Rohrmodelle des Sprechtraktes

Analyse, Parameterschätzung und Syntheseexperimente

Dissertation
zur Erlangung des Doktorgrades
der Naturwissenschaften

vorgelegt beim Fachbereich Physik
der Johann Wolfgang Goethe-Universität
in Frankfurt am Main

von
Karl Schnell
aus
Frankfurt am Main

Frankfurt am Main 2003
(DF1)

vom Fachbereich Physik der Johann Wolfgang Goethe-Universität
als Dissertation angenommen.

Dekan: Prof. Dr. H. Schmidt-Böcking

Gutachter: Prof. Dr.-Ing. A. Lacroix, Prof. Dr. H. Reininger

Datum der Disputation: 10.9.2003

Inhaltsverzeichnis

Kapitel 1

Einleitung

Die in der Natur auftretenden Prozesse weisen oft eine hohe Komplexität auf, so daß zu deren mathematischen Beschreibung meistens eine vereinfachte Modellbildung zweckmäßig ist. Dies trifft auch für die menschliche Sprachproduktion zu, welche in mehrere Teilprozesse unterteilt werden kann. Eine möglichst exakte Beschreibung dieser einzelnen Prozesse erfordert Modelle, die in der Regel viele Parameter aufweisen, welche in einer komplizierten Beziehung zu den Merkmalen der Sprachsignale stehen. Daher kann aus den Modellparametern zwar auf die Merkmale geschlossen werden, aber meistens nicht umgekehrt. Dies ist allerdings für viele Aufgabenstellungen in der Sprachverarbeitung erforderlich.

Die Sprachanalyse und -erzeugung mittels Sprachproduktionsmodellen wird durchgeführt, indem die Sprachspektren bzw. deren spektrale Einhüllende möglichst gut durch die Modellbetragsgänge approximiert werden. Um dies zu erreichen, werden die Modellparameter aus dem Sprachsignal geschätzt. Für diesen Zweck ist eine möglichst einfache mathematische Modellbeschreibung vorteilhaft, welche mit einer Komplexitätsreduzierung der Modelle einhergeht. Dafür wird das gesamte Sprachproduktionsmodell auf die wesentlichen Produktionsmechanismen beschränkt. Als die grundlegenden Bestandteile der Sprachproduktion werden das akustische Höhlungssystem des Sprechtraktes, das sich von der Glottis bis zu den Lippen und/oder den Nasenlöchern erstreckt, und dessen Anregung angesehen. Die Sprachlaute entstehen, indem die Anregung durch den Sprechtrakt gefiltert und von den Öffnungen abgestrahlt wird. Das Höhlungssystem des Sprechtraktes kann in erster Näherung als lineares System angesehen werden, das infolge der Artikulation auch zeitvariabel ist. Wird zusätzlich die Anregung unabhängig vom Sprechtrakt angenommen, entsteht das bekannte Quelle-Filter Modell. Die theoretischen Grundlagen für diese Modelle ergeben sich formal aus der Systemtheorie. Da die Systemidentifikation von allgemeinen linearen Systemen nur in Spezialfällen optimal gelöst werden kann, werden für die anwendungsorientierten Sprechtraktmodelle in der Regel einfache Nur-Pole-Modelle verwendet, die aus der linearen Prädiktion resultieren. Diese Systeme können den Sprechtrakt allerdings nur zum Teil modellieren, auch wenn sie als Kreuzgliedketten-Filter dargestellt werden. Für genauere Sprechtraktmodelle existieren nur teilweise adäquate Schätzalgorithmen. Die vorliegende Arbeit versucht deshalb den Mangel an Schätzalgorithmen für erweiterte Modelle des Sprechtraktes zu beheben, die über das übliche Nur–Pole-Modell hinausgehen und eine adäquate Schätzung aus dem Sprachsignal erzielen.

Im Folgenden wird eine Gliederung der Arbeit gegeben:

Im zweiten Kapitel „Menschliche Spracherzeugung" werden wesentliche Teilprozesse

der Sprachproduktion erläutert. Die Betrachtung bleibt dafür im Wesentlichen auf die Lauterzeugung beschränkt und behandelt den Aufbau des Sprechtraktes und dessen Anregungsarten.

Im dritten Kapitel „Rohrmodelle" werden Modelle der Sprachproduktion behandelt. Als grundlegender Modellansatz des Sprechtraktes wird das zeitdiskrete Rohrmodell verwendet, welches die eindimensionale Ausbreitung ebener Wellen beschreibt, die durch Impedanzsprünge infolge von Rohrquerschnittssprüngen und Rohrverzweigungen gestört wird. Dafür werden zuerst die akustischen Grundlagen des Rohrmodells mit seinen mathematischen Vereinfachungen behandelt, wonach darauf aufbauend die grundlegenden Elemente des Rohrmodells beschrieben werden, welche aus Adaptoren, Rohrelementen und Rohrabschlüssen bestehen. Die Berücksichtigung des Nasaltraktes erfordert Verzweigungen, die durch Dreitor-Adaptoren beschrieben werden. Die Öffnungen des Sprechtraktes an den Lippen, Nasenlöchern und an der Glottis lassen sich durch Rohrabschlüsse beschreiben. Durch Verbinden dieser Elemente können zahlreiche unterschiedliche Rohrstrukturen gebildet werden. Ringstrukturen werden ebenfalls diskutiert, welche eine Verzweigung mit einer dahinterliegenden Kopplung darstellen. Neben Modellen mit einem Systemeingang und -ausgang kommen auch Rohrmodelle mit zwei Systemausgängen in Betracht. Diese werden für die Analyse von nasalierten Vokalen verwendet, um die Schallabstrahlung der Lippen und Nasenlöcher separat zu modellieren. Der Nasaltrakt selbst besitzt mit seinen beiden Nasenlöchern zwei Öffnungen für die Schallabstrahlung, wodurch auch hierfür Modelle mit zwei Systemausgängen Verwendung finden. Um diese vielfältigen Rohrstrukturen im Zeit- und im Bildbereich, im Folgenden mit Z-Bereich bezeichnet, einheitlich behandeln zu können, wird eine Vorgehensweise diskutiert, mit Hilfe derer die Übertragungsfunktion und das Zeitverhalten einer nahezu beliebigen Rohrstruktur algorithmisch berechnet werden kann. Dafür analysiert der vorgestellte Algorithmus die zuerst unbekannte Rohrstruktur, welche als Graph betrachtet wird. Das Sprachproduktionsmodell wird für kurze Zeitabschnitte als stationär angesehen, da sich die Artikulatoren in den betrachteten Zeitabschnitten kaum verändern. Eine Ausnahme bildet die Einbeziehung der stimmhaften Anregung in das Vokaltraktmodell. Der Rohrabschluß an der Glottis wird daher wegen der schnellen Stimmbandschwingungen bei Phonation auch zeitvariabel modelliert.

Im vierten Kapitel „Analyse" werden Schätzalgorithmen für die Bestimmung der Modellparameter aus dem Sprachsignal vorgestellt. Nur für den einfachsten Fall eines unverzweigten Rohres mit einem Reflexionskoeffizienten ± 1 am Rohrabschluß stehen Standardalgorithmen für die optimale Parameterbestimmung zur Verfügung, da bei diesem System die Struktur eines Prädiktions-Modells vorliegt. Für die lineare Prädiktion wird der Spezialfall für periodische Signale [Sn96] mittels einer geometrischen Betrachtung hergeleitet, welche auch als inverse Filterung eines unverzweigten Rohrmodells interpretiert werden kann, analog zur Burg-Methode. Dieser Spezialfall findet für eine periodensynchrone Analyse von stimmhaften Sprachsignalen Verwendung. Die Parameterbestimmung von erweiterten Rohrmodellen wird durch Minimierung eines spektralen Abstandsmaßes zwischen dem Rohrmodell und dem Sprachsignal vollzogen. Hierfür existieren schon Ansätze, in denen allerdings ein Fehlermaß verwendet wird, das nicht direkt die Systemfunktion des Modells verwendet. So wird z.B. in [Rah93, Schr94] ein cepstrales Abstandsmaß und in [Fr95a, Fr95b] ein Vergleich von Polstellen verwendet. Dies hat den Nachteil, daß durch die Transformation der Sy-

stemfunktion ein Informationsverlust entsteht; ein Fehlermaß, welches ausschließlich Polstellen verwendet, kann Pol-Nullstellen-Systeme nicht vollständig beschreiben. Ein grundlegender Gegenstand dieser Arbeit ist die Entwicklung von Fehlermaßen, in denen die Übertragungsfunktion vollständig berücksichtigt wird. Dadurch läßt sich die Parameterbestimmung auch systemtheoretisch konsistent darstellen, da die Pole und Nullstellen des Modells vollständig berücksichtigt werden. Als Grundlage des Fehlermaßes wird die inverse Filterung herangezogen. Es zeigt sich, daß für eine adäquate Schätzung das Fehlermaß der inversen Filterung nicht unmittelbar auf erweiterte Rohrmodelle erfolgreich angewandt werden kann. Deshalb wird eine modifizierte Fehlerdefinition der inversen Filterung vorgestellt, durch deren Minimierung eine adäquate Schätzung erreicht werden kann. Die Modifikation der Definition besteht aus einem Korrekturfaktor, der von den konstanten Termen des Zähler- und Nennerpolynoms der Übertragungsfunktion abhängig ist. Die Minimierung des Fehlers wird durch ein gradientenbasiertes Optimierungsverfahren mit adaptiver Schrittweite vollzogen. Bei der Parameterschätzung wird zusätzlich zu einer guten Modellierung des Sprachspektrums auch ein realistischer Verlauf der geschätzten Flächen angestrebt. Erst durch einen realistischen Flächenverlauf ergibt sich ein zutreffendes akustisches Modell; vorteilhaft ist hierbei der Bezug der Modellparameter zur Artikulation. Neben Analysen von Modellen mit einem Systemausgang werden auch Schätzung von verzweigten Rohrmodellen mit zwei Systemausgängen behandelt, welche für die Analyse nasalierter Sprachlaute und für eine separate Analyse des Nasaltraktes verwendet werden. Für die Analyse nasalierter Vokale mit einem verzweigten Rohrmodell, das die Nasenlöcher- und Lippenabstrahlung jeweils mit einem eigenen Systemausgang modellieren, werden getrennt aufgenommene Mund- und Nasensignale verwendet. Die Separation der Mund- und Nasensignale wird durch eine Dämmplatte erreicht, die den Kopf des Sprechers umschließt und zwei Räume trennt.

Neben diesen zeitinvarianten Modellen werden auch zeitvariable Rohrmodelle analysiert. Die Zeitvariabilität entsteht durch die Verwendung eines zeitveränderlichen Glottisabschlusses. Für die Parameterbestimmung zeitvariabler Modelle muß die Fehlerdefinition angepaßt werden. Da für die Berechnung des Fehlermaßes die konstanten Terme des Zähler- und Nennerpolynoms der Übertragungsfunktion benötigt werden und die Übertragungsfunktion im Z-Bereich infolge der Zeitvariabilität nicht zur Verfügung steht, wird der Ansatz verfolgt, die benötigten Terme aus dem Ausgangssignal des Rohrmodells zu ermitteln. Durch eine solch angepaßte Darstellung der Fehlerdefinition sollen auch Schätzungen von zeitvariablen Rohrmodellen ermöglicht werden.

Wird der Fehler mittels universell anwendbarer Optimierungsverfahren minimiert, so resultieren daraus ein verhältnismäßig hoher Rechenaufwand und mögliche Probleme mit lokalen Minima. Daher werden zusätzlich Verfahren vorgestellt, die den Rechenaufwand der Schätzung verringern. Dies wird durch Verwendung von speziellen Operationen erreicht, welche suboptimale Lösungen erzielen. Suboptimale Lösungen stellen eine optimale Schätzung für eine Untergruppe von Modellparametern dar, während die übrigen Parameter als bekannt vorausgesetzt werden. Hierbei liegt der Vorteil darin, daß die suboptimalen Lösungen durch geschlossene Formeln ermittelt werden können, wodurch quasi eine analytische Lösung gegeben ist. Für die Erzielung einer Gesamtlösung unter Berücksichtigung aller Parameter wird ein Satz von Operationen iterativ angewendet. Diese Verfahren können allerdings nicht auf alle erweiterten Rohrmodelle sinnvoll angewandt werden und müssen speziell auf die Struktur des Rohrmodells

angepaßt werden. Dieser Ansatz wird zuerst für eine ARMA-Schätzung verwendet, in der Pole und Nullstellen abwechselnd geschätzt werden, womit für stimmhafte Sprachsignalperioden sehr gute Resultate erzielt werden. Die Analysen von Testsignalen ergeben perfekte Ergebnisse. Der Optimierungsansatz mittels suboptimaler Lösungen kann auch auf unverzweigte Rohre mit einem oder zwei vorgegebenen Abschlüssen angewendet werden. Rohrmodelle mit einer Verzweigung können mit diesem Ansatz ebenfalls analysiert werden, wofür zuerst die Koeffizienten des Seitenzweiges geschätzt werden, welche die Nullstellen des Systems repräsentieren und anschließend die restlichen Koeffizienten bestimmt werden. Nasale und nasalierte Vokale sowie deren getrennte Mund- und Nasensignale werden damit untersucht. Die Sprachspektren lassen sich hiermit gut durch die geschätzten Betragsgänge der verzweigten Modelle approximieren. Im Vergleich zu [Lim96] und [Liu96], in denen auch Schätzalgorithmen für verzweigte Rohrmodelle vorgestellt werden, erzielen die in dieser Arbeit entwickelten Algorithmen auch für Sprachsignale gute Ergebnisse, und nicht nur für Testsignale. Im Gegensatz zu [Liu96] werden darüber hinaus sämtliche Flächen des Rohrmodells geschätzt.

Im fünften Kapitel werden die geschätzten Modellparameter für die Spracherzeugung verwendet. Mit Hilfe der geschätzten Flächen können bei entsprechender Anregung des Rohrmodells stationäre Laute erzeugt werden. Durch Interpolation von Sätzen von Vokaltraktflächen können Lautübergänge modelliert werden. Diesbezüglich werden Vokal-Konsonant-Vokal Übergänge behandelt, in denen die Konsonanten durch Konstriktionen in den geschätzten Vokaltraktflächen gekennzeichnet sind. Neben der Generierung von Lautübergängen und Lautketten wird eine Lauttransformation vorgestellt, in der unnasalierte Vokale nachträglich eine künstliche Nasalierung erhalten. Dafür wird das analysierte unverzweigte Rohrmodell des unnasalierten Lautes durch die verzweigten Rohrmodelle aus dem Mund- und Nasensignal des nasalierten Lautes in der Weise erweitert, daß eine künstliche Nasalierung herbeigeführt wird.

Die Qualität des synthetisierten Sprachsignals wird neben dem Filtermodell auch von der Anregung beeinflußt. Bei der stimmhaften Anregung können durch Verwendung von gleichförmigen Anregungsperioden, wie sie z.B. bei Impulsfolgen vorkommen, nicht die Fluktuationen eines natürlichen Sprachsignals modelliert werden. Um diese Unregelmäßigkeiten in der Anregung zu berücksichtigen, wird ein Satz von benachbarten und modifizierten Residualperioden, welche aus dem Restsignal der inversen Filterung stammen, für die Anregung aller stimmhafter Laute verwendet. Resynthetisierte stimmhafte Sprachsignale zeigen auf, daß dadurch die Natürlichkeit der erzeugten Sprachsignale erhöht wird.

Kapitel 2

Menschliche Spracherzeugung

2.1 Sprechtrakt

Die menschliche Spracherzeugung wird durch akustische Anregung eines zeitvariablen Hohlraumsystems ermöglicht. Die maßgeblichen Hohlräume sind dabei der Rachen und der Mundraum, welche den Vokaltrakt darstellen, und im Falle des abgesenkten Velums der angekoppelte Nasaltrakt (Bild 2.1). Diese Hohlräume bilden den Sprechtrakt und erstrecken sich von den Stimmbändern bis zu den Lippen bzw. den Nasenlöchern.An

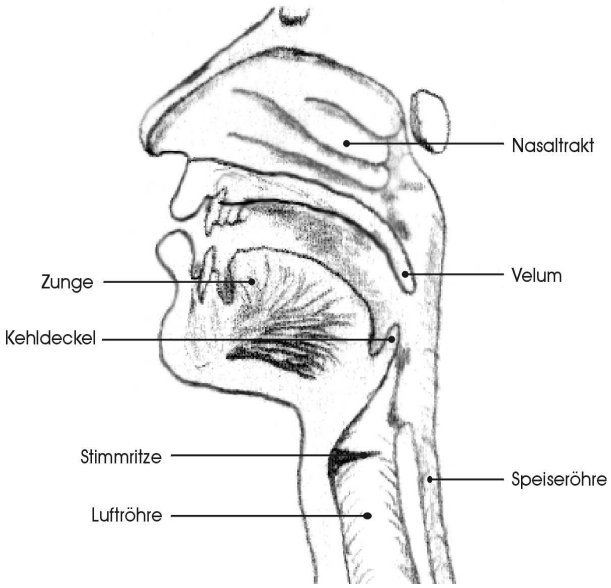


Bild 2.1: Sagittalschnitt des Sprechtrakts.

der Lauterzeugung sind noch weitere Resonanzräume beteiligt, da bei geöffneter Glottis die Lunge mit dem Sprechtrakt durch die Luftröhre verbunden ist. Dieses Hohlraumsystem unterhalb der Stimmbänder stellt den subglottalen Trakt dar. Die Glottis bzw. die Stimmritze stellt die Öffnungsfläche zwischen den Stimmbändern dar. Für eine Lauterzeugung wird der Sprechtrakt angeregt, wobei dafür die Energie aus einem erhöhten Druck in den Lungen resultiert, welcher durch Muskelkraft erzeugt und aufrechterhalten wird. Durch den erhöhten subglottalen Druck resultiert ein Teilchenfluß durch den Sprechtrakt, der an den Nasenlöchern bzw. der Mundöffnung entweicht. Der Luftstrom

kann das System der Stimmbänder zu Schwingungen anregen, wodurch die stimmhafte Anregung resultiert. Neben der stimmhaften Anregung der Stimmbänder existieren auch stimmlose Anregungsmechanismen. Durch eine Verengung des Vokaltrakts kann infolge eines schnellen Teilchenstroms eine turbulente Strömung resultieren, wodurch eine rauschhafte stimmlose Anregung entsteht. Eine weitere Anregungsart kann durch Lösen eines totalen Verschlusses im Vokaltrakt erzeugt werden, wodurch der aufgestaute Druck im Vokaltrakt als impulsartiger Schall sehr schnell abgebaut wird. Im Folgenden werden die wichtigsten Anregungsarten erläutert.

Stimmhafte Anregung

Die stimmhafte Anregung (Phonation) wird durch Schwingen der Stimmbänder bzw. Stimmlippen im Kehlkopf (Bild 2.2) erzeugt. Bei Phonation sind die Stimmbandmuskeln angespannt, so daß eine Rückstellkraft vorherrscht, welche die Stimmlippen zusammendrückt. Gleichzeitig wird durch Anspannen des Zwergfells und der Brustmuskeln ein Druck auf die Lunge ausgeübt. Dieser wird als subglottaler Druck über die Luftröhre bis an die geschlossenen Stimmbänder weitergeleitet. Infolge des subglottalen Druckes wandert die geschlossene Verbindungsstelle der beiden Stimmlippen nach oben, bis das Stimmlippenpaar auseinander gedrückt wird und die Glottis somit geöffnet wird, wie in Bild 2.3 zu sehen ist.Durch die offene Glottis kann Luft aus dem

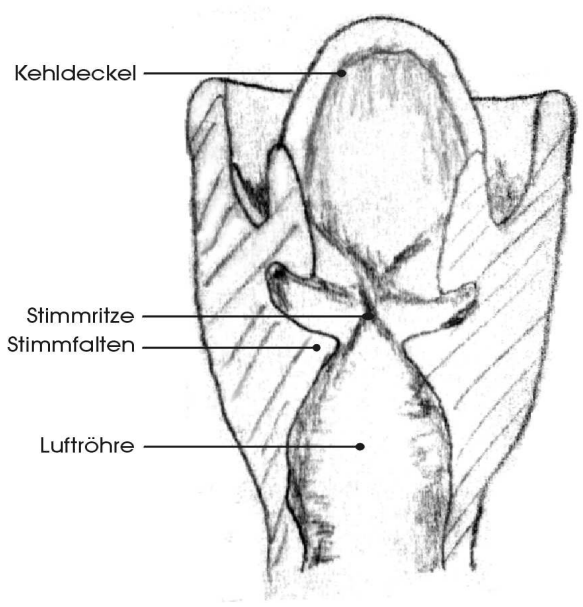


Bild 2.2: Kehlkopf mit Glottis.

subglottalen Bereich in den Rachenraum strömen. Da die Glottisöffnungsfläche klein ist, sind die resultierenden Teilchengeschwindigkeiten in der Glottis relativ hoch. Die maximale Öffnungsfläche der Glottis eines Mannes kann etwas über 20 mm^2 betragen während der Phonation. Die Länge der Glottis liegt etwas über 2 cm, wobei die Form oval- bis dreieckförmig ist. Durch die erhöhte kinetische Energie der Teilchen in der Glottis fällt der statische Druck ab. Dies folgt aus der Energieerhaltung nach dem Gesetz von Bernoulli. Dieser Druckabfall bewirkt, daß die Stimmlippen nicht

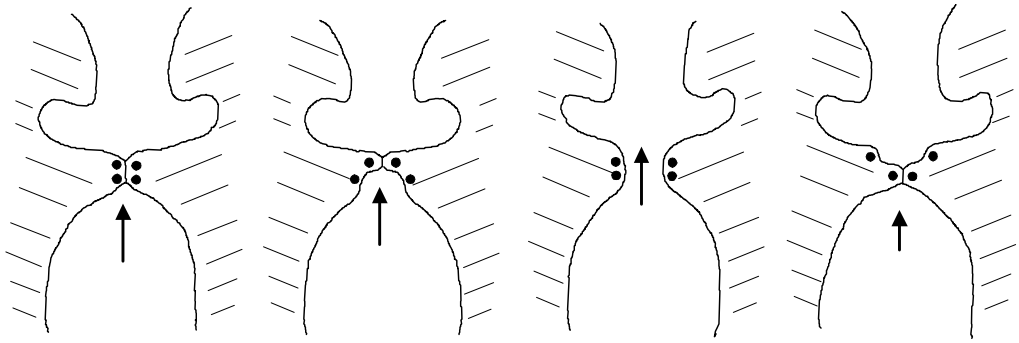


Bild 2.3: Zyklus einer Glottisschwingung mit zwei Massen (schwarze Punkte) auf jeder Seite.

mehr so stark auseinander gedrückt werden. Durch die Rückstellkräfte der Stimmbandmuskeln schließt sich die Glottis wieder, wodurch der Druck vor der geschlossenen Glottis sich wieder aufbaut. Eine Periode der stimmhaften Anregung ist durch das Öffnen und Schließen der Glottis beschrieben. Die schwingenden Massen der Stimmbänder, die Rückstellkräfte der Stimmbandmuskeln und die Antriebskräfte durch den subglottalen Druck stellen Komponenten eines schwingenden Systems dar. Diese vereinfachte Modellbetrachtung der Phonation ist die Grundlage für die Ein-, Zwei- und m-Massen Modelle der selbstschwingenden Glottismodelle. Wie in Bild 2.3 zu sehen ist, können die beiden Punkte an den Stimmlippen die bewegten Massen darstellen. Das Zwei-Massenmodell [Is72] beschreibt die Bewegung dieser beiden Massen infolge der auftretenden Kräfte. Es ist zu bemerken, daß neben einer horizontalen Bewegung auch eine vertikale Bewegung der Stimmlippen auftritt, wie z.B. in [Ii89] an Hand von Simulationen zu sehen ist. Wie in der Bewegung von Bild 2.3 kommt es bei dieser Schwingungsform vor, daß die obere Masse der unteren Masse mit einer zeitlichen Verzögerung folgt. Dadurch kann die Bewegung mit Hilfe nur eines Massenparameters beschrieben werden [Av01]. Durch das Ausströmen der Luft aus der Glottis in den Rachenraum können sich Wirbel bilden, die eine Erklärung für die vorhandenen Rauschanteile in den stimmhaften Sprachsignalen geben. Die Entstehung der Wirbel kann durch numerische Simulationen beobachtet werden [Li91, Zh02]. Insbesondere im höheren Spektralbereich wird das Sprachsignal zunehmend unperiodisch. Durch verschiedene Stimmarten kann das zur periodischen Anregung überlagerte Rauschen variieren [Chi91]. Eine Abweichung von der Periodizität wird auch durch eine ungleichmäßige Schwingung verursacht, die sich auf die Schwingungsamplitude, wie auch auf die Schwingungsdauer auswirkt. Diese beiden Effekte sind unter den Begriffen Jitter und Shimmer bekannt, welche auch von der Vokaltraktgeometrie abhängig sind, wie in [Kr91] auch durch interagierende Glottis- und Vokaltraktmodelle beobachtet werden kann. Die stimmhafte Anregung ist daher selbst bei konstant gehaltener Grundfrequenz und Lautstärke nicht exakt periodisch. Der Sprechtrakt wird durch den Schallfluß aus der Glottis angeregt. Es wird davon ausgegangen, daß die Öffnungsfläche der Glottis und der Schallfluß stark korrelieren. Die Glottisfunktion kann durch parametrische Modelle beschrieben werden [Fa86a, Ol93]. Ein Beispiel der Glottisfunktion von [Ol93] ist in Bild 2.4 dargestellt. Es ist zu sehen, daß das Schließen der Glottis schneller erfolgt als das Öffnen. Der tatsächliche Glottisfluß besitzt in Wirklichkeit oft einen

Mindestwert, wie in [Ro73] beschrieben. Darüber hinaus weist die Form der Glottisfunktion in der geöffneten Phase zusätzliche Höcker (ripple) in Folge von Interaktionen zwischen Stimmlippen und Vokaltrakt auf [Ba94]; womöglich sind auch Nichtlinearitäten beteiligt. Da das Sprachsignal durch die Lippenabstrahlung eine Hochpaßfilterung erfährt, existieren auch parametrische Modelle der Glottisfunktion in der ersten zeitlichen Ableitung [Fa86a]. Die Parameter können dabei die Form des Signals verändern, wie z.B. das Verhältnis zwischen offener und geschlossener Glottisphase. Diese Modelle beschreiben wie erwähnt nicht alle Merkmale der Glottisfunktion. Meist wird für eine Modellierung angenommen, daß die Glottisanregung unabhängig vom Vokaltrakt ist, obwohl Abhängigkeiten existieren wie z.B. in [Bi86] beschrieben.

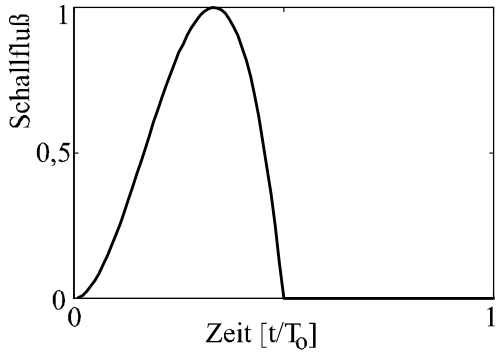


Bild 2.4: Glottisfunktion nach Oliveira.

Turbulente Rauschanregung

Eine im Sprechtrakt auftretende turbulente Strömung bewirkt eine rauschhafte Anregung, welche allerdings erst bei hohen Teilchengeschwindigkeiten und deren Geschwindigkeitsdifferenzen entsteht. Wird im Vokaltrakt der Luftkanal stark verengt, so entstehen insbesondere hinter der Verengung Turbulenzen. Die Wirbel bewegen sich nach ihrer Entstehung mit dem Luftstrom. Der Anregungsort ist bei rauschhafter Anregung daher kurz hinter der Konstriktion anzunehmen. Sind die Stimmbandmuskeln kaum oder gar nicht angespannt, so entsteht infolge der weit geöffneten Glottis durch einen hohen subglottalen Druck ein stetiger starker Teilchenfluß durch den Vokaltrakt. Durch diesen entstehen für Reibelaute infolge der Konstriktion Wirbel, welche eine nahezu stationäre Rauschanregung bewirken. Diese Rauschanregung kann auch bei gleichzeitiger Phonation stattfinden, so daß zwei Anregungsmechanismen an unterschiedlichen Vokaltraktstellen zugleich auftreten. Bei Phonation stellt die Glottis eine zeitvariable Verengung dar, so daß sich auch Turbulenzen hinter der Glottis bilden können. Dadurch besitzt die stimmhafte Anregung zusätzlich einen schwachen Rauschanteil. Für die Analyse von Frikativen bzw. Reibelauten mittels Sprechtraktmodellen ist es problematisch, daß die spektrale Einhüllende des Rauschspektrums, welches die Anregung bildet, in der Regel nicht bekannt ist. In [Sh91] ist zu sehen, daß die spektrale Verteilung unter anderem von der Vokaltraktgeometrie abhängt.

Anregung durch Explosion

Die stimmhafte und rauschhafte Anregung kann weitgehend als stationär angesehen werden. Im Gegensatz dazu stellt die Anregung der Explosive, welche durch einen einmaligen Druckabbau entsteht, eine zeitlich instationäre Anregung dar. Wird der Vokaltrakt an einer Stelle zu einem Verschluß verengt, so baut sich bei gleichzeitigem Anspannen des Zwerchfells vor dem Verschluß ein Druck auf, der dem subglottalen Druck entspricht. Dafür muß das Velum geschlossen sein, damit sich der Druck nicht über die Nasenlöcher abbauen kann. Vor dem Explosionsgeräusch herrscht daher eine kurze Ruhephase. Beim Lösen des Verschlusses baut sich der Druck plötzlich ab, wodurch ein impulsartiger Schall (burst) erzeugt wird. Dieser Druckabfall ist so schnell, daß er durch ein sehr kurzzeitiges Rauschen oder sogar durch einen Impuls modelliert werden kann. Die Einhüllende des Anregungsspektrums kann daher annäherungsweise als konstant angesehen werden. Nach dem sehr schnellen Druckabfall folgt eine kurzzeitige nahezu stationäre Phase mit Rauschen, welche durch den turbulenten Luftstrom verursacht wird. Bei stimmhaften Explosiven erfolgt mit der Lösung des Verschlusses gleichzeitig ein Stimmeinsatz, wobei die Glottis schon vor dem Lösen des Vokaltraktverschlusses schwingen kann [Lade96]. Vor der Verschlußlösung des stimmhaften Explosivs kann eine gedämpfte niederfrequente Schwingung beobachtet werden, welche durch Körperschall abgestrahlt wird.

Nasaltrakt

Durch Senken des Velums ist das Höhlungssystem des Nasaltraktes mit dem Vokaltrakt verbunden. Der Nasaltrakt verzweigt sich hinter dem Velum in einen rechten und linken Nasengang, dessen Ausgänge die beiden Nasenlöcher bilden. Die beiden Nasengänge weisen eine Unsymmetrie auf, wodurch auch das Übertragungsverhalten des Nasaltraktes beeinflußt wird [Da94]. An den Nasengängen sind jeweils mehrere Nebenhöhlen durch dünne Kanäle angekoppelt. Die wichtigsten Nebenhöhlen sind die Stirnhöhlen, Kieferhöhlen und Keilbeinhöhlen. Diese sind jeweils paarweise und näherungsweise symmetrisch angeordnet. Die Kieferhöhlen befinden sich rechts und links vom Nasengang. Die Stirnhöhlen sind vorne oberhalb der Nasenhöhle angeordnet, während sich die Keilbeinhöhlen im Inneren des Kopfes befinden. Neben diesen Höhlen existieren noch weitere Nebenhöhlen wie z.B. die Siebbeinzellen. Die Geometrie des Nasaltraktes kann während des Sprechens als überwiegend unveränderbar angesehen werden. Nur im Bereich des Velums treten Änderungen in der Nasengeometrie während des Sprechens auf, so daß in Rohrmodellen des Nasaltraktes die Querschnitte nahe dem Velum als variabel angenommen werden können [Ma82]. Über längere Zeitabschnitte hinweg kann sich die Geometrie des Nasaltraktes infolge von Schleimlösung und Schleimbildung sowie durch An- und Abschwellen des Gewebes verändern. Obwohl die Gestalt der Nase recht kompliziert ist, existieren Rohrmodelle für ebene Wellenausbreitung, welche den Nasaltrakt modellieren. Eine mögliche Modellierung der Nebenhöhlen kann dabei durch Verzweigungen realisiert werden. Das eindimensionale Modell mit möglichen Verzweigungen des Nasaltraktes ist z.B. in [Ma82, Mey89, Da94, Da96b, Fe96, Kr98, SnL99a] für die Analyse und zur Spracherzeugung verwendet.

Systematik der Sprachlaute

Die Sprachlaute werden phonetisch nach der Anregungsart und nach den Einstellungen der Artikulatoren klassifiziert. Die Sprechtraktgeometrie wird durch die Artikulatoreinstellungen bestimmt. So können die Vokale nach der Position des höchsten Punktes des Zungenrückens in zwei Dimensionen klassifiziert werden, wobei durch die Lippenrundung noch eine zusätzliche Dimension möglich ist. Mit anderen Parametern sind noch weitere Klassifikationen möglich [Lade96]. Die Konsonanten werden nach der Stelle der Vokaltraktkonstriktion eingeteilt, welche stark mit der stimmlosen Anregungsstelle korreliert. Zusätzlich wird noch durch die Anregungsart in stimmhafte und stimmlose Konsonanten unterteilt, wobei die stimmhaften Konsonanten auch eine zusätzliche stimmlose Anregung aufweisen können. Die verschiedenen Positionen der Konstriktionen im Vokaltrakt sind in Bild 2.5 dargestellt, wobei hier nur die wichtigsten Stellen aufgeführt sind.Diese Orte können sich infolge Koartikulation für Realisierun-

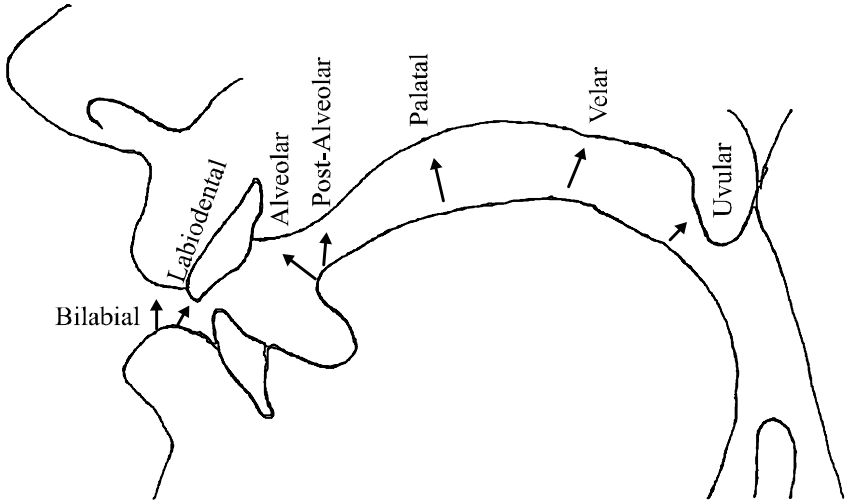


Bild 2.5: Verengungsstellen im Vokaltrakt (Auswahl).

gen in bestimmten Lautfolgen verschieben. Die Konstriktionsstelle kann für bestimmte Konsonanten mit Sensoren am Palatum, welche den Kontakt mit der Zunge feststellen, direkt gemessen werden [En01a]. Im Anhang werden Beispiele der Lautschrift in SAMPA-Notation aufgelistet, welche sich nach IPA richtet. Die abstrakten Lautklassen werden als Phoneme bezeichnet. Da Sprachäußerungen nicht exakt reproduzierbar sind und sich somit immer unterscheiden, kann immer nur eine Realisierung eines Phonems in der Form der Sprachlaute akustisch gemessen werden.

Vokaltraktgeometrie

Die Zunge ist der wichtigste und zugleich variabelste Artikulator für die Veränderung der Vokaltraktgeometrie. Die Einstellung der Zunge wird meist im Sagittalschnitt betrachtet. In dieser Betrachtungsweise kann die Zunge schon mittels einem oder zwei Parametern dargestellt werden, wenn der eine Parameter den Zungenmittelpunkt und der zweite Parameter die Zungenspitze beschreibt. Für eine genauere Modellierung sind mehr als zwei Parameter notwendig. Neben der Modellierung des Sagittalschnittes der Zunge existieren auch dreidimensionale Zungenmodelle [En99, En01b, Da99, Pas99],

die allerdings für die eigentliche Lauterzeugung in ein eindimensionales Vokaltraktmodell transformiert werden. Außer der Zunge wirken als wichtige Artikulatoren noch die Lippen, das Velum, der Unterkiefer und gegebenenfalls das Zungenbein. Es existieren Ansätze, welche die Einstellungen der Artikulatoren mehr oder weniger genau auf den Sagittalschnitt des Vokaltraktes abbilden. Das Übertragungsverhalten des Vokaltraktes ist allerdings maßgeblich von seiner Querschnittsflächenfunktion abhängig. Die Vokaltraktgeometrie kann durch Röntgenaufnahmen [Fa70] oder NMR-Aufnahmen [St96] ermittelt werden, aus denen die Querschnittsfunktionen folgen. Eine Auswahl dieser Querschnittsflächen ist im Anhang für einen Vergleich mit den geschätzten Flächen dargestellt. In [St98] sind Flächen aus NMR-Aufnahmen einer weiblichen Sprecherin gezeigt und in [St01] sind ebenfalls anhand von NMR-Aufnahmen Flächen für unterschiedliche Sprechweisen aufgezeigt. Für artikulatorische Modelle ist statt der komplizierten dreidimensionalen Geometrie der Sagittalschnitt hilfreich, welcher allerdings nicht direkt die Querschnittsfläche angibt. In [Jo83, Su87] ist anhand von Tomographieaufnahmen vom Sagittalschnitt auf die Querschnittsfläche geschlossen worden. Es gibt verschiedene einfache mathematische Modellfunktionen, die das Verhältnis zwischen Höhe bzw. Breite d im Sagittalschnitt und der Querschnittsfläche A angeben. Mit das einfachste ist das $\boldsymbol{\alpha}$-$\boldsymbol{\beta}$-Modell, welches durch

$$A = \boldsymbol{\alpha} \cdot d^{\beta} \tag{2.1}$$

definiert ist und dessen Parameter in [Soq96] an Tomographie-Daten angepaßt sind. Weiterhin existiert auch ein Polynomansatz. In [Gr82] ist z.B. für den Bereich des Rachens und Mundraums ein Polynom 3. Grades als ausreichend erachtet worden, dessen Polynomkoeffizienten an MRT-Aufnahmen angepaßt sind. In [Soq02] ist für Parametereinstellungen unterschiedlicher Modelle eine Fehlerbewertung anhand von Tomographieaufnahmen durchgeführt. Dafür muß die dreidimensionale Geometrie des Vokaltraktes durch ein Rohr mit veränderlichem Querschnitt ersetzt werden. Es zeigt sich, daß die Güte der Modelle von dem betrachteten Vokaltraktbereich abhängt. Die Vokaltraktflächen weisen an unterschiedlichen Positionen unterschiedliche Flächenumrisse auf. Die Umrisse können zum Teil stark variieren, was auch in [Fa70, Soq02, Jo83, Su87] zu sehen ist. Zusätzlich verändern sich auch die Formen der Flächenumrisse in der Regel für unterschiedlich große Flächen an einer festen Vokaltraktposition unter Berücksichtigung des Skalierungsfaktors. Bei einer eindimensionalen Betrachtung des Vokaltraktes ist anzumerken, daß er im mittleren Bereich gebogen ist. Sondhi hat in [So86] abgeschätzt, daß die Sprechtraktbiegung nur einen geringen Einfluß auf das akustische Verhalten hat. Es existieren auch parametrische Modelle der Querschnittsfunktion des Vokaltraktes. In [Fa97] ist ein Modell vorgestellt, welches den Vokaltrakt in Abschnitte einteilt, welche durch unterschiedliche Ansätze parametrisiert werden.

Koartikulation

Eine der Schwierigkeiten in der Sprachverarbeitung wird dadurch verursacht, daß die Laute in unterschiedlichen Wörtern und Sätzen verschiedene Realisierungen aufweisen. Die Variationen der realisierten Laute hängen insbesondere von ihrer Lautumgebung ab. Dies läßt sich dadurch erklären, daß eine Äußerung als eine zusammenhängende Sprechtraktbewegung angesehen wird. Um eine fließende Bewegungen beim Sprechen zu ermöglichen, werden nicht alle Zielstellungen (targets) der Artikulatoren für die entsprechenden Laute erreicht. Dies liegt daran, daß die Bewegungen zum Teil abgekürzt

werden oder manche Artikulatorstellungen von der lautlichen Umgebung sogar völlig übernommen werden. Dabei muß beachtet werden, ob die Artikulatorstellung für die Charakterisierung des Lautes zwingend vorliegen muß. Falls dies nicht der Fall ist, kann die Einstellung des Artikulators infolge Koartikulation variabel verändert werden, was in [Kr92] unter dem Begriff eines passiven Artikulators für die artikulatorische Sprachsynthese verwendet wird. Daß die Velumstellung der Vokale stark von benachbarten Nasalen abhängt, ist in [BSnL02] anhand von Leistungsvergleichen von getrennten Mund- und Nasensignalen gut zu sehen. Die ineinander übergehenden Bewegungen der Artikulatoren können analog zu der Motorik des menschlichen Bewegungsapparates gesehen werden. Dies bedeutet, daß die Artikulatoren so eingestellt werden, daß sie sich möglichst wenig ändern und eine weitgehend flüssige Bewegung ermöglichen. Nicht alle Artikulatoren können gleich schnell eingestellt werden, wodurch die Artikulatorbewegungen für die Laute nicht immer zur gleichen Zeit stattfinden. So kann es z.B. vorkommen, daß manche Artikulatorstellungen viel weiter vor dem zugehörigen Laut eingestellt werden, als es für andere Artikulatoren der Fall ist. Es können sich genau genommen alle Laute in einer Lautkette gegenseitig beeinflussen, obwohl die Beeinflussung zwischen benachbarten Lauten im allgemeinen am stärksten ausgeprägt ist. Dabei wird der nachfolgende Laut vom vorangegangenen Laut in der Regel stärker beeinflußt als umgekehrt [Ke77]. Aus den genannten Ursachen der Koartikulation folgt, daß für kurze Vokale die Koartikulation stärker ausgeprägt ist als für lange Vokale, da weniger Zeit für die Sprechtraktbewegung zur Verfügung steht. Die Koartikulation führt dazu, daß infolge der modifizierten Artikulatorstellungen die Laute mit unterschiedlichen akustischen Merkmalen in fließender Sprache auftreten. Die Lautmodifikationen werden vom Menschen zum Teil nicht wahrgenommen. Der Mensch ist durch seine unbewußt ablaufende Sprachsignalverarbeitung des Innenohrs und Gehirns größtenteils in der Lage die unterschiedlichen Realisierungen der Laute in ihrer lautlichen Umgebung als die richtigen Phoneme zu erkennen, wie sie vom Sprecher beabsichtigt wurden. Hier spielen auch semantische Gesichtspunkte eine Rolle. Der Mensch hat die verschiedenen Realisierungen der Laute in Lautketten erlernt, so daß er Lautfolgen ohne Koartikuationseffekte als unnatürlich empfindet oder gar nicht versteht. Obwohl der Mensch sich der Koartikulation nicht direkt bewußt ist und sie nur bedingt wahrnimmt, stellt sie für technische Anwendungen der Sprachverarbeitung große Probleme dar. Dies trifft für die Spracherkennung wie für die Sprachsynthese zu. Um dieser Problematik zu begegnen lassen sich für die Spracherzeugung zwei Lösungsansätze aufzeigen. Ersterer ist der modellbasierte Ansatz, in dem versucht wird ein Sprachproduktionsmodell zu verwenden, in dem die Regeln der Koartikulation größtenteils inhärent enthalten sein sollen. Der zweite Ansatz geht nur von den Produktionsresultaten aus und verwendet daher die vielfältigen Realisierungen der Sprachsignale und ihrer akustischen Merkmale, welche durch statistische Modelle untersucht werden können.

Kapitel 3

Zeitdiskretes Rohrmodell

3.1 Akustische Grundlagen der Modellierung des Sprechtraktes

Die für die Akustik relevanten grundlegenden physikalischen Größen sind Dichte, Druck, die mittlere Teilchengeschwindigkeit und gegebenenfalls die Temperatur. Für eine Berechnung der Schallfelder reichen in den meisten Fällen Druck und/oder die mittlere Teilchengeschwindigkeit aus. Schallereignisse finden statt, wenn das Gleichgewicht dieser beiden Größen gestört ist. Dies liegt vor, wenn der Druck oder der Teilchenfluß des Mediums nicht homogen verteilt sind. Liegt eine Druckdifferenz zwischen zwei Gebieten vor, so resultiert eine Kraft auf die Teilchen des dazwischen liegenden Bereichs. Die Teilchen werden beschleunigt und strömen in ein benachbartes Gebiet mit geringerem Druck. Dadurch erhöht sich in Folge der eindringenden Teilchen die Dichte, wodurch sich auch der Druck dort erhöht. Eine Schallwelle breitet sich daher durch die wechselseitige Beziehung zwischen Schalldruck und Teilchenfluß aus. Daher verknüpfen die Grundgleichungen der Akustik den Druck mit dem Teilchenfluß bzw. Schnelle. Zwei unabhängige Gleichungen werden benötigt, um die zeitliche und örtliche Entwicklung einer Größe isoliert diskutieren zu können. Da es nur auf die Druckdifferenzen ankommt ist nicht der absolute Druck p_a maßgeblich, sondern die Differenz zum zeitlich mittlerem Druck p_0. Der Geschwindigkeitsvektor $\boldsymbol{v}_a$ wird Schallschnelle genannt und kann als Mittelung über alle Teilchengeschwindigkeiten in einem sehr kleinen Gebiet angesehen werden. Die Druckdifferenzen p werden als Schallwechseldruck bezeichnet und werden hier im Gegensatz zum absoluten Druck p_a ohne Index dargestellt. Die absolute Dichte $\boldsymbol{\rho}_a$ und die Schallschnelle werden analog zum Druck in zeitveränderliche und konstante Größen aufgeteilt:

$$p_a = p_0 + p \tag{3.1}$$

$$\boldsymbol{v}_a = \boldsymbol{v}_0 + \boldsymbol{v} \tag{3.2}$$

$$\boldsymbol{\rho}_a = \boldsymbol{\rho}_0 + \boldsymbol{\rho}. \tag{3.3}$$

Für die quantitative Beschreibung existieren die Grundgleichungen der Akustik in Form der Euler Gleichung und der Kontinuitätsgleichung. Hier werden sie wie so oft als Näherung in linearisierter Form verwendet. Die Euler Gleichung:

$$-\boldsymbol{\nabla} p_a = \boldsymbol{\rho}_a \left(\frac{\partial \boldsymbol{v}_a}{\partial t} + (\boldsymbol{v}_a \cdot \boldsymbol{\nabla}) \boldsymbol{v}_a \right) \tag{3.4}$$

ist die allgemeine Bewegungsgleichung einer reibungsfreien Flüssigkeit oder eines Gases. Auf der linken Seite der Euler Gleichung steht die Kraft in differentieller Form, während auf der rechten Seite die Dichte mal der Beschleunigung steht. Da die Kraft auf ein sich bewegendes Teilchens bzw. Volumenelementes in einer Strömung betrachtet wird, kann das Teilchen bei einer infinitesimalen Bewegung eine Veränderung der Kraft im Vergleich zum vorherigen Ort erfahren. Dies wird durch den Term $(\boldsymbol{v}\cdot\boldsymbol{\nabla})\boldsymbol{v}$ berücksichtigt. Mit zusätzlichen dissipativen Termen, wie z.B. $\boldsymbol{\eta}\triangle\boldsymbol{v}$ mit der Viskosität $\boldsymbol{\eta}$, erhält man aus der Eulergleichung die Navier-Stokes Gleichung. Diese kann für numerische Berechnungen der glottalen Anregung und des Sprechtraktes verwendet werden, hat allerdings den Nachteil, daß sie sehr rechenintensive Lösungswege beinhaltet. Daher werden die Navier-Stokes Gleichungen in [Ii89, Li91] nur zweidimensional behandelt für eine Untersuchung der Glottis. In [Li91] läßt sich eine Wirbelbildung beobachten. Eine numerische Lösung in Zylinderkoordinaten wird in [Zh02] vorgestellt, in der die Wirbelbildung während einer Glottisschwingung diskutiert wird. Den Einfluß der Einbeziehung der Navier-Stokes Gleichung auf die Lösung und der resultierenden Parameter, wie zum Beispiel die Grundfrequenz, ist in [Vr02] diskutiert. Für die Modellierung des Sprechtraktes ohne Einbeziehung der Glottis ist die Verwendung der Euler Gleichung angebracht. Für praktische Berechnungen verwendet man Näherungen für die Euler Gleichung (3.4). Da die Dichteschwankungen meist klein sind gegenüber dem Mittelwert $\boldsymbol{\rho}_0$, kann der Mittelwert statt dem absoluten Wert in (3.4) verwendet werden. Weiterhin nimmt $\boldsymbol{v}$ in der Regel verhältnismäßig kleine Werte an, so daß die höheren Potenzen von $\boldsymbol{v}$ vernachlässigt werden können, wodurch der Term $(\boldsymbol{v}\cdot\boldsymbol{\nabla})\boldsymbol{v}$ weggelassen wird. Diese akustische Näherung wirkt sich nicht auf alle Bereiche der Sprachproduktion gleich aus. In der Glottis können durch die kleinen Öffnungsflächen infolge des subglottalen Druckes hohe Geschwindigkeiten auftreten, so daß der Term $(\boldsymbol{v}\cdot\boldsymbol{\nabla})\boldsymbol{v}$ einen nicht zu vernachlässigenden Beitrag liefert. Die Energie des glottalen Flusses kann nur teilweise in einen akustischen Fluß des Vokaltraktes umgewandelt werden. Für die Akustik des Vokaltraktes, der sich von der Glottis bis zu den Lippen erstreckt, werden die Abweichungen durch die Näherungen als gering angesehen. Durch die erläuterten Vereinfachungen erhält man die linearisierte Euler Gleichung:

$$-\boldsymbol{\nabla}p = \boldsymbol{\rho}_0\frac{\partial\boldsymbol{v}}{\partial t}. \tag{3.5}$$

Durch die örtlichen und zeitlichen Ableitungen können p_0 und $\boldsymbol{v}_a$ durch die Wechselgrößen p und $\boldsymbol{v}$ ersetzt werden. Die zweite Grundgleichung ist die Kontinuitätsgleichung:

$$\boldsymbol{\nabla}\cdot(\boldsymbol{\rho}_a\boldsymbol{v}_a) = -\frac{\partial\boldsymbol{\rho}_a}{\partial t}, \tag{3.6}$$

welche eine Bilanzgleichung darstellt. In der Eulergleichung kommt das Variablenpaar Druck und Schallschnelle vor, so daß es von Vorteil ist, wenn in der Kontinuitätsgleichung das selbe Variablenpaar auftritt. Da nur sehr kleine Druckschwankungen auftreten, kann die Beziehung zwischen p und $\boldsymbol{\rho}$ linearisiert werden, so daß p proportional zu $\boldsymbol{\rho}$ angenommen wird. Die Proportionalitätskonstante kann mit Hilfe der Adiabatengleichung bestimmt werden. Die Druckschwankungen sind in der Regel so schnell, daß kein Temperaturaustausch stattfindet. Aus dem Adiabatengesetz und der Annahme, daß die absoluten Dichten und Volumina umgekehrt proportional mit der

Konstanten k durch $V_a = k/\boldsymbol{\rho}_a$ sind, folgt für zwei Volumina V_0 und $V' = V_0 + V$:

$$p_0 V_0^{\varkappa} = (p_0 + p)\,(V_0 + V)^{\varkappa} \tag{3.7}$$

bzw.

$$p_0 \left(\frac{k}{\boldsymbol{\rho}_0}\right)^{\varkappa} = (p_0 + p) \left(\frac{k}{\boldsymbol{\rho}_0 + \boldsymbol{\rho}}\right)^{\varkappa}, \tag{3.8}$$

wobei $\varkappa$ den Adiabatenexponent darstellt. Durch Trennung der Dichten und Drücke wird (3.8) umgestellt zu

$$\left(1 + \frac{p}{p_0}\right) = \left(1 + \frac{\boldsymbol{\rho}}{\boldsymbol{\rho}_0}\right)^{\varkappa}. \tag{3.9}$$

Da $\boldsymbol{\rho} \ll \boldsymbol{\rho}_0$ gilt, kann der rechte Term durch eine Taylorreihenentwicklung:

$$\left(1 + \frac{\boldsymbol{\rho}}{\boldsymbol{\rho}_0}\right)^{\varkappa} \simeq 1 + \varkappa \frac{\boldsymbol{\rho}}{\boldsymbol{\rho}_0} \tag{3.10}$$

mit Abbruch nach dem linearen Term nahezu perfekt approximiert werden, woraus die Beziehung zwischen Druck und Dichte folgt:

$$\frac{p}{\boldsymbol{\rho}} = \varkappa \frac{p_0}{\boldsymbol{\rho}_0} \qquad \Longrightarrow \qquad p = c^2 \boldsymbol{\rho} \;\text{ mit }\; c = \sqrt{\frac{\varkappa p_0}{\boldsymbol{\rho}_0}}. \tag{3.11}$$

Die Proportionalitätkonstante ist das Quadrat der Schallgeschwindigkeit c , was bekanntlich erst durch weitere Ableitungen gesehen werden kann. Wird zusätzlich auf der linken Seite der Kontinuitätsgleichung (3.6) $\boldsymbol{\rho}_a$ durch $\boldsymbol{\rho}_0$ ersetzt für $\boldsymbol{\rho} \ll \boldsymbol{\rho}_0$ und werden infolge der Ableitungen Wechselgrößen verwendet, so läßt sich die Kontinuitätsgleichung darstellen als

$$\boldsymbol{\nabla} \cdot (\boldsymbol{\rho}_0 \boldsymbol{v}) = -\frac{1}{c^2} \frac{\partial p}{\partial t}. \tag{3.12}$$

Mit (3.5) und (3.12) stehen zwei vereinfachte Gleichungen zur Verfügung, die Schalldruck und Schallschnelle verbinden. Zweckmäßig ist das Vorhandensein von nur einer Gleichung, die nur noch eine Größe enthält und somit in Abhängigkeit von Zeit und Ort diskutiert werden kann. Eliminiert man z.B die Schallschnelle durch geeignetes Differenzieren und Gleichsetzen von (3.5) und (3.12), so ergibt sich die Wellengleichung:

$$\frac{\partial^2 p}{\partial t^2} = c^2 \triangle p. \tag{3.13}$$

Diese besitzt Lösungen der Form

$$\sum_{\boldsymbol{n}} f_{\boldsymbol{n}}(\boldsymbol{n} \cdot \boldsymbol{r} - ct) \tag{3.14}$$

wobei $\boldsymbol{r}$ der Ort und $\boldsymbol{n}$ ein Einheitsvektor ist, der die Ausbreitungsrichtung der ebenen Welle darstellt. $f_{\boldsymbol{n}}$ sind beliebige Funktionen. Reelle ebene Wellen einer Frequenz können in der komplexen Schreibweise durch den Realteil:

$$\mathrm{Re}\left\{\underline{p} \exp(j\boldsymbol{k} \cdot \boldsymbol{r} - j\boldsymbol{\omega} t)\right\} \quad \text{mit} \quad \boldsymbol{\omega} = |\boldsymbol{k}|\, c = kc \tag{3.15}$$

dargestellt werden, wobei $\underline{p}$ die komplexe Amplitude der Mode ist. Für die eindimensionale Ausbreitung ebener Wellen entlang der x-Achse existieren zwei Lösungen der Wellengleichung

$$\frac{\partial^2 p}{\partial t^2} = c^2 \frac{\partial^2 p}{\partial x^2} \tag{3.16}$$

mit $|\boldsymbol{n}| = 1$. Das Signal p wird beschrieben als Überlagerung einer hinlaufenden Wellen p^+ und einer rücklaufenden Welle p^-:

$$p(x,t) = p^+(x - ct) + p^-(x + ct), \tag{3.17}$$

wobei p^+ und p^- wieder beliebige Funktionen sein können. Die Größe $c = \sqrt{\varkappa p_0/\boldsymbol{\rho}_0}$ stellt die Schallgeschwindigkeit dar. Aus (3.5) und (3.12) kann ebenso die Wellengleichung für die Schallschnelle erhalten werden, die eine analoge Lösung

$$v(x,t) = v^+(x - ct) - v^-(x + ct) \tag{3.18}$$

liefert, wobei v die mittlere Geschwindigkeit der Teilchen entlang der x-Achse ist. Hierbei muß beachtet werden, daß die Teilchengeschwindigkeit eine vektorielle Größe ist und hier nur die Beträge der hin- und rücklaufenden Welle verwendet werden. Für die Berechnung der Schallschnelle v wird dies durch ein Minuszeichen der rücklaufenden Welle berücksichtigt. Für weitere Berechnungen ist es zweckmäßig eine Beziehung zwischen Druck- und Schallschnellewellen herzustellen. Dies geschieht durch Ableitungen von Druck und Schnelle nach Ort und Zeit, welche in die Euler bzw. Kontinuitätsgleichung eingesetzt werden; anschließendes Integrieren liefert

$$p^+(x - ct) = c\boldsymbol{\rho}_0 \cdot v^+(x - ct) \qquad \text{und} \qquad p^-(x - ct) = c\boldsymbol{\rho}_0 \cdot v^-(x - ct). \tag{3.19}$$

Das Verhältnis zwischen der Druckwelle und Schnellewelle für die hin- und rücklaufenden Anteile der ebenen Welle wird Schallkennimpendanz genannt, welche mit Z_0 dargestellt wird:

$$Z_0 = \frac{p^+}{v^+} = \frac{p^-}{v^-} = c\boldsymbol{\rho}_0. \tag{3.20}$$

Die Welle p^+ mit nur einer Frequenzkomponente kann durch die komplexe Amplitude $\underline{p}^+$ mit

$$p^+(x,t) = \mathrm{Re}\left\{\underline{p}^+ \cdot e^{jkx} e^{-j\omega t}\right\} \tag{3.21}$$

dargestellt werden. Die Impedanz $\underline{Z}$ als komplexe Zahl wird auf die komplexen Amplitudenwerte angewendet mit:

$$p^+(x,t) = \mathrm{Re}\left\{\underline{Z} \cdot \underline{v}^+ \cdot e^{jkx} e^{-j\omega t}\right\} \qquad \text{bzw.} \qquad \underline{p}^+ = \underline{Z} \cdot \underline{v}^+ \tag{3.22}$$

$$\text{und} \qquad p^-(x,t) = \mathrm{Re}\left\{\underline{Z} \cdot \underline{v}^- \cdot e^{-jkx} e^{-j\omega t}\right\} \qquad \text{bzw.} \qquad \underline{p}^- = \underline{Z} \cdot \underline{v}^-. \tag{3.23}$$

Da die Impedanz $\underline{Z}$ der ebenen Welle mit $Z_0 = c\boldsymbol{\rho}_0$ reell ist, sind die Druck- und Schnellewellen für ebene Wellen in Phase.

Modellierung des Sprechtraktes

Die vereinfachte Betrachtung des Schallfeldes durch ebene Wellen besitzt in der Sprachverarbeitung für die Modellierung der Akustik des Sprechtraktes einen festen Stellenwert. Der Sprechtrakt ohne Ankopplung des Nasaltrakts bildet in erster Näherung ein

unverzweigtes langgestrecktes Rohrsystem mit veränderlichem Querschnitt. Wie z.B. in [Fa70, Soc02] zu sehen ist, verändert sich auch die Querschnittsform entlang des Vokaltraktes. Zusätzlich kann sich die Form des Vokaltraktquerschnitts in Abhängigkeit von der Fläche an einer bestimmten Vokaltraktposition ändern. Diese Geometrie kann nur durch eine numerische Lösung der dreidimensionalen Wellengleichung (3.13) vollständig berücksichtigt werden. Es hat sich gezeigt, daß die Lösungen einer vereinfachten Geometrie mit einer anschließenden Dimensionsreduktion sich als praktikable Näherungen erweisen. Das Höhlungssystem des Vokaltraktes kann dafür als Rohr mit veränderlicher Querschnittsfläche interpretiert werden. Die Randbedingungen an den Wänden werden in erster Näherung als schallhart angenommen, wodurch die Schallschnelle an den Wänden verschwindet. Die Randbedingung kann durch eine von der Wand reflektierten Welle erfüllt werden, welche eine Schallschnelle mit umgekehrten Vorzeichen aufweist. Zusätzlich zu den ebenen Wellen, die sich längs des Rohres ausbreiten, ist es möglich, daß sich noch Wellen quer bzw. schräg dazu ausbilden. Um den Einfluß dieser Querwellen abzuschätzen, werden Lösungen von vereinfachten dreidimensionalen Geometrien herangezogen. Das kreisrunde und rechteckige Rohr können durch ihre Symmetrie analytisch gelöst werden. Dafür werden geeignete Koordinaten gewählt, die die Symmetrie berücksichtigen, so daß ein Produktansatz zur Lösung führt. Für ein Rohr mit rechteckigem Querschnitt, der Seitenlängen a und b, ergeben sich mit den kartesischen Koordinaten die Lösungen als Moden:

$$p(x,y,z,t) = \mathrm{Re}\left\{\left(\underline{p_x}^{\pm} e^{jk'x} \cdot p_y e^{jym\pi/a} \cdot p_z e^{jzn\pi/b}\right) \cdot e^{-j\omega t}\right\}, \tag{3.24}$$

wobei n und m natürliche Zahlen sind. Die Lösungen quer zur Ausbreitungsrichtung ergeben stehende Wellen infolge der Überlagerung mit den reflektierten Wellen. Mit (3.15) gilt

$$|\boldsymbol{k}| = \frac{\boldsymbol{\omega}}{c} = \frac{2\boldsymbol{\pi}}{\boldsymbol{\lambda}} = \sqrt{k'^2 + k_y^2 + k_z^2} = \sqrt{k'^2 + \left(\frac{m\boldsymbol{\pi}}{a}\right)^2 + \left(\frac{m\boldsymbol{\pi}}{b}\right)^2} \tag{3.25}$$

woraus folgt, daß sich Querwellen für Wellenlängen $\boldsymbol{\lambda}$ bzw. Frequenzen $\boldsymbol{\nu}$ nur für

$$\boldsymbol{\lambda} \leq 2 \cdot \max(a,b) \qquad \text{bzw.} \qquad \boldsymbol{\nu} \geq \frac{c}{2 \cdot \max(a,b)} \tag{3.26}$$

ausbilden können. Neben dem rechteckigen Rohr weist das kreisrunde Rohr ebenfalls eine einfache Symmetrie auf. Für einen Zylinder mit der Länge l und dem Innenradius r_i kann die Wellengleichung in Zylinderkoordinaten

$$\frac{\partial^2 p}{\partial r^2} + \frac{1}{r}\frac{\partial p}{\partial r} + \frac{1}{r^2}\frac{\partial^2 p}{\partial \boldsymbol{\varphi}^2} + \frac{\partial^2 p}{\partial x^2} = \frac{1}{c^2}\frac{\partial^2 p}{\partial t^2} \tag{3.27}$$

mit r als Radius, $\boldsymbol{\varphi}$ als Winkel und x als Symmetrieachse durch einen Produktansatz

$$\underline{p}(r,\boldsymbol{\varphi},x,t) = R(r) \cdot W(\boldsymbol{\varphi}) \cdot X(x) \cdot T(t) \tag{3.28}$$

gelöst werden. Die einzelnen Lösungen ergeben sich zu:

$$R(r) = C_j J_m(\boldsymbol{\gamma}_{m,n} r) \tag{3.29}$$

$$W(\boldsymbol{\varphi}) = C_{w1} e^{jm\boldsymbol{\varphi}} + C_{w2} e^{-jm\boldsymbol{\varphi}} \tag{3.30}$$

$$X(x) = C_{x1} e^{jk'x} + C_{x2} e^{-jk'x} \tag{3.31}$$

$$T(t) = e^{-j\boldsymbol{\omega}t}. \tag{3.32}$$

Für $R(r)$ ergeben sich als physikalisch sinnvolle Lösungen die ungedämpften Wellenmoden in Form der Besselfunktionen $J_m(\boldsymbol{\gamma}_{m,n}r)$. Für die Einhaltung der Randbedingungen sind dabei $\boldsymbol{\gamma}_{m,n}$ diejenigen Wellenzahlen, die $j'_{m,n} = \boldsymbol{\gamma}_{m,n}r_i$ erfüllen, wobei $j'_{m,n}$ die Nullstellen der abgeleiteten Besselfunktionen J'_m sind. Für die ebenen Wellen in x-Richtung ergeben sich Lösungen wie mit (3.21), nur daß die Wellenzahl k' ähnlich zu m und n abhängig ist. Analog zu den rechteckigen Rohren zeigt sich, daß ungedämpfte Querwellen sich in kreisrunden Rohren nur unter der Bedingung

$$\boldsymbol{\lambda} \leq 1,72 \cdot d_i \qquad \text{bzw.} \qquad \boldsymbol{\nu} \geq \frac{c}{1,72 \cdot d_i} \tag{3.33}$$

ausbilden können [He99], wobei $d_i = 2r_i$ der Innendurchmesser des Rohres ist. Die beiden Beispiele zeigen, daß sich erst ab einer bestimmten Grenzfrequenz Quermoden ausbilden können bei einem endlichen Querschnitt. Die Abhängigkeiten der Grenzfrequenzen von den größten Distanzen im Querschnitt sind für rechteckige und kreisrunde Rohre durch (3.26) und (3.33) ähnlich. Dies kann anschaulich dadurch erklärt werden, daß etwa mindestens für eine halbe Wellenperiode bzw. Wellenlänge Raum im Rohr zur Verfügung stehen sollte, damit die Randbedingungen erfüllt werden können und sich eine Schwingung ausbildet. Die Abschätzungen lassen vermuten, daß für ovalförmige und kompliziertere Querschnitte, wie sie im Vokaltrakt auch vorkommen, ähnliche Abhängigkeiten wie durch (3.26) und (3.33) bestehen. Der informationshaltige Frequenzbereich der Sprachsignale liegt im unteren Teil des hörbaren Frequenzbereichs, so daß Quermoden in diesem kaum zu vermuten sind. Bei Ausbildung der ersten Quermoden kann das Wellenfeld immer noch überwiegend aus ebenen Wellen längs des Rohres bestehen. Es ist zu beachten, daß sich die Quermoden erst bei entsprechender Anregung ausbilden. Bei stimmhaften Lauten bewirkt die Anregung hauptsächlich Schallflüsse in Richtung längs des Vokaltraktes. Da Quermoden erst bei höheren Frequenzen überhaupt auftreten können, wird die Schallausbreitung im Vokaltrakt mit einem Ansatz ebener Wellen modelliert. Bei den Betrachtungen werden die Wände des Rohres meist als schallhart gewählt und zusätzlich eine verlustlose Wellenausbreitung angenommen, was für den Sprechtrakt nur näherungsweise gilt.

Verluste

Im Sprechtrakt treten unterschiedliche Verluste auf, welche sich bezüglich ihrer funktionalen Abhängigkeiten von Rohrdurchmesser und Frequenz unterscheiden. Die Verluste im Vokaltrakt werden als relativ gering eingeschätzt, können allerdings dennoch die Formantfrequenzen und insbesondere die Bandbreiten der Formanten beeinflussen [So74]. Die Verluste für den weitverzweigten Nasaltrakt mit seinen zum Teil sehr kleinen Wandabständen sind stärker als im Vokaltrakt anzunehmen. In [Li85] ist eine Aufstellung von Verlusten für den Vokaltrakt gezeigt, die in Tabelle 1 zu sehen ist. Die Angaben der Dämpfungen D sind in Neper pro Meter angegeben. Ein Neper ist im Gegensatz zu einem Dezibel der natürliche Logarithmus der Größenverhältnisse (1 Neper $\simeq$ 8,7 Dezibel). Für die Formeln ist A die Rohrquerschnittsfläche in m^2, U der Volumenstrom in m^3/s und f die Frequenz in Hertz.

Tabelle 1: Verluste des Vokaltraktes nach [Li85].

Verlustursache		Wert $\left[\frac{\text{Neper}}{\text{m}}\right]$
Viskosität	D_{vsc}	$3,6 \cdot 10^{-5} \cdot \sqrt{\frac{f}{A}}$
Wärmeleitung	D_{heat}	$1,6 \cdot 10^{-5} \cdot \sqrt{\frac{f}{A}}$
Wandvibrationen	D_{wal}	$537 \cdot f^{-2} \sqrt{\frac{1}{A}}$
Wandabsorbtion	D_{wab}	$4,7 \cdot 10^{-4} \cdot \sqrt{\frac{1}{A}}$

In [Li85] ist noch ein Dämpfungsterm D_{lam} für tiefe Frequenzen bei laminarer Strömung und D_{trb} für turbulente Strömung angegeben, die allerdings vernachlässigt werden können. Generell nehmen die Verluste zu für kleinere Rohrdurchmesser. Dies läßt sich folgenderweise erklären: Wirken die Verluste überwiegend an den Wänden, so kann der Verlust proportional zum Rohrumfang angenommen werden. Da bei kreisrunden Rohren der Umfang proportional zum Radius r ist und die schalldurchströmte Fläche A proportional zu r^2 ist, resultiert eine zu r^{-1} bzw. $\sqrt{A}^{-1}$ proportionale Dämpfung. Wie in [Fa70, Soc02] zu sehen ist, besitzen die Umrisse der Vokaltraktflächen insbesondere im Rachenbereich kompliziertere Umrisse, so daß sich bei gleicher Fläche ein größerer Umfang als beim Kreis ergibt. Dadurch sind die Verluste bei gleicher Fläche stärker anzunehmen und können zusätzlich von der Proportionalität $\sqrt{A}^{-1}$ etwas abweichen. Um die Bedeutung der Verluste für ein Rohrmodell des Vokaltraktes abzuschätzen, werden die Verluste einzeln und zusammen mittels D_{ges} für die Frequenzen 0,25 kHz, 0,5 kHz, 1 kHz, 4 kHz und 10 kHz jeweils mit einer Querschnittsfläche von 2 cm^2 in Tabelle 2 gezeigt. Die Dämpfung wird als Verlustfaktor für eine Rohrlänge von 17,5 cm angegeben, was der Länge des Vokaltraktes entspricht.

Tabelle 2: Verteilte Verluste im Vokaltrakt.

	f=250 Hz	f=500 Hz	f=1 kHz	f=4 kHz	f=10 kHz
$e^{-0.175 \cdot D_{vsc}}$	0,993	0,990	0,986	0,972	0,956
$e^{-0.175 \cdot D_{heat}}$	0,997	0,996	0,994	0,988	0,980
$e^{-0.175 \cdot D_{wab}}$	0,994	0,994	0,994	0,994	0,994
$e^{-0.175 \cdot D_{wal}}$	0,899	0,974	0,993	0,999	0,999
$e^{-0.175 \cdot D_{ges}}$	0,884	0,953	0,966	0,953	0,931

Die Dämpfung infolge von Wandvibrationen mit D_{wal} ist für tiefe Frequenzen sehr dominant, wohingegen sie schon ab 1 kHz vernachlässigt werden kann. Für hohe Frequenzen machen sich besonders die viskose Reibung und auch die Wärmeleitung bemerkbar. Diese quantitativ dargestellten Verluste treten im Vokaltrakt verteilt auf. Es kann für die Wandvibrationen auch eine lokale Realisierung an der Glottis angenommen werden, da die Effekte der Wandvibrationen sich hauptsächlich auf tiefe Frequenzen

auswirken. Eine lokale Wirkung hat die Vokaltraktbiegung, die nach Sondhi [So86] als gering einzuschätzen ist. Die Glottis stellt auch einen lokalen Verlust dar, und insbesondere die Ausströmung aus der Glottis in den Vokaltrakt. Durch den Bernoulli-Effekt geht ein Druckverlust einher infolge einer erhöhten Teilchengeschwindigkeit. Dieser kann hinter der Glottis nicht wieder in einen vollständigen Druckaufbau zurückgewandelt werden. Es wird angenommen, daß durch die hohen Teilchengeschwindigkeiten vor allem hinter der Glottis Turbulenzen entstehen. Diese können auch im Zusammenhang mit der Bildung eines Luftstrahls einhergehen, da angenommen wird, daß nicht der gesamte Glottisfluß in einen akustischen Anteil in den Vokaltrakt eingeht. Es kann sich noch ein nicht akustisch wirksamer Fluß als Strahl (jet) bilden, der nur einen Teil des Querschnitts im Vokaltrakt durchströmt. Diese Erkenntnisse stützen sich auf örtlich verteilte Messungen des Flusses im Vokaltrakt, welche erstmals von Teager [Te80] durchgeführt wurden. Da hier keine physikalische Modellierung der Stimmbänder erfolgt, werden die Effekte der Glottis nur teilweise berücksichtigt.

3.2 Rohrmodelle

Für Rohre deren Durchmesser nicht zu groß sind gegenüber den betrachteten Wellenlängen erfolgt die dominante Schallausbreitung in ebenen Wellen. Für die Ausbreitung ebener Wellen, welche hier eindimensional beschrieben wird, existieren mit (3.17) zwei Lösungen der Wellengleichung. Das Drucksignal p wird dargestellt als Überlagerung einer hinlaufenden Welle p^+ und einer rücklaufenden Welle p^-. Diese Signale sind vom Ort und der Zeit abhängig. In einem homogenen Rohr breitet sich eine ebene Welle in der Zeit t über die Strecke $s = ct$ aus. Zeit und Raum sind somit für die Wellenausbreitung durch die Schallgeschwindigkeit c fest verknüpft. Da im zeitdiskreten Fall nur Zeitpunkte zur Verfügung stehen, die Vielfache einer Zeiteinheit $\boldsymbol{\tau}$ sind, ist auch der Ort neben der Zeit quantisiert. Daher stehen im zeitdiskreten Rohrmodell nur die Wellengrößen an den Orten zur Verfügung, an denen die Welle eine Strecke von einer Einheitsrohrlänge bzw. deren Vielfachen zurück gelegt hat. Aus der Zeitdiskretisierung resultiert eine Ortsdiskretisierung. Ein Rohr wird zeitdiskret daher durch eine Verkettung von Einheitsrohrelementen gleicher Länge beschrieben, wie in Bild 3.1 zu sehen. Das Drucksignal p im k'ten Rohrelement besteht aus der vorwärtslaufen Welle p_k^+ und der rückwärtslaufenden Welle p_k^-. Die Wellen p_k^+ und p_k^- können zu einem Vektor

$$\boldsymbol{p}_k = \begin{pmatrix} p_k^+ \\ p_k^- \end{pmatrix} \tag{3.34}$$

zusammengefaßt werden. Da die Werte an beiden Enden eines Einheitsrohres benötigt werden, bezeichnet ein zweiter Index die Wellengröße am rechten Ende mit r und am linken Ende des Einheitsrohres mit l, wie in Bild 3.1 zu sehen ist. Die Vektoren $\boldsymbol{p}_k$ werden durch Matrizen verknüpft, die jeweils Querschnittssprünge und/oder Einheitsrohrelemente enthalten. Das zeitdiskrete Rohrmodell mit Verwendung in der Sprachverarbeitung wurde erstmals von Kelly und Lochbaum Anfang der sechziger Jahre vorgestellt [Kel62].

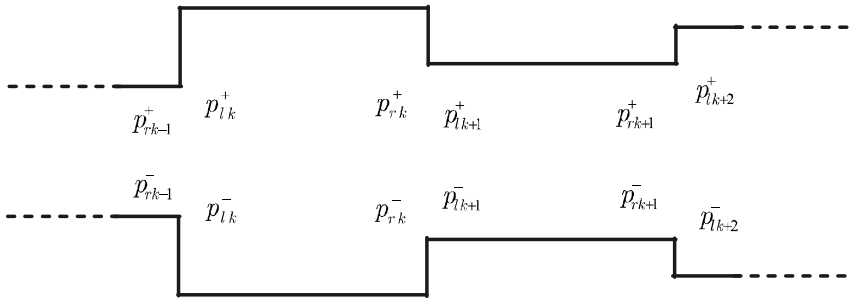


Bild 3.1: Zeitdiskretes Rohrmodell der ebenen Wellenausbreitung.

3.3 Elemente von Rohrsystemen

Die einfachste Beschreibung eines Einheitsrohrelementes ist ein Rohr mit homogenem Querschnitt und verlustloser Wellenausbreitung, wodurch die Wellensignale nur eine feste Zeitverzögerung erfahren. Die Verzögerung repräsentiert die Zeitdauer, in der die Welle durch ein Einheitsrohrelement propagiert ist. Im z-Bereich kann dies durch eine Multiplikation mit einer Potenz von z beschrieben werden. Die Z-Transformierten der Wellengrößen werden hier mit großen Buchstaben gekennzeichnet. Um in der weiteren Ableitung auf eine möglichst geringe Laufzeit pro Rohrelement zu kommen wird wie üblich eine Verzögerung von $z^{-\frac{1}{2}}$ für das Einheitsrohrstück verwendet. Damit ist die Verknüpfungsmatrix für das k'te Einheitsrohrelement im z-Bereich beschrieben mit

$$\begin{pmatrix} P^+_{kl} \\ P^-_{kl} \end{pmatrix} = \begin{pmatrix} z^{+\frac{1}{2}} & 0 \\ 0 & z^{-\frac{1}{2}} \end{pmatrix} \begin{pmatrix} P^+_{kr} \\ P^-_{kr} \end{pmatrix} = z^{+\frac{1}{2}} \begin{pmatrix} 1 & 0 \\ 0 & z^{-1} \end{pmatrix} \begin{pmatrix} P^+_{kr} \\ P^-_{kr} \end{pmatrix}. \tag{3.35}$$

Durch das Vorziehen des Faktor $z^{+\frac{1}{2}}$ aus der Matrix kommen in den Matrixelementen nur Terme z^{-1} vor; dies ist wichtig, da nur ganzzahlige Potenzen von z^{-1} im Zeitbereich realisiert werden können. Analog dazu kann auch der Faktor $z^{-\frac{1}{2}}$ vorgezogen werden. In einem Kettenfilter unterscheiden sich diese beiden Möglichkeiten durch einen Zustandsspeicher im oberen oder unteren Signalflußpfad des Filters [La96]. Auch reelle Verlustfaktoren mit $0 < \boldsymbol{\alpha} < 1$ können unmittelbar realisiert werden

$$\begin{pmatrix} P^+_{kl} \\ P^-_{kl} \end{pmatrix} = z^{+\frac{1}{2}} \begin{pmatrix} \frac{1}{\boldsymbol{\alpha}} & 0 \\ 0 & \boldsymbol{\alpha} z^{-1} \end{pmatrix} \begin{pmatrix} P^+_{kr} \\ P^-_{kr} \end{pmatrix} = \boldsymbol{\alpha}^{-1} z^{+\frac{1}{2}} \begin{pmatrix} 1 & 0 \\ 0 & \boldsymbol{\alpha}^2 z^{-1} \end{pmatrix} \begin{pmatrix} P^+_{kr} \\ P^-_{kr} \end{pmatrix}, \tag{3.36}$$

wobei die unterschiedlichen Ausbreitungsrichtungen der oberen und unteren Wellengrößen beachtet werden muß. Hierfür läßt sich auch ein Faktor separieren, so daß das linke obere Matrixelement unabhängig von dem Dämpfungsfaktor ist. Diese Beschreibung kann für eine inverse Filterung benutzt werden, da der Gesamtverstärkungsfaktor separiert ist, dessen Bedeutung für die Parameterbestimmung im nächsten Kapitel erläutert wird. In (3.36) wird angenommen, daß die Welle infolge der Ausbreitung durch das Einheitsrohrelement eine frequenzunabhängige Dämpfung erfährt. Die Einheitsrohrelemente werden durch Matrizen miteinander verknüpft. Da Rohrsegmente mit unterschiedlichen Impedanzen verknüpft werden können, bezeichnet man diese Verknüpfungsmatrizen auch als Adaptoren.

