Entwicklung eines dreidimensionalen Artikulatormodells für die Sprachsynthese

Diplomarbeit

am Institut für Computergraphik, Fachbereich Informatik, Universität Rostock
und
am Fraunhofer-Institut für Graphische Datenverarbeitung, Rostock

Bearbeiter: Peter Birkholz
geboren am 08.04.1978 in Rostock
Matrikelnr.: 097201776

Betreuer: Prof. Dr.-Ing. habil. Dietmar Jackèl
Universität Rostock, Fachbereich Informatik

Dipl.-Ing. Christian Peter
Fraunhofer IGD, Rostock

Gutachter: Prof. Dr.-Ing. habil. Dietmar Jackèl
Universität Rostock, Fachbereich Informatik

Prof. Dr.-Ing. habil. Erika Müller
Universität Rostock, Fachbereich Elektrotechnik

Abgabedatum: 2. Juli 2002

Danksagung

An dieser Stelle möchte ich mich ganz herzlich bei allen Personen bedanken, die zum Gelingen der Arbeit beigetragen haben.
Meinen Betreuern Dietmar Jackèl und Christian Peter möchte ich besonders dafür danken, dass ich das Thema dieser Arbeit selbst auswählen durfte und dass sie mich in jeder Hinsicht bei der Bearbeitung unterstützt haben.
Ralf Winkler und Mohan Sondhi danke ich für die Diskussionen, aus denen viele Anregungen hervorgegangen sind und die meine Augen für eine kritische und distanzierte Betrachtung meiner eigenen Arbeit geschärft haben.
Christine Mooshammer und Christian Peter waren so freundlich, sich die Zeit für eine Korrekturlesung der Arbeit zu nehmen.
Großer Dank gebührt auch Kevin Munhall, der mir das Röntgenfilmmaterial für die eigenen Untersuchungen zur Verfügung gestellt hat.

Zusammenfassung

In dieser Arbeit wird ein System zur Synthese stimmhafter Sprachsignale mit einem dreidimensionalen Artikulatormodell vorgestellt. Sowohl die geometrische Konstruktion des Artikulatormodells als auch die Methoden für die akustische Synthese werden ausführlich behandelt.
Die Geometrie des Modells wurde aus Röntgenfilmaufnahmen des Kopfes und anatomischen Daten aus der Literatur abgeleitet. Die Positionen der Artikulatoren (Unterkiefer, Lippen, Zunge, weicher Gaumen und Kehlkopf) lassen sich mit mehreren Parametern steuern. Aus der Geometrie des Artikulatormodells wird die akustische Übertragungsfunktion des Vokaltrakts berechnet.
Mit der Übertragungsfunktion und der Anregung des Sprechtrakts durch ein selbstschwingendes Modell der Stimmlippen wird das akustische Sprachsignal erzeugt. Eine einfache zeitliche Steuerung der Modellparameter ermöglicht schließlich die Synthese zusammenhängender Äußerungen aus Vokalen und Nasallauten.

Abstract

A system for the synthesis of voiced speech signals by means of a three-dimensional articulatory model is presented. The geometrical construction of the articulatory model as well as the acoustic synthesis are the topics of this work.
The geometry of the model was derived from midsagittal x-ray tracings and from anatomical data in the literature. The positions of the articulators (jaw, lips, tongue, velum and larynx) are controlled by a set of parameters. The acoustic transfer function of the vocal tract is calculated from the geometry of the articulatory model.
The transfer function, together with the excitation of the vocal tract from a two-mass model of the vocal cords are used for computing the acoustic speech signal. Finally, with a simple method for controlling the model parameters in time, continuous speech with vowels and nasal sounds can be synthesized.

Inhaltsverzeichnis

1 Einleitung **1**
- 1.1 Motivation . . . 1
- 1.2 Einführung in die Sprachschallerzeugung . . . 2
  - 1.2.1 Akustische Eigenschaften des Sprachschalls bei stimmhafter Anregung . . . 2
  - 1.2.2 Sekundäre Schallerzeugung . . . 3
  - 1.2.3 Nasale . . . 4
- 1.3 Methoden der Sprachsynthese . . . 5
  - 1.3.1 Formantsynthese . . . 5
  - 1.3.2 Konkatenative Synthese . . . 6
  - 1.3.3 Artikulatorische Synthese . . . 7
- 1.4 Überblick über die Arbeit . . . 8

2 Physiologie des Sprechapparats **11**
- 2.1 Nasenhöhle . . . 12
  - 2.1.1 Lage und Aufbau . . . 12
  - 2.1.2 Funktion . . . 13
- 2.2 Mundhöhle . . . 13
  - 2.2.1 Lage und Aufbau . . . 14
  - 2.2.2 Funktion . . . 14
- 2.3 Rachen . . . 14
- 2.4 Zunge . . . 15
  - 2.4.1 Lage und Aufbau . . . 15
  - 2.4.2 Funktion . . . 15
- 2.5 Kehlkopf . . . 16
  - 2.5.1 Lage und Aufbau . . . 16
  - 2.5.2 Phonation . . . 17
- 2.6 Untere Atemwege . . . 20
  - 2.6.1 Aufbau . . . 20
  - 2.6.2 Funktion . . . 21

3 Artikulatormodelle aus der Literatur **23**
- 3.1 Sechs-Parameter-Modell von Ishizaka . . . 24
- 3.2 Modell von Mermelstein . . . 25
  - 3.2.1 Geometrie . . . 25
  - 3.2.2 Vokaltraktlänge und Querschnittsfunktion . . . 27

3.3 3D-Vokaltraktmodell von Engwall 28
3.3.1 Altes Modell 28
3.3.2 Neues Modell 29
3.4 Diskussion 30

4 Entwicklung eines neuen Artikulatormodells 33
4.1 Geometrie und Parameter des Mund- und Rachenraums 33
4.1.1 Die Mantelgitter 35
4.1.2 Das Zungengitter 41
4.1.3 Beschränkung der Parameter 46
4.2 Berechnung der Querschnittsfunktion 51
4.2.1 Mittellinie des Ansatzrohrs 51
4.2.2 Schnitt- und Flächenberechnung 53
4.3 Geometrie des Nasenraums 61

5 Akustische Synthese 65
5.1 Überblick 65
5.1.1 Größen und Einheiten 65
5.1.2 Vorgehensweise 66
5.1.3 Diskretisierung 67
5.2 Berechnung der Impulsantworten 69
5.2.1 Elektrisches Ersatzschaltbild des Vokaltrakts 69
5.2.2 Nasennebenhöhlen 73
5.2.3 Abstrahlungsimpedanz 75
5.2.4 Eingangsimpedanz und Übertragungsfunktion 76
5.3 Modell für die stimmhafte Anregung 78
5.3.1 Einführung 78
5.3.2 Zwei-Massen-Modell 79
5.3.3 Numerische Simulation 85

6 Steuerung der Artikulatoren 89

7 Schlussfolgerungen 93
7.1 Zusammenfassung der eigenen Beiträge 93
7.2 Diskussion der Ergebnisse 94
7.3 Ausblick 94
7.3.1 Artikulatormodell 94
7.3.2 Akustische Synthese 95
7.3.3 Steuerung der Parameter 96

A Technische Realisierung 97

Literaturverzeichnis 99

Kapitel 1

Einleitung

1.1 Motivation

> Man muss jedem Hindernis Geduld, Beharrlichkeit und eine sanfte Stimme entgegenstellen.
> *Thomas Jeffersen*

Sprache ist die wichtigste Kommunikationsform des Menschen. Die voranschreitende technische Entwicklung wird in Zukunft zweifellos auch die verbale Verständigung zwischen Menschen und Maschinen zur Folge haben, die von der Mensch-zu-Mensch-Kommunikation nicht mehr zu unterscheiden ist. Dazu sind sowohl Fortschritte auf dem Gebiet der künstlichen Intelligenz als auch der Sprachsynthese und -erkennung erforderlich.
Die vorliegende Arbeit widmet sich der Sprachsynthese. Schon vor der Entwicklung der Computer gab es ernsthafte Versuche, den Prozess der menschlichen Spracherzeugung nachzubilden. Die ersten dokumentierten Sprechapparate wurden bereits gegen Ende des 18. Jahrhunderts gebaut [26].
Heute wird synthetische Sprache fast ausschließlich mit dem Computer erzeugt. Die Sprachsynthetisatoren werden hauptsächlich zur Wiedergabe geschriebener Texte und in automatischen Auskunfts- und Dialogsystemen eingesetzt. Die Methoden für die automatische Spracherzeugung sind vielfältig. In Abschnitt 1.3 werden die wichtigsten drei von ihnen kurz vorgestellt.
Viele Forscher widmen sich momentan der konkatenativen Synthese, die erfolgreich in fast allen kommerziellen Sprachsynthetisatoren eingesetzt wird. Durch die Verwendung voraufgenommener Sprachbausteine stößt diese Art der Synthese jedoch schnell an ihre Grenzen. Es ist damit z.B. äußerst schwierig, die Stimme eines generischen Sprechers beliebigen Alters und Geschlechts zu erzeugen. Auch die Änderung der Stimmqualität bzw. des Gemütszustands eines bestimmten Sprechers wirft erhebliche Probleme auf.
Auf lange Sicht besitzt die artikulatorische Sprachsynthese ein weitaus höheres Potential als die konkatenative Synthese, auch wenn sie uns heute noch vor viele Probleme stellt. Dennoch basierte schon der erste vollautomatische Sprachsynthetisator von 1968 [70] auf einem artikulatorischen Synthetisator. Auch die vorliegende Arbeit beschäftigt sich mit der artikulatorischen Sprachsynthese, wobei das Hauptaugenmerk auf die Konstruktion eines dreidimensionalen Artikulatormodells gelegt wird.

1.2 Einführung in die Sprachschallerzeugung

In diesem Abschnitt soll eine kurze Einführung in die Sprachschallerzeugung durch einen menschlichen Artikulationstrakt gegeben werden.[1]
Nach der klassischen Quelle-Filter-Theorie [21] lässt sich die Spracherzeugung als ein zweistufiger Prozess auffassen. Die erste Stufe (Quelle) erzeugt ein Schallsignal mit einem bestimmten Spektrum, welches in der zweiten Stufe einen Resonator (Filter) durchläuft, der das Quellspektrum mit einer bestimmten Übertragungsfunktion „verformt".
Im Folgenden sollen die stimmhafte und stimmlose Schallerzeugung getrennt voneinander betrachtet werden.

1.2.1 Akustische Eigenschaften des Sprachschalls bei stimmhafter Anregung

In diesem Abschnitt geht es um die Erzeugung von Sprachlauten, bei denen der Vokaltrakt eine durchgängig freie Passage darstellt (keine Engstellen oder Verschlüsse) und das Quellsignal eine Folge von Luftstrompulsen ist, die durch das quasiperiodische Öffnen und Schließen der Stimmbänder entstehen.[2] Dazu zählen die Vokale und vokalähnlichen Laute. Das akustische Filter ist in diesem Fall der gesamte Sprechtrakt vom Kehlkopf bis zur Mund- oder Nasenöffnung. Da es sich hier um ein lineares System handelt, erhält man das Spektrum des am Mund abgestrahlten Schallsignals durch die Multiplikation des Quellspektrums mit der Übertragungsfunktion des Filters (siehe Abb. 1.1). Ein mathematisch äquivalenter Weg wäre die Faltung des Quellsignals mit der Impulsantwort des Filters im Zeitbereich.

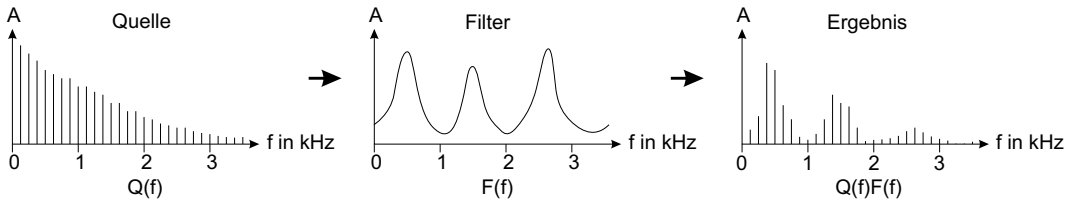

Abbildung 1.1: Prinzip der Filterung des Quellsignals durch den Vokaltrakt. Nach [21].

Die Übertragungsfunktion des Vokaltrakts ist dadurch charakterisiert, dass sie an bestimmten Frequenzen eine deutlich höhere Amplitude als im restlichen Spektrum aufweist. Bei den Gipfeln im Spektrum handelt es sich um die *Formanten*. Sie entstehen an den Resonanzfrequenzen des Vokaltrakts.
Um die Entstehung der Formanten zu verstehen, kann man sich den Vokaltrakt anatomisch näherungsweise als ein Rohr mit veränderlichem Querschnitt entlang seiner Mittellinie vorstellen. Das Rohr ist ca. 17 cm lang und am vorderen Ende geöffnet. Oftmals wird es auch als *Ansatzrohr* bezeichnet.

[1] Eine ausführlichere Darstellung des Themas findet man z.B. in [73].

[2] Die Anatomie des Vokaltrakts und die Funktion der Stimmbänder wird in Abschnitt 2 eingehend behandelt.

In dem Rohr befindet sich eine Luftsäule, die durch die Quelle in Schwingungen versetzt wird. Durch die Anwendung der *Websterschen Horngleichung*[3] lässt sich zeigen, dass sich innerhalb des Rohres stehende Wellen bei bestimmten Resonanzfrequenzen, eben den Formanten, ausbilden. Eine stehende Welle ist nichts anderes als eine Schallwelle, die die spezifische Geometrie ihrer Umgebung als Randbedingung erfüllt und daher in dieser Geometrie eine Resonanzwelle bildet.
Betrachten wir zunächst das querschnittsneutrale Rohr als einfachsten Fall, d.h. ein Rohr mit einer gleichmäßigen Querschnittsfläche von z.B. 5 cm^2 über seine gesamte Länge. Für diesen Fall lässt sich zeigen, dass die Resonanzfrequenzen bei einer Rohrlänge von 17 cm etwa 500 Hz, 1500 Hz und 2500 Hz für die Formanten $F1$, $F2$ und $F3$ betragen. Bei einer Abweichung vom neutralen Querschnitt, wie es bei den meisten Sprachlauten der Fall ist, ändert sich auch die Lage der Formanten. Die Auswirkungen von Querschnittsänderungen des Ansatzrohres auf die Formantfrequenzen wurde z.B. von Ungeheuer [71] genauer untersucht. Hervorgerufen werden diese Querschnittsänderungen durch die Bewegung der Artikulatoren (siehe Abschnitt 2).

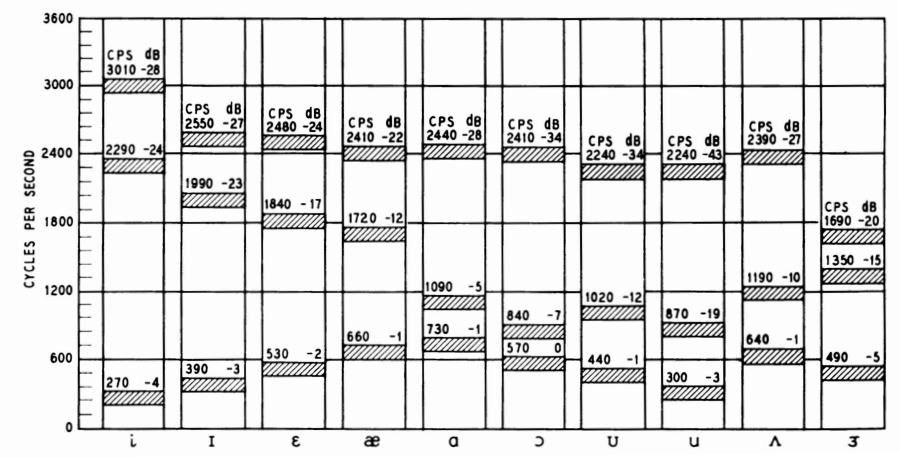

Abbildung 1.2: Über 33 Sprecher gemittelte Frequenzen und relative Amplituden der ersten drei Formanten für die englischen Vokale. Quelle: [26].

Die ersten drei Formantfrequenzen für die verschiedenen Vokale des Englischen sind in Abb. 1.2 dargestellt. Sie wurden über jeweils 33 Sprecher gemittelt. Es ist erwiesen, dass bereits die ersten zwei oder drei Formanten für die Unterscheidbarkeit und Verständlichkeit von Vokalen ausreichend sind. Die höheren Formanten enthalten typischerweise nur noch sprecherspezifische Informationen.

1.2.2 Sekundäre Schallerzeugung

Neben der stimmhaften Anregung durch die Stimmbänder gibt es noch zwei andere Möglichkeiten, im Vokaltrakt Schall zu erzeugen.
Eine Möglichkeit besteht in der kritischen Engebildung an einer bestimmten Stelle des Vokaltrakts, während die Stimmritze (Öffnung zwischen den Stimmbändern) weit geöff-

[3] Auf eine nähere Betrachtung der Gleichung sei an dieser Stelle verzichtet. Der interessierte Leser findet sie z.B. in [71] erläutert.

net ist. An der Engstelle erhöht sich die Strömungsgeschwindigkeit der Luft so weit, dass die normalerweise laminare Luftströmung in eine chaotisch-turbulente Verwirbelung umschlägt. Dadurch entsteht ein Schallsignal, welches sich von der Engstelle aus in beide Richtungen im Vokaltrakt ausbreitet. Dieses Anregungssignal ist im Gegensatz zur stimmhaften Schallerzeugung aperiodisch und wird Rauschen genannt. Durch die Filterung des Rauschsignals im Vokaltrakt entstehen die verschiedenen Frikative bzw. Reibelaute ([s], [f], [sch], ...). Welcher Laut dabei konkret entsteht, ist von der Position der Engstelle im Ansatzrohr abhängig.
Grundsätzlich gelten die akustischen Gesetzte für die Bildung stimmhafter Laute auch hier, d.h. das Rauschspektrum wird an den Formanten des nachgeschalteten Resonators (Teil des Ansatzrohres zwischen Engstelle und Mundöffung) verstärkt. Ein Sonderfall ist in diesem Zusammenhang das Flüstern. Das Rauschen entsteht dabei an einer Verengung zwischen den Stellknorpeln des Kehlkopfes (siehe Abschnitt 2.5.2).
Wenn der Vokaltrakt sowohl stimmhaft (an den Stimmlippen) als auch stimmlos (an einer kritischen Verengung) angeregt wird, spricht man von gemischter Anregung. Dadurch entstehen die stimmhaften Reibelaute, wie in „Garage“ oder in „See“.

Eine weitere Möglichkeit zur sekundären Schallerzeugung besteht in der Bildung und dem plötzlichen Lösen eines Verschlusses im Vokaltrakt. Dadurch entstehen die Plosivlaute, die analog zu den Frikativen sowohl stimmlos ([p], [t], [k]) als auch stimmhaft ([b], [d], [g]) gebildet werden können. Die kurze (40-100 ms) Verschlussphase ist bei den stimmlosen Plosiven durch das Fehlen jeglicher Anregung charakterisiert. Dagegen ist die Verschlussphase der stimmhaften Plosive zumindest durch eine zeitweilige stimmhafte Anregung gekennzeichnet. Bei der „Sprengung“ des Verschlusses entsteht eine mehr oder weniger stark ausgeprägte „impulsartige Druckwelle“ (Burst). Auch hier ist der Ort der Verschlussbildung für den von uns wahrgenommenen Plosivlaut entscheidend. Die Unterscheidung erfolgt jedoch weniger anhand des Spektrums des Bursts, sondern anhand der Formantübergänge zu den benachbarten Lauten.
Die Resultate sekundärer Schallerzeugung werden auch als *Geräuschlaute* bezeichnet.

1.2.3 Nasale

Bisher wurde die Akustik des Vokaltrakts ohne den Nasenraum betrachtet, d.h. wir haben vorausgesetzt, dass das Gaumensegel (siehe Abschnitt 2) angehoben war und den Nasenraum vom übrigen Vokaltrakt akustisch abgetrennt hat.
Durch das Senken des Gaumensegels wird der Nasenraum akustisch an den Vokaltrakt angekoppelt. Je nach der Stellung der übrigen Artikulatoren ergeben sich – bei stimmhafter Anregung – die folgenden Möglichkeiten:

- Der Mundraum bildet eine freie Passage. In dieser Situation bilden sich sowohl stehende Wellen entlang des Schallpfades Rachen-Nasenraum als auch entlang des Pfades Rachen-Mundraum aus. Durch die Resonanzen des Nasenraumes wird dem restlichen Ansatzrohr Energie bei bestimmten Frequenzen entzogen. Im Spektrum des von den Lippen abgestrahlen Schallsignals zeigt sich dies in Form von *Antiresonanzen*, d.h. Einbrüchen bei diesen Frequenzen. Von den Nasenlöchern wird ein Schallsignal abgestrahlt, in dem genau diese Frequenzen betont werden, das

aber insgesamt sehr viel schwächer ist. Das auditive Ergebnis dieses Vorgangs sind *nasalierte Vokale*, wie sie z.B. im Französischen vorkommen.

- Die Lippen sind geschlossen und die Zunge ist gesenkt. Hier findet eine akustische Abstrahlung nur noch von den Nasenlöchern statt. In diesem Fall bildet der gesamte Mundraum einen akustischen Resonator, der parallel zum Rachen-Nasenraum-Schallpfad geschaltet ist und diesem Energie bei bestimmten Antiformanten entzieht. Das auditive Ergebnis ist der Nasal [m].

- Die Lippen sind geöffnet und die Zunge bildet am Zahndamm einen Verschluss. Das Volumen der als Resonanzraum wirksamen Mundhöhle ist geringer als beim [m]. Die Lage der Antiformanten ändert sich und es entsteht der auditive Eindruck des Nasals [n].

- Die Lippen sind geöffnet und die Zunge bildet einen Verschluss mit dem weichen Gaumen. Hier entfällt der akustische Einfluss des Mundraums komplett und es entsteht ein Nasal wie im Wort „lang“.

1.3 Methoden der Sprachsynthese

Es existieren eine Vielzahl verschiedener Methoden für die Sprachsynthese, wobei fast alle in eine der folgenden Kategorien fallen: Formantsynthese, konkatenative Synthese und artikulatorische Synthese. Diese werden in den folgenden Unterpunkten kurz vorgestellt.

1.3.1 Formantsynthese

Ein Großteil des akustisch-phonetischen Wissens liegt heute in Form von Formantfrequenzen und Formantübergängen vor. Dieses Wissen lässt sich leicht in Regeln umwandeln, um einen Formantsynthetisator zu steuern. Die Formantsynthese benutzt ein stark vereinfachtes Quelle-Filter-Modell, bei dem angenommen wird, dass Quelle und Filter komplett unabhängig voneinander existieren. Dementsprechend werden sie auch unabhängig voneinander modelliert.

Für die stimmhafte Anregung wird die Quelle häufig durch eine Folge von Impulsen oder einfachen Pulsformen angenähert, die durch die Sprachgrundfrequenz f_0 gesteuert werden. Das Vokaltraktfilter steuert in diesem Fall explizit die Frequenz, Bandbreite und Amplitude jedes Formanten.
Für die stimmlose oder gemischte Anregung wird zusätzlich ein Rauschgenerator benutzt, dessen Spektrum mit Hilfe eines Filters gemäß den vorgegebenen Regeln geformt wird.

Die von Formantsynthetisatoren erzeugte Sprache ist in der Regel sehr verständlich, aber wenig natürlich. Ein Vorteil ist der geringe Bedarf an Speicherplatz und Rechenleistung.
Prominentestes Beispiel für diese Art der Synthese ist der Klatt-Formantsynthetisator [36]. Es handelt sich dabei um einen sehr komplex strukturierten Synthetisator, der eine hohe Flexibilität aufweist. Andererseits wird es durch die große

Menge von 20 einstellbaren Parametern zunehmend erschwert, die dazugehörigen Regeln zu formulieren.

1.3.2 Konkatenative Synthese

Konkatenative Synthese verkettet kurze, natürlichsprachliche Bausteine, die aus Sprachaufnahmen herausgeschnitten und in einer Sprachdatenbank abgelegt wurden, zu neuen Äußerungen. Bei den Bausteinen handelt es sich in der Regel um elementare Einheiten wie Phone, Diphone oder Halbsilben.
Phone sind die kleinsten lautlichen Einheiten, die sich durch die Segmentierung eines gesprochenen Textes gewinnen lassen. Ein reines Phoninventar ist mit nur 40-50 Einheiten sehr klein.
Diphone sind die Übergänge zwischen zwei Phonen und erstrecken sich von der Mitte des ersten bis zur Mitte des zweiten Phons. Prinzipiell ist die Anzahl der Diphone einer Sprache gleich dem Quadrat der Anzahl der Phone, wenn meistens auch nicht alle Kombinationen von Phonen in der Sprache vorkommen.
Halbsilben entstehen durch die Aufspaltung von Silben in eine initiale und eine finale Halbsilbe und sind damit verhältnismäßig große Einheiten. Im Deutschen gibt es ungefähr 5500 Halbsilben.

Der Sprachsynthetisator hat die Aufgabe, aus den vorliegenden Basiseinheiten diejenigen auszuwähen, die zusammen eine gegebene Zieläußerung bilden sollen, und diese entsprechend zu verketten. Dabei sind in der Regel Änderungen der prosodischen Attribute (Dauer, Grundfrequenz und Amplitude) der ausgewählten Bausteine nötig, um eine gewünschte Intonation zu erzielen. Weiterhin müssen spektrale Anpassungen an den Verknüpfungspunkten der Einheiten vorgenommen werden, um „weiche“ Segmentübergänge zu erzeugen.
Durch die Tatsache, dass *natürlichsprachliche* Einheiten verkettet werden, kann diese Form der Synthese potenziell sehr natürlich klingen. Bis vor einigen Jahren basierten die neu entwickelten konkatenativen Synthesesysteme meistens auf einer Diphon-Datenbasis, in der zu jedem Diphon eine Instanz abgespeichert war. In den letzten 4-5 Jahren hat hier ein schleichender Umbruch eingesetzt, der heute die ersten wirklich qualitativ hochwertigen Synthesesysteme ermöglicht hat.
Die Idee ist es, zu jedem Grundbaustein statt nur einer viele hundert Instanzen abzuspeichern, die in verschiedenen Kontexten erzeugt wurden. Für die Verkettung müssen nun aus den vielen Instanzen jeweils diejenigen ausgewählt werden, die am ehesten der prosodischen Zielstellung entsprechen und an den Verkettungsstellen die geringsten spektralen Unterschiede aufweisen. Dadurch sind nachträgliche Änderungen der prosodischen Parameter kaum mehr nötig. Erkauft wird dieser Vorteil jedoch durch den erhöhten Aufwand bei der Erstellung der Datenbasis sowie durch einen erhöhten Speicherplatz- und Rechenzeitbedarf.
Für die Segmentierung und Etikettierung großer Sprachdatenmengen gibt es heute aber automatische Werkzeuge, die zum großen Teil den Forschungen in der Spracherkennung zu verdanken sind. Außerdem sind jetzt effiziente Suchtechniken auf großen Datenbeständen verfügbar, die es erlauben, in Echtzeit aus Millionen von potenziellen Segmentsequenzen die für eine Äußerung optimale herauszufinden.

Trotz der vielversprechenden Erfolge bleibt die konkatenative Synthese sprecher- und sprachabhängig, da sie stets auf den Sprachdaten eines Sprechers basiert.

1.3.3 Artikulatorische Synthese

Die artikulatorische Sprachsynthese ahmt den Prozess der Spracherzeugung auf physiologischer und akustischer Ebene direkt nach. Zum heutigen Zeitpunkt ist sie jedoch noch eher auf Grundlagenforschung als auf praktische Anwendung hin orientiert. Es gibt noch keinen artikulatorisch gesteuerten Synthetisator, der beliebige Sätze mindestens einer Sprache mit hoher Qualität erzeugen kann. Dies liegt zum Einen an unserem begrenzten Wissen über die genauen physiologisch-motorischen Abläufe bei der Sprachproduktion, andererseits aber auch an der unzureichend erforschten Problematik der Entstehung und Ausbreitung von turbulenten Signalen im Vokaltrakt. Außerdem fehlen umfassende Modelle des Vokaltrakts und der Stimmbänder, die alle Aspekte der Anregung und Filterung in detaillierter Weise nachbilden können.
Dennoch wurden gerade in den letzten Jahren große Fortschritte in dieser Richtung gemacht, die eine Änderung der Situation erhoffen lassen.

Artikulatorische Sprachsynthese basiert auf einem Modell des Vokaltrakts und der Steuerung der Artikulatoren (Lippen, Zunge, Kehlkopf, Gaumensegel, Unterkiefer). Das Vokaltraktmodell dient der Formung des Ansatzrohres, während die Artikulatorsteuerung der zeitlichen Veränderung der Rohrgeometrie dient. Auch hier bildet der Vokaltrakt einen akustischen Filter, der von einer oder mehreren Quellen angeregt wird.
Mit Hilfe der Rohrgeometrie und einem geeigneten Signal für die Quelle wird dann das akustische Schalldrucksignal an der Mundöffnung und den Nasenlöchern berechnet.

Für die akustischen Berechungen im Ansatzrohr macht man üblicherweise folgende vereinfachende Annahmen:

- Die Schallausbreitung im Ansatzrohr findet nur eindimensional statt, also nur entlang der Rohrmittellinie. Diese Annahme trifft zu, wenn die Wellenlängen im Vergleich zum maximalen Rohrdurchmesser groß sind. Das ist bei der Sprachproduktion für die uns interessierenden Frequenzen näherungsweise der Fall.

- Die Krümmung des Ansatzrohres wird vernachlässigt. Dies ist nach einer theoretischen Untersuchung von Sondhi [57] gerechtfertigt. Er zeigt, dass die Resonanzfrequenzen in einem gekrümmten und einem ungekrümmten Vokaltrakt unterhalb von 4 kHz nur um ca. 2%-8% voneinander abweichen.

- Es wird von der exakten Formung des Vokaltrakts abstrahiert. Laut Ungeheuer [71] ist für die Lage der Formanten nur der Verlauf der Querschnittsfläche als Funktion ihrer Position auf der Rohrmittellinie entscheidend. Deshalb wird der Vokaltrakt oft als ein Rohr mit kreisrundem Querschnitt und veränderlichem Radius approximiert.

Die Aufgabe eines Artikulatormodells ist es also, zu jedem Zeitpunkt einer Äußerung die *Querschnittsfunktion* $A(x)$ des Ansatzrohres zu bestimmen, die den Flächeninhalt des Rohrquerschnitts über der Position x auf der Mittellinie (Rohrachse) abträgt.

Die Querschnittsfunktion bildet die Grundlage für die Berechnung der Übertragungsfunktion des Vokaltrakts.
Je nachdem, welche der folgenden Methoden man zur akustischen Simulation benutzt, muss man $A(x)$ entlang der x-Achse abtasten und gelangt zum klassischen diskreten Röhrenmodell, welches das Ansatzrohr als eine Aneinanderreihung von kurzen zylindrischen Rohrabschnitten mit variablen Querschnitten approximiert.

1. Eine Möglichkeit besteht darin, die Webstersche Horngleichung nach den stehenden Wellen im Ansatzrohr zu lösen [71] und die gefundenen Formantfrequenzen in einen Formantsynthetisator zu geben. Die Lösung der Horngleichungen muss jedoch sehr zeitaufwendig approximiert werden, da eine geschlossene Lösung nur für sehr wenige Sonderfälle existiert. Außerdem werden Energieverluste in der Horngleichung nicht berücksichtigt.

2. Die akustischen Eigenschaften der Quelle und des Vokaltrakts werden durch Differenzialgleichungen beschrieben. Dies führt zu einer großen Anzahl linearer und nichtlinearer Gleichungen, die für jeden Abtastwert numerisch gelöst werden müssen. Beispiele für diese Vorgehensweise findet man z.B. in [27] und [32]. Auch diese Methode ist äußerst rechenintensiv.

3. Kelly und Lochbaum [35] haben 1962 eine erheblich schnellere Methode als in 1.) und 2.) vorgestellt. Sie wird als Wellendigitalfilter bezeichnet und basiert auf vor- und zurücklaufenden Wellen in einer digitalen Übertragungsleitung, die den Vokaltrakt repräsentiert. Probleme stellen jedoch frequenzabhängige Verluste in der Leitung dar, da die Methode im Zeitbereich operiert. Außerdem müssen die zylindrischen Rohrabschnitte der Übertragungsleitung alle die gleiche Länge besitzen, die von der Abtastrate des Sprachsignals abhängt. Dadurch kann auch die Gesamtlänge des Rohres nur ein ganzzahliges Vielfaches der Rohrabschnittslänge sein.

4. Die Quelle wird im Zeitbereich und die Übertragungsfunktion im Frequenzbereich berechnet [58]. Dieser *hybride* Ansatz nutzt gleichermaßen die Vorteile von Zeit- und Frequenzbereichsverfahren und wird auch in der vorliegenden Arbeit verwendet.

1.4 Überblick über die Arbeit

Was Sie in den nächsten Abschnitten erwartet, zeigt die folgende Kurzübersicht:

- Abschnitt 2 gibt einen Überblick über die Anatomie der Sprechorgane und ihre Funktion bei der Spracherzeugung.

- Hauptbestandteil dieser Arbeit ist die Entwicklung eines dreidimensionalen Artikulatormodells. Um die Grundlage für den Vergleich mit anderen Modellen zu schaffen, werden in Abschnitt 3 ein eindimensionales, ein zweidimensionales und ein dreidimensionales Artikulatormodell aus der Literatur vorgestellt. Am Ende des Abschnitts werden ihre Vor- und Nachteile diskutiert.

- Der Kern der vorliegenden Arbeit besteht aus den Abschnitten 4 und 5. In Abschnitt 4 wird auf die geometrische Konstruktion und die Parametrisierung des neuen, dreidimensionalen Artikulatormodells eingegangen. Anschließend werden die Verfahren vorgestellt, mit denen die Querschnitts- und Umfangsfunktion des Vokaltrakts aus dem Modell gewonnen werden.

  Beide Funktionen sind die Eingabedaten für die akustische Synthese, die in Abschnitt 5 behandelt wird. Die akustische Synthese umfasst auch die numerische Simulation der Stimmlippenschwingung mit dem klassischen Zwei-Massen-Modell von Ishizaka und Flanagan [32].

- In Abschnitt 6 wird ein einfaches aber dennoch flexibles Verfahren vorgestellt, mit dem manuell die Parameter des Artikulatormodells und der akustischen Synthese im zeitlichen Verlauf gesteuert werden können. Dies ermöglicht die Synthese zusammenhängender Äußerungen mit verschiedenen stimmhaften Lauten. Für die Wörter „Oma“ und „Jana“ werden die Sonagramme des gesprochenen und (re)synthetisierten Sprachsignals gegenübergestellt.

- In Abschnitt 7 werden die eigenen Beiträge dieser Arbeit zur artikulatorischen Sprachsynthese zusammengefasst. Außerdem werden die erzielten Ergebnisse diskutiert und Möglichkeiten für die Weiterentwicklung des Modells und die Synthese aufgezeigt.

- Das Artikulatormodell, die akustische Synthese und die Parametersteuerung wurden in einem Computerprogramm umgesetzt. Einige Details zur technischen Realisierung werden in Anhang A gegeben.

Kapitel 2

Physiologie des Sprechapparats

Dieser Abschnitt gibt einen Überblick über die an der Sprachproduktion beteiligten Organe. Es wird ihre Anatomie und ihre Funktion bei der Spracherzeugung behandelt. Die anatomischen Fakten stammen zum großen Teil aus dem Buch [6].

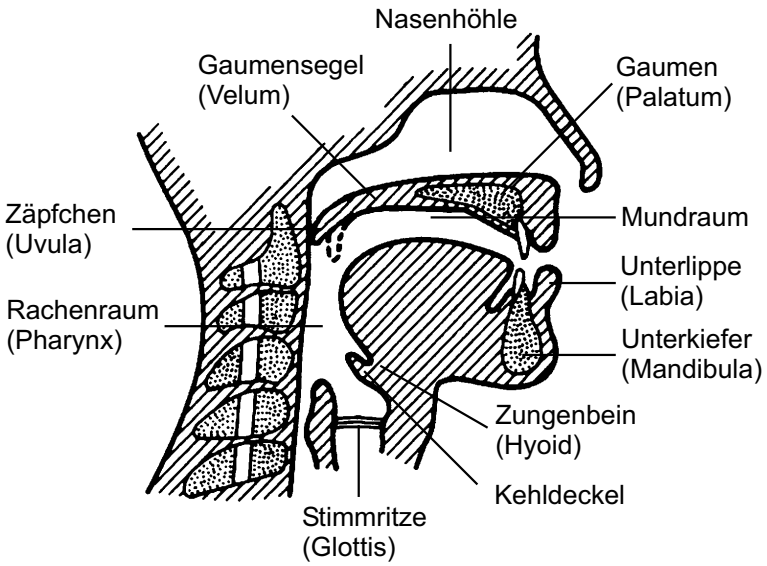

Abbildung 2.1: Anatomie des Vokaltrakts. Quelle: [73].

Abb. 2.1 zeigt schematisch den Aufbau des Sprechtrakts vom Kehlkopf aufwärts. Neben der Spracherzeugung sind seine Hauptaufgaben die Atmung und die Nahrungsaufnahme. Der Sprechtrakt lässt sich grob in drei Kavitäten unterteilen: den Rachen-, den Mund- und den Nasenraum. Seine Form wird in erster Linie von den aktiv bewegbaren Artikulatoren Lippen, Zunge, Unterkiefer, Velum und Larynx (Kehlkopf) diktiert. Dabei unterliegt er verschiedenen geometrischen Veränderungen:

- Verlängerung/Verkürzung durch die Senkung/Hebung des Kehlkopfes sowie die Spreizung/Vorstülpung der Lippen;
- Querschnittsänderung je nach dem Grad der Kieferöffnung und der Lage der Zunge;

- Lokale Enge- bzw. Verschlussbildung, z.B. beim [s] mit alveolarer (am Zahndamm) Engebildung und beim [t] mit alveolarem Verschluss;
- Zu-/Abschaltung des Nasenraums durch abgesenktes bzw. gehobenes Velum;

Diese geometrischen Veränderungen wirken sich auf das akustische Verhalten des Sprechtrakts aus und bestimmen dadurch den Klang gesprochener Laute.

2.1 Nasenhöhle

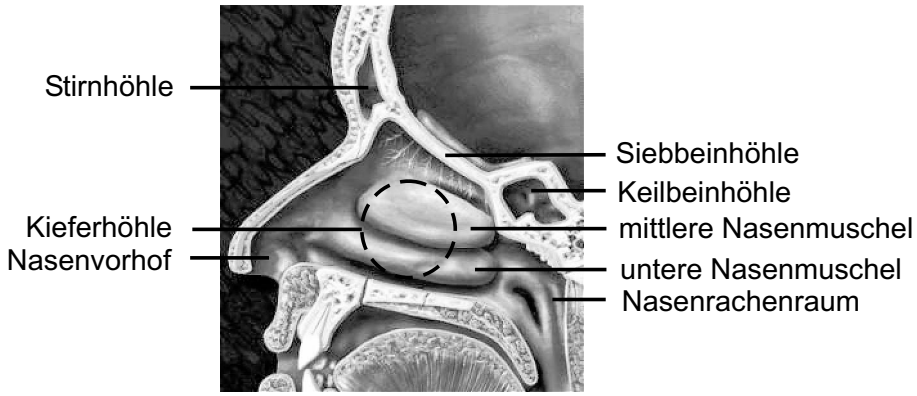

Abbildung 2.2: Nasenraum mit Nebenhöhlen (Stirnhöhle, Keilbeinhöhle, Kieferhöhle, Siebbeinhöhle).

2.1.1 Lage und Aufbau

Unsere Nasenhöhle bildet den ersten Abschnitt der Atemwege. Im Vergleich zur Mundhöhle ist ihre Form relativ statisch. Durch eine senkrechte Trennwand, die Nasenscheidewand, wird sie in zwei paarig angelegte, schmale Hohlräume unterteilt. Die Nasenscheidewand besteht aus einem hinteren knöchernen Teil, sowie einem vorne liegenden Teil, der aus Knorpel und straffem Bindegewebe aufgebaut ist.
Die Nasenhöhle besitzt zwei Öffnungen. Die beiden vorderen Nasenöffnungen (Nasenlöcher) verbinden die beiden Nasenhohlräume mit der Außenwelt. Über die inneren Öffnungen stehen die Hohlräume mit dem Nasenrachenraum in Verbindung.
Von den beiden lateralen Nasenwänden ragen jeweils drei dünne Knochenplatten in die Nasenhöhle hinein. Bei diesen mit Schleimhautepithel überzogenen Knochenplatten handelt es sich um die unteren, mittleren und oberen[1] Nasenmuscheln.

An verschiedenen Stellen außerhalb der eigentlichen Nasenhöhle liegen die Nasennebenhöhlen. Dabei handelt es sich um Hohlräume in bestimmten Schädelknochen, die mit Schleimhaut ausgekleidet und mit Luft gefüllt sind. Die Nasennebenhöhlen kommen paarweise vor und sind über kleine Zugänge mit dem jeweiligen Nasenhohlraum verbunden. Abb. 2.2 zeigt die Lage der Nebenhöhlen. Es gibt

[1] Die oberen Nasenmuschen sind in Abb. 2.2 nicht dargestellt.

- eine rechte und linke Stirnhöhle (*Sinus frontalis*), die über die beiden mittleren Nasengänge mit dem entsprechenden Nasenhohlraum verbunden sind. Sie liegen jeweils hinter und oberhalb der Augenbrauen im Stirnbein. Sie sind durch eine dünne Wand voneinander getrennt. Die Trennwand liegt nur selten in der Kopfmitte, weshalb beide Höhlen zumeist unterschiedlich groß sind.

- eine rechte und linke Kieferhöhle (*Sinus maxillaris*). Diese sind unter dem jeweiligen Auge in der Wange zu finden und breiten sich fast im gesamten Oberkieferknochen aus. Sie bilden die größte Nasennebenhöhle.

- eine rechte und linke Siebbeinhöhle (*Sinus ethmoidalis*). Diese setzen sich aus kleinen, ungefähr erbsengroßen und miteinander verbundenen Hohlräumen – den sogenannten Siebbeinzellen – zusammen. Auf jeder Seite liegt die Höhle zwischen dem jeweiligen Nasenbein und dem inneren Augenwinkel.

- eine rechte und linke Keilbeinhöhle (*Sinus sphenoidalis*), die nahezu die gesamten jeweiligen Keilbeinknochen einnehmen. Die Keilbeine befinden sich tief im Schädelinnern oberhalb des Rachendachs.

2.1.2 Funktion

Während der Atmung dient die Nase der Anfeuchtung, Erwärmung und Filterung der angesaugten Luft. Außerdem enthält die Nase das Riechorgan. Während der Sprachproduktion dient sie der Erzeugung der Nasallaute. Die Nasennebenhöhlen sind dabei mit Helmholtz-Resonatoren (siehe Abschnitt 5.2.2 zur akustischen Synthese) vergleichbar, die an verschiedenen Stellen durch kleine Öffnungen an den Nasaltrakt gekoppelt sind [13, 14]. Dadurch werden verschiedene Antiresonanzen erzeugt, die u.a. für den typischen Klang der Nasallaute verantwortlich sind.

2.2 Mundhöhle

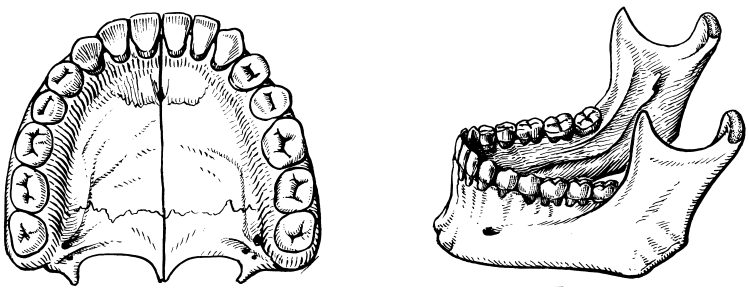

Abbildung 2.3: Harter Gaumen (links) und Unterkiefer (rechts). Quelle: [6].

