Zur Hörbarkeit von digitalen Clipping-Verzerrungen
(The audibility of digital clipping artefacts)

Frank Schultz, Vladimir Cholakov*, Hans-Joachim Maempel**

* Technische Universität Berlin, Fachgebiet Audiokommunikation

schueinh@mailbox.tu-berlin.de, v_cholakov@hotmail.com, hans-joachim.maempel@tu-berlin.de

Kurzfassung

In einem Hörversuch wurden Erkennungsschwellen für digitales Clipping ermittelt. Die digitale Übersteuerung der 7 Hörbeispiele aus den Bereichen Sprache, Musik und Geräusch war jeweils in einem Bereich von 0 bis 20 dB in Schritten von 0,25 dB variiert. Die verzerrten Stimuli wurden nach ITU-R BS.1770 an die Lautheit der unverzerrten Signale angeglichen. Der Hörversuch war gemäß dem 3AFC-Paradigma und der adaptiven *The Best PEST*-Methode angelegt. Dieses psychometrische Verfahren erfasst im Unterschied zu den klassischen Verfahren die sensorischen Schwellen ohne den psychologischen Einfluss der Antwortneigung der Versuchsperson und weist zudem eine hohe Reliabilität auf. Die im Versuch ermittelten Schwellen variieren interindividuell sowie mit dem Audiomaterial deutlich. Einzelne Versuchspersonen konnten bei bestimmten Signaltypen bereits eine digitale Übersteuerung von 0,25 dB sicher von der Referenz unterscheiden.

1. Einleitung

1.1. Problemstellung

Man könnte annehmen, dass digitale Clippingverzerrungen in heutigen Digitalsystemen aufgrund ihrer hohen (Re)-Quantisierung ein eher nachrangiges Problem darstellen. Aber selbst ungeachtet der Praxis des Lautheitskrieges [1] gibt es noch häufig technische Vorgänge, bei denen – mitunter unbemerkt – digitale Übersteuerungen auftreten. Der Versuch, unterschiedliche Pegelmessvorschriften für den Studio- und Broadcastbereich zu vereinheitlichen [2], wahrnehmungsrelevante technische Maße zu finden und geeignete Messinstrumente zu etablieren, dauert seit vielen Jahren an. So kann es vorkommen, dass angezeigte Messwerte falsch interpretiert werden und infolgedessen digitale Systeme übersteuert werden. Desweiteren können bestimmte Signalverarbeitungsalgorithmen in Soft- oder Hardware (z.B. SRC, Dynamics, ΔΣ-DA-Wandler) digitale Verzerrungen erzeugen (vgl. Kap. 2 in [3]). Daher ist es von Nutzen, die Empfindlichkeit des menschlichen Gehörs für Clippingverzerrungen zu kennen.

1.2. Analoge und digitale Übersteuerung

Bei analogen Systemen wird die Aussteuerungsgrenze durch einen gerade noch zulässigen Klirrfaktor definiert. Bei ihrer Überschreitung steigt er stetig weiter an. Ein digitales System

hingegen reagiert bei Überschreitung der durch die Wortbreite fest gegebenen Aussteuerungsgrenze mit einem sprunghaften Anstieg des Klirrfaktors. *Abb. 1* zeigt die Abnahme der Klirrdämpfung a_{k3} schematisch für analoge, magnetische Bandaufzeichnung und messtechnisch für einen 20-Bit-A/D-Wandler [4].

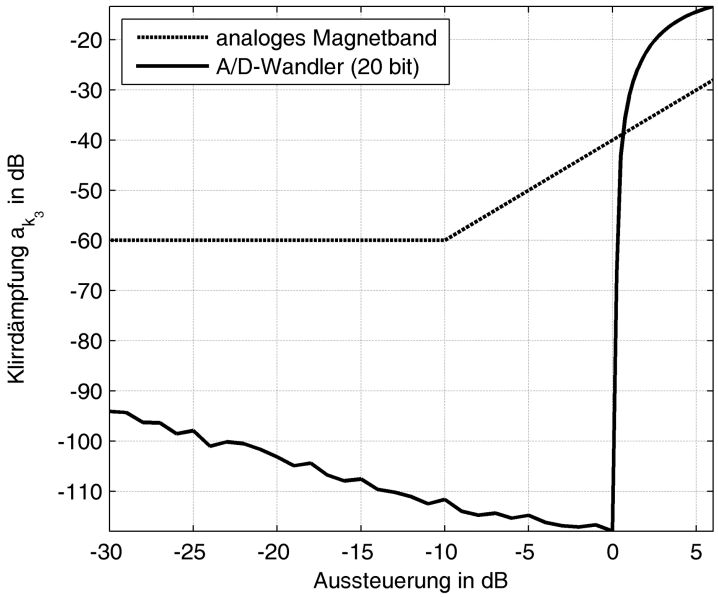

Abb. 1: Übersteuerungsverhalten analoger und digitaler Systeme nach [4] (S. 554)

