

# Artikulatorische und akustische Phonetik

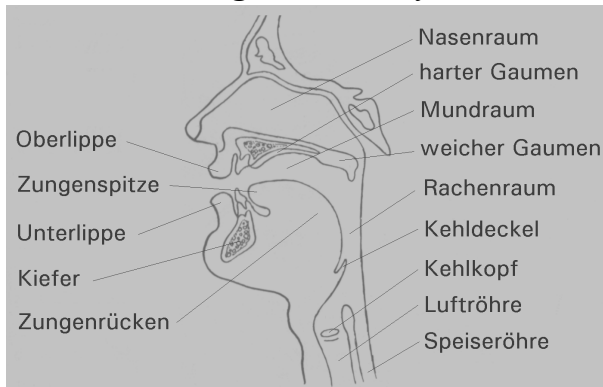
## Ein Kurzüberblick

Bernd J. Kröger

<http://www.speechtrainer.eu/>

### 1 Einführung

Der Sprechapparat des Menschen kann funktional wie anatomisch in 3 Bereiche unterteilt werden. Die **Lunge** stellt aerodynamische Energie (Luftdruck und Luftstrom) zur Verfügung



(Initiation, siehe Catford 1988). Im **Kehlkopf** wird mittels dieser Energie ein akustisches Signal produziert (Phonation) und im **Ansatzrohr** (Rachen-, Mund- und Nasenraum) wird dieses Signal moduliert (Artikulation). Somit kann der Sprechapparat in den **sublaryngalen Bereich** (Lunge, Bronchien und Luftröhre), den **Larynx** (Kehlkopf)

**Abbildung 1.** Der Sprechapparat des Menschen

und den **supralaryngalen Bereich** (Rachen-, Mund- und Nasenraum) unterteilt werden (s. Abb. 1).

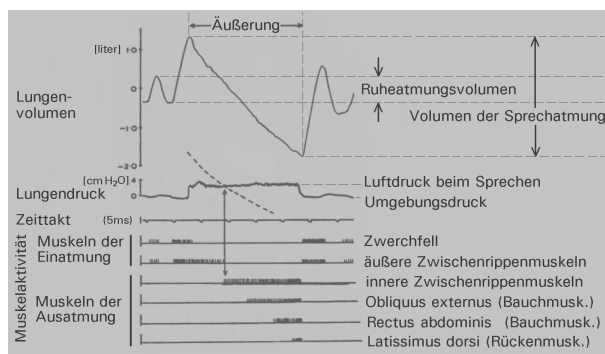
### 2 Initiation

Die Lungenaktivität während des Sprechens hat – neben den primär lebenserhaltenden Funktionen (Gasaustausch mit dem Blutkreislauf) - aus der Sicht der Sprachproduktion zwei wichtige Funktionen: Zum einen die Herstellung und Aufrechterhaltung eines zeitlich konstanten **Luftüberdrucks** in der Lunge, zum anderen die Bereitstellung einer bestimmten Luftmenge als Basis für den zum Sprechen nötigen **Luftstrom**. Die Aktivität der Gesamtheit der Atmungsmuskulatur ist während des Sprechens darauf ausgerichtet, den Lungenluftdruck annähernd konstant zu halten (siehe Ohala 1990).

Anatomisch wird die Lunge vom Brustkorb (Brustbein und Rippen) und nach unten vom Zwerchfell gegenüber dem Bauchraum abgegrenzt. Entsprechend unterscheiden wir **Brustatmung** (Hebung und Senkung des Brustkorbes) und **Bauchatmung** (mit Zwerchfellaktivität verknüpfte Bauchwandbewegung). Wichtige Muskeln der Einatmung sind der Zwerchfell-

muskel (Diaphragma), dessen Kontraktion die Absenkung des Zwerchfells bewirkt, und die äußeren Zwischenrippenmuskeln (Mm. Intercostales externi), deren Kontraktion die Hebung des Brustkorbes bewirken. Darüber hinaus tragen Teile der Hals-, Schulter- und Rückenmuskulatur zur Einatmung bei. Wichtige Muskeln der Ausatmung sind die inneren Zwischenrippenmuskeln (Mm. intercostales interni) aber auch Teile der Bauch- und Rückenmuskulatur. (siehe Abb. 2). Die Muskulatur dient der Änderung des Lungenvolumens und steuert so den Luftdruck in der Lunge.

Die Aktivität der Atmungsmuskulatur beim Sprechen dient der Aufrechterhaltung eines **zeitlich konstanten Überdrucks** in der Lunge. Nach dem Einatmen ist der Brustkorb gedehnt. Zusätzlich zu den muskulären Kräften der Ein- bzw. Ausatmung sind elastische Kräfte wirksam. Der zu den elastischen Kräften korrespondierende Druck sinkt mit Abnahme des Lungenvolumens (Abb. 2, gestrichelte Linie). Zur Realisierung eines zeitlich konstanten Drucks beim Sprechen werden die elastischen Kräfte zusammen mit den inspiratorischen und expiratorischen muskulären Kräften stetig ausbalanciert (Abb. 2).



**Abbildung 2.** Lungenvolumen, Lungenluftdruck, Zeittakt eines Zeitgebers und die Aktivität der Atmungsmuskulatur während der Produktion einer Äußerung (nach Draper, Ladefoged et al. 1959). Die Zeitachse verläuft horizontal von links nach rechts.

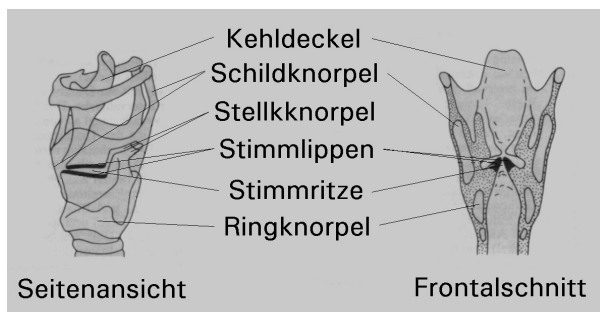
Der Lungendruck wird über Änderungen des Lungenvolumens gesteuert. Das maximale Volumen der Lunge eines erwachsenen Menschen mittleren Alters liegt bei rund 4-6 Litern (Maximalvolumen). Nach Ausatmen mit maximaler Muskelkraft liegt ein Lungenvolumen von rund 1-2 Litern vor (Residualvolumen). Die Differenz von 3-5 Liter wird als **Vitalkapazität** bezeichnet. Die Kapazität der **Ruheatmung** liegt bei rund 0,5-2 Litern, die der **Sprechatmung** liegt je nach Lautstärke zwischen Ruheatmung (leises Sprechen) und Vitalkapazität (sehr lautes Sprechen).

