

- In diesem Kapitel werden semantische Netzwerke, das mentale Lexikon, der mentale Silbenspeicher sowie die Grundlagen der Artikulation und der Generierung des akustischen Sprachsignals erläutert. Es werden alle lexikalischen Repräsentationsformen (Konzept, Lemma und phonologische Form) sowie alle Repräsentationsformen im Silbenspeicher (motorische Form oder Motorplan, auditive Form, somatosensorische Form) und der Aufbau von Motorplänen aus Sprechbewegungseinheiten erläutert. Danach wird erklärt, wie Motorpläne artikulatorische Abläufe für Lippen, Zunge, Gaumensegel und den Kehlkopf bzw. die Aryknorpel aktivieren können und wie im Sprechapparat das akustische Sprachsignal entsteht. Am Ende des Kapitels werden produktionsbedingte Sprach- und Sprechstörungen erläutert.

2.1 Wörter, Silben und Sprachlaute

Die grundlegenden Einheiten gesprochener Sprache sind Wörter, Silben Sprachlaute. Beginnen wir zunächst mit den Wörtern, die aber erst im Kontext eines Satzes komplexe Bedeutungen transportieren können.

2.1.1 Konzepte und semantische Netzwerke

Wollen wir eine sprachliche Äußerung realisieren, so ist der Ausgangspunkt hierzu eine kommunikative Intention, bzw. ein spezifischer Inhalt, der zunächst auf der semantischen Ebene bzw. im **semantischen neuronalen Netzwerk** aktiviert wird. Inwieweit diese Ebene bzw. dieses Netzwerk rein metasprachlicher Natur ist oder inwieweit diese Ebene bzw. dieses Netzwerk bereits sprachlich determiniert ist, soll im Rahmen dieses Buches

nicht diskutiert werden. Wir gehen hier davon aus, dass zumindest Teile dieser Ebene bzw. dieses Netzwerkes eng mit dem erlernten Sprachsystem interagieren. Dieses semantisch-linguistische Netzwerk funktioniert wie folgt: Möchte ein Sprecher beispielsweise den Satz „Der Hund jagt die Katze“ produzieren, so wird zunächst das **Konzept**, also die Bedeutung <Hund>, <jagen> und <Katze>, aktiviert. Dies weist bereits auf die Konzeptebene des mentalen Lexikons und aktiviert die zugehörigen Lexeme. Zusätzlich muss auf semantischer Ebene, also auf der Bedeutungsebene, zur Festlegung der richtigen Satzbedeutung definiert sein, wie die Beziehung zwischen den Nomen „Hund“ und „Katze“ und dem Verb „jagen“ aussieht:

$$\text{Subjekt : } \langle \text{Hund} \rangle + \text{Objekt : } \langle \text{Katze} \rangle + \text{Verb : } \langle \text{jagen} \rangle \quad (2.1)$$

Mittels grammatikalischem und syntaktischem Wissen wird dann das Verb in Relation zum Subjekt auf 3. Person Singular, also auf die Form „jagt“, gebeugt, und es werden noch die Funktionswörter „der“ und „die“ hinzugefügt, bevor der gesamte Satz dann in seiner phonologischen Form /dɛr hʊnt ja:kt di: kat tsə/ aktiviert und silbifiziert werden kann. Die phonetischen Schriftzeichen werden in [Abschn. 2.3.1](#) erläutert. Im Silbenspeicher werden dann die den einzelnen Silben zugehörigen Motorpläne aktiviert und als Sequenz artikulatorisch realisiert.

Um zunächst die Rolle des metasprachlichen Teils des semantischen Netzwerkes noch genauer zu beleuchten, nehmen wir im Folgenden ein einfacheres Beispiel, nämlich das einer Wortproduktion durch ein Kleinkind, das noch im Begriff ist, seine Muttersprache zu erlernen. Im Fall der Intention eines Kleinkindes, einen Kommunikationspartner durch die Äußerung des Wortes „Ball“ auf einen runden, roten, in einer Ecke des Raumes liegenden Gegenstand aufmerksam zu machen, kann dieser Prozess wie folgt ablaufen: Zunächst sieht das Kind den in der Ecke eines Raumes liegenden Ball und aktiviert in seinem Gehirn neben den visuell ausgelösten *metasprachlichen semantischen Merkmalen* <rund>, <rot> noch weitere aus seiner Erfahrung (seinem Vorwissen) herrührende semantische Merkmalskomplexe, z. B. <rollt>, <spielen>, <werfen>. Im metasprachlichen semantischen Netzwerk des Kindes werden all diese **semantischen Merkmale** bzw. *Konzepte* zusammen mit dem Konzept <Ball> aktiviert.

Aufgrund seiner durch Kommunikation erworbenen Erfahrung weiß das Kind, auch wenn es noch nicht schreiben kann, dass dieses Konzept <Ball> auch direkt durch das Lexem „Ball“ sprachlich benannt werden kann. Nicht immer ist eine Eins-zu-eins-Zuordnung zwischen Konzepten und Lexemen so einfach realisierbar. So kann das Konzept <Fahrrad> kontextuell bedingt zur Aktivierung unterschiedlicher Lexeme wie z. B. „Rad“, „Bike“ oder auch „Fahrrad“ führen. Kehren wir zurück zum Beispiel des Kindes, welches das Konzept <Ball> als Wort realisieren möchte. Falls es dieses Wort noch nicht fehlerfrei realisieren kann, wird es evtl. /ba/ sagen. Ab einem gewissen Alter wird aber die zum Lexem gehörende korrekte Realisierung ausgesprochen. Somit führt ein im semantischen Netzwerk aktiviertes Konzept zur Aktivierung und damit zum Abruf des zugehörigen Lexems aus dem mentalen Lexikon. Die mit dem Lexem definierte Lautfolge des Wortes (z. B. / bal/) kann dann geäußert werden. Der Prozess des Aktivierens eines bestimmten

Lexems aufgrund einer vorliegenden Voraktivierung eines Konzeptes im semantischen Netzwerk besteht somit aus der Aktivierung des Eintrags bzw. Items im mentalen Lexikon, welcher dann zur Produktion ausgewählt und dessen Lautfolge anschließend aus dem mentalen Lexikon abgerufen wird.

Einzelne semantische Merkmale (z. B. <beweglich>, <rund>, <Spielzeug>, <werfen>, <fangen>) wie auch daraus entstehende Komplexe von semantischen Merkmalen wie z. B. <Ball> bilden in ihrer Gesamtheit auf der metasprachlichen Ebene ein *semantisches Netzwerk*, in dem dann das gesamte Weltwissen des Sprechers als Vernetzung von Konzepten realisiert ist. Ein Konzept ist aber immer über *semantische Relationen* mit anderen Konzepten verbunden. So wird das Konzept <Ball> erst verständlich, wenn die oben genannten weiteren Konzepte wie <beweglich>, <rund> etc. durch semantische Relationen wie z. B. [... ist ...], [... ist ein ...], [... kann man ...] miteinander verbunden werden:

$$\langle \text{Ball} \rangle [\text{ist}] \langle \text{beweglich} \rangle \quad (2.2a)$$

$$\langle \text{Ball} \rangle [\text{ist}] \langle \text{rund} \rangle \quad (2.2b)$$

$$\langle \text{Ball} \rangle [\text{ist ein}] \langle \text{Spielzeug} \rangle \quad (2.2c)$$

$$\langle \text{Ball} \rangle [\text{ist ein}] \langle \text{Objekt} \rangle \quad (2.2d)$$

$$\langle \text{Ball} \rangle [\text{kann man}] \langle \text{werfen} \rangle \quad (2.2e)$$

$$\langle \text{Ball} \rangle [\text{kann man}] \langle \text{fangen} \rangle \quad (2.2f)$$

Diese Relationen führen von einem Konzept zu einem anderen und sind somit immer gerichtet. [Abb. 2.1](#) stellt einen kleinen Ausschnitt eines semantischen Netzwerkes dar. Hier repräsentieren die von Ovalen umgebenen Ausdrücke Konzepte (z. B. <Gegenstand>, <Hund> ...), während die Pfeile gerichtete Konzeptrelationen wie [ist ein] oder [benötigt] darstellen. Die Art der Relationen wird in [Abb. 2.1](#) durch die am Pfeil angegebenen Ausdrücke definiert. Implizite Relationen zwischen Konzepten, wie z. B. <Hund>[benötigt]<Wasser> müssen in semantischen Netzwerken nicht unbedingt explizit vermerkt werden (Weiterführendes zu semantischen Netzwerken: Steyvers und Tenenbaum 2005 und [Abschn. 7.4](#)).

2.1.2 Mentales Lexikon und mentaler Silbenspeicher

Das metasprachliche semantische Netzwerk kann als Teil des vom Menschen erworbenen Weltwissens ([Abb. 2.1](#)) angesehen werden. Ein *Wort* hingegen ist dadurch definiert, dass ein Konzept mit einer definiert Lautfolge in direkter Verbindung steht. Ein Wort bzw. ein Lexem verbindet also eine bestimmte Bedeutung mit einer definierten Lautfolge und ist

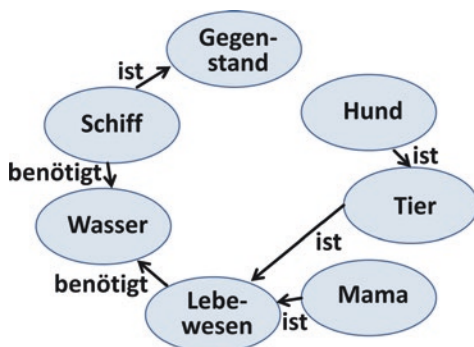


Abb. 2.1 Kleines semantisches Netzwerk

Bestandteil des mentalen Lexikons. Das **mentale Lexikon** ist damit ein Wissensspeicher, in dem beim Erwachsenen Menschen ca. 60.000 Wörter gespeichert sind. Das mentale Lexikon verbindet demnach Konzepte als Teile des semantischen Netzwerkes mit lautlicher Struktur (vgl. Levelt et al. 1999).

Wie aber erlernt das Kind ein Wort? Einem Kind, welches ein bestimmtes Wort (z. B. „Ball“) sprechen kann, muss nicht unbedingt klar sein, dass dieses Wort aus der Folge von drei Lauten, nämlich /ba/ besteht. Oder anders formuliert: die **phonologische Bewusstheit**, d. h. das Wissen um den Aufbau eines Wortes aus unterschiedlichen Sprachlauten (phonologische Form des Wortes), entsteht erst nach und nach. Stellen wir uns aber ein Szenario vor, welches das Kleinkind initiieren könnte, um das Wort „Ball“ aussprechen zu lernen. Wenn ein Kleinkind seinen Kommunikationspartner beispielsweise auf einen im Raum liegenden Ball aufmerksam macht, weil es gerade an diesem Gegenstand interessiert ist, und wenn dieses Kleinkind darüber hinaus vielleicht gerade auch Lust auf Kommunikation hat und sich fragt, ob es durch Aussprache eines Wortes direkt verstanden wird, und wenn ihm dann der Ball vielleicht auch gereicht wird, so wird das Kleinkind vielleicht den Laut /a/, die Lautfolge /ba/ oder Ähnliches äußern und evtl. auch in Richtung des Balles zeigen. Wird es von der Bezugsperson nicht verstanden, d. h., wird die Bezugsperson nicht auf den Ball aufmerksam, so ist denkbar, dass das Kleinkind fortwährend Lautfolgen produzieren wird, die das Wort Ball repräsentieren sollen. Vielleicht kann sich das Kind dazu auch in Erinnerung rufen, wie das Wort Ball klingt, weil es dieses Wort zuvor bereits einmal von der Bezugsperson gehört hat. Irgendwann wird die Bezugsperson auf den Ball aufmerksam und dem Kleinkind das Wort auch noch einmal vorsagen, und das Kind wird sich die korrekte **phonologische Form** merken. Vielleicht könnte es in Tage später stattfindenden Kommunikationsszenarien wieder um den Ball gehen, und irgendwann wird das Kind in der Lage sein, eine für die Bezugsperson befriedigende Aussprache des Wortes Ball zu produzieren. Es ist wahrscheinlich, dass das Kind dann für die erste gelungene Produktion des Wortes besonders gelobt wird, sodass sich ihm diese Lautfolge und die zugehörige Artikulation als Realisierung des Wortes besonders einprägt.

Das hier beschriebene Lernszenario umfasst somit immer eine Interaktion mit einem Kommunikationspartner (z. B. der Mutter). Stellen wir uns vor, das Kleinkind sitzt auf dem

Arm der Mutter und schaut herum und sieht einen Ball im Zimmer herumliegen. Hat das Kind nun Lust, das Wort Ball auszusprechen, aber noch nicht sicher gelernt, wie dieses Wort auszusprechen ist, könnte es mit irgendeiner sprachlichen Äußerung (z. B. /a:/) beginnen. Es ist wahrscheinlich, dass die Mutter die Äußerung nicht versteht. Das Kind könnte dann zum Ball zeigen und danach direkt die Mutter angucken, ohne etwas zu sagen. Ein solches Szenario kann von der Mutter so interpretiert werden, dass das Kind wissen möchte, wie dieser Gegenstand heißt, und die Mutter wird dem Kind sagen: „Ball!“ Damit hat das Kind eine *auditive Vorlage* für dieses Wort, die es imitieren kann. Das Kind wird nun evtl. versuchen, das Wort „Ball“ selbst auszusprechen (das von der Mutter gesprochene Wort zu imitieren). Und es wird evtl. über Tage und Wochen so viele Versuche unternehmen, bis es das Wort Ball verständlich aussprechen kann.

Das alles führt aber nicht nur zum Erlernen und Speichern der **motorischen Form** (der Artikulation) dieses Wortes, sondern gleichzeitig auch zum Erlernen und Speichern der **auditiven Form** (des Gehörseindrucks, des Klangs) dieses Wortes. Die Speicherung der auditiven Form für jedes Wort ist wichtig, damit der Mensch in seinem weiteren Leben kontrollieren kann, ob er das Wort immer richtig ausspricht. Da der Mensch natürlicherweise dazu neigt, seine Bewegungsabläufe (und damit auch artikulatorische Bewegungsabläufe des Sprechens) im Laufe der Zeit zu ökonomisieren und damit zu vereinfachen, muss er auf der Hut sein, dass er nicht zu stark vereinfacht und damit seine Aussprache unverständlich wird.

Zusätzlich zur auditiven Form wird über den Zeitverlauf des Erlernens der motorischen Form einer Silbe bzw. eines Wortes auch noch eine weitere sensorische Form gespeichert, nämlich die **somatosensorische Form**. Die Somatosensorik beschreibt die **taktile Wahrnehmung** und die **propriozeptive Wahrnehmung**, also die Wahrnehmung der Gelenkstellung (z. B. Unterkiefer) und die Wahrnehmung der Muskelspannung und dadurch z. B. die Wahrnehmung der Lage der Zunge während des Sprechens. Der Sprechende kann über die schnelle somatosensorische Rückkopplung bereits erkennen, ob der Artikulationsablauf „normal“ war, d. h. ob sich die Produktion der Silbe „wie sonst auch anfühlte“, bevor die langsame auditive Rückmeldung stattfindet. Es wird im nächsten Abschnitt noch näher auf die Wichtigkeit der somatosensorischen Rückkopplung bei der Sprachproduktion eingegangen.

Konzepte sind abstrakte symbolische Entitäten, und das gesamte semantische Netzwerk kann daher im Gehirn sehr effektiv (d. h. „platzsparend“ bzw. nur wenige Neuronen in Anspruch nehmend) gespeichert werden. Es ist für das Gehirn kein Problem, 60.000 Konzepte und deren Relationen zueinander dauerhaft auf ein paar Quadratmillimeter kortikaler Oberfläche zu speichern. Hier handelt es sich um *kognitive Entitäten*. Anders sieht dies aber bei der Speicherung von motorischen Abläufen (z. B. motorische Form eines Wortes) und sensorischen Eindrücken (z. B. auditive Form eines Wortes) aus. Diese Daten sind nicht abstrakt kognitiver, sondern motorischer und sensorischer Natur. Da es allerdings so ist, dass Wörter bzw. Äußerungen in Silben zerlegt werden können und ca. 95 % dessen, was wir täglich sprechen, im Deutschen auf ca. 2000 unterschiedliche Silben zurückgeführt werden kann, ist auch dies wiederum kein Problem. Die motorischen Pläne und die

zugehörigen sensorischen Eindrücke von ca. 2000 unterschiedlichen Silben (Silben sind ca. 250 ms lang) können ebenfalls auf ein paar Quadratmillimeter kortikaler Oberfläche gespeichert werden. Die Verbindung der Lautfolge jeder Silbe mit ihrem Motorplan und der auditiven Form wird als *mentaler Silbenspeicher* bezeichnet (vgl. Levelt et al. 1999; Cholin 2008). Der Übergang von der symbolisch kognitiven Ebene der Sprachverarbeitung (mentales Lexikon plus Grammatikregeln) zur motorischen und sensorischen Ebene gesprochener Sprache (*sensorisch-motorische Ebene*, mentaler Silbenspeicher) wird auf der Seite der Produktion über die *Silbifizierung* vollzogen (Abb. 1.1).

Abschließend sei noch angemerkt, dass im weiteren Verlauf des Spracherwerbs neben dem Konzept und der phonologischen Form eines Wortes zu jedem Lexem auch noch sein grammatikalischer Status (Substantiv, Verb, ..., männlich, weiblich, Singular, Plural, ...) vermerkt werden muss, um das Wort im Satz richtig flektieren zu können und um evtl. andere Wörter (z. B. ein Verb) an den Status (z. B. Singular oder Plural) eines Substantives anzupassen, wenn dieses Substantiv im Satz als Subjekt steht. In einigen Theorien wird diese grammatikalische Definition der Wortform als **Lemma** bezeichnet. Demnach ist jedes Wort im mentalen Lexikon auf drei Ebenen zu definieren, nämlich auf der Konzeptebene, auf der Lemmaebene und auf der phonologischen Ebene (**Lexemebene**).

