

# OPTIMIERUNG DER NUMERIK EINES LINEAREN GLEICHUNGSSYSTEMS FÜR DIE SIMULATION DES SCHALLFELDES IM VOKALTRAKT

*Johann August Marwitz, Simon Stone, Peter Birkholz*

*Institut für Akustik und Sprachkommunikation, Technische Universität Dresden*

*Johann\_August.Marwitz@mailbox.tu-dresden.de*

**Kurzfassung:** Im Gegensatz zur in Forschung und Industrie weit verbreiteten konkativen Sprachsynthese besitzt die artikulatorische Sprachsynthese alle Freiheiten, die auch ein natürlicher Sprechapparat besitzt. Durch die vollständige aeroakustische Simulation der Sprache entsteht nachteilig ein hoher Rechenaufwand. Der weltweit eingesetzte Synthesizer VocalTractLab [1, 2] erreichte bisher auf aktueller PC-Hardware nur einen Echtzeitfaktor von ca. 4,5. In diesem Paper wird dargestellt, wie die Rechenzeit durch ein angepasstes numerisches Verfahren auf Faktor 2 reduziert werden kann. In der Simulation im VocalTractLab wird der Sprechapparat durch ein System von 97 diskreten Rohrabschnitten angenähert. Die Rohrabschnitte sind zeitlich in ihrer Länge, Querschnittsfläche sowie ihrem Umfang variabel. Zur Simulation des Schallfeldes in diesem zeitveränderlichen Rohrsystem wird in jedem Zeitschritt der Simulation ein lineares Gleichungssystem (LGS) mit 97 Unbekannten erstellt und nach den unbekanntem Volumenströmen (ein Volumenstromwert pro Rohrabschnitt) aufgelöst. Bei einer Abtastrate von 44,1 kHz, muss dieses LGS 44100 mal pro Sekunde simulierter Zeit gelöst werden. Dadurch entsteht ein hoher Rechenaufwand, der durch die Auswahl eines effizienten Algorithmus massiv reduziert werden kann. Die in Betracht kommenden Verfahren werden in dieser Arbeit vorgestellt, verglichen und die Ergebnisse mit dem optimalen Algorithmus präsentiert. Algorithmen aus drei großen Verfahrensklassen (direkte Verfahren, Splittingverfahren und Projektionsmethoden) wurden zu diesem Zweck verglichen. Für die in diesen Klassen enthaltenen Näherungsverfahren wird als Evaluationskriterium die Anzahl der Iterationen angesetzt, die für ein hinreichend kleines Residuum bzw. eine hinreichend genaue Lösung benötigt wird. Außerdem wurden verschiedene Vorkonditionierungen erprobt, mit denen die benötigte Anzahl der Iterationen um bis zu Faktor 26 reduziert werden konnte. Ein beispielhafter Residuenverlauf ist in Abbildung 2 zu sehen. Da die benötigte Zeit zur Durchführung einer Iteration in den unterschiedlichen Verfahrensweisen teilweise nicht konstant ist und Vergleichbarkeit mit den direkten Verfahren (bei denen die exakte Lösung des LGS bestimmt wird) gewährleistet werden sollte, wurde ein Vergleich der Berechnungszeit angestellt. Bei der Beurteilung der Berechnungszeit zeigte sich, dass besonders eine ebenfalls vorgestellte effiziente Speicherstrategie des Gleichungssystems entscheidende Verbesserung brachte. Insgesamt ist es gelungen, mittels der Cholesky-Zerlegung und einer auf den Algorithmus sowie auf das LGS angepassten Speichermethode die Berechnungszeit um das Fünf- bis Sechsfache gegenüber dem bisher verwendeten Gauß-Seidel-Relaxationsverfahren zu verkürzen. Verglichen mit der Dauer des simulierten akustischen Signals nimmt die Berechnung des LGS nur noch ca. die Hälfte der simulierten Zeit in Anspruch - womit die gesamte Synthese deutlich näher an die Echtzeitfähigkeit rückt.

