

Gerhard Spikofski [spikofski@irt.de](mailto:spikofski@irt.de)

Institut für Rundfunktechnik GmbH München

## **Lautstärkemessung im Rundfunk – Stand der internationalen Standardisierung**

### **Loudness Measurement in Broadcast – Status of the International Standardisation**

#### **1 Einleitung**

Das Problem der Lautstärkeunterschiede bei Radio- und Fernsehprogrammen ist seit langem bekannt. Mit der Einführung neuer digitaler Übertragungstechniken und der damit verbundenen (teilweise unvermeidlichen) parallelen Ausstrahlung (auch) identischer Programme über digitale und analoge Übertragungswege (z. B. DAB/FM-Hörfunk; DVB-S, DVB-T, DVB-C/Analog-TV) ist das Thema Lautstärkeunterschiede bei Radio- und Fernsehprogrammen wieder aktuell geworden.

So kommt es immer häufiger zu Beschwerden der Rundfunkhörer und Fernsehzuschauer über störende Lautstärkeunterschiede bei Radio- und Fernsehprogrammen. Die Lautstärkeunterschiede oder Lautstärkesprünge sind besonders auffällig und entsprechend störend beim „zappen“ durch die verschiedenen empfangbaren Hörfunk- oder Fernsehprogramme, z. B. der digitalen DVB-Satelliten-Kanäle.

Extrem störend können die Lautstärkeunterschiede innerhalb eines Programms z. B. in einem Film beim Übergang von einem Filmdialog zu einem stark komprimierten Werbeblock sein.

Die auftretenden Lautstärkesprünge können mehr als 10 dB betragen, mit anderen Worten, die Lautstärke kann bei den Übergängen mehr als verdoppelt werden.

Die Gründe für die beobachteten Pegel- und Lautstärkeunterschiede bei Radio- und Fernsehprogrammen sind u. a.

- Offensichtliche Missverständnisse in Bezug auf die Aussteuerung von Audio-Kanälen
- Einsatz sehr unterschiedlicher und teilweise nicht standardisierter Aussteuerungsmesser
- Kein standardisierter Studio-Lautstärkemesser verfügbar
- Übertragung von Archivmaterial (analog und digital) ohne individuelle Pegelanpassung an den betreffenden Übertragungskanal

## **2 Lautstärkemessung im deutschen Rundfunk**

Am IRT wurden in den vergangenen Jahren Untersuchungen zu dieser Problematik durchgeführt, um Methoden zur Vermeidung der beschriebenen Lautstärkeunterschiede bei Radio- und Fernsehprogrammen zu entwickeln.

Die Untersuchungen konzentrierten sich auf die Verifizierung der bestehenden Rundfunk-Aussteuerungsempfehlungen und die Entwicklung eines Studio-Lautstärkemessers, dessen Messwerte eine ausreichend hohe Korrelation ( $\geq 90\%$ ) mit der subjektiven Lautstärke aufweisen.

### **2.1 Aussteuerungsempfehlung und Lautstärkealgorithmus**

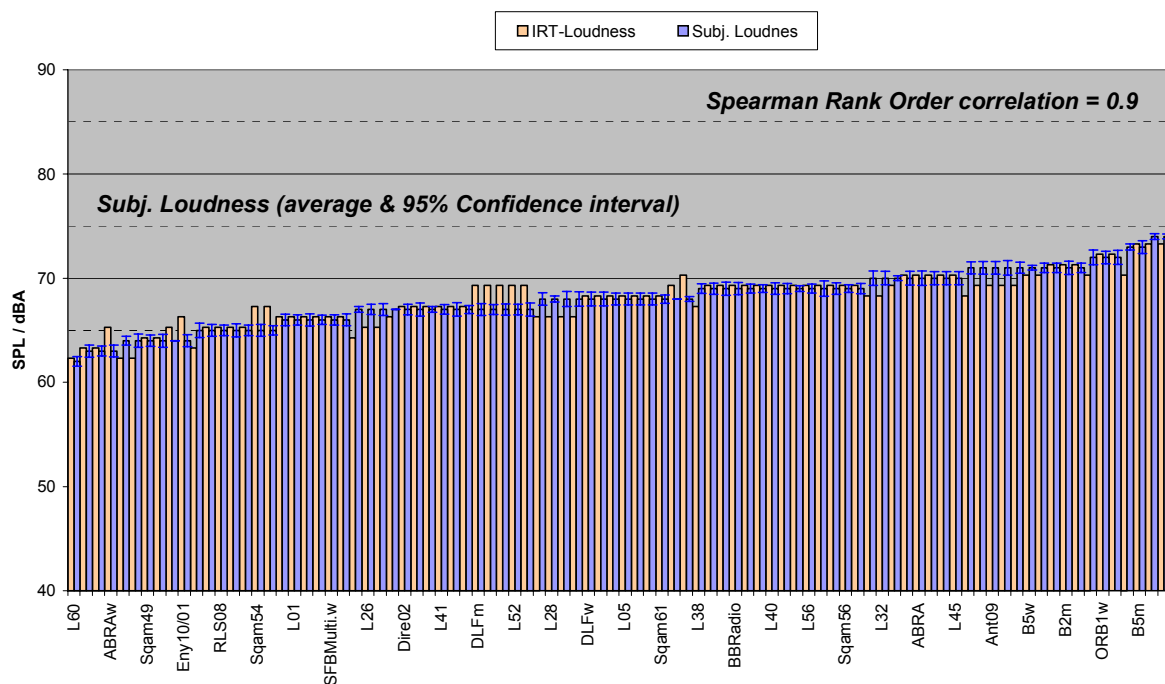
Es konnte eine Lösung erarbeitet werden, die beiden Aspekten gerecht wird. In Bezug auf die Aussteuerung digitaler Audio-Produktionen und – Übertragungen sind die (identischen) Empfehlungen der FS-Produktionsrichtlinien [ 1 ] und der ARD-HFBLK-Empfehlung 15 IRT „Headroom bei digitalen Tonsignalen“ [ 2 ] maßgebend.

Danach sollen die maximalen Signalpegel QPPM (Quasi Peak Programme Meter) bei digitalen Audioproduktionen  $-9$  dB FS (Full Scale) nicht überschreiten. Das bedeutet, unter Verwendung eines Aussteuerungsmessers entsprechend [ 3 ] mit einer Integrationszeit von 10 ms soll ein Headroom von 9 dB eingehalten werden.

Der IRT-Lautstärke-Algorithmus basiert prinzipiell auf der QPPM-Messung, wobei sich der Lautstärkepegel LSM aus der Berechnung des Medianwerts der QPPM-Pegel über ein definiertes Zeitintervall von 3 sec ergibt. [ 4, 5, 6 ].

Es stehen zurzeit zwei Studio-Aussteuerungsmesser (Prototypen) mit IRT-Lautstärke-Implementierung zur Verfügung. Der Lautstärkemesser DAG2000 (Digital Peak Programme Meter 478-220B) sowie der PC-basierte Lautstärkemesser Penguin PG-AM Pro.

