TECHNISCHE UNIVERSITÄT ILMENAU

# Shannon verstehen

Eine Erläuterung von C. Shannons mathematischer Theorie der Kommunikation

**Rüdiger Grimm**

**Nr. 15** **Dezember 2004**

# Diskussionsbeiträge

INSTITUT FÜR MEDIEN- UND KOMMUNIKATIONSWISSENSCHAFT

# Shannon verstehen

## Eine Erläuterung von C. Shannons mathematischer Theorie der Kommunikation

Rüdiger Grimm

Nr. 15 Dezember 2004

Herausgeber: Der Rektor der Technischen Universität Ilmenau
Redaktion: Institut für Medien- und Kommunikationswissenschaft,
Prof. Dr. Rüdiger Grimm
ISSN 1617-9048
Kontakt: Rüdiger Grimm, Tel.: +49 3677 69 47 35
E-Mail: ruediger.grimm@tu-ilmenau.de

# – Shannon verstehen –

# Eine Erläuterung von C. Shannons mathematischer Theorie der Kommunikation

R: Grimm / 19.11.2004

## *Vorbemerkung*

*Der hier vorgelegte Artikel „Shannon verstehen – Eine Erläuterung von C. Shannons mathematischer Theorie der Kommunikation“ ist Teil eines geplanten Lehrbuchs über „Digitale Kommunikation“ (2005). Es wird sich an Studierende, Wissenschaftler und interessierte Laien wenden, die sich für das Zusammenspiel von Kommunikation, Sprache und ihre informatorisch-technische Unterstützung interessieren. Das Lehrbuch wird Nicht-Informatikern erklären, wie das Internet funktioniert, und den Informatikern die Sprechakttheorie. Shannons mathematische Theorie der Kommunikation ist zwar ein Kernelement der Kommunikationsinformatik, wird aber von den geistes- und sozialwissenschaftlich orientierten Kommunikationswissenschaftlern fleißig rezipiert – und häufig missverstanden. Dieser Artikel soll helfen, Shannon richtig zu verstehen.*

*Diese Fassung ist ein „Diskussionsbeitrag“, der – bei aller Sorgfalt – keinen Anspruch auf Freiheit von Irrtümern und Fehlern erhebt. Er wird aufgrund weiterer Diskussion und Kritik weiterentwickelt und dabei verbessert werden.*

# 1 Gegenstand

Shannon (1948) und Shannon/Weaver (1949) behandeln in ihrer „Mathematischen Theorie der Kommunikation“ ein konstruktives Ingenieurs-Problem zur effizienten Übertragung von Zeichen über diskrete und kontinuierliche Kanäle mit und ohne Störung. Durch einen Bezug zum Informationsbegriff provozieren sie die Kommunikationswissenschaft zu einer Überinterpretation der Erklärungskraft dieser Theorie für den semantisch-pragmatischen Aspekt der Information.

Shannons Theorie behandelt drei Dinge:

1. Sie entwickelt mathematische Definitionen für fundamentale Begriffe der Übertragung von Zeichen. Darunter sind diskrete und kontinuierliche Übertragung, Kanalkapazität, Sende- und Empfangsrate, Störgeräuschrate, Korrekturrate, Redundanz der Kodierung, effiziente Kodierung, Kompression. Dafür werden Wörter verwendet wie Auswahl, Unsicherheit, Entropie und Information.

2. Sie leitet aus statistischen Eigenschaften der Datenströme mathematische Regeln für die effiziente Kodierung der Rohdaten für ihre Übertragung ab.

3. Sie zeigt, dass die Redundanz der Kodierung das Maß für die verlustfreie Kompressionsfähigkeit der Kodierung bzw. für die Robustheit gegenüber Störsignalen bildet. Sie interpretiert umgekehrt die Redundanzfreiheit der Kodierung als Maß für die Freiheit der Auswahl unter möglichen Zeichen (Choice) und für die Unsicherheit des Empfangs möglicher Zeichen (Uncertainty). Die Grundlage dazu bilden Wahrscheinlichkeitsverteilungen für das Auftreten von Zeichen (Entropy).

# 2 Hintergrund und Wirkung

Der Mathematiker Claude Elwood Shannon (1916-2001) hat im Jahre 1948 als Mitarbeiter der Bell Telephone Laboratories in New Jersey eine „mathematische Theorie der Kommunikation“ formuliert (Shannon 1948). Dieser Aufsatz erschien ein Jahr später zusammen mit einer Erläuterung des damaligen Direktors der Natural Science Foundation Warren Weaver, ebenfalls Mathematiker, in einem Buch unter fast demselben Titel: Aus „eine Theorie…“ wurde „*die* Theorie…“ (Shannon, Weaver 1949). Die Theorie erregte vom ersten Tag an bis heute großes Aufsehen in der wissenschaftlichen Öffentlichkeit, auch außerhalb der Mathematik und Informatik, besonders in der Kommunikationswissenschaft und Philosophie (erstmals Schramm 1954, in jüngerer Zeit z.B. Wagenknecht 1995, Capurro 2001). Sie wird heute sozusagen als ein „kanonischer Text einer Theorie der Information“ (Capurro 2001) angesehen und fehlt in keinem Lehrbuch über Kommunikationstheorien, etwa knapp und treffend in (Noelle-Neumann/Schulz/Wilke 2004), ausführlich in (Kloock/Spahr 2000, mit einigen Mängeln), und kritisch, aber nur modellmäßig in (Bentele/Beck 1994) und (Merten 1999).

Die von Shannon vorgelegte Theorie entwickelt Regeln für effizientes Kodieren diskreter und analoger Signale auf Übertragungskanälen mit und ohne Störung. Sie beweist mit mathematischen Mitteln, dass diese Regeln optimal sind: Aufgrund statistischer Analysen des Input- und Outputstroms können effiziente Kodierungsregeln für die Symbole und Signale spezifiziert werden, die aus physikalisch-mathematischen Gründen nicht verbessert werden können. In diesem Sinne ist Shannons Kommunikationstheorie abgeschlossen. Das allein ist eine wissenschaftliche Leistung ersten Ranges. Aber Shannons Beitrag reicht noch weiter.

Der Aufsatz ist brillant geschrieben, sowohl mathematisch, als auch sprachlich. Er enthält Beispiele und graphischen Erläuterungen. Die abstrakten Aussagen werden interpretiert und sowohl auf diskrete Übertragung von Morse-Telegraf und Fernschreiber (5-Bit-Code, ohne Einschränkung auf ASCII-8-Code erweiterbar), als auch auf analoge Radio- und Fernsehsignale angewendet. Dennoch ist dieser Aufsatz keine leichte Lektüre, er beruht auf glasharter Mathematik. Ich bezweifle, dass alle, die Shannon zitieren, ihn auch wirklich verstanden oder überhaupt gelesen haben. Denn neben seinen nicht misszuverstehenden ingenieurwissenschaftlichen Erkenntnissen auf dem Gebiet der Kodierung von Symbolen und Signalen unterliegt die philosophische Interpretation seiner Theorie manchem Missverständnis (nicht so Merten, der genau diese Beobachtung anstellt, Merten 1999, S. 74 und 151).

Es gibt zwei immer wiederkehrende Missverständnisse. Das eine beruht auf der Verknüpfung des Informationsbegriffs mit dem der Unsicherheit. Shannon meint mit der Unsicherheit die des Empfängers der zu erwartenden Botschaft:

> Can we find a measure of how much choice is involved in the selection of the event or of how uncertain we are of the outcome?
>
> (Shannon 1948, Abschnitt 6, Choice, Uncertainty and Entropy)

Dagegen beziehen manche Autoren die Unsicherheit auf die Wirkung einer Nachricht selbst, so etwa Capurro: „Information als Synonym von Ungewissheit“ und „Eine Botschaft bringt dem Empfänger etwas Neues oder Überraschendes, verursacht dadurch Ungewissheit“ in (Capurro 2001, S.2 u. 5).

Das andere Missverständnis liegt darin, dass Shannon niemals einzelnen Symbolen oder Nachrichten, sondern nur dem Informationskanal insgesamt einen Entropiewert zuweist, und zwar aufgrund der statistischen Eigenschaften aller in ihm möglichen Nachrichten. An keiner Stelle des originalen Papiers definieren Shannon/Weaver ein numerisches Informationsmaß für eine einzelne Botschaft, siehe etwa ausführlich (Shannon/Weaver 1949, S.14) und noch einmal

> The concept of information […] deals not with a single message but rather with the statistical character of a whole ensemble of messages.
>
> (Weaver in Shannon/Weaver 1949, Abschnitt 3.2, S. 27)

 © Grimm, TU Ilmenau, Dez 2004

Manche Interpreten nehmen dagegen fälschlicherweise an, einzelne Nachrichten trügen einen Entropiewert und damit einen Informationswert (z.B. Wagenknecht 1995; Kümmel wiederholt in Kloock/Spahr 2001, S. 216, 225; wikipedia.de 2004 gravierend unter dem Stichwort „Entropie“). Das ist nicht der Fall. Das wäre in Bezug auf den semantischen und pragmatischen Aspekt von Information unsinnig, und es ist sogar in Bezug auf den syntaktischen Aspekt nach Shannon inhaltsleer. Ein solches numerisches Maß hätte keinen weiteren Nutzen für die Kodierungstheorie, aus diesem Grunde hütet Shannon sich davor. Obwohl für einen bestimmten Informationskanal der Empfang des Buchstabens „e“ eine höhere Wahrscheinlichkeit hat als der des Buchstabens „x“, wie das etwa in der deutschen Sprache der Fall ist, unterscheiden sie sich nach Shannon nicht in ihrem Informationswert, der sich allein auf die Entropie aller möglichen Datenströme einschließlich dessen bezieht, dem sie beide entstammen; in diesem Beispiel also auf die gesamte deutsche Sprache.

Im Sinne von Bentele und Bystina (1978, S. 96 ff.) drückt die Entropie die „potentielle Information“ der Informationsquelle aus, aber nicht (niemals) die „aktuelle Information“, die eine individuelle Botschaft bei einem Empfänger erzeugt.

Für Überinterpretation bis hin zu Missverständnissen haben Shannon und Weaver durch den Hinweis, eine Theorie der Information geschaffen zu haben, vielleicht selbst gesorgt. Zwar lassen die Autoren keinen Zweifel daran, dass die von Shannon vorgelegte Theorie den Informationsbegriff auf die „blinde“ Auswahlfreiheit unter mehreren möglichen Ereignissen beschränkt. Shannon ist sich dieses Unterschiedes bewusst und grenzt sich gleich am Anfang klar ab:

> The fundamental problem of communication is that of reproducing at one point either exactly or approximately a message selected at another point. Frequently the messages have *meaning*; that is they refer to or are correlated according to some system with certain physical or conceptual entities. These semantic aspects of communication are irrelevant to the engineering problem. The significant aspect is that the actual message is one selected from a set of possible messages. The system must be designed to operate for each possible selection, not just the one which will actually be chosen since this is unknown at the time of design.
>
> (Shannon 1948, Introduction)

Fundamental sind die Aussagen: „These semantic aspects of communication are irrelevant to the engineering problem”, und: “The system must be designed to operate for each possible selection", mit anderen Worten, Shannon löst ein ingenieurspraktisches Problem zum Design eines Übertragungsapparates. So beschreibt es Weaver in (Shannon/Weaver 1949, Abschnitte 1.2 und 1.3, S. 4 ff.), wo er die drei semiotischen Ebenen Syntax, Semantik und Pragmatik voneinander unterscheidet und die Theorie von Shannon ausdrücklich auf die syntaktische Ebene beschränkt:

> The mathematical theory of the engineering aspects of communication […] admittedly applies in the first instance only to problem A, namely, the technical problem of accuracy and of transference of various types of signals from sender to receiver.
>
> Weaver in (Shannon/Weaver 1949, Abschnitt 1.3, S. 6); aber man beachte die Provokation "in the first instance"!

Das hätte die Welt außerhalb der Ingenieurslaboratorien kaum bewegt. Sie bewegt, dass Shannon und Weaver aufgrund ihrer Erkenntnisse eine provokative Interpretation des Phänomens der Information wagen. Und diese ragt in die Semantik von Sprache hinein. Unter „Information" verstehen der Philosoph und der Kommunikationswissenschaftler mehr als ein Maß der Redundanz seiner syntaktischen Kodierung. Information ist für sie ein semantischer Begriff und spricht die Bedeutung der übertragenen Symbole und Signale an, die im jeweiligen Handlungskontext der kommunizierenden Personen pragmatische Wirkung erzielt.

Und genau diese Provokation unternimmt Weaver in (Shannon/Weaver 1949, Abschnitt 3, S. 24 ff.), indem er die Forschung dazu anregt, Shannon's Theorie auf die semiotischen Ebenen der Semantik und Pragmatik auszudehnen. Shannon selbst enthält sich dieser Provokation zwar explizit, aber er impliziert sie doch mit seiner Wortwahl „Entropie" statt „Optimale Kodierungsrate", „Equivocation" statt „Korrekturrate" und „Information" statt „Botschaft" oder „Zeichenfolge".

Gleichwohl halte ich die Provokation für fruchtbar, da sie uns zwingt, die Grenze zwischen der syntaktischen und semantischen Natur von Sprache zu studieren. Meines Erachtens ist folgende Beschreibung des semantischen Informationsbegriffs zulässig:

> Eine *Botschaft* ist eine Zeichenfolge, die für Sender und Empfänger eine (im Allgemeinen nicht deckungsgleiche) semantische Bedeutung hat.
>
> Eine *Information im semantisch-pragmatischen Sinn* ist eine Botschaft, die das Wissen des Empfängers erhöht und die auf Seiten des Senders mit einer geringen Emotionalität verbunden ist.

 © Grimm, TU Ilmenau, Dez 2004

Die „geringe Emotionalität des Senders“ berücksichtigt lediglich den Gebrauch des Wortes „Information“ in der Alltagssprache. Diese Beschreibung enthält eine Reihe vager Bestimmungen. Ein „Mehr“ und „Weniger“ von „Wissen“ und „Emotionalität“ ist schwer oder gar nicht numerisch fassbar, außer mit Mitteln von Tests und Laborversuchen, die notwendig einen relativen pragmatischen Kontext einbeziehen müssen. Zwar scheint intuitiv ein Zusammenhang zwischen „Unsicherheit über den Empfang der Botschaft“ und ein „Mehren von Wissen“ zu bestehen, der sogar noch durch eine *notwendige* Beziehung verstärkt wird: *Wenn* eine Botschaft keine Information im syntaktischer Hinsicht enthält (weil der Informationskanal determiniert ist), *dann* hat der Empfänger „nichts Neues“ erfahren und kann daher auch sein Wissen nicht aufgrund der Botschaft erhöhen. Das gilt aber selbst in dieser Schlussrichtung nur für den vollständig determinierten Informationskanal (Entropie exakt null), und er gilt überhaupt nicht in umgekehrter Schlussrichtung. Daher gibt Shannon’s Theorie nicht mehr als diese intuitive Analogie her.

Auf die Begriffe Information und Kommunikation im sozialwissenschaftlichen Sinne gehen (Bentele/Beck 1994) ein. Zum Informationsbegriff in der Politikwissenschaft siehe (Vowe 2004).

Es folgt eine Darstellung der wichtigsten Elemente der Shannon’schen Theorie.

# 3 Das Grundmodell und die Grundbegriffe

Shannon stellt sich die Aufgabe, eine Theorie der Übertragung analoger und diskreter Signale zu entwerfen, die praktisch dazu dient, effiziente Übertragungssysteme zu konstruieren. Dafür setzt er auf statistische Eigenschaften der Signale, die sinnvollerweise nur dann zum Tragen kommen, wenn die Signale nicht vorab bestimmt sind. Auf diese Art der „Unsicherheit“ bezieht Shannon den Begriff „Information“, nämlich auf die rechnerische Auswahlmöglichkeit von Nachrichtenströmen aus einer Menge vorgegebener Zeichensätze bzw. Frequenzbereiche.

> Can we define a quantity which will measure, in some sense, how much information is “produced” by such a process, or better, at what rate information is produced? […] Can we find a measure of how much “choice” is involved in the selection of the event or of how uncertain we are of the outcome?
>
> (Shannon 1948, Abschnitt 6, Choice, Uncertainty and Entropy)

„Units of Information“ in dem gefundenen Maß sind dabei rein syntaktisch die Zeichen eines Alphabets („The resulting units of information will be called natural units“, Shannon 1948, Introduction).

- Die „Units of Information“ der natürliche Sprache sind die Buchstaben und Satzzeichen, je nach Ausprägung 32 (Teletype), 128 (7-Bit-ASCII) oder 256 (ASCII) Zeichen.
- Die „Units of Information“ der reellen Zahlen in Dezimalbruchdarstellung sind die Ziffern 0-9 und das Komma.
- Die „Units of Information“ der Computer sind „Binary Digits“, sogenannte Bits. Sie gelten als „kleinste Informationseinheiten“, weil sie nur die Auswahl (sic!) zwischen zwei verschiedenen Zuständen zulassen. Bei weniger als zwei Zuständen, nämlich nur einem Zustand, gibt es keine Auswahlmöglichkeit mehr und daher auch keine Information.

Diese „Informationseinheiten“ haben nichts mit der inhaltlichen Bedeutung der aus ihnen zusammengesetzten Wörter und Sätze zu tun. Einen Bezug zur Semantik von Information schließt Shannon ausdrücklich und mit Recht aus:

> Frequently the messages have *meaning* […]. These semantic aspects of communication are irrelevant to the engineering problem.
>
> (Shannon 1948, Introduction)

Also ein „Engineering Problem”, ein Konstruktionsproblem für Übertragungssysteme ist das Ziel der Theorie. Und in der Tat wird jeder Begriff so definiert, dass er für das Konstruktionsproblem verwertbar ist. Das zu Grunde liegende Kommunikationsmodell ist von der Kommunikationswissenschaft übernommen und für die Erklärung rückbezüglicher, pragmatischer und semantischer Aspekte der Kommunikation erweitert worden (etwa Merten 1977 u. 1999, Bentele/Beck 1994).


