

Versuchsanordnung zur Erzeugung von Stimmsignalen

Tobias Fürtjes¹, Tobias Frauenrath², Malte Kob²

¹ Institut für Technische Akustik, RWTH Aachen, Neustrasse 50, 52066 Aachen

² Lehr- und Forschungsgebiet für Phoniatrie und Pädaudiologie, RWTH – Universitätsklinikum Aachen

Einleitung

Für die Lehre der menschlichen Stimmerzeugung ist es hilfreich, den komplexen Vorgang der Phonation mit Modellen zu erklären. Während zur Vokalunterscheidung die Vokaltrakteinstellung ein wichtiger Parameter für die Ausprägung der Energiemaxima (Formanten) im Stimmsignal ist, bestimmt die Stimmlippschwingung und die dabei auftretenden Strömungs- und Druckvariationen sowohl die Grundfrequenz als auch die Qualität der Stimme. Eine einfache mathematische Beschreibung des bei der regulären Stimmlippschwingung modulierten Volumenstroms ist in [Rosenberg 1970] zu finden. Er vergleicht mehrere mathematische Beschreibungen des Signals in Abhängigkeit der Parameter Amplitude, Frequenz, Öffnungsquotient (Open Quotient, OQ) sowie Geschwindigkeitsquotient (Skew).

In diesem Beitrag werden die Ergebnisse einer Studienarbeit beschrieben, deren Gegenstand es ist, diese Schwingung vorerst mit MatLAB zu simulieren und anschließend mittels Mikrokontroller und Signalprozessor elektrisch zu erzeugen. Das elektrische Signal der Grundschwingung soll hierbei dynamisch änderbar und in einer Hardware PC-unabhängig generiert werden.

Das fertige Exponat besteht aus der Controllereinheit, einer Verstärkerschaltung mit angeschlossenem Horntrieb und aus MRT-Daten gewonnen, realistischen Vokaltraktmodellen für die jeweiligen Vokale. Der Aufbau wird mit Kontrollreglern für die Steuerung ausgestattet und erlaubt eine direkte auditive Kontrolle der Auswirkung von Parameteränderungen.

Stimmsignal-Modelle

Die am unteren Ende des Vokaltrakts anliegende Grundschwingung der menschlichen Stimme ist durch verschiedene mathematische Modelle beschreibbar. Besonders gebräuchlich sind hierbei die Definitionen von [Fant 1960], [Rosenberg 1970], [Liljencrants u. a. 1985], [Klatt und Klatt 1990] und [Veldhuis 1998]. In allen Ansätzen wird die Grundschwingung in drei Phasen aufgeteilt: Öffnungsphase, Schließphase und Ruhephase. Die signalformbeschreibenden Parameter „Öffnungsquotient“, „Skew“, „Amplitude“ und „Frequenz“ werden in allen vier Definitionen genau festgelegt. Abbildung 1 zeigt eine EGG-Aufnahme, an der diese Parameter bei realen Aufnahmen bestimmt werden können.

Im Folgenden soll jedoch lediglich der Ansatz von Rosenberg verfolgt werden, da sein Ansatz besonders praktikabel für eine Implementierung in einem Mikrokontroller ist. Rosenberg gibt in seiner Arbeit verschiedene Si-

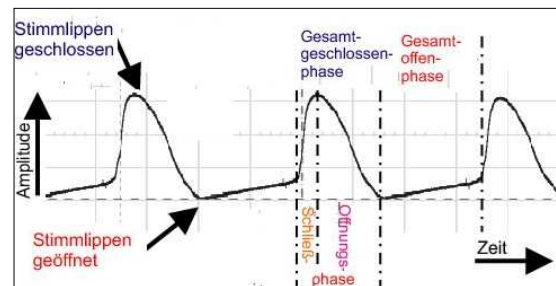


Abbildung 1: EGG-Messung einer Stimmaufnahme. Bei der EGG-Messung entsprechen positive Amplitudenwerte der geschlossenen Phase, negative Werte der offenen Phase der Glottisschwingung

gnalbeschreibungen an. Dabei variiert die Gültigkeit der Annäherung an die Realität stark. Die verschiedenen Varianten der Rosenbergmodellierung sind in Abb. 2 gezeigt und in Tabelle 1 mathematisch beschrieben.

