

Modellierung des subglottalen Systems für die artikulatorische Sprachsynthese

Peter Birkholz, Dietmar Jackèl

Institut für Informatik, Universität Rostock, 18051 Rostock, Email: piet@informatik.uni-rostock.de

Einleitung

Diese Arbeit beschreibt die Nachbildung des subglottalen Systems durch ein stark vereinfachtes Rohrmodell für die Verwendung in der artikulatorischen Sprachsynthese. In der Literatur existieren unterschiedliche Ansätze zur Modellierung des subglottalen Systems. Einerseits lässt es sich analog zum supraglottalen System mit Hilfe einer Flächenfunktion beschreiben, die weitestgehend die akustischen Eigenschaften des Systems bestimmt [6]. Dabei werden die Querschnittsflächen der verzweigten Luftwege in einem bestimmten Abstand von der Glottis jeweils zu einem Gesamtquerschnitt zusammengefasst und der Querschnittsverlauf in ein diskretes Rohrmodell überführt. Alternativ kann das subglottale System direkt durch *RLC*-Schwingkreise nachgebildet werden, die jeweils einen subglottalen Formanten erzeugen [1]. Diese Methode ist zwar sehr schnell in der Simulation, sie unterscheidet sich aber konzeptionell von der üblichen „indirekten“ akustischen Modellierung des supraglottalen Vokaltrakts in Form eines diskreten Rohrmodells [3].

Zielstellung

Das Ziel der vorliegenden Arbeit war die effiziente und physiologisch plausible Modellierung des subglottalen Systems als diskretes Rohrmodell mit möglichst wenigen Rohrabschnitten, das sowohl realistische Werte für den DC-Strömungswiderstand als auch für die Frequenzen und Amplituden der ersten drei subglottalen Formanten erzeugt. Frühere Nachbildungen des subglottalen Systems [4, 6] verwenden verhältnismäßig viele Rohrabschnitte und sind dadurch rechenintensiv.

Der Strömungswiderstand der unteren Luftwege ist laut Vincent *et al.* [8] eine Funktion der Vitalkapazität (VC) der Lunge und liegt für eine VC zwischen 80% und 20% zwischen 1.0 und 2.0 cgs Ω (1 cgs Ω = 1 cm H₂O/l/s). Unser Rohrmodell soll einen Strömungswiderstand von ca. 1.5 cgs Ω besitzen. Als Zielwerte für die ersten drei Formantfrequenzen verwenden wir die von Badin *et al.* [1] angegebenen Werte von $F_1 = 615$ Hz, $F_2 = 1355$ Hz und $F_3 = 2110$ Hz, die im Einklang mit Messungen in verschiedenen Studien stehen [5, 6]. Die Amplitude der subglottalen Eingangsimpedanz an der ersten Resonanzfrequenz liegen laut Ishizaka *et al.* [6] zwischen 20 und 40 cgs Ω und laut Habib *et al.* [5] bei etwa 50 cgs Ω . Für die Impedanz der zweiten Resonanz ist gemäß beiden Untersuchungen ein Wert von 50 cgs Ω typisch, während die Impedanz der dritten Resonanz stark um 50 cgs Ω herum streut. Die Zielwerte für unser Rohrmodell orientieren sich mit den Impedanzen $Z_1 = 30$ cgs Ω , $Z_2 = 50$ cgs Ω

und $Z_3 = 50$ cgs Ω an den Messungen von Ishizaka *et al.* [6].

Vorgehensweise

Im Unterschied zum Rachen- und Mundraum handelt es sich bei den unteren Luftwegen um ein stark verzweigtes Luftwegesystem aus annähernd zylindrischen Zweigrohren, das sich über 23 Generationen näherungsweise wie ein Binärbaum verzweigt [6] (siehe Abb. 1a). Die

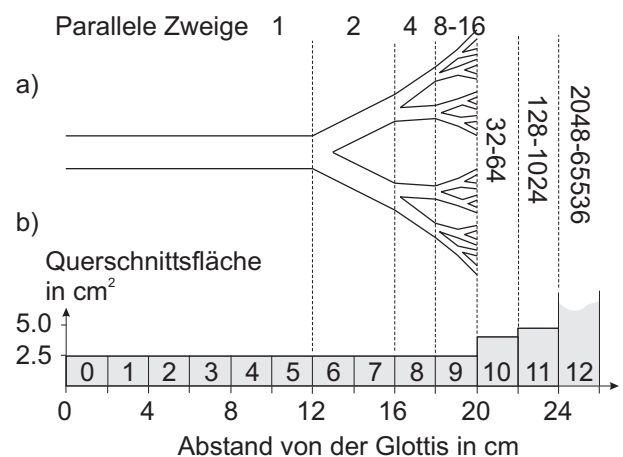


Abbildung 1: Baumstruktur des subglottalen Systems und Nachbildung als diskretes Rohrmodell.