Die Stärke der während des Sprechens durch Kehlkopf und Ansatzrohr fließenden Luftstroms hängt aber nicht nur von der Aktivität der Atmungsmuskulatur sondern auch wesentlich von dem **Grad der Engebildung** (Konstriktionsbildung) im Bereich des Kehlkopfes und des Ansatzrohres ab. Bei geringer Engebildung im Kehlkopf und im Ansatzrohr kann in kurzer Zeit

eine (zu) große Luftmenge ausströmen. Deshalb tritt beim Sprechen zu fast allen Zeiten eine Engebildung entweder im Kehlkopf oder im Ansatzrohr auf.

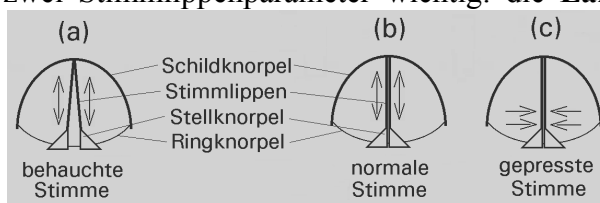
### 3 Phonation

Im Kehlkopf wird ein Teil der aerodynamische Energie (Überdruck unterhalb des Kehlkopfes und Luftstrom durch den Kehlkopf) in **Schwingungen der Stimmlippen** und in ein **akustisches Signal** (Schwingungen der Luftsäule im Ansatzrohr) gewandelt. Die Entstehung von Stimmlippenschwingungen und damit die Entstehung eines akustischen Signals im Kehlkopf (Stimmton) wird als **Phonation** bezeichnet. Der Kehlkopf befindet sich zwischen Luftröhre und Rachenraum (Abb. 1). Er besteht aus Knorpeln, Muskulatur und Schleimhaut (Abb. 3).



**Abbildung 3.** Seitenansicht und hinterer Frontalschnitt des Kehlkopfes

Der Ringknorpel (Cricoid) bildet den oberen Abschluss der Luftröhre. Mit dem Ringknorpel verbunden sind der nach vorne kippbare Schildknorpel (Thyroid) und zwei Stellknorpel (Arytenoid). Zwischen Schild- und Stellknorpel verlaufen die Stimmlippen. Die Stimmlippen bestehen aus Muskelfasern (M. vocalis oder M. thyroarytenoideus, TA) und mucosa (Schleimhaut). Durch Änderung des Abstandes zwischen den Stellknorpeln ändert sich der Abstand der Stimmlippen zueinander (Abb. 4). Die Öffnung zwischen den Stimmlippen wird auch als **Stimmritze** (Glottis) bezeichnet. Aus funktional phonetischer Sicht sind zwei Stimmlippenparameter wichtig: die **Längsspannung der Stimmlippen** und der **Abstand der Stimmlippen** zueinander. Durch Kippen des Schildknorpels und durch die Aktivität des M. vocalis kann die Längsspannung der Stimmlippen und damit die Höhe des Stimmtons geändert werden. Die Positionierung von Schild- und Stellknorpel geschieht durch die innere Kehlkopfmuskulatur. Wichtige Kehlkopfmuskeln zur Positionierung der Stellknorpel (und damit zur Einstellung des Abstandes der Stimmlippen zueinander) sind der Stimmlippenöffner (M. cricoarytenoideus posterior, PCA) und die Stimmlippen-schließer (M. cricoarytaenoideus lateralis, LCA, M. interarytenoideus, IA). Die Positionierung des Schildknorpels wird in erster Linie vom Stimmlippenspanner (M. cricothyroideus, CT) geleistet.

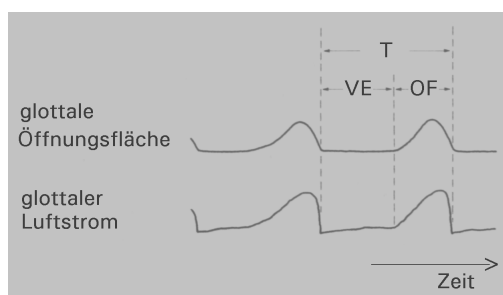


**Abbildung 4.** Aufsicht auf die Stimmritze. Die Pfeile stellen die parallel bzw. senkrecht zur Stimmritze wirkenden Kräfte dar.

geschieht durch die innere Kehlkopfmuskulatur. Wichtige Kehlkopfmuskeln zur Positionierung der Stellknorpel (und damit zur Einstellung des Abstandes der Stimmlippen zueinander) sind der Stimmlippenöffner (M. cricoarytenoideus posterior, PCA) und die Stimmlippen-schließer (M. cricoarytaenoideus lateralis, LCA, M. interarytenoideus, IA). Die Positionierung des Schildknorpels wird in erster Linie vom Stimmlippenspanner (M. cricothyroideus, CT) geleistet.

Für die Erzeugung von regelmäßigen Schwingungen der Stimmlippen (Phonation) ist neben der Bereitstellung von aerodynamischer Energie auch eine geeignete Positionierung der Stimmlippen, nämlich der lockere Glottisschluss und eine tonhöhenabhängige Vorspannung der Stimmlippen längs der Stimmritze nötig (Abb. 4b). Je höher die Längsspannung der Stimmlippen, um so höher die Tonhöhe des erzeugten Schallsignals. Man unterscheidet aktive Spannung der Stimmlippen (Aktivität des TA) und passive Spannung der Stimmlippen (durch Kippen des Schildknorpels, Aktivität des CT). Der richtige Abstand der Stimmlippen zueinander (Stimmlippenabstand) entscheidet über das Auftreten von Phonation. Bei sehr festem Verschluss der Stimmlippen, d.h. beim Zusammenpressen der Stimmlippen ist das Schwingen der Stimmlippen nicht mehr möglich. Aber auch bei zu weitem Öffnen der Stimmritze erlischt die Schwingung der Stimmlippen. Somit ist der lockere Glottisschluss Voraussetzung für die Phonation.

