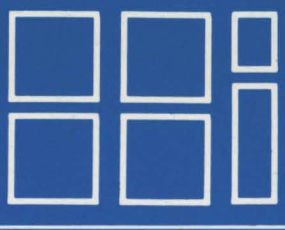


Heinrich-Hertz-Institut
für Nachrichtentechnik
Berlin GmbH

TECHNISCHER BERICHT NR. 213

BEURTEILUNG DER BILDGÜTE IN KABELFERNSEHANLAGEN

von

Dipl.-Ing. Helmut Bünning

14. Dezember 1979

Technischer Bericht Nr. 213

Beurteilung der Bildgüte in Kabelfernsehanlagen

Zusammenfassung:

Die für die Bildgütebeurteilung geeigneten Verfahren werden einander gegenüber gestellt und auf Anwendbarkeit geprüft. Es zeigt sich, daß die Urteile über die Güte gestörter Fernsehbilder vorteilhafterweise mit Hilfe der Einzelbildbeurteilung mit fünfstufiger Güteskala beschafft werden können. Es wird eine Auswertungsmethode beschrieben, die je Störart einen einfachen Zusammenhang zwischen den Urteilsmittelwerten und den zugeordneten Störspannungspegeln unter Verwendung der logistischen Funktion aufdeckt. Es folgen ausführliche Angaben zur Durchführung des Beurteilungsverfahrens. Abschließend werden einige auf die vorliegende Aufgabenstellung anwendbare Ergebnisse aus der Literatur diskutiert.

Bearbeiter:

H. Bünning

(Dipl.-Ing. Helmut Bünning)

WISS.TECHN.GESCHÄFTSFÜHRER
(Dr. Horst Ohnsorge)

ABTEILUNGSLEITER
(Dr. Wolf Herold)

Berlin, den 14. Dezember 1979

JSSN 0172 - 8873

Inhaltsverzeichnis

		Seite
1	Einleitung mit Aufgabenstellung	3
2	Methoden zur Beurteilung gestörter Fernsehbilder (Datenerhebungsverfahren)	5
	2.1 Grundlagen	5
	2.1.1 Dimensionalität des Merkmalsraumes	5
	2.1.2 Skaleneigenschaften	6
	2.1.3 Mehrfach- und Einzelreizmethoden	7
	2.2 Einzelbildbeurteilung (opinion rating)	8
	2.2.1 Wahrnehmbarkeitsschwelle	11
	2.3 Bildpaarbeurteilung	11
	2.3.1 Beurteilung der Äquivalenz von gestörten Fernsehbildern	12
	2.3.2 Beurteilung bezüglich der Unterschiedlichkeit bzw. Ähnlichkeit	12
	2.3.3 Beurteilung nach der Präferenz (Paarvergleich)	13
	2.3.4 Beurteilung nach gestufter Präferenz (comparison method)	14
	2.3.5 Verhältnisschätzung (ratio estimation)	14
3	Auswertung der subjektiven und objektiven Daten	16
	3.1 Verknüpfung der subjektiven Werte (Urteilsmittelwert) mit den objektiven Werten (Störspannungspegel)	17

		Seite
	3.2 Beschreibung des subjektiven Datensatzes durch Urteilsverteilungen	21
4	Planung des Bildbeurteilungsverfahrens	24
	4.1 Übersicht	24
	4.2 Auswahl und Instruktion der Beurteilungspersonen	26
	4.3 Auswahl der Testbilder	27
	4.4 Festlegung der Beobachtungsbedingungen	28
	4.4.1 Beobachtungsplatz	28
	4.5 Apparativer Aufbau	29
	4.5.1 Messung des Störspannungspegels	30
	4.5.2 Erzeugung der Störarten, Wahl der Störstufenanzahl	31
5	Diskussion von Ergebnissen aus der Literatur	32
	5.1 Weißes Rauschen und Intermodulationsrauschen	33
	5.2 Kreuzmodulation	35
6	Zusammenfassung	37
7	Anhang	39
	Literatur	42
	Anlagen 1 und 2	
	Bilder 1 bis 3	

1 Einleitung mit Aufgabenstellung

Mambo und Coll / 1 / berichteten Anfang 1979 aus Kanada, daß die Anzahl der durch Kabelfernsehsysteme übertragenen Programme in den letzten Jahren ständig zugenommen hat und daß die Kanalkapazität in den Großstädten bereits voll ausgelastet ist. Die Fernsehgesellschaften suchen deshalb nach Möglichkeiten, die Kapazität für die Übertragung weiterer Fernsehsignale zu erweitern, ohne die Bildqualität in den bestehenden Kanälen zu beeinträchtigen. Durch die zunehmende Auslastung der Kabelfernsehbänder steigen die Störspannungspegel an, da durch nichtlineare Verzerrungen neue Intermodulations- (IM) und Kreuzmodulationsprodukte (XM) in den einzelnen Fernsehkanälen entstehen. Dadurch wird die Bildqualität je nach Kanallage unterschiedlich verschlechtert.

Nach der offiziellen Einführung des Kabelfernsehens in der Bundesrepublik Deutschland ist in den folgenden Jahren mit einer ähnlichen Entwicklung zu rechnen, so daß bei der Untersuchung der Auswirkungen von Störungen auf die Bildqualität von der vollen Kanalbelegung (hier: 30 Kanäle im regulären 7 MHz-Raster) ausgegangen werden muß. Die Bildqualität kann durch gesteigerten technischen und damit wirtschaftlichen Aufwand beispielsweise durch kürzere Verstärkerfeldlängen bei geringerer Aussteuerung der Breitbandstreckenverstärker verbessert werden. Da dieser Aufwand möglichst klein gehalten werden soll, kommt es darauf an, die von den Fernsehteilnehmern noch tolerierbare Bildqualität durch eine repräsentative Gruppe von Beurteilungspersonen (Bpn) subjektiv bestimmen zu lassen und durch Objektivierung meßbar zu machen. Gelingt dieses Vorhaben, so kann die subjektiv empfundene Bildqualität aufgrund physikalischer Meßwerte vorhergesagt werden. Wird diese Tolerierbarkeitsgrenze z. B. durch größere Kanaldichte überschritten, so sind technische Maßnahmen zur Verbesserung der Bildqualität zu ergreifen. Das wird durch ein im Rahmen dieses Vorhabens entwickeltesVerfahren erreicht, das eine Reduzierung der IM- und XM-Störungen durch phasenverkoppelte Bildträger bewirkt (Krick / 2 /).

Die Aufgabe der elektrovisuellen Übertragungskette besteht im Prinzip darin, das Originalbild am Ort des Zuschauers im Aufnahmeraum (z.B. Szene im FS-Studio) bzw. im Freien möglichst naturgetreu am Ort des Zuschauers im Wiedergaberaum (z. B. Wohnzimmer) zu reproduzieren. Sieht man von der bisher nicht angestrebten bzw. nicht realisierbaren Bildgröße und von der Bildtiefe (stereoskopisches Fernsehbild) ab, so wird die Bildgüte im wesentlichen durch das Umfeld, die Wiedergabe der Leuchtdichte (Luminanz-Signal), der Farbart (Chrominanz-Signal), der Bildschärfe und durch die Technik des Bildaufbaus (Zeilenstruktur, Flimmern durch periodische Bildwiederholung) bestimmt. Daneben entstehen im Übertragungsweg besonders bei hoher Aussteuerung der Streckenverstärker merkbare nichtlineare Verzerrungen, die die Bildqualität je nach Art und Intensität erheblich beeinflussen können. Sollen diese Bildstörungen allein und reproduzierbar erfaßt werden, so müssen die Aufnahme- (Farbfernsehkamera), die Übertragungs- (Modulatoren, Umsetzer, Streckenverstärker, Demodulatoren) und die Wiedergabetechnik (Farbfernsehmonitor) höchsten Anforderungen genügen bzw. standardisiert sein. Diese Bedingungen werden durch die im Haus installierte Kabelfernsehanlage erfüllt.
Ursprünglich sah die Planung die Bearbeitung der folgenden beiden Teilaufgaben im Rahmen dieses Vorhabens vor:

a) Um wieviele Bildgütestufen kann die Bildgüte durch Anwendung des Verfahrens der phasenstarr gekoppelten Bildträger verbessert werden (höhere Bildgüte) bzw. um welchen Faktor lassen sich die Verstärkerfeldabstände bei gleicher Bildgüte vergrößern (größere Reichweite)?

b) Auswirkung der verschiedenen Störarten (IM, XM, IM + XM) auf die Bildqualität bei gleichen Testbildern (Farbdias).

Nach dem Abschluß der Planungsphase zeigte es sich, daß die praktische Durchführung des o. g. Programms aus zeitlichen und gerätetechnischen Gründen nicht mehr möglich war. Deshalb wird

unter Punkt 5 versucht, diese Fragen, soweit wie möglich, anhand von geeigneten Literaturstellen zu beantworten.

2 Methoden zur Beurteilung gestörter Fernsehbilder (Datenerhebungsverfahren)

2.1 Grundlagen

Die unterschiedlich gestörten Fernsehbilder stellen Reize dar (Reizkontinuum), die von den Augen der Bpn aufgenommen (Reizerfassung) und von den Sehnerven bis zum Sehzentrum weitergeleitet werden (Reizübertragung). Im Sehzentrum erfolgt eine Verarbeitung des durch die Erfassung und Übertragung modifizierten Reizes, deren Ergebnis eine Empfindung ist (Empfindungskontinuum). Das von den Bpn abgegebene Urteil kann als Abbild der Einstufung dieser Empfindung auf einer inneren Merkmals-Skala (eindimensionales Urteil) oder auf mehreren orthogonalen Skalen (Urteil im mehrdimensionalen Merkmalsraum) aufgefaßt werden (Urteilskontinuum). Die Empfindung läßt sich also nur indirekt anhand des Urteils erfassen.

Die bei Akzeptanzuntersuchungen angewendeten Skalierungsmethoden unterscheiden sich voneinander durch die Dimensionalität des Merkmalsraumes (ein- bzw- mehrdimensionale Verfahren), durch die Skaleneigenschaften und durch die Anzahl der gleichzeitig dargebotenen Reize.

2.1.1 Dimensionalität des Merkmalsraumes

Aus Untersuchungen der Akzeptanz akustischer Darbietungen ist bekannt, daß der Beurteilung komplexer Hörereignisse (Wahrnehmungen) ein mehrdimensionaler Merkmalsraum zugrunde liegt, in dem die Urteile bzw. Wahrnehmungen gewisse Positionen einnehmen (vgl. Eisler / 3 /, McDermott / 4 /, v. Bismarck / 5 /, Wilkens / 6 /). Die wenigen diesen akustischen Perzeptionsraum aufspannenden orthogonalen Achsen heißen Beurteilungsaspekte, wie z. B. Schärfe, Volumen und Transparenz des Klanges. Nach Jost / 7 /,

Wilkens / 6 / kann jedes Güteurteil als Projektion des Urteils im mehrdimensionalen Merkmalsraum auf eine eindimensionale Güteachse aufgefaßt werden, deren Lage im Raum durch den Geschmack der Bpn bestimmt wird.

Im Gegensatz zu diesem kurz angedeuteten Wissensstand ist aus der Literatur zur Beurteilung von Fernsehbildern keine Veröffentlichung bekannt, in der die Dimensionalität eines visuellen Merkmalsraumes untersucht wird. Wendet man das oben gesagte auf visuelle Wahrnehmungen an, so ist die Kenntnis des visuellen Merkmalsraumes nach Anzahl und Art der Dimensionen auch nicht von Bedeutung, wenn man sich nur für die Einstufung gestörter Fernsehbilder bezüglich der eindimensionalen "Über-Alles-Qualität" (Güte) interessiert, ohne gleichzeitig aufdecken zu wollen, aus welchen ganz spezifischen Eigenschaften die Bildqualität zusammengesetzt ist. Für die Einstufung von gestörten Fernsehbildern nach der Bildgüte genügen demnach eindimensionale Datenerfassungs- und Auswertungsverfahren.

2.1.2 Skaleneigenschaften

Die für die Einstufung von Empfindungen verwendeten Skalen weisen entweder Ordinal- oder Intervallskalen-Niveau auf.

Der Ordinal- bzw. Rangskala werden in bestimmter Reihenfolge geordnete Zahlen zugewiesen, ohne daß über deren Abstand eine Aussage gemacht wird.Mit Hilfe dieser Skala läßt sich nur die Rangfolge der beurteilten Objekte bezüglich bestimmter Urteilskriterien (z. B. Farbsättigung, Kontrast, Qualität, Ausmaß der Störungen) festlegen.

Die Intervall- bzw. Einheitenskala wird in mehrere Intervalle gleicher Breite (Einheit) unter Zuweisung geordneter Zahlen unterteilt. Sie liefert zusätzlich zur Rangskala Maßzahlen für das Ausmaß der im Reiz vorhandenen spezifischen Eigenschaft bzw. Qualität. Durch die Zuordnung von Adjektiva zu den abstrakten Zahlen wird die Beurteilung der Objekte durch die Verankerung der Skala im Sprachschatz der Bpn vereinfacht. Es besteht jedoch bei falscher

Wahl der Benennung der Skalenstufen die Gefahr der Skalenverzerrung, wenn die den Adjektiva zugeordneten Empfindungen nicht äquidistant sind (Überprüfung nach Prosser et al. / 11 /). Die Festlegung der Skalenendpunkte kann durch verbale Beschreibung und durch Vorführung extremer Bildbeispiele vor dem Test erfolgen. Eine ausführliche Beschreibung von Skalierungstechniken findet sich bei Jackson / 8 /.

