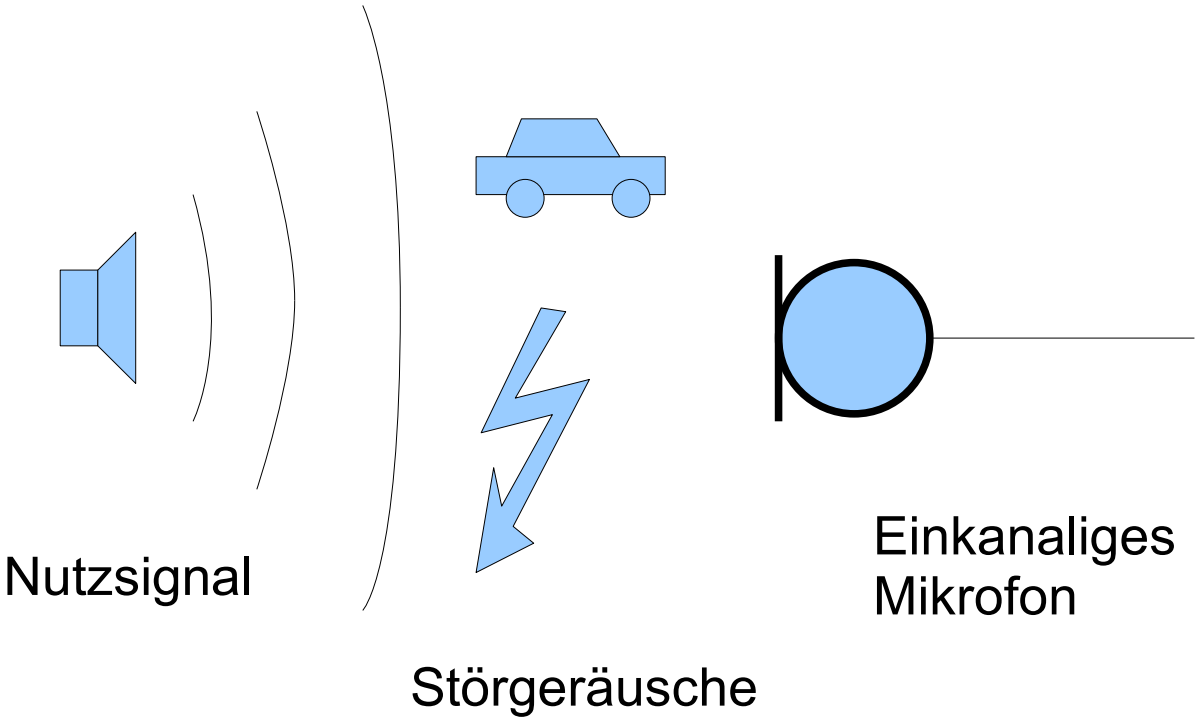
Störgeräuschreduktion bei stimmhaften Sprachsignalen



Allgemein: einkanalige Störgeräuschreduktion

Gliederung

- Störgeräuschreduktion: Arten und Einsatzgebiete
- Arten der Anregung des Sprechtrakts
  - Lineares Modell der Erzeugung stimmhafter Sprache
- Analyse
  - Fensterung des Sprachsignals
  - Spektrum eines gefenst. stimmhaften Sprachsignals
- Der Algorithmus zur Störgeräuschreduktion
  - Beispiel: stimmhaftes Sprachsignal
  - Spektrum bei hohen Frequenzen
  - Grundfrequenzschätzung
  - Grenzen des Algorithmus, Ausblick

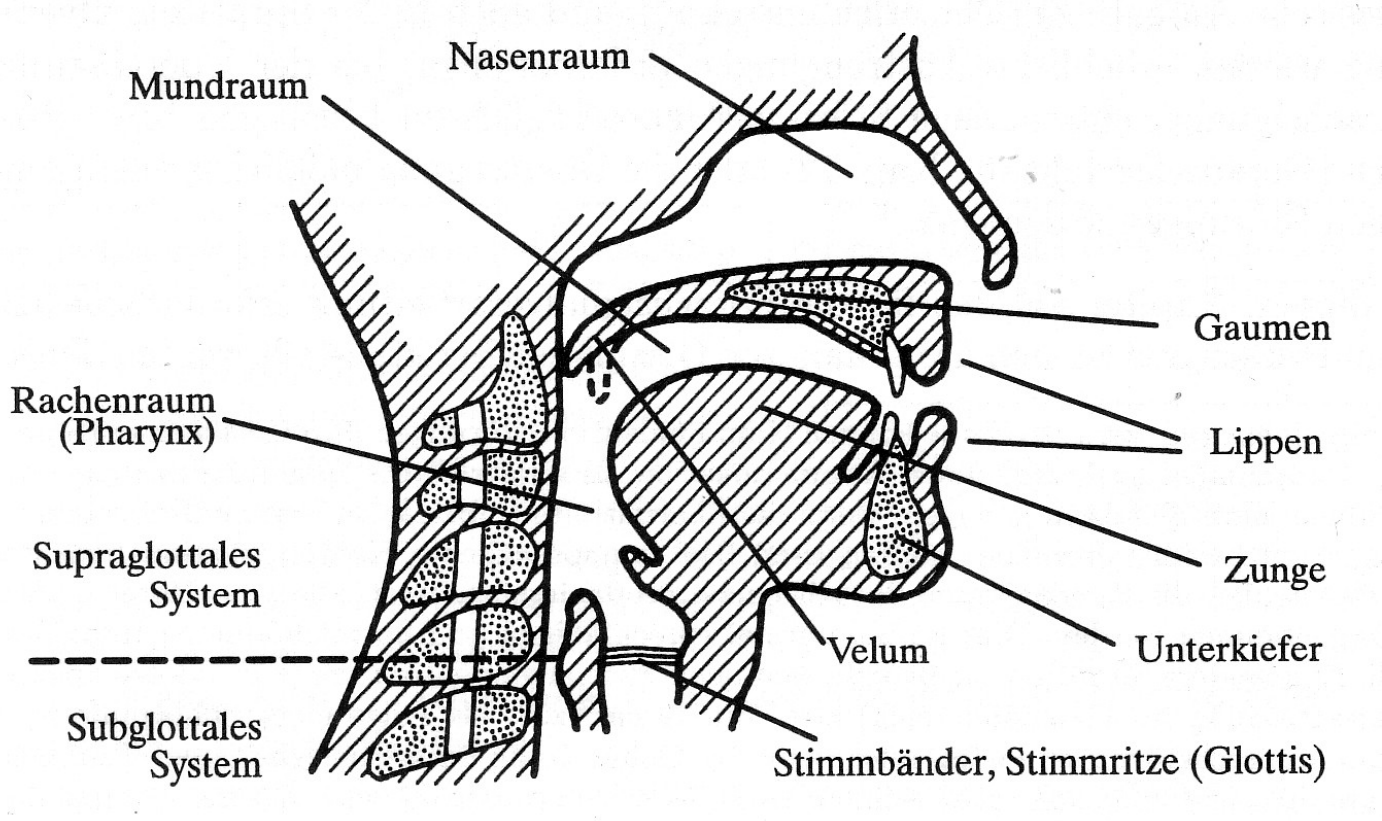
Störgeräuschreduktion: Arten und Einsatzgebiete

- Einkanalige Verfahren
  - Statistische Eigenschaften der Störung:

    stationär/nichtstationär
  - Optimalfilter (Wienerfilter): Ermitteln des Leistungsdichtespektrums des Störgeräusches in Sprachpausen
  - Spektrale Subtraktion
- Einsatzgebiete einkanaliger Verfahren
  - Automatische Spracherkennungssysteme
  - Verbesserung der Sprachqualität z.B. im Mobilfunk

Arten der Anregung des Sprechtrakts (1/2)

- Stimmlose und transiente Anregung



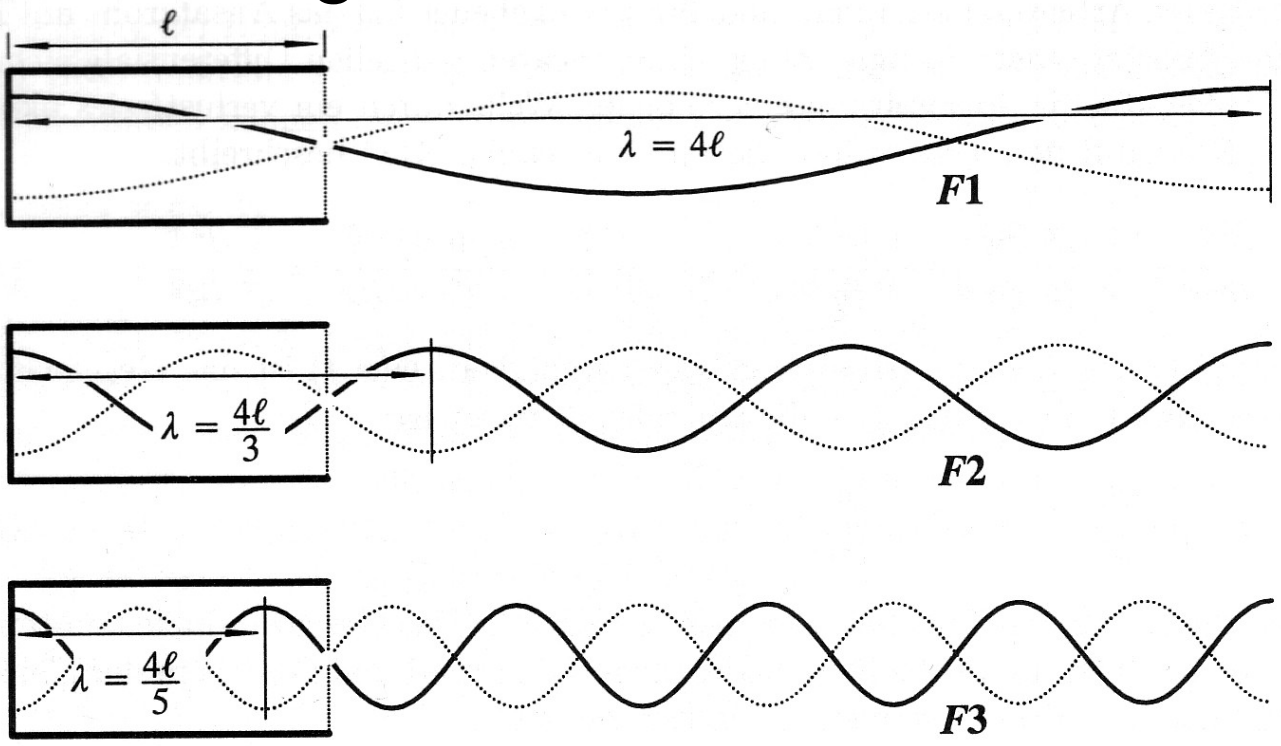
Arten der Anregung des Sprechtrakts (2/2)

- **Stimmhafte Anregung** der Stimmbänder (Phonation):
  - Druckabfall aufgrund des Bernoulli-Effekts führt zu abruptem Verschluss der Glottis

