# Einführung in die Informationstheorie*)

Mitteilung aus dem Institut für Nachrichtenvermittlung und Datenverarbeitung der Techn. Hochschule Stuttgart

Von **A. Lotze**, Stuttgart

Mit 3 Bildern

**DK 621.391.001.11**

In den folgenden Ausführungen sollen die wichtigsten Begriffe der Informationstheorie für den in der Praxis stehenden Ingenieur der Nachrichtentechnik anhand anschaulicher Beispiele erläutert werden. Zur Ergänzung sei auf die in dieser Zeitschrift erschienenen Arbeiten [1, 2] und die dort angegebenen Literaturstellen hingewiesen.

## 1. Funktion von Quelle und Empfänger eines Nachrichtenkanals

Kein Nachrichtenmittel vermag unmittelbar Denkinhalte zu übertragen. Es erhebt sich die Frage, was eigentlich bei der elektrischen Nachrichtenübertragung geschieht.

Jene Stelle am Eingang eines Nachrichtenkanals, an der man sich die Nachricht entspringend denken kann, soll als Quelle und die Aufnahmeeinrichtung am anderen Ende des Kanals als Empfänger bezeichnet werden (Bild 1). Beide besitzen eine zweispaltige Zuordnungsliste zwischen einem Symbolalphabet $Q$ und einer Zeichenliste $X$ bzw. einer Zeichenliste $Y$ und einem Symbolalphabet $E$. Angelieferte Symbole einer Liste $Q$ werden in Form von Zeichen einer Liste $X$ übertragen. Die empfangenen Zeichen werden beim Empfänger in eine Liste $Y$ eingeordnet. Jedem Zeichen aus $Y$ wird dann das entsprechende Symbol eines Alphabets $E$ zugeordnet.

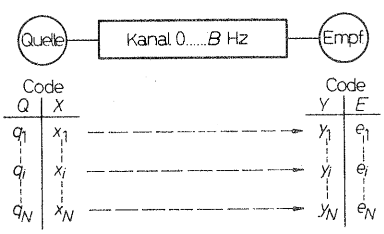


Bild 1. Übertragung einer Nachricht

### Beschränkung auf endlichen Zeichenvorrat der Quelle

Die folgenden Betrachtungen beschränken sich auf diskrete Symbolalphabete mit endlichem Umfang und dementsprechend auf einen endlich großen Zeichenvorrat $N$ in der Liste $X$. Eine wesentliche Einschränkung in der praktischen Anwendbarkeit der hier angestellten Überlegungen auf die elektrische Nachrichtenübertragung ergibt sich dadurch nicht. Wenn z. B. angenommen wird, daß die angelieferten Symbole der Liste $Q$ aus einer kontinuierlichen, niederfrequenten Signalfunktion der Bandbreite $B$ bestehen, so kann man sich diese Signalfunktion nach dem Abtasttheorem ersetzt denken durch deren Abtastwerte im Abstand $\frac{1}{2B}$. Weil aber jeder technisch realisierbare Übertragungskanal zumindest durch Rauschen gestört ist, wird dem Empfänger nur die Unterscheidung endlich vieler Amplitudenstufen der Abtastwerte möglich sein. Im folgenden sei deshalb angenommen, daß eine der Quelle angelieferte kontinuierliche Signalfunktion abgetastet und sogleich in ihren Abtastwerten endlich fein quantisiert werde. Dann genügt stets ein endlicher Umfang des Symbolalphabets $Q$ und des Zeichenvorrats $X$. Der Grenzübergang zum kontinuierlichen Kanal nach *C. E. Shannon* [3] wird also im folgenden nicht behandelt.

Die Zeichenliste $X$ soll $N$ verschiedene Zeichentypen umfassen. Jedes Zeichen soll von der Quelle aus $s$ Stromschritten gebildet werden und jeder Stromschritt soll mit $z$ verschiedenen Amplitudenstufen gesendet werden können. Der Zeichenvorrat $X$ umfaßt also

$$N = z^s \tag{1}$$

verschiedene Zeichentypen und kann aufgefaßt werden als ein Zahlensystem zur Basis $z$ mit $s$ Stellen je Zahl.

Mit den Elementen aus diesem Zeichenvorrat werden die Symbole der Liste $Q$ sozusagen „numeriert". Beim Empfänger wird das empfangene Zeichen — quasi die Symbol-Nummer — wieder umgesetzt in das zugehörige Symbol der Liste $E$.

### Entscheidungsgehalt

Als Maß für den Aufwand, der notwendig ist, um mit irgendeinem derartigen digitalen Code die $N$ Symboltypen eines Alphabets $Q$ unverwechselbar zu unterscheiden („zu numerieren"), wird je Zeichen diejenige einheitliche Stellenzahl $s$ gewählt, die bei einer Numerierung mit Binärzahlen (d. h. bei $z = 2$) notwendig wäre. Man nennt dieses Maß den Entscheidungsgehalt $H_0$. Er ist gleich dem Zweierlogarithmus aus der Zahl $N$ des Zeichenvorrats:

$$H_0 = \mathrm{ld}\, N\,^{1)}\,. \tag{2}$$

$H_0$ wird also gemessen in jener Anzahl von Binärentscheidungen (Bit), welche die Quelle fällen muß, um für ein angeliefertes Symbol $q_i$ die zugehörige Zeichentype $x_i$ aus der Liste $X$ auszuwählen (jede Stelle entspricht einer Binär-Entscheidung).

Wenn die mittlere Übertragungszeit je Zeichen $\tau$ ist, kann man auch den Entscheidungsfluß angeben, welcher dem Kanal zugeführt wird. Er beträgt dann

$$H_0^* = \frac{H_0}{\tau}\,. \tag{3}$$

und wird angegeben in bit/sec.

Dazu sollen zwei einfache Beispiele betrachtet werden:

a) Ein digitaler Code zur Basis $z = 8$ und mit $s = 3$ Stellen je Zeichen soll übertragen werden. Dann umfaßt der Vorrat in der Liste $X$ $z^s = 8^3 = 512$ Zeichentypen, und der Entscheidungsgehalt beträgt $H_0 = \mathrm{ld}\, N = 9$ bit.