3.3.1 Adaptoren

Werden Rohrelemente mit unterschiedlichen Querschnittsflächen A verknüpft, so ändert sich die Impedanz infolge des Querschnittssprungs. Für die Schallausbreitung in Rohren ist die Verwendung des Schallflusses $u = Av$ statt der Schallschnelle zweckmäßig, da er für Strömungen in einem Rohr unabhängig von der Querschnittsfläche ist. Das Verhältnis zwischen Druck und Schallfluß wird als akustische Impedanz bezeichnet und wird hier auch mit Z gekennzeichnet. Die akustische Impedanz Z ebener Wellen in einem Rohr mit der Querschnittsfläche A ist somit nach (3.20) gegeben durch:

$$Z = \frac{p^+}{u^+} = \frac{p^-}{u^-} = \frac{Z_0}{A} = \frac{c\boldsymbol{\rho}_0}{A}. \tag{3.37}$$

Der Schallfluß ist wie die Schallschnelle eine vektorielle Größe von der nur die Beträge notiert werden, so daß für den Gesamtfluß $u = u^+ - u^-$ gilt. Für die Gebiete der Einheitsrohrelemente existieren Lösungen der Differentialgleichung, welche an den Grenzen zueinander stetig überführt werden müssen. An der Schnittstelle zwischen zwei Rohrsegmenten müssen daher Druck und Fluß stetig verlaufen. Eine Welle kann sich in einem Rohrsegment nur ausbreiten, wenn ein bestimmtes Verhältnis zwischen Druck und Fluß eingehalten wird. Im benachbarten Rohrsegment mit unterschiedlicher Impedanz breitet sich die Welle mit einem anderen Verhältnis aus. Daher kann eine Welle nicht ihre ganze Energie in das benachbarte Rohr übertragen, da sonst Druck bzw. Fluß unstetig verlaufen würden. Die nicht übertragbare Energie wird in das ursprüngliche Rohrsegment als reflektierte Welle zurückgestreut, wodurch die Stetigkeitsbedingungen an dem Querschnittssprung erfüllt werden können. Aus den Stetigkeitsbedingungen folgen für Druck und Fluß an der Schnittstelle zwischen zwei Rohrsegmenten der Stelle k:

$$p^+_{rk-1} + p^-_{rk-1} = p^+_{lk} + p^-_{lk} \tag{3.38}$$

$$u^+_{rk-1} - u^-_{rk-1} = u^+_{lk} - u^-_{lk}. \tag{3.39}$$

Mit Hilfe der akustischen Impedanz läßt sich in der zweiten Gleichung der Fluß eliminieren

$$\frac{1}{Z_{k-1}}\left(p^+_{rk-1} - p^-_{rk-1}\right) = \frac{1}{Z_{k+1}}\left(p^+_{lk} - p^-_{lk}\right). \tag{3.40}$$

Aus (3.38) und (3.40) ergibt sich durch Elimination von p^+_{k+1}die Beziehung

$$p^-_{rk-1} = \frac{Z_k - Z_{k-1}}{Z_k + Z_{k-1}} p^+_{rk-1} + \frac{2Z_{k-1}}{Z_k + Z_{k-1}} p^-_{lk}. \tag{3.41}$$

bzw. mit der reellen Rohrimpedanz (3.37)

$$p^-_{rk-1} = \frac{A_{k-1} - A_k}{A_{k-1} + A_k} p^+_{rk-1} + \frac{2A_k}{A_{k-1} + A_k} p^-_{lk}. \tag{3.42}$$

Die beiden Faktoren von p^+_{rk-1} und p^-_{lk} stellen jeweils die Anteile von p^-_{rk-1} dar, die am Querschnittssprung reflektiert und transmitiert werden. Daran ist zu erkennen, daß eine einfallende Welle p^+_k mit dem Reflexionsfaktor

$$r_k = \frac{Z_k - Z_{k-1}}{Z_k + Z_{k-1}} = \frac{A_{k-1} - A_k}{A_{k-1} + A_k} \tag{3.43}$$

infolge der Impedanzänderung zurückgestreut wird. Analog zu p^-_{rk-1} kann auch p^+_{lk} durch die anderen Torgrößen dargestellt werden, wodurch die Streumatrix $\boldsymbol{S}$ beschrieben wird:

$$\begin{pmatrix} P^-_{rk-1} \\ P^+_{lk} \end{pmatrix} = \begin{pmatrix} r_k & 1-r_k \\ 1+r_k & -r_k \end{pmatrix} \begin{pmatrix} P^+_{rk-1} \\ P^-_{lk} \end{pmatrix}. \tag{3.44}$$

Sie stellt die ausfallenden Torgrößen durch die einfallenden Größen dar. Werden die Torgrößen des einen Rohrsegements durch die Torgrößen des anderen Rohrsegements beschrieben, ergibt sich die Darstellung der Betriebskettenmatrix $\boldsymbol{T}$:

$$\begin{pmatrix} P^+_{rk-1} \\ P^-_{rk-1} \end{pmatrix} = \frac{1}{1+r_k} \begin{pmatrix} 1 & r_k \\ r_k & 1 \end{pmatrix} \begin{pmatrix} P^+_{lk} \\ P^-_{lk} \end{pmatrix}. \tag{3.45}$$

Der Vorteil dieser Matrixdarstellung besteht darin, daß Torgrößen, die über mehrere Zweitore getrennt sind, durch das Produkt der Betriebskettenmatrizen, welche die dazwischenliegenden Tore verknüpfen, mit einander in Beziehung gebracht werden können. Da Querschnittssprünge und Einheitsrohrelemente abwechselnd auftreten, werden beide in einer einzigen Betriebskettenmatrix $\boldsymbol{T}_k$ des k'ten Rohrelements zusammengefaßt:

$$\begin{pmatrix} P^+_{rk-1} \\ P^-_{rk-1} \end{pmatrix} = z^{\frac{1}{2}} \cdot d(r_k) \begin{pmatrix} 1 & r_k \cdot z^{-1} \\ r_k & z^{-1} \end{pmatrix} \begin{pmatrix} P^+_{rk} \\ P^-_{rk} \end{pmatrix} = z^{\frac{1}{2}} \boldsymbol{T}_k \begin{pmatrix} P^+_{rk} \\ P^-_{rk} \end{pmatrix}. \tag{3.46}$$

Der Faktor $z^{\frac{1}{2}}$ ist nicht in der Betriebskettenmatrix $\boldsymbol{T}_k$ integriert, wodurch $\boldsymbol{T}_k$ unmittelbar zeitdiskret realisiert werden kann. Durch die Zusammenfassung von Rohrelement und Querschnittssprung reichen die im Rohrstück rechts befindlichen Wellengrößen für eine Beschreibung aus, wodurch der zusätzliche Index r weggelassen werden kann

$$\begin{pmatrix} P^+_{k-1} \\ P^-_{k-1} \end{pmatrix} = z^{\frac{1}{2}} \boldsymbol{T}_k \begin{pmatrix} P^+_{k} \\ P^-_{k} \end{pmatrix}. \tag{3.47}$$

Für die verschiedenen Wellengrößen existieren unterschiedliche Betriebskettenmatrizen. Die Matrix $\boldsymbol{T}_k'$ stellt mit

$$\boldsymbol{T}_k = d(r_k)\boldsymbol{T}_k' = d(r_k) \begin{pmatrix} 1 & r_k \cdot z^{-1} \\ r_k & z^{-1} \end{pmatrix} \tag{3.48}$$

eine Grundform dar, durch welche sich mit dem Vorfaktor $d(r)$ die Betriebskettenmatrizen der unterschiedlichen Wellendarstellungen ergeben. Eine Verkettung von N Querschnittssprüngen mit den dazugehörigen Rohrelementen kann durch eine Multiplikation der Matrizen $\boldsymbol{T}_k$ ausgedrückt werden

$$\begin{aligned} \begin{pmatrix} P^+_0 \\ P^-_0 \end{pmatrix} &= z^{\frac{N}{2}} \prod_{k=1}^{N} d(r_k) \begin{pmatrix} 1 & r_k \cdot z^{-1} \\ r_k & z^{-1} \end{pmatrix} \begin{pmatrix} P^+_N \\ P^-_N \end{pmatrix} \\ &= z^{\frac{N}{2}} \prod_{k=1}^{N} \boldsymbol{T}_k \begin{pmatrix} P^+_N \\ P^-_N \end{pmatrix}. \end{aligned} \tag{3.49}$$

Der Faktor $z^{\frac{N}{2}}$ wird in den folgenden Berechnungen weggelassen, da er nur eine Gesamtverzögerung darstellt, so daß die vereinfachte Beziehung

$$\begin{pmatrix} \widetilde{P}^+_0 \\ \widetilde{P}^-_0 \end{pmatrix} = \prod_{k=1}^{N} \boldsymbol{T}_k \begin{pmatrix} \widetilde{P}^+_N \\ \widetilde{P}^-_N \end{pmatrix} \tag{3.50}$$

besteht. Die Größen $\widetilde{P}$ und P sind für den Fall identisch, falls die Rohrstücke zwischen den Größen gleich viele Zustandsspeicher oben und unten im Signalflußpfad besitzen. $\widetilde{P}^+$ und $\widetilde{P}^-$ werden ab jetzt ebenfalls mit P^+ und P^- bezeichnet, um die Schreibweise zu vereinfachen

$$\begin{pmatrix} P_0^+ \\ P_0^- \end{pmatrix} = \prod_{k=1}^{N} \boldsymbol{T}_k \begin{pmatrix} P_N^+ \\ P_N^- \end{pmatrix}. \tag{3.51}$$

Durch Herausziehen des Verzögerungsfaktor $z^{\frac{N}{2}}$ ist es möglich, Einheitsrohrlängen zu verwenden, die sonst nur mit der doppelten Abtastrate zu realisieren wären. Eine weitere Kürzung der Einheitsrohrlängen ist auf direktem Wege nur durch Änderung der Abtastrate möglich. Es ist immer zu prüfen, ob für ein bestimmtes Rohrsystem der weggelassene Verzögerungsfaktor relevant ist. Dies trifft zum Beispiel für den Fall einer Ringstruktur zu, der in einem nachfolgenden Abschnitt behandelt wird. In der Produktmatrix $\prod \boldsymbol{T}_k$ sind die diagonal liegenden Elemente von einander abhängig, da die Polynome invers zueinander sind, was einem umgedrehten Polynomkoeffizientenvektor entspricht. Diese Relationen können durch vollständige Induktion bewiesen werden [Sn96].

In den Gleichungen (3.38) und (3.39) kann auch der Druck eliminiert werden, woraus sich die Adaptoren für Flußwellen ergeben. Sie unterscheiden sich von den Druckwellenadaptoren durch den Vorfaktor $d(r)$; vgl. [La96]. Neben Druck- und Flußwellen sind Leistungswellen l [Kub85] gebräuchlich, welche definiert sind durch

$$l = \sqrt{p \cdot u} \quad \text{bzw.} \quad l = p\sqrt{\frac{A}{c\boldsymbol{\rho}_0}}. \tag{3.52}$$

Wird die rechte Beziehung von (3.52) in (3.38) und (3.40) substituiert, so resultiert der Zweitor-Adaptor für Leistungswellen. Bei Leistungswellen ist zu beachten, daß das Quadrat der Wellengrößen die Leistung der Welle wiedergibt. Die Betriebskettenmatrizen des Querschnittsprungs, welche hier zusammen mit einem Rohrelement dargestellt werden, unterscheiden sich für die verschiedenen Wellengrößen durch den Faktor $d(r)$ in (3.48), welcher in Tabelle 3 aufgelistet ist; zusätzlich sind die Streumatrizen dargestellt ohne Berücksichtigung der Zustandsspeicher bzw. Rohrelemente. Neben diesen physikalisch motivierten Wellendarstellungen wird durch den Vorfaktor $d(r) = 1$ eine Grundform definiert, deren Wellengröße keiner physikalischen Größe unmittelbar entspricht und für die $\boldsymbol{T}_k = \boldsymbol{T}_k'$ gilt. Durch die akustische Impedanz $Z = c\boldsymbol{\rho}_0/A$ sind Druck und Flußwellen miteinander verknüpft, wodurch die Leistung der Welle mit nur einer Größe berechnet werden kann. Daraus ergibt sich die Leistung in Abhängigkeit von Druck oder Schallfluß zu

$$p \cdot u = \frac{c\boldsymbol{\rho}_0}{A} u^2 = \frac{A}{c\boldsymbol{\rho}_0} p^2. \tag{3.53}$$

Durch Verwendung von nur einer Wellengröße hängt die Leistung der Welle von der Querschnittsfläche ab. Dadurch bewirkt eine Änderung der Fläche A bei einem Signal p oder u eine Leistungsänderung der Welle nach (3.53).

Tabelle 3: Betriebskettenmatrizen und Streumatrizen.

	Betriebskettenmatrix $\boldsymbol{T}_k = d(r) \cdot \boldsymbol{T}_k'$	Streumatrix der Adaptoren
Grundform	$\boldsymbol{T}_k'$	$\begin{pmatrix} r & 1-r^2 \\ 1 & -r \end{pmatrix}$
Druckwellen	$\frac{1}{1+r} \cdot \boldsymbol{T}_k'$	$\begin{pmatrix} r & 1-r \\ 1+r & -r \end{pmatrix}$
Flußwellen	$\frac{1}{1-r} \cdot \boldsymbol{T}_k'$	$\begin{pmatrix} r & 1+r \\ 1-r & -r \end{pmatrix}$
Leistungswellen	$\frac{1}{\sqrt{1+r^2}} \cdot \boldsymbol{T}_k'$	$\begin{pmatrix} r & \sqrt{1-r^2} \\ \sqrt{1-r^2} & -r \end{pmatrix}$

Bei zeitvariablen Flächen kann durch eine zusätzliche Korrektur die Energieerhaltung erfüllt werden. Verändert sich die Fläche von A zu $A' = A + \triangle A$ so werden die neuen Druckgrößen p' und Flußgrößen u' bei Energieerhaltung beschrieben mit

$$p' = \sqrt{\frac{A}{A'}}p \quad \text{und} \quad u' = \sqrt{\frac{A'}{A}}u. \tag{3.54}$$

Somit gilt

$$\frac{c\boldsymbol{\rho}_0}{A}u^2 = \frac{c\boldsymbol{\rho}_0}{A'}u'^2 \quad \text{und} \quad \frac{A}{c\boldsymbol{\rho}_0}p^2 = \frac{A'}{c\boldsymbol{\rho}_0}p'^2 \tag{3.55}$$

für die Leistungen der Wellen nach (3.53). Da die Energie der Leistungswellen $l = \sqrt{p \cdot u}$ unabhängig von der Fläche A ist, ist für sie keine Korrektur bei zeitvariablen Flächen erforderlich. Für die Korrekturen von Druck und Fluß nach (3.54) müssen die vorwärts und rückwärts gerichteten Wellenanteile eines Rohrelementes bekannt sein, die bei Rohrkettenfiltern mit einem Laufzeitglied pro Rohrelement nicht direkt zur Verfügung stehen. Dies ist explizit nur bei Kettenfiltern mit Laufzeitgliedern oben und unten gegeben, welche allerdings eine doppelt so große Rohrlänge aufweisen. Die Überlegungen des Zustandsspeicherproblems bei zeitvariablen Rohrmodellen gehen hier auf Eichler und Lacroix zurück. Eichler hat in [EiL96, Ei96] durch Rechnersimulationen mittels Kettenfilter mit zwei Zustandsspeichern pro Rohrelement gezeigt hat, daß das Verhalten der zeitvariablen Filter mit Leistungswellen identisch ist mit dem Verhalten von zeitvariablen Filtern mit korrigierten Druck- und Flußwellen. Dieses Verhalten ist konform mit der Theorie. Bei schnellen und starken Flächenänderungen können dadurch störende Einschwingvorgänge vermieden werden. Da hier allerdings nur Rohrelemente mit einem Laufzeitglied verwendet werden, sind Leistungswellen durch ihre Energieerhaltung vorteilhaft. Da die Vokaltraktflächen sich nur verhältnismäßig langsam ändern, ist eine Korrektur auch in der Druck- und Flußdarstellung nicht unbedingt erforderlich. Dies gilt nicht für die Modellierung der Glottis. Dafür hat Strube unter anderem in [Str82] einen zeitvariablen Adaptor vorgestellt, der unter der Annahme hergeleitet ist, daß zeitlich vor und nach der Wellenreflexion eines kurzzeitigen Wellensignals unterschiedliche konstante Impedanzen bzw. Rohrquerschnitte vorherrschen.

Der Adaptor kann dann durch die Stetigkeitsbedingungen von Druck und Fluß hergeleitet werden. Es exitieren noch weitere Modifikationen des Rohrmodells. In [Vä94a] sind auch konische Rohrelemente vorgeschlagen und in [Vä94b, Vä00, Str75b, Str00] sind Rohrelemente mit variabler Länge diskutiert. In [L78, L79] wird eine Modellierung von Sprachsignalen mit einem Rohrmodell mit weniger Reflexionskoeffizienten als Rohrelemente vorgeschlagen, da einige Koeffizienten recht klein sind und kaum zur Modellierung beitragen, womit sie weggelassen werden können.

Dreitor Adaptoren

Analog zu den Zweitor-Adaptoren können auch Mehrtor-Adaptoren hergeleitet werden. Sie folgen aus den Stetigkeitsbedingungen für Druck und Schallfluß an den Toren. Insbesondere Dreitor-Adaptoren sind von Interesse, da mit ihnen einfache Rohrverzweigungen realisiert werden können. Werden die Drücke und Flüsse entsprechend

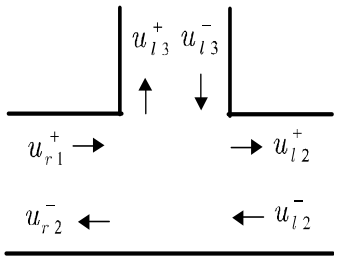


Bild 3.2: Rohrverzweigung mit Schallfl üssen.

den Nummern der Tore von eins bis drei indiziert, so ergeben sich die Bedingungsgleichungen:

$$\begin{aligned} p_{r1}^{+} + p_{r1}^{-} = p_{l2}^{+} + p_{l2}^{-} = p_{l3}^{+} + p_{l3}^{-} \\ u_{r1}^{+} + u_{l2}^{-} + u_{l3}^{-} = u_{r1}^{-} + u_{l2}^{+} + u_{l3}^{+}. \end{aligned} \tag{3.56}$$

Hierbei sind entsprechend Bild 3.2 die Summe der einströmenden Flüsse betragsmäßig gleich der Summe der ausströmenden Flüsse. Der Schallfluß kann mittels der Impedanz in den Druck umgerechnet werden:

$$A_1 \left(p_{r1}^{+} - p_{r1}^{-}\right) = A_2 \left(p_{l2}^{+} - p_{l2}^{-}\right) + A_3 \left(p_{l3}^{+} - p_{l3}^{-}\right). \tag{3.57}$$

Die Streumatrix $\boldsymbol{S}$ des Dreitores beschreibt die drei ausfallenden Größen durch die einfallenden Größen:

$$\begin{pmatrix} p_{r1}^{-} \\ p_{l2}^{+} \\ p_{l3}^{+} \end{pmatrix} = \boldsymbol{S} \begin{pmatrix} p_{r1}^{+} \\ p_{l2}^{-} \\ p_{l3}^{-} \end{pmatrix}, \tag{3.58}$$

welche mit den drei Gleichungen aus (3.56) und (3.57) ermittelt werden kann und in Tabelle 4 gezeigt ist. Durch die Substitutionen

$$p_i = u_i \frac{c\boldsymbol{\rho}_0}{A_i} \quad \text{und} \quad p_i = l_i \sqrt{\frac{c\boldsymbol{\rho}_0}{A_i}} \tag{3.59}$$

können aus der Streumatrix für Druckwellen die Streumatrizen für Fluß- und Leistungswellen erhalten werden, welche in Tabelle 4 ebenfalls gezeigt sind. Die Streumatrizen

besitzen den gemeinsamen Vorfaktor $\boldsymbol{\lambda} = \frac{1}{A_1+A_2+A_3}$. Mittels der Substitutionen (3.59) können auch aus der Betriebskettenmatrix für Druckwellen die Matrizen $\boldsymbol{T}_k$ der anderen physikalischen Wellenformen hergeleitet werden, wofür der Reflexionskoeffizient in Abhängigkeit der Flächen ausgedrückt werden muß (siehe Tabelle 3).

Tabelle 4: Streumatrizen der Dreitor-Adaptoren.

	Streumatrix $\boldsymbol{S}$
Druckwellen	$\boldsymbol{\lambda} \cdot \begin{pmatrix} A_1 - A_2 - A_3 & 2A_2 & 2A_3 \\ 2A_1 & A_2 - A_1 - A_3 & 2A_3 \\ 2A_1 & 2A_2 & A_3 - A_1 - A_2 \end{pmatrix}$
Flußwellen	$\boldsymbol{\lambda} \cdot \begin{pmatrix} A_1 - A_2 - A_3 & 2A_1 & 2A_1 \\ 2A_2 & A_2 - A_1 - A_3 & 2A_2 \\ 2A_3 & 2A_3 & A_3 - A_1 - A_2 \end{pmatrix}$
Leistungsw.	$\boldsymbol{\lambda} \cdot \begin{pmatrix} A_1 - A_2 - A_3 & 2\sqrt{A_1A_2} & 2\sqrt{A_1A_3} \\ 2\sqrt{A_1A_2} & A_2 - A_1 - A_3 & 2\sqrt{A_2A_3} \\ 2\sqrt{A_1A_3} & 2\sqrt{A_2A_3} & A_3 - A_1 - A_2 \end{pmatrix}$

Für die Beschreibung des Dreitor-Adaptors sind analog zum Zweitor-Adaptor nur die relativen Flächenverhältnisse von Bedeutung, so daß statt der drei Flächen A_i zwei Parameter $\boldsymbol{\rho}_i$ ausreichen, um das Verhalten des Dreitores eindeutig zu bestimmen. Eine gebräuchliche Definition der zweiparametrigen Beschreibung ist mittels

$$\boldsymbol{\rho}_i = \frac{2A_i}{A_1 + A_2 + A_3} \qquad i = 1, 2\ (, 3) \tag{3.60}$$

gegeben, wobei nur $\boldsymbol{\rho}_1$ und $\boldsymbol{\rho}_2$ verwendet werden, da $\boldsymbol{\rho}_3$ durch die beiden anderen Parameter mit

$$\boldsymbol{\rho}_3 = 2 - \boldsymbol{\rho}_1 - \boldsymbol{\rho}_2 \tag{3.61}$$

ausgedrückt werden kann. Die Streumatrix des zweiparametrigen Druckwellenadaptors gestaltet sich somit zu:

$$\boldsymbol{S} = \begin{pmatrix} \boldsymbol{\rho}_1 - 1 & \boldsymbol{\rho}_2 & 2 - \boldsymbol{\rho}_1 - \boldsymbol{\rho}_2 \\ \boldsymbol{\rho}_1 & \boldsymbol{\rho}_2 - 1 & 2 - \boldsymbol{\rho}_1 - \boldsymbol{\rho}_2 \\ \boldsymbol{\rho}_1 & \boldsymbol{\rho}_2 & 1 - \boldsymbol{\rho}_1 - \boldsymbol{\rho}_2 \end{pmatrix}. \tag{3.62}$$

Der Dreitor-Adaptor für akustische Druckwellen entspricht dem Dreitor-Paralleladaptor für die von Fettweis eingeführten Wellendigitalfilter, welcher eine Parallelschaltung in einem elektrischen Netzwerk ermöglicht [Sü92].

Rohrabschlüsse

Am Ende des Rohres wird ein Abschluß benötigt, der den Impedanzsprung vom Rohrende zum Außenraum beschreibt. Infolge des Impedanzsprunges werden die mit dem Abschlußkoeffizienten gewichteten Anteile der auslaufenden Wellen in das Rohr zurück reflektiert. Die Möglichkeit von einfallenden Wellen, welche vom Außenraum in das

Rohr hinein gelangen, wird hier nicht betrachtet. Zwei Grenzfälle für den Abschlußkoeffizienten sind mit ± 1 gegeben, wobei $+1$ einen schallharten Rohrabschluß und -1 einen schallweichen Abschluß beschreibt. Der schallweiche Abschluß entspricht einem angehängten Zweitor-Adaptor mit einem Reflexionsfaktor von -1, bei dem der rechte untere Eingang unbeachtet bleibt. Daher beschreibt ein schallweicher Abschluß von -1 einen idealisierten Übergang von einer endlichen Fläche zu einer unendlich großen Querschnittsfläche, wodurch dieser Rohrabschluß auch an den Lippen oder and den Nasenlöchern als Näherung verwendet wird. Dieser Betrachtung liegt eine Ausbreitung ebener Wellen zugrunde. Die aus dem Rohr austretenden ebenen Wellen gehen in einem größeren Außenraum näherungsweise in Kugelwellen über. Dieser Übergang findet verstärkt für hohe Frequenzen statt. Die Impedanz für Kugelwellen mit dem Radius r und der Wellenzahl k ist

$$Z_K = \frac{p}{v} = c\boldsymbol{\rho}_0 \frac{ikr}{1+ikr}, \tag{3.63}$$

welche im Gegensatz zur Impedanz einer ebenen Welle komplex und frequenzabhängig ist. Die Impedanz (3.63) führt zur Abstrahlimpedanz einer atmenden Kugel. Die Lippenabstrahlung kann dadurch nur näherungsweise beschrieben werden. Für kleine Lippenöffnungen ist die Beschreibung eines Kolbenstrahlers in einer unendlich ausgedehnten Wand eine recht gute Näherung. Eine bessere Approximation für alle Öffnungsflächen an den Lippen liefert ein Kolbenstrahler in einer Kugel, dessen mathematische Behandlung recht umfangreich ist. Für kleine Kugeln nähert sich dieser der atmenden Kugel an, während er sich für sehr große Kugeln dem Verhalten eines Kolbenstrahlers in einer unendlich ausgedehnten Wand annähert [Fl72]. Laine hat mehrere Modelle im Z-Bereich vorgestellt, welche die Kolbenstrahlerimpedanz in einer Kugel modellieren [Lai82]. Die Impedanz ist von der Öffnungsfläche A in cm^2 abhängig. Die Modellparameter wurden durch einen Optimierungsalgorithmus ermittelt. Das hier verwendete Modell von Laine ist das Pol-Nullstellen System Z_{pz} für die normalisierte Strahlungsimpedanz, welche definiert ist durch

$$Z_{pz}(z) = \frac{a(1-z^{-1})}{1-bz^{-1}} \tag{3.64}$$

mit

$$a = 0,0779 + 0,2373\sqrt{A} \quad \text{und} \quad b = 0,843 - 0,3062\sqrt{A}. \tag{3.65}$$

Diese Formel ist nicht für verschiedene Abtastraten parametrisiert. In [Kr98] sind weitere Parameter a und b für das Modell (3.64) von Laine angegeben, welche für eine Abtastrate um 20 kHz optimiert sind. Der Lippenabschluß ist mit (3.43) und der Berücksichtigung, daß es sich bei Z_{pz} um eine normalisierte akustische Impedanz handelt, gegeben durch:

$$R_{lip}(z) = \frac{Z_{pz}-1}{Z_{pz}+1} = \frac{(a-1)-(a-b)z^{-1}}{(a+1)-(a+b)z^{-1}}. \tag{3.66}$$

In Bild 3.3 ist der Betrag und die Phase bzw. das Argument vom Lippenabschlußkoeffizienten $R_{lip}(z)$ gezeigt. Es ist zu erkennen, daß für kleine Öffnungsflächen und sehr tiefe Frequenzen der Abschluß gegen -1 strebt.

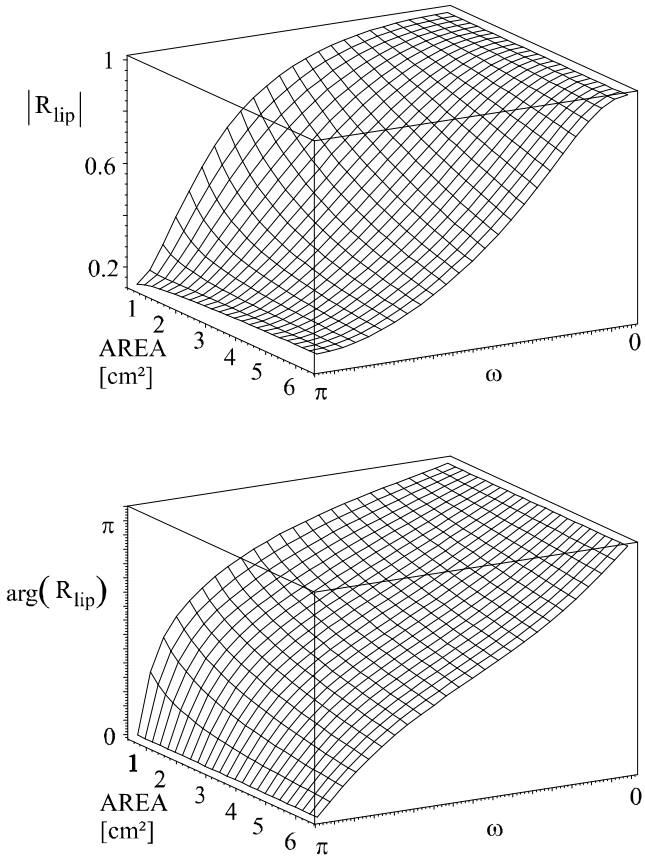


Bild 3.3: Lippenabschluß mit dem Impedanz-Modell nach Laine.

3.3.2 Übertragungsfunktion von Rohrsystemen

Mit den beschriebenen Rohrelementen, Adaptoren und Abschlüssen können unterschiedliche Rohrsysteme erstellt werden. Adaptoren verknüpfen dabei jeweils die Rohrelemente und beschreiben infolge der Impedanzanpassungen Querschnittssprünge. Durch Dreitor-Adaptoren ist die Möglichkeit von Rohrverzweigungen gegeben. Für die Beschreibung eines solchen Rohrmodells als zeitdiskretes System muß der Ein- und Ausgang festgelegt werden. Die Übertragungsfunktion zwischen dem Ein- und Ausgang des Rohrsystems muß individuell bestimmt werden. Deshalb wird eine allgemeine Vorgehensweise vorgestellt, die eine algebraische Berechnung der Übertragungsfunktion von Rohrsystemen ermöglicht. Die Vorgehensweise wird durch einen Algorithmus definiert, durch den die Übertragungsfunktion vieler Rohrstrukturen zur Laufzeit eines Programms berechnet werden kann. Die Übertragungsfunktion $H(z)$ des zeitinvarianten Rohrmodells ist durch das Zähler- und Nennerpolynom beschrieben, welche auch für die Parameterbestimmung des Modells verwendet werden. Zuerst werden einfache Rohrsysteme behandelt, in denen keine Verzweigungen auftreten.

Unverzweigtes Rohrsystem

Der Fall eines unverzweigten Rohrsystems wird zuerst behandelt, in dem sich der Systemausgang am Rohrende befindet, das durch ein System:

$$C(z) = \frac{C_n(z)}{C_d(z)} \tag{3.67}$$

mit beliebigem Zählerpolynom $C_n(z)$ und Nennerpolynom $C_d(z)$ abgeschlossen ist, wie in Bild 3.4 zu sehen ist. Die Größe x stellt eine beliebige Wellengröße dar und wird durch die verwendeten Adaptoren festgelegt. Mit C kann ein frequenzabhängiger Lippenabschluß des Vokaltraktes beschrieben werden. Der Systemeingang befindet sich am Rohranfang, welcher hier einen reflexionsfreien Rohrabschluß besitzt. Das Rohr selbst habe eine Länge von N Rohrelementen. Die Zustandsspeicher des Rohrmodells

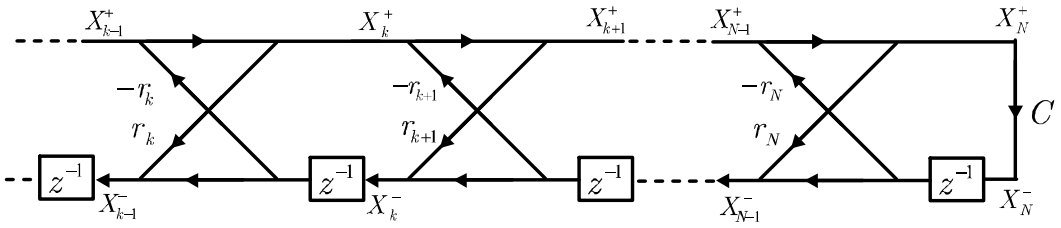


Bild 3.4: Ausschnitt aus dem Signalflußpfad des Rohrmodells für $\boldsymbol{T}_k = \boldsymbol{T}_k'$.

befinden sich in dem Beispiel von Bild 3.4 im unteren Signalflußpfad. Das Rohrmodell von Bild 3.4 kann mittels der Betriebskettenmatrizen beschrieben werden, wie in Bild 3.5 gezeigt. Das Rohr zwischen Systemeingang und -ausgang wird durch das Produkt

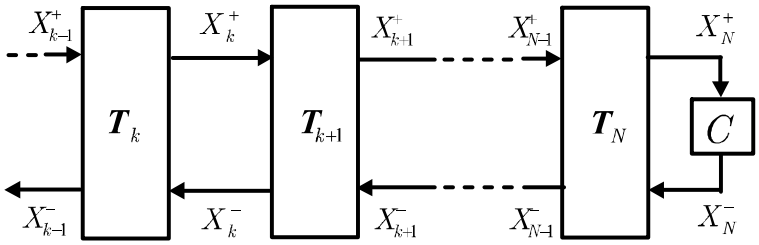


Bild 3.5: Rohrmodell in der Darstellung der Betriebskettenmatrizen.

$\boldsymbol{T}_g$ der N Betriebskettenmatrizen $\boldsymbol{T}_k$ beschrieben, welche jeweils das k'te Rohrelement mit dem dazugehörigen Adaptor darstellen:

$$\boldsymbol{T}_g = \begin{pmatrix} T_g^{11} & T_g^{12} \\ T_g^{21} & T_g^{22} \end{pmatrix} = \prod_{k=1}^{N} \boldsymbol{T}_k. \tag{3.68}$$

Die Beziehungen zwischen dem Systemeingang und -ausgang können aus den Gleichungen

$$\boldsymbol{X}_0 = \boldsymbol{T}_g \, \boldsymbol{X}_N \quad \text{und} \quad X_N^- = C \cdot X_N^+ \tag{3.69}$$

ermittelt werden, wobei $\boldsymbol{X}$ entsprechend (3.34) einen Vektor aus vor- und rücklaufenden Wellen darstellt. Daraus ergibt sich die Übertragungsfunktion

$$H(z) = \frac{X_N^+}{X_0^+} = \frac{1}{T_g^{11} + CT_g^{12}} = \frac{C_d}{C_d T_g^{11} + C_n T_g^{12}}, \tag{3.70}$$

bei der Zähler- und Nennerpolynom explizit dargestellt sind. Falls der Rohrabschluß ein nicht rekursives Teilsystem darstellt, besitzt die Übertragungsfunktion nur Pole. Für das Synthesefilter der LPC-Analyse bzw. der Burg-Methode weist der Abschluß die Form $C = \pm 1$ auf, woraus sich die Übertragungsfunktion

$$H(z) = 1/(T_g^{11} \pm T_g^{12}) \tag{3.71}$$

ergibt. Bei einem reellen Abschluß C kann dieser auch durch eine zusätzliche Betriebskettenmatrix $\boldsymbol{T}_{N+1}$ beschrieben werden, so daß $H(z) = 1/T_g^{11}$ gilt. Es existieren noch weitere Fälle von unverzweigten Rohrstrukturen, die hier als Spezialfall bei der Behandlung der verzweigten Rohrsysteme auftreten.

Einfach verzweigtes Rohrsystem

Für eine Verzweigung des Rohres wird die Streumatrix (3.62) des Dreitor-Paralleladaptor verwendet. Im Falle des unverzweigten Rohres kann die Übertragungsfunktion $H(z)$ durch die Verwendung der 2×2 Betriebskettenmatrizen berechnet werden. Daher wird die Rohrverzweigung auch durch eine 2×2 Betriebskettenmatrix $\boldsymbol{T}_D'$ dargestellt. Der angekoppelte Seitenzweig ist in der Matrix $\boldsymbol{T}_D'$ integriert und darf keinen Ein- oder Ausgang des Systems enthalten. Wie in Bild 3.6 zu sehen ist, stellt $\boldsymbol{T}_D'$ den Dreitor-Adaptor mit dem Seitenzweig $\widetilde{H}$ dar. Der Seitenzweig wird durch die Teilübertragungsfunktion

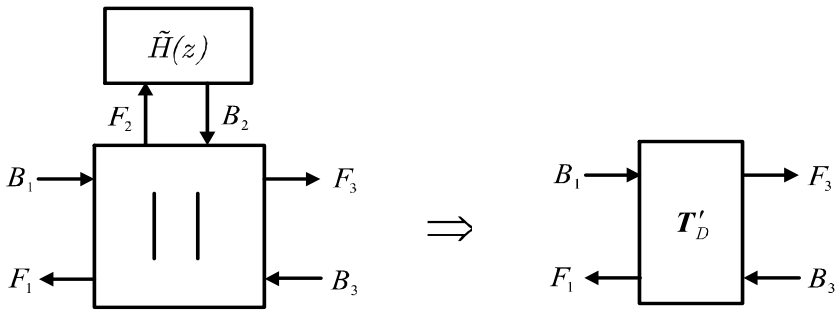


Bild 3.6: Darstellung des Dreitors mit Seitenzweig als Zweitor.

$$\widetilde{H}(z) = \frac{Q(z)}{P(z)} \tag{3.72}$$

allgemein beschrieben und kann selbst wieder Verzweigungen aufweisen. Q und P sind das Zähler- und Nennerpolynom von $\widetilde{H}$, welche durch den Seitenzweig bestimmt sind. $\widetilde{H}(z)$ beschreibt die Filterung des Ausgangssignals f_2 des Dreitores zum Eingang b_2 des Dreitores infolge des Seitenzweiges. Mit

$$\begin{pmatrix} F_1 \\ F_2 \\ F_3 \end{pmatrix} = \boldsymbol{S} \begin{pmatrix} B_1 \\ B_2 \\ B_3 \end{pmatrix} \quad \text{und} \quad B_2 = \widetilde{H}(z) F_2 \tag{3.73}$$

stehen vier Gleichungen zur Verfügung. Durch Eliminieren von zwei Gleichungen kann die gewünschte Struktur der Betriebskettenmatrix

$$\begin{pmatrix} B_1 \\ F_1 \end{pmatrix} = \boldsymbol{T}_D' \begin{pmatrix} F_3 \\ B_3 \end{pmatrix} \tag{3.74}$$

erhalten werden. $\boldsymbol{T}_D'$ kann durch $\widetilde{H}(z)$ als allgemeine Formel [SnL98a] angegeben werden:

$$\boldsymbol{T}_D' = \begin{pmatrix} \{\boldsymbol{\rho}_1 + \boldsymbol{\rho}_2 - 1 \mid 1\} & \{\boldsymbol{\rho}_2 - 1 \mid 1 - \boldsymbol{\rho}_1\} \\ \{1 - \boldsymbol{\rho}_1 \mid \boldsymbol{\rho}_2 - 1\} & \{1 \mid \boldsymbol{\rho}_1 + \boldsymbol{\rho}_2 - 1\} \end{pmatrix} \tag{3.75}$$

$$\text{mit der Abkürzung: } \{\boldsymbol{\nu} \mid \boldsymbol{\mu}\} := \frac{\boldsymbol{\nu} \cdot Q + \boldsymbol{\mu} \cdot P}{\boldsymbol{\rho}_2 (Q + P)} \quad .$$

Mit (3.75) ist zu sehen, daß alle Matrixelemente das gleiche Nennerpolynom $\boldsymbol{\rho}_2(Q+P)$ aufweisen. Dieser Nenner wird aus der Matrix als Vorfaktor herausgezogen, so daß die Matrixelemente nur Zählerpolynome aufweisen. Für die Berechnung der Übertragungsfunktion werden die $\boldsymbol{T}$-Matrizen miteinander multipliziert. Die Betriebskettenmatrizen (3.48) stellen jeweils einen Adaptor und ein Rohrelement dar, welche hier durch einen Zustandspeicher unten repräsentiert sind. Daher wird auch für $\boldsymbol{T}_D'$ an der rechten Seite ein Rohrelement berücksichtigt. Die Betriebskettenmatrix $\boldsymbol{T}_D$ beschreibt einen Dreitor-Paralleladaptor mit Seitenzweig und einem zusätzlichen Rohrelement:

$$\boldsymbol{T}_D = \frac{1}{\boldsymbol{\rho}_2 (Q + P)} \begin{pmatrix} \{\boldsymbol{\rho}_1 + \boldsymbol{\rho}_2 - 1, 1\} & \{\boldsymbol{\rho}_2 - 1, 1 - \boldsymbol{\rho}_1\} \cdot z^{-1} \\ \{1 - \boldsymbol{\rho}_1, \boldsymbol{\rho}_2 - 1\} & \{1, \boldsymbol{\rho}_1 + \boldsymbol{\rho}_2 - 1\} \cdot z^{-1} \end{pmatrix} \tag{3.76}$$

$$\text{mit der Abkürzung: } \{\boldsymbol{\nu}, \boldsymbol{\mu}\} := \boldsymbol{\nu} \cdot Q + \boldsymbol{\mu} \cdot P.$$

Für die Multiplikation mit anderen $\boldsymbol{T}$-Matrizen kann das vorgezogene Nennerpolynom dabei mit anderen vorangestellten Nennerpolynomen multipliziert werden.

Verzweigtes Rohrsystem als Ringstruktur

Neben dem Aufspalten der Rohrstruktur durch Verzweigungen ist es möglich, daß die aufgespalteten Zweige wieder durch einen weiteren Dreitor verbunden werden, wodurch eine Ringstruktur entsteht; vgl. Bild 3.7. Die Übertragungsfunktion kann hierfür nicht mit Hilfe von $\boldsymbol{T}_D$ bestimmt werden, da mit der Wiedervereinigung der Pfade von Tor zwei und drei die Größen B_3 und F_3 mit den Größen B_2 und F_2 in (3.73) gegenseitig abhängig sind, wie in Bild 3.7 zu sehen ist. Deshalb wird die komplette Ringstruktur durch eine 2×2 Betriebskettenmatrix $\boldsymbol{T}_R'$ beschrieben [SnL00a] (siehe Bild 3.7) mit:

$$\begin{pmatrix} B_1 \\ F_1 \end{pmatrix} = \boldsymbol{T}_R' \begin{pmatrix} C_3 \\ D_3 \end{pmatrix}. \tag{3.77}$$

Die beiden Pfade zwischen den Dreitoren werden allgemein durch die Matrizen $\boldsymbol{G}$ und $\boldsymbol{H}$ dargestellt. Diese Matrizen $\boldsymbol{G}$ und $\boldsymbol{H}$ verknüpfen die entsprechenden Torgrößen der

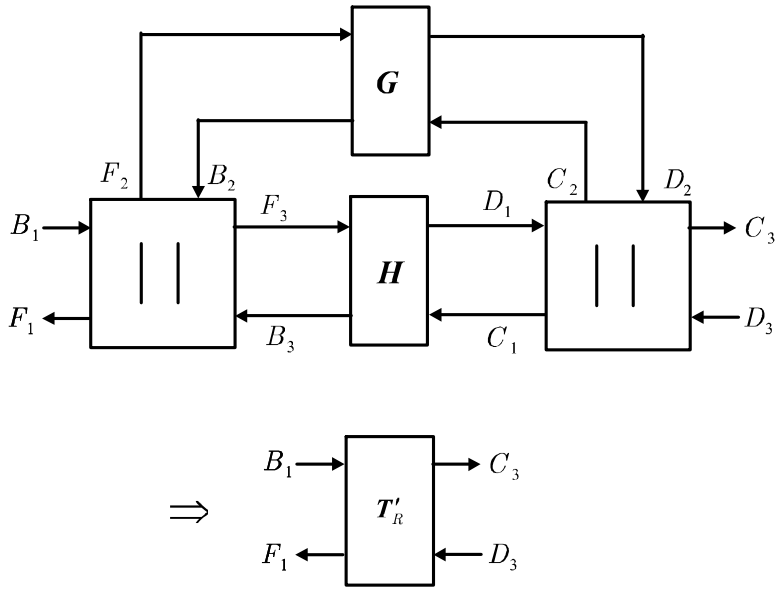


Bild 3.7: Ringstruktur als 2×2 Betriebskettenmatrix.

Dreitore miteinander durch

$$\begin{pmatrix} F_2 \\ B_2 \end{pmatrix} = \begin{pmatrix} G_{11} & G_{12} \\ G_{21} & G_{22} \end{pmatrix} \begin{pmatrix} D_2 \\ C_2 \end{pmatrix} \tag{3.78}$$

und

$$\begin{pmatrix} F_3 \\ B_3 \end{pmatrix} = \begin{pmatrix} H_{11} & H_{12} \\ H_{21} & H_{22} \end{pmatrix} \begin{pmatrix} D_1 \\ C_1 \end{pmatrix}.$$

Mit den Streumatrizen der beiden Dreitor-Paralleladaptoren:

$$\begin{pmatrix} F_1 \\ F_2 \\ F_3 \end{pmatrix} = \begin{pmatrix} u-1 & v & 2-u-v \\ u & v-1 & 2-u-v \\ u & v & 1-u-v \end{pmatrix} \begin{pmatrix} B_1 \\ B_2 \\ B_3 \end{pmatrix} \tag{3.79}$$

und

$$\begin{pmatrix} C_1 \\ C_2 \\ C_3 \end{pmatrix} = \begin{pmatrix} r-1 & s & 2-r-s \\ r & s-1 & 2-r-s \\ r & s & 1-r-s \end{pmatrix} \begin{pmatrix} D_1 \\ D_2 \\ D_3 \end{pmatrix}$$

stehen mit (3.78) zusammen zehn Gleichungen zur Verfügung, aus denen $\boldsymbol{T}'_R$ ermittelt werden kann. $\boldsymbol{T}'_R$ verknüpft die äußeren Signale B_1, F_1, C_3 und D_3 des Ringes miteinander. Die übrigen acht inneren Signale müssen unter Zuhilfenahme der zehn Gleichungen eliminiert werden, so daß sich zwei Gleichungen für die Beschreibung von $\boldsymbol{T}'_R$ ergeben. Wegen der damit großen Anzahl von auftretenden Termen wird für das Lösen der Gleichungen ein Computer-Algebra System verwendet. Für diesen Zweck wird das Algebrasystem Maple V benutzt. Von den Lösungen werden nur die beiden oberen Elemente T'^{11}_R und T'^{12}_R von $\boldsymbol{T}'_R$ gezeigt, da die beiden unteren Elemente sich aus den oberen berechnen lassen. Die Zähler- und Nennerpolynome von T'^{11}_R und T'^{12}_R werden explizit für den Fall angegeben, daß die Matrixelemente von $\boldsymbol{G}$ und $\boldsymbol{H}$ nur

Zählerpolynome enthalten. Diese Bedingung ist erfüllt, wenn in den Pfaden $\boldsymbol{G}$ und $\boldsymbol{H}$ keine Verzweigungen auftreten; ansonsten müssen die Zähler- und Nennerpolynome noch ermittelt werden. Die Elemente in $\boldsymbol{T}_R'$ weisen alle den gemeinsamen Nenner auf:

$$T'_{nR} = u(r(G_{11} - G_{22} + G_{21} - G_{12}) + s(H_{11} - H_{22} + H_{21} - H_{12})), \tag{3.80}$$

so daß $T_R'^{11}$ und $T_R'^{12}$ durch die Polynombrüche

$$T_R'^{11} = \frac{T_{zR}'^{11}}{T'_{nR}} \qquad \text{und} \qquad T_R'^{12} = \frac{T_{zR}'^{12}}{T'_{nR}} \tag{3.81}$$

dargestellt werden kann. Die beiden Zähler der oberen Elemente ergeben sich zu

$$\begin{aligned} T_{zR}'^{11} = {} & (H_{11} - rvG_{22} + (r-1)H_{12} + (u-1+v)H_{21} + (ru - r + u + \\ & 1 - v + rv)H_{22})G_{11} + (rvG_{21} + (s-1)H_{11} + (1-r-s)H_{12} + (1 + \\ & us - v - u + vs - s)H_{21} + (s + v + r - us + u - ru - vs - 1 - rv)H_{22} \\ &)G_{12} + +((1-v)H_{11} + (r + v - 1 - rv)H_{12} + (u-1)H_{21} + (ru - u + \\ & 1 - r)H_{22})G_{21} + ((s - 1 - vs + v)H_{11} + (rv - r + 1 - s - v + vs)H_{12} \\ & + (us - s + 1 - u)H_{12} + (r - us - 1 + s + u - ru)H_{22})G_{22} + (vs - 2s \\ & + us)H_{22}H_{11} + (2s - us - vs)H_{21}H_{12} \end{aligned} \tag{3.82}$$

und

$$\begin{aligned} T_{zR}'^{12} = {} & (-rvG_{22} + (s - 1 + r)H_{11} + (1-s)H_{12} + (1 - r - v - u + \\ & vs + rv - s + us + ru)H_{21} + (s - 1 - vs + v + u - us)H_{22})G_{11} + \\ & (rvG_{21} + (1-r)H_{11} - H_{12} + (v - 1 + u + r - ru - rv)H_{21} + (1 - \\ & u - v)H_{22})G_{12} + ((s - 1 - rv + r - vs + v)H_{11} + (vs - v - s + \\ & 1)H_{12} + (us - s + ru - u + 1 - r)H_{21} + (s - 1 - us + u)H_{22})G_{21} \\ & + ((1 - v - r + rv)H_{11} + (v-1)H_{12} + (r - 1 - ru + u)H_{21} + (1 - \\ & u)H_{22})G_{22} + (vs - 2s + us)H_{22}H_{11} + (2s - us - vs)H_{21}H_{12}. \end{aligned} \tag{3.83}$$

Die beiden unteren Matrixelemente in $\boldsymbol{T}_R'$ ergeben sich aus den jeweils diagonal liegenden oberen Elementen. Der Nenner ist für alle Matrixelemente identisch mit T'_{nR}. Der Zähler $T_{zR}'^{21}$ ergibt sich aus $T_{zR}'^{12}$ und $T_{zR}'^{22}$ aus $T_{zR}'^{11}$, indem die Indizes von allen Termen in $T_{zR}'^{11}$ und $T_{zR}'^{22}$ vertauscht werden mit den Korrespondenzen:

$$11 := 22, \quad 22 := 11, \quad 12 := 21 \quad \text{und} \quad 21 := 12. \tag{3.84}$$

Zusätzlich werden alle auftretenden Minus- und Pluszeichen vertauscht, so daß zum Beispiel aus

$$T_{zR}'^{11} = (H_{11} - rvG_{22} + (r-1)H_{12} + (u-1+v)H_{21} \ldots$$

mit der beschriebenen Transformation

$$T_{zR}'^{22} = (-H_{22} + rvG_{11} + (1-r)H_{21} + (1-u-v)H_{12} \ldots$$

erhalten wird. Da die Betriebskettenmatrizen $\boldsymbol{T}$ noch ein Rohrelement aufweisen, wird analog zu $\boldsymbol{T}_D$ ein zusätzliches Rohrelement rechts an die Ringstruktur angefügt, wodurch die Betriebskettenmatrix

$$\boldsymbol{T}_R = \begin{pmatrix} T_R'^{11} & T_R'^{12} \cdot z^{-1} \\ T_R'^{21} & T_R'^{22} \cdot z^{-1} \end{pmatrix} \tag{3.85}$$

resultiert. Für einen symmetrischen Ring, in dem die beiden Zweige identisch sind und somit $\boldsymbol{H} = \boldsymbol{G}$ gilt, vereinfachen sich die Matrixelemente in $\boldsymbol{T}_R'$ zu

$$T_R'^{11} = \frac{G_{11} + (s+r-1)G_{12} + (v-1)G_{21} + (1+sv+rv-v-r-s)G_{22}}{(r+s)v} \tag{3.86}$$

und

$$T_R'^{12} = \frac{(s+r-1)G_{11} + G_{12} + (1+rv+vs-v-s-r)G_{21} + (v-1)G_{22}}{(r+s)v}. \tag{3.87}$$

Der skalare Nenner läßt sich dadurch erklären, daß im symmetrischen Fall die Pole sich auf den Positionen von Nullstellen befinden und diese Pol-Nullstellenpaare sich somit aufheben. Im Folgenden wird ein Beispiel behandelt, das einen Übergang von einer symmetrischen Ringstruktur mit $\boldsymbol{G} = \boldsymbol{H}$ zu einer unsymmetrischen Ringstruktur mit $\boldsymbol{G} \neq \boldsymbol{H}$ beschreibt. Die beiden unverzweigten Rohrzweige $\boldsymbol{G}$ und $\boldsymbol{H}$ enthalten dabei jeweils vier Rohrelemente mit drei Reflexionskoeffizienten. Der Systemeingang befindet sich mit b_1 am linken Dreitor von Bild (3.8), während sich der Systemausgang mit c_3 am rechten Dreitor befindet.Dadurch hängt die Übertragungsfunktion nur

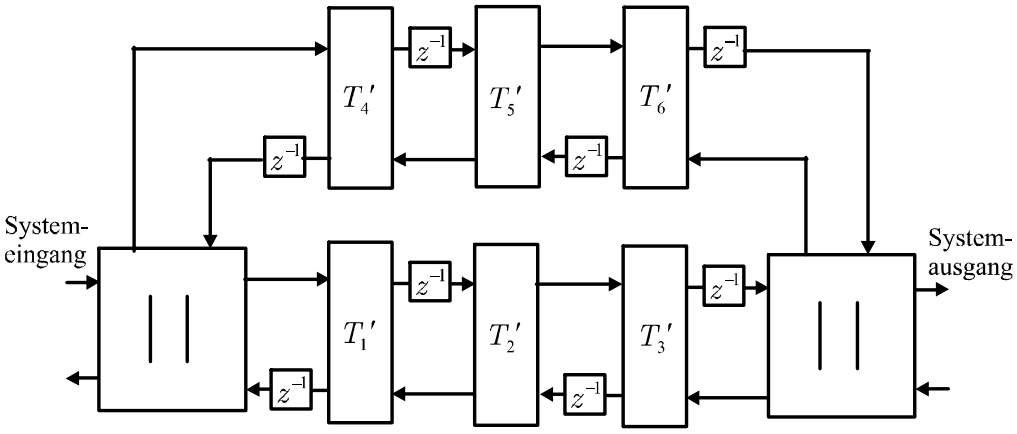


Bild 3.8: Rohrsystem mit Ringstruktur.

von $T_{zR}'^{11}$ und T_{nR}' ab. Die vier Betragsgänge in Bild 3.9 zeigen einen Parameterübergang von einer symmetrischen Rohrstruktur zu einer unsymmetrischen Rohrstruktur, durch welchen die Aufspaltung der Pol-Nullstellen Paare zu erkennen ist.Die Pole und Nullstellen der ersten und vierten Konfiguration sind in Bild 3.10 zu sehen. Im symmetrischen Fall (linkes Bild) liegen vier Nullstellen und Pole aufeinander, während sich im unsymmetrischen Fall Pol-Nullstellen Paare aufgespalten haben. Dieses Beispiel wurde im Frequenzbereich und Zeitbereich realisiert, was das Übertragungsverhalten bestätigt hat. Im Zeitbereich sind zwei Realisierungsbedingungen einzuhalten, um das Rohrmodell konsistent umzusetzen. Die Reflexionskoeffizienten können im Ring nicht willkürlich eingestellt werden. Ist eine Fläche im Ring definiert, so sollten sich die anderen Flächen aus den Reflexionskoeffizienten ergeben. Werden ausgehend von einer Fläche die anderen Flächen im Ring nacheinander berechnet, so ist dabei zu achten, daß die letzte berechnete Fläche den Ring wieder mit dem Wert der Ersten konsistent schließt. Eine Ringstruktur mit N Flächen enthält N Reflexionskoeffizienten und

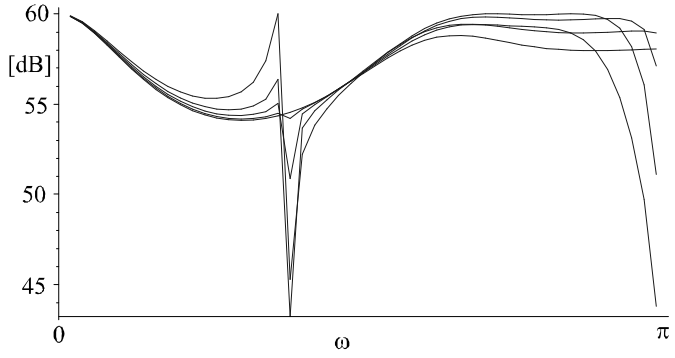


Bild 3.9: Betragsgänge als Polygonzug des Überganges einer symmetrischen Ringstruktur in eine unsymmetrische Struktur.

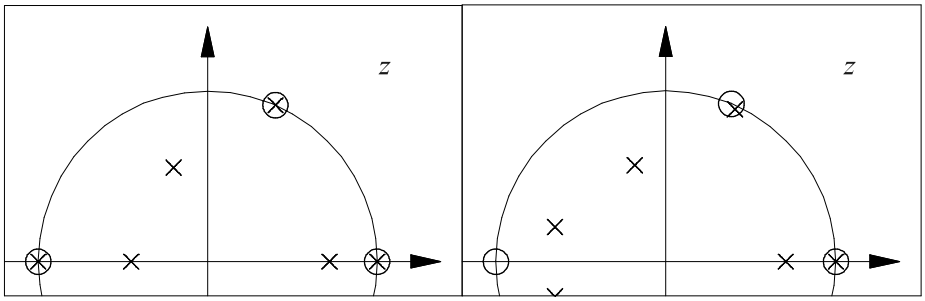


Bild 3.10: Pole und Nullstellen der Ringstruktur: Symmetrischer Fall (linkes Bild) und unsymmetrischer Fall (rechtes Bild).

Dreitorparameter, welche Querschnittssprünge im Ring bewirken. Somit ist ein Reflexionskoeffizient im Ring durch die anderen bestimmt. Weiterhin ist in der Ringstruktur auf die Positionen der Laufzeitglieder zu achten, welche die Rohrelemente darstellen. Diese werden mit einem Zustandsspeicher oben oder unten im Signalflußpfad realisiert. Hier macht sich der Unterschied zwischen $\widetilde{P}$ und P gemäß (3.49)-(3.51) bemerkbar. Deshalb wird in diesem Beispiel eine geradzahlige Anzahl von vier Rohrelementen für $\boldsymbol{G}$ und $\boldsymbol{H}$ verwendet, in denen sich die Zustandsspeicher abwechselnd oben und unten im Signalflußpfad befinden. Als Bedingung ist es ausreichend, wenn jeweils in $\boldsymbol{G}$ und $\boldsymbol{H}$ sich in den unteren und oberen Pfaden gleich viele Zustandsspeicher befinden.

Übertragungsfunktion mit Systemeingang am Rohranfang

In den vorangegangenen Abschnitten wurde gezeigt, wie mit $\boldsymbol{T}_D$ und $\boldsymbol{T}_R$ durch (3.76) und (3.85) eine Rohrverzweigung oder eine Ringstruktur in eine 2×2 Betriebskettenmatrix umgewandelt werden kann. Im Folgenden wird gezeigt, wie nach Festlegung des Systemeingangs und -ausgangs mit Hilfe der 2×2 Matrizen und den Rohrabschlüssen die Übertragungsfunktion bestimmt werden kann. Zuerst wird der Fall behandelt, in dem sich der Systemeingang am Anfang der Rohrstruktur mit dem Rohrabschluß $D(z)$ befindet und der Ausgang X_N^+ am Rohrende mit dem Abschluß $C(z)$. Die Abschlüsse

werden allgemein durch Polynombrüche dargestellt:

$$C(z) = \frac{C_n(z)}{C_d(z)} \quad \text{und} \quad D(z) = \frac{D_n(z)}{D_d(z)}. \tag{3.88}$$

Die Betriebskettenmatrizen zwischen den beiden Rohrabschlüssen werden zu der Gesamtmatrix $\boldsymbol{T}_g$ multipliziert. Das gemeinsame Nennerpolynom T_{ng} der Elemente von $\boldsymbol{T}_g$ wird vor die Matrix gezogen, so daß $\boldsymbol{T}_g = \boldsymbol{T}_{zg}/T_{ng}$ gilt, wobei $\boldsymbol{T}_{zg}$ nur Zählerpolynome enthält. Das Eingangssignal q wird am Rohranfang im oberen Signalflußpfad zu x_0^+ addiert. Damit ergibt sich die Matrixgleichung:

$$\begin{pmatrix} X_0^+ + Q \\ X_0^- \end{pmatrix} = \frac{1}{T_{ng}} \begin{pmatrix} T_{zg}^{11} & T_{zg}^{12} \\ T_{zg}^{21} & T_{zg}^{22} \end{pmatrix} \begin{pmatrix} X_N^+ \\ X_N^- \end{pmatrix} \tag{3.89}$$

und mit den Gleichungen für die Rohrabschlüsse

$$X_N^- = CX_N^+ \quad \text{und} \quad X_0^+ = DX_0^- \tag{3.90}$$

resultiert die Übertragungsfunktion:

$$H = \frac{X_N^+}{Q} = \frac{C_d D_d T_{ng}}{C_d D_d T_{zg}^{11} + C_n D_d T_{zg}^{12} - D_n (C_d T_{zg}^{21} + C_n T_{zg}^{22})}. \tag{3.91}$$

In (3.91) sind das Zähler- und Nennerpolynom von $H(z)$ explizit angegeben. Für ein unverzweigtes Rohr mit nur einem Abschluß C am Rohrausgang ergibt sich aus (3.91) die Formel (3.70) mit $T_{ng} = 1$ und $D = 0$ (d.h. $D_d = 1$ und $D_n = 0$).

Systemeingang zwischen den Rohrabschlüssen

Der Systemeingang muß sich nicht direkt an einem Rohrabschluß befinden, sondern kann auch innerhalb des Rohrsystems plaziert werden. Dadurch existieren Betriebskettenmatrizen vor und hinter dem Systemeingang, welche zu $\boldsymbol{T}_A$ und $\boldsymbol{T}_B$ aufmultipliziert werden. $\boldsymbol{T}_A$ beschreibt die Rohrkonfiguration vor dem Systemeingang mit dem Abschluß C und $\boldsymbol{T}_B$ beschreibt die Rohrkonfiguration hinter dem Systemeingang, die den Systemausgang am Rohrabschluß D enthält. Das Eingangssignal q wird mit dem Faktor k_+ zu dem oberen Signalpfad x_v^+ und mit dem Faktor k_- zu dem unteren Signalpfad x_v^- addiert. Werden aus den beiden resultierenden Betriebskettenmatrizen $\boldsymbol{T}_A$ und $\boldsymbol{T}_B$ die jeweiligen Nennerpolynome vorangestellt, ergeben sich die Gleichungen

$$\begin{pmatrix} X_0^+ \\ X_0^- \end{pmatrix} = \frac{1}{T_{nA}} \begin{pmatrix} T_{zA}^{11} & T_{zA}^{12} \\ T_{zA}^{21} & T_{zA}^{22} \end{pmatrix} \begin{pmatrix} X_v^+ \\ X_v^- + k_- \cdot Q \end{pmatrix} \tag{3.92}$$

und

$$\begin{pmatrix} X_v^+ + k_+ \cdot Q \\ X_v^- \end{pmatrix} = \frac{1}{T_{nB}} \begin{pmatrix} T_{zB}^{11} & T_{zB}^{12} \\ T_{zB}^{21} & T_{zB}^{22} \end{pmatrix} \begin{pmatrix} X_N^+ \\ X_N^- \end{pmatrix}. \tag{3.93}$$

Zusammen mit den Gleichungen (3.90) für die Rohrabschlüsse ergibt sich die Übertragungsfunktion

$$H = \frac{X_N^+}{Q} = \frac{H_n}{H_d} \tag{3.94}$$

mit dem Zähler

$$H_n = T_{nB}(D(k_- T_{zA}^{22} - k_+ T_{zA}^{21}) - k_- T_{zA}^{12} + k_+ T_{zA}^{11}) \tag{3.95}$$

und dem Nenner

$$\begin{aligned} H_d = {} & T_{zA}^{11}(T_{zB}^{11} + CT_{zB}^{12}) + T_{zA}^{12}(T_{zB}^{21} + CT_{zB}^{22}) \\ & - D(T_{zA}^{21}(T_{zB}^{11} + CT_{zB}^{12}) + T_{zA}^{22}(T_{zB}^{21} + CT_{zB}^{22})). \end{aligned} \tag{3.96}$$

Für einen Systemeingang direkt an dem Rohrabschluß C ist $\boldsymbol{T}_A$ gleich der Einheitsmatrix, woraus dann für $k_+ = 1$ und $k_- = 0$ die Übertragungsfunktion (3.91) resultiert. Folglich ist (3.91) und (3.70) ein Spezialfall von (3.94). Für den Fall mit $k_+ = 1$, $k_- = 0$, $D = 0$ und $C = \pm 1$ ergibt sich für ein unverzweigtes Rohr die Übertragungsfunktion

$$H = \frac{T_A^{11}}{T_A^{11}(T_B^{11} \pm T_B^{12}) + T_A^{12}(T_B^{21} \pm T_B^{22})}. \tag{3.97}$$

Treten keine Rohrverzweigungen auf, sind T_{nA} und T_{nB} konstant Eins. Dieses Rohrsystem kann als einfaches Modell für stimmlose Frikative und Explosive angesehen werden, da bei diesen Lauten sich die Anregung im Inneren des Vokaltraktes befindet. Für eine solche Rohrkonfiguration ist aus (3.97) zu erkennen, daß die hintere Höhle $\boldsymbol{T}_A$, welche sich vor dem Systemeingang befindet, die Nullstellen des Systems bestimmt. Die Pole des Systems werden hingegen durch die vordere und hintere Höhle bzw. mit $\boldsymbol{T}_A$ und $\boldsymbol{T}_B$ festgelegt. Bei Rohrabschlüssen mit Pol- und Nullstellen tragen diese auch zu den Nullstellen bzw. Polen bei.

Bestimmung der Übertragungsfunktion

Mit $\boldsymbol{T}_D$ von (3.76) kann eine Rohrabzweigung und mit $\boldsymbol{T}_R$ von (3.85) eine Ringstruktur durch eine 2×2 Betriebskettenmatrix dargestellt werden. Falls keine ineinander greifende Ringstrukturen auftreten, können durch diese Operationen alle Strukturen mit Dreitoren durch die 2×2 Matrizen beschrieben werden. Sind alle Rohrabzweigungen und Ringstrukturen in 2×2 Matrizen umgewandelt, so kann die Formel (3.94) verwendet werden, um die Übertragungsfunktion zu berechnen. Falls sich der Systemeingang am Rohranfang befindet, kann auch gleich (3.91) verwendet werden. Diese Umwandlung von Dreitor-Strukturen in 2×2 Betriebskettenmatrizen mit anschließender Anwendung von (3.94) und ihren vereinfachten Formeln kann durch einen Algorithmus spezifiziert werden. Dem Algorithmus ist die Rohrstruktur zu Anfang unbekannt, welche hierfür als Graph dargestellt wird. Die Elemente des Graphen bestehen aus den Adaptoren mit den dazugehörigen Rohrelementen und den Rohrabschlüssen. Diese Elemente sind miteinander verknüpft, wodurch jedes Element mindestens ein und höchsten drei Nachbarelemente aufweist. Wenn in dem Graph der Systemeingang und -ausgang gekennzeichnet sind, kann er analysiert werden. Im Folgenden werden die einzelnen Schritte der Analyse beschrieben, wofür in Bild (3.11) ein Beispiel gegeben ist:

1. Abschlüsse markieren: Als erstes werden die Rohrabschlüsse im Graphen durch den Algorithmus markiert, deren unverzweigte Zweige bzw. Teilrohrsysteme keinen Systemeingang oder -ausgang besitzen. Diese Endpunkte werden zusammen mit dem Rohrabschluß gespeichert.

2a. Umwandlung der Rohrabzweige: Da die markierten Rohrabschlüsse keinen Ein- oder Ausgang besitzen, stellen sie einen Seitenzweig eines Dreitors dar und können mittels $\boldsymbol{T}_D$ in eine 2×2 Matrix umgewandelt werden. Dadurch erscheinen die Verzweigungen mit den Abzweigen nicht mehr explizit in der Struktur der Betriebskettenmatrizen. Der Abzweig ist von einer Struktur mehrerer 2×2 Matrizen mit einem möglichen Rohrabschluß. Diese werden zu einer einzigen Matrix aufmultipliziert, so daß die Polynome Q und P in $\boldsymbol{T}_D$ bestimmt werden können. Dabei ist zu beachten, daß die 2×2 Matrizen des Abzweigs auch Nennerpolynome besitzen können, da sie durch eine oder mehrere durchgeführte Umwandlungen selbst Verzweigungen implizit beinhalten können.

2b. Umwandlung der Ringe: Ringstrukturen werden mittels $\boldsymbol{T}_R$ in eine 2×2 Matrix umgewandelt, so daß zusammen mit $\boldsymbol{T}_D$ sämtliche Verzweigungen abgearbeitet sind. Dadurch liegen nur noch verkettete 2×2 Betriebskettenmatrizen und Rohrabschlüsse vor. Der Systemeingang und -ausgang liegen an bestimmten Betriebskettenmatrizen.

3. Berechnung der Übertragungsfunktion: Die Matrizen zwischen Ein- und Ausgang werden zu der Matrix $\boldsymbol{T}_B$ aufmultipliziert. Sind Matrizen vor dem Systemeingang vorhanden, werden diese zu $\boldsymbol{T}_A$ aufmultipliziert. Es können verschiedene Fälle unterschieden werden, bezüglich der Existenz der Matrix $\boldsymbol{T}_A$ und der Rohrabschlüsse. Befindet sich der Systemeingang direkt an einem Rohrabschluß, dann wird keine Matrix $\boldsymbol{T}_A$ benötigt. Die verschiedenen Fälle der Kettenstruktur können mit den Bezeichnungen R für Abschluß, M für Matrix und E bzw. A für den Ein- bzw. Ausgang, welche in einer bestimmten Reihenfolge auftreten, identifiziert werden. Zwei Beispiele sind mit

$$\begin{gathered} R - M - E - M - A \\ \text{und} \\ E - M - A - R \end{gathered} \tag{3.98}$$

gegeben. Oberer Fall der Kettenstruktur muß mit (3.94) gelöst werden, während unterer Fall schon mit (3.91) verarbeitet werden kann. Beim Fehlen der Matrix $\boldsymbol{T}_A$ kann diese durch die Einheitsmatrix in der Formel ersetzt werden. Bei den Rohrabschlüssen kann analog vorgegangen werden, wobei der Zähler zu Null und der Nenner zu Eins gewählt wird. Alternativ können auch alle Fälle als Formeln jeweils einzeln vorliegen und nach einer Fallunterscheidung ausgewählt werden. Der Fall einer Matrix hinter dem Ausgang $\ldots - A - M - R$ ist nicht direkt durch die gezeigten Formeln abgedeckt, kann aber analog hergeleitet werden. Dafür kann auch die Formel einer Struktur mit $\ldots - A - R$ verwendet werden, wenn die Matrix und der Abschluß $-A - M - R$ in einen Rohrabschluß $-A - R'$ zusammengefaßt werden. Dies wäre für $R - M - E \ldots$ entsprechend am Rohranfang ebenfalls möglich.

Objektorientierte Implementierung

Für die Realisierung mittels eines Computerprogramms wurde ein objektorientierter Entwurf in der Programmiersprache C++ vorgenommen. Die Struktur wird dabei

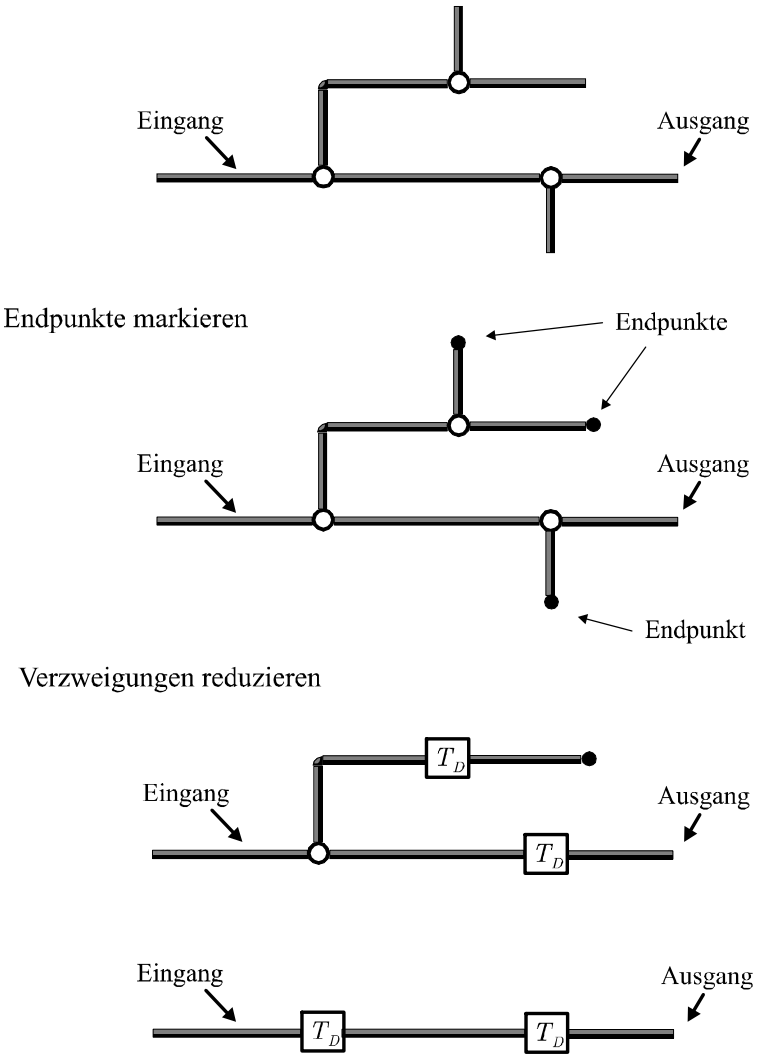


Bild 3.11: Reduzierung der Rohrstruktur für die Berechnung der Übertragungsfunktion des Rohrsystems.

im Wesentlichen durch eine geeignete Definition der Klassen und ihren Beziehungen zueinander bestimmt. Eine Klasse kann vereinfacht beschrieben werden durch einen heterogenen Datenbehälter bzw. Container, der noch zusätzlich Funktionen beinhaltet, welche direkt auf die Elemente bzw. Daten des Objekts zugreifen können. Auf eine Unterscheidung zwischen Klasse und Objekt wird hier verzichtet, da streng genommen die Objekte Instanzen der Klassen sind. Während die Elemente den Zustand eines Objekts beschreiben, rufen die Funktionen bzw. Methoden die Funktionsweisen des Objekts auf. Dies ist eine Abstrahierung von beliebigen Objekten der Welt, die durch verschiedene Eigenschaften beschrieben sind und auf die man nur objektspezifische Operationen anwenden kann. Im Folgenden werden die grundlegenden Klassen skizziert. Als einfachste Klassen bieten sich Zweitore, Dreitore und Abschlüsse an, welche als Tor-Element bezeichnet werden. Diese sind selber Elemente der Klasse Tor-Kette, welche eine Rohrstruktur darstellt. Tor-Element ist in Tor-Kette in einem Feld deklariert, so daß beliebig viele Tor-Elemente enthalten sein können. Dies kann auch durch eine Container-Klasse umgesetzt werden. Die Tor-Element Klassen besitzen als

Elemente die Ein- und Ausgänge der Tore; die Ausgänge **aus* liegen dabei als Zeiger bzw. Pointer vor, welche mit einem Stern (*) gekennzeichnet sind. Durch die Verwendung von Zeigern können zwei Elemente miteinander verknüpft werden, indem die Ausgänge auf die jeweiligen Eingänge des benachbarten Tor-Elementes zeigen. Dies wird unter anderem durch die Funktion *KettenElemente_verbinden()* durchgeführt. Durch das Element *nachbar* sind die Tor-Elemente miteinander verbunden und somit untereinander bekannt. Weitere Elemente sind die Modellparameter wie zum Beispiel die Reflexionskoeffizienten oder Querschnittsflächen und die Filterkoeffizienten für die Rohrabschlüsse. Die Anzahl der Tore ist durch *anz* festgelegt.

Tabelle 5: Beispiele von Klassendefinitionen für das Rohrmodell.

Klasse: Tor-Element

Elemente:
koeff, *Flächen*
ein[anz], **aus*[anz]
**nachbar*[anz]
// int anz: Toranzahl 1-3
PolyBruchMatrix *mT*
//Template(Matrix)
FktPtr.: *calc_Tor()*
⋮

Methoden: *set_Wellendarstellung()* ⋮

Klasse: Tor-Kette
Elemente: Tor-Element **torElemente* int **calc_Reihenfolge* **systemEin*, **systemAus* Tor-Element **endPunkte* **ParameterListe* // fürOptimierungsalg. ⋮
Methoden: *calc_Hz()* // z-Bereich *calc_Kette()* // Zeitbereich *KettenElemente_verbinden()* ⋮

Die Berechnung eines Tores im Zeitbereich wird durch einen Funktionszeiger aufgerufen, der entsprechend der gewählten Wellendarstellung auf die jeweilige Methode (Funktionen der Klasse) zeigt. Mit dieser Funktion werden die Eingänge *ein* auf die Ausgänge *aus* durch die Streumatrix abgebildet. Für die Berechnung im Z-Bereich wird die Betriebskettenmatrix mT verwendet. Die Elemente von mT können entweder als Brüche von Polynomen oder nur als Polynome implementiert werden. Bei einem gemeinsamen Nennerpolynom kann dies einzeln abgespeichert werden, was die Berechnung verkürzt. Die Polynome selber besitzen reelle Koeffizienten und komplexe Argumente. Template-Klassen bieten sich für ein solches Problem in C++ an. Templates besitzen die Eigenschaft, daß die Typen der Variablen in der Klassendefinition unbestimmt sind und erst bei der Deklaration bestimmt werden. So kann eine Matrixklasse definiert werden, die beim Gebrauch beliebige Elementtypen besitzen kann [Ladd96]. Die Template-Klassen TK_Polynom, TK_Bruch und TK_Matrix können dann wie folgt dargestellt werden kann:

```
typedef TK_Polynom<double,complex> dcPoly;
typedef TK_Bruch<dcPoly> PolyBruch;
typedef TK_Matrix<PolyBruch> PolyBruchMatrix.
```

PolyBruchMatrix wird tatsächlich selber noch zu einer weiteren Klasse erweitert oder abgeleitet, welche eine Betriebskettenmatrix darstellt, so daß zum Beispiel auch der Reflexionskoeffizient einfach in der Matrix gesetzt werden kann. Damit kann die

allgemeine PolyBruchMatrix den Elementen und Methoden einer Betriebskettenmatrix angepaßt werden. Das Rechnen mit den Polynom- und Matrix-Objekten kann in der üblichen mathematischen Notation durchgeführt werden, da in C++ die mathematischen Operatoren wie zum Beispiel Plus '+' und Minus '-' überladen werden können. Um die Übertragungsfunktion zu berechnen, greift die Funktion *calc_Hz()* von Tor-Kette auf diese Matrizen von Tor-Element zurück. Die verschiedenen möglichen Rohrstrukturen werden über eine graphische Benutzerschnittstelle in Form eines Gitters eingegeben, wie in Bild 3.12 zu sehen ist.

