Forschungsberichte des
Instituts für Phonetik und
Sprachliche Kommunikation
der Universität München
(FIPKM) 35 (1997) 3-15

DIE MENSCH-MASCHINE-KOMMUNIKATION[1]

Hans G. Tillmann

Institut für Phonetik und Sprachliche Kommunikation
Ludwig-Maximilians-Universität München
Schellingstr. 3
D-80799 Munich, Germany

Man-Machine-Communication

The first contribution of this volume of FIPKM is the manuscript of an invited talk given at the Technical University of Munich (Lehrstuhl für Mensch-Maschine-Kommunikation). The first part contains three very generally formulated considerations of principle concerning the psychophysical nature of human communication.

(i) It starts by considering what constitutes natural communication.

(ii) It discusses the interrelation of measurable time-functions on the one hand and meaningful categories on the other.

(iii) It introduces the role of the self and its personally given body in producing communicative actions.

The second part asks and discusses three questions:

(a) What is the role played by the naturally given double asymmetry that is so fundamental to human interactions?

(b) Why should one start to consider the very complex time relations of everything that can be perceived in a human world?

(c) Can machines learn to master the phonetically given phenomenological/categorical richness of any natural speech act?

[1]Vollständige (und durch Anmerkungen ergänzte) Fassung eines beim aktuellen Vortrag am 22.11.96 im Institut für Mensch-Maschine-Kommunikation der TU München auf 30 Minuten gekürzten Manuskripts, das in der hier vorliegenden Form Herrn Prof. Dr. Manfred Lang in Dankbarkeit für die freundschaftliche Zusammenarbeit der letzten Jahre gewidmet wurde.

Sehr verehrte Damen und Herren,

Spektabilis,

lieber Herr Lang,

mein Kollege in Kiel, der das am KTH in Stockholm entwickelte Infovox-Text-to-speech-System für das Deutsche eingerichtet hat, würde an meiner Stelle hier jetzt auf einen Knopf drücken, und der Text, den ich vortrage, würde nicht aus meinem Mund, sondern einem Lautpsrecher ertönen, und zwar so, daß eine leblose Maschine - und kein Mensch wie ich - auf phonetisch erstaunlich natürlich klingende Weise Herrn Prof. Lang Glück- und Segenswünsche für seine weitere Arbeit aussprechen würde. -- Also, sehr verehrter lieber Herr Lang, ich lasse es mir natürlich nicht nehmen, Ihnen selbst - höchstpersönlich und von ganzem Herzen - Gesundheit, Glück und Segen zu wünschen. Was ich hiermit tue, einfach, indem ich dies selbst sage.

Auf das hier angesprochene Selbst (und Redewendungen wie "ich lasse es mir nicht nehmen", "höchstpersönlich", "von ganzem Herzen") komme ich ausführlicher zurück; und auch das erste Beispiel für eine sogenannte Mensch-Maschine-Kommunikation habe ich hier nur darum an den Anfang gestellt, um darauf später, unter ganz verschiedenen Gesichtspunkten, wieder zurückzukommen.

Mein Thema versteht sich als Frage: Die Mensch-Maschine-Kommunikation - was ist darunter zu verstehen? Gibt es die - ich wiederhole ausdrücklich: *die* - Mensch-Maschine-Kommunikation überhaupt als einheitlichen Gegenstand[2] der wissenschaftlichen Forschung? Und wie läßt sich ein solcher Gegenstand in nicht nur rein technischen Worten begreifen und genauer kennzeichnen, d.h. explizit machen? Deshalb sei diese Frage im folgenden aus der Sicht eines Geisteswissenschaftlers behandelt, der die mit ihr verbundenen ingenieurswissenschaftlichen Aspekte sehr ernst nimmt. Mein Beitrag hat zwei Teile. Im ersten sollen - nach einleitenden Bemerkungen zum Wort Kommunikation und seinem Gebrauch im Deutschen - einige eher philosophische Grundfragen angeschnitten und in dem hier gegebenen begrenzten Zeitrahmen soweit entwickelt werden, daß der Hintergrund deutlich wird, vor dem ich im zweiten Teil dann drei speziellere Probleme ansprechen will, die meines Erachtens für den Kern der Mensch-Maschine-Kommunikation auch aus technischer Sicht zu bedenken sind:

a) Die Asymmetrie der natürlichen Kommunikation;

b) die Berücksichtigung der menschlichen Zeitverhältnisse;

c) das, was ich die Natürlichkeit nenne und am phänomenalen (bzw. kategorialen) Reichtum der Sprachäußerung beispielhaft erläutern möchte.

I. Das Wort Kommunikation

Im Englischen ist das Wort *communication* seit 1690 belegt, und zwar ausdrücklich in der Bedeutung: exchange of ideas[3]. Auch im Deutschen hat das Wort Kommunikation eine lange Tradition, doch es war schon nicht mehr in Gebrauch, als es Anfang der 50iger Jahre von Meyer-Eppler (im Anschluß an eine kleine Konferenz von wenigen Spezialisten am MIT) nach Deutschland

[2]Der Artikel *die* soll hier, linguistisch gesehen, also durchaus eine definite Kennzeichnung zum Ausdruck bringen.

[3]Für diesen Hinweis danke ich Frau Prof. Dr. Ina Schabert.

reimportiert wurde. Als 1957 ein hochgebildeter und sehr weltoffener Oberstudiendirektor mit Deutsch als Lehrfach beim Abitur erfuhr, daß sein Lieblingsschüler sich an der Universität Bonn für das Fach Kommunikationsforschung einschreiben würde, war dies eine absolut unverständliche Vokabel, mit der auch die anderen Studienräte nichts anzufangen wußten. Der Duden z.B. verzeichnet noch in der Auflage von 1961 beim Verb *kommunizieren*: "1. veraltet für mitteilen, 2. kirchlich: die Kommunikation empfangen, 3. zusammenhängen, in Verbindung stehen" (und es folgt ein Hinweis auf die kommunizierenden Röhren der Physik). Doch das Bild ändert sich dann rasch. Denn kaum mehr als 10 Jahre später hat das Wort Kommunikation auch im Deutschen eine geradezu unglaubliche argumentative Kraft gewonnen. So wird an der Universität - man denke hier nur an Chomskys damals aktuelle Kompetenz-Performanz-Dichothomie - heftigst darüber diskutiert, daß Sprache ja auch Kommunikation beinhalte, und sogar in den Abiturklassen werden Aufsatzthemen gestellt, die der Auseinandersetzung mit dem in der doppelten Verneinung etwas paradox klingenden Axiom von Watzlawick gewidmet sind, demzusolge es unmöglich sei, nicht zu kommunizieren. Nach etwa weiteren 10 Jahren hat sich diese Aufregung inzwischen längst wieder gelegt. Das Wort Kommunikation gewinnt eine völlig sachlich-neutrale Bedeutung, und so bezeichnet es heute ein recht breit gefächertes Sachgebiet, das zum wissenschaftlich zentralen Thema von sehr unterschiedlichen Fachdisziplinen geworden ist[4]. Besonders deutlich wird dies ja auch daran, daß sogar die Mensch-Maschine-Kommunikation zu einem aktuellen sehr ernst zu nehmenden Problem im Bereich der Ingenieurswissenschaften geworden ist, mit dem man sich im Hinblick auf industrielle Anwendungen systematisch und analytisch auseinandersetzen muß. Mit anderen Worten: Jeder gebildete Deutsche versteht sofort, was unter der Mensch-Maschine-Kommunikation zu verstehen ist und warum man dafür heute sogar einen Lehrstuhl braucht. - - Doch wie läßt sich das, was mit der so selbstverständlich gegebenen Bezeichnung schon so einfach vorgegeben ist, als Gegenstand der Untersuchung genauer kennzeichnen?

