Eidgenössische Technische Hochschule Zürich
Swiss Federal Institute of Technology Zurich

Sprachverarbeitung I / 4   HS 2016

Lineare Prädiktion

Buch: Kapitel 4.5

Beat Pfister

Sprachverarbeitung I / 4

Vorlesung: Lineare Prädiktion

- Einführung in die lineare Prädiktion
- Einsatz der LP in der Sprachverarbeitung
- Interpretation des LP-Ansatzes

Übung:

- Berechnung des Prädiktors
- spektrale Eigenschaften des Prädiktors

Lineare Prädiktion

Tatsache: Aufeinanderfolgende Abtastwerte von Sprachsignalen sind nicht statistisch unabhängig

Grundidee: Vorhersage für $s(n)$ aus $s(n-1),\ s(n-2), \ldots$

$\longrightarrow\ \tilde{s}(n)$ >>>

linearer Prädiktor: $$\tilde{s}(n) = -\sum_{k=1}^{K} a_k\, s(n-k)$$

Frage: Wie werden die Prädiktorkoeffizienten bestimmt?

Bestimmen der Prädiktorkoeffizienten

Aufgabe: Prädiktion von N Abtastwerten eines Sprachsignals mit

$$\tilde{s}(n) = -\sum_{k=1}^{K} a_k\, s(n-k)$$ >>>

(Gleichungssystem mit N Gleichungen und K Unbekannten)

Lösung für $N = K$: möglich, jedoch uninteressant

Lösung für $N \gg K$: Gleichungssystem überbestimmt

$\longrightarrow$ fehlerfreie Prädiktion i.a. nicht möglich

$\longrightarrow$ Optimierung so, dass Prädiktionsfehler minimal

Ermitteln der Prädiktorkoeffizienten

Optimierung: Bestimmung der Prädiktorkoeffizienten so, dass die Energie des Fehlersignals minimal

linearer Prädiktor: $\tilde{s}(n) = -\sum_{k=1}^{K} a_k\, s(n-k)$

Prädiktionsfehler: $e(n) = s(n) - \tilde{s}(n) = s(n) + \sum_{k=1}^{K} a_k\, s(n-k)$

Optimierung der a_k damit Fehlerenergie minimal:

$$E = \sum_n e^2(n) = \sum_n \left[s(n) + \sum_{k=1}^{K} a_k\, s(n-k) \right]^2 \stackrel{!}{=} \text{min}$$

Ermitteln der Prädiktorkoeffizienten (2)

Für Minimum von E gilt:

$$\frac{\partial E}{\partial a_i} = 2 \sum_n \left[s(n) + \sum_{k=1}^{K} a_k \, s(n-k) \right] s(n-i) \stackrel{!}{=} 0 \qquad 1 \leq i \leq K$$

(System aus K Gleichungen mit K Unbekannten)

$\longrightarrow$ Normalgleichungen:

$$\sum_{k=1}^{K} a_k \sum_n s(n-k) \, s(n-i) = -\sum_n s(n) \, s(n-i) \qquad 1 \leq i \leq K$$

Wofür kann der lineare Prädiktor gebraucht werden?

Eigenschaften des linearen Prädiktors

Prädiktionsfehler im Zeitbereich: (Differenzengleichung)

$$e(n) = s(n) - \tilde{s}(n) = s(n) + \sum_{k=1}^{K} a_k \, s(n-k) = \sum_{k=0}^{K} a_k \, s(n-k)$$

mit $a_0 = 1$

Abtastwert $e(n)$ des Ausgangssignals ist gewichtete Summe aus den Abtastwerten $s(n-k)$ des Eingangssignals, mit $k = 0 \ldots K$.

$\longrightarrow$ Die a_k sind die Koeffizienten eines Transversalfilters $A(z)$

Frage: Wie sieht die Übertragungsfunktion dieses Filters aus? >>>

Vom Prädiktor abgeleitete Filter

Prädiktionsfehler: $e(n) = \sum_{k=0}^{K} a_k \, s(n-k)$:

z-Transformation: $E(z) = \sum_{k=0}^{K} a_k \, z^{-k} S(z) = A(z)\, S(z)$

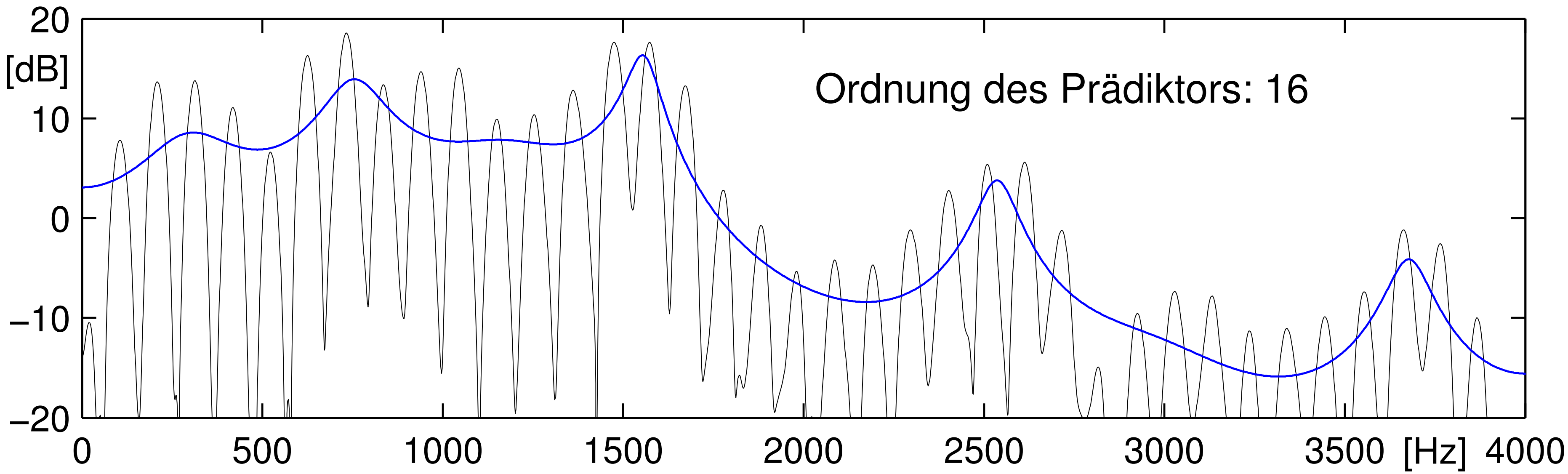
$\longrightarrow$ $S(z)$ gefiltert mit $A(z) = 1 + a_1 z^{-1} + \ldots + a_K \, z^{-K}$ ergibt $E(z)$

und daraus folgt: $S(z) = \dfrac{1}{A(z)} E(z) = H(z)\, E(z)$

$\longrightarrow$ $E(z)$ gefiltert mit $H(z) = 1/A(z)$ ergibt $S(z)$

$H(z)$ heisst Synthesefilter $A(z)$ heisst inverses Filter >>>

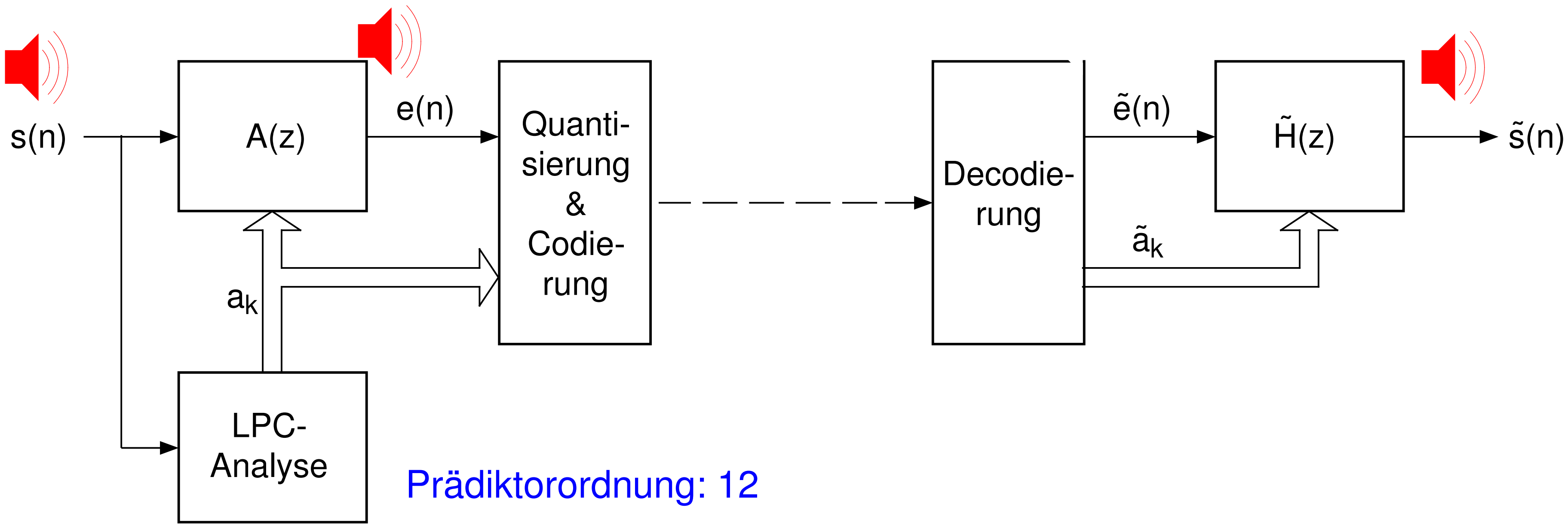
Approximation des Signalspektrums durch $H(z)$



Verwendung:
- Beschreibung des groben Verlaufs des Spektrums >>>
- Schätzung der Formanten (Position, Höhe und Güte)
- effiziente Sprachübertragung (Sprachmodellierung)

Sprachmodellierung mittels linearer Prädiktion

Zielsetzung: Effiziente Übertragung von Sprachsignalen



Prädiktorordnung: 12

zu übertragende Daten:

- Fehlersignal $e(n)$ und
- Prädiktorkoeffizienten a_k pro Analyseabschnitt

Sprachmodellierung mittels linearer Prädiktion

LPC-Modellierung lohnt sich dann, wenn $e(n)$ und a_k
zusammen eine kleinere Datenmenge ergeben als $s(n)$ allein.