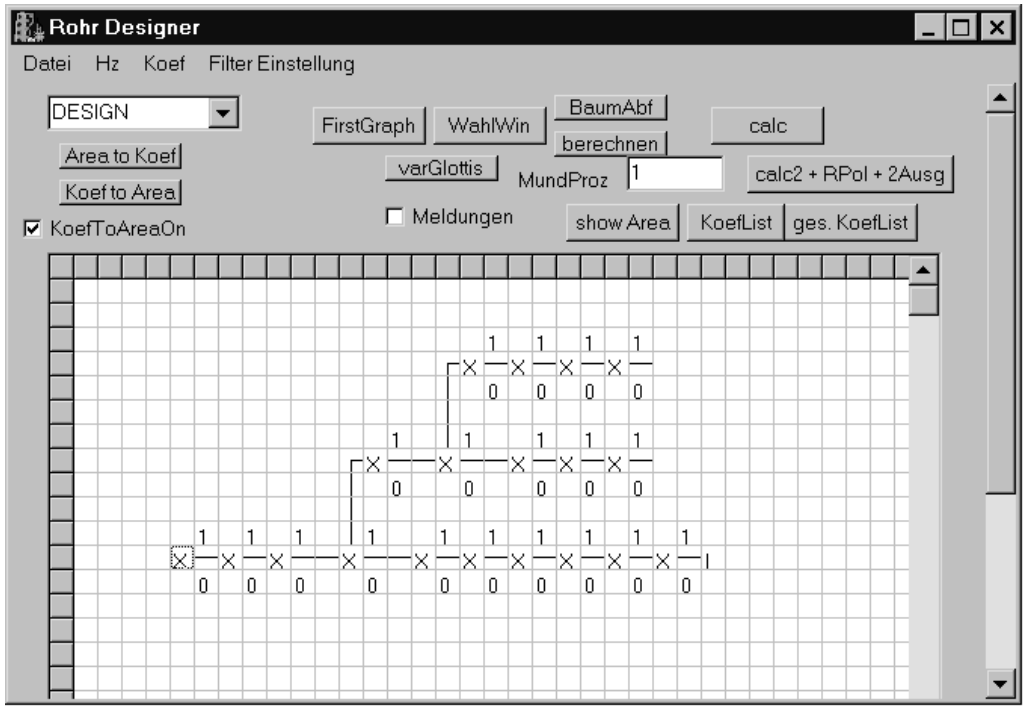


Bild 3.12: Eingabefenster für die Gestaltung der Rohrstruktur.

In das Gitter können Zweitor- oder Dreitor-Elemente sowie Rohrabschlüsse plaziert werden, welche durch Setzen von Verbindungslinien miteinander verknüpft werden können. Die Zahlen Null und Eins stellen die Anzahl von Zustandsspeicher im oberen und unteren Signalflußpfad dar, während die Kreuze einen Querschnittssprung und/oder eine Verzweigung darstellen. Die graphischen Elemente im Eingabefenster werden durch Klassen realisiert, die jeweils mit den Elementen der Tor-Kette mittels Zeiger verbunden sind, wodurch zur Laufzeit interaktiv auf Methoden von *torElemente* zugegriffen werden kann. Hier hat sich nachträglich herausgestellt, daß es teilweise praktischer wäre, ganze unverzweigte Teilrohrsysteme, d.h. mehrere Rohrstücke auf einmal, als Elemente zu plazieren, da dadurch weniger Elemente zu setzen sind. Nachdem die Rohrstruktur im Eingabefenster festgelegt wurde, kann sie automatisch analysiert werden, um die Übertragungsfunktion zu ermitteln. Zuerst fährt der Algorithmus die Rohrstruktur mit einem Zeiger ab, wobei die Eckpunkte der Rohrstruktur gespeichert werden entsprechend Schritt 1 "Abschlüsse markieren". Die Abzweige mit den entsprechenden Eckpunkten können dann mit T_D im zweiten Schritt (2a) vereinfacht werden. Dazu hat das Tor-Element des Dreitores eine zusätzliche Matrix als Element, welche

T_D repräsentiert. Wenn alle Dreitore in 2×2 Matrizen umgewandelt und die resultierenden Matrizen aufmultipliziert sind, wird der Fall der Kettenstruktur ermittelt. Damit kann entsprechend Schritt 3 die passende Formel für die Übertragungsfunktion ausgewertet werden.

Realisierung des Zeitbereichs

Für die Berechnung des Filters im Zeitbereich werden die Methoden *calc_ Tor()* der Objekte von Torelemente aufgerufen, die jeweils die entsprechenden Streumatrizen auswerten, um die Eingänge auf die Ausgänge abzubilden. Die Laufzeitglieder sind in den Streumatrizen und in *calc_ Tor()* nicht explizit enthalten. Es ist zu beachten, daß die Laufzeitglieder nur durch die Berechnungsreihenfolge der Torelemente realisiert werden. Die Speicherstelle des Ausgangs des einen Tores T_1 ist mit dem Eingang des benachbarten Tores T_2 identisch, was durch Zeiger realisierbar ist. Wird zuerst das Tor T_1 mit dem Ausgang berechnet und dann das benachbarte Tor T_2 mit dem darauf-liegenden Eingang, so kann der Wert in einer Iterationsstufe bzw. einem Taktzyklus von einem Tor zum Nächsten überführt werden, womit keine Laufzeit entsteht. Wird hingegen zuerst das benachbarte Tor T_2 mit dem Eingang berechnet, so übergibt danach der Ausgang von Tor T_1 einen Wert an T_2, der erst in der nächsten Iteration in die Berechnung der Streumatrix von T_2 mit eingeht. Durch diese zeitlich verzögerte Auswirkung des übergebenen Wertes ist ein Laufzeitglied realisiert. Daher wird durch die Berechnungsreihenfolge von zwei benachbarten Toren bestimmt, ob der dazwischenliegende Zustandsspeicher sich im oberen oder unteren Signalflußpfad befindet. Liegen für eine Rohrstruktur die Positionen der Zustandsspeicher fest, so kann durch einen Algorithmus die Berechnungsreihenfolge der Streumatrizen bestimmt werden. Dazu werden wiederholt in einer Schleife die Torelemente nacheinander überprüft, ob sie als nächstes berechnet werden könnten. Die ausgewählten Elemente werden in dieser Reihenfolge in eine Liste aufgenommen, bis alle eingetragen sind, wobei die Elemente nur einmal darin vorkommen dürfen. Ein Element kann nur dann als nächstes in die Liste eingetragen werden, wenn an allen seinen Toren mindestens eine der folgenden Bedingungen erfüllt ist: Die Bedingung ist für ein Tor erfüllt, wenn der Eingang des Tores an einem Zustandsspeicher liegt. Befindet sich der Zustandsspeicher an einem Ausgang, so muß das zu dem Tor benachbarte Element schon bereits in der Liste aufgenommen worden sein, was gleichbedeutend ist, daß es vorher berechnet wird. Befindet sich kein Nachbarelement an dem Tor ist die Bedingung von vornherein erfüllt. Sind diese Bedingungen für sämtliche Tore des Elementes erfüllt, so kann dieses für die nächste Berechnung durch einen Eintrag in die Liste vorgesehen werden. Diese Liste wird einmal erstellt und legt die Berechnungsreihenfolge für eine Rohrstruktur fest. Werden die Positionen der Zustandsspeicher verändert, so muß die Berechnungsreihenfolge wieder neu ermittelt werden.

Zum Testen der Funktionalität des Programms in der Entwicklungsphase hat sich ein Vergleich zwischen dem Betragsgang der berechneten Übertragungsfunktion und dem Betragsspektrum des Rohrsystemausgangs mit einer Anregung durch eine Impulsfolge bewährt.

3.3.3 Zeitvariable Abschlüsse

Da die Glottis nur in Zeitabschnitten als stationär angenommen werden kann, die kleiner als eine Grundperiode sind, erfordert eine genaue Modellierung des Rohrmodells an der Glottis einen zeitvariablen Rohrabschluß. Dafür soll der Rohrabschluß nur den Einfluß der Glottis auf die Wellenausbreitung im Vokaltrakt modellieren. Die rücklaufenden Wellen im Vokaltrakt, die auf den Glottisabschluß auftreffen, werden zum Teil in den Vokaltrakt wieder reflektiert, was durch den Übergang von der Vokaltraktimpedanz zur Glottisimpedanz verursacht wird. Die Glottisimpedanz $Z_G = R + i\boldsymbol{\omega}L$ hat neben einem Realteil R auch einen kleineren Imaginärteil [Fla72], der hier für den zeitvariablen Abschluß unberücksichtigt bleibt. Der Abschluß wird durch die Verengung der konstanten Vokaltraktfläche A_v zur zeitlich veränderlichen Glottisfläche A_g beschrieben. Daraus folgt für einen reellen Rohrabschlußkoeffizienten r_g an der Glottis:

$$r_g(t) = k_1 \cdot \frac{A_g(t) - A_v}{A_g(t) + A_v} \tag{3.99}$$

$$\text{mit} \quad k_1 \leq 1. \tag{3.100}$$

k_1 ist dabei ein reeller Faktor der zusätzliche Verluste modelliert. Die Verwendung von absoluten Flächen kann durch den Gebrauch von Flächenverhältnissen vermieden werden. Ist $\max(A_g)$ die maximale Fläche der Glottisöffnung in einer Periode, so kann das Verhältnis von $\max(A_g)$ zur größeren Vokaltraktfläche durch die Konstante k_2 bestimmt werden:

$$\frac{\max(A_g)}{A_v} = k_2 < 1. \tag{3.101}$$

Wird die Glottisfläche selbst auf ein Maximum von Eins normiert:

$$\overline{A}_g(t) = \frac{A_g(t)}{\max(A_g)}, \tag{3.102}$$

so kann der Glottiskoeffizient [SnL99b] ohne absolute Flächen angegeben werden mit

$$r_g(t) = k_1 \cdot \frac{k_2 \cdot \overline{A}_g(t) - 1}{k_2 \cdot \overline{A}_g(t) + 1}. \tag{3.103}$$

Für das Signal x_0 im Vokaltrakt direkt an der Glottis ergibt sich die Streuung durch den Rohrabschluß mittels

$$\begin{aligned} x_0^+ &= r_g(t) \cdot x_0^- \\ &\text{bzw.} \\ x_0^+ &= r_g(t) \cdot x_0^- + q, \end{aligned} \tag{3.104}$$

falls das Eingangssignal q an der Glottis eingespeist wird. Durch den zeitvariablen Abschluß ist das gesamte Rohrsystem zwar noch linear, aber nicht mehr zeitinvariant. Die Zeitvariabilität hat zur Folge, daß eine Übertragungsfunktion im Z-Bereich nicht mehr verfügbar ist, was sich auf die Parameterbestimmung auswirkt.

Kapitel 4

Analyse von Systemen und zeitdiskreten Rohrmodellen

Für die Erzeugung synthetischer Sprachsignale unter Verwendung von Rohrmodellen werden Modellparameter benötigt, mittels der die spektrale Einhüllende der Sprachlaute durch den Modellbetragsgang approximiert wird. Eine mögliche Vorgehensweise besteht darin Vokaltraktflächen zu verwenden, welche aus Röntgen- oder NMR-Aufnahmen gewonnen wurden. Da mit diesen Aufnahmetechniken überwiegend nur stationäre Laute recht aufwendig analysiert werden können, sind diesem Verfahren Grenzen gesetzt. Nachteilig ist, daß die Vereinfachungen des Sprechtraktmodells in den aus Röntgen- oder NMR-Aufnahmen ermittelten Parametern nicht berücksichtigt werden.

In dieser Arbeit wird eine Herangehensweise verfolgt, in der die Flächen aus dem natürlichen Sprachsignal geschätzt werden. Dies hat den Vorteil, daß die Übertragungsfunktion die akustischen Merkmale der natürlichen Sprache sehr gut modellieren kann, da die Modellparameter gerade nach diesem Fehlerkriterium geschätzt werden. Erwünscht ist darüber hinaus, daß die ermittelten Sprechtraktflächen auch die realen Verhältnisse der Sprechtraktgeometrie wiedergeben. Die Rohrmodelle werden aus systemtheoretischer Sicht betrachtet, so daß die Parameterbestimmung auf Prinzipien von Schätzverfahren allgemeiner linearer Systeme beruhen, wobei neue Methoden entwickelt werden müssen, da die Standardschätzverfahren nur auf den einfachsten Fall eines unverzweigten Rohrmodells angewandt werden können und die vorhandenen Schätzverfahren für erweiterte Systeme nur bedingt akzeptable Ergebnisse liefern. Ein Ansatz der Parameterbestimmung ist, mittels eines Optimierungsverfahrens einen definierten Fehler zu minimieren, der die Güte der Schätzung darstellt. Hierfür ist neben der Auswahl des Optimierungsverfahrens insbesondere die Fehlerdefinition entscheidend, wie in [Fr95a, Le83] auch schon erwähnt. Für die Definition des Fehlers werden in [Fr95a, Fr95b] Vergleiche von einzelnen Resonanzen vorgenommen, wobei sich der Nachteil ergibt, daß nicht alle Resonanzen der Laute berücksichtigt werden und die Resonanzen einzeln zugeordnet werden müssen. In [Gu93, Le83] werden LPC-Spektren für Vergleiche verwendet, auch wenn die Modelle Pole und Nullstellen besitzen, wodurch die Nullstellen nicht unmittelbar in der Schätzung berücksichtigt werden. In [Rah93, Schr94] werden unter anderem Cepstralkoeffizienten für ein Abstandsmaß benutzt. Für eine Schätzung des Vokaltraktes existieren auch Ansätze die Anregung mit in die Schätzung einzubeziehen, wobei für die Anregung parametrisierte Glottisfunktionen verwendet werden, wie in [Di96] diskutiert.

In dieser Arbeit wird aufbauend auf der Leistungsminimierung der inversen Filterung ein Fehlermaß für Pol- und Nullstellen Systeme vorgestellt, welches für die Schätzung von erweiterten Rohrmodellen verwendet werden kann. Die Parameter der Rohrmodelle sind im Gegensatz zu allgemeinen Pol-Nullstellensysteme schwieriger zu bestimmen, da bei diesen Systemen die Pole und Nullstellen unter der Bedingung der Minimalphasigkeit frei geschätzt werden können, während bei Rohrmodellparametern zusätzliche Restriktionen eingehalten werden müssen. Die Restriktionen bewirken, daß für erweiterte Rohrmodelle nicht jede minimalphasige Pol-Nullstellen Konfiguration in die Rohrparameter umgewandelt werden kann. Für allgemeine Pol-Nullstellen Schätzverfahren existieren Ansätze, die auf Prony zurückgehen und Ende des 18. Jahrhunderts diskutiert wurden. Bei diesen und anderen Verfahren besteht oft das Problem, daß der Systemeingang des zu schätzenden Systems benötigt wird, obwohl mit dem Sprachsignal nur ein Ausgangssignal vorliegt. Dies kann gelöst werden, indem eine Impulsantwort h' geschätzt wird. Die Koeffizienten des Pol-Nullstellen Modells mit der Impulsantwort h sollen dann in der Weise bestimmt werden, daß die Summe

$$\sum_k \left(h'_k - h_k\right)^2 \tag{4.1}$$

minimal wird. Die mit Prony verwandten Methoden: Shank's, Kalman's und Steiglitz's Methode versuchen auf unterschiedliche Weise den Ausdruck (4.1) zu minimieren [Kum82, Stei77]. Ein weiterer Ansatz besteht darin, zuerst ein Nur-Pole Modell hoher Ordnung zu schätzen, aus dem dann die Nullstellen bestimmt werden, was auch in Durbin's zweite Methode verwandt wird [Song80, Bro99]. In dieser Arbeit werden Pol-Nullstellen Schätzverfahren vorgestellt, die auf einem anderen Ansatz beruhen.

4.1 Allgemeine lineare Systeme

Für die Parameterbestimmung wird ein allgemeines lineares zeitinvariantes System betrachtet. Durch solche Systeme können physikalische Modelle mit Resonatoren dargestellt werden. Die Polstellen des Systems beschreiben dabei die Resonatoren, während die Nullstellen mögliche Antiresonatoren darstellen. Durch lineare Systeme kann das Frequenzverhalten und die Zeitentwicklung von physikalischen Vorgängen beschrieben werden, die in der zeitkontinuierlichen Darstellung einer linearen Differentialgleichung genügen. Für eine zeitdiskrete Modellierung wird die Differentialgleichung durch eine Differenzengleichung ersetzt. Die Übertragungsfunktion $H(z)$ eines linearen zeitdiskreten Systems ist durch das Nenner- und Zählerpolynom charakterisiert:

$$H(z) = \frac{b_0 + \sum_{i=1}^{M} b_i z^{-1}}{1 + \sum_{i=1}^{N} a_i z^{-1}}. \tag{4.2}$$

Wird dieses System mit einem unkorrelierten Signal $w(n)$, wie z.B. weißes Rauschen oder ein Impuls angeregt, so ergibt sich die Ausgangsfolge $y(n)$ als Lösung der Differenzengleichung:

$$y(n) + \sum_{i=1}^{N} a_i \cdot y(n-i) = \sum_{i=0}^{M} b_i \cdot w(n-i). \tag{4.3}$$

Wird w als stochastisches Signal angesehen, so kann dies auch als Verarbeitung eines Zufallsprozesses angesehen werden, der hier als stationär angenommen wird. Viele

physikalische Vorgänge lassen sich durch einen Zufallsprozeß interpretieren, in dem ein lineares Modell durch einen unkorrelierten Prozeß bzw. durch das resultierende unkorrelierte Signal angeregt wird. Da meist nur der gefilterte Prozeßausgang vorhanden ist, müssen die Modellparameter aus dem Ausgangssignal $y(n)$ ermittelt werden. Ein Lösungsansatz besteht darin, aus $y(n)$ abgeleitete statistische Größen, wie etwa die Autokorrelationsfunktion (AKF), zu verwenden. Da die AKF Produkte von zeitverschobenen Zufallsgrößen beinhaltet, wird Gleichung (4.3) mit $y(n-k)$ auf beiden Seiten multipliziert; anschließende Erwartungswertbildung führt auf:

$$\mathrm{E}\left[y(n)y(n-k)\right] + \sum_{i=1}^{N} a_i \mathrm{E}\left[y(n-i)y(n-k)\right] = \sum_{i=0}^{M} b_i \mathrm{E}\left[w(n-i)y(n-k)\right]. \qquad (4.4)$$

Die Erwartungswerte können durch die Autokorrelationsfunktion r_{yy} und die Kreuzkorrelationsfunktion r_{wy} dargestellt werden, wobei Stationarität berücksichtigt wird

$$r_{yy}(k) + \sum_{i=1}^{N} a_i \cdot r_{yy}(k-i) = \sum_{i=0}^{M} b_i \cdot r_{wy}(k-i). \qquad (4.5)$$

Da nur das Ausgangssignal y vorliegt, läßt sich nur r_{yy} unmittelbar berechnen, wohingegen für die Auswertung von r_{wy} Annahmen über das Eingangssignal gemacht werden müssen. Um die Kreuzkorrelationsfunktion in der Gleichung zu beseitigen, wird $y = w * h$ substituiert. Dabei ist h die Impulsantwort des kausalen Systems. Berücksichtigt man das $w(n)$ unkorreliert sein soll, so folgt mit der Konstanten $\boldsymbol{\sigma}_w$:

$$r_{ww}(i) = \boldsymbol{\sigma}_w^2 \cdot \boldsymbol{\delta}(i) \qquad \Longrightarrow \qquad r_{wy}(k-i) = \boldsymbol{\sigma}_w^2 \cdot h(i-k). \qquad (4.6)$$

Für ein kausales System kann daher mit

$$c_k = \sum_{i=k}^{M} b_i \cdot h(i-k) = \sum_{i=0}^{M-k} b_{i+k} \cdot h(i) \qquad (4.7)$$

die Gleichung (4.5) geschrieben werden als

$$r_{yy}(k) + \sum_{i=1}^{N} a_i \cdot r_{yy}(k-i) = \begin{cases} \boldsymbol{\sigma}_w^2 c_k & \text{für } 0 \le k \le M \\ 0 & \text{für } k > M. \end{cases} \qquad (4.8)$$

Das Gleichungssystem (4.8) wird auch als Yule-Walker Gleichung bezeichnet [Ha96], welche eine Beziehung zwischen den Filterkoeffizienten und der AKF herstellt. Die Produkte $b_{i+k} \cdot h(i)$ in (4.7) stellen die Nichtlinearität des Gleichungssystems und gleichzeitig die Auswirkung des nicht rekursiven Teils des Systems dar. Bei einem rein rekursiven System treten diese Produkte nicht mehr auf, so daß im Folgenden der wichtige Spezialfall des Nur-Pole Modells behandelt wird.

4.1.1 Analyse von rein rekursiven Systemen

Bei einem rein rekursiven System ist die Übertragungsfunktion

$$H(z) = \frac{b_0}{1 + \sum_{i=1}^{N} a_i z^{-1}} \qquad (4.9)$$

nur durch das Nennerpolynom bestimmt, abgesehen von dem Faktor b_0. Die zugehörige Differenzengleichung ergibt sich mit der unkorrelierten Anregung $w(n)$ zu

$$y(n) + \sum_{i=1}^{N} a_i \cdot y(n-i) = b_0 \cdot w(n). \tag{4.10}$$

Die Yule-Walker Gleichungen für das Nur-Pole Modell lassen sich dann mit (4.8) und $c_0 = b_0 h_0 = b_0^2$ bzw. $c_k = b_0^2 \boldsymbol{\delta}(k)$ schreiben als

$$r_{yy}(k) + \sum_{i=1}^{N} a_i \cdot r_{yy}(k-i) = b_0^2 \boldsymbol{\sigma}_w^2 \cdot \boldsymbol{\delta}(k)\ , \quad \text{für } 0 \leq k \leq N. \tag{4.11}$$

b_0 tritt nur in der ersten Gleichung für $k = 0$ auf und stellt einen frequenzunabhängigen Verstärkungsfaktor des Systems dar. (4.11) stellt $N+1$ Gleichungen mit dem Koeffizienten b_0 und den N Koeffizienten a_i dar. Wird der Verstärkungsfaktor b_0 nicht beachtet, so werden nur noch die Gleichungen für $k = 1 \ldots N$ benötigt, um nach den unbekannten Koeffizienten aufzulösen. Mit Hilfe der Autokorrelationsmatrix $\boldsymbol{R}$ und dem Autokorrelationsvektor $\boldsymbol{r}_{yy}$ können die Gleichungen für $k = 1 \ldots N$ in (4.11) durch eine Matrixgleichung ausgedrückt werden. Dabei wird berücksichtigt, daß die Autokorrelationsfunktion symmetrisch ist. Für die Bestimmung des Koeffizientenvektors $\boldsymbol{a}$ wird die Matrix $\boldsymbol{R}$ invertiert:

$$\boldsymbol{R} = \begin{pmatrix} r_{yy}(0) & r_{yy}(1) & \cdots & r_{yy}(N-1) \\ r_{yy}(1) & r_{yy}(0) & \cdots & r_{yy}(N) \\ \vdots & \vdots & \ddots & \vdots \\ r_{yy}(N-1) & r_{yy}(N-2) & \cdots & r_{yy}(0) \end{pmatrix}, \qquad \boldsymbol{r}_{yy} = \begin{pmatrix} r_{yy}(1) \\ r_{yy}(2) \\ \vdots \\ r_{yy}(N) \end{pmatrix} \tag{4.12}$$

$$\boldsymbol{R} \cdot \boldsymbol{a} = -\boldsymbol{r}_{yy} \quad \Longrightarrow \quad \boldsymbol{a} = -\boldsymbol{R}^{-1} \boldsymbol{r}_{yy}. \tag{4.13}$$

Gleichung (4.13) wird ebenfalls als Yule-Walker Gleichung bezeichnet. Das Gleichungssystem (4.13) kann durch den Durbin Algorithmus effizient gelöst werden, indem die spezielle Matrixform von $\boldsymbol{R}$ ausgenutzt wird, welche eine Töplitz Struktur ist. Neben einem Effizienzvorteil durch die schnelle rekursive Berechnung, erhält man zusätzlich durch den Durbin Algorithmus die Parcor Koeffizienten, welche eine neue Koeffizientendarstellung beinhalten. Diese ermöglichen ein einfaches Stabilitätkriterium des Nur-Pole Modells und korrespondieren mit den Reflexionskoeffizienten eines unverzweigten Rohrmodells mit einem Abschluß ± 1.

4.1.2 Lineare Prädiktion und inverse Filterung

Die lineare Prädiktion [MG76] eines Signals x stellt eine Schätzung der Signalwerte $x(n)$ durch eine Linearkombination vergangener Signalwerte $x(n-k)$ dar. $\widehat{x}(n)$ ist der Schätzwert von $x(n)$ mittels der Beziehung:

$$\widehat{x}(n) = \sum_{k=1}^{N} a_k x(n-k). \tag{4.14}$$

Der Fehler der Vorhersage wird als Prädiktionsfehler $e(n)$ bezeichnet und ergibt sich durch die Abweichung der Schätzung zum tatsächlichen Wert

$$e(n) = x(n) - \widehat{x}(n) = x(n) - \sum_{k=1}^{N} a_k x(n-k). \tag{4.15}$$

Das Prädiktionsfehlerfilter P berechnet $e(n)$ und ist in diesem Fall ein FIR-Filter mit der Übertragungsfunktion

$$P(z) = 1 - \sum_{k=1}^{N} a_k z^{-k}. \tag{4.16}$$

Der Prädiktor wird als optimal angesehen, wenn die Ausgangsleistung des Prädiktionsfehlerfilters minimal wird, wobei für statistische Signale der Erwartungswert minimiert wird:

$$\mathrm{E}[e(n)^2] \longrightarrow \min. \tag{4.17}$$

Der unter diesem Kriterium optimale Koeffizientensatz ist für die Analyse und Synthese von Signalen von großem Nutzen. Um ein Extremum von (4.17) zu erhalten, müssen die Ableitungen nach den Koeffizienten verschwinden, wodurch sich die Bedingungsgleichungen

$$\frac{\partial \mathrm{E}[e(n)^2]}{\partial a_k} = 0, \quad \text{für } k = 1 \ldots N \tag{4.18}$$

ergeben. Die Gleichungen (4.18) führen unter Verwendung der Autokorrelationsmatrix auf:

$$\boldsymbol{R} \cdot \boldsymbol{a} = \boldsymbol{r}_{xx} \quad \Longrightarrow \quad \boldsymbol{a} = \boldsymbol{R}^{-1} \boldsymbol{r}_{xx}. \tag{4.19}$$

Die Gleichung in (4.19) ist bis auf ein Vorzeichen identisch mit (4.13). Die optimalen Prädiktionskoeffizienten sind daher, vom Vorzeichen abgesehen, gleich den geschätzten Parametern eines AR-Prozesses. Folglich läßt sich die lineare Prädiktion zur Parameterschätzung von rein rekursiven Systemen nutzen. Die Übertragungsfunktion des Synthesefilters der linearen Prädiktion hat die Form

$$H_P(z) = \frac{1}{P(z)} = \frac{1}{1 - \sum_{i=1}^{N} a_i z^{-1}}, \tag{4.20}$$

wobei ein reeller Verstärkungsfaktor noch hinzugefügt werden kann. H_p entspricht dem System (4.9) für $b_0 = 1$. Da das Synthesefilter H_P das inverse System des Analysefilters P ist, spricht man auch von inverser Filterung. Nicht jedes Filter ist allerdings ohne Modifikationen für die inverse Filterung geeignet, wie noch diskutiert wird.

4.1.3 Lineare Prädiktion für periodische Signale

Eine einzelne Periode enthält bis auf die Kenntnis eines Phasenversatzes die vollständige Information eines periodischen Signals, so daß es in diesem Fall ausreicht für die Schätzung eine einzelne Periode zu analysieren. Da stimmhafte Sprachsignale abschnittsweise als annähernd periodisch angesehen werden können, wird hier die inverse Filterung für periodische Signale behandelt. In [Sn96] wurde dieser Ansatz verfolgt, der hier vollständig durch eine geometrische Interpretation dargestellt werden kann. Die zu analysierende Periode der Länge L wird dabei als reellwertiger Vektor $\boldsymbol{v}$ der

Dimension L aufgefaßt. Werte außerhalb des Analysefensters bzw. der Periode können durch periodische Fortsetzung ermittelt werden. Diese außerhalb liegenden Werte werden benötigt, wenn die Periode $\boldsymbol{v}$ um einen Wert verschoben wird. Der letzte Wert v_L rückt bei einer Linksverschiebung an die erste Stelle des Vektors. Als Ersatz für den Verzögerungs- bzw. Verschiebungsoperator z^{-1} wird der Operator zv^{-1} definiert, der eine zyklische Verschiebung darstellt mit

$$zv^{-1}(\boldsymbol{v}) = zv^{-1}(v_{1,}v_{2,\ldots}v_{L-1}, v_L) = (v_{L,}, v_{1,}v_{2,\ldots}v_{L-1}). \tag{4.21}$$

Eine mehrfache Anwendung von zv^{-1} spiegelt sich im Exponenten des Operators wieder. Die um i zyklisch verschobenen Perioden bzw. Vektoren werden durch einen unteren Index gekennzeichnet:

$$\boldsymbol{v}_i \equiv zv^{-i}(\boldsymbol{v}) = \underbrace{zv^{-1}(zv^{-1}(\ldots zv^{-1}}_{i-\text{mal}}(\boldsymbol{v}))\ldots). \tag{4.22}$$

Das Skalaprodukt $\langle||\rangle$ von zwei Vektorperioden $\boldsymbol{w}$ und $\boldsymbol{v}$ und die Norm $||||$ wird für reellwertige endlich dimensionale Vektorräume definiert durch

$$\langle \boldsymbol{w} \,|\boldsymbol{v}\rangle = \sum_{i=1}^{L} w_i v_i \ , \qquad\qquad \|\boldsymbol{v}\| = \sqrt{\langle \boldsymbol{v} \,|\boldsymbol{v}\rangle}. \tag{4.23}$$

Die Skalarprodukte von zyklisch verschobenen Vektoren weisen spezielle Eigenschaften auf, die für die Berechnung der Prädiktion verwendet werden. Diese Eigenschaften ergeben sich dadurch, daß nur die relative Verschiebung zwischen zwei Vektoren für ihr Skalarprodukt entscheidend ist, und darüber hinaus nur die Anzahl der Verschiebungen wichtig ist, aber nicht die Richtung der Verschiebung:

$$\langle \boldsymbol{v}_k \,|\boldsymbol{v}_{k+j}\rangle = \langle \boldsymbol{v} \,|\boldsymbol{v}_j\rangle \ , \qquad \langle \boldsymbol{v} \,|\boldsymbol{v}_j\rangle = \langle \boldsymbol{v} \,|\boldsymbol{v}_{-j}\rangle \,. \tag{4.24}$$

Die rechte Gleichung in (4.24) ergibt sich aus der Linken mit Hilfe der Symmetrie des Skalarprodukts. Weiterhin sind die Normen von allen zyklisch verschobenen Vektoren identisch

$$\|\boldsymbol{v}\| = \|\boldsymbol{v}_i\| \,, \tag{4.25}$$

da dieselben Werte in $\boldsymbol{v}_i$ wie in $\boldsymbol{v}$ auftreten, allerdings in einer anderen Permutation. Steht für die lineare Prädiktion eine Periode $\boldsymbol{v}$ zur Verfügung, so kann der geschätzte Vektor $\widehat{\boldsymbol{v}}$ mit dem Operator zv^{-1} beschrieben werden durch

$$\widehat{\boldsymbol{v}} = \sum_{k=1}^{N} a_k \cdot zv^{-k}(\boldsymbol{v}) = \sum_{k=1}^{N} a_k \cdot \boldsymbol{v}_k. \tag{4.26}$$

$\widehat{\boldsymbol{v}}$ soll dabei durch eine geeignete Linearkombination der Vektoren $\boldsymbol{v}_k$ den Vektor $\boldsymbol{v}$ möglichst gut approximieren. Die Prädiktion kann nur perfekt gelingen, wenn $\boldsymbol{v}$ ein Element der linearen Hülle ist, die durch die Vektoren $\boldsymbol{v}_k$ aufgespannt wird; andernfalls ist die Norm des Fehlervektors $\boldsymbol{e}$ mit:

$$\boldsymbol{e} = \boldsymbol{v} - \widehat{\boldsymbol{v}} = \boldsymbol{v} - \sum_{k=1}^{N} a_k \boldsymbol{v}_k, \tag{4.27}$$

die den Abstand zwischen dem geschätzten Vektor $\widehat{\boldsymbol{v}}$ und dem tatsächlichen Vektor $\boldsymbol{v}$ beschreibt, immer größer als Null. Die Approximation von $\boldsymbol{v}$ durch $\widehat{\boldsymbol{v}}$ wird als optimal angesehen, wenn die Norm $\|\boldsymbol{e}\|$ minimal ist. Dieses Optimum liegt vor, wenn die Skalarprodukte

$$\langle \boldsymbol{v} - \widehat{\boldsymbol{v}} \,|\boldsymbol{v}_j\rangle = \langle \boldsymbol{v} - \textstyle\sum_{k=1}^{N} a_k \boldsymbol{v}_k \,|\boldsymbol{v}_j\rangle = 0 \qquad \text{für } \; j = 1 \ldots N \tag{4.28}$$

verschwinden, was aus folgender Betrachtung gesehen werden kann. Wird $\boldsymbol{v}$ in einen parallelen und einen orthogonalen Anteil zur Linearen Hülle der N Vektoren $\boldsymbol{v}_k$ aufgespalten, so kann $\widehat{\boldsymbol{v}}$ nur den parallelen Anteil darstellen, da $\widehat{\boldsymbol{v}}$ selbst aus den Vektoren $\boldsymbol{v}_k$ zusammengesetzt ist. Für ein Optimum stellt $\widehat{\boldsymbol{v}}$ genau den parallelen Anteil dar.

Sukzessive Berechnung der Koeffizienten a_k

Die Gleichung (4.27) läßt sich zu

$$\boldsymbol{v} = \sum_{k=1}^{N} a_k \boldsymbol{v}_k + \boldsymbol{e} \tag{4.29}$$

umstellen, woran unmittelbar gesehen werden kann, daß $\boldsymbol{v}$ nach den zyklisch verschobenen Vektoren $\boldsymbol{v}_k$ entwickelt wird. Da nicht davon auszugehen ist, daß die Vektoren $\boldsymbol{v}_k$ eine orthogonale Basis bilden, können die verallgemeinerten Fourierkoeffizienten bzw. Skalarprodukte $\langle \boldsymbol{v} \,|\boldsymbol{v}_k\rangle \,/\, \|\boldsymbol{v}_k\|^2$ nicht direkt als Entwicklungskoeffizienten verwendet werden, um eine optimale Lösung zu erzielen. Die Entwicklung wird daher rekursiv vollzogen. Zuerst wird nur ein verschobener Vektor für die Entwicklung berücksichtigt, wonach in den folgenden Schritten immer mehr Vektoren berücksichtigt werden. Der obere Index der Koeffizienten a_k gibt dafür im Folgenden die Anzahl der verwendeten Vektoren für die Schätzung wieder. Entsprechend einem Orthogonalisierungsverfahren wird $\boldsymbol{v}$ in jedem Schritt mit Hilfe eines weiteren Vektors $\boldsymbol{v}_k$ approximiert. Steht eine Lösung unter Verwendung von m Vektoren $\boldsymbol{v}_1$ bis $\boldsymbol{v}_m$ mit den Koeffizienten a_1^m bis a_m^m zur Verfügung, so kann in der Regel die Approximation von $\boldsymbol{v}$ verbessert werden, wenn zusätzlich der nächste Vektor $\boldsymbol{v}_{m+1}$ für die Linearkombination in $\widehat{\boldsymbol{v}}$ verwendet wird. Für die Berechnung von a_{m+1}^{m+1} kann nicht direkt das Skalarprodukt $\langle \boldsymbol{v} \,|\boldsymbol{v}_{m+1}\rangle$ verwendet werden, da berücksichtigt werden muß, daß $\boldsymbol{v}$ mit $\widehat{\boldsymbol{v}} = \sum_{k=1}^{m} a_k^m \boldsymbol{v}_k$ schon durch die ersten Vektoren $\boldsymbol{v}_1$ bis $\boldsymbol{v}_m$ approximiert wird und $\boldsymbol{v}_{m+1}$ auch Anteile dieser Vektoren $\boldsymbol{v}_1$ bis $\boldsymbol{v}_m$ enthalten kann. Diese Anteile müssen in $\boldsymbol{v}_{m+1}$ beseitigt werden, wenn für die Berechnung des Entwicklungskoeffizienten das Skalarprodukt mit $\boldsymbol{v}_{m+1}$ gebildet werden soll. Die Bedingungsgleichungen sind dafür gegeben durch

$$\langle \boldsymbol{v}_{m+1} - \textstyle\sum_{k=1}^{m} b_k^m \boldsymbol{v}_k \,|\boldsymbol{v}_j\rangle = 0 \;, \qquad \text{für } j = 1 \ldots m. \tag{4.30}$$

Diese sind den Bedingungen in (4.28) für $N = m$ ähnlich. Durch die Eigenschaften (4.24) der Skalarprodukte gilt auch $\langle \boldsymbol{v}_k \,|\boldsymbol{v}_j\rangle = \langle \boldsymbol{v}_{-k} \,|\boldsymbol{v}_{-j}\rangle$. Damit gehen die Gleichungen (4.30), welche auch mit

$$\langle \boldsymbol{v}_{m+1} \,|\boldsymbol{v}_j\rangle - \textstyle\sum_{k=1}^{m} b_k^m \langle \boldsymbol{v}_k \,|\boldsymbol{v}_j\rangle = 0 \;, \qquad \text{für } j = 1 \ldots m$$

dargestellt werden können, über in

$$\langle \boldsymbol{v}_{-m-1} - \textstyle\sum_{k=1}^{m} b_k^m \boldsymbol{v}_{-k} \,|\boldsymbol{v}_{-j}\rangle = 0 \;, \qquad \text{für } j = 1 \ldots m. \tag{4.31}$$

Durch Verschieben der Vektorindizes um $m+1$ nach (4.24) erhält man

$$\langle \boldsymbol{v} - \textstyle\sum_{k=1}^{m} b_k^m \boldsymbol{v}_{m+1-k} \,|\boldsymbol{v}_{m+1-j}\rangle = 0 \ , \qquad \text{für } j = 1 \ldots m. \tag{4.32}$$

Mit $k' = m+1-k$ und $j' = m+1-j$ resultiert weiterhin:

$$\langle \boldsymbol{v} - \textstyle\sum_{k'=1}^{m} b_{m+1-k'}^m \boldsymbol{v}_{k'} \,|\boldsymbol{v}_{j'}\rangle = 0 \ , \qquad \text{für } j' = 1 \ldots m. \tag{4.33}$$

Durch Vergleich mit (4.28) für $N = m$ kann man sehen, daß die Beziehung

$$a_k^m = b_{m+1-k}^m, \qquad \text{für } k = 1 \ldots m \tag{4.34}$$

gilt. $\boldsymbol{v}_{m+1} - \sum_{k=1}^{m} b_k^m \boldsymbol{v}_k$ ist nach (4.30) der Anteil von $\boldsymbol{v}_{m+1}$ der orthogonal zur Linearen Hülle der Vektoren $\boldsymbol{v}_1$ bis $\boldsymbol{v}_m$ ist. Dieser orthogonale Anteil von $\boldsymbol{v}_{m+1}$ wird statt $\boldsymbol{v}_{m+1}$ selbst für die Entwicklung verwendet. Dadurch kann der optimale Entwicklungskoeffiezient a_{m+1}^{m+1} durch den verallgemeinerten Fourierkoeffizienten berechnet werden:

$$a_{m+1}^{m+1} = \frac{\langle \boldsymbol{v} \,|\boldsymbol{v}_{m+1} - \sum_{k=1}^{m} b_k^m \boldsymbol{v}_k\rangle}{\left\|\boldsymbol{v}_{m+1} - \sum_{k=1}^{m} b_k^m \boldsymbol{v}_k\right\|^2}, \tag{4.35}$$

der mit (4.34) auch dargestellt werden kann als

$$a_{m+1}^{m+1} = \frac{\left\langle \boldsymbol{v} \,\middle|\, \boldsymbol{v}_{m+1} - \sum_{k=1}^{m} a_{m+1-k}^m \boldsymbol{v}_k\right\rangle}{\left\|\boldsymbol{v}_{m+1} - \sum_{k=1}^{m} a_{m+1-k}^m \boldsymbol{v}_k\right\|^2}. \tag{4.36}$$

Die Entwicklung von $\boldsymbol{v}$ ergibt sich mit den $m+1$ Vektoren $\boldsymbol{v}_k$ zu

$$\widehat{\boldsymbol{v}} = \textstyle\sum_{k=1}^{m} a_k^m \boldsymbol{v}_k + a_{m+1}^{m+1} \cdot (\boldsymbol{v}_{m+1} - \sum_{k=1}^{m} b_k \boldsymbol{v}_k), \tag{4.37}$$

oder mit (4.34) und nach $\boldsymbol{v}_k$ sortiert zu

$$\widehat{\boldsymbol{v}} = \textstyle\sum_{k=1}^{m} (a_k^m - a_{m+1}^{m+1} a_{m+1-k}^m) \boldsymbol{v}_k + a_{m+1}^{m+1} \boldsymbol{v}_{m+1}. \tag{4.38}$$

Damit lassen sich die optimalen Koeffizienten unter Verwendung von $m+1$ Vektoren darstellen durch a_{m+1}^{m+1} und den m Koeffizienten a_k^m, welche die optimale Lösung mit nur m Vektoren repräsentiert ohne Berücksichtigung des Vektors $\boldsymbol{v}_{m+1}$. Die neuen optimalen Koeffizienten a_k^{m+1} ergeben sich nach (4.38) durch die Beziehungen:

$$a_k^{m+1} = a_k^m - a_{m+1}^{m+1} a_{m+1-k}^m \quad \text{für} \quad k = 1 \ldots m \tag{4.39}$$

$$a_{m+1}^{m+1} \qquad\qquad\qquad \text{für} \quad k = m+1. \tag{4.40}$$

Lösungsweg unter Verwendung der Vektoren $\boldsymbol{v}_k$

Für $m = 1$ ist die Lösung des ersten Koeffizienten durch (4.35) mit $m = 0$ gegeben durch:

$$a_1^1 = \frac{\langle \boldsymbol{v} \,|\boldsymbol{v}_1\rangle}{\|\boldsymbol{v}_1\|^2}, \tag{4.41}$$

da noch keine Koeffizienten b_k benötigt werden. Mit (4.41) beginnend können dann die Lösungen für $m = 2 \ldots N$ mit (4.36) und (4.39) sukzessive berechnet werden, wodurch sich letztlich die optimalen Koeffizienten a_k^N ergeben. Hierfür besteht in den Lösungsfomeln eine Analogie zum Durbin-Algorithmus.

Lösungsweg unter Verwendung orthogonaler Anteile

Die zukzessive Berechnung läßt sich auch mit Hilfe der Fehlervektoren darstellen, wie im Folgenden diskutiert. Es wird angenommen, daß $\boldsymbol{v}$ schon durch m Vektoren $\boldsymbol{v}_k$ mit den entsprechenden Koeffizienten approximiert wird. Für die weitere Entwicklung von $\boldsymbol{v}$ mit $m+1$ Vektoren kann der Anteil von $\boldsymbol{v}_{m+1}$ verwendet werden, der orthogonal zur linearen Hülle $\mathrm{Lin}\{\boldsymbol{v}_1 \dots \boldsymbol{v}_m\}$ ist, wie in (4.37) mit $\boldsymbol{v}_{m+1} - \sum_{k=1}^{m} b_k \boldsymbol{v}_k$ zu erkennen ist. Daher kann statt der Vektoren $\boldsymbol{v}_k$ selbst, die Anteile für die Lösung verwendet werden, die orthogonal zu den ersten $k-1$ Vektoren $\boldsymbol{v}_i$ mit $i < k$ sind. Dafür ist es zweckmäßig den Vektor $\boldsymbol{v}_k^{\perp n}$zu definieren, welcher den orthogonalen Anteil von $\boldsymbol{v}_k$ zu $\mathrm{Lin}\{\boldsymbol{v}_1 \dots \boldsymbol{v}_n\}$ darstellt. Es werden für die weiteren Berechnungen nur die beiden Fälle $k = 0$ und $k = n+1$ von $\boldsymbol{v}_k^{\perp n}$ benötigt, welche durch

$$\begin{aligned} \boldsymbol{v}_0^{\perp m} &= \boldsymbol{v} - \textstyle\sum_{k=1}^{m} a_k^m \boldsymbol{v}_k \\ \boldsymbol{v}_{m+1}^{\perp m} &= \boldsymbol{v}_{m+1} - \textstyle\sum_{k=1}^{m} b_k^m \boldsymbol{v}_k \end{aligned} \tag{4.42}$$

dargestellt werden können. Der orthogonale Anteil ist der Anteil des Vektors, der senkrecht zu den Elementen von $\mathrm{Lin}\{\boldsymbol{v}_1 \dots \boldsymbol{v}_m\}$ ist, und wird entsprechend (4.27) somit als Fehlervektor interpretiert. $\boldsymbol{v}_0^{\perp m}$ entspricht $\boldsymbol{e}$ aus (4.27) für $m = N$. Sind die Fehlervektoren $\boldsymbol{v}_0^{\perp m}$ und $\boldsymbol{v}_{m+1}^{\perp m}$ bekannt, so kann der Fehler $\boldsymbol{v}_0^{\perp m+1}$ der nächsten Lösung mit (4.37) dargestellt werden zu

$$\begin{aligned} \boldsymbol{v}_0^{\perp m+1} &= \boldsymbol{v} - \widehat{\boldsymbol{v}} \\ &= \boldsymbol{v} - \textstyle\sum_{k=1}^{m} a_k^m \boldsymbol{v}_k + a_{m+1}^{m+1} \cdot (\boldsymbol{v}_{m+1} - \sum_{k=1}^{m} b_k \boldsymbol{v}_k) \\ &= \boldsymbol{v}_0^{\perp m} + a_{m+1}^{m+1} \cdot \boldsymbol{v}_{m+1}^{\perp m}. \end{aligned} \tag{4.43}$$

Der neue Entwicklungskoeffizient a_{m+1}^{m+1} wird als optimal angesehen, wenn die Norm

$$\left\| \boldsymbol{v}_0^{\perp m+1} \right\| = \left\| \boldsymbol{v}_0^{\perp m} + a_{m+1}^{m+1} \cdot \boldsymbol{v}_{m+1}^{\perp m} \right\| \tag{4.44}$$

minimal wird. Dies kann erreicht werden, indem der parallele Anteil von $\boldsymbol{v}_{m+1}^{\perp m}$ zu $\boldsymbol{v}_0^{\perp m}$ in (4.44) von $\boldsymbol{v}_0^{\perp m}$ abgezogen wird, womit sich der optimale Koeffizient ergibt:

$$\left\| \boldsymbol{v}_0^{\perp m+1} \right\| \to \min \qquad \Longrightarrow \qquad a_{m+1}^{m+1} = -\frac{\left\langle \boldsymbol{v}_0^{\perp m} \,\middle|\, \boldsymbol{v}_{m+1}^{\perp m} \right\rangle}{\left\| \boldsymbol{v}_{m+1}^{\perp m} \right\|^2}. \tag{4.45}$$

Wird der Koeffizient a_{m+1}^{m+1} in der letzten Gleichung von (4.43) verwendet, so kann $\boldsymbol{v}_0^{\perp m+1}$ unmittelbar erhalten werden. Neben $\boldsymbol{v}_0^{\perp m+1}$ läßt sich auch $\boldsymbol{v}_{m+2}^{\perp m+1}$ ermitteln. Für $\boldsymbol{v}_{m+2}^{\perp m+1}$ gilt mit (4.42):

$$\boldsymbol{v}_{m+2}^{\perp m+1} = \boldsymbol{v}_{m+2} - \textstyle\sum_{k=1}^{m+1} b_k^{m+1} \boldsymbol{v}_k = z v^{-1} (\boldsymbol{v}_{m+1} - \sum_{k=1}^{m+1} b_k^{m+1} \boldsymbol{v}_{k-1}). \tag{4.46}$$

Mit (4.39) und (4.34) folgt:

$$b_1^{m+1} = a_{m+1}^{m+1} \qquad \text{und} \qquad b_k^{m+1} = a_{m+2-k}^{m} - a_{m+1}^{m+1} \cdot a_{k-1}^{m} \quad \text{für} \quad k = 2 \dots m+1, \tag{4.47}$$

womit $\boldsymbol{v}_{m+2}^{\perp m+1}$ geschrieben werden kann zu

$$\boldsymbol{v}_{m+2}^{\perp m+1} = z v^{-1} (\boldsymbol{v}_{m+1} - \textstyle\sum_{k=2}^{m+1} (a_{m+2-k}^{m} - a_{m+1}^{m+1} \cdot a_{k-1}^{m}) \boldsymbol{v}_{k-1} + a_{m+1}^{m+1} \boldsymbol{v}). \tag{4.48}$$

Mittels (4.34) und Erniedrigen des Index k in den Summen ergibt sich

$$\boldsymbol{v}_{m+2}^{\perp m+1} = zv^{-1}(\boldsymbol{v}_{m+1} - \textstyle\sum_{k=1}^{m} b_k^m \boldsymbol{v}_k + a_{m+1}^{m+1}(\boldsymbol{v} - \sum_{k=1}^{m} a_k^m \boldsymbol{v}_k)), \tag{4.49}$$

wobei die rechte Seite mittels $\boldsymbol{v}_0^{\perp m}$ und $\boldsymbol{v}_{m+1}^{\perp m}$ ausgedrückt werden kann. Daraus folgt

$$\begin{aligned} & & \boldsymbol{v}_{m+2}^{\perp m+1} &= zv^{-1}(\boldsymbol{v}_{m+1}^{\perp m} + a_{m+1}^{m+1}\boldsymbol{v}_0^{\perp m}), \\ & \text{und zusammen mit} & \boldsymbol{v}_0^{\perp m+1} &= \boldsymbol{v}_0^{\perp m} + a_{m+1}^{m+1} \cdot \boldsymbol{v}_{m+1}^{\perp m} \end{aligned} \tag{4.50}$$

aus (4.43) ergeben sich die benötigten Beziehungen, womit die orthogonalen Anteile mit $m+1$ Vektoren aus den orthogonalen Anteilen mit m Vektoren bestimmt werden können. Die neuen Koeffizienten a_k^{m+1} mit $k = 1 \ldots m+1$ können mit (4.45) und (4.39) durch die orthogonalen Anteile ermittelt werden. Zu Beginn der Schätzung ist der Fehler $\boldsymbol{v}_0^{\perp 0}$ maximal und gleich dem zu approximierenden Vektor selbst, da noch kein Vektor für die Linearkombination $\widehat{\boldsymbol{v}}$ zur Verfügung steht. Dasselbe gilt für $\boldsymbol{v}_1^{\perp 0}$ nach (4.42), womit gilt:

$$\boldsymbol{v}_0^{\perp 0} = \boldsymbol{v}\ , \qquad \boldsymbol{v}_1^{\perp 0} = \boldsymbol{v}_1 = zv^{-1}(\boldsymbol{v}). \tag{4.51}$$

Durch (4.45) wird der letzte Koeffizient bestimmt, wodurch die anderen Koeffizienten nicht mehr explizit auftreten. Daher werden die letzten Koeffizienten a_k^k, die jeweils zu der Lösung mit Verwendung von k Vektoren gehören, als einen neuen Koeffizientensatz definiert. Diese erhalten mit

$$r_k = a_k^k \tag{4.52}$$

eine eigene Bezeichnung. Mit (4.51) als Anfangsbedingungen lassen sich mittels (4.45) und (4.39) für $m = 1 \ldots N$ alle Koeffizienten r_k bzw. a_k^k nacheinander bestimmen zusammen mit der Berechnung der orthogonalen Anteile durch die Formeln (4.50). Damit läßt sich der zu $\mathrm{Lin}\{\boldsymbol{v}_1 \ldots \boldsymbol{v}_N\}$ parallele Anteil von $\boldsymbol{v}$ als Entwicklung nach den Vektoren $\boldsymbol{v}_k^{\perp k-1}$ darstellen mit den Entwicklungskoeffizienten r_k:

$$\widehat{\boldsymbol{v}} = \sum_{k=1}^{N} a_k \cdot \boldsymbol{v}_k = \sum_{k=1}^{N} r_k \cdot \boldsymbol{v}_k^{\perp k-1}. \tag{4.53}$$

Die Vektoren $\boldsymbol{v}_k^{\perp k-1}$ sind im Gegensatz zu den Vektoren $\boldsymbol{v}_k$ orthogonal zueinander und stellen somit eine orthogonale Basis von $\mathrm{Lin}\{\boldsymbol{v}_1 \ldots \boldsymbol{v}_N\}$ dar, falls die Vektoren $\boldsymbol{v}_k$ linear unabhängig sind. Bei der Berechnung des Koeffizienten r_{k+1} bleiben die ersten Koeffizienten r_1 bis r_k unverändert infolge der Orthogonalität der Vektoren $\boldsymbol{v}_k^{\perp k-1}$ untereinander, während die Koeffizienten a_1 bis a_k durch die Berechnung des zusätzlichen Koeffizienten a_{k+1} in der Regel verändert werden müssen. Die Bestimmung der Parameter r_k kann als Graph dargestellt werden, wie in Bild 4.1 zu sehen ist. Gemäß (4.45) werden die Koeffizienten r_k durch das Skalaprodukt

$$r_{k+1} = -\frac{\left\langle \boldsymbol{v}_0^{\perp k} \middle| \boldsymbol{v}_{k+1}^{\perp k} \right\rangle}{\left\| \boldsymbol{v}_{k+1}^{\perp k} \right\|^2} \tag{4.54}$$

berechnet beginnend mit dem Koeffizienten r_1. Da $\widehat{\boldsymbol{v}}$ als Prädiktor für die Werte in $\boldsymbol{v}$ angesehen wird, stellen die orthogonalen Anteile $\boldsymbol{v}_0^{\perp m}$ und $\boldsymbol{v}_{m+1}^{\perp m}$ durch (4.42) die Schätzfehler zu den wahren Werten in den Vektoren $\boldsymbol{v}$ und $\boldsymbol{v}_{m+1}$ dar. Daher kann der

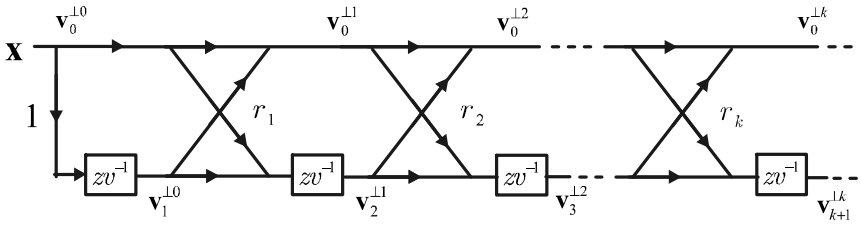


Bild 4.1: Signalflußgraph für die Verarbeitung der orthogonalen Anteile mit Kreuzgliedern.

Fehler der Vorwärtsprädiktion $\boldsymbol{e}_k^v$ und der Fehler der Rückwärtsprädiktion $\boldsymbol{e}_k^r$ durch die orthogonalen Anteile ausgedrückt werden:

$$\boldsymbol{e}_m^v = \boldsymbol{v}_0^{\perp m} = \boldsymbol{v} - \sum_{k=1}^{m} a_k^m \boldsymbol{v}_k \tag{4.55}$$

$$\boldsymbol{e}_m^r = zv^{+1}(\boldsymbol{v}_{m+1}^{\perp m}) = \boldsymbol{v}_m - \sum_{k=1}^{m} b_k^m \boldsymbol{v}_{k-1}. \tag{4.56}$$

Dabei ist zu beachten, daß die Rückwärtsprädiktion üblicherweise die Werte in $\boldsymbol{v}_m$ schätzt, so daß auf $\boldsymbol{v}_{m+1}^{\perp m}$ noch eine zyklische Verschiebung angewandt wird. Der Graph in Bild 4.2 mit den Prädiktionsfehlervektoren ist mit dem Graph von Bild 4.1 mit den orthogonalen Anteilen nahezu identisch, nur daß einmal das Signal vor der zyklischen Verschiebung zv^{-1} und einmal das dahinter liegende Signal betrachtet wird . Die

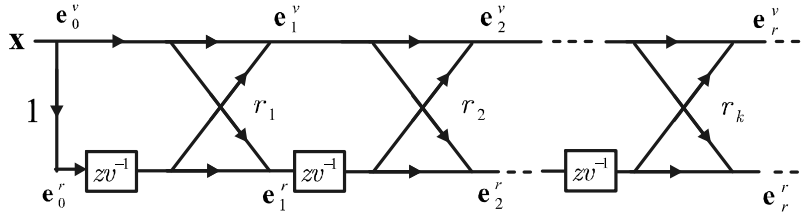


Bild 4.2: Signalflußgraph für die Verarbeitung der Prädiktionsfehler.

Berechnungsformel für die Koeffizienten r_k ergibt sich entsprechend (4.54) zu

$$r_{k+1} = -\frac{\langle \boldsymbol{e}_k^v \,|zv^{-1}(\boldsymbol{e}_k^r)\rangle}{\|zv^{-1}(\boldsymbol{e}_k^r)\|^2}. \tag{4.57}$$

Diese Berechnungen entsprechen der Burg-Methode [Bu68], angepaßt auf periodische Signale. Statt statistischer Größen und Operationen werden hier vektorielle Operationen und Eigenschaften unter Einbeziehung des Skalarproduktes benutzt. Durch die Verwendung der zyklischen Verschiebung können die Schätzungen optimal und eindeutig durchgeführt werden, was durch die ausschnittweise Betrachtung von unendlich langen Signalen in der Regel nicht gegeben ist. Dadurch ergeben sich für bestimmte Modifikationen der Schätzformel (4.45), wie z.B. mittels des Itakura-Koeffizienten, keine unterschiedlichen Resultate. Weiterhin ist das Schätzergebnis nur vom Betragsspektrum der analysierten Periode abhängig. Die Phase hat überhaupt keinen Einfluß auf das Ergebnis. Dies ist dadurch bedingt, da durch die zyklische Verschiebung die

Randwerte an den Perioden für die weitere Berechnungsstufe nicht verloren gehen. Diese vektorielle Betrachtungsweise der linearen Prädiktion, angepaßt auf die Analyse periodischer Signale, findet grundlegende Verwendung für die Schätzalgorithmen der inversen Filterung im Zeitbereich, die im Abschnitt "Parameterbestimmung durch Verwendung von suboptimalen Lösungen" vorgestellt werden. Bei der Verarbeitung stimmhafter Sprachsignale muß angemerkt werden, daß die stationären Sprachsignalabschnitte infolge von Rauschen und Fluktuationen des Anregungsmechanismus nicht exakt periodisch sind.

4.2 Rohrsysteme

Im Gegensatz zur Behandlung von allgemeinen Pol-Nullstellen Modellen werden bei der Parameterbestimmung von Rohrsystemen zwei Ziele gleichzeitig verfolgt. Erstens soll der geschätzte Betragsgang das Sprachspektrum approximieren und zweitens sollen die geschätzten Querschnittsflächen mit den tatsächlichen Flächen des Sprechtraktes Übereinstimmungen aufweisen. Die zur analysierten Sprachaufnahme gehörigen Sprechtraktflächen sind unbekannt, so daß für den Vergleich Flächen aus Röntgen- oder NMR-Aufnahmen aus der Literatur verwendet werden. Für die Schätzung der Flächen stellt sich die grundlegende Frage, ob es überhaupt möglich ist die Querschnittsflächen eindeutig aus dem Sprachsignal oder aus der Lippenimpedanz zu schätzen. Diese Fragestellung wird unter anderem in [Srö67, Mer67, Str75a, Str76, At78, So79, Wa79, Schr94] diskutiert. Geschätzte Vokaltraktflächen aus der Impedanz an den Lippen sind in [Srö67] gezeigt, welche recht glaubwürdig erscheinen. Die Verwendung der Impedanz ist auch in [So79] teilweise favorisiert, wobei die Vorgehensweise den Nachteil hat, daß sie das Sprachsignal nur indirekt berücksichtigt. Für eine Schätzung aus dem Sprachsignal sind häufig die Formanten verwendet worden, wie in [Mer67], wobei neben den Frequenzen der Formanten auch die Bandbreiten benötigt werden [Wa79]. Die Formanten stellen nur einen Teil der Information des Sprachsignals dar, woraus sich Probleme mit der Eindeutigkeit ergeben können. In [At78] ist gezeigt, daß unterschiedliche Querschnittsflächenverläufe dieselben Formanten erzeugen können, wodurch die Eindeutigkeit nicht mehr gegeben ist. Bei diesen Beispielen waren allerdings die Systemfunktionen, bis auf die Formanten, nicht exakt identisch. In der Literatur werden auch Beispiele aufgezeigt, die für unterschiedliche Flächenverläufe die gleiche Übertragungsfunktion ergeben. Dies ist allerdings nur bei zwei Rohrabschlüssen vom Betrage Eins vorgestellt worden, was eigentlich kein realistischer Fall ist, da die Reflexionsfaktoren infolge der Verluste vom Betrage kleiner als Eins sein sollten. Selbst den Glottisabschluß in der geschlossenen Glottisphase wird man vom Betrage kleiner als Eins annehmen. Voraussetzung für die Eindeutigkeit ist, daß der Lippenabschluß fest vorbestimmt ist. Für eine Schätzung des Sprechtraktes muß aus dem Sprachsignal der Einfluß der Abstrahlung und Anregung separiert werden, dessen Problematik in [Str76, So79] angesprochen wird, da die zu separierenden Größen nicht exakt bekannt sind. In [Schr94] wird versucht die Probleme der Eindeutigkeit zu lösen, indem die zeitliche Entwicklung der Flächen einer ganzen Lautsequenz geschätzt wird. Es werden dabei nur die Flächen als realistisch erachtet, die eine kontinuierliche Sprechtraktbewegung ermöglichen.

In dieser Arbeit geht in die Schätzung des Rohrmodells die gesamte Information des Betragsspektrums der analysierten Sprachlaute ein, so daß die Probleme der Ein-

deutigkeit, wie sie in [At78] beschrieben sind, nicht unmittelbar auftreten.

4.2.1 Vorfilterung des Sprachsignals für die Analyse des Sprechtraktes

Die Parameter des Rohrmodells sollen aus dem Sprachsignal bestimmt werden, wobei das Rohrmodell nur den Sprechtrakt darstellen soll. Der Sprechtrakt besteht aus dem Vokaltrakt und gegebenenfalls auch aus dem Nasaltrakt bei gesenktem Velum. Im Sprachproduktionsprozeß existieren noch weitere Einflußgrößen, die das Sprachspektrum beeinflussen. Dazu gehören das stimmhafte Anregungssignal g der Glottis bei Phonation und die Abstrahlungscharakteristik R der Lippen bzw. Nasenlöcher. Damit das geschätzte Rohrmodell nur den Einfluß des Vokal- und Nasaltraktes nachbildet, muß für die Schätzung der Einfluß der Anregung und Abstrahlung aus dem Sprachsignal s separiert werden. Dafür wird angenommen, daß das Sprachsignal s im Spektralbereich als Multiplikation von Anregung G, Sprechtrakt H und Abstrahlung R dargestellt werden kann

$$S = G \cdot H \cdot R. \tag{4.58}$$

Zu beachten ist, daß die einzelnen Systeme nicht unabhängig voneinander sind. Die unterschiedliche Abschlußimpedanz des Vokaltraktes bei verschiedenen Lauten wirkt sich zum Beispiel auf das selbstschwingende System der Stimmbänder aus, und damit auf das Glottissignal [Al85, Ba94, Fa86a, Fa86b, Ti97]. Es wird vereinfachend angenommen, daß die spektrale Einhüllende von GR keine Resonanzen innerhalb des Frequenzbereichs aufweist, sondern nur einen monotonen Abfall zu hohen Frequenzen hin beschreibt. Das Übertragungsverhalten des Sprechtraktes hingegen weist Resonanzen auf, besitzt dafür aber keinen starken spektralen Abfall oder Anstieg über den gesamten Frequenzbereich hinweg. Diese Aufteilung des Spektrums in Resonanzen und spektralen Abfall läßt sich durch komplexe und reelle Pole vollziehen. Die komplexen Polstellen korrespondieren zu den Resonanzen, während die reellen Pole nur einen monotonen spektralen Abfall modellieren können. Ein spektraler Abfall läßt sich praktisch auch durch eine komplexe Resonanz nahe bei der Frequenz Null beschreiben, was allerdings für diese Betrachtungsweise hier nicht herangezogen wird. Das Modell der Anregung und Abstrahlung GR enthält daher zweckmäßigerweise nur reelle Pole:

$$G \cdot R = \prod_{i=1}^{M} \frac{1}{1 - k_i^p \cdot z^{-1}}. \tag{4.59}$$

Eine Schätzung des Sprachsignals mit reellen Polen modelliert nur den Abfall des stimmhaften Sprachspektrums und damit den Effekt der Anregung und Abstrahlung. Um diesen Einfluß zu beseitigen wird das Sprachsignal mit den reellen Nullstellen $\frac{1}{GR}$ gefiltert. Dazu wird eine Burg-Minimierung erster Ordnung durchgeführt. Im Gegensatz zu [Ge96, GeL96] wird hier die Burg-Methode mit der Filterung $1 - k_i^p \cdot z^{-1}$ wiederholt angewendet mit $i = 1 \ldots M$, so daß mehrere reelle Polstellen geschätzt werden. Die Burgmethode erster Ordnung wird bis zu dreimal wiederholt. Falls ein Koeffizient k_i^p vom Betrag negativ ist, kann abgebrochen werden, da dieser einen Anteil mit einem spektralen Anstieg darstellt. Das gefilterte Sprachsignal s beinhaltet dann

überwiegend nur noch den Einfluß des Sprechtraktes H:

$$H = \frac{S}{GR} \sim \prod_{i=1}^{M} \left(1 - k_i^p \cdot z^{-1}\right) S \quad \text{mit} \quad M \leq 3,$$

wobei die Grundfrequenz außer Acht gelassen wird. Das Linienspektrum der stimmhaften Anregung wird durch diese Filterung nicht beeinflußt. Die Präemphasekoeffizienten k_i^p hängen von Faktoren wie dem Mikrophonabstand zum Mund und dem artikulierten Laut selbst ab. Die Lautabhängigkeit kann damit erklärt werden, daß die Abstrahlung von der Lippenöffnungsfläche abhängt und eine um so stärkere Hochpaß-Charakteristik aufweist, desto kleiner die Mundöffnung ist, vgl. (3.64) und (3.66). Dieser Zusammenhang läßt sich mit der Abhängigkeit des Präemphasekoeffizienten zur Lippenrundung in [GeL96] bestätigen. Das Spektrum bzw. die spektrale Einhüllende G des Glottissignals weist eine Tiefpaß-Charakteristik auf, die eine viel stärkere Ausprägung besitzt als die Hochpaß-Charakteristik der Abstrahlung, wodurch GR insgesamt einen Tiefpaß beschreibt. Bild 4.3 zeigt den geschätzten spektralen Abfall, welcher durch eine dreifach hintereinander ausgeführte Burgmethode erster Ordnung von einer Sprachsignalperiode des Nasals /n/ gewonnen wurde.

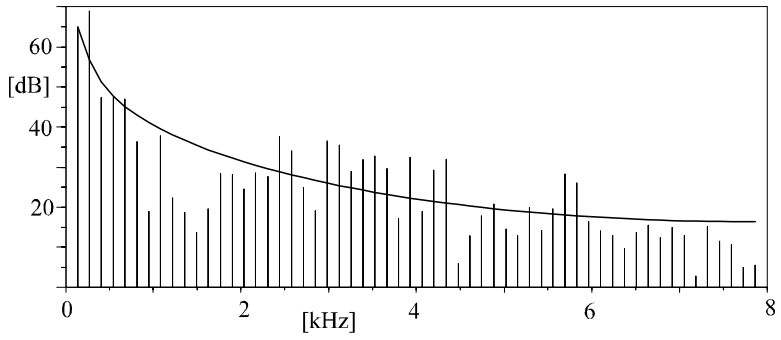


Bild 4.3: Schätzung mit reellen Polen: Adaptive Präemphase (durchgezogene Linie) und DFT des analysierten Nasals /n/ (Linienspektrum).

4.2.2 Parameterbestimmung von linearen zeitinvarianten Rohrsystemen

Für die modellbasierte Analyse werden die Modellparameter p_k in der Weise eingestellt, daß der Betragsgang des untersuchten Modells das Betragsspektrum des analysierten Signals $x(n)$ approximiert. Die Abweichung der Modellierung wird mathematisch durch eine Fehlerfunktion $e(p_k, x)$ definiert, welche mittels eines Optimierungsalgorithmus minimiert wird. Bei den Untersuchungen hat sich herausgestellt, daß der Herleitung der Fehlerdefinition, welche für erweiterte Rohrmodelle und Systeme verwendet wird besondere Aufmerksamkeit geschenkt werden muß.

Definition der Fehlerfunktion

Die hier verwendete Fehlerfunktion wird mit einer Modifikation aus dem Prinzip der inversen Filterung bzw. der linearen Prädiktion abgeleitet. Die inverse Filterung kann

für allgemeine Systeme nur unter bestimmten Voraussetzungen ohne Modifikationen erfolgreich verwendet werden. Bei der inversen Filterung wird die Ausgangsleistung des Analysefilters minimiert, welches das inverse Synthesesystem darstellt. Die Übertragungsfunktion $H(z)$ des linearen zeitinvarianten Synthesefilters kann allgemein mit

$$H(z) = \frac{b_0 + \sum_{i=1}^{M} b_i z^{-1}}{a_0 + \sum_{i=1}^{N} a_i z^{-1}}, \qquad \text{bzw.} \quad H''(z) = \frac{b_0'' + \sum_{i=1}^{M} b_i'' z^{-1}}{1 + \sum_{i=1}^{N} a_i'' z^{-1}} \tag{4.60}$$

dargestellt werden, wobei der Koeffizient a_0 gleich Eins gesetzt werden kann. Die Polynomkoeffizienten selbst sind wieder von den Modellparametern abhängig, wie zum Beispiel den Reflexionskoeffizienten. Die Ausgangsleistung des inversen Filters, welches das Fehlersignal repräsentiert, kann mit Hilfe des Parsevalschen Theorems im Frequenzbereich formuliert werden:

$$e = \frac{1}{2\pi} \int_0^{2\pi} \left| \frac{X(\mathrm{e}^{\mathrm{j}\omega})}{H(\mathrm{e}^{\mathrm{j}\omega})} \right|^2 \mathrm{d}\omega. \tag{4.61}$$

Das Integral der Fehlerdefinition (4.61) läßt sich für die Analyse von Zeitabschnitten endlicher Länge als Summe darstellen:

$$e = \frac{1}{N} \sum_{k=0}^{N-1} \left| \frac{X(\mathrm{e}^{\mathrm{j}2\pi k/N})}{H(\mathrm{e}^{\mathrm{j}2\pi k/N})} \right|^2. \tag{4.62}$$

Der Faktor $2\pi k/N$ im Exponenten stellt die Frequenzmoden der Fensterbreite N dar. Die Summanden in (4.62) müssen für eine Berechnung nur bis zur Hälfte von N aufsummiert werden, da das Spektrum für reellwertige Signale symmetrisch ist. In früheren Untersuchungen [Sn96, SnL97] stellte sich heraus, daß für verzweigte Rohrmodelle diese Fehlerdefinition e zu schlechten Ergebnissen führt. Für die Parameterbestimmung selbst wird der Fehler e mit einem gradientenbasierten Optimierungsverfahren minimiert. Die Einbeziehung der Dreitor-Parameter verschlechterten die Ergebnisse teilweise sehr stark, obwohl der Fehler e gleichzeitig abnimmt. Dies bedeutet, daß der Wert des Fehlers e von (4.61) kein adäquates Gütemaß für die Modellierung von erweiterten Rohrmodellen darstellt. Daher muß die Fehlerdefinition (4.61) modifiziert werden. Die erste modifizierte Version des Fehlers e wird im Folgenden gezeigt, dessen Analyseergebnisse in [Sn96, SnL97] vorgestellt sind. Für die Fehlerdefinition werden Signale mit endlicher Länge betrachtet. Die Definition des Fehlers wird durch eine Fallunterscheidung von Einzelfehlern e_z für jede Frequenz individuell zusammengesetzt, welche zu dem resultierenden Gesamtfehler e_a führen:

$$\begin{aligned}
e_z(k) &= \frac{1}{N} \left| \frac{X(\mathrm{e}^{\mathrm{j}2\pi k/N})}{Y(\mathrm{e}^{\mathrm{j}2\pi k/N})} \right| \qquad \text{falls } \left|X(\mathrm{e}^{\mathrm{j}2\pi k/N})\right| > \left|Y(\mathrm{e}^{\mathrm{j}2\pi k/N})\right| \\
&\quad \text{oder} \\
e_z(k) &= \frac{1}{N} \left| \frac{Y(\mathrm{e}^{\mathrm{j}2\pi k/N})}{X(\mathrm{e}^{\mathrm{j}2\pi k/N})} \right| \qquad \text{falls } \left|X(\mathrm{e}^{\mathrm{j}2\pi k/N})\right| \leq \left|Y(\mathrm{e}^{\mathrm{j}2\pi k/N})\right| \\
e_a &= \sum_{k=0}^{N-1} e_z(k).
\end{aligned} \tag{4.63}$$

Statt H kann hierfür der Systemausgang $y = h * x$ des Synthesefilters verwendet werden, wobei für die Anregung x eine Impulsfolge benutzt wird. Der Gesamtfehler e_a führt zu wesentlich besseren Ergebnissen als die Verwendung von (4.61). Der Fehler wurde in dieser Weise modifiziert, da für (4.61) gilt, daß für eine bestimmte Frequenz $\boldsymbol{\psi}$ bei der $\left|H(\mathrm{e}^{\mathrm{j}\psi})\right|$ größer als $\left|X(\mathrm{e}^{\mathrm{j}\psi})\right|$ ist, der Beitrag zum Fehler (4.61) allein durch ein größeren Wert von $\left|H(\mathrm{e}^{\mathrm{j}\psi})\right|$ kleiner wird, aber nur durch große Werte von $|H|$ das Betragsspektrum $|X|$ nicht modelliert wird. Für unendlich großes $\left|H(\mathrm{e}^{\mathrm{j}\psi})\right|$ für alle Frequenzen $\boldsymbol{\psi}$ ist der Fehler offensichtlich Null für sämtliche im Wertebereich beschränkten Signale, wodurch der Zweck des Fehlermaßes zur Spektralanpassung nicht erfüllt wird. Die Fallunterscheidung berücksichtigt diesen Effekt. In [Sn96, SnL97] ist zu sehen, daß mit Hilfe der Minimierung der Fehlerdefinition e_a die Sprachsignalspektren durch den Betragsgang eines erweiterten Rohrmodells relativ gut modelliert werden. Durch die Fallunterscheidung ist die Fehlerdefinition kein geschlossener Ausdruck. Dies kann für die mathematische Behandlung von Nachteil sein, wenn zum Beispiel der Fehler durch algebraische Gleichungen minimiert werden soll, die durch Ableitung des Fehlermaßes gewonnen werden. Die mathematische Gestalt der Fehlerdefinition e_a deutet darauf hin, daß die Fehlerdefinition theoretisch nicht konsistent ist. Dies soll bedeuten, daß durch den Trick der Fallunterscheidung möglicherweise die Unzulänglichkeiten der Fehlerdefinition (4.61) überdeckt werden; das Problem an sich allerdings nicht vollständig gelöst ist. Dafür wird die Fehlerdefinition der linearen Prädiktion im Folgenden genauer betrachtet. An der Struktur des Prädiktionsfehlerfilters

$$\begin{aligned} P(z) &= 1 - \sum_{i=1}^{N} a_i z^{-i} \\ &= \sum_{i=0}^{N} a_i z^{-i} \end{aligned}$$

ist auffällig, daß der Koeffizient a_0 gleich Eins ist und damit als bekannt vorausgesetzt wird, während alle anderen Koeffizienten frei geschätzt werden. Wird der Koeffizient a_0 in $P(z)$ wie die anderen Koeffizienten für die Parameterbestimmung geschätzt, ergibt sich keine sinnvolle Schätzung des Nur-Pole Modells. Dieser Sachverhalt gibt die Motivation die Fehlerdefinition (4.61) in der Weise zu ändern, daß die Nennerstruktur des Analysefilters der Struktur von $P(z)$ angepaßt wird. Gleiches gilt für das Zählerpolynom des Analysefilters, wenn es nicht schon wie in H''^{-1} diese Struktur aufweist. Diese Modifikation stellt sich in vielen Analysen von Test- und Sprachsignalen als erfolgreich heraus. Die Veränderung der Fehlerdefinition kann auch von einer anderen Sichtweise motiviert werden. Entwickelt man das Nennerpolynom des inversen Systems H^{-1} durch eine Reihe von Potenzen z^{-i} für $i \geq 0$, so ergibt sich:

$$\begin{aligned} \frac{1}{b_0 + \sum_{i=1}^{M} b_i z^{-1}} &= \sum_{i=0}^{\infty} c_i z^{-i} \\ &= \frac{1}{b_0} + \sum_{i=1}^{\infty} c_i z^{-i}. \end{aligned} \tag{4.64}$$

Die Koeffizienten c_i auf der rechten Seite stellen die Impulsantwort des rein rekursiven Teils von H^{-1} dar. Der konstante Term $c_0 = b_0^{-1}$ ergibt sich aus der oberen Gleichung

in (4.64) für einen Grenzübergang von $|z|$ gegen Unendlich. Mit (4.64) läßt sich das inverse System von (4.60) auch als Reihe darstellen:

$$H^{-1}(z) = \sum_{i=0}^{\infty} d_i z^{-i} = \frac{a_0}{b_0} + d_1 z^{-1} + d_2 z^{-2} + \ldots \quad . \tag{4.65}$$

Aus einem Koeffizientenvergleich mit $P(z)$ folgt, daß a_0/b_0 in (4.65) und b_0 in (4.64) gleich Eins sind, wenn die Systeme die Struktur eines Prädiktors aufweisen sollen. Um das inverse Filter H^{-1} daran anzupassen, muß der Faktor b_0/a_0 hinzugefügt werden. Das korrigierte inverse Filter H_k^{-1} ist damit definiert durch:

$$H_k^{-1} = \frac{b_0}{a_0} \cdot \frac{a_0 + \sum_{i=1}^{N} a_i z^{-1}}{b_0 + \sum_{i=1}^{M} b_i z^{-1}} = \frac{1 + \sum_{i=1}^{N} a'_i z^{-1}}{1 + \sum_{i=1}^{M} b'_i z^{-1}}. \tag{4.66}$$

Für die Parameterbestimmung sollte folglich nicht die Ausgangsleistung des inversen Filters H^{-1} minimiert werden, sondern die des korrigierten Analysefilters H_k^{-1}. Die Fehlerdefinition des allgemeinen Systems (4.60) ergibt sich mit der Modifikation [SnL98a] zu

$$e_k = \frac{1}{2\pi} \left| \frac{b_0}{a_0} \right|^2 \cdot \int_0^{2\pi} \left| \frac{X(\mathrm{e}^{\mathrm{j}\omega})}{H(\mathrm{e}^{\mathrm{j}\omega})} \right|^2 \mathrm{d}\omega. \tag{4.67}$$

Der angehängte Index k in e_k gibt an, daß die Fehlerdefinition im Gegensatz zu (4.61) korrigiert ist. Der Korrekturterm $|b_0/a_0|^2$ vor dem Integral ist keine einfache Konstante. Stellt H ein verzweigtes oder ein anderes erweitertes Rohrmodell dar, so sind die Polynomkoeffizienten $a_0(p_k)$ und $b_0(p_k)$ Funktionen der Rohrparameter p_k. Das Integral in der Fehlerdefinition (4.67) läßt sich für Signalabschnitte x der Länge N als Summe darstellen:

$$e_k = \frac{1}{N} \left| \frac{b_0}{a_0} \right|^2 \cdot \sum_{k=0}^{N-1} \left| \frac{X(\mathrm{e}^{\mathrm{j}2\pi k/N})}{H(\mathrm{e}^{\mathrm{j}2\pi k/N})} \right|^2 . \tag{4.68}$$

Die Fehlerdefinition als endliche Summe (4.68) kann bei Kenntnis des Betragsganges und der Koeffizienten a_0 und b_0 berechnet werden. An der Fehlerdefinition ist zu erkennen, daß sich nur die Beträge der Spektralwerte auswirken, während die Phase unberücksichtigt bleibt. Dies läßt sich dadurch erklären, daß die Fehlerdefinition auf die lineare Prädiktion zurückgeführt werden kann, in der die Autokorrelationswerte auftreten. Die Autokorrelationswerte lassen sich wiederum aus dem Betragsspektrum mit Hilfe des Wiener-Kintschin Theorems ableiten, wofür das Betragsspektrum mit dem Leistungsdichtespektrum gleichgesetzt werden muß. Ist das Signal durch Anregung eines linearen physikalischen Prozesses entstanden, so kann dieser durch Resonatoren und Antiresonatoren beschrieben werden, durch dessen Betragsgang ein bestimmter Phasengang zugehörig ist. Werden durch den Schätzalgorithmus die Pole und Nullstellen richtig geschätzt, so ist dadurch gleichzeitig der Phasengang des Modells korrekt, bedingt durch die Beziehung zwischen Betrags- und Phasengang. Hierbei wird vorausgesetzt, das die Lösungen der Modellschätzungen minimalphasig sind. (4.67) und (4.68) erfüllt nicht die mathematische Definition eines Maßes bzw. einer Abstandsfunktion einer Metrik, da das Minimum nicht bei Null liegt für $X = H$. Bei der Minimierung spielt das allerdings keine große Rolle. Bedeutender könnte die Tatsache sein, daß die Fehlerdefinition nicht symmetrisch ist in Bezug auf die beiden Argumente X und H.

