

Physikalische Modellierung der menschlichen Stimme

Malte Kob und Bernd J. Kröger
Klinik für Phoniatrie und Pädaudiologie
Universitätsklinikum, RWTH, 52074 Aachen
E-mails: mkob@ukaachen.de, bkroeger@ukaachen.de

1 Einleitung

Die Nachbildung der Stimmerzeugung beim Menschen durch mathematische Beschreibung der stimmerzeugenden Organe dient sowohl zum Verständnis der Funktion der Stimmorgane als auch zur Nachbildung individueller Eigenschaften des Sprachschalls, die durch Unterschiede der Physiologie des Menschen begründet sind. Sowohl für die medizinische Diagnostik als auch für die Kodierung von Sprachschall wäre es wünschenswert, sprecherspezifische Stimmcharakteristiken durch entsprechende Parametrisierung des Modells erzeugen zu können. Die Stimm-synthese ist durch hinreichend genaue Modellierung der Stimmorgane im Zeitbereich mittlerweile in recht guter Qualität machbar, wenn auch eine qualitativ hochwertige Echtzeit-Synthese noch nicht möglich ist.

Verschiedene Algorithmen für Modelle der an der Stimmerzeugung beteiligten Organe wurden in MATLAB implementiert, wobei eine Kombination der Module in einem waveguide-Modell die Berücksichtigung der Rückkopplung der Komponenten untereinander möglich macht.

Die in Abb. 1 dargestellte graphische Oberfläche erlaubt den Zugriff auf die wichtigsten Parameter während der Berechnung und stellt eine Vielzahl von Visualisierungsmöglichkeiten bereit, wie eine 3D-Darstellung der Stimmlip-penschwingung oder die Schallausbreitung im Ansatzrohr.

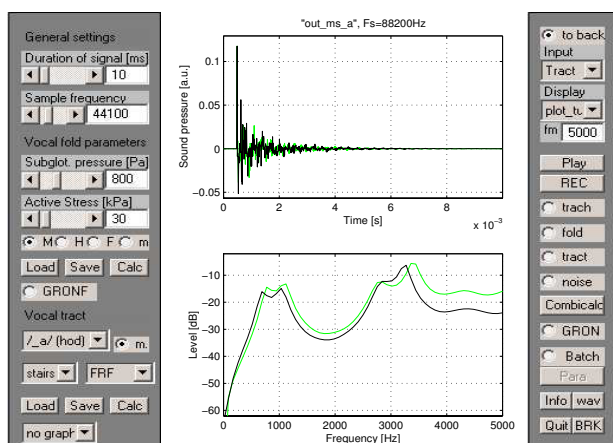


Abbildung 1: Grafische Benutzeroberfläche des Hauptprogramms.

2 Modellkomponenten

Die Prozesse der Stimmschallgenerierung können verschiedenen Stimmorganen zugeordnet werden. Die Lunge stellt im wesentlichen einen konstanten subglottalen Druck bereit, der die Phonation erst ermöglicht. Die Schallerzeugung und -formung durch die Funktionen von Stimmlippen und Ansatzrohr (Raum zwischen Stimmlippen und Mundöffnung) wird im folgenden näher beschrieben. Auf die Richtungs- und Frequenzabhängigkeit der Abstrahlung des Stimmschalls am Mund wird hier nicht näher eingegangen. Eine ausführliche Beschreibung findet sich in [1].

2.1 Stimmlippen

Die Stimmlippen können durch eine Anordnung gekoppelter mechanischer Oszillatoren beschrieben werden. Wird das klassische 2-Massenmodell in Längsrichtung der Stimmlippe weiter unterteilt [2], kann dieser weitere Freiheitsgrad für die Modellierung von Stimmstörungen oder Stimmqualitäten (z.B. Behauchtheit, gepresste Stimme) genutzt werden. Der hier vorgestellte Ansatz ermöglicht die Diskretisierung in beliebig viele Massenpaare. Weiterhin wurde die bei neueren Modellen [3] berücksichtigte Jetablösung in der divergenten Phase der Stimmlippen-schwingung implementiert.

2.2 Ansatzrohr

Für die artikulatorische Stimm-synthese ist im Zeitbereich insbesondere der eindimensionale Wellenleiter-Ansatz gebräuchlich [4]. Basis für die Berechnung der Transferfunktion von der Glottis bis zur Mundöffnung sind meist äquivalente Flächenfunktionen (EAF), die aus Magnetresonanzbildgebung gewonnen werden. Die hohe Zahl der Berechnungen läßt sich durch Interpolation der EAF in Kegelschnitte reduzieren [1]. In Abb. 2 ist der Verlauf beider EAFs für den Vokal [a] dargestellt. Die Steuerung der Parametrisierung des Wellenleiter-Modells durch ein dynamisches Artikulationsmodell ermöglicht dann die Synthese von Lauten und Sätzen [5].