2.1.3 Mehrfach- und Einzelreizmethoden

Die Datenerhebungsverfahren lassen sich auch nach der Anzahl der gleichzeitig den Bpn zur Beurteilung angebotenen Beurteilungsobjekte (Reize) in Mehrfach- und Einzelreizmethoden unterteilen. Zwei Gesichtspunkte spielen bei der Auswahl der für unseren Anwendungsfall geeigneten Methode eine Rolle:

a) Die Anzahl der insgesamt zu beurteilenden Objekte, die den Zeit- und Kostenaufwand für das Beurteilungsverfahren bestimmt, und

b) die organisatorische und technische Realisierbarkeit im Rahmen eines vertretbaren technischen Aufwandes.

Zu a) Die Anzahl n der zur k-ten Klasse zu kombinierenden gleichzeitig darzubietenden Bildkombinationen (k gleiche Fernsehbilder mit n_S verschiedenen Störspannungsstufen) ergibt sich aus:

$$n = n_B \cdot \binom{n_S}{k} \qquad (1)$$

wobei n_B = Anzahl der Testbilder (Farbdias)
$k = 2$ beim Paarvergleich

Für $k \geq 2$ steigt die Anzahl der Beurteilungsobjekte überproportional mit n_s an. Vom Zeitaufwand her gesehen ist also eine Einzelreizmethode (s. Pkt. 2.2) vorzuziehen, die nur $n = n_B \cdot n_s$ Beurteilungen je Störart verlangt.

zu b) Der technische Aufwand steigt mit der Anzahl k der gleichzeitig zur Beurteilung vorzustellenden unterschiedlich gestörten Bilder derselben Vorlage. Bei echtem Simultanbetrieb der KTV-Anlage müßte die gerätetechnische Ausstattung von der Eichleitung an (s. Bild 1) k mal vorhanden sein (d.h. k Eichleitungen, Streckenverstärker, Mischer, Demodulatoren und Farbmonitore). Ein Zeitmultiplexbetrieb kann schon deshalb ausgeschlossen werden, da er die Bildqualität erheblich beeinflussen dürfte.

Da nur eine KTV-Anlage vorhanden ist, läßt sich bei Echtzeitbetrieb (d.h. ohne Bildspeicher) nicht einmal ein Paarvergleich mit simultaner Darbietung gleicher Bilder mit unterschiedlichen Störstufen (derselben Störart) realisieren. Es muß also auf eine sukzessive Darbietung ausgewichen werden. Die Reihenfolge der Bilder muß dann zufällig sein.

Es hat sich also gezeigt, daß aus zeitlichen und technischen Gründen nur zwei Verfahren im Hinblick auf die Art der Darbietung der Beurteilungsobjekte in die engere Wahl kommen:

die Einzelbildbeurteilung und
die sukzessive Bildpaarbeurteilung (Paarvergleich).
Diese sollen in den folgenden Abschnitten besprochen werden.

2.2 Einzelbildbeurteilung (opinion rating)

Bei dieser Methode wird jedes Objekt (Bild-Störstufenkombination) in zufälliger Reihenfolge einzeln anhand von Intervallskalen bezüglich der Bildgüte und / oder der Bildstörung beurteilt.

Das Urteil entsteht durch die Einstufung einer vom Reiz herrührenden Empfindung auf einer inneren Referenzskala, deren Gegenstück die externe Skala ist.

Wegen des relativ geringen Aufwands und wegen der Verfügbarkeit eines an einer Kategorieskala orientierten "Absoluturteils" ist dieses Verfahren für die Beurteilung von Fernsehbildern in der Literatur bevorzugt angewendet worden (/ 1 /, / 11 /, / 13 /, / 19 / bis / 27 /). Durch die Verwendung unterschiedlicher Skalen (quality-scale, impairment-scale) und hauptsächlich durch die unterschiedliche Anzahl und Benennung der Skalenstufen wurde lange Zeit eine vergleichende Interpretation der Forschungsergebnisse erschwert (Zusammenstellung bei Prosser et al. / 11 / und im C.C.I.R.-Report 405-2 / 12 /). Die C.C.I.R-Empfehlung 500 / 12 / schlägt zur Vereinheitlichung neben der Anzahl und Art der Bpn (s. Pkt. 4.2), der Testbilder und den Beobachtungsbedingungen (s. Pkt. 4.4) eine fünfstufige Qualitätsskala und eine fünfstufige Störungsskala vor.

Für den deutschen Sprachraum erscheint die folgende Übersetzung der Skalenstufen in Anlehnung an die o.g. C.C.I.R.-Empfehlung sinnvoll zu sein, wobei sich gemäß Schneeberger / 13 / für die Güteskala eine den Schulnoten entsprechende Bewertung der Skalenstufen als zweckmäßig erwiesen hat (s. Anlage 1).

Note	Beschreibung der Bildqualität	Beschreibung der Bildstörung
1	sehr gut	nicht sichtbar
2	gut	sichtbar, aber nicht störend
3	mittelmäßig	etwas störend
4	schlecht	störend
5	sehr schlecht	stark störend

Die Skalen werden durch die Endpunkte "bestmögliche Qualität" (sehr gut) bzw. "Bildstörung nicht sichtbar" und die "schlechtestmögliche Qualität" (sehr schlecht) bzw. "Bildstörung stark störend"

begrenzt. Die Stufenbeschreibung sollte so gewählt sein, daß die zugehörigen subjektiven Intervalle gleich breit sind. Außerdem ist zu beachten, daß die in den Bpn begründete etwas unterschiedliche Auffassung der die Skalenstufen beschreibenden Adjektiva zu einer erhöhten Urteilsvarianz zwischen den Bpn je Reiz führen kann. Diesem Einfluß kann durch die Anweisung an die Bpn entgegengewirkt werden, bevorzugt die abstrakte (als äquidistant angenommene) Notenskala zu verwenden, wobei die Adjektiva als erläuternde Beschreibung, nicht als Definitionen, aufzufassen sind (vgl. / 11 /).

Die Bpn zeigen gemäß Allnatt und Corbett / 14 / die Tendenz, ihre Urteile in die Mitte des angebotenen Reizfeldes zu legen, wodurch die Vergleichbarkeit von Ergebnissen bei Verwendung von nach Lage und Umfang unterschiedlichen Reizfeldern nur bedingt gegeben ist. Gemäß / 12 / läßt sich der Reizumfang durch das Angebot von sehr guten (nicht gestörten) und sehr schlechten (stark gestörten) Bildern nahezu standardisieren.

Die innere Referenzskala ist nicht nur personenabhängig (geringfügig) unterschiedlich, sondern sie kann sich auch durch Anpassungsvorgänge beim Test und durch Ermüdung ändern. Durch diese Einflüsse eignet sich das Verfahren der Einzelbildbeurteilung nur für mittlere bis große Reizunterschiede. Wegen des relativ geringen Aufwandes bei der Datenerhebung,kann eine relativ große Anzahl von Objekten beurteilt werden. Geht man davon aus, daß sich die inneren Referenzskalen der Bpn durch einen Lernprozeß vor dem Test stabilisiert haben, so sind die am festen inneren Maßstab orientierten Urteile als "Absoluturteile" anzusehen, was als weiterer Vorteil zu werten ist.

Wählt man die Reizunterschiede zwischen den Objekten groß genug, so kann das Verfahren aufgrund der o. g. Vorteile zur Anwendung kommen.

2.2.1 Wahrnehmbarkeitsschwelle (WS)

Es wäre denkbar, die WS der am kritischsten beurteilten Störart als Gütekriterium für Fernsehbilder zu verwenden. Der Aufwand bei der Datenerfassung ist relativ gering, das durch die WS definierte Qualitätsniveau aber sehr hoch, so daß mit relativ hohen Kosten für die technische Realisierung gerechnet werden muß.

Für die Ermittlung der WS sollte ein Graubild wegen der hohen Sensibilität gegenüber Störungen benutzt werden. Die Erfassung der WS für verschiedene Störarten kann auf zwei Arten unter Einhaltung der Beobachtungsbedingungen gemäß Pkt. 4.4 erfolgen:

a) Vorstellung eng benachbarter Störstufen in der Nähe der WS, Beurteilung anhand der Skala mit den Stufen "Störung sichtbar" bzw. "Störung nicht sichtbar"

b) Die Bpn stellt den Störabstand an einer Eichleitung so ein, daß die Störung erstens gerade sichtbar und zweitens gerade unsichtbar wird. Die zugehörigen Störspannungspegel werden meßtechnisch erfaßt. Diese Prozedur ist zur Erhöhung der Aussagesicherheit mehrfach zu wiederholen.

2.3 Bildpaarbeurteilung

Zwei unterschiedlich gestörte Fernsehbilder derselben Vorlage werden den Bpn gleichzeitig bzw. kurz nacheinander zur Beurteilung vorgestellt. Beim klassischen Paarvergleich werden alle möglichen Bildpaarkombinationen gezeigt. Andere Verfahren arbeiten mit einem festen Referenzbild, das häufig in die Mitte des Reizfeldes gelegt wird.

Die Urteile ergeben sich bei den eindimensionalen Verfahren aus dem Vergleich zweier Empfindungen bezüglich eines Urteilskriteriums (einer Skala: z.B. Präferenz, Schärfe etc.). Im mehrdimensionalen Merkmalsraum wird die Unterschiedlichkeit als umfassendes Urteilskriterium verwendet. Je nach der Anzahl der Skalenstufen werden die Rangfolge der Beurteilungsobjekte (besser - schlechter) bzw.die Distanzen zwischen den Beurteilungsobjekten (drei- und mehrstufige Skalen) ermittelt. Die Verhältnisschätzung liefert ebenfalls Relativurteile, die als Quotienten zweier Empfindungen auf einer Skala mit Nullpunkt aufgefaßt werden können.

Mit Hilfe der o. g. Verfahren werden vorhandene Unterschiede zwischen gestörten Fernsehbildern beurteilt. Sieht man nun eine von der Bp einstellbare Störgröße für ein Fernsehbild vor, mit der die Urteilsdistanz zum zweiten Bild gegen Null verändert werden kann, so kommt man zu einem Verfahren, bei dem die Beurteilung der Gleichwertigkeit zweier Störungen gefordert wird.

2.3.1 Beurteilung der Äquivalenz von gestörten Fernsehbildern

Den Bpn werden zwei Bilder derselben Vorlage angeboten, wobei auf der einen Seite (linke Bildschirmseite bzw. linker Monitor) das durch eine bzw. mehrere Störarten gestörte Bild und auf der anderen Seite (rechte Bildschirmseite bzw. rechter Monitor) ein Referenzbild mit einer variablen Störgröße (z.B. additives weißes Rauschen) gezeigt werden. Die Bpn haben nun die Störgröße so einzustellen, daß beide Bilder bezüglich der Qualität bzw. der Störwirkung gleichwertig sind.

Mit dieser Methode können verschiedenartige Bildstörungen (z.B. aus mehreren Störarten zusammengesetzte Störungen) im Verhältnis zur Qualität bzw. Störwirkung einer bestimmten anderen Störart (z.B. weißes Rauschen) untersucht werden.

Als Nachteil erweist sich, daß die Beurteilung zweier Störarten auf Gleichheit umso schwieriger wird, je stärker sich diese voneinander unterscheiden. Vorteilhaft ist der geringe Aufwand bei der Datenerhebung, da nur ein Urteil je Objekt abzugeben ist. Außerdem liefert das Verfahren mit dem Rauschabstand einen objektiven Maßstab für die Bildqualität bezüglich der Richtung (besser - schlechter). Solange jedoch mit dem weißen Rauschen als Referenzstörart keine Einzelbildbeurteilung (mit der Qualitätsskala) durchgeführt worden ist, fehlt der Bezug zur "absoluten" subjektiven Qualitätsskala.

2.3.2 Beurteilung bezüglich der Unterschiedlichkeit bzw. Ähnlichkeit

Als Urteilskriterium wird die "Unterschiedlichkeit" verwendet, wobei davon ausgegangen wird, daß alle wesentlichen Aspekte zur Urteilsbildung herangezogen werden. Für die Datenerhebung kann

z.B. die folgende sechsstufige Skala verwendet werden:

Bewertung	1	2	3	4	5	6
Beschreibung der Unterschiedlichkeit	keine	geringe	deutliche	sehr deutliche	große	extrem große

Die Anzahl der Dimensionen des den Urteilen zugrunde liegenden Merkmalsraumes läßt sich durch ein mehrdimensionales Skalierungsverfahren ("INDSCAL" nach Carroll & Chang / 15 /, s.a. Bortz / 18 /) bestimmen. Die Art der Fragestellung nach der Unterschiedlichkeit (bzw. Ähnlichkeit) läßt keine direkte Interpretation der Dimensionen zu.

2.3.3 Beurteilung nach der Präferenz (Paarvergleich)

Die beiden unterschiedlich gestörten Fernsehbilder werden bezüglich der Präferenz (A besser B oder umgekehrt) beurteilt, wobei davon ausgegangen wird, daß alle wesentlichen Beurteilungsaspekte zur Urteilsbildung herangezogen werden. Die Auswertung erfolgt mit dem "Law of Comparative Judgment" (Guilford / 17 /, angewendet bei Ulene / 9 /). Es ergeben sich Einstufungen der Beurteilungsobjekte auf einer eindimensionalen Güteskala.

Im Gegensatz zu den "Absoluturteilen" aus der Einzelbildbeurteilung kann der Paarvergleich nur eine relative Zuordnung der Urteile zueinander erbringen. Die vorliegende Fragestellung nach der Bevorzugung läßt keine Aussagen über spezielle Eigenschaften der Fernsehbilder zu. Bei der Beurteilung von Bildpaaren (gleiche Bilder mit verschiedenen Störstufen) wird ein gestörtes Bild zum Testbild und das andere zum Referenzbild. Durch die Gegenwärtigkeit der Referenz sind die Paarvergleichsurteile gut reproduzierbar. Bei zu großen Reizunterschieden gibt das Urteil "A besser B" die große Reizdistanz nur ungenau wieder, so daß sich der klassische Paarvergleich nur für kleine bis mittlere Qualitätsunterschiede eignet. Da die Anzahl der zu beurteilenden verschieden gestörten Bildpaare bei Berücksichtigung aller möglichen Paarkombinationen je Testbild mit $n_s \cdot (n_s-1)/2$ (n_s = Anzahl der Störstufen) ansteigt, läßt sich der Paarvergleich nur bei einer kleinen Anzahl von Beurteilungsobjekten anwenden.