- Schwingung mit Grundfrequenz F_0

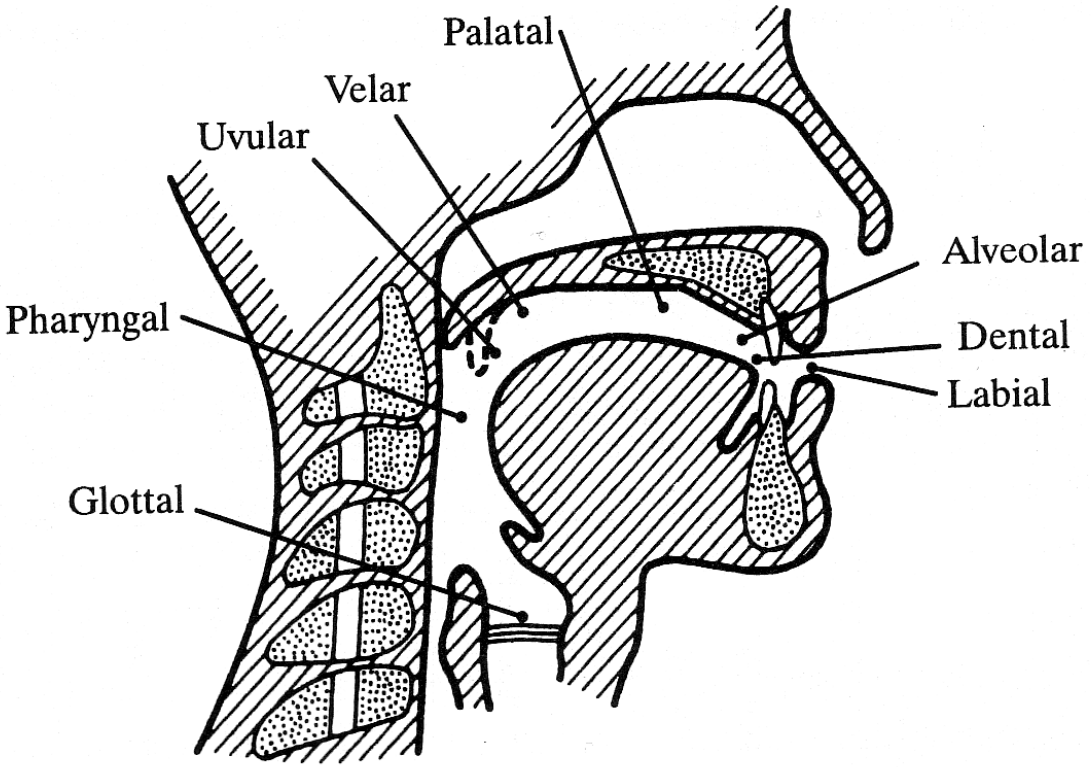
Vokaltrakt als akustisches Rohr

- Für folgende Wellenlängen ergeben sich stehende Wellen im näherungsweise zylinderförmigen Vokaltrakt

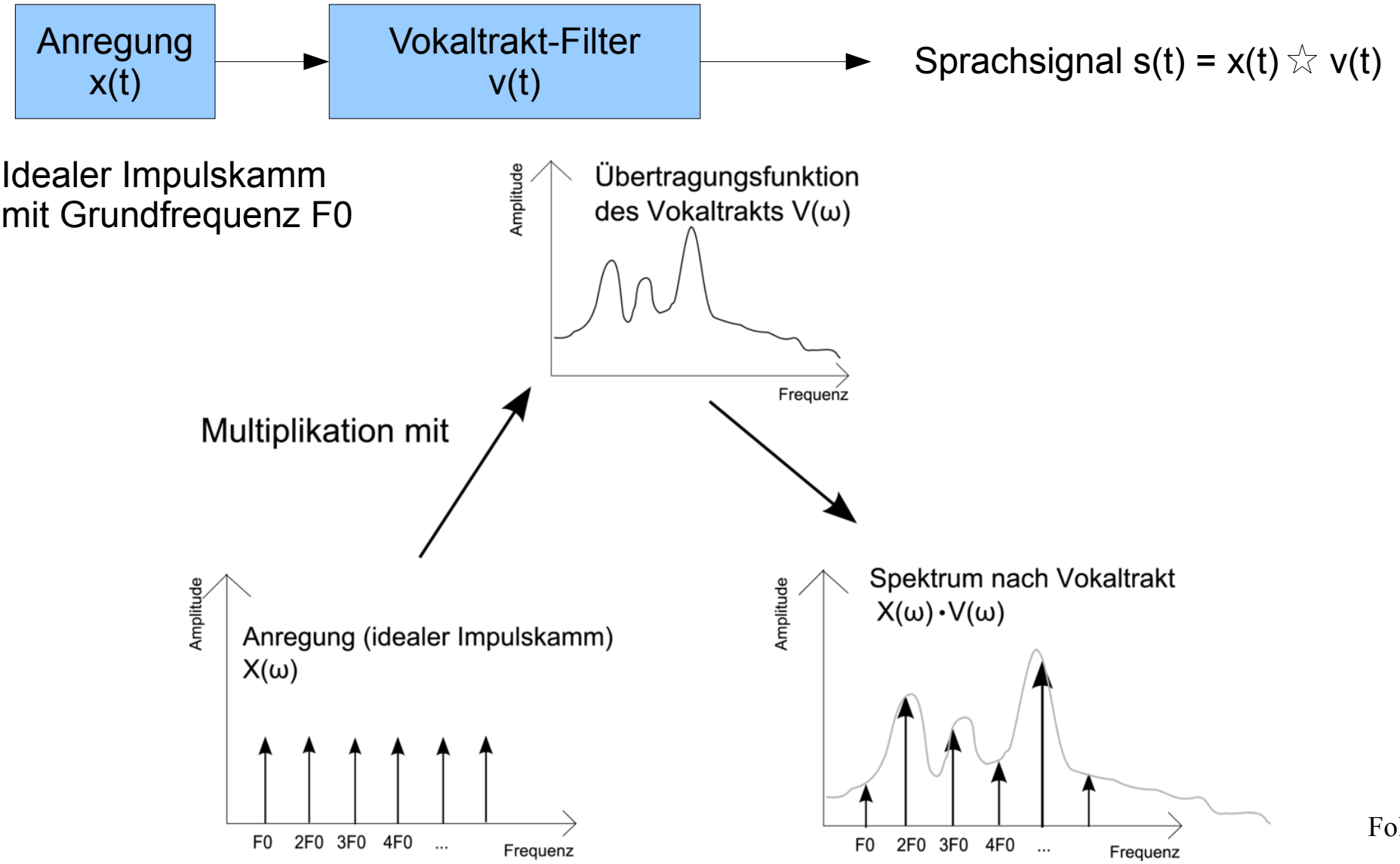


Vokaltrakt

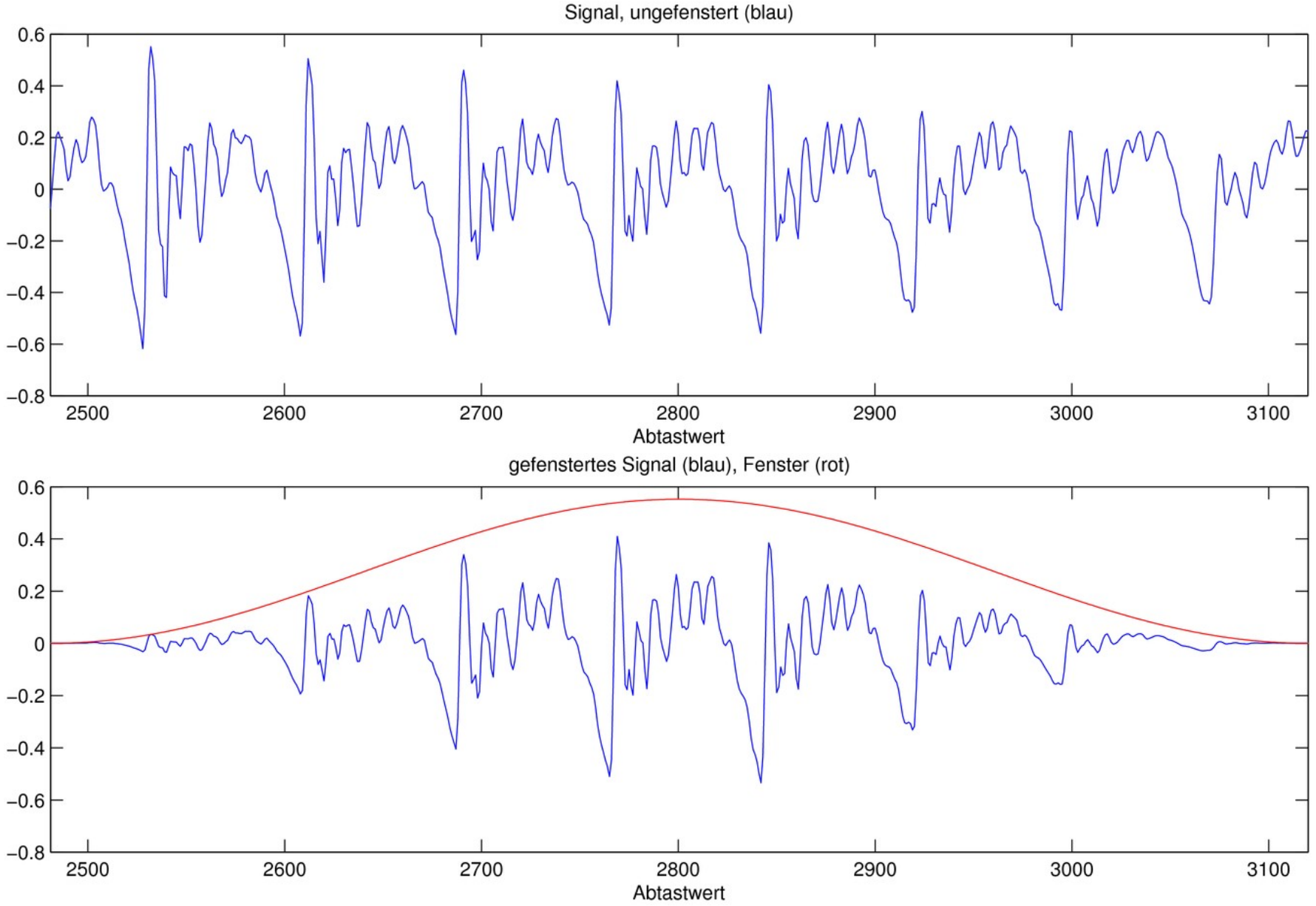
- Filterung durch den Vokaltrakt:
  - Resonanzfrequenzen bei $(2k-1) \cdot 500\text{Hz}$, $k=1,2,...$ (Formanten)
  - Durch Änderung der Form (Röhrenmodell)