Das folgende ist ein erster Versuch, die Mensch-Maschine-Kommunikation erneut in Frage zu stellen. Dabei kann es nicht schon um endgültige Antworten gehen, sondern es ist mein Ziel, einen neuen und neuartigen Dialog zwischen Naturwissenschaften und Geisteswissenschaften, insbesondere aber zwischen der Ingenieurs- und der Sprachwissenschaft über dieses für die zukünftige Entwicklung unserer Kultur so grundlegende Thema anzuregen.

Vielleicht ist es hier nützlich, nach der Wortgeschichte in einem zweiten Rückblick auch die Entwicklung des Begriffs der Kommunikation kurz zu rekapitulieren. Zentral war zunächst die Tatsache, daß bei jeder de facto stattfindenden Kommunikation materielle Prozesse zu beobachten sind, die man unter dem Gesichtspunkt der Übertragung von Information analysieren kann. Für den Autor der 1959 zum erstenmal erschienenen *Grundlagen und Anwendungen der Informationstheorie*, Meyer-Eppler, ist dieser Gesichtspunkt ausschlaggebend, und dabei wird der Begriff der Kodierung zum zentralen Kern. Claude Shannon hatte ja schon vor 1950 in seiner *Mathematical Theory of Communication* bewiesen, daß es keine Nachricht gibt, die sich nicht effektiv kodieren und fehlerfrei übertragen läßt. Die Entdeckung, daß die Entropieformel (also die mit minus eins

[4]Der hier skizzierte Ablauf der Wortgeschichte von Kommunikation läßt sich, sprachtheoretisch, also in drei Schritten wie folgt zusammenfassen: es tritt Mitte der fünfziger Jahre neu ins Lexikon des gebildeten Deutschen mit der Markierung "von wenigen Spezialisten verwendetes Fachwort (im Kontext der Informationstheorie)"; etwa zehn Jahre später ist der pragmatische Kontext im Lexikon markiert als "intellektuelles Modewort (beinhaltend einen neu entdeckten, bisher vernachlässigten höchst relevanten Aspekt der menschlichen Existenz)"; nach weiteren zehn Jahren ist diese Aktualität verblaßt, so daß es im Lexikon fortan pragmatisch unmarkiert ist, was für die alltägliche Praxis bedeutet, daß Kommunikation ein einfaches Sachwort wie Haus und Brücke etc. geworden ist.

multiplizierte Summe, die wir über die Produkte aus den relativen Wahrscheinlichkeiten der einzelnen Symboleinheiten und den dyadischen Logorithmen eben dieser Wahrscheinlichkeiten bilden können) exakt die Anzahl der bits liefert, die für eine optimale Übertragung benötigt werden, - sie war 1949 eine wissenschaftliche Sensation. Wobei Shannon darüber hinaus mit dem zweiten Hauptsatz seiner mathematischen Kommunikationstheorie ja dann zudem auch noch beweisen konnte, daß es selbst bei jedem beliebig gestörten Kanal eine Kodierung gibt, mit der sich jede beliebige Nachricht auf optimale Weise fehlerfrei übertragen läßt.

Das Konzept von Kodierung und Dekodierung wurde schon Anfang der 50iger Jahre auch für die linguistische Theorie, insbesondere für die Phonologie in der von Roman Jakobson, Gunnar Fant und Morris Halle am MIT weiterentwickelten Form[5] zentral. Die linguistische Information bestand nach dieser Theorie aus einer Phonemkette, die beim Sprechvorgang in (artikulatorisch und akustisch definierte) distinktive binäre Merkmale kodiert wird, welche, vom akustischen Signal zum empfangenden Hörer übertragen, dort wieder in die linguistische, d.h. phonematische, Information dekodiert werden. Daß diese einfache Theorie falsch sein mußte, wurde klar, sobald die ersten entsprechend konstruierten Text-to-speech-Systeme so gut wie völlig unverständliche Sprache produzierten, und das Problem verschärfte sich dann noch, als man bei den ersten maschi-nellen Spracherkennungsversuchen feststellen mußte, daß die phonematisch kodierte Information keineswegs automatisch aus dem gegebenen Sprachsignal zu dekodieren war. Man hatte zwar eine schöne Theorie der Informationsübertragung, aber man konnte nicht mehr sagen, worin denn eigentlich die Information bestehe.

Von der frühen Informations- und Kommunikationstheorie muß man wohl nur die Erkenntnis festhalten, daß alle Information endlich ist, somit quantisierbar und in hinreichend vielen bit speicherbar. Doch dieser Aspekt tritt in den Hintergrund, sobald man sich für die Information als solche interessiert und die damit verbundenen Inhalte. Es geht dann weniger um die Verbindung zwischen den Kommunikatoren (sie wird als funktionierend vorausgesetzt), vielmehr geht es um die Kommunikatoren selbst und ihren Seelenzustand. Während es im Bereich der Humanwissenschaften bei der Kommunikation, insbesondere im therapeutischen Bereich, zunehmend um das Selbst und seine emotionale Befindlichkeit sowie eine mögliche Veränderung derselben geht, wird auf der technischen Seite zunehmend versucht, kognitive Funktionen (wie das Wahrnehmen und Identifizieren von Objekten, von Relationen, von Konstellationen und ganzen Szenarien einschließlich deren zeitlicher Veränderung und den daraus dann ableitbaren, darauf aufbaubaren sogenannten Wissensbasen) auf zwingende Weise zu modellieren, wobei solche Wissensbasen - ich denke hier z.B. ganz konkret an Programmiersprachen wie Prolog - Weltwissen repräsentieren, aus dem ganz automatisch Schlüsse abgeleitet werden können, die keineswegs trivial sind. Dies gilt insbesondere immer dann, wenn es um konkurrierende Alternativen geht, unter denen eine Entscheidung zu treffen ist.