© Grimm, TU Ilmenau, Dez 2004

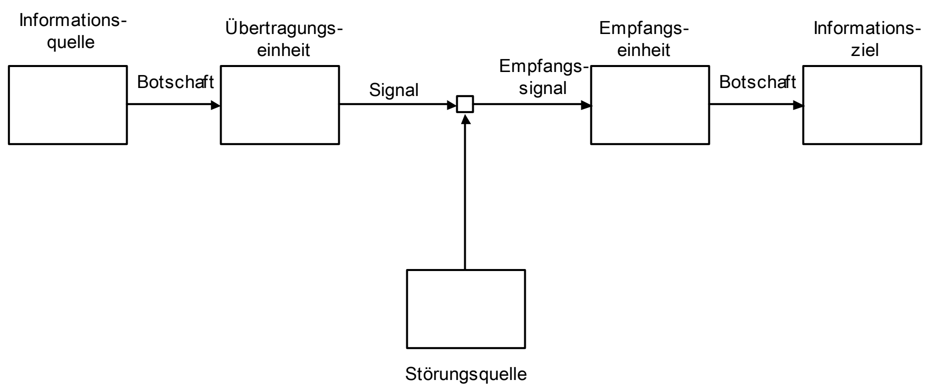


*Abbildung 1:* Schematisches Diagramm eines allgemeinen Kommunikationssystems im Sinne von Shannon (Shannon 1948, Introduction).

Die Informationsquelle produziert Signale in analoger oder diskreter Form.

- Beispiele für analoge Signale: Radio, Fernsehen, und Mobilfunk senden elektromagnetische Wellen in Frequenzbereichen (weit) unterhalb des sichtbaren Lichts. Analoges Telefon sendet Schwachstrom. Klingeln, Hupen, Musikinstrumente und die menschliche Stimme senden Schallwellen.

- Beispiele für diskrete Signale: ISDN-Telefon sendet Schwachstrom in Portionen von 64 Bits pro Sekunde. Leuchttürme senden Lichtsignale in Sequenzen von Lang- und Kurzportionen (auf dem Medium analoger Lichtwellen). Geschriebene Sprache besteht aus einzelnen Buchstaben und Satzzeichen. Das Internet überträgt Bits. E-Mails und Webseiten bestehen aus ASCII-Zeichen, das sind 8-Bit-Sequenzen, sogenannte Bytes.

Die Übertragung analoger Signale kann mathematisch-technisch auf die Übertragung diskreter Signale zurückgeführt werden. Beide Übertragungsformen können mit Hilfe von Transformationsfunktionen ineinander überführt werden (s.u. Fourier und Nyquist). Für die digitale Kommunikation ist die diskrete Datenübertragung der wichtigere Fall, da diese der Datenverarbeitung in Computern und der Kommunikation im Internet zu Grunde liegt.

Die Übertragungseinheit („Transmitter“) wandelt die Signale der Quelle in solche Signale um, die das Übertragungsmedium transportieren kann. Beim Telefon beispielsweise produziert der Sprecher als Informationsquelle stimmliche Laute, die das Telefongerät des Sprechers als Übertragungseinheit in analoge elektrische Impulse transformiert. Beim Internet transformiert der lokale Protokollautomat eines Computers in der Rolle der Übertragungseinheit den Datensatz einer E-Mail in die Bits und Bytes des Übertragungsprotokolls, die das Internet als Medium transportieren kann.

Auf Seiten des Empfängers „decodiert“ die Empfangseinheit („Receiver“) die Signale des Übertragungsmediums in solche Signale, die das Informationsziel weiterverarbeiten kann. Im Beispiel des Telefons ist die Empfangseinheit das Telefongerät mit dem Lautsprecher des Hörers und das Informationsziel ist der hörende Gesprächspartner, der die Signale seines Hörers als akustische Signale über sein Ohr aufnimmt. Im E-Mail-System des Internets ist die Empfangseinheit der Protokollautomat des empfangenden E-Mail-Service, und das Informationsziel ist die Mailbox des adressierten Nutzers.

Das Übertragungsmedium transportiert die Signale vom Sender zum Empfänger. Es kann einfach und direkt sein, wie im Gespräch von Angesicht zu Angesicht, bei dem der luftgefüllte und sichtbare Zwischenraum das Medium darstellt. Oder das Übertragungsmedium kann durch eine komplexe Organisation realisiert sein wie bei der Telefonie oder im Internet, bei der das Übertragungsmedium jeweils ein hochorganisiertes Netz aus Apparaten, Leitungen und Personen ist.


© Grimm, TU Ilmenau, Dez 2004

# 4 Zusammenhang zwischen analogen und diskreten Signalen (Fourier und Nyquist)

## 4.1 Fourier

Bereits im frühen 19. Jahrhundert hat Fourier gezeigt, dass sich alle „vernünftigen" kontinuierlichen Funktionen $g(t)$ als Summe von Sinus- und Kosinus-Funktionen schreiben lassen, die ihrerseits durch die diskreten Parameter Amplitude und Frequenz bestimmt sind:

(1) $$g(t) = \frac{1}{2}c + \sum_{n=1}^{\infty} a_n \sin(2\pi nft) + \sum_{n=1}^{\infty} b_n \cos(2\pi nft)$$

Der Wert $g(t)$ wird als Ausschlag der Welle zum Zeitpunkt $t$ interpretiert.

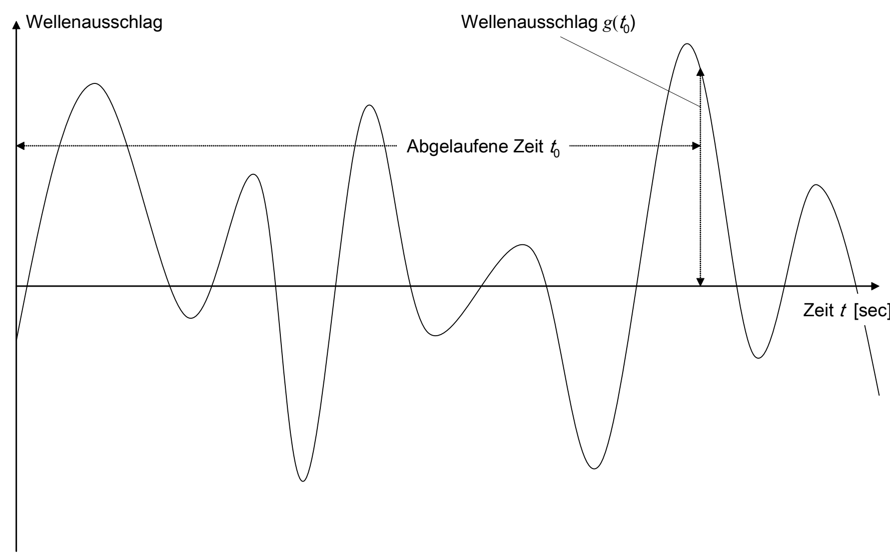


*Abbildung 2:* Graph einer Wellenfunktion „Zeit $t \rightarrow$ Ausschlag $g(t)$"

$f$ ist die Grundfrequenz (Anzahl Schwingungen pro Sekunde) aller Terme; $a_n$ und $b_n$ sind die Amplituden (maximale Ausschläge) der n-ten Terme $a_n \sin(2\pi nft)$ bzw. $b_n \cos(2\pi nft)$. Die Dauer $T$ einer Schwingung und ihre Frequenz sind über die Gleichung $f=1/T$ miteinander verbunden.

Bei Kenntnis der Periode $T$ und der Funktion $g(t)$ lassen sich die Parameter der Fourierterme bestimmen durch

$$(2) \qquad a_n = \frac{2}{T}\int_0^T g(t)\sin(2\pi nft)dt \quad b_n = \frac{2}{T}\int_0^T g(t)\cos(2\pi nft)dt \quad c = \frac{2}{T}\int_0^T g(t)dt$$

Umgekehrt kann jede kontinuierliche Funktion bei Kenntnis der Werte $c$, $a_1$, $b_1$, $a_2$, $b_2$,… aus der Fouriergleichung (1) rekonstruiert werden. D.h. jede kontinuierliche Funktion ist als Menge diskreter Werte bestimmbar. Bei einer festgelegten Obergrenze $W$ für Frequenzen (Tiefpassfilter) ist $a_n$=$b_n$=0 für alle $n>W/f$ und daher die Menge der Fourierparameter endlich. Aber auch ohne eine absolute Frequenzobergrenze kommt man mit endlich vielen Fourierparametern aus, indem man mit endlichen vielen Fouriertermen bereits eine gute Approximation von $g(t)$ erreicht.

Genauer siehe etwa bei Tanenbaum (2003, „The Physical Layer").

## 4.2 Nyquist und Shannon

Die Fourierparameter einer Funktion, die ein analoges Signal beschreibt, stellen alle Informationen dar, die das Signal in sich trägt, denn mit ihnen ist das Signal vollständig bestimmt. Die Fourierparameter stellen damit eine diskretisierte Datenübertragung dar, die dem Wellensignal entnommen werden. Bei einer Frequenzobergrenze ist diese Menge endlich. Jetzt ist die Frage, wie einer periodischen Funktion, die ein analoges Signal ausdrückt, die Fourierparameter entnommen werden können, mit anderen Worten, wie viele diskrete Informationen ein analoges Signal transportiert. Durch Nyquist (1924, 1928) und (wie unten ausgeführt) in von Shannon erweiterter Form kennen wir die maximale Übertragungsrate:

Wenn ein analoges ungestörtes Signal durch einen Tiefpassfilter auf maximale Frequenz $W$ beschränkt ist, dann kann das Signal durch genau $2W$ Abtastungen pro Sekunde vollständig rekonstruiert werden. Mehr Abtastungen liefern keine zusätzliche Information, da höhere Werte der Fourierreihe durch den Frequenzfilter verloren sind. Dieses ist zugleich die maximale Übertragungsrate des analogen Signals. Das gilt für zwei verschiedene Levels (immer dasselbe Maximum und die Null). Wenn das analoge Signal $L$ verschiedene Levels hat ($L$-1 verschiedene Maxima plus die Null), dann werden entsprechend $2WL$ diskrete Werte pro Sekunde übertragen. Umgerechnet auf eine binäre Darstellung der Werte lautet die Nyquist-Formel


© Grimm, TU Ilmenau, Dez 2004

(3) Nyquist: maximale Datenrate für einen analogen ungestörten Kanal ist

$2W \cdot \log_2 L$ Bit/sec

Der Logarithmus zur Basis 2 von $L$ besagt lediglich, wie viele Bits notwendig sind, um die ganze Zahl $L$ in binärer Form darzustellen. Ein Beispiel: Ein geräuschloses Signal mit seiner Frequenz von 3.000 Hz (Schwingungen pro Sekunde) und nur einem Amplitudenlevel ($L$=2) kann höchstens $2 \cdot 3.000 \cdot \log_2(2) = 6.000$ Bits/Sekunde übertragen.

Wie unten (Formeln 34-37) ausgeführt werden wird, hat Shannon diesen Satz für einen analogen Kanal mit Störrauschen verallgemeinert (Shannon 1948, Abschnitt 26):

Wenn ein analoges Signal gestört ist, dann kann es die Übertragungsrate eines ungestörten Signals nicht erreichen. Gleichwohl liefern die Amplituden der analogen Übertragung („Lautstärke" bei Schallwellen) eine Bandbreite von diskreten Werten. Shannon hat dafür eine Obergrenze angegeben. Wenn ein analoges Signal mit durchschnittlicher Sendeleistung von höchstens $S$ durch ein Störsignal mit durchschnittlicher Leistung $N$ überlagert wird und durch einen Tiefpassfilter auf maximale Frequenz $W$ beschränkt ist, dann gilt

(4) Shannon: maximale Datenrate für einen analogen gestörten Kanal ist

$W \cdot \log_2(1 + S/N)$ Bit/sec

Hierbei ist $S/N$ die „Lautstärke" des Signals über dem Störgeräusch („signal-to-noise ratio"), die industriell in der Regel nicht direkt, sondern in der Form $10 \cdot \log_{10}(S/N)$ in *Decibel dB* angegeben wird. Näheres hierzu siehe bei Tanenbaum (2003, „The Physical Layer"), wo auch folgendes erhellendes Beispiel steht:

Im klassischen analogen Telefonsystem mit einer Frequenz von 3.000 Hz (Schwingungen pro Sekunde) beträgt die durchschnittliche Übertragungslautstärke etwa 60 dB, das entspricht der „signal-to-noise ratio" $S/N$= 1.000.000, da $\log_{10}(10^6)=6$. Deshalb kann man über das analoge Telefon niemals mehr als $3.000 \cdot \log_2(1 + S/N) = 3.000 \cdot \log_2(1.000.001) \approx 3.000 \cdot 20 = 60.000$ Bit/Sekunde verlustfrei (!) übertragen, unabhängig davon, wie viele oder wie wenige verschiedene Maxima das Signal annimmt und wie oft das Signal abgetastet wird. Da das ein Ergebnis mathematischer und informationstechnischer Regeln ist, kann man Gegenbeispiele dazu wie Perpetuum-Mobile-Apparate behandeln: sie sind nicht möglich. Allerdings weist Tanenbaum mit

Recht darauf hin, dass auch diese Rate nur eine Obergrenze darstellt, die in der Praxis kaum erreicht wird. Bekanntlich schaffen Modems für das Analog-Telefon bequem 56 KBit/s.

# 5 Der diskrete Kanal ohne Störung

## *5.1 Kanalkapazität*

Unter der Übertragung diskreter Signale versteht man einen dynamischen Prozess, bei dem ein Sender in regelmäßigen oder unregelmäßigen Zeitabständen Signale aussendet, die der Empfänger im ungestörten Fall genauso aufnimmt. Alle für den Ingenieur interessanten Beispiele wie der altmodische Morse-Telegraf oder der Fernschreiber, aber auch das moderne E-Mail und das Web (Buchstaben, Satz- und Kontrollzeichen aus dem ASCII-Vorrat) arbeiten mit einem endlichen Zeichenvorrat, den man „Alphabet" nennt. Die Elemente des Alphabets nennt man Zeichen oder Symbole. Ein diskreter Kanal ist gekennzeichnet durch das Alphabet, dessen Zeichen er übertragen kann, und durch die Kapazität, in welcher Geschwindigkeit er Zeichen übertragen kann:

- Alphabet = Endliche Menge von Zeichen (auch Symbole genannt), die ein Kanal übertragen kann
- |Alphabet| = Länge des Alphabets = Anzahl der *verschiedenen* Zeichen, die der Kanal zur Übertragung akzeptiert
- Übertragungsrate = Anzahl übertragener Zeichen pro Sekunde. Sie wird in Symbole/s, oder im Falle der binären Übertragung, in Bit/s angebeben
- Kapazität = Maximale Übertragungsrate eines Übertragungsmediums, zum Beispiel bei ISDN-Telefonie 64.000 Bit/s

Falls ein Zeichen binär kodiert ist, wie bei ASCII, dann kann die Angabe Symbole/s in Bit/s umgerechnet werden über die Anzahl Bits, die für die Darstellung eines Symbols gebraucht wird. Da jedes ASCII-Zeichen in 8 Bits dargestellt wird, entspricht eine Übertragungsrate von $n$ ASCII-Zeichen pro Sekunde einer binären Übertragungsrate von $8n$ Bit/s. Jedes der 256 *verschiedenen* ASCII-Symbole wird durch jeweils $\log_2(256) = 8$ Bits ausgedrückt. So kommt es in (Shannon 1948, Abschnitt 1) zur Formel

(5) $$C = \lim_{T \to \infty} \frac{\log_2 N(T)}{T} \text{ Bit/s}$$

wobei $N(T)$ diejenige Anzahl *möglicher* Zeichen ist (bei jedem tatsächlich übertragenen ASCII-Zeichen waren insgesamt 256 Zeichen möglich gewesen), die pro Zeitintervall $T$ übertragen werden könnten. $N(T)$ ist also gleich $|\text{ASCII-Zeichensatz}|^{n(T)}$, wo $n(T)$ die Anzahl der tatsächlich übertragenen Zeichen pro $T$ sind. Shannon argumentiert (Abschnitt 1), dass die so definierte Kapazität in allen interessanten Fällen endlich ist und im Falle eines regelmäßigen Zeichenstroms in die einfache Formel (Anzahl Symbole/Sekunde)·(Anzahl Bits/Symbol) übergeht:

$\log_2(N)=\log_2(|\text{ASCII-Zeichensatz}|^n)=n\cdot\log_2(|\text{ASCII-Zeichensatz}|)=n\cdot\log_2(256)= 8n$

## 5.2 Kleiner Exkurs über den Logarithmus

$\log_{10}(n)$ ist der Logarithmus von der Zahl $n$ zur Basis 10 und gibt die Anzahl Dezimalziffern an, die erforderlich ist, um die Zahl $n$ darstellen zu können. $\text{Log}_{10}$(1 bis 9) liegt zwischen 0 und 1, also kann man die Zahlen 0-9 mit einer einzigen Ziffer darstellen, $\text{Log}_{10}$(10 bis 99) liegt zwischen 1 und 2, also kann man alle Zahlen 10-99 sicher mit zwei Dezimalziffern darstellen, usw.

$\log_2(n)$ ist der Logarithmus von der Zahl n zur Basis 2 und gibt die Anzahl Binärziffern (nur 0 und 1) an, die erforderlich ist, um die Zahl $n$ binär darstellen zu können. $\text{Log}_2(1)=0$, denn ein einziger möglicher Zustand braucht keine Darstellung. $\text{Log}_2(2)=1$, denn mit einem Bit 0 oder 1 kann man zwei verschiedene Zustände darstellen. $\text{Log}_2(4)=2$, denn mit 2 Bits kann man vier verschiedene Zustände darstellen, nämlich durch 00, 01, 10 und 11. Und so weiter. Die Zwischenwerte zwischen „glatten“ Logarithmuswerten werden entsprechend interpoliert, also gilt zum Beispiel $0=\log_2(1)<\log_2(2)=1<\log_2(3)<\log_2(4)=2<\log_2(5)<\log_2(6)<\log_2(7)<\log_2(8)=4<\log_2(9)$ … usw. Man kann das so interpretieren: zur Darstellung von $n$ verschiedenen Zeichen sucht man sich die nächst höhere Zweierpotenz $2^m \geq n$ und kann dann sicher sein, dass man mit $\log_2(2^m)=m$ Bits auskommt. Wenn $n$ selbst schon eine Zweierpotenz ist, dann schöpft man dabei alle Binärkombinationen aus, andernfalls bleiben einige Binärkombinationen unberücksichtigt.