Lineares Modell der Erzeugung stimmhafter Sprache

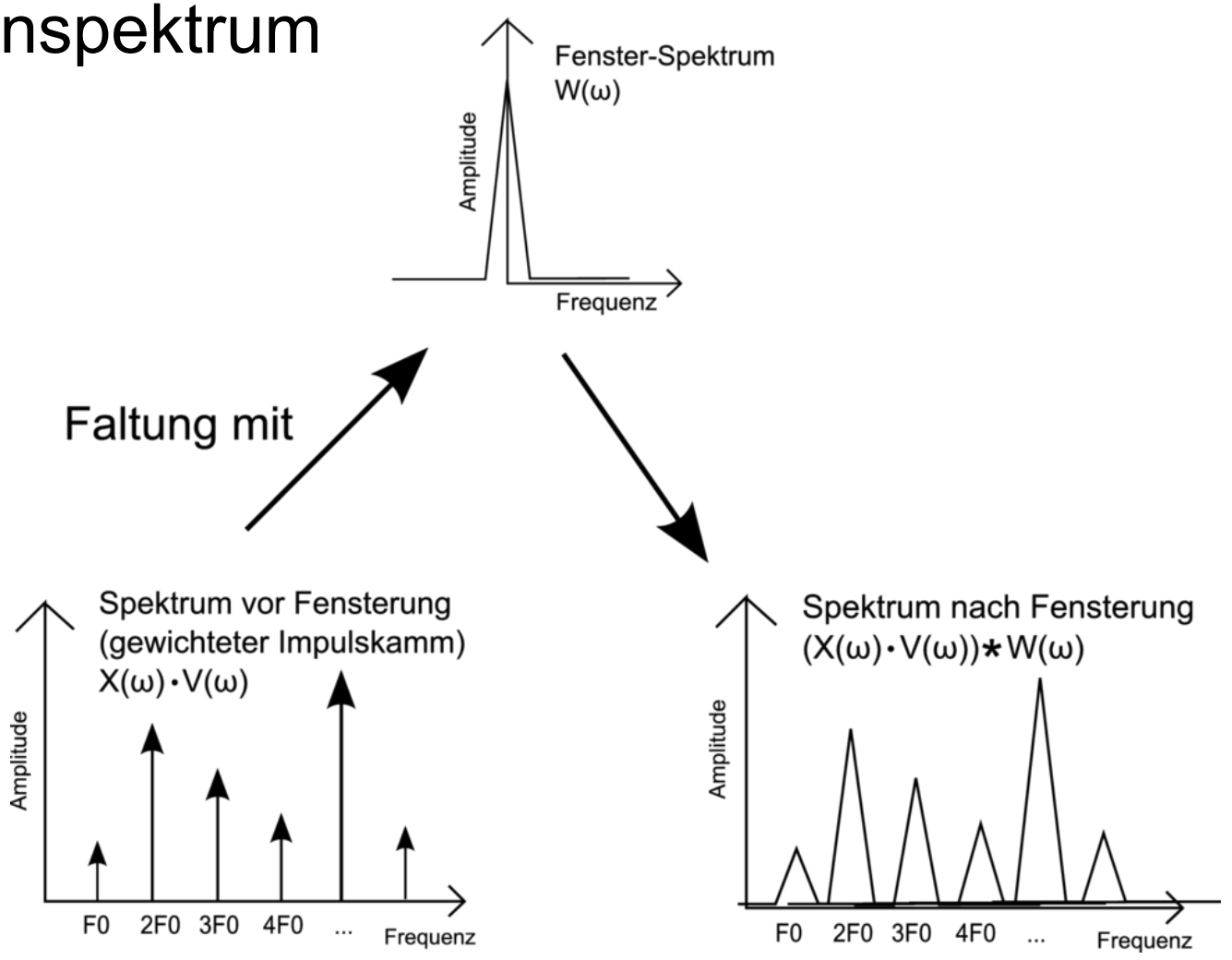


Analyse: Fensterung des Sprachsignals



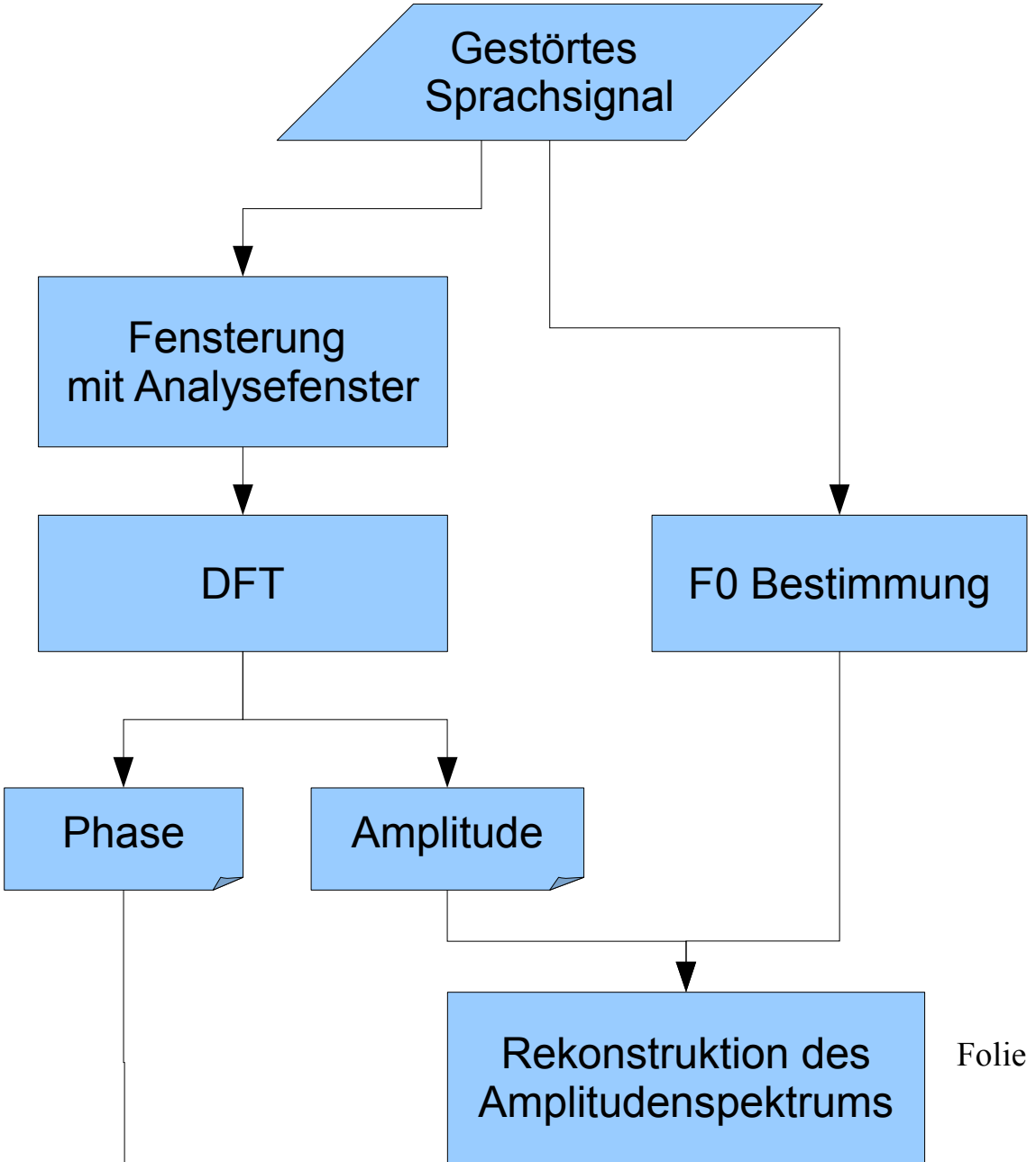
Spektrum eines gefensterten stimmhaften Sprachsignals

- Quasiperiodisches Signal hat näherungsweise Linienspektrum



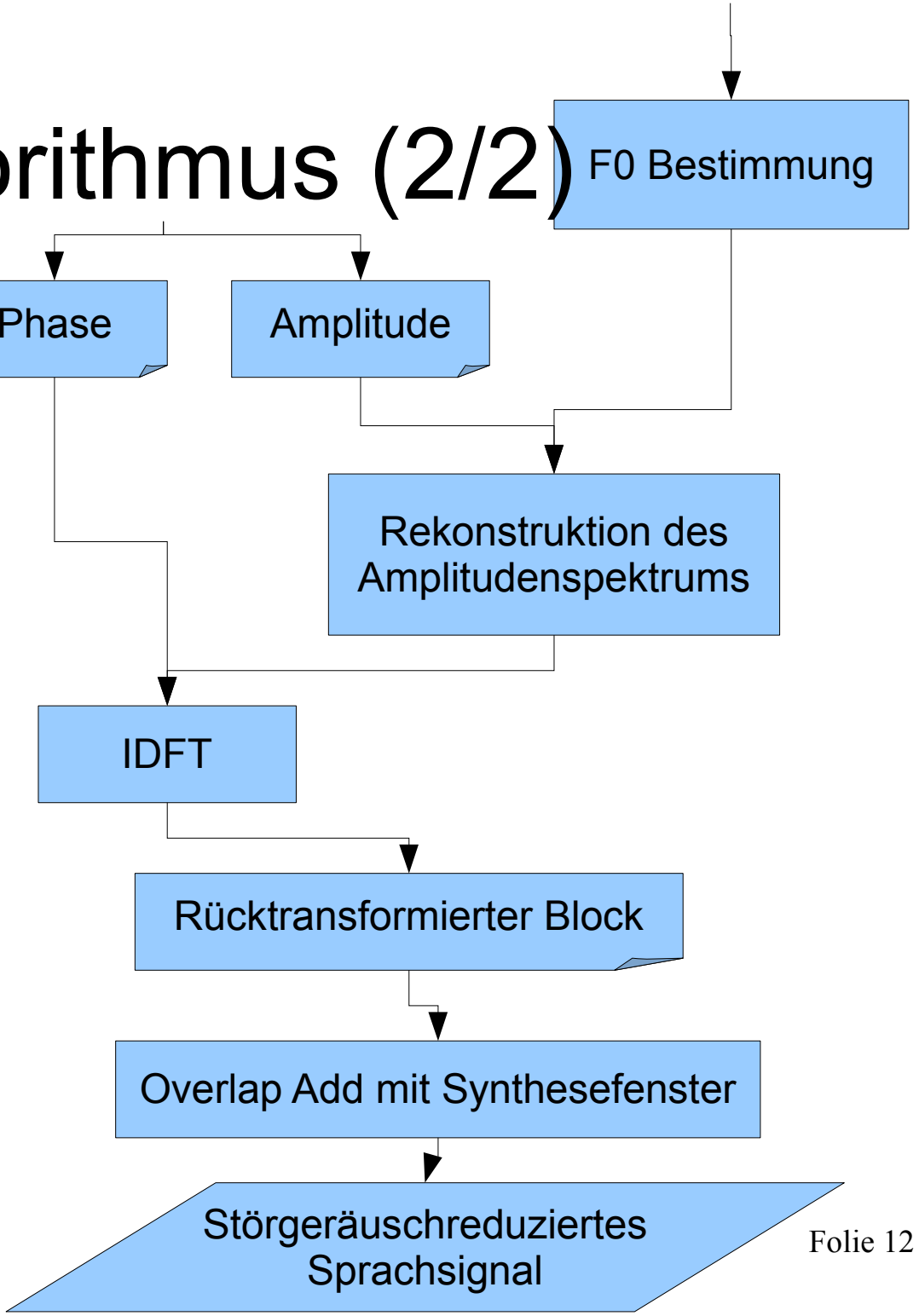
Der Algorithmus (1/2)

- Blockweise Verarbeitung des Sprachsignals
- Grundfrequenz-bestimmung (F0)
- Diskrete Fourier-Transformation (DFT) mit Analysefenster, nennt sich auch Short Time Fourier Transfom (STFT)



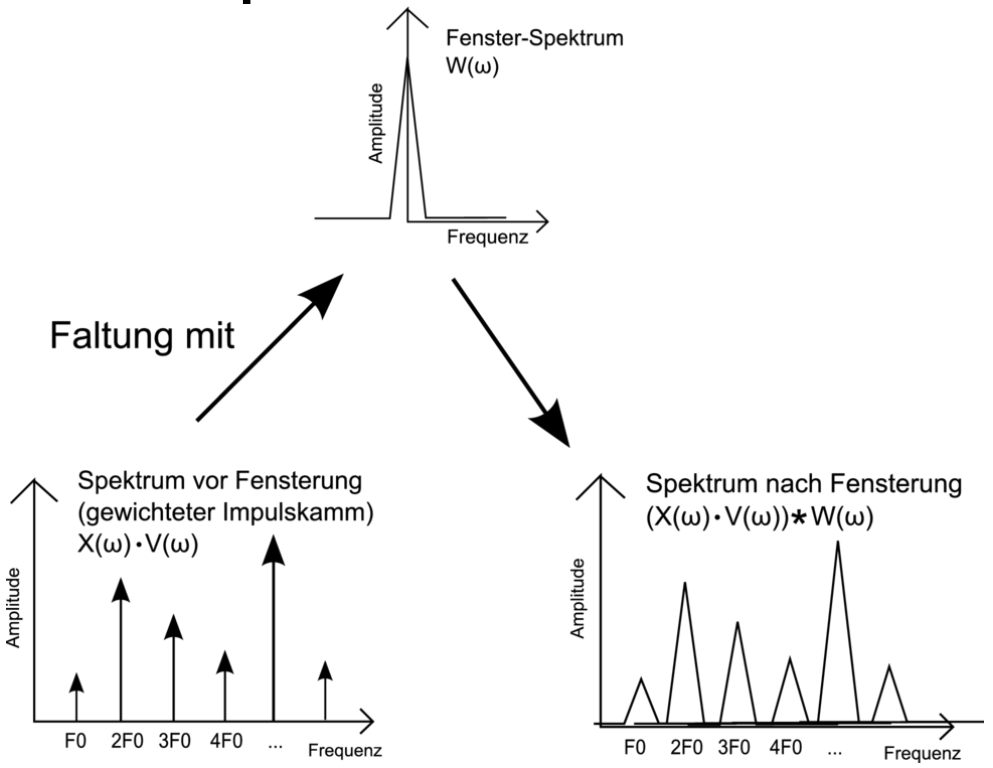
Der Algorithmus (2/2)