2.2.1 Lage und Aufbau

Die Mundhöhle lässt sich in den Mundvorhof und die eigentliche Mundhöhle unterteilen. Der Vorhof wird vom Raum zwischen den Lippen, den Wangen und den Zahnreihen gebildet. Die eigentliche Mundhöhle erstreckt sich von den Zahnreihen bis zum Rachen, und wird vom Gaumen überdacht. Sie wird unten durch die Zunge und den Mundboden und seitlich durch die Wangen begrenzt.
Als Mundboden bezeichnet man die Weichteile zwischen dem Unterkieferkörper und dem Zungenbein.
Der Unterkiefer (*Mandibula*) besteht aus einem U-förmigen Knochen, der das „Fundament“ für die unteren Zähne bildet (siehe Abb. 2.3). Das vordere geschlossene Ende des Kiefers bildet das Kinn, während die Knochenfortsätze des hinteren offenen Endes zusammen mit dem Schläfenbein das Kiefergelenk bilden.
Der Gaumen (*Palatum*) trennt die Mundhöhle von der Nasenhöhle. Er wird in den harten Gaumen und den weichen Gaumen unterteilt. Der harte Gaumen umfasst die vorderen 2/3 des Gaumens und besteht aus dem Oberkieferknochen, der mit einer Schleimhaut bedeckt ist. Der weiche Gaumen (siehe Abb. 2.1) füllt das hintere Drittel des Gaumens aus und bildet das Gaumensegel (*Velum*). Der kurze Fortsatz am hinteren Ende des Gaumensegels ist das Gaumenzäpfchen (*Uvula*).
Die Lippen (*Labia oris*) werden in die Ober- und die Unterlippe unterteilt. Sie bestehen zum größten Teil aus Muskeln, aber auch aus Bindegewebe, Blutgefäßen, Drüsen und Nerven.

2.2.2 Funktion

Der Unterkiefer lässt sich durch die Kiefermuskeln heben, senken, vor- und zurückverlagern sowie seitlich verschieben. Damit ist er aktiv an der Formung der Mundhöhle beteiligt. Für die Sprachproduktion ist besonders die Hebung und Senkung wichtig.
Das Velum ist durch zwei Muskeln mit dem Schädel verbunden. Durch die Aktivität der beiden Muskeln wird das Velum angehoben und der Nasenrachenraum gegenüber dem Mundrachenraum abgeschlossen. Dies ist bei der „normalen“ (oralen) Sprachproduktion der Fall. Nur bei der Erzeugung von Nasallauten wird das Velum gesenkt, und der Nasenraum als Schallpfad zugeschalten. Das Zäpfchen (*Uvula*) am hinteren Ende des Velums ist ein muskulöser Fortsatz und spielt bei der Produktion des uvularen Trills [R] eine Rolle.
Die Lippen dienen bei der Spracherzeugung der Sprechtraktverlängerung bzw. -verkürzung bei Vokalen (z.B. [u] und [i]) und der Enge- bzw. Verschlußbildung bei Konsonanten (z.B. [f] und [p]).

2.3 Rachen

Der Rachen (*Pharynx*) ist ein ca. 12 cm langer, mit Schleimhaut ausgekleideter Muskelschlauch, der an der Schädelbasis aufgehängt ist. Er liegt hinter der Nasenhöhle und der Mundhöhle und bildet somit den gemeinsamen Abschnitt von Atem- und Speiseweg.
Der Rachen ist sehr wenig beweglich. Er kann lediglich durch drei Muskeln seitlich verengt oder versteift werden. Die Versteifung der Rachenwände hat großen Einfluss auf

die Resonanzklangfarbe der Sprachlaute.
Der Rachen gliedert sich in drei übereinander liegende Abschnitte:

- Der Nasenrachenraum (*Epipharynx*) ist der oberste Abschnitt und liegt hinter der Nasenhöhle. Am Dach des Nasenrachenraums befinden sich die Öffnungen zu den Ohrtrompeten (Eustachschen Röhren), die den Rachen mit den Mittelohren verbinden.

- Der Mundrachenraum (*Oropharynx*) liegt hinter der Mundhöhle zwischen dem weichen Gaumen und dem Kehlkopfdeckel. Die Hebung des weichen Gaumens schließt den Mundrachenraum und den Nasenrachenraum gegeneinander ab.

- Der Kehlkopfrachenraum (*Hypopharynx*) liegt hinter dem Kehlkopf und reicht vom Kehlkopfdeckel bis zum Beginn der Speiseröhre.

2.4 Zunge

2.4.1 Lage und Aufbau

Die Zunge besteht zum größten Teil aus Muskelgewebe und füllt im Ruhezustand bei normaler Atmung den gesamten Mundraum aus. Sie ist von einer Bindegewebsschicht mit eingebetteten Drüsen und Lymphknötchen umgeben, die wiederum von einer alles umgebenden Schleimhaut bedeckt sind. Die Zunge besteht zu einem Drittel aus dem pharyngalen Teil, der durch Muskeln fest mit dem Zungenbein und dem Schläfenbein verbunden ist. Die restlichen zwei Drittel – der orale Teil – ist frei beweglich.
Die Zungenoberfläche wird von vorne nach hinten in die Zungenspitze (*Apex*), das Zungenblatt (*Lamina*), den Zungenrücken (*Dorsum*) und die Zungenwurzel (*Radix*) unterteilt.
Ihre große Beweglichkeit verdankt die Zunge den äußeren und inneren Zungenmuskeln. Die äußeren Zungenmuskeln entspringen den Knochenteilen des Schädels oder des Zungenbeins und enden an der Zungensehnenplatte (unter der Zungenoberfläche gelegene Durchflechtung der Zungenmuskeln). Diese dient den inneren Zungenmuskeln zur Befestigung.

2.4.2 Funktion

Die Zunge gehört zu den wichtigsten Organen bei der Lautbildung, da sich ihre Lage- und Formänderung auf große Teile des Sprechtrakts auswirkt. Die Zungenmuskeln dienen in ihrem komplexen Zusammenspiel den verschiedensten Funktionen. Dazu gehören die Verschmälerung und Verbreiterung der Zunge, die Anhebung der Seitenränder, die Vor-/Zurück- und die Auf-/Abbewegung sowie die Hebung und Senkung der Zungenspitze.
Für eine nähere Beschreibung der Zungenmuskeln und ihrer Funktionen sei auf [6] verwiesen.

2.5 Kehlkopf

2.5.1 Lage und Aufbau

Am unteren Ende des Sprechtrakts (unterhalb des Zungenbeins) befindet sich der Kehlkopf (*Larynx*), der den Rachen mit der Luftröhre verbindet.

Abb. 2.4 zeigt das Skelett des Kehlkopfes, welches aus verschiedenen großen und kleinen Knorpeln besteht, die durch Bindegewebe, Muskeln und Bänder miteinander verbunden sind.

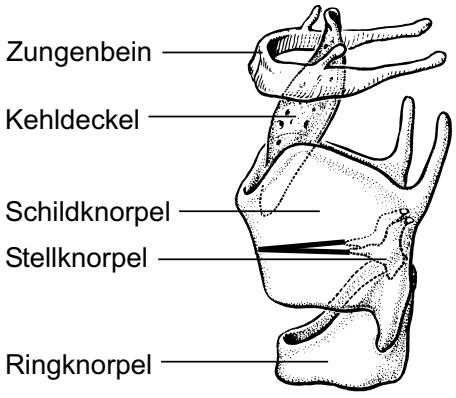

Abbildung 2.4: Knorpelskelett des Kehlkopfes. Die Lage der Stimmbänder ist durch zwei dicke schwarze Striche markiert. Quelle: [6].

- Der Ringknorpel (*Cricoid*) befindet sich am unteren Ende des Kehlkopfes. Seine Form ist mit der eines Siegelrings vergleichbar. Am hinteren oberen Ende des Knorpels sind die zwei kleinen Stellknorpel gelenkig aufgesetzt. Auf der unteren Außenfläche befindet sich beiderseits je eine Gelenkfläche für die Verbindung mit dem unteren Horn des Schildknorpels.

- Der Schildknorpel (*Thyroid*) besteht aus zwei großen seitlichen Knorpelplatten, die am vorderen Ende über einen Winkel von ca. $90 - 120°$ miteinander in Verbindung stehen. Beim Mann tritt er als Adamsapfel deutlich hervor.

- Die Form des Kehldeckelknorpels (*Epiglottis*) ist mit der eines Fahrradsattels vergleichbar. Die Spitze des Knorpels zeigt nach unten vorne und die seitlichen Ränder sind etwas nach hinten umgeschlagen.

- Die paarigen kleinen Stellknorpel (*Arytenoid*, „Aryknorpel“) ähneln einer dreiseitigen Pyramide. Sie besitzen eine Gelenkverbindung mit dem Ringknorpel und dienen der Befestigung der Stimmbänder.

Der Innenraum des Kehlkopfes besteht aus drei unterschiedlich hohen und breiten Stockwerken. Das oberste Stockwerk – der Vorhof (*Vestibulum laryngis*) – reicht vom Kehlkopfeingang bis zu den nach innen vorgewulsteten Taschenfalten („falsche

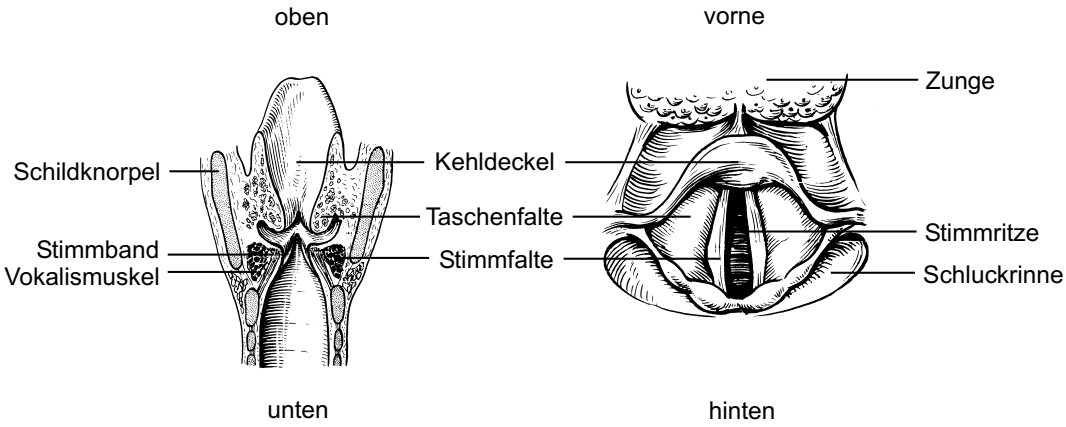

Abbildung 2.5: Innenraum des Kehlkopfes von hinten (linkes Bild) und von oben (rechtes Bild) gesehen. Quelle: [6].

Stimmlippen“). Unter den Taschenfalten befinden sich die Stimmfalten („Stimmlippen“), die die Stimmritze (*Glottis*) links und rechts begrenzen. Sie verlaufen beiderseits zwischen den Stellknorpeln und der Innenfläche der Schildknorpelplatten. Der zur Seite und nach oben ausgebuchtete Raum zwischen den Taschen- und Stimmfalten bildet das mittlere Stockwerk (*Ventriculus laryngis*) des Kehlkopfinnenraums. Das unterste Stockwerk erstreckt sich von den Stimmlippen abwärts bis zum Kehlkopfausgang. In der Wand unter der Schleimhaut dieses (kegelförmigen) Raumes befindet sich eine Membran aus elastischem Bindegewebe, dessen obere dichte Randzone die Stimmbänder bildet.

Die Muskulatur des Kehlkopfes lässt sich in die Stellmuskeln und die Spannmuskeln unterteilen. Durch die Aktivität der Stellmuskeln kann die Weite der Stimmritze verändert werden, während die Spannmuskeln der Veränderung der Spannung der Stimmbänder dienen. Beide Prozesse sind für die Stimmproduktion ausgesprochen wichtig.
Zu den Spannmuskeln zählen der *Musculus cricothyroideus*, der vom Unterrand des Schildknorpels bis zum Bogen des Ringknorpels reicht, sowie der Vokalismuskel, dessen Fasern parallel zu den Stimmbändern verlaufen. Durch die Aktivität des M. cricothyroideus werden die Stimmlippen gespannt, indem sich der Schildknorpel nach vorne dreht. Der Vokalismuskel steuert die innere Elastizität der Stimmlippen und ermöglicht eine Feinabstimmung der Stimmbänderspannung.
Die Stellmuskeln umfassen vier Muskeln für das Schließen und einen Muskel für die Öffnung der Stimmritze.
Neben den Spann- und Stellmuskeln ermöglicht die obere und untere Zungenmuskulatur ein Heben und Senken des Kehlkopfes.

2.5.2 Phonation

Die Stimmritze wird beidseitig von zwei ungleich großen Abschnitten begrenzt: im vorderen (längeren) Abschnitt durch die membranartigen Stimmfalten und im hinteren (kürzeren) Abschnitt durch die Aryknorpel. Die Länge der Stimmfalten reicht von $1,5$ cm bei der Frau bis zu $2,5$ cm beim Mann. Ihre vorderen Enden liegen an der Innenseite

der Schildknorpelplatten dicht nebeneinander. Ihre hintere Befestigung befindet sich an den kleinen Fortsätzen der Stellknorpel. Die Weite und die Form der Stimmritze ist von der jeweils geforderten Funktion abhängig und wird durch die Bewegung der Stellknorpel und der auf sie wirkenden Muskeln verändert.
Abb. 2.6 zeigt die Stellung der Stimmbänder bei verschiedenen Funktionen. Die Teilbilder (a) und (b) zeigen die Stimmlippen bei normaler und tiefer Einatmung. Dabei ist die Stimmritze weit geöffnet.

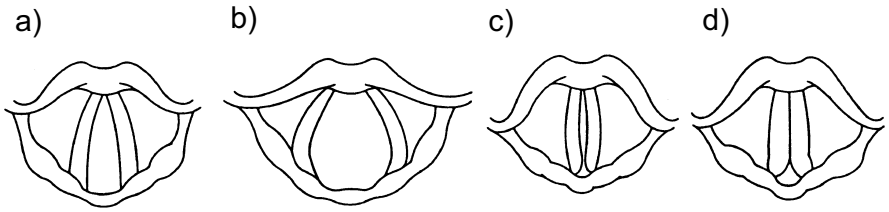

Abbildung 2.6: Stellung der Stimmbänder bei verschiedenen Funktionen: (a) normale Atmung, (b) tiefes Einatmen, (c) Phonation, (d) Flüstern. Quelle: [73].

Abb. 2.6 (c) zeigt die Stimmbänder während der stimmhaften Anregung des Vokaltrakts (Phonation). Die stimmhafte Anregung wird durch die myoelastisch-aerodynamische Theorie [72] erklärt. Dabei wird der pulmonale (von den Lungen kommende) Luftstrom durch das Schwingen der Stimmbänder so modifiziert, das eine quasiperiodische Anregung des Sprechtraktes für die Produktion stimmhafter Äußerungen entsteht. In der Voreinstellung sind die Stimmlippen leicht gespannt und die Stimmritze ist geschlossen (Abb. 2.7-(a),(b) Figur 4). Der subglottale Druck reicht dazu aus, die Stimmbänder zu öffnen, so dass Luft durch die Stimmritze entweichen kann (Abb. 2.7-(a),(b) Figuren 5 und 6). Deren Querschnitt ist jedoch im Vergleich zu den Querschnitten über und unter dem Kehlkopf gering, so dass die Luft dort mit großer Geschwindigkeit strömt. Dies erzeugt zwischen den Stimmlippen einen Unterdruck (Bernoulli-Effekt), der zusammen mit der natürlichen Rückstellkraft der Stimmbänder dafür sorgt, dass sie sich wieder aufeinanderzu bewegen. Der Verschluss der Glottis erfolgt abrupt (Abb. 2.7 Figuren 1-3), und es kann sich erneut ein genügend hoher subglottaler Druck aufbauen, um die Stimmbänder ein weiteres Mal auseinander zu sprengen. Dieser Zyklus kann solange fortgesetzt werden, wie der Luftstrom aus der Lunge aufrecht erhalten werden kann.
Wenn man die Stimmlippen im Längsschnitt wie in Abb. 2.7 (b) betrachtet, stellt man fest, dass sie sich entlang ihrer vertikalen Ausdehnung nicht gleichmäßig öffnen und schließen. Sowohl beim Öffnen als auch beim Schließen geht der untere Teil der Stimmlippen dem oberen voran, er schließt und öffnet sich also *eher* als der obere Teil. Auch in der mediosagittal-horizontalen Richtung – also entlang der Ausdehnung der Stimmritze selbst – schließen und öffnen sich beide Seiten nicht über die gesamte Länge zur gleichen Zeit. In [8] wird die Öffnung als ein kontinuierlicher Prozess beschrieben, bei dem sich zuerst der hintere Teil öffnet. Von dort aus werden die Stimmbänder ähnlich einem Reißverschluss bis nach vorne auseinandergezogen, bis sie komplett geöffnet sind. Der Verschluss verläuft analog von vorne nach hinten, wobei die Stimmlippen wie ein Reißverschluss zugezogen werden.

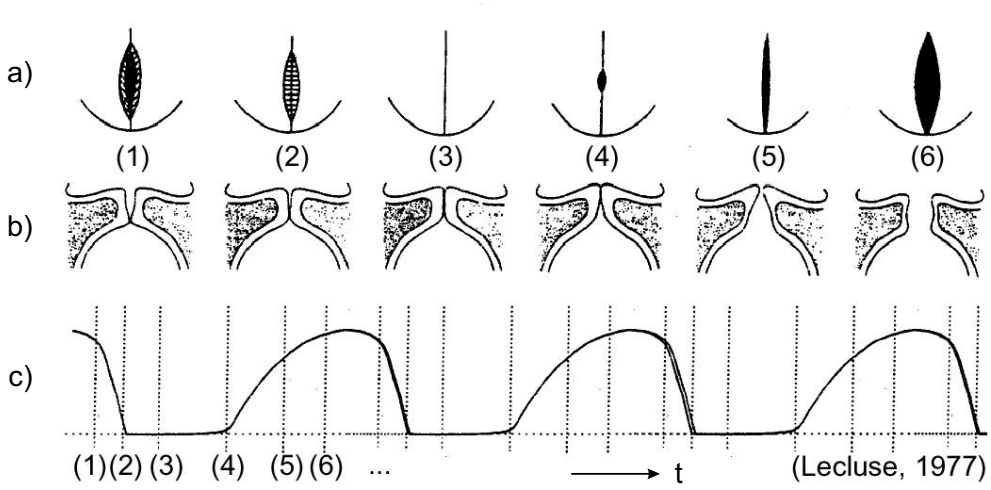

Abbildung 2.7: Schematische Ansicht der Bewegung der Stimmbänder. Bild (a) zeigt die Stimmritze von oben, (b) zeigt die Stimmlippen von vorne und (c) die Anregungsfunktion (Volumenstrom durch die Stimmritze). Quelle: [73].

Für das Sprachsignal ist es sehr wichtig, dass das Schließen im Gegensatz zum Öffnen sehr schnell bzw. abrupt passiert. Dadurch entsteht ein „Knick“ im Signalverlauf der Anregungsfunktion (siehe Abb. 2.7 (c)), der bewirkt, dass im stimmhaften Sprachsignal Frequenzen bis hinauf zu mehreren kHz merklich vorhanden sind.
Die Zyklusdauer der Stimmbandschwingung wird als Sprachgrundperiode T_0 bezeichnet. Die Sprachgrundfrequenz f_0 berechnet sich aus $f_0 = \frac{1}{T_0}$ und reicht bei normaler Phonation bei Männern von etwa 80 bis 200 Hz und bei Frauen von 150 bis 300 Hz.

Der Mensch ist in der Lage, verschiedene Stimmqualitäten bzw. Phonationstypen hervorzubringen. Diese unterscheiden sich bezüglich der phonatorischen Voreinstellungen der Muskelspannungen im Kehlkopf.

- Modale Stimmqualität (modal voice): Diese Stimmqualität ist durch eine gleichmäßige Schwingung der Stimmbänder gekennzeichnet. Dabei treten keine hörbaren Reibungsgeräusche an den Stimmlippen auf. Typischerweise sind die Stimmlippen während eines Teils des Stimmzyklus komplett geschlossen. Dieser Phonationstyp wird oft auch als „Bruststimme“ bezeichnet.

- Falsett (falsetto voice): Die f_0 ist gegenüber der modalen Stimme merklich erhöht. Die Stimmbänder sind stark gestreckt und dadurch dünner. Dies führt zu einer geringeren schwingenden Masse und folglich einer größeren Tonhöhe. Das Falsett wird häufig auch als „Kopfstimme“ bezeichnet.

- Rauhe Stimme (harsh voice): Die rauhe Stimme ist durch aperiodische Schwingungszyklen und stark hörbare Friktionsgeräusche gekennzeichnet. Die f_0 entspricht etwa der modalen Stimmqualität.

- Knarrstimme (creaky voice): Die Stimmlippen sind stark zusammengepresst, besitzen aber eine geringe Längsspannung. Dies führt zu einer starken Verdickung der

Stimmlippen, die nur am vorderen Ende und mit einer sehr geringen, unregelmäßigen Grundfrequenz schwingen.

- Behauchte Stimme (breathy voice): Die behauchte Stimme als Modifikation der modalen Stimme klingt ruhig und sanft. Während der Schwingzyklen wird die Glottis nicht vollständig geschlossen und es treten leichte Reibungsgeräusche auf.

- Flüsterstimme (whisper): Abb. 2.6 (d) zeigt die Stellung der Stimmlippen beim Flüstern. Dabei ist der membranartige Teil komplett verschlossen, während der knorpelartige Teil eine dreiecksförmige Öffnung besitzt. Während der Entweichung des pulmonalen Luftstroms durch das „Flüsterdreieck“ entsteht ein Friktionsgeräusch, welches als stimmlose Anregung des Vokaltrakts dient.

2.6 Untere Atemwege

2.6.1 Aufbau

Bei der Phonation bilden die Stimmlippen grundsätzlich die engste Stelle im gesamten Bereich der Sprechorgane (von der Lunge bis zur Mundöffnung). Die Sprechorgane werden durch sie in das *subglottale* System (untere Atemwege) und das *supraglottale* System (Rachen-, Mund- und Nasenraum) geteilt. Den obersten Abschnitt der unteren Atemwege bildet die Luftröhre (*Trachea*). Sie beginnt direkt unterhalb des Kehlkopfes und ist etwa 12 cm lang und 1,3 bis 2 cm weit. Sie besteht aus ca. 15-20 nach hinten offenen Knorpelspangen, die durch Bänder miteinander verbunden sind und den Luftgang offen halten sollen. Die Luftröhre endet in der Brusthöhle mit ihrer Teilung in die zwei Hauptbronchien, die zur linken und rechten Lunge führen (siehe Abb. 2.8). Der rechte Hauptbronchus verläuft steiler und ist kürzer und weiter als der linke Hauptbronchus. Er teilt sich in drei Äste, die zu den drei Lungenlappen der rechten Körperhälfte führen. Der linke Hauptbronchus teilt sich nur in zwei Äste für die zwei linken Lungenlappen. Die Verzweigungen setzen sich für jeden Ast über mehrere Stufen bis zu den Bronchiolen fort. Diese messen nur noch ca. 1 mm im Durchmesser.

Auch die Bronchiolen verzweigen sich weiter. Sie gehen in die sehr feinen Ästchen der *Bronchioli respiratorii* über, die wiederum direkt in die Lunge münden, in dem die eigentliche Atmung stattfindet. Dieser Bereich ist durch die Alveolargänge mit den Lungenbläschen (Alveolen) gekennzeichnet. Die Lungenbläschen liegen traubenförmig wie bei einem Weintraubenzweig um die Alveolargänge und die Bronchioli respiratorii. Insgesamt hat ein Mensch ungefähr 300 Millionen Lungenbläschen. Jedes einzelne ist von einem feinen Netz von Blutgefäßen umgeben, die für den Gasaustausch sorgen.

Von der Luftröhre bis zu den Bronchiolen erstreckt sich die baumartige Verzweigung über etwa 19 Stufen und von den Bronchiolen bis zu den Alveolargängen über ca. 4 weitere Stufen. Das gesamte Volumen der Lunge beträgt ungefähr 4800 ml, wovon 3150 ml auf die Alveolen entfallen.

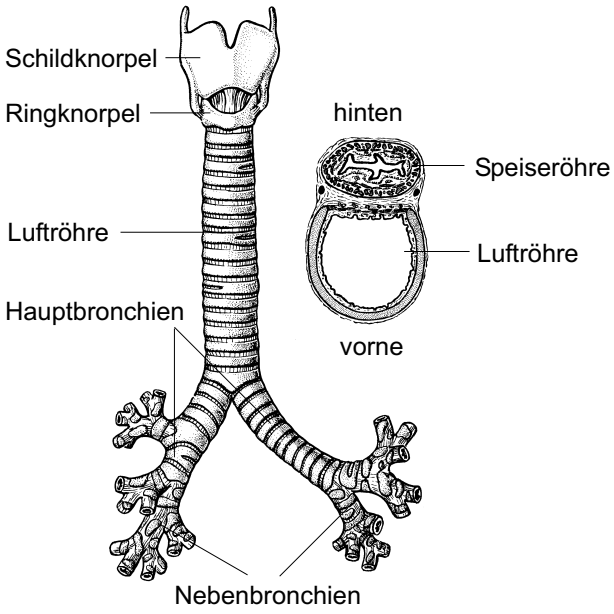

Abbildung 2.8: Luftröhre und Bronchien. Quelle: [6].

2.6.2 Funktion

Die unteren Atemwege dienen der Weiterleitung und Reinigung der Atemluft sowie dem Gasaustausch mit Kapillarblut.
Bei der Spracherzeugung sind sie für die Entstehung subglottaler Formanten verantwortlich [22, 34, 11], also für akustische Resonanzen, die sich unterhalb der Stimmlippen ausbilden. Da die unteren Atemwege gegenüber dem supraglottalen System eine weitestgehend statische Geometrie besitzen, haben auch die subglottalen Formanten konstante Frequenzen und Bandbreiten. In [34] wurden für die ersten drei subglottalen Formantfrequenzen etwa 640 Hz, 1400 Hz und 2100 Hz gemessen. Subglottale Formanten spielen im Sprachspekrum in der Regel eine untergeordnete Rolle, da bei normaler Phonation die glottale Impedanz sehr hoch ist. Schwache und behauchte Stimmen würden dagegen mehr Spuren subglottaler Formanten aufweisen.

Kapitel 3

Artikulatormodelle aus der Literatur

Es wurden bis heute eine Vielzahl unterschiedlicher Artikulatormodelle in der Literatur vorgestellt. Ihre Aufgabe ist es, Sprechtraktgeometrien zu generieren, aus denen die Querschnittsfunktion abgeleitet werden kann. Kröger [37] nimmt eine Einteilung der Artikulatormodelle in fünf Gruppen vor:

- Bei den *geometrisch orientierten* Modellen steht die direkte Parametrisierung der Formung des Ansatzrohres im Mittelpunkt. Dies bedeutet z.B. die direkte Vorgabe der Zylinderquerschnittsflächen im klassischen Röhrenmodell, wie es Fant [21] für eine Zwei-Sektionen-Modell und ein Vier-Sektionen-Modell getan hat.

- Bei den *effektorisch orientierten* Modellen steht die Modellierung der Artikulatoren im Vordergrund, die letztendlich die Sprechtraktform bestimmen. Modelle dieser Art stammen z.B. von Coker [9], Lindblom [38] und Mermelstein [42]. Auch das in dieser Arbeit entwickelte Modell fällt in diese Kategorie.

- Bei den *statistisch orientierten* Modellen ist die Formung des Sprechtrakts und die Anzahl der artikulatorischen Parameter das Resultat statistischer Auswertungen über einer Reihe phonetischer Messdaten. Ein Beispiel dafür ist das „quasiartikulatorische“ Modell von Meyer, Wilhelms und Strube [43]. Es liefert direkt die segmentierte Querschnittsfunktion als Linearkombination von nur wenigen Basisfunktionen zurück, die aus einer statistischen Analyse gewonnen wurden.

- Die *physiologisch orientierten* Modelle basieren auf einer physiologisch/biomechanischen Nachbildung der Sprechwerkzeuge. Wilhelms-Tricarico hat beispielsweise im Jahr 1995 Methoden zur Modellierung weicher Gewebestrukturen (Zunge, Lippen) vorgestellt, die auf Berechnungen mit finiten Elementen basieren [76]. Diese Methoden hat er erstmals auf die Simulation eines 3D-Modells der Zunge angewendet.

- *Akustisch orientierte* Artikulatormodelle basieren auf dem Wissen über die Zusammenhänge zwischen der Formung des Ansatzrohres und der Lage der Formantfrequenzen. Es wird dabei versucht, aufgrund einer geeigneten Modifikation des querschnittsneutralen Ansatzrohres die gewünschten akustischen Eigenschaften herbeizuführen. Die Theorie dafür wurde u.a. von Ungeheuer [71] entwickelt.

Im Folgenden soll lediglich zwischen den ein-, zwei- und dreidimensionalen Artikulatormodellen unterschieden werden. In den kommenden Abschnitten wird aus jeder Gruppe jeweils ein Vertreter aus der Literatur vorgestellt. Im Abschnitt 3.4 werden diese dann einer kurzen Beurteilung unterzogen.
Besonderes Augenmerk soll auf das Modell von Mermelstein (Abschnitt 3.2) gelegt werden, da dieses der Ansatzpunkt für die vorliegende Arbeit war.

3.1 Sechs-Parameter-Modell von Ishizaka

Ein eindimensionales Artikulatormodell mit sechs Freiheitsgraden wurde 1980 von J. L. Flanagan, K. Ishizaka und K. L. Shipley veröffentlicht [28]. Nach der Einteilung von Kröger gehört es in die Klasse der geometrisch orientierten Modelle. Es entstand im Rahmen eines Projektes mit dem Ziel, Sprache mit sehr niedrigen Bitraten zu kodieren. Das dazu entwickelte Computersystem versucht, für eine gegebene Spracheingabe eine möglichst genaue Replikation durch das Modell zu erzeugen. Dazu werden sowohl die Parameter ihres Modells als auch die Voreinstellungen der Glottis im zeitlichen Verlauf ständig neu geschätzt und optimiert. Die akustische Synthese erfolgt mit Hilfe des bereits 1972 von Ishizaka und Flanagan veröffentlichten Zwei-Massen-Modells der Stimmbänder [32] sowie einem elektrischen Analogon des Vokaltraktes.

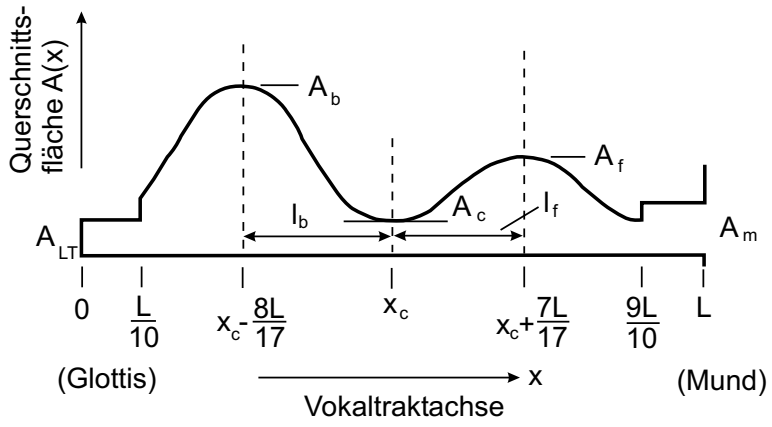

Abbildung 3.1: Sechs-Parameter-Modell. Nach [28].

Abb. 3.1 zeigt die Geometrie des eindimensionalen Modells. Es handelt sich um eine unmittelbare Parametrisierung der Querschnittsfunktion. Die Querschnittsfläche innerhalb des Vokaltrakts ist eine Funktion der Variablen x, die entlang der Vokaltraktachse von den Stimmbändern in Richtung der Lippen verläuft.
Die sechs Parameter des Modells sind A_b, A_f, A_c, x_c, A_m und L. L ist die Länge des Vokaltrakts und A_m der Flächeninhalt zwischen den Lippen. An der Position x_c befindet sich eine Engstelle mit dem Flächeninhalt A_c, die den Vokaltrakt in einen vorderen und einen hinteren Hohlraum unterteilt. Man kann sich den hinteren Hohlraum näherungsweise als Rachen und den vorderen Hohlraum als Mundhöhle vorstellen. Der Engpass entsteht dabei in der Regel dort, wo sich die Zunge am dichtesten der hinteren bzw. oberen Vokaltraktbewandung nähert. Mit den Flächeninhalten A_b und A_f werden der hintere

und vordere Teilraum dimensioniert, wobei die Positionen der maximalen Querschnitte in Abhängigkeit von L und x_c gegeben sind.
Die Querschnittsfläche $A(x)$ ist durch

$$A(x) = \begin{cases} \left(\frac{A_b+A_c}{2}\right) - \left(\frac{A_b-A_c}{2}\right)\cos\left(\pi \cdot \frac{x_c-x}{l_b}\right) & \text{falls } x \leq x_c \\ \left(\frac{A_f+A_c}{2}\right) - \left(\frac{A_f-A_c}{2}\right)\cos\left(\pi\left[0.4+0.6\left(\frac{x-x_c}{l_f}\right)\right]\left(\frac{x-x_c}{l_f}\right)\right) & \text{sonst} \end{cases} \tag{3.1}$$

gegeben, wobei $l_b = \frac{8L}{17}$ und $l_f = \frac{7L}{17}$. Außerdem gelten die folgenden Ungleichungen:

$$A_c \leq A_b, A_f, \quad A_c > 0, \quad 13 \leq L \leq 21 \text{ cm}, \quad \frac{L}{10} < x_c < \frac{9L}{10}$$

Der Querschnitt des hintersten (laryngalen) Abschnittes wurde auf 2 cm^2 gesetzt. Seine Länge wird – wie die des Mundvorhofs – mit $\frac{1}{10}$ der Vokaltraktlänge als konstant angesehen.
Die Geometrie des Modells wurde aus einer Reihe von sagittalen Röntgenaufnahmen von Gunnar Fant [21] abgeleitet und wird in [28] sowohl für die Repräsentation von vokalischen als auch konsonantischen Querschnittsfunktionen eingesetzt.

3.2 Modell von Mermelstein

Das 1973 von P. Mermelstein vorgestellte Artikulationsmodell [42] ist eine geometrische Beschreibung der Vokaltraktkontur in der mediosagittalen Ebene und fällt damit in die Klasse der effektorisch orientierten Modelle. Als Vorlage für das Modell dienten Röntgenaufnahmen vom Sprechtrakt während der Artikulation. Mermelsteins Arbeit zeigt, dass die Vokaltraktkontur – mit Hilfe verschiedener Variablen, die die Position des Unterkiefers, des Zungenkörpers, der Zungenspitze, der Lippen, des Velums und des Zungenbeins beschreiben – akkurat repräsentiert werden kann.
In der Arbeit von Mermelstein wurde mit dem Modell auch die temporale Änderung der Variablen während einer Äußerung durch Röntgenfilmmaterial bestimmt. Daraus wurde die mediosagittale Konturänderung des Vokaltraktes abgeleitet und eine Resynthese des Gesprochenen durchgeführt.
An dieser Stelle sollen nur die statischen Aspekte seines Modells betrachtet werden, während auf die Dynamik der Artikulation bewusst verzichtet wird.

3.2.1 Geometrie

Die von Mermelstein gewählten artikulatorischen Variablen bestimmen einzeln oder paarweise die Position der oben genannten Artikulatoren in der mediosagittalen Ebene. Lippen, Unterkiefer, Zunge, Velum und Zungenbein werden als bewegliche und teilweise voneinander abhängige Strukturen betrachtet. Die folgenden Beschreibungen beziehen sich auf Abb. 3.2 (a).
Vom Unterkiefer wird angenommen, dass er um einen festen Punkt F außerhalb des eigentlichen Vokaltraktes rotieren kann. Die Lage des Unterkiefers wird durch den Punkt J

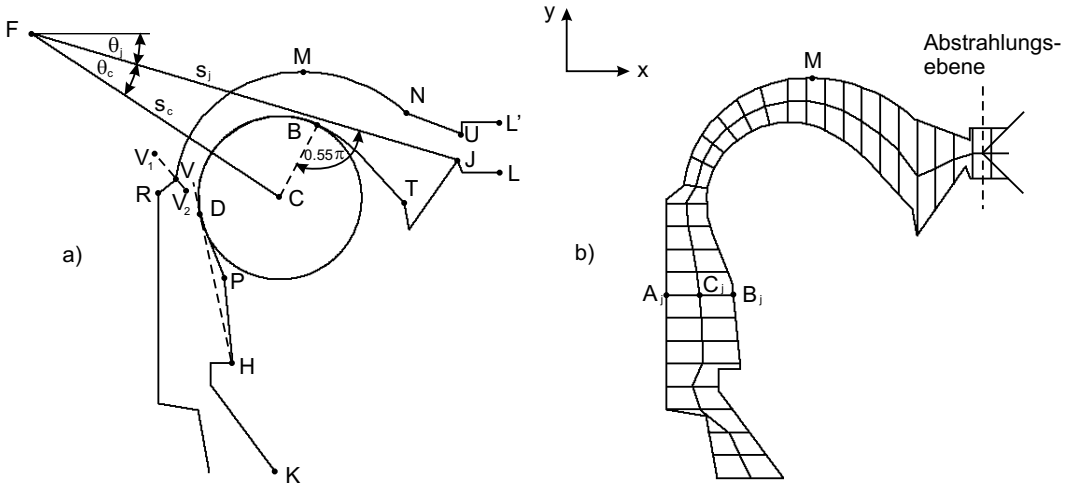

Abbildung 3.2: Modell von Mermelstein [42]. Bild (a) verdeutlicht die Konstruktion des Modells und Bild (b) die Berechnung der Mittellinie.

beschrieben, dessen Position relativ zum Punkt F in Polarkoordinaten (s_j, θ_j) angegeben wird. Der Abstand s_j zwischen beiden Punkten wird konstant bei 11.3 cm gehalten.
Das Zungenbein liegt am Punkt H und wird in absoluten Koordinaten angegeben. Bei einer Auf-/Abbewegung von H bewegt sich der gesamte unter H liegende Teil im gleichen Maße mit. Die Stimmbänder am Punkt K liegen konstant 2.7 cm unterhalb von H. Eine Vor-/Zurückbewegung von H zieht eine halb so weite Bewegung von K in die gleiche Richtung nach sich.
Der Zungenkörper wird durch einen Kreis mit beweglichem Mittelpunkt C und einem konstanten Radius von 2 cm repräsentiert. Die Position von C wird relativ zur Strecke $\overline{FJ}$ in den Polarkoordinaten (s_c, θ_c) angegeben.
Die Kontur der vorderen Rachenwand wird durch die Position von sowohl dem Zungenbein als auch dem Zungenkörper bestimmt. Wenn D der Berührungspunkt einer Tangente an die Rückseite des Zungenkreises ist, die außerdem durch H verläuft, dann wird der Punkt P auf der Normalen der Strecke $\overline{HD}$ abgetragen, die genau in der Mitte zwischen H und D beginnt. Der Abstand p des Punktes P zur Strecke $\overline{HD}$ ist eine Funktion von $\|\overrightarrow{HD}\|$ und berechnet sich nach $p = 0.57 \cdot (\|\overrightarrow{HD}\| - 3.48)$. Alle Angaben sind hier in cm gemacht.
Zungenspitze und -blatt bewegen sich relativ zum Punkt B auf dem Zungenkreis. B liegt so auf dem Kreis, dass $\overline{BC}$ und $\overline{FJ}$ einen konstanten Winkel von 0.55π einschließen. Die Position der Zungenspitze T wird relativ zu B in Polarkoordinaten (s_t, θ_t) angegeben, wobei θ_t sowohl eine Komponente für die aktive Anhebung/Senkung der Zungenspitze enthält, aber auch passiv von von θ_j und der Position des Zungenkörpers abhängig ist. Der Wert für s_t wird mit 3.4 cm für Vokale als konstant angegeben. Die Form des Zungenblatts wird durch eine Kurve beschrieben, die den Zungenkreis tangiert und durch T verläuft.
Die Form der Lippen wird durch zwei Variablen (p_l, h_l) beschrieben. Beide Variablen wirken sich gleichermaßen auf die Unter- und die Oberlippe aus. p_l ist ein Maß für die Vorstülpung bzw. die Zurückgezogenheit und h_l für die Offenheit bzw. Geschlossenheit der Lippen. L und L' werden folgendermaßen berechnet:

$$L = J + \begin{pmatrix} p_l \\ -h_l \end{pmatrix} \qquad L' = U + \begin{pmatrix} p_l \\ h_l \end{pmatrix} \tag{3.2}$$

Die Öffnungsfläche zwischen den Lippen ist also sowohl von der Position des Unterkiefers als auch dem Parameter h_l abhängig.
Der Punkt V markiert die Position des Zäpfchens. Der Zustand des Velums wird durch die Lage dieses Punktes beschrieben, der sich beliebig entlang der Strecke $\overline{V_1V_2}$ bewegen kann. Aus der Position von V wird die Öffnungsfläche vom Rachen- zum Nasenraum mittels der Formel $(V - V_1)^2$ berechnet.

Die restlichen Teile des Modells beschreiben die Kontur des Oberkiefers und der Rachenrückseite. Sie sind unveränderlich. M bezeichnet den höchsten Punkt des harten Gaumens und N teilt die Strecke von M nach U bei $\frac{2}{3}$ ihrer Länge. Die Kurven VM und MN sind Kreisbögen, deren Mittelpunkt auf einer vertikalen Linie durch M liegt. Die Punkte N und U werden durch eine Linie verbunden. Die Kontur der Rachenrückseite ist eine vertikale Linie durch den Punkt R, dessen Position fest vorgegeben ist.

Die zeitliche Steuerung der Modellparameter (s_j, θ_j, s_c, θ_c, ...) erfolgt mit a priori definierten Funktionen und kann in [42] nachgelesen werden.

3.2.2 Vokaltraktlänge und Querschnittsfunktion

Die Länge des Vokaltrakts und die Querschnittsfunktion werden mit Hilfe eines Gittersystems wie in Abb. 3.2 (b) bestimmt. Es besteht aus horizontalen Gitterlinien im pharyngalen Teil, aus radialen Linien im velaren Teil und aus vertikalen Linien im oralen Teil. Das Zentrum der radialen Gitterlinien liegt 3.85 cm senkrecht unterhalb des Punktes M, wobei sie dort jeweils im Winkel von 10° auseinander liegen. Die parallelen Gitterlinien haben jeweils einen Abstand von 0.5 cm.
Für jede der insgesamt N Gitterlinien wird der Schnittpunkt A_j mit der oberen/hinteren Kontur und der Schnittpunkt B_j mit der vorderen/unteren Kontur berechnet. Der Index j läuft von 0 (Stimmbänder) bis $N-1$ (Mundöffnung). Die Mittellinie des Vokaltrakts wird nun als Folge gerader Liniensegmente approximiert, die die Mittelpunkte C_j der Schnittlinien $\overline{A_jB_j}$ miteinander verbinden. Durch die Aufsummierung der Liniensegmentlängen ergibt sich die Vokaltraktlänge.
Die Länge einer Schnittlinie $\overline{A_jB_j}$ sei im Folgenden die „mediosagittale Distanz“ d_j genannt. Da es sich hier um ein zweidimensionales Modell handelt, ist es notwendig, von der mediosagittalen Distanz auf den Flächeninhalt für die Querschnittsfunktion zu schließen. Dies wird mit der Funktion $t(j, d_j)$ getan, die die mediosagittale Distanz an der Schnittlinie j mit Hilfe von empirischen Formeln auf die Flächeninhalte abbildet.
Da wir für die Querschnittsfunktion jedoch nicht an dem Flächeninhalt eines Schnittes durch das Modell *parallel zu den Gitterlinien* interessiert sind, sondern am Flächeninhalt eines Schnittes *senkrecht zur Mittellinie*, muß der Wert der Funktion $t(j, d_j)$ noch mit dem Faktor $\cos(\alpha)$ „korrigiert“ werden.

$$\hat{a}_j = t(j, d_j) \cdot \cos(\alpha) \tag{3.3}$$

Der Winkel α ist die Abweichung der Normalen der Mittellinie im Punkt C_j zur Ausrichtung der Schnittlinie $\overline{A_jB_j}$. Die Flächeninhalte $\hat{a}_j$ repräsentieren ungleichmäßig verteilte Abtastwerte der Querschnittsfunktion, die auf der Mittellinie um die linearen Abstände zwischen den Mittelpunkten C_j auseinanderliegen.
Die Transformation von der mediosagittalen Distanz zur Querschnittsfläche $t(j,d_j)$ wurde folgendermaßen definiert:

- Die Querschnittsfläche im pharyngalen Bereich wird durch eine Ellipse approximiert, deren eine Achse d_j ist und deren andere Achse vom Punkt *K* bis zum Punkt *V* von 1.5 bis 3 cm zunimmt.

