



**University of
Zurich** ^{UZH}

**Zurich Open Repository and
Archive**

University of Zurich
University Library
Strickhofstrasse 39
CH-8057 Zurich
www.zora.uzh.ch

Year: 2010

Die Wahrnehmung der Sprachlaute

Schmid, Stephan

Other titles: Wissenschaftsgeschichtliche Streiflichter auf Phonetik und Phonologie

Posted at the Zurich Open Repository and Archive, University of Zurich
ZORA URL: <https://doi.org/10.5167/uzh-33283>
Book Section

Originally published at:
Schmid, Stephan (2010). Die Wahrnehmung der Sprachlaute. In: Egli, W M; Tomkowiak, I. Sinne.
Zürich: Chronos, 123-145.

Werner M. Egli und Ingrid Tomkowiak (Hg.)

Sinne

Redaktionelle Mitarbeit: Helene Mühlestein

Sonderdruck

CHRONOS

Der vorliegende Sammelband enthält Beiträge der Ringvorlesung
der Privatdozierenden an der Universität Zürich im Herbstsemester 2008.

Gefördert mit Mitteln der Hochschulstiftung Zürich und
des Zürcher Universitätsvereins (ZUNIV).

Informationen zum Verlagsprogramm:
www.chronos-verlag.ch

© 2010 Chronos Verlag, Zürich
ISBN 978-3-0340-0983-6

Inhalt

Sinn, Sinne, Sinnessysteme – Bedeutungen, Perspektiven, Themen <i>Werner M. Egli</i>	7
Ohne Sinne kein Sinn und keine Sinne ohne Sinn <i>Erwin Sonderegger</i>	23
Sinneswahrnehmung und Moral. Phänomenologische und theologische Überlegungen zur Ethik <i>Markus Huppenbauer</i>	45
Die molekularen Mechanismen des Sehens <i>Oliver Zerbe</i>	63
Blicke aus Bildern. Fallstudien zu Franz von Lenbach, Candida Höfer und Franz Gertsch <i>Wolfgang Kersten</i>	79
Hören – eine entwicklungspsychologische Perspektive <i>Stefanie Stadler Elmer</i>	97
Die Wahrnehmung der Sprachlaute <i>Stephan Schmid</i>	123
Bambus als Achse des Sinnessystems in einer indigenen Psychologie des Himalajas. Die Sunuwar und ihr Sinn für Kommunikation <i>Werner M. Egli</i>	147
Körper und Sinne in der Kochkunst. Konzepte mittelalterlicher Nahrungsmitteldiätetik <i>Dorothee Rippmann</i>	167
Mit Herzen, Mund und Händen. Liedkultur und Biografie im christlichen Kontext <i>J. Jürgen Seidel</i>	197

Patient und technikorientierte Medizin am Beispiel der Urologie.
Verlieren wir unsere Sinne?

Hubert John

217

Die Sinne des Tieres. Ein neuer Ansatz im Tierschutz

Michael Hässig

223

Autorinnen und Autoren

229

Stephan Schmid

Die Wahrnehmung der Sprachlaute

Wissenschaftsgeschichtliche Streiflichter auf Phonetik und Phonologie

Alle Menschen sind befähigt mittels Sprache(n) miteinander zu interagieren. Dazu verfügen sie über ein implizites abstraktes Wissen, welches unter anderem aus einem Wortschatz sowie aus grammatikalischen Formen und Regeln besteht. Wenn nun in der alltäglichen Kommunikation solche im Langzeitgedächtnis gespeicherte mentale Strukturen in konkrete Mitteilungen umgesetzt werden, benötigen die an der Interaktion teilnehmenden Individuen – die man kommunikationstheoretisch als ‹Sender› und ‹Empfänger› bezeichnen kann – für die Übertragung der verbalen Äusserungen auch noch eine materielle Dimension: den ‹Kanal›.

Sprachliche Kommunikation manifestiert sich in zwei verschiedenen Übertragungsformen: Für die gesprochene Sprache brauchen wir den auditiv-akustischen Kanal, für die geschriebene Sprache den grafisch-visuellen Kanal. Der auditiv-akustische Kanal ist insofern primär, als alle hörfähigen Menschen natürlicherweise lernen, eine oder mehrere Sprachen zu verstehen und zu sprechen.¹ Lesen und Schreiben sind hingegen Kulturtechniken, die in der Regel durch schulisches Lehren vermittelt werden; zudem gibt es in der Welt Sprachen, die gar nicht verschriftet worden sind.² Somit wird klar, dass in der sprachlichen Kommunikation auch zwei verschiedene Sinne angesprochen werden: der Hörsinn für die primäre mündliche Sprache und der Gesichtssinn für die sekundäre geschriebene Sprache.

Der vorliegende Beitrag beschäftigt sich mit der Wahrnehmung der Sprachlaute und geht somit der Frage nach, welche besonderen Prozesse beim Hören von sprachlichen Äusserungen im Spiel sind. Dabei kann es keineswegs darum gehen, die gesamte Komplexität dieser Vorgänge zu beschreiben; vielmehr soll anhand einiger wissenschaftsgeschichtlicher Streiflichter gezeigt werden, mit welchen Theorien und Methoden die hier angesprochenen Phänomene in der Phonetik und in der Phonologie analysiert worden sind. Gleichzeitig werden wir in einzelnen Seitenblicken auch der Frage nachgehen, welche Beziehungen sich bei der Wahrnehmung von Sprachlauten zwischen Hörsinn und Gesichtssinn

ergeben; neben Hinweisen auf Fachliteratur werden auch zwei literarische Texte zur Veranschaulichung herbeigezogen.

Der Aufsatz ist im Wesentlichen in vier Abschnitte gegliedert. Im ersten Abschnitt wird zunächst der Erkenntnisgegenstand von Phonetik und Phonologie anhand der heutigen Einteilung und Abgrenzung vorgestellt, um anschliessend die historische Entwicklung und Ausdifferenzierung dieser beiden Teilgebiete der Linguistik nachzuvollziehen; in einem Exkurs wird das Phänomen der Synästhesie angesprochen. Der zweite Teil liefert elementare physiologische und physikalische Grundlagen zum Hören von Sprachlauten, bevor im dritten Teil verschiedene Probleme und Theorien der Lautwahrnehmung im engeren Sinn erörtert werden; dabei wird auch der Zusammenhang zwischen visueller und akustischer Perzeption angesprochen. Der vierte Teil bietet einen Ausblick in einzelne Aspekte der Perzeption von Prosodie, bevor einige kurze Schlussbemerkungen unsere Betrachtungen zu Ende führen.

Phonetik und Phonologie. Einteilung und Geschichte

Wenn wir in einem Lexikon der Sprachwissenschaft die Begriffe ‹Phonetik› und ‹Phonologie› nachschlagen, so erfahren wir, dass die Phonetik physiologische und physikalische Phänomene im Zusammenhang mit der Erzeugung von Vokalen und Konsonanten studiert, während die ‹Phonologie› die Funktionsweise von lautlichen Strukturen in bestimmten Sprachsystemen untersucht.³ Die Aufgabenteilung zwischen den beiden Schwesterdisziplinen sowie die interne Ausgliederung der Phonetik kann am besten anhand eines Schemas aufgezeigt werden, das auf die bereits erwähnten Faktoren des Kommunikationsprozesses ‹Sender› (Sprecher), ‹Empfänger› (Hörer), ‹Kanal› (Schallwellen) und ‹Kode› (Sprache) zurückgreift:

Die Phonologie entwirft Modelle zur Beschreibung und Erklärung des gemeinsamen Wissens über die Laute einer bestimmten Sprache, das gewissermassen in den Köpfen von Sprecher und Hörer gespeichert ist und diesen beiden erst die Bildung und Interpretation von Lautstrukturen ermöglicht. Sie stellt somit eine Beschreibungsebene der Linguistik im engeren Sinne dar – vergleichbar mit der Morphologie, der Syntax und der Semantik, welche die Strukturen von Wörtern, Sätzen und Bedeutungen analysieren.⁴

Die Phonetik untersucht hingegen die konkreten Vorgänge, welche die Produktion, Übertragung und Perzeption von Sprachlauten ermöglichen. Sie zerfällt in drei Teildisziplinen: die artikulatorische Phonetik beschreibt die Bewegungen der Sprechorgane, die akustische Phonetik misst die physikalischen Eigenschaften von Sprachschall, die auditive Phonetik untersucht die Aufnahme und

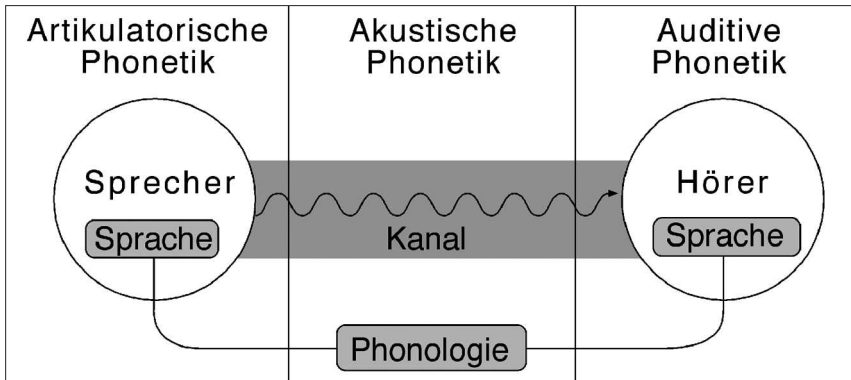


Abb. 1: Phonetik und Phonologie.

