

Einführung in die Signalverarbeitung

Phonetik und Sprachverarbeitung, 2. Fachsemester,
Block Sprachtechnologie I

Florian Schiel

Institut für Phonetik und Sprachverarbeitung, LMU München

Signalverarbeitung - Teil 5

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 I

Anwendung der Methoden auf die Verarbeitung von Sprache.

- Höhenanhebung (pre-emphasis)
- Kurzzzeitanalyse (short-time analysis)
- Verwendung von Fensterfunktionen (windowing)
- DFT: Interpolation durch *zero-padding*

Höhenanhebung (pre-emphasis)

Stimmhaftes Sprachspektrum fällt im Mittel mit ca. -6dB/Oktave zu hohen Frequenzen ab.

Grund:

Spektrum des Glottissignals hat -12dB/Oktave,
Übertragungsfunktion der Kugelabstrahlung am Mund hat +6dB/Oktave

→ insgesamt -6dB/Oktave (z.B. *Harrington et al, 1999*)

Für ausgewogenes Spektrum (Formanten besser sichtbar):

→ Filterung mit Hochpass +6dB/Oktave (pre-emphasis)

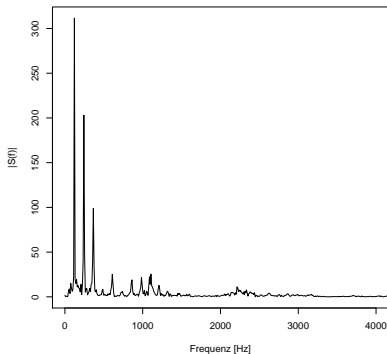
Filterformel:

$$\check{s}(t_n) = s(t_n) - 0.95s(t_{n-1})$$

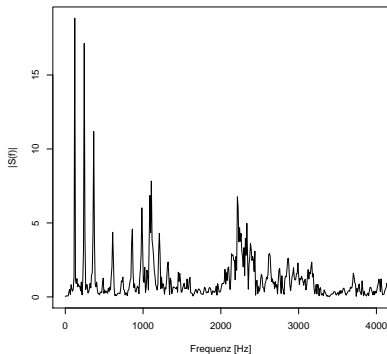
Höhenanhebung (pre-emphasis)

Beispiel: Normales Amplitudenspektrum und nach Höhenanhebung

Amplitudenspektrum



Pre-Emphasis Amplitudenspektrum



Kurzzzeitanalyse

Unterscheidung:

- *Langzeitanalyse* : Berechnungen über längeren Zeitbereich (z.B. Wort, Äußerung)
→ Merkmal m *zeitunabhängig* (siehe Teil 2, *long term feature*)
- *Kurzzzeitanalyse* : Berechnungen zu einem bestimmten Zeitpunkt t_k
→ Merkmale $m(t_k)$ *zeitabhängig*

Beispiel: *Sonagramm*

Das Sonagramm zeigt Amplitudenspektren (DFT) im zeitlichen Verlauf. Jedes einzelne Spektrum wird zu einem bestimmten Zeitpunkt berechnet.

→ Sonagramm ist zweidimensional: $|S(f_n, t_k)|$

Kurzzzeitanalyse

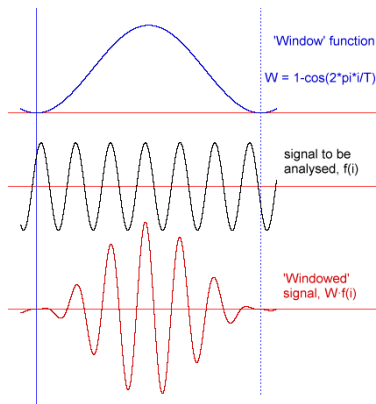
Berechnung zu einem Zeitpunkt t_k :
Signalstück wird symmetrisch um t_k aus dem Signal
herausgeschnitten.

Schneiden = Multiplikation mit
einer Fensterfunktion $f(t_{n-k})$:

$$s_f(t_n) = s(t_n) \cdot f(t_{n-k})$$

$s_f(t_n)$ ist dann der Input für die
eigentliche Berechnung.