- Im Bereich des weichen Gaumens sind die Querschnittsflächen $2 \cdot d_j^{1.5}$ cm^2 und im Bereich des harten Gaumens $1.6 \cdot d_j^{1.5}$ cm^2. Im Abschnitt des Zahndamms (zwischen dem Punkt *N* und den Schneidezähnen) gilt:

$$t(j,d_j) = \begin{cases} 1.5d_j & \text{für } d_j < 0.5 \\ 0.75 + 3(d_j - 0.5) & \text{für } 0.5 < d_j < 2 \\ 5.25 + 5(d_j - 2) & \text{für } d_j > 2 \end{cases} \tag{3.4}$$

- Im Bereich der Lippen wird von einer elliptischen Querschnittsfläche ausgegangen. Die Ellipse hat eine horizontale Ausdehnung von $2 + 1.5(s_l - p_l)$ und eine vertikale Ausdehnung von d_j. s_l ist die vertikale Distanz zwischen der Ober- und Unterlippe.

Wie man in Abb. 3.2 (b) erkennt, ist das Vokaltraktende nicht durch die tangentiale Begrenzung von Ober- und Unterlippe gegeben, sondern etwas ins Innere des Vokaltrakts zu den Mundwinkeln verlegt. Mermelstein schätzt die Position der Mundwinkel, indem er einen rechten Winkel derart in den Mund zeichnet, dass der Scheitel auf der Mittellinie liegt und die Schenkel an der Ober- und Unterlippe anliegen. Senkrecht durch diesen Scheitelpunkt verläuft die Schallabstrahlungsebene des Vokaltrakts.

3.3 3D-Vokaltraktmodell von Engwall

1999 wurde von Olov Engwall in [18] der damalige Entwicklungsstand eines dreidimensionalen, effektorisch orientierten Modells des Vokaltrakts vorgestellt. In den folgenden Jahren wurde das Modell weiter entwickelt und an die konkreten anatomischen Daten eines Referenzsprechers angepasst [19, 20]. In den folgenden beiden Abschnitten werden das „alte“ Modell und das „neue“ Modell (gegenwärtiger Entwicklungsstand) vorgestellt.

3.3.1 Altes Modell

Das alte Modell besteht aus einem Polygonnetz, welches in fünf Bereiche unterteilt ist, die die Vokal- und Nasaltraktwände, die Lippen, die Zähne und die Zunge repräsentieren. Das Netz wird durch 750 Punkte, die durch rund 1000 Polygone miteinander verbunden sind, definiert. Um die Komplexität zu reduzieren, ist es bezüglich der mediosagittalen Ebene symmetrisch angelegt.

Das Drahtgittermodell ist leicht mit den am TMH[1] entwickelten „Talking Heads“[2] kombinierbar, und soll damit in erster Linie als visuelle Unterstützung beim Training für hörgeschädigte Kinder und beim Lernen einer Fremdsprache dienen. Außerdem soll es für die artikulatorische Sprachsynthese eingesetzt werden.

Das Modell besitzt zehn artikulatorische Parameter, davon fünf für die Zunge und jeweils einen für die Larynxhöhe, die Kieferöffnung, die Lippenvorstülpung, die Lippenrundung und die Position des Velums. Die Beschreibung der Parameter folgt zum großen Teil den Ideen von Mermelstein [42], wurde aber den Gegebenheiten durch die Hinzunahme der dritten Dimension angepasst. Die Änderung der Parameterwerte wirken sich als Deformationen des Polygonnetzes aus. Dies geschieht durch die parametergesteuerte Bewegung eines virtuellen Punktes von einer Quell- zu einer Zielposition, der die in der Nachbarschaft liegenden Punkte des Netzes – je nach Gewichtung mehr oder weniger stark – in die Richtung der Zielposition zieht. Durch die Hinzunahme eines zweiten Pivotpunktes sind auch Rotationsbewegungen der Gitterpunkte möglich (hier im Falle der Kieferöffnung).
Die fünf Parameter der Zunge erlauben eine Vor-/Zurück und Auf-/Abbewegung des Zungenkörpers, eine Anhebung von Zungenspitze und -blatt, eine Hebung und Senkung der Zungenseiten gegenüber dem Zungenkörper und die Erzeugung einer Rille/Vertiefung auf dem Zungenrücken.
Die restlichen fünf Parameter des Modells entsprechen etwa denen des Mermelstein-Modells.

Die Berechnung der Querschnittsfunktion für die akustische Synthese ist unmittelbar aus dem 3D-Modell möglich. Zuerst wird dazu die Mittellinie mit Hilfe von 30 Paaren von Referenzpunkten bestimmt, die auf der oberen und unteren Kontur der Vokaltraktwand in der mediosagittalen Ebene liegen. Die Verbindung der Mittelpunkte jedes Referenzpunktpaares ergibt den stückweise linearen Verlauf der Mittellinie. Anschließend werden darauf 21 Punkte im gleichen Abstand von je 0.875 cm berechnet, die das Zentrum von 21 senkrecht zur Mittellinie orientierten Schnittebenen bilden. Eine Ausnahme sind nur der erste und letzte Punkt, deren Lage von der Länge des Vokaltraktes abhängig sind. Aus den Schnittflächen durch das Polygonnetz kann direkt die Querschnittsfunktion abgeleitet werden.

3.3.2 Neues Modell

Die Weiterentwicklung des im vorigen Abschnitt beschriebenen Modells bestand hauptsächlich in der Anpassung der Geometrie an die konkreten anatomischen Daten eines Referenzsprechers [19, 20]. Dazu wurden mit Hilfe eines Magnetresonanztomografen die Volumendaten seines Sprechtrakts bei der Artikulation von 13 schwedischen Vokalen und 10 Konsonanten erfasst. Aus den MRI-Daten wurden die Zunge, das Palatum, der Unterkiefer und die Vokaltraktwände dreidimensional rekonstruiert. Auf den 3D-

[1] Department of Speech, Music and Hearing (TMH) am Royal Institute of Technology (KTH) in Stockholm

[2] Talking Heads sind Computeranimationen sprechender Köpfe.

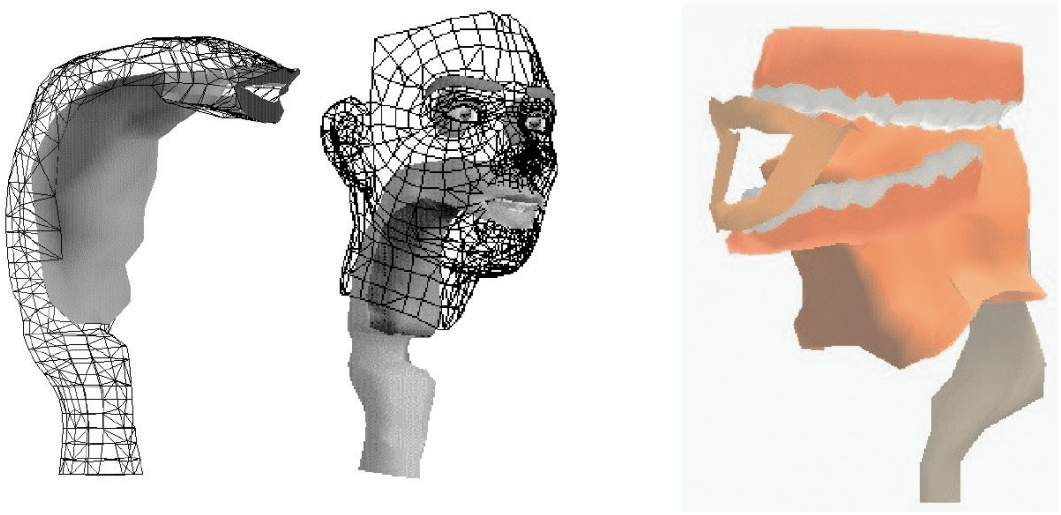

Abbildung 3.3: Altes Modell von O. Engwall ohne und mit Gesichtsnetz (links und Mitte) sowie neues Modell (rechts).

Datensätzen für die Zunge wurde eine lineare Komponentenanalyse durchgeführt, die sechs Parameter für die Rekonstruktion aller gemessenen Zungenformen zurückgeliefert hat. Die Vokaltraktwände, das Palatum und der Unterkiefer inklusive der oberen und unteren Zahnreihe wurden als statische Drahtgitternetze für die äußere Begrenzung des Vokaltraktes rekonstruiert (siehe Abb. 3.3). Einzig die Lippen wurden nicht aus den MRI-Daten rekonstruiert, sondern den Modellen für die Gesichtsanimation am KTH angepasst.

Weiterhin wurden die benötigten Algorithmen entwickelt, die die Kontaktbehandlung der Zunge mit den Wänden des Vokaltraktes abwickeln. Die Berechnung der Querschnittsfunktion geschieht analog zu dem im letzten Abschnitt beschriebenen Verfahren. Diese wurden in [20] einer akustischen Synthese zugeführt, um zu prüfen, inwiefern die künstlich erzeugten Formantfrequenzen mit denen des Referenzsprechers übereinstimmen. Es wurde ermittelt, dass das Modell in der Lage ist, die Vokale des Referenzsprechers in relativ guter Qualität zu replizieren, und dass sie in den meisten Fällen von Testhörern akzeptiert wurden. Größere Fehler wurden nur bei den vorderen Vokalen /i:/ und /e:/ gemacht, die auch von den Testhörern schlecht erkannt wurden. Die Formanten $F3$ und $F4$ lagen bei fast allen Vokalen zu niedrig.

3.4 Diskussion

An dieser Stelle soll eine Einschätzung der drei vorgestellten Modelle vorgenommen werden.

Das eindimensionale Modell von Ishizaka besitzt zweifellos den Vorteil der Einfachheit. Auch die Anzahl von nur sechs artikulatorischen Parametern kann als nahezu minimal angesehen werden. Trotz all der radikalen Vereinfachungen ist das Modell in der Lage, die in Fants Buch [21] angegebenen Querschnittsfunktionen mit akzeptabler

Genauigkeit zu approximieren.
Andererseits kann es jedoch viele geometrische Feinheiten, die auch akustisch bedeutsam sind, nicht akkurat nachbilden. Dies ist insofern nicht tragisch, als dass die Querschnittsfunktion $A(x)$ für die akustische Synthese entlang der x-Achse sehr grob abgetastet wird. Sie wird auf zehn Abschnitte konstanten Querschnitts der Länge $\frac{L}{10}$ reduziert. Eine akustische Simulation mit nur zehn Rohrabschnitten ist eine sehr grobe Approximation der wahren Verhältnisse, die durch weitere geometrische Details im Ausgangsmodell kaum verbessert werden kann.

Das zweidimensionale Modell von Mermelstein hat in der Praxis eine weite Verbreitung gefunden. Es ist exakter als ein eindimensionales Modell und die artikulatorischen Variablen besitzen einen direkten Bezug zur Physiologie des Vokaltrakts.
Durch die Beschränkung auf zwei Dimensionen müssen jedoch empirische Funktionen verwendet werden, die die mediosagittalen Distanzen aus dem Modell auf Flächeninhalte für die Querschnittsfunktion abbilden. Für diese Transformationen gibt es in der Literatur verschiedene Vorschläge [4, 42, 30], die in der Regel auf jeweils einen speziellen Sprecher zugeschnitten sind. Ein weiterer Nachteil ist die Tatsache, dass in zwei Dimensionen laterale Passagen an den Rändern der Zunge nicht modelliert werden können. Weiterhin wäre für eine exakte Berechnung der Schalldämpfung in der akustischen Simulation die Kenntnis vom Umfang des Vokaltrakts entlang der Mittellinie wünschenswert.

Das Vokaltraktmodell von Engwall gehört in der artikulatorischen Sprachsynthese zu den ersten Vorstößen in die dritte Dimension. Ab diesem Punkt ist es prinzipiell möglich, Sprachausgabe sehr hoher Qualität zu erzielen. Engwall hat außerdem gezeigt, dass auf heutigen PC's auch mit einem 3D-Modell Sprachsynthese in Echtzeit möglich ist. Dies wird auch im Rahmen der vorliegenden Arbeit bestätigt. Die Abweichungen der Formantfrequenzen zwischen den Vokalen des Modells und des Referenzsprechers sind wahrscheinlich eher auf Unzulänglichkeiten der akustischen Synthese als auf das Modell selbst zurückzuführen.

Die positiven Ergebnisse von Engwalls Untersuchungen – aber auch eigene Experimente mit dem Modell von Mermelstein – haben die vorliegende Arbeit zum großen Teil motiviert.

Kapitel 4

Entwicklung eines neuen Artikulatormodells

In diesem Abschnitt wird Schritt für Schritt das dreidimensionale Modell eines männlichen Vokaltrakts mit steuerbaren Artikulatoren entwickelt. Es gibt kein direktes menschliches Pendant für dieses Modell. Die Geometrie ist zum Teil dem Mermelstein-Modell nachempfunden und basiert zum anderen Teil auf eigenen Auswertungen von Röntgenfilmmaterial.[1]
Die Hauptaufgabe des Modells ist es, für jede beliebige Stellung der Artikulatoren die Querschnittsfunktion des Ansatzrohrs zu bestimmen. Bei dem hier vorgestellten Modell wird außerdem die Umfangsfunktion berechnet, die für jede Position auf der Mittellinie den Umfang des Ansatzrohrs angibt. Sie ist für die Berechnung der Energieverluste bei der akustischen Simulation von Bedeutung [26], wie in Abschnitt 5 gezeigt wird.
In Abschnitt 4.1 wird nun zunächst detailliert auf die Parameter und die Geometrie des Mund- und Rachenraums eingegangen. Anschließend werden in Abschnitt 4.2 die Methoden zur Berechnung der Mittellinie und der Querschnitts- und Umfangsfunktion vorgestellt. Der Nasenraum wird am Schluss getrennt behandelt.

4.1 Geometrie und Parameter des Mund- und Rachenraums

Die Geometrie des Modells hat verschiedene Anforderungen zu erfüllen. Einerseits soll sie die Form des Vokaltrakts realistisch wiederspiegeln und sich durch geeignete Parameter in physiologisch sinnvoller Weise verändern lassen. Die Artikulatoren sollen sowohl unabhängig voneinander bewegt werden können als auch Abhängigkeiten untereinander aufweisen. Das heißt z.B., dass eine Bewegung des Unterkiefers auch eine absolute Lageänderung der mit ihr verbundenen Zunge nach sich zieht, dass aber die Zunge auch aktiv ihre relative Position gegenüber dem Unterkiefer ändern kann.
Für die Berechnung der Querschnittsfunktion ist es erforderlich, dass das Vokaltraktmodell an jeder Position auf der Mittellinie mit einer zu ihr senkrechten Ebene geschnitten

[1] „X-ray Film Database for Speech Research“, bereitgestellt von Kevin Munhall (Queen’s University, Kingston, Canada).

werden kann, um aus den Schnittbildern die Flächeninhalte an den entsprechenden Stellen zu berechnen.

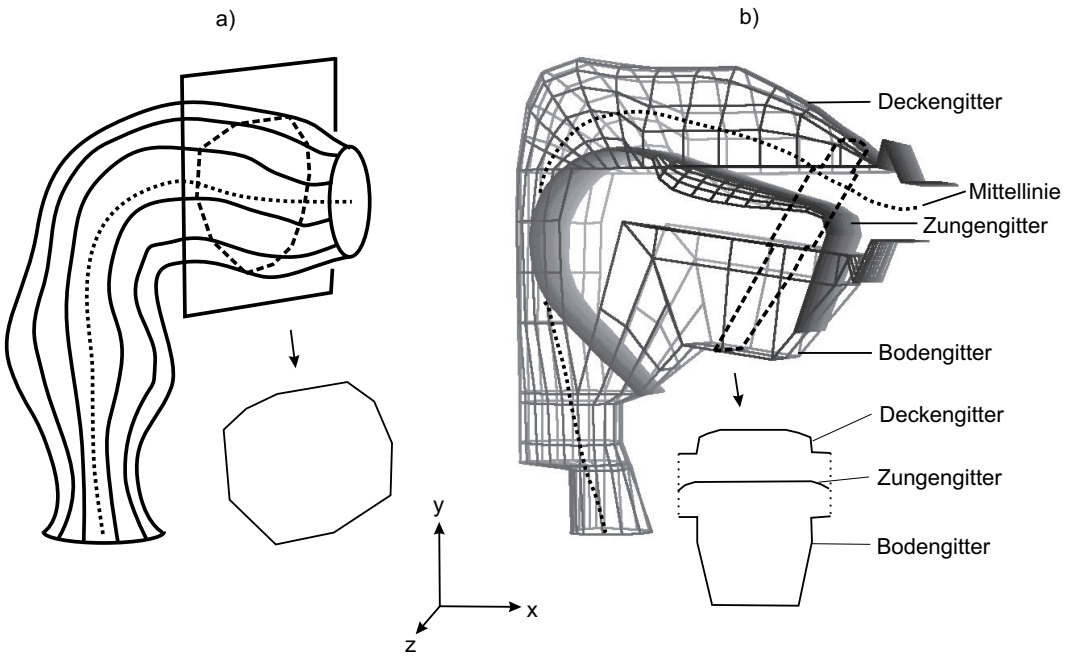

Abbildung 4.1: Von der Vorstellung des Vokaltrakts als Schlauch mit variablem Querschnitt in (a) zum Modell aus drei Punktgittern in (b).

Um diesen Forderungen in möglichst einfacher Weise nachzukommen, wurde der folgende Ansatz gemacht: Der Vokaltrakt wird direkt als ein gekrümmtes Rohr mit veränderlichem Querschnitt entlang seiner Mittellinie modelliert. Seine Mantelfläche wird durch mehrere dicht beieinanderliegende Linien definiert, die sich jeweils vom Anfang bis zum Ende des Rohres erstrecken und sich dabei nicht überschneiden (siehe Abb. 4.1 (a)). Sie werden im Folgenden als *Hauptlinien* bezeichnet. Jede von ihnen ist die stückweise lineare Verbindung von der gleichen Anzahl an Stützpunkten. Wenn N die Anzahl dieser Punkte ist und M die Anzahl der Hauptlinien, dann benötigt man genau $M \times N$ Punkte für die Beschreibung der Rohrmantelform. Die stückweise lineare Verbindung der jeweils gleichen Punkte jeder Hauptlinie wird als *Querlinie* bzw. als *Kranz* bezeichnet. Damit entspricht die Anzahl der Stützpunkte eines Kranzes der Anzahl der Hauptlinien und umgekehrt. Die Hauptlinien und Querlinien bilden zusammen ein *Gitter*.
Es hat sich als günstig herausgestellt, den Vokaltrakt statt mit nur einem mit drei solcher Gitter zu modellieren, da so eine übersichtliche Zuordnung von Haupt- und Querlinien zu den Artikulatoren möglich ist. Abb. 4.1 (b) zeigt die drei Vokaltraktgitter in einer perspektivischen Projektion.
Die Rückseite des Kehlkopfes und des Rachens, der harte und der weiche Gaumen sowie die Oberlippe werden vom *Deckengitter* geformt. Die Kehlkopf- und Rachenvorderseite, der Unterkiefer und die Unterlippe werden vom *Bodengitter* gebildet. Das *Zungengitter* dient der Formung der Zunge.
Für die Berechnung der Querschnittsfläche muss nun jedes der drei Gitter mit der senkrecht zur Mittellinie ausgerichteten Ebene zum Schnitt gebracht werden. Wenn man sich

die Schnittpunkte mit dem Boden- und Deckengitter zu einem einzigen Polygon verbunden denkt (mit Hilfe der gestrichelten Linien in Abb. 4.1 (b), unten), dann ist der gesuchte Flächeninhalt der Teil des Polygons über dem Zungengitter. Gemeinsam werden das Boden- und Deckengitter auch als *Mantelgitter* bezeichnet.
Die Gitterpunkte werden in einem kartesischen Koordinatensystem definiert, d.h. die x-Achse verläuft von links nach rechts, die y-Achse von unten nach oben und die z-Achse zeigt in Richtung des Betrachters. Es wird angenommen, dass der Vokaltrakt symmetrisch bezüglich der mediosagittalen Ebene aufgebaut ist. Dies wurde zwar durch MRI-Messungen widerlegt (besonders was die Zunge betrifft), hat aber nach allgemeiner Auffassung geringen Einfluss auf die Resonanzen des Vokaltrakts. Die mediosagittale Ebene entspricht in unserem Koordinatensystem der xy-Ebene durch den Koordinatenursprung. Deshalb genügt es, mit Hilfe der drei Gitter nur die linke Hälfte des Vokaltrakts zu modellieren. Die rechte Hälfte ergibt sich dann durch Spiegelung der Gitterpunkte.

	Anzahl Hauptlinien	Anzahl Querlinien
Bodengitter	5	18
Deckengitter	7	27
Zungengitter	12	36

Tabelle 4.1: Größe der Vokaltraktgitter (Hauptlinien$\times$Querlinien). Durch sie wird die linke Hälfte des Artikulatormodells repräsentiert.

Die Anzahl der Haupt- und Querlinien des Boden-, Decken- und Zungengitters sind in Tab. 4.1 gegeben.

4.1.1 Die Mantelgitter

Für die Beschreibung der Mantelgitter ist es sinnvoll, diese zuerst in Abschnitte zu zerlegen, die sich leicht getrennt voneinander betrachten lassen. Abb. 4.2 (a) zeigt die Vokaltraktkontur in der mediosagittalen Ebene. Die Unterteilung in Abschnitte (ohne die Zungenkontur) ist in Teilbild (b) gegeben. Ab dieser Stelle wollen wir eine Unterscheidung zwischen dem *Modellkoordinatensystem* und den *lokalen Koordinatensystemen* vornehmen. Die oben genannten Abschnitte werden teilweise zunächst in einem lokalen Koordinatensystem definiert und dann ins Modellkoordinatensystem transformiert, wo sie zu den Mantelgittern zusammengesetzt werden.
In Tab. 4.2 sind die geometriebestimmenden Parameter für den Vokaltraktmantel aufgelistet. Zu jedem Parameter ist außerdem das Intervall angegeben, durch das der jeweilige Wertebereich eingeschränkt wird. Die oberen und unteren Schranken der Wertebereiche habe ich mit Hilfe von Röntgenfilmaufnahmen eines Kopfes bei der Artikulation bestimmt.
Abb. 4.3 zeigt die Geometrie des harten und des weichen Gaumens in lokalen Koordinatensystemen. Da sie beide ein Teil des Deckengitters sind, muss die Anzahl der Hauptlinien jeweils mit der des Deckengitters übereinstimmen. Die Hautlinien werden in dieser

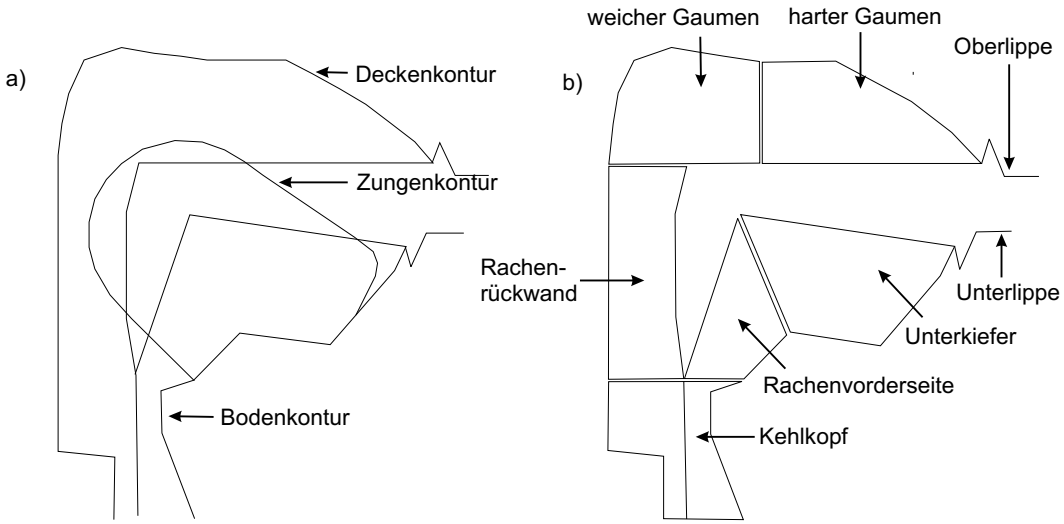

Abbildung 4.2: Zerlegung des Modells in seine Elemente (Darstellung in der mediosagittalen Ebene).

Bezeichnung	Wertebereich	Einheit	Bestimmung
LH	$[-0.8, 0.5]$	cm	Öffnung der Lippen
LP	$[0.8, 1.6]$	cm	Vorstülpung der Lippen
JP	$[-0.5, 0.3]$	cm	Vor-/Zurückverlagerung des Unterkiefers
JA	$[-0.15, 0.0]$	rad	Öffnungswinkel des Unterkiefers
VO	$[0.0, 1.0]$		Öffnungsgrad des velopharyngealen Ports
HX	$[2.0, 2.75]$	cm	absolute x-Position des Zungenbeins
HY	$[-7.0, -6.1]$	cm	absolute y-Position des Zungenbeins

Tabelle 4.2: Parameter für die Formung der Mantelgitter mit ihren oberen und unteren Schranken.

und den folgenden Abbildungen für jedes Teilgitter mit $H0, H1, \ldots$ bezeichnet. Die Querlinien werden analog in jedem Teilgitter mit $Q0, Q1, \ldots$ bezeichnet.
Der harten Gaumen ist unveränderlich und kann unmittelbar aus Abb. 4.3 rekonstruiert werden.
Vom weichen Gaumen dagegen sind zwei Zustände in der Seitenansicht abgebildet, von denen einer für den maximal geöffneten und einer für den maximal geschlossenen velopharyngalen Port (Öffnung zwischen Mund- und Nasenrachenraum) steht. Mittels des Parameters VO kann der aktuelle Zustand des Velums linear zwischen den beiden Extremzuständen interpoliert werden. Die Form von jedem der Kränze wird als eine viertel Ellipse approximiert, auf deren Umfang die Kranzpunkte gleichmäßig verteilt sind. Die eine Halbachse der Ellipse hat dabei die auf der Zeichnung erkennbare mediosagittale Länge und Ausrichtung des betroffenen Kranzes und die andere Halbachse nimmt von $Q0$ bis $Q6$ in ihrer Länge linear von 1.5 bis 1.8 cm zu und erstreckt sich in Richtung der negativen z-Werte. Die Kranzpunkte auf $H0$ besitzen hier analog zum Palatum die kleinsten z-Werte.

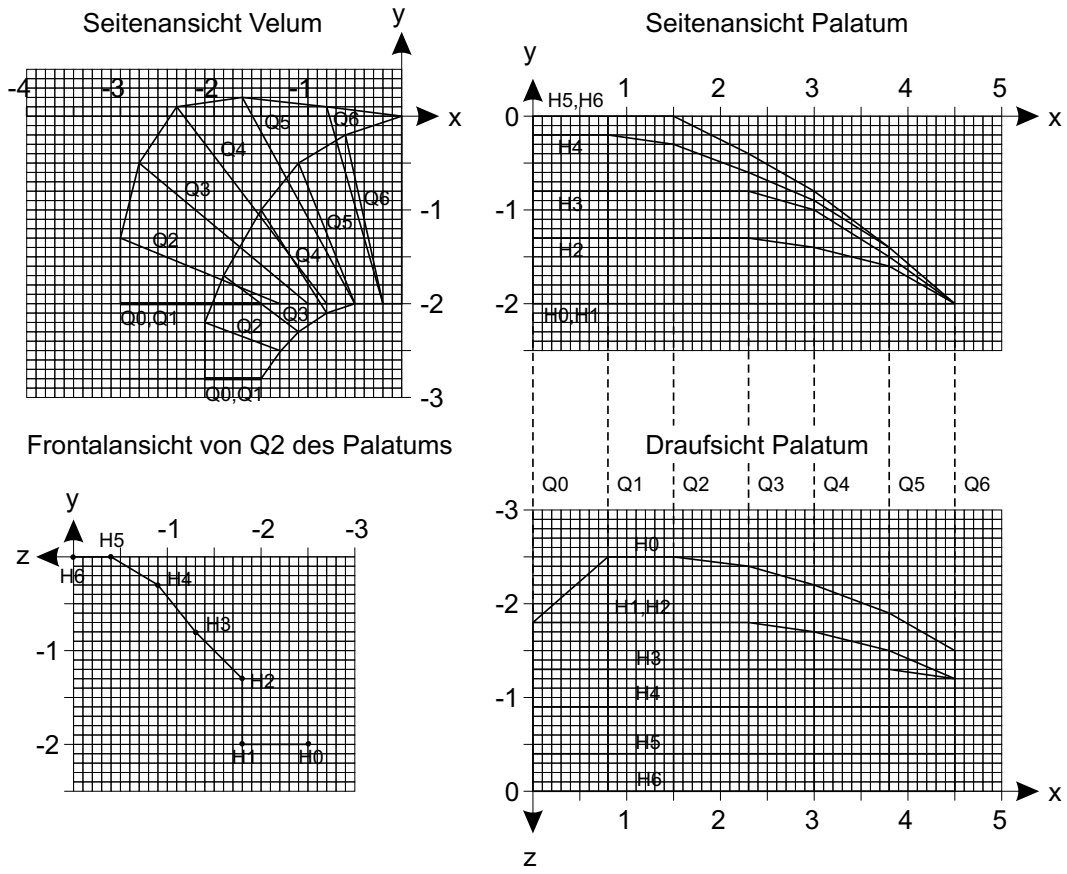

Abbildung 4.3: Geometrie des harten und weichen Gaumens.

Man beachte, dass die Kränze $Q0$ und $Q1$ für beide Zustände in der selben Ebene liegen. Für das geschlossene Velum sind sie identisch und für das geöffnete Velum unterscheiden sie sich in den x- und z-Koordinaten ihrer Punkte.

Als nächstes betrachten wir die Geometrie des Kehlkopfes in Abb. 4.4, wobei nur der Teil des Kehlkopfes ab den Stimmbändern aufwärts modelliert wird. Der Kehlkopf setzt sich aus zwei Teilgittern zusammen, von denen eins zum Deckengitter und das andere zum Bodengitter gehört. In der Abbildung ist die Grenze zwischen ihnen als eine gestrichelte Linie eingezeichnet. Sowohl vom Boden- als auch vom Deckengitter formt der Kehlkopf die ersten sechs Kränze.

Auch wenn das Kehlkopfskelett im Allgemeinen nicht sehr beweglich ist, kann man auf Röntgenaufnahmen leichte Variationen in der horizontalen Ausdehnung erkennen. Analog zum Velumgitter wird die aktuelle Kehlkopfgeometrie deshalb in Abhängigkeit vom Wert des Parameters HX zwischen zwei Extremzuständen interpoliert. Der Zustand in Abb. 4.4 (a) wird erreicht, wenn sich das Zungenbein maximal weit links befindet, bzw. HX seinen Minimalwert annimmt. Das Teilbild (b) zeigt den Zustand für den maximalen Wert von HX.

Im Teilbild (c) erkennt man, dass die Kranzformen von jedem der beiden Teilgitter auch hier als viertel Ellipsen approximiert werden. Die eine Halbachse der Ellipse ist dabei horizontal ausgerichtet und reicht von der gestrichelten Linie bis zur jeweiligen

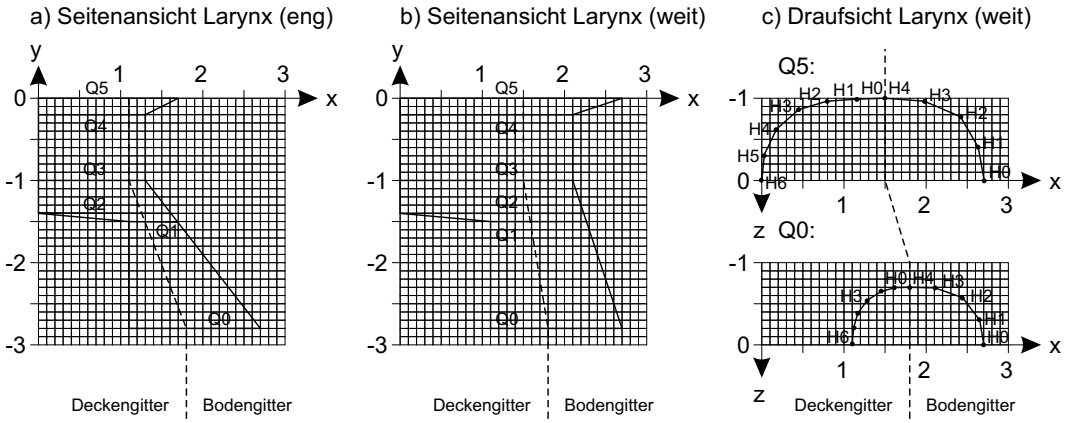

Abbildung 4.4: Geometrie des Kehlkopfes.

Kontur des Decken- oder Bodengitters. Die zweite Halbachse zeigt in die (negative) *z*-Richtung und ihre Länge wird linear zwischen 0.7 cm ($Q0$) und 1.0 cm ($Q5$) interpoliert.

Die Lippen des Modells bilden jeweils die letzten vier Kränze des Boden- und Deckengitters. Dabei ist der letzte Kranz des Palatums mit dem ersten Kranz der Oberlippe und der letzte Kranz des Unterkiefers mit dem ersten Kranz der Unterlippe identisch. Abb. 4.5 zeigt die Geometrie der Lippen relativ zu ihrem Ansatzpunkt am Palatum bzw. Unterkiefer in einem lokalen Koordinatensystem. Daraus ist ersichtlich, dass die Unterlippe lediglich eine Spiegelung der Oberlippe an der x-Achse ist. Außerdem erkennt man, dass sich die Punkte jedes Kranzes nur bezüglich ihrer *z*-Koordinate unterscheiden und deshalb für die Betrachtungen in der mediosagittalen Ebene auf jeweils einen einzelnen Punkt reduziert werden können.
Die Lage von $Q1$ relativ zu $Q0$ ist für beide Lippen konstant. Der Modellparameter LP gibt den horizontalen Abstand zwischen $Q1$ und $Q3$ an und entspricht der Länge $p+q$. Dabei stehen p und q in einem konstanten Verhältnis von 1:2. Die aktive Öffnung der Lippen wird durch den Parameter LH ausgedrückt und entspricht dem Abstand h. Durch die Lage des Unterkiefers wird die Lippenöffnungsfläche indirekt beeinflusst. Um auch bei einem gesenkten Unterkiefer einen Lippenverschluss zu ermöglichen, kann LH auch negative Werte annehmen.
Die Ausdehnung der Lippenkränze in Richtung der *z*-Achse wird im Folgenden als Halbbreite bezeichnet (die Gitter beschreiben nur die linke Hälfte des Vokaltrakts). Auf Abb. 4.5 sind die Halbbreiten in der Draufsicht zu erkennen. $Q0$ besitzt eine konstante Halbbreite von 1.5 cm und entspricht damit der Halbbreite von dem jeweils letzten Kranz des harten Gaumens und des Unterkiefers. Die Halbbreite von $Q3$ wird durch den experimentell angepassten Ausdruck

$$w = 1.5 - 1.3 \cdot \frac{LP - LP_{min}}{LP_{max} - LP_{min}} \tag{4.1}$$

berechnet. LP_{min} und LP_{max} sind die obere und untere Schranke des Wertebereichs von LP. Daraus ist ersichtlich, dass w einen Wert von 0.2 cm annimmt, wenn LP sein Maximum erreicht. Mit anderen Worten sind die Lippen an der Öffnung $2 \cdot 0.2$ cm $= 4$ mm breit, wenn sie maximal vorgestülpt sind. Die Halbbreiten der zwischen $Q0$ und $Q3$

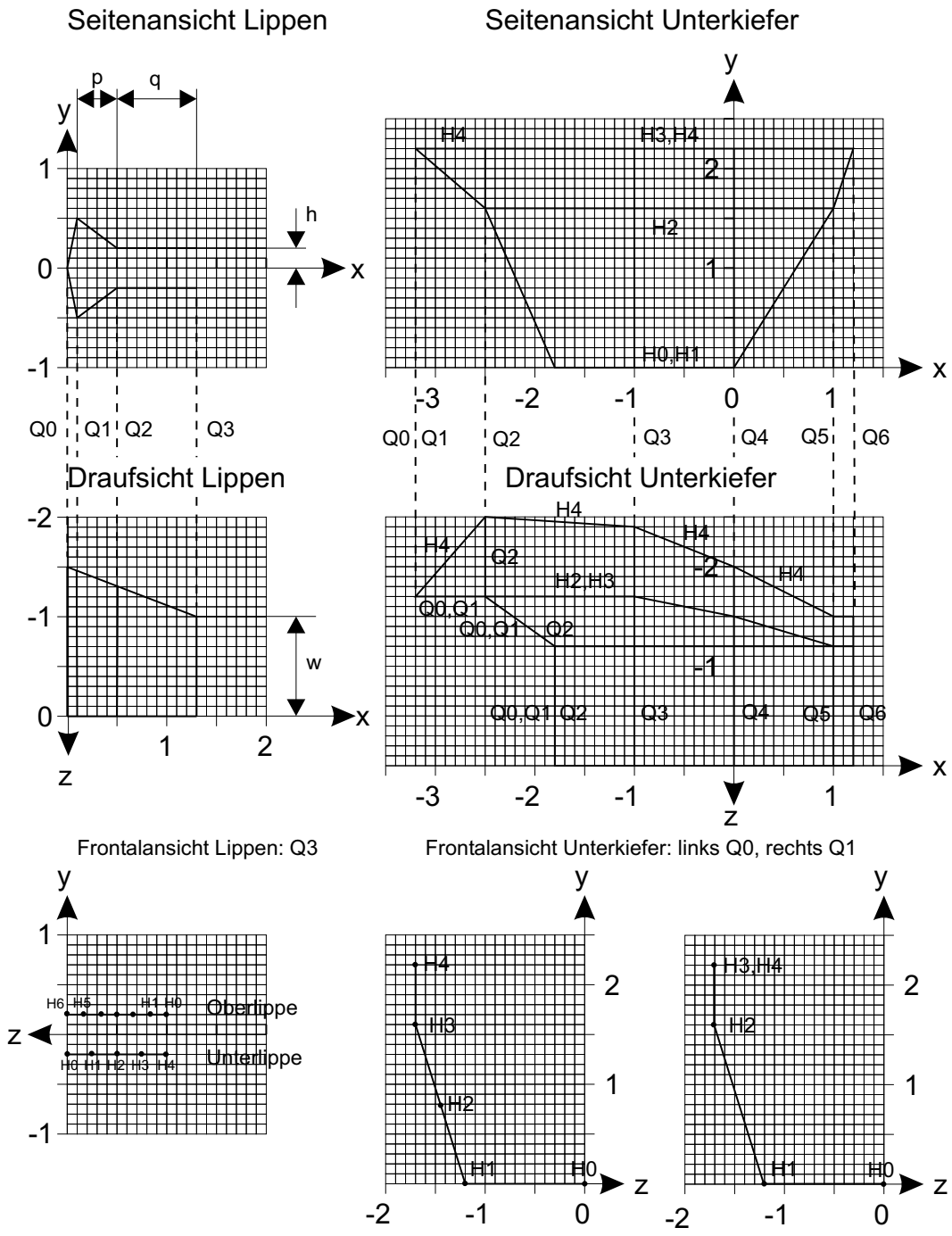

Abbildung 4.5: Geometrie der Lippen und des Unterkiefers.

liegenden Kränze werden linear interpoliert. Die Verteilung der Punkte auf jedem Kranz erfolgt gleichmäßig über die gesamte Halbbreite, analog zur Frontalansicht in Abb. 4.5.

Der Unterkiefer ist wie der harte Gaumen eine statische Konstruktion und wird in Abb. 4.5 auf der rechten Seite in einem lokalen Koordinatensystem definiert.
Das Kinn des Modells befindet sich im Koordinatenursprung. Die einzige Besonderheit sind die Kränze $Q0$ und $Q1$. Die durch sie definierten Linienzüge sind exakt identisch.

Dennoch erkennt man in der Frontalansicht, dass die Koordinaten der Stützpunkte auf $Q0$ von denen auf $Q1$ abweichen. Das sorgt für eine gleichmäßigere Verteilung der Hauptlinien zwischen der Kehlkopfvorderseite und dem Unterkiefer.

Wir wollen uns nun ansehen, wie die bis hier beschriebenen Teilgitter im Modellkoordinatensystem zum Boden- und Deckengitter zusammengeführt werden.

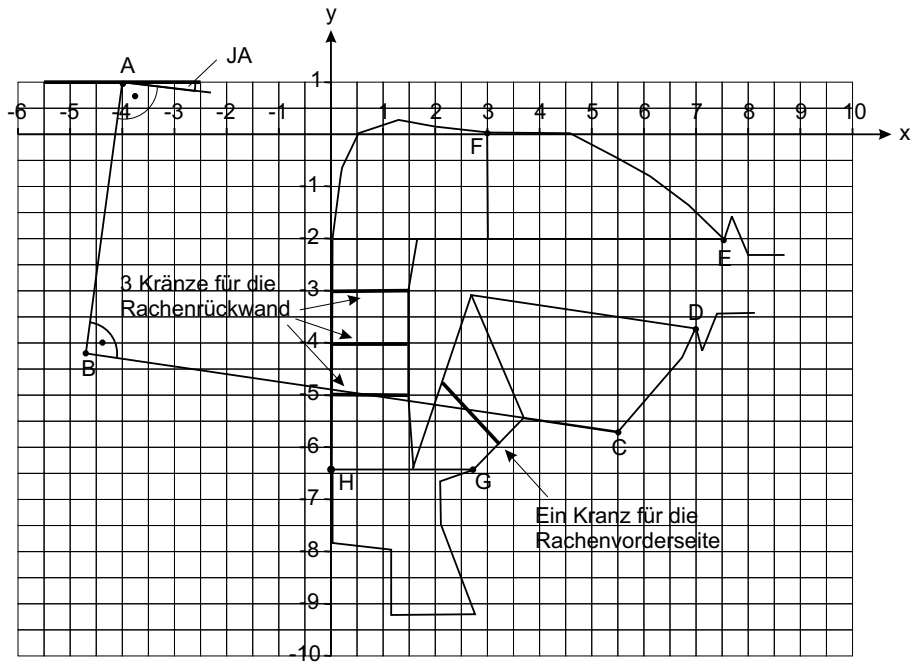

Abbildung 4.6: Komplette Kontur der Mantelgitter im Sagittalschnitt.