$\ln = \log_e$ ist der „natürliche Logarithmus zur Basis $e$“ und spielt für die Kommunikationstheorie keine Rolle. Er wird hier deswegen nicht weiter behandelt.

Die Logarithmusfunktion $\log_a(x)$ und die Exponentialfunktion $a^x$ zur selben Basis $a$ sind immer invers zueinander: Für alle $x$ gilt $\log_a(a^x)=x$. Für alle $x>0$ gilt zusätzlich $a^{\log a(x)}=x$. Insbesondere gelten $\log_a(1)=0$ und $a^0=1$.

## 5.3 Wahrscheinlichkeitssteuerung des Übertragungsprozesses

Die Buchstaben in geschriebenen natürlichen Sprachen haben eine für ihre Sprache typische Häufigkeit. In der deutschen und englischen Sprache etwa ist der Buchstabe *e* am Häufigsten, *n* am Zweithäufigsten, während *j* und *q* relativ selten vorkommen.

Die Effizienz der Kodierung und die sich daraus ergebende Auslastung eines Kanals mit der höchst möglichen Übertragungsrate von Zeichen pro Sekunde orientiert sich an dem Gedanken, dass man einen möglichst knappen Bitvorrat wählt und den derart auf die zu übertragenden Zeichen verteilt, dass selten genutzte Zeichen eine aufwändigere Kodierung erhalten und häufiger genutzte Zeichen eine knappere. So verfährt etwa das Morsealphabet, das für das häufige *e* einen einfachen Punkt, für das ebenfalls häufige *n* einen Strichpunkt **–.** und für das seltene *q* die aufwändige Kombination **– –.–** festlegt. Um aber zu allgemein gültigen Aussagen über die Effizienz und Kapazitätsauslastung zu kommen, muss die Wahrscheinlichkeitsverteilung der Zeichen in den Übertragungsströmen untersucht und auf sinnvolle Weise beschrieben werden. „Sinnvoll" heißt dabei, dass jede Definition für weitere Berechnungen verwertbar ist.


© Grimm, TU Ilmenau, Dez 2004

| Buchstabe | Relative Häufigkeit | Position | Buchstabe | Relative Häufigkeit | Position |
|---|---|---|---|---|---|
| A | 6,51 | 6. | N | 9,78 | 2. |
| B | 1,89 | 16. | O | 2,51 | 15. |
| C | 3,06 | 12. | P | 0,79 | 21. |
| D | 5,08 | 8. | Q | 0,02 | 26. |
| E | 17,40 | 1. | R | 7,00 | 5. |
| F | 1,66 | 18. | S | 7,27 | 4. |
| G | 3,01 | 13. | T | 6,15 | 7. |
| H | 4,67 | 9. | U | 4,35 | 10. |
| I | 7.55 | 3. | V | 0,67 | 22. |
| J | 0,27 | 23. | W | 1,89 | 17. |
| K | 1,21 | 19. | X | 0,03 | 25. |
| L | 3,44 | 11. | Y | 0,04 | 24. |
| M | 2,53 | 14. | Z | 1,13 | 20. |

(Wikipedia.de 2004) „Das lateinische Alphabet“

Buchstaben nach Häufigkeit:
ENISRATDHULCGMOBWFKZPVJYXQ (Wikipedia.de)
ENRISTDHAULCGMOBZWFKVPJYQX (Fischer Lexikon Technik Bd. 4, 2004)

*Abbildung 3:* Wahrscheinlichkeitsverteilung der Buchstaben in der deutschen Sprache (Wikipedia 2004 und Fischer Lexikon, Technik, Bd.4).

| Buchstabe | Relative Häufigkeit | Position | Buchstabe | Relative Häufigkeit | Position |
|---|---|---|---|---|---|
| A | 6,73 | 7. | N | 7,18 | 5. |
| B | 1,13 | 20. | O | 8,28 | 3. |
| C | 3,02 | 11. | P | 2,16 | 17. |
| D | 2,94 | 12. | Q | 0,06 | 25. |
| E | 11,37 | 1. | R | 7,15 | 6. |
| F | 2,79 | 13. | S | 6,61 | 8. |
| G | 1,58 | 19. | T | 9,65 | 2. |
| H | 3,06 | 10. | U | 2,60 | 14. |
| I | 7,82 | 4. | V | 0,98 | 21. |
| J | 0,15 | 24. | W | 1,67 | 18. |
| K | 0,17 | 23. | X | 0,36 | 22. |
| L | 3,94 | 9. | Y | 2,29 | 15. |
| M | 2,18 | 16. | Z | 0,04 | 26. |

(Eigene Zählung 2004)

Buchstaben nach Häufigkeit:
ETOINRASLHCDFUYMPWGBVXKJQZ (eigene Zählung 2004)
ENRISTDHAULCGMOBZWFKVPJYQX (Mergenthaler 1884)

*Abbildung 4:* Wahrscheinlichkeitsverteilung der Buchstaben in der englischen Sprache (eigene Auszählung 2004 und Mergenthaler 1884).

Dazu stellt Shannon zunächst fest, dass die Zeichenströme der diskreten Kanäle, die uns interessieren, stochastische Prozesse vom ergodischen Markoff'schen Typ sind und deshalb die einschlägigen mathematischen Sätze gelten. Was bedeutet das?

*Stochastischer Prozess*: Serie von Ereignissen, die durch eine Wahrscheinlichkeitsfunktion gesteuert werden, d.h. für die eine Wahrscheinlichkeitsaussage getroffen werden kann. Beispiele:

- die zufällige Aussendung von 0 und 1 mit 0,5-Wahrscheinlichkeit jedes Ereignisses;
- in Melodien von Volksmusik werden mittlere Töne häufiger als hohe und tiefe Töne gespielt;
- in Zeichenfolgen von Sätzen natürlicher Schriftsprachen ist die Wahrscheinlichkeit des Auftretens jedes Buchstabens durch die relative Buchstabenhäufigkeit gegeben;
- Vollkommen vorher bestimmte Ereignisse sieht man nicht als stochastischen Prozess an.

Wenn an einer Stelle des Prozesses die Ereignisse $S_1$, $S_2$, …, $S_n$ möglich sind (z.B. an einer Satzstelle die Buchstaben A-Z oder a-z oder ein Satzzeichen), dann bezeichnet man die zugehörigen Wahrscheinlichkeiten mit $p(1)$, $p(2)$, ..., $p(n)$. Im Beispiel der deutschen Sprache wären etwa die Wahrscheinlichkeiten $p(\mathrm{e})\approx0{,}17$, $p(\mathrm{n})\approx0{,}10$ usw.

*Markoff'scher Prozess*: Ein stochastischer Prozess ist vom Markoff'schen Typ, wenn die Wahrscheinlichkeit eines Ereignisses von vorherigen Ereignissen abhängt. Beispiele:

- die zufällige Aussendung von 0 und 1 mit 0,5-Wahrscheinlichkeit jedes Ereignisses ist *kein* Markoff'scher Prozess, da jedes Ereignis unabhängig vom vorherigen Ereignis immer die Wahrscheinlichkeit 0,5 hat; Roulette und Lotto dürfen ausdrücklich keine Markoff'schen Prozesse sein, d.h. selbst wenn die Kugel beim Roulette zehnmal auf Rot gefallen ist, so ist die Wahrscheinlichkeit, dass sie beim nächsten Wurf auf Schwarz fällt, wieder nur 0,5 (sogar etwas weniger wegen der Möglichkeit des Zero).
- Melodien von Volksmusik sind Markoff'sche Prozesse, indem bestimmte Folgen von aufeinander folgenden Tönen oder Intervallen (Sekunden, Terzen, Quarten usw.) bevorzugt werden;
- Zeichenfolgen von Sätzen natürlicher Sprachen sind Markoff'sche Prozesse, da bestimmte Buchstabenfolgen in Zweierkombinationen (Digramme), Dreierkombinationen (Trigramme), usw. bevorzugt werden. Beispielsweise kommt die Buchstabenfolge „di“ in der deutschen Sprache häufiger vor, als die Buchstabenfolge „qu“, während „xx“ fast niemals vorkommt.


© Grimm, TU Ilmenau, Dez 2004

Die Abhängigkeit des Auftretens eines Ereignisses $S_j$ an einer Stelle des Übertragungsstroms von einem davor liegenden Ereignis $S_i$ kann in zweierlei Hinsicht gemessen werden. Zum einen bezeichnet man die Wahrscheinlichkeit, dass die Ereignisse $S_i$ und $S_j$ hintereinander eintreten, mit $p(ij)$, oder ausführlicher als $p(S_iS_j)$. Die Wahrscheinlichkeit dagegen, dass $S_j$ eintritt, nachdem feststeht, dass unmittelbar vorher $S_i$ eingetreten war, schreibt man als $p_i(j)$ oder auch als $p(i \Rightarrow j)$. Zum Beispiel bezeichnet man die relative Häufigkeit der Digramme „an", „qu" und „xx" in der deutschen Schriftsprache mit $p(\text{an})$, $p(\text{qu})$ bzw. $p(\text{xx})$. Die Wahrscheinlichkeit dagegen, dass ein n auftritt, nachdem a bereits sicher aufgetreten ist (bzw. u, wenn q bereits aufgetreten ist, bzw. x, wenn x bereits aufgetreten ist), bezeichnet man mit $p_a(n)$, $p_q(u)$, $p_x(x)$, oder mit $p(a \Rightarrow n)$, $p(q \Rightarrow u)$, $p(x \Rightarrow x)$. Diese beiden Wahrscheinlichkeiten sind durchaus verschieden! Zum Beispiel ist „qu" ein nicht besonders häufiges Digramm der deutschen Schriftsprache, aber wenn „q" bereits aufgetreten ist, dann ist die Wahrscheinlichkeit, dass nun ein „u" kommt, fast 100 Prozent. Allgemein gilt

$$(6) \qquad p(ij)=p(i)p_i(j)$$

*Ergodischer Prozess*: Ein stochastischer Prozess ist ergodisch, wenn die Wahrscheinlichkeitsverteilung der einzelnen Ereignisse bei zunehmender Länge der Ereignisfolgen stabile Grenzwerte annimmt. Es ist eine Art „Regel der großen Zahl": je größer eine Ereignissequenz, desto näher an die Grenzwerte kommen die darin gemessenen relativen Häufigkeiten der Ereignisse. Beispiele:

- Melodien von Volksmusik sind ergodische Markoff'sche Prozesse, da die typischen Tonfolgen in längeren Teilstücken immer dominanter werden.
- Zeichenfolgen von Sätzen natürlicher Sprachen sind ergodische Markoff'sche Prozesse; je länger ein Beispieltext gewählt wird, desto näher kommen die konkret ausgezählten relativen Häufigkeiten von Buchstaben, Digrammen und Trigrammen usw. an stabile Werte, die die Wahrscheinlichkeiten repräsentieren. Das ist unabhängig vom Texttyp, so lange man in der natürlich geschriebenen Sprachumgebung bleibt. Ernst Jandls Klangedichte führen zu anderen Häufigkeiten.
- Texte, in denen häufig die Sprache gewechselt wird, sind zwar Markoff'sche Prozesse, aber nicht ergodisch. Bei jedem Sprachwechsel beginnt eine neue Wahrscheinlichkeitsverteilung.

Shannon untersucht *diskrete Kanäle als stochastische Prozesse vom ergodischen Markoff'schen Typ*. Damit erfasst er alle interessanten Fälle der Praxis, wie die Datenübertragung von natürlicher Sprache und die Übertragungsprotokolle des Internets mit ihren Inhalts- und Steuerzeichen.

## 5.4 Effektive Kodierung

Das Ziel einer effektiven Kodierung richtet Shannon in seinem grundlegenden Abschnitt 6 über „Choice, Uncertainty and Entropy" an der Wahrscheinlichkeit des Auftretens einzelner Symbole aus. Dabei beobachtet Shannon zwei gegenläufige Tendenzen, die miteinander in Einklang zu bringen sind.

- Erstens die Trennschärfe von Zeichen mit Hilfe reichhaltigen Codes: Um alle Symbole eines Alphabets trennscharf voneinander zu bezeichnen, muss jedes Symbol ein eigenes Bitmuster haben. Dafür braucht man genügend viele Bits pro Symbol. Zum Beispiel, um 100 verschiedene Symbole zu kodieren, kommt man mit 7 Bits pro Symbol ganz gut aus (28 Kombinationen blieben dabei unberücksichtigt), aber auch eine Bitlänge von 8, 9 oder 10 (und mehr) Bits wäre gut geeignet, und hätte sogar den Vorteil, dass für zukünftige Erweiterungen des Zeichensatzes noch viel Platz wäre. Dabei würden nämlich viele Kombinationen unberücksichtigt bleiben, d.h. man hätte ziemlich viel Redundanz im Code.
- Zweitens die Einsparung von Code zur effektiven Nutzung des Übertragungskanals: Je weniger Bits man für ein Zeichen braucht, desto mehr Zeichen kann man auf einem Kanal übertragen, wenn man realistischer Weise davon ausgeht, dass alle Übertragungskanäle endliche Kapazitäten haben. Wenn also zum Beispiel ein Kanal maximal 60.000 Bits pro Sekunde übertragen kann, dann würde er bei einer Kodierung mit 7-Bits/Zeichen 8.571 Zeichen pro Sekunde übertragen, aber bei einer großzügigeren Kodierung von 8, 9 oder 10 Bits/Zeichen nur 7.500, 6.666 bzw. 6.000 Zeichen pro Sekunde. Wollte man zur Erhöhung der Übertragungsrate von Zeichen zu ihrer Kodierung nur noch 6 Bits/Zeichen verwenden, dann erreichte man zwar tatsächlich 10.000 Zeichen/Sekunde, aber es könnten nicht mehr alle Zeichen trennscharf voneinander unterschieden werden: mehrere Zeichen müssten denselben 6-Bits-Code miteinander teilen. Ein berühmtes Beispiel für die Verknappung von Code auf Kosten der Trennschärfe ist der „MPEG Audio Layer 3"-Kompressionscode für Audio-Signale, bekannt als MP3 (MPEG 1992/1994).

 © Grimm, TU Ilmenau, Dez 2004

Shannons Ziel ist es, das Optimum dieser beiden Tendenzen zu bestimmen: Präzise Trennschärfe *und* maximale Auslastung der Übertragungskapazität. Einen Code, der diese beiden Eigenschaften erfüllt, nennt man einen *effektiven Code*.

Der Ansatzpunkt zur Bestimmung der effektiven Kodierung ist die Wahrscheinlichkeit des Auftretens der Zeichen. Shannon geht bei der Begriffsbestimmung in zwei Schritten vor.

1. Zunächst definiert er eine Wahrscheinlichkeitsfunktion eines *einzelnen Ereignisses* aus der Menge der an dieser Stelle möglichen Ereignisse. Ein einzelnes Ereignis ist zum Beispiel das einmalige Aussenden eines Buchstabens, der aus einem Alphabet von sechsundzwanzig Buchstaben ausgewählt wird. Diese Wahrscheinlichkeitsfunktion nennt er „Entropie eines Ereignisses" (Abschnitt 6).

2. Danach definiert Shannon die Wahrscheinlichkeitsverteilung einer *Folge von Ereignissen* als eine nach den Wahrscheinlichkeiten der bisher erreichten Zustände gewichteten Summe der Entropien der dann jeweils folgenden Einzelereignisse. Die Ereignisfolge nennt Shannon eine „Informationsquelle" und nennt ihre Wahrscheinlichkeitsverteilung die „Entropie der Informationsquelle"(Abschnitt 7).

Aus der Entropie einer Informationsquelle leitet Shannon die effektive Kodierung zur maximalen Auslastung einer Kanalkapazität ab (Abschnitt 9) und veranschaulicht sein Vorgehen mit Beispielen (Abschnitt 10).

Warum nun gerade der Begriff Entropie gewählt wurde, der ja aus der Physik als Maß der Unordnung und als Richtungsgeber der Zeit bekannt ist, hängt mit der Form der Entropieformel und der Rolle der Wahrscheinlichkeit zusammen. In Folge dieser Analogie kommt es dann noch zu einigen anderen philosophischen Spekulationen, die aber nicht mehr Teil der Shannon'schen Theorie sind. Die Shannon'sche Entropie gibt in Bits/Symbol an, wie viele Bits ein Zeichen im Durchschnitt kostet, wenn der Zeichensatz passgenau für diese Informationsquelle effizient kodiert ist: das ist die minimale Bitlänge, die idealer Weise für diesen Zeichensatz pro Zeichen durchschnittlich zu erreichen ist.

© Grimm, TU Ilmenau, Dez 2004 

## 5.5 Wahlmöglichkeit, Unsicherheit und Entropie

Die effektive Kodierung hängt direkt von der Wahrscheinlichkeit des Auftretens eines Zeichens aus einer Menge möglicher Zeichen ab: Je häufiger ein Zeichen auftritt, desto weniger Bits sollte es kosten. Die Wahrscheinlichkeit eines Auftretens kann interpretiert werden als die Wahlfreiheit des (unbefangenen) Senders und ebenso als die Unsicherheit des Empfängers über das zu erwartende Ereignis.