Fehlerminimierung

Für die Parameterbestimmung kann eine Fehlerminimierung durch allgemeine Optimierungsalgorithmen erreicht werden, wofür hier ein gradientenbasiertes Verfahren verwendet wird. Die Ableitungen des Gradienten $\nabla \boldsymbol{e}_k(\boldsymbol{p})$ werden durch kleine Variationen $\boldsymbol{\varepsilon}$ der Parameterwerte p_k angenähert:

$$\frac{\partial \boldsymbol{e}_k(\boldsymbol{p})}{\partial p_k} \simeq \frac{\boldsymbol{e}_k(\boldsymbol{p}+\boldsymbol{\varepsilon} \cdot \boldsymbol{p}_k)}{\boldsymbol{\varepsilon}} \qquad \text{mit} \qquad \boldsymbol{p} = (p_1, p_2, \ldots p_N)^{\mathrm{T}} \,. \tag{4.69}$$

Der Parametervektor $\boldsymbol{p}$ wird schrittweise so weit in die negative Richtung des Gradienten verschoben bis der Fehler ansteigt. Dies wird zuerst mit einer großen Schrittweite so oft wie möglich durchgeführt. Danach wird mit immer kleineren Schrittweiten fortgefahren. Bei der kleinsten vorgegebenen Schrittweite endet eine Iteration. Mit der nächsten Iteration wird der Gradient neu geschätzt, so daß der Parametervektor in eine neue Richtung schreiten kann. Als Modellparameter werden die Reflexionskoeffizienten der Zweitore und die Parameter $\boldsymbol{\rho}_i$ der Dreitore von (3.60) verwendet. Bei Veränderungen des Parametervektors $\boldsymbol{p}$ ist zu achten, daß der Wertebereich der Parameter konsistent mit dem Modell eingehalten wird. Die Reflexionskoeffizienten müssen daher vom Betrag kleiner als Eins sein, die beiden Dreitorparameter $\boldsymbol{\rho}_1$ und $\boldsymbol{\rho}_2$ müssen positiv sein und die Summe $\boldsymbol{\rho}_1 + \boldsymbol{\rho}_2$ darf nicht größer als Zwei werden, da sich sonst negative Flächen ergeben würden. Während die Reflexionskoeffizienten einzeln überprüft werden können, müssen die Dreitorparameter paarweise kontrolliert werden. Dies wurde im Computerprogramm berücksichtigt, indem eine Schnittstellenklasse definiert wird, die die Verbindung vom Rohrmodell zu einer Optimierungsklasse herstellt. In dieser Schnittstelle werden neben dem Parametervektor $\boldsymbol{p}$ in Form von Zeigern auch die Art der Koeffizienten übermittelt, aus denen die Wertebereiche abgeleitet werden können. In einer Tabelle eines GUI-Fensters können die einzelnen Werte p_k von $\boldsymbol{p}$ und der Wertebereiche nach durchgeführten Iterationen beobachtet und verändert werden. Insbesondere können dadurch auch einzelne Parameter während der Optimierung konstant gehalten oder gegebenenfalls neu eingestellt werden. Dies wird unter anderem benötigt, falls ein Rohrabschluß durch einen Reflexionskoeffizienten im Vektor $\boldsymbol{p}$ dargestellt wird. Auch die Startkonfiguration der Werte von $\boldsymbol{p}$ kann dadurch leicht auf beliebige Werte gesetzt werden. Für die Analyse von Rohrmodellen, muß zuerst eine Rohrstruktur ausgewählt werden, und dann ein Startwert für den Parametervektor $\boldsymbol{p}$ bestimmt werden, der die Rohrkonfiguration darstellt. Im Gegensatz zur Rohrkonfiguration beschreibt die Rohrstruktur hier die allgemeine Struktur des Rohrsystems unter Verwendung von Verzweigungen, Torelementen und Positionen der Ein- und Ausgänge, womit die Querschnittsflächen noch unbestimmt bleiben. Erst durch die Rohrkonfiguration werden die Werte der Flächen einer Rohrstruktur bestimmt. Für den Optimierungsalgorithmus bleibt die Rohrstruktur unverändert bestehen. Die Rohrabschlüsse sind dabei vorgegeben und werden nicht vom Algorithmus verändert. Die Rohrabschlüsse stellen somit für die Optimierung Nebenbedingungen dar.

Analyse von Testsignalen einfach verzweigter Rohrsysteme

Für die Bewertung von Schätzalgorithmen ist die Analyse von Testsignalen hilfreich, da dadurch ermittelt werden kann, ob der Algorithmus unter optimalen Bedingungen das

globale Minimum finden kann. Optimale Bedingungen heißt in diesem Zusammenhang, daß für das Rohrsystem ein Parametervektor existiert, dessen zugehörige Übertragungsfunktion das Betragsspektrum des zu analysierenden Signals perfekt nachbildet. Diese Bedingung ist erfüllt, wenn das zu analysierende Signal durch dieselbe Rohrstruktur erzeugt wurde, welche auch für die Analyse verwendet wird. Im folgenden Beispiel wird ein Testsignal t mittels eines verzweigten Rohrsystems erzeugt, welches mit einer Impulsfolge angeregt wird. Für die Analyse wird ein Zeitabschnitt des Testsignals verwendet, welches eine Ausgangsperiode t_p im eingeschwungenen Zustand darstellt. Der Schätzalgorithmus startet von einer neutralen Rohrkonfiguration, in der alle Rohrflächen gleich groß sind. Dies bedeutet, daß zu Anfang alle Reflexionskoeffizienten Null sind und alle Dreitorparameter den Wert 2/3 besitzen. Im Folgenden werden Resultate von der Analyse des Testsignals t_p gezeigt, welches durch ein Rohrsystem mit einer Verzweigung und einem festen reellen Rohrabschluß am Ausgang erzeugt wurde. Außer dem Torabschluß am Ausgang werden alle Koeffizienten unter Einhaltung der Konsistenz des Rohmodells frei geschätzt. Bild 4.4 zeigt die Ergebnisse der Analyse von t_p nach einer Minimierung des modifizierten Fehlers e_k (4.68) mit dem Gradientenverfahren, beginnend von einer neutralen Startkonfiguration. Wie in Bild 4.4 zu

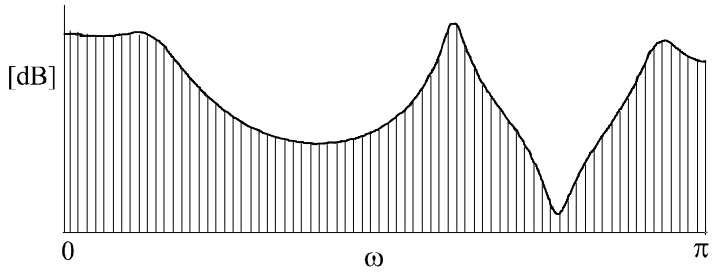


Bild 4.4: Analyse eines Testsignals durch Minimierung der korrigierten Fehlerdefinition e_k aus neutraler Position: Geschätzter Betragsgang (durchgezogene Linie), DFT des Testsignals (Linienspektrum).

sehen ist, stimmen die diskreten Frequenzen von dem Linienspektrum des Testsignals mit denen des geschätzten Betragsgangs nahezu perfekt überein. Der Algorithmus hat das globale Minimum gefunden. Bild 4.5 zeigt im Gegensatz zu Bild 4.4 die Resultate unter Verwendung des Fehlers e ohne Modifikation (4.62) aus derselben neutralen Startkonfiguration beginnend. Es ist zu erkennen, daß das globale Minimum keinesfalls erreicht wird. Da der Algorithmus in ein lokales Minimum geraten sein könnte, startet der Algorithmus in einem weiteren Versuch von einer Startkonfiguration, die die angestrebte optimale Lösung schon darstellt. Das Resultat nach etlichen Iterationen ist in Bild 4.6 zu sehen. Hier ist zu erkennen, daß durch die Minimierung des Fehler e ohne Korrektur der Algorithmus sich sogar aus der optimalen Konfiguration bewegt [SnL98a]. Dieses Konvergenzverhalten aus der optimalen Lösung heraus läßt sich durch die von den Modellparametern abhängigen Polynomkoeffizienten $a_0(p_k)$ und $b_0(p_k)$ erklären. Sind diese beiden Koeffizienten nicht von den zu schätzenden Rohrparametern p_k abhängig, so liefern die Fehlerdefinitionen e und e_k gleich gute Ergebnisse. Sind im Rohrmodell allerdings z.B. Rohrverzweigungen vorhanden, ist dies schon nicht mehr gegeben. Diese Erkenntnis des notwendigen Korrekturfaktors erklärt, warum selbst für unverzweigte Rohre mit reellem Abschluß ± 1 die Rohrelemente des Analysefilters

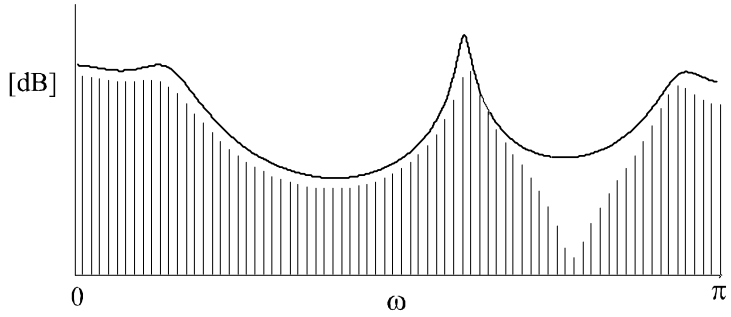


Bild 4.5: Analyse eines Testsignals durch Minimierung der Fehlerdefinition e der inversen Filterung: Geschätzter Betragsgang (durchgezogene Linie), DFT des Testsignals (Linienspektrum).

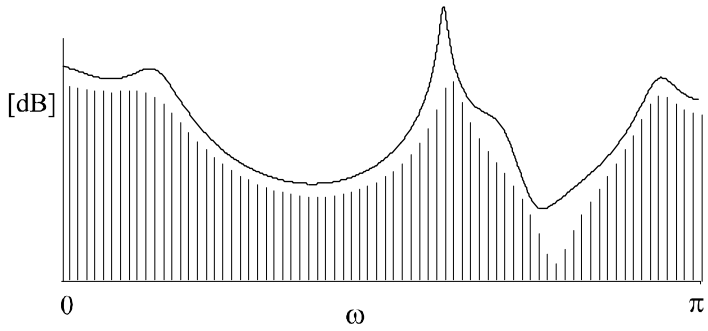


Bild 4.6: Analyse eines Testsignals durch Minimierung der Fehlerdefinition e aus optimaler Position.

immer in der Grundstruktur (siehe Tabelle 3) verwendet werden sollten, obwohl im Synthesefilter eine andere Wellendarstellung verwendet werden kann. Die Betriebskettenmatrizen der Rohrelemente mit Querschnittssprung unterscheiden sich von der Grundform $\boldsymbol{T}_k{}'$ mit

$$\boldsymbol{T}_k{}' = \begin{pmatrix} 1 & r_k \cdot z^{-1} \\ r_k & z^{-1} \end{pmatrix}$$

hinsichtlich der Wellendarstellung durch einen Vorfaktor $d(r_k)$, der in Tabelle 3 für die verschiedenen Wellenformen aufgelistet ist. Nach (3.71) ergibt sich die Übertragungsfunktion für ein unverzweigtes Rohr mit dem Abschluß ± 1:

$$H(z) = 1/(T_g^{11} \pm T_g^{12})$$

mit den Elementen der Matrix $\boldsymbol{T}_g$, welche das Produkt sämtlicher Betriebskettenmatrizen repräsentiert:

$$\boldsymbol{T}_g = \left(\prod_{k=1}^{N} d(r_k) \right) \cdot \boldsymbol{T}_g{}' = \prod_{k=1}^{N} d(r_k) \boldsymbol{T}_k{}'$$

$$\text{mit} \quad \boldsymbol{T}_g{}' = \prod_{k=1}^{N} \boldsymbol{T}_k{}'$$

und bei der der Faktor $\prod_k d(r_k)$ vor die Matrix gezogen werden kann. $\boldsymbol{T}_g'$ stellt dabei das Produkt der Betriebskettenmatrizen in Grundform dar. Das inverse Filter ergibt sich damit zu:

$$\begin{aligned} H^{-1} &= T_g^{11} \pm T_g^{12} \\ &= \left(\prod_{k=1}^{N} d(r_k) \right) \cdot \left(T_g'^{11} \pm T_g'^{12} \right) . \end{aligned}$$

An den beiden oberen Elementen in $\boldsymbol{T}_k'$ ist zu erkennen, daß das Polynom $T_g'^{11}$ in H^{-1} einen konstanten Term mit dem Wert Eins besitzt, während in $T_g'^{12}$ nur Potenzen von z^{-1} vorkommen. Das resultierende Polynom ergibt sich daher zu:

$$H^{-1} = \sum_{i=0}^{N} c_i z^{-i} = \left(\prod_{k=1}^{N} d(r_k) \right) + c_1 z^{-1} + c_2 z^{-2} + \ldots \ .$$

Folglich ergibt sich für den Korrekturfaktor $b_0(r_k)/a_0(r_k)$ von H das Produkt $\prod_k d(r_k)^{-1}$, so daß sich das modifizierte inverse Filter H_k^{-1} nach (4.66) ergibt zu:

$$\begin{aligned} H_k^{-1} &= \left(\prod_{k=1}^{N} d(r_k) \right)^{-1} \cdot H^{-1} = \left(T_g'^{11} \pm T_g'^{12} \right) \\ &= \ 1 \ + c_1' z^{-1} + c_2' z^{-2} + \ldots \ . \end{aligned}$$

Daran ist zu erkennen, daß nach der Modifikation das korrigierte inverse Filter durch die Betriebskettenmatrizen in Grundform $\boldsymbol{T}_k'$ beschrieben ist, obwohl das Synthesefilter eine andere Wellendarstellung aufweist.

Analyse von Rohrmodellen mit frequenzabhängigem Abschluß

In LPC-Modellen in Kreuzgliedstruktur ist der Rohrabschluß an den Lippen mit ± 1 gewählt, wodurch der Lippenabschluß nur für tiefe Frequenzen und kleine Öffungsflächen gut modelliert werden kann. Um eine bessere Modellierung auch für die höheren Frequenzen zu erreichen wird der Rohrabschluß an den Lippen mit dem Impedanzmodell von Laine realisiert, woraus der Abschluß $R_{lip}(z)$ in (3.66) folgt. R_{lip} enthält als Parameter die Lippenöffnungsfläche A. Dieser Flächenparameter muß vor der Analyse für den jeweiligen Laut eingestellt werden und wird durch den Optimierungsalgorithmus nicht mehr verändert. Der Rohrabschluß $R_{lip}(z)$ wird hier mit einem reellen Faktor $\boldsymbol{\alpha}$ mit $0 < \boldsymbol{\alpha} < 1$ multipliziert, der zusätzliche Verluste modelliert. Diese können sich aus den Verlusten innerhalb des Vokaltraktes ergeben. Diese inneren Verluste werden hier durch die verwendeten Betriebskettenmatrizen $\boldsymbol{T}$ nicht berücksichtigt, obgleich sie auch frequenzunabhängig durch Rohrelemente von (3.36) modelliert werden könnten. Um das Modell für die Analysen einfach zu halten, wird hier mit einem Verlustfaktor $\boldsymbol{\alpha}$ am Rohrabschluß operiert, woraus der Lippenabschluß

$$L(z) = \boldsymbol{\alpha} \cdot R_{lip}(z) \tag{4.70}$$

resultiert. $\boldsymbol{\alpha}$ besitzt für die gezeigten Analysen in der Regel Werte, die zwischen 0,85 und 0,95 liegen. Bei den Analysen zeigt sich insgesamt, daß wenn die Verluste höher

angesetzt werden als die theoretischen Literaturwerte, der geschätzte Flächenverlauf oft realistischer ausfällt. Dies könnte damit zusammenhängen, daß für die theoretischen Verlustwerte meist einfache Geometrien angenommen werden. Für die Analyse von Vokalen und Nasalen werden in den folgenden Beispielen Rohrmodelle verwendet, welche den frequenzabhängigen Rohrabschluß $L(z)$ am Systemausgang aufweisen [SnL98b]. Die Schätzung wird durch Minimierung des Fehlermaßes (4.68) mit dem Gradientenverfahren durchgeführt. Die Anfangseinstellung des Parametervektors ist eine neutrale Rohrkonfiguration. Der Lippenabschuß ist für die Schätzung fest voreingestellt und dessen Öffnungsflächenparameter ist den Lauten entsprechend angepaßt. Dadurch weist z.B. der Vokal /i:/ eine größere Öffnungsfläche auf als /u:/. Für die Nasenlöcher wird eine kleine Öffnungsfläche verwendet. Die analysierten stimmhaften Sprachsignale haben eine Abtastrate von 16 kHz und werden jeweils mit einer adaptiven Präemphase vorgefiltert. Wie an den Spektren der analysierten Sprachsignale zu erkennen, sind für diese Analysen längere Zeitabschnitte des stätionären Sprachsignals gewählt, welche mehrere Sprachperioden beinhalten.

Analyse von Vokalen

Bild 4.7 oben zeigt den geschätzten Betragsgang mit Präemphase im Vergleich zum Vokalspektrum. Wie besonders gut beim Vokal /i:/ zu erkennen ist, modelliert der Betragsgang des Rohrmodells in erster Linie den Verlauf der harmonischen Frequenzen der Grundfrequenz. Im höheren Frequenzbereich wird der Sprachlaut zunehmend unperiodisch. Der Verlauf der geschätzten Betragsgänge ist mit denen von LPC-Spektren vergleichbar. Im Gegensatz zum LPC-Modell besitzt hier das Rohrmodell den frequenzabhängigen Rohrabschluß. In Bild 4.7 unten sind zu den Analysen die geschätzten Vokaltraktflächen gezeigt, welche sich mit Flächen aus Röntgen- [Fa70] oder NMR-Aufnahmen [Sto96] vergleichen lassen (siehe Anhang). Der Vokaltrakt ist infolge einer Lippenvorstülpung und Kehlkopfabsenkung für den Vokal /u:/ länger als für den Laut /i:/. Bei den Analysen zeigt sich, daß für den Vokal /u:/ im Gegensatz zu den meisten anderen Vokalen die geschätzten Flächen teilweise nicht so gut ausfallen, da die hintere Vokaltrakthöhle im Rachen sich in den Schätzungen oft nur schwach ausbildet. Dies ist insbesondere der Fall, wenn mit Lippenabschlüssen analysiert wird, die keine oder nur eine geringe Dämpfung aufweisen. Ein ähnlicher Effekt ist in [Srö67] zu sehen.

Analyse von Nasalen durch verzweigte Rohrmodelle mit identischer Nasenstruktur

Für die Analyse der Nasale /m/ und /n/ wird jeweils dieselbe Flächenkonfiguration des Nasaltraktes verwendet [SnL98b]. Die verwendeten Flächen des Nasaltraktes sind aus dem Nasal /N/ geschätzt. Bei diesem Nasal liegt der Verschluß im Mundraum am dichtesten am Velum. Die Nasaltraktflächen sind in Bild 4.8 zu sehen und stellen die ersten dreizehn Flächen jeweils hinter den Nasenlöchern dar. Diese Flächen sind für die Analyse von /m/ und /n/ fest vorgegeben. Die geschätzten Betragsgänge sind für beide Nasale im Bild 4.8 zu sehen. Es ist zu erkennen, daß die Nebenbedingungen der Nasaltraktflächen aus /N/ sich hier für den Nasal /m/ ungünstiger auswirken als für den Nasal /n/. Die geschätzten Flächen des Rachens und der Mundhöhle mit dem vorgegebenen Nasaltrakt sind ebenfalls in Bild 4.8 zu sehen.

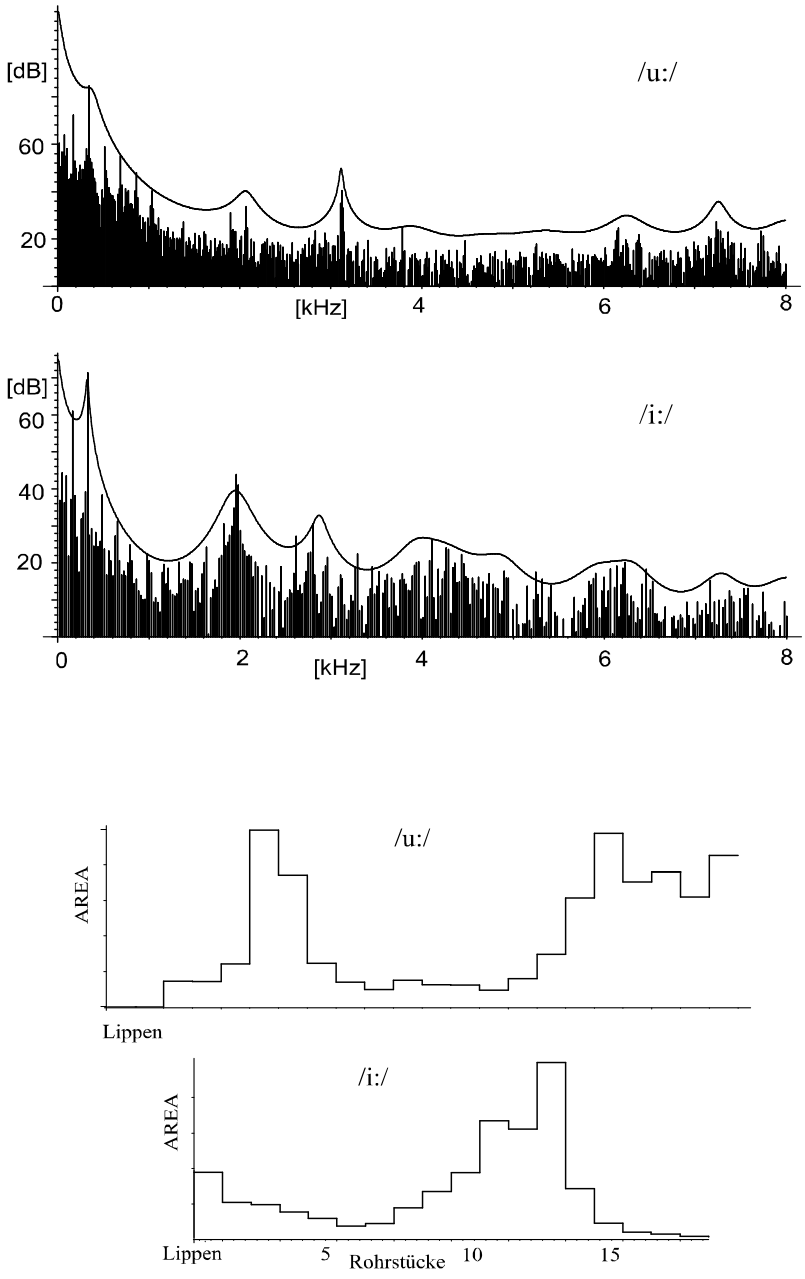


Bild 4.7: Analyse von Vokalen durch Rohrmodelle mit frequenzabhängigem Lippenabschluß: (oben) Geschätzter Betragsgang (durchgezogene Linie), DFT des analysierten Sprachsignals (Linienspektrum). Geschätzte Vokaltraktflächen der Vokale (unten).

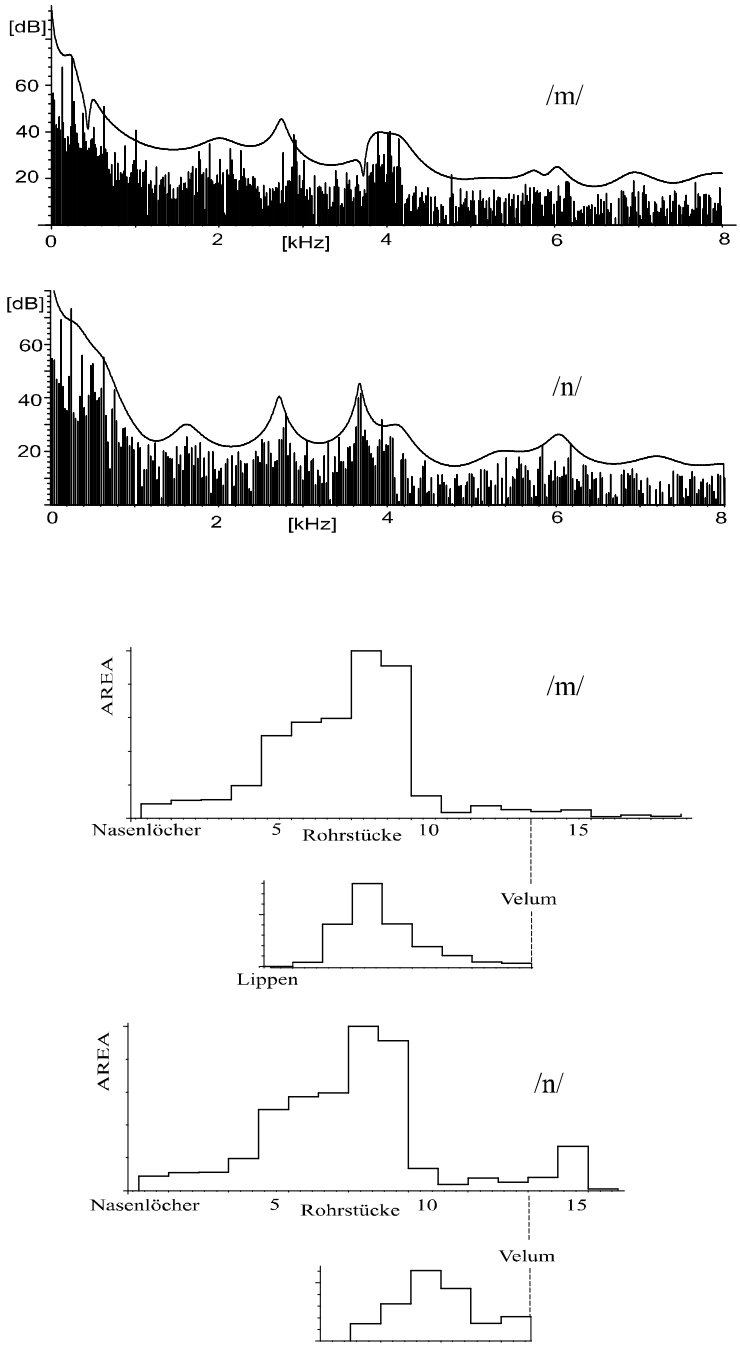


Bild 4.8: Analyse von Nasalen durch verzweigte Rohrmodelle mit fester Nasenstruktur, welche aus dem Nasal /N/ geschätzt ist: Geschätzte Betragsgänge (oben); geschätzte Mund- und Rachenflächen aus dem Nasal /m/ und /n/ mit Nasaltraktflächen, welche aus dem Nasal /N/ geschätzt sind.

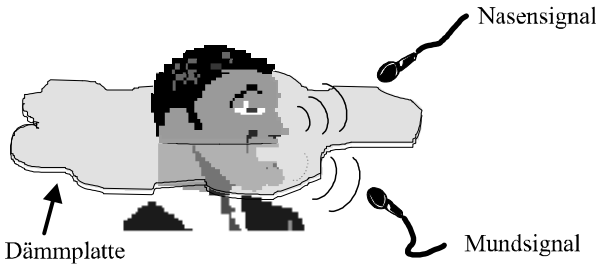


Bild 4.9: Akustische Trennung von Mund- und Nasensignal durch eine Dämmplatte [SnL99a].

4.2.3 Analyse von Rohrmodellen mit zwei Systemausgängen

Für nasalierte Vokale ist das Velum gesenkt, so daß sich das Sprachsignal als Überlagerung aus Mund- und Nasensignal zusammensetzt. Die beiden Signale sollen hier in einem Modell jeweils durch einen separaten Systemausgang beschrieben werden, welche die Öffnungen an den Lippen und Nasenlöchern repräsentieren. Dabei wird vernachlässigt, daß die Mund- bzw. Nasensignale durch die Nasenlöcher bzw. Lippen wieder in den Sprechtrakt eindringen können. Die beiden einzelnen Signale werden für eine Modellanalyse verwendet, wofür sie separat vorliegen müssen. Das einzelne Mund- und Nasensignal aus dem gemischten Sprachsignal zu extrahieren erscheint äußert schwierig, so daß hier das Mund- und Nasensignal schon bei der Sprachaufnahme akustisch getrennt wird.

Trennung von Mund- und Nasensignal

Eine Schwierigkeit der gleichzeitigen Aufnahme von Mund- und Nasensignal besteht darin, daß bei einer natürlichen Schallausbreitung sich Nasen- und Mundsignale überlagern und somit gemischt vorliegen, wohingegen durch eine mechanische Separation die natürliche Schallausbreitung gestört wird. Für die Durchführung der Signaltrennung wird folgender Ansatz [SnL99a] verfolgt: Der Kopf des Sprechers wird von einer Dämmplatte umgeben, die Mund- und Nasenlöcher trennt, wie in Bild 4.9 zu sehen ist. Dafür besteht die Platte aus zwei zusammengesetzten Teilen, die jeweils eine Aussparung in Form eines Halbkreises aufweisen, die an die Kopfform angepaßt ist. Befindet sich der Kopf dann in der zusammengesetzten Dämmplatte, so kann durch eine leichte Neigung des Kopfes erreicht werden, daß mögliche Luftritze zwischen Kopf und Platte geschlossen werden. Um zu verhindern, daß die Mund- und Nasensignale über einen Umweg um die Platte herum sich mischen können, wurde die Platte in einen Türrahmen integriert. Dadurch befindet sich der Kopf zwischen zwei Räumen, so daß sich die Nasenlöscher in einem Raum befinden, während sich der Mund in einem anderen Raum befindet. Die Schallausbreitung findet relativ ungestört statt, da Mund und Nase jeweils einen eigenen großen Raum für die Schallausbreitung zur Verfügung haben. Durch zwei Mikrophone kann dann das Nasen- und Mundsignal in den jeweiligen Räumen getrennt aufgenommen werden. Die Störungen der Schallausbreitung infolge der Trennplatte sind für die Nasenlöcher stärker anzunehmen, da der Schall aus den Nasenlöchern direkt auf die Trennwand quer zur Ausbreitungsrichtung fällt. Beim Mundsignal befindet sich die Trennwand günstigerweise längs zur Schallausbreitung.

Als Material für die Trennwände werden Dämmplatten verwendet, die einen faserartigen Aufbau aus Zellulose besitzen. Die Vorgehensweise kann die natürliche Sprechweise durch zwei Effekte beeinflussen. Erstens liegt der Sprecher, obwohl üblicherweise in einer aufrechten Haltung gesprochen wird; zweitens hört der Sprecher überwiegend sein Nasensignal und sehr wenig vom Mundsignal. Das Mundsignal wird nur dumpf und sehr schwach wahrgenommen. Diese Wahrnehmung umfaßt das sehr stark gedämpfte Schallsignal durch das Medium Luft, wie den durch Körperschall übertragenen Anteil. Ein neuer Versuchsaufbau in [BSnL02] ermöglicht eine Artikulation in aufrechter Position.

Die Trennung von Mund und Nasensignal läßt sich mit den beschriebenen Maßnahmen gut erreichen. Die aufgenommenen getrennten Signale hören sich unverfälscht an, so daß anzunehmen ist, daß die Vorrichtung das Signal in ihrem Informationsgehalt nur wenig beeinträchtigt (siehe Hörbeispiele). In Bild 4.10 ist das getrennte Mund- und Nasensignal der Äußerung "Orange" gezeigt. Es ist darin zu erkennen, daß im letzten Abschnitt des artikulierten Wortes das Velum kurzzeitig gesenkt wurde. Im Nasensignal ist auch bei geschlossenem Velum ein schwaches Signal zu erkennen. Dies kann durch Übersprechen durch die Dämmplatte erfolgen. Eine weitere Möglichkeit ist ein Übersprechen durch Vibrationen des geschlossenen Velums infolge des Schallfeldes im Vokaltrakt, wie in [Da96b] erwähnt wird.

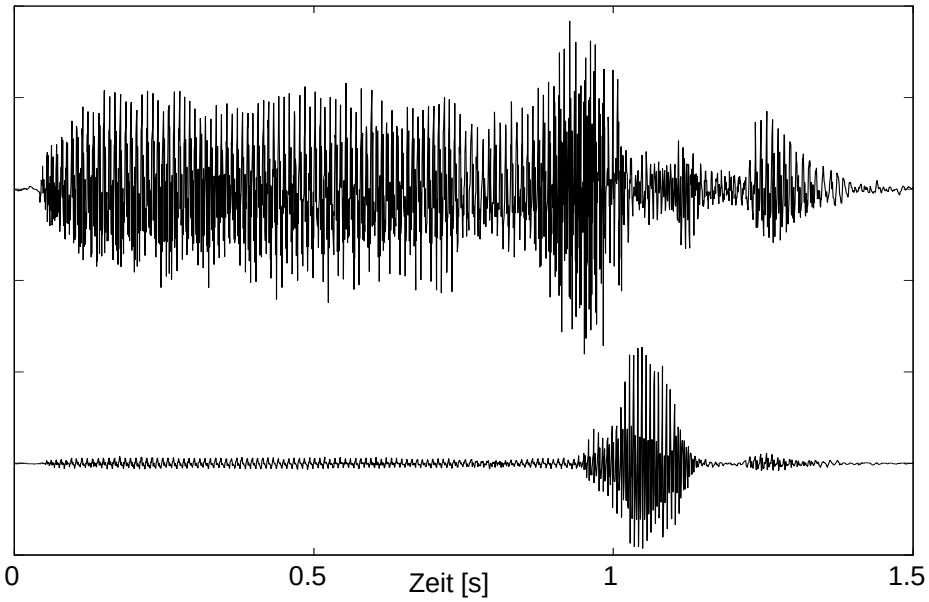


Bild 4.10: Separiertes Mund- und Nasensignal des Wortes "Orange": Mundsignal (oben) und Nasensignal (unten).

Analyse von nasalierten Vokalen

Für die Analyse von nasalierten Vokalen wird ein Rohrmodell mit einer Verzweigung verwendet, das für das Mund- und Nasensignals zwei Systemausgänge besitzt. Diese Ausgänge befinden sich jeweils an den Rohrenden der beiden Seitenzweige. Wird mittels eines Schätzalgorithmus darauf hin optimiert, daß aus Mund- und Nasensignal

gemischte Signal der Nasalvokale nachzubilden, können die Rohrteilsysteme, die den Nasal- und Vokaltrakt repräsentieren, nicht eindeutig geschätzt werden. Die Wirkungen der Seitenzweige bzw. der durch sie hervorgerufenen Resonanzen und Antiresonanzen vermischen sich in den zusammengefaßten Systemausgängen, womit keine eindeutige Zuordnung gewährleistet ist. Deshalb sollen die beiden Rohrmodellausgänge jeweils explizit das Mund- und Nasensignal nachbilden. Die beiden Systemausgänge müssen zwei verschiedene Signale gleichzeitig modellieren, wodurch starke Restriktionen für die Analyse vorliegen. Das Rohrsystem mit den beiden Systemausgängen x_M^+ für die Lippenöffnung und x_N^+ für die Nasenlöcher ist in Bild 4.11 zu sehen. Im verzweig-

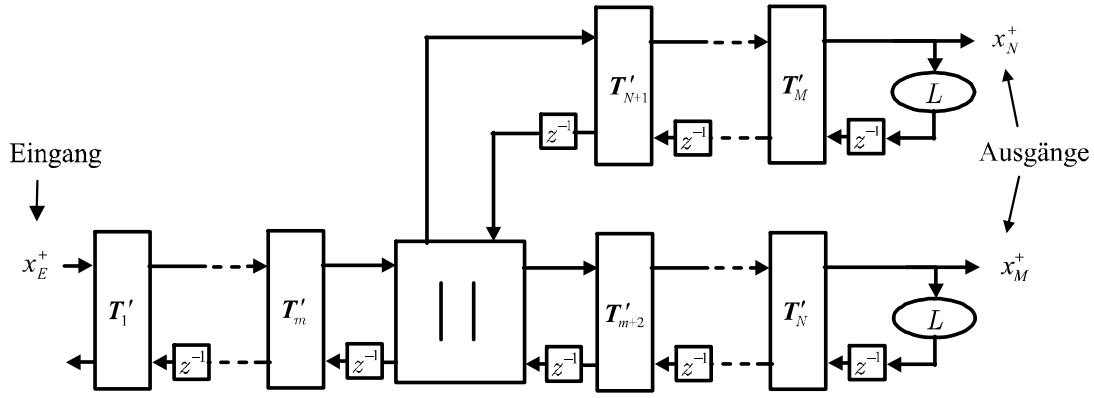


Bild 4.11: Rohrmodell mit zwei Ausgängen an den Lippen und Nasenlöchern.

ten Rohrmodell beschreibt das Teilsystem vor dem Dreitor den Rachen. Der untere Abzweig hinter dem Dreitor beschreibt den Mundraum und der obere Abzweig den Nasaltrakt. Der Seitenzweig des Nasaltraktes weist entsprechend der Anatomie mehr Rohrelemente auf als der Seitenzweig des Mundraums. Mit der Modellierung des Nasaltraktes durch ein unverzweigtes Rohr können die Nebenhöhlen des Nasaltraktes und dessen Aufspaltung durch das Septum nicht direkt modelliert werden. Der Rohrabschluß an den Lippen und an den Nasenlöchern wird jeweils mit dem frequenzabhängigen Abschluß $L(z)$ in (4.70) beschrieben, wobei für den Nasenausgang eine kleinere Fläche eingestellt wird. Für die Analyse stehen die getrennten Mundsignale s'_M und Nasensignale s'_N von isoliert gesprochenen Nasalvokalen zur Verfügung, die wie zuvor beschrieben durch die mechanische Vorrichtung einer Dämmplatte akustisch separiert wurden. Die Abtastrate der getrennt aufgenommenen Signale liegt bei 32 kHz. Für die Beseitigung des Einflusses der stimmhaften Anregung und Abstrahlung werden die Signale mit einer mehrfach angewandten Burg-Methode erster Ordnung vorgefiltert. Die vorgefilterten Mund- und Nasensignale s_M und s_N werden für die Analyse verwendet. Der Systemausgang x_M^+ am Rohrabschluß der Lippen soll das Mundsignal s_M modellieren während gleichzeitig der Systemausgang x_N^+ am Nasenabschluß das Nasensignal s_N beschreiben soll. Durch die zwei Systemausgänge existieren zwei Übertragungsfunktionen mit $H_M(z)$ für das Mundsignal und $H_N(z)$ für das Nasensignal:

$$H_M(z) = \frac{X_M^+}{X_E^+} \quad \text{und} \quad H_N(z) = \frac{X_N^+}{X_E^+}, \tag{4.71}$$

wobei X_E^+ den Systemeingang am Rohranfang des Rachens darstellt. Die gleichzeitige Modellierung der beiden Signale bewirkt, daß die beiden Übertragungsfunktionen

$H_M(z)$ und $H_N(z)$ durch einen einzigen Koeffizientensatz bestimmt sind. Dies erschwert die Analyse, da es dadurch möglich ist, daß durch eine Veränderung eines Koeffizienten zwar das eine Ausgangssignal besser approximiert wird, aber das andere Signal gleichzeitig schlechter modelliert werden kann. Die für die Analyse verwendete Fehlerdefinition setzt sich aus den Einzelfehlern e_M und e_N der beiden Signale zusammen:

$$e_M = \frac{1}{L}\left|\frac{b_0^M}{a_0^M}\right|^2 \cdot \sum_{k=0}^{L-1}\left|\frac{S_M(\mathrm{e}^{\mathrm{j}2\pi k/L})}{H_M(\mathrm{e}^{\mathrm{j}2\pi k/L})}\right|^2 \tag{4.72}$$

und

$$e_N = \frac{1}{L}\left|\frac{b_0^N}{a_0^N}\right|^2 \cdot \sum_{k=0}^{L-1}\left|\frac{S_N(\mathrm{e}^{\mathrm{j}2\pi k/L})}{H_N(\mathrm{e}^{\mathrm{j}2\pi k/L})}\right|^2 . \tag{4.73}$$

S_M und S_N sind die Spektren der vorgefilterten Mund- und Nasensignale, die jeweils eine Periode beinhalten. Die Vorfaktoren a_0 und b_0 beziehen sich entsprechend ihrem oberen Index jeweils auf die konstanten Terme in H_M und H_N. Die Summe beider Einzelfehler ergibt den Gesamtfehler e_g mit

$$e_g = e_M + e_N. \tag{4.74}$$

Der Optimierungsalgorithmus minimiert den Fehler e_g. Dabei werden alle Modellparameter bis auf die beiden Rohrabschlüsse geschätzt. Die Bilder 4.12 und 4.13 zeigen die Ergebnisse der Minimierung von e_g von einer neutralen Startkonfiguration aus. Das Bild 4.12 oben zeigt die geschätzten Betragsgänge des nasalierten Vokals /ã/, während das Bild 4.13 die des nasalierten Vokals /ĩ/ zeigen. Die nasalierten Vokale wurden für die Analyse isoliert gesprochen. An den Spektren ist zu erkennen, daß eine gleichzeitige Modellierung von zwei Signalen die Problematik mit sich bringt, daß die Modellierung des einen Signals die des anderen Signals negativ beeinflussen kann. Ein mehrfach verzweigter Nasaltrakt könnte die Modellierung verbessern, da einerseits die Nasenanatomie besser wiedergegeben würde und gleichzeitig für die Modellierung mehr Pole und Nullstellen zur Verfügung ständen. Bild 4.12 unten zeigt die geschätzten Flächen von Nasal- und Vokaltrakt des nasalierten Vokals /ã/. Die geschätzten Flächen des Lautes /ã/ sehen glaubhaft aus, wenn man die aus Röntgenaufnahmen ermittelten Flächen von Fant [Fa70] als Referenz heranzieht. Die geschätzten Flächen variieren allerdings zum Teil in Abhängigkeit von den Sprachproben, den Einstellungen der Abschlußkoeffizienten und den Längen der Seitenzweige, die den Nasaltrakt und den Mundraum modellieren.

Analyse von mehrfach verzweigten Rohrsystemen als Modell des Nasaltrakts

Die Modellierung des Nasaltrakts durch ein unverzweigtes Rohrmodell stellt eine starke Vereinfachung der Nasenanatomie dar. Der Nasaltrakt besitzt mehrere Nebenhöhlen, die zum Teil über sehr dünne Kanälchen mit dem Nasengang verbunden sind. Weiterhin gibt es eine Verzweigung in einen rechten und linken Nasengang mit Ausgängen an den beiden Nasenlöchern. Der Nasengang weist teilweise äußert komplizierte Querschnittsformen auf [Soq02], woraus sich stärkere Verluste als im Vokaltrakt für die Wellenausbreitung ableiten lassen. In [Ra99] wird ein numerischer Ansatz gewählt, um das

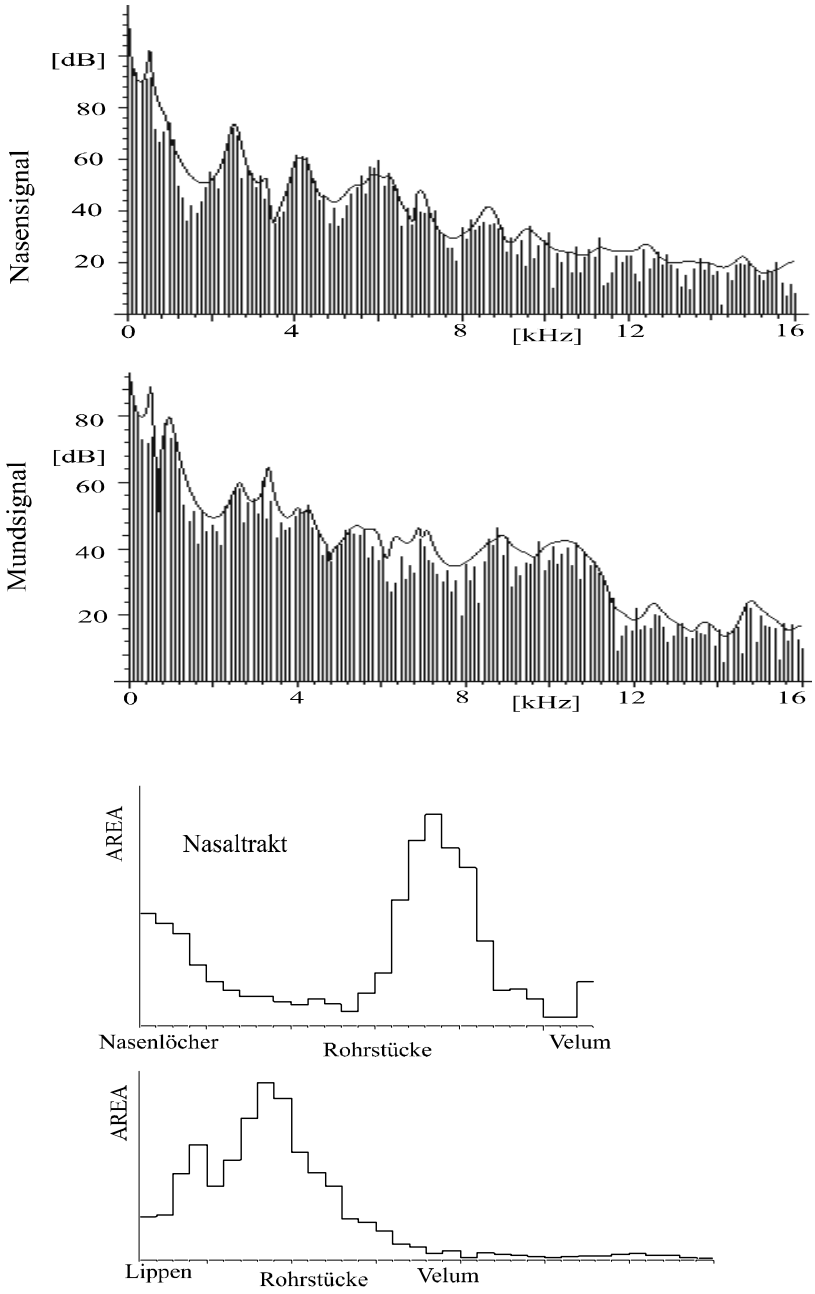


Bild 4.12: Analyse des nasalierten Vokals /ã/: Geschätzte Übertragungsfunktion zu den Nasenlöchern bzw. Lippen mit Präemphase (durchgezogene Linie) und DFT des Nasensignals bzw. Mundsignals (oben); geschätzte Nasaltrakt- und Vokaltraktflächen des nasalierten Vokals /ã/ (unten).

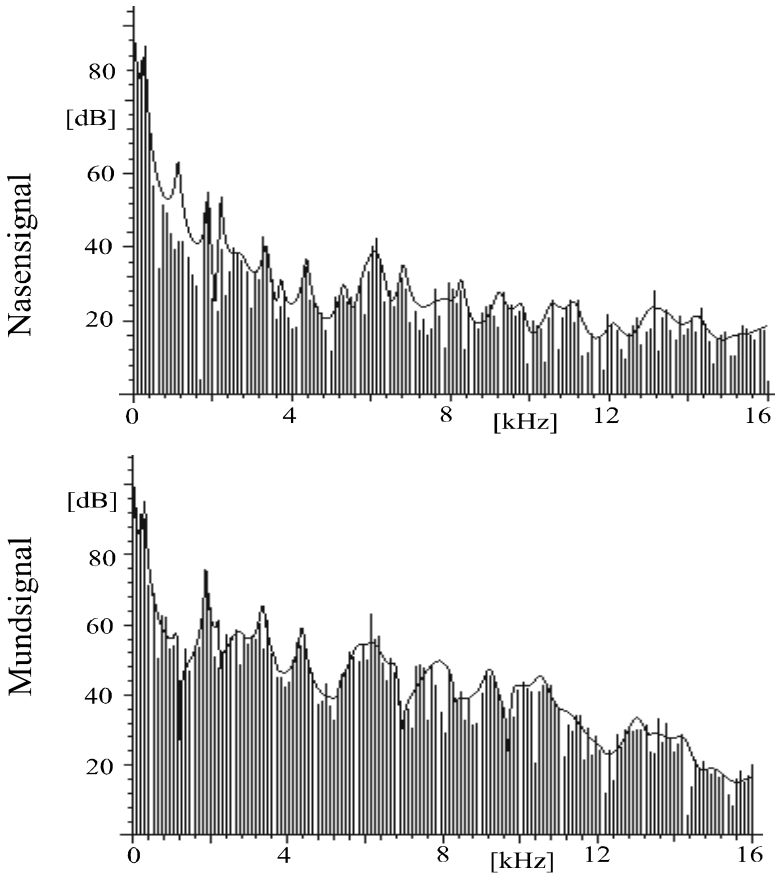


Bild 4.13: Analyse des nasalierten Vokals /$\widetilde{\text{i}}$/: Geschätzte Übertragungsfunktion zu den Nasenlöchern mit Präemphase (durchgezogene Linie) und DFT des Nasensignals (Linienspektrum) in der oberen Graphik; geschätzte Übertragungsfunktion zu den Lippen mit Präemphase (durchgezogene Linie) und DFT des Mundsignals (Linienspektrum) in der unteren Graphik.

akustische Übertragungsverhalten des Nasaltraktes bzw. dessen Übertragungsfunktion aus der dreidimensionalen Geometrie zu bestimmen. Dazu wird die Geometrie eines Nasaltraktes verwendet, welche aus digitalisierten Kryo Schnitten eines Menschen (post mortem) gewonnen wurde. Die Wellenausbreitung in diesem Modell wird durch die Lösung der dreidimensionalen Wellengleichung mittels finiter Differenzen berechnet. In [RaL00] werden computertomographische Daten eines lebenden Menschen verwendet, wodurch eine Verbesserung zu [Ra99] erzielt wird. In [RaL02] wird zusätzlich eine verlustlose Schallausbreitung berücksichtigt, welche durch eine Dämpfung der Oberflächenschicht in der Nase realisiert wird, wodurch auch gleichzeitig die Dämpfung in einem realistischen Zusammenhang zum Querschnitt abhängig ist. In [Suz96] wurde ein numerischer Ansatz mittels finiter Elemente gewählt, dessen Geometrie allerdings eine gröbere Auflösung aufweist. Mit Hilfe einer Impulsanregung im Bereich des Velums des dreidimensionalen Modells von [Ra99] kann das Übertragungsverhalten zu den beiden Nasenlöchern durch die Simulation der Schallausbreitung mittels finiter Differenzen ermittelt werden. Dafür werden die Impulsantworten n_l und n_r jeweils an den Positionen der beiden Nasenlöcher im Modell erfaßt, wonach die zwei Übertragungsfunktionen für das linke Nasenloch N_l und rechte Nasenloch N_r mittels DFT ermittelt wird. Der Systemeingang befindet sich jeweils am Velum. Da das dreidimensionale Gitternetz des Nasaltraktes relativ fein ist ($\sim$ 1mm), ist die Berechnung sehr rechenintensiv. Diese numerischen Untersuchungen wurden von Ranostaj durchgeführt [Ra99]. Diese beiden Übertragungsfunktionen N_l und N_r werden nun benutzt, um die Parameter eines mehrfach verzweigten Rohrmodells zu bestimmen, welches den Nasaltrakt modelliert [RaSnL99]. Entsprechend der Aufspaltung des Nasenganges durch das Septum besitzt das Rohrmodell eine Verzweigung in einen linken und rechten Nasengang mit jeweils einem Systemausgang für die Nasenlöcher. Um Nebenhöhlen zu berücksichtigen werden Abzweige an die beiden Nasengänge mittels Dreitor-Adaptoren angekoppelt. Es wurde dabei mit unterschiedlichen Anzahlen und Längen von Abzweigen bzw. Nebenhöhlen experimentiert. Durch längere Abzweige besitzt das verzweigte System des Nasaltraktes mehr Pole und Nullstellen, wodurch eine bessere Modellierung erreicht werden kann, auch wenn die Abzweige unnatürlich lang gewählt sind. Eine andere Möglichkeit den Systemgrad zu erhöhen besteht darin mehrere Abzweige zu verwenden, was sich für die Modellierung als vorteilhaft herausgestellt hat. Es werden deshalb Ergebnisse von analysierten Rohrmodellen gezeigt mit jeweils einem Abzweig und drei Abzweigen pro Nasengang. Dabei muß angemerkt werden, daß im untersuchten dreidimensionalen Modell nur eine Nebenhöhle angekoppelt ist, da die Auflösung der Kryo Schnitte für die sehr dünnen Kanälchen zu den Nebenhöhlen größtenteils nicht ausreichte. Dies konnte allerdings in nachfolgenden Volumenmodellen berücksichtigt werden. Die Rohrstrukturen der gezeigten Analysen sind oben in den Bildern 4.14 und 4.15 schematisch abgebildet. Die Analyse wird analog zu der von den Nasalvokalen des letzten Abschnittes durch Minimierung eines Gesamtfehlers

$$e_g = e_l + e_r \tag{4.75}$$

erreicht, der sich aus den Einzelfehlern

$$e_l = \frac{1}{L}\left|\frac{b_0^l}{a_0^l}\right|^2 \cdot \sum_{k=0}^{L-1}\left|\frac{N_l(\mathrm{e}^{\mathrm{j}2\pi k/L})}{H_l(\mathrm{e}^{\mathrm{j}2\pi k/L})}\right|^2 \tag{4.76}$$

und

$$e_r = \frac{1}{L}\left|\frac{b_0^r}{a_0^r}\right|^2 \cdot \sum_{k=0}^{L-1}\left|\frac{N_r(\mathrm{e}^{\mathrm{j}2\pi k/L})}{H_r(\mathrm{e}^{\mathrm{j}2\pi k/L})}\right|^2 \tag{4.77}$$

zusammensetzt. $H_r(z)$ und $H_l(z)$ sind die beiden Übertragungsfunktionen des verzweigten Rohrmodells zwischen dem Systemeingang x_v^+ an der Position des Velums und den beiden Systemausgängen x_l^+ und x_r^+, welche jeweils das linke und rechte Nasenloch repräsentieren:

$$H_r(z) = \frac{X_r}{X_v} \quad \text{und} \quad H_l(z) = \frac{X_l}{X_v}. \tag{4.78}$$

Die Minimierung des Fehlers e_g wird durch das Gradientenverfahren erreicht, welches aus einer neutralen Rohrkonfiguration startet. Die geschätzten Betragsgänge der beiden Ausgangssignale sind für die beiden Rohrstrukturen unten in den Bildern 4.14 und 4.15 zu sehen. Hierbei besteht wieder die Problematik zwei Systemausgänge gleichzeitig zu modellieren analog zu der Analyse der Nasalvokale des letzten Abschnitts.

Der Seitenzweig des Rohrmodells in Bild 4.14 ist länger gewählt worden als die von Bild 4.15, da dessen Rohrmodell mehrere Seitenzweige aufweist. Damit sollen die Systemgrade der beiden Modelle angeglichen werden. Der Systemgrad des Modells mit den drei Abzweigen pro Hauptzweig ist allerdings immer noch höher als der des Modells mit jeweils einem Abzweig. Die Modellierung der Signalspektren N_l und N_r mit drei Seitenzweig-Paaren ist größtenteils besser als die mit nur einem Seitenzweig, wie in den Schätzergebnissen zu sehen ist. Auch wenn die Übertragungsfunktionen N_l und N_r zum linken und rechten Nasenloch nicht genau als die wahren Übertragungsverhältnisse des Nasaltraktes anzunehmen sind, zeigen die Analyseergebnisse, daß der Algorithmus in der Lage ist, auch akzeptable Lösungen für komplizierte Rohrstrukturen zu erzielen. Dies belegt besonders das Resultat in Bild 4.15 mit mehreren Seitenzweigen. Für Analysen von solch mehrfach verzweigten Rohrsystemen mit unterschiedlichen Variationen der Rohrstruktur, hat sich der allgemeine Ansatz der entwickelten Algorithmen für die Bestimmung der Übertragungsfunktion erweiteter Rohrmodelle als zweckmäßig erwiesen. Die gewählte Implementierung unterstützt die Vielfältigkeit der Rohrmodelle und deren Analysen.

4.3 Analyse von zeitvariablen Rohrsystemen

Bisher wurden Analysen von zeitinvarianten Systemen behandelt. Die Vokaltraktflächen verändern sich beim Sprechen relativ langsam, so daß für die Parameterbestimmung näherungsweise ein stationäres Übertragungsverhalten des Vokaltraktes angenommen werden kann, wenn die Analysefenster nicht zu groß gewählt sind. Für den stimmhaften Anregungsmechanismus trifft dies nicht zu, da der Zyklus einer Stimmbandschwingung selbst schon ein instationäres Verhalten aufweist. Für die stimmhafte Anregung kann Stationarität angenommen werden, wenn man als Analysezeitabschnitt die geschlossene Glottisphase wählt. Dies wird hier nicht vorgenommen, da in diesem kurzen Abschnitt der Einfluß der geöffneten Phase fehlt. Würden mit den so ermittelten Flächen Laute synthetisiert, so wären die Spektren verfälscht, da nur der spektrale Einfluß der geschlossenen Glottisphase berücksichtigt wird. Für solche Analysen müßte die Lage der geschlossenen Phase bekannt sein. Wie in [Tir79, Wo79] zu sehen ist, sind

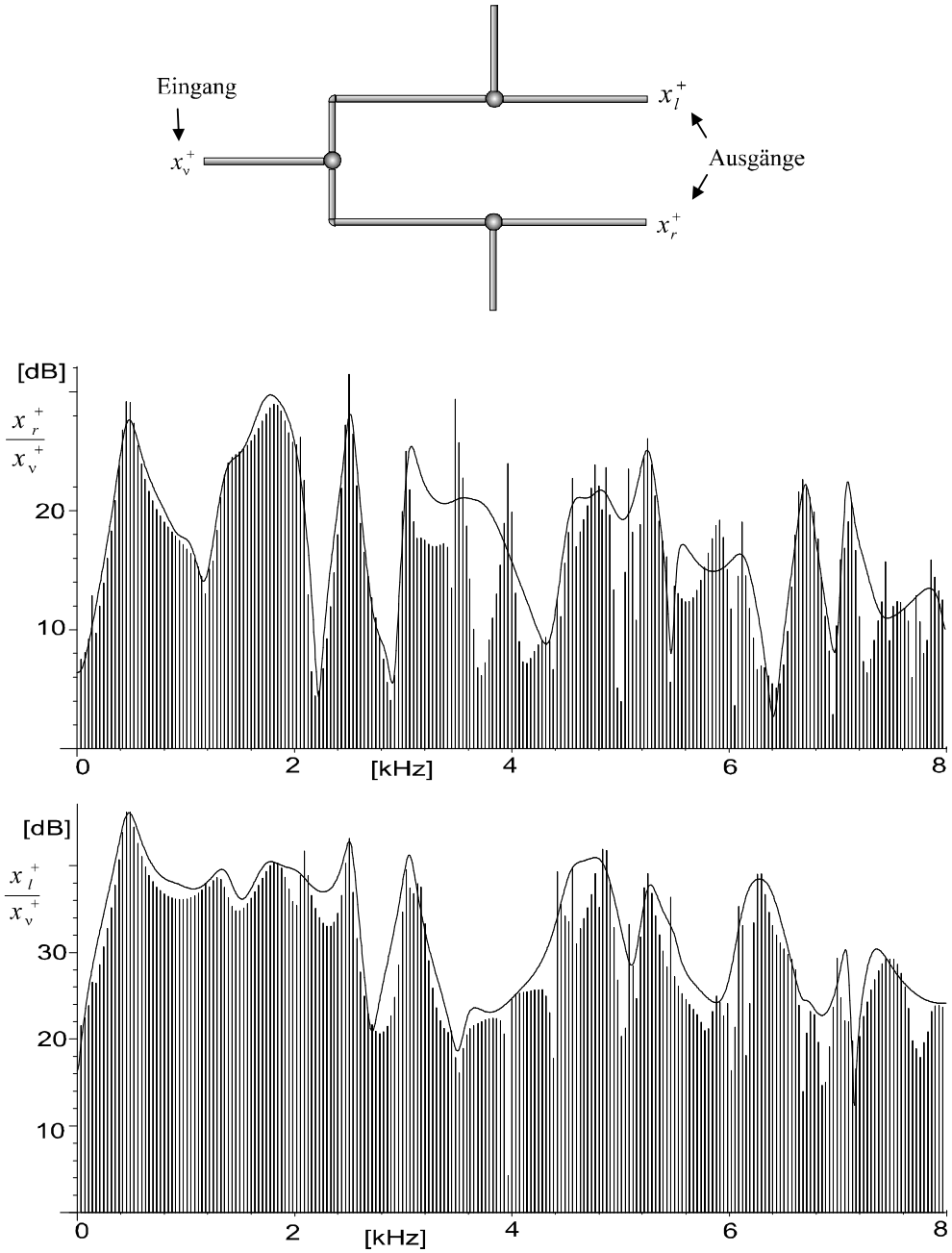


Bild 4.14: Analyse mittels eines verzweigten Rohrmodells mit einem Seitenzweig für jeweils den linken und rechten Nasengang, welches oben schematisch dargestellt ist. Geschätzte Betragsgänge vom Velum zum rechten und linken Nasenloch mit einem verzeigten Nasenmodell mit einem Seitenzweigpaar und DFT der jeweils vorgegebenen Impulsantwort (unten).

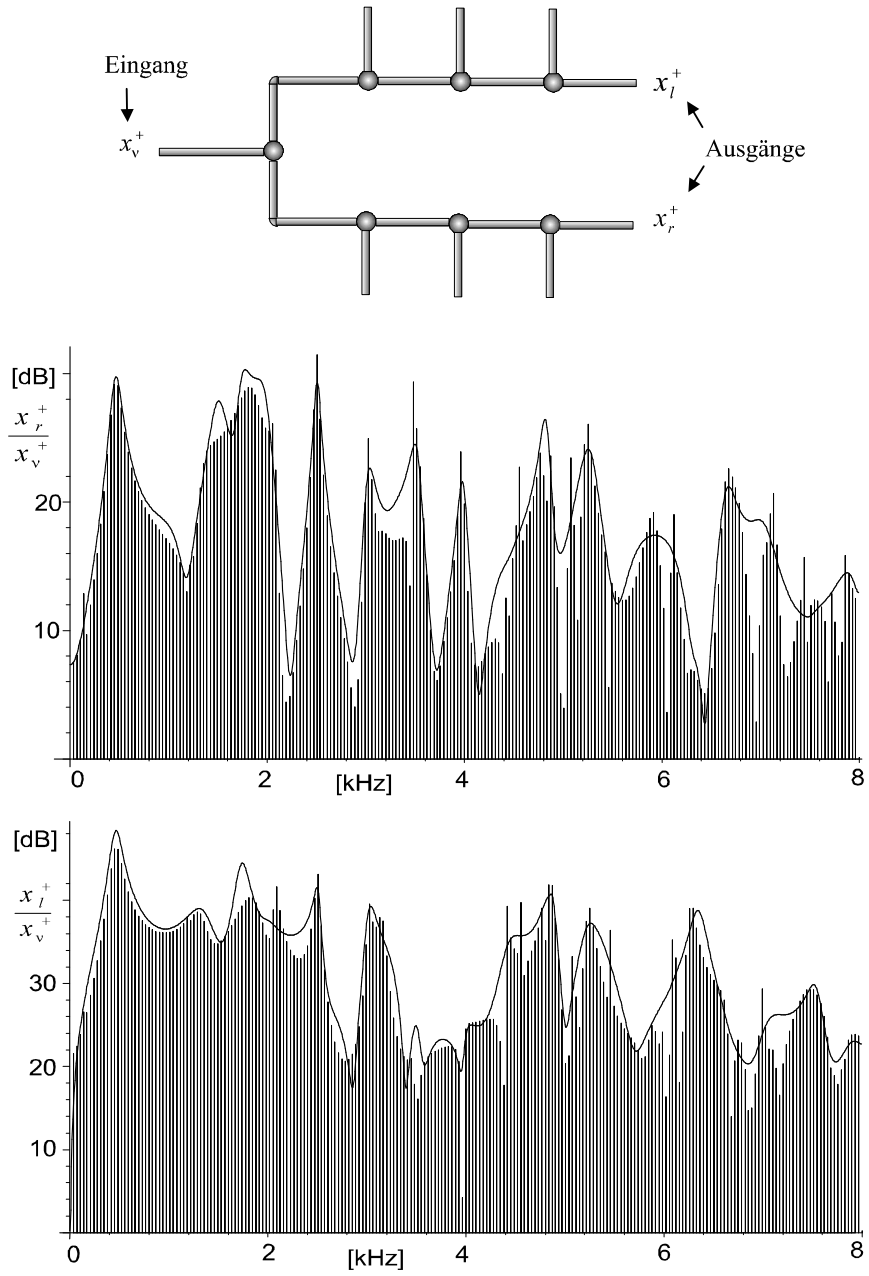


Bild 4.15: Analyse mittels eines verzweigten Rohrmodells mit drei Seitenzweigen für jeweils den linken und rechten Nasengang, welches oben schematisch dargestellt ist. Geschätzte Betragsg änge vom Velum zum rechten und linken Nasenloch mit einem verzeigten Nasenmodell mit drei Seitenzweigpaaren und DFT der jeweils vorgegebenen Impulsantwort (unten).

die Übertragungsfunktionen bei geöffneter und geschlossener Glottis unterschiedlich. In der geöffneten Glottis-Phase sollen sie Bandbreiten größer und die Formantfrequenzen tiefer werden im Vergleich zur geschlossenen Phase [Wo79]. In [Str74, Tir79, Ko02] sind Algorithmen für die Erkennung der geschlossenen Phase unter Verwendung des Sprachsignals vorgestellt, wobei die Problematik der Vorgehensweise bei Strube angesprochen wird. In [Ko02] sind bessere Ergebnisse vorgestellt als in [Str74]. Da die hier zu analysierenden Zeitabschnitte mindestens eine Grundperiode enthalten, wird das Verhalten der Glottis für die im Folgenden behandelten Analysen als zeitvariabel angesehen. Die Instationarität kann durch selbstschwingende Glottismodelle berücksichtigt werden, was hier allerdings nicht vorgenommen wird. Eine Interaktion zwischen Glottis und Vokaltrakt kann nach [Al85] hörbare Verbesserungen erbringen. Modelle der Glottis für die stimmhafte Anregung von Sprechtraktmodellen werden in [So87, Mey89] diskutiert. In diesem hier diskutierten Ansatz wird nicht der Versuch unternommen den kompletten Anregungsmechanismus zu modellieren; sondern es wird nur der Einfluß der Glottis in Bezug auf den Abschluß des Vokaltraktes berücksichtigt. Dafür wird der Rohrabschluß durch den zeitvariablen Glottiskoeffizienten $r_g(t)$ von (3.104) realisiert [SnL99b]. Durch die Verwendung eines zeitvariablen Systems ist die Übertragungsfunktion im Z-Bereich nicht mehr verfügbar. Das zeitvariable Rohrsystem liefert bei Anregung ein Ausgangssignal y, dessen Spektrum Y statt der Übertragungsfunktion H verwendet werden kann. Für die Auswertung der Fehlerdefinition e_k werden jedoch der Betragsgang und die konstanten Terme (die Koeffizienten von z^0) der Polynome in $H(z)$ benötigt. Um den Fehler e_k auswerten zu können, muß der Korrekturfaktor berechnet werden. Dieser läßt sich im zeitinvarianten Fall aus der algebraischen Form bzw. den Polynomkoeffizienten von $H(z)$ gewinnen. Da die Übertragungsfunktion im Z-Bereich nicht vorliegt, wird der Korrekturfaktor aus dem Ausgangssignal y des Rohrmodells berechnet [SnL99b]. Der Korrekturfaktor b_0^2/a_0^2 ist das Quadrat des konstanten Terms in $H(z)$. Das Szego-Kolmogoroff Theorem liefert eine Beziehung zwischen dem konstanten Term l_0 der Laurentreihe eines minimalphasigen Systems

$$L(z) = l_0 + \sum_{k=1}^{\infty} l_k \cdot z^{-k} \tag{4.79}$$

und dem Betragsgang mittels

$$\frac{1}{2\pi} \int_0^{2\pi} \ln\left(\left|L(\mathrm{e}^{\mathrm{j}\omega})\right|^2\right) \mathrm{d}\omega = \ln\left(l_0^2\right). \tag{4.80}$$

Ein Beweis der Beziehung (4.80) findet sich in [Pa91] (Appendix 13). Dadurch kann der Korrekturfaktor bestimmt werden mit

$$\frac{1}{2\pi} \int_0^{2\pi} \ln\left(\left|H(\mathrm{e}^{\mathrm{j}\omega})\right|^2\right) \mathrm{d}\omega = \ln\left(\frac{b_0^2}{a_0^2}\right), \tag{4.81}$$

bzw. für ein endliches Analysesignal der Länge N mit

$$\frac{1}{N} \sum_{k=0}^{N-1} \ln\left|H(\mathrm{e}^{\mathrm{j}2\pi k/N})\right|^2 = \ln\left(\frac{b_0^2}{a_0^2}\right). \tag{4.82}$$

Wird die Exponentialfunktion auf die Summe in (4.82) angewandt, so ergibt sich ein Produkt von Exponentialfunktionen, welche sich mit dem Logarithmus aufheben. Dadurch resultiert ein Produkt für den Korrekturfaktor:

$$\frac{b_0^2}{a_0^2} = \prod_{k=0}^{N-1} \left| H(\mathrm{e}^{\mathrm{j}2\pi k/N}) \right|^{\frac{2}{N}} . \tag{4.83}$$

Es sei angemerkt, daß in [SnL99b] und anderen der Term $1/N$ im Exponenten weggelassen wurde, da in (4.82) die Summe auf der linken Seite schon als Mittelwert angenommen wurde. Substituiert man den Frequenzgang H des Modells in (4.67) durch das Spektrum des Systemausgangs y so ergibt sich:

$$e_k = \frac{1}{2\pi} \cdot \exp\left(\frac{1}{2\pi} \int_0^{2\pi} \ln\left(\left| Y(\mathrm{e}^{\mathrm{j}\omega}) \right|^2 \right) \mathrm{d}\boldsymbol{\omega} \right) \cdot \int_0^{2\pi} \left| \frac{X(\mathrm{e}^{\mathrm{j}\omega})}{Y(\mathrm{e}^{\mathrm{j}\omega})} \right|^2 \mathrm{d}\boldsymbol{\omega}$$

bzw. für endliche Signale:

$$e_k = \frac{1}{N} \cdot \exp\left(\frac{1}{N} \sum_{k=0}^{N-1} \ln \left| Y(\mathrm{e}^{\mathrm{j}2\pi k/N}) \right|^2 \right) \cdot \sum_{k=0}^{N-1} \left| \frac{X(\mathrm{e}^{\mathrm{j}2\pi k/N})}{Y(\mathrm{e}^{\mathrm{j}2\pi k/N})} \right|^2 . \tag{4.85}$$

Die Produktform des Korrekturfaktors liefert für endliche Signale die Fehlerdarstellung:

$$e_k = \frac{1}{N} \cdot \prod_{k=0}^{N-1} \left| Y(\mathrm{e}^{\mathrm{j}2\pi k/N}) \right|^{\frac{2}{N}} \cdot \sum_{k=0}^{N-1} \left| \frac{X(\mathrm{e}^{\mathrm{j}2\pi k/N})}{Y(\mathrm{e}^{\mathrm{j}2\pi k/N})} \right|^2 . \tag{4.86}$$

Die Fehlerdefinition (4.85) bzw. (4.86) ist nur von dem Betragsspektrum des zu analysierenden Signals x und des Systemausgangs y des zeitvariablen Rohrmodells abhängig. Folglich wird für die Fehlerdefinition keine algebraische Form der Übertragungsfunktion benötigt. Für die Analyse des Signals x der Länge N sollte y ebenfalls dieselbe Länge N aufweisen. Um Einschwingvorgänge zu vermeiden, wird das Rohrmodell mit mehreren Anregungsperioden v der Länge N angeregt. Der Anfangszustand des Rohrmodells ist dabei so gewählt, daß die Zustandsspeicher gelöscht sind. Die letzte Ausgangsperiode wird für die Analyse verwendet, wodurch eine Ausgangsperiode y der Länge N im eingeschwungenen Zustand vorliegt. Durch die DFT-Werte der Periode y kann der Fehler e_k berechnet werden. Der Wert des Fehlers wird wie im zeitinvarianten Fall durch ein Optimierungsverfahren minimiert. Die Auswertungen der Fehlerfunktion im Zeitbereich sind rechenaufwendiger als die im Z-Bereich, welche mit Hilfe der Übertragungsfunktion durchgeführt werden können. Für die Wahl der Anregungsperioden v werden zwei Varianten verwendet. Da der Glottisabschlußkoeffizient mit einer vorgegeben Glottisfunktion g gesteuert wird, kann g auch selbst die Anregung des Rohrmodells bilden. Als Glottisfunktion wird eine parametrisierte Zeitfunktion von Oliveira [Ol93] verwendet, welche in Bild 2.4 zu sehen ist. Da das Betragsspektrum von g in der Regel nicht denselben spektralen Abfall aufweist, wie das der spektralen Einhüllenden des zu analysierenden Signals, wird ein zusätzliches Nullstellenfilter $1 - k_g z^{-1}$ am Anfang oder Ende des Rohrmodells zugefügt. Dieses System mit dem Parameter k_g kann die Differenz der spektralen Abfälle zwischen G und X ausgleichen. k_g wird dabei durch den Optimierungsalgorithmus mitbestimmt und wird dafür in den Parametervektor $\boldsymbol{p}$

integriert. Die zweite Variante besteht darin, das inverse System der Präemphase als Anregungsmodell zu verwenden. Dazu werden aus dem Sprachsignal x durch eine wiederholte Burgmethode erster Ordnung die M Präemphasekoeffizienten r_i^p geschätzt. Diese ergeben das Anregungsfilter:

$$G_p(z) = \prod_{i=1}^{M} \frac{1}{1 - r_i^p \cdot z^{-1}} \tag{4.87}$$

des Rohrsystems. Das Anregungsmodell $G_p(z)$ wird selbst mit einem Signal angeregt, das eine Konstante als spektrale Einhüllende besitzt. Für die Analyse wird für die Anregung von G_p eine Impulsfolge verwendet. Die Abstände der Impulse sind von der Länge der analysierten Periode abhängig. Die Steuerung des Glottiskoeffizienten r_g mit der Glottisfunktion von Oliveira wird mit der Anregungsperiode synchronisiert in Bezug auf zeitlich gleicher Öffnungsphase. In Bild 4.16 ist die Parameterbestimmung für zeitvariable Rohrmodelle skizziert.