Dies ist möglich, weil $e(n)$ spezielle Eigenschaften hat: >>>

- ebene spektrale Enveloppe
- periodisch oder rauschartig

Als grobe Beschreibung für $e(n)$ sind also zu übertragen:

- ⋆ die Periode T_0 bzw. die Grundfrequenz F_0 >>>
- ⋆ die Signalstärke G (die Quadratwurzel der Leistung)

Grobe Approximation des Prädiktionsfehlers

$$\tilde{e}(n) = G\,u(n)$$

wobei

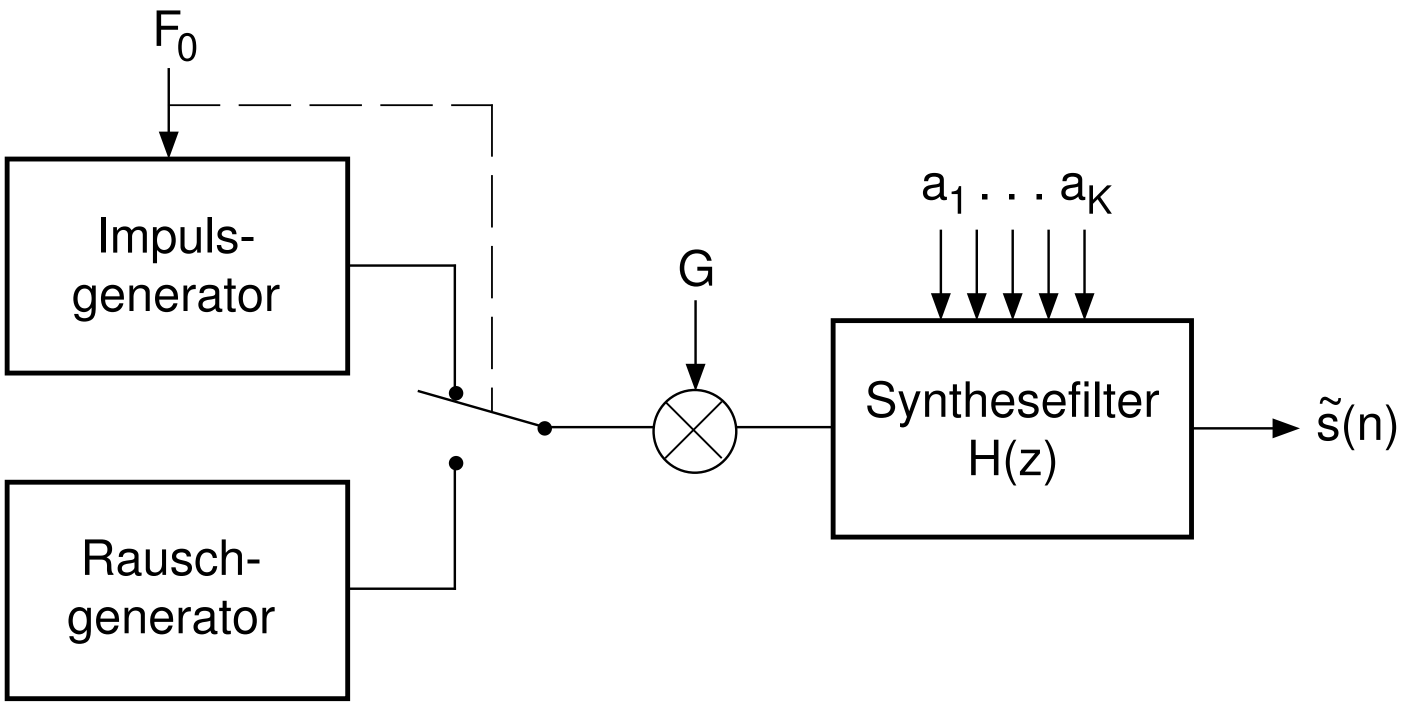
$$u(n) = \begin{cases} \sqrt{T_0/T_s}\sum\limits_m \delta(n - mT_0/T_s) & \text{falls } e(n) \text{ periodisch mit } T_0/T_s \\ \mathcal{N}_0 & \text{sonst (weisses Rauschen).} \end{cases}$$

$$G = \sqrt{\frac{1}{N}\sum_n e^2(n)}$$

$\longrightarrow$ nur 2 Werte pro Analyseabschnitt von $e(n)$ sind zu übertragen!

Signalrekonstruktion aus den LPC-Parametern

$G,\, F_0,\, a_k$



>>>

Syntheselänge und F_0 unabhängig variierbar!

>>>

Analogie zum menschlichen Sprechapparat

Feststellung: LPC-Sprachproduktions-Modell unterscheidet auch
- Signalproduktion und
- Klang- bzw. spektrale Formung

Frage: Sind somit aus den a_k Rückschlüsse auf die Stellung der Artikulatoren möglich? >>>

Antwort: Ja, wenn LPC-Analyse aus Signal $s'(n)$

Präemphase: $S'(z) = P(z)\,S(z)$ wobei $P(z) = 1 - 0.98z^{-1}$

$H(z) \longrightarrow$ akustisches Filter >>>

Zusammenfassung

Lineare Prädiktion: Voraussage des aktuellen Wertes aus vorangehenden

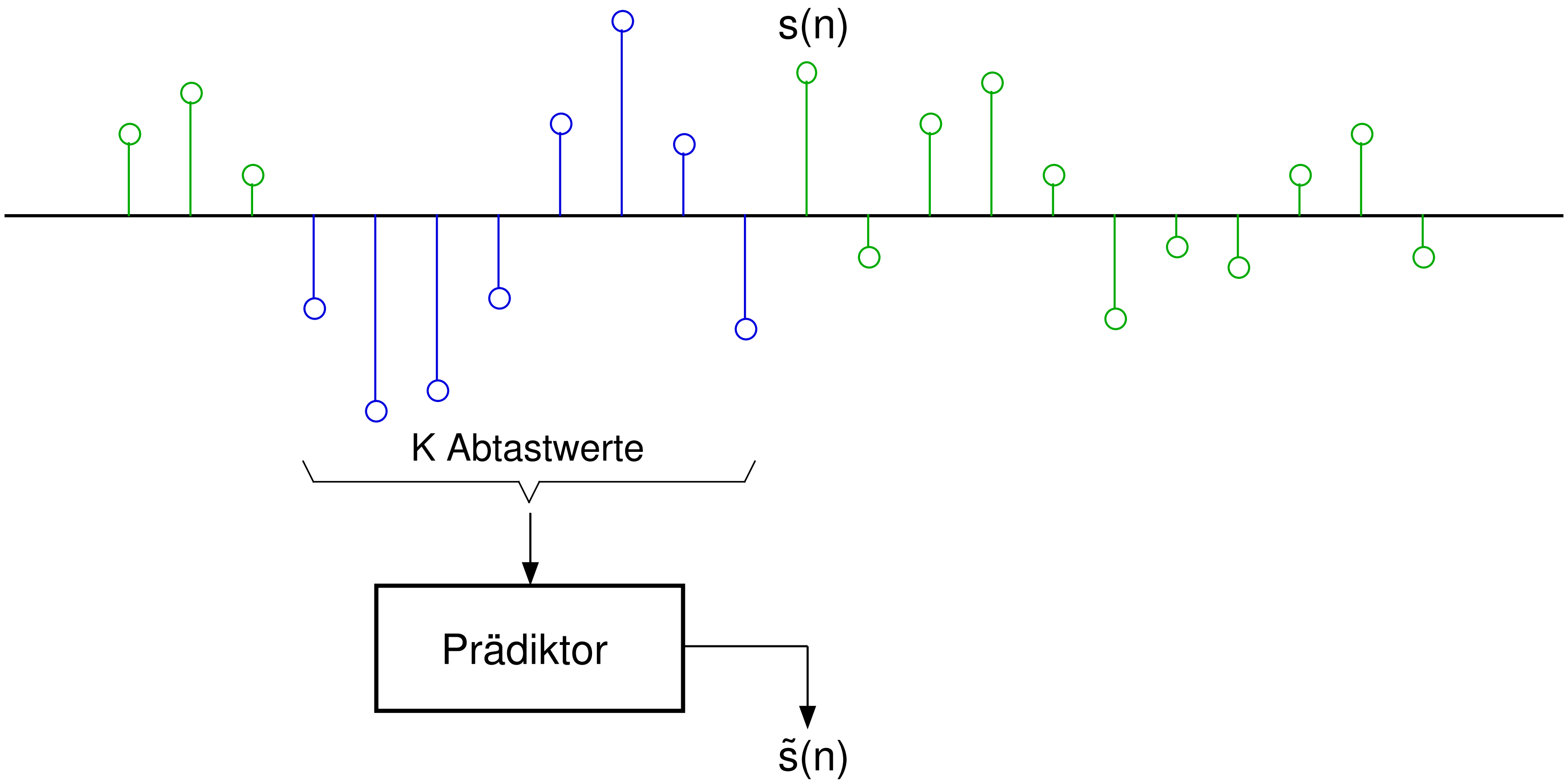
- Prädiktor approximiert Enveloppe des Spektrums
  - höhere Ordnung $\longrightarrow$ bessere Approximation
  - LPC-Analyse ist eine Kurzzeitanalyse
- Anwendungen:
  - Effiziente Sprachübertragung (Sprachmodellierung)
  - Trennnung von Anregung und Klangformung
  - Veränderung von Dauer und/oder Grundfrequenz