Beispiel zum Verständnis der Wahlfreiheit des Senders: Wegen des ergodischen Charakters der deutschen Schriftsprache wird ein (unbefangener) Sprecher eines deutschen Satzes sich den Wahrscheinlichkeiten der Buchstaben im Satz annähern, und zwar umso mehr, je mehr er spricht. Er kann sozusagen gar nicht anders, indem er als „Informationsquelle“ die deutsche Sprache verwendet. Dass einzelne Sätze eine andere Häufigkeit enthalten können und dass sogar ein bewusster Sprecher absichtlich unwahrscheinliche Buchstabenfolgen spricht, etwa Karl Valentins „Wrdlbrmpf“, bleibt dabei unberücksichtigt, denn es würde in einem Meer von natürlichen Sätzen statistisch irrelevant bleiben. Es handelt sich also um die effektive Kodierung einer Informationsquelle insgesamt, eines Sendesystems, und nicht einzelner Botschaften!

Auf der Seite des Empfängers kann die Wahrscheinlichkeit des Auftretens eines Zeichens als die Unsicherheit darüber interpretiert werden, welches Zeichen wirklich kommt. Die Wahrscheinlichkeit liefert ein Maß dieser Unsicherheit, wie man das ja aus allen anderen Anwendungen der Wahrscheinlichkeitsrechnung gewohnt ist.

Beispiel für die Unsicherheit des Empfängers: Der Hörer eines deutschen Satzes erwartet als erstes Ereignis einen Buchstaben, und indem er die Wahrscheinlichkeiten der Buchstabenhäufigkeiten kennt, kann er diesem unsicheren Ereignis einen Wahrscheinlichkeitswert zuordnen: „Dass jetzt ein „e“ kommt, hat die Wahrscheinlichkeit 0,17, dass ein „n“ kommt, hat die Wahrscheinlichkeit 0,10 usw.“

In diesem Sinne ist die Wahrscheinlichkeit des Auftretens eines Ereignisses ein Maß für die Wahlfreiheit (des Senders) und der Unsicherheit (des Empfängers). Das ist strikt zu unterscheiden von der semantische Freiheit des Sprechers zu sagen, was er will, und den emotionalen Unsicherheiten des Hörers über eventuell bang erwartete Nachrichten.


© Grimm, TU Ilmenau, Dez 2004

Jetzt ist die entscheidende Frage, an der Shannon sein Genie zeigt, wie denn nun dieses Wahrscheinlichkeitsmaß für die Wahlfreiheit und Unsicherheit (was dasselbe Maß ist) genau formuliert werden soll, mathematisch gesprochen: welche Maßformel $H(p_1,\ldots p_n)$ für die Wahlfreiheit (bzw. Unsicherheit) über die Ereignisse $S_1,\ldots S_\mathrm{n}$ denn nun dazu geeignet ist, eine effiziente Kodierung zu bestimmen.

Interessanteweise argumentiert Shannon intuitiv in der Gedankenwelt der Wahlfreiheit und Unsicherheit und nicht mit Blick auf das zu erzielende Ergebnis. Dadurch bekommt sein Ergebnis eine philosophische Tiefe, die zeigt, wie die effiziente Kodierung mit der Wahrscheinlichkeitsverteilung *tatsächlich* zusammenhängt. Er argumentiert so:

Eine Formel $H(p_1,\ldots p_n)$, die die Wahlfreiheit des Senders und, in gleicher Weise, die Unsicherheit des Empfängers eines Ereignisses (etwa eines gesendeten und gelesenen Buchstabens) beschreibt, hängt von der Wahrscheinlichkeitsverteilung $p_1,\ldots p_n$ für das Eintreten eines der Ereignisse $S_1,\ldots S_\mathrm{n}$ ab und sollte dabei drei Eigenschaften haben:

1. Sie sollte stetig von den Argumenten $p_1,\ldots p_n$ abhängen, d.h. kleine Veränderungen der $\mathrm{p}_i$ sollten auch nur kleine Veränderungen des Maßes $H$ bewirken und keine Sprünge.

2. Wenn alle Ereignisse gleich wahrscheinlich sind, dann sollte das Maß $H$ monoton von der Anzahl der Ereignisse abhängen: je mehr Ereignisse, desto höhere Wahlfreiheit bzw. Unsicherheit.

3. Wenn eine Untermenge möglicher Ereignisse eine gemeinsame akkumulierte Wahrscheinlichkeit des Eintretens gegenüber allen anderen Ereignissen hat, und wenn zusätzlich innerhalb der Untermenge die Einzelwahrscheinlichkeiten gemessen werden können, dann sollte das zugehörige Wahlfreiheits- bzw. Unsicherheitsmaß $H$ die nach den Einzelwahrscheinlichkeiten gewichtete Summe der einzelnen $H$-Werte sein. Beispiel: Wenn unter drei Buchstaben A, B und C die Eintrittswahrscheinlichkeit für A 1/2 ist und damit das Auftreten von „B oder C“ die Restwahrscheinlichkeit 1/2 hat, und wenn sich zusätzlich die Wahrscheinlichkeiten für B oder C unter der Voraussetzung, dass es nicht A ist, zu 2/3 für B und 1/3 für C ergeben, dann verteilen sich die Eintrittswahrscheinlichkeiten auf die einzelnen Buchstaben wie folgt: A zu 1/2, B zu 1/3 und C zu 1/6. Die Forderung an $H$ lautet dann so: $H(\frac{1}{2},\frac{1}{3},\frac{1}{6}) = \frac{1}{2}H(\frac{1}{2},\frac{1}{2}) + \frac{1}{2}H(\frac{2}{3},\frac{1}{3})$ (Beispiel aus Shannon 1948, Abschnitt 6).

Für diese drei Forderungen gibt es nur eine einzige mathematische Formel, und die lautet

(7) $$H(p_1,...,p_n) = -K\sum_{i=1}^{n} p_i \log p_i$$ (Shannon 1948, Abschnitt 6, Theorem 2)

wobei $K$ eine beliebige Konstante ungleich 0 darstellt. Da diese keine Aussagekraft hat, normiert sie Shannon gleich zu 1. Das Minuszeichen soll den Leser nicht erschrecken, es sorgt lediglich dafür, dass der Wert von $H$ immer positiv ist: da nämlich alle $p_i \leq 1$ sind, sind alle log $p_i \leq 0$, so dass das Minuszeichen die Gesamtsumme wieder positiv dreht.

Diese Formel elektrisiert den physikalisch vorgebildeten Leser, der darin eine Ähnlichkeit zur Boltzmann'sche Entropieformel für die Wahrscheinlichkeitsverteilung $W$ der Gasmoleküle in einem abgeschlossenen Gasbehälter wiederentdeckt:

$$\Delta S = -k \ln W$$ (Boltzmann-Formel für die Entropie von Gasen)

$k$ ist dabei eine durch Experimente festgestellte Naturkonstante, die so genannte „Boltzmann-Konstante". $S$ ist die Entropie und stellt eine Zustandsgröße in Abhängigkeit von Volumen und Temperatur des Gases dar. $\Delta S$ ist demnach die Differenz der Entropien des Gases in zwei Zuständen. Diese Entropieformel gilt bekanntlich allgemein für alle geschlossenen Systeme: „Die Entropie eines Zustandes ist proportional dem natürlichen Logarithmus seiner Wahrscheinlichkeit". Im Verbund mit dem 2. Hauptsatze der Thermodynamik, der besagt, dass ohne Zuführung von äußerer Energie, also in geschlossenen Systemen, die Entropie zunimmt, besagt die Boltzmann-Formel, dass Systeme von selbst vom unwahrscheinlichen Zustand in einen wahrscheinlicheren Zustand übergehen, niemals umgekehrt. Daran knüpfen sich verschiedene Spekulationen, etwa über die Richtung von Zeit und den Wärmetod des Weltalls. Diese Spekulationen sind hier nicht von Belang. Diese Zusammenhänge stehen in jedem Lehrbuch der Physik, ich habe sie zum Beispiel in einem Lehrbuch meiner Schulzeit (Kuhn 1976) noch einmal nachgelesen.

Wichtig für die Informationstheorie von Shannon ist die Assoziation der physikalischen Entropie mit dem Mangel an detaillierter Information über die innere Struktur des Systems, die über die Wahrscheinlichkeitssaussage kompensiert wird. Diese „Unsicherheit der Information", die sich im physikalischen Entropiebegriff verbirgt, zusammen mit der äußeren Ähnlichkeit der Formeln, waren Anlass für Shannon, das Maß der „Wahlfreiheit" und „Unsicherheit" einer ergodischen Informationsquelle „Entropie" zu nennen. Zu diesem Zweck wählt er aus pragmatischen Gründen

 © Grimm, TU Ilmenau, Dez 2004

den Logarithmus zur Basis 2 (der ja bekanntlich für die Anzahl erforderlicher Bits zur binären Darstellung des Arguments steht), und den frei wählbaren Koeffizienten K normiert er zu 1.

(8) Definition (Shannon): $H(p_1,...,p_n) = -\sum_{i=1}^{n} p_i \log_2 p_i$ bezeichnet die *Entropie einer Wahrscheinlichkeitsverteilung* $p_1,\ldots,p_n$ für ein einzelnes Ereignis aus einer Menge möglicher Ereignisse $S_1,\ldots,S_n$. Wenn dieses Ereignis durch eine Zufallsvariable $x$ beschrieben wird (zum Beispiel „ein Buchstabe"), dann schreibt man für $H$ auch $H(x)$ (zum Beispiel $H$(Buchstabe an $k$-ter Position eines Satzes)). $H$ wird in „Bits" angegeben.

Achtung: Bisher ist nur die Entropie der Möglichkeit eines einzelnen Ereignisses definiert worden und noch nicht einer Folge von möglichen Ereignissen. Das folgt im nächsten Abschnitt über Informationsquellen.

Shannon erläutert die Formel anhand eines einfachen Beispieles: ein Ereignis bestehe aus zwei möglichen Antworten „Ja" oder „Nein", wobei für das Auftreten von „Ja" die Wahrscheinlichkeit $p$ angenommen wird und für das Auftreten von „Nein" die Restwahrscheinlichkeit 1-$p$. In diesem Fall ist $H$ eine Funktion einer Variablen $p$ des Einheitsintervalls auf das Einheitsintervall, $H$:[0,1]→[0,1], mit $H(p)=p\log_2(p)-(1-p)\log_2(1-p)$ und kann in einem zweidimensionalen Graphen dargestellt werden.

$p$=0 bedeutet, dass „Ja" nicht eintreten kann und dass „Nein" sicher eintritt. Die Wahlmöglichkeit des Senders und entsprechend die Erwartungsunsicherheit des Empfängers sind in diesem Fall null. Falls im anderen Extrem „Ja" die Wahrscheinlichkeit $p$=1 hat und „Nein" die Restwahrscheinlichkeit 1-1=0, dann trifft „Ja" sicher ein und „Nein" ist ausgeschlossen. Wiederum hat der Sender keine Wahl und der Empfänger ist sicher darin, was er empfängt. Folglich ist in beiden Fällen die Entropie Null, was die Formel bestätigt: $H(1,0) = -1\cdot\log_2 1-0\cdot\log_2 0 = 0$ und ebenso $H(0,1)=-0\cdot\log_2 0-1\cdot\log_2 1 = 0$. (Anmerkung über „0·log0": gewiss ist „log0" nicht definiert, da $\lim_{x\to 0}\log x = \to -\infty$ ist, aber die log-Funktion wächst bei Annäherung des Argumentes an Null langsamer gegen -∞, als die lineare Funktion $x$ gegen 0, daher ist $\lim_{x\to 0}(x\log x) = 0$).

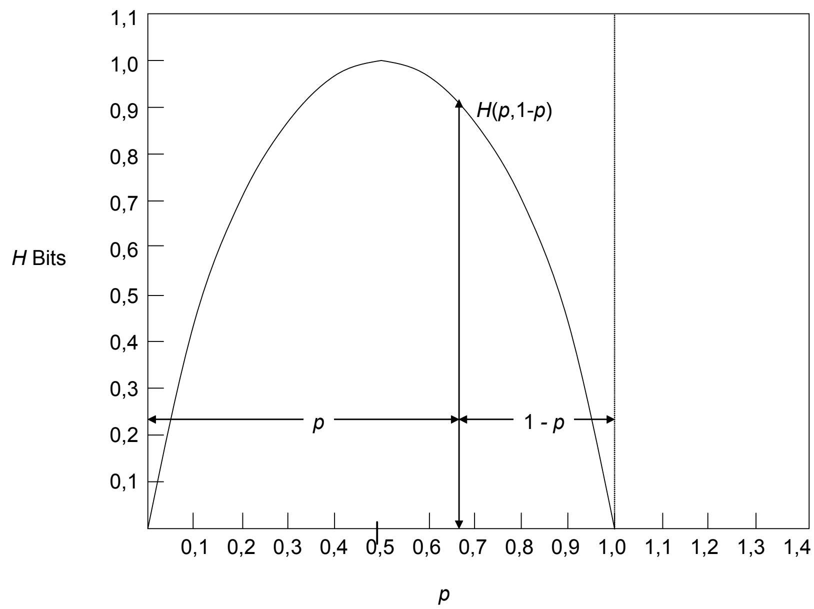


*Abbildung 5:* Entropie eines Ereignisses mit zwei Möglichkeiten, die mit den Wahrscheinlichkeiten $p$ bzw. 1-$p$ eintreten. (Shannon 1948, Abschnitt 6)

$p$=1/2 bedeutet, dass „Ja“ und „Nein“ beide gleich wahrscheinlich sind. In diesem Fall nimmt die Entropie ein Maximum bei $-1/2\cdot\log_2(1/2)-1/2\cdot\log_2(1/2) = 1/2\cdot\log_2(2)+1/2\cdot\log_2(2) = 1/2\cdot1+1/2\cdot1 =1$ an. Auch das entspricht der Intuition, dass in diesem Falle der Gleichwahrscheinlichkeit der Sender ein Maximum an Auswahlfreiheit und der Empfänger ein Maximum an Unsicherheit hat.

Diese drei diskutierten Spezialfälle gelten übrigens allgemein auch für mehrdimensionale Ereignisse: Wenn alle $p_1,\ldots,p_n$ bis auf ein $p_k$ Null sind, dann ist dieses eine $p_k$=1 sicher. Wie man leicht sieht, gilt dann $H$=0, da alle Summanden $p_i\log_2(p_i)$ null sind, denn alle $p_i$ bis auf das eine $k$ sind null, während bei diesem der Log-Wert null ist: $p_k\log_2(p_k) = 1\cdot\log_2(1) =1\cdot0=0$. Das ist konsequent, weil das Eintreten dieses einen $k$-ten Ereignisses sicher ist. Wenn dagegen alle $p_i$ gleichmäßig verteilt den Wert $1/n$ annehmen, dann sollte die Entropie einen positiven Maximalwert annehmen, da die Auswahl vollkommen frei und entsprechend die Empfangserwartung maximal unsicher ist, und in der Tat nimmt $H$ hier das Maximum $\log_2(n)$ an: $H(1/n,\ldots,1/n)=n\cdot(1/n\cdot\log_2(n)) = \log_2(n)$.

## 5.6 Entropie einer Informationsquelle

Um nun zu einer Folge von Ereignissen übergehen zu können, definiert Shannon (noch in Abschnitt 6) die *bedingten Entropien* von zwei aufeinander folgenden Ereignissen, die im Allgemeinen voneinander abhängen, zum Beispiel erst der Buchstabe q, dann der Buchstabe u. Wir unterstellen für beide Ereignisse dasselbe Alphabet (d.h. jedes Mal können dieselben Buchstaben auftreten) und nutzen die Wahrscheinlichkeiten für das einzelne Auftreten eines Zeichens und die beiden Wahrscheinlichkeiten für das Hintereinanderauftreten zweier Zeichen. $p(i)$ bezeichnet die Wahrscheinlichkeit des Ereignisses $S_i$. $p(ij)$ ist die Wahrscheinlichkeit für das Digramm $S_iS_j$ und $p_i(j)$ ist die Wahrscheinlichkeit dafür, dass $S_j$ auftritt, nachdem feststeht, dass $S_i$ schon aufgetreten ist. Es gilt immer $p(ij)=p(i)p_i(j)$.

Die Entropie des Ereignisses eines Digramms ist nach Definition (8) die gewichtete Summe der Wahrscheinlichkeiten der Digramme, also

(9) $$H(x,y) = -\sum_{i,j} p(i,j) \cdot \log_2 p(i,j)$$

Aus elementaren Zusammenhängen der bedingten Wahrscheinlichkeiten $p_i(j)$ mit den absoluten Wahrscheinlichkeiten $p(i)$ ergibt sich, dass die Entropie der Folge zweier Ereignisse nicht höher als die Summe der Einzelentropien der Ereignisse sein kann:

(10) $H(x,y) \leq H(x) + H(y)$ mit Gleichheit genau dann, wenn die Ereignisse unabhängig sind, d.h. wenn alle $p(j) = p(i \Rightarrow j)$.

Dieses Ergebnis bestätigt wiederum die gelungene Begriffswahl von $H$, da es die Entropie (Wahlfreiheit, Unsicherheit) intuitiv zum Ausdruck bringt: Abhängigkeiten der Ereignisse $x$ und $y$ erniedrigen ihre Auswahlfreiheit gegenüber der unabhängigen Wahl der Ereignisse. Die *bedingte Entropie* eines Ereignisses $y$ nach einem Ereignis $x$, $H_x(y)$, definiert Shannon entsprechend als den Durchschnitt der Entropien von $y$ für jeden Wert von $x$, gewichtet nach der Wahrscheinlichkeit, dass dieses spezielle $x$ eintritt:

(11) Definition: $H_x(y) = -\sum_{i,j} p_i \cdot (p_i(j) \cdot \log_2 p_i(j))$ ist die *bedingte Entropie* eines Ereignisses $y$ in Abhängigkeit des davor eintretenden Ereignisses $x$.

Wegen der Beziehung (6) $p_i \cdot p_i(j)=p(i,j)$, und dem Distributivgesetz ergibt das

(11') $H_x(y) = -\sum_{i,j} p(i,j) \cdot \log_2 p_i(j)$ für die *bedingte Entropie*.