Das Kehlkopfgitter bildet wie breits erwähnt die ersten sechs Kränze von jedem der beiden Gitter. Zu diesem Zweck wird einfach der Koordinatenursprung des lokal definierten Kehlkopfes im Modellkoordinatensystem an den Punkt $H = (0, HY)^T$ verschoben. Der Parameter HY drückt also aus, wie stark der Kehlkopf bzw. das Zungenbein gesenkt oder gehoben ist.
Die nächsten drei Kränze des Deckengitters formen die Rachenrückwand. Ihre unveränderliche Lage ist in Abb. 4.6 eingezeichnet. Es ist selbstverständlich, dass die Anzahl der Hauptlinien durch diese Kränze genau der Anzahl der Hauptlinien des Deckengitters entsprechen muss. Es wird davon ausgegangen, dass die Kränze wieder die Form von viertel Ellipsen besitzen, deren horizontale Halbachsen 1.5 cm lang sind und deren Halbachsen parallel zur z-Achse eine Länge von 1.2, 1.4 und 1.5 für den unteren, mittleren bzw. oberen Kranz haben. Die darauf folgenden Kränze des Deckengitters werden durch das Velum und Palatum definiert. Dazu werden einfach beide lokal definierten Teilgitter mit ihrem Koordinatenursprung ins Modellkoordinatensystem an die Position $F = (3, 0)^T$ verschoben. An den letzten Kranz des harten Gaumens im Punkt E wird dann nahtlos die Oberlippe angesetzt. Insgesamt bekommen wir damit die in Tab. 4.1 angegebene Anzahl an Querlinien für das Deckengitter.

Für das Bodengitter folgt nach den sechs Kränzen des Kehlkopfes zunächst ein

einzelner Kranz für die Rachenvorderseite. Seine Punkte sind die Mittelwerte der Punkte des unmittelbaren Vorgängerkranzes ($Q5$ der Kehlkopfvorderseite) und des unmittelbaren Nachfolgerkranzes ($Q0$ des Unterkiefers). Er dient lediglich der deutlicheren Visualisierung der Rachenvorderseite und könnte theoretisch auch weggelassen werden. Die darauf folgenden Kränze des Bodengitters werden vom Unterkiefer geformt. Bevor das lokal definierte Unterkiefermodell mit seinem Ursprung an den Punkt C in Abb. 4.6 verschoben wird, führt es eine Rotation um den Winkel JA aus. Der Modellparameter JA (in Radianten) kann nur Null oder negative Werte annehmen. Die Lage des Punktes C ergibt sich, indem der am Punkt A „aufgehängte“ rechten Winkel ABC um den Winkel JA rotiert wird. Dabei gilt, dass $\|\overrightarrow{AB}\| = 5.2$ cm und $\|\overrightarrow{BC}\| = 10.3 + JP$ cm, d.h ein Vor- oder Zurückverlagerung des Unterkiefers wird mit Hilfe des Parameters JP ausgedrückt. An den letzten Kranz des Unterkiefers wird anschließend die Unterlippe angefügt. Damit kommen wir auch für das Bodengitter auf die in Tab. 4.1 genannte Anzahl an Querlinien.

4.1.2 Das Zungengitter

An das Modell einer Zunge werden bezüglich der Flexibilität besondere Anforderungen gestellt. Deshalb wird die Zungengeometrie derzeit mit Hilfe von 11 Parametern beschrieben, die in Tabelle 4.3 aufgeführt sind.

Bezeichnung	Wertebereich	Einheit	Bestimmung
TCA	$[1.47, 3.14]$	rad	Polarwinkel des Zungenkreismittelpunkts
TCL	$[2.5, 6.0]$	cm	Polarradius des Zungenkreismittelpunkts
TCR	$[1.5, 2.25]$	cm	Radius des Zungenkreises
TBA	$[3.14, 4.71]$	rad	Polarwinkel des Zungenspitzenmittelpunkts
TBL	$[1.9, 5.0]$	cm	Polarradius des Zungenspitzenmittelpunkts
TBX	$[0.2, 0.8]$		Ort der max. Wölbung des Zungenblattes
TBY	$[-2.0, 2.0]$	cm	Stärke der Wölbung des Zungenblattes
TAX	$[0.2, 0.8]$		Ort der max. Wölbung des Zungenrückens
TAY	$[-2.0, 2.0]$	cm	Stärke der Wölbung des Zungenrückens
RTW	$[0.0, 1.0]$		Breite der Zungenspitze
RTG	$[0.0, 1.0]$		Furchenbildung am Zungenblatt

Tabelle 4.3: Parameter für die Formung des Zungengitters mit den jeweiligen Wertebereichen.

Im Vergleich zu anderen Zungenmodellen ist das eine sehr hohe Anzahl, denn in der Regel reichen bereits zwei bis vier Parameter aus, um die wesentlichen mediosagittalen Zungenkonturen für vokalische Laute zu determinieren. Mit dem vorgestellten Modell sollen jedoch nach Möglichkeit die für *alle* Sprachlaute benötigten Zungengeometrien nachgebildet werden können. Dazu gehören auch Merkmale, die auf die Hinzunahme der dritten Dimension zurückzuführen sind, also z.B. die lateralen Passagen beim /l/ oder die Furchenbildung entlang der Zungenmittellinie beim /s/ oder /sch/. Dennoch ist die vorgeschlagene Parametrisierung mit Sicherheit umfangreicher als unbedingt notwendig,

da Abhängigkeiten der Parameter untereinander bisher noch nicht berücksichtigt wurden. Bei einer Weiterentwicklung des Modells könnte die Parameteranzahl also noch reduziert werden.

Die Berechnung des Zungengitters ist im Wesentlichen ein zweistufiger Prozess. Im ersten Schritt wird eine mediosagittale Kontur berechnet, wie es bei den herkömmlichen zweidimensionalen Modellen getan wird. Im zweiten Schritt werden dann die Zungenkranzformen für die verschiedenen Positionen entlang der Kontur berechnet und zu einem dreidimensionalen Modell zusammengeführt.
Die folgende Beschreibung bezieht sich auf Abb. 4.7, die die Berechnung der Zungenkontur in der *xy*-Ebene verdeutlicht. Die Kontur setzt sich aus zwei Kreisbögen und zwei quadratischen Splines zusammen. Die Kreisbögen sind Teil eines großen und eines kleinen Kreises. Der große Kreis repräsentiert den „Zungenkörper“ und der kleine Kreis die Zungenspitze. Der Zungenrücken wird durch den ersten Spline gebildet, der das Zungenbein (S_0) mit dem großen Kreis verbindet. Der zweite Spline repräsentiert das Zungenblatt und verbindet den großen mit dem kleinen Kreis.

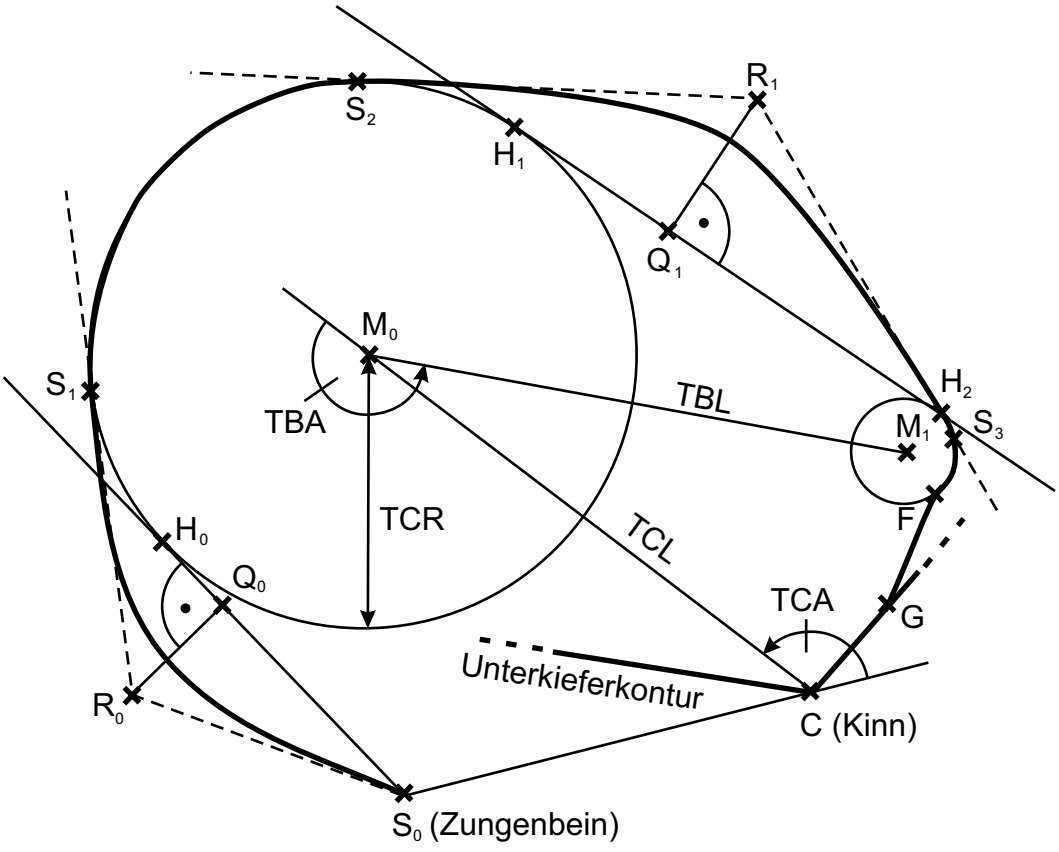

Abbildung 4.7: Kontur der Zunge in der mediosagittalen Ebene.

Zuerst werden die Mittelpunkte der beiden Kreise berechnet. Dabei gehen wir von den Punkten S_0 und C aus, die auf der ersten Hauptlinie des Bodengitters liegen. Das Zungenbein S_0 liegt auf dem letzten Kranz der Vorderseite des Kehlkopfgitters. C repräsentiert das Kinn und liegt auf $Q5$ des Unterkiefergitters.

Wenn man den Vektor $\overrightarrow{S_0C}$ um den Winkel TCA dreht, auf die Länge TCL bringt und zum Punkt C addiert erhält man den Mittelpunkt M_0 des ersten Kreises. Wenn man dann den Vektor $\overrightarrow{CM_0}$ um den Winkel TBA dreht, auf die Länge TBL bringt und zu M_0 addiert, erhält man den Mittelpunkt M_1 des zweiten Kreises. Der große Kreis besitzt den Radius TCR, während der Radius des kleinen Kreises mit 4 mm konstant ist.
Die Position des Zungenkörpers wird also mit Hilfe der Parameter TCA und TCL relativ zur Position des Zungenbeins und des Unterkiefers definiert. Analog ist die absolute Lage der Zungenspitze sowohl von den Parametern TBA und TBL als auch von der Position des Zungenkörpers abhängig. Mittels des Parameters TCR wurde der Radius des Zungenkörpers variabel gestaltet.
Nun können die zwei Splines berechnet werden. H_0 ist der Berührungspunkt der Tangente durch den Punkt S_0 an die „Rückseite“ des großen Kreises. Der Parameter TAX definiert die Lage von Q_0 mittels der Gleichung $Q_0 = TAX \cdot H_0 + (1 - TAX) \cdot S_0$. Mit Hilfe von Q_0 und dem Parameter TAY kann der Punkt R_0 berechnet werden. Dazu wird der Vektor $\overrightarrow{S_0H_0}$ um 90° nach links gedreht, normiert, mit dem Wert von TAY multipliziert und anschließend zu Q_0 addiert. S_1 ist nun wiederum der Berührungspunkt der Tangente durch R_0 an die Rückseite des großen Kreises. Die Punkte S_0, R_0 und S_1 definieren den ersten Spline.
Der zweite Spline wird berechnet, indem zuerst eine Gerade tangential an die Oberseite der beiden Kreise gelegt wird. Mit Hilfe der Berührungspunkte H_1 und H_2 kann $Q_1 = TBX \cdot H_2 + (1 - TBX) \cdot H_1$ berechnet werden. Davon ausgehend wird analog zum ersten Spline R_1 berechnet, nur statt mit TAY mit TBY. Die Berührungspunkte der Tangenten durch R_1 an die Oberseite der beiden Kreise seien S_2 und S_3. S_2, R_1 und S_3 definieren den zweiten Spline.
Die beiden Splines zusammen mit den Kreisbögen S_1S_2 und S_3F ergeben die eigentliche Zungenkontur. Die Punkte S_3 und F liegen auf dem kleinen Kreis um 90° auseinander.

Nun werden in gleichmäßigen Abständen 33 Punkte auf der Konturlinie berechnet. Zu jedem Punkt wird zusätzlich ein normierter Vektor bestimmt, der in dem Punkt senkrecht auf der Kontur steht. Die Vektoren zeigen dabei in Richtung des Deckengitters. Die Punkte und Vektoren seien im Folgenden mit P_i bzw. $\overrightarrow{v_i}$ $(0 \leq i < 33)$ bezeichnet. In Abb. 4.7 entspricht P_0 dem Punkt S_0 und P_{32} dem Punkt F.

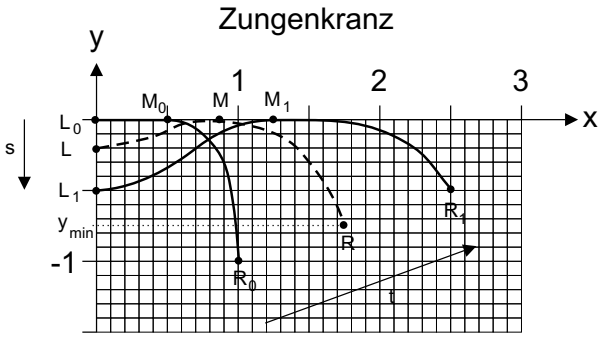

Abbildung 4.8: Berechnung eines Zungenkranzes (gestrichelte Linie). Der Parameter *s* bestimmt die Tiefe der Furche in der Zungenmitte und der Parameter *t* ist ein Maß für die Breite des Zungenkranzes.

Für jedes P_i wird dann die Querlinie des Zungengitters an der entsprechenden Stelle berechnet.
Abb. 4.8 zeigt, wie die Zungenkränze in einem lokalen Koordinatensystem definiert werden. Die Lage der Punkte L_0, R_0 und R_1 ist konstant. Die Parameter s und t sind für jeden Zungenkranz variabel. t ist ein Maß für die Breite des Zungenkranzes und s ein Maß für die Tiefe der Furche bzw. Rille in der Zungenmitte.
Jede Kranzlinie wird durch zwei Punkte – L und R (die Endpunkte der gestrichelten Linie in Abb. 4.8) – eindeutig beschrieben. Die zwei Parameter s und t (beide $\in [0,1]$) bestimmen die Lage dieser Punkte. Dabei gelten folgende Beziehungen:

$$\begin{aligned} R &= R_0 + t(R_1 - R_0) \\ L_1 &= (0, -0.5t)^T \\ L &= L_0 + s(L_1 - L_0) \end{aligned} \tag{4.2}$$

Mit dem Hilfspunkt $M = (0.5R_x, 0)^T$ in Abb. 4.8 wird der Kranz in einen linken und einen rechten Abschnitt unterteilt.
Der linke Teil der Linie wird durch ein Polynom $f(x)$ dritten Grades beschrieben. Unter der Voraussetzung, dass $f(x)$ durch die Punkte L um M verläuft, und dass die erste Ableitung der Funktion in beiden Punkten gleich Null ist, ergibt sich

$$f(x) = \frac{2L_y}{M_x^3}x^3 - \frac{3L_y}{M_x^2}x^2 + L_y. \tag{4.3}$$

Der rechte Teil der Linie wird durch eine Parabel g(x) beschrieben, die durch die Punkte M und R geht und deren erste Ableitung im Punkt M gleich Null ist:

$$g(x - M_x) = \frac{R_y}{(R_x - M_x)^2}(x - M_x)^2. \tag{4.4}$$

Nun kommen die Zungenparameter RTW und RTG ins Spiel. RTW ist genau der Parameter t und RTG der Parameter s für den Zungenkranz 32, also den letzten Kranz auf der Zungenkontur. Für die Zungenkränze 0 bis 15 werden s und t derzeit mit $t = 1$ und $s = 0$ konstant gehalten. Für die Kränze 16 bis 31 werden s und t linear zwischen den Werten für Kranz 15 und 32 interpoliert.
Der Parameter RTW drückt damit aus, wie spitz die Zunge nach vorne hin zuläuft und RTG ist ein Maß für die Furchenbildung des vorderen Zungenteils. Mit Hilfe dieser Parameter lassen sich die lateralen Passagen beim [l] und die typischen Furchen bei den vorderen Zischlauten nachbilden.

Als nächstes muss vermieden werden, dass ein Zungenkranz breiter ist als die dortigen Vokaltraktwände es zulassen. Dazu werden die Mantelgitter für jeden Kranz i mit einer Ebene durch den Punkt P_i geschnitten, so dass aus der Querschnittsfläche die maximal erlaubte Breite des Kranzes ermittelt werden kann. Anschließend werden die Kranzpunkte mit Hilfe der Funktionen $f(x)$und $g(x)$ berechnet. Der erste Kranzpunkt hat dabei im lokalen Koordinatensystem den x-Wert Null – liegt also in der Mitte der Zunge – und der letzte Punkt hat als x-Wert die maximale Breite des Kranzes.

Die Transformation eines beliebigen Punktes $Q = (x,y)^T$ des Kranzes i im lokalen Koordinatensystem zum Punkt $Q' = (x',y',z')^T$ im Modellkoordinatensystem ist durch

$$\begin{aligned} x' &= P_{i,x} + \vec{v}_{i,x} \cdot y \\ y' &= P_{i,y} + \vec{v}_{i,y} \cdot y \\ z' &= -x \end{aligned} \tag{4.5}$$

gegeben.
Dabei kann es jedoch in Bereichen starker Krümmung der Zungenkontur dazu kommen, dass sich in der mediosagittalen Ebene benachbarte Kränze überschneiden. Dies kann besonders dann passieren, wenn ein Kranz (im lokalen Koordinatensystem betrachtet) sehr schmal und tief ist, also einen kleinen Wert für y_{min} besitzt. Abb. 4.9 (b) zeigt, wie in solchen Fällen der Vektor $\vec{v}_i$ derart „korrigiert“ wird, dass keine Überschneidung mehr auftritt. Wenn für den Kranz i festgestellt wird, dass seine mediosagittale Projektion einen gemeinsamen Schnittpunkt mit einem vorhergehenden oder nachfolgenden Kranz aufweist, dann wird $\vec{v}_i$ minimal dahingehend verändert, dass sich die geschnittenen Kränze nur noch in einem Punkt berühren. Jede Korrektur verändert einen Vektor $\vec{v}_i$ so, dass er nicht mehr genau senkrecht im Punkt P_i auf der Zungenkontur steht. Die Anzahl der notwendigen Korrekturen sollte also möglichst gering gehalten werden.

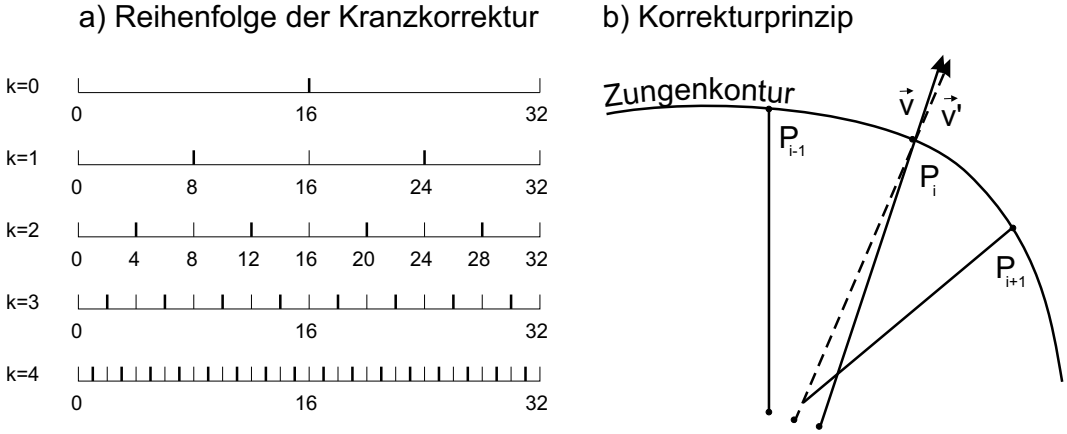

Abbildung 4.9: In Bild (a) sind für jeden Schritt $k = 0,1,\ldots$ jeweils diejenigen Kränze fett gezeichnet, die auf eine Überschneidung mit den dünn gezeichneten Nachbarkränzen getestet werden. Bild (b) zeigt die Korrektur des mittleren Kranzes, der seinen rechten Nachbarn schneidet. Das Ergebnis der Korrektur ist die gestrichelte Linie.

Ein wichtiger Punkt ist dabei die Reihenfolge, in der die 33 Zungenkränze auf Konflikte mit einem linken und rechten Nachbarn getestet werden. Wenn die Kränze vom ersten bis zum letzten *nacheinander* durchgetestet werden würden, könnte sich unter Umständen ein „Fehler“ fortpflanzen, d.h. eine Anpassung des Kranzes i an seinen linken Nachbarn könnte im nächsten Schritt eine Anpassung des Kranzes $i+1$ an den Kranz i erfordern.
Deshalb habe ich mich entschieden, die Anpassungen in fünf Schritten vorzunehmen (siehe Abb. 4.9 (a)). Im ersten Schritt ($k = 0$) wird nur der Kranz 16 auf Schnittpunkte mit dem ersten und dem letzten Kranz getestet. Im nächsten Schritt ($k = 1$) werden die Kränze 8 und 24 gegenüber ihren linken und rechten Nachbarn 0 und 16 bzw. 16 und 32

korrigiert. Dieses Verfahren wird bis $k = 4$ weiter fortgesetzt. Das erklärt auch, warum ich mich für genau 33 Kränze entlang der Konturlinie entschieden habe.

Anschließend werden noch die drei letzten Kränze 33, 34 und 35 des Zungengitters berechnet, die eine „saubere“ Verbindung der Zungenspitze mit dem Unterkiefer herstellen sollen.
Abb. 4.10 (b) zeigt ihre Lage in einer 3D-Projektion. Für jeden dieser Kränze sind jeweils die *x*-und *y*-Koordinaten aller Punkte gleich.

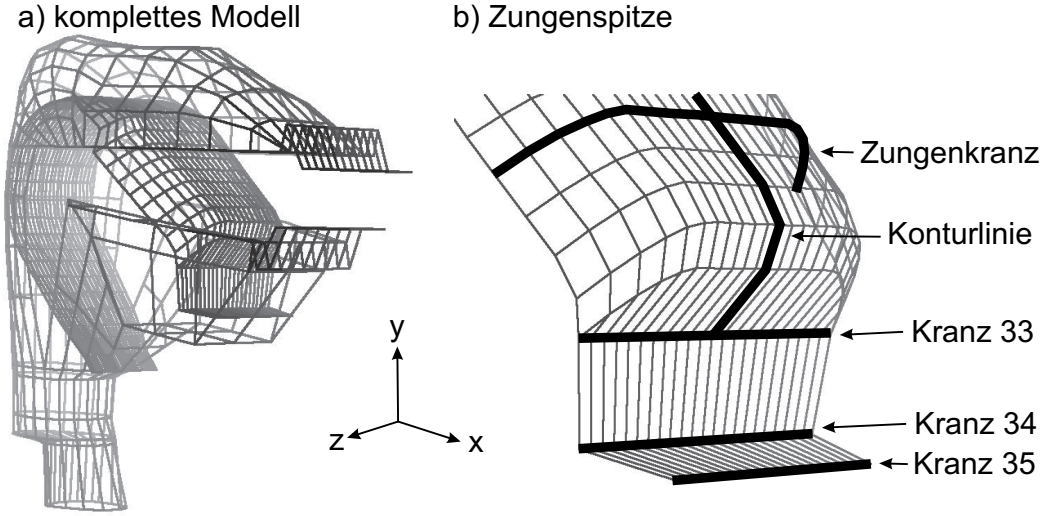

Abbildung 4.10: Verdeutlichung der Ausrichtung der Zungenkränze gegenüber der Zungenkonturline

Kranz 33 liegt so, dass er die Unterseite von Kranz 32 abschließt. Kranz 34 liegt senkrecht unter Kranz 33 und einen halben cm über dem Unterkieferboden. Auch der Kranz 35 liegt einen halben cm über dem Unterkieferboden und „berührt“ die Unterkiefervorderseite. Die Breite der Zungenkränze an diesen Stellen wird den Vorgaben durch die Mantelgitter angepasst.

4.1.3 Beschränkung der Parameter

Eine wichtige Rolle spielt die Einschränkung der Wertebereiche der Mantel- und Zungenparameter. Dadurch wird sicher gestellt, dass der Vokaltrakt nur physiologisch sinnvolle Geometrien annehmen kann, also dass z.B. der Zungenkörper oder die Zungenspitze nicht aus den Mantelgittern herausragen können. Erst durch diese Einschränkungen kann eine fehlerfreie Berechnung der Vokaltraktmittellinie und der Querschnittsflächen gewährleistet werden.
Die Mantelparameter müssen lediglich mit den in Tab. 4.2 vorgegebenen oberen und unteren Schranken verglichen und gegebenenfalls darauf eingeschränkt werden. Einzig die untere Schranke des Parameters *LH* muss so angepasst werden, dass sich die Lippen nur soweit schließen können, bis sie aufeinander treffen. Das ist jeweils von der momentanen Ausrichtung des Unterkiefers abhängig.

Etwas komplizierter verhält es sich mit den Zungenparametern. Ihre oberen und unteren Schranken müssen zum Teil erst noch bestimmt werden. Dadurch wird sichergestellt, dass der große und der kleine Zungenkreis nicht aus dem Mund herausragen und dass keiner der Punkte für die Definition der beiden Zungensplines innerhalb der Kreisflächen liegt.

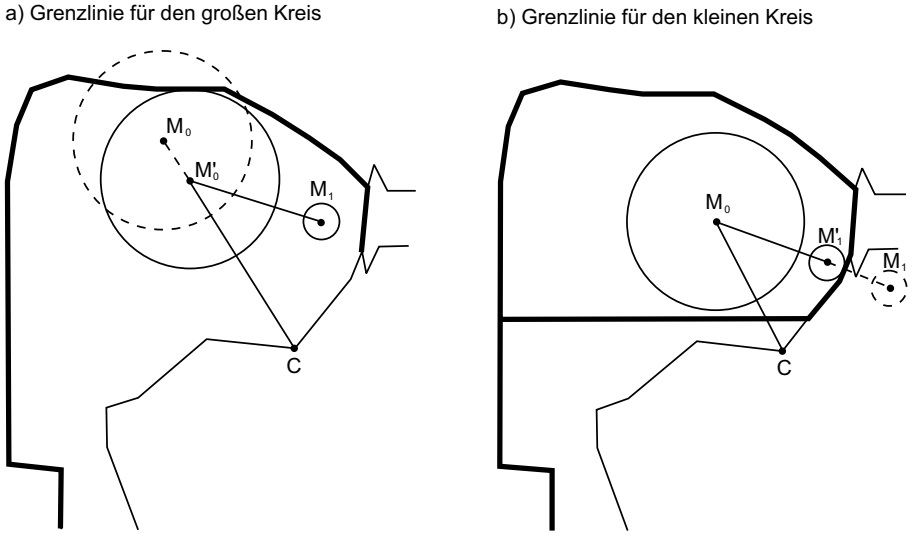

Abbildung 4.11: Die fett gedruckten Linien begrenzen den Aufenthaltsbereich für den großen Zungenkreis (links) und den kleinen Zungenkreis (rechts).

In Abb. 4.11 markieren die dicken schwarzen Linien die Grenzen, über die der große bzw. der kleine Kreis nicht hinaustreten dürfen. Beide Grenzlinien verlaufen entlang der gesamten Kontur des Deckengitters und verbinden die oberen mit den unteren Schneidezähnen. Die Grenzlinie für den kleinen Kreis verläuft anschließend noch ein Stück weiter abwärts entlang der Unterkiefervorderseite und dann parallel zur x-Achse nach hinten. Die Höhe des letzten, horizontalen Linienabschnitts ist nicht kritisch. Die Linie wurde 0.5 cm über dem Kinn im Punkt C positioniert.
Abb. 4.11 (a) zeigt ein Beispiel für den Fall, dass der Parameter *TCL* so groß gewählt wurde, dass der gestrichelte große Kreis aus dem Mundraum heraustritt. Der durchgezogene große Kreis zeigt die Situation, nachdem *TCL* soweit beschränkt wurde, dass keine Überschneidung mit der Grenzlinie mehr auftritt. Eine analoge Situation für den kleinen Zungenkreis zeigt Abb. 4.11 (b).
Die Punkte, durch deren stückweise gradlinige Verbindung die fettgedruckten Grenzlinien in Abb. 4.11 definiert werden, wollen wir im Folgenden mit U_i bezeichnen. Ob ein Konflikt zwischen dem großen Kreis und seiner Grenzlinie vorliegt lässt sich überprüfen, indem jeder ihrer gradlinigen Abschnitte $\overline{U_iU_{i+1}}$ auf Schnittpunkte mit dem Kreis getestet wird. Für den Fall, dass der Kreis weit jenseits der Grenzlinie liegt, muss jeder gradlinige Grenzlinienabschnitt auch noch auf Schnittpunkte mit der Strecke $\overline{CM_0}$ getestet werden. Die Überprüfung für den kleinen Kreis verläuft analog.
Die Einschränkung der Zungenparameter wird in drei Schritten vorgenommen:

1. Wenn ein Konflikt des großen Kreises mit seiner Grenzlinie vorliegt:

   - Reduziere *TCL* soweit, bis der Konflikt aufgelöst ist oder *TCL* seine feste untere Schranke (siehe Tab. 4.3) erreicht hat.

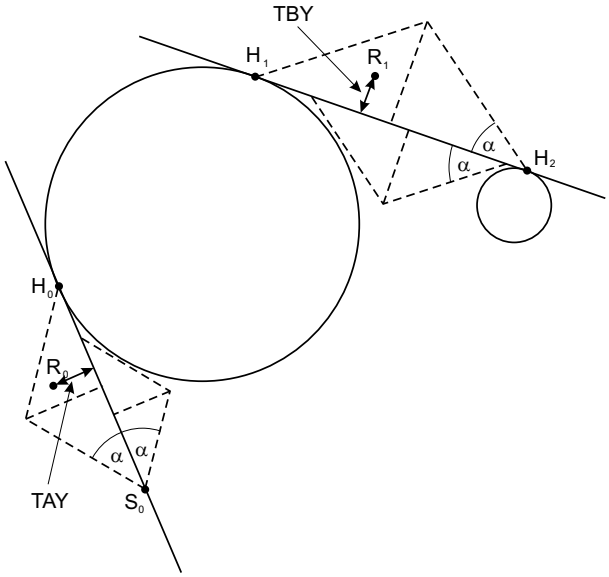

Abbildung 4.12: Beschränkung der Lage der Punkte R_0 und R_1 durch die Flächen der gestrichelt gedruckten Dreiecke.

- Falls immer noch ein Konflikt vorliegt, erhöhe den Winkel *TCA* soweit, bis der Konflikt aufgelöst ist oder *TCA* seine obere Schranke erreicht hat. Die obere Schranke von *TCA* ist dadurch gegeben, dass der tiefste Punkt des Kreises stets 2 mm oberhalb des Zungenbeins liegen muss.

2. Wenn ein Konflikt des kleinen Kreises mit seiner Grenzlinie vorliegt:

   - Reduziere *TBL* soweit, bis der Konflikt aufgelöst ist oder *TBL* seine untere Schranke (Summe der Radien beider Kreise) erreicht hat.
   - Falls der Kreismittelpunkt unterhalb des letzten (horizontalen) Grenzlinienabschnitts liegt, erhöhe *TBA* soweit, dass der Mittelpunkt bei gleichbleibendem *TBL auf* diesem Abschnitt liegt.
   - Reduziere *TBL* gegebenenfalls erneut.

3. Beschränke die Parameter *TAY* und *TBY* soweit, dass die Punkte R_0 und R_1 innerhalb der gestrichelten Dreiecke in Abb. 4.12 liegen. Die oberen und unteren Schranken für *TAY* und *TBY* hängen also von den Werten der Parameter *TAX* und *TBX* ab.

Wir wollen nun genauer die Vorgehensweise bei der Beschränkung des Parameters *TCL* untersuchen, wenn ein Konflikt des großen Kreises mit seiner Grenzlinie vorliegt. Für die Beschränkung von *TBL* ist die Vorgehensweise dann analog.
Falls für einen gegebenen Wert für *TCL* ein Konflikt festgestellt wurde, dann gehen wird davon aus, dass ein niedrigerer Wert für *TCL* existiert, für den der Zungekreis die Grenzlinie in genau einem Punkt berührt und komplett innerhalb der Begrenzung liegt. Der Berührungspunkt kann entweder auf einem der gradlinigen Linienabschnitte $\overline{U_iU_{i+1}}$ liegen, oder es kann sich direkt um einen der Punkte U_i handeln.

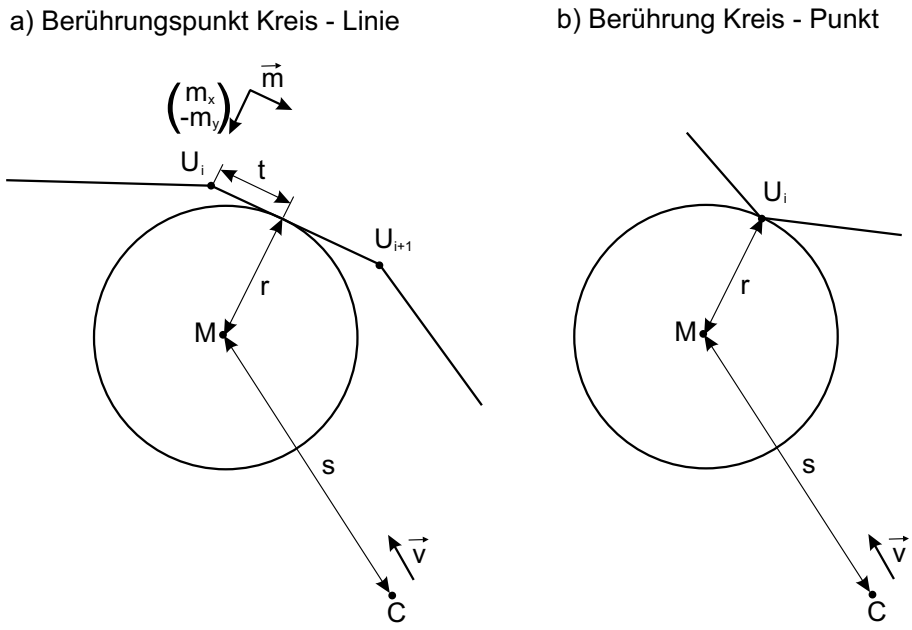

Abbildung 4.13: Beschränkung der Länge s (TCL) im Falle eines Konflikes zwischen Kreis und Grenzlinie.

Der erste Fall muss für jeden Linienabschnitt untersucht werden und ist in Abb. 4.13 (a) verdeutlicht. Gegeben seien folgende Vektoren:

$$\vec{m} = \frac{\overrightarrow{U_iU_{i+1}}}{\|\overrightarrow{U_iU_{i+1}}\|}, \quad \vec{v} = \frac{\overrightarrow{CM}}{\|\overrightarrow{CM}\|} \quad \text{und} \quad \vec{w} = \overrightarrow{U_iC}.$$

Mit dem Ansatz

$$U_i + t\vec{m} + r\begin{pmatrix} m_y \\ -m_x \end{pmatrix} - s\vec{v} = C \tag{4.6}$$

ergeben sich für s und t die Lösungen

$$s = \frac{w_x m_y - w_y m_x - r}{m_x v_y - m_y v_x} \quad \text{und} \quad t = \frac{w_x v_y - w_y v_x - r\vec{m}\,\vec{v}}{m_x v_y - m_y v_x}. \tag{4.7}$$

Falls $t \in [0, \|\overrightarrow{U_iU_{i+1}}\|]$ und $s < TCL$, dann wird TCL der Wert von s zugewiesen, sofern dieser die untere Schranke nicht unterschreitet.
Der zweite Fall (Abb. 4.13 (b)) muss für jeden Punkt U_i der Grenzlinie untersucht werden. Die Vektoren $\vec{v}$ und $\vec{w}$ sind die gleichen wie im ersten Fall. Hier kann der Ansatz

$$r^2 = \|C + s\vec{v} - U_i\|^2 \tag{4.8}$$

gemacht werden. Damit ergibt sich die quadratische Gleichung

$$s^2 + 2\vec{w}\,\vec{v}\,s + \|\vec{w}\|^2 - r^2 = 0, \tag{4.9}$$

die nach s aufgelöst werden kann. Wenn reelle Lösungen für s existieren, und die kleinste von ihnen kleiner ist als TCL, dann wird TCL der Wert von diesem s zugewiesen, sofern

er die untere Schranke nicht unterschreitet.

Als nächstes wollen wir die Vorgehensweise bei der Beschränkung des Winkels *TCA* untersuchen, wenn nach der Beschränkung von *TCL* noch immer ein Konflikt des großen Kreises mit seiner Grenzlinie vorliegt. Hier gehen wir davon aus, dass es einen größeren Wert für *TCA* gibt, für den der Zungenkreis die Grenzlinie in genau einem Punkt berührt und komplett innerhalb der Begrenzung liegt. Auch dabei kann der Berührungspunkt entweder auf einem der gradlinigen Linienabschnitte $\overline{U_iU_{i+1}}$ liegen, oder genau auf einen der Punkte U_i fallen.

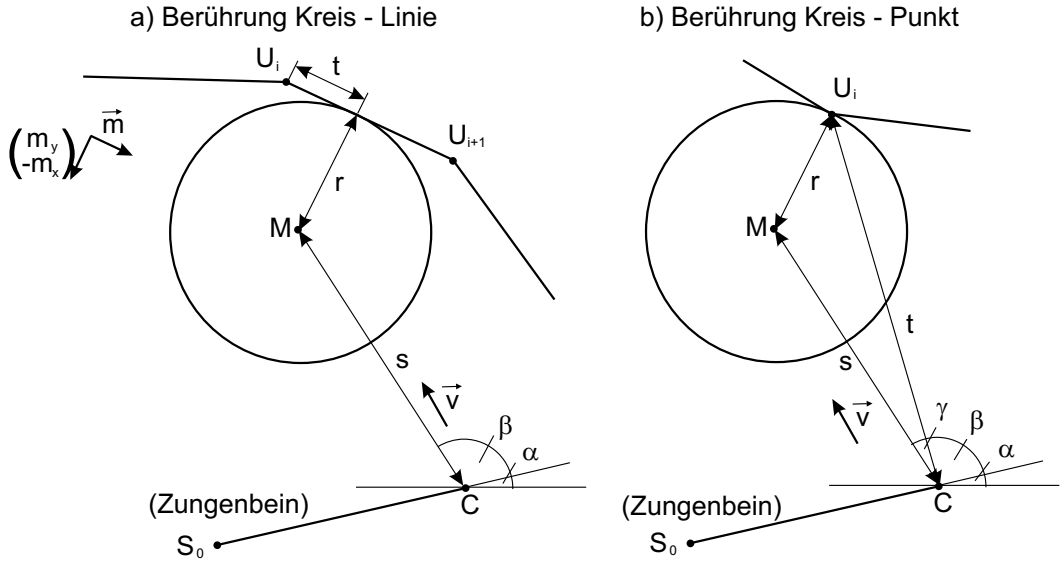

Abbildung 4.14: Beschränkung des Winkels *TCA* im Falle eines Konflikes zwischen Kreis und Grenzlinie.

Abb. 4.14 (a) verdeutlicht den ersten Fall. Die Vektoren $\vec{m}$ und $\vec{v}$ seien wieder

$$\vec{m} = \frac{\overrightarrow{U_iU_{i+1}}}{\|\overrightarrow{U_iU_{i+1}}\|} \quad \text{und} \quad \vec{v} = \frac{\overrightarrow{CM}}{\|\overrightarrow{CM}\|}.$$

Mit dem Ansatz

$$\|U_i + t\vec{m} + r\begin{pmatrix} m_y \\ -m_x \end{pmatrix} - C\|^2 = s^2 \tag{4.10}$$

und der Ersetzung

$$\vec{q} = U_i - C + r\begin{pmatrix} m_y \\ -m_x \end{pmatrix} \tag{4.11}$$

ergibt sich die quadratische Gleichung

$$t^2 + 2\vec{q}\,\vec{m}t + \|\vec{q}\,\|^2 - s^2 = 0. \tag{4.12}$$

Für beide Lösungen für t muss geprüft werden, ob sie reell sind, und wenn ja, ob sie in den Grenzen $[0, \|\overrightarrow{U_iU_{i+1}}\|]$ liegen. Wenn für ein t beide Bedingungen erfüllt sind, dann lässt sich daraus leicht M berechnen und mit den Koordinaten von M der Winkel β.

$$M = U_i + t\vec{m} + r\begin{pmatrix} m_y \\ -m_x \end{pmatrix}, \quad \beta = \arctan\left(\frac{M_y - A_y}{M_x - A_x}\right) - \alpha \tag{4.13}$$

Der Winkel α ist bekannt, da die Koordinaten der Punkte S_0 und C bekannt sind. Falls $\beta > TCA$, dann wird TCA der Wert von β zugewiesen, sofern er die obere Schranke für TCA nicht überschreitet.

Der zweite Fall (Abb. 4.13 (b)) muss wieder für jeden Punkt U_i der Grenzlinie untersucht werden. Sei $t = \|\overrightarrow{U_iC}\|$. Mit Hilfe des Kosinussatzes

$$r^2 = s^2 + t^2 - 2st\cos\gamma \tag{4.14}$$

kann der Winkel γ bestimmt werden:

$$\gamma = \arccos\left(\frac{s^2 + t^2 - r^2}{2st}\right). \tag{4.15}$$

Außerdem ergibt sich β nach der Formel

$$\beta = \arctan\left(\frac{U_{i,y} - C_y}{U_{i,x} - C_x}\right), \tag{4.16}$$

wobei α wie im ersten Fall aus der Position von S_0 und C abgeleitet werden kann.
Falls $\beta + \gamma > TCA$, dann wird TCA der Wert von $\beta + \gamma$ zugewiesen.
Nachdem alle Zungenparameter in ihrem Wertebereich eingeschränkt wurden, kann zur Berechnung der Querschnittsfunktion übergegangen werden.

4.2 Berechnung der Querschnittsfunktion

4.2.1 Mittellinie des Ansatzrohrs

Um die Querschnittsfunktion berechnen zu können, muss zunächst der Verlauf der Vokaltraktmittellinie bestimmt werden. Die Mittellinie sollte den Pfad der Schallausbreitung im Vokaltrakt möglichst genau widerspiegeln. In Abschnitt 3.2 wurde beschrieben, wie die Mittellinie für das Modell von Mermelstein berechnet wird. Da mir diese Methode allein zu ungenau erschien, habe ich sie im Rahmen dieser Arbeit weiter entwickelt. Wegen der Symmetrie des Vokaltraktmodells wird auch hier davon ausgegangen, dass die Mittellinie in der mediosagittalen Ebene liegt.
Im Folgenden werden die Schritte zur Berechnung der Mittellinie ausführlich beschrieben.
Zunächst wird als erste, grobe Approximation der Mittellinie eine μ-Linie wie in Abb. 4.15 (a) definiert. Die mediosagittale Ebene ist dort durch die gestrichelten Linien in die drei Sektoren (I), (II) und (III) unterteilt. Durch jeden der Sektoren verläuft ein Abschnitt der μ-Linie. Im Sektor (I) ist der Verlauf der Linie vertikal, im Sektor (II) radial und im Sektor (III) horizontal. Der radiale Teil der μ-Linie besteht aus einem viertel Kreisbogen mit dem Radius des großen Zungenkreises (TCR).

