

6.2 Interpretative Auswertung

Über die rein deskriptive Auswertung hinaus soll in diesem Kapitel neben einigen allgemeinen Bemerkungen zur Diskussion um hochauflösende Aufzeichnungs- oder Kodierungsverfahren vor allem der Versuch einer möglichen Erklärung der vier – außerhalb des deutlich erkennbaren Trends liegenden – Testergebnisse unternommen werden⁸³. Es sei jedoch ausdrücklich betont, dass die Gültigkeit der Ergebnisse dieser vier Testpersonen keinesfalls angezweifelt wird; im Folgenden soll lediglich mit dem Wissen um die beim Testsetup entstandenen Schwierigkeiten ein möglicher Erklärungsansatz geliefert werden.

6.2.1 Allgemeine Vorbemerkungen

Immer wieder steht die Hörbarkeit von Ultraschall-Komponenten oberhalb des Audiobandes im Mittelpunkt von Diskussionen, in denen die Existenzberechtigung hoher Quantisierungs- und Samplingraten in Bezug auf die zwangsläufig entstehenden enormen (unkomprimierten) Datenmengen in Frage gestellt bzw. gerechtfertigt wird.

Einerseits gilt eine obere menschliche Wahrnehmungsgrenze von durchschnittlich 16-17kHz als allgemein akzeptiert, wobei sich diese Werte rein auf periodische Stimuli beziehen und Einzelpersonen bis ca. 20kHz oder auch darüber wahrnehmen können. Andererseits sind Schallereignisse mit unperiodischer Struktur für das menschliche Hörvermögen selbstverständlich im musikalischen Alltag weitaus verbreiteter, durch deren komplexe Strukturen jedoch ungeeigneter zur Feststellung einer oberen Wahrnehmungsgrenze. Des Weiteren können impulshafte Signale durch die steilen Anstiegsflanken in ihrem Einschwingvorgang im Zeitbereich Frequenzen oberhalb des Audiobandes entsprechen.

Daraus folgt die Vermutung, dass zwar bezogen auf periodische Signale eine individuell verschiedene Hörschwelle zwischen ca. 16 und 20kHz besteht, während des Einschwingvorgangs sehr steiler Impulse mit spektralen Anteilen oberhalb dieser Hörschwelle die Stereocilien⁸⁴ allerdings so stimuliert werden, dass diese Frequenzen

⁸³ Zumal bei 100 durchgeführten Surround-Tests kein Ergebnis im Bereich des Signifikanzniveaus lag.

⁸⁴ Beim Eintreffen eines Schallereignisses werden in der Cochlea im Innenohr jene auf der Basilarmembran befindlichen Stereocilien (Haarzellen) auf den Rezeptoren stimuliert, die auf die Weiterleitung über afferente Nervenfasern einer definierten Frequenz ausgerichtet sind.

6 Auswertung der Versuchsergebnisse

detektiert und verarbeitet werden können – bezüglich dieser schnellen Einschwingvorgänge stehen aber noch immer viele Fragen offen.

Um sich der menschlichen Perzeption bezüglich Frequenzen oberhalb des hörbaren Bereichs zu nähern, wurden bereits einige Hörversuche durchgeführt. Bedingt durch die Nichtlinearität von Lautsprechersystemen können sich bei der Wiedergabe über einen Lautsprecher jedoch Intermodulationsprodukte im wahrnehmbaren Bereich bilden, so dass sich ein hörbarer Unterschied rein auf technische Unzulänglichkeiten des Wiedergabesystems zurückführen ließe.