Verarbeitung des Sprachsignals im Ohr und im Gehirn. Selbstverständlich ergeben sich Beziehungen zwischen den verschiedenen Phasen der lautsprachlichen Kommunikation; für unseren Zusammenhang ist insbesondere die Frage von Belang, inwieweit die Wahrnehmung durch motorische Muster oder durch sprachspezifisches Wissen geprägt wird.

Betrachtet man die historische Entwicklung des Wissens über Sprachlaute, so stellt man ein gewisses Primat des phonologisch-sprachlichen und des artikulatorisch-physiologischen Ansatzes fest. Begriffe wie ‹Vokal›, ‹Konsonant› und ‹Silbe› gehen letztendlich auf die antiken Grammatiker und insbesondere auf Aristoteles (382–322 v. C.) zurück. Auch lieferte der Grieche Claudius Galenus (129–199), Leibarzt des römischen Kaisers Marcus Aurelius, bereits detaillierte Beschreibungen des Sprechapparats; die ältesten überlieferten Zeichnungen des Kehlkopfs stammen allerdings erst von Leonardo da Vinci (1452–1519).⁵ Dieses Primat der Beschreibung von Sprache und Artikulation gegenüber Akustik und Perzeption dürfte mit der phänomenologisch-funktionalen Plausibilität der Lautstrukturen für die verbale Kommunikation sowie mit der visuell-taktilen Zugänglichkeit von Sprechbewegungen zu tun haben.

Ein amüsantes Beispiel für die Anschaulichkeit artikulatorischer Bewegungen findet sich in der Komödie *Le bourgeois gentilhomme* von Molière (1670), und zwar im Dialog zwischen *Monsieur Jourdain* und dem *Maître de philosophie* (vierte Szene des zweiten Akts):⁶

MAÎTRE DE PHILOSOPHIE: Et là-dessus j'ai à vous dire que les lettres sont divisées en voyelles, ainsi dites voyelles parce qu'elles expriment les voix; et en consonnes, ainsi appelées consonnes parce qu'elles sonnent avec les voyelles, et

ne font que marquer les diverses articulations des voix. Il y a cinq voyelles ou voix: a, e, i, o, u.

MONSIEUR JOURDAIN: J'entends tout cela.

MAÎTRE DE PHILOSOPHIE: La voix A se forme en ouvrant fort la bouche: A. MONSIEUR JOURDAIN: A, A. Oui.

MAÎTRE DE PHILOSOPHIE: La voix E se forme en rapprochant la mâchoire d'en bas de celle d'en haut: A, E.

MONSIEUR JOURDAIN: A, E, A, E. Ma foi! oui. Ah! que cela est beau!

MAÎTRE DE PHILOSOPHIE: Et la voix I en rapprochant encore davantage les mâchoires l'une de l'autre, et écartant les deux coins de la bouche vers les oreilles: A, E, I.

MONSIEUR JOURDAIN: A, e, i, i, i, i. Cela est vrai. Vive la science!

Die Bildung der Sprachlaute kann also mindestens zum Teil mit dem Gesichts- und dem Tastsinn von jedem Sprecher propriozeptiv nachvollzogen werden. Es war wohl nicht zuletzt diese Einsichtigkeit des Sprechvorgangs, welche den Gelehrten und Automatenbauer Wolfgang von Kempelen (1734–1804) ab 1769 zur ersten artikulatorischen Modellierung inspirierte – anhand der Konstruktion einer Sprechmaschine, deren zugrunde liegende Prinzipien er in einem Traktat darlegte.⁷ Die Begründung der Phonetik als eigenständige wissenschaftliche Disziplin erfolgte jedoch erst im 19. Jahrhundert aufgrund der komplementären geistesgeschichtlichen Strömungen von Romantik und Positivismus.

Die romantische Suche nach dem Ursprung der europäischen (Volks-)kultur bildete den Hintergrund für die Entstehung der historisch-vergleichenden Sprachwissenschaft, welche die genetische Zusammengehörigkeit von Sprachfamilien durch die Rekonstruktion einer gemeinsamen «Ursprache» erschliessen wollte. Ein wichtiges methodisches Werkzeug für die Ausgliederung von Sprachfamilien und Sprachräumen ist der Begriff des «Lautgesetzes», wodurch Entwicklungen zwischen verschiedenen historischen Sprachstufen erfasst werden. Die Bedeutung der lautlichen Dimension erforderte allmählich eine Problematisierung der Schrift als nur bedingt verlässliches Zeugnis der gesprochenen Sprache, so dass aus dem Interesse an lautlichen Erscheinungen schliesslich die Notwendigkeit eines besseren Verständnisses der Artikulationsvorgänge erwuchs.⁸

Unter dem Vorzeichen des Positivismus standen hingegen die Fortschritte der Naturwissenschaften in der zweiten Hälfte des 19. Jahrhunderts. In diesen Zeitraum fallen sogar die ersten Vorläufer der akustischen und auditiven Phonetik. So formulierte der Physiologe und Physiker Hermann von Helmholtz (1821–1894) anhand von Experimenten mit kugelförmigen Resonatoren eine mathematische Erklärung der Klangfarbe aufgrund von Obertönen (vgl. *infra*).⁹ Der Ohrenheilarzt Oskar Wolf führte wiederum Untersuchungen zur Hörbarkeit

von Sprachlauten durch, indem er eine Versuchsperson auf einer Allee auf einen Sprecher zugehen liess.¹⁰

An ein sprachwissenschaftliches Publikum wandte sich schliesslich die Monografie *Grundzüge der Lautphysiologie* des Leipziger Philologen Eduard Sievers (1850–1932): Der Titel der Erstausgabe von 1876 bringt einerseits die positivistische Orientierung an den «exakten» Wissenschaften zum Ausdruck, während der Untertitel «zur Einführung in das Studium der Lautlehre der indogermanischen Sprachen» gleichzeitig die Verbindung zur («romantisch» motivierten) historisch-vergleichenden Sprachwissenschaft herstellt.¹¹ Wissenschaftshistorisch bemerkenswert ist die Tatsache, dass Sievers den Titel der zweiten Auflage (1881) in *Grundzüge der Phonetik* abänderte und somit eine terminologische Neuschöpfung übernahm, welche der Engländer Henry Sweet in seinem *Handbook of Phonetics* (1877) vorgeschlagen hatte.¹² Die neue Disziplin und ihre Bezeichnung fanden ihre endgültige Etablierung 1886 am Pariser Gründungskongress der *Association Internationale de Phonétique* (A.P.I), welchen der Franzose Paul Passy (1859–1940) organisiert hatte; zwei Jahre später wurde die erste Version des phonetischen Alphabets vorgelegt, welches nach mehreren Revisionen heute noch verbreitet ist.¹³ Die positivistische Ausrichtung der damaligen Phonetik zeichnete sich einerseits in der Verwendung von Untersuchungsinstrumenten wie etwa dem durch den Physiologen Carl Ludwig (1812–1895) entwickelten Kymographen,¹⁴ andererseits in ihrem Streben nach praktischen Anwendungsmöglichkeiten; so waren einige Phonetiker Fremdsprachenlehrer, während andere für eine Orthografiereform des Englischen eintraten.

Im Grossen und Ganzen beschäftigte sich die Phonetik am Ende des 19. Jahrhunderts also mit der experimentellen Untersuchung des Sprechvorgangs, aber auch mit der Beschreibung von Sprachen mittels phonetischer Zeichen. Diese linguistische Stossrichtung sollte in der ersten Hälfte des 20. Jahrhunderts an Bedeutung gewinnen aufgrund eines Paradigmenwechsels, der durch den Genfer Sprachwissenschaftler Ferdinand de Saussure (1857–1913) eingeleitet wurde. Saussure prägte in seinem *Cours de linguistique générale* (1916 posthum von seinen Schülern veröffentlicht) die Auffassung, dass die Sprachwissenschaft ihren eigentlichen Erkenntnisgegenstand nur aufgrund von dem virtuellen Sprachsystem inhärenten Oppositionen erschliessen konnte und sich somit nicht mit der materiellen Bedingtheit des Sprachgebrauchs auseinanderzusetzen brauchte: «la linguistique a pour unique et véritable objet la langue envisagée en elle-même et pour elle-même».¹⁵

Der strukturalistische Ansatz de Saussures wurde anfangs der Dreissiger Jahre von den Mitgliedern des Prager Linguistischen Zirkels weiterentwickelt und insbesondere auf die Analyse von Lautstrukturen angewendet. Eine *Summa* dieser Forschungsbemühungen legte der Russe Nikolaj S. Trubetzkoy (1890–1939)

in seinem Todesjahr mit dem Werk *Grundzüge der Phonologie* (man beachte den Titel!) vor, das zu einem Klassiker der Linguistik werden sollte. Darin wird gleich zu Beginn die Arbeitsteilung zwischen Phonetik und Phonologie programmatisch festgehalten: «Es empfiehlt sich daher, statt einer einzigen zwei verschiedene ‹Lautlehren› einzuführen [...]: die Sprechaktlautlehre, die mit konkreten physikalischen Erscheinungen zu tun hat, muss naturwissenschaftliche, die Sprachgebildelautlehre dagegen rein sprach- (bzw. geistes- oder sozial-) wissenschaftliche Methoden gebrauchen. Wir bezeichnen die Sprechaktlautlehre mit dem Namen Phonetik, die Sprachgebildelautlehre mit dem Namen Phonologie.»¹⁶

Eine phonologische Analyse nach strukturalistischen Prinzipien versteht die Lautgestalt einer Sprache als ein System von Gegensätzen. Jede Sprache trifft aus den vielfältigen Artikulationsmöglichkeiten, welche der menschliche Sprechapparat hervorbringen kann, eine kleine Auswahl und benutzt diese, um Bedeutungen voneinander zu unterscheiden. Es gibt vertauschbare und unvertauschbare Sprachlaute; nicht alle Schalleigenschaften sind für die Sprache relevant. Grundeinheit des phonologischen Systems ist nicht der konkret geäußerte Sprachlaut, sondern das ‹Phonem›, welches mittels der so genannten Austauschprobe ermittelt wird. So können wir aufgrund des ‹Minimalpaars› *Kasse* und *Gasse* zwei Phoneme /k/ und /g/ mit bedeutungsscheidender Funktion ermitteln. Hingegen spielt es keine Rolle, ob im Wort *rot* der Konsonant /r/ mit der Zungenspitze oder dem Halszäpfchen gebildet wird; es handelt sich in diesem Fall also um zwei Varianten desselben Phonems.

