

Hauptseminararbeit zum Thema

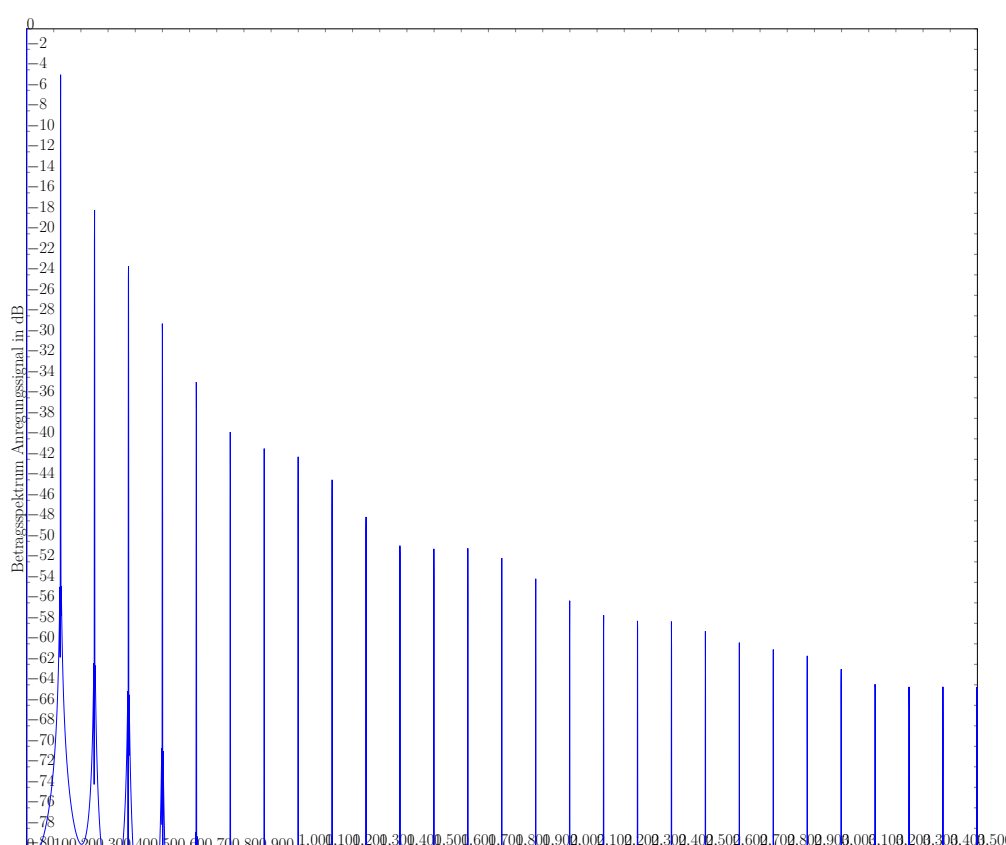
Synthese von Sprachsignalen

Karl-Ludwig Besser, Zhongjiu Li, Franz Marcus Schüffny, Peter Steiner

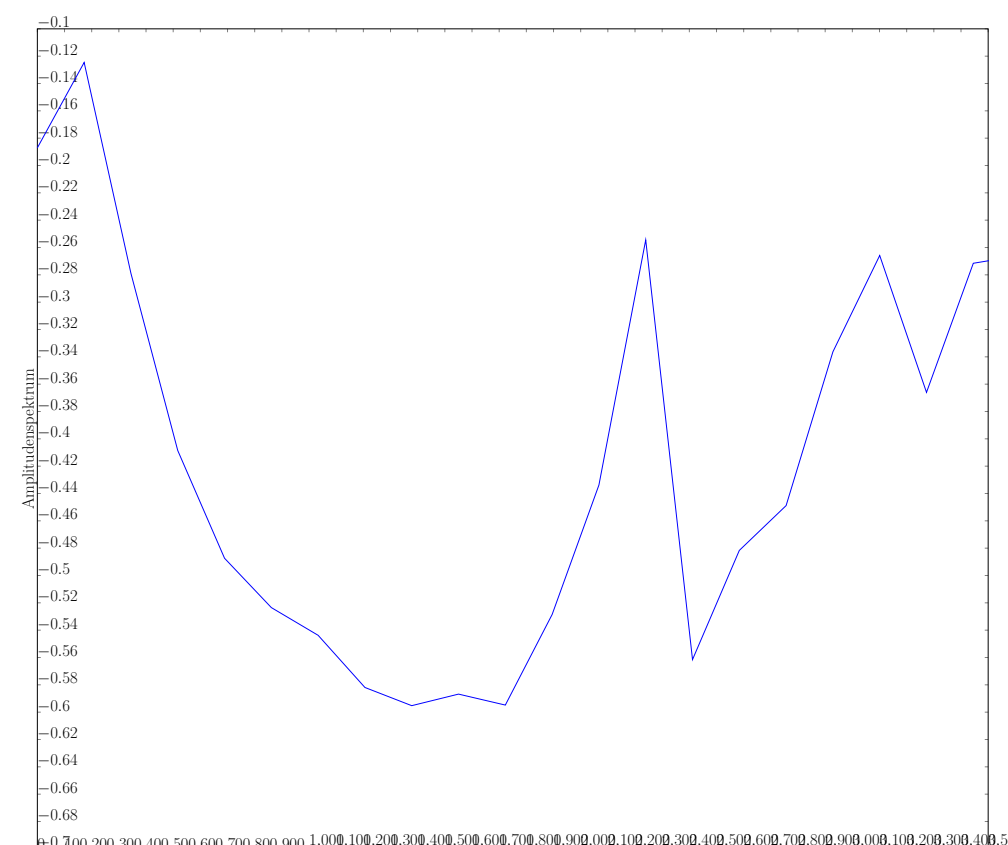
Betreuer:
PD Dr.-Ing. Ulrich Kordon
Dipl.-Ing. Steffen Kürbis

