Einführung in die Signalverarbeitung

Phonetik und Sprachverarbeitung, 2. Fachsemester, Block Sprachtechnologie I

Florian Schiel

Institut für Phonetik und Sprachverarbeitung, LMU München

Signalverarbeitung - Teil 4

Allgemeines

- Unterrichtssprache ist Deutsch (englische Fachbegriffe in Klammern)
- Fragen am besten sofort; besser einmal zuviel gefragt
- Literatur/Videos:
  - Barry van Veen: Video-Turorial Introduction to the Z-transform https://www.youtube.com/watch?v=NjRiqHSGmB0.
  - Pfister B, Kaufmann T (2008): Sprachverarbeitung - Grundlagen und Methoden der Sprachsynthese und Spracherkennung. Springer-Verlag Berlin Heidelberg.

Abtastung des Signals

Bisher: Fourier von analogen Signalen
Frage: Was passiert mit dem Spektrum beim digitalen (abgetasteten) Signal?

Abtastung = Multiplikation mit periodischem Impulssignal

$$s(t_n) = s(t) \cdot i(t)$$

mit

$$i(t) = \begin{cases} 1 & : \quad t = 0, T_{abt}, 2T_{abt}, 3T_{abt} \dots \\ 0 & : \quad \text{sonst} \end{cases}$$

$\rightarrow$ Spektrum des digitalen Signals

$$S_n(f) = \mathcal{F}\{s(t) \cdot i(t)\} = S(f) \star I(f)$$

Abtastung des Signals

Was aber ist das Spektrum des Impulssignals $i(t)$?
Impulssignal = periodisches Signal mit Periodendauer T_{abt}
$\rightarrow$ Spektrum $I(f)$ muss ein Linienspektrum sein mit Linienabstand $f_{abt} = \frac{1}{T_{abt}}$
(in diesem Fall sind die Linien alle gleich groß.)

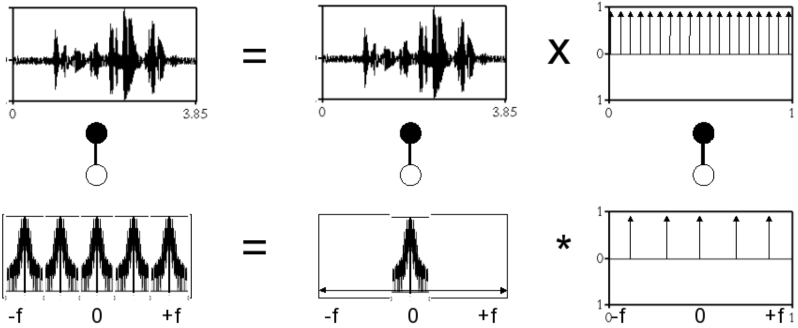
Was bedeutet Faltung mit Linienspektrum?

Wdh: Faltung eines Signals mit einem Impuls = Reproduktion des Signals

$\rightarrow$ Faltung mit vielen Impulsen = Mehrfache (versetzte) Reproduktion des Signals

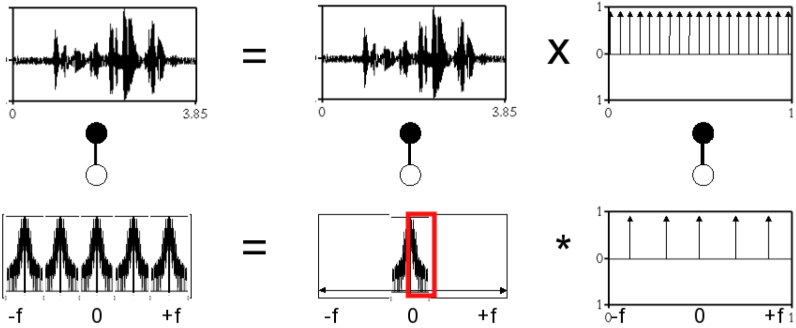
Abtastung des Signals

Die Abtastung des Signals $s(t)$ bewirkt also eine Replikation des Spektrums bei ganzzahligen Vielfachen von $f_{abt} = \frac{1}{T_{abt}}$:



Abtastung des Signals

Die Abtastung des Signals $s(t)$ bewirkt also eine Replikation des Spektrums bei ganzzahligen Vielfachen von $f_{abt} = \frac{1}{T_{abt}}$:



Da die wiederholten Spektren alle identisch und symmetrisch, betrachtet man nur den positiven Frequenzbereich $0 \dots \frac{f_{abt}}{2}$

Abtasttheorem II

Was passiert, wenn das Signal Frequenzen $f > \frac{f_{abt}}{2}$ enthält?

$\rightarrow$ Hohe Frequenzen benachbarter Spektren überlagern sich
$\rightarrow$ Abtastbedingung nicht erfüllt
$\rightarrow$ *aliasing effect* (Spiegelfrequenzen hörbar)

Bitte beunruhigen Sie sich nicht wegen des Begriffs 'negative Frequenzen'. Diese entstehen durch die Fouriertransformation, haben aber keine tiefere Bedeutung, außer dass sie als Spiegelfrequenzen eine Störung verursachen können.