Abb. 2 zeigt eine nicht übersteuerte und eine durch Übersteuerung geclippte Sinusschwingung im Zeitbereich und Frequenzbereich (Amplitudenspektrum). Bei 1 dB Clipping liegen a_{k3} und a_{k5} um -30 dB_{rel}, entsprechend *Abb. 1*.

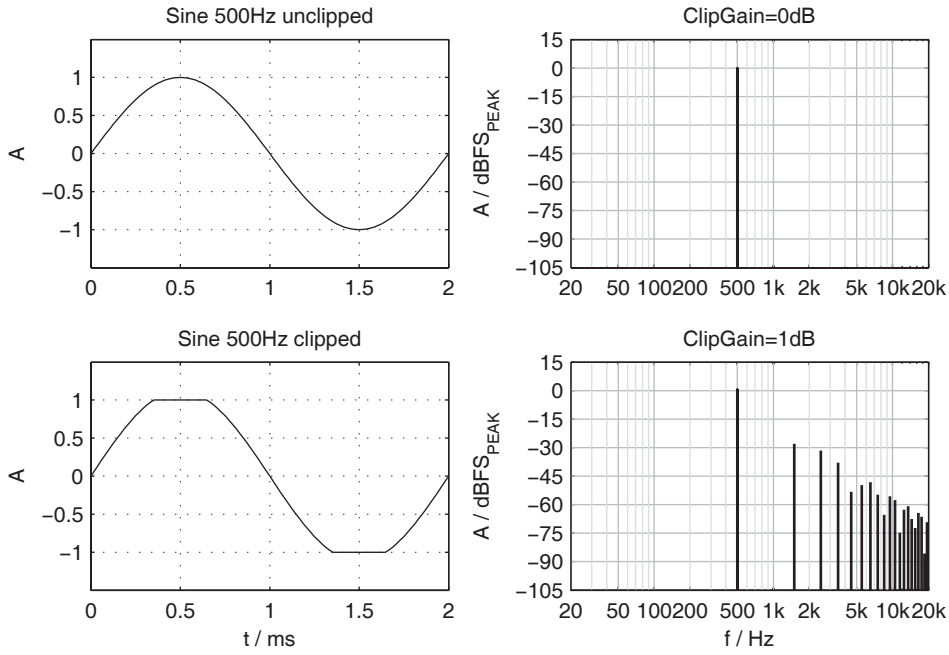

Abb. 2: Digitale Übersteuerung eines Sinussignals im Zeit- (links) und Frequenzbereich (rechts)

1.3. Bisherige Untersuchungen zur Hörbarkeit von Clippingverzerrungen

Untersuchungen zur Hörbarkeit nichtlinearer Verzerrungen in Übertragungssystemen befassen sich überwiegend mit anderen, speziellen Typen wie beispielweise Quantisierungs- oder Lautsprecherverzerrungen. Die Hörbarkeit von digitalen Clippingverzerrungen untersuchte 1984 Jakubowski [5]. Im Hörversuch wurden 35 Programmbeispiele jeweils sukzessive digital verzerrt, bis der Unterschied zur Referenz wahrnehmbar wurde. Es zeigte sich, dass eine Überschreitung der Aussteuerungsgrenze eines digitalen Systems in gewissen Grenzen ohne hörbare Klangänderung vorgenommen werden kann und diese Grenzen audioprogrammabhängig variieren. Die ebenmerklichen Aussteuerungspegel ausgewählter Programmbeispiele sind in *Abb. 3* dargestellt. Von Interesse sind hier die mit einer Integrationszeit von 0 ms gemessenen Spitzenpegel (schwarz). Die digitale Aussteuerungsgrenze konnte im Falle des Programmbeispiels Hi-Hat um 12 dB überschritten werden, bis ein Unterschied hörbar wurde, während z.B. Audioinhalte wie ein Sinussignal oder ein Klavier nicht bzw. kaum übersteuert werden durften. Die hellen Balken zeigen, dass bei der Peakpegelmessung mit 10 ms Integrationszeit die Hörbarkeit von digitalen Verzerrungen für einige Audiosignale bereits deutlich unter 0 dBFS erreicht wird.

