Analyse und Verwendung des Rohrmodells für die Spracherzeugung

K. Schnell, A. Lacroix

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

Einleitung

Das zeitdiskrete Rohrmodell, realisiert durch Kreuzgliedkettenfilter, kann verwendet werden, um den Sprechtrakt zu modellieren. Das in diesem Beitrag aufgezeigte Modell weist vorgeschriebene frequenzabhängige Rohrabschlüsse auf, so daß keine Standardschätzalgorithmen für die Analyse herangezogen werden können. Die Parameterbestimmung basiert auf dem Prinzip der inversen Filterung. Dafür wird ein modifizierter iterativer Schätzalgorithmus bereitgestellt. Neben der Analyse von Vokalen ist besonders die Analyse von Konsonanten interessant, da mit den geschätzten Konstriktionen Vokal-Konsonant-Vokal Übergänge modelliert werden können.

Rohrmodell

Ein Modell für die Ausbreitung akustischer ebener Wellen in einem Rohrsystem ist zeitdiskret durch ein Kreuzglied-Kettenfilter realisiert. Das Rohrmodell kann durch Betriebskettenmatrizen $\boldsymbol{T}_i$ beschrieben werden, die jeweils den i'ten Querschnittssprung, mit einem nachfolgendem Rohrelement enthalten. Der Querschnittssprung wird durch einen 2-Tor Adaptor beschrieben mit dem dazugehörigen Reflexionskoeffizient r_i

$$\boldsymbol{T}_i = k(r_i)\cdot\begin{pmatrix} 1 & r_i\cdot z^{-1} \\ r_i & z^{-1} \end{pmatrix}. \qquad (1)$$

$\boldsymbol{T}_i$ verknüpft die beiden Torgrößen vor und hinter einem Querschnittssprung mit Rohrelement. $k(r_i)$ ist dabei ein Vorfaktor, der die Art der Wellengrößen beschreibt. Das hier behandelte Rohrmodell ist unverzweigt und enthält frequenzabhängige Rohrabschlüsse an der Lippenöffnungsfläche auf der einen Seite und an der Verengung des Sprechtraktes an der Glottis auf der anderen Seite. Der Lippenabschluß kann durch ein Pol-Nullstellen-System modelliert werden [1]. Der Rohrabschluß an der Glottis wird in [2] durch einen festen reellen Reflexionkoeffizienten dargestellt. Hier wird ein Rohrabschluß durch ein System mit einer Nullstelle verwendet, so daß der Reflexionfaktor frequenzabhängig wird.

Parameterbestimmung

Um die Parameter des Rohrmodells aus dem Sprachsignal zu schätzen, ist es vorteilhaft, den Einfluß der Anregung und Abstrahlung aus dem Sprachsignal zu separieren. Dies wird durch eine adaptive Präemphase realisiert, die auch wiederholt angewendet werden kann. Für die Analyse des Rohrmodells werden einzelne Perioden von stimmhafter Sprache verwendet, die wegen der Separation von Anregung und Abstrahlung durch ein System mit reellen Nullstellen vorgefiltert werden. Die Analyse des Rohrmodells wird vollzogen durch eine Leistungsminimierung des Ausgangs von dem inversen Filter. Dabei muß berücksichtigt werden, daß die von z

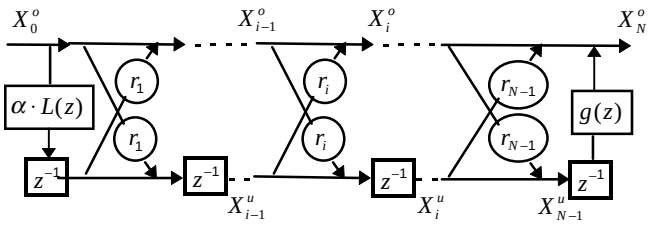

Bild 1: Inverses Filter des Rohrmodells.