2.3.4 Beurteilung nach gestufter Präferenz (comparison method)

Diese Methode erlaubt nicht nur die Beurteilung, welches gestörte Bild bezüglich der Qualität besser ist (Präferenz), sondern gibt zusätzlich eine Information, um wieviele Stufen das Bild A besser oder schlechter als das Bild B ist. Die folgende Skala wird in der C.C.I.R. Rec. 500 / 12 / angegeben:

Bewertung	Bild A .....(als) Bild B
+ 3	viel besser
+ 2	besser
+ 1	etwas besser
0	gleich
- 1	etwas schlechter
- 2	schlechter
- 3	viel schlechter

Verwendet man, wie bei Hopf/Kiefer / 16 /, ein Referenzbild mit einer mittleren definierten Störung (durch einen frequenzmodulierten Tonträger gestörtes Schwarzweißbild der absoluten Gütestufe "ziemlich gut"), so kann man die Relativurteile direkt in Abhängigkeit von der variablen Störgröße darstellen. Diese Methode eignet sich für den Vergleich von verschiedenen Störarten. Ein Kategorie-Urteil ist hier ohne Anwendung der "Einzelbildbeurteilung" nicht möglich.

2.3.5 Verhältnisschätzung (ratio estimation)

Bei der Verhältnisschätzung wird davon ausgegangen, daß die Bpn in der Lage sind, das Ausmaß einer Eigenschaft eines Fernsehbildes (z.B. Farbsättigung, Qualität, Störwirkung) im Verhältnis zum Ausmaß derselben Eigenschaft eines Referenzbildes derselben Vorlage zu beurteilen.
Taylor et al. / 10 / haben diese Methode sehr erfolgreich für die Beurteilung von Fernsehbildern eingesetzt, die durch "Weißes Rauschen" und "Intermodulationsrauschen" verschieden stark gestört waren. Bei Taylor et al. wies das auf der linken Seite des Monitors gezeigte Bild einen mittleren Störspannungspegel auf (Referenzbild), während

das Bild auf der rechten Seite des Monitors in zufälliger Reihenfolge durch eine Anzahl von Störspannungspegeln gestört wurde. Die Bpn mußten zwei Urteile abgeben:

a) welches Bild wirkt stärker gestört (A oder B)?

b) wieviel mal mehr ist A (B) gegenüber B (A) gestört?

Hieraus ergeben sich Aussagen, die sich immer auf einen Referenzstörspannungspegel beziehen, wie z. B.:

> Ein mit einem Störspannungspegel von 35 dB gestörtes Bild (IM) wirkt doppelt so stark gestört, wie ein mit 39 dB gestörtes Bild.

Dieses Verfahren liefert zwar gut reproduzierbare Urteile, läßt aber keine an einer Kategorie-Skala orientierten Urteile ("Absoluturteile") zu.

Zusammenfassend kann zu den Bildpaarbeurteilungsverfahren gesagt werden, daß sie sich für den vorliegenden Anwendungsfall aus zwei Gründen nicht eignen:

a) Die Reizunterschiede können so groß gewählt werden, daß die Anwendung der z.T. aufwendigen Verfahren für kleine Reizunterschiede nicht erforderlich ist.

b) Man erhält nur Relativurteile, auch wenn man definiert gestörte Referenzbilder bei der Beurteilung verwendet.

Zum hier erforderlichen Güteurteil gelangt man nur über die Einzelbildbeurteilung mit Güteskala.

3 Auswertung der subjektiven und objektiven Daten

Nach der Datenerhebung liegt eine große Anzahl von Urteilen über gestörte Fernsehbilder vor, die von mehreren Bpn anhand von Kategorieskalen (s. Pkt 2.2) abgegeben worden sind. Die Aufgabe des Analyseverfahrens besteht nun darin, die große Anzahl von Einzelurteilen (Rohdaten) ohne erheblichen Informationsverlust auf wenige aussagekräftige Kennwerte zu reduzieren (Datenreduktion). Diese sind dann mit den entsprechenden physik. Meßwerten (Störspannungspegel, Definition s. Pkt. 4.5.1) zu verknüpfen. Ist der Zusammenhang zwischen den subjektiven und objektiven Daten für die interessierende Störart bekannt, so kann die Bildgüte ohne aufwendige (subjektive) Beurteilungsverfahren aus den physik. Meßwerten vorhergesagt werden. Die o. g. Forderungen erfüllt ein von Weaver / 19 / prinzipiell beschriebenes und später von Prosser et al. / 11 / erheblich erweitertes Verfahren, das unter den Punkten 3.1 und 3.2 näher besprochen wird.

Zuvor sei noch auf eine von Mertz et al. / 20 /, Cavanaugh und Lessman / 21 / u.a. benutzte Auswertungsmethode hingewiesen, die auch von den relativen Urteilshäufigkeiten je Beurteilungsobjekt (Summe über alle Skalenstufen = 1) ausgehen. Diese relativen Häufigkeiten werden getrennt nach den Objekten (Störspannungspegeln) über eine bestimmte Anzahl von Skalenstufen summiert und den Skalenstufengruppen 1, 2 oder besser, 3 oder besser bis 6 oder besser (bei einer siebenstufigen Skala) zugeordnet. Die zu jeder Skalenstufengruppe gehörenden relativen Summenhäufigkeiten werden dann durch je eine komplementäre, kumulative Normalverteilung nach der Methode der kleinsten Fehlerquadratsumme approximiert. Bei sieben Skalenstufen ergeben sich sechs kumulative Normalverteilungen, aus denen z.B. abzulesen ist, wieviel Prozent der Bpn bei welchem Störspannungspegel z.B. mit "Stufe 2 oder besser" (impairment just perceptible) geurteilt haben.

Nach Cavanaugh und Lessman sind zwei Skalenstufen von besonderer Bedeutung:

a) Stufe 2 (impairment just perceptible)
b) Stufe 4 (impairment to picture, but not objectionable)

Die Stufe 2 entspricht etwa der Wahrnehmbarkeitsschwelle und die "Stufe 4 oder besser" umfaßt die gestörten Fernsehbilder, die gerade noch nicht beanstandet werden.

Die Handhabung der beiden o. g. Auswertungsverfahren wird an einem unter Punkt 7 (Anhang) behandelten Beispiel demonstriert.

Das wie folgt erläuterte Auswertungsverfahren nach Prosser et al. beschreibt die anhand einer fünfstufigen Skala über gestörte Fernsehbilder gefällten Urteile durch ein geschlossenes Verteilungsmodell. Für jede Störart (z.B. Echostörungen) ergibt sich je Beurteilungsobjekt eine Binomialverteilung (s. Pkt. 3.2), die durch den Verteilungsmittelwert p (bei z.B.5 Skalenstufen)eindeutig bestimmt ist. Außerdem können die normierten Urteilsmittelwerte $\bar{p}$ je Beurteilungsobjekt über alle Bpn berechnet und in Abhängigkeit von den Störspannungspegeln dargestellt werden. Für Echostörungen, IM-Rauschen und weißes Rauschen ergibt sich in guter Näherung eine logistische Funktion (s.Pkt. 3.1 und 7) mit den Parametern M und S.

Wegen der übersichtlichen Darstellung der Ergebnisse und der Verfügbarkeit weniger aussagekräftiger Kennwerte wird dieses Verfahren in der Literatur häufig verwendet.

3.1 Verknüpfung der subjektiven Werte (Urteilsmittelwert) mit den objektiven Werten (Störspannungspegel)

Durch die Zuordnung der fünf Bewertungsziffern (Schulnoten) zu den Skalenstufen werden die Wahrnehmungen der Bpn quantitativ in Form von Urteilen erfaßbar gemacht. Der beste Schätzwert für die dem Einzelreiz (gestörtes Fernsehbild) entsprechende "wahre" Wahrnehmung ist der arithm. Mittelwert $\bar{p}$ über alle Bpn-Urteile. Dieser ist mit einer gewissen Unsicherheit behaftet, da die Bpn nur eine Stichprobe aus der großen Gruppe der Fernsehteilnehmer darstellt. Der Vertrauensbereich des Mittelwertes μ läßt sich bei z.B. 95 %iger

Sicherheit ($z = 1{,}96$) angenähert aus der Standardabweichung der Stichprobe ($s_{\bar{x}}$) und dem Stichprobenumfang ($n_{\bar{x}}$) abschätzen:

$$\mu = \bar{p} \pm 1{,}96 \cdot \frac{s_{\bar{x}}}{\sqrt{n_{\bar{x}}}} \qquad (1)$$

Gemäß Prosser et al. wird der normierte Urteilsmittelwert ($0 \leqq \bar{p} \leqq 1$) für jedes gestörte Testbild wie folgt berechnet:

$$\bar{p} = p_1 + 0{,}75\ p_2 + 0{,}50\ p_3 + 0{,}25\ p_4 \qquad (2)$$

wobei $p_{1\ldots5}$ = relativen Urteilshäufigkeiten je Skalenstufe s

Die Summe der relativen Häufigkeiten ist je Beurteilungsobjekt = 1. Die Bewertungsfaktoren ergeben sich aus $(5 - s) / 4$.

Bildgüte	sehr gut	gut	mittel-mäßig	schlecht	sehr schlecht
Skalenstufe s	1	2	3	4	5
Bewertung	1,00	0,75	0,50	0,25	0,00

Die über die Bp^n gemäß (2) berechneten Urteilsmittelwerte können zwecks weitergehender Datenreduktion zusätzlich über die Testbilder erstreckt werden, wenn mit Hilfe der Varianzanalyse nachgewiesen werden kann, daß die Bildinhalte die Urteile nicht signifikant beeinflussen bzw., daß die Varianz der Urteile fast ausschließlich durch die unterschiedlichen Störspannungspegel hervorgerufen wird.

Der nächste Schritt bei der Datenanalyse führt zur Aufdeckung des Zusammenhanges der je Bild und Störstufe (Beurteilungsobjekt) abgegebenen Urteilsmittelwerte $\bar{p}$ mit den entsprechenden Störspannungsmeßwerten in dB. Dieser Zusammenhang kann für einige Störarten (z.B.

Rauschen, Echostörungen) durch die logistische Funktion bzw. die Gaußfunktion gut angenähert werden. Wegen der einfacheren mathematischen Form wird die logistische Funktion bevorzugt verwendet. Sie ist durch die Parameter M und S vollständig bestimmt.

$$\bar{p}\ (x) = \left[1 + exp\left(\frac{M - x}{S}\right)\right]^{-1} \qquad (3)$$

x = Störspannungspegel / dB
M = der dem mittleren Urteil (Skalenstufe 3) zugeordnete Störspannungspegel (median observer)
S = Steigung der Funktion im Wendepunkt (slope index)

Es hat sich gezeigt, daß selbst bei störungsfreien Fernsehbildern nur ein Teil der Bpn das Urteil "sehr gut" (Stufe 1) erteilen. Die angenäherte logistische Funktion läuft also nicht asymptotisch gegen 1, sondern gegen Werte zwischen 0,8 bis 0,9. Zu den Ursachen dieses Effektes kann man die mangelnde technische Perfektion des Fernsehbildes (Zeilenstruktur, Flimmern), die Qualität der Testbilder und als unangenehm empfundene Bildinhalte zählen. Diese Reststörung (residual impairment) wird in der logistischen Funktion durch den Term 1/H berücksichtigt:

$$\bar{p}\ (x) = \left[\frac{1}{H} + exp\left(\frac{M - x}{S}\right)\right]^{-1} \qquad (4)$$

H entspricht mit guter Genauigkeit dem Urteilsmittelwert $\bar{p}$ für das beste, störungsfreie Fernsehbild (s. Flor / 22 /). Die verbleibenden Parameter M und S lassen sich mit der Methode der kleinsten Fehlerquadratsumme aus dem folgenden Ansatz ermitteln:

$$\sum_i \left\{\bar{p}_i - \left[\frac{1}{H} + exp\left(\frac{M - x_i}{S}\right)\right]^{-1}\right\}^2 = \text{Minimum} \qquad (5)$$

wobei x_i = Störspannungspegel in dB

$\bar{p}_i$ = zugehörige Urteilsmittelwerte

Die Güte der Anpassung kann mit dem Chi-Quadrat-Test überprüft werden.
Die Arbeiten einiger Autoren lassen erkennen, daß die logistische Funktion zur Beschreibung des Zusammenhangs zwischen der Bildqualität und den Störspannungspegeln bei den in dieser Arbeit interessierenden Störarten Rauschen (/ 19 /, / 25 /, / 26 /, / 27 /), Inter- und Kreuzmodulation (/ 22 /, / 23 /, / 24 /) verwendet werden kann.

Häufig treten zwei oder mehr verschiedene Störarten gleichzeitig auf. Sind diese hinreichend voneinander unabhängig (vgl. / 31 /), so läßt sich die durch die resultierende Störung hervorgerufene Bildqualität $\bar{p}$ anhand einer von Prosser et al. / 11 / gefundenen empirischen Formel aus den zu bestimmten Störspannungspegeln gehörenden Urteilsmittelwerten $\bar{p}_r$ der Einzelstörungen berechnen:

$$\frac{1-\bar{p}}{\bar{p}} = \sum_{r=o}^{n} \left(\frac{1-\bar{p}_r}{\bar{p}_r} \right) \quad (6) \quad \text{bzw.} \quad I = \sum_{r=o}^{n} I_r \qquad (7)$$

r = Laufindex, n = Anzahl der Störarten, $\bar{p}_o$ = Urteilsmittelwert der Reststörung, I = impairment-unit in imps.