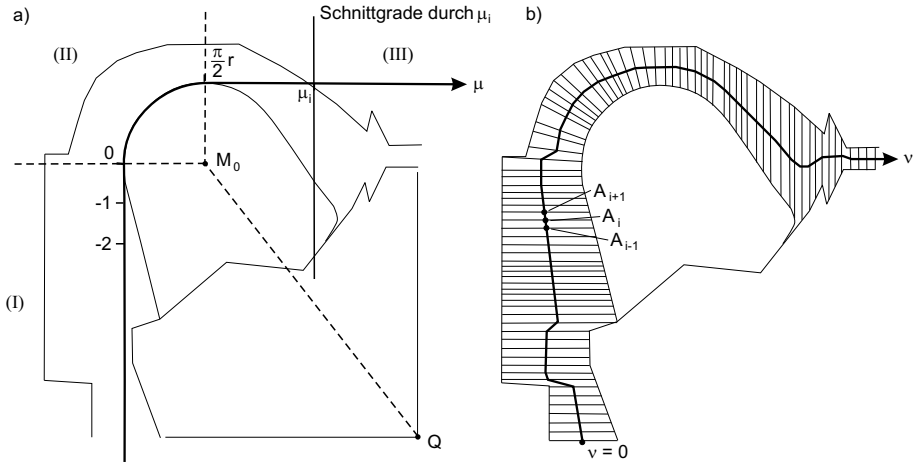

Abbildung 4.15: Erste und zweite Näherung für die Mittellinie.

Mit der Definition der μ-Linie ist es möglich, jedem Stützpunkt der Boden-, Decken- und Zungenkontur[2] einen eindeutigen μ-Wert zuzuweisen, indem auf den Abschnitt der μ-Linie im Sektor dieses Punktes das Lot gefällt wird. Die μ-Werte aller Stützpunkte dieser Konturen werden aufsteigend sortiert in einer Liste abgelegt.
Um eine gleichmäßigere Verteilung der μ-Werte in der Liste zu erreichen, wird darauf geachtet, dass keine zwei benachbarten Werte eine Differenz von mehr als 2.5 mm oder weniger als 1 mm aufweisen. Gegebenenfalls müssen Werte gelöscht oder zusätzlich eingefügt werden.
Für jeden Wert μ_i in der Liste lässt sich eindeutig eine Gerade bestimmen, die an der entsprechenden Stelle senkrecht auf der μ-Linie steht. Die Verbindungslinien der Schnittpunkte dieser Geraden mit der Deckenkontur sowie der Boden- bzw. Zungenkontur sind in Abb. 4.15 (b) dargestellt. Die stückweise lineare Verbindung der Mittelpunkte A_i dieser Verbindungslinien ergibt die zweite Näherung für die Mittellinie, und wird im Folgenden als ν-Linie bezeichnet.
Durch die ν-Linie wird der Verlauf der Mittellinie schon viel genauer angenähert, jedoch besitzt sie an einigen Stellen noch deutlich sichtbare Knicke. Deshalb wird sie im nächsten Schritt geglättet. Das Resultat ist dann die fein abgetastete ξ-Linie.
Zuerst wird die Gesamtlänge L der ν-Linie durch die Aufsummierung aller ihrer Streckenabschnitte ermittelt. Dann werden nach der Formel

$$\nu_i = L \cdot i/(N-1), \quad 0 \leq i < N \tag{4.17}$$

N Positionen in gleichmäßigen Abständen auf der Linie berechnet. Die Stützpunkte R_i der ξ-Linie erhält man, indem jeweils der Schwerpunkt eines kleinen Abschnittes der ν-Linie um die Position ν_i gebildet wird.
Abb. 4.16 (a) verdeutlicht die Berechnung des Schwerpunkts. Die Stützpunkte der ν-Linie sind als A_k bezeichnet und der Linienzug aus den Punkten $B_0B_1B_2...B_{M-1}$ beschreibt den

[2]Tatsächlich wird hier nicht die Hauptlinie $H0$ des Zungengitters als Konturlinie betrachtet, sondern $H6$. Dadurch wird letztlich eine genauere Annäherung der Mittellinie im Falle von stark abfallenden Zungenrändern, wie z.B. bei den Lateralen im oralen Teil, erreicht.

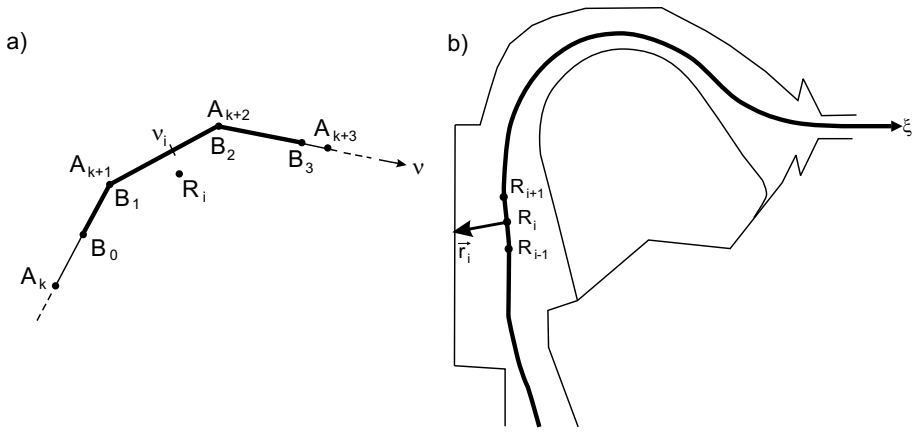

Abbildung 4.16: Berechnung der entgültigen Mittellinie (ξ-Linie) durch die Glättung des Verlaufs der ν-Linie. Die Stützpunkte R_i der ξ-Linie sind jeweils der arithmetische Mittelwert der Punkte eines kurzen Abschnitts (fett gedruckter Linienabschnitt in (a)) der ν-Linie.

Teil der Linie, der von der Position $\nu_i - \frac{w}{2}$ bis zur Position $\nu_i + \frac{w}{2}$ reicht, wobei w die Länge des Abschnittes auf der ν-Linie ist, für die der Schwerpunkt gebildet wird.
Der Stützpunkt R_i der ξ-Linie berechnet sich dann nach der Formel

$$R_i = \frac{1}{w} \sum_{j=0}^{M-2} \frac{1}{2}(B_j + B_{j+1}) \|\overrightarrow{B_j B_{j+1}}\|. \tag{4.18}$$

Je größer w gewählt wird, desto stärker wird die Mittellinie geglättet und desto mehr Details des ursprünglichen Verlaufs gehen verloren. Ein Wert von 3 cm für w hat sich als günstig herausgestellt.
Die ξ-Linie ist das Endergebnis dieser Berechnungen und wird im Folgenden als Mittellinie bezeichnet. Zuletzt wird noch für jeden Stützpunkt R_i der Mittellinie der Normalenvektor $\overrightarrow{r_i}$ berechnet, indem die Vektoren $\overrightarrow{R_{i-1}R_i}$ und $\overrightarrow{R_i R_{i+1}}$ addiert werden, der Ergebnisvektor normiert und um 90° entgegen dem Uhrzeigersinn gedreht wird (siehe Abb. 4.16 (b)). Die Vorberechnung dieser Vektoren erlaubt später eine effiziente Bestimmung der Schnittebenen.

4.2.2 Schnitt- und Flächenberechnung

Anhand des Verlaufs der Mittellinie kann als nächstes die Ausrichtung der Ebenen bestimmt werden, mit denen das Artikulatormodell geschnitten wird. Die Flächeninhalte der dabei entstehenden Schnittflächen – aufgetragen über der Position entlang der Mittellinie – ergeben dann die Querschnittsfunktion. Analog wird die Umfangsfunktion berechnet, indem jeweils der Umfang der Schnittflächen abgetragen wird. Die dazu benötigten Verfahren werden in diesem Abschnitt entwickelt.
Begonnen wird mit der Berechnung der Schnittebenen. Da sie senkrecht zur Mittellinie ausgerichtet sind, stehen sie auch immer senkrecht auf der *xy*-Ebene. Deshalb erscheinen sie in der mediosagittalen Ebene als Schnitt*linien*.

Häufig werden die Schnittlinien zweidimensionaler Artikulatormodelle ermittelt, indem ca. 20-40 Punkte in gleichmäßigen Abständen auf der Mittellinie abgetastet werden, in denen die Mittelliniennormalen gebildet werden. Diese Vorgehensweise ist sehr einfach, hat aber zwei Nachteile:

1. Einerseits könnte das Modell in Abschnitten, in denen sich der tatsächliche Querschnitt stark verändert, unterabgetastet werden, d.h. große, örtlich dicht beieinanderliegende Querschnittsänderungen werden in der Querschnittsfunktion ungenau erfasst. Besonders bei kontinuierlichen Artikulationsgesten könnte sich die Querschnittsfunktion im zeitlichen Verlauf an den betroffenen Stellen sprunghaft ändern, was sich auch akustisch bemerkbar machen würde.

2. Andererseits könnte es passieren, dass sich die Schnittebenen im Bereich des Ansatzrohres überschneiden. Das ist besonders dort zu beobachten, wo viele Schnittebenen dicht beieinander liegen und die Mittellinie stark gekrümmt ist.

Beide Probleme werden mit dem hier vorgestellten Verfahren gelöst.
Dazu wird jede Schnittlinie durch ein Tupel

$$(\xi_P, P, \overrightarrow{v}, min, max, quality)$$

repräsentiert. ξ_P ist der Abstand von der Stimmritze (gemessen entlang der Mittellinie), P ist der Ansatzpunkt auf der Mittellinie und $\overrightarrow{v}$ der normierte Richtungsvektor. $\overrightarrow{v}$ zeigt in Richtung des Deckengitters, also von der Mittellinie aus gesehen 90° nach links. Jede Schnittlinie hat einen Anfangspunkt $P + max\overrightarrow{v}$ und einen Endpunkt $P + min\overrightarrow{v}$. Der Anfangspunkt ist jeweils der Schnittpunkt mit der Konturlinie des Deckengitters, markiert also in der mediosagittalen Ebene die maximale Ausdehnung des Rohrquerschnitts nach oben/hinten. Der Endpunkt ist analog die maximale Ausdehnung des Rohrquerschnitts nach vorne/unten. *quality* ist ein Maß dafür, inwiefern $\overrightarrow{v}$ einem Normalenvektor auf der μ-Linie entspricht und wird an späterer Stelle eine Rolle spielen.

Die Unterabtastung der Querschnittsfunktion wird vermieden, indem sichergestellt wird, dass durch jeden Stützpunkt der Decken-, Boden- und Zungenkontur mindestens jeweils eine Schnittlinie verläuft. Dadurch werden alle abrupten Querschnittsänderungen – zumindest in der mediosagittalen Ebene – garantiert erfasst.
Es werden also zu jedem Konturstützpunkt Q alle möglichen Schnittlinien gesucht, die senkrecht auf der Mittellinie stehen und durch Q verlaufen. Wenn für eine Schnittlinie der Abstand von Q zur Mittellinie kleiner ist als 4 cm, dann wird das Schnittlinientupel für die spätere Weiterverarbeitung in eine Liste eingetragen.
Die folgenden Punkte beschreiben, wie die Tupelwerte für die in Frage kommenden Schnittlinien durch einen der Punkte Q ermittelt werden.

1. Die Berechnung der Werte ξ_P und P der Schnittlinien durch den Punkt Q verdeutlichen Abb. 4.17 (a) und der folgende Algorithmus:

   1: N ist die Anzahl der Stützpunkte der ξ-Linie
   2: **for** $i := 0$ to $N-2$ **do**

3: **if** (Q liegt rechts von $R_i + n\overrightarrow{r_i}$) und ($Q$ liegt links von $R_{i+1} + n\overrightarrow{r_{i+1}}$) **then**
4: $\overrightarrow{w} := \overrightarrow{R_iR_{i+1}}$
5: S_0 ist der Schnittpunkt von $Q + n\overrightarrow{w}$ und $R_i + m\overrightarrow{r_i}$
6: S_1 ist der Schnittpunkt von $Q + n\overrightarrow{w}$ und $R_{i+1} + m\overrightarrow{r_{i+1}}$
7: $t := \|\overrightarrow{S_0Q}\|/\|\overrightarrow{S_0S_1}\|$
8: $P := R_i + t\overrightarrow{w}$
9: {Die Schnittlinie ist nur dann gültig, wenn der Abstand von Q zur Mittellinie kleiner als 4 cm ist}
10: **if** $\|\overrightarrow{PQ}\| < 4$ cm **then**
11: $\xi_P := \xi_i + t(\xi_{i+1} - \xi_i)$
12: Berechne $\overrightarrow{v}, min, max$ und *quality* und trage das Tupel für die Schnittlinie in die Liste ein
13: **end if**
14: **end if**
15: **end for**

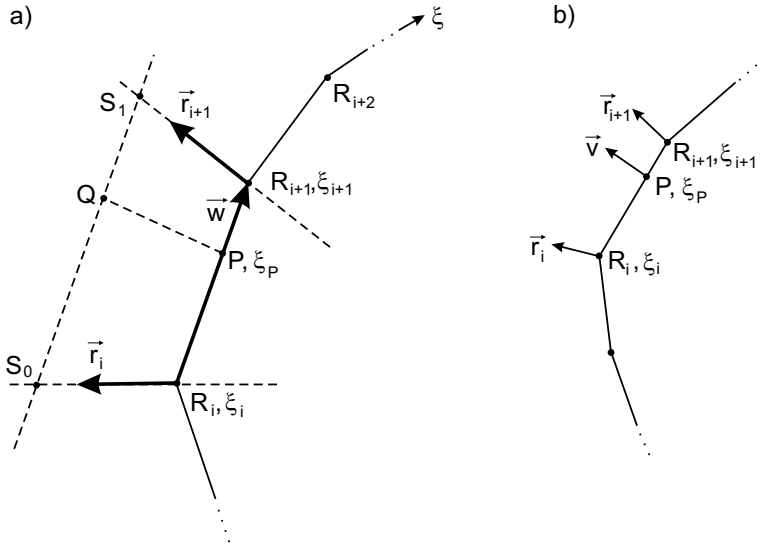

Abbildung 4.17: Gegeben ist ein beliebiger Punkt Q. Gesucht sind der Punkt P, der sich durch die senkrechte Projektion von Q auf die Mittellinie ergibt und seine Position ξ_P auf der Linie in (a) sowie der Normalenvektor v im Punkt P in (b).

2. Der Normalenvektor $\overrightarrow{v}$ im Punkt P auf der Mittellinie wird berechnet, indem zunächst das Liniensegment $\overline{R_iR_{i+1}}$ der Mittellinie gesucht wird, für das $\xi_i \leq \xi_P \leq \xi_{i+1}$ gilt. Die Suche lässt sich beschleunigen, indem am Anfang für i die Abschätzung $i = \frac{\xi_P}{\xi_{N-1}}(N-1)$ getroffen wird (N ist die Anzahl der Stützpunkte der Mittellinie). $\overrightarrow{v}$ kann dann gemäß Abb. 4.17 (b) linear zwischen r_i und r_{i+1} interpoliert werden.

3. Die *min*- und *max*-Werte einer Schnittlinie werden berechnet, indem die Gerade $P + n\overrightarrow{v}$ mit der Deckenkontur, der Bodenkontur sowie der ersten und letzten Hauptlinie des Zungengitters geschnitten wird. Die Schnittpunkte S_i ($i = 0,1,2,3$) sind

in Abb. 4.18 (a) eingezeichnet, und lassen sich als $S_i = P + d_i \vec{v}$ schreiben. Der gesuchte *max*-Wert ist d_0. Der gesuchte *min*-Wert ist das Minimum aus d_1 und d_2, falls die Zunge geschnitten wurde und d_3, wenn nicht die Zunge, dafür aber die Bodenkontur geschnitten wurde.

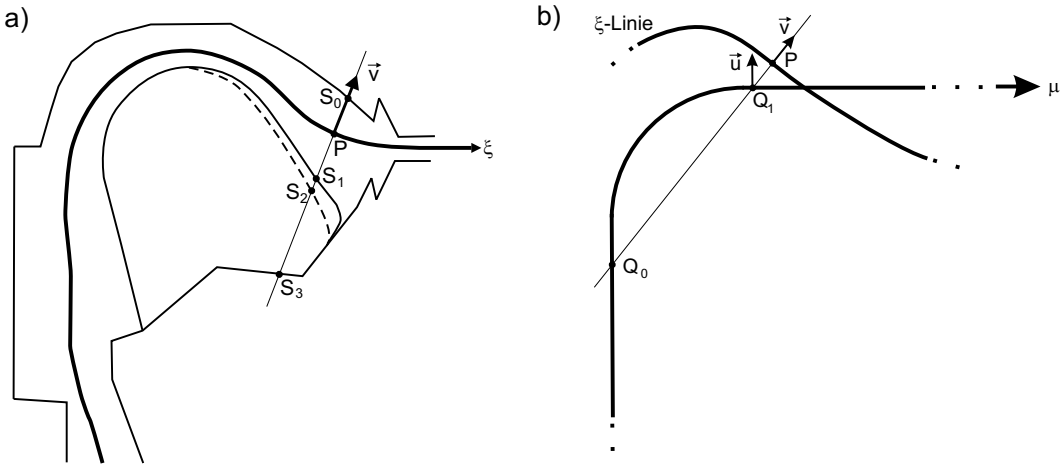

Abbildung 4.18: Berechnung der Werte für *min* und *max* in (a) sowie für *quality* in (b). Die Zunge in (a) wird durch zwei Konturlinien dargestellt (durchgezogen und gestrichelt), die die erste und die letzte Hauptlinie des Zungengitters repräsentieren.

4. Der Wert für *quality* ist ein Maß für die „Qualität“ der Schnittlinie. Er ist umso höher, je ähnlicher der Verlauf der Mittellinie im Punkt P dem Verlauf der μ-Linie ist. Entsprechend Abb. 4.18 (b) werden zuerst die Schnittpunkte Q_0, Q_1 und Q_2 mit den drei Abschnitten (horizontal, vertikal, radial) der μ-Linie gebildet. Von den existierenden Schnittpunkten wird derjenige mit dem geringsten Abstand zu P ausgewählt. In diesem wird der Normalenvektor $\vec{u}$ ermittelt. *quality* ist das Skalarprodukt aus $\vec{u}$ und $\vec{v}$.

Nachdem die Tupel für die Schnittlinien durch alle Stützpunkte der Boden-, Decken- und Zungenkontur in die Liste eingetragen wurden, müssen noch zwei zusätzliche Schnittlinien hinzugefügt werden, die den Anfang und das Ende des Ansatzrohres definieren.
Das Ansatzrohr beginnt stets an der Stimmritze (Punkt R_0 der ξ-Linie), so dass die dazugehörige horizontale Schnittlinie direkt als das Tupel $(0, R_0, (-1,0)^T, -2, 2, 1)$ in die Liste eingetragen werden kann. Die Wahl der *min*- und *max*-Werte ist für diese Linie nicht kritisch.
Die Ebene der Schallabstrahlung bildet das Ende des Ansatzrohres. Sie ist in der Regel nicht die tangentiale Begrenzung von Ober- und Unterlippe, sondern ein Stück ins Innere des Mundraumes verlegt. Mermelstein [42] schlägt vor, die Abstrahlungsebene in Höhe der Mundwinkel anzusetzen und zeichnet deshalb einen rechten Winkel derart in den Mundraum, dass die Schenkel die Ober- und Unterlippe berühren und der Scheitelpunkt auf der Mittellinie liegt. Dieses Verfahren wird auch in dieser Arbeit für die Berechnung der letzten Schnittlinie verwendet.
Abb. 4.19 (a) verdeutlicht die Berechnung des ξ_P-Wertes der entsprechenden Schnittlinie mit Hilfe des Satz des Thales. Zunächst muss der Linienabschnitt $\overline{R_i R_{i+1}}$ der Mittellinie

gefunden werden, der den Kreis mit dem Mittelpunkt M und dem Radius $r = \|\overrightarrow{QS}\|/2$ schneidet. Mit dem Ansatz

$$\begin{aligned} \overrightarrow{w} &= \overrightarrow{R_i R_{i+1}} \\ \overrightarrow{u} &= \overrightarrow{M R_i} \\ r &= \|R_i + t\overrightarrow{w} - M\| \end{aligned} \tag{4.19}$$

ergibt sich für t die quadratische Gleichung

$$t^2 + t\frac{2\overrightarrow{u}\,\overrightarrow{w}}{\|\overrightarrow{w}\|^2} + \frac{\|\overrightarrow{u}\|^2 - r^2}{\|\overrightarrow{w}\|^2} = 0. \tag{4.20}$$

Wählt man die kleinere der beiden Lösungen für t ergibt sich $\xi_P = \xi_i + t(\xi_{i+1} - \xi_i)$.
Unter besonderen Umständen kann es passieren, dass ξ_P entweder sehr weit in den Mundraum hineinreicht oder einen so großen Wert besitzt, dass die Schnittlinie die Boden- oder Deckenkontur nicht mehr schneiden würde. Um das zu verhindern, wird ξ_P folgendermaßen eingeschränkt:

$$max\{\xi_A, \xi_B\} \leq \xi_P \leq min\{\xi_Q, \xi_S, \xi_{N-1}\}. \tag{4.21}$$

ξ_A, ξ_B, ξ_Q und ξ_S sind die Positionen der senkrechten Projektionen der Punkte A, B, Q und S auf die Mittellinie, während ξ_{N-1} der ξ-Wert des letzten Stützpunktes der Mittellinie ist.

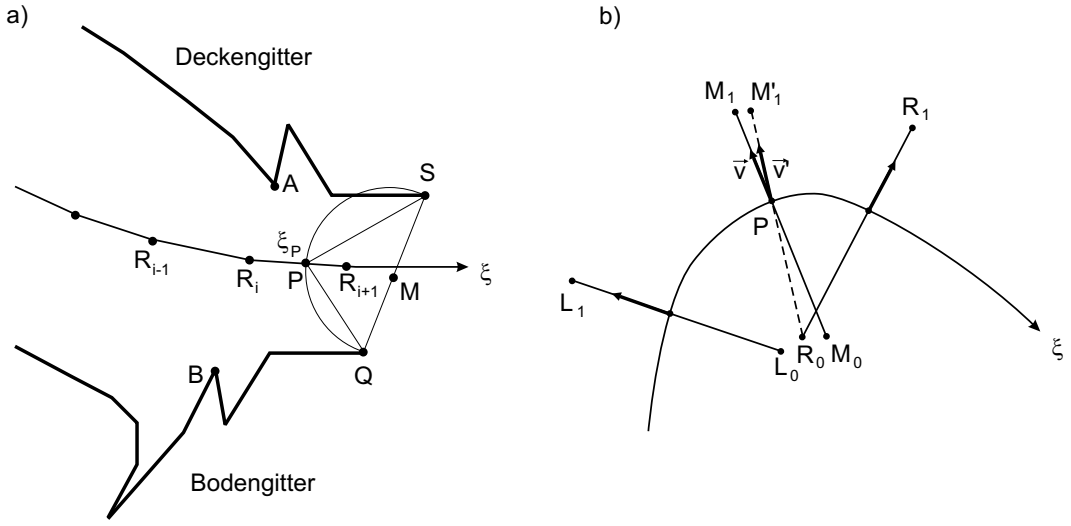

Abbildung 4.19: Position der Abstrahlungsebene im Punkt P (a) und Korrektur der Schnittlinienausrichtung (b). In (a) markieren die Punkte A und B die oberen und unteren Schneidezähne und S und Q die vordere Begrenzung der Ober- und Unterlippe.

Nachdem die beiden Schnittlinien vom Ansatzrohranfang und -ende in die Liste eingefügt wurden, werden alle Tupel in der Liste nach dem *quality*-Wert sortiert. Angefangen beim Tupel mit dem höchsten *quality*-Wert werden die Schnittlinien nun nacheinander aus der ersten Liste ausgelesen und in eine zweite (verkettete) Liste einsortiert. Die Sortierung in der zweiten Liste erfolgt nach aufsteigenden ξ-Werten.

An dieser Stelle erfolgt die Vermeidung von Überschneidungen benachbarter Schnittlinien im Bereich des Ansatzrohres. Immer wenn eine Schnittlinie aus der ersten in die zweite Liste übertragen wird, dann wird sie mit ihrem linken und rechten Nachbarelement in der zweiten Liste auf Überschneidungen getestet. Die zweite Liste ist umso voller, je später eine Schnittlinie dort eingetragen wird, also desto geringer ihr *quality*-Wert ist. Damit steigt die Wahrscheinlichkeit, dass die Ausrichtung dieser Schnittlinie geringfügig geändert werden muss, damit keine Überschneidungen mit der linken oder rechten Nachbarlinie auftreten. So wird sichergestellt, dass am ehesten Schnittlinien „geringer Qualität“ korrigiert werden, da die Korrekturen zur Folge haben, dass sie nicht mehr exakt senkrecht zur Mittellinie ausgerichtet sind.

Das Korrekturverfahren funktioniert ähnlich dem der Zungenkränze und ist in Abb. 4.19 (b) dargestellt. Wenn $(\xi_P, P, \overrightarrow{v}, min, max, quality)$ die Schnittlinie ist, die neu in die zweite Liste einsortiert wird, dann sind ihre Endpunkte $M_0 = P + min\overrightarrow{v}$ und $M_1 = P + max\overrightarrow{v}$. Analog können die Endpunkte der Schnittlinien links und rechts von ihr (also eine Position davor und danach in der zweiten Liste) bestimmt werden. Der folgende Algorithmus löst eventuelle Überschneidungskonflikte auf, indem die Ausrichtung von $\overrightarrow{v}$ angepasst wird.

1: Bilde den Schnittpunkt von $P + t\overrightarrow{v}$ mit $\overline{L_0L_1}$.
2: **if** der Schnittpunkt existiert **then**
3: **if** $(t < 0)$ und $(t \geq min)$ **then**
4: $\overrightarrow{v} := \overrightarrow{L_0P}/\|\overrightarrow{L_0P}\|$
5: **end if**
6: **if** $(t > 0)$ und $(t \leq max)$ **then**
7: $\overrightarrow{v} := \overrightarrow{PL_1}/\|\overrightarrow{PL_1}\|$
8: **end if**
9: **end if**
10: Bilde den Schnittpunkt von $P + t\overrightarrow{v}$ mit $\overline{R_0R_1}$.
11: **if** der Schnittpunkt existiert **then**
12: **if** $(t < 0)$ und $(t \geq min)$ **then**
13: $\overrightarrow{v} := \overrightarrow{R_0P}/\|\overrightarrow{R_0P}\|$
14: **end if**
15: **if** $(t > 0)$ und $(t \leq max)$ **then**
16: $\overrightarrow{v} := \overrightarrow{PR_1}/\|\overrightarrow{PR_1}\|$
17: **end if**
18: **end if**
19: Berechne *min* und *max* neu.

Schnittlinien aus der ersten Liste werden nur dann in die zweite Liste übernommen, wenn ihr ξ_P-Wert nicht größer ist als derjenige der Abstrahlungsebene! Nachdem alle Tupel in die zweite Liste übertragen wurden, wird noch dafür gesorgt, dass die ξ-Werte benachbarter Schnittlinien nicht zu weit auseinanderliegen. Bei Abständen von mehr als 5 mm werden zusätzlich Schnittlinien interpoliert.

Nun wollen wir uns anschauen, wie das Modell mit den durch die Schnittlinien beschriebenen Ebenen geschnitten wird und daraus die Flächeninhalte berechnet werden.

Die Querschnittsfläche des Ansatzrohres an einer bestimmten Position entlang der Mittellinie erhält man, indem jede Hauptlinie des Boden-, Decken- und Zungengitters mit der dortigen Schnittebene geschnitten wird. Die Verbindung der Schnittpunkte zu einem Linienzug ist für einen Rohrquerschnitt innerhalb der Mundhöhle in Abb. 4.20 (a) dargestellt.

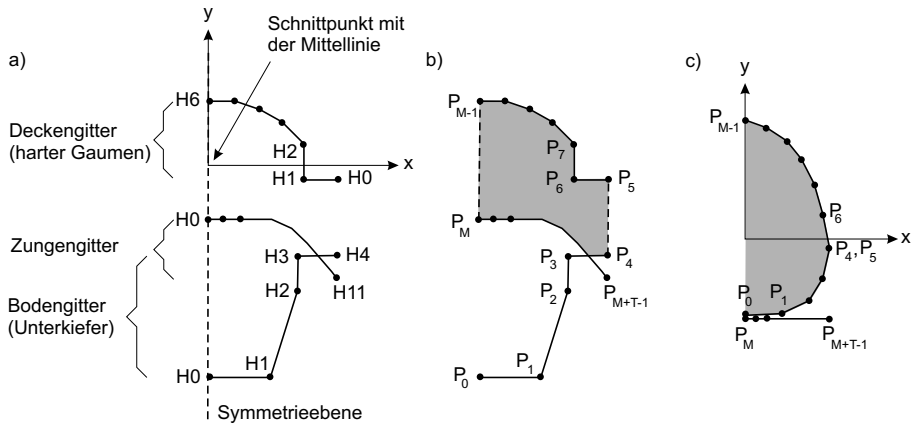

Abbildung 4.20: Bild (a) und (b) zeigen den Querschnitt durch den Vokaltrakt im Bereich der Mundhöhle und Bild (c) im Bereich des Kehlkopfes.

Die Stützpunkte des Linienzugs sollen im Folgenden wie in Abb. 4.20 (b) bezeichnet werden: P_0 bis P_{M-1} sind die Schnittpunkte mit den Mantelgittern und daran schließen sich die Schnittpunkte P_M bis P_{M+T-1} mit dem Zungengitter an. Die Schnittpunkte sind in einem lokalen Koordinatensystem definiert, dessen Ursprung der Schnittpunkt mit der Mittellinie ist. Der für die Querschnittsfunktion gesuchte Flächeninhalt ist grau unterlegt. Der gesuchte Umfang ist die Länge der Randlinie der grauen Fläche.
An einigen Stellen im Vokaltrakt kommt es vor, dass nicht alle Gitter oder alle Hauptlinien eines Gitters von der Schnittebene „getroffen“ werden. Beispielsweise wird es niemals Schnittpunkte mit dem Zungengitter für eine horizontale Schnittfläche unterhalb des Zungenbeins geben. Für den nachfolgenden Algorithmus zur Berechnung des Flächeninhaltes ist es jedoch erforderlich, dass stets alle Punkte $P_0 \ldots P_{M+T-1}$ definiert sind. Deshalb muss sichergestellt werden, dass die von der Schnittebene nicht geschnittenen Hauptlinien trotzdem einen Punkt P_i zugewiesen bekommen.
Für die Schnittpunkte der Mantelgitter P_0 bis P_{M-1} geschieht das, indem ein undefinierter Punkt linear zwischen seinen nächsten linken und rechten Nachbarn interpoliert wird. Falls der erste oder letzte Punkt nicht definiert ist, werden sie auf feste Werte gesetzt.
Für die Zungenpunkte P_M bis P_{M+T-1} wird zunächst geprüft, ob der erste und letzte Punkt definiert sind. Wenn der erste Punkt undefiniert ist, wird er auf $(0, y_{min})^T$ gesetzt, wobei y_{min} der minimale y-Wert aller Mantelpunkte ist. Wenn der letzte Zungenpunkt undefiniert ist, dann wird er auf $(x_{max}, y_{min})^T$ gesetzt, wobei x_{max} der maximale x-Wert aller Mantelpunkte ist. Die zwischen P_M und P_{M+T-1} liegenden Zungenpunkte werden wieder interpoliert. Für den Fall, dass das Zungengitter nicht von der Schnittebene geschnitten wurde, heißt das also, dass die Zungenpunkte eine horizontale Linie unterhalb der Mantelpunkte bilden (siehe Abb. 4.20 (c)). Für den folgenden Algorithmus zur Flächeninhaltsberechnung ist es außerdem strikt erforderlich, dass jeder der Zungenpunkte

definitiv rechts von seinem Vorgänger liegt und dass der x-Wert von P_{M+T-1} größer oder gleich x_{max} ist.

Den Flächeninhalt einer Querschnittsfläche erhält man durch die Aufsummierung kleiner Trapezflächen.

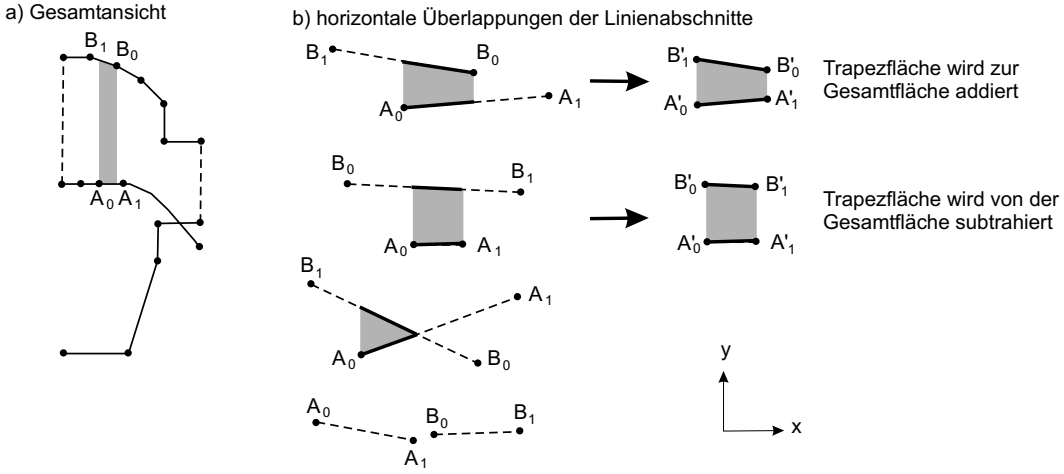

Abbildung 4.21: Bild (a) zeigt eine Trapezfläche, die einen Teil des gesuchten Flächeninhalts ausmacht. Bild (b) zeigt verschiedene Möglichkeiten, wie sich die Linienabschnitte $\overline{A_0A_1}$ und $\overline{B_0B_1}$ der Zunge und des Mantels überlappen können.

Abb. 4.21 und der folgende Algorithmus verdeutlichen die Vorgehensweise.

1: $Flächeninhalt := 0$
2: {Alle Teilstrecken der Zunge durchlaufen}
3: **for** $i := M$ to $M+T-1$ **do**
4: **if** $(i = M+T-1)$ **then**
5: $A_0 := P_i$
6: $A_1 := P_i$
7: $A_{1,x} := A_{1,x} + 10$
8: **else**
9: $A_0 := P_i$
10: $A_1 := P_{i+1}$
11: **end if**
12:
13: {Alle Teilstrecken des Mantels durchlaufen}
14: **for** $k := 0$ to $M-2$ **do**
15: $B_0 := P_k$
16: $B_1 := P_{k+1}$
17: Reduziere die Strecken $\overline{B_0B_1}$ und $\overline{A_0A_1}$ auf die Streckenabschnitte $\overline{B_0'B_1'}$ und $\overline{A_0'A_1'}$, in denen sie sich bezüglich ihrer x-Werte überlappen und $\overline{B_0B_1}$ oberhalb von $\overline{A_0A_1}$ liegt. Abb. 4.21 (b) zeigt einige Beispiele für die „beschnittenen“ Linien.
18:
19: $Umfang := Umfang + \|\overrightarrow{B_0'B_1'}\|$

20:
21: {Trapezfläche zwischen $\overline{B'_0B'_1}$ und $\overline{A'_0A'_1}$ addieren}
22: **if** $(B'_{1,x} < B'_{0,x})$ **then**
23: $Flächenihnalt := Flächeninhalt + \frac{1}{2}(A'_{1,x} - A'_{0,x})(B'_{1,y} - A'_{0,y} + B'_{0,y} - A'_{1,y})$
24: $Umfang := Umfang + \|\overrightarrow{A'_0A'_1}\|$
25: **end if**
26:
27: {Trapezfläche zwischen $\overline{B'_0B'_1}$ und $\overline{A'_0A'_1}$ subtrahieren}
28: **if** $(B'_{1,x} > B'_{0,x})$ **then**
29: $Flächenihnalt := Flächeninhalt - \frac{1}{2}(A'_{1,x} - A'_{0,x})(B'_{0,y} - A'_{0,y} + B'_{1,y} - A'_{1,y})$
30: $Umfang := Umfang - \|\overrightarrow{A'_0A'_1}\|$
31: **end if**
32: **end for**
33: **end for**
34: $Flächeninhalt := Flächeninhalt \cdot 2$
35: $Umfang := Umfang \cdot 2$

Am Ende des Algorithmus werden sowohl der Umfang als auch der Flächeninhalt verdoppelt, da sie jeweils nur für eine Hälfte des Vokaltrakts berechnet wurden. Der Umfang darf verdoppelt werden, da nach dem obigen Algorithmus in ihm *nicht* der Abstand zwischen den Punkten P_{M-1} und P_M (siehe Abb. 4.20) enthalten ist.
Wenn die Flächeninhalte und Umfangslängen der Rohrquerschnitte für alle Schnittebenen berechnet wurden, lassen sie sich wie in Abb. 4.22 über ihrer Position auf der Mittellinie auftragen. Die Flächeninhalte und Umfangslängen werden für die ξ-Werte zwischen den berechneten Werten linear interpoliert und anschließend diskretisiert.

Die Diskretisierung muss für die akustische Simulation durchgeführt werden. Dort wird das Ansatzrohr als eine Aneinanderreihung kurzer Rohrabschnitte mit unterschiedlichen Querschnittsflächen und Umfangslängen betrachtet. Konkret wird sowohl der Rachen- als auch der Mundraum in jeweils 32 Rohrabschnitte gleicher Länge unterteilt. Dies erscheint mir als ein guter Kompromiss zwischen einer möglichst feinen Abtastung und vertretbarem Rechenaufwand für die akustische Synthese. Als Grenze zwischen dem Rachen- und Mundraum wird die Schnittlinie durch den zweiten Kranz des Velumgitters angenommen, also das untere Ende des Gaumensegels.
Die Diskretisierung selbst besteht nicht einfach in einer gleichmäßigen Abtastung der linear interpolierten Funktion, sondern für jeden Abtastwert wird die entsprechende Fläche unter der linearen Funktion berechnet und durch das Abtastintervall dividiert.
Abb. 4.23 zeigt die Umfangs- und Querschnittsfunktionen für die Laute /a:/, /u:/, /l/ und /i:/ zusammen mit der jeweiligen Länge des Vokaltraktes.

4.3 Geometrie des Nasenraums

Der Nasenraum ist geometrisch nur im hinteren Teil, dem Nasenrachenraum, veränderlich. Ein Absenken des Velums führt zu einer Vergrößerung und eine Anhebung zu einer Verkleinerung des Nasenrachenraums. Gegenläufig führt ein gesenktes Velum zu einer

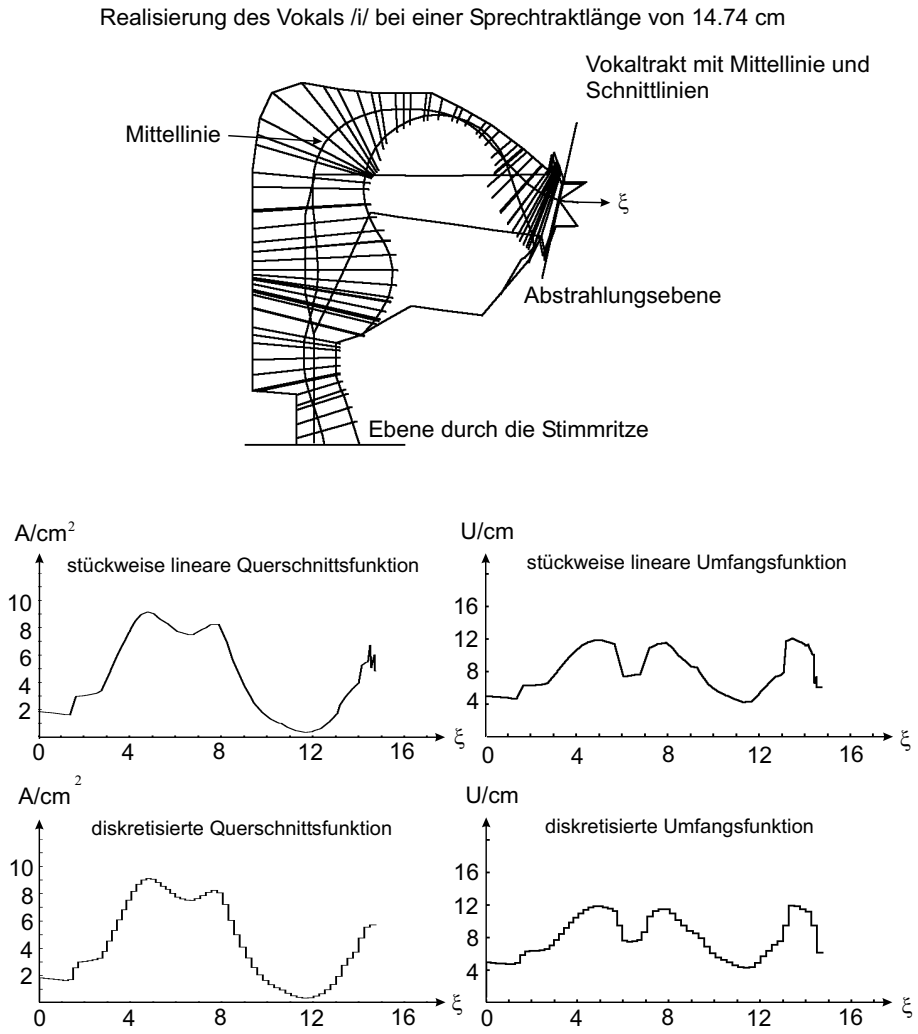

Abbildung 4.22: Diskretisierung der Umfangs- und Querschnittsfunktion für den Vokal /i/.

Verengung und ein gehobenes Velum zu einer Weitung des hinteren Mundraums. Wie bereits im Zusammenhang mit den Mantelgittern beschrieben wurde, wird der Zustand des Velums mit dem Parameter *VO* (velum opening) ausgedrückt. In dieser Arbeit wird vereinfachend angenommen, dass ein Absenken des Velums stets auch eine Vergößerung der Öffnungsfläche zwischen Mundrachen- und Nasenrachenraum nach sich zieht. Damit kann also der Grad der Nasalierung gesteuert werden. Im Modell entspricht die Öffnung der Fläche zwischen den Querlinien $Q0$ und $Q1$ des Velumgitters und kann leicht berechnet werden.