Hochschullehrer:
Jun.-Prof. Dr.-Ing. Peter Birkholz

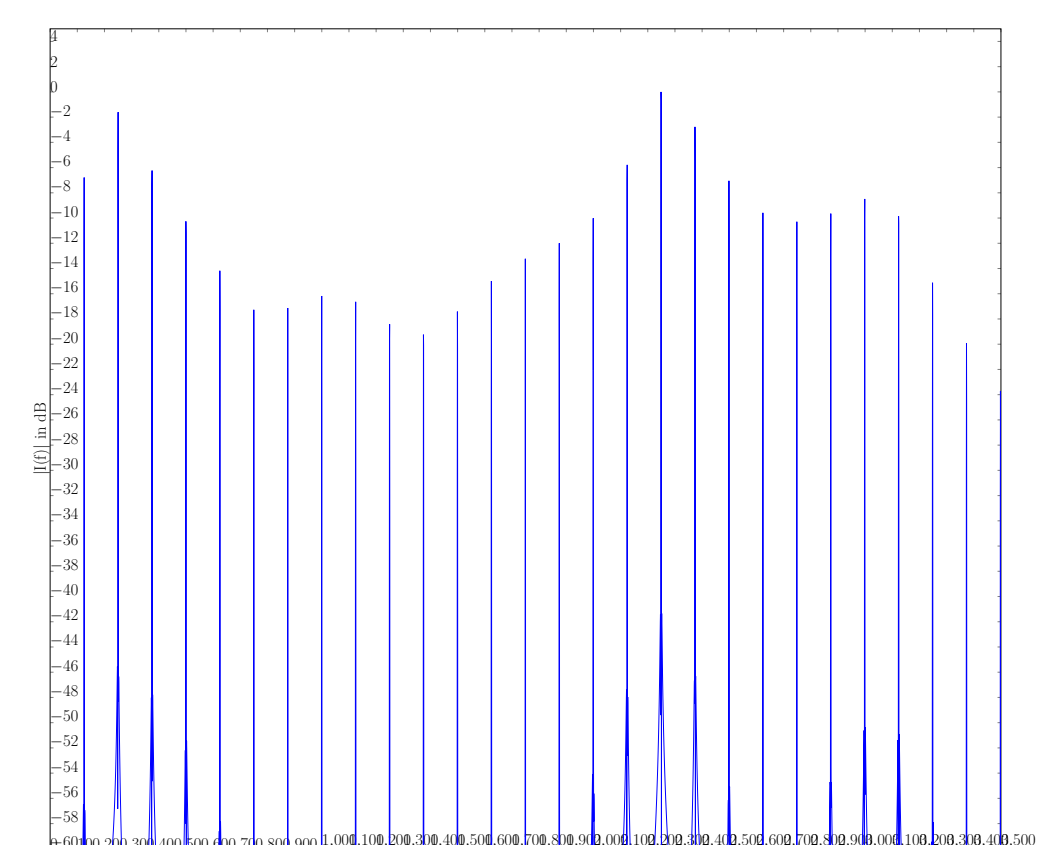
Verteidigungsdatum:
14.07.2015



Spektrum des Anregungssignals



Spektrum des aufgenommenen Signals