Der Phonembegriff ist eine wesentliche Errungenschaft der strukturalistischen Phonologie und wird bei Trubetzkoy rein operationell und sprachintern definiert, unter Verwahrung gegen jegliche psychologische Interpretation. Der Autor distanziert sich ausdrücklich von Termini wie ‹Lautvorstellung› oder ‹Lautabsicht›, die er früher selbst verwendet hatte: «[...] denn das Phonem ist ein sprachwissenschaftlicher und nicht ein psychologischer Begriff. Jede Bezugnahme auf das ‹Sprachbewusstsein› muss bei der Definition des Phonems ausgeschaltet werden.»¹⁷ Entgegen dieser wissenschaftssoziologisch nachvollziehbaren Abgrenzung der Linguistik gegenüber der Psychologie finden sich aber auch in den *Grundzügen der Phonologie* Textpassagen, welche Evidenz für die kognitive Realität von Phonemen und insbesondere für ihre Rolle bei der Wahrnehmung fremdsprachlicher Lautunterscheidungen liefern. So schreibt Trubetzkoy im Abschnitt *Falsche Beurteilung der Phoneme einer fremden Sprache*: «Das phonologische System einer Sprache ist gleichsam ein Sieb, durch welches alles Gesprochene hindurchgelassen wird. [...] Der Mensch eignet sich das System seiner Muttersprache an. Wenn er aber eine andere Sprache reden hört, wendet er unwillkürlich bei der Analyse

des Gehörten das ihm geläufige ‹phonologische Sieb› seiner Muttersprache an.»¹⁸ Die strukturalistische Begrifflichkeit legt also nahe, dass die Lautwahrnehmung im Wesentlichen phonemisch ist. Die Erfahrung zeigt auch, dass lautliche Unterscheidungen mit bedeutungsscheidender Funktion dem metasprachlichen Bewusstsein eher zugänglich sind als solche, welche reine Varianten eines Phonems betreffen.

Die strukturalistische Phonologie wurde von einem anderen Mitglied der ‹Prager Schule›, dem Russen Roman Jakobson (1896–1982) weiterentwickelt. Auf Jakobson geht insbesondere die Idee zurück, dass das Phonem als kleinste phonologische Einheit in weitere Bestandteile, die so genannten ‹distinktiven Merkmale› zerlegt werden kann. So unterscheiden sich zum Beispiel die beiden Phoneme /k/ und /g/, welche aus dem Minimalpaar *Kasse* vs. *Gasse* gewonnen wurden, artikulatorisch nur in einer einzigen Eigenschaft: Beide werden durch einen Verschluss am weichen Gaumen gebildet, aber bei [g] vibrieren zusätzlich die Stimmlippen, was bei [k] nicht der Fall ist. Die Stimmhaftigkeit ist in der Tat ein distinktives Merkmal, welches in den meisten Sprachen der Welt Phoneme voneinander unterscheidet.

Eine kohärente Theorie der distinktiven Merkmale wurde von Jakobson nach dem zweiten Weltkrieg zusammen mit Morris Halle und Gunnar Fant erarbeitet.¹⁹ Sie postuliert einen Katalog von vermutlich universellen, binären (zweiwertigen) Merkmalen, die nicht nur artikulatorisch, sondern zum Teil auch akustisch und auditiv definiert sind, wie aus den Bezeichnungen [gespannt/ungespannt], [kompakt/diffus] und [dunkel/hell] ersichtlich wird.²⁰

Sprachlaute hören und sehen (I). Synästhesie

Eigentlich hatte Jakobson das Merkmalskonzept schon in einer Studie angelegt, die er während des Zweiten Weltkriegs im schwedischen Exil veröffentlicht hatte.²¹ Seine Überlegungen zu Kindersprache und Aphasie (die ebenfalls zu den Klassikern des linguistischen Strukturalismus zählt) sind für unsere Belange insofern von Interesse, als sie explizit einen Vergleich zwischen Lautwahrnehmung und ‹Gesichtsempfindungen› herstellen: Vokale werden aufgrund der Gegensätze [dunkel/hell] und [farbig/farblos] definiert, während Konsonanten nur hell oder dunkel sind.²² Die beiden Dimensionen des Gesichtsinns sind hier nicht als rein terminologische Metaphern gemeint, sondern weisen durchaus auf phänomenologische Übereinstimmungen hin, wie Jakobson in einem Exkurs über die *audition colorée*, das Farbenhören, darzulegen versucht; so fügt er als Beispiele die Vokal-Farbzusammenordnungen von zwei Versuchspersonen mit unterschiedlicher Muttersprache an:²³

	Versuchsperson S. P.	Versuchsperson K. L.
a	rot	rot
o	rotblau	blau
u	dunkelblau	schwarzbraun
e	hellgrün	gelb
i	kanariengelb	silberweiss

Tab. 1: Farbenhören nach Jakobson (1944).

Wie aus Tab. 1 hervorgeht, bestehen zwar individuelle Unterschiede zwischen den beiden Versuchspersonen, aber man kann auch bestimmte Tendenzen in den Zuordnungen von Vokalen zu Farbempfindungen erkennen. Die Gemeinsamkeiten betreffen weniger die Dimension der Farbe (von den wenig gesättigten Farben der Hochzungenvokale [i] und [u] zur starken Farbe des Tiefzungenvokals [a]) als jene der Helligkeit (vom dunklen Hinterzungenvokal [u] zu den hellen Vorderzungenvokalen [i] und [e]). Die Unterscheidung zwischen «dunklen» und «hellen» Vokalen scheint in der Tat über eine grössere perzeptive Plausibilität zu verfügen, ist sie doch in der Umgangssprache eher verbreitet als die Vorstellung von «farbigen» und «farblosen» Vokalen (man denke etwa an die Bezeichnung der verschiedenen a-Qualitäten in den schweizerdeutschen Dialekten).

Bekanntlich ist das Farbenhören eine spezielle Ausprägung einer allgemeineren Erscheinung, der Synästhesie, d.h. der Auslösung einer Sinnesempfindung aufgrund eines Reizes in einem anderen Körpersinn. So können Klänge Wärmeempfindungen hervorrufen, Düfte mit Farben gekoppelt sein usw.; in der Forschung sind über dreissig verschiedene Erscheinungsformen von Synästhesie beschrieben worden, wobei die Zuordnung von Farben zu Buchstaben offenbar am häufigsten ist.²⁴

«Helle Töne» und «kalte Farben»: Synästhetische Metaphern treten in unserer Alltagssprache als feste Redewendungen auf, so dass sie uns eigentlich gar nicht bewusst sind. Der Terminus «Synästhesie» bezeichnet ja nicht nur ein wahrnehmungpsychologisches Phänomen, sondern auch eine rhetorische Figur, die vor allem in literarischen Texten häufig eingesetzt wird. In der Dichtung besteht das Wesen der Synästhesie allerdings meist in ihrem nicht-konventionellen Charakter, wie das berühmte Sonett *Voyelles* des französischen Dichters Arthur Rimbaud (1854–1891) zeigt:²⁵

Voyelles

A noir, E blanc, I rouge, U vert, O bleu: voyelles,

Je dirai quelque jour vos naissances latentes:

A, noir corset velu des mouches éclatantes

Qui bombinent autour des puanteurs cruelles,

Golfes d'ombre; E, candeurs des vapeurs et des tentes,
Lances des glaciers fiers, rois blancs, frissons d'ombelles;
I, pourpres, sang craché, rire des lèvres belles
Dans la colère ou les ivresses pénitentes;

U, cycles, vibrement divins des mers virides,
Paix des pâtis semés d'animaux, paix des rides
Que l'alchimie imprime aux grands fronts studieux;

O, suprême Clairon plein des strideurs étranges,
Silences traversés des Mondes et des Anges:
O l'Oméga, rayon violet de Ses Yeux!

Die Zuordnung von Farben zu Vokalen scheint sich hier ganz den gängigen synästhetischen Mustern zu entziehen, als wolle sie die Arbitrarität des sprachlichen Zeichens und die schöpferische Freiheit des Künstlers zelebrieren und somit neue Sinneszusammenhänge erschliessen. Doch überlassen wir die Analyse dieses faszinierenden poetischen Textes lieber den Fachleuten und kehren wir zu unserem Thema zurück, indem wir uns den materiellen Bedingungen der Wahrnehmung der Sprachlaute zuwenden.

