



# Einführung in die Signalverarbeitung

Phonetik und Sprachverarbeitung, 2. Fachsemester,  
Block Sprachtechnologie I

Florian Schiel

Institut für Phonetik und Sprachverarbeitung, LMU München

Signalverarbeitung - Teil 7

# Allgemeines

- Unterrichtssprache ist Deutsch (englische Fachbegriffe in Klammern)
- Fragen am besten sofort; besser einmal zuviel gefragt
- Literatur:
  - Jurafsky D, Martin J H (2000): Speech and Language Processing. Prentice Hall, Kap I.7.
  - Schröder E (1980): Signalverarbeitung
  - Pfister B, Kaufmann T (2008): Sprachverarbeitung - Grundlagen und Methoden der Sprachsynthese und Spracherkennung. Springer-Verlag Berlin Heidelberg.
  - Rabiner, Lawrence R., Schafer R W (1978): Digital Processing of Speech Signals. Prentice-Hall, New Jersey, USA.
  - Hess W (1993): Digitale Filter. Teubner Studienbücher, B.G.Teubner, Stuttgart.
  - Harrington J, Cassidi St (1999): Techniques in Speech Acoustics. Kluwer Academic Publishers, Dordrecht/Boston/London.

# Sprachsignalverarbeitung III

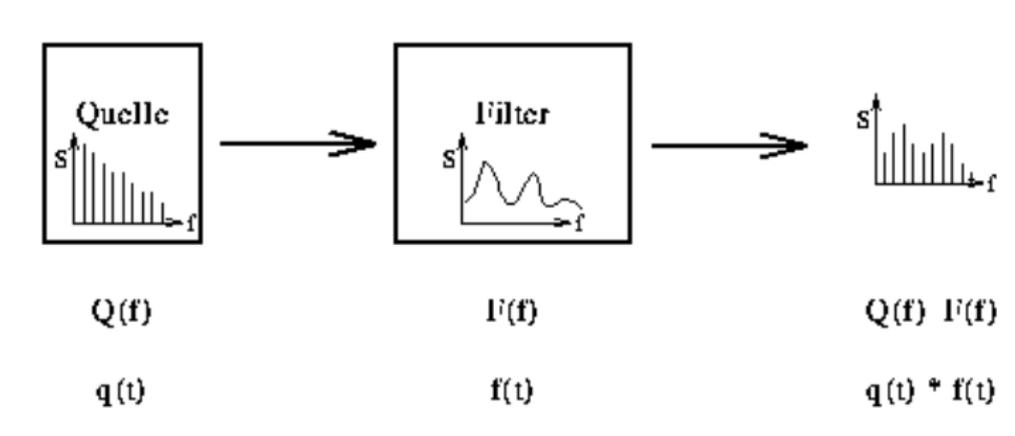
*Anwendung der Methoden auf die Verarbeitung von Sprache.*

- Lineare Prädiktion
- Cepstrum

# Lineare Prädiktion (LP, linear prediction)

## Quelle-Filter-Modell (Fant 1960)

Sprachsignal entsteht durch Signal der Quelle (z.B. Glottis) gefiltert von einem linearen Filter (z.B. Ansatzrohr).



# Lineare Prädiktion (LP, linear prediction)

Ziel der *linearen Prädiktion* (linear prediction):  
Trennung von Quellsignal  $q(t)$  und Filterfunktion  $F(f)$

Für allgemeine Quellensignale nicht möglich!

Annahme: Quellensignal ist ein Impulssignal  
→ Trennung ist möglich durch *lineare Prädiktion*

Idee der linearen Prädiktion:

*Für ein Signalstück (Fenster) suche das Filter  $P(f)$ , welches das Signal in ein minimales Signal (Impulse) wandelt. Das inverse Filter von diesem Filter ist das gesuchte Vokaltraktfilter.*

$$F(f) = P^{-1}(f)$$

# Lineare Prädiktion (LP, linear prediction)

1. Signal fenstern; für jedes Fenster tue:
2. Prädiktorformel versucht, *den nächsten Abtastwert aus den  $K$  vorangegangenen Abtastwerten vorherzusagen*:

$$s'(t_n) = -a_1 s(t_{n-1}) - a_2 s(t_{n-2}) - \dots - a_K s(t_{n-K})$$

$$s'(t_n) = - \sum_{k=1}^K a_k s(t_{n-k})$$

3. Voraussage kann nicht optimal sein. Der Fehler ist:

$$e(t_n) = s(t_n) - s'(t_n) = s(t_n) + \sum_{k=1}^K a_k s(t_{n-k}) = \sum_{k=0}^K a_k s(t_{n-k})$$