Ich will diese stark vereinfachte Skizze, die absichtlich vieles offenläßt, hier abbrechen, um mich auf einen speziellen Punkt zu konzentrieren, nämlich die einzigartige Relation zwischem dem Selbst und der physikalisch gegebenen Welt. Und auch an dieser Relation interessiert mich im folgenden nur ein höchst spezieller Aspekt (alles andere bleibt außer Betracht), nämlich der Zusammenhang zwischen kommunikativ relevanten Kategorien und den beim Kommunikationsprozeß abzuleitenden Zeitfunktionen. Ich möchte dies am Beispiel von natürlichen Sprechakten veranschaulichen, aber das

[5]Roman Jakobson, Gunnar M. Fant, Morris Halle: Preliminaries to Speech Analysis. The Distinctive Features and Their Correlates. *9th Printing*, Cambrige, Mass., MIT-Press 1969

gilt natürlich sehr viel genereller und betrifft die Grundlagen jeder Art von menschlicher Kommunikation (und somit aller Anthropologie). Im Grunde geht es um Psychophysik in ihrer höchsten Form.

Rein datenmäßig betrachtet sind uns Wissenschaftlern die Kategorien durch Symbole immer nur indirekt gegeben, z.B. durch die orthographische Verschriftung einer im Sprechakt mitgeschnittenen Äußerung, während wir die digitalisierte Oszillation des Luftdrucks als Signal, in diesem Fall als Sprachsignal, in Form der jeweils empirisch gemessenen Zeitfunktion datenmäßig direkt abspeichern können.

Kategorien sind immer an das Selbst eines wahrnehmenden Subjekts gebunden, das sie durch symbolische Äußerungen für sich selbst und andere identifziert. Dieses Selbst hat in seinem Kern ein Ich. Zweitens: Jedes solche Ich ist an seinen eigenen Körper gebunden. Drittens: Mit seinem Körper schafft das Ich selbst- und fremdbeobachtbare Tatsachen, um auf diese Weise in Kommunikation zu treten. Beispielsweise spricht der Mund von Tillmann dem Kollegen Lang Glückwünsche aus. Für sich selbst betrachtet, haben die höchst komplex ablaufenden Artikulationsbewegungen (mit denen zudem das akustische Feld so gestört wird, daß wir über die Reaktion der Membran eines Mikrofons die Zeitfunktion des Sprachschalles erhalten) mit dem Glückwunsch als solchem inhaltlich gar nichts zu tun. Logisch gesehen sind die beiden Tatsachen, daß nämlich zum einen ein Glückwunsch geäußert (und verstanden) wurde, und daß dabei zum andern meßbare Zeitfunktionen realisiert wurden, völlig kontingent. Der individuelle, historisch einmalige Glückwunsch ist eine extrem komplexe Kategorie, und auch die sich beim Glückwünschen abspielenden materiellen Prozesse, z.B. in den Gehirnen von Tillmann, Lang und den anderen Ohrenzeugen, sind ebenfalls extrem komplex. Beide Komplexe stehen in keinem analytischen Verhältnis: Im Prinzip könnten nämlich die Beschreibungen des einen und des anderen in ihren jeweiligen Einzelheiten unabhängig voneinander wahr oder falsch sein. Und doch kann das eine nicht ohne das andere existieren. Feigl hat dies zum erstenmal zwingend analysiert und uns den Begriff der empirischen Identifikation geschenkt - ich selbst spreche deshalb in früheren Publikatonen vom Feiglschen Prinzip der empirischen Identifikation, z.B. zwischen linguistischen Einheiten und Sprachsignalen. Nur dieses Prinzip kann den Zusammenhang stiften zwischen naturwissenschaftlichen und geisteswissenschaftlichen Theorien, jedenfalls im linguistischen Bereich.

Im Hinblick auf MMK, die Mensch-Maschine-Kommunikation, muß ich noch einmal wenigstens kurz auch auf die menschlichen Emotionen und Gefühle zurückkommen. Das Interesse der Sprachphilosophen an Gefühlen, insbesondere dem Schmerz, liegt nach dem soeben Gesagten auf der Hand. Denn es gehört notwendig zum Schmerz, daß er nur existiert, sobald ihn jemand empfindet. Zwar sind Zahnschmerzen eine Kategorie, die über das Feiglsche Prinzip mit Vorgängen am kranken Zahn empirisch auf eine reproduzierbare Weise verknüpft ist, doch es gehört zur Existenzform dieser Tatsache, daß es ein Subjekt gibt, das sie "höchstpersönlich", in der Form des Selbst-Erlebens verspürt. Es ist immer ein Ich, das körperlichen Schmerz empfindet und dies - verbal oder nonverbal - zum Ausdruck bringt.

Die empirische Grenze zwischen Empfindungen und Signalen ist somit also selbst unmittelbar im Körperlichen begründet - und ist somit durch das körperlich gegebene Selbst der kommunizierenden Subjekte festgelegt. Auf diesem Hintergrund möchte ich argumentieren, daß die kategoriale Grenze der Psychophysik orthogonal zum gesamten Signalphonetischen Band und nicht nur zu einzelnen Bereichen der Signalmanifestation zu sehen ist. Diese Grenze separiert, jedenfalls methodisch, die immer nur persönlich erfahrbare Leiblichkeit nicht nur des eigenen, sondern auch der anderen beseelten Körper auf der einen Seite von der physischen Realität auf der anderen Seite und den dort rein physikalisch gegebenen Prozessen, die in der Form von Zeitfunktionen an den materiell

vorliegenden Strukturen abzuleiten sind. Und von diesen raumzeitlichen Strukturen können wir uns im Prinzip (natürlich nicht faktisch) immer auch eine Kopie hergestellt denken, die in jedem Atom mit dem Original übereinstimmt. Es ist die so konzipierbare Tatsache des auch atomar gegebenen eigenen Körpers (auch wenn sie uns im Alltag in der Regel völlig transphänomenal bleibt), die wir mit unserem in phänomenaler Form gegebenen Körper *empirisch* identifizieren. Nur dank der synchronen Koinzidenz von mentalen Kategorien und physischen Signalen kann der rein empirisch gegebene Zusammenhang genauer erforscht werden. Darauf basiert z.B. auch die maschinelle Spracherkennung und die maschinelle Sprachsynthese.[6]

Zum Schluß von römisch eins noch eine Nachbemerkung, um zu betonen, daß die hier angestellten Überlegungen keineswegs neu und nicht notwendig an das Wort Kommunikation geknüpft sind. Die Feststellung, daß die Information, auf die es ankommt, nicht übertragen, sondern vom Sender intendiert, vom Empfänger, der diese Intention erkennt, erschlossen wird, läßt sich radikalisieren. Das tut z.B. Karl Bühler, der 1927 in *Die Krise der Psychologie* eine Theorie der menschlichen Kommunikation entwickelt, ohne dieses Wort zu verwenden. Statt dessen ist vom Sender und vom Empfänger die Rede, einem Zweiersystem von Zeichengeber und Zeichenempfänger. Die Bühlersche Theorie wird nicht zuletzt gerade auch im Rahmen von MMK-Entwicklungen heute wieder aktuell, weil der Sender über ein Modell des Empfängers und seiner Welt, der Empfänger über ein Modell des Senders und seiner Welt verfügen muß, damit der Zeichenverkehr semantisch funktionieren kann.

