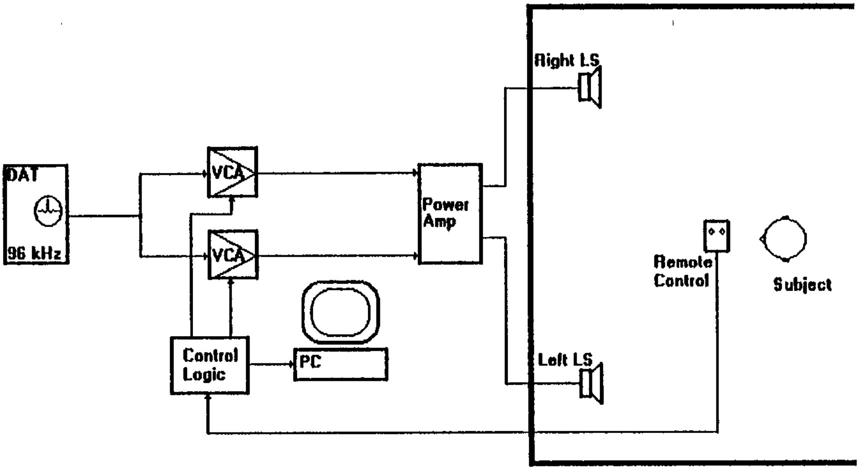

Figure 7 Block diagram of the experimental set-up used to conduct experiment 2 (minimum audible angle)

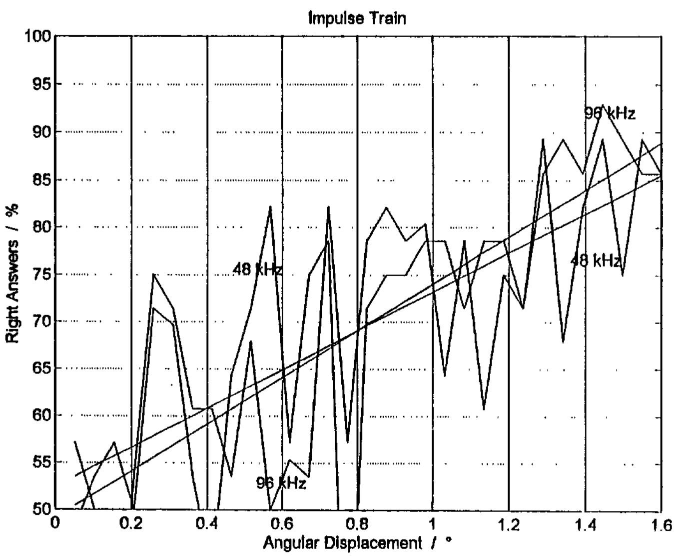

Figure 8 Percentage of right answers versus angular displacement for impulse trains at 48 kHz and 96 kHz sampling rate

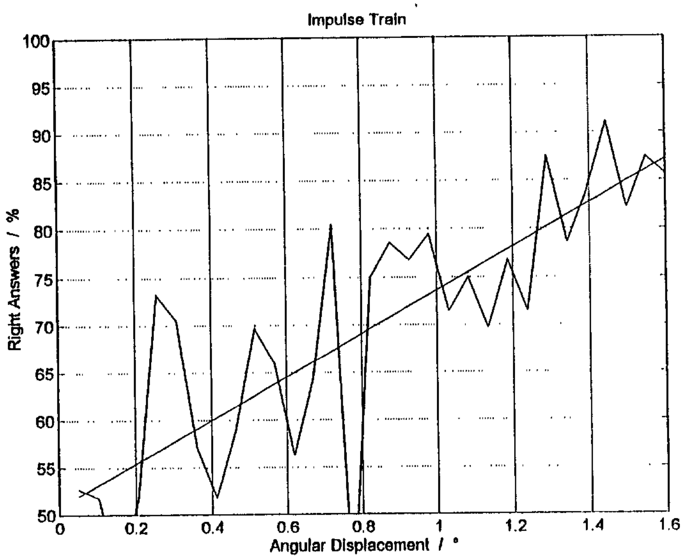

Figure 9 Percentage of right answers versus angular displacement for impulse trains (results from 96 kHz and 48 kHz added)