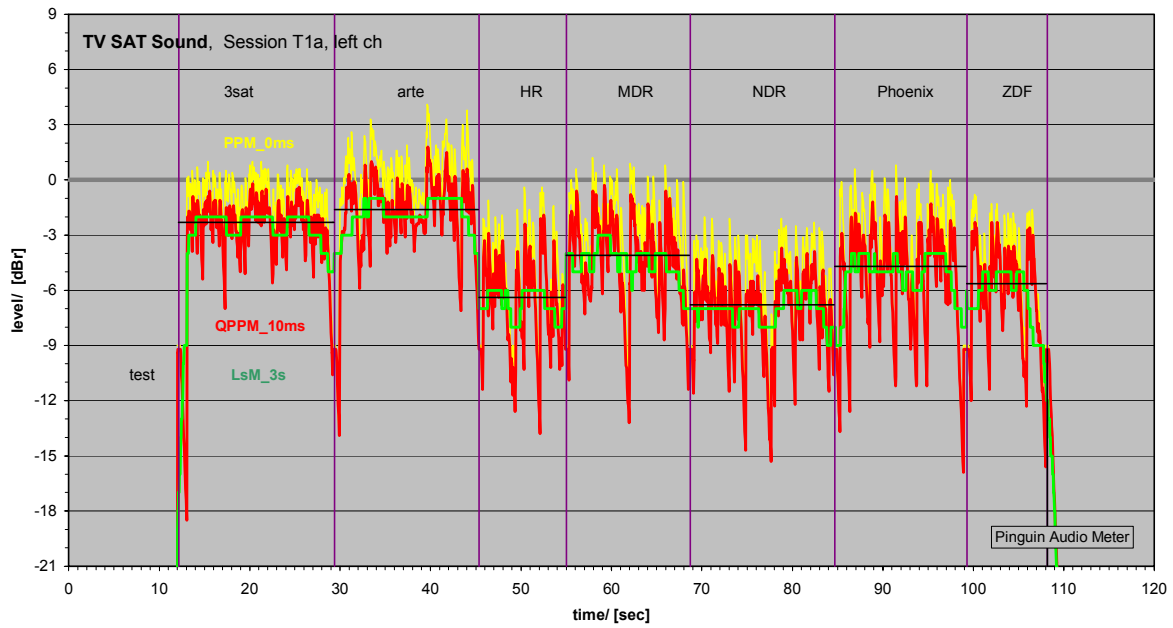
Entsprechende Messungen zur Bestimmung der Korrelation zwischen subjektiver und objektiver Lautstärke auf der Basis von 90 Hörfunkprogrammbeispielen zeigen, dass die subjektiven und - mit Hilfe des IRT-Lautstärkealgorithmus gemessenen - objektiven Lautstärkepegel 90%ig miteinander korrelieren ( Abb. 1 ).



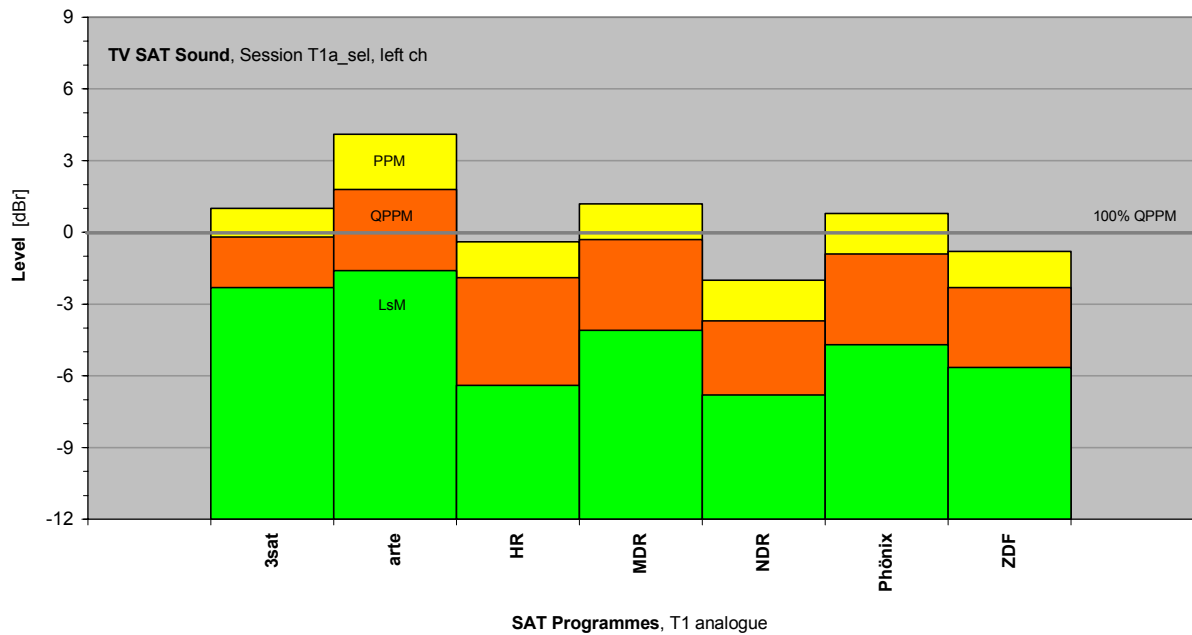
**Abb. 1:** Korrelation zwischen subjektiver und objektiver Lautstärke (IRT-Lautstärkealgorithmus) von Hörfunkprogrammen

## 2.2 Prinzip der Analyse und praktische Ergebnisse

Um die bewährte Methode der Signal- und Lautstärke-Pegelmessung zu verdeutlichen, wurden in Abb. 2 beispielhaft die an 7 Testsequenzen gemessenen Pegel dargestellt. Im Einzelnen sind der Spitzenpegel PPM\_0ms (gelb) (Peak Programme Meter, Integrationszeit 0 ms), der Quasispitzenpegel QPPM\_10ms (orange) und der Lautstärkepegel LSM LSM\_3s (grün) dargestellt.



**Abb. 2a:** Prinzip der Auswertung – Darstellung der Pegel PPM\_0ms, QPPM\_10ms und LSM\_3s über der Zeit für 7 ausgewählte Testsequenzen



**Abb. 2 b:** Prinzip der Auswertung – Darstellung der Maximalpegel PPM, QPPM und des LSM-Medianwertes

Aus den über der Zeit dargestellten Pegelverläufen in Abb. 2a wurden jeweils die relevanten individuellen Pegelwerte jeder Testsequenz berechnet: PPM- und QPPM-Maximalwert, LSM-Medianwert (=repräsentativer Lautstärkepegel) (Abb. 2b).

In entsprechenden Untersuchungen zu Lautstärke- und Qualitätsunterschieden bei DAB-Programmen und Fernsehprogrammen [ 6, 7 ] konnte gezeigt werden, dass die am IRT erarbeiteten Strategien geeignet sind, die Aussteuerungs- und Lautstärkeproblematik messtechnisch zu erfassen und damit zu vermeiden.