Aus den Formeln leitet man unmittelbar ab, dass die bedingte Entropie von $y$ die Differenz zwischen der Digramm-Entropie für $(x,y)$ und der Entropie des ersten Ereignisses $x$ aus dem Digramm ist:

(12) $H(x,y) = H(x) + H_x(y)$, sowie $H(x,y) = H(y) + H_y(x)$

Aus (10) und (12) ergibt sich die intuitiv nachvollziehbare Ungleichung

(13) $H(y) \geq H_x(y)$

die besagt, dass die Kenntnis eines Ereignisses $x$ die Entropie (Unsicherheit) des folgenden Ereignisses erniedrigt, da ja die Wahrscheinlichkeitsbeziehung zwischen $x$ und $y$ einen gewissen Rückschluss auf das Ereignis $y$ zulässt, es sei denn, dass die beiden Ereignisse $x$ und $y$ unabhängig voneinander sind, in dem Falle sind die Entropien gleich:

(14) $H(y) = H_x(y)$ genau dann, wenn $x$ und $y$ unabhängige Ereignisse sind,
d.h. wenn alle $p(j) = p(i \Rightarrow j)$.

Eine *Informationsquelle* ist definiert als ergodischer Markoff'scher Prozess aus Ereignissen. Jedes stattgefundene Ereignis – z.B. nachdem ein Buchstabe ausgesendet wurde – hinterlässt das System – die Informationsquelle – in einem Zustand $S_i$, der dadurch gekennzeichnet ist, dass $i$ Ereignisse stattgefunden haben – die Buchstabenfolge $a_1, \ldots, a_i$ ausgesendet wurde – und das System dadurch in die Zustände $S_1, \ldots, S_i$ überführt hatte. Nun ist das System in Zustand $S_i$ und erwartet ein weiteres Ereignis – das Aussenden des nächsten Buchstabens. Das *Folge*ereignis ist nun ein einzelnes Ereignis und hat eine Entropie, die im Zustand $S_i$ mit $H_i$ bezeichnet wird. Sie ist nach Definition (8) die Summe der $p\log(p)$ aus der Wahrscheinlichkeitsverteilung des erwarteten Ereignisses – des Folgebuchstabens. Da aber das System in einem bestimmten Zustand ist – bestimmte Buchstaben bereits ausgesendet wurden –, handelt es sich bei den Wahrscheinlichkeiten um *bedingte* Wahrscheinlichkeiten in Abhängigkeit des bestehenden Zustands, die wir mit $p_i(j)$ bezeichnen. Das Folgeereignis im Zustand $S_i$ hat also die Entropie $H_i = \sum_j p_i(j) \log_2 p_i(j)$, wobei mit $j$ über alle möglichen Ereignisse an dieser Stelle (alle möglichen Buchstaben, die als *nächstes* gesendet werden können) summiert wird. Der Zustand $S_i$ hat seinerseits eine Wahrscheinlichkeit, dass er entstanden ist, die mit $P_i$ bezeichnet wird. Im Beispiel der Informationsquelle ist $P_i=p(a_1,\ldots,a_i)$ die Auftrittswahrscheinlichkeit der bisher gesendeten Buchstabenfolge $a_1,\ldots,a_i$.

> Definition: Die *Entropie der Informationsquelle* ist definiert als die Summe der Entropien der Einzelereignisse $H_i$ nach allen möglichen Zuständen $S_i$, gewichtet nach der Wahrscheinlichkeit $P_i$ des Auftretens des Zustandes $S_i$ an der $i$-ten Stelle:
>
> (15) $$H = \sum_i P_i H_i = -\sum_{i,j} P_i p_i(j) \log_2 p_i(j)$$

Falls die Folgeereignisse voneinander unabhängig sind, gilt $p_i(j)=p(j)$ und deshalb wie bei der Entropie eines einzelnen Ereignisses $H = -\sum_j p(j)\log_2 p(j)$ (s. Definition 8 oben), denn dann ist $\sum_{i,j} P_i p_i(j)\log_2 p_i(j) = \sum_{i,j} P_i p(j)\log_2 p(j) = (\sum_i P_i)(\sum_j p(j)\log_2 p(j)) = \sum_j p(j)\log_2 p(j)$, weil ja $\Sigma_i P_i=1$, d.h. die Wahrscheinlichkeiten aller möglichen gegenwärtigen Zustände zusammen immer 1 ergeben. In diesem Fall der voneinander unabhängigen Folgeereignisse ist übrigens die Entropie einer Nachrichtenquelle ungefähr gleich dem Logarithmus der umgekehrten Auftrittswahrscheinlichkeit $p$ einer typischen langen Botschaft (Zeichenfolge), geteilt durch die Länge $N$ der Botschaft, und zwar je länger und typischer, desto eher: $(1/N)\cdot\log_2(1/p)\rightarrow H$. (Shannon 1948, Abschnitt 7 über die Entropie einer Informationsquelle).

Die Entropie einer Informationsquelle bemisst sich in „Bits pro Symbol“ und gibt, wie wir sehen werden, die durchschnittliche Bitlänge eines Symbols an, wenn alle Symbole in Bezug auf diese Informationsquelle möglichst effektiv, d.h. trennscharf und ohne übrig bleibende Bitkombinationen so kodiert sind, dass seltenere Zeichen längere und häufigere Zeichen kürzere Bitkombinationen lang sind. Die Entropie gibt mithin die minimale Bitlänge an, die idealer Weise für diesen Zeichensatz pro Zeichen durchschnittlich zu erreichen ist.

Die Entropie einer Informationsquelle kann man selbstverständlich auch als zeitlichen Prozess beschreiben und in „Bits pro Sekunde“ angeben. Diese „dynamische Entropie“ bezeichnet Shannonn mit *H'* und definiert sie als

(16) $$H' = \sum_i f_i H_i \text{ Bit/s}$$

wobei $f_i$ die Frequenz des Ereigniswechsels vom *i*-ten auf das *i*+1-te Ereignis ist. Die beiden Entropieformate hängen über die durchschnittliche Übertragungsrate einer Informationsquelle von *m* Symbolen pro Sekunde zusammen:

(17) $$H' = mH \text{ Bit/s}$$

Wie misst man nun die Entropie einer Informationsquelle, die doch verschieden lange Zeichenketten produzieren kann? Hier kommt der ergodische Charakter der Informationsquelle zum Tragen: je länger eine Botschaft, desto „typischer“ ihr statistisches Verhalten. Shannon zeigt folgende Konvergenzen für Folgen von Zeichenketten (Shannon 1948, Abschnitt 7):

Einer Botschaft (Zeichenfoge) *B* der Länge *N* kann man eine Auftrittswahrscheinlichkeit *p*(*B*) zuordnen. Dann konvergiert die Folge $G_N := -\frac{1}{N}\sum_{|B|=N} p(B)\log p(B)$ mit wachsenden *N* monoton fallend gegen den Entropiewert *H* der Informationsquelle, aus der die *B* stammen: $\lim_{N\to\infty} G_N = H$.

Eine noch bessere Konvergenz wird erzielt, wenn man für die Wahrscheinlichkeitswerte der Botschaften die bedingten Wahrscheinlichkeiten nachfolgender Symbole berücksichtigt. Wenn man mit $p(B,S_j)$ die Wahrscheinlichkeit der Botschaft *B* gefolgt von dem Zeichen $S_j$ bezeichnet und mit $p_B(S_j) = p(B,S_j)/p(B)$ die bedingte Wahrscheinlichkeit von $S_j$, wenn davor sicher *B* aufgetreten ist, dann konvergiert die Folge der Zahlen $F_N := -\sum_{j,|B|=N} p(B,S_j)\log p_B(S_j)$ ebenfalls monoton fallend gegen *H*, und zwar schneller als $G_N$: $\lim_{N\to\infty} F_N = H$. Außerdem gilt: $H \le F_N \le G_N$ für alle *N*.

An dieser Stelle sei noch einmal in aller Deutlichkeit darauf hingewiesen, dass eine individuelle Botschaft (Zeichenfolge) *B* zwar eine Auftrittswahrscheinlichkeit *p*(*B*), aber keine Entropie *H* hat. Entropien kommen nur der Informationsquelle insgesamt zu.


© Grimm, TU Ilmenau, Dez 2004

## 5.7 Effizienz der Kodierung und Hauptsatz

Ein Zeichensatz erzielt bei völliger Gleichverteilung der Auftrittswahrscheinlichkeit seiner Zeichen eine maximale Entropie. Das Verhältnis der tatsächlichen (womöglich ungünstigen) zur völlig ausgeglichenen Wahrscheinlichkeitsverteilung einer Informationsquelle zeigt an, wie viel „Platz“ zur Kompression des Codes bei Erhalt der Trennschärfe noch übrig ist. Diesen nennt man Redundanz:

(18) Definition: Die *relative Entropie einer Informationsquelle* ist der Quotient aus der tatsächlichen Entropie geteilt durch die maximale Entropie desselben Zeichensatzes bei angenommener Gleichverteilung der Zeichenwahrscheinlichkeiten.

(19) Definition: Die *Redundanz einer Informationsquelle* ist definiert als „Eins minus ihrer relativen Entropie“.

Dass $H$ in der Tat die „beste Kodierung“ von Symbolen eines Zeichensatzes anzeigt, d.h. dass die „beste Kodierung“ des Zeichensatzes im Durchschnitt $H$ Bits pro Symbol kostet, beweist Shannon in seinem

(20) *Hauptsatz über den diskreten ungestörten Kanal*: Eine Informationsquelle mit Entropie $H$ Bits/Symbol kann durch einen vorgegebenen Kanal mit Kapazität $C$ Bits/Sekunde verlustfrei niemals schneller als mit der Übertragungsrate $C/H$ Symbole/Sekunde übertragen werden. Die Übertragungsrate $C/H$ kann durch geschickte Kodierung (Kompression zur Eliminierung der Redundanz) beliebig nahe erreicht werden. (Shannon 1948, Abschnitt 9, Theorem 9)

Der zweite Teil der Aussage wird konstruktiv bewiesen. Dazu dient beispielsweise das Fano’sche Kodierungsverfahren (Fano 1949): Man ordne die Zeichenfolgen einer festen Länge N nach sinkender Wahrscheinlichkeit. Diese geordnete Menge von Zeichenfolgen teile man in zwei Teilmengen derselben Wahrscheinlichkeit und gebe allen Folgen der oberen Teilmenge das Bit 0 als erstes Bit und allen Zeichenfolgen der unteren Teilmenge das Bit 1 als erstes Bit. Dann teile man jede der beiden Teilmengen wiederum in jeweils zwei Teilmengen der gleichen Wahrscheinlichkeit und gebe ihnen entsprechend als zweites Bit eine 0, wenn sie zum oberen Teil gehören und eine 1, wenn sie zum unteren Teil gehören. So fahre man fort, bis die Zeichenfolgen all einzeln als Bitmuster kodiert sind (Shannon 1948, Ende von Abschnitt 9).

Shannon führt in Abschnitt 10 ein erhellendes Beispiel aus: Ein Zeichensatz aus vier Buchstaben *A*, *B*, *C* und *D* würde man unbefangen mit zwei Bits gleichmäßig kodieren durch

| | |
|---|---|
| *A* | 00 |
| *B* | 01 |
| *C* | 10 |
| *D* | 11 |

Ein Kanal mit 300 Bit/s (wie die alten Akustikkoppler der 1980-er Jahre) würde diese Zeichen zum Beispiel immer mit 150 Zeichen pro Sekunde übertragen. Nun nehmen wir aber mal an, dass eine Informationsquelle aus diesen vier Zeichen eine ungleichmäßige Wahrscheinlichkeitsverteilung bevorzugen würde, und zwar treten die Buchstaben *A*, *B*, *C* und *D* unabhängig voneinander mit der Wahrscheinlichkeit 1/2, 1/4, 1/8 bzw. 1/8 auf, dann ist die Entropie dieser Quelle

$H = -\frac{1}{2}\log\frac{1}{2} - \frac{1}{4}\log\frac{1}{4} - \frac{1}{8}\log\frac{1}{8} - \frac{1}{8}\log\frac{1}{8} = \frac{1}{2} + \frac{2}{4} + \frac{6}{8} = \frac{7}{4}$ Bits/Symbol. Die Fano'sche Methode liefert nun einen effizienteren Code, und zwar bekommt *A* das Bit 0; *B*, *C* und *D* beginnen mit Bit 1; *B* bekommt als zweites Bit 0; *C* und *D* bekommen als zweites Bit 1; *C* bekommt als drittes Bit 0 und *D* bekommt als drittes Bit 1:

| | |
|---|---:|
| *A* | 0 |
| *B* | 10 |
| *C* | 110 |
| *D* | 111 |

Die durchschnittliche Bitzahl pro Symbol bei einer durchschnittlichen Übertragung ist nicht etwa (1+2+3+3)/4=9/4, sondern hängt ja vom wahrscheinlichen Auftreten der Zeichen ab, sie ist also (1/2·1 Bit + 1/4·2 Bits + 2/8·3 Bits) = (1/2+1/2+3/4) Bits = 7/4 Bits, das ist genau der Entropiewert. Damit überträgt unser 300 Bit/s-Kanal im Durchschnitt *C*/*H* = 300/(7/4) = 171,4 Zeichen/s, und das ist eine Steigerung gegenüber 150 um 21,4/171,4 also um ein Achtel. Und das ist auch genau der Redundanzwert dieser Informationsquelle: Die maximale Entropie des Beispieles {*A*,*B*,*C*,*D*} oben wäre -4·(1/4·log(1/4))=log(4)=2. Die tatsächliche Entropie ist 7/4. Die relative Entropie ist damit (7/4)/2=7/8 und die Redundanz damit 1-(7/8)=1/8.

Die Einsparquote im Rahmen der Redundanz ist übrigens unabhängig von der Kanalkapazität und gilt bei einer Kapazität von 300 B/s genauso wie für 1 MB/s oder 1 GB/s.

Man kann die Redundanz eines Codes auch durch eine geeignete Kompression des vorhandenen Zeichensatzes erreichen. In Folge von Shannons eigenem Beispiel in Abschnitt 10 kann man das

Alphabet {*A*,*B*,*C*,*D*} aus dem Beispiel oben auf ein anderes Alphabet {*A'*,*B'*,C',*D'*} komprimieren, indem man nun rückwärts jeweils Bit-Digramme wie folgt zuordnet:

| | |
|---|---|
| 00 | *A'* |
| 01 | *B'* |
| 10 | *C'* |
| 11 | *D'* |

Zum Beispiel würde nun eine originale 4-Zeichen-Botschaft *ABBA*, die im „Fano-Code" als 0 10 10 0 übertragen wird, im neuen Code als *B'B'A'*, also nur in drei Zeichen und mithin schneller übertragen werden. Die tatsächliche Kompressionsrate ist mit 1/8 allerdings etwas geringer als 1/4, daher klappt das nicht mit jeder 4-Zeichen-Folge. Aber im Durchschnitt spart man doch bei je 8 Zeichen eines ein. Soweit der konstruktive Teil von Shannons Hauptsatz.

Der absolute Teil besagt, dass man selbst mit einer noch so ausgeklügelten Kodierung diese Kompression nicht mehr verlustfrei überbieten kann. Die maximale Übertragungsrate für eine Informationsquelle mit Entropie $H$ Bits/Symbol über einen Kanal mit Kapazität $C$ Bit/s beträgt $C/H$ Zeichen pro Sekunde, was unmittelbar einleuchtet, denn jedes Zeichen „kostet" im Durchschnitt $H$ Bits. In dem Beispiel oben haben wir also durch effizientere Kodierung bzw. Kompression bereits die maximale Übertragungsrate in Zeichen/Sekunde erreicht.

Die Entropien $H$ Bits/Symbol und $H'$ Bits/Sekunde hängen über die Formel (17) $H'=mH$ Bit/s zusammen, wobei $m$ die Anzahl Zeichen pro Sekunde ist, die die Informationsquelle über den Kanal überträgt. Der Hauptsatz für ungestörte diskrete Kanäle (20) besagt nun, dass $\max(m)=C/H$ die bestmögliche Übertragungsrate für alle Informationsquellen über diesen Kanal ist, also gilt

$$(21) \qquad \max(H') = C \text{ Bit/s}$$

wobei das Maximum über alle Informationsquellen und alle denkbaren Kodierungen für diesen Kanal gebildet wird. Dieser Zusammenhang wird für die Definition der Kapazität eines diskreten gestörten Kanals ausgenutzt, s.u. Formel (25).

Die Sensation durch die Erfindung des MP3-Audiocodes (Brandenburg et al. 1989 und 1994, MPEG 1992/1994) bestand ja darin, dass man Musik in guter digitaler Qualität über das Telefon übertragen konnte, was auf den ersten Blick wie eine Durchbrechung der Shannon'schen „Naturgesetze" im Range eines Perpetuum Mobile erscheint. In Wirklichkeit wird diese „Überbietung der Shannon'schen Grenze" durch die Aufgabe der Trennschärfe, also mit dem Verlust von Zeichen erreicht. Der Hauptsatz stellt die Grenze nur für verlustfreie Übertragung fest. Die Kunst der MP3-Erfinder bestand gerade darin, den Verlust so zu organisieren, dass der Hörer ihn nicht empfindet, und das erforderte neben dem Studium von Kodierungsregeln ein biologisches und wahrnehmungspsychologisches Verständnis des menschlichen Ohres.

Abschließend sei noch auf die Redundanz natürlicher Schriftsprachen hingewiesen. Shannon beobachtet, dass Kreuzworträtsel ohne Lücken problemlos möglich sind, wenn eine Sprache redundanzfrei ist, weil dann jede Buchstabenfolge ein sinnvolles Wort ergäbe. Je mehr Redundanz in einer Sprache ist, desto schwerer ist die Konstruktion eines Kreuzworträtsels, weil es dann immer wenige sinnvolle Buchstabenfolgen gibt. Die deutsche Sprache hat etwa 50% Redundanz, *wi mn drch Wglssn vn viln Bchstbn as enm Tx fststlln knn*. Man kann sich fragen, wieso natürliche Sprachen eigentlich eine so hohe Redundanz herausgebildet haben. Der Grund liegt in der Robustheit, die die Datenübertragung durch Kompensation von Geräuschstörungen mit Redundanz erreicht. Das behandelt Shannon in einem eigenen Kapitel über diskrete Kanäle mit Störung.