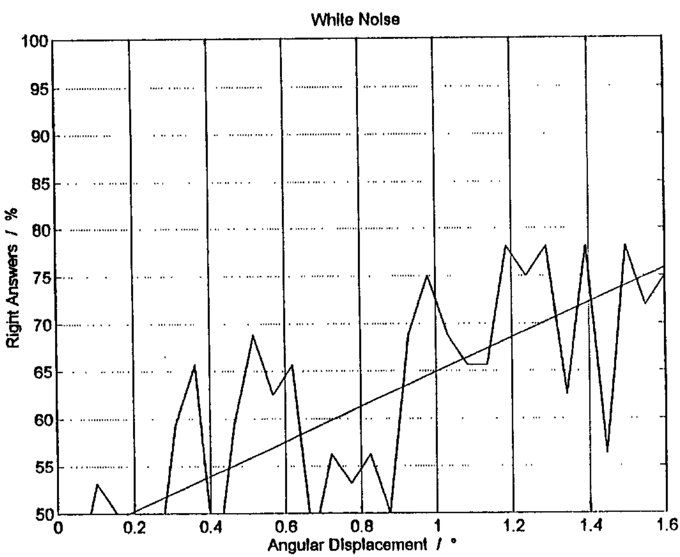

Figure 10 Percentage of right answers versus angular displacement for white noise (results from 96 kHz and 48 kHz added)

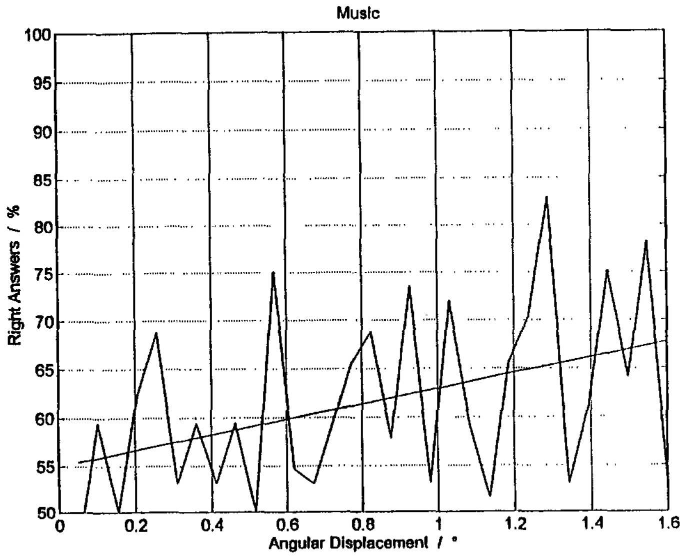

Figure 11 Percentage of right answers versus angular displacement for music (results from 96 kHz and 48 kHz added)

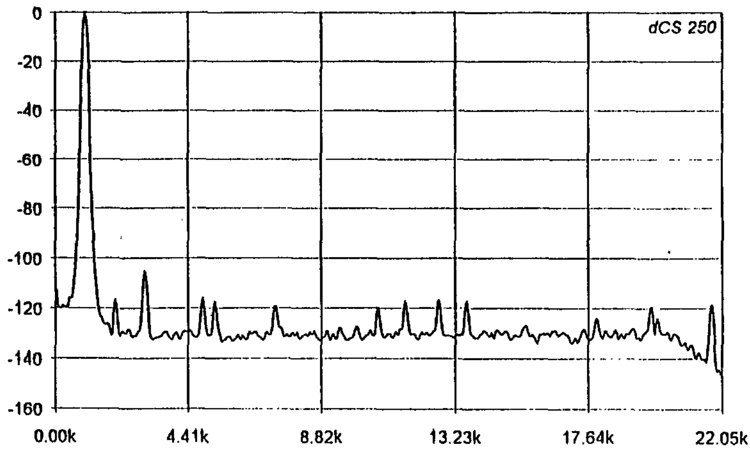

Figure 12 Distortion spectrum of a 1 kHz sinewave conversion from 96 kHz to 44.1 kHz using the dCS 972 sample rate converter, 96 kHz to 48 kHz conversion is likely to be better