

Inverse Filterung von Rohrmodellen mit verteilten Verlusten für die Schätzung des Sprechtraktes

Karl Schnell, Arild Lacroix

Institut für Angewandte Physik, J. W. Goethe-Universität Frankfurt am Main

Email: {Schnell, Lacroix}@iap.uni-frankfurt.de

Einleitung

Mit zeitdiskreten Rohrmodellen kann die Ausbreitung ebener Wellen im Sprechtrakt modelliert werden; daher sind sie für die Sprachverarbeitung von großem Interesse. Die Wellenausbreitung wird für die Modelle meist verlustlos angenommen. In diesem Beitrag wird eine Modellierung von verteilten frequenzabhängigen Verlusten mittels eines Pol-Nullstellen-Systems vorgestellt, welche die Wandvibrationen, viskose Reibung und Wärmeleitung berücksichtigt. Für ein solches verlustbehaftetes Rohrmodell werden die Reflexionskoeffizienten aus dem Sprachsignal mittels eines Optimierungsalgorithmus bestimmt. Die Schätzungen zeigen auf, daß das Sprachspektrum durch den Modellbetragsgang gut approximiert werden kann und daß die geschätzten Vokaltraktflächen einen realistischen Verlauf aufweisen.

Rohrelemente mit frequenzabhängigen Verlusten

Das zeitdiskrete Rohrmodell besteht aus einer Aneinanderreihung von Rohrelementen gleicher Länge, welche im verlustlosen Fall durch Laufzeitglieder realisiert werden. Da die Welle beim Propagieren durch ein Rohrelement eine Dämpfung erfährt, wird jedes Laufzeitglied z^{-1} mit einer Systemfunktion $V(z)$ multipliziert, das frequenzabhängige Verluste modelliert. Die Betriebskettenmatrix T beschreibt die linken Tor- bzw. Wellengrößen x_l^+ und x_l^- durch die rechten x_r^+ und x_r^- , wobei x^+ die vorlaufende und x^- die rücklaufende Welle darstellt. Für verlustbehaftete Rohrelemente ergibt sich die Betriebskettenmatrix zu:

$$T'_V = \begin{pmatrix} V^{-1/2} z^{1/2} & 0 \\ 0 & V^{1/2} z^{-1/2} \end{pmatrix} = V^{-1/2} z^{1/2} \begin{pmatrix} 1 & 0 \\ 0 & V z^{-1} \end{pmatrix} \quad \text{eq. 1}$$

$$T_V = z^{-1/2} T'_V \quad \begin{pmatrix} x_l^+ \\ x_l^- \end{pmatrix} = T_V \begin{pmatrix} x_r^+ \\ x_r^- \end{pmatrix}$$

Der Vorfaktor $z^{1/2}$ der Matrix T'_V stellt nur eine Gesamtlaufzeit dar und kann weggelassen werden, wodurch die Betriebskettenmatrix T_V verwendet wird mit nur ganzzahligen Potenzen von z^{-1} . Da die Laufzeit eines Rohrelementes $z^{-1/2}$ entspricht, beschreibt der Verlustfaktor $V(z)$ die Dämpfung für die doppelte Länge eines Rohrelementes. Der Betragsgang für den theoretischen Verlustfaktor W nach [1] infolge von zwei Rohrelementen mit einer Querschnittsfläche von 2 cm² ist für eine Abtastrate von 22 kHz in Bild 1 (unterbrochene Linie) gezeigt. Dabei sind die Verluste infolge Wandvibrationen, welche sich bei tiefen Frequenzen auswirken, sowie infolge viskoser Reibung und Wärmeleitung berücksichtigt, die einen schwachen spektralen Abfall zu hohen Frequenzen hin bewirken. Die Zahlenwerte der Verlustparameter wurden aus [1] übernommen, wobei der Verlustfaktor für sehr tiefe Frequenzen angehoben wurde. Für eine Modellierung der frequenzabhängigen Verluste soll das System $V(z)$ den theoretischen Verlustfaktor W approximieren. Dafür werden die Koeffizienten des rekursiven Systems $V'(z) = C(z)/D(z)$ durch einen Optimierungsalgorithmus ermittelt, welcher das spektrale Abstandsmaß