Dies bestätigen Untersuchungen, die im Rahmen der 110th AES-Convention in Amsterdam von Ashihara Kaoru und Kiryu Shogo präsentiert wurden.⁸⁵ Die beiden Autoren stellten fest, dass alle Probanden nur dann in der Lage waren, zwischen einem gegebenen Stimulus⁸⁶ mit und ohne Frequenzkomponenten oberhalb von 22kHz zu differenzieren, wenn dieser über einen einzelnen Lautsprecher wiedergegeben wurde. Sobald der Stimulus – aufgeteilt in 6 Frequenzbänder – zur Reduktion von Intermodulations-Verzerrungen über sechs, den einzelnen Frequenzbändern zugeordnete Lautsprecher wiedergegeben wurde, war es den Probanden nicht mehr möglich, Ultraschallkomponenten zu detektieren. Daraus zogen sie die Schlussfolgerung, dass das Hinzufügen von Ultraschallkomponenten eventuell Klangeindrücke hervorruft, die durch nichtlineare Produkte in den Lautsprechern verursacht werden.

In einer auf der 115th AES-Convention in New York vorgestellten Studie der NHK Science & Technical Research Laboratories, Tokyo⁸⁷ wurde mit Hilfe von Hörversuchen die Unterscheidbarkeit von Musik mit und ohne Frequenzanteilen oberhalb des Audiobandes geprüft. Um Artefakte durch Intermodulations-Verzerrungen zu unterbinden, wurden zwei vollkommen unabhängige Wiedergabepfade verwendet: Zum einen für das hörbare Audioband, zum anderen für die Ultraschallfrequenzen. Jedes der 20 für den Test angebotenen Musikbeispiele wurde mittels eines digitalen FIR-Tief- bzw. Hochpassfilters mit sehr steiler Roll-Off-Charakteristik in zwei Frequenzbänder unterteilt. Während der Hörversuche wurde das hörbare Audioband permanent, das Ultraschallband nur temporär wiedergegeben. Die Ergebnisse zeigen, dass von 36

⁸⁵ AES Convention Paper #5401 2001.

⁸⁶ „A harmonic complex tone consisted of only odd number harmonics was used as the stimulus. Its fundamental frequency (f0) was 2kHz. The lower 5 components that were the fundamental, 3rd, 5th, 7th and 9th harmonics were defined as non-target components and the higher 5 components that were the 11th, 13th, 15th, 17th and 19th harmonics were defined as target-components. Frequency of the highest harmonics was 38kHz. The duration of the stimulus was 2000ms including linear onset and offset ramps of 100ms each.“ [AES Convention Paper #5401 2001: 1].

⁸⁷ AES Convention Paper #5876 2003.

6 Auswertung der Versuchsergebnisse

Testpersonen lediglich eine in der Nähe der kritischen Wahrscheinlichkeit lag. Bei einer Wiederholung mit 6 unterschiedlichen Beispielen war es dieser Probandin jedoch nicht möglich, eine Differenzierung zwischen den einzelnen Signalen (mit und ohne Ultraschallanteilen) im Bereich des Signifikanzniveaus vorzunehmen. Hieraus schlossen die Autoren, dass die Möglichkeit der Differenzierbarkeit beider Signale durch Probanden weder bestätigt noch verneint werden kann.

Darüber hinaus sei aus hörphysiologischer Sicht angemerkt, dass unser hauptsächlich zur Verstärkung der eintreffenden Reize dienendes Mittelohr, bestehend aus den Gehörknöchelchen Hammer, Amboss und Steigbügel, einen sehr steilen mechanischen Tiefpass darstellt. Dadurch kann nur sehr wenig Ultraschallenergie vom Trommelfell überhaupt zum Innenohr vordringen, somit auch nur in einem geringen Maße von der Cochlea aufgenommen und über die afferenten Nervenfasern weitergeleitet werden.