Sprachlaute hören. Physiologische und physikalische Grundlagen

Damit ein Hörer die verbale Mitteilung eines Sprechers verstehen kann, muss er den Sprachschall empfangen und verarbeiten. Die Aufnahme des akustischen Signals (Rezeption) erfolgt – gleich wie bei allen Klängen oder Geräuschen – über die drei Teile des Ohrs: Aussenohr, Mittelohr und Innenohr. Die Luftdruckschwankungen der in die Ohrmuschel eintretenden Schallwellen treffen am Ende des äusseren Gehörgangs auf das Trommelfell; dort werden sie in Festkörperschwingungen umgewandelt und über die drei Gehörknöchel des Mittelohrs (Hammer, Amboss und Steigbügel) an das ovale Fenster (den Übergang zum Innenohr) geleitet, wobei der Schwingungsdruck gleichzeitig verstärkt wird. Im mit Lymphe gefüllten Schlauchsystem des Innenohrs – dem eigentlichen Gehörorgan – bahnen sich die Schallwellen dann ihren Weg durch die zweieinhalb Windungen der Hörschnecke. Am Abschluss der Cochlea befindet sich die Basilarmembran, wo über 20'000 Haarzellen eine Art «Analyse» der übertragenen Frequenzen (vgl. *infra*) durchführen und die empfangenen Reize schliesslich in elektrische Impulse umwandeln.²⁶

Von der Cochlea gelangen diese chemoelektrischen Signale schliesslich über den

Hörnerv zur auditiven Endverarbeitung ins Gehirn. Die zerebralen Hörzentren befinden sich seitlich in der Tiefe der sylvischen Spalte der Grosshirnrinde; in der linken Hirnhälfte liegt diese Hörrinde zwischen den beiden Arealen, die man traditionellerweise als «Sprachzentren» betrachtet, dem Broca-Areal und dem Wernicke-Areal. Die beiden Gehirnregionen sind benannt nach dem französischen Arzt Paul Broca (1824–1880) und dem deutschen Gelehrten Carl Wernicke (1848–1905). Broca stellte bei Verletzungen in der dritten Windung des Stirnlappens Aphasieformen fest, die insbesondere die Sprachproduktion (und somit auch die Artikulation der Sprachlaute) behinderten; Wernicke beschrieb hingegen bei der «sensorischen» Aphasie Störungen des Assoziationsvermögens und somit des Sprachverstehens, die er auf Läsionen in der ersten Windung des Schläfenlappens zurückführte.²⁷

Die funktionale Spezialisierung der beiden Hirnhälften, die so genannte «Lateralisierung» spielt eine Rolle bei der Verarbeitung von sprachlichen Signalen. Da gleich wie beim Sehen die Übertragung der Sinnesreize «gekreuzt» erfolgt (d.h. vom linken Ohr zur rechten Hirnhälfte und vom rechten Ohr zur linken Hirnhälfte), besteht für die Lautwahrnehmung ein Rechts-Ohr-Vorteil, wie die kanadische Neurologin Doreen Kimura mittels der Technik des so genannten «dichotischen» Hörens feststellte. Wenn man nämlich mit einem Kopfhörer die beiden Ohren mit unterschiedlichen sprachlichen Stimuli beschallt, so wird in der Regel die im rechten Ohr empfangene Information besser wahrgenommen.²⁸ Es würde zu weit führen, an dieser Stelle die sehr komplexen neurologischen Vorgänge bei der Wahrnehmung der Sprachlaute genau erörtern zu wollen.²⁹ Versuchen wir vielmehr, uns den akustischen Eigenschaften von Sprachschall zu nähern, indem wir zunächst rein phänomenologisch die verschiedenen Dimensionen des Hörens betrachten. So wie der Gesichtssinn gleichzeitig mehrere Aspekte eines Gegenstands erfasst (Grösse, Form, Farbe, Helligkeit, Lage im Raum, Bewegung), so nehmen wir von einem Klangereignis neben der räumlichen Position der Schallquelle zusätzlich Dauer, Lautstärke, Tonhöhe und Klangfarbe (*timbre*) wahr. Die deutschen Bezeichnungen dieser Dimensionen des Hörens weisen zum Teil einen metaphorischen Charakter auf: Wenn der zweite Teil des Kompositums «Lautstärke» in einem gewissen Sinn an den Tastsinn (oder zumindest an Festkörperdynamik) erinnert, so sind «Höhe» und «Farbe» klar der visuellen Wahrnehmung entlehnt. Weniger als mit Synästhesie haben wir es hier aber mit einer grundsätzlichen Eigenschaft sprachlicher Konzeptualisierung zu tun, die der räumlichen Erfahrung eine gewisse kognitive Priorität einräumt.³⁰

Alle fünf Dimensionen des Hörens (Position der Schallquelle, Dauer, Lautstärke, Tonhöhe und Klangfarbe) können anhand der akustischen Korrelate von Schallwellen beschrieben werden, d.h. von Longitudinalwellen, die man sich als

eine räumliche Abfolge von Druckminima und -maxima vorstellen kann.³¹ Die Schallausbreitung erfolgt in der Form einer Sinoidalschwingung, die sich wie ein Pendel in unterschiedlicher Richtung durch die Ruhelage bewegt. Die das Hörereignis auslösende Schallquelle kann anhand der unterschiedlichen «Phasen» der Schwingung (bzw. der jeweiligen Verschiebung ihres Startpunktes gegenüber der Ruhelage) lokalisiert werden; von Nutzen ist dabei die Tatsache, dass Menschen über zwei Ohren verfügen. Die Dauer von Sprachlauten misst man in der Regel an deren zeitlichen Ausdehnung in Millisekunden. Die Lautstärke hängt mit der Energie der von der Schallquelle ausgelösten Luftdruckschwankung zusammen und ist bei der Schwallwelle an der maximalen Auslenkung vom Nullpunkt, bzw. deren Amplitude, zu erkennen; als Messeinheit für den Schallpegel wird heute meist Dezibel (dB) verwendet.³² Das physikalische Korrelat der Tonhöhe ist die Grundfrequenz der Schallwelle, d.h. die Frequenz des niedrigsten Teiltons einer komplexen Schallwelle; diese wird in Hertz (Hz) gemessen.³³ Von den fünf Dimensionen des Hörens ist die Klangfarbe die komplexeste, weshalb zunächst deren linguistische Funktionalität sowie der Umfang des menschlichen Hörvermögens betrachtet werden sollen.

Mit Ausnahme der Phase sind nämlich alle akustischen Eigenschaften relevant für die Kodifizierung sprachlicher Information. Für die Wahrnehmung der Sprachlaute im engeren Sinn (also der einzelnen «Segmente» in der Lautkette) ist die Klangfarbe am wichtigsten: Sie ermöglicht die Unterscheidung zwischen bestimmten Vokalen (z.B. [a] und [e]) oder Konsonanten (z.B. [f] und [s]). Auch die Dauer von Sprachlauten kann distinktiv sein, insbesondere in Form einer phonemischen Opposition zwischen kurzen und langen Vokalen, wie man sie etwa im Deutschen anhand des Minimalpaars *Bahn* vs. *Bann* erkennen kann. Mittels unterschiedlicher Lautstärke können betonte Silben gegenüber unbetonten Silben hervorgehoben werden (wie zum Beispiel im deutschen Minimalpaar *úmfahren* vs. *umfáhren*); zur Kennzeichnung der phonologischen Kategorie «Akzent» wird allerdings auch die Modulation der Dauer und der Tonhöhe verwendet.³⁴ Schliesslich kann der Tonhöhenverlauf eines ganzen Satzes dessen Sinn ebenfalls verändern: Bei einer Frage (*Gehen wir?*) liegt am Satzende eher ein steigender Tonhöhenverlauf vor, während eine Aufforderung (*Gehen wir!*) eher mit einem hohen Ton zu Beginn geäussert wird.³⁵

In welchem Bereich der Lautstärke und der Tonhöhe (bzw. der Energie und der Frequenz) bewegt sich nun der Sprachschall? Die Energie von Klangereignissen erstreckt sich auf der logarithmischen Dezibelskala von 0 dB bis über 150 dB. Die Hörschwelle wird in der Regel bei 0 dB angesetzt, auch wenn das minimale Energieniveau für die Wahrnehmung von auditiven Reizen in Abhängigkeit des Frequenzbereichs variiert. Beim Flüstern erreichen wir ein Niveau von ungefähr 30 dB, während sich die Lautstärke für normale Konversation im Bereich von

50 dB bis 60 dB bewegt. Die Schmerzgrenze liegt in etwa zwischen 120 dB und 130 dB; ein lauter akustisches Ereignis, wie z. B. ein Gewehrschuss mit 170 dB, führt zu einer Schädigung des Gehörs.³⁶ Die Tonhöhe von gesprochener Sprache hängt vom Geschlecht der Sprecher sowie von weiteren individuellen Gegebenheiten ab; bei Männern liegt die mittlere Sprechstimmlage zwischen 100 und 150 Hz, bei Frauen zwischen 200 und 300 Hz, bei Kindern zwischen 300 und 450 Hz.³⁷ Beim Singen werden natürlich viel höhere Töne produziert; so erreicht die Sopranstimme der «Königin der Nacht» in Mozarts Oper *Die Zauberflöte* das «dreigestrichene f» mit einer Frequenz von ca. 1400 Hz. Das Hörvermögen erstreckt sich bei einem Neugeborenen über einen Frequenzbereich von gegen 20'000 Hz und nimmt dann mit zunehmendem Alter ab.