### **3 Internationale Standardisierung eines Lautstärkemessers**

#### **3.1 SRG3-Testprozedur**

Seit Mitte 2002 beschäftigt sich die Arbeitsgruppe (Special Rapporteur Group SRG3) innerhalb der ITU-R damit, einen objektiven Lautstärkemesser für Rundfunkanwendungen zu standardisieren. Die durchgeführte weltweite Studie, deren Abschlussbericht Mitte 2003 vorgelegt wurde, [ 8 ] beinhaltete die folgenden Schwerpunkte

- Sammlung und Auswahl rundfunktypischer Audio-Testbeispiele
- Subjektive Lautstärkemessungen in unterschiedlichen Testlabors
- Definition der Anforderungen an die zu testenden Lautstärkealgorithmen einschließlich der entsprechenden Ausschreibung
- Objektive Messungen der eingesandten Lautstärkemesser
- Auswertung der Ergebnisse der subjektiven und objektiven Messungen und Bewertung der getesteten Lautstärkemesser

Das Testmaterial, das für die subjektiven und objektiven Lautstärkemessungen verwendet wurde, besteht aus Programmsequenzen von Fernseh- und Rundfunkanstalten auf der ganzen Welt. Deshalb beinhalten die Sprachsequenzen eine Vielfalt unterschiedlicher Sprachen. Außerdem beinhalten die Sequenzen Musik, Audiobeispiele aus Fernseh- und Kinoproduktionen, Sportveranstaltungen, Nachrichten, Sound Effekte und Werbung.

Die subjektiven Tests wurden in fünf verschiedenen Testlabors auf der ganzen Welt durchgeführt, wobei insgesamt 97 Probanden an den Tests teilnahmen. In Einzelnen waren die folgenden Testlabors beteiligt: Australian Broadcasting Corporation (Australien), British Broadcasting Corporation (England), Communication Research Centre (Canada), National Acoustic Laboratories (Australien) und National Film Board (Canada).

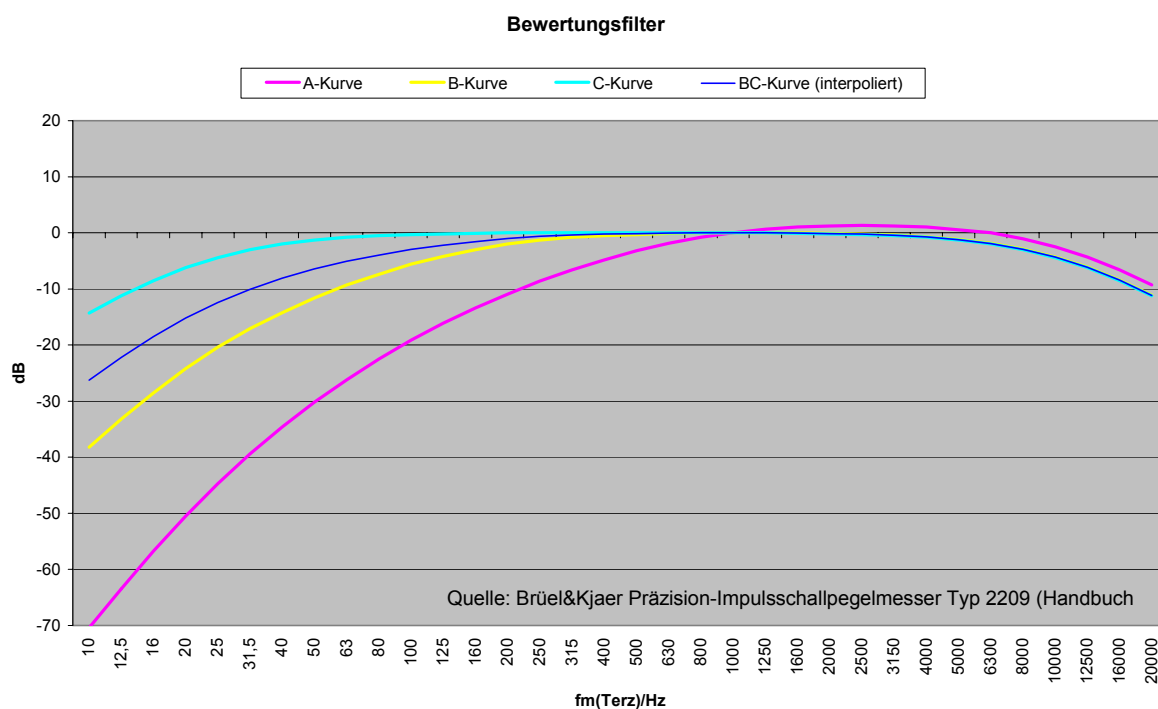
Das von einem Panel aus 3 Mitarbeitern der SRG3 ausgewählte Testmaterial beinhaltet 48 Mono-Audiosequenzen (16 bit/48 khz PCM) einschließlich eines Referenzsprachbeispiels (Weibliche Englische Sprache, SPL 60 dbA). Diese 48 Testbeispiele wurden im Experiment zweimal mit unterschiedlichem Pegel angeboten, so dass insgesamt 96 Testbeispiele zu beurteilen waren. Das Testmaterial beinhaltet folgende Kategorien

- Sprache ohne Hintergrundgeräusche (16)
- Hörspiel (Dialog mit Umgebungsgeräuschen) (4)
- Sprache mit Hintergrundgeräuschen (22)
- Sprache mit Hintergrundgeräuschen (Interview/Sport) (28)
- Instrumentalmusik (14)
- Musik mit Leadsänger (6)
- Gesang ohne Instrumente (4)
- Sound Effekte, Umgebungsgeräusche ohne Sprache (2)
- 

Bei den subjektiven Experimenten handelte es sich um einen individuellen Lautstärkevergleich zwischen dem Referenzsprachbeispiel und den getesteten Audiosequenzen. Dabei wurden die zu vergleichenden Audiosequenzen über einen Lautsprecher vor dem Probanden angeboten. Aufgabe des Probanden war es, die Lautstärke der Testsequenzen mit Hilfe eines Testsoftware integrierten Pegelstellers auf die gleiche subjektive Lautstärke einzupegeln wie die Referenz. Während des Experimentes konnte der Proband mit Hilfe eines Testsoftware integrierten Umschalters beliebig zwischen den Testsequenzen und der Referenz hin und her schalten.

Auf die Ausschreibung der ITU-R wurden insgesamt 10 Lautstärkemesser der folgenden sieben Privatfirmen und Forschungseinrichtungen zur Teilnahme an den Messungen eingesandt DAG2000/IRT, Dolby Laboratories, Dorrough, NHK/Yamaki, Opticom, Pinguin and TC Electronic.

Neben den eingesandten Lautstärkemessern wurden vier Varianten eines einfachen Lautstärkealgorithmus auf der Basis der RMS- bzw. Leq-Messung mit den drei Bewertungskurven (A-, B-, BC-Kurve) und ohne Bewertung einbezogen (Abb. 3).



**Abb. 3:** Audio-Bewertungsfilter

Davon waren zwei Varianten der Leq-Messung, die A- und B-Bewertung, in zwei der offiziell beteiligten Lautstärkemesser implementiert. Im Rahmen der Durchführung der Messungen und Veröffentlichung der Ergebnisse wurden alle Verfahren streng anonym mit [ A ] ... [ M ] bezeichnet. Lediglich die Leq-Algorithmen wurden identifiziert ([ B ] LeqA , [ G ] LeqB (Dolby); [ L ] Leq, [ M ] LeqBC – (CRC/Souludre)).