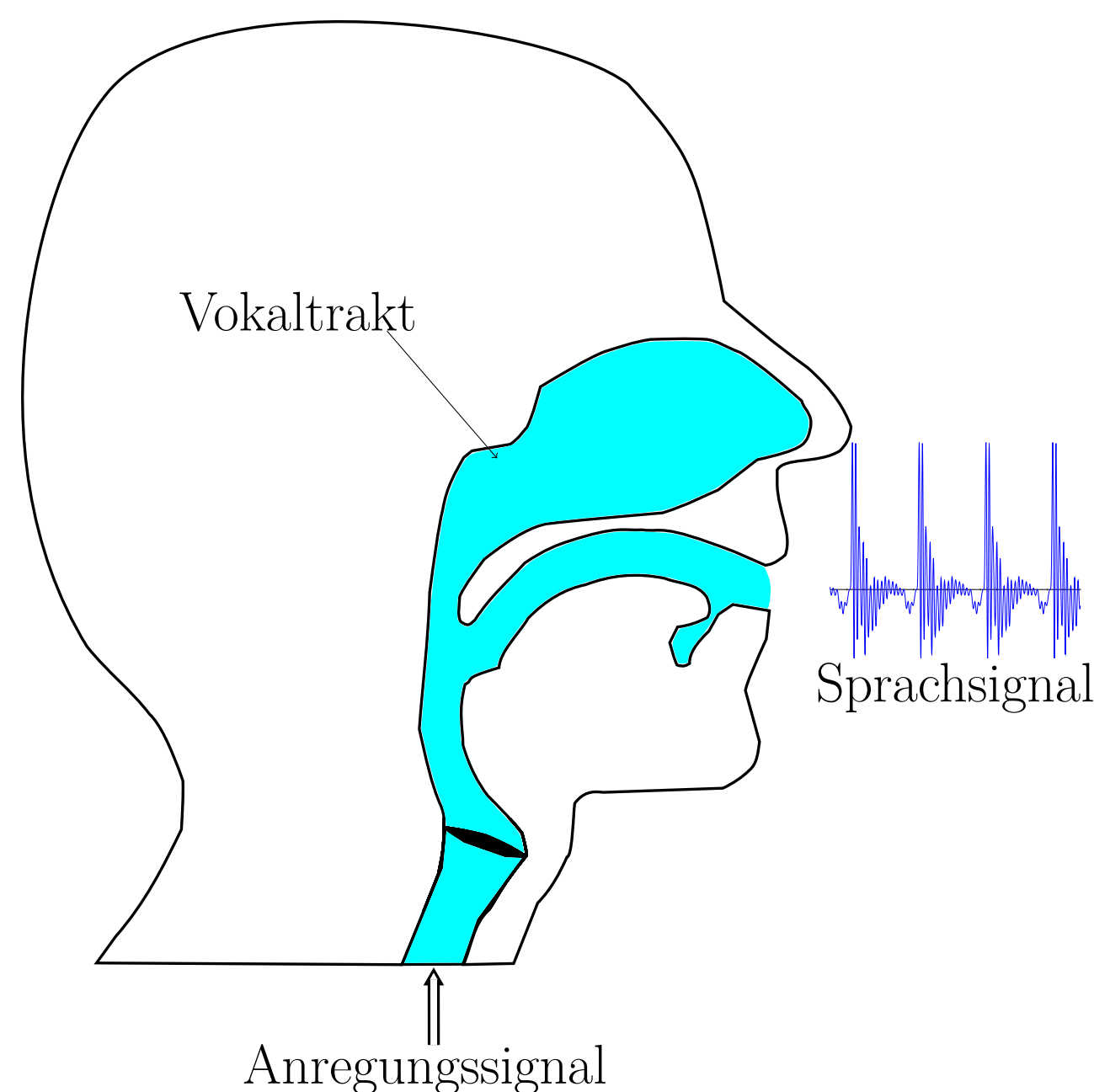
Spektrum des synthetisierten Signals

Sprachsignalanalyse

Zur Parametrisierung des verwendeten Quelle-Filter-Modells müssen reale Sprachsignale analysiert werden.