Bei der **Phonation** werden die Stimmlippen durch die Kehlkopfmuskulatur wie oben beschrieben eingestellt und gespannt. Die Schwingung der Stimmlippen führt zu einem ständigen Wechsel von phonatorischer Glottisöffnung und Glottisschluss (Abb. 5). Die Stimmlippen-schwingung entsteht aber nicht direkt aufgrund der Aktivität der Kehlkopfmuskulatur (schnelle Abfolge von Muskelkontraktionen), sondern aufgrund der Bereitstellung aerodynamischer Energie zusammen mit der richtigen laryngalen Voreinstellung (lockerer nicht zu fester Verschluss der Stimmlippen und Längsspannung). Es sind insbesondere zwei Mechanismen der Energieübertragung zu nennen, die zeitlich geschickt (d.h. „phasenrichtig“) zusammenarbeiten, um die Schwingung einzuleiten und aufrecht zu erhalten. (1) Der Luftdruck unterhalb der Stimmlippen drückt die geschlossenen Stimmlippen auseinander (Öffnungskraft innerhalb der Verschlussphase). (2) Die hohe Strömungsgeschwindigkeit der Luft zwischen den (dann) geöffneten Stimmlippen sorgt für die Ausbildung eines Unterdrucks zwischen den Stimmlippen (Bernoulli-Effekt) und damit für eine Anziehung der Stimmlippen (Schließkraft innerhalb der Offenphase). Der Vorgang beginnt dann erneut.



**Abbildung 5.** Öffnungsfläche der Stimmritze und zugehöriger Luftstrom als Funktion der Zeit. T ist die Schwingungsdauer einer Schwingungsperiode, OF kennzeichnet die Offenphase der Stimmlippen und VE die Verschlussphase der Stimmlippen innerhalb einer Schwingungsperiode.

**Anmerkung:** Dieser Mechanismus zur Erzeugung und Aufrechterhaltung der Stimmlippen-schwingung beim Sprechen (auch beim Singen) existiert in vergleichbarer Weise auch bei der

Tonerzeugung des Blechbläfers mit seinem Blechblasinstrument. Hier wird nun die Funktion der Stimmlippen von den Lippen des Bläfers übernommen.

**Anmerkung:** Der phonationsbedingte Wechsel von Öffnungs- und Schließbewegung der Stimmlippen ist luftstrombedingt (aerodynamisch initiiert) und klar von muskulär initiierten glottalen Öffnungs- und Schließbewegungen (Abduktions- und Adduktionsbewegungen) zu trennen. Abduktionsbewegungen dienen der Erreichung der Atemstellung; Adduktionsbewegungen dienen der Einstellung der Stimmlippen zur Phonation. Eine Abduktions- und nachfolgend eine Adduktionsbewegung treten auch bei der Realisierung eines stimmlosen Lautes im Kontext stimmhafter Laute auf. Muskulär initiierte Ab- und Adduktionsbewegungen laufen langsamer ab als die aerodynamisch initiierten Öffnungs- und Schließbewegungen der Stimmlippen.

Die **Tonhöhe** der Stimme wird in erster Linie durch die Stärke der Stimmlippenspannung (Längsspannung der Stimmlippen) determiniert, die **Lautstärke** hauptsächlich durch die Höhe des Lungenluftdrucks (und damit des Luftdrucks unterhalb der Stimmlippen). Je höher der Lungenluftdruck, um so lauter erklingt die Stimme.

Der Abstand der Stimmlippen hingegen determiniert die **Stimmqualität**. So wird die Stimme als „knarrend“ oder „gepresst“ bezeichnet, wenn starke Verschlusskräfte senkrecht zur Stimmritze wirken (Abb. 4c). Dies tritt auch bei der **Glottalisierung** auf. Sind diese Verschlusskräfte sehr stark, kann keine Schwingung der Stimmlippen mehr auftreten. Die Stimmlippen sind nun fest verschlossen. Dieser feste Verschluss der Stimmritze ist das Merkmal des **Glottisverschlusslautes** [ʔ] und tritt z.B. im Wörtern mit anlautendem Vokal auf (z.B. das Gegensatzpaar „Heiner“ und „Einer“ am Wortanfang von „Einer“). (Alle hier benutzten phonetischen Lautsymbole folgen den Vorgaben des Handbook of IPA, 1999). Sind die Stimmlippen hingegen leicht geöffnet (und zwar bereits vor Einsetzen der Schwingung der Stimmlippen), so ergibt sich die Stimmqualität „behaucht“ (Abb. 4a). Hier dürfen die Stimmlippen allerdings nicht zu weit geöffnet sein, da sonst keine Stimmlippenschwingung mehr auftritt. Das Vorhandensein bzw. das Nichtvorhandensein von Stimmlippenschwingungen unterscheidet genau die Lautkategorien **stimmhafte und stimmlose Laute** (stimmlos vs. stimmhaft: z.B. [t] vs. [d] in „Tank“ kontra „Dank“). So ist im Fall der **stimmlosen Laute** (mit Ausnahme des [ʔ]) eine weite Öffnung der Stimmritze gegeben (Abb. 4a). Normale Stimmqualität wird, wie oben beschrieben, bei locker verschlossenen Stimmlippen produziert (Abb. 4b).

**Anmerkung:** Auch im Fall des Parameters der Stimmlippenspannung und der Öffnungsweite der Stimmritze gibt es wieder Analogien zur Tonerzeugung bei Blechblasinstrumenten. Die Lippenspannung des Bläfers ist (neben dem Drücken der Ventile) zur Einstellung der Tonhöhe wichtig und der Grad der Lippenöffnung im Mundstück entscheidet über die Tonqualität.

Neben der Erzeugung von Sprechschall aufgrund der Phonation (z.B. bei der Produktion von Vokalen) kann Sprechschall auch durch **Friktion** erzeugt werden (z.B. bei der Produktion von stimmlosen **Frikativen** wie [s] oder [f]). Diese Form des Sprachschalls entsteht im Ansatzrohr aufgrund einer starken Engebildung (z.B. [f]: Enge zwischen Oberlippe und unterer Zahnreihe; [s]: Enge zwischen Zungenspitze und Zahndamm). Aufgrund einer solchen Enge wird die durch das Ansatzrohr fließende Luft stark beschleunigt und nach Ausströmen aus der Enge entstehen aufgrund der Luftreibung Luftwirbel, die wiederum hörbare Luftdruckschwankungen verursachen. Ein wesentlicher Unterschied zwischen Phonations- und Friktionsschall ist die Periodizität. Phonationsschall ist periodisch (Abb. 5). Es handelt sich um einen Klang mit definierter Tonhöhe. Friktionsschall hingegen ist ein Rauschen, also ein nichtperiodisches Signal.