Doch kann ein Computer Gefühle zeigen und so programmiert werden, daß er zu einem echten Zeichengeber und -nehmer im vollen Bühlerschen Sinne würde? Natürlich kann ich heute - das ist ja keine Utopie mehr - ein Computerprogramm schreiben, das ein Prologprädikat wie

Tillmann_vonGanzemHerzen_an_Lang_(Geburtstag (Glückwunsch (Artikulation...

(die schließenden Klammern lasse ich weg) erzeugt und dann auch ein Sprachsignal generiert, das als phonetisch perfekt gesprochener Glückwunsch - und vielleicht sogar darüber hinaus eindeutig mit der Stimme von Tillmann - wahrgenommen wird. Doch wie Sie, Herr Lang - und mit dieser Überlegung darf ich den ersten Teil beenden - auf ein solches Sprachsignal reagieren würden, wird wohl entscheidend davon abhängen, was in einem solchen Fall wirklich der Fall wäre. Wenn es für Sie zutrifft, daß Tillmann - so wie Dennis Klatt - einem Text-to-speech-System seine Stimme geliehen hätte und sich Ihnen nach einer Kehlkopfoperation auf diese Weise mitteilen würde, dann wäre das sicher ein Glückwunsch, der in seiner vollen Kategorialität vollzogen würde. Etwas völlig anderes wäre der Fall, wenn Tillmann die Geburtstage von Kollegen in seinem Rechner abgespeichert hätte, damit zum betreffenden Zeitpunkt eine entsprechende Voice-Mail abgeschickt wird. Ich bestreite, daß mit einem vom Betriebssystem des Rechners termingerecht produzierten Sprachsignal wirklich ein Glückwunsch ausgesprochen würde. In diesem Zusammenhang wird wohl auch deutlich, wie gut es ist, daß ihre Lehrstuhlbezeichnung Mensch-Maschine- und nicht etwa Mensch-Rechner-Kommunikation heißt.

[6]Beim ersten ist das Signal gegeben, gesucht wird die komplexe Kategorie, unter die es für die wahrnehmenden Subjekte einer Sprachgemeinschaft fällt, beim zweiten ist die symbolische, d.h. rein kategoriale Kennzeichnung einer Äußerung gegeben, gesucht ist ein Signal, das diese Kennzeichnung für die Mitglieder der Sprachgemeinschaft erfüllt.

II Drei ausgewählte Problemfelder

Die drei eingangs thematisch genannten Teilaspekte der menschlichen Kommunikation definieren Forschungsfelder, bei denen mir eine grundlegende und fächerübergreifende Diskussion nicht nur von der geisteswissenschaftlichen Seite aus, sondern sogar auch unter ingenieurswissenschaftlichen Gesichtspunkten wichtig scheint; jedenfalls dürfte es sich für die zukünftige Entwicklung der Informationstechnologie lohnen, wenn beide Seiten in ihrer wissenschaftlichen Arbeit hier gemeinsame Zielsetzungen definieren könnten. -- Von sehr grundlegender Bedeutung ist wohl in erster Linie:

a) Die natürliche Asymmetrie der Sender-Empfänger-Relation

In der üblichen Kommunikationstheorie wird, nicht zuletzt unter linguistischem Einfluß, der gemeinsame Symbolvorrat und die - in einer Sprachtheorie zu kennzeichnende - gemeinsam zu beherrschende Menge von Sprachstrukturen viel zu stark hervorgehoben. Die bei tatsächlich ablaufenden Kommunikationsakten so wesentliche wie notwendige Asymmetrie der Sender-Empfänger-Relation wird dabei vernachlässigt, und auch für die sprachliche Mensch-Maschine-Kommunikation wird sie m.E. immer noch zu wenig beachtet. Diese Asymmetrie, die schon für Bühler beim begrifflichen Ausbau des Zweiersystems von Zeichengeber und Zeichenempfänger konstitutiv war, tritt noch weitaus radikaler in Erscheinung bei Leibniz, der ja eine prinzipielle Erklärung dafür finden mußte, wie es bei fensterlosen Monaden möglich ist, daß sie miteinander kommunizieren.

Jedes unteilbare - individuelle! - Selbst, also die Monade, enthält bei Leibniz eine Repräsentation des gesamten Universums, in der Form von mehr oder weniger klaren Vorstellungen. Wenn also eine Monade mit einer anderen in Kommunikation tritt, so geschieht dies nach der Theorie von Leibniz immer jeweils in den beteiligten Monaden selbst, also z.B. derart, daß die Nicht-selbst-Monade des Hörers innerhalb der Selbst-Monade des Sprechers angesprochen wird, während sich die Selbst-Monade des Hörers von der in ihr repräsentierten Nicht-Selbst-Monade des aktuellen Sprechers angesprochen fühlt.

Die bei Leibniz völlig abstrakt abgeleitete Modellbildung der monadenintern ablaufenden Kommunikation läßt sich heute übrigens direkt auch hirnphysiologisch deuten. Denn wir gehen ja davon aus, daß z.B. in meinem Gehirn eine Art Bild[7] von Lang bestehen muß, und zugleich auch im Gehirn von Lang ein Bild von Tillmann, wenn das eine Selbst dem anderen Selbst Glückwünsche ausspricht, und dabei muß natürlich auch die körperliche Tätigkeit des Glückwünschens wiederum in beiden Monaden zur Darstellung kommen. Leibniz erklärt das faktische Zustandekommen von Kommunikation sehr elegant durch den Gleichlauf der Prozesse in den beteiligten Monaden (und dafür sorgt bei ihm die prästabilierte Harmonie). Bei den an einem Kommunikationsprozess beteiligten Gehirnen freilich sorgt die Erzeugung und Übertragung von Sprachsignalen und den vielen anderen Zeitfunktionen in einer *gemeinsam zugänglichen Form* nicht nur für die notwendige Synchronisation, sondern auch für das Zustandekommen der Selbst- und Fremdwahrnehmungen in den beteiligten Systemen.

