

### 6 Auswertung der Versuchsergebnisse

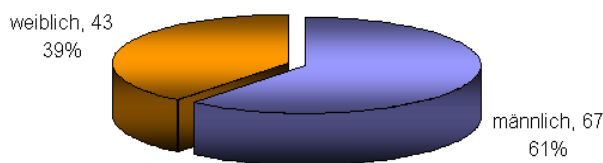
Ziel des vorliegenden Hörvergleichs ist, herauszufinden, ob es Probanden nachweisbar möglich ist, zwischen den zwei digitalen Kodierungsverfahren DSD und PCM (176,4kHz/24Bit) zu differenzieren. Dies sollte anhand von acht Stereo- und zwölf Surround-Musikbeispielen, die eigens für diesen Hörvergleich aufgenommen wurden, mit Hilfe eines ABX-Doppelblindtests untersucht werden. Die mathematische Auswertung basiert auf dem stochastischen Modell der Binomialverteilung.

Neben einer rein deskriptiven Auswertung soll in Kapitel 6.2 ergänzend eine interpretative Auswertung auf der Grundlage wahrnehmungspsychologischer Überlegungen erfolgen – es sei jedoch an dieser Stelle bereits deutlich herausgestellt, dass der Schwerpunkt der Auswertung eindeutig auf dem Kapitel 6.1 liegt.

#### 6.1 Deskriptive Auswertung

Die Hörtests fanden innerhalb von 28 Testtagen statt. In diesem Zeitraum konnten 145 Tests mit 110 Testpersonen<sup>79</sup> durchgeführt werden. Dadurch wird auf jeden Fall den ITU-Richtlinien entsprochen, die besagen, dass bereits aufgrund der Ergebnisse von 20 Personen Schlussfolgerungen gezogen werden können.<sup>80</sup>

Das Testpublikum bestand aus 43 weiblichen und 67 männlichen Testpersonen, deren



Altersstruktur aus Grafik Nr.43 ersichtlich wird. Im arithmetischen Mittel ergibt sich daraus ein Durchschnittsalter der Probanden von 32,9 Jahren.