Beispiel: Weist die Störart A bei einem Störspannungspegel von 32 dB den Urteilsmittelwert $\bar{p}_1 = 0{,}68$ (fast gut) und die Störart B bei 23 dB den Wert $\bar{p}_2 = 0{,}45$ (mittelmäßig) auf, so erhält man gemäß Gleichung (6) und (7):

$$\underline{I} = I_1 + I_2 = 0{,}47 + 1{,}22 = \underline{1{,}69 \text{ imps}}$$

Die aus den Störarten A und B zusammengesetzte Gesamtstörung führt also zu einem Urteilsmittelwert von $\bar{p} = 0{,}37$ (Bildqualität mittelmäßig bis schlecht).

Die Imp-Skala reicht von I = 0 für ein perfektes Bild ($\bar{p} = 1$) über I = 1 für ein mittelmäßiges Bild ($\bar{p} = 0{,}5$) bis I = ∞ für ein sehr schlechtes Bild ($\bar{p} = 0$). Prosser et al. verwenden als Bildgüte-Kennzahlen 1/8, 1/4 (hoher Güte-Standard), 1/2 und 1 Imp (= mittleres Urteil, niedriger Güte-Standard).

Die Gültigkeit dieses Gesetzes der "Summierbarkeit subjektiver Störeinheiten (subjective impairment units)" ist für die Störungskombinationen "Weißes Rauschen" - "2 µs-Echostörungen" (in NTSC-Schwarzweißbildern, s. / 11 /) und für geformtes Rauschen im Luminanz- und Chrominanz-Kanal (s. / 26 /) geprüft worden. Die Anwendbarkeit auf IM und XM läßt sich durch geeignete subjektive Tests kontrollieren.

3.2 Beschreibung des subjektiven Datensatzes durch Urteilsverteilungen

Weaver / 19 / hat mit breitbandigem Rauschen gestörte Fernsehbilder durch eine homogene Bpn-Gruppe anhand einer 6-stufigen Störungsskala beurteilen lassen. Es zeigte sich, daß die (relativen) Urteilshäufigkeiten in den Skalenstufen je Beurteilungsobjekt durch eine Binomialverteilung mit dem Parameter p (Verteilungsmittelwert) gut angenähert werden konnten.

Da gemäß Pkt. 2.2 eine fünfstufige Skala vorgesehen ist, beziehen sich die folgenden Erläuterungen auf diese Stufenanzahl.

Für die Binomialverteilung gilt:

$$P(x) = \binom{4}{x} p^{x} \left(1 - p\right)^{4 - x} \qquad (9)$$

$P(x)$ = Wahrscheinlichkeit für das Auftreten von Urteilen in der Skalenstufe s (1 bis 5)

x = Zufallsvariable ($x = 5 - s$)

Für jedes gestörte Fernsehbild ergibt sich eine diskrete Verteilung mit dem Parameter p in Abhängigkeit von der Zufallsvariablen x bzw. der Skalenstufe s.

Die gemäß (9) berechneten Wahrscheinlichkeiten lassen sich durch die Summanden des binomischen Lehrsatzes für $(p+q)^{s-1}$ bei $s = 5$ Skalenstufen ersetzen:

$$
\begin{aligned}
p_1 &= p^4 \\
p_2 &= 4p^3\,(1-p) \\
p_3 &= 6p^2\,(1-p)^2 \\
p_4 &= 4p\,(1-p)^3 \\
p_5 &= (1-p)^4
\end{aligned}
\qquad (10)
$$

Setzt man diese theoretischen relativen Häufigkeiten in Gleichung (2) ein, so ergibt sich, daß $p = \bar{p}$ ist. Der Verteilungsmittelwert p kann also durch den normierten Urteilsmittelwert $\bar{p}$ ersetzt werden.

An dem folgenden theoretischen Beispiel läßt sich zeigen, daß auf den Urteilsmittelwert nur 42 % der Urteile entfallen, wenn die Bpn im Mittel mit $\bar{p} = 0{,}73$ urteilen.

Bildgüte	sehr gut	gut	mittel-mäßig	schlecht	sehr schlecht
Skalenstufe s	1	2	3	4	5
relat. Häufigkeit p_s	0,286	0,420	0,232	0,057	0,005

Für die grobe Einstufung der gestörten Fernsehbilder eignet sich eine zweistufige Skala mit den Benennungen "annehmbar" und "unannehmbar". Auf das Beispiel bezogen entfallen selbst bei einem unter "gut" auf der fünfstufigen Skala eingestuften Fernsehbild noch 12,5 % der Urteile in die Kategorie "unannehmbar". Ein hohes

Qualitätsniveau kann sichergestellt werden, wenn höchstens 5 % der Bpn-Urteile in der Stufe "unannehmbar" zugelassen werden (5 %-Punkt). Dem entspricht ein $\bar{p}$ von 0,811 (besser als "gut"). Der zugeordnete Störspannungspegel folgt aus der zutreffenden logistischen Funktion.

Weaver / 19 / hatte für eine homogene Bpn-Gruppe Urteilsverteilungen ermittelt, die gut durch eine Binomialverteilung erster Ordnung von Typ $(p + q)^{s-1}$ bei $s = 6$ angenähert werden konnten. Abweichend davon haben Prosser et al. / 11 / aufgedeckt, daß die Urteilsverteilungen von Nichtexperten besser durch eine Binominalverteilung zweiter Ordnung vom Typ $(p + q)^{2s-1}$ bei $s = 5$ approximiert werden können. Hierbei werden die zehn Terme paarweise zu fünf Gruppen zusammengefaßt, so daß sich die relativen Häufigkeiten der Urteile je Skalenstufe wie folgt errechnen lassen:

$$
\begin{aligned}
p_1 &= p^9 + 9\,p^8\,(1-p) \\
p_2 &= 36\,p^7\,(1-p)^2 + 84\,p^6\,(1-p)^3 \\
p_3 &= 126\,p^5\,(1-p)^4 + 126\,p^4\,(1-p)^5 \\
p_4 &= 84\,p^3\,(1-p)^6 + 36\,p^2\,(1-p)^7 \\
p_5 &= 9\,p\,(1-p)^8 + (1-p)^9
\end{aligned}
\qquad (11)
$$

Wird (11) in (2) eingesetzt, so folgt daraus, daß der Verteilungsmittelwert p nur für $p = 0$, 0,5 und 1 mit dem normierten Urteilsmittelwert $\bar{p}$ übereinstimmt. Es gilt:

$$
\bar{p} = p^8\,(9-8p) + p^6\,(1-p)^2\,(63-36\,p) + 63\,p^4\,(1-p)^4 + p^2\,(1-p)^6\,(9-12\,p) \qquad (12)
$$

Die Approximation der Urteilsverteilungen durch eine Binomialverteilung zweiter Ordnung liefert je Beurteilungsobjekt einen

p-Wert, woraus die theoretischen relativen Häufigkeiten der Urteile $p_{1...5}$ gemäß (11) und $\bar{p}$ gemäß (12) berechnet werden können. Neben dem "mittleren Urteil" ($\bar{p}$ = 0,5) ist der o.g. "5 %-Punkt" (wobei $\bar{p}$ = 0,78) ein wichtiges Kriterium für die Festlegung einer Standardbildqualität. In diesem Fall urteilen 83 % der Bpn mit "sehr gut" bzw. "gut" und nur 1 % mit "schlecht" bzw. "sehr schlecht". Aufgrund der in dieser Arbeit als Bpn vorgesehenen Nichtexperten dürfte eine Anwendung der Binomialverteilung zweiter Ordnung für die Beschreibung der Urteilsverteilungen je Beurteilungsobjekt geeignet sein.

4 Planung des Bildbeurteilungsverfahrens

Zum Zeitpunkt der Berichterstattung war abzusehen, daß sich die praktische Bildgütebeurteilung aus zeitlichen und gerätetechnischen Gründen nicht mehr durchführen ließ. Daher beschränkt sich dieser Bericht auf die Ergebnisse der Planungsphase.

4.1 Übersicht

Gemäß Aufgabenstellung soll festgestellt werden, um wieviele Bildgütestufen die Bildqualität bei Anwendung des Verfahrens zur Reduzierung von IM- und X.1-Störungen durch phasenverkoppelte Bildträger in Vielkanal-Kabelfernsehsystemen (Krick / 2 /)verbessert wird und in welchem Maße sich die verschiedenen Störarten (IM, XM, Rauschen) auf die Bildgütebeurteilung auswirken.

Beide Ziele können vom Aufwand und von der Art und Sicherheit der Aussagen her durch die Einzelbildbeurteilung mit einer 5-stufigen Bildgüteskala (quality-scale) gemäß Pkt. 2.1 erreicht werden. Alternativ kann die ebenfalls 5-stufige Bildstörungsskala (impairment-scale) oder beide Skalen verwendet werden (s. Anlage 1).

Im Hauptversuch haben 15 Bpn (3 Gruppen à 5 Bpn) fünf von Störungen überlagerte Farbdias (s. Pkt. 4.3) unter gleichen Beobachtungsbedingungen (s. Pkt. 4.4) anhand der o. g. Skalen zu beurteilen.

Jedes Dia wird kombiniert mit sechs Störstufen zur Vermeidung von Zieheffekten in zufälliger Reihenfolge für jede Störart getrennt (IM, XM, IM + XM, Rauschen) auf dem Farbmonitor gezeigt. Die Störstufen sind so zu wählen, daß sie Urteile von "sehr gut" bis "sehr schlecht" ermöglichen. Wegen umfangreicher Einstellarbeiten beim Wechsel der Störart wird jede Störart für sich mit 5 x 6 = 30 Bild-Störstufen-Kombinationen (Beurteilungsobjekte) angeboten. Es ist darauf zu achten, daß nicht zwei gleiche Dias aufeinander folgen. Aus Zeitgründen sind die Urteile zu beiden Skalen (s. Pkt 2.1) aufgrund einer Bildvorstellung nacheinander abzugeben. Die Bpn werden instruiert, beide Urteile völlig unabhängig voneinander zu fällen.Die Anzahl der Urteile beträgt 30 Beurteilungsobjekte x 4 Störarten x 15 Bpn = 1800 Urteile je Skala.

Der Zeitplan ergibt sich aus den folgenden Hinweisen.

Das Einzelbild wird 10 Sekunden zwecks Beurteilung gezeigt. In den folgenden 10 Sekunden wird der Störspannungspegel gemessen, während den Bpn ein fehlerfreies Graubild (Pausenbild) gezeigt wird. Für 30 gestörte Bilder werden je Störart 10 Minuten benötigt. Es folgen 5 Minuten Pause, die z. T. für die Einstellung der nächsten Störart verwendet werden. Bei vier Störarten ergibt sich demnach der folgende Zeitplan für eine Bpn-Gruppe:

Test / Pause (alternierend):
10 / 5 / 10 / 5 / 10 / 5 / 10 Minuten

Zu den 55 Minuten kommt noch eine vor dem Haupttest durchzuführende Lernperiode von ca. 15 Minuten Dauer, so daß sich für jede Bpn-Gruppe eine Testdauer von 70 Minuten ergibt. Daraus folgt die Gesamtdauer des Bildbeurteilungsverfahrens für alle drei Gruppen (= 15 Bpn) zu 210 Minuten Testdauer plus ca. 20 Minuten Zwischenpausen = 230 Minuten. Für die Vorauswahl der Testbilder (s. Pkt. 4.3) werden fünf Bpn für weitere 75 Minuten benötigt.

Wegen der relativ geringen Anzahl von Bildvorstellungen und wegen der Verschlechterung der Bildqualität durch ein Videoaufzeichnungsgerät wird auf eine Vorproduktion der Bildfolgen auf Videoband ver-

zichtet. Die deshalb manuell vorzunehmenden Geräteeinstellungen müssen mit großer Sorgfalt durchgeführt werden.

Der Beobachtungsplatz wird nach Mankopf / 28 / in Anlehnung an den Platz von Bühler / 29 / aufgebaut. Im Abstand von 6 x Bildschirmhöhe finden 2 Bpn in der ersten Reihe und 3 Bpn in der zweiten Reihe auf erhöhten Podesten angeordneten Sitzen Platz (s. Bild 1). Die Platznummern sind für die Erfassung eines eventuell vorhandenen Platzeinflusses auf die Urteile (Varianzanalyse) auf dem Bildbeurteilungsbogen (Anlage 1) zu notieren.

Der für die Bildbeurteilung erforderliche apparative Aufbau (Messung der Störspannungspegel, Erzeugung der Störarten) wird unter Punkt 4.5 beschrieben.

4.2 Auswahl und Instruktion der Bpn

In unserem Fall sind Experten nicht als Bpn einsetzbar, da sie zu kritisch urteilen und deshalb ein zu hohes Bildqualitätsniveau festlegen würden. Dieses sollte durch eine repräsentative Gruppe von Bpn ermittelt werden, die sich aus Fernsehteilnehmern zusammensetzt.
Allnatt und Corbett zeigen in einer Arbeit / 14 /, daß diese in der Bildgütebeurteilung ungeübte Beobachter (Nichtexperten) nur dann stabile, reproduzierbare Urteile abgeben, wenn sie vor dem Test ein gewisses Maß an Übung erreicht haben (Lernperiode). Die C.C.I.R. Rec. 500 / 12 / empfiehlt, mindestens 10 bis 20 angelernte, normalsichtige Nichtexperten für das Beurteilungsverfahren einzusetzen. Als Kompromiß bietet sich der Einsatz von 15 Bpn an. Diese werden vor dem Test anhand von später nicht mehr verwendeten Diapositiven mit dem Beurteilungsprozeß (Urteilsfindung unter Verwendung der Skalen) vertraut gemacht. Um außerdem eine Angleichung der Urteilsbereiche der Bpn an das möglichst breite Reizfeld (gestörte Dias) zu erreichen (Standardisierung), werden nicht gestörte (sehr gute) und sehr stark gestörte (sehr schlechte) Dias vorgeführt.