2.1.3 Mentaler Silbenspeicher und phonologische Bewusstheit

Werden allmählich mehr und mehr Silben erlernt, erkennt das Kind intuitiv bestimmte Strukturprinzipien. So scheinen Silben immer über einen Vokal zu verfügen. Davor und nach können kein, ein oder mehrere Konsonanten stehen. Das Kind lernt also intuitiv zwischen Konsonanten und Vokalen zu unterscheiden. Darüber hinaus erkennt das Kind, dass die Menge der vom Kommunikationspartner verwendeten Vokale und Konsonanten begrenzt ist. Das Kind wird nach dem Erlernen vieler Silben irgendwann erkennen, dass z. B. das Wort „Ball“ aus drei Lauten /ba/ aufgebaut ist, da diese drei Laute auch in anderen Silben auftreten (z. B. /b/ in /ba/ von „Banane“, /a/ in /ma/ von „Mama“, /l/ in /la/ in „Laterne“). Nach und nach wird das Kind durch das Vergleichen von Wortpaaren auch unterschiedliche Vokale erkennen. So können beispielsweise /e/ vs. /a/ in ansonsten gleichlautenden Wörtern durchaus bedeutungsunterscheidend, d. h. *phonemisch* wirken (z. B. „Quelle“ vs. „Qualle“; sogenannte *Minimalpaare*). In Tab. 2.1 ist das Lautsystem des Deutschen für alle Vokale und Konsonanten in phonetischer Lautschrift gelistet, und in Tab. 2.2 sind Realisierungsbeispiele für alle Laute bzw. **Phoneme** in Form von **Minimalpaaren** angegeben.

Abb. 2.2 zeigt neben den Kategorien Vokal vs. Konsonant noch weitere Unterkategorien. So gibt es im Deutschen Lang- und Kurzvokale (unterschieden durch den Doppelpunkt) wie auch Diphthonge (/aɪ/, /aʊ/, /ɔɪ/ nicht in der Tabelle aufgeführt), d. h. Vokale mit inhärentem Bewegungsablauf (z. B. von /a/ nach /ɪ/ in /aɪ/). Wie in der Abbildung gezeigt, können die Vokale in zwei Dimensionen (Zungenposition vorne-hinten und hoch-tief) angeordnet werden. Eine dritte Dimension ist hier durch Schrägstrich angedeutet und unterscheidet gerundete von ungerundeten Vokalen (Lippenrundung). Konsonanten kann man zunächst

Tab. 2.1 Minimalpaarliste Vokale

Lang- vs. Kurzvokal:	Nur Quantität:	/ɛ/ – /e:/ wie in „quellen“ – „quälen“ /a/ – /a:/ wie in „Schall“ – „Schaal“
	Quantität und Qualität:	/i:/ – /ɪ/ wie in „Miete“ – „Mitte“ /e:/ – /e/ wie in „Beet“ – „Bett“ /u:/ – /ʊ/ wie inw „spuken“ – „spucken“ /y:/ – /ʏ/ wie in „Fühler“ – „Füller“ /ø:/ – /œ/ wie in „Röslein“ – „Rösslein“ /o:/ – /ɔ/ wie in „Robe“ – „Robbe“
Langvokale:	Unterschiedlicher Öffnungsgrad:	/i:/ – /e:/ wie in „Tier“ – „Teer“ /e:/ – /ɛ:/ wie in „sehen“ – „sähen“ /ɛ:/ – /a:/ wie in „äsen“ – „aasen“ /u:/ – /o:/ wie in „Brut“ – „Brot“ /o:/ – /a:/ wie in „Mode“ – „Made“ /y:/ – /ø:/ wie in „rühren“ – „rören“ /ø:/ – /a:/ wie in „Öl“ – „Aal“
	Vordere ungerundete vs. hintere:	/i:/ – /u:/ wie in „Stiel“ – „Stuhl“ /e:/ – /o:/ wie in „leben“ – „loben“
	Vordere ungerundete vs. vordere gerundete:	/y:/ – /u:/ wie in „spülen“ – „spulen“ /ø:/ – /O:/ wie in „Bögen“ – „Bogen“
Kurzvokale:	Unterschiedlicher Öffnungsgrad:	/ɪ/ – /e/ wie in „Stille“ – „Stelle“ /e/ – /a/ wie in „Leck“ – „Lack“ /ʊ/ – /ɔ/ wie in „Glucke“ – „Glocke“ /ɔ/ – /a/ wie in „Motte“ – „Matte“ /ʏ/ – /œ/ wie in „knüpfen“ – „knöpfen“ /œ/ – /a/ wie in „Hölle“ – „Halle“
	Vordere ungerundete vs. hintere:	/ɪ/ – /ʊ/ wie in „Kippe“ – „Kuppe“ /e/ – /ɔ/ wie in „Geld“ – „Gold“
	Vordere ungerundete vs. vordere gerundete:	/ɪ/ – /ʏ/ wie in „Gericht“ – „Gerücht“ /e/ – /œ/ wie in „Kerner“ – „Körner“
	Vordere gerundete vs. hintere:	/ʏ/ – /ʊ/ wie in „drücken“ – „drucken“ /œ/ – /ɔ/ wie in „rösten“ – „rosten“

nach Lauten mit und ohne Stimmbeteiligung trennen (stimmhafte vs. stimmlose Laute, z. B. /b/ vs. /p/ oder /s/ vs. /z/), aber auch nach der Artikulationsart (Plosive, Frikative, Nasale ...) oder nach ihrem Artikulationsort. Die In [Abb. 2.2](#) gezeigten Laute werden erst richtig deutlich, wenn man die dazu in [Tab. 2.1](#) aufgelisteten Beispiele nachvollzieht.

Neben dem intuitiven Erlernen der Anzahl der bedeutungsunterscheidenden Lautkategorien erlernt das Kind auch noch, welche Silbenstrukturen in einer Sprache zugelassen sind (z. B. /CV/, /CVC/, /CCV/, /CCVC/, ... mit C = Konsonant und V = Vokal) und dass in einer Silbe nicht jede Lautfolge vorkommen darf. So ist im Deutschen die Lautfolge /CCV/ = /bla:/ erlaubt, nicht aber die Lautfolge /lba:/. Beispiele sind /CV/ = /ba:/, /li:/, ..., /CCV/ = /ta:l/, /la:k/ ...; /CCV/ = /kla:/, /ʃpa:/, ..., /CCVC/ = /blas/, /tru:k/ ... etc.

Tab. 2.2 Minimalpaarliste Konsonanten

Konsonanten:	Stimmlos vs. stimmhaft:	/p/ – /b/ wie in „Pass“ – „Bass“ /f/ – /v/ wie in „falten“ – „walten“ /t/ – /d/ wie in „Tank“ – „Dank“ /s/ – /z/ wie in „reißen“ – „reisen“ /k/ – /g/ wie in „Kunst“ – „Gunst“
	Nasal vs. nicht nasal	/m/ – /b/ wie in „Mitte“ – „Bitte“ /n/ – /d/ wie in „Name“ – „Dame“ /ŋ/ – /g/ wie in „Enge“ – „Egge“
	Plosive: Artikulationsort:	/p/ – /t/ wie in „peilen“ – „teilen“ /t/ – /k/ wie in „Tante“ – „Kante“ /b/ – /d/ wie in „Born“ – „Dorn“ /d/ – /g/ wie in „Daumen“ – „Gaumen“
	Frikative: Artikulationsort:	/f/ – /s/ wie in „Haff“ – „Hass“ /s/ – /ʃ/ wie in „Fleiß“ – „Fleisch“ /ʃ/ – /ç/ wie in „tauschen“ – „tauchen“ /v/ – /z/ wie in „wollen“ – „sollen“ /z/ – /j/ wie in „sagen“ – „jagen“
	Nasale: Artikulationsort	/m/ – /n/ wie in „kämmen“ – „kennen“ /n/ – /ŋ/ wie in „bannen“ – „bangen“



b

	Bilabial	Labiodental	Alveolar	Postalveolar/ palatoalveolar	Palatal	Velar	Uvular	Glottal
Plosive	p b		t d			k g		ʔ
Frikative		f v	s z	ʃ ʒ	ç j	x	ʁ	h
Nasale	m		n			ŋ		
Lateral			l					
Vibranten		r					R	
Approximanten		ʋ			j			

Realisierungen des Phonems /x/ sind: [ç] und [x]
Realisierungen des Phonems /r/ sind: [r], [ʀ] und [ʁ] und nach Vokal auch [ʁ]
Realisierungen des Phonems /v/ sind: [v] und [ʋ]
Realisierungen des Phonems /j/ sind: [j] und [ɥ]

Abb. 2.2 Lautsystem des Deutschen. **a** Vokalsystem (der Schrägstrich trennt ungerundete und gerundete Vokale); **b** System der Konsonanten (pro Kästchen: links stimmlose, rechts stimmhafte Konsonanten). Vibranten und Laterale stellen eine spezielle Gruppe von Approximanten dar

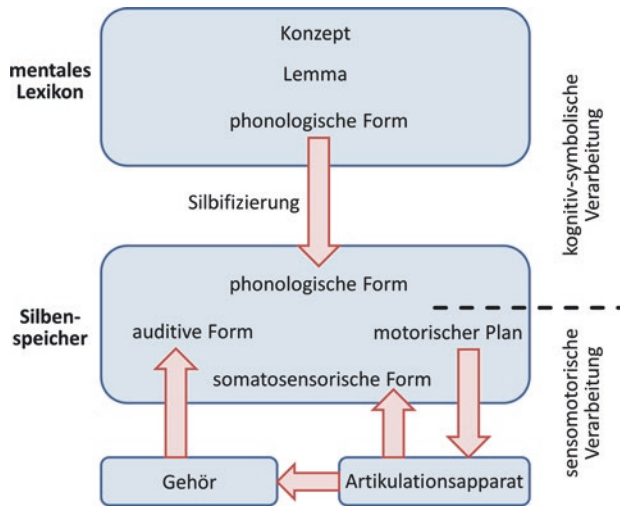


Abb. 2.3 Ablauf der Sprachproduktion unter Nutzung des mentalen Lexikons und des mentalen Silbenspeichers. Im mentalen Lexikon werden semantisches Konzept, syntaktische Wortspezifikationen (Lemma) und phonologische Form (Lautfolge) assoziiert. Im mentalen Silbenspeicher werden die Lautfolgen einzelner Silbe mit dem Motorplan der Silbe und der auditiven Form der Silbe (d. h., wie die Silbe klingt) und auch mit der somatosensorischen Form der Silbe (d. h., wie sich die Artikulation der Silbe taktil und propriozeptiv anfühlt) assoziiert. In manchen Ansätzen wird die Konzept- und Lemmaebene auch zusammengefasst. Dies ist aber problematisch, da Syntax wie phonologische Form stark sprachspezifischer Natur sind, während dies für die Konzeptebene nicht gilt

Somit ist das Kind nach und nach in der Lage, die lautliche Struktur, d. h. die *phonologische Struktur* der Silben, zu erkennen, und es ist dann auch in der Lage, komplexe Wörter in Silben zu zerlegen (z. B. „lachen“ -> /la+/xən/). Um die Speicherkapazität des Gehirns effektiv zu nutzen, werden nun auch kognitiv symbolische Repräsentationen für die in der Sprache auftretenden unterschiedliche Laute angelegt, und im mentalen Lexikon wird nun eine Verbindung zwischen dem metasprachlichen Konzept eines Wortes und seiner Lautfolge angelegt (Abb. 2.3). Die Speicherung von Wörtern geschieht somit nun rein kognitiv und damit aus neuronaler Sicht sehr platzsparend. Nachfolgend werden Wörter bei der Produktion silbifiziert und deren motorische Formen (Motorplan) und das „sensorische Soll“ (also wie sich die produzierte Silbe anzuhearsen hat) dann einfach aus dem Silbenspeicher abgerufen (Abb. 2.3). Dies realisiert dann die phonetische Form des Wortes oder der Äußerung.

Fazit zu Abschn. 2.1

Das semantische Netzwerk umfasst eine Menge von Konzepten und eine Menge von Relationen zwischen Konzepten. Konzepte wiederum umfassen ein oder mehrere semantische Merkmale. Während Konzepte metasprachlich sein können, sind Wörter (Lemmas und Lexeme) sprachspezifisch. Viele Konzepte können aber auch sprachspezifisch als Wortbedeutungen angesehen werden. Die Aussprache eines Wortes definiert

seine phonologische Form. Der grammatikalische Status eines Wortes ist auf der Lemmaebene spezifiziert. Das Lautinventar einer Sprache (Phoneminventar) erlaubt die phonologische Spezifikation jedes Wortes. Im mentalen Lexikon sind alle Wörter einer Sprache als symbolisch-kognitive Entitäten abgespeichert (ca. 60.000 Wörter). Im Silbenspeicher sind auditive, somatosensorische und motorische Repräsentationen (sensorisch-motorische Repräsentationen) aller häufigen Silben einer Sprache abgelegt (ca. 2000 Silben). Damit können alle aus häufigen Silben bestehenden Wörter und Sätze aufgrund der abgespeicherten motorischen Repräsentationen (Motorpläne) direkt realisiert werden. Dies sind im Deutschen oder Englischen ca. 95 % aller täglich gesprochenen Äußerungen. Die gespeicherten auditiven und somatosensorischen Formen (sensorischen Formen) der häufigen Silben dienen der Kontrolle der artikulatorischen Ausführung. Seltene Silben können aus häufigen Silben, die eine ähnliche Struktur aufweisen, realisiert werden.

2.2 Die Artikulation

Die *Silbe* ist die Grundeinheit der Artikulation. Sprechen wir beispielsweise Buchstaben des Alphabets aus, so tun wir das in Silbenform. Wir sagen nicht /a:/ sondern /ʔa:/ (glottale Verschlusslösung vor dem /a:/); nicht /b/ (wie immer das auch gehen sollte), sondern /be:/ etc. Wir könnten sagen /n:/ oder /s:/; letztlich sagen wir aber /ʔen/ und /ʔes/. Wir bedienen uns also gerne einer silbischen Struktur, nämlich im einfachen Falle der /CV/-Struktur oder der /CVC/-Struktur. Die Debatte, ob einige im Deutschen auftretenden Laute wie z. B. der Glottisverschlusslaut /ʔ/ Phonemstatus haben oder nicht (also bedeutungsunterscheidend wirken können oder nicht), ist im Kontext dieses Buches unwichtig. Unstrittig ist, dass bereits auf der Ebene der motorischen Planung ausgehend von der Lautfolge der Silbe (Abb. 2.3) im Deutschen immer zunächst ein anlautender Konsonant erwartet wird. Dies Grundprinzip lernt das Kind intuitiv und ist Teil seiner phonologischen Bewusstheit.

Wie aber vollzieht sich der Übergang von der lautsymbolisch-kognitiven Form einer Silbe zu ihrer motorischen Realisierung? Im Fall der häufigen Silben ist die Antwort verblüffend einfach. Denken wir zurück an das Beispiel „Ball“. Das Kind hat viele Versuche der Realisierung dieses Wortes (bzw. dieser Silbe) gemacht und hat parallel mit dem Entstehen seiner phonologischen Bewusstheit die symbolisch kognitive Form der Lautfolge (das Konzept) wie auch die Artikulation dieses Wortes gelernt und beide Formen miteinander assoziiert, also neuronal verbunden. Ein hierbei zu leistender wichtiger und nicht trivialer Schritt ist das Erwerben der Fertigkeiten zur Artikulation von Silben und Wörtern.

2.2.1 Motorische Ablaufpläne und Sprechbewegungseinheiten

Das motorische System ist hierarchisch aufgebaut. Zunächst wird eine Aktion (z. B. eine Hand-Arm-Bewegung) als kognitive Einheit abgerufen; z. B. „Winke!“. Dann wird diese kognitiv definierte Aktion evtl. noch in zeitlich nacheinander ablaufende diskrete

Unteraktionen differenziert. Mit jeder Aktion bzw. Unteraktion wird dann ein im Verlauf der Entwicklung des Kindes gelernter motorische Plan aktiviert, z. B. „hebe Arm“, „schwenke Hand“, „senke Arm“. Erst dann werden Muskelgruppen zeitlich genau koordiniert aktiviert, um die zeitliche Sequenz der motorischen Pläne der (Unter-)Aktionen auszuführen. Ähnlich ist die *Motorhierarchie der Sprechartikulation* aufgebaut. Zunächst wird die zu realisierende Silbe kognitiv spezifiziert (die Phonem- bzw. Lautfolge der Silbe bzw. der Silbenfolge eines Wortes). Danach wird der **Motorplan** jeder Silbe aus dem Silbenspeicher heraus abgerufen bzw. zeitlich nacheinander aktiviert. Erst dann wird die eigentliche Artikulation, d. h. die Aktivierung der Muskulatur ausgeführt.

Auf der Ebene der *motorischen Planung* ist die Artikulation der Silbe noch nicht dergestalt spezifiziert, dass bereits ein zeitlicher Ablaufplan für die Aktivierung einzelner Muskeln bzw. Muskelgruppen für die Ansteuerung jedes Artikulators (Oberlippe, Unterlippe, Zunge etc.) vorliegt. Allerdings ist hier bereits die *zeitliche Planung* der für die artikulatorische Realisierung der Silbe wesentlichen Artikulationsbewegungen dergestalt definiert, dass z. B. im Falle des Wortes „Ball“ Folgendes klar ist:

- Ich benötige anfangs ein Zeitintervall für den Lippenverschluss zur Realisierung des /b/.
- Danach benötige ich ein Zeitintervall für die Realisierung der Mundöffnung und die Einstellung der Zunge zur Realisierung des /a/.
- Danach benötige ich ab einem bestimmten Zeitpunkt ein Zeitintervall für die Hebung der Zungenspitze an den Gaumen zur Realisierung des /l/, wobei die Zungenspitze nach der Hebung für ein definiertes Zeitintervall den hinter den Zähnen liegenden Zahndamm berühren muss.

Ein zeitlicher Ablaufplan für eine ähnliche CVC-Silbe ist in [Abb. 2.7](#) dargestellt.