Akustik eines derart verzweigten Rohrsystems wäre nur mit hohem Rechenaufwand direkt zu simulieren. Deshalb ist es üblich, dieses System analog zum supraglottalen System mit Hilfe eines einzelnen unverzweigten Rohres nachzubilden. Wenn man von einem reibungsfreien Rohr mit schallhart reflektierenden Wänden ausgeht, so müssen für ein akustisch äquivalentes Einzelrohr lediglich die Querschnitte aller „parallelen“ Zweige des Bronchialbaums in einem bestimmten Abstand zur Glottis jeweils zu einem Gesamtquerschnitt addiert werden. Das Rohrmodell, das diese Flächenfunktion repräsentiert, haben wir aus 13 Rohrabschnitten mit jeweils konstantem Querschnitt zusammengesetzt (siehe Abb. 1b). Jeder Rohrabschnitt hat eine Länge von 2 cm, so dass sich eine Gesamtlänge von 26 cm ergibt, die typisch für die unteren Luftwege ist. Der Querschnitt der ersten 10 Rohrabschnitte beträgt etwa 2.5 cm² und kann direkt auf das Rohrmodell übertragen werden [6]. Die drei letzten Rohrabschnitte überdecken dagegen jeweils mehrere Teilungsgenerationen mit stark variierenden Gesamtquerschnitten und müssen jeweils einen geeigneten mittleren Querschnitt annehmen, um das gewünschte akusti-

sche Verhalten des subglottalen Systems hervorzubringen. Für eine realistische akustische Simulation ist es auch notwendig, den Reibungswiderstand und die Elastizität der Rohrwände zu berücksichtigen. Hierbei ist zu beachten, dass n zylindrische Zweigrohre mit einem Gesamtquerschnitt A insgesamt einen größeren Umfang besitzen als ein *einzelnes* zylindrisches Rohr mit dem Querschnitt A . Daher gehen wir in unserem Modell gegenüber den üblichen Ansätzen mit runden Rohrquerschnitten von elliptischen Querschnitten mit separat einstellbarem Flächeninhalt *und* Umfang aus. Wenn ein Einzelrohr mit der Fläche A , das n gleichartige parallele Zweigrohre mit den Querschnitten A/n repräsentiert, den Umfang $S = \sqrt{n} \cdot 2 \cdot \sqrt{A} \cdot \pi$ besitzt, dann ergeben sich die gleichen Werte für die akustische Impedanz der elastischen Wände und für den Strömungswiderstand wie für das entsprechende Bündel aus Zweigrohren. Dementsprechend lässt sich über die Exzentrizität der elliptischen Querschnitte der subglottalen Rohrabschnitte die Parallelität der Zweigrohre des Bronchialbaums erfassen. Für die Rohrabschnitte 0...8 unseres Rohrmodells lässt sich obige Gleichung für den Umfang direkt anwenden. Für die übrigen Rohrabschnitte muss wieder ein geeigneter Mittelwert für den Umfang gefunden werden, damit die gewünschten akustischen Eigenschaften entstehen. Einen starken Einfluss auf die subglottalen Resonanzen besitzen auch die mechanischen Eigenschaften der elastischen Rohrwände des subglottalen Systems, insbesondere die Wandmasse M_w und die „interne Wandreibung“ B_w (jeweils pro Flächeneinheit) [6]. Um die Anzahl der einstellbaren Parameter klein zu halten, verwenden wir für alle Rohrabschnitte die gleichen Wandeigenschaften. Insgesamt hatten wir also 9 Parameter hinsichtlich der gewünschten akustischen Eigenschaften zu optimieren (die Querschnittsflächen A_{10} , A_{11} und A_{12} , die Umfangslängen S_9 bis S_{12} sowie M_w und B_w).

Die akustischen Eigenschaften für einen bestimmten Satz an Parameterwerten haben wir aus dem Spektrum der Eingangsimpedanz bestimmt, das mit Hilfe des analogen elektrischen Netzwerks für das Rohrmodell berechnet wurde (ähnlich der Methode in [2]). Am peripheren Ende haben wir das analoge elektrische Netzwerk für das subglottale System mit einem Kurzschluss abgeschlossen, der vereinfacht die hohe akustische Kapazität der Lungenbläschen nachbildet.

