PHONETIK UND SPRACHSIGNALVERARBEITUNG

Hans G. Tillmann

IPS (LMU München)
tillmann@phonetik.uni-muenchen.de

Abstract: Dies ist in schriftlicher Form ein Ausblick auf die geplanten mündlichen Ausführungen zum Thema *Phonetik und Sprachsignalverarbeitung*. Er besteht wie auch der Vortag aus zwei separaten Teilen. Im ersten geht es zunächst um die Frage, was hier unter Phonetik zu verstehen ist. Deshalb möchte ich darlegen, welche Bedeutung ich mit dieser Bezeichnung im Titel verbinde. Sie steht weniger für ein isolierbares Fachgebiet, sondern eher für das zu untersuchende sehr komplexe Sachgebiet, das mit der gesprochenen Sprache gegeben ist. In diesem Sinne ist Phonetik ein interdisziplinär zu betrachtendes Forschungsfeld, das in der Kooperation von vielen Fachdisziplinen gemeinsam zu bearbeiten ist. Was demgegenüber unter dem Fachgebiet SSV zu verstehen ist, muss hier nicht erläutert werden. Deshalb geht es im zweiten Teil meines Vortrages darum, an ausgewählten Beispielen zu zeigen, welche verschiedene Rollen Sprachsignale und ihre Verarbeitung im Rahmen der zuvor gekennzeichneten Phonetik im Einzelnen spielen und in der Zukunft noch spielen können.

1 Erster Teil: Phonetik und lautsprachlicher Kommunikationsprozess

Ohne Sprachsignale gäbe es keine Phonetik. Doch gab es von Anfang an – und gibt es auch heute noch – eine Phonetik ohne jeden expliziten Bezug auf diese Sprachsignale und ihre Verarbeitung. Als wichtigstes Beispiel sei die sogenannte „Ohrenphonetik“ erwähnt, die vor allem in England mit „ear training“ sowie „narrow and broad transcription“ (mit den Buchstaben der IPA) eine lang gepflegte Tradition besitzt. Dieser Punkt wird in den folgenden Ausführungen eine zentrale Bedeutung gewinnen, wenn es darum geht, den Begriff von den phonetischen Tatsachen, ohne die kein Sprechakt ablaufen kann, einzuführen und zu erläutern.

1.1 Paul Menzerath und Werner Meyer-Eppler: Ein kurzer Rückblick auf die kommunikationstheoretische Wende im Paradigma der phonetischen Sprachforschung an der Universität Bonn nach 1945

Mit diesem kurzen Rückblick möchte ich beginnen, weil ich die ehrenvolle Einladung (mit den hiermit präsentierten Ausführungen beisteuern zu dürfen zur 26. ESSV-Konferenz) ausdrücklich der positiven Reaktion auf einen Beitrag zur Heike-Festschrift verdanke, der im letzten Jahr erschienen ist und in dem ich den hier angesprochenen Paradigmenwechsel in der phonetischen Sprachforschung ausführlich thematisiert habe (S.1-14 im Band 45 der Halleschen Schriften zur Sprechwissenschaft und Phonetik; um mich darauf zu beziehen, werde ich (statt „a. a. O.“ hinzuschreiben) einfach Heike-Beitrag sagen). Da ich das alles hier nicht wiederholen will, beschränke ich mich darauf, drei Aspekte hervorzuheben, damit ich in der folgenden Argumentation unten darauf zurückkommen kann.

Vor der Bonner Wende war die Phonetik das an vielen Universitäten etablierte Fachgebiet, das (neben anderen Aufgaben, wie z.B. im Ausspracheunterricht) insbesondere auch für die Linguistik die Sprachdaten zu erfassen und bereitzustellen hatte, die für die theoretische Weiterverarbeitung benötigt wurden. Hierbei ging es sowohl um die systematische Entwicklung allgemeiner Sprachtheorien als auch um die Theorie von Einzelsprachen

einschließlich ihrer Dialekte und Soziolekte. Nach der kommunikationstheoretischen Wende kehrt sich das Verhältnis um. Die von Linguisten formulierten Sprachtheorien werden daraufhin untersucht, inwieweit und in welcher Form sie zur Erklärung des tatsächlichen Funktionierens von lautsprachlichen Kommunikationsprozessen beitragen können.

Zweitens muss gesagt werden, dass die klassische Phonetik, die sich ganz selbstverständlich als „Lehre von den Sprachlauten" verstehen durfte, durch Befunde der sich neu entwickelnden Instrumental- und Experimentalphonetik vor der Bonner Wende zunehmend in eine Krise geraten war, was (wie unten unter 2.2 noch zu erläutern sein wird) dazu führte, dass es dann schließlich um 1930 zu einer strikten Trennung von Phonetik und Phonologie kam. Was diese Trennung von naturwissenschaftlicher Phonetik und geisteswissenschaftlicher Phonologie betrifft, so gelingt es Meyer-Eppler, die beiden Disziplinen im Rahmen der von ihm selbst weiterentwickelten mathematischen Informationstheorie wiederzuvereinigen. Dies zeigt sich am Modell der Kommunikationskette (mit den beiden Kästchen für den Sender und den Empfänger, die miteinander durch einen Strich für den Kanal verbunden sind). Die kategorial gesicherten linguistischen Einheiten befinden sich im „gemeinsamen Symbolvorrat von Sender und Empfänger", so wie er als Venn-Diagramm in den beiden sich überschneidenden Kreisen über dem Kanalstrich verbildlicht wird. Die Phonetik befasst sich dann mit der Kodierung und Dekodierung der zu übertragenden Symbole. Somit aber treten die Sprachsignale und ihre Verarbeitung durch Mensch und Maschine ins Zentrum auch der phonetischen Forschung. Das sei noch durch ein Zitat belegt.

Als Herausgeber der Reihe „Informationstheorie und Kybernetik in Einzeldarstellungen", in der als erster Band 1959 sein Buch „Grundlagen und Anwendungen der Informationstheorie" erscheint, schreibt Meyer-Eppler im Vorwort zur Reihe:

> „Aufgabe der Informationstheorie ist es, die Kommunikation von Mensch zu Mensch, die sich als Zeichenverkehr manifestiert, oder die Kommunikation des Menschen mit der Welt, die auf eine Beobachtung hinausläuft, einer quantitativen und strukturellen Erfassung zugänglich zu machen, während die Kybernetik als „science of relations" (N.WIENER) die regulären Verhaltensweisen von hochkomplexen energetisierten „Systemen" (d.h. von informationsverarbeitenden „Maschinen", Lebewesen und Gruppen von Lebewesen) mit mathematischen Methoden studiert" (ohne Seitenzahl).

Drittens soll die interessante Tatsache festgehalten werden, dass die Protagonisten der Bonner Wende aus ganz anderen Fachgebieten zur Phonetik kamen. Der in Psychologie habilitierte Menzerath hatte in Bonn ein Labor für Instrumental- und Experimentalphonetik aufgebaut (und bereits im Jahrzehnt vor der Bonner Wende seine berühmte Endeckung der ***Koartikulation*** von Konsonanten und Vokalen im Kopf der Silbe sowie der ***Steuerung*** des Vokals durch den Folgekonsonanten im Reim der Silbe publiziert). Der viel jüngere Meyer-Eppler hatte sich mit einer mathematischen Arbeit über ***Periodenforschung*** in Physik habilitiert und brachte auf Einladung von Menzerath den Arbeitsschwerpunkt der elektronischen Nachrichtentechnik in die Bonner Phonetik. Nach einem gemeinsamen Besuch von Sprachforschungszentren in den USA wurde aus dem experimentalphonetischen Labor ein „Institut für Phonetik und Kommunikationsforschung".

1.2 Sprechakt und Äußerung

Es gibt keinen natürlichen (d.h. mündlich realisierten) Sprechakt ohne die lautsprachliche Äußerung, und damit wird das Funktionieren von Sprechakten zur phonetisch zu untersuchenden Tatsache. Aus meiner Sicht ist mit dieser so trivialen wie grundlegenden Feststellung das Sachgebiet der Phonetik auf eine klare und eindeutige Weise vorgegeben.

> Es sei nicht verschwiegen, dass dieser Sicht der Dinge von prominenter Seite widersprochen wurde: Als ich bei der Begrüßung zu einer Veranstaltung an der VIU den berühmten ersten Satz, mit dem

Trubetzkoy seine 1939 erschienenen Prinzipien der Phonologie einleitet, zitiert hatte („Jedes Mal, wenn ein Mensch einem anderen etwas sagt, liegt ein Sprechakt vor"), um ihn in die Form meiner Tatsachenbehauptung zu übersetzen, schüttelte in der ersten Reihe ein bekannter Sprachphilosoph protestierend seinen Kopf. Auf meine Frage, was er gegen diese Feststellung einzuwenden habe, schüttelte er erneut seinen Kopf und äußerte gleichzeitig in wörtlicher Rede die Gegenfrage: „Ist das kein Sprechakt?". Wir haben uns dann darauf geeinigt, dass aus sprachphilosophischer Sicht der Sprechakt der Verneinung auch durch eine nonverbale Aktion wie Kopf- oder Zeigefingerschütteln ausgedrückt werden kann, insofern nämlich diese Aktion vom Gesprächspartner im Sinne einer Verneinung verstanden wird. (Als Konsequenz dieser kleinen Auseinandersetzung schränke ich die für die Phonetik relevanten natürlichen Sprechakte auf die Menge der mündlich realisierten, also in wörtlicher Rede wiederzugebenden ein.)

Im Vorblick auf meine Ausführungen unten zum Thema „Äußerung und Bedeutung" sei noch auf einen nicht nur aus sprachphilosophischer Sicht semantisch interessanten Unterschied im zweimaligen Kopfschütteln von Prof. Manfred Pinkal aufmerksam gemacht. Beim erstenmal hat das Kopfschütteln die Bedeutung des Neinsagens, jedenfalls wird damit auf unmittelbar verständliche Weise ein Widerspruch ausgedrückt, beim zweiten Mal jedoch wird das erste Kopfschütteln gewissermaßen zitiert und zum Zwecke seiner Demonstration auf identische (d.h. kategoriale) Weise reproduziert, damit sich darauf die Frage beziehen lässt, ob dies kein Sprechakt sei. In diesem Sinne werden wir unten zwischen heteronymen und autonymen phonetischen Äußerungen unterscheiden.

Damit ist schon angedeutet, dass das zu untersuchende Sachgebiet der phonetischen Sprachforschung sich nicht beschränken lässt auf die mit der mündlichen Realisierung gegebenen Äußerungen per se, so wie sie in einer weiten oder engen phonetischen Transkription festgehalten werden können. Es gilt auch in umgekehrter Richtung: Es gibt keine reguläre, d.h. natürlich produzierte Äußerung ohne einen funktionierenden Sprachakt. Unter diesem Gesichtspunkt rückt das phonetisch erfolgreiche Funktionieren von Sprechakten ins Zentrum der Forschung.