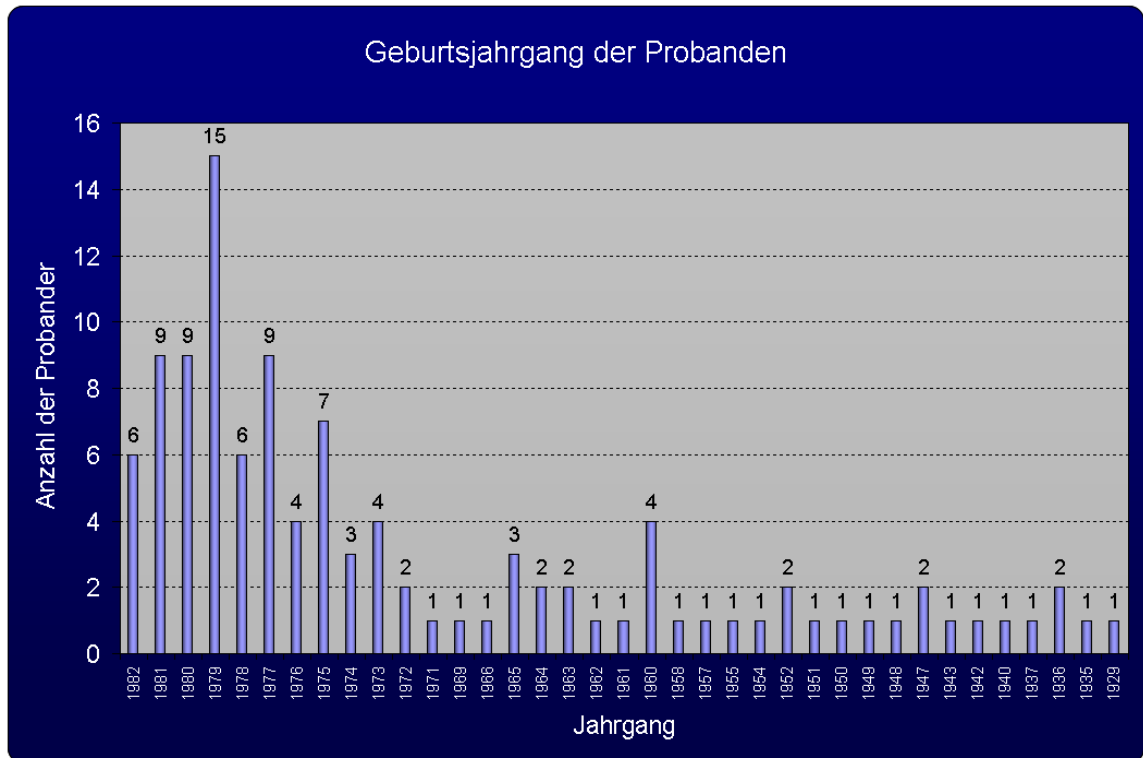
**Grafik Nr.42:**

Prozentuale Zusammensetzung des Testpublikums nach deren Geschlecht.

<sup>79</sup> Manche Probanden führten den Test mit zwei Musikbeispielen entweder direkt aneinander anschließend oder auch an unterschiedlichen Tagen durch.

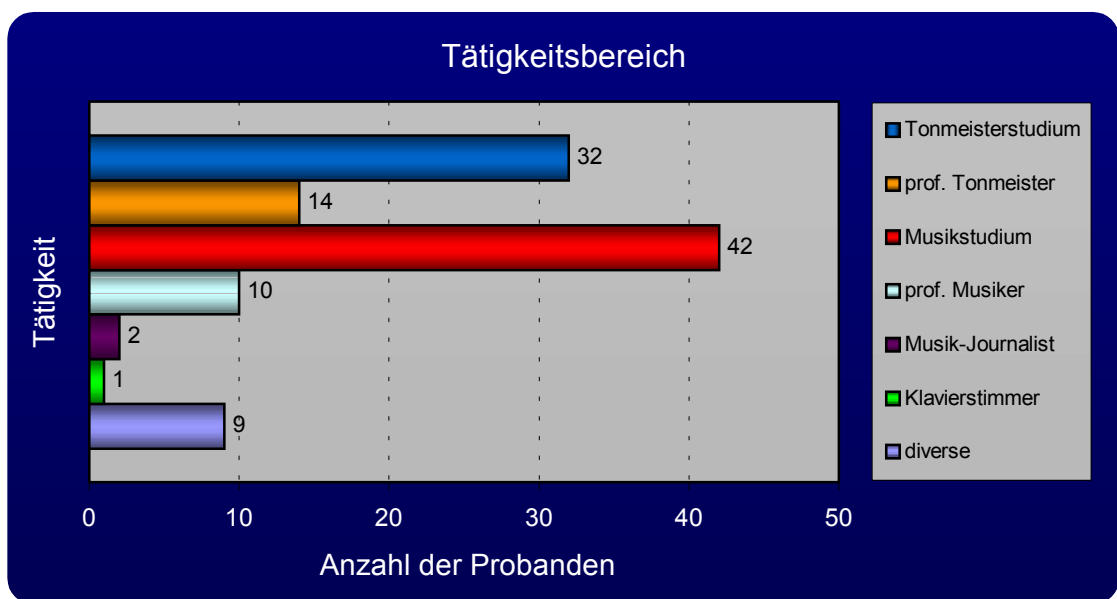
<sup>80</sup> ITU-R BS.1116-1 1997: 5.