Thema der nächsten Lektion:

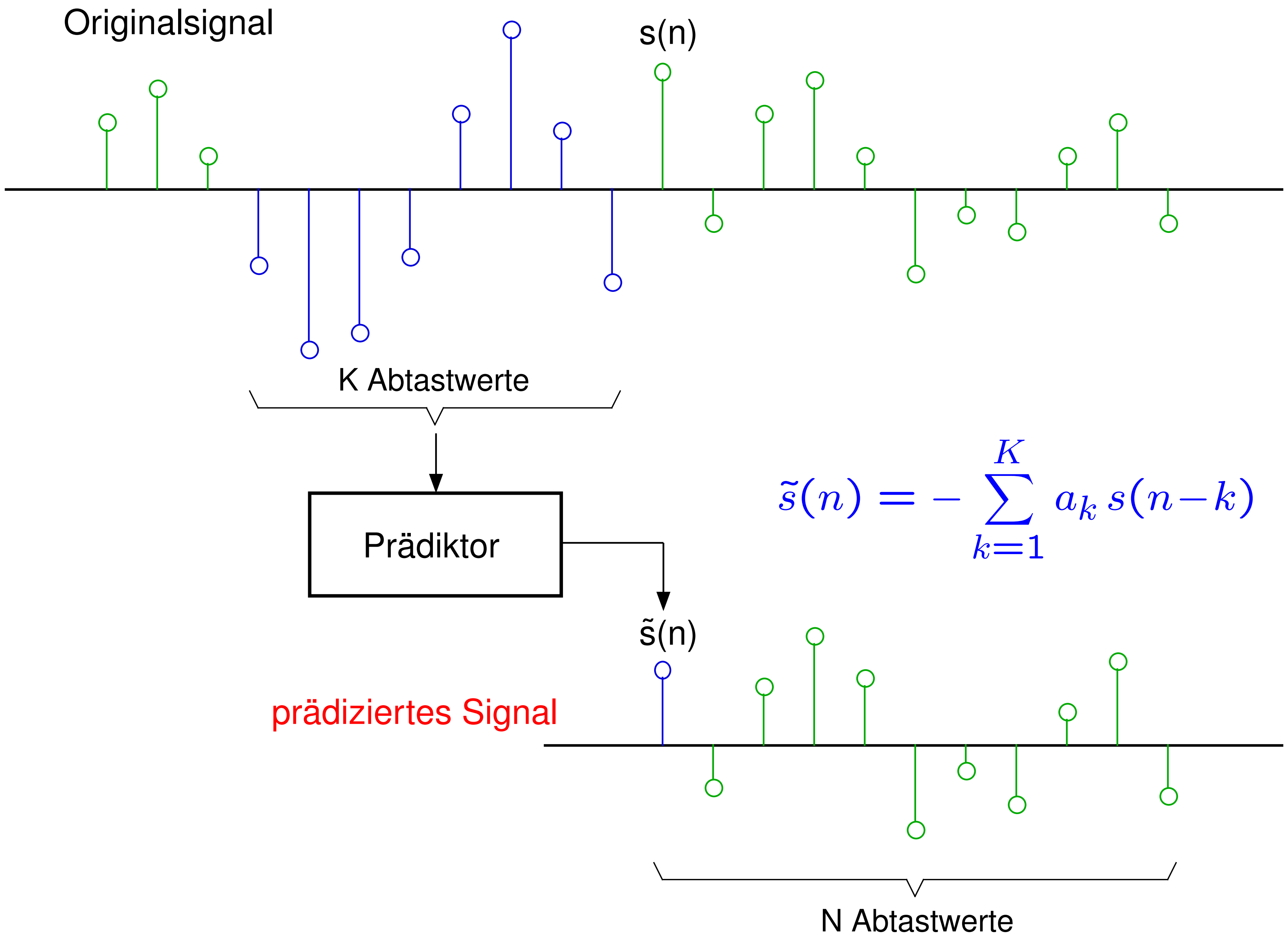
Homomorphe Analyse

Zur Übersicht der Vorlesung *Sprachverarbeitung I* >>>

Idee der Prädiktion

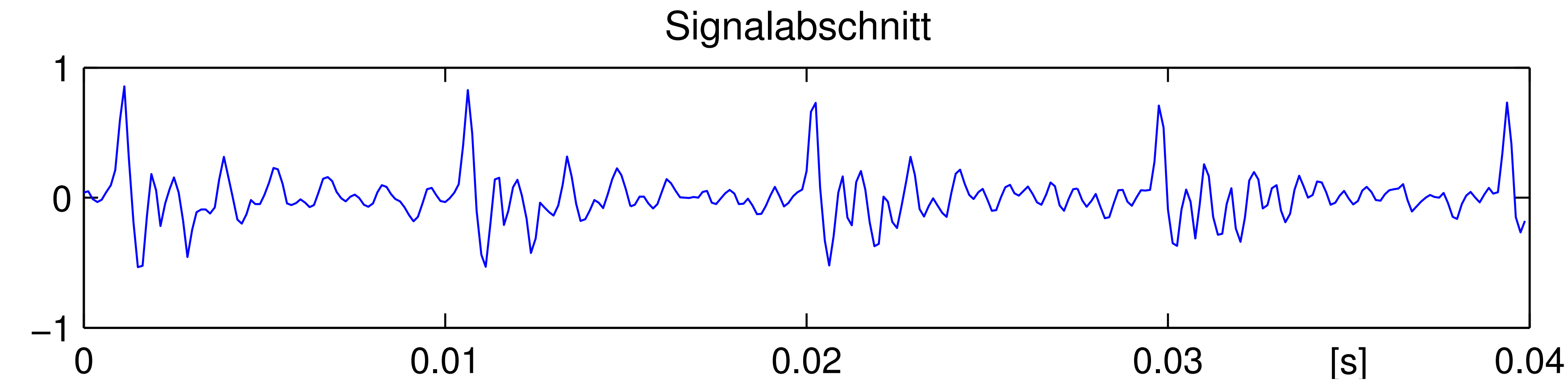


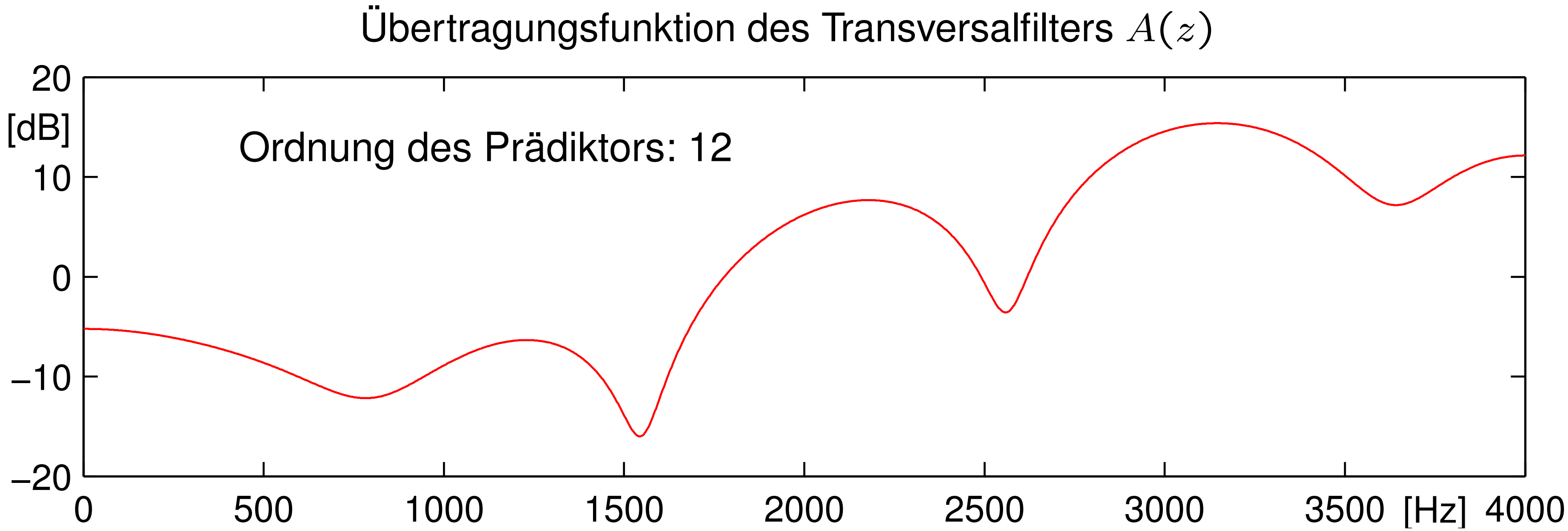
<<<



$$\tilde{s}(n) = -\sum_{k=1}^{K} a_k \, s(n-k)$$

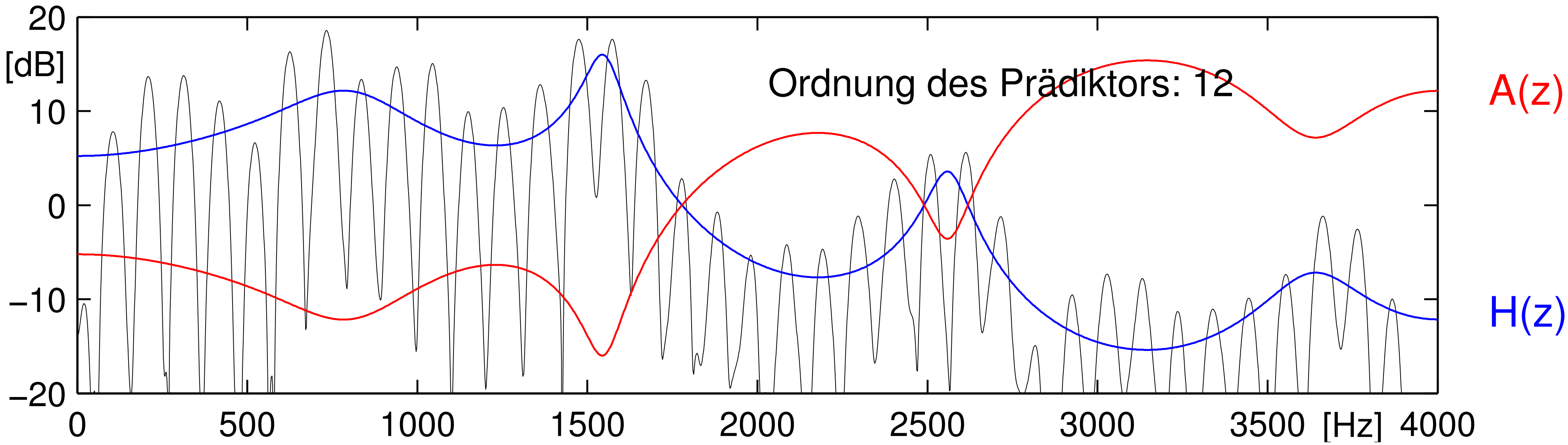
<<<


Signalabschnitt
1
0
–1
0
0.01
0.02
0.03
[s]
0.04


Übertragungsfunktion des Transversalfilters $A(z)$
