

# SYNTHETISCHE TIEFPAß-ERWEITERUNG VON TELEFONSPRACHE

Ulrich Kornagel

Signal Theorie, Institut für Nachrichtentechnik

Technische Universität Darmstadt, Merckstrasse 25, D-64283 Darmstadt

E-Mail: [ulrich.kornagel@nt.tu-darmstadt.de](mailto:ulrich.kornagel@nt.tu-darmstadt.de)

## 1. EINLEITUNG

Sprachsignale werden in der Regel mit einer Bandbreite von 0.3 kHz bis 3.4 kHz durch Telefonsysteme übertragen, was zu der bekannten Qualitätsminderung führt. Als Abhilfe kann auf der Empfängerseite ein Algorithmus zur synthetischen Bandbreiten-Erweiterung eingesetzt werden. Das Ziel hiervon ist die Erzeugung spektraler Komponenten in den Frequenzbereichen unterhalb 0.3 kHz (die Tiefpaß-Erweiterung) und von 3.4 kHz typischerweise bis zu 8 kHz (die Hochpaß-Erweiterung), ohne zusätzliche Information zu übertragen. In den letzten Jahren wurden verschiedene Ansätze untersucht (z.B. [2, 5, 6]) die sich auf die Hochpaß-Erweiterung beziehen. Andere Veröffentlichungen (z.B. [3, 7]) berücksichtigen auch die Tiefpaß-Erweiterung. Speziell in [3] wird die Tiefpaß-Erweiterung auf Grundlage eines Codebuch-Ansatzes diskutiert. Die Artefakte, die bei diesem Ansatz auftreten, können durch den Autor dieser Veröffentlichung bestätigt werden und sind die Motivation für die hier präsentierte Idee der Tiefpaß-Erweiterung. Dieses Verfahren kann jedem Hochpaß-Erweiterungssystem angehängt werden.

## 2. SYNTHETISCHE SPEKTRALERWEITERUNG

Der Ausgangspunkt der Tiefpaß-Erweiterung ist das Sprachsignal, das von einem traditionellen Hochpaß-Erweiterungsverfahren geliefert wird. Hier wird angenommen, daß es Frequenzkomponenten im Bereich 0.3 kHz bis 8 kHz aufweist. Es kann mittels der LPC-Analyse rahmenweise (typischerweise mit 20 ms Rahmenlänge und 10 ms Rahmenüberlapp) in die spektrale Einhüllende und das spektral flache Fehlersignal zerlegt werden. (Diese Komponenten können ggf. schon aus dem Hochpaß-Erweiterungsverfahren übernommen werden.) Aus diesen beiden Komponenten werden das Tiefpaß-Anregungssignal (siehe Kapitel 3) und das Tiefpaß-Formfilter (siehe Kapitel 4) unterhalb 300 Hz berechnet. Der gesuchte Tiefpaß-Signalanteil wird gemäß dem linearen Modell der Spracherzeugung durch Anregung des Formfilters mit dem Anregungssignal erzeugt. Das bandbreitenerweiterte Signal erhält man schließlich durch Addition des synthetischen Tiefpaß-Signals zum Eingangssignal des Algorithmus.

## 3. TIEFPAß-ANREGUNGSSIGNAL

Das Tiefpaß-Anregungssignal muß spektral flach sein und soll ausschließlich Spektralkomponenten unterhalb 300 Hz aufweisen. Weiterhin soll für stimmhafte Laute die harmonische Pitchstruktur so gut wie möglich erhalten werden. Ein möglicher Ansatz ist die Anwendung einer Nichtlinearität. Die quadratische Funktion z.B. läßt sich leicht handhaben, vgl. [4]. Das so erhaltene Tiefpaß-Anregungssignal muß derart verstärkt werden, daß seine Leistung, bezogen auf 0.3kHz, gleich der Leistung des Fehlersignals, bezogen auf 7.7 kHz (8 kHz - 0.3 kHz), ist.