# 6 Der diskrete Kanal mit Störung

## 6.1 Störung, Korrektur und Übertragungsrate

Wenn ein Übertragungskanal gestört ist, dann kommen nicht mehr alle Sendesymbole $x$ in ihrer originalen Form an. Der Sender weiß also nicht sicher, was aus seinem Input $x$ wird, und der Empfänger weiß nicht mehr sicher, ob sein Output $y$ mit dem Input übereinstimmt. Während bei einem ungestörten Kanal die Eingangs- und Ausgangsentropien gleich waren, $H(x)=H(y)$, so gilt das für den gestörten Kanal nicht mehr: Es besteht eine Störwahrscheinlichkeit, die sich aus Sicht des Senders in einem bedingten Entropiewert des Outputs in Abhängigkeit vom Input, $H_x(y)$, ausdrücken lässt. Auf Empfängerseite besteht die Störwahrscheinlichkeit in der unsicheren Abhängigkeit des Inputs vom Output, $H_y(x)$. Diese beiden bedingten Entropien sind im ungestörten Kanal null, d.h. Input und Output sind gleich und determinieren sich damit gegenseitig vollkommen. Im gestörten Kanal gilt dagegen allgemeiner

$$(22) \qquad H(x,y) = H(x) + H_x(y) = H(y) + H_y(x)$$

$H_x(y)$ kann man als Entropie des Störgeräusches interpretieren, und $H_y(x)$ als erforderliche Korrekturrate des fehlerhaften Outputs. Alle diese Werte können pro Symbol und pro Sekunde ange-

 © Grimm, TU Ilmenau, Dez 2004

geben werden. Zum Vergleich mit der Kanalkapazität werden im Folgenden die Entropien in Bits/Sekunde betrachtet.

Die Korrekturrate $H_y(x)$ nennt Shannon *Equivocation*, das heißt so viel wie „Gleichlaut". Ich finde den Begriff „Korrekturrate" für $H_y(x)$ zur Kompensation der „Störung" $H_x(y)$ treffender, denn sie gibt an, wie viele Bits pro Sekunde am Empfangsteil hinzugefügt werden müssen, um dort die Botschaft aus dem Input wiederherzustellen.

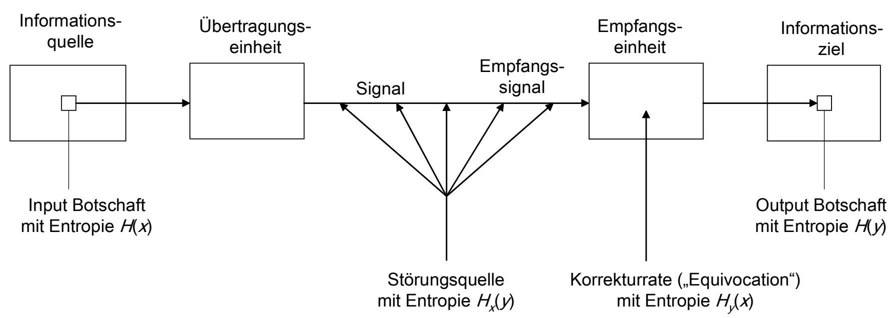


*Abbildung 6:* Stör- und Korrekturrate („Equivocation") eines diskreten gestörten Kanals.

Shannon entscheidet sich dafür, als Übertragungsmaß der Information einer Informationsquelle über einen diskreten Kanal mit Störung die Sendeentropie $H(x)$ als Ausgangspunkt für den Input zu nehmen, von dem die *Empfangs*unsicherheit $H_y(x)$ abgezogen werden muss (Shannon 1948, Abschnitt 12).

(23) Definition: Die *Übertragungsrate einer Informationsquelle* ist definiert als $R := H(x) - H_y(x)$ Bit/s

Das kann man so interpretieren: Der Kanal überträgt nur soviel an "echter Information", als er im Input aufnimmt minus dessen, was der Empfänger von sich aus hinzufügen muss, um das Originalsignal wieder herzustellen. Nach (22) ist das äquivalent zu

(24) $R = H(y) - H_x(y)$ Bit/s

Die Übertragungsrate besagt in dieser Formulierung, dass soviel an „echter Information" übertragen wird, wie das Outputsignal hergibt minus de Störung, die das Outputsignal enthält. Beide Versionen (23) und (24) sind gleichwertig und intuitiv nachvollziehbar.

Exkurs in die Verschlüsselungstheorie (Kryptologie): In einer späteren Arbeit über „Secrecy Systems" (1949) bemerkt Shannon einen fundamentalen Zusammenhang zwischen Störung und Verschlüsselung. Die Verschlüsselung kann als Störung des stochastischen Prozesses der Informationsquelle interpretiert werden. Während die Störung $H_x(y)$ bei der Übertragung unerwünscht ist und beim Empfang mit einer Korrektur $H_y(x)$ ausgeglichen werden muss, ist die Verschlüsselung $H_x(y)$ erwünscht. Sie muss beim Empfang durch eine Entschlüsselung $H_y(x)$ wieder aufgehoben werden. Je stärker die „Störung" der Verschlüsselung, desto besser: In Bezug auf den Kryptoanalytiker, der den Datenstrom der verschlüsselten Nachricht nach statistischen Gesichtspunkten auswertet, um Anhaltspunkte für den Klartext zu erhalten, ist die Verschlüsselung statistisch „perfekt", wenn aus Kenntnis des Kryptogramms $H(y)$ im Output kein Rückschluss auf den Klartext $H(x)$ im Input möglich ist, wenn also Input $x$ und Output $y$ statistisch völlig unabhängig sind. In diesem Fall sind nach Formel (10) $H(x,y)=H(x)+H(y)$, sowie nach (14) $H_x(y)=H(y)$ und $H_y(x)=H(x)$. Letzteres besagt sehr anschaulich: Zu einem gegebenen Kryptogramm $y$ passt jeder Klartext $x$ mit derselben Wahrscheinlichkeit wie die freie Auswahl eines Klartextes $x$, mit anderen Worten: aus Kenntnis des Kryptogramms $y$ kann man statistisch keine Rückschlüsse auf einen zugehörigen Klartext $x$ ziehen.

 © Grimm, TU Ilmenau, Dez 2004

## 6.2 Beispiel einer Störung

Das Genie von Shannon zeigt sich wiederum in der gelungenen Begriffswahl für das Maß der Stör- und Übertragungsrate. Denn naiv würde man vielleicht den Prozentsatz der umkippenden Bits als Störmaß annehmen. Aber wenn 50% der Bits im Durchschnitt umfallen, dann kann man sich eigentlich auf *kein einziges* Bit im Output mehr verlassen, d.h. der Informationswert liegt dann bereits bei 0% und die Störrate sollte 100% betragen. Genau das leistet Shannon's „Equivocation“ alias „Korrekturrate“. Dazu ein Beispiel (Shannon 1949, Abschnitt 12):

Ein diskreter Kanal, bei dem aufgrund einer dauernden Störung 1% der Bits umfallen, hat folgende Wahrscheinlichkeitsverteilung seiner Zeichen, wobei wir hier jedes Bit als eigenes Zeichen ansehen: Der Empfang ($y$) einer Null lässt mit Wahrscheinlichkeit 0,99 auf das Senden ($x$) einer Null schließen und mit Wahrscheinlichkeit 0,01 auf das Senden einer Eins. Für den Empfang einer Eins gelten dieselben Wahrscheinlichkeiten analog; in Formeln: Für $p_y(x)$: $p_0(0)=0{,}99$, $p_0(1)=0{,}01$, $p_1(0)=0{,}01$, $p_1(1)=0{,}99$. Wenn man für die Informationsquelle in unserem Beispiel weiterhin annimmt, dass sie alle Zeichen (hier: Bits) mit gleicher Wahrscheinlichkeit aussendet, dann tritt die Eins bzw. die Null jeweils mit Wahrscheinlichkeit 0,5 auf, d.h. sowohl für $p(x)$ als auch für $p(y)$ gilt $p(0)=p(1)=0{,}5$. Für die Störrate $H_y(x)$ gilt nach Definition (11):

$$\begin{aligned} H_y(x) &= -[0{,}5\cdot(0{,}99\cdot\log_2(0{,}99)+0{,}01\cdot\log_2(0{,}01)) + 0{,}5\cdot(0{,}01\cdot\log_2(0{,}01)+0{,}99\cdot\log_2(0{,}99))] \\ &= 0{,}5\cdot 0{,}081+0{,}5\cdot 0{,}081 = 0{,}081 \end{aligned}$$

Die Inputrate $H(x)$ hat wegen der Gleichverteilung den Maximalwert $\log_2(2)=1$. Insgesamt ist die erforderliche Korrekturrate 8,1%, die Übertragungsrate also $R=H(x)-H_y(x)=1-0{,}081=0{,}919$ entsprechend 91,9%.

Wenn statt 1% im Schnitt sogar 10% der Bits umfallen, dann ist die erforderliche Korrekturrate $H_y(x)=-0{,}9\cdot\log_2(0{,}9)-0{,}1\cdot\log_2(0{,}1)=0{,}470=47\%$, und als Übertragungsrate bleibt $R$=53%. Bei 30% umfallender Bits liegt die zu korrigierende Fehlerrate $H_y(x)=-0{,}7\cdot\log_2(0{,}7)-0{,}3\cdot\log_2(0{,}3)=0{,}88=88\%$ bereits erheblich über der Informationsrate von R=12%.

Und wenn sogar 50% der Bits umfallen, beträgt die erforderliche Korrekturrate $H_y(x)=-0{,}5\cdot\log_2(0{,}5)-0{,}5\cdot\log_2(0{,}5)=1=100\%$, und als Übertragungsrate bleibt null übrig, wie erwartet. Interessanterweise steigt die Übertragungsrate mit wachsender Umfallrate der Bits jenseits der 50% wieder an, da dann aus den empfangenen Bits mit höherer Wahrscheinlichkeit auf die gesendeten Bits, nämlich auf den jeweils anderen Wert, geschlossen werden kann. Man kann sich sozusagen wieder mehr darauf verlassen, dass eine Eins gesendet wurde, wenn eine Null empfangen wird und umgekehrt. Auch das ist intuitiv nachvollziehbar und bestätigt Shannons geglückte Begriffswahl.

## 6.3 Effizienz der Kodierung und Hauptsatz

Der Hauptsatz für den diskreten Kanal *ohne* Störung in der Formulierung von (21) besagt, dass die maximale Entropie in Bit/s über alle Informationsquellen gerade die Kanalkapazität ist: $\max(H)=C$. In Verallgemeinerung dieses Sachverhalts definiert Shannon die Kanalkapazität eines diskreten Kanals *mit* Störung als die maximale Übertragungsrate $R=H(x)–H_y(x)$ über alle Informationsquellen für diesen Kanal (Shannon 1948, Abschnitt 12):

(25) Definition: Die *Kapazität C eines diskreten Kanals* ist definiert als

$$C := \max(H(x) - H_y(x))$$
$$= \max(H(y) - H_x(y))$$

mit dem Maximum über alle Informationsquellen für diesen Kanal.

Wenn der Kanal ungestört ist, dann gelten $H_y(x)=H_x(y)=0$, sowie $H(x)=H(y)=H$, so dass sich die Definition (25) auf den Hauptsatz (20, 21) für den ungestörten Kanal reduziert. Diese Definition gibt die Aufgabe vor, wie die Kapazität konkreter Übertragungskanäle zu berechnen ist: nämlich als Optimierungsaufgabe, für die gegebene Formel der Übertragungsrate $R=H(x)-H_y(x)$ das Maximum zu bilden, indem man die geeignete Informationsquelle und ihre maximale Entropie findet.

Der von Shannon entwickelte “Fundamentalsatz für diskrete gestörte Kanäle“ gibt nun an, wie viel Korrektur man mindestens in den Informationsstrom einfügen muss, um einen fehlerfreien Output zu erzeugen (Shannon 1948, Abschnitt 13, Theorem 11).


© Grimm, TU Ilmenau, Dez 2004

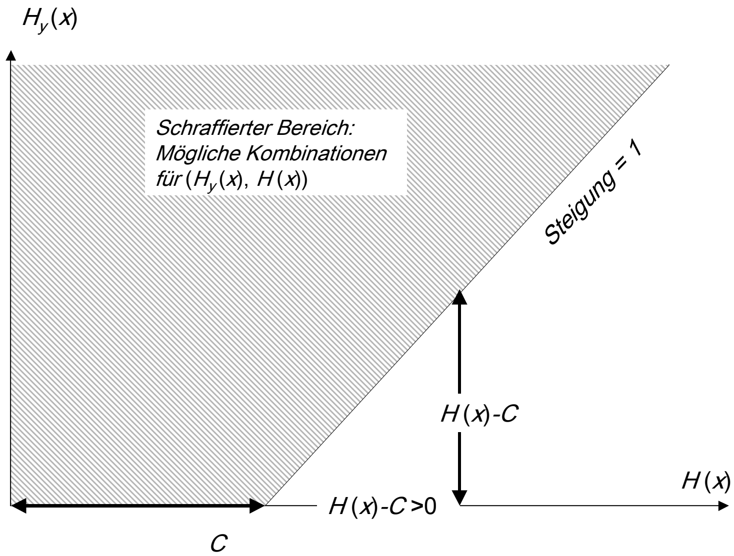


*Abbildung 7:* Zu Hauptsatz (26): Mögliche Kombination aus Input-Entropie $H(x)$ und Korrekturrate $H_y(x)$ bei einem diskreten gestörten Kanal. Nach Shannon 1948, Abschnitt13, Abb. 9.

(26) *Hauptsatz über den diskreten gestörten Kanal*: Eine Informationsquelle mit Input-Entropie $H$ Bit/s kann durch einen vorgegebenen Kanal mit Kapazität $C$ Bit/s fehlerfrei übertragen werden, wenn $H \leq C$, denn dann ist genügend Platz vorhanden, um die Informationsquelle effizient zu kodieren und ohne Korrekturrate ($H_y(x)=0$) fehlerfrei zu übertragen. Wenn dagegen $H>C$, d.h. wenn die Information mehr Platz braucht, als der Kanal an Kapazität hergibt, dann kann man die Informationsquelle so effizient kodieren, dass man mit einer Korrekturrate $H_y(x) \geq H-C$ auskommt, aber niemals mit weniger.

Der Hauptsatz bezeichnet exakt die Grenze für mögliche Kombinationen aus der Input-Entropie $H(x)$ und der Korrekturrate $H_y(x)$ bei einem diskreten gestörten Kanal, bei denen eine störungsfreie Übertragung ohne zusätzlichen Übertragungskanal möglich ist, sofern die Kodierung geschickt genug gewählt wird, genügend Redundanz in den Quellcode einzufügen, s. Abb. 7.

Der Beweis des Hauptsatzes ist nicht konstruktiv und daher unanschaulich. Anders als im störfreien Fall mit der Fano-Methode (s.o.) gibt es hier auch keine generelle Methode zur Entwicklung eines effizienten Codes bei Störung. Etwas aufwändigere Beispiele rechnet Shannon in (Shannon 1948, Abschnitte 15-17) durch.

## 6.4 Interpretation des Hauptsatzes

Die Korrekturrate („Equivocation“) $H_y(x)$ ist derjenige Informationsstrom in Bit/s, den der Empfänger in den Empfangsstrom mit der Entropie $H(y)$ einfügen muss, um das originale Signal wieder herzustellen, das der Empfänger nicht erhalten hatte, da die Störung der Quellenentropie $H(x)$ den Informationsanteil mit der Störentropie $H_x(y)$ weggenommen hatte. Wenn ein Kanal genügend Kapazität hat, $C \geq H$, dann kann die Korrektur noch innerhalb des Inputstroms eingefügt werden, und zwar in Form von Redundanz, der in den Code eingeht. Wie das funktioniert, demonstrieren beispielsweise der *Hamming Code* und der *CRC – Cyclic Redundancy Code*, etwa bei (Tanenbaum 2003, Abschnitt 3.2 über „Error Detection and Correction“).

Wenn der Kanal nicht genügend Platz hat, $H>C$, dann muss die Korrekturinformation auf andere Weise zugeführt werden, etwa über einen zusätzlichen Kanal, der mindestens über die Bandbreite $H–C$ verfügen muss.

Redundanz kann Störung kompensieren. Redundanz kann also für die Robustheit der Übertragung gegenüber Störung sorgen. Das ist der Grund für die etwa 50% Redundanz der natürlichen Sprachen, da die Kapazitäten der Übertragungskanäle beliebig groß sind: Papier und Tinte stehen praktisch unbegrenzt zur Verfügung, so dass ein geschriebener Text von einem flüchtigen Leser bequem verstanden werden kann, auch wenn er dabei vieles überliest. Und bei der gesprochenen Sprache ist es effizienter, langsam zu sprechen und dabei die Korrekturrate des Hörers in Anspruch zu nehmen, als auf ständige Rückfragen hin zu wiederholen. Es kann vorkommen, dass eine Umgebung so laut ist, dass man sich trotz Rufens nur noch unklar verständigen kann. In diesem Fall ist die Kanalkapazität zu klein zur störungsfreien Übertragung. Zusätzliche Kapazität kann dadurch geschaffen werden, dass der Rufer ein Blatt Papier hochhält mit den wichtigsten Wörtern, die er mitteilen wollte. Dieser enthält demnach Korrekturinformation, die durch den normalen Kanal, das gesprochene Wort, nicht übertragen werden kann.