*) Als Vortrag gehalten anläßlich der NTG-Fachtagung „Informationstheorie" am 4. April 1963 in Stuttgart.

¹) Man kann auch schreiben

$$H_0 = \mathrm{ld}\, N \text{ bit}.$$

Bei „bit" handelt es sich — wie z. B. auch bei dB und Np — um eine Pseudo-Einheit, eine „Zähleinheit", die nach Belieben hinzugesetzt oder auch weggelassen werden kann.

J7

Bei einer Übertragungszeit je Zeichen von $\tau = 1/500$ s wird der Entscheidungsfluß

$$H_0^* = 9 \cdot 500 = 4500 \text{ bit/s}$$

b) Bei den Symbolen der Liste $Q$ soll es sich um ein niederfrequentes Signal der Bandbreite $B = 3500$ Hz handeln. Seine Amplitude werde im Abstand $\tau = \frac{1}{2B}$ abgetastet, also 7000mal je Sekunde, und die Abtastwerte werden sogleich quantisiert in 32 Stufen (etwa „Telefonqualität"). Jeder Abtastwert hat also einen Entscheidungsgehalt von

$$H_0 = \text{ld}\, 32 = 5 \text{ bit.}$$

Entsprechend der Abtastfrequenz wird $\tau = \frac{1}{7000}$ s und der Entscheidungsfluß $H_0^* = H_0/\tau = 5 \cdot 7000 = 35\,000$ bit/s. Über den echten Nachrichtenwert einer solchen Übertragung für den Empfänger machen aber Entscheidungsgehalt $H_0$ und Entscheidungsfluß $H_0^*$ noch keinerlei Aussagen. Dies zeigt folgende Überlegung:

Nimmt man z. B. an, die Quelle sende endlos, d. h. in stets wiederholter, geordneter Folge, die Zeichentypen für die Buchstaben A bis Z, dann ist dieser Entscheidungsfluß für den Empfänger vollkommen wertlos. Er könnte den Kanal ebensogut abtrennen und sich diese Zeichenfolge oder ABC-Folge selbst erzeugen — mit einer Ausnahme:

Wenn Zeichen durch Übertragungsfehler gefälscht würden! Solche gefälschten Zeichen wären vom Empfänger nicht vorhersehbar, sie würden ihn überraschen! Daraus erkennt man, daß ein irgendwie gearteter Zuwachs an Kenntnis, an Information, beim Empfänger nur entstehen kann, wenn er über das jeweils nächste eintreffende Zeichen im Unsicheren ist, m. a. W., wenn das eintreffende Zeichen einen gewissen Überraschungswert für ihn hat, also seine Kenntnis bereichert (und zwar entweder über die Nachrichten-Quelle oder über eine Geräuschquelle).

Demnach kann also Information von einer Quelle nur durch einen sog. stochastischen Prozeß geliefert werden, d. h. durch einen Prozeß, bei dem eine Vorhersage über den weiteren Verlauf nur mit Wahrscheinlichkeit, nicht mit Sicherheit, möglich ist.

Nur solche Zeichen, deren Eintreffen nicht mit Sicherheit, sondern nur mit Wahrscheinlichkeit vorhersagbar ist, vermehren die Information des Empfängers, und zwar um so mehr, je seltener, je überraschender sie auftreten.

## Mittlerer Informationsgehalt (Entropie)

Um nun den Begriff Informationsgehalt zu definieren, sei ein Code betrachtet, dessen Zeichenliste $X$ einen Vorrat von $N$ statistisch unabhängig auftretenden Zeichentypen $x_1, \ldots x_i \ldots x_N$ besitzen möge. Der Entscheidungsgehalt der Zeichenmenge beträgt also nach Gl. (2) $H_0 = \text{ld}\, N$. Diese $N$ Zeichentypen sollen von der Quelle mit den unterschiedlichen Wahrscheinlichkeiten $p(x_i)$ gesendet werden. Den Kehrwert $1/p(x_i)$ der Wahrscheinlichkeiten kann man dann anschaulich als den „Überraschungswert" der einzelnen Zeichentypen bezeichnen.

Falls es sich um einen Binärcode handelt, würde sich als einheitliche Stellenzahl $s$ je Zeichen wiederum $s = \text{ld}\, N$, also $s = H_0$, ergeben (von Aufrundungen auf ganze Stellen abgesehen). Es sollen aber unterschiedliche Wahrscheinlichkeiten für das Auftreten der einzelnen Zeichentypen angenommen werden, deshalb wird man im Mittel an Stellenzahl je Binärzeichen sparen können, wenn man deren Wahrscheinlichkeit, also deren Überraschungswert, berücksichtigt und mit ungleicher Stellenzahl je Zeichentype in folgender Weise codiert:

Man gibt den häufig auftretenden Zeichentypen mit kleinerem Überraschungswert weniger Stellen und den seltener auftretenden Zeichentypen, die größeren Überraschungswert haben, dementsprechend mehr Binärstellen.

Dies ist ein Prinzip, das lange vor der Begründung der Informationstheorie schon *Samuel Morse* benutzte, als er seinen nach ihm benannten Code aufstellte.

Der Erwartungswert der Binär-Stellenzahl je Zeichen wird dann

$$\bar{s} = \sum_i p(x_i) \cdot s_i \,, \tag{4}$$

wobei $s_i$ die Stellenzahl von $x_i$ ist.