## 4. TIEFPAß-FORMFILTER

Als Tiefpaß-Formfilter wird ein ARMA Filter mit einer festgelegten Durchlaß-Charakteristik unterhalb 300 Hz verwendet. Der einzige Freiheitsgrad ist ein zusätzlicher Faktor  $\xi$  zur Anpassung der

Verstärkung. Dieser wird so berechnet, daß die spektrale Neigung zu Frequenzen unterhalb 300 Hz mittels des angepaßten ARMA Filters fortgesetzt wird. (Die prinzipielle Idee der fortgesetzten spektralen Neigung findet sich auch in [1].) Die spektrale Neigung wird aus der spektralen Einhüllenden (siehe Kapitel 2) extrahiert, vgl. Abb. 1.

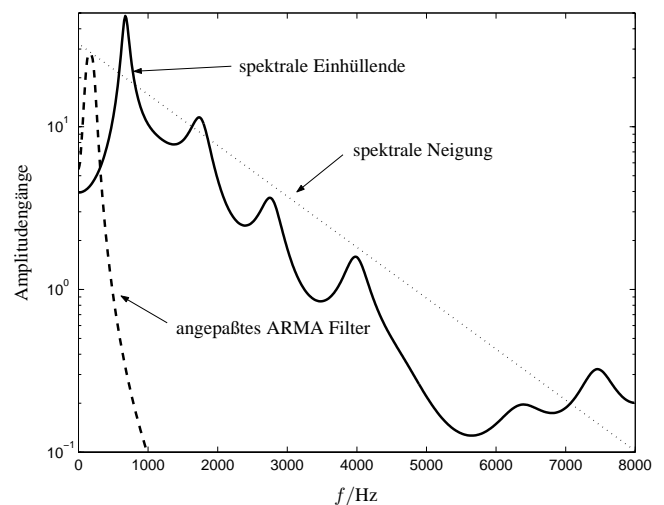


Abb. 1. Fortsetzung der spektralen Neigung zu den Frequenzen unterhalb 300Hz mittels des angepaßten ARMA Filters.

## 4.1. ARMA Filter

Das ARMA Filter  $G_{ARMA}(z)$  hat drei Nullstellen bei  $z = 0$ , eine Nullstelle bei  $z = 0.98$ , zwei konjugiert komplexe Polstellen bei  $z = 0.98e^{\pm j(2\pi 120Hz)/f_s}$ , und zwei konjugiert komplexe Polstellen bei  $z = 0.98e^{\pm j(2\pi 215Hz)/f_s}$  mit  $f_s = 16$  kHz (vgl. Gl. 1), motiviert durch die mittlere Pitchfrequenz von 120 Hz bei Männern bzw. 215 Hz bei Frauen für deutsche Sprache.

$$G_{ARMA}(z) = \frac{1.0 - 0.98z^{-1}}{1.0 - 3.9108z^{-1} + 5.7445z^{-2} - 3.7560z^{-3} + 0.9224z^{-4}} \quad (1)$$

Die Dämpfung oberhalb 300 Hz spielt keine entscheidende Rolle, da das Tiefpaß-Anregungssignal bereits auf das Frequenzband unterhalb 300 Hz beschränkt ist.

## 4.2. Anpassung der Verstärkung

Der Faktor  $\xi$  für das angepaßte ARMA Filter

$$G_{ad}(e^{j\Omega}) = \xi \cdot G_{ARMA}(e^{j\Omega})$$

wird separat für jeden Rahmen berechnet. Den Ausgangspunkt hierzu bildet die spektrale Einhüllende (siehe Kapitel 2), die durch ein All-Pol Filter  $H_{16}(e^{j\Omega})$  16. Ordnung (vgl. Gl. 2) beschrieben werden kann.

$$H_{16}(e^{j\Omega}) = \frac{1.0}{1.0 + \sum_{i=1}^{16} a_i e^{-j\Omega i}} \quad (2)$$

Der Anpassungsfaktor  $\xi$  wird nur durch die spektrale Neigung definiert, die aus  $H_{16}(e^{j\Omega})$  gewonnen wird. Hierzu wird zunächst die Ordnung des Filters  $H_{16}(e^{j\Omega})$  auf *eins* reduziert, d.h. das All-Pol Filter  $H_1(e^{j\Omega})$  erster Ordnung wird berechnet, vgl. Gl. 3.