Die Lautstärkemesser wurden an CRC-Audio Perception Lab geschickt, wo die objektiven Messungen unter der Leitung von Gilbert Soulodre durchgeführt wurden. Vor Beginn der eigentlichen Messungen wurde für jeden Lautstärkemesser mit Hilfe drei ausgewählter Testsignale ein Verifikationstest durchgeführt, um sicher zu stellen, dass die Testmessungen im CRC die selben Ergebnisse liefern wie die eigenen Messungen der Entwickler.

Danach wurden für jeden Lautstärkemesser die gemessenen Lautstärkewerte - ein Bezugswert je Testbeispiel - der 96 Testsequenzen registriert. Um Vergleichbarkeit der Messwerte zu erreichen, wurden die Messwerte jedes Lautstärkemessers normalisiert, indem die Messwerte der einzelnen Testsequenzen jeweils auf den Messwert der Referenz bezogen wurden.

### 3.2 SRG3-Ergebnisse

Um die Leistungsfähigkeit der untersuchten Lautstärkemesser möglichst objektiv bestimmen zu können, wurde von der SRG3 ein Satz von neun mathematisch definierten Qualitätsparametern festgelegt, die alle Aspekte der Lautstärkemesserbewertung weitgehend abdecken. Die detaillierten Berechnungen der festgelegten folgenden Qualitätsparameter werden in [ 8 ] angegeben.

R	-	Correlation
Sprmrho	-	Spearman's Rank Order Correlation
RMSE	-	Root Mean Square Error against the Average subjective scores
MAE	-	Maximum Absolute Error against the Average subjective scores
AAE	-	Average Absolute Error against the Average subjective scores
LPI	-	Loudness Performance Index
ALPI	-	Advanced Loudness Performance Index
ProdSubiDev	-	Product Subjective Deviation
MeanSubiDev	-	Mean Subjective Deviation

	<i>R</i>	<i>Sparman rho</i>	<i>RMSE</i> (dB)	<i>MAE</i> (dB)	<i>AAE</i> (dB)	<i>LPI</i>	<i>ALPI</i>	<i>Prod SubjDev</i>	<i>Mean SubjDev</i>
A	0.944[10]	0.9165[10]	2.37[11]	6.37[10]	1.88[11]	0.1139[11]	0.1177[11]	0.0000[10]	0.9608[11]
B [Leq(A)]	0.929[11]	0.8888[11]	2.19[10]	6.39[11]	1.77[10]	0.1754[10]	0.1830[10]	0.0000[10]	0.8945[10]
C	0.955 [9]	0.9522 [8]	1.75 [9]	5.76 [9]	1.35 [9]	0.3574 [9]	0.3737 [9]	0.6239 [8]	0.6722 [9]
D	0.976 [3]	0.9580 [5]	1.31 [3]	4.70 [5]	0.99 [3]	0.5960 [3]	0.6325 [3]	0.6793 [4]	0.5152 [4]
F	0.965 [8]	0.9508 [9]	1.55 [8]	3.61 [1]	1.28 [8]	0.5059 [7]	0.5296 [7]	0.6200 [9]	0.6686 [8]
G [Leq(B)]	0.972 [5]	0.9524 [7]	1.37 [6]	4.19 [4]	1.07 [5]	0.5770 [5]	0.6131 [5]	0.6695 [5]	0.5390 [5]
H	0.848[12]	0.8407[12]	3.33[12]	6.89[12]	2.90[12]	0.0092[12]	0.0093[12]	0.0000[10]	1.4698[12]
I	0.972 [5]	0.9602 [3]	1.36 [5]	4.80 [6]	1.09 [6]	0.5889 [4]	0.6218 [4]	0.6637 [6]	0.5513 [6]
J	0.968 [7]	0.9547 [6]	1.51 [7]	4.97 [7]	1.17 [7]	0.4910 [8]	0.5193 [8]	0.6500 [7]	0.5926 [7]
K	0.975 [4]	0.9582 [4]	1.33 [4]	5.13 [8]	0.99 [3]	0.5754 [6]	0.6105 [6]	0.6800 [3]	0.5140 [3]

**Tab.:** Performance der untersuchten Lautstärkemesser. Werte in Klammern [ ] kennzeichnen die Rangfolge für jeden Parameter [ 8 ]



Diese neun Qualitätsparameter erlauben eine differenzierte mehrdimensionale Bewertung der Leistungsfähigkeit der untersuchten Lautstärkemesser. Insbesondere ergibt sich in Bezug auf jeden einzelnen Parameter eine Rangfolge, die es gestattet, die Leistungsfähigkeit der Lautstärkemesser individuell einzustufen. Die resultierenden Ergebnisse sind für die untersuchten Lautstärkemesser einschließlich der Parameter bezogenen relativen Rangfolge (in Klammern [ ]) (Tab.) dargestellt.

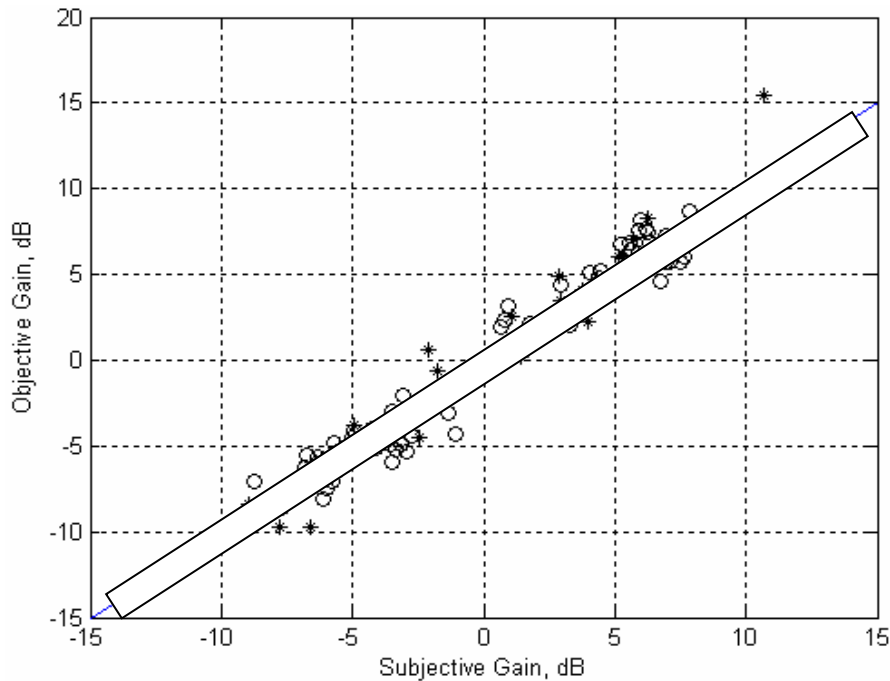
Neben dieser tabellarischen Performancematrix wird in [ 8 ] für jeden Lautstärkemesser zusätzlich die Verteilung der gemessenen objektiven über den subjektiven Lautstärkewerten grafisch dargestellt. Diese Darstellung erlaubt eine gute optische Interpretation in Bezug auf die Leistungsfähigkeit der untersuchten Lautstärkemesser. In Abb. 4&5 sind die entsprechenden Diagramme beispielhaft für die Lautstärkemesser [ I ] und [ L ] dargestellt. In den Diagrammen erkennt man, dass die Leistungsfähigkeit des betrachteten Lautstärkemessers durch die Streubreite der Messwerte beschrieben wird. Je enger sich die Messwerte entlang der Diagonalen häufen, desto besser bilden die objektiven die subjektiven Lautstärkewerte ab.