1.3 Äußerung und Bedeutung

Ein jedes Mal, wenn ein Mensch einem anderen etwas sagt, indem er mit seiner lautlich wahrnehmbaren Äußerung ganz bestimmte phonetische Tatsachen in die Welt setzt, treten diese Tatsachen als solche in den Hintergrund, indem der am Sprechakt beteiligte Sprecher und seine Hörer die Äußerung als solche automatisch transzendieren in die konkreten oder abstrakten Situationen hinein, die, wenn Sprecher und Hörer dieselbe Sprache beherrschen, mit den lexikalisch verfügbaren Mitteln semantisch beherrschbar werden. Dies ist der Fall, wenn der Hörer versteht, was der Sprecher ihm sagen will.

Die Absichten des Sprechers spielen für den logisch orientierten Sprechaktphilosophen die zentrale Rolle. Das Bedeutungsproblem wird zunächst auf die Frage zurückgeführt, wie es möglich sei, dass jemand (in Anführungszeichen) „x" äußert, um damit x auszudrücken, und die Antwort, die dann noch weiter ausgeführt wird, lautet im ersten Schritt, dass der Hörer den Sprecher versteht, indem er die Intentionen erkennt, die dieser mit seiner Äußerung verbindet. Diese Intentionen spielen auch bei der rein phonetischen Betrachtung der vom Sprecher in einem Sprechakt erzeugten Äußerung eine wesentliche Rolle.

Neben das Sachgebiet der Phonetik tritt bei der Erforschung der menschlichen Sprache das zweite Sachgebiet der Semantik. Und es wird in der zukünftigen Sprachforschung entscheidend darauf ankommen, die sehr komplexen und sehr engen Querbeziehungen, die zwischen diesen beiden Sachbereichen, der Phonetik und Semantik von Sprechakten bestehen, aufzudecken und theoretisch explizit zu machen.

Auf das sehr verwickelte Sachgebiet der Semantik und die verschiedenen aktuell darin arbeitenden linguistischen und nichtlinguistischen (d. h. logischen, psychologischen und anthropologischen) Forschungsdisziplinen kann hier nicht eingegangen werden. Ausdrücklich festhalten möchte ich

jedoch, dass die phonetische Seite mit den Anführungszeichen in der sprachphilosophischen Fragestellung als völlig trivial vorausgesetzt und damit gleichsam bagatellisiert wird. Die zukünftige Forschung wird zeigen, dass die neuronalen Prozesse der Sprachsignalverarbeitung in den Gehirnen von Sprechern und Hörern sich von den semantischen Prozessen nicht wesentlich unterscheiden.

Was die phonetischen Tatsachen betrifft, so kann ein Sprecher mit seiner Äußerung viele unterschiedliche Intentionen verbinden. Im ersten Fall produziert er die Äußerung, um damit etwas ganz anderes zu meinen und dies dem Hörer mitzuteilen. In diesem Fall spreche ich von einer ***heteronym*** produzierten Äußerung.

Ein Sprecher kann aber auch jederzeit mit seinem Sprechakt keine andere Absicht verbinden als die, seine Äußerung als solche selbst zu meinen. Er liefert dann ein kategorial reproduzierbares Muster derselben. In diesem Fall sage ich, dass die Äußerung ***autonym*** gemeint ist, als Repräsentant ihrer eigenen Identität. (Ein berühmtes Beispiel liefert jene Sprachaufnahme von Daniel Jones, bei welcher er für den phonetischen Transkriptionunterricht der Reihe nach seine Kardinalvokale produziert, zunächst in lang angehaltener Form und dann als mehrfach repetierte Kurzvokale.)

Als ein weiteres Beispiel sei das berühmte Experiment von Wolfgang Köhler genannt, der seinen Versuchspersonen in Berlin die Aufgabe stellte, den beiden als „maluma“ und „takete“ autonym zu reproduzierenden Äußerungen zwei Arten von Zeichnungen zuzuordnen, eine mit Ecken und Spitzen, die andere mit runden Formen. Die Zuordnung war eindeutig, da „maluma“ eher als rund, „takete“ eher als spitz wahrgenommen wurde.

Wir halten erstens fest, dass die Kategorie der phonetischen Tatsachen, die ein Mensch mit seiner Äußerung in die Welt setzt, allein durch phonetisch autonyme Sprechakte zu vermitteln ist und auch nur so effektiv vermittelt werden kann. Damit werden autonym reproduzierte phonetische Tatsachen selbst zum Gegenstand und damit zum Inhalt der sprachlichen Kommunikation. Insofern macht es wenig Sinn, zwischen objektsprachlicher und meta-sprachlicher Kommunikation zu unterscheiden. Der einzige Unterschied zur „normalen Kommunikation“ liegt in der Intention des Sprechers.

Heteronym und autonym produzierte Äußerungen unterscheiden sich in ihrer *phonetischen* Form. Deshalb erkennen wir in der Regel am Telefon sofort, ob wir den Angerufenen selbst antreffen oder nur die Stimme vom Anrufbeantworter hören. Hier liegt ein Forschungsfeld noch weitgehend brach und wartet darauf, dass die phonetischen Tatsachen aufgedeckt werden, die zeigen, wie diese beiden Formen von Sprechakten sich z.B. am akustischen Sprachsignal unterscheiden lassen.

Zusammenfassend dürfen wir feststellen, dass sich bei der Erforschung des tatsächlichen Funktionierens von Sprechakten Phonetik und Semantik nicht ganz so einfach voneinander trennen lassen, wie es manche Vertreter der philosophischen Sprechakttheorie wahrhaben wollen. Phonetik und Semantik sind zwei miteinander verwandte Sachgebiete, die für die empirische Sprachforschung eng zusammengehören und deshalb gerade auch bezüglich der rein in der Sache bestehenden gegenseitigen Beziehungen untersucht werden müssen. Nur wenn die Sachgebiete Phonetik und Semantik zu einem gemeinsamen Forschungsprojekt von interdisziplinär zusammenarbeitenden Fachleuten der einschlägigen Einzeldisziplinen werden, kann sich eine anthropologisch umfassende komplexe Theorie über das tatsächliche Funktionieren von Sprechakten beim lautsprachlichen Kommunikationsprozess in einer empirisch verifizierbaren Form entwickeln. Auf der kategorialen Seite dürfen wir nämlich nicht vergessen, dass eine von einem Sprecher produzierte lautsprachlich Äußerung, auditiv wahrgenommen, im Bewusstsein des Hörers nicht anders in Erscheinung tritt als der Sprecher beim Akt des Sprechens oder als auch die ganze übrige Welt.

2 Zweiter Teil : Sprachsignale als „Phonetische Tatsachen“

Mehr noch als die frühen Experimentalphonetiker misstraute auch Meyer-Eppler den natürlichen Wahrnehmungsleistungen der Ohrenphonetik als Grundlage der Entwicklung einer streng wissenschaftlichen Theorie der lautsprachlichen Kommunikation. Aus diesem Grunde führte er expressis verbis die Instanz des externen Beobachters ein:

> „Die in einer Kommunikationskette sich abspielenden Prozesse können nur von einem außerhalb der Kette stehenden *externen Beobachter* hinreichend exakt beschrieben werden, einem Beobachter, dem *sämtliche* Glieder der Kette zugänglich sind. Zur Beschreibung des Beobachteten bedient er sich einer wissenschaftlichen *Metasprache*, die nicht mit der zwischen dem Expedienten und Perzipienten vereinbarten *Objektsprache* übereinstimmt. Alle informationstheoretischen Ausführungen der folgenden Kapitel sind in der Metasprache des externen Beobachters formuliert“ (Meyer-Eppler 1959, S.5f).

Um den Konflikten, die sich aus Unterscheidungen wie der von Objekt- und Metasprache ergeben können, aus dem Weg zu gehen, habe ich in *Tillmann mit Mansell (1980)* (kurz: *TmM*) darauf hingewiesen, dass der Begriff der Äußerung (engl.: utterance) auf eine systematische Weise mehrdeutig ist. Um dies zum Ausdruck zu bringen, wurde eine klare terminologische Trennung zwischen zwei mit jedem Sprechakt gegebenen Empirien eingeführt und zwischen wahrnehmbaren ***Phonetischen Ereignissen*** und mess- und damit digitalisierbaren ***Phonetischen Vorgängen*** unterschieden. Vorher bereits hatte ich ganz in diesem Sinne auf Einladung der Fakultät, die in München einen neuen Lehrstuhl für „Allgemeine Phonologie und Phonetik“ zu besetzen hatte, in einem Vortrag über „Symbolphonetik und Signalphonetik“ die Frage thematisiert, wie man Sprachsignale als Zeitfunktionen und die wahrgenommen Äußerungen in ihrer symbolisch beschreibbaren Form in einer empirisch verifizierbaren Zuordnungstheorie miteinander verknüpfen kann.

> Diese beiden Teilfachbezeichnungen haben sich auch ohne zitierbare Quelle allein durch Mundpropaganda erstaunlich rasch durchgesetzt. Nur Kohler spricht bei der Signalphonetik von „Messphonetik“, was die Sache ja auch nicht schlecht trifft.

Im Rahmen der Entwicklung großer Sprachdatenbanken für das VERBMOBIL-Projekt (1993-2000) habe ich mit dem Begriff der Phonetischen Tatsache auf eine vehemente Diskussion der Frage reagiert, ob und wann spontansprachliche Daten auch wirklich authentische Daten sind. Die Antwort bestand einfach darin festzulegen, dass es sich bei den bereitgestellten Daten um ***Phonetische Tatsachen*** handelt, sobald den digitalisierten Sprachsignalen von den im PHONDAT-Projekt verbundenen Phonetikinstituten bezüglich vorher festgelegter Kategorien eine eindeutige Annotation zugeordnet wurde, aus der u.a. hervorgeht, wie ein Wort der deutschen Sprache im Vergleich zu der im Aussprachewörter notierten Zitierform beim gegebenen Sprachsignal phonetisch tatsächlich ausgesprochen wurde (vgl. dazu auch Wolfgang Hess et.al.1995 sowie vorher Tillmann et.al. 1993). Damit wird die Trennung von Ereignissen und Vorgängen in gewisser Weise durch die eine empirische Identifikation überbrückt.