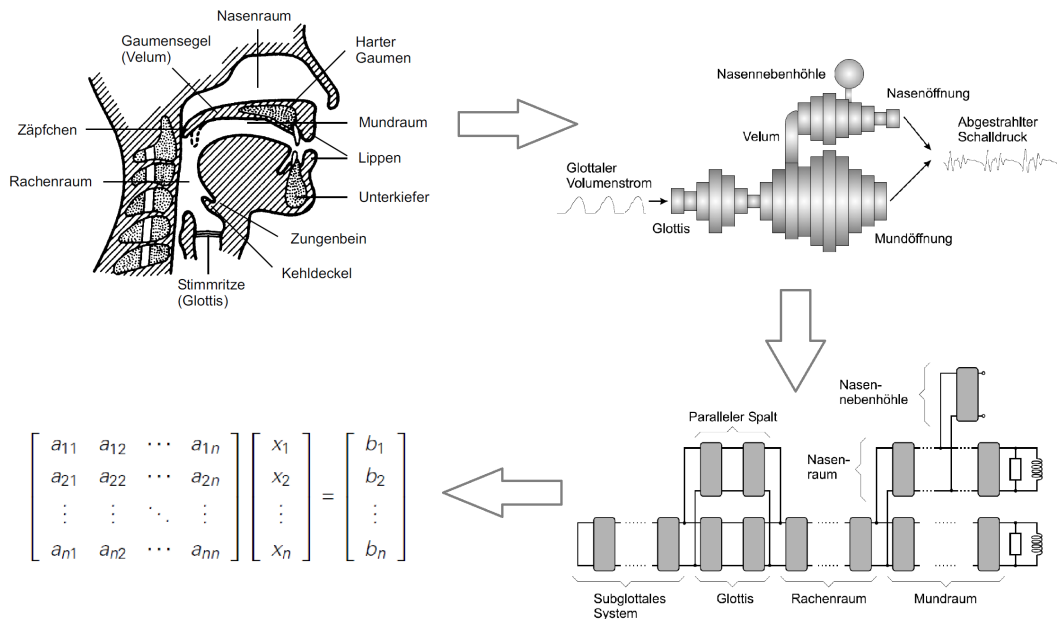


Abbildung 1 – Vom Sprechapparat zum linearen Gleichungssystem (Abbildung nach [3]).

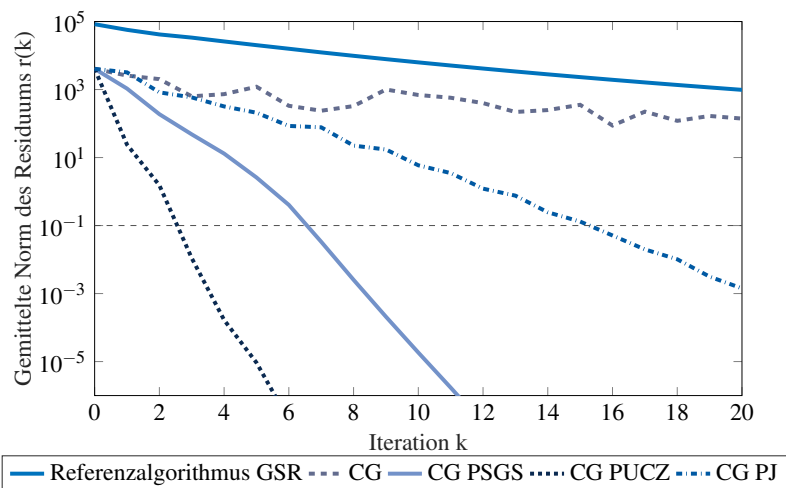


Abbildung 2 – Vorkonditionierte CG-Verfahren im Vergleich. Die horizontale gestrichelte Linie markiert die Grenze, ab der eine ausreichend genaue Lösung des LGS zu erwarten ist.

## Literatur

- [1] BIRKHOLZ, P.: *Modeling consonant-vowel coarticulation for articulatory speech synthesis*. *PLOS ONE*, 8(4), S. e60603, 2013.
- [2] BIRKHOLZ, P.: *Enhanced area functions for noise source modeling in the vocal tract*. In *10th International Seminar on Speech Production, Köln*, S. 1–4. 2014.
- [3] BIRKHOLZ, P.: *3D-Artikulatorische Sprachsynthese*. Universität Rostock, 2005.