

Auf Basis des informationstheoretischen Modells und einiger zusätzlicher Annahmen zeigt dieses Kapitel mit Hilfe einiger Sätze und Korollare, daß über den gestörten Kanal genau dann ein Maximum an Information übertragen werden kann, wenn die lineare Transformation eine Datenorthonormalisierung durchführt. In diesem Fall wären die Daten der Verbundquelle Y dekorreliert und alle Komponenten der Daten hätten die gleiche normierte Varianz¹.

Das bedeutet für die Korrelationsmatrix² von Y

$$C_{YY} = \langle y y^T \rangle = \begin{pmatrix} \sigma_{y_1}^2 & \dots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \dots & \sigma_{y_n}^2 \end{pmatrix} = P_Y$$

und für jede einzelne Komponente $i, j = 1, 2, \dots, n$

$$C_{YY} \text{ } i, j = \langle y_i y_j \rangle = \begin{cases} 0 & i \neq j \quad (\text{Dekorrelation}) \\ \sigma_{y_i}^2 & i = j \quad (\text{Normierung}) \end{cases}$$

Es ist also das Ziel dieses Kapitels zu zeigen, daß in dem informationstheoretischen Modell die Korrelationsmatrix C_{YY} die obige Form besitzen muß, damit ein Maximum an Information über den gestörten Kanal übertragen wird. In diesem Fall bietet die Datenorthonormalisierung einen wesentlichen Vorteil bei der Datenübertragung.

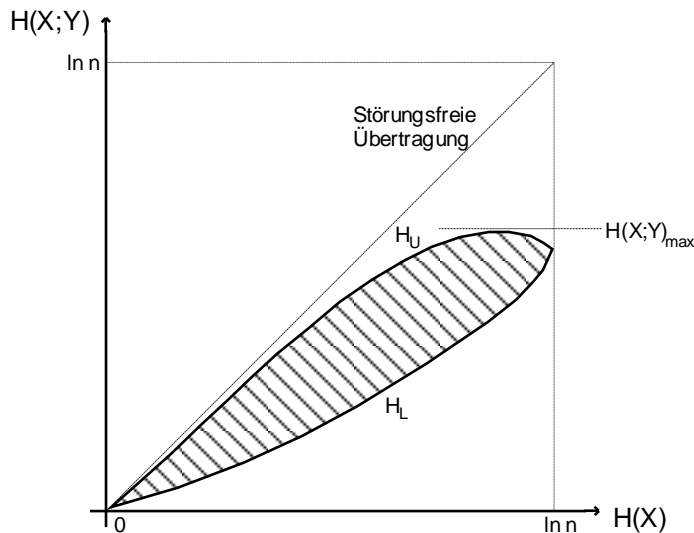
Allerdings untersuchte Shannon sein Modell mit Mitteln der Signaltheorie und erhält dabei für die Kanalkapazität (siehe folgenden Abschnitt) eine einfache Formel, die nur noch von einem Signal-Rauschabstand abhängig ist und gewisse physikalische Begebenheiten berücksichtigt.

Dies war nicht mein Ziel. Mein Modell und meine Untersuchungen beruhten auf der Vermutung, daß Datenworte mit einer einheitlichen Varianz in jeder Komponente durch additive Störungen weniger beeinflußt werden, als Datenworte mit unterschiedlichen Varianzen in den Komponenten. Denn bei einer Übertragung von Datenworten geht die Information von Komponenten mit sehr geringen Varianzen durch additive Störungen gänzlich verloren. Um zu zeigen, daß eine Datenorthonormalisierung für dieses Problem eine Lösung bietet, mußte ich einen sehr langen Weg durch die Informationstheorie und Signaltheorie zurücklegen. Meinem Ziel kam ich durch den Bericht [PLU93] ein ganzes Stück näher. Dort wurden genau die Untersuchungen durchgeführt, mit denen ich mich auch beschäftigte. Leider war die mathematische Herleitung von [PLU93] sehr unvollständig und teilweise schwer nachvollziehbar. Nachdem ich mich längere Zeit mit dem Bericht beschäftigt hatte, entschloß ich mich jetzt endgültig die mathematische Herleitung selbst durchzuführen, natürlich auf Basis der Berichte [SHA49] und [PLU93].