Des Öfteren wird bei Diskussionen das zeitliche Auflösungsvermögen eines digitalen Kodierungsformates, bedingt durch dessen Samplingrate, erwähnt und auf die Wahrnehmungsfähigkeit des menschlichen Gehörs im Bereich von nur wenigen Mikrosekunden verwiesen. Während eine Samplingrate von 44,1kHz alle 22,67 Mikrosekunden eine „Datenerneuerung“ durchführt, sinkt diese Zeitspanne bei einer Abtastrate von 192kHz bereits auf 5,2µsec – bei einer Abtastrate von 2,8224MHz liegt sie im Bereich von ca. 354nsec. Von der Tatsache abgesehen, dass die Phasendarstellung nicht durch das Zeitraster der Abtastfrequenz beeinflusst wird, sondern beliebig genau erfasst werden kann, entspricht die menschliche frequenzabhängige Vornelokalisationsunschärfe von zwei Grad bei Sinussignalen einem zeitlichen Auflösungsvermögen von ca. 20µsec.⁸⁸ Angemerkt sei ferner, dass bei der Überführung einer mechanischen Bewegung der Stereocilien in neuronale Aktivität ein elektrischer Impuls innerhalb von (nur) etwa 10µsec generiert wird.⁸⁹

Die vorliegenden Erkenntnisse der Forschung bezüglich der Wahrnehmbarkeit von Ultraschall sowie des zeitlichen Auflösungsvermögens sind jedoch teilweise unklar und werden äußerst kontrovers diskutiert.

⁸⁸ Zwicker, Fastl 1999: 309f.

⁸⁹ Spitzer 2002: 61.

6.2.2 Wahrnehmungspsychologische Überlegungen

Wie in den Ergebnissen der Auswertung des Hörvergleichs festgestellt wurde, war es insgesamt vier Probanden möglich, bei jeweils unterschiedlichen Musikbeispielen, jedoch übereinstimmend bei Kopfhörerwiedergabe (also Stereobeispielen), innerhalb der kritischen Wahrscheinlichkeit von 5% zu liegen. Im Folgenden soll mit dem Wissen um die technischen Probleme, die das umfangreiche Versuchsssetup mit sich gebracht hat, ein möglicher spekulativer Erklärungsansatz dieser vier – außerhalb des klar erkennbaren Trends liegenden – Fälle beschrieben werden:

Systembedingt produziert das Non-Audio-Format bei Funktionen, die direkt oder auch indirekt eine „Stop“- oder „Play“-Funktion beinhalten, beim DSD-Modus ein im Gegensatz zum PCM-Modus teilweise minimal unterschiedlich klingendes, sehr kurzes Knackgeräusch, das vermutlich nur durch einen Fade-In oder Fade-Out auf digitaler Ebene hätte vermieden werden können. Dies würde einerseits natürlich der Voraussetzung widersprechen, dass das Material in keiner Weise bearbeitet sein darf; der essentiellere Aspekt jedoch wäre andererseits, dass ein solcher nur schwer in einem Non-Audio-Format realisierbar ist. Trotz intensiver Beschäftigung mit diesem Problem und Beratung durch einige der den Hörvergleich unterstützenden Firmen, sowie einer Umstrukturierung des ursprünglich geplanten Rechnerkonzepts, war zum damaligen Zeitpunkt eine Lösung durch eine zusätzliche Soft- oder Hardware nicht greifbar und wäre auch momentan nur schwer umzusetzen.

Dass dieses Knacken nicht bewusst zur Kenntnis genommen wurde und der Test somit in keiner Weise seiner Aussagekraft enthoben wird, zeigt die Auswertung der Fragebögen, in denen ein Knackgeräusch ausnahmslos nicht erwähnt bzw. nicht eindeutig einem der beiden Signale zugewiesen wurde. Dies bestätigt die Annahme, dass durch die Anlage und die Zielsetzung des Hörvergleichs eine bewusste Hörfokussierung auf klangliche Parameter im Mittelpunkt stand, wohingegen es beiden Testleitern wiederholt möglich war, unabhängig vom gewählten Musikbeispiel durch eine bewusst gewählte auditive Aufmerksamkeitslenkung auf das Knackgeräusch eine eindeutige Zuordnung des Signals „X“ zum jeweiligen Quellsignal „A“ oder „B“ vorzunehmen. Aus diesem Grund soll im Folgenden näher auf Aufmerksamkeit und Aufmerksamkeitslenkung eingegangen werden.