In der Literatur gibt es verschiedene Vorschläge für die Querschnittsfunktion des Nasenraumes [21, 40, 13]. Im Gegensatz zu den anderen Publikationen wurden die Daten in [13] mit Hilfe der Magnetresonanztomographie an lebenden Personen gewonnen. Sie unterscheiden sich besonders im mittleren Nasenabschnitt von früher publizierten Daten, die häufig aus Untersuchungen an Leichnamen abgeleitet wurden. Die älteren Daten zeigen in der Regel zu große Querschnitte im Mittelteil der Nase, was wahrscheinlich auf

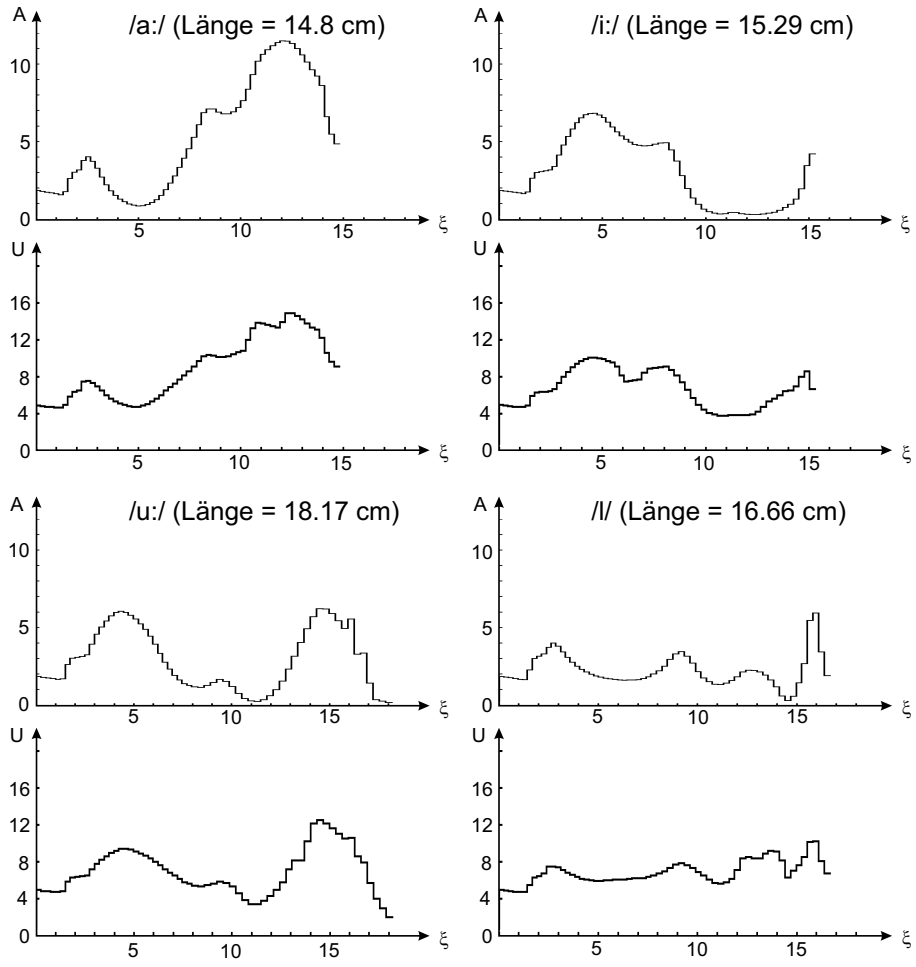

Abbildung 4.23: Diskretisierte Umfangs- und Querschnittsfunktionen für die Laute /a:/, /u:/, /l/ und /i:/.

ausgetrocknete Schleimhäute zurückzuführen ist.

In dieser Arbeit stütze ich mich deshalb auf die Daten aus [13]. Aus den von vier Sprechern gemessenen Querschnitts- und Umfangsfunktionen habe ich jeweils einen „mittleren“ Funktionsverlauf abgeleitet.

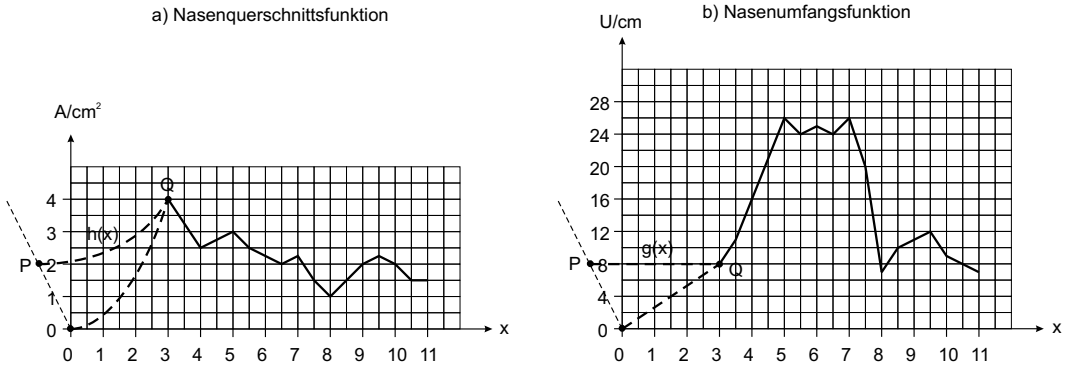

Abbildung 4.24: Querschnitts- und Umfangsfunktion des Nasenraumes.

Sehen wir uns zunächst die Querschnittsfunktion in Abb. 4.24 (a) an. Im vorderen Teil wird sie durch eine Funktion $h(x) = ax^2 + bx + c$ beschrieben, während sie vom Punkt Q bis zu den Nasenlöchern statisch ist. Die Werte für a, b und c werden aus der Öffnungsfläche A des Velums abgeleitet. Ich mache außerdem die Annahme, dass sich der Nasenraum linear mit Zunahme der Öffnungsfläche ein kleines Stück in Richtung des Rachenraums verlängert. Dadurch beginnt die Querschnittsfunktion an der Stelle $P_x = fA$ statt an der Stelle $P_x = 0$, wobei ich f auf den Wert $-\frac{1}{2}$ gesetzt habe. Wenn wir annehmen, dass die erste Ableitung der Funktion $h(x)$ im Punkt P Null ist und $P_y = A$, dann erhält man für a, b und c die folgenden Ausdrücke:

$$a = \frac{Q_y - A}{(P_x - Q_x)^2}, \tag{4.22}$$

$$b = -2aP_x, \tag{4.23}$$

$$c = Q_y - aQ_x^2 - bQ_x. \tag{4.24}$$

Abb. 4.24 (b) bildet die Umfangsfunktion des Nasenraums ab. Auch hier ist der vordere Teil veränderlich. Er wird durch die lineare Funktion $g(x) = mx + n$ beschrieben. Mit $P = (fA, 5\sqrt{A})^T$ ergeben sich für m und n die folgenden Ausdrücke:

$$m = \frac{Q_y - P_y}{Q_x - P_x}, \tag{4.25}$$

$$n = Q_y - mQ_x. \tag{4.26}$$

Auch der Nasenraum wird durch 32 Rohrabschnitte gleicher Länge approximiert. In Abb. 4.25 sind die diskretisierten Querschnitts- und Umfangsfunktionen des Mund-, Rachen- und Nasenraumes für die Nasallaute /m/ und /n/ abgebildet. Der Nasenraum, dessen Einlass jeweils weit geöffnet ist, wurde grau hinterlegt.

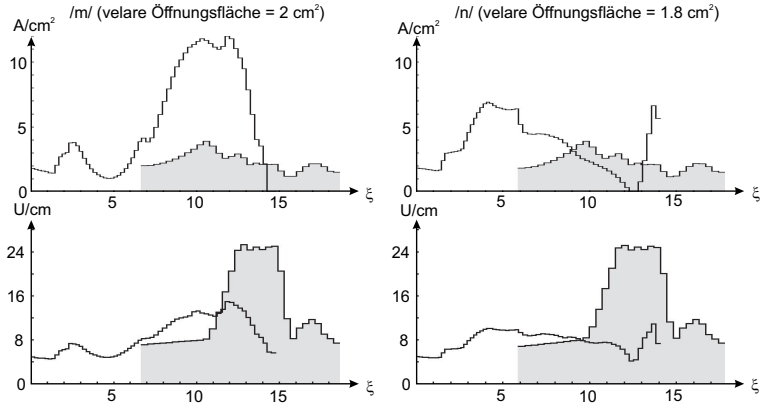

Abbildung 4.25: Gemeinsame Darstellung der Umfangs- und Querschnittsfunktion für den Nasen-, Rachen- und Mundraum für die Nasale /m/ und /n/.

Kapitel 5

Akustische Synthese

Die akustische Synthese hat die Aufgabe – ausgehend von einer konkreten Formung des Ansatzrohres – das an der Mundöffnung und den Nasenlöchern abgestrahlte Schallsignal zu generieren. Die Formung des Ansatzrohrs wird durch die Querschnitts- und Umfangsfunktion repräsentiert, deren Berechnung in den letzten Abschnitten besprochen wurde.
Das in diesem Abschnitt vorgestellte Verfahren für die akustische Synthese ist in der Lage, stimmhafte Sprachlaute mit sehr natürlichem Klang zu erzeugen. Die Produktion von Geräuschlauten dagegen wird hier nicht behandelt, da bisher kein Verfahren existiert, welches diese physikalisch exakt mit vertretbarem Rechenaufwand und in hoher Qualität erzeugen kann.
Die in diesem Abschnitt vorgestellten Methoden sind unabhängig von einem bestimmten Artikulatormodell, da die Eingabegrößen lediglich die Querschnitts- und Umfangsfunktion sind, die aus beliebigen Quellen stammen können.

5.1 Überblick

5.1.1 Größen und Einheiten

Zunächst soll ein kurzer Überblick über die physikalischen Größen gegeben werden, die für die Spracherzeugung eine Rolle spielen [5]. Dabei werden durchgängig die Einheiten des cgs-Systems[1] verwendet. Zur besseren Unterscheidung werden außerdem die zeitabhängigen Größen mit kleinen Buchstaben und die frequenzabhängigen Größen mit großen Buchstaben bezeichnet.
Eine Schallwelle, die sich eindimensional in einem Rohr ausbreitet, lässt sich durch die orts- und zeitabhängigen Größen Schalldruck $p(x,t)$ in $\mathrm{dyn}/\mathrm{cm}^2$ und Schallschnelle $v(x,t)$ in cm/s beschreiben. Der Schalldruck ist die lokale Änderung des Luftdrucks gegenüber dem Normaldruck, während die Schallschnelle die Geschwindigkeit angibt, mit der die Teilchen der Luft um ihre Ruhelage oszillieren. Der Volumenstrom (Schallfluss) $u(x,t)$ ist das Produkt aus der Schallschnelle $v(x,t)$ und der Rohrquerschnittsfläche an der Stelle x zum Zeitpunkt t und hat die Einheit cm^3/s. Schalldruck und Volumenstrom sind

[1] Im cgs-System werden für die Länge, die Masse und die Zeit die Einheiten Zentimeter (cm), Gramm (g) und Sekunde (s) verwendet. Die Einheit für die Kraft ist $[F] = \mathrm{g} \cdot \mathrm{cm}/\mathrm{s}^2 = \mathrm{dyn}$.

über die akustische Impedanz $z(x,t) = p(x,t)/u(x,t)$ miteinander verknüpft. Die Einheit der akustischen Impedanz ist $\mathrm{g/cm^4 s}$.
Im Folgenden wird die Ortsabhängigkeit der Größen oft als Index angegeben: z.B. $p_s(t)$ für den subglottalen Schalldruck (direkt unterhalb der Stimmritze) oder $u_g(t)$ für den glottalen Volumenstrom durch die Stimmritze.

5.1.2 Vorgehensweise

Dieser Abschnitt soll ein Verständnis für die hybride Vorgehensweise bei der akustischen Synthese schaffen. Die hybride Methode macht im Gegensatz zu den reinen zeitdiskreten Simulationen sowohl von Zeitbereichs- als auch von Frequenzbereichsverfahren gebrauch, um das Schallsignal zu berechnen. Dabei wird von einer eindimensionalen Schallausbreitung im Vokaltrakt ausgegangen.
In Anlehnung an die Methode von Sondhi und Schroeter [58] wird die Anregungsfunktion des Vokaltrakts im Zeitbereich und die Übertragungsfunktion im Frequenzbereich berechnet. Die Anregungsfunktion ist der glottale Volumenstrom $u_g(t)$, während die Übertragungsfunktion $H(f)$ das Verhältnis des abgestrahlten Schalldrucks $P_{out}(f)$ zur Fouriertransformierten $U_g(f)$ der Anregungsfunktion ist. Durch die inverse Fouriertransformation der Übertragungsfunktion erhält man die Impulsantwort des Vokaltrakts, die gefaltet mit $u_g(t)$ den abgestrahlten Schalldruck $p_{out}(t)$ ergibt.
Die Anregungsfunktion wird mit dem selbstschwingenden Glottismodell von Ishizaka und Flanagan [32] berechnet. Selbstschwingend bedeutet, dass die Schwingungsform der Stimmlippen nicht vorgegeben wird. Statt dessen ergibt sie sich ganz automatisch aus den Druck-, Trägheits- und Federkräften, die an den Stimmlippen angreifen. Diese resultieren aus dem Volumenstrom durch die Stimmritze, der seinerseits von der Öffnungsfläche der Stimmritze und den darüber und darunter herrschenden Drücken abhängt. Es handelt sich bei der Glottis also um ein sehr komplexes, nichtlineares System, welches am besten im Zeitbereich beschrieben und simuliert wird.

Durch das periodische Öffnen und Schließen der Stimmritze wird an den Stimmlippen aerodynamische Strömungsenergie in Schallenergie umgewandelt. Die Schallenergie wird durch den Vokaltrakt zu der Mundöffnung und den Nasenlöchern übertragen. Das Ansatzrohr ist also eine akustische Übertragungsleitung.
Die Übertragungsfunktion wird aus der Ansatzrohrgeometrie abgeleitet. Dazu wird der Vokaltrakt als eine Aneinanderreihung kurzer Rohrabschnitte aufgefasst (siehe Abb. 5.1). Die Länge, die Querschnittsfläche und der Umfang der Rohrabschnitte gehen direkt aus der diskreten Querschnitts- und Umfangsfunktion des Vokaltrakts hervor.
Das Röhrenmodell lässt sich in ein elektrisches Ersatzschaltbild überführen, mit dessen Hilfe sich die Übertragungsfunktion im Frequenzbereich berechnen lässt. Dabei können – im Gegensatz zu Zeitbereichsverfahren – besonders gut frequenzabhängige Energieverluste im Vokaltrakt berücksichtigt werden. Diese entstehen z.B. durch Wandhaftung, Schallabstrahlung und Energieabsorption durch die Vokaltraktwände.

In Abb. 5.2 sind die Schritte zur Berechnung des Schallsignals am Beispiel des Vokals /a:/ dargestellt. Die oberen vier Bilder zeigen noch einmal verkürzt die Generierung der Querschnitts- und Umfangsfunktion, wie sie in den letzten Abschnitten

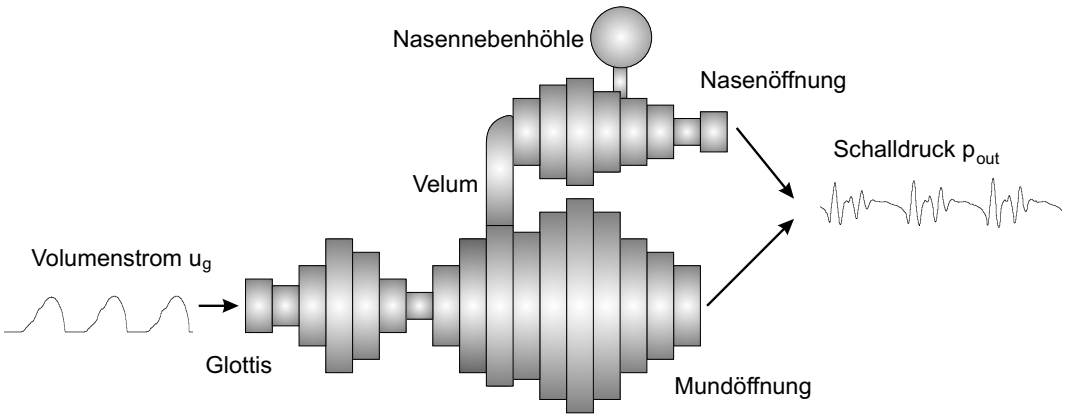

Abbildung 5.1: Diskretes Röhrenmodell des Vokaltrakts. Der runde, an den Nasenraum angekoppelte Resonator symbolisiert die Nasennebenhöhlen.

beschrieben wurde.
Aus der Querschnitts- und Umfangsfunktion werden sowohl die Übertragungsfunktion $H(f)$ als auch die Eingangsimpedanz $Z_{in}(f)$ des Vokaltrakts im Frequenzbereich abgeleitet. Durch die inverse Fouriertransformation werden beide Funktionen in den Zeitbereich überführt. Die Impulsantwort der Übertragungsfunktion ist $h(t)$ und die Impulsantwort der Eingangsimpedanz $z_{in}(t)$. Eine detaillierte Ableitung der Übertragungsfunktion und der Eingangsimpedanz wird in Abschnitt 5.2 gegeben.
Die Faltung der Impulsantworten mit dem Volumenstrom $u_g(t)$ ergeben den Schalldruck $p_{out}(t)$ an der Mund- und Nasenöffnung sowie den supraglottalen Druck $p_0(t)$.
Der supraglottale Druck ist der Schalldruck direkt oberhalb der Stimmritze. Von ihm hängt der Druckabfall über der Stimmritze ab, aus dem sich das Schwingungsverhalten der Stimmlippen sowie der glottale Volumenstrom ergeben.[2]
Ansonsten hängt die Anregung des Vokaltrakts von der Voreinstellung der Stimmlippen durch den Kehlkopf und dem subglottalen Druck ab. Das Glottismodell wird ausführlich in Abschnitt 5.3 vorgestellt.

5.1.3 Diskretisierung

Die Umsetzung der Berechnungen auf dem Computer erfordert eine Digitalisierung bzw. Diskretisierung der betroffenen Funktionen. Diskrete Funktionen werden durch einen Abtastindex in eckigen Klammern hinter dem Funktionssymbol gekennzeichnet. $u_g[n]$ ist z.B. die diskrete Version des Volumenstroms $u_g(t)$. Die Abtastfrequenz $f_A = 1/T$ wurde auf 22 kHz gesetzt. Nach dem Nyquist-Theorem darf somit die höchste im abgetasteten Schallsignal vorkommende Frequenz 11 kHz betragen, was für die Sprachproduktion völlig ausreichend ist.
Das Ziel der akustischen Synthese ist die Berechnung der Abtastwerte $p_{out}[n]$. Die Berechnung erfolgt für jeden Abtastwert nach dem gleichen Schema. Der folgende Pseudo-Code verdeutlicht den Syntheseprozess für eine beliebige Äußerung.

[2] Die Abhängigkeit der Anregungsfunktion von der Ansatzrohrgeometrie (indirekt über den supraglottalen Druck) wurde als wichtiger Faktor für die Natürlichkeit von Vokalen erkannt [3].

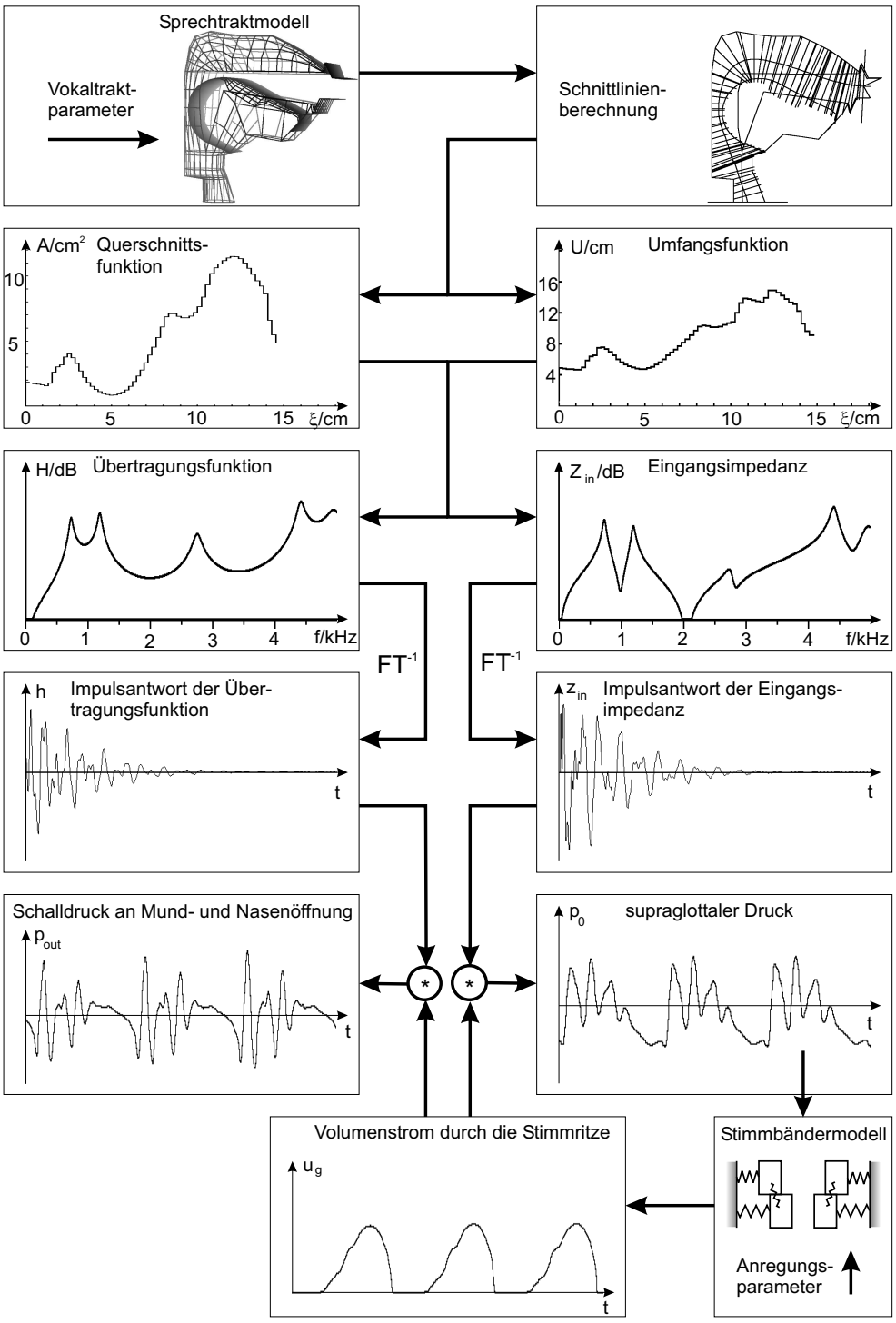

Abbildung 5.2: Informationsflüsse für die akustische Synthese. Das Zeichen $*$ steht für die Faltung. Weitere Erläuterungen werden im Text gegeben.

1: L ist die Länge der Impulsantworten in Abtastwerten
2: M ist die Länge der Äußerung in Abtastwerten
3:
4: **for** $n := 0$ to $M - 1$ **do**

5: Berechne die Impulsantworten $h[i]$ und $z_{in}[i]$, $0 \leq i < L$
6: Berechne die aktuellen Abtastwerte des Volumenstroms $u_g[n]$ und des supraglottalen Drucks $p_0[n]$.
7: {Falte die Volumenstromfunktion mit $h[i]$}
8: $p_{out}[n] = \sum_{i=0}^{L-1} u_g[n-i]h[i]$
9: **end for**

Zunächst werden also für jeden Abtastwert die Impulsantworten $h[i]$ und $z_{in}[i]$ ermittelt. Diese hängen über mehrere Zwischenstufen von den Vokaltraktparametern ab. Ein Vorschlag für eine einfache zeitliche Steuerung der Parameter wird in Abschnitt 6 gemacht. Eine komplette Neuberechnung der Impulsantworten für jeden Abtastwert würde einen unvertretbar hohen Rechenaufwand nach sich ziehen. Deshalb werden sie nur ca. alle 10 ms wirklich neu aus der Querschnitts- und Umfangsfunktion berechnet und für die dazwischenliegenden Zeitpunkte linear interpoliert. Die lineare Interpolation der Abtastwerte der Impulsantworten entspricht zwar nicht exakt einer Interpolation der Querschnittsfunktion, ist aber laut [58] eine gute Näherung.
Im nächsten Schritt wird der aktuelle Abtastwert $u_g[n]$ des glottalen Volumenstroms ermittelt. Dieser ist zusammen mit $p_0[n]$ das Ergebnis der numerischen Simulation des Glottismodells.
Im letzten Schritt wird dann der gesuchte Schalldruckabtastwert durch die Faltung von u_g mit h berechnet.

5.2 Berechnung der Impulsantworten

In diesem Abschnitt folgt eine ausführliche Beschreibung der Methoden, mit denen die Übertragungsfunktion $H(f)$ und die Eingangsimpedanz $Z_{in}(f)$ des Vokal- und Nasaltrakts im Frequenzbereich berechnet werden. Aus der Repräsentation der Funktionen im Frequenzbereich lassen sich dann die Impulsantworten $h(t)$ und $z_{in}(t)$ im Zeitbereich durch die inverse Fourier-Transformation berechnen.

5.2.1 Elektrisches Ersatzschaltbild des Vokaltrakts

Mit Hilfe elektroakustischer Analogien lässt sich die Theorie elektrischer Schaltkreise für die Lösung akustischer Probleme anwenden [5]. Dazu muss das akustische System – in diesem Fall das Ansatzrohr – in ein analoges elektrisches Ersatzschaltbild überführt werden.
Die Grundgrößen eines elektrischen Stromkreises – Strom und Spannung – werden den akustischen Größen Volumenstrom (Schallfluss) und Schalldruck gleichgesetzt. Die physikalischen Eigenschaften der Luft – Trägheit, Kompressibilität und Viskosität – entsprechen der Induktivität, der Kapazität und dem Ohmschen Widerstand im elektrischen Stromkreis.
Schematisch lässt sich ein akustisches System also mit Hilfe von elektrischen Bauelementen (Spule, Kondensator und Widerstand) darstellen. Die Spule repräsentiert die „akustische Masse“ eines Volumenelements der Luft, die durch ihre Trägheit ihrer Beschleunigung einen (induktiven) Widerstand entgegensetzt. Der Kondensator repräsentiert die

Nachgiebigkeit eines Volumenelements der Luft (ähnlich der Nachgiebigkeit einer Feder), die eine Kompression oder Expansion des Luftvolumens ermöglicht. Der Ohmsche Widerstand repräsentiert analog zum elektrischen Stromkreis Energieverluste durch Reibung.
Abb. 5.3 zeigt, wie ein kurzer Abschnitt einer akustischen Rohrleitung mit gleichmäßigem Querschnitt und nachgiebigen Wänden in ein elektrisches Schaltbild überführt werden kann.

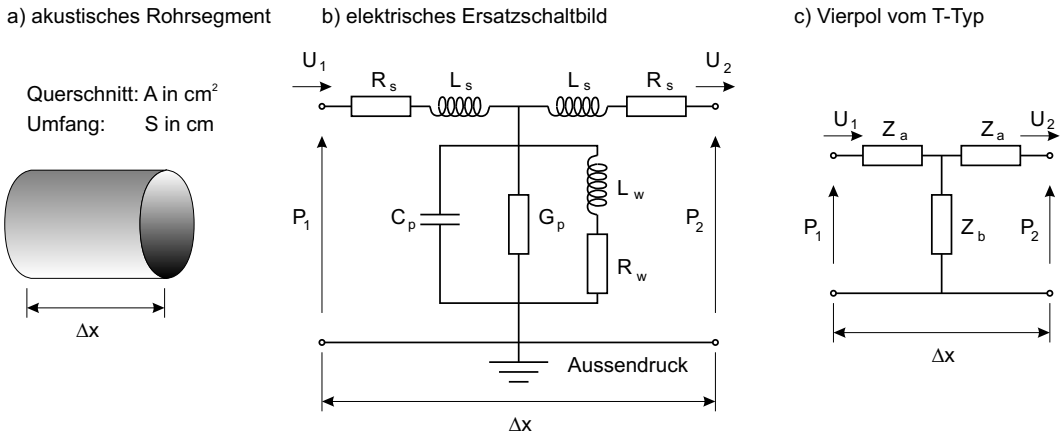

Abbildung 5.3: Überführung eines Abschnitts einer akustischen Rohrleitung in ein elektrisches Ersatzschaltbild. Quelle: [26].

Am Eingang des Rohrabschnitts (a), welcher durch das Schaltbild (b) repräsentiert wird, existiert der Schalldruck P_1 und der Volumenstrom U_1. Am Ausgang sind Schalldruck und Volumenstrom P_2 bzw. U_2. Die Schalldruckdifferenz zwischen Eingang und Ausgang bewirkt eine Beschleunigung der Luftmasse im Rohrabschnitt in Richtung des niedrigeren Drucks. Die Trägheit dieser Luftmasse wird durch die Induktivitäten L_s repräsentiert. Die Widerstände R_s repräsentieren einen Energieverlust proportional zu U^2, der aufgrund der Reibung an den Rohrwänden entsteht.
Parallel zu R_s und L_s sind die Kapazität C_p, die Leitfähigkeit G_p und der komplexe Widerstand $R_w + j\omega L_w$ geschaltet. C_p ermöglicht die Kompression oder Expansion der Luft in dem Rohrabschnitt, indem ein Teil des Volumenstroms in oder aus dem Kondensator fließt. G_p repräsentiert einen Energieverlust, der proportional zum Quadrat des lokalen Schalldrucks ist. Dieser Verlust entsteht durch die Ableitung der Kompressionswärme [3] über die Rohrwand. Die Impedanz $R_w + j\omega L_w$ steht für den mechanischen Widerstand, den die nachgiebigen Rohrwände dem Schalldruck entgegensetzen.
Tab. 5.1 gibt die Formeln an, mit deren Hilfe man die genannten (frequenzabhängigen) Größen für einen Rohrabschnitt der Länge Δx, der Querschnittsfläche A und des Umfangs S berechnen kann. Die Herleitung der Formeln findet man in [26], und die Werte für R_w und L_w wurden aus [33] übernommen.
Das Schaltbild in Abb. 5.3 (b) lässt sich in den Vierpol in (c) überführen, indem die entsprechenden Komponenten in (b) zu den komplexen Widerständen Z_a und Z_b zusammengefasst werden:

[3] Die lokalen Druckänderungen in einer Schallwelle erfolgen adiabatisch.

Formel	Einheit	Bedeutung
$R_s = (\Delta x S/2A^2)\sqrt{\rho\mu\omega/2}$	g/cm^4s	Reibungsverluste an der Rohrwandung
$L_s = \Delta x\rho/2A$	g/cm^4	Trägheit der Luftmasse
$C_p = \Delta x A/\rho c^2$	cm^4s^2/g	Kompressibilität der Luft
$G_p = (\Delta x S(\eta-1)/\rho c^2)\sqrt{\lambda\omega/2\xi\rho}$	cm^4s/g	Wärmeverluste
$R_w = 1600/\Delta x S$	g/cm^4s	Realanteil der Wandimpedanz
$L_w = 1.5/\Delta x S$	g/cm^4	Imaginäranteil der Wandimpedanz
Δx	cm	Länge des Rohrabschnitts
A	cm^2	Querschnittsfläche des Rohrabschnitts
S	cm	Umfang des Rohrabschnitts
$\omega = 2\pi f$	s^{-1}	Kreisfrequenz
$\rho = 1.14\cdot 10^{-3}$	g/cm^3	Dichte der Luft
$c = 35300$	cm/s	Schallgeschwindigkeit
$\mu = 1.86\cdot 10^{-4}$	$dyn\cdot s/cm^2$	Viskositätskoeffizient
$\eta = 1.4$		adiabatische Gaskonstante
$\lambda = 5.5\cdot 10^{-5}$	$cal/cm\cdot s°C$	Koeff. der Wärmeleitfähigkeit
$\xi = 0.24$	$cal/g°C$	spezifische Wärme

Tabelle 5.1: Komponenten in Abb. 5.3 (b). Quellen: [26, 75].

$$Z_a = R_s + j\omega L_s, \tag{5.1}$$

$$Z_b = \frac{1}{j\omega C_p + G_p + \frac{1}{R_w + j\omega L_w}}. \tag{5.2}$$

Ausgehend davon kann man den Zusammenhang zwischen den Größen P_2 und U_2 am Ausgang sowie P_1 und U_1 am Eingang des Rohrabschnitts wie folgt formulieren:

$$\begin{pmatrix} P_2 \\ U_2 \end{pmatrix} = \begin{pmatrix} A & B \\ C & D \end{pmatrix} \cdot \begin{pmatrix} P_1 \\ U_1 \end{pmatrix} = K \cdot \begin{pmatrix} P_1 \\ U_1 \end{pmatrix}. \tag{5.3}$$

Mit Hilfe der Schaltungs- und Systemtheorie (siehe z.B. [7]) lässt sich zeigen, dass die Matrix

$$K = \begin{pmatrix} 1 + Z_a/Z_b & -2Z_a - Z_a^2/Z_b \\ -1/Z_b & 1 + Z_a/Z_b \end{pmatrix} \tag{5.4}$$

die gewünschte Beziehung herstellt.

Wenden wir uns nun dem gesamten Vokaltrakt zu. Wie bereits erwähnt lassen sich der Rachen- als auch der Mund- und Nasenraum getrennt voneinander als eine Aneinanderreihung mehrerer Rohrabschnitte auffassen, deren Querschnitte und Umfangslängen aus der Querschnitts- und Umfangsfunktion hervorgehen. Abb. 5.4 (a) zeigt das entsprechende elektrische Ersatzschaltbild des gesamten Vokaltrakts.

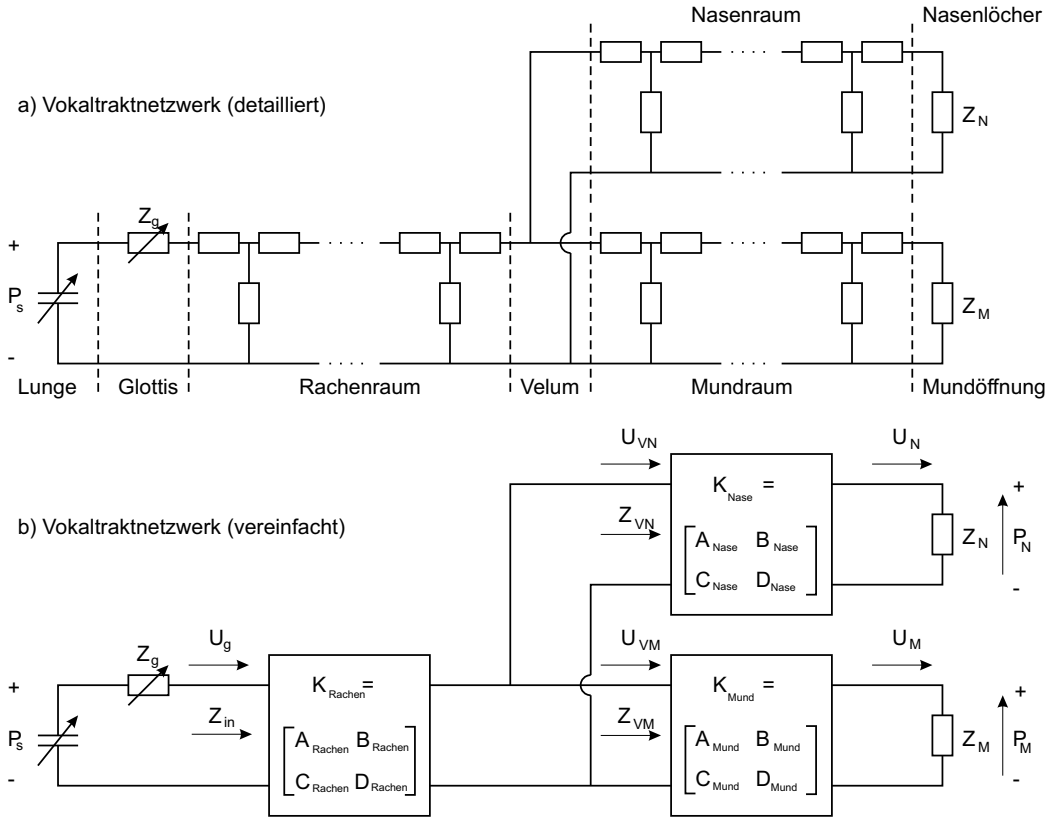

Abbildung 5.4: Elektrisches Ersatzschaltbild des gesamten Vokaltrakts. Nach [27, 75]. P_s ist der subglottale Druck und Z_g die glottale Impedanz (siehe Abschnitt 5.3.2). Die restlichen Größen werden in diesem Abschnitt behandelt.

Die Matrix für ein beliebiges Rohr aus mehreren Abschnitten erhält man, indem man die Matrizen der einzelnen Abschnitte miteinander multipliziert. Sei N die Anzal der Abschnitte und seien K_1 bis K_N die Matrizen der Rohrabschnitte. Wenn P_1 und U_1 der Schalldruck und der Volumenstrom am Eingang des ersten Rohrabschnitts sind und P_2 und U_2 die gleichen Größen am Ausgang des letzten Rohrabschnitts, dann erhält man die Matrix K für das gesamte Rohr durch die Multiplikation der Matrizen K_1 bis K_N in umgekehrter Reihenfolge.

$$K = K_N K_{N-1} \cdot \ldots \cdot K_2 K_1 \tag{5.5}$$

Darauf aufbauend lässt sich jeweils eine Matrix für den Rachen-, Mund- und Nasenraum berechnen, die wir mit K_{Rachen}, K_{Mund} und K_{Nase} bezeichnen wollen.

In Abb. 5.4 (b) sind die Abschnitte für die drei Teilräume zu jeweils einem einzigen Vierpol mit der dazugehörigen Matrix zusammengefasst.

5.2.2 Nasennebenhöhlen

Im Nasenraum befinden sich die Zugänge zu den Nasennebenhöhlen. Es ist erwiesen, dass die Nebenhöhlen deutlichen Einfluss auf die Klangfarbe von Nasallauten haben [39, 40]. Man geht davon aus, dass sie für die Nullstellen bzw. Antiresonanzen verantwortlich sind, die in der Übertragungsfunktion des Nasenraums gemessen werden können.
Eine Nasennebenhöhle ist ein luftgefüllter Hohlraum mit einem engen Hals und somit ein typischer Helmholtz-Resonator. Solch ein Resonator hat die Eigenschaft, eine bestimmte Frequenz aus einem Geräusch herauszufiltern. Bei der herausgefilterten Frequenz handelt es sich um die Eigenfrequenz f_0 des Resonators, die wie folgt berechnet wird [13]:

$$f_0 = \frac{c}{2\pi}\sqrt{\frac{A}{l \cdot V}}. \tag{5.6}$$

Darin ist V das Volumen des Hohlraums, l die Länge des Halses[4], A die Querschnittsfläche des Halses und c die Schallgeschwindigkeit (siehe Tab. 5.1).
Dang und Honda haben in [14] ein Modell für die Nasennebenhöhlen vorgeschlagen, das hier übernommen werden soll. Sie haben im Übertragungsspektrum von Nasallauten durchschnittlich vier Antiresonanzen lokalisiert, die sie jeweils einer Nebenhöhle zugeordnet haben. Die Nebenhöhlen wurden als Helmholtz-Resonatoren modelliert. Ihre Abmessungen sowie die Positionen, an denen sie an den Nasenraum angekoppelt sind, sind in Tab. 5.2 zusammengefasst.

	l. Keilbeinhöhle	r. Keilbeinhöhle	Kieferhöhle	Stirnhöhle
Volumen V (cm^3)	11.3	6.8	33	6.2
Halslänge l (cm)	0.3	0.3	0.45	1.0
Halsquerschnitt A (cm^2)	0.185	0.185	0.145	0.11
Abstand der Öffnung von den Nasenlöchern (cm)	6.5	6.5	5.0	4.0
Resonanzfrequenz (Hz)	1305	1682	552	749
Bandbeite (Hz)	188	216	108	124

Tabelle 5.2: Daten zu den vier Nebenhöhlen des Modells von Dang und Honda [14].

Die Anbindung der Resonatoren an den Nasenraum erfolgt mit Hilfe spezieller Kopplungsmatrizen. Diese werden zwischen den entsprechenden Rohrabschnitten in das Produkt der Matrizen K_1 bis K_N des Nasenraums eingefügt.
Abb. 5.5 zeigt das elektrische Ersatzschaltbild für zwei Rohrabschnitte des Nasenraums, zwischen denen ein Helmholtz-Resonator angekoppelt wurde. Das Netzwerk des Resonators besteht aus der akustischen Masse L_{sin}, dem Rohrwiderstand R_{sin}, der akustischen Nachgiebigkeit C_{sin} und der Leitfähigkeit G_{sin}, wobei der Index *sin* für *Sinus* (=Nebenhöhle) steht.
Die Größen werden mit den folgenden Formeln berechnet [21]:

[4] Nach [13] braucht hier nicht die übliche Endkorrektur der Halslänge vorgenommen werden.

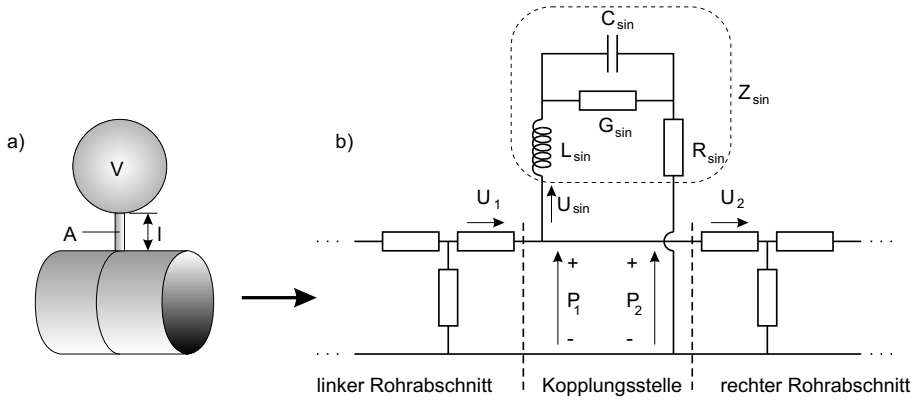

Abbildung 5.5: Einbindung einer Nasennebenhöhle als Helmholtz-Resonator mit der Impedanz Z_{sin} zwischen zwei Rohrabschnitte des Nasenraums.

$$L_{sin} = \frac{\rho \cdot l}{A} \tag{5.7}$$

$$C_{sin} = \frac{V}{\rho c^2} \tag{5.8}$$

$$G_{sin} = \frac{A_{Hohlraum}}{6500} \tag{5.9}$$

$$R_{sin} = \frac{2 \cdot l}{A}\sqrt{\frac{\pi\rho\mu\omega}{2A}}. \tag{5.10}$$

Die Konstanten ρ, μ und c sind in Tab. 5.1 gegeben. A ist wieder die Halsquerschnittsfläche, l die Halslänge und V das Hohlraumvolumen. $A_{Hohlraum}$ ist die Öberfläche des Resonatorhohlraums und lässt sich für einen kugelförmigen Resonator wie in Abb. 5.5 (a) mit der Formel

$$A_{Hohlraum} = 4\pi\left(\frac{3V}{4\pi}\right)^{\frac{2}{3}} \tag{5.11}$$

berechnen. Die Konstante 6500 in der Gleichung (5.9) für die akustische Leitfähigkeit wurde von Dang und Honda [14] experimentell bestimmt, um die Bandbreiten der Antiresonanzen mit den gemessenen Bandbreiten in Übereinstimmung zu bringen.
Der Gesamtwiderstand Z_{sin} des Resonators beträgt

$$Z_{sin} = R_{sin} + j\omega L_{sin} + \frac{1}{G + j\omega C_{sin}}. \tag{5.12}$$

Die Kopplungsmatrix K_{sin} erhält man, indem man in Abb. 5.5 (b) P_2 und U_2 in Beziehung zu P_1 und U_1 setzt. Mit dem Ansatz

$$P_2 = P_1 \tag{5.13}$$

$$U_2 = U_1 - U_{sin} = U_1 - \frac{P_1}{Z_{sin}} \tag{5.14}$$

erhält man für K_{sin} sofort

$$K_{sin} = \begin{pmatrix} 1 & 0 \\ -\frac{1}{Z_{sin}} & 1 \end{pmatrix}. \tag{5.15}$$

Durch die Einbindung aller vier Resonatoren aus Tabelle 5.2 erhält man sehr realistische Übertragungsfunktionen für die Nasallaute.

5.2.3 Abstrahlungsimpedanz

In Abb. 5.4 ist der Vokaltrakt an der Mund- und Nasenöffnung durch die Impedanzen Z_M und Z_N abgeschlossen. Diese sind für die Druckverluste verantwortlich, die durch die Schallabstrahlung entstehen.

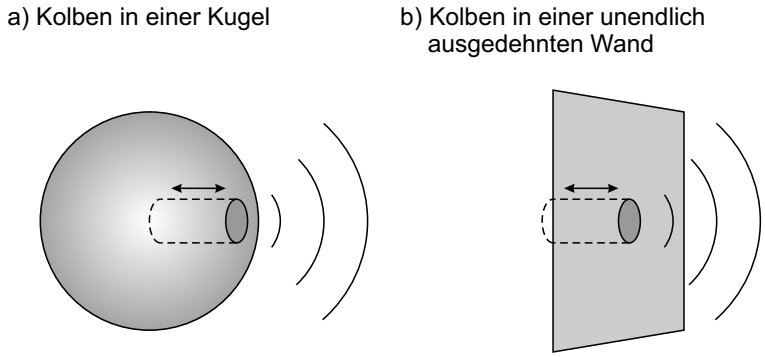

Abbildung 5.6: Schallabstrahlung von einem schwingenden Kolben in einer Kugel (a) und in einer unendlich ausgedehnten Wand (b). Die Doppelpfeile geben die Schwingungsrichtung der Kolben an. Quelle: [26].