Im Laufe der Zeit fand ich mit Hilfe von [PAP91] meinen eigenen Lösungsweg, so daß die mathematische Herleitung im Abschnitt **Error! Reference source not found.** eine eigenständige Entwicklung von mir darstellt, die sich auf die Literatur [PAP91] und [SHA49] sowie [MIL90] stützt. Vor der eigentlichen mathematischen Herleitung werden noch einige wichtige informationstheoretische Grundbegriffe eingeführt.

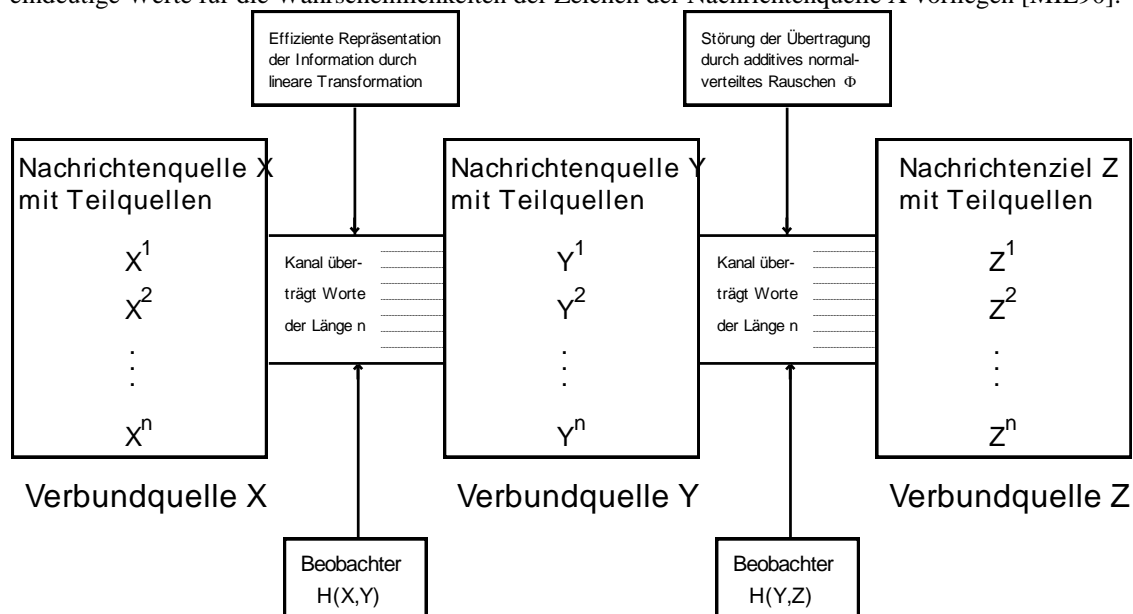
¹ Da Y als mittelwertfrei vorausgesetzt wird, gilt $\langle y \rangle = 0$. Damit vereinfacht sich die Varianz von Y zu $\sigma_Y^2 = \langle (y - \langle y \rangle)^2 \rangle = \langle y^2 \rangle$. Für die Varianz einer Komponente i eines Wortes $y \in Y$ gilt $\sigma_{y_i}^2 = \langle y_i^2 \rangle = P_{Y,i}$ und entspricht damit der mittleren Leistung von Y in der i -ten Komponente. Die mittlere Gesamtleistung von Y ist dann $P_Y = \sum_i P_{Y,i}$, siehe auch Definition **Error! Bookmark not defined.**

² In der klassischen Sichtweise der Datenorthonormalisierung werden die Varianzen der einzelnen Komponenten auf 1 normiert. Die hier angegebene Definition stellt eine Verallgemeinerung dar, die für den Spezialfall $P_Y = n$ der klassischen Sichtweise entspricht.



Zwischen der Entropie der Nachrichtenquelle $H(X)$ und der Transinformation $H(X;Y)$ besteht im allgemeinen kein Zusammenhang. Dennoch kann man für einen gegebenen Wert $H(X)$ einen unteren Wert H_L und einen oberen Wert H_U für die Transinformation angeben. Die Größe der Transinformation hängt davon ab, wie optimal die Nachrichtenquelle an den Übertragungskanal angepaßt wurde.