$$e(W, V') = \frac{1}{\pi} \int_0^\pi \left| \frac{W(\omega)}{V'(\omega)} \right|^2 d\omega \rightarrow \min \quad \text{eq. 2}$$

durch ein Gradientenverfahren minimiert. e stellt die Ausgangsleistung einer spektralen inversen Filterung dar. Der resultierende Betragsgang für $V(z) = \beta \cdot V'(z)$ mit einer Nullstelle und zwei Polen ist in Bild 1 (durchgezogene Linie) gezeigt. Die Gruppenlaufzeit des Systems weist ein Maximum von 0,04 auf und fällt für sehr tiefe Frequenzen ab. Um die Beträge der Verluste genau zu modellieren, wird für das in Bild 1 dargestellte V ein $\beta = 0,9925$ verwendet, das durch die Minimierung nach eq. 2 nicht bestimmt werden kann. Da die Verluste auch von der Rohrquerschnittsfläche abhängen, kann der reelle Faktor β lautabhängig eingestellt werden, entsprechend der mittleren Querschnittsfläche des jeweiligen Lautes. Das aus der Optimierung resultierende Zähler- und Nennerpolynom für $V(z) = \beta \cdot C(z)/D(z)$ ergibt sich zu:

$$C(z) = 1 - 0,9151 \cdot z^{-1} \quad \text{eq. 3}$$

$$D(z) = 1 - 0,91654 \cdot z^{-1} + 0,00329 \cdot z^{-2}$$

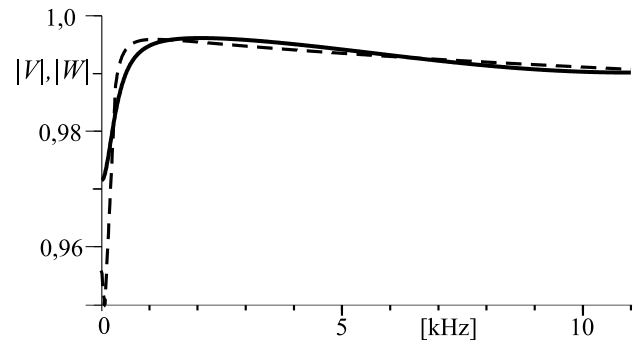


Bild 1: Betragsgang des Verlustfaktors $V(z)$ mit 2 Polen und einer Nullstelle (durchgezogene Linie) und des theoretischen Verlustfaktors W (unterbrochene Linie) für 22 kHz Abtastrate.

Rohrmodell des Vokaltraktes

Die Einheitsrohrelemente werden durch Adaptoren verknüpft, welche die Querschnittssprünge beschreiben. Die Betriebskettenmatrix eines Querschnittssprunges und eines Rohrelementes ergibt sich zu:

$$T_i = k(r_i) \cdot V^{-1/2} \begin{pmatrix} 1 & r_i V z^{-1} \\ r_i & V z^{-1} \end{pmatrix} \quad \text{mit} \quad \begin{pmatrix} x_i^+ \\ x_i^- \end{pmatrix} = T_i \begin{pmatrix} x_{i+1}^+ \\ x_{i+1}^- \end{pmatrix}, \quad \text{eq. 4}$$

wobei x_i^+ und x_i^- die vor- und rücklaufenden Wellen des i -ten Rohrelementes und r_i den dazugehörigen Reflexionskoeffizienten repräsentieren. Der reelle Faktor $k(r_i)$ spezifiziert die benutzte Wellenanalgie: Druck-, Fluß- oder Leistungswellen. Für die Modellierung des Vokaltraktes wird ein unverzweigtes Rohrmodell verwendet, das am rechten Systemausgang mit $\alpha \cdot L(z)$ abgeschlossen ist, um die Lippenabstrahlung zu berücksichtigen. Dafür wird das Pol-Nullstellensystem $L(z) = F(z)/G(z)$ von Laine [2] verwendet angepaßt für die verwendete Abtastrate. α beschreibt einen zusätzlichen reellen Verlustfaktor. Für die Berechnung der