2.0 Das Prinzip der empirischen Identifikation

Wir müssen davon ausgehen, dass der Zusammenhang zwischen dem Sprachsignal einer Äußerung und der kategorialen Seite rein empirisch ist. Beide Seiten der Zuordnung sind logisch unabhängig voneinander, weil die Beschreibung des einen und des anderen unabhängig voneinander wahr oder falsch sein können. Somit gibt es zwischen den Zeitfunktionen von Sprachsignalen und den Kategorien, die symbolsprachlich festgehalten werden, keinen analytischen Zusammenhang. Wer unter einer Lupe das Oszillogramm in der Rille einer Schallplatte studiert, kann prima vista nicht wissen, was er beim Abspielen hören wird.

Doch auf der rein empirischen Seite ist der Zusammenhang eines Signals mit der zugeordneten kategorialen Kennzeichnung in den hier interessierenden Fällen in der Regel immer sehr streng. Die Zuornung ist experimentell reproduzierbar. Wird das Oszillogramm auf einen Lautsprecher gegeben, dann werden wir auf der kategorialen Seite jedes Mal Thomas Mann einen bestimmten Satz aus Felix Krull lesen hören oder z.B. den Walter Gieseking, der einen ganz bestimmten Takt aus einer Mozartsonate spielt.

> Herbert Feigl, ein Mitglied des Wiener Kreises, hat in "The 'Mental' and the 'Physical' für den Fall der Geltung des Prinzip der empirischen Identifikation das Gleichheitszeichen eingeführt und am Beispiel der auf der zahnmedizinischen Seite zu beobachteten Vorgänge am kranken Zahn erläutert. Diese werden auch vom Zahnarzt gleichgesetzt mit den vom Patienten kategorial erlebten Zahnschmerzen. Diese Gleichsetzung von Vorgängen und Ereignissen kann ich nicht akzeptieren. Darauf komme ich bei der lautphysiologischen Sprachlauttheorie noch einmal zurück.
>
> Und so gibt es auch manchen Phonetiker, der bei der Betrachtung eines Breitbandsonagramms die Feststellung trifft: „Dieser Vokal hier ist ein tiefes a".

Eine Zuordnungsrelation gilt, wenn sie experimentell reproduzierbar ist. Nur weil uns die phonetischen Tatsachen auf diese Weise gegeben sind, wird es möglich, die Zuordnungsrelation durch wissenschaftliche Forschungsarbeit in der Form von empirisch verifizierbaren Zuordnungstheorien explizit zu machen. Im akustischen Signalbereich finden wir z.B. eine solche Zuordnungstheorie gewissermaßen rein technisch implementiert, sobald eine speech-to-text- oder eine text-to-speech-soft- oder hardware zufriedenstellend funktioniert. Das liegt auf der Hand.

Wenn es jedoch darum geht, das tatsächliche Funktionieren von natürlichen Sprechakten in allen wesentlichen Aspekten und somit allen wesentlichen damit verbundenen phonetischen Tatsachen explizit zu machen, so steckt die interdisziplinäre Sprachforschung noch in den Kinderschuhen. Andererseits wissen wir schon sehr viel. Ich nenne hier ein einfaches Beispiel, weil ich später darauf zurückkommen muss.

> Wie lässt es sich z.B. erklären, dass wir bei einem dargebotenen akustischen Signal sofort erkennen, dass es sich um gesprochene Sprache handelt, auch wenn wir die Sprache des Sprechers vorher noch nie gehört haben. Die Begründung liegt in dem, was ich in TmM die A-, B- und C-Prosodie genannt habe. Die damit gegebene prosodische Form ist universell und bei jeder einzelsprachlich geregeltem Äußerung phonetisch realisiert. Unter die A-Prosodie fällt der relativ langsame F0-Verlauf, der sich als Sprechmelodie auditiv gut verfolgen lässt. Unter die schneller ablaufende B-Prosodie fällt die rhythmische Silbensequenz, die sich noch mitklopfen lässt. Viel zu rasch für Auge und Ohr jedoch laufen die mit den Vokalen koartikulierten Sprechbewegungen der C-Prosodie ab, mit der die Konsonanten am Silbenrand reproduziert werden.

2.0.1 Kategoriale Einheit und die Vielheit der Vorgänge

Der mentalen Einheit auf der kategorialen Seite der Äußerung entspricht auf der physikalischen Seite eine Vielheit von Manifestationsbereichen, in denen sich die Vorgänge abspielen, die sich in der Form von Zeitfunktionen unterschiedlicher Dimensionalität als Sprachsignale entlang der Kommunikationskette zwischen sprechendem und hörendem Nervensystem mehr oder weniger einfach oder aufwendig darstellen lassen. Wie lässt sich diese Vielheit mit der Idee von verifizierbaren Zuordnungstheorien in Einklang bringen? Diese Frage kann heute niemand beantworten. Doch ich möchte im folgenden auf einige Szenarien der bisherigen Forschungsgeschichte eingehen, die zeigen, wie man das Problem von Einheit und Vielheit in den Griff bekommen könnte.

> In den folgenden Abschnitten 2.1 bis 2.4 fragen wir, welche interessanten Sachverhalte im Kontext des empirischen Identifikationsprinzips durch die phonetische Sprachforschung bei der Untersuchung der durch Zeitfunktionen darstellbaren Vorgänge in der Kommunikationskette aufgedeckt werden konnten. Wir beginnen mit der Akustischen Phonetik, schauen dann zurück auf die Artikulatorische Phonetik als „Lehre von den Sprachlauten". Drittens gehen wir auf die phonetische Kanaltheorie ein,

die erklärt, wie die phonetisch relevante Information durch das akustische Sprachsignal übertragen wird, um schließlich zu diskutieren, welchen Sinn es macht, die lautsprachliche Kommunikationskette in mehrere Meyer-Epplersche Beobachtungsketten aufzuteilen.

2.1 Sprachsignalverarbeitung und Akustische Phonetik

Auch für die Akustische Phonetik hat der Übergang von der analogen zur digitalen SSV eine entscheidende Rolle gespielt, weil nämlich damit der Computer zum wichtigsten Instrument der Instrumentalphonetik wurde. Zunächst aber wurde das analoge Sprachsignal in seiner spektralen Darstellung mit dem Sonagraphen schon kurz nach Bonner Wende verfügbar und im Sinne der empirischen Zuordnungstheorie unter die Lupe genommen. Das sei an zwei Beispielen erläutert, einer viel zu optimistischen akustischen Merkmaltheorie zur Kennzeichnung der Phoneme aller Sprachen (von der nach den anfänglich großen Erfolgen heute keiner mehr spricht) und der von den Haskinsforschern entwickelten Methode einer "Analyse durch Synthese", die bis heute in den Experimenten zur kategorialen Wahrnehmung eingesetzt wird.

Wenn man bedenkt, dass der gemeinsame Symbolvorrat von Sender und Empfänger für Meyer-Eppler aus Phonemsequenzen bestand, dann versteht man auch den Optimismus, mit dem Roman Jakobson und seine beiden damaligen Studenten Fant und Halle auf die Idee kommen konnten, anhand der Betrachtung von Sonagrammen in einer einsemestrigen Seminarveranstaltung herauszufinden, wie sich die Phoneme aller Sprachen der Welt im akustischen Manifestationsbereich durch binäre akustische Merkmale kennzeichnen lassen. Eine binäre Kodierung schien auch linguistisch interessant, weil sich z.B. mit +/-vokalisch und +/-konsonantisch dann statt der üblichen zwei Lautklassen zwei weitere darstellen lassen. Die Gleitlaute sind weder richtige Vokale noch richtige Konsonanten, und die Sonoranten (zumal wenn sie einen konsonantischen Silbenkern bilden) haben auch gewisse vokalische Eigenschaften und können somit auch bestimmte vokalische Funktionen übernehmen. Die distinctive-feature-Theorie hatte einen unglaublich großen Erfolg und wurde auch in der empirischen Sprachforschung sehr ernst genommen. Doch wie wenig man auf der Seite der technischen Anwendungen damit anfangen konnte, zeigte sich schon bei den ersten Versuchen, mit den Methoden der analogen Sprachsignalverarbeitung einen automatischen Phonemerkenner zu bauen. Diese Form einer phonetischen Zuordnungstheorie konnte empirisch nicht verifiziert werden.

Eine wichtige phonetische Erkenntnis, auf die ich im Heikebeitrag ausführlich eingegangen bin, bestand darin, dass die kategoriale Seite einer Äußerung in erster Linie durch ihre lexikalischen Einheiten bestimmt wird. Damit wurden das Wort und seine phonetisch auf so unterschiedliche Weise zu realisierenden Formen zu einem zentralen Thema. Weil bei der kognitiven Verarbeitung des akustischen Sprachsignals top-down-Prozesse dominieren, fällt es dem Hörer gar nicht auf, wenn ein Satz, der mit den Wörtern „ich bin mit dem Auto …" beginnt, durch eine Lautfolge wie „chbimim …" geäußert wird. Sobald man Sonagrammen eine phonetische Transkription zuordnete, erkannte man, das darin ganze Wörter einfach von der Bildfläche verschwunden waren.

Zur gleichen Zeit wurde demgegenüber mithilfe der analogen Sprachsignalverarbeitung eine sehr viel interessantere und tatsächlich auch empirisch verifizierbare Zuordnungstheorie auf der Syntheseseite konzipiert. Um eine Lesemaschine für Blinde zu entwickeln wurde in New York mit dem Geld einer reichen Stiftung eine nach dem Stifter benannte Forschergruppe finanziert, die sich unter der Leitung von Frank Cooper an die Arbeit machte, um diese Lesemaschine für Blinde auch wirklich in Angriff zu nehmen. Aus diesem ersten *text-to-speech*-Projekt sind die Haskins Laboratories hervorgegangen.

Auch hier war das „Lesen von Sonagrammen" der Ausgangspunkt, um auf der technischen Seite ein Verfahren zu realisieren, mit dem handgemalte Formantverläufe, also künstlich erzeugte Sonagramme, optisch abgetastet und hörbar gemacht werden konnten. Die F0-Frequenz des Stimmtones konnte durch eine entsprechend oszillierende Lichtquelle er-

zeugt werden. So entstand das berühmte Pattern-Playback-System, das nicht nur ein großer technischer Erfolg war, sondern dann sogar weltweit zu einem ganz neuen Paradigma der phonetischen Sprachforschung geführt hat, der „Analyse durch Synthese“. So konnte man gleich am Anfang, also lange vor der digitalen Erzeugung von künstlichen Sprachstimuli, mit dem Pattern-play-back-Verfahren die Tatsache aufdecken, dass kurze Formant-transitionen am Anfang oder Ende eines isolierten Vokalsegments dem hörenden Nervensystem die Information über die konsonantischen Artikulationsstellen am Rand des vokalischen Silbenkerns vermitteln. In diesem Zusammenhang wurde das Phänomen der Kategorialen Wahrnehmung entdeckt und experimentell erforschbar gemacht.