Für die Wahrnehmung von Sprache sind Frequenzbereiche bis zu 10'000 Hz relevant, die somit weit über den als Tonhöhe wahrgenommenen Grundfrequenzen liegen. Sprachschall ist in der Tat komplexer als eine einfache Sinuswelle. Erstens sind die Schallwellen von Sprachlauten nicht immer periodisch. Dies trifft im Prinzip nur für «stimmhafte» Laute zu, bei denen sich die Stimmlippen regelmässig öffnen und schliessen: Die quasi-periodischen Vokale und gewisse Konsonanten (z.B. [m] oder [l]) sind «Klänge» und verfügen somit über eine als Tonhöhe wahrnehmbare Grundfrequenz; man kann damit Melodien singen. In allen Sprachen gibt es aber auch «stimmlose» Konsonanten (z.B. [k] oder [s]), deren Schallwellen a-periodisch sind und die auditiv den Charakter von Geräuschen haben. Zweitens besteht Sprachschall aus der Überlagerung der Grundwelle mit vielen «Obertönen» (so genannten Harmonischen, deren Frequenzen ganzzahlige Vielfache des ersten Teiltons sind). Je nach Formgebung des «Vokaltrakts», d.h. der oberhalb der Stimmlippen sich befindlichen Hohlräume, werden gewisse Obertonbereiche verstärkt und andere abgeschwächt; der Vokaltrakt wirkt also als «Filter» für den in den Stimmlippen erzeugten Rohschall, woraus schliesslich die für die einzelnen Sprachlaute charakteristische «Klangfarbe» entsteht, welche den akustischen Unterschied zwischen den Vokalen [i] und [u] oder zwischen den Konsonanten [b] und [d] ausmacht.³⁸

Untersuchungen zur Akustik der Sprache stellen heute wohl den grössten Teil der phonetischen Forschung dar. Zwar waren die physikalischen Gesetzmässigkeiten komplexer Schwingungen schon seit dem Theorem von Jean-Baptiste Fourier (1768–1830) bekannt; einen wesentlichen Auftrieb erlangte dieser Forschungszweig jedoch mit der Möglichkeit, die akustischen Eigenschaften von Sprachlauten (bzw. die unterschiedliche Energieverteilung in deren Obertonbereichen) anhand von so genannten «Spektrogrammen» sichtbar zu machen. War das 1939 auf der Weltausstellung in New York vorgestellte Verfahren des *Visible Speech* – ebenso wie die nach dem Zweiten Weltkrieg entwickelten Nachfolgegeräte – aufgrund der relativ hohen Kosten noch wenigen Spezialisten zugäng-

lich, so hat seit der Jahrtausendwende der technologische Fortschritt zu einer wesentlichen Demokratisierung des Wissenschaftsbetriebs geführt. Insbesondere seit der Verfügbarkeit des *Freeware*-Programms *Praat*, das von den Holländern Paul Boersma und David Weenink entwickelt wurde, reicht ein gewöhnlicher Computer, um akustische Analysen gesprochener Sprache durchzuführen.³⁹

Sprachlaute wahrnehmen. Probleme und Theorien

Ein Vergleich zwischen den Teilgebieten der Phonetik zeigt, dass die auditive Phonetik gegenüber der artikulatorischen und akustischen nach wie vor ein Stiefkind darstellt. Die auditive Phonetik ist nicht nur die jüngste der drei Disziplinen, da sie sich praktisch erst nach dem Zweiten Weltkrieg entwickelte (gewissermaßen im Schlepptau der akustischen Phonetik und des sprachtechnologischen Fortschritts); der Wahrnehmung von Sprachlauten wird in Einführungen in die Phonetik meist auch am wenigsten Platz eingeräumt. Dies hängt nicht zuletzt damit zusammen, dass Einsichten in solche Vorgänge nur schwierig zu gewinnen sind, weshalb die in den letzten Jahrzehnten vorgeschlagenen theoretischen Modelle der Sprachwahrnehmung zum Teil auch relativ spekulativ blieben.

Eine erste zentrale Frage lautet in diesem Zusammenhang, inwieweit die Wahrnehmung der Sprachlaute sich von derjenigen anderer Höreindrücke unterscheidet. Die sensorische Aufnahme in den peripheren Hörorganen (Rezeption) funktioniert ja im Prinzip für alle akustischen Signale gleich; dennoch scheint die Schallübertragung im Ohr für Sprachsignale besonders gut eingerichtet zu sein. Viele Phonetiker würden in der Tat die Frage nach der Spezifität der Lautwahrnehmung bejahen. Anlass dazu geben nicht nur die bereits erwähnten Beobachtungen zum ›dichotischen Hören‹ aufgrund der Lateralisierung des Gehirns, sondern auch weitere, anhand von Perzeptionsexperimenten gewonnene Erkenntnisse.

Ein ›klassisches‹ Phänomen der Psychophysik ist insbesondere dasjenige der ›kategorialen Wahrnehmung‹.⁴⁰ Für nicht-sprachliche auditive Stimuli kommt offenbar eine ›kontinuierliche Wahrnehmung‹ zum Zug, die es ermöglicht, eine zunehmende Veränderung der Tonhöhe oder der Lautstärke in gleichmässigen Schritten festzustellen. Bei Sprachlauten hat man hingegen beobachtet, dass eine kontinuierliche Veränderung des akustischen Signals zunächst gar nicht, dann aber sprunghaft bemerkt wird. So modifizierte man zum Beispiel mittels Sprachsynthese Silben vom Typus [ba], [da] und [ga] derart, dass die spektralen Eigenschaften des Übergangs vom Konsonanten zum Vokal in mehreren Schritten gleichmässig verändert wurden. Spielte man nun Versuchspersonen ein Kontinuum von Silben vor, das sich fortlaufend vom Pol [ba] zum Pol [da] bewegte,

so änderte sich die Identifikation der Stimuli schlagartig an einem bestimmten Punkt des Kontinuums; daraus folgerte man, dass die Hörer die akustischen Signale von Sprachlauten in diskrete Kategorien einteilen.⁴¹

Auch wenn der empirische Befund der kategorialen Wahrnehmung eine wesentliche Erkenntnis liefert, so sind damit noch lange nicht alle Fragen zur kognitiven Verarbeitung von Sprachlauten gelöst. Welcher Art sind denn die relevanten Kategorien? Handelt es sich um distinktive Merkmale, um Phoneme oder gar um Silben? Sind diese Kategorien rein phonetisch-phonologisch oder sind andere Sprachebenen, vielleicht sogar andere Sinnesmodalitäten im Spiel? Welche Rolle hat dabei der Hörer: Dekodiert er einfach passiv das Signal, indem er die akustischen Informationen bestimmten Kategorien zuordnet, oder rekonstruiert er die Sprachlaute, indem er die Sinnesreize mit aus der Erfahrung bekannten Mustern vergleicht?

Hinzu kommen weitere Probleme, welche die auditive Phonetik in Betracht ziehen muss. Oft können nämlich die Sprachlaute im akustischen Signal gar nicht als einzelne Segmente isoliert werden, da sie sich zu einem grossen Mass zeitlich überlappen. Diese zum Teil fehlende Linearität des Sprachsignals hängt mit dem Mechanismus der «Koartikulation» zusammen: Vokale und Konsonanten werden bei der Sprachproduktion nicht aneinandergereiht wie Perlen auf einer Kette; vielmehr verändern die Sprechorgane aufgrund der grossen Geschwindigkeit der Artikulationsbewegungen kontinuierlich ihre Lage, so dass es zu verschiedenen Übergangs- und Verschmelzungserscheinungen kommt. Auf das Phänomen der Koartikulation lässt sich zudem ein weiteres fundamentales Problem der Lautwahrnehmung zurückführen: die fehlende akustische Invarianz. Vergleicht man etwa in den Spektrogrammen der Silben [da] und [di] den jeweiligen Übergang zwischen Konsonant und Vokal, so zeigt das Geräusch des Konsonanten [d] zwei stark verschiedene Formen der Energieverteilung in Abhängigkeit des folgenden Vokals.

Mit verschiedenen Theorien hat man versucht, die hier angesprochenen Fragen und Probleme zu ergründen. Im Sinne einer Auswahl sollen an dieser Stelle vier verschiedene Modelle erwähnt werden, wobei aus Platzgründen nur die jeweiligen Grundgedanken skizziert werden können. Es handelt sich um (i) die akustisch-auditive Theorie nach Gunnar Fant, (ii) die Kohorten-Theorie nach Marslen-Wilson/Tyler, (iii) das H&H-Modell nach Lindblom sowie die «motorische Theorie der Lautwahrnehmung» nach Liberman.⁴²

Das Lebenswerk des kürzlich verstorbenen Schweden Gunnar Fant (1919–2009) ist eigentlich eine akustische Theorie der Sprachproduktion, wie der Titel seines grundlegenden Werkes aus dem Jahr 1960 besagt.⁴³ Die Erkenntnisse zur Akustik der Sprachlaute haben aber durchaus auch Implikationen für die Lautwahrnehmung, welche Fant 1964 an einer Tagung und im Kapitel 10 einer späteren Mo-

nographie dargelegt hat.⁴⁴ Die akustisch-auditive Theorie der Sprachperzeption legt nahe, dass der Hörer die Sprachlaute anhand einer Reihe von akustischen Eigenschaften kategorisieren kann, die im Wesentlichen distinktiven Merkmalen entsprechen.⁴⁵ Die Kategorisierung ergibt sich aus binären Entscheidungen folgender Art: Ist das Signal periodisch oder aperiodisch? Falls es periodisch ist, kann es aufgrund der Amplitude als Vokal identifiziert werden? Ein solches Modell beschränkt sich auf die phonetisch-phonologische Ebene der Sprache, ohne übergeordnete (z.B. semantische oder syntaktische) Faktoren zu berücksichtigen; die Lautwahrnehmung erfolgt von den kleinsten Elementen zu den grösseren Einheiten (*bottom-up*) und schreibt dem Hörer eine grösstenteils ‹passive› Rolle zu.