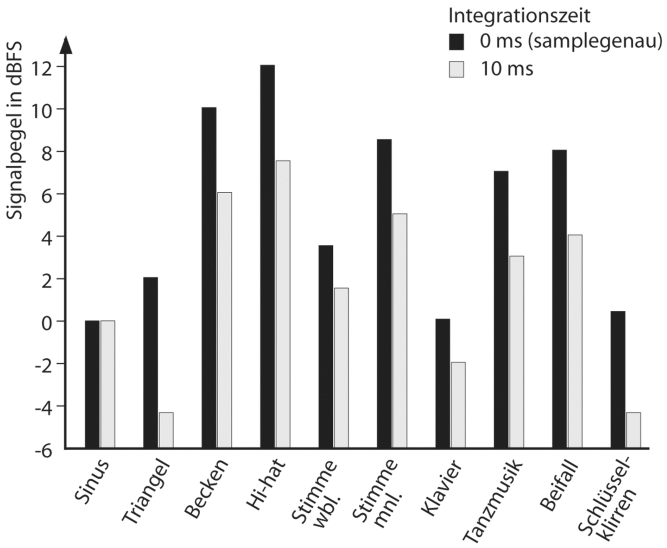

Abb. 3: Hörbarkeitsschwelle für digitale Übersteuerungen nach [5], zit. in [4] (S. 554)

1.4. Erneute Untersuchung der Hörbarkeit von Clippingverzerrungen

Seit 1984 haben sich die technischen Möglichkeiten der Durchführung von Hörversuchen verbessert, die psychometrischen Verfahren weiterentwickelt und die Hörgewohnheiten der Menschen verändert. Da zudem die Untersuchung von Jakubowski [5] in der methodischen Darstellung intransparent bleibt (z.B. unbekannte Anzahl von Versuchspersonen) und das dort eingesetzte klassische psychometrische Verfahren (Herstellungs- oder Grenzverfahren) prinzipielle Mängel aufweist (Habituations- bzw. Antizipationsfehler, Kriterienproblem), ist eine erneute Bestimmung der Erkennungsschwellen von Clippingverzerrungen sinnvoll. Zum einen werden durch eine zeitgemäße DA-Wandler- und Wiedergabetechnik andersartige Verzerrungen als Störvariable stärker reduziert, zum anderen ermöglicht der Einsatz eines Forced-Choice-Verfahrens eine schnelle, kriterienfreie und zuverlässige Bestimmung der fraglichen Unterschiedsschwellen.

2. Hörversuch

2.1. Stimuli

2.1.1. Audioinhalte

Es wurden sieben natürliche Audioinhalte von Referenz-CDs und hochwertigen Kunstmusikproduktionen ausgewählt, die den Bereich typischer Signalcharakteristika (stationär vs. transientenlastig) abdecken und keine Verzerrungen aufgrund digitaler Übersteuerung enthalten: Beifall, Perkussion, Stimme weiblich, Stimme männlich, Klavier, Trompete/Orchester, Cello/Klavier. Normalisierte, ca. dreisekündige, sinnvolle Ausschnitte der Inhalte dienten als Referenz-Stimuli.

2.1.2. Erzeugung der Clipping-Stimuli

Die PCM-Daten der Referenzsignale wurden in Matlab eingelesen und in das 64-Bit-Floating-Point-Format gewandelt, in dem die gesamte folgende Signalbearbeitung vorgenommen wurde. Die Referenzsignale wurden verstärkt und die Signalspitzen jenseits des positiven bzw. negativen Referenzspitzenwerts abgeschnitten. Die Übersteuerungspegel wurden für jedes Referenzsignal in einem Bereich von 0 bis 20 dB mit einer Auflösung von 0,25 dB variiert, so dass sich pro Audioinhalt 80 verzerrte Stimuli und ein unverzerrter Referenzstimulus ergeben.