Der Schwerpunkt liegt hierbei auf der Bestimmung der Filterparameter. Es werden primär Bandpassfilter 2. Ordnung eingesetzt. Die drei charakteristischen Parameter sind Mittenfrequenz, Bandbreite und Grundverstärkung.

Die Mittenfrequenz wird als Formantfrequenz bezeichnet. Um diese zu bestimmen, wurden verschiedene Analysemethoden verwendet. Von der Software Praat wurde der fertig implementierte Burg-Algorithmus bereitgestellt, selbst nachprogrammiert wurde der Cepstrum-Algorithmus. Dabei wird unter Annahme des Quelle-Filter-Modells eine Trennung von Quellsignal und Übertragungsfunktion des Filters vorgenommen.



Syntheseergebnisse

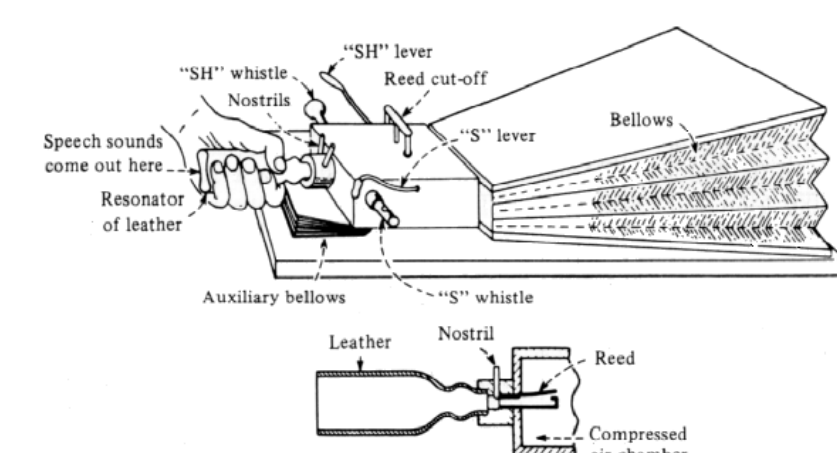
Implementiert wurden folgende Laute: Vokale, Diphthonge (/au/, /ei/, /eu/), stimmlose Frikative (/ch/, /f/, /s/, /sch/), stimmhafte Frikative (/w/), stimmhafte Plosive (/b/, /d/, /g/), stimmlose Plosive (/k/, /t/, /p/), Liquide (/l/, /r/) sowie Nasale (/m/, /n/). Allerdings können stimmhafte Plosive nur in Verbindung mit einem darauffolgenden Vokal synthetisiert werden.

Sprachsignalsynthese

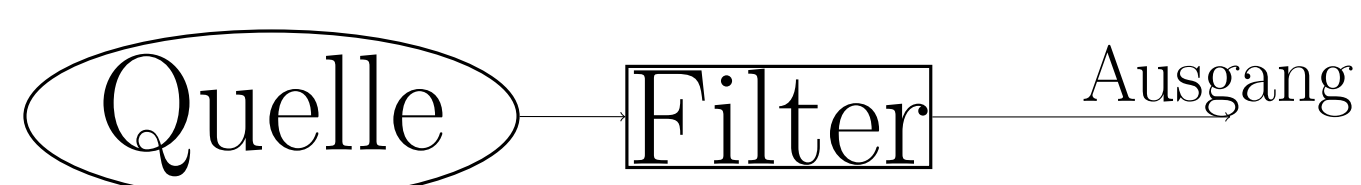
Zur Synthese von Vokallauten wird ein periodisches breitbandiges Anregungssignal als Quelle verwendet. Die Signalformung wird mittels Bandpassfilter in Reihenschaltung realisiert.

Zur Synthese von Zischlauten wird ein Rauschgenerator als Quelle verwendet. Die Signalformung wird mittels Bandpassfilter in Parallelschaltung realisiert.

Bei der Lautkombination werden einzelne Laute mit einem Von-Hann-Fenster gefenstert und ineinander verschoben. Damit erreicht man neben verständlichen Einzellaute auch eine gute Verständlichkeit der ganzen Lautkombination.



Von-Kempelen's Sprachautomat



Quelle-Filter-Modell

Das Quelle-Filter-Modell versucht eine Zerlegung von Sprachsignalen in Anregungssignale und Filterstrukturen. Durch geeignete Wahl der Modellparameter soll eine möglichst gute Modellierung des menschlichen Artikulationstrakts erreicht werden.