

# MODELLE DER NASALIERUNG IN SPEAK

*F. Ranostaj, A. Lacroix*

*Institut für Angewandte Physik, Johann Wolfgang Goethe-Universität  
lacroix@iap.uni-frankfurt.de*

**Abstract:** SPEAK ist ein Experimentalsystem zur Untersuchung und Demonstration der Sprechakustik. In diesem Beitrag wird gezeigt, wie man mit SPEAK unterschiedlicher Modelle des Nasaltrakts realisieren kann: ein Modell des Nasaltrakts basiert auf Wellendigitalfilter, während das andere Modell Impulsantworten verwendet, die durch eine dreidimensionale numerische Simulation der Schallausbreitung im Nasaltrakt bestimmt werden. Verschiedene Effekte der Nasalierung werden mit den Analysefunktionen von SPEAK demonstriert.

## 1 Einleitung

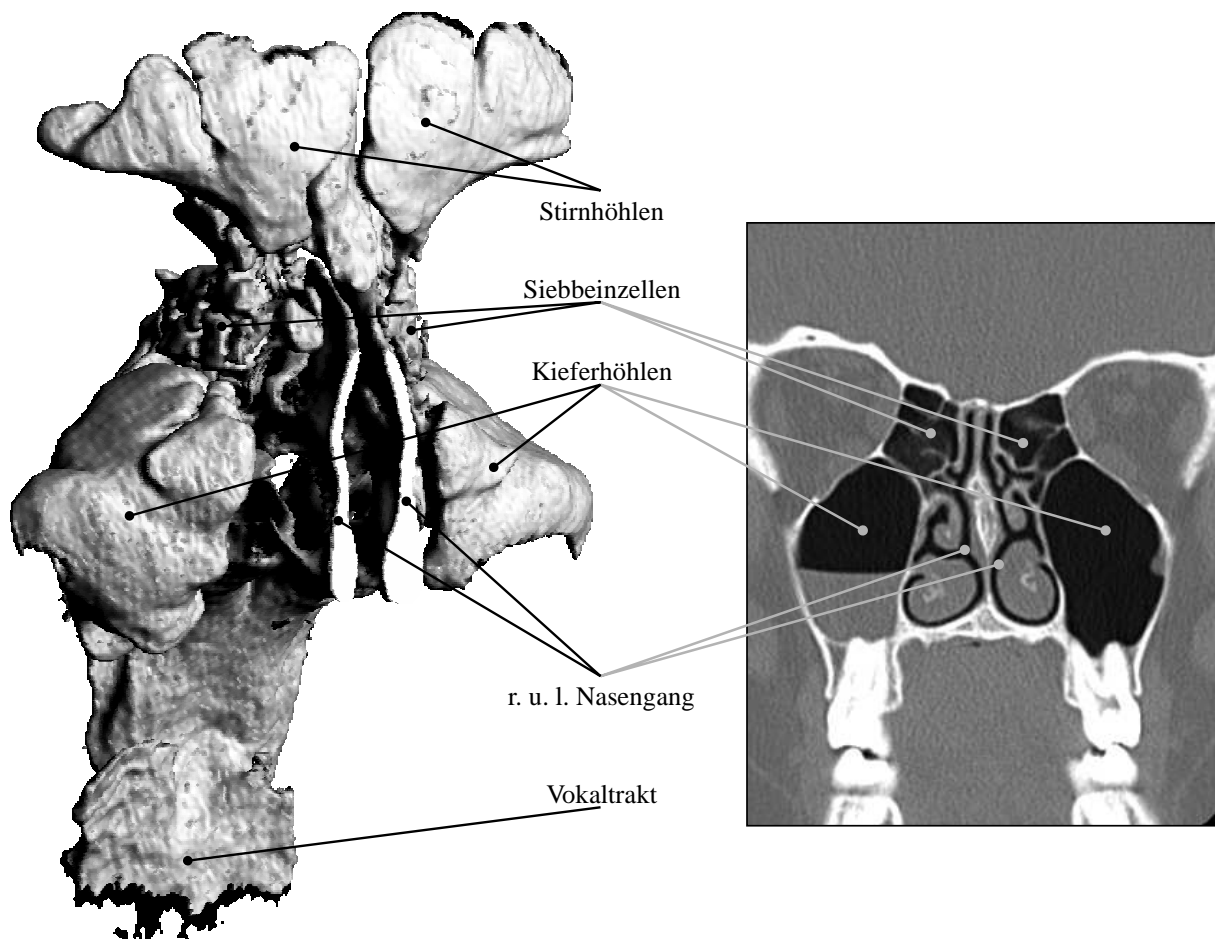
Zum Verständnis der menschlichen Sprachproduktion sind genaue Kenntnisse der akustischen Vorgänge im Sprechtrakt notwendig [1]. Der Sprechtrakt lässt sich hierfür in Vokal- und Nasaltrakt untergliedern, für die jeweils geeignete Verfahren zur Untersuchung der akustischen Vorgänge eingesetzt werden können. Die damit gewonnenen Beschreibungen des Vokal- und Nasaltrakts lassen sich in dem Softwaresystem SPEAK kombinieren und untersuchen. Im folgenden werden diese Verfahren vorgestellt und die Umsetzung der Beschreibungen in SPEAK demonstriert.