Übertragungsfunktion werden sämtliche Betriebskettenmatrizen T_i zur Matrix T_g aufmultipliziert. Mit den beiden oberen Elementen T_g^{11} und T_g^{12} der Matrix T_g ergibt sich die Übertragungsfunktion für ein Vokaltraktmodell der Länge N zu:

$$H(z) = \frac{G}{GT_g^{11} + \alpha FT_g^{12}} \quad \text{eq. 5}$$

Das Zähler- und Nennerpolynom von H weist infolge hoher Potenzen von C und D Polynomkoeffizienten auf, die sehr starke Unterschiede in ihren Größenordnungen aufweisen. Dadurch ergeben sich numerische Fehler ab Rohrlängen von ca. 18 und darüber, auch wenn mit doppelter Genauigkeit (double) gerechnet wird. Die numerischen Fehler können vermieden werden, falls für die Auswertung von $H(e^{j\omega})$ die Werte $C(e^{j\omega})$ und $D(e^{j\omega})$ als Quotient $C(e^{j\omega})/D(e^{j\omega})$ berechnet werden. Dies kann effizient gestaltet werden, da viele Terme nur einmal berechnet werden müssen. Für die Realisierung im Zeitbereich ist eine abwechselnde Anordnung der verlustbehafteten Laufzeitglieder oben und unten im Signalflußpfad empfehlenswert, um numerische Fehler auszuschließen.

Vokaltraktsschätzungen aus Sprachsignalen

Für die Analyse von Sprachlauten werden die Reflexionskoeffizienten des Rohrmodells aus dem Sprachsignal geschätzt. Das Sprachsignal s' wird mit einer zum Teil mehrfach angewandten adaptiven Präemphase vorgefiltert, damit der Einfluß der stimmhaften Anregung und der Abstrahlung auf die spektrale Einhüllende beseitigt wird. Das resultierende gefilterte Sprachsignal s wird für die Analyse verwendet. Vor der Schätzung müssen die Anzahl der Rohrelemente und der Lippenabschluß $\alpha \cdot L(z)$ festgelegt werden. Die Reflexionskoeffizienten werden durch eine Minimierung des spektralen Abstandsmaßes $e(S,H)$ analog zu eq.2 geschätzt. Dafür müssen die Vorfaktoren $k(r_i)$ in T_i gleich Eins gesetzt werden, da sonst nach [3] ein Korrekturfaktor für die Fehlerdefinition vonnöten ist. Bild 2 zeigt den geschätzten Betragsgang des verlustbehafteten Rohrmodells und das Betragsspektrum des analysierten Vokals /u:/. Der geschätzte Betragsgang des Rohrmodells im verlustlosen Fall mit $V=1$ unterscheidet sich kaum von dem des verlustbehafteten Modells. Wie in Bild 3 zu sehen ist, weisen jedoch die geschätzten Vokaltraktflächen des Lautes /u:/ mit und ohne ver-

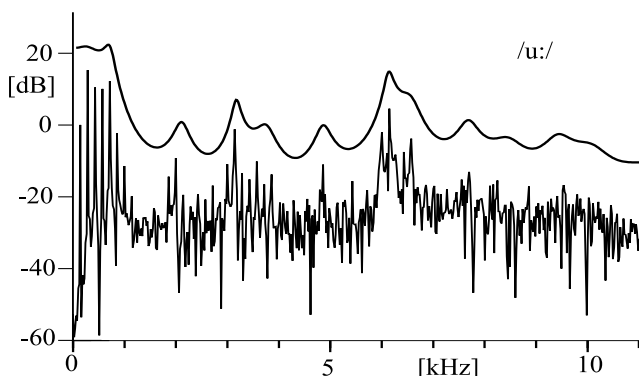


Bild 2: Geschätzter Betragsgang des verlustbehafteten Rohrmodells (obere Linie) und Betragsspektrum des analysierten Vokals /u:/ mit Präemphase (unten).