Worauf es mir in diesem sehr komplex aufgespannten Rahmen allein ankommt, ist die darin begründete Asymmetrie. Sie gilt auf allen Ebenen, von der höchsten Ebene der kommunikativen Intentionen bis hinunter in die kleinsten phonetischen Details eines jeden natürlich vollzogenen Sprechaktes. Das sei hier kurz beleuchtet, und zwar zuerst unter einem semantischen "high-level"-

[7] Der Bildbegriff wird hier natürlich in einer völlig abstrakten, d.h. mathematischen Weise verwendet.

und dann auch an einem phonetischen "low-level"-Aspekt der lautsprachlichen Kommunikation.

Was zunächst die höchste rationale Ebene der menschlichen Kommunikation betrifft, so hat uns H. P. Grice darauf aufmerksam gemacht, wie komplex die Bedeutungsintention in Wirklichkeit strukturiert ist. Es genügt nämlich durchaus nicht, einfach zu sagen, daß ein Sprecher in einem Sprechakt des Glückwünschens seine Glückwünsche *in der Intention* formuliert, durch den Vollzug einer passenden Äußerung den Glückwunsch zum Ausdruck zu bringen. Vielmehr ist damit die *zweite, ganz andere Intention* verknüpft, daß der Hörer erkennen möge, daß der Sprecher diese Intention mit dem Vollzug der Äußerung verfolgt. Und auch dieses reicht noch nicht aus für eine vollständige Kennzeichnung des rationalen Glückwünschens. Denn der Sprecher hat auch noch die wiederum *ganz andere dritte Intention*, daß ihn der Hörer versteht, weil er die soeben genannte zweite Intention als solche erkennt (daß der Sprecher nämlich mit seiner Äußerung die besagte erste Absicht verbindet). Lassen sich auf einer Maschine diese von Grice aufgedeckten Sprecherintentionen - und die damit verbundenen immer nur im zwischenmenschlichen Kommunikationsakt entstehenden "nichtnatürlichen" Bedeutungen - implementieren? Ich lasse die Frage hier offen. Denn sie muß von beiden Seiten, der sprachphilosophischen wie der maschinellen, sehr sorgfältig analysiert werden. Jedenfalls wird man hier gewissen Zukunftsvisionen im Bereich der künstlichen Intelligenz, wie sie von einigen bekannten AI-Protagonisten artikuliert werden, auf den Grund gehen müssen.

Die asymmetrische Relation, die grundsätzlich zwischen dem immer individuellen Sender und den jeweils gegebenen individuellen Empfängern besteht, sei noch an einem weiteren, ganz konkreten phonetischen Beispiel verdeutlicht. Dazu betrachten wir die individuellen Merkmale, die jeder Sprecher rein akustisch seinen lautsprachlichen Äußerungen auf phonetisch unverwechselbare Weise aufprägt. Daß diese Merkmale vom Sprecher selbst völlig anders wahrgenommen werden als von seinen Hörern, erfährt jeder, der seine eigene Stimme zum erstenmal in der Aufzeichnung, also als Hörer, zur Kenntnis nimmt. Grundsätzlich nimmt jedes in einer gemeinsamen Welt handelnde Individuum diese Welt allein aufgrund seines aktiven Handelns ganz anders - d.h. in einer anderen Form - wahr als alle anderen nur beobachtenden Individuen. Wie verschieden und unterschiedlich z.B. sind jetzt, in diesem Augenblick, die Wahrnehmungen des hier Vortragenden von denen seiner Zuhörer.

Was also auf der untersten Ebene der lautsprachlichen Kommunikation die rein phonetische Wahrnehmung betrifft, so spricht vieles sogar dafür, daß unsere eigene individuelle Charakteristik bei der Bewertung der fremden Charakteristik eine Rolle spielt, derart, daß z.B. der Sopran eine andere Sopran- oder Tenorstimme anders hört als ein Tenor den anderen Tenor oder einen Sopran, Alt, Bariton oder Baß.

Auf diesem sehr komplexen Hintergrund wird man die Frage klären müssen: Trifft es zu, daß, weil der Maschine die Individualität der menschlichen Kommunikatoren fehlt, die bei jeder Mensch-Maschine-Kommunikation - insbesondere auf intentionalem Niveau - gegebene Leibnizsche Asymmetrie bei der Mensch-Maschine-Kommunikation zu vernachlässigen ist? Oder muß in einem tieferen nichttechnischen Sinne zwischen Mensch-Maschine- und Maschine-Mensch-Kommunikation, also zwischen MeMaK und MaMeK, unterschieden werden? Daß rein technisch die Richtung des Informationsflusses bei der Entwicklung von Anwendungen eine entscheidende Rolle spielt, heißt ja nicht, daß sich die doppelte Asymmetrie der menschlichen Kommunikation auf die Mensch-Maschine-Kommunikation vererbt. Ich denke jedenfalls, daß Sie, lieber Herr Kollege Lang, darauf verzichten werden, unter dem Dach der MMK zwei eigene Memak- und Mamek-Abteilungen zu etablieren. Die Asymmetrie wäre demnach nur einfach, also zwischen Mensch und Maschine, gegeben und nicht, im Leibnizschen Sinne, doppelt.

Jedoch muß man ganz klar sehen, daß diese beim Menschen aus den jeweils individuellen Selbst- und Fremdbeobachtungskomponenten sich ergebende Asymmetrie bei den Memak- und Mamek-Prozessen nur soweit vernachlässigt werden dürfen, wie man darauf verzichten wird, jedem maschinellen System eine eigene menschenähnliche Individualität und Identität zu inkorporieren. Die Frage, die hier offenbleibt, ist die strukturelle Tiefe und die semantische Reichweite der gegebenenfalls zu versuchenden "Inkorporation von menschenähnlicher Individualität", falls dies überhaupt zukünftig eine Rolle spielen wird.