$$H_1(e^{j\Omega}) = \frac{\tilde{b}_0}{1.0 + \tilde{a}_1 e^{-j\Omega}} \quad (3)$$

Für stimmhafte Laute klingt die spektrale Einhüllende zu höheren Frequenzen hin ab. Stimmlose Laute zeigen dagegen tendenziell eine mit zunehmender Frequenz ansteigende spektrale Einhüllende. Zum Beispiel zeigen Abb. 2 und Abb. 3 die Amplitudengänge von  $H_{16}(e^{j\Omega})$  und  $H_1(e^{j\Omega})$  jeweils für einen stimmhaften Laut ( $H_{16}^{(v)}(e^{j\Omega})$ ,  $H_1^{(v)}(e^{j\Omega})$ ) und einen stimmlosen Laut ( $H_{16}^{(uv)}(e^{j\Omega})$ ,  $H_1^{(uv)}(e^{j\Omega})$ ).

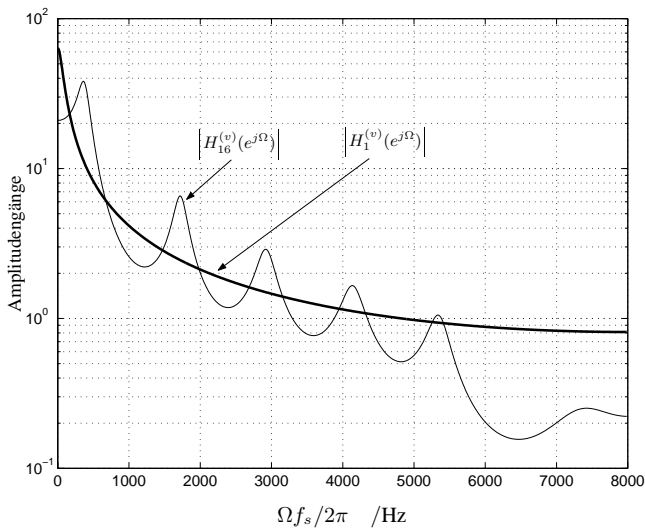


Abb. 2. Amplitudengänge  $|H_{16}^{(v)}(e^{j\Omega})|$  und  $|H_1^{(v)}(e^{j\Omega})|$  für einen stimmhaften Laut.

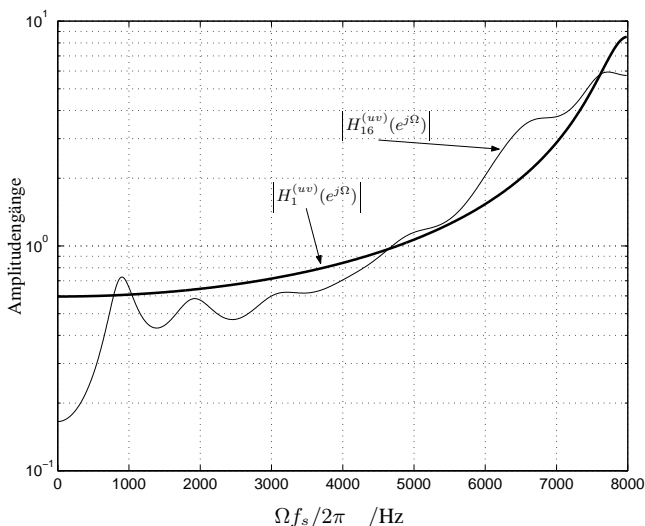


Abb. 3. Amplitudengänge  $|H_{16}^{(uv)}(e^{j\Omega})|$  und  $|H_1^{(uv)}(e^{j\Omega})|$  für einen stimmlosen Laut.