Zunächst aber soll etwas genauer auf den **Sprechapparat** und die für das Sprechen wichtigen **Artikulatoren** eingegangen werden ([Abb. 2.4](#)). Von außen sichtbar sind Ober- und Unterlippe und die beim Sprechen auftretenden Bewegungen des Unterkiefers. Manchmal sehen wir auch das vordere Stück der Zungenspitze. Wichtig für die Lautartikulation ist aber die gesamte Zunge, die je nach Vokal (oder auch je nach Konsonant) sehr unterschiedlich geformt sein kann und ihre Gesamtmasse im Mund- und Rachenraum in weiten Bereichen verschieben kann. Deshalb werden die Vokale auch gerne nach den phonetischen Merkmalen hoch-tief bzw. vorne-hinten eingeteilt. Das Gaumensegel kann gehoben (nicht nasale Artikulation) oder abgesenkt sein. Im Fall der Absenkung kann die Luft nicht nur durch den Mundraum, sondern auch durch den Nasenraum (Nasaltrakt) strömen, und es können dann Nasallaute (Nasale) gebildet werden. Die Stimmritze ist im Fall der stimmlosen Laute geöffnet, im Fall der stimmhaften Laute aber locker verschlossen, um in diesem Falle schwingen zu können ([Abb. 2.5](#)) und dadurch **Phonationsschall** zu erzeugen (roter Punkt in [Abb. 2.4](#)).

Drei artikulatorische Momentaufnahmen für die in der Silbe /bal/ enthaltenen Laute sind in [Abb. 2.6](#) dargestellt. Dabei muss aus der Sicht der motorischen Planung der Silbe /bal/ sichergestellt sein, dass über die Gesamtzeit der Realisierung dieser Silbe das Gaumensegel angehoben und damit die **velopharyngeale Pforte** (d. h. der Übergang zwischen

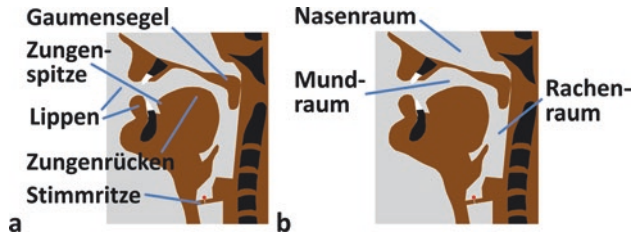


Abb. 2.4 Der menschliche Sprechapparat. **a** die Artikulatoren; **b** die Hohlräume (roter Punkt: Ort der Entstehung des Phonationsschalls). Gebräuchliche lateinische Bezeichnungen: Stimmritze → Glottis; Gaumensegel → Velum; Rachenraum → Pharynx; Zungenrücken → Dorsum; Zungenspitze → Apex

Rachen- und Nasenraum) verschlossen ist. Das Gaumensegel ist also für die gesamte Dauer dieser Silbe angehoben, damit der Nasenraum nicht akustisch mit dem Mundraum verbunden ist und nur nicht nasale Laute realisiert werden. Eine Verbindung durch Senkung des Gaumensegels existiert vor allem im Fall der Nasallaute (z. B. /m/ und /n/). Darüber hinaus muss sichergestellt sein, dass die Stimmklappen mittels Einstellbewegung der Aryknorpel nah beieinander liegen, um die lockere Verschließung der Stimmritze zu gewährleisten, um so die Stimmklappenschwingung und damit die Phonation zu ermöglichen (Abb. 2.5).

Es ist aber nicht ausreichend, nur die in Abb. 2.6 angegebenen „Momentaufnahmen“ der Lautbildung (Lautmittelpunkte) im motorischen Plan einer Silbe zu definieren. Vielmehr müssen auch Aussagen über die Dauer von Verschlussbildungen, über den Beginn von Artikulationsbewegungen etc. im Motorplan einer Silbe festgelegt sein. Es kann angenommen werden, dass im Fall des Sprechens wie bei anderen Bewegungen (z. B. beim Winken) nicht nur das Ziel der Bewegung oder Aktion (im Fall des Sprechens z. B. die Aktion der konsonantischen Verschlussbildung) zu definieren ist, sondern dass weitere Parameter der Aktion wie z. B. die Dauer der Bewegung insgesamt von ihrem Beginn an bis hin zum Erreichen des Ziels im Motorplan festgeschrieben werden müssen. Aus der Sicht der Artikulation definieren wir daher nicht nur eine Folge von Lauten, sondern

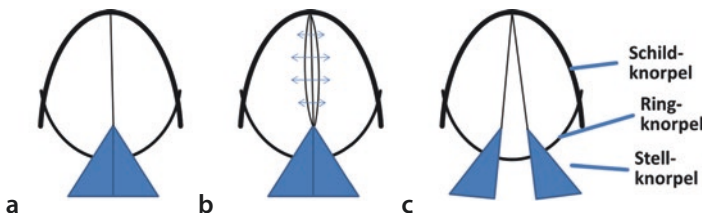


Abb. 2.5 Stimmklappen locker geschlossen zur Phonationsstellung für stimmhafte Laute (**a** keine Phonation; **b** mit Phonation: schnell schwingende Stimmklappen) und **c** geöffnete Stellung der Stimmklappen zur Realisierung stimmloser Laute. Dicke schwarze Linie oben: Schildknorpel, der oft auch als Adamsapfel sichtbar wird; mitteldicke schwarze Linie unten: Ringknorpel. Dünne vertikale Linien: Stimmklappen; blau: Aryknorpel (Stellknorpel)

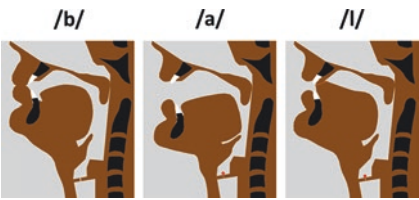


Abb. 2.6 Momentaufnahmen zur Artikulation der Laute des Wortes „Ball“

auf der Ebene der motorischen Planung eine Menge von zeitlich miteinander koordinierten und sich zeitlich überlappenden Aktionen, nämlich der **Sprechbewegungseinheiten** (*speech action units*, vgl. Goldstein et al. 2007). Eine Liste der im Deutschen (Standarddeutsch bzw. Hochdeutsch) auftretenden Sprechbewegungseinheiten (SBEs) ist in [Tab. 2.3](#) dargestellt.

Tab. 2.3 Das System der Sprechbewegungseinheiten für das Deutsche. Es können vier Typen von Sprechbewegungseinheiten unterschieden werden: vokalische, enge- bzw. verschlussbildende, pharyngeale und glottale SBEs

Name	SBE zur ...	z. B. zur Realisierung von ...
hhzr	hohen Hebung des Zungenrückens	/i:/, /u:/
hezr	Hebung des Zungenrückens	/e:/, /o:/
sezt	Senkung des Zungenrückens	/ɛ:/
tszt	Tiefen Senkung des Zungenrückens	/a:/
rvzt	Rückverlagerung des Zungenrückens	/u:/, /o:/
vvzt	Vorverlagerung des Zungenrückens	/i:/, /e:/
ruli	Rundung der Lippen	/u:/, /y:/
vbli	Verschlussbildung durch die Lippen	/b/, /p/, /m/
vbzs	Verschlussbildung durch die Zungenspitze	/d/, /t/, /n/
vbzt	Verschlussbildung durch den Zungenrücken	/g/, /k/
eazs	Alveolaren Engebildung durch die Zungenspitze	/s/, /z/
epzs	Postalveolaren Engebildung durch die Zungenspitze	/ʃ/, /ʒ/
oegl	Geste zur Öffnung der Glottis	stimmlose Laute
phgl	Geste zur lockeren Verschließung der Glottis (Phonation)	stimmhafte Laute
vbgl	Geste zur festen Verschließung der Glottis	/ʔ/
vbvp	Geste zur Verschchlussbildung der velopharyngealen Pforte	Obstruenten (Plosive und Frikative)
oevp	Geste zur Senkung des Gaumensegels (Öffnung der velopharyngealen Pforte)	Nasale

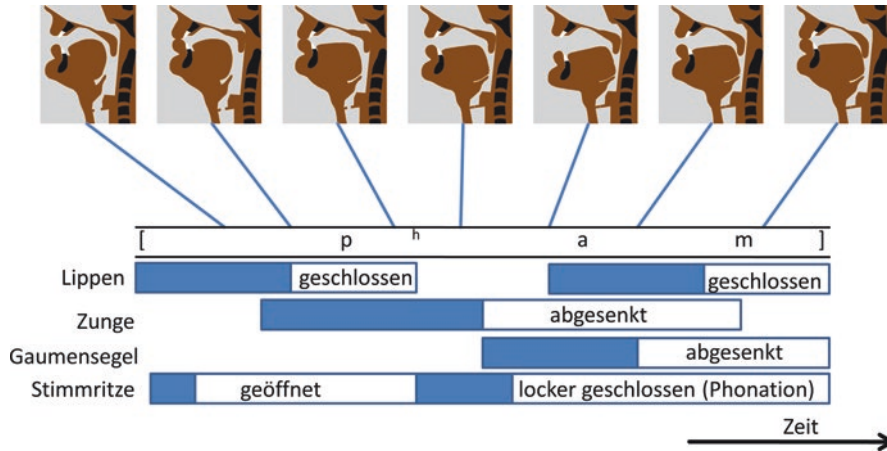


Abb. 2.7 Motorischer Ablaufplan der Silbe /pam/. Die blauen Bereiche (Bewegungsphase) und dahinter liegenden blau umrahmten Kästchen (Zielphase) kennzeichnen Bewegungs- und Zielphase und damit die gesamte Zeitdauer jeder Sprechbewegungseinheit (SBE). Namen der Gesten nach Tab. 2.3: Lippen: vbli, vbli; Zunge: tszr; Gaumensegel: vbvp (nicht gezeigt), oevp; Stimmritze: oegl, phgl

Abb. 2.7 stellt eine Visualisierung der zeitlichen Koordinierung der Sprechbewegungseinheiten für die Silbe /bal/ (z. B. im Wort „Pam_pel_mu_se“) dar. Die Wahl fiel auf diese (im Deutschen nicht unbedingt häufige) Silbe, um auch die Einbeziehung der glottalen und der velopharyngealen Artikulation, also die Einbeziehung der Artikulatoren Glottis und Gaumensegel (Velum) bei der Sprechartikulation deutlich zu machen. In der Silbe /bal/ ist der erste Laut stimmlos. Deshalb muss hier die Stimmritze weit geöffnet sein (oegl, Tab. 2.3). Bei der Realisierung der nachfolgenden stimmhaften Lauten hingegen muss die Stimmritze locker geschlossen sein, um so ein Schwingen der Stimmlippen und damit die Phonation zu ermöglichen (phgl). Der Laut /m/ ist ein Nasal, sodass hier das Gaumensegel abgesenkt sein muss, da die Realisierung dieses Lautes die Ankopplung des Nasenraumes an den Mundraum verlangt (oevp). Diese glottalen und velopharyngealen SBEs müssen mit oralen SBEs zeitlich koordiniert sein. Zur Realisierung der Konsonanten treten hier labiale Verschlussbildungen auf (vbli). Zur Realisierung des Vokals ist hier eine tiefe Senkung des Zungenrückens erforderlich (tszr).

2.2.2 Artikulatorische Kompensation

Warum aber wird nicht bereits auf hoher motorischer Ebene – d. h. auf der Ebene der motorischen Planung – detailliert die zeitliche Abfolge der neuronalen Aktivierung jedes einzelnen Muskels bzw. jeder einzelnen Muskelgruppe für aller Artikulatoren definiert? Die Antwort lautet, dass dies die Flexibilität und Redundanz der Artikulation einschränken würde. Stellen wir uns einmal vor, dass wir unfall- oder krankheitsbedingt den Unterkiefer nur noch eingeschränkt bewegen können. Wir können dann dennoch sofort, also

ohne umfangreiches „artikulatorisches Nachlernen“, verständlich sprechen, obwohl die Aktivierungsmuster der einzelnen am Sprechen beteiligten Muskeln bzw. Muskelgruppen nun ganz anders aussehen müssen. Wir können dies leicht durch die Ausführung eines Experimentes belegen. Fixieren wir das Ende eines Schreibstiftes seitlich (also im linken oder rechten Mundwinkel) zwischen den Zähnen, ohne dass der Stift aber weiter in den Mundraum hineinragt, also ohne dass die Artikulation der Zunge hierdurch allzu stark eingeschränkt wird. Der Stift soll im Wesentlichen den Effekt haben, die Position des Unterkiefers zu fixieren. Dieses experimentelle Paradigma wird auch als *Beißblockexperiment* bezeichnet. Wir können feststellen, dass wir trotz Stift sofort in verständlicher Weise weitersprechen können. Dies passiert, obwohl hier doch eigentlich eine vollständige „Neuprogrammierung“ der muskulären Steuerung erfolgen müsste, da die fehlende Unterkieferbewegung nun durch stärkere Bewegungen der Zunge und der Lippen kompensiert werden muss. So kann nun beispielsweise der Verschluss des /b/ in „Ball“ nicht mehr unter Zuhilfenahme der Anhebung des Unterkiefers sondern nur noch allein durch Unter- und Oberlippenbewegung realisiert werden. Dieses Phänomen bezeichnet man als *sofortige motorische Kompensation* (vgl. Kelso und Tuller 1983).

In ähnlicher Weise muss die Einstellung des Mundraumes zu einem /a/ nun durch eine stärkere Lippen- und Zungenbewegung erfolgen, da die Hilfestellung leistende gleichzeitige Absenkung des Unterkiefers nun entfällt. Neben der eingeschränkten Funktionsfähigkeit des Kiefergelenkes können wir uns aber auch andere Paradigmen vorstellen, z. B. eine eingeschränkte Funktionsfähigkeit eines bestimmten Muskels oder einer Muskelgruppe eines bestimmten Artikulators. In diesem Fall können wir in ähnlicher Weise auch schnell wieder normal sprechen, also kompensieren, da die motorische Planung der Silben ja erhalten geblieben ist, und jetzt „nur“ die muskuläre Ausführung durch andere Muskelgruppen angepasst werden muss, um das Ziel der Planung – nämlich das Erreichen einer bestimmten EndEinstellung des Artikulators für eine bestimmte SBE – nicht zu gefährden.

Auf „hoher Planungsebene“ werden somit nicht muskuläre Aktivierungsmuster, sondern nur das Zeitmuster der SBEs (ihre Bewegungs- und Zielzeitintervalle) sowie die Ziele der SBEs selbst, nämlich die Produktion eines konsonantischen Verschlusses, einer konsonantischen Enge, einer bestimmten vokalischen Hohlraumformung im Mund- und Rachenraum oder die Einstellung einer bestimmten Öffnungsweite oder eines Verschlusses für die velopharyngeale Pforte und für die Stimmritze festgelegt. Dies definiert den *Motorplan* einer Silbe. Genau diese Information generiert eindeutige akustische Muster für jede Silbe und macht dadurch jede Silbe auditiv erkennbar und damit jedes Wort und jede Äußerung verstehbar.

Da also letztlich das akustische bzw. auditive Resultat das Ziel der Sprachproduktion und damit das Ziel der Gesamtheit aller eine Silbe realisierenden SBEs ist, ist eine enge Kopplung zwischen Produktion und auditiver Feedbackwahrnehmung notwendig. Diese *auditive Rückkopplung* wird insbesondere während des Spracherwerbs genutzt. Das aus der Artikulation entstehende akustische Signal sowie die bei der Artikulation entstehenden somatosensorischen Rückkopplungssignale (*somatosensorische Rückkopplung*) werden an die Planungsebene zurückgemeldet, und es wird anhand dieser Information überprüft, ob der motorische Plan korrekt ist. Dabei ist die Rückmeldung somatosensorischer

Information, d. h. der Information über die gegenseitige Berührung von Artikulatoren (z. B. Ober- und Unterlippen) oder von Artikulator mit Sprechtraktwand (z. B. Zungenspitze mit Zahndamm) oder die Information über den momentanen Öffnungsgrad des Unterkiefers oder der augenblicklichen Stellung der Zunge (via muskulärer Sensoren zum muskulären Spannungszustand oder via Gelenkrezeptoren zum momentanen Einstellwinkel des Gelenks) derart schnell, dass diese Information auch während der Artikulation unterhalb der Ebene der motorischen Planung dazu genutzt werden kann, zu prüfen, ob eine SBE im zeitlichen Plan liegt und korrekt ausgeführt wird, d. h. ob ein in einer SBE definiertes artikulatorisches Ziel (z. B. ein Lippenverschluss) auch tatsächlich zur vom Motorplan vorgesehenen Zeitpunkt erreicht wird. Ist aufgrund der Planung anzunehmen, dass die Aktivierung der SBE zu früh endet, kann somit aufgrund der somatosensorischen Rückkopplung noch während der Ausführung der SBE „on the fly“ die Ausführungsdauer der SBE verlängert oder die Artikulationsgeschwindigkeit in Richtung des artikulatorischen Zieles der SBE erhöht werden. Somatosensorisches Feedback kann somit während der Produktion zur Korrektur der Artikulation (z. B. der Ausführung einer SBE, aber auch zur Änderung der zeitlichen Ausdehnung gerade aktiver SBEs) herangezogen werden, während die durch das langsamere auditive Feedback gegebene Information in erster Line dergestalt genutzt werden kann, dass man evtl. den Sprachfluss stoppt, wenn auditiv ein Produktionsfehler rückgemeldet wurde. Das aktuell fehlerhaft realisierte Wort kann dann neu und korrigiert realisiert werden.

2.2.3 Artikulatorische Merkmale unterschiedlicher Lautklassen

Als Lautklassen unterscheiden wir im Deutschen nicht nur Vokale und Konsonanten, sondern wir können entsprechend [Abb. 2.2](#) die Konsonanten in Plosive, Frikative, Nasale, Laterale, Vibranten und Approximanten unterteilen (**Artikulationsart**). Darüber hinaus können Plosive und Frikative an jeder Artikulationsstelle bzw. an jedem **Artikulationsort** (bilabial bis glottal; s. die Spalten in der Konsonantentabelle in [Abb. 2.2](#)) noch in stimmhafte und stimmlose Laute unterschieden werden (rechts und links in den Kästen der Konsonantentabelle in [Abb. 2.2](#)). Im Folgenden werden die artikulatorischen und aerodynamischen Merkmale dieser im Deutschen vorkommenden Lautklassen dargestellt (vgl. Raphael et al. 2007; Grassegger 2016).