4. Wiederholung für alle Abtastwerte im Fenster der Länge  $N$   
 $\rightarrow (N - K)$  Gleichungen  $e(t_n)$  mit  $K$  Unbekannten  
 $a_0, a_1, a_2, \dots, a_K$

# Lineare Prädiktion (LP, linear prediction)

5. Überspezifiziertes Gleichungssystem ( $N \gg K$ )  
→ keine analytische Lösung, aber Optimierungsverfahren (root mean square minimization) liefern die LP Koeffizienten (LP coefficients, LPC), welche ein minimales Fehlersignal  $e(t_n)$  im ganzen Fenster ergeben  
→  $K$  LP-Koeffizienten  $a_k$  für jedes Fenster in Signal  
→ sog. *inverses Filter* pro Fenster:

$$e(t_n) = \sum_{k=0}^K a_k s(t_{n-k}) \quad (\text{FIR-Filter})$$

6. Die Z-Transformierte dieses *inversen Filters* ist ein einfaches Polynom nach  $z$ :

$$P(z) = \sum_k a_k z^{-k} \quad E(z) = P(z)S(z)$$

# Lineare Prädiktion (LP, linear prediction)

- Invertierung dieses Filters = geschätzte Filterfunktion des Ansatzrohres zum Zeitpunkt des Fensters im Signal.  
Invertieren im Frequenzbereich: Kehrwert bilden

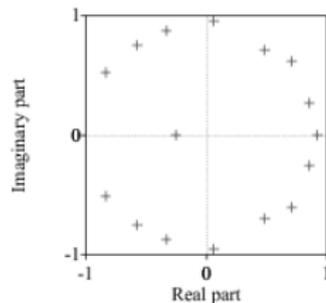
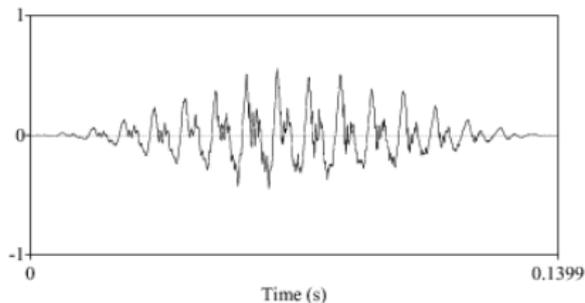
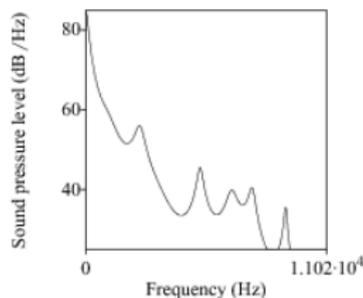
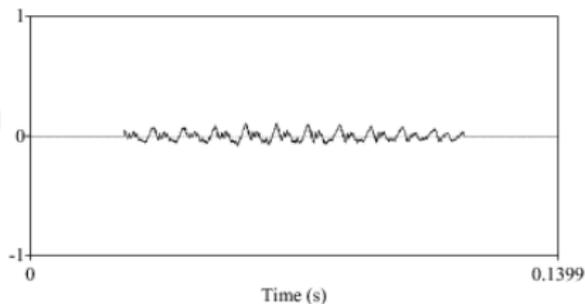
$$F(z) = P^{-1}(z) = \frac{1}{\sum_k a_k z^{-k}}$$

- Das Fehlersignal  $e(t_n)$ , also die Filterung von  $s(t_n)$  im Fenster mit  $P(z)$  ist das gesuchte Anregungssignal  $q(t)$

## LPC-Analyse

Die LPC-Analyse liefert für ein Sprachsignal pro Fensterung eine Schätzung für das Fant'sche Filter  $F(f)$  und eine Schätzung für das Quellensignal  $q(t)$  in diesem Fenster.