Fourier in der Praxis

In der Praxis kann man keine unendlichen Signale oder Spektren berechnen.
Jedes digitale Signal ist zeitlich begrenzt und hat N Abtastwerte.

Die *diskrete Fouriertransformation* (DFT):

- implizite Annahme: das (zeitlich begrenzte) Signal $s(t_n) \quad (n = 0 \ldots N)$ ist nur eine 'Periode' eines unendlichen periodischen Signals
- Fourierreihenentwicklung liefert N 'Linien' (weil Linienspektrum) mit Abstand $\Delta f = \frac{f_{abt}}{N}$ symmetrisch im Bereich $-\frac{f_{abt}}{2} \cdots + \frac{f_{abt}}{2}$
- Jede Linie besteht aus 2 Werten: Amplitude und Phase der Sinusschwingung

Diskrete Fouriertransformation (DFT)

- Je länger also das Signalstück (N) desto feiner die *Frequenzauflösung* des berechneten, diskreten (weil Linien) Spektrums
- In der Praxis genutzt wird nur das positive Spektrum, also $\frac{N}{2}$ Linien
- Da jede Linie aus zwei Werten, Amplitude und Phase, besteht, sind es wieder genau N Werte wie im Signal $s(t_n)$.

Die DFT berechnet also aus N Abtastwerten ein komplexes Spektrum für $\frac{N}{2}$ positive Frequenzwerte, die sich von $f = 0 \ldots \frac{f_{abt}}{2} Hz$ erstrecken.
Jeder Frequenzwert besteht aus Amplitude und Phase der im Signal enthaltenen Sinusschwingungen.

(Phase wird oft vernachlässigt. Beachte aber: aus dem Amplitudenspektrum allein lässt sich $s(t_n)$ nicht rekonstruieren!)

Diskrete Fouriertransformation (DFT)

Formeln der DFT

Üblicherweise wird die DFT nicht in Abhängigkeit der kontinuierlichen Frequenz f sondern der diskreten *Kreisfrequenz* ω_k berechnet:

$$S_D(\omega_k) = \sum_{n=0}^{N-1} s(t_n) e^{-j2\pi kn/N}$$

$$\omega_k = \frac{2\pi k}{NT_{abt}} \qquad k = 0 \ldots N-1$$

Reziprozitätsgesetz der Systemtheorie

Zeitliche und spektrale *Auflösung* (= Messgenauigkeit) sind voneinander abhängig

$$\Delta f \cdot \Delta t \equiv \text{Konstante} \qquad (\Delta t \neq T_{abt}!)$$

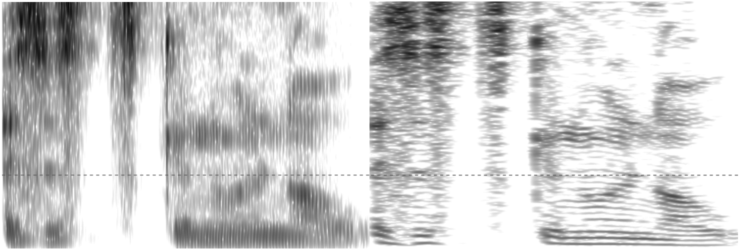
Anschauliche Erklärung:

- Zeitsignal kürzer $\rightarrow$ Zeitauflösung Δt kleiner (genauer, weil wir genauer wissen, wo das Signal stattfindet!)
- $\rightarrow$ N wird kleiner (weniger Abtastwerte)
- $\rightarrow$ weniger Werte ($\frac{N}{2}$) im Spektrum $S_D(\omega_k)$
- Frequenzbereich $0 \ldots \frac{f_{abt}}{2}$ bleibt aber gleich, weil die Abtastrate f_{abt} gleich bleibt
- $\rightarrow$ weniger Spektralwerte für gleichen Frequenzbereich
- $\rightarrow \Delta f$ wird größer, Frequenzauflösung schlechter

Reziprozitätsgesetz der Systemtheorie

Das Reziprozitätsgesetz findet sich beim Sonagramm wieder in dem Unterschied

- *Schmalbandsonagramm* : Zeitfenster ist groß ($> 25msec$)
  $\rightarrow$ Zeitauflösung ist schlecht (Glottisschläge nicht sichtbar)
  $\rightarrow$ Frequenzauflösung ist gut (Obertöne sichtbar)
- *Breitbandsonagramm* : Zeitfenster ist klein ($< 25msec$)
  $\rightarrow$ Zeitauflösung ist gut (Glottisschläge sichtbar)
  $\rightarrow$ Frequenzauflösung ist schlecht (Obertöne nicht sichtb.)



Z-Transformation : Motivation

Ein lineares zeitdiskretes System (z.B. ein digitales Filter) lässt sich nicht immer zufriedenstellend durch seine Übertragungsfunktion $G(f)$ beschreiben.
(ein Signal schon!)

