Spracherzeugung – Akustik, Modelle und Anwendungen

Arild Lacroix
Johann Wolfgang Goethe-Universität
Institut für Angewandte Physik
Frankfurt am Main
Lacroix@iap.uni-frankfurt.de

1 Einleitung

Die gesprochene Sprache kann als Ergebnis eines in Bild 1 dargestellten komplexen Regelungsprozesses angesehen werden, an dem auch das Hören, die Wahrnehmung und die Informationsverarbeitung im Nervensystem und im Gehirn beteiligt sind [1]. In diesem Beitrag wird nur der Vorwärtszweig des rückgekoppelten Systems behandelt, mit Einschluß der Schallanregung, der Wellenausbreitung im akustischen Bereich der Spracherzeugung und der Abstrahlung vom Mund und/oder den Nasenlöchern [1, 2]. Hinsichtlich der verwendeten Fachtermini sei auf [3] hingewiesen.

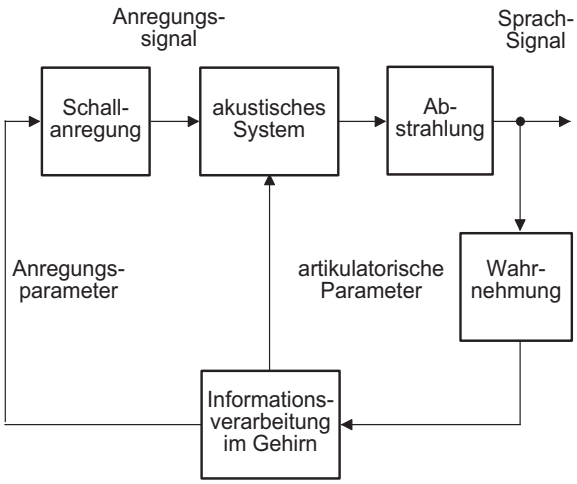

Bild 1: Rückkopplungssystem der Spracherzeugung

2 Mechanismus der Spracherzeugung

Im folgenden werden anhand eines Sagittalschnittes durch den menschlichen Kopf die Organe und deren Funktion erläutert, soweit sie an der Spracherzeugung beteiligt sind. Bild 2 zeigt dieses Schnittbild in stark vereinfachter Form mit den Artikulatoren *Zunge, Unterkiefer, Lippen* und *Gaumensegel (Velum)* , die aktiv zur Artikulation beitragen.

Die Vokale und die stimmhaften Konsonanten werden näherungsweise periodisch angeregt, in dem ein Luftstrom aus der *Lunge* die *Stimmbänder* zum Schwingen bringt. Zwischen den Stimmbändern befindet sich die *Stimmritze (Glottis)*, die infolge der schwingenden Stimmbänder abwechselnd geöffnet oder geschlossen wird. Das Ergebnis ist ein impulsförmiges Signal, welches das akustische System anregt. Das Spektrum dieses breitbandigen Signals hat wegen seiner Periodizität harmonische Komponenten und nimmt bei höheren Frequenzen mit etwa 6 dB/Oktave ab. Diese Art der Anregung wird auch als *Phonation* bezeichnet.

Die Schallanregung erzeugt eine akustische Welle, die sich durch ein System von Röhren und Hohlräumen komplizierter Geometrie ausbreitet. Bei den *Vokalen* ist das Gaumensegel gewöhnlich angehoben, sodaß das akustische System im wesentlichen aus der Rachenhöhle und der daran anschließenden Mundhöhle besteht. Abhängig von der Stellung der Artikulatoren besitzt das Höhlungssystem eine für den jeweiligen Vokal charakteristische Geometrie. Für die Vokale ergibt sich aus der Stellung des Unterkiefers (offen, halboffen, eng), der Position

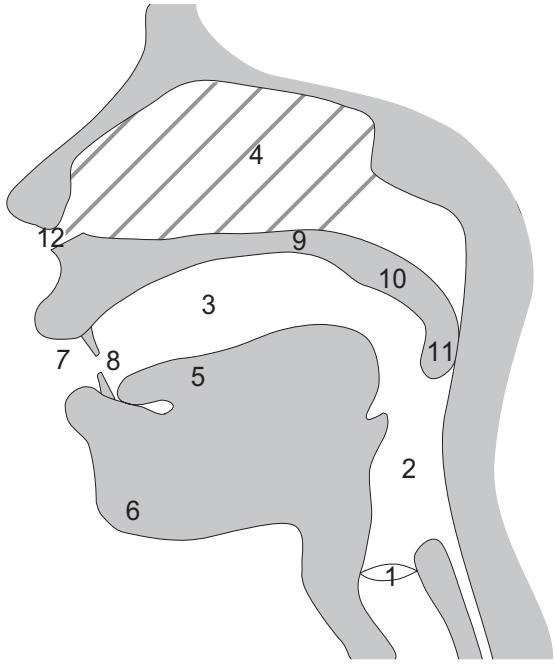

Bild 2: Sagittalschnitt durch den menschlichen Kopf: 1 Stimmbänder, 2 Rachenhöhle, 3 Mundhöhle, 4 Nasenhöhle (Nasenscheidewand schraffiert), 5 Zunge, 6 Unterkiefer, 7 Lippen, 8 Zähne, 9 Gaumen, 10 Gaumensegel, 11 Zäpfchen, 12 Nasenlöcher

des Zungenrückens (vorne, mitten, hinten) und der Lippenform (rund, breit) ein eindeutiger Zusammenhang mit den jeweiligen Lauten, auf den schon FORCHHAMMER hingewiesen hat [4] und der auch die Basis des heute allgemein akzeptierten Vokalsystems ist [5]. Die spektralen Darstellungen der Vokale zeigen mehrere auffällige Energiemaxima, die sogenannten *Formanten*, die von den *Resonanzen* des Höhlungssystems herrühren.