20
[dB]
10
0
–10
–20
Ordnung des Prädiktors: 12
0
500
1000
1500
2000
2500
3000
3500
[Hz]
4000

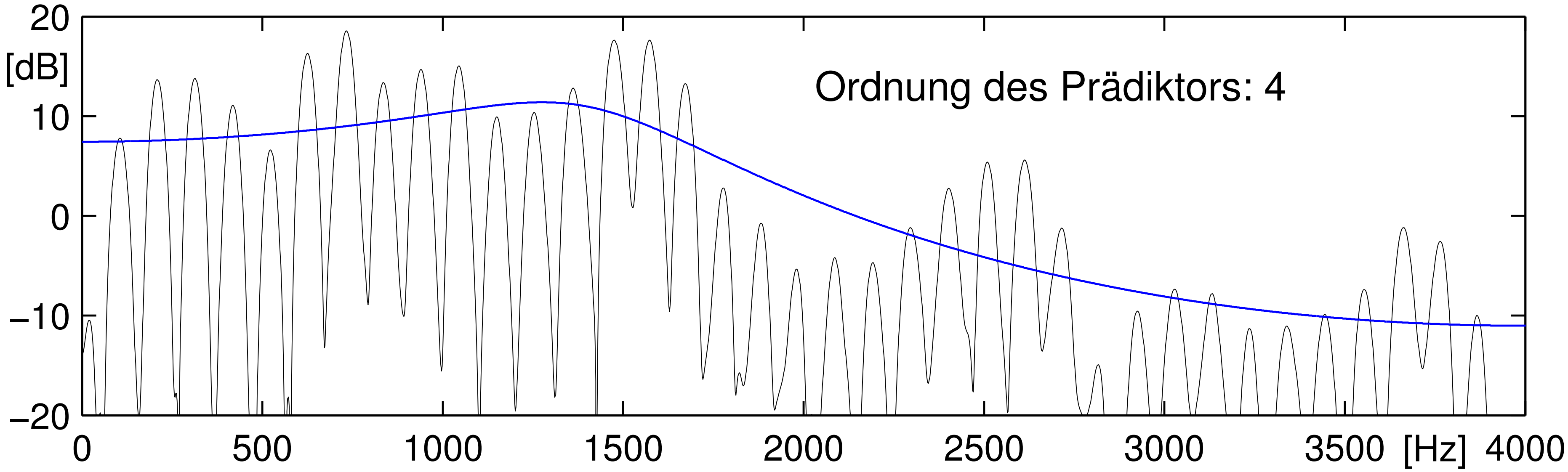
<<<

Übertragungsfunktion von $H(z)$ und $A(z)$

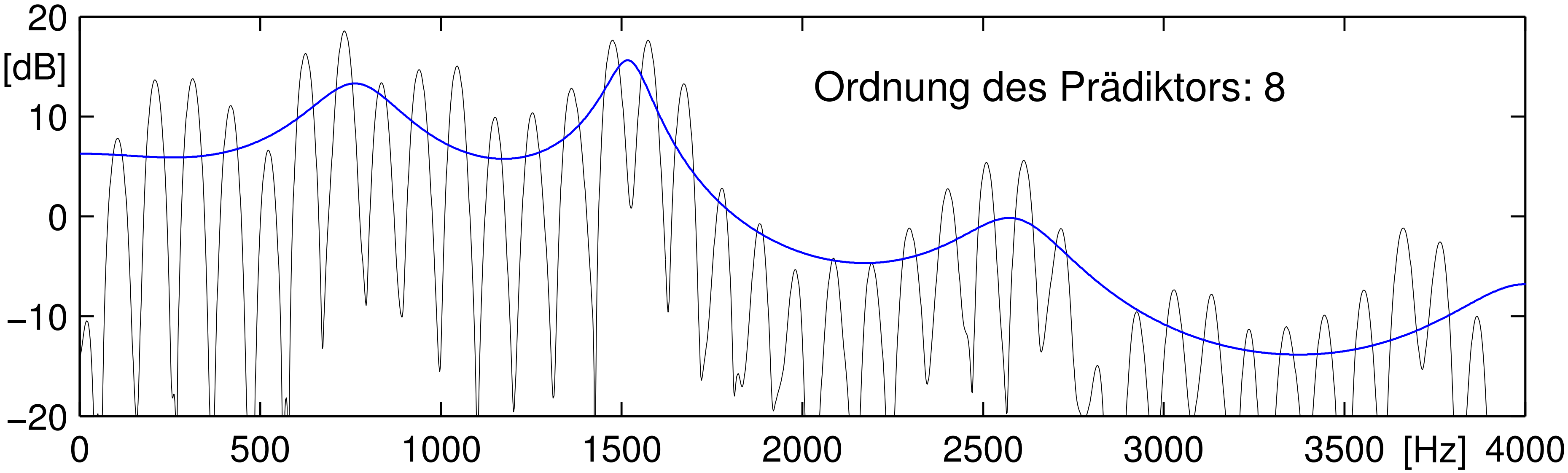


H(z) beschreibt den groben Verlauf des Spektrums (≈ Enveloppe)

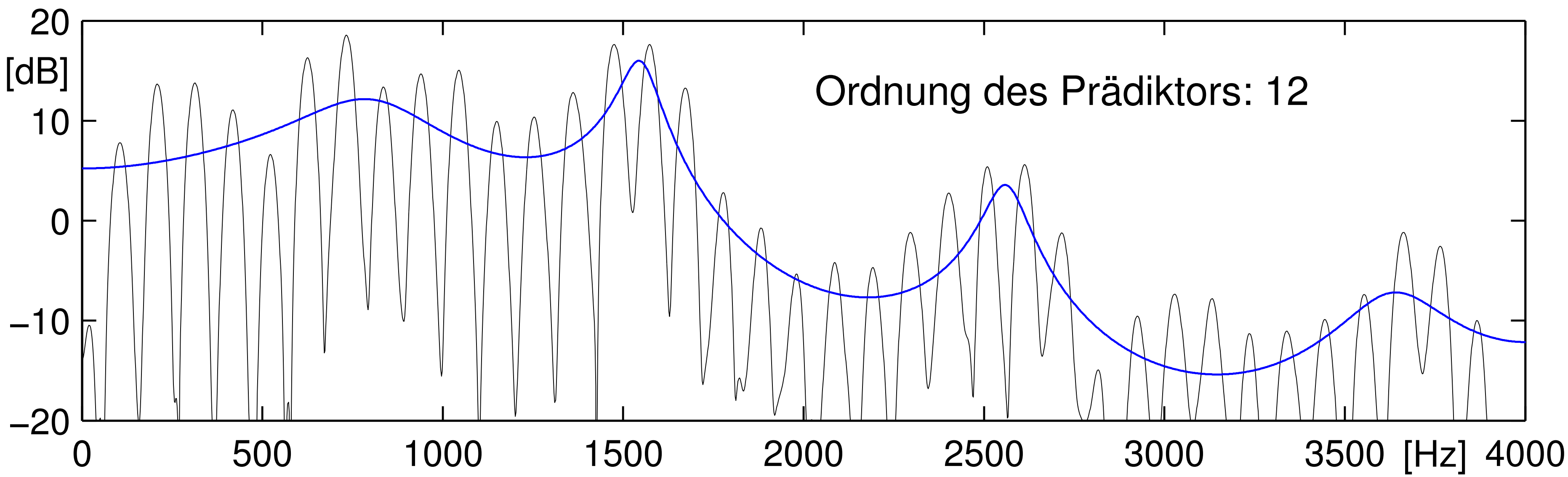
<<<


20
[dB]
10
0
−10
−20
Ordnung des Prädiktors: 4
0
500
1000
1500
2000
2500
3000
3500
[Hz]
4000