Die Schlussfolgerungen in [ 8 ] lauten sinngemäß: LeqBC kann unter Berücksichtigung aller definierten Qualitätsparameter als bester Lautstärkemesser eingestuft werden. Daneben muss Leq (ohne Bewertung) als fast gleich gut eingestuft werden. Diese Ergebnisse bestätigen, dass eine einfache, Energie basierte Lautstärkemessung bei rundfunktypischem Programmmaterial stabiler ist als komplexe Maße, die auf detaillierten psychoakustischen Modellen basieren können.

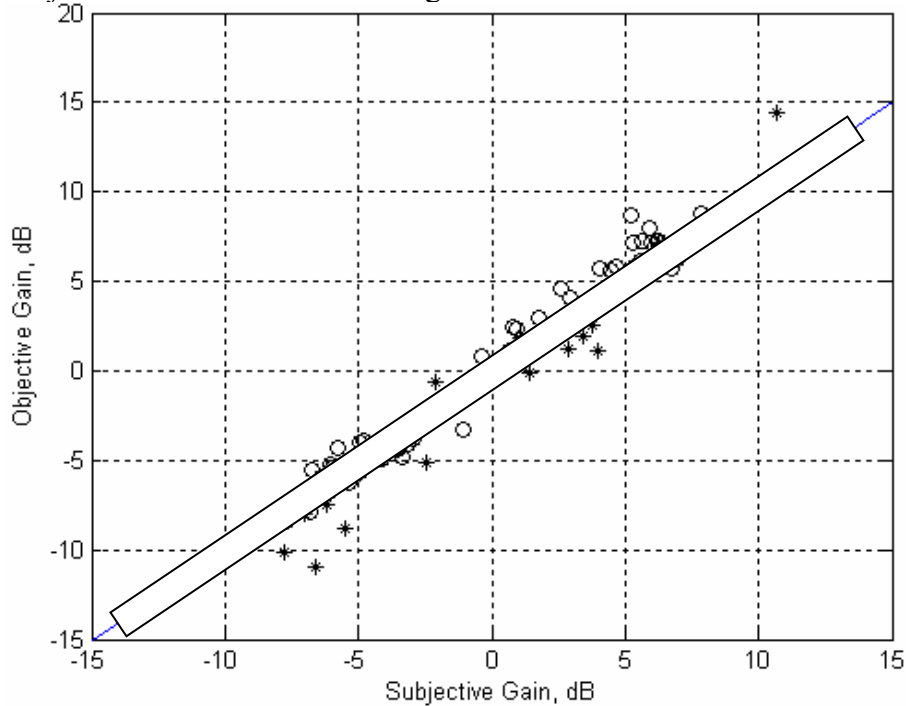
#### **4 Anmerkungen zu den Ergebnissen SRG-Studie**

Im Hinblick auf die beiden in der Auswertung zu Grunde gelegten Korrelationsparameter (R, Spearmanrho) (Tab.) fällt zunächst auf, dass insgesamt relativ hohe Korrelationen mit den subjektiven Beurteilungen erzielt wurden. Betrachtet man z. B. Spearmanrho, so erreichen 9 von 12 Lautstärkemessern Korrelationsfaktoren  $> 0.95$ . Das scheint – aus Sicht des Autors - ein Ergebnis zu sein, das für Rundfunkanwendungen durchaus ausreichend ist. Es mag hier

die Frage erlaubt sein, ob die abgeleitete qualitative Rangfolge, die sich formal aus den berechneten Zahlenwerten ergibt, tatsächlich in der Praxis signifikant ist oder ob es sich um einen eher akademischen Unterschied handelt.



**Abb. 4:** Einzelergebnisse für Meter [ I ]. Kennzeichnung des 95%-VB des Mittelwertes der subjektiven Lautstärkebeurteilungen



**Abb. 5:** Einzelergebnisse für Meter [ L ] (= Leq ohne Bewertung). Kennzeichnung des 95%-VB des Mittelwertes der subjektiven Lautstärkebeurteilungen

#### **4.1 Berücksichtigung des 95%-VB des Mittelwertes der subjektiven Lautstärkebeurteilungen**

Die Ergebnisse in Bezug auf einen wichtigen Stichprobenparameter der subjektiven Datenbasis, der 95%-Vertrauensbereich des Mittelwertes (95%-VB), werden in [ 8 ] zwar angegeben ( $95\text{-VB} < 2 \text{ db}$ ), bei der Berechnung der neun relevanten Qualitätsparameter bleibt dieser Parameter jedoch unberücksichtigt.

Um zu testen, in welcher Weise sich die Berücksichtigung des 95%-VB bei der Qualitätsrangfolge auswirkt, wurde für die Lautstärkemesser [ I ] und [ L ] eine Neuberechnung der Parameter RMSE, AAE durchgeführt.

Wie bekannt, definiert der 95%-VB einen Toleranzbereich (Mittelwerte  $\pm 95\text{-VB}$ ) für den gesuchten Erwartungswert (Mittelwert der Grundgesamtheit bei z. B. Stichproben  $n > 1000$ ). D. h. bei unterschiedlichen Stichproben werden sich die Mittelwerte jedes Mal leicht unterscheiden. Es kann mit einer statistischen Sicherheit von 95% jedoch vorausgesagt werden, dass der relevante Erwartungswert innerhalb des 95%-VB liegt. Für die Auswertung kann man daraus folgern, dass Werte innerhalb des 95%-VB ignoriert werden können.

Der entsprechende Toleranzbereich, der im Wesentlichen von der Verteilung, der Anzahl und der statistischen Fehlerwahrscheinlichkeit abhängt, liegt etwa bei  $\pm 1 \text{ db}$ , wie aus vergleichbaren eigenen Messungen bekannt ist.

Der 95%-VB sollte z. B. bei der Berechnung der Parameter RMSE, AAE, in der Weise berücksichtigt werden, dass nicht alle absoluten Unterschiede bzw. Fehler zwischen subjektiver und objektiver Messung berechnet, sondern nur die subjektiven Daten außerhalb des 95%-VB berücksichtigt werden.

Um die Unterschiede, die sich mit und ohne Berücksichtigung des 95%-VB angeben zu können, wurden die Meter [ I ] und [ L ] entsprechend analysiert.

Für diese Analyse wurde ein 95%-VB von  $\pm 1 \text{ dB}$  angenommen. Das bedeutet, es wurden bei beiden Systemen nur die Abweichungen der objektiven Werte berücksichtigt, die außerhalb des 95%-VB liegen (Abb. 4&5).

Für die relevanten Werte wurden die Parameter RMSE und AAE neu berechnet. Die entsprechenden Ergebnisse sind in Abb. 6 dargestellt.