Es ist ausdrücklich darauf hinzuweisen, daß nicht der Bildinhalt (= natürliches Bild als Träger der Störung), sondern die durch die Störung beeinträchtigte Bildqualität (Qualitätsskala) bzw. die Auswirkung der speziellen Störart (Störungsskala) zu beurteilen sind.

Nach Schneeberger / 13 / ermüden die Testbeobachter nach spätestens zwei Stunden, so daß diese Zeit incl. der Pausen nicht überschritten werden sollte (hier: ca. 70 Minuten Testdauer). Um einerseits den Zeitaufwand für das Beurteilungsverfahren gering zu halten und andererseits die Beobachtungsbedingungen (s. Pkt. 4.3) einzuhalten, wurden Gruppen zu je fünf Bpn gebildet, die gleichzeitig die Bildgütebeurteilung durchführen. Eine gegenseitige Beeinflussung sollte vermieden werden.

4.3 Auswahl der Testbilder

Die Testbilder dienen als Träger der im Übertragungsweg entstehenden Störungen.

Da die Qualität von normalen Fernsehbildern zu beurteilen ist, sollten für den Test solche Bildinhalte ausgesucht werden, die auch im Fernsehprogramm vorkommen (natürliche Bilder). Nach Taylor et al. / 10 / fällt die Wahl auf Standbilder (hier Farbdias), da diese gegenüber Bewegtbildern die Erkennbarkeit von Störungen verbessern. Die Bilder sollten gemäß C.C.I.R.Rec. 500 / 12 / auf die verschiedenen Störarten empfindlicher reagieren als durchschnittliche Bilder. Um fünf solche Farbdias für den Haupttest bereitstellen zu können, ist in einem Vortest eine Vorauswahl zu treffen. Wegen der Reproduzierbarkeit der Ergebnisse und der Vergleichbarkeit mit Testergebnissen anderer Autoren werden 15 Farbdias aus Standardserien (SMPTE, EBU, IRT) ausgewählt. Jedes Farbdia wird mit drei verschiedenen Störstufen (bei insgesamt vier Störarten) zur Beurteilung gezeigt (= 45 Beurteilungsobjekte je Störart). Für den Vortest genügt eine Gruppe von fünf Bpn. Es ergeben sich also 900 Urteile und ein Zeitbedarf von 75 Minuten. Es kommt das Einzelbildbeurteilungsverfahren zur Anwendung.

4.4 Festlegung der Beobachtungsbedingungen

Die Bildbeurteilung sollte möglichst wenig durch die Beobachtungsbedingungen beeinflußt werden. Die vorhandenen Einflüsse sollten wegen der Vergleichbarkeit verschiedener Arbeitsergebnisse bei allen Autoren gleich sein. Die C.C.I.R. Rec. 500 / 12 / schlägt zwecks Standardisierung u.a. Beobachtungsbedingungen vor, die diesem Test zugrunde gelegt werden sollen:

a) Beobachtungsabstand = 6 x Bildschirmhöhe.

b) Die Spitzenleuchtdichte des Schirms soll 220 $\pm$ 30 asb (70 $\pm$ 10 cd/m^2) erreichen.

c) Das Verhältnis der Leuchtdichte des dunkelgetasteten Bildschirms zur Spitzenleuchtdichte soll $\leq$ 0,02 sein.

d) Das Verhältnis der Leuchtdichte von Bildschwarz zu Bildweiß soll bei vollständig abgedunkeltem Raum ungefähr 0,01 betragen.

e) Die Umfeldleuchtdichte hinter dem Bildschirm soll bei 10 % der Spitzenleuchtdichte liegen.

f) Der Testraum soll schwach beleuchtet und weiß sein.

g) Das Verhältnis des Umfeldraumwinkels zum Bildschirmraumwinkel soll $\geq$ 9 betragen.

4.4.1 Beobachtungsplatz

Die o. g. Bedingungen werden durch einen von Bühler / 29 / beschriebenen und später durch Mankopf / 28 / modifizierten Beobachtungsplatz eingehalten. Der Beobachtungsplatz nach Bühler ist gekennzeichnet durch ein parabelförmiges indirekt beleuchtetes Umfeld mit zum seitlichen Rand hin abnehmender Leuchtdichte. Damit wird für den Beobachter ein quasiunendliches Umfeld konstanter Leuchtdichte hinreichend gut simuliert. Der Monitor steht im Parabelbogen.

Da der Aufwand beim Aufbau dieses Beobachtungsplatzes groß ist, hat Mankopf die Konzeption einer frontal beleuchteten, diffus reflek-

tierenden Umgebung des Bildschirms wieder aufgegriffen und einen Beobachtungsplatz mit rechteckigem Umfeld realisiert, das durch einen Overhead-Projektor gleichmäßig ausgeleuchtet wird. Die Aufhellung des Bildschirms wird durch eine entsprechend gestaltete Folie verhindert.

4.5 Apparativer Aufbau

Die hier für die Bildbeurteilung zu simulierenden Störarten sollten den in der Praxis vorkommenden die Bildgüte beeinträchtigenden Störungen möglichst weitgehend entsprechen. Diese Forderung kann mit Hilfe einer im Hause vorhandenen fast ausschließlich aus kommerziellen KTV-Komponenten bestehenden Versuchsanlage für 30 TV-Kanäle (s. Bild 1) erfüllt werden.

Im Prinzip werden die Videosignale (Nutzbild, Störbild) mit Modulatoren durch Restseitenband-Amplitudenmodulation in das ZF-Frequenzband (38,9 MHz) gebracht (Negativ-Modulation), um anschließend mit 30 Frequenzumsetzern in das HF-Frequenzband(bis 300 MHz)transponiert zu werden. Hinsichtlich der Bildträgeranordnung interessiert hier nur das reguläre Raster mit den Bildträgerfrequenzen:

$$f_{BT} = 0{,}25 + i \cdot f_o \quad \text{in MHz}$$

wobei $i = 1...n$, $f_o = 7$ MHz sind, n = Anzahl der Bildträger.

Der Umsetzer erlaubt u.a. die Betriebsarten mit freilaufenden (einzeln quarzstabilisierten) oder phasenverkoppelten Bildträgern. Bei der ersten Betriebsart entstehen Bildstörungen durch IM- und XM-Produkte, bei der zweiten nur durch XM-Produkte, deren Größe stark von den Phasenbeziehungen zwischen den Bildträgern abhängig ist. Die Beschreibung eines Minimierungsverfahrens für die Phasen findet sich bei Krick / 2 /.

Als Übertragungsstrecke dienen 16 kaskadierte breitbandige Verstärker mit zwischengeschalteten Dämpfungsgliedern bzw. Koaxkabeln (8 x 70 m). Alle Kanäle weisen den gleichen Signalpegel auf. Auf

der Empfangsseite wird der gewünschte Nutzkanal mit dem Testbild zuerst in das ZF-Frequenzband und dann durch einen Nyquist-Demodulator in das Video-Frequenzband umgesetzt und auf einem professionellen Farbmonitor (Barco) dargestellt.

4.5.1 Messung des Störspannungspegels

Die Messung des Störspannungspegels erfolgt nach der Darbietung des Testbildes auf einem Graubild (50 % vom Weißpegel) im VF-Band mit einem Video-Störspannungsmesser mit Video-Tiefpaß (s. Anlage 2). Das Graubild entspricht einem mittleren simulierten Schwarzweiß-Testbild.

Gemessen wird der echte Effektivwert der Störspannung im Bildbereich des Videosignals. Die während der Austastung der Synchron- und der Austastimpulse vom Meßgerät nicht erfaßte Störenergie wird bei der Anzeige berücksichtigt. Diese ist in dB geeicht. Die Bezugsspannung beträgt 0,700 V.

$$\text{Anzeige} = 20 \log \left(\frac{U_{eff}}{0{,}700} \right) \text{ in dB.}$$

Ein Rauschbewertungsfilter nach C.C.I.R.Rec. 421-2 erlaubt die Berücksichtigung der Augenempfindlichkeit gegenüber Fremdspannungen. Es dient der Unterdrückung der hochfrequenten Anteile des weißen Rauschens bei der Messung der relativ niederfrequenten IM- bzw. XM-Störspannungen.

Nachteilig ist, daß die Störspannungspegel-Messung zeitlich nach der Beurteilung des gestörten Testbildes erfolgen muß. Mit der Prüfzeilenmeßtechnik wäre es im Prinzip möglich, die Störspannungspegel während der Vorführung des gestörten Testbildes zu messen. Das im Haus verfügbare Prüfzeilenmeßgerät (UPF, Rohde & Schwarz) läßt leider nicht die Messung der tieffrequenten IM und XM zu, da es einen Hochpaß mit einer Grenzfrequenz von 200 kHz enthält.

4.5.2 Erzeugung der Störarten, Wahl der Störstufenanzahl

Die Störarten IM und XM entstehen durch nichtlineare Verzerrungen in den Breitbandverstärkern (HF-Frequenzband).

Es handelt sich hauptsächlich um Verzerrungen dritter Ordnung, deren Störspannungspegel mit der Aussteuerung der Streckenverstärker nach dem 2 : 1-Gesetz ansteigt. Die IM-Störspannungen sind Verzerrungsprodukte der Bildträger. Sie fallen bei äquidistanter Trägeranordnung größtenteils in die Nähe der Bildträger und wirken nach der Demodulation wie ein niederfrequentes Rauschen im Fernsehbild des Nutzkanals. Für die Übertragung des Nutzbildes (Farbdia) wird der oberste Kanal S 20 ausgewählt, da in diesem der Störspannungspegel (IM + XM + Rauschen) maximal wird. Die übrigen 29 Kanäle dienen als Störkanäle. Die Bildinhalte dieser Kanäle verursachen die XM-Produkte. Wegen der Reproduzierbarkeit der Ergebnisse und um gleichzeitig eine "worst-case"-Bedingung zu erreichen, werden die 29 Kanäle synchron mit einem Schwarz-Weiß-Rechteck-Signal (f = 109,375 kHz) moduliert (m = 100 %). Die so simulierte XM zeigt sich im Nutzbild (Kanal S 20) durch ein Muster senkrecht stehender Balken mit etwas erhöhter Helligkeit. Diese Leuchtdichte-Vergrößerung entsteht während der Schwarzintervalle des Störbildes infolge eines i.A. negativen Koeffizienten dritter Ordnung für die Taylorreihe der Verstärkerkennlinie.

Innerhalb jeder Störart werden sechs Störstufen vorgesehen, da für die Darstellung des Zusammenhanges zwischen den Bildgüteurteilen und den Störspannungspegeln mindestens sechs etwa äquidistante Werte erforderlich sind (logistische Funktion). Diese sollten den subjektiven Urteilsbereich von "sehr gut" bis "sehr schlecht" abdecken. Erzeugt werden die Störstufen durch Über- (IM + XM) bzw. Untersteuerung (Rauschen) der Breitbandverstärker mit Hilfe einer vorgeschalteten regelbaren Eichleitung.

In der folgenden Tabelle sind die Störarten, deren Erzeugung und Messung zusammengestellt worden:

Störart	Betriebsart des Umsetzers	Bildinhalt der 29 Störkanäle	Bildinhalt des Nutzkanals bei der Messung
Rauschen IM	freilaufend	Schwarzpegel (m = 100 %)	
IM + XM			50 % vom Weißpegel
XM	phasenverkoppelt	Schwarz-Weiß-Rechteck-Signal (m = 100 %)	(Graubild)

Über dieses Programm hinaus könnte es noch von Interesse sein, den Einfluß der Tonträger auf die Bildqualität zu untersuchen. Das gleiche gilt auch für die der Praxis besser entsprechende asynchrone Kreuzmodulation.

5 Diskussion von Ergebnissen aus der Literatur

In den folgenden zwei Abschnitten werden die Ergebnisse aus zwei Arbeiten zitiert und diskutiert, die sich mit der Beurteilung der Störwirkung von Störarten (anhand von Kategorieskalen) auseinandergesetzt haben, wie sie auch in der vorliegenden Aufgabenstellung vorgesehen waren.

Bei Taylor et al. / 10 / hatten 700 Bpn anhand einer fünfstufigen Störungs-Skala mehrere durch die Störarten weißes Rauschen und IM-Rauschen gestörte Farbdias (Küchen-Szene, Susan u.a.) unter Beobachtungsbedingungen zu beurteilen, die mit dem Vorschlag der C.C.I.R.Rec.500 /12/ vergleichbar sind. Die Auswertung erfolgte u.a. mit dem bei Cavanaugh und Lessmann / 21 / beschriebenen Verfahren.

Groenveld und van Dijk / 24 / berichten über die Beurteilung von Schwarzweiß- und Farb-Fernsehbildern anhand einer fünfstufigen Güte-Skala. Die Fernsehbilder wurden durch Kreuzmodulation in verschiedenen Erscheinungsformen gestört. Die Auswertung erfolgte nach Prosser et al. / 11 /.

Die beiden in den o. g. Arbeiten verwendeten Skalen entsprechen den in dieser Arbeit angegebenen Skalen.

5.1 Weißes Rauschen und IM-Rauschen

Das weiße Rauschen (flat noise) wurde bei Taylor et al. dem Testbild in der Videolage zugesetzt. Für eine praxisgemäße Simulation des in der HF-Lage entstehenden Rauschens mußte ein Spezialfilter eingesetzt werden, das dem Einfluß des Nyquist-Filters im Demodulator Rechnung trägt. Außerdem ist noch ein Rauschbewertungsfilter nach der C.C.I.R.Rec.421, System M, zur Berücksichtigung der Augenempfindlichkeit gegenüber Rauschstörungen im Fernsehbild verwendet worden. Um diese auf amerikanische Fernsehnorm (System M) abgestimmten Untersuchungen auf europäische Verhältnisse übertragen zu können, sind die folgenden Umrechnungen erforderlich:

a) infolge der unterschiedlichen Rauschbewertungsfilter und

b) wegen der unterschiedlichen äquivalenten Rauschbandbreiten (USA: 3,8 MHz, Europa: 4,4 MHz)

ergibt sich ein Korrekturwert von + 1,2 dB, der zu den bei Taylor et al. angegebenen Störabständen zu addieren ist.