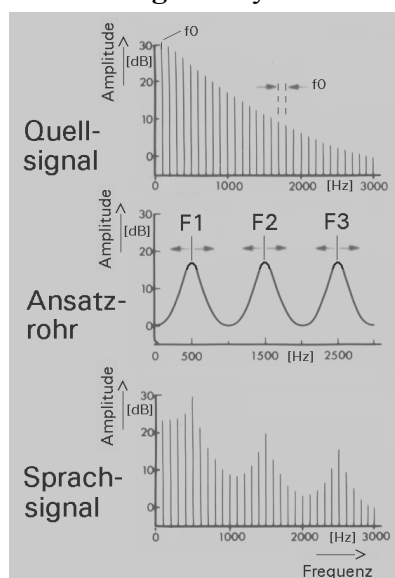
Die Parameter der Initiation und Phonation wie z.B. Tonhöhe, Lautstärke und Stimmqualität (z.B. gepresste, normale, behauchte Stimme) kodieren **prosodische** Information (z.B. Frage kontra Aussage, emotionaler Zustand des Sprechers etc.). Daneben existiert die (oft als primär betrachtete) **segmentale** Information. Diese wird in erster Linie durch die Parameter der Artikulation kodiert. Durch unterschiedliche Einstellungen und Bewegungen der supralaryngalen Artikulationsorgane (Lippen, Zunge, etc.) werden Laute realisiert und damit Bedeutungsunterschiede auf Wortebene signalisiert.

#### *4 Artikulation und Akustogenese*

Das an der Stimmritze erzeugte akustische Signal – also die akustische Konsequenz der Stimmlippenschwingung - wird durch das Ansatzrohr moduliert und von Mund und Nase abgestrahlt (Akustogenese). Die Stimmritze stellt eine **akustische Quelle** und das Ansatzrohr (Rachen-, Mund- und Nasenraum, Abb. 1) ein **akustisches Filter** dar. Die Filterwirkung des Sprechtraktes ist frequenzabhängig. Die spektrale Analyse des Quellsignals ergibt ein Linienspektrum mit kontinuierlich abnehmender Amplitude der Spektrallinien (Abb. 6 oben). Die **Tonhöhe** ( $f_0$ ) des hier dargestellten Klanges kann an der Frequenz des ersten Teiltones (erste Spektrallinie) oder auch am Abstand jeweils zweier benachbarter Spektrallinien abgelesen werden. Die **Klangfarbe** hingegen wird durch die Gesamtform des Spektrums, also durch den

Gesamtverlauf der Amplituden der Spektrallinien determiniert. Die vom Ansatzrohr bewirkte akustische Modulation des Quellsignals ist ein Resultat der Ausbildung von stehenden Schallwellen im Ansatzrohr. Diese stehenden Wellen werden auch Resonanzen des Ansatzrohres oder **Formanten** genannt. Die Frequenz jedes Formanten hängt von der Geometrie des Ansatzrohres, d.h. von der Länge des Ansatzrohres zwischen Stimmritze und Lippen und von der Formung des Ansatzrohres aufgrund von Lippen-, Gaumensegel- und Zungenstellung ab. Eine Resonanz des Ansatzrohres (Formant) bedingt ein akustisches Amplitudenmaximum. Die Amplitudenmaxima der ersten drei Formanten eines Ansatzrohres können anhand der Filtercharakteristik des Ansatzrohres (Abb. 6 mitte) erkannt werden. Die zugehörigen Formantfrequenzen ( $F_1$ ,  $F_2$ ,  $F_3$ ) können ebenfalls anhand dieser Maxima ermittelt werden. Im Unterschied zu der kontinuierlich abnehmenden Amplitude der Spektrallinien des Quellsignals bewirkt die akustische Filterung durch das Ansatzrohr, dass die Amplitudenmaxima und damit die Formanten auch anhand des Spektrums des vom Munde abgestrahlten Schallsignals abgelesen werden können (Abb. 6 unten). Die Modulation (Filterung) des Quellsignals durch das Ansatzrohr ändert somit den Gesamtverlauf der Amplituden der Spektrallinien und damit insbesondere den Klang des akustischen Signals. Die Tonhöhe und damit die Lage der Spektrallinien auf der Frequenzachse wird hingegen vom Ansatzrohr nicht geändert.

**Anmerkung:** Ein systematischer Unterschied in der Klangerzeugung beim Sprechen und Singen gegenüber der Klangerzeugung bei Blasinstrumenten ist die Unabhängigkeit bzw. Abhängigkeit der Tonhöhe von der Länge des Rohres. Während die Länge des Rohres bei Blasinstrumenten die Tonhöhe determiniert, ist dies beim Sprechen und Singen kaum der Fall. Hier wird die Tonhöhe überwiegend vom pulmonalen und laryngalen System determiniert. Grund hierfür ist, dass beim Sprechen und Singen keine starke Kopplung zwischen Stimmlippen und Ansatzrohr besteht, während bei der instrumentalen Tonerzeugung zwischen den Lippen des Bläusers und dem Rohr des Instrumentes starke Kopplungen auftreten. Die Grundtonhöhe  $f_0$  ist beim Sprechen somit unabhängig von den Formantfrequenzen  $F_1$ ,  $F_2$ ,  $F_3$  usw. Wohl determiniert  $f_0$  den Abstand der Teiltöne (Abb. 6 oben und unten). Insbesondere muss einer



**Abbildung 6.** Darstellung des Spektrums des Quellsignals (oben), der Filtercharakteristik des Ansatzrohres (mitte) und des Sprachsignals (unten).

Formantfrequenz auch nicht ein ganzzahliges Vielfaches von  $f_0$  sein und die Formantfrequenzen müssen auch untereinander nicht in einem bestimmten Verhältnis zueinander stehen.

Vielmehr können sie aufgrund der Variabilität der Ansatzrohrform um die hier dargestellten Mittelwerte variieren (Abb. 6 Mitte).

Die Formänderung des Ansatzrohres wird durch Lageänderung der aktiv beweglichen **Artikulatoren** (Artikulationsorgane) – vornehmlich der **Zunge**, der **Lippen** und des **Gaumensegels** (Abb. 1) – bewirkt. Zunge und Lippen sind komplex aufgebaute Muskelsysteme. Die **Zunge** wird in intrinsische und extrinsische Zungenmuskulatur unterteilt. Die intrinsische Zungenmuskulatur baut die Zunge auf und bewirkt hauptsächlich eine Änderung der Zungenform. Beispielsweise kann eine Hebung der **Zungenspitze** zur **konsonantischen Artikulation** realisiert werden (M. longitudinalis). Die extrinsische Zungenmuskulatur verbindet die Zunge mit Kiefer, Schädelknochen und über das Zungenbein auch mit dem Kehlkopf. Sie bewirkt vor allem die Änderung der Zungenlage. Wichtige extrinsische Zungenmuskeln sind z.B. der die Zunge mit dem Kieferknochen verbindende M. genioglossus, und der die Zunge mit dem Schädel verbindende M. styloglossus. Die **Lippen** werden hauptsächlich vom Lippenringmuskel (M. orbicularis oris) aufgebaut, der die Lippenverschlussbildung sowie die Lippenrundung realisiert. Weitere Lippenmuskeln realisieren die Oberlippenhebung, Unterlippenenkung, Lippenstülpung und das Zurückziehen und Anheben der Mundwinkel. Das **Gaumensegel** (Velum) umfasst dem Bereich des **weichen Gaumens** (Abb. 1). Am Ende des Gaumensegels befindet sich das **Zäpfchen** (Uvula). Das Gaumensegel wird durch den M. levator palatini und den M. tensor palatini angehoben und gestrafft. Seine Funktion beim Sprechen ist vor allem die Verbindung bzw. Trennung des Nasenraumes vom Mund- und Rachenraum. Ist das Velum gesenkt (Abb. 1), so ist eine akustische Ankoppelung des Nasenraumes gegeben. Diese ist z.B. bei der Realisierung von **Nasalen** (z.B. [m], [n]) wichtig.