Im Vergleich zu den SRG3-Werten (Tab.) ergibt sich für den Qualitätsunterschied zwischen Meter [ I ] und [ L ] eine deutlich andere Beurteilung, nämlich (I“ besser als L“).

Meter [ I ]: RMSEoVB=1,36; RMSEmVB=1,2

Meter [ L ]: RMSEoVB=1,26; RMSEmVB=2,7

## 4.2 Praktische Relevanz der SRG3-Qualitätsparameter

Im Folgenden wird versucht, die Signifikanz der formellen qualitativen Rangfolge der untersuchten Lautstärkemesser zu beleuchten. Aus praxisorientierter Sicht werden von den insgesamt neun definierten Parametern insbesondere die Parameter Sprmnrho, RMSE und AAE als relevant betrachtet. Der Korrelationsparameter Sprmnrho ist für diese Betrachtung u. a. deshalb geeignet, weil Sprmnrho eine Variable der „nicht parametrischen Statistik“ ist und damit keine Analyse der Verteilung erforderlich ist.

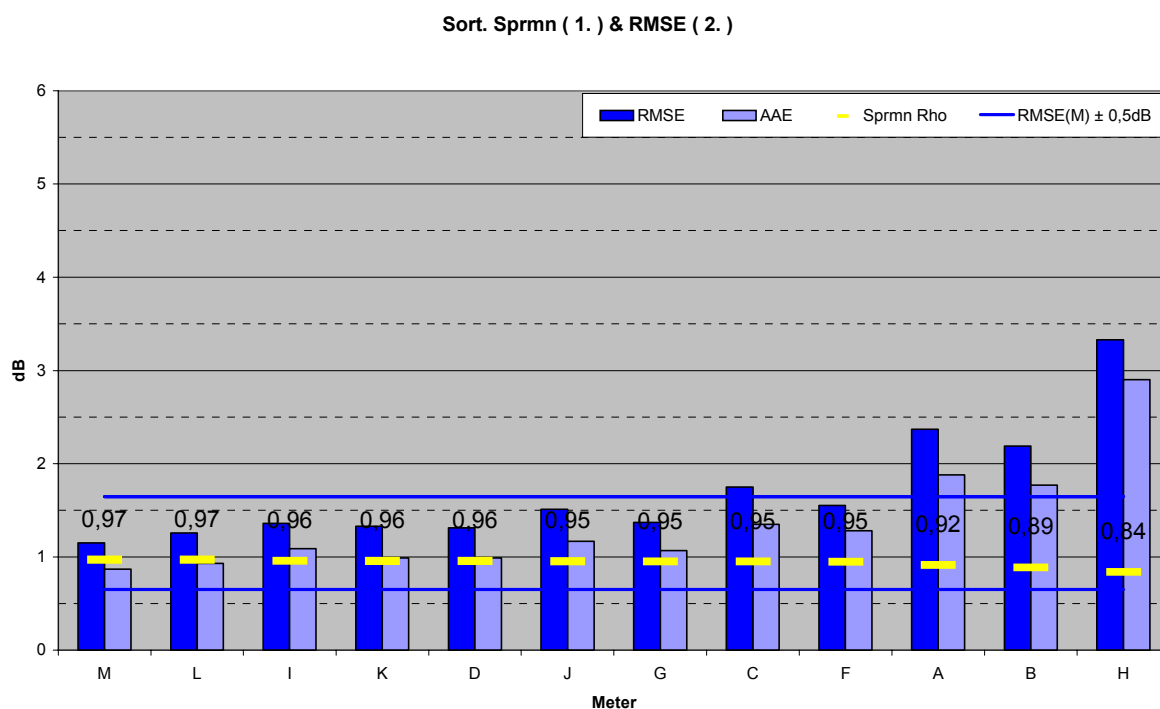
AAE ist als Parameter, der die absoluten Fehler erfasst, als grundlegend anzusehen. Das gilt ebenso für RMSE, der sich prinzipiell durch die quadratische Bewertung von AAE ergibt, die kleine Fehler geringer und große Fehler stärker bewertet.

Zunächst wird eine Rangfolge für die hier betrachteten Parameter angenommen. Diese lautet Sprmnrho ( 1 ), RMSE ( 2 ). Danach werden die Ergebnisse unter Bezug auf die festgelegte Rangfolge sortiert.

Formell ergibt sich daraus eine bestimmte Rangfolge der untersuchten Meter (Abb. 6). [ M ] der „beste“ und [ H ] der „schlechteste“ Lautstärkemesser.

Diese - mit den angegebenen Kriterien - berechnete Rangfolge sollte natürlich in Bezug auf Signifikanz beurteilt werden. Dazu bieten sich elementare statistische Methoden an, wenn man auf die Datenbasis zurückgreifen kann (z. B. Wilcoxon-Test).

Steht einem der komplette Basis-Datensatz nicht zur Verfügung, kann man den 95%-VB heranziehen, um Abschätzungen über die Signifikanz der dargestellten Unterschiede zu machen. In Abb. 6 wurde z. B. ein realistischer Wert von  $\pm 0,5$  dB als Toleranzbereich - bezogen auf den RMSE-Wert des Meters [ M ] - zu Grunde gelegt. Daraus würde sich ergeben, dass sich die Meter [ M ], [ L ], [ I ], [ K ], [ D ], [ J ], [ G ] und [ F ] nicht signifikant unterscheiden.



**Abb. 6:** Sprmnrho, RMSE & AAE (sortiert nach Sprmnrho (1.) & RMSE (2.)) – Toleranzbereich RMSE(M)  $\pm 0,5$  dB

### 4.3 Testsignaleigenschaften

Abschließend eine wichtige Anmerkung zu den verwendeten Testsignalen der SRG3-Studie. Es handelt sich ausschließlich um monofone Testsequenzen, die beim Lautstärkevergleich über einen Lautsprecher reproduziert wurden. Das bedeutet, dass alle Lautstärkealgorithmen im Test nur auf der Analyse des zur Verfügung stehenden monofonen Signals basieren konnten. Da heute in der Audioproduktion sowohl im Rundfunk und Fernsehen als auch in der kommerziellen Produktion monofone Signale kaum noch eine Rolle spielen, ja heute neben Stereo- bereits Mehrkanal-Musikproduktionen (5.1-Produktionen) stark verbreitet sind, ist ein universeller Lautstärkemesser für Stereo- und Mehrkanal-Audioproduktionen erforderlich. Ob die für monofone Signale getesteten (und möglicherweise entwickelten) Lautstärkealgorithmen auch für Stereo- und Mehrkanalproduktionen geeignet sind, bleibt offen.

## **5 Orientierende Experimente zur Lautstärke von 5.0 Surroundsound- und 2.0 Stereo-Downmix-Signalen**

Für den IRT-Lautstärke-Algorithmus gilt jedenfalls, dass er stereotauglich ist. Denn die Messung basiert auf der separaten Messung des linken und rechten Stereokanals, wobei die resultierenden Lautstärken des rechten und linken Kanals über eine einfache Beziehung verknüpft werden. Die im Rahmen der Algorithmus-Entwicklung durchgeführten subjektiven und die in letzter Zeit durchgeführten und veröffentlichten Lautstärkemessungen mit dem IRT-Lautstärkemesser [ 6, 7 ] basierten immer auf stereofonen Audiosignalen.