## 2 Nasaltrakt

Bei dem Nasaltrakt handelt es sich um ein verzweigtes akustisches System. Der Schall breitet sich darin von dem Velum kommend über die beiden Nasengänge zu den Nasenlöchern aus. Die Aufspaltung der Schallwege und die Gestalt der Nasengänge beeinflussen die Schallausbreitung maßgeblich. Wie in Bild 1 zu sehen ist, wird die Gestalt von jeweils drei muschelförmigen Einbuchtungen (Conchen) dominiert. Aufgrund der sich daraus ergebenden geringen Weite der Nasengänge tritt außerdem eine signifikante Dämpfung der Schallwelle auf. Des Weiteren wird die Schallausbreitung im Nasaltrakt durch Helmholtzresonanzen mehrerer Nebenhöhlen beeinflusst, die über dünne Kanäle mit den Nasengängen verbunden sind.

Zur Bestimmung der akustischen Eigenschaften wird für ein dreidimensionales Modell des Nasaltrakts [2] die Schallausbreitung numerisch ermittelt. Hierfür wird das Modell des Nasaltrakts an der Velumöffnung mit einem Schallimpuls angeregt. Die Schallausbreitung wird durch eine nach der Methode der finiten Differenzen diskretisierten Wellengleichung

$$\begin{aligned} \frac{\partial^2 p}{\partial t^2} &= c^2 \Delta p \\ \approx \quad p_{t+1,x,y,z} - 2p_{t,x,y,z} + p_{t-1,x,y,z} &= k^2 \left( p_{t,x+1,y,z} - 2p_{t,x,y,z} + p_{t,x-1,y,z} \right. \\ &\quad \left. + p_{t,x,y+1,z} - 2p_{t,x,y,z} + p_{t,x,y-1,z} \right. \\ &\quad \left. + p_{t,x,y,z+1} - 2p_{t,x,y,z} + p_{t,x,y,z-1} \right) \\ \iff \quad p_{t+1,x,y,z} &= 2p_{t,x,y,z} - p_{t-1,x,y,z} + k^2 \left( \dots \right) \end{aligned}$$



**Bild 1** - Links: Oberflächendarstellung der Hohlräume des Nasaltrakts. Die Velumöffnung liegt direkt über dem hier sichtbaren Teil des Vokaltrakts, dem Ansatz des Rachens. Rechts: Computertomographisches Schnittbild des Nasaltrakts. Schwarze Bereiche stellen Hohlräume dar. Mittig zu sehen ist die Nasenscheidewand, die die beiden Nasengänge trennt. Die Verengung des rechten und linken Nasenganges durch die drei Conchen wird deutlich.