2.1.3. Lautheitsanpassung

Da im Hörversuch die Unterscheidung zwischen Referenz- und verzerrtem Stimulus nur aufgrund der nichtlinearen Verzerrung, nicht hingegen aufgrund unterschiedlicher Signalleistung erfolgen soll, wurden die Lautheiten aller verzerrten Stimuli gemäß der ITU-Empfehlung BS.1770 [9] an die Lautheiten der jeweiligen Referenzstimuli angepasst. Schließlich wurden die so erzeugten Stimuli im PCM-Format (16 Bit, 44,1 kHz, Stereo) exportiert und standen als einzelne Wave-Audiodateien zum Aufruf durch die Hörversuchs-Ablaufsteuerung bereit.
Für die Herstellung der Stimuli, die Durchführung des Hörversuchs und die Auswertung und Visualisierung der erhobenen Daten wurde die Software Matlab R2007a mit Filter Design Toolbox 4, Signal Processing Toolbox 6 und Statistics Toolbox 7 (alles The MathWorks, Inc.) verwendet.

2.2. Testverfahren

2.2.1. Paradigma

Der Hörversuch erfolgte gemäß dem 3-Alternative-Forced-Choice-Paradigma (3AFC), d.h. der Versuchsperson (Vp) werden in einem Durchgang (Trial) nacheinander drei Stimuli (Alternativen) dargeboten: Zweimal der Referenzstimulus und einmal der manipulierte Stimulus (Testreiz) an unbekannter, zufälliger Position. Die Vp muss nun entscheiden, welcher Stimulus der manipulierte war, notfalls durch Raten. Die Ratewahrscheinlichkeit beträgt demnach $p_R=1/3$. Durch das Erzwingen einer Entscheidung (Forced Choice) geht die Antwortneigung der Vp nicht in die Erhebung der Erkennungsleistung ein (kriterienfreies Verfahren).

2.2.2. Methode

Die Richtigkeit der Entscheidung der Vp bestimmt, welche Merkmalsausprägung (hier: Übersteuerungspegel) im nächsten Trial für den manipulierten Stimulus ausgewählt wird (adaptives Verfahren). Für die Adaptionsregel wurde wegen der großen Variationsbreite der unabhängigen Variable keine Staircase- sondern die parametrische Methode *The BEST Pest* gewählt. Nach [10] benötigt sie zudem etwa die Hälfte weniger Trials als andere Adaptionsregeln, um die Schwelle zu bestimmen und gilt damit als besonders effizient (vgl. auch [11] und [12]). Auf der Grundlage einer sigmoiden psychometrischen Modellfunktion wird unter Berücksichtigung aller absolvierten Trials mittels Maximum-Likelihood-Schätzung die wahrscheinliche Erkennungsschwelle der Vp vorausgesagt. Im nächsten Trial wird ihr der nächstliegende Stimulus als Testreiz dargeboten.

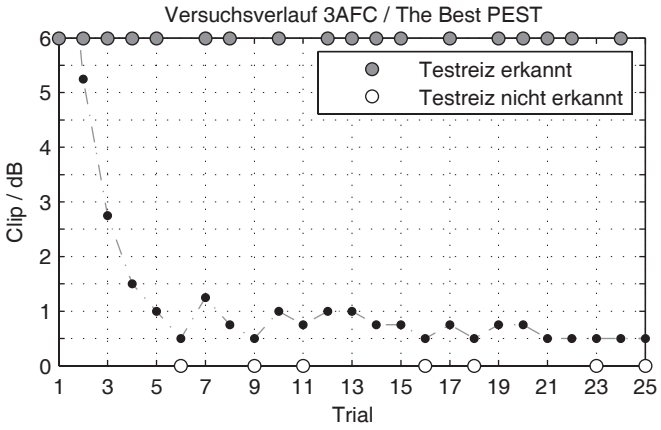

Abb. 4: Konvergenzvorgang der adaptiven *The Best PEST* – Prozedur für eine Vp (Cello/Klavier)

Mit steigender Anzahl von Trials nimmt die Zuverlässigkeit der Schätzung zu und konvergiert gegen die gesuchte Erkennungsschwelle (vgl. exemplarisch *Abb. 4*). Da die Steilheit der psychometrischen Modellfunktion die Geschwindigkeit der Konvergenz entscheidend beeinflusst, wurde sie für jedes Referenzsignal gesondert im Vorversuch bestimmt. Bei einigen Vpn traten am Ende des Konvergenzvorgangs Schwankungen der Schätzungwerte auf, die auf eine individuell nicht optimale Steilheit der Modellfunktion zurückzuführen sind sowie auf Lapsingfehler infolge versuchsdauerbedingter Konzentrationsverminderung. Daher wurde zur Sicherung der Reliabilität des Messverfahrens der Medianwert der letzten zehn Schätzwerte als Messwert der Erkennungsschwelle betrachtet.