Grundsätzlich entgegengesetzt, nämlich *top-down*, verfährt die Kohorten-Theorie, die eigentlich ein psycholinguistisches Modell der Worterkennung darstellt.⁴⁶ Sprachlaute werden darin als Bestandteile von phonotaktischen Mustern (Silben, Konsonantenverbindungen usw.) aufgefasst, die ihrerseits Wortformen, also bedeutungstragende Elemente bilden. Nun wird angenommen, dass Wörter im mentalen Lexikon aufgrund formaler Gemeinsamkeiten in ‹Kohorten› gespeichert sind; das Wort *Mandel* bildet zum Beispiel eine Kohorte mit phonetisch ähnlichen Wörtern wie *Mann*, *Mansarde*, *Mannequin*, *Mantel*, *Mangel* usw. Bei der Spracherkennung vergleicht der Hörer also die auditive Information mit vorhandenen Wortmustern und identifiziert schlussendlich das richtige Wort; die Wahrnehmung der Sprachlaute ist somit eingebunden in einen interaktiven Prozess, wobei der Hörer eine relativ aktive Rolle spielt.

Aktiv vergleichend ist der Hörer auch bei Lindbloms H&H-Theorie, welche – ähnlich wie diejenige von Fant – auch in Schweden als Modell der Sprachproduktion entwickelt wurde.⁴⁷ Die Abkürzung H&H steht für die beiden englischen Begriffe *Hyperspeech* und *Hypospeech*, die man auf Deutsch mit ‹Hyperartikulation› und ‹Hypoartikulation› übersetzen könnte. Ausgangspunkt von Lindbloms Überlegungen ist das Problem der fehlenden akustischen Invarianz, das in der realen lautsprachlichen Kommunikation noch viel akuter ist: Derselbe Sprecher kann ja das gleiche Wort völlig verschieden aussprechen, je nachdem ob er langsam und betont deutlich spricht (*Hyperspeech*) oder ob er schnell und undeutlich artikuliert (*Hypospeech*). Vor allem die Hypoartikulation bereitet der Lautwahrnehmung Schwierigkeiten, da manchmal ganze Phoneme oder sogar Silben im akustischen Signal gar nicht auszumachen sind.

Als Beispiel für die Hyperartikulation mag Abb. 2 dienen, welche die Wortfolge *der in einen warmen Mantel* zeigt, so wie sie der Phonetiker Klaus Kohler für die Illustration des Deutschen im Handbuch des Weltlautschriftvereins gelesen hat.⁴⁸ In diesem mit dem Computerprogramm *Praat*⁴⁹ erstellten Schaubild sieht man oben ein so genanntes ‹Oszillogramm›, welches die Wellenform des kontinuier-

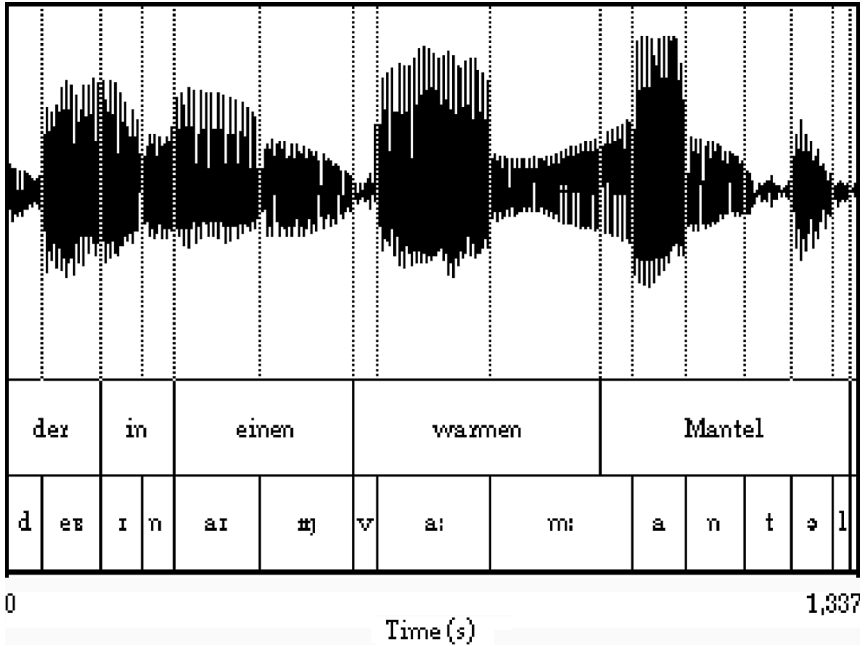


Abb. 2: Hyperartikulation im Deutschen.

lichen Sprachschalls abbildet; unten habe ich zweimal versucht, das Signal zu segmentieren – zunächst in Wörter, dann in Sprachlaute. Die mangelnde Übereinstimmung der beiden Segmentierungen zeigt sich vor allem bei den beiden Wörtern *einen* und *warmen*, wo in der Tat aufgrund der Hyperartikulation die Nebensilben zu einfachen Konsonanten reduziert wurden.

Da gemäss der H&H-Theorie der Hörer über ein implizites Wissen bezüglich der artikulatorischen Variation verfügt (er ist ja selbst auch ein Sprecher), kann er die im Sprachsignal nicht vorhandene Information kompensieren und in der effektiven Lautwahrnehmung die entsprechenden Phoneme ergänzen. Im Gegensatz zur akustisch-auditiven Theorie und zur Kohortentheorie stellt dieses Modell also eine explizite Beziehung zwischen Artikulation und Perzeption her; selbstverständlich ist der Hörer dabei aktiv.

Die Verbindung zwischen Sprachproduktion und Perzeption ist noch ausgeprägter in der <motorischen Theorie der Lautwahrnehmung> (*motor theory of perception*), welche am Haskins-Labor in New York von einer Forschergruppe um Alvin Liberman entwickelt wurde.⁵⁰ Das vielleicht bekannteste (weil zugleich provokativste) Modell geht in der Tat davon aus, dass der Hörer bei der Lautwahrnehmung das Signal aufgrund seines eigenen Artikulationsvermögens

bewertet, indem er die akustischen Sinnesreize aktiv mit inneren Sprechbewegungen vergleicht und dadurch zu einer Kategorisierung der Sprachlaute gelangt. Evidenz für diese Annahmen hat man darin gesehen, dass beim Hören von Sprachlauten manchmal gewisse Bewegungen der peripheren Sprechmotorik ausgelöst werden oder dass sogar im Gehirn neuronale Mechanismen aktiviert werden, die eigentlich für die Sprachproduktion zuständig sind.

Es würde zu weit führen, die Argumente für und wider die *motor theory of perception* hier *in extenso* auszuführen. Der Zusammenhang zwischen Artikulation und Perzeption soll hingegen unter einem anderen Aspekt erwähnt werden, der erneut den Gesichtssinn ins Spiel bringt: Welche Rolle spielt eigentlich die visuelle Wahrnehmung der Sprechbewegungen für die Kategorisierung der Sprachlaute durch den Hörer?

Sprachlaute hören und sehen (II). Der McGurk-Effekt

Im Gegensatz zur oben besprochenen Synästhesie geht es hier nicht um durch Klänge hervorgerufene Farbvisionen, sondern um die tatsächliche Hilfestellung, die das Beobachten der Sprechbewegungen für die Wahrnehmung der artikulierten Laute bieten kann. Im Prinzip handelt es sich dabei um Wissen, das den meisten Menschen implizit zur Verfügung steht. So bemerken wir manchmal bei Filmsynchronisierungen, dass ein Gesichtsausdruck nicht mit den gesprochenen Wörtern übereinstimmt: Sagt ein Schauspieler *I love you*, so rundet er beim Vokal des letzten Wortes den Mund, während der Vokal des entsprechenden deutschen Wortes *dich* mit gespreizter Lippenstellung gebildet wird.

Während also die meisten Menschen unbewusst zumindest teilweise die visuelle Information der Sprechorgane für die Wahrnehmung der Sprachlaute benutzen, erfährt bei Schwerhörigen oder Gehörlosen die Technik des Lippenlesens eine ganz andere Bedeutung, indem die fehlende oder mangelhafte Information des akustischen Signals durch visuelle Erfahrungen ergänzt oder ersetzt wird. Allerdings sind nicht alle Sprechvorgänge sichtbar: So werden z. B. [m] und [b] mit geschlossenen Lippen gebildet, während die die beiden Konsonanten unterscheidende Bewegung – das Heben oder Senken des Gaumensegels – mit bloßem Auge nicht zu erkennen ist.

In der theoretischen Diskussion zur Sprachverarbeitung hat die Frage der <heteromodalen> Perzeption seit den 70er Jahren des 20. Jahrhunderts an Bedeutung gewonnen, ausgelöst durch eine zufällige Entdeckung des englischen Entwicklungspsychologen Harry McGurk.⁵¹ Der nach ihm benannte Wahrnehmungseffekt kann in einem einfachen Experiment überprüft werden. Wenn Versuchspersonen in einer Filmaufnahme einen Sprecher sehen, der dreimal

die Silbe [ga] artikuliert, während sie gleichzeitig dreimal mit der Silbe [ba] beschallt werden, so nehmen viele Hörer die Silbe [da] wahr. Die Lautwahrnehmung verbindet offenbar die widersprüchlichen Sinneseindrücke und erzeugt so etwas wie einen Kompromiss zwischen den beiden Stimuli: Der Artikulationsort des alveolaren Verschlusslauts [d] liegt in der Tat zwischen denjenigen des bilabialen [b] und des velaren [g], so wie der relevante spektrale Index des rekonstruierten Konsonanten, der so genannte ‹Locus›, zwischen demjenigen der beiden anderen liegt.