Beim nächsten Problemfeld geht es mir um die Berücksichtigung der für das menschliche Wahrnehmen, Denken und Handeln relevanten Zeitstrukturen, kurz:

(b) Die menschlichen Zeitverhältnisse

Dieses zweite Problemfeld ist weitaus weniger theoretisch und abstrakt als das letzte, doch darum keineswegs von einer weniger grundlegenden Bedeutung. Da die empirische Gleichsetzung von kategorialen Bewußtseinsinhalten und objektiv analysierbaren Signalparametern allein durch die synchrone Koinzidenz gewährleistet ist, mit der beides gleichzeitig zu beobachten ist, wird die Aufmerksamkeit der Forschung ganz von selbst - und in der Regel naiv-unkritisch[8] - auf die jeweils gegebenen Zeitverhältnisse gelenkt. Man denke hier nur an das Sonagramm von lautsprachlichen Äußerungen, also die sogenannten "visible-speech-Diagramme", bei denen die im Kurzzeitspektrogramm analysierten Signalzustände in ihrem zeitlichen Ablauf zur Darstellung kommen.[9]

Was nun einerseits die hier angesprochenen Zeitverhältnisse und andererseits die großen Erfolge der heute praktizierten automatischen Spracherkennung betrifft, so beobachten wir hier die höchst paradoxe Situation, daß der Erfolg der letzteren auf Verfahren der Mustererkennung wie ‘Dynamic Programming’, ‘Time Warping’ und ‘Hidden Markov Modelling’ zurückzuführen ist, deren Witz gerade darin besteht, die tatsächlich gegebenen Zeitverhältnisse zu ignorieren. Denn das eigentliche Ziel dieser Verfahren besteht darin, die Unterschiede in den einzelnen Lautdauern einer jeweils gegebenen Äußerung als "störende Variation" auszuschalten und damit als mögliche weitere Informationsquelle zu vernachlässigen. Bei der Links-rechts-Sequenz einer Hidden-Markov-Modellierung bleibt man halt solange im Zustand eines Knotens, bis die Emissionswahrscheinlichkeiten auf einen neuen Zustand führen, also einen der nächsten Knoten weiter rechts.[10]

Um dafür zu werben, die menschlichen Zeitverhältnisse speziell im Bereich der Sprachtechnologie - die ja durch Ihren Lehrstuhl auf höchstem internationalem Niveau vertreten ist - bei der zukünftigen Arbeit stärker zu berücksichtigen, möchte ich zuerst an einem sehr einfachen

[8]Darauf kommen wir in der folgenden Anmerkung zurück.

[9]Weit davon entfernt, das Feiglsche Prinzip zu reflektieren, hören wir den Betrachter eines Sonagramms z.B. sagen "hier haben wir das [a]", und dabei identifiziert er die Kategorie einer zunächst nur phänomenal gegebenen alphabetischen Einheit mit einem bestimmten Lautsegment im Zeitfrequenz-Diagramm, das sich allein nur aufgrund seiner charakteristischen Formantstruktur (und durch nichts anderes) abgrenzen läßt.

[10]Daß hiermit die Ausschaltung der menschlichen Zeitverhältnisse zum Erfolg führt, ist natürlich nur scheinbar paradox. Denn man kann leicht zeigen, daß dies nur bei sogenannter Lesesprache gut funktioniert. Bei lebendiger Spontansprache versagen die klassischen Methoden, wie es sich ja auch dann beim Verbmobil-Projekt des BMBF mit einem dramatischen Rückgang der Erkennungsraten auf zunächst weniger als 50% gezeigt hat.

Beispiel belegen, welch ausschlaggebende Rolle die Zeitverhältnisse bei der Realisierung einer zusammenhängend gesprochenen Wortkette tatsächlich spielen, um dann auf die Tatsache einzugehen, daß die mit den unterschiedlichen Zeitverhältnissen gegebenen phonetischen Ausprägungen von lexikalisch vorgegebenen Einheiten für den sprechenden und hörenden Menschen im Kommunikationsprozeß zwar völlig verborgen bleiben, für das damit verbundene Sprachverstehen aber dennoch unverzichtbar sind.

Zuerst also das einfache Beispiel: Die folgenden Wörter kann ich in der Form einer Liste sprechen, indem ich sie einzeln aufzähle: "Kohl", "sagte", "Strauß", "wird" (usw.). Daß Sie hier eine Wortliste hören, also das Aufzählen einzelner Wörter, obwohl ich zwischen den Wortsignalen gar keine Signalpause realisiert habe, liegt allein daran, daß die Lautdauern in der wortfinalen Silbenkoda nach dem Prinzip des 'prepausal lengthening' zeitlich gedehnt sowie mit einer abfallenden F0-Kontur realisiert worden sind. Werden die hier aufgezählten Wörter nicht phonetisch isoliert, sondern syntaktisch miteinander verknüpft gesprochen, so macht es einen großen Unterschied, ob ich die besagte Lautdehnung am Wortende auf das erste oder das zweite Wort derselben Kette anwende. Ohne daß ich das vom Sprachzentrum im Gehirn zu verfolgende Sprachsignal mit einer physikalischen Pause versehe, hören Sie im ersten Fall

"Kohl, sagte Strauß, wird nie Kanzler",

Und im zweiten

"Kohl sagte: Strauß wird nie Kanzler"

Natürlich hätte ich hier, etwa im Sprechstil des Lesens, nach "Kohl" oder nach "sagte" ohne weiteres auch eine Pause einlegen können.[11]

Uns bleiben die tatsächlich gegebenen Zeitverhältnisse, mit denen wir die Wörter unserer Sprache bei zusammenhängender Rede realisieren und dabei phonetisch auf sehr unterschiedliche Weise ausprägen, beim natürlichen Kommunikationsprozeß in aller Regel verborgen. Dies läßt sich möglicherweise auf die Tatsache zurückführen, daß wir in unserem mentalen Lexikon die Wörter in genau der Form abspeichern, in der wir sie üblicherweise zitieren und demonstrieren. Es mag strittig sein, ob wir mental die Wörter in der sehr klar artikulierten Form abgespeichert haben, mit der wir sie isoliert produzieren, als typische Einwort-Satzäußerung (wie z.B. im Sprachunterricht bei Sprechakten des Vorführens einzelner Wörter). Doch es steht fest, daß die Wörter nur bei solchen speziellen Sprechakten ihre volle segmentelle Ausprägung aufweisen. Und es ist zweitens noch einmal festzuhalten, daß die tatsächlich gegebenen Zeitverhältnisse, die sich bei fließender Rede ergeben, so gut wie absolut unauffällig bleiben. Wir ahnen nicht, daß von der phonematisch definierten Segmentstruktur einer lexikalischen Einheit nach Koartikulation, Assimilation, Elision und weiteren artikulatorischen Verschleifungen in vielen Fällen, insbesondere bei spontan getätigten Sprachäußerungen, oft nur sehr wenig übrig bleibt.[12]

---

[11]Doch läßt sich, wie beim Vortrag vorgeführt, die Wahrnehmung einer Pause allein durch den zeitlichen Ablauf einer Artikulationsbewegung auslösen, die im Tempo abgebremst wird, ohne daß es zum Stillstand kommt. Dieses in jüngster Zeit als phonetisches "prepausal lengthening" wiederentdeckte Prinzip war übrigens schon im letzten Jahrhundert in seiner sehr viel generelleren Form bekannt und wurde vergessen. Es ist grundsätzlich mit dem Ende, d.h. mit dem "Auslaufen" von menschlichen Handlungen verbunden. Vgl. dazu meinen Beitrag zur frühen Experimental- und Instrumentalphonetik in der ENCYCLOPEDIA OF LANGUAGE AND LINGUISTICS (S. 3082-3095).