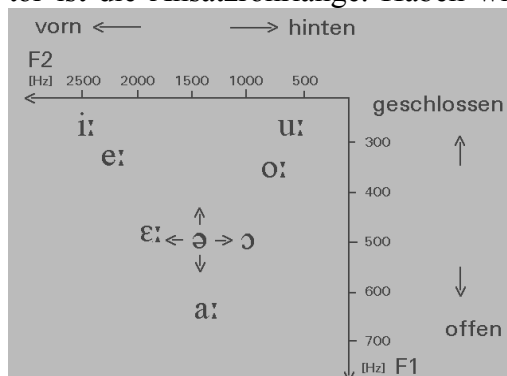
Unterschiedliche Laute werden durch unterschiedliche Bewegungen der supralaryngalen Artikulatoren realisiert. **Konsonanten** werden nach Artikulationsort und Artikulationsart unterschieden (siehe z.B. Pompino-Marschall 1995 oder Grassegger 2001).

Als **Artikulationsorte** werden u.a. **bilabial** (an Unter- und Oberlippe gebildet, z.B. [p, b, m] in „pellen“, „bellen“, „melden“), **labiodental** (an Unterlippe und oberer Zahnreihe gebildet, z.B. [f, v] in „falten“, „walten“), **alveolar** (am Zahndamm gebildet, z.B. [t, d, n, s, z, l], in „Tank“, „Dank“, „Name“, „reißen“, „reisen“, „Lamm“), postalveolar (an Zahndamm bis hartem Gaumen gebildet, z.B. [ʃ, ʒ] in „Schein“, „Genie“), **palatal** (am harten Gaumen gebildet, z.B. [ç, (j), j]) in „ich“, „jagen“), **velar** (am weichen Gaumen gebildet, z.B. [k, x] in „Kohl“, „ach“), **uvular** (am Zäpfchen gebildet, z.B. [R], gerolltes Zäpfchen-r in „Radio“) und **glottal** (an der Stimmritze gebildet, z.B. [ʔ, h] in „Einer“, „Heiner“) unterschieden.

Bezüglich der **Artikulationsart** werden u.a. **Plosivlaute (Plosive)** (vollständiger Verschluss, z.B. [p, b, t, d]), **Frikativlaute (Frikative)** (Engebildung mit Luftverwirbelung und Geräuschbildung, z.B. [f, v, s, z]), **Nasallaute (Nasale)** (gesenktes Velum, z.B. [m, n]), r-Laute (Vibranten) (vibrierender Artikulator, z.B. Zungenspitze [r] oder Zäpfchen [ʀ]) und **Laterallaute (Laterale)** (seitlich abgesenkter Zungenrand, z.B. [l]) unterschieden. Darüber hinaus differenziert die Beteiligung der Stimme das System der Konsonanten in **stimmhafte und stimmlose Laute** (z.B. [b] vs. [p] oder [z] vs. [s]).

Während Konsonanten artikulatorisch durch lokale Engebildung im Ansatzrohr charakterisiert werden, muss bei **Vokalen** die Formung des gesamten Ansatzrohres festgelegt werden. Diese globale Ansatzrohrform wird vor allem durch die **Zungenlage (Höhe und Grad der Vor- bzw. Rückverlagerung)** und die **Lippenformung (Grad der Rundung bzw. Spreizung)** determiniert. Beispielsweise unterscheiden wir im Deutschen ungerundete Vorderzungenvokale (hoch, halbhoch, halbtief und tief: [i:, e:, ε:, a:] in „Miete“, „Meter“, „mähen“, „Made“), gerundete Vorderzungenvokale (hoch und halbhoch: [y:, ø:] in „Mühe“, „mögen“) und gerundete Hinterzungenvokale (hoch und halbhoch: [u:, o:] in „Mut“, „Not“). Eine kompakte Darstellung des Lautsystems des Deutschen gibt Grassegger (2001).

Die Frequenzen der **Formanten** werden durch die **Artikulation** geändert. Ein wichtiger Faktor ist die Ansatzrohrlänge. Haben wir ein Ansatzrohr konstanten Querschnitts (dies ist am



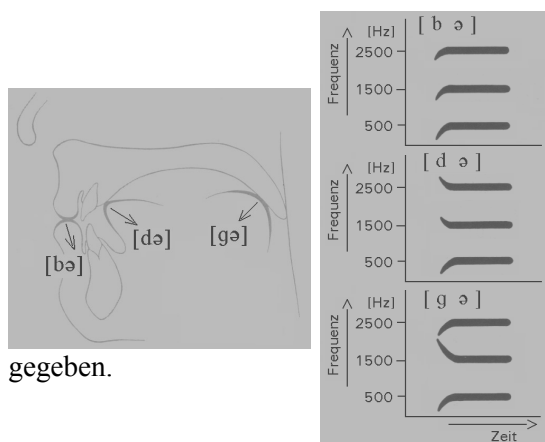
**Abbildung 7.** Zum Zusammenhang zwischen akustischem und artikulatorischem Vokalraum. Durchschnittsfrequenzen des ersten und zweiten Formanten von acht Vokalen des Deutschen und zugehörige artikulatorische Kategorien sind gekennzeichnet (Neben den im Text beschriebenen Langvokalen wurde hier noch der Kurzvokal [ɔ] („Hocke“) und der Schwa-laut eingezeichnet.

ehesten beim „Schwa-Laut“ oder „Murmelvokal“ [ə] erfüllt), so ergeben sich die Formantfrequenzen bei einer typischen Ansatzrohrlänge von 17,5 cm in regelmäßigen Abständen von 1000 Hz beginnend mit dem ersten Formanten bei 500 Hz (Abb. 6 Mitte).