 © Grimm, TU Ilmenau, Dez 2004

# 7 Analoge Informationsquellen mit Störung

## 7.1 Analoge Signale

Analoge Signale werden durch stetige Funktionen beschrieben, die nach Fourier (s.o.) als Summen von Sinus- und Cosinus-Schwingungen, also als periodische Funktionen aufgefasst werden können. In den Frequenz- und Maxima-Werten der einzelnen Fourierterme enthalten die analogen Signale diskrete Werte. Über Frequenz- und Amplitudenmodulationen können analoge Signalquellen als diskrete Informationsquellen genutzt werden. Das ist mit den modernen Übertragungsmedien wie Lichtwellen, digitales Radio, digitales Fernsehen über Antenne, ISDN oder T-DSL heute der Regelfall.

Viele Ergebnisse der diskreten Kanäle übertragen sich auf analoge Kanäle. Darüber hinaus gibt es einige spezielle Phänomene analoger Informationsquellen. Shannon stellt in (Shannon 1948, Teile III-V) keine allgemeine Theorie wie für den diskreten Fall vor, sondern entwickelt einige Ergebnisse für wichtige Spezialfälle analoger Informationsquellen, etwa für Signale mit Frequenzobergrenzen.

Ein Beispiel für eine kontinuierliche Funktionsfamilie, die analoge Signalquellen beschreibt, ist eine Menge phasenverschobener periodischer Funktionen, z.B. $\{\sin(\theta+t) | \, 0 \leq \theta \leq 2\pi\}$. Andere Beispiele bilden Funktionen mit Frequenz und Amplitudenbeschränkung oder Lautsignale gesprochener Sprachen.

Für die Untersuchung beschränkt man sich ebenfalls auf „ergodische Funktionsfamilien“ $\{f_\alpha\}$. Das sind solche Familien von Funktionen, die erstens „stationär“ sind, d.h. unter jeder Zeitverschiebung $f_\alpha(t) \rightarrow f_\alpha(t+\theta)$ wird dieselbe Funktionsfamilie erzeugt, und die zweitens *keine* stationäre Unterfamilie enthalten, die mit Wahrscheinlichkeit größer null und kleiner eins auftritt. Zum Beispiel ist die Familie $\{\sin(\theta+t) | \, 0 \leq \theta \leq 2\pi\}$ stationär und ergodisch, während die Familie $\{\alpha\sin(\theta+t) | \, 0 \leq \theta \leq 2\pi, \, 0 \leq \alpha \leq 1\}$ zwar stationär, aber nicht ergodisch ist, da für jedes feste $a$ mit $0<a<1$ eine Unterfamilie $\{\alpha\sin(\theta+t) | \, 0 \leq \theta \leq 2\pi, \, 0 \leq \alpha \leq a\}$ mit Wahrscheinlichkeit größer null und kleiner eins auftritt, die ihrerseits stationär ist.

## 7.2 Entropie einer kontinuierlichen Wahrscheinlichkeitsverteilung

Eine Funktion, deren Frequenzobergrenze $W$ ist, ist vollständig determiniert durch seine Funktionswerte in $\frac{1}{2W}$ Sekunden Abstand (Nyquist, s.o. Gleichung 3; und Shannon 1948, Abschnitt 19, Theorem 13). In diesem Zeitintervall zwischen 0 und $\frac{1}{2W}$ kann eine Funktion als eine eindimensionale Wahrscheinlichkeitsverteilung $p(x)$ aufgefasst werden, deren Funktionswert die Wahrscheinlichkeit des Auftretens für den in diesem Intervall abgetasteten Wert gibt.

Für ein Zeitintervall zwischen 0 und $T = \frac{n}{2W}$ ist die Funktion nach Nyquist durch $n$ verschiedene diskrete Werte determiniert und kann daher als $n$-dimensionale Wahrscheinlichkeitsverteilung $p(x_1, x_2, \ldots, x_n)$ angesehen werden.

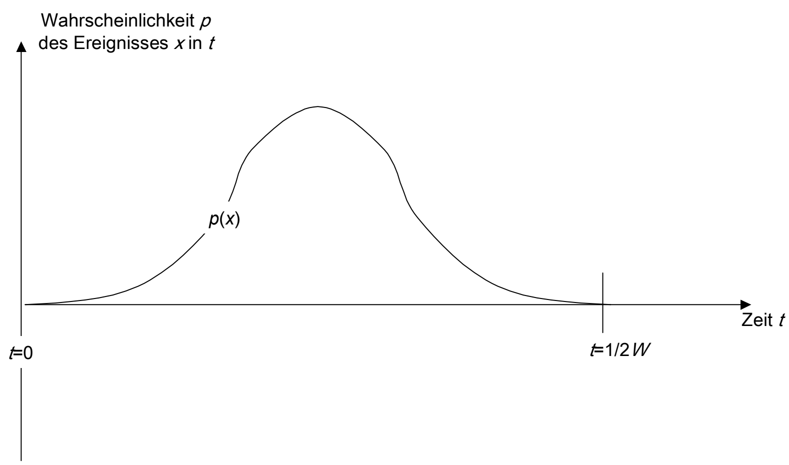


*Abbildung 8:* Eindimensionale Wahrscheinlichkeitsverteilung $p(x)$ in Gauss'scher Normalform im Intervall $[0,1/2W]$

Analog zur Entropie diskreter Wahrscheinlichkeiten $H(p_1,\ldots,p_n) = -\sum_{i=1}^{n} p_i \log_2 p_i$ (s.o. 8) wird für die stetige eindimensionale Wahrscheinlichkeitsverteilung $p(x)$ die Entropie wie folgt definiert:

(27) Definition: $H(x) = -\int_{-\infty}^{\infty} p(x)\log_2 p(x)dx$ bezeichnet die *Entropie einer eindimensionalen stetigen Wahrscheinlichkeitsverteilung* $p(x)$ im Intervall zwischen 0 und $T = \frac{1}{2W}$. $H$ wird in „Bits" angegeben.

Bei einem größeren Zeitintervall zwischen 0 und $T = \frac{n}{2W}$ und einer zugehörigen $n$-dimensionalen stetigen Wahrscheinlichkeitsverteilung $p(x_1, x_2, \ldots, x_n)$ wird die Entropie wie folgt definiert:

(28) Definition: $H(x_1, x_2, ..., x_n) = -\int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} p(x_1, x_2, ..., x_n)\log_2 p(x_1, x_2, ..., x_n)dx_1 dx_2 ... dx_n$ bezeichnet die *Entropie einer n-dimensionalen stetigen Wahrscheinlichkeitsverteilung* $p(x_1, x_2, \ldots x_n)$ im Intervall zwischen 0 und $T = \frac{n}{2W}$. $H$ wird in „Bits" angegeben.

Im 2-dimensionalen Fall, also im Intervall zwischen $T$=0 und $T$=1/$W$ und mit der Wahrscheinlichkeitsverteilung $p(x,y)$ gilt

(29) $$H(x, y) = -\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} p(x, y)\log_2 p(x, y)dxdy$$

$$H_y(x) = -\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} p(x, y)\log_2 \frac{p(x, y)}{p(y)} dxdy$$

$$H_x(y) = -\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} p(x, y)\log_2 \frac{p(x, y)}{p(x)} dxdy$$

mit $p(x) = -\int_{-\infty}^{\infty} p(x, y)dy$ und $p(y) = -\int_{-\infty}^{\infty} p(x, y)dx$.

Die Integrale brauchen gar nicht über das ganze Kontinuum (-∞,∞) gebildet zu werden, sondern nur im positiven Bereich der Funktionen zwischen 0 und $\frac{n}{2W}$.

Die kontinuierlichen Entropien haben viele der Eigenschaften diskreter Entropien, aber auch einige besondere Eigenschaften. Insbesondere gelten (vgl. Shannon 1948, Abschnitt 20):

1. $H(x,y) \leq H(x)+H(y)$ mit Gleichheit genau dann, wenn $p(x,y)=p(x)\cdot p(y)$, d.h. wenn die beiden Ereignisse unabhängig voneinander sind. Vergleiche den diskreten Fall (10).

2. $H(x,y)=H(x)+H_x(y)=H(y)+H_y(x)$, sowie $H_x(y)\leq H(y)$ mit Gleichheit genau dann, wenn $p(x,y)=p(x)\cdot p(y)$, d.h. wenn die beiden Ereignisse unabhängig voneinander sind. Vergleiche die diskreten Fälle (12,13).

3. Unter allen eindimensionalen Wahrscheinlichkeitsverteilungen $p(x)$ mit einer festen Volumenbeschränkung $V$, d.h. für $\int_{-\infty}^{\infty} p(x)dx = V$ nimmt $H(x)$ für dasjenige $p(x)$ den größten Wert an, das konstant $1/V$ ist, d.h. für das gilt $p(x) = 1/V$ für $x \in [0, \frac{1}{2W}]$ und $p(x)=0$ außerhalb davon. Vergleiche den diskreten Fall der gleichverteilten diskreten Wahrscheinlichkeitswerte und der dadurch erzielten maximalen Entropie $\log(n)$.

4. Je gleichmäßiger eine Wahrscheinlichkeitsverteilung ausgeprägt ist, desto größer ist ihre Entropie.

5. Unter allen eindimensionalen Wahrscheinlichkeitsverteilungen mit Standardabweichung $\sigma$ nimmt $H(x)$ für die Gauss'sche Normalverteilung den größten Wert an, d.h. für $p(x) = \frac{1}{\sqrt{2\pi}\sigma} e^{-(\frac{x^2}{2\sigma^2})}$, für das bekanntlich gilt: $\int_{-\infty}^{\infty} p(x)dx = 1$ und $\int_{-\infty}^{\infty} p(x)x^2 dx = \sigma^2$. Entsprechende Aussagen gelten für mehrdimensionale Wahrscheinlichkeitsverteilungen.

6. Für die Gauss'sche Normalverteilung in 5. oben ergibt der maximierte Entropiewert $H(Gauss) = \log_2 \sqrt{2\pi e}\sigma$. Vergleiche den diskreten Fall der gleichverteilten diskreten Wahrscheinlichkeitswerte und der dadurch erzielten maximalen Entropie $\log(n)$.

7. Die Entropiewerte stetiger Wahrscheinlichkeitsverteilungen sind koordinatenabhängig. Allgemein geht der Koordinatenwechsel über die Jacobische Determinante $|\partial y/\partial x|$ in die Entropie ein. Speziell für einen linearen Koordinatenwechsel $y_i=\sum_j a_{ij}x_j$ gilt $H(y)=H(x)+\log_2|a_{ij}|$, wobei $|a_{ij}|$ die Determinante der Matrix $(a_{ij})$ ist.

 © Grimm, TU Ilmenau, Dez 2004

## 7.3 Übertragungsrate und Kanal-Kapazität

Ein analoger Übertragungskanal („continuous channel“, Shannon 1948, Teil IV) hat auf der Sendeseite als Input eine Familie stetiger Funktionen $\{f_\alpha(t)\}$ und auf der Empfängerseite als Output eine gestörte Version dieser Funktionen (Shannon 1948, Abschnitt 24). Bei einer Frequenzbeschränkung nach oben durch eine maximale Frequenz $W$ transportiert diese Familie im Zeitintervall $[0,T]$ $T{\cdot}2W$ diskrete Signale, d.h. alle $\frac{1}{2W}$ Sekunden einen anderen Wert $x_i$. Über einen Zeitintervall $[0,\frac{n}{2W}]$ lässt sich die Funktionenfamilie als $n$-dimensionale stetige Wahrscheinlichkeitsverteilung $P(x_1,x_2,\ldots,x_n)$ beschreiben. Die Übertragungsrate der Information des gestörten analogen Kanals kann wie im diskreten Fall (23, 24) definiert werden als

$$(30)\qquad \begin{aligned} R &:= H(x) - H_y(x) \ \text{Bit/s} \\ &= H(y) - H_x(y) \ \text{Bit/s} \end{aligned}$$

Die Kapazität eines analogen Übertragungskanals wird ebenfalls wie im diskreten Fall (25) als das Maximum der Übertragungsrate über alle möglichen Informationsquellen definiert:

$$(31)\qquad C := max(R)$$

Wenn man abkürzend $P(x)=P(x_1,x_2,\ldots,x_n)$ für die $n$-dimensionale Input-Wahrscheinlichkeitsverteilung und $P(y)= P(y_1,y_2,\ldots,y_n)$ für die $n$-dimensionale Output-Wahrscheinlichkeitsverteilung schreibt, berechnet sich die Kanalkapazität aus den Input- und Output-Wahrscheinlichkeiten wie folgt:

$$(32)\qquad C = \lim_{T\to\infty} \max_{P(x)} \frac{1}{T} \iint P(x,y) \log \frac{P(x,y)}{P(x)P(y)} dxdy$$

Wie im diskreten Fall hilft hier die ergodische Eigenschaft der Funktionenfamilien, dass der Wert mit wachsendem Zeitintervall stabiler wird. Die Kanalkapazität ist vom gewählten Koordinatensystem für $x$ und $y$ unabhängig.

Für $\log_2$ zur Basis 2, d.h. für eine Kodierung von $H(x)$, $H_y(x)$, $H(y)$ und $H_x(y)$ in *Bits*, stellt die Kanalkapazität $C$ die maximal größte Anzahl an Bits pro Sekunde dar, die eine Informationsquelle ohne (genauer: „mit beliebig kleiner“) Korrekturrate $H_y(x)$ über den Kanal senden kann, ganz wie im diskreten Fall (s.o. Hauptsatz 26).

Die Sachlage vereinfacht sich erheblich für den Spezialfall, dass die Störung $n$ („$n$" für „Noise") und das Signal $x$ im statistischen Sinne voneinander unabhängig und einander additiv hinzugefügt sind. In diesem Fall hat die Störung eine unabhängige Wahrscheinlichkeitsverteilung $Q(n)$ und dementsprechend eine unabhängige Entropie $H(n)$. Aus den Definitionen (30, 31) und der Interpretation von $H_x(y)$ als Störung ergeben sich dann als Übertragungsrate $R$ und Kapazität $C$

$$(33) \qquad R = H(y) - H(n)$$
$$C = \max_{P(x)}(H(y) - H(n))$$

wobei $y$ die Wahrscheinlichkeitsverteilung des Outputs und $n$ die Wahrscheinlichkeitsverteilung der Kanalstörung sind. Der Beweis von (33) ist einfach: Es gilt $y=x+n$, da Quellsignal und Störung sich additiv überlagern, also gilt $H(x,y)=H(x,n)$. Außerdem gilt $H(x,n)= H(x)+H(n)$, da $x$ und $n$ unabhängig voneinander sind, s.o. Eigenschaft 1 gestörter analoger Signale. Schließlich gilt, wie immer, $H(x,y)=H(y)+H_y(x)$, s.o. Eigenschaft 2. Wenn man diese Gleichungen zusammensetzt, ergibt sich $H(y)+H_y(x)=H(x,y)=H(x,n)=H(x)+H(n)$, und entsprechend durch Subtraktion von $H_y(x)$ und $H(n)$ auf beiden Seiten $H(y)–H(n) =H(x)–H_y(x)= R$, wie behauptet.

## *7.4 Kanal-Kapazität für eine Informationsquelle mit begrenzter Sendeleistung*

Formel (33) gilt, wie gesagt, für den speziellen Fall einer statistisch unabhängiger Störung, die einem Signal additiv zugefügt ist. Daraus ergibt sich einer der wichtigsten Spezialfälle für analoge gestörte Informationsquellen mit Frequenzobergrenze $W$ und Begrenzung der *durchschnittlichen* Sendeleistung nach oben durch $S$. Wenn ein solches Signal durch ein gleichmäßiges, normalverteiltes Rauschen („white thermal noise") der Leistung $N$ gestört ist, gilt die einfache Formel einer Kanalkapazität

$$(34) \qquad C = W \cdot \log(1 + \frac{S}{N})$$

Dieser Fall wird in (Shannon 1948, Abschnitt 25) behandelt. Die Ableitung ist einfach: Weil die Gauss'sche Normalverteilung die maximale Übertragungsrate garantiert (Eigenschaften 5 und 6 für analoge gestörte Signale) und diese zur Berechnung der Kanalkapazität heranzuziehen ist (Definition 31), ist das Empfangssignal $y$ als Gauss'sche Normalverteilung anzusehen. Es hat die

 © Grimm, TU Ilmenau, Dez 2004

durchschnittliche Empfangsleistung *N*+*S*. Damit ergeben sich für die Entropien $H(y)=W\log(\sqrt{2\pi e}(N+S))$ und $H(n)=W\log(\sqrt{2\pi e}N)$. Aus Formel (33) *C*=max(*R*)=*H*(*y*)–*H*(*n*) ergibt sich die behauptete Formel (34).

Der andere wichtige Spezialfall, der hier aufgeführt werden soll, handelt von analogen gestörten Informationsquellen mit Frequenzobergrenze *W* und Begrenzung der *absoluten* Sendeleistung durch *S*. Das heißt, kein einziges Signalelement übertrifft die Leistung *S*, während in (34) oben der durchschnittlichen Begrenzung einzelne Signale *S* durchaus übertreffen können, wenn der gesamte Durchschnitt darunter bleibt. Der Fall der Begrenzung der *absoluten* Sendeleistung ist leider nicht ganz so glatt zu formulieren, sondern erlaubt nur Abschätzungen nach oben und unten. Dieser Fall wird in (Shannon 1948, Abschnitt 26) behandelt:

Wenn ein analoges Signal mit Frequenzobergrenze *W* und Begrenzung der *absoluten* Sendeleistung von höchstens *S* durch ein gleichmäßiges, normalverteiltes Rauschen („white thermal noise“) der Leistung *N* gestört ist, gilt für die Abschätzung der Kanalkapazität nach unten:

(35) $$C \geq W \cdot \log \frac{2}{\pi e^3} \frac{S}{N}$$

Das heißt, dass *C* in jedem Fall *mindestens* diesen Wert annimmt oder überschreitet. Für die Abschätzung nach oben, d.h. welchen Wert *C* *höchstens* annehmen kann, muss man zwei Fälle unterscheiden: erstens, dass die höchste Signalleistung *S* die Störung *N* sehr stark übertrifft, *S*>>N, bzw. *S*/*N* sehr groß; zweitens, dass die höchste Signalleistung *S* sehr viel schwächer als die Störung *N* ist, *S*<<N, bzw. *S*/*N* fast null:

(36) $$C \leq W \cdot \log(1 + \frac{2}{\pi e} \frac{S}{N})$$, wenn *S*/*N* sehr groß ist

(37) $$C \rightarrow W \cdot \log(1 + \frac{S}{N})$$, je kleiner *S*/*N* ist.