## Lineare Prädiktion (LP, linear prediction)

Beispiel: LPC mit  $K = 16$  in einem /u:/Signal  
 $s(t)$ Pole in der  
 $z$ -  
EbeneAnregung  
 $q(t)$ Filter  
 $F(f)$

# Lineare Prädiktion (LP, linear prediction)

## Diskussion des LP-Modells:

- Nach LP hat das Vokaltraktfilter nur Polstellen, keine Nullstellen; die Pole müssen in der  $z$ -Ebene innerhalb des Einheitskreises liegen, sonst Filter instabil  
→ diese Modellannahme ist eher unwahrscheinlich; z.B. entstehen durch die Ankoppelung des Nasenraums *Antiformanten*, welche nur durch Nullstellen in der Filterfunktion modelliert werden können.
- Nach LP ist das Anregungssignal minimal  
→ in der Realität besteht das Glottissignal nicht nur aus idealen Impulsen  
(durch ein Pre-emphasis-Filter kann dies jedoch kompensiert werden; s. Pfister & Kaufmann, 2008, S. 87)
- LP-Modell funktioniert nicht bei impulsartigen oder stochastischen Anregungssignalen.

# Lineare Prädiktion (LP, linear prediction)

Warum macht man (trotzdem) LPC-Analyse?

- Grundfrequenzbestimmung: leichter aus dem Anregungssignal (weil Impulssignal)
- Bestimmung der Formanten: leichter aus der Filterfunktion, weil Obertöne nicht mehr vorhanden
- Kompression: leichter das Anregungssignal und das Filter zu einem Handy zu übertragen als das Sprachsignal
- Manipulation: durch Veränderung der Anregung und anschließender Re-synthese lässt sich z.B. die Intonation ändern (z.B. für Sprachsynthese)
- Spracherkennung: leichter Sprache aus der Filterfunktion zu erkennen als aus dem Spektrum, weil Obertöne nicht mehr stören

# Lineare Prädiktion (LP, linear prediction)

## Demo: Praat - Signal-Manipulation

- *Signal laden*
- *To Manipulation* : LPC, Grundfrequenz wird bestimmt
- *Edit Manipulation* : Abstände der Impulse im Anregungssignal werden manipuliert
- *Play (LPC)* : LPC-Resynthese mit manipuliertem Anregungssignal
- *Manipulation : Extract Pulses* : erzeugt Pulsfolge aus Anregung (Point process)
- *Point Process : To Sound (hum)* : Resynthese mit neutralem Vokaltrakt

# Lineare Prädiktion (LP, linear prediction)

## Demo: Praat - Signal-Manipulation

- *Sprachsignal laden; Resample auf 16kHz*
- *'Analyse Spectrum / To LPC (autocorr.)'* : LPC-Filter werden berechnet (order 16)
- *Zikadensignal laden; Resample auf 16kHz; auf ungefähr gleiche Länge schneiden ('Convert / Extract part ...')*
- *LPC Re-synthese* : LPC Objekt und geschn. Zikadensignal markieren und *'Filter ...'*
- *Play* : LPC-Resynthese mit Zikadenanregung; perfekt zu verstehen

Fazit: linguistische Information hauptsächlich im LPC Filter

# Cepstrum

## Cepstrum

Das Cepstrum eines Signals beschreibt die Energie von periodische Formen ('Welligkeiten') im Amplitudenspektrum.

Cepstrum = **S**pectrum mit invertiertem 'spec'

Spektrum :  $N$  Spektralwerte (DFT)

Cepstrum :  $N$  cepstrale Werte

$$C(q_0), C(q_1), C(q_2), \dots, C(q_{N-1})$$

wobei die  $q_n$ ,  $n = 0 \dots N - 1$  hier keine diskreten Frequenzen bezeichnen sondern *quefrequencies* (gemessen in  $\frac{1}{\text{Hz}}$ )

# Cepstrum

Jede *quefreny* steht für eine bestimmte trigonometrische Form im Spektrum:

Die quefreny  $q_0$  bezeichnet einen 'unendlich langen Cosinus', also den *Gleichanteil* im Spektrum (= Gesamtenergie).

$q_1$  steht für eine Welligkeit, die einer halben Cosinus-Schwingung über das ganze Spektrum entspricht (= abfallendes Spektrum).

$q_2$  steht für eine Welligkeit, die einer ganzen Cosinus-Schwingung über das ganze Spektrum entspricht, also ein Spektrum in Form einer Senke u.s.w.

Cepstrum = spektrale Grobstruktur (Filter) bei niedrigen quefrequencies + spektrale Feinstruktur (Quelle) bei hohen quefrequencies

# Cepstrum

Cepstrum entspricht *qualitativ* einer Discrete Cosine Transform (DCT) des logarithmierten Amplitudenspektrums.

