Einführung in die Signalverarbeitung

Phonetik und Sprachverarbeitung, 2. Fachsemester, Block Sprachtechnologie I

Florian Schiel

Institut für Phonetik und Sprachverarbeitung, LMU München

Signalverarbeitung - Teil 2

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.
  - Schrüfer 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.

Wiederholung

Digitales Signal ist eine Tabelle von Zahlen.
Jede Zahl repräsentiert einen Abtastwert $s(t_n)$

Beispiel:
od -s signal.wav | more

Verschiebung (*shift*)

Digitale Signal können mit einer konstanten Zahl V addiert werden

$$v(t_n) = V + s(t_n)$$

Ist $V > 0.0$: Verschiebung zu pos. Werten
Ist $V < 0.0$: Verschiebung zu neg. Werten

Beispiel:
Verschiebung des Signals um 165 nach unten:

$$v(t_n) = s(t_n) - 165$$

Frage: Was passiert bei Addition von 0?

Skalierung

Digitale Signal können mit einer konstanten Zahl M multipliziert werden

$$m(t_n) = M \cdot s(t_n)$$

Ist $M > 1.0$: Verstärkung (lauter)
Ist $M < 1.0$: Abschwächung (leiser)

Beispiel:
Verdoppelung der Amplitude (entspricht +6dB):

$$b(t_n) = 2.0 \cdot s(t_n)$$

Frage: Was passiert bei Skalierung mit 0.0?
Frage: Was passiert bei Skalierung mit -2.0?

Addition

Digitale Signal können werteweise addiert werden

$$s(t_n) = a(t_n) + b(t_n)$$

Voraussetzung: $a(t_n)$ und $b(t_n)$ sind gleich lang $n = 0...N$ und haben die gleiche Abtastrate

Beispiel:
Umwandlung eines Stereo-Signals (Kanal $l(t_n)$ und $r(t_n)$) in ein Monosignal $m(t_n)$:

$$m(t_n) = \frac{1}{2}[l(t_n) + r(t_n)]$$

Frage: Warum die Skalierung mit $\frac{1}{2}$*?*

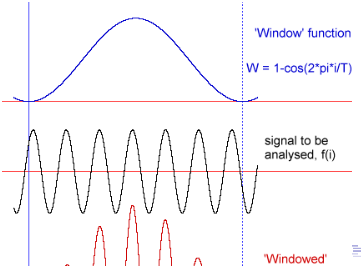
Multiplikation

Digitale Signal können werteweise multipliziert werden

$$s(t_n) = a(t_n)b(t_n)$$

Voraussetzung: *müssen nicht gleich lang sein bzw. bei unterschiedlicher Länge wird angenommen, das Signal ist außerhalb Null.*

Beispiel:
'Weiches' Ausschneiden eines Signalstücks aus längerem Signal



Faltung (*convolution*)

Signal $a(t_n)$ wird mit $b(t_n)$ zu einem neuen Signal $g(t_n)$ 'gefaltet' (*convoluted*):

$$g(t_n) = \sum_k a(t_k)b(t_{n-k}) = a(t_n) \star b(t_n)$$

Meistens ist $a(t_n)$ länger als $b(t_n)$ (nicht zwingend)
Der Summenparameter k läuft über den gesamten Bereich, in dem die Überlappung der beiden Signale nicht Null ist.

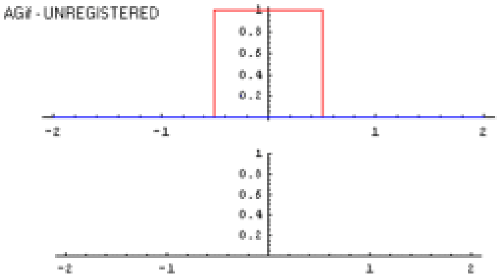
Anschaulich:
Signal $b(t_n)$ wird an der Y-Achse gespiegelt $(-k)$ und von links über das Signal $a(t_k)$ geschoben $(n-k)$. In jeder möglichen Verschiebestellung n wird k-mal multipliziert und alle Ergebnisse aufaddiert. Das ergibt dann den Wert von g an der Stelle t_n.

Faltung

$$g(t_n) = a(t_n) \star b(t_n) =$$

$$= \sum a(t_k) b(t_{n-k})$$

rotes Rechteck : $a(t_n)$
grünes Rechteck : $b(t_{n-k})$
gelbe Fläche : $a(t_k)b(t_{n-k})$
unten : $g(t_n)$



Langzeitanalyse

Langzeitanalyse (long term feature, LTF)

Berechnung von Werten aus einem längeren Signal (z.B. einer Äußerung)

$$F = L(s(t_n)) \qquad L() : \text{Funktion der Langzeitanalyse}$$

D.h. Wert F aus Langzeitanalyse ist nicht vom Zeitpunkt t_n abhängig, sondern gilt für eine ganze Äußerung.