Wahrnehmung von Prosodie

Zum Schluss unserer Ausführungen sollen kurz zwei weitere Aspekte angesprochen werden. Zum einen haben wir uns bis jetzt vor allem auf die Wahrnehmung einzelner Laute konzentriert und dabei die Tatsache vernachlässigt, dass Vokale und Konsonanten ja in grösseren zusammenhängenden Einheiten auftreten; neben der segmentalen Ebene gibt es auch eine suprasegmentale Ebene, zu welcher die sprachlichen Erscheinungen der Prosodie (Akzent, Intonation usw.) gehören. Zum anderen wurde die distinktive Funktion von Sprachlauten, bzw. von Phonemen, in den Vordergrund gerückt und somit die Verwendung von Sprache als Kommunikationswerkzeug, das der Unterscheidung und Übertragung von Bedeutungen und Inhalten dient. Durch die Verwendung bestimmter Sprachformen geben wir uns aber auch als Angehörige einer sozialen Gruppe zu erkennen; im Schweizerdeutschen können wir z.B. aufgrund der Klangfarbe des Vokals [a] Vermutungen über die geographische Herkunft des Sprechers anstellen. Zwei Experimente sollen hier vorgestellt werden: Das erste untersuchte die Wahrnehmung des Sprachrhythmus bei Kleinkindern, das zweite die Wahrnehmung dialektaler Prosodie bei Erwachsenen.

Die Frage, ob und wie sich Sprachen bezüglich ihrer rhythmischen Struktur unterscheiden, beschäftigt die Forschung in Phonetik und Phonologie seit mehreren Jahrzehnten; so hat man z. B. angenommen, dass im Spanischen die Silben eine mehr oder weniger gleichmässige Dauer aufweisen, während im Englischen die Akzentgruppen tendenziell gleich lang dauern.⁵² Interessante Hinweise zur Wahrnehmung des Sprachrhythmus liefern Ergebnisse aus der psycholinguistischen Forschung, die belegen, dass bereits Säuglinge auf Sprachen mit unterschiedlicher rhythmischer Struktur reagieren. In so genannten ‹Schnullerexperimenten› werden Neugeborenen Sätze aus verschiedenen Sprachen vorgespielt, während gleichzeitig ein in einem Schnuller eingebautes Messgerät die Saugrate misst; bekanntlich reagieren Säuglinge auf neue Reize mit einer erhöhten Saugrate, während diese bei gleich bleibender Umgebung

konstant bleibt oder abnimmt. Wie dem Forschungsbericht des Psycholinguisten Franck Ramus zu entnehmen ist, blieb die Saugrate bei französischen Neugeborenen gleich, wenn sie Englisch und Holländisch oder Spanisch und Italienisch hörten; hingegen nahm die Saugrate zu, wenn Spanisch mit Holländisch oder Englisch mit Italienisch abwechselte.⁵³ Daraus hat man u.a. gefolgert, dass die Sprachen der Welt in «Rhythmusklassen» eingeteilt werden können und dass der Rhythmus eine fundamentale Rolle für die Organisation der Prosodie spielt. Die Frage, inwieweit Erwachsene verschiedene Dialekte nur aufgrund prosodischer Merkmale unterscheiden können, wurde von einer Forschergruppe der Universität Bern untersucht, wobei auch das Phonetische Laboratorium der Universität Zürich bei der Datenerhebung beteiligt war.⁵⁴ In diesem Experiment wurden Sprachsignale aus vier schweizerdeutschen Dialekten (Bern, Wallis, Zürich, Graubünden) verändert, indem die für die Klangfarbe und somit für die einzelnen Sprachlaute massgebenden Frequenzbänder herausgefiltert wurden, so dass nur noch Lautstärke und Tonhöhe wahrnehmbar waren (der auditive Effekt ist vergleichbar mit der Situation, in der man in einem Hotelzimmer die Nachbarn sprechen hört, ohne dass man die einzelnen Wörter versteht). Studierende der Universitäten Bern und Zürich mussten anschliessend den Dialekt der manipulierten Stimuli identifizieren. Die allgemeine Erkennungsrate der vier Dialekte lag bei beiden Populationen knapp über dem Zufallsniveau von 25%; hingegen zeigten die Berner eine signifikant höhere Erkennungsrate für den Walliser Dialekt, während umgekehrt die Zürcher den Bündner Dialekte klar besser erkannten. Diese Ergebnisse belegen einerseits, dass es möglich ist, Sprachen oder Dialekte aufgrund von rein prosodischen Indizien zu erkennen. Andererseits weisen sie aber auch darauf hin, dass diese Fähigkeit durch Erfahrung gefördert werden kann; die Studierenden der Universität Bern sind offenbar eher mit Walliserdeutsch in Kontakt gekommen, während die Zürcher Studierenden besser mit der Bündner Prosodie vertraut sind.

Schlussbemerkungen

Es war nicht das Ziel dieses Beitrags, bahnbrechende Erkenntnisse über die Wahrnehmung der Sprachlaute zu liefern. Vielmehr ging es darum, einen Einblick in Fragestellungen und Resultate zu geben, die in der phonetischen und phonologischen Forschung diskutiert werden. Der wissenschaftsgeschichtliche Abriss hat einerseits deutlich gemacht, dass die Perzeption nach wie vor jenes Moment der lautsprachlichen Kommunikation bleibt, das am schwierigsten zu verstehen ist. Die miteinander konkurrierenden theoretischen Modelle weisen somit zwangsläufig einen etwas spekulativen Charakter auf; immerhin kommt

ihnen das Verdienst zu, die Forschung durch die Formulierung von Hypothesen anzuregen. Vielleicht ist die Sprachperzeption aber auch ein dermassen komplexes Phänomen, dass die verschiedenen Modelle einfach unterschiedliche Aspekte beleuchten; eine Kombination verschiedener Theorien dürfte am ehesten die Gesamtheit der beteiligten Faktoren beleuchten können. Hervorheben muss man an dieser Stelle auch die Bedeutung des technologischen Fortschritts, der insbesondere durch die Möglichkeiten der akustischen Analyse und Resynthese von Sprachsignalen neue Forschungsparadigmen eröffnet.

Anmerkungen

- 1 Einen Sonderfall stellen die Gebärdensprachen dar, die in den Gemeinschaften der Gehörlosen als Primärkode erworben werden. Wie die Schrift benutzt auch die Gebärdensprache den Gesichtssinn, allerdings auf eine komplexere und dynamischere Art und Weise. Schriftzeichen werden erkannt als Vordergrund vor einem Hintergrund (meist aufgrund von Helligkeits- und weniger aufgrund von Farbunterschieden), während die Gebärdensprache eine Reihe von Parametern wie Art und Richtung der Bewegung, Position gegenüber des Sprechers usw. verwendet; vgl. David Crystal: Die Cambridge Enzyklopädie der Sprache. Zürich 1993, 220–225 (Engl. Orig. 1987).
- 2 Bekanntlich kann man die verschiedenen Schrifttypen in zwei grobe Oberkategorien einteilen: logografische Schriftzeichen (Piktogramme und Ideogramme) verweisen direkt auf Begriffe oder Bedeutungen, während phonografische Schriftzeichen (Buchstaben oder Silbenzeichen) im Wesentlichen die Lautstruktur der gesprochenen Sprache abbilden; vgl. Crystal, Die Cambridge Enzyklopädie (wie Anm. 1), 194–203. Phonografische Schriften sind also in einem zusätzlichen Sinn sekundär gegenüber der mündlichen Kommunikation.
- 3 Vgl. Hadumod Bussmann: Lexikon der Sprachwissenschaft. 2. Auflage. Stuttgart 1990, 589–582; Helmut Glück (Hg.): Lexikon Sprache. Stuttgart 1993, 465–466.
- 4 Vgl. Angelika Linke/Markus Nussbaumer/Paul R. Portman: Studienbuch Linguistik. 5., erweiterte Auflage. Tübingen 2004, Kapitel 2 und 4.
- 5 Vgl. Maria Schubiger: Einführung in die Phonetik. 2. Auflage. Berlin 1977, 7–12; Bernd Pompino-Marschall: Einführung in die Phonetik. 2. Auflage. Berlin 2003, 5–9.
- 6 Zitiert aus der Larousse-Ausgabe, Molière: Le bourgeois gentilhomme. Paris 1970, 48. *Molière* ist das Pseudonym von Jean-Baptiste Poquelin (1622–1637).
- 7 Wolfgang von Kempelen: Mechanismus der menschlichen Sprache nebst Beschreibung einer sprechenden Maschine. Stuttgart 1970, 500 (Orig. 1791).
- 8 Allerdings überschrieb Jacob Grimm (1785–1863) das erste Buch seiner vierbändigen *Deutschen Grammatik* von 1822 noch mit *Über die Buchstaben*, obwohl die darin beschriebenen Erscheinungen eigentlich lautlicher Natur sind; vgl. Jacob Grimm: Deutsche Grammatik. 4 Bände. Göttingen, 1822–1837.
- 9 Hermann von Helmholtz: Die Lehre von den Tonempfindungen als physiologische Grundlage für die Theorie der Musik. Hildesheim 1968, 668 (Orig. Braunschweig 1862).
- 10 Oskar Wolf: Sprache und Ohr. Physiologisch-akustische und pathologische Studien. Braunschweig 1871, 71.
- 11 Eduard Sievers: Grundzüge der Lautphysiologie zur Einführung in das Studium der indogermanischen Sprachen. Hildesheim 1980, 150 (Orig. Leipzig 1876).
- 12 Die historische Persönlichkeit von Henry Sweet (1845–1912) wird oft mit der Figur des