Den Ergebnissen von Taylor et al. wurden drei bewertete Videostörspannungspegel für "Weißes Rauschen" entnommen und auf unsere Verhältnisse umgerechnet:

1) - 35 dB für den "mittleren Beobachter", wobei 50 % der Bpn das betreffende gestörte Fernsehbild mit dem o.g. Störspannungspegel besser als Stufe 3 und 50 % schlechter als Stufe 3 bewerten (geringe Bildgüte).

2) - 41 dB für das mittlere "Stufe 2-Urteil" (can see some interference, but it's not annoying"), wobei 50 % der Bpn das betreffende gestörte Fernsehbild besser als Stufe 2 und 50 % schlechter als Stufe 2 beurteilen.

3) - 42 dB für den "5 %-Punkt", wobei 5 % der Bpn das betreffende gestörte Fernsehbild in der Kategorie "unsatisfactory pictures" einstufen (hohe Bildgüte).

Das IM-Rauschen (composite triple beat)wurde,wie in unserem Fall geplant, durch 28 unmodulierte quarzstabilisierte Bildträger erzeugt, die über hoch ausgesteuerte Breitbandstreckenverstärker übertragen wurden. Der Bildträger und die IM-Produkte desKanals 8 wurden demoduliert und in der Videolage dem Testbild additiv hinzugefügt. Das derart simulierte IM-Rauschen enthielt kaum Anteile über 100 kHz. Gemessen wurde der Effektivwert der Störspannung mit der Bezugsspannung von 0,714 Volt (hier: 0,700 Volt). Die Störspannungspegel können direkt von Taylor et al. übernommen werden. Analog zum Rauschen ergeben sich für IM-Rauschen die folgenden interpolierten Werte:

1) ca. - 40 dB für den "mittleren Beobachter"
2) ca. - 46 dB für das mittlere "Stufe 2-Urteil"
3) ca. - 47 dB für den "5 %-Punkt".

Das schmalbandige IM-Rauschen erfordert also für dieselbe Bildgüte im Verhältnis zum breitbandigen (weißen) Rauschen um ca. 5 dB höhere Störabstände. Die "IM" beeinträchtigt also die Bildgüte stärker als das "Weiße Rauschen".

Neben dem Beurteilungsverfahren mit Störungsskala wurde von Taylor et al. noch eine Verhältnisschätzung zwischen verschieden stark gestörten Testbildern durchgeführt. Jedes gestörte Bild wurde einem Referenzbild mit 39 dB Störabstand (für Rauschen und IM) gegenübergestellt.

Es zeigte sich, daß der Störeindruck verdoppelt wurde, wenn der Störabstand bei "Weißem Rauschen" um 4,5 dB und bei "IM-Rauschen" um 4,0 dB abnahm. Die Störabstände, bei denen kaum noch eine Bildverbesserung trotz steigenden Störabstandes eintrat, lag beim "Weißen Rauschen" bei umgerechnet ca. 40 dB und beim "IM-Rauschen" bei ca. 45 dB. In beiden Fällen urteilen 79 % der Bpn mit "Stufe 2 oder besser".

5.2 Kreuzmodulation

In Vielkanal-Kabelfernsehsystemen erzeugen die Synchronimpulse wegen der Negativmodulation maximale Störspannungen (XM), die als vertikale stehende oder bewegte hellere Balken im Fernsehbild in Erscheinung treten. Es ergeben sich stehende unterschiedlich über das Fernsehbild verteilte Balken, wenn die Zeilenfrequenzen der störenden Kanäle und des gestörten Kanals genau übereinstimmen (synchrone XM).

Groenveld und van Dijk / 24 / haben verschiedene Muster stehender Balken bezüglich ihrer Beeinflussung der Bildqualität untersucht.

Der Störspannungspegel wurde in Anlehnung an die NCTA-Definition / 30 / gemessen, wonach der Spitze-Spitze-Wert der Störspannung in der HF-Lage gemessen und auf den Spitzenwert des unmodulierten Bildträgers bezogen wird. Da es sich hierbei um kein normgerechtes Fernsehsignal handelt, erscheint eine Umrechnung neben anderen Gründen problematisch. Im folgenden werden deshalb die Originaldaten aus der o.g. Arbeit zitiert.

Aus der Beurteilung der Störwirkung eines einzelnen Balkens in Schwarzweiß- und Farbfernsehbildern ergab sich, daß die Störabstände für Schwarzweißfernsehbilder, gemittelt über sechs Testbilder, für ein "Gut-Urteil" um 2 dB größer waren, als bei Farbfernsehbildern. Da Störungen in Schwarzweißfernsehbildern zu kritischeren Urteilen führten, wurden die folgenden Bildbeurteilungen nur mit Schwarzweißbildern durchgeführt.

Für das kritischste Testbild mit störendem Einzelbalken (2 Kanäle)

ergab sich ein Störspannungspegel von -48 dB, bei dem 80 % der Bpn mit "gut oder besser" urteilten. Ein aus mehreren teilweise konizidierenden Balken zusammengesetzer Einzelbalken (Vielkanalsystem) wirkt bei gleichem Störspannungspegel weniger störend als der scharf begrenzte Einzelbalken. Für die gleiche Bildqualität("Gut-Urteil") kann der Störspannungsabstand gegenüber dem Einzelbalken um 6 dB verringert werden.

Die stärkste Störwirkung geht von einem senkrechten Balkengitter mit fünf äquidistanten Balken aus. Um die gleiche Bildgüte wie beim scharfbegrenzten Einzelbalken zu erzielen, muß der Störspannungsabstand um 5 dB vergrößert werden. Viele gleichmäßig über das Fernsehbild verteilte senkrechte Balken vermindern die Störwirkung. Sie wirken wie eine Schwarzwertanhebung ($n \geq 10$ Kanäle).

Die z.T. interpolierten Werte lassen sich aus der folgenden Tabelle ablesen:

Nr.	Störart		mittlerer Beobachter	Stufe 2-Urteil	5 %-Punkt	Meßwertdefinition	Skala
1	weißes Rauschen		-35 dB	-41 dB	-42 dB	$20\log\left(\frac{U_{eff}}{0,700}\right)$	5-stufige Störungsskala
2	IM-Rauschen		-40 dB	-46 dB	-47 dB		
3	Kreuzmodulation	Einzelbalken	-39 dB	-48 dB		s.Text	5-stufige Güteskala
		koinzidier.Einzelbalken		-42 dB			
		Balkengitter (5 Balken)		-53 dB			

Die Werte der Störarten 1 und 2 sind mit der Störart 3 aus mehreren Gründen nicht direkt vergleichbar. Es seien hier nur genannt: Verwendung verschiedener Skalen und Testbilder, Messung der Störspannungspegel nach verschiedenen Definitionen und in verschiedenen Frequenzlagen (amerikanische, europäische Norm).

Zur Beantwortung der Fragen a) und b) aus Punkt 1 sind also eigene Messungen möglichst unter weitgehend standardisierten Bedingungen unerläßlich.

6 Zusammenfassung

In KTV-Anlagen wird die Güte gestörter Fernsehbilder mit steigender Kanal-Anzahl zunehmend durch nichtlineare Verzerrungen in den Breitbandverstärkern (Inter- und Kreuzmodulation) verschlechtert. Eine Verbesserung erfordert einen mit der Bildgüte steigenden wirtschaftlichen Aufwand. Um diesen möglichst klein zu halten, ist die vom Fernsehteilnehmer gerade noch tolerierbare Bildgüte zu ermitteln. Die für die Bildgütebeurteilung anwendbaren Verfahren werden diskutiert und auf Anwendbarkeit geprüft. Aus zwei Gründen bietet sich die Einzelbildbeurteilung mit fünfstufiger Güteskala an:

a) Sie erfordert nur ein Urteil je Beurteilungsobjekt (gestörtes Fernsehbild)

b) Sie liefert durch Adjektiva gekennzeichnete Urteile, die sich am "inneren Güte-Maßstab" der Bpn orientieren.

Die Auswertung der Urteile führt je Störart zu einer Darstellung des Zusammenhanges zwischen den je Beurteilungsobjekt ermittelten Urteilsmittelwerten und den zugehörigen Störspannungspegeln durch die logistische Funktion. Ist diese je Störart bekannt, so lassen sich die mittleren Güteurteile anhand der entsprechenden Störspannungspegel vorhersagen, ohne daß jedesmal aufwendige subj. Beurteilungsverfahren durchgeführt werden müssen. Man gelangt zu einer Aussage

über die gerade noch tolerierbare Bildgüte, wenn man die Güteskala in die gleich großen Kategorien "annehmbar" und "unannehmbar" aufteilt und einen gewissen Prozentsatz von Urteilen (z.B. 5 %) der Kategorie "unannehmbar" für das Fernsehbild mit der Grenzgüte zuläßt. Das zum "5 %-Punkt" gehörende mittlere Güteniveau liegt auf der fünfstufigen Güteskala bei etwas besser als "gut". Hieraus läßt sich dann anhand der bekannten logistischen Funktion der noch zulässige Störspannungspegel angeben.

Zur Durchführung des Einzelbildbeurteilungsverfahrens werden ausführliche Angaben gemacht.

Aus zeitlichen und gerätetechnischen Gründen ließ sich der Bildgütetest nicht mehr durchführen, so daß einige auf die vorliegende Aufgabenstellung anwendbare Ergebnisse aus der Literatur entnommen und diskutiert worden sind.

7 Anhang

Um die unter Punkt 3 diskutierten beiden Auswertungsverfahren anhand eines speziellen Anwendungsbeispiels demonstrieren zu können, wurden die benötigten relativen Urteilshäufigkeiten aus einem bei Prosser et al. / 11 / angegebenen Testergebnis rekonstruiert. Dort sind 2/us-Echostörungen in Schwarzweiß-Testbildern (625 Zeilen) anhand einer fünfstufigen Güteskala beurteilt worden. Für jeden Signal-/Echostörabstand x in dB ergab sich ein normierter Urteilsmittelwert $\bar{p}$ gemäß (2) (bei Prosser: p_5). Der Zusammenhang zwischen $\bar{p}$ und x wurde durch eine logistische Funktion gemäß (3) mit den Parametern M = 19,5 dB und S = 4,8 dB angenähert (ohne Restströrung):

$$\bar{p} = \left[1 + e \; x \; p \left(\frac{19{,}5 - x}{4{,}8} \right) \right]^{-1} \qquad (3a)$$

Darüber hinaus wurden die relativen Häufigkeiten der Urteile je Störspannungsabstand durch eine Binomialverteilung 2. Ordnung mit dem Parameter p approximiert.

Da die o.g. Literaturstelle nur $\bar{p}$ angibt, ist p aus Gleichung (12) zu bestimmen und daraus die relativen Urteilshäufigkeiten unter Verwendung von (11) zu berechnen. Die folgende Tabelle 1 zeigt die zusammengestellten Werte:

Tabelle 1

rel. Häufigkeit je Skalenstufe	Signal-/Echostörabstand x in dB						Bildqualität
	10	15	20	25	30	35	
p_1	-	0,001	0,027	0,256	0,627	0,876	sehr gut
p_2	0.001	0,027	0,274	0,540	0,344	0,122	gut
p_3	0,042	0,254	0,488	0,189	0,028	0,002	mittelmäßig
p_4	0,395	0,531	0,197	0,015	0,001	-	schlecht
p_5	0,562	0,187	0,014	-	-	-	sehr schlecht
$\bar{p}$	0,121	0,281	0,526	0,759	0,899	0,962	
p	0,162	0,305	0,523	0,731	0,857	0,923	

Bild 2a zeigt die logistische Funktion gemäß (3a). 50 % der Urteile entfallen bei M = 19,5 dB (mittlerer Beobachter) auf die Güteklassen 1 bis 3 (annehmbar) und 50 % auf die Güteklassen 3 bis 5 (unannehmbar). Man erhält einen hohen Gütestandard, wenn nur 5 % Urteile der Kategorie "unannehmbar" zugelassen werden. Der zugehörige Störabstand (hier 25,6 dB) kann aus Bild 2a für $\overline{p}$ = 0,780 entnommen werden.

Cavanaugh und Lessman / 24 / summieren die relativen Urteilshäufigkeiten je Beurteilungsobjekt über eine bestimmte Anzahl von Skalenstufen und approximieren die Summenhäufigkeiten je Skalenstufengruppe durch eine kumulative Normalverteilung mit den Parametern $\overline{x}$ (Mittelwert) und s (Standardabweichung). Diese Methode wurde in / 24 / auf die Beurteilung von Störungen in NTSC-Farbfernsehbildern angewendet, die durch lineare Verzerrungen des Luminanzsignals (Differentielle Amplitude und Phase) verursacht werden. Zum Vergleich mit der zuerst behandelten Methode wird dieses Verfahren auf den Datensatz nach Tabelle 1 angewendet. Das Ergebnis zeigt die Tabelle 2:

Tabelle 2

Summenhäufigkeit in Skalenstufen-Gruppe		Signal-/Echostörabstand in dB						Bildqualität
	Obergrenze	10	15	20	25	30	35	
1	1,5	-	-	0,03	0,26	0,63	0,88	sehr gut
≦ 2	2,5	-	0,03	0,30	0,80	0,97	1,00	gut o.besser
≦ 3	3,5	0,04	0,28	0,79	0,99	1,00	1,00	mittelmäßig oder besser
≦ 4	4,5	0,44	0,81	0,99	1,00	1,00	1,00	schlecht ode besser

Aus Bild 2b läßt sich z.B. ablesen, wieviel % der Urteile je Störabstand zu den Kategorien Stufe 1, Stufe 2, 3 und 4 oder besser abgegeben worden sind. Das Bild 3 zeigt dieselben Kurven als Geraden im Summenhäufigkeitspapier für kumulativ normalverteilte Zufallsvariable. Hieraus lassen sich die beiden Parameter der Normalver-

teilung leicht ablesen (z.B. $\bar{x} \pm s = 22{,}0 \pm 3{,}9$ dB für Stufe 2 oder besser). Als Parameter sind die Obergrenzen der Skalenstufengruppen zu verwenden, da die Summenhäufigkeit über die Skalenstufen (Klassen) ermittelt worden sind. Die strichlierte Gerade teilt die Urteile in 50 % Urteile der Kategorie "annehmbar" und 50 % der Kategorie "unannehmbar". Läßt man 5 % der Urteile der 2. Kategorie zu, so ergibt sich ein Störabstand von 25,6 dB. Bei diesem Wert urteilen 84 % der Bpn "gut oder besser" bzw. 30 % "sehr gut". Bei 50 % läßt sich aus derselben Geraden der Wert für M entnehmen.