Langzeitanalyse

Beispiele

- Durchschnittliche Energie
  (average energy, root mean square, RMS):
  Summe über alle quadrierten Abtastwerte, auf Anzahl der Abtastwerte normieren, Wurzel ziehen

$$\bar{E} = \sqrt{\frac{1}{N}\sum_{n=0}^{N} s(t_n)^2}$$

Frage: Warum muss quadriert werden?

Langzeitanalyse

Beispiele

- Signalrauschabstand (signal to noise ratio, SNR):
  Die Differenz zwischen mittlerem Pegel des Nutzsignals (Sprache) und dem Pegel des Hintergrund-/Störgeräuschs

$$L_{SNR} = L_{signal} - L_{noise} = 10 \cdot log_{10} \left[\frac{\bar{E}_{signal}}{\bar{E}_{noise}} \right]$$

Z.B. $L_{SNR} = 20dB$ bedeutet: Das Nutzsignal ist 20dB stärker als das Hintergrundgeräusch.
$L_{SNR} = 0dB$ bedeutet: Das Nutzsignal ist genauso stark wie das Hintergrundgeräusch.

Langzeitanalyse

Beispiele

- Mittelwertsbereinigung (remove bias):
  Viele Signale haben einen Gleichanteil (bias), d.h. sie sind nicht genau symmetrisch um den Nullwert. Für viele Analysen muss dieser Gleichwert aus dem Signal entfernt werden:

  $$\text{Bias} \qquad B = \frac{1}{N} \sum_{n=0}^{N} s(t_n)$$

  Verschiebung des Signals um den negativen Gleichanteil:

  $$\text{Bereinigtes Signal} \qquad \hat{s}(t_n) = s(t_n) - B$$

Langzeitanalyse

Probleme

- Richtige Normierung, so dass Länge keinen Einfluss hat
- Lange Pausen können Ergebnisse (wegen der Zeitnormierung) verfälschen, oft in Aufnahmen mit langer Pause vor/nach dem Signal
- Manche Analysen machen nur in bestimmten Signalbereichen Sinn, z.B.
  - Grundfrequenzanalyse (nur stimmhafte Bereiche)
  - Formantanalyse (nur Vokale)

Daher ist oft eine Vorsegmentierung notwendig

Digitale Filter

Filter

Methode zur Dämpfung/Verstärkung von definierten Frequenzbereichen aus einem Signal. Filter sind *lineare Systeme*, d.h. es dürfen keine zusätzlichen Frequenzen produziert werden, die nicht schon in das Filter eingespeist wurden

Digitales Filter

Rechenmethode zur Realisierung eines Filters für digitale Signale. Da lineares System, dürfen nur Addition und Multiplikation verwendet werden

Digitale Filter

Impulsantwort (impulse response)

Signal am Ausgang eines (digitalen) Filters, das entsteht, wenn am Eingang ein Impuls der Energie 1 eingespeist wird

Wir unterscheiden:

- Filter mit endlicher Impulsantwort (finite impulse response filter, FIR)
- Filter mit unendlicher Impulsantwort (infinite impulse response filter, IIR

Filter mit endlicher Impulsantwort, FIR Filter

Ein Abtastwert des gefilterten Signals $\check{s}(t_n)$ wird aus den K vergangenen Werten des Eingangssignals $s(t_{n-k})$ berechnet:

$$\check{s}(t_n) = \sum_{k=0}^{K} a_k s(t_{n-k})$$

a_k sind die (konstanten) *Filterkoeffizienten* (filter coefficients); $K+1$ ist die *Filterordnung*, z.B.

$$\begin{aligned}
\check{s}(t_n) &= 0.5s(t_n) + 0.2s(t_{n-1}) - 0.5s(t_{n-2}) \\
a_0 &= 0.5 \\
a_1 &= 0.2 \\
a_2 &= -0.5 \\
\text{Filterordnung} &= 3
\end{aligned}$$

Filter mit endlicher Impulsantwort, FIR Filter

Wir erkennen sofort:

Filterformel	Faltungsformel
$\check{s}(t_n) = \sum a_k s(t_{n-k})$	$g(t_n) = \sum a(t_k) b(t_{n-k})$

Die Filterformel ist eine Faltung von Signal $s(t_n)$ mit einem sehr kurzen Signal $a(t_k)$

Filter mit endlicher Impulsantwort, FIR Filter

Sei das Eingangssignal ein Impuls bei $t = 0$

$$s(t_n) = 1, 0, 0, \text{...} \text{ für } n = 0 \ldots \infty$$

dann berechnet sich das Ausgangssignal $\check{s}(t_n)$ zu

$$\begin{aligned}
\check{s}(t_0) &= 0.5 \cdot 1 + 0.2 \cdot 0 - 0.5 \cdot 0 = 0.5 \\
\check{s}(t_1) &= 0.5 \cdot 0 + 0.2 \cdot 1 - 0.5 \cdot 0 = 0.2 \\
\check{s}(t_2) &= 0.5 \cdot 0 + 0.2 \cdot 0 - 0.5 \cdot 1 = -0.5 \\
\check{s}(t_3) &= 0.5 \cdot 0 + 0.2 \cdot 0 - 0.5 \cdot 0 = 0 \\
\check{s}(t_4) &= 0.5 \cdot 0 + 0.2 \cdot 0 - 0.5 \cdot 0 = 0 \\
&\ldots
\end{aligned}$$