Vokale zeichnen sich gegenüber Konsonanten vor allem dadurch aus, dass der Fluss des Luftstroms ausgehend von den Lungen durch die Luftröhre (Trachea) und dann durch die Stimmritze (Glottis) oberhalb der Stimmritze, also im Rachen- und Mundraum, nicht behindert wird. Dies gilt aber auch für Nasale und Laterale, sodass die Definition von Vokalen noch durch den Zusatz zu ergänzen ist, dass dabei keine Verschlussbildung im Mundraum (wie bei Nasalen) und auch keine Behinderung des Luftstroms in der Mitisagittalebene des Mundraumes (z. B. wie bei sagittaler Hebung der Zungenspitze bei /l/) auftreten darf. Wir haben hier bewusst nicht die Definition von Vokalen als sogenannte „Öffnungslaute“ gewählt, da insbesondere bei den Eckvokalen /i:/, /y:/, /a:/ und /u:/ (s. Vokaldiagramm in [Abb. 2.2](#)) sehr wohl vokalische Engebildungen im Ansatzrohr



Abb. 2.8 Die Mediosagittalschnitte der Eckvokale /a:/, /i:/, /u:/ (vgl. Abb. 2.2).

vorliegen (s. Abb. 2.8). Diese Engebildungen sind jedoch nicht so stark, dass das Strömen der Luft behindert würde. Im Falle der Vokale tritt die stärkste Engebildung an der Stimmritze auf. Die stärkste Engebildung ist diejenige, die aus aerodynamischer Sicht bei der Lautproduktion die jeweils ausströmende Luftmenge determiniert.

Das Vokaldiagramm in Abb. 2.2 spiegelt die Ordnung der Vokale nach artikulatorischen Kategorien wider. So liegt die vokalische Enge bei den vorderen Vokalen auch im vorderen Mundbereich (palatal), bei hinteren Vokalen im hinteren Mundbereich (velar). Darüber hinaus ist die Anordnung der Vokale auch nach Zungenhöhe (bzw. Öffnungsgrad des Mundes und damit auch nach Grad der Senkung des Unterkiefers) angegeben. So liegt die Zunge bei /i:/, /y:/ und /u:/ sehr hoch, während /a:/ den tiefsten Vokal darstellt. Eine dritte Dimension (neben vorne-hinten und hoch-tief) stellt die vokalische Dimension gerundet-ungerundet dar. So sind die hinteren Vokale im Deutschen alle gerundet, während die vorderen Vokale ungerundet (/i:/, /ɪ/, /e:/, /ɛ:/, /ɐ/), aber auch gerundet (/y:/, /ʏ/, /ø:/, /œ/) auftreten können. Eine vierte Dimension speziell im Deutschen ist die Unterscheidung zwischen Kurz- und Langvokalen. Diese ist durch das Längungszeichen /:/ gekennzeichnet. Zur Unterscheidung von Lang- und Kurzvokalen siehe insbesondere die in Tab. 2.1 angegebenen Minimalpaare. Neben den Lang- und Kurzvokalen gibt es auch noch die reduzierten Vokale /ə/ und /ɐ/ (Schwalaute und r-gefärbter Schwalaute, wie z. B. am Ende der Wörter „Miete“ vs. „Mieter“). Dieser Vokaltyp kann nur in unbetonten Silben auftreten. Als weitere Unterkategorie der Langvokale können im Deutschen noch die Diphthonge /aɪ/, /aʊ/ und /ɔɪ/ genannte werden. Dies sind Vokale mit hörbarer oder unverdeckter Artikulationsbewegung und treten beispielsweise in den Wörtern „Mai“, „lau“ und „neun“ auf.

Nasale sind artikulatorisch recht einfach beschreibbar. Bei diesen Lauten ist das Gaumensegel abgesenkt, sodass der gesamte Luftstrom durch die geöffnete velopharyngeale Pforte und damit vom Rachenraum in den Nasenraum geführt wird, während im Mundraum ein bilabialer, alveolarer oder velarer Verschluss gebildet wird. Der Luftstrom wird im Mundraum somit vollständig gestoppt. Beim bilabialen Verschluss wird Ober- und Unterlippe zum Verschluss geführt. Beim alveolaren Verschluss wird die Zungenspitze (Apex) zum Zahndamm (Alveolen) geführt. Beim velaren Verschluss wird der Zungenrücken (Dorsum) an den weichen Gaumen (Velum) geführt (Abb. 2.9).

Plosive stoppen den durch den Kehlkopf geführten Luftstrom vollständig. Im Deutschen kann der bilabiale, der alveolare und der velare Vollverschluss im Mundraum gebildet werden. Im Unterschied zu den Nasalen ist nun aber auch das Gaumensegel vollständig angehoben und damit die velopharyngeale Pforte verschlossen, sodass der Luftstrom



Abb. 2.9 Mediosagittalschnitt des bilabialen, alveolaren und velaren Nasals. Der Luftstrom wird hierbei durch Rachen- und Nasenraum geführt. Die velopharyngeale Pforte ist geöffnet, da das Gaumensegel z. B. im Vergleich insbesondere zu den Frikativen und Plosiven stark abgesenkt ist



Abb. 2.10 Mediosagittalschnitt der stimmhaften Plosive /b/, /d/, /g/

vollständig im Rachen- und Mundraum gestaut wird (Abb. 2.10). Somit herrscht auch im Mundraum ein vom Lungensystem durch die Stimmritze übertragener Überdruck während der Verschlussphase eines Plosivlautes, der dann mit der anschließend erfolgenden Öffnung des Verschlusses, der sogenannten Verschlusslösung, abgebaut wird. Mit der Verschlusslösung geht auch ein sofortiger Abbau des Überdrucks einher. Luft strömt jetzt schnell nach vorne, also in Richtung Mund und auch vor den Mund. Daher rührt der Name Plosiv für diese Laute. Bei den stimmlosen Plosiven /p/, /t/ und /k/ wird der Lungenluftdruck wegen der weit geöffneten Stimmritze vollständig in den Mundraum übertragen. Bei stimmhaften Plosiven /b/, /d/ und /g/ muss dies nicht der Fall sein, da insbesondere bei Auftreten eines solchen Lautes zwischen zwei Vokalen die Phonation oftmals während der Verschlussphase des Plosivs aufrechterhalten bleibt und damit ein gewisser Abfall eines Teils des oberhalb der Lungen gebildeten Luftüberdrucks auch über der glottalen Enge stattfindet. Ansonsten würde die Stimmlippenschwingung nämlich während der oralen Verschlussphase enden. In der Tat ist die Aufrechterhaltung der Stimmlippenschwingung bei Plosiven wegen der oralen Unterbrechung des Luftstroms schwierig. Auf die Wichtigkeit des Luftstroms für die Aufrechterhaltung der glottalen Schwingung kommen wir in Abschn. 2.1.3 zurück.

Frikative werden ähnlich wie Plosive gebildet. Die velopharyngeale Pforte ist hier wie bei den Plosivlauten vollständig verschlossen. Im Mundraum wird jetzt aber kein Verschluss, sondern nur eine den Luftstrom hemmende Engstelle realisiert. Die Enge ist so stark, dass der Luftdruck zwischen Stimmritze und Enge höher ist als der Luftdruck vor der Enge, also am Mund. Der Luftstrom ist somit gezwungen, diese Enge zu passieren. Es entstehen dabei wegen der Erhöhung der Strömungsgeschwindigkeit der Luftpartikel in der Enge Turbulenzen „stromabwärts“, also vor der Enge in Richtung Mund, die wiederum



Abb. 2.11 Mediosagittalschnitte der stimmlosen Frikative im Deutschen. Aufbau des Luftdrucks hinter der Enge (Hohlraum mit rotem Kreuz) und Entstehung der sekundären Schallquelle zur Rauschgenerierung vor der Enge, also stromabwärts (roter Punkt). Die primäre Schallquelle ist die Stimmritze

als Geräusch, also als Rauschen hörbar werden. Dieses Rauschen können wir uns bewusst machen, indem wir den Laut „s“ /s:/ oder „sch“ /ʃ:/ lange anhalten. Das Geräusch wird im Mundraum kurz vor (also stromabwärts) der Engstelle gebildet. Frikative können im Deutschen an vielen Artikulationsstellen gebildet werden; nämlich bilabial, alveolar, palatoalveolar, palatal und velar (Abb. 2.11). Der Unterschied zwischen alveolar und palatoalveolar ist der Unterschied zwischen „s“ /s/ und „sch“ /ʃ/ (s. auch Minimalpaarliste Tab. 2.2), wobei das „sch“ etwas weiter hinten gebildet wird als das „s“. Der Unterschied zwischen /ç/, z. B. in „ich“ und /x/, z. B. in „ach“ liegt ebenfalls in der Änderung der Artikulationsstelle von palatal (harter Gaumen) zu velar (weicher Gaumen). In Abb. 2.11 ist ebenfalls das /h/ der Gruppe der stimmlosen Frikative hinzugefügt worden. Hier liegt die Verschlussbildung glottal und nicht um Mundraum. Das Gleiche gilt nach Tab. 2.2 auch für den Glottisverschlusslaut /ʔ/, der der Gruppe der stimmlosen Plosive hinzugefügt wurde.

Da im Mundraum ein Überdruck zur Realisierung der oralen Geräuschbildung an der Engstelle vorliegt, ist die Aufrechterhaltung der Stimmhaftigkeit bei stimmhaften Frikativen (s. Abb. 2.2 und Tab. 2.2) ähnlich der bei Plosiven nicht einfach. Deshalb werden alle Laute, die nicht Frikative oder Plosive sind, auch als *Sonoranten* bezeichnet. Bei den Lauten dieser Lautgruppen (Vokale, Nasale, Laterale, Approximanten) ist die Aufrechterhaltung der Stimmlippenschwingung (Phonation) kein Problem, und diese Laute werden deshalb auch immer stimmhaft (oder sonor) gebildet. Die Frikative und Plosive werden auch als *Obstruenten* zusammengefasst. Bei Bildung dieser Lautgruppe tritt immer eine dermaßen große Behinderung des Luftstroms im Mundraum auf, dass diese zu einem Aufbau des Luftüberdrucks stromaufwärts hinter dieser Engstelle, also hinter der Obstruktion oder Konstriktion, führt und damit auch die Phonation erschwert. Dieser Aufbau eines Luftüberdrucks im Mundraum ist das eigentlich wichtige Lautmerkmal der Frikative und Plosive und für die Ausbildung der zugehörigen akustischen Lautmerkmale, nämlich des turbulenten Rauschens, unabdingbar.

Laterale bzw. laterale Approximanten sind durch die Ausbildung einer Berührung der Zunge mit dem Zahndamm oder einem weiter hinten liegenden Artikulationsort (harter/

weicher Gaumen) in der Mittelebene gekennzeichnet. Seitlich (also lateral) liegt aber keine Verschiebung vor, sodass der Luftstrom ohne Widerstand lateral strömen kann. Im Deutschen kennen wir nur den alveolaren Laterallaut /l/, der durch eine mediosagittale Berührung von Zungenspitze mit Zahndamm gekennzeichnet ist (Mediosagittalschnitt also deckungsgleich zu dem des /d/ in Abb. 2.10), wobei die seitlichen Zungenränder aber abgesenkt sind, sodass der Luftstrom hier ungehindert, d. h. ohne Aufbau eines Luftüberdrucks im Mundraum, passieren kann.

Approximanten stellen eine Art Zwischenlautgruppe zwischen Konsonanten und Vokalen dar. Der Hintergrund ist, dass die Enge im Mundraum hier nur derart kurzzeitig ausgebildet wird, dass dadurch die Luft nicht gestaut und der Luftstrom nicht behindert wird. So kann der labiodentale Approximant /v/ als sehr kurzer (oder anders ausgedrückt: als flüchtig realisierter) labiodentaler Frikativ /v/ und der palatale Approximant /j/ als sehr kurze Realisierung des palatalen Frikativs /j/ angesehen werden. Im Deutschen können beiden Realisierungen je nach Sprecher z. B. am Anfang der Wörter „Vase“ oder „Julia“ auftreten.

Vibranten gehören auch zur Gruppe der Vibranten und treten im Deutschen fast nur noch dialektal auf (z. B. im Bairischen). Im Hochdeutschen ist das Rollen der Zungenspitze beim /r/ bzw. das Schwingen des am Gaumensegel anhängenden Zäpfchens beim /ʀ/ fast vollständig zugunsten von Realisierungen mittels Frikativen gewichen (z. B. in „treten“ oder „Kragen“; und s. auch Abb. 2.2).

Fazit zu Abschn. 2.2

Die Silbe ist die Grundeinheit der Artikulation. Insbesondere können einige Konsonanten nicht isoliert produziert werden. Wollen wir dennoch die Artikulation einer Silbe in kleinste Bestandteile zerlegen, so sind diese Bestandteile nicht die Laute, sondern die Sprechbewegungseinheiten (SBEs). Sprechbewegungseinheiten sind die Grundelemente der motorischen Planung und der motorischen Steuerung. Sie spezifizieren die Artikulation auf hoher Planungsebene aus der Sicht der motorischen Steuerung, nämlich auf der Ebene der auditiven Resultate. Die Aktivierung einzelner Muskeln der unterschiedlichen Artikulatoren (Lippen, Zunge, Gaumensegel ...) wird erst auf einer tieferen Ebene innerhalb der Hierarchie der motorischen Steuerung spezifiziert. Dies ermöglicht sofortige artikulatorische Kompensation im Falle von artikulatorischen oder auditiven Behinderungen. Zur Lautartikulation: Silben sind aus Konsonanten und Vokalen aufgebaut. Bei Vokalen ist die Luftströmung von der Lunge zum Mund nicht behindert. Konsonanten hingegen bilden Engstellen oder Verschlüsse im Mundraum aus. Wir unterscheiden Plosive (Verschlusslaute, z. B. /p/, /b/, /t/, /d/, /k/ und /g/), Frikative (Reibelaute, z. B. /f/ und /s/), Nasale (z. B. /m/ und /n/), Laterale (z. B. /l/), Vibranten (z. B. /r/) und weitere. Darüber hinaus können Konsonanten nach ihrem Artikulationsort unterschieden werden (labial, apikal, dorsal bzw. genauer: bilabial, labiodental, dental, alveolar, palatoalveolar, palatal, velar, uvular, pharyngeal, glottal).

2.3 Das akustische Sprachsignal

Ziel der Artikulation ist die Realisierung bestimmter zielgerichteter Artikulatorbewegungen, die wir im vorhergehenden Abschnitt als Sprechbewegungseinheiten definiert haben. Das Ziel des Zusammenspiels aller Sprechbewegungseinheiten in einer Silbe, einem Wort oder einer ganzen Äußerung ist aber nicht die Generierung der Bewegungsfolge selbst, sondern die Generierung eines daraus resultierenden akustischen Signals, welches das Kriterium erfüllen muss, dass es für den Kommunikationspartner verstehbar ist bzw. darüber hinaus auch alle vom Sprecher beabsichtigten Intentionen an den Hörer übermitteln sollte. So dient beispielsweise die Einstellung eines bestimmten Stimmlippenabstandes durch die Aryknorpel der Erzeugung oder Nichterzeugung der Phonation, also der Stimmhaftigkeit oder Stimmlosigkeit. Die Einstellung oraler Engen oder Verschlüsse dient der Realisierung bestimmter akustischer Eigenschaften bei Konsonanten und die Lage und Formung der gesamten Zunge zusammen mit der Mundöffnung der Realisierung bestimmter akustischer Eigenschaften bei Vokalen.

Betrachten wir zunächst den Fall der vokalischen Artikulation. Das akustische Basissignal entsteht aufgrund der Stimmlippenschwingung im Kehlkopf (Larynx) kurz oberhalb der Stimmritze (Glottis) und wird als **Phonationsschall** bezeichnet. Die Stimmlippen schwingen beim Sprechen in einem Frequenzbereich um 150 Hz, jedoch je nach Sprechertyp (Mann, Frau, Kind) auch tiefer oder höher. Der im Kehlkopf produzierte Phonationsschall wird durch den darüber liegenden Rachen- und Mundraum (auch als **Ansatzrohr** bezeichnet; Abb. 2.12) je nach Formung dieser Räume gefiltert und dann vom Mund abgestrahlt. In der akustisch

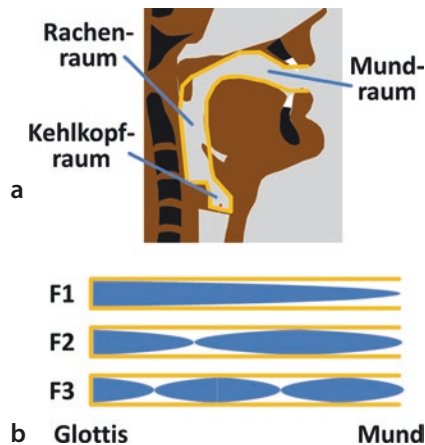


Abb. 2.12 a Quelle (kleiner roter Punkt) und Filter (Ansatzrohr, bestehend aus Larynx-, Rachen- und Mundraum, gelb umrandet); b Ausbreitung der stehenden Welle des ersten, zweiten bzw. dritten Formanten F1, F2 und F3 im Ansatzrohr. Die gelbe Umrandung im oberen und unteren Teil repräsentiert das Ansatzrohr. Im unteren Modellteil ist somit die Krümmung des Ansatzrohres nicht berücksichtigt

orientierten Literatur wird die Stimmritze und der dort generierte Phonationsschall auch als **Quelle** und das darüber liegende Ansatzrohr als **Filter** bezeichnet (vgl. Kent 1993, 1997). Die Filterwirkung entsteht dadurch, dass sich aufgrund des Phonationsschalls im Ansatzrohr stehende Schallwellen ausbilden können, welche durch die Resonanzeigenschaften des Ansatzrohres festgelegt sind (Abb. 2.12). Im Folgenden werden zunächst typische Eigenschaften des Phonationsschalls (der Quelle) beschrieben. Nachfolgend werden die akustischen Eigenschaften insbesondere die Resonanzen des Ansatzrohres (des Filters) erläutert.