[12]In diesem Zusammenhang wäre ich gerne auf mein Paradebeispiel, auf das Wort "und" und seine vielen möglichen Reduktionsformen bei zusammenhängender Rede eingegangen. Die volle Ausprägung ergibt sich nur bei einem Fragesatz wie "Und?". Nur in der hier gezeigten alphabetisch expliziten Form besitzt es fünf Laute; es beginnt mit dem 'Glottalstop', dem

Oft bleibt von der mental repräsentierten Lautstruktur nur noch ein prosodischer Rest, etwa in der Form von minimalen Daueunterschieden, die jedoch völlig ausreichen, damit der Hörer das betreffende Wort noch eindeutig wahrnimmt. Tilgt man am Signal den phonetischen Rest, der von der vollen Zitierform übriggeblieben ist, so kann unter Umständen nicht nur eine ungrammatische Äußerung resultieren, sondern sogar eine semantisch falsche Interpretation.[13]

Mein letztes Problemfeld

c) Natürlichkeit und kategorialer/phänomenaler Reichtum

läßt sich am besten auf dem Hintergrund der heutigen Text-to-speech-Systeme entwickeln, also am Beispiel der sprechenden Maschinen. Denn damit ist genau das angesprochen, was die künstliche Sprachschallsynthese bis heute vermissen läßt.

Was die Natürlichkeit betrifft, so beruht diese auf dem phänomenalen Reichtum, mit dem wir jede von Menschen in natürlichen Sprechakten produzierte Sprachäußerung wahrnehmen. Ich verwende hier kategorial und phänomenal nebeneinander, weil das phänomenal gegebene Erscheinungsbild kategorial fundiert ist; was sich allein darin erweist, daß wir beim Vergleich zweier Äußerungen in der Lage sind, die kategoriale Differenz zu bestimmen. Diese kann null

---

harten Stimmeinsatz, der im Unterschied zum weichen durch die Orthographie im Deutschen nicht ausgedrückt wird (man vergleiche nur das erste Lautsegment im Minimalpaar /?unth/ und /hunth/, also den orthographisch als "und" und "Hund" zu schreibenden Wörtern). Selbst bei einer sorgfältigen Aussprache des Zahlwortes 99 [sprich: "neun und neunzig"] wird der t-Plosiv ohne h-Aspiration realisiert. Vielmehr geht die Sprechbewegung der Zungenspitze ohne die Unterbrechung einer Verschlußlösung zum homorgan gebildeten n-Nasal des Folgewortes über. So läßt sich in phonetischer Lautschrift (hier in SAMPA-Notation) eine ganze Reihe möglicher Reduktionen festhalten:

[QUnth] -> [QUnt] -> [Unt] -> [Un] -> [@n] -> [n:]-> [:] -> [].

Zwischen 9,90 [sprich "neun_neunzig] und dem Wort 99 wird im Extremfall [sprich:"neun_en_neunzig", "neun_n_neunzig", "neun_neunzig"] kein Unterschied mehr gemacht. Die Aussprache mit @ (dem reduzierten Neutralvokal [sprich "neun_en_neunzig"]) ist bei Nachrichtensprechern absolut üblich.

[13]In meinem Beitrag zur Kohler-Festschrift (vgl. dazu Abb. 1 S. 154 in "Kleine Phonetik und Große Phonetik", Phonetica 52, 144-159, 1995) bin ich etwas ausführlicher auf ein Beispiel eingegangen, bei dem das Wort "und" nur noch durch eine minimale glottale Restgeste ausgedrückt wird. Bei einem spontansprachlichen Satz "Ich will nach Hamburg und möchte...." aus dem PHONDAT-Korpus war das Wort "und" zwar verschluckt, doch zeigte sich im Oszillogramm am Anfang des darauffolgenden Wortes "möchte" nach den ersten Stimmtonperioden des m-Lautes eine kleine Delle in der Einhüllenden, die sich über drei bis vier Stimmtonperioden erstreckt und die offensichtlich durch eine kurzfristig intermittierende Entspannung der Stimmlippen hervorgerufen wird. Bei genauerer Inspektion zeigt sich, daß die Stimmlippen am Ende der kurzen Entspannungsphase nicht nur erneut anziehen, sondern auch ihre Grundfrequenz um etwa einen ganzen Ton erhöhen. Dieser Tonsprung bleibt dem Ohr zunächst völlig verborgen. Er wird erst deutlich hörbar, sobald man das Schallsegment des m-Lautes in der besagten Delle in zwei Teile schneidet und die beiden kurzen Segmente in isolierter Abfolge darbietet. Wird die Gesamtäußerung ohne das kurze initiale Segmentstück des m-Lautes dargeboten, wird der ungrammatische Satz "Ich will nach Hamburg möchte...." wahrgenommen. Nur mit der besagten etwa 30 msec langen Initialphase des m-Lautes wird aus dem Tonsprung der tatsächlich geäußerte Satz "Ich will nach Hamburg und möchte ...". Das Wort "und" wird also hier durch ein an sich ganz unauffälliges, nur unter dem Mikroskop des Oszillogramms zu entdeckendes minimales Entspannungs- und erneutes Spannungsverhalten der Glottis ausgedrückt. Man kann dies vielleicht auch so umschreiben, daß die Glottis mit dem Wort "und" hier noch die große Entstimmungsgeste assoziert, also die ballistisch ablaufende Öffnungsbewegung, mit der bei alphabetisch expliziter Aussprache der stimmlos aspirierte finale t-Plosiv produziert wird. Und was die menschlichen Zeitverhältnisse betrifft, so genügt bei dem hier angeführten Beispiel offenbar der kleine prosodische Reflex einer solchen segmentalen Entstimmungsgeste, um beim Hörer, im Kontext der gegebenen Wortsequenz, die lexikalische Einheit "und" zu evozieren.

werden[14]. Andererseits führt jede systematisch durchgeführte Kategorisierung auf eine fachsprachlich immer nur reduziert repräsentierbare Empirie, die in dieser Form dem tatsächlich gegebenen Reichtum nie wirklich gerecht werden kann. Auf diesen Punkt, der für die zukünftige Entwicklung der Mensch-Maschine-Kommunikation möglicherweise eine wichtige Rolle spielen wird, komme ich abschließend noch einmal zurück.