Häufige Interpretationen der niedrigen Cepstralkoeffizienten:

$q_0$  : Gesamtenergie des Signals : immer  $\geq 0$

$q_1$  : Neigung des Spektrums :  $> 0 \Rightarrow$  fallend,  $< 0 \Rightarrow$  steigend

$q_2$  : Krümmung des Spektrums :  $> 0 \Rightarrow$  Senke/Tal,  $< 0 \Rightarrow$  Berg

$q_3$  : Asymmetrie des Spektrums :  $> 0 \Rightarrow$  mehr Energie bei hohen Frequenzen,  $< 0 \Rightarrow$  mehr Energie bei tiefen Frequenzen

# Cepstrum

## Berechnung des Cepstrums

1. Berechne das Amplitudenspektrum

$$|S(f_n)| = |\mathcal{F}\{s(t_n)\}|$$

2. Logarithmiere das Spektrum (jeder einzelne Spektralwert wird logarithmiert)

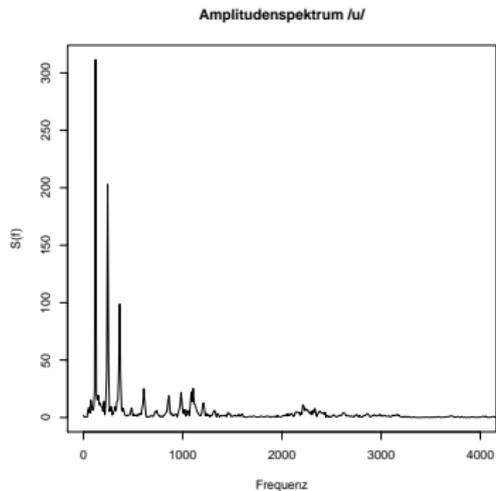
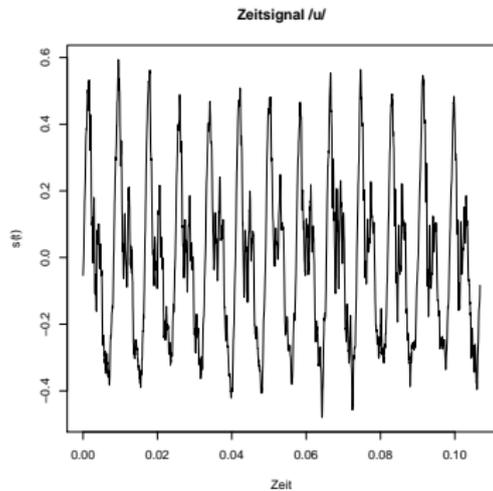
$$\log[|S(f_n)|]$$

3. Inverse Fouriertransformation

$$C(n) = \mathcal{F}^{-1}\{\log[|S(f_n)|]\}$$

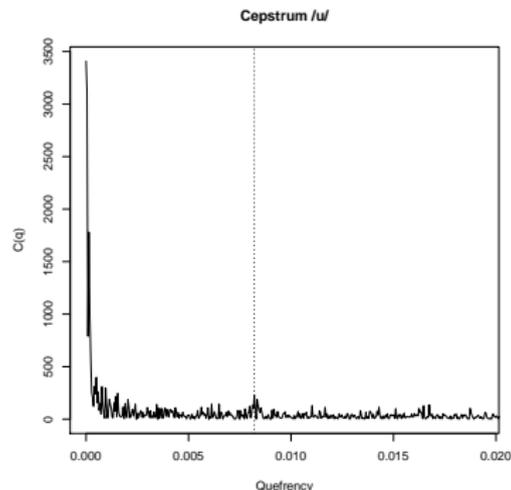
# Cepstrum

## Beispiel

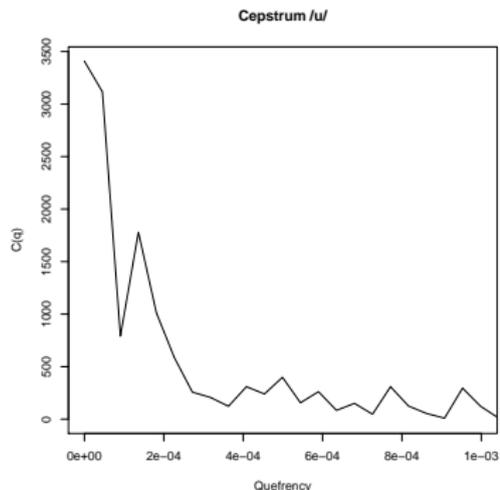


## Cepstrum

$$|\text{Cepstrum}| \quad q = 0 \dots 0.02 \frac{1}{\text{Hz}}$$



$$|\text{Cepstrum}| \quad q = 0 \dots 0.001 \frac{1}{\text{Hz}}$$



Das Maximum bei  $q_f = 0.008 \frac{1}{\text{Hz}}$  stammt von der Grundfrequenz ( $f_0 = \frac{1}{q_f} \approx 120\text{Hz}$ ).