2.3.1 Phonation

Die Entstehung des Phonationssignals ist bereits recht komplex. Im zeitlichen Ablauf einer Äußerung tritt die Phonation abschnittsweise auf. Nämlich immer dann, wenn ein oder mehrere stimmhafte Laute hintereinander gebildet werden müssen. Zur Gruppe der stimmhaften Laute zählen die Vokale und die stimmhaften Konsonanten wie z. B. /b/, /z/, /m/, /l/ (Abschn. 2.1). Stimmlose Laute hingegen zeichnen sich durch die Abwesenheit von Phonationsschall aus. Die Gruppe der stimmlosen Lauten besteht im Deutschen und vielen anderen Sprachen aus stimmlosen Plosiven und stimmlosen Frikativen wie z. B. /t/ oder /s/. Stimmhaftigkeit wird erzeugt, indem die **Stimmritze** durch die Einstellung der Aryknorpel locker verschlossen ist (vgl. Abb. 2.5). Die Schließung ist also nur so fest, dass die Stimmlippen aufgrund aerodynamischer Energieübertragung noch schwingen, d. h. sich schnell öffnen und schließen könnten. Der durch eine entsprechende Bewegung der Aryknorpel erzeugte *lockere Verschluss* der Stimmritze (Abb. 2.6) erlaubt ein Schwingen der Stimmlippen. Dabei werden nach Schließen der Stimmritze aufgrund eines durch Lungenaktivität (pulmonale Aktivität) aufgebauten *subglottalen Drucks* die Stimmlippen kurzzeitig auseinandergedrückt. Aufgrund des dann entstehenden Luftstroms, der die glottal Enge mit hoher Geschwindigkeit passiert, genauer gesagt aufgrund des daraus resultierenden *dynamischen Unterdrucks* in der Stimmritze zusammen mit den elastischen Rückstellkräften der Stimmlippen, wird dann die Wiederverschließung der Stimmritze eingeleitet. Dies beschreibt den Ablauf eines Zyklus einer fortlaufenden Stimmlippen-schwingung (Abb. 2.13). In Abb. 2.13 kann auch erkannt werden, dass die Schwingung des unteren Teils der Stimmlippen immer der des oberen Teils der Stimmlippen vorausseilt (vgl. Hanson et al. 2001 und vgl. Abschn. 9.3.1).

Der Phonationsvorgang wird somit durch Muskelaktivierung initiiert. Dies ist das Zusammenführen der Stimmlippen an den Aryknorpeln zu einem lockeren Verschluss und die Bereitstellung eines Luftüberdrucks durch die Lungenmuskulatur unterhalb des Kehlkopfes. Der Phonationsvorgang wird auch wieder durch eine muskulär kontrollierte an den Aryknorpeln realisierte Öffnungsbewegung der Stimmritze beendet (vgl. Abb. 2.5). Dies geschieht, wenn nach einem stimmhaften Laut oder nach einer Folge von stimmhaften Lauten ein stimmloser Laut oder eine Folge von stimmlosen Lauten produziert werden soll oder aber wenn das Äußerungsende vorliegt. Der Phonationsvorgang selbst, also das schnelle Schwingen der Stimmlippen in der Frequenz des Stimmtons (um ca. 150 Hz), wird aus aerodynamischer Energie gespeist. Dies ist die schnelle Abfolge von

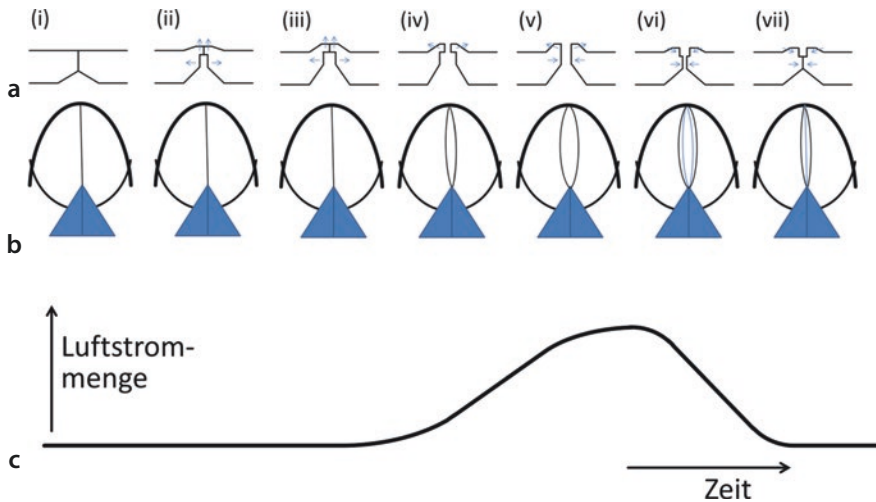


Abb. 2.13 Schwingungsvorgang der Stimmritze bei der Phonation (**a** und **b**) und der resultierende Luftstrom als Funktion der Zeit für einen Schwingungszyklus (**c**). **a** Frontalschnitt der Stimmritze (Sicht von vorne); **b** Aufsicht auf die Stimmritze in sieben verschiedenen Phasen eines glottalen Schwingungszyklus; **c** Stärke des glottalen Luftstroms als Funktion der Zeit in zeitlicher Koordination zu den darüber abgebildeten sieben Zeitpunkten. Phase (i)–(iv): Der subglottale Druck bewirkt die Öffnung der Stimmritze. Phase (v)–(vii): Die Strömungsgeschwindigkeit des durch die Stimmritze strömenden Luftstroms ist insbesondere gegen Ende der glottalen Offenphase (Phase vi) sehr hoch. Die hier auftretende hohe Strömungsgeschwindigkeit der Luft führt zu einem lokalen Unterdruck zwischen den Stimmlippen und damit zusammen mit den elastischen Rückstellkräften der Stimmritze zur erneuten Schließung der Stimmritze. Danach wiederholt sich der Vorgang

subglottalem Überdruck bei verschlossener Stimmritze einerseits mit interglottalem Unterdruck aufgrund der schnellen Luftströmung in der glottalen Enge andererseits.

Die an den Aryknorpeln bewirkte und muskulär initiierte Schließung der Stimmritze zwischen dem Übergang von stimmlosen zu stimmhaften Passagen und die Öffnung der Stimmritze im Übergang von stimmhaften zu stimmlosen Passagen verläuft im Vergleich zur Öffnung und Schließung der Stimmritze bei der Phonation vergleichsweise langsam ([Abb. 2.14](#); hier ist die schnelle phonatorische Stimmlippenschwingung blau eingezeichnet). Diese langsame Öffnungs- und Schließbewegung wird durch definierte glottale Sprechbewegungseinheiten realisiert (oegl, phgl, [Tab. 2.2](#)) und ist zeitlich exakt mit den bei der Lautbildung parallel dazu auftretenden Sprechbewegungseinheiten zur Verschuss- oder Engebildung im Mundraum koordiniert.

Somit speist sich die Stimmlippenschwingung bei der Phonation aus aerodynamischer Energie, nämlich aus dem von den Lungen nach dem Einatmen erzeugten Luftüberdruck unterhalb der Stimmritze und der daraus resultierenden Luftströmung, wenn die Stimmritze geöffnet ist. Die Lunge muss somit auch immer genügend Luft zum Ausströmen entlang der glottalen Engstelle und des Rachen-, Mund- und ggf. auch Nasenraums zur Verfügung stellen. Deshalb ist die Dauer einer Äußerung durch die maximal mögliche Ausatemlänge

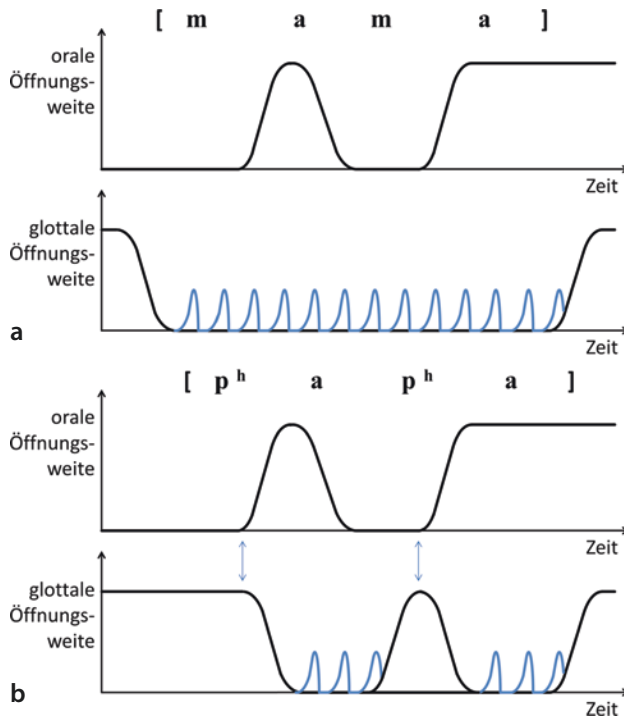
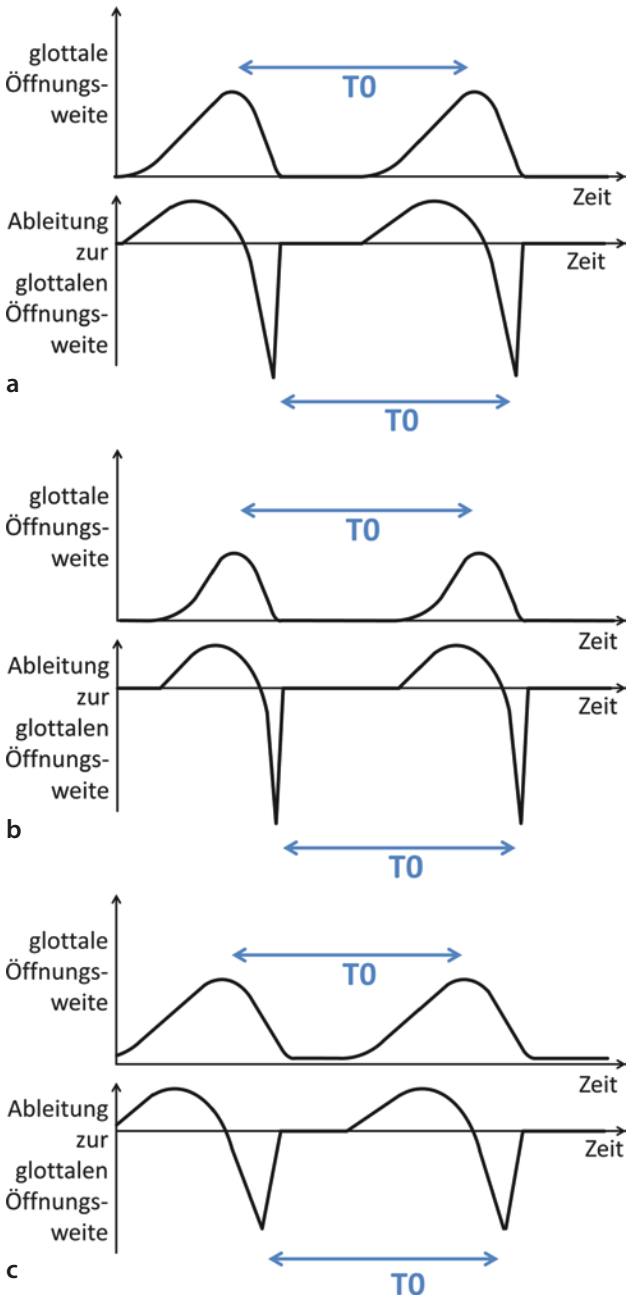


Abb. 2.14 Langsame (artikulatorische) und schnelle (phonatorische) Öffnung und Schließung der Stimmritze und deren zeitliche Koordinierung mit oralen Sprechbewegungseinheiten bei der Produktion der Wörter „Mama“ (a) und „Papa“ (b). Neben der Öffnung der Stimmritze (Weitstellung der Stimmlippen) zur Realisierung stimmloser Laute ist hier auch die Öffnung der Stimmlippen am Anfang und Ende der Äußerung, also die Atemstellung, zu erkennen. Die schnelle phonatorische Öffnungs- und Schließbewegung ist blau bzw. grau eingezeichnet. Zur Erreichung maximaler Behauchung bei stimmlosen Plosiven muss die Stimmritze genau zum Zeitpunkt der oralen Verschlusslösung maximal geöffnet sein (s. blaue vertikale Doppelpfeile)

beim Sprechen begrenzt. Darüber hinaus muss je nach *Sprechstärke*, also nach der momentanen mittleren Lautstärke, mit der eine Person spricht (d. h. je nachdem, ob diese beispielsweise in einem kleinen Raum nur zu einer weiteren Person oder in einem großen Raum zu einem großen Zuhörerkreis sprechen muss), der subglottale Druck entsprechend für die gesamte geplante Äußerung voreingestellt werden. Diese Voreinstellung des Lungen- bzw. des subglottalen Drucks ist hoch bei lautem, aber niedrig bei leisem Sprechen. Während einer Äußerung selbst bleibt dieser Druck auf dem voreingestellten Niveau annähernd konstant. Lediglich zum Ende der Äußerung hin kann er ein wenig abfallen.

Der Phonationsschall resultiert aus der während der Offenphase der glottalen Phonationsschwingung in das Ansatzrohr abgegebenen Luftimpuls (Abb. 2.15). Am besten wird die akustische Anregung des Ansatzrohres während der Phonation durch die *zeitliche Ableitung* des glottalen Luftstroms beschrieben (Abb. 2.15). Wir erkennen, dass hier genau im Moment des Verschließens der Stimmritze die stärkste (negative) Amplitude

Abb. 2.15 Darstellung der Luftstrommenge (oben) und ihrer zeitlichen Ableitung (unten) während zweier glottalen Schwingungsperioden bei (a) normaler Phonation, (b) gepresster Phonation und (c) behauchter Phonation. Wir unterscheiden innerhalb einer glottalen Schwingungsperiode die glottale Offenphase und die Verschlussphase. Bei geöffneter Stimmritze (Offenphase) kann noch zwischen Öffnungs- und Verschlussphase unterschieden werden. Die Dauer einer Phonationschwingung ist mit T_0 (Periodendauer) gekennzeichnet: Grundfrequenz des Stimmtons: $F_0 = 1 / T_0$



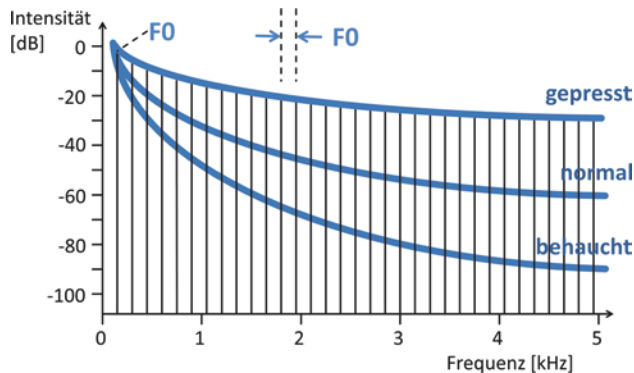


Abb. 2.16 Linienspektrum (Grund- und Obertöne) einer Phonationsschwingung von 150 Hz bei gepresster (harter), normaler und behauchter (weicher) Stimmqualität. Die dicken blauen bzw. grauen Linien markieren *spektrale Hüllkurven* und geben damit die Maximalamplituden der Teiltöne bei verschiedenen Stimmqualitäten an (monotoner spektraler Abfall). Der erste senkrechte Strich des Linienspektrums repräsentiert den Grundton, hier mit $F_0 = 150$ Hz, die weiteren senkrechten Striche die Obertöne $f_n = n \cdot F_0$ mit $n = 1, 2, 3, \dots$. Der Abstand zwischen zwei benachbarten Obertönen beträgt immer genau F_0 . Als Teiltöne werden Grundton plus Obertöne bezeichnet (erster Teilton = Grundton; zweiter Teilton = erster Oberton usw.)

auftritt. Zu dem Zeitpunkt einer glottalen Verschlussbildung wird das Ansatzrohr somit akustisch maximal angeregt. Das **Frequenzspektrum** des Phonationsschalls ist monoton abfallend, und zwar mit einer Rate von ca. 12 dB/Oktave (Abb. 2.16). Das bedeutet, dass sich die Leistung des Phonationsschalls pro Oktave um 12 dB (also auf ca. ein Sechzehntel) verringert. Dieser spektrale Abfall erscheint stark. Dennoch kann unser Gehör den Phonationsschall und die daraus resultierende akustische Anregung des Ansatzrohres noch gut bis ca. 3 kHz, somit also über ca. 5 Oktaven (im Fall von $F_0 = 150$ Hz: 150 Hz–300 Hz–600 Hz–1200 Hz–2400 Hz–4800 Hz), wahrnehmen. Dies resultiert aus der Tatsache, dass Lautstärke näherungsweise proportional zum Logarithmus der Schallintensität wahrgenommen wird (Abschn. 3.1). Wir werden sehen, dass dies dem Frequenzbereich der für die Sprachverarbeitung wichtigen ersten drei Formanten entspricht.

Im Rahmen dieses Buches kann nicht detailliert auf den Zusammenhang zwischen der Zeitdarstellung, also dem Oszillogramm einer (Schall-)Schwingung, und seinem Spektrum, also der Aufgliederung der Schwingung nach seinen Frequenzanteilen, eingegangen werden. Es sei hier nur so viel gesagt: Eine periodische Schwingung der Frequenz f (wie z. B. die glottale Phonationsschwingung) zerfällt immer in Grundton (Frequenz $f_1 = f$) und Obertöne bzw. Teiltöne (Frequenz $f_n = n \cdot f$, $n = 2, 3, 4, \dots$). Je nachdem, wie die periodische Grundschwingung geformt ist, können Grundton und Obertöne unterschiedlich intensitätsstark sein. Bei harter Stimme (auch *gepresste Stimme* genannt) sind die Obertöne intensitätsstärker als bei normaler Stimme. Dies ist durch einen vergleichsweise langsamen Abfall der Intensität der Obertöne mit 6 dB/Oktave gekennzeichnet. Bei einer weichen Stimme (auch als *behauchte Stimme* bezeichnet) kann der Intensitätsabfall der Obertöne vergleichsweise steil ausfallen (bis zu 18 dB/Oktave). Bei der behauchten Stimme wären also hohe

Frequenzanteile im Stimmklangspektrum nur gering angeregt. Da bei behauchter Stimme die Glottis neben dem Phonationsschall aber auch hochfrequentes Rauschen (glottales Rauschen) erzeugt, kann dieses Rauschen aufgrund der niedrigen Intensität der Obertöne des Phonationsschalls dann gut wahrgenommen werden. Dies führt dazu, dass die weiche Stimme oft auch als behauchte Stimme bezeichnet wird. Die akustische Erzeugung von Rauschen an der Glottis wie auch an anderen Engstellen im Ansatzrohr wird weiter unten in diesem Kapitel diskutiert.