> Dem Analyse-durch-Synthese-Ansatz verdankt die phonetische Sprachforschung wichtigste Erkenntnisse über das tatsächlichen Funktionieren von Sprechakten. Das betrifft zum Beispiel das Phänomen der sogenannten kategorialen Wahrnehmung und die Entwicklung des entsprechenden Testverfahren, bei dem durch die systematische Variation eines akustischen Parameters wie z.B. der Startfrequenz des zweiten Formanten oder der Dauer der *voice-onset-time* ein theoretisches Kontinuum definiert wird, das dann mit äquidistanten Teststimuli belegt werden kann. Diese werden, entsprechend randomisiert, der Versuchsperson zweifach präsentiert: erstens in einem Identifikationstest, wo die Kategorie des jeweils einzeln dargebotenen Stimulus benannt und angekreuzt wird, sowie zweitens in einem Diskriminationstest, bei dem im Kontinuum benachbarte Paare dargeboten werden. Hier wird die Versuchsperson instruiert mitzuteilen, ob ein gehörtes Paar gleich oder verschieden ist.

Was die Ermittlung von phonetischen Tatsachen im Rahmen einer empirisch expliziten Zuordnungstheorie betrifft, so gelingt es mit diesem Testverfahren nicht nur, kategorial vorgegebne Einheiten auf einem physikalisch gegebenen Kontinuum zu extensionalisieren, sondern auf diesem theoretischen Kontinuum eines gegebenen Signalparameters auch die Kategoriengrenzen zu ermitteln. Denn die Diskriminationskurve zeigt gewissermaßen wie ein Finger genau an den Stellen des Kontinuums, an denen es bei der Identitätskurve zu einem raschen Kategorienwechsel kommt, ein spitzes Maximum. Auf der kategorialen Seite werden die Stimuli im Diskriminationstest innerhalb der *extensionalisierten Kategorie* demgegenüber als gleich bewertet.

Ohne auf die Vielzahl der Einzelergebnisse eingehen zu müssen, lässt sich an dieser Stelle verdeutlichen, dass mit der ***digitalen*** Sprachsignalverarbeitung der Computer zum wichtigsten Instrument der Instrumentalphonetik wurde und dass dies zu einer neuen Blüte der experimentalphonetischen Sprachforschung geführt hat. Denn mit der digitalen Technik wurde es plötzlich sehr einfach, alle interessierenden Signalparameter als unabhängige Testvariable zu isolieren, systematisch zu verändern und alle denkbaren Teststimuli herzustellen. Doch die kategoriale Wahrnehmung funktioniert nicht nur bei so elementaren Beispielen wie der Artikulationsstelle von Plosiven oder ihrer Stimm-haftigkeit, sondern auch bei so komplexen Kategorien wie der individuellen Sprecher-identität. Darauf werde ich im mündlichen Vortrag zurückkommen und ein entsprechendes Kontinuum vorführen.

> Welche großen Fortschritte die Phonetik dem Einsatz der digitalen Sprachtechnologie in der Entwicklung von Werkzeugen und deren Einsatz in der Forschungsarbeit im einzelnen verdankt, sei hier wenigstens am Rande angedeutet und durch drei Beispiele aus dem Münchner Institut belegt. Das bekannte *MAUS*-System, das von Hartmut Pfitzinger entwickelte Verfahren zur Messung des mit der A-Prosodie variierenden *lokalen Sprechtempos* sowie die *elektromagnetische dreidimensionale Artikulographie* hätten ohne den sehr aufwendigen Einsatz von digitaler SSV nie verwirklicht werden können.

Ursprünglich wollte ich an dieser Stelle noch etwas ausführlicher auf den unterschiedlich motivierten Einsatz der SSV in der industriellen Sprachtechnologie und der phonetischen Sprachforschung eingehen. Ich beschränke mich jedoch auf zwei kurze Bemerkungen.

> In einem phonetischen Wahrnehmungsexperiment kann immer nur eine einfache Hypothese geprüft werden, nicht aber eine komplette Zuordnungstheorie. Wie weit wir davon - selbst bei einer Einschränkung auf die systematische Aufarbeitung der mit den BAS-Daten verfügbar gemachten

phonetischen Tatsachen - noch entfernt sind, zeigt sich daran, wie wenig von der in Tillmann/Pompino-Marschall (1993) vorgeschlagenen "complete phonetic theory of spoken German" realisiert werden konnte.

Die Phonetik konnte der Industrie bei der technischen Entwicklung der automatischen Spracherkennung wenig Hilfe bieten. Optimistische Anfangsversuche haben zu der sprichwörtlich gewordenen Verschlechterung von Erkennungsraten geführt. Auf der Syntheseseite aber war die Kooperation hingegen an vielen Stellen fruchtbarer.

Andererseits darf man abschließend sagen, dass erfolgreiche speech-to-text-Systeme eine phonetische Zuordnungstheorie auf eine implizite Weise implementieren, deren verifizierender Anspruch durch ihr Funktionieren befriedigt wird. Leider fällt dabei für eine phonetisch explizit formulierbare Zuordnungstheorie bis heute wenig ab.

2.2 Frühe Erfolge und das spätere Scheitern der klassischen Lauttheorie

Genau 100 Jahre vor der Bonner Wende waren es philologisch ausgebildete Geisteswissenschaftler, die mit der artikulatorischen Phonetik eine neue Art von Sprachforschung begründeten. Sie nannten sich Lautphysiologen; doch in Wahrheit waren sie schon echte Behavioristen, als es den Behaviorismus als Verhaltenspsychologie noch gar nicht gab, und sie hatten völlig andere Interessen. Ihr philologisch motiviertes Thema war nämlich die Verschriftung der gesprochenen Sprache.

Ihre optimistisch vereinfachende Annahme bestand darin, man könne den Buchstaben bei der alphabetischen Verschriftung von lautsprachlichen Äußerungen einen artikulatorischen Inhalt zusprechen. Dieser auf der kategorialen Seite zu beschreibende artikulatorische Inhalt wurde, und das ist die behavioristische Annahme, auf der nicht-kategorialen Seite empirisch identifiziert mit den artikulatorisch ablaufenden Vorgängen, wie sie nur einem externen Beobachter zugänglich sind und dabei durch Zeitfunktionen beschreibbar werden. Wissenschaftstheoretisch bezeichnet man heute ein derart behavioristisches Programm als „naiven Realismus". Doch hat auch das spätere Scheitern der lautphysiologischen „Theorie der Sprachlaute" nicht nur diesen methodischen, sondern auch einen sachlichen Grund.

Das Ziel der Lautphysiologen bestand in einer Verbesserung der Orthographie, um, wie Brücke 1856 auf der ersten Seite seiner „Grundzüge der Physiologie und Systematik der Sprachlaute für Linguisten und Taubstummenlehrer" schreibt, diese „mehr als bisher mit der Aussprache in Übereinstimmung zu bringen, anstatt uns von diesem Ziele alles Schreibens noch weiter zu entfernen".

Für A. M. Bell, dem Begründer der angelsächsischen Tradition, dessen Buch „Visible Speech" 1865 erscheint, steht, wie der Untertitel zeigt, die Verschriftung von Kolonialsprachen im Vordergrund, denn dieser lautet: „Universal alphabetics or self-interpreting physiological letters for the writing of all languages in one alphabet".

Zunächst ist es natürlich auch heute noch eine bestechende Idee, den Buchstaben des Alphabets eine artikulatorische Semantik zu geben. Vereinfacht lässt sich das Vorgehen der Lautphysiologen etwa so beschreiben, dass sie die einzelnen Buchstaben wie wörtliche Rede in einer Art autonym produzierter Zitierform aussprechen. Bei der Selbstbeobachtung kann der am Schreibtisch sitzende Sprachforscher dank der kategorialen Reproduzierbarkeit die artikulatorisch auffälligen Unterschiede systematisch festhalten und protokollieren.

Dass die dabei entdeckten rein artikulatorischen Lautmerkmale auch tatsächlich kommunikativ relevant sind, zeigt sich beispielsweise, wenn ein Sprecher in einer bestimmten Situation die Absicht hat, seinen Hörer über die Possessivverhältnisse von „Mein" und „Dein" aufzuklären. Nicht nur aus Sicht der Lautphysiologen muss er bei der Äußerung von „Dies hier ist MEIN Geld" an der betreffenden Stelle im artikulatorischen Ablauf der Sprechbewegung dafür sorgen, dass sich bei gleichzeitig gesenktem Gaumensegel die Lippen schließen, während hingegen bei der Äußerung „Dies hier ist DEIN Geld" die Zungenspitze für eine Verschlussbildung in Aktion treten muss und das Velum zugleich den Nasenraum abschließt. Sonst ergäbe sich auf der alphabetischen Seite kein „d", sondern ein „n". (In den

Anführungszeichen stehen hier, logisch gesehen, Namen von Buchstaben, die in wörtlicher Rede wiedergegeben werde können.)

Wie naheliegend es tatsächlich ist, in einer Theorie der Sprachlaute statt der abstrakten phonologischen Merkmale, wie sie z.B. in der Generativen Phonologie postuliert werden, die von den Lautphysiologen entdeckten rein artikulatorischen Beschreibungskategorien zu verwenden, lässt sich am besten mit den bekannten McGurk-Effekten demonstrieren, beispielsweise indem man einen Sprecher filmt, der „ba" sagt, und zu dieser Lippenbewegung dann auf der Tonspur ein gesprochenes „la" hinzu synchronisiert. Schaut man bei der Wiedergabe auf das Videobild des Sprechers, hört man „bla", schließt man die Augen, hört man das tatsächlich gesprochene „la". Dieser Effekt ist robust, er verschwindet nicht, wenn der gesehene Sprecher männlich ist und die gehörte Stimme weiblich oder kindlich. (Auf diese visuell induzierten auditiven Täuschungen bei Sprachwahrnehmungsexperimenten komme ich vielleicht beim Vortrag in Eichstätt noch einmal zurück.)

Der große Erfolg der Lautphysiologie liegt in der Entwicklung der elementaren phonetischen Beschreibungssprache, wie sie bis heute im Transkriptionsunterricht vermittelt wird. Sie funktioniert nur auf der kategorial semantischen Seite, denn sie wird den Studenten in speziellen Sprechakten gezielt vermittelt. Die Semantik dieser rein phonetischen Sprechakte hat zwei Seiten. Zum einen wird die Bedeutung, (Brücke sagt: der Lautwert) eines Transkriptionssymbols durch autonymes Reproduzieren demonstriert, was in der Terminologie der Logiker auf eine ostensive Definition hinausläuft. Nachdem aber das so definierte Lautzeichen kategorial als identisch reproduzierbar gesichert, kann dessen Kategorie auch durch die elementaren artikulatorischen Merkmale gekennzeichnet werden. Wie ich in TmM gezeigt habe, tritt logisch gesehen an die Stelle der ostensiven Definition die durch eine definite Kennzeichnung. Diese hat die Form eines analysierten Ausdrucks. Diese zweifache Semantik entspricht aber genau dem, was Phonologen in ihrem naiven Realismus „Segmentation" und „Dekomposition" nennen.