Die beiden Koeffizienten  $\tilde{a}_1$  und  $\tilde{b}_0$  von  $H_1(e^{j\Omega})$  können mittels der step-down Rekursion berechnet werden. Diese bildet die 16 Filterkoeffizienten ( $a_1 \dots a_{16}$ ) von  $H_{16}(e^{j\Omega})$  auf die 16 Reflexionskoeffizienten ( $\Gamma_1 \dots \Gamma_{16}$ ) ab. Die Koeffizienten von  $H_1(e^{j\Omega})$

sind dann gemäß Gl. 4 und 5 definiert.

$$\tilde{a}_1 = \Gamma_1 \quad (4)$$

$$\tilde{b}_0 = \frac{1.0}{\sqrt{\prod_{i=2}^{16} (1.0 - |\Gamma_i|^2)}} \quad (5)$$

Der Anpassungsfaktor  $\xi$  wird schließlich als Verhältnis der Amplituden  $|H_1(e^{j\Omega_c})|$  und  $|G_{ARMA}(e^{j\Omega_c})|$  bei einer bestimmten Frequenz  $\Omega_c$  unterhalb 300 Hz berechnet, z.B. bei  $\Omega_c = 150\text{Hz } 2\pi/f_s$ , siehe Gl. 6.

$$\xi = \frac{|H_1(e^{j\Omega_c})|}{|G_{ARMA}(e^{j\Omega_c})|} \quad (6)$$

## 5. NACHVERARBEITUNG

Aufgrund von Schätzungenauigkeiten weist der Faktor  $\xi$  manchmal zeitliche Schwankungen auf, was zu rumpelnden Störungen im synthetisierten Signal führen würde. Eine einfache Abhilfe hierzu ist die zeitliche Glättung des Faktors. Da der Einfluß von  $\xi$  auf den Erweiterungsvorgang offensichtlich ist, können weitere, anspruchsvollere Nachverarbeitungsvarianten einfach realisiert werden.

## 6. ERGEBNIS UND ZUSAMMENFASSUNG

Die vorgestellte Methode erlaubt die spektrale Erweiterung von Telefonsprache zu Frequenzen unterhalb 300 Hz. Hierzu wird ein ARMA Filter mit einer Durchlaß-Charakteristik unterhalb 300 Hz derart angepaßt, daß bei einer bestimmten Frequenz unterhalb 300 Hz seine Amplitude und die durch die spektrale Neigung definierte Amplitude gleich sind. Die bei einem Codebuch-Ansatz beobachteten rumpelnden Störgeräusche treten bei der vorgeschlagenen Methode nur noch sehr abgeschwächt auf. Weiterhin ist eine Nachverarbeitung wegen des offensichtlichen Einflusses der angepaßten ARMA Filterverstärkung auf die synthetisierte Sprache einfach möglich. Da der vorgeschlagene Algorithmus in sich abgeschlossen ist, kann er mit jedem Hochpaß-Erweiterungssystem verbunden werden.

## 7. LITERATUR

- [1] H. Yasukawa: *Restoration of Wide Band Signal From Telephone Speech Using Linear Prediction Error Processing*, Proc. International Conference on Spoken Language Processing, Vol. II, 901 - 904, 1996.
- [2] N. Enbom, W. B. Kleijn: *Bandwidth Expansion of Speech Based on Vector Quantization of the MEL Frequency Cepstral Coefficients*, Proc. IEEE Workshop on Speech Coding 1999, 171 - 173, 1999.
- [3] G. Miet, A. Gerrits, J. C. Valière: *Low-Band Extension of Telephone-Band Speech*, Proc. ICASSP 2000, Vol. III, 1851 - 1854, 2000.
- [4] U. Kornagel: *Spectral Widening of the Excitation Signal for Telephone-Band Speech Enhancement*, Proc. IWAENC 2001, 215 - 218, 2001.
- [5] P. Jax, P. Vary: *Enhancement of Band-Limited Speech Signals*, 10th Aachen Symposium on Signal Theory, 331 - 336, 2001.
- [6] J. A. Fuemmeler, R. C. Hardie, W. R. Gardner: *Techniques for the Regeneration of Wideband Speech from Narrowband Speech*, EURASIP Journal on Applied Signal Processing, Vol. 2001, Nr. 4, 266 - 274, 2001.
- [7] U. Kornagel: *Spectral Widening of Telephone Speech Using an Extended Classification Approach*, Proc. EUSIPCO 2002, Vol. II, 339 - 342, 2002.