## 6 Auswertung der Versuchsergebnisse



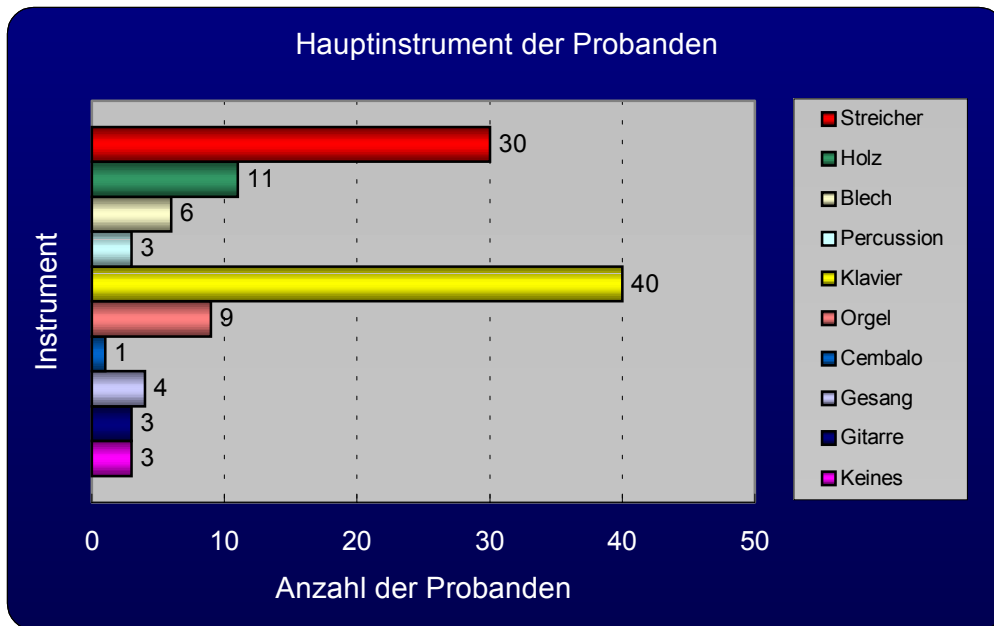
**Grafik Nr.43:** Altersstruktur der Testpersonen.

Grafik Nr.44 verdeutlicht, aus welchem beruflichen Betätigungsfeld sich die Testhörerschaft zusammensetzte. Ergänzend dazu bietet Grafik Nr.45 eine Übersicht über das jeweilige Hauptinstrument der Testpersonen. Aus beiden Grafiken lässt sich erkennen, dass das Testpublikum in der Mehrzahl auf jeden Fall (professionell-) musikalisches und kritisch-analytisches Hören gewohnt war.



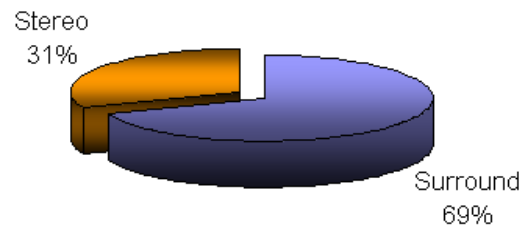
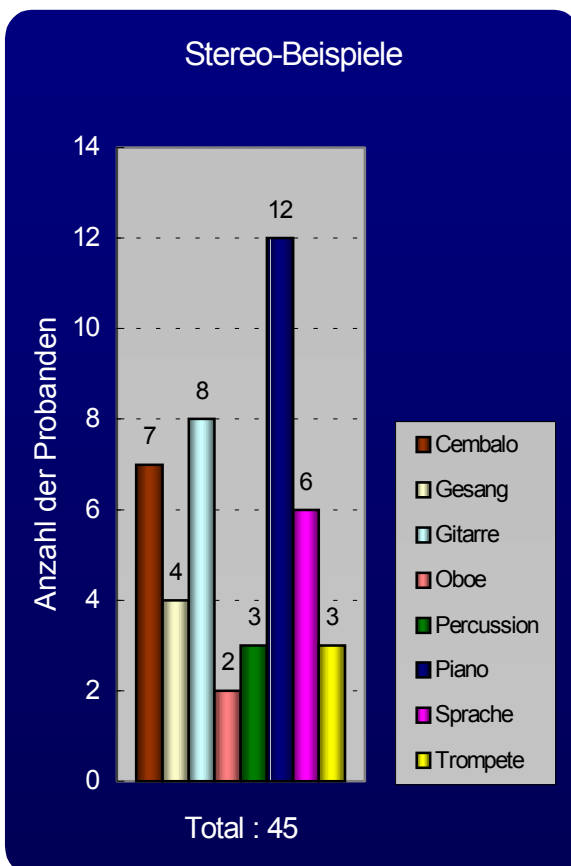
**Grafik Nr.44:** Beruflicher Tätigkeitsbereich der Probanden.