unabhängigen Terme in der Übertragungsfunktion des inversen Filters nicht von den zu schätzenden Modellparametern abhängen; dies kann durch einen zusätzlichen Faktor sichergestellt werden [3]. Das inverse Rohrmodell ist in Bild 1 skizziert. Darin ist mit $L(z)$ der Rohrabschluß an den Lippen dargestellt, welcher durch einen Pol und eine Nullstelle modelliert werden kann [1]. Ein zusätzlicher Faktor α kleiner Eins kann Verluste im Sprechtrakt berücksichtigen. Dieser Rohrabschluß ist von der Mundöffnungsfläche abhängig und wird vor der Analyse individuell für jeden Laut eingestellt. Am Ausgang des inversen Filters ist der fest eingestellte Rohrabschluß an der Glottis durch eine Nullstelle in $g(z)$ realisiert. Diese beiden Rohrabschlüsse stellen feste Randbedingungen für die Parameterbestimmung dar, womit eine Analyse durch die Burg-Methode [4] in der Regel zu keinen zufriedenstellenden Ergebnissen führt. Um einen Reflexionskoeffizienten zu schätzen, wird hier im Gegensatz zu Burg nicht direkt hinter dem zu schätzenden Zweitor die Ausgangsleistung minimiert, sondern am Ausgang des inversen Filters [2]. Damit läßt sich der i'te Reflexionskoeffizient durch R_i optimal schätzen, unter der Bedingung das die anderen Koeffizienten als gegeben vorausgesetzt sind. In R_i sind die Erwartungswerte durch Zeitmittelwerte ersetzt worden. Die Berechnungsformel ergibt sich aus

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

zu

$$R_i = -\left(\overline{o_i^{11}\cdot o_i^{12}} + \overline{o_i^{11}\cdot u_i^{11}} + \overline{u_i^{12}\cdot o_i^{12}} + \overline{u_i^{12}\cdot u_i^{11}}\right) \Big/ \left(\overline{o_i^{12}\cdot o_i^{12}} + 2\cdot\overline{o_i^{12}\cdot u_i^{11}} + \overline{u_i^{11}\cdot u_i^{11}}\right) \qquad (2)$$

mit

$$\boldsymbol{F}_i = \begin{pmatrix} F_i^{11} & F_i^{12} \\ F_i^{21} & F_i^{22} \end{pmatrix} = \begin{pmatrix} 1 & g(z)\cdot z^{-1} \\ 0 & 0 \end{pmatrix} \cdot \prod_{k=1}^{N-i-1} \begin{pmatrix} 1 & r_{N-k}\cdot z^{-1} \\ r_{N-k} & z^{-1} \end{pmatrix} \qquad (3)$$

und $o_i^{uv}(n) = f_i^{uv}(n) * x_{i-1}^o(n)$, $u_i^{uv}(n) = f_i^{uv}(n) * x_{i-1}^u(n-1)$.

Für $i = N-1$ wird in (3) auf der rechten Seite das Produkt gleich der Einheitsmatrix gesetzt. Zu Beginn des Algorithmus' werden alle r_i für $i = 1 \ldots N-1$ gleich Null gesetzt. Die Koeffizienten werden dann nach (2) von r_1 beginnend bis r_{N-1} berechnet, was eine Iteration darstellt. Bei der nächsten Iteration werden die Ergebnisse der vorherigen in Bezug auf die r_i benutzt, wodurch sich die Ergebnisse weiter verbessern. Für die Analyse werden einzelne Perioden analysiert, die links und rechts für die Berechnung periodisch fortgesetzt werden, so daß der Schätzalgorithmus unabhängig von der Phase des zu analysierenden Signals ist.

Analyse von Testsignalen

Durch Analyse von Testsignalen kann überprüft werden, ob der Algorithmus in der Lage ist unter vordefinierten Bedingungen das globale Minimum zu erreichen. Durch Anregung eines Rohrsystems durch eine Impulsfolge wird ein Ausgangssignal erzeugt, welches als Testsignal fungiert. Dies wird mit der selben Struktur des Rohrmodell analysiert, wobei die Koeffizienten r_i unbekannt sind und nur die

gleichen Rohrabschlüsse wie bei der Generierung der Testsignale verwendet werden. In Bild 2 sind die Betragsgänge der geschätzten Rohrmodelle versetzt gezeigt im Vergleich zum Betragsspektrum des Testsignals. Die erste Iteration liefert noch unzureichende Ergebnisse, während nach mehreren Iterationen ein nahezu perfektes Ergebnis vorliegt.