2.3. Durchführung

2.3.1. Technischer Versuchsaufbau

Der technische Versuchsaufbau bestand aus einem elektrostatischen Kopfhörer STAX SRS-202 und Kopfhörerverstärker STAX SRM-252II, einem Audiointerface PreSonus EASERA GATEWAY (IEEE 1394) und einem PC. Die grafische Benutzeroberfläche, die Ablaufsteuerung und die Datenerfassung wurden durch Matlab-Skripte realisiert.

2.3.2. Stichprobe

Die Stichprobe wurde aus Personen gezogen, die aufgrund musikalischer Ausbildung oder der Erfahrung in der Teilnahme an Hörversuchen als Expertenhörer einzustufen sind. Insgesamt nahmen 17 Vpn überwiegend aus studentischem Umfeld, an dem Hörversuch teil.

2.3.3. Ablauf

Der Hörversuch begann für jede Vp mit einer Einführung in die grafische Benutzeroberfläche. Danach wurden in zufälliger Reihenfolge die Tests für die sieben Audioinhalte durchgeführt, die jeweils mit einer Trainingsphase begannen, in der sich die Vp auf die Verzerrungsartefakte einstellen und die Lautstärke frei wählen konnte. Die eigentliche Hörtestprozedur umfasste pro Audioinhalt eine feste Anzahl Trials, die im Vorversuch bestimmt wurde. Die gesamte Hörversuchsdauer pro Vp betrug durchschnittlich 60 Minuten.

3. Ergebnisse

Mit Abschluss der Hörversuche liegt die Erkennungsschwelle, also der eben merkliche Unterschied zwischen Referenzsignal und verzerrtem Signal als Übersteuerungspegel, für jede Vp sieben Mal vor – entsprechend sieben Audioinhalten. *Abb. 5* zeigt in Form von Boxplots für jeden Audioinhalt Verteilungskennwerte der Schwellen innerhalb der Stichprobe. Dargestellt sind Median (Strich), Interquartilbereich (Box), Interdezilbereich (Whisker) und Extremwerte (Punkte).