Bei vernünftiger Anwendung dieses Prinzips wird die Binärstellenzahl

$$\bar{s} < \text{ld}\, N = H_0 \,, \tag{5}$$

also kleiner als die einheitliche Stellenzahl $s$. Es läßt sich außerdem zeigen, daß $\bar{s}$ dann seinen Minimalwert erreicht, wenn es gelingt, jeder der $N$ Zeichentypen gerade

$$s_i = \text{ld}\, 1/p(x_i) \tag{6}$$

Binärstellen zu geben, d. h. jene Stellenzahl, welche gleich dem Zweierlogarithmus ihres Überraschungswertes ist. Man erhält dann als Minimalwert

$$\bar{s}_{\min} = \sum_i p(x_i) \cdot \text{ld}\, 1/p(x_i) = H \,. \tag{7}$$

Diesen Wert nennt man — unabhängig vom verwendeten Code — die Entropie der Quelle oder den mittleren Informationsgehalt, den das Auftreten irgendeines der $N$ Zeichen liefert. Dementsprechend wird der mittlere Informationsfluß

$$H^* = \frac{H}{\tau} \,. \tag{8}$$

Man kann diesen mittleren Informationsgehalt $H$ auch interpretieren als den kleinstmöglichen mittleren Aufwand an Binärentscheidungen, der seitens der Quelle notwendig ist, um für ein angeliefertes Symbol die zugehörige Zeichentype $x_i$ aus dem Zeichenvorrat $N$ auszuwählen. Noch anschaulicher könnte man dies auch so formulieren: Die Entropie $H$ ist der kleinstmögliche mittlere Aufwand, um das Auftreten eines bestimmten von den $N$ möglichen Ereignissen $x_1 \ldots x_i \ldots x_N$ zu beschreiben. Damit wird der mittlere Informationsgehalt $H$ einer von $N$ möglichen Zeichentypen gleichgesetzt dem Mindestaufwand, der für die eindeutige Unterscheidung dieser Zeichentype von allen übrigen $(N-1)$ Zeichentypen notwendig ist.

Demgegenüber bezeichnet man mit

$$I_i = \mathrm{ld}\, 1/p\,(x_i) \tag{9}$$

jenen Informationsgehalt, den das Auftreten eines einzelnen Ereignisses, d. h. eines einzelnen Zeichens $x_i$, liefert. Treten alle $N$ Zeichentypen gleichwahrscheinlich auf, so wird einheitlich

$$p\,(x_i) = 1/N \text{ und } I_i = \mathrm{ld}\, 1/p\,(x_i) = \mathrm{ld}\, N = s\,. \tag{10}$$

Die Entropie $H$ erreicht dann ihren Maximalwert, es wird

$$\sum_1^N 1/N \cdot \mathrm{ld}\, N = \mathrm{ld}\, N = H_0 = H_{\max}\,. \tag{11}$$

Die Entropie (mittlerer Informationsgehalt) $H$ und der Entscheidungsgehalt $H_0$ sind dann gleich groß. Auch dieser Höchstwert des mittleren Informationsgehalts im Falle der Gleichwahrscheinlichkeit aller Zeichentypen ist einleuchtend, wenn man bedenkt, daß im Falle ungleicher Wahrscheinlichkeiten ein Beobachter bei der Quelle noch mit Erfolg eine Wette abschließen könnte, welches Zeichen als nächstes gesendet wird. Er müßte nur auf jene Zeichentype setzen, welche die größte Wahrscheinlichkeit $p\,(x_i)$ hat. Durch die Kenntnis der ungleichen Wahrscheinlichkeiten besäße dieser Beobachter also ein gewisses Maß an „Vor-Information", er würde dementsprechend (im Mittel) weniger überrascht.

Man nennt deshalb einen Code, dessen Entropie $H$ kleiner ist als der Maximalwert $H_0$, redundant, d. h. weitschweifig, und definiert als Redundanz

$$R = H_0 - H \text{ bzw. } r = \frac{H_0 - H}{H_0}\,. \tag{12}$$

## 2. Der Übertragungskanal

Bisher wurden nur Vorgänge bei der Quelle beschrieben. Als nächstes stellt sich nun die Frage nach dem Transport von Nachrichten über einen Übertragungskanal. Dazu muß untersucht werden, bis zu welchem Grad ein Kanal gegebener Eigenschaften (z. B. Bandbreite und Störabstand) überhaupt in der Lage ist, einen von der Quelle mit irgendeinem Code gelieferten Informationsfluß so zu transportieren, daß der Empfänger die von der Quelle gesendeten Zeichen eindeutig wiedererkennen, d. h. eindeutig decodieren kann.

Die Kapazität $C$ eines Nachrichtenkanals ist deshalb definiert als jener obere Grenzwert des mittleren Informationsflusses $H^*$, der von einem Kanal gegebener Eigenschaften — bei beliebigem Aufwand für die vorteilhafteste Codierung — gerade noch so übertragen werden kann, daß dem Empfänger die sichere Wiedererkennung, d. h. fehlerfreie Decodierung — ebenfalls mit beliebig hohem Aufwand — noch möglich ist.

### Kanalkapazität bei fehlerfreier Übertragung

Zunächst soll von der vereinfachenden Vorschrift ausgegangen werden, daß kein Zeichen bei der Übertragung verfälscht werden darf [4]. Dazu werde ein Kanal der Bandbreite $B$ betrachtet mit dem empfangsseitigen Störabstand $r$:

$$r = 20 \cdot \log \frac{U_{N\max}}{U_{G\max}} \approx 6 \cdot \mathrm{ld} \frac{U_{N\max}}{U_{G\max}}\,, \tag{13}$$

wobei mit $U_N$ die Nutz- und mit $U_G$ die Geräuschspannung bezeichnet werden sollen.