Durch Änderungen der Lippenform und Zungenlage können die Formanten nun auf der Frequenzachse verschoben werden (Pfeile in Abb. 6 Mitte). Die Abhängigkeit der Formantfrequenzen von der Geometrie des Ansatzrohres ist sehr komplex. Es können aber zwei Faustregeln angegeben werden: Zum einen bewirkt die zunehmende Mundöffnung (Kie-fersenkung und auch Zungensenkung) eine Absenkung der Frequenz des ersten Formanten. So zeichnet den Vokal [a:] eine tiefe, die Vokale [i:] und [u:] hingegen eine hohe Frequenz des ersten Formanten aus (Abb. 7, vertikale Pfeile).

Zum anderen bewirkt die Vorverlagerung des Zungenrückens von hinten nach vorne eine Anhebung und die Rückverlagerung von vorne nach hinten eine Absenkung der Frequenz des zweiten Formanten. So zeichnet sich der Vokal [i:] durch eine hohe, der Vokale [u:] hingegen durch eine niedrige Frequenz des zweiten Formanten aus (Abb. 7, horizontale Pfeile).

Ähnliche Gesetzmäßigkeiten können auch für die **lokale Engebildungen** und damit für die Produktion von Konsonanten aufgestellt werden. Durch eine lokale Enge- bzw. Verschlussbildung kommt es ebenfalls zu Änderungen (Verschiebungen) von Formantfrequenzen. Da im zeitlichen Ablauf der Artikulation in allen Übergangsbereichen von Vokalen zu Konsonanten und von Konsonanten zu Vokalen eine Zunahme bzw. Abnahme des Grades der Engebildung vorliegt, entsteht eine kontinuierliche Änderungen der Formantfrequenzen: Es treten **Formanttransitionen** auf (Abb. 8 rechts). Die Formanttransitionen in Abb. 8 stellen das akustische Abbild der **Artikulatorbewegung** zur Lösung einer artikulatorischen Enge von Konsonant zu Vokal dar (Abb. 8 links), nämlich der Öffnungsbewegung der Lippen bei [b], der Senkung der Zungenspitze bei [d] und des Zungenrückens bei [g]. Es zeigt sich, dass je nach Lage der artikulatorischen Enge (Artikulationsort labial, alveolar oder palatal/velar) Änderungen der Formantfrequenzen in unterschiedliche Richtung erfolgen. So ändert sich der erste Formant zwar immer gleichförmig (Anstieg mit abnehmender Engebildung) der zweite Formant zeigt aber beispielsweise bei der Rückbildung der labialen Enge einen Anstieg, ansonsten aber eine Absenkung. Der dritte Formant zeigt nur bei einer alveolaren Enge eine Absenkung und eine Anhebung sonst (Abb. 8 rechts). Somit können insbesondere diese drei konsonantischen **Artikulationsorte** allein über die **Richtung der Formanttransitionen** der ersten drei Formanten unterschieden (kodiert) werden.



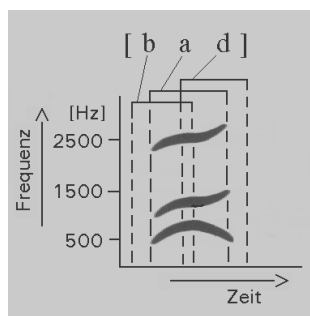
**Abbildung 8.** Die Formanttransitionen beim Übergang vom labialen, apikalen bzw. velaren Verschluss zum Murrelvokal. Die Pfeile im linken Bildteil deuten die artikulatorische Öffnungsbewegung beim Übergang vom Plosiv in den Vokal an. Die zugehörigen Formanttransitionen sind im rechten Bildteil angegeben.

**Anmerkung:** Setzen wir die Konsonanten [b, d, g] nun aber in den Kontext anderer Vokale (oben ist der Kontext immer der Schwa-Laut), so können sich durchaus auch andere Richtun-

gen für die Formanttransitionen ergeben. Dies führt zum Problem der fehlenden Invarianz akustischer Lautmerkmale und wird im nächsten Absatz diskutiert.

### 5 Zur phonetischen Realisierung ganzer Äußerungen

Die Artikulationsorgane sind während der Lautproduktion in ständiger **Bewegung**. Statische Artikulatorpositionen treten während der Produktion von fließender Rede (Wörter oder Sätze) weder in konsonantischen noch in vokalischen Bereichen der Äußerung auf. Betrachten wir die Frequenzverläufe der Formanten einer Silbe (Abb. 9), so sehen wir, dass sich die Formantfrequenzen als Folge von kontinuierlichen Artikulatorbewegungen und damit als Folge einer ständigen Wandlung der Ansatzrohrform kontinuierlich ändern. Darüber hinaus ist die artikulatorische, aber auch die akustische Realisierung von Lauten stark **kontextabhängig**. So kann insbesondere die artikulatorische und akustische Gestalt von Konsonanten stark von den umliegenden Vokalen abhängen. Aber auch die prosodischen Eigenschaften wie Sprechtempo oder Betonungsgrad beeinflussen Lautrealisierungen.



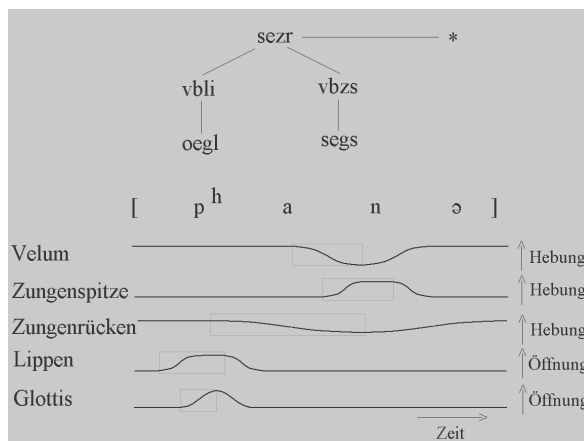
**Abbildung 9.** Verläufe des ersten, zweiten und dritten Formanten der ersten drei Laute des Wortes „baden“. Oben eingezeichnet: Zeitbereiche der akustischen Merkmale der Laute.

Somit ist das (durch die Lautschrift evtl. naheliegendes) Modell einer nicht überlappenden Aneinanderreihung diskreter und invarianter Einheiten in der akustischen wie auch in der artikulatorischen Domäne nicht haltbar. Neben der gegenseitigen Beeinflussung haben wir auf der artikulatorischen aber auch auf der akustischen Ebene ein **zeitliches Überlappen von Lautmerkmalen**. So kodiert beispielsweise eine Formanttransition den Artikulationsort eines Plosives. Andererseits gehört in einer Plosiv-Vokal-Plosiv-Verbindung der Bereich der Phonation zwischen den konsonantischen Verschlussphasen (i.e. der Bereich des Verlaufes der Formanten, Abb. 9) auch zum Vokal.