sukzessive berechnet, wobei  $p$  den Druck des Schallfeldes an entsprechenden Orten,  $c$  die Schallgeschwindigkeit und  $k$  das Verhältnis zwischen räumlicher und zeitlicher Diskretisierungsweite bezeichnet. Die feinen Strukturen im Nasaltrakt erfordern eine räumlich isometrische Diskretisierung von 0,5 mm. Für die Simulation werden die Wände des Nasaltrakts in guter Näherung als schallhart angenommen. Die Dämpfung der Schallausbreitung entlang der Wände des Nasaltrakts wird durch eine Erweiterung des Finiten-Differenzen-Operators berücksichtigt, wie in [3] ausgeführt. Die sich ergebenden Impulsantworten werden an den Nasenlöchern und an der Velumöffnung aufgezeichnet. Auf diese Weise werden Transmittanz und Reflektanz erfasst.

### 3 Vokaltrakt

Der Vokaltrakt ist einfacher geformt, er lässt sich durch ein Rohr mit variierendem Querschnitt modellieren. Aufgrund des geringen Durchmessers genügt es, die Wellenausbreitung eindimensional, entlang der Rohrachse, zu betrachten. Konzentriert man die Querschnittsänderungen an äquidistanten Stellen, lässt sich die Schallausbreitung in dem Rohrmodell mit einem Wellendi-

gitalfilter, dem Kreuzglied-Kettenfilter, beschreiben:

$$\mathbf{T} = \mathbf{T}_{Q_1} \mathbf{T}_R \mathbf{T}_{Q_2} \mathbf{T}_R \cdots \mathbf{T}_{Q_n} \mathbf{T}_R$$

mit

$$\mathbf{T}_{Q_i} = \frac{1}{1+r_i} \begin{pmatrix} 1 & r_i \\ r_i & 1 \end{pmatrix}, \quad \mathbf{T}_R = \begin{pmatrix} z^{-1/2} & 0 \\ 0 & z^{1/2} \end{pmatrix}.$$

Die Betriebskettenmatrizen beziehen sich auf in dem Rohr hin- und zurücklaufende Schallwellen. Die Matrix  $\mathbf{T}_{Q_i}$  beschreibt die Streuung der Schallwellen an dem  $i$ te Querschnittsprung, wobei der Reflexionsfaktor  $r_i = \frac{A_i - A_{i+1}}{A_i + A_{i+1}}$  ein Maß des Verhältnisses der angrenzenden Querschnittsflächen bzw. akustischen Impedanzen ist. Mit der Matrix  $\mathbf{T}_R$  wird die Laufzeit der Schallwelle in dem dazwischenbefindlichen homogenen Rohrsegment beschrieben. Die Beschreibung erfolgt im  $z$ -Bereich, wobei die Länge  $l$  eines Rohrsegments die verwendete Abtastrate  $f_s = c/2l$  bestimmt. Das Übertragungsverhalten eines Rohrmodells  $H(z) = 1/T_{22}$  ergibt sich aus dem Produkt  $\mathbf{T}$  der Betriebskettenmatrizen.

Der Vorteil dieses Modells liegt darin, dass die Filterparameter bzw. die Rohrquerschnitte aus Sprachsignalen geschätzt werden können. Dadurch ist es möglich, die Rohrkonfigurationen einer Vielzahl von Lauten und Lautübergängen zu erfassen. Die Parameterschätzung kann mittels der Burg-Methode erfolgen, die die Reflexionkoeffizienten so wählt, dass mit jedem Rohrsegment die Korrelation des Signals in einem inversen Filter sukzessive minimiert wird. Hierfür wird für einen Signalabschnitt  $x$  ein Hilfssignal  $v_n = -x_n$  gebildet. In jedem Schritt wird ein Reflexionskoeffizient  $r_i$  bestimmt und die Signale werden entsprechend dekorreliert:

$$r_i = \frac{-2 \sum x_n v_n}{\sum x_n^2 + \sum v_n^2}, \quad \begin{aligned} x_n &\leftarrow x_n + r_i v_{n-1} \\ v_n &\leftarrow v_{n-1} + r_i x_n \end{aligned}$$

Die Topologie des Wellenleiters kann durch Einfügen von Verzweigungen erweitert werden: So ist die Ankopplung des Nasaltrakts in diesem Modell durch einen Dreitoradaptor möglich.