- Rekonstruktion des Amplitudenspektrums
- Verwenden der Original-Phase zur Rücktransformation (IDFT)
- Overlap Add mit Synthesefenster



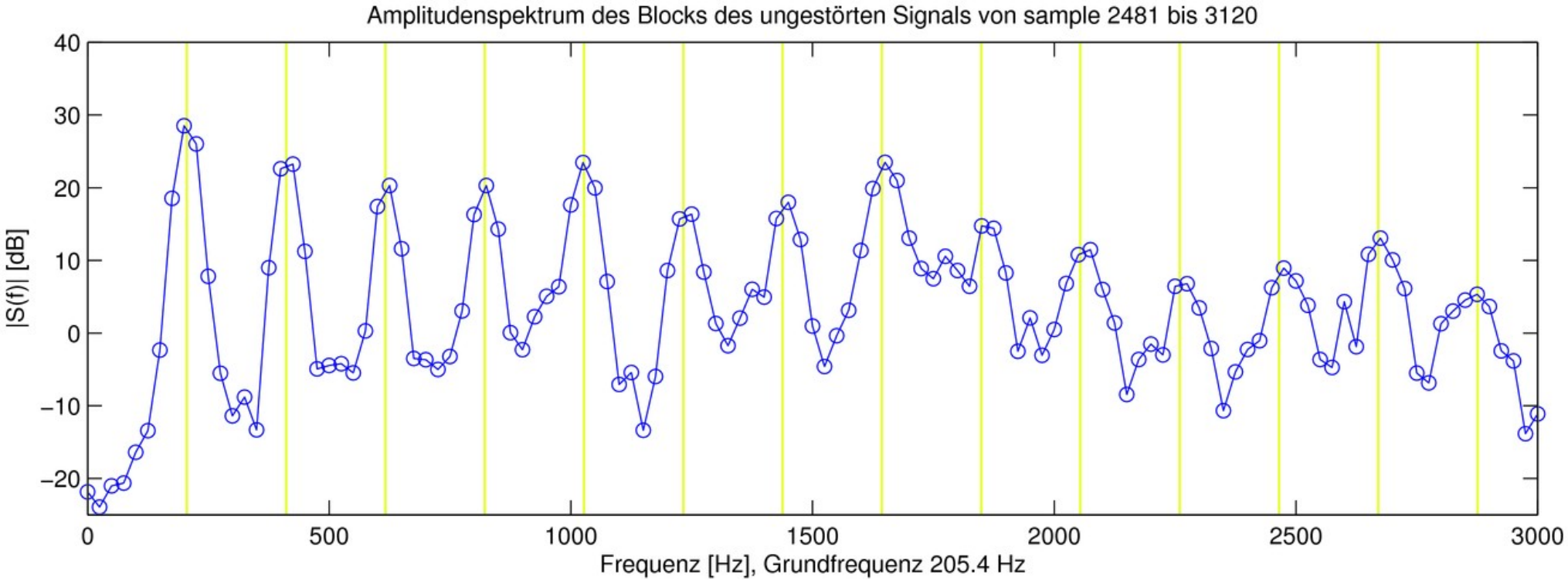
Rekonstruktion des Amplitudenspektrums

- Bestimmung der Höhe des Spektrums an den Vielfachen der Grundfrequenz F0
- Faltung des so erhaltenen Linienspektrums mit dem Fenster-Spektrum



Beispiel: stimmhaftes Sprachsignal[1]

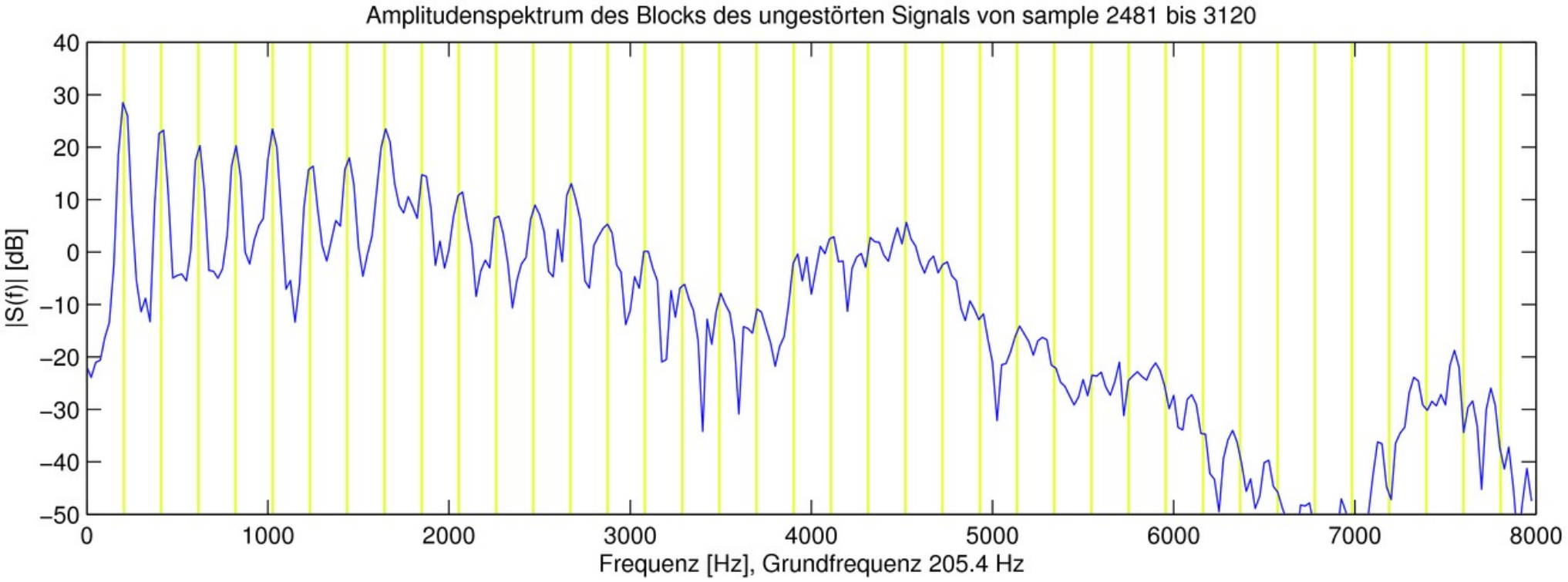
- Spektrum des ungestörten Sprachsignals
  - Blocklänge für DFT: 40ms
  - Grundfrequenzschätzung: autocorr2
  - Fensterfunktion: von Hann-Fenster



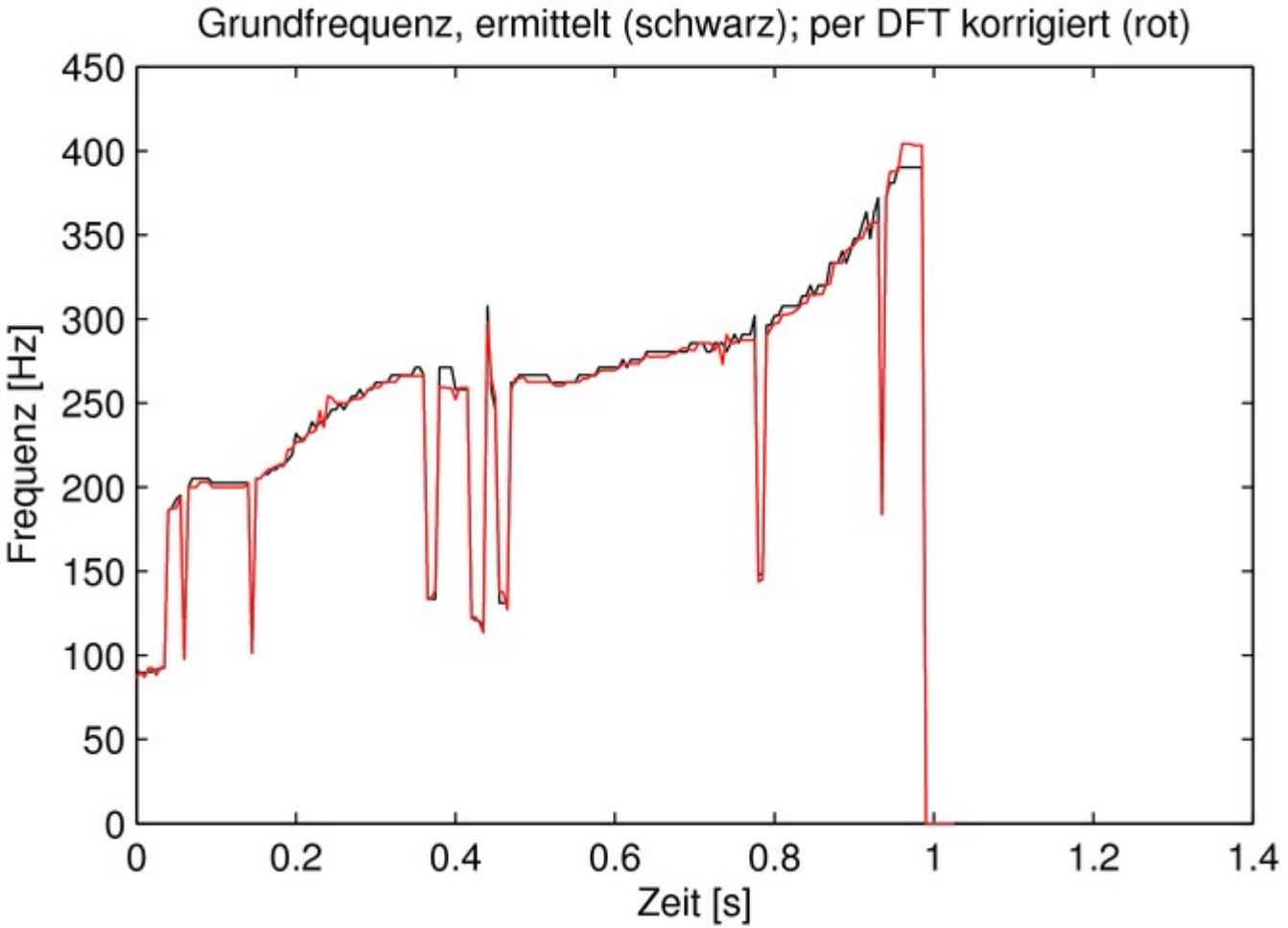
Spektrum an der gleichen Stelle des Sprachsignals[1] bis 8kHz

Man erkennt:

- Bei höheren Frequenzen weniger harmonisch



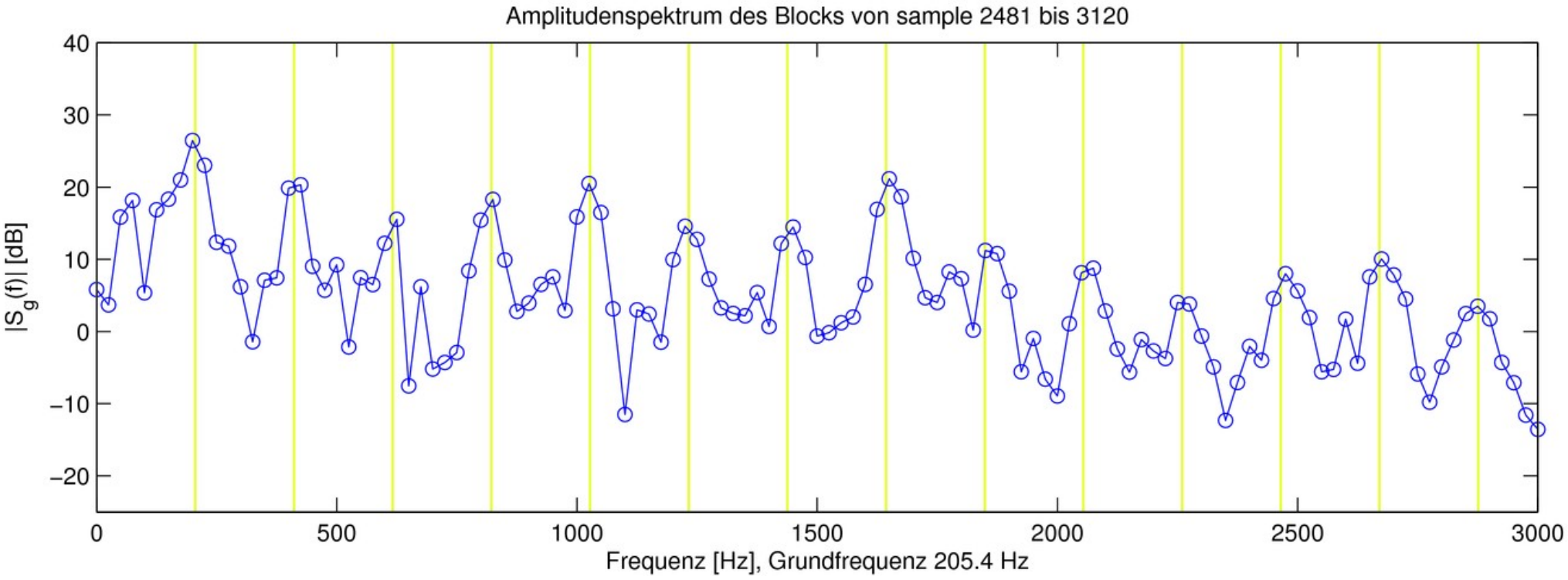
Grundfrequenz, ermittelt aus gestörtem Sprachsignal[1]