Das Scheitern der lautphysiologischen Theorie der Sprachlaute ist auf wissenschaftlichen Fortschritt zurückzuführen. Denn auf der methodischen Seite wurde mit der Instrumentalphonetik das Artikulationsverhalten extern beobachtbar. Bei den graphisch aufgezeichneten sogenannten „Sprachkurven" handelte es sich ja tatsächlich um Zeitfunktionen. Wie später beim Sonagramm versuchte man beim Studium der Sprachkurven es zu lernen, die durch die Graphik repräsentierten Zeitfunktionen im Hinblick auf die damit sichtbar gemachten artikulatorischen Abläufe als Sprache in ihrer gesprochenen Form lesen zu können. Man glaubte sich in der Lage, die phonetischen Tatsachen durch eine Zuordnungstheorie auf eine empirisch verifizierbare Weise dingfest zu machen. Dass und warum das scheitern musste, kann man im Heike-Beitrag nachlesen.

Trotzdem war die Forschungsarbeit der frühen Instrumentalphonetiker im Aufdecken von vielen bis heute wichtigen phonetischen Tatsachen höchst erfolgreich. Von der Vielzahl und Vielfalt der Einzelergebnisse vermittelt Scripture 1903 in „The Elements of Experimental Phonetics" ein beeindruckendes Bild. Auf dem Höhepunkt der instrumentalphonetischen Sprachforschung präsentiert Scripture einen umfassenden Bericht über den *State of the Art* der Artikulatorischen Phonetik Da findet man schon die VOT, die unterschiedliche Dauer von Vokalen in betonten und unbetonten Silben, das prepausal/prefinal lengthening und sogar eine von Scripture selbst erfundene FFT-Methode zur Spektralanalyse von graphisch vorliegenden Vokalperioden, die er bei der Vergrößerung der Rillenoszillation von Schallplatten gewinnt. Über die Dauer dieser Perioden lässt sich der F0-Verlauf bestimmen und die abfallende Intonation am Äußerungsende.

Das für mich wichtigste Ergebnis der frühen Instrumentalphonetik ist eine ganz grundlegende Entdeckung von Rousselot (1881a,b), die er „Les modifications phonétiques du langage" benennt. Statt der erwarteten kategorialen Konstanz von Sprachlauten wird beim Lesen von Sprachkurven deren fehlende Invarianz als systematische Veränderung erkennbar, und genau diese systematische Variation (etwa die Verkürzung der Vokaldauern als Funktion der Anzahl der Silben pro Wort) wird von nun an ins Zentrum der instrumentalphonetischen Forschung gerückt.

Die Fragestellung nach den systematischen und somit theoretisch beherrschbaren Modifikationen in der phonetisch beschreibbaren Form von lautsprachlichen Äußerungen, wie sie durch einzelsprachlich geregelte Sprechbewegungen bei jedem Sprechakt beobachtet werden, hat durch das instrumentalphonetische Scheitern der klassischen Lauttheorie nichts von ihrer Aktualität verloren, ganz im Gegenteil. Erst mit den Methoden der digitalen SSV und nicht zuletzt dank der (durch die DFG ermöglichten) Entwicklung der dreidimensionalen elektromagnetischen Artikulographie sind wir erst heute in der Lage, diese Fragestellung hypothesengeleitet in Angriff zu nehmen. Um dies mit einfachen Worten auch umgangssprachlich erläutern: Wie ändert sich ein isoliert gesprochener Sprachlaut, wenn er in einem bestimmten Wort gesprochen wird? Wie ändert er sich, wenn nicht in einer alphabetisch expliziten Zitierform realisiert wird, sondern in fließender Rede? Wie ändert sich die kanonische Form einer lexikalischen Einheit in unterschiedlichen Sprechstilen? Wie wird die phonetische Form einer lexikalischen Einheit modifiziert, wenn sich die Sprechgeschwindigkeit in einer bestimmten Richtung ändert? Die phonetische Sachlage hier wird beliebig kompliziert.

> Im mündlichen Vortrag möchte ich wenigstens kurz auf das zusammen mit Hartmut Pfitzinger entwickelte Forschungsparadigma einer „Synthese durch Analyse“ eingehen, das wir zuerst bei der ICSLP2000 in Peking auf einem Poster zum PHD-Projekt (*Parametric High-Definition*-Sprachsynthese) vorgestellt haben.
>
> Zweitens bietet es sich in diesem Zusammenhang an, die Reduktion der kanonischen Formen, wie sie in den BAS-Daten ja direkt als phonetisch gegebene Tatsache dokumentiert sind, auch theoretisch in ihrer Systematik explizit zu machen. Diese Idee wurde in der Startphase des PHONDAT-Verbundvorhaben in Tillmann & Pompino-Marschall (1993) auf der EUROSPEECH in Berlin mit der allzu optimistischen Idee verbunden, anhand der neu aufzubauenden sehr großen Sprachsignaldatenbanken so etwas wie eine CPT (complete phonetic theory) des heute gesprochenen Deutsch zu entwickeln.

Die ausschlaggebenden Tatsache, die allein erklärt, warum das ABC in der Menschheitsgeschichte nur ein einziges Mal erfunden wurde, habe ich in TmM ausführlicher beschrieben: Die alphabetische Form von Äußerungen, die phonetisch durch eine wohlartikulierte B- und C-Prosodie realisiert wird, läuft für die menschliche Wahrnehmung viel zu rasch ab, als dass sie in all ihren artikulatorischen Einzelheiten direkt beobachtet werden könnte. Wäre dies nicht der Fall und hätte Brücke mit seiner Methode der „directen Beobachtung der Sprachlaute“ die B- und C-prosodischen Abläufe wie unter einer Zeitlupe beobachten können, so wäre er nie auf die Idee gekommen, die Schrift besser als bisher mit der Aussprache in Übereinstimmung zu bringen. Ja sogar auch das Alphabet selbst wäre vor 3000 Jahren vermutlich nie von den Phöniziern entdeckt bzw. erfunden worden. Darauf möchte ich in Eichstätt mit einer akustischen Demonstration der A-, B- und C-Prosodie zurückkommen.

Nach den hier beschriebenen Erfolgen sei am Ende dieses Abschnitts das späte Scheitern der klassischen „Phonetik als Lehre von den Sprachlauten“ aus der Sicht eines ihrer führenden Vertreters wiedergegeben. Scripture schreibt, nachdem er 1932 bei Menzerath den ersten sagital aufgenommenen Röntgenfilm der beim Sprechen ablaufenden artikulatorischen Vorgänge gesehen hatte:

> „Die kleinsten Bewegungen der Lippen, der Zunge, des Gaumensegels, des Zungenbeins, der Kehlkopfknorpel usw. spielen sich vor dem Auge ab“. Scripture nennt den Eindruck eines solchen Films „überwältigend.“ Denn da sehe man einen „Schattenmenschen […], wie er spricht, atmet und schluckt. Die Sprechwerkzeuge stehen nicht für einen Augenblick still, jeder Sprechakt ist die Summe der Bewegungen aller Organe des Mundes, des Rachens, des Kehlkopfes usw., und diese Summe spielt sich in der Zeit ab. Lautstellungen gibt es überhaupt nicht: es kommt alles auf Lautbewegungen hinaus. Man begreift sofort, daß die bisherige Lautphysiologie nur eine Irrlehre sein kann, und wartet gespannt auf neues.“(S.. 173)

Als die Instrumental- und Experimentalphonetiker in ihren Sprachkurven statt der gesuchten Lautsegmente „unendlich viele Lautnuancen“ fanden und nachdem Panconcelli-Calzia

die phonetische Existenz von Sprachlauten und Silben strikt verneinte und sie für eine Erfindung, ja für eine „Fiktion der Linguisten“ erklärte, übernahmen diese ihrerseits das Kommando und präsentierten ihre eigene „rein geisteswissenschaftliche“ Theorie der Sprachlaute. Die Einzelheiten darf ich hier als bekannt voraussetzen. Am Ende dieses langen Abschnitts sind zum Abschluss des ganzen nur noch zwei Bemerkungen erforderlich. Als phonematische Einheiten sind Sprachlaute immer nur als abstrakte einzelsprachliche zu definierende Größen gegeben. Die Wörter einer Sprache unterscheiden sich allein durch ihre phonematischen Einheiten, wobei sich beim Vergleich von Minimalpaaren (wie Fisch und Tisch, Muße und Muse usw.) das Phonem als *kleinste* phonematische Einheit nachweisen lässt. Diese Form der linguistischen Sprachlautanalyse konvergiert für jede bisher beschriebene Sprache sehr rasch auf ein sehr kleines Repertoire. Damit lässt sich jedes Wort dieser Sprache alphabetisch repräsentieren und jede in wörtlicher Rede produzierte Äußerung dieser Sprache auf eindeutige Weise verschriften.

Das heißt im Klartext: Die klassische Phonologie ist eine reine Wortphonologie. Und so wird auch für die geisteswissenschaftliche Sprachlauttheorie das Wort zur maßgebenden Größe. Beim der Vergleichen von artikulatorisch reproduzierter Minimalpaare interessiert sich der Phonologe nur für die lautliche Seite der lexikalischen Einheiten in ihrer alphabetisch explizierten Form, so wie sie nur bei isolierter kategorialer Reproduktion zu beobachten ist - sei es zur Demonstration einer Zitierform, sei es zum Zwecke ihrer alphabetischen Analyse. Von einer Realisationsphonologie im Sinne Theo Vennemanns kann hier noch keine Rede sein.