Die logistische Funktion ist eine Kurve von Mittelwerten. Sie läßt sich deshalb aus Bild 2b anhand der zu den 50 %-Werten gehörenden Störabstände je Skalenstufengruppe rekonstruieren. Es ist dabei zu beachten, daß die Skalenstufenobergrenzen anzusetzen sind.

Die beiden Verfahren sind also als äquivalent anzusehen. Die Wahl des Verfahrens wird nur durch die Form der angestrebten Aussagen bestimmt.

Literatur

/ 1 / Mambo, P.L., Coll, D.C.
Perceived Picture Quality in CATV Systems with Impairments, IEEE Transactions on Cable Television, Vol. CATV-4, No. 1, January 1979

/ 2 / Krick, W.
Verbesserung der Kabelfernsehübertragung durch ein optimales kohärentes Trägersystem, Rundfunktechnische Mitteilungen, 1979, Heft 5

/ 3 / Eisler, H.
Measurement of Perceived Acoustic Quality of Sound-Reproducing Systems by Means of Faktor Analysis, JASA 39 (1966), p. 484 - 492

/ 4 / McDermott, B.J.
Multidimensional Analyses of Circuit Quality Judgments, JASA 45 (1969), p. 774 - 781

/ 5 / v. Bismarck, G.V.
Zur mehrdimensionalen Beschreibung der Wahrnehmung komplexer Schalle, DAGA (1973)

/ 6 / Wilkens, H.
Mehrdimensionale Beschreibung subjektiver Beurteilungen der Akustik von Konzertsälen, Acustica 38 (1977), S. 10 - 23

/ 7 / Jost, E.
Über die Klangeigenschaften von Lautsprechern, Jahrbuch des Staatlichen Instituts für Musikforschung (1972), Merseburger, S. 175 - 202

/ 8 / Jackson, J.E.
Scaling Techniques for Subjective Judgments of Image Quality, Journal of the SMPTE, Vol. 83, Nov. 1974

/ 9 / Ulene, R.A.
Subjective Effects of Bit Errors in a PCM Color Television System, RCA Review, Vol. 37, Sept. 1976

/ 10 / Taylor, A.S., Malarkey, P.E., Welch, R.E.
TV Picture Interference Study, 10^{th} international television Symposium, Montreux, 1977, Session $J\text{-}27_{I,II}$, post deadline paper

/ 11 / Prosser, R.D., Allnatt, J.W., Lewis, N.W.
Quality grading of impaired television pictures, Proc. IEE, Vol. 111, Nr. 3, March 1964

/ 12 / C.C.I.R.
Method for the subjective assessment of the quality of television pictures, Rec. 500 and Report 405-2, Section 11 C, XIIIth Plenary Assembly, Geneva, 1974, Vol. XI

/ 13 / Schneeberger, F.
Qualitätsparameter bei der Bewertung digitaler Bildübertragungssysteme, Diss. 1976, TH Aachen

/ 14 / Allnatt, J.W., Corbett, J.M.
Adaption in Observers During Television Quality-Grading Tests, Ergonomics, 1972, Vol. 15, No. 4, 353 - 366

/ 15 / Carroll, J.D., Chang, J.J.
Analysis of individual differences in multidimensional scaling via a N-way generalization of "Eckart-Young" decomposition, Psychometrika 35 (1970), p. 283 - 319

/ 16 / Hopf, H., Kiefer, D.
Über die Störwirkung der Intermodulationsprodukte zwischen Ton- und Farbträger beim PAL-Verfahren, Rundfunktechnische Mitteilungen, Band 11 (1967), H. 2

/ 17 / Guilford, J.P.
Psychometric Methods, McGraw-Hill Book Company, INC., New York (1954), p. 154 - 177

/ 18 / Bortz, J.
Das INDSCAL-Verfahren als Methode zur Differenzierung kognitiver Strukturen, Zeitschr. Für experimentelle und angewandte Psychologie, 1975, Bd.XXII,H.1, S.33-46

/ 19 / Weaver, L.E.
Subjective Impairment of Television Pictures, Electronic Radio Engineer, May 1959

/ 20 / Mertz, P., Fowler, A.D., Christopher, H.N.
Quality Rating of Television Images, Proc. of the I.R.E., Vol. 38, Nov. 1950, p. 1269 - 1283

/ 21 / Cavanaugh, J.R., Lessmann, A.M.
Subjective Effects of Differential Phase Distortions in NTSC Color Television Pictures, Journal of the SMPTE, Vol. 80, Aug. 1971

/ 22 / Flor, W.
Beurteilung intermodulationsgestörter Fernsehbilder durch Nichtexperten, Fernseh- und Kino-Technik, 31. Jahrgang, Nr. 2/1977, S. 46 - 48

/ 23 / Flor, W.
Beurteilung intermodulationsgestörter Fernsehbilder, Rundfunktechnische Mitteilungen, Jahrgang 21 (1977), H. 2, S. 62 - 67

/ 24 / Groenveld, P., van Dijk, B.P.
Subjective Evaluation of Cross-Modulation Effects in CATV Systems, 10th international television symposium, montreux 1977, Session J 26

/ 25 / Prosser, R.D., Allnatt, J.W.
Subjective Quality of television pictures impaired by random noise, Proc. IEE, Vol. 112, No. 6, June 1965

/ 26 / Allnatt, J.W., Prosser, R.D.
Subjective quality of colour television pictures impaired by random noise, Proc. IEE, Vol. 113,No. 4, April 1966

/ 27 / Allnatt, J.W., Lewis, N.W.
Subjective and Objective Impairments in Television Pictures, IEE. CONF. PUBL.46, 1968

/ 28 / Mankopf, M.
Realisierung standardisierter Beobachtungsbedingungen für die Bildgütebeurteilung, Fernseh- und Kino-Technik, 33. Jahrgang, Nr. 6 / 1979

/ 29 / Bühler, J.
Einrichtung zur subjektiven Beurteilung von Fernsehbildern, Fernseh- und Kino-Technik 1972, Nr. 7

/ 30 / Simons, K.E.
Technical Handbook for CATV Systems, Jerrold Electronics Corp. 3rd Edition, 1968, pp. 40 - 45

/ 31 / Lewis, N.W., Allnatt, J.W.
Subjective Quality of Television Pictures with Multiple Impairments, Electronics Letters, Sept. 65, Vol. 1, No. 7

Note	1	2	3	4	5
Bild-Nr.	Bildqualität				
	sehr gut	gut	mittel mäßig	schlecht	sehr schlecht
1	☐	☐	☐	☐	☐
2	☐	☐	☐	☐	☐
3	☐	☐	☐	☐	☐
4	☐	☐	☐	☐	☐
5	☐	☐	☐	☐	☐
6	☐	☐	☐	☐	☐
7	☐	☐	☐	☐	☐
8	☐	☐	☐	☐	☐
9	☐	☐	☐	☐	☐
10	☐	☐	☐	☐	☐
11	☐	☐	☐	☐	☐
12	☐	☐	☐	☐	☐
13	☐	☐	☐	☐	☐
14	☐	☐	☐	☐	☐
15	☐	☐	☐	☐	☐

Bild-Nr.	Bildstörung				
	nicht sichtbar	sichtbar, nicht störend	etwas störend	störend	stark störend
1	☐	☐	☐	☐	☐
2	☐	☐	☐	☐	☐
3	☐	☐	☐	☐	☐
4	☐	☐	☐	☐	☐
5	☐	☐	☐	☐	☐
6	☐	☐	☐	☐	☐
7	☐	☐	☐	☐	☐
8	☐	☐	☐	☐	☐
9	☐	☐	☐	☐	☐
10	☐	☐	☐	☐	☐
11	☐	☐	☐	☐	☐
12	☐	☐	☐	☐	☐
13	☐	☐	☐	☐	☐
14	☐	☐	☐	☐	☐
15	☐	☐	☐	☐	☐

Anlage 2: Geräte-Liste

Nr.	Benennung	Typenbezeichnung	Hersteller
1	Dia-Abtaster	DC 40 LP 49 B	Bosch
2	Graustufen-Generator	TF 2909	Marconi
3	Kontroll-Monitor	MC 51 BA	Bosch
4	Signal-Generator	F 47	I E C
5	Eintast-Testsignal-Generator	148	Tektronix
6	Modulator	PM 5580	Philips
7	Umsetzer	PM 5581	Philips
8	Eichleitung	DPU BN 180433/75	Rohde & Schwarz
9	Breitbandverstärker	S 43115-P0001u.2-A	Siemens
10	Bandpass-Filter	TTA 250-3-5 EE	Telonic
11	Mischer	ZAJ 1	Mini-Circuits-Laboratory
12	HF-Generator	8640 B	Hewlett-Packard
13	TV-Demodulator	Modell 6250	Scientific-Atlanta
14	Video-Störspannungsmesser	UPSF	Rohde & Schwarz
15	Farbmonitor	CTVM 2/51	Barco