Spektrum des gestörten Sprachsignals[1]

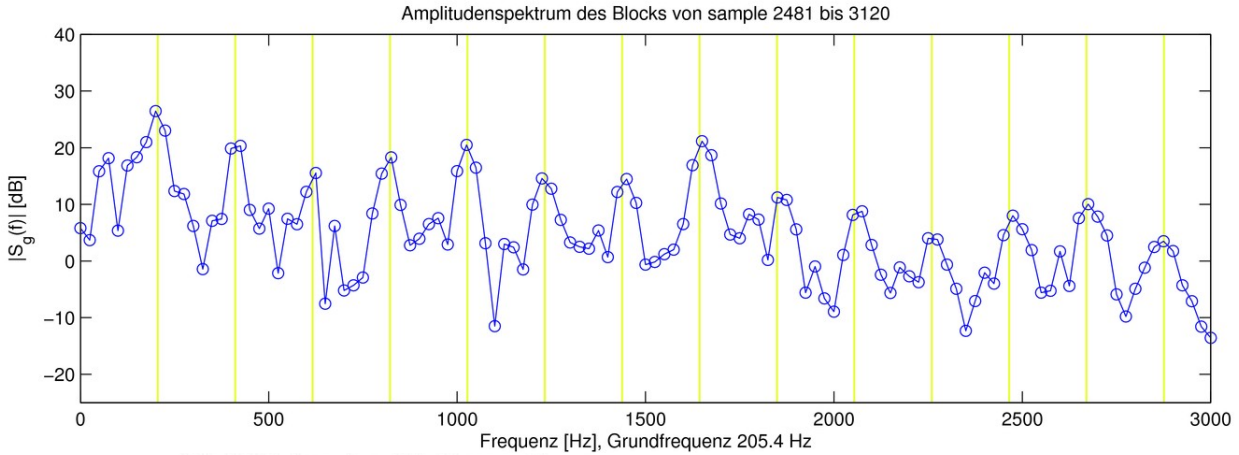
$$H_1(z) = \frac{1}{1 - 0.9z^{-1}} \quad H_2(z) = \frac{1}{1 - 0.6z^{-1}}$$

$$H(z) = H_1(z) \cdot H_2(z)$$

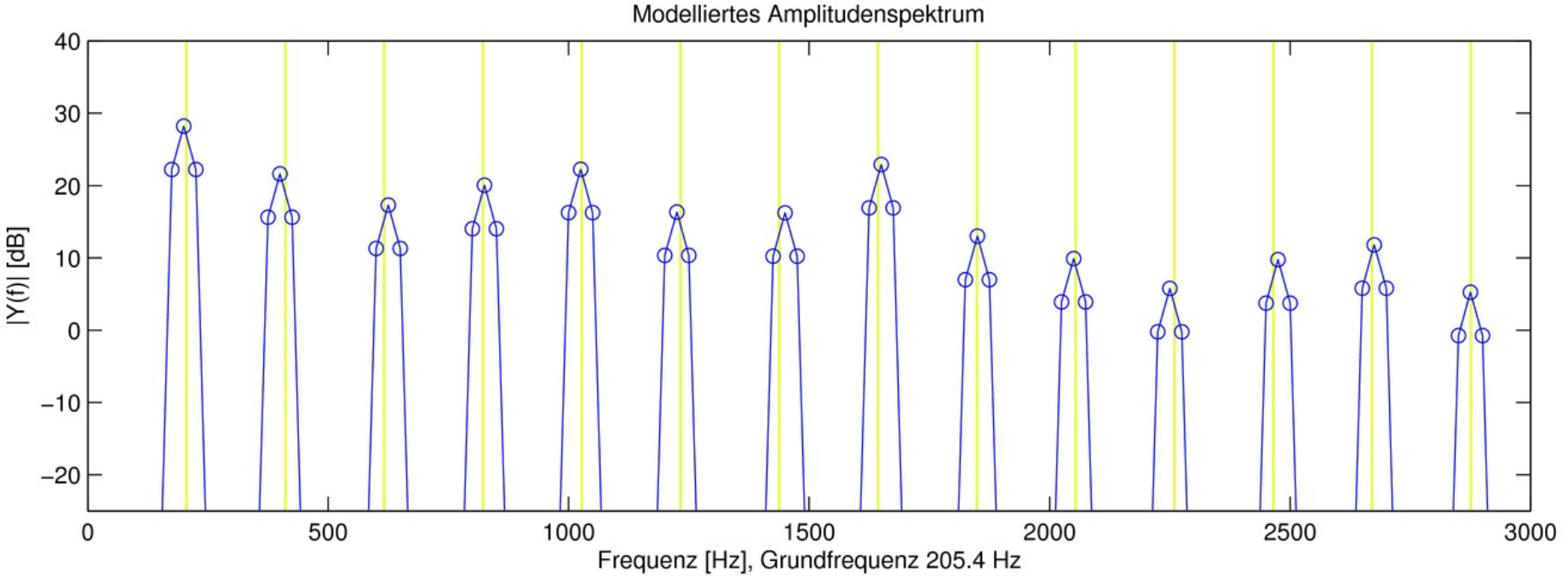


Spektrum des verarbeiteten Sprachsignals[1]

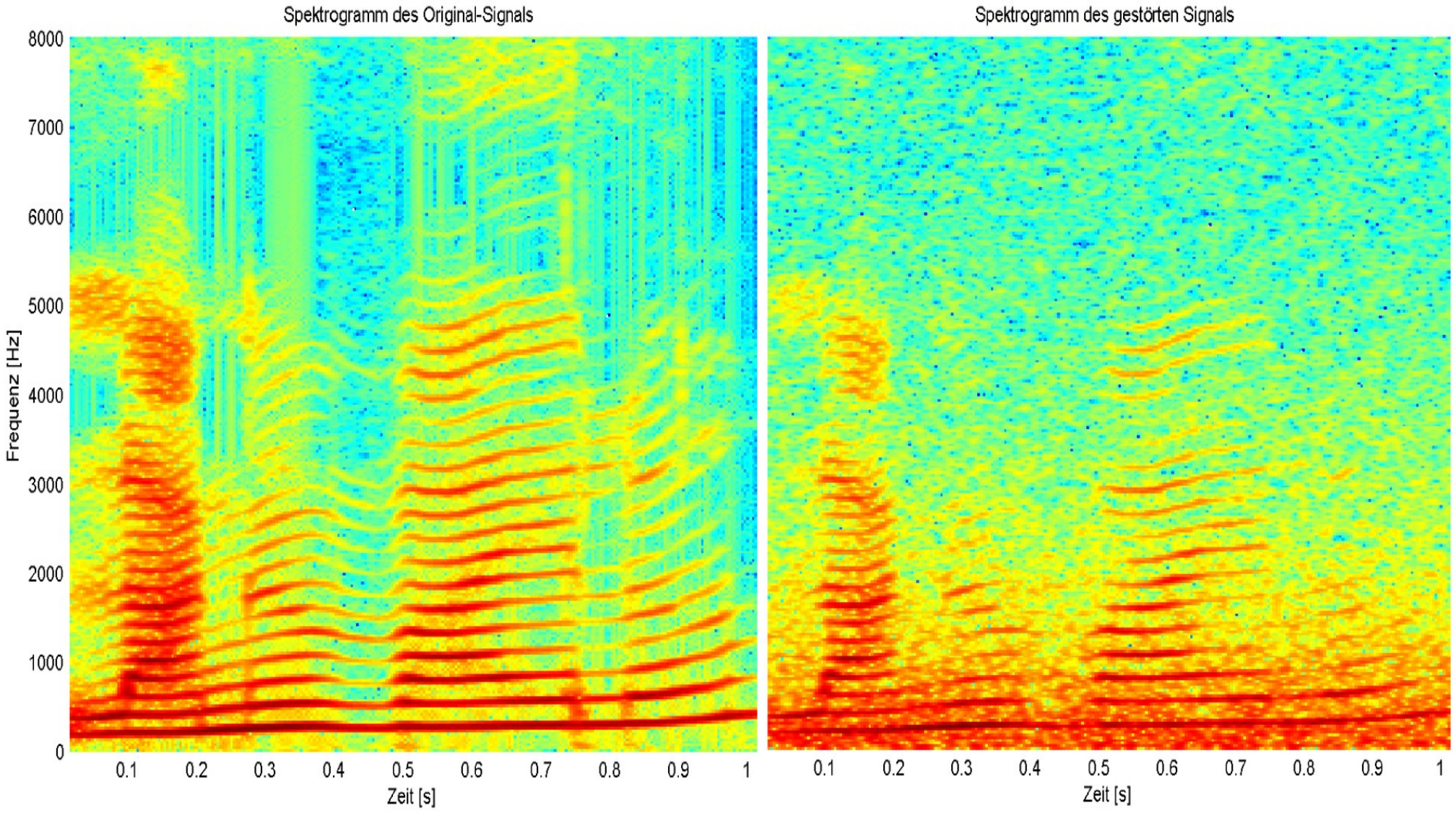
gestört:



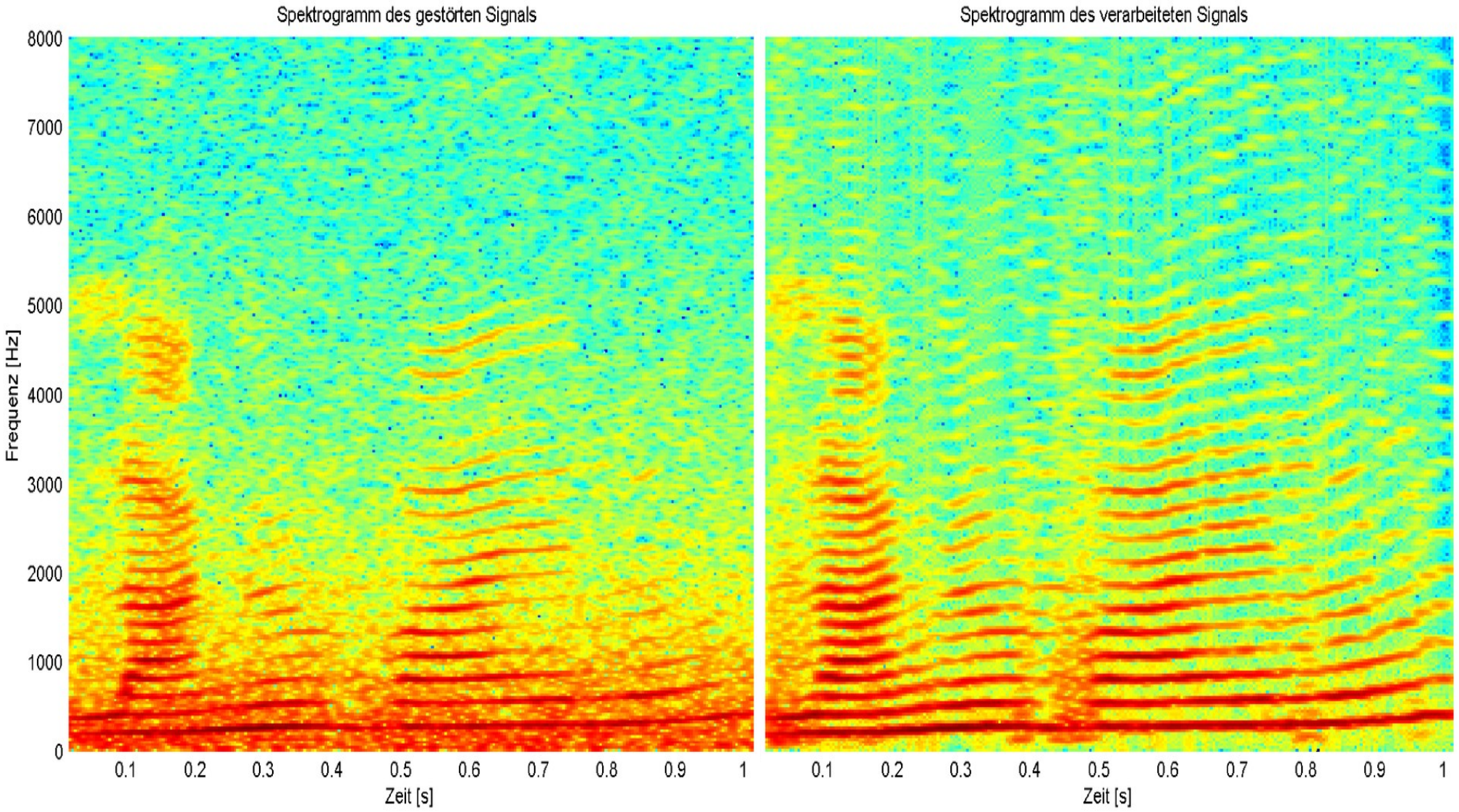
verarbeitet:



Spektrogramme des Sprachsignals[1] (1/2)

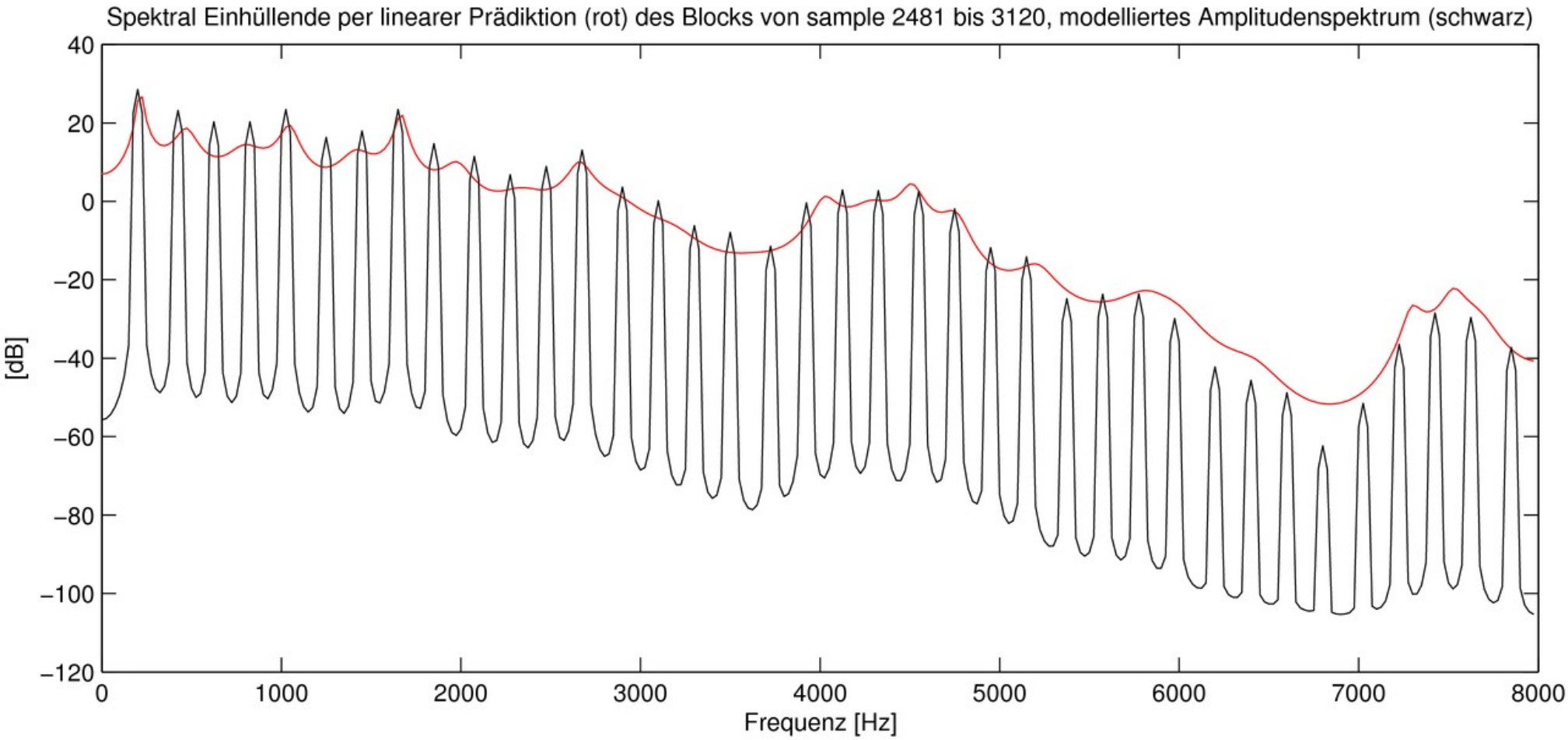


Spektrogramme des Sprachsignals[1] (2/2)



Spektrum bei hohen Frequenzen

lin.Präd.: lin. Präd. ab 5 kHz: Ohne lin. Präd.: gestört:



Grundfrequenzschätzung

- AMDF (average magnitude difference function)

$$AMDF(\lambda) = \sum_{i=1}^{N} |x(i) - x(i+\lambda)|, \quad T_0 = \frac{1}{f_A} \arg\min_{\lambda} AMDF(\lambda)$$

- AKF (Autokorrelationsfunktion)

$$\widetilde{AKF}(\lambda) = \sum_{i=1}^{N} s(i)s(i+\lambda)$$

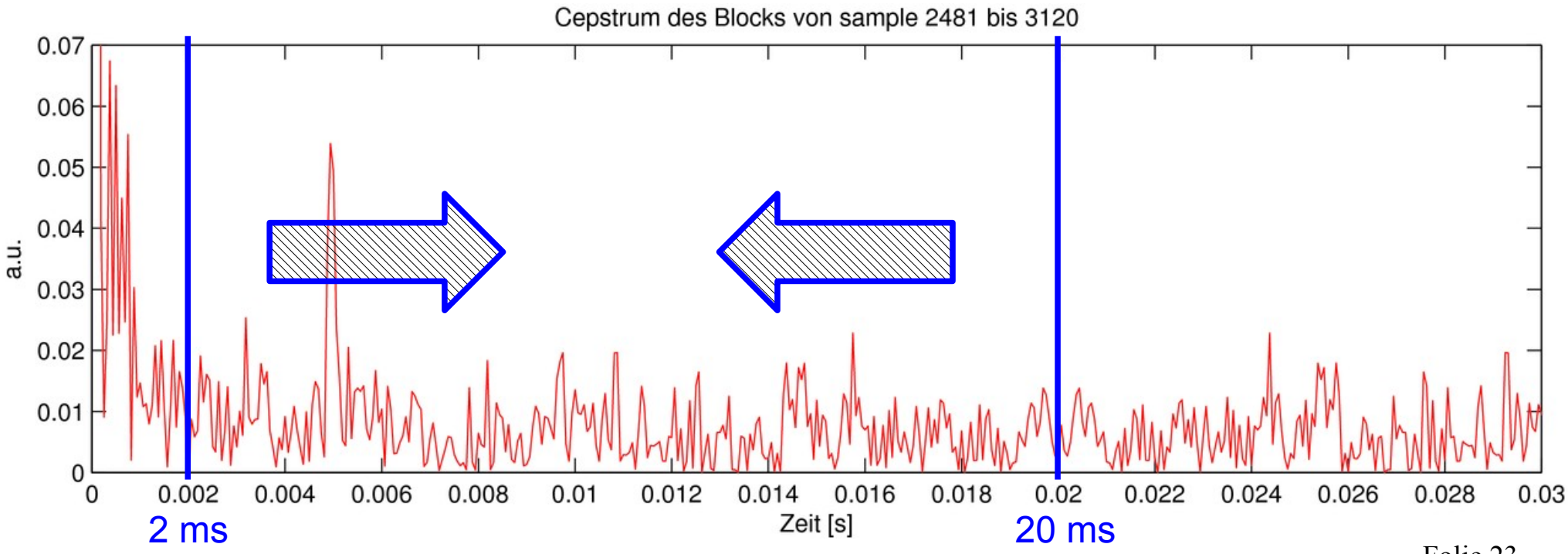
- Kumulantfolge 3. Ordnung

$$c_{xxx2}(k) = \sum_{i=0}^{L-k-1} s^2(i) \cdot s(i+k), \quad k = -(L-1), ..., L-1$$

Grundfrequenzschätzung per Cepstrum

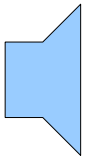
$$X_\mu = \mathrm{DFT}\{x_n\} = \sum_{n=0}^{N-1} x_n e^{-i\frac{2\pi}{N}\mu n} \qquad \mu = 0, \ldots, N-1$$

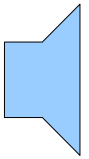
$$C_x(n) = \mathrm{IDFT}\{\ln |X_\mu|\}$$

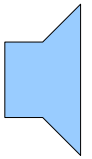


Grenzen des Algorithmus

- Fehlschlagen der Grundfrequenzbestimmung bei starken Störungen

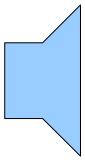

gestört


cumul2_bandpass

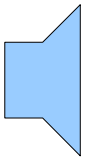

cheat_gf_ma

- Bei nichtstationären Störgeräuschen im verarbeiteten Signal immer noch nichtstationäre Reststörung enthalten

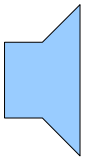
Originalsignal



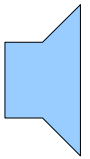
gestörtes Signal



verarbeitetes Signal



verarbeitetes Signal
mit Grundfrequenz aus Original



Ausblick

- Programmierung eines VUD (Voiced Unvoiced Detector) ist für die Anwendung auf allgemeine Sprachsignale nötig
- Verbesserung der Grundfrequenzschätzung würde zu besseren Ergebnissen führen
- Echtzeit-Implementierung
- Code optimieren um Rechenzeit zu verringern – je nach Einstellung der „hopsize“ und Methode zur Grundfrequenzschätzung bis zu 50 Mal langsamer als Echtzeit