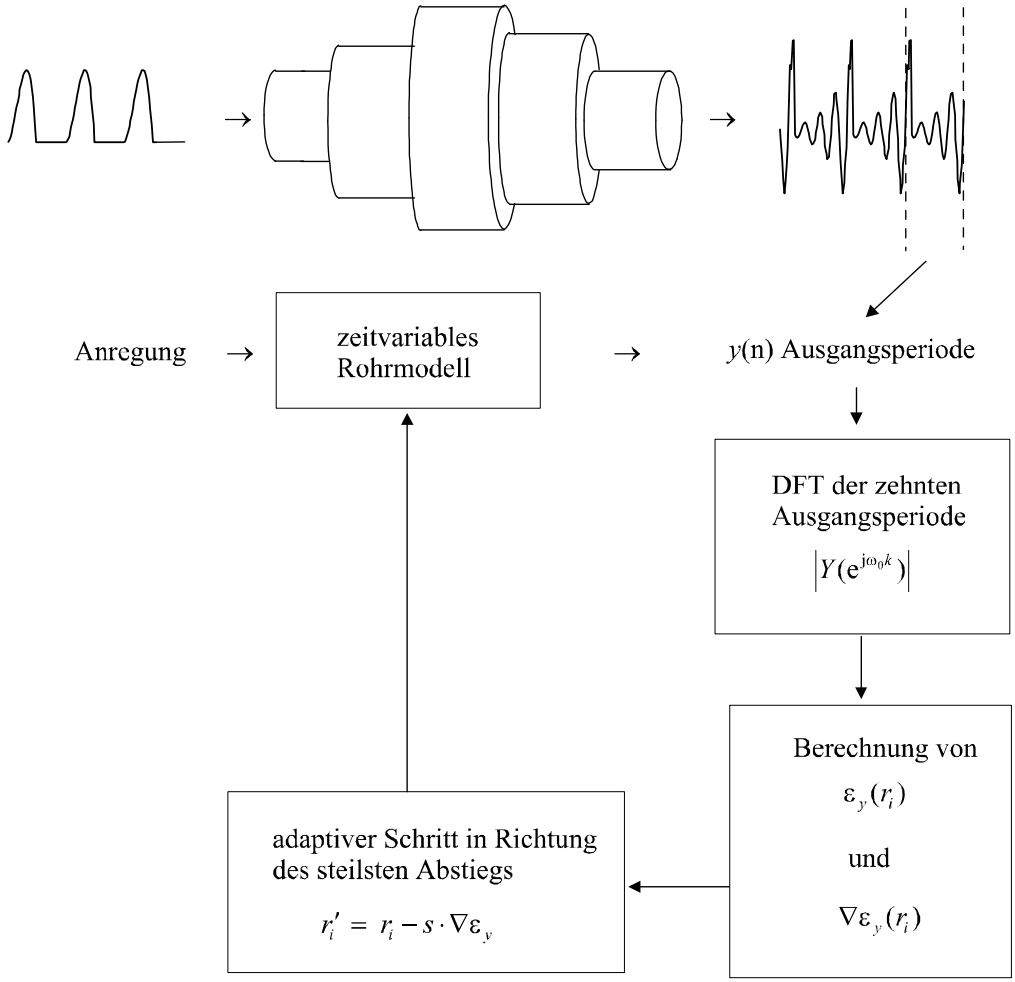


Bild 4.16: Schätzung eines zeitvariablen Rohrmodells durch Verwendung des Filterausgangssignals.

4.3.1 Analyse von Einzellauten

Das Schätzverfahren von Bild 4.16 wird dazu verwendet stimmhafte Sprachlaute mit einem zeitvariablen Rohrsystem zu analysieren. Es werden für die Analyse einzelne Perioden der stimmhaften Sprachlaute verwendet, welche das Signal x in (4.85) repräsentieren. Der Optimierungsalgorithmus startet aus einer neutralen Stellung, wobei alle Parameter bis auf die Rohrabschlüsse geschätzt werden. Die Abschlüsse sind die zeitinvarianten Abschlüsse an Mund und/oder Nase und der zeitvariable Rohrabschluß $r_g(t)$ an der Glottis. Die Abschlüsse an den Lippen und Nasenlöchern sind analog zu vorherigen Analysen frequenzabhängig gewählt. Die Anzahl der verwendeten Rohrelemente zwischen Lippen- und Glottisabschluß repräsentiert die Vokaltraktlänge und ist vor der Analyse festzulegen. Da die tatsächliche Vokaltraktlänge nur vermutet werden kann, werden mehrere Analysen mit unterschiedlichen Vokaltraktlängen durchgeführt, die einen kleinen Bereich um die angenommene Länge abdecken. Eine Vokaltraktlänge muß nach den Analysen ausgewählt werden. Dies wird anhand der geschätzten Flächen und des erreichten Fehlerwertes bestimmt.

Vokale

Die Vokaltraktflächen in Bild 4.17 sind mit dem Glottissignal g und dem Nullstellen-System der Anregung geschätzt worden [SnL99b]. Die Abtastrate der analysierten Sprachsignale liegt bei 32 kHz. Es ergeben sich zum Teil große Ähnlichkeiten mit den geschätzten Vokaltraktflächen aus NMR Untersuchungen von Story [St96], wobei für einen Vergleich auf die unterschiedlichen Sprecher hingewiesen werden muß. Die DFT-Spektren der Sprachsignale und der Ausgangsperiode des geschätzten zeitvariablen Rohrmodells ist in Bild 4.18 zu sehen.

Die Vokaltraktflächen im Bild 4.19 wurden mit demjenigen System geschätzt, daß das Anregungssystem $G_p(z)$ beinhaltet, welches durch eine wiederholte Burgmethode aus dem Sprachsignal geschätzt worden ist [SnL00c]. Die Abtastrate dieser analysierten Sprachsignale liegt bei 22 kHz.

Nasale

Die Flächen des Nasals /n/ im Bild 4.20 wurden mittels eines verzweigten Rohrmodells geschätzt, welches mit einer Anregung des Glottissignals g mit dem System einer angepaßten reellen Nullstelle angeregt wurde. Die geschätzten Nasaltraktflächen besitzen gewisse Ähnlichkeiten mit denen von Fant [Fa70]. Die Abtastrate der analysierten Sprachsignale der Nasale liegt bei 32 kHz. Die Verwendung eines Glottisabschlusses für Nasale erweist sich als problematischer als für Vokale, da durch das unverzweigte Rohrmodell für die Nase das Modell stark vereinfacht ist und durch den festen Glottisabschluß eingeschränkt ist.

Konsonanten

Ein besonderes Merkmal der Konsonanten ist die starke Verengung des Vokatraktes an einer bestimmten Stelle. Im Gegensatz zu den Vokalen weisen die Konsonanten eine meist weniger ausgeprägte Stationarität auf. Die Explosivlaute können z.B. gar nicht stationär artikuliert werden, was sich als hinderlich für die Analyse erweist. Für diese

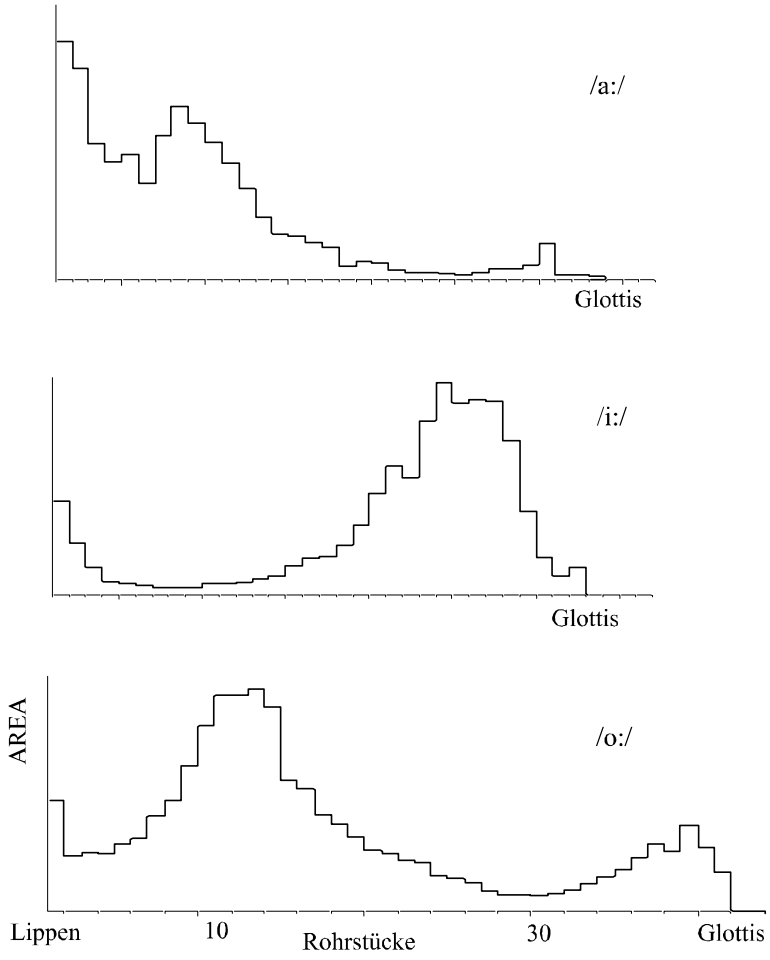


Bild 4.17: Geschätzte Vokaltraktflächen von /a:/, /i:/ und /o:/ (von oben nach unten) mit Verwendung der Glottisfunktion g als Anregung.

Laute ist die Sprechtraktbewegung ein wesentliches Element der Lautcharakteristik. Insbesondere bei den stimmhaften Explosivlauten ist der Konsonant stark durch den Übergang zum nachfolgenden stimmhaften Laut charakterisiert. Aus welcher Vokaltraktstellung die Bewegung ihren Ursprung hatte kann als wesentlich für den Explosiv angesehen werden. Für die Konstriktionsstelle der Konsonanten existiert die Vorstellung einer Target- oder Zielkonfiguration, die in einer Lautkette für den Konsonanten angestrebt wird, aber oft nicht erreicht wird. Das Problem des instationären Verhaltens der Explosive wird für die folgenden Analysen gelöst, indem für die Untersuchungen stimmhafte Laute artikuliert werden, die eine Vokaltraktstellung kurz nach der Verschlußlösung über einen längeren Zeitabschnitt halten. Im Gegensatz zu den natürlich artikulierten Lauten werden bei ihnen die interessierende Vokaltraktstellung stationär bei gleichzeitiger Phonation artikuliert. Aus Sprachsignalen dieser artikulierten Laute werden die Vokaltraktflächen geschätzt, welche im Bild 4.21 zu sehen sind. Der Glottisabschluß des unverzweigten Rohrmodells ist dabei zeitvariabel. Anhand der Analyseergebnisse von Bild 4.21 ist zu erkennen, daß die Konstriktionen in den geschätzten Flächen beobachtet werden können; dabei liegen die Positionen der Verengungen in einem realistischen Bereich für die entsprechenden Laute.

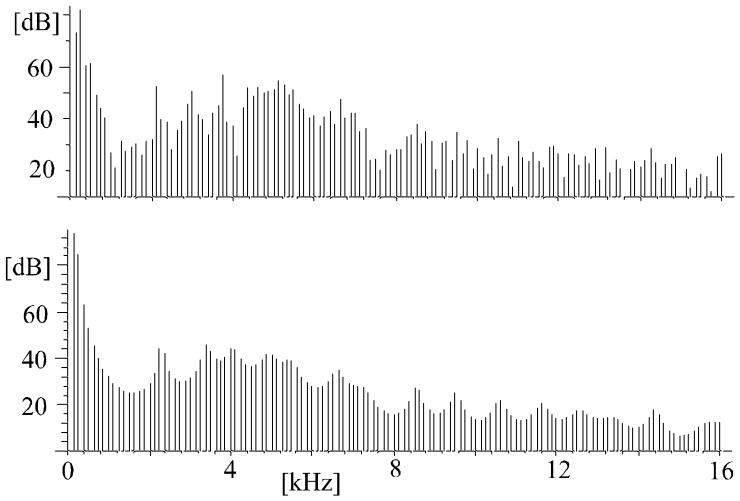


Bild 4.18: Analyse des Vokals /i:/: Betragsspektrum einer Periode des Rohrmodellausgangs (unten) und DFT der analysierten Sprachperiode (oben).

4.3.2 Alternative Fehlerdefinition in reiner Produktform

Alle zuvor beschriebenen Fehlerdefinitionen beruhen letztlich auf dem Prinzip der inversen Filterung bzw. linearen Prädiktion mit Modifikation. Es kann hilfreich sein, wenn für die Parameterschätzung alternative Fehlerdefinitionen zur Verfügung stehen, die nicht auf der gleichen Grundlage basieren wie die zuvor diskutierten. Hier stellt sich die Frage nach welchen Kriterien alternative Fehlerdefinitionen gesucht werden sollen. Die Fehlerdefinition e_k in der Darstellung von (4.86) besitzt durch (4.83) ein Produkt. Diese Feststellung motiviert die Suche nach einer Fehlerdefinition [SnL00b], die statt Summen nur Produkte aufweist wie in (4.68). Der Fehler e_k in (4.68) besteht aus einer Summe von Produkten zwischen den Betragswerten $\left|X(e^{\mathrm{j}2\pi k/N})\right|$ und $\left|Y(e^{\mathrm{j}2\pi k/N})\right|^{-1}$. Die formalen mathematischen Beziehungen zwischen den beiden Operationen Addition und Multiplikation sollen beibehalten werden, wobei die Addition durch eine Multiplikation ersetzt werden soll, da eine Produktform gesucht wird. Dafür werden Produkte von Potenzen, welche $\left|X(e^{\mathrm{j}2\pi k/N})\right|$ und $\left|Y(e^{\mathrm{j}2\pi k/N})\right|$ beinhalten, verwendet, da ein Produkt sich formal zur Addition verhält wie eine Potenz zur Multiplikation. Dies läßt sich dadurch erklären, da z.B. das Produkt $n \cdot x = 0 + x + x + \ldots + x$ eine n-fach ausgeführte Addition darstellt und eine Potenz $x^n = 1 \cdot x \cdot x \cdot \ldots \cdot x$ eine n-fach ausgeführte Multiplikation mit x, wobei das Null- bzw. Eins-Element zusätzlich verwendet wird. Diese formale Betrachtungsweise liefert keinen Beweis oder logische Herleitung, sondern soll eine Herangehensweise darstellen, um zu neuen Kandidaten von möglichen Fehlerdefinitionen zu gelangen. Evaluationen von einigen möglichen Variationen der

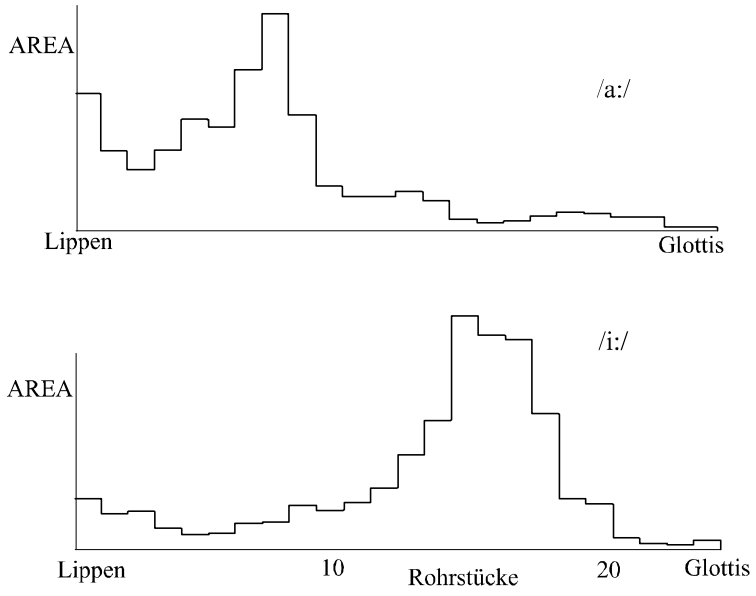


Bild 4.19: Geschätzte Flächen von Vokalen mit $G_p(z)$ als Anregungssystem.

Fehlerdefinition führte auf das Fehlermaß:

$$e_p = \prod_{k=0}^{N-1} \left| \boldsymbol{\lambda} \frac{X\left(\mathrm{e}^{\mathrm{j}2\pi k/N}\right)}{Y\left(\mathrm{e}^{\mathrm{j}2\pi k/N}\right)} \right|^{\left| \boldsymbol{\lambda} \frac{X\left(\mathrm{e}^{\mathrm{j}2\pi k/N}\right)}{Y\left(\mathrm{e}^{\mathrm{j}2\pi k/N}\right)} \right|} \tag{4.88}$$

$$\text{mit} \quad \boldsymbol{\lambda} = \prod_{k=0}^{N-1} \left| \frac{Y\left(\mathrm{e}^{\mathrm{j}2\pi k/N}\right)}{X\left(\mathrm{e}^{\mathrm{j}2\pi k/N}\right)} \right| .$$

Durch Anwendung von exp(ln) erhält man aus (4.88) für e_p die Form

$$e_p = \exp\left(\sum_{k=0}^{N-1} \left| \boldsymbol{\lambda} \frac{X(\mathrm{e}^{\mathrm{j}2\pi k/N})}{Y(\mathrm{e}^{\mathrm{j}2\pi k/N})} \right| \cdot \ln\left(\left| \boldsymbol{\lambda} \frac{X(\mathrm{e}^{\mathrm{j}2\pi k/N})}{Y(\mathrm{e}^{\mathrm{j}2\pi k/N})} \right| \right) \right), \tag{4.89}$$

welche eine gewisse Ähnlichkeit mit der Formel der Entropie aufweist. Die Beträge $|X/Y|$ in e_p können auch durch Quadrate zu $|X/Y|^2$ ersetzt werden, was zu einer Fehlerdefinition mit vergleichbaren Resultaten für die Parameterbestimmung führt. Die Fehlerdefinitionen werden darauf getestet, ob sie für eine Übertragungsfunktion (4.60) mit $H(a_0, a_1, \ldots a_N, b_0, b_1, \ldots b_M)$ für X und Y einen Fehler aufweisen, der in einer kleinen Umgebung der optimalen Lösung $(a_0, a_1, \ldots a_N, b_0, b_1, \ldots b_M)$ auch tatsächlich ein Fehlerminimum darstellt. Dafür werden die Fehler mit den Substitutionen

$$X = H(a_0, a_1, \ldots a_N, b_0, b_1, \ldots b_M), \tag{4.90}$$

$$Y = H(a_0, a_1 \pm \boldsymbol{\varepsilon}, \ldots a_N, b_0, b_1, \ldots b_M) \tag{4.91}$$

berechnet, wobei ein sehr kleines $\boldsymbol{\varepsilon}$ jeweils zu jedem Koeffizienten addiert bzw. subtrahiert wird. Das $\boldsymbol{\varepsilon}$ in (4.91) ist nur beispielhaft bei dem Koeffizienten a_1. Damit

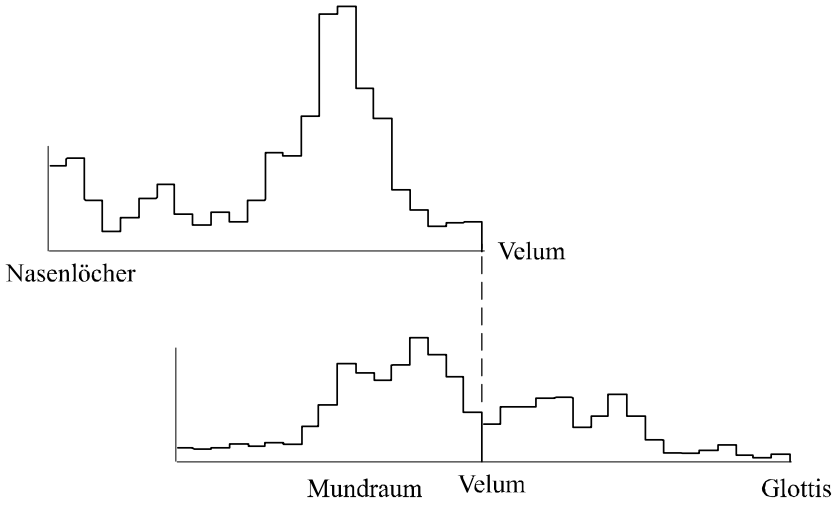


Bild 4.20: Geschätzte Flächen des Nasals /n/.

kann abgeschätzt werden, ob $(a_0, a_1, \ldots a_N, b_0, b_1, \ldots b_M)$ ein Minimum für die zu testende Fehlerdefinition darstellt. Für viele getestete Fehlervarianten trifft dies nicht zu. Die Fehlerdefinition e_p mit (4.88) kann genauso wie e_k mit (4.85) für die Analyse mit zeitvariablen Rohrsystemen verwendet werden. Bild 4.22 zeigt die resultierenden Flächen nach der Analyse durch Minimierung des Fehlers e_p von dem Sprachsignal des Vokals /i:/ mit einer Abtastrate von 32 kHz. Bild 4.23 zeigt die geschätzten Flächen für den Vokal /O/, dessen analysiertes Sprachsignal mit einer Abtastrate von 22 kHz aufgenommen wurde. Das für die Schätzung verwendete Rohrmodell weist den zeitvariablen Rohrabschluß $r_g(t)$ an der Glottis und den frequenzabhängigen Abschluß $L(z)$ von (4.70) an den Lippen auf, wie zuvor für die analysierten Explosive. Das Anregungssignal des zeitvariablen Rohrsystems ist eine Impulsfolge, welche durch das Anregungssystem $G_p(z)$ gefiltert wird. Die untere Graphik in Bild 4.24 zeigt das Betragsspektrum der synthetisierten Periode nach der Analyse. Zum Vergleich ist in Bild 4.24 oben das Betragsspektrum der analysierten Sprachperiode des Lautes /O/ gezeigt. Es ist zu sehen, daß die Formantstruktur des Vokals wiedergegeben wird. Durch die Fehlerdefinition e_p ist eine Alternative zum Fehler e_k gegeben. Daß dies eine Verbesserung bedeuten kann, illustriert folgendes Beispiel, in dem ein zeitinvariantes einfach verzweigtes Rohrmodell untersucht wird. Dafür wird von dem verzweigten Rohrmodell durch Anregung mit einer Impulsfolge ein Testsignal generiert. Bei der Analyse von verzweigten Rohrsystemen kann es vorkommen, daß das Gradientenverfahren in einem lokalen Minimum verharrt. Dies ist in Bild 4.25 zu sehen, in dem das Testsignal durch Minimierung von e_k mit der dazugehörigen verzweigten Rohrstruktur analysiert wurde. Die Schätzung startet von einer neutralen Rohrkonfiguration und konvergiert wie zu sehen nicht in das globale Minimum. Bild 4.26 zeigt die Resultate nach der Analyse desselben Testsignals mit der entsprechenden Rohstruktur, allerdings mit Verwendung der Fehlerdefinition e_p statt wie für die Resultate in Bild 4.25 mit e_k. Auch hierfür startet der Algorithmus aus einer neutralen Rohrkonfiguration, in der alle Flächen gleich groß sind. Dieses Beispiel soll nicht aufzeigen, daß der Fehler e_p für das Gradientenverfahren grundsätzlich besser ist als e_k, sondern nur daß er bei möglichen schlechten Schätzresultaten mit Verwendung von e_k, welche auch durch die Startkonfiguration

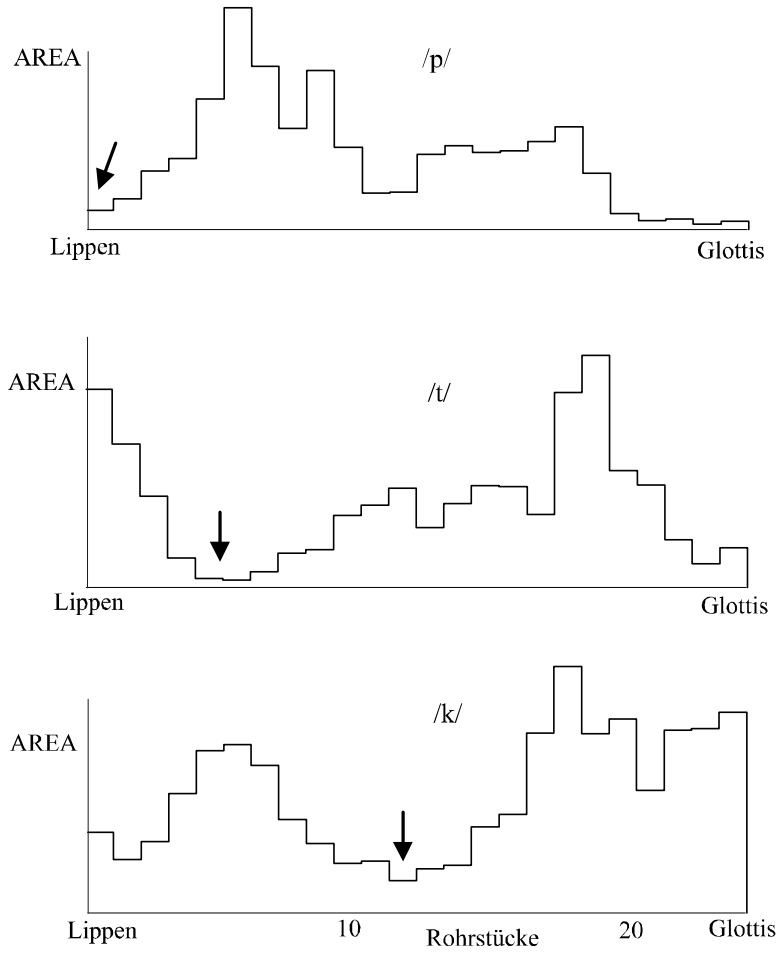


Bild 4.21: Geschätzte Vokaltraktflächen von Explosiven mit $G_p(z)$ als Anregungssystem. Pfeile markieren die jeweiligen Konstriktionen.

bedingt sind, eine bessere Alternative darstellen kann.

4.3.3 Analyse von Lautübergängen

Mit der Analyse von sehr kurzen Zeitabschnitten können nur Momentaufnahmen von Sprachäußerungen beschrieben werden. Dies ist für stationäre Laute ausreichend. Lautübergänge können dadurch nur ausschnittweise erfaßt werden. Deshalb wird hier eine Sequenz von benachbarten Zeitabschnitten analysiert, die einen stimmhaften Lautübergang beschreiben [SnL00f]. Als Zeitabschnitte werden einzelne Perioden verwendet, die an den Nulldurchgängen segmentiert sind. Die benachbarten Perioden werden nacheinander mit einem Rohrmodell analysiert. Für die Analyse einer einzelnen Periode kann das Vorwissen benutzt werden, das durch das Analyseergebnis der zeitlich vorangegangen Periode gewonnen wurde. Deshalb wird die Analyse einer Periode mit einer Startkonfiguration durchgeführt, die die Analyseresultate der vorhergehenden Periode beinhaltet, wodurch weniger Iterationen für das Konvergieren des Optimierungsalgorithmus benötigt werden, da sich die Anfangskonfiguration schon nahe am Minimum befindet. Die Analyse der ersten Periode in der Sequenz wird von einer neutralen Startkonfiguration mit mehreren Iterationen vollzogen. Dies geschieht analog zu

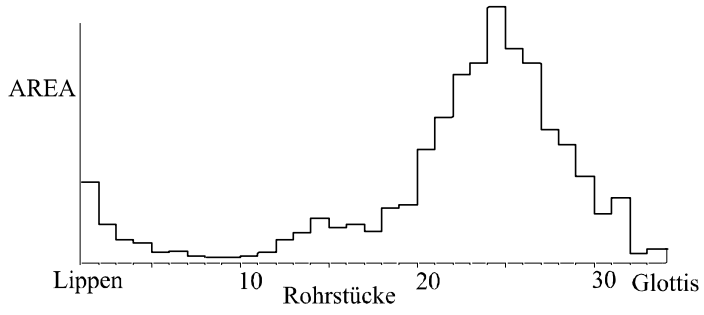


Bild 4.22: Geschätzte Vokaltraktflächen des Vokals /i:/ mit 32 kHz Abtastrate.

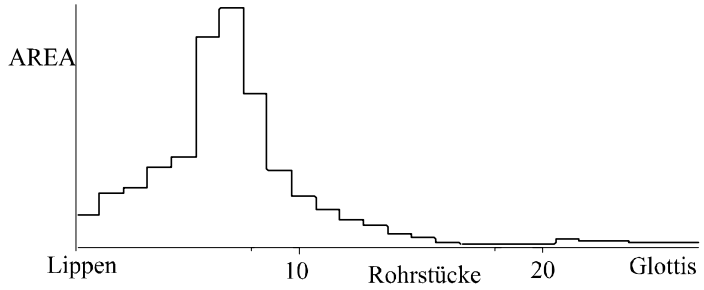


Bild 4.23: Geschätzte Vokaltraktflächen des Vokals /O/ mit 22 kHz Abtastrate.

der Analyse einer isolierten Periode, da noch kein Vorwissen vorhanden ist. Werden weniger Iterationen für die nachfolgenden Perioden verwendet, kann der Flächenverlauf in Richtung der Zeitachse geglättet werden, da das Minimum jeder Periode nicht vollständig erreicht wird. Fluktuationen von Periode zu Periode können so vermindert werden. Ursachen für diese Fluktuationen können in kleinen Fehlern der Periodensegmentierung oder in den Schwankungen der stimmhaften Anregungsperioden selbst liegen. Bild 4.27 zeigt den Flächenübergang aus einer Schätzung des Diphthongs /aI/. Jeder Flächensatz stellt dabei eine analysierte Periode dar. Für die Analyse wird ein Rohrmodell mit dem zeitvariablen Rohrabschluß $r_g(t)$ an der Glottis und dem frequenzabhängigen Abschluß $L(z)$ an den Lippen verwendet. Bei der Analyse des Flächenübergangs ist die Vokaltraktlänge konstant angenommen. Das Fehlermaß e_k wird für den Optimierungsalgorithmus verwendet. Bild 4.28 zeigt die Ergebnisse des analysierten Übergangs [za]. Die Konstriktion des Konsonanten ist in den geschätzten Flächen ansatzweise zu sehen.

Halbautomatische Segmentierung der Perioden

Für die Analyse von Lautübergängen oder ganzen Lautketten muß das Sprachsignal in eine Sequenz von Zeitabschnitten zerlegt werden, welche dann einzeln analysiert werden können. Ist die Länge der einzelnen Zeitabschnitte für die Analyse fest vorgegeben, so stimmen die resultierenden Markierungen in der Regel nicht mit den Perioden der stimmhaften Laute überein. In den gezeigten Beispielen wurden für die Analyse die Perioden einzeln segmentiert, so daß eine periodensynchrone Analyse möglich ist, was einen höheren Aufwand beinhaltet. Die Segmentierung der einzelnen Perioden

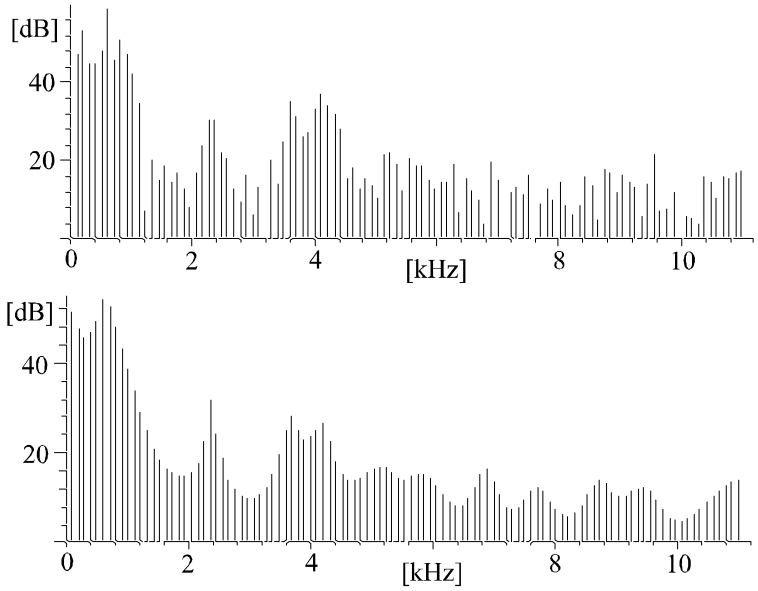


Bild 4.24: Analysiertes und synthetisiertes Signalspektrum des Vokals /O/ : DFT des analysierten Sprachsignals (oberes Linienspektrum), DFT des synthetisierten Sprachsignals (unteres Linienspektrum).

ist nicht immer eindeutig. Für die hier benutzte Segmentierung werden die Perioden nur an den Nulldurchgängen markiert. Dabei wird zusätzlich festgelegt, ob die Positionen der Nullstellen sich nur an den Nulldurchgängen mit positiver oder negativer Steigung befinden dürfen. Dies kann für verschiedene Lautübergänge unterschiedlich günstig sein. Die Segmentierung wird halbautomatisch durchgeführt. Zuerst wird von Hand die erste Periode markiert. Danach wird immer eine Schätzung der nächsten Markierung für die rechts anliegende Periode automatisch vom Computerprogramm vorgeschlagen. Dafür wird die Länge der vorherigen Periode als erster Anhaltspunkt verwendet. Die Nullstelle, die dieser ersten Schätzung am nächsten ist, wird als nächste Periodenmarkierung vorgeschlagen. Dabei können wieder nur Nullstellen auftreten, die entweder eine positive oder negative Steigung aufweisen. Die vorgeschlagene Periodenmarke kann dann angenommen werden oder von Hand korrigiert werden, wobei einfach der nächstliegende rechte oder linke Nulldurchgang gewählt werden kann. Mit dieser Vorgehensweise werden alle Perioden nacheinander markiert. Die Überprüfung der geschätzten Nullstellen durch eine Person ist deshalb notwendig, da vermeintlich falsche Schätzungen infolge von Grundfrequenzänderungen oder sich verändernde Periodenformen auftreten können. Welche Markierungen die Richtigen sind kann allerdings nicht immer eindeutig beurteilt werden. Für manche Verläufe ist es günstiger mit Nulldurchgängen zu operieren, die eine positive bzw. negative Steigung aufweisen. Dies hängt damit zusammen, wie sich die Gestalt der Zeitsignale von Periode zu Periode ändern, da es vorkommen kann, daß steile Flanken im Zeitsignal während des Übergangs verschwinden und/oder sich stark verschieben.

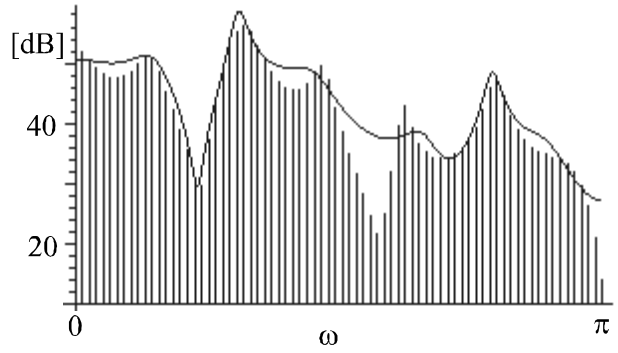


Bild 4.25: Analyse eines Testsignals mittels eines verzweigten Rohrmodells durch Minimierung von e_k: DFT des analysierten Testsignals (Linienspektrum), geschätzter Betragsgang (durchgezogene Linie).

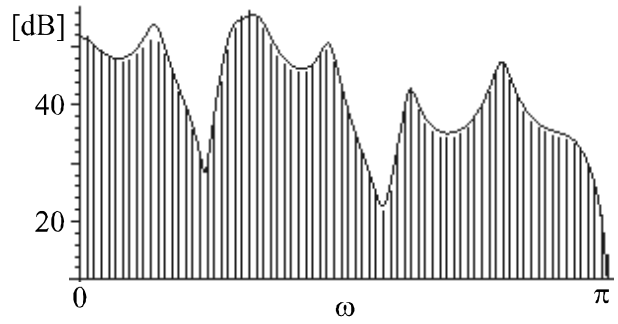


Bild 4.26: Analyse eines Testsignals mittels eines verzweigten Rohrmodells durch Minimierung von e_p: DFT des analysierten Testsignals (Linienspektrum), geschätzter Betragsgang (durchgezogene Linie).

Markierung einer Diphondatenbank

Durch dieses Verfahren wurden nicht nur die Perioden der gezeigten Beispiele markiert, sondern auch eine komplette Diphondatenbank. Diese wurde von einer weiblichen Sprecherin bei einer Abtastrate von 16 kHz am Institut für Phonetik (Uni-Frankfurt) aufgenommen. Dabei zeigte sich, daß stimmhafte Laute mit Rauschanteilen meistens schwieriger zu markieren sind; insbesondere wenn Rauschanteile auch an den Flanken der Perioden auftreten. Vokale konnten oft besser markiert werden als stimmhafte Konsonanten, da diese in der Regel keine markanten steilen Flanken im Zeitsignal besitzen. Schwierigkeiten bereiten die Übergänge von stimmhaften Lauten zur Stille oder umgekehrt, da die Amplituden dort sehr schwach sind und die Grundfrequenz sich stark ändert. Die markierten Sequenzen der gesamten Diphondatenbank sind auch mittels eines einfachen Rohrmodells analysiert worden. Die gewonnenen Koeffizienten wurden dann für anfängliche Syntheseexperimente verwendet, welche allerdings nur ein Versuchsstadium darstellten. Die Anregung der stimmhaften Laute wurde mittels einer Impulsfolge realisiert. Die stimmlosen Zeitsignale der Diphondatenbank wurden unverändert für die Synthese verwendet. Für die stimmhaften Laute zeigt sich, daß es vorteilhaft sein kann, die ersten und letzten Perioden einer Sequenz zur Stille hin für

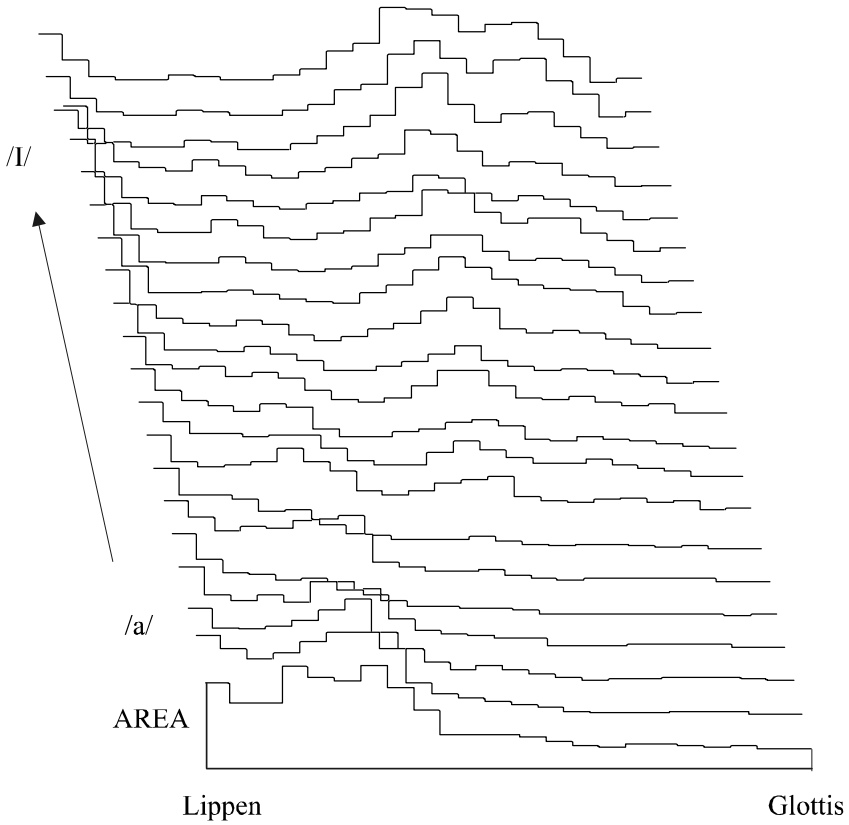


Bild 4.27: Geschätzter Übergang der Vokaltraktflächen des Diphthongs /aI/.

die Synthese nicht zu berücksichtigen. Es zeigte sich, daß es zu Schwierigkeiten in der Amplitudensteuerung bei einigen Lautübergängen gekommen ist. Die Erkenntnisse, die im Abschnitt "Resynthese" aufgezeigt werden, konnten in diese zuvor durchgeführten Syntheseexperimente nicht einfließen.

4.4 Parameterbestimmung durch Verwendung von suboptimalen Lösungen

Die Parameterbestimmung von Rohrmodellen wird durch Minimierung eines Fehlermaßes erreicht. In den vorangegangenen Abschnitten wurde ein allgemeines Optimierungsverfahren verwendet, um den Fehler zu minimieren. Solche allgemeine Verfahren verwenden keine Informationen über die inneren Strukturen des Modells und behandeln daher das System selbst als Black Box. Diese Vorgehensweise wird oft dann verfolgt, wenn die mathematische Behandlung sich zu kompliziert darstellt. Die Anwendung allgemeiner Optimierungsverfahren sind in der Regel rechenintensiv. Es werden nun schnellere Algorithmen vorgestellt, die allerdings nur auf einen Teil der möglichen Rohrstrukturen sinnvoll anwendbar sind. Diese Verfahren greifen auf Formeln zurück, die Teillösungen unmittelbar bereitstellen. Diese wurden in einer mathematischen Skriptsprache implementiert, wofür Scilab [Go99] verwendet wurde, welches an der INRIA in Frankreich entwickelt wurde und Ähnlichkeiten zu Matlab aufweist. Es muß angemerkt werden, daß Scriptsprachen in der Ausführung von Programmen oft deutlich

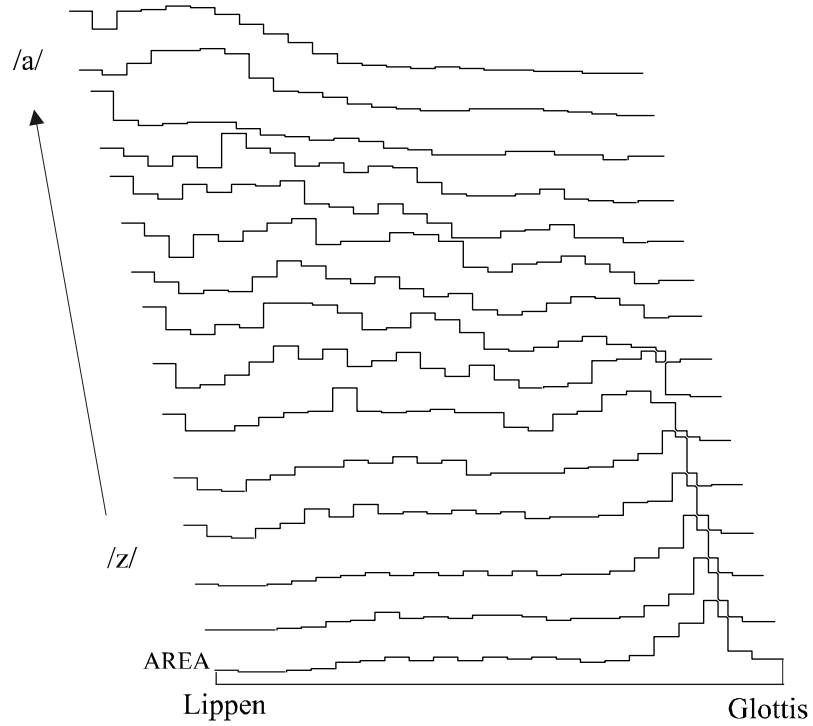


Bild 4.28: Geschätzter Übergang der Vokaltraktflächen vom stimmhaften Frikativ /z/ zum Vokal /a/.

langsamer sind als compilierte Sprachen, allerdings dafür eine schnellere Entwicklungs- und Testzeit beinhalten. Für die Parameterbestimmung werden im Gegensatz zum Gradientenverfahren Lösungsformeln verwendet, die das Schätzproblem unter Nebenbedingungen lösen. Diese Nebenbedingungen sollen die mathematische Behandlung der Schätzung so weit vereinfachen, daß geschlossene Lösungsformeln hergeleitet werden können. Dadurch wird das Problem in Teilprobleme zerlegt, die man mit einem determinierten Algorithmus lösen kann. Die Teilprobleme sind allerdings nicht unabhängig voneinander, wodurch die Schätzung iterativ erfolgt. Das gesamte Problem kann als die Suche nach dem optimalen Parametervektor $\boldsymbol{p}$ in Bezug auf den definierten Fehler angesehen werden. $\boldsymbol{p}$ ist Element des Vektorraums V der Dimension K, die der Anzahl der zu schätzenden Modellparameter entspricht. Der Parametervektor $\boldsymbol{p}$ mit den einzelnen Parametern p_k wird dargestellt durch

$$\boldsymbol{p} = \begin{pmatrix} p_1 \\ p_2 \\ \vdots \\ p_K \end{pmatrix} = \sum_{k=1}^{K} p_k \cdot \boldsymbol{u}_k = \sum_{k \in I} \boldsymbol{p}_k \tag{4.92}$$

$$\text{mit} \quad \boldsymbol{p}_k = p_k \cdot \boldsymbol{u}_k \quad \text{für } k \in I \text{ und } \quad I = \{1 \ldots K\}\,, \tag{4.93}$$

wobei der Index k ein Element der Indexmenge I ist. Als mögliche Lösungen des Gesamtproblems kommen nur minimalphasige Lösungen in Betracht, die darüber hinaus das Modell konsistent darstellen sollen. Die Minimalphasigkeit des Modells schließt dessen Stabilität ein. Die Konsistenz des Rohrmodells wird durch negative Querschnittsflächen verletzt. Die Menge Q der zugelassenen Lösungen stellt eine Untermenge des

Vektorraums V dar:

$$Q = \{\boldsymbol{p} \mid \boldsymbol{p} \Rightarrow \text{minimalphasiges System} \wedge \boldsymbol{p} \Rightarrow \text{konsistentes Modell}\} \tag{4.94}$$

$$Q \subset \{V\} \tag{4.95}$$

Die Einschränkungen durch Q bedeuten für diejenigen Parameter p_k, welche Reflexionskoeffizienten darstellen, daß sie vom Betrag kleiner als Eins sein müssen. Für die zweiparametrige Darstellung des Dreitores sind die Restriktionen der beiden Parameter nach (3.60) nicht unabhängig voneinander. Die Menge Q erfüllt in der Regel nicht die Eigenschaften eines Vektorraums. Für die Parameterschätzung gilt es den optimalen Parametervektor zu finden, für den die Fehlerdefinition $e(\boldsymbol{p})$ minimal wird, wobei $\boldsymbol{p}$ Element von Q sein muß. Dieser optimale Vektor stellt die Gesamtlösung des Schätzproblems dar. Es wird hier von dem Fall ausgegangen, daß keine Lösungsformeln für die optimale Schätzung sämtlicher Parameter verfügbar sind. Deshalb werden einige Parameter als bekannt vorausgesetzt, so daß sich durch die vereinfachte mathematische Darstellung Lösungsformeln herleiten lassen. Diese ergeben nur Teillösungen, da nicht alle Parameter p_k bestimmt werden. Für die Zerlegung des Gesamtproblems in Teilprobleme wird die Indexmenge I der Parameter in M disjunkte Teilmengen I_λ aufgeteilt:

$$I_{\boldsymbol{\lambda}} \cap I_{\boldsymbol{\mu}} = \emptyset\ , \quad \forall\ \boldsymbol{\lambda} \neq \boldsymbol{\mu} \qquad \text{und} \qquad \bigcup_{\boldsymbol{\lambda}=1}^{M} I_{\boldsymbol{\lambda}} = I, \quad M \leq K. \tag{4.96}$$

Die komplementäre Indexmenge $\overline{I_\lambda}$ enthält alle Indexwerte bis auf die von I_λ:

$$\overline{I_{\boldsymbol{\lambda}}} = I \diagup I_{\boldsymbol{\lambda}}. \tag{4.97}$$

Durch die Zerlegung der Indexmenge I können die Untermengen Q_λ von Q mit

$$Q_{\lambda} = \left\{ \boldsymbol{p} = \sum_{k \in I_{\lambda}} \boldsymbol{p}_k \mid \boldsymbol{p} \in Q \right\}, \tag{4.98}$$

gebildet werden, die aus Elementen mit Indizes aus I_λ zusammengesetzt werden. Die Menge, die durch ausschließliche Verwendung der Elemente mit den restlichen Indizes entsteht, wird beschrieben durch:

$$\overline{Q_{\lambda}} = \left\{ \boldsymbol{p} = \sum_{k \in \overline{I_{\lambda}}} \boldsymbol{p}_k \mid \boldsymbol{p} \in Q \right\}. \tag{4.99}$$

Das Gesamtproblem besteht darin den Fehler für alle Parameter $\boldsymbol{p}_k$ zu minimieren mit dem gesuchten Parametervektor

$$\boldsymbol{p} \mid e(\boldsymbol{p}) = \min{}_{\text{glob}} \quad \text{für } \boldsymbol{p} \in Q. \tag{4.100}$$

Es wird hierbei unterstellt, daß genau eine Lösung existiert. Teillösungen bestimmen nur die Parameter mit den Indizes der Untermenge $I_{\boldsymbol{\lambda}}$. Die restlichen Parameter $\boldsymbol{p}_k$ mit $k \in \overline{I}_{\boldsymbol{\lambda}}$ sind vorbestimmt, womit die optimale Teillösung von einem vorgegebenen

Parametervektor $\boldsymbol{q_\lambda} \in \overline{Q_\lambda}$ abhängig ist. Die Lösungsmenge ist mit den vorbestimmten Parametern aus $\boldsymbol{q_\lambda}$ zu der Menge $Q_\lambda^{\boldsymbol{q_\lambda}}$ eingeschränkt:

$$Q_\lambda^{\boldsymbol{q_\lambda}} = \{\boldsymbol{p} = \boldsymbol{r_\lambda} + \boldsymbol{q_\lambda} \mid \boldsymbol{r_\lambda} \in Q_\lambda \wedge \boldsymbol{p} \in Q\} \tag{4.101}$$
$$\text{mit vorgegebenem } \boldsymbol{q_\lambda} \in \overline{Q_\lambda}\ .$$

Eine optimale Teillösung $\widehat{\boldsymbol{p}}_\lambda^{\boldsymbol{q}} \in Q_\lambda^{\boldsymbol{q_\lambda}}$ wird durch eine noch zu definierende Operation $\Theta_\lambda(\boldsymbol{q_\lambda})$ erhalten, die von der Vorgabe $\boldsymbol{q_\lambda}$ abhängt:

$$\Theta_\lambda(\boldsymbol{q_\lambda}) \to \widehat{\boldsymbol{p}}_\lambda^{\boldsymbol{q}} \quad \text{mit} \quad \widehat{\boldsymbol{p}}_\lambda^{\boldsymbol{q}} \mid e\left(\widehat{\boldsymbol{p}}_\lambda^{\boldsymbol{q}}\right) \leq e\left(\boldsymbol{p}\right)\ \forall\ \boldsymbol{p}, \widehat{\boldsymbol{p}}_\lambda^{\boldsymbol{q}} \in Q_\lambda^{\boldsymbol{q_\lambda}} \tag{4.102}$$
$$\text{und} \quad \boldsymbol{q_\lambda} \in \overline{Q_\lambda}.$$

$\widehat{\boldsymbol{p}}_\lambda^{\boldsymbol{q}}$ stellt die optimale Teillösung unter den Nebenbedingungen, welche durch $\boldsymbol{q_\lambda}$ beschrieben sind, dar. Damit stellt $\widehat{\boldsymbol{p}}_\lambda^{\boldsymbol{q}}$ in der Regel keine Gesamtlösung dar, in der alle Parameter optimal geschätzt sind. Die Operationen Θ_λ können in einem iterativen Algorithmus dazu verwendet werden Gesamtlösungen zu erzielen, die in das Optimum konvergieren. Dazu muß zuerst eine Zerlegung I_λ von I mit den entsprechenden Operationen Θ_λ vorliegen. Es wird nach einer Zerlegung gesucht, von der die optimalen Teillösungen $\widehat{\boldsymbol{p}}_\lambda^{\boldsymbol{q}}$ berechnet werden können. Dies soll bedeuten, daß Operationen Θ_λ existieren, welche einen determinierten Algorithmus darstellen, der in vorgeschriebenen endlichen Schritten zur optimalen Teillösung führt. Diese Teillösungen werden iterativ ausgeführt, da sie nur unter den Nebenbedingungen $\boldsymbol{q_\lambda}$ ein Minimum angeben und die Nebenbedingungen von Resultaten anderer Teillösungen abhängen. Durch die Abhängigkeit von den Nebenbedingungen sind die Teillösungen gekoppelt. Die Operationen Θ_λ können an sich nur sequentiell abgearbeitet werden. Daher werden die geschätzten Parameter einer Teillösung als vorgegebene Parameter einer anderen Teillösung übergeben, um den Fehler e zu verkleinern. Wendet man auf einen Parametersatz $\{\boldsymbol{p}_k^a = p_k^a \cdot \boldsymbol{u}_k\}$, der aus einer zuvor angewandten Operation $\Theta_{\boldsymbol{\alpha}}$ entstanden ist mit

$$\boldsymbol{p}^a = \Theta_{\boldsymbol{\alpha}}(\boldsymbol{q_\alpha})\ , \quad \boldsymbol{q_\alpha} \in \overline{Q_\alpha}, \tag{4.103}$$

die Operation Θ_λ an, so resuliert der Parametervektor $\boldsymbol{p}^b$:

$$\boldsymbol{p}^b = \Theta_\lambda(\boldsymbol{q}_\lambda^a) \tag{4.104}$$
$$\text{mit} \quad \boldsymbol{q}_\lambda^a = \sum_{k \in \overline{I_\lambda}} \boldsymbol{p}_k^a\ .$$

Der Fehler des neuen Parametervektors $\boldsymbol{p}^b$ kann sich im Vergleich zum vorherigen Parametervektor $\boldsymbol{p}^a$ nur verbessern oder bleibt zumindest gleich:

$$e\left(\boldsymbol{p}^b\right) \leq e\left(\boldsymbol{p}^a\right). \tag{4.105}$$

Die Ungleichung (4.105) folgt aus (4.102) angewandt auf (4.104). Die Teillösung $\boldsymbol{p}^b$ der Operation Θ_λ ist durch die Parameterübergabe von der Lösung $\boldsymbol{p}^a$ der zuvor durchgeführten Operation $\Theta_{\boldsymbol{\alpha}}$ abhängig. In welcher Reihenfolge die Teillösungen abgearbeitet werden, wird durch einen Iterationsalgorithmus festgelegt. Für die Schätzung sämtlicher Parameter werden mehrere Iterationen durchgeführt. Jede Iteration besteht aus

allen Operationen Θ_λ der Zerlegung $\{Q_\lambda\}$ von Q, da abgesehen von günstigen Bedingungen nur durch Verwendung sämtlicher Operationen Θ_λ das globale Minimum gefunden werden kann. Der einfachste Iterationsalgorithmus ist die einmalige Anwendung aller Operationen in einer vorgegebenen Reihenfolge, wobei die Resultate einer Operation immer in die Nebenbedingungen der nächsten Operation einfließen. Damit folgt ganz allgemein mit der Ungleichung (4.105) eine monoton fallende Fehlerentwicklung. Es ist dadurch allerdings nicht für alle Modelle theoretisch gewährleistet, daß das globale Minimum von e erreicht wird, da die Fehlerentwicklung entweder in einem Parametervektor verharren kann, welcher ein lokales Minimum darstellt, oder ungünstigerweise nur mit sehr kleinen Schritten vorwärts kommt. Ob diese Fälle eintreten können, ist von der Fehlerfunktion $e(\boldsymbol{p})$ des gegebenen Problems abhängig.

4.4.1 Schätzung allgemeiner rekursiver Systeme

Hinsichtlich der Nebenbedingungen des zugelassenen Lösungsraums sind allgemeine rekursive Systeme im Vergleich zu erweiterten Rohrmodellen einfacher zu behandeln. Dies hängt mit den Parameterrestriktionen zusammen, da die Pole und Nullstellen eines allgemeinen rekursiven Systems nur die Bedingungen der Minimalphasigkeit erfüllen müssen, aber sonst keine weiteren Einschränkungen erfahren. Das Fehlermaß e der inversen Filterung läßt sich aus (4.66) und (4.67) für allgemeine rekursive Systeme mit der Übertragungsfunktion

$$H(z) = \frac{B(z)}{A(z)} = \frac{1 + \sum_{i=1}^{M} b_i z^{-1}}{1 + \sum_{i=1}^{N} a_i z^{-1}} \tag{4.106}$$

ableiten zu

$$e = \frac{1}{2\boldsymbol{\pi}} \int_0^{2\boldsymbol{\pi}} \left| \frac{1 + \sum_{i=1}^{N} a_i z^{-1}}{1 + \sum_{i=1}^{M} b_i z^{-1}} X(\mathrm{e}^{\mathrm{j}\boldsymbol{\omega}}) \right|^2 \mathrm{d}\boldsymbol{\omega}. \tag{4.107}$$

x stellt das zu analysierende Signal dar. Wie schon mit (4.67) erläutert wird für die Parameterschätzung ein Korrekturfaktor verwendet. Dieser ist hier schon in dem rekursiven System (4.106) berücksichtigt, da die Koeffizienten b_0 und a_0 in $H(z)$ gleich Eins gesetzt sind. Mit Hilfe des Parsevalschen Theorems läßt sich erkennen, daß die Fehlerdefinition (4.107) einen Prädiktionsfehler e mit

$$e(n) = x(n) - \widehat{x}(n) = x(n) - \sum_{k=1}^{N} a_k x(n-k) + \sum_{k=1}^{M} b_k e(n-k) \tag{4.108}$$

darstellt. Dieser Prädiktor berücksichtigt zur Schätzung des nächsten Wertes $x(n)$ auch vergangene Prädiktionsfehlerwerte $e(n-k)$ im Gegensatz zur gewöhnlichen linearen Prädiktion (4.15). Wegen der Minimalphasigkeit des Systems müssen sich sämtliche Pole und Nullstellen im Einheitskreis der Z-Ebene befinden. Dies kann durch eine Darstellung mit Reflexionskoeffizienten r_k^b und r_k^a des Zähler- und Nennerpolynoms einfach kontrolliert werden, da diese dafür vom Betrage kleiner Eins sein müssen. Mit den Polynomkoeffizienten b_k und a_k bzw. den entsprechenden Reflexionskoeffizienten werden nach (4.109) die Parameter $\boldsymbol{p}_k$ bestimmt

$$p_k = b_k \quad \text{für} \quad k = 1 \ldots M \qquad \text{und} \qquad \{b_k\} \longleftrightarrow \left\{r_k^b\right\} \tag{4.109}$$

$$p_{k+M} = a_k \quad \text{für} \quad k = 1 \ldots N \qquad \text{und} \qquad \{a_k\} \longleftrightarrow \{r_k^a\} \ . \tag{4.110}$$

Für die Schätzung der Pole existiert mit der Burg-Methode ein determinierter Algorithmus, der eine optimale minimalphasige Lösung erzielt. Die Burg-Methode wird als Basisalgorithmus für das Pol-Nullstellen Problem verwendet [SnL99c, SnL01b], wobei für die Nullstellenbestimmung die Burg-Methode nicht direkt auf das zu analysierende Signal angewendet werden kann. Für die Teillösungen wird eine Zerlegung der Indexmenge I in zwei Indexmengen I_1 und I_2 vorgenommen, die jeweils die Parameter für die Pole und Nullstellen darstellen:

$$I_2 = \{1 \dots M\} \tag{4.111}$$

$$I_1 = \{M+1 \dots M+N\}\,. \tag{4.112}$$

Für die Teilmengen I_1 und I_2 werden die beiden Operationen $\Theta_{\boldsymbol{\lambda}}$ im Folgenden erklärt.

1. Teillösung Θ_1

Für die erste Teillösung werden die Pole geschätzt, welche durch die Parameter mit den Indizes aus I_1 beschrieben werden. Die Lösungsmenge ist durch $Q_1^{\boldsymbol{q}_1}$ gegeben. Die Parameter der vorgegeben Nullstellen mit $\boldsymbol{q}_1 \in \overline{Q_1}$ stellen die Nebenbedingung für die optimalen Pole dar. Der Einfluß der Nullstellen kann aus dem zu schätzenden Signal durch eine Filterung separiert werden. Dies kann im Zeit- oder Frequenzbereich realisiert werden. Der für die Operation Θ_1 zu minimierende Fehler der ersten Teillösung ergibt sich zu

$$e = \frac{1}{2\boldsymbol{\pi}} \int_0^{2\boldsymbol{\pi}} \left| \left(1 + \sum_{i=1}^{N} a_i z^{-1}\right) \frac{X(\mathrm{e}^{\mathrm{j}\boldsymbol{\omega}})}{\overline{B}(\mathrm{e}^{\mathrm{j}\boldsymbol{\omega}})} \right|^2 \mathrm{d}\boldsymbol{\omega}. \tag{4.113}$$

$\overline{B}(z)$ stellt das vorgegebene Zählerpolynom bzw. die Nullstellen aus $\boldsymbol{q}_1 \in \overline{Q_1}$ dar. Dieser Fehler e kann mittels der Burg-Methode minimiert werden, welche auf das Signal

$$x_B = \mathrm{IDFT}\left(\frac{X(\mathrm{e}^{\mathrm{j}\boldsymbol{\omega}})}{\overline{B}(\mathrm{e}^{\mathrm{j}\boldsymbol{\omega}})}\right) \tag{4.114}$$

angewandt wird. Der Fehler kann mit (4.114) dargestellt werden zu

$$e = \frac{1}{2\boldsymbol{\pi}} \int_0^{2\boldsymbol{\pi}} \left| \left(1 + \sum_{i=1}^{N} a_i z^{-1}\right) X_B(\mathrm{e}^{\mathrm{j}\boldsymbol{\omega}}) \right|^2 \mathrm{d}\boldsymbol{\omega}. \tag{4.115}$$

Die erste partielle Lösung ergibt sich damit zu:

$$\Theta_1(\overline{B}) \equiv \text{Burg-Methode angewendet auf } x_B, \tag{4.116}$$

wobei im Signal x_B der Einfluß der vorbestimmten Nullstellen $\overline{B}$ in x beseitigt ist. Es ist zweckmäßig für die Analyse von stimmhaften Sprachsignalen die Burg-Methode unter der Annahme durchzuführen, daß das Signal x_B als periodisch fortgesetzt angesehen wird. Dafür werden in der Burgmethode die Zustandsspeicher als zyklische Verschiebungen (4.21) realisiert, wodurch die Burgmethode unabhängig von der Phase des zu analysierenden Signals ist. Bei der IDFT in (4.114) kann dann die Phase willkürlich gewählt werden.

2. Teillösung Θ_2

In der zweiten Teillösung werden mit den Parametern der Indexmenge I_2 sämtliche Nullstellen geschätzt. Für die Schätzung werden die Parameter der Indexmenge I_1 durch den Nenner $\overline{A}$ vorgegeben. Der Fehler der zweiten Teillösung ist eigentlich durch

$$e = \frac{1}{2\pi} \int_0^{2\pi} \left| \frac{\overline{A}(\mathrm{e}^{\mathrm{j}\omega}) \cdot X(\mathrm{e}^{\mathrm{j}\omega})}{\left(1 + \sum_{i=1}^{M} b_i z^{-1}\right)} \right|^2 \mathrm{d}\boldsymbol{\omega} \tag{4.117}$$

definiert. Hierfür fehlt es allerdings an der entsprechenden Operation Θ_2. Die Burg-Methode kann nicht direkt angewandt werden, da sie zur Schätzung der Polstellen bestimmt ist. Wenn der Integrand in (4.117) reziprok aufgestellt wird, werden die Nullstellen zu Polstellen umgewandelt und umgekehrt. Das daraus resultierende neue Fehlerintegral ergibt sich dann zu

$$e' = \frac{1}{2\pi} \int_0^{2\pi} \left| \frac{\left(1 + \sum_{i=1}^{M} b_i z^{-1}\right)}{\overline{A}(\mathrm{e}^{\mathrm{j}\omega}) \cdot X(\mathrm{e}^{\mathrm{j}\omega})} \right|^2 \mathrm{d}\boldsymbol{\omega}. \tag{4.118}$$

Der Einfluß der Pole kann analog zu (4.114) bezeitigt werden mit

$$x_A = \mathrm{IDFT}\left(\frac{1}{\overline{A}(\mathrm{e}^{\mathrm{j}\omega}) X(\mathrm{e}^{\mathrm{j}\omega})}\right). \tag{4.119}$$

Mit (4.119) resultiert der Fehler

$$e' = \frac{1}{2\pi} \int_0^{2\pi} \left| \left(1 + \sum_{i=1}^{M} b_i z^{-1}\right) \cdot X_A(\mathrm{e}^{\mathrm{j}\omega}) \right|^2 \mathrm{d}\boldsymbol{\omega}, \tag{4.120}$$

welcher durch die Burg-Methode minimiert werden kann. Die zweite partielle Lösung wird daher durch:

$$\Theta_2'(\overline{A}) \equiv \text{Burg-Methode angewendet auf } x_A \tag{4.121}$$

vertreten. Θ_2' stellt durch die Veränderung der Fehlerdefinition von (4.117) in (4.118) nur einen Ersatz für Θ_2 dar. Auch hier wird die Burg-Methode, angepaßt für periodische Signale, verwendet. Die beiden partiellen Lösungen Θ_1 und Θ_2' mit (4.114) und (4.119) werden iterativ angewendet, um sämtliche Koeffizienten des Zähler- und Nennerpolynoms zu schätzen.

Gesamtlösung mit alternierender Berechnung

Für eine Gesamtlösung werden die beiden Teillösungen Θ_1 und Θ_2' abwechselnd angewendet und die dabei jeweils geschätzten Koeffizienten beim nächsten Schritt berücksichtigt [SnL99c]. Der Algorithmus beginnt mit der Schätzung der Pole, so daß zuerst Θ_1 angewandt wird. Da keine Kenntnis über die Lage der Nullstellen vorhanden ist, wird zu Anfang $\overline{B} = 1$ gesetzt. Das zugehörige Flußdiagramm ist in Bild 4.29 dargestellt.Wie im Diagramm von Bild 4.29 zu sehen ist, wird die Entwicklung

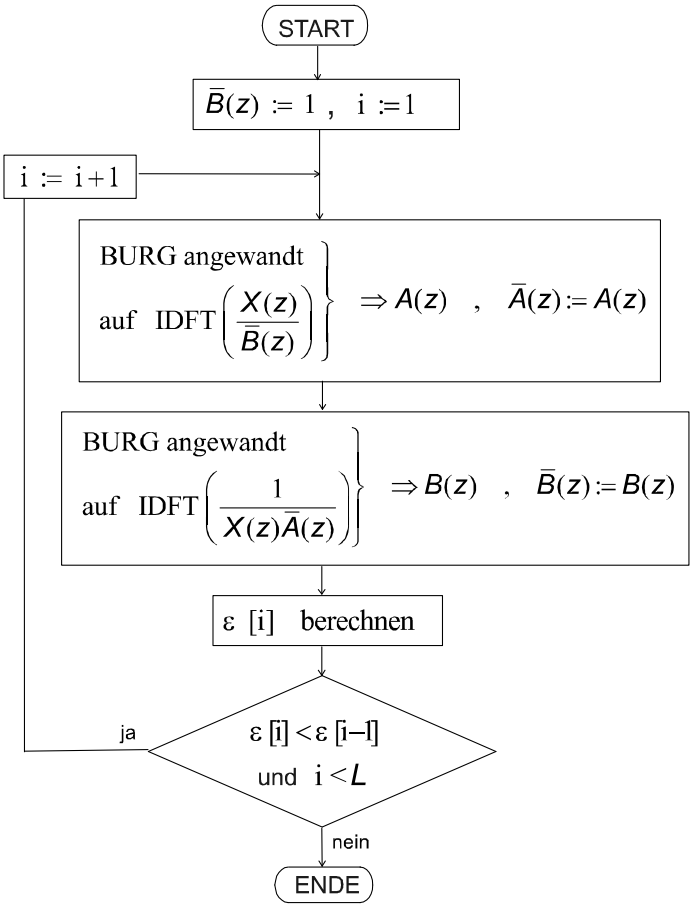


Bild 4.29: Flußdiagramm der Gesamtlösung mit alternierender Berechnung von Θ_1 und Θ_2'.

des Fehlers e überwacht. Der Algorithmus endet, falls eine vorgegebene Anzahl von Iterationen erreicht ist oder wenn sich der Fehler bei einer Iteration im Vergleich zur Vorherigen verschlechtert hat. Ein Überwachen des Fehlers ist erforderlich, da es in durchgeführten Schätzungen vorkommt, daß der Fehler ab einer bestimmten Iteration auf einmal anwächst. Ein Fehleranstieg sollte bei Einhaltung der Forderungen nach den vorangegangenen Abschnitten theoretisch durch (4.105) ausgeschlossen sein. Eine Ursache kann darin liegen, daß mit der ersten Teillösung e minimiert wird, und mit der zweiten Teillösung e', was durch die Verwendung von Θ_2' statt Θ_2 bedingt ist. Die Operationen Θ_2 und Θ_2' liefern für sich alleine betrachtet gute Ergebnisse, minimieren allerdings nicht dieselbe Fehlerdefinition. Dies ist allerdings für die zuvor behandelten Betrachtungen vorausgesetzt worden. Dadurch kann theoretisch die Verwendung von Θ_2' eine Verbesserung des Fehlers e' mit einer gleichzeitigen Verschlechterung des Fehlers e bewirken, so daß mit

$$\boldsymbol{p}^b = \Theta_2'(\boldsymbol{p}^a) \tag{4.122}$$

zwar auf jeden Fall

$$e'\left(\boldsymbol{p}^b\right) \leq e'\left(\boldsymbol{p}^a\right) \tag{4.123}$$

gilt, aber gleichzeitig

$$e\left(\boldsymbol{p}^b\right) \leq e\left(\boldsymbol{p}^a\right) \qquad \text{oder} \qquad e\left(\boldsymbol{p}^b\right) > e\left(\boldsymbol{p}^a\right) \tag{4.124}$$

gelten kann. Gleiches gilt für die Operation

$$\boldsymbol{p}^c = \Theta_1(\boldsymbol{p}^b) \tag{4.125}$$

für die zwar nach (4.105) die Ungleichung

$$e\left(\boldsymbol{p}^c\right) \leq e\left(\boldsymbol{p}^b\right) \tag{4.126}$$

gilt, aber auch gleichzeitig die beiden Ungleichungen

$$e'\left(\boldsymbol{p}^c\right) \leq e'\left(\boldsymbol{p}^b\right) \qquad \text{oder} \qquad e'\left(\boldsymbol{p}^c\right) > e'\left(\boldsymbol{p}^b\right) \tag{4.127}$$

theoretisch möglich sind. Eine gegenseitige Beeinflussung durch die abwechselnde Anwendung der Operationen Θ_1 und Θ_2' statt Θ_2 ist daher nicht auszuschließen. Glücklicherweise ist dieses mögliche Anwachsen des Fehlers bei Testsignalen nur nach Erreichen der optimalen Gesamtlösung anzutreffen und bei Sprachsignalen nach Erreichen einer guten Lösung. Brinker von Philips Research Eindhoven hat etwas später einen ähnlichen Algorithmus in [Br00] vorgestellt. Diese Algorithmen sind unabhängig voneinander entwickelt worden. Bei den Arbeiten von Brinker ist auch ein mögliches Anwachsen des Fehlers festgestellt worden, für das keine Erklärung gefunden wurde. Auch in dieser Arbeit wird das Ansteigen des Fehlerwertes durch Überwachen desselbigen vermieden. Mit Hilfe der theoretischen Betrachtung im vorherigen Abschnitt kann eine Erklärung für den Fehleranstieg gegeben werden. Durch die Unterscheidung der Operationen Θ_2' und Θ_2 kann erklärt werden, warum der Fehler bei diesen Fehlerdefinitionen und Operationen ansteigen kann. Es könnten allerdings auch Ungenauigkeiten infolge der wertdiskreten Realisierung des Algorithmus in Betracht kommen, wie zum Beispiel numerische Rundungsfehler. Eine fehlerhafte Implementierung ist nicht sehr wahrscheinlich, da der Algorithmus in unterschiedlichen Programmiersprachen implementiert wurde, und sogar in Maple mit einer erhöhten Genauigkeit der Fließkommazahlen. Darüber hinaus hat Brinker unabhängig davon auch die Beobachtungen des Fehleranstiegs gemacht. Durch die Verwendung einer Operation Θ_2 statt Θ_2' ist es unter Umständen möglich das Ansteigen des Fehlers e zu vermeiden.

Analyse von Testsignalen

Um die Leistungsfähigkeit des Algorithmus abzuschätzen, werden für die Analyse zuerst Testsignale verwendet. Das in dem dargestellten Beispiel verwendete Testsignal wird durch ein vorgeschriebenes minimalphasiges Pol-Nullstellen System mit 5 Nullstellen und 10 Polstellen erzeugt, welches mit einer Impulsfolge angeregt wird. Eine Periode des Ausgangssignals wird analysiert, wobei die Ordnung des Analysesystems der des erzeugenden Systems entspricht. Der Schätzalgorithmus bestimmt nach dem Flußdiagramm von Bild 4.29 alle Pole und Nullstellen aus einer Periode des Testsignals. Der obere Graph in Bild 4.30 zeigt die Resultate der Schätzung nach der ersten Iteration, während der untere Graph die Ergebnisse nach der vierten Iteration aufzeigt. In Bild 4.30 ist zu sehen, daß die erste Iteration ein Pol-Nullstellen Paar noch nicht modellieren kann. Dies ändert sich schnell mit den nachfolgenden Iterationen, so daß mit der dritten bzw. vierten Iteration die Approximation als nahezu perfekt angesehen werden kann. Es hat sich bei vielen Versuchsbeispielen gezeigt, daß der Algorithmus gerade bei dichten und stark ausgeprägten Pol-Nullstellen Paaren mehrere

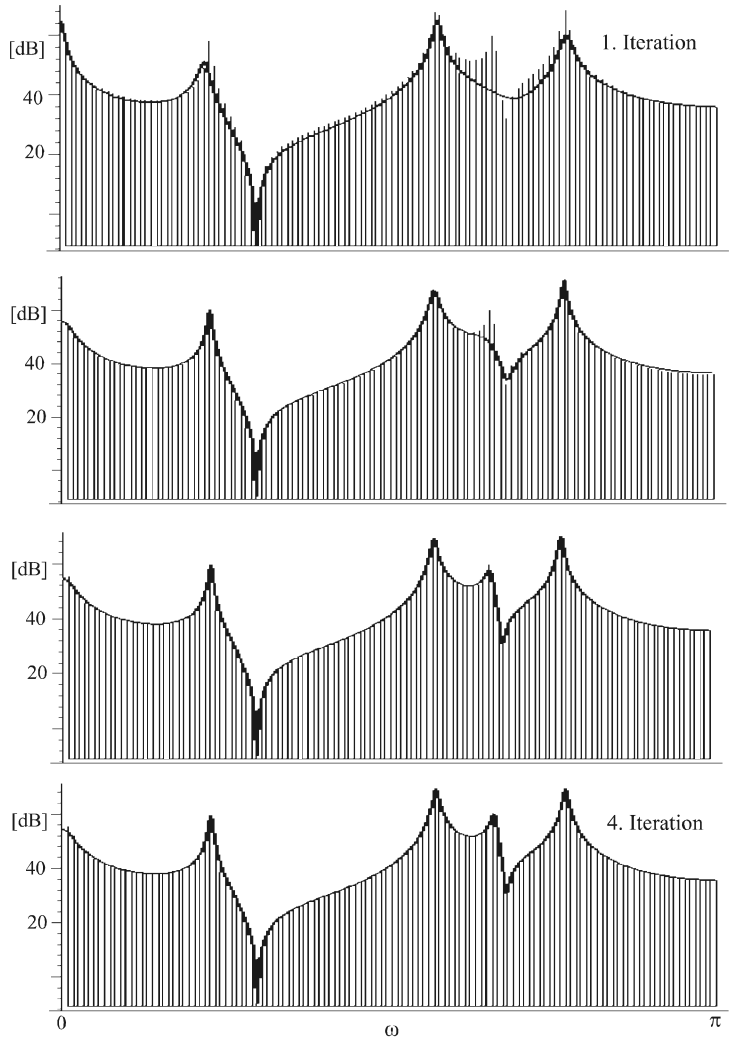


Bild 4.30: Die ersten vier Iterationen der Analyse eines Testsignals mit der ersten Iteration oben beginnend: Geschätzter Betragsgang (durchgezogene Linie), DFT der analysierten Testperiode (Linienspektrum).

Iterationen benötigt, um das Optimum zu finden; der Algorithmus hat allerdings selbst in den durchgeführten Schätzungen der vermeintlich ungünstigsten Fälle die optimale Gesamtlösung immer erreicht.

Analyse von Sprachsignalen

Die Ordnungen des Analyse- und Synthesesystems sind für die Sprachsignalanalysen nicht identisch, im Gegensatz zum zuvor behandelten Beispiel eines Testsignals. Das Sprachsignal ist durch das Synthesesystem des realen Sprechtrakts generiert worden, wodurch das Synthesesystem einen wesentlich höheren Systemgrad aufweist, wie das für die Analyse verwendete. Das System der realen Spracherzeugung enthält genau betrachtet auch Nichtlinearitäten. Für die Analyse muß der Systemgrad vor der Schätzung festgelegt werden. Bild 4.30 zeigt Analysen von einer Periode des Vokals /i:/ mit unterschiedlichem Systemgrad. Die Abtastrate liegt für die analysierten Sprachsignale

bei 16 kHz. Eine Erhöhung des Systemgrades verbessert insbesondere bei niedrigen

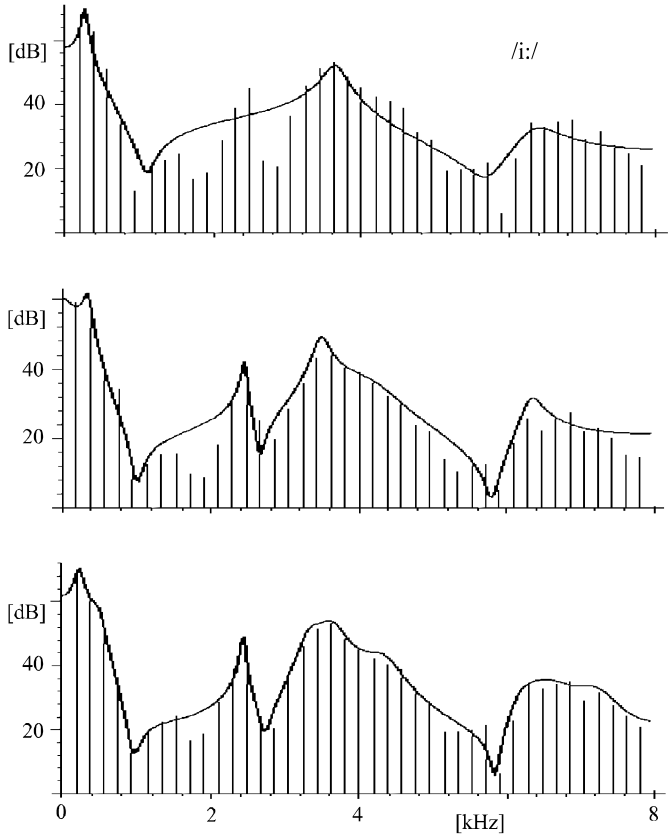


Bild 4.31: Analyse des Vokals /i:/: System mit zwei Nullstellen-Paaren und vier Pol-Paaren (oberes Bild), drei Nullstellen-Paaren und sechs Pol-Paaren (mittleres Bild) sowie vier Nullstellen-Paaren und acht Pol-Paaren (unteres Bild).

Systemgraden die Modellierung. Bild 4.32 zeigt die Resultate von analysierten Perioden von Vokalen und Nasalen, die belegen, daß die Betragsspektren der stimmhaften Sprachlaute größtenteils sehr gut modelliert werden können. Die durchschnittliche Anzahl der Iterationen liegt bei Sprachsignalen etwas über 10.

Gesamtlösung mit parallelen Blöcken

Für den zuvor beschriebenen Algorithmus muß festgelegt werden, ob mit der Schätzung der Pole oder der Nullstellen begonnen werden soll. Diese Unsymmetrie im Algorithmus zwischen den Nullstellen und Polen wird im Folgenden beseitigt. Dafür werden in zwei parallelen Blöcken die Pole und Nullstellen genau einmal geschätzt, aber jeweils mit der umgekehrten Reihenfolge [SnL01b]. Die beiden parallelen Blöcke sind im Flußdiagramm von Bild 4.33 zu sehen. Da kein Vorwissen über die Pole und Nullstellen vorhanden ist, wird mit Koeffizientensätzen $\overline{B} = 1$ und $\overline{A} = 1$ begonnen. Während Block 1 mit der Schätzung der Pole beginnt, beginnt Block 2 mit der Schätzung der Nullstellen. Diese beiden Blöcke werden parallel verarbeitet, da keine Resultate eines Blockes in den anderen Block einfließen, wodurch sie unabhängig voneinander sind. Nachdem beide Blöcke abgearbeitet sind, werden die Fehler e_1 und e_2 von den geschätzten Parametervektoren der jeweiligen Blöcke berechnet. Die Fehlerwerte werden

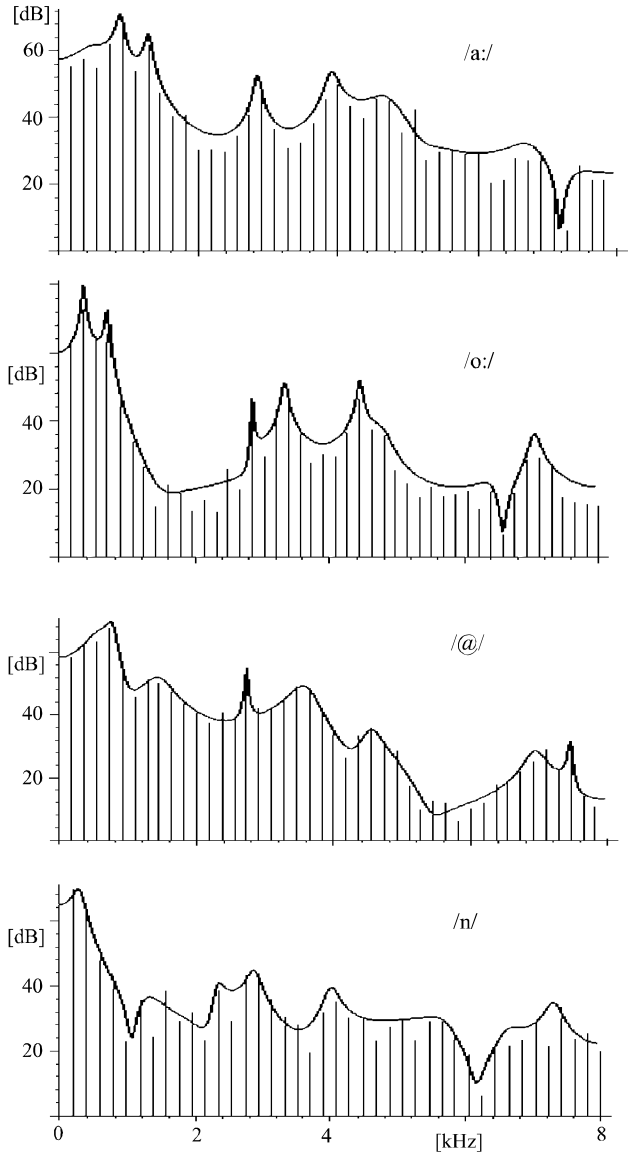


Bild 4.32: Analyse von Sprachlauten mit 20 Polstellen und 10 Nullstellen.

dann mit dem Fehler $e[i]$ der vorherigen i-ten Iteration verglichen. Der Algorithmus endet, falls beide Fehler größer als der vorherige Fehler $e[i]$ sind. Ebenfalls endet der Algorithmus falls eine maximale Anzahl N von Iterationen erreicht ist. Hat sich nur ein Fehler verschlechtert, so wird der Koeffizientensatz mit dem besseren Fehlerwert für die nächste Iteration verwendet. Bei einer Verbesserung beider Fehler e_1 und e_2 im Vergleich zu $e[i]$ werden die resultierenden Koeffizientensätze von beiden Blöcken kombiniert [SnL01b]. Dies wird durch das arithmetische Mittel der Reflexionskoeffizienten

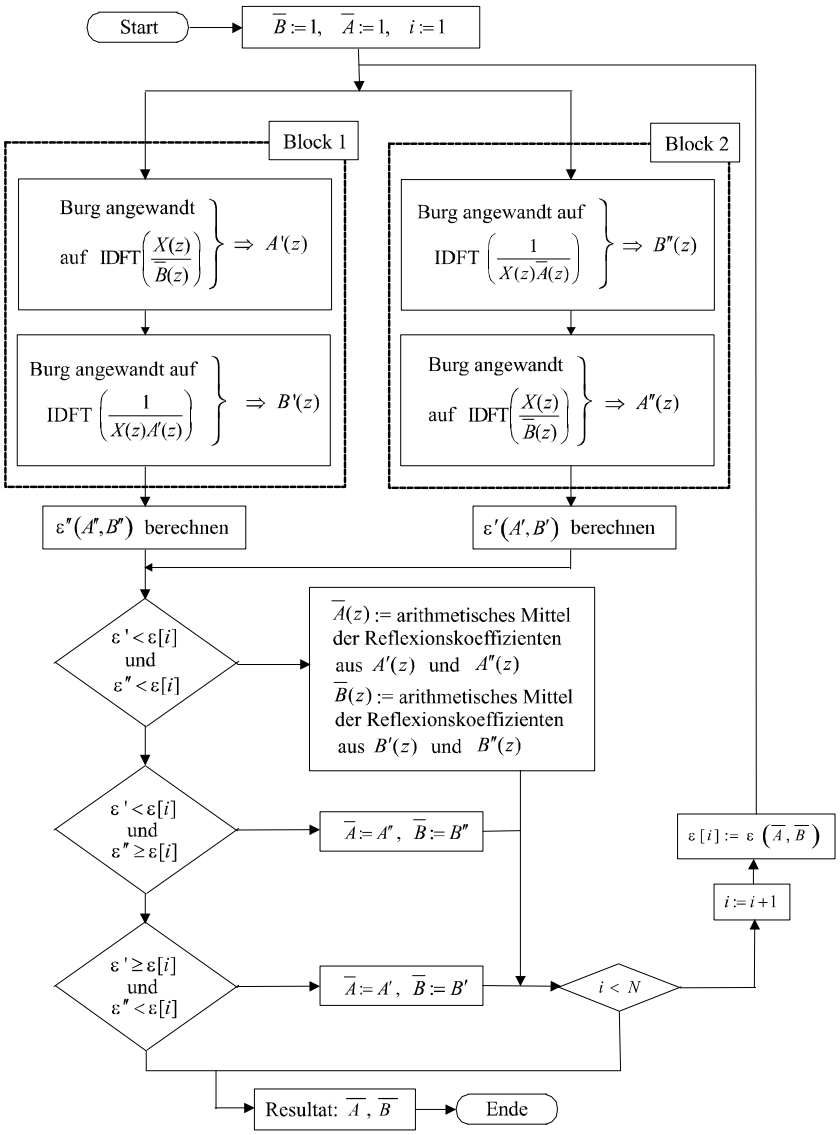


Bild 4.33: Flußdiagramm der Gesamtlösung mit Verwendung von zwei parallelen Blöcken.

realisiert:

$$\{a_i'\} \rightarrow \{r_i'^a\} \quad \text{und} \quad \{a_i''\} \rightarrow \{r_i''^a\} \tag{4.128}$$

$$\widehat{r}_i^a = \frac{r_i'^a + r_i''^a}{2}$$

$$\{b_i'\} \rightarrow \{r_i'^b\} \quad \text{und} \quad \{b_i''\} \rightarrow \{r_i''^b\}$$

$$\widehat{r}_i^b = \frac{r_i'^b + r_i''^b}{2}$$

$$\{\widehat{r}_i^a\} \rightarrow \{\widehat{a}_i\} \quad \text{und} \quad \{\widehat{r}_i^b\} \rightarrow \{\widehat{b}_i\}. \tag{4.129}$$

Die Koeffizienten mit einem und zwei Strichen stellen die jeweiligen Schätzergebnisse der Blöcke dar. Aus den resultierenden Polynomkoeffizienten $\widehat{a}_i$ und $\widehat{b}_i$ sind $\overline{A}$ und $\overline{B}$ gegeben. Durch die Mittelung der Reflexionskoeffizienten ist gewährleistet, daß die

Lösung wieder minimalphasig ist. Diese Mittelung könnte auch in den Flächen oder in anderen Darstellungen erfolgen, welche von den Reflexionskoeffizienten abhängen. Der Graph in Bild 4.34 oben zeigt das Ergebnis einer Analyse eines Testsignals nach der ersten Iteration, während in Bild 4.34 unten das Ergebnis nach der 15. Iteration dargestellt ist. Die Systemordnung der Analyse der Testsignale richtet sich nach der

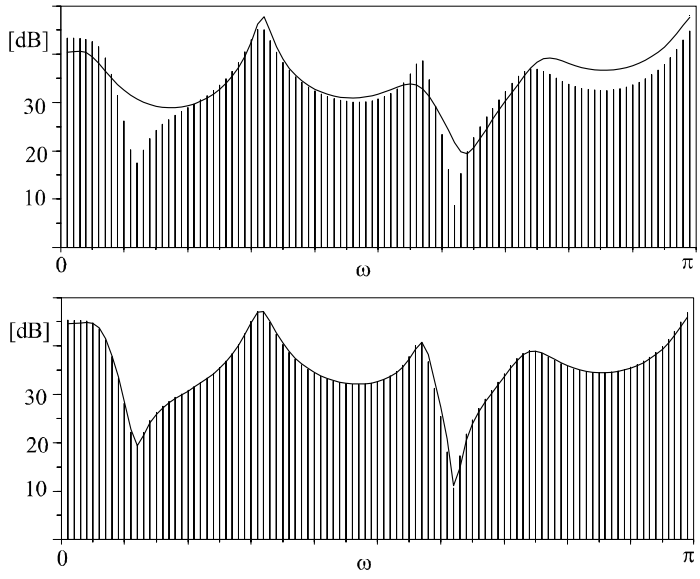


Bild 4.34: Analyse eines Testsignals mit Verwendung von zwei parallelen Blöcken: Ergebnisse nach der ersten Iteration (oben) und nach der 15. Iteration (unten).

Ordnung des erzeugenden Systems. Im Folgenden werden Analysen von Sprachsignalen mit einer Abtastrate von 16 kHz gezeigt. In Bild 4.35 ist das Ergebnis der Analyse des Nasals /n/ mit 20 Polstellen und 10 Nullstellen zu sehen. Bei Nasalen und nasalierten

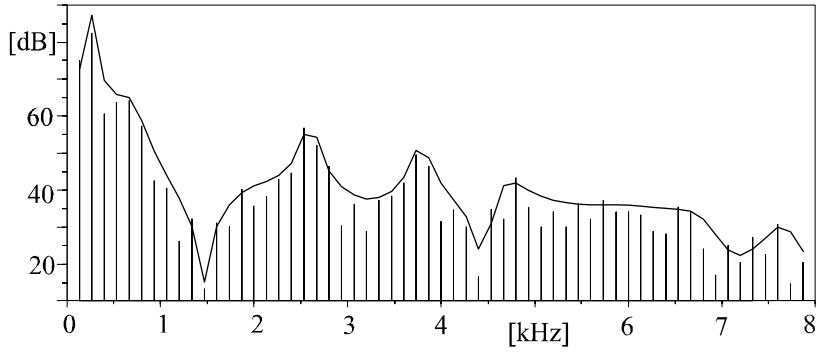


Bild 4.35: Analyse des Nasals /n/ mit 20 Polen und 10 Nullstellen.

Lauten ist der Sprechtrakt verzweigt, wodurch Nullstellen in der Übertragungsfunktion vorhanden sind. Für Vokale mit geschlossenem Velum ist die Rohrstruktur an sich unverzweigt, wenn mögliche kleine Nebenhöhlen (z.B. Recessus piriformis / piriform fossa) vernachlässigt werden. Dennoch ist eine Analyse mit Polen und Nullstellen auch für Vokale sinnvoll, da einerseits der Vokaltrakt selbst eine Struktur aufweist,

die nicht eine einfache Röhre darstellt, und andererseits die Anregung mit berücksichtigt werden muß. Bei Phonation kann der subglottale Trakt kurzzeitig während der geöffneten Glottisphase [Fr94, Tir79] den Einfluß von Nullstellen bewirken. Diese lassen sich an dem vereinfachten Beispiel des Rohrmodells mit einer Anregung in der Rohrmitte durch (3.97) erklären. Nasalierte Vokale besitzen durch die Ankopplung des Nasaltraktes oft ausgeprägte Nullstellen. Neben einer möglichen Nasaltraktankopplung existieren noch zwei symmetrisch angeordnete kleine Nebenhöhlen (Recessus piriformis / piriform fossa) kurz vor der Glottis, welche nach [Da96a, Da96b] Nullstellen bei 4 bis 5 kHz bewirken. In Bild 4.36 ist ein Beispiel aufgezeigt mit der Analyse einer Periode des Vokals /i:/ bei dem Nullstellen die Modellierung des Vokalspektrums im Vergleich zum Nur-Pole Modell verbessern. In [Vi98] sind auch Unterschiede zwischen Sprachanalysen mittels Nur-Pole und Pol-Nullstellen Modellen gezeigt. In [Vi98] wird die ARMA-Schätzung aus einer Reihenentwicklung der Cepstralkoeffizienten gewonnen. In Bild 4.36 oben ist die Periode des Vokals /i:/ mit 30 Polstellen analysiert.

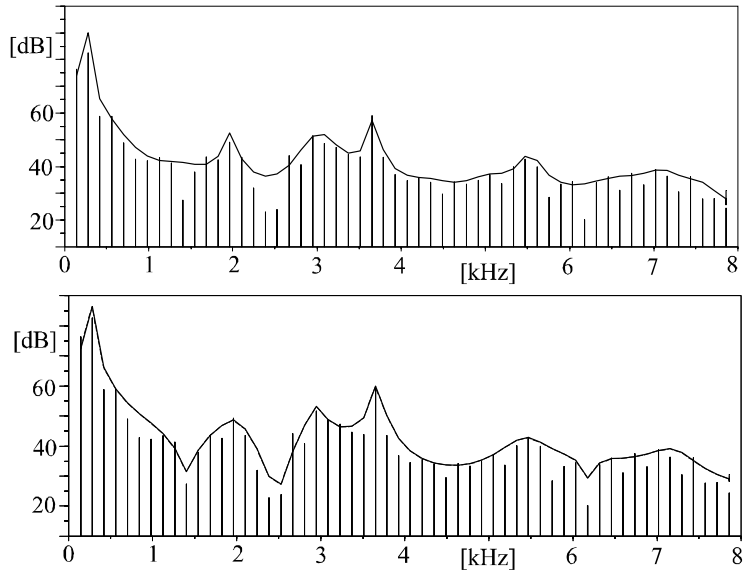


Bild 4.36: Analyse des Vokals /i:/: System mit 30 Polen (oben), System mit 20 Polen und 10 Nullstellen (unten).

Es ist zu sehen, daß die Einbuchtungen im Betragsspektrum durch Polstellen allein nicht ausreichend modelliert werden können, wenn nicht zu sehr hohen Systemgraden zurückgegriffen wird. In Bild 4.37 ist das Ergebnis der Analyse des nasalierten Vokals /$\tilde{\text{a}}$/ und in Bild 4.38 das Analyseergebnis des stimmhaften Frikativs /z/ gezeigt, welche einen ähnlichen Effekt beschreiben. Stimmlose Sprachlaute können auch durch den Algorithmus analysiert werden, wie das Beispiel des Frikativs /S/ in Bild 4.39 zeigt. Das analysierte Sprachsignal des Frikativs wurde mit einem von Hann Fenster gewichtet. Bei der Analyse von stimmlosen Signalen besteht das Problem, daß der unstetige Verlauf des Spektrums der rauschhaften Anregung zuzuschreiben ist und nicht vom Rohrsystem modelliert werden soll. Durch die Rauschanregung entsteht ein zackiges Betragsspektrum, in dem die schmalen Nischen durch Nullstellen modelliert werden können. Um dies zu verhindern muß das Sprachspektrum geglättet werden. Dies kann zum Beispiel durch eine homomorphe Analyse mit einer Lifterung im cepstral-Bereich

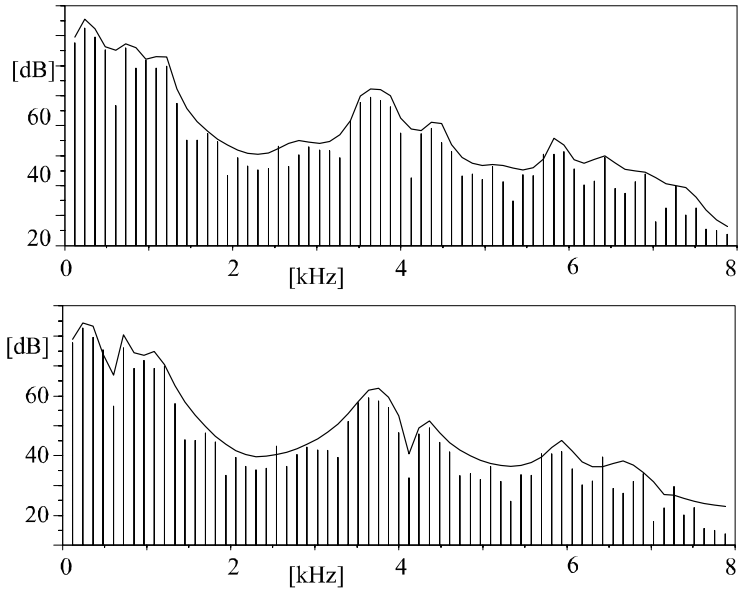


Bild 4.37: Analyse des nasalierten Vokals /$\tilde{\text{a}}$/ : System mit 30 Polen (oben), System mit 20 Polen und 10 Nullstellen (unten).

erreicht werden.

Vergleich der Algorithmen

Es zeigt sich, daß der Algorithmus mit den parallelen Blöcken bei Sprachsignalen überwiegend bessere Resultate erzielt, als der Algorithmus mit den abwechselnden Berechnungen. Dies ist insbesondere bei Spektralverläufen zu beobachten, die weniger glatt verlaufen. Der Algorithmus mit den alternierenden Berechnungen konvergiert schneller in eine Lösung als die andere Methode. Bei der Analyse von Testsignalen liefern beide Algorithmen nahezu perfekte Resultate, wobei der Algorithmus mit den parallelen Blöcken eine höhere Anzahl von Iterationen für die optimale Lösung benötigt. Die gezeigten Beispiele der allgemeinen rekursiven Systeme besitzen außer der Minimalphasigkeit des Systems keine Einschränkungen der Pole und Nullstellen. In den folgenden Beispielen werden Rohrstrukturen untersucht, welche die möglichen Pol-Nullstellen Konfigurationen stärker einschränken. Dies wirkt sich auf den Algorithmus in der Weise aus, daß für die Zerlegung von I wesentlich mehr Untermengen benötigt werden.

4.4.2 Analyse von Modellen mit zwei vorgegebenen Rohrabschlüssen

Im Folgenden wird das unverzweigte Rohrmodell für den Vokaltrakt behandelt. Der Lippenabschluß wird entweder frequenzabhängig mit $L(z)$ von (4.70) [SnL01a, SnL01c] oder frequenzunabhängig mit einem reellen Abschlußkoeffizienten [SnL00e] realisiert. Neben diesem Lippenabschluß soll auch ein Rohrabschluß an der Glottis vorgegeben werden, der durch eine reelle Konstante beschrieben oder frequenzabhängig modelliert wird. Eine Zeitabhängigkeit des Glottisabschlusses wird wegen der benötigten subop-

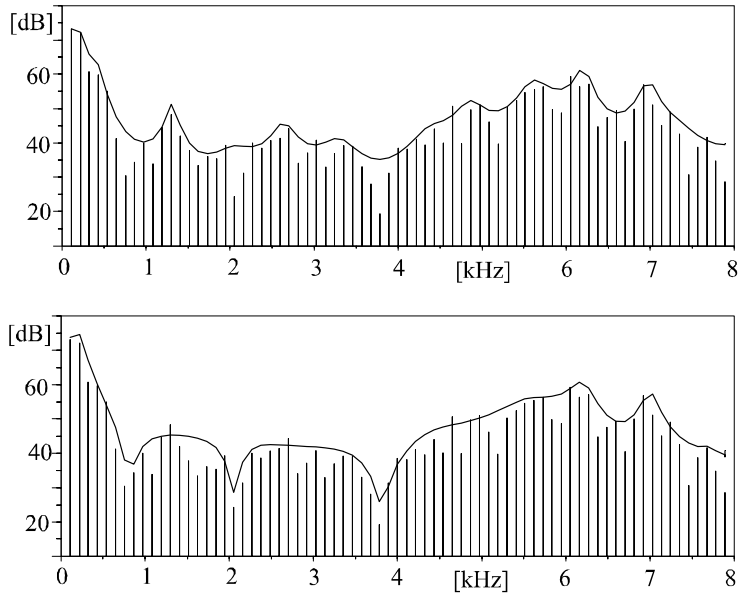


Bild 4.38: Analyse des Lautes /z/ : System mit 30 Polen (oben), System mit 20 Polen und 10 Nullstellen (unten).

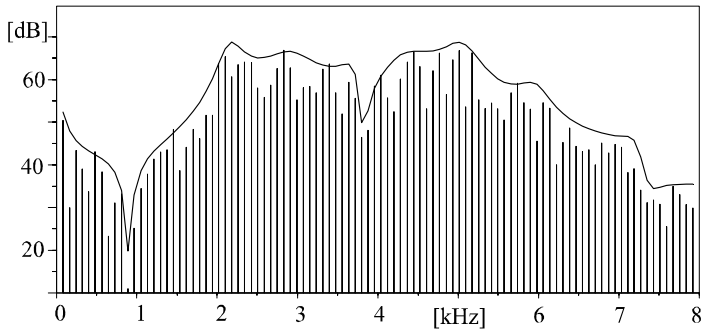


Bild 4.39: Analyse des stimmlosen Frikativs /S/ mit 20 Polen und 10 Nullstellen.

timalen Schätzoperationen Θ_i in diesem Fall nicht berücksichtigt. Das Schätzproblem besteht darin, die Reflexionskoeffizienten zwischen den beiden vorgegebenen Abschlüssen zu bestimmen. Dazu wird das inverse Rohrfilter benutzt, welches in Bild 4.40 dargestellt ist. Das inverse Filter besteht neben den Abschlüssen aus einer Verkettung von Zweitoren $\boldsymbol{T}_i$ angepaßt an periodische Signale:

$$\boldsymbol{T}_i = \begin{pmatrix} 1 & r_i \cdot zv^{-1} \\ r_i & zv^{-1} \end{pmatrix}. \tag{4.130}$$

Durch die Analyse von Perioden wird die zyklische Verschiebung zv^{-1} statt z^{-1} verwendet. Der Vorfaktor $d(r)$ in (3.48) ist konstant zu Eins gewählt, um die Modifikation (4.67) für die inverse Filterung zu berücksichtigen. Die oberen und unteren Wellengrößen im inversen Filter können an der i-ten Stelle als Vektor $\boldsymbol{x}_i$ dargestellt werden mit

$$\boldsymbol{x}_i = \begin{pmatrix} x_i^o \\ x_i^u \end{pmatrix}, \tag{4.131}$$

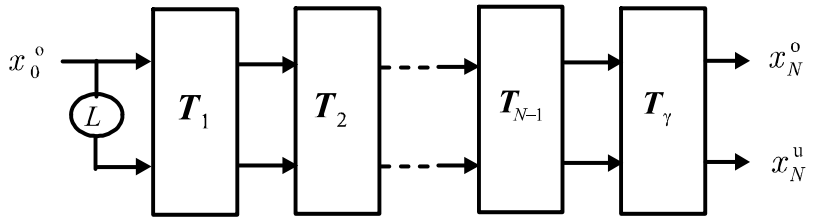


Bild 4.40: Inverses Rohrmodell mit zwei festen Rohrabschlüssen.

welche durch die Betriebskettenmatrizen $\boldsymbol{T}_i$ mittels

$$\boldsymbol{x}_i = \boldsymbol{T}_i \cdot \boldsymbol{x}_{i-1} \quad \text{und} \quad \boldsymbol{x}_N = \boldsymbol{T}_\gamma \cdot \boldsymbol{x}_{N-1} \tag{4.132}$$

miteinander verknüpft werden. $\boldsymbol{T}_\gamma$ beschreibt den Glottisabschluß als Zweitor. Im Gegensatz zum Synthesefilter befindet sich der Lippenabschluß $L(z)$ am Systemeingang und der Glottisabschluß G am Systemausgang. Es sei angemerkt, daß die Nummerierung der Zweitore des inversen Filters hier in umgekehrter Reihenfolge im Vergleich zum Synthesefilter vorgenommen werden. Die Anzahl der Rohrelemente zwischen den beiden Abschlüssen bestimmt die Vokaltraktlänge und muß vor der Parameterschätzung festgelegt werden. Die ermittelten Vokaltraktlängen aus NMR- oder Röntgenaufnahmen aus der Literatur werden dafür wieder als Anhaltspunkte verwendet. Die Mundöffnungsfläche des Laine-Modells wird ebenfalls vor der Parameterschätzung lautabhängig eingestellt. Durch den vorgegebenen Lippenabschluß liefert der Burg-Algorithmus keine optimalen Ergebnisse mehr, da der untere und obere Pfad im inversen Filter nicht mehr die Vorwärts- und Rückwärtsprädiktionsfehler darstellen. Der festgelegte Glottisabschluß kann durch die Burg-Methode nicht berücksichtigt werden. Für die Bestimmung der Reflexionskoeffizienten muß eine Minimierung des Fehlers (4.67) durchgeführt werden. Der Filterausgang ist mit x_N^o gegeben, womit die Parameterbestimmung durch inverse Filterung erfolgt mit dem Fehlerkriterium:

$$e_o = \quad \mathrm{E}\left[(x_N^o)^2\right] \rightarrow \min. \tag{4.133}$$

Analog zu dem Burgkoeffizienten wird auch alternativ die Verwendung des arithmetischen Mittels von x_N^o und x_N^u getestet mit:

$$e_b = \quad \mathrm{E}\left[(x_N^o)^2 + (x_N^u)^2\right] \rightarrow \min. \tag{4.134}$$

Die Minimierung von (4.134) erfüllt allerdings nur noch in Spezialfällen das Kriterium der inversen Filterung, da nicht allein die Ausgangsleistung des inversen Filters berücksichtigt wird. Im Gegensatz zur Burgmethode wird durch (4.133) und (4.134) für die Schätzungen aller Koeffizienten die Ausgangsleistung hinter dem letzten Zweitor minimiert. Die beiden Erwartungswerte (4.133) und (4.134) können für einen einzelnen Reflexionskoeffizienten r_i minimiert werden unter der Voraussetzung, daß die übrigen Koeffizienten vorgegeben sind. Dadurch folgt eine Zerlegung der Indexmenge I in N bzw. $N-1$ Untermengen falls das letzte Tor einen festen Glottisabschluß besitzt:

$$I_i = \{i\} \quad , \text{ für } \; i = 1 \ldots N \;\; \text{bzw. } 1 \ldots N-1. \tag{4.135}$$

Für die Schätzung eines Koeffizienten r_i werden die beiden Eingänge x_{i-1}^o und x_{i-1}^u des zu schätzenden Zweitores benötigt, welche in der Weise bis zum letzten Zweitor gefiltert

werden sollen, so daß dessen Ausgangsleistung minimal wird. Daran ist zu erkennen, daß die Resultate des zu schätzenden Zweitors $\boldsymbol{T}_i$ von allen dahinterliegenden Zweitoren abhängen, welche durch die Matrix

$$\boldsymbol{F}_i = \begin{pmatrix} F_i^{11} & F_i^{12} \\ F_i^{21} & F_i^{22} \end{pmatrix} \tag{4.136}$$

beschrieben wird mit

$$\boldsymbol{x}_N = \boldsymbol{F}_i \cdot \boldsymbol{T}_i \cdot \boldsymbol{x}_{i-1}. \tag{4.137}$$

Die Matrix $\boldsymbol{F}_i$ ergibt sich nach (4.137) zu:

$$\boldsymbol{F}_i = \begin{cases} \boldsymbol{T}_\gamma \cdot \prod_{k=1}^{N-i-1} \boldsymbol{T}_{N-k} & \text{für } i < N-1 \\ \boldsymbol{T}_\gamma & \text{für } i = N-1\ . \end{cases} \tag{4.138}$$

$\boldsymbol{T}_\gamma$ stellt das letzte Zweitor im inversen Filter mit dem Glottisabschluß dar:

$$\boldsymbol{T}_\gamma = \begin{pmatrix} 1 & G(z) \cdot zv^{-1} \\ G(z) & zv^{-1} \end{pmatrix}. \tag{4.139}$$

Die beiden unteren Matrixelemente von $\boldsymbol{T}_\gamma$ haben auf das Synthesefilter keine Auswirkung, da bei diesem Filter am Glottisabschluß nur der Systemausgang im oberen Signalflußpfad aktiv ist. Die beiden unteren Elemente werden an die Struktur einer Betriebskettenmatrix angepaßt, so daß sich für einen reellen Glottisabschluß mit $G(z) = r_N$ eine Betriebskettenmatrix $\boldsymbol{T}_N$ ergibt. Für die Fehlerdefinition (4.133) sind die unteren Elemente von $\boldsymbol{T}_\gamma$ ohne Bedeutung. Die Elemente der Matrix $\boldsymbol{F}_i$ sind Polynome in zv^{-1}:

$$F_i^{\lambda\beta} = \sum_k = f_i^{\lambda\beta}(k) \cdot zv^{-k}. \tag{4.140}$$

Die Polynomkoeffizienten $f_i^{\lambda\beta}(k)$ hängen von $G(z)$ und den Reflexionskoeffizienten r_k für $k > i$ ab und stellen eine Impulsantwort dar. Falls $G(z)$ ein rekursives System darstellt, besitzen die Elemente $F_i^{\lambda\beta}$ auch ein Nennerpolynom. Dieser Fall wird hier nicht betrachtet, da $G(z)$ als Konstante oder FIR-System angenommen wird, womit die Impulsantworten $f_i^{\lambda\beta}(k)$ ein nicht rekursives Teilsystem darstellen und somit Einschwingvorgänge nicht berücksichtigt werden müssen. Die Kriterien (4.133) und (4.134) für die optimal geschätzten Koeffizienten $\widehat{r}_i$ und $\widehat{r}_i^B$ sind durch die Bedingungsgleichungen

$$\frac{\partial \mathrm{E}\left[\left(x_N^o\right)^2\right]}{\partial r_i} = 0 \qquad \Rightarrow \widehat{r}_i \tag{4.141}$$

und

$$\frac{\partial \left(\mathrm{E}\left[\left(x_N^o\right)^2\right] + \mathrm{E}\left[\left(x_N^u\right)^2\right]\right)}{\partial r_i} = 0 \qquad \Rightarrow \widehat{r}_i^B \tag{4.142}$$

gegeben. Das Zweitor $\boldsymbol{T}_i$ in (4.137) hängt von dem zu schätzenden Koeffizienten r_i ab, während $\boldsymbol{F}_i$ vorgegeben ist. Damit läßt sich mit Hilfe von (4.137) die Bedingungsgleichungen (4.141) und (4.142) nach der unbekannten Variable r_i auflösen, womit $\widehat{r}_i$ und $\widehat{r}_i^B$ vorliegen. Die suboptimalen Lösungen ergeben sich nach (4.141) [SnL00e] durch den Koeffizienten

$$\widehat{r}_i = -\frac{\overline{o_i^{11} o_i^{12}} + \overline{o_i^{11} u_i^{11}} + \overline{u_i^{12} o_i^{12}} + \overline{u_i^{12} u_i^{11}}}{\overline{o_i^{12} o_i^{12}} + 2 \cdot \overline{o_i^{12} u_i^{11}} + \overline{u_i^{11} u_i^{11}}} \tag{4.143}$$

und für (4.142) durch

$$\widehat{r}_i^B = -\frac{\overline{o_i^{21}o_i^{22}} + \overline{o_i^{21}u_i^{21}} + \overline{u_i^{22}o_i^{22}} + \overline{u_i^{22}u_i^{21}} + \overline{o_i^{11}o_i^{12}} + \overline{o_i^{11}u_i^{11}} + \overline{u_i^{12}o_i^{12}} + \overline{u_i^{12}u_i^{11}}}{\overline{o_i^{22}o_i^{22}} + 2 \cdot \overline{o_i^{22}u_i^{21}} + \overline{u_i^{21}u_i^{21}} + \overline{o_i^{12}o_i^{12}} + 2 \cdot \overline{o_i^{12}u_i^{11}} + \overline{u_i^{11}u_i^{11}}} \tag{4.144}$$

mit den zyklisch gefalteten Signalen

$$\begin{aligned} o_i^{\lambda\beta} &= f_i^{\lambda\beta}(n) * x_{i-1}^o(n) \\ u_i^{\lambda\beta} &= f_i^{\lambda\beta}(n) * x_{i-1}^u(n-1). \end{aligned} \tag{4.145}$$

Die Erwartungswerte E[x] sind in (4.143) und (4.144) durch Mittelwerte $\overline{x}$ ersetzt. In den folgenden Beispielen wird immer Periodizität des analysierten Signals angenommen. Die Verzögerung in $x_{i-1}^u(n-1)$ von $u_i^{\lambda\beta}$ in (4.145) wird durch den unteren Zustandsspeicher von $\boldsymbol{T}_i$ verursacht und wird durch eine zyklische Verschiebung realisiert. $\widehat{r}_i$ und $\widehat{r}_i^B$ repräsentieren für zwei verschiedene Fehlerdefinitionen die Operationen Θ_i, welche den i'ten Koeffizienten optimal berechnen unter der Voraussetzung, daß die anderen Koeffizienten bekannt sind. Es existieren $N-1$ Operationen Θ_i entsprechend den $N-1$ zu schätzenden Reflexionskoeffizienten. Die Operationen werden nacheinander angewandt beginnend bei Θ_1 bis Θ_{N-1}, wobei in der nächsten Iteration wieder mit Θ_1 begonnen wird. Die Ergebnisse der vorherigen Iteration gehen in die Matrix $\boldsymbol{F}_i$ ein, was sich auf die Berechnung von $\widehat{r}_i$ und $\widehat{r}_i^B$ durch (4.145) auswirkt. Jeder Reflexionskoeffizient wird in einer Iteration genau einmal geschätzt. Für die Startkonfiguration sind alle Koeffizienten r_i zu Null gesetzt.

Korrektur der geschätzten Koeffizienten

Mit der Berechnungsformel (4.143) ist nicht immer gewährleistet, daß der geschätzte Reflexionskoeffizient vom Betrage kleiner Eins ist. Dies kann durch den Lippenabschluß erklärt werden, der von den Werten ± 1 abweicht. Als erläuterndes Beispiel wird angenommen, daß ein Testsignal analysiert wird, welches durch Anregung des stabilen Systems

$$H'(z) = \frac{1}{1 - 0,9 \cdot z^{-1}} \tag{4.146}$$

mit weißem Rauschen oder einer Impulsfolge erzeugt wurde. Wird der Lippenabschluß mit -0,8 für das Analysefilter vorgegeben und bleibt der Glottisabschluß mit $G(z) = 0$ unberücksichtigt, so ergibt sich für ein Rohr mit genau einem Reflexionskoeffizienten der optimale Koeffizient r_1 zu 9/8, der vom Betrage größer als Eins ist. Durch den angenommenen Lippenabschluß von -0,8 ergibt sich dennoch ein stabiles Synthesefilter mit der Übertragungsfunktion

$$H(z) = \frac{1}{1 - 0,8 \cdot 9/8 \cdot z^{-1}} = \frac{1}{1 - 0,9 \cdot z^{-1}}. \tag{4.147}$$

Das System H hat dieselbe Polstelle wie H' innerhalb des Einheitskreises. Um das Rohrmodell konsistent zu halten, müssen die Reflexionskoeffizienten vom Betrage kleiner gleich Eins sein, da negative Flächen dem physikalischen Modell widersprechen. Deshalb werden die nach $\widehat{r}_i$ geschätzten Koeffizienten bei Verletzung der Konsistenz vom Betrag auf 0,99 gesetzt bzw. einem vergleichbaren Wert kleiner Eins. Diese Korrektur kommt bei Sprachsignalen hauptsächlich dann vor, wenn ohne Präemphase analysiert wird, da für die Modellierung des starken spektralen Abfalls der Einhüllenden

betragsmäßig große Werte der Reflexionskoeffizienten benötigt werden. Mit Verwendung einer Präemphase tritt eine Korrektur nur gelegentlich auf. Im Folgenden werden die Schätzalgorithmen auf Rohrmodelle mit verschiedenen Abschlüssen angewandt.

Reeller Rohrabschluß an der Glottis

Als erstes wird der Fall eines reellen Glottiskoeffizienten mit

$$G(z) = r_g \tag{4.148}$$

behandelt. Dadurch kann $\boldsymbol{T}_\gamma$ als gewöhnliches Zweitor mit Querschnittssprung beschrieben werden, wodurch r_g mit einem negativen Wert anzunehmen ist. Bei geschlossener Glottis würde sich als verlustloser harter Abschluß idealisiert ein Glottiskoeffizient von -1 ergeben. Der Betrag des Abschlusses r_g ist allerdings selbst in diesem Falle durch Berücksichtigung von Verlusten kleiner als Eins anzunehmen. Dabei muß zusätzlich bedacht werden, daß die Glottis bei Phonation auch eine geöffnete Phase besitzt, wodurch die Verluste stärker anzunehmen sind. Der Rohrabschluß an der Glottis kann analog zum Lippenabschluß nicht nur den Einfluß des Abschlusses berücksichtigen, sondern auch Dämpfungen innerhalb des Vokaltraktes, was den Betrag des Glottiskoeffizienten zusätzlich verkleinert.

Analyse von Vokalen

Die folgenden Resultate von Sprachsignalanalysen wurden mit $r_g = -0,5$ und $\boldsymbol{\alpha} = 0,85$ für $L(z)$ erzielt. Die Sprachsignale wurden dafür mit der adaptiven Präemphase vorgefiltert. Die Bilder 4.41 bis 4.45 zeigen Ergebnisse von analysierten Vokalen mit einer sequentiellen Anwendung der partiellen Lösungen Θ_i durch (4.143) bzw. (4.144). Im Bild 4.41 sind die Schätzungen der Betragsgänge nach der ersten und zwanzigsten Iteration für den Vokal /a:/ zu sehen. Es ist deutlich zu erkennen, daß die erste Iteration noch kein annehmbares Ergebnis erzielt. Die Fehlerentwicklung in Abhängigkeit der Iterationsanzahl ist in den Bildern 4.43 und 4.44 für die beiden Fehlerdefinitionen gezeigt. Der Startfehler ist dafür auf Eins normiert worden. Es ist zu sehen, daß die Verwendung des Fehlers e_o eine schnellere Konvergenz ermöglicht als die Fehlerdefinition e_b. Die Verwendung des Fehlers e_b liefert ähnliche Ergebnisse wie e_o, allerdings mit dem Unterschied daß die geschätzten Resonanzen oft zu schwach ausgeprägt sind, wie auch in Bild 4.41 zu erkennen ist. Demzufolge ist die Fehlerdefinition e_o zu favorisieren, was sich durch die Einhaltung des Fehlerkriteriums (4.67) auch theoretisch begründen läßt. Hierfür ist nicht der Korrekturfaktor das Entscheidende, sondern daß nur die Leistung des Systemausgangs allein minimiert wird. Durch den vordefinierten Glottisabschluß ergibt sich die Vokaltraktlänge durch die Anzahl der Rohrelemente, welche vor der Parameterschätzung festgelegt werden muß. Die Abtastrate der analysierten Sprachsignale liegt bei 16 kHz. Da die genaue Sprechtraktlänge unbekannt ist, werden mehrere Analysen mit unterschiedlichen Vokaltraktlängen durchgeführt. Die resultierenden Fehlerwerte sind in Bild 4.45 in Abhängigkeit der Rohrelementeanzahl dargestellt. Es ist zu erkennen, daß die Vokale /i:/ und /a:/ ein erstes Fehlerminimum bei 15 und 16 Rohrelementen aufweisen und ein zweites bei 20. Da die tatsächliche Vokaltraktlänge bei diesen Lauten deutlich unter 20 Rohrelemente liegt, wird ein Wert um das erste Minimum für die Vokaltraktlänge verwendet. Bei dem Vokal /o:/ stellt

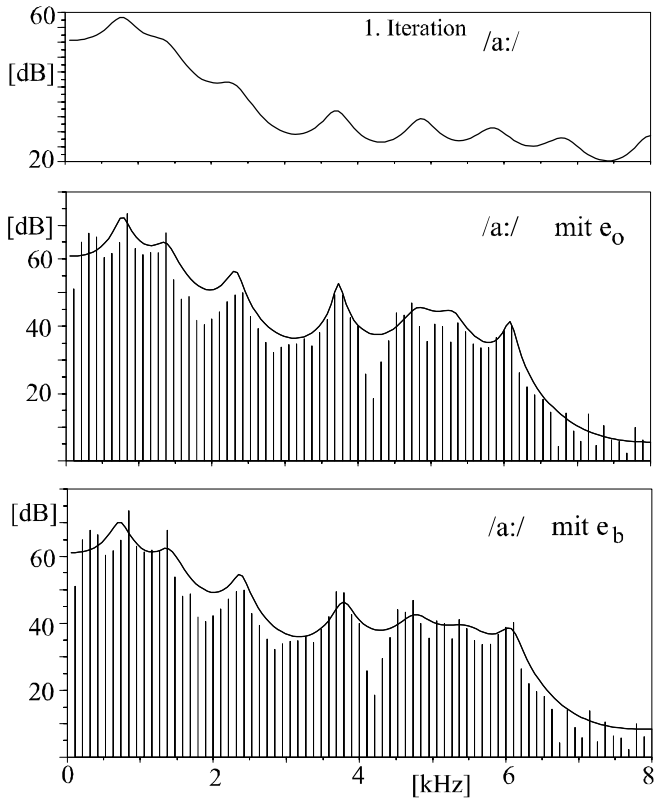


Bild 4.41: Analyse des Vokals /a:/ durch ein Rohrmodell mit festem Glottisabschluß r_g: Minimierung des Fehlers e_o nach der ersten Iteration (oben) und nach mehreren Iterationen (mittleres Bild), Minimierung des Fehlers e_b nach mehreren Iterationen (unten). Geschätzter Betragsgang (durchgezogene Linie), DFT der analysierten gemittelten Periode (Linienspektrum).

das erste Minimum eine längere Vokaltraktlänge dar als bei den Lauten /i:/ und /a:/, was durch die Lippenvorstülpung und Glottisabsenkung des Vokals /o:/ erklärt werden kann. Die Länge für /o:/ ist allerdings etwas zu groß. Der Fehler wird bei einer deutlich höheren Anzahl von verwendeten Rohrelementen absinken, da mehr Koeffizienten zur Verfügung stehen. Daß bei den Analysen der Sprachlaute sich Minima in der Nähe der phonetisch vorhergesagten Vokaltraktlängen bilden können, unterstützt die vorgenommene Modellbildung mit den beiden Rohrabschlüssen. Es ist zu bemerken, daß es zum Beispiel für den analysierten Laut /i:/ mit Verwendung von 17 oder 18 Rohrelementen eine schlechtere Modellierung erzielt wird als mit 16, obwohl mehr Reflexionskoeffizienten für die Modellierung des Sprachsignals zur Verfügung stehen. Durch die Analyseergebnisse unterschiedlicher Vokaltraktlängen kann allerdings nur eine Tendenz für die tatsächliche Länge erkannt werden. Hierfür könnte es vorteilhaft sein nur den Sprachsignalabschnitt mit geschlossener Glottisphase zu analysieren. Dafür müßte dieser Abschnitt der Sprachperiode allerdings mit Schätzverfahren erst ermittelt werden, was wieder fehlerbehaftet ist.

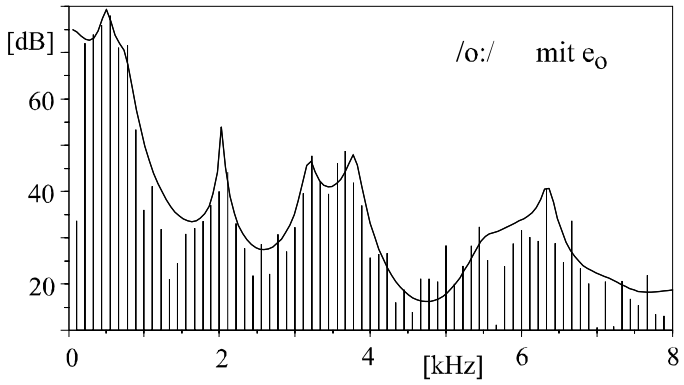


Bild 4.42: Analyse des Vokals /o:/ durch ein Rohrmodell mit festem Glottisabschluß r_g: Minimierung des Fehlers e_o nach mehreren Iterationen.

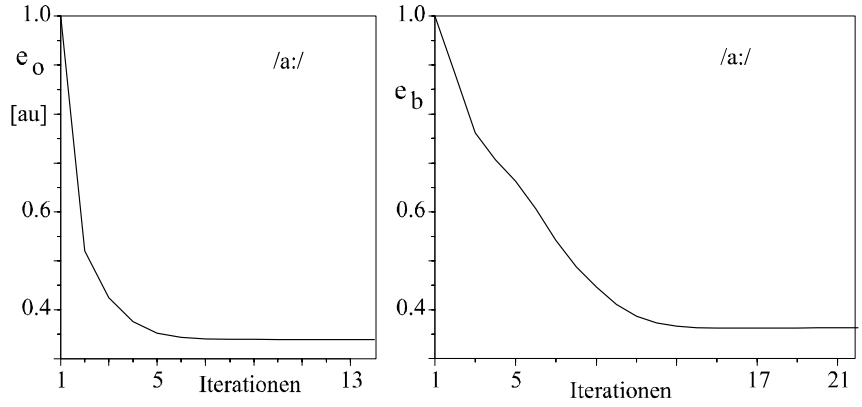


Bild 4.43: Entwicklung der beiden Fehlerdefinitionen in Abhängigkeit der Iterationsanzahl für den Vokal /a:/: Minimierung von e_o (links) und e_b (rechts).

Frequenzabhäniger Glottisabschluß

Bei geöffneter Glottis bewirkt die Glottisimpedanz durch den Einfluß des subglottalen Trakts einen frequenzabhängigen Abschluß an der Glottis [Fl72]. Neben dem Einfluß der Glottis können auch noch die Verluste des Vokaltraktes mit dem Glottisabschluß berücksichtigt werden. Der Glottisabschluß $G(z)$ wird frequenzabhängig als kleines Filter realisiert. Für die Durchführung des Analysealgorithmus ist es von Vorteil, wenn dieser Abschluß durch ein nicht rekursives System dargestellt wird, da somit die Elemente der Matrix $\boldsymbol{F}_i$ nur Zählerpolynome aufweisen. Damit ist eine schnelle Filterung mit den finiten Impulsantworten $f_i^{\lambda\beta}(n)$ möglich. Bei einem rekursiven System müssten die Signalperioden mehrmals durchgefiltert werden, um Einschwingvorgänge zu berücksichtigen. Der Betragsgang des Rohrabschlusses an der Glottis läßt sich schwer einschätzen, da die Zeitvariabilität der Glottisimpedanz und frequenzabhängige Rohrverluste im Vokaltrakt berücksichtigt werden sollten. Deshalb wird hier mit einfachen Tief- und Hochpässen für den Glottisabschluß $G(z)$ experimentiert. Die Verwendung von zwei vorgegebenen Rohrabschlüssen bewirkt eine periodische Resonanzstruktur im Übertragungsverhalten eines uniformen Rohrmodells. Diese Resonanzfrequenzen des

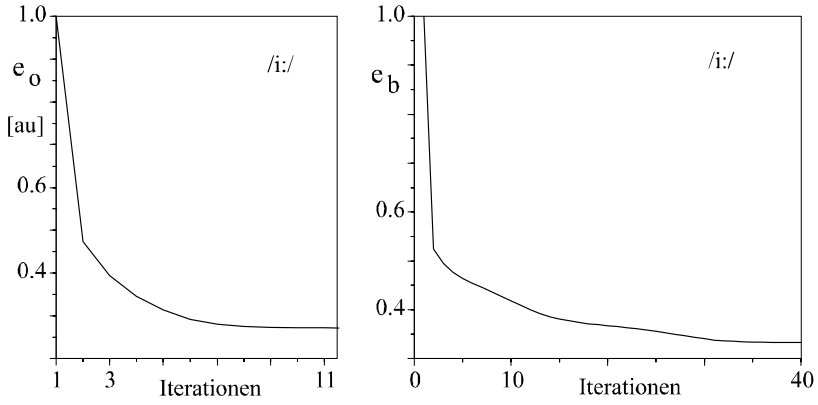


Bild 4.44: Entwicklung der beiden Fehlerdefinitionen in Abhängigkeit der Iterationsanzahl für den Vokal /i:/: Minimierung von e_o (links) und e_b (rechts).

uniformen Rohres werden durch unterschiedliche Vokaltraktformen verschoben, wobei zusätzlich die Bandbreiten verändert werden. Im höheren Frequenzbereich des stimmhaften Sprachspektrums sind allerdings kaum nennenswert ausgeprägte Resonanzen zu beobachten. Dies bedeutet, daß die Bandbreiten im höheren Spektralbereich größer sind. Dies kann durch einen betragsmäßig kleinen Glottisabschluß für den entsprechenden Frequenzbereich erreicht werden. Daher wurde mit mehreren Tiefpässen experimentiert, obwohl in [Fl72] nach Flanagan die Glottisimpedanz mit steigender Frequenz betragsmäßig zunimmt. Es können allerdings noch die Effekte der Wandvibrationen des Vokaltrakt berücksichtigt werden, die eine Tiefpaß-Charakteristik aufweisen. Um diese Resonanzstruktur zu berücksichtigen wurde ein Tiefpaß mit

$$G(z) = -0,37 \cdot (1 + 0,135 \cdot z^{-1}) \tag{4.149}$$

für den Rohrabschluß an der Glottis angenommen.

Analyse von Testsignalen

Um den Schätzalgorithmus für Rohrmodelle mit zwei frequenzabhängigen Rohrabschlüssen zu beurteilen, werden Testsignale analysiert. Für dieses Analysesystem werden, wie für das Synthesesystem, dieselben frequenzabhängigen Rohrabschlüsse verwendet mit $L(z)$ an den Lippen und dem Tiefpaß $G(z)$ (4.149) an der Glottis, sowie die gleiche Anzahl von Rohrelementen. Zu Begin der Schätzung besitzen die Reflexionskoeffizienten den Wert Null. Bild 4.46 zeigt die Resultate einer Analyse eines Testsignals nach dem Kriterium (4.133). Darin ist zu sehen, daß der iterative Algorithmus nach mehreren Iterationen sehr nahe an die optimale Lösung gelangt, was sich auch in weiteren Analysen von Testsignalen bestätigt. Da die gesamte spektrale Information in die Schätzung eingeht, entstehen auch nicht die Mehrdeutigkeiten wie sie in [So79] beschrieben werden.

Analyse von Sprachsignalen

Die Analysen der folgenden Sprachlaute wurden jeweils mit dem Fehlerkriterium (4.133) durchgeführt. Die Abtastrate der analysierten Sprachsignale beträgt 22 kHz. Die

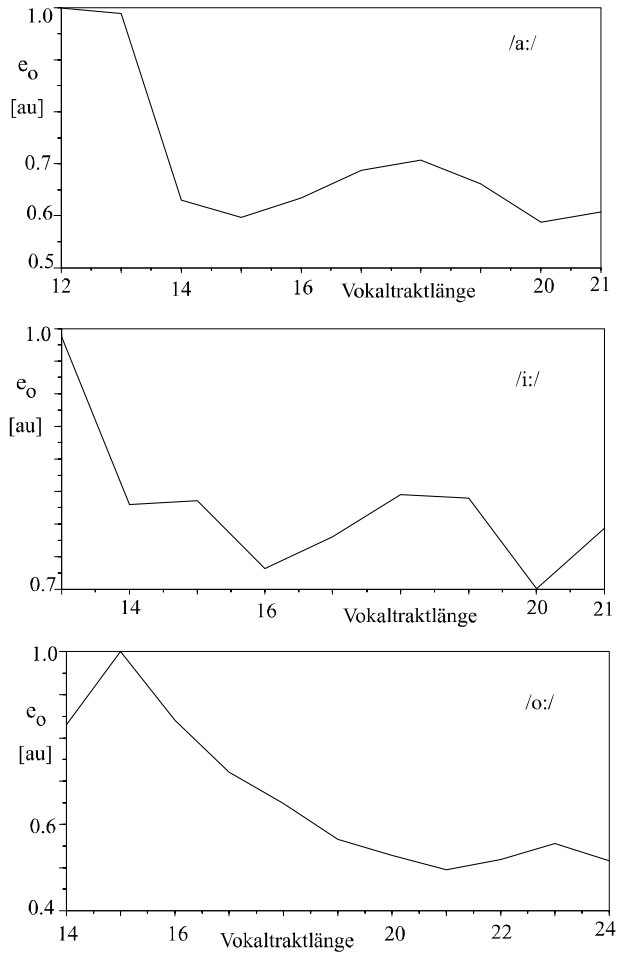


Bild 4.45: Fehler e_o in Abhängigkeit der Analyseergebnisse für unterschiedliche Vokaltraktlängen.

Sprachsignale werden mit einer adaptiven Präemphase vorgefiltert. Bild 4.47 zeigt das Resultat der Schätzung des Vokals /a:/ für ein Rohrmodell mit dem Glottisabschluß $G(z)$ (4.149) und Lippenabschluß $L(z)$, welcher an die analysierten Sprachlaute angepaßt ist. Wie dieses Beispiel aufzeigt, können die Spektren von Sprachlauten gut approximiert werden unter der Nebenbedingung des vorgegebenen Glottisabschlusses.Im Folgenden werden stimmhafte Konsonanten analysiert, welche auch für die Erzeugung von VCV Übergängen verwendet werden [SnL01c]. Bei den Konsonanten sind insbesondere ihre Verengungsstellen im Vokaltrakt interessant. Im Bild 4.48 sind die geschätzten Vokaltraktflächen im logarithmischen Maßstab der stimmhaften Frikative /z/ und /Z/ zu sehen. Diese beiden Laute sind die stimmhaften Varianten von /s/ und /S/. Die Konstriktionen sind darin gut zu erkennen.Wie phonetisch anzunehmen ist, befindet sich die Konstriktion des Lautes /Z/ von den Lippen aus betrachtet hinter der des Lautes /z/. Der Laut /z/ wurde auch in unterschiedlichen Vokalumgebungen analysiert. Dabei ergaben sich Konstriktionen an der gleichen oder unmittelbar benachbarten Stellen. Im Bild 4.49 sind geschätzte logarithmierte Vokaltraktflächen der stimmhaften Explosive /b/, /d/ und /g/ gezeigt. Die Stellen der Konstriktionen sind an verschiedenen Positionen zu beobachten entsprechend der bilabialen Position

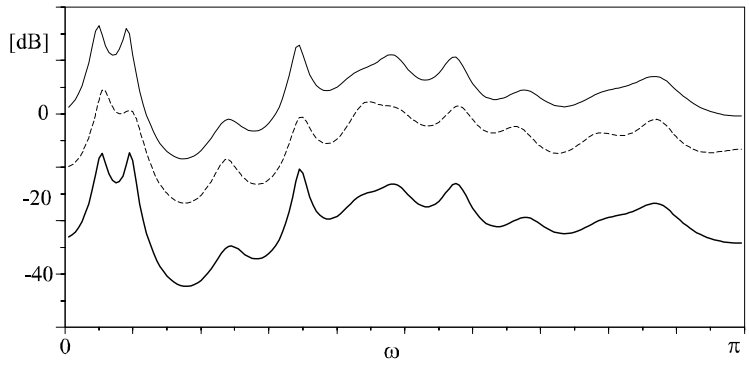


Bild 4.46: Iterative Analyse eines Testsignals durch ein Rohrmodell mit frequenzabhängigem Lippenabschluß und Glottisabschluß $G(z)$: Geschätzter Betragsgang nach mehreren Iterationen (obere durchgezogene Linie), geschätzter Betragsgang nach der ersten Iteration (gestrichelte Linie), DFT als Polygonzug der analysierten Testperiode (untere durchgezogene Linie).

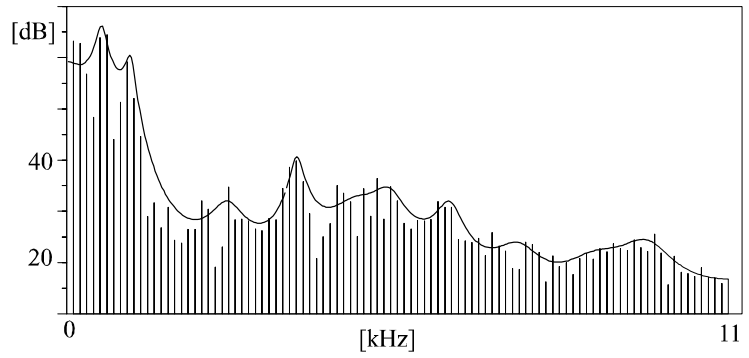


Bild 4.47: Analyse des Vokals /a:/ durch ein Rohrmodell mit frequenzabhängigem Lippenabschluß und Glottisabschluß $G(z)$: Betragsgang mit Berücksichtigung der Präemphase nach mehreren Iterationen (durchgezogene Linie), DFT der analysierten Sprachperiode (Linienspektrum).

für /b/, der alveolaren Position für /d/ und der velaren Position für /g/.Die Analyse von stimmhaften Explosiven ist insofern problematisch, da diese Laute instationär sind und teilweise durch den Übergang zum nächsten Laut charakterisiert sind. Hier werden für die Analyse von Explosiven im Gegensatz zu den Analysen der vorangegangenen Abschnitte Sprachsignale von natürlich artikulierten Explosiven in der Umgebung des Schwa-Lautes verwendet. Die Lautketten wurden allerdings mit einer eher langsamen Sprechgeschwindigkeit artikuliert. Für die Analyse werden die ersten Perioden nach dem Stimmeinsatz des Explosivs benutzt, welche im Spektralbereich zu einer Periode gemittelt werden. Um realistische Flächenverhältnisse aus der Analyse zu erhalten, kann es von Vorteil sein, die Präemphasekoeffizienten des nachfolgenden Vokals für die Vorfilterung des stimmhaften Explosivs zu verwenden. Dies kann dadurch erklärt werden, da sich der analysierte Sprachsignalabschnitt des Explosivs in einem instationären Prozeß befindet, der durch Glottiseinschwingvorgänge und der Vokaltraktbewegung charakterisiert ist. Die unterschiedliche Vorfilterung beeinflußt die Stärke der

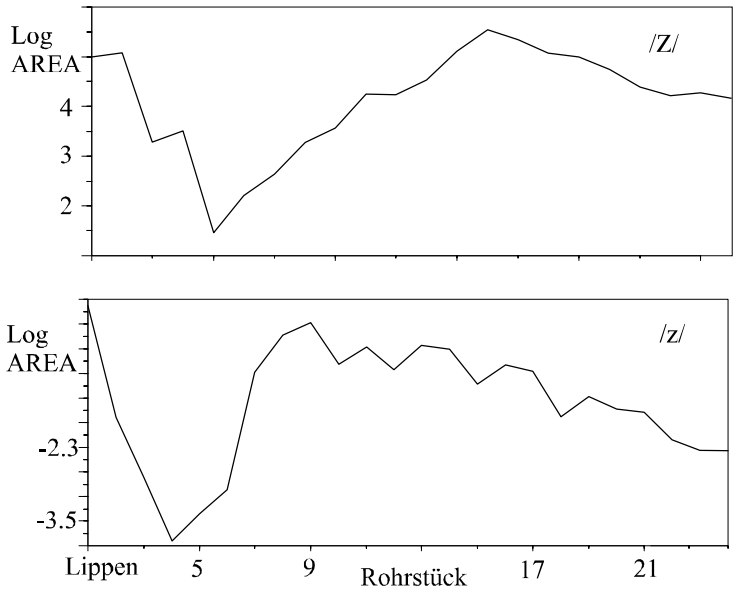


Bild 4.48: Geschätzte logarithmierte Vokaltraktflächen der stimmlosen Frikative /Z/ (oben) und /z/ (unten) für ein Rohrmodell mit frequenzabhängigem Lippenabschluß und Glottisabschluß $G(z)$.

ersten Reflexionskoeffizienten, so daß in den absoluten Flächen die Konstriktion nicht immer gut zu erkennen ist. Deshalb werden hier logarithmierte Flächen verwendet, da die Verhältnisse der Vokaltraktflächen durch ein logarithmisches Maß besser wiedergegeben werden. Insgesamt zeigen die Resultate deutlich, daß die Konstriktionen sich an Vokaltraktstellen befinden, die mit den phonetisch angenommenen Stellen korrespondieren. Für die Positionen des Explosivs /d/ und insbesondere /b/ kann davon ausgegangen werden, daß sie kaum von der vokalischen Umgebung abhängen. Beim Explosiv /g/ ist die Position stärker von dem nachfolgenden Vokal abhängig. Dies kann auch experimentell durch Sensoren am oberen Rand des Mundraumes (Palatum) belegt werden [En01a].

4.4.3 Verzweigtes Rohrsystem

Für die Analyse von Sprachlauten bei deren Produktion der Nasaltrakt beteiligt ist kommen auch Rohrmodelle mit einer Verzweigung zum Einsatz. Für den verzweigten Fall besitzen die Rohrmodelle nun einen vorgegeben reellen Rohrabschluß am Systemausgang. Dieser Abschluß ist hier reell gewählt, so daß die Nullstellen des Systems nur durch den Seitenzweig verursacht werden. Die Analyse des verzweigten Rohrmodells erweist sich durch die auftretenden Nullstellen schwieriger zu behandeln als das unverzweigte Rohrsystem. Im Gegensatz zum allgemeinen ARMA-Modell sind hier die möglichen Pol-Nullstellen Konfigurationen eingeschränkt, da unter anderem für das verzweigte Rohrsystem mehr Pole und Nullstellen existieren als Modellparameter vorhanden sind. Die Parameter des Seitenzweiges bestimmen nicht nur die Nullstellen des Systems, sondern liefern zusätzlich einen Beitrag zu den Polstellen, wie mittels (3.76) zu erkennen ist. Die Modellparameter können in die Pol- und Nullstellen des Systems umgerechnet werden, was umgekehrt nicht gewährleistet werden kann. Die

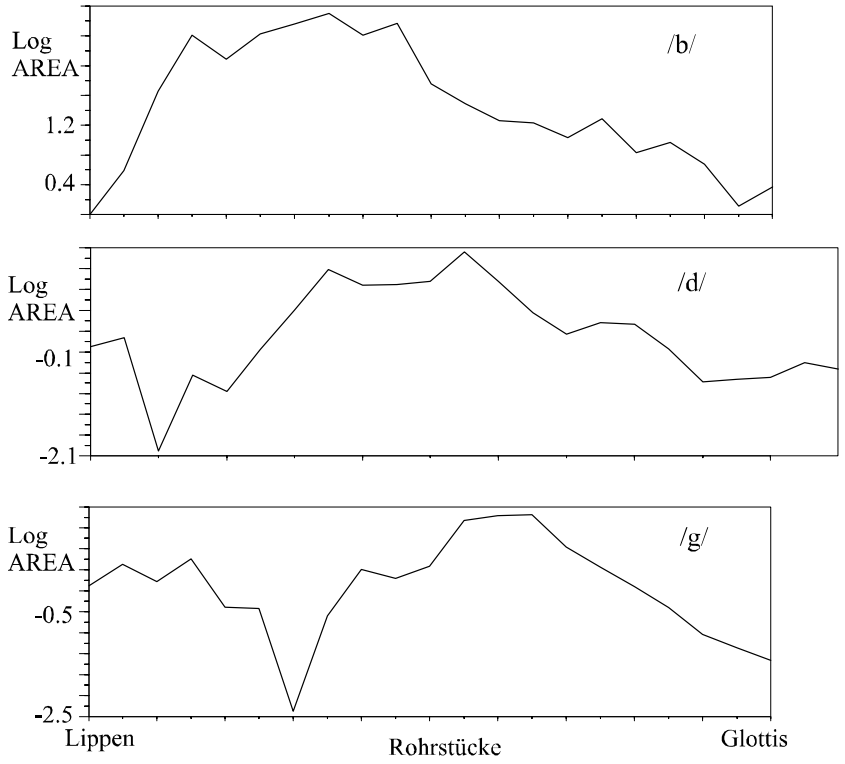


Bild 4.49: Geschätzte logarithmierte Vokaltraktflächen der Explosive /b/ (oben), /d/ (mittig) und /g/ (unten) für ein Rohrmodell mit frequenzabhängigem Lippenabschluß und Glottisabschluß $G(z)$.

Nullstellen, alleine betrachtet, können allerdings in die Parameter des Seitenzweiges transformiert werden. Dies ist unter anderem deshalb möglich, da ebenso viele Nullstellen wie Seitenzweigparameter vorhanden sind. Da durch den reellen Rohrabschluß sämtliche Nullstellen dem Seitenzweig zugerechnet werden können, kann der Schätzalgorithmus in zwei Schritte aufgeteilt werden. Zuerst werden die Nullstellen durch eine allgemeine ARMA Schätzung ermittelt und in die Reflexionskoeffizienten des Seitenzweiges umgerechnet. Anschließend werden die restlichen Reflexionskoeffizienten durch inverse Filterung iterativ [SnL02a, SnL02b] mit der Formel für $\widehat{r}_i$ geschätzt; zum Vergleich wird auch nach Burg geschätzt [SnL00d]. Die beiden Parameter des Dreitors werden analog zu den Reflexionskoeffizienten ermittelt. Neben den hier vorgestellten Algorithmen existieren schon Ansätze der Parameterschätzung von verzweigten Rohrmodellen, welche allerdings für die Sprachsignalanalyse nur bedingt verwendbar sind. Eine Aufspaltung der Schätzung in Pole und Nullstellen wurde auch in [Lim96] diskutiert, wobei die Parameter mittels einer Minimierung eines euklidischen Abstandsmaßes zwischen Polynomkoeffizienten bestimmt wurden. In [LiuL96a, LiuL96b, Liu98] wurde eine inverse Filterung des verzeigten Rohrmodells mittels der Burg-Methode vorgenommen, bei der direkt hinter dem zu schätzenden Zweitor die Ausgangsleistung minimiert wird. Im Schätzalgorithmus von [LiuL96a, LiuL96b] werden nur zwei Reflexionskoeffizienten des Seitenzweiges geschätzt und die Parameter des Dreitors sind zusätzlich eingeschränkt. Der Seitenzweig wird bei dieser Arbeit durch eine Art allgemeines Optimierungsverfahren mitgeschätzt. Während die Analyseergebnisse der Testsignale

in [LiuL96a, LiuL96b] sehr gut ausfallen, ergeben die Analysen von Sprachsignalen nicht so gute Resultate. Dies hängt einmal damit zusammen, daß nur die wenigsten Reflexionskoeffizienten des Seitenzweiges geschätzt werden und andererseits die Burg-Methode, welche eine Minimierung der Ausgangsleistung des zu schätzenden Zweitores beinhaltet, bei dieser Art von Rohrmodellen keine optimalen Resultate liefert. In [Lim96] werden keine Ergebnisse von Analysen von Sprachsignalen gezeigt, da für Sprachanalysen das Schätzverfahren als problematisch erachtet wurde. Neben den Simulationen mit Systemen wurde in [Da98] unter Verwendung eines realen Objekts der Nasaltrakt bzw. Mundraum als Seitenzweig untersucht. Für akustische Messungen wurde in [Da98] ein entsprechend geformtes Rohr mit einem Seitenzweig nachgebaut, welches den verzweigten Sprechtrakt darstellen soll.

Bestimmung der Nullstellen

Die Nullstellen werden mit einem allgemeinen Pol-Nullstellen System durch die iterative Verwendung von suboptimalen Lösungen geschätzt. Dazu wird das Verfahren aus Bild 4.33 mit zwei parallelen Blöcken benutzt. Die Ordnung des Zählerpolynoms korrespondiert mit der Anzahl der Reflexionskoeffizienten des Seitenzweiges. Die Ordnung des Nennerpolynoms muß hingegen nicht exakt mit der Ordnung des Nenners der Übertragungsfunktion des verzweigten Rohrmodells übereinstimmen, da die geschätzten Polstellen nicht weiter verwendet werden. Die geschätzten Pole sollen den Einfluß der Resonanzen in dem zu analysierenden Signal berücksichtigen. Bei ausschließlicher Verwendung von Nullstellen für die Schätzung würden die Nullstellen auch den Einfluß der Polstellen approximieren. Der Zähler des verzweigten Rohrmodells kann mit (3.76) dargestellt werden als:

$$B = \sum_{i=0} b_i \cdot z^{-1} = \boldsymbol{\rho}_{\,2} B' = \boldsymbol{\rho}_{\,2} \left(Q(z) + P(z)\right) \tag{4.150}$$

$$\text{mit} \qquad B' = 1 + \sum_{i=1} b'_i \cdot z^{-1}. \tag{4.151}$$

Q und P stellen das Zähler- und Nennerpolynom der Teilübertragungsfunktion des Seitenzweiges dar, bei der sich der Ein- und Ausgang am Dreitor befindet, entsprechend Bild 3.6. Das Polynom $Q(z) + P(z)$ ist von den Reflexionskoeffizienten des Seitenzweiges abhängig. Der Faktor $\boldsymbol{\rho}_2$ geht in den Polynomkoeffizienten b_0 ein, so daß er durch den Korrekturfaktor (4.67) in B für die Schätzung unterdrückt werden muß. Die Koeffizienten b'_i sind durch den ARMA-Ansatz bestimmt und können durch eine rekursiv definierte Funktion in die Reflexionskoeffizienten des Seitenzweiges umgewandelt werden, womit Q und P bestimmt sind.

Inverse Filterung

Die Parameter des Seitenzweiges sind durch die Ergebnisse der allgemeinen ARMA-Schätzung bestimmt, während die restlichen Parameter nun durch inverse Filterung geschätzt werden, wofür das inverse Filter in einen rein rekursiven und einen nicht-rekursiven Anteil zerlegt wird. Die Betriebskettenmatrizen $\boldsymbol{T}_i$ besitzen als Elemente nur Zählerpolynome, wodurch sie nur zum nicht-rekursiven Anteil beitragen. Die Matrix $\boldsymbol{T}_D$, welche den Seitenzweig und den Dreitor mit einem Rohrstück darstellt, kann

faktorisiert werden zu

$$\boldsymbol{T}_D = T_D{}^r \, \boldsymbol{T}_D^n. \tag{4.152}$$

Der Vorfaktor $T_D{}^r$ beschreibt den rein rekursiven Anteil mit

$$T_D{}^r = \frac{1}{B} = \frac{1}{\boldsymbol{\rho}_2(Q+P)}, \tag{4.153}$$

während die Matrix $\boldsymbol{T}_D^n$ den nicht-rekursiven Teil darstellt mit

$$\boldsymbol{T}_D^n = \begin{pmatrix} \{\boldsymbol{\rho}_1 + \boldsymbol{\rho}_2 - 1, 1\} & \{\boldsymbol{\rho}_2 - 1, 1 - \boldsymbol{\rho}_1\} \cdot zv^{-1} \\ \{1 - \boldsymbol{\rho}_1, \boldsymbol{\rho}_2 - 1\} & \{1, \boldsymbol{\rho}_1 + \boldsymbol{\rho}_2 - 1\} \cdot zv^{-1} \end{pmatrix} \tag{4.154}$$

und der Abkürzung $\{x, y\} := x \cdot Q + y \cdot P$.

Da hier Periodizität der Analysesignale angenommen wird, sind die Laufzeitglieder durch die zyklische Verschiebung realisiert. $T_D{}^r$ beschreibt nicht nur den rein rekursiven Anteil von $\boldsymbol{T}_D$ sondern auch gleichzeitig den des gesamten inversen Filter, wodurch der rein rekursive Teil zuerst behandelt werden kann. Dies wird durch Filterung im Zeitbereich oder durch eine Multiplikation im Frequenzbereich mit anschließender inverser DFT vollzogen:

$$x' = \mathrm{IDFT}\left(\frac{X}{B'}\right) = \mathrm{IDFT}\left(\frac{X}{Q+P}\right). \tag{4.155}$$

Stellt das Eingangssignal x des inversen Filters eine Periode dar, so wird im Zeitbereich die Periode x mehrmals durch das rekursive System $T_D{}^r$ gefiltert, um Einschwingvorgänge zu vermeiden. Mit der Durchführung der Operation im Frequenzbereich finden keine Einschwingprobleme statt. Für die Analyseergebnisse sind praktisch keine Unterschiede zu erkennen, ob die Operation im Zeitbereich oder im Frequenzbereich durchgeführt wurde. Durch die Verwendung von zyklischen Verschiebungen ist für die inverse Filterung die Phase des Analysesignals x' nicht relevant, wie zuvor im Falle des unverzweigten Rohres. Im Signal x' ist durch die Vorverarbeitung (4.155) der Einfluß der geschätzten Nullstellen des verzweigten Rohres beseitigt, welche die Pole des inversen Filters darstellen. Die restlichen Koeffizienten werden aus x' durch eine Minimierung der Ausgangsleistung des nicht-rekursiven Teils des inversen Filters geschätzt, welcher in Bild 4.50 dargestellt ist. Für das nicht-rekursive System werden die Verknüpfungsmatrizen

$$\boldsymbol{x}_i = \boldsymbol{T}_i \cdot \boldsymbol{x}_{i-1} \quad \text{und} \quad \boldsymbol{x}_{M+1} = \boldsymbol{T}_D^n \cdot \boldsymbol{x}_M \tag{4.156}$$

$$\tag{4.157}$$

$$\text{mit} \quad \boldsymbol{x}_i = \begin{pmatrix} x_i^o \\ x_i^u \end{pmatrix} \tag{4.158}$$

verwendet. Der obere und untere Pfad am Eingang des inversen Filters wird mit Hilfe des Abschlusses R initialisiert:

$$x_0^o = x' \,, \quad x_0^u = R \cdot x'. \tag{4.159}$$

R stellt im Synthesefilter den reellen Rohrabschluß am Systemausgang dar. Für die inverse Filterung des nicht rekursiven Systems nach Bild 4.50 werden zwei Ansätze diskutiert, die sich bezüglich der Orte im inversen Rohrmodell unterscheiden, an denen die Ausgangsleistungen minimiert werden.

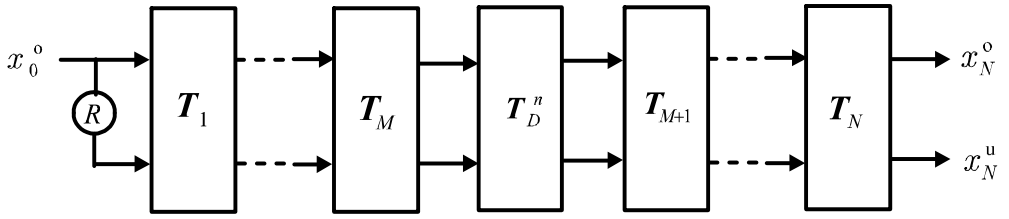


Bild 4.50: Nicht-rekursiver Anteil des inversen Filters des verzweigten Rohrmodells.

Inverse Filterung mittels Burg-Kriterium

Als erster Ansatz werden die Reflexionskoeffizienten einfach mit dem Burg-Koeffizienten geschätzt. Die Parameter des Dreitores können dazu nach demselben Kriterium nach Burg ermittelt werden mit den Bestimmungsgleichungen

$$\frac{\partial \mathrm{E}\left[\left(x_{M+1}^{o}\right)^2\right] + \mathrm{E}\left[\left(x_{M+1}^{u}\right)^2\right]}{\partial \boldsymbol{\rho}_i} = 0 \quad \text{für } i = 1, 2 \ , \tag{4.160}$$

wie in [SnL00d] beschrieben. Die Reflexionskoeffizienten werden mit dem ersten beginnend genau einmal berechnet. In Bild 4.51 ist die Analyse des Nasal /n/ mit einer Abtastrate von 11 kHz gezeigt. Die geschätzte Nullstelle ist gut zu erkennen. Es zeigt

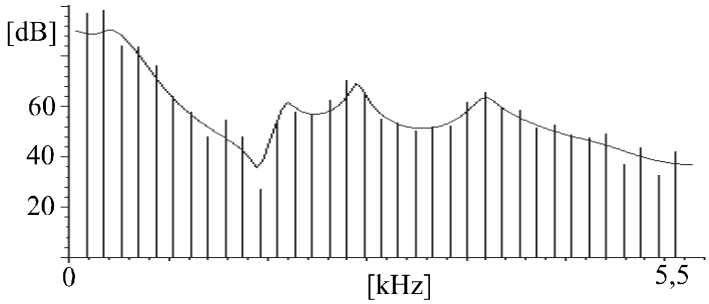


Bild 4.51: Analyse des Nasals /n/ mittels dem Burg-Kriterium für die inverse Filterung des verzweigten Rohrmodells.

sich allerdings, daß gerade bei stärker ausgeprägten Nullstellen, der Algorithmus nicht so gut in der Lage ist, die Einschränkungen des Lösungsraumes der Polstellen durch den vorgegebenen Seitenzweig zu berücksichtigen. In Bild 4.52 sind die geschätzten Nasaltraktflächen eines verzweigten Rohrmodells von Sprachsignalen der Nasale /n/ und /m/, welche eine Abtastrate von 16 kHz aufweisen. Wie zu sehen ist, sind die geschätzten Nasaltraktflächen der beiden Nasale ähnlich. Da die Schätzung der Koeffizienten nach Burg keine Parameter hinter dem zu schätzenden Zweitor berücksichtigen, führt eine wiederholte Berechnung der Modellparameter zu keinen veränderten Resultaten. Die Burg-Methode erzielt für verzweigte Rohre in dieser Weise keine optimalen Ergebnisse. Um eine Verbesserung zu erzielen, wird deshalb im Gegensatz zur Burg-Methode eine Leistungsminimierung am Ausgang des inversen Filter vollzogen, die iterativ durchgeführt wird.

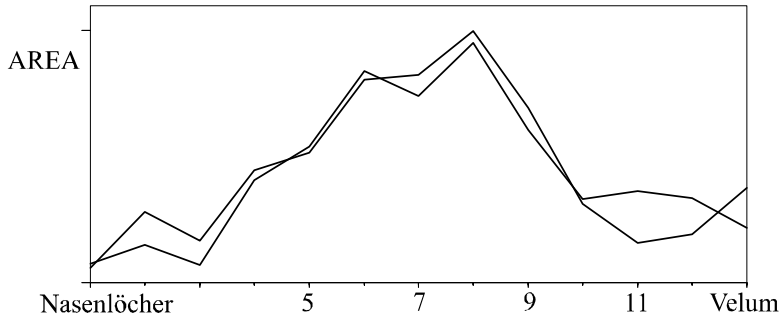


Bild 4.52: Geschätzte Nasaltraktflächen eines verzweigten Rohrmodells aus Sprachsignalen mit 16 kHz Abtastrate der Nasale /n/ und /m/.

Iterative inverse Filterung

Im zweiten Ansatz werden die Modellparameter durch Leistungsminimierung des Filterausgangs x_N^o nach (4.133) bestimmt. Das Kriterium (4.141) der optimalen Reflexionskoeffizienten wird auch durch:

$$\frac{\partial \mathrm{E}\left[\left(x_N^o\right)^2\right]}{\partial \boldsymbol{\rho}_i} = 0 \qquad \Rightarrow \widehat{\boldsymbol{\rho}}_i \tag{4.161}$$

für die beiden Koeffizienten $\boldsymbol{\rho}_i$ des Dreitores verwendet. Der optimale Koeffizient $\widehat{\boldsymbol{\rho}}_i$ kann als geschlossene Formel hergeleitet werden, wenn die Eingangssignale x_M^o und x_M^u von $\boldsymbol{T}_D^n$ gegeben sind. Nach Auflösen von (4.161) nach $\boldsymbol{\rho}_i$ ergeben sich verhältnismäßig viele Terme. Durch eine geeignete Darstellung des Signals x_N^o kann die Anzahl der auftretenden Terme reduziert werden. Dies kann mit einer Entwicklung von x_N^o nach den Koeffizienten $\boldsymbol{\rho}_i$ mit

$$x_N^o = \boldsymbol{\chi} + \boldsymbol{\rho}_1 \cdot \boldsymbol{\chi}_1 + \boldsymbol{\rho}_2 \cdot \boldsymbol{\chi}_2 \tag{4.162}$$

erreicht werden. Dadurch ergeben sich nach (4.161) die optimalen Koeffizienten zu

$$\widehat{\boldsymbol{\rho}}_1 = -\frac{\overline{\boldsymbol{\chi} \cdot \boldsymbol{\chi}_1} + \boldsymbol{\rho}_2 \cdot \overline{\boldsymbol{\chi}_2 \cdot \boldsymbol{\chi}_1}}{\overline{\boldsymbol{\chi}_1 \cdot \boldsymbol{\chi}_1}} \tag{4.163}$$

$$\widehat{\boldsymbol{\rho}}_2 = -\frac{\overline{\boldsymbol{\chi} \cdot \boldsymbol{\chi}_2} + \boldsymbol{\rho}_1 \cdot \overline{\boldsymbol{\chi}_1 \cdot \boldsymbol{\chi}_2}}{\overline{\boldsymbol{\chi}_2 \cdot \boldsymbol{\chi}_2}}$$

mit

$$\begin{aligned} \boldsymbol{\chi}_1 &= o_Q^{11} - u_P^{11} + u_Q^{12} - o_Q^{12} \\ \boldsymbol{\chi}_2 &= u_Q^{12} + o_Q^{11} + u_Q^{11} + o_P^{12} \\ \boldsymbol{\chi} &= -o_Q^{11} - u_Q^{11} - u_Q^{12} + u_P^{11} + o_P^{11} + u_P^{12} - o_P^{12} + o_Q^{12} \end{aligned} \tag{4.164}$$

und

$$\begin{aligned} o_Q^{\lambda\beta} &= q(n) * f_{M+1}^{\lambda\beta}(n) * x_M^o(n) \\ u_Q^{\lambda\beta} &= q(n) * f_{M+1}^{\lambda\beta}(n) * x_M^u(n-1) \\ o_P^{\lambda\beta} &= p(n) * f_{M+1}^{\lambda\beta}(n) * x_M^o(n) \\ u_P^{\lambda\beta} &= p(n) * f_{M+1}^{\lambda\beta}(n) * x_M^u(n-1). \end{aligned} \tag{4.165}$$

Die Zeitverschiebung von $x_M^u(n-1)$ ist durch den unteren Zustandsspeicher in $\boldsymbol{T}_D^n$ bedingt. Die Faltungen wie die Zeitverschiebungen werden mit der Annahme der Periodizität zyklisch ausgeführt. Die Impulsantworten $q(n)$ und $p(n)$ sind die Polynomkoeffizienten von Q und P:

$$\begin{aligned} Q &= \sum_n q(n) \cdot zv^{-n} \\ P &= \sum_n p(n) \cdot zv^{-n}, \end{aligned} \tag{4.166}$$

welche sich aus der Teilübertragungsfunktion $\widetilde{H}(z)$ des Seitenzweiges ergeben. Die Matrix $\boldsymbol{F}_i$ stellt nach (4.137) die Zweitore hinter dem zu schätzenden i'ten Zweitor dar und ergibt sich im Falle des verzweigten Rohrmodells zu

$$\boldsymbol{F}_i = \begin{cases} \prod\limits_{k=0}^{N-M-1} \boldsymbol{T}_{N-k} \cdot \boldsymbol{T}_D^n \cdot \prod\limits_{k=0}^{M-i-1} \boldsymbol{T}_{M-k} & \text{für } i = 1 \ldots M-1 \\ \prod\limits_{k=0}^{N-M-1} \boldsymbol{T}_{N-k} \cdot \boldsymbol{T}_D^n & \text{für } i = M \\ \prod\limits_{k=0}^{N-i-1} \boldsymbol{T}_{N-k} & \text{für } i = M+1 \ldots N-1. \end{cases} \tag{4.167}$$

$\boldsymbol{F}_N$ stellt für den Fall $i = N$ die Einheitsmatrix dar, da sich hinter dem N'ten Zweitor kein weiteres Tor und auch kein vorgegebener Glottisabschluß befindet. Für die Schätzungen der Reflexionskoeffizienten wird die Berechnungsformel (4.143) angewandt, wobei für die gefilterten Signale $o^{\lambda\beta}$ und $u^{\lambda\beta}$ von (4.145) und für die Elemente $f^{\lambda\beta}$ die Matrix (4.167) verwendet wird. Dies entspricht für den Algorithmus einer Zerlegung der Indexmenge in $I_i = \{i\}$ für alle Indizes i außer der Indizes, die die Parameter des Seitenzweiges darstellen und durch das ARMA-Verfahren bestimmt sind. Die Modellparameter werden nacheinander geschätzt entsprechend ihrer Reihenfolge im inversen Filter. Im Falle negativer Flächen werden die Modellparameter korrigiert. Für die Ermittlung der beiden Dreitorkoeffizienten muß die Berechnungsreihenfolge untereinander festgelegt werden. Die Anfangsrohrkonfiguration für den zweiten Teil des Schätzalgorithmus beinhaltet gleich große Flächen für die Parameter, abgesehen von den ermittelten Flächen des Seitenzweiges. Es kann sich als günstig erweisen, die Dreitorparameter erst nach den ersten Iterationen für eine Schätzung frei zu geben, so daß sie z.B. in den ersten vier Iterationen immer den Wert 2/3 aufweisen.

Analyse von Testsignalen

Das zu analysierende Testsignal stellt den Ausgang eines verzweigten Rohrmodells mit einem Rohrabschluß $R = -0,9$ dar, welches mit einer Impulsfolge anregt wurde. Die Resultate der Analyse einer Testsignalperiode sind in Bild 4.53 gezeigt. Für die Analyse sind die Anzahl der Rohrstücke und der Rohrabschluß vom Synthesefilter des Testsignals übernommen worden. Sämtliche Reflexionskoeffizienten und die beiden Dreitorparameter werden wie zuvor beschrieben dabei geschätzt. Es zeigt sich, daß der Algorithmus eine Lösung in der Nähe des globalen Minimums gefunden hat. Je besser die Nullstellen durch den allgemeinen ARMA-Ansatz geschätzt werden können, um so besser werden in der Regel die restlichen Koeffizienten geschätzt.

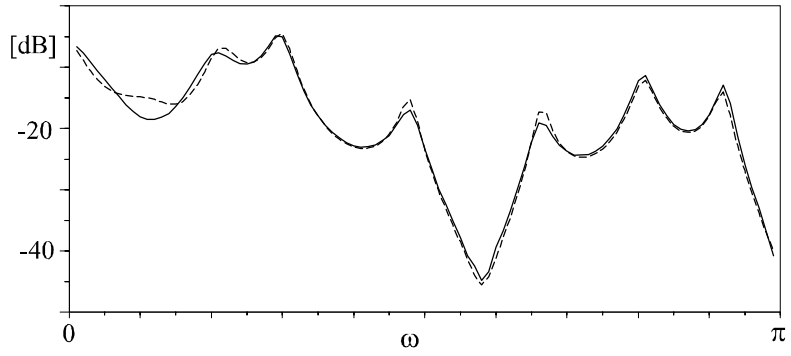


Bild 4.53: Analyse eines Testsignals mittels der iterativen inversen Filterung eines verzweigten Rohrmodells: Spektrum des Testsignals als Polygonzug (durchgezogene Linie), geschätzter Betragsgang (gestrichelte Linie).

Analyse von Sprachsignalen

Für die Parameterbestimmung der verzweigten Rohrmodelle ergibt sich die Schwierigkeit, daß die geschätzten Nullstellen des Seitenzweiges ungünstige Restriktionen für die Polstellen darstellen können. Im Falle von Analysen von Testsignalen beeinträchtigen die geschätzten Nullstellen den Lösungsraum der Pole hinsichtlich der optimalen Lösung kaum bzw. gar nicht, da die Struktur und der Systemgrad für die Analyse und Synthese der Testsignale identisch sind. Dies kann bei Sprachsignalen nicht gewährleistet werden, da Analysefilter und das System des realen Sprechtrakts sich hinsichtlich Struktur und Systemgrad unterscheiden. Die Schätzung der restlichen Modellparameter muß die Polstellen unter den beschriebenen Einschränkungen, bedingt durch die geschätzten Polynome Q und P, bestimmen. In [Lim96] ist ein Algorithmus vorgestellt, der die restlichen Parameter durch Minimierung eines euklidischen Abstandsmaßes der Polynomkoeffizienten schätzt. Für dieses Verfahren werden nur Resultate von analysierten Testsignalen vorgestellt, da für Sprachsignale der Algorithmus nach [Lim96] problematisch ist. Eine Erklärung dafür könnte die beschriebene Einschränkung der Pole sein sowie das verwendete Fehlermaß. Mit der hier vorgestellten iterativen inversen Filterung können hingegen die Pole bzw. die restlichen Parameter unter den Nebenbedingungen eines vorbestimmten Seitenzweiges sehr gut geschätzt werden, wie die folgenden Beispiele aufzeigen. Für die Analyse werden mehrere benachbarte Sprachperioden im Spektralbereich gemittelt. Danach werden die gemittelten Sprachperioden durch eine adaptive Präemphase vorgefiltert.

Analyse von Nasalen

Im Falle von Nasalen repräsentiert der Seitenzweig des Rohrmodells die Mundhöhle, die für den Nasal /m/ länger angenommen wird als für den Nasal /n/. In den Bildern 4.54 und 4.55 sind die Resultate der Analysen der Nasale /n/ und /m/ gezeigt [SnL02b]. Die Nasenlöcher werden durch einen Abschlußkoeffizienten von -0.9 modelliert. Die Abtastrate der analysierten Sprachsignale der Nasale beträgt 22 kHz, woraus die Längen des Rachens und Nasaltraktes sowie die der Mundhöhle folgen. Die Analyse wird, wie in den darauf folgenden Beispielen, mit der iterativen inversen Filterung durchgeführt. In Bild 4.54 sind für die Nasale /m/ und /n/ die geschätzten Betrags-

gänge der verzweigten Rohrmodelle im Vergleich zu den DFT-Spektren der gemittelten Sprachperioden mit Präemphase gezeigt. Zusätzlich ist der Beitrag der Nullstellen des Rohrmodells dargestellt, welcher durch den Seitenzweig verursacht wird. In Bild 4.55

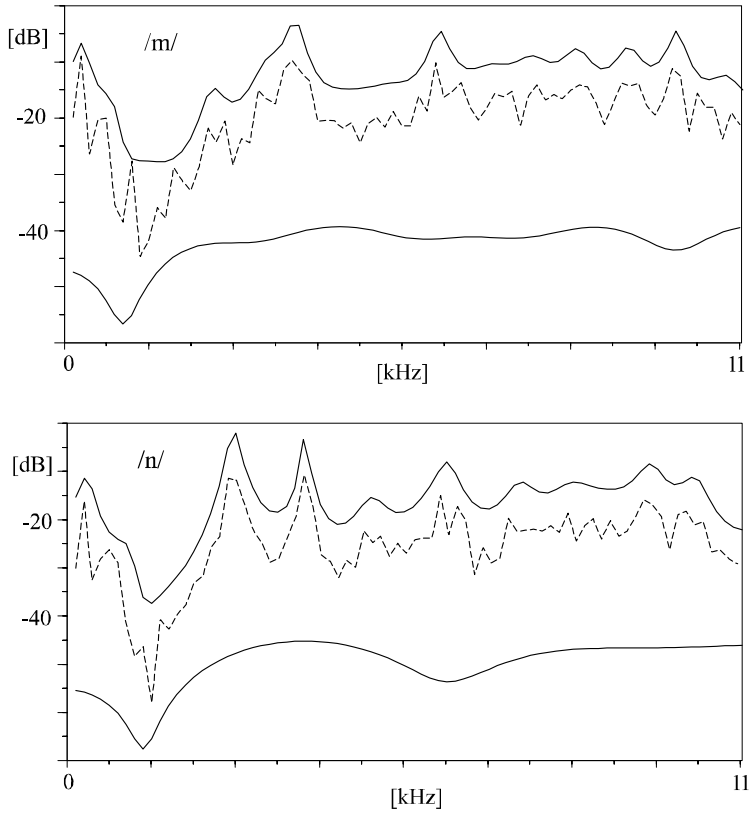


Bild 4.54: Analyse der Nasale /m/ (oben) und /n/ (unten) mittels der iterativen inversen Filterung eines verzweigten Rohrmodells: Geschätzter Betragsgang (obere durchgezogene Linie), DFT der analysierten Periode (mittlere gestrichelte Linie) und Beitrag der Nullstellen (untere durchgezogene Linie).

sind die geschätzten Flächen der Seitenzweige aus den beiden Nasalen der Analysen des Bildes 4.54 gezeigt. Entsprechend der unterschiedlichen Längen und Formen der Mundhöhlen ergibt sich eine dominante Nullstelle im Bereich um 1 kHz bzw. etwas darüber. Die Nullstelle liegt für den Nasal /m/ infolge der längeren Mundhöhle bei einer niedrigeren Frequenz als für den Nasal /n/. Die Form der geschätzten Flächen und die unterschiedlichen Positionen der Nullstellen lassen darauf schließen, daß diese Nullstellen des Seitenzweiges auch tatsächlich die Mundhöhle modellieren. Dies ist nicht unbedingt gewährleistet, da der Nasaltrakt selbst auch Nebenhöhlen besitzt und verzweigt ist, und damit auch Nullstellen zur Übertragungsfunktion beiträgt. Durch einen Vergleich mit dem Übertragungsverhalten von Nasalen aus [Da94], welcher aus anatomischen Daten ermittelt wurde, ist zu erkennen, daß die dominanten Nullstellen im Bild 4.54 den Einfluß der Mundhöhle beschreiben. Darüber hinaus kann durch einen Vergleich mit [Da94] gesehen werden, daß die Einbuchtungen im Betragsspektrum der Nasale im Bild 4.54 unterhalb der Frequenz der dominanten Nullstelle durch die Nebenhöhlen verursacht werden.

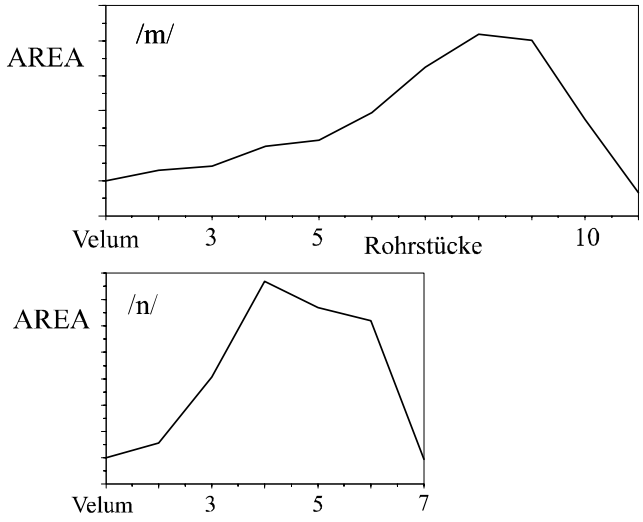


Bild 4.55: Geschätzte Mundhöhlen aus den Nasalen /m/ (oben) und /n/ (unten), welche den Seitenzweig des verzweigten Rohrmodells darstellen.

Analyse von nasalierten Vokalen

Nasalierte Vokale werden analog zu den Nasalen mit einem verzweigten Rohrmodell analysiert. Der Seitenzweig stellt im Gegensatz zu den Analysen der Nasale den Nasaltrakt dar. Das Nasensignal der nasalierten Vokale wird in diesem Modell nicht explizit berücksichtigt, da nur ein Systemausgang für die Modellierung des Sprachsignals verwendet wird. Die Abtastrate der analysierten Sprachaufnahmen der nasalierten Vokale beträgt 16 kHz, wonach sich die Anzahl der Rohrelemente richtet. In Bild 4.56 sind die geschätzten Betragsgänge des nasalierten Vokals /$\widetilde{\text{a}}$/ und des nasalierten Schwa-Lautes gezeigt und in Bild 4.57 die des nasalierten Vokals /$\widetilde{\text{i}}$/ [SnL02c]. Für den Vokal /$\widetilde{\text{a}}$/ ist eine stark ausgeprägte Nullstelle zu sehen, die sich in vielen Messungen in dieser Art ausgeprägt hat. In Bild 4.58 sind die Flächen des geschätzten Seitenzweiges zu sehen, welche den Einfluß des Nasaltraktes wiedergeben. In Bild 4.59 oben sind die geschätzten Flächen des Vokaltraktes von /$\widetilde{\text{a}}$/ und unten die des nasalierten Vokals /$\widetilde{\text{i}}$/ gezeigt, dessen zugehöriger Seitenzweig in Bild 4.58 oben darstellt ist. Die Vokaltraktflächen sind mit denen eines unnasalierten Vokals grob vergleichbar, obwohl die Höhle sich für /$\widetilde{\text{i}}$/ zu nahe an den Lippen befindet. Die Vokaltraktflächen lassen sich aus dem verzweigten Rohrmodell für die nasalierten Vokale weniger robust schätzen, als mit einem unverzweigten Rohrmodell für die Vokale. Es ist zu beachten, daß der Ausgang des Rohrsystems das Mund- und Nasensignal der nasalierten Laute modellieren muß. Für eine genauere Modellierung müssen Mund- und Nasensignal separat behandelt werden. Dies kann mit Hilfe von gewöhnlichen Sprachaufnahmen nicht ohne weiteres erreicht werden.

Analyse von Mund- und Nasensignalen

Unter Verwendung einer Trennwand, die den Kopf umschließt, lassen sich Mund- und Nasensignale schon bei der Aufnahme separieren. In einem vorangegangenen Abschnitt wurde ein verzweigtes Rohrmodell mit zwei Ausgängen verwendet um Mund- und Nasensignal gleichzeitig zu modellieren. Da ein Koeffizientensatz des Modells dadurch

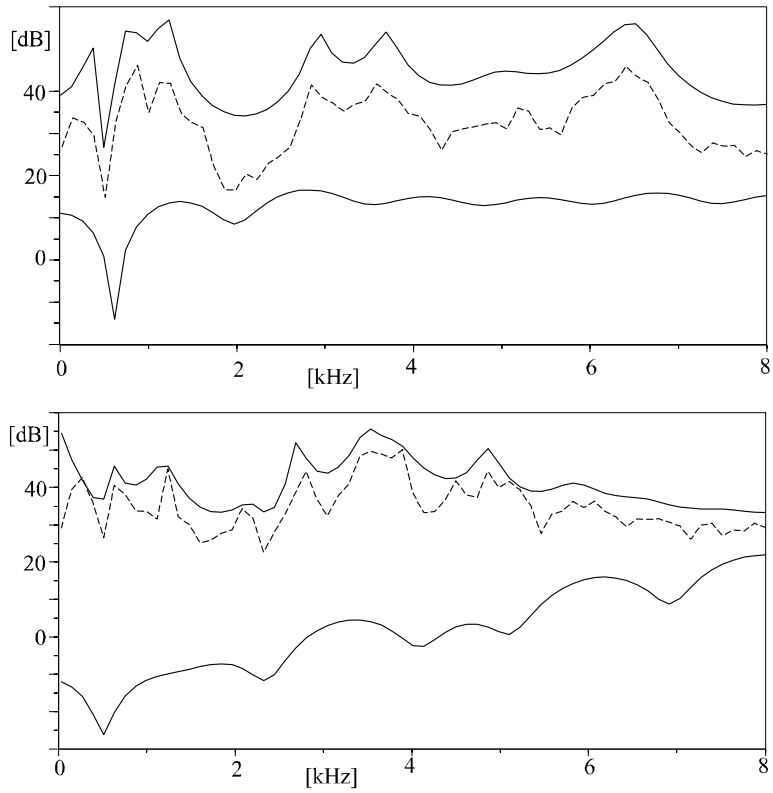


Bild 4.56: Analyse des nasalierten Vokals /ã/ (oben) und des nasalierten Schwa-Lautes (unten) mittels der iterativen inversen Filterung eines verzweigten Rohrmodells: Geschätzter Betragsgang (obere durchgezogene Linie), DFT der analysierten Periode (mittlere gestrichelte Linie) und Beitrag der Nullstellen (untere durchgezogene Linie).

zwei Signale berücksichtigen muß, können die jeweiligen Spektren des Mund- und Nasensignals in der Regel nicht so gut durch die Modellbetragsgänge approximiert werden wie im Falle einer Schätzung unter Berücksichtigung nur eines Systemausgangs. Dies resultiert unter anderem aus der komplexen Struktur des Nasaltraktes, womit das verwendete unverzweigte Rohrmodell eine starke Vereinfachung der tatsächlichen Nasenstruktur darstellt. Um eine bessere Modellierung der Signalspektren zu erreichen wird für jedes Signal ein eigenes Modell verwendet. Ein verzweigtes Rohrmodell wird für die Modellierung des Mundsignals verwendet und ein weiteres Rohrmodell für das Nasensignal [SnL02d, SnL03]. Die Seitenzweige der beiden Rohrmodelle stellen dabei jeweils den Nasaltrakt bzw. den Mundraum dar. Die beiden Modelle weisen für die Zweige hinter dem Dreitor-Adaptor sowie für den Seitenzweig eine unterschiedliche Anzahl von Rohrstücken auf, da Nasaltrakt und Mundraum in den beiden Modellen vertauscht sind. Die Modellparameter dieser beiden Rohrmodelle werden unabhängig voneinander aus dem Mund- und Nasensignal geschätzt. Die Abtastrate der Signale beträgt 16 kHz (ursprünglich 32 kHz). Im Bild 4.61 sind die Resultate der analysierten Mund- und Nasensignale des nasalierten Vokals /ã/ gezeigt und in Bild 4.60 die Resultate für den nasalierten Schwa-Laut /@̃/. Im Spektrum der Nasensignale ist eine Resonanz unter 500 Hz zu erkennen. In [Lin76] wurde durch sweep-tone Messungen in der Nase versucht das Übertragungsverhalten des Nasaltraktes zu messen, wofür das Velum während der

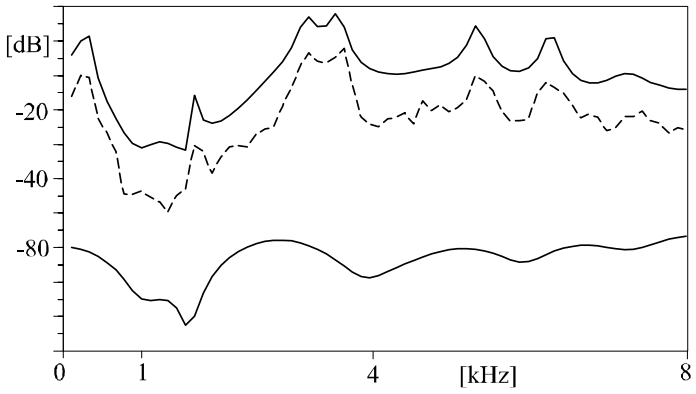


Bild 4.57: Analyse des nasalierten Vokals /ĩ/ mittels der iterativen inversen Filterung eines verzweigten Rohrmodells.

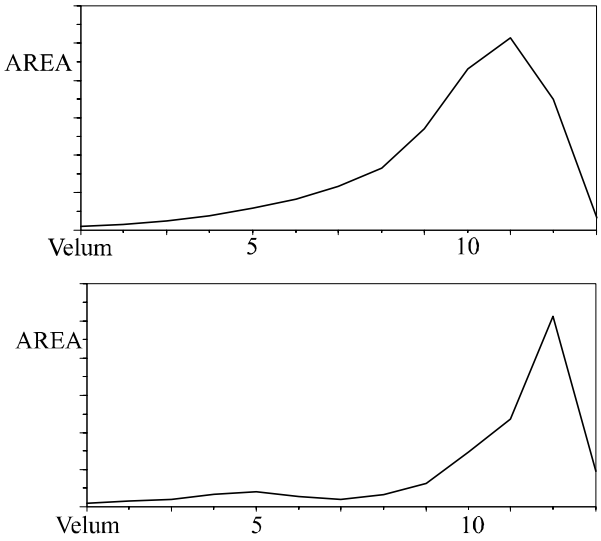


Bild 4.58: Geschätzte Flächen des Seitenzweiges aus den nasalierten Vokalen /ã/ (oben) und /ĩ/ (unten).

Messung geschlossen ist. Dabei ergab sich ein Nasalformant bei ca. 300 Hz, der mit den hier geschätzten Resonanzen aus dem Nasensignal unter 500 Hz korrespondiert. Bei einem Vergleich von Spektren nasalierter Vokale muß angemerkt werden, daß die Nasalierung der Vokale in verschiedenen Sprachen nicht notwendigerweise gleich ausgeprägt ist [Che97]. In Bild 4.62 links sind die Flächen des Seitenzweiges gezeigt, welche aus dem Mundsignal von /ã/ geschätzt wurden und in Bild 4.62 rechts die Rohrstücke zwischen Dreitor und Systemausgang, welche aus dem Nasensignal von /ã/ geschätzt wurden. Beide Teilrohrsysteme repräsentieren den Einfluß des Nasaltrakts. Während aus verschiedenen Nasensignalaufnahmen vergleichbare Nasenformen geschätzt werden können, sind die geschätzten Nasaltraktformen des Seitenzweiges aus dem Mundsignal weniger einheitlich. Dies hängt unter anderem damit zusammen, daß in dem einen Fall die Nullstellen die Flächenformen bestimmen, während im anderen Fall im wesentlichen die geschätzten Polstellen die Form bestimmen. Die durchgeführten Analysen belegen, daß durch die Verwendung von zwei Rohrmodellen die Signale hinreichend

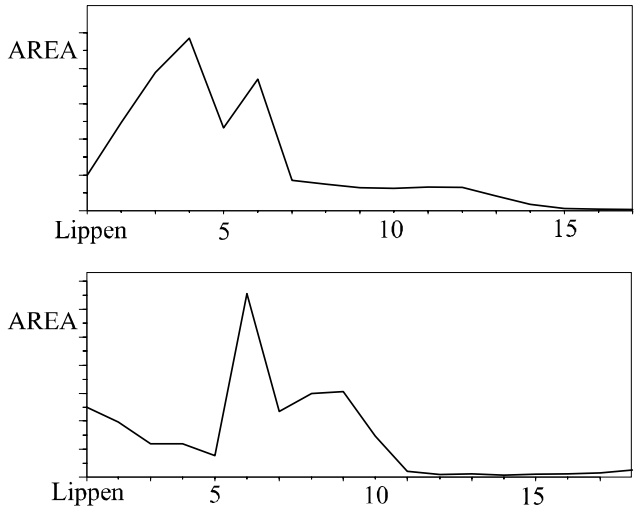


Bild 4.59: Geschätzte Vokaltraktflächen des verzweigten Rohrmodells aus den nasalierten Vokalen /$\widetilde{\text{a}}$/ (oben) und /$\widetilde{\text{i}}$/ (unten).

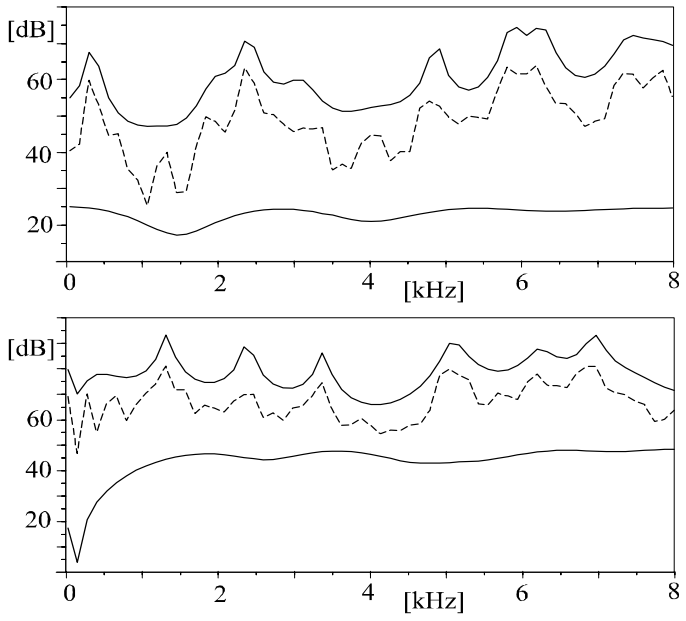


Bild 4.60: Analyse des Nasensignals (oben) und Mundsignals (unten) des nasalierten Schwa-Lautes mit jeweils einem verzweigten Rohrmodell: Geschätzter Betragsgang (obere durchgezogene Linie), DFT der analysierten Periode (mittlere gestrichelte Linie) und Beitrag der Nullstellen (untere durchgezogene Linie).

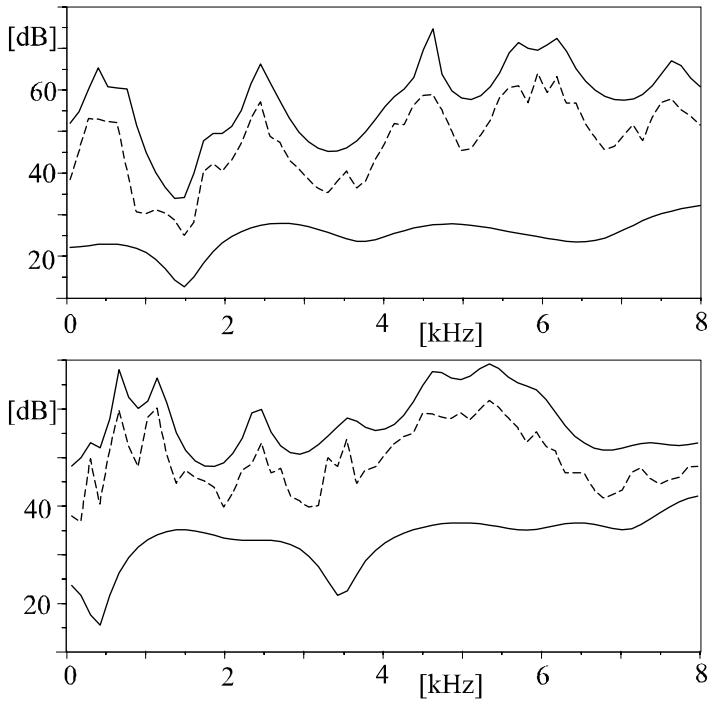


Bild 4.61: Analyse des Nasensignals (oben) und Mundsignals (unten) des nasalierten Vokals /ã/ mit jeweils einem verzweigten Rohrmodell: Geschätzter Betragsgang (obere durchgezogene Linie), DFT der analysierten Periode (mittlere gestrichelte Linie) und Beitrag der Nullstellen (untere durchgezogene Linie).

gut modelliert werden können im Vergleich zur Verwendung von nur einem Rohrmodell, welches zwei Systemausgänge besitzt. Die Verwendung eines einzigen Modells für das Mund- und Nasensignal stellt allerdings durch die einheitliche Modellierung des Sprechtraktes ein konsistentes Modell dar. Für eine bessere Modellierung beider Signale mit nur einem Modell wäre unter Umständen eine Erweiterung des Nasaltraktes durch Verzweigungen erforderlich.

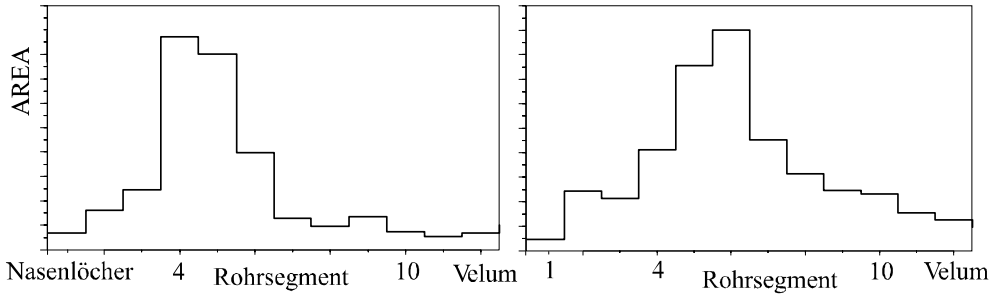


Bild 4.62: Geschätzte Flächen des Nasaltraktes: Rohrstücke zwischen Dreitor und Systemausgang aus dem Nasensignal von /ã/ (links); Seitenzweig des verzweigten Rohres aus dem Mundsignal von /ã/ (rechts).

Kapitel 5

Syntheseexperimente

Im vorherigen Kapitel wurden Methoden vorgestellt, welche die Modellparameter aus dem Sprachsignal schätzen. Die geschätzten Parameter werden nun für die Erzeugung von Sprachsignalen verwendet. Ein geschätzter Koeffizientensatz repräsentiert einen stationären Abschnitt des Sprachsignals. Erst durch einen Übergang der Modellparameter in einen weiteren Parametersatz können auch Lautübergänge und damit Lautketten synthetisiert werden. Da zwischen zwei Parameterätzen interpoliert wird und der Leistungs- und Spektralverlauf der interpolierten Parametersätze von der Parameterdarstellung abhängig sind, beeinflußt die Parameterdarstellung den Übergang. Für die Parameterdarstellung sollte gewährleistet sein, daß jeder interpolierte Parametersatz wieder ein stabiles System darstellt. Als Parameter bieten sich die Reflexionskoeffizienten, die daraus erhaltenen Flächen bzw. logarithmierten Flächen oder die Log Area Ratios an. Im Gegensatz zu diesen besitzen die LSF bzw. LSP (line spectrum frequencies / pairs) keine direkte Beziehung zu der Geometrie des Sprechtraktes, weisen aber gute Interpolationseigenschaften auf [Du97]. Die Steuerung des Modells der Spracherzeugung sieht Stützparametersätze vor, zwischen denen interpoliert wird. Die Wahl der Stützparametersätze und deren Anzahl wirkt sich neben einem geeigneten Interpolationsalgorithmus auf die Sprachqualität aus. Eine sehr grobe Auflösung der Stellen der Stützparametersätze ist bei einer Stützstelle pro Sprachlaut gegeben. Eine feine Auflösung liegt hingegen vor, wenn ähnlich der Sprachcodierung das Sprachsignal in kleine, quasi stationäre, benachbarte Segmente zerlegt wird, aus denen die Parametersätze ermittelt werden. Je gröber die zeitliche Auflösung der Analyse ausfällt, desto wichtiger wird der Interpolationsalgorithmus. Die fehlende Information über die dazwischen liegenden Stützstellen muß bei einer gröberen Auflösung durch den Interpolationsalgorithmus ersetzt werden. Instationäre Sprachabschnitte benötigen eine höhere Anzahl von Stützstellen als stärker stationäre Abschnitte. Als minimale Anzahl von Stützstellen kann ein Parametersatz pro Laut angesehen werden, wodurch die Lautübergänge vollständig durch den Interpolationsalgorithmus aus zwei Lautstützstellen gewonnen werden müssen. Dabei sind besonders Übergänge von Vokalen zu Konsonanten und umgekehrt interessant, da gerade Konsonanten durch ihre Übergänge charakterisiert werden.

5.1 Erzeugung von VCV-Übergängen

Der Vokal-Konsonant-Vokal Übergang, welcher hier mit VCV (englischer Sprachgebrauch) abgekürzt wird, kann als wesentliche Struktur von Lautketten angesehen werden. Daraus ergibt sich eine Darstellung der Sprache als abwechselnde Folge von Konsonanten und Vokalen, wobei die Konsonanten und Vokale hierfür auch durch eine Gruppe von mehreren Lauten ersetzt werden können. Während die Vokale einen verhältnismäßig uneingeschränkten Schallfluß durch den Vokaltrakt zulassen, sind die Konsonanten durch eine starke Konstriktion im Vokaltrakt charakterisiert. Durch die VCV-Struktur ist der Sprechvorgang von abwechselnden Vokaltraktkonfigurationen mit und ohne Konstriktionen und der damit resultierenden Übergänge beschrieben. Die Mannigfaltigkeit von VCV-Übergängen kann durch eine Abbildung der Konsonanten auf ihre Konstriktionsstelle verringert werden. Bei den Übergängen ist insbesondere der stimmhafte Übergang von Explosiven zu Vokalen interessant, da dieser eine starke Beeinflussung der Vokalformanten beinhaltet, was in [Oh66] an vielen Beispielen gezeigt ist. Für die Modellierung der Übergänge werden unverzweigte Rohrmodelle verwendet, deren Parameter die Querschnittsflächen selbst oder ihre logarithmierten Werte darstellen. Die Vokaltraktflächen der Konsonanten werden durch ihre benachbarten Vokale beeinflußt. Um solche Koartikulationseffekte zu berücksichtigen wird eine Gewichtsfunktion $\boldsymbol{w}$ eingeführt, welche den Einfluß der Konstriktionsflächen im zeitlichen Verlauf des Überganges bewertet. Diese Gewichtsfunktion $w_i(t)$ kann neben der Zeit t auch von der Position im Vokaltrakt abhängig sein, so daß die Flächen des Konsonanten im Bereich der Konstriktion einen stärkeren Einfluß besitzen. Für die Erzeugung eines VCV-Flächenübergangs wird zuerst ein Übergang zwischen den Vokalen gebildet, welche den Konsonant einschließen. Dieser VV-Übergang habe zur Zeit t den Parametervektor $\boldsymbol{p}^{vv}(t)$, welcher die Flächen bzw. die logarithmierten Flächen des Rohrmodells darstellt. Bild 5.1 zeigt hierfür als Beispiel einen Übergang von /a/ zu /i/. Zu diesem Übergang wird die Flächenkonfiguration der Konstriktion eingeblendet, die durch einen Parametervektor $\boldsymbol{p}^c$ dargestellt wird. Die Gewichtsfunktion $w(i,t)$ stellt dabei die Stärke der Einblendung von $\boldsymbol{p}^c$ dar [SnL00c]. Der resultierende Parametervektor $\boldsymbol{p}^{vcv}(t)$ der VCV-Sequenz ergibt sich zu

$$p_i^{vcv}(t) = p_i^{vv}(t) \cdot (1 - w_i(t)) + p_i^c \cdot w_i(t), \tag{5.1}$$

wobei der Index i die Position im Vokaltrakt angibt. Für die analysierten Flächen des Rohrmodells mit dem zeitabhängigen Glottisabschluß zeigt Bild 5.2 einen Flächenübergang von /a/ zu /k/ zu /i/. Die Flächen des Explosivs sind die ermittelten Flächen aus Bild 4.21. Die Vokaltraktflächen in Bild 5.2 stellen nur den stimmhaften Übergang dar, da das Explosionsgeräusch für den stimmlosen Explosiv durch ein Zeitsignal realisiert wird. Dieser Zeitabschnitt kann in der Nähe der Konstriktion in das Rohrmodell als Eingang eingespeist werden, so daß er durch das Modell eine Filterung erfährt. Dies führt allerdings nicht zwangsläufig zu einer besseren Sprachqualität. Die stimmhafte Anregung des Rohrmodells ist eine Impulsfolge, die mit einem System reeller Pole gefiltert wird, welche sich aus der Präemphase ergeben. Die synthetisierte Lautsequenz ist verständlich. Neben /k/ wurden auch /p/ und /t/ mit den Flächen aus Bild 4.21 verwendet, was zu ähnlichen Ergebnissen führte (siehe Hörbeispiele). Die stimmhaften Explosive sind mit den Flächen von Bild 4.21 weniger gut zu bilden wie die stimmlosen Explosive. Neben den zeitvariablen Rohrmodellen ist auch ein zeitinvariantes

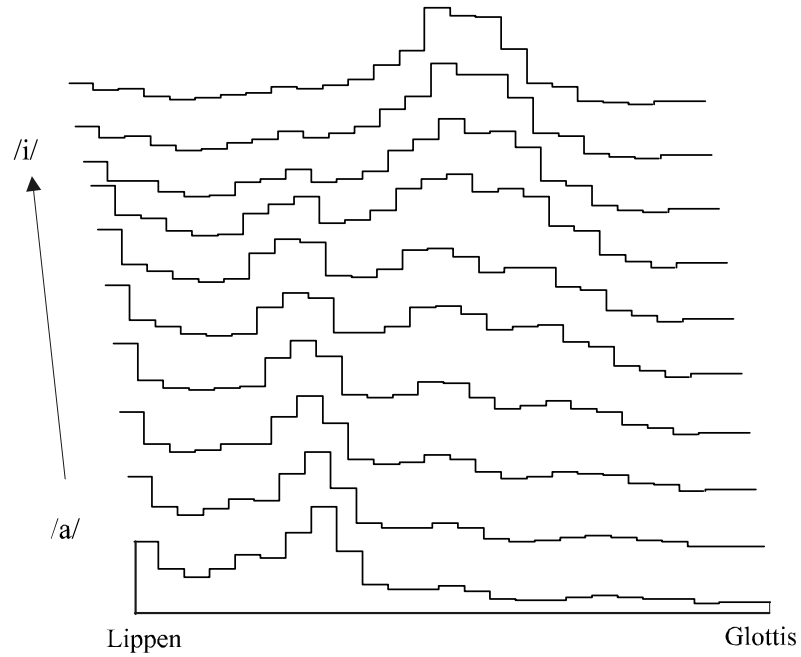


Bild 5.1: Generierter Flächenübergang vom Vokal /a/ zum Vokal /i/.

Rohrmodell benutzt worden, um VCV Sequenzen zu generieren. Für die Synthese von stimmhaften Explosiven werden die Parameter nun nicht aus Lauten gewonnen, die ein stationäres Halten der entsprechenden Vokaltraktkonfiguration verlangen, sondern aus Lautübergängen stimmhafter Konsonanten. Dazu werden die geschätzten logarithmierten Flächen der Bilder 4.49 verwendet, welche aus den Äußerungen [b@], [d@] und [g@] geschätzt worden sind [SnL01c]. Das in diesem Fall hier verwendete Rohrmodell ist im Gegensatz zum vorherigen Beispiel zeitinvariant und besitzt einen frequenzabhängigen Rohrabschluß an der Glottis. Das Bild 5.3 zeigt die Spektrogramme der synthetisierten Sprachsignale der Lautketten [aba], [ada] und [aga] (siehe auch Hörbeispiele). Die Formantabbiegungen infolge der Konstriktionen der Vokaltraktflächen sind gut zu erkennen. In Bild 5.4 sind beispielhaft Spektrogramme von natürlichen Sprachsignalen desselben Sprechers gezeigt, in denen die Übergänge der Formanten zu sehen sind. Diese weisen Ähnlichkeiten mit denen der synthetisierten Sprachsignale von Bild 5.3 auf. Es ist dabei zu beachten, daß die Vokaltraktflächen des Konsonanten für die Synthese nicht aus einem Übergang zum Vokal /a/ gewonnen wurden. Da die Explosive /g/ und /k/ dieselbe Konstriktionsstelle aufweisen, müssen für diese beiden Konsonanten die Formantabbiegungen vergleichbar sein. Das Spektrogramm der natürlichen Äußerung [aka] aus Bild 5.4 ist deshalb mit dem Spektrogramm der synthetisierten Lautkette [aga] in Bild 5.3 ähnlich. Auch bei Übergängen mit beteiligten Frikativen geht der Vokaltrakt im stimmhaften Bereich von der Stellung des Vokals in eine Vokaltraktstellung des Konsonanten über und umgekehrt. Dies ist nicht so stark ausgeprägt wie im Falle der stimmhaften Explosive, da unter anderem die Artikulationsbewegung nur teilweise im stimmhaften Abschnitt erfolgt. Die Bewegungen der Resonanzen infolge der Formantabbiegungen richten sich zu bestimmten Frequenzen, welche mit der Konstrikionsstelle korrelieren, entsprechend der Locus-Theorie. Dadurch sind die Formantabbiegungen für die Konsonanten /b/ und /p/ vergleichbar,

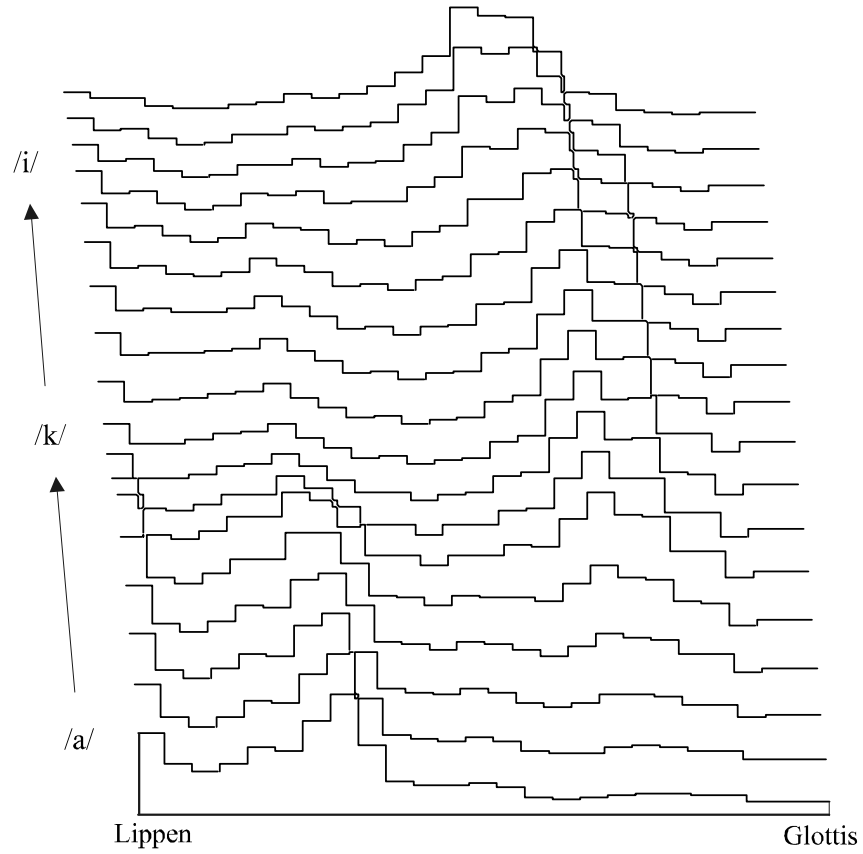


Bild 5.2: Generierte Vokaltraktflächen des Vokal-Konsonant-Vokal Überganges [aki].

wie auch für die Konsonanten /s/ und /z/, da die Konstriktionen jeweils an denselben Stellen auftreten. In [Na95] ist anhand von MRT-Aufnahmen zu sehen, daß die Vokaltraktflächen von stimmhaften und stimmlosen Frikativen vergleichbar sind, während für verschiedene Sprecher sich die gezeigten Flächen stärker unterscheiden. In Bild 5.5 sind Spektrogramme der natürlichen Äußerungen [aSa] und [asa] gezeigt, in dem ebenfalls Formantabbiegungen zu erkennen sind. Um den stimmhaften Teil der Lautsequenz [aSa] zu synthetisieren werden die Flächen von /Z/ aus Bild 4.48 verwendet. Der Flächenübergang für den stimmhaften Teil ist im Bild 5.6 dargestellt, in dem die Konstriktion in der Mitte zu erkennen ist. Das Spektrogramm des synthetisierten Sprachsignals [aSa], welches durch diesen Flächenübergang erzeugt wurde, ist in Bild 5.7 zu sehen. Der rauschhafte Anteil wird mit einem Zeitsignal realisiert, das in der Mitte des Übergangs eingespielt wurde. Zwei Beispiele von Zeitsignalen synthetisierter Lautübergänge sind in Bild 5.8 gezeigt, welche durch einen Flächenübergang eines Konsonanten zu einem Vokal erzeugt wurden. Es muß angemerkt werden, daß sich die Vokaltraktflächen bei einem VCV-Übergang kurz vor und hinter einem Explosiv oft unterscheiden, da sich während des stimmlosen Abschnittes die Artikulatoren bewegen. Die Bewegungen der einzelnen Artikulatoren müssen dabei nicht zeitlich synchron sein [Ga77, Lö99], da sich zum Beispiel in der Regel erst die Zunge und dann die anderen Artikulatoren verändern. In [Ga77] ist durch Messungen beobachtet worden, daß sich im stimmlosen Abschnitt die Zunge in Richtung des nachfolgenden Vokals bewegt.

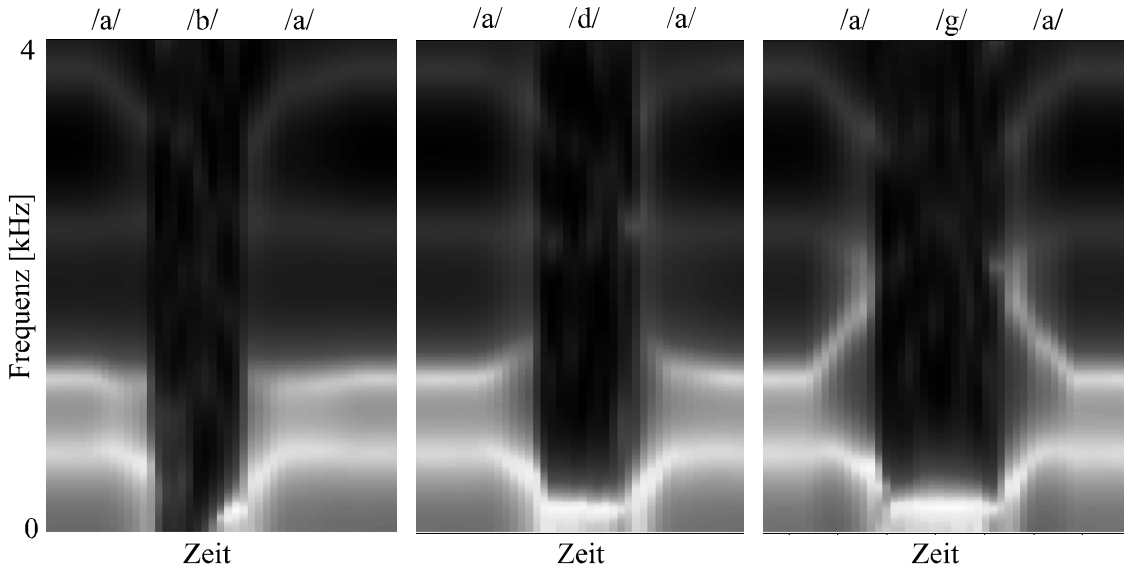


Bild 5.3: Spektrogramme der synthetisierten Lautketten [aba] (links), [ada] (mitte) und [aga] (rechts).

Übergänge zu unterschiedlichen Vokaltraktlängen

Für die Analyse von Einzellauten kann das Rohrmodell durch Verwendung eines Glottisabschlusses unterschiedliche Vokaltraktlängen aufweisen. Dies hat für die Synthese zur Folge, daß sich Vokaltraktlängen während der Lautübergänge ändern können. Die Zeitdiskretisierung des Rohrmodells läßt aber nur ganzzahlige Vielfache der Einheitsrohrlänge explizit zu. Die Länge eines Einheitsrohres ist wiederum von der Abtastrate abhängig. Eine kontinuierliche Rohrverkürzung bzw. Rohrstreckung wäre über eine Variation der Abtastrate möglich. Es können auch kleine digitale Filter statt der Zustandsspeicher verwendet werden, die eine variable Zeitverzögerung modellieren. Dafür eignen sich z.B. Allpaßfilter, da sie sich nur auf die Phase des Signals auswirken. Eine Interpolation mit FIR-Filtern ist in [Vä94b, Str75b] diskutiert, wobei in [Vä94b] auch der erhöhte Rechenaufwand insbesondere für hohe Frequenzen angemerkt wird, da dort infolge der kürzeren Wellenlängen größere Fehler auftreten. Ein Vergleich und eine Übersicht ist in [Vä00, Str00] gegeben. Diese (fractional-delay) Methoden erfordern einen erhöhten Rechenaufwand, so daß eine einfachere Realisierung für die Synthesebeispiele benutzt wird. Mit der Verwendung von einzelnen Zustandsspeichern für die Rohrelemente ist nur eine diskrete Vokaltraktlängenänderung möglich. Wenn für einen Laut L_1 mit einer Vokaltraktlänge N_1 der nachfolgende Laut L_2 eine Länge von $N_2 = N_1 - 1$ aufweist, so muß im Übergang ein Rohrelement eliminiert werden. Dafür werden für den Laut L_1 zwei Modellparametersätze bereitgestellt: $\boldsymbol{p}_1^1$ und $\boldsymbol{p}_1^2$ mit den Längen N_1 und N_2. Gleiches wird für den nachfolgenden Laut L_2 vollzogen, so daß $\boldsymbol{p}_2^1$ mit der Länge N_1 und $\boldsymbol{p}_2^2$ mit Länge N_2 ermittelt werden. $\boldsymbol{p}_1^1$ und $\boldsymbol{p}_2^2$ sind die vom Schätzalgorithmus zur Verfügung gestellten Parametersätze. $\boldsymbol{p}_1^2$ und $\boldsymbol{p}_2^1$ werden aus diesen gewonnen, indem in $\boldsymbol{p}_1^2$ ein Zweitor an einer zu bestimmenden Stelle weggenommen wird und an der gleichen Stelle in $\boldsymbol{p}_2^1$ ein Zweitor eingefügt wird, wobei der dazugehörige Reflexionskoeffizient den Wert Null besitzt. Für diese Operation wird eine Position bzw. ein Zweitor im Rohrmodell gewählt, die einen Mindestabstand zu den

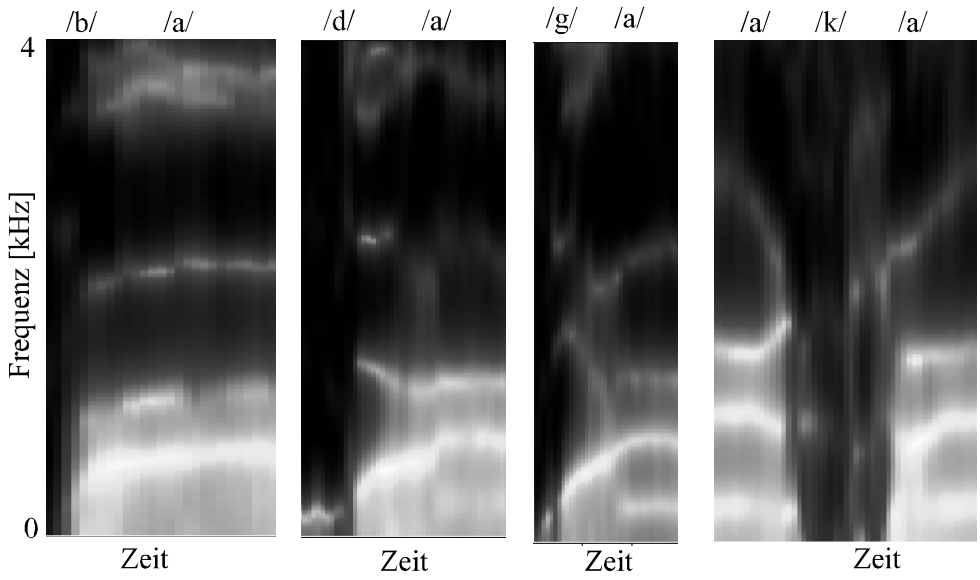


Bild 5.4: Spektrogramme von Sprachaufnahmen der Lautketten [ba], [da], [ga] und [aka] (von links nach rechts).

Rohrabschlüssen aufweist und einen betragsmäßig kleinen Reflexionskoeffizienten besitzt. Da dieser Reflexionskoeffizient verschwinden soll, ist es von Vorteil, wenn dieser einen möglichst kleinen Wert besitzt. Um einen zeitkontinuierlichen Verlauf der Koeffizienten zu gewährleisten, wird ein Parametersatz $\widetilde{\boldsymbol{p}}_1^1$ definiert, der mit $\boldsymbol{p}_1^1$ idenstisch ist mit der Ausnahme, daß der Koeffizient an der herauszunehmenden Stelle Null ist. Damit kann der Übergang in zwei Schritten vollzogen werden. Zuerst wird ein Übergang mit der Vokaltraktlänge N_1 durchgeführt, welcher beschrieben wird durch:

$$(1-\boldsymbol{\lambda}/2)\cdot\left(\boldsymbol{\lambda}\cdot\widetilde{\boldsymbol{p}}_1^1+(1-\boldsymbol{\lambda})\boldsymbol{p}_1^1\right)+\boldsymbol{\lambda}/2\cdot\boldsymbol{p}_2^1 \tag{5.2}$$

$$\text{für}\quad \boldsymbol{\lambda}=0\longrightarrow 1. \tag{5.3}$$

Darauffolgend wird an der entsprechenden Stelle das Zweitor entfernt. Der herauszunehmende Reflexionskoeffizient hat zu diesem Zeitpunkt den Wert Null erreicht. Danach wird mit einem Übergang mit der Rohrlänge N_2:

$$(1-\boldsymbol{\lambda})/2\cdot\boldsymbol{p}_1^2+(1+\boldsymbol{\lambda})/2\cdot\boldsymbol{p}_2^2 \tag{5.4}$$

$$\text{für}\quad \boldsymbol{\lambda}=0\longrightarrow 1 \tag{5.5}$$

der zweite Teil des Übergangs vollendet. Durch diese Realisierung des Überganges sind im synthetisierten Zeitsignal keine Amplitudensprünge oder Artefakte zu beobachten. Ein Übergang mit einer Rohrverlängerung läßt sich analog vollziehen. Unterscheiden sich die Vokaltraktlängen in mehreren Rohrstücken, so muß der Übergang in mehrere Übergänge zerlegt werden, in denen jeweils ein Rohrstück hinzugenommen bzw. eliminiert wird.

5.2 Resynthese

Werden, wie im vorherigen Abschnitt, sehr wenige Parametersätze als Stützstellen für die Synthese der Lautketten verwendet, so müssen die Übergänge ganz oder teilweise

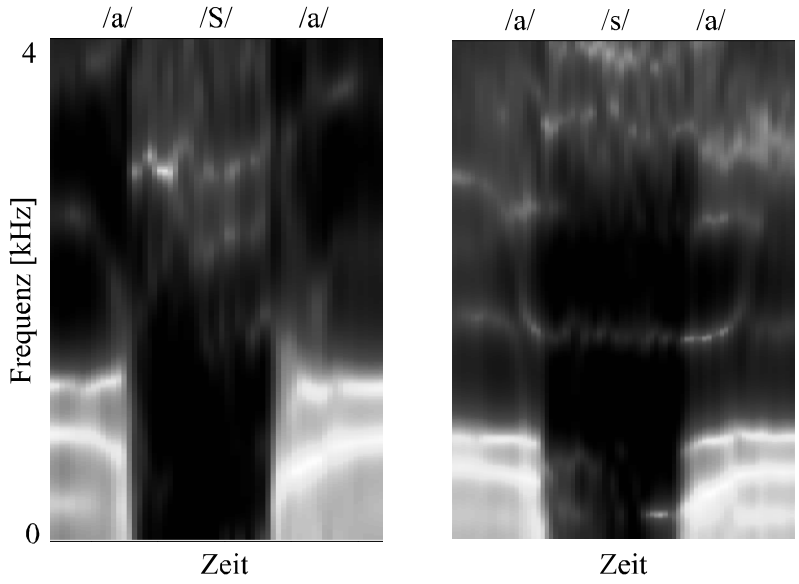


Bild 5.5: Spektrogramme der original Sprachsignale [aSa] (links) und [asa] (rechts).

durch die interpolierten Parametersätze modelliert werden, was sich auf die Sprachqualität auswirkt. Da die Systemanregung, die Parameterdarstellung, die Interpolation der Flächenübergange, die Auswahl der Sprachlaute sowie die Stützstellen die Qualität des synthetisierten Sprachsignals beeinflussen, können deren Auswirkungen nur bedingt differenziert beurteilt werden. Um die Qualität der erzeugten Sprachsignale in Bezug auf die verwendeten Filter und Anregungsmodelle zu testen, wird eine Resynthese durchgeführt. Abgesehen davon kann die Resynthese auch für andere Zwecke verwendet werden, da z.B. eine Verkettung von resynthetisierten Diphonen einen Teilaspekt einer Sprachsynthese abdecken. Für die Resynthese werden die Stützstellen dicht gewählt, um eine genaue Modellierung des analysierten Sprachsignals zu ermöglichen. Im Gegensatz zu den synthetisierten VCV-Übergängen des letzten Abschnitts stammen die Parametersätze nicht aus unterschiedlichen Sprachäußerungen, so daß Koartikulationseffekte schon in der Datenbasis berücksichtigt sind. Die Resynthese ist daher mit der Codierung verwandt, allerdings unterscheidet sie sich von dieser durch eine zusätzliche Variation des Sprachsignals in der Grundfrequenz, Lautdauer und Leistung. Durch das original Sprachsignal ist ein Referenzsignal für eine Beurteilung der Sprachqualität vorhanden. Für die Resynthese sollen nur die Modellparameter der analysierten Signale verwendet werden, so daß das Residualsignal nicht für die Anregung des Synthesesytems verwendet wird. Es wird eine Anregung angestrebt, welche weitgehend unabhängig von dem analysierten Sprachsignal ist. Eine Abhängigkeit der Anregung zu der zu synthetisierenden Phonemkette und zu prosodischen Werten kann allerdings sinnvoll sein, da die stimmhafte Anregung durch den Anregungsmechanismus und der Kopplung zwischen Glottis und Vokaltrakt von den Phonemen und der Prosodie beeinflußt wird. Diese Abhängigkeit kann sich dabei auf das Spektrum und den Rauschanteil der stimmhaften Anregung beziehen. Resyntheseexperimente sind in früheren Untersuchungen in [Chi94, Rah89, So83, Fra86, Gu93] vorgestellt worden, welche sich hinsichtlich der Modelle, der Anregung sowie der Parameterbestimmung aus dem Sprachsignal unterscheiden. In sehr frühen Untersuchungen [So83], in denen

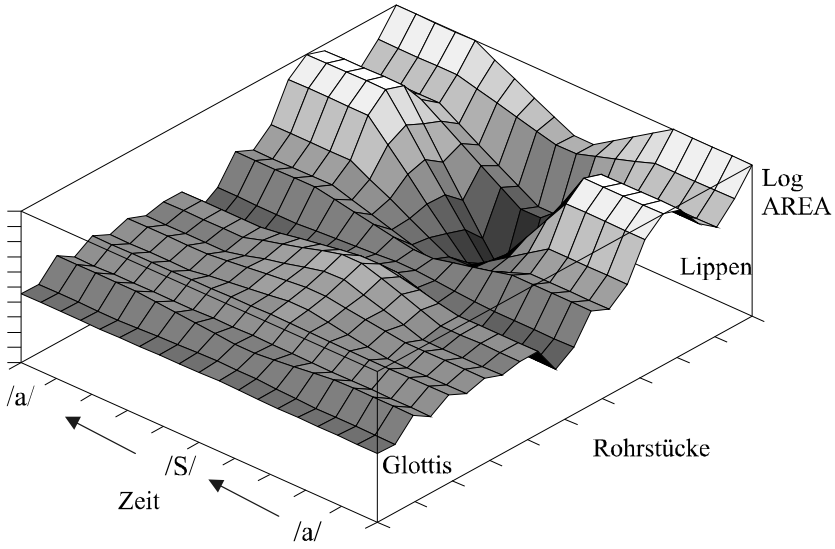


Bild 5.6: Generierte logarithmierte Vokaltraktflächen als Polygonzug des VCV-Überganges [aSa]. Rohrmodell mit frequenzabh ängigem Glottisabschluß.

die ermittelten Flächen direkt für die Synthese eingesetzt werden, wurde noch von einer mäßigen Sprachqualität berichtet. In [Rah89] wird ein neuronales Netz verwendet, um die Sprachspektren auf die Flächen eines Sprechtraktmodells abzubilden. In [Chi94] wird eine Glottisfunktion als Anregung benutzt, die durch ein Polynom approximiert wird; mit Verwendung eines Codebuches wurde der Begriff GELP statt CELP abgeleitet. In [FraL86] wurden verteilte Verluste in das unverzweigte Rohrmodell integriert. Bei dem Beispiel in [Chi94], wie auch in [Fra86], ist die Anregung allerdings vom analysierten Signal stark abhängig, so daß [Chi94, Fra86] eher eine Codierung darstellen.

5.2.1 Periodensynchrone Analyse des stimmhaften Sprachsignals

Das Sprachsignal wird für die Analyse in stimmhafte und stimmlose Abschnitte unterteilt, wobei nur die stimmhaften Abschnitte analysiert werden. Für die Analyse und Resynthese wird ein unverzweigtes Rohrmodell ohne Glottisabschluß verwendet, womit die Vokaltraktlänge im Verlauf der Lautäußerung nicht adaptiert werden muß. Für Nasale wird die gleiche Struktur des Rohrmodells verwendet, was überraschenderweise in einer ersten Beurteilung zu keiner merklichen Verschlechterung der Synthese für die Nasale im Vergleich zu den Vokalen führte. Der Rohrabschluß am Systemausgang wird betragsmäßig kleiner als Eins gewählt, um Verluste zu berücksichtigen. Neben dem reellen Rohrabschluß wird alternativ auch der frequenzabhängige Abschluß des Laine-Modells verwendet. Durch diese Rohrabschlüsse wird durch die Burg-Methode keine optimale Schätzung erreicht, so daß die iterative inverse Filterung verwendet wird. Die Perioden der stimmhaften Sprachabschnitte sind für die Analyse an den Nulldurchgängen markiert. Vor der Analyse werden die Perioden mit einer festen Präemphase von -0,95 vorgefiltert. Eine von Periode zu Periode adaptive Präemphase hat sich in den durchgeführten Syntheseexperimenten noch nicht bewährt. Um Fluktuationen zwischen den Sprachperioden zu vermindern, werden benachbarte Perioden im

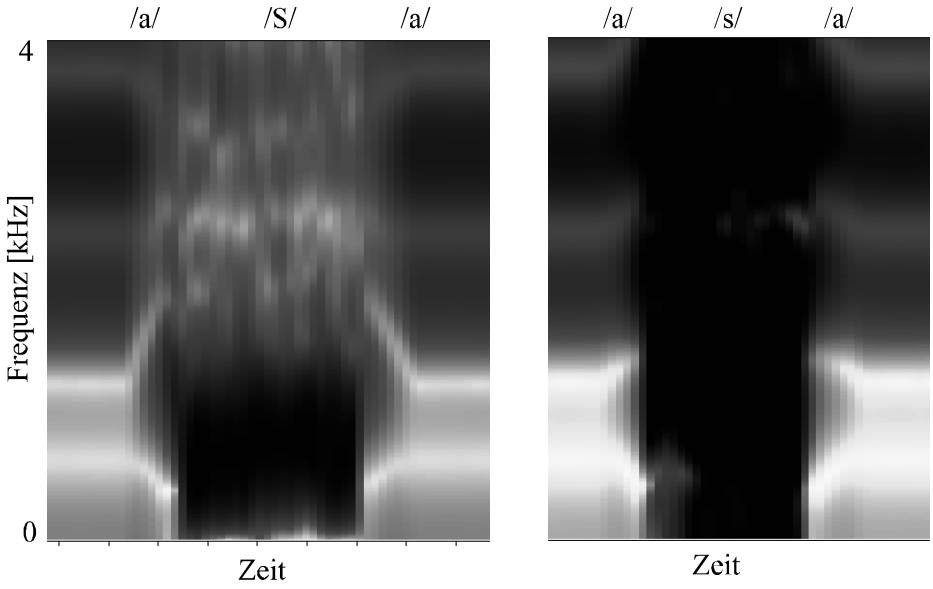


Bild 5.7: Spektrogramme der synthetisierten Lautketten [aSa] (links) und [asa] rechts.

Spektralbereich gemittelt. Diese gemittelten Perioden werden dann nacheinander mit der iterativen inversen Filterung analysiert. Dazu können die Formeln (4.143) der Teillösungen verwendet werden, wobei der Glottiskoeffizient in $\boldsymbol{F}_i$ den Wert Null besitzt. Nach der Analyse der ersten gemittelten Periode werden die nächsten Perioden mit weniger Iterationen geschätzt, da die Schätzung mit einer Startkonfiguration beginnt, welche das Analyseresultat der vorherigen Periode darstellt. Für die erste Periode wird die Analyse von einer neutralen Position der Modellparameter begonnen, bei der alle Reflexionskoeffizienten gleich Null sind. Bild 5.9 zeigt den geschätzten Flächenübergang aus dem analysierten Lautübergang [va] aus [vaI]. Dieser stellt nur einen Ausschnitt aus der ganzen analysierten Sprachäußerung dar. Der Rohrabschluß für die Analyse ist mit -0,9 festgelegt. Jeder Flächensatz in Bild 5.9 repräsentiert eine analysierte Periode, wobei anzumerken ist, daß die Sprachperioden vor der Analyse mit ihren Nachbarperioden gemittelt wurden. Der Übergang von [v] zu dem Diphthong [aI] ist in der Flächenentwicklung gut zu erkennen. Die geschätzten Flächen werden für die Resynthese des Lautübergangs bzw. der Lautketten verwendet; das zugehörige Residualsignal der Analyse wird allerdings nicht für die Synthese berücksichtigt. Die Anregung ist somit nicht von dem analysierten Sprachsignal abhängig, abgesehen von prosodischen Parametern des Grundfrequenzverlaufs.

5.2.2 Anregung des Rohrmodells

Für die Resynthese wird hier für alle stimmhaften Laute das gleiche Anregungsmodell verwendet. Als einfachster Ansatz kann für die stimmhafte Anregung eine Impulsfolge verwendet werden. Diese wird noch durch das rekursive System mit den Präemphasekoeffizienten gefiltert. Das resultierende Anregungssignal ist streng periodisch, was für das natürliche Sprachsignal nur annäherungsweise gilt. In [To01] sind für diese Abweichungen der Periodizität auch Nichtlinearitäten der Glottisanregung als Begründung angegeben. Eine Ursache der Abweichung zur exakten Periodizität

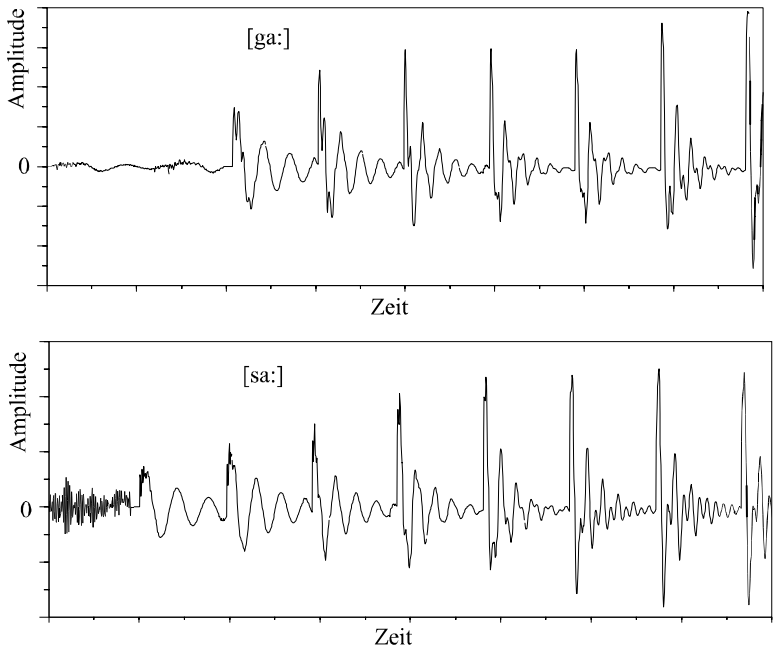


Bild 5.8: Zeitsignale der synthetisierten Lautketten [ga:] und [sa:] durch Flächenübergänge.

ist auch durch das Rauschen infolge von Wirbeln hinter der Glottis begründet, welche auch durch Lösungen der Navier-Stokes Gleichung beobachtet werden können [Ii89, Li91, Zh02]. In [Qi00] wird eine Bestimmung des Rauschanteils in stimmhafter Sprache diskutiert. Eine streng periodische Anregung verleit der Synthese den bekannten maschinenhaften Klang [Sa78]. Um diesen unnatürlichen Klangeindruck zu vermeiden, muß das Anregungssignal kleine Variationen von Periode zu Periode aufweisen. Dafür können Amplitude und Periodenlänge periodenweise leicht variiert werden. Das natürliche stimmhafte Sprachsignal weist allerdings in verschiedenen Frequenzbändern eine unterschiedliche Periodizität auf, da in der Regel zu höheren Frequenzen hin das Sprachsignal eine geringere Periodizität aufweist. Um eine natürlich klingende Anregung zu erhalten, wird die Anregung aus dem Residualsignal gewonnen, das aus einer Burg-Analyse einer einzelnen Sprachäußerung ermittelt wurde. Als Äußerung wird eine Aufnahme des Schwa-Lautes verwendet, welche mit monotoner Grundfrequenz gesprochen wurde. Es werden wenige benachbarte Perioden aus diesem Signal entnommen, die die Datenbasis für die stimmhafte Anregung bilden [SnL01d]. Bild 5.10 zeigt fünf benachbarte Perioden aus der Äußerung des Schwa-Lautes bei einer Abtastrate von 22 kHz.Die Perioden des Sprachsignals werden nicht unmittelbar übernommen, sondern erfahren noch eine Vorverarbeitung. Der geschätzte Betragsgang der Burg-Analyse kann die spektrale Einhüllende des analysierten Signals nicht exakt reproduzieren. Daher werden die Perioden im Spektralbereich in der Weise vorverarbeitet, daß die über alle Perioden gemittelten Spektralwerte für alle Frequenzen gleiche Mittelwerte aufweisen, was sich insbesondere im unteren Frequenzbereich auswirkt. Die Fluktuationen von Periode zu Periode bleiben dabei erhalten, da nur die Mittelwerte verändert werden. Die Perioden werden zusätzlich mit ihren Nachbarn zu einer Periode gemittelt. Diese gemittelten Perioden werden dann nacheinander verwendet, um die stimmhafte

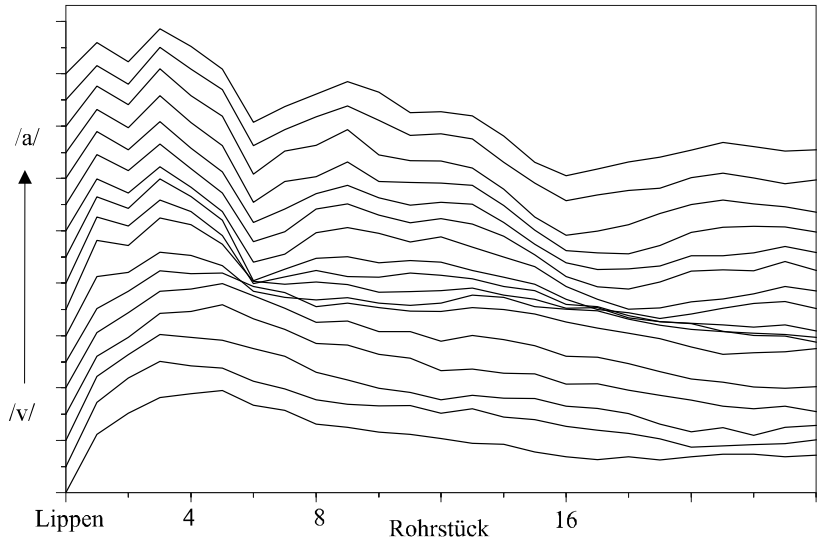


Bild 5.9: Ausschnitt aus einem gesch ätzten Flächenübergang des Lautübergangs /v/ zu /a/.

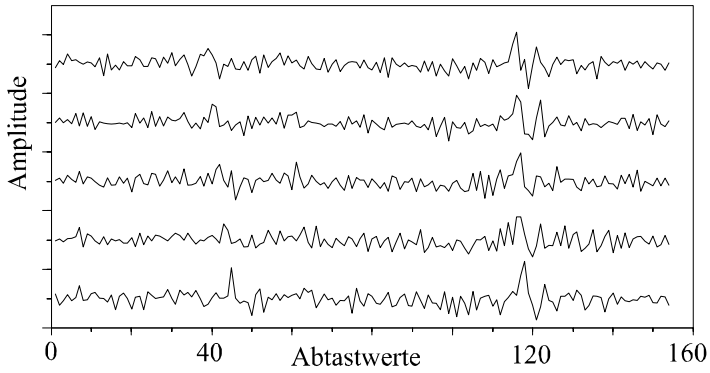


Bild 5.10: Benachbarte Residualperioden des Schwa-Lautes.

Anregung zu bilden. Beim Erreichen der letzen Periode wird wieder mit der ersten Periode zyklisch fortgefahren. Die Perioden müssen allerdings noch für den vorgegebenen Grundfrequenzverlauf auf die benötigte Länge gebracht werden, was durch eine Operation im Spektralbereich durchgeführt wird. Für eine Verlängerung der Perioden können die Frequenzmoden mit der Höchsten beginnend abwärtszählend für die neu hinzugekommenen Moden doppelt verwendet werden, wohingegen für eine Periodenverkürzung einfach höhere Frequenzmoden weggelassen werden. Da nur wenige Perioden (<10) benutzt werden, können diese ohne großen Speicherbedarf mit verschiedenen Längen abgespeichert werden, so daß sie während der Synthese nicht berechnet werden müssen. Im Bild 5.11 oben sind fünf vorverarbeitete Perioden gezeigt, deren Periodenlänge zusätzlich um zehn Abtastwerte verringert wurde. In Bild 5.11 unten sind zusätzlich die Betragsspektren der Perioden dargestellt.Die Fluktuationen von Periode zu Periode der modifizierten Residualsignale sind im Gegensatz zu den original Residualperioden schwächer geworden. In den Spektren von Bild 5.11 ist zu erkennen, daß die spektralen Schwankungen im oberen Spektralbereich stärker ausgeprägt sind.

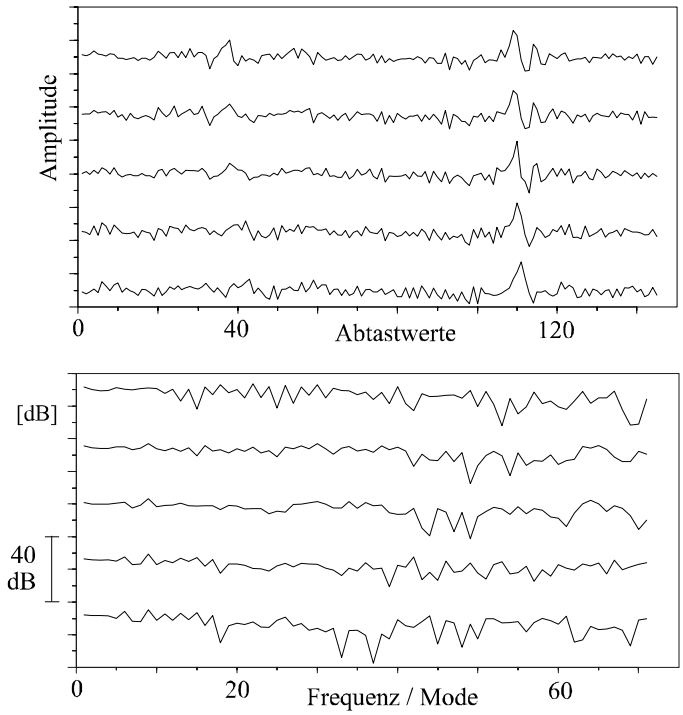


Bild 5.11: Modifizierte Residualperioden des Schwa-Lautes: Zeitsignale der Perioden (oben), DFT der Perioden (unten).

5.2.3 Resynthesebeispiele

Das verwendete Filter für die Resynthese ist ein unverzweigtes Rohrmodell in der Leistungswellendarstellung. Die Leistung einer durch das Rohrmodell gefilterten Anregungsperiode ist von der Wellendarstellung und den Rohrmodellparametern abhängig. Um die Ausgangsleistung bei der Synthese für die einzelnen Parametersätze zu normieren, wird für jeden Parametersatz die Leistung einer, mit einer Impulsfolge angeregten, synthetisierten Ausgangsperiode abgespeichert. Diese Berechnung muß für jedes analysierte Signal nur einmal durchgeführt werden. Für die Resynthese werden die Filterparameter von Periode zu Periode linear in der Darstellung der logarithmierten Flächen von einem Koeffizientensatz zum Nächsten überführt. Für die Anregung werden die vorverarbeitenden Residualperioden benutzt, die auf die entsprechende Länge der Grundfrequenz gestreckt oder gestaucht werden. Wie erläutert sind die Anregungsperioden unabhängig von der zu resynthetisierenden Lautäußerung gewonnen worden. Der mittlere Graph in Bild 5.12 zeigt einen mit dieser Anregung resynthetisierten Lautübergang [va], welcher durch den Flächenübergang von Bild 5.9 erzeugt wurde.Die Anregungsperioden müssen für jeden Grundfrequenzverlauf angepaßt werden, da sie nicht von der zu resynthetisierenden Äußerung entnommen worden sind. Dadurch weisen resynthetisierte Sprachäußerungen mit verändertem Grundfrequenzverlauf im Prinzip dieselbe Sprachqualität auf als solche, bei denen der Originalverlauf f_0 verwendet wurde. Dabei ist vorausgesetzt, daß der veränderte Grundfrequenzverlauf nicht als solches unnatürlich ist. Für stimmlose Sprachabschnitte wird einfach das Originalsignal für die Resynthese übernommen. Die so resynthetisierte Sprache kann als natürlich klingend eingestuft werden und ist zum größten Teil über Lautsprecher nicht als Resynthese

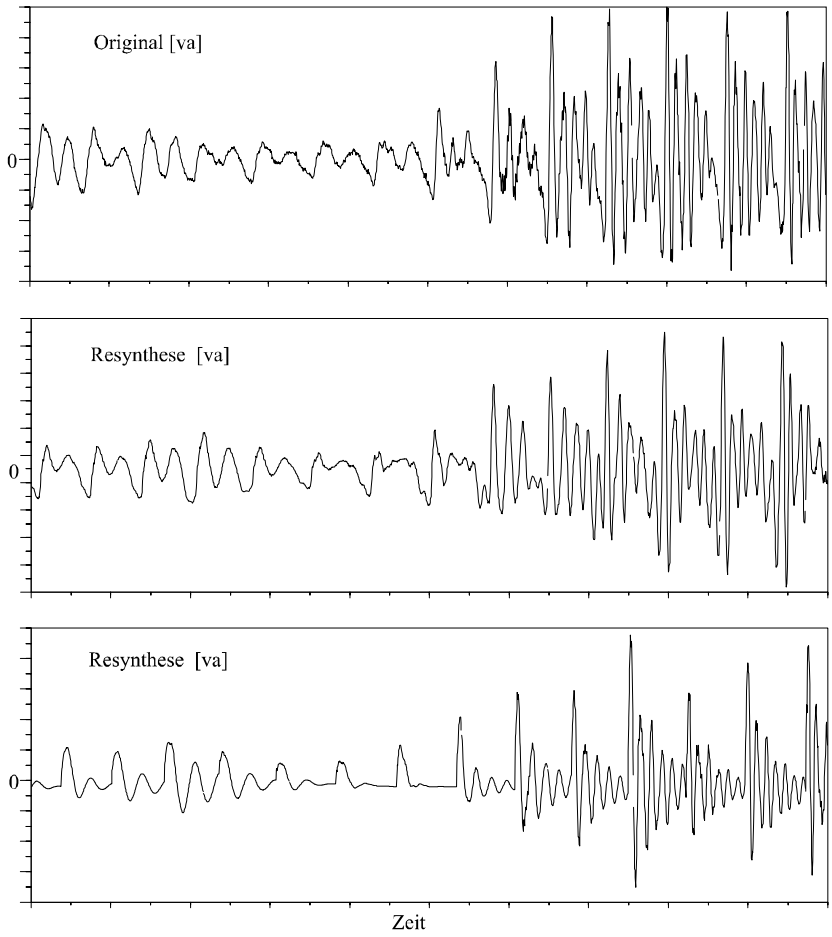


Bild 5.12: Original und resynthetisierter Lautübergang [va]: Original Sprachsignal (oben), Resynthese mit Modellanregung der modifizierten Residualperioden (mittig), Resynthese mit Impulsfolgen-Anregung (unten).

zu erkennen (siehe Hörbeispiele). Im Vergleich zum Originalsignal ist dennoch ein Unterschied hörbar. Im Vergleich mit den Residualperioden führt die Verwendung von einer Impulsfolge als Anregung auch mit Amplituden- und Periodenlängenvariationen zu einer schlechteren Sprachqualität. Im Bild 5.12 unten ist das resynthetisierte Sprachsignal mittels einer Anregung mit einer Impulsfolge gezeigt. Für die Analyse und Resynthese wird für die Nasallaute, wie für die anderen Laute, auch ein unverzweigtes Rohrmodell verwendet, wodurch allerdings keine schlechtere Sprachqualität der Nasale im Vergleich zu den übrigen stimmhaften Lauten erzielt wird. Dadurch kann allerdings nicht allgemein die Schlußfolgerung gezogen werden, daß verzweigte Rohrmodelle bzw. andere Pol-Nullstellen Systeme keine zusätzliche Verbesserung erbringen. Durchgeführte Resyntheseexperimente mit allgemeinen Pol-Nullstellen Systemen, in denen die Nullstellen durch Kreuzgliedstrukturen realisiert sind und resynthetisierte Perioden überlappend ineinander übergehen, führten zu keinen besseren Ergebnissen. Dies kann damit zusammenhängen, daß die geschätzten Nullstellen, und dadurch auch die Pole, von Analysesegment zu Analysesegment teilweise unstetig hin und her springen. Eine Verbesserung könnte hier eine ARMA-Schätzung erbringen, die die benachbarten Analysesegmente in die Schätzung mit einbezieht, so daß ein stetigen Verlauf der Bewegung

der Pole und Nullstellen gefördert wird. Das hier verwendete unverzweigte Rohrmodell enthält keine Nullstellen, abgesehen von einem möglichen frequenzabhängigen Rohrabschluß. Die abrupten Übergänge zwischen Nasallauten und den benachbarten stimmhaften Lauten stellen in beiden Richtungen keine Probleme dar. In Bild 5.13 ist ein resynthetisierter Übergang des Nasals /m/ zum Vokal /a/ im Vergleich zum original Sprachsignal zu sehen. In Bild 5.14 ist ein resynthetisierter Übergang des Vokals /a/ zum Nasal /N/ zu sehen .

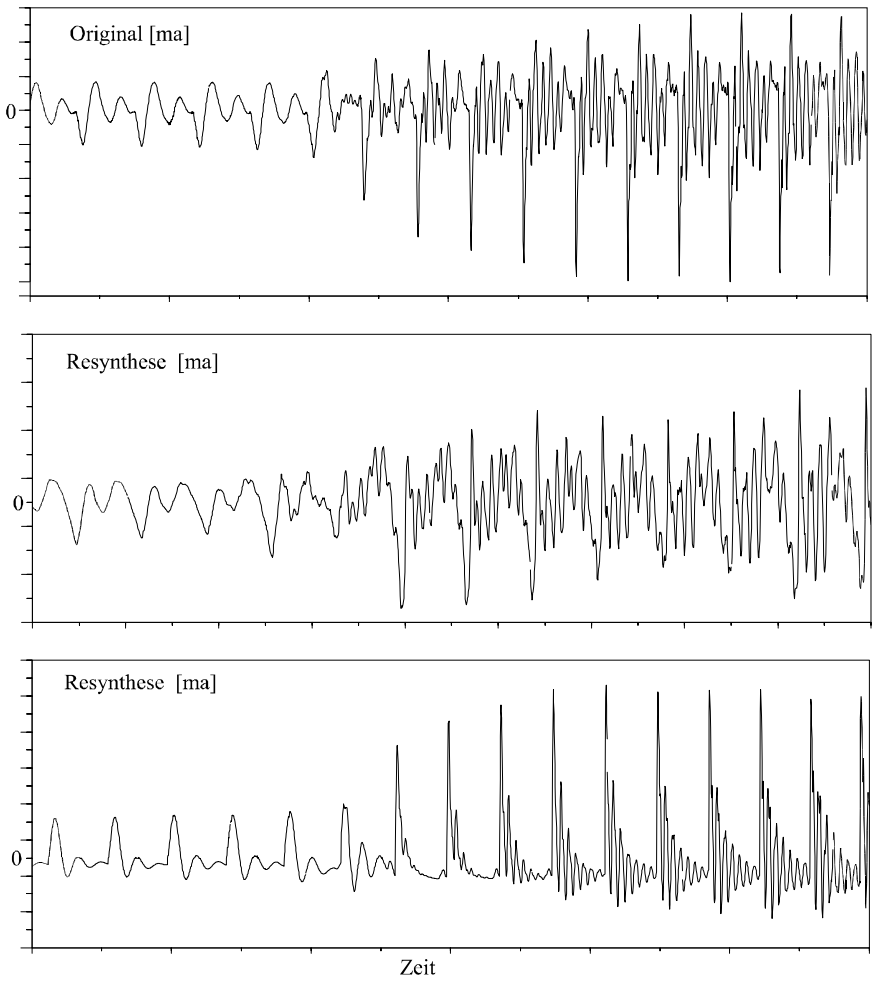


Bild 5.13: Original und resynthetisierter Lautübergang [ma]: Original Sprachsignal (oben), Resynthese mit Modellanregung der modifizierten Residualperioden (mittig), Resynthese mit Impulsfolgen-Anregung (unten).

5.2.4 Künstliche Nasalierung von unnasalierten Vokalen

Die Analyseergebnisse von Mund- und Nasensignalen nasalierter Vokale können dazu verwendet werden, Sprachsignalen unnasalierter Vokale nachträglich eine künstliche Nasalierung aufzuprägen [SnL02d, SnL03]. Dafür wird zuerst der unnasalierte Vokal mittels eines unverzweigten Rohrmodells analysiert. Als Beispiel hierfür soll die Nasalierung des unnasalierten Vokals /a/ behandelt werden, wofür die aus /a/ geschätzten

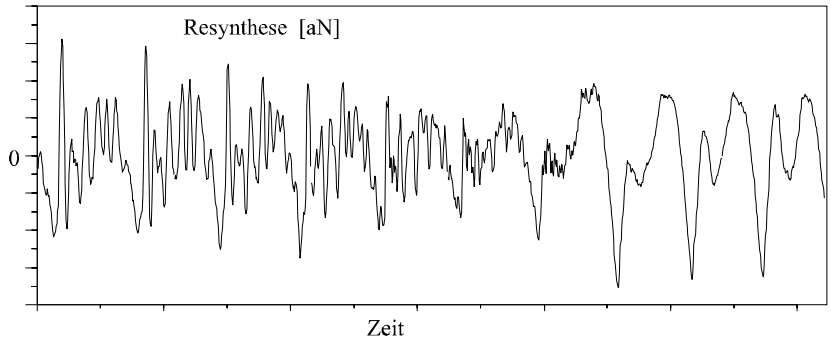


Bild 5.14: Resynthetisierter Lautübergang [aN] mit Modellanregung der modifizierten Residualperioden.

Flächen eines unverzweigten Rohrmodells verwendet werden, welche in Bild 5.15 zu sehen sind. Das Rohrmodell besitzt einen vorgegebenen reellen Lippenabschluß mit dem Wert -0,9; die Modellparameter werden mit der iterativen inversen Filterung geschätzt.

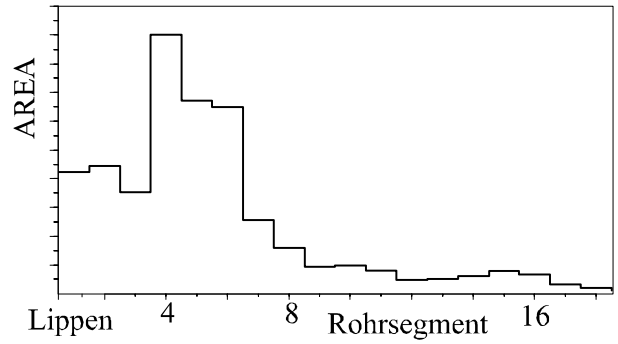


Bild 5.15: Geschätzte Vokaltraktflächen des unnasalierten Vokals /a/ mit einem unverzweigten Rohrmodell.

Für die Anregung der angestrebten nasalierten Lauterzeugung wird aus dem Sprachsignal des unnasalierten Vokals das Residualsignal x_r ermittelt. Für die Nasalierung werden die ermittelten Flächen (Nase in Bild 4.62) aus dem Mund- und Nasensignal des nasalierten Vokals /$\widetilde{\text{a}}$/ (Bild 4.61) benutzt. Diese Analysen stammen von Sprachaufnahmen desselben Sprechers. An das Rohrmodell des Vokals /a/ von Bild 5.15 wird mittels eines eingefügten Dreitores der Seitenzweig von Bild 4.62 angekoppelt, der aus dem Mundsignal des nasalierten Vokals geschätzt wurde. Diese Ankopplung berücksichtigt den Einfluß des Nasaltrakts auf das Mundsignal bei gesenktem Velum. Im Bild 5.16 oben ist der Betragsgang des unverzweigten Rohrmodells des Vokaltraktes und unten der des Modells mit Ankopplung des Seitenzweiges dargestellt, welcher den Einfluß des Nasaltrakts darstellt. In Bild 5.16 ist zu erkennen, daß die Ankopplung des Seitenzweiges eine erkennbare Nullstelle im unteren Frequenzbereich bewirkt. In [Da96b] ist mittels Sprechtraktsimulationen infolge einer Nasaltraktankopplung für den Vokal /a/ eine Nullstelle ungefähr an derselben Stelle zu sehen, was die ermittelte Übertragungsfunktion in Bild 5.16 bestätigt. Zusätzlich zu dem verzweigten Modell des Mundsignals wird für eine vollständige Nasalierung noch ein Nasensignal berücksichtigt. Für das

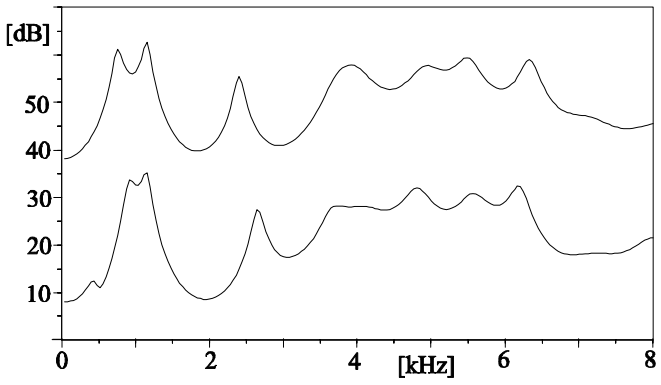


Bild 5.16: Betragsgang des unverzweigten Rohrmodells (oben), geschätzt aus dem unnasalierten Vokal /a/, und Betragsgang des Rohrmodells mit Ankopplung des Nasaltraktes (unten).

Modell des Nasensignals wird das verzweigte Rohrmodell verwendet, welches aus dem Nasensignal von /ã/ geschätzt wurde. Diese beiden verzweigten Rohrmodelle werden jeweils mit dem Residualsignal x_r des unnasalierten Vokals angeregt, so daß die Ausgänge der beiden Modelle das Mundsignal y_M und das Nasensignal y_N ergeben. Das Sprachsignal y des künstlich nasalierten Vokals wird durch eine Linearkombination der beiden Signale gebildet:

$$y = (\gamma - 1) \cdot y_M + \gamma \cdot y_N \qquad \text{mit} \quad 0 < \gamma < 1. \tag{5.6}$$

γ gibt das Mischungsverhältnis des Mund- und Nasensignals wieder. Die gesamte Vorgehensweise ist in Bild 5.18 skizziert. In Bild 5.17 sind die Betragsgänge von verschiedenen Linearkombinationen der Übertragungsfunktionen der beiden verzweigten Rohrmodelle gezeigt. Es ergibt sich dabei eine Nullstelle zwischen 500 Hz und 1 kHz.

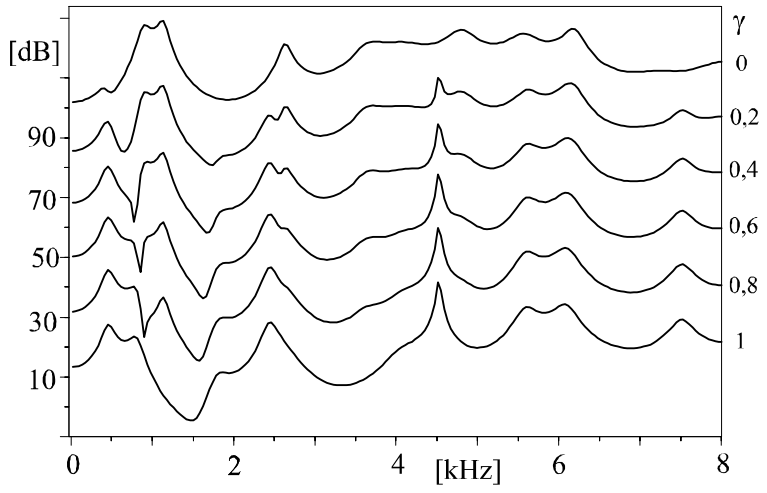


Bild 5.17: Betragsg änge von Linearkombinationen der Übertragungsfunktionen für das Mund- und Nasensignal, welche jeweils durch ein verzweigtes Rohrmodell dargestellt werden. Von oben nach unten wird der Einfluß des Nasensignals stärker.

Die resultierende Nullstelle läßt sich auch in Sprachspektren und deren Analysen von

natürlich nasalierten Vokalen beobachten, wie das Sprachspektrum von Bild 4.56 belegt, welches von Sprachaufnahmen desselben Sprechers gewonnen wurde. Diese Nullstelle ist auch in theoretischen Berechnungen von verzweigten Rohrmodellen für /ã/ in [Stev98] zu sehen. In [De93] ist eine ähnliche Nullstelle für den selben Vokal durch die Nasaltraktankopplung gezeigt. Für ein realistisches Mischungsverhältnis sollte das Mundsignal stärker gewichtet sein als das Nasensignal. Diese Verhältnisse sind durch den zweiten, dritten und eventuell vierten Betragsgang von oben in Bild 4.56 gegeben. Die synthetisierten Signale klingen durch die beschriebene Verarbeitung deutlich nasaliert (siehe Hörbeispiele). Der Grad der Nasalierung kann durch das Mischungsverhältnis verändert werden. Bei einer zu starken und unrealistischen Gewichtung des Nasensignals wird der Vokal perzeptiv nicht mehr erkannt. Neben dem Vokal /a/ wurde auch der Schwa-Laut künstlich nasaliert, bei dem die Nasalierung des zuvor unnasalierten Lautes auch eindeutig auditiv wahrzunehmen ist. Vergleicht man die Betragsgänge in Bild 5.17 mit denen des natürlich nasalierten Vokals, so sind insbesondere im unteren Frequenzbereich Ähnlichkeiten zu erkennen. Diese bezieht sich auf die Nullstelle und auf die darausfolgende Abschwächung des ersten Formanten. Durch die Anregung mittels der Residualsignale klingen die künstlich nasalierten Vokale sehr natürlich.

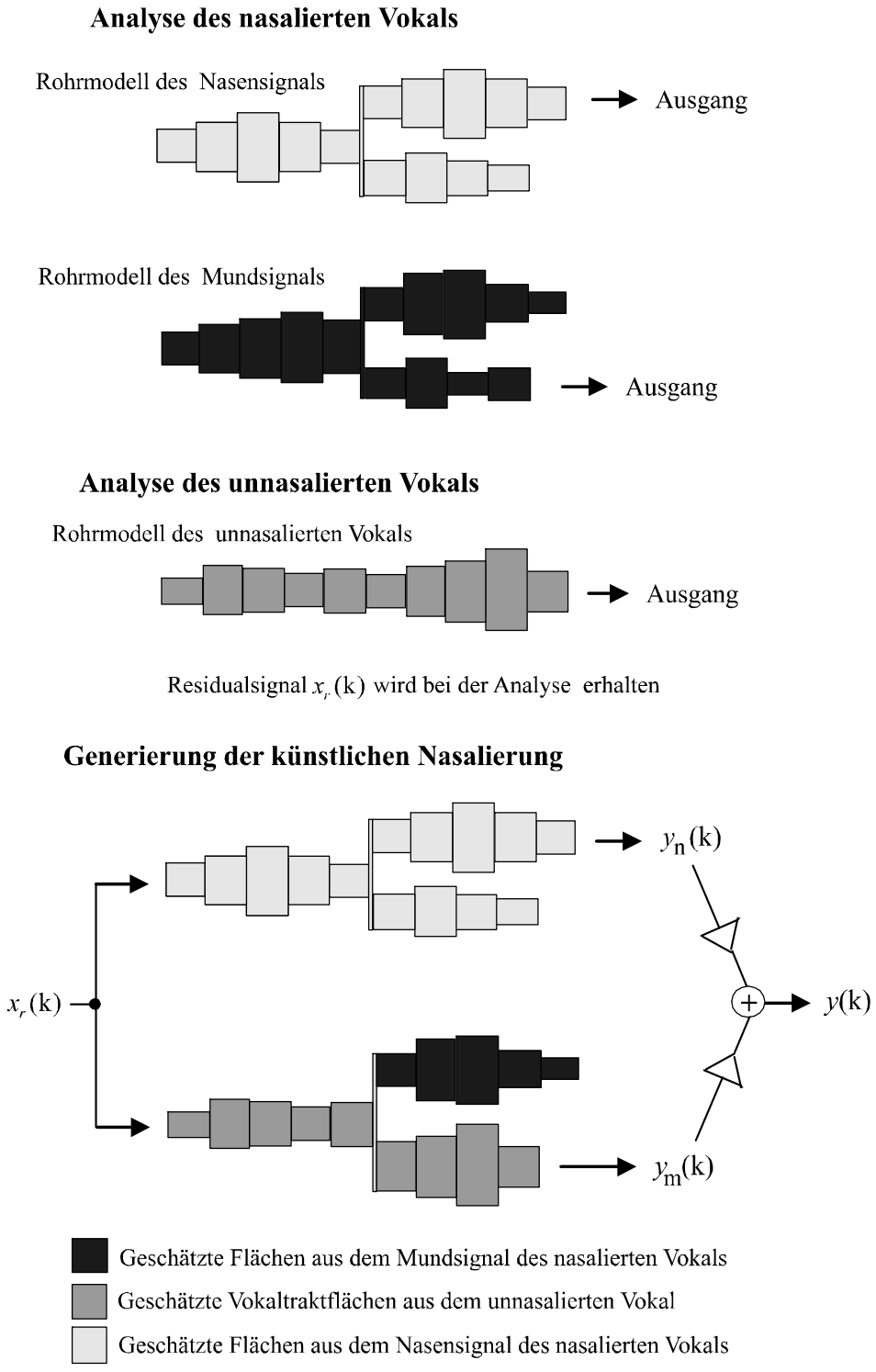


Bild 5.18: Schematische Darstellung der künstlichen Nasalierung eines unnasalierten Vokals mit verzweigten Rohrmodellen.

Kapitel 6

Zusammenfassung

In dieser Arbeit werden Verfahren für die Schätzung erweiterter Rohrmodelle zur Sprachproduktion vorgestellt, mit denen eine Vielzahl an möglichen Rohrmodellstrukturen einheitlich behandelt werden können. Dies betrifft die Darstellung der unterschiedlichen Rohrstrukturen im Z-Bereich und die Parameterbestimmung aus dem Sprachsignal. Die algorithmische Behandlung von nahezu beliebigen Rohrstrukturen im Zeit- und Frequenzbereich kann mittels einer sukzessiven Reduktion der Rohrstruktur mit Hilfe der Darstellung in Form von Betriebskettenmatrizen vollzogen werden. Die Parameterbestimmung wird durch eine Minimierung eines spektralen Abstandsmaßes mittels eines allgemeinen Optimierungsverfahrens erreicht. Hierbei hat sich die Fehlerdefinition als wichtig für eine gute Modellierung erwiesen. Neben einer Fehlerdefinition im zeitinvarianten Fall mit Hilfe des Z-Bereichs konnte die Fehlerdefinition auch auf zeitvariable Systeme übertragen werden, welche z.B. bei Verwendung eines zeitvariablen Glottisabschlusses benötigt wird. Über diese auf allgemeine lineare Systeme anwendbaren Schätzalgorithmen hinaus werden auch schnellere Algorithmen vorgestellt, die eine iterative inverse Filterung beinhalten. Das Grundprinzip bei diesen Schätzalgorithmen besteht darin, daß Formeln verwendet werden, aus denen sich suboptimale Lösungen direkt erzielen lassen. Diese schätzen eine Untermenge der Modellparameter optimal unter der Nebenbedingung, daß die restlichen Parameter bekannt sind. Die iterative inverse Filterung wird auf allgemeine Pol-Nullstellen Systeme, auf unverzweigte Rohrmodelle mit ein oder zwei Rohrabschlüssen und auf einfach verzweigte Rohrmodelle angewandt. Aus den Schätzungen der Modellparameter lassen sich die Systemfunktion und die Querschnittsflächen bestimmen, wobei sich für beide Schätzresultate nicht zwangsläufig gleich gute Ergebnisse einstellen müssen. Der Schätzalgorithmus ist in erster Linie bestrebt eine gute spektrale Approximation des Modells in Bezug auf das Sprachspektrum zu erzielen, wobei sich Modellrestriktionen durchaus negativ bemerkbar machen können. Die Resultate zeigen, daß die Schätzalgorithmen eine gute spektrale Approximation ermöglichen. Problematisch können Signale mit vielen Minima in der spektralen Einhüllenden sein. Dadurch ist es möglich, daß der Algorithmus die ungünstigen Fälle der möglichen Nullstellen schätzt. Diese Problematik betrifft allerdings weniger den Schätzalgorithmus selbst, als die Darstellung des Signals, wofür auch eine angepaßte Vorverarbeitung hinzugezählt werden muß. Die Modellierung des Sprachspektrums kann als die wesentliche Anforderung an die Parameterschätzung angesehen werden; dies ist allerdings nicht das alleinige Ziel, da eine gute spektrale Approximation des Sprachsignals durch das Modell nur dann zwangsläufig einen realistischen Flächenverlauf ergibt, falls das Modell den Sprechtrakt hinreichend gut model-

liert. Zusätzlich muß eine gewisse Eindeutigkeit bei der Schätzung gewährleistet sein. Für diesen Zweck müssen Nebenbedingungen wie Rohrabschlüsse vorgegeben sein. Da einige Parameter, wie die Abschlüsse oder auch Längen des Nasal- und Vokaltraktes, für die Schätzung vorausgesetzt werden müssen, ergeben sich bei Variationen dieser Nebenbedingungen mehrere vergleichbare Lösungen. Dadurch ist noch eine Unsicherheit vorhanden, da die Nebenbedingungen nicht exakt bekannt sind. Dies hat sich in den Untersuchungen bei verzweigten Rohrmodellen und Modellen mit Glottisabschluß gezeigt. Um dieses Auswahlproblem zu lösen, werden die Lösungen mit denjenigen Vokaltrakt- und Nasaltraktlängen gewählt, die einen kleinen Schätzfehler aufweisen und dem phonetischen Vorwissen entsprechen. Für eine Beurteilung der geschätzten Flächen muß beachtet werden, daß die Flächen des Sprechtraktes für unterschiedliche Sprachaufnahmen variieren können. Dies liegt im nicht exakt reproduzierbaren Sprachproduktionsprozeß begründet, worin die Variationen der Vokaltraktstellungen und der Anregung eingeschlossen sind. Dies wird durch Analysen von mehreren Äußerungen des gleichen Lautes bzw. der gleichen Lautketten berücksichtigt; durch sich wiederholende Schätzergebnisse ergibt sich eine gewisse Sicherheit und Konstanz in den Resultaten. Daher sind mehrere Analysen vonnöten, um repräsentative Ergebnisse zu erzielen. Die Beurteilung der geschätzten Flächen kann nur bis zu einem gewissen Grad erfolgen, da die wahren Querschnittsflächen unbekannt sind und die aus der Literatur zum Vergleich stehenden Flächen, welche durch MRT- und Röntgen-Aufnahmen ermittelt wurden, selbst von Sprecher zu Sprecher unterschiedlich sind. Eine gemeinsame Struktur der beteiligten Sprechtraktformen ist dennoch für die jeweiligen Laute zu erkennen, welche auch in den aus dem Sprachsignal geschätzten Flächen als solche zu erkennen ist. Die Analysen mit den verwendeten Rohrmodellen legen nahe, daß der Vokaltrakt sich im vorderen und mittleren Bereich besser schätzen läßt als im hinteren Rachenbereich. Dies hängt damit zusammen, daß der Vokaltrakt sich im hinteren Rachenbereich durch die birnenförmige Vertiefungen am Kehlkopf (Recessus piriformis / piriform fossa) und insbesondere durch die zeitvariable Glottis komplizierter gestaltet. Für die Schätzung von Nasalen und nasalierten Lauten wird die an den Vokaltrakt angekoppelte Nasenstruktur durch ein unverzweigtes Rohr realisiert, wodurch die Nasennebenhöhlen nicht unmittelbar durch die Rohrstruktur modelliert werden können. Trotz der Einfachheit der Nasenstruktur ergeben sich verhältnismäßig gute Ergebnisse. Die durchgeführten Analysen der mehrfach verzweigten Nasenstruktur belegen, daß auch komplizierte Nasaltrakt-Strukturen mit einer guten spektralen Modellierung erzielt werden können. Neben den Schätzalgorithmen, die nur einen Systemausgang modellieren, sind auch Parameterschätzungen möglich, mit denen zwei Systemausgänge gleichzeitig modelliert werden. Diese finden für eine Modellierung von Mund- und Nasensignal bei nasalierten Lauten sowie bei einer Nasaltraktmodellierung unter Berücksichtigung der beiden Nasengänge eine Verwendung.

Neben der Analyse von Rohrmodellen werden auch Ansätze für eine Verwendung der geschätzten Flächen für die Spracherzeugung diskutiert; damit ist es möglich, die Ergebnisse auditiv zu überprüfen. Hierfür werden insbesondere Vokal-Explosiv-Vokal Übergänge behandelt, da sich diese in früheren Spachsyntheseexperimenten [Sn96, Ei96, Ge96] als kritisch erwiesen haben. Die Übergänge werden mit nur drei Flächensätze realisiert, welche jeweils die beteiligten Laute repräsentieren. Die verwendeten Vokaltraktflächen aus der Schätzung der Explosive weisen eine Konstriktion an der den Konsonanten charakterisierenden Vokaltraktstelle auf. Trotz der verwendeten geringen

Datenbasis können durch einen gewichteten Flächenübergang verständliche Lautketten produziert werden, wobei dies nur für eine Teilmenge der möglichen Lautkombinationen von Vokal-Explosiv-Vokal Übergängen untersucht wurde. Hier könnte unter Umständen eine abgeänderte Flächeninterpolation die Ergebnisse noch verbessern. Daß durch die Verwendung von Modellen, welche die Sprechtraktstruktur und -geometrie wiedergeben, auch Lauttransformationen ermöglicht werden, zeigt daß Beispiel der künstlichen Nasalierung von zuvor unnasalierten Vokalen, bei der die Analyseergebnisse der getrennten Mund- und Nasensignale benutzt werden. Dadurch ist eine modellbasierte Realisierung von Koartikulationseffekten gegeben. Für die Spracherzeugung hat neben dem Sprechtraktmodell auch die Anregung einen Einfluß auf die Sprachqualität. Für die stimmhafte Anregung werden wenige benachbarte Residualperioden verwendet, wobei die Perioden noch vorverarbeitet worden sind. Dadurch lassen sich weitgehend natürlich klingende Sprachsignale erzeugen. Diese klingen besser als bei einer Anregung mit einer Impulsfolge, da die natürlichen Fluktuationen der Anregung berücksichtigt werden.

Für die Analyse von Sprachproduktionsmodellen bzw. Pol-Nullstellen-Systemen ergibt sich durch die vorgestellte Arbeit eine erfolgreiche Erweiterung der inversen Filterung für die Schätzung von erweiterten Rohrmodellen. Neben einem auf nahezu beliebige Rohrmodellstrukturen anwendbaren Ansatz wird auch ein effizienter Ansatz durch die iterative inverse Filterung diskutiert. Die vorgestellten Schätzverfahren haben sich größtenteils als äußert erfolgreich erwiesen und ermöglichen spektrale Approximationen von Sprachspektren durch erweiterte Rohrmodelle, wie sie vergleichbar nur von der linearen Prädiktion, allerdings für Nur-Pole-Modelle, erzielt werden. Die für die Parameterbestimmung von linearen Systemen gewonnenen Erkenntnisse können durch ihre grundlegende Betrachtung über die Sprachverarbeitung hinaus auch für andere Gebiete etwa der Systemidentifikation verwendet werden. Für die Spracherzeugung ergibt sich mit Hilfe der diskutierten Verfahren ein möglicher Mittelweg zwischen der rein artikulatorischen Spracherzeugung, in der zwar die Anatomie der Sprachproduktion gut wiedergegeben wird, allerdings oft mit Problemen in der spektralen Modellierung zu rechnen ist, und den konkatenativen Spracherzeugungsansätzen, in denen Sprachsignalabschnitte direkt oder parametrisiert verwendet werden, wodurch nur ein geringer Bezug zum Sprachproduktionsprozeß vorhanden ist.

Kapitel 7

Anhang

7.1 SAMPA-Notation der Lautschrift

Lange Vokale

Symbol	Beispielwort	Transkription
i:	Lied	li:t
e:	Beet	be:t
E:	spät	SpE:t
a:	Tat	ta:t
o:	rot	ro:t
u:	Blut	blu:t
y:	süß	zy:s
2:	blöd	bl2:t

kurze Vokale
(einige kurze Vokale ergeben sich durch Weglassen des Doppelpunktes)

Symbol	Beispielwort	Transkription
I	Sitz	zIts
E	Gesetz	g@zEts
a	Satz	zats
O	Trotz	trOts
U	Schutz	SUts
Y	hübsch	hYpS
9	plötzlich	pl9tslIC
@	bitte	bIt@

(Nasalierung wird mit ~ dargestellt)

Diphthonge

Symbol	Beispielwort	Transkription
aI	Eis	aIs
aU	Haus	haUs
OY	Kreuz	krOYts

Explosive

Symbol	Beispielwort	Transkription
p	Pein	paIn
b	Bein	baIn
t	Teich	taIC
d	Deich	daIC
k	Kunst	kUnst
g	Gunst	gUnst

Frikative und Gleitlaute

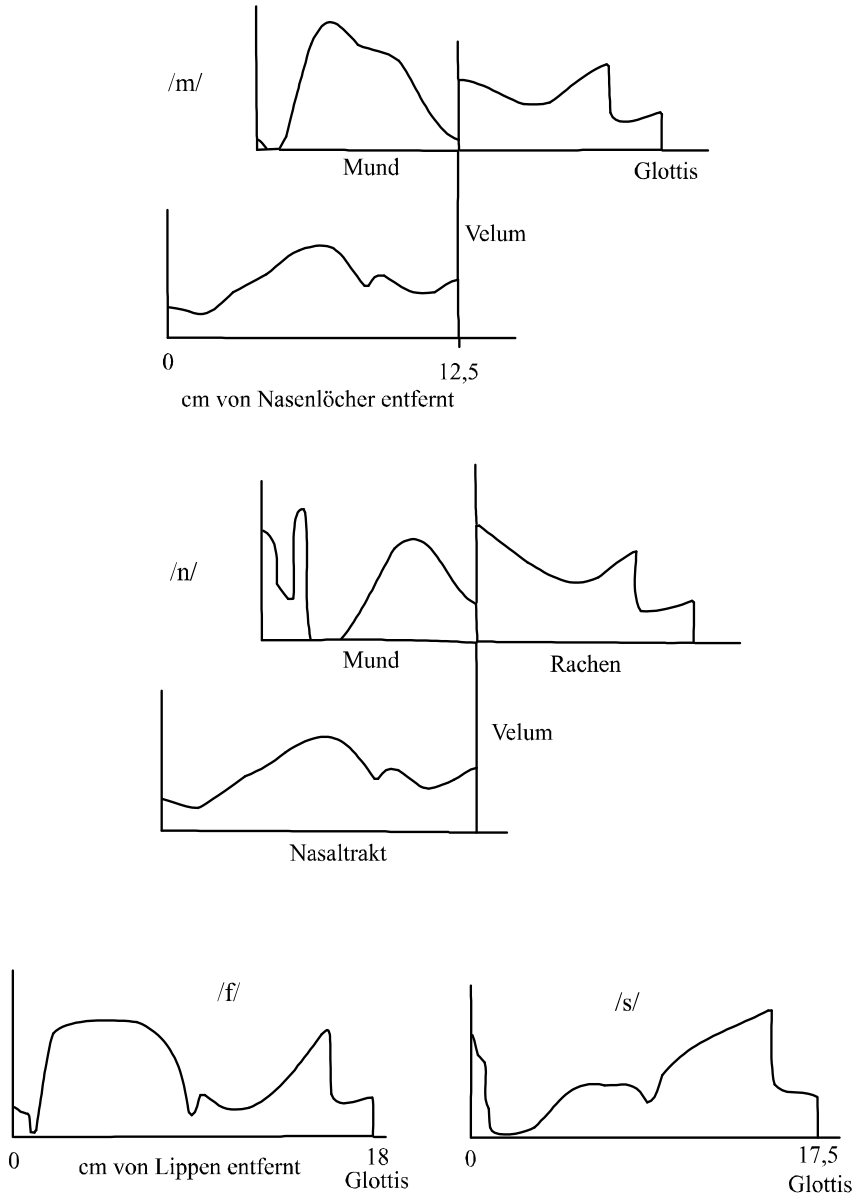
Symbol	Beispielwort	Transkription
f	fast	fast
v	was	vas
s	Tasse	tas@
z	Hase	ha:z@
S	waschen	vaSn
Z	Genie	Zeni:
C	sicher	zIC6
j	Jahr	ja:6
x	Buch	bu:x
h	Hand	hant

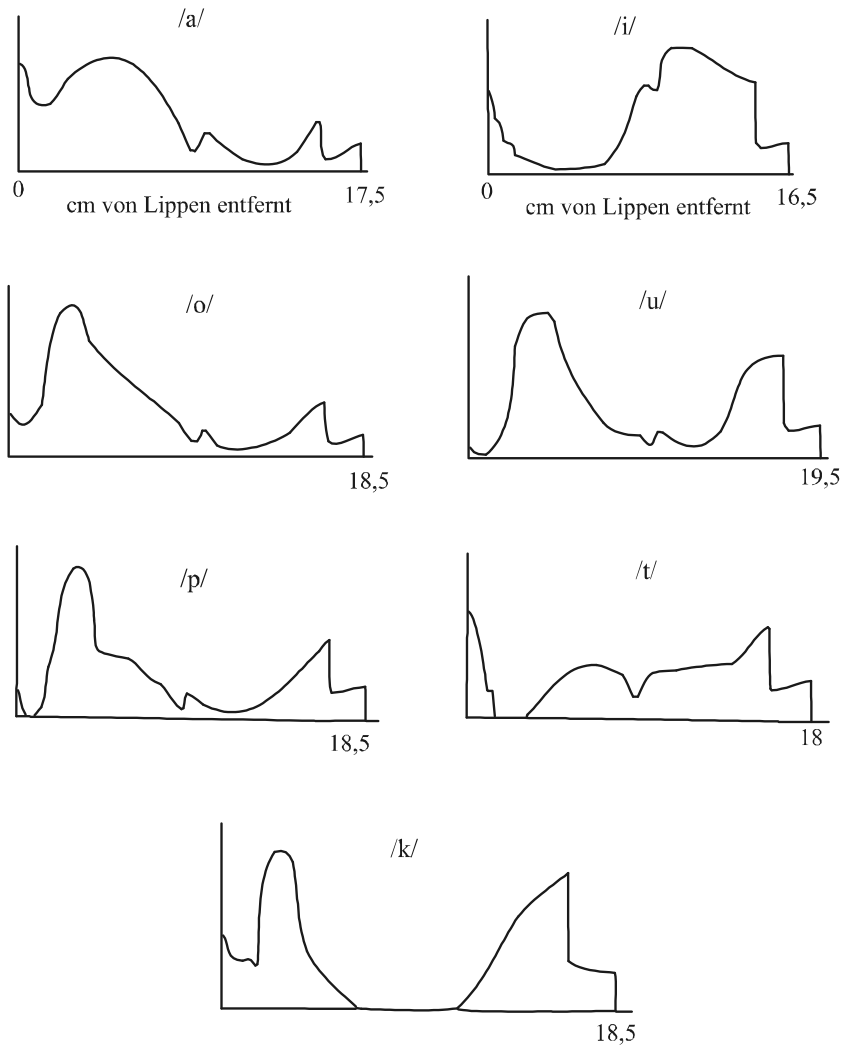
Nasale und Lateral

Symbol	Beispielwort	Transkription
m	mein	maIn
n	nein	naIn
N	Ding	dIN
l	Leim	laIm

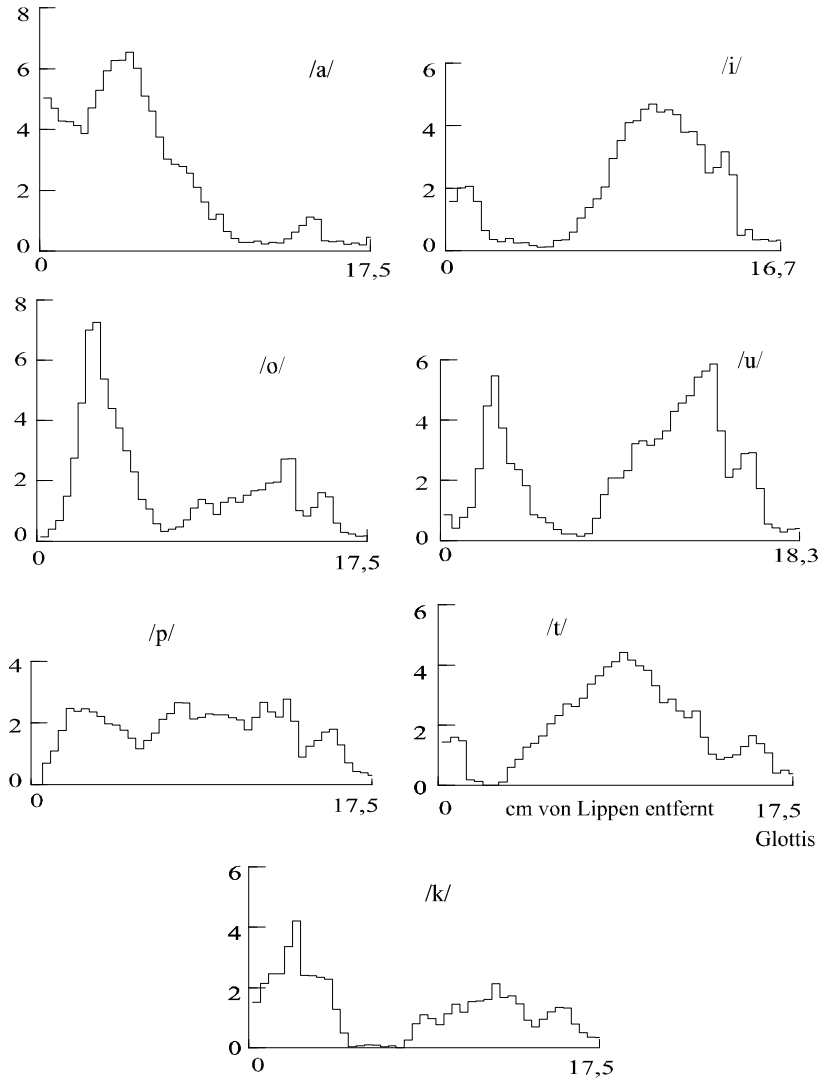
7.2 Querschnittsflächen aus der Literatur

Ermittelte Querschnittsflächen des Sprechtraktes aus Röntgenaufnahmen nach Fant [Fa70].




/a/
0
cm von Lippen entfernt
17,5
/i/
0
cm von Lippen entfernt
16,5
/o/
18,5
/u/
19,5
/p/
18,5
/t/
18
/k/
18,5

Ermittelte Querschnittsflächen des Vokaltraktes aus MRT-Aufnahmen nach Story [St96].



7.3 Hörbeispiele

Hörbeispiele befinden sich auf der CD im wav-Format.

1. Äußerungen von getrennten Mund- und Nasensignalen.
2. Synthetisierte VCV-Sequenzen mit stimmhaften und stimmlosen Explosiven.
3. Resynthesebeispiele.
4. Künstliche Nasalierung des Vokals /a/.

Kapitel 8

Literatur

[Al85] Allen J.; Strong W.J.: „A model for the synthesis of natural sounding vowels", J.Acoust.Soc.Am. 1985 (78) pp. 58-69.

[At78] Atal B.S.; Chang J.J.; Mathews M.V. Tukey J.W.: „Inversion of articulatory-to-acoustic transformation in the vocal tract by a computer-sorting technique", J.Acoust.Soc.Am. 1978 (63) pp. 1535-1555.

[Av01] Avanzini F.; Alku P.; Karjalainen M.: „One-delayed-mass model for efficient synthesis of glottal flow", Proc. EUROSPEECH-2001, Aalborg Danmark, Vol I. pp. 51-54 , 2001.

[Ba94] Bavegard M.; Fant G.: „Notes on glottal source interaction ripple", STL-QPSR (4) 1994, Royal Inst. of Technology Stockholm, pp. 63-77.

[Bi86] Bickley C.A.; Stevens K.N.: „Effects of a vocal-tract constriction on the glottal source: experimental and modelling studies", Journal of Phonetics (14) 1986, pp. 373-382.

[BSnL02] Bettinelli M.; Schnell K.; Lacroix, A.: „Separate Messung und Analyse von Mund- und Nasensignalen bei natürlicher Sprache" , Studientexte zur Sprachkommunikation: Band 24, 13. Konferenz ESSV-2002, Dresden, S. 237-244, 2002.

[Bu68] Burg, J.: „A new Analysis Technique for Time Series Data", NATO Advanced Study Inst. on Signal Processing, Enschede, 1968.

[Br00] den Brinker A.C.; Oomen A.W.J.: „Fast ARMA modelling of power spectral density functions", Proc. X. European Signal Processing Conference EUSIPCO-2000, Tampere, pp. 1229-1232.

[Bro99] Broersen P.M.T.: „Accurate ARMA Models with Durbin's Second Method", Proc. ICASSP'99, 1999.

[Che97] Chen M.Y.: „Acoustic correlates of English and French nasalized vowels", J.Acoust.Soc.Am. 1997 (102) pp. 2360-2370.

[Chi91] Childers D.G.; Lee C.K.: „Vocal quality factors: Analysis, synthesis, and perception", J.Acoust.Soc.Am. 1991 (90) pp. 2394-2410.

[Chi94] Childers D.G.; Hu T.H.: „Speech Synthesis by Glottal Excited linear prediction", J.Acoust.Soc.Am. 1994 (96) pp. 2026-2036.

[Da94] Dang, J.; Honda K.; Suzuki H.: „Morphological and Acoustical Analysis of the Nasal and the Paranasal Cavities", J.Acoust.Soc.Am., Vol. 96, No. 4, pp. 2088-2100, October 1994.

[Da96a] Dang J.; Honda K.: „Local and global effects of the piriform fossa on speech spectra", 130th Meeting: ASA, contr. paper 3aSC8, J.Acoust.Soc.Am. (98) No. 5, (abstract) pp. 2931, 1996.

[Da96b] Dang J.; Honda K.: „An Improved Vocal Tract Model of Vowel Production Implementing Piriform Resonance and Transvelar Nasal Coupling", Proc. ICSLP'96, pp. 965-968, 1996.

[Da98] Dang J.;Shadle C.H.; Kawanishi Y.; Honda K.;Suzuki H.: „An experimental study of the open end correction coefficient for side branches within an acoustic tube", J.Acoust.Soc.Am. 1998 (104) pp. 1075-1084.

[Da99] Dang J.; Honda K.: „Speech synthesis of VCV sequence using a physioligal articulatory model", Joint Meeting: ASA/EAA/DEGA, Berlin 1999, CD-ROM; contr. paper 2pSCa2, J.Acoust.Soc.Am. (105) No.2, (abstract) pp. 1091, 1999.

[De93] Deller J. R.; Proakis J. G.; Hansen J. H. L.: „Discrete-Time Processing of Speech Signals, Macmillan Publishing Company", New York 1993, (p. 138).

[Di96] Ding W.; Hideki Kasuya H.: „A Novel Approach to the Estimation of Voice Source and Vocal Tract Parameters from Speech Signals", Proc. ICSLP'96, pp. 1257-1260, 1996.

[Du97] Dutoit T.: „An Introduction to Text-to-Speech Synthesis", Kluwer Academic Publishers, Dordrecht/Boston/London, 1997.

[Ei96] Eichler M.: „Zeitvariable Rohrmodelle für eine artikulatorisch parametrisierte Sprachsynthese", Diplomarbeit, Johann Wolfgang Goethe-Universität, Frankfurt am Main 1996.

[EiL96] Eichler M.; Lacroix, A.: „Schallausbreitung in zeitvariablen Rohrsystemen", Tagungsband DAGA'96, Bonn, S. 506-507, 1996.

[En99] Engwall O.: „Modeling of the Vocal Tract in Three Dimensions", Proc. EUROSPEECH'99, Budapest 1999.

[En01a] Engwall O.: „Using Linguopalatal Contact Patterns to Tune a 3D Tongue Model", Proc. EUROSPEECH'01, pp. 1475-1478 ,Aalborg Dänemark 2001.

[En01b] Engwall O.: „Synthesizing static vowels and Dynamic sounds using a 3D vocal tract model", Proc. 4th ISCA Tutorial and Research Workshop on Speech Synthesis", September, 2001.

[Fa70] Fant G.: „Acoustic Theory of Speech Production, Mouton", The Hague - Paris, Second Edition 1970.

[Fa86a] Fant G.; Liljencrants J.; Lin Q.G.: „A four parameter model of glottal flow", STL-QPSR 4, pp. 1-13.

[Fa86b] Fant G.: „Glottal Flow: Models and interaction", Journal of Phonetics, (14), pp. 393-399.

[Fa97] Fant G.; Bavegard M.: „Parametric model of VT area functions: vowels and consonants", TMH-QPSR 1/1997.

[Fe96] Feng G.; Castelli E.: „Some acoustic features of nasal and nasalized vowels: A target for vowel nasalization", J.Acoust.Soc.Am., Vol. 99, No. 6, pp. 3694-3706, June 1996.

[Fl72] Flanagan J. L.: „Speech Analysis, Synthesis, and Perception", 2nd ed., Berlin-Heidelberg New York 1972.

[Fra86] Frank W.; Lacroix A.: „Multi Tube Models for Speech Synthesis", Proc. EUSIPCO-86, Hague Netherlands 1986, pp. 373-376.

[Fr94] Fröhlich M.; Warneboldt H.; Strube H.W.: „Berücksichtigung des subglottalen Systems bei der artikulatorischen Sprachsynthese", Tagungsband DAGA'94, 1994, S. 1305-1308.

[Fr95a] Fröhlich M.; Strube H.W.: „Schätzung von artikulatorischen Parametern durch einen genetischen Algorithmus", Tagungsband DAGA'95, 1995, S. 1203-1206.

[Fr95b] Fröhlich M.: „Erweiterung und akustische Optimierung eines artikulatorischen Sprachsynthesesystems", Diplomarbeit, Drittes Physikalisches Institut, Georg-August-Universität Göttingen, 1995.

[Ga77] Gay T.: "Articulatory movements in VCV sequences", J.Acoust.Soc. Am. (62), pp. 183-193, 1977.

[Ge96] Gerland C.: „Zur Gewinnung der Steuerparameter aus fließender Sprache für ein Sprachsynthesesystem auf der Basis von Rohrmodellen", Diplomarbeit, Johann Wolfgang Goethe-Universität, Frankfurt am Main 1996.

[GeL96] Gerland C.; Lacroix, A.: „Estimation of Vocal Tract Areas from Fluently Spoken Text", ITG-Fachbericht 139/ Sprachkommunikation, VDE-Verlag, ITG-Fachtagung Sprachkommunikation, Frankfurt/Main, S. 113-116, 1996.

[Go99] Gomez C.: „Engineering and Scientific Computing with Scilab", Birkhäuser, Boston Basel Berlin,1999.

[Gr82] Greisbach R.: „Ermittlung der Querschnittsfläche und des midsagittalen Durchmessers des Sprechtrakts sowie deren funktioneller Zusammenhang im Bereich von Gaumen und Rachen aus Röntgen-Computer-Tomogrammen", 1982 IP-Köln Bericht Nr. 12, S.43-60.

[Gu93] Gupta S.K.; Schroeter J.: „Pitch-synchronous frame-by-frame and segment-based articulatory analysis by synthesis", J.Acoust.Soc.Am. 1993 (94) pp. 2517-2530.

[Ha96] Hayes M.H.: „Statistical Digital Signal Processing and Modelling", John Wiley & Sons, Inc, New York, 1996.

[He99] Henn H.; Sinambari G.R.; Fallen M.: „Ingenieurakustik", 2. Auflage, Vieweg-Verlag, Braunschweig/Wiesbaden, 1999.

[Ii89] Iijima H.; N. Miki.; Nagai N.: „Fundamental Consideration of Finite Element Method for the Simulation of the Vibration of Vocal Cords", Proc. ICASSP'89 1998 Glsgow (Scotland), Vol I, pp. 246-249.

[Is72] Ishizika K.; Flanagan J.L.: „Synthesis of voiced sounds from a two-mass model of the vocal cords", Bell Systems Tech. J., 51 (6), pp. 1233-1268, 1972.

[Jo83] Johansson C.; Sundberg J.; Wilbrand H.: „From sagittal distance to area", STL-QPSR (4), Royal Inst. of Technology Stockholm, pp. 39-49.

[Kel62] Kelly J.L.; Lochbaum C.C.: „Speech Synthesis", Proc. Int. Congres on Acoustics, Paper G 42, pp. 1-4, 1962, reprinted in: „Speech Synthesis", Flanagan J.L.; Rabiner L.R. (Editors), Dowden, Hutchinson and Ross, Stoudsburg, pp. 127-130.

[Ke77] Kent R.D.; Minifie F.D.: „Coarticulation in recent speech production models", Journal of Phonetics (4) 1977, pp. 115-133.

[Ko02] Kounoudes A.; Naylor P.A.; Brookes M.: „Automatic Epoch Extraction for Closed-Phase Analysis of Speech", 14th Int. Conf. on Digital Signal Processing, DSP-2002, Santorini 2002 (Greece), Vol. II, pp. 979-984.

[Kr91] Kröger B.J.: „Zur Auswirkung der Glottis-Sprechtrakt-Kopplung auf die Stimmreinheit", Sprache-Stimme-Gehör Zeitschrift für Kommunikationsstörungen. Heft 4, 15. Jahrgang, Dez. 1991, S. 127-178.

[Kr92] Kröger B.J.: „Minimal Rules for Articulatory Speech Synthesis", Proc. EUSIPCO-92, Brussels Belgium 1992, pp. 331-334.

[Kr98] Kröger B. J.: „Ein phonetisches Modell der Sprachproduktion", Linguistische Arbeiten 387, Max Niemeyer Verlag Tübingen, 1998.

[Kub85] Kubin, G.: „Wave Digital Filters: Voltage, Current, or Power Waves?", Proc. Int. Conf. Acoust., Speech, and Sig. Processing ICASSP'85, Tampa, pp. 69-72, 1985.

[Kum82] Kumaresan R; Tufts D.W.: „Estimation the Parameters of Exponentially Damped Sinusoids and Pole-Zero Modeling in Noise", IEEE Trans. ASSP-30 (1982), pp. 833-840 .

[L78] Lacroix A.: „Source Coding of Speech Signals by Improved Modeling of the Voice Source", NTG Int. Conf. Information and System Theory in Digital Communications, Berlin 1978, NTG-Fachberichte Band 65, VDE-Verlag Berlin, pp. 103-108.

[L79] Lacroix A.; Makai B.: „A Novel Vocoder Concept Based on Discrete Time Acoustic Tubes", Proc. Int. Conf. Acoust., Speech, and Signal Processing, ICASSP'79, Washington D.C. 1979, pp. 73-76.

[L96] Lacroix, A.: „Digitale Filter", 4. Aufl. Oldenbourg Verlag München Wien 1996.

[Ladd96] Ladd R.S.: „C++ Templates and Tools", sec. ed. M&T Books/MIS: Press, Inc., New York, 1996.

[Lade96] Ladefoged P.; Maddieson I.: „The Sounds of the World's Languages", Blackwell Publishers, Oxford Uk & Cambridge USA, 1996.

[Lai82] Laine, U.K.: „Modeling of lip radiation impedance in the z-domain", Proc. Int. Conf. Acoust., Speech, and Sig. Processing ICASSP'82, Paris, pp. 1992-1995.

[Le83] Levinson S.E.; Schmidt C.E.: „Adaptive Computation of articulatory parameters from the speech signal", J.Acoust.Soc.Am. 1983 (74) pp. 1145-1154.

[Li85] Liljencrants J.: „Speech Synthesis with a Reflection-Type Line Analog", Dissertation, Royal Institute of Technology, Stockholm 1985.

[Li91] Liljencrants J.: „Numerical simulations of glottal flow", Proc. EUROSPEECH-91, 1991 Genova (Italy), Vol. I, pp. 255-258.

[Lim96] Lim I.T.; Lee B.G.: „Lossy Pole-Zero Modeling of Speech Signals", IEEE Trans. Speech and Audio Processing, Vol. 4, No. 2, pp. 81-88, March 1996.

[Lin76] Lindgvist-Gauffin J.; Sundberg J.: „Acoustic properties of the nasal tract", Phonetica (33) 1967, pp. 161-168.

[LiuL96a] Liu M.; Lacroix A.: „Improved Vocal Tract Model for Speech Synthesis", Proc. Eigth European Signal Processing Conference, EUSIPCO'96, Trieste, Italy, pp. 1055-1058, 1996.

[LiuL96b] Liu M.; Lacroix A.: „Improved Vocal Tract Model for the Analysis of Nasal Speech Sounds", Proc. Int. Conf. Acoust., Speech, and Sig. Processing ICASSP'96, Atlanta, USA, pp. 801-804, 1996.

[Liu98] Liu M.: „Zeitdiskrete Modelle für den Stimmtrakt auf der Basis akustischer Rohrsysteme", Dissertation, Johann Wolfgang Goethe-Universität, Frankfurt am Main 1998.

[Lö99] Löfqvist A.; Gracco V.L.: „Interarticulator programming in VCV sequences: Lip and tongue movements", J.Acoust.Soc.Am. 1999 (105) pp. 1864-1876.

[Ma82] Maeda S.: „The Role of Sinus Cavities in the Production of Nasal Vowels", Proc. ICASSP-82, Paris, pp. 911-914, 1982.

[Mer67] Mermelstein P.: „Determination of the Vocal-Tract Shape from Measured Formant Frequencies", J.Acoust.Soc.Am. 1967 (41) pp. 1283-1294.

[Mey89] Meyer P.; Wilhelms R.; Strube H.W.: „A quasiarticulatory speech synthesizer for German language running in realtime", J.Acoust.Soc.Am. 1989 (86), pp. 523-539.

[MG76] Markel J.D.; Gray A.H.: „Linear Prediction of Speech", Springer-Verlag: Berlin - Heidelberg - New York 1976.

[Na95] Narayanan S.; Alwan A.; Haker K.: „An articulatory study of fricative consonants using magnetic resonance imaging", J.Acoust.Soc.Am. 1995 (98) pp. 1325-1347.

[Oh66] Ohman S.E.G.: „Coarticulation in VCV utterances: Spectographic measurements", J.Acoust.Soc.Am. 1966 (39) pp. 151-168.

[Ol93] Oliveira L. C.: „Estimation of Source Parameters by Frequency Analysis", Proc. Eurospeech, Berlin, Vol 1, 99-102, 1993.

[Pa91] Papoulis A.: „Probability, Random, Variables, and Stochastic Processes", McGraw-Hill, New York, 3rd Edition 1991.

[Pas99] Pascal P.; Payan Y.; Perkell J.; Zandipour M.; Matthies M.: „VCV synthesis from muscle commands", Joint Meeting: ASA/EAA/DEGA, Berlin 1999; contr. paper 2pSCa1, J.Acoust.Soc.Am. (105) No.2, (only abstract) pp. 1091, 1999.

[Qi00] Qi Y.; Hillman R. E.: „Temporal and spectral estimations of harmonics-to-noise ratio in human voice signals", J.Acoust.Soc.Am. 1997 (102) pp. 537-543.

[Rah89] Rahim M.G.; Goodyear C.C.: „Articulatory Synthesis with the Aid of a Neural Net", ICASSP'89 1998 Glasgow (Scotland), Vol I, pp. 227-230.

[Rah93] Rahim M.G.; Goodyear C.C.; Kleijn W.B.; Schroeter J.; Sondhi M.M.: „On the use of neural networks in articulatory speech synthesis", J.Acoust.Soc.Am. 1993 (93) pp. 1109-1121.

[Ra99] Ranostaj F.: „Entwicklung neuer Methoden zur Modellierung des Nasaltraktes in Sprachsynthesesystemen", Diplomarbeit, Johann Wolfgang Goethe-Universität, Frankfurt am Main 1999.

[RaSnL99] Ranostaj F.; Schnell K.; Lacroix A.: „Modellierung des Nasaltrakts", Studientexte zur Sprachkommunikation: Band 16, Zehnte Konferenz ESSV'99 , Görlitz, S. 58-63, 1999.

[RaL00] Ranostaj F.; Lacroix A.: „Bestimmung des Übertragungsverhaltens des Nasaltrakts aus computertomographischen Daten", ITG-Fachbericht 161/ Sprachkommunikation, VDE-Verlag, ITG-Fachtagung Sprachkommunikation / KONVENS, Ilmenau 2000, S. 131-134, 2000.

[RaL02] Ranostaj F.; Lacroix A.: „Messung und Simulation der Akustik des Nasaltraktes", Tagungsband ESSV-2002 Dresden, Studientexte zur Sprachkommunikation Band 24, S. 245-252, 2002.

[Ro73] Rothenberg M.: „A new inverse-filtering technique for deriving the glottal airflow waveform during voicing", J.Acoust.Soc.Am. 1973 (53) pp.1632-1645.

[Sa78] Sambur M.R.; Rosenberg A.E.; Rabiner L.R.; McGonegal C.A.: „On reducing the buzz in LPC synthesis", J.Acoust.Soc.Am. (63), pp. 918-924, 1978.

[Schr94] Schroeter J.; Sondhi M.M.: „Techniques for Estimating Vocal-Tract Shapes from Speech Signal", IEEE Trans. on Speech and Audio Processing, Vol. 2, No. 1, pp. 233-150, 1994.

[Schü94] Schüßler H.W.: „Digitale Signalverarbeitung 1", 4. Auflage, Springer-Verlag Berlin Heidelberg New York, 1994.

[Sh91] Shadle C.H.: „The effect of geometry on source mechanisms of fricative consonants", Journal of Phonetics (19) 1991, pp 409-424.

[Sn96] Schnell K.: „Sprachsynthese mit erweiterten Rohrmodellen", Diplomarbeit, Johann Wolfgang Goethe-Universität, Frankfurt am Main 1996.

[SnL97] Schnell K.; Lacroix A.: „Parameterbestimmung für erweiterte Rohrmodelle zur Sprachsynthese", Tagungsband DAGA 1997, pp 541-542.

[SnL98a] Schnell K.; Lacroix A.: „Erweiterte Rohrmodelle für die Sprachproduktion", Tagungsband DAGA'98, Zürich, pp 384-385, 1998.

[SnL98b] Schnell K.; Lacroix A.: „Parameterbestimmung von Rohrmodellen für die Spracherzeugung" , ITG Fachbericht 152 / Sprachkommunikation, VDE-Verlag, 5.ITG-Fachtagung Sprachkommunikation und 9. Konferenz ESSV, Dresden, S. 101-104, 1998.

[SnL99a] Schnell K.; Lacroix A.: „Parameter Estimation from Speech Signals for Tube Models", Joint Meeting: ASA/EAA/DEGA, Berlin 1999, CD-ROM; contr. paper 2pSCa4, J.Acoust.Soc.Am. (105) No.2, (abstract) pp. 1091, 1999.

[SnL99b] Schnell K.; Lacroix A.: „Parameter Estimation for Tube Models with Time Dependent Glottis Impedance", Proc. of the second EURASIP Conference ECMCS'99, Krakow, CD-ROM, 1999.

[SnL99c] Schnell K.; Lacroix A.: „Parameterbestimmung für Pol-Nullstellen-Modelle", Studientexte zur Sprachkommunikation: Band 16, Zehnte Konferenz ESSV'99 , Görlitz, S. 64-71, 1999.

[SnL00a] Schnell K.; Lacroix A.: „Akustische Rohrsysteme mit Abzweigungen, Verzweigungen und Kopplungen", Tagungsband DAGA-2000, Oldenburg, S. 358-359.

[SnL00b] Schnell K.; Lacroix A.: „Bestimmung von Rohrmodellparametern aus Sprachsignalen", Tagungsband DAGA-2000, Oldenburg, S. 374-375.

[SnL00c] Schnell K.; Lacroix A.: „Realisation of a Vowel-Plosive-Vowel Transition by a tube model", Proc. X. European Signal Processing Conference, EUSIPCO-2000, Tampere, Finland, pp. 757-760, 2000.

[SnL00d] Schnell K.; Lacroix A.: „Parameter Estimation for Branched Tube Systems", ITG-Fachbericht 161, VDE-Verlag, Konvens 2000/ ITG-Fachtagung Sprachkommunikation, S. 127-130, 2000.

[SnL00e] Schnell K.; Lacroix A.: „Vokaltrakt-Schätzung unter Berücksichtigung einer reellen Glottisimpedanz", ITG-Fachbericht 161, VDE-Verlag, Konvens 2000/ ITG-Fachtagung Sprachkommunikation, S. 279-284, 2000.

[SnL00f] Schnell K.; Lacroix A.: „Parameter Estimation for Time Variable Tube Models from Speech Signals " in: Papers in Phonetics and Speech Processing, Forum Phoneticum 70 (2000), main Editors: Palkova Z, Wodarz H.-W., Hector Frankfurt am Main, pp. 137-148, 2000.

[SnL01a] Schnell K.; Lacroix A.: „Analyse und Verwendung des Rohrmodells für die Spracherzeugung", Tagungsband, DAGA 2001, S. 566-567.

[SnL01b] Schnell K.; Lacroix A.: „Pole Zero Estimation from Speech Signals by an Iterative Procedure", Proc. Int. Conf. Acoust., Speech, and Sig. Processing ICASSP-2001, Salt Lake City, USA, pp. 109-112, 2001.

[SnL01c] Schnell K.; Lacroix A.: „Inverse Filtering of Tube Models with Frequency Dependent Tube Terminations", Proc. EUROSPEECH-2001, Aalborg Denmark, pp. 2467-2470, 2001.

[SnL01d] Schnell K.; Lacroix A.: „Resynthese von Sprachsignalen mit Kettenfiltern durch periodensynchrone Analyse und lautunabhängiger Anregung", Studientexte zur Sprachkommunikation: Band 22, 12. Konferenz ESSV-2001, Bonn 2001, S. 244-249.

[SnL02a] Schnell K.; Lacroix A.: „Analyse von Sprachlauten auf der Basis verzweigter Rohrmodelle", Tagungsband DAGA-2002 - Fortschritte der Akustik, DEGA, Bochum, S. 652-653, 2002.

[SnL02b] Schnell K.; Lacroix A.: „Parameter Estimation of Branched Tube Models by Iterative Inverse Filtering", Proc. 14$^{\text{th}}$ Int. Conf. on Digital Signal Processing, DSP-2002, Santorini 2002, Greece, Vol. I, pp. 333-336.

[SnL02c] Schnell K.; Lacroix A.: „Analysis of Nasals and Nasalized Vowels Based on Branched Tube Models", Proc. XI. European Signal Processing Conference EUSIPCO-2002, Toulouse, Vol. III, pp. 65-68, 2002.

[SnL02d] Schnell K.; Lacroix A.: „Analyse und Erzeugung von Nasalvokalen mittels verzweigter Rohrmodelle", Studientexte zur Sprachkommunikation: Band 24, 13. Konferenz ESSV-2002, Dresden, S. 229-236, 2002.

[SnL03] Schnell K.; Lacroix A.: „Generation of Nasalized Speech Sounds Based on Branched Tube Models Obtained from Separate Mouth and Nose Outputs", Proc. Int. Conf. Acoust., Speech, and Sig. Processing ICASSP-2003, Hong Kong 2003 (in print).

[So74] Sondhi M.M.: „Model for wave propagation in a lossy vocal tract", J.Acoust.Soc.Am. 1974 (55) pp. 1070-1075.

[So79] Sondhi M.M.: „Estimation of Vocal-Tract Areas: The Need for Acoustical Measurements", IEEE Trans. ASSP-27, No. 3, pp. 268-273, 1979.

[So83] Sondhi M.M.; Resnick J.R.: „The inverse problem of the vocal tract: Numerical method acoustical experiments, and speech synthesis", J.Acoust. Soc. Am. 1983 (73) pp. 985-1002.

[So86] Sondhi M.M.: „Resonances of a bent vocal tract", J.Acoust.Soc.Am. 1986 (79) pp. 1113-1116.

[So87] Sondhi M.; Schroeter J.: „A hybrid time-frequency domain articulatory speech synthesizer", IEEE Trans. on Acoustics, Speech and Signal Processing, ASSP-35 (1987), pp. 1070-1075.

[Song80] Song K.H.; Un C. K.: „On Pole-Zero Modelling of Speech", Proc. ICASSP'80, pp. 162-165, 1980.

[Soq96] Soquet A. ;Lecuit V; Metens T. ;Demolin D.: „From Sagittal Cut to Area Function: An RMI Investigation", Proc. ICSLP'96, pp. 1205-1208, 1996.

[Soq02] Soquet A. ;Lecuit V; Metens T. ;Demolin D.: „Mid-sagittal cut to area function transformation: Direct measurements of mid-sagittal distance and area with MRI", Speech Communication Vol. 36, 2002, pp. 169-180.

[St96] Story B.H. et al.: „Vocal Tract Area Functions from Magnetic Resonance Imaging" , J.Acoust.Soc.Am. Vol. 100 (1996), pp. 537-554.

[St98] Story B.H.; Titze I.R.; Hoffman E.A.: „Vocal tract functions for an adult female speaker based on volumetric imaging", J.Acoust.Soc.Am. 1998 (104) pp. 471-487.

[St01] Story B.H.; Titze I.R.; Hoffman E.A.: „The relationship of vocal tract shape to three voice qualities", J.Acoust.Soc.Am. 2001 (109) pp.1651-1667.

[Srö67] Schroeder M.R.: „Determination of the Geometry of the Human Vocal Tract by Acoustical Measurements", J.Acoust.Soc.Am. 1967 (41) 1002-1010.

[Stei77] Steiglitz K.: „On the Simultaneous Estimation of Poles and Zeros in Speech Analysis", IEEE Trans. ASSP-25, pp. 229-234, 1977.

[Stev98] Stevens K. N.: „Acoustic Phonetics", MIT Press, Cambridge London, 1998.

[Str74] Strube H.W.: „Determination of the instant of glottal closure from speech wave", J.Acoust.Soc.Am. (56), 1625-1629.

[Str75a] Strube H.W.: „Zur Bestimmung der Querschnittsfunktion des menschlichen Stimmkanals aus dem Sprachsignal", Tagungsband DAGA'75, 1975, S. 373-376.

[Str75b] Strube H.W.: „Sampled-data representation of non-uniform lossless tube of continuously variable length", J.Acoust.Soc.Am. 1975 (57) 256-257.

[Str76] Strube H.W.: „Can the Area Function of the Human Vocal Tract be Determined from the Speech Wave?", U.S.-Japan Joint Seminar on Dynamic Aspects of Speech Production, Tokyo 1976.

[Str82] Strube H.W.: „Time-Varying Wave Digital Filters for Modeling Analog Systems", IEEE Trans. ASSP, Dec. 1982, pp. 864-868.

[Str00] Strube H.W.: „The meaning of the Kelly-Lochbaum acoustic-tube model", J.Acoust.Soc.Am. 2000 (108) pp. 1850-1855.

[Su87] Sundberg J.; Johansson C.; Wilbrand H.; Ytterbergh C.: „From sagittal distance to area: A study of transverse, vocal tract cross-sectional area", Phonetica (44) 1987, 76-90.

[Suz96] Suzuki H.; Nakai T.; Sakakibara H.: „Analysis of Acoustic Properties of the Nasal Tract Using 3-D FEM", Int. Conf. on Spoken Language Processing ICSLP'96, Philadelphia 1996, Vol. 2, pp. 1285-1288.

[Sü92] Schüßler H.W.: „Digitale Signalverarbeitung", Band I, 3. Auflage, Springer-Verlag Berlin 1992.

[Te80] Teager H.M.: „Some Observations on Oral Flow During Phonation", IEEE Trans. ASSP-28, No. 5, 1980, pp. 599-601.

[Ti97] Titze I.R.; Story B.H.: „Acoustic interaction of the voice source with the lower vocal tract", J.Acoust.Soc.Am. 1997 (101) pp. 2234-2243.

[Tir79] Tirupattur V.; Ananthapadmanabha; Yegnanarayana B.: „Epoch Extraction from Linear Prediction Residual for Identification of Closed Glottis Intercal", IEEE Trans. ASSP-27, No. 4, pp. 309-319, 1979.

[To01] Tokuda I.; Miyano T.; Aihara K.: „Surrogate analysis for detecting nonlinear dynamics in normal vowels", J.Acoust.Soc.Am. 2001 (110) pp. 3207-3217.

[Vä94a] Välimäki V.; Karjalainen: „Improving the Kelly-Lochbaum Vocal Tract Model Using Conical Tube Selections and Fractional Delay Filtering Techniques", Int. Conf. on Spoken Language Processing ICSLP'94, Yokohama 1994.

[Vä94b] Välimäki V.; Karjalainen; Kuisma T: „Articulatory Speech Synthesis Based on Fractional Delay Waveguide Filters", Proc. ICASSP'94 Australia, 1994.

[Vä00] Välimäki V.; Laakso T.I.: „Principles of Fractional Delay Filters", Proc. Int. Conf. on Acoustics, Speech, and Signal Processing. ICASSP'00, Istanbul 2000.

[Vi98] Vich R.; Smekal Z.: „All-Pole and Zero-Pole Speech Modelling", Proc. BIOSIGNAL'98, Brno Tschechien, 1998.

[Vr02] Vries M.P.; Schutte H.K.; Veldman A.E.P.; Verkerke G.J.: „Glottal flow through a two-mass model: Comparison of Navier-Stokes solutions with simplified models", J.Acoust.Soc.Am. (111), pp. 1847-1853, 2002.

[Wa79] Wakita H.: „Estimation of Vocal-Tract Shapes from Acoustical Analysis of the Speech Wave: The State of the Art", IEEE Trans. ASSP-27, No. 3, pp. 281-285, 1979.

[Wo79] Wong D. Y.; Markel J. D.; Gray A. H. Jr.: „Least Squares Glottal Inverse Filtering from the Acoustic Speech Waveform", IEEE Trans. ASSP-27, No. 4, pp. 350-355, 1979.

[Zh02] Zhao W.; Zhang C.; Frankel S.F.; Mongeau L.: „Computational aeroacoustics of phonation, Part I,Computational methods and sound generation mechanisms", J.Acoust.Soc.Am. (112), pp. 2134-2146, 2002.

Lebenslauf von

Karl Schnell

geboren am 31.01.1968 in Frankfurt am Main

1974-1978	Grundschule (Mühlbergschule/Ffm)
1978-1987	Freiherr vom Stein Schule Gymnasium in Frankfurt am Main (Abitur 1987)
1.7.1987-30.9.1988	Grundwehrdienst Standort Wetzlar (2. PzGrenBtl. 133)
1.10.1988-30.9.1989	WS und SS des Ingenieur-Studienganges Holztechnik der Fachhochschule Rosenheim
ab 10.1989	Studium der Physik an der Johann Wolfgang Goethe- Universität Frankfurt am Main mit Abschluß als Diplom-Physiker (Vordiplom 1991, Diplom 1996)
ab 1.1997	Doktorand im Fachbereich Physik der J. W. Goethe-Universität
ab 12.1998	Wissenschaftlicher Angestellter (halbtags/nebentätig) im Institut für Angewandte Physik der J. W. Goethe-Universität Frankfurt

Akademische Lehrer in alphabetischer Reihenfolge:

H. Ast, W. Greiner, R.-J. Jelitto, H. Klein, A. Lacroix, W. Martienssen, R. Mester, E. Mohler, J. Reinhardt, H. Reininger, B. Schürmann, R. Stock, R. Tetzlaff, W. Unger, J. Wolfart, D. Wolf.

Danksagung

Prof. Lacroix möchte Ich danken für die Betreuung der Dissertation und Diplomarbeit und insbesondere für seine Bereitschaft und Unterstützung die erzielten Ergebnisse einem breiteren Publikum vorzustellen. Prof. Kummer möchte Ich danken für die zur Verfügungstellung von Räumlichkeiten und durch seine motivierende Haltung.

Den damaligen Kollegen aus der Diplomzeit Carsten Gerland, Martin Eichler und Markus Lausser danke Ich für die gute Zusammenarbeit und die gute Atmosphäre innerhalb der Arbeitsgruppe. Während der Zeit als Doktorand möchte Ich den Kollegen Frank Ranostaj und (Dr.) Ralf Thomas Pietsch für die gute Zusammenarbeit danken; insbesondere gilt Herrn Dipl. Phys. Ranostaj für seine Diskussionsbereitschaft dank.