Der Wert bei  $q = 0 \frac{1}{\text{Hz}}$  ist die logarithmierte Gesamtenergie.

Der hohe Wert bei  $q_1$  entsteht durch das stark abfallende Spektrum.

# Cepstrum

## Motivation für das Cepstrum

### Quelle-Filter-Modell (Fant 1960)

Sprachsignal entsteht durch Signal der Quelle (z.B. Glottis) gefiltert von einem linearen Filter (z.B. Ansatzrohr).

→ Sprachspektrum ist Produkt aus Quelle und Filter

$$S(f) = Q(f) \cdot F(f)$$

Logarithmus: Multiplikation wird zu Addition

$$\log[S(f)] = \log[Q(f)] + \log[F(f)]$$

# Cepstrum

Inverse Fouriertransformation: Addition bleibt erhalten  
(weil lineares System):

$$\mathcal{F}^{-1}\{\log[S(f)]\} = \mathcal{F}^{-1}\{\log[Q(f)]\} + \mathcal{F}^{-1}\{\log[F(f)]\}$$

→ im Cepstrum sind Quelle und Filter additiv verbunden und nicht mehr gefaltet.

→ ungewünschte Teile können einfach subtrahiert werden

→ *liftering* (abgeleitet von **filtering**)

D.h. man kann z.B. im Cepstrum die Werte bei hohen Frequenzen (= Quelle) einfach löschen, das Cepstrum zurück in ein Spektrum transformieren und erhält die Übertragungsfunktion des Ansatzrohres (= Filter).

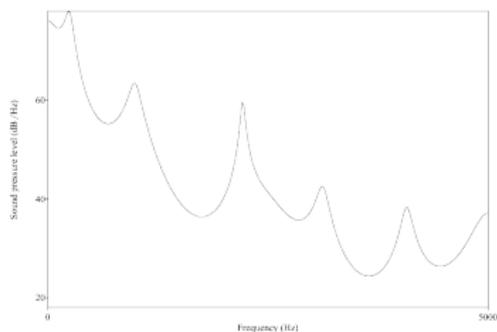
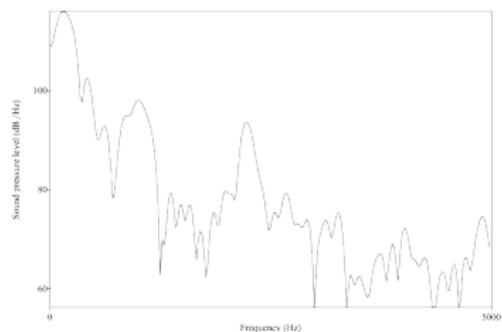
# Cepstrum

## Wofür verwendet man das Cepstrum?

- Bestimmung der Grundfrequenz (= Position des Maximums)
- Cepstrale Glättung (cepstral smoothing):  
Im Cepstrum werden die Werte bei hohen Frequenzen entfernt (= *liftering*) und wieder zurücktransformiert  
→ Spektrum ohne Obertöne
- Automatische Spracherkennung:  
Cepstral-Werte bei niedrigen Frequenzen eignen sich sehr gut als Muster (weil sie die Grobstruktur des Spektrums beschreiben);  
noch besser: *Mel Frequency Cepstral Coefficients* (MFCC):  
Cepstrum berechnet aus einer Mel-Filterbank.

# Cepstrum

Beispiel: Sprachsignal /u:/  
Normales Spektrum und cepstral geglättetes Spektrum  
(liftering der ersten 16 cepstralen Koeffizienten)



# Fragen

*Was versucht die sog. Prädiktorformel der Linearen Prädiktion?*

*Warum verwendet die LP ein Optimierungsverfahren, um die LP-Coeffizienten zu bestimmen?*

*Sie haben ein Sprachsignal mit 1sec Länge und eine Fensterfunktion mit 50msec Länge (nicht-überlappend). Sie machen eine cepstrale Analyse und liftern auf die ersten 16 Cepstral-Koeffizienten. Welche/wieviele Daten erhalten Sie aus Ihrer Analyse?*

*Was unterscheidet das Cepstrum vom Sprachsignal?*

*Warum funktioniert cepstrale Glättung besser als eine einfache Glättung des Spektrums?*