Die Formel (37) entspricht der Formel (34), gilt hier aber nur für sehr schwache Signale bei hohem Störgeräusch. Je stärker einzelne „Signaltöne“ sind, desto schwächer schlägt der (logarithmische) Summand *S*/*N* auf die Kapazität durch (Faktor $2/\pi e$ ist etwa gleich 0,117).

An dieser Stelle sei noch einmal auf das Beispiel der Informationsrate des analogen Telefons hingewiesen. Im klassischen analogen Telefonsystem mit einer Frequenz von 3 kHz beträgt die durchschnittliche Übertragungslautstärke etwa 60 dB, das entspricht dem „Signal-Rausch-Abstand“ $S/N= 10^6$, da $\log_{10}(10^6)=6$. Deshalb kann man über das analoge Telefon *verlustfrei* niemals mehr als $3.000 \cdot \log_2(1 + S/N) = 3.000 \cdot \log_2(1.000.001) \approx 3.000 \cdot 20 = 60$ KBits/Sekunde übertragen. Da CD-Qualität eine Übertragungsrate von 1,4 MBit/s erfordert, ist klar, dass die Anforderung einer HiFi-Übertragung über das analoge Telefon verlustfrei unmöglich ist (Shannon 1948, Gleichungen 34-37 oben), auch nicht für das digitale ISDN-Telefon mit 64 KBit/s. Die Leistung der MP3-Erfinder bestand denn auch darin, eine Verlustrate einzuführen, die das Ohr nicht wahrnimmt (Brandenburg et al. 1989 und 1994, MPEG 1992/1994). Für analoges Telefon mit 3 kHz Mono wird das Audio-Signal um 1:96 komprimiert, für ISDN-Telefon (11 kHz Stereo und bis zu 64 KBit/s) um 1:24 mit immerhin UKW-Qualität; und CD-Qualität (über 15 kHz Stereo und bis zu 128 KBit/s) erreicht man bei einer Kompressionsrate von 1:12. MP3 ist demnach nicht nur eine wissenschaftliche Leistung der Nachrichtentechnik, sondern auch der Biologie des Ohres und der Wahrnehmungspsychologie.


© Grimm, TU Ilmenau, Dez 2004

# 8 Übungsaufgaben

1. Welchen Gegenstand hat die Shannon'sche Kommunikationstheorie?

2. Welche der semiotischen Ebenen behandelt die Shannon'sche Kommunikationstheorie, welche werden nicht behandelt?

3. Welche maximale Datenrate kann ein ungestörter analoger Kanal mit zwei Levels (eine Amplitude und Null) erreichen, dessen maximale Frequenz 3 kHz ist?

4. Im klassischen analogen Telefonsystem mit einer Frequenz von 3 kHz beträgt die durchschnittliche Übertragungslautstärke etwa 60 dB, das entspricht der „signal-to-noise ratio" $S/N$=1.000.000, da $\log_{10}(10^6)=6$. (Zur Erinnerung 1dB = $10\cdot\log_{10}(S/N)$.) (a) Mit Hilfe welcher Shannon'schen Formel wird die maximale Datenrate in Bit/s über diesen Kanal bestimmt? (b) Wie hoch ist die maximale Datenrate in diesem Kanal? (c) Wie viele Dezibel und welche zugehörige „Signal-to-Noise-Ratio" würden für den Kanal durchschnittlich ausreichen, wenn man mit der halben Datenrate zufrieden wäre? (d) Alternativ, für dieselbe durchschnittliche Lautstärke 60 dB, welche Frequenz würde für den Kanal ausreichen, um die halbe Datenrate zu erreichen?

5. Wie viele verschiedene Zeichen kann man mit einer Kombination aus 8 Bits kodieren?

6. Nennen Sie aus dem Bereich natürlicher Schriftsprachen ein Beispiel für einen ergodischen Markoff'schen Prozess, sowie einen Markoff'schen Prozess, der nicht ergodisch ist.

7. Definieren Sie die „Entropie eines Ereignisses", das als „A" oder „B" (erweiterte Aufgabe: als „A" oder „B" oder „C") auftreten kann. Welche Wahrscheinlichkeitsverteilung führt zur minimalen Entropie, und welche zur maximalen Entropie? Welches sind die Werte der minimalen und maximalen Entropie?

8. Eine Informationsquelle sende sechs Zeichen A, B, C, D, E und F mit folgender Auftrittswahrscheinlichkeit aus: A 10%, B 5%, C 40%, D 20%, E 5%, F 20%. Dabei sei jedes Sendeereignis vom vorherigen unabhängig. (a) Berechnen Sie Entropie, relative Entropie und Redundanz dieser Informationsquelle (nutzen Sie dabei die Werte $\log_2(10)=3{,}32$ und $\log_2(6)=2{,}58$). (b) Kodieren Sie die Symbole nach der Methode von Fano (es gibt dabei mehrere Möglichkeiten, die gleichwertig sind). Berechnen Sie die zugehörige durchschnittliche Bitlänge pro Symbol. (c) Wieso bleibt die durchschnittliche Bitlänge aus (b) etwas über dem Idealwert der Entropie aus (a)?

9. Berechnen Sie die Korrektur- und Übertragungsrate einer Informationsquelle, die die Symbole A und B sendet, wobei in 8% des Datenstroms A und B aufgrund einer Störung vertauscht werden. Die Wahrscheinlichkeiten des Auftretens eines A oder B seien gleich (nutzen Sie dabei die Werte $\log_2(10)=3{,}32$, $\log_2(92)=6{,}52$).

10. Eine digitale Musikquelle erzeugt CD-Qualität mit einer Übertragungsrate von 1,4 MBit/s und soll über einen ISDN-Kanal mit 64 KBit/s übertragen werden. (a) Welche Entropie müsste ein zusätzlicher Korrekturkanal haben, um das empfangene Signal wieder auf die originale CD-Qualität zurück zu bringen? (b) Wie hoch muss die verlustbehaftete Kompressionsrate sein, wenn der ISDN-Kanal ohne Korrekturkanal genutzt werden muss (was die Regel ist)?

# 9 Literaturangaben

Bentele, Günter; Beck, Klaus (1994): Information – Kommunikation – Massenkommunikation: Grundbegriffe und Modelle der Publizistik und Kommunikationswissenschaft. In: Jarren, Otfried (Hrsg.): Medien und Journalismus 1. Westdeutscher Verlag, Opladen 1994, S. 16-50.

Bentele, Günter; Bystrina, Ivan (1978): Semiotik. Grundlagen und Probleme. W. Kohlhammer Verlag, Stuttgart, 1978.

Brandenburg, Karlheinz; Grill, B.; Jonuscheit, H.; Kapust, R.; Seitzer, D.; Sporer, Th. (1989): „Übertragung von hochwertigen Tonsignalen mit Datenraten im Bereich 64 - 144 kbit/s." Rundfunktechnische Mitteilungen, 1989, H. 5, S. 209-213 (zuerst veröffentlicht in ITG-Fachtagung „Hörrundfunk" 1988).

Brandenburg, Karlheinz; Stoll, G. (1994): „The ISO/MPEG-1 Audio Codec: A Generic Standard for Coding of High Quality Digital Audio." Journal of the AES, Oktober 1994, S. 780 - 792 (zuerst veröffentlicht als Konferenzbeitrag zur 92. AES-Convention, Wien 1992).

Capurro, Rafael (2001): Theorie der Botschaft. Beitrag zum Symposion „Transziplinäre Kommunikation. Aktuelle Be-Deutung des Phänomens Kommunikation im fächerübergreifenden Dialog." Universität Salzburg, Österreich, 25.-26. April 2001, 21 S. Proceedings. Online von http://sammelpunkt.philo.at:8080/archive/00000062/ [download 11.10.2004].

Fano, R.M. (1949) in: M.I.T, The Research Laboratory of Electronics, Technical Report No. 65, March 17, 1949.

Hartley, R.V.L. (1928): Transmission of Information. Bell System Technical Journal, 7. July 1928, 535-563.

Jarren, Otfried (Hrsg., 1994):Medien und Journalismus 1. Eine Einführung. Wetsdeutscher Verlag, Opladen 1994, 330 Seiten.

Kloock, Daniela; Spahr, Angela (2000): Medientheorien – Eine Einführung. UTB, Werner Fink, München, 2. Auflage 2000. Darin: Kümmel, Albert: Mathematische Medientheorien, S. 205-236.


© Grimm, TU Ilmenau, Dez 2004

Kuhn, Wilfried (1976): Physik, Band III B, Thermodynamik und Statistik. Westermann, Braunschweig, 1976, 180 S.

Merten, Klaus (1977): Kommunikation. Eine Begriffs- und Prozessanalyse. Westdeutscher Verlag, Opladen 1977.

Merten, Klaus (1999). Einführung in die Kommunikationswissenschaft. Grundlagen der Kommunikationswissenschaft. Lit-Verlag, Münster usw. 1999, 585 Seiten.

MPEG-1 (1992): ISO/IEC JTC1/SC29/WG11 (MPEG), International Standard ISO/IEC 11172-3, Coding of moving pictures and associated audio for digital storage media at up to about 1.5 MBit/s. Part-3: Audio. 1992. („MPEG-1 Audio". Beschreibt eine Familie von drei Audiocodierverfahren („Layer"). „MP3" ist „MPEG-1 Audio" + „MPEG-2 Audio low sampling rates").

MPEG-2 (1994): ISO/IEC JTC1/SC29/WG11 (MPEG), International Standard ISO/IEC 13818-3, Generic coding of moving pictures and associated audio. Part 3: Audio. 1994. („MPEG-2 Audio". Enthält auch die „halben Abtastraten" (u.a. mehr). „MP3" ist „MPEG-1 Audio" + „MPEG-2 Audio low sampling rates").

Nyquist, H. (1924): Certain Factors Affecting Telegraph Speed. Bell System Technical Journal, 4. April 1924, 324 ff.

Nyquist, H. (1928): Certain Topics in Telegraph Transmission Theory. A.I.E.E. Trans., v. 47, April 1928, 617 ff.

Schramm, Wilbur (1954): The Process and Effect of Mass Communication. University of Illinois Press. Urbana 1954.

Shannon, Claude E. (1948): A Mathematical Theory of Communication. The Bell System Technical Journal. Vol. 27, July 1948, pp. 379-423; and October 1948, pp. 623-656. Reprint with Corrections at http://cm.bell-labs.com/cm/ms/what/shannonday/paper.html [download 25.10.04].

Shannon, Claude E.; Weaver, Warren (1949): The Mathematical Theory of Communication. University of Illinois Press: Urbana and Chicago 1949. E.g., in: Paperback 1949/1998, Library of Congress Card Catalog, USA, No. 49-11922, 125 pages. Auf Deutsch erschienen als „Mathematische Grundlagen der Informationstheorie", übersetzt von H. Dreßler, Oldenbourg, München, Wien 1976.

Shannon, Claude E. (1949): Communication Theory of Secrecy Systems. The Bell System Technical Journal, Vol 28, October 1949, pp. 656-715.

Tanenbaum, Andrew S. (2003): Computer Networks. Fourth Edition. Prentice-Hall International Editions, Englewood Cliffs, N.J., 2003, 912 Seiten.

Vowe, Gerhard (2004): Der Informationsbegriff in der Politikwissenschaft. Diskussionsbeiträge des Institut für Medien- und Kommunikationswissenschaft, TU Ilmenau, Nr. 10, ISSN 1617-9048, Januar 2004, 25 S.

Wagenknecht, Achim (1995): Einführung in die politische Philosophie Hanna Arendts. Tectum Verlag, 1995, 128 S.

# Inhaltsverzeichnis

1 Gegenstand .... 1
2 Hintergrund und Wirkung .... 2
3 Das Grundmodell und die Grundbegriffe .... 6
4 Zusammenhang zwischen analogen und diskreten Signalen (Fourier und Nyquist) .... 10
4.1 Fourier .... 10
4.2 Nyquist und Shannon .... 11
5 Der diskrete Kanal ohne Störung .... 13
5.1 Kanalkapazität .... 13
5.2 Kleiner Exkurs über den Logarithmus .... 14
5.3 Wahrscheinlichkeitssteuerung des Übertragungsprozesses .... 15
5.4 Effektive Kodierung .... 19
5.5 Wahlmöglichkeit, Unsicherheit und Entropie .... 21
5.6 Entropie einer Informationsquelle .... 26
5.7 Effizienz der Kodierung und Hauptsatz .... 30
6 Der diskrete Kanal mit Störung .... 33
6.1 Störung, Korrektur und Übertragungsrate .... 33
6.2 Beispiel einer Störung .... 36
6.3 Effizienz der Kodierung und Hauptsatz .... 37
6.4 Interpretation des Hauptsatzes .... 39
7 Analoge Informationsquellen mit Störung .... 40
7.1 Analoge Signale .... 40
7.2 Entropie einer kontinuierlichen Wahrscheinlichkeitsverteilung .... 41
7.3 Übertragungsrate und Kanal-Kapazität .... 44
7.4 Kanal-Kapazität für eine Informationsquelle mit begrenzter Sendeleistung .... 45
8 Übungsaufgaben .... 48
9 Literaturangaben .... 49

 © Grimm, TU Ilmenau, Dez 2004

01 Rüdiger Grimm, „Vertrauen im Internet – Wie sicher soll E-Commerce sein?“, April 2001, 22 S.
TU Ilmenau, Institut für Medien- und Kommunikationswissenschaft, ruediger.grimm@tu-ilmenau.de

02 Martin Löffelholz, „Von Weber zum Web – Journalismusforschung im 21. Jahrhundert: theoretische Konzepte und empirische Befunde im systematischen Überblick“, Juli 2001, 25 S.
TU Ilmenau, Institut für Medien- und Kommunikationswissenschaft, martin.loeffelholz@tu-ilmenau.de

03 Alfred Kirpal, „Beiträge zur Mediengeschichte – Basteln, Konstruieren und Erfinden in der Radioentwicklung“, Oktober 2001, 28 S.
TU Ilmenau, Institut für Medien- und Kommunikationswissenschaft, alfred.kirpal@tu-ilmenau.de

04 Gerhard Vowe, „Medienpolitik: Regulierung der medialen öffentlichen Kommunikation“, November 2001, 68 S.
TU Ilmenau, Institut für Medien- und Kommunikationswissenschaft, gerhard.vowe@tu-ilmenau.de

05 Christiane Hänseroth, Angelika Zobel, Rüdiger Grimm, „Sicheres Homebanking in Deutschland – Ein Vergleich mit 1998 aus organisatorisch-technischer Sicht“, November 2001, 54 S.
TU Ilmenau, Institut für Medien- und Kommunikationswissenschaft, ruediger.grimm@tu-ilmenau.de

06 Paul Klimsa, Anja Richter, „Psychologische und didaktische Grundlagen des Einsatzes von Bildungsmedien“, Dezember 2001, 53 S.
TU Ilmenau, Institut für Medien- und Kommunikationswissenschaft, paul.klimsa@tu-ilmenau.de

07 Martin Löffelholz, „Von ‚neuen Medien’ zu ‚dynamischen Systemen’, Eine Bestandsaufnahme zentraler Metaphern zur Beschreibung der Emergenz öffentlicher Kommunikation“, Juli 2002, 29 S.
TU Ilmenau, Institut für Medien- und Kommunikationswissenschaft, martin.loeffelholz@tu-ilmenau.de

08 Gerhard Vowe, „Politische Kommunikation. Ein historischer und systematischer Überblick der Forschung“, September 2002, 43 S.
TU Ilmenau, Institut für Medien- und Kommunikationswissenschaft, gerhard.vowe@tu-ilmenau.de

09 Rüdiger Grimm (Ed.), „E-Learning: Beherrschbarkeit und Sicherheit“, November 2003, 90 S.
TU Ilmenau, Institut für Medien- und Kommunikationswissenschaft, ruediger.grimm@tu-ilmenau.de

10 Gerhard Vowe, „Der Informationsbegriff in der Politikwissenschaft“, Januar 2004, 25 S.
TU Ilmenau, Institut für Medien- und Kommunikationswissenschaft, gerhard.vowe@tu-ilmenau.de

11 Martin Löffelholz, David H. Weaver, Thorsten Quandt, Thomas Hanitzsch, Klaus-Dieter Altmeppen, „American and German online journalists at the beginning of the 21st century: A bi-national survey“, Januar 2004, 15 S.
TU Ilmenau, Institut für Medien- und Kommunikationswissenschaft, martin.loeffelholz@tu-ilmenau.de

12 Rüdiger Grimm, Barbara Schulz-Brünken, Konrad Herrmann, „Integration elektronischer Zahlung und Zugangskontrolle in ein elektronisches Lernsystem“, Mai 2004, 23 S.
TU Ilmenau, Institut für Medien- und Kommunikationswissenschaft, ruediger.grimm@tu-ilmenau.de

13 Alfred Kirpal, Andreas Ilsmann, „Die DDR als Wissenschaftsland? Themen und Inhalte von Wissenschaftsmagazinen im DDR-Fernsehen“, August 2004, 21 S.
TU Ilmenau, Institut für Medien- und Kommunikationswissenschaft, alfred.kirpal@tu-ilmenau.de

14 Paul Klimsa, Torsten Konnopasch, „Der Einfluss von XML auf die Redaktionsarbeit von Tageszeitungen“, September 2004, 30 S.
TU Ilmenau, Institut für Medien- und Kommunikationswissenschaft, paul.klimsa@tu-ilmenau.de

15 Rüdiger Grimm, „Shannon verstehen. Eine Erläuterung von C. Shannons mathematischer Theorie der Kommunikation“, Dezember 2004, 51 S.
TU Ilmenau, Institut für Medien- und Kommunikationswissenschaft, ruediger.grimm@tu-ilmenau.de