In erster Näherung lässt sich die Schallabstrahlung an der Mund- oder Nasenöffnung mit der eines vibrierenden Kolbens in einer Kugel vergleichen (siehe Abb. 5.6 (a)). Die Kolbenquerschnittsfläche ist gleich der Öffnungsfläche vom Mund bzw. den Nasenlöchern, während die Kugel den Kopf repräsentiert. In [45] wurde die Abstrahlungsimpedanz für diesen Fall abgeleitet. Sie ist sowohl eine Funktion von der Frequenz als auch von der relativen Größe zwischen Kolben und Kopf. Die Funktion lässt sich jedoch nicht als geschlossener analytischer Ausdruck aufschreiben.
Ein Grenzfall liegt laut [26] aber vor, wenn die Größe des Kolbens sehr klein im Vergleich zum Kugeldurchmesser ist. Dann lässt sich die Abstrahlungsimpedanz als die eines Kolbens in einer ebenen, unendlich ausgedehnten Wand auffassen (siehe Abb. 5.6 (b)). Sie lässt sich in geschlossener Form als

$$Z = \frac{\rho c}{A}\left[1 - \frac{J_1(2ka)}{ka} + j\frac{S_1(2ka)}{ka}\right] \tag{5.16}$$

aufschreiben, wobei a der Radius des Kolbens, $k = \frac{2\pi f}{c}$ die Wellenzahl, A die Kolbenfläche und ρ und c die Dichte der Luft und die Schallgeschwindigkeit sind. $J_1(x)$ ist die Besselfunktion erster Ordnung und $S_1(x)$ die Struve-Funktion erster Ordnung.

$$J_1(x) = \frac{x}{2} - \frac{x^3}{2^3 \cdot 1! \cdot 2!} + \frac{x^5}{2^5 \cdot 2! \cdot 3!} - \frac{x^7}{2^7 \cdot 3! \cdot 4!} + \ldots \quad (5.17)$$

$$S_1(x) = \frac{2}{\pi}\left(\frac{x^2}{3} - \frac{x^4}{3^2 \cdot 5} + \frac{x^6}{3^2 \cdot 5^2 \cdot 7} - \frac{x^8}{3^2 \cdot 5^2 \cdot 7^2 \cdot 9} + \ldots\right) \quad (5.18)$$

Für kleine Werte für *ka* kann man Z durch die ersten zwei Terme von $J_1(x)$ und den ersten Term von $S_1(x)$ approximieren, und man erhält

$$Z = \frac{\rho c}{A}\left[\frac{(ka)^2}{2} + j\frac{8(ka)}{3\pi}\right]. \quad (5.19)$$

Diese Approximation für die Abstrahlungsimpedanz wird in der Literatur häufig verwendet.
Ein komplett anderer Ausdruck für Z wird z.B. in [75] vorgeschlagen:

$$Z = \frac{\rho\omega^2}{4\pi c}K(\omega) + j\frac{8\rho\omega}{3\pi\sqrt{\pi A}} \quad (5.20)$$

$$K(\omega) = \begin{cases} \frac{0.6\omega}{2\pi 1600} + 1 & \text{für } 0 \le \omega < 2\pi \cdot 1600 \\ 1.6 & \text{für } \omega \ge 2\pi \cdot 1600 \end{cases}$$

Ein objektiver Vergleich verschiedener Formeln für die Abstrahlungsimpedanz ist mir nicht bekannt. Deshalb habe ich mich (willkürlich) für Gleichung (5.16) entschieden, wobei $J_1(x)$ und $S_1(x)$ bis zum 8. Glied aufsummiert werden.

5.2.4 Eingangsimpedanz und Übertragungsfunktion

Mit Hilfe der Abstrahlungsimpedanzen Z_N (Nasenlöcher) und Z_M (Mundöffnung) lassen sich nun die Eingangsimpedanzen für den Nasenraum (Z_{VN}) und den Mundraum (Z_{VM}) berechnen (siehe Abb. 5.4 (b)). Die Eingangsimpedanz ist das Verhältnis vom Schalldruck zum Volumenstrom am Eingang einer akustischen Rohrleitung, also $\frac{P_1}{U_1}$. Für die Eingangsimpedanz des Nasenraums erhält man

$$Z_{VN} = \frac{D_{Nase}Z_N - B_{Nase}}{A_{Nase} - C_{Nase}Z_N}, \quad (5.21)$$

wobei A_{Nase} bis D_{Nase} die Elemente der Matrix K_{Nase} sind. Analog erhält man Z_{VM}, indem man A_{Nase} bis D_{Nase} durch die entsprechenden Matrix-Elemente von K_{Mund} ersetzt und Z_N mit Z_M austauscht.
Mit Z_{VN} und Z_{VM} können nun die globalen Matrizen K_N für den Nasaltrakt (Rachen- und Nasenraum) und K_V für den Vokaltrakt (Rachen- und Mundraum) berechnet werden.

$$K_N = K_{Nase} \cdot \begin{pmatrix} 1 & 0 \\ -\frac{1}{Z_{VM}} & 1 \end{pmatrix} \cdot K_{Rachen} \quad (5.22)$$

$$K_V = K_{Mund} \cdot \begin{pmatrix} 1 & 0 \\ -\frac{1}{Z_{VN}} & 1 \end{pmatrix} \cdot K_{Rachen} \quad (5.23)$$

Der mittlere Faktor auf der rechten Seite der Gleichungen ist jeweils die Kopplungsmatrix, mit der am Velum die Verbindung zum jeweils abzweigenden Schallpfad hergestellt wird. Die Kopplungsmatrix in Gleichung (5.23) muss natürlich nur dann einbezogen werden, falls der Nasenraum am Velum überhaupt geöffnet ist.
Die Übertragungsfunktion vom Volumenstrom durch die Stimmritze zum Schalldruck an der Mundöffnung kann nun mit

$$H_V = \frac{P_M}{U_G} = \frac{Z_M}{A_V - C_V Z_M} \tag{5.24}$$

berechnet werden, wobei A_V und C_V Elemente der Matrix K_V sind. H_V wird auf Null gesetzt, falls ein Verschluss im Mundraum existiert.
Die Übertragungsfunktion vom Volumenstrom durch die Stimmritze zum Schalldruck an den Nasenlöchern wird analog mit

$$H_N = \frac{P_N}{U_G} = \frac{Z_N}{A_N - C_N Z_N} \tag{5.25}$$

berechnet. H_N wird auf Null gesetzt, wenn der Nasaltrakt nicht an den Vokaltrakt angekoppelt ist, also die Querschnittsfläche des ersten Rohrabschnitts des Nasenraumes Null ist. Die gesamte Übertragungsfunktion des Sprechtrakts ist die Summe aus H_V und H_N:

$$H(f) = \frac{P_{Sprache}(f)}{U_G(f)} = H_V(f) + H_N(f). \tag{5.26}$$

Um die diskrete Impulsantwort $h[n]$ der Übertragungsfunktion zu erhalten, muss auf $H(f)$ die inverse Fourier-Transformation angewendet werden. Da $H(f)$ eine kontinuierliche Funktion ist, ist die kontinuierliche Impulsantwort $h(t)$ theoretisch unendlich lang. Durch die Energieverluste im Vokaltrakt (Wandhaftung, Abstrahlung,...) klingt die Impulsantwort jedoch relativ schnell ab. Deshalb genügt es, $h[n]$ z.B. für eine Länge von 512 Abtastwerten zu berechnen. Bei einer Abtastfrequenz von 22 kHz entspricht das einer Zeitdauer von 23.22 ms.
Um $h[n]$ für 512 Abtastwerte zu berechnen, muss $H(f)$ für 256 Frequenzen ermittelt werden. Dabei setze ich jedoch $H(f)$ für Frequenzen oberhalb von 4 kHz auf Null, da für sie die Annahme von der eindimensionalen Schallausbreitung im Vokaltrakt in der Regel nicht mehr zutrifft. Anschließend wird $h[n]$ wie in [58] mit der rechten Hälfte eines Hamming-Fensters multipliziert, um ein „sauberes“ Abklingen der Impulsantwort sicherzustellen. Ein (vollständiges) Hamming-Fenster $w[k]$ der Länge N erhält man nach [73] mit

$$w[k] = 0.54 - 0.46 \cdot \cos\left(k\frac{2\pi}{N-1}\right), \quad k \in \{0,1,\ldots,N-1\}. \tag{5.27}$$

Neben $h[n]$ müssen wir auch die Impulsantwort $z_{in}[n]$ der Eingangsimpedanz $Z_{in}(f)$ des Vokaltrakts berechen. $z_{in}[n]$ ist die invers Fouriertransformierte der Eingangsimpedanz $Z_{in}(f)$, die wie folgt berechnet wird:

$$Z_{in} = \frac{D_{Rachen} Z_{front} - B_{Rachen}}{A_{Rachen} - C_{Rachen} Z_{front}}. \tag{5.28}$$

Z_{front} ist die Eingangsimpedanz des Mund- und Nasenraums und damit die Parallelschaltung von Z_{VN} und Z_{VM}:

$$Z_{front} = \frac{Z_{VN} Z_{VM}}{Z_{VN} + Z_{VM}}. \tag{5.29}$$

Auch $Z_{in}(f)$ wird tiefpassgefiltert (alle Frequenzen über 4 kHz auf Null) und $z_{in}[n]$ mit der rechten Hälfte eines Hamming-Fensters multipliziert.

5.3 Modell für die stimmhafte Anregung

5.3.1 Einführung

Die stimmhafte Anregung (Phonation) des Sprechtrakts bei der Sprachproduktion erfolgt an den Stimmlippen (siehe Abschnitt 2.5.2). Dort findet die Umwandlung aerodynamischer Strömungsenergie in akustische Energie statt, indem eine regelmäßige Folge von Luftstrompulsen erzeugt wird. Die Energieumwandlung findet über die Zwischenstufe der mechanisch-dynamischen Stimmlippenschwingung statt. Die Stimmlippenschwingung führt zu einer zeitlich modulierten Öffnungsfläche der Stimmritze, und diese wiederum zu einem zeitlich modulierten Luftstrom. Die sich dadurch zeitlich ändernden Druckverhältnisse in der Glottis sorgen – zusammen mit einer geeigneten Voreinstellung der Stimmlippen durch den Kehlkopf – für eine Aufrechterhaltung der Schwingung.

Es gibt verschiedene Möglichkeiten für die Berechnung des Volumenstroms durch die Stimmritze. Neben einer umfassenden physiologischen Modellierung des Kehlkopfes bieten sich eine aerodynamisch-mechanische Simulation der Stimmlippenschwingung oder eine direkte Parametrisierung der glottalen Öffnungsfläche oder des Volumenstroms an.

Zu einer streng physiologisch basierten Simulation des Kehlkopfes existieren bisher nur wenige Ansätze (z.B. [24, 25]), da sie aufgrund des komplexen Zusammenspiels der Knorpel, Muskeln und Bänder sehr schwierig erscheint. Eine Parametrisierung der glottalen Öffnungsfläche wurde z.B. in [67] und [69] vorgeschlagen. Eine Abstraktionsstufe weiter geht die unmittelbare Beschreibung der Volumenstromfunktion mit Hilfe analytischer Ausdrücke für die steigende und fallende Flanke eines Luftstrompulses. Solche Formeln wurden z.B. in [51] und [74] veröffentlicht.

Die selbstschwingenden Glottismodelle enthalten eine Simulation der Stimmlippenschwingung aufgrund der Luftstrom- und Druckverhältnisse in der Glottis. Diese Modelle basieren auf der myoelastisch-aerodynamischen Theorie der Phonation, nach der die Schwingungsenergie der Stimmlippen aus der aerodynamischen Energie gewonnen wird.

Starke Verbreitung hat das von Ishizaka und Flanagan [32] entwickelte Zwei-Massen-Modell der Stimmlippen gefunden. Der Erfolg dieses Modells beruht auf dem gelungenen Kompromiss zwischen Einfachheit und Vollständigkeit. Deshalb wird es auch in dieser Arbeit verwendet und die Einbindung in die akustische Synthese beschrieben.

Weitere selbstschwingende Modelle sind z.B. das Drei-Massen-Modell von Story und Titze [64] und das kontinuierlich-mechanische Modell von Titze [65].

5.3.2 Zwei-Massen-Modell

Laut [58] wurde in verschiedenen Studien nachgewiesen, dass das Glottismodell von Ishizaka und Flanagan trotz seiner Einfachheit sehr realistische Eigenschaften besitzt.
Dieser und der nächste Abschnitt sollen dazu dienen, das Modell in seinen Grundzügen vorzustellen und die Formeln zu seiner numerischen Simulation abzuleiten. Einige Verbesserungsvorschläge für das Modell aus [58] wurden übernommen.
Hauptaufgabe des Glottismodells ist es, den zeitlich veränderlichen Volumenstrom $u_g(t)$ durch die Glottis zu berechnen. Dieser hängt neben der Ansatzrohrgeometrie von den Voreinstellungen der Stimmlippen durch die Kehlkopfmuskulatur und dem subglottalen Druck ab. Der subglottale Druck und die Voreinstellung der Stimmlippen sind als „Anregungsparameter“ in Tab. 5.3 zusammengefasst.

Bezeichnung	Wertebereich	Einheit	Bestimmung
p_s	[0, 9810]	$\mathrm{dyn/cm^2}$	subglottaler Druck
A_{Ruhe}	[0, 1]	$\mathrm{cm^2}$	neutrale Öffnungsfläche der Glottis
q	[0.1, 1.2]		Stimmlippenspannung
g_s^2	[0.15, 1.85]		Widerstandsregler

Tabelle 5.3: Anregungsparameter des Glottismodells. Quellen: [32, 58].

Betrachten wir nun den Aufbau des Glottismodells anhand von Abb. 5.7 (a). Jede Stimmlippe besteht aus zwei Massestücken m_1 und m_2, die übereinander angeordnet sind und den unteren bzw. oberen Teil der Stimmlippe repräsentieren. Da die linke und rechte Stimmlippe symmetrisch zur mediosagittalen Ebene verlaufen, genügt es, eine Seite zu modellieren. Die Auslenkung der Massen aus ihrer Ruhelage ist nur in horizontaler Richtung erlaubt und sei mit x_1 bzw. x_2 bezeichnet. Jede Masse ist durch eine Feder und ein Dämpfungsglied mit dem Schildknorpel verbunden. Die Federkräfte sind $s_1(x_1)$ und $s_2(x_2)$ und die Reibungskoeffizienten sind r_1 und r_2. Zusätzlich sind beide Massestücke über eine Feder mit der Federkonstanten k_c verbunden. Die vertikalen Ausdehnungen der Massestücke sind d_1 und d_2 und die Länge der Stimmritze bzw. die Tiefe der Massen ist l_g. Die Länge der Stimmritze ist identisch mit der Länge der Stimmlippen, d.h. es kommt beim Aufeinandertreffen der jeweils linken und rechten unteren oder oberen Massen zum vollständigen Verschluss der Stimmritze.
A_0 ist der Flächeninhalt des ersten Rohrabschnitts des Vokaltrakts. Die Flächeninhalte A_1 und A_2 sind Funktionen von x_1 und x_2 und bestimmen die glottale Impedanz Z_g, also den Widerstand, den die Stimmritze dem Volumenstrom entgegensetzt.
Der Volumenstrom durch die Glottis ist – analog zum Ohmschen Gesetz – der Quotient aus der Differenz der Drücke direkt unter und über den Stimmlippen (Druckabfall Δp) sowie der glottalen Impedanz:

$$u_g(t) = \frac{\Delta p(t)}{Z_g(t)} = \frac{p_s(t) - p_0(t)}{Z_g(t)}. \tag{5.30}$$

p_s ist der subglottale Druck und p_0 ergibt sich durch die Faltung von u_g mit der Impulsantwort z_{in} der Eingangsimpedanz des Vokaltrakts. u_g und p_0 sind also voneinander abhängig

und müssen simultan berechnet werden.

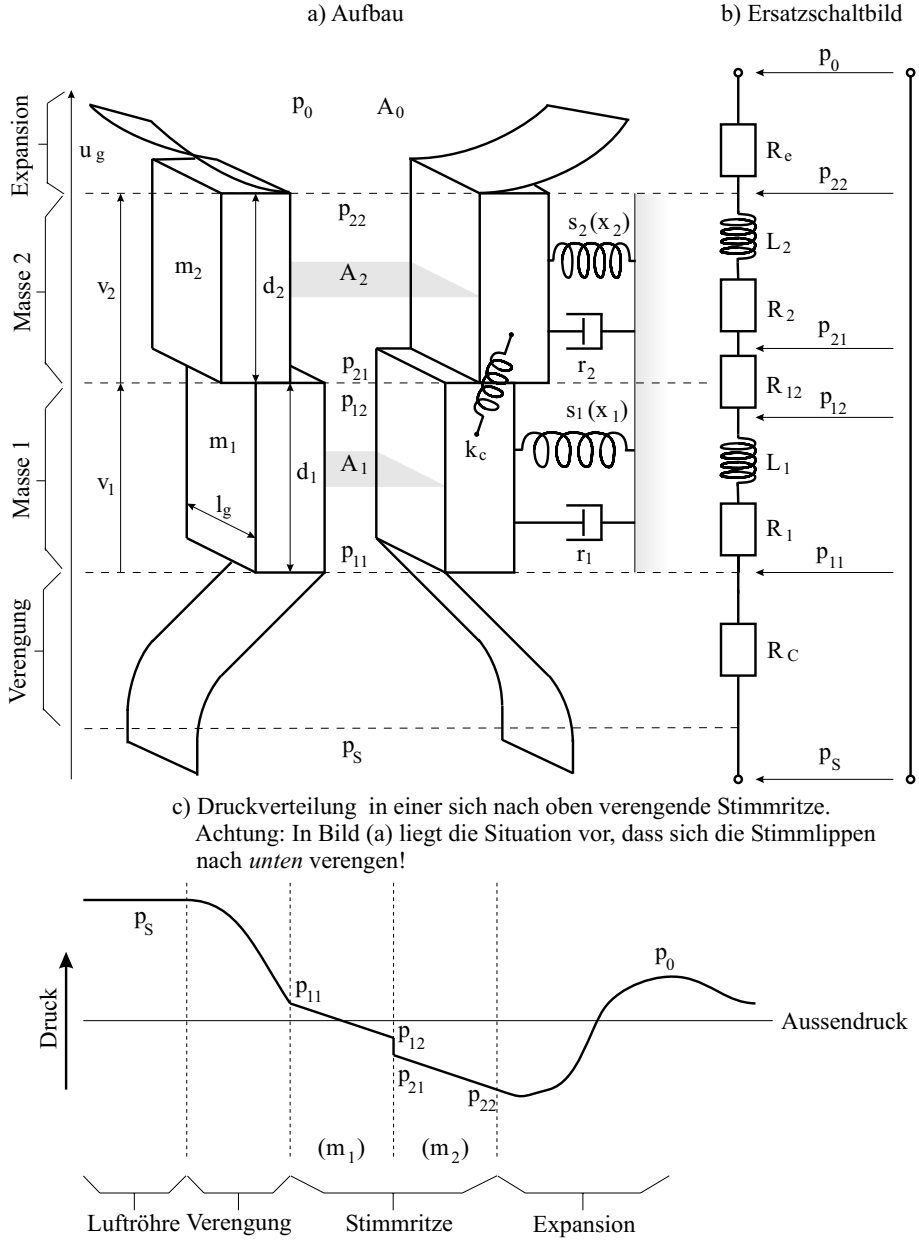

Abbildung 5.7: Stimmbändermodell von Ishizaka und Flanagan [32].

Schauen wir uns nun die glottale Impedanz Z_g genauer an. Abb. 5.7 (b) zeigt das elektrische Ersatzschaltbild der Stimmritze mit mehreren Widerständen. Die Widerstände sind jeweils in der Höhe angeordnet, wo sie in Abb. 5.7 (a) entstehen. Z_g ist die Summe dieser Widerstände und besitzt somit einen ohmschen und einen induktiven Anteil. Bei der Ermittlung der Einzelwiderstände gehen wir von einer stabilen, eindimensionalen Luftströmung durch die Glottis aus. Deshalb kann die Bernoulli-Gleichung benutzt werden, um die Druckverteilung im Luftstrom zu berechnen. Aus den Druckdifferenzen können die Widerstände abgeleitet werden.
Die Strömungsgeschwindigkeiten im Bereich der unteren und oberen Masse betragen

$v_1 = \frac{u_g}{A_1}$ und $v_2 = \frac{u_g}{A_2}$. Im Folgenden sind ρ und μ wieder die Dichte und die Viskosität der Luft.

1. *Verengung/Einlassbereich*
   Der Druckverlust im unteren Eingangsbereich zur Glottis beträgt

$$p_s - p_{11} = (1+0.37)\frac{\rho}{2}v_1^2 = 1.37\frac{\rho}{2}\frac{u_g^2}{A_1^2} \tag{5.31}$$

   und ist damit um den Faktor 0.37 größer als der reine Druckverlust durch die Zunahme der Strömungsgeschwindigkeit von 0 cm/s auf v_1. Der zusätzliche Verlustfaktor wurde aus Strömungsexperimenten an einem Kehlkopfmodell abgeleitet [72]. Für R_c erhalten wir

$$R_c = 1.37\frac{\rho}{2}\frac{|u_g|}{A_1^2}. \tag{5.32}$$

   Die Trägheit der Luftmassen im Einlassbereich ist laut [32] vernachlässigbar.

2. *Unterer Teil der Stimmlippen (Masse 1)*
   Der Druckabfall im unteren Teil der Stimmritze ist auf die viskose Reibung innerhalb der Strömung und die Trägheit der Luftmassen zurückzuführen. Die Druckabnahme erfolgt linear über die gesamte Höhe des unteren Abschnitts, und lässt sich durch die Formel

$$p_{11} - p_{12} = u_g d_1 \frac{12\mu l_g^2}{A_1^3} + \frac{\rho d_1}{A_1}\cdot\frac{du_g}{dt} \tag{5.33}$$

   ausdrücken [72]. Für den Widerstand R_1 und die Induktivität L_1 gilt also

$$R_1 = d_1\frac{12\mu l_g^2}{A_1^3}, \tag{5.34}$$

$$L_1 = \frac{\rho d_1}{A_1}. \tag{5.35}$$

3. *Übergang vom unteren zum oberen Teil*
   An der Grenze zwischen den Massen m_1 und m_2 ist u_g kontinuierlich, aber die Teilchengeschwindigkeit ändert sich von v_1 nach v_2. Daraus resultiert die Druckänderung

$$p_{12} - p_{21} = \frac{\rho}{2}u_g^2\left(\frac{1}{A_2^2} - \frac{1}{A_1^2}\right) \tag{5.36}$$

   bzw. der Widerstand

$$R_{12} = \frac{\rho}{2}|u_g|\left(\frac{1}{A_2^2} - \frac{1}{A_1^2}\right). \tag{5.37}$$

4. *Oberer Teil der Stimmlippen (Masse 2)*
   Der Druckabfall und die Widerstände im oberen Teil der Stimmlippen werden analog zu denen im unteren Teil berechnet.

$$p_{21} - p_{22} = u_g d_2\frac{12\mu l_g^2}{A_2^3} + \frac{\rho d_2}{A_2}\cdot\frac{du_g}{dt} \tag{5.38}$$

$$R_2 = d_2 \frac{12\mu l_g^2}{A_2^3} \tag{5.39}$$

$$L_2 = \frac{\rho d_2}{A_2} \tag{5.40}$$

5. *Expansion/Auslassbereich*
   Die Druckänderung im Auslassbereich beträgt laut [32]

$$p_{22} - p_0 = -\frac{\rho}{2}\left(\frac{u_g}{A_2}\right)^2 \cdot 2 \cdot \frac{A_2}{A_0}\left(1 - \frac{A_2}{A_0}\right). \tag{5.41}$$

   Daraus folgt für den Widerstand R_e

$$R_e = -\frac{\rho}{A_2 A_0} \cdot \left(1 - \frac{A_2}{A_0}\right)|u_g|. \tag{5.42}$$

Abb. 5.7 (c) zeigt beispielhaft die Druckverteilung in einer Stimmritze, die im oberen Teil enger ist als im unteren.
Für die Summe R der ohmschen Widerstände und die Summe L der induktiven Widerstände erhält man die folgenden Ausdrücke:

$$R = 12\mu l_g^2\left(\frac{d_1}{A_1^3} + \frac{d_2}{A_2^3}\right) + \frac{\rho}{2}|u_g|\left(\frac{0.37}{A_1^2} + \frac{1}{A_2^2} - \frac{2}{A_2 A_0} + \frac{2}{A_0^2}\right) = R_A + |u_g| R_B \tag{5.43}$$

$$L = \rho\left(\frac{d_1}{A_1} + \frac{d_2}{A_2}\right). \tag{5.44}$$

Gleichung (5.30) lässt sich nun also in der Form

$$p_s - p_0 = u_g \cdot R + \frac{du_g}{dt} \cdot L \tag{5.45}$$

schreiben.

Aus der Druckverteilung in der Stimmritze lassen sich jetzt die auf ihre Seitenflächen wirkenden Kräfte ableiten.
Abb. 5.8 (a) zeigt die Stimmlippen in ihrer Ruheposition, also wenn sowohl der Volumenstrom durch die Stimmritze als auch die Auslenkungen x_1 und x_2 Null sind. Dieser Zustand ist die Voreinstellung der Stimmlippen durch den Kehlkopf und wird durch die „neutrale“ Öffnungsfläche A_{Ruhe} (grau schraffiert) charakterisiert. Für beliebige Auslenkungen x_1 und x_2 lassen sich die Querschnittsflächen im unteren und oberen Teil der Stimmritze nach der Formel

$$A_i = \begin{cases} A_{Ruhe} + 2 l_g x_i & \text{falls } x_i > x_{min} \\ 0 & \text{sonst} \end{cases}$$

berechnen, wobei $i \in \{0,1\}$ und $x_{min} = -A_{Ruhe}/2l_g$.

Betrachten wir zuerst die Kräfte F_1 und F_2, die durch den (statischen) Luftdruck

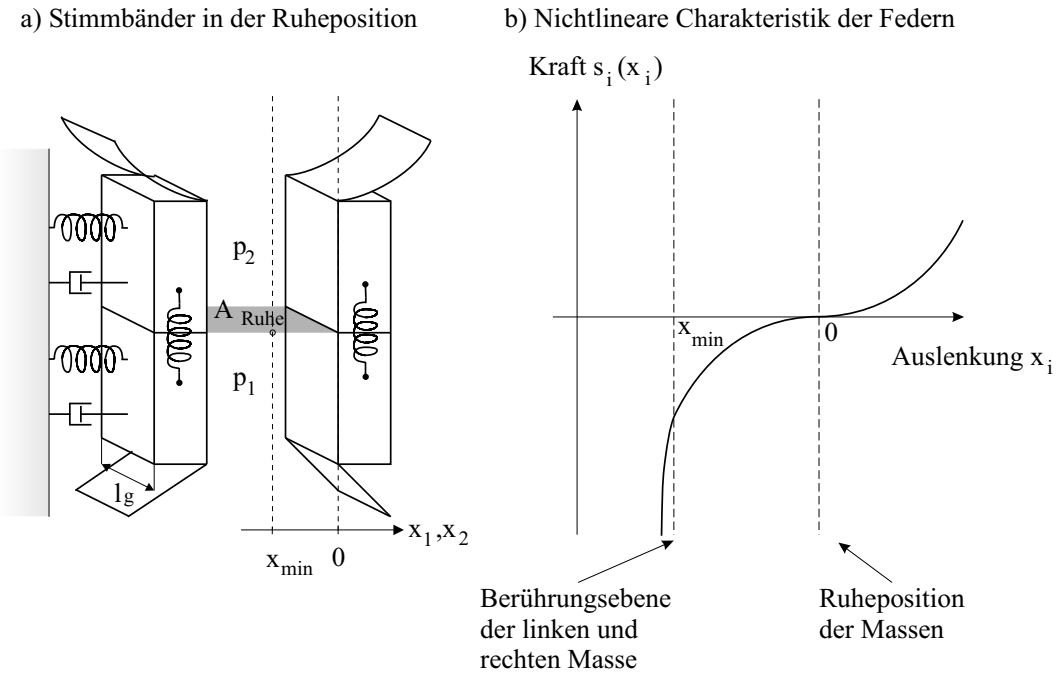

Abbildung 5.8: Dargestellt sind die Stimmlippen im Ruhezustand (a) sowie die nichtlinearen Federkräfte (b).

auf die Seitenflächen der Stimmritze einwirken. Die mittleren Drücke p_1 und p_2 im unteren und oberen Teil der Stimmritze sind

$$\begin{aligned} p_1 &= \frac{1}{2}(p_{11}+p_{12}) \\ &= p_{11}-\frac{1}{2}(p_{11}-p_{12}) \\ &= p_s - 1.37\frac{\rho}{2}\left(\frac{u_g}{A_1}\right)^2 - \frac{1}{2}\left(R_1 u_g + L_1\frac{du_g}{dt}\right) \end{aligned}$$

und

$$\begin{aligned} p_2 &= \frac{1}{2}(p_{21}+p_{22}) \\ &= p_1 - \frac{1}{2}(p_{11}-p_{12}) - (p_{12}-p_{21}) - \frac{1}{2}(p_{21}-p_{22}) \\ &= p_1 - \frac{\rho}{2}u_g^2\left(\frac{1}{A_2^2}-\frac{1}{A_1^2}\right) - \frac{1}{2}\left((R_1+R_2)u_g + (L_1+L_2)\frac{du_g}{dt}\right). \end{aligned}$$

Für die tatsächlich wirksamen Kräfte muss die Unterscheidung vorgenommen werden, ob sich die linke und rechte Stimmlippe im oberen bzw. unteren Teil berühren oder nicht. Tab. 5.4 fasst die verschiedenen Fälle zusammen.

Neben den Kräften F_1 und F_2 wirken auch die Federkräfte $s_1(x_1)$ und $s_2(x_2)$ auf die beiden Massestücke. Die Federn besitzen eine nichtlineare Charakteristik, die laut [32] an herausgeschnittenen menschlichen Stimmbändern gemessen wurde. Für den Fall, dass die Stimmritze geschlossen ist, muss ein zusätzlicher Term zur Federkraft addiert werden, der

x_1	x_2	$F_1/l_g d_1$	$F_2/l_g d_2$
$x_1 > x_{min}$	$x_2 > x_{min}$	p_1	p_2
$x_1 \leq x_{min}$	$x_2 > x_{min}$	p_s	0
$x_1 > x_{min}$	$x_2 \leq x_{min}$	p_s	p_s
$x_1 \leq x_{min}$	$x_2 \leq x_{min}$	p_s	0

Tabelle 5.4: Kräfte auf das untere und obere Massestück unter verschiedenen Bedingungen. $l_g d_1$ und $l_g d_2$ sind die Seitenflächeninhalte der Stimmritze im unteren bzw. oberen Teil.

die Rückstellkraft durch das an der Kontaktstelle deformierte Gewebe repräsentiert.

$$s_i(x_i) = \begin{cases} k_i(x_i + \eta_{ki} \cdot x_i^3) & \text{für } x_i > x_{min} \\ k_i(x_i + \eta_{ki} \cdot x_i^3) + h_i\left((x_i - x_{min}) + \eta_{hi}(x_i - x_{min})^3\right) & \text{für } x_i \leq x_{min} \end{cases} \tag{5.46}$$

Ein Graph der Funktion $s_i(x_i)$ ist in Abb. 5.8 (b) dargestellt.
Eine zusätzlich Feder mit der linearen Federkonstante k_c verbindet das untere mit dem oberen Massestück. Dadurch wird die Biegesteifigkeit der Stimmlippen in die laterale Richtung repräsentiert.

Jedes Massestück besitzt einen Dämpfer, der die viskose Reibung innerhalb der Stimmlippen nachbildet. Die Widerstände r_1 und r_2 sind

$$r_1 = \frac{2\xi_1\sqrt{m_1 k_1}}{g_s^2} \quad \text{und} \quad r_2 = \frac{2\xi_2\sqrt{m_2 k_2}}{g_s^2}, \tag{5.47}$$

wobei k_1 und k_2 die linearen Komponenten der Federsteifigkeit in Gleichung (5.46) sind. Die Division durch den Anregungsparameter g_s^2 war in der ursprünglichen Veröffentlichung von Ishizaka und Flanagan nicht vorgesehen, sondern wurde erst später von Sondhi und Schroeter [58] (bzw. Coker) vorgeschlagen, um realistische Übergänge von Vokalen zu stimmlosen Plosiven und Frikativen zu ermöglichen.
Für ξ_1 und ξ_2 gilt:

$$\xi_1 = \begin{cases} 0.1 & \text{für } x_1 > x_{min} \\ 1.1 & \text{für } x_1 \leq x_{min} \end{cases} \quad \text{und} \quad \xi_2 = \begin{cases} 0.6 & \text{für } x_2 > x_{min} \\ 1.6 & \text{für } x_2 \leq x_{min} \end{cases}.$$

Nun können die Bewegungsgleichungen für die beiden Massestücke formuliert werden:

$$F_1 = m_1\ddot{x}_1 + r_1\dot{x}_1 + s_1(x_1) + k_c(x_1 - x_2), \tag{5.48}$$
$$F_2 = m_2\ddot{x}_2 + r_2\dot{x}_2 + s_2(x_2) + k_c(x_2 - x_1). \tag{5.49}$$

Die Lösungen dieser zwei voneinander abhängigen Differenzialgleichungen sind die Auslenkungen x_1 und x_2.
Die Werte aller im Text noch nicht gegebenen Variablen sind in Tab. 5.5 zusammengefasst.

Variable	Bedeutung	Wert	Einheit
m_1	untere Masse der Stimmlippen	$0.125/q$	g
m_2	obere Masse der Stimmlippen	$0.025/q$	g
d_1	Dicke der unteren Masse	$0.25/q$	cm
d_2	Dicke der oberen Masse	$0.05/q$	cm
η_{k1}	nichtlinearer Federkoeffizient	100	—
η_{k2}	nichtlinearer Federkoeffizient	100	—
η_{h1}	nichtlinearer Federkoeffizient	500	—
η_{h2}	nichtlinearer Federkoeffizient	500	—
h_1	nichtlinearer Federkoeffizient	$3k_1$	dyn/cm
h_2	nichtlinearer Federkoeffizient	$3k_2$	dyn/cm
k_1	linearer Federkoeffizient	$80000 \cdot q$	dyn/cm
k_2	linearer Federkoeffizient	$8000 \cdot q$	dyn/cm
k_c	Kopplungsfederkoeffizient	$25000 \cdot q^2$	dyn/cm

Tabelle 5.5: Variablen des Zwei-Massen-Modells. Quelle: [58].

Der Anregungsparameter q ist ein Maß für die Längsspannung der Stimmlippen. Über ihn wird die Grundfrequenz gesteuert. Eine Zunahme von q bewirkt gemäß Tab. 5.5 eine Abnahme der Dicke und der Massen der Stimmlippen sowie eine Zunahme der Federsteifigkeiten. Dadurch erhöht sich die Grundfrequenz in etwa proportional zu q. Umgekehrt führt eine Verringerung des Parameters q zu einer Abnahme der Grundfrequenz.
Für den Vokal /a:/ ergibt sich mit den Anregungsparametern $q = 0.9$, $g_s^2 = 0.5$, $A_{Ruhe} = 0.05$ cm^2 und $p_s = 8000$ dyn/cm^2 ungefähr eine Grundfrequenz von 142 Hz.

5.3.3 Numerische Simulation

Für die numerische Simulation werden die Differenzialgleichungen aus dem vorigen Abschnitt durch Differenzengleichungen angenähert. Dazu werden die ersten und zweiten Ableitungen einer beliebigen Funktion $f(t)$ durch

$$\frac{df(t)}{dt} \simeq \frac{f[n]-f[n-1]}{T} \quad \text{und} \quad \frac{d^2f(t)}{dt^2} \simeq \frac{f[n]-2f[n-1]+f[n-2]}{T^2} \tag{5.50}$$

ausgedrückt, wobei T das Abtastintervall ist.
Schauen wir uns zuerst die Berechnung des aktuellen Volumenstromabtastwertes $u_g[n]$ an. Ausgangspunkt dafür ist die Diskretisierung der Gleichungen (5.43) und (5.45).

$$p_s[n] - p_0[n] = R[n]u_g[n] + L[n]\frac{u_g[n]-u_g[n-1]}{T} \tag{5.51}$$

$$R[n] = R_A[n] + u_g[n]R_B[n] \tag{5.52}$$

Der supraglottale Druck p_0 ergibt sich aus der Faltung von u_g mit der Impulsantwort der Eingangsimpedanz z_{in}. Der aktuelle Abtastwert für den supraglottalen Druck lässt sich

damit folgendermaßen berechnen:

$$p_0[n] = u_g[n]z_{in}[0] + \sum_{i=1}^{N-1} u_g[n-i]z_{in}[i] = C[n]u_g[n] + D[n]. \tag{5.53}$$

N ist die Länge der (bekannten) Impulsantwort $z_{in}[i]$. Ansonsten kommen nur der aktuelle und zurückliegende Abtastwerte von u_g in der Formel vor.
Das Einsetzen von (5.52) und (5.53) in (5.51) ergibt

$$p_s - Cu_g[n] - D = R_A u_g[n] + R_B u_g^2[n] + L \cdot \frac{u_g[n] - u_g[n-1]}{T}. \tag{5.54}$$

Außer für u_g wurde die Abhängigkeit der restlichen Größen von n nicht mehr explizit angegeben. Das Umstellen von (5.54) ergibt die quadratische Gleichung

$$0 = u_g^2[n] + u_g[n]\frac{TR_A + L + TC}{TR_B} + \frac{TD - Tp_s - Lu_g[n-1]}{TR_B}. \tag{5.55}$$

Von den beiden Lösungen wird die größere für $u_g[n]$ ausgewählt. Anschließend kann $p_0[n]$ mit Gleichung (5.53) berechnet werden.

Nun lassen sich die aktuellen Auslenkungen $x_1[n]$ und $x_2[n]$ der beiden Massestücke ermitteln, indem die Gleichungen (5.48) und (5.49) diskretisiert und gelöst werden. Die Diskretisierung liefert

$$\begin{aligned} F_1 &= m_1\frac{x_1[n] - 2x_1[n-1] + x_1[n-2]}{T^2} + r_1\frac{x_1[n] - x_1[n-1]}{T} + s_1(x_1[n-1]) + k_c(x_1[n] - x_2[n]) \\ F_2 &= m_2\frac{x_2[n] - 2x_2[n-1] + x_2[n-2]}{T^2} + r_2\frac{x_2[n] - x_2[n-1]}{T} + s_2(x_2[n-1]) + k_c(x_2[n] - x_1[n]). \end{aligned}$$

F_1 und F_2 lassen sich mit Hilfe von $u_g[n]$ und Tab. 5.4 berechnen. m_1, m_2, T und k_c sind bekannt. Die Reibungswiderstände r_1 und r_2 ergeben sich aus Gleichung (5.47). Man beachte, dass die Kraftfunktionen s_1 und s_2 um einen Abtastwert verzögert wurden, um sie bereits ohne die Kenntnis von $x_1[n]$ und $x_2[n]$ berechnen zu können.
Die diskreten Bewegungsgleichungen lassen sich als das lineare Gleichungssystem

$$\begin{pmatrix} A & B \\ C & D \end{pmatrix} \cdot \begin{pmatrix} x_1[n] \\ x_2[n] \end{pmatrix} = \begin{pmatrix} E \\ F \end{pmatrix} \tag{5.56}$$

schreiben, in dem die Größen A bis F durch

$$\begin{aligned} A &= m_1 + r_1T + k_cT^2 \\ B &= -k_cT^2 \\ C &= -k_cT^2 \\ D &= m_2 + r_2T + k_cT^2 \\ E &= F_1T^2 + 2m_1x_1[n-1] - m_1x_1[n-2] + r_1Tx_1[n-1] - s_1(x_1[n-1])T^2 \\ F &= F_2T^2 + 2m_2x_2[n-1] - m_2x_2[n-2] + r_2Tx_2[n-1] - s_2(x_2[n-1])T^2 \end{aligned}$$

gegeben sind.
Für $x_1[n]$ und $x_2[n]$ erhält man dann:

$$x_1[n] = \frac{\begin{vmatrix} E & B \\ F & D \end{vmatrix}}{\begin{vmatrix} A & B \\ C & D \end{vmatrix}}, \qquad x_2[n] = \frac{\begin{vmatrix} A & E \\ C & F \end{vmatrix}}{\begin{vmatrix} A & B \\ C & D \end{vmatrix}}. \tag{5.57}$$

Für die Implementierung des Glottismodells auf dem Computer bietet es sich an, für jede der Größen x_1, x_2, u_g und p_0 einen Ringpuffer anzulegen, um bei der numerischen Simulation bequem auf ihre jeweils letzten Abtastwerte zugreifen zu können.

Kapitel 6

Steuerung der Artikulatoren

Das in Abschnitt 4 erarbeitete Modell erlaubt es, den Zustand der Artikulatoren zu einem bestimmten Zeitpunkt mit Hilfe der Vokaltraktparameter auszudrücken. Für eine zeitliche Steuerung der Artikulation sind alle Parameter als Funktionen der Zeit aufzufassen. Für einzelne gehaltene stimmhafte Laute können die Funktionswerte der Parameter für den Zeitraum der Äußerung als konstant angesehen werden. Für den Übergang zwischen zwei verschiedenen Lauten müssen die Parameter dagegen auf geeignete Weise ihren Wert anpassen. In diesem Abschnitt wird eine einfache Methode vorgestellt, mit der die Funktionsverläufe der Parameter über mehrere Laute hinweg generiert werden können.

Wenn ein ruhender Artikulator – wie z.B. die Zunge – von seiner derzeitigen Position an eine neue Position bewegt werden soll, geschieht dies in der Regel durch die Aktivität der angreifenden Muskeln. Man kann sich vorstellen, dass die Bewegung des Artikulators nicht gleichförmig erfolgt, sondern dass er von seiner Ausgangsposition aus in Richtung der Zielposition zuerst beschleunigt wird, um anschließend rechtzeitig wieder abgebremst zu werden. Obwohl der Sachverhalt in der Realität komplizierter ist, lässt er sich dennoch im Falle der Übergänge (Transitionen) zwischen vokalischen Lauten leicht auf die Parameterfunktionen übertragen. Mermelstein [42] definiert die Transition des Parameters $X(t)$ zwischen zwei Vokalen als

$$X(t) = \frac{1}{2}\left[X(t_a) + X(t_b) + (X(t_a) - X(t_b))\cos\left(\pi\frac{t - t_a}{t_b - t_a}\right)\right], \tag{6.1}$$

wobei $X(t_a)$ und $X(t_b)$ der Start- bzw. Zielwert des Parameters sind und t_a und t_b der Start- und Endzeitpunkt der Transition.
Die Übergänge von und zu Plosiven und Frikativen lassen sich in der Regel nicht so einfach beschreiben, da hier noch andere Faktoren die Dynamik der Artikulation beeinflussen. Diese sollen uns an dieser Stelle jedoch nicht interessieren.

Im Rahmen dieser Arbeit wurde eine Methode entwickelt, mit der die einzelnen Parameterfunktionen für eine beliebig lange zusammenhängende Kette vokalischer Laute erzeugt werden können. Diese Methode wird sowohl für die Vokaltraktparameter als auch die Anregungsparameter verwendet. Dabei bietet sie für die Transitionen mehr Flexibilität als Gleichung (6.1).