teilte Verluste einen deutlichen Unterschied auf, was mit den untersten zwei Formanten des Vokals /u:/ zusammenhängt, da niederfrequente Resonanzen durch die Wandvibrationen in ihren Bandbreiten und Frequenzen beeinflusst werden [4]. Ein Vergleich mit den

Vokaltraktflächen in Bild 3 (c) des Lautes /u:/, welche durch MRT ermittelt wurden [5], zeigt, daß die geschätzten Flächen mit Berücksichtigung der verteilten Vokaltraktverluste einen realistischen Verlauf aufweisen im Gegensatz zum verlustlosen Fall. Die Flächen aus [5] wurden von einem anderen Sprecher ermittelt und dienen hier nur als erster Anhalt. Es sei angemerkt, daß bei anderen untersuchten Lauten nicht derart gravierende Unterschiede in den Vokaltraktflächen beobachtet wurden.

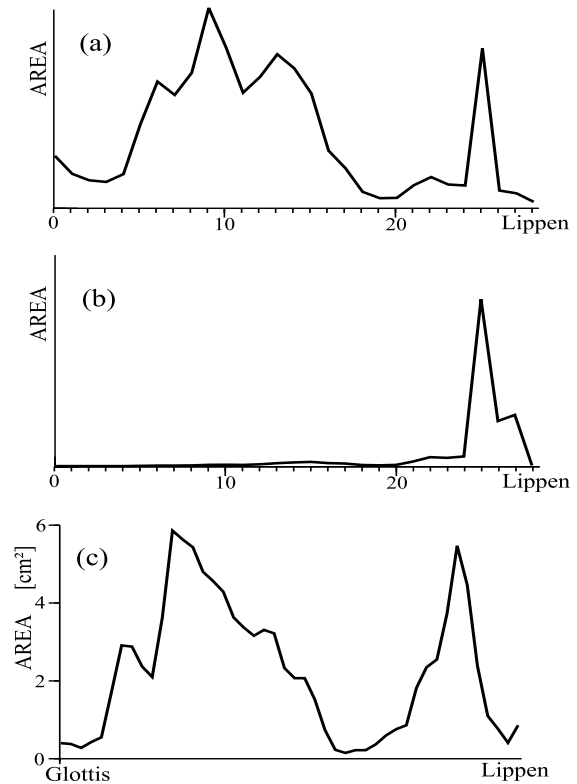


Bild 3: Vokaltraktflächen des Vokals /u:/: Geschätzt mit dem verlustbehafteten Rohrmodell (a), mit dem verlustlosen Rohrmodell (b) und aus MRT-Aufnahmen [5] ermittelt (c).

Zusammenfassung

Frequenzabhängige Verluste im Vokaltrakt können durch zusätzliche Pol-Nullstellen-Systeme im Rohrmodell beschrieben werden, welche mit den Laufzeitgliedern kombiniert werden. Die Parameter der Verlustsysteme können durch einen Optimierungsalgorithmus an Literaturwerte angepaßt werden. Die Modellierung mit bzw. ohne Verlusten führt bei den geschätzten Betragsgängen in der Regel nur zu geringen Unterschieden. Bei den geschätzten Vokaltraktflächen können jedoch lautabhängig zum Teil gravierende Unterschiede auftreten, so daß die Berücksichtigung von verteilten Verlusten insgesamt zu realistischen Querschnittsverläufen führt.

Referenzen

- [1] Liljencrants J.: „Speech Synthesis with a Reflection-Type Line Analog“, Dissertation, Royal Inst. of Technology, Stockholm 1985.
- [2] Laine, U.K.: „Modeling of lip radiation impedance in the z-domain“, Proc. ICASSP'82, Paris 1982, pp. 1992-1995.
- [3] Schnell K.; Lacroix A.: „Erweiterte Rohrmodelle für die Sprachproduktion“, Tagungsband DAGA'98, pp 384-385, 1998.
- [4] Sondhi M.M.: „Model for wave propagation in a lossy vocal tract“, J.Acoust.Soc.Am. 1974 (55) pp. 1070-1075.
- [5] Story B.H. et al.: „Vocal Tract Area Functions from Magnetic Resonance Imaging“, J.Acoust.Soc.Am. 1996 (100), pp. 537-554.