2.3 Rauschgenerator

Während Friktionsrauschen nur bei der Erzeugung entsprechender Konsonanten auftritt (z.B. [ʃ] in **Schnee**), ist das Aspirationsrauschen Bestandteil des natürlich klingenden stimmhaften Lauts. Ursache für dieses gepulste Rauschen ist die Bildung von Wirbeln bei der Ablösung des Jets von den Stimmlippen. Ein Modell, daß die

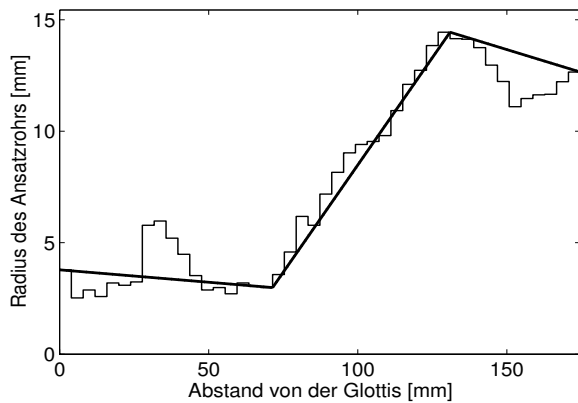


Abbildung 2: Berechnung der Transferfunktion mit Zylinder- (dünn) oder Kegelsegmenten (dick).

Wirbel- und die Rauschsignalerzeugung nachbildet, wurde auf Basis des Modells von Sinder [6] implementiert.

3 Anwendungen

Der Einsatz des Programms für die Nachbildung der Singstimme und insbesondere außergewöhnlicher Gesangsstile wie dem Obertonsingen ist bereits früher diskutiert worden [8]. Beispiele für die Synthese von Klängen finden sich auf dem Internet [9].

3.1 Artikulatorische Synthese

Das dynamische Artikulationsmodell mit nachgeschalteter einfacher akustischer Modellierung wurde bereits erfolgreich in Vergleich mit anderen Syntheseprogrammen getestet [7]. Sowohl die Synthese einzelner Laute als auch die Erzeugung zusammenhängender Sätze ist möglich.

3.2 Nachbildung von Stimmstörungen

Als Beispiel für die Synthese von Stimmsignalen mit dem Mehrmassenmodell wurde beispielhaft die Stimmlippenschwingung bei einer pathologischen lokalen Massenzunahme der Stimmlippe (Sängerknötchen) berechnet. Das Sängerknötchen wurde durch Erhöhung der Masse eines Stimmlippensegments nachgebildet.

Die resultierende irreguläre Stimmlippenschwingung wies eine Halbierung der Grundfrequenz auf. Durch Erhöhung der Stimmlippenspannung konnte beim Modell diese Irregularität korrigiert werden, während beim Sänger dieser Prozess vermutlich unbewußt abläuft.

In Abb. 3 ist das Spektrogramm des Signals am Mund für die Vokaltraktkonfiguration des Vokals [a:] gezeigt. Die Stimmlippenspannung wurde linear von 30 kPa auf 60 kPa erhöht. Durch Erhöhung der Stimmlippenlängsspannung kann neben der Zunahme der Grundfrequenz auch der Übergang von unterschiedlichen glottalen Schwingungsformen nachgebildet werden.

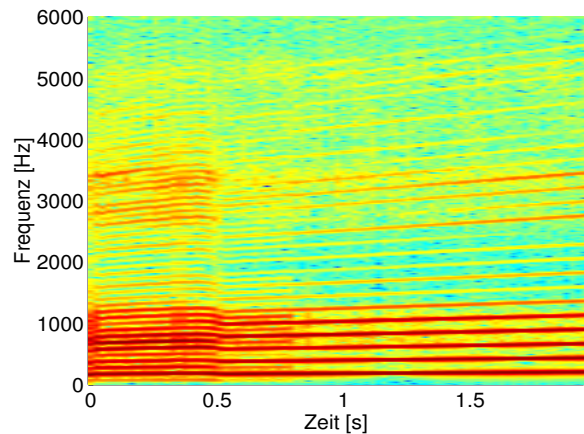


Abbildung 3: Spektrogramm Modenübergang bei Sängerknötchen.

4 Ausblick

Das Programm hat sich als Plattform für die Implementierung neuer Algorithmen sowie als didaktisches Hilfsmittel bewährt. Die Anwendung zur Modellierung von Stimm- und Sprechstörungen wird gegenwärtig erprobt. Eine Kombination des dynamischen Artikulationsmodells mit dem selbstschwingenden Mehrmassenmodell der Stimmlippen, der akustischen Modellierung des Ansatzrohres und dem Modell für die Rauschgenerierung ist geplant.

Literatur

- [1] M. Kob (2002): Physical modeling of the singing voice. Dissertation, Rheinisch-Westfälische Technische Hochschule Aachen, im Druck.
- [2] I.R. Titze (1973): The Human Vocal Cords: A Mathematical Model Part I. *Phonetica* **28** 129-170.
- [3] X. Pelorson *et al.* (1995): Description of the flow through in-vitro models of the glottis during phonation. *acta acustica* **3** 191-202.
- [4] J.L. Kelly and C.C. Lochbaum (1962): Speech synthesis. Proceedings of the 4th International Congress on Acoustics, 1-4.
- [5] B.J. Kröger (1998): Ein phonetisches Modell der Sprachproduktion. Niemeyer, Tübingen.
- [6] D.J. Sinder (1999): Speech Synthesis Using an Aeroacoustic Fricative Model. Ph.D. Thesis, University of New Jersey.
- [7] K. Fellbaum, H. Klaus, J. Sotscheck. (1994): Hörversuche zur Beurteilung von Sprachqualität von Synthesystemen für die deutsche Sprache. *Fortschritte der Akustik – DAGA 94*, 117-122.
- [8] M. Kob (2001): Untersuchung der Eigenschaften des Obertongesangs. *Fortschritte der Akustik – DAGA 01*, 436-437.
- [9] URL www.akustik.rwth-aachen.de/~malte/vox