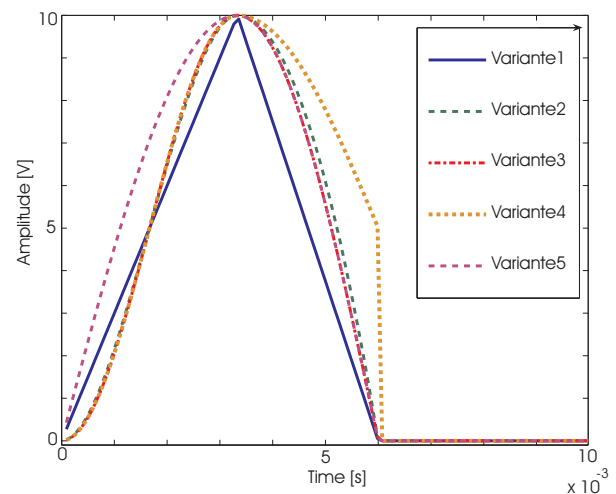


Abbildung 2: Die verschiedenen Varianten des Rosenberg-Modells im Überblick

Simulation

Zur Simulation der Machbarkeit der Synthese kommt MATLAB zum Einsatz. Eine Benutzerschnittstelle bieten dem Nutzer vielfältige Bedienmöglichkeiten. In der graphischen Benutzerschnittstelle (GUI) kann der Benutzer mit Schieberegler und Menüs die verschiedenen Parameter einstellen und das Ergebnis direkt auf dem Bildschirm betrachten. In der entstandenen GUI können Frequenz, Amplitude, Öffnungsquotient sowie Speedquotient per Slider reguliert und auf die fünf verschiede-

Tabelle 1: Mathematische Beschreibung der Varianten des Rosenberg-Modells

Variante	Phasenbeschreibung	
	öffnend	schließend
1	$a(\frac{t}{T_p})$	$a(1 - \frac{(t-T_p)}{T_n})$
2	$a(3(\frac{t}{T_p})^2 - 2(\frac{t}{T_p})^3)$	$a(1 - (\frac{t-T_p}{T_n})^2)$
3	$\frac{a}{2}(1 - \cos(\frac{t}{T_p}\pi))$	$a \cos(\frac{(t-T_p)\pi}{2})$
4	$\frac{a}{2}(1 - \cos(\frac{t}{T_p}\pi))$	$\frac{a}{2}(1 + \cos(\frac{(t-T_p)\pi}{2}))$
5	$a \sin(\frac{t}{T_p}\frac{\pi}{2})$	$a \cos(\frac{(t-T_p)\pi}{2})$

nen Varianten des Rosenbergmodells angewendet werden. Zusätzlich kann dem Signal ein in dB relativ zum periodischen Signal skalierter Rauschanteil zugemischt werden um so die Simulation um das real vorhandene natürliche Rauschen zu ergänzen. Des Weiteren sind in die Simulation eine Soundausgabe und eine erweiterte Plotfunktion integriert, welche neben dem eigentlichen Signal auch das Spektrum und den zeitlichen Verlauf des Signals grafisch darstellt. Bei der Programmierung wurde möglichst ressourcensparend programmiert um für die Realisierung in Hardware eine optimale Ausgangssituation zu schaffen.

Hardwareumsetzung

Die gewünschte Hardwarelösung soll eine dynamische Parameterveränderung bieten, da nur so eindrucksvoll der Vorgang der Sprachgeneration gezeigt wird. Durch die erreichte PC Unabhängigkeit ist die Hardware, und somit auch später das komplette Exponat, intuitiv ohne Einarbeitung durch fachfremde Interessenten beispielsweise an einem Demonstrationsstand zu bedienen.