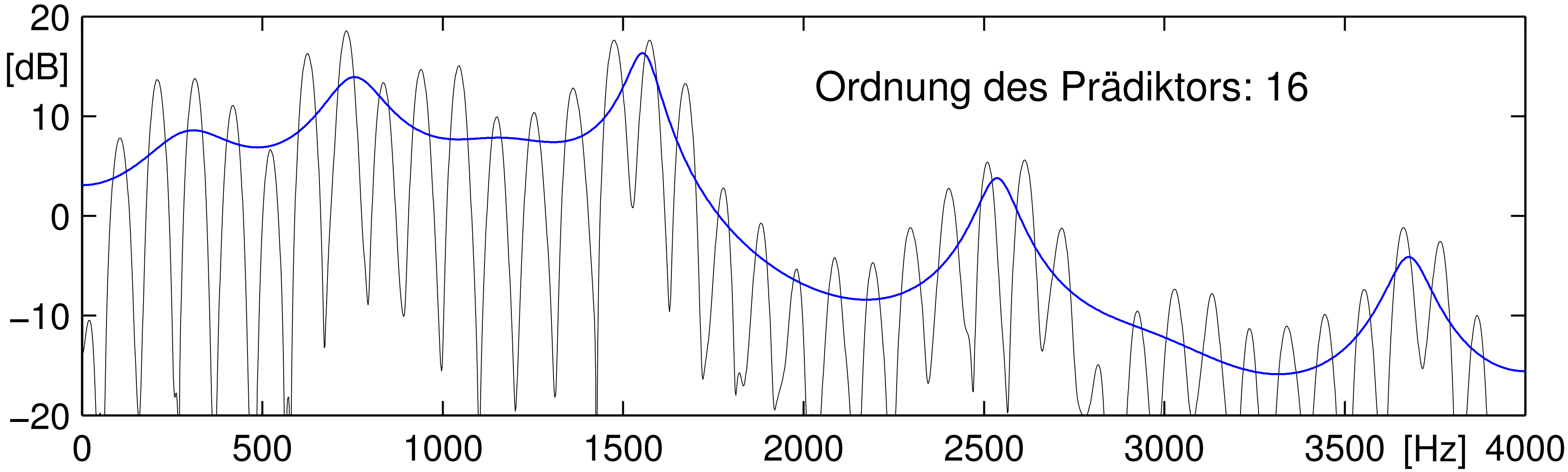
<<<


20
[dB]
10
0
−10
−20
0
500
1000
1500
2000
2500
3000
3500
[Hz]
4000
Ordnung des Prädiktors: 8

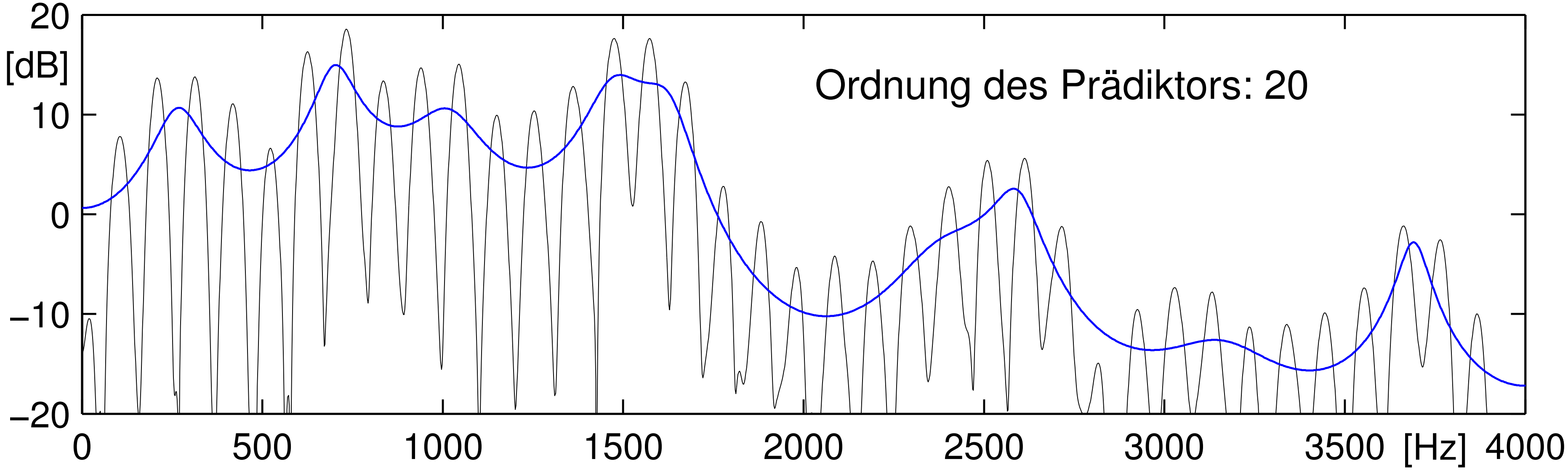
<<<


20
[dB]
10
0
–10
–20
0
500
1000
1500
2000
2500
3000
3500
[Hz]
4000
Ordnung des Prädiktors: 12

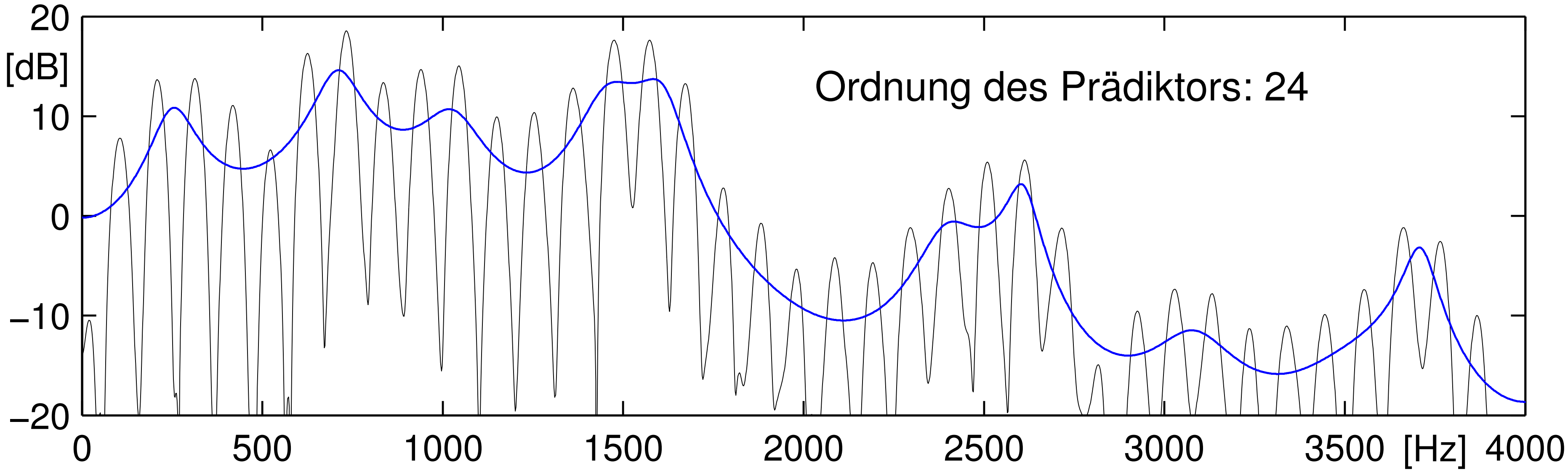
<<<


20
[dB]
10
0
−10
−20
0
500
1000
1500
2000
2500
3000
3500
[Hz]
4000
Ordnung des Prädiktors: 16

<<<


20
[dB]
10
0
−10
−20
0
500
1000
1500
2000
2500
3000
3500
[Hz]
4000
Ordnung des Prädiktors: 20

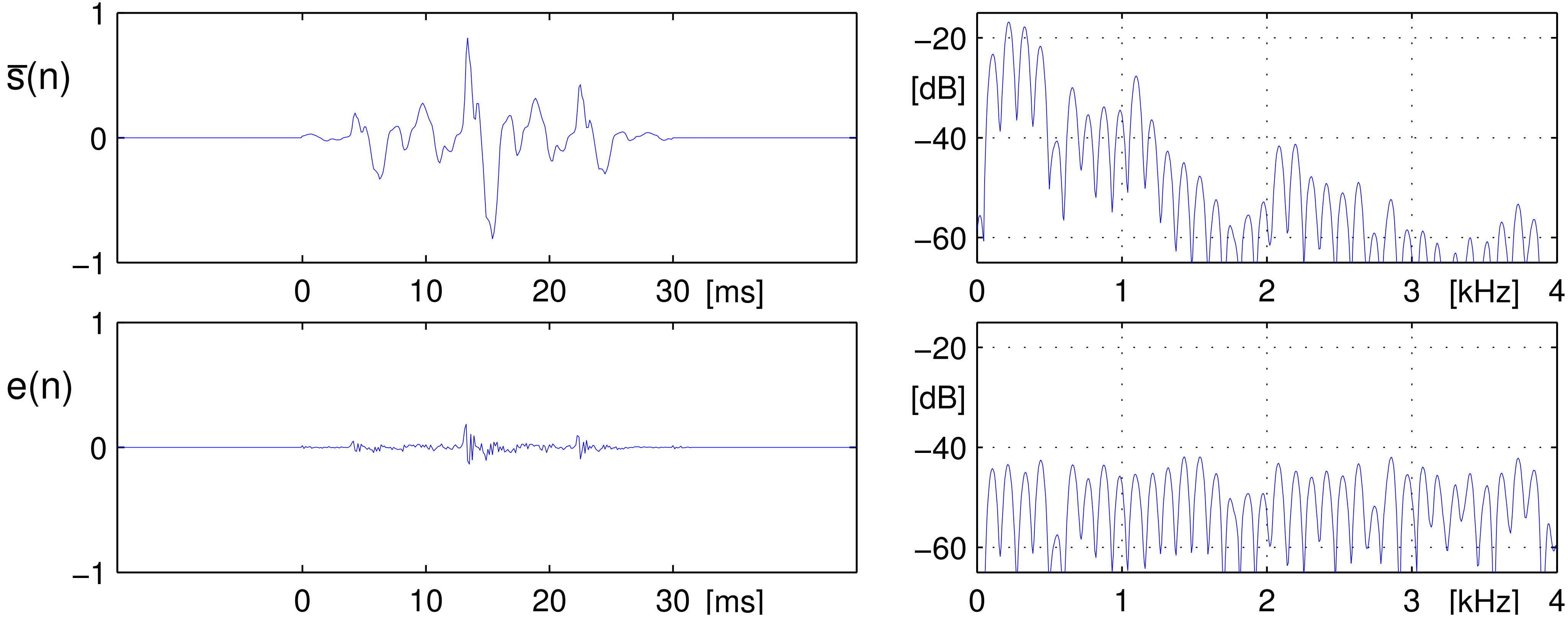
<<<


20
[dB]
10
0
−10
−20
Ordnung des Prädiktors: 24
0
500
1000
1500
2000
2500
3000
3500
[Hz]
4000