$\rightarrow$ die Impulsantwort ist gleich den Filterkoeffizienten a_k

Filter mit endlicher Impulsantwort, FIR Filter

Sei das Eingangssignal ein Impuls bei $t = 0$

$$s(t_n) = 1, 0, 0, \text{...} \text{ für } n = 0 \ldots \infty$$

dann berechnet sich das Ausgangssignal $\check{s}(t_n)$ zu

$$\left.\begin{aligned}
\check{s}(t_0) &= 0.5 \cdot 1 + 0.2 \cdot 0 - 0.5 \cdot 0 = 0.5 \\
\check{s}(t_1) &= 0.5 \cdot 0 + 0.2 \cdot 1 - 0.5 \cdot 0 = 0.2 \\
\check{s}(t_2) &= 0.5 \cdot 0 + 0.2 \cdot 0 - 0.5 \cdot 1 = -0.5 \\
\check{s}(t_3) &= 0.5 \cdot 0 + 0.2 \cdot 0 - 0.5 \cdot 0 = 0 \\
\check{s}(t_4) &= 0.5 \cdot 0 + 0.2 \cdot 0 - 0.5 \cdot 0 = 0 \\
&\ldots
\end{aligned}\right\} a_k$$

Filtern heißt also: *Faltung mit der Impulsantwort* a_k

Filter mit unendlicher Impulsantwort, IIR Filter

Ein Abtastwert des gefilterten Signals $\check{s}(t_n)$ wird aus den K vergangenen Werten des Eingangssignals $s(t_{n-k})$ *und aus den M vergangenen bereits berechneten Ausgangswerten* $\check{s}(t_{n-m})$ berechnet. Dadurch entsteht eine *rekursive* Rechenvorschrift:

$$\check{s}(t_n) = \sum_{k=0}^{K} a_k s(t_{n-k}) + \sum_{m=1}^{M} b_m \check{s}(t_{n-m})$$

a_k und b_m sind die *Filterkoeffizienten*; die *Filterordnung* ist das Maximum aus $K+1$ und M, z.B.

$$\begin{aligned}
\check{s}(t_n) &= 0.5s(t_n) + 0.2s(t_{n-1}) - 0.5s(t_{n-2}) + 0.1\check{s}(t_{n-1}) \\
a_0 &= 0.5 \\
a_1 &= 0.2 \\
a_2 &= -0.5 \\
b_1 &= 0.1 \qquad \text{Filterordnung} = 3
\end{aligned}$$

Filter mit unendlicher Impulsantwort, IIR Filter

Beachte: Der rekursive Teil der IIR-Formel beginnt nie bei $m = 0$.
(Warum?)

Rekursivität $\rightarrow$ IIR-Filter haben theoretisch eine unendlich lange Impulsantwort
$\rightarrow$ IIR-Filter können eine anwachsende (selbstverstärkende) Impulsantwort haben
$\rightarrow$ *instabiles Filterverhalten*

Das Design von IIR-Filtern ist schwieriger als das von FIR-Filtern.
IIR-Filter sind mächtiger als FIR-Filter, d.h. sie haben mehr Möglichkeiten, Frequenz und Phasen zu beeinflussen.

Design von Digitalen Filtern

Viele Verfahren zum Filter-Design
(An der TUM gibt es einen Lehrstuhl nur für dieses Thema.)

Einfachste Methode:

- Entwurf der gewünschten *Übertragungsfunktion* des Filters (transfer function)
- Ermitteln der Impulsantwort aus der Übertragungsfunktion durch Fouriertransformation
(*siehe nächster Abschnitt 'Spektrum'*)
- Verwendung der Impulsantwort als Filterkoeffizienten in einem FIR-Filter

Demo

Tutoriat: Filter in praat
Filter / Filter (Hann pass band)

Tiefpass bei 2000Hz

`From: 0Hz`
`To: 2000Hz`

Hochpass bei 500Hz

`From: 500Hz`
`To: 20000Hz`

Demo

Tutoriat: Pass Filter with praat

Telefonqualität = Bandpass von 300-3300Hz

`From: 330Hz`

`To: 3300Hz`

Fragen

Wieviele Additionen muss mein Rechner durchführen, wenn er zwei Signale der Länge 2sec und einer Abtastrate von 8kHz addieren will?

Wenn man einen Impuls (1 bei $t = 0$*, 0 sonst) mit der Rechteckfunktion faltet, was ist dann das Ergebnis?*

Mit einer Langzeitanalyse ermitteln wir die Grundfrequenzen einer Dialektgruppe (männliche und weibliche Sprecher). Was muss bei der Mittelung dieser Werte beachtet werden?

Sie haben bei Aldi ein billiges IIR-Filter erstanden und stellen zu Hause fest, dass es leider instabil ist.
Wenn Sie nur ein Nullsignal auf den Eingang des Filter geben, wird es dann auch instabil?