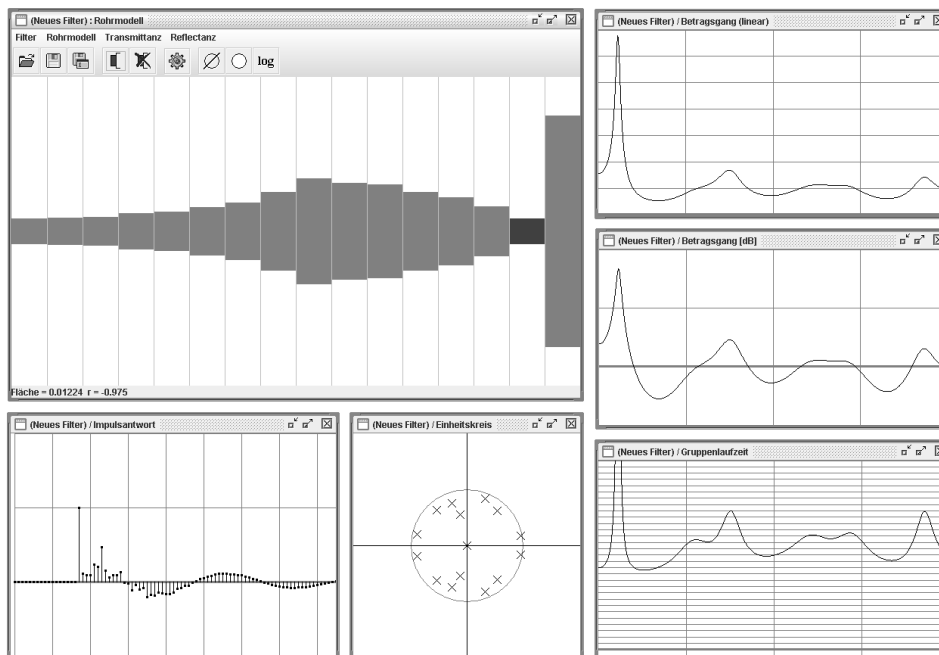
## 4 Schallabstrahlung und Anregung

Der Umfang von Mundöffnung und Nasenlöchern ist klein verglichen mit der Wellenlängen des für die Sprache relevanten Schalls. Deshalb können zur Modellierung der Schallabstrahlung an Mund und Nasenlöchern in guter Näherung ein schallweicher Abschluß und ein Differenzierer verwendet werden; alternativ kann ein geeigneter Querschnittsprung am Ende des Rohrmodells angefügt werden.

Für stimmhafte Laute erfolgt die Anregung durch ein Signal, welches aus dem zeitlichen Verlauf der Glottisöffnung ermittelt wird. Für die Erzeugung von stimmlosen Lauten kann weißes oder korreliertes Rauschen verwendet werden.

## 5 Experimentalsystem SPEAK

Das Experimentalsystem für Sprechakustik, abgekürzt SPEAK, erlaubt die Untersuchung und Demonstration verschiedener Modelle der Sprachproduktion. Um die Wirkung der genannten Elemente einzeln und in Kombination untersuchen zu können, erweist sich ein modular aufgebautes System als vorteilhaft. Hierin lassen sich die einzelnen Elemente oder Komponenten physikalisch beziehungsweise akustisch korrekt zusammenfügen und gestatten so die Realisierung verschiedener Modelle der Sprachproduktion. SPEAK ist mit einem Echtzeit-Synthesystem ausgestattet, das eine auditive Beurteilung von dem realisiertem Modell erlaubt.