### **5.1 Durchgeführte Experimente**

Um zu testen, wie relevant die mit dem – für Stereosignale entwickelten - IRT-Algorithmus gemessenen Lautstärken in Bezug auf mehrkanalige 5.0-Aufnahmen sind, wurden folgende orientierenden Experimente durchgeführt.

Zunächst wurden für die Experimente 10 Beispiele aus 5.0-Produktionen unterschiedlicher Genres (durch entsprechende Blenden) mit einer durchschnittlichen Länge von 10 sec ausgewählt. Im Einzelnen wurden folgende Testbeispiele verwendet: Rock, Folk, Jazz (2), Bigband, Hörspiel, Atmo, Orchester (3).

Um eine definierte Aussteuerung aller Beispiele zu gewährleisten, wurden alle Sequenzen einheitlich entsprechend [ 1, 2 ] angesteuert, also mit  $QPPM_{max} = -9$  dB FS. Für die so normalisierten Testsequenzen wurde der 2.0-Downmix entsprechend [ 9 ] erzeugt.

$$L' = L + 0,7C + 0,7LS$$

$$R' = R + 0,7C + 0,7RS$$

Dieser Stereo-Downmix war der Bezug für die orientierenden subjektiven und objektiven Lautstärkemessungen.

Dazu wurden die ausgewählten 5.0-Sequenzen mit den entsprechenden 2.0-Downmix unter dem Aspekt „gleiche subjektive Lautstärke“ miteinander verglichen. Dementsprechend hatten die Probanden im Experiment die Aufgabe, das Stereosignal mit Hilfe eines Stereopegelstellers auf die gleiche Lautstärke einzupegeln wie das 5.0-Surroundsignal. Das Experiment wurde unter Studioabhörbedingungen entsprechend [ 10 ] mit fünf

Expertenhörern durchgeführt, wobei vor dem eigentlichen Lautstärkevergleich für jede 5.0-Testsequenz die optimale Abhörlautstärke festgelegt wurde.

Die resultierenden Dämpfungswerte bzw. Pegelwerte der Stereo-Downmix sind also ein Maß für die subjektiven Lautstärken der 5.0-Surroundsignale.

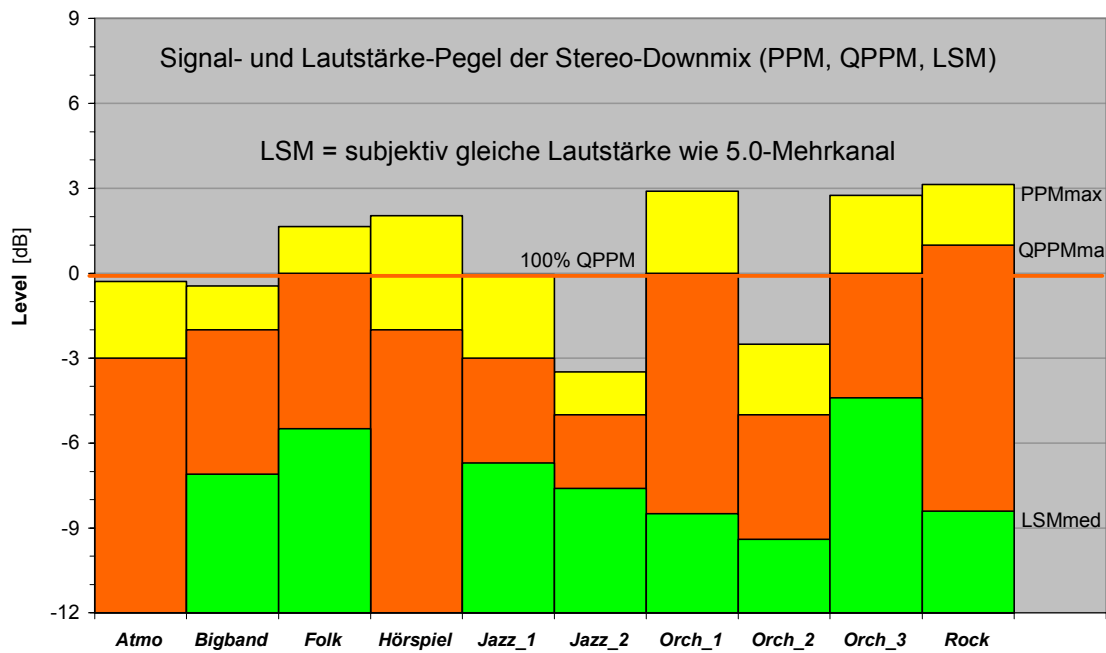
Zur Messung der Referenz- Lautstärkepegel kann selbstverständlich der IRT-Lautstärkemesser eingesetzt werden, mit dem die Lautstärkepegel LSM der Stereo-Downmix gemessen werden können. Die entsprechenden Messergebnisse sind in Abb. 7 dargestellt, wobei die bekannte Darstellung der Pegel PPMmax, QPPmax und LSMmedian verwendet wird.

Um nun einen Bezug zwischen den Lautstärkepegeln der Stereo-Downmix und den 5.0-Surroundsignalen herzustellen, werden außerdem die Lautstärkepegel der Surroundsignale kanalweise mit dem IRT-Lautstärkemesser gemessen, und zwar LSM (Lf R f), LSM (C) und LSM (Ls Rs). Die gemessenen Kanal-Lautstärkepegel sind in Abb. 8 dargestellt. Gleichzeitig werden in Abb. 8 die Referenz- Lautstärkepegel LSS der „gleich lauten“ Downmix dargestellt.

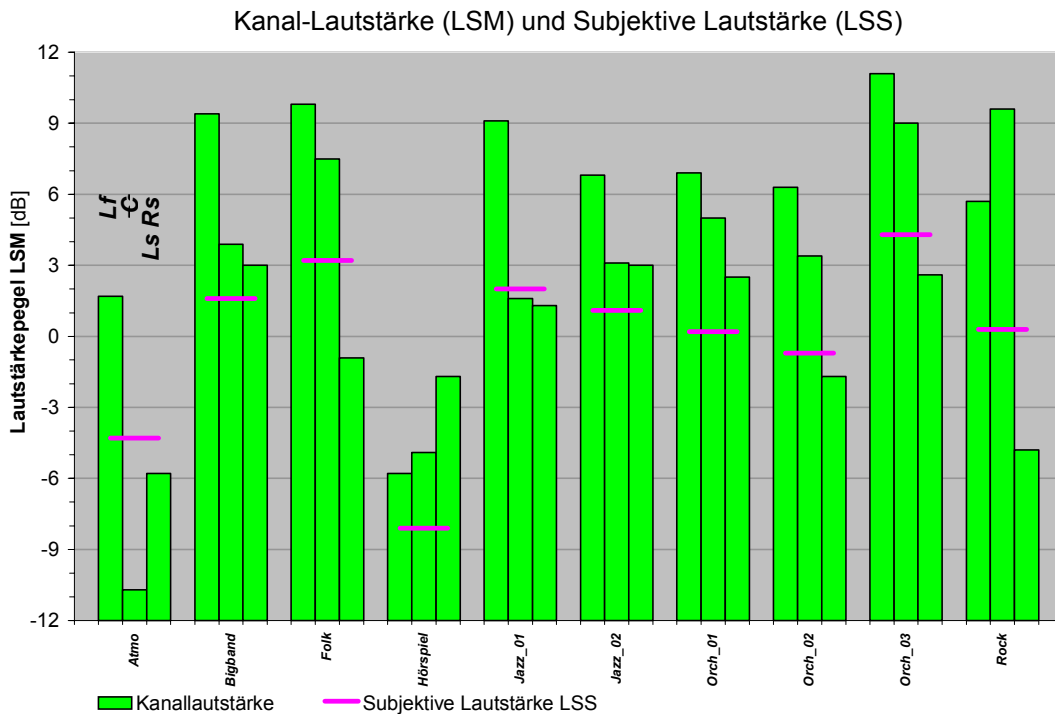
Um aus diesen Ergebnissen einen Lautstärke-Algorithmus für 5.0-Surroundsingnale ableiten zu können, müssen die Kanal-Lautstärkepegel so miteinander verknüpft werden, dass sich der jeweilige dargestellte Referenz- Lautstärkepegel LSS mathematisch ergibt. Mit anderen Worten, es geht um eine Optimierung der folgenden Lautstärkesummierung bzw. der Faktoren  $\alpha$ ,  $\beta$ ,  $\gamma$ :