Es sei noch angemerkt, dass der Grundton der Sprech- oder Singstimme auch häufig mit F_0 angegeben wird, um diesen Grundton in eine Reihe mit den Frequenzen der Formanten F_1 , F_2 , F_3 , ... zu setzen. Es ist aber so, dass Formanten und Grundton bzw. die daraus resultierenden Obertöne ganz unterschiedlichen Ursprungs sind (zu den Formanten s. [Abschn. 2.3.2](#)), auch wenn diese in überlappenden Frequenzbereichen auftreten.

Glottales Rauschen bedeutet, dass nun neben der periodischen glottalen Schwingung, und damit neben dem zugehörigen periodischen Klangsignal, an der Stimmritze insbesondere im Bereich der Aryknorpel auch noch ein Rauschsignal generiert wird. Darüber hinaus können Rauschsignale auch an Engstellen im Rachen- und Mundraum entstehen. An der Glottis und im Rachen- oder Mundraum generierte Rauschsignale (nichtperiodisch) haben in der Regel Frequenzanteile, die im Unterschied zum Linienspektrum kontinuierlich über den Frequenzbereich von ca. 2 bis ca. 20 kHz verteilt sind. Rauschsignale sind im Zeitbereich nichtperiodisch. Somit wird bei behauchter Stimme der Frequenzbereich oberhalb von ca. 2 kHz von glottalem Rauschen gefüllt (gleichförmig geschwärzter Bereich in **Spektrogrammen**, vgl. [Abb. 2.17](#)), während der darunterliegende Frequenzbereich eher von den Teiltönen des periodischen Phonationssignals dominiert wird. Ein

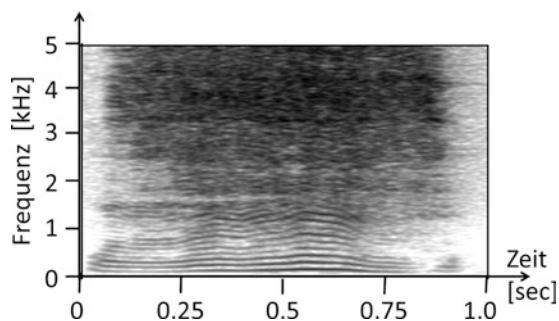


Abb. 2.17 Spektrogramm eines lang angehaltenen stimmhaften Frikativs /z/. y-Achse: Frequenz von 0 Hz bis 5000 Hz, x-Achse: Zeitintervall (0 sec bis 1 sec), Schwärzungsgrad: Intensität des akustischen Signals. Der Grundton (ca. 100 Hz) und ca. 10 Obertöne sind erkennbar. Der Grundton schwankt während der Produktion (Grundtonmaximum bei ca. 0,6 sec). Über dem Grundton und den Obertönen ist ein kontinuierliches Rauschsignal (von ca. 1000 Hz bis 5000 Hz zu erkennen). Im Unterschied zu den unten gezeigten Spektrogrammen (z. B. [Abb. 2.22](#)) liegt hier eine hohe Frequenzauflösung (bei gleichzeitiger schlechter Zeitauflösung) vor. Damit sind hier Grund- und Obertöne erkennbar, während bei geringerer Frequenzauflösung gekoppelt mit höherer Zeitauflösung eher die Formanten sowie einzelne glottale Schwingungsperioden im Zeitbereich erkennbar werden (vgl. [Abb. 2.22](#))

ähnliches Klangspektrum mit Teiltönen bis ca. 2 kHz und Rauschen darüber tritt auch bei stimmhaften Frikativlauten auf (Abb. 2.17). Dort ist allerdings die Rauschquelle dann nicht mehr an der Glottis, sondern liegt im Rachen- oder Mundraum, nämlich im Bereich des weichen Gaumens für /χ/, im Bereich des harten Gaumens für /j/, im Bereich des Übergangs von hartem Gaumen zum Zahndamm für /z/, im Bereich des Zahndamms für /z/ und im Bereich von Schneidezähnen und Unterlippe für das /v/.

2.3.2 Formanten

Der an der Stimmritze entstehende Phonationsschall breitet sich im Ansatzrohr, also im Rachen-, Mund- und bei Nasallauten auch im Nasenraum aus. Aus akustischer Sicht bedeutet dies aber nicht, dass sich die mit der Frequenz des Stimmtons (Grundfrequenz $F_0 = 1 / T_0$, Abb. 2.15) oberhalb der Stimmritze entstehende Folge an Luftimpulsen nur einmalig in Richtung Mund wandert. Eine solche Vorstellung würde vielleicht passen, wenn kein Ansatzrohr, also kein Rachen- und Mundraum oberhalb der Stimmritze, anzutreffen wäre, sondern wenn sich der Stimmtone direkt in einem Raum ausbreiten könnte. Was tatsächlich aber passiert, ist die Ausbreitung *stehender Wellen* im Ansatzrohr aufgrund der Reflexion des Schallsignals an Glottis und Mund (Abb. 2.12). Intuitiv ist die Entstehung stehender akustischer Wellen im Fall des geöffneten Mundes schwer zu verstehen. Am *geschlossenen Ende*, d. h. an der Glottis, wird die Schallwelle während der glottalen Verschlussphase vollständig reflektiert. Am *offenen Ende* des Ansatzrohres, d. h. am Mund, wird nur ein Teil der Schallenergie nach außen weitergegeben. Ein Großteil wird auch hier reflektiert. Die im Frequenzbereich von 100 Hz bis 3000 Hz und auch noch höher liegenden Frequenzanteile des Stimmtons werden somit zwischen geschlossenem und offenem Ende des Ansatzrohres hin und her geworfen und formen so stehende Schallwellen. Hintergrund zur Reflexion an beiden Enden, also am geschlossen wie am offenen Ende des Ansatzrohres, ist, dass sich an beiden Enden die Querschnittsfläche A des Ansatzrohres abrupt ändert. An der Glottis auf $A = 0 \text{ cm}^2$ auf einen endlichen Wert im cm^2 -Bereich und am Mund von einem endlichen Wert auf unendlich, da vor dem Mund keine Begrenzung mehr für die Schallwelle vorliegt. In beiden Fällen führt dies zu einer enormen akustischen Impedanzsprung.

Die sich im Ansatzrohr z. B. bei der Vokalproduktion ausbildenden stehenden Wellen repräsentieren die Resonanzen des Ansatzrohres und damit die Resonanzen von Pharynx-, Mund- und evtl. Nasenraum. Jede Resonanz zeigt ein akustisches Energiemaximum an einer definierten Frequenz, nämlich der *Resonanzfrequenz*. Diese Resonanzen des Ansatzrohres werden auch als *Formanten* bezeichnet. Am geschlossenen Ende des Ansatzrohres (an der Glottis) können sich besonders gut Druckschwankungen (eine Abfolge von Druckmaxima und -minima) ausbilden, während am offenen Ende (Mund) die stehende Welle näherungsweise mit dem schwankungsfreien Umgebungsdruck übereinstimmen muss. An der Glottis kann sich somit ein *Druckbauch* ausbilden, während am Mund immer ein *Druckknoten* auftreten muss (vgl. Abb. 2.12). Somit können sich im Ansatzrohr diejenigen

stehenden Wellen etablieren, deren Wellenlänge bei $1/4$, $3/4$, $5/4$ usw. der Länge des Ansatzrohres (Abb. 2.12) liegen.

Dass der Mund bei der Vokalproduktion als offenes Ende angesehen werden kann, ist aufgrund der bei Vokalen vorliegenden Mundöffnung offensichtlich. Warum aber kann die Stimmritze als geschlossenes Ende betrachtet werden? Hier sei daran erinnert, dass die Hauptanregung des Ansatzrohres durch einen Luftstromimpuls zum Zeitpunkt der glottalen Verschlussbildung erfolgt, also zum Zeitpunkt des absoluten Maximums bei der zeitlichen Ableitung des glottalen Volumenstroms (vgl. Abb. 2.15). Zu diesem Zeitpunkt und danach ist die Stimmritze innerhalb jedes glottalen Schwingungszyklus zunächst geschlossen. In diesem Zeitintervall des glottalen Verschlusses findet somit die Bestärkung der stehenden Wellen (der Formanten) durch den an der Glottis produzierten eingegangenen Luftstromimpuls statt. Während der nun während des glottalen Schwingungszyklus folgenden glottalen Öffnung klingen die stehenden Wellen dagegen leicht ab. Dies ist der Hauptfaktor für die Dämpfung der Formanten.

Das akustische Resultat einer stehenden Welle, nämlich die Bündelung der akustischen Energie im Bereich einer Resonanzfrequenz, wird auch als **Formant** bezeichnet. Nach Abb. 2.12 können wir die Wellenlängen der Formanten L aus der Länge l des Ansatzrohres bestimmen. L ist viermal länger als das Ansatzrohr für F_1 . Erst für F_2 und F_3 entspricht die Wellenlänge der Formanten, die je nach Person und je nach gerade realisiertem Laut zwischen ca. 15 und ca. 20 cm liegt, ungefähr der Länge des Ansatzrohres. Der Zusammenhang zwischen L für eine stehende Welle und der zugehörigen Frequenz $f = 1/T$ ist über die Schallgeschwindigkeit c gegeben:

$$c = L / T \quad (2.3a)$$

und damit:

$$f = 1 / T \quad (2.3b)$$

und

$$f = c / L \quad (2.3c)$$

Setzt man die Länge des Ansatzrohres mit $l = 17,5$ cm und die Schallgeschwindigkeit mit 350 m/s an, so ergeben sich für die ersten drei Formanten, also mit $L = 4 \cdot l$ für F_1 , mit $L = (4 \cdot l)/3$ und mit $L = (4 \cdot l)/5$ für F_3 , die Frequenzen $F_1 = 500$ Hz, $F_2 = 1500$ Hz und $F_3 = 2500$ Hz.

Bei einem querschnittskonstanten Ansatzrohr, wie es näherungsweise beim Neutralvokal /ə/ (z. B. am Ende des Wortes „Bitte“ oder lang gehalten als Pausenlaut) entsteht, tritt näherungsweise das *Formantmuster* $F_1 = 500$ Hz, $F_2 = 1500$ Hz und $F_3 = 2500$ Hz auf. Die im Bereich dieser Resonanzfrequenzen liegenden Teiltöne werden angehoben, während die zwischen den Formanten, hier also die um 1000 Hz und um 2000 Hz liegenden Teiltöne abgeschwächt werden (Abb. 2.18). Das Ansatzrohr hat also aufgrund der Formanten eine spezielle Frequenzcharakteristik, die von Akustikern auch als *Übertragungsfunktion* oder *Filterfunktion* bezeichnet wird (vgl. Abb. 2.18c).

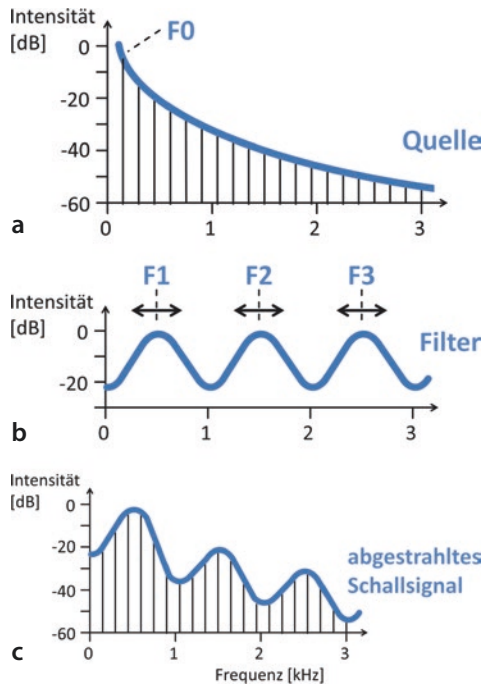


Abb. 2.18 Das Quellsignal (Phonationssignal, **a**) wird durch die Übertragungsfunktion des Ansatzrohres (Filter, **b**) „überformt“ und dann als hörbares Sprachsignal am Mund abgestrahlt (**c**). Es treten Formantfrequenzen bei $F_1 = 500$ Hz, $F_2 = 1500$ Hz und $F_3 = 2500$ Hz auf, die aufgrund der Form- und Längenänderung des Ansatzrohres während der Artikulation um diese Zentralfrequenzen herum variieren können (s. Pfeile oberhalb der Filter- bzw. Übertragungsfunktion Abbildung **b**)

Kommen wir zurück zum Quelle-Filter-Modell, welches wir am Anfang dieses Kapitels eingeführt haben. Das Ansatzrohr fungiert also als Filter und hebt diejenigen Teiltöne des Phonationssignals hervor, die im Frequenzbereich eines Formanten liegen. Dabei muss die Zentralfrequenz eines Formanten (also eine Resonanzfrequenz des Ansatzrohres) nicht mit der Frequenz eines Teiltons übereinstimmen. Da Länge und Form des Ansatzrohres bei gesprochener Sprache fortlaufend variieren, variieren auch die Formanten ständig um ihre mittlere Frequenzlage herum (siehe Pfeile im Bereich der Formanten F_1 , F_2 , und F_3 der Übertragungsfunktion, also des Filters in [Abb. 2.13c](#)). Beim Sprechen treten Längen- und Formänderung des Ansatzrohres gleichzeitig auf. Betrachten wir die Artikulation des /i:/ und des /a:/ ([Abb. 2.19](#)), so erkennen wir, dass beim /i:/ der vordere Bereich des Ansatzrohres verengt und der hintere Teil geweitet ist. Beim /u:/ erkennen wir sogar zwei Engbildungen, nämlich an den Lippen und im Bereich des Gaumensegels, während sowohl der Mund- wie auch der Rachenraum einen Hohlraum bilden. Beim /u:/ kommt auch noch hinzu, dass das Ansatzrohr aufgrund der Lippenrundung und der Absenkung des Kehlkopfes deutlich länger ist als beispielsweise beim /i:/ oder /a:/ ([Abb. 2.19](#) und [2.20](#)).

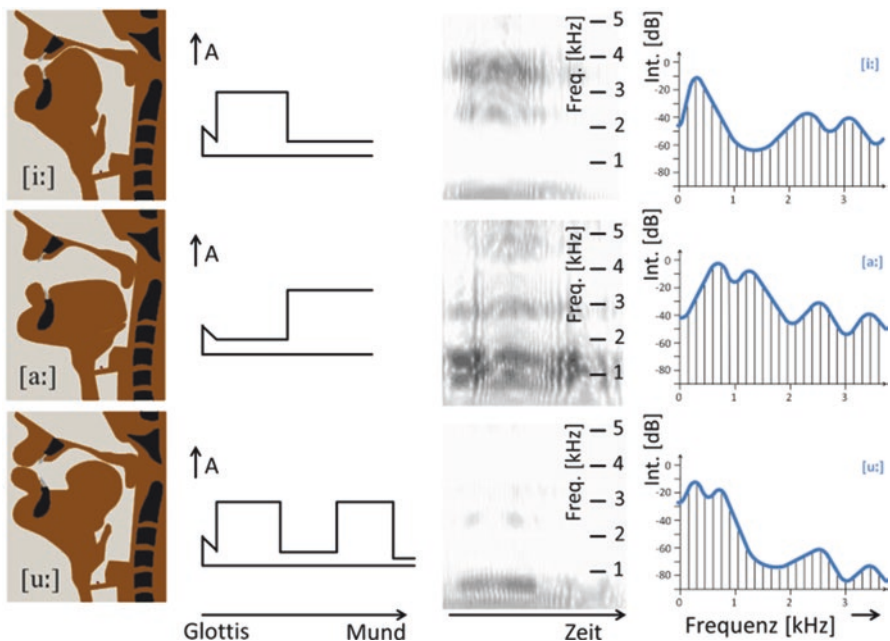


Abb. 2.19 Mediosagittalschnitt (links), stilisierte Querschnitts- oder Areafunktion (Querschnittsfläche A) von Stimmritze bis Mund (Mitte links), Spektrogramm (Mitte rechts: Frequenz als Funktion der Zeit) und Kurzzeitspektrum (rechts: Intensität als Funktion der Frequenz) für die drei Vokale [i:], [a:], und [u:]

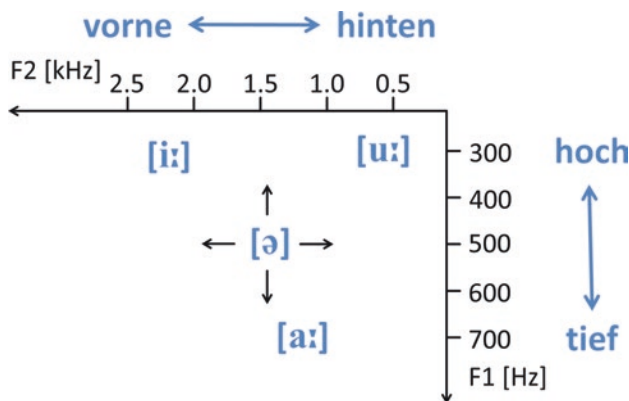


Abb. 2.20 Positionierung der drei Eckvokale und des Schwa-Lauts im F1-F2-Vokalraum. Zusätzlich sind artikulatorische Kategorien der Vokale angegeben: Tiefe Vokale zeigen einen hohen F1-Wert, hohe Vokale hingegen einen niedrigen F1-Wert. Vordere Vokale zeigen einen hohen F2-Wert, hintere (und gerundete) Vokale hingegen einen niedrigen F2-Wert