## 6 Auswertung der Versuchsergebnisse



**Grafik Nr.45:** Hauptinstrument der Probanden.

Die 145 durchgeführten Tests teilen sich in einem Verhältnis von 1:2,2 auf 45 Stereo-Beispiele (davon 30 mit Kopfhörerwiedergabe) und 100 Surround-Beispiele auf.

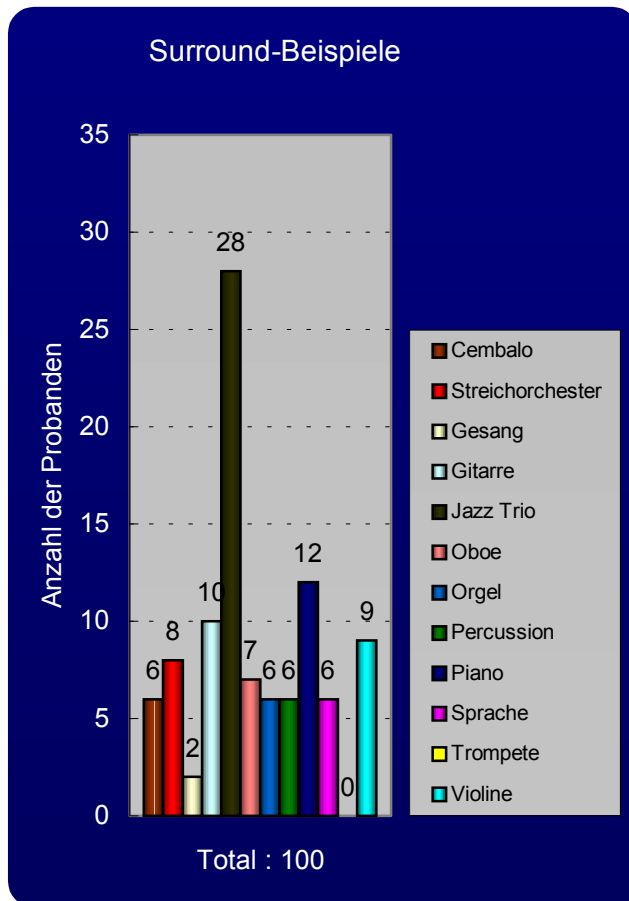


**Grafik Nr.46:** Verhältnis von Stereo- und Surround-Beispielen.

Aus den Grafiken Nr.47a und 47b lässt sich entnehmen, mit welchem Verteilungsverhältnis sich die Probanden für die 20 zur Auswahl stehenden Musikbeispiele in der eigentlichen Testphase entschieden haben.

**Grafik Nr.47a:** Verteilung der insgesamt 45 durchgeführten Stereo-Tests auf die zur Verfügung stehenden Musikbeispiele.

## 6 Auswertung der Versuchsergebnisse

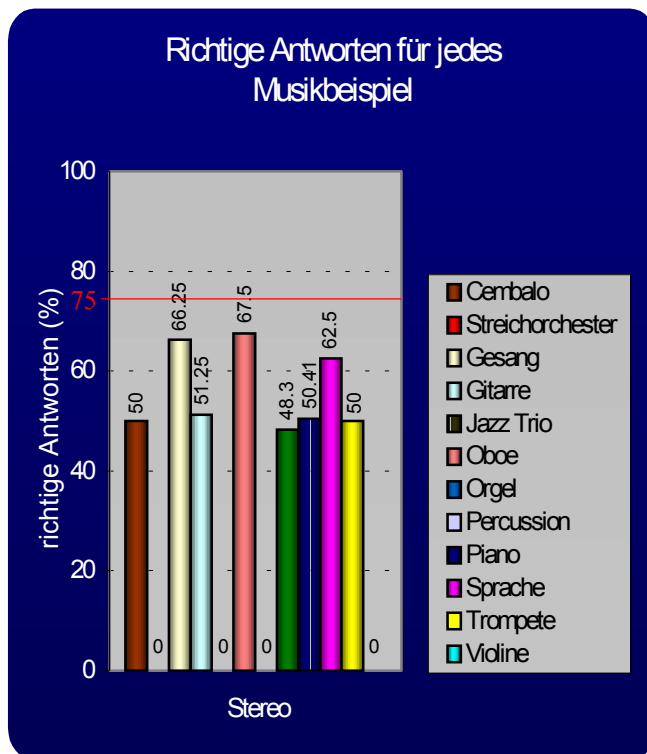


**Grafik Nr.47b:**

Verteilung der insgesamt 100 durchgeführten Surround-Tests auf die zur Verfügung stehenden Musikbeispiele.