- Professor Higgins in George Bernard Shaws Theaterstück *Pygmalion* (1914) in Verbindung gebracht. Der Autor schreibt dazu in seinem Vorwort: «Pygmalions Higgins is not a portrait of Sweet [...]; still, as will be seen, there are touches of sweet in the play» (zitiert nach der von Jacqueline Fisher betreuten Ausgabe bei Longman, Bernard Shaw: *Pygmalion*, London 1991, 4).
- 13 Der Weltlautschriftverein trägt heute den englischen Namen *International Phonetic Association* (IPA); vgl. *Handbook of the International Phonetic Association*. Cambridge 1999, 204; www.langsci.ucl.ac.uk/ipa/.
 - 14 Dieser «Wellenschreiber» überträgt die am Mund oder an der Nase austretenden Luftdruckschwankungen mittels einer Kapsel und eines Schlauchs auf eine Feder, deren Auslenkungen auf Russpapier aufgezeichnet werden, das auf eine sich drehende Trommel gespannt ist; vgl. Eugen Dieth: *Vademekum der Phonetik*. Bern 1950, 26.
 - 15 Zitiert nach der kritischen Ausgabe von Tullio De Mauro (Ferdinand de Saussure: *Cours de linguistique générale*. Paris 1986, 317).
 - 16 Nikolaj S. Trubetzkoy: *Grundzüge der Phonologie*. 7. Auflage. Göttingen 1989, 7 (Orig. 1939).
 - 17 Trubetzkoy, *Grundzüge der Phonologie* (wie Anm. 17), 37–38.
 - 18 Trubetzkoy, *Grundzüge der Phonologie* (wie Anm. 17), 47.
 - 19 Siehe Roman Jakobson/ Gunnar Fant/Morris Halle: *Preliminaries to speech analysis. The distinctive features and their correlates*. Massachusetts 1951, 64.
 - 20 Die deutschen Bezeichnungen für die Jakobsonschen Merkmale entnehmen wir Klaus Kohler: *Einführung in die Phonetik des Deutschen*. 2. Auflage. Berlin 1995, 128.
 - 21 Roman Jakobson: *Kindersprache, Aphasie und allgemeine Lautgesetze*. Frankfurt am Main 1969, 143 (Orig. Uppsala 1944).
 - 22 Vgl. Jakobson, *Kindersprache* (wie Anm. 21), 101–105.
 - 23 Vgl. Jakobson, *Kindersprache* (wie Anm. 21), 114–116.
 - 24 Eine Übersicht über verschiedene Formen der Synästhesie findet man auf der Web-Seite des Lehrstuhls für Neuropsychologie der Universität Zürich: www.psychologie.uzh.ch/fachrichtungen/neuropsych/Forschung/KonkreteForschungsthemen/Synaesthesia.html
 - 25 Zitiert aus folgender Ausgabe: Arthur Rimbaud: *Poésies. Une saison en enfer. Illuminations*. Paris 1965, 78–79.
 - 26 Für eine genauere Beschreibung des Gehörs sei verwiesen auf Pompino-Marschall, *Einführung in die Phonetik* (wie Anm. 5), 143–146, sowie auf Henning Reetz: *Artikulatorische und akustische Phonetik*. Trier 1999, 149–170.
 - 27 Die neuere Hirnforschung lehnt diese Einteilung jedoch als zu schematisch ab und betont insbesondere auch die Rolle der rechten Hirnhälfte für die Wahrnehmung der «Prosodie» (vgl. *infra*); siehe Cyrill Ott/Martin Meyer: *Functions of the left and right posterior temporal lobe during segmental and suprasegmental speech perception*. Im Druck in: Stephan Schmid (Hg.): *La dimensione temporale del parlato/The temporal dimension of speech*. Torriana, 2010.
 - 28 Doreen Kimura: *Cerebral dominance and the perception of verbal stimuli*. In: *Canadian Journal of Psychology* 15, 156–165; vgl. Pompino-Marschall, *Einführung in die Phonetik* (wie Anm. 5), 169–170.
 - 29 Die Funktionsweise des Nervensystems wird u.a. beschrieben in Magnús Pétursson/Joaachim Neppert: *Elementarbuch der Phonetik*. 2. Auflage. Hamburg 1996, 177–187; vgl. auch Pompino-Marschall, *Einführung in die Phonetik* (wie Anm. 5), 149–151.
 - 30 Vgl. John Lyons: *Semantics. Volume II*, Cambridge 1977 (deutsche Übersetzung: *Semantik*. Band II. München 1983, 322 ff.).
 - 31 Eine ausführlichere Darstellung der akustischen Grundlagen findet man z.B. bei Pompino-Marschall, *Einführung in die Phonetik* (wie Anm. 5), 87–133 sowie bei Reetz, *Artikulatorische und akustische Phonetik* (wie Anm. 26), 3–99; siehe auch Joachim Neppert: *Elemente einer akustischen Phonetik*. 4. Auflage. Hamburg 1999, 19–254.
 - 32 Das Dezibel ist ein Zehntel eines Bel. Eigentlich handelt es sich bei der Einheit Bel – deren Namen auf den Erfinder des Telefons, Alexander Graham Bell (1847–1922), zurückgeht – um

- ein Mass für die Übertragung elektrischer Energie und nicht um ein Mass für Luftdruckschwankungen.
- 33 1 Hz = 1 Zyklus der Schallwelle pro Sekunde; die Frequenz ist somit indirekt proportional zur «Periode», also zur Zeit, die verstreicht, bis die Schallwelle zweimal durch den Nullpunkt geht. Der Name der Masseinheit geht auf den deutschen Physiker Heinrich Hertz (1857–1894) zurück.
 - 34 In den so genannten Tonsprachen werden durch Tonhöhenbewegungen sogar einzelne Wörter unterschieden: im Chinesischen hat die Lautfolge /ma/ je nach Grundtonverlauf vier verschiedene Bedeutungen; siehe Pompino-Marschall, Einführung in die Phonetik (wie Anm. 5), 244.
 - 35 Die drei «suprasegmentalen» oder «prosodischen» Eigenschaften Dauer, Lautstärke und Tonhöhenverlauf üben neben sprachlichen auch paraverbale Funktionen aus, indem sie Emotionen ausdrücken oder die kommunikative Interaktion mit dem Gesprächspartner regeln.
 - 36 Für die Lautstärke als auditive Grösse wird oft auch die Einheit «Phon» verwendet: vgl. Pompino-Marschall, Einführung in die Phonetik (wie Anm. 5), 153–154; Neppert, Elemente einer akustischen Phonetik (wie Anm. 31), 84.
 - 37 Vgl. Neppert, Elemente einer akustischen Phonetik (wie Anm. 31), 125.
 - 38 Für eine eingehende Erläuterung dieses «Quelle-Filter-Modells» (einer akustischen Theorie des Sprechens) siehe Pompino-Marschall, Einführung in die Phonetik (wie Anm. 5), 99–132 sowie Pétursson/Neppert, Elementarbuch der Phonetik (wie Anm. 29), 125–149.
 - 39 Siehe die Webseite <http://www.fon.hum.uva.nl/praat/> sowie die Anleitung in Pompino-Marschall, Einführung in die Phonetik (wie Anm. 5), 133–139.
 - 40 Vgl. Stevan R. Harnad (Hg.): *Categorical perception: the groundwork of cognition*. Cambridge 1987.
 - 41 Alvin M. Liberman/Katherine S. Harris/Howard S. Hoffman/Belver C. Griffith: The discrimination of speech sounds within and across phoneme boundaries. In: *Journal of Experimental Psychology* 54 (1957), 358–368. Zur kategorialen Wahrnehmung von Sprachlauten siehe auch Neppert, Elemente einer akustischen Phonetik (wie Anm. 31), 288–290 sowie Pompino-Marschall, Einführung in die Phonetik (wie Anm. 5), 162–164.
 - 42 Eine eingehendere Erläuterung von verschiedenen Perzeptionstheorien findet man bei Pétursson/Neppert, Elementarbuch der Phonetik (wie Anm. 29), 191–205 sowie bei Pompino-Marschall, Einführung in die Phonetik (wie Anm. 5), 160–172.
 - 43 Gunnar Fant: *Acoustic theory of speech production*. Den Haag 1960, 328.
 - 44 Gunnar Fant: *Auditory patterns of speech*. In: Weiant Wather-Dunn (Hg.): *Models for the perception of speech and visual form*. Cambridge (Massachusetts) 1967, 111–125; Gunnar Fant: *Speech sounds and features*. Cambridge (Massachusetts) 1973, 160–170.
 - 45 Zusammen mit Jakobson war Fant ein wesentlicher Wegbereiter der Theorie der distinktiven Merkmale (vgl. Anm. 19).
 - 46 William Marslen-Wilson/Lorraine Tyler: The temporal structure of spoken language understanding. In: *Cognition* 8 (1980), 1–71.
 - 47 Björn Lindblom: Explaining phonetic variation: a sketch of the H&H theory. In: William J. Hardcastle/Alain Marchal (Hg.): *Speech production and speech modelling*. Dordrecht 1990, 403–439.
 - 48 Klaus Kohler: German. In: *Handbook of the IPA* (wie Anm. 13), 86–89.
 - 49 Vgl. Anm. 39.
 - 50 Alvin M. Liberman/Franklin S. Cooper/Donald P. Shankweiler/Michael Studdert-Kennedy: Perception of the speech code. In: *Psychological review* 74 (1967), 431–461.
 - 51 H. McGurk/John MacDonald: Hearing lips and seeing voices. In: *Nature* 164 (1976), 746 bis 748.
 - 52 Vgl. Pompino-Marschall, Einführung in die Phonetik (wie Anm. 5), 248–249.
 - 53 Franck Ramus/Marina Nespor/Jacques Mehler: Correlates of linguistic rhythm in the speech signal. In: *Cognition* 73 (1999), 285.

-
- 54 Adrian Leemann/Beat Siebenhaar: Perception of dialectal prosody. In: Proceedings of Interspeech 2008, Brisbane 2008, 524–527.