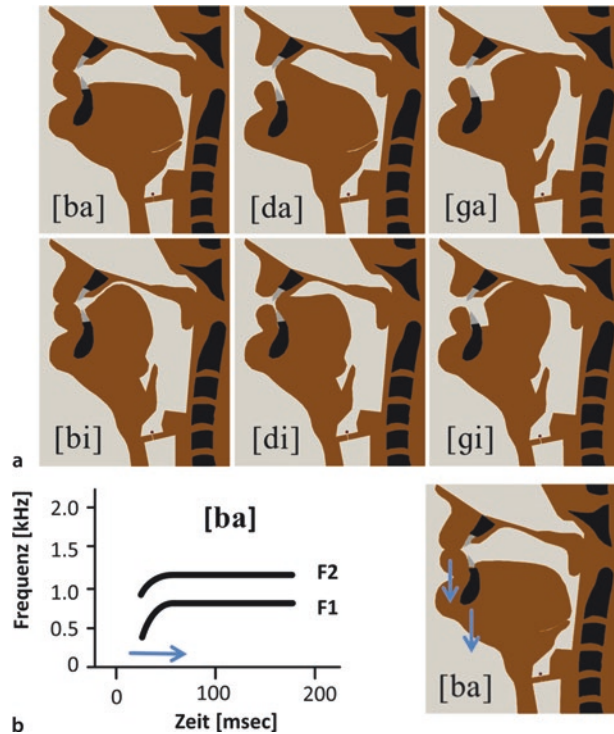
Bevor wir die Akustik der Konsonanten beschreiben, sei zur Vokalartikulation abschließend noch bemerkt, dass der oft herangezogene Vergleich, dass die Stimmritze dem Mundstück eines Blasinstrumentes und das Ansatzrohr dem gesamten Rest des Blasinstrumentes entspricht, nicht richtig ist. Bei Blasinstrumenten, egal ob Blech- oder Holzblasinstrument, wird die Tonhöhe nicht am Mundstück alleine ausgebildet, während die Stimmtonhöhe in erster Näherung nur durch das Stimmlippensystem determiniert wird. Bei Musikinstrumenten determiniert genau die Länge des folgenden Teils, also des Rohres, die Grundtonhöhe. Also wird die Länge des Rohres bei Musikinstrumenten nicht zur Klangänderung wie bei der Vokalproduktion durch den Menschen genutzt, sondern zur Einstellung der Tonhöhe. Dabei wird die effektive akustische Länge beim Holzblasinstrument durch die Lage und Anzahl der verschlossenen Bohrungen bzw. der geschlossenen Klappen und beim Blechblasinstrument durch die Stellung der Ventile bzw. des Zuges und damit auch durch die Gesamtlänge des Rohres determiniert. Beim Sprechen ist die Situation komplexer. Hier ist neben der Variation der Tonhöhe durch die Stimmlippen auch noch die Variation des Vokalklanges durch die Variation der Form und Länge des Ansatzrohres möglich. Eine Klangvariation beim Musikinstrument ist verglichen mit der Vokalklangänderung beim Menschen nur eingeschränkt möglich.

2.3.3 Formanttransitionen und sekundäre Schallquellen

Bei der Artikulation der Konsonanten (vgl. [Abschn. 2.2](#)) werden im Mundraum Engen oder Verschlüsse, sogenannte *Konstriktionen*, gebildet. Die Stimmritze ist bei stimmlosen Konsonanten geöffnet und bei stimmhaften Konsonanten zur Phonation locker verschlossen. Gleiches gilt für die velopharyngeale Pforte zur Ausbildung nasaler bzw. nicht nasaler Laute. Bei der Vokalartikulation spielt die *Gesamtform* des Ansatzrohres eine wichtige Rolle. Das *vokalische Formantmuster*, d. h. die im zeitlichen Zentrum des Vokals vorliegende Formantfrequenzverteilung F1-F2-F3, hängt sowohl von der Länge als auch von der detaillierten Formung des Ansatzrohres ab. Somit ist die Positionierung aller Artikulatoren (Lippen, Zunge, Gaumensegel und Kehlkopf) für die artikulatorische wie akustische Realisierung jedes Vokals maßgeblich. Konsonantische Enge- oder Verschlussbildungen sind *lokal*, da sie entweder mittels der Lippen, der Zungenspitze oder dem Zungenrücken gebildet werden. Somit fällt die Ansatzrohrform bei Konsonanten in unterschiedlichen vokalischen Kontexten (z. B. /di/ vs. /da/) bis auf die Enge- oder Verschlussbildung auch sehr unterschiedlich aus ([Abb. 2.21](#)). Dementsprechend sind auch die akustischen Louteigenschaften der Konsonanten nicht so einfach definierbar wie die der Vokale.

Während Vokale in Isolation statisch, also ohne jede Bewegung der Artikulatoren, produziert werden können (z. B. ein lang ausgehaltenes /a:/), funktioniert dies bei Konsonanten nur eingeschränkt. Zwar können auch viele Konsonanten in Isolation gesprochen werden (z. B. /n:/, /l:/, /r:/, /s:/, ...), die Verständlichkeit dieser Laute steigt aber enorm, wenn sie nicht in Isolation, sondern im vokalischen Kontext (CV oder CVC; z. B. beim Buchstabieren: /be:/ oder /ʔen/) realisiert werden. Dies resultiert aus der Tatsache, dass

Abb. 2.21 **a** Mediosagittalschnitte stimmhafter Plosive während ihrer Verschlussphase in unterschiedlichem Vokalkontext. **b** Zusammenhang von Formanttransitionen (stilisiert) und Artikulationsbewegungen für [ba]



dadurch die *Bewegung* des Enge- oder verschlussbildenden Artikulators mit in die Produktion des akustischen Signals einbezogen wird. Insbesondere kann man dies leicht bei der Produktion der Nasale /m/ und /n/ erkennen, die isoliert gesprochen eher verwechselt werden als im vokalischen Kontext. Nicht möglich ist die Artikulation von Plosiven in Isolation. Hier ist immer insbesondere die Lösung der Verschlussbildung in den Vokal hinein wichtig. Und es sind insbesondere die mit dieser Verschlusslösung verbundenen akustischen Korrelate, die erst die Unterscheidung des Artikulationsortes von Plosivlauten ermöglichen. Somit spielt bei Konsonanten die *Artikulationsbewegung* der Enge- oder der Verschlussbildenden Artikulatoren in die Konstriktion hinein und aus ihr heraus eine wesentliche Rolle (vgl. Abschn. 3.3.1). Diese Bewegung der Artikulatoren und die daraus folgende Änderung der Formung des Ansatzrohres lässt sich direkt aus den Änderungen der Frequenzverläufe der Formanten über die Zeit, d. h. aus den *Formanttransitionen* ableiten (Abb. 2.22). Etwas überspitzt kann man formulieren: So wie ich Hand-Arm-Bewegungen visuell wahrnehmen kann, so kann ich Artikulatorbewegungen (z. B. Zungenbewegungen) anhand der Formanttransitionen auditiv wahrnehmen.

Es sei an dieser Stelle angemerkt, dass wir bei der perceptiven Analyse eines gesprochenen Wortes immer eine durch einen bestimmten Sprecher realisierte an einem bestimmten Zeitpunkt gemachte phonetische Realisierung des Wortes bzw. der Äußerung analysieren.

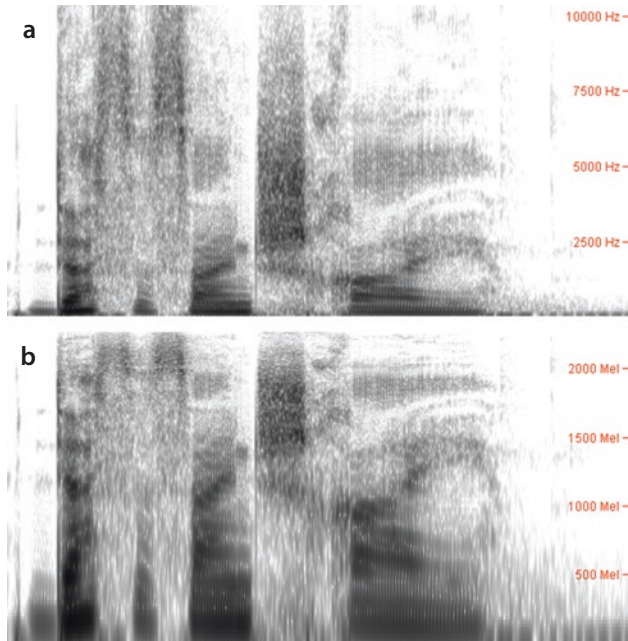
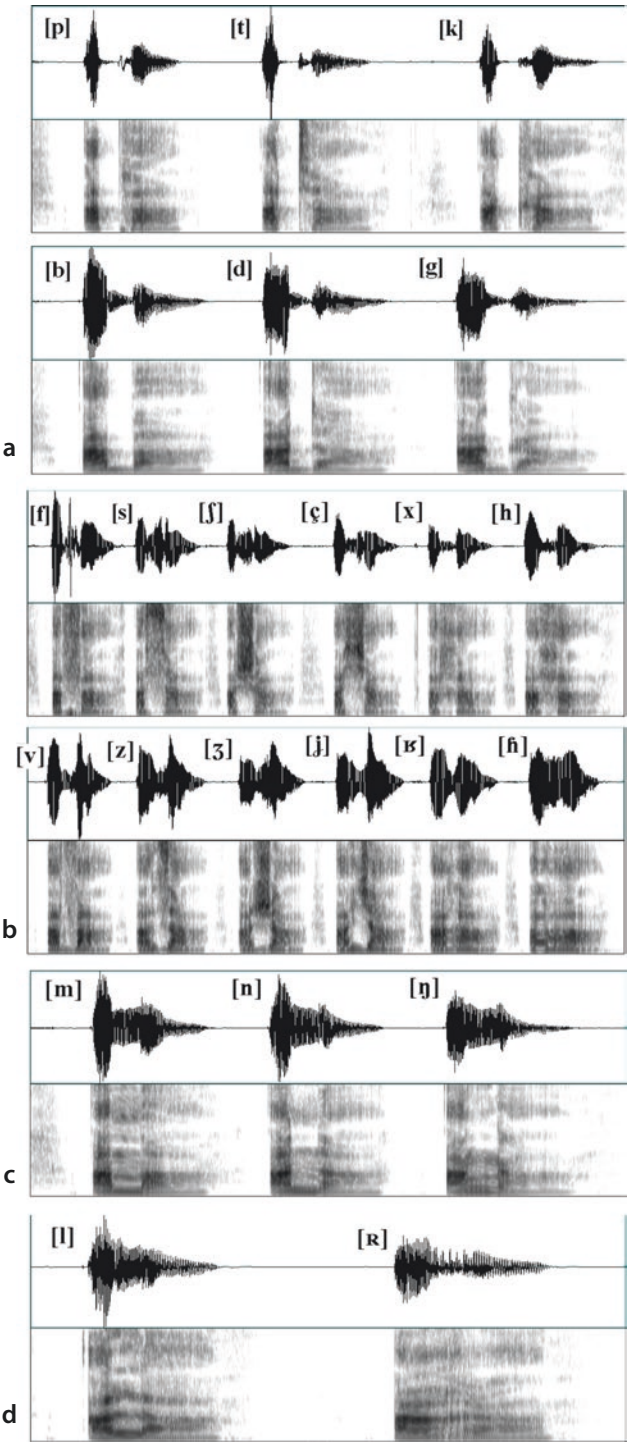


Abb. 2.22 a Spektrogramm des Satzes „Das is(t) ein Schrei“ ([dasIsaInʃrai]). Im Spektrogramm stellt die x-Achse die Zeit, die y-Achse die Frequenz in Hz (**a** linear und **b** logarithmisch; zur Mel-Skala s. Abschn. 3.1.1) und der Schwärzungsgrad die Intensität des Schallsignals dar. Die Verläufe des ersten, zweiten und dritten (in Ansätzen auch des vierten und fünften Formanten) über die Zeit sind gut erkennbar. In den stimmhaften Bereichen kennzeichnen die vertikalen Linien jeweils eine glottale Periode (Zeitpunkte der glottalen Verschießung und damit der maximalen akustischen Anregung des Sprechtraktes). Friktionsrauschen im Fall der Frikativlaute ist ebenfalls gut erkennbar. In der logarithmischen Frequenzskalierung tritt die Formantbewegung von F1 und F2 besser hervor als in der linearen Frequenzskalierung

Um der phonologisch-phonetischen Tradition zu folgen, schreiben wir die Lauttranskription von Realisierungen in eckige Klammern ([]: phonetische Transkription) und unterscheiden diese damit von der phonologischen Lautfolge (von Schrägstrichen umgeben: //). Phonologische Symbole repräsentieren immer die sprachspezifisch bedeutungsunterscheidende Ebene.

Analysiert man auf der akustischen Ebene verschiedene Konsonanten im vokalischen Kontext, also [C] im Kontext [VCV], so erscheinen die Bewegungen der enge- oder verschlussbildenden Artikulatoren als *Formanttransitionen* (Abb. 2.21–2.23). Zusätzlich können auch Geräusche durch die Anregung einer sekundären Schallquelle hinzutreten. Dies sind Geräuschbildungen aufgrund von Engebildungen im Mundraum. In Abb. 2.23 sind die wesentlichen akustischen Eigenschaften verschiedener Konsonanttypen im [VCV]-Kontext mit [V] = [e] in einer Folge von Spektrogrammen zusammengestellt. Im jeweiligen Vokalbereich sind die Formanten gut erkennbar. Zum Konsonanten hin sehen

Abb. 2.23 Oszillogramme und darunter jeweils die zugehörigen Spektrogramme für verschiedene Lauttypen bzw. Lautklassen. **a** Plosive (stimmhaft und stimmlos), **b** Frikative (stimmhaft und stimmlos), **c** Nasale (stimmhaft) und **d** Lateral und Vibrant (stimmhaft). Die Zeitachse ist bei den Frikativen am stärksten gestaucht. Alle Konsonanten [C] wurden im [VCV]-Kontext mit [C] = [e] realisiert



wir Formanttransitionen. Zwischen den Vokalen erkennen wir bei Plosiven einen *Bereich mit wenig Schallenergie* aufgrund der Verschlussbildung im Mundraum, bei Frikativen einen *Bereich mit Rauschenergie* (graue Bereiche bei höheren Frequenzen) aufgrund der Engebildung, bei Nasalen sehr *intensitätsschwache Formanten* mit *diskontinuierlichem Übergang* zu den Formanten der umliegenden Vokale und bei Lateral und r-Laut (Vibrant) ebenfalls intensitätsschwache Formanten, die hier aber in kontinuierlichem Übergang zu den Vokalformanten stehen. Darüber hinaus ist zu erkennen, dass die Formanttransitionen vor und nach Enge- bzw. Verschlussbildung, also vom vorhergehenden Vokal in den Konsonanten hinein und vom Konsonanten in den nachfolgenden Vokal hinein, für jeden Artikulationsort bei allen Konsonantarten (Plosiv bis Lateral) für jeden Artikulationsort nahezu identisch sind, da sie eben genau den Artikulationsort perzeptiv kodieren (vgl. [Abschn. 3.3.1](#)).

Die Frequenzbänder des bei Frikativen im Sprechtrakt erzeugten **Rauschens** ([Abb. 2.23](#)) variieren von Artikulationsstelle zu Artikulationsstelle und stellen somit ebenfalls ein akustisches Merkmal des Artikulationsortes dar. So liegt der Frequenzbereich des /s/ höher als der des /ʃ/, wie wir auch bereits vom Hören her vermuten können. Ein vergleichbares, aber viel kürzer andauerndes Rauschsignal finden wir bei den stimmlosen Plosivlauten zum Zeitpunkt der Verschlusslösung. Hier entsteht, wie wir auch direkt hören können, nach Lösen des Verschlusses ein kurzzeitiges im Mundraum generiertes Schallsignal („Rauschimpuls“ oder *noise burst*), das beispielsweise im Fall des /t/ lauter und höher klingt als im Fall des /k/.

Die Transition des ersten Formanten ist vor allem bei stimmhaften Konsonanten sehr intensitätsstark und damit im Spektrogramm gut erkennbar (hoher Schwärzungsgrad). Bei stimmlosen Obstruenten (Plosiven und Frikativen) hingegen setzt die Stimme erst später nach der Verschlusslösung ein. Hier ist die Stimmritze zum Zeitpunkt der konsonantischen Verschlusslösung meist noch weit geöffnet, und es wird eine gewisse Zeit benötigt, um die Stimmritze mittels der Aryknorpelbewegung zu schließen und dadurch Phonation zu erlauben. Somit können bei stimmlosen Lauten die Formanttransitionen nicht direkt nach der oralen Verschlusslösung bzw. Lösung der Enge mittels phonatorischer Schallenergie „gefüllt“ werden. Dennoch sind bei stimmlosen Lauten die Transitionen höherer Formanten (F2 und F3) direkt mit Schallenergie gefüllt und damit hörbar, da diese Formanten direkt durch sekundäre Schallquellen akustisch angeregt werden.

Bei allen stimmhaften Konsonanten lässt sich darüber hinaus während der Verschluss- oder Engebildung (Konstriktionsbildung) ein horizontaler akustischer Energiebereich bei sehr tiefen Frequenzen, d. h. im Bereich der Grundfrequenz F0, also nahe der Nulllinie des Spektrogramms erkennen. Der Grundton ist damit während der Enge- oder Verschlussphase als durchgehender Stimmtön hörbar. Dies können wir im Selbstversuch testen, indem wir bei geschlossenem Mund phonieren. Die tiefsten Teiltöne sind trotz verschlossenen Mundraums hörbar.

Im Fall des Laterals und des Vibranten sehen wir einen vokalähnlichen kontinuierlichen Verlauf der Formanten, allerdings energieschwächer. Die Eingruppierung von Approximanten zwischen Konsonanten und Vokalen lässt sich somit auch akustisch gut

untermauern. Die konsonantische Enge wird kaum ausgebildet. Es findet nur die Bewegung in diese Enge hinein und dann sofort wieder aus ihr heraus statt, ohne eine starke konsonantische Enge auszubilden. Das führt dazu, dass die Formanten von Vokal zu Vokal durch den dazwischenliegenden Approximanten nicht unterbrochen, sondern nur geschwächt werden. Approximanten sind in erster Linie dynamische Laute. Wie die Plosivlaute definieren sie sich insbesondere über ihre akustischen (und damit auch perzeptiv relevanten) Merkmale und damit über die Bewegungen der die konsonantische Enge ausbildenden Artikulatoren.