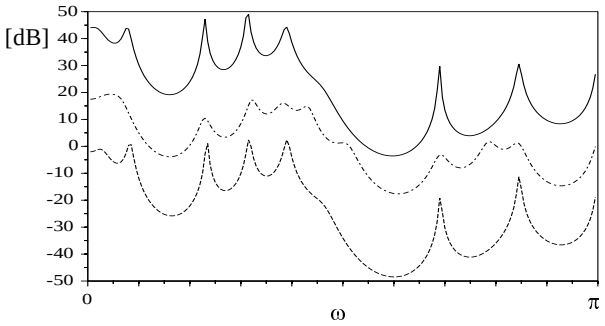

Bild 2: Durchgezogene Linie oben: Betragsspektrum des Testsignals; punktierte gestrichelte Linie mittig: Geschätzter Betragsgang des Rohrmodells nach der ersten Iteration; gestrichelte Linie unten: Geschätzter Betragsgang nach 50 Iterationen.

Analyse von Sprachlauten

Im Gegensatz zu den Vokalen weisen die Konsonanten in ihren stimmhaften Perioden mehr Schwankungen auf. Um diese Instationarität zu berücksichtigen, werden mehrere benachbarte Perioden, die zuvor auf die selbe Periodenlänge gebracht wurden, im Spektralbereich gemittelt. Anschließend wird durch eine IDFT ein Zeitsignal generiert, welches nun für die Analyse zur Verfügung steht. Für einen Sprachlaut können durch verschiedene Werte von N Analysen mit unterschiedlichen Vokaltraktlängen vorgenommen werden. Es wird die Vokaltraktlänge ausgewählt in Übereinstimmung mit dem betrachteten Laut bei einer kleinen Ausgangsleistung des inversen Filters. Es sind Fälle aufgetreten, bei denen die Vorfilterung korrigiert wurde, um die Flächenrelationen besser sichtbar zu machen. Bild 3 zeigt die geschätzten logarithmierten Querschnittsflächen des stimmhaften Lautes /z/, die zu dem stimmlosen Laut /s/ korrespondieren. Die Darstellung der Flächen ist logarithmisch gewählt, weil in dieser Darstellung die Konstriktion besonders gut zu erkennen ist. In Bild 4 ist der ermittelte Vokaltrakt des Lautes /Z/ dargestellt, der zu /S/ korrespondiert, jeweils in SAMPA Notation.

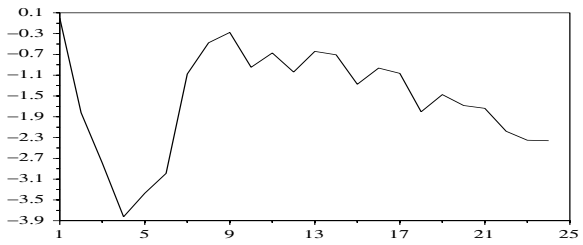

Bild 3: Geschätzte logarithmierte Vokaltraktflächen von /z/, links ist die Lippenöffnung und rechts die Glottis.

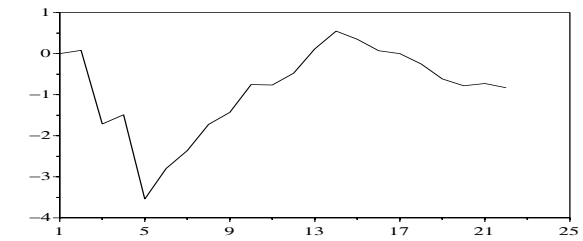

Bild 4: Geschätzte logarithmierte Vokaltraktflächen von /Z/, links ist die Lippenöffnung und rechts die Glottis.