Typisch: Verschiebung um 5-25ms
Wiederholung über ganzes Signal



Fensterfunktionen

Für manche Berechnungen spielt die Fensterform keine Rolle (z.B. Energie, Nulldurchgangsrate)

Problem

Werden in der Kurzzzeitanalyse *spektrale* Berechnungen durchgeführt, beeinflusst die Form der Fensterfunktion das Spektrum!

Denn: Multiplikation im Zeitbereich entspricht Faltung im Spektralbereich:

$$s(t_n) \cdot f(t_{n-k}) \quad \circ - \bullet \quad S_D(f_n, t_k) \star F_D(f_n)$$

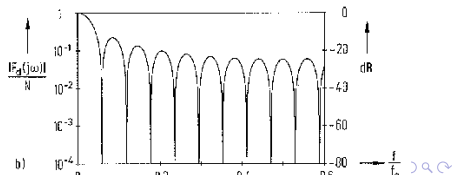
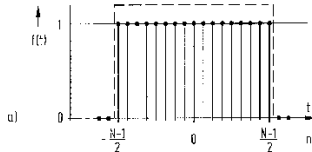
d.h. statt gewünschtem Spektrum $S_D(f_n, t_k)$ messen wir dieses *gefaltet* mit der DFT der Fensterfunktion.

Fensterfunktionen

Beispiel: Rechteckfenster

Aus Teil 3 (Systemtheorie / Digitale Filter) wissen wir:
Rechteck $\circ - \bullet$ Si-Funktion

Si-Funktion hat starke
Nebengipfel
→ keine ideale Impulsform
→ bei Faltung werden
benachbarte Spektralwerte
geglättet
→ Original-Spektrum wird
verschmiert



Fensterfunktionen

Gesucht: Fensterfunktion mit möglichst schmalen Spektrum

Viele verschiedene (keine ist ideal):

- Bartlett (Dreiecksfunktion)
- von Hann-Fenster (Cosinus-Halbschwingung)
- Hamming-Fenster (geplättete Cosinus-Halbschwingung)

$$f_{hamming}(t) = 0.54 + 0.46 \cos(2\pi t)$$

- Welsh (quadratische Funktion)
- Blackman (mehrere Cosinus-Terme)
- Kaiser (modifiziert Besselfunktion)

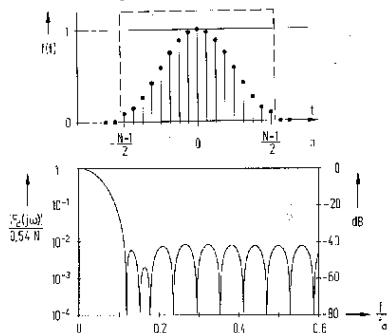
Häufig verwendet: *Hamming* und *Blackman*

Fensterfunktionen

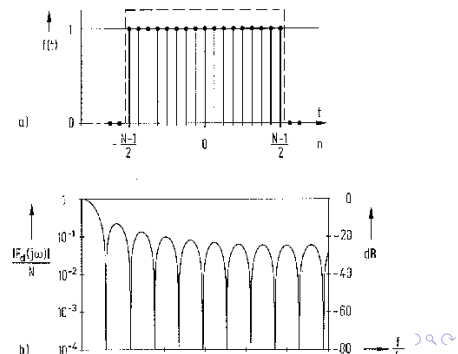
Beispiel: Hamming-Fenster (hamming window)

$$f_{\text{hamming}}(t) = 0.54 + 0.46 \cos(2\pi t)$$

Hamming-Fenster



Rechteckfenster



DFT: Interpolation durch *zero-padding*

Reziprozitätsgesetz:

”Zeitauflösung umgekehrt proportional zur Frequenzauflösung”

und:

Diskrete Fouriertransformation (DFT):