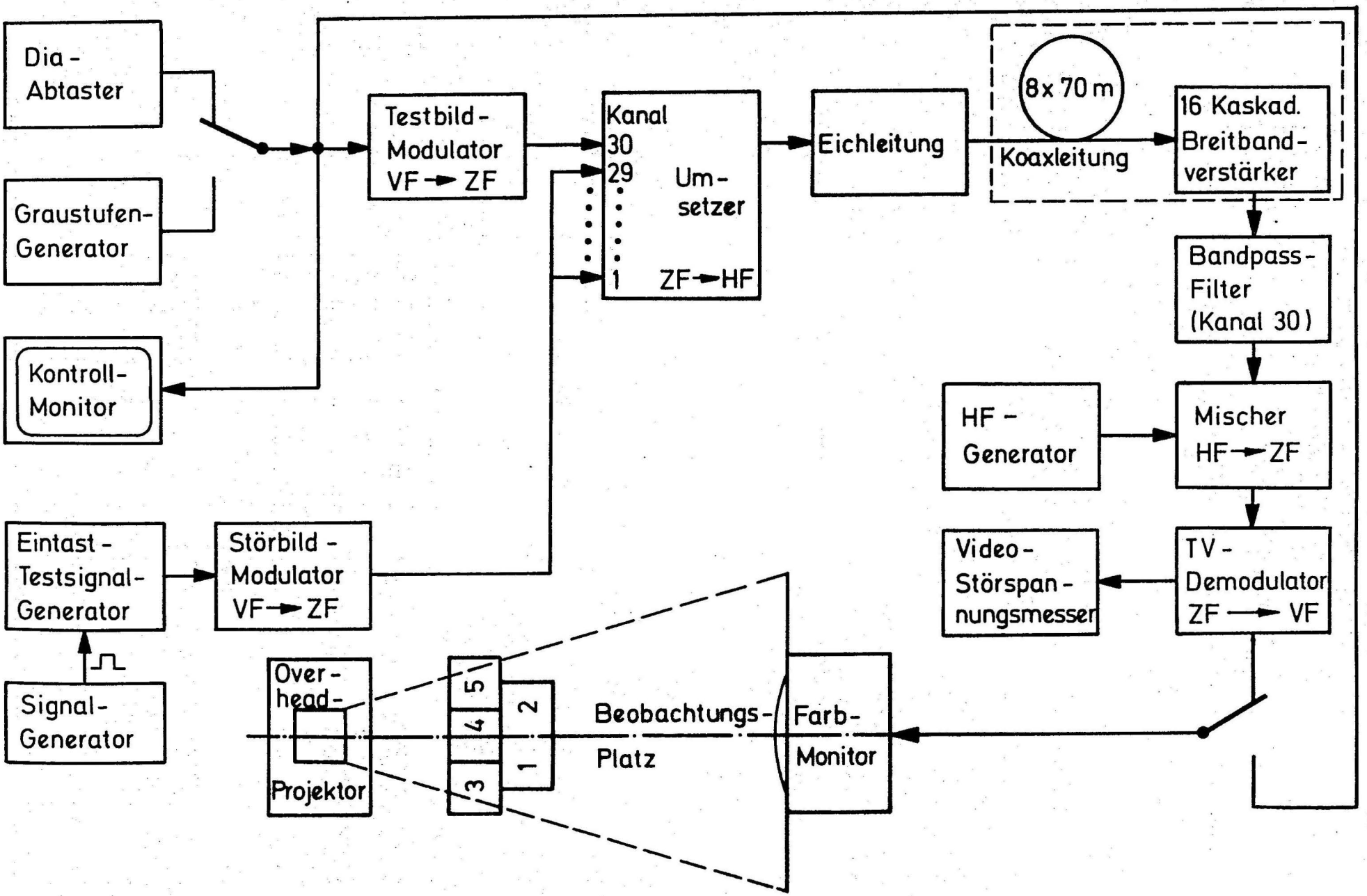


Bild 1: Apparativer Aufbau für die Bildgüte - Beurteilung

Bild 2a

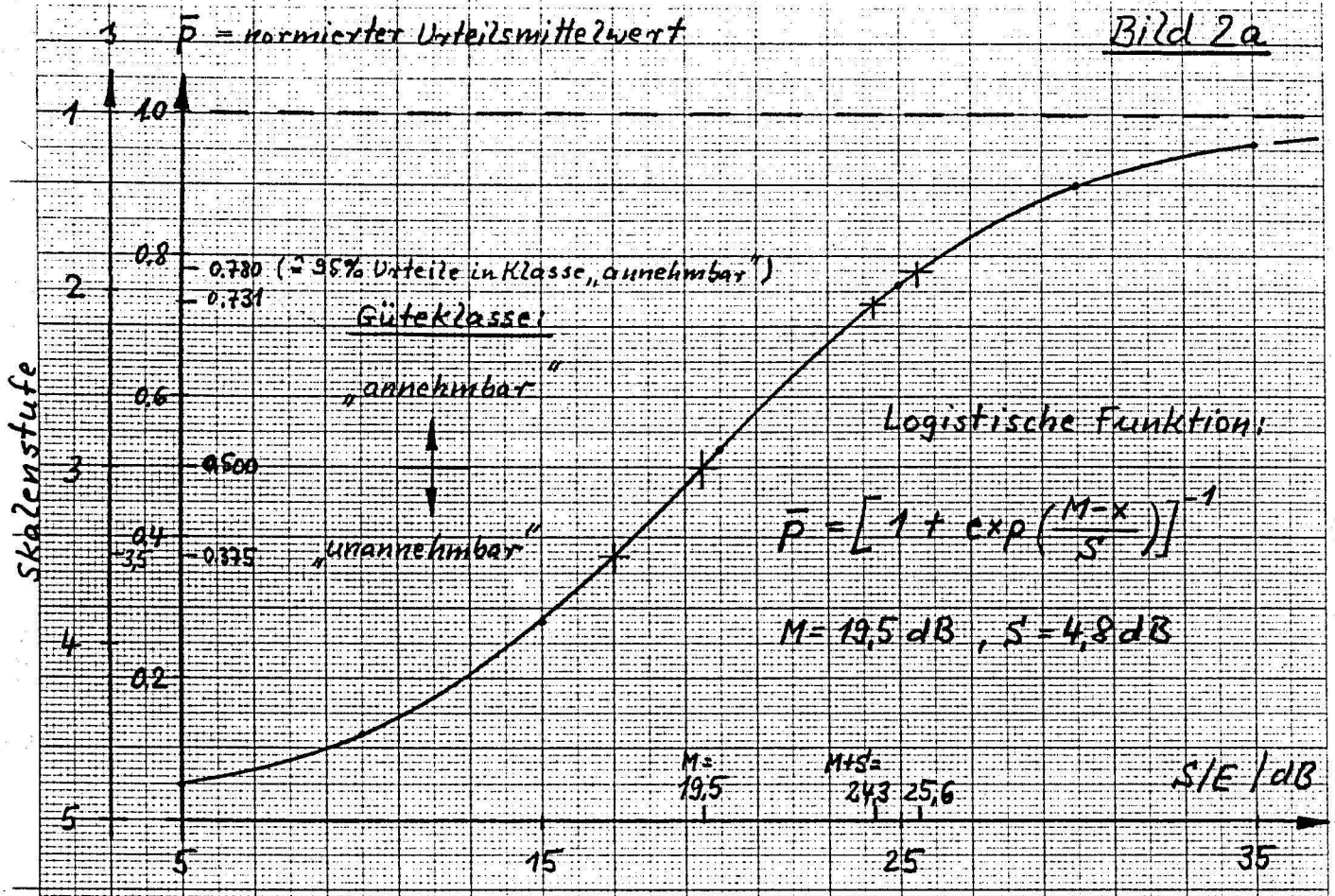
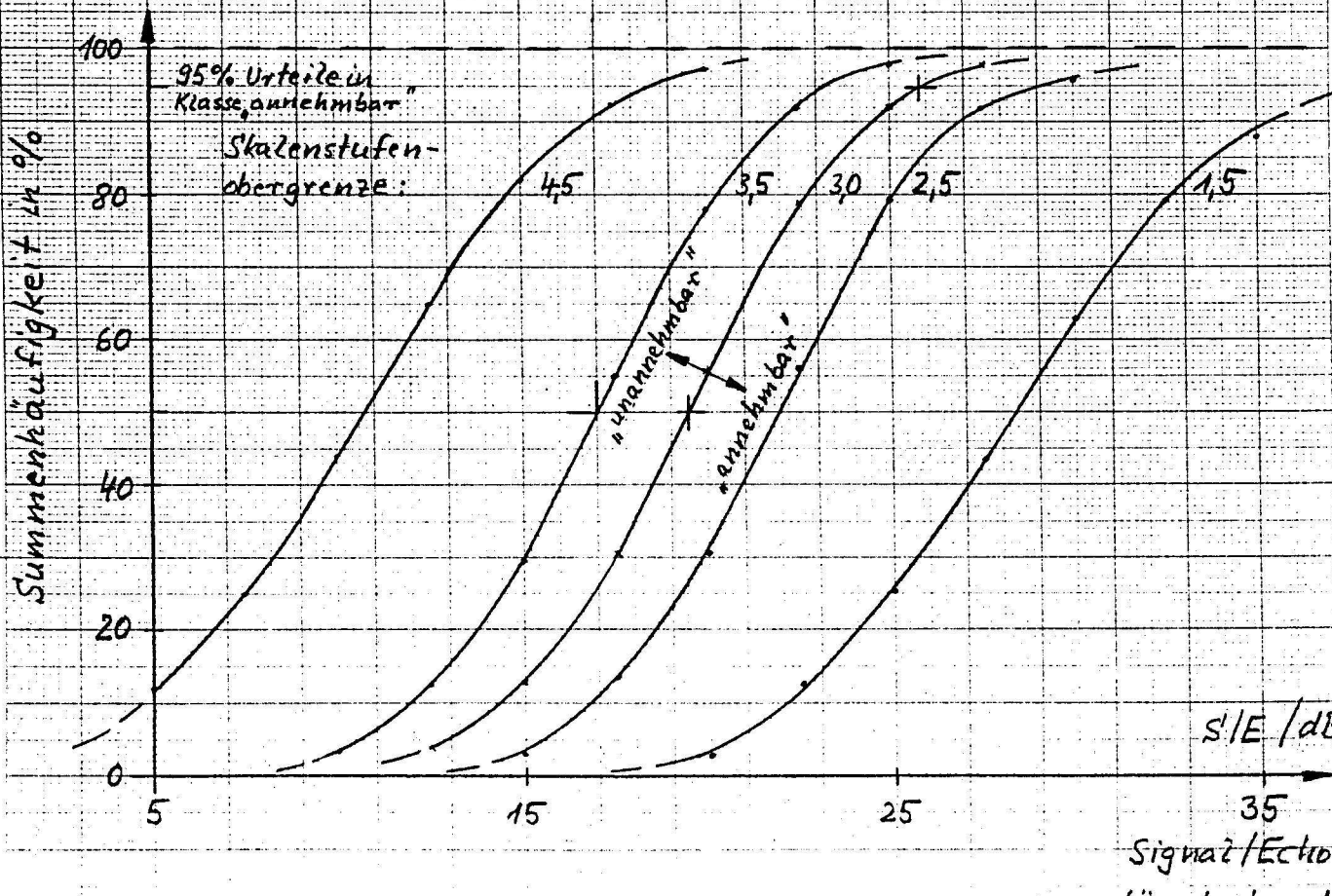
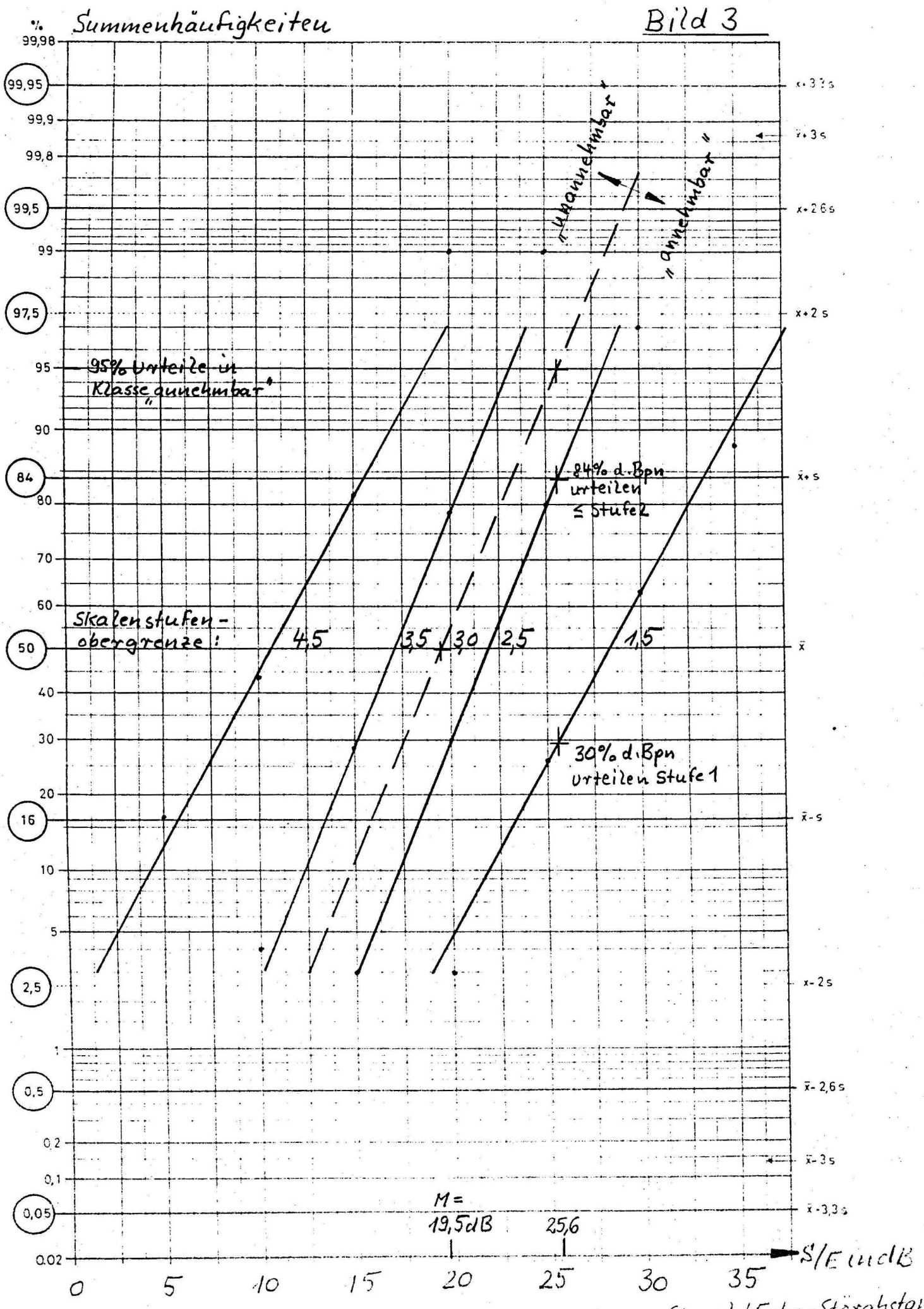
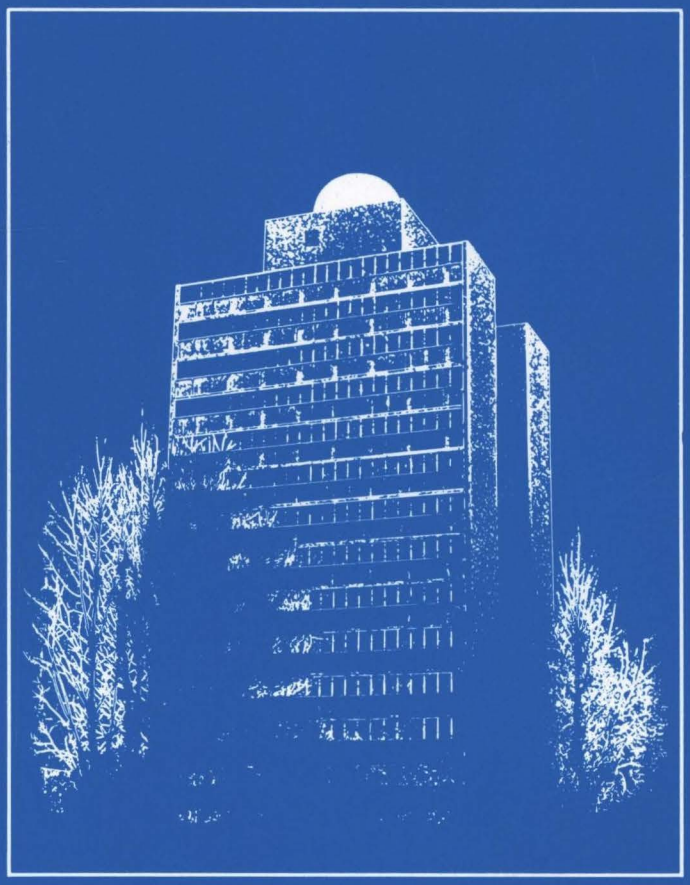


Bild 2b




% Summenhäufigkeiten
Bild 3
99,98
99,95
99,9
99,8
99,5
99
97,5
95
90
84
80
70
60
50
40
30
20
16
10
5
2,5
1
0,5
0,2
0,1
0,05
0,02
x̄+3s
x̄+2,6s
x̄+2s
x̄+s
x̄
x̄-s
x̄-2s
x̄-2,6s
x̄-3s
x̄-3,3s
„unannehmbar"
„annehmbar"
95% Urteile in Klasse „annehmbar"
84% d. Bpn urteilen ≤ Stufe 2
30% d. Bpn urteilen Stufe 1
Skalenstufen-obergrenze:
4,5
3,5
3,0
2,5
1,5
M =
19,5 dB
25,6
S/E in dB
0
5
10
15
20
25
30
35
Signal / Echo-Störabstand

Heinrich-Hertz-Institut
für Nachrichtentechnik
Berlin GmbH