Aufmerksamkeit ist ein „Zustand der gesteigerten Wachheit (Vigilanz) und Anspannung, der der selektiven Orientierung des Wahrnehmens, Denkens und Handelns zugrunde liegt. [...] Mit *Aufmerksamkeits-Regulation* bzw. *-Steuerung* [...] werden neurophysiologische Prozesse bezeichnet, welche die Aufnahme, Verarbeitung und Nutzung bestimmter Informationen fördern, der Auswahl (Selektion) der individuell bedeutsamen Situationselemente bei gleichzeitiger Dämpfung irrelevanter Eindrücke Nachdruck verleihen, [...]. Aufmerksamkeit gilt als Selektionsfunktion der Bewußtseinstätigkeit, die das Klarheitsrelief der Eindrücke in Analogie zum Wahrnehmungsfeld bestimmt. Im fokalen, d.h. zentralen Bereich erscheinen Einzelheiten klar und deutlich, so daß sie in ihrer gegenständlichen Bedeutung aufgefaßt werden können (*Apperception*); im peripheren Bereich dagegen gibt es bestenfalls mitbewußte, unklare Eindrücke (*Perceptionen*). Selektive Aufmerksamkeit bzw. Unaufmerksamkeit sind Voraussetzungen des geordneten Erkennens und Denkens, da Klarheit und Deutlichkeit notwendig sind und Aufmerksamkeit und Bewußtseinsumfang dem Erkennen deutliche Grenzen setzen.“⁹⁰

Zurückgehend auf den Ansatz von Broadbent wurde in den 1950er Jahren angenommen, dass Filter im Wahrnehmungssystem für die Verringerung der Informationsvielfalt zuständig sind, indem sie nicht alle Nervenreize weiterleiten. Diese Theorien werden als sogenannte Flaschenhals-Theorien (bottleneck theories) bezeichnet. Ausgangspunkt für diese ist die eingeschränkte Kapazität der zentralen Informationsverarbeitung.

Aufgrund der Ergebnisse von Untersuchungen zum dichotischen Hören⁹¹ formulierte Broadbent 1958 ein Filtermodell der selektiven auditiven Wahrnehmung: Seiner Meinung nach war die Aufmerksamkeit nach dem Prinzip eines selektiven Filters organisiert, das die Menge der eingehenden sensorischen Daten entweder durch Weitergabe wichtiger Informationen an das Bewusstsein, oder durch Sperrung unerwünschter Informationen steuert. Dabei gelangen die eintreffenden Daten vor dem eigentlichen Filter zunächst in eine Pufferzone. Werden Daten vom Filter blockiert, verbleiben sie im Pufferspeicher. Broadbent vermutete, dass das entscheidende Kriterium für das Passieren einer Mitteilung durch den Filter auf der Basis physischer Merkmale geschieht. Darüber hinaus nahm er an, dass nur Daten aus einem einzigen

⁹⁰ dtv-Wörterbuch zur Psychologie 1994: 75ff.

⁹¹ Die Organisation eines Experiments nach dem dichotischen Prinzip bedeutet die simultane Darbietung zweier unterschiedlicher Reize, wobei jedem Ohr jeweils ein Reiz dargeboten wird. Die Versuchspersonen sollen dabei einen Reiz von dem anderen unterscheiden. Die Fähigkeit zu dieser Leistung wird als "Beschatten" ("shadowing") bezeichnet.

6 Auswertung der Versuchsergebnisse

Kommunikationskanal⁹² gleichzeitig durchgelassen werden. Dieser Kanal ist in seiner Kapazität begrenzt.

Eine Modifikation der Filtertheorie ergab sich 1960 aufgrund von Beschattungsexperimenten durch Treisman, da manche Sachverhalte (wie z.B. der Cocktail-Party-Effekt) durch die Broadbent'sche Theorie nicht mit den Beobachtungen in Übereinstimmung zu bringen waren.

Treisman schaltete bei den Versuchspersonen eine auf dem linken Ohr zu beschattende Mitteilung plötzlich zum rechten, eigentlich zu ignorierenden Ohr um, während die auf dem linken Ohr zu hörende Mitteilung in eine zufällige, sinnlose Wortfolge wechselte. Dabei stellte er fest, dass die Versuchspersonen einige wenige Worte dieser Mitteilung zu reproduzieren in der Lage waren, obwohl diese dem nichtbeschatteten Ohr dargeboten wurden.