Zusammenfassung

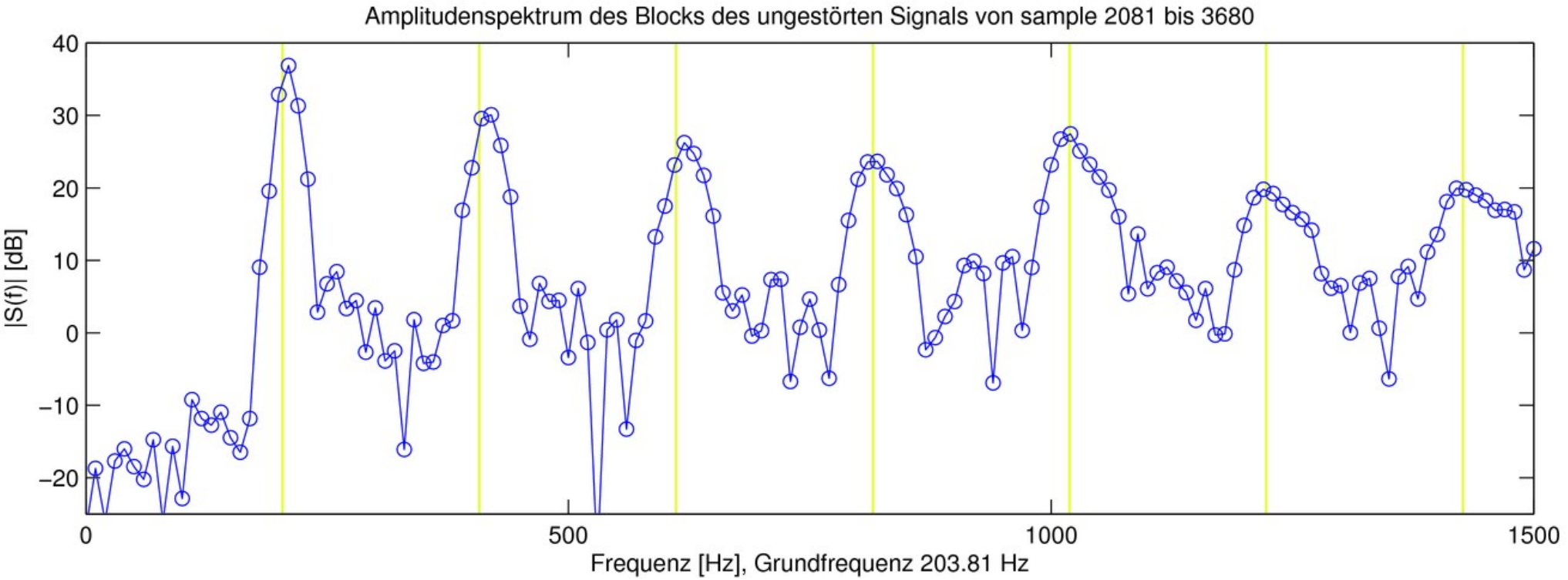
- Einkanalige Störgeräuschreduktion
- Modelle der Sprachproduktion
  - akustisches Rohr
  - Röhrenmodell
  - Lineares Modell
- STFT-basierter Algorithmus zur Störgeräuschreduktion stimmhafter Sprachsignale
  - Beispiele: Spektren, Hörproben, Spektrogramme
  - Methoden der Grundfrequenzschätzung

Fragen?

Vielen Dank für Ihre Aufmerksamkeit.

Beispiel 2: Ungestörtes stimmhaftes Sprachsignal[1] mit längeren Blöcken

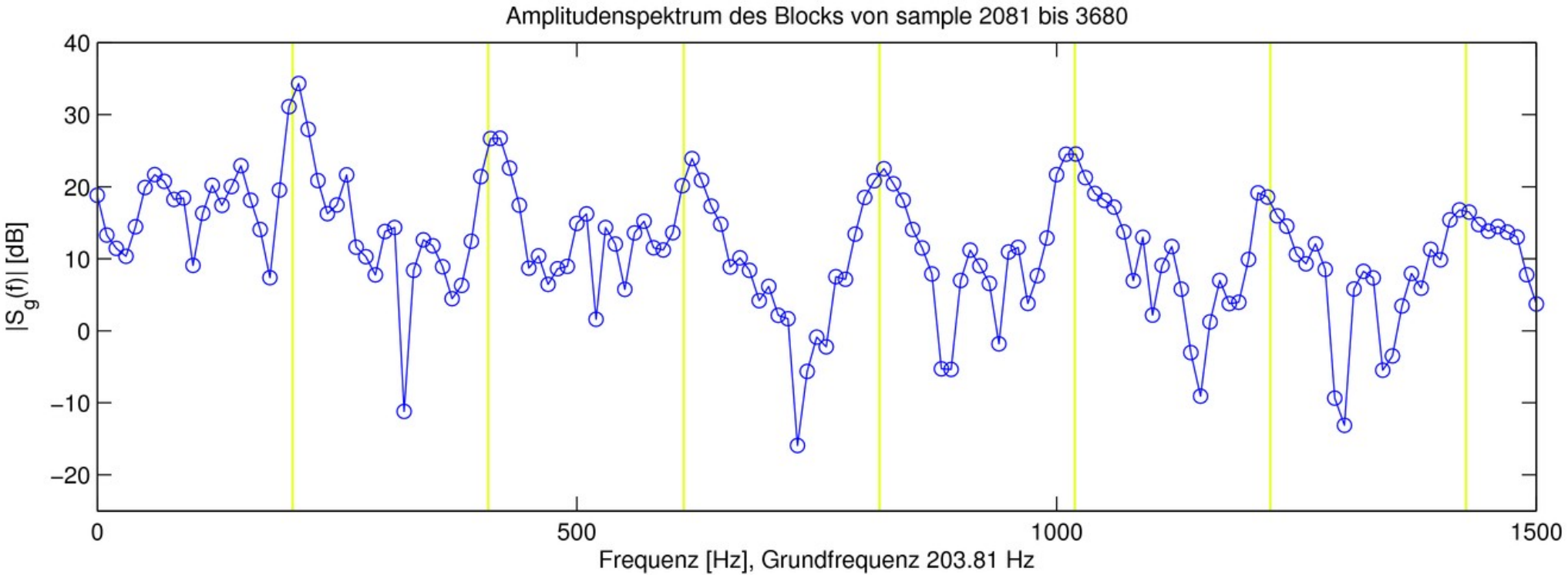
- Spektrum des ungestörten Sprachsignals
  - Blocklänge für DFT: 100ms
  - Grundfrequenzbestimmung: autocorr2
  - Fensterfunktion: von Hann-Fenster