<<<

Zusammenhang zwischen Sprachsignal und Prädiktionsfehler

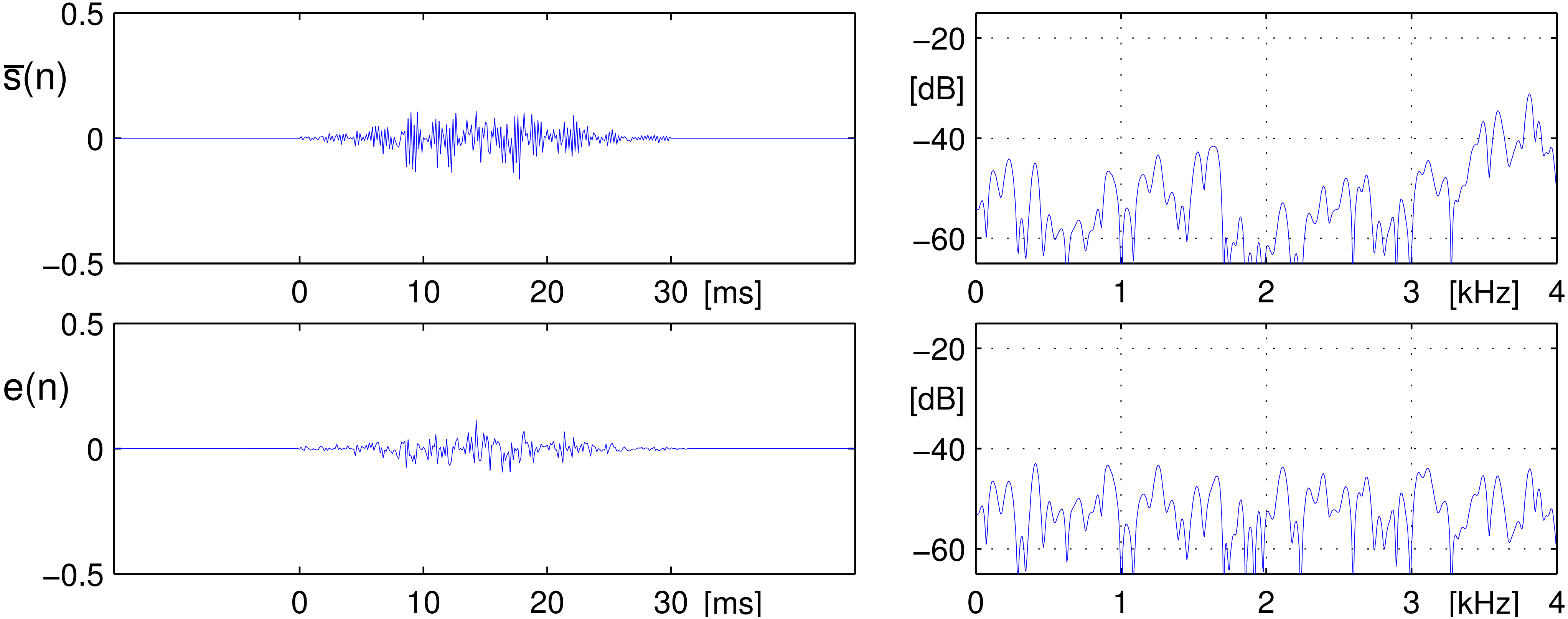
stimmhaftes Sprachsegment / Prädiktor 12. Ordnung



>>

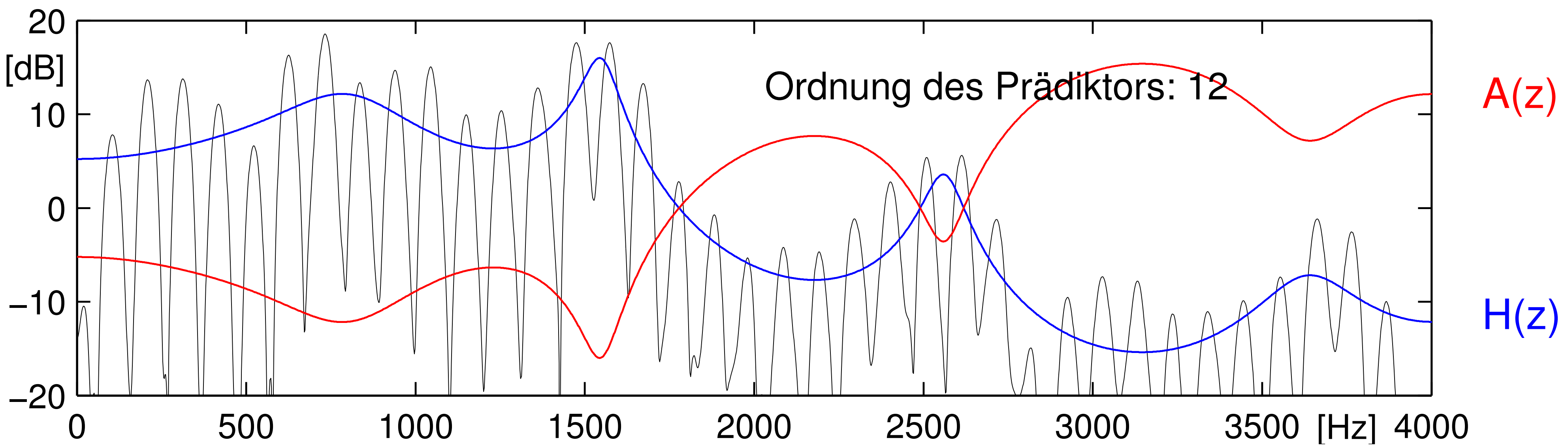
Zusammenhang zwischen Sprachsignal und Prädiktionsfehler

stimmloses Sprachsegment / Prädiktor 12. Ordnung



<<<

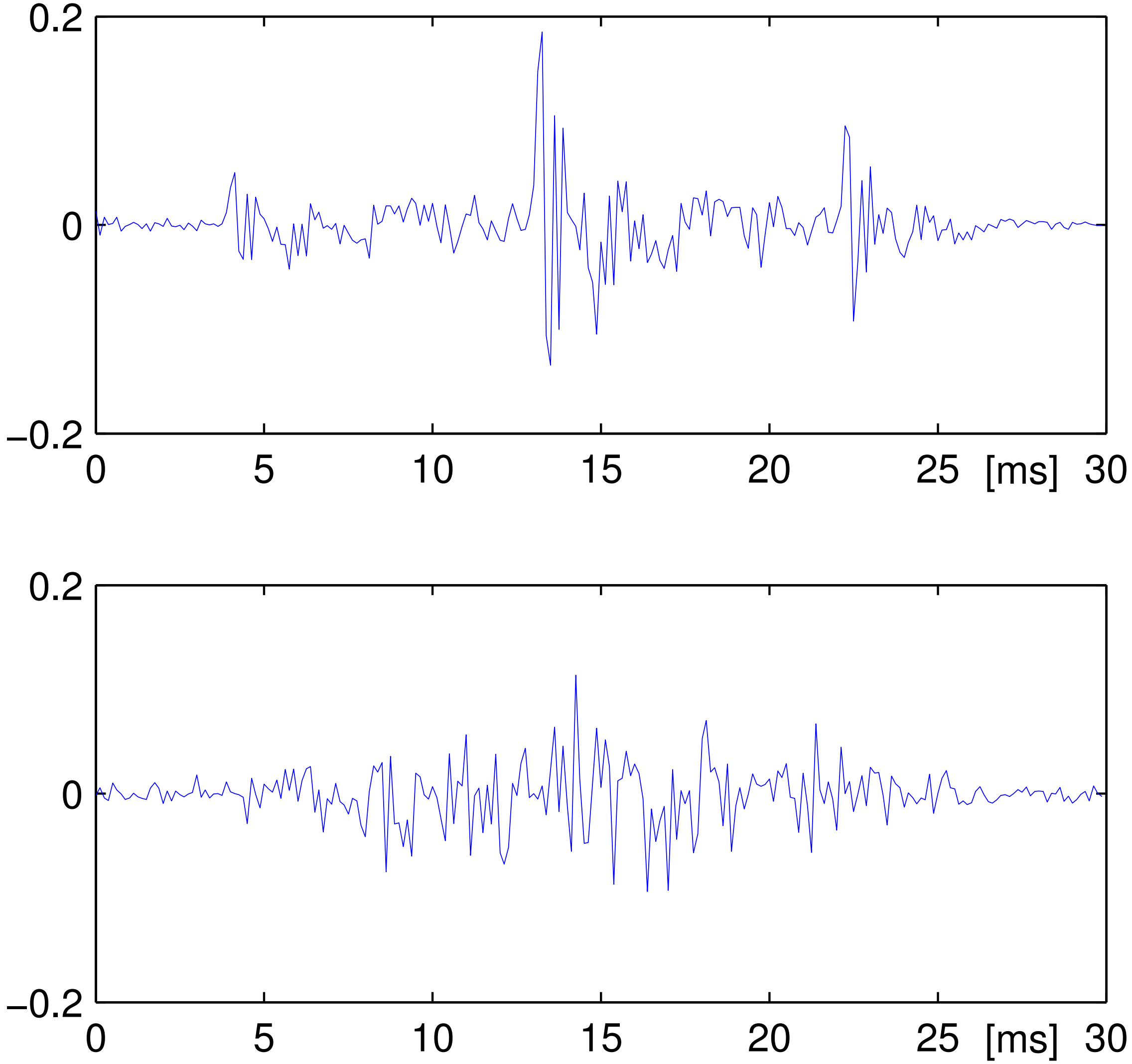
Übertragungsfunktion von $H(z)$ und $A(z)$