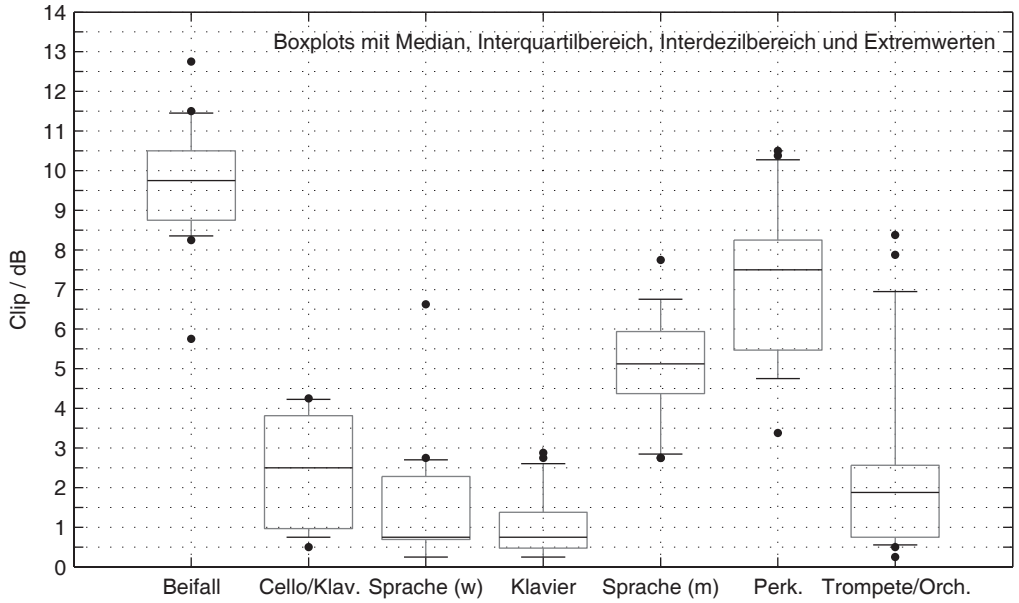

Abb. 5: Boxplots der Erkennungsschwellen für die getesteten Audioinhalte

Die Maße sowohl der Streuung als auch der zentralen Tendenz hängen deutlich vom zugrundeliegenden Audioinhalt ab. Perkussion und Cello/Klavier zeigen im Vergleich zu den anderen Audioinhalten deutlich größere Streubreiten. Geringe mittlere Erkennungsschwellen treten bei Cello/Klavier, weiblicher Stimme, Klavier und Trompete/Orchester auf. Nachdem die empirischen Verteilungen weitgehend normalverteilt sind, ist es zulässig, aus der Stichprobe die Populationsparameter zu schätzen. *Abb. 6* zeigt die geschätzten Populationsmittel-

werte und die 95%-Konfidenzintervalle. In ihnen liegen 95% der Parameter der Populationen, aus denen die entsprechend beobachteten Stichprobenkennwerte stammen können.

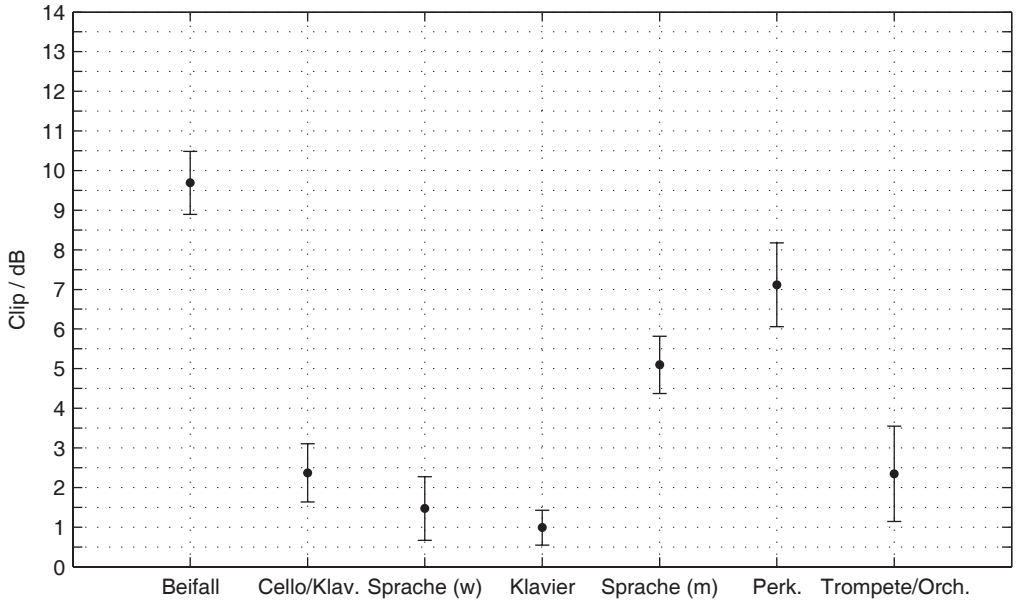

Abb. 6: Mittlere Erkennungsschwellen und 95%-Konfidenzintervalle für die getesteten Audioinhalte

Auch nach dieser Betrachtung, die für die Grundgesamtheit der Expertenhörer generalisierbar ist, fallen Cello/Klavier, weibliche Stimme, Klavier und Trompete/Orchester als die Audioinhalte auf, die mit einer mittleren Schwelle zwischen 1 und 2,5 dB für die Hörbarkeit digitaler Übersteuerungen besonders empfindlich sind. Die Populationsmittelwerte für Beifall, Perkussion und männliche Stimme liegen mit rund 10 dB, 7 dB und 5 dB wesentlich höher.