2.3 Die Übertragung der lautsprachlichen Information vom Sprecher zum Hörer durch das akustische Sprachsignal: die *phonetische* Kanaltheorie

Anhänger der ***Motor Theory of Speech Perception***, zu denen ich nicht zähle, weisen gerne auf die sehr beeindruckenden Experimente zum Phänomen „*duplex perception“* hin, in denen überraschenderweise gezeigt werden kann, dass unser auditives NS über einen höchst speziellen eigenen **speech mode** verfügt**,** in den es dann umschaltet, sobald ein hörbares Ereignis als gesprochene Sprache gehört wird. Ein und dasselbe Schallsignal kann also auf zweifache Weise gehört werde, einmal als gesprochene Sprache und einmal als nichtsprachliches Schallereignis. Diesen Effekt werde ich im Vortrag an einem eigenen Beispiel vorführen. Dabei habe ich in einem kurzen Satz (der übrigens, um den Effekt nicht zu gefährden, hier nicht zitiert werden darf) die stimmhaften Passagen des Sprachsignals durch die Repetition der einzelnen Stimmtonperioden etwa zwanzig- bis vierzigfach gedehnt. Bei der Wiedergabe des so erzeugten Signals hört man nur die periodischen Teile als eine Folge von Klängen und die nicht gedehnten konsonantischen Signalstücke bleiben dabei unauffällig, sie werden gewissermaßen ausgeblendet. In meiner Terminologie fallen die gedehnten Teilstücke in den Bereich der A-Prosodie, die man in ihrem Verlauf auditiv sehr gut verfolgen kann. Will man das auditive System des NS in den Modus der Verarbeitung von gesprochener Sprache bringen, muss man dem Hörer zuerst die Aufnahme im Originalton vorspielen und sodann schrittweise in einer zwei-, vier-, acht- (usw.) bis zwanzig- oder vierzigfachen Dehnung präsentieren. Hier bleibt das auditive Nervensystem im sog. speech mode. Der Hörer nimmt den artikulatorischen Ablauf der durch das akustischen Signal abgebildeten Sprechbewegung auch in der schrittweise verzögerten Form wahr. Die ungedehnten konsonantischen Anteile bleiben gewissermaßen in einer Art von Topdown-Prozess in der Wahrnehmung präsent, werden also nicht wie bei der ersten Präsentation maskiert.

Mit diesem Beispiel sollen zwei phonetische Sachverhalte verdeutlicht werden. Erstens kann hier gezeigt werden, dass der artikulatorische Inhalt, der durch das akustische Signal zum hörenden Nervensystem übertragen wird, bei der schrittweisen Dehnung der

periodischen Anteile des Signals nicht zum Verschwinden gebracht werden kann, wenn nämlich das kognitive Beobachtungssystem im hörenden Nervensystem schon „weiß“, dass es sich beim Gehörten um gesprochene Sprache handelt.

Zweitens zeigt sich, dass die für das menschliche Ohr viel zu rasch ablaufende B- und C-Prosodie unter die Zeitlupe der A-Prosodie geholt und damit in ihrem artikulatorischen Ablauf beobachtbar gemacht werden kann. Das kognitive Beobachtungssystem „weiß“ also nicht nur, dass es sich um gesprochene Sprache handelt, sondern auch, dass es sich um dieselbe Äußerung handelt. Das erklärt, warum dann selbst auf der **kategorialen** Seite auch in der gedehnten Form immer dieselbe Äußerung wahrgenommen wird.

> Wir wissen heute, wie das am Munde ins akustische Feld abgestrahlte Sprachsignal durch die beim Sprechakt ablaufenden artikulatorischen Vorgänge auf berechenbare Weise erzeugt wird. Das mathematisch einfachste Modell liefert die Z-Transformation. Liegen die glottale Anregungsfunktion g(t) und die Impulsantwort h(t) des Ansatzrohres (ohne oder mit dem zugeschalteten Nasenraum) in digitalisierter Form vor, dann erhalten wir mit dem Produkt ihrer beider Z-Transformierten, G(z) und H(z), exakt die Z-Transformierte F(z), aus der sich das am Munde ins akustische Feld abgestrahlten Sprachsignals f(t) in seiner digitalisierten Form leicht exakt berechnen lässt. (unter der vereinfachenden Annahme, dass nichtlineare Effekte zu vernachlässigen sind).
>
> Etwas komplizierter sind die vierpoltheoretischen Modelle der Nachrichtentechnik, wobei hier nur der Name von Gunnar Fant fallen muss.

Was das Sachgebiet Phonetik betrifft, so haben Meyer-Eppler und Ungeheuer die für unseren Zusammenhang grundlegende ***physikalische*** Theorie der akustischen Artikulation entwickelt. Auf die Einzelheiten (wie den Einsatz der Websterschen Horngleichung usw.) näher einzugehen, ist hier kein Raum. Maßgebend für unseren Zusammenhang ist die Sigmafunktion $\sigma(z)$die den Verlauf der Querschnittsfläche der schwingenden Luftsäule im Ansatzrohr bei der Vokalproduktion entlang der z-Achse von den Stimmlippen bis zur Mundöffnung beschreibt (in Abb. **Abb.1** auf der nächsten Seite von links nach rechts). Im Kontext der Digitalen SSV wird $\sigma(z)$ auch *area-function* genannt.

Als Ausgangspunkt wählte Ungeheuer in seiner Dissertation 1957 das kreiszylindrische Rohr, dessen Impulsantwort (bei einem nicht zu großen Durchmesser) die äquidistanten Formantfrequenzwerte liefert, die ohrenphonetisch nicht nur einen Neutralvokal ergeben, sondern diesen Schwa-Laut der Lautphysiologen geradezu in einer akustisch reiner Form.

> Bei gemessenen Formatfrequenzen von zum Beispiel
>
> F1 = 500 Hz
> F2 = 1500 Hz
> F3 = 2500 Hz
> etc.
>
> und einer Schallgeschwindigkeit von 340 m/sec beträgt die Länge des Ansatzrohres genau 17 cm. Die Weite des konstanten Querschnittverlaufs spielt kleine Rolle, solange sie nur bezüglich der Wellenlänge der größten interessierenden Resonanzfrequenz klein bleibt. Deshalb funktioniert die ***quantal theory of speech*** von Ken Stevens. Die findet man jedoch schon bei Gerold Ungeheuer, der darauf hinweist, und das eine solche Änderung von $\sigma(z)$ die Frequenz der stehenden harmonischen Wellen, die sich mathematisch als Lösung der Horngleichung bei den gegeben Randbedingungen im Ansatzrohr ergeben, keinen Einfluss hat.

Der akustisch gegebene ideale Schwa-Laut mit den äquidistanten Resonanzfrequenzen ist auch ein geradezu ideales Kodiersystem zur Übertragung von phonetischer Information. Durch ein gezieltes Abweichen vom konstanten Querschnittsverlauf der Sigmafunktion lassen sich insbesondere die unteren drei Formanten aus dem neutralen Zentrum in Richtung der Eckvokale des artikulatorisch vorgegebenen Vokalvierecks verschieben.

> In TmM wurde diese Tatsache herangezogen, um erstens zu erklären, wie sich die bei der Vokalartikulation ergebenden Abweichungen vom neutralen Ansatzrohr auswirken auf die resultierenden

Formantfrequenzen. Man sieht am „Formantverschiebungsschema“ in sofort, warum z.B. eine Lippenrundung, also die Verengung des Querschnittsverlaufs an dieser Artikulationsstelle, die Formantfrequenzen absenkt.

Zweitens kann man an diesem „Formantverschiebungsschema“ insbesondere aber auch verstehen, welche Auswirkungen die mit den CV- und -VC-Bewegungen der Zunge gegebenen C-prosodischen Abweichungen des Querschnittsverlaufs im Ansatzrohr auf die Formanttransitionen am silbischen Vokalrand auf die Formantfrequenzen eines Vokals zur Folge hat.

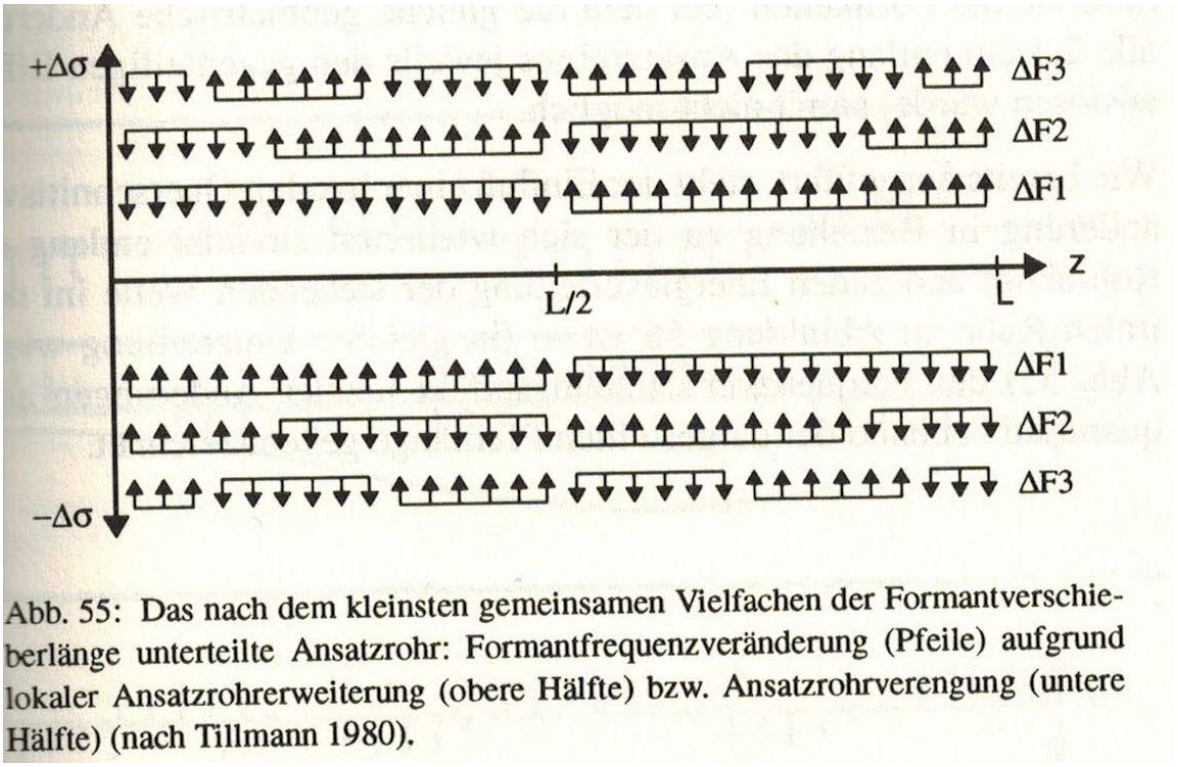

Abb. 55: Das nach dem kleinsten gemeinsamen Vielfachen der Formantverschieberlänge unterteilte Ansatzrohr: Formantfrequenzveränderung (Pfeile) aufgrund lokaler Ansatzrohrerweiterung (obere Hälfte) bzw. Ansatzrohrverengung (untere Hälfte) (nach Tillmann 1980).

Abb. 1 Das Schema der Resonanzverschiebung (aus Pompino-Marschall (2003))

Aus der akustischen Theorie der Vokalartikulation folgt, dass die Abbildung der sich beim B- und C-prosodischen Artikulieren ergebenden Änderungen der Querschnittsgeometrie auf das akustische Signal eindeutig ist. Bei zeitlich unveränderter Geometrie ist die Abbildung nicht eineindeutig umkehrbar. Max Hadersberg hat aber in einer von mir bei der Informatik betreuten Dissertation gezeigt, dass sich dies das grundsätzlich ändert, sobald die Querschnittsgeometrie in Bewegung versetzt wird. Er war zwar weit davon entfernt, das Umkehrproblem der akustischen Artikulationstheorie zu lösen, das darin besteht, aus gemessenen Formantfrequenzverläufen auf die artikulierende Geometrie zurückrechnen zu können. Aber er hat gezeigt, dass unterschiedliche Geometrien, die dasselbe akustische Bild liefern, voneinander zu unterscheiden sind, wenn man ihre sich bei zeitlicher Änderung ergebenden Umgebungen betrachtet.