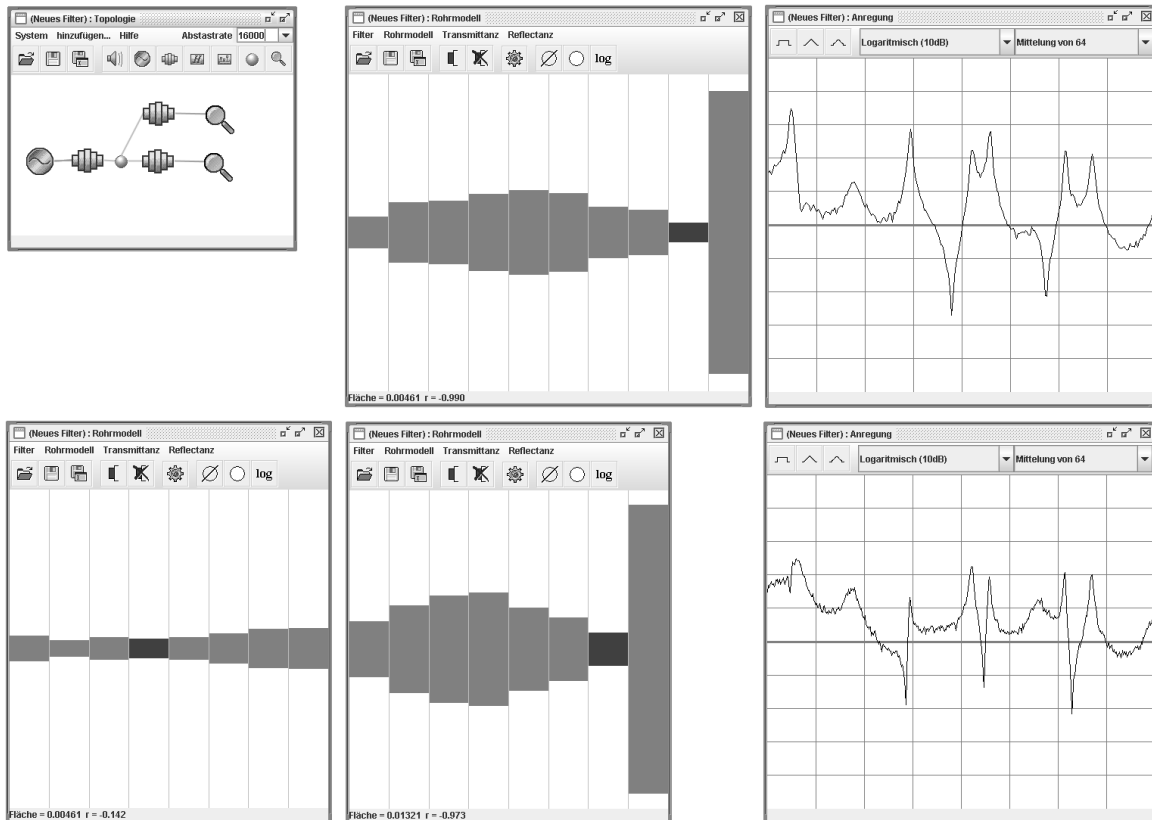


**Bild 2** - Unverzweigtes Rohrmodell mit verschiedenen Analysefunktionen. Rechts, von oben nach unten: Linearer und logarithmierter Betragsgang, Gruppenlaufzeit. Links unten: Impulsantwort; Mitte unten: Pol-Schema in der  $z$ -Ebene.

Hierbei wurde darauf geachtet, dass die einzelnen Komponenten sich über eine graphische Schnittstelle einfach parametrisieren lassen und diese Parametrisierung ohne wahrnehmbaren Zeitversatz audiovisuell umgesetzt wird. Die Darstellung einiger Eigenschaften beispielsweise eines Filterelements sind in Bild 2 für ein Rohrmodell gezeigt. Darüber hinaus ist auch der Phasengang und der dreidimensional dargestellte Betrag der Übertragungsfunktion in der  $z$ -Ebene verfügbar.

