

Analytical Report: Verification of Perceptual Differences Between CD (16-bit/44.1 kHz) and HD (24-bit/96 kHz) Audio Formats

I. Statistical Analysis of the Results from a Methodological Perspective

The conducted experiment was a blind test utilizing the Two-Alternative Forced Choice (2AFC) model. From a formal standpoint, assuming the null hypothesis (H_0), which states: *Listeners do not hear the difference, and their answers are pure guessing*, the probability (p) of a correct indication in each trial is exactly 0.5.

Based on the analysis of the binomial distribution, the collected data was evaluated:

1. Consistency with the binomial distribution (the guessing effect) With 4 trials, the probability distribution for pure guessing predicts the following results, which we compare with the empirical data from the experiment ($N=223$):

- **0 correct indications:** expected: 6.25% | Actual: 5.83% (13 individuals)
- **1 correct indication:** expected: 25.00% | Actual: 20.63% (46 individuals)
- **2 correct indications:** expected: 37.50% | Actual: 46.64% (104 individuals)
- **3 correct indications:** expected: 25.00% | Actual: 17.04% (38 individuals)
- **4 correct indications:** expected: 6.25% | Actual: 9.87% (22 individuals)

Conclusions: the empirical distribution shows a strong resemblance to the theoretical distribution for random guessing. We observe certain deviations (an overrepresentation of 2/4 and 4/4 results and an underrepresentation of 1/4 and 3/4); however, from a holistic perspective, they do not indicate the presence of a real discriminatory ability within the surveyed population.

2. Aggregation of trials (Binomial test) Treating each listening session as an independent trial, we obtain:

- **Total number of trials:** 892 ($223 \text{ subjects} \times 4 \text{ pairs}$)
- **Total number of correct indications:** 456
- **Overall percentage accuracy:** 51.12%

The conducted binomial test for $p=0.5$ yielded a p -value of 0.5247. From a scientific standpoint, this means that the deviation from the level of perfect guessing (50%) is statistically completely insignificant. We have absolutely no grounds to reject the null hypothesis.

II. Evaluation of the Research Sample Size ($N=223$)

Are 223 participants a sufficient number to draw categorical conclusions? The answer is: Yes, this is a highly satisfactory sample size for this type of psychoacoustic experiment.

From a methodological perspective:

- **Margin of Error:** Given $N=223$ and a 95% confidence level, the margin of error is $\pm 6.56\%$. This means that if we were to repeat this study on the entire Earth's population, the result would fall within this range with a 95% probability. With the

overall result (51.12% correct answers across all trials), this error poses not the slightest risk of approaching a threshold that we would consider "conscious recognition" (e.g., scores above 70-80%).

- **Statistical Power:** The study generated a total of 892 independent measurements. In the field of psychoacoustics, where the effect size of the difference between CD and HD is assumed to be negligible (or zero), a sample of 892 measurements provides robust statistical power, eliminating the risk of a Type II error (i.e., a false non-rejection of the H0 hypothesis).
- **No Need for Homogeneity:** The design of Winer's test (a home-based test) assumed a diversity of equipment and conditions. The sample of 223 individuals is large enough to neutralize the impact of outliers, such as damaged equipment or extremely specific acoustic conditions.

III. Research Conclusion

In light of the statistical evaluation, the results of the presented experiment demonstrate a categorical lack of ability within the surveyed population to perceptually discriminate between digital recordings with parameters of 16-bit/44.1 kHz and 24-bit/96 kHz.

The distribution of responses is characterized by parameters typical for random sampling (random noise). The research sample of N=223 with 892 generated decisions is fully adequate to prove this thesis with a high level of scientific reliability. Claims of a measurable acoustic advantage of high-resolution formats in blind tests (while maintaining the same mastering quality) find no confirmation in the analyzed empirical material.