Eine beliebige Äußerung (*PhoneChain*) wird als Verkettung von N Phonen aufgefasst

$$PhoneChain = Phone_0 Phone_1 ... Phone_{N-1}, \tag{6.2}$$

während die Phone $Phone_0$ bis $Phone_{N-1}$ durch jeweils ein Tupel

$$Phone = (t_s, duration, \{Param_0, Param_1, ..., Param_{M-1}\}) \tag{6.3}$$

repräsentiert werden.
$\{Param_0, Param_1, ..., Param_{M-1}\}$ ist die Menge der M Anregungs- und Vokaltraktparameter und t_s und $duration$ sind der Startzeitpunkt und die Dauer des Phons. Die Laute werden also zeitlich exakt voneinander abgegrenzt, wobei der Verlauf der Parameterfunktionen im entsprechenden Zeitintervall durch die Tupel $Param_0$ ... $Param_{M-1}$ definiert wird.
Jedes dieser Tupel hat die Form

$$Param = (isFloating, y_m, y_l, y_r, c_l, c_r, s_l, s_r). \tag{6.4}$$

Die Bedeutung der einzelnen Tupelwerte ergibt sich aus den nachfolgenden Beschreibungen.
In Abb. 6.1 (b) und (c) sieht man, wie der Funktionsverlauf eines Parameters im Zeitintervall eines Phons durch die Punkte Q, R, S und T bestimmt wird. Jeder Punkt besitzt jeweils eine t- und eine y-Koordinate (z.B. $Q = (Q_t, Q_y)^T$), deren Werte sich aus dem *Param*-Tupel ergeben.

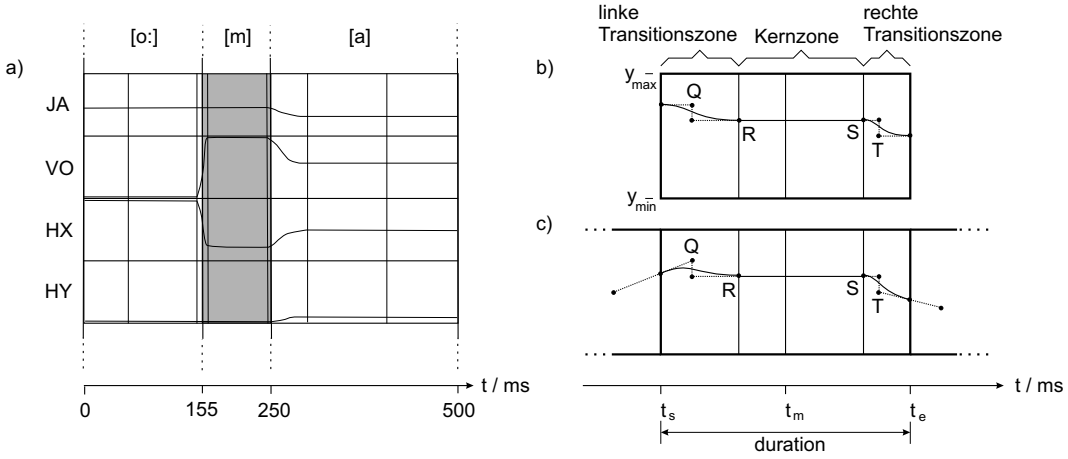

Abbildung 6.1: Vokaltraktparameter als Funktion der Zeit für Worte aus mehreren Phonen. Die in (a) dargestellten Parameter sind der Kieferöffnungswinkel (*JA*), die Öffnung des velopharyngalen Ports (*VO*) sowie die horizontale (*HX*) und vertikale (*HY*) Position des Zungenbeins.

Die y-Werte der Punkte sind jeweils durch die (festen) unteren und oberen Schranken des entsprechenden Parameters beschränkt, die mit y_{min} und y_{max} bezeichnet sind. Durch die Punkte R und S wird das Zeitintervall $[t_s, t_s + duration]$ in drei Teile unterteilt: die linke Transitionszone, die Kernzone und die rechte Transitionszone. In der Kernzone ist der Funktionswert des Parameters konstant und in den beiden Transitionszonen wird der

Funktionsverlauf durch jeweils einen kubischen Spline beschrieben. Die vier Stützpunkte der Splines liegen auf den entsprechenden gepunkteten Linien.
Die Koordinaten der Punkte Q, R, S und T werden mit Hilfe der Werte des *Param*-Tupels folgendermaßen berechnet:

$$t_m = t_s + 0.5 \cdot duration \tag{6.5}$$

$$R = \begin{pmatrix} t_m - 0.5 \cdot duration \cdot c_l \\ y_m \end{pmatrix} \tag{6.6}$$

$$S = \begin{pmatrix} t_m + 0.5 \cdot duration \cdot c_r \\ y_m \end{pmatrix} \tag{6.7}$$

$$Q = \begin{pmatrix} t_s + (R_t - t_s) \cdot s_l \\ y_l \end{pmatrix} \tag{6.8}$$

$$T = \begin{pmatrix} S_t + (t_s + duration - S_t) \cdot s_r \\ y_r \end{pmatrix}. \tag{6.9}$$

Der Wertebereich von c_l, c_r, s_l und s_r ist $[0,1]$. Mit c_l und c_r wird die linke und rechte Ausdehnung der Kernzone im Verhältnis zu *duration* angegeben, während mit s_l und s_r die Position der Punkte Q und T innerhalb der Transitionszonen bestimmt werden.
Für die Stützpunkte des kubischen Splines in der linken Transitionszone muss unterschieden werden, ob es sich bei dem aktuellen Laut um den ersten in der *PhoneChain* handelt, oder ob er einen Vorgänger besitzt. Wenn es sich um den ersten Laut in der Kette handelt, dann sind die Stützpunkte des Splines $(t_s, Q_y)^T$, Q, $(Q_t, R_y)^T$ und R. Wenn der aktuelle Laut einen Vorgänger besitzt, dann muss der erste Stützpunkt so abgeändert werden, dass er auf der Linie liegt, die den Punkt T des Vorgängerlauts mit dem Punkt Q des aktuellen Lauts verbindet. Dadurch wird ein glatter Übergang der Parameterfunktionen zwischen benachbarten Lauten sichergestellt. Abb. 6.1 (b) und (c) veranschaulichen die beiden Fälle. Analog zur linken Transitionszone verhält es sich mit dem Spline der rechten Transitionszone. Hier muss für den Stützpunkt bei $t = t_s + duration$ unterschieden werden, ob der aktuelle Laut einen Nachfolger hat oder nicht.
Das eben beschriebene Verfahren wird angewendet, wenn der Wert des boolschen Parameters *isFloating* im *Param*-Tupel auf *FALSE* gesetzt ist. Dadurch sind die Werte der Parameterfunktion innerhalb der Kernzone stets konstant. Das ist aber nicht immer erwünscht, da z.B. der Grundfrequenzverlauf „im Zentrum“ eines stimmhaften Lauts bei einer Äußerung selten konstant ist. Auch koartikulatorische Effekte lassen sich so schwer realisieren.
Das kann umgangen werden, wenn *isFloating* auf *TRUE* gesetzt wird. In diesem Fall werden die y-Werte der Punkte R und S so abgeändert, dass sie auf der Verbindungslinie $\overline{QT}$ liegen. Die Parameterfunktion wird dann in der linken und rechten Transitionszone durch jeweils einen quadratischen Spline beschrieben und in der Kernzone durch die Verbindungslinie der Punkte R und S. Der erste der drei Stützpunkte des linken Splines ergibt sich dann analog zum Fall $isFloating = FALSE$ und die anderen beiden Stützpunkte sind Q und R. Die Stützpunkte des rechten Splines sind S, T und der Punkt bei $t = t_s + duration$.

Abb. 6.1 (a) zeigt die Zeitfunktionen der Parameter *JA* (Kieferöffnungswinkel),

VO (Öffnung des velopharyngalen Ports), *HX* (horizontale Position des Zungenbeins) und *HY* (vertikale Position des Zungenbeins) für das Wort „Oma“. Der Laut [m] ist zur besseren Abgrenzung zu den Vokalen grau hinterlegt. In Abb. 6.2 sind jeweils die Sonagramme zu den gesprochenen und synthetisierten Wörtern „Oma“ und „Jana“ dargestellt.

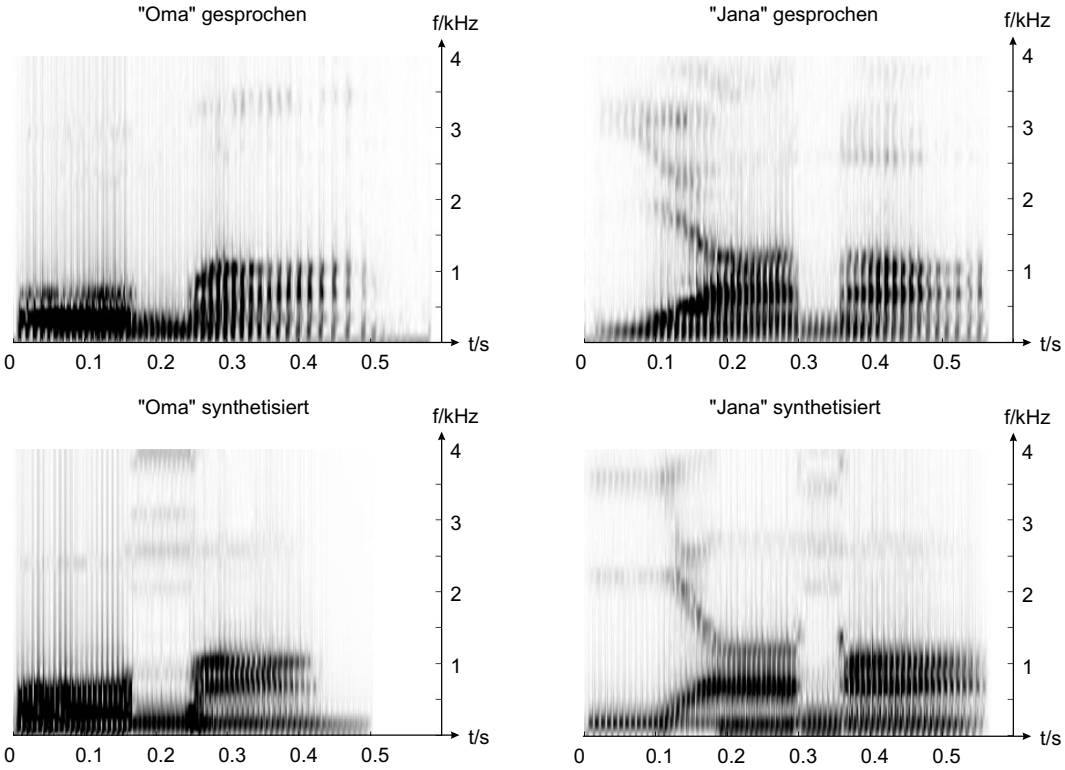

Abbildung 6.2: Vergleich der Sonagramme von zwei jeweils gesprochenen und resynthetisierten Äußerungen.

Man erkennt für die ersten drei Formanten eine gute Übereinstimmung zwischen den gesprochenen und synthetisierten Wörtern. Auch akustisch sind die synthetischen Äußerungen gut zu verstehen. Die Parameterfunktionen wurden für beide Wörter manuell so angepasst, dass sie der Akustik des Originals möglichst nahe kommen.

Kapitel 7

Schlussfolgerungen

7.1 Zusammenfassung der eigenen Beiträge

An dieser Stelle sollen die wichtigsten Beiträge der vorliegenden Arbeit zusammengefasst werden.
Der Kern der Arbeit ist die Präsentation eines neuen, dreidimensionalen Artikulatormodells für die Sprachsynthese. Es ist speziell dafür ausgelegt, in jeder Situation zuverlässig eine genaue Querschnitts- und Umfangsfunktion des Ansatzrohres bereitzustellen. Die Geometrie baut dabei zum Teil auf dem Modell von Mermelstein auf (siehe Abschnitt 3.2).
Besonderer Wert wird auf einen „weichen“ Verlauf der Mittellinie des Ansatzrohres gelegt und darauf, dass sich die Schnittebenen im Bereich des Ansatzrohres nicht überschneiden. Die Schnittebenen werden entlang der Mittellinie so positioniert, dass bei der Schnittflächenberechnung alle Unstetigkeitsstellen in der Vokaltraktkontur erfasst werden. Das sorgt dafür, dass während der Artikulation keine „Sprünge“ in der Querschnittsfunktion auftreten.
Die Positionen der Artikulatoren lassen sich mit mehreren Parametern flexibel festlegen. Die Methoden zur Beschränkung der Parameterwerte auf physiologisch sinnvolle Grenzen werden hergeleitet.
Außerdem wird auch die akustische Synthese vorgestellt, anhand der sich die vom Artikulatormodell produzierten Umfangs- und Querschnittsfunktionen testen lassen. Die Synthese ist eine Kombination mehrerer aus der Literatur bekannter Techniken. Sie basiert auf einem hybriden Verfahren, wie es z.B. in [58] oder [3] vorgeschlagen wird. Die Quelle wird im Zeitbereich mit dem Zwei-Massen-Modell der Stimmlippen von Ishizaka und Flanagan modelliert, während die Übertragungsfunktion im Frequenzbereich berechnet wird. Die (inverse) Fouriertransformation und die Faltung stellen die Verbindung der beiden Ansätze her. Für die mechanische Impedanz der Vokaltraktwände wurden die Ergebnisse aus [33] verwendet.
Die Geometrie des Nasenraums lehnt sich an die Messungen von Dang und Honda [13] an, während das Modell für die Nasennebenhöhlen aus [14] stammt. Zusätzlich habe ich einen eigenen einfachen Mechanismus umgesetzt, der den Nasenraum beim Absenken des Velums ein Stück in Richtung des Rachens verlängert. Inwiefern sich das akustisch auswirkt, bleibt zu untersuchen.

Um das Artikulatormodell zusammen mit der akustischen Synthese auch für zusammenhängende Äußerungen aus mehreren stimmhaften Lauten testen zu können, wird außerdem ein einfaches aber flexibles Verfahren für die zeitliche Steuerung der Vokaltrakt- und Anregungsparameter vorgestellt.

7.2 Diskussion der Ergebnisse

Das Ergebnis der Arbeit ist ein System zur artikulatorischen Synthese vokalischer und vokalähnlicher Äußerungen. Die tragende Rolle dabei spielt das dreidimensionale Artikulatormodell für die Generierung der Querschnitts- und Umfangsfunktionen. Anders als bei den zweidimensionalen Modellen muss hier nicht mehr auf empirische Formeln für die Hinzurechnung der dritten Dimension zurückgegriffen werden. Die mit dem 3D-Modell berechneten Querschnitts- und Umfangsfunktionen einiger Laute wurden präsentiert (siehe Abb. 4.23 und 4.25).
Auch die lateralen Passagen beim Laut /l/ lassen sich auf natürliche Weise modellieren. Tatsächlich ist jedoch gerade beim /l/ die Akustik stark von der dreidimensionalen Wellenausbreitung im Vokaltrakt abhängig. Durch die Berechnung der akustischen Übertragungsfunktion in einer Dimension ist das /l/ also auch hier nur eine Approximation. Dennoch zeigen die Formanttransitionen von und zum /l/ einen realistischen Verlauf.
Für Vokale und Nasale ist die eindimensionale akustische Behandlung völlig ausreichend. Die akustische Ähnlichkeit, die mit dem vorgestellten System zwischen gesprochenen und synthetisierten Äußerungen erreicht werden kann, erkennt man an den Sonagrammen in Abb. 6.2. Informelle Hörtests haben die Verständlichkeit der Syntheseergebnisse bestätigt. Die Natürlichkeit der synthetischen Äußerungen ist stark von den Werten der Anregungsparameter abhängig. In der Regel erreicht sie aber noch nicht ganz die Qualität natürlicher Sprache. Im folgenden Abschnitt werden Vorschläge gemacht, wie die Natürlichkeit weiter gesteigert werden könnte.
Wie bereits erwähnt, ist das Artikulatormodell speziell für die Sprachsynthese entwickelt worden. Dennoch ließe es sich auch als visuelle Unterstützung im Phonetik- oder Fremdsprachenunterricht einsetzen, um die Spracherzeugung besser zu verstehen oder die Artikulation „schwieriger“ Laute zu verdeutlichen.
Das Artikulatormodell ist so einfach aufgebaut, dass damit auf schnellen Computern (ab etwa 2 GHz Taktfrequenz) Sprachsynthese in Echtzeit möglich ist.

7.3 Ausblick

Das Artikulatormodell, die akustische Synthese und die Steuerung der Parameter bieten viele Ansatzpunkte für eine Weiterentwicklung. In diesem Abschnitt werden in loser Reihenfolge mehrere Verbesserungen und Erweiterungen vorgeschlagen.

7.3.1 Artikulatormodell

Für das 3D-Modell wäre es sinnvoll, die Anzahl der Parameter auf ein „vernünftiges“ Maß zu reduzieren. Eine Reduktion von derzeit 18 auf 14 bis 16 Parameter halte ich oh-

ne Einbußen bei der Flexibilität des Modells für machbar. Besonders die vier Parameter *TAX*, *TAY*, *TBX* und *TBY* für die Beschreibung der Splines der Zungenkontur könnten zum Teil direkt von der Position des großen und kleinen Zungenkreises abhängen. Eine Untersuchung anhand von Röntgenaufnahmen des Kopfes für eine große Anzahl von Lauten könnte das klären. Auch der Radius des Zungenkreises könnte von der Position der Zunge abhängen.
Außerdem ließe sich die Geometrie des Modells mit Hilfe von MRI-Daten an beliebige Sprecher anpassen. Das würde eine objektive Bewertung der akustischen Synthese ermöglichen. Durch das Vorliegen der dreidimensionalen Ansatzrohrgeometrie ließen sich mit einer entsprechenden akustischen Simulation auch dreidimensionale Wellenausbreitungsphänomene im Vokaltrakt untersuchen.
Die Schalldämpfung durch das Mitschwingen der Wände hat einen großen Einfluss auf die Resonanzen des Vokaltrakts. Die Berechnungen zur Dämpfung ließen sich präzisieren, wenn man in dem Modell eine Unterscheidung zwischen verschiedenen Gewebetypen vornehmen würde, die das Ansatzrohr begrenzen (Zungengewebe dämpft stärker als der harte Gaumen oder die Zähne). In der Literatur sind jedoch nur wenige Messungen zur mechanischen Impedanz der Vokaltraktwände zu finden.

7.3.2 Akustische Synthese

Dang und Honda [15] haben den akustischen Effekt der Schluckrinne (*piriformis fossa*) untersucht und festgestellt, dass sie die Übertragungsfunktion offener Vokale (z.B. /a:/) merklich beeinflusst. Sie untersuchten außerdem den Einfluss der transvelaren Ankopplung des Nasenraumes an den Rachen- und Mundraum. Während normalerweise die akustische Ankopplung des Nasenraums nur bei geöffnetem velopharyngalen Port beachtet wird, tritt sie tatsächlich auch durch die Vibration des Velums selbst auf. Das hat hauptsächlich einen Einfluss auf den ersten Formanten geschlossener Vokale (z.B. beim /i:/). Sowohl die Schluckrinne als auch die transvelare Ankopplung des Nasenraums lassen sich leicht in das Modell der akustischen Übertragungsleitung einarbeiten.
Sondhi [56] schlägt eine weitere „Korrektur“ des akustischen Übertragungsverhaltens durch das Einfügen zusätzlicher Induktivitäten an den Übergängen zwischen benachbarten Rohrabschnitten vor. Der Wert der Induktivitäten hängt jeweils vom Verhältnis der Querschnittsflächen der benachbarten Rohrabschnitte ab. Diese Maßnahme bewirkt eine Verschiebung einiger Formantfrequenzen um bis zu 15%.
Neben der stimmhaften Schallerzeugung sollte auch die Produktion von Geräuschlauten simuliert werden. Geräuschlaute entstehen durch die Verwirbelung der Luft an starken Verengungen im Ansatzrohr oder an scharfen Kanten, wie z.B. den Zähnen. Damit fallen sie in den Bereich der Strömungsakustik, ein Grenzgebiet zwischen der Strömungsmechanik und der Akustik. Zur ernsthaften Anwendung der Strömungsakustik auf die Erzeugung von Geräuschlauten im Vokaltrakt existieren zur Zeit nur wenige Arbeiten (z.B. [48] und [54]).
Bisher wurde für die Erzeugung der Geräuschlaute ein- und derselbe Ansatz in verschiedenen Variationen gemacht: das Einfügen von Rauschquellen in den entsprechend engen Regionen des Vokaltrakts. Dadurch wird der physikalische Mechanismus der Entstehung von Geräuschlauten zwar nachgeahmt, aber nicht modelliert. Beispiele für dieses Vorgehen findet man in [60] und [58]. Die verschiedenen Verfahren für die Erzeugung der

Geräuschlaute müssten auf ihre Qualität und Anwendbarkeit für das vorgestellte System getestet werden.
Neben den Geräuschlauten könnten auch die unteren Atemwege in die akustische Synthese integriert werden. Vorschläge für die Modellierung des subglottalen Systems findet man z.B. in [34] und [22].
Das in dieser Arbeit vorgestellte Glottismodell von Ishizaka und Flanagan hat die Eigenschaft, nur unter bestimmten physikalischen Bedingungen zu oszillieren. Dadurch gestaltet es sich manchmal als schwierig, die richtigen Werte für die Anregungsparameter zu finden, um eine bestimmte Grundfrequenz zu erzeugen. Um die Grundfrequenz und weitere Eigenschaften der Anregung exakt steuern zu können, würde es sich anbieten, statt auf ein selbstschwingendes Glottismodell auf eine Parametrisierung der Stimmritzenöffnungfläche zurückzugreifen [67, 69]. Die Interaktion zwischen der Anregung und dem Vokaltraktfilter über den supraglottalen Druck würde dabei trotzdem erhalten bleiben.
Ideal wäre ein Modell der Stimmbänder, mit dem man auch leicht einen gewünschten Phonationstyp einstellen kann; z.B. über einen Parameter, der die Phonation kontinuierlich von einer behauchten Stimme, über die modale Stimme zu einer Knarrstimme ändern kann.
Auch die Trills, z.B. das gerollte /R/, sollten in die akustische Synthese einbezogen werden.

7.3.3 Steuerung der Parameter

Sobald alle Laute einer Sprache zufriedenstellend mit Hilfe des Artikulatormodells und der akustischen Synthese produziert werden können, ist die automatische Synthese längerer zusammenhängender Äußerungen denkbar. Die automatische Steuerung der Anregungs- und Vokaltraktparameter ist dabei eine große Herausforderung. Schon jetzt lassen sich für beliebige Laute jeweils alle Vokaltraktparameter abspeichern und mit wenig Aufwand zu zusammenhängenden Äußerungen zusammenfügen.
Denkbar wäre auch ein System zur automatischen Nachahmung beliebiger aufgenommener Wörter oder Sätze mit der Stimme des Artikulatormodells. Erste Ansätze in dieser Richtung existieren bereits [27, 49].

Anhang A

Technische Realisierung

In diesem Abschnitt wird kurz auf die technische Realisierung des Systems eingegangen, welches in dieser Arbeit beschrieben wurde.
Das Programm wurde in der Programmiersprache C++ für das Betriebssystem Windows (ab Windows 98) geschrieben. Für die Sound- und die Grafikausgabe wird das DirectX-SDK eingesetzt, speziell DirectSound (ab Version 5.0) und DirectDraw. Ansonsten werden softwareseitig nur das Windows-API und die Standard-C++ Bibliotheken benutzt.
Hardwareseitig ist für eine „ruckelfreie“ grafische Darstellung ein PC mit 400 MHz ausreichend, und für die Soundausgabe kann jede gängige Soundkarte benutzt werden. Die Soundkarte wird vom Programm sowohl für die Ein- als auch die Ausgabe mit 22050 Hz, 16 Bit (mono) betrieben.
Außerdem nutzt das Programm die von C++ bereitgestellten Techniken zur Umsetzung objektorientierter Konzepte. Die wichtigsten Einheiten des Programms sind als Klassen realisiert. Dazu gehören:

- die Klasse Cords: Hier sind alle Funktionen zur numerischen Behandlung der Stimmlippenschwingung mit dem Zwei-Massen-Modell enthalten.
- die Klasse Tract: Diese Klasse enthält alle Funktionen, um aus einem Satz von Vokaltraktparametern die Querschnitts- und Umfangsfunktion des Vokaltrakts abzuleiten. Dazu gehört die geometrische Konstruktion des Boden-, Decken- und Zungengitters sowie die Berechnung der Mittellinie, der Schnittgeraden und der Querschnittsflächen und Umfangslängen. Außerdem ist hier die Diskretisierung der beiden Funktionen umgesetzt.
- die Klasse Head: Diese Klasse ist speziell für die grafische Ausgabe des in der Klasse Tract berechneten Modells gedacht. Hier werden die typischen Aufgaben einer grafischen Ausgabe-Pipeline erledigt (geometrische Transformation, Clipping, Projektion, ...). Zusätzlich stellt diese Klasse die Methoden für die grafische Manipulation der Vokaltraktparameter am dreidimensionalen Modell bereit.
- die Klasse Speaker: Diese Klasse repräsentiert einen „Sprecher“ und besitzt als Member-Variablen sowohl eine Instanz der Klasse Tract als auch der Klasse Cords. Hier wird zusammen mit dem Modell der Stimmlippen und des Vokaltrakts die akustische Synthese durchgeführt. Dazu wird auch das in Abschnitt 6 beschriebene Steuermodell für die Vokaltrakt- und Anregungsparameter verwendet.

Für das Programm wurde eine umfangreiche grafische Oberfläche entwickelt, die mehrere „Hauptansichten“ für die verschiedenen Programmteile besitzt. Dazu gehören

- eine Oszillogramm-Ansicht, in der aufgenommene oder synthetisierte Schallsignalverläufe betrachtet und analysiert werden können. Zu den Verfahren der Sprachanalyse gehören unter anderem die Spektralanalyse, die Berechnung von LPC-Koeffizienten und Methoden zur automatischen Grundfrequenzbestimmung.

- eine Sonagramm-Ansicht, in der die Sonagramme beliebiger Schallsignale berechnet (in Echtzeit oder nachträglich) und dargestellt werden können.

- eine Vokaltrakt-Ansicht, in der alles dargestellt wird, was mit dem dreidimensionalen Artikulatormodell zu tun hat. Es wird eine grafische Projektion des Modells als Drahtgitter angezeigt und gleichzeitig die Querschnitts- und Umfangsfunktion für die aktuellen Parameter. Die Parameter können intuitiv durch das „Ziehen“ von Kontrollpunkten mit der Maus in der Darstellungsfläche des Modells geändert werden. Dabei sieht man auch gleichzeitig die entsprechende Übertragungsfunktion des Vokaltrakts bzw. das Spektrum des synthetisierten Schallsignals. Außerdem werden hier Funktionen zum Laden und Abspeichern der Vokaltraktparameter sowie die Möglichkeit zum Vergleich des synthetischen Spektrums mit beliebigen anderen Spektren bereitgestellt.

- ein Editor zur Generierung zusammenhängender Äußerungen, in dem per Maus die Parameter und ihre Übergänge für mehrere Laute bearbeitet werden können. Mit diesem Editor wurden auch die Parameterfunktionen für die Äußerungen „Jana“ und „Oma“ (siehe Abschnitt 6) erstellt.

- eine Ansicht zur Visualisierung der Stimmlippenschwingung. Hier werden während der Synthese die Signalverläufe mehrerer Funktionen dargestellt: der subglottale und supraglottale Druck, die Auslenkung der Massen des Stimmlippenmodells, der glottale Volumenstrom und der Schalldruck an der Mundöffung. Außerdem wird eine Animation der schwingenden Massen gezeigt.

Literaturverzeichnis

[1] Alipour-Haghighi, F.; Titze, I.R.: „Elastic models of vocal fold tissues“, *Journal of the Acoustical Society of America*, vol. 90(3), 1326-1331, 1991

[2] Alipour, F.; Fan, C.; Scherer, R.C.: „A numerical simulation of laryngeal flow in a forced-oscillation glottal model“, *Computer Speech and Language*, 10, 75-93, 1996

[3] Allen, D.R.; Strong, W.J.: „A model for the synthesis of natural sounding vowels“, *Journal of the Acoustical Society of America*, vol. 78(1), 58-69, 1985

[4] Baer, T.; Gore, J.C.; Gracco, L.C.; Nye, P.W.: „Analysis of vocal tract shape and dimensions using magnetic resonance imaging: Vowels“, *Journal of the Acoustical Society of America*, vol. 90(2), 799-828, 1991

[5] Beranek, L.L.: *Acoustics*, Woodbury, NY: Acoustical Society of America, 1986

[6] Bertolini, R.; Leutert, G.; Rother, P., Scheuner, G.; Wendler, D.: *Systematische Anatomie des Menschen*, 3. Auflage, VEB Verlag, Volk und Gesundheit Berlin, 1987

[7] Boylestad, R. L.: *Introductory Circuit Analysis*, Seventh Edition, Macmillan Publishing Company, New York, 1994

[8] Childers, D.G.; Hicks, D.M.; Moore, G.P.; Alsaka, Y.A.: „A model for vocal fold vibratory motion, contact area, and the electroglottogram“, *Journal of the Acoustical Society of America*, vol. 80(5), 1309-1320, 1986

[9] Coker, C.H.: „A model of articulatory dynamics and control“, *Proceedings of the IEEE*, vol. 64, no. 4, 452-460, 1976

[10] Cranen, B.; Boves, L.: „Pressure measurements during speech production using semiconductor miniature pressure transducers: Impact on models for speech production“, *Journal of the Acoustical Society of America*, vol. 77(4), 1543-1551, 1985

[11] Cranen, B.; Boves, L.: „On subglottal formant analysis“, *Journal of the Acoustical Society of America*, vol. 81(3), 734-746, 1987

[12] Cranen, B.; Schroeter, J.: „Modeling a leaky glottis“, *Journal of Phonetics*, 23, 165-177, 1995

[13] Dang, J.; Honda, K.: „Morphological and acoustical analysis of the nasal and the paranasal cavities“, *Journal of the Acoustical Society of America*, vol. 96(4), 2088-2100, 1994

[14] Dang, J.; Honda, K.: „Acoustic characteristics of the human paranasal sinuses derived from transmission characteristic measurement and morphological observation“, *Journal of the Acoustical Society of America*, vol. 100(5), 3374-3383, 1996

[15] Dang, J.; Honda, K.: „An improved vocal tract model of vowel production implementing piriform fossa resonance and transvelar nasal coupling“, *ICSLP-96* (Philadelphia, USA), 965-968, 1996

[16] Dang, J.; Honda, K.: „Acoustic characteristics of the piriform fossa in models and humans“, *Journal of the Acoustical Society of America*, vol. 101(1), 456-465, 1997

[17] Dunn, H. K.: „The calculation of vowel resonances, and an electrical vocal tract“, *Journal of the Acoustical Society of America*, vol. 22, number 6, 1950

[18] Engwall, O.: „Vocal tract modeling in 3D“, *TMH-QPSR*, 1-8, 1/1999

[19] Engwall, O.: „Replicating three-dimensional tongue shapes synthetically“, *TMH-QPSR*, 1-12, 2-3/2000

[20] Engwall, O.: „Synthesizing static vowels and dynamic sounds using a 3D vocal tract model“, *TMH-QPSR*

[21] Fant, G.: *Acoustic Theory of Speech Production*, Mouton, The Hague, 1960

[22] Fant, G.; Ishizaka, K.; Lindqvist, J.; Sundberg, J.: „Subglottal formants“, *STL-QPSR*, 1-12, 1/1972

[23] Fant, G.; Nord, L.; Branderud, P.: „A note on the vocal tract wall impedance“, *STL-QPSR*, 13-20, 4/1976

[24] Farley, G.R.: „A quantitative model of F_0 control“, *Journal of the Acoustical Society of America*, vol. 95(2), 1017-1029, 1994

[25] Farley, G.R.: „Control of voice F_0 by an artificial neural network“, *Journal of the Acoustical Society of America*, vol. 96(3), 1374-1379, 1994

[26] Flanagan, J.L.: *Speech analysis, synthesis and perception*, Springer-Verlag Berlin, 1965

[27] Flanagan, J. L.; Ishizaka, K.; Shipley, K. L.: „Synthesis of speech from a dynamic model of the vocal cords and vocal tract“, *The Bell System Technical Journal*, vol. 54, no. 3, 1975

[28] Fanagan, J.L.; Ishizaka, K.; Shipley, K.L.: „Signal models for low bit-rate coding of speech“, *Journal of the Acoustical Society of America*, vol. 68(3), 780-791, 1980

[29] Fujimura, O.: „Sweep-tone measurements of vocal-tract characteristics“, *Journal of the Acoustical Society of America*, vol. 49, number 2 (part 2), 541-557, 1971

[30] Greisbach, R.; Philipp, J.: „Die Querschnittsfläche des Sprechtraktes im Bereich des harten Gaumens als Funktion des midsaggitalen Durchmessers“, *Institut für Phonetik der Universität zu Köln*, Berichte Nr. 10, S. 33-40, 1980

[31] Hanson, H.M.; Stevens, K.N.; Kuo, H.K.J.; Chen, M.Y.; Slifka, J.: „Towards models of phonation“, *Journal of Phonetics*, 29, 451-480, 2001

[32] Ishizaka, K.; Flanagan, J.L.: „Synthesis of voiced sounds from a two-mass model of the vocal cords“, *The Bell System Technical Journal*, vol. 51, no. 6, 1233-1268, 1972

[33] Ishizaka, K.; French, J.C.; Flanagan, J.L.: „Direct determination of vocal tract wall impedance“, *IEEE Transactions on Acoustics, Speech and Signal Processing*, vol. ASSP-23, no. 4, 370-373, 1975

[34] Ishizaka, K.; Matsudaira, M.; Kaneko, T.: „Input acoustic-impedance measurement of the subglottal system“, *Journal of the Acoustical Society of America*, vol. 60, No. 1, 190-197, 1976

[35] Kelly, J.L.; Lochbaum, C.C.: „Speech synthesis“, *Proc. Fourth Int. Congr. Acoust.*, Paper G42, 1-4, Kopenhagen, 1962

[36] Klatt, D.: „Software for a cascade/parallel formant synthesizer“, *Journal of the Acoustical Society of America*, vol. 67(3), 971-994, 1980

[37] Kröger, B.: *Ein phonetisches Modell der Sprachproduktion*, Niemeyer Verlag, Tübingen, 1998

[38] Lindblom, B.; Sundberg, J.: „Acoustical consequences of lip, tongue, jaw, and larynx movement“, *Journal of the Acoustical Society of America*, vol. 50, 1166-1179, 1971

[39] Lindqvist-Gauffin, J.; Sundberg, J.: „Acoustic properties of the nasal tract“, *Phonetica*, vol. 33, 161-168, 1976

[40] Maeda, S.: „The role of the sinus cavities in the production of nasal vowels“, *IEEE International Conference of Acoustics, Speech, Signal Processing*, vol. 2, 911-914, 1982

[41] Mathews, M.V.; Miller, J.E.; David Jr., E.E.: „Pitch synchronous analysis of voiced sounds“, *Journal of the Acoustical Society of America*, vol. 33, number 2, 179-186, 1961

[42] Mermelstein, P.: „Articulatory model for the study of speech production“, *Journal of the Acoustical Society of America*, vol. 53, number 4, 1070-1082, 1973

[43] Meyer, P.; Wilhelms, R.; Strube, H.: „A quasiarticulatory speech synthesizer for German language running in real time“, *Journal of the Acoustical Society of America*, vol. 86(2), 523-540, 1989

[44] Miller, R.L.: „Nature of the vocal cord wave“, *Journal of the Acoustical Society of America*, vol. 31, number 6, 667-677, 1959

[45] Morse, P.; Ingard, K.U.: *Theoretical Acoustics*, Princeton University Press, 1987

[46] Narayanan, S.S.; Alwan, A.A.; Haker, K.: „An articulatory study of fricative consonants using magnetic resonance imaging“, *Journal of the Acoustical Society of America*, vol. 98(3), 1325-1347, 1995

[47] Narayanan, S.S.; Alwan, A.A.; Haker, K.: „Towards articulatory-acoustic models for liquid approximants based on MRI and EPG data. Part I. The laterals“, *Journal of the Acoustical Society of America*, vol. 101(2), 1064-1089, 1997

[48] Narayanan, S.; Alwan, A.: „Noise source models for fricative consonants“, *IEEE Transactions on Speech and Audio Processing*, vol. 8, no. 2, 328-344, 2000

[49] Rahim, M.G.; Goodyear, C.C.; Kleijn, W.B.; Schroeter, J.; Sondhi, M.M.: „On the use of neural networks in articulatory speech synthesis“, *Journal of the Acoustical Society of America*, vol. 93(2), 1109-1121, 1993

[50] Rosen, G.: „Dynamic analog speech synthesizer“, *Journal of the Acoustical Society of America*, vol. 30, number 3, 201-209, 1958

[51] Rosenberg, A.E.: „Effect of glottal pulse shape on the quality of natural vowels“, *Journal of the Acoustical Society of America*, vol. 49, number 2 (part 2), 583-590, 1971

[52] Rubin, P.; Baer, T.; Mermelstein, P.: „An articulatory synthesizer for perceptual research“, *Journal of the Acoustical Society of America*, vol. 70(2), 321-328, 1981

[53] Schroeter, J.; Cranen, B.: „Physiologically-motivated modeling of the voice source in articulatory analysis/synthesis“, *Proceedings of the 3rd European Conference on Speech Communication and Technology (Eurospeech 1993)*, vol. 1, 95-98, Berlin, 1993

[54] Sinder, D.J.: „Speech synthesis using an aeroacoustic fricative model“, *Dissertation at the Graduate School - New Brunswick Rutgers, the State University of New Jersey*, 1999

[55] Smith, S.: *The Scientist and Engineer's Guide to Digital Signal Processing*, Second Edition, California Technical Publishing, 1999

[56] Sondhi, M.M.: „An improved vocal tract model“, *Proceedings of the 11th ICA*, vol. 4, 167-170, Paris, France, 1983

[57] Sondhi, M.M.: „Resonances of a bent vocal tract“, *Journal of the Acoustical Society of America*, vol. 79(4), 1113-1116, 1986

[58] Sondhi, M.M.; Schroeter, J.: „A hybrid time-frequency domain articulatory speech synthesizer“, *IEEE Transactions on Acoustics, Speech and Signal Processing*, vol. ASSP-35, no. 7, 955-967, 1987

[59] Stevens, K. N.; Kasowski, S.; Fant, G.: „An electrical analog of the vocal tract“, *Journal of the Acoustical Society of America*, vol. 25, number 4, 734-742, 1953

[60] Stevens, K.N.: „Airflow and turbulence noise for fricative and stop consonants: Static considerations“, *Journal of the Acoustical Society of America*, vol. 50, number 4 (part 2), 1180-1191, 1971

[61] Stone, M.: „A three-dimensional model of tongue movement based on ultrasound and x-ray microbeam data“, *Journal of the Acoustical Society of America*, vol. 87(5), 2207-2217, 1990

[62] Stone, M.; Lundberg, A.: „Three-dimensional tongue surface shapes of English consonants and vowels“, *Journal of the Acoustical Society of America*, vol. 99 (6), 3728-3737, 1996

[63] Stone, M.; Davis, E.P.; Douglas, A.S.; Aiver, M.N.; Gullapalli, R.; Levine, W.S.; Lundberg, A.J.: „Modeling tongue surface contours from Cine-MRI Images“, *Journal of Speech, Language, and Hearing Research*, vol. 44, 1026-1040, 2001

[64] Story, B.H.; Titze, I.R.: „Voice simulation with a body-cover model of the vocal folds“, *Journal of the Acoustical Society of America*, vol. 97(2), 1249-1260, 1995

[65] Titze, I.R.: „The human vocal cords: A mathematical model, part I“, *Phonetica 28*, 129-170, 1973

[66] Titze, I.R.: „The human vocal cords: A mathematical model, part II“, *Phonetica 29*, 1-21, 1974

[67] Titze, I.R.: „Parameterization of the glottal area, glottal flow, and vocal fold contact area“, *Journal of the Acoustical Society of America*, vol. 75(2), 570-580, 1984

[68] Titze, I.R.: „The physics of small-amplitude oscillation of the vocal folds“, *Journal of the Acoustical Society of America*, vol. 83(4), 1536-1552, 1988

[69] Titze, I.R.: „A four-parameter model of the glottis and vocal fold contact area“, *Speech Communication*, 8, 191-201, 1989

[70] Umeda, N.; Matsui, E.; Suzuki, T., Omura, H.: „Synthesis of fairy tales using an analog vocal tract“, *Proceedings of 6th International Congress on Acoustics*, B159-162, Tokyo, Japan, 1968

[71] Ungeheuer, G.: *Elemente einer akustischen Theorie der Vokalartikulation*, Springer-Verlag, Berlin, 1962

[72] Van Den Berg, Jw.; Zantema, J.T.; Doornenbal Jr., P.: „On the air resistance and the Bernoulli effect of the human larynx“, *Journal of the Acoustical Society of America*, vol. 29, number 5, 626-631, 1957

[73] Vary, P.; Heute, U.; Hess, W.: *Digitale Sprachsignalverarbeitung*, B. G. Teubner Stuttgart, 1998

[74] Veldhuis, R.: „A computationally efficient alternative for the Liljencrants-Fant model and its perceptual evaluation“, *Journal of the Acoustical Society of America*, vol. 103(1), 566-571, 1998

[75] Wakita, H.; Fant, G.: „Toward a better vocal tract model“, *STL-QPSR*, 9-29, 1/1978

[76] Wilhelms-Tricarico, R.: „Physiological modeling of speech production: Methods for modeling soft-tissue articulators“, *Journal of the Acoustical Society of America*, vol. 97(5), 3085-3098, 1995

Selbstständigkeitserklärung

Hiermit erkläre ich, Peter Birkholz, dass ich die vorliegende Arbeit selbstständig und nur unter Verwendung der angegebenen Literatur und Hilfsmittel angefertigt habe.

Rostock, den Peter Birkholz

Thesen

1. Die Sprachsynthese mit dem Computer hat in den letzten Jahren große Fortschritte gemacht. Dennoch wird die Natürlichkeit der menschlichen Sprache noch nicht vollkommen erreicht.

2. Durch den stetigen Anstieg der Rechenleistung und Speicherkapazität der Computer erscheint die artikulatorische Sprachsynthese zunehmend als vielversprechende Alternative zu den verbreiteten konkatenativen Sprachsyntheseverfahren.

3. Die Artikulation der Sprache durch den Vokaltrakt ist ein sehr komplexer Prozess. Dieser Prozess lässt sich ausgehend von einem Modell des Vokaltrakts in natürlicher Weise nachbilden. Der Erfolg eines solchen Artikulatormodells hängt davon ab, wie genau und wie einfach sich die Geometrien des Vokaltrakts für die verschiedenen Laute erzeugen lassen.

4. Viele Artikulatormodelle beschränken sich auf die Modellierung der Geometrie in einer oder zwei Dimensionen. Dabei geht zwangsläufig Information verloren. Das in dieser Arbeit entwickelte Modell zeigt, dass ein dreidimensionales Modell – sowohl von der Geometrie als auch von der Steuerung her – nicht viel komplizierter sein muss als ein herkömmliches Modell.

5. Die Sprachsynthese mit einem dreidimensionalen Artikulatormodell ist auf heutigen Computern in Echtzeit möglich. Außerdem kann die Querschnitts- und Umfangsfunktion ohne zusätzliche Informationen direkt aus dem Modell gewonnen werden.

6. Die Integration von Zeitbereichs- und Frequenzbereichstechniken für die akustische Synthese kann sehr natürlich klingende Sprache produzieren.

7. Die Erzeugung von Zisch- und Plosivlauten sowie die Steuerung der Sprechwerkzeuge sind in der artikulatorischen Sprachsynthese heute noch unvollständig gelöste Probleme.