Bild 5 und 6 zeigen die geschätzten logarithmierten Vokaltraktflächen der Laute /g/ und /d/. Der nachfolgende Vokal bei den hier untersuchten stimmhaften Explosiven ist der Schwa-Laut. In den Bildern 3-6 sind die Konstriktionen in ihrer Ausprägung sehr gut zu erkennen und befinden sich an unterschiedlichen Positionen. Entsprechend zu den Artikulationsstellen der Sprachlaute befinden sich die Konstriktionen von /Z/ und /g/ hinter denen von den Lauten /z/ bzw. /d/. Die Abtastrate der analysierten Sprachlaute lag bei 22kHz.

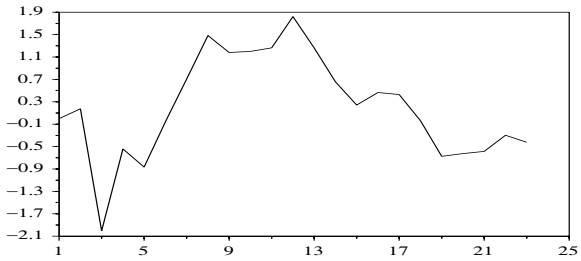

Bild 5: Geschätzte logarithmierte Vokaltraktflächen von /d/, links ist die Lippenöffnung und rechts die Glottis.

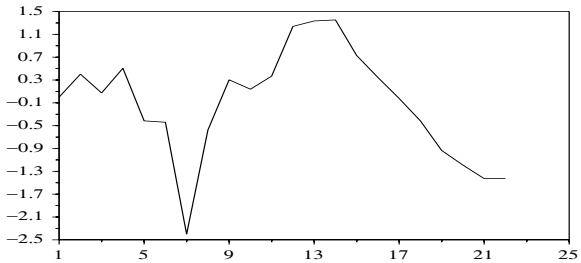

Bild 6: Geschätzte logarithmierte Vokaltraktflächen von /g/, links ist die Lippenöffnung und rechts die Glottis.

VCV Übergänge

Beispiele von synthetisierten Vokal-Konsonant-Vokal Übergängen zeigen, daß die gewonnenen Flächen mit den Konstriktionen sich in einer vokalischen Umgebung als Übergangselement einsetzen lassen. Die Integration der Konsonanten in den Flächenübergang geschieht im Wesentlichen entsprechend zu der in [5] vorgestellten Methode.

Zusammenfassung

Der in [2] vorgestellte iterative Schätzalgorithmus wird auf ein Rohrmodell mit zwei statt nur einem frequenzabhängigen Rohrabschluß angewandt. Dabei zeigen Testsignale nahezu optimale Ergebnisse. Bei der Analyse von Frikativen und stimmhaften Explosiven können in den geschätzten Flächen die Konstriktionsstellen im Vokaltrakt an den entsprechenden Stellen beobachtetet werden. Für die Analyse von instationären stimmhaften Konsonanten ist es von Vorteil, einzelne Perioden im Spektralbereich zu mitteln. Mit den ermittelten Vokaltraktflächen der Konsonanten ist es möglich, VCV Übergänge zu erzeugen.

Literatur

[1] Laine, U.K.: „Modeling of lip radiation impedance in the z-domain“, Proc. of ICASSP-82, Paris, pp. 1992-1995.

[2] Schnell, K.; Lacroix, A.: „Vokaltrakt-Schätzung unter Berücksichtigung einer reellen Glottisimpedanz“, ITG Fachbericht 161, Konvens-2000/ Sprachkommunikation , 2000, VDE-Verlag, pp. 279-284.

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

[4] Burg, J.: „A new Analysis Technique for Time Series Data“, NATO Advanced Study Inst. on Signal Proc., Enschede, 1968.

[5] Schnell, K.; Lacroix, A.: „Realisation of a Vowel-Plosive-Vowel Transition by a tube model“, Proc. of EUSIPCO-2000, Tampere, Finland, 2000, pp. 757-760.