Bei den *nasalierten Vokalen* ist das Gaumensegel abgesenkt, sodaß an der spektralen Formung des Sprachsignals neben der Mund- und Rachenhöhle auch die Nasenhöhle beteiligt ist. Das Sprachsignal wird sowohl von der Mundöffnung wie auch von den Nasenlöchern abgestrahlt. Infolge Interferenz treten in der Übertragungsfunktion neben Polen, die von den Resonanzen verursacht sind, auch Nullstellen, die sogenannten *Antiformanten* auf. Bei den *Nasalen* ist die Mundhöhle an den Lippen, hinter der oberen Zahnreihe oder am Gaumen nach vorne nahezu vollständig verschlossen, sodaß die Abstrahlung des Schalls vorwiegend von den Nasenlöchern erfolgt. Die Mundhöhle ist im Bereich des Gaumensegels mit den übrigen Höhlungen verbunden und wirkt akustisch als gekoppelter Resonator. Auch bei den Nasalen treten Pole und Nullstellen auf. Die *Konsonanten* sind durch eine mehr oder weniger ausgeprägte Verengung des Stimmkanals gekennzeichnet. Bei den stimmhaften Konsonanten erfolgt die Anregung wie bei den Vokalen durch Phonation. Die Position der Verengung (Artikulationsstelle) legt die jeweilige Konsonantenklasse fest (beispielsweise *Dental, Velar*). Zusätzlich zur Phonation kann abhängig vom Grad der Verengung auch turbulentes Rauschen als Anregung auftreten. Bei den *Frikativen* ist das Rauschen die alleinige Lautanregung. Im Fall von *Flüstersprache* wird das turbulente Rauschen an der Stimmritze erzeugt. Eine Sonderstellung unter den Konsonanten nehmen die *Explosivlaute* ein, bei deren Artikulation ein vollständiger Verschluß des Mundraums nach vorne herbeigeführt wird. Nach einer kurzen Staupause wird der Verschluß gelöst und es entsteht ein impulsartiger Schall, der bei stimmhaften Explosiven von Phonation begleitet ist; Explosive können zusätzlich auch *aspiriert* (behaucht) sein.

Die Konsonanten können nach der Artikulationsart und der Artikulationsstelle systematisch entsprechend dem IPA-Konsonanten-Schema eingeordnet werden [5].

3 Akustik der Spracherzeugung

Die Akustik der Schallausbreitung in dem dargestellten Höhlungssystem wird durch die Wellengleichung, eine partielle Differentialgleichung, beschrieben. Da der Querschnitt des Höhlungssystems längs der Ausbreitungsrichtung variiert, ist die WEBSTERsche Horngleichung für diese Fragestellung geeignet [6]. Infolge der Bewegung der Artikulatoren ist die Querschnittsfunktion der Differentialgleichung zeitabhängig. Die Abschlußimpedanzen an der Glottis und am Mund sind ebenfalls zeitabhängig. Abhängig von dem Sprachlaut wechselt die Anregungsstelle und der Anregungstyp (Puls, turbulentes Rauschen, Explosionsgeräusch). Verluste entstehen durch viskose Reibung an den Wänden und infolge Wärmeleitung durch die Wände; dazu kommen noch Verluste durch Vibrationen etwa der Wangen und der Lippen. Geschlossene Lösungen der zeitinvarianten Horngleichung existieren nur für vergleichsweise einfache Querschnittsverläufe, sodaß man für realistische Querschnittsverläufe auf numerische Lösungsverfahren angewiesen ist.

In dem interessierenden Frequenzbereich (Hörfrequenzbereich) genügt die Betrachtung ebener Wellen. Die Analyse der Wellenausbreitung wird weiterhin vereinfacht, wenn das komplizierte Höhlungssystem durch eine Aneinanderreihung homogener Rohrstücke genähert dargestellt wird. Für hinreichend kurze Zeitintervalle kann Zeitinvarianz angenommen werden. Unter diesen Voraussetzungen lassen sich Signalflußgraphen für die Beschreibung der Wellenausbreitung herleiten, die neben ihrer Einfachheit auch den Vorzug besitzen, zeitdiskret realisiert oder implementiert werden zu können.

4 Modelle der Spracherzeugung auf der Basis zeitdiskreter akustischer Rohre

Im folgenden wird das Konzept der Wellengrößen benutzt, wobei eingeprägte Wellen mit *A* und reflektierte Wellen mit *B* bezeichnet werden. Entsprechend der unterschiedlichen physikalischen Dimension werden Druckwellen, Flußwellen und Leistungswellen (eigentlich Wurzel aus der Leistung) verwendet. Bild 3 zeigt den Signalflußgraphen für Druckwellen, der die Verbindung zweier verlustloser homogener Rohre mit den Längen l_0 beschreibt. Für einheitliche Rohrlängen l_0 der beteiligten Rohrstücke führt die Zuordnung $\tau_0 = l_0/c_0 = T/2$ mit der Schallgeschwindigkeit c_0 zu einer zeitdiskreten Realisierung mit der Abtastperiode T. Damit wird die direkte Implementierung auf einem Digitalrechner oder auf Signalprozessor–Hardware möglich und erlaubt auf diesem Wege eine kostengünstige und effiziente Realisierung. Für unterschiedliche Querschnittsflächen miteinander verbundener Rohrstücke hat der Reflexionskoeffizient einen von Null verschiedenen Wert und beschreibt die Reflexion und die Transmission der Wellen *A*, die von beiden Seiten auf die Rohrverbindung treffen. In Bild 4 ist der Zusammenhang zwischen den Rohrquerschnittsflächen S_1 und S_2 dargestellt.

Abhängig von dieser Relation nimmt der Reflexionskoeffizient Werte zwischen minus und plus Eins an. Für $S_1 = S_2$ wird $r = 0$ und es tritt keine Reflexion auf. Für $S_2 \to \infty$ (offenes Rohrende) wird der Reflexionskoeffizient bekanntlich frequenzabhängig [7]. Falls die Frequenzabhängigkeit von r nicht berücksichtigt wird, sollte zumindest eine Längenkorrektur am offenen Rohrende entsprechend der Mündungskorrektur nach Lord Rayleigh vorgenommen werden [8].

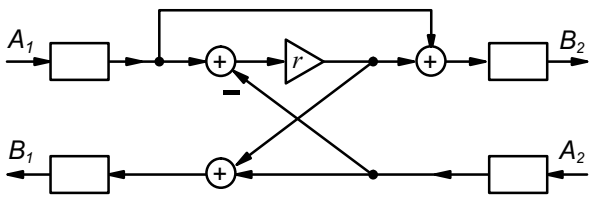

Bild 3: Signalflußgraph zur Beschreibung von Druckwellen für eine Verbindung zweier geschlossener Rohre

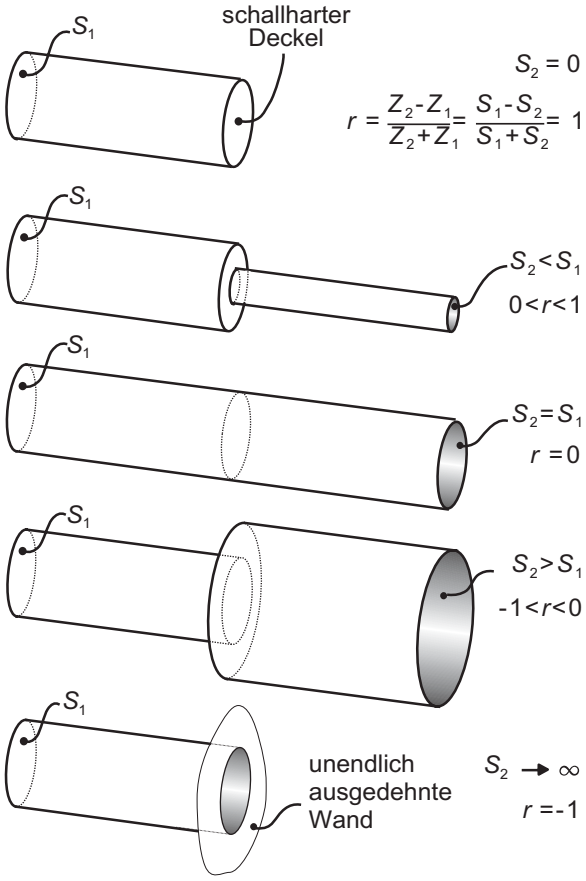

Bild 4: Verbindung zweier homogener akustischer Rohre

Ein Modell des Höhlungssystems der Spracherzeugung, bestehend aus gleich langen homogenen Rohrstücken, ist in Bild 5 dargestellt. Zwischen Rachen- und Mundhöhle sowie der Nasenhöhle ist eine Verbindung wirksam, falls das Gaumensegel gesenkt ist. Die Nasenhöhle hat verglichen mit der Rachen- und Mundhöhle eine komplizierte Struktur:

– Die Nasenhöhle ist durch die Nasenscheidewand in zwei näherungsweise spiegelbildliche Höhlen geteilt.

– Drei Paare von Nebenhöhlen sind über dünne Kanäle mit den Nasenhöhlen verbunden.

– Im Bereich der Konchen und der Siebbeinzellen weisen die Nasenhöhlen eine verwickelte Topologie auf.

Daher ist die in Bild 5 gezeigte Höhle nur ein sehr grobes Modell der Nasenhöhle.

Die in Bild 3 benutzte Struktur zur Verbindung zweier verlustloser homogener Rohre ist von den Wellendigitalfiltern als Adaptor bekannt [9]. Wie schon erwähnt, wird in Bild 3 der Druckwellen-Adaptor benutzt. Bild 6 zeigt eine Anzahl unterschiedlicher Zweitor-Adaptoren, von denen ein Teil physikalisch relevante Wellengrößen verknüpfen [10]. Ein Adaptor mit vier Multiplizierern wurde übrigens schon in [11] veröffentlicht. Bild 7 zeigt drei äquivalente Signalflußgraphen der Rohrelemente, die sich in der Laufzeit in beiden Ausbreitungsrichtungen

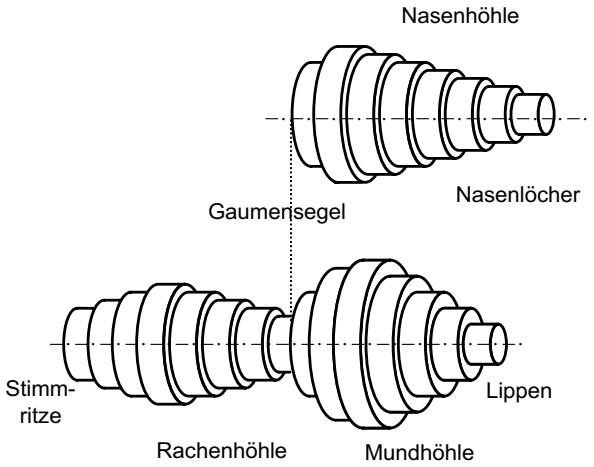

Bild 5: Rohrmodell des Höhlungssystems der Spracherzeugung. An der punktierten Linie gibt es eine Verbindung zwischen Rachen- und Mundhöhle sowie der Nasenhöhle, falls das Gaumensegel gesenkt ist

unterscheiden [10]. Die Laufzeitglieder mit der halben Abtastperiode lassen sich durch die Verwendung der anderen beiden Varianten vermeiden. Wenn die physikalisch korrekte Ausbreitungszeit in beiden Richtungen von Bedeutung ist, dann sollten die beiden Varianten mit den Laufzeitgliedern T abwechselnd verwendet werden. Das Gaumensegel wird durch einen Dreitor-Adapter nachgebildet, der die Verteilung der Schallwellen in die drei beteiligten Höhlungen [12] regelt.

Mittels Rohrelementen, Zweitor- und Dreitor-Adaptoren lassen sich aneinander gefügte und verzweigte Rohrsysteme modellieren, wie das in Bild 5 dargestellte. Für stimmhafte Sprache wird das Rohrmodell an der Stimmritze durch einen näherungsweise periodischen Puls geeigneter Form angeregt. Die Pulsperiode entspricht der Stimmband-Grundfrequenz, die eine maßgebliche Größe für die korrekte Intonation ist. Für geflüsterte Sprache erfolgt die Anregung an der Stimmritze mit einem rauschähnlichen Signal. Bei den stimmlosen Konsonanten wird das Rauschen an der Artikulationsstelle eingespeist.

Die Zeitabhängigkeit kann für das Rohrmodell durch die Verwendung adäquater Adaptoren berücksichtigt werden [13, 14]. In praktischen Versuchen mit Sprachsignalen hat sich jedoch gezeigt, daß auch mit den herkömmlichen Adaptoren bereits sehr gute Resultate erzielt werden [14, 15] Frequenzabhängige Verluste lassen sich durch Dämpfungskoeffizienten in beiden Ausbreitungsrichtungen der Rohrelemente in Bild 7 berücksichtigen. Verluste können auch durch konzentrierte Impedanzen mit resistivem Anteil dargestellt werden [16, 17]. Die Abstrahlung kann durch vergleichsweise einfache Filter modelliert werden [18].

Drei der verschiedenen Typen von Zweitor-Adaptoren ermöglichen die genaue physikalische Modellierung für Druck-, Fluß- oder Leistungswellen [19]. Unabhängig davon können aber auch andere Zweitor-Adaptor-Realisierungen verwendet werden, jedoch treten während der Wellenausbreitung unterschiedliche Verstärkungen auf.

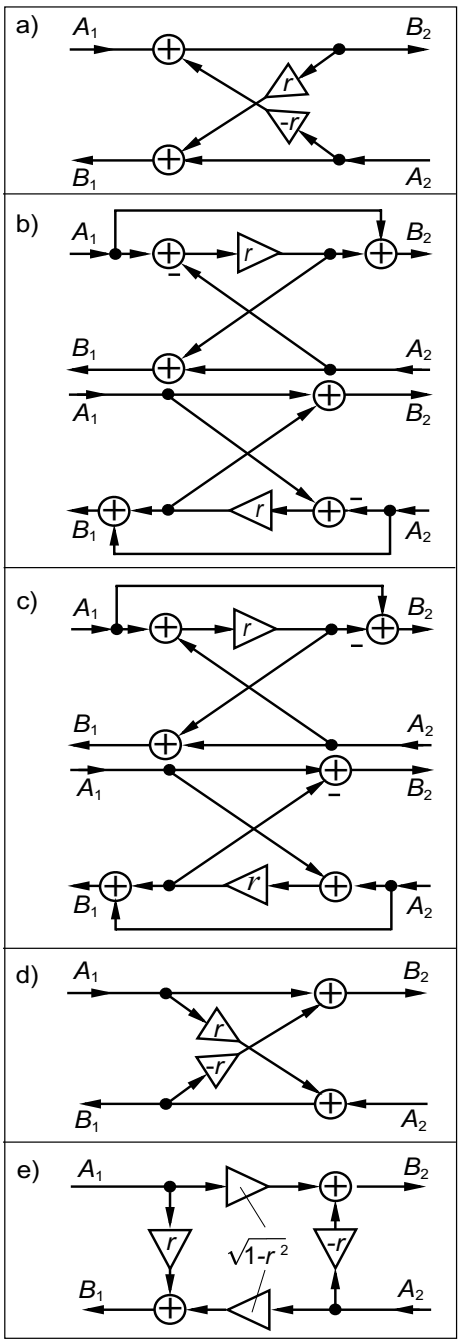

Bild 6: Verschiedene Realisierungen von Zweitor-Adaptoren: b) Druckwellen-Adaptor, c) Flußwellen-Adaptor, e) Leistungswellen-Adaptor. Die Adaptoren a) und d) weisen dasselbe Übertragungsverhalten wie die anderen Adaptoren auf, jedoch mit einer Verstärkung, die vom Reflexionskoeffizienten abhängt; aus [10]

Das interaktive Programm TUBE DESIGNER [20] ermöglicht die simultane Berechnung und Darstellung vorgegebener Querschnitte eines Rohrsystems und dessen Übertragungsfunktion. Bild 8 zeigt dazu ein Beispiel.

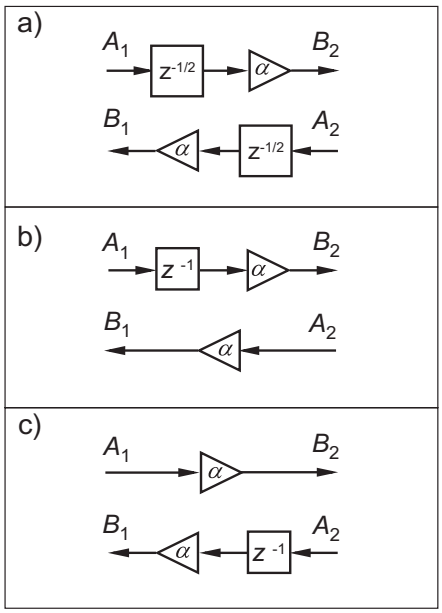

Bild 7: Zeitdiskrete Äquivalente des homogen akustischen Rohrs mit Einschluß von Verlusten. a) Laufzeit in beiden Ausbreitungsrichtungen $\tau_0 = T/2$ mit der Abtastperiode T, b) Laufzeit vorwärts $2\tau_0 = T$, c) Laufzeit rückwärts $2\tau_0 = T$; aus [10]

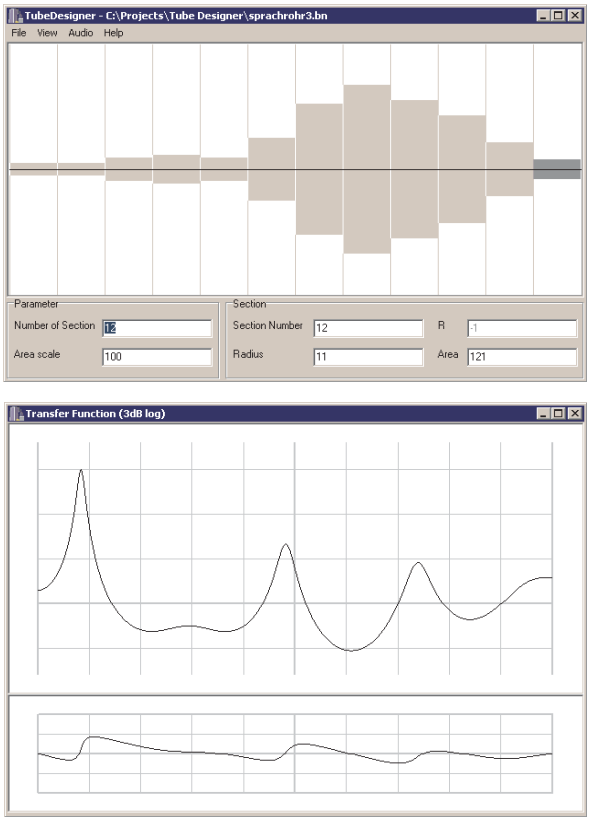

Bild 8: Rohrquerschnitt und Frequenzgang, interaktiv und simultan berechnet mit TUBE DESIGNER

5 Schätzung der Parameter

Die Schätzung der Reflexionskoeffizienten für den Fall einer Aneinanderreihung homogener Rohrstücke derselben Länge ohne Verzweigung wird durch die *lineare Prädiktion* geleistet, die sich auf eine berühmte Arbeit des Franzosen PRONY zurückführen läßt [21]. Die Prädiktion ist später weiterentwickelt worden [22]-[28] und erlaubt die Berechnung des Reflexionskoeffizienten auf zwei Wegen:

– Nach der Berechnung der Autokorrelationsfunktion des Sprachsignals wird ein lineares Gleichungssystem, dessen Koeffizientenmatrix TOEPLITZ-Struktur aufweist, mit dem LEVINSON/DURBIN-Algorithmus gelöst [24].

– *Inverse Filterung* des Sprachsignals durch eine Kettenschaltung nicht rekursiver Zweitore, die invers zu den zeitdiskreten Rohrmodellen arbeiten [24]-[28] die Berechnung des Reflexionskoeffizienten jeder Stufe wird mit dem Ziel der Minimierung der Signalleistung vorgenommen, wobei zwei Beziehungen verwendet werden, die von BURG [25] beziehungsweise ITAKURA und SAITO [26] angegeben wurden.

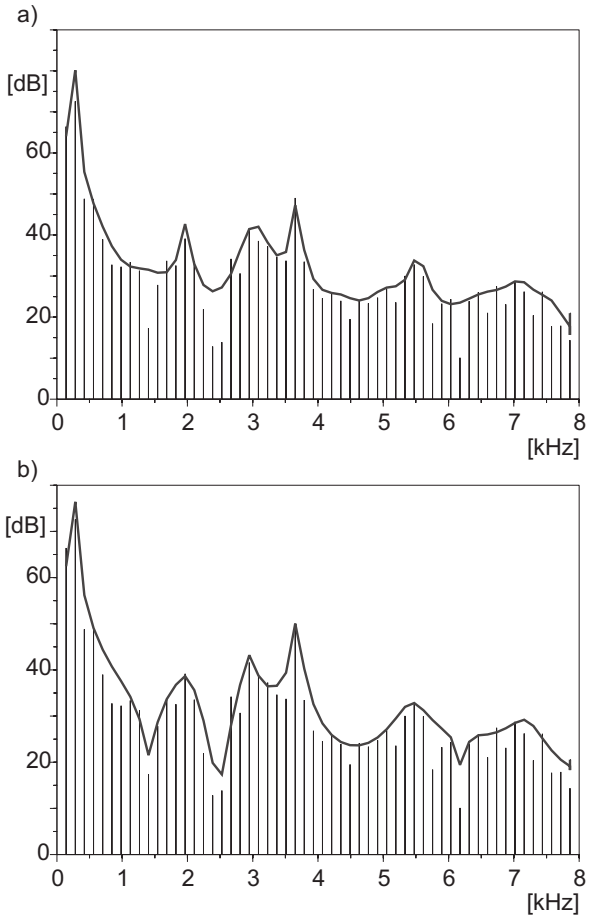

Bild 9: Betragsgang (durchgezogene Linie) für ein a) Nur-Pole-Modell von Grad 30, b) Pol-Nullstellen-Modell mit 20 Polen und 10 Nullstellen im Vergleich mit dem DFT-Spektrum des Vokals [i:]; aus [29]

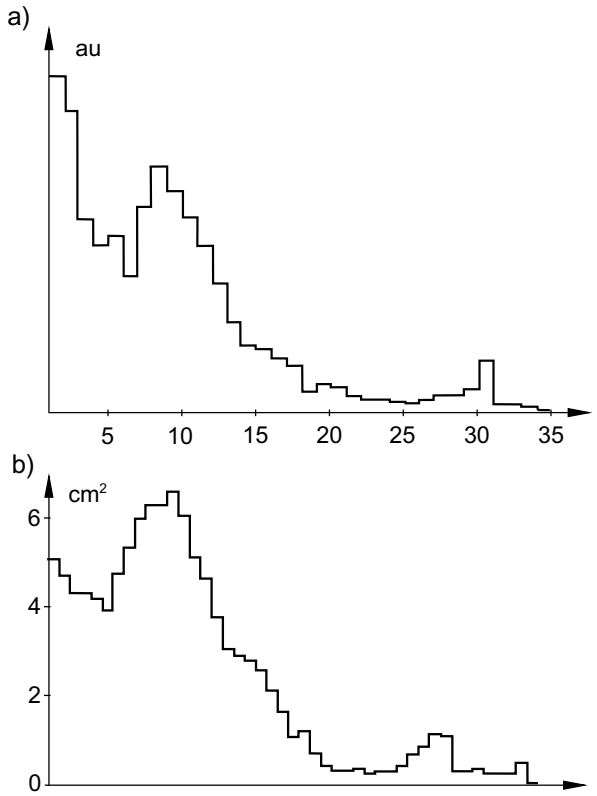

Bild 10: Querschnittsflächen aufgetragen über der Koordinate der Ausbreitungsrichtung (Mund links, Stimmritze rechts a) für ein 33-stufiges Rohrmodell des Vokals [a:] (aus [30]), b) aus NMR-Daten für denselben Vokal jedoch von einem anderen Sprecher; aus [32]

Trotz der unterschiedlichen Herleitung liefern beide Verfahren für Sprachsignale nahezu identische Resultate.

Die Übertragungsfunktion unverzweigter Rohrmodelle ist durch das Fehlen von Nullstellen gekennzeichnet; man spricht daher von Nur-Pole-Modellen. Wendet man die Prädiktion auf Sprachsignale an, so erhält man zumindest für Vokale mit dem Nur-Pole-Modell befriedigende Resultate. Jedes Polpaar repräsentiert eine Resonanzstelle im Spektrum. Vergleicht man die Betragsgänge der Nur-Pole-Modelle mit dem DFT-Spektren der zugrunde liegenden Sprachsignale, so ergeben sich selbst für Vokale Abweichungen. Bild 9 a) verdeutlicht diesen Sachverhalt an dem Vokal [i:]. Trotz des hohen Systemgrades von 30 bleiben an mehreren Stellen des Betragsganges deutliche Abweichungen. Erst mit einem Pol-Nullstellen-Modell können diese Abweichungen behoben werden, wie in Bild 9 b) dargestellt. Der dafür verwendete Schätzalgorithmus ist in [29] beschrieben. Die in Bild 9 erkennbaren Nullstellen im DFT-Spektrum sind auf die Anregung zurückzuführen.

Bild 10 a) zeigt die Querschnittsflächen, die aus den Reflexionskoeffizienten eines unverzweigten Rohrmodells berechnet werden. Der Rohrabschluß an den Lippen ist frequenzabhängig und differiert von Laut zu Laut; zusätzlich wird ein zeitvariabler Abschluß an der Stimmritze benutzt [30, 31]. Der Einfluß von Anregung und Abstrahlung wird durch bis zu dreifache inver-

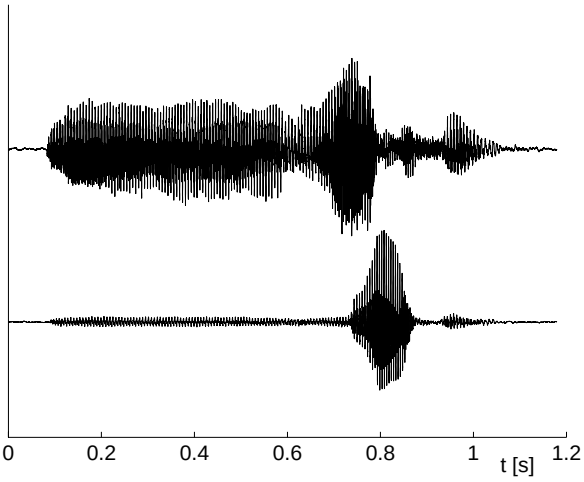

Bild 11: Mundsignal (oben), Nasensignal (unten) der Äußerung „Orange“

se Filterung erster Ordnung separiert. Diese Querschnittsflächen können mit NMR-Daten desselben Sprachlauts, jedoch geäußert von einer anderen Person, verglichen werden [32]. Auch wenn eine Übereinstimmung nicht möglich ist, wird in beiden Fällen tendenziell ein vergleichbarer Kurvenverlauf erzielt; ähnlich gute Resultate wurden auch für andere Sprachlaute ermittelt [30].

Die Parameterschätzung wird erheblich schwieriger, wenn der Nasaltrakt an der Spracherzeugung beteiligt ist. Die Schätzung von Polen und Nullstellen ähnlich wie in dem in Bild 9 dargestelltem Beispiel ist auch hier möglich, jedoch ist der Zusammenhang zwischen Polen und Nullstellen und den Parametern eines verzweigten Rohrmodells nicht eindeutig [33]. Für Nasale wurden gute Resultate erzielt, wenn die Länge der Mundhöhle bekannt ist [34]. Verfahren der inversen Filterung können angewandt werden, wenn die Signale aus der Mundöffnung und den Nasenlöchern getrennt verfügbar sind. Dafür ist ein Algorithmus nötig zur Trennung zweier Quellen aus einem Signal, wie beispielsweise in [35] beschrieben. Wegen der hohen Korrelation zwischen Mund- und Nasensignal sind derartige Verfahren jedoch wenig erfolgversprechend. Wir haben in unserem Sprachlabor die Möglichkeit geschaffen, Mund- und Nasensignal separat aufzuzeichnen. In Bild 11 sind für die Äußerung „Orange“Mund- und Nasensignal dargestellt. Die zeitliche Lage des nasalierten Vokals [$\tilde{a}$] ist gut zu erkennen. Bei einer auditiven Beurteilung zeigt sich, daß nur eine Mischung der beiden Signale den korrekten Höreindruck der Äußerung wiedergibt. Die inverse Filterung der beiden Signale führt zu Schätzungen der Mund- und Nasenhöhle [36]. In [37] wurde ein Algorithmus angegeben, um die Parameter der gesamten Rohrkonfiguration zu bestimmen.

Es wurde bereits erwähnt, daß die Topologie des Nasenraums unter anderem wegen der gekoppelten drei Paare von Nebenhöhlen außerordentlich kompliziert ist. Bild 12 veranschaulicht in einer 3D-Ansicht diese komplexe Struktur. Die Daten stammen aus einer CT-Aufnahme, die uns von unserem Universitätsklinikum zur Verfügung gestellt wurden. Für dieses 3D-Modell wurde die Wellenausbreitung mit dem Verfahren der

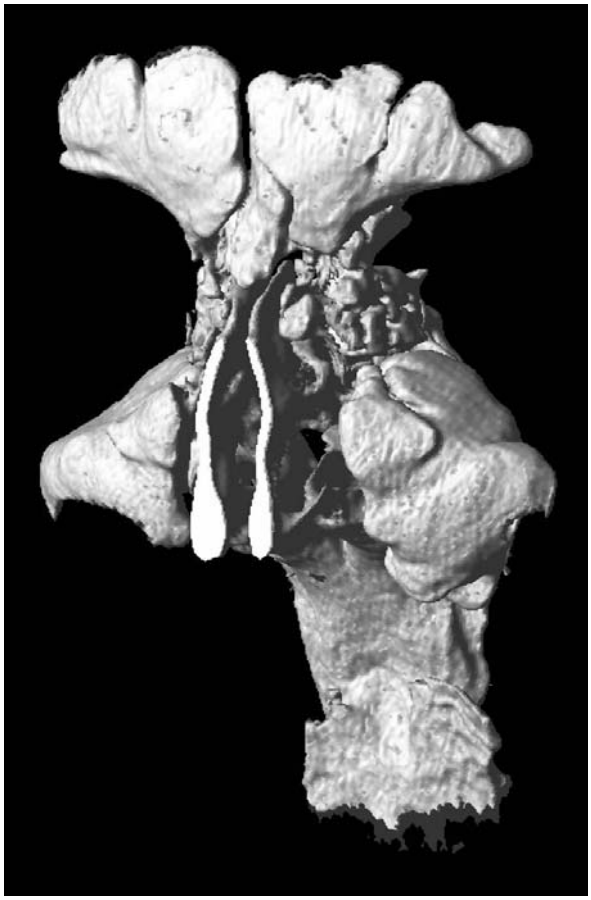

Bild 12: 3D-Ansicht der Nasenhöhle zwischen Nase (weisse Fläche vorne) und Gaumensegel (unten) mit Einschluß dreier Paare von Nebenhöhlen: Stirnhöhlen (oben), Kiefernhöhlen (links und rechts), Keilbeinhöhlen (hinten, nicht sichtbar); aus [38]

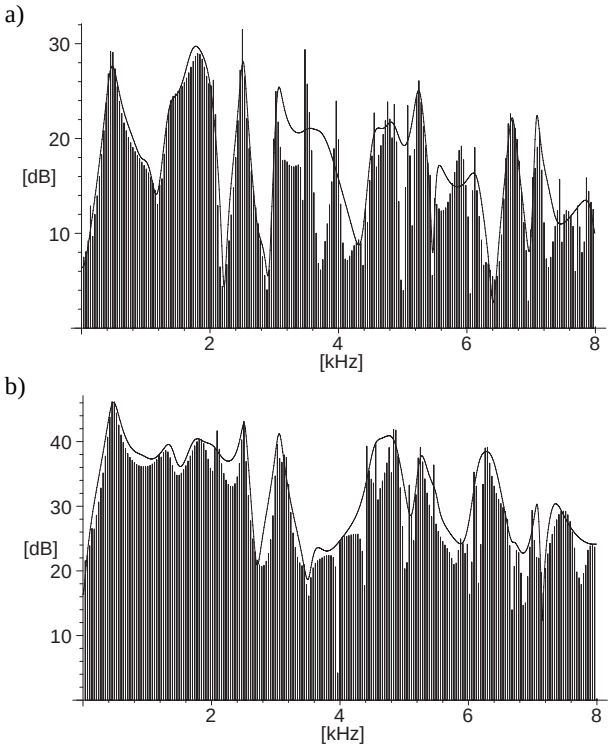

Bild 13: Betrag des Frequenzgangs des Modells der Nasenhöhle mit Einschluß eines Nebenhöhlenpaares (durchgezogene Linie) a) rechter Zweig, b) linker Zweig, im Vergleich mit DFT-Spektren, entstanden durch Auswertung der Wellengleichung mittels finiter Differenzen; aus [39])

finiten Differenzen analysiert [38]. Es resultieren Impulsantworten beider Nasengänge, deren DFT-Spektren mit Modell-Frequenzgängen beider Nasengänge mit je einer Nebenhöhle in Bild 13 verglichen werden [39]. Die Approximation ist schon vergleichsweise gut und kann voraussichtlich durch die Berücksichtigung von mehr als nur einem Nebenhöhlenpaar noch weiter verbessert werden.

6 Anwendungen

Akustische Rohrmodelle sind in der Sprachverarbeitung bisher vorzugsweise auf zwei Gebieten eingesetzt worden: *Sprachsynthese* und *Sprachcodierung*.

In der Sprachsynthese gibt es zahlreiche Untersuchungen, die auch teils zu kompletten Synthese-Systemen geführt haben [40]-[47]. Als vorteilhaft hat sich dabei die vergleichsweise enge Relation zwischen den Rohrparametern und den artikulatorischen Parametern erwiesen. Hinsichtlich der bislang erzielten Sprachqualität sind zukünftig noch Verbesserungen zu erwarten.

In der Sprachcodierung sind Rohrmodelle in Gestalt der Kreuzglied-Kettenfilter zumindest auf der Empfängerseite (Decoder) fest etabliert. Der Grund ist in dem schnell durchführbarem Stabilitätstest des Empfängerfilters zu sehen, der wegen der üblicherweise nötigen groben Quantisierung der Übertragungsparameter wichtig ist. Darüberhinaus ist die Übertragungsfunktion des Rohrmodells unempfindlich gegenüber der Quantisierung der Rohrparameter (Reflexionskoeffizienten oder logarithmierte Flächenverhältnisse). Beispiele für Codierverfahren basierend auf Rohrmodellen sind der GSM-Standard [48] und der LPC-Vocoder [49].

Anwendungen in der Spracherkennung sind möglich, wenn die Modellierung des Stimmappartes so gut gelingt, daß sprecherspezifische Eigenheiten hinreichend gut wiedergegeben werden.

Auch die Spracherkennung kann von den Rohrmodellen profitieren, sofern die zuverlässige Trennung unterschiedlicher Lautklassen gelingt.

7 Zusammenfassung

Die Spracherzeugung wird behandelt mit Einschluß der Schallanregung und der Wellenausbreitung im Höhlungssystem des Stimmapparates. Für die verschiedenen Kategorien der Sprechlaute wird die Artikulation erläutert. Unter der Annahme der Ausbreitung ebener Wellen werden Modelle angegeben für das Spracherzeugungssystem. Mittels modularer Komponenten wie Rohr-Element, Zweitor- und Dreitor-Adaptor können vollständige Systeme der Spracherzeugung nachgebildet werden, die für stimmhafte Sprechlaute mit näherungsweise periodischen Pulsen oder mit einem rauschähnlichen Signal für stimmlose Laute angeregt werden. Für ein unverzweigtes Rohrmodell liefert die lineare Prädiktion die Reflexionkoeffizienten als Schätzwerte des Modells, aus denen sich die Querschnittsflächen sehr einfach berechnen lassen. Die Parameterschätzung ist ungleich schwieriger für verzweigte Rohrmodelle, die für Nasale, nasalierte Vokale, Liquide und auch für stimmlose Konsonanten adäquat sind; Verfahren zur Parameterschätzung werden angegeben. Die verwickelte Geometrie des Nasaltrakts wird genauer dargestellt. Existierende und potentielle Anwendungen der Rohrmodelle werden erläutert und diskutiert.

* * *

Den Herren Diplom-Physiker Frank Ranostaj and Diplom-Physiker Karl Schnell danke ich für zahlreiche Diskussionen und für die Unterstützung bei der Vorbereitung des Manuskripts.

Literatur

[1] FLANAGAN, J. L.: *Speech Analysis, Synthesis, and Perception*. Berlin–Heidelberg–New York, 2nd ed. 1972.

[2] FANT, G.: *Acoustic Theory of Speech Production*. The Hague-Paris, 2nd ed. 1970.

[3] SOTSCHECK, J. et al.: *Terminologie der Sprachakustik. ITG-Empfehlung 4.3.1-01, Berlin–Offenbach 1996, 33 pp*.

[4] FORCHHAMMER, J.: *Die Grundlage der Phonetik* . Heidelberg 1924.

[5] International Phonetic Association: *Report on the 1989 Kiel Convention*. Journal of the Int. Phonetic Asoc. 19

(1989) no. 2, pp. 67-80. Compare also IPA chart, revised 1993 in Journal of the Int. Phonetic Assoc. 23 (1993) no. 1, center page unnumbered.

[6] UNGEHEUER, G.: *Elemente einer akustischen Theorie der Vokalartikulation*. Berlin–Göttingen–Heidelberg, 1962.

[7] SCHÖNBACH, B.: *Schallausbreitung in gekoppelten Rohrsystemen*. VDI-Fortschrittsberichte Reihe 7, Nr. 176, VDI-Verlag, Düsseldorf 1990, 189 pp.

[8] RAYLEIGH, J. W. S.: *The Theory of Sound*. Vol. II §312, London, 2nd ed. 1896.

[9] FETTWEIS, A.: *Digital Filter Structures Related to Classical Filter Networks*. Archiv der Elektronik und Übertragungstechnik 25 (1971), pp. 79-89.

[10] LACROIX, A.: *Digitale Filter.* München–Wien, 4. Aufl. 1996.

[11] KELLY, J. L.; LOCHBAUM, C. C.: *Speech Synthesis*. Proc. 4. ICA Copenhagen 1962, paper G42 pp. 1-4.

[12] LACROIX, A.: *Source Coding of Speech Signals by Improved Modeling of the Voice Source*. Proc. lTG Conf. Information and System Theory in Digital Communications, Berlin 1978, pp. 103-108.

[13] STRUBE, H. W.: *Time-Varying Wave Digital Filters for Modeling Analog Systems.* IEEE Trans. on Acoustics, Speech, and Signal Processing ASSP-30 (1982), pp. 864-868.

[14] MEYER, P.; STRUBE, H. W.: *Calculations on the Time-Varying Vocal Tract.* Speech Communication 3 (1984), pp. 109-122.

[15] EICHLER, M.; LACROIX, A.: *Schallausbreitung in zeitvariablen Rohrsystemen.* Tagungsband DAGA, Bonn 1996, pp. 506-507.

[16] KARAL, E. C.: *The Analogous Acoustical Impedance for Discontinuities and Constrictions of Circular Cross Section*. J. Acoust. Soc. of America 25,5 (1953), pp. 327-334.

[17] LILJENCRANTS, J.: *Speech Synthesis with a Reflection-Type Line Analog*. Thesis Royal Inst. of Technology, Stockholm, 1985.

[18] LAINE, U. K.: *Modelling of Lip Radiation Impedance in the Z-Domain.*. Proc. lCASSP, Paris 1982, pp. 1992-1995.

[19] KUBIN, G.: *Wave Digital Filters: Voltage, Current or Power Waves.* Proc. ICASSP, Tampa 1985, pp. 69-72.

[20] RANOSTAJ, F.: *Programm* TUBE DESIGNER *zum interaktiven Entwurf und zur Analyse von Rohrsystemen*. Institut für Angewandte Physik, Frankfurt am Main 2002.

[21] PRONY, R.: *Essai Experimental et Analytique.* Journal de l' École Polytechnique ou Bulletin du Travail fait a cette école; Deuxieme Cahier, Paris an IV (1975), pp. 24-76.

[22] ATAL, B. S.; HANAUER, L. S.: *Speech Analysis and Synthesis by Linear Prediction*. J. Acoust. Soc. of America 50 (1971), pp. 637-655.

[23] MAKHOUL, J.: *Linear Prediction: A Tutorial Review*. Proc. IEEE 63 (1975), pp. 561-580.

[24] MARKEL, J. D.; GRAY, A. H.: *Linear Prediction of Speech*. Berlin–Heidelberg–New York, 2nd printing 1980.

[25] BURG, J.: *A New Analysis Technique for Time Series Data*. NATO Advanced Study Institute on Signal Processing, Enschede 1968.

[26] ITAKURA, F.; SAITO, S.: *Digital Filtering Techniques for Speech Analysis and Synthesis*. Proc. 7. ICA, Budapest 1971, pp. 261-265.

[27] WAKITA, H.: *Direct Estimation of the Vocal Tract Shape by Inverse Filtering of Acoustic Speech Waveforms*. IEEE Trans. on Audio and Electroacoustics, AU-21 (1973), pp. 417-427.

[28] MAKHOUL, J.: *Stable and Efficient Lattice Methods for Linear Prediction*. IEEE Trans. Acoustics, Speech, and Signal Processing ASSP-25 (1977), pp. 423-428.

[29] SCHNELL, K.; LACROIX, A.: *Pole-Zero Estimation from Speech Signals by an Iterative Procedure*. Proc. ICASSP, Salt Lake City 2001, CD-ROM.

[30] SCHNELL, K.; LACROIX, A.: *Parameter Estimation for Models with Time Dependent Glottis Impedance*. Proc. 2nd Eurosip Conf. ECMCS '99, Krakow 1999, CD-ROM.

[31] SCHNELL, K.; LACROIX, A.: *Realization of a Vowel-Plosive-Vowel Transition by a Tube Model*. Proc. Eusipco, Tampere 2000, pp. 757-760.

[32] STORY, B. H. et al.: *Vocal Tract Functions from Magnetic Resonance Imaging*. J. Acoust. Soc. of America 100 (1996), pp. 537-554.

[33] SCHNELL, K.; LACROIX, A.: *Parameter Estimation for Branched Tube Systems*. Tagungsband ITG/Konvens Tagung Sprachkommunikation, Ilmenau 2000, pp. 127-130.

[34] LIU, M.; LACROIX, A.: *Improved Vocal Tract Model for the Analysis Nasal Speech Sounds*. Proc. ICASSP, Atlanta 1996, pp. II 801-804.

[35] JUTTEN, C.; HERAULT, J.: *Blind Separation of Sources, Part I: An Adaptive Algorithm Based on Neuromimetic Architecture*. Signal Processing 24(1991), pp. 1-10.

[36] LIU, M.; LACROIX, A.: *Analysis of Acoustic Models of the Vocal Tract Including the Nasal Cavity*. Proc. 45. Int. Kolloq. Ilmenau 1998, Vol. I pp. 433-438.

[37] SCHNELL, K.; LACROIX, A.: *Parameter Estimation from Speech Signals for Tube Models*. Proc. 45. ASA/EAA Conf. Berlin 1999, CD-ROM.

[38] RANOSTAJ, F.; LACROIX, A.: *Bestimmung des Übertragungsverhalten des Nasaltraktes aus computertomographischen Daten*. Tagungsband ITG/Konvens Tagung Sprachkommunikation, Ilmenau 2000, pp. 131-134.

[39] RANOSTAJ, F.; SCHNELL, K.; LACROIX, A.: *Modellierung des Nasaltrakts*. Tagungsband 10. Konf. ESSV, Görlitz 1999, pp. 58-63.

[40] ITAKURA, F. et al.: *An Audio Response Unit Based on Partial Autocorrelation*. IEEE Trans. Communications COM-20 (1972), pp. 792-797.

[41] JONSSON, A.; HEDELIN, P.: *A Swedish Text–to–Speech System Based on an Area Function Model*. Proc. ICASSP, Boston 1983, pp. 1340 - 1343.

[42] SONDHI, M. M.: *An Improved Vocal Tract Model*. Proc. 11. ICA, Paris 1983, pp. 167 - 170.

[43] MEYER, P.; STRUBE, H. W.; WILHELMS, R.: *Anpassung eines stilisierten Vokaltraktmodelles an stationäre Sprachlaute*. Tagungsband DAGA, Darmstadt 1984, pp. 825 - 828.

[44] HEIKE, G.; PHILIPP, J.: *Artikulatorische Sprachsynthese: Das Programmsystem LISA*. Tagungsband DAGA, Darmstadt 1984, pp. 833 - 836.

[45] DORFFNER, G.; KOMMENDA, M; KUBIN, G.: *GRAPHON – The Vienna Speech Synthesis System for Arbitrary German Text*. Proc. ICASSP, Tampa 1985, pp. 744 - 747.

[46] CARRÉ, R.; CHENNOUKH, S.; MRAYATI, M.: *Vowel-Consonant-Vowel Transitions: Analysis, Modeling, and Synthesis*. Proc. ICSLP, Banff 1992.

[47] EICHLER, M.; LACROIX, A.: *Ein Experimentalsystem zur Sprachsynthese mit einem zeitdiskreten Rohrmodell*. Tagungsband DAGA, Bonn 1996, pp. 508 - 509.

[48] *Special Issue GSM-Standard*, Speech Communication, January 1988.

[49] MARKEL, J. D.; GRAY, A. H.:*Fixed-Point Truncation Arithmetic Implementation of a Linear Prediction Autocorrelation Vocoder*. IEEE Trans. on Acoustics, Speech, and Signal Processing, ASSP-22 (1974), pp. 273 - 282.