Eine Antwort auf die Frage nach den invarianten Einheiten der Sprachproduktion gibt die Theorie der artikulatorischen **Geste** (auch „Artikulatorische Phonologie“ nach Browman und Goldstein 1989, 1990 und 1992). Die invarianten Einheiten gesprochener Sprache sind aus der Sicht der Sprachproduktion nicht statische Artikulatorpositionierungen sondern charakteristische zielgerichtete **Artikulatorbewegungen**. Betrachten wir beispielsweise die Realisie-

nung des Wortes „Panne“, so geschieht dies durch die zeitliche Koordinierung mehrerer Gesten (siehe Abb. 10). Die Realisierung des [p] geschieht durch eine zeitliche Überlappung einer Geste zur Verschlussbildung der Lippen (vbli, siehe auch Tab. 1) mit einer Geste zur Öffnung der Glottis (oegl). Die Realisierung des Vokals geschieht durch die Überlappung einer Geste zur Zungenrückensenkung (sezr) usw. Zusätzlich ist festgelegt, dass im Falle des Fehlens von gestischer Aktivität die Zunge die Stellung des Murrevokals [ə] realisiert, also das Gaumensegel gehoben ist und die Stimmlippen in Phonationsstellung sind. So wird im Falle von Vokalen auf die explizite Bezeichnung einer glottalen Geste und im Falle von nicht nasalierten Lauten auf die Bezeichnung einer Geste des Gaumensegels verzichtet. Zur Realisierung des gesamten Wortes ist die geschickte zeitliche Koordinierung all dieser Gesten erforderlich. Dabei kommt es immer auch zur zeitlichen Überlappung von Gesten unterschiedlicher Laute.

**Anmerkung:** Der Schwa-Laut zeichnet sich durch das Fehlen von zur bezeichnenden Gesten aus. Hier wird lediglich zur Wahrung der richtigen zeitlichen Struktur der Äußerung auf der phonologischen Ebene ein Platzhalter (\*) gesetzt.



**Abbildung 10.** Die zur Realisierung des Wortes „Panne“ benötigten Gesten. Phonologische Ebene (oben), phonetische Transkription (Mitte) und Ebene der artikulatorischen Realisierung (unten). Die Rechtecke kennzeichnen die zeitliche Ausdehnung (Dauer) einer Geste. Die Linienverläufe kennzeichnen die Bewegung einzelner Artikulatoren. Achtung: Abb. 10: Ersetze Lippen : Öffnung durch Lippen: Verschluss (oder kehre die Richtung des Pfeiles um)

Abbildung 10 zeigt zwei Arten der gestischen Spezifikation des Wortes „Panne“. Auf der **phonologischen Ebene** können Gesten als abstrakte Einheiten, die in bestimmter Relation zueinander stehen, dargestellt werden. Gesten werden ähnlich wie in autosegmental-phonologischen Konzepten auf drei Ebenen angeordnet (vokalische Gesten, konsonantische Gesten, velische und glottale Öffnungsgesten). Die **Assoziationslinien** verbinden Gesten, die miteinander zeitlich koordiniert werden. Auf der **Ebene der artikulatorischen Realisierung** sind die Gesten dann konkret ausführbare Ereignisse (Aktionen), eben definierte Artikulatorbewegungen, die mit definierter Dauer versehen auf definierte Artikulatoren wirken.

Die Geste tritt somit in diesem Konzept sowohl als **abstrakte linguistisch-phonologische Größe** als auch als **konkrete Größe der phonetisch-artikulatorischen Realisierung** auf. Auf der phonologischen Ebene kann ein System diskreter vierbuchstabiger Gestensymbole realisiert werden (Tab. 1). Die symbolische Darstellung grenzt unterschiedliche Gesten in diskreter Weise voneinander ab und führt in ähnlicher Weise wie die bekannten segmentalphonologischen Systeme zu einem bedeutungsunterscheidenden (distinktiven) System.

	Geste zur....	z.B. zur Realisierung von ...
<b>hhzr</b>	hohen Hebung des Zungenrückens	/i:/, /u:/
<b>hezr</b>	Hebung des Zungenrückens	/e:/, /o:/
<b>sezr</b>	Senkung des Zungenrückens	/a:/
<b>rvzr</b>	Rückverlagerung des Zungenrückens	/u:/, /o:/
<b>vvzr</b>	Vorverlagerung des Zungenrückens	/i:/, /e:/
<b>ruli</b>	Rundung der Lippen	/u:/, /y:/
<b>vbli</b>	Verschlussbildung durch die Lippen	/b/, /p/, /m/
<b>vbzs</b>	Verschlussbildung durch die Zungenspitze	/d/, /t/, /n/
<b>vbzr</b>	Verschlussbildung durch den Zungenrücken	/g/, /k/
<b>eazs</b>	alveolaren Engebildung durch die Zungenspitze	/s/, /z/
<b>epzs</b>	postalveolaren Engebildung durch die Zungenspitze	/ʃ/, /ʒ/
<b>oegl</b>	Geste zur Öffnung der Glottis	stimmlosen Lauten
<b>segs</b>	Geste zur Senkung des Gaumensegels	Nasalen

**Tabelle 1.** Ausschnitt aus dem phonologischen System der Gesten des Deutschen

Gleichzeitig können dieser symbolischen Darstellung einer Geste aber auch konkrete Werte z.B. über Dauer und räumlich-zeitlichen Verlauf der durch die Geste realisierten Artikulatorbewegungen zugeordnet werden. Somit löst dieser Ansatz auch die **Problematik der phonetisch-phonologischen Schnittstelle** in eleganter Weise: Die Geste ist sowohl diskrete linguistische Einheit als auch konkrete quantifizierbare artikulatorische Größe. Aufgrund der Quantifizierbarkeit konnte dieses Konzept auch bereits erfolgreich zur Steuerung von Sprachsynthesatoren eingesetzt werden (z.B. Kröger 1998). Auch löst dieser Ansatz das **Problem der fehlenden Invarianz von akustischen Lautmerkmalen**: Perzeptive Untersuchungen ergaben, dass z.B. im Fall stimmhafter Plosive vor allem die Formant*transitionen* das akustische Merkmal des Artikulationsortes sind. Andererseits ist bekannt, dass beispielsweise die Transition des zweiten Formanten für den Artikulationsort „alveolar“ je nach Vokalkontext sehr unterschiedlich sein kann (Öhmann 1966). Es ist also eine fehlende Invarianz zwischen der phonetisch-phonologischen Kategorie und dem akustischen Merkmal zu beklagen. Erst in der artikulatorischen Domäne tritt das invariante Lautmerkmal zu Tage, nämlich eben der artikula-

torische Zielpunkt der Geste zur Verschlussbildung durch die Zungenspitze (vbzs). Diese Resultate führten zu der Annahme einer engen Kopplung zwischen Artikulation und Wahrnehmung: **Artikulatorische Gesten sind die Einheiten auf die unser Wahrnehmungsmechanismus trainiert ist.** (Lieberman und Mattingly 1985, Mattingly 1990).