Die untere und obere Begrenzungslinie für $H(X;Y)$ treffen bei $H(X) = \ln(n)$ wieder zusammen, da in diesem Fall eindeutige Werte für die Wahrscheinlichkeiten der Zeichen der Nachrichtenquelle X vorliegen [MIL90].



Dieser Abschnitt zeigt, daß für ein bestimmtes informationstheoretisches Modell bei der Übertragung von Informationen über einen gestörten Kanal genau dann eine maximale Transinformation erreicht werden kann, wenn die zu übertragenden Daten orthonormalisiert sind. Dieses Modell wird im folgenden dargestellt.

Die Übertragung von Informationen über einen gestörten Kanal erfolgt dabei von der Nachrichtenquelle Y zum Nachrichtenziel Z , wobei die Daten der Nachrichtenquelle Y bereits linear transformiert wurden. Das Ziel ist es herzustellen, wie die Daten der Nachrichtenquelle Y beschaffen sein müssen, damit die Transinformation $H(Y;Z)$ ihren Maximalwert $H(Y;Z)_{\max}$ annimmt. Dies entspricht einer Anpassung der Nachrichtenquelle Y an den Übertragungskanal. Vor der eigentlichen Herleitung scheint es sinnvoll, alle Bedingungen genau zu beschreiben, die an das obige informationstheoretische Modell gestellt werden.

Bezieht man obige Sachverhalte nun auf die Hotelling Transformation, dann handelt es sich bei den Basisbildern um Eigenvektoren, und diese werden nach [WIN72] als Eigenbilder bezeichnet. Wie stellen sich Eigenbilder in der Realität dar ?

Um die Frage zu klären, wird in dieser Diplomarbeit eine Visualisierung von Eigenbildern in Graustufen durchgeführt. Dies entspricht dem Datenmaterial aus [WIN72], so daß es sich bei der folgenden Visualisierung um eine Überprüfung von Versuchsergebnissen handelt.

Zur Durchführung der Hotelling Transformation muß aus vorhandenem Bildmaterial eine $n^2 \times n^2$ Kovarianzmatrix C_X bestimmt und deren Eigenwerte und Eigenvektoren berechnet werden. Für die

Visualisierung werden die Eigenvektoren gemäß ihrer Eigenwerte absteigend sortiert. Dabei besitzt jeder der n^2 Eigenvektoren genau n^2 Komponenten.

Anschließend werden die Eigenvektoren in Teilstücke der Länge n zerlegt und in $n \times n$ Matrizen A'_{ij} mit $i, j = 1, 2, \dots, n$ zeilenweise abgelegt. Jeder einzelne der n^2 Eigenvektoren beschreibt damit eines der n^2 Eigenbilder A'_{ij} . Um eine aufwendige Verarbeitung von Bilddaten zu vermeiden, wird für die Visualisierung eine *synthetische* Korrelationsmatrix C_{XX} mit der Korrelationsfunktion aus [HAB71] berechnet. Diese Funktion wird im Abschnitt **Error! Reference source not found.** vorgestellt. Die für das Experiment entwickelte Software beruht auf der von mir selbständig erstellten objektorientierten Klassenbibliothek MAC, welche in Kapitel **Error! Reference source not found.** ausführlich beschrieben wird.

Für die folgende Abbildung wurden die Eigenbilder einer 64×64 Korrelationsmatrix C_{XX} berechnet. Zur Bestimmung der Korrelationsmatrix wurde die oben genannte Korrelationsfunktion mit den Parametern $\alpha = 0.125$ und $\beta = 0.249$ verwendet. Diese Parameter liefern eine gute Approximation für die Korrelationen des Bildes eines Kameramanns aus [HAB71]. Die den Eigenbildern zugeordneten Eigenwerte werden in der nachfolgenden Abbildung zeilenweise von links oben nach rechts unten kleiner.

Abbildung 0-1 Visualisierte Eigenbilder mit 128 Graustufen einer 64×64 Korrelationsmatrix C_{XX}
Nach diesem Versuch stellte sich heraus, daß die Ergebnisse meiner Diplomarbeit und aus [WIN72] nicht identisch sind. Sie sind aber durchaus ähnlich. Aussagen über die Differenzen der Ergebnisse sind nur schwer möglich, da das Experiment aus [WIN72] nur sehr ungenau beschrieben wurde.