Wenn man die Beispiele, die auf der kategorialen Seite das Wahrnehmen des artikulatorischen Inhalts von gesprochenen Äußerungen illustrieren, in Verbindung bringt mit den Erkenntnissen über die Abbildung von ablaufenden Sprechbewegungen auf das dabei produzierte akustische Signal, dann kann man die Frage, wann aus Schall Sprachschall wird, so beantworten, dass ein akustisches Signal als gesprochene Sprache gehört wird, sobald diesem die mit den einzelsprachlich geregelten Sprechbewegungen verbundene dreifache Prosodie aufgeprägt wird. Das geschieht z.B. beim Vocoder, wenn eine Orgel zum Sprechen gebracht wird. Ein weiteres Beispiel ist die Tatsache, das selbst so extrem verzerrte Sprache wie clipped speech, bei der nur die Nulldurchgänge erhalten bleiben,

Sprache bleibt und sogar verstanden werden kann, wenn nur noch Reste der A-, B- und C-Prosodie für das hörende Nervensystem erkennbar bleiben.

Die phonetische Kanaltheorie besagt, das in der Kommunikationskette keine Symbole vom Sender zum Empfänger geschickt werden.

2.4 Meyer-Epplersche Beobachtungsketten beim kognitiven Ablauf von Sprechakten

Wenn das Kanalmodell die akustische Übertragung der phonetischen Information im lautsprachlichen Kommunikationsprozess erklärt, dann erklärt das Modell der kognitiven Beobachtungsketten den Rest. Im Rahmen von empirischen Zuordnungstheorien dürfen wir nämlich davon ausgehen, dass sich im individuellen Bewusstsein von Sprechern und Hörern immer genau dann und nur dann ein einzelsprachlich geregelter Sprechakt ergibt, wenn auf der kognitionstheoretischen Seite durch die neuronalen Selbst- und Fremdbeobachtungssysteme ein prosodisch nachvollziehbares kortikales Bild der ablaufenden Sprechbewegung erzeugt wird.

Die psychologische Modellvorstellung, die der Beobachtungskette im Sinne Meyer-Epplers entspricht, ist nicht an der auditiven, sondern an der *visuellen* Wahrnehmung orientiert. Es gibt drei Hauptkomponenten. Auf der Hirnrinde einsteht nur dann ein ***kortikales Bild***, wenn es in der physikalischen Wirklichkeit erstens ein so genanntes ***kognitives Objekt*** gibt (z.B. eine Tasse) und davon zweitens an der sensorischen Peripherie des Nervensystems, also auf der Netzhaut des Auges, ein ***retinales Bild*** erzeugt wird. Dieses optische Bild wird reiztransformiert und durch die Signalverarbeitungsprozesse im Gehirn auf dem Weg zur Hirnrinde so transformiert, dass dort das kortikales Bild des kognitiven Objekt resultiert, das mit der gesehenen Tasse empirisch identifiziert werden kann. Bei der visuellen Sprachwahrnehmung ist das kognitive Objekt nichts anderes als die gesehene Sprechbewegung, und dass das kortikale Bild derselben sich auch auf die neurologische Verarbeitung der auditiven Seite auswirkt, zeigen die McGurk-Effekte.

Wenn wir den Begriff der **retinalen Repräsentation des kognitiven Objekts,** das mit den **beim Sprechakt ablaufenden Sprechbewegungen** gegeben ist**,** auch auf die phonetische Wahrnehmung von lautsprachlichen Äußerungen anwenden wollen, müssen wir von einer ***dreifach gegebenen retinalen Repräsentation*** ausgehen. Das kortikale Bild der *gehörten und gefühlten* Sprechbewegung beim motorisch aktiven NS des Senders unterscheidet sich von dem der *nur gehörten* beim Empfänger.

Dank der beim Sprechakt über den akustischen Kanal übertragenen phonetischen Information entsteht auf der Basilarmembran des Gehörsinnes das retinale Bild einer gerade ablaufenden Sprechbewegung. Da das Ohr des Sprechers nicht nur durch den Luftschall stimuliert wird, unterscheiden sich die retinalen Bilder schon an der akustischen Peripherie, auch wenn sie in der prosodischen Form so gut wie identisch ablaufen. Die kortikalen Bilder jedoch, die beim Sender und Empfänger entstehen, sind sehr verschieden. Insbesondere spielen beim Sprechen auf der sensorischen Seite motorisch erzeugte Reafferenzen eine entscheidende Rolle. Auch wird beim sprechenden NS der motorisch geplante Ablauf der Sprechbewegung mit dem auditiv resultierenden kortikalen Bild verglichen.

> Doch das ist erst die halbe Wahrheit. Denn an dieser Stelle muss die Frage gestellt werden, in welcher Form der artikulatorische Ablauf auf der ***artikulatorische Retina*** zur Darstellung kommt. Wie „sieht" das sprechende Nervensystem sein eigenes artikulierendes Oberflächenverhalten (OV), das ja auf seiner sensorischen Oberfläche eine taktile und propriozeptive Reiztransformation (RT) auslöst? Bei B- und C-prosodisch wohlartikulierten Lautsequenzen ist die Abbildung von OV auf RT extrem nichtlinear.

2.4.1 Kortikale Wohlartikuliertheit dank der B- und C-Prosodie

Phonetisch wohlartikulierte Sprache lässt sich durch CVCVCV-Sequenzen beschreiben, wobei V für einen Vokal, Diphthong, Triphthong oder eine Abfolge derselben steht, während C durch einen oder mehrere Konsonanten realisiert wird. Fehlt zwischen zwei vokalischen Silbenkernen ein konsonantischer Übergang, so werden sie durch nur ein V repräsentiert. Andererseits werden Konsonantenfolgen zwischen zwei Silbenkernen zu einem C zusammengefasst.

> Der Begriff der ***phonetischen Wohlartikuliertheit*** von gesprochener Sprache wurde 1871 vom Lautphysiologen Friedrich Techmer in seiner Leipziger Antrittsvorlesung über „Die naturwissenschaftliche Analyse und Synthese der hörbaren Sprache“ im Zusammenhang mit seiner Diskussion von Selbst- und Mitlauten eingeführt, um die „innige Verbindung“ bei der CV- und VC-Artikulation zu erklären.

Dieser alte Begriff erhält bei der artikulatorischen Retina eine ganz neue Bedeutung. Dank der B- und C-Prosodie kommt es auf der artikulatorischen Retina zu einem abrupten Wechsel der konsonantischen Artikulationsstellen, an denen jeweils immer nur ein artikulierendes Organ genau einen Artikulationsmodus realisiert. Ein solcher Wechsel käme nicht zustande ohne die mit offenem Vokaltrakt produzierten vokalischen Elemente in CV{…CV…}CV-Sequenzen.

> Zwar lässt sich jede beliebig lange konsonantische CCCCC…CCC-Sequenz von jedem geschulten Ohrenphonetiker auch ohne eingeschobene V-Elemente aussprechen. Sie wäre aber im Sinne von Techmer nicht wohl artikuliert. Bei wohlartikulierten CVCVC-Sequenzen jedoch „springt“ das den Konsonanten produzierende Artikulationsorgan zwischen den vokalischen Silbenkernen an eine bestimmte Stelle der artikulatorischen Retina (die ja beim V-Element bis auf den vokalischen Rand bei den mit hoher Zungenlage artikulierten Vokoiden blank bleibt), um dort einen der Modi aus der Menge der verfügbaren Artikulationsarten (z.B. plosiv, frikativ, lateral, gerollt etc.) zu demonstrieren.
>
> In den Tabellen der IPA ist jeder Konsonant durch genau einen Artikulationsort, ein artikulierendes Organ, einen Artikulationsmodus (etc.) definiert. Damit werden Sprachlaute mit Bezug auf die artikulatorische Retina zu regulären Vektoren, die in jedem Beschreibungsparameter immer nur einen der möglichen Werte realisieren. So erklärt sich auch die Tatsache, dass die Kennzeichnung der alphabetischen Lautzeichen durch die IPA-Tabellen als *reguläre Vektoren* bezeichnet werden können, die in jedem Beschreibungsparameter immer nur einen Wert annehmen. (Die Klick-Laute afrikanischer Sprachen, die mit Ihre zwei Artikulationsstellen keine regulären Vektoren sind, können als eine die Regel bestätigende Ausnahme akzeptiert (oder zu den unklaren Fällen gezählt) werden.)
>
> Informationstheoretisch gesehen ist dies eine sehr primitive Art der Kodierung mit maximalen Hammingdistanzen..

Was die proximale Wahrnehmung von wohlartikulierten Sprechbewegungen betrifft, darf man also sagen, dass die phonetisch relevante Information, die das Nervensystem an seiner motorischen Peripherie als “overt behavior“ in der Form des artikulatorischen Oberflächenverhalten der Sprechwerkzeuge realisiert, über die artikulatorische Retina in einer wohl artikulierten diskreten Form ins sensorische Nervensystem zurückkehrt. Bei im Sinne von Techmer nicht wohlartikulierten CCCC-Sequenzen wäre dies nicht der Fall.

> Ein an drei Artikulationsstellen gleichzeitig gerolltes R ist zwar ohne weiteres demonstrierbar, ist aber im IPA-System ist nicht vorgesehen. Das wäre nämlich kein regulärer Vektor und könnte auch nicht in einer wohlartikulierten Form als eine intervokalische VCV-Sequenz gesprochen werden.

Dank der nichtlinearen Abbildung von kontinuierlich ablaufendem OV und der damit verbundenen diskretisierenden RT auf der artikulatorischen Retina kommt es zu dem kortikalen Bild, das der im Sprachakt wahrgenommen Äußerung im Sinne der der empirischen Zuordnungstheorie entspricht

2.4.2 Das Zusammenwirken der Nah- und Fernsinne beim phonetischen Spracherwerb

Die aktuelle Forschung (bei Google findet man zu "oral perception" fast 8000 Einträge) wird wohl bald zeigen, dass es sich beim Konzept der artikulatorischen Retina nur um eine viel zu optimistische extreme Vereinfachung handeln kann. Trotzdem ist sie nützlich, um den phonetischen Erwerb einer Muttersprache zu kommentieren und damit den Erstsprachenerwerb in den Monaten nach der Geburt kommunikationstheoretisch zu erklären.