$$LS(5.0) = \alpha(LSM(Lf Rf)) + \beta(LSM(C)) + \gamma(LSM(Ls Rs))$$

Diese Optimierung scheint nicht unlösbar. Wir sind deshalb optimistisch, auf der Basis des stereofonen IRT-Lautstärkemessers einen Mehrkanal-Lautstärkemesser zu entwickeln, der die Lautstärke von 5.1-Surroundsignalen - mit der gleichen hohen Korrelation mit der subjektiven Lautstärke wie bei Stereosignalen - messen kann. Weiterführende Experimente werden darüber Aufschluss geben.



**Abb. 7:** Signal- und Lautstärke-Pegel der Stereo-Downmix mit gleicher Lautstärke wie die 5.0-Testsequenzen



**Abb. 8:** Kanal-Lautstärken (LSM) und Subjektive Lautstärken (LSS) für ausgewählte 5.0-Testsequenzen



## 6 Ausblick

Selbstverständlich wurden die Ergebnisse der SRG3-Studie eingehend diskutiert. Daraus ergaben sich folgende Schlussfolgerungen. Die Leistungsfähigkeit der relativ einfachen RMS- oder Leq-Messung (BC- /ohne Bewertung) im Vergleich mit den anderen komplexen, weitgehend auf psychoakustischen Prinzipien basierenden, Algorithmen wird in Frage gestellt. Im Wesentlichen wurden zwei Ursachen für dieses Ergebnis diskutiert, eine vermutete zu geringe Varianz der Signaleigenschaften innerhalb der Testsequenzen und die Beschränkung auf das monofone Audioformat.

Als Konsequenz dieser intensiven Diskussion wurde beschlossen, als Voraussetzung für eine internationale Standardisierung eines Rundfunklautstärkemessers eine ergänzende SRG3-Studie durchzuführen, die auf neuem Testmaterial mit größeren Signalunterschieden basiert und insbesondere Stereo- und Mehrkanal-Aufnahmen einbezieht. Zurzeit - November 2004 - arbeitet SRG3 an der Planung der neuen Messungen.

Selbstverständlich wird das IRT die weitere Arbeit der SRG3 beobachten und unterstützen. Kommt es nach Abschluss der neuen SRG3-Studie tatsächlich zu einer internationalen Standardisierung eines Rundfunklautstärkemessers, z. B. auf der Basis der RMS-Messung, so könnte der bisher in den beiden Prototypen (DAG2000, Pinguin) implementierte IRT-Algorithmus ohne Probleme durch den neuen standardisierten Algorithmus ersetzt werden.

Dadurch dass der IRT-Algorithmus als praxistauglicher Prototyp realisiert wurde, ergeben sich entscheidende Vorteile für dieses Konzept. Um den Lautstärkemesser tatsächlich in der Praxis einsetzen zu können, müssen neben dem eigentlichen Algorithmus eine Reihe von zusätzlichen Konstruktionsmerkmalen einbezogen und realisiert werden. Das sind unter anderem die Echtzeitanzeige, das Skalenlayout, die Möglichkeit der Registrierung der Messwerte und die Mischpultintegration. Von den Prototypen des IRT-Lautstärkemessers wurden diese zusätzlichen Anforderungen bereits erfüllt, so dass diese Geräte prinzipiell in der Praxis einsetzbar sind und bereits mit Erfolg eingesetzt wurden.

Von der SRG3 wurden diese Anforderungen bisher nur ansatzweise diskutiert. Bei den bisher durchgeführten Messungen waren diese Anforderungen deshalb nicht relevant, weil nur

statisch gemessen wurde, d. h. die getesteten Lautstärkemesser mussten je Testbeispiel jeweils nur einen (mittleren) Bezugswert anzeigen. Bei der internationalen Standardisierung müssen die genannten konstruktiven Voraussetzungen selbstverständlich berücksichtigt werden.

## 7 Literatur

- [ 1 ] Technische Richtlinien zur Herstellung von Fernsehproduktionen für ARD, ZDF und ORF. München: Institut für Rundfunktechnik, Mai 2003
- [ 2 ] Empfehlung 15 IRT der ARD-Hörfunkbetriebsleiterkonferenz: Headroom bei digitalen Tonsignalen. München: Institut für Rundfunktechnik, Okt. 1994
- [ 3 ] Technische Pflichtenhefte der öffentlich-rechtlichen Rundfunkanstalten in der Bundesrepublik Deutschland. Pflichtenheft Nr. 3/6 „Aussteuerungsmesser“
- [ 4 ] Spikofski, Gerhard: Lautstärkemessung im Rundfunk-Sendestudio (Loudness measurement in broadcast studios). Tonmeistertagung <21, 2000, Hannover>: Bericht. München: Saur, 2001, S. 604 – 618
- [ 5 ] Gerhard Spikofski and Siegfried Klar: Levelling and loudness in radio and television broadcasting. EBU-Technical Review No. 297, January 2004
- [ 6 ] Gerhard Spikofski and Siegfried Klar: DAB and CD quality – reality or illusion? EBU-Technical Review No. 296, October 2003
- [ 7 ] Gerhard Spikofski, Siegfried Klar: Lautstärkeunterschiede bei digitalen Fernsehproduktionen über Satellit. Tonmeister-Informationen VDTmagazin, Heft 3/2004, S. 16-23
- [ 8 ] Gilbert A. Souludre: Evaluation of Objective Loudness Meters. Convention Paper presented at the 116th Convention 2004 May 8-11 Berlin, Germany
- [ 9 ] Recommendation ITU-R BS.775-1: Multichannel Stereophonic Sound Systems With And Without Accompanying Picture.
- [ 10 ] EBU Tech. 3276 (2nd edition and Supplement 1): Listening conditions for the assessment of sound programme material: Monophonic and two-channel stereophonic. [http://www.ebu.ch/tech\\_t3276.pdf](http://www.ebu.ch/tech_t3276.pdf)