Für die Optimierung der Parameter wurde die Methode des Simulated Annealing [7] verwendet. Minimiert wurde die Summe der quadratischen Differenzen zwischen den tatsächlichen und den gewünschten akustischen Parametern.

Ergebnisse

Das beste Ergebnis aus mehreren Durchläufen des Optimierungsalgorithmus mit unterschiedlichen Startwerten und Verfahrensparametern ist die in Abb. 2 dargestellte Eingangsimpedanz des subglottalen Systems. Die ersten drei Formanten haben die Frequenzen 587 Hz, 1332 Hz und 2031 Hz und die Amplituden 29.8 cgs Ω , 51.3 cgs Ω

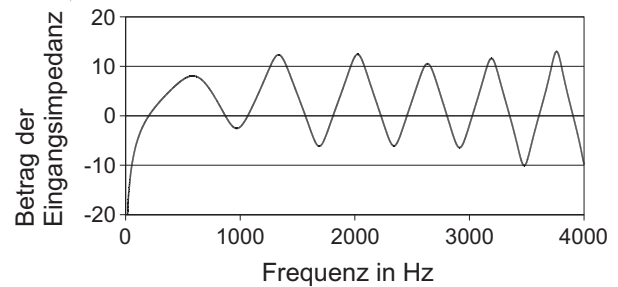


Abbildung 2: Betrag der Eingangsimpedanz des simulierten subglottalen Systems.

und 53.1 cgs Ω . Der DC-Strömungswiderstand beträgt 1.52 cgs Ω . Damit sind die Abweichungen zu den Zielformantfrequenzen 4.6%, 0.2% und 11.3% und zu den Zielimpedanzen 0.7%, 2.6% und 6.2%. Auch der DC-Strömungswiderstand liegt dicht beim mittleren Wert von 1.5 cgs Ω . Die optimierten Parameterwerte sind: $A_{10} = 4 \text{ cm}^2$, $A_{11} = 4.7 \text{ cm}^2$, $A_{12} = 20 \text{ cm}^2$, $S_9 = 4 \cdot \hat{S}_9$, $S_{10} = 8 \cdot \hat{S}_{10}$, $S_{11} = 64 \cdot \hat{S}_{11}$, $S_{12} = 256 \cdot \hat{S}_{12}$, $M_w = 2.7 \text{ kg/m}^2$ und $B_w = 10000 \text{ kg/m}^2\text{s}$. \hat{S}_i ist jeweils der Umfang für ein *zylindrisches* Rohr mit dem Querschnitt A_i . Unsere Ergebnisse zeigen, dass sich die wesentlichen akustischen Eigenschaften des subglottalen Systems durch ein stark vereinfachtes aber dennoch physiologisch plausibles Rohrmodell nachbilden lassen. Gegenüber den bisherigen Rohrmodellen für das subglottale System mit 20-30 Rohrabschnitten [4, 6] benötigt unser Modell nur 13 Abschnitte (Rechenzeiteinsparung) und liefert nicht nur für die Resonanzen realistische Werte, sondern auch für den DC-Strömungswiderstand.

Literatur

- [1] P. Badin and G. Fant. Notes on vocal tract computation. *STL-QPSR*, 2-3:53-108, 1984.
- [2] P. Birkholz and D. Jackèl. Influence of temporal discretization schemes on formant frequencies and bandwidths in time domain simulations of the vocal tract system. In *Interspeech 2004-ICSLP*, Jeju, Korea, 2004.
- [3] P. Birkholz and D. Jackèl. Artikulatorische Sprachsynthese mit dem Programm TractSyn - Ein Überblick. In *Fortschritte der Akustik, DAGA '05*, 2005.
- [4] P. Boersma. *Functional Phonology*. PhD thesis, University of Amsterdam, 1998.
- [5] R. H. Habib, R. B. Chalker, B. Suki, and A. C. Jackson. Airway geometry and wall mechanical properties estimated from subglottal input impedance in humans. *Journal of Applied Physiology*, 77(1):441-451, 1994.
- [6] K. Ishizaka, M. Matsudaira, and T. Kaneko. Input acoustic-impedance measurement of the subglottal system. *Journal of the Acoustical Society of America*, 60(1):190-197, 1976.
- [7] S. Kirkpatrick, C. D. Gelatt, and M. P. Vecchi. Optimization by simulated annealing. *Science*, 220:671-680, 1983.
- [8] N. J. Vincent, R. Knudson, D. E. Leith, P. T. Macklem, and J. Mead. Factors influencing pulmonary resistance. *Journal of Applied Physiology*, 29(2):236-243, 1970.