Ein weiteres Problem segmental-phonologischer Konzepte ist die Beschreibung von **reduzierten Formen**: Es ist bekannt, dass bei schneller Sprechweise eine Vielzahl von Assimilations- und Elisionseffekten auftreten können (Kohler 1990). Beispielsweise kann die Äußerung „Er hat mir geholfen“ bei langsamer Sprechweise als [ʔɛʰat<sup>h</sup>miɐ̯gəʰɔlfən] und bei schneller Sprechweise als [ɛʰbmɐ̯gəʰɔlfm] realisiert werden. Wir sehen hier beispielsweise einen **Assimilationsprozess** des Artikulationsortes wie auch der Stimmhaftigkeit den Übergang von [t] in „hat“ nach [b] vor nachfolgendem [m]. Darüber hinaus können auch **Elisionen** (Wegfall von Segmenten) erkannt werden, da sich z.B. das Wort „hat“ auf ein Segment und das Wort „mir“ auf zwei Segmente reduziert. Somit ist also gerade das **Segment** - als Realisierung der zugrundeliegenden Einheit Phonem - extremen Änderungen aufgrund des (nicht-phonologischen!) Faktors Sprechtempo ausgesetzt. Die segmentale Darstellung ist somit nicht invariant gegenüber Änderungen des Sprechtempos. Im gestischen Ansatz kann hingegen gezeigt werden, dass bei Sprechtempoänderungen die zugrundeliegende **gestische Struktur** erhalten bleibt (Browman und Goldstein 1990, Kröger 1998). **Hier werden die Reduktionen allein durch Änderung der zeitlichen Koordinierung und Änderung der zeitlichen Ausdehnung von Gesten erreicht.**

Neben der Organisation der zeitlichen Abläufe zur Lautbildung tritt die Produktion von **prosodischen Merkmalen** wie **Akzentuierung (Betonung)** und der **Intonation**. Aus phonetischer Sicht soll unter Akzent (Betonung) die lokale Hervorhebung einer Silbe gegenüber benachbarten Silben innerhalb einer Äußerung verstanden werden. Diese Hervorhebung erfolgt in den meisten Sprachen in erster Linie über markante Änderungen der Grundfrequenz und Erhöhung der Silbendauer und damit Änderung der Dauer der in der Silbe auftretenden Laute. Die Intonation einer Äußerung manifestiert sich phonetisch in erster Linie im Gesamtverlauf der Grundfrequenz ( $f_0$ ). In vielen Sprachen ist beispielsweise der Anstieg der Grundfrequenz am Äußerungsende ein wichtiges Merkmal der Frageintonation, ein Abfallen der Grundfrequenz hingegen signalisiert das Ende einer Aussage. Hier sind auch bereits die organisatorischen Einheiten der Sprachproduktion oberhalb des Lautes, nämlich **Silbe** und **Äußerung** genannt. Man kann zusätzlich noch die Einheit der **rhythmischen Gruppe** annehmen, die jeweils von einer betonten Silbe bis zur nächsten reicht.

Die Domäne der Intonation ist die Äußerung. Eine Äußerung kann aus einer oder mehreren rhythmischen Gruppen aufgebaut sein. Eine rhythmische Gruppe wiederum besteht aus einer oder mehreren Silben. Die Domäne der Akzentuierung ist die Silbe. Eine Silbe wiederum ist aus mehreren Lauten aufgebaut. Es ist anzumerken, dass sich Intonation und Akzentuierung gegenseitig beeinflussen und bedingen. Darüber hinaus gibt es auch Wechselwirkungen mit der Lautbildung. So konnte beispielsweise gezeigt werden, dass die Artikulation von Lauten in betonten Silben mit höherer Präzision und mit höherem artikulatorischem Aufwand verbunden ist. Die Artikulatoren nähern sich in betonten Silben den (fiktiven) artikulatorischen Zielpunkten der Laute stärker als in unbetonten Silben. Somit sollte eine Theorie der Lautproduktion die prosodischen Aspekte gesprochener Sprache nicht unbeachtet lassen (Kröger 1998, S. 223ff).

### *Literatur*

- Browman, C.P., Goldstein, L. (1989): "Articulatory gestures as phonological units", *Phonology* **6**, 201-251.
- Browman, C.P., Goldstein, L. (1990): "Tiers in articulatory phonology, with some implications for casual speech", in: J. Kingston, M.E. Beckman (Hrsg.), *Papers in Laboratory Phonology I: Between the Grammar and Physics of Speech* (Cambridge University Press, Cambridge), S. 341-376. Auch in: *Haskins Laboratories Status Report on Speech Research* **SR-92** (1987), 1-30.
- Browman, C.P., Goldstein, L. (1992): "Articulatory phonology: An overview", *Phonetica* **49**, 155-180.
- Catford, J.C. (1988): *A Practical Introduction to Phonetics* (Oxford University Press, New York).
- Draper, M.H., Ladefoged, P., Whitteridge, D. (1959): "Respiratory muscles in speech", *Journal of Speech and Hearing Research* **2**, 16-27.
- Grassegger, H. (2001): *Phonetik – Phonologie*. (Schulz-Kirchner Verlag, Idstein)
- Handbook of the International Phonetic Association (1999). Cambridge: Cambridge University Press.
- Kohler, K.J. (1990): "Segmental reduction in connected speech in German: Phonological facts and phonetic explanations", in: W.J. Hardcastle, A. Marchal (Hrsg.), *Speech Production and Speech Modelling* (Kluwer Academic Press, Dordrecht), S. 69-92.
- Kröger, B.J. (1998): *Ein phonetisches Modell der Sprachproduktion* (Niemeyer, Tübingen).
- Lieberman, A.M., Mattingly, I.G. (1985): "The motor theory of speech perception revised", *Cognition* **21**, 1-36.
- Mattingly, I.G. (1990): "The global character of phonetic gestures", *Journal of Phonetics* **18**, 445-452.
- Öhman, S.E.G. (1966): "Coarticulation in VCV utterances", *Journal of the Acoustical Society of America* **39**, 151-168.
- Ohala, J.J. (1990): "Respiratory activity in speech", in: W.J. Hardcastle, A. Marchal (Hrsg.), *Speech Production and Speech Modelling* (Kluwer Academic Press, Dordrecht), S. 23-54.
- Pompino-Marschall, B. (1995) *Phonetik und Phonologie*. (DeGruyter, Berlin)