> Die sogenannte Lall- bzw. Babbelphase, die auch bei gehörlos geborenen Kindern stattfindet, dient offenbar zunächst (zusammen mit den anderen sog. 'mass movements') dem planlosen Spielverhalten der späteren Sprechwerkzeuge, das über die taktilen und propriozeptiven Sinne auf den sensorischen Kortex kognitiv zurückgemeldet wird. Beim Babbeln werden sowohl wohl- als auch nicht-wohlartikulierte lautliche Äußerungen produziert. Während das gehörlose Nervensystem am Ende dieser Entwicklungsphase verstummt, hat das hörende es in dieser Zeit gelernt, die mit dem akustisch produzierten Signal über auditive Retina des kognitiven Selbstbeobachtungssystems erzeugten Hörereignisse mit den proximal über die artikulatorische Retina zurückgemeldeten Mundbewegungen (nach dem lerntheoretischen Prinzip der Reizgeneralisation) miteinander zu verknüpfen, man kann sogar geradezu sagen, eine empirische Zuordungstheorie zu implementieren.
>
> Die Konsequenzen liegen auf der Hand. Sobald das lautproduzierende Oberflächenverhalten der motorischen Seite auf der sensorischen antizipierbar wird, wird es auch faktisch antizipiert, und damit zugleich planbar. So wird aus planlosem geplantes Verhalten. (Auf diesen Mechanismus hat u. a. Arnold Gehlen in seiner biologisch orientierten Anthropologie hingewiesen.)

Die auditive Retina ist die Retina eines Fernsinnes. Nachdem in der Lallphase das proximale Babbeln auch auditiv als proximales, also selbsterzeugtes Verhalten wahrgenommen wird, ändert sich die Situation, sobald für das kindliche Nervensystem über die auditive Beobachtungskette auch das distale lautproduzierende Artikulationsverhalten seiner sprechenden Umgebung mit dem proximalen eigenen verglichen und somit imitierbar wird. Wird das Imitieren gelernt, so wird mit dem Imitieren auch das Imitierte gelernt, als plan- und damit reproduzierbares Artikulationsverhalten. Wenn das phonetisch Imitierte dann auch auf der semantischen Seite als sprachlich einsetzbares Äußerungsverhalten verfügbar geworden ist, wird aus dem Imitierten Muttersprache in einer ersten kleinkindlichen Form.

Chomsky hat geradezu sprachpäpstlich verkündigt, eine Sprache könne nicht durch Imitation gelernt wird. Darauf muss man antworten, dass seine Sprachtheorie (und die Unterscheidung von Kompetenz und Performanz) nicht einmal erklären kann, wie das Imitieren gelernt wird.

3 Abschließende Bemerkungen zur Hirnforschung

Hirnforscher, die als externe Beobachter durch bildgebende Verfahren die Vorgänge sichtbar machen, die auf der kortikalen Seite während eines Sprechaktes in der Form von sehr komplexen Zeitfunktionen ablaufen, können nicht das sichtbar machen, was Sprecher und Hörer als gesprochene Sprache erleben und verstehen. Sie können aber z.B. zeigen, dass und warum sich die kortikalen Bilder der Beobachtungsketten beim Sender und Empfänger der Kommunikationskette so drastisch unterscheiden. Und was den Zusammenhang von Äußerung und Bedeutung betrifft, so wird beim realen Sprechakt die Äußerung nie kortikal isoliert vom individuellen Sprecher und der gegebenen Situation betrachtet werden dürfen und auch nicht unabhängig von dem, was man Gedächtnis nennt.

Im Gehirn hängt alles zusammen, was auch faktisch auf der kategorialen Seite zusammengehört. Wie Jeff Hawkins in seinem wunderbaren Buch über das – wie ich einschränkend sagen möchte: *vermutliche* – Funktionieren des Gehirns und insbesondere über das Geschehen in den einzelnen Schichten von Nervenzellen in der Hirnrinde zeigt, ist die Hirnforschung von ihren ehrgeizigen Zielsetzungen noch meilenweit entfernt. (Auf

die mehr als dreihundertachtzigtausend. Einträge zu diesem Buch bei Google sei hier nur kurz hingewiesen). Die Hirnforschung kann sicher eines Tages verständlich machen, wie die Phonetik und Semantik von Sprechakten hirnphysiologisch miteinander verbunden sind. Doch Vorgänge bleiben Vorgänge, wie sie durch Zeitfunktionen zu beschreiben sind. Ihre kategoriale Identität lässt sich daraus nicht ableiten.

Der Zusammenhang zwischen den extrem komplexen Vorgängen im Gehirn und der auf der kategorialen Seite gegebenen Welt ist nur durch empirische Identifikation gegeben, aber nicht selbst empirisch direkt zu beobachten. Und so ist auch die vereinfachende Annahme der Vertreter einer „motorischen" Sprachwahrnehmungstheorie, jede geplante lautsprachliche Äußerung läge im Sprachzentrum der Hirnrinde in einer phonologisch kontextfreien diskreten Form vor und würde erst durch die zu langsame periphere Maschinerie in eine kontinuierlich ablaufende kontextsensitive Sprechbewegung transformiert, nicht nur naiv, sondern auch falsch. Eine linguistische Beschreibung besitzt keine Kausalität, sie kann nicht durch sich selbst erklärt werden. Sprachlaute sind buchstäblich bezeichenbare Einheiten, aber das erklärt nicht ihre Existenz. Wie ein Gehirn „lernen" kann, mit Hilfe dieser Einheiten den Text von lautsprachlichen Äußerungen hinzuschreiben, ist eine ganz andere Frage. Die „Vereinfachung" eines Wortes auf seine buchstäbliche Form ist nicht seine Erklärung, sondern für den externen Beobachter immer mit zusätzlichen Vorgängen im Sehzentrum des Gehirns verbunden. Aber die Weiterentwicklung der lautsprachlichen zur schriftsprachlichen Kommunikation ist hier kein Thema.

Zum Abschuß ein dreifaches Danksagen:

Danken möchte ich an erster Stelle den Veranstaltern dieser 26. ESSV-Konferenz, und zwar ausdrücklich namentlich Prof. Wirsching ! Ohne ihre Einladung hätte ich diesen Text nicht als Vortrag geschrieben (den man deshalb auch wie einen Vortrag lesen sollte (mit entsprechenden Betonungen und Pausen)).

Zweitens danke ich Rüdiger Hoffmann für die Anregung, die im Heike-Beitrag behandelte Thematik im größeren Zusammenhang der Verbindung von geistes- und naturwissenschaftlicher Sprachforschung zu diskutieren. Wenn Phonetik und Semantik Sachgebiete sind, die aus der Sicht ganz verschiedener Fachgebiete betrachtet werden müssen, dann werden die Fachgrenzen zum Verschwinden gebracht. Sie stehen gewissermaßen senkrecht auf den Sachen, auf die der fachliche Blick bei der interdisziplinären Kooperation gerichtet wird. Dabei spielen die Methoden der Sprachsignalverarbeitung eine zentrale Rolle.

Drittens danke ich meinem jungen Freund Tim Anthony Alexander für seine technische Hilfe.

Literatur

Bell, A. M.: Visible Speech. Universal alphabetics or self-interpreting physiological letters for the writing of all languages in one alphabet. London/New York 1867

Bissiri, M.P. and H. R. Pfitzinger: Italian speakers learn lexical stress of German morphologically complex words. Speech Communication 51, 933–947, 2009.

Brücke, E. W. v.: Untersuchungen über die Lautbildung und das natürliche System der Sprache. Sitzungsber. der königl. Akad. der Wissenschaften. Mathem.-Naturwiss. Classe II, 182-208, Wien 1849

Brücke, E. W. v.: Grundzüge der Physiologie und Systematik der Sprachlaute für Linguisten und Taubstummenlehrer. Wien !956

Hammarström, G.: Linguistische Einheiten im Rahmen der modernen Sprachwissenschaft. Berlin-Göttingen-Heidelberg 1966

Heike, G.: Über das phonologische System der Stadtkölner Mundart. Zeitschr. Für Phonetik, Sprachwissenschaft und Phonetik 14, 1-20, 1961

Heike, G.: Suprasegmentale Merkmale der Stadtkölner Mundart. Phonetica 8, 147-165, 1962

Heike, G.: Sprachliche Kommunikation und linguistische Analyse. Heidelberg 1969

Hess, W., Kohler, K., Tillmann, H. G. ``The PhonDat-Verbmobil Speech Corpus", Proceedings of EUROSPEECH, pp. 863-866, Madrid 1995

Menzerath, Paul und A. de Lacerda: Koartikulation, Steuerung und Lautabgrenzung. Berlin-Bonn 1933

Menzerath, Paul: Gedanken über Kern- und Wendepunkte der Phonetik. Arch. f. vergl. Phonetik 6, 89-102, 1942

Meyer, E. A.: Beiträge zur deutschen Metrik. Die neueren Sprachen 6, 1-37, 122-40, 1897

Meyer-Eppler, W.: Grundlagen und Anwendungen der Informationstheorie. Berlin - Göttin gen - Heidelberg 1961

Meyer-Eppler, W. und G. Ungeheuer: Die Vokalartikulation als Eigenwertproblem. Zeit schrift für Phonetik 10, 245-257, 1957

Pompino-Marschall, B.: Einführung in die Phonetik. Berlin 2003

Rousselot, P.J.: Les modifications phonétiques du langage. Revue des patois gallo-romans 4, 65-208, 1981a

Rousselot, P.J.: La méthode graphique appliqué à la recherche des transformations inconscientes du langage. Revue des patois gallo-romans 4, 209-13,1981b

Scripture, E. W.: The Elements of Experimental Phonetics, New York/ London 1902

Scripture, E. W.: Referate. Zeitschrift für Experimental-Phonetik I (3/4), 171-88, 1932

Techmer, F.: Naturwissenschaftliche Analyse und Synthese der hörbaren Sprache. Intern. Zeitschr. f. allgemeine Sprachwissenschaft I, 69-170, 1884

Tillmann, H. G., G. Heike, H. Schnelle und G. Ungeheuer: DAWID I – Ein Beitrag zur automatischen Spracherkennung, Beitrag A12, 5. Intern. Akustikkongress, Lüttich 1965

H.G. Tillmann, B. Pompino-Marschall: Theoretical Principles Concerning Segmentation, Labelling and Levels of Categorical Annotation for Spoken Language Database Systems, Proceedings of EUROSPEECH 1993, pp. 1691 - 1694, Berlin (1993).

Ungeheuer, G. : Elemente einer akustischen Theorie der Vokalartikulation. Berlin 1962

Vennemann, Th. und Joachim Jacobs: Sprache und Grammatik. Darmstadt 1982