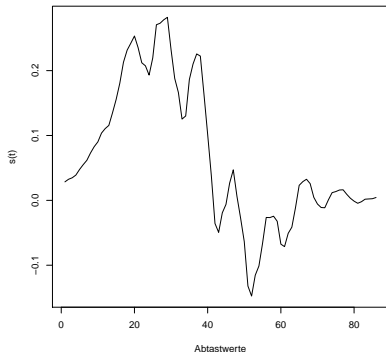
”DFT berechnet aus N Abtastwerten $\frac{N}{2}$ Frequenzwerte”

Beispiel:

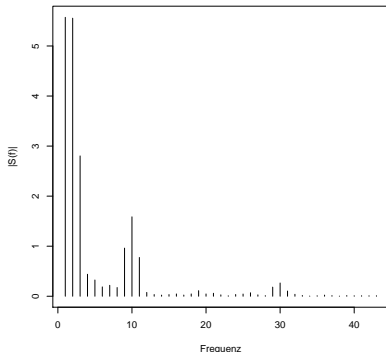
Abtastrate sei 16000Hz → Nyquistfrequenz ist somit 8000Hz
Fensterung mit Hamming-Fenster der Breite 80 Samples (5ms)
und DFT liefert Breitbandspektrum mit 40 Linien.
Frequenzabstand der Linien ist: $8000/40 = 200\text{Hz}$

DFT: Interpolation durch *zero-padding*

Gefensterteres Zeitsignal der Länge 5ms



Amplitudenspektrum



Das gleiche Spektrum (also immer noch Breitband!) aber *mit mehr Linien* erhält man durch *zero-padding* des Zeitsignals.

DFT: Interpolation durch *zero-padding*

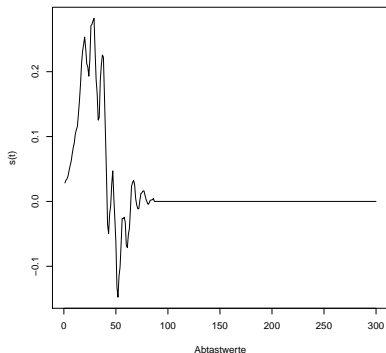
Dem Zeitsignal werden Null-Werte hinzugefügt

→ N wird größer

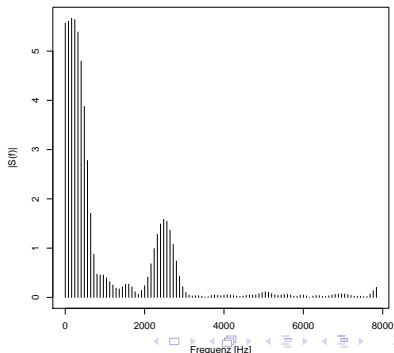
→ Spektrum enthält mehr Linien ($\frac{N}{2}$)

→ Interpolation zwischen den ursprünglichen 40 Spektrallinien

Zeitsignal mit zero padding



Amplitudenspektrum zero-padded



DFT: Interpolation durch *zero-padding*

Beachte:

Die Qualität des Spektrums wird durch *zero-padding* *nicht* verbessert. Das Spektrum wird lediglich mit mehr Linien, d.h. feiner abgetastet, dargestellt.

Nach der gleichen Methode kann man die Abtastrate eines Signals erhöhen, indem man das Spektrum mit Nullwerten verlängert. Auch hier hat man dann zwar eine nominell höhere Abtastrate f_{abt} , aber der dadurch gewonnene Spektralbereich bleibt natürlich leer.

(vgl. Kapitel *re-sampling* in Teil 1)

Fragen

Wie sieht die Impulsantwort des Filter der Höhenanhebung (pre-emphasis) aus?

Ist ein Sonogramm das Ergebnis einer Langzeitanalyse? Warum?

Welche mathematischen Operationen werden bei der Kurzeitanalyse angewandt? Addition, Multiplikation oder Faltung?

Was für eine Fouriertransformierte hat eine Rechteckfunktion? Warum ist es daher keine geeignete Fensterfunktion für spektrale Berechnungen?

Wird durch die Anwendung des zero-paddings die Frequenzauflösung des Spektrums verbessert? Warum?