4. Diskussion

Neben der Variation von Streuung und Mittelwert der Erkennungsschwellen mit dem Audioinhalt fallen die niedrigen Schwellen einzelner Vpn auf. Demnach können empfindliche Individuen bereits eine digitale Übersteuerung von 0,25 dB (weibliche Stimme, Klavier und Trompete/Orchester) resp. 0,5 dB (Cello/Klavier) zuverlässig detektieren. Die höchste individuelle Schwelle liegt bei ca. 12,75 dB und wurde für den Audioinhalt Beifall beobachtet. Dass für transiente Audiosignale Erkennungsschwellen im Bereich um 10 dB liegen, steht in Übereinstimmung mit [5] (vgl. *Abb. 3*, Becken, Hi-hat und Beifall), ebenso die Niederschwelligkeit der Erkennung digitaler Verzerrungen bei stark tonhaltigen Signalen (vgl. *Abb. 3*, Sinus, Triangel, Klavier und Schlüsselklirren). Eine Varianzanalyse für Messwiederholungen und A-posteriori-Vergleiche [13] ergaben, dass sich die Schwellenmittelwerte für Beifall und Perkussion zueinander und auch zu denen der anderen Audioinhalte signifikant unterscheiden. 78% der Varianz der Schwellen sind durch die Variation des Audioinhaltes erklärbar, 9 % sind Urteilervarianz und 13 % Fehlervarianz.

Die im Einzelfall sehr niedrigen Schwellen zeigen, dass für einige Programmbeispiele die Stufung der unabhängigen Variable mit 0,25 dB noch zu groß gewählt war. Daher wurde unter Verwendung einer feineren Übersteuerungsabstufung von 0,0005 dB die Erkennungsschwelle einer empfindlichen Vp nochmals für den Audioinhalt Klavier sowie für ein Sinus-

signal mit einer Frequenz von 500 Hz ermittelt. Sie liegt bei 0,2 dB (Klavier) und 0,015 dB (Sinussignal). Dieses zwar nicht verallgemeinerbare aber gleichwohl reliable Individualergebnis zeigt wie in [14] die bemerkenswerte Empfindlichkeit des menschlichen Gehörs für Klangfarbendifferenzen und die vergleichsweise große perzeptive Wirkung speziell digitaler Übersteuerungen.

5. Literatur

[1] Lund, T. (2004). ”Distortion to the people”. In: *Bericht der 23. Tonmeistertagung.* S. 57-64.

[2] Klar, S. & G. Spikofski (2002). “On levelling and loudness problems at television and radio broadcast studios”. In: *112th Convention of the AES*. Preprint 5538.

[3] Müller, S. (1999). *Digitale Signalverarbeitung für Lautsprecher.* Diss. Aachen: RWTH.

[4] Weinzierl, S. (Hg.) (2008). „Aufnahmeverfahren“. In: *Handbuch der Audiotechnik.* Berlin, Heidelberg: Springer. S. 551-607.

[5] Jakubowski, H. (1984). “Aussteuerungsmessung in der digitalen Tonstudiotechnik“. In: *Rundfunktechnische Mitteilungen* 28 (5). S. 213-219.

[6] Soulodre, A. G. (2004). ”Evaluation of Objective Loudness Meters”. In: *116th Convention of the AES*. Preprint 6161.

[7] Spikofski, G. (2004). “Lautstärkemessung im Rundfunk – Stand der internationalen Standardisierung”. In: *Bericht der 23. Tonmeistertagung*. S. 34-52.

[8] Skovenborg, E. & S. H. Nielsen (2004). “Evaluation of Different Loudness Models with Music and Speech Material“. In: *117th Convention of the Audio Engineering Society*. Preprint 6234.

[9] International Telecommunication Union (Hg.) (2006). *Algorithms to measure audio programme loudness and true-peak audio level*. Rec. ITU-R BS.1770.

[10] Pentland, A. P. (1980). ”Maximum likelihood estimation: The best PEST”. In: *Perception & Psychophysics* 28 (4). S. 377-379.

[11] Treutwein, B. (1995). “Minireview: Adaptive Psychophysical Procedures“. In: *Vision Research* 35 (17). S. 2503-2522.

[12] Lieberman, H. R. & A. P. Pentland (1982). ”Microcomputer-based estimation of psychophysical thresholds: The Best PEST”. In: *Behavior Research Methods & Instrumentation* 14 (1). S. 21-25.

[13] Bortz, J. (2005). *Statistik für Human- und Sozialwissenschaftler.* 6. vollst. bearb. u. akt. Aufl. Heidelberg: Springer Medizin Verl.

[14] Brunner S., H.-J. Maempel & S. Weinzierl (2006). “On the Audibility of Comb Filter Distortions“. In: *Bericht der 24. Tonmeistertagung*. S. 321-329.