Der Empfänger wird also maximal

$$z_0 = \frac{U_{N\max}}{U_{G\max}} \tag{14}$$

Amplitudenstufen sicher unterscheiden können. Dann ist $z_0$ die größte zulässige Basis des digitalen Code einer Zeichenliste $X$ bei der Quelle. Der damit je Stromschritt mögliche Entscheidungsgehalt werde als die Dynamik $D_0$ des Kanals bezeichnet, für die gilt

$$D_0 = \mathrm{ld}\, z_0\,. \tag{15}$$

Mit der nach dem Abtast-Theorem maximal erlaubten Anzahl von Stromschritten je Sekunde erhält man deshalb den maximalen Entscheidungsfluß, der ungefälscht übertragen werden kann:

$$H^*_{0\max} = 2\,B \cdot \mathrm{ld}\, z_0 = 2\,B \cdot D_0\,. \tag{16}$$

Bei einem redundanz-freien derartigen Code werden Entscheidungsgehalt und Entropie gleich, also auch

$$H^*_{\max} = H^*_{0\max} = 2\,B \cdot D_0 = C_0\,. \tag{17}$$

Dieser Wert werde als die Kanalkapazität $C_0$ bezeichnet. Sie gilt für den speziellen Fall einer fehlerfreien Übertragung — und damit auch fehlerfreien Decodierung.

### Kanalkapazität bei fehlerbehafteter Übertragung

Der allgemeinste Fall ist aber der Kanal mit fehlerfreier Decodierung trotz gestörter, fehlerbehafteter Übertragung. Ein Teil der von der Quelle gesendeten Zeichen kommt also verfälscht beim Empfänger an.

Für diesen Fall werde eine Quelle betrachtet, deren Zeichenliste $X$ einen Code zu einer beliebigen Basis $z \geqq 2$ und mit einer einheitlichen Stellenzahl $s$ je Zeichen besitzt. Die Redundanz $R$ sei Null, d. h. alle möglichen Zeichen sollen gleichwahrscheinlich und statistisch unabhängig voneinander auftreten. Die Liste $X$ umfaßt demnach einen Zeichenvorrat $N_Q = z^s$. Der Entscheidungsgehalt $H_0$ und der mittlere Informationsgehalt $H$ sind also gleich groß. Sie betragen

$$H_0\,(X) = H\,(X) = \mathrm{ld}\, N_Q\,. \tag{18}$$

Die Ursache der Übertragungsfehler kann man sich in einer Geräuschquelle $G$ denken, welche unabhängig von der Quelle ebenfalls nach einem stochastischen Prozeß Information erzeugt und in den Kanal einspeist (Bild 2). Dadurch werden die von der Quelle gesendeten Zeichen in statistisch verteilten Abständen mehr oder weniger verfälscht.

Der Empfänger kann deshalb von einem empfangenen Zeichen $y_k$ nicht mehr mit Sicherheit rückschließen auf das gesendete Zeichen $x_i$. Er ist unsicher über die Herkunft eines empfangenen Zeichens $y_k$ aus der Liste $X$ der Quelle. Mit einer im Einzelfall verschieden großen Wahrscheinlichkeit kann jedes gesendete Zeichen $x_i$ unter dem Einfluß der Geräuschquelle übergegangen sein in das empfangene Zeichen $y_k$.

### Äquivokation, Irrelevanz, Transinformationsgehalt

Die Wahrscheinlichkeit dafür, daß ein bereits empfangenes Zeichen $y_k$ ein bestimmtes Zeichen $x_i$ zur Ursache hatte, nennt man deshalb die bedingte Rückschluß-Wahrscheinlichkeit $p(x_i | y_k)$.

Der zugehörige Informationsgehalt dieses Ereignisses ist dann

$$I_{x_i | y_k} = \mathrm{ld}\, 1/p(x_i | y_k)\,. \tag{19}$$

Dies ist jener Informationsgehalt, der dem Empfänger im Einzelfall zum eindeutigen Rückschluß von $y_k$ auf das gesendete Zeichen $x_i$ fehlt. Nur ein hypothetischer Beobachter, der Quelle und Empfänger gleichzeitig übersehen kann, kommt in den Besitz dieses Informationsgehaltes.

Man nennt ferner die Wahrscheinlichkeit für das Auftreten eines bestimmten solchen „Pärchens", eines Verbundereignisses, die Verbundwahrscheinlichkeit $p(x_i, y_k)$.

Damit kann man auch den Erwartungswert jenes Rückschlußinformationsgehaltes angeben, der dem Empfänger im Mittel zum sicheren Rückschluß, d. h. zur sicheren Wiedererkennung eines gesendeten Zeichens $x_i$ fehlt. Er beträgt

$$H(X | Y) = \sum_i \sum_k p(x_i, y_k) \cdot I_{x_i | y_k}\,. \tag{20}$$

Diesen Entropiewert nennt man auch die Äquivokation (vom englischen „Vieldeutigkeit").

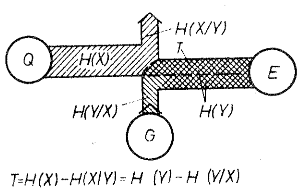


Bild 2. Übertragung von der Nachrichtenquelle zum Nachrichtenempfänger über einen gestörten Nachrichtenkanal. Der Einfluß der Kanalstörungen auf die Nachrichtenübertragung

In Bild 2 ist dargestellt, was bei der Übertragung geschieht: Die Nachrichtenquelle liefert je Zeichen den mittleren Informationsgehalt $H(X)$. Davon wird im Mittel der Teil $H(X | Y)$ während der Übertragung durch den Einfluß der Geräuschquelle „zerstört", sozusagen „aus dem Kanal hinausgeworfen". Statt dessen liefert die Geräuschquelle „sinnlose Überraschungen" an den Empfänger, deren irrelevanter, mittlerer Informationsgehalt durch die bedingte Entropie $H(Y | X)$, die sog. Irrelevanz, gemessen wird. Jenen unvollständigen mittleren Informationsgehalt $T$, der den Empfänger über ein bei der Nachrichtenquelle eingetretenes Ereignis (das Auftreten von $x_i$) im Mittel noch erreicht, nennt man den

$$\text{Transinformationsgehalt } T = H(X) - H(X | Y). \tag{21a}$$

Dementsprechend wird der Transinformationsfluß

$$T^* = H^*(X) - H^*(X | Y)\,. \tag{22}$$

Aus Bild 2 ersieht man, daß $T$ auch noch anders definiert werden kann:

Beim Empfänger ergibt sich aufgrund der Überraschungswerte bzw. der Wahrscheinlichkeiten der eintreffenden (richtigen oder gefälschten) Zeichen ein scheinbarer mittlerer Informationsgehalt $H(Y)$ je Zeichen, in dem die Irrelevanz-Entropie der Geräuschquelle mit enthalten ist. Also gilt auch

$$T = H(Y) - H(Y | X)\,. \tag{21b}$$

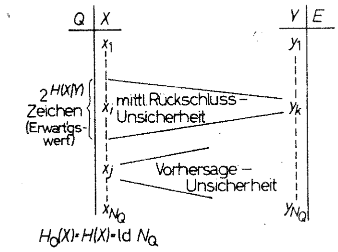


Bild 3. Rückschlußunsicherheit und Vorhersageunsicherheit bei der Nachrichtenübertragung über einen gestörten Kanal

Eine etwas andere Darstellung desselben Tatbestands zeigt Bild 3. Man betrachtet wieder die beiden Code-Tabellen $Q/X$ beim Sender und $Y/E$ beim Empfänger.

Die mittlere Rückschlußunsicherheit des Empfängers über die Herkunft eines empfangenen Zeichens $y_k$ drückt sich in Bild 3 aus in einem Fächer, der vom Empfänger zur Liste $X$ der Quelle hin offen ist. Dabei wird vorausgesetzt, daß die Zeichen $x_i$ der Liste $X$ so geordnet sind, daß benachbarte Zeichen jeweils größte Ähnlichkeit miteinander besitzen.

In gleicher Weise läßt sich auch die Irrelevanz $H(Y | X)$ darstellen durch einen Fächer, der von einem gesendeten Zeichen $x_j$ zur Liste $Y$ des Empfängers hin geöffnet ist. Er kennzeichnet die Unsicherheit eines Beobachters bei der Quelle über die Ankunft eines gesendeten Zeichens $x_j$ in der Liste $Y$ des Empfängers.

Dem Empfänger nützt nun zunächst der unzerstört übertragene Anteil $T$ von $H(X)$ wenig, da er ja wegen des Abmangels $H(X | Y)$ nicht ausreicht, um die von der Quelle gesendeten Zeichen sicher wiederzuerkennen.

### Bedeutung der Redundanz für die Fehlererkennung und Fehlerkorrektur

Kein Codierungsverfahren ist in der Lage, bei einem redundanzfreien Entscheidungsgehalt $H_0(X) = H(X) = \mathrm{ld}\, N_Q$ der Quelle den im Kanal vernichteten Anteil $H(X | Y)$ auf der Empfängerseite wieder zu regenerieren, sozusagen den „Fächer der Rückschluß-Unsicherheit zusammenzuklappen". Es ist deshalb wichtig zu wissen, wie viele Zeichen der Liste $X$ als mögliche Ursachen eines empfangenen Zeichens $y_k$ der in Bild 3 eingezeichnete Fächer der mittleren Rückschlußunsicherheit umfaßt:

Ein fehlender mittlerer Informationsgehalt $H(X | Y)$ entspricht jenem Entscheidungsgehalt, der zur Auswahl eines bestimmten unter $2^{H(X | Y)}$ gleichwahrscheinlichen Ereignissen $x_i$ notwendig ist. [Das ist eben jene Anzahl von Auswahl-Entscheidungen, die der Empfänger nicht treffen kann, weil ihm $H(X | Y)$ fehlt.] Der Fächer der mittleren Rückschluß-Unsicherheit umfaßt demnach $2^{H(X | Y)}$ Zeichen der Liste $X$ als mögliche Ursachen eines empfangenen Zeichens $y_k$, nämlich „ein richtiges und $[2^{H(X | Y)} - 1]$ falsche Zeichen" der Liste X. Nach *C. E. Shannon* [3] muß es aber möglich sein, als oberen Grenzwert in einem gestörten Kanal dennoch je Zeichen einen mittleren Informationsgehalt $T < H(X)$ so zu übertragen, daß er noch fehlerfrei decodiert werden kann, m. a. W., daß jede Fälschung der übertragenen Zeichen beseitigt wird.

Ein richtig decodierbarer Informationsgehalt $T$ je gesendetes Zeichen $x_i$ der Quelle ergibt sich offenbar nur dadurch, daß man sich in der Benützung des gesamten Zeichenvorrats $N_Q$ der Liste $X$ Beschränkungen auferlegt. Es soll deshalb nur noch eine Untergruppe $X_T$ mit $N_T < N_Q$ Zeichentypen (auch gleichwahrscheinliche) benutzt werden, so daß die Entropie der Quelle nur noch

$$H(X_T) = \operatorname{ld} N_T \tag{23}$$

beträgt. Es soll ferner $N_T$ so gewählt werden, daß gilt:

$$N_T = \frac{N_Q}{2^{H(X|Y)}} \,. \tag{24}$$

Der Gesamtvorrat $N_Q$ der Liste $X$ wird also geteilt durch die mittlere Fächerbreite der Rückschluß-Unsicherheit. Dementsprechend muß man sich auch in der Symbolliste $Q$ auf nur $N_T$ Symbole beschränken $(q_1 \ldots q_{N_T})$. Ferner sei angenommen, daß diese $N_T$ benutzten Zeichentypen der Untergruppe $X_T$ so geschickt aus dem Gesamtvorrat $N_Q$ ausgewählt wurden, daß in jedem Fächer der mittleren Rückschlußunsicherheit, der von einem empfangenen Zeichen $y_k$ ausgeht, nur eines der benutzten $N_T$ Zeichen fallen kann.

Dann muß künftig zwischen zwei verschiedenen Entropiewerten der Quelle unterschieden werden, und zwar zwischen

a) der Entropie $H(X) = H_0(X) = \operatorname{ld} N_Q$,
die für den bisherigen Fall gilt, daß der Gesamtvorrat aller $N_Q$ Zeichen gleichwahrscheinlich benutzt wird; da der Code nicht geändert wird, behält dieser Entropiewert in der Bedeutung als Entscheidungsgehalt $H_0(X)$ weiterhin Gültigkeit;

und

b) der — kleineren — Entropie $\operatorname{ld} N_T = H(X_T)$,
welche nunmehr, bei der Beschränkung auf die Benutzung der Untergruppe $N_T$, tatsächlich gilt (und jetzt am Ausgang der Quelle statistisch meßbar wäre).