$G(f)$ ist die Fouriertransformierte der Impulsantwort $g(t)$
Was aber, wenn die Impulsantwort unendlich lang ist wie bei einem IIR-Filter?

Wie kann man bei einem solchen System aus der Filterformel die Übertragungsfunktion berechnen?

Z-Transformation : Motivation

Warum ist das für die Phonetik wichtig?

Vokaltrakt = Rohr = akustisches Filter mit Resonanzen
Resonanzen (Formanten) = Frequenzen werden verstärkt =
= IIR Filter

Vokaltrakt + Nasenraum = Rohr mit Verzweigung =
= akustisches Filter mit Formanten und Antiformanten =
= IIR + FIR Filter

Um also den Vokaltrakt als akustisches Filter zu modellieren, braucht man ein Filter, das FIR und IIR Anteile enthält.

Z-Transformation : Motivation

Z-Transformierte

Die *Z-Transformierte* beschreibt beliebige zeitdiskrete lineare Systeme und man kann aus ihr die Übertragungsfunktion leicht ableiten.

Die Z-Transformierte ist definiert als eine Funktion der komplexen Variable z: $F(z)$
D.h. die abhängige Variable ist zweidimensional in der komplexen Ebene, der sog. *Z-Ebene* (z-plane).
In der Z-Ebene ist die X-Achse der Realteil von z, die Y-Achse der Imaginärteil von z. Über der Z-Ebene erhebt sich die Z-Transformierte $F(z)$ wie ein dreidimensionales Gebirge.

Z-Transformation : Filterformel

Die Z-Transformierte eines linearen Filters besteht aus zwei Polynomen in Nenner und Zähler:

$$F(z) = \frac{a_0 + a_1 z^{-1} + a_2 z^{-2} + \dots}{b_1 z^{-1} + b_2 z^{-2} + \dots} = \frac{a_0 + \sum_k a_k z^{-k}}{\sum_j b_j z^{-j}}$$

wobei die a_k die FIR- und die b_j die IIR-Filterkoeffizienten sind.

Beide Polynome können für bestimmte z zu Null werden (z.B. hat das Polynom z^{-2} bei $z = 0$ eine Nullstelle).

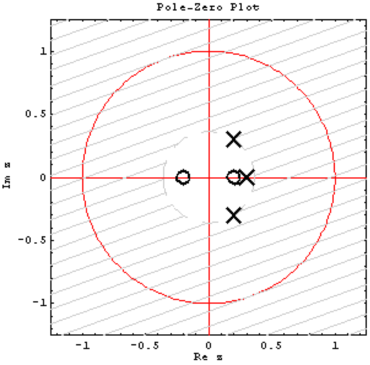
Wenn $F(z_0) = 0$ spricht man von einer *Nullstelle* bei z_0;
wenn $F(z_p) \to \infty$ spricht man von einer *Polstelle* bei z_p.
(weil das Polynom im Nenner zu Null wird; $1/0 \to \infty$.)

Z-Transformation : Filterformel

Nullstellen kann man sich als *Senken*, Polstellen als sehr hohe *Spitzen* in der Z-Ebene vorstellen.
Zusammen bilden sie das 'Gebirge' der Z-Transformierten $F(z)$.

Z-Ebene mit eingezeichneten Nullstellen ($\circ$) und Polstellen ($\times$) eines einfachen Filters

Pol- und Nullstellen sind i.d.R. *konjugiert-komplexe* Paare (d.h. je 2 symmetrisch zur X-Achse).
Polstellen dürfen nur innerhalb des *Einheitskreises* (Radius 1) liegen.



Z-Transformation : Übertragungsfunktion

Die Übertragungsfunktion $F(f = 0 \ldots \frac{f_{abt}}{2})$ des Filters mit der Z-Transformierten $F(z)$ ist der Verlauf von $F(z)$ über dem halben Einheitskreis (von 0 bis 180°), also für

$$z = e^{j\phi} \quad \text{für} \quad \phi = 0 \ldots \pi$$

Entsprechend findet man auch die DFT auf dem Einheitskreis, aber an diskreten Punkten:

$$z_D = e^{j\frac{2\pi k}{N}} \quad \text{für} \quad k = 0 \ldots \frac{N}{2}$$

$\rightarrow$ Lage von Pol- und Nullstellen abhängig von Filterkoeff.
$\rightarrow$ Pol- und Nullstellen bestimmen die Form der Übertragungsfunktion: *in der Nähe von Polen ist sie hoch, in der Nähe von Nullstellen ist sie niedrig.*

Fragen

Warum ist das Spektrum eines regelmäßig wiederholten Impulses ein Linienspektrum?

Wie kann man die Abtastbedingung systemtheoretisch erklären?

Warum besteht jede Spektrallinie in der DFT aus zwei Werten?

Wenn man die Länge eines Analysefensters verdoppelt, was passiert dann im DFT-Spektrum?

Was ist der Vorteil der Z-Transformation bei der Beschreibung von digitalen Filtern?