Auffällig ist, dass das Jazz Trio überdurchschnittlich oft als Testbeispiel gewählt wurde. Begründet wurde das von den Probanden damit, dass durch die deutliche Panorama-Verteilung der Instrumente (Klavier ↔ L, R; Bass ↔ C; Schlagzeug ↔ LS, RS) die Aufnahme sehr durchsichtig sei und gleichzeitig sowohl unterschiedliche klangliche als auch räumliche Aspekte beinhalte und somit gute Anhaltspunkte zum Hören biete.

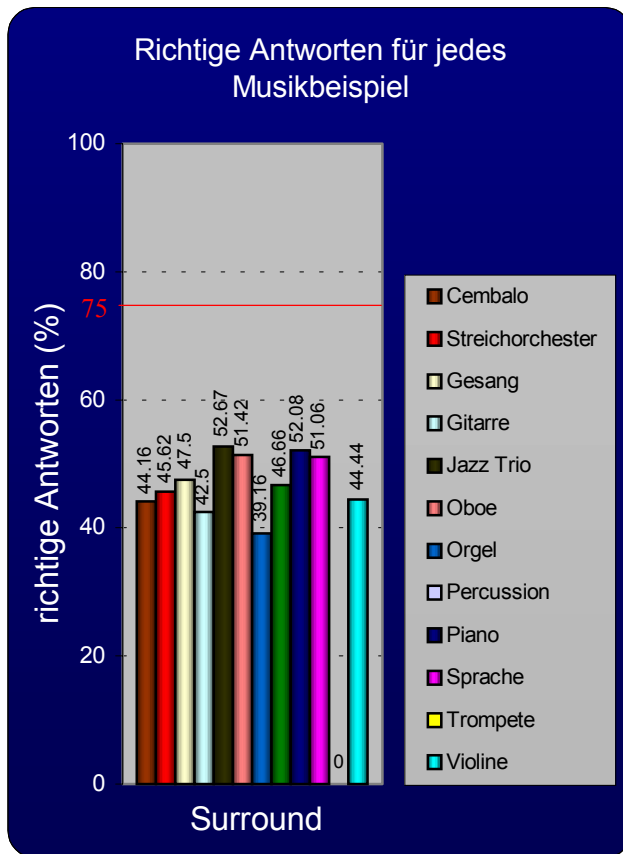
Wie sich die Anzahl der richtigen Antworten pro durchgeführtem Test im arithmetischen Mittel auf die einzelnen, verfügbaren Musikbeispiele in Stereo und Surround verteilt, ist aus den Grafiken Nr.48a und 48b ersichtlich.



Die horizontal rot eingetragene Linie markiert bei 75% (mit 20 Entscheidungen sind nur jeweils 5%-Schritte möglich) die untere Grenze der mindestens 15 richtigen Entscheidungen, bei denen die Eintrittswahrscheinlichkeit  $P(X \geq 15)$  im Bereich der kritischen Wahrscheinlichkeit von 5% liegt.