An dem Verhalten der Geräuschquelle $G$, damit auch am Fächer der Äquivokation in Bild 3, hat sich dadurch nichts geändert. Aber der Empfänger kennt nun den verabredeten kleineren Zeichenvorrat $N_T$. Nun möge zunächst die unrealistische Annahme gelten, daß der beim Empfang eines Zeichens $y_k$ fehlende Rückschluß-Informationsgehalt stets gleich dessen Erwartungswert $H(X|Y)$ sei. Unter allen möglichen $2^{H(X|Y)}$ Ursachen eines empfangenen Zeichens $y_k$ befindet sich dann in der Liste $X$ der Quelle jeweils nur ein Zeichen $x_i$, welches gleichzeitig der Untergruppe $X_T$ angehört. Der Empfänger kann also das empfangene Zeichen $y_k$ mit Sicherheit dieser einzig möglichen Ursache $x_i$ zuordnen. Damit würde der Fächer der mittleren Rückschlußunsicherheit zwar nicht „zusammengeklappt", aber bis auf eine mögliche Ursache $x_i$ entleert — was auf das gleiche hinausläuft. Dieser Tatbestand zeigt sich auch formal, denn es wird

$$H(X_T) = \operatorname{ld} N_T = \operatorname{ld} \frac{N_Q}{2^{H(X|Y)}}$$
$$= \operatorname{ld} N_Q - H(X|Y) = H(X) - H(X|Y) = T \,. \tag{25}$$

Es wird also tatsächlich der mittlere Transinformationsgehalt $T$ gleich groß wie die Entropie $H(X_T)$ des von der Quelle benutzten beschränkten Zeichenvorrats $X_T$. Das bedeutet aber, daß in Bezug auf das verkleinerte Symbol- und Zeichenalphabet $X_T$ mit dem Vorrat $N_T$ die Äquivokation, d. h. die Vieldeutigkeit empfangener Zeichen bzw. die Rückschluß-Unsicherheit, zu Null wird:

$$H(X_T|Y) = 0 \,. \tag{26}$$

Die Redundanz der Quelle bezüglich der Entropie der benutzten Untergruppe $X_T$ wird dann

$$R_T = \operatorname{ld} N_Q - H(X_T) = H(X) - T = H(X|Y) \,. \tag{27}$$

Der redundante Anteil $R_T$ des Entscheidungsgehalts $H_0$ pro Zeichen der Untergruppe $X_T$ aus $N_T$ Zeichentypen ist also gerade so groß wie jener mittlere Rückschluß-Informationsgehalt $H(X|Y)$, der im Kanal durch den Einfluß der Geräuschquelle pro Zeichen im Mittel verloren geht.

Dieses Ergebnis kann man auch so formulieren:

Durch die Redundanz in der (unverändert beibehaltenen) Codierung des verkleinerten Zeichenvorrats $N_T$ wird dem Empfänger so viel zusätzlicher Entscheidungsgehalt pro Zeichen mitgeliefert, daß er trotz der Fälschung eines empfangenen Zeichens $y_k$ sich noch für das richtige Herkunftszeichen $x_i$ entscheiden kann.

Diesen Überlegungen lag allerdings die oben erwähnte unrealistische Annahme zugrunde, wonach stets mit der mittleren Breite des Fächers der Rückschlußunsicherheit, m. a. W. mit der mittleren Anzahl möglicher Herkunftszeichen und nicht mit deren Augenblickswerten gerechnet wurde. Später wird noch gezeigt werden, inwieweit dadurch in der Praxis das Erreichen dieses *Shannon*schen oberen Grenzwerts $T$ bzw. $T^*$ eingeschränkt wird.

### Gegenüberstellung der Kanalkapazität bei fehlerfreier bzw. fehlerbehafteter Übertragung

Zunächst soll noch die von *Shannon* angegebene Kapazität $C_{\mathrm{Sh}}$ eines gestörten Nachrichtenkanals angegeben werden. Die Kanalkapazität war bereits definiert worden als jener obere Grenzwert des sicher decodierbaren Informationsflusses, der bei vorteilhaftester Codierung erreicht werden kann. Also wird

$$C_{\mathrm{Sh}} = T^*_{\max} = {}_{\max}\{H^*(X) - H^*(X|Y)\}\,^{*)} \,. \tag{28}$$

*) In der Diskussion wurde die Frage angeschnitten, ob es zulässig sei, für die Herleitung von $C_{\mathrm{Sh}}$ ein Symbolalphabet und dessen Zeichenvorrat $N_T$ zu verwenden, bei dem alle $N_T$ Zeichentypen gleichwahrscheinlich und statistisch unabhängig auftreten. Dies ist deshalb erlaubt, weil die Kanalkapazität $C_{\mathrm{Sh}}$ definitionsgemäß lediglich eine Eigenschaft des Kanals (unabhängig von der Nachrichtenquelle) darstellt. $C_{\mathrm{Sh}}$ ist also unabhängig von der irgendwie gearteten Redundanz $R$ irgendeiner „Urquelle". Wenn $C_{\mathrm{Sh}}$ weitmöglichst ausgenutzt werden

Die Äquivokation $H(X|Y)$ hängt aber nicht allein von der Geräuschquelle ab, sondern auch von der Störanfälligkeit des verwendeten Code gegenüber den speziellen Eigenschaften der jeweiligen Geräuschquelle. Deshalb kann der obere Grenzwert $C_{Sh}$ offenbar nur erreicht bzw. angestrebt werden mit einer Nachrichtenquelle, deren Codierungsverfahren für die Liste $X$ optimal an die Eigenschaften des Kanals und seine Geräuschcharakteristik angepaßt ist. Um die Gleichung (17) für die Kanalkapazität $C_0$ und Gleichung (28) für die Kanalkapazität $C_{Sh}$ besser vergleichen zu können, wird Gleichung (28) etwas umgeformt:

Da mit der nach dem Abtasttheorem höchstmöglichen Anzahl von Stromschritten pro Sekunde gesendet wird, kann man mit

$$h = \frac{H^*}{2B} \quad (29)$$

die Entropie je Stromschritt bezeichnen und erhält dann

$$C_{Sh} = 2B \cdot {}_{\max}\{h(x) - h(x|y)\}. \quad (30\,A)$$

Kürzt man ab

$$D_{\max} = {}_{\max}\{h(x) - h(x|y)\}, \quad (31)$$

so erhält Gleichung (30 A) folgende Form:

$$C_{Sh} = 2B \cdot D_{\max}. \quad (30\,B)$$

$D_{\max}$ bedeutet dann den Maximalwert des mittleren Transinformationsgehalts je Stromschritt.

Zum Vergleich sei nochmals die Gleichung (17) angeführt, welche für die Kanalkapazität bei f e h l e r f r e i e r Übertragung gilt:

$$C_0 = 2B \cdot D_0 = 2B \cdot \mathrm{ld}\, z_0. \quad (17)$$

Wählt man als Basis des zu übertragenden Code einen Wert $z_0$ gemäß den Gleichungen (17) und (14), so entspricht dies einer Abstufung der gesendeten Amplitude je Stromschritt, welche so grob ist, daß sie durch die Störspannungen des Kanals mit Sicherheit noch nicht verfälscht wird. Mit diesem Code kann also ein mittlerer Informationsgehalt $D_0 = \mathrm{ld}\, z_0$ je Stromschritt fehlerfrei übertragen werden. Wählt man dagegen als Basis des gesendeten Code einen Wert $z_x > z_0$, so bedeutet dies eine feinere Abstufung der Amplitude gesendeter Stromschritte. Der Empfänger wird deshalb wegen der Störspannungen des Übertragungskanals die gesendeten Amplitudenstufen nicht mehr mit Sicherheit wiedererkennen.

Die Erhöhung des gesendeten Entscheidungsgehalts um den Betrag ($\mathrm{ld}\, z_x - \mathrm{ld}\, z_0$) je Stromschritt wird aber bei geeigneter Codierung nicht ausschließlich für fehlerkorrigierende Redundanz benötigt, sondern ein Teil hiervon dient der Übertragung z u s ä t z l i c h e n Informationsgehalts. Der Empfänger erhält deshalb je Stromschritt einen mittleren Transinformationsgehalt $D > D_0$, welcher bei optimaler Codierung den oberen Grenzwert $D_{\max}$ erreicht.

Dies heißt: Ein Code, welcher die *Shannon*sche Kanalkapazität anstrebt, muß auf jeden Fall den Kanal bezüglich der Anzahl $z_0$ s i c h e r unterscheidbarer Amplitudenstufen ü b e r s t e u e r n [5].

## Grenzen der Fehlererkennung und Fehlerkorrektur

Zum Schluß soll untersucht werden, inwieweit in der Praxis die völlig gesicherte Übertragung eines Transinformationsflusses gemäß den Gleichungen (22) und (28)

$$T^* \leqq T^*_{\max} = C_{Sh} \quad (32)$$

angenähert werden kann.

Wie erwähnt, war für die Herleitung der Gleichungen (25), (26) und (28) immer mit dem E r w a r t u n g s w e r t $H(X|Y)$ des fehlenden Rückschluß-Informationsgehaltes, also mit einem M i t t e l wert für die Anzahl möglicher Herkunftszeichen $x_i$ als Ursache eines empfangenen Zeichens $y_k$, gerechnet worden. Der Erwartungswert $H(X|Y)$ entspricht aber auch einem Erwartungswert $f_m$ der Fehlerzahl pro Zeichen.

Die t a t s ä c h l i c h e Fehlerzahl pro empfangenes Zeichen ist jedoch in Wirklichkeit eine Z u f a l l s v a r i a b l e, deren Augenblickswerte vom Verhalten der Geräuschquelle abhängen.

Als einfachstes Beispiel werde dazu ein Binärcode betrachtet mit einer einheitlichen Stellenzahl $s$ pro Zeichen, also einem Gesamt-Zeichenvorrat $N_Q = 2^s$, und dem Entscheidungsgehalt $H_0 = s$. Im Einzelfall gilt für die Wahrscheinlichkeit $p(f)$, daß $f$ Fehler pro Zeichen auftreten:

$$p(f) > 0 \text{ für jeden Wert } 0 \leqq f \leqq s. \quad (33)$$

Nur für $s \to \infty$ würde

$$f \to f_m = \text{const} = s \cdot p_E \quad (34)$$

werden (wobei $p_E$ die Schrittfehlerwahrscheinlichkeit bedeutet). Für z. B. $f$ Fehler pro Binärzeichen in beliebiger Anordnung innerhalb der $s$ Stellen ergeben sich $U_f = \binom{s}{f}$ mögliche Herkunftszeichen in der Liste $X$ der Quelle. Um sich für das eine richtige unter diesen $U_f$ Zeichen zu entscheiden, muß der Empfänger aus der mitgelieferten Redundanz (mindestens) einen Entscheidungsgehalt $R = \mathrm{ld}\, U_f$ bit entnehmen können.