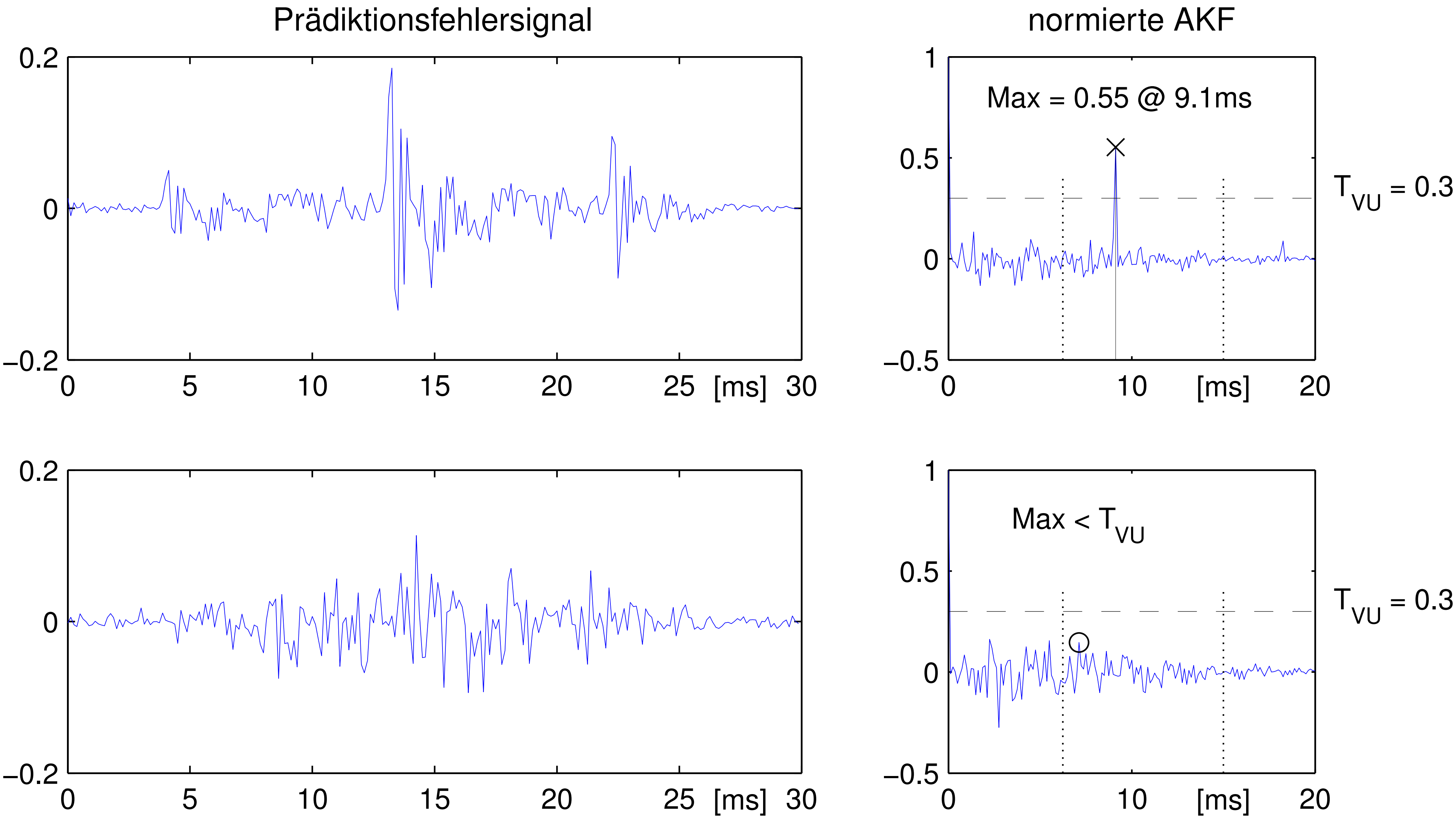
H(z) beschreibt den groben Verlauf des Spektrums (≈ Enveloppe)

<<<

Prädiktionsfehlersignal

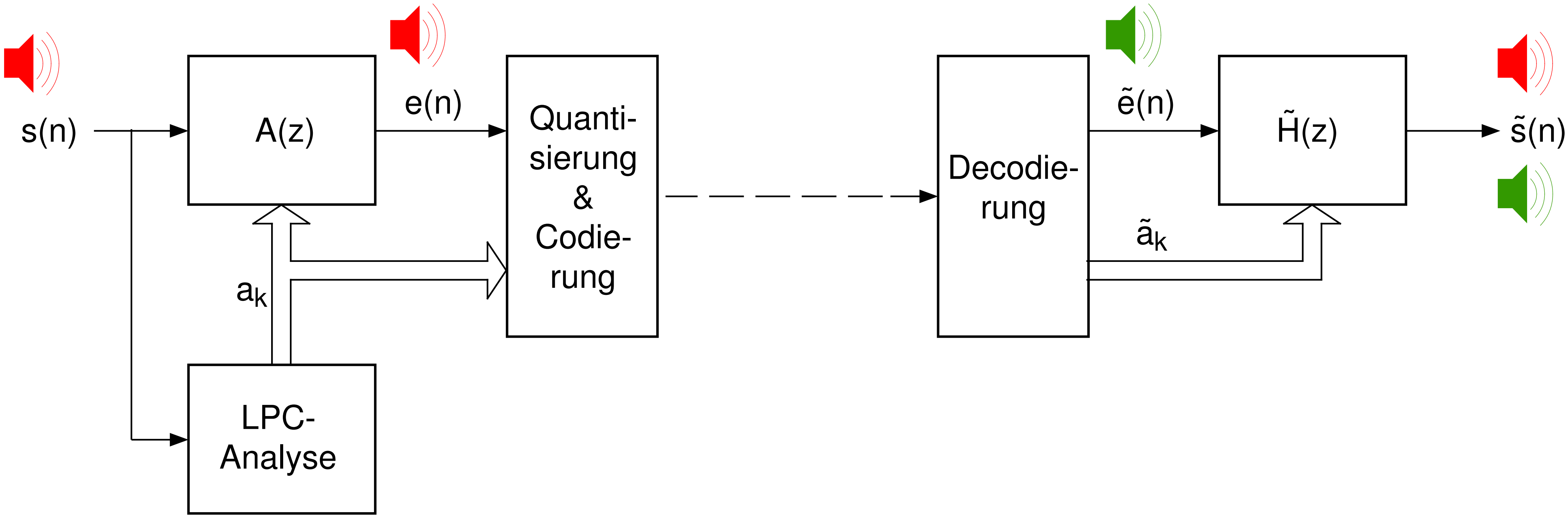


Wie lässt sich die Periode bestimmen?


Prädiktionsfehlersignal
normierte AKF
Max = 0.55 @ 9.1ms
$T_{VU} = 0.3$
$\text{Max} < T_{VU}$
$T_{VU} = 0.3$
[ms]
[ms]
[ms]
[ms]

<<<

Effiziente Übertragung von Sprachsignalen mittels LPC_{12}



Datenrate für LPC (zu übertragen $\tilde{e}(n)$ und a_k): 2–6 kBit

$\longrightarrow$ hängt von Analyserate, Prädiktor-Ordnung und Quantisierung ab!

Einfluss der Ordnung des Prädiktors

Signal	Ordnung des Prädiktors			
	2	6	12	18
$e(n)$				
$\tilde{e}(n)$				
$\tilde{s}(n)$				