**Grafik Nr.48a:** Anzahl der richtigen Antworten im arithmetischen Mittel für jedes Stereo-Musikbeispiel.

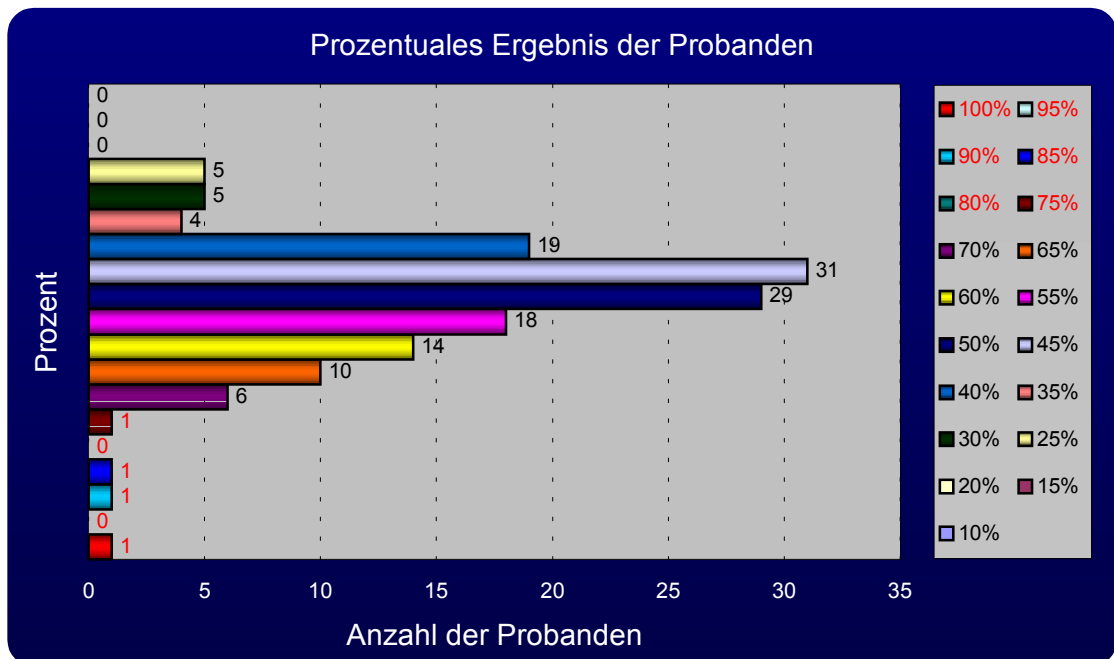
## 6 Auswertung der Versuchsergebnisse



**Grafik Nr.48b:**

Anzahl der richtigen Antworten im arithmetischen Mittel für jedes Surround-Musikbeispiel.

Aus Grafik Nr.49 lässt sich die Verteilung der 145 prozentualen Ergebnisse der Probanden entnehmen, die im Verlaufe der Testdurchführung erzielt wurden.



**Grafik Nr.49:** Anzahl der Probanden in Abhängigkeit von deren erreichtem prozentualen Ergebnis (Anzahl der richtigen Entscheidungen).

## 6 Auswertung der Versuchsergebnisse

---

Wie in Kapitel 5 ausführlich erklärt wurde, muss der Proband bei 20 Entscheidungen mindestens 75% (also 15 von 20) korrekte Antworten gegeben haben, um die kritische Wahrscheinlichkeit zu erreichen. Die Wahrscheinlichkeit hierfür beträgt  $p = 0,021$ .

Aus der Übersichtsgrafik Nr.49 wird deutlich, dass 4 Testpersonen im Bereich der kritischen Wahrscheinlichkeit liegen (rot unterlegt). Dies entspricht lediglich 2,76% der Gesamttests. Gemeinsam ist diesen 4 Ergebnissen, dass sie mit Stereo-Beispielen stattgefunden haben, die über Kopfhörer, also unter weitestgehender Ausblendung von ablenkenden Umwelt- und Raumeinflüssen, wiedergegeben wurden. Jeder dieser Probanden hatte dabei allerdings ein anderes Beispiel gewählt:

- **Oboe:** Stereo Kopfhörer **75%** richtige Antworten  $\rightarrow p = 0,021$
- **Sprache:** Stereo Kopfhörer **85%** richtige Antworten  $\rightarrow p < 0,005$
- **Gitarre:** Stereo Kopfhörer **90%** richtige Antworten  $\rightarrow p < 0,005$
- **Gesang:** Stereo Kopfhörer **100%** richtige Antworten  $\rightarrow p < 0,000...$