Basierend auf einem weiteren Beschattungsexperiment konnte Treisman 1964 seine nach dem ersten Experiment aufgestellte Annahme verifizieren, dass Mitteilungen nicht allein nach physischen, sondern auch nach semantischen Kriterien verarbeitet werden und formulierte die sogenannte Dämpfungstheorie der selektiven auditiven Aufmerksamkeit.

Die Kernaussage dieser Theorie ist, dass eingehende Daten auf Basis physischer Kriterien gedämpft, aber nicht völlig ausgefiltert werden; die semantischen Kriterien werden auf alle Informationen angewendet, ob beachtet oder nicht.

Ziel des vorliegenden Hörvergleichs ist, bei den zwei unterschiedlichen Kodierungsverfahren DSD und High-Resolution-PCM (176,4kHz/24Bit) in der klanglichen Wiedergabe Elemente der Differenzierung aufzufinden, so dass eine reproduzierbare Unterscheidung beider gegebenenfalls möglich ist. Dabei spielen klangliche Parameter wie z.B. Räumlichkeit, Stereo-Abbildung (Surround-Abbildung), Transparenz, Klangbalance, Klangfarbe, Impulsverhalten, etc. eine entscheidende Rolle.

Die Aufmerksamkeit richtet sich dabei also nicht auf semantische Aspekte der Musik (im Sinne einer Bedeutungsebene), sondern auf Aspekte des Klanges und seiner zugehörigen Elemente (u.a. auch Knacken, Rauschen, etc.).

⁹² Darunter versteht man z.B. Reizdaten mit bestimmten Eigenschaften, Daten aus einem bestimmten Sinnesgebiet bzw. -organ, oder Daten, die in einem Bedeutungszusammenhang stehen.

6 Auswertung der Versuchsergebnisse

In Analogie zu den sinnvollen Wortzusammenhängen im Fall des Treisman-Beschattungsexperiments lässt sich eine tonmeisterspezifische semantische Ebene – also eine Bedeutungsebene des Klanges – annehmen: Diese wird während der Ausbildung sowie der späteren Berufspraxis entwickelt und verfeinert. Ihre inhaltliche Bedeutung ist für die Aufmerksamkeitslenkung von TonmeisterInnen entscheidend.

Da alle vier im Bereich der kritischen Wahrscheinlichkeit liegenden Probanden Tonmeister(-studierende) waren und die jeweiligen Musikbeispiele bezeichnenderweise über Kopfhörer gehört und sich demzufolge noch einmal ausschließlicher auf das Hören eingelassen haben (störende Raum- bzw. Umwelteinflüsse werden deutlicher ausgeschlossen), könnte diese Bedeutungsebene hier besonders wirksam werden.

An dieser Stelle sei hinzugefügt, dass es mittels Gedächtnis-, Konditionierungs- und Reaktionszeituntersuchungen gelungen ist, einige Ergebnisse bezüglich der Verarbeitung der sekundären (der zu ignorierenden) Mitteilung zu erhalten. Auch wenn diese teilweise nicht eindeutig sind, lässt sich feststellen, dass die zu ignorierenden eingehenden Daten unterhalb der Bewusstseinschwelle auf bedeutungsrelevante Informationen hin überwacht werden und die selektive auditive Aufmerksamkeit somit auch von vorhandenen Wissensstrukturen des Individuums beeinflusst wird.⁹³

Betrachtet man das Testergebnis im Zusammenhang gerade mit der Dämpfungstheorie nach Treisman, die heute als weitestgehend anerkannt gilt, können auch klangliche Elemente wie ein Knacken als – tonmeisterspezifisch – semantische Inhalte gelten und eine Entscheidungsfindung unbewusst oder mitbewusst beeinflussen.

⁹³ Zimbardo 1995: 228.