Für diese Arbeit wird das Evaluation Board ADSP-21065L der Firma Analog Devices eingesetzt. Dieses Board bietet verschiedene Vorteile: Es ist ausreichend schnell um die Rechenoperationen in Echtzeit zu ermöglichen und wird bereits an der RWTH bei Praktika des Elektrotechnikstudiums eingesetzt, was die Einarbeitungszeit deutlich herabsetzt. In Kooperationen mit dem Institut für Nachrichtengeräte und Datenverarbeitung der RWTH konnte somit eine schnelle Implementierung erfolgen. Der bereits enthaltene AD/DA Wandler und Vorverstärker erspart weitere, zur direkten Signalausgabe nötige Hardware.

In dieser Arbeit wird ein getrenntes Programmier- und Bedien-Schnittstellenkonzept verfolgt. So ist es möglich, Parameteränderungen direkt über ein Bedieninterface via RS232 Schnittstelle einzugeben und die Programmierung und das Debuggen davon unabhängig aber gleichzeitig über die Programmierschnittstelle (JTAG) zu realisieren. Praktiziert wird dieses „debugging“ über einen externen „JTAG zu USB Adapter“. Die fertig evaluierte Software wird zur PC-unabhängigen Nutzung der Hardware auf den einen Festspeicher (EPROM) gebrannt.

Vorerst steht zur dynamischen Parameterwahl ein unter Visual C++ entwickeltes GUI, das auf einem PC ausgeführt wird, zur Verfügung. Um auch im jetzigen Stadium PC-unabhängig arbeiten zu können, sind die auf dem Board vorhandenen Taster so programmiert, dass es möglich ist, durch Drücken einer Tastenkombination die oben genannten Parameter zu verändern.

Fazit und Ausblick

Die elektronische Sprachsignalsynthese durch eine PC-unabhängige Hardware wird in dieser Arbeit gezeigt. Die Softwaresimulation des Rosenberg-Modells unter MATLAB dient zur Untersuchung der Implementierbarkeit in Hardware und als PC-gestützte Variante während der Entwicklung. Das ausgegebene Signal ist der an der menschlichen Glottis produzierten Wellenform sehr ähnlich. Generell kann der Prozess der menschlichen Sprachgeneration sehr gut anhand des vorgestellten Konzeptes erklärt werden. Das Modell soll jedoch nicht nur Neugierde wecken sondern auch zu Unterrichtszwecken eingesetzt werden.

Darüber hinaus dient das Modell nicht nur zur Erzeugung von Signalen der gesunden Stimme. Zur Synthese von Stimmstörungen können typische Parameter an diesem Modell eingestellt werden, um deren Auswirkungen auf die Sprache direkt hörbar zu machen. Dies kann sowohl in der Logopädischen Schulung als auch zur Symptomerklärung für betroffene Patienten nützlich sein.

In einem Folgeprojekt soll die Stimmsynthese mit einem individuell und intuitiv angefertigten Bedienportal über die RS232-Schnittstelle gesteuert werden. Denkbar ist auch eine Ansteuerung über weitere externe Eingabegeräte, wie z. B. einen Midi-Controller. Ein weiteres Ziel ist die Hardware-Implementierung des im Stimmsignal enthaltenen natürlichen Rauschanteils. Die ausgewählte DSP-Hardware bietet für diese Erweiterungen genügend freie Ressourcen und erlaubt somit eine künftige Implementierung weiterer Komponenten in das Exponat.

Literatur

- [Fant 1960] FANT, Gunnar: Acoustic theory of speech production. (1960), Nr. 2-4
- [Klatt und Klatt 1990] KLATT, D. H. ; KLATT, L. C.: Analysis, synthesis, and perception of voice quality variations among female and male talkers. In: *J Acoust Soc Am* 87 (1990), Feb, Nr. 2, S. 820–857
- [Liljencrants u. a. 1985] LILJENCRAINTS, J. ; LIN, Q. ; FANT, G.: A four-parameter model of glottal flow. In: *STL-QPSR* (1985)
- [Rosenberg 1970] ROSENBERG, A.E.: Effect of Glottal Shape on the Quality of Natural Vowels. In: *The Journal of the Acoustical Society of America* (1970)
- [Veldhuis 1998] VELDHUIS, R.: A computationally efficient alternative for the Liljencrants-Fant model and its perceptual evaluation. In: *J Acoust Soc Am* 103 (1998), Jan, Nr. 1, S. 566–571