Würde man verlangen, daß der Empfänger sich in jedem denkbaren Einzelfall (d. h. bei $0 \ldots s$ Fehlern je Zeichen) richtig entscheidet, so würde er pro Zeichen folgenden redundanten Entscheidungsgehalt für die Auswahl unter allen möglichen Fehlermustern aus $0 \leqq f \leqq s$ Fehlern pro Zeichen benötigen:

$$R_T = R_{T_{\max}} = \mathrm{ld} \sum_{f=0}^{f=s} \binom{s}{f} = \mathrm{ld}\, 2^s = s. \quad (35)$$

Damit würde aber die Entropie

$$H_T = H_0 - R_{T_{\max}} = s - s = 0.$$

soll, so muß in jedem Fall gefordert werden, daß die Meldungen dieser Urquelle derart umcodiert werden (z. B. durch Zusammenfassung sehr langer Folgen von Meldungen zu einem neuen Symbol), daß ein neues Alphabet aus $N_T$ Symboltypen entsteht, welche gleichwahrscheinlich und statistisch unabhängig auftreten. Diese $N_T$ Symboltypen sind dann durch einen Zeichenvorrat $X_T$ als Untergruppe eines Zeichenvorrats $X$ in der Weise zu codieren, wie in der obigen Herleitung erläutert wurde.

Man sieht daraus, daß eine vollkommene Fehlerkorrektur bei endlicher Zeichenlänge $s$ — also mit endlichem Aufwand für Codierung und Decodierung — unmöglich ist.

In der Praxis legt man deshalb die Redundanz $R_{\text{ist}}$ derart fest, daß nur eine bestimmte Höchstfehlerzahl $f_{\text{grenz}} < s$ pro Zeichen, also ein bestimmter Augenblickswert $I_{x_i|y_k}$ noch beherrscht wird. Damit wird

$$R_{\text{ist}} \geqq \operatorname{ld} \sum_{f=0}^{f_{\text{gr}}} \binom{s}{f} . \tag{36}$$

In der Regel wird $R_{\text{ist}} > H(X|Y)$ sein,

also

$$N_{T_{\text{ist}}} \leqq \frac{N_Q}{2^{R_{\text{ist}}}} < \frac{N_Q}{2^{H(X|Y)}} = N_{T_{\max}} \text{ (für } s \to \infty). \tag{37}$$

Die Restfehlerwahrscheinlichkeit muß also in der Praxis stets $> 0$ bleiben.

Häufig wird für den zu fordernden Wert $f_{\text{grenz}}$ die Redundanz $R_{\text{ist}}$ bei Fehlerkorrektur noch immer unwirtschaftlich groß, und man fordert deshalb vom Empfänger nicht mehr, sämtliche Fehler auszumerzen, d. h. zu korrigieren. Man begnügt sich mit einem Code kleinerer Redundanz. Der kleinere redundante Entscheidungsgehalt je Zeichen wird so bemessen, daß er für den Empfänger noch ausreicht, um das Vorhandensein von maximal $f_{\text{grenz}}$ Fehlern zu erkennen, aber nicht mehr, um das richtige Herkunftszeichen zu bestimmen. Der Empfänger weist dann solche als fehlerhaft erkannten Zeichen zurück und läßt sie von der Quelle wiederholen. Man kann sagen: Der Empfänger fügt weitere Redundanz durch Wiederholungsanforderungen bei Bedarf ein. Die mit der praktischen Ausführung fehlerkorrigierender oder fehlererkennender Code zusammenhängenden Fragen werden in der Arbeit von *E. R. Berger* behandelt, die in einem nachfolgenden Heft dieser Zeitschrift erscheinen wird.

*Schrifttumsverzeichnis*

[1] Entwurf NTG 0102/1962: Informationstheorie, Begriffe. Nachrichtentechn. Z. 16 (1963), S. 46.

[2] *E. R. Berger* und *H. Piloty:* Erläuterungen zu den in der Informationstheorie verwendeten Begriffen der mathematischen Statistik. Nachrichtentechn. Z. 16 (1963), S. 49.

[3] *C. E. Shannon:* A mathematical theory of communication. Bell Syst. techn. J. 27 (1947), S. 379 ff. und 623 ff.

[4] *K. Küpfmüller:* Informationstheorie. Jahrbuch des Elektr. Fernmeldewesens (1954/55), S. 25 ff.

[5] *W. Endres:* Drittes Symposium über Informationstheorie. Nachrichtentechn. Z. 9 (1956), S. 261 ff.

Weitere Literatur siehe [2]. Außerdem werden, ohne Anspruch auf Vollständigkeit, folgende zur Einführung in das Gebiet geeignete Arbeiten in der Reihenfolge ihres Erscheinens genannt.

[6] *R. V. L. Hartley:* Transmission of information. Bell Syst. techn. J. 7 (1928), S. 535.

[7] *C. E. Shannon:* Neuere Entwicklungen in der Nachrichtentechnik. Techn. Mitt. PTT 28 (1950), S. 337.

[8] *R. Piloty:* Über die Beurteilung der Modulationssysteme mit Hilfe des nachrichtentheoretischen Begriffs der Kanalkapazität. Arch. elektr. Übertr. 4 (1950), S. 493.

[9] *F. A. Fischer:* Die Grundgedanken der modernen Theorie der Nachrichtenübertragung. Fernmelde-Ing. (1951), Heft 4.

[10] *K. Küpfmüller:* Die Entropie der deutschen Sprache. Fernmeldetechn. Z. 7 (1954), S. 265.

[11] *H. Frühauf:* Über die Grundprobleme der Theorie der Nachrichtenübertragung. Nachrichtentechn. 4 (1954), S. 49.

[12] *H. Frühauf:* Zum Stand der Informationstheorie. Nachrichtentechn. 6 (1956), S. 433 ff.

[13] *F. A. Fischer:* Informationstheorie, aus: Handbuch für Hochfrequenz- und Elektro-Techniker, Bd. IV, S. 25—44. Verlag Radio-Foto-Kinotechnik GmbH. Berlin-Borsigwalde.

[14] *P. Neidhardt:* Grundlagen und Anwendung der Informationstheorie in der Elektrotechnik. Dtsch. Elektrotechnik 1956, S. 411 ff. und S. 455 ff.

[15] *P. Neidhardt:* Einführung in die Informationstheorie. Verlag Technik/Berliner Union Stuttgart/Berlin 1957.

[16] *K. Küpfmüller:* Nachrichtenübertragung und Nachrichtenverarbeitung. Naturwiss. 48 (1961), S. 177.

(Eingeg.: 20. Mai 1963)