Abschließend sei deshalb nochmals betont, wie wichtig insbesondere die *zeitliche Änderung der Formantfrequenzen* bzw. auch aller anderen akustischen Merkmale bei der Sprachproduktion ist. Dies werden wir auch aus perzeptiver Sicht noch einmal ausführlich diskutieren: Lebewesen sind darauf trainiert, *Reizänderungen* wahrzunehmen, während konstante Reize auf die Dauer ermüden.

Darüber hinaus soll noch angemerkt werden, dass die Richtung der Formanttransitionen vom Vokal in die konsonantische Konstriktion und umgekehrt auch für ein und denselben Artikulationsort je nach vorhergehendem und nachfolgendem Vokal stark variieren kann. Der Artikulationsort der Konsonanten ist also in Hinblick auf die Formanttransitionen *komplex enkodiert*. Dies führte zur Entwicklung der Motortheorie der Sprachwahrnehmung (Kap. 3).

Fazit zu Abschn. 2.3

Das akustische Sprachsignal stimmhafter Laute basiert auf dem an der Stimmritze (auch Quelle genannt) gebildeten Primärschall (auch Phonationsschall). Dieses Schallsignal regt das aus Rachen-, Mund und Nasenraum bestehende Ansatzrohr akustisch an bzw. wird vom Ansatzrohr überformt oder gefiltert und dann von Mund und Nase abgestrahlt (Quelle-Filter-Theorie). Bei Vokalen bilden sich im Ansatzrohr Formanten, deren Frequenzlagen ein Formantmuster ausbilden, welches mit der Form des Ansatzrohres, also mit der Artikulation, variiert. Die Artikulationsbewegungen zur Ausbildung der konsonantischen Engstellen bzw. Verschlüsse resultieren akustisch in Formanttransitionen. Die unterschiedlichen Verläufe dieser Transitionen charakterisieren unterschiedliche konsonantische Artikulationsorte. Im Ansatzrohr sowie an der Glottis kann aufgrund turbulenter Luftströmung auch Rauschen entstehen. Eine solche Rauschquelle wird als sekundäre Schallquelle bezeichnet. Rauschquellen produzieren nichtperiodische Signale. Ihr Frequenzspektrum ist kontinuierlich und liegt bei gesprochener Sprache zumeist im Bereich von 2 kHz bis 10 kHz. Phonationsschall hingegen ist periodisch und liefert ein Linienspektrum. Die Grundfrequenz des Phonationsschalls liegt etwa im Bereich von 80 Hz bis 400 Hz, und die Obertöne sind genaue Vielfache der Grundfrequenz. Die Amplitude der Obertöne wird von der Lage der Formanten beeinflusst; die Obertöne verleihen den Formanten akustische Energie. Deshalb ist der Phonationsschall bei gesprochener Sprache insbesondere im Bereich der ersten drei Formanten, also im Bereich von ca. 100 Hz bis ca. 3000 Hz, energiereich.

2.4 Störungen der Sprachproduktion

Die am Anfang dieses Kapitels beschriebene Einteilung der Hierarchie der Sprachproduktion in einen kognitiven und einen nachgeordneten sensorisch-motorischen Teil und die weitere Unterteilung des kognitiven Teils in Konzeptualisierung, Wortfindung (also lexikalische Aktivierung) und grammatikalische Strukturierung des zu produzierenden Satzes bzw. des sensorisch-motorischen Teils in motorische Planung und motorische Ausführung kann auch als Basis für die Einteilung der Störungen der Sprachproduktion herangezogen werden. Folgen wir der durch die American Speech Language & Hearing Association gegebenen Klassifizierungen (American Speech, Language & Hearing Association 2016), so können wir darüber hinaus noch Störungen im Kindesalter und Störungen im Erwachsenenalter unterscheiden. Bei Erwachsenen handelt es sich dabei vorwiegend um durch organische oder psychische Störungen bzw. Erkrankungen bedingte Sprechstörungen: Aphasien, Sprechapraxie, Dysarthrien, Stottern, Artikulationsstörungen und Stimmstörungen. Bei Kindern treten entwicklungsbedingte Sprach- und Sprechstörungen hinzu: Sprachentwicklungsverzögerungen und Sprachentwicklungsstörungen (Abschn. 4.3).

Aphasien treten vor allem nach Schlaganfällen auf und betreffen vor allen die Fähigkeit, Sprache zu produzieren, aber auch gesprochene Sprache wahrzunehmen. Darüber hinaus beeinträchtigen Aphasien auch das Sprachsystem insgesamt, also das gesammelte Wissen über Wörter, ihre Bedeutungen und die Grammatik, und führen damit zu Problemen bei der Konzeptualisierung, Satzstrukturierung und beim Sprachverständnis. Damit führen Aphasien auch zu Einschränkungen beim Lesen und Schreiben bis hin zur völligen Schreib- und Leseunfähigkeit. Während bei der Wernicke-Aphasie hauptsächlich das Sprachverständnis eingeschränkt ist, ist es bei der Broca-Aphasie die Sprachproduktion. Dies bedeutet aber nicht, dass ein unter einer Wernicke-Aphasie leidender Patient fehlerfrei sprechen kann. Da das Sprachverständnis eng an die Konzeptualisierung und an den lexikalischen Zugriff gekoppelt ist, kann der Wernicke-Aphasiker zwar sprechen, bleibt aber häufig unverständlich, da er häufig Wortverwechslungen macht, falsche Wörter verwendet (semantische Paraphasien) oder auch Wortneuschöpfungen realisiert (Neologismen). Der Broca-Aphasiker hingegen spricht angestrengt, da die Konzept-zu-Lexem-Aktivierung wie auch die Produktionsprozesse der Satzbildung und der Silbifizierung betroffen sein können. Aufgrund des gut erhaltenen Sprachverständnisses hat der Broca-Aphasiker ein ausgeprägtes Störungsbewusstsein.

Sprechapraxien treten ebenfalls häufig nach Schlaganfall auf und sind daher in vielen Fällen auch mit Aphasien gekoppelt. Eine reine Sprechapraxie zeichnet sich dadurch aus, dass das Sprachsystem und damit die Fähigkeit des Patienten, Wörter und Sätze bis hin zur phonologischen Form „innerlich“ produzieren zu können, nicht gestört ist. Allerdings ist der Patient unfähig, eine adäquate motorische Planung für Lautverbindungen in Silben, also auch für ganze Silben und ganze Wörter zu realisieren. Die Symptome eines reinen Sprechapraktikers liegen auf der lautlichen Ebene, und die dahinterliegende Störung der motorischen Planung ist oftmals nur schwer zu erkennen. Viele Lautfehler eines Sprechapraktikers sehen aus symptomatischer Sicht oft so aus, als wenn er einen Laut durch einen

anderen Laut ersetzen würde. Sieht es z. B. so aus, als wenn ein stimmhafter durch einen stimmlosen Laut ersetzt wird, so könnte der Sprechfehler auch als phonologischer Fehler, und damit als neuronale Fehlaktivierung auf kognitiver Ebene, interpretiert werden. Es ist nur schwer zu erkennen, dass bei den sprechpraktischen Fehlproduktionen, also z. B. bei der Produktion eines stimmhaften Lautes anstelle eines stimmlosen Lautes, oft nur das Timing zweier SBEs nicht korrekt ist. So ist bei einem sprechpraktisch fehlgebildeten stimmlosen Plosiv als stimmhafter Plosiv oftmals wohl die die Stimmlosigkeit hervorruhende SBE (glottale Öffnungs-SBE) vorhanden, aber derart zeitlich falsch koordiniert, dass diese im Zeitintervall der oralen Verschlussbildung liegt und damit nicht das akustische Ziel eine hörbare Stimmlosigkeit nach der Verschlusslösung realisiert. Bis heute gibt es aber keine Verfahren, die diese vermeintlich phonologischen Fehler eines Sprachapraktikers als apraktische Fehler erkennen könnten. Häufig zeigen Sprechapraktiker aber auch andere Symptome wie artikulatorisches Suchverhalten oder eine starke Inkonsistenz in der Realisierung von Lautfehlern, anhand der dann die Sprechapraxie diagnostiziert werden kann.

Dysarthrien treten bei Patienten auf, die z. B. unter Morbus Parkinson oder amyotropher Lateralsklerose (ALS) leiden. Bei Dysarthrien ist die Weiterleitung primär-motorischer neuronaler Aktivierungen an die Muskulatur und damit die Ausführung von Artikulationsbewegungen gestört. Das Sprachsystem des Dysarthriepatienten kann also durchaus intakt sein, d. h. seine Fähigkeit, Sätze zu bilden, ist nicht gestört. Auch ist die motorische Planung der zeitlichen Koordinierung der Sprechbewegungseinheiten (SBE) nicht gestört, wohl aber ihre motorische Ausführung. Dysarthrien können in sehr unterschiedlichen Formen vorliegen. Bei der häufig auftretenden Form der rigid-hypokinetischen Dysarthrie (bei 90 % aller Parkinson-Patienten) sind die Bewegungen insgesamt zu gering. Die Muskulatur ist starr und die Bewegungsmöglichkeiten sind stark eingeschränkt. Bei der spastischen Dysarthrie (häufig nach Schlaganfall oder Schädel-Hirn-Trauma) kann es hingegen zu ausladenden Artikulationsbewegungen kommen. Die Muskulatur ist in diesem Fall fortwährend überstark angespannt (hyperton), und es kommt daher vielfach auch noch zu einem gepressten Stimmton und zu stoßartiger Artikulation. Aber auch der entgegengesetzte Fall, nämlich der einer hypotonen Dysarthrie, kann nach Schlaganfall und Schädel-Hirn-Trauma auftreten. In dem Fall bleibt die Muskulatur schlaff, und das Sprechen klingt meist müde, leise und monoton.

Es sollen hier auch *Artikulationsstörungen aufgrund fehlgebildeter Artikulationsorgane* (z. B. Gaumenspalten) erwähnt werden. Diese Artikulationsstörungen sind wegen der guten Versorgungslage in vielen Ländern kaum noch anzutreffen. So werden z. B. im Fall der Lippen-Kiefer-Gaumenspalten frühzeitig operative Eingriffe hinsichtlich der diese Sprechstörungen auslösenden organischen Probleme durchgeführt. Im Erwachsenenalter werden allerdings z. B. im Falle einer Krebserkrankung manchmal Zungenteilresektionen, d. h. die Entfernung von bestimmten Teilbereichen der Zunge notwendig. Intuitiv ist anzunehmen, dass solche Patienten nur sehr langsam und damit erst nach und nach wieder verständlich sprechen können. Es ist aber bekannt, dass diese Patienten sehr schnell die neue Situation hinsichtlich ihrer reduzierten Artikulationsorgane erfassen und sofort kompensatorische und adaptive Strategien entwickeln, um ihre Artikulation an die

neuen Gegebenheiten des Artikulationssystems anzupassen. Hintergrund hierzu ist, dass die im Gehirn vorliegende Kompetenz der motorische Planung und Ausführung ja vollständig erhalten geblieben ist und dass auch Kompetenz hinsichtlich kompensatorischer und adaptiver Strategien z. B. als Resultat des Spracherwerbs vorliegt.

Im Kindesalter auftretende *nicht organisch bedingte Artikulationsstörungen* hingegen gehören in den Bereich der Sprach- und Sprechentwicklungsstörungen. So kann es sein, dass Kinder bestimmte Sprachlaute (z. B. einige Frikativlaute) auch im Alter von 6 Jahren noch falsch oder zumindest nicht klar aussprechen können. Dies liegt dann daran, dass die korrekte Artikulation dieser Laute noch nicht erworben wurde. Das Kind hat also noch nicht die Fähigkeit, die zugehörige Sprechbewegungseinheit oder das Bündel zugehöriger SBEs korrekt zu realisieren. Da Kinder die SBEs erst nach und nach über die ersten sechs Lebensjahre lernen, muss darauf geachtet werden, dass eine Artikulationsstörung erst als solche bezeichnet werden kann, wenn ein Kind bestimmte Laute auch noch ein bis mehrere Jahre nach dem typischen Erwerbsalter dieser Laute noch nicht realisieren kann.

Im Kindesalter auftretende *phonologische Störungen* sind wiederum von Artikulationsstörungen abzugrenzen. Artikulationsstörungen kennzeichnen Defizite der motorischen Planung bzw. Ausführung. Dagegen beruht eine rein phonologische Störung auf einem sprachlich-kognitiven Defizit, nämlich auf der systematisch falschen Anwendung eines Lautes in einem definierten Kontext. Ein typisches Zeichen ist z. B. die *Vorverlagerung*, bei der in bestimmten Kontexten, z. B. in CV- oder CVC-Silben, ein velarer silbeninitialer Konsonant im Wortkontext immer postalvelar realisiert wird (z. B. „Kanne“ -> „Tanne“; „Gabel“ -> „Dabel“), während die Bildung des /k/- und /g/-Lautes, z. B. in Einzelsilben wie /ka/ oder /ga/, problemlos klappt.

Stimmstörungen können organisch und psychisch bedingt sein. Organische Stimmstörungen resultieren beispielsweise aus Knotenbildung auf den Stimmlippen nach zu extremem Stimmgebrauch oder durch auf Stimmlippen sitzenden Tumoren. Dadurch wird das Schwingungsverhalten der Stimmlippen beeinträchtigt. Es ergeben sich daraus unterschiedlichste Formen der Heiserkeit bis hin zur Aphonie. Vor allem ist es wichtig, dass die organische Ursache nicht die lockere Verschließung der Stimmritze verhindert, da dies wiederum die effektive Phonation beeinträchtigen würde. Eine Behinderung des Bewegungsverhaltens der Aryknorpel kann insbesondere auch eine Folge von im Kehlkopfbereich auftretenden Lähmungen sein. Psychogen bedingte Stimmstörungen hingegen resultieren z. B. aus psychischen Störungen wie Depression oder aber aus psychischen Belastungen wie Stress. Diese Störungen sind nicht einfach zu systematisieren, zeigen oftmals aber eine ähnliche Symptomatik wie andere Stimmstörungen, nämlich unterschiedliche Formen und Grade von Heiserkeit bis hin zur Aphonie.

Stottern ist die Sprechstörung, die dem Laien vielleicht am ehesten bekannt ist. Dieses Störungsbild wird aber immer noch sehr kontrovers diskutiert. So werden sowohl organische wie auch psychische Ursachen angenommen. Dagegen ist die Symptomatik relativ klar. Stottern kann aus symptomatischer Sicht als eine Sprechstörung definiert werden, in der Unflüssigkeiten in der Produktion eines Wortes oder einer Äußerung auftreten. Diese Unflüssigkeiten

können sich durch ein Langziehen von Lauten oder durch häufige und lange Sprechpausen, aber auch durch eine repetitive Realisierung von Lauten und Silben manifestieren.

Fazit zu Abschn. 2.4

Auf der Basis der Sprachproduktionshierarchie können die zugehörigen Sprach- und Sprechstörungen wie folgt unterschieden werden. Broca-Aphasien beschreiben Störungen der kognitiv-symbolischen Verarbeitung bei der Sprachproduktion (organische oder hirnfunktionelle Störungen im Bereich des semantischen Netzwerks, des mentalen Lexikons, der Grammatikmodule). Sprechapraxien beschreiben organische oder hirnfunktionelle Störungen im Bereich der motorischen Planung. Dysarthrien beschreiben organische oder funktionelle Störungen im Bereich der motorischen Ausführung (vor allem im Bereich Hirnstamm, Nervenbahnen, Muskeln). Artikulationsstörungen können aufgrund fehlgebildeter Artikulationsorgane oder im Kindesalter als Entwicklungsstörungen auftreten. Phonologische Störungen zeigen phonologisch regelhafte Fehler wie z. B. die permanente Substitution eines Lautes durch einen anderen. Stimmstörungen beschreiben Störungen der Phonation und können organisch oder auch psychisch bedingt sein. Stottern ist ein sehr komplexes Störungsbild. Hier werden sowohl organische als auch hirnfunktionelle sowie psychische Ursachen diskutiert.

Literatur

- American Speech, Language & Hearing Association (2016) Classification of speech and language disorders: <http://www.asha.org/public/speech/disorders/>. Zugegriffen: 27. Okt. 2017
- Cholin J (2008) The mental syllabary in speech production: an integration of different approaches and domains. *Aphasiology* 22:1127–1141
- Goldstein L, Pouplier M, Chen L, Saltzman E, Byrd D (2007) Dynamic action units slip in speech production errors. *Cognition* 103:386–412
- Grassegger H (2016) *Phonetik/Phonologie*. Schulz Kirchner Verlag, Idstein
- Hanson HM, Stevens KN, Kuo HKJ, Chen MY, Slifka J (2001) Towards models of phonation. *J Phonetics* 29:451–480
- Kelso JAS, Tuller B (1983) „Compensatory articulation“ under conditions of reduced afferent information. *J Speech Lang Hear R* 26:217–224
- Kent RD (1993) Vocal tract acoustics. *J Voice* 7:97–117
- Kent RD (1997) *The Speech Sciences*. Singular Publishing, San Diego, CA
- Levett WJM, Roelofs A, Meyer AS (1999) A theory of lexical access in speech production. *Behav Brain Sci* 22:1–75
- Raphael LJ, Bordon GJ, Harris KS (2007) *Speech science primer: physiology, acoustics, and perception of speech*. Lippincott Williams & Wilkins, Baltimore, MD
- Steyvers M, Tenenbaum JB (2005) The large-scale structure of semantic networks: statistical analyses and a model of semantic growth. *Cognitive Sci* 29:1551–6709

Neuronale Modellierung der Sprachverarbeitung und
des Sprachlernens

Eine Einführung

Kröger, B.J.

2018, XIII, 290 S. 141 Abb., 54 Abb. in Farbe., Softcover

ISBN: 978-3-662-55458-6