Sprachsignal[1], gestört mit längeren Blöcken

$$H_1(z) = \frac{1}{1 - 0.9z^{-1}} \quad H_2(z) = \frac{1}{1 - 0.6z^{-1}}$$

$$H(z) = H_1(z) \cdot H_2(z)$$

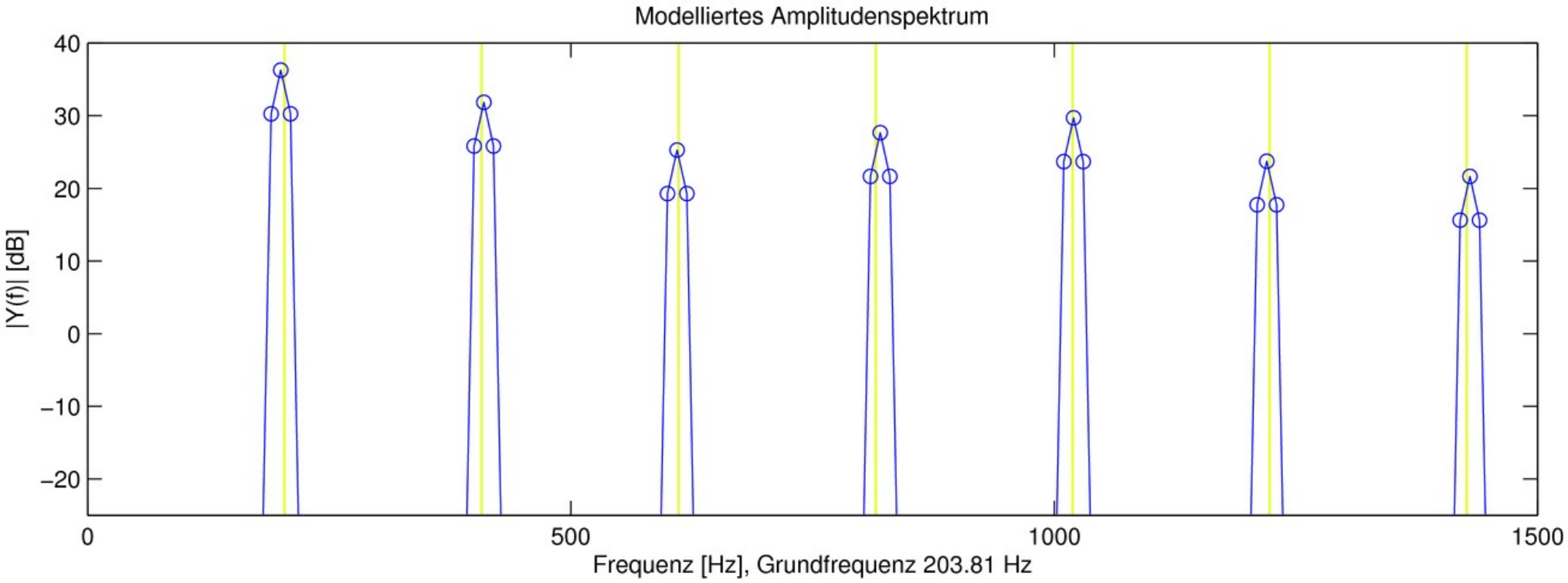


Verarbeitetes gestörtes Signal[1] mit längeren Blöcken

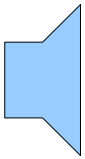
Verarbeitet
mit 100 ms Blocklänge

Verarbeitet
mit 40ms Blocklänge

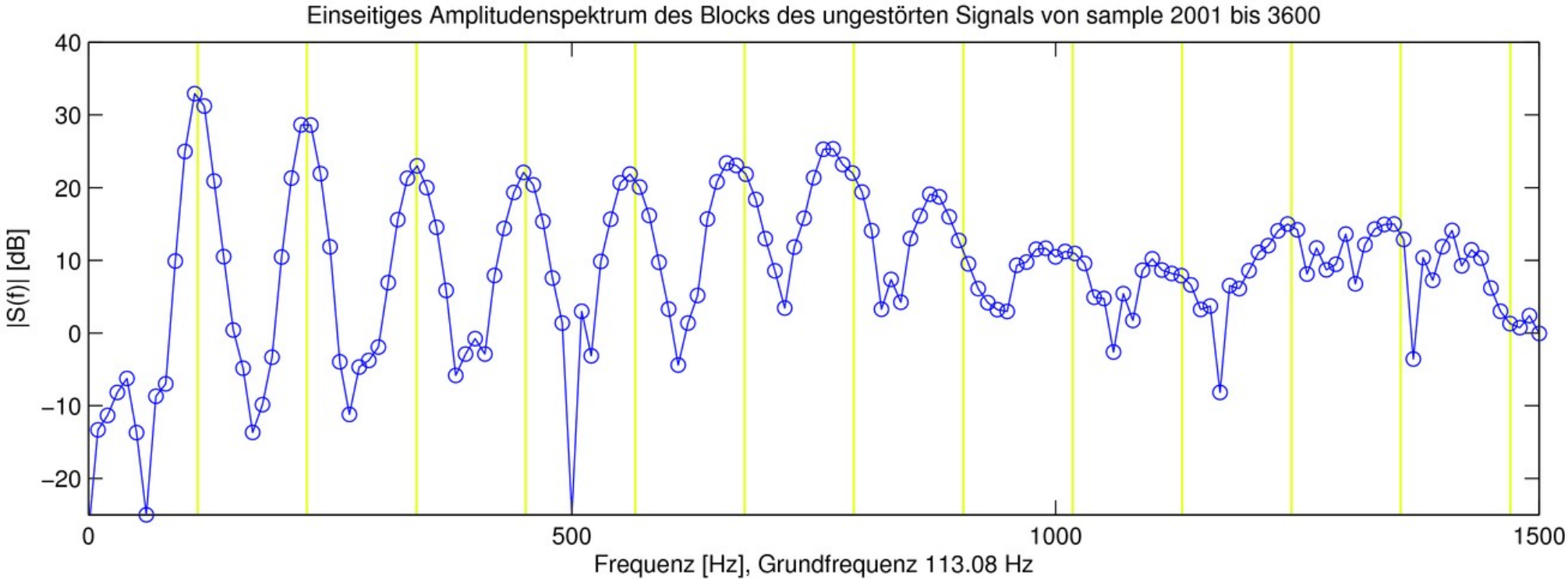
gestört



Beispiel 3: Sprachsignal^2 eines männlichen Sprechers



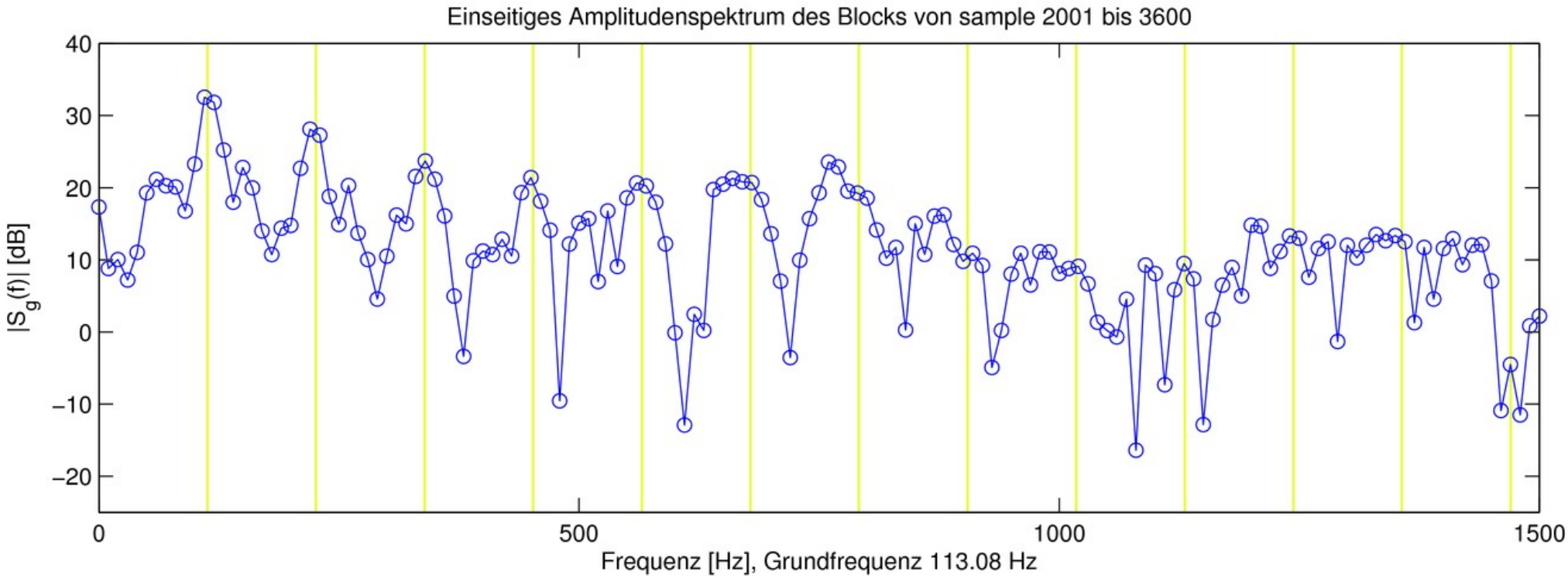
- Blocklänge für DFT: 100ms



Spektrum des stimmhaften Sprachsignals[2], gestört

$$H_1(z) = \frac{1}{1 - 0.9z^{-1}} \quad H_2(z) = \frac{1}{1 - 0.6z^{-1}}$$

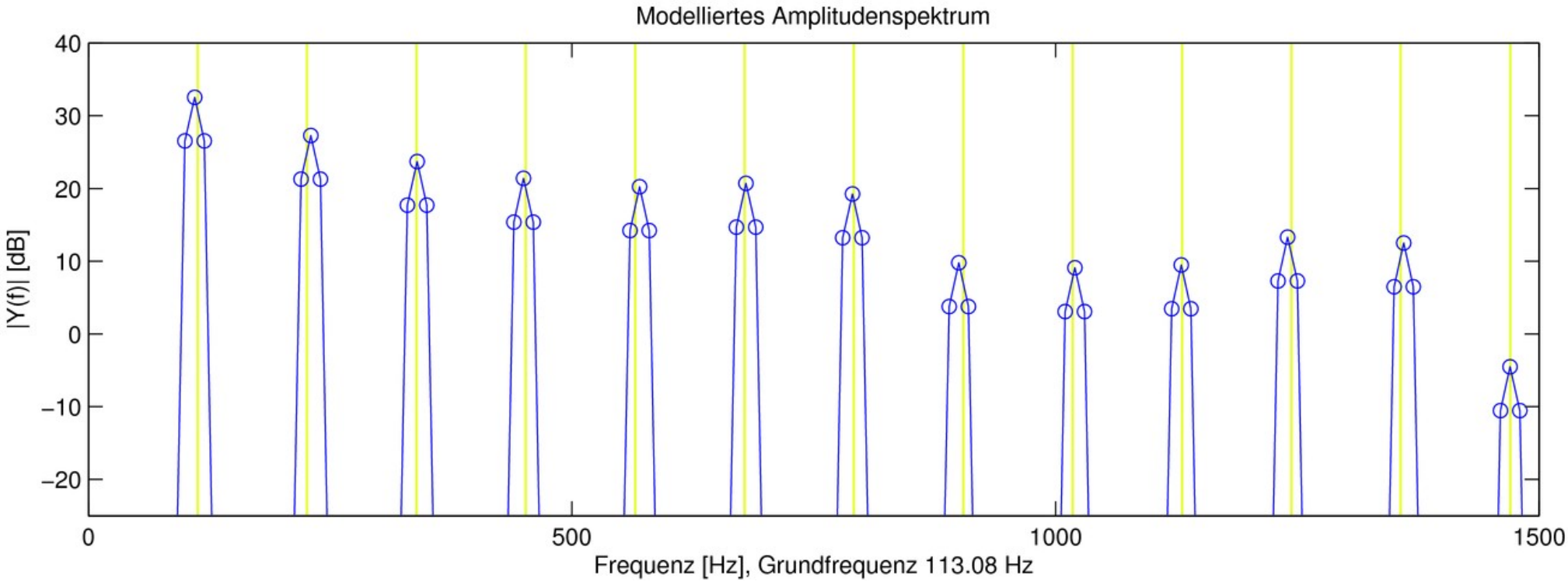
$$H(z) = H_1(z) \cdot H_2(z)$$



Spektrum des verarbeiteten gestörten Sprachsignals[2]

verarbeitet

gestört

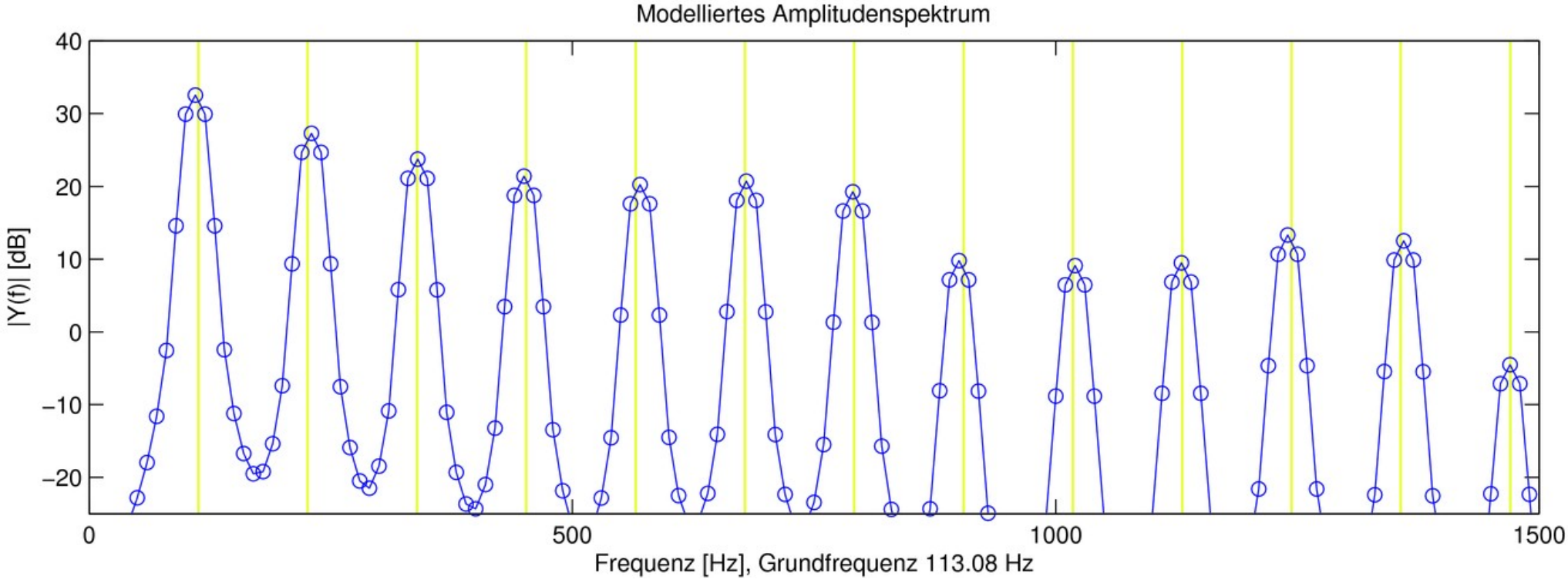


Auswirkungen des Fenster-Spektrums[2]

Modelliertes Amplitudenspektrum mit Fenster-Spektrum bestimmt per

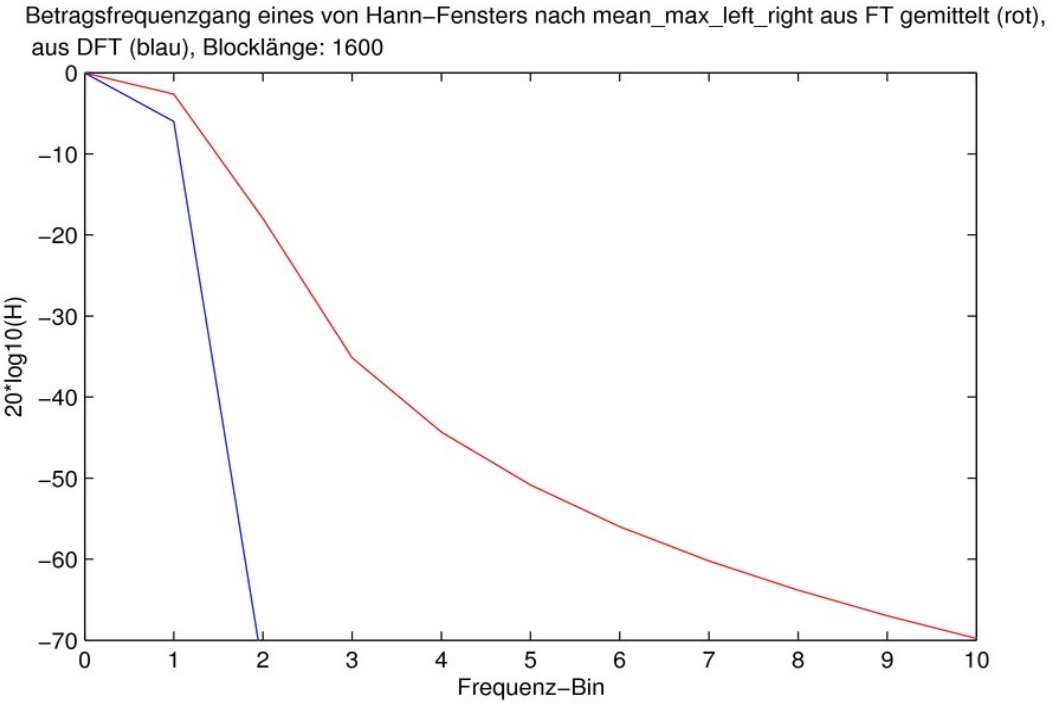
dft

mean_max_left_right



Bestimmung des Fenster-Spektrums

- Per DFT:
  $W(\omega)=DFT\{w(t)\}$
- Per 10fach enger abgetasteter DTFT: und Verwendung des Mittelwertes der Maxima nach links und nach rechts jeweils innerhalb eines Frequenzintervalls der Breite des halben Abstands zweier DFT-Frequenzen (mean_max_left_right)



Auswirkungen des Fenster-Spektrums bei Sprachsignal[1]

Modelliertes Amplitudenspektrum mit Fenster-Spektrum bestimmt per

dft

mean_max_left_right