<<<

Veränderung von Dauer und Grundfrequenz via LPC

		Grundfrequenz		
		70 %	100 %	140 %
	60 %			
Dauer	100 %			
	150 %			

Gegensatz zu Tonband

>>>

<<<

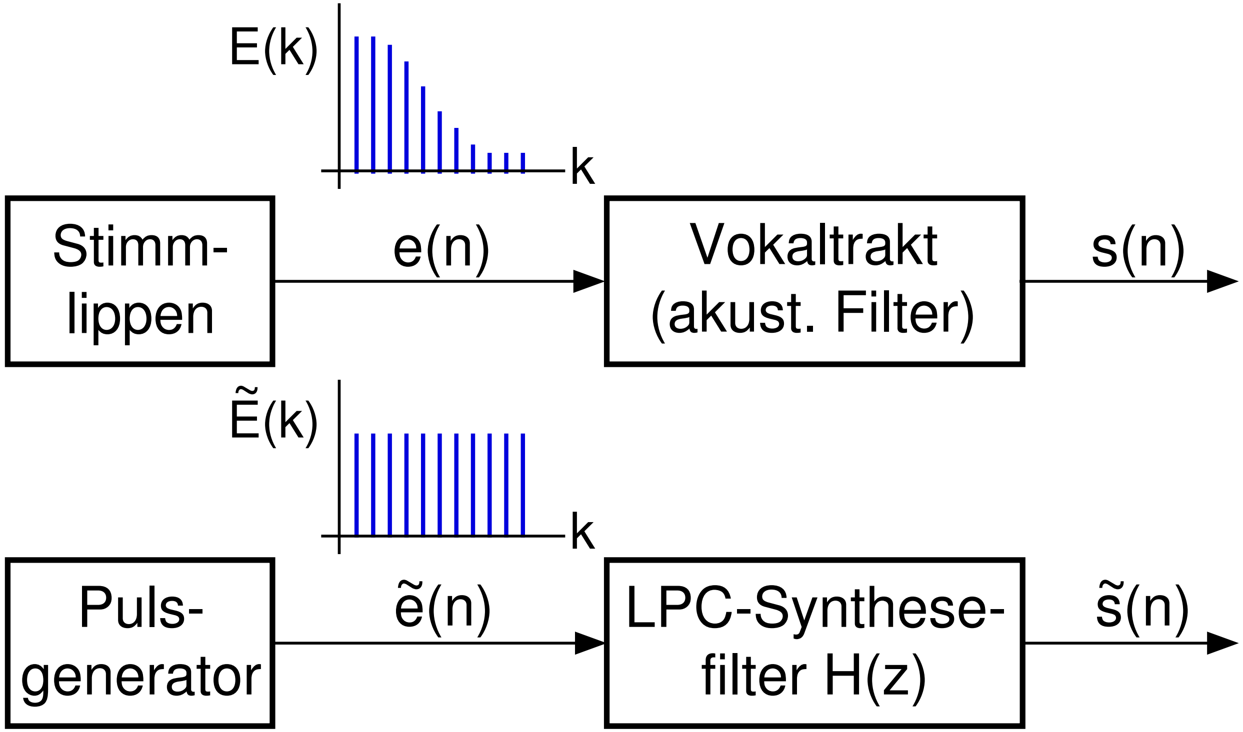
Abspielen einer Tonbandaufnahme

Bandgeschwindigkeit	20 % zu hoch	
	normal	
	20 % zu tief	

<<<

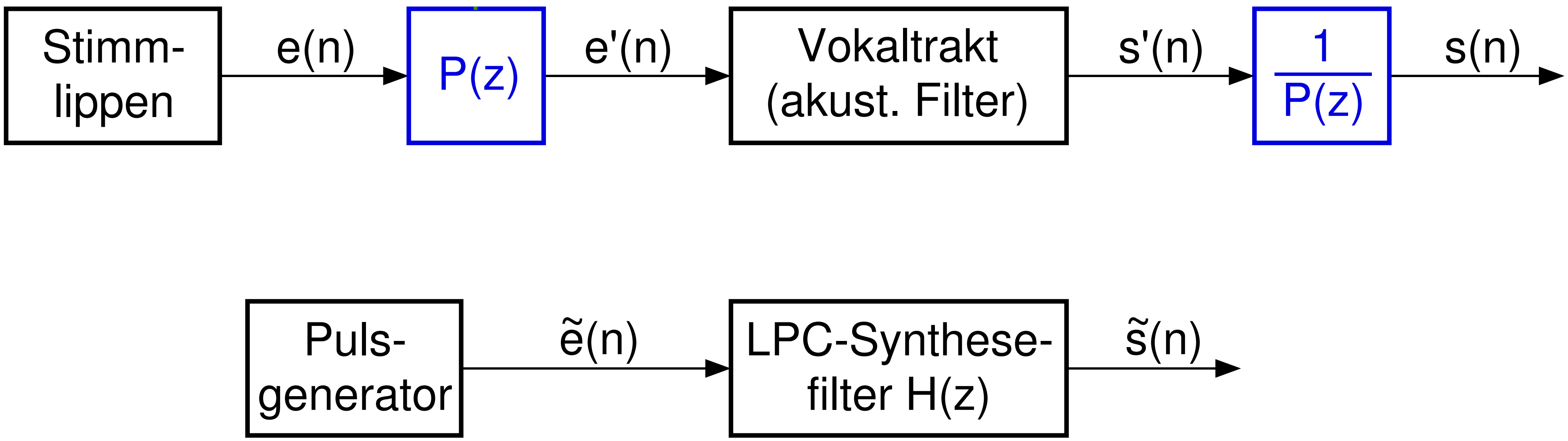
Menschl. Sprechapparat vs. LPC-Sprachproduktions-Modell

nur für stimmhafte Laute



Menschl. Sprechapparat vs. LPC-Sprachproduktions-Modell

nur für stimmhafte Laute

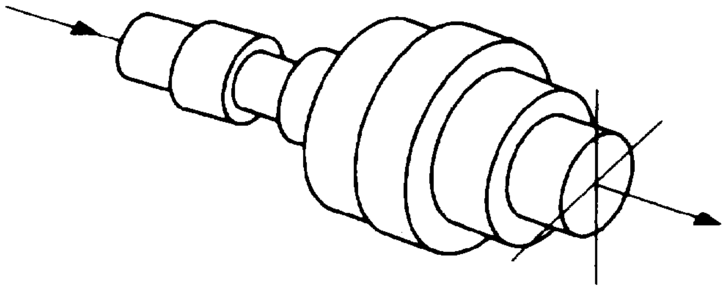


<<<

Hochpassfilter P(z) $\rightarrow$ e'(n) flaches Spektrum

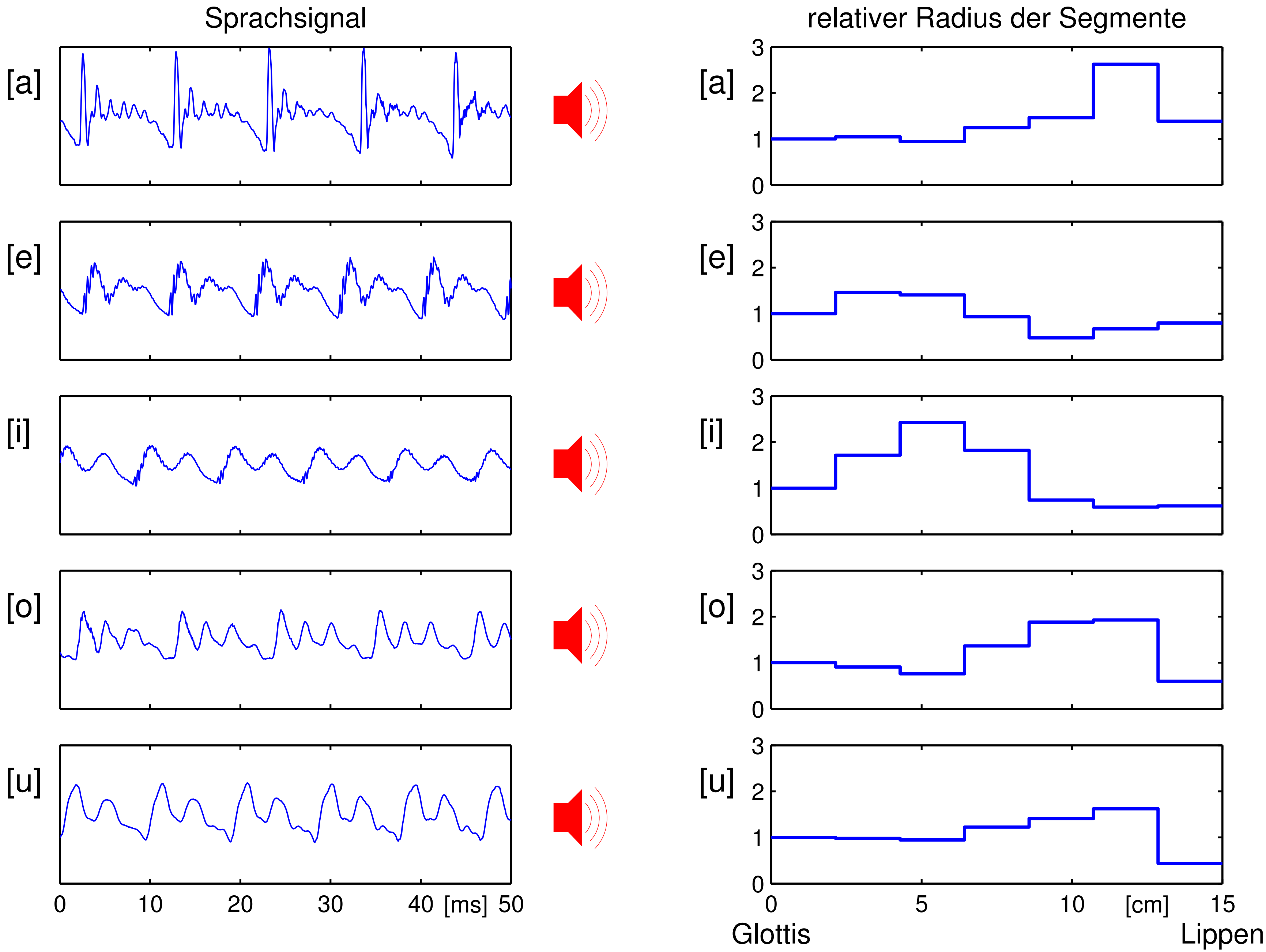
Zu $H(z)$ äquivalentes akustisches Filter

Röhrensegmente



Verhältnis aufeinanderfolgender Röhrenquerschnitte: $\frac{A_{i+1}}{A_i} = \frac{1-k_i}{1+k_i}$

(k_i: Reflexionskoeffzienten aus Durbin-Algorithmus; Buch S. 80)


Sprachsignal
relativer Radius der Segmente
[a]
[e]
[i]
[o]
[u]
0
10
20
30
40
[ms]
50
0
1
2
3
0
5
10
[cm]
15
Glottis
Lippen

<<<