Leider wurden alle 4 Tests innerhalb der letzten vier Testtage absolviert, so dass es aufgrund einer hundertprozentigen Frequentierung der Testtermine in dieser Zeit nicht mehr möglich war, mit diesen Personen jeweils einen nachfolgenden Verifikationstest durchzuführen. Dies wäre mit Sicherheit interessant gewesen, jedoch nach Meinung der Testleiter keine zwingend notwendige Maßnahme, da die Aussagekraft der statistischen Auswertung aufgrund der – gegenüber den Empfehlungen – schon erhöhten Anzahl der zu treffenden Entscheidungen bereits größer als gefordert ist und somit reine „Zufallsergebnisse“ weitestgehend ausgeschlossen werden können (Zufallsergebnisse sind natürlich nicht ganz auszuschließen, die Wahrscheinlichkeit für diese ist jedoch äußerst gering). Für diese 4 Fälle soll im kommenden Unterkapitel 6.2.2 ein möglicher Erklärungsansatz skizziert werden, der aber bei der rein deskriptiven Auswertung keine Berücksichtigung findet.

Die Schlussfolgerung muss also lauten, dass bei diesen 4 Fällen im Bereich der kritischen Wahrscheinlichkeit aufgrund der Entscheidungsregel die Hypothese H (es existieren keine wahrnehmbaren Unterschiede zwischen der Quelle A und der Quelle B) verworfen und die Gegenhypothese G (es existieren wahrnehmbare Unterschiede zwischen A und B) angenommen werden kann. In diesen Fällen könnte davon ausgegangen werden, dass die Probanden einen Unterschied zwischen den Quellen A und B wahrgenommen haben.

## 6 Auswertung der Versuchsergebnisse

---

Dahingegen kann bei 141 von 145 Testergebnissen (entspricht 97,24%) die Hypothese H nicht abgelehnt, sondern muss beibehalten werden. In diesen Fällen könnte die Vermutung geäußert werden, dass ein Unterschied zwischen den Quellen A und B vom Probanden nicht wahrgenommen wurde.

Die Grafiken Nr. 48a und 48b zeigen sehr anschaulich, wie sich die prozentuale Verteilung bei den einzelnen Musikbeispielen in der Nähe des Zufalls-Niveaus von 50 Prozent bewegt (bei den Surround-Beispielen noch erheblich deutlicher als bei den Stereo-Beispielen). Bestärkt werden diese Beobachtungen, wenn alle falschen und richtigen Entscheidungen zusammen addiert werden: So ergibt sich bei insgesamt 145 Tests (multipliziert mit jeweils 20 Entscheidungen) eine Gesamt-Entscheidungszahl von 2900. Mit Hilfe der Grafik Nr.49 lässt sich eine Verteilung von 1454 richtigen zu 1446 falschen Entscheidungen berechnen, wobei eine bemerkenswerte Genauigkeit bezüglich des statistischen Erwartungswertes<sup>81</sup> von 1450<sup>82</sup> richtigen bzw. falschen Entscheidungen erreicht wird – bei nur 4 (nicht zu verwechseln mit den 4 signifikanten Testergebnissen) von diesem Erwartungswert abweichenden „Differenzentscheidungen“ entspricht das einer Abweichung von 0,28%.

Wie im Kapitel 4.4 erwähnt, wurde nach den ersten 10 getroffenen Zuweisungen von „X“ ein Zwischenergebnis abgespeichert, um eine eventuelle Entwicklung innerhalb des Testdurchlaufs festzustellen. Hier ließ sich jedoch keinerlei Regelmäßigkeit entdecken.

Da eine Differenzierbarkeit nur bei Stereo-Beispielen erfolgt ist, kann resümierend gesagt werden, dass Unterschiede bei den Surround-Beispielen, also unter Einbeziehung räumlicher Komponenten, in keinem von 100 Tests nachweislich wahrgenommen wurden. Ebenso konnte eine Unterscheidung perkussiver Elemente (Percussion und Cembalo), die sehr schnelle Signalanstiegsflanken enthalten, nicht vorgenommen werden. Es muss zusammenfassend festgestellt werden, dass bei dem vorliegenden Hörvergleich in der Regel nicht nachzuweisen war, dass klangliche Unterschiede zwischen DSD und High-Resolution-PCM (176,4kHz/24Bit) gehört wurden.

---

<sup>81</sup> Arithmetischer Mittelwert einer Zufallsvariablen.

<sup>82</sup> Bei Betrachtung des Tests als Zufallsexperiment.