Der kategoriale Reichtum zeigt im Kontext der maschinellen Sprach- und Informationsverarbeitung ganz unterschiedliche Gesichter. Zum einen wird darauf verwiesen, daß die vielen individuellen und situativen Merkmale, die das Erscheinungsbild jeder natürlich produzierten Sprachäußerung prägen, als Störgrößen aufzufassen sind, sobald es um eine robuste automatische Spracherkennung geht. Auf der anderen Seite aber sind diese sogenannten Störgrößen auf eine Variabilität in den lautsprachlichen Äußerungen zurückzuführen, die ihrerseits wertvolle Information enthält. Wenn man für die maschinelle Sprachverarbeitung eine Kategorisierung von möglichen Sprechsituatonen bzw. von möglichen Sprechertypen oder sogar individuellen Sprechern heranzieht, ließe sich im Prinzip auch die damit verbundene Information berücksichtigen und aus dem jeweils vorliegenden Sprachsignal ableiten. So könnte also beispielsweise die durch die individuellen Sprecher hervorgerufene Variabilität in den Signaleigenschaften, welche sich bei der Spracherkennung so störend auswirkt, z.B. für die Sprechererkennung ausgenutzt werden. In diesem Fall würde also auch auf der technischen Seite aus der Störung eine wertvolle Information.

Der phänomenale Reichtum, durch den sich jede uns begegnende vom Menschen produzierte Sprachäußerung auszeichnet, läßt sich natürlich nicht so ohne weiteres in kategorialer Form erfassen, also systematisch kennzeichnen und somit explizit machen. Ich möchte das nun noch am Beispiel der unterschiedlichen Sprechstile verdeutlichen.

Die heute übliche Unterscheidung zwischen einem förmlichen und einem ungezwungenen Sprechstil (formal vs. casual speech) reicht bei weitem nicht aus. Auch genügt es nicht, den Sprechstil als Funktion einer Distanz zwischen Sprecher und Hörer zu erklären: Bei geringer sozialer Distanz z.B. ergibt sich ein familiärer Ton, bei größerer räumlicher Distanz wird lauter gesprochen usw. Ich möchte mich im folgenden allein auf die Vielzahl möglicher Sprechstile beschränken (ohne auf den dabei jeweils gegebenen Ausdruck der Sprechstimme einzugehen[15]).

Vergleichen wir eine spontansprachlich produzierte Äußerung mit einer solchen, die deren Text nur reproduziert, so tritt der Unterschied zwischen Original und Imitation sogleich sehr deutlich in Erscheinung. Wir erkennen sofort, ob es sich bei einer Äußerung um eine sogenannte authentisch spontane oder eine nur nachgestellte handelt.

Selbst wenn das Original von einem Schauspieler - wie man dann sagt - perfekt imitiert wird, können wir dennoch beim unmittelbaren Vergleich in der Regel jede Schauspielerstimme von der eines Laien unterscheiden. Wenn ein Schauspieler den anderen imitiert, wird es natürlich problematisch. Doch dies ist wiederum ein kategorial ganz eigener Fall. Ein dem direkt entsprechender weiterer Fall liegt nämlich auch bei der sogenannten authentischen Lesesprache vor.

---

[14]Dies ist z.B. der Fall, wenn wir hören, daß ein Minimalpaar aus identischen Wörtern besteht, die zudem vom selben Sprecher auf identische Weise geäußert werden.

[15]Der stimmliche Ausdruck ist eine sehr komplexe Funktion der inneren und äußeren Situation, in der sich der Produzent einer lautsprachlichen Äußerung gerade befindet. Diese äußere und innere Situation kann sich im Ablauf einer Äußerung sehr plötzlich ändern, und das hören wir sofort. Ein bekanntes Beispiel sind die den Sprachschall plötzlich überlagernden "lächelnden Lippen".

In der Regel erkennen wir ja sofort, ob ein Text frei gesprochen oder nur gelesen ist.

Doch gibt es auch bei der Lesesprache wieder extrem viele kategorial zu unterscheidende Varianten. So hat jeder Hochschulprofessor seinen eigenen Vortragsstil. Und dabei verfügt auch hier jeder einzelne wieder über viele kategorial differenzierbare, also eindeutig wahrnehmbare Untervarianten. Bei mir selbst fällt auf, wie stark sich mein rein phonetischer Sprechstil ändert, wenn ich nicht, wie hier, einen vorbereiteten Text vortrage, sondern Folien kommentiere. In beiden Fällen wird vorher viel Zeit investiert, damit der Sprecher sich nachher kurz fassen kann. Demgegenüber entstehen bei der spontanen Rede die textlichen Formulierungen erst beim Sprechen, und auch das kann man immer hören. Ich habe als Student in einem überfüllten kleinen Hörsaal direkt hinter einem berühmten Philosophen gestanden und konnte den vorgetragenen Text im Manuskript mitlesen. Bei diesem Vortragsstil hatte der Hörer dann allerdings den Eindruck, als würden die Gedanken wirklich erst im Augenblick der Äußerung entstehen, und dieser Kontrast zwischen dem so einfach vorgegebenen Text und der mit geradezu schauspielerischen Mitteln vorgetragenen Anstrengung beim Ausdrücken von Gedanken war in der Tat, rein phonetisch, höchst spannend zu beobachten.

Wieviele Sprechstile sich auch immer im direkten Vergleich unterscheiden lassen, zu jedem einzelnen gibt es dann die dazu passenden situativen und individuellen Merkmale, und so erhalten wir schon rein phonetisch einen Reichtum, der sich in seinem phänomenalen Umfang einer kategorialen Systematisierbarkeit zu entziehen scheint. Vor diesem Hintergrund muß am Ende die vielleicht wichtigste Frage gestellt werden: Wollen wir wirklich, daß uns Maschinen mit dem ganzen phänomenalen Reichtum der lebendigen menschlichen Kommunikationspartner entgegentreten? Oder ist es vielleicht gerade sogar dieser rein phänomenal gegebene, wissenschaftlich nicht bis auf den letzten Rest durchzukategorisierende Reichtum in seinem ganzen Umfang, von dem wir Menschen uns von jeder uns zu Diensten stehenden Maschine unterscheiden? Besteht also am Ende der eigentliche Kategoriensprung zwischen Mensch und Maschine gerade in diesem phänomenalen Reichtum, für deren Produktion es immer eines lebendigen, persönlich beseelten Körpers bedarf? Hier läge dann die durch nichts zu überwindende Grenze zwischen maschineller Künstlichkeit und menschlicher Natürlichkeit.

Auf der anderen Seite werden wir Wissenschaftler uns keine Grenze setzen beim Versuch, die im phänomenalen Reichtum gegebene Information auch maschinell auf vollständige Weise zu verarbeiten. Dieses Ziel ist uns vorgegeben. Erreichen werden wir es nie. Doch die Grenzen, die zwischen den Menschen und den Maschinen bestehen bleiben, lassen sich dank unserer Arbeit auf eine immer genauere Weise bestimmen. Und nur darauf kommt es wohl am Ende - anthropologisch gesehen - auch an.

Ich danke für Ihre Geduld.