In SPEAK sind zwei Arten von Komponenten vorhanden, für die es jeweils geeignete Verbindungstypen gibt. Zum einem gibt es Komponenten deren Ein- und Ausgänge sich auf den Wellenformalismus beziehen. Zur Verbindung dieser digitalen Wellenfilter ist ein Verbindungstyp vorgesehen, in dem neben hin- und rücklaufende Wellen auch die Impedanzen der angebotenen Wellenfilter zusammengefasst werden. Mit diesem wird eine physikalisch korrekte Realisierung von Rohrverzweigungen durch Mehrtoradaptoren gewährleistet. Gerichtete Verbindungen, ähnlich einem Signalflussgraphen, stehen zur Verbindung anderer Komponenten, wie FIR-Filter, Signalquellen und Elemente zur Signalanalyse, zur Verfügung. Diese Elemente können damit untereinander und an Wellenfilter verbunden werden, wobei ein Ausgang auch mehreren Komponenten zugänglich ist. Die Signalflussrichtung wird farblich hervorgehoben.

Neben verschiedenen Filtersystemen stehen eine Reihe von Anregungen zur Verfügung wie Impulse, weißes Rauschen, oder Glottisschwingungen nach [4]. Auch diese können interaktiv parametrisiert werden, wobei sich die Parameter unmittelbar auf die erzeugte Wellenform auswirken.



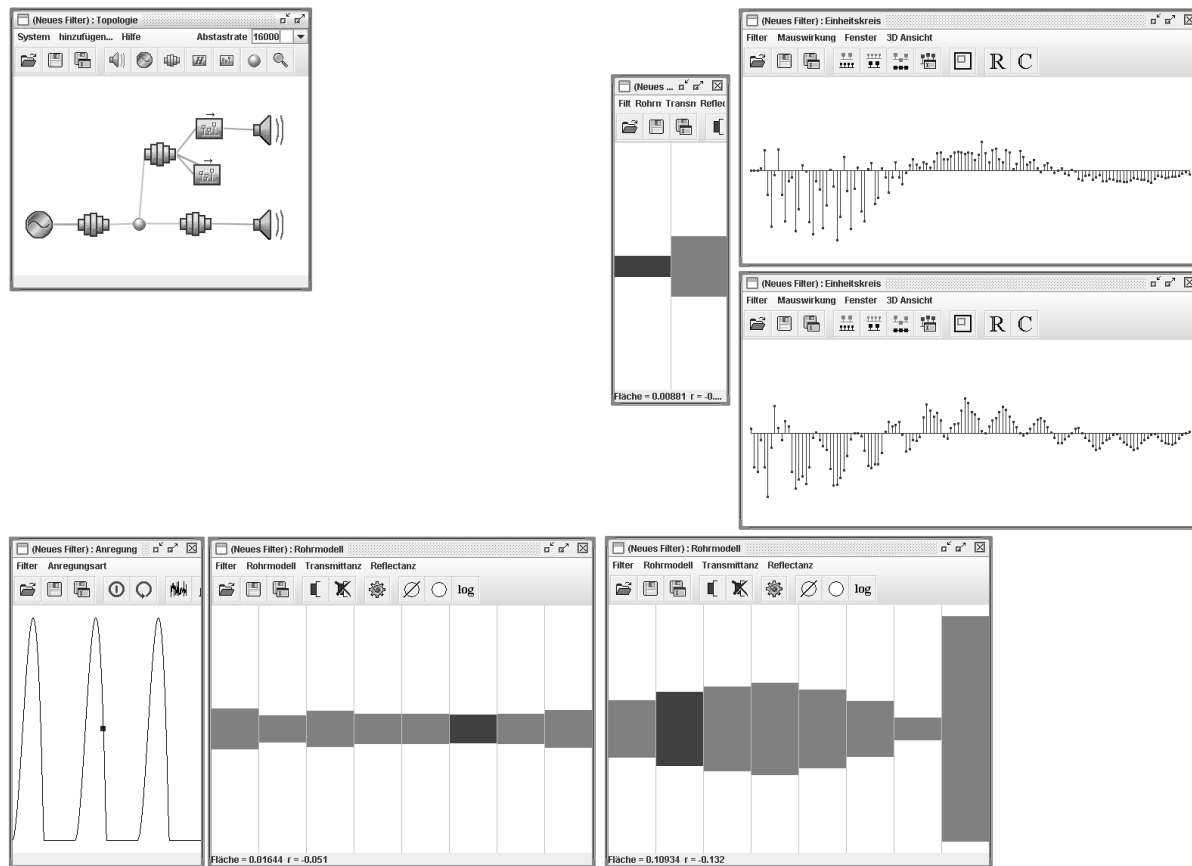
**Bild 3** - Verzweigtes Sprechtraktmodell zu Untersuchung grundlegender Effekte der Nasalierung. Oben links: Schema der Verknüpfung der Komponenten; die Anregung erfolgt mit weißem Rauschen. Oben Mitte: Rohrmodell des Nasaltrakts. Oben rechts: Gemittetes Spektrum des Nasensignals. Unten links, unten Mitte: Rohrmodell des Vokaltrakts. Unten rechts: gemittetes Spektrum des Mundsignals.

## 6 Beschreibung des Sprechtrakts durch ein verzweigtes Rohrsystem

Die Schätzung des Querschnittsverlaufs des Vokaltrakts kann nach den Verfahren von Burg oder Itakura für ein unverzweigtes Rohrmodell durchgeführt werden. Diese Verfahren sind in SPEAK implementiert, wobei eine simultane Visualisierung der Analyseresultate möglich ist. Dämpfungen im Vokaltrakt werden in dem Abschlusskoeffizienten zusammengefasst, sein Wert liegt geringfügig über -1; dieser wird durch eine Erweiterung der genannten Verfahren berücksichtigt [5]. Nach vorheriger mehrfacher Anwendung einer adaptiven Präemphase ergeben sich realistische Flächenverläufe. Neben unmittelbarer Spracheingabe können auch gespeicherte Sprachproben analysiert werden

Die Ankopplung des Nasaltrakts an den Vokaltrakt wird über einen Dreitoradaptor an der Position des Velums vorgenommen, wobei die erste Querschnittsfläche gemäß der Velumöffnung gewählt wird. Dabei wird mit der Größe der Velumöffnung der Grad der Nasalierung eines Lautes vorgegeben. Lässt man die Verzweigungen innerhalb des Nasaltrakts unberücksichtigt, kann man seine akustischen Eigenschaften durch ein Rohr variierenden Querschnitts approximieren. Der Verlauf der Querschnittsfläche des Nasaltrakts wird hierfür entsprechend den anatomischen Gegebenheiten festgelegt.

Mit dem so gebildeten einfach verzweigten Rohrmodell des Sprechtrakts lässt sich die Auswirkung der Nasalierung auf den von Mund und Nase abgestrahlten Schall untersuchen. Man sieht



**Bild 4** - Spechtraktmodell mit detaillierter Beschreibung der Akustik des Nasaltrakts durch FIR-Filter. Oben links: Schema der Verknüpfung der Komponenten. Oben mitte: Ansatzrohr des Nasaltrakts. Oben rechts: Impulsantwort bzw. FIR-Koeffizienten der Transmittanz des Nasaltrakts. Mitte rechts: Impulsantwort bzw. FIR-Koeffizienten der Reflektanz des Nasaltrakts. Die Darstellungen sind auf 128 Koeffizienten beschränkt. Unten links: Zeitfunktion der Anregung. Unten Mitte, unten rechts: Rohrmodell des Vokaltrakts

beispielsweise die Entstehung des Nasalformanten bei 400 Hz in dem Nasensignal. Es lassen sich die Auswirkungen der Resonanzen (Pole) eines Nebenzweiges als Nullstellen in der Übertragungsfunktion des betrachteten Zweiges wiederfinden. Oder es kann untersucht werden, in wie weit Formanten des Mundsignals auch in dem Nasensignal zu erwarten sind, was in Rahmen von [6] interessiert. In allen Fällen stehen hier nicht nur Frequenzgänge zu Verfügung, wie Bild 3 zeigt, sondern die synthetisierten Laute können auch auditiv beurteilt werden.

## 7 Beschreibung des Nasaltrakts durch FIR-Filter

Man gewinnt über die eingangs beschriebene Methode der Finite-Differenzen-Simulation eine sehr detaillierte Beschreibung des akustischen Verhaltens des Nasaltrakts in Form der zwei Impulsantworten. Aufgrund der erforderlichen Rechenleistung ist es nicht möglich, die Finite-Differenzen-Simulation in Echtzeit in eine Synthese zu integrieren. Eine realisierbare Implementierung in SPEAK basiert auf der Verwendung der vorberechneten Impulsantworten, da der Nasaltrakt als zeitlich konstant angesehen werden kann. Da die Impulsantworten relativ schnell abfallen, genügt es, sich auf die ersten 60 ms zu beschränken und man kann Filter mit endlicher

Impulsantwort einsetzen. Für die Verwendung in SPEAK wird die Abtastrate der Impulsantwort an die des Sprechtraktmodells angepasst; in dieser Untersuchung werden 16 kHz verwendet. Die Ankopplung an den Vokaltrakt erfolgt über einen Dreitoradaptor gefolgt von einem kurzen Rohrabschnitt. Dieser Rohrabschnitt besteht aus zwei Rohrsegmenten, wovon das erste der Velumöffnung entspricht, während das zweite die akustische Impedanz des Nasaltrakts definiert. Diese Realisierung ist in Bild 4 dargestellt, wobei das zweite FIR-Filter die Rückwirkung des Nasaltrakts auf den Vokaltrakt bestimmt. Die Eingangsgröße beider Filter ist die hinlaufende Schalldruckwelle des letzten Rohrsegments.

## 8 Zusammenfassung

In diesem Beitrag werden exemplarisch zwei verschiedene Beschreibungen des Nasaltrakts erläutert. Diese Modelle des Nasaltrakts werden mit dem Rohrmodell des Vokaltrakts gekoppelt und das resultierende Übertragungsverhalten wird untersucht. Die durch Nasalisierung hervorgerufenen Resonanzen und Antiresonanzen des Sprechtrakts werden mittels verschiedener Analysefunktionen von SPEAK dargestellt.

Die Implementierung von SPEAK in Java erweist sich als vorteilhaft: Neben einer ansprechenden und einfach zu bedienenden graphischen Benutzerschnittstelle ist eine Analyse und Synthese von Audiosignalen in Echtzeit möglich. Darüber hinaus besteht die Möglichkeit, ohne langwierige Installationen die Applikation via Internet verfügbar zu machen; sie läßt sich direkt von der Homepage unserer Arbeitsgruppe starten [7].

## Literatur

- [1] Lacroix, A.: Spracherzeugung — Akustik, Modelle und Anwendungen; Tagungsband DAGA: Bochum 2002, CD-ROM.
- [2] Ranostaj, F., Lacroix, A.: Bestimmung des Übertragungsverhaltens des Nasaltrakts aus computertomographischen Daten: Konvens 2000 Sprachkommunikation (ITG-FB 161), Illmenau 2000, S. 131-134.
- [3] Ranostaj, F., Lacroix, A.: Bestimmung der Randbedingungen von numerischen Simulationen der Schallausbreitung aus computertomographischen Daten; Tagungsband DAGA, Aachen 2003, CD-ROM.
- [4] Oliveira, L. C.: Estimation of Source Parameters by Frequency Analysis; Proc. Eurospeech: Berlin 1993, pp. 99-102.
- [5] Schnell, K., Ranostaj, F., Lacroix, A.: Parameter Estimation of Vocal Tract Models for SPEAK; Proc. Joint CFA/DAGA: Straßburg 2004, CD-ROM.
- [6] Geueke, M., Lacroix, A.: Analyse nasaliertter Vokale unter Berücksichtigung getrennter Mund- und Nasensignale; in diesem Band.
- [7] <http://www.rz.uni-frankfurt.de/~lacroix/index.html>