Analyse von Sprachlauten auf der Basis verzweigter Rohrmodelle

K. Schnell, A. Lacroix

Institut für Angewandte Physik, Johann Wolfgang Goethe-Universität
Robert-Mayer-Straße 2-4, D-60325 Frankfurt am Main

Einleitung

Die Ausbreitung ebener Wellen durch den Sprechtrakt kann durch ein Rohrmodell beschrieben werden, das im zeitdiskreten Fall durch ein Kreuzglied-Kettenfilter realisiert wird. Dabei kommen verzweigte Rohrmodelle für die Analyse von Nasalen und nasalierten Vokalen in Betracht. In diesem Beitrag werden zeitdiskrete Rohrmodelle mit einem Seitenzweig behandelt, wobei der Seitenzweig für Nasale und nasalierte Vokale eine unterschiedliche Bedeutung besitzt. Um die Parameter des Modells aus Sprachsignalen zu schätzen, wird eine inverse Filterung vorgestellt, spezialisiert auf das entsprechende System.

Verzweigte Rohrmodelle

Die Struktur des verzweigten Rohrmodells ist in Bild 1 beispielhaft dargestellt. Der Seitenzweig stellt im Falle von Nasalen den Mundraum dar, während er für nasalierte Vokale den Einfluß des Nasaltraktes modelliert.

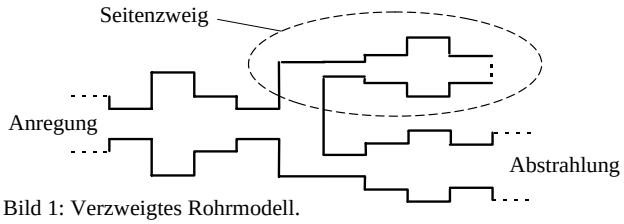

Bild 1: Verzweigtes Rohrmodell.

Für die Beschreibung des zeitdiskreten Rohrmodells werden die vorwärtsgerichteten Wellen x^f und rücklaufenden Wellen x^b der verschiedenen Rohrelemente durch Adaptoren verknüpft. Die Betriebskettenmatrix T_i verknüpft die Wellengrößen des i-ten Rohrelementes mit denen des benachbarten Rohrs. Dabei beschreibt T_i einen Querschnittssprung mit einem zusätzlichem Rohrstück:

$$\boldsymbol{T}_i = \begin{pmatrix} 1 & r_i z^{-1} \\ r_i & z^{-1} \end{pmatrix}, \quad \begin{pmatrix} x_i^f \\ x_i^b \end{pmatrix} = \boldsymbol{T}_i \begin{pmatrix} x_{i+1}^f \\ x_{i+1}^b \end{pmatrix}. \qquad (1)$$

Die Rohrverzweigung wird durch einen Dreitoradaptor realisiert. Wie in [1] beschrieben, kann das Dreitor mit dem Seitenzweig durch eine 2×2 Betriebskettenmatrix $\boldsymbol{T}_D$ beschrieben werden:

$$\boldsymbol{T}_D = \frac{1}{\rho_2(Q+P)} \cdot \begin{pmatrix} \langle \rho_1+\rho_2-1, 1\rangle & \langle \rho_2-1, 1-\rho_1\rangle z^{-1} \\ \langle 1-\rho_1, \rho_2-1\rangle & \langle \rho_1+\rho_2-1, 1\rangle z^{-1} \end{pmatrix}, \qquad (2)$$

mit der Abkürzung $\langle x, y\rangle := x\cdot Q + y\cdot P$.

Die in Bild 2 auftretende rationale Teilübertragungsfunktion $\tilde{H}(z)$ mit den Polynomen $Q(z)$ und $P(z)$ beschreibt den Seitenzweig am

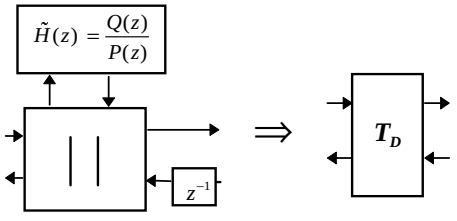

Bild 2: Reduziertes Dreitor.

Dreitorparalleladaptor und bestimmt die Nullstellen des Rohrmodells. Die beiden Parameter ρ_1 und ρ_2 des Dreitoradaptors beschreiben die Kopplung. Die Reflexions-koeffzienten des Seitenzweiges bestimmen nicht nur die Nullstellen des Rohrmodells, sondern liefern auch einen Beitrag zu den Polstellen; das führt zu einer Verkopplung der Pole und Nullstellen, was die Parameterbestimmung kompliziert gestaltet [2].

Parameterbestimmung

Für die Bestimmung der Parameter aus dem Sprachsignal wird eine Minimierung der Ausgangsleistung des inversen Rohrsystems durchgeführt. Der rekursive Anteil des inversen Systems kann dafür separiert werden, so daß dieser Teil zuerst geschätzt und invers gefiltert werden kann. $\boldsymbol{T}_D$ kann dazu in einen rein rekursiven Faktor B^{-1} und eine Matrix $\boldsymbol{T}_D'$ faktorisiert werden, die nur Zählerpolynome als Elemente besitzt:

$$\boldsymbol{T}_D = \frac{1}{B(z)} \cdot \boldsymbol{T}_D', \qquad B(z) = \rho_2(Q+P) = \rho_2 B'(z). \qquad (3)$$

Wie für die inverse Filterung von allgemeinen Rohrsystemen in [1] beschrieben, erhält das inverse Filter einen Korrekturfaktor, damit der konstante Term der Übertragungsfunktion nicht von den zu schätzenden Parametern abhängt. Deswegen wird der Gesamtfaktor ρ_2 in $B(z)$ für die Parameterbestimmung weggelassen. Die Nullstellen $B' = Q + P$ des Rohrmodells, die den rekursiven Teil des inversen Filters repräsentieren, werden mittels eines ARMA Schätzverfahrens gewonnen, welches in [3] beschrieben ist. Zusätzlich zu den Nullstellen werden dabei auch Pole geschätzt, die den Einfluß der Polstellen des Rohrmodells berücksichtigen, aber weiterhin nicht verwendet werden. Der rein rekursive Anteil wird zuerst mittels DFT im Frequenzbereich durchgeführt. Das Eingangssignal x' des nicht-rekursiven inversen Filters in Bild 3 ergibt sich aus:

$$x' = \mathrm{IDFT}\,(X \cdot B'^{-1}).$$

Die Betriebskettenmatrizen in

$$\boldsymbol{x}_i = \boldsymbol{T}_i \cdot \boldsymbol{x}_{i-1}, \quad \boldsymbol{x}_{M+1} = \boldsymbol{T}_D' \cdot \boldsymbol{x}_M; \quad \boldsymbol{x}_i = \begin{pmatrix} x_i^o \\ x_i^u \end{pmatrix}$$

werden für die Beschreibung des inversen Filters benutzt, wie in Bild 3 gezeigt. Für die Analyse einzelner Perioden werden die Signalabschnitte für die Operationen der Zustandsspeicher periodisch fortgesetzt, sodaß die Zustandsspeicher eine zyklische Verschiebung realisieren.

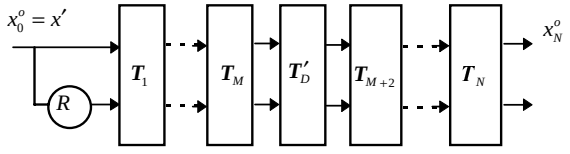

Bild 3: Nichtrekursiver Teil des inversen Filters.

R ist der Rohrabschluß und wird mit $R = -0{,}9$ angenommen. Für die Parameterbestimmung der Pole des Rohrmodells wird die Leistung des Filterausgangs x_N^o des inversen Filters minimiert. Die Koeffizienten jedes Zweitores werden nacheinander bestimmt, beginnend bei T_1 für den Reflexionskoeffizienten r_1 bis T_N; ρ_1 und ρ_2 sind die zu schätzenden Parameter von $\boldsymbol{T}_D'$. Die Koeffizienten können optimal ermittelt werden, unter der Voraussetzung daß die anderen bekannt sind. Dafür müssen die beiden Eingangssignale des

zu bestimmenden Zweitors zu dem Filterausgang x_N^o in der Weise gefiltert werden, daß die Ausgangsleistung minimal wird. Die Bedingungsgleichungen ergeben sich zu:

$$\mathrm{E}\left[x_N^{o\,2}\right] \rightarrow \min \quad \Rightarrow \quad \frac{\partial \mathrm{E}\left[x_N^{o\,2}\right]}{\partial r_i} = 0\,, \quad \frac{\partial \mathrm{E}\left[x_N^{o\,2}\right]}{\partial \rho_i} = 0\,. \qquad (4)$$

Die Lösungen von (4) sind abhängig von den Parametern der Zweitore, die sich hinter dem zu schätzenden Zweitor befinden. Diese können durch die Matrizen

$$\boldsymbol{F}_i = \begin{pmatrix} F_i^{11} & F_i^{12} \\ F_i^{21} & F_i^{22} \end{pmatrix} = \begin{cases} \prod_{k=i+1}^{M} \boldsymbol{T}_k \cdot \boldsymbol{T}_D' \cdot \prod_{k=M+2}^{N} \boldsymbol{T}_k & \text{für } i = 1...M \\ \prod_{k=i+1}^{N} \boldsymbol{T}_k & \text{für } i = M+1...N-1 \end{cases} \qquad (5)$$

dargestellt werden. Für die optimalen Reflexionskoeffizienten $\hat{r}_i$ ist in [4, 5] ein Ausdruck angegeben. Um für die optimalen Parameter $\hat{\rho}_1$ und $\hat{\rho}_2$ eine kompakte Formel zu erhalten, wird das Signal x_N^o in Terme entwickelt, welche abhängig von ρ_1 und ρ_2 sind:

$$x_N^o = o + \rho_1 \cdot o_1 + \rho_2 \cdot o_2\,.$$

Daraus folgen die optimalen Koeffizienten mit (4):

$$\hat{\rho}_1 = -\frac{\overline{o \cdot o_1} + \rho_2 \cdot \overline{o_2 \cdot o_1}}{\overline{o_1 \cdot o_1}}\,,\ \hat{\rho}_2 = -\frac{\overline{o \cdot o_2} + \rho_1 \cdot \overline{o_1 \cdot o_2}}{\overline{o_2 \cdot o_2}}\,. \qquad (6)$$

Die Signale $o\,, o_1$ und o_2 können aus den Polynomen Q und P zusammen mit den Polynomen $F_{M+1}^{\lambda\beta}$, welche die Elemente der Matrix $\boldsymbol{F}_{M+1}$ sind, berechnet werden durch

$$o_1 = u_Q^{11} - l_P^{11} + l_Q^{12} - u_Q^{12}\,, \qquad o_2 = l_Q^{12} + u_Q^{11} + l_Q^{11} + u_P^{12}\,,$$

$$o = -u_Q^{11} - l_Q^{11} - l_Q^{12} + l_P^{11} + u_P^{11} + l_P^{12} - u_P^{12} + u_Q^{12}$$

mit

$$u_Q^{\lambda\beta}(n) = q(n) * f_{M+1}^{\lambda\beta}(n) * x_M^u(n)\,,\ l_Q^{\lambda\beta}(n) = q(n) * f_{M+1}^{\lambda\beta}(n) * x_M^l(n-1)\,,$$

$$u_P^{\lambda\beta}(n) = p(n) * f_{M+1}^{\lambda\beta}(n) * x_M^u(n)\,,\ l_P^{\lambda\beta}(n) = p(n) * f_{M+1}^{\lambda\beta}(n) * x_M^l(n-1)\,.$$

Die $q(n)$, $p(n)$ und $f_{M+1}^{\lambda\beta}(n)$ sind Polynomkoeffizienten von $Q = \sum_k q(k) \cdot z^{-k}$, $P = \sum_k p(k) \cdot z^{-k}$ und $F_{M+1}^{\lambda\beta} = \sum_k f_{M+1}^{\lambda\beta}(k) \cdot z^{-k}$.

Da ein Parameter nur unter der Bedingung optimal geschätzt wird, daß die übrigen Koeffizienten bekannt sind, müssen mehrere Iterationen verwendet werden, um ein Leistungsminimum zu erreichen. Mit jeder neuen Iteration sind die vorgegeben Koeffizienten in den Matrizen $\boldsymbol{F}_i$ besser vorbestimmt, so daß die Parameter im Vergleich zur vorherigen Iteration genauer geschätzt werden können.

Analyse von Sprachsignalen

Die stimmhaften Sprachsignale werden mit einer adaptiven Präemphase vorgefiltert, um die Anregung und Abstrahlung aus dem Sprachsignal zu separieren. Diese Vorfilterung wird mit einer wiederholten Prädiktion erster Ordnung realisiert. Als Sprachsignal für die Analyse werden mehrere benachbarte Perioden verwendet, die im Spektralbereich zu einer einzelnen Periode gemittelt werden, so daß die Analyse weitgehend unabhängig von den Fluktuationen der Perioden ist. In Bild 4a) sind die Ergebnisse der Analyse des Nasals /n/ zu sehen. In Bild 4b) und c) sind die Resultate der Analyse des naslierten Vokals / $\tilde{a}$ / und des nasalierten Schwalautes gezeigt. Die in Bild 4a) dominante Nullstelle resultiert aus dem Seitenzweig, der den angekoppelten Mundraum repräsentiert, während bei den nasalierten Vokalen in b) und c) der Seitenzweig den Nasaltrakt modelliert und somit eine andere Nullstellenverteilung aufweist. Die ausgeprägte Nullstelle in Bild 4b) und c) um 600 Hz der nasalierten Vokale ist in der DFT wie im geschätzten Betragsgang gut zu erkennen. Die gezeigten Ergebnisse wurden mit 50 Iterationen erzielt. Die Abtastrate der analysierten Sprachsignale beträgt 16 kHz. Obwohl die geschätzten Nullstellen der ARMA Schätzung die Lage der Pole einschränken, können die Pole durch die iterative Schätzung hinreichend gut bestimmt werden.

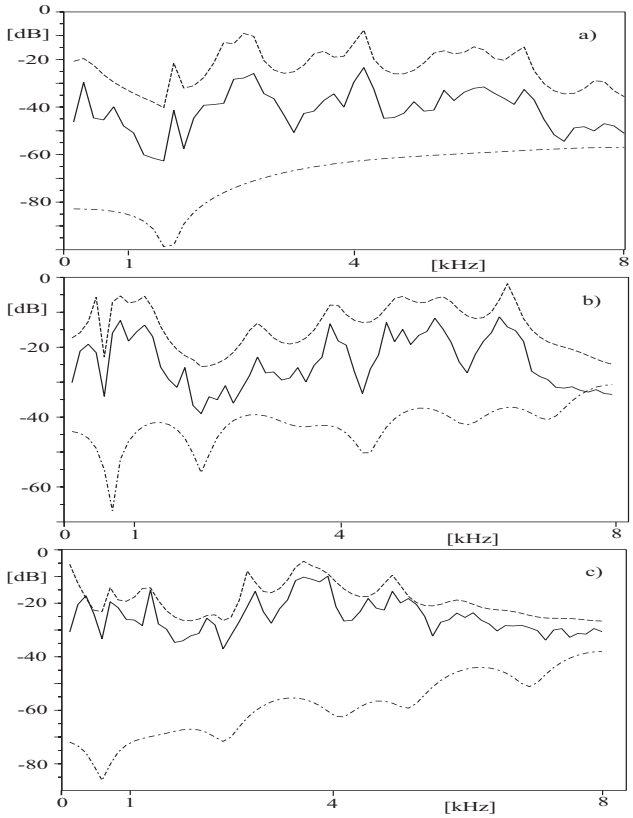

Bild 4: Analyse a) des Nasals /n/, b) nasalierten Vokals / $\tilde{a}$ / und c) des nasalierten Schwalautes: Geschätzter Betragsgang (unterbrochene obere Linie), DFT der gemittelten Periode mit Präemphase (durchgezogene mittlere Linie) und geschätzter Betragsgang der Nullstellen (strich-punktierte untere Linie).

Zusammenfassung

Für die Parameterbestimmung können die Nullstellen des verzweigten Rohrsystem zuerst durch eine ARMA Schätzung ermittelt werden. Dadurch sind die Parameter des Seitenzweigs bestimmt. Die restlichen Parameter werden iterativ mittels inverser Filterung geschätzt. Im Gegensatz zur Burg-Methode wird die Ausgangsleistung dabei nicht direkt hinter dem zu schätzenden Tor, sondern am Filterausgang minimiert. Während für Nasale der Seitenzweig den Mundraum beschreibt, gibt der Seitenzweig für nasalierte Vokale den Einfluß des Nasaltrakts wieder. Die Analysen von Sprachsignalen zeigen an Hand von Spektren, daß durch die iterative inverse Filterung eine hinreichend gute Modellierung erreicht wird.

Literatur

[1] Schnell, K.; Lacroix, A.: „Erweiterte Rohrmodelle für die Sprachproduktion“, Tagungsband DAGA 1998, pp. 384-385.

[2] 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.

[3] Schnell, K.; Lacroix, A.: “Pole Zero Estimation from Speech Signals by an Iterative Procedure“, Proc. ICASSP-2001, Salt Lake City, USA, Vol. I, pp. 109-112, 2001.

[4] Schnell, K.; Lacroix, A.: „Analyse und Verwendung des Rohrmodells für die Spracherzeugung“, Tagungsband, DAGA 2001, pp. 566-567.

[5] Schnell, K.; Lacroix, A.: “Inverse Filtering of Tube Models with Frequency Dependent Tube Terminations“, Proc. EUROSPEECH-